



Contents lists available at SciVerse ScienceDirect

Journal of Multivariate Analysis

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

Regression when both response and predictor are functions

F. Ferraty^a, I. Van Keilegom^b, P. Vieu^{c,*}^a Institut de Mathématiques de Toulouse, Université de Toulouse, 31062 Toulouse cedex 9, France^b Institute of Statistics, Université catholique de Louvain, B-1348 Louvain-la-Neuve, Belgium^c Institut de Mathématiques de Toulouse, Université Paul Sabatier, 31062 Toulouse cedex 9, France

ARTICLE INFO

Article history:

Received 27 May 2011

Available online 1 March 2012

AMS subject classifications:

62G10

62G20

Keywords:

Asymptotic normality

Functional data

Functional response

Kernel estimator

Naive bootstrap

Nonparametric regression

Wild bootstrap

ABSTRACT

We consider a nonparametric regression model where the response \mathcal{Y} and the covariate \mathcal{X} are both functional (i.e. valued in some infinite-dimensional space). We define a kernel type estimator of the regression operator and we first establish its pointwise asymptotic normality. The double functional feature of the problem makes the formulas of the asymptotic bias and variance even harder to estimate than in more standard regression settings, and we propose to overcome this difficulty by using resampling ideas. Both a naive and a wild componentwise bootstrap procedure are studied, and their asymptotic validity is proved. These results are also extended to data-driven bases which is a key point for implementing this methodology. The theoretical advances are completed by some simulation studies showing both the practical feasibility of the method and the good behavior for finite sample sizes of the kernel estimator and of the bootstrap procedures to build functional pseudo-confidence area.

© 2012 Elsevier Inc. All rights reserved.

1. Introduction

The familiar nonparametric regression model can be written as follows:

$$\mathcal{Y} = r(\mathcal{X}) + \varepsilon, \quad (1)$$

where \mathcal{Y} is a response variable, \mathcal{X} is a covariate and the error ε satisfies $E(\varepsilon|\mathcal{X}) = 0$. The nonparametric feature of the problem comes from the fact that the only restrictions on r are smoothness restrictions. In the past four decades, the literature on this kind of models has been impressively large but mostly restricted to the standard multivariate situation where both \mathcal{X} and \mathcal{Y} are real or multivariate. On the other hand, recent technological advances on collecting and storing data have put statisticians in front of situations where the datasets are of functional nature (curves, images, etc.) with the need to develop new models and methods (or for adapting standard ones) to this new kind of data. This field of research, known as Functional Data Analysis (FDA) has been popularized by Ramsay and Silverman [25,26] and the first advances in nonparametric FDA are described in [16] (see also the recent Oxford Handbook of FDA by Ferraty and Romain [14]). In a natural way, the regression problem (1) took part in the interest for nonparametric FDA and has been the object of various studies in the past decade. However, as pointed out in the recent bibliographical discussion by Ferraty and Vieu [17] the existing literature is concentrated on the situation where the response variable \mathcal{Y} is scalar (i.e. when \mathcal{Y} takes values on the real line \mathbb{R}).

* Corresponding author.

E-mail addresses: ferraty@math.univ-toulouse.fr (F. Ferraty), ingrid.vankeilegom@uclouvain.be (I. Van Keilegom), vieu@math.univ-toulouse.fr, vieu@cict.fr (P. Vieu).

When both the response \mathcal{Y} and the explanatory variable \mathcal{X} are of functional nature, mostly functional linear aspects have been developed, as can be seen for instance in [10,7] or in the bibliographical work of Chiou et al. [5] and the references therein (see however two recent theoretical works on nonparametric methods by Ferraty et al., [12] and [21]). Several real examples have been studied in the literature in order to emphasize the usefulness of such a linear functional approach. For instance, in the precursor work of Ramsay and Dalzell [24], a functional linear regression of average monthly precipitation curves on average monthly temperature curves is proposed. Müller et al. [23] show the ability of the functional linear regression to emphasize the relationships that exist between temporary gene expression profiles for different *Drosophila* (fly specy) life cycle phases. In the setting of functional time series, the reader can find in [3] lots of theoretical developments whereas Antoch et al. [1] applied the functional linear model to forecast electricity consumption curves.

Our contribution in this paper is to provide various advances in nonparametric regression when both the response \mathcal{Y} and the explanatory variable \mathcal{X} are of functional nature. When from an asymptotic point of view this paper completes the recent theoretical advances presented by Ferraty et al. [12] and [21], our contribution is the first one to develop methodological background for dealing with computational and applied issues. The mathematical background for modeling such infinite dimensional settings is stated in Section 2. Then, the nonparametric model and its associated kernel estimator are constructed in Section 3. The asymptotic behavior of the procedure is studied in Section 4 by means of an asymptotic normality result. As it is very often the case in nonparametric high dimensional problems, the parameters of such an asymptotic distribution of the estimator are very complicated and hardly usable in practice. To overcome this difficulty, a componentwise bootstrap method is introduced in Section 5 in order to approximate this theoretical asymptotic distribution by an empirical easily usable one. This componentwise bootstrap procedure needs to consider some basis and Section 6 will give some results allowing to use data-driven bases. The feasibility of the whole procedure will be illustrated in Section 7 by means of some Monte Carlo experiments. In Section 8, some general conclusions will be drawn. Finally, the Appendix contains the proof of the main theoretical results.

2. The functional background

Let \mathcal{E} be a functional space, \mathcal{X} be a random variable suitably defined on the measurable space $(\mathcal{E}, \mathcal{A})$, and let the hilbertian random variable \mathcal{Y} live in a measurable separable Hilbert space $(\mathcal{H}, \mathcal{B})$ with \mathcal{H} endowed with inner product $\langle \cdot, \cdot \rangle$ and corresponding norm $\| \cdot \|$ (i.e. $\|g\|^2 = \langle g, g \rangle$), and with orthonormal basis $\{e_j : j = 1, \dots, \infty\}$. Moreover, \mathcal{E} is endowed with a semi-metric¹ $d(\cdot, \cdot)$ defining a topology to measure the proximity between two elements of \mathcal{E} and which is disconnected of the definition of \mathcal{X} in order to avoid measurability problems. This setting is quite general, since it includes the classical finite dimensional framework, but also the case where \mathcal{E} and/or \mathcal{H} are functional spaces, like the space of continuous functions: L^p -spaces like Sobolev or Besov spaces. At this stage it is worth noticing that this kind of modelization includes the setting that occurs quite often in practice (but, of course, not always) when $\mathcal{E} = \mathcal{H}$. However, even in this simpler situation, one still needs to introduce two different topological structures: a semi-metric structure d that will be a key tool for controlling the good behavior of the estimators and also some separable Hilbert structure which is necessary for studying the operator r component by component. Because the semi-metric d is not a metric in most cases, it is not possible to derive a norm and a corresponding inner product to define correctly any Hilbert space. This is why these two structures are needed even if \mathcal{X} and \mathcal{Y} live in the same space. So, we decided to present our results in the general framework where \mathcal{E} is not necessarily equal to \mathcal{H} .

Each of these two structures endows the corresponding space with some specific topology. Concerning the semi-metric topology defined on \mathcal{E} we will use the notation

$$B_{\mathcal{E}}(\chi, t) = \{\chi_1 \in \mathcal{E} : d(\chi_1, \chi) \leq t\}$$

for the ball in \mathcal{E} with center χ and radius t , while in the Hilbert space \mathcal{H} we will denote

$$B_{\mathcal{H}}(z, r) = \{y \in \mathcal{H} : \|z - y\| \leq r\}$$

for the ball in \mathcal{H} with center z and radius r .

In the sequel, we will also need the following notations. We denote

$$F_{\mathcal{X}}(t) = P(d(\mathcal{X}, \chi) \leq t) = P(\mathcal{X} \in B(\chi, t)),$$

which is usually called in the literature the small ball probability function when t is a decreasing sequence to zero. Further, define for any $k \geq 1$:

$$\varphi_{\mathcal{X},k}(s) = E[\langle r(\mathcal{X}) - r(\chi), e_k \rangle \mid d(\mathcal{X}, \chi) = s] = E[r_k(\mathcal{X}) - r_k(\chi) \mid d(\mathcal{X}, \chi) = s],$$

where

$$r_k(\chi) = E[\langle \mathcal{Y}, e_k \rangle \mid \mathcal{X} = \chi] = \langle r(\chi), e_k \rangle,$$

¹ A semi-metric (sometimes called pseudo-metric) $d(\cdot, \cdot)$ is a metric which allows $d(x_1, x_2) = 0$ for some $x_1 \neq x_2$.

and define for any $k, \ell \geq 1$:

$$\psi_{\chi, k\ell}(s) = E[s_{k\ell}(\mathcal{X}) - s_{k\ell}(\chi) \mid d(\mathcal{X}, \chi) = s],$$

where

$$s_{k\ell}(\chi) = \text{Cov}[\langle \mathcal{Y}, e_k \rangle, \langle \mathcal{Y}, e_\ell \rangle \mid \mathcal{X} = \chi].$$

Also, let

$$\tau_{h\chi}(s) = F_\chi(hs)/F_\chi(h) = P(d(\mathcal{X}, \chi) \leq hs \mid d(\mathcal{X}, \chi) \leq h) \quad \text{for } 0 < s \leq 1,$$

and

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

Remark 2.1. It is worth noting that the semi-metric plays a major role in this kind of methodology. This is why it is crucial to fix a well adapted semi-metric. The reader will find in [16] useful discussions and guidelines about the way of choosing a semi-metric dealing with practical aspects (see Chapter 3) as well as theoretical purposes (see Chapter 13).

3. Construction of the estimator

Let $\delta = \{(\mathcal{X}_1, \mathcal{Y}_1), \dots, (\mathcal{X}_n, \mathcal{Y}_n)\}$ be a sample of i.i.d. data drawn from the distribution of the pair $(\mathcal{X}, \mathcal{Y})$. The estimator of the regression operator r is given by

$$\widehat{r}_h(\chi) = \frac{\sum_{i=1}^n \mathcal{Y}_i K(h^{-1}d(\mathcal{X}_i, \chi))}{\sum_{i=1}^n K(h^{-1}d(\mathcal{X}_i, \chi))},$$

where χ is a fixed element of \mathcal{E} . Here, K is a probability density function (kernel) and h is a bandwidth sequence, tending to zero when n tends to infinity. This estimator is a functional version of the familiar Nadaraya–Watson estimator, and has been recently introduced for functional covariates (but for scalar response \mathcal{Y}) in [16]. Basically, this estimator is an average of the observed response \mathcal{Y}_i for which the corresponding \mathcal{X}_i is close to the new functional element χ at which the operator r has to be estimated. The size of the neighborhood around χ that will be used for the estimation of $r(\chi)$ is controlled by the smoothing parameter h . The shape of the neighborhood is linked to the topological structure introduced on the space \mathcal{E} ; in other words the semi-metric d will play a major role in the behavior of the estimator. The semi-metric d will act on the asymptotic behavior of $\widehat{r}_h(\chi)$ through the functions F_χ and $\tau_{0\chi}$ and more specifically through the following quantities:

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

$$M_{1\chi} = K(1) - \int_0^1 K'(s) \tau_{0\chi}(s) ds,$$

and

$$M_{2\chi} = K^2(1) - \int_0^1 (K^2)'(s) \tau_{0\chi}(s) ds.$$

4. Asymptotic normality of $\widehat{r}_h(\chi)$

From now on, χ is a fixed functional element in the space \mathcal{E} . Consider the following assumptions:

- (C1) For each $k, \ell \geq 1$, $r_k(\cdot)$ and $s_{k\ell}(\cdot)$ are continuous in a neighborhood of χ , and $F_\chi(0) = 0$.
- (C2) For some $\delta > 0$, all $0 \leq s \leq \delta$ and all $k \geq 1$, $\varphi_{\chi, k}(0) = 0$, $\varphi'_{\chi, k}(s)$ exists, and $\varphi'_{\chi, k}(s)$ is uniformly Lipschitz continuous of order $0 < \alpha \leq 1$, i.e. there exists a $0 < L_k < \infty$ such that $|\varphi'_{\chi, k}(s) - \varphi'_{\chi, k}(0)| \leq L_k s^\alpha$ uniformly for all $0 \leq s \leq \delta$. Moreover, $\sum_{k=1}^{\infty} L_k^2 < \infty$ and $\sum_{k=1}^{\infty} \varphi'_{\chi, k}(0)^2 < \infty$.
- (C3) For some $\delta > 0$, all $0 \leq s \leq \delta$ and all $k \geq 1$, $\psi_{\chi, kk}(0) = 0$, and $\psi_{\chi, kk}(s)$ is uniformly Lipschitz continuous of order $0 < \beta \leq 1$, i.e. there exists a $0 < N_k < \infty$ such that $|\psi_{\chi, kk}(s)| \leq N_k s^\beta$ uniformly for all $0 \leq s \leq \delta$. Moreover, $\sum_{k=1}^{\infty} N_k < \infty$.
- (C4) The bandwidth h satisfies $h \rightarrow 0$, $nF_\chi(h) \rightarrow \infty$ and $(nF_\chi(h))^{1/2} h^{1+\alpha} = o(1)$.
- (C5) The kernel K is supported on $[0, 1]$, K has a continuous derivative on $[0, 1]$, $K'(s) \leq 0$ for $0 \leq s < 1$ and $K(1) > 0$.
- (C6) For all $0 \leq s \leq 1$, $\tau_{0\chi}(s)$ exists, $\sup_{0 \leq s \leq 1} |\tau_{h\chi}(s) - \tau_{0\chi}(s)| = o(1)$, $M_{0\chi} > 0$ and $M_{1\chi} > 0$.

(C7) $\text{Var}(\|\mathcal{Y}\| \mid \mathcal{X} = \chi) < \infty$, and for all $k \geq 1$, $E(\langle \mathcal{Y} - r(\mathcal{X}), e_k \rangle^4 \mid \mathcal{X} = \chi) < \infty$.

Let

$$\widehat{g}(\chi) = (nF_\chi(h))^{-1} \sum_{i=1}^n \mathcal{Y}_i K\left(\frac{d(\mathcal{X}_i, \chi)}{h}\right)$$

and

$$\widehat{f}(\chi) = (nF_\chi(h))^{-1} \sum_{i=1}^n K\left(\frac{d(\mathcal{X}_i, \chi)}{h}\right),$$

and note that $\widehat{r}(\chi) = \widehat{g}(\chi)/\widehat{f}(\chi)$. Define

$$\mathcal{B}_{n\chi} = h \frac{M_{0\chi}}{M_{1\chi}} \sum_{k=1}^{\infty} \varphi'_{\chi,k}(0) e_k$$

and let C_χ be the operator characterized by

$$\langle C_\chi f, e_\ell \rangle = \sum_{k=1}^{\infty} \langle f, e_k \rangle a_{k\ell},$$

where

$$a_{k\ell} = \text{Cov}(\langle \mathcal{Y}, e_k \rangle, \langle \mathcal{Y}, e_\ell \rangle \mid \mathcal{X} = \chi) \frac{M_{2\chi}}{M_{1\chi}^2}.$$

We are now ready to state the asymptotic normality of the estimator $\widehat{r}(\chi)$. While there are already various results of this kind when the response is scalar (see for instance [22,13] or [9]), this is the first result on the asymptotic normality in nonparametric kernel regression when both the response and the explanatory variable are functional.

Theorem 4.1. Assume (C1)–(C7). Then, for any $\chi \in \mathcal{E}$,

$$(nF_\chi(h))^{1/2} \left(\widehat{r}(\chi) - r(\chi) - \mathcal{B}_{n\chi} \right) \xrightarrow{\mathcal{L}} \mathcal{W}_\chi,$$

where \mathcal{W}_χ follows a normal distribution on \mathcal{H} with zero mean and covariance operator C_χ .

Remark 4.1. The small ball probability $F_\chi(h)$ plays a major role since it appears in the standardization. In fact, the same standardization is systematically used in all theorems. This is why condition (C4) acting on $F_\chi(h)$ is crucial and deserves some comments. From a probabilistic viewpoint, considering standard processes (i.e. exponential-type) and usual metrics (i.e. derived from the supremum norm, L_2 -norm, etc.) leads us to choose the bandwidth $h = O((\log n)^{-\eta})$ for some $\eta > 0$ in order to satisfy (C4). In this case, one is able to get only a logarithmic rate of convergence for the nonparametric regression operator (i.e. the standard polynomial rate is not reached). However, if one adopts the statistical viewpoint, one has at hand a powerful tuning parameter which is the semi-metric; a relevant choice for the semi-metric may lead to the polynomial rate which reduces the curse of dimensionality impact. For instance, [16, p. 213] showed that any projection-based semi-metric (i.e. a semi-metric derived from a distance between projected elements onto some finite-dimensional subspace) allow us to get the polynomial rate. The data-driven semi-metric $d_j(\cdot, \cdot)$ involved in the simulation study (see Section 7) is a special case of such useful projection-based semi-metric. Of course, other relevant families of semi-metrics can be interestingly used in practice (see again [16] or [27]).

5. Componentwise bootstrap approximation

Both naive and wild bootstrap procedures have been successfully used in the literature to approximate the asymptotic distribution in functional regression. The most recent result was provided by Ferraty et al. [15] when the explanatory variable is functional and the response is real. In the following we propose an extension of these bootstrap procedures to the new situation studied here, i.e. when both variables are functional. The idea is to state that, for any fixed basis element e_k , when one projects onto e_k , the bootstrap approximation has a good theoretical behavior, which is the aim of Theorem 5.1. This is why one introduces the terminology “componentwise bootstrap”.

Naive bootstrap. We assume here that the model is homoscedastic, i.e. the conditional covariance operator of ε given \mathcal{X} does not depend on \mathcal{X} : for any $g, h \in \mathcal{H}$, $E(\langle \varepsilon, g \rangle \langle \varepsilon, h \rangle \mid \mathcal{X}) = E(\langle \varepsilon, g \rangle \langle \varepsilon, h \rangle)$. The bootstrap procedure consists of several steps:

(1) For all $i = 1, \dots, n$, define $\widehat{\varepsilon}_{i,b} = \mathcal{Y}_i - \widehat{r}_b(\mathcal{X}_i)$, where b is a second smoothing parameter.

(2) Let $\widehat{\varepsilon}_b = n^{-1} \sum_{i=1}^n \widehat{\varepsilon}_{i,b}$. Define new i.i.d. random elements $\varepsilon_1^{\text{boot}}, \dots, \varepsilon_n^{\text{boot}}$ by:

$$P^\delta \left(\varepsilon_i^{\text{boot}} = \widehat{\varepsilon}_{j,b} - \widehat{\varepsilon}_b \right) = n^{-1}$$

for $i, j = 1, \dots, n$, where P^δ is the probability conditionally on the original sample $(\mathcal{X}_1, \mathcal{Y}_1), \dots, (\mathcal{X}_n, \mathcal{Y}_n)$.

(3) For $i = 1, \dots, n$, let

$$\mathcal{Y}_i^{\text{boot}} = \widehat{r}_b(\mathcal{X}_i) + \varepsilon_i^{\text{boot}},$$

and denote $\mathcal{S}^{\text{boot}} = (\mathcal{X}_i, \mathcal{Y}_i^{\text{boot}})_{i=1, \dots, n}$.

(4) Define $\widehat{r}_{hb}^{\text{boot}}(\chi) = \frac{\sum_{i=1}^n \mathcal{Y}_i^{\text{boot}} K(h^{-1}d(\mathcal{X}_i, \chi))}{\sum_{i=1}^n K(h^{-1}d(\mathcal{X}_i, \chi))}$.

Wild bootstrap. We assume here that the model can be heteroscedastic. With respect to the naive bootstrap, we need to change the second step: define $\varepsilon_i^{\text{boot}} = \widehat{\varepsilon}_{i,b} V_i$, where V_1, \dots, V_n are i.i.d. real valued random variables that are independent of the data $(\mathcal{X}_i, \mathcal{Y}_i)$ ($i = 1, \dots, n$) and that satisfy $E(V_1) = 0$ and $E(V_1^2) = 1$.

Before stating the next theorem which is a direct consequence of Theorem 1 in [11,15], let us first enumerate briefly some additional assumptions (where α is defined in (C2)):

- (i) For any $k = 1, 2, \dots$, $E(|\langle \mathcal{Y}, e_k \rangle| | \mathcal{X} = \cdot)$ is continuous in a neighborhood of χ and $\sup_{d(\mathcal{X}_1, \chi) < \epsilon} E(|\langle \mathcal{Y}, e_k \rangle|^m | \mathcal{X} = \chi_1) < +\infty$ for some $\epsilon > 0$ and for all $m \geq 1$.
- (ii) $F_{\chi_1}(t)/F_\chi(t)$ is Lipschitz continuous of order α in χ_1 , uniformly in t in a neighborhood of 0.
- (iii) $\forall \chi_1 \in \mathcal{E}, \forall 0 \leq s \leq 1, \tau_{0\chi_1}$ exists and $\sup_{\chi_1 \in \mathcal{E}, 0 \leq s \leq 1} |\tau_{h\chi_1}(s) - \tau_{0\chi_1}(s)| = o(1)$; $M_{0\chi} > 0, M_{2\chi} > 0, \inf_{d(\mathcal{X}_1, \chi) < \epsilon} M_{1\chi_1} > 0$ for some $\epsilon > 0$, and $M_{j\chi_1}$ is Lipschitz continuous of order α for $j = 0, 1, 2$.
- (iv) $h, b \rightarrow 0, h/b \rightarrow 0, h(F_\chi(h))^{1/2} = O(1), b^{1+\alpha}(nF_\chi(h))^{1/2} = o(1), F_\chi(b+h)/F_\chi(b) \rightarrow 1, (F_\chi(h)/F_\chi(b)) \log n = o(1)$ and $bh^{\alpha-1} = O(1)$.
- (v) For each n , there exist $r_n \geq 1, l_n > 0$ and t_{1n}, \dots, t_{rn} such that $B_\mathcal{E}(\chi, h) \subset \cup_{k=1}^n B_\mathcal{E}(t_{kn}, l_n)$, with $r_n = O(n^{b/h})$ and $l_n = o(b(nF_\chi(h))^{-1/2})$.

More details about these assumptions can be found in [11,15].

Theorem 5.1. For any $k = 1, 2, \dots$, and any bandwidths h and b , let $\widehat{r}_{k,h}(\chi) = \widehat{r}_h(\chi), e_k$ and $\widehat{r}_{k,hb}^{\text{boot}}(\chi) = \widehat{r}_{hb}^{\text{boot}}(\chi), e_k$. If, in addition to (C1), (C2) and (C5), conditions (i)–(v) hold, then, for the wild bootstrap procedure and for any $k = 1, 2, \dots$, we have:

$$\sup_{y \in \mathbb{R}} \left| P^\delta \left((nF_\chi(h))^{1/2} \left\{ \widehat{r}_{k,hb}^{\text{boot}}(\chi) - \widehat{r}_{k,b}(\chi) \right\} \leq y \right) - P \left((nF_\chi(h))^{1/2} \left\{ \widehat{r}_{k,h}(\chi) - r_k(\chi) \right\} \leq y \right) \right| \xrightarrow{\text{a.s.}} 0,$$

where P^δ denotes probability, conditionally on the sample \mathcal{S} (i.e. $(\mathcal{X}_i, \mathcal{Y}_i), i = 1, \dots, n$). In addition, if the model is homoscedastic, then the same result holds for the naive bootstrap.

Indeed, for a fixed k , the problem reduces to a one-dimensional response problem, and hence we can directly apply the bootstrap result obtained in that case (Theorem 1 in [11,15]).

6. Data-driven basis

The main result of Section 5 investigates the asymptotic behavior of the componentwise bootstrap errors associated to an orthonormal basis e_1, e_2, \dots . However, in order to implement this bootstrap procedure, it is necessary to determine this orthonormal basis. From a statistical point of view, it is reasonable to implement a data-driven basis i.e. an orthonormal basis estimated from the data $(\mathcal{X}_i, \mathcal{Y}_i), i = 1, \dots, n$.

In the next theorem, we give a third important result allowing the use of data-driven bases. Any generic data-driven basis will be denoted by $\{\widehat{e}_k : k = 1, \dots, \infty\}$. We use the notation $r_{\widehat{e}_k}(\chi)$ to denote the k -th component of the function $r(\chi)$ with respect to this estimated basis, and similarly the estimators $\widehat{r}_{k,h}(\chi)$ and $\widehat{r}_{k,hb}^{\text{boot}}(\chi)$ are defined. We show below that Theorem 5.1 remains valid when the basis $\{\widehat{e}_k : k = 1, \dots, \infty\}$ is employed.

Theorem 6.1. Assume the conditions of Theorem 5.1 hold, and assume in addition that $\|\widehat{e}_k - e_k\| = o(1)$ a.s. and $(h/b)(nF_\chi(h))^{1/2} \|\widehat{e}_k - e_k\| = o(1)$ a.s. for $k = 1, 2, \dots$. Then, for the wild bootstrap procedure, we have:

$$\sup_{y \in \mathbb{R}} \left| P^\delta \left((nF_\chi(h))^{1/2} \left\{ \widehat{r}_{k,hb}^{\text{boot}}(\chi) - \widehat{r}_{k,b}(\chi) \right\} \leq y \right) - P \left((nF_\chi(h))^{1/2} \left\{ \widehat{r}_{k,h}(\chi) - r_{\widehat{e}_k}(\chi) \right\} \leq y \right) \right| \xrightarrow{\text{a.s.}} 0.$$

In addition, if the model is homoscedastic, then the same result holds for the naive bootstrap.

Let us now focus on a natural way of building an orthonormal basis from the sample in order to implement our bootstrap procedure. Inspired by functional principal component analysis (see [8], or [24], among others for early results and for instance, [4,28,6], or [18], for recent contributions where functional principal component analysis plays a major role), we introduce the second order moment regression operator $\Gamma_{r(\mathcal{X})}(\cdot) = E(\langle r(\mathcal{X}), \cdot \rangle r(\mathcal{X}))$ which maps \mathcal{H} onto \mathcal{H} . The orthonormal eigenfunctions e_1, e_2, \dots of $\Gamma_{r(\mathcal{X})}(\cdot)$ are relevant directions for the hilbertian variable $r(\mathcal{X})$. Indeed, for any fixed strictly positive integer K , the eigenfunctions e_1, \dots, e_K associated to the K largest eigenvalues minimize the quantity

$$E \left(\left\| r(\mathcal{X}) - \sum_{k=1}^K \langle r(\mathcal{X}), \psi_k \rangle \psi_k \right\|^2 \right)$$

over any orthonormal sequence ψ_1, \dots, ψ_K . But here, the regression operator $r(\cdot)$ is unknown and we propose now two examples of useful data-driven bases.

Example 1. One can use the functional response \mathcal{Y} instead of $r(\mathcal{X})$. This amounts to consider the orthonormal eigenfunctions $\tilde{e}_1, \tilde{e}_2, \dots$ associated to the eigenvalues $\mu_1 \geq \mu_2 \geq \dots$ of the second order moment operator $\Gamma_{\mathcal{Y}} = E(\langle \mathcal{Y}, \cdot \rangle \mathcal{Y})$ which can be estimated by its empirical version

$$\Gamma_{\mathcal{Y},n}(\cdot) = \frac{1}{n} \sum_{i=1}^n \langle \mathcal{Y}_i, \cdot \rangle \mathcal{Y}_i.$$

The sequence of orthonormal eigenfunctions $\hat{e}_1, \hat{e}_2, \dots$ of $\Gamma_{\mathcal{Y},n}$ is a well known estimate of the orthonormal basis $e_1 = \text{sign}(\langle \hat{e}_1, \tilde{e}_1 \rangle) \tilde{e}_1, e_2 = \text{sign}(\langle \hat{e}_2, \tilde{e}_2 \rangle) \tilde{e}_2, \dots$. However, in order to simplify notations, from now on, we do not distinguish $\hat{e}_1, \hat{e}_2, \dots$ and e_1, e_2, \dots .

Proposition 1. Assume the conditions of Theorem 5.1 hold. Assume in addition that there exists a $M > 0$ such that $E\|\mathcal{Y}\|^{2m} < M m! < \infty$ for all $m \geq 1$, and that the eigenvalues of $\Gamma_{\mathcal{Y}}$ satisfy $\mu_1 > \mu_2 > \dots$. Then, Theorem 6.1 is valid for the empirical orthonormal eigenfunctions $\hat{e}_1, \hat{e}_2, \dots$.

Example 2. Consider now the orthonormal eigenfunctions e_1, e_2, \dots associated to the eigenvalues (in descending order) of $\Gamma_{r(\mathcal{X})}$. If the regression operator $r(\cdot)$ would be known as is the case in the simulations, one gets the same result with the orthonormal eigenfunctions $\Gamma_{r(\mathcal{X}),n}(\cdot) = \frac{1}{n} \sum_{i=1}^n \langle r(\mathcal{X}_i), \cdot \rangle r(\mathcal{X}_i)$ as soon as $r(\mathcal{X})$ is almost surely bounded (i.e. $\|r(\mathcal{X})\| \leq C$ a.s. for some $0 < C < \infty$). When the regression operator $r(\cdot)$ is unknown, which is the usual statistical situation, a more sophisticated way of building a data-driven basis consists in using the eigenfunctions $\hat{e}_1, \hat{e}_2, \dots$ of the estimated second order moment regression operator $\Gamma_{\hat{r}_h(\mathcal{X})} = \frac{1}{n} \sum_{i=1}^n \langle \hat{r}_h(\mathcal{X}_i), \cdot \rangle \hat{r}_h(\mathcal{X}_i)$.

Proposition 2. Assume the conditions of Theorem 5.1 hold. Assume in addition that there exists a $C > 0$ such that $\|r(\mathcal{X})\| \leq C < \infty$ a.s. and that $(nF_{\mathcal{X}}(h))^{1/2} \|\Gamma_{\hat{r}_h(\mathcal{X})} - \Gamma_{r(\mathcal{X}),n}\|_{\infty} = o(b/h)$ a.s. (where $\|U\|_{\infty} = \sup_{\|x\|=1} \|U(x)\|$). Then, Theorem 6.1 is valid for the empirical orthonormal eigenfunctions $\hat{e}_1, \hat{e}_2, \dots$.

Remark 6.1. There exist at least two ways of studying the quantity $\|\Gamma_{\hat{r}_h(\mathcal{X})} - \Gamma_{r(\mathcal{X}),n}\|_{\infty}$, which are still open questions. They use the following decomposition:

$$\begin{aligned} \Gamma_{\hat{r}_h(\mathcal{X})} - \Gamma_{r(\mathcal{X}),n} &= \frac{1}{n} \sum_{i=1}^n \left(\langle \hat{r}_h(\mathcal{X}_i), \cdot \rangle \hat{r}_h(\mathcal{X}_i) - \langle r(\mathcal{X}_i), \cdot \rangle r(\mathcal{X}_i) \right) \\ &= \frac{1}{n} \sum_{i=1}^n \left(\langle \hat{r}_h(\mathcal{X}_i) - r(\mathcal{X}_i), \cdot \rangle (\hat{r}_h(\mathcal{X}_i) - r(\mathcal{X}_i)) + \langle r(\mathcal{X}_i), \cdot \rangle (\hat{r}_h(\mathcal{X}_i) - r(\mathcal{X}_i)) \right. \\ &\quad \left. + \langle \hat{r}_h(\mathcal{X}_i) - r(\mathcal{X}_i), \cdot \rangle r(\mathcal{X}_i) \right). \end{aligned}$$

A first way consists of focusing on the following inequality:

$$\|\Gamma_{\hat{r}_h(\mathcal{X})} - \Gamma_{r(\mathcal{X}),n}\|_{\infty} \leq \sup_{x \in S} \|\hat{r}_h(x) - r(x)\|^2 + 2C \sup_{x \in S} \|\hat{r}_h(x) - r(x)\|.$$

Hence, the uniform result of Ferraty et al. [11,15] needs to be extended to the case of hilbertian responses and assumptions need to be added in order that $(nF_{\mathcal{X}}(h))^{1/2} \sup_{x \in S} \|\hat{r}_h(x) - r(x)\| = o(b/h)$ a.s.

A second way consists of investigating the asymptotic properties of the quantity

$$\frac{1}{n} \sum_{i=1}^n \langle r(\mathcal{X}_i), \cdot \rangle (\hat{r}_h(\mathcal{X}_i) - r(\mathcal{X}_i)),$$

which is an average of hilbertian variables. This is not a trivial problem, because $\hat{r}_h(\cdot)$ depends on the whole sample.

Although these are still open problems, one might expect that such results will be stated in the near future.

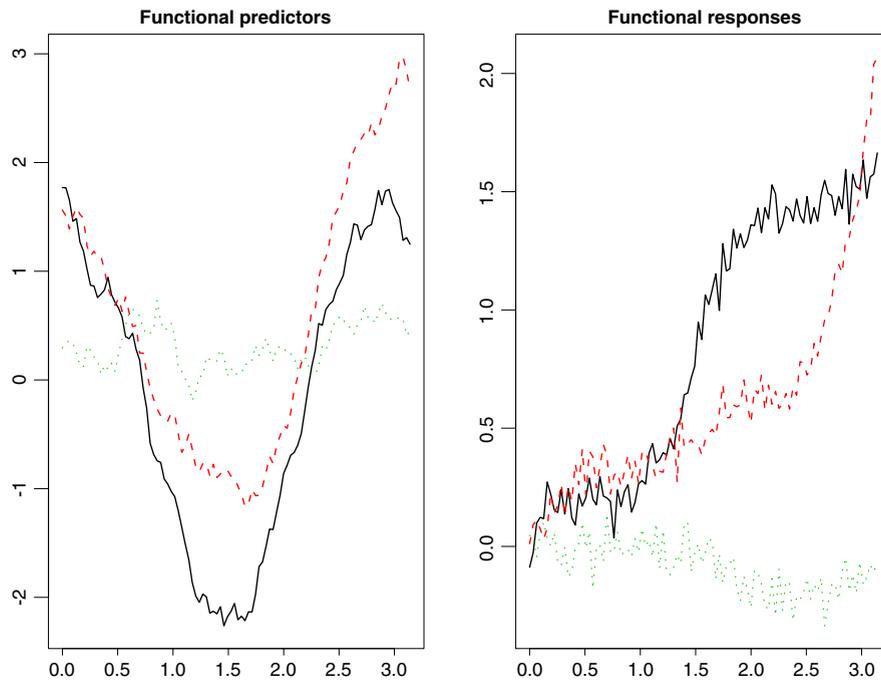


Fig. 1. Left panel: 3 functional predictors; right panel: the 3 corresponding functional responses.

7. Simulations

This section aims at illustrating the potentialities of using the bootstrap method in our double functional setting (functional response and functional explanatory variable). We first detail the simulation of both functional predictors and functional responses. It is worth emphasizing that this is the first time that the nonparametric functional model is used in such a setting. In the second part, we focus on the ability of the nonparametric functional regression to predict functional responses from functional predictors. The third part illustrates the bootstrap methodology and we will see how the theoretical results parallel the practical experiment. A last part is devoted to building and visualizing functional pseudo-confidence areas, which is a very interesting new tool for assessing the accuracy of predictions when both response and regressor are functional. All R routines as well as command lines for reproducing simulations, figures, etc., are downloadable at <http://www.math.univ-toulouse.fr/staph/npfd> (click on the “Additional methods” item in the left menu).

Simulating functional responses and functional predictors. Let $\mathcal{X}_1, \dots, \mathcal{X}_n$ be $n = 250$ functional predictors such that

$$\mathcal{X}_i(t_j) = a_i \cos(\omega_i t_j) + \sum_{k=1}^j W_{ik},$$

where a_1, \dots, a_n (resp. $\omega_1, \dots, \omega_n$) are n independent real random variables (r.r.v.) uniformly distributed over $[1.5; 2.5]$ (resp. $[0.2; 2]$), $0 = t_1 < t_2 < \dots < t_{99} < t_p = \pi$ are $p = 100$ equally spaced measurements and the W_{ik} 's are i.i.d. realizations of $N(0, \sigma^2)$ with $\sigma^2 = 0.01$. The additional sum of Gaussian r.r.v. is just to make the functional predictor quite rough. The left panel of Fig. 1 displays 3 functional predictors ($\mathcal{X}_1, \mathcal{X}_2$ and \mathcal{X}_3). The right panel plots the corresponding responses by using a mechanism described later on. The regression operator $r(\cdot)$ is defined such that, for all $i = 1, \dots, n$ and $j = 1, \dots, 100$:

$$r(\mathcal{X}_i)(t_j) = \int_0^{t_j} \mathcal{X}_i(u)^2 du.$$

The building of the functional response \mathcal{Y} follows the following scheme for all $i = 1, \dots, n$ and $j = 1, \dots, 100$:

$$\mathcal{Y}_i(t_j) = r(\mathcal{X}_i)(t_j) + \varepsilon_i(t_j),$$

and we discuss now the way of simulating the errors $\varepsilon_i(t_j)$'s. One proposes to use the richness of the functional responses setting to mix two kinds of errors: standard additive noise and structural perturbation. The standard additive noise is defined as

$$\varepsilon_i^{\text{add}}(t_j) \sim N\left(\mathbf{0}, \sqrt{\text{snr}} \times \text{tr}(\Gamma_{r_c(\mathcal{X}, n)}^*)\right),$$

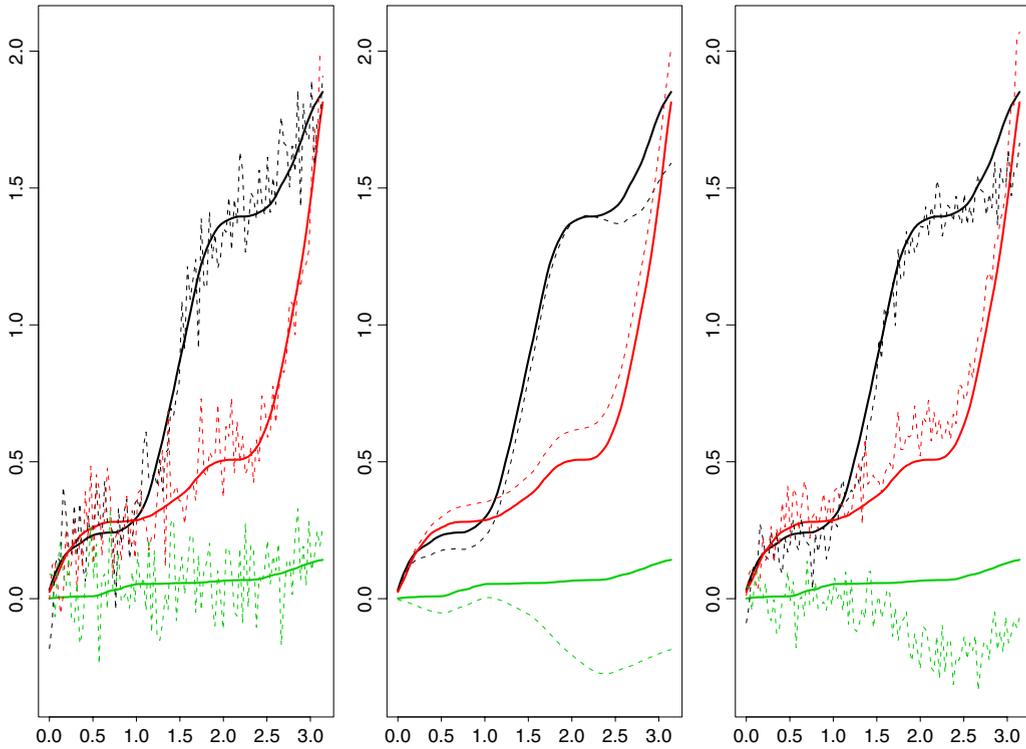


Fig. 2. (a) Additive error; (b) structural error; (c) additive + structural error. For each panel, we plot $r(\mathcal{X}_1)$, $r(\mathcal{X}_2)$, $r(\mathcal{X}_3)$ (solid lines) and \hat{y}_1 , \hat{y}_2 , \hat{y}_3 (dashed lines).

where $\Gamma_{r_c(\mathcal{X}),n} = 1/n \sum_{i=1}^n \langle r_c(\mathcal{X}_i), \cdot \rangle r_c(\mathcal{X}_i)$ is the empirical covariance operator of $r(\mathcal{X})$ (with $r_c(\mathcal{X}_i) = r(\mathcal{X}_i) - (1/n) \sum_{i=1}^n r(\mathcal{X}_i)$), $\text{tr}(\cdot)$ stands for the standard trace operator and snr is the signal-to-noise ratio. Let $\Gamma_n^{\text{add}} = \frac{1}{n} \sum_{i=1}^n \langle \varepsilon_i^{\text{add}}, \cdot \rangle \varepsilon_i^{\text{add}}$ be the empirical covariance operator of the additive error. Then one has:

$$\text{tr}(\Gamma_n^{\text{add}}) = snr \times \text{tr}(\Gamma_{r_c(\mathcal{X}),n}).$$

So, snr controls the ratio between the amount of variance in ε^{add} (i.e. $\text{tr}(\Gamma_n^{\text{add}})$) and in $r(\mathcal{X})$ (i.e. $\text{tr}(\Gamma_{r_c(\mathcal{X}),n})$). This is why snr is termed signal-to-noise ratio.

Another way of perturbing the regression operator $r(\cdot)$ is to use the eigenvalues $\lambda_{1,n} > \lambda_{2,n} > \dots$ and corresponding eigenfunctions $e_{1,n}, e_{2,n}, \dots$ of $\Gamma_{r(\mathcal{X}),n}$ to build what one calls structural errors as follows:

$$\varepsilon_i^{\text{struct}}(t_j) = \sum_{k=1}^p \eta_{ik} e_{k,n}(t_j),$$

where, for $k = 1, 2, \dots$, the r.r.v. $\eta_{1k}, \dots, \eta_{nk}$ are independent and identically distributed as $N(0, \sqrt{snr \times \lambda_{k,n}})$. It is easy to check that

$$\text{tr}(\Gamma_n^{\text{struct}}) = snr \times \text{tr}(\Gamma_{r(\mathcal{X}),n}),$$

where Γ_n^{struct} is the covariance operator of the structural error.

Now, the last step consists in building a third error by mixing both previous ones:

$$\varepsilon_i^{\text{mix}}(t_j) = \sqrt{\rho} \varepsilon_i^{\text{add}}(t_j) + \sqrt{1 - \rho} \varepsilon_i^{\text{struct}}(t_j),$$

with $\rho \in (0, 1)$ and the covariance operator Γ_n^{mix} of the mixed error satisfies:

$$\text{tr}(\Gamma_n^{\text{mix}}) = snr \times \text{tr}(\Gamma_{r_c(\mathcal{X}),n}).$$

Fig. 2 gives an idea of how the error-type affects the regression. Finally, in our simulations, we use the mixed error with $snr = 5\%$ and $\rho = 0.3$. In order to validate the theoretical property of our bootstrap methodology, for both next paragraphs (focusing on implementation and bootstrap aspects), one considers the situation described at the beginning of **Example 2** (Section 6). This corresponds to the case when the regression operator r is known: e_1, e_2, \dots (resp. $\hat{e}_1, \hat{e}_2, \dots$) are the orthonormal eigenfunctions associated to the eigenvalues (in descending order) of $\Gamma_{r_c(\mathcal{X})}$ (resp. $\Gamma_{r_c(\mathcal{X}),n}$). As a by-product, a third

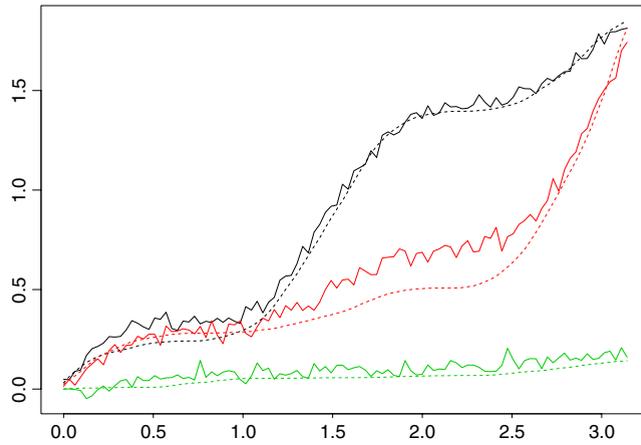


Fig. 3. 3 predicted functional responses (solid lines); corresponding observed functional responses (dashed lines).

paragraph proposes an heuristic and practical way of building functional pseudo-confidence areas where the considered data-driven basis $\widehat{e}_1, \widehat{e}_2, \dots$ is derived from the covariance operator $\Gamma_{\widehat{r}_c(\mathcal{X})}$ where $\widehat{r}_c(\mathcal{X}) = \widehat{r}_h(\mathcal{X}) - 1/n \sum_{i=1}^n \widehat{r}_h(\mathcal{X}_i)$.

Implementing functional nonparametric regression. Recall that this is the first time that one implements nonparametric regression when both predictor and response are functional variables. So, before going ahead with the bootstrap method, it is important to emphasize the quality of prediction reached by our nonparametric regression in this “double” functional setting. The first step is to fix the semi-metric $d(\cdot, \cdot)$. According to the simulated data, the projection-based semi-metric is a good candidate:

$$d_j(\chi_1, \chi_2) = \sqrt{\sum_{j=1}^J \langle \chi_1 - \chi_2, v_{j,n} \rangle^2},$$

where $v_{1,n}, v_{2,n}, \dots$ are the eigenvectors associated with the largest eigenvalues of the empirical covariance operator of the functional predictor \mathcal{X} :

$$\Gamma_{\mathcal{X},n} = 1/200 \sum_{i=1}^{200} \langle \mathcal{X}_i, \cdot \rangle \mathcal{X}_i.$$

This kind of semi-metric is especially well adapted when the functional predictors are rough (for more details about the interest of using semi-metrics, see [16]). The original sample is split into two subsamples: the learning sample (i.e. $(\mathcal{X}_i, \mathcal{Y}_i)_{i=1, \dots, 200}$) and the testing sample (i.e. $(\mathcal{X}_i, \mathcal{Y}_i)_{i=201, \dots, 250}$). The learning sample allows us to compute the kernel estimator (with optimal parameters h and k by using a standard cross-validation procedure). A first way of assessing the quality of prediction is to compare predicted functional responses (i.e. $\widehat{r}_h(\chi)$ for any χ in the testing sample) versus the true regression operator (i.e. $r(\mathcal{X})$) as in Fig. 3. However, if one wishes to assess the quality of prediction for the whole testing sample, it is much better to see what happens direction by direction. In other words, displaying the predictions onto the direction $e_{k,n}$ amounts to plotting the 50 points $(\langle r(\mathcal{X}_i), \widehat{e}_k \rangle, \langle \widehat{r}_h(\mathcal{X}_i), \widehat{e}_k \rangle)_{i=201, \dots, 250}$. Fig. 4 proposes a componentwise prediction graph for the four first components (i.e. $k = 1, \dots, 4$). The percentage of variance explained by these 4 components is 99.7% (i.e. $0.997 = (\sum_{k=1}^4 \widehat{\lambda}_k) / (\sum_{k=1}^{100} \widehat{\lambda}_k)$, where $\widehat{\lambda}_1 > \widehat{\lambda}_2 > \dots$ denotes the eigenvalues of $\Gamma_{\widehat{r}_c(\mathcal{X}),n}$). The quality of componentwise predictions is quite good for each component. *Investigating the bootstrap method.* We illustrate now Theorem 5.1 by comparing, for $k = 1, \dots, 4$, the density function $f_{k,\chi}^{\text{boot}}$ of the componentwise bootstrapped error

$$\langle \widehat{r}_{hb}^{\text{boot}}(\chi) - \widehat{r}_b(\chi), \widehat{e}_k \rangle$$

with the density function $f_{k,\chi}^{\text{true}}$ of the componentwise true error

$$\langle \widehat{r}_h(\chi) - r(\chi), \widehat{e}_k \rangle.$$

So, one has to estimate both density functions and we will do that for any fixed functional predictor $\chi \in \{\mathcal{X}_{201}, \dots, \mathcal{X}_{250}\}$. A standard Monte-Carlo scheme is used to estimate $f_{k,\chi}^{\text{true}}$:

1. build 200 samples $\{(\mathcal{X}_i^s, \mathcal{Y}_i^s)_{i=1, \dots, 200}\}_{s=1, \dots, 200}$,
2. carry out 200 estimates $\{\langle \widehat{r}_h^s(\chi) - r(\chi), \widehat{e}_k \rangle\}_{s=1, \dots, 200}$, where \widehat{r}_h^s is the functional kernel estimator of the regression operator $r(\cdot)$ derived from the sth sample $(\mathcal{X}_i^s, \mathcal{Y}_i^s)_{i=1, \dots, 200}$.

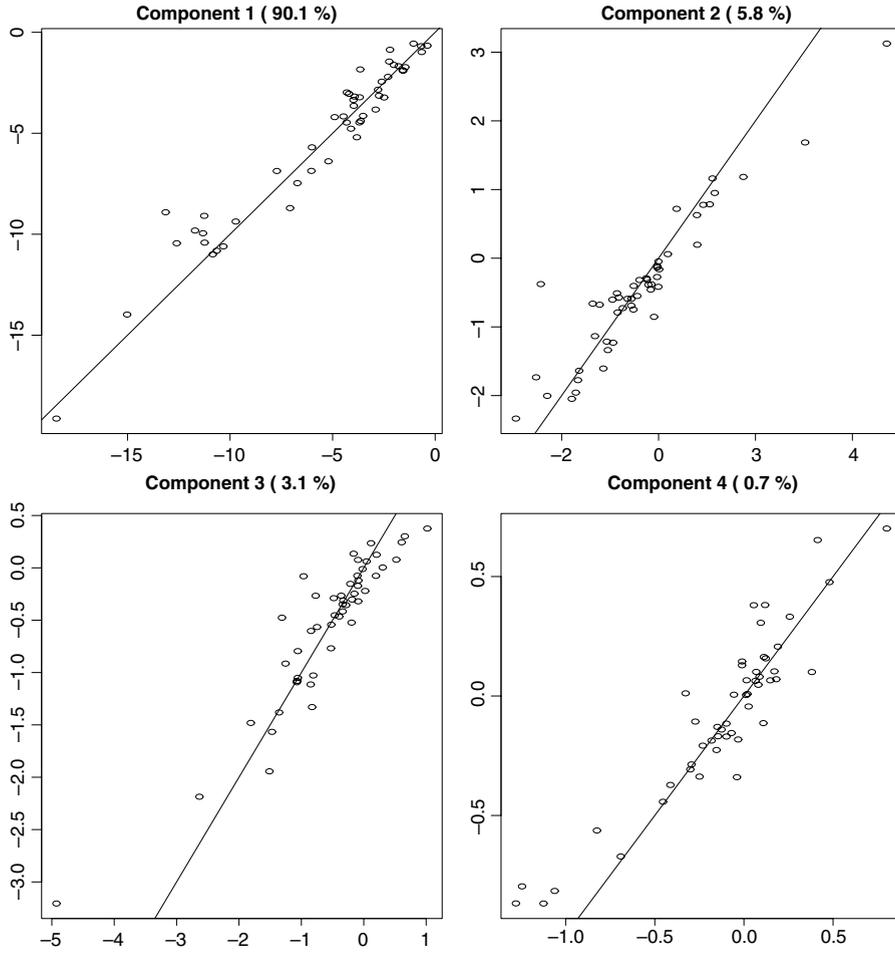


Fig. 4. Componentwise prediction quality.

3. compute a standard density estimator over the 200 values

$$\{(\widehat{r}_h^s(\chi) - r(\chi), \widehat{e}_k)\}_{s=1, \dots, 200}.$$

Concerning the estimation of $f_{k,\chi}^{\text{boot}}$, we use the same wild bootstrap procedure as described in [11,15]:

1. compute $\widehat{r}_b(\chi)$ over the initial sample $\mathcal{S} = (\mathcal{X}_i, \mathcal{Y}_i)_{i=1, \dots, 200}$,
2. repeat 200 times the bootstrap algorithm over \mathcal{S} by using i.i.d. random variables V_1, V_2, \dots , drawn from the two Dirac distributions

$$0.1(5 + \sqrt{5})\delta_{(1-\sqrt{5})/2} + 0.1(5 - \sqrt{5})\delta_{(1+\sqrt{5})/2},$$

which ensures that $E(V_1) = 0$ and $E(V_1^2) = E(V_1^3) = 1$,

3. estimate the density $f_{k,\chi}^{\text{boot}}$ by using again any standard estimator over the 200 values $(\widehat{r}_{hb}^{\text{boot}1}(\chi) - \widehat{r}_b(\chi), \widehat{e}_k), \dots, (\widehat{r}_{hb}^{\text{boot}200}(\chi) - \widehat{r}_b(\chi), \widehat{e}_k)$.

The kernel estimator uses the asymmetric quadratic kernel and the semi-metric $d_4(\cdot, \cdot)$. The bandwidth h is selected via a cross-validation procedure and we set $b = h$. Fig. 5 compares the estimated $f_{k,\chi}^{\text{boot}}$ with the corresponding estimated $f_{k,\chi}^{\text{true}}$ for the four first components (i.e. $k = 1, \dots, 4$). The first row corresponds to the fixed curves $\chi = \mathcal{X}_{201}$, the second to $\chi = \mathcal{X}_{202}, \dots$, the fifth to $\chi = \mathcal{X}_{205}$. It is clear that both densities are very close for each component. In order to assess the overall quality of the bootstrap method, one computes the variational distance between the estimation of $f_{k,\chi}^{\text{true}}$ and $f_{k,\chi}^{\text{boot}}$ (i.e. $\text{dist}_{\chi,k} = 0.5 \int |f_{k,\chi}^{\text{true}}(t) - f_{k,\chi}^{\text{boot}}(t)| dt \in [0, 1]$), at a fixed curve χ in $\{\mathcal{X}_{201}, \dots, \mathcal{X}_{250}\}$ and for the four components.

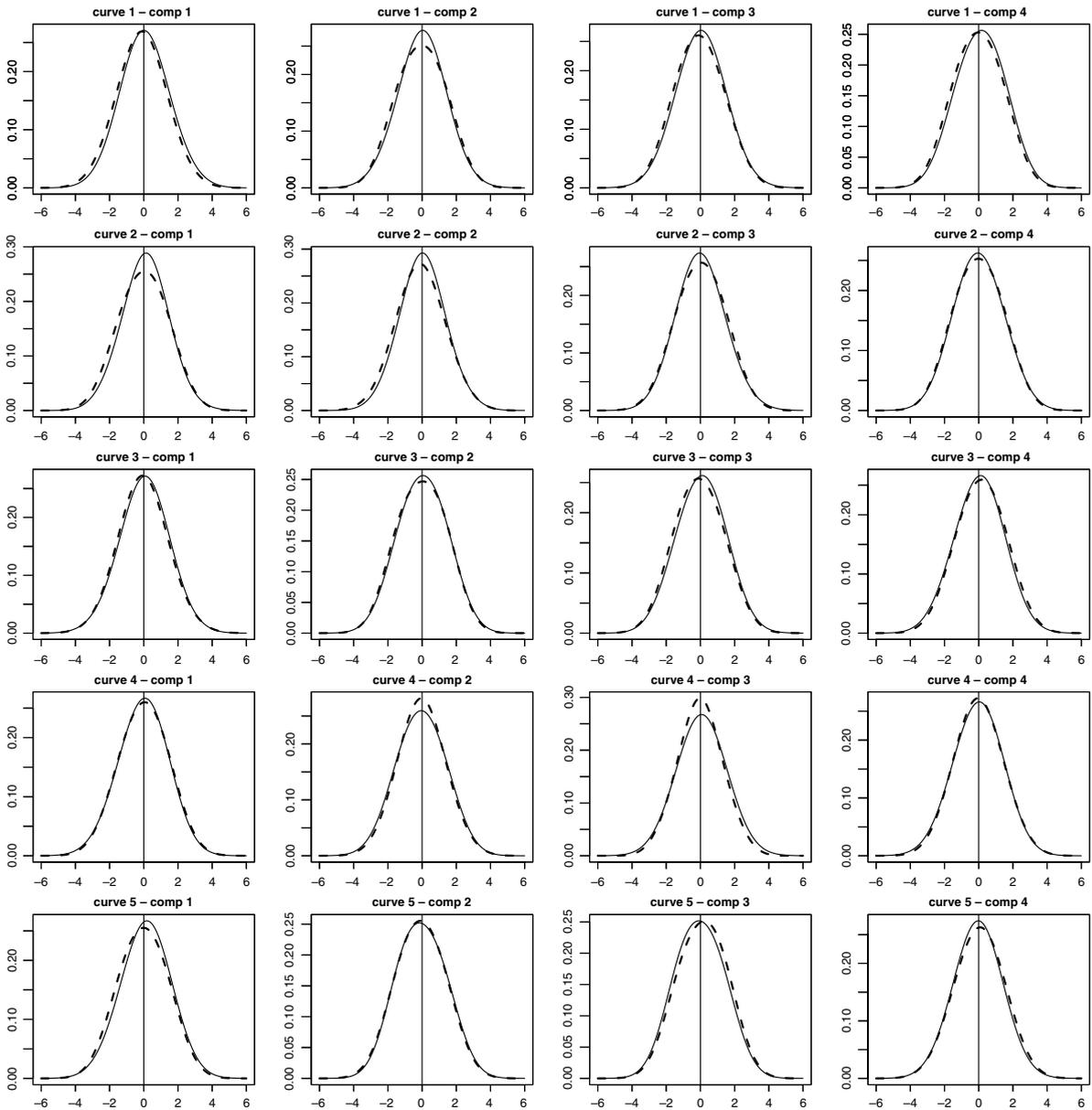


Fig. 5. Estimated densities: solid line → true error; dashed line → bootstrap error.

Fig. 6 displays for each component $k = 1, 2, 3, 4$, the boxplot derived from the 50 values $\text{dist}_{\mathcal{X}_{201},k}, \dots, \text{dist}_{\mathcal{X}_{250},k}$. It appears clearly that the true errors can be very well approximated by the bootstrapped errors. *Toward functional pseudo-confidence areas.* According to the previous developments, one is able to produce componentwise confidence intervals named, for any component k ($k = 1, 2, \dots$), confidence intervals $\widehat{I}_k^{\alpha_k}$ such that

$$P(\widetilde{r}_k(\chi) \in \widehat{I}_k^{\alpha_k}) = 1 - \alpha_k,$$

where $\widetilde{r}_k(\chi) = \langle r(\chi), \widehat{e}_k \rangle$ with $\widehat{e}_1, \dots, \widehat{e}_K$, the K eigenfunctions associated to the K largest eigenvalues of $\Gamma_{\widehat{\mathcal{T}}_c(\mathcal{X})}$ (i.e. $\widehat{e}_1, \widehat{e}_2, \dots$ is a data-driven orthonormal basis). So, for a finite fixed number K of components, one gets

$$P\left(\bigcap_{k=1}^K [\widetilde{r}_k(\chi) \in \widehat{I}_k^{\alpha_k}]\right) \geq 1 - \alpha,$$

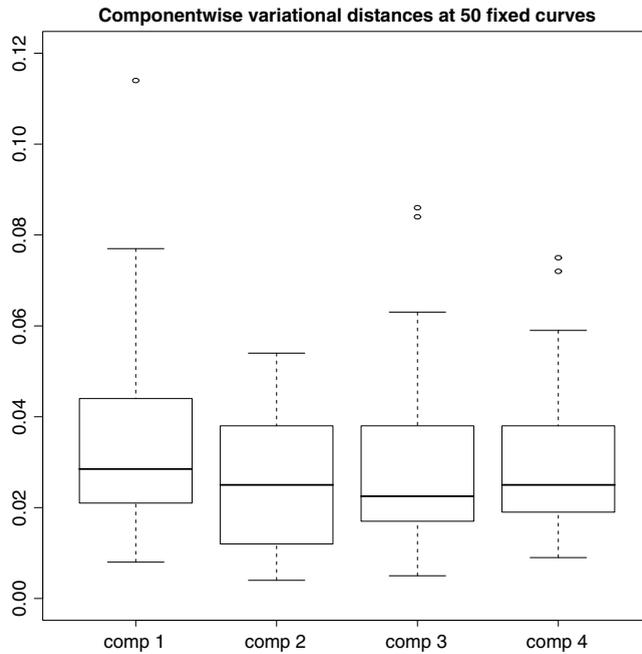


Fig. 6. Componentwise variational distances between estimation of $f_{k,\chi}^{\text{true}}$ and of $f_{k,\chi}^{\text{boot}}$.

as soon as, for $k = 1, \dots, K$, one sets $\alpha_k = \alpha/K$ with $\alpha \in (0, 1)$. This last inequality amounts to the following one:

$$P\left(\sum_{k=1}^K \tilde{r}_k(\chi) \hat{e}_k(\cdot) \in \hat{E}^\alpha\right) \geq 1 - \alpha,$$

with $\hat{E}^\alpha = \left\{ \sum_{k=1}^K a_k \hat{e}_k(\cdot), (a_1, \dots, a_K) \in \hat{I}_1^{\alpha_1} \times \dots \times \hat{I}_K^{\alpha_K} \right\}$. This means that one is able to produce a functional pseudo-confidence area for the projection $\tilde{r}^K(\chi)$ of $r(\chi)$ onto the K -dimensional subspace of \mathcal{H} spanned by the K data-driven basis functions $\hat{e}_1, \dots, \hat{e}_K$. Fig. 7 displays this functional pseudo-confidence area for 9 different fixed curves extracted from the testing sample with $\alpha = 0.05$ and $K = 4$ (recall that the four first components contain 99.7% of the variance of $r(\mathcal{X})$). The shape of these pseudo-confidence areas is quite natural since the starting point is zero for each simulated functional response. Moreover, one can remark that $r(\chi)$ and its K -dimensional projection onto $\hat{e}_1, \dots, \hat{e}_K$ are very close for this example. Of course, when one replaces the data-driven basis with the eigenfunctions of Γ_y , one gets very similar functional pseudo-confidence areas.

8. Conclusions

This paper proposes significant advances for analyzing a nonparametric regression model when both the response and the predictor are functional variables. We show that the kernel estimator provides good predictions under this model. One of the main contributions of this work is that we allow for random bases, which is important for implementing our method in practice. We also show that the bootstrap methodology remains valid in this double functional setting, from both theoretical and practical point of view. Consequently, one is able to plot functional pseudo-confidence areas, which is a very interesting tool for assessing the quality of prediction.

Acknowledgments

All the participants of the STAPH group on Functional Statistics in the Mathematical Institute of Toulouse are greatly acknowledged for their helpful comments on this work (activities of this group are available at the webpage <http://www.lsp.ups-tlse/staph>). Reviewers and Editors of the journal are also greatly acknowledged for their pertinent comments on an earlier version of this paper. In addition, the second author acknowledges support from IAP research network nr. P6/03 of the Belgian government (Belgian Science Policy), and from the European Research Council under

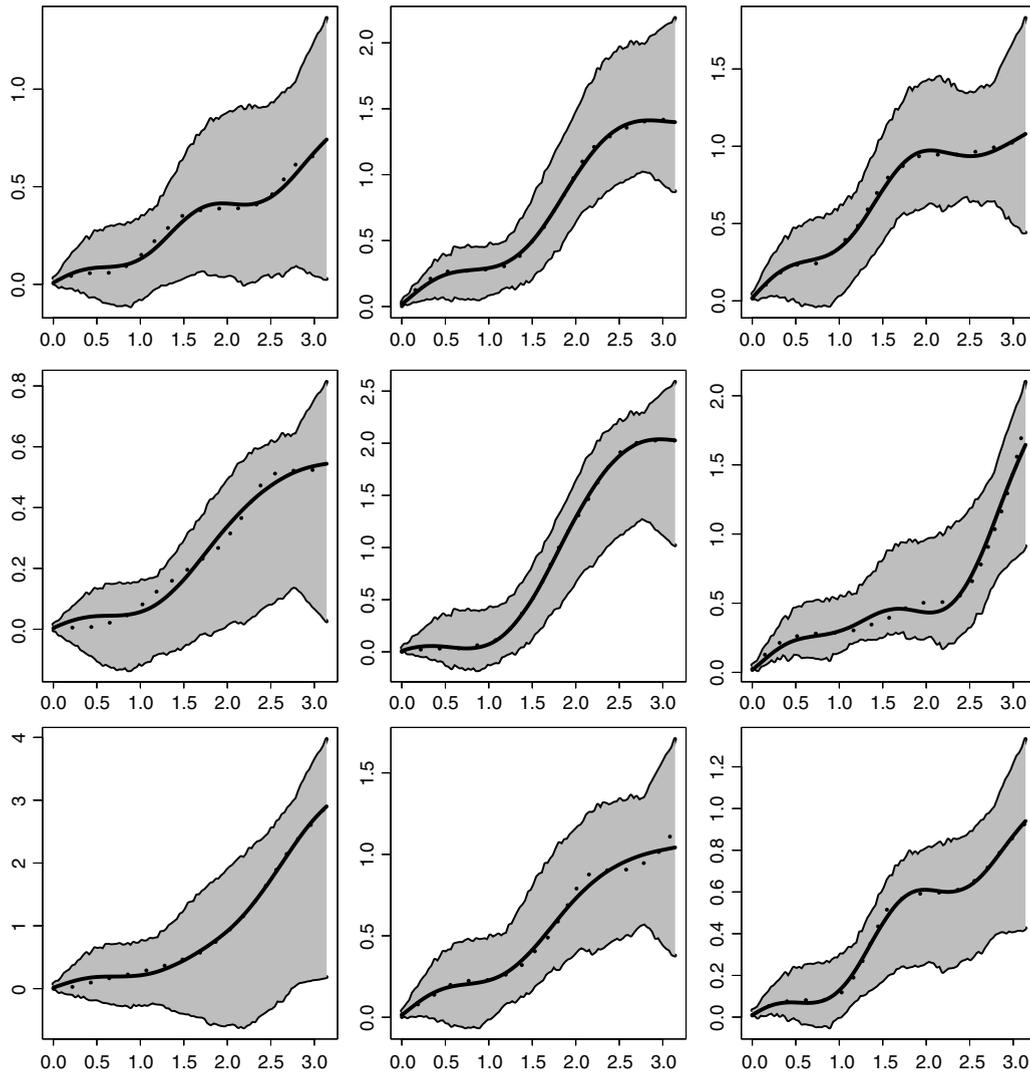


Fig. 7. Functional pseudo-confidence area: thick solid line $\rightarrow \tilde{r}(\chi)$; thick dotted line $\rightarrow r(\chi)$; thin solid line \rightarrow lower/upper functional bounds; gray color \rightarrow 95% functional pseudo-confidence area.

the European Community’s Seventh Framework Programme (FP7/2007–2013) / ERC Grant agreement No. 203650. She also acknowledges the financial support delivered by the Université Paul Sabatier as invited Professor.

Appendix. Appendix section

A.1. Proof of Theorem 4.1

Write

$$\begin{aligned}
 & (nF_\chi(h))^{1/2} \left[\frac{\widehat{g}(\chi)}{\widehat{f}(\chi)} - r(\chi) - \mathcal{B}_{n\chi} \right] \\
 &= (nF_\chi(h))^{1/2} \left[\frac{\widehat{g}(\chi)}{\widehat{f}(\chi)} - \frac{E\widehat{g}(\chi)}{E\widehat{f}(\chi)} \right] + (nF_\chi(h))^{1/2} \left[\frac{E\widehat{g}(\chi)}{E\widehat{f}(\chi)} - r(\chi) - \mathcal{B}_{n\chi} \right].
 \end{aligned}
 \tag{A.1}$$

We will first show that the second term in right hand side of (A.1) is negligible. Consider

$$\begin{aligned} \frac{E\widehat{g}(\chi)}{E\widehat{f}(\chi)} - r(\chi) &= \frac{E\left(\mathcal{Y} - r(\chi)\right)K\left(\frac{d(\mathcal{X}, \chi)}{h}\right)}{E\left(K\left(\frac{d(\mathcal{X}, \chi)}{h}\right)\right)} \\ &= \frac{\sum_{k=1}^{\infty} E\left(\varphi_{\chi,k}(d(\mathcal{X}, \chi))K\left(\frac{d(\mathcal{X}, \chi)}{h}\right)\right)e_k}{E\left(K\left(\frac{d(\mathcal{X}, \chi)}{h}\right)\right)}. \end{aligned} \tag{A.2}$$

For a fixed direction e_k , we can follow the lines of proof of Lemmas 1 and 2 in [13] (replacing φ by $\varphi_{\chi,k}$), which shows that (A.2) equals

$$h \sum_{k=1}^{\infty} \left(\varphi'_{\chi,k}(0) \frac{M_{0\chi}}{M_{1\chi}} + \mathcal{R}_{n,k} \right) e_k,$$

where the remainder term $\mathcal{R}_{n,k}$ satisfies

$$\begin{aligned} |\mathcal{R}_{n,k}| &= \frac{h \left| \int [\varphi'_{\chi,k}(\xi_{t,k}) - \varphi'_{\chi,k}(0)] K(t) dP^{d(\mathcal{X}, \chi)/h}(t) \right|}{\int K(t) dP^{d(\mathcal{X}, \chi)/h}(t)} \\ &\leq h^{1+\alpha} L_k \frac{\left| \int t^\alpha K(t) dP^{d(\mathcal{X}, \chi)/h}(t) \right|}{\int K(t) dP^{d(\mathcal{X}, \chi)/h}(t)} \\ &\leq h^{1+\alpha} L_k, \end{aligned}$$

with $|\xi_{t,k}| \leq ht$ for any $k \geq 1$ and $0 \leq t \leq 1$, and where $0 < \alpha < 1$ and L_k are defined in assumption (C2). Now define $\mathcal{R}_{n,\chi} = \sum_{k=1}^{\infty} \mathcal{R}_{n,k} e_k$. Then, it follows that the second term in right hand side of (A.1) equals $(nF_\chi(h))^{1/2} \mathcal{R}_{n,\chi}$. Next, note that

$$nF_\chi(h) \|\mathcal{R}_{n,\chi}\|^2 = nF_\chi(h) \sum_{k=1}^{\infty} \mathcal{R}_{n,k}^2 \leq nF_\chi(h) h^{2(1+\alpha)} \sum_{k=1}^{\infty} L_k^2 = o(1),$$

by assumptions (C2) and (C4). Hence, by Slutsky's theorem, it suffices to prove the weak convergence of

$$(nF_\chi(h))^{1/2} \left[\frac{\widehat{g}(\chi)}{\widehat{f}(\chi)} - \frac{E\widehat{g}(\chi)}{E\widehat{f}(\chi)} \right] = \frac{(nF_\chi(h))^{1/2}}{\widehat{f}(\chi)E\widehat{f}(\chi)} \left[\{\widehat{g}(\chi) - E\widehat{g}(\chi)\}E\widehat{f}(\chi) - \{\widehat{f}(\chi) - E\widehat{f}(\chi)\}E\widehat{g}(\chi) \right].$$

Again by Slutsky's theorem and by noting that $\widehat{f}(\chi) - M_{1\chi} \xrightarrow{P} 0$ (see Lemma 4 in [13]), and that $E\widehat{g}(\chi)/E\widehat{f}(\chi) - r(\chi) = \mathcal{B}_{n\chi} + O(h^{1+\alpha}) = o(1)$, the latter expression has the same asymptotic distribution as

$$\begin{aligned} &\frac{(nF_\chi(h))^{1/2}}{M_{1\chi}} \left[\widehat{g}(\chi) - E\widehat{g}(\chi) - \{\widehat{f}(\chi) - E\widehat{f}(\chi)\}r(\chi) \right] \\ &= \frac{1}{M_{1\chi}} (nF_\chi(h))^{-1/2} \sum_{i=1}^n \left[\mathcal{Y}_i K\left(\frac{d(\mathcal{X}_i, \chi)}{h}\right) - E\left\{ r(\mathcal{X}) K\left(\frac{d(\mathcal{X}, \chi)}{h}\right) \right\} \right. \\ &\quad \left. - r(\chi) K\left(\frac{d(\mathcal{X}_i, \chi)}{h}\right) + r(\chi) E\left\{ K\left(\frac{d(\mathcal{X}, \chi)}{h}\right) \right\} \right] \\ &= \sum_{i=1}^n (Z_{ni} - EZ_{ni}), \end{aligned}$$

where for $1 \leq i \leq n$,

$$Z_{ni} = \frac{1}{M_{1\chi}} (nF_\chi(h))^{-1/2} \left[\mathcal{Y}_i K\left(\frac{d(\mathcal{X}_i, \chi)}{h}\right) - r(\chi) K\left(\frac{d(\mathcal{X}_i, \chi)}{h}\right) \right].$$

Let us now calculate the covariance operator of $S_n = \sum_{i=1}^n Z_{ni}$:

$$\begin{aligned}
 E \left[\langle S_n - ES_n, e_k \rangle \langle S_n - ES_n, e_\ell \rangle \right] &= \sum_{i=1}^n E \left[\langle Z_{ni} - EZ_{ni}, e_k \rangle \langle Z_{ni} - EZ_{ni}, e_\ell \rangle \right] \\
 &= \sum_{i=1}^n E \left[\langle Z_{ni} - E(Z_{ni} | \mathcal{X}_i), e_k \rangle \langle Z_{ni} - E(Z_{ni} | \mathcal{X}_i), e_\ell \rangle \right] \\
 &\quad + \sum_{i=1}^n E \left[\langle E(Z_{ni} | \mathcal{X}_i) - EZ_{ni}, e_k \rangle \langle E(Z_{ni} | \mathcal{X}_i) - EZ_{ni}, e_\ell \rangle \right] \\
 &= \frac{1}{M_{1\mathcal{X}}^2 F_{\mathcal{X}}(h)} E \left[\langle \mathcal{Y} - r(\mathcal{X}), e_k \rangle \langle \mathcal{Y} - r(\mathcal{X}), e_\ell \rangle K^2 \left(\frac{d(\mathcal{X}, \mathcal{X})}{h} \right) \right] \\
 &\quad + \frac{1}{M_{1\mathcal{X}}^2 F_{\mathcal{X}}(h)} \text{Cov} \left[\langle r(\mathcal{X}) - r(\mathcal{X}), e_k \rangle K \left(\frac{d(\mathcal{X}, \mathcal{X})}{h} \right), \right. \\
 &\quad \left. \times \langle r(\mathcal{X}) - r(\mathcal{X}), e_\ell \rangle K \left(\frac{d(\mathcal{X}, \mathcal{X})}{h} \right) \right]. \tag{A.3}
 \end{aligned}$$

The second term of (A.3) is equal to

$$\begin{aligned}
 &\frac{1}{M_{1\mathcal{X}}^2 F_{\mathcal{X}}(h)} \text{Cov} \left[\varphi_{\mathcal{X},k}(d(\mathcal{X}, \mathcal{X})) K \left(\frac{d(\mathcal{X}, \mathcal{X})}{h} \right), \varphi_{\mathcal{X},\ell}(d(\mathcal{X}, \mathcal{X})) K \left(\frac{d(\mathcal{X}, \mathcal{X})}{h} \right) \right] \\
 &= \frac{1}{M_{1\mathcal{X}}^2 F_{\mathcal{X}}(h)} \left[\int_0^1 \varphi_{\mathcal{X},k}(ht) \varphi_{\mathcal{X},\ell}(ht) K^2(t) dP^{d(\mathcal{X}, \mathcal{X})/h}(t) \right. \\
 &\quad \left. - \int_0^1 \varphi_{\mathcal{X},k}(ht) K(t) dP^{d(\mathcal{X}, \mathcal{X})/h}(t) : \int_0^1 \varphi_{\mathcal{X},\ell}(ht) K(t) dP^{d(\mathcal{X}, \mathcal{X})/h}(t) \right] \\
 &= a_{nkl,2} \quad (\text{say}).
 \end{aligned}$$

On the other hand, the first term of (A.3) can be written as

$$\frac{1}{M_{1\mathcal{X}}^2 F_{\mathcal{X}}(h)} E \left\{ s_{k\ell}(\mathcal{X}) K^2 \left(\frac{d(\mathcal{X}, \mathcal{X})}{h} \right) \right\} := a_{nkl,1} \quad (\text{say}).$$

Let $a_{nkl} = a_{nkl,1} + a_{nkl,2}$. It now follows from Theorem 1 in [19] that it suffices to show the following three conditions:

- (i) $\lim_{n \rightarrow \infty} a_{nkl} = a_{kl}$ for all $k, \ell \geq 1$.
- (ii) $\lim_{n \rightarrow \infty} \sum_{k=1}^{\infty} a_{nkk} = \sum_{k=1}^{\infty} a_{kk} < \infty$.
- (iii) $\forall \delta > 0, \forall k \geq 1: \lim_{n \rightarrow \infty} \sum_{i=1}^n E(\langle Z_{ni}, e_k \rangle^2 I\{|\langle Z_{ni}, e_k \rangle| > \delta\}) = 0$.

Proof of (i). Using the continuity of $\varphi_{\mathcal{X},k}$ and $\varphi_{\mathcal{X},\ell}$ it is easily seen that $a_{nkl,2} = o(1)$. Moreover, the continuity of $s_{k\ell}(\mathcal{X})$ leads to

$$\begin{aligned}
 a_{nkl,1} &= \frac{1}{M_{1\mathcal{X}}^2 F_{\mathcal{X}}(h)} s_{k\ell}(\mathcal{X}) E \left\{ K^2 \left(\frac{d(\mathcal{X}, \mathcal{X})}{h} \right) \right\} (1 + o(1)) \\
 &= \frac{1}{M_{1\mathcal{X}}^2 F_{\mathcal{X}}(h)} s_{k\ell}(\mathcal{X}) F_{\mathcal{X}}(h) M_{2\mathcal{X}} (1 + o(1)) \\
 &= a_{kl} + o(1),
 \end{aligned}$$

where the second equality follows from Lemma 5 in [13].

Proof of (ii). For the second condition, a more refined derivation is needed, since the remainder terms in the calculation leading to a_{kl} should be summable. This can be achieved using arguments similar to those used for the bias term $\mathcal{B}_{n\mathcal{X}}$. In fact, using assumption (C3) we have that

$$\begin{aligned}
 a_{nkk,1} &= \frac{1}{M_{1\mathcal{X}}^2 F_{\mathcal{X}}(h)} E \left[\psi_{\mathcal{X},kk}(d(\mathcal{X}, \mathcal{X})) K^2 \left(\frac{d(\mathcal{X}, \mathcal{X})}{h} \right) \right] + \frac{M_{2\mathcal{X}} s_{kk}(\mathcal{X})}{M_{1\mathcal{X}}^2} (1 + o(1)) \\
 &\leq \frac{1}{M_{1\mathcal{X}}^2 F_{\mathcal{X}}(h)} N_k E \left[(d(\mathcal{X}, \mathcal{X}))^\beta K^2 \left(\frac{d(\mathcal{X}, \mathcal{X})}{h} \right) \right] + \frac{M_{2\mathcal{X}} s_{kk}(\mathcal{X})}{M_{1\mathcal{X}}^2} (1 + o(1)).
 \end{aligned}$$

Hence,

$$\begin{aligned} \sum_{k=1}^{\infty} a_{nkk,1} &= O(h^\beta) \sum_{k=1}^{\infty} N_k + \frac{M_{2\mathcal{X}} \sum_{k=1}^{\infty} s_{kk}(\mathcal{X})}{M_{1\mathcal{X}}^2} (1 + o(1)) \\ &= \left(\sum_{k=1}^{\infty} a_{kk} \right) (1 + o(1)), \end{aligned}$$

since $\sum_{k=1}^{\infty} N_k < \infty$ by assumption (C3). Similarly, we have that

$$\sum_{k=1}^{\infty} a_{nkk,2} = o(1),$$

provided that $\sum_{k=1}^{\infty} \varphi'_{\mathcal{X},k}(0)^2 < \infty$ and $\sum_{k=1}^{\infty} L_k^2 < \infty$ (see condition (C2)). This shows that (ii) is valid, since

$$\sum_{k=1}^{\infty} a_{kk} = \sum_{k=1}^{\infty} \text{Var}(\langle \mathcal{Y}, e_k \rangle | \mathcal{X} = \mathcal{X}) \frac{M_{2\mathcal{X}}}{M_{1\mathcal{X}}^2} = \text{Var}(\|\mathcal{Y}\| | \mathcal{X} = \mathcal{X}) \frac{M_{2\mathcal{X}}}{M_{1\mathcal{X}}^2},$$

which is finite by assumptions (C6) and (C7).

Proof of (iii). Finally, for condition (iii) above, consider

$$\begin{aligned} &\sum_{i=1}^n E \left(\langle Z_{ni} - EZ_{ni}, e_k \rangle^2 I_{\{|\langle Z_{ni} - EZ_{ni}, e_k \rangle| > \delta\}} \right) \\ &\leq \sum_{i=1}^n \left[E \left(\langle Z_{ni} - EZ_{ni}, e_k \rangle^4 \right) P\{|\langle Z_{ni} - EZ_{ni}, e_k \rangle| > \delta\} \right]^{1/2} \\ &\leq 16F_{\mathcal{X}}(h)^{-1} \left[E \left(\langle \mathcal{Y} - r(\mathcal{X}), e_k \rangle^4 K^4 \left(\frac{d(\mathcal{X}, \mathcal{X})}{h} \right) \right) P \left(\left| \langle \mathcal{Y} - r(\mathcal{X}), e_k \rangle K \left(\frac{d(\mathcal{X}, \mathcal{X})}{h} \right) \right. \right. \right. \\ &\quad \left. \left. \left. - E \left\{ \langle \mathcal{Y} - r(\mathcal{X}), e_k \rangle K \left(\frac{d(\mathcal{X}, \mathcal{X})}{h} \right) \right\} \right| > \delta(nF_{\mathcal{X}}(h))^{1/2} \right) \right]^{1/2} \\ &\leq 16F_{\mathcal{X}}(h)^{-1} \left\{ E \left(\langle \mathcal{Y} - r(\mathcal{X}), e_k \rangle^4 | \mathcal{X} = \mathcal{X} \right) 1 + o(1) \right\}^{1/2} E \left\{ K^4 \left(\frac{d(\mathcal{X}, \mathcal{X})}{h} \right) \right\}^{1/2} \\ &\quad \times \text{Var} \left\{ \langle \mathcal{Y} - r(\mathcal{X}), e_k \rangle K \left(\frac{d(\mathcal{X}, \mathcal{X})}{h} \right) \right\}^{1/2} \delta^{-1} (nF_{\mathcal{X}}(h))^{-1/2}, \end{aligned}$$

and this tends to zero since $E(\langle \mathcal{Y} - r(\mathcal{X}), e_k \rangle^4 | \mathcal{X} = \mathcal{X}) < \infty$, $E\{K^4(\frac{d(\mathcal{X}, \mathcal{X})}{h})\} = O(F_{\mathcal{X}}(h))$ and $\text{Var}\{\langle \mathcal{Y} - r(\mathcal{X}), e_k \rangle K(\frac{d(\mathcal{X}, \mathcal{X})}{h})\} = O(F_{\mathcal{X}}(h))$.

This shows that all the conditions of Theorem 1 in [19] are satisfied and hence $\sum_{i=1}^n [Z_{ni} - EZ_{ni}]$ converges to a zero mean normal limit with covariance operator given by $C_{\mathcal{X}}$. \square

A.2. Proof of Theorem 6.1

We give the proof for the wild bootstrap procedure. For the naive bootstrap, the arguments for the calculation of the variance need to be slightly adapted (see the end of the proof for more details). Write (with $a_n = (nF_{\mathcal{X}}(h))^{1/2}$)

$$\begin{aligned} &P^\delta \left(a_n \left\{ \widehat{r}_{k,hb}^{\text{boot}}(\mathcal{X}) - \widehat{r}_{k,b}(\mathcal{X}) \right\} \leq y \right) - P \left(a_n \left\{ \widehat{r}_{k,h}(\mathcal{X}) - \widehat{r}_k(\mathcal{X}) \right\} \leq y \right) \\ &= P^\delta \left(a_n \left\{ \widehat{r}_{k,hb}^{\text{boot}}(\mathcal{X}) - \widehat{r}_{k,b}(\mathcal{X}) \right\} \leq y \right) - P^\delta \left(a_n \left\{ \widehat{r}_{k,hb}^{\text{boot}}(\mathcal{X}) - \widehat{r}_{k,b}(\mathcal{X}) \right\} \leq y \right) \\ &\quad + P^\delta \left(a_n \left\{ \widehat{r}_{k,hb}^{\text{boot}}(\mathcal{X}) - \widehat{r}_{k,b}(\mathcal{X}) \right\} \leq y \right) - P \left(a_n \left\{ \widehat{r}_{k,h}(\mathcal{X}) - r_k(\mathcal{X}) \right\} \leq y \right) \\ &\quad + P \left(a_n \left\{ \widehat{r}_{k,h}(\mathcal{X}) - r_k(\mathcal{X}) \right\} \leq y \right) - P \left(a_n \left\{ \widehat{r}_{k,h}(\mathcal{X}) - \widehat{r}_k(\mathcal{X}) \right\} \leq y \right). \end{aligned} \tag{A.4}$$

The second term of (A.4) is $o(1)$ a.s. by Theorem 5.1. For the third term, note that if we write $Z_n = a_n(\widehat{r}_h(\chi) - r(\chi))$, then $a_n(\widehat{r}_{k,h}(\chi) - r_k(\chi)) = \langle Z_n, \widehat{e}_k \rangle$ and $a_n(\widehat{r}_{k,h}(\chi) - r_k(\chi)) = \langle Z_n, e_k \rangle$. Hence,

$$\langle Z_n, \widehat{e}_k \rangle = \langle Z_n, e_k \rangle + \langle Z_n, \widehat{e}_k - e_k \rangle \leq \langle Z_n, e_k \rangle + \|Z_n\| \|\widehat{e}_k - e_k\| = \langle Z_n, e_k \rangle + o_p(1),$$

since $\|\widehat{e}_k - e_k\| = o_p(1)$ and $\|Z_n\| = O_p(1)$ (this last point follows from $E\|Z_n\|^2 < \infty$ under the conditions of Theorem 5.1). Hence the third term of (A.4) is $o(1)$. It remains to consider the first term of (A.4). It follows from the proof of Theorem 1 in [11,15] that this term is equal to

$$\Phi\left(\frac{y - a_n\{E^\delta[\widehat{r}_{k,hb}^{\text{boot}}(\chi)] - \widehat{r}_{k,b}(\chi)\}}{a_n\sqrt{\text{Var}^\delta[\widehat{r}_{k,hb}^{\text{boot}}(\chi)]}}\right) - \Phi\left(\frac{y - a_n\{E^\delta[\widehat{r}_{k,hb}^{\text{boot}}(\chi)] - \widehat{r}_{k,b}(\chi)\}}{a_n\sqrt{\text{Var}^\delta[\widehat{r}_{k,hb}^{\text{boot}}(\chi)]}}\right) + o(1)$$

a.s., and that this is $o(1)$ a.s. uniformly in y provided

$$a_n|E^\delta[\widehat{r}_{k,hb}^{\text{boot}}(\chi)] - \widehat{r}_{k,b}(\chi) - E^\delta[\widehat{r}_{k,hb}^{\text{boot}}(\chi)] + \widehat{r}_{k,b}(\chi)| = o(1), \quad \text{a.s.}, \tag{A.5}$$

and

$$\frac{\text{Var}^\delta[\widehat{r}_{k,hb}^{\text{boot}}(\chi)]}{\text{Var}^\delta[\widehat{r}_{k,hb}^{\text{boot}}(\chi)]} \rightarrow 1, \quad \text{a.s.} \tag{A.6}$$

For the proof of (A.5) note that the expression between absolute values equals (where the subindex h in $\widehat{f}_h(\chi)$ is added to make clear which bandwidth we are using)

$$\begin{aligned} & \frac{(nF_\chi(h))^{-1}}{\widehat{f}_h(\chi)} \sum_{i=1}^n \left\{ \widehat{r}_{k,b}(\mathcal{X}_i) - \widehat{r}_{k,b}(\chi) - \widehat{r}_{k,b}(\mathcal{X}_i) + \widehat{r}_{k,b}(\chi) \right\} K(h^{-1}d(\mathcal{X}_i, \chi)) \\ &= \frac{(nF_\chi(h))^{-1}}{\widehat{f}_h(\chi)} \sum_{i=1}^n (\widehat{r}_b(\mathcal{X}_i) - \widehat{r}_b(\chi), \widehat{e}_k - e_k) K(h^{-1}d(\mathcal{X}_i, \chi)) \\ &= \frac{(nF_\chi(h))^{-1}}{\widehat{f}_h(\chi)} \sum_{i,j=1}^n K_h(d(\mathcal{X}_i, \chi)) \left[\frac{K_b(d(\mathcal{X}_j, \mathcal{X}_i))}{\sum_\ell K_b(d(\mathcal{X}_\ell, \mathcal{X}_i))} - \frac{K_b(d(\mathcal{X}_j, \chi))}{\sum_\ell K_b(d(\mathcal{X}_\ell, \chi))} \right] \times \langle \mathcal{Y}_j, \widehat{e}_k - e_k \rangle, \end{aligned} \tag{A.7}$$

where for any h , $K_h(u) = K(h^{-1}u)$. For fixed values of i and j ,

$$\begin{aligned} & K_h(d(\mathcal{X}_i, \chi)) \left[\frac{K_b(d(\mathcal{X}_j, \mathcal{X}_i))}{\sum_\ell K_b(d(\mathcal{X}_\ell, \mathcal{X}_i))} - \frac{K_b(d(\mathcal{X}_j, \chi))}{\sum_\ell K_b(d(\mathcal{X}_\ell, \chi))} \right] \\ &= K_h(d(\mathcal{X}_i, \chi)) \left\{ K_b(d(\mathcal{X}_j, \mathcal{X}_i)) \left[\frac{1}{nF_{\mathcal{X}_i}(b)\widehat{f}_b(\mathcal{X}_i)} - \frac{1}{nF_\chi(b)\widehat{f}_b(\chi)} \right] \right. \\ & \quad \left. + \frac{1}{nF_\chi(b)\widehat{f}_b(\chi)} \left[K_b(d(\mathcal{X}_j, \mathcal{X}_i)) - K_b(d(\mathcal{X}_j, \chi)) \right] \right\} \\ &\leq \left\{ CK_h(d(\mathcal{X}_i, \chi)) I(d(\mathcal{X}_j, \chi) \leq b+h) \left[(nF_\chi(h))^{-1/2} + h^\alpha \right] (nF_\chi(b))^{-1} \right. \\ & \quad \left. + \widehat{f}_b^{-1}(\chi) K_h(d(\mathcal{X}_i, \chi)) I(d(\mathcal{X}_j, \chi) \leq b+h) \frac{h}{b} (nF_\chi(b))^{-1} \right\} (1 + o(1)), \end{aligned}$$

for some $0 < C < \infty$, since it follows from Lemmas 5 and 6 in [11,15] that

$$\sup_{d(\mathcal{X}_1, \chi) \leq h} |\widehat{f}_b(\chi_1) - \widehat{f}_b(\chi)| = o\left((nF_\chi(h))^{-1/2}\right) \quad \text{a.s.}$$

and since it follows from the Lipschitz continuity of $F_{\mathcal{X}_1}(t)/F_\chi(t)$ of order α uniformly in t that

$$\sup_{d(\mathcal{X}_1, \chi) \leq h} |F_{\mathcal{X}_1}(b) - F_\chi(b)| = o\left(h^\alpha F_\chi(b)\right) \quad \text{a.s.},$$

where α is defined in condition (C2). It now follows that the absolute value of (A.7) multiplied by $(nF_\chi(h))^{1/2}$ is asymptotically bounded by

$$\begin{aligned} & \left[C(nF_\chi(h))^{-1/2} + Ch^\alpha + \widehat{f}_b^{-1}(\chi) \frac{h}{b} \right] (nF_\chi(h))^{-1/2} (nF_\chi(b))^{-1} \\ & \quad \times \widehat{f}_h^{-1}(\chi) \sum_{i,j=1}^n K_h(d(\mathcal{X}_i, \chi)) I(d(\mathcal{X}_j, \chi) \leq b+h) \|\mathcal{Y}_j\| \cdot \|\widehat{e}_k - e_k\| \\ & \leq \left[C + \frac{h}{b} (nF_\chi(h))^{1/2} \right] \|\widehat{e}_k - e_k\| E(\|\mathcal{Y}\| | \mathcal{X} = \chi) \frac{F_\chi(b+h)}{F_\chi(b)} (1 + o(1)), \end{aligned}$$

since $h^\alpha = O(h/b)$, and this is $o(1)$ a.s. by the assumptions given in the statement of the theorem. Next, consider the verification of (A.6):

$$\begin{aligned} & \text{Var}^\delta [\widehat{r}_{k,hb}^{\text{boot}}(\chi)] - \text{Var}^\delta [\widehat{r}_{k,hb}(\chi)] \\ & = \frac{1}{(nF_\chi(h))^2 \widehat{f}_h^2(\chi)} \sum_{i=1}^n \left[\langle \mathcal{Y}_i - \widehat{r}_b(\mathcal{X}_i), \widehat{e}_k \rangle^2 - \langle \mathcal{Y}_i - \widehat{r}_b(\mathcal{X}_i), e_k \rangle^2 \right] K^2(h^{-1}d(\mathcal{X}_i, \chi)) \\ & = \frac{1}{(nF_\chi(h))^2 \widehat{f}_h^2(\chi)} \sum_{i=1}^n \langle \mathcal{Y}_i - \widehat{r}_b(\mathcal{X}_i), \widehat{e}_k - e_k \rangle \langle \mathcal{Y}_i - \widehat{r}_b(\mathcal{X}_i), \widehat{e}_k + e_k \rangle K^2(h^{-1}d(\mathcal{X}_i, \chi)) \\ & \leq \frac{1}{(nF_\chi(h))^2 \widehat{f}_h^2(\chi)} \sum_{i=1}^n K^2(h^{-1}d(\mathcal{X}_i, \chi)) \|\mathcal{Y}_i - \widehat{r}_b(\mathcal{X}_i)\|^2 \|\widehat{e}_k - e_k\| (\|\widehat{e}_k\| + \|e_k\|) \\ & \leq \frac{M_{2\chi}}{nF_\chi(h)} E(\|\mathcal{Y} - r(\mathcal{X})\|^2 | \mathcal{X} = \chi) \|\widehat{e}_k - e_k\| (\|\widehat{e}_k\| + \|e_k\|) (1 + o(1)), \end{aligned}$$

which is $o((nF_\chi(h))^{-1})$ a.s., since $\|\widehat{e}_k - e_k\| = o(1)$ a.s. Note that when the naive bootstrap is used, we have that (where $\widehat{\sigma}_{\varepsilon,k}^2 = n^{-1} \sum_{i=1}^n \langle \widehat{\varepsilon}_{i,b} - \widehat{\varepsilon}_b, e_k \rangle^2$ and similarly for $\widehat{\sigma}_{\varepsilon,\widehat{k}}^2$):

$$\begin{aligned} \text{Var}^\delta [\widehat{r}_{k,hb}^{\text{boot}}(\chi)] - \text{Var}^\delta [\widehat{r}_{k,hb}(\chi)] & = \frac{1}{(nF_\chi(h))^2 \widehat{f}_h^2(\chi)} \sum_{i=1}^n K^2(h^{-1}d(\mathcal{X}_i, \chi)) [\widehat{\sigma}_{\varepsilon,\widehat{k}}^2 - \widehat{\sigma}_{\varepsilon,k}^2] \\ & \leq \frac{M_{2\chi}}{nF_\chi(h)} E(\|\mathcal{Y} - r(\mathcal{X})\|^2) \|\widehat{e}_k - e_k\| (\|\widehat{e}_k\| + \|e_k\|) (1 + o(1)) = o((nF_\chi(h))^{-1}). \end{aligned}$$

The proof is now complete. \square

A.3. Proof of Proposition 1

According to Bosq [2], for any fixed k , there exists a constant C_k ($0 < C_k < \infty$) such that

$$\|\widehat{e}_k - e_k\| \leq C_k \|\Gamma_{\mathcal{Y},n} - \Gamma_{\mathcal{Y}}\|_\infty, \tag{A.8}$$

where $\|\cdot\|_\infty$ is the standard operator norm (i.e. $\|U\|_\infty = \sup_{\|x\|=1} \|U(x)\|$). The proof of Proposition 1 is then based on the following intermediate result:

Lemma 1. *Let $\mathcal{Y}_1, \dots, \mathcal{Y}_n$ be an i.i.d. sequence of Hilbertian random variables and assume that there exists $M > 0$ such that $E\|\mathcal{Y}_1\|^{2m} \leq Mm!$. Then, we have:*

$$\|\Gamma_{\mathcal{Y},n} - \Gamma_{\mathcal{Y}}\|_\infty = O_{\text{a.co.}} \left(\sqrt{\frac{\log n}{n}} \right). \tag{A.9}$$

The proof of this lemma is based on Yurinskii's Corollary [29, p. 491] and is quite standard (this is why we omitted its proof which is available on request). For more details and references on the literature dealing with this kind of results, see for instance [3] or [20].

Now, (A.9) combined with (A.8) leads to, for any fixed k ,

$$\|\widehat{e}_k - e_k\| = O_{\text{a.co.}} \left(\sqrt{\frac{\log n}{n}} \right).$$

In order to end the proof of Proposition 1, it remains to prove that $\sqrt{\log n/n} = o((b/h)(nF_\chi(h))^{-1/2})$ or equivalently $(h/b)\sqrt{\log n F_\chi(h)} = o(1)$.

Case 1: $\log n F_{\chi}(h) = O(1)$. The result is trivial since $h/b = o(1)$.

Case 2: $\log n F_{\chi}(h) \rightarrow +\infty$. **Theorem 5.1** assumes $b^{1+\alpha} \sqrt{n F_{\chi}(h)} \rightarrow 0$ which implies that $(n/\log n)b^{2+2\alpha} = o(1)$ and hence $b = o((\log n/n)^{\epsilon})$, with $\epsilon > 0$ ($\alpha \in (0, 1]$). On the other hand, **Theorem 5.1** assumes also that $b h^{\alpha-1} = O(1)$, which is equivalent to $h = O(b^{1+\epsilon'})$ with $\epsilon' > 0$ and this leads us to $h/b = o((\log n/n)^{\epsilon''})$ with $\epsilon'' > 0$. Then, it is easy to see that $(h/b)\sqrt{\log n} = o(n^{-\epsilon''}(\log n)^{\epsilon''+1/2}) = o(1)$ and $(h/b)\sqrt{\log n F_{\chi}(h)} = o(1)$ holds. \square

A.4. Proof of Proposition 2

The proof is based on the following decomposition:

$$\|F_{\hat{r}_h(\mathcal{X})} - F_{r(\mathcal{X})}\|_{\infty} \leq \|F_{\hat{r}_h(\mathcal{X})} - F_{r(\mathcal{X}),n}\|_{\infty} + \|F_{r(\mathcal{X}),n} - F_{r(\mathcal{X})}\|_{\infty}. \quad (\text{A.10})$$

By using (A.9) with $r(\mathcal{X})$ instead of \mathcal{Y} , we obtain:

$$\|F_{r(\mathcal{X}),n} - F_{r(\mathcal{X})}\|_{\infty} = O_{\text{a.co.}} \left(\sqrt{\frac{\log n}{n}} \right). \quad (\text{A.11})$$

Taking the hypotheses of Proposition 2 together with (A.10) and (A.11), one gets:

$$\|F_{\hat{r}_h(\mathcal{X})} - F_{r(\mathcal{X})}\|_{\infty} = o \left(\frac{b}{h\sqrt{n F_{\chi}(h)}} \right) + O \left(\sqrt{\frac{\log n}{n}} \right), \quad \text{a.s.}$$

Remember that $\sqrt{\log n/n} = o(b/(h\sqrt{n F_{\chi}(h)}))$ (see the end of the proof of Proposition 1), and for any fixed k , there exists a $0 < C_k < \infty$ such that $\|\hat{e}_k - e_k\| \leq C_k \|F_{\hat{r}_h(\mathcal{X})} - F_{r(\mathcal{X})}\|_{\infty}$, from which Proposition 2 follows. \square

References

- [1] J. Antoch, L. Prchal, M. De Rosa, P. Sarda, Electricity consumption prediction with functional linear regression using spline estimators, *J. Appl. Stat.* 37 (12) (2010) 2027–2041.
- [2] D. Bosq, Modelization, nonparametric estimation and prediction for continuous time processes, in: G. Roussas (Ed.), *Nonparametric Functional Estimation and Related Topics*, in: NATO, ASI Series, 1991, pp. 509–529.
- [3] D. Bosq, *Linear Processes in Functional Spaces. Theory and Applications*, in: *Lecture Notes in Statistics*, vol. 149, Springer Verlag, New York, 2000.
- [4] H. Cardot, A. Mas, P. Sarda, CLT in functional linear regression models, *Probab. Theory Related Fields* 138 (2007) 325–361.
- [5] J.-M. Chiou, H.-G. Müller, J.-L. Wang, Functional response models, *Statist. Sinica* 14 (2004) 659–677.
- [6] C. Crambes, A. Kneip, P. Sarda, Smoothing splines estimators for functional linear regression, *Ann. Statist.* 37 (2009) 35–72.
- [7] A. Cuevas, M. Febrero, R. Fraiman, Linear functional regression: the case of fixed design and functional response, *Canad. J. Statist.* 30 (2002) 285–300.
- [8] J. Dauxois, A. Pousse, Y. Romain, Asymptotic theory for the principal component analysis of a random vector function: some application to statistical inference, *J. Multivariate Anal.* 12 (1982) 136–154.
- [9] L. Delsol, Advances on asymptotic normality in nonparametric functional time series analysis, *Statistics* 43 (2009) 13–33.
- [10] J. Fan, J.T. Zhang, Functional linear models for longitudinal data, *J. R. Stat. Soc. Ser. B* 39 (1998) 254–261.
- [11] F. Ferraty, A. Laksaci, A. Tadj, P. Vieu, Rate of uniform consistency for nonparametric estimates with functional variables, *J. Statist. Plann. Inference* 140 (2010) 335–352.
- [12] F. Ferraty, A. Laksaci, A. Tadj, P. Vieu, Kernel regression with functional response, *Electron. J. Stat.* 5 (2011) 159–171.
- [13] F. Ferraty, A. Mas, P. Vieu, Nonparametric regression on functional data: inference and practical aspects, *Aust. N. Z. J. Stat.* 49 (2007) 267–286.
- [14] F. Ferraty, Y. Romain (Eds.), *The Oxford Handbook of Functional Data Analysis*, Oxford University Press, 2010.
- [15] F. Ferraty, I. Van Keilegom, P. Vieu, On the validity of the bootstrap in nonparametric functional regression, *Scandinavian J. Statist.* 37 (2010) 286–306.
- [16] F. Ferraty, P. Vieu, *Nonparametric Functional Data Analysis: Theory and Practice*, Springer, New York, 2006.
- [17] F. Ferraty, P. Vieu, Kernel regression estimation for functional data, in: F. Ferraty, Y. Romain (Eds.), *Oxford Handbook of Functional Data Analysis*, Oxford University Press, 2010.
- [18] W. Gonzalez-Manteiga, A. Martinez-Calvo, Bootstrap in functional linear regression, *J. Statist. Plann. Inference* 141 (2011) 453–461.
- [19] S. Kundu, S. Majumdar, K. Mukherjee, Central limit theorems revisited, *Statist. Probab. Lett.* 47 (2000) 265–275.
- [20] M. Ledoux, M. Talagrand, *Probability in Banach Spaces*, Springer-Verlag, Berlin, 1991.
- [21] H. Lian, Convergence of functional k -nearest neighbor regression estimate with functional responses, *Electron. J. Stat.* 5 (2011) 31–40.
- [22] E. Masry, Nonparametric regression estimation for dependent functional data: asymptotic normality, *Stochastic Process. Appl.* 115 (2005) 155–177.
- [23] H.-G. Müller, J.-M. Chiou, X. Leng, Inferring gene expression dynamics via functional regression analysis, *BMC Bioinformatics* 9 (2008) 60.
- [24] J. Ramsay, C. Dalzell, Some tools for functional data analysis (with discussion), *J. R. Stat. Soc. Ser. B* 53 (1991) 539–572.
- [25] J. Ramsay, B. Silverman, *Applied Functional Data Analysis: Methods and Case Studies*, Springer, New York, 2002.
- [26] J. Ramsay, B. Silverman, *Functional Data Analysis*, second ed., Springer, New York, 2005.
- [27] C. Timmermans, L. Delsol, R. von Sachs, A new semimetric and its uses for nonparametric functional data analysis, in: F. Ferraty (Ed.), *Recent Advances in Functional Data Analysis and Related Topics*, in: *Contributions to Statistics*, Physica-Verlag, Springer, Heidelberg, 2011.
- [28] F. Yao, Functional principal component analysis for longitudinal and survival data, *Statist. Sinica* 17 (2007) 965–983.
- [29] V.V. Yurinskii, Exponential inequalities for sums of random vectors, *J. Multivariate Anal.* 6 (1976) 473–499.