Recursive estimation of nonparametric regression with functional covariate.

Aboubacar AMIRI[‡]   Christophe CRAMBES[*†]   Baba THIAM[‡]

Abstract

The main purpose is to estimate the regression function of a real random variable with functional explanatory variable by using a recursive nonparametric kernel approach. The mean square error and the almost sure convergence of a family of recursive kernel estimates of the regression function are derived. These results are established with rates and precise evaluation of the constant terms. Also, a central limit theorem for this class of estimators is established. The method is evaluated on simulations and real data set studies.

1 Introduction

Functional data analysis is a branch of statistics that has been studied frequently and developed over recent years. This type of data appears in many practical situations such as continuous phenomena. Thus, the possible application fields that favor the use of functional data are broad and include the following: climatology, economics, linguistics, medicine, and so on. Because the pioneering work of authors such as Ramsay and Dalzell (1991) and Frank and Friedman (1993), many developments have been investigated to build theories and methods regarding functional data; for instance, how can the mean or the variance of functional data be defined? What type of model can be considered with functional data? These papers also emphasize the drawback of merely using multivariate methods with this type of

---

[*]Corresponding author: christophe.crambes@univ-montp2.fr

[†]Université Montpellier 2, Institut de Mathématiques et de Modélisation de Montpellier, Place Eugène Bataillon, 34090 Montpellier, France. Email: christophe.crambes@univ-montp2.fr

[‡]Université Lille Nord de France, Université Lille 3, Laboratoire EQUIPPE, EA 4018, Villeneuve d'Ascq, France. Emails: aboubacar.amiri@univ-lille3.fr, baba.thiam@univ-lille3.fr

data; rather, they suggest considering these data as objects that belong to a specific functional space. The monographs of Ramsay and Silverman (2005, 2002) provide and overview of both the theoretical and practical aspects of functional data analysis.

Regression is one of the most studied functional models. In this model, the variable of interest Y is real and the covariate $\mathcal{X}$ belongs to a functional space $\mathcal{E}$ endowed with a semi-norm $\|\cdot\|$. Thus, the regression model is written

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

where $r : \mathcal{E} \longrightarrow \mathbb{R}$ is an operator and ε is the random error term. Much research has investigated this model when the operator r is supposed to be linear, which has contributed to the popularity of the so-called functional linear model. In this linear context, the operator r is written as $\langle \alpha, . \rangle$ where $\langle ., . \rangle$ denotes the inner product of the space $\mathcal{E}$ and α belongs to $\mathcal{E}$. Thus, the goal is to estimate the unknown function α. We refer the reader to Cardot et al. (2003) and Crambes et al. (2009) for different methods of estimating α. Another way to approach the (1) model is to think in a nonparametric way. Many authors have also investigated this direction. Recent advances on the topic have been the subject of a bibliographical review in Gonzàlez Manteiga and Vieu (2007) and monographs by Ferraty and Vieu (2006), Ferraty and Romain (2010), thereby providing the theoretical and practical properties of a kernel estimator of the operator r. Specifically, if $(\mathcal{X}_i, Y_i)_{i=1,\dots,n}$ is a sample of independent and identically distributed couples with the same law as $(\mathcal{X}, Y)$, then, this kernel estimator is defined for all $\chi \in \mathcal{E}$ by

$$r_n(\chi) := \frac{\displaystyle\sum_{i=1}^{n} Y_i K\left(\frac{\|\chi - \mathcal{X}_i\|}{h}\right)}{\displaystyle\sum_{i=1}^{n} K\left(\frac{\|\chi - \mathcal{X}_i\|}{h}\right)}, \tag{2}$$

where K is a kernel and $h > 0$ is a bandwidth. Masry (2005) considered the asymptotic normality of (2) in the dependent case, while Ling and Wu (2012) obtained almost sure convergence. This nonparametric regression estimator raises several problems because the choice of the semi-norm $\|\cdot\|$ of the space $\mathcal{E}$ and the choice of the bandwidth, ... With regard to bandwidth, many solutions have been considered when the covariate is real (e.g., cross validation). Recently, Amiri (2012) studied an estimator using a sequence of bandwidths in the multivariate setting that allowed the computation of this estimator in a recursive way thereby generalizing previous research (Devroye and Wagner (1980) and Ahmad and Lin (1976)). This estimator has acceptable theoretical properties (from the point of view of the mean square error and almost sure convergence). It is also of practical interest: for instance, it presents a computational time gain when researchers want to predict new

values of the variable of interest when new observations appear. This case is not true for the basic kernel estimator, which must be computed again using the whole sample. The purpose of the current study is to adapt the recursive estimator studied in Amiri (2012) to a case in which the covariate is functional.

The remainder of the paper is organized as follows. Section 2 defines the recursive estimator of the operator r when the covariate $\mathcal{X}$ is functional and presents the asymptotic properties of this estimator. Section 3 evaluates the performance of our estimator using a simulation study and a real dataset. Finally, the proofs of the theoretical results are presented in Section 4.

2 Functional regression estimation

2.1 Notations and assumptions

Let $(\mathcal{X}, Y)$ be a pair of random variables defined in $(\Omega, \mathcal{A}, P)$, with values on $\mathcal{E} \times \mathbb{R}$, where $\mathcal{E}$ is a Banach space endowed with a semi-norm $\|\cdot\|$. Assume that $(\mathcal{X}_i, Y_i)_{i=1,\ldots,n}$ is a sample of n random variables independent and identically distributed, having the same distribution as $(\mathcal{X}, Y)$. The (1) model is then rewritten as

$$Y_i = r(\mathcal{X}_i) + \varepsilon_i, \quad i = 1, \ldots, n,$$

where for any $i = 1, \ldots, n$, ε_i is a random variable such that $\mathrm{E}(\varepsilon_i | \mathcal{X}_i) = 0$ and $\mathrm{E}(\varepsilon_i^2 | \mathcal{X}_i) = \sigma_\varepsilon^2(\mathcal{X}_i) < \infty$.

Nonparametric regression aims to estimate the functional $r(\chi) := \mathrm{E}(Y | \mathcal{X} = \chi)$, for $\chi \in \mathcal{E}$. To this end, let us consider the family of recursive estimators indexed by a parameter $\ell \in [0, 1]$, and defined by

$$r_n^{[\ell]}(\chi) := \frac{\sum\limits_{i=1}^{n} \frac{Y_i}{F(h_i)^\ell} K\left(\frac{\|\chi - \mathcal{X}_i\|}{h_i}\right)}{\sum\limits_{i=1}^{n} \frac{1}{F(h_i)^\ell} K\left(\frac{\|\chi - \mathcal{X}_i\|}{h_i}\right)},$$

where K is a kernel, (h_n) a sequence of bandwidths and F the cumulative distribution function of the random variable $\|\chi - \mathcal{X}\|$. Our family of estimators is a recursive modification of the estimate defined in (2) and can be computed recursively by

$$r_{n+1}^{[\ell]}(\chi) = \frac{\left[\sum\limits_{i=1}^{n} F(h_i)^{1-\ell}\right] \varphi_n^{[\ell]}(\chi) + \left[\sum\limits_{i=1}^{n+1} F(h_i)^{1-\ell}\right] Y_{n+1} K_{n+1}^{[\ell]}(\|\chi - \mathcal{X}_{n+1}\|)}{\left[\sum\limits_{i=1}^{n} F(h_i)^{1-\ell}\right] f_n^{[\ell]}(\chi) + \left[\sum\limits_{i=1}^{n+1} F(h_i)^{1-\ell}\right] K_{n+1}^{[\ell]}(\|\chi - \mathcal{X}_{n+1}\|)},$$

with

3

$$\varphi_n^{[\ell]}(\chi) = \frac{\sum\limits_{i=1}^{n} \frac{Y_i}{F(h_i)^\ell} K\left(\frac{\|\chi - \mathcal{X}_i\|}{h_i}\right)}{\sum\limits_{i=1}^{n} F(h_i)^{1-\ell}}, \ \ f_n^{[\ell]}(\chi) = \frac{\sum\limits_{i=1}^{n} \frac{1}{F(h_i)^\ell} K\left(\frac{\|\chi - \mathcal{X}_i\|}{h_i}\right)}{\sum\limits_{i=1}^{n} F(h_i)^{1-\ell}}, \quad (3)$$

and $K_i^{[\ell]}(\cdot) := \frac{1}{F(h_i)^\ell \sum\limits_{j=1}^{i} F(h_j)^{1-\ell}} K\left(\frac{\cdot}{h_i}\right)$.

More precisely, $r_n^{[\ell]}(\chi)$ is the adaption to the functional model of the finite-dimensional recursive family of estimators introduced by Amiri (2012), which includes the well-known recursive ($\ell = 0$) and semi recursive ($\ell = 1$) estimators. The recursive property of this estimator class is clearly useful in sequential investigations and for large sample sizes because the addition of a new observation means that the non-recursive estimators must be recomputed. In addition, we need to store extensive data to calculate them.

We assume that the following assumptions hold:

H1 The operators r and σ_ε^2 are continuous on a neighborhood of χ, and $F(0) = 0$. Moreover, the function $\varphi(t) := \mathrm{E}\left[\{r(\mathcal{X}) - r(\chi)\} | \|\mathcal{X} - \chi\| = t\right]$ is assumed to be derivable at $t = 0$.

H2 K is nonnegative bounded kernel with support on the compact $[0, 1]$ such that $\inf\limits_{t \in [0,1]} K(t) > 0$.

H3 For any $s \in [0, 1], \tau_h(s) := \frac{F(hs)}{F(h)} \to \tau_0(s) < \infty$ as $h \to 0$.

H4 (i) $h_n \to 0, nF(h_n) \to \infty$ and $A_{n,\ell} := \frac{1}{n} \sum\limits_{i=1}^{n} \frac{h_i}{h_n} \left[\frac{F(h_i)}{F(h_n)}\right]^{1-\ell} \to \alpha_{[\ell]} < \infty$ as $n \to \infty$.

(ii) $\forall r \leq 2, B_{n,r} := \frac{1}{n} \sum\limits_{i=1}^{n} \left[\frac{F(h_i)}{F(h_n)}\right]^{r} \to \beta_{[r]} < \infty$, as $n \to \infty$.

Assumptions **H1**, **H2** and the first part of **H4** are common in nonparametric regression. They have been used by Ferraty et al. (2007) and are the same as those classically used in finite-dimensional settings. The conditions $A_{n,\ell} \to \alpha_{[\ell]} < \infty$ and **H4(ii)** are particular to the recursive problem and the same as those used in finite-dimensional cases. Note that F plays a crucial role in our calculus: Its limit at zero, and a fixed χ is known as 'small ball probability'. Before announcing our results, let us provide typical examples of bandwidths and small ball probabilities that satisfy **H3** and **H4** (see Ferraty et al. (2007) for more details).

If $\mathcal{X}$ is a fractal (or geometric) process, then the small ball probabilities are

of the form $F(t) \sim c'_\chi t^\kappa$, where c'_χ and κ are positive constants, and $\|\cdot\|$ might be a supremum norm, an L^p norm or a Besov norm. The choice of bandwidth $h_n = An^{-\delta}$ with $A > 0$ and $0 < \delta < 1$ implies $F(h_n) = c''_\chi n^{-\delta\kappa}$, $c''_\chi > 0$. Thus, **H3** and **H4** hold when $\delta\kappa < 1$. In fact, assumption **H3** and the first part of **H4** are clearly non-restrictive because they are the same as those used in the non-recursive case. With regard to **H4(ii)**, if $\delta\kappa r < 1$, then $\sum_{i=1}^{n} i^{-\delta\kappa r} \sim \frac{n^{1-\delta\kappa r}}{1-\delta\kappa r}$. Thus, the condition is satisfied when $\beta_{[r]} = \frac{1}{1-\delta\kappa r}$. The same argument is also valid for $A_{n,\ell}$, if $\max\{\kappa r, 1 + \kappa(1-\ell)\} < 1/\delta$.

2.2 Main results

Let us introduce the following notations from Ferraty et al. (2007):

$$
\begin{aligned}
M_0 &= K(1) - \int_0^1 (sK(s))'\tau_0(s)ds, \; M_1 = K(1) - \int_0^1 K'(s)\tau_0(s)ds, \\
M_2 &= K^2(1) - \int_0^1 (K^2(s))'\tau_0(s)ds.
\end{aligned}
$$

We can establish the asymptotic mean square error of our recursive estimate.

Theorem 1 *Under the assumptions* **H1** − **H4**, *we have*

$$
\begin{aligned}
E\left[r_n^{[\ell]}(\chi)\right] - r(\chi) &= \varphi'(0)\frac{\alpha_{[\ell]}}{\beta_{[1-\ell]}}\frac{M_0}{M_1}h_n[1 + o(1)] + O\left[\frac{1}{nF(h_n)}\right], \\
Var\left[r_n^{[\ell]}(\chi)\right] &= \frac{\beta_{[1-2\ell]}}{\beta_{[1-\ell]}^2}\frac{M_2}{M_1^2}\sigma_\varepsilon^2(\chi)\frac{1}{nF(h_n)}[1 + o(1)].
\end{aligned}
$$

Theorem 1 is an extension of Ferraty et al. (2007) result to the class of recursive estimators. Using a bias-variance representation and an additional condition, the asymptotic mean square error of our estimators is established in the following result:

Corollary 1 *Assume that the assumptions of Theorem 1 hold. If there exists a constant $c > 0$ such that $nF(h_n)h_n^2 \to c$, as $n \to \infty$, then*

$$
\lim_{n\to\infty} nF(h_n)E\left[\left(r_n^{[\ell]}(\chi) - r(\chi)\right)^2\right] = \left[\frac{\beta_{[1-2\ell]}}{\beta_{[1-\ell]}^2}\frac{M_2\sigma_\varepsilon^2(\chi)}{M_1^2} + \frac{c\alpha_{[\ell]}^2}{\beta_{[1-\ell]}^2}\frac{\varphi'(0)^2M_0^2}{M_1^2}\right].
$$

In particular, if $\mathcal{X}$ is fractal (or geometric process) with $F(t) \sim c'_\chi t^\kappa$, then the choice $h_n = An^{-\frac{1}{\kappa+2}}$, $A, \kappa > 0$, implies that

$$
\lim_{n\to\infty} n^{\frac{2}{2+\kappa}}E\left[\left(r_n^{[\ell]}(\chi) - r(\chi)\right)^2\right] = \left[\frac{\beta_{[1-2\ell]}}{\beta_{[1-\ell]}^2}\frac{M_2\sigma_\varepsilon^2(\chi)}{c'_\chi A^\kappa M_1^2} + \frac{\alpha_{[\ell]}^2}{\beta_{[1-\ell]}^2}\frac{\varphi'(0)^2M_0^2A^2}{M_1^2}\right].
$$

Bosq and Cheze-Payaud (1999) established a similar result for the Nadaraya-Watson estimator using the finite-dimensional setting and continuous time processes.

To obtain the almost sure convergence rate of our estimator, we assume that the following additional assumptions hold.

H5 There exist $\lambda > 0$ and $\mu > 0$ such that $\mathrm{E}\left[\exp\left(\lambda|Y|^{\mu}\right)\right] < \infty$.

H6 $\lim\limits_{n \to +\infty} \dfrac{nF(h_n)(\ln n)^{-1-\frac{2}{\mu}}}{(\ln \ln n)^{2(\alpha+1)}} = \infty$ for some $\alpha \geq 0$ and $\lim\limits_{n \to +\infty} (\ln n)^{\frac{2}{\mu}} F(h_n) = 0$.

Assumption **H5** is clearly met if Y is bounded, which implies that

$$\mathrm{E}\left(\max_{1 \leq i \leq n} |Y_i|^p\right) = O[(\ln n)^{p/\mu}], \forall p \geq 1, n \geq 2. \tag{4}$$

In fact, if we set $M = \begin{cases} \left(\frac{p-\mu}{\lambda\mu}\right)^{1/\mu} & \text{if } p > \mu \\ 0 & \text{else} \end{cases}$, then one is able to write:

$$\mathrm{E}\left(\max_{1 \leq i \leq n} |Y_i|^p\right) \leq M^p + \mathrm{E}\left(\max_{1 \leq i \leq n} |Y_i|^p \mathbb{1}_{\{|Y_i|>M\}}\right).$$

Because the function $x \mapsto (\ln x)^{p/\mu}$ for all $p \geq 1$ is concave down on set $]\max\{1, \exp(\frac{p}{\mu}-1)\}, +\infty[$, Jensen's inequality (via assumption **H5**) implies that

$$
\begin{aligned}
\mathrm{E}\left(\max_{1 \leq i \leq n} |Y_i|^p \mathbb{1}_{\{|Y_i|>M\}}\right) &\leq \left[\ln \mathrm{E}\exp\left(\lambda \max_{1 \leq i \leq n} |Y_i|^\mu \mathbb{1}_{\{|Y_i|>M\}}\right)\right]^{p/\mu} \\
&\leq \left[\ln \sum_{i=1}^{n} \mathrm{E}\exp\left(\lambda|Y_i|^\mu\right)\right]^{p/\mu} = O[(\ln n)^{p/\mu}],
\end{aligned}
$$

and (4) follows. An example of a sequence of random variables, Y_i satisfying **H5** (and then (4)) is the standard Gaussian distribution, with $\lambda = 1$ and $\mu = 2$. Bosq and Cheze-Payaud (1999) used relation (4) in the multivariate framework to establish the optimal quadratic error of the Nadaraya-Watson estimator. Assumption **H6** is satisfied when $\mathcal{X}$ is a fractal or non-smooth, whereas condition $\lim\limits_{n \to \infty} F(h_n)(\ln n)^{\frac{2}{\mu}} = 0$ is not necessary when $\mu \geq 2$. We can write the following theorem for our estimator of the regression operator.

Theorem 2 *Assume that* **H1** $-$ **H6** *hold. If* $\lim\limits_{n \to +\infty} nh_n^2 = 0$*, then*

$$\limsup_{n \to \infty} \left[\frac{nF(h_n)}{\ln \ln\ n}\right]^{1/2} \left[r_n^{[\ell]}(\chi) - r(\chi)\right] = \frac{\left[2\beta_{[1-2\ell]}\sigma_\varepsilon^2(\chi)M_2\right]^{1/2}}{\beta_{[1-\ell]}M_1} \quad a.s.$$

6

The choices of bandwidths and small ball probabilities previously provided are typical examples that satisfy the condition $\lim_{n \to +\infty} nh_n^2 = 0$. Case $\ell = 1$ of Theorem 2 is an extension to the functional setting of Roussas (1992) result concerning the almost sure convergence of Devroye-Wagner's estimator. Note, that in the non-recursive framework, the rate of convergence obtained is of the form $\left[\frac{nF(h_n)}{\ln n}\right]^{1/2}$ (see Lemma 6.3 in Ferraty and Vieu (2006)). Unlike the non-recursive case, the rate of convergence of the recursive estimators are obtained with exact upper bounds.

To obtain asymptotic normality, we make the following additional assumption, which was clearly verified by the choices of bandwidths and the small ball probabilities above.

H7 For any $\delta > 0$, $\lim_{n \to \infty} \frac{(\ln n)^\delta}{\sqrt{nF(h_n)}} = 0$.

Theorem 3 *Assume that* **H1** $-$ **H5** *and* **H7** *hold. If there exists $c \geq 0$ such that $\lim_{n \to \infty} h_n \sqrt{nF(h_n)} = c$, then*

$$\sqrt{nF(h_n)} \left(r_n^{[\ell]}(\chi) - r(\chi)\right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(c\frac{\alpha_{[\ell]}}{\beta_{[1-\ell]}}\frac{M_0}{M_1}\varphi'(0), \ \frac{\beta_{[1-2\ell]}}{\beta_{[1-\ell]}^2}\frac{M_2}{M_1^2}\sigma_\varepsilon^2(\chi)\right).$$

Note that the choices of bandwidths and small ball probabilities above imply that $\beta_{[1-2\ell]}/\beta_{[1-\ell]}^2 < 1$. Thus, the recursive estimators are more efficient than classical estimators in this case in the sense that their asymptotic variance is small for a given M_1, M_2 and $\sigma_\epsilon(\chi)$.

Because the result of Theorem 3 depends on unknown quantities, we derived a usable asymptotic distribution for the following corollary of case $\ell = 0$.

Corollary 2 *Assume that* **H1** $-$ **H5** *and* **H7** *hold. If $\lim_{n \to \infty} h_n\sqrt{nF(h_n)} = 0$ and $\lim_{n \to \infty} \frac{\ln n}{nF^2(h_n)} = 0$, then for any consistent estimators $\widehat{M_i}, i = 1, 2$ and $\widehat{\sigma}_\varepsilon(\chi)$ of $M_i, i = 1, 2$ and $\sigma_\varepsilon(\chi)$ respectively, we have*

$$\sqrt{n\widehat{F}(h_n)}\sqrt{\frac{\widehat{\beta}_{[1]}\widehat{M_1}^2}{\widehat{M_2}\widehat{\sigma}_\varepsilon^2(\chi)}} \left(r_n^{[0]}(\chi) - r(\chi)\right) \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1),$$

where $\widehat{F}$ and $\widehat{\beta}_{[1]}$ are the empirical counterparts of F and $\beta_{[1]}$ defined by

$$\widehat{F}(t) = \frac{1}{n}\sum_{i=1}^n \mathbb{1}_{\{\|\mathcal{X}_i - \chi\| \leq t\}} \ \text{and} \ \widehat{\beta}_{[1]} = \frac{1}{n}\sum_{i=1}^n \frac{\widehat{F}(h_i)}{\widehat{F}(h_n)}.$$

Corollary 2 is similar to the result obtained by Ferraty et al. (2007) in the non-recursive case. The assumptions used to establish this result are fulfilled

by the choices of bandwidths and small ball probabilities above. If we consider the uniform kernel $K(\cdot) = \mathbb{1}_{[0,1]}(\cdot)$, then the asymptotic distribution in Corollary 2 can be rewritten as

$$\sqrt{\frac{n\widehat{F}(h_n)\widehat{\beta}_{[1]}}{\widehat{\sigma}_\varepsilon^2}} \left(r_n^{[0]}(\chi) - r(\chi) \right) \xrightarrow{\mathcal{D}} \mathcal{N}(0,1).$$

According to Corollary 2, the asymptotic confidence band of $r(\chi)$ with level $1 - \alpha$ is given by

$$\left[r_n^{[0]}(\chi) - z_{1-\alpha/2}\sqrt{n\widehat{F}(h_n)}\sqrt{\frac{\widehat{\beta}_{[1]}\widehat{M_1}^2}{\widehat{M_2}\widehat{\sigma}_\varepsilon^2(\chi)}} ; r_n^{[0]}(\chi) + z_{1-\alpha/2}\sqrt{n\widehat{F}(h_n)}\sqrt{\frac{\widehat{\beta}_{[1]}\widehat{M_1}^2}{\widehat{M_2}\widehat{\sigma}_\varepsilon^2(\chi)}} \right],$$

where $z_{1-\alpha/2}$ is the quantile of order $1 - \alpha/2$ of the standard normal distribution.

3 Simulation study and real dataset example

To observe the behavior of our recursive estimator in practice, this section considers a simulation study. We simulated our data in the following way: The curves $\mathcal{X}_1, \ldots$, and $\mathcal{X}_n$ are standard Brownian motions on $[0,1]$, with $n = 100$. Each curve is discretized into $p = 100$ equidistant points on $[0,1]$. The operator r is defined by $r(\chi) = \int_0^1 \chi(s)^2 \, ds$. The error ε is simulated as a Gaussian random variable with mean 0 and standard deviation 0.1. The simulations were repeated 500 times to compute the prediction errors for a new curve χ, which was also simulated as a standard Brownian motion on $[0,1]$.

In this functional context, the estimator depends on the choice of many parameters: the semi-norm $\|\cdot\|$ of the functional space $\mathcal{E}$, the sequence of bandwidths (h_n), the kernel K, the parameter ℓ and the distribution function F in case $\ell \neq 0$. Because the choice of kernel K is not crucial, we used the quadratic kernel defined by $K(u) = \left(1 - u^2\right)\mathbb{1}_{[0,1]}(u)$ for all $u \in \mathbb{R}$, which behaves correctly in practice and is easy to implement. We estimated the distribution function F, using the empirical distribution function, which is uniformly convergent.

3.1 Choice of the bandwidth

In this simulation, the semi-norm was based on the principal components analysis of the curves that retained 3 principal components (see Besse et al. (1997) for a description of this semi-norm), whereas ℓ is fixed and equal to 0. We show below that this parameter ℓ has a negligible influence on the

behavior of the estimator.

We chose to take a sequence of bandwidths $h_i = C \max\limits_{i=1,\ldots,n} \|\mathcal{X}_i - \chi\| \, i^{-\nu}$, for $1 \leq i \leq n$, with $C \in \{0.5, 1, 2, 10\}$ and $\nu \in \left\{\frac{1}{10}, \frac{1}{8}, \frac{1}{6}, \frac{1}{5}, \frac{1}{4}, \frac{1}{3}, \frac{1}{2}, 1\right\}$.

At the same time, we also computed the estimator (2) introduced by Ferraty and Vieu (2006). Following Rachdi and Vieu (2007), we introduced an automatic selection of the bandwidth with a cross validation procedure. We used this procedure for the estimator of Ferraty and Vieu (2006). For the recursive estimator, we denoted $h_i = h_i(C, \nu)$ with $C \in \{0.5, 1, 2, 10\}$ and $\nu \in \left\{\frac{1}{10}, \frac{1}{8}, \frac{1}{6}, \frac{1}{5}, \frac{1}{4}, \frac{1}{3}, \frac{1}{2}, 1\right\}$, and we considered the cross validation criterion

$$CV(C, \nu) = \frac{1}{n} \sum_{i=1}^{n} \left(Y_i - r_n^{[\ell],[-i]}(\mathcal{X}_i)\right)^2,$$

where $r_n^{[\ell],[-i]}$ represents the recursive estimator of r using the $(n-1)$-sample that corresponds to the initial sample without the ith observation $(\mathcal{X}_i, Y_i)$ for $1 \leq i \leq n$. Then, we selected the values C_{CV}, ν_{CV} of C and ν that minimized $CV(C, \nu)$. Our estimator was $r_n^{[\ell]}$ using these selected values of C and ν.

Table 1 presents the mean and standard deviations of the prediction error over 500 simulations for the optimal values of C and ν with respect to the CV criterion (these optimal values were $C_{CV} = 1$ and $\nu_{CV} = 1/10$ for our estimator). Specifically, denoting $\widehat{Y}^{[j]} = r_n^{[\ell],[j]}(\chi^{[j]})$ the predicted value at the jth iteration of the simulation ($j = 1, \ldots, 500$) for a new curve $\chi^{[j]}$, we provided the mean ($MSPE$) and the standard deviations of the quantities $\left(\widehat{Y}^{[j]} - Y^{[j]}\right)^2$. The errors were computed for our estimator (label (1) in the table) and the estimator from Ferraty and Vieu (2006) (label (2) in the table), both of which were adapted using the Rachdi and Vieu (2007) procedure. From these results, we observed that the estimator from Ferraty and Vieu (2006) is slightly better than our estimator for the $MSPE$ criterion. As we will observe in (see Subsection 3.4), the advantage of our estimator is its computational time. This section also examines the behavior of the prediction errors when the sample size increases. We used $n = 100$, $n = 200$ and $n = 500$; as expected, the errors decreased when the sample size increased.

3.2 Choice of the semi-norm

In this simulation, the parameter ℓ was fixed equal to 0 and we chose bandwidth $h_i = \max\limits_{i=1,\ldots,n} \|\mathcal{X}_i - \chi\| \, i^{-1/10}$. Next, we aimed to compare the influence of the choice of the semi-norm by considering the following:

- The semi-norm $[PCA]$ based on the principal components analysis of

	$n = 100$	$n = 200$	$n = 500$
(1)	0.3022	0.2596	0.1993
	(0.6887)	(0.6275)	(0.5430)
(2)	0.2794	0.2143	0.1368
	(0.5512)	(0.5055)	(0.4208)

Table 1: Mean and standard deviation of the square prediction error computed via 500 simulations for different values of n with the optimal values of bandwidth provided by C_{CV} and ν_{CV}

the curves retained $q = 3$ principal components; specifically,

$$\|\mathcal{X}_i - \chi\|_{PCA} = \sqrt{\sum_{j=1}^{q} \langle \mathcal{X}_i - \chi, \nu_j \rangle^2},$$

where $\langle ., . \rangle$ is the usual inner product of the space of square-integrable functions and (ν_j) is the sequence of the eigenfunctions of the empirical covariance operator Γ_n defined by $\Gamma_n u := \frac{1}{n} \sum_{i=1}^{n} \langle \mathcal{X}_i, u \rangle u$.

• The semi-norm $[FOU]$ was based on a decomposition of the curves in a Fourier basis, with $b = 8$ basis functions; specifically,

$$\|\mathcal{X}_i - \chi\|_{FOU} = \sqrt{\sum_{j=1}^{b} (a_{\mathcal{X}_i,j} - a_{\chi,j})^2},$$

where $(a_{\mathcal{X}_i,j})$ and $(a_{\chi,j})$ are the coefficient sequences of the respective Fourier approximations of curves $\mathcal{X}_i$ and χ.

• The semi-norm $[DERIV]$ was based on a cubic spline approximation comparison of the second derivatives of the curves (with numerous interior knots $k = 8$ for the cubic splines); specifically,

$$\|\mathcal{X}_i - \chi\|_{DERIV} = \sqrt{\langle \widetilde{\mathcal{X}}_i - \widetilde{\chi}, \widetilde{\mathcal{X}}_i - \widetilde{\chi} \rangle},$$

where $\widetilde{\mathcal{X}}_i$ and $\widetilde{\chi}$ are the spline approximations of the second derivatives of curves $\mathcal{X}_i$ and χ.

• The semi-norm $[PLS]$ in which the data are projected on the space determined by a PLS regression on the curves applied $K = 5$ PLS basis functions; specifically,

$$\|\mathcal{X}_i - \chi\|_{PLS} = \sqrt{\sum_{j=1}^{K} \langle \mathcal{X}_i - \chi, p_j \rangle^2},$$

where (p_j) is the sequence of PLS basis functions.

The results are provided in Table 2. For these simulated data, the semi-norms $[PCA]$ and $[PLS]$ show better results. As Ferraty and Vieu (2006) indicated, however, a universal norm does not exist that would overcome the others. The choice of the semi-norm depends on the treated data.

norm	$[PCA]$	$[FOU]$	$[DERIV]$	$[PLS]$
MSPE	0.3936	0.4506	0.4527	0.3887
	(1.5190)	(1.5624)	(1.5616)	(1.5098)

Table 2: Mean and standard deviation of the square prediction error computed on 500 simulations for different norms.

3.3 Choice of the parameter ℓ

In this simulation, we used $h_i = \max_{i=1,\dots,n} \|\mathcal{X}_i - \chi\| \, i^{-1/10}$ and the semi-norm based on the principal components analysis of the curves that retained 3 principal components. The parameter ℓ varied into $\{0, \frac{1}{4}, \frac{1}{2}, \frac{3}{4}, \text{ and } 1\}$. The results are provided in Table 3. We showed that the values of the $MSPE$ criterion are close; thus, this parameter does not appear to significantly affect the quality of the prediction, even if the mean square error decreases with respect to ℓ as in the multivariate setting.

ℓ	0	0.25	0.5	0.75	1
MSPE	0.4054848	0.4054814	0.4054786	0.4054764	0.4054746
	(1.372965)	(1.372930)	(1.372896)	(1.372863)	(1.372831)

Table 3: Mean and standard deviation of the square prediction error computed on 500 simulations for different values of ℓ.

3.4 Computational time

This subsection highlights an important advantage of the recursive estimator compared with the initial one from Ferraty and Vieu (2006)), regarding the increase in computational time needed to predict a response when new values of the explanatory variable are sequentially added to the database. In fact, when a new observation $(\mathcal{X}_{n+1}, Y_{n+1})$ appears, the computation of the recursive estimator $r_{n+1}^{[\ell]}$ requests another iteration of the algorithm via its value computed with the sequence $(\mathcal{X}_i, Y_i)_{i=1,\dots,n}$, whereas the initial estimator must be recalculated using the whole sample $(\mathcal{X}_i, Y_i)_{i=1,\dots,n+1}$. The following illustrates the computation time difference between both estimators in these types of situations. From an initial sample $(\mathcal{X}_i, Y_i)_{i=1,\dots,n}$ with size $n = 100$, we considered N additional observations for different values of

N. We compared the cumulative computational times to obtain the recursive and non-recursive estimators when adding these new observations. The characteristics of the computer we used to perform these computations were CPU: Duo $E4700$ 2.60 GHz, HD: 149 Go, Memory: 3.23 Go. The simulation was performed under the following conditions: the curves $\mathcal{X}_1, \dots,$ and $\mathcal{X}_n$ as well as the new observations $\mathcal{X}_{n+1}, \dots,$ and $\mathcal{X}_{n+N}$, which are standard Brownian motions on $[0, 1]$ with $n = 100$ and $N \in \{1, 50, 100, 200, 500\}$. The semi-norm, bandwidth sequence and the parameter ℓ were chosen as in the previous cases.

The computational times are displayed in Table 4. Here, our estimator shows a clear advantage with regard to computational time compared with the estimator from Ferraty and Vieu (2006).

N	1	50	100	200	500
comp. time for $r_{n+1}^{[\ell]}, \dots, r_{n+N}^{[\ell]}$	0.125	0.484	0.859	1.563	3.656
comp. time for $r_{n+1}, \dots, r_{n+N}$	0.047	1.922	5.594	21.938	152.719

Table 4: Cumulative computational times in seconds for the recursive and Ferraty and Vieu (2006) estimators when adding N new observations for different values of N.

3.5 First real dataset example

This subsection applies our estimator to a real dataset. Functional data are particularly adapted to studying a time series. We illustrate this fact using the El Niño time series[1] that provides the monthly sea surface temperature from January 1982 to December 2011 (360 months). These data are plotted in Figure 1. From this time series, we extracted the 30 annual curves $\mathcal{X}_1, \dots,$ and $\mathcal{X}_{30}$ from 1982 to 2011, discretized into $p = 12$ points. These curves are plotted in Figure 2. The variable of interest at month j of year i is the sea temperature $\mathcal{X}_{i+1}$ for month j; in other words, for $j = 1, \dots, 12$ and for $i = 1, \dots, 29$, $Y_i^{[j]} = \mathcal{X}_{i+1}(j)$.

We predicted the values of $Y_{29}^{[1]} \dots,$ and $Y_{29}^{[12]}$ (in other words, the values of curve $\mathcal{X}_{30}$). The recursive estimator and the estimator from Ferraty and Vieu (2006) were computed by choosing the semi-norm, the bandwidth sequence and the parameter ℓ as in previous cases.

We analyzed the results by computing the mean square prediction error through 2011, which was provided by

$$MSPE = \frac{1}{12} \sum_{j=1}^{12} \left(\widehat{Y}_{29}^{[j]} - Y_{29}^{[j]} \right)^2,$$

---

[1]available online at http://www.math.univ-toulouse.fr/staph/npfda/
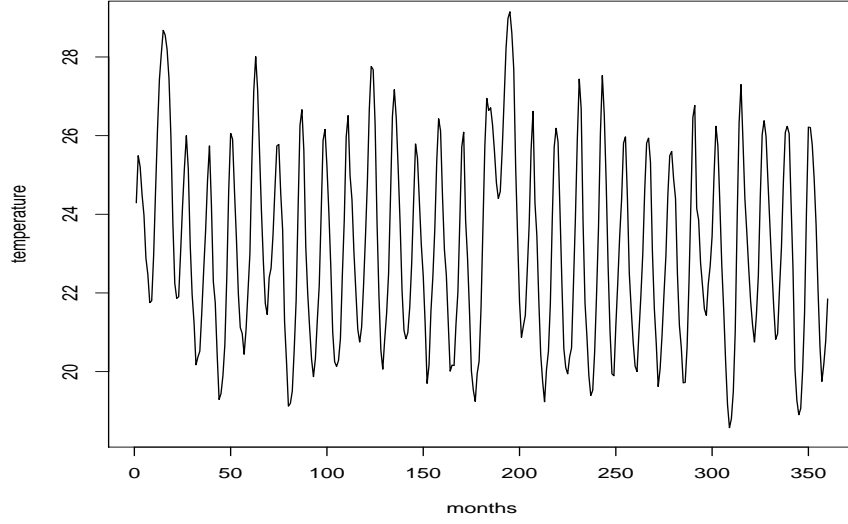
Figure 1: El Niño monthly temperature time series from January 1982 to December 2011
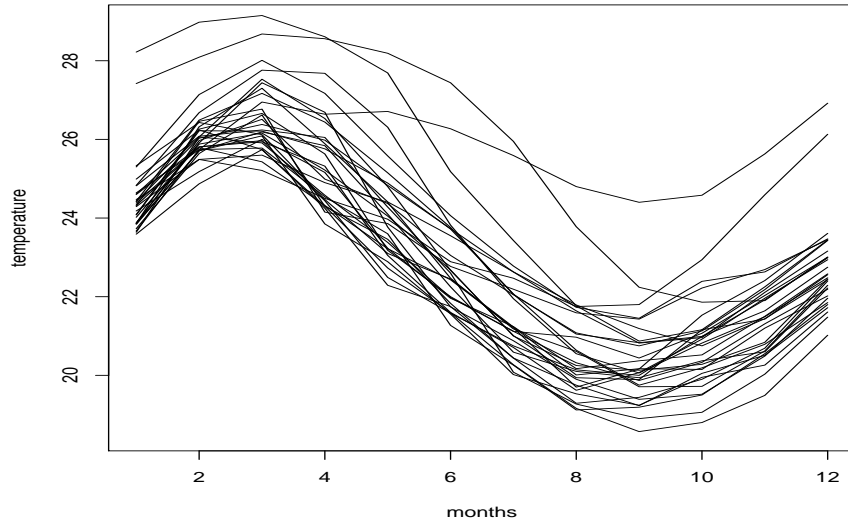

Figure 2: El Niño annual temperature curves from 1982 to 2011

where $\widehat{Y}_{29}^{[j]}$ was computed with either the recursive estimator (result = 0.5719) or the estimator from Ferraty and Vieu (2006) (result: 0.2823). The estimator from Ferraty and Vieu (2006) again revealed its advantage with regard to

13

prediction, whereas our estimator behaved well and had an advantage with regard to computational time (as highlighted in the previous subsection). The computational time (in seconds) needed for our estimator to predict twelve values (the final year) was 0.128s, whereas the computational time for the Ferraty and Vieu (2006) estimator was 0.487s.

To construct the confidence intervals described at the end of Section 2, we used the example from 2011. The constants involved in Corollary 2 were estimated by

$$\widehat{M_1} = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{\widehat{F}(h_i)} K\left(\frac{\|\chi - \mathcal{X}_i\|}{h_i}\right) \text{ and } \widehat{M_2} = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{\widehat{F}(h_i)} K^2\left(\frac{\|\chi - \mathcal{X}_i\|}{h_i}\right).$$

The consistency of $\widehat{M_i}, i = 1, 2$ is the same as the lines of the Lemmas 3, and 5 proofs with the help of the Glivenko-Cantelli theorem. In addition, the estimation of the conditional variance σ_ε^2 was performed using a nonparametric kernel regression procedure:

$$\widehat{\sigma}_\varepsilon^2(\chi) = \frac{\sum_{i=1}^{n} Y_i^2 K\left(\frac{\|\chi - \mathcal{X}_i\|}{h_i}\right)}{\sum_{i=1}^{n} K\left(\frac{\|\chi - \mathcal{X}_i\|}{h_i}\right)} - \left(r_n^{[0]}(\chi)\right)^2.$$

The different values of $\widehat{\sigma}_\varepsilon^2$ for each month of last year are shown in Table 5.

month	1	2	3	4	5	6
$\widehat{\sigma}_\varepsilon^2$	1.00551	0.60462	0.81971	1.60694	2.07112	2.02370
month	7	8	9	10	11	12
$\widehat{\sigma}_\varepsilon^2$	2.01471	1.65082	1.25991	1.24542	1.27177	1.08482

Table 5: Recursive estimations of σ_ε^2 for each month on the last year.

The estimations of the other quantities in the confidence intervals were $\widehat{\beta}_{[1]} = 1.04966$, $\widehat{M_1} = 0.73367$ and $\widehat{M_2} = 0.62716$.
We compared the confidence intervals for 2011 in Table 6 with the recursive estimator (label (1) in the table) and the estimator from Ferraty and Vieu (2006) (label (2) in the table). We noticed that the confidence interval with the estimator from Ferraty and Vieu (2006) was still better, mainly due to the estimation of σ_ε and the terms M_1 and M_2 (which were estimated by the respective non-recursive counterparts of $\widehat{\sigma}_\varepsilon^2$, and $\widehat{M_1}$, and $\widehat{M_2}$ was obtained by replacing h_i with h_n in the non-recursive case), whereas the term $\widehat{\beta}_{[1]}$ decreased the width of the confidence interval in the recursive case.
The corresponding true curve, predicted curve and 95% confidence intervals over the year 2011 are plotted in Figure 3.

month	1	2	3	4
(1)	[23.24; 27.38]	[24.68; 27.89]	[25.10; 28.83]	[23.62; 28.86]
(2)	[22.80; 26.24]	[24.83; 27.49]	[25.07; 28.17]	[23.42; 27.76]
month	5	6	7	8
(1)	[21.96; 27.90]	[21.04; 26.91]	[19.91; 25.77]	[18.76; 24.06]
(2)	[21.80; 26.74]	[20.44; 25.32]	[19.18; 24.05]	[18.42; 22.83]
month	9	10	11	12
(1)	[18.59; 23.22]	[18.65; 23.26]	[19.22; 23.87]	[20.59; 24.89]
(2)	[18.37; 22.22]	[18.81; 22.63]	[19.54; 23.40]	[20.88; 24.45]

Table 6: Confidence intervals for 2011 with the recursive estimator (label (1)) and the estimator from Ferraty and Vieu (2006) (label (2)).
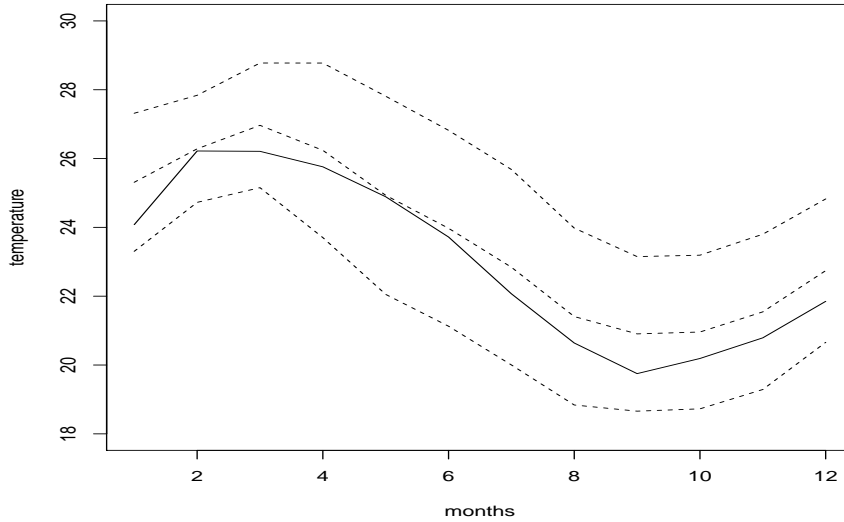

Figure 3: El Niño true and predicted temperature curves for 2011. The solid line denotes the true curve. The dashed lines represent the predicted curve and the 95% confidence intervals with the recursive estimator.

3.6   Second real dataset example

This subsection highlights the performances of our estimator with regard to computational time gain. The dataset we used consisted of information collected by ORAMIP[2], a French organization that studies air quality. We disposed of a sample of 474 daily ozone pollution measurements curves (in μg/m^3). The variable Y_i of interest was the daily maximum of ozone. The ozone curve the day before (from 6 : 00 pm to 5 : 00 pm the day after)

---

[2]Observatoire Régional de l'Air en Midi-Pyrénées

was used as a functional explicative variable $\mathcal{X}_i$. Specifically, each $\mathcal{X}_i$ was observed at p = 24 equidistant points that corresponded to hourly measurements. The sample was divided into a learning sample of 332 and a test sample of 142. The ozone curve from the learning sample is plotted in Figure 4. Every value of the test sample was predicted using our method and the
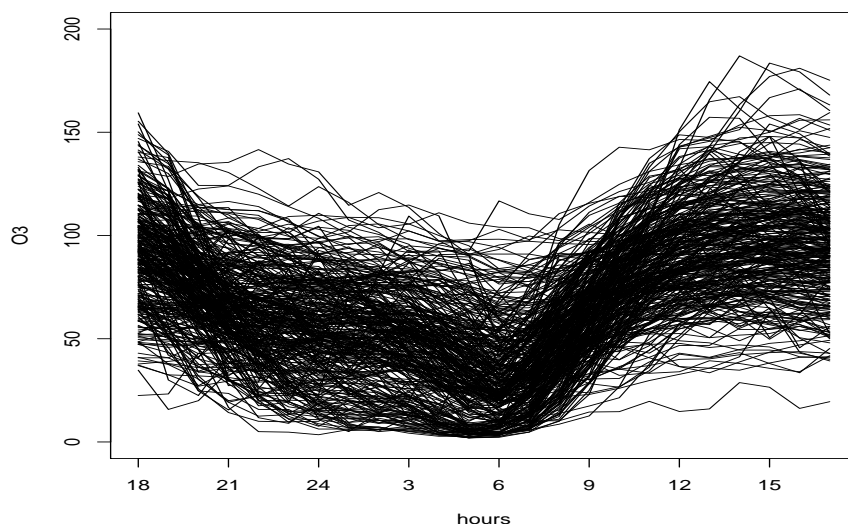

Figure 4: Daily ozone curves for the learning sample

method from Ferraty and Vieu (2006). The learning sample and the curves of the test sample were predicted as if they had arrived in real time. Again, the estimator from Ferraty and Vieu (2006) showed its advantage over our estimator with regard to mean prediction error (25.797 vs. 26.171). However, our estimator performed best when predicting several values and using recursivity, especially when many values need to be predicted (which was the case here, given that we predicted 142 values). The computational time for our estimator (in seconds) was 2.025, whereas it was 25.478 for the Ferraty and Vieu method; however, we are conscious that this advantage was relative in this case because we dealt with daily measurements. In any case, the choice is up to the user: a better prediction error or a faster computational time.

16

4 Proofs

Throughout the proofs, we denote γ_i a sequence of real numbers going to zero as i goes to ∞. The kernel estimate $r_n^{[\ell]}$ writes

$$r_n^{[\ell]}(\chi) \;=\; \frac{\varphi_n^{[\ell]}(\chi)}{f_n^{[\ell]}(\chi)},$$

where $\varphi_n^{[\ell]}$ and $f_n^{[\ell]}$ are defined in (3).

4.1 Proof of Theorem 1

To prove the first assertion of Theorem 1, let us use the following decomposition

$$\mathrm{E}\left[r_n^{[\ell]}(\chi)\right] \;=\; \frac{\mathrm{E}\left[\varphi_n^{[\ell]}(\chi)\right]}{\mathrm{E}\left[f_n^{[\ell]}(\chi)\right]} - \frac{\mathrm{E}\left\{\varphi_n^{[\ell]}(\chi)\left[f_n^{[\ell]}(\chi) - \mathrm{E}f_n^{[\ell]}(\chi)\right]\right\}}{\left\{\mathrm{E}\left[f_n^{[\ell]}(\chi)\right]\right\}^2}$$
$$+ \frac{\mathrm{E}\left\{r_n^{[\ell]}(\chi)\left[f_n^{[\ell]}(\chi) - \mathrm{E}f_n^{[\ell]}(\chi)\right]^2\right\}}{\left\{\mathrm{E}\left[f_n^{[\ell]}(\chi)\right]\right\}^2}.$$

The first part of Theorem 1 is then a direct consequence of Lemmas 1-4 below.

Lemma 1 *Under assumptions **H1-H4**, we have*

$$\frac{E\left[\varphi_n^{[\ell]}(\chi)\right]}{E\left[f_n^{[\ell]}(\chi)\right]} - r(\chi) = h_n \varphi'(0) \frac{\alpha_{[\ell]}}{\beta_{[1-\ell]}} \frac{M_0}{M_1}\left[1 + o(1)\right].$$

Lemma 2 *Under assumptions **H1-H4**, we have*

$$E\left\{\varphi_n^{[\ell]}(\chi)\left[f_n^{[\ell]}(\chi) - Ef_n^{[\ell]}(\chi)\right]\right\} \;=\; O\left[\frac{1}{nF(h_n)}\right],$$
$$E\left\{r_n^{[\ell]}(\chi)\left[f_n^{[\ell]}(\chi) - Ef_n^{[\ell]}(\chi)\right]^2\right\} \;=\; O\left[\frac{1}{nF(h_n)}\right].$$

Lemma 3 *Under assumptions **H1-H4**, we have*

$$E\left(f_n^{[\ell]}(\chi)\right) = M_1\left[1 + o(1)\right] \quad and \quad E\left(\varphi_n^{[\ell]}(\chi)\right) = r(\chi)M_1\left[1 + o(1)\right].$$

17

The study of the variance term in Theorem 1 is based on the following decomposition which can be found in Collomb (1976).

$$
\begin{aligned}
\mathrm{Var}\left[r_n^{[\ell]}(\chi)\right] \;=\;& \frac{\mathrm{Var}\left[\varphi_n^{[\ell]}(\chi)\right]}{\left\{\mathrm{E}\left[f_n^{[\ell]}(\chi)\right]\right\}^2} - 4\frac{\mathrm{E}\left[\varphi_n^{[\ell]}(\chi)\right]\mathrm{Cov}\left[f_n^{[\ell]}(\chi),\varphi_n^{[\ell]}(\chi)\right]}{\left\{\mathrm{E}\left[f_n^{[\ell]}(\chi)\right]\right\}^3} \\
&+3\mathrm{Var}\left[f_n^{[\ell]}(\chi)\right]\frac{\left\{\mathrm{E}\left[\varphi_n^{[\ell]}(\chi)\right]\right\}^2}{\left\{\mathrm{E}\left[f_n^{[\ell]}(\chi)\right]\right\}^4} + o\left[\frac{1}{nF(h_n)}\right]. \qquad (5)
\end{aligned}
$$

The second assertion of Theorem 1 follows from (5) and Lemma 4 below. $\square$

Lemma 4 *Under assumptions **H1-H4**, we have*

$$
\begin{aligned}
Var\left[f_n^{[\ell]}(\chi)\right] \;&=\; \frac{\beta_{[1-2\ell]}}{\beta_{[1-\ell]}^2}M_2\frac{1}{nF(h_n)}\left[1+o(1)\right]. \\
Var\left[\varphi_n^{[\ell]}(\chi)\right] \;&=\; \frac{\beta_{[1-2\ell]}}{\beta_{[1-\ell]}^2}\left[r^2(\chi)+\sigma_\epsilon^2(\chi)\right]M_2\frac{1}{nF(h_n)}\left[1+o(1)\right]. \\
Cov\left[f_n^{[\ell]}(\chi),\varphi_n^{[\ell]}(\chi)\right] \;&=\; \frac{\beta_{[1-2\ell]}}{\beta_{[1-\ell]}^2}r(\chi)M_2\frac{1}{nF(h_n)}\left[1+o(1)\right].
\end{aligned}
$$

Now let us prove Lemmas 1-4.

4.1.1 Proof of Lemma 1

Observe that

$$
\frac{\mathrm{E}\left[\varphi_n^{[\ell]}(\chi)\right]}{\mathrm{E}\left[f_n^{[\ell]}(\chi)\right]} - r(\chi) \;=\; \frac{\sum\limits_{i=1}^{n}\frac{1}{F(h_i)^\ell}\mathrm{E}\left[(Y_i-r(\chi))K\left(\frac{\|\chi-\mathcal{X}_i\|}{h_i}\right)\right]}{\sum\limits_{i=1}^{n}\frac{1}{F(h_i)^\ell}\mathrm{E}\left[K\left(\frac{\|\chi-\mathcal{X}_i\|}{h_i}\right)\right]}.
$$

Obviously

$$
\begin{aligned}
\mathrm{E}\left[(Y_i-r(\chi))K\left(\frac{\|\chi-\mathcal{X}_i\|}{h_i}\right)\right] \;&=\; \mathrm{E}\left[(r(\mathcal{X})-r(\chi))K\left(\frac{\|\mathcal{X}-\chi\|}{h_i}\right)\right] \\
&=\; \mathrm{E}\left[\varphi(\|\mathcal{X}-\chi\|)K\left(\frac{\|\mathcal{X}-\chi\|}{h_i}\right)\right] \\
&=\; \int_0^1\varphi(h_it)K(t)d\mathrm{P}^{\|\mathcal{X}-\chi\|/h_i}(t).
\end{aligned}
$$

Therefore, a Taylor's expansion of φ around 0 ensures that

$$
\mathrm{E}\left[(Y_i-r(\chi))K\left(\frac{\|\chi-\mathcal{X}_i\|}{h_i}\right)\right] \;=\; h_i\varphi'(0)\int_0^1 tK(t)d\mathrm{P}^{\|\mathcal{X}-\chi\|/h_i}(t)+o(h_i).
$$

18

From the proof of Lemma 2 in Ferraty et al. (2007), **H2** and Fubini's Theorem imply that

$$\int_0^1 tK(t)d\mathrm{P}^{\|\mathcal{X}-\chi\|/h_i}(t) \;=\; F(h_i)\left[K(1) - \int_0^1 (sK(s))'\tau_{h_i}(s)ds\right], \quad (6)$$

and

$$\mathrm{E}K\left(\frac{\|\mathcal{X}-\chi\|}{h_i}\right) = \int_0^{h_i} K\left(\frac{t}{h_i}\right) d\mathrm{P}^{\|\mathcal{X}-\chi\|}(t) = F(h_i)\left[K(1) - \int_0^1 K'(s)\tau_{h_i}(s)ds\right]. \quad (7)$$

Also, combining (6) and (7), we have

$$\frac{\mathrm{E}\left[\varphi_n^{[\ell]}(\chi)\right]}{\mathrm{E}\left[f_n^{[\ell]}(\chi)\right]} - r(\chi) \;=\; \frac{\sum\limits_{i=1}^n h_i F(h_i)^{1-\ell}\left\{\varphi'(0)\left[K(1) - \int_0^1 (sK(s))'\tau_{h_i}(s)ds\right] + \gamma_i\right\}}{\sum\limits_{i=1}^n F(h_i)^{1-\ell}\left[K(1) - \int_0^1 K'(s)\tau_{h_i}(s)ds\right]}$$

$$\;:=\; \frac{D_1}{D_2}.$$

Finally, by virtue of **H3** we get from Toeplitz's Lemma (see Masry (1986)) that

$$\frac{D_1}{nh_n F(h_n)^{1-\ell}} = \alpha_{[\ell]}\varphi'(0)M_0[1+o(1)], \quad \frac{D_2}{nF(h_n)^{1-\ell}} = \beta_{[1-\ell]}M_1[1+o(1)],$$

and Lemma 1 follows. □

4.1.2 Proof of Lemma 3

Equality (7) allows to write

$$\mathrm{E}\left[f_n^{[\ell]}(\chi)\right] \;=\; \frac{1}{\sum\limits_{i=1}^n F(h_i)^{1-\ell}}\sum_{i=1}^n \frac{1}{F(h_i)^\ell}\mathrm{E}\left[K\left(\frac{\|\chi-\mathcal{X}_i\|}{h_i}\right)\right]$$

$$\;=\; \frac{\sum\limits_{i=1}^n \frac{F(h_i)^{1-\ell}}{nF(h_n)^{1-\ell}}\left[K(1) - \int_0^1 K'(s)\tau_{h_i}(s)ds\right]}{B_{n,1-\ell}} = M_1[1+o(1)],$$

where the last equality follows from assumptions **H3**, **H4** and Toeplitz's Lemma. Now, conditioning on $\mathcal{X}$, we have

$$\mathrm{E}\left[Y_i K\left(\frac{\|\chi-\mathcal{X}_i\|}{h_i}\right)\right] \;=\; \mathrm{E}\left\{[r(\mathcal{X}) - r(\chi) + r(\chi)]K\left(\frac{\|\chi-\mathcal{X}_i\|}{h_i}\right)\right\} =: A_i + B_i,$$

where

$$A_i := \mathrm{E}\left\{[r(\mathcal{X}) - r(\chi)]K\left(\frac{\|\chi-\mathcal{X}_i\|}{h_i}\right)\right\} \leq \sup_{\chi'\in\mathcal{B}(\chi,h_i)}\left|r(\chi') - r(\chi)\right|\mathrm{E}K\left(\frac{\|\chi-\mathcal{X}_i\|}{h_i}\right),$$

19

and $B_i := r(\chi)\mathrm{E}K\left(\frac{\|\chi-\mathcal{X}_i\|}{h_i}\right)$. Since r is continuous (**H1**), then

$$\mathrm{E}\left[Y_iK\left(\frac{\|\chi-\mathcal{X}_i\|}{h_i}\right)\right] = [r(\chi)+\gamma_i]\,\mathrm{E}K\left(\frac{\|\chi-\mathcal{X}_i\|}{h_i}\right) = F(h_i)M_1\left[r(\chi)+\gamma_i\right]. \quad (8)$$

We deduce from (8), with the help of assumptions **H3** and **H4**, and applying again Toeplitz's Lemma, that

$$\mathrm{E}\left[\varphi_n^{[\ell]}(\chi)\right] \;=\; \frac{1}{\sum\limits_{i=1}^{n}F(h_i)^{1-\ell}}\sum_{i=1}^{n}\frac{1}{F(h_i)^{\ell}}\mathrm{E}\left[Y_iK\left(\frac{\|\chi-\mathcal{X}_i\|}{h_i}\right)\right] = r(\chi)M_1\left[1+o(1)\right],$$

that proves Lemma 3. $\qquad\square$

4.1.3 Proof of Lemma 4

First, notice that as in (7), we have

$$\mathrm{E}\left[K^2\left(\frac{\|\chi-\mathcal{X}\|}{h_i}\right)\right] \;=\; F(h_i)\left[K^2(1)-\int_0^1(K^2)'(s)\tau_{h_i}(s)ds\right]. \quad (9)$$

Relation (7) and assumption **H3** ensure that

$$\mathrm{E}^2\left[K\left(\frac{\|\chi-\mathcal{X}\|}{h_i}\right)\right] \;=\; O\left[F(h_i)^2\right],$$

thus

$$\mathrm{Var}\left[K\left(\frac{\|\chi-\mathcal{X}\|}{h_i}\right)\right] = M_2F(h_i)\left[1+\gamma_i\right].$$

It follows that

$$\begin{aligned}\mathrm{Var}\left[f_n^{[\ell]}(\chi)\right] &\;=\; \frac{1}{\left(\sum\limits_{i=1}^{n}F(h_i)^{1-\ell}\right)^2}\sum_{i=1}^{n}F(h_i)^{1-2\ell}M_2\left[1+\gamma_i\right]\\ &\;=\; \frac{\beta_{[1-2\ell]}}{\beta_{[1-\ell]}^2}\frac{1}{nF(h_n)}M_2\left[1+o(1)\right],\end{aligned}$$

and the first step of Lemma 4 follows. In a similar manner, for the second step, let us write

$$\mathrm{Var}\left[\varphi_n^{[\ell]}(\chi)\right] \;=\; \frac{1}{\left(\sum\limits_{i=1}^{n}F(h_i)^{1-\ell}\right)^2}\sum_{i=1}^{n}F(h_i)^{-2\ell}\mathrm{Var}\left[Y_iK\left(\frac{\|\chi-\mathcal{X}_i\|}{h_i}\right)\right].$$

Next, one obtains by conditioning on $\mathcal{X}$,

$$\mathrm{E}\left[Y_i^2 K^2\left(\frac{\|\chi - \mathcal{X}_i\|}{h_i}\right)\right] = \mathrm{E}\left[r^2(\mathcal{X})K^2\left(\frac{\|\chi - \mathcal{X}_i\|}{h_i}\right)\right] + \mathrm{E}\left[\sigma_\varepsilon^2(\mathcal{X})K^2\left(\frac{\|\chi - \mathcal{X}_i\|}{h_i}\right)\right].$$

Assumption **H1** and (9) ensure that

$$\begin{aligned}
\mathrm{E}\left[Y_i^2 K^2\left(\frac{\|\chi - \mathcal{X}_i\|}{h_i}\right)\right] &= \left[r^2(\chi) + \sigma_\varepsilon^2(\chi)\right]\mathrm{E}\left[K^2\left(\frac{\|\chi - \mathcal{X}_i\|}{h_i}\right)\right][1 + \gamma_i] \\
&= \left[r^2(\chi) + \sigma_\varepsilon^2(\chi)\right]M_2 F(h_i)\left[1 + \gamma_i\right],
\end{aligned}$$

and then from Toeplitz's Lemma, with **H3** and **H4**, it follows that

$$\begin{aligned}
\mathrm{Var}\left[\varphi_n^{[\ell]}(\chi)\right] &= \frac{1}{\left(\sum\limits_{i=1}^n F(h_i)^{1-\ell}\right)^2}\sum_{i=1}^n F(h_i)^{1-2\ell}\left[r^2(\chi) + \sigma_\varepsilon^2(\chi)\right]M_2\left[1 + \gamma_i\right] \\
&= \frac{\beta_{[1-2\ell]}}{\beta_{[1-\ell]}^2}\left[r^2(\chi) + \sigma_\varepsilon^2(\chi)\right]M_2\frac{1}{nF(h_n)}\left[1 + o(1)\right],
\end{aligned}$$

which proves the second assertion of Lemma 4. The covariance term writes

$$\begin{aligned}
\mathrm{Cov}\left[f_n^{[\ell]}(\chi), \varphi_n^{[\ell]}(\chi)\right] &= \frac{1}{\left(\sum\limits_{i=1}^n F(h_i)^{1-\ell}\right)^2}\left\{\mathrm{E}\left[\sum_{i=1}^n\sum_{j=1}^n\frac{Y_iK\left(\frac{\|\chi-\mathcal{X}_i\|}{h_i}\right)K\left(\frac{\|\chi-\mathcal{X}_j\|}{h_j}\right)}{F(h_i)^\ell F(h_j)^\ell}\right]\right. \\
&\quad\left. - \sum_{i=1}^n\frac{\mathrm{E}\left[Y_iK\left(\frac{\|\chi-\mathcal{X}_i\|}{h_i}\right)\right]}{F(h_i)^\ell}\sum_{j=1}^n\frac{\mathrm{E}K\left(\frac{\|\chi-\mathcal{X}_j\|}{h_j}\right)}{F(h_j)^\ell}\right\} \\
&= \frac{1}{\left(\sum\limits_{i=1}^n F(h_i)^{1-\ell}\right)^2}\sum_{i=1}^n\frac{\mathrm{E}\left[Y_iK^2\left(\frac{\|\chi-\mathcal{X}_i\|}{h_i}\right)\right]}{F(h_i)^{2\ell}} \\
&\quad - \frac{1}{\left(\sum\limits_{i=1}^n F(h_i)^{1-\ell}\right)^2}\sum_{i=1}^n\frac{\mathrm{E}\left[Y_iK\left(\frac{\|\chi-\mathcal{X}_i\|}{h_i}\right)\right]\mathrm{E}K\left(\frac{\|\chi-\mathcal{X}_i\|}{h_i}\right)}{F(h_i)^{2\ell}} := I - II.
\end{aligned}$$

Notice that by (6) and (8), one may write

$$II = O\left[\frac{1}{n}\left(B_{n,1-\ell}\right)^{-2}B_{n,2(1-\ell)}\right] = O\left(\frac{1}{nF(h_n)}\right).$$

Next, from assumption **H1** and conditioning on $\mathcal{X}$, we have

$$\mathrm{E}\left[Y_iK^2\left(\frac{\|\chi - \mathcal{X}_i\|}{h_i}\right)\right] = M_2 F(h_i)\left[r(\chi) + \gamma_i\right].$$

It follows that

$$I = \frac{(B_{n,1-\ell})^{-2}}{nF(h_n)}\sum_{i=1}^n\frac{F(h_i)^{1-2\ell}}{nF(h_n)^{1-2\ell}}M_2 r(\chi)\left[1 + \gamma_i\right],$$

and the third assertion of Lemma 4 follows again by applying Toeplitz's Lemma. $\square$

4.1.4 Proof of Lemma 2

Lemma 2 is a direct consequence of Lemmas 3 and 4. $\qquad\qquad\square$

4.2 Proof of Theorem 2

Let us consider the following decomposition

$$r_n^{[\ell]}(\chi) - r(\chi) = \frac{\tilde{\varphi}_n^{[\ell]}(\chi) - r(\chi)f_n^{[\ell]}(\chi)}{f_n^{[\ell]}(\chi)} + \frac{\varphi_n^{[\ell]}(\chi) - \tilde{\varphi}_n^{[\ell]}(\chi)}{f_n^{[\ell]}(\chi)}, \qquad (10)$$

where $\tilde{\varphi}_n^{[\ell]}(\chi)$ is a truncated version of $\varphi_n^{[\ell]}(\chi)$ defined by

$$\tilde{\varphi}_n^{[\ell]}(\chi) = \frac{1}{\sum\limits_{i=1}^{n} F(h_i)^{1-\ell}} \sum_{i=1}^{n} \frac{Y_i}{F(h_i)^{\ell}} \mathbb{1}_{\{|Y_i| \le b_n\}} K\left(\frac{\|\chi - \mathcal{X}_i\|}{h_i}\right), \qquad (11)$$

b_n being a sequence of real numbers which goes to $+\infty$ as $n \to \infty$. Next, for any $\varepsilon > 0$, we have for the residual term of (10)

$$P\left\{\left|\varphi_n^{[\ell]}(\chi) - \tilde{\varphi}_n^{[\ell]}(\chi)\right| > \varepsilon \left[\frac{\ln\ln n}{nF(h_n)}\right]^{\frac{1}{2}}\right\} \le P\left(\bigcup_{i=1}^{n}\{|Y_i| > b_n\}\right) \le \mathrm{E}\left[e^{\lambda|Y|^{\mu}}\right] n^{1-\lambda\delta},$$

where the last inequality follows by setting $b_n = (\delta \ln n)^{\frac{1}{\mu}}$, with the help of Markov's inequality. Assumption **H5** ensures that for any $\varepsilon > 0$,

$$\sum_{n=1}^{\infty} P\left\{\left|\varphi_n^{[\ell]}(\chi) - \tilde{\varphi}_n^{[\ell]}(\chi)\right| > \varepsilon \left[\frac{\ln\ln n}{nF(h_n)}\right]^{\frac{1}{2}}\right\} < \infty \text{ if } \delta > \frac{2}{\lambda},$$

which together with the Borel-Cantelli Lemma imply that

$$\left[\frac{nF(h_n)}{\ln\ln n}\right]^{1/2} \left|\varphi_n^{[\ell]}(\chi) - \tilde{\varphi}_n^{[\ell]}(\chi)\right| \to 0 \text{ a.s, as } n \to \infty. \qquad (12)$$

The main term in (10) writes

$$\begin{aligned}
\tilde{\varphi}_n^{[\ell]}(\chi) - r(\chi)f_n^{[\ell]}(\chi) &= \left\{\tilde{\varphi}_n^{[\ell]}(\chi) - r(\chi)f_n^{[\ell]}(\chi) - \mathrm{E}\left[\tilde{\varphi}_n^{[\ell]}(\chi) - r(\chi)f_n^{[\ell]}(\chi)\right]\right\} \\
&\quad + \left\{\mathrm{E}\left[\tilde{\varphi}_n^{[\ell]}(\chi) - r(\chi)f_n^{[\ell]}(\chi)\right]\right\} := N_1 + N_2. \qquad (13)
\end{aligned}$$

Theorem 2 will be proved if Lemmas 5 and 6 below are established. Indeed, from Lemma 3 we have $\mathrm{E}\left(f_n^{[\ell]}(\chi)\right) = M_1[1 + o(1)]$ and it can be shown (following the same lines of the proof of Lemma 5) that

$$f_n^{[\ell]}(\chi) - \mathrm{E}f_n^{[\ell]}(\chi) = O\left(\sqrt{\frac{\ln\ln n}{nF(h_n)}}\right) \text{ a.s.}$$

$\qquad\qquad\square$

Lemma 5 *Under assumptions* **H1** $-$ **H6**, *we have*

$$\varlimsup_{n\to\infty} \left[\frac{nF(h_n)}{\ln\ln\ n}\right]^{1/2} N_1 = \frac{\left[2\beta_{[1-2\ell]}\sigma_\varepsilon^2(\chi)M_2\right]^{1/2}}{\beta_{[1-\ell]}} \quad a.s.$$

Lemma 6 *Assume that* **H1** $-$ **H5** *hold. If* $\lim_{n\to+\infty} nh_n^2 = 0$, *then*

$$\varlimsup_{n\to\infty} \left[\frac{nF(h_n)}{\ln\ln\ n}\right]^{1/2} N_2 = 0.$$

4.2.1 Proof of Lemma 5

Let us set

$$W_{n,i} = \frac{1}{F(h_i)^\ell} K\left(\frac{\|\chi - \mathcal{X}_i\|}{h_i}\right) \left[Y_i \mathbb{1}_{\{|Y_i|\le b_n\}} - r(\chi)\right] \quad \text{and} \quad Z_{n,i} = W_{n,i} - \mathrm{E}W_{n,i},$$

and define

$$S_n = \sum_{i=1}^n Z_{n,i} \quad \text{and} \quad V_n = \sum_{i=1}^n \mathrm{E}Z_{n,i}^2.$$

Observe that

$$
\begin{aligned}
V_n &= \sum_{i=1}^n F(h_i)^{-2\ell}\left\{\mathrm{E}\left(K^2\left(\frac{\|\chi - \mathcal{X}\|}{h_i}\right)[Y - r(\chi)]^2\right)\right.\\
&\quad \left.+\mathrm{E}\left(K^2\left(\frac{\|\chi - \mathcal{X}\|}{h_i}\right)Y\left[2r(\chi) - Y\right]\mathbb{1}_{\{|Y|>b_n\}}\right)\right\}\\
&\quad -\sum_{i=1}^n F(h_i)^{-2\ell}\mathrm{E}^2\left(K\left(\frac{\|\chi - \mathcal{X}\|}{h_i}\right)\left[Y\mathbb{1}_{\{|Y|\le b_n\}} - r(\chi)\right]\right)\\
&:= A_1 + A_2 - A_3. \tag{14}
\end{aligned}
$$

A_1 writes

$$
\begin{aligned}
A_1 &= \sum_{i=1}^n F(h_i)^{-2\ell}\mathrm{E}\left\{K^2\left(\frac{\|\chi - \mathcal{X}\|}{h_i}\right)\mathrm{E}\left[(Y - r(\chi))^2|\mathcal{X}\right]\right\}\\
&= \sum_{i=1}^n \frac{\sigma_\varepsilon^2(\chi)\mathrm{E}K^2\left(\frac{\|\chi-\mathcal{X}\|}{h_i}\right)}{F(h_i)^{2\ell}} + \frac{\mathrm{E}\left[K^2\left(\frac{\|\chi-\mathcal{X}\|}{h_i}\right)\{\sigma_\varepsilon^2(\mathcal{X}) - \sigma_\varepsilon^2(\chi)\}\right]}{F(h_i)^{2\ell}}\\
&:= A_{11} + A_{12}.
\end{aligned}
$$

From **H2**, by applying Fubini's Theorem, we have

$$A_{11} = \sum_{i=1}^{n} F(h_i)^{1-2\ell} \sigma_\varepsilon^2(\chi) \left[K^2(1) - \int_0^1 (K^2(s))' \tau_{h_i}(s) ds \right].$$

Applying again Toeplitz's Lemma, **H3** and **H4** allow to get

$$\frac{A_{11}}{nF(h_n)^{1-2\ell}} \to \beta_{[1-2\ell]} \sigma_\varepsilon^2(\chi) M_2, \quad \text{as} \ \ n \to +\infty. \tag{15}$$

The second term of the decomposition of A_1 is bounded using (9) as

$$A_{12} \le \sum_{i=1}^{n} F(h_i)^{1-2\ell} \sup_{\chi' \in B(\chi, h_i)} |\sigma_\varepsilon^2(\chi') - \sigma_\varepsilon^2(\chi)| \left[K^2(1) - \int_0^1 (K^2(s))' \tau_{h_i}(s) ds \right],$$

while the continuity of σ_ε^2 (**H1**) with Toeplitz's Lemma ensure that

$$\frac{A_{12}}{nF(h_n)^{1-2\ell}} \to 0, \ \text{as} \ n \to +\infty. \tag{16}$$

Now, let us study the term A_2 appearing in the decomposition of V_n. Using Cauchy-Schwartz's inequality, and denoting $\|K\|_\infty := \sup_{t \in [0,1]} K(t)$, we get

$$
\begin{aligned}
A_2 \ &\le \ \|K\|_\infty^2 \sum_{i=1}^{n} F(h_i)^{-2\ell} \left\{ \mathrm{E}\left(Y^2 [2r(\chi) - Y]^2 \right) P\left(|Y| > b_n \right) \right\}^{\frac{1}{2}} \\
&\le \ 3Q_n \|K\|_\infty^2 \sum_{i=1}^{n} F(h_i)^{-2\ell},
\end{aligned}
$$

where $Q_n = \left(\max \left\{ \mathrm{E}\left(Y^4 \right), 4|r(\chi)| \mathrm{E}|Y|^3, 4r^2(\chi) \mathrm{E}\left(Y^2 \right) \right\} P\left(|Y| > b_n \right) \right)^{\frac{1}{2}}$.
We deduce from **H4** and **H5**, with the choice $b_n = (\delta \ln n)^{1/\mu}$, that

$$\frac{A_2}{nF(h_n)^{1-2\ell}} = O\left[\frac{e^{-\frac{\lambda b_n^\mu}{2}} (\ln n)^{\frac{2}{\mu}}}{F(h_n)} \right] \to 0, \ \text{as} \ n \to +\infty \ \text{with} \ \delta > \frac{2}{\lambda}. \tag{17}$$

The last term A_3 is bounded

$$|A_3| \le b_n^2 [1 + o(1)] \sum_{i=1}^{n} F(h_i)^{2-2\ell} \left[K(1) - \int_0^1 (K'(s)) \tau_{h_i}(s) ds \right]^2.$$

It follows from **H6** that

$$\frac{A_3}{nF(h_n)^{1-2\ell}} = O\left[F(h_n)(\ln n)^{\frac{2}{\mu}} \right] \to 0, \ \text{as} \ n \to +\infty. \tag{18}$$

relations (15), (16), (17) and (18) can be used to derive the following equivalence

$$V_n \sim nF(h_n)^{1-2\ell}\beta_{[1-2\ell]}\sigma_\varepsilon^2(\chi)M_2, \text{ as } n \to +\infty. \tag{19}$$

Next, since $\frac{\ln F(h_n)}{\ln n} \to 0$ as $n \to +\infty$, then the first part of **H6** implies that

$$\frac{nF(h_n)(\ln n)^{-\frac{2}{\mu}}}{\ln\left[nF(h_n)^{1-2\ell}\right]\left\{\ln\ln\left[nF(h_n)^{1-2\ell}\right]\right\}^{2(\alpha+1)}} \to \infty, \quad \text{as } n \to +\infty.$$

Setting $b_n = (\delta \ln n)^{\frac{1}{\mu}}$, there exists $n_0 \geq 1$ such that for any $i \geq n_0$, we have

$$\frac{iF(h_i)(\ln i)^{-\frac{2}{\mu}}}{\ln\left[iF(h_i)^{1-2\ell}\right]\left\{\ln\ln\left[iF(h_i)^{1-2\ell}\right]\right\}^{2(\alpha+1)}} > \frac{2\|K\|_\infty^2 \max\left\{|r(\chi)|^2, (\delta\ln i)^{\frac{2}{\mu}}\right\}}{F(h_i)^{2\ell}} \geq Z_{n,i}^2.$$

Hence, the event $\left\{Z_{n,i}^2 > \frac{iF(h_i)^{1-2\ell}}{\ln\left[iF(h_i)^{1-2\ell}\right]\left\{\ln\ln\left[iF(h_i)^{1-2\ell}\right]\right\}^{2(\alpha+1)}}\right\}$ is empty for $i \geq n_0$. We deduce from (19) that

$$\sum_{i=1}^{\infty} \frac{(\ln\ln V_i)^\alpha}{V_i}\mathrm{E}\left(Z_{n,i}^2 \mathbb{1}_{\left\{Z_{n,i}^2 > \frac{V_i}{\ln V_i(\ln\ln V_i)^{2(\alpha+1)}}\right\}}\right) < \infty.$$

Let S be a random function defined on $[0, +\infty[$ such that for any $t \in [V_n, V_{n+1}[, S(t) = S_n$. Using Theorem 3.1 in Jain et al. (1975), there exists a Brownian motion ξ such that

$$\left|\frac{S(t) - \xi(t)}{(2t\ln\ln t)^{\frac{1}{2}}}\right| = o\left[(\ln\ln t)^{-\frac{\alpha}{2}}\right] a.s., \quad \text{as } t \to \infty, \text{ for any } t \in [V_n, V_{n+1}[.$$

It follows that

$$\overline{\lim_{t\to\infty}} \frac{S(t)}{(2t\ln\ln t)^{\frac{1}{2}}} = \overline{\lim_{t\to\infty}}\left[\frac{S(t) - \xi(t)}{(2t\ln\ln t)^{\frac{1}{2}}} + \frac{\xi(t)}{(2t\ln\ln t)^{\frac{1}{2}}}\right] = 1 \ a.s.$$

Then we have

$$\frac{S_n}{\sqrt{2V_n\ln\ln V_n}} \to 1 \ a.s., \text{ as } n \to \infty, \tag{20}$$

by virtue of the definition of S and the fact that $\frac{V_{n+1}}{V_n} \to 1$ as $n \to \infty$. From (19), we have

$$\lim_{n\to\infty} \frac{\left\{nF(h_n)^{1-2\ell}\ln\ln\left[nF(h_n)^{1-2\ell}\right]\right\}^{1/2}B_{n,1-\ell}}{(2V_n\ln\ln V_n)^{\frac{1}{2}}} = \frac{\beta_{[1-\ell]}}{\left\{2\beta_{[1-2\ell]}\sigma_\varepsilon^2(\chi)M_2\right\}^{1/2}}.$$

Lemma 5 follows from the last convergence and the fact that $S_n = N_1 \sum_{i=1}^{n} F(h_i)^{1-\ell}$, with the help of (20). $\qquad\square$

4.2.2 Proof of Lemma 6

We have

$$
\begin{aligned}
N_2 &= \frac{1}{\sum\limits_{i=1}^{n} F(h_i)^{1-\ell}} \sum_{i=1}^{n} F(h_i)^{-\ell} \left\{ \mathrm{E}\left[K\left(\frac{\|\chi - \mathcal{X}\|}{h_i} \right) (r(\mathcal{X}) - r(\chi)) \right] \right. \\
&\left. - \mathrm{E}\left[K\left(\frac{\|\chi - \mathcal{X}\|}{h_i} \right) Y \mathbb{1}_{\{|Y|>b_n\}} \right] \right\} := A + B.
\end{aligned} \tag{21}
$$

As in the proof of Lemma 1, we can write

$$
A = h_n \frac{\alpha_{[\ell]}}{\beta_{[1-\ell]}} \varphi'(0) M_0 \left[1 + o(1) \right], \tag{22}
$$

and then,

$$
\left[\frac{nF(h_n)}{\ln \ln n} \right]^{1/2} A = \left[\frac{nF(h_n)}{\ln \ln n} \right]^{1/2} h_n \frac{\alpha_{[\ell]}}{\beta_{[1-\ell]}} \varphi'(0) M_0 \left[1 + o(1) \right] = o(1),
$$

where the last equality follows from the condition $nh_n^2 \to 0$. For the second term of the right-hand-side in (21), using Cauchy-Schwartz's inequality and the boundedness of the kernel K, we get

$$
|B| \leq \frac{\|K\|_\infty}{\sum\limits_{i=1}^{n} F(h_i)^{1-\ell}} \sum_{i=1}^{n} F(h_i)^{-\ell} \left\{ \mathrm{E}\left[Y_i^2 \right] \mathrm{P}\left[|Y_i| > b_n \right] \right\}^{1/2}.
$$

From Markov's inequality combined with (4), it follows that

$$
\begin{aligned}
|B| &\leq \frac{\|K\|_\infty}{\sum\limits_{i=1}^{n} F(h_i)^{1-\ell}} \sum_{i=1}^{n} F(h_i)^{-\ell} \left\{ \mathrm{E}\left[Y_i^2 \right] \mathrm{E}\left[e^{\lambda |Y_i|^\mu} \right] e^{-\lambda b_n^\mu} \right\}^{1/2} \\
&= O\left(\frac{1}{nF(h_n)} \frac{B_{n,-\ell}}{B_{n,1-\ell}} n^{1-\lambda\delta} (\ln n)^{2/\mu} \right), \tag{23}
\end{aligned}
$$

which gives

$$
\left[\frac{nF(h_n)}{\ln \ln n} \right]^{1/2} B = O\left(\frac{1}{\sqrt{\ln \ln n}} \frac{1}{\sqrt{nF(h_n)}} n^{1-\lambda\delta} (\ln n)^{2/\mu} \right) = o(1) \text{ if } \delta > \frac{1}{\lambda},
$$

and Lemma 6 is proved. $\qquad \square$

4.3 Proof of Theorem 3

Using the decomposition (10), we have to show that

$$\sqrt{nF(h_n)}\left[\varphi_n^{[\ell]}(\chi) - \tilde{\varphi}_n^{[\ell]}(\chi)\right] \to 0 \text{ a.s.} \qquad (24)$$

and

$$\sqrt{nF(h_n)}\left[\tilde{\varphi}_n^{[\ell]}(\chi) - r(\chi)f_n^{[\ell]}(\chi)\right] \xrightarrow{\mathcal{D}} \mathcal{N}\left(\frac{cM_0\varphi'(0)\alpha_{[\ell]}}{\beta_{[1-\ell]}}, \frac{\beta_{[1-2\ell]}M_2\sigma_\varepsilon^2(\chi)}{\beta_{[1-\ell]}^2}\right), (25)$$

where $\tilde{\varphi}_n^{[\ell]}$ is defined in (11) and c is such that $\lim\limits_{n\to\infty}\sqrt{nF(h_n)}h_n = c$, since $f_n^{[\ell]}(\chi) \xrightarrow{P} M_1$. This later follows from the first parts of Lemmas 3 and 4. For (24), we follow the same lines of proof of (12) substituting $\left[\dfrac{\ln\ln n}{nF(h_n)}\right]^{1/2}$ by $\dfrac{1}{\sqrt{nF(h_n)}}$ to get the desired result. About (25), using the decomposition (13), it remains to prove Lemmas 7 and 8 below.

Lemma 7 *Assume that Assumptions* **H1** $-$ **H5** *and* **H7** *hold. Then*

$$\sqrt{nF(h_n)}N_1 \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \frac{\beta_{[1-2\ell]}}{\beta_{[1-\ell]}^2}\sigma_\varepsilon^2(\chi)M_2\right).$$

Lemma 8 *Assume that Assumptions* **H1** $-$ **H5** *hold. If there exists $c \geq 0$ such that $\lim\limits_{n\to\infty} h_n\sqrt{nF(h_n)} = c$, then*

$$\lim_{n\to\infty}\sqrt{nF(h_n)}N_2 = c\frac{\alpha_{[\ell]}}{\beta_{[1-\ell]}}\varphi'(0)M_0.$$

4.3.1 Proof of Lemma 7

Setting

$$W'_{n,i} = \frac{\sqrt{nF(h_n)}}{\sum_{i=1}^n F(h_i)^{1-\ell}}W_{n,i} \quad \text{and} \quad Z'_{n,i} = W'_{n,i} - \mathrm{E}W'_{n,i},$$

where $W_{n,i}$ is defined in the proof of Theorem 2, then

$$\sqrt{nF(h_n)}N_1 = \sum_{i=1}^n Z'_{n,i}.$$

To prove Lemma 7, we first prove that

$$\lim_{n\to\infty}\sum_{i=1}^n \mathrm{E}(Z'^2_{n,i}) = \frac{\beta_{[1-2\ell]}}{\beta_{[1-\ell]}^2}\sigma_\varepsilon^2(\chi)M_2, \qquad (26)$$

and then check that $W'_{n,i}$ satisfies the Lyapounov's condition. Next, from (19) we have

$$
\begin{aligned}
\sum_{i=1}^{n} \mathrm{E}(Z'^2_{n,i}) &= \frac{nF(h_n)}{\left(\sum_{i=1}^{n} F(h_i)^{1-\ell}\right)^2} V_n = \frac{1}{nF(h_n)^{1-2\ell}} \frac{1}{B^2_{n,1-\ell}} V_n \\
&= \frac{\beta_{[1-2\ell]}}{\beta^2_{[1-\ell]}} \sigma^2_\varepsilon(\chi) M_2 \left[1 + o(1)\right]
\end{aligned}
$$

which proves (26). To check the Lyapounov's condition, set $p > 2$, we have

$$
\sum_{i=1}^{n} \mathrm{E}\left(|Z'_{n,i}|^p\right) = \sum_{i=1}^{n} \mathrm{E}\left(|Z'_{n,i}|^{p-2} Z'^2_{n,i}\right).
$$

Since

$$
|W'_{n,i}| \leq \frac{\|K\|_\infty \sqrt{nF(h_n)}}{\sum_{i=1}^{n} F(h_i)^{1-\ell}} F(h_i)^{-\ell} |b_n - r(\chi)|,
$$

it follows that

$$
\sum_{i=1}^{n} \mathrm{E}\left(|Z'_{n,i}|^p\right) \leq \frac{(nF(h_n))^{\frac{p}{2}} \sum_{i=1}^{n} F(h_i)^{-p\ell} Var\left(K\left(\frac{\|\chi - \mathcal{X}_i\|}{h_i}\right)\left(Y_i \mathbb{1}_{\{|Y_i| \leq b_n\}} - r(\chi)\right)\right)}{2^{2-p} \|K\|_\infty^{2-p} |b_n - r(\chi)|^{2-p} \left(\sum_{i=1}^{n} F(h_i)^{1-\ell}\right)^p}.
$$

Using the same decomposition as in (14), we have

$$
\begin{aligned}
&\sum_{i=1}^{n} F(h_i)^{-p\ell} Var\left(K\left(\frac{\|\chi - \mathcal{X}_i\|}{h_i}\right)\left(Y_i \mathbb{1}_{\{|Y_i| \leq b_n\}} - r(\chi)\right)\right) \\
&= \sum_{i=1}^{n} F(h_i)^{-p\ell} \left\{ \mathrm{E}\left(K^2\left(\frac{\|\chi - \mathcal{X}\|}{h_i}\right)[Y - r(\chi)]^2\right) \right. \\
&\quad \left. + \mathrm{E}\left(K^2\left(\frac{\|\chi - \mathcal{X}\|}{h_i}\right) Y \left[2r(\chi) - Y\right] \mathbb{1}_{\{|Y| > b_n\}}\right) \right\} \\
&\quad - \sum_{i=1}^{n} \frac{\mathrm{E}^2 K\left(\frac{\|\chi - \mathcal{X}\|}{h_i}\right) \left[Y \mathbb{1}_{\{|Y| \leq b_n\}} - r(\chi)\right]}{F(h_i)^{p\ell}} := B_1 + B_2 - B_3. \quad (27)
\end{aligned}
$$

Setting $b_n = (\delta \ln n)^{\frac{1}{\mu}}$ for some $\delta, \mu > 0$ and following the same lines as in the proof of (15), (16), (17) and (18) substituting the exponent 2 by p in all the expressions, we have

$$
B_1 = O\left(nF(h_n)^{1-p\ell}\right).
$$

From Toeplitz's Lemma, we can write

$$\frac{(nF(h_n))^{\frac{p}{2}}}{(\sum_{i=1}^n F(h_i)^{1-\ell})^p} |b_n - r(\chi)|^{p-2} B_1 = O\left(\left(\frac{(\ln n)^{\frac{1}{\mu}}}{\sqrt{nF(h_n)}}\right)^{p-2}\right) = o(1).$$

Next, for the second expression B_2 of (27), we get

$$\frac{B_2}{nF(h_n)^{1-p\ell}} = O\left(\frac{\exp\left(-\frac{\lambda b_n^\mu}{2}\right)(\ln n)^{\frac{2}{\mu}}}{F(h_n)}\right) = o(1) \text{ with } \delta > \frac{2}{\lambda}.$$

It follows again from Toeplitz's Lemma that

$$\frac{(nF(h_n))^{\frac{p}{2}}}{(\sum_{i=1}^n F(h_i)^{1-\ell})^p} |b_n - r(\chi)|^{p-2} B_2 = O\left(\left(\frac{(\ln n)^{\frac{1}{\mu}}}{\sqrt{nF(h_n)}}\right)^{p-2}\right) = o(1).$$

In the same manner as in the proof of (18), we have

$$\frac{B_3}{nF(h_n)^{1-p\ell}} = O\left[F(h_n)(\ln n)^{\frac{2}{\mu}}\right].$$

Then

$$\frac{(nF(h_n))^{\frac{p}{2}}}{(\sum_{i=1}^n F(h_i)^{1-\ell})^p} |b_n - r(\chi)|^{p-2} B_3 = O\left(\left(\frac{(\ln n)^{\frac{1}{\mu}}}{\sqrt{nF(h_n)}}\right)^{p-2}\right) = o(1),$$

which concludes the proof of Lemma 7. □

4.3.2 Proof of Lemma 8

We go back to the decomposition of (21) in the proof of Lemma 6.
On one hand, from (22), we write

$$\sqrt{nF(h_n)}A = \sqrt{nF(h_n)}h_n \frac{\alpha_{[\ell]}}{\beta_{[1-\ell]}} \varphi'(0) M_0 \left[1 + o(1)\right]. \tag{28}$$

On the other hand from (23), we get

$$\sqrt{nF(h_n)}B = O\left(\frac{1}{\sqrt{nF(h_n)}}\right) = o(1) \text{ if } \delta > \frac{1}{\lambda}, \tag{29}$$

and Lemma 8 follows from the combination of (28) and (29). □

4.3.3 Proof of Corollary 2

From the standard Glivenko-Cantelli type results, we have

$$\frac{F(h_n)}{\widehat{F}(h_n)} \xrightarrow{\mathbb{P}} 1. \tag{30}$$

Then, to prove Corollary 2, it suffices to show that $\widehat{\beta}_{[1]} \xrightarrow{\mathbb{P}} \beta_{[1]}$. To see this, we use the following result

$$\sup_{x \in \mathbb{R}} \left| \widehat{F}(x) - F(x) \right| = O((\ln n / n)^{1/2}),$$

established by Devroye (1977) and write

$$
\begin{aligned}
\widehat{\beta}_{[1]} - \beta_{[1]} &= \frac{1}{n\widehat{F}(h_n)} \sum_{i=1}^{n} \left(\widehat{F}(h_i) - F(h_i) \right) + \frac{F(h_n)}{\widehat{F}(h_n)} B_{n,1} - \beta_{[1]} \\
&\leq \frac{F(h_n)}{\widehat{F}(h_n)} \frac{1}{nF(h_n)} \sup_{x \in \mathbb{R}} \left| \widehat{F}(x) - F(x) \right| + \left| \frac{F(h_n)}{\widehat{F}(h_n)} B_{n,1} - \beta_{[1]} \right| = o_{\mathbb{P}}(1),
\end{aligned}
$$

by virtue of **H4(ii)** and (30). □

References

Ahmad, I. and Lin, P.E. (1976). Nonparametric sequential estimation of a multiple regression function. *Bull. Math. Statist.*, **17**, 63-75.

Amiri, A. (2012). Recursive regression estimators with application to nonparametric prediction. *J. Nonparametr. Stat.* **24** (1), 169-186.

Besse, P., Cardot, H. and Ferraty, F. (1997). Simultaneous nonparametric regressions of unbalanced longitudinal data. *Comp. Statist. Data Anal.*, **24**, 255-270.

Bosq, D., and Cheze-Payaud, N. (1999). Optimal asymptotic quadratic error of nonparametric regression function estimates for a continuous-time process from sampled-data. *Statistics*, **32** (3), 229–247.

Cardot, H., Ferraty, F. and Sarda, P. (2003). Splines estimators for the functional linear model. *Statistica Sinica*, **13**, 571-591.

Collomb, G. (1976). Estimation nonparamétrique de la régression par la méthode du noyau. *Unpublished PhD Thesis*, Université Paul Sabatier, Toulouse, France.

Crambes, C., Kneip, A. and Sarda, P. (2009). Smoothing splines estimators for functional linear regression, *Ann. Statist.*, **37**, 35-72.

Devroye, L., (1977). A uniform bound for the deviation of empirical distribution functions. *J. Mult. Anal.*, **7**, 594-597.

Devroye, L. and Wagner, T.J. (1980). Distribution-free consistency results in nonparametric discrimination and regression function estimation. *Ann. Statist.*, **8**, 231-239.

Ferraty, F. Mas, A. and Vieu, P. (2007). Nonparametric regression on functional data: Inference and practical aspects. *Aust. N. Z. J. Stat.,* **49** (3), 267–286.

Ferraty, F. and Romain, Y. (2010). Handbook on functional data analysis and related fields. *Oxford University Press*, Oxford.

Ferraty, F. and Vieu, P. (2006). Nonparametric Modelling for Functional Data. Methods, Theory, Applications and Implementations. *Springer-Verlag*, London.

Frank, I.E. and Friedman, J.H. (1993). A statistical view of some chemometrics regression tools. *Technometrics*, **35**, 109–135.

Gonzàlez Manteiga, W. and Vieu, P. (2007). Statistics for functional data. *Comp. Stat. and Data Anal.*, **51**, 4788-4792.

Jain, C., Jogdeo, K. and Stout, W. (1975). Upper and lower functions for martingales and mixing processes, *Ann. Probab.* **3** (1), 119–145.

Ling, N. and Wu, Y. (2012). Consistency of modified kernel regression estimation for functional data, *Statistics* **46** (2), 149–158.

Masry, E. (1986). Recursive probability density estimation for weakly dependent stationary processes, *IEEE Trans. Inform. Theory* **32** (2), 254–267.

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

Rachdi, M. and Vieu, P. (2007). Nonparametric regression for functional data: automatic smoothing parameter selection. *J. Statist. Plann. Inference*, **137**, 2784-2801.

Ramsay, J.O. and Dalzell, C.J. (1991). Some tools for functional data analysis (with discussion), *J. Roy. Statist. Soc., Ser. B*, **53**, 539-572.

Ramsay, J.O. and Silverman, B.W. (2002). Applied functional data analysis. *Springer-Verlag*, New-York.

Ramsay, J.O. and Silverman, B.W. (2005). *Functional data analysis (2nd Ed.)*. Springer-Verlag, New-York.

Roussas, G. G.(1992). Exact rates of almost sure convergence of a recursive kernel estimate of a probability densiy function: Application to regression and hazard rate estimation, *J. Nonparametr. Stat.,* **1** (3), 171– 195.