

Real-time estimation for functional stochastic regression models

Amir Aboubacar & Mohamed Chaouch

To cite this article: Amir Aboubacar & Mohamed Chaouch (2020): Real-time estimation for functional stochastic regression models, Journal of Statistical Computation and Simulation, DOI: [10.1080/00949655.2020.1746786](https://doi.org/10.1080/00949655.2020.1746786)

To link to this article: <https://doi.org/10.1080/00949655.2020.1746786>



Published online: 02 Apr 2020.



Submit your article to this journal [↗](#)



Article views: 16



View related articles [↗](#)



View Crossmark data [↗](#)



Real-time estimation for functional stochastic regression models

Amir Aboubacar^a and Mohamed Chaouch^{b*}

^aUMR 9221-LEM-Lille Economie Management, Université de Lille, Lille, France; ^bDepartment of Mathematics, Statistics and Physics, Qatar University, Doha, Qatar

ABSTRACT

In this paper, a heteroscedastic functional regression model with martingale difference errors is considered. We are interested in real-time estimation of the regression as well as the conditional variance operators when the response is a real-valued random variable and the covariate belongs to an infinite-dimensional space. A Robbins–Monro-type estimator of the conditional variance is introduced when a sample is collected from an underlying stationary and ergodic process. First, a local uniform L^q -consistency (for $q \geq 2$) rate of the recursive estimator of the regression operator is established. Then, a pointwise mean-square consistency rate of the conditional variance is given when the regression function is supposed to be known and when it is estimated recursively. Simulation studies are conducted to assess the proposed estimator's performance, in terms of reducing the computational time without affecting significantly the accuracy, compared to its natural competitor. An application to real environmental data is also carried out to illustrate the real-time on day ahead prediction of the maximum ozone concentration in Mexico city as well as its volatility.

ARTICLE HISTORY

Received 8 October 2018
Accepted 19 March 2020

KEYWORDS

Conditional variance; ergodic processes; functional data; Robbins–Monro algorithms; nonparametric estimation; kernel methods



2010 MATHEMATICS SUBJECT CLASSIFICATIONS

62G05; 62L20; 62G20; 62G08

1. Introduction

Nonparametric estimation, based on independent and identically distributed (i.i.d.) or weakly dependent stationary processes, of the regression function has been widely studied for real or vectorial-valued regressors. See, for example, [1–4] and the references therein. Recently, an increasing interest is given to the regression function estimation when the response is scalar and the predictor takes values in some abstract functional space (e.g. Banach or Hilbert space). The book by Ramsay and Silverman [5] introduced parametric models for functional data and studied their properties, whereas the monograph by Ferraty and Vieu [6] discussed limit theorems for the nonparametric estimators. Diverse case studies in several areas including criminology, economics and neurophysiology can be found in [5,6] and recently [7] for a survey of recent advances on functional data analysis.

In this paper, we consider (X, Y) a pair of random variables defined on the probability space $(\Omega, \mathcal{A}, \mathbb{P})$, where Y is a real-valued random variable and X a functional random

CONTACT Mohamed Chaouch  mchaouch@qu.edu.qa; mo.chaouch@gmail.com  Department of Mathematics, Statistics and Physics, Qatar University, Doha, Qatar

variable that takes values in some abstract functional space \mathcal{E} endowed with a semi-metric $d(\cdot, \cdot)$. Let $(X_i, Y_i)_{i=1, \dots, n}$ be a *stationary and ergodic process* distributed as the random pair (X, Y) . We suppose that the data is generated following the *nonlinear heteroscedastic functional regression* model defined below:

$$Y_i = r(X_i) + \sigma(X_i)\varepsilon_i, \quad i = 1, \dots, n, \quad (1)$$

where $r(X) := \mathbb{E}(Y|X)$ and $\sigma^2(X) := \text{Var}(Y|X)$ are, respectively, the regression and the conditional variance operators of Y given X . Here, we consider that the sequence of random variable $\{\varepsilon_i\}$ form a *martingale difference sequence* such that

$$\mathbb{E}(\varepsilon_i | \mathcal{G}_{i-1}) = 0 \quad \text{a.s.} \quad \text{and} \quad \text{Var}(\varepsilon_i | \mathcal{G}_{i-1}) = 1 \quad \text{a.s.}, \quad (2)$$

where \mathcal{G}_{i-1} is the σ -field generated by $\{(X_1, Y_1), \dots, (X_{i-1}, Y_{i-1}), X_i\}$.

A Kernel-type estimators of the regression operator $r(x) = \mathbb{E}(Y|X = x)$ and the conditional variance operator $\sigma^2(x) = \text{Var}(Y|X = x)$, for $x \in \mathcal{E}$, are defined, respectively, as follows (see for instance [6,8]):

$$\hat{r}_n(x) = \frac{\sum_{i=1}^n Y_i K(d(x, X_i)/h_n)}{\sum_{i=1}^n K(d(x, X_i)/h_n)}, \quad (3)$$

and

$$\hat{\sigma}_n^2(x) = \frac{\sum_{i=1}^n (Y_i - \hat{r}_n(X_i))^2 K(d(x, X_i)/h_n)}{\sum_{i=1}^n K(d(x, X_i)/h_n)}, \quad (4)$$

where $K(\cdot)$ is a kernel function, $(h_n)_n$ a sequence of positive real numbers which decreases to zero as n tends to infinity. The kernels, bandwidths and the semi-metrics may be different between (3) and (4). Many authors investigated several asymptotic aspects of the regression estimator (3). See for instance the monographs by Ferraty and Vieu [6] and Ferraty and Romain [7]. Among the asymptotic properties that have been studied, note the mean-squared convergence, with rates, as well as the asymptotic normality obtained by Ferraty et al. [9]. While a mixing process is considered, rates of almost sure uniform consistency, over a compact set, of (3) were established in [10]. Masry [11] obtained the mean-squared convergence and the asymptotic normality under the same framework.

All the asymptotic results mentioned above concerning the regression were established under mixing assumptions. However, it is well known that the mixing conditions may be viewed as an asymptotic independence (see [12]), and they are very restrictive assumptions since some linear processes (such as autoregressive processes) do not even satisfy them (see [13–16] for more details and examples). Consequently, it is relevant to find similar asymptotic results concerning the nonparametric regression estimator under a more general and less restrictive dependence assumptions. Recently, Laïb and Louani [17,18] studied

the conditional bias of the regression estimator (3) and established the almost sure uniform consistency rate as well as the asymptotic normality for stationary and ergodic processes. For further theoretical and practical motivations to consider ergodic data, reader is advised to see [12,17,19–22], among others.

Unlike the regression function, less attention was given to the nonparametric conditional variance estimator, given by (4), when the regressor is functional. Estimation of the conditional variance is an important statistical task since it measures the risk associated to the regression estimation. The conditional variance also appears in the form of the regression confidence interval and therefore its estimation as well as the study of its asymptotic properties is clearly relevant. In finite-dimensional case ($X \in \mathbb{R}^d, d \geq 1$) several authors were interested to the question (see for instance [23–26]). As far as we know, in the infinite-dimensional case (X is functional), Hu [8] studied the convergence in probability (without rate) under a strong mixing assumption. More recently, Chaouch [27] investigated the uniform consistency rate as well as the asymptotic normality of the conditional variance when data are generated from an heteroscedastic functional regression model with martingale difference errors.

Motivated by Chaouch [27] and Stadtmüller and Tsybakov [28], our purpose in this paper is to introduce a recursive estimator of the regression and the variance operators based on stochastic approximation algorithm. Nowadays, with the progress of measurement apparatus and the development of automatic sensors, we can get access to large samples of observations taking values in high dimensional spaces. Therefore, within this new framework of ‘massive data’ classical estimation approaches, such as maximum likelihood or least squares, are time consuming. To deal with this constraint, recursive techniques allow to perform the conditional variance estimation without any full data storage requirement. When the data arrive sequently the value of each successive estimator is obtained from its value at the previous step by a simple adjustment that takes into account only the recently received data.

The use of stochastic algorithms has been introduced by Robbins and Monro [29] and adopted by Révész [30] for the nonparametric regression estimation. Several authors, such as [31–33] studied the asymptotic properties of the recursive regression estimator in the finite-dimensional framework. Recently, Amiri et al. [34] and Amiri and Thiam [35] defined this estimator (for a particular choice of the step size) when the regressor takes values in an infinite-dimensional space with i.i.d and α -mixing data respectively. In this paper, we generalize the [35] estimator to a wider class of dependent processes (satisfying an ergodic assumption) and for any step size. In contrast to the regression operator, and to the best of our knowledge, a recursive estimator of the conditional variance has not been introduced in the statistical literature when the covariate is infinite dimensional. In the finite-dimensional case ($X \in \mathbb{R}^d$) and when the data is i.i.d., [28] studied the Robins–Monro-type estimator of the regression function as well as the conditional variance. Our results may be seen as a generalization of the results obtained in the last reference.

This paper is organized as follows. In Section 2, Robbins–Monro-type estimators of the regression and the conditional variance are introduced. Section 3, contains the main results which are the uniform, in a neighbourhood of $x \in \mathcal{E}$, L^q -convergence ($q \geq 2$) with rate of the nonparametric regression estimator and the pointwise mean-squared convergence rate of the conditional variance estimator. Section 4, is devoted to a simulation study which

show that the recursive nonparametric estimator of the conditional variance remains as accurate as the nonrecursive one while its computation is less time consuming when data is received online. Finally, the proofs of the theoretical results are stated in Section 6.

2. Real-time estimation of the conditional variance

In order to argue the definitions given in (5)–(6), we recall the stochastic approximation algorithm proposed by Robbins and Monro [29]. Let $M(\cdot)$ be a given function such that the equation $M(z) = 0$ has a unique zero, say $z = \vartheta$. The purpose is to estimate ϑ . Robbins and Monro [29] achieved this by maintaining a running estimate $\widehat{\vartheta}_n$ at step ‘ $n + 1$ ’ given by $\widehat{\vartheta}_{n+1} = \widehat{\vartheta}_n - a_n \eta_n$, where a_n is a step-size tending to zero and η_n is an appropriate estimator of the functional $M(\cdot)$, which depends on $\widehat{\vartheta}_n$ and the information received at ‘step $n + 1$ ’. Keeping the above definition in mind, we are interested in the asymptotic properties of the regression and conditional variance estimators defined as follows. For $x \in \mathcal{E}$, an estimator of the regression operator $r(x)$, using the sample $(X_1, Y_1), \dots, (X_n, Y_n)$, is defined as follows:

$$r_n(x) = r_{n-1}(x) + \theta_n [Y_n - r_{n-1}(x)] \quad (5)$$

where $(\theta_n)_n$ is a sequence of positive numbers tending to zero as n goes to infinity. Given (5), a sequential estimator of $\sigma^2(x)$, can be defined through a stochastic approximation algorithm as follows:

$$\sigma_n^2(x) = \sigma_{n-1}^2(x) + \tau_n \{ [Y_n - r_{n-1}(X_n)]^2 - \sigma_{n-1}^2(x) \} \quad (6)$$

where $(\tau_n)_n$ is a sequence of positive numbers tending to zero as n goes to infinity, for any $x \in \mathcal{E}$. In the representations of the estimators in (5) and (6), Y_n and $(Y_n - r_{n-1}(X_n))^2$ can be seen as the estimators of $r(x)$ and $\sigma^2(x)$ based on the only data recorded at ‘step n ’. Therefore, our estimators follow from the stochastic approximation algorithms, by choosing $M(z) = z - r(x)$ for the regression and $M(z) = z - \sigma^2(x)$ for the conditional variance, with respective step sizes θ_n and τ_n . In practice, the initial values $r_0(x)$ and $\sigma_0^2(x)$ may be chosen arbitrarily or at random. In the statistical literature, several choices of θ_n and τ_n are possible. A popular choice of θ_n and τ_n is $\theta_n = \theta_0 n^{-1} \Delta_n F_x^{-1}(h_n)$ and $\tau_n = \tau_0 n^{-1} \Delta_n F_x^{-1}(h_n)$, where

$$\Delta_n := \Delta_n(x) := K [d(x, X_n)/h_n] \quad \text{and} \quad F_x(h) := \mathbb{P} [d(x, X) \leq h], \quad (7)$$

θ_0 and τ_0 are positive constants. In this case, the algorithms in (5) and (6) are very similar to the valuable stochastic approximation approach introduced in [30]. However, it is easy to see that the recursive form of the regression and the conditional variance estimators in Equations (5) and (6) are not fully computable with the choices above because of the presence of $F_x(h_n)$ which is unknown. As discussed in simulations below, this issue can be solved by making a suitable choice of the θ_n and τ_n (as functions of $F_x(h_n)$) or replacing $F_x(\cdot)$ with its empirical counterpart. As another example, if

$$\theta_n = \tau_n = \frac{\Delta_n(x) F_x(h_n)^{-\ell}}{\sum_{i=1}^n \Delta_i(x) F_x(h_i)^{-\ell}}, \quad \ell \in [0, 1], \quad (8)$$

then from (5), one gets

$$r_n(x) := \frac{\varphi_{n,1}^{[\ell]}(x)}{\varphi_{n,0}^{[\ell]}(x)}, \quad \text{where } \varphi_{n,s}^{[\ell]}(x) = \sum_{i=1}^n \frac{Y_i^s \Delta_i(x)}{F_x(h_i)^\ell}, \quad s \in \{0, 1\}. \tag{9}$$

The estimators (9) was proposed by Amiri et al. [34] as a recursive counterpart of the functional regression estimate (3). The asymptotic properties, such as the mean square error and the almost sure convergence, of $r_n(x)$ was derived by Amiri et al. [34]. Moreover, the same choice of θ_n and τ_n given in (8), and an easy computation allow to represent the estimator (6) as follows:

$$\sigma_n^2(x) = \frac{\sum_{i=1}^n \left(Y_i - r_{i-1}^{[\ell]}(X_i) \right)^2 F_x(h_i)^{-\ell} K(d(x, X_i)/h_i)}{\sum_{i=1}^n F_x(h_i)^{-\ell} K(d(x, X_i)/h_i)}. \tag{10}$$

Thus, in case $\ell = 0$, (10) is a sequential version of (4) which is obtained by considering sequences of bandwidth parameters $(h_i)_i$ and regression estimators $(r_{i-1}(\cdot))_i$ rather than a single bandwidth and single regression estimator value.

3. Assumptions and asymptotic results

In order to state our results, let us introduce some additional notations. Let \mathcal{F}_{n-1} be the σ -field generated by $\{(X_1, Y_1), \dots, (X_{n-1}, Y_{n-1})\}$. In what follows, we consider a fixed curve $x \in \mathcal{E}$. We denote by $\mathcal{B}(x, h)$ a ball centred at $x \in \mathcal{E}$ with radius h . Working on the probability space $(\Omega, \mathcal{A}, \mathbb{P})$, let $F_x^{\mathcal{F}_{i-1}}(h) = \mathbb{P}(X_i \in \mathcal{B}(x, h) \mid \mathcal{F}_{i-1})$ the conditional distribution function, given the σ -field \mathcal{F}_{i-1} , of the real random variable $d(x, X_i)$. Denote by $o_{a.s.}(h)$ (resp. $\mathcal{O}_{a.s.}(h)$) a real random function ω such that $\omega(h)/h$ converges to zero almost surely as $h \rightarrow 0$ (resp. $\omega(h)/h$ is almost surely bounded).

In the sequel, we consider $q \geq 2$ satisfying the assumptions (H0) and (H3). Our first result consists in establishing the uniform, in a neighbourhood of x , L^q -consistency rate of the recursive estimator of the regression operator. Then, a pointwise mean square consistency with rate of the conditional variance operator is given.

- (H0) $\mathbb{E}(|\varepsilon_n|^t \mid \mathcal{G}_{n-1}) = \mathbb{E}(|\varepsilon_n|^t \mid X_n) \leq \mu_t < \infty$, for any $2 \leq t \leq q$.
- (H1) K is nonnegative bounded kernel of class \mathcal{C}^1 over its support $[0, 1]$ and satisfied the condition $K'(t) < 0, \forall t \in [0, 1]$ and $|\int_0^1 (K^j)'(u) du| < \infty$ for $j \geq 1$.
- (H2) There exists a sequence of nonnegative random functionals $(f_{i,1}(\cdot))_{i \geq 1}$ almost surely bounded by a sequence of deterministic quantities $(b_i(\cdot))_i$, such that $\sup_{i,x} b_i(x) < \infty$. Let $(g_{i,x}(\cdot))_{i \geq 1}$ be a sequence of random functions, $f_1(\cdot)$ a deterministic positive functional and a nonnegative real function $\phi(\cdot)$ tending to zero, as its argument tends to 0, such that:
 - (i) $F_x(h) = \phi(h)f_1(x) + o(\phi(h))$ as $h \rightarrow 0$, where $o(\phi(h))$ is uniform in x .
 - (ii) For any $i \in \mathbb{N}$, $F_x^{\mathcal{F}_{i-1}}(h) = \phi(h)f_{i,1}(x) + g_{i,x}(h)$, with $g_{i,x}(h) = o_{a.s.}(\phi(h))$ as $h \rightarrow 0$, and $g_{i,x}(h)/\phi(h)$ is almost surely bounded uniformly in x and $o_{a.s.}(\phi(h))$ is uniform in x .

- (iii) There exists a nondecreasing bounded function τ_0 such that, uniformly in $s \in [0, 1]$, $\phi(hs)/\phi(h) = \tau_0(s) + o(1)$, as $h \downarrow 0$ and, for $j \geq 1$, $\int_0^1 (K^j(t))' \tau_0(t) dt < \infty$.
- (H3) (i) There exist $\alpha > 0$ and $c_1 > 0$ such that $\sup_{z \in \mathcal{B}(x, h_n)} |r(z) - r(x)| \leq c_1 d^\alpha(z, x)$.
- (ii) There exist $\beta > 0$ and $c_2 > 0$ such that $\sup_{z \in \mathcal{B}(x, h_n)} |\sigma^t(z) - \sigma^t(x)| \leq c_2 d^\beta(z, x)$, for any $0 < t \leq q$.
- (iii) The operator $f_1(\cdot)$ is uniformly bounded on $\mathcal{B}(x, h_n)$ and $0 < C_0 \leq \inf_{n \geq 1} \inf_{z \in \mathcal{B}(x, h_n)} f_1(z)$.

Let

$$C^* := M_1 \frac{\sup_{n \geq 1} \sup_{z \in \mathcal{B}(x, h_n)} b_n(z)}{\inf_{n \geq 1} \inf_{z \in \mathcal{B}(x, h_n)} f_1(z)}, \quad (11)$$

where, as in [9], we define: $M_j = K^j(1) - \int_0^1 (sK^j(s))' \tau_0(s) ds$, for $j = 1, 2$.

Comments on assumptions: Condition (H0) is of Markov-type condition, whereas (H1) is classical in nonparametric estimation. Assumptions (H3)(i)–(iii) are smoothness conditions of the regression and the variance functions, respectively. Similar conditions have been assumed by Laïb and Louani [18] and by Stadtmüller and Tsybakov [28]. Hypotheses (H2) involve the ergodic nature of the data and the small ball techniques used in this paper. Laïb and Louani [18] discussed several examples of processes satisfying the condition (H2)(ii) and they gave an explicit form of their conditional, with respect to the σ -field \mathcal{F}_{i-1} , small ball probability.

Theorem 3.1: *Let $r_n(x)$ be the estimator defined in (5) with $\theta_n = \rho_n \Delta_n F_x^{-1}(h_n)$, Δ_n and F_x defined in (7). Suppose the following conditions are satisfied:*

$$(i) h_n \downarrow 0, \quad \rho_n \rightarrow 0, \quad (ii) \rho_n \phi^{-1}(h_n) \rightarrow 0 \text{ as } n \rightarrow \infty \quad \text{and} \quad (iii) \sum_{n=1}^{\infty} \rho_n = \infty. \quad (12)$$

Then, for any $q \geq 2$ such that the assumptions (H0)–(H3) hold true, we have

$$\sup_{z \in \mathcal{B}(x, h_n)} \mathbb{E} |r_n(z) - r(z)|^q \rightarrow 0, \text{ as } n \rightarrow \infty. \quad (13)$$

Moreover, let

$$\phi(h) = \phi_0 h^\nu, \text{ and } \rho_n = \rho_0 n^{-1}, \text{ with } \rho_0 > \frac{\alpha}{C^* (2\alpha + \nu)}, \quad (14)$$

ϕ_0 and ν are positive constants. If $h_n = h_0 n^{-(1/2\alpha + \nu)}$, $h_0 > 0$, then we have

$$\sup_{z \in \mathcal{B}(x, h_n)} \mathbb{E} |r_n(z) - r(z)|^q = \mathcal{O}\left(n^{-(q\alpha/(2\alpha + \nu))}\right), \text{ as } n \rightarrow \infty. \quad (15)$$

Remark 1: The L^2 -convergence of $r_n(x)$ was derived by Amiri et al. [34]. In Theorem 3.1, we generalized their results and established the uniform, with respect to $z \in \mathcal{B}(x, h_n)$, L^q -convergence ($q \geq 2$) of $r_n(z)$ under an ergodic assumption. Notice that, since we are considering a general dependence framework, proofs of our results are based on martingale difference techniques which allow to provide a unified nonparametric time series analysis setting.

In the next theorem, we state the L^2 -convergence of the conditional variance estimator.

Theorem 3.2: *Given a fixed $x \in \mathcal{E}$, let $r_n(x)$ and $\sigma_n^2(x)$ stand for the estimators defined in (5) and (6) with the choices: $\theta_n = \rho_n \Delta_n F_x^{-1}(h_n)$ and $\tau_n = \gamma_n \Delta_n F_x^{-1}(h_n)$, Δ_n and F_x defined in (7). Under assumptions (H0)–(H3) and if the following conditions:*

$$(i) \gamma_n \rightarrow 0, \gamma_n \phi^{-1}(h_n) \rightarrow 0 \quad (ii) \gamma_n \phi^{-2}(h_n) \rightarrow 0 \text{ as } n \rightarrow \infty \quad \text{and} \quad (iii) \sum_{n=1}^{\infty} \gamma_n = \infty, \tag{16}$$

and those given in (12) are also satisfied, then, one gets

$$\mathbb{E} [\sigma_n^2(x) - \sigma^2(x)]^2 \rightarrow 0, \text{ as } n \rightarrow \infty.$$

Furthermore, by considering choices in (14) and

$$\gamma_n = \gamma_0 n^{-1}, \text{ with } \gamma_0 > \frac{\min \left\{ \frac{8\alpha - 3\nu}{2}, 2\alpha \right\}}{2C^* (2\beta + \nu)} \quad \text{and} \quad \nu \leq \frac{8\alpha}{3}, \tag{17}$$

if $h_n = h_0 n^{-1/(2\beta + \nu)}$, $h_0 > 0$, we obtain

$$\mathbb{E} [\sigma_n^2(x) - \sigma^2(x)]^2 = \mathcal{O} \left(n^{-((\min\{8\alpha - 3\nu, 2\beta\}) / (2\beta + \nu))} \right), \text{ as } n \rightarrow \infty. \tag{18}$$

Corollary 3.2.1: *In the case the regression operator $r(\cdot)$ is known, we get, under assumptions and conditions of Theorem 3.1,*

$$\mathbb{E} [\sigma_n^2(x) - \sigma^2(x)]^2 = \mathcal{O} \left(n^{-(2\beta / (2\beta + \nu))} \right), \text{ as } n \rightarrow \infty. \tag{19}$$

Discussion of the results: Observe that our results given in the above theorems are established under conditions (12)–(17) which are classical in recursive estimation based on stochastic algorithms. Similar conditions are used by Stadtmüller and Tsybakov [28] in the finite-dimensional context and for independent data. Moreover, the choice made for the small ball probability in (14) is satisfied for some functional random variables such as the case when X is a fractal process with respect to $d(\cdot)$ (see for instance [9]) or when X is generated by a functional Autoregressive model (see [18] for more details).

Now, we are interested in comparing our result about the variance with the one given in [8] in the i.i.d. case and using the same choice of $h_n = h_0 n^{-1/2\beta + \nu}$, $h_0 > 0$, and $\phi(h_n)$ given in (14). An easy calculation shows that, when $\alpha > 3\nu/4$, the rate of convergence for the recursive estimator of the variance given in (18) is better than the Hu’s nonrecursive one, even if Hu’s convergence results are in probability. Observe that the convergence rate obtained in Corollary 3.2.1 can be reached also when the regression function is unknown whenever $\alpha > (3\nu + 4\beta)/8$.

4. Simulation studies

In this simulation study, we are interested in the evaluation of the accuracy of the recursive conditional variance estimator. A sample $(X_i, Y_i)_{i=1, \dots, 500}$ is simulated where the functional

covariates are generated as:

$$X(t) = A(2 - \cos(\pi t\omega)) + (1 - A) \cos(\pi t\omega), \quad t \in [0, 1],$$

where $\omega \rightsquigarrow \mathcal{N}(0, 1)$ and A is a Bernoulli random variable with parameter equal to $1/2$. In practice we do not observe the curves continuously over the interval $[0, 1]$. A discretized version of them is rather observed. Therefore, all curves are sampled at the same grid of 100 equidistant points on $[0, 1]$.

The response variable Y_i is simulated according to the heteroscedastic regression model below:

$$Y_i = \int_{-1}^1 tX_i(t)dt + \left(\int_{-1}^1 |t| X_i^2(t)dt \right) \varepsilon_i, \quad i = 1, \dots, n,$$

where the residuals ε_i are generated according to the following models:

model_1 : $\varepsilon_i = 1/\sqrt{2}(\varepsilon_{i-1} + \eta_i)$, where η_i iid centred Gaussian random variables with mean equal zero and variance equal to 5.

model_2 : ε_i are independent Bernoulli random variables with mean equal to $1/2$.

We split this sample into two parts: a learning subsample $\{(X_i, Y_i); i = 1, \dots, 450\}$ which is used to find the 'optimal' smoothing parameters as well as the 'optimal' step sizes to calculate the real-time and the static estimators. The testing subsample $\{(X_j, Y_j); j = 451, \dots, 500\}$ is rather used to evaluate the performance of the predictor.

Observe that the estimator of the variance function can be seen as a regression function where the response variables are the square of the observed residuals. Therefore, and by analogy as for the conditional mean, the conditional variance is seen as the predictor of the true errors $[Y_i - r(X_i)]^2$ in the testing subsample. Since we investigate the case of real-time estimation, our purpose is to initiate the estimation of the conditional variance based on $n_0 = 30$ observations. Such initial value is obtained via Hu's nonrecursive estimator in (4) and performed with the Ferraty and Vieu's R-routine called *funopare.knn.lcv*.¹

The next step consists in updating the initial value of the estimator until the final sample size $30 < N \leq 500$ is reached. The updated values are obtained by using the recursive estimator given in (6) and compared, in terms of accuracy and computation time, to the nonrecursive estimator as defined in (4). Under the functional data framework, considered here, the computation of Δ_n depends on some tuning parameters. To this end, the semi-metric $d(\cdot, \cdot)$ is based on the L_2 distance of the second derivatives of the curves. More precisely

$$d(X_i, x) = \sqrt{\frac{1}{\sqrt{2}} \langle X_i^{(2)} - x^{(2)}, X_i^{(2)} - x^{(2)} \rangle}, \quad i = 1, \dots, N,$$

where $X_i^{(2)}$ and $x^{(2)}$ are, respectively, the second derivative of the curves X_i and x and $\langle \cdot, \cdot \rangle$ is the L_2 inner product.

As discussed in [6], this choice is argued by the fact that the curves are smooth enough. The choice of $K(\cdot)$ is not crucial in nonparametric estimation, thus we use the quadratic kernel defined by $K(u) = \frac{3}{2}(1 - u^2)\mathbb{1}_{[0,1]}(u)$, $u \in \mathbb{R}$, which behaves correctly in practice and is easy to implement.

4.1. Accuracy with respect to θ_n and τ_n choices

To perform the estimators (5) and (6), we consider the following cases $\theta_n, \tau_n \in \{\Delta_n/\sum_{i=1}^n \Delta_i, \Delta_n F_n(h_n)/\sum_{i=1}^n \Delta_i F_i(h_i), \Delta_n/n F_n(h_n), 1/n\}$, where F_n is the empirical version of F_x and $(h_i)_{i=1, \dots, n}$ a sequence of bandwidth selected such that

$$h_n = C \max_{1 \leq i \leq n} d(X_i, x) n^{-\nu}, \quad n = n_0 + 1, \dots, N. \tag{20}$$

The constants C and ν are such that $0 < C < 10, 0 < \nu < 1$ and they minimize the following cross-variation criterion:

$$CV(C, \nu) := \frac{1}{n} \sum_{j=1}^n [Y_j - \widehat{r}_{n,[-j]}(X_j)]^2, \tag{21}$$

where $\widehat{r}_{n,[-j]}$ indicates the regression estimator based on the sample $(X_1, Y_1) \dots (X_n, Y_n)$ leaving out the pair (X_j, Y_j) from the sample. Then, we apply the same method at a second stage to select the ‘optimal’ bandwidth associated to the conditional variance estimator given in (6).

Observe that the estimator of the conditional variance can be seen as a regression function where the response variables are the square of the observed residuals. Therefore, and by analogy as for the conditional mean, the conditional variance is seen as the predictor of the true errors $[Y_i - r(X_i)]^2$ in the testing subsample. In this subsection, we consider $N = 100$ and the simulations are repeated over 100 replications.

Let us denote by $\sigma_{n,k}^2(x)$ the value of the conditional variance estimator computed from the k th replication of the Monte–Carlo procedure and evaluated at a fixed curve x . The prediction accuracy (at step k) of the estimators is measured by the average mean square error $(AMSE_k)_{k=1, \dots, 100}$ such that

$$AMSE_k = \frac{1}{50(N - n_0)} \sum_{n=n_0+1}^N \sum_{j=451}^{500} \left\{ [Y_j - r(X_j)]^2 - \sigma_{n,k}^2(X_j) \right\}^2, \quad k = 1, \dots, 100. \tag{22}$$

The prediction errors (the box-plots of the sequence $(AMSE_k)_{k=1, \dots, 100}$) are displayed in Figure 1. It can be observed that, for the considered choice of bandwidth h_n , the prediction errors are comparable for the different values of step-size parameters θ_n and τ_n considered in this simulation study. Thus, for a specific choice of bandwidth, the step-size parameters do not have a major influence on the quality of the prediction even though the smallest AMSEs were recorded for $(\theta_n, \tau_n) = (\Delta_n/\sum_{i=1}^n \Delta_i, 1/n)$.

4.2. Comparison of recursive and nonrecursive estimators

In this subsection, the step-seizes are selected, as per the previous study, such that $\theta_n = \Delta_n/\sum_{i=1}^n \Delta_i$ and $\tau_n = 1/n$. The purpose of this simulation study is to assess the accuracy of the recursive conditional variance estimator and compare it to its natural competitor introduced by Hu [8].

Following the study of Rachdi and Vieu [36], an automatic selection of the bandwidth with cross-validation procedure is considered to compute the nonrecursive estimators of



Figure 1. Boxplots of the average (over the $N - n_0$ estimation steps and the test sample) mean square errors (computed at the various iteration $k = 1, \dots, 100$ of the Monte–Carlo procedure) of the conditional variance estimator defined in (6), using different values of θ_n and τ_n , $n = 31, \dots, 100$ in models_1 and models_2.

the regression and the conditional variance defined, respectively, in (3) and (4). For the recursive estimators of the regression and the conditional variance, the bandwidths are selected using the criteria described in (20).

In this comparison, we consider model_1 and model_2 described above and the final sample sizes are 50, 100 and 200. We consider 100 replications of the Monte–Carlo procedure and compute the average mean square error defined in (22). The results are displayed in Figure 2. For a fixed sample size N , a comparison of the AMSEs shows that the non-recursive estimator is slightly better than the recursive one. Moreover, it can be noticed that prediction errors decrease when the sample size N increases. As it can be observed later,

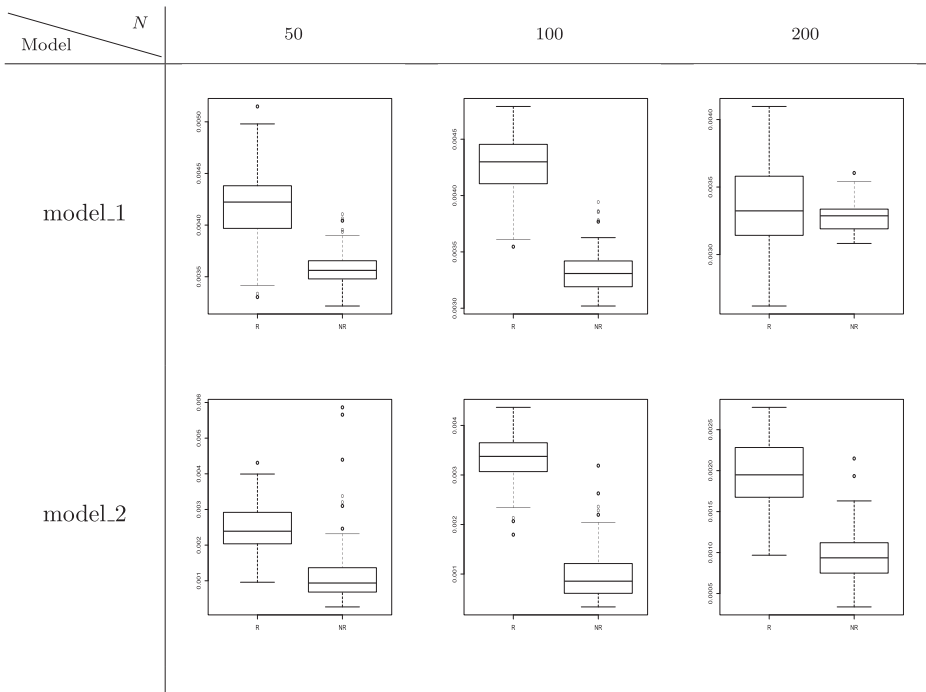


Figure 2. Boxplots of the average (over the $N - n_0$ estimation steps and the test sample) mean square errors (computed over 100 iterations of a Monte–Carlo simulation) of the conditional variance estimator defined in (6), using $\theta_n = \Delta_n / \sum_{i=1}^n \Delta_i$ and $\tau_n = 1/n, n = 31, \dots, N$ in models_1 and models_2.

the main advantage of the recursive estimator is its significantly reduced computational time.

5. Application to massive ozone concentration data

This section aims to illustrate the developed methodology using the air quality data in Mexico City metro area. The R package **aire.zmv**² is an open source tool which provides a real-time, daily maximum, minimum, or hourly average data for each pollution measuring station or geographical zones in the Zona Metropolitana del Valle de Mexico (greater Mexico City).

In this application, we are interested in real-time (online) one day-ahead forecasting the average (over stations located in Mexico city) daily maximum ozone concentration. The raw data comes in the form of a time series observed at each hour from 1986 to 2019. Since our purpose is to provide an illustration of the developed methodology, we consider only the data from 01 January 2005 to 15 July 2019. There are some missing measurements due to sensors failure at some hours, such a data are removed from the study for simplicity.

The first step in this analysis consists in transforming the main discrete time series into a functional time series where curves displayed in Figure 3 are realizations of the functional random variable X . In total, there are $n = 4988$ daily curves, say X_1, \dots, X_n , discretized at $p = 24$ points (corresponding the hours of the day). The response variable of interest Y in this study is the average daily maximum ozone concentration and the covariate X

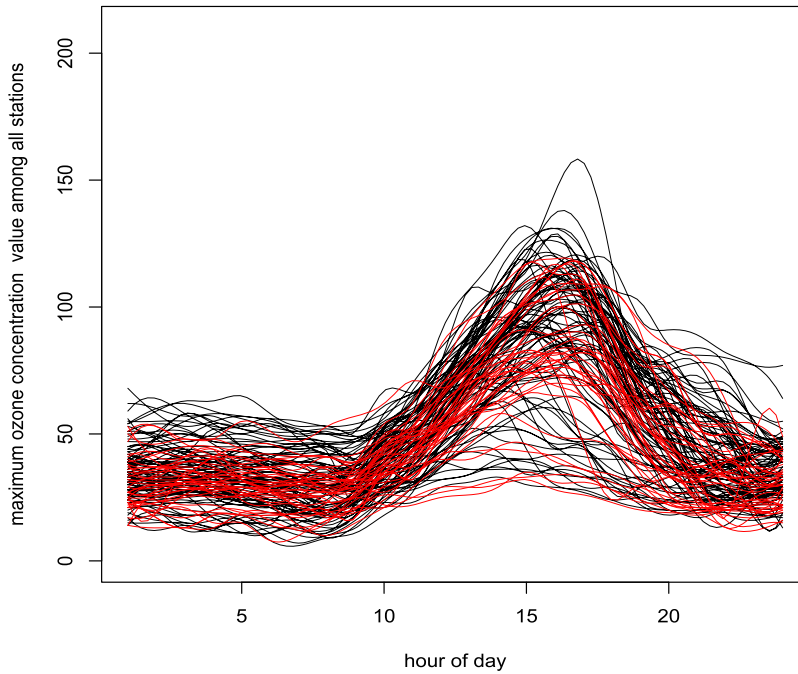


Figure 3. A sample of intraday average maximum ozone concentration curves in Mexico city.

used to predict Y is the previous day average maximum ozone concentration curve. Here, it is assumed that the current average maximum ozone concentration curve gives enough information to predict the next day average peak ozone concentration. Such assumption is known as Markov property which is well considered in short-term time series forecasting.

More formally, for a specific day d , Y_d is such that $Y_d = \max_{1 \leq j \leq 24} X_d(j)$ where $X_d(j)$ is the average maximum ozone concentration of the day d at the hour j . Therefore, the available sample is $(X_1, Y_2), (X_2, Y_3), \dots, (X_{n-1}, Y_n)$.

Our purpose in this application is to use the regression function as a day-ahead predictor of the unknown average daily ozone concentration peak given the last observed intraday ozone concentration curve. In other words, the prediction of the average peak ozone concentration in Mexico city at the day $n+1$ will be denoted by \widehat{Y}_n^{nr} which is calculated by using the whole available sample till the day n and by considering the non-recursive regression estimator defined in (3) as a predictor. Moreover, the risk associated with such prediction is assessed by the conditional variance denoted \widehat{R}_n^{nr} which is calculated nonrecursively using (4). On the other hand, the average daily peak ozone concentration as well as its volatility are computed recursively (denoted respectively $\widehat{Y}_n^r := r_{n-1}(X_n)$ and $\widehat{R}_n^r := \sigma_{n-1}(X_n)$) using the recursive regression and conditional variance estimators defined, respectively, in (5) and (6).

Days from 11 April 2005 to 14 July 2019 are used as a testing subsample allowing to assess the forecasts, whereas the rest of the data are considered as a training subsample used to select the ‘optimal’ smoothing parameters. Notice that, under this online forecasting framework, the training subsample is sequentially updated (each day) by one additional

Algorithm 1 Online one day-ahead volatility forecasting (recursive)

1: **procedure** ONLINE ONE DAY-AHEAD FORECASTING OF PEAK OZONE CONCENTRATION AND ITS VOLATILITY

2: **Initialization** consider the first $n_0 = 100$ pairs $\{(X_1, Y_2), \dots, (X_{n_0-1}, Y_{n_0}), X_{n_0}\}$ and use the Ferraty and Vieu's **R**-routine to

- (i) forecast Y_{n_0+1} using the regression estimator defined in (3). Such predicted value is denoted $\widehat{Y}_{n_0}^{init} := r_{n_0-1}(X_{n_0})$. Let $h_{n_0-1}^Y$ be the bandwidth used to make such prediction;
- (ii) calculate the risk associated with $\widehat{Y}_{n_0}^{init}$ by using the offline conditional variance estimator given by (4) and let us denoted by $\widehat{R}_{n_0}^{init} := \sigma_{n_0-1}^2(X_{n_0})$. Let $h_{n_0-1}^R$ be its associated bandwidth parameter.
- (iii) Then calculate the small ball probabilities and the step-sizes (with $h_{n_0-1} \in \{h_{n_0-1}^Y, h_{n_0-1}^R\}$) as:

$$\widehat{F}_{n_0-1} = \frac{1}{n_0 - 1} \sum_{j=1}^{n_0-1} \mathbb{1}\{d(X_j, X_{n_0}) \leq h_{n_0-1}\},$$

$$\theta_{n_0-1} = \tau_{n_0-1} = \frac{1}{(n_0 - 1)\widehat{F}_{n_0-1}} K(d(X_{n_0-1}, X_{n_0})/h_{n_0-1}).$$

3: **for** each j in $(n_0 + 1)$ **to** $n = 4988$ **do**

4: select the parameters (C_Y, ν_Y) and (C_R, ν_R) using the cross-validation procedure (21) on the grid of candidates $\{0.5, 1, 2, 10, 50\} \times \{1/300, 1/200, 1/100, 1/50, 1/10, 1/2\}$. Then calculate the new bandwidth parameters $h_{j-1} = C \max_{1 \leq i \leq j-1} d(X_i, X_j) (j - 1)^{-\nu}$, $(C, \nu) \in \{C_Y, C_R\} \times \{\nu_Y, \nu_R\}$.

5: update the small ball probabilities, the step-sizes and the last forecasts \widehat{Y}_{j-1}^r and \widehat{R}_{j-1}^r by considering only the last received pair (X_{j-1}, Y_j) using the following formulas:

$$\widehat{F}_{j-1} = \widehat{F}_{j-2} + \frac{1}{j - 1} \left[\mathbb{1}\{d(X_{j-1}, X_j) \leq h_{j-1}\} - \widehat{F}_{j-2} \right], \tag{23}$$

$$\theta_{j-1} = \tau_{j-1} = \frac{1}{(j - 1)\widehat{F}_{j-1}} K(d(X_{j-1}, X_j)/h_{j-1}),$$

$$\widehat{Y}_j^r = \widehat{Y}_{j-1}^r + \theta_{j-1} (Y_j - \widehat{Y}_{j-1}^r), \tag{24}$$

$$\widehat{R}_j^r = \widehat{R}_{j-1}^r + \tau_{j-1} \left[(Y_j - \widehat{Y}_{j-1}^r)^2 - \widehat{R}_{j-1}^r \right]. \tag{25}$$

6: **end for**

7: **end procedure**

observation which is removed from the testing subsample. Moreover, as in the simulation study, the quadratic kernel is considered in this application and given the degree of smoothness of the curves (see Figure 3) the L_2 distance between the second derivatives of the curves is considered as a semi-metric for both regression and conditional variance estimators. The 'optimal' bandwidths, for the recursive and the nonrecursive regression

and conditional variance estimators, are obtained by using the same procedures detailed in subsection 4.1.

Algorithms 1 and 2 describe the procedures used to obtain the online (recursive) as well as the offline (nonrecursive) day-ahead forecasts of the daily average peak ozone concentration and its volatility.

Algorithm 2 Offline one day-ahead volatility forecasting (nonrecursive)

- 1: **procedure** OFFLINE ONE DAY-AHEAD FORECASTING OF PEAK OZONE CONCENTRATION AND ITS VOLATILITY
 - 2: **Initialization** consider the first $n_0 = 100$ pairs $\{(X_1, Y_2), \dots, (X_{n_0-1}, Y_{n_0}), X_{n_0}\}$ and use the Ferraty and Vieu's **R**-routine to
 - (i) forecast Y_{n_0+1} using the regression estimator defined in (3). Such predicted value is denoted $\widehat{Y}_{n_0}^{init} := r_{n_0-1}(X_{n_0})$. Let $h_{n_0-1}^Y$ be the bandwidth used to make such prediction;
 - (ii) calculate the risk associated to $\widehat{Y}_{n_0}^{init}$ by using the offline conditional variance estimator given by (4) and let us denoted by $\widehat{R}_{n_0}^{init} := \sigma_{n_0-1}^2(X_{n_0})$. Let $h_{n_0-1}^R$ be its associated bandwidth parameter.
 - 3: **for** each j in $(n_0 + 1)$ **to** $n = 4988$ **do**
 - 4: select the bandwidth h_{j-1} by a cross-validation procedure over a sequence of 20 candidates of bandwidths, based on the full sample $\{(X_1, Y_2), \dots, (X_{j-1}, Y_j), X_j\}$. This procedure is performed using the **R** routines *funopare.kernel.cv* (a global bandwidth is automatically selected with a cross-validation procedure) or *funopare.knn.lcv* (a local bandwidth (i.e. local number of neighbours) is selected by a cross-validation procedure).
 - 5: again use the full sample $\{(X_1, Y_2), \dots, (X_{j-1}, Y_j), X_j\}$ and the selected bandwidth to calculate \widehat{Y}_j^{nr} and \widehat{R}_j^{nr} the nonrecursive (static) forecasts of the average peak load and its volatility using, respectively, nonrecursive estimators of the regression and the variance operators defined in (3) and (4). This is done with the help of the **R** routine *funopare.kernel*.
 - 6: **end for**
 - 7: **end procedure**
-

Remark 2: Observe that if we replace, in Equations (24) and (25), \widehat{Y}_j^r and \widehat{R}_j^r by their values, one gets, for $j = n_0 + 1, \dots, n$:

$$r_{j-1}(X_j) = r_{j-2}(X_{j-1}) + \theta_{j-1} [Y_j - r_{j-2}(X_{j-1})] \quad (26)$$

and

$$\sigma_{j-1}^2(X_j) = \sigma_{j-2}^2(X_{j-1}) + \tau_{j-1} \left\{ [Y_j - r_{j-2}(X_{j-1})]^2 - \sigma_{j-2}^2(X_{j-1}) \right\}. \quad (27)$$

Therefore, a comparison, to the recursive estimators defined in Equations (5) and (6), shows that the argument where the regression and the conditional variance operators are applied is not fixed under the forecasting context. Indeed, in (5) and (6) the regression and the conditional variance are evaluated at a fixed curve x whereas in (26) and (27) the estimators are evaluated at different curves. This effect is due to the online forecasting

framework considered in this study. As a consequence, the predictors used in this application to obtain the one day-ahead forecasts do not exactly fit to the theoretical framework developed in Section 3. Nevertheless, the actual application aims to illustrate how the recursive estimators defined in (5) and (6) might be useful to provide real-time forecasts even though the proxies of the above estimators are used.

The theoretical study of the predictors defined in (26) and (27), where the argument of the regression and the variance operators moves over time, will be the purpose of future studies. Notice that, the same problem also occurs to update, recursively, the small ball probability estimator given in Equation (23). Indeed, one can observe that the argument, which is the bandwidth, in the empirical distribution function changes from an iteration to another. A nonrecursive representation of the empirical distribution function can be used but a supplementary computational complexity will be added to the approach.

5.1. On the assessment of the accuracy of online and offline one day-ahead forecasts

To assess the performance, in terms of forecast accuracy, of online and offline forecasts, the relative absolute prediction error is considered. Thus, for each step $j \in \{101, \dots, 4988\}$

$$RAE_j^r = \frac{|Y_j - \widehat{Y}_{j-1}^r|}{Y_j} \quad \text{and} \quad RAE_j^{nr} = \frac{|Y_j - \widehat{Y}_{j-1}^{nr}|}{Y_j}.$$

Observe that, in contrast to the simulation study where the form of the conditional variance is known, in practice we cannot assess the prediction performance of the volatility. Figure 4 displays the relative absolute prediction errors of the average ozone concentration peak as well as the associated predicted volatilities. The results are clustered by type of the day in the week. Overall, the nonrecursive method provides more accurate predictions than the recursive one. Moreover, one can observe smaller predicted volatilities with the nonrecursive method compared to the recursive one.

In the following subsections 5.2–5.4, we are interested in answering the following questions:

- (1) Does considering recursive small ball probability estimator reduce the computation time of the online forecasts?
- (2) After separating the time needed to select the bandwidths, how fast are the online forecasts compared to the offline ones?
- (3) Which bandwidth criterion selection reduces more the computation time of online forecasts?

5.2. The effect of considering a recursive small ball probability in reducing the online forecasts’ computation time

In this subsection, we are interested in answering the question (1) above, which consists in assessing the computation time reduction of the online forecasts when the small ball probability is also estimated recursively as per Equation (23). Figure 5 displays the sequence of time (daily updating time and cumulated time) needed to calculate the one-day ahead

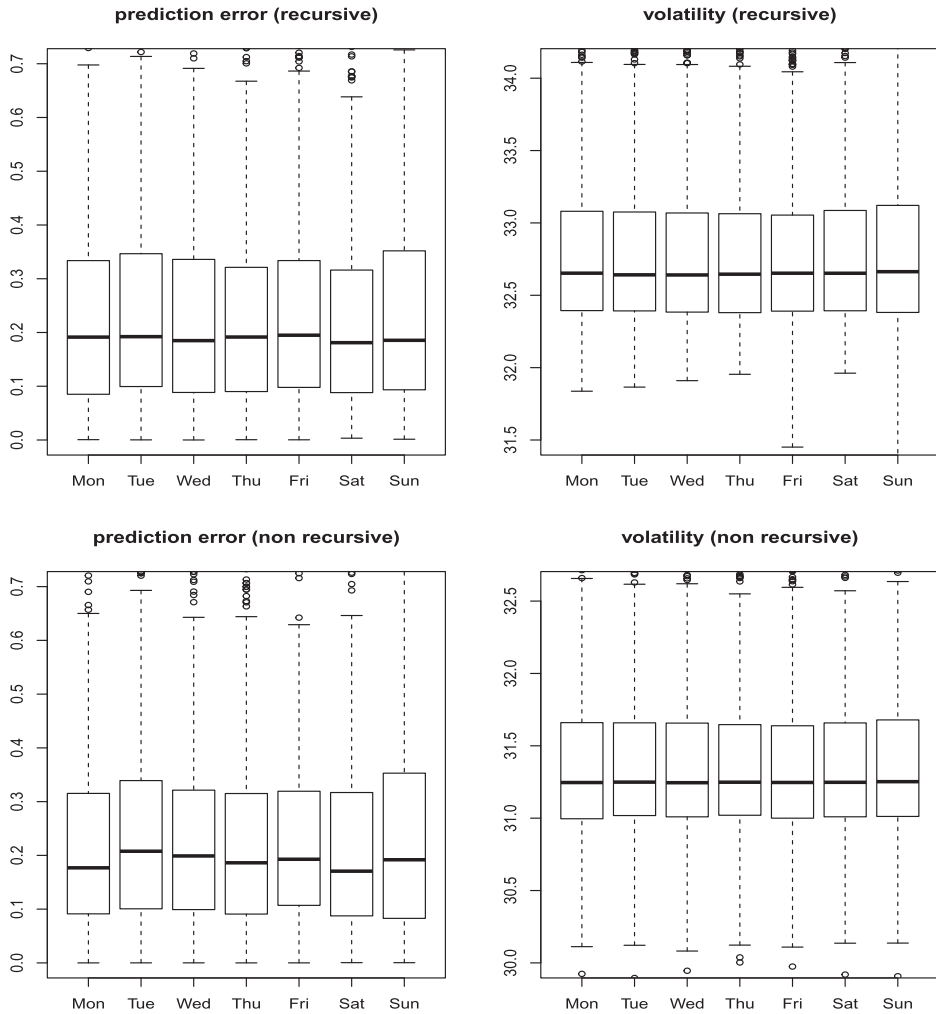


Figure 4. Relative absolute prediction errors, clustered by day type, of average peak ozone concentration in Mexico city obtained with recursive and nonrecursive approaches (left panels). The predicted volatility, clustered by type of day, using recursive and nonrecursive approaches (right panels).

forecasts of the peak ozone concentration as well as its volatility using the recursive estimators whenever the small ball probability is estimated recursively or not. It is worth noting, here, that the time needed to select the bandwidth is excluded from the comparison. In Figure 5 recursive estimators of the regression and the conditional variance involving recursive estimation of the small ball probability are called *totally recursive*. Whereas, those using a nonrecursive estimator of the small ball probability are called *partially recursive*.

5.3. Time efficiency comparison between online and offline predictors when small ball probability is estimated recursively

The purpose of this subsection is to compare the time efficiency of the *total recursive* estimators to the *nonrecursive* ones after excluding the time elapsed to select the smoothing

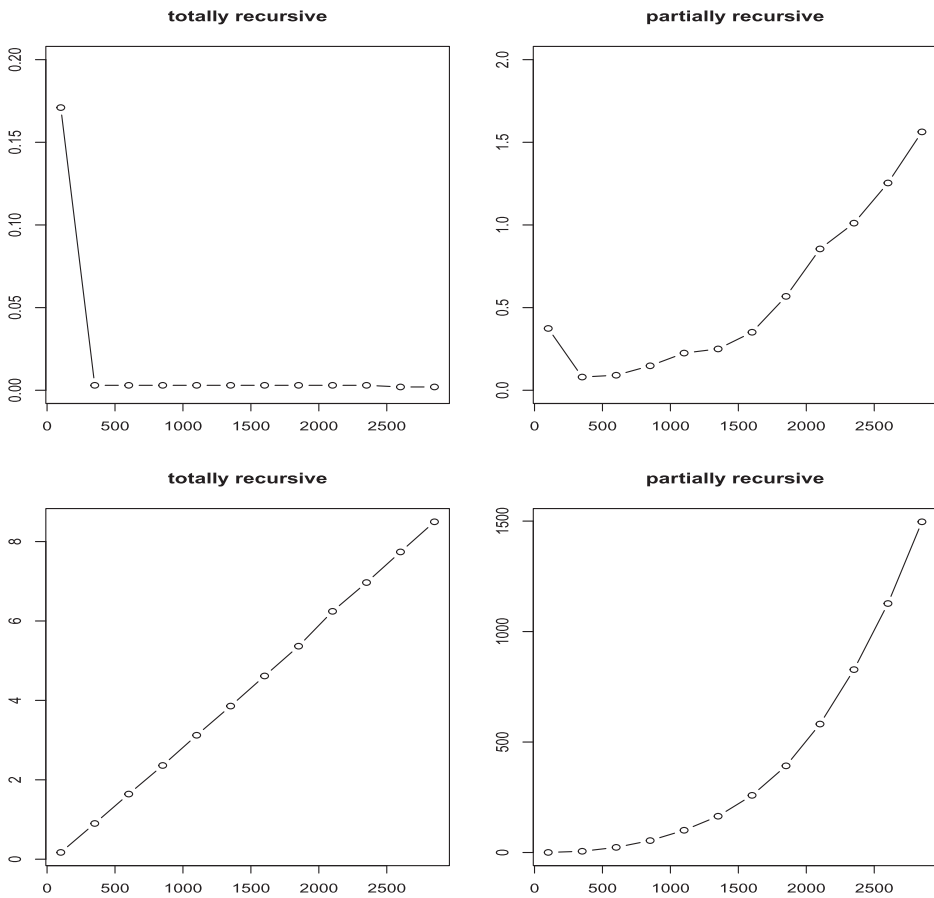


Figure 5. Daily computational times in seconds (as a function of the number of observations in the testing sample) needed to obtain one-day ahead forecasts (top two panels) and the cumulated computational time needed to obtain all the predictions until the actual testing sample size (bottom two panels). Left panels: recursive predictor with a recursive small ball probability and right panels: recursive predictor with a nonrecursive small ball probability.

parameters. More precisely, we first select the smoothing parameters of the *totally recursive* and the *nonrecursive* estimators according to steps 4 detailed in Algorithms 1 and 2. Then, we start assessing the computation time efficiency of the *totally recursive* predictions and the *nonrecursive* ones. In other words, we compare the time elapsed to execute the steps 5 for the online forecasts (Algorithm 1) and for offline forecasts (Algorithm 2).³

Figure 6 displays the computation time elapsed to obtain the one-day ahead forecasts, as well as the cumulative time needed to cover all testing sample, using the *totally recursive* and the *nonrecursive* estimators. We recall that the time needed to select the smoothing parameter is excluded from computation time assessment. One can notice that predictions with totally recursive approach are much faster than the nonrecursive one. The gain in speed of the recursive predictor is substantial when the step size in step 5 of Algorithm 1 is provided using a recursive proxy of the small ball probability.

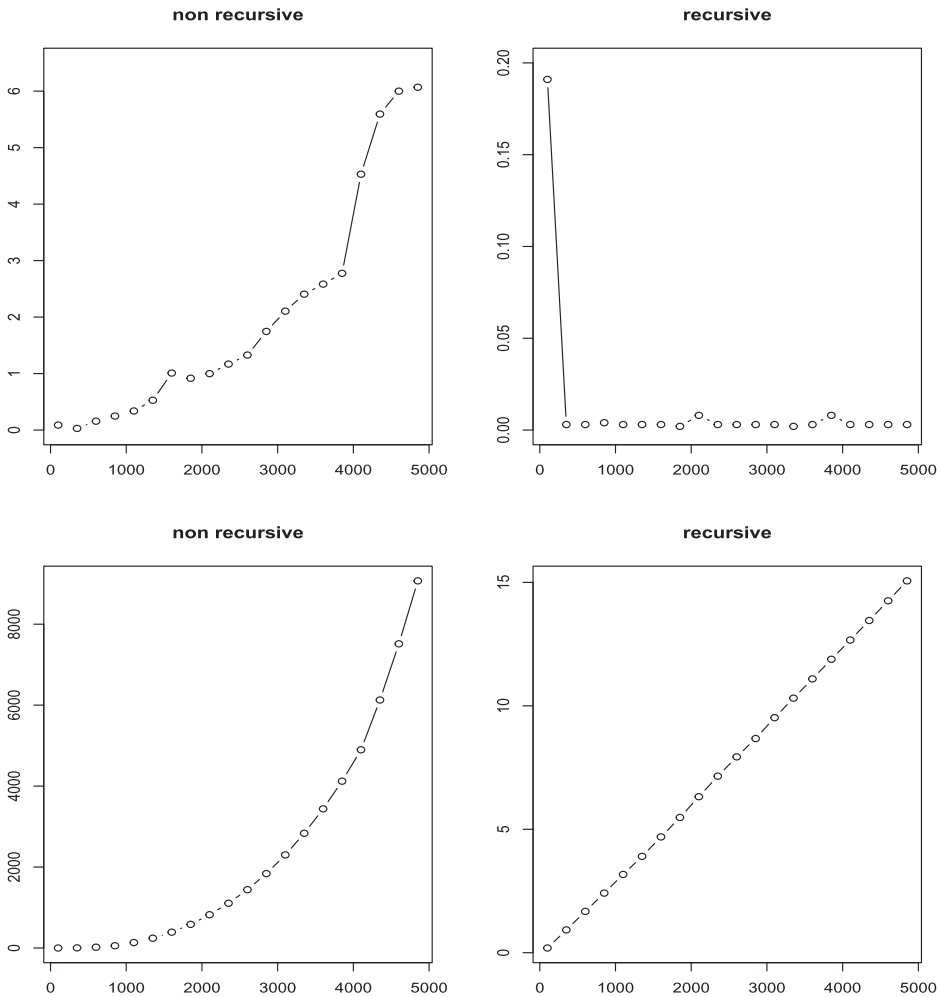


Figure 6. Daily computational times in seconds (as a function of the number of observations in the testing sample) needed to obtain one-day ahead forecasts (top two panels) and the cumulated computational time needed to obtain all the predictions until the actual testing sample size (bottom two panels). Left two panels: totally recursive predictor and right two panels: nonrecursive predictor.

5.4. Discussion on the smoothing parameter selection criterion

In this subsection, we are interested in comparing the smoothing parameter selection criterion detailed in steps 4 of Algorithms 1 and 2. We are trying to answer the question (3) above and assessing the two bandwidth selection criterion in reducing the computation time of the totally recursive estimator. In order to conduct a fair comparison, a grid of 20 possible bandwidths is considered in performing steps 4 in the two algorithms.

Figure 7 displays the time elapsed, for each daily bandwidth selection using the cross-validation procedures described in steps 4 of Algorithms 1 and 2. One can observe that the bandwidth criterion given in Algorithm 1 is more efficient, in reducing the computation time of the totally recursive estimators, compared to the one used in Algorithm 2.

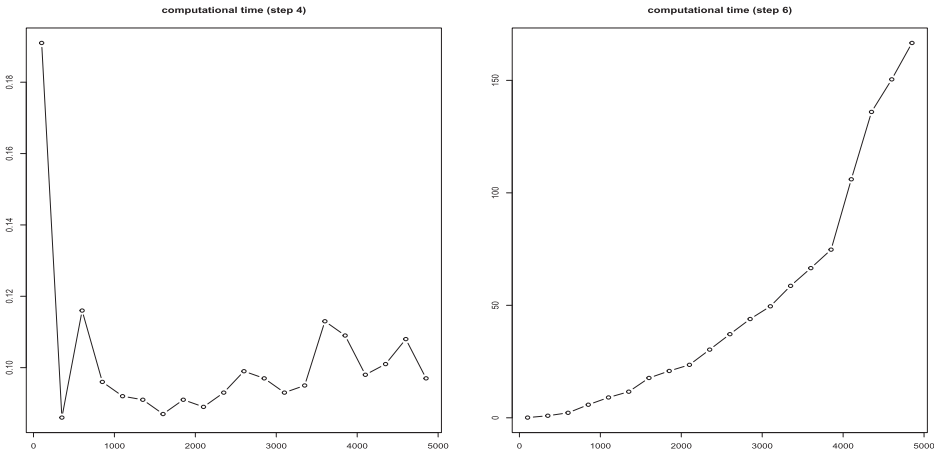


Figure 7. Daily computational time in seconds (as a function of the number of observations in the testing sample) needed to select the bandwidth using the cross-validation procedures described in steps 4 of Algorithm 1 (left panel) and Algorithm 2 (right panel).

6. Proof of main results

Here and later, C will denote a positive constant whose value is not important and may vary from line to line.

Proof of Theorem 3.1: For any $q \geq 2$ and $z \in \mathcal{B}(x, h_n)$, let us denote by

$$V_n(z) := |r_n(z) - r(z)|^q \quad \text{and} \quad \delta_n := \delta_n(z) = \theta_n [Y_n - r_{n-1}(z)]$$

with $\theta_n := \theta_n(z)$ defined in Theorem 3.1.

Using the recursive form of the regression operator given by Equation (5) and a Taylor expansion of the function $u \mapsto |u|^q$ over the points $u_0 = r_{n-1}(z) - r(z)$ and $u_0 + \delta_n$ one gets:

$$\begin{aligned} V_n(z) &= V_{n-1}(z) + q\delta_n V_{n-1}^{(q-2)/q}(z) [r_{n-1}(z) - r(z)] \\ &\quad + \frac{1}{2}q(q-1)\delta_n^2 \left[V_{n-1}^{1/q}(z) + |\delta_n|v \right]^{q-2}, \end{aligned} \tag{28}$$

with $0 < v < 1$.

First, by applying the c_r -inequality, we obtain:

$$\mathbb{E} \left\{ \delta_n^2 \left[V_{n-1}^{1/q}(z) + |\delta_n|v \right]^{q-2} \middle| \mathcal{F}_{n-1} \right\} \leq c_{q-1} \mathbb{E} \left\{ \delta_n^2 \left[V_{n-1}^{(q-2)/q}(z) + |\delta_n|^{q-2} \right] \middle| \mathcal{F}_{n-1} \right\}. \tag{29}$$

Next, observe that

$$\begin{aligned} \mathbb{E}(\delta_n | \mathcal{F}_{n-1}) &= \mathbb{E} \{ [r(X_n) + \sigma(X_n)\varepsilon_n - r_{n-1}(z)] \theta_n | \mathcal{F}_{n-1} \} \\ &= \frac{\rho_n}{F(h_n)} \left(\mathbb{E} \{ [r(X_n) - r(z) + r(z) - r_{n-1}(z)] \Delta_n(z) | \mathcal{F}_{n-1} \} \right. \\ &\quad \left. + \mathbb{E} \{ \sigma(X_n)\varepsilon_n \Delta_n(z) | \mathcal{F}_{n-1} \} \right) \\ &=: \frac{\rho_n}{F(h_n)} (I_1 + I_2). \end{aligned}$$

Using condition (2) and by a double conditioning with respect to the σ -field \mathcal{G}_{n-1} , one can write

$$I_2 = \mathbb{E} \{ \mathbb{E} [\sigma(X_n) \varepsilon_n \Delta_n(z) | \mathcal{G}_{n-1}] | \mathcal{F}_{n-1} \} = \mathbb{E} \{ \sigma(X_n) \Delta_n(z) \mathbb{E}(\varepsilon_n | \mathcal{G}_{n-1}) | \mathcal{F}_{n-1} \} = 0.$$

For the first term I_1 , we have by assumption (H3)(i)

$$I_1 \leq \left\{ \sup_{y \in B(x, h)} |r(y) - r(z)| + [r(z) - r_{n-1}(z)] \right\} \mathbb{E}(\Delta_n(z) | \mathcal{F}_{n-1}).$$

Making use of Lemma A.0.1 and assumptions (H2)–(H3) one obtains, uniformly in z , $\mathbb{E}(\Delta_n(z) | \mathcal{F}_{n-1}) \leq C^* F(h_n)$, for n large enough, where C^* is the positive constant independent of z defined in (11). Therefore, we have

$$\begin{aligned} \mathbb{E}(\delta_n | \mathcal{F}_{n-1}) &\leq \rho_n C^* \left\{ \sup_{y \in B(x, h)} |r(y) - r(z)| + [r(z) - r_{n-1}(z)] \right\} \\ &\leq \rho_n C^* \{ 2c_1 h_n^\alpha + [r(z) - r_{n-1}(z)] \}. \end{aligned}$$

Hence,

$$\begin{aligned} \mathbb{E}(\delta_n [r_{n-1}(z) - r(z)] | \mathcal{F}_{n-1}) &\leq \rho_n C^* \left\{ 2c_1 h_n^\alpha |r_{n-1}(z) - r(z)| - V_{n-1}^{2/q}(z) \right\} \\ &\leq \rho_n C^* \left[c_1 h_n^\alpha V_{n-1}^{1/q}(z) - V_{n-1}^{2/q}(z) \right]. \end{aligned} \quad (30)$$

Now, let $\xi_n = Y_n - r(X_n)$, then, for $1 \leq t \leq q$, one obtains, through c_r -inequality

$$\begin{aligned} &\mathbb{E}(|\delta_n|^t | \mathcal{F}_{n-1}) \\ &\leq \frac{\rho_n^t}{F^t(h_n)} \mathbb{E} \left\{ [c_t^2 |r(X_n) - r(z)|^t + c_t^2 |r(z) - r_{n-1}(z)|^t + c_t |\xi_n|^t] \Delta_n^t(z) | \mathcal{F}_{n-1} \right\} \\ &\leq \sup_{y \in B(x, h_n)} |r(y) - r(z)|^t \times c_t^2 \frac{\rho_n^t}{F^t(h_n)} \mathbb{E}(\Delta_n^t(z) | \mathcal{F}_{n-1}) \\ &\quad + \frac{\rho_n^t}{F^t(h_n)} V_{n-1}^{t/q}(z) \mathbb{E}(\Delta_n^t(z) | \mathcal{F}_{n-1}) \\ &\quad + c_t \frac{\rho_n^t}{F^t(h_n)} \mathbb{E} \left\{ \mathbb{E} [\sigma^t(X_n) |\varepsilon_n|^t \Delta_n^t(z) | \mathcal{G}_{n-1}] | \mathcal{F}_{n-1} \right\}. \end{aligned}$$

By using assumptions (H0) and (H3) one gets

$$\begin{aligned} \frac{\mathbb{E} \left\{ \mathbb{E} [\sigma^t(X_n) |\varepsilon_n|^t \Delta_n^t(z) | \mathcal{G}_{n-1}] | \mathcal{F}_{n-1} \right\}}{\rho_n^{-t} F^t(h_n)} &\leq c_t \rho_n^t \mu_t \left[\sup_{y \in B(x, h_n)} |\sigma^t(y) - \sigma^t(z)| + \sigma^t(z) \right] \\ &\quad \times \frac{\mathbb{E}(\Delta_n^t(z) | \mathcal{F}_{n-1})}{F^t(h_n)}. \end{aligned}$$

From assumption (H2), assumption (H3) allows to write

$$\begin{aligned} \mathbb{E}(|\delta_n|^t | \mathcal{F}_{n-1}) &\leq \rho_n^t \left\{ c_1 c_t^2 h_n^{t\alpha} + c_t^2 V_{n-1}^{t/q}(z) + \mu_t(c_2 h_n^\beta + C) \right\} \frac{\mathbb{E}(\Delta_n^t(z) | \mathcal{F}_{n-1})}{F^t(h_n)} \\ &\leq C^* \rho_n^t \left\{ c_1 c_t^2 h_n^{t\alpha} + c_t^2 V_{n-1}^{t/q}(z) + \mu_t(c_2 c_t^2 h_n^\beta + C) \right\} \phi^{1-t}(h_n). \end{aligned} \quad (31)$$

By substituting (29)–(31) in (28) and taking the conditional expectation, with respect to \mathcal{F}_{n-1} , in both sides of the inequality, one obtains

$$\begin{aligned} \mathbb{E}(V_n(z) | \mathcal{F}_{n-1}) &\leq V_{n-1}(z) \left\{ 1 - \rho_n \left[qC^* - c_{13} \frac{\rho_n}{\phi(h_n)} - c_{13} \rho_n^{q-1} \phi^{1-q}(h_n) \right] \right\} \\ &\quad + c_{14} \rho_n \left(\begin{aligned} &h_n^\alpha V_{n-1}^{(q-1)/q}(z) + \frac{\rho_n}{\phi(h_n)} h_n^{2\alpha} V_{n-1}^{(q-2)/q}(z) \\ &+ \frac{\rho_n}{\phi(h_n)} V_{n-1}^{(q-2)/q}(z) + \frac{\rho_n}{\phi(h_n)} h_n^{q\alpha} + \rho_n^{q-1} \phi^{1-q}(h_n) \end{aligned} \right), \end{aligned} \quad (32)$$

where c_{13} and c_{14} are positive constants independent of z . Let $0 < \kappa_0 < qC^*/3$ be a constant, define

$$\kappa_s = (q - s - 1)/q \quad \text{and} \quad a_s = \left[\frac{\kappa_0 q \phi^s(h_n)}{c_{14} \rho_n^s (s + h_n^\alpha) (q - s - 1)} \right]^{-q/(s+1)}, \quad s \in \{0, 1\}.$$

Because n is large enough, so that

$$c_{13} \frac{\rho_n}{\phi(h_n)} + c_{13} \left(\frac{\rho_n}{\phi(h_n)} \right)^{q-1} < \kappa_0. \quad (33)$$

Using the mean value Theorem applied to the function $u \mapsto u^\kappa$, we get the inequality

$$V_n^\kappa(z) \leq (1 - \kappa) a_s^\kappa + \kappa a_s^{\kappa-1} V_n(z),$$

which together with (32) and (33) lead to

$$\begin{aligned} \mathbb{E}(V_n(z) | \mathcal{F}_{n-1}) &\leq V_{n-1}(z) [1 - \rho_n(qC^* - 3\kappa_0)] \\ &\quad + c_{15} \rho_n \left(\begin{aligned} &\rho^{q-1} \phi^{1-q}(h_n) h_n^{q\alpha} + \left[\frac{\rho_n}{\phi(h_n)} \right]^{q-1} \\ &+ h_n^{q\alpha} + \left[\frac{\rho_n}{\phi(h_n)} h_n^{2\alpha} \right]^{q/2} + \left[\frac{\rho_n}{\phi(h_n)} \right]^{q/2} \end{aligned} \right), \end{aligned} \quad (34)$$

Now, take expectation of both sides of (34) and denote

$$\mathcal{W}_n(x) = \sup_{z \in \mathcal{B}(x, h_n)} \mathbb{E}[V_n(z)].$$

Now, because the constants in the right-side in (34) are uniformly on $z \in \mathcal{B}(x, h_n)$, then we get

$$\mathcal{W}_n(x) \leq \mathcal{W}_{n-1}(x) [1 - \rho_n(qC^* - 3\kappa_0)] + c_{15} \rho_n \left(\begin{aligned} &(h_n^{q\alpha} + 1) \left[\frac{\rho_n}{\phi(h_n)} \right]^{q-1} \\ &+ (h_n^{q\alpha} + 1) \left[\frac{\rho_n}{\phi(h_n)} \right]^{q/2} + h_n^{q\alpha} \end{aligned} \right). \quad (35)$$

Since it is assumed that $h_n \rightarrow 0$ and by condition (12) we have $\rho_n \phi^{-1}(h_n) \rightarrow 0$, as $n \rightarrow \infty$, then Lemma A.0.2 allows to conclude the proof of the first part of this Theorem.

Now, we are interested in establishing the consistency rate given by (15) in Theorem 3.1. For this purpose, observe that for n large and κ_0 small enough, inequality (35) becomes

$$\mathcal{W}_n(x) \leq \mathcal{W}_{n-1}(x) (1 - \rho_n q C^*) + c_{15} \rho_n \left((h_n^{q\alpha} + 1) \left[\frac{\rho_n}{\phi(h_n)} \right]^{q-1} + (h_n^{q\alpha} + 1) \left[\frac{\rho_n}{\phi(h_n)} \right]^{q/2} + h_n^{q\alpha} \right).$$

Next, for a particular choice of ρ_n, h_n and $\phi(h_n)$ given in (14), the following inequality:

$$\mathcal{W}_n(x) \leq \mathcal{W}_{n-1}(x) \left(1 - \frac{\rho_0 q C^*}{n} \right) + D n^{-1-(q\alpha/(2\alpha+\nu))}, \quad (36)$$

is satisfied, where D is a positive constant. Finally, the application of Lemma A.0.3 permits us to conclude. \blacksquare

Proof of Theorem 3.2: Set $U_n = \sigma_n^2(x) - \sigma^2(x)$. From (6), we have that:

$$\begin{aligned} \mathbb{E}(U_n^2 | \mathcal{F}_{n-1}) &= U_{n-1}^2 + 2U_{n-1} \left(\mathbb{E} \left[(Y_n - r_{n-1}(X_n))^2 \tau_n | \mathcal{F}_{n-1} \right] - \sigma_{n-1}^2(x) \mathbb{E}(\tau_n | \mathcal{F}_{n-1}) \right) \\ &\quad + \mathbb{E} \left\{ [(Y_n - r_{n-1}(X_n))^2 - \sigma_{n-1}^2(x)]^2 \tau_n^2 | \mathcal{F}_{n-1} \right\} \\ &= U_{n-1}^2 + \frac{2\gamma_n U_{n-1} A_n}{F_x(h_n)} + \left[\frac{\gamma_n}{F_x(h_n)} \right]^2 B_n. \end{aligned} \quad (37)$$

we are now interested in the term

$$\begin{aligned} A_n &= \mathbb{E} \left[(Y_n - r_{n-1}(X_n))^2 \Delta_n(x) | \mathcal{F}_{n-1} \right] - \sigma_{n-1}^2(x) \mathbb{E}(\Delta_n(x) | \mathcal{F}_{n-1}) \\ &= \mathbb{E}(\xi_n^2 \Delta_n(x) | \mathcal{F}_{n-1}) + 2\mathbb{E} \{ [r(X_n) - r_{n-1}(X_n)] \xi_n \Delta_n(x) | \mathcal{F}_{n-1} \} \\ &\quad + \mathbb{E} \{ [r(X_n) - r_{n-1}(X_n)]^2 \Delta_n(x) | \mathcal{F}_{n-1} \} - \sigma_{n-1}^2(x) \mathbb{E}(\Delta_n(x) | \mathcal{F}_{n-1}). \end{aligned}$$

Observe that a double conditioning with respect to the σ -field \mathcal{G}_{n-1} allows us to write

$$\begin{aligned} &\mathbb{E} \{ [r(X_n) - r_{n-1}(X_n)] \xi_n \Delta_n(x) | \mathcal{F}_{n-1} \} \\ &= \mathbb{E} \{ \mathbb{E} \{ [r(X_n) - r_{n-1}(X_n)] \sigma(X_n) \varepsilon_n \Delta_n(x) | \mathcal{G}_{n-1} \} | \mathcal{F}_{n-1} \} \\ &= \mathbb{E} \{ [r(X_n) - r_{n-1}(X_n)] \sigma(X_n) \Delta_n(x) \mathbb{E}(\varepsilon_n | \mathcal{G}_{n-1}) | \mathcal{F}_{n-1} \} = 0, \end{aligned}$$

by condition (2). Then, by considering again condition (2), and using assumption (H3), one gets, almost surely,

$$\begin{aligned} A_n &= \mathbb{E}(\xi_n^2 \Delta_n(x) | \mathcal{F}_{n-1}) - \sigma_{n-1}^2(x) \mathbb{E}(\Delta_n(x) | \mathcal{F}_{n-1}) \\ &\quad + \mathbb{E} \{ [r(X_n) - r_{n-1}(X_n)]^2 \Delta_n(x) | \mathcal{F}_{n-1} \} \\ &= \mathbb{E} \{ [\sigma^2(X_n) - \sigma_{n-1}^2(x)] \mathbb{E}(\varepsilon_n^2 | \mathcal{G}_{n-1}) \Delta_n(x) | \mathcal{F}_{n-1} \} + A_{n1} \\ &\leq \left\{ \sup_{z \in B(x, h_n)} |\sigma^2(z) - \sigma^2(x)| - U_{n-1} \right\} \times \mathbb{E}(\Delta_n(x) | \mathcal{F}_{n-1}) + A_{n1}. \end{aligned}$$

Therefore, we have, by assumption (H3)

$$A_n = \left[-U_{n-1} + \mathcal{O}\left(h_n^\beta\right)\right] \mathbb{E}\left(\Delta_n(x) \mid \mathcal{F}_{n-1}\right) + A_{n1}. \tag{38}$$

The term A_{n1} in the right-side of (38) will be studied latter.

Now, let us focus on the term B_n that appears in Equation (37). Using successively the c_r -inequality, we get

$$\begin{aligned} B_n &\leq \mathbb{E} \left\{ \left[2\xi_n^2 + 2\left(r(X_n) - r_{n-1}(X_n)\right)^2 - \sigma_{n-1}^2(x) \right]^2 \Delta_n^2(x) \mid \mathcal{F}_{n-1} \right\} \\ &\leq 2\mathbb{E} \left\{ \left[2\xi_n^4 + 4\left(r(X_n) - r_{n-1}(X_n)\right)^4 + \left(\xi_n^2 - \sigma_{n-1}^2(x)\right)^2 \right] \Delta_n^2(x) \mid \mathcal{F}_{n-1} \right\} \\ &\leq 8\mathbb{E} \left\{ \left[r(X_n) - r_{n-1}(X_n) \right]^4 \Delta_n^2(x) \mid \mathcal{F}_{n-1} \right\} + 4\mathbb{E} \left[\xi_n^4 \Delta_n^2(x) \mid \mathcal{F}_{n-1} \right] \\ &\quad + 2\mathbb{E} \left\{ \left[\xi_n^2 - \sigma_{n-1}^2(x) \right]^2 \Delta_n^2(x) \mid \mathcal{F}_{n-1} \right\} =: 8B_{n1} + 4B_{n2} + 2B_{n3}. \end{aligned} \tag{39}$$

The term B_{n1} will be considered later, now let us study B_{n2} and B_{n3} . Similar to the control of A_n , by a double conditioning with respect to the σ -field \mathcal{G}_{n-1} and making use of the assumptions (H0) and (H3) and because $\sigma(x)$ is bounded, one can easily show that

$$\begin{aligned} B_{n,2} &= \mathbb{E} \left[\sigma^4(X_n) \varepsilon_n^4 \Delta_n^2(x) \mid \mathcal{F}_{n-1} \right] \\ &= \mathbb{E} \left\{ \left[\sigma^4(X_n) - \sigma^4(x) \right] \mathbb{E} \left(\varepsilon_n^4 \mid \mathcal{G}_{n-1} \right) \Delta_n^2(x) \mid \mathcal{F}_{n-1} \right\} \\ &\quad + \sigma^4(x) \mathbb{E} \left[\mathbb{E} \left(\varepsilon_n^4 \mid \mathcal{G}_{n-1} \right) \Delta_n^2(x) \mid \mathcal{F}_{n-1} \right] \\ &\leq \mu_4 \mathbb{E} \left(\Delta_n^2(x) \mid \mathcal{F}_{n-1} \right) \left\{ \sup_{z \in B(x, h_n)} \left| \sigma^4(z) - \sigma^4(x) \right| + \sigma^4(x) \right\} \\ &= \left[\mathcal{O}\left(h_n^\beta\right) + C \right] \mathbb{E} \left(\Delta_n^2(x) \mid \mathcal{F}_{n-1} \right). \end{aligned} \tag{40}$$

Turning to the term B_{n3} , the same arguments as in B_{n2} , with the help of c_r -inequality lead to

$$\begin{aligned} B_{n,3} &\leq 2\mathbb{E} \left\{ \left[\xi_n^2 - \sigma^2(X_n) \right]^2 \Delta_n^2(x) \mid \mathcal{F}_{n-1} \right\} + 2\mathbb{E} \left\{ \left[\sigma^2(X_n) - \sigma_{n-1}^2(x) \right]^2 \Delta_n^2(x) \mid \mathcal{F}_{n-1} \right\} \\ &=: \left[\mathcal{O}\left(h_n^{2\beta}\right) + C \right] \mathbb{E} \left(\Delta_n^2(x) \mid \mathcal{F}_{n-1} \right) + U_{n-1}^2 \mathbb{E} \left(\Delta_n^2(x) \mid \mathcal{F}_{n-1} \right). \end{aligned} \tag{41}$$

Consequently, Equations (39)–(41) allow to write

$$B_n = 8B_{n1} + \left[\mathcal{O}\left(h_n^{2\beta}\right) + \mathcal{O}\left(h_n^\beta\right) + U_{n-1}^2 + C \right] \mathbb{E} \left(\Delta_n^2(x) \mid \mathcal{F}_{n-1} \right). \tag{42}$$

Now, we replace, in the decomposition (37), A_n and B_n by their expressions, given by (38) and (42), and one obtains

$$\begin{aligned}
\mathbb{E} \left(U_n^2 | \mathcal{F}_{n-1} \right) &= U_{n-1}^2 + \frac{2\gamma_n U_{n-1} A_n}{F_x(h_n)} + \left(\frac{\gamma_n}{F_x(h_n)} \right)^2 B_n \\
&\leq U_{n-1}^2 + 2\gamma_n U_{n-1} \left\{ \frac{\left[-U_{n-1} + \mathcal{O}(h_n^\beta) \right] \mathbb{E} \left(\Delta_n(x) | \mathcal{F}_{n-1} \right) + \frac{A_{n1}}{F_x(h_n)}}{F_x(h_n)} \right\} \\
&\quad + \gamma_n^2 \left\{ \frac{B_{n1}}{F_x^2(h_n)} + \frac{\left[\mathcal{O}(h_n^{2\beta}) + \mathcal{O}(h_n^\beta) + U_{n-1}^2 + C \right] \mathbb{E} \left(\Delta_n^2(x) | \mathcal{F}_{n-1} \right)}{F_x^2(h_n)} \right\} \\
&= U_{n-1}^2 \left\{ 1 - 2\gamma_n \frac{\mathbb{E} \left(\Delta_n(x) | \mathcal{F}_{n-1} \right)}{F_x(h_n)} + \gamma_n^2 \frac{\mathbb{E} \left(\Delta_n^2(x) | \mathcal{F}_{n-1} \right)}{F_x^2(h_n)} \right\} \\
&\quad + U_{n-1} \gamma_n \left\{ \mathcal{O}(h_n^\beta) \frac{\mathbb{E} \left(\Delta_n(x) | \mathcal{F}_{n-1} \right)}{F_x(h_n)} + \frac{A_{n1}}{F_x(h_n)} \right\} \\
&\quad + \gamma_n^2 \left\{ \frac{B_{n1}}{F_x^2(h_n)} + \frac{\left[\mathcal{O}(h_n^{2\beta}) + \mathcal{O}(h_n^\beta) + C \right] \mathbb{E} \left(\Delta_n^2(x) | \mathcal{F}_{n-1} \right)}{F_x^2(h_n)} \right\}. \quad (43)
\end{aligned}$$

Let us set

$$q_n = \left[\frac{\mathcal{O}(h_n^\beta) \mathbb{E} \left(\Delta_n(x) | \mathcal{F}_{n-1} \right) + A_{n1}}{F_x(h_n)} \right] q_0^{-1}, \quad q_0 \in (0, 1).$$

Because $U_{n-1} \leq \frac{1}{2}(q_n + (U_{n-1}^2/q_n))$, one gets

$$U_{n-1} \leq \frac{1}{2} \left\{ \frac{\mathcal{O}(h_n^\beta) \mathbb{E} \left(\Delta_n(x) | \mathcal{F}_{n-1} \right) + A_{n1}}{F_x(h_n) q_0} + \frac{F_x(h_n) q_0 U_{n-1}^2}{\mathcal{O}(h_n^\beta) \mathbb{E} \left(\Delta_n(x) | \mathcal{F}_{n-1} \right) + A_{n1}} \right\}. \quad (44)$$

By replacing (44) in (43), one obtains

$$\begin{aligned}
\mathbb{E} \left(U_n^2 | \mathcal{F}_{n-1} \right) &\leq U_{n-1}^2 \left\{ 1 - 2\gamma_n \frac{\mathbb{E} \left(\Delta_n(x) | \mathcal{F}_{n-1} \right)}{F_x(h_n)} + \gamma_n^2 \frac{\mathbb{E} \left(\Delta_n^2(x) | \mathcal{F}_{n-1} \right)}{F_x^2(h_n)} + \frac{\gamma_n q_0}{2} \right\} \\
&\quad + \frac{\gamma_n}{2q_0} \left\{ \frac{\mathcal{O}(h_n^\beta) \mathbb{E} \left(\Delta_n(x) | \mathcal{F}_{n-1} \right)}{F_x(h_n)} + \frac{A_{n1}}{F_x(h_n)} \right\}^2 \\
&\quad + \gamma_n^2 \left\{ \frac{B_{n1}}{F_x^2(h_n)} + \frac{\left[\mathcal{O}(h_n^{2\beta}) + \mathcal{O}(h_n^\beta) + C \right] \mathbb{E} \left(\Delta_n^2(x) | \mathcal{F}_{n-1} \right)}{F_x^2(h_n)} \right\}.
\end{aligned}$$

Now, observe that Lemma A.0.1 allows us to bound the following quantities:

$$\frac{\mathbb{E} \left(\Delta_n(x) \middle| \mathcal{F}_{n-1} \right)}{F_x(h_n)} \leq C^*, \quad \frac{\mathbb{E} \left(\Delta_n^2(x) \middle| \mathcal{F}_{n-1} \right)}{F_x^2(h_n)} \leq C^* \phi^{-1}(h_n) \text{ and } \frac{\mathbb{E}^2 \left(\Delta_n(x) \middle| \mathcal{F}_{n-1} \right)}{F_x(h_n)} \leq C^*,$$

where C is a positive constant independent on x . Then, one can easily get

$$\begin{aligned} \mathbb{E} \left(U_n^2 \middle| \mathcal{F}_{n-1} \right) &\leq U_{n-1}^2 \left\{ 1 - \gamma_n \left[2C^* - \frac{q_0}{2} - C^* \gamma_n \phi^{-1}(h_n) \right] \right\} \\ &\quad + \gamma_n \left[\mathcal{O}(h_n^{2\beta}) + \mathcal{O}(\phi^{-2}(h_n)) A_{n1}^2 \right] \\ &\quad + \gamma_n^2 \left[\mathcal{O}(\phi^{-2}(h_n)) B_{n1} + \mathcal{O}(h_n^{2\beta} \phi^{-1}(h_n)) \right. \\ &\quad \left. + \mathcal{O}(h_n^\beta \phi^{-1}(h_n)) + \mathcal{O}(\phi^{-1}(h_n)) \right]. \end{aligned}$$

Let us apply the expectation on both sides of the previous inequality and denote $\mathcal{K}_n(x) := \mathbb{E}(U_n^2)$. Then, since it is supposed that $\gamma_n \phi^{-1}(h_n) \rightarrow 0$, we obtain

$$\begin{aligned} \mathcal{K}_n(x) &\leq \mathcal{K}_{n-1}(x) \left\{ 1 - \gamma_n \left[2C^* - \frac{q_0}{2} \right] \right\} + \mathcal{O}(\gamma_n \phi^{-2}(h_n)) \mathbb{E}(A_{n1}^2) \\ &\quad + \mathcal{O}(\gamma_n^2 \phi^{-2}(h_n)) \mathbb{E}(B_{n1}) + \mathcal{O}(\gamma_n h_n^{2\beta}) \\ &\quad + \mathcal{O}(h_n^{2\beta} \gamma_n^2 \phi^{-1}(h_n)) + \mathcal{O}(h_n^\beta \gamma_n^2 \phi^{-1}(h_n)) + \mathcal{O}(\gamma_n^2 \phi^{-1}(h_n)). \end{aligned} \quad (45)$$

Now, to achieve the proof of this theorem, we need to control, asymptotically, the quantities

$$\mathbb{E}(A_{n1}^2) := \mathbb{E} \left\{ \mathbb{E}^2 \left[\left(r(X_n) - r_{n-1}(X_n) \right)^2 \Delta_n(x) \middle| \mathcal{F}_{n-1} \right] \right\}$$

and

$$\mathbb{E}(B_{n1}) := \mathbb{E} \left\{ \mathbb{E} \left[\left(r(X_n) - r_{n-1}(X_n) \right)^4 \Delta_n^2(x) \middle| \mathcal{F}_{n-1} \right] \right\}.$$

To this end, observe that by the Cauchy-Schwarz inequality and the properties of the conditional expectation, one can write

$$\begin{aligned} \mathbb{E}(A_{n1}^2) &\leq \mathbb{E}(B_{n1}) = \mathbb{E} \left\{ \mathbb{E} \left[\left(r_{n-1}(X_n) - r(X_n) \right)^4 \Delta_n^2(x) \middle| \mathcal{F}_{n-1} \right] \right\} \\ &= \mathbb{E} \left[\left(r_{n-1}(X_n) - r(X_n) \right)^4 \Delta_n^2(x) \right] \\ &\leq \left\{ \sup_{z \in \mathcal{B}(x, h_n)} \mathbb{E} \left[r_{n-1}(z) - r(z) \right]^8 \right\}^{1/2} \times \mathbb{E}^{1/2} \left(\Delta_n^4(x) \right). \end{aligned} \quad (46)$$

Theorem 3.1 insures that

$$\sup_{z \in \mathcal{B}(x, h_n)} \mathbb{E} \left[r_{n-1}(z) - r(z) \right]^8 \rightarrow 0 \text{ as } n \rightarrow \infty$$

and, by Lemma A.0.1 and assumption (H2), we have

$$\mathbb{E} \left(\Delta_n^4(x) \right) = \mathcal{O} \left(\phi(h_n) \right).$$

Then we conclude that $\mathbb{E}(A_{n1}^2) \rightarrow 0$ and $\mathbb{E}(B_{n1}) \rightarrow 0$ as $n \rightarrow \infty$. Finally, under conditions (16), and by making use Lemma A.0.2, one gets $\mathcal{K}_n(x) \rightarrow 0$ as $n \rightarrow \infty$.

Next, to provide the convergence result with rate, stated in (18), observe that if $h_n = h_0 n^{-1/2\beta+\nu}$, then similar to (15), we have that:

$$\mathbb{E} \left(A_{n1}^2 \right) \leq \mathbb{E} \left(B_{n1} \right) = \mathcal{O} \left(\phi(h_n) n^{-\frac{4\alpha}{2\beta+\nu}} \right),$$

which together with (46) and (17) imply that for n large and q_0 small enough

$$\begin{aligned} \mathcal{K}_n(x) &\leq \mathcal{K}_{n-1}(x) \left(1 - 2C^* \gamma_n \right) + \mathcal{O} \left(\gamma_n n^{-(4\alpha/(2\beta+\nu))} \phi^{-3/2}(h_n) \right) \\ &\quad + \mathcal{O} \left(\gamma_n h_n^{2\beta} \right) + \mathcal{O} \left(\gamma_n^2 \phi^{-1}(h_n) \right) \\ &\leq \mathcal{K}_{n-1}(x) \left(1 - \frac{2C^* \gamma_0}{n} \right) + \frac{D}{n} \left[n^{(-8\alpha+3\nu)/(2(2\beta+\nu))} + n^{-(2\beta/(2\beta+\nu))} \right]. \\ &\leq \mathcal{K}_{n-1}(x) \left(1 - \frac{2C^* \gamma_0}{n} \right) + \frac{D}{n^{1+\min\{(8\alpha-3\nu)/(2(2\beta+\nu)), (2\beta/(2\beta+\nu))\}}}. \end{aligned} \quad (47)$$

where D is a positive constant. The application of Lemma A.0.3 conducts to the desired result. ■

Proof of Corollary 3.2.1: One can easily prove the result of the corollary by following similar steps as those in Theorem 3.2. ■

Notes

1. Available on the website: www.lsp.ups-tlse.fr/staph/npfda
2. <https://cran.r-project.org/web/packages/aire.zmvm/aire.zmvm.pdf>
3. The characteristics of the computer we used to perform these forecasts were CPU: Duo E4700 2.60 GHz, HD: 149 Go, Memory: 3.23 Go.

Disclosure statement

No potential conflict of interest was reported by the author(s).

References

- [1] Bosq D. Nonparametric statistics for stochastic processes. Estimation and prediction. New York: Springer-Verlag; 1998. (Lecture notes in statistics; 110).

- [2] Collomb G, Härdle W. Strong uniform convergence rates in robust nonparametric time series analysis and prediction: kernel regression estimation from dependent observations. *Stochastic Process Appl.* **1986**;23(1):77–89.
- [3] Rosenblatt M. Conditional probability density and regression estimators. *Multivariate Analysis, II. Proceeding of International Symposium*; Dayton, OH; New York: Academic Press; 1969. p. 25–31.
- [4] Stone CJ. Optimal global rates of convergence for nonparametric regression. *Ann Statist.* **1982**;10(4):1040–1053.
- [5] Ramsay JO, Silverman BW. *Functional data analysis*. New York: Springer Series in Statistics; **2005**.
- [6] Ferraty F, Vieu P. *Nonparametric modelling for functional data, methods, theory, applications and implementations*. London: Springer-Verlag; **2006**.
- [7] Ferraty F, Romain Y. *The Oxford handbook of functional data analysis*. Oxford: Oxford University Press; **2011**.
- [8] Hu Y. Nonparametric estimation of variance function for functional data under mixing conditions. *Comm. Statistics – Theory and Methods.* **2013**;42(10):1774–1786.
- [9] Ferraty F, Mas A, Vieu P. Nonparametric regression on functional data: inference and practical aspects. *Aust N Z J Stat.* **2007**;49(3):267–286.
- [10] Ferraty F, Vieu P. Nonparametric models for functional data, with applications in regression, time series prediction and curve discrimination. *J Nonparametr Stat.* **2004**;16:111–125. *The International Conference on Recent Trends and Directions in Nonparametric Statistics.*
- [11] Masry E. Nonparametric regression estimation for dependent functional data: asymptotic normality. *Stochastic Process Appl.* **2005**;115:155–177.
- [12] Rosenblatt M. Uniform ergodicity and strong mixing. *Z Wahrs Verw Gebiete.* **1972**;24:79–84.
- [13] Gastwirth JL, Rubin H. The asymptotic distribution theory of the empiric cdf for mixing stochastic processes. *Ann Statist.* **1975**;3(4):809–824.
- [14] Andrews DWK. Non-mixing autoregressive processes. *J Appl Probab.* **1984**;21:930–934.
- [15] Chernick MR. A limit theorem for the maximum of autoregressive processes with uniform marginal distributions. *Ann Probab.* **1981**;9:145–149.
- [16] Pham TD, Ran LT. Some mixing properties of time series models. *Stochastic Process Appl.* **1985**;19:297–303.
- [17] Laïb N, Louani D. Nonparametric kernel regression estimation for functional stationary ergodic data: asymptotic properties. *J Multivariate Anal.* **2010**;101(10):2266–2281.
- [18] Laïb N, Louani D. Rates of strong consistencies of the regression function estimator for functional stationary ergodic data. *J Statist Plann Inference.* **2011**;141(1):359–372.
- [19] Bouzebda S, Chaouch M, Laïb N. Limiting law results for a class of conditional mode estimates for functional stationary ergodic data. *Math Meth Stat.* **2016**; 25(3):168–195.
- [20] Delecroix M. *Sur l'estimation et la prévision nonparamétrique des processus ergodiques*. Doctorat d'état, Université des Sciences et Techniques de Lille Flandre-Artois; 1987.
- [21] Krengel U. *Ergodic theorms*. Berlin: Walter de Gruyter & Co; **1985**.
- [22] Ling N, Liang L, Vieu P. Nonparametric regression estimation for functional stationary ergodic data with missing at random. *J Statist Plann Inference.* **2015**; 162: 75–87.
- [23] Cai TT, Wang L. Adaptive variance function estimation in heteroscedastic nonparametric regression. *Ann Statist.* **2008**;36(5):2025–2054.
- [24] Fan JQ, Yao Q. Efficient estimation of conditional variance functions in stochastic regression. *Biometrika.* **1998**;85(3):645–660.
- [25] Laïb N. Kernel estimates of the mean and the volatility functions in a nonlinear autoregressive model with ARCH errors. *J Statist Plann Inference.* **2005**;134:116–139.
- [26] Hall P, Carroll RJ. Variance function estimation in regression, the effect of estimating the mean. *J Roy Statist Soc Ser B-Methodol.* **1989**;51(1):3–14.
- [27] Chaouch M. Volatility estimation in a nonlinear heteroscedastic functional regression model with martingale difference errors. *J Multivariate Anal.* **2019**;170:129–148.
- [28] Stadtmüller U, Tsybakov AB. Nonparametric recursive variance estimation. *Statistics.* **1995**;27:55–63.

- [29] Robbins R, Monro SA. A stochastic approximation method. *Ann Statist.* 1951;22(3):400–407.
- [30] Révész P. Robbins-Monro procedure in a Hilbert space and its application in the theory of learning processes I. *Studia Sci Math Hungar.* 1973;8:391–398.
- [31] Amiri A. Recursive regression estimators with application to nonparametric prediction. *J Nonparam Stat.* 2012;24(1):169–186.
- [32] Györfi L, Kholer M, Krzyzak A, Walk H. A distribution-free theory of nonparametric regression. New York: Springer-Verlag; 2002.
- [33] Wang L, Liang HY. Strong uniform convergence of the recursive regression estimator under φ -mixing conditions. *Metrika.* 2004;59(3):245–261.
- [34] Amiri A, Crambes C, Thiam B. Recursive estimation of nonparametric regression with functional covariate. *Comp Stat and Data Anal.* 2014;69:154–172.
- [35] Amiri A, Thiam B. Consistency of the recursive nonparametric regression estimation for dependent functional data. *J Nonparam Statist.* 2014;26(13):471–487.
- [36] Rachdi M, Vieu Ph. Nonparametric regression for functional data: automatic smoothing parameter selection. *J Statist Plann Inference.* 2017;137:2784–2801.
- [37] Polyak BT. Introduction to optimization. New York: Optim. Software; 1987.

Appendix

We recall the following technical lemma which is proved by Laïb and Louani [17].

Lemma A.0.1 (see [17]): *Assume that conditions (H1), (H2)(i), (H2)(ii) and (H2)(iii) hold true. For any real numbers $j \geq 1$ and $1 \leq k \leq 2 + \delta$ with $\delta > 0$, as $n \rightarrow \infty$, and for any fixed $i \in 1, \dots, n$, we have*

- (i) $1/\phi(h)\mathbb{E}[\Delta_i^j(x) \mid \mathcal{F}_{i-1}] = M_{jf_{i,1}}(x) + \mathcal{O}_{a.s.}(g_{i,x}(h)/\phi(h))$.
- (ii) $1/\phi(h)\mathbb{E}[\Delta_i^j(x)] = M_{jf_1}(x) + o(1)$.
- (iii) $1/\phi^k(h)[\mathbb{E}(\Delta_1(x))]^k = M_{1f_1^k}(x) + o(1)$.

Next, let us introduce the following Lemmas proved in [37] which play a key role in our proofs.

Lemma A.0.2 (see [37]): *Let $u_{k+1} \leq q_k u_k + \alpha_k$, $0 \leq q_k < 1$, $\alpha_k \geq 0$, $\sum_{k=0}^{\infty} (1 - q_k) = \infty$, $\alpha_k / (1 - q_k) \rightarrow 0$. Then $\lim_{k \rightarrow \infty} u_k \leq 0$. In particular if $u_k \geq 0$, then $u_k \rightarrow 0$.*

Lemma A.0.3 (Chung's Lemma, [37]):

Let $u_k \geq 0$ and

$$u_{k+1} \leq \left(1 - \frac{c}{k}\right) u_k + \frac{d}{k^{p+1}}, \quad d > 0, \quad p > 0, \quad c > 0.$$

Then,

$$u_k \leq d(c - p)^{-1} k^{-p} + o(k^{-p}), \quad \text{for } c > p.$$