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

**A B S T R A C T**

We consider a nonparametric regression model where the response \(Y\) and the covariate \(X\) 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:

\[
Y = r(X) + \epsilon,
\]

(1)

where \(Y\) is a response variable, \(X\) is a covariate and the error \(\epsilon\) satisfies \(E(\epsilon | X) = 0\). 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 \(X\) and \(Y\) 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 \(Y\) is scalar (i.e. when \(Y\) takes values on the real line \(\mathbb{R}\)).
When both the response $Y$ and the explanatory variable $X$ 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 $Y$ and the explanatory variable $X$ 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 $E$ be a functional space, $X$ be a random variable suitably defined on the measurable space $(\mathcal{E}, \mathcal{A})$, and let the hilbertian random variable $Y$ live in a measurable separable Hilbert space $(H, B)$ with $H$ endowed with inner product $\langle \cdot, \cdot \rangle$ and corresponding norm $\|g\|_E^2 = \langle g, g \rangle$, and with orthonormal basis $\{e_j : j = 1, \ldots, \infty \}$. Moreover, $E$ is endowed with a semi-metric $d(\cdot, \cdot)$ defining a topology to measure the proximity between two elements of $E$ and which is disconnected of the definition of $X$ in order to avoid measurability problems. This setting is quite general, since it includes the classical finite dimensional framework, but also the case where $E$ and/or $H$ 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 $E = H$. 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 $X$ and $Y$ live in the same space. So, we decided to present our results in the general framework where $E$ is not necessarily equal to $H$.

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

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

for the ball in $E$ with center $\chi$ and radius $t$, while in the Hilbert space $H$ we will denote

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

for the ball in $H$ with center $z$ and radius $r$.

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

$$ F_\varepsilon(t) = P(d(X, \chi) \leq t) = P(X \in B(\chi, t)), $$

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

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

where

$$ r_k(\chi) = E[(Y, e_k) \mid X = \chi] = (r(\chi), e_k). $$

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 $k, \ell \geq 1$:

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

where

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

Also, let

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

and

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

**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 $\varepsilon = \{((X_1, Y_1), \ldots, (X_n, Y_n))\}$ be a sample of i.i.d. data drawn from the distribution of the pair $(X, Y)$. The estimator of the regression operator $r$ is given by

$$
\hat{r}_h(\chi) = \frac{\sum_{i=1}^{n} Y_i K(h^{-1}d(X_i, \chi))}{\sum_{i=1}^{n} K(h^{-1}d(X_i, \chi))},
$$

where $\chi$ is a fixed element of $\varepsilon$. 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 $Y$) in [16]. Basically, this estimator is an average of the observed response $Y_i$ for which the corresponding $X_i$ is close to the new functional element $\chi$ at which the operator $r$ has to be estimated. The size of the neighborhood around $\chi$ 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 $\hat{r}_h(\chi)$ through the functions $F_X$ and $\tau_{0\chi}$ and more specifically through the following quantities:

$$
M_0X = K(1) - \int_0^1 (sK(s))' \tau_{0\chi}(s) ds,
$$

$$
M_1X = K(1) - \int_0^1 K'(s) \tau_{0\chi}(s) ds,
$$

and

$$
M_2X = K^2(1) - \int_0^1 (K^2)'(s) \tau_{0\chi}(s) ds.
$$

### 4. Asymptotic normality of $\hat{r}(\chi)$

From now on, $\chi$ 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 $\chi$, and $F_X(0) = 0$.

(C2) For some $\delta > 0$, all $0 \leq s \leq \delta$ and all $k \geq 1$, $\psi_{\chi,k}(0) = 0$, $\psi_{\chi,k}'(s)$ exists, and $\psi_{\chi,k}'(s)$ is uniformly Lipschitz continuous of order $0 < \alpha \leq 1$, i.e. there exists a $0 < L_k < \infty$ such that $|\psi_{\chi,k}'(s) - \psi_{\chi,k}'(0)| \leq L_k s^\alpha$ uniformly for all $0 \leq s \leq \delta$. Moreover, $\sum_{k=1}^{\infty} L_k < \infty$ and $\sum_{k=1}^{\infty} \psi_{\chi,k}'(0)^2 < \infty$.

(C3) For some $\delta > 0$, all $0 \leq s \leq \delta$ and all $k \geq 1$, $\psi_{\chi,k,k}(0) = 0$, and $\psi_{\chi,k,k}(s)$ is uniformly Lipschitz continuous of order $0 < \beta \leq 1$, i.e. there exists a $0 < N_k < \infty$ such that $|\psi_{\chi,k,k}(s)| \leq N_k s^\beta$ uniformly for all $0 \leq s \leq \delta$. Moreover, $\sum_{k=1}^{\infty} N_k < \infty$.

(C4) The bandwidth $h$ satisfies $h \to 0$, $nF_X(h) \to \infty$ and $h^{1+\alpha} F_X(h) = o(1)$.

(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, $\sup_{0 \leq s \leq 1} |\tau_{0\chi}(s) - \tau_{0\chi}(s)| = o(1)$, $M_0X > 0$ and $M_1X > 0$. 
(C7) \( \text{Var}(\|Y\| \mid X = \chi) < \infty \), and for all \( k \geq 1 \), \( E((Y - r(X), e_k)^4 \mid X = \chi) < \infty \).

Let

\[
\widehat{g}(\chi) = (nF_X(h))^{-1} \sum_{i=1}^{n} y_i k \left( \frac{d(X_i, \chi)}{h} \right)
\]

and

\[
\widehat{f}(\chi) = (nF_X(h))^{-1} \sum_{i=1}^{n} k \left( \frac{d(X_i, \chi)}{h} \right),
\]

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

\[
B_{n\chi} = h \frac{M_0}{M_1} \sum_{k=1}^{\infty} \varphi'_{\chi,k}(0) e_k
\]

and let \( C_{\chi} \) be the operator characterized by

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

where

\[
a_{kl} = \text{Cov}(\langle y, e_\ell \rangle, \langle y, e_k \rangle \mid X = \chi) \frac{M_2}{M_1^2}.
\]

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 \( \chi \in \mathcal{E} \),

\[
(nF_X(h))^{1/2} \left( \widehat{r}(\chi) - r(\chi) - B_{n\chi} \right) \xrightarrow{d} W_{\chi},
\]

where \( W_{\chi} \) follows a normal distribution on \( \mathcal{H} \) with zero mean and covariance operator \( C_{\chi} \).

**Remark 4.1.** The small ball probability \( F_X(h) \) 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 \( F_X(h) \) 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 \( h = O \left( (\log n)^{-\eta} \right) \) 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 \( d_{ij}(\cdot, \cdot) \) 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 \( e \) given \( X \) does not depend on \( X \): for any \( g, h \in \mathcal{H} \), \( E(\langle e, g \rangle \langle e, h \rangle \mid X) = E(\langle e, g \rangle \langle e, h \rangle) \). The bootstrap procedure consists of several steps:

1. For all \( i = 1, \ldots, n \), define \( \widehat{e}_{i,b} = y_i - \widehat{\gamma}_b(X_i) \), where \( b \) is a second smoothing parameter.
(2) Let \( \varepsilon_b = n^{-1} \sum_{i=1}^{n} \hat{e}_{i,b} \). Define new i.i.d. random elements \( \varepsilon^\text{boot}_1, \ldots, \varepsilon^\text{boot}_n \) by:

\[
P^\varepsilon (\varepsilon^\text{boot}_i = \hat{e}_{i,b} - \varepsilon_b) = n^{-1}
\]

for \( i, j = 1, \ldots, n \), where \( P^\varepsilon \) is the probability conditionally on the original sample \((X_1, Y_1), \ldots, (X_n, Y_n)\).

(3) For \( i = 1, \ldots, n \), let

\[
y^\text{boot}_i = \hat{r}_b(X_i) + \varepsilon_i^\text{boot}.
\]

and denote \( y^\text{boot} = (X_i, y^\text{boot}_i)_{i=1,\ldots,n} \).

(4) Define \( \hat{r}^\text{boot}_{hb}(\chi) = \frac{\sum_{i=1}^{n} y^\text{boot}_i \iota(h^{-1}(X_i, \chi))}{\sum_{i=1}^{n} \iota(h^{-1}(X_i, \chi))} \).

### 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 \( \varepsilon^\text{boot}_i = \hat{e}_{i,b} V_i \), where \( V_1, \ldots, V_n \) are i.i.d. real valued random variables that are independent of the data \((X_i, Y_i) (i = 1, \ldots, n)\) and that satisfy \( E(V_i) = 0 \) and \( E(V_i^2) = 1 \).

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 \( \alpha \) is defined in (C2)):

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

\[
\sup_{y \in \mathbb{R}} \left| P^\varepsilon \left( (nF_X(h))^{1/2} \left| \hat{r}^\text{boot}_{hb}(\chi) - \hat{r}_{k,b}(\chi) \right| \leq y \right) - P \left( (nF_X(h))^{1/2} \left| \hat{r}_{k,b}(\chi) - r(\chi) \right| \leq y \right) \right| \overset{a.s.}{\rightarrow} 0,
\]

where \( P^\varepsilon \) denotes probability, conditionally on the sample \( \varepsilon \) (i.e. \((X_i, Y_i), i = 1, \ldots, n\)). 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 \( \{e_1, e_2, \ldots\} \). 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 \((X_i, Y_i), i = 1, \ldots, n\).

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 \( \{\hat{e}_k : k = 1, \ldots, \infty\} \). We use the notation \( r_k(\chi) \) to denote the \( k \)-th component of the function \( r(\chi) \) with respect to this estimated basis, and similarly the estimators \( \hat{r}_{k,b}(\chi) \) and \( \hat{r}^\text{boot}_{hb}(\chi) \) are defined. We show below that Theorem 5.1 remains valid when the basis \( \{\hat{e}_k : k = 1, \ldots, \infty\} \) is employed.

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

\[
\sup_{y \in \mathbb{R}} \left| P^\varepsilon (nF_X(h))^{1/2} \left| \hat{r}^\text{boot}_{hb}(\chi) - \hat{r}_{k,b}(\chi) \right| \leq y \right) - P (nF_X(h))^{1/2} \left| \hat{r}_{k,b}(\chi) - r(\chi) \right| \leq y) \right| \overset{a.s.}{\rightarrow} 0.
\]

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 moment operator regression operator \( \Gamma_{r(X)}() = E(\langle r(X),.\rangle r(X)) \) which maps \( H \) onto \( H \). The orthonormal eigenfunctions \( e_1, e_2, \ldots \) of \( \Gamma_{r(X)}() \) are relevant directions for the hilbertian variable \( r(X) \). Indeed, for any fixed strictly positive integer \( K \), the eigenfunctions \( \psi_1, \ldots, \psi_K \) associated to the \( K \) largest eigenvalues minimize the quantity

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

over any orthonormal sequence \( \psi_1, \ldots, \psi_K \). 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 \( Y \) instead of \( r(X) \). This amounts to consider the orthonormal eigenfunctions \( \hat{e}_1, \hat{e}_2, \ldots \) associated to the eigenvalues \( \mu_1 \geq \mu_2 \geq \cdots \) of the second order moment operator \( \Gamma_Y = E(\langle Y, .\rangle Y) \) which can be estimated by its empirical version

\[
\hat{\Gamma}_{Y,n}(\cdot) = \frac{1}{n} \sum_{i=1}^{n} (Y_i, \cdot) Y_i.
\]

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

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

**Example 2.** Consider now the orthonormal eigenfunctions \( e_1, e_2, \ldots \) associated to the eigenvalues (in descending order) of \( \Gamma_{r(X)}() \). 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 \( \Gamma_{r(X),n}(\cdot) = \frac{1}{n} \sum_{i=1}^{n} (r(X_i), \cdot) r(X_i) \) as soon as \( r(X) \) is almost surely bounded (i.e. \( E\|r(X)\| < C \) 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 \( \hat{e}_1, \hat{e}_2, \ldots \) of the estimated second order moment regression operator \( \hat{\Gamma}_{\hat{r}(X)} = \frac{1}{n} \sum_{i=1}^{n} (\hat{r}(X_i), \cdot) \hat{r}(X_i) \).

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

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

\[
\hat{\Gamma}_{\hat{r}(X)} - \hat{\Gamma}_{\hat{r}(X),n} = \frac{1}{n} \sum_{i=1}^{n} \left( (\hat{r}(X_i), \cdot) \hat{r}(X_i) - (r(X_i), \cdot) r(X_i) \right)
\]

\[
= \frac{1}{n} \sum_{i=1}^{n} \left( (\hat{r}(X_i) - r(X_i), \cdot) (\hat{r}(X_i) - r(X_i)) + (r(X_i), \cdot) (\hat{r}(X_i) - r(X_i)) \right)
\]

\[
+ (\hat{r}(X_i) - r(X_i), \cdot) r(X_i). \right)
\]

A first way consists of focusing on the following inequality:

\[
\|\hat{\Gamma}_{\hat{r}(X)} - \hat{\Gamma}_{\hat{r}(X),n}\|_{\infty} \leq \sup_{\|x\|=1} \|\hat{r}(x) - r(x)\|^2 + 2 C \sup_{\|x\|=1} \|\hat{r}(x) - r(x)\|.
\]

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 \( (nF_{\hat{r}}(h))^{1/2} \sup_{\|x\|=1} \|\hat{r}(x) - r(x)\| = o(b/h) \) a.s.

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

\[
\frac{1}{n} \sum_{i=1}^{n} (r(X_i), \cdot) (\hat{r}(X_i) - r(X_i))
\]

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

Although these are still open problems, one might expect that such results will be stated in the near future.
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/npfda (click on the "Additional methods" item in the left menu).

Simulating functional responses and functional predictors. Let \( X_1, \ldots, X_n \) be \( n = 250 \) functional predictors such that

\[
X_i(t_j) = a_i \cos(\omega_i t_j) + \sum_{k=1}^{j} W_k,
\]

where \( a_1, \ldots, a_n \) (resp. \( \omega_1, \ldots, \omega_n \)) are \( n \) independent real random variables (r.r.v.) uniformly distributed over \([1.5; 2.5]\) (resp. \([0.2; 2]\)), \( 0 = t_1 < t_2 < \cdots < t_{99} < t_p = \pi \) are \( p = 100 \) equally spaced measurements and the \( W_k \)'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 \( \{X_1, X_2, X_3\} \). 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 \( i = 1, \ldots, n \) and \( j = 1, \ldots, 100 \):

\[
r(X_i)(t_j) = \int_0^{t_j} X_i(u)^2 \, du.
\]

The building of the functional response \( Y \) follows the following scheme for all \( i = 1, \ldots, n \) and \( j = 1, \ldots, 100 \):

\[
y_i(t_j) = r(X_i)(t_j) + \varepsilon_i(t_j),
\]

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

\[
\varepsilon_i^{\text{add}}(t_j) \sim N\left(0, \sqrt{\text{snr} \times \text{tr}(T_{\tau_i(X_i,n)})}\right).
\]
where $\Gamma_{r(X),n} = \frac{1}{n} \sum_{i=1}^{n} (r(X_i), \cdot) r(X_i)$ is the empirical covariance operator of $r(X)$ (with $r_i(X_i) = r(X_i) - (1/n) \sum_{i=1}^{n} r(X_i)$), tr($\cdot$) stands for the standard trace operator and $\text{snr}$ is the signal-to-noise ratio. Let $\Gamma_{\text{add}}^n = \frac{1}{n} \sum_{i=1}^{n} (\epsilon_{\text{add}i}, \cdot) \epsilon_{\text{add}i}$ be the empirical covariance operator of the additive error. Then one has:

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

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

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

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

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

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

where $\Gamma_{\text{struct}}^n$ is the covariance operator of the structural error.

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

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

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

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

Fig. 2 gives an idea of how the error-type affects the regression. Finally, in our simulations, we use the mixed error with $\text{snr} = 5\%$ and $\rho = 0.3$. In order to valid 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: $e_1, e_2, \ldots$ (resp. $\tilde{e}_1, \tilde{e}_2, \ldots$) are the orthonormal eigenfunctions associated to the eigenvalues (in descending order) of $\Gamma_{r(X)}$ (resp. $\Gamma_{r(X),n}$). As a by-product, a third
Theorem 5.1 by comparing, for each component, the quality of componentwise predictions is quite good:

\[ \text{quality of componentwise predictions} = 99.7\% \]

where \( k \) is the direction in which the testing sample is incremented. The true regression operator is represented 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 \( e_i \) amounts to plotting the 50 points \( (r(X_i), \hat{\sigma}_i, (\hat{\rho}_h(X_i), \hat{\sigma}_i)) \) for \( i = 201, \ldots, 250 \). Fig. 4 proposes a componentwise prediction graph for the four first components. The percentage of variance explained by these 4 components is 99.7% (i.e. 0.997 = \( \frac{\sum_{k=1}^{4} \lambda_k}{\sum_{k=4}^{100} \lambda_k} \), where \( \lambda_1 > \lambda_2 > \cdots \) denotes the eigenvalues of \( \Gamma_h(X_i,n) \). The quality of componentwise predictions is quite good for each component. Investigating the bootstrap method. We illustrate now Theorem 5.1 by comparing, for \( k = 1, \ldots, 4 \), the density function \( f_{k,X}^{\text{true}} \) of the componentwise bootstrapped error

\[ f_{k,X}^{\text{boot}}(\chi) - f_{k,X}^{\text{true}}(\chi) \]

with the density function \( f_{k,X}^{\text{true}} \) of the componentwise true error

\[ f_{k,X}^{\text{true}}(\chi) - r(\chi, \hat{\sigma}_i) \]

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

1. build 200 samples \( \{(X_i^1, Y_i^1)_{i=1,\ldots,200}\} \) for \( i = 1, \ldots, 200 \)
2. carry out 200 estimates \( \{(\hat{\rho}_h(X_i) - r(\chi, \hat{\sigma}_i))_{i=1,\ldots,200}\} \) where \( \hat{\rho}_h \) is the functional kernel estimator of the regression operator \( r(\cdot, \cdot) \) derived from the \( j \)-th sample \( (X_i^j, Y_i^j)_{i=1,\ldots,200} \).
Fig. 4. Componentwise prediction quality.

3. compute a standard density estimator over the 200 values
\[ \{ \hat{r}_{b}^{(s)}(\chi) - r(\chi), \hat{r}_{b}(\chi) \}_{s=1,...,200}. \]

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

1. compute \( \hat{r}_{b}(\chi) \) over the initial sample \( \delta = (X_{i}, y_{i})_{i=1,...,200} \).
2. repeat 200 times the bootstrap algorithm over \( \delta \) by using i.i.d. random variables \( V_{1}, V_{2}, \ldots \), 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 \( \{ \hat{r}_{b}^{\text{boot1}}(\chi) - \hat{r}_{b}(\chi), \hat{e}_{k} \}, \ldots, \{ \hat{r}_{b}^{\text{boot200}}(\chi) - \hat{r}_{b}(\chi), \hat{e}_{k} \} \).

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. \( k = 1, \ldots, 4 \)). The first row corresponds to the fixed curves \( \chi = X_{201}, \) the second to \( \chi = X_{202}, \ldots \), the fifth to \( \chi = X_{205} \). 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. \( \text{dist}_{k,\chi} = 0.5 \int |f_{k,\chi}^{\text{true}}(t) - f_{k,\chi}^{\text{boot}}(t)| \, dt \in [0, 1] \)), at a fixed curve \( \chi \) in \{ \&_{201}, \ldots, \&_{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 $\text{dist}_{X_{201},k}, \ldots, \text{dist}_{X_{250},k}$. 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 $k (k = 1, 2, \ldots)$, confidence intervals $I_k$ such that

$$ P \left( \hat{r}_k(\chi) \in I_k \right) = 1 - \alpha_k, $$

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

$$ P \left( \bigcap_{k=1}^{K} \hat{r}_k(\chi) \in I_k \right) \geq 1 - \alpha, $$
as soon as, for $k = 1, \ldots, K$, one sets $\alpha_k = \alpha/K$ with $\alpha \in (0, 1)$. This last inequality amounts to the following one:

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

with $\tilde{E}^\alpha = \left\{ \sum_{k=1}^{K} a_k \tilde{e}_k(\cdot), (a_1, \ldots, a_K) \in \tilde{E}_1^{\alpha_1} \times \cdots \times \tilde{E}_K^{\alpha_K} \right\}$. 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 $\tilde{e}_1, \ldots, \tilde{e}_K$. 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 $r(X)$). 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 $\tilde{e}_1, \ldots, \tilde{e}_K$ are very close for this example. Of course, when one replaces the data-driven basis with the eigenfunctions of $\mathcal{H}$, 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 $\rightarrow r(\chi)$; thick dotted line $\rightarrow \tilde{r}(\chi)$; 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

$$\left( nF_\chi(h) \right)^{1/2} \left[ \frac{\hat{g}(\chi)}{f(\chi)} - r(\chi) - B_n \right]$$

$$= \left( nF_\chi(h) \right)^{1/2} \left[ \frac{\hat{g}(\chi)}{f(\chi)} - \frac{E\hat{g}(\chi)}{Ef(\chi)} \right] + \left( nF_\chi(h) \right)^{1/2} \left[ \frac{E\hat{g}(\chi)}{Ef(\chi)} - r(\chi) - B_n \right]. \quad (A.1)$$
We will first show that the second term in right hand side of (A.1) is negligible. Consider

\[
\frac{E\hat{g}(\chi)}{Ef(\chi)} - r(\chi) = \frac{E \left( (Y - r(\chi))K \left( \frac{d(X, \chi)}{h} \right) \right)}{E \left( K \left( \frac{d(X, \chi)}{h} \right) \right)} = \sum_{k=1}^{\infty} \frac{E \left( \varphi_{X, k}(d(X, \chi))K \left( \frac{d(X, \chi)}{h} \right) \right) \epsilon_k}{E \left( K \left( \frac{d(X, \chi)}{h} \right) \right)}. \tag{A.2}
\]

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

\[
h \sum_{k=1}^{\infty} \left( \varphi_{X, k}(0) \frac{M_{0x}}{M_{1x}} + R_{n,k} \right) \epsilon_k,
\]

where the remainder term \( R_{n,k} \) satisfies

\[
|R_{n,k}| = h \left| \frac{\int \left[ \varphi_{X, k}(\xi_t, k) - \varphi_{X, k}(0) \right] K(t) dP_{d(X, \chi)/h}(t)}{\int K(t) dP_{d(X, \chi)/h}(t)} \right| \leq h^{1+\alpha} L_k \left| \frac{\int t^n K(t) dP_{d(X, \chi)/h}(t)}{\int K(t) dP_{d(X, \chi)/h}(t)} \right| \leq h^{1+\alpha} L_k,
\]

with \(|\xi_t| \leq \alpha \) 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 \( R_{n,x} = \sum_{k=1}^{\infty} R_{n,k} \epsilon_k \). Then, it follows that the second term in right hand side of (A.1) equals \( (nF_{x}(h))^{1/2} R_{n,x} \). Next, note that

\[
nF_{x}(h) \|R_{n,x}\|^2 = nF_{x}(h) \sum_{k=1}^{\infty} R_{n,k}^2 \leq nF_{x}(h) \sum_{k=1}^{\infty} \sum_{k=1}^{\infty} \epsilon_k^2 = o(1),
\]

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

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

Again by Slutsky’s theorem and by noting that \( \hat{f}(\chi) \to M_{1x} \) a.s (see Lemma 4 in [13]), and that \( E\hat{g}(\chi)/Ef(\chi) - r(\chi) = B_{nx} + O(h^{1+\alpha}) = o(1) \), the latter expression has the same asymptotic distribution as

\[
\left( \frac{nF_{x}(h)}{M_{1x}} \right)^{1/2} \left[ \frac{\hat{g}(\chi) - E\hat{g}(\chi)}{\hat{g}(\chi) - Ef(\chi)} \right] \left( \hat{f}(\chi) - Ef(\chi) \right) \left( Ef(\chi) \right) \left( \frac{\hat{g}(\chi) - E\hat{g}(\chi)}{\hat{g}(\chi) - Ef(\chi)} \right) \left( \hat{f}(\chi) - Ef(\chi) \right) \left( Ef(\chi) \right).
\]

where for \( 1 \leq i \leq n \),

\[
Z_{ni} = \frac{1}{M_{1x}} (nF_{x}(h))^{1/2} \left[ \frac{\hat{g}(\chi) - E\hat{g}(\chi)}{\hat{g}(\chi) - Ef(\chi)} \right] \left( \hat{f}(\chi) - Ef(\chi) \right) \left( Ef(\chi) \right) \left( \frac{\hat{g}(\chi) - E\hat{g}(\chi)}{\hat{g}(\chi) - Ef(\chi)} \right) \left( \hat{f}(\chi) - Ef(\chi) \right) \left( Ef(\chi) \right).
\]
Let us now calculate the covariance operator of $S_n = \sum_{i=1}^{n} Z_i$:

$$E \left[ \langle S_n - ES_n, e_k \rangle \langle S_n - ES_n, e_\ell \rangle \right] = \sum_{i=1}^{n} E \left[ \langle Z_i - EZ_i, e_k \rangle \langle Z_i - EZ_i, e_\ell \rangle \right]$$

$$= \sum_{i=1}^{n} E \left[ \langle Z_i - E(Z_i|X_i), e_k \rangle \langle Z_i - E(Z_i|X_i), e_\ell \rangle \right]$$

$$+ \sum_{i=1}^{n} E \left[ \langle E(Z_i|X_i) - EZ_i, e_k \rangle \langle E(Z_i|X_i) - EZ_i, e_\ell \rangle \right]$$

$$= \frac{1}{M_1^2 F_X(h)} E \left[ (y - r(\chi), e_k)(y - r(\chi), e_\ell)K^2 \left( \frac{d(\chi, \chi)}{h} \right) \right]$$

$$+ \frac{1}{M_2^2 F_X(h)} \text{Cov} \left[ (r(\chi) - r(\chi), e_k)K \left( \frac{d(\chi, \chi)}{h} \right) \right] \times (r(\chi) - r(\chi), e_\ell)K \left( \frac{d(\chi, \chi)}{h} \right)$$

(A.3)

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

$$\frac{1}{M_1^2 F_X(h)} \text{Cov} \left[ \varphi_{X,k}(d(\chi, \chi))K \left( \frac{d(\chi, \chi)}{h} \right), \varphi_{X,\ell}(d(\chi, \chi))K \left( \frac{d(\chi, \chi)}{h} \right) \right]$$

$$= \frac{1}{M_1^2 F_X(h)} \left[ \int_{0}^{1} \varphi_{X,k}(ht)\varphi_{X,\ell}(ht)K^2(t) \, dp^{d(\chi,\chi)/h}(t) \right]$$

$$- \int_{0}^{1} \varphi_{X,k}(ht)K(t) \, dp^{d(\chi,\chi)/h}(t) \times \int_{0}^{1} \varphi_{X,\ell}(ht)K(t) \, dp^{d(\chi,\chi)/h}(t)$$

$$= a_{nk \ell 2} \quad \text{(say)}.$$

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

$$\frac{1}{M_1^2 F_X(h)} E \left[ s_{nk}(\chi)K^2 \left( \frac{d(\chi, \chi)}{h} \right) \right] := a_{nk \ell 1} \quad \text{(say)}.$$

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

(i) $\lim_{n \to \infty} a_{nk \ell} = a_{nk \ell 1}$ for all $k, \ell \geq 1$.

(ii) $\lim_{n \to \infty} \sum_{k=1}^{\infty} a_{nk \ell k} = \sum_{k=1}^{\infty} a_{kk} < \infty$.

(iii) $\forall \delta > 0, \forall k \geq 1: \lim_{n \to \infty} \sum_{i=1}^{n} E \left( |Z_i, e_k|^2 1\{|Z_i, e_k| > \delta\} \right) = 0$.

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

$$a_{nk \ell 1} = \frac{1}{M_1^2 F_X(h)} s_{nk}(\chi) E \left[ K^2 \left( \frac{d(\chi, \chi)}{h} \right) \right] (1 + o(1))$$

$$= \frac{1}{M_1^2 F_X(h)} s_{nk}(\chi) F_X(h) M_2 X (1 + o(1))$$

$$= a_{kk} + o(1),$$

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 $a_{kk}$ should be summable. This can be achieved using arguments similar to those used for the bias term $B_{nX}$. In fact, using assumption (C3) we have that

$$a_{nk \ell 1} = \frac{1}{M_1^2 F_X(h)} E \left[ \varphi_{X,kk}(d(\chi, \chi))K^2 \left( \frac{d(\chi, \chi)}{h} \right) \right] + \frac{M_2 X s_{kk}(\chi)}{M_1^2 F_X(h)} (1 + o(1))$$

$$\leq \frac{1}{M_1^2 F_X(h)} N_k E \left[ (d(\chi, \chi))^{\beta} K^2 \left( \frac{d(\chi, \chi)}{h} \right) \right] + \frac{M_2 X s_{kk}(\chi)}{M_1^2 F_X(h)} (1 + o(1)).$$
Hence,

$$\sum_{k=1}^{\infty} a_{bkk,1} = O(h^p) \sum_{k=1}^{\infty} N_k + \frac{M_{2X}}{M_{1X}^2} \sum_{k=1}^{\infty} s_{kk}(\chi) (1 + o(1))$$

$$= \left( \sum_{k=1}^{\infty} a_{bk} \right) (1 + o(1)),$$

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

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

provided that $$\sum_{k=1}^{\infty} \Phi_k(k)^2 < \infty$$ and $$\sum_{k=1}^{\infty} I_k^2 < \infty$$ (see condition (C2)). This shows that (ii) is valid, since

$$\sum_{k=1}^{\infty} a_{bk} = \sum_{k=1}^{\infty} \text{Var}(Y, e_k | X = \chi) \frac{M_{2X}}{M_{1X}^2} = \text{Var}(\|y\| | X = \chi) \frac{M_{2X}}{M_{1X}^2}.$$

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

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

$$\sum_{i=1}^{n} E \left( (Z_{ni} - EZ_{ni}, e_k)^2 I[|Z_{ni} - EZ_{ni}, e_k| > \delta] \right)$$

$$\leq \sum_{i=1}^{n} \left[ E \left( (Z - EZ, e_k)^4 \right) P[|Z - EZ, e_k| > \delta] \right]^{1/2}$$

$$\leq 16F_X(h)^{-1} \left[ E \left( (Y - r(\chi), e_k)^4 K \left( \frac{d(\chi, \chi)}{h} \right) \right) P \left( (Y - r(\chi), e_k) K \left( \frac{d(\chi, \chi)}{h} \right) > \delta (nF_X(h))^{1/2} \right) \right]^{1/2}$$

$$\leq 16F_X(h)^{-1} \left[ E \left( (Y - r(\chi), e_k)^4 | X = \chi \right) \right]^{1/2} + o(1) \left[ E \left( K^4 \left( \frac{d(\chi, \chi)}{h} \right) \right) \right]^{1/2}$$

$$\times \text{Var} \left( (Y - r(\chi), e_k) K \left( \frac{d(\chi, \chi)}{h} \right) \right) \delta^{-1} (nF_X(h))^{-1/2},$$

and this tends to zero since $$E((Y - r(\chi), e_k)^4 | X = \chi) < \infty, E[K^4(d(\chi, \chi)/h)] = O(F_X(h))$$ and $$\text{Var}((Y - r(\chi), e_k) K(d(\chi, \chi)/h)) = O(F_X(h)).$$

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 $$C_X.$$

**A.2. Proof of Theorem 6.1**

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

$$P^4 \left( a_n \left( \hat{\tau}_{k,h}^\text{boot}(\chi) - r_{k,h}(\chi) \right) \leq y \right) - P \left( a_n \left( \hat{\tau}_{k,h}(\chi) - r_{k}(\chi) \right) \leq y \right)$$

$$= P^4 \left( a_n \left( \hat{\tau}_{k,h}^\text{boot}(\chi) - \hat{\tau}_{k,h}(\chi) \right) \leq y \right) - P \left( a_n \left( \hat{\tau}_{k,h}^\text{boot}(\chi) - \hat{\tau}_{k,h}(\chi) \right) \leq y \right)$$

$$+ P^4 \left( a_n \left| \hat{\tau}_{k,h}^\text{boot}(\chi) - \hat{\tau}_{k,h}(\chi) \right| \leq y \right) - P \left( a_n \left| \hat{\tau}_{k,h}(\chi) - r_{k}(\chi) \right| \leq y \right)$$

$$+ P \left( a_n \left| \hat{\tau}_{k,h}(\chi) - r_{k}(\chi) \right| \leq y \right) - P \left( a_n \left| \hat{\tau}_{k,h}(\chi) - r_{k}(\chi) \right| \leq y \right). \quad (A.4)$$
The second term of (A.4) is $o(1)$ a.s. by Theorem 5.1. For the third term, note that if we write $Z_n = a_n (\hat{\tau}_h(\chi) - r(\chi))$, then $a_n (\hat{\tau}_{k,b}(\chi) - \hat{\tau}_h(\chi)) = (Z_n, \hat{e}_k)$ and $a_n (\hat{\tau}_{k,b}(\chi) - r(\chi)) = (Z_n, e_k)$. Hence,

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

since $\|\hat{e}_k - e_k\| = o_p(1)$ 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

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

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

$$a_n [E^{\hat{\tau}_{k,b}(\chi)}] - \hat{\tau}_{k,b}(\chi) - E^{\hat{\tau}_{k,b}(\chi)} + \hat{\tau}_{k,b}(\chi) = o(1), \quad \text{a.s.,}$$

and

$$\frac{\text{Var}[\hat{\tau}_{k,b}(\chi)]}{\text{Var}[\hat{\tau}_{k,b}(\chi)]} \to 1, \quad \text{a.s.}$$

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

$$\frac{(nF_x(h))^{-1}}{\hat{b}_h(\chi)} \sum_{i=1}^n \left\{ \hat{\tau}_{k,b}(\chi_i) - \hat{\tau}_{k,b}(\chi) - \hat{\tau}_{k,b}(\chi_i) + \hat{\tau}_{k,b}(\chi) \right\} K \left( h^{-1} d(\chi_i, \chi) \right)$$

$$= \frac{(nF_x(h))^{-1}}{\hat{b}_h(\chi)} \sum_{i=1}^n \left\{ \hat{\tau}_{b}(\chi_i) - \hat{\tau}_{b}(\chi), \hat{e}_k - e_k \right\} K \left( h^{-1} d(\chi_i, \chi) \right)$$

$$= \frac{(nF_x(h))^{-1}}{\hat{b}_h(\chi)} \sum_{i=1}^n \left[ K_b(d(\chi_i, \chi)) - K_b(d(\chi_j, \chi)) \right] \left( \sum_{\ell} K_b(d(\chi_\ell, \chi)) - \sum_{\ell} K_b(d(\chi_\ell, \chi)) \right) \times \langle y_j, \hat{e}_k - e_k \rangle,$$
where $\alpha$ is defined in condition (C2). It now follows that the absolute value of (A.7) multiplied by $(nF_X(h))^{1/2}$ is asymptotically bounded by
\[
\left[ C(nF_X(h))^{-1/2} + Ch^\alpha + \hat{f}_b^{-1}(\chi) \frac{h}{b} \right] (nF_X(h))^{-1/2} (nF_X(b))^{-1} \\
\times \hat{f}_b^{-1}(\chi) \sum_{i=1}^n \left( d(\chi, \chi) \right) \left( \left\| y_i \right\| \left\| \hat{\epsilon}_k - e_k \right\| \right)
\]
\[
\leq \left[ C + \frac{h}{b} (nF_X(h))^{1/2} \right] \left\| \hat{\epsilon}_k - e_k \right\| E \left( \left\| y \right\| \right) = \frac{F_X(b+h)}{F_X(b)} (1 + o(1)),
\]
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):
\[
\text{Var}^4[\hat{\epsilon}_{k,b,\chi}(\chi)] - \text{Var}^4[\hat{\epsilon}_{k,hb,\chi}(\chi)]
\]
\[
= \frac{1}{(nF_X(h))^{2f_\chi(\chi)}} \sum_{i=1}^n \left[ \left( y_i - \hat{\epsilon}_b(\chi_i), \hat{\epsilon}_b \right)^2 - \left( y_i - \hat{\epsilon}_b(\chi_i), e_k \right)^2 \right] K^2(h^{-1}d(\chi, \chi))
\]
\[
= \frac{1}{(nF_X(h))^{2f_\chi(\chi)}} \sum_{i=1}^n \left( y_i - \hat{\epsilon}_b(\chi_i), \hat{\epsilon}_b - e_k \right) (y_i - \hat{\epsilon}_b(\chi_i), \hat{\epsilon}_b + e_k) K^2(h^{-1}d(\chi, \chi))
\]
\[
\leq \frac{1}{(nF_X(h))^{2f_\chi(\chi)}} \sum_{i=1}^n K^2(h^{-1}d(\chi, \chi)) \left( \left\| y_i - \hat{\epsilon}_b(\chi_i) \right\| \left\| \hat{\epsilon}_b - e_k \right\| \left( \left\| e_k \right\| \right) \right)
\]
\[
\leq \frac{M_{2X}^2}{nF_X(h)} E \left( \left\| y - r(h) \right\| \right)^2 \left( \left\| \hat{\epsilon}_b - e_k \right\| \left( \left\| e_k \right\| \right) \right) (1 + o(1)),
\]
which is $o((nF_X(h))^{-1})$ a.s., since $\left\| \hat{\epsilon}_b - e_k \right\| = o(1)$ a.s. Note that when the naive bootstrap is used, we have that (where $\hat{\sigma}_{r,k}^2 = n^{-1} \sum_{i=1}^n (\hat{\epsilon}_i - \hat{\epsilon}_k)^2$ and similarly for $\hat{\sigma}_{r,k}^2$):
\[
\text{Var}^4[\hat{\epsilon}_{k,b,\chi}(\chi)] - \text{Var}^4[\hat{\epsilon}_{k,hb,\chi}(\chi)] = \frac{1}{(nF_X(h))^{2f_\chi(\chi)}} \sum_{i=1}^n K^2(h^{-1}d(\chi, \chi)) (\hat{\sigma}_{r,k}^2 - \hat{\sigma}_{r,k}^2)
\]
\[
\leq \frac{M_{2X}^2}{nF_X(h)} E \left( \left\| y - r(h) \right\| \right)^2 \left( \left\| \hat{\epsilon}_b - e_k \right\| \left( \left\| e_k \right\| \right) \right) (1 + o(1)) = o((nF_X(h))^{-1}).
\]
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
\[
\left\| \hat{\epsilon}_k - e_k \right\| \leq C_k \left\| y_{\chi} - y_{\chi} \right\|, \tag{A.8}
\]
where $\left\| \cdot \right\|_{\infty}$ is the standard operator norm (i.e. $\left\| U \right\|_{\infty} = \sup_{\left|x \right| = 1} \left\| U(x) \right\|$). The proof of Proposition 1 is then based on the following intermediate result:

**Lemma 1.** Let $y_1, \ldots, y_n$ be an i.i.d. sequence of hilbertian random variables and assume that there exists $M > 0$ such that
\[
E \left( \left\| y_1 \right\|^{2m} \right) \leq Mm!. \text{ Then, we have:}
\]
\[
\left\| y_{\chi} - y_{\chi} \right\| = O_{a.s.} \left( \sqrt{\frac{\log n}{n}} \right). \tag{A.9}
\]

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 $k$,
\[
\left\| \hat{\epsilon}_k - e_k \right\| = O_{a.s.} \left( \sqrt{\frac{\log n}{n}} \right).
\]
In order to end the proof of Proposition 1, it remains to prove that $\sqrt{\log n} n = o \left( (b/h)(nF_X(h))^{-1/2} \right)$ or equivalently $(h/b) \sqrt{\log n F_X(h)} = o(1)$. 

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

Case 2: $\log n F_X(h) \to +\infty$. Theorem 5.1 assumes $b^{1+\epsilon} \sqrt{n F_X(h)} \to 0$ 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 $b h^{2-1} = O(1)$, which is equivalent to $h = (b h^{-1} + \varepsilon)$ with $\varepsilon > 0$ and this leads us to $h/b = o\left((\log n/n)^{-\epsilon}\right)$ with $\epsilon" > 0$. Then, it is easy to see that $(h/b)^{\epsilon} / \sqrt{\log n} = o\left(n^{-\epsilon"} (\log n)^{-\epsilon"+1/2}\right) = o(1)$ and $(h/b)\sqrt{\log n F_X(h)} = o(1)$ holds. $\Box$

### A.4. Proof of Proposition 2

The proof is based on the following decomposition:

$$\|G_{\hat{h}(X)} - G_{r(X)}\|_\infty \leq \|G_{\hat{h}(X)} - G_{r(X),n}\|_\infty + \|G_{r(X),n} - G_{r(X)}\|_\infty. \quad (A.10)$$

By using (A.9) with $r(X)$ instead of $y$, we obtain:

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

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

$$\|G_{\hat{h}(X)} - G_{r(X)}\|_\infty = o\left(b_{h/\sqrt{n F_X(h)}} + O\left(\sqrt{\frac{\log n}{n}}\right), \quad \text{a.s.}ight.$$  

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

### References


