



**HAL**  
open science

## Recursive estimation of nonparametric regression with functional covariate.

Aboubacar Amiri, Christophe Crambes, Baba Thiam

► **To cite this version:**

Aboubacar Amiri, Christophe Crambes, Baba Thiam. Recursive estimation of nonparametric regression with functional covariate.. 2012. hal-00750894

**HAL Id: hal-00750894**

**<https://hal.science/hal-00750894>**

Preprint submitted on 12 Nov 2012

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Recursive estimation of nonparametric regression with functional covariate.

Aboubacar AMIRI<sup>‡\*</sup>; Christophe CRAMBES<sup>†</sup>    Baba THIAM<sup>‡</sup>

November 12, 2012

## Abstract

The main purpose of this work 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 a real data set study.

*Keywords* : Functional data, recursive kernel estimators, regression function, quadratic error, almost sure convergence, asymptotic normality.

*MSC* : 62G05, 62G07, 62G08.

## 1 Introduction

Functional data analysis is a branch of statistics that has been the object of many studies and developments these last years. This kind of data appears in many practical situations, as soon as one is interested on a continuous phenomenon for instance. For this reason, the possible application fields propitious for the use of functional data are very wide: climatology, economics, linguistics, medicine, . . . Since the pioneer works ([Ramsay and Dalzell(1991)], [Frank and Friedman (1993)]), many developments have been investigated, in order to build theory and methods around functional data, for instance how it is possible to define the mean, or the variance of functional data, what kind of model it is possible to consider with functional data, and so on

---

\*Corresponding author: aboubacar.amiri@univ-lille3.fr

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

<sup>‡</sup>Université Lille Nord de France, Université Lille 3, Laboratoire EQUIPPE EA 4018, Villeneuve d'Ascq, France.

... These papers also highlight the drawback of a mere use of multivariate methods with this kind of data, and on the contrary suggest to consider these data as objects belonging to some functional space. The monographs of [Ramsay and Silverman(2005), Ramsay and Silverman(2002)] present an overview on both theoretical and practical aspects of functional data analysis.

One of the most studied models in this functional setting is the regression model when the variable of interest  $Y$  is real and the covariate  $\mathcal{X}$  belongs to some functional space  $\mathcal{E}$ , endowed with a semi-norm  $\|\cdot\|$ . Then, the regression model writes

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

where  $r : \mathcal{E} \rightarrow \mathbb{R}$  is an operator and  $\varepsilon$  is an error random variable. Many works have been done around this model when the operator  $r$  is supposed to be linear, contributing to the popularity of the so-called functional linear model. In this linear context, the operator  $r$  writes  $\langle \alpha, \cdot \rangle$  where  $\langle \cdot, \cdot \rangle$  stands for an inner product of the space  $\mathcal{E}$  and  $\alpha$  belongs to  $\mathcal{E}$ . The goal is then to estimate the unknown function  $\alpha$ . We refer the reader for instance to the works of [Cardot et al.(2003)] or [Crambes et al.(2009)] for different methods to estimate  $\alpha$ . Another way to approach the model (1) is to think in a nonparametric way. This direction has also been investigated by many authors. Recent advances on the topic have been the object of monographs by [Ferraty and Vieu(2006)], [Ferraty and Romain(2010)], giving theoretical and practical properties of a kernel estimator of the operator  $r$ . More precisely, if  $(\mathcal{X}_i, Y_i)_{i=1, \dots, n}$  is a sample of independent and identically distributed couples with the same law as  $(\mathcal{X}, Y)$ , this kernel estimator is defined, for all  $\chi \in \mathcal{E}$ , by

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

where  $K$  is a kernel and  $h > 0$  is a bandwidth. In the dependent case, [Masry(2005)] have considered the asymptotic normality of (2), while the almost sure convergence was obtained by [Ling and Wu (2012)]. This nonparametric regression estimator raises several problems, as the choice of the semi-norm  $\|\cdot\|$  of the space  $\mathcal{E}$ , the choice of the bandwidth, ... Concerning the bandwidth, when the covariate is real, many solutions has been considered, like for instance cross validation. Recently, in the multivariate setting, [Amiri(2012)] studied an estimator using a sequence of bandwidths that allows to compute this estimator in a recursive way, generalizing previous works of [Devroye and Wagner(1980)], [Ahmad and Lin(1976)]. This esti-

mator shows good theoretical properties, from the point of view of mean square error and almost sure convergence. It also have some practical interests: for instance, it presents a computational gain of time when one wants to predict new values of the variable of interest when new observations appear. It is not the case for the basic kernel estimator which has to be computed again on the whole sample. The purpose of this work is to adapt the recursive estimator studied in [Amiri(2012)] to the case where the covariate is of functional nature.

The remainder of the paper is organized as follows. In section 2, we define the recursive estimator of the operator  $r$  when the covariate  $\mathcal{X}$  is functional and we present the asymptotic properties of this estimator. In section 3, we evaluate the performances of our estimator with a simulation study and the treatment of a real dataset. Finally, the proofs of the theoretical results are postponed to 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  $(\mathcal{X}_i, Y_i)_{i=1, \dots, n}$  is a sample of  $n$  random variables independent and identically distributed, having the same distribution as  $(\mathcal{X}, Y)$ . The model (1) is then rewritten as

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

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

Nonparametric regression aims to estimate the functional  $r(\chi) := E(Y | \mathcal{X} = \chi)$ , 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

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

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

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

with

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

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

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 famous ones, say recursive ( $\ell = 0$ ) and semi recursive ( $\ell = 1$ ) estimators. The recursive property of this class of estimators is clearly useful in sequential investigations and also for large sample size, since addition of a new observation means the non-recursive estimators must be entirely recomputed. Besides, we are required to store extensive data in order to calculate them.

We will assume that the following assumptions hold.

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

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

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

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

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

Assumptions **H1**, **H2** and the first part of **H4** are classical in nonparametric regression. They have been used by [Ferraty et al.(2007)] and are the same as those classically used in the finite-dimensional setting. The conditions  $A_{n,\ell} \rightarrow \alpha_{[\ell]} < \infty$  and **H4(ii)** are particular to the recursive problem and they are also the same as the ones used in the finite-dimensional case. Note that  $F$  plays a crucial role in our calculus, its limit at zero, and for a fixed  $\chi$  is known as ‘small ball’ probability. Then, before announcing our results, let us give typical examples of bandwidths and small ball probabilities satisfying

**H3** and **H4** (see [Ferraty et al.(2007)] for more details).

If  $\mathcal{X}$  is fractal (or geometric) process, then the small ball probabilities are of the form  $F(t) \sim c'_\chi t^\kappa$ , where  $c'_\chi$  and  $\kappa$  are positive constants, and  $\|\cdot\|$  may be a supremum norm, an  $L^p$  norm or a Besov norm. In this case, the choice of bandwidth  $h_n = An^{-\delta}$  with  $A > 0$  and  $0 < \delta < 1$  implies that  $F(h_n) = c''_\chi n^{-\delta\kappa}$ ,  $c''_\chi > 0$ . Then, **H3** and **H4** hold as soon as  $\delta\kappa < 1$ . Indeed, assumption **H3** and the first part of **H4** are clearly unrestrictive, since they are the same as those used in the non-recursive case. Concerning **H4(ii)**, if  $\delta\kappa r < 1$ , then  $\sum_{i=1}^n i^{-\delta\kappa r} \sim \frac{n^{1-\delta\kappa r}}{1-\delta\kappa r}$ , so that, the condition is satisfied as soon as  $\beta_{[r]} = \frac{1}{1-\delta\kappa r}$ . The same argument is also valid for  $A_{n,\ell}$ , if  $\max\{\kappa r, 1 + \kappa(1 - \ell)\} < 1/\delta$ .

## 2.2 Main results

As in [Ferraty et al.(2007)], let us introduce the following notations:

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

Now, we can establish the asymptotic mean square error of our recursive estimate.

**Theorem 1** *Under the assumptions **H1** – **H4**,*

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

Theorem 1 is an extension of the work of [Ferraty et al.(2007)] to the class of recursive estimators. Using the bias-variance representation, with the help of 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  $nF(h_n)h_n^2 \rightarrow c$ , as  $n \rightarrow \infty$ , then*

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

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

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

In the finite-dimensional setting and for continuous time processes, a similar result was established by [Bosq and Cheze-Payaud(1999).] for the Nadaraya-Watson estimator.

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

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

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

Assumption **H5** is clearly checked if  $Y$  is bounded and implies that

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

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

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

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

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

and (4) follows. An example of sequence of random variables  $Y_i$  satisfying **H5** (and then (4)) is the standard gaussian, with  $\lambda = 1$  and  $\mu = 2$ . Relation (4) have been used in the multivariate framework by [Bosq and Cheze-Payaud(1999).] in order to establish the optimal quadratic error of the Nadaraya-Watson estimator. Assumption **H6** is satisfied as soon as  $\mathcal{X}$  is fractal or non smooth, while the condition  $\lim_{n \rightarrow \infty} F(h_n)(\ln n)^{\frac{2}{\mu}} = 0$  is not necessary when  $\mu \geq 2$ .

Now, we can write the following theorem for our estimator of the regression operator.

**Theorem 2** Assume that **H1** – **H6** hold. If  $\lim_{n \rightarrow +\infty} nh_n^2 = 0$ , then

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

The choices of bandwidths and small ball probabilities given previously are typical examples satisfying the condition  $\lim_{n \rightarrow +\infty} nh_n^2 = 0$ . The case  $\ell = 1$  of Theorem 2 is an extension to the functional setting of the result of [Roussas(1992)] 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)]). Also conversely to the non recursive case, the rate of convergence of the recursive estimators are obtained with exact upper bounds.

To get the asymptotic normality, we will suppose the following additional assumption, which is clearly verified by the choices of bandwidths and small ball probabilities given above.

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

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

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

This result is similar to the one obtained by [Ferraty et al.(2007)] in the non recursive case. Let us mention that, the choices of bandwidths and small ball probabilities given above imply that  $\frac{\beta_{[1-2\ell]}}{\beta_{[1-\ell]}^2} < 1$ . Then, the recursive estimators are more efficient than classical estimators, in the sense that their asymptotic variance is small.

In practice, if we need to construct confidence bands for the regression function  $r$ , the constants involved in Theorem 3 need to be estimated. In particular, as mentioned in [Ferraty et al.(2007)], if we choose the simple uniform kernel, we can find explicit values of the constants  $M_1$  and  $M_2$ . About conditional variance  $\sigma_\varepsilon^2(\chi)$  it may be estimated by mean of the functional kernel regression technique since it can be rewritten as

$$\sigma_\varepsilon^2(\chi) = \mathbb{E}(Y^2 | \mathcal{X} = \chi) - (\mathbb{E}(Y | \mathcal{X} = \chi))^2.$$



### 3 Simulation study and real dataset example

In order to see the behavior of our recursive estimator in practice, we consider in this section a simulation study. We simulate our data in the following way. The curves  $\mathcal{X}_1, \dots, \mathcal{X}_n$  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  $\varepsilon$  is simulated as a gaussian random variable with mean 0 and standard deviation 0.1. The simulations are repeated 500 times in order to compute the prediction errors for a new curve  $\chi$ , also simulated as a standard Brownian motion on  $[0, 1]$ .

In our functional context, our 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  $\ell$  and the distribution function  $F$  in the case  $\ell \neq 0$ . Since the choice of the kernel  $K$  is not crucial, we use the quadratic kernel, defined by  $K(u) = (1 - u^2) \mathbb{1}_{[0,1]}(u)$  for all  $u \in \mathbb{R}$ , which is known to behave correctly in practice, and easy to implement. About the distribution function  $F$ , we estimate it by the empirical distribution function, which is known to be uniformly convergent.

#### 3.1 Choice of the bandwidth

In this simulation, the semi-norm is based on the principal components analysis of the curves, keeping 3 principal components (see [Besse et al.(1997)] for a description of this semi-norm), while  $\ell$  is fixed equal to 0. We will see below that this parameter  $\ell$  is not much influent in the behavior of the estimator.

We choose to take a sequence of bandwidths  $h_i = C \max_{i=1, \dots, n} \|\mathcal{X}_i - \chi\| i^{-\nu}$ ,

for  $i = 1, \dots, n$ , with  $C \in \{0.5, 1, 2, 10\}$  and  $\nu \in \{\frac{1}{10}, \frac{1}{8}, \frac{1}{6}, \frac{1}{5}, \frac{1}{4}, \frac{1}{3}, \frac{1}{2}, 1\}$ .

At the same time, we also compute the estimator (2) introduced by [Ferraty and Vieu(2006)].

Following [Rachdi and Vieu(2007)], we introduce an automatic selection of the bandwidth, with a cross validation procedure. We use this procedure for the estimator of [Ferraty and Vieu(2006)]. For our recursive estimator, we denote  $h_i = h_i(C, \nu)$  with  $C \in \{0.5, 1, 2, 10\}$  and  $\nu \in \{\frac{1}{10}, \frac{1}{8}, \frac{1}{6}, \frac{1}{5}, \frac{1}{4}, \frac{1}{3}, \frac{1}{2}, 1\}$ , and we consider 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 corresponding to the initial sample without the  $i^{\text{th}}$  observation  $(\mathcal{X}_i, Y_i)$ , for  $i = 1, \dots, n$ . Then we select the values  $C_{CV}$  and  $\nu_{CV}$  of  $C$  and  $\nu$  that minimize  $CV(C, \nu)$ , and our estimator is  $r_n^{[\ell]}$  using these selected values of  $C$  and

$\nu$ .

Table 1 presents the mean and standard deviations of the prediction error over 500 repeated simulations, for the optimal values of  $C$  and  $\nu$  with respect to the  $CV$  criterion (these optimal values are  $C_{CV} = 1$  and  $\nu_{CV} = 1/10$  for our estimator). More precisely, denoting  $\widehat{Y}^{[j]} = r_n^{[\ell],[j]}(\chi^{[j]})$  the predicted value at the  $j^{\text{th}}$  iteration of the simulation ( $j = 1, \dots, 500$ ) for a new curve  $\chi^{[j]}$ , we give the mean ( $MSPE$ ) and the standard deviations of the quantities  $(\widehat{Y}^{[j]} - Y^{[j]})^2$ . The errors are computed for our estimator (label (1) in the table) and the estimator from [Ferraty and Vieu(2006)] (label (2) in the table), both adapted with [Rachdi and Vieu(2007)] procedure. We can see on these results that the estimator from [Ferraty and Vieu(2006)] is a little better than our estimator for the  $MSPE$  criterion. As we will see later (see subsection 3.4), the advantage of our estimator is from the point of view of computational time. We also look at the behaviour of the prediction errors when the sample size increases: we took  $n = 100$ ,  $n = 200$  and  $n = 500$ : as expected, the errors decrease when the sample size increases.

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

Table 1: Mean and standard deviation of the square prediction error, computed on 500 repeated simulations, for different values of  $n$ , with the optimal values of bandwidth given from  $C_{CV}$  and  $\nu_{CV}$ .

### 3.2 Choice of the semi-norm

In this simulation, the parameter  $\ell$  is fixed equal to 0 and we choose to take a bandwidth  $h_i = \max_{i=1, \dots, n} \|\mathcal{X}_i - \chi\| i^{-1/10}$ . The aim is now to compare the influence of the choice of the semi-norm, considering the following ones:

- the semi-norm [PCA] based on the principal components analysis of the curves, keeping  $q = 3$  principal components, more precisely

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

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

- the semi-norm [*FOU*] based on a decomposition of the curves in a Fourier basis, with  $b = 8$  basis functions, more precisely

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

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

- the semi-norm [*DERIV*] based on a comparison of cubic splines approximations of the second derivatives of the curves, (with a number of interior knots  $k = 8$  for the cubic splines), more precisely

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

where  $\tilde{\mathcal{X}}_i$  and  $\tilde{\chi}$  are the spline approximations of the curves  $\mathcal{X}_i$  and  $\chi$ ,

- the semi-norm [*PLS*] where the data are projected on a space determined by a PLS regression on the curves, taking  $K = 5$  PLS basis functions, more precisely

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

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

The results are given in Table 2. For these simulated data, the semi-norms [*PCA*] and [*PLS*] show better results. However, as pointed out in [Ferraty and Vieu(2006)], there is no universal norm that would overcome the others. The choice of the semi-norm depends on the data to be treated.

norm	[ <i>PCA</i> ]	[ <i>FOU</i> ]	[ <i>DERIV</i> ]	[ <i>PLS</i> ]
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 repeated simulations, for different choices of norms.

### 3.3 Choice of the parameter $\ell$

In this simulation, we choose to take  $h_i = \max_{i=1,\dots,n} \|\mathcal{X}_i - \chi\| i^{-1/10}$  and the semi-norm based on the principal components analysis of the curves, keeping 3 principal components. The parameter  $\ell$  is varying into  $\{0, \frac{1}{4}, \frac{1}{2}, \frac{3}{4}, 1\}$ . The results are given in Table 3. We can see that the values of the *MSPE*

criterion are really close, so this parameter does not seem to have an important influence on the quality of the prediction, even if we observe as in the multivariate setting the decreasing of the mean square error according  $\ell$ .

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

Table 3: Mean and standard deviation of the square prediction error, computed on 500 repeated simulations, for different values of  $\ell$ .

### 3.4 Computational time

In this subsection, we highlight an important advantage of the recursive estimator compared to the initial one, from [Ferraty and Vieu(2006)]. This concerns the gain of computational time in the prediction of the response, when new values of the explanatory variable are sequentially added to the database. Indeed, when a new observation  $(\mathcal{X}_{n+1}, Y_{n+1})$  appears, the computation of the recursive estimator  $r_{n+1}^{[\ell]}$  just asks another iteration of the algorithm, by using its value computed with the sequence  $(\mathcal{X}_i, Y_i)_{i=1, \dots, n}$ , while the initial estimator from [Ferraty and Vieu(2006)] must be recomputed on the whole sample  $(\mathcal{X}_i, Y_i)_{i=1, \dots, n+1}$ . This explains the computation time difference between both estimators in this kind of situations, as illustrated in the following. From an initial sample  $(\mathcal{X}_i, Y_i)_{i=1, \dots, n}$  with size  $n = 100$ , we consider  $N$  additional observations, for different values of  $N$ . We compare the cumulated computational times to obtain the recursive and the non recursive estimators, when adding these  $N$  new observations. The characteristics of the computer on which the computations have been done are: CPU: Duo E4700 2.60 GHz, HD: 149 Go, Memory: 3.23 Go. The simulation is done in the following conditions: the curves  $\mathcal{X}_1, \dots, \mathcal{X}_n$ , as well as the new observations  $\mathcal{X}_{n+1}, \dots, \mathcal{X}_{n+N}$ , are standard Brownian motions on  $[0, 1]$ , with  $n = 100$  and  $N \in \{1, 50, 100, 200, 500\}$ . The semi-norm, the sequence of bandwidths and the parameter  $\ell$  are chosen as each particular previous case.

The computational times are collected in Table 4. Here, our estimator shows its clear advantage in terms of computational time compared to the estimator from [Ferraty and Vieu(2006)].

### 3.5 A real dataset example

In this subsection, we use our estimator in a situation of a real dataset. Functional data are particularly adapted when one wants to study a time

$N$	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: Cumulated computational times in seconds for the recursive and [Ferraty and Vieu(2006)] estimators when adding  $N$  new observations, for different values of  $N$ .

series. We illustrate this fact with El Niño time series<sup>1</sup> which gives the monthly sea surface temperature from January, 1982 up to December, 2011 (360 months) and plotted on Figure 1. From this time series, we extract the 30 yearly curves  $\mathcal{X}_1, \dots, \mathcal{X}_{30}$  from 1982 to 2011, discretized into  $p = 12$  points. These yearly curves are plotted on Figure 2. The observation of the variable of interest at a certain month  $j$  of the year  $i$  is the value of the sea temperature  $\mathcal{X}_{i+1}$  for the month  $j$ , in other words, for  $j = 1, \dots, 12$  and for  $i = 1, \dots, 29$ ,  $Y_i^{[j]} = \mathcal{X}_{i+1}(j)$ .

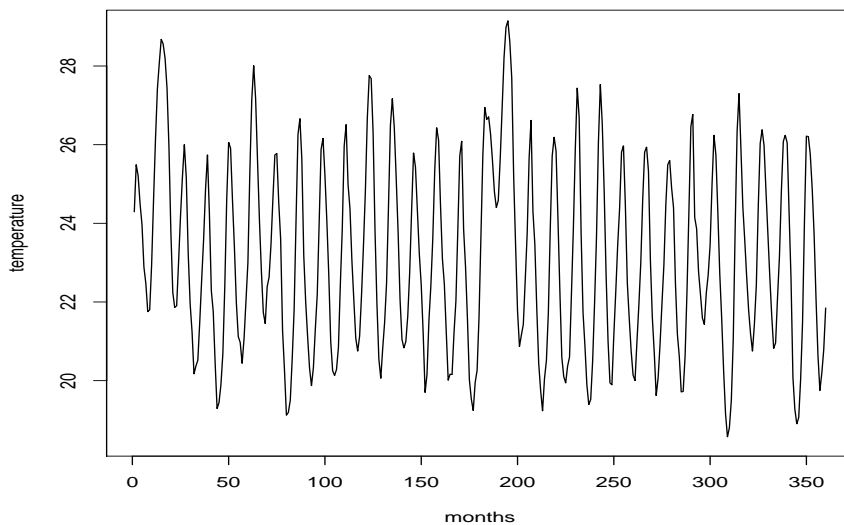


Figure 1: El Niño monthly temperature time series from January, 1982 up to December, 2011.

We predict the values of  $Y_{29}^{[1]}, \dots, Y_{29}^{[12]}$  (in other words, the values of the curve  $\mathcal{X}_{30}$ ). The recursive estimator and the estimator from [Ferraty and Vieu(2006)] are computed by choosing the semi-norm, the sequence of bandwidths and the parameter  $\ell$  as each particular previous case.

<sup>1</sup>available online at <http://www.math.univ-toulouse.fr/staph/npfda/>

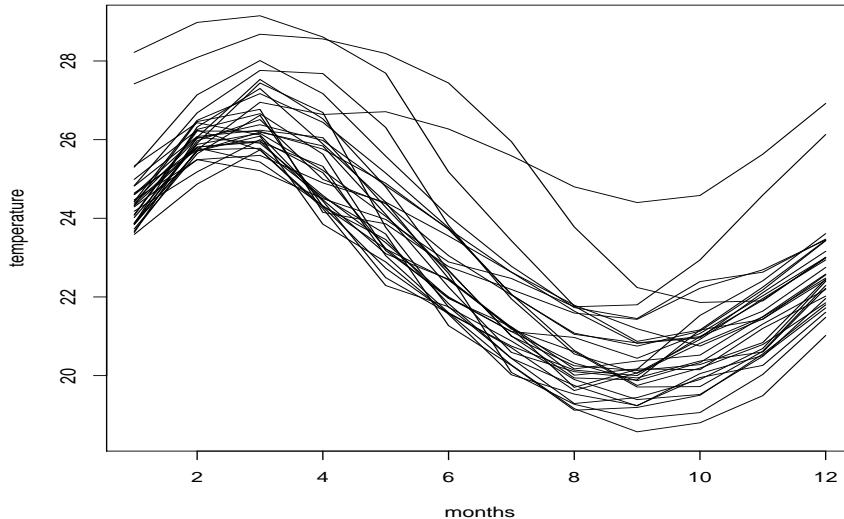


Figure 2: El Niño yearly curves temperatures from 1982 up to 2011.

We analyze the results by computing the mean square prediction error over the year 2011, given by

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

where  $\widehat{Y}_{29}^{[j]}$  is computed either with the recursive estimator (result: 0.5719) or the estimator from [Ferraty and Vieu(2006)] (result: 0.2823). The corresponding true curve and predicted curves over the year 2011 are plotted on Figure 3. The estimator from [Ferraty and Vieu(2006)] shows again its advantage in terms of prediction, while our estimator behaves quite well and has the advantage of computational time as highlighted in previous subsection. Here, for the prediction of twelve values (the final year), the computational time (in seconds) for our estimator is 0.128 while the computational time for the estimator from [Ferraty and Vieu(2006)] is 0.487.

## 4 Proofs

Throughout the proofs, we denote by  $\gamma_i$  a sequence of real numbers going to zero as  $i$  tends to  $\infty$ . The kernel estimate  $r_n^{[\ell]}$  can be written as

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

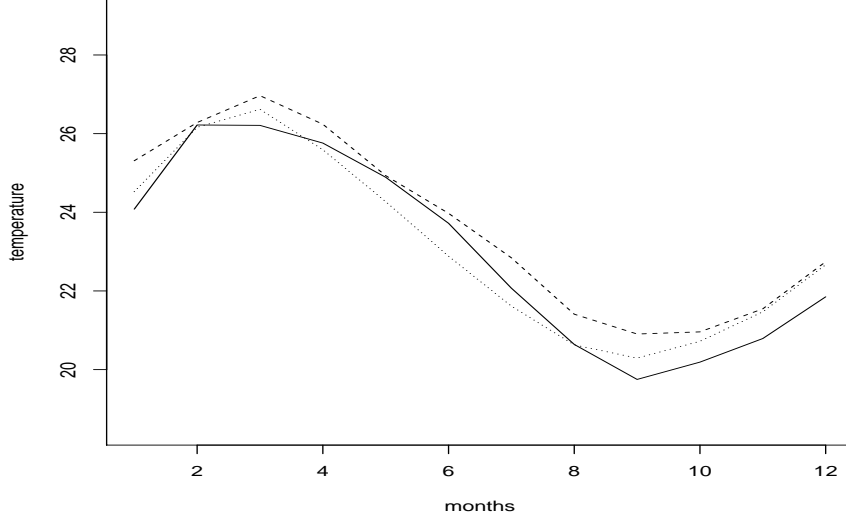


Figure 3: El Niño true and predicted temperature curves for the year 2011. The solid line is the true curve. The dashed line is the predicted curve with the recursive estimator. The dotted line is the predicted curve with the estimator from Ferraty and Vieu [Ferraty and Vieu(2006)].

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, we use the following decomposition

$$\begin{aligned} E \left[ r_n^{[\ell]}(\chi) \right] &= \frac{E \left[ \varphi_n^{[\ell]}(\chi) \right]}{E \left[ f_n^{[\ell]}(\chi) \right]} - \frac{E \left\{ \varphi_n^{[\ell]}(\chi) \left[ f_n^{[\ell]}(\chi) - E f_n^{[\ell]}(\chi) \right] \right\}}{\left\{ E \left[ f_n^{[\ell]}(\chi) \right] \right\}^2} \\ &\quad + \frac{E \left\{ r_n^{[\ell]}(\chi) \left[ f_n^{[\ell]}(\chi) - E f_n^{[\ell]}(\chi) \right]^2 \right\}}{\left\{ E \left[ f_n^{[\ell]}(\chi) \right] \right\}^2}. \end{aligned}$$

The first part of Theorem 1 is then a direct consequence of the following lemmas.

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

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

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

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

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

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

To study the variance term in Theorem 1, we use the following decomposition which can be found in [Collomb(1976)].

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

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

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

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

Now let us prove Lemmas 1-4.

#### 4.1.1 Proof of Lemma 1

Observe that

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



Noting that

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

a Taylor's expansion of  $\varphi$  around 0 ensures that

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

From the proof of Lemma 2 in [Ferraty et al.(2007)], it follows from **H2** and Fubini's Theorem that

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

and

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

Combining (6) and (7), we have

$$\begin{aligned}
\frac{\mathbb{E} \left[ \varphi_n^{[\ell]}(\chi) \right]}{\mathbb{E} \left[ f_n^{[\ell]}(\chi) \right]} - r(\chi) &= \frac{\sum_{i=1}^n h_i F(h_i)^{1-\ell} \left\{ \varphi'(0) \left[ K(1) - \int_0^1 (sK(s))' \tau_{h_i}(s) ds \right] + \gamma_i \right\}}{\sum_{i=1}^n F(h_i)^{1-\ell} \left[ K(1) - \int_0^1 K'(s) \tau_{h_i}(s) ds \right]} \\
&:= \frac{D_1}{D_2}.
\end{aligned}$$

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.  $\square$

#### 4.1.2 Proof of Lemma 3

From (7), we can write

$$\begin{aligned}
\mathbb{E} \left[ f_n^{[\ell]}(\chi) \right] &= \frac{1}{\sum_{i=1}^n F(h_i)^{1-\ell}} \sum_{i=1}^n \frac{1}{F(h_i)^\ell} \mathbb{E} \left[ K \left( \frac{\|\chi - \mathcal{X}_i\|}{h_i} \right) \right] \\
&= \frac{\sum_{i=1}^n \frac{F(h_i)^{1-\ell}}{nF(h_n)^{1-\ell}} \left[ K(1) - \int_0^1 K'(s) \tau_{h_i}(s) ds \right]}{B_{n,1-\ell}} = M_1 [1 + o(1)],
\end{aligned}$$

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

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

where

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

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

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

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

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

that proves Lemma 3.  $\square$

#### 4.1.3 Proof of Lemma 4

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

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

The relation (7) and assumption **H3** ensure that

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

then we get

$$\text{Var} \left[ K \left( \frac{\|\chi - \mathcal{X}\|}{h_i} \right) \right] = M_2 F(h_i) [1 + \gamma_i].$$

We obtain that

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

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

$$\text{Var} \left[ \varphi_n^{[\ell]}(\chi) \right] = \frac{1}{\left( \sum_{i=1}^n F(h_i)^{1-\ell} \right)^2} \sum_{i=1}^n F(h_i)^{-2\ell} \text{Var} \left[ Y_i K \left( \frac{\|\chi - \mathcal{X}_i\|}{h_i} \right) \right].$$

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

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

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

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

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

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

which proves the second assertion of Lemma 4. For the covariance term, this can be written as

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

Notice that from (6) and (8), we have

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

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

$$\mathbb{E} \left[ Y_i K^2 \left( \frac{\|\mathcal{X} - \mathcal{X}_i\|}{h_i} \right) \right] = M_2 F(h_i) [r(\mathcal{X}) + \gamma_i],$$

it follows that

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

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.  $\square$

## 4.2 Proof of Theorem 2

We have the following decomposition

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

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

$$\tilde{\varphi}_n^{[\ell]}(\mathcal{X}) = \frac{1}{\sum_{i=1}^n F(h_i)^{1-\ell}} \sum_{i=1}^n \frac{Y_i}{F(h_i)^\ell} \mathbf{1}_{\{|Y_i| \leq b_n\}} K \left( \frac{\|\mathcal{X} - \mathcal{X}_i\|}{h_i} \right), \quad (11)$$

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

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

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  $\varepsilon > 0$ ,

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

and by Borel-Cantelli's lemma we get

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

For the principal term in (10), we can write

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

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

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

□

**Lemma 5** *Under assumptions H1 – H6, we have*

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

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

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

#### 4.2.1 Proof of Lemma 5

Let us set

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

and define

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

Observe that

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

We can write

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

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

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

and from Toeplitz's Lemma, by virtue of **H3** and **H4**, we get

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

For the second term of the decomposition of  $A_1$ , from (9) we have

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

The continuity of  $\sigma_\varepsilon^2$  (**H1**) with again Toeplitz's lemma ensure that

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

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

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

where

$$Q_n = (\max \{ \mathbb{E}(Y^4), 4|r(\chi)|\mathbb{E}|Y|^3, 4r^2(\chi)\mathbb{E}(Y^2) \}) P(|Y| > b_n)^{\frac{1}{2}}.$$

We deduce from **H4** and **H5**, again with the choice  $b_n = (\delta \ln n)^{1/\mu}$ , that

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

Next, for the last term  $A_3$ , we have

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

It follows from **H6** that

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

We deduce from (15), (16), (17) and (18) that

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

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

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

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

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

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

We deduce from (19) that

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

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  $\xi$  such that

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

It follows that

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

and then we have

$$\frac{S_n}{\sqrt{2V_n \ln \ln V_n}} \rightarrow 1 \quad a.s., \text{ as } n \rightarrow \infty, \quad (20)$$

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

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

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).  $\square$

#### 4.2.2 Proof of Lemma 6

We have

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

As in the proof of Lemma 1, we can write

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

and then,

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

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

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

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

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



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.  $\square$

### 4.3 Proof of Theorem 3

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

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

and

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

where  $\tilde{\varphi}_n^{[\ell]}$  is defined in (11) and  $c$  is such that  $\lim_{n \rightarrow \infty} \sqrt{nF(h_n)}h_n = c$ , since  $f_n^{[\ell]}(\chi) \xrightarrow{P} M_1$ . This later follows from the first parts of Lemmas 3 and 4.

For (24), following the same lines of proof of (12) with substituting  $\left[ \frac{\ln \ln n}{nF(h_n)} \right]^{1/2}$  by  $\frac{1}{\sqrt{nF(h_n)}}$  gives the desired results. 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  $\lim_{n \rightarrow \infty} h_n \sqrt{nF(h_n)} = c$ , then*

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

#### 4.3.1 Proof of Lemma 7

Setting

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

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

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

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

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

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

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

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

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

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

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

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) with substituting the exponent 2 by  $p$  in all the expressions, we have

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

so that 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 from (18), we have

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

so that

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

which concludes the proof of Lemma 7.  $\square$

### 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

$$\sqrt{nF(h_n)}A = \sqrt{nF(h_n)}h_n \frac{\alpha^{[\ell]}}{\beta_{[1-\ell]}} \varphi'(0)M_0 [1 + o(1)]. \quad (28)$$

On the other hand from (23), we get

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

and Lemma 8 follows from the combination of (28) and (29).  $\square$

## References

- [Ahmad and Lin(1976)] Ahmad, I. and Lin, P.E. (1976). Nonparametric sequential estimation of a multiple regression function. *Bull. Math. Statist.*, **17**, 63-75.
- [Amiri(2012)] Amiri, A. (2012). Recursive regression estimators with application to nonparametric prediction. *J. Nonparametr. Stat.* **24** (1), 169-186.
- [ Besse et al.(1997)] Besse, P., Cardot, H. and Ferraty, F. (1997). Simultaneous nonparametric regressions of unbalanced longitudinal data. *Comp. Statist. Data Anal.*, **24**, 255-270.
- [Bosq and Cheze-Payaud(1999).] 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 et al.(2003)] Cardot, H., Ferraty, F. and Sarda, P. (2003). Splines estimators for the functional linear model. *Statistica Sinica*, **13**, 571-591.
- [Collomb(1976)] Collomb, G. (1976). Estimation nonparamétrique de la régression par la méthode du noyau. *Unpublished PhD Thesis*, Université Paul Sabatier, Toulouse, France.
- [Crambes et al.(2009)] Crambes, C., Kneip, A. and Sarda, P. (2009). Smoothing splines estimators for functional linear regression, *Ann. Statist.*, **37**, 35-72.
- [Devroye and Wagner(1980)] Devroye, L. and Wagner, T.J. (1980). Distribution-free consistency results in nonparametric discrimination and regression function estimation. *Ann. statist.*, **8**, 231-239.
- [Ferraty et al.(2007)] 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 and Romain(2010)] Ferraty, F. and Romain, Y. (2010). Handbook on functional data analysis and related fields. *Oxford University Press*, Oxford.
- [Ferraty and Vieu(2006)] Ferraty, F. and Vieu, P. (2006). Nonparametric Modelling for Functional Data. Methods, Theory, Applications and Implementations. *Springer-Verlag*, London.
- [Frank and Friedman (1993)] Frank, I.E. and Friedman, J.H. (1993). A statistical view of some chemometrics regression tools. *Technometrics*, **35**, 109–135.

- [Jain et al.(1975)] Jain, C., Jogdeo, K. and Stout, W. (1975). Upper and lower functions for martingales and mixing processes, *Ann. Probab.* **3** (1), 119–145.
- [Ling and Wu (2012)] Ling, N. and Wu, Y. (2012). Consistency of modified kernel regression estimation for functional data, *Statistics* **46** (2), 149–158.
- [Masry(1986)] Masry, E. (1986). Recursive probability density estimation for weakly dependent stationary processes, *IEEE Trans. Inform. Theory* **32** (2), 254–267.
- [Masry(2005)] Masry, E. (2005). Nonparametric regression estimation for dependent functional data: Asymptotic normality, *Stochastic Process. Appl.* **115** 155–177.
- [Rachdi and Vieu(2007)] Rachdi, M. and Vieu, P. (2007). Nonparametric regression for functional data: automatic smoothing parameter selection. *J. statist. Plann. Inference*, **137**, 2784-2801.
- [Ramsay and Dalzell(1991)] 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 and Silverman(2002)] Ramsay, J.O. and Silverman, B.W. (2002). Applied functional data analysis. *Springer-Verlag*, New-York.
- [Ramsay and Silverman(2005)] Ramsay, J.O. and Silverman, B.W. (2005). *Functional data analysis (2<sup>nd</sup> Ed.)*. Springer-Verlag, New-York.
- [Roussas(1992)] Roussas, G. G.(1992). Exact rates of almost sure convergence of a recursive kernel estimate of a probability density function: Application to regression and hazard rate estimation, *J. Nonparametr. Stat.*, **1** (3), 171– 195.