Generalized regression estimation for continuous time processes with values in functional spaces

Bertrand Maillot, Christophe Chesneau

Abstract. We consider two continuous time processes; the first one is valued in a semi-metric space, while the second one is real-valued. In some sense, we extend the results of F. Ferraty and P. Vieu in "Nonparametric models for functional data, with application in regression, time-series prediction and curve discrimination" (2004), by establishing the convergence, with rates, of the generalized regression function when a real-valued continuous time response is considered. As corollaries, we deduce the convergence of the conditional distribution function as well as conditional quantiles. Note that a parametric rate of convergence in probability is reached while working with a naive kernel.

Keywords: continuous time process; regression function estimation; conditional distribution function

Classification: 62G07, 62C05, 62E20

1. Introduction

Prediction of variables is an important problem in statistics and particularly in time series analysis. To this aim, the estimation of the conditional expectation, which is not only useful for prediction, is a relevant solution which has been largely studied for scalar and vectorial regressors and for discrete or continuous time processes, see, for example, [28], [25], [27], [8], [21], [26] and [1]. Recently, functional data have been more and more studied because of the wide fields of application. The interest for functional data is not new, but authors first used discretization of the functional data, see [17] for a survey. Then, the functional nature of the data have been considered with models such as the linear functional models, see [23], [5], [6] and the references therein, and later on, general nonparametric approaches have been developed, see [12], [22] and [16]. An important application is the prediction of a continuous time process, but all the works performed on the subject interpreted the continuous time process as a functional variable observed at discrete time, see, e.g., [3].

DOI 10.14712/1213-7243.2022.003

In this paper, we consider a function of a real valued stochastic variable and an explanatory variable taking values in a semi-metric vectorial space, both observed at a continuous time. Functional data that are observables at continuous time are frequent in practice. As an example, we quote the isotherm or isobar curves in meteorology. Our aim in this paper is to study the convergence, with rates, of the generalized regression function estimator and to deduce corollaries on the convergence of the conditional distribution function and the conditional quantile function estimators. Under stronger conditions on the dependence structure, we reach the parametric rate of convergence in probability while working with a naive kernel. This rate of convergence has been already obtained by [7] in the nonparametric statistical framework.

The contents of the paper are organized as follows: we present the statistical framework and hypotheses in Section 2. The results are provided in Section 3, with discussions. Some examples of applications are developed in Section 4. Section 5 is devoted to the proofs. A summary section is given in Section 6.

2. Statistical framework and assumptions

This section presents the overall mathematical setting of the study.

2.1 Statistical framework. Let $\{X_t, Y_t\}_{t\in\mathbb{R}^+}$ be a continuous time process defined on a probability space $(\Omega, \mathcal{F}, P)$ and observed for $t \in [0, T]$, where Y_t is real valued and X_t takes values in a semi-metric vectorial space $\mathcal{H}$ equipped with the semi-metric $d(\cdot,\cdot)$. We suppose that the law of (X_t, Y_t) does not depend on t and that there exists a regular version of the conditional probability distribution of Y_t given X_t, see [19], [20] and [18] for conditions giving the existence of the conditional probability. Throughout this paper, $\mathcal{C}$ denotes a compact set of $\mathcal{H}$ and $\mathcal{S}$ is a compact of $\mathbb{R}$. Let Ψ be a real valued Borel function defined on $\mathcal{S} \times \mathbb{R}$ and consider for any $y \in \mathcal{S}$ the generalized regression function, supposed to exist for any $x \in \mathcal{C}$, and defined by

$$r(x, y) = E(\Psi(y, Y_0)|X_0 = x).$$

In the sequel, we use a positive bounded kernel K with support $[0, 1]$, a bandwidth h_T decreasing to 0 and define the generalized regression function estimate by

$$\widehat{r}_T(x, y) = \frac{\int_{t=0}^{T} \Psi(y, Y_t) K(h_T^{-1} d(x, X_t))\, \mathrm{d}t}{\int_{t=0}^{T} K(h_T^{-1} d(x, X_t))\, \mathrm{d}t}$$

when the denominator is not null, otherwise, we take

$$\widehat{r}_T(x,y) = \frac{\int_{t=0}^{T} \Psi(y, Y_t)\,\mathrm{d}t}{T}.$$

2.2 Assumptions. For $x \in \mathcal{H}$ and $h > 0$, we denote by $\mathcal{B}(x,h)$ the ball of center x and radius h. In the following, for a real variable Z we use the notation $\|Z\|_p := \big(E(|Z|^p)\big)^{1/p}$ and $\|Z\|_\infty := \sup\{x \in \mathbb{R}\colon P(Z > x) > 0\}$.

We first introduce a local Hölderian condition on the generalized regression function r which is needed for all our results.

There exist three constants $c_1 > 0$, $C < \infty$, and $\eta > 0$, such that for any $x \in \mathcal{C}$, any $(u,v) \in \mathcal{B}(x,c_1)^2$ and any $y \in \mathcal{S}$

$$|r(u,y) - r(v,y)| \le C d(u,v)^\eta. \tag{1}$$

Since most of the useful properties related to the process under study are needed only locally, the constant c_1 introduced above is used below to set several hypotheses. As usual, in order to obtain uniform results, we need an hypothesis on the compact $\mathcal{C}$. We suppose that there exists two positive constants C_1 and d_1 such that

$$\forall\, \nu \in]0,1[\,,\ \mathcal{C} \text{ can be covered by } L_\nu \le \frac{C_1}{\nu^{d_1}} \text{ balls of radius } \nu. \tag{2}$$

Since the notion of density is not so natural for semi-metric-space valued random variable as compared to the random vector case (even if it can be defined and estimated for functional variables, see [9] or [10]), an hypothesis upon the probability of small balls is fundamental to establish properties and results. Throughout this paper, we assume that there exists a function φ and two constants $(\beta_1, \beta_2) \in \mathbb{R}_+^2$ such that for any $x \in \mathcal{C}$ and any $h \in]0, c_1]$,

$$0 < \beta_1 \varphi(h) \le P(X_0 \in \mathcal{B}(x,h)) \le \beta_2 \varphi(h). \tag{3}$$

In the following set of hypotheses on the distribution of the processes for a set $\mathcal{A}$, $\mathbb{1}_{\mathcal{A}}$ is the indicator function.

(D.1) (a) There exists $M_1 > 0$ such that for any $(s,t) \in \mathbb{R}_+^2$, $x \in \mathcal{C}$, and $y \in \mathcal{S}$, we have $E(|\Psi(y,Y_s)\Psi(y,Y_t)| \mathbb{1}_{\mathcal{B}(x,c_1)^2}(X_s,X_t) | X_s, X_t) \le M_1$.

(b) There exist $p \in]4, \infty]$ and $Q > 0$, such that for any $n \in \mathbb{N}$, and any $u > 0$, $P(\sup_{t \in [n,n+1[} \sup_{y \in \mathcal{S}} |\Psi(y,Y_t)| > u) \le Q u^{-p}$ and $\|\sup_{y \in \mathcal{S}} \Psi(y, Y_0)\|_p < \infty$.

(c) There exist $c_2 > 0$, $\eta_1 > 1/2$, $\gamma > 0$, and a constant p' such that for any $y \in \mathcal{S}$ and $u \le c_2$, $\|\sup_{y' \in [y-u,y+u] \cap \mathcal{S}} \Psi(y,Y_t) - \Psi(y',Y_t)\|_{p'} \le \gamma u^{\eta_1}$. And one of the two following conditions is fulfilled:

($\star$) $p' > 2$.

($\star\star$) For any $y \in \mathbb{R}$, $\Psi(\cdot, y)$ is an increasing (or decreasing) function and $p' = \eta_1 = 1$.

(d) There exist $\beta_3 > 0$ and $\delta > 0$ such that for any $(s,t) \in \mathbb{R}_+^2$ with $|s-t| > \delta$ for any $x \in \mathcal{C}$ and any $h \in]0, c_1]$, $P(X_s \in \mathcal{B}(x,h), X_t \in \mathcal{B}(x,h)) \leq \beta_3 \varphi(h)^2$.

(e) The process $(X_t, Y_t)_{t \geq 0}$ is α-mixing, see [4] for a presentation of mixing conditions, and mixing coefficient verifies $\alpha(u) \leq cu^{-a}$ where $c \geq 1$, and $a > \max(3p/(p-4), p'/(p'-2))$.

(K) The positive kernel K, with support [0,1], is differentiable and has a bounded derivative.

Finally, we impose the following condition upon the bandwidth:

(H) (a) For any $n \in \mathbb{N}$, the bandwidth h_T is continuous and differentiable on the interval $[n, n+1[$, where

$$\left|\frac{\frac{\mathrm{d}h_T}{\mathrm{d}T}}{h_T}\right| = \left|\frac{h_T'}{h_T}\right| = \mathcal{O}\Big(\frac{1}{T}\Big) \qquad \text{and} \qquad \frac{\varphi(h_T)T}{\ln(T)}$$

is increasing.

(b) We choose h_T such that

$$\varphi(h_T)T > \ln(T)^{\xi}\Big(\frac{T^{d_1+p_1+2}}{h_T^{d_1}}\Big)^{2(a+p)/((a+1)p)}$$

where p, a and d_1 are defined in the conditions (D.1) (b), (D.1) (e) and (2), respectively, $\xi > 1$ and $p_1 > 2/(2\eta_1 - 1)$.

Let us now discuss the hypotheses above. Usually, while studying the regression function in continuous time framework, authors assume that the process of interest is bounded, see [2]. Here, we consider hypotheses (D.1) (a)–(b) to control the covariance terms and to avoid imposing the process to be bounded when we handle the discretization phase.

Hypothesis (D.1) (c) is a regularity condition upon the function Ψ. It is necessary to obtain a uniform result over the compact $\mathcal{S}$.

The condition (D.1) (d) introduces a constraint on the joint distribution of the process X when considered in small balls.

In the condition (D.1) (e), we impose a polynomial decreasing mixing coefficient which is less restricting than the geometrically decreasing one.

The condition (K) is very standard in nonparametric function estimation.

The constraints with respect to the bandwidth h_T along with its relationship to the function φ are considered in the conditions (H) (a)–(b). As compared to the vectorial process case, the hypothesis (H) (a) seems to be natural, see [2]. The condition (H) (b) allows us to have a small enough variance of the generalized regression estimator.

3. Results

We are now in position to present the main results of the study, beginning with the convergence of $\widehat{r}_T(x,y)$.

Theorem 3.1. *Under the conditions* (1), (2), (3), (D.1), (H) *and* (K) *there exists a constant* $L > 0$ *such that*

$$\limsup_{T\to\infty} \sup_{y\in\mathcal{S}} \sup_{x\in\mathcal{C}} \frac{|\widehat{r}_T(x,y) - r(x,y)|}{h_T^\eta + \sqrt{\ln(T)/(T\varphi(h_T))}} \le L \qquad \text{a.s.} \tag{4}$$

The paper [13] obtained the convergence of the regression function estimator when both the response variable and the explanatory one are functional while working with discrete and independent variables: we extended their results to (continuous) α-mixing processes, even if the form of our response variable is a little less general than theirs, which takes values in a general abstract Banach space. When $\mathcal{S}$ is reduced to one point and the response is a scalar, our result can be compared to the discrete time version obtained by F. Ferraty and P. Vieu, see Theorem 3.1 in [15]. We can note that an alternative method of prediction has been studied for α-mixing processes with a functional valued predictor: [11] investigated the conditional mode estimation.

We now stress our attention on the conditional distribution function

$$F(y|x) = P\big(Y_0 \le y | X_0 = x\big) \qquad \text{where } \ y \in \mathbb{R} \ \text{ and } \ x \in \mathcal{C}$$

which can be estimated by

$$\widehat{F}_T(y|x) := \widehat{r}_T(x,y) \qquad \text{with } \ \Psi(y,Y_t) = \mathbb{1}_{]-\infty,y]}(Y_t). \tag{5}$$

Corollary 3.1. *Under the conditions of Theorem* 3.1, *there exists a constant* $L_1 > 0$ *such that*

$$\limsup_{T\to\infty} \sup_{y\in\mathcal{S}} \sup_{x\in\mathcal{C}} \frac{|\widehat{F}_T(y|x) - F(y|x)|}{h_T^\eta + \sqrt{\ln(T)/(T\varphi(h_T))}} < L_1 \qquad \text{a.s.} \tag{6}$$

When $\mathcal{C}$ is reduced to one point, [14] obtained a similar result while considering a smoothed estimator of the conditional distribution function in the discrete time process case.

An additional condition is required to define the conditional quantile and to state the corollary involving the conditional quantiles estimator.

(D.2) Suppose that there exists $\eta_2 \geq 1$, $C' > 0$ and $u_0 < 1$ such that for any $x \in \mathcal{C}$ and any $(u, u') \in \mathcal{S}^2$ with $|u - u'| < u_0$, $F(.|x)$ is continuous and we have $|F(u|x) - F(u'|x)| > C'|u - u'|^{\eta_2}$.

For θ in $]0, 1[$, define the conditional quantile $u_\theta(x)$ by the following equation

$$F(u_\theta(x)|x) = \theta.$$

The conditional quantile estimator is then defined by

$$\widehat{u_\theta(x)} = \sup_{u \in \mathbb{R}}(u, \widehat{F}(u|x) < \theta). \tag{7}$$

Corollary 3.2. *Assume that condition* (D.2) *holds with* $\mathcal{S} = [a_1, b_1]$. *Under conditions of Theorem* 3.1, *whenever* $a_1 < \inf_{x \in \mathcal{C}} u_{\theta_1}(x) \leq \sup_{x \in \mathcal{C}} u_{\theta_2}(x) < b_1$ *for some* $(\theta_1, \theta_2) \in]0, 1[$, *there exists a constant* L_2 *such that,*

$$\limsup_{T \to \infty} \sup_{\theta \in [\theta_1, \theta_2]} \sup_{x \in \mathcal{C}} \frac{\left|\widehat{u_\theta(x)} - u_\theta(x)\right|}{\left(h_T^\eta + \sqrt{\ln(T)/(T\varphi(h_T))}\right)^{1/\eta_2}} \leq L_2 \qquad \text{a.s.} \tag{8}$$

Remark 3.1. When $\theta_1 = \theta_2$ and $\mathcal{C} = x$, the paper [14] has already obtained the convergence of the quantile estimator for discrete time processes. Here, we do not have to suppose that F is differentiable and Corollary 3.2 is still valid when η_2 is not an integer.

In order to present the conditions under which the parametric rate is reached, we need to introduce the following notation:

$$\text{For any } \; t \geq 0 \qquad \varepsilon_t(y) := \Psi(y, Y_t) - E(\Psi(y, Y_t)|X_t). \tag{9}$$

While working with random vectors, we usually impose conditions on the joint density to obtain the parametric rate. Hereafter, we express the same kind of dependence structure with the probability of small balls.

(D.3) (a) There exists $c_2 > 0$ and a function g_0 integrable on $]0, \infty[$ such that for any $x \in \mathcal{C}$, any $s > t \geq 0$, and any $h \in [0, c_2]$, $|P((X_t, X_s) \in \mathcal{B}(x, h)^2) - P(X_t \in \mathcal{B}(x, h))^2| \leq g_0(s - t)\varphi(h)^2$.

(b) There exists an integrable function g_1 on $[0, \infty[$ such that for any $y \in \mathcal{S}$ and any $(s, t) \in \mathbb{R}+^2$

$$\max\{|E(\varepsilon_s(y)|X_s, X_t)|, |E(\varepsilon_s(y)\varepsilon_t(y)|X_s, X_t)|\} \leq g_1(|s - t|).$$

Now, for any integer $m > 1$, set $\ln_m(\cdot) := \ln(|\ln_{m-1}(\cdot)|)$ with $\ln_1(\cdot) = \ln(\cdot)$. Theorem 3.2 and Corollaries 3.3–3.4 give parametric rates of convergence in probability for the estimators defined above.

Theorem 3.2. *Under* (1), (2), (3) *and* (D.3) *if we choose* $K = \mathbb{I}_{[0,1]}$ *and* $h_T = T^{-\ln_2(\lfloor T \rfloor)}$ *then for any integer* $m \geq 1$, *any* $x \in \mathcal{C}$ *and* $y \in \mathcal{S}$, *we have*

$$\lim_{T\to\infty} \sqrt{\frac{T}{\ln_m(T)}}\, |\widehat{r}_T(x,y) - r(x,y)| \xrightarrow{p} 0. \tag{10}$$

The following corollaries are direct consequences of Theorem 3.2.

Corollary 3.3. *Under the conditions of Theorem* 3.2 *for any integer* $m \geq 1$, *any* $x \in \mathcal{C}$ *and any* $y \in \mathcal{S}$, *we have*

$$\lim_{T\to\infty} \sqrt{\frac{T}{\ln_m(T)}}\, \left|\widehat{F}_T(y|x) - F(y|x)\right| \xrightarrow{p} 0. \tag{11}$$

Corollary 3.4. *Let* $\mathcal{S} = [a_1, b_1]$. *Assume that conditions of Theorem* 3.2 *and the hypothesis* (D.2) *hold. If* $a_1 < \inf_{x\in\mathcal{C}} u_{\theta_1}(x) \leq \sup_{x\in\mathcal{C}} u_{\theta_2}(x) < b_1$ *for some* $(\theta_1, \theta_2) \in]0,1[$, *then for any integer* $m \geq 1$, *any* $x \in \mathcal{C}$, *any* $y \in \mathcal{S}$, *and any* $\theta \in [\theta_1, \theta_2]$, *we have*

$$\lim_{T\to\infty} \left(\sqrt{\frac{T}{\ln_m(T)}}\right)^{1/\eta_2} \left|\widehat{u_\theta(x)} - u_\theta(x)\right| \xrightarrow{p} 0. \tag{12}$$

4. Examples

Before anything else, since we do not impose a stationarity hypothesis, let us point out that, excepted the parametric rates, our results remain valid in the discrete time setting. In fact, it suffices to take $(X_t, Y_t) := (U_i, V_i)$ whenever $t \in [i, i+1[$, where (U_i, V_i) is the discrete time process of interest. Therefore, examples given in [15] still hold in the framework of this paper. Moreover, taking the abstract space as $\mathbb{R}^n$ with $n \in \mathbb{N} \setminus 0$, our results are also valid for vectorial processes.

The rest of this section is divided into several examples related to our main results, namely functional parametric rate, time series prediction and autoregressive process.

4.1 Functional parametric rate. At first, we give an example of functional valued continuous time process that satisfies the main condition necessary to obtain the parametric rate, that is the condition (D.3) (a). Let Υ be an unknown real valued function defined on $\mathbb{R}^n \times [0,1]$ and let $(Z_t)_{t\in\mathbb{R}^+}$ be a non observed $\mathbb{R}^n$-valued process fulfilling condition H in [2, page 116]. Now, denote by f the bounded density function of the random variable Z_t, and by $f_{s,t}$ the joint density function of the random vector (Z_s, Z_t). For any $(x,y) \in \mathbb{R}^{2n}$, set $g_{s,t}(x,y) :=$

$f_{s,t}(x,y) - f(x)f(y)$. Let $(\mathcal{E}, d)$ be the semi-metric space of real valued functions defined on $[0,1]$, with the underlying semi-metric $d(\cdot,\cdot)$. For any $t \in [0,T]$, the observed process is given by $X_t := (\Upsilon(Z_t,u))_{u\in[0,1]}$. It is obvious that for any $x \in \mathbb{R}^n$, $\Upsilon(x,\cdot) \in \mathcal{E}$ and so, (X_t) is a functional valued continuous time process. We define the function υ: $x \mapsto \Upsilon(x,\cdot)$. Let $\mathcal{C}$ be a compact set of $\mathbb{R}^n$. Suppose that for any $x \in \mathcal{C}$ and any $x_0 \in \mathbb{R}^n$ such that $d(\Upsilon(x,\cdot),\Upsilon(x_0,\cdot)) < \delta$ then

$$a\|x - x_0\|_{\mathbb{R}^n} \leq d(\Upsilon(x,\cdot),\Upsilon(x_0,\cdot)) \leq b\|x - x_0\|_{\mathbb{R}^n},$$

where a, b and δ are positive constants and $\|\cdot\|_{\mathbb{R}^n}$ stands for the euclidean norm on $\mathbb{R}^n$. Then, we have

$$\begin{aligned}
&\left|P((X_t,X_s) \in \mathcal{B}(\upsilon(x),h)^2) - P(X_t \in \mathcal{B}(\upsilon(x),h))^2\right| \\
&\quad = \left|P((Z_t,Z_s) \in \upsilon^{-1}(\mathcal{B}(\upsilon(x),h))^2) - P\big(Z_t \in \upsilon^{-1}(\mathcal{B}(\upsilon(x),h))\big)^2\right| \\
&\quad = \int_{\upsilon^{-1}(\mathcal{B}(\upsilon(x),h))^2} |g_{s,t}(x,y)|\,\mathrm{d}x\,\mathrm{d}y \\
&\quad \leq \|g_{s,t}\|_\infty \frac{\pi^n (2h)^{2n}}{\Gamma(n/2+1)^2 a^{2n}},
\end{aligned}$$

where Γ is the Euler's gamma function. Moreover, if f is bounded from below by a positive constant in a neighborhood of $\mathcal{C}$, then there exists two positive constants β_1 and β_2 such that

$$0 < \beta_1\varphi(h) \leq P(X_0 \in \mathcal{B}(x,h)) \leq \beta_2\varphi(h),$$

and it is obvious that condition (D.3) (a) is fulfilled with $\varphi(h) = h^n$.

4.2 Time series prediction. Let (Z_t, Y_t) be a bivariate continuous time process. Here, we aim at predicting Y_t given the past of the process Z_t, say $(Z_u)_{u\in[t-a;t]}$ or $(Z_u - Z_{t-a})_{u\in[t-a;t]}$. Despite the fact that it seems natural and useful to study the relation between the two processes for any t, this topic has been considered only at discrete times. Setting for any t, $U_t := (Z_u - Z_{t-1})_{u\in[t-1;t]}$, we can apply our results to the process (U_t, Y_t). Considering the sup-norm on the functional space, if Z_t is a bilateral Wiener process, we have for any $t \in \mathbb{R}$,

$$P(U_t \in \mathcal{B}(0,h)) = \exp\Big(\frac{-\pi^2}{8h^2}(1 + o(1))\Big).$$

Moreover, whenever $|t - s| > 1$,

$$P((U_s,U_t) \in \mathcal{B}(0,h)^2) = P(U_t \in \mathcal{B}(0,h))^2,$$

and conditions (3) and (D.1) (d) are satisfied for $\varphi(h) = \exp(-\pi^2/8h^2)$.

4.3 Autoregressive process. We give now an example of autoregressive functional valued continuous time process. Consider the Hilbert space $L^2([0,1])$, and denote by $\mathcal{L}$ the space of continuous linear operator from $L^2([0,1])$ to $L^2([0,1])$, equipped with the usual norm $\|\cdot\|_{\mathcal{L}}$. Let ε_t be a continuous time process valued in $L^2([0,1])$, such that for some $h > 0$ and any $t \in \mathbb{R}$, the discrete time process $(\varepsilon_{t-ih})_{i\in\mathbb{N}}$ is a white noise, see [3] for a definition of Hilbertian white noise. It is obvious that such a process exists since the (U_t) in the last example verifies such a property. Then, for any $\varrho \in \mathcal{L}$ such that $\|\varrho\|_{\mathcal{L}} < 1$, the process

$$X_t := \sum_{i=0}^{\infty} \varrho^i(\varepsilon_{t-ih})$$

defines almost surely an autoregressive process which is the solution of the equation

$$X_t := \varrho(X_{t-ih}) + \varepsilon_t.$$

5. Proofs

To prove our results, we will establish a sequence of lemmas splitting up the whole proof into several steps. Set

$$\Delta_{T,t}(x) = K(h_T^{-1} d(x, X_t))$$

and introduce the notations

$$\widehat{r}_{1,T}(x) := \frac{1}{TE(\Delta_{T,0}(x))} \int_{t=0}^{T} \Delta_{T,t}(x)\,\mathrm{d}t,$$

$$\widehat{r}_{2,T}(x,y) := \frac{1}{TE(\Delta_{T,0}(x))} \int_{t=0}^{T} \Psi(y, Y_t)\Delta_{T,t}(x)\,\mathrm{d}t$$

and

$$s_T(x,y) = \int_{(s,t)\in[0,T]^2} |\mathrm{Cov}(\Psi(y,Y_s)\Delta_{T,s}(x), \Psi(y,Y_t)\Delta_{T,t}(x))|\,\mathrm{d}s\,\mathrm{d}t.$$

Then, one can observe that, when $\widehat{r}_{1,T}(x) \neq 0$,

$$\widehat{r}_T(x,y) = \frac{\widehat{r}_{2,T}(x,y)}{\widehat{r}_{1,T}(x)}.$$

The following lemmas will be useful to prove our main result. In particular, Lemma 5.1 studies the behavior of the bias of the generalized regression function estimator. It is useful to prove both Theorem 3.1 and Theorem 3.2. Lemma 5.2

is needed to prove Theorem 3.1 and Lemma 5.3 provides an upper bound for the generalized regression function estimator variance under conditions of Theorem 3.2.

Lemma 5.1. *Under conditions* (1) *and* (3), *we have for T large enough,*

$$\sup_{y\in\mathcal{S}}\sup_{x\in\mathcal{C}} |E\widehat{r}_{2,T}(x,y) - r(x,y)| \le Ch_T^{\eta}, \tag{13}$$

where η and C are introduced in Hypothesis 1.

Lemma 5.2. *Under assumptions of Theorem 3.1, we have*

$$\sup_{x\in\mathcal{C}}\sup_{y\in\mathcal{S}} s_T(x,y) = \mathcal{O}(T\varphi(h_T)). \tag{14}$$

Lemma 5.3. *Under the conditions of Theorem* 3.2, *we have*

$$\sup_{x\in\mathcal{C}}\Big(\sup_{y\in\mathcal{S}}\big(\mathrm{Var}(\widehat{r}_{2,T}(x,y))\big) + \mathrm{Var}(\widehat{r}_{1,T}(x))\Big) = \mathcal{O}\Big(\frac{1}{T}\Big). \tag{15}$$

In the proofs of Lemmas 5.1–5.3, we fix $(x,y)\in\mathcal{B}\times\mathcal{S}$ and when no confusion is possible, use the notation Ψ_t, $\Delta_{T,t}$ and s_T instead of $\Psi(y,Y_t)$, $\Delta_{T,t}(x)$ and $s_T(x,y)$.

Proof of Lemma 5.1: Observe that

$$\begin{aligned} E\widehat{r}_{2,T}(x,y) = \frac{E(\Psi_0\Delta_{T,0})}{E\Delta_{T,0}} &= \frac{E\big(E\big(r(X_0,y)+\varepsilon_0(y)|X_0\big)\Delta_{T,0}\big)}{E\Delta_{T,0}} \\ &= \frac{E(r(X_0,y)\Delta_{T,0})}{E\Delta_{T,0}}. \end{aligned}$$

Thus, making use of hypothesis (1), we have the desired relation, i.e.,

$$\begin{aligned} |E\widehat{r}_{2,T}(x,y) - r(x,y)| &\le \frac{E(|r(X_0,y)-r(x,y)|\Delta_{T,0}(x))}{E\Delta_{T,0}(x)} \\ &\le \sup_{u\in\mathcal{B}(x,h)} \frac{E(|r(u,y)-r(x,y)|\Delta_{T,0}(x))}{E\Delta_{T,0}(x)} \le Ch_T^{\eta}. \end{aligned}$$

Lemma 5.1 is proved. □

PROOF OF LEMMA 5.2: Set $\Gamma := [0,T]^2$ and $v_T := \varphi(h_T)^{-1}$. Below, we give a decomposition of the upper bound of s_T into three terms, and treat these terms separately

$$\begin{aligned} s_T \leq & \int_{\Gamma\cap\{|t-s|<\delta\}} |\mathrm{Cov}(\Psi_t\Delta_{T,t}, \Psi_s\Delta_{T,s})| \,\mathrm{d}t\,\mathrm{d}s \\ & + \int_{\Gamma\cap\{\delta\leq|t-s|\leq v_T\}} |\mathrm{Cov}(\Psi_t\Delta_{T,t}, \Psi_s\Delta_{T,s})| \,\mathrm{d}t\,\mathrm{d}s \\ & + \int_{\Gamma\cap\{v_T<|t-s|\}} |\mathrm{Cov}(\Psi_t\Delta_{T,t}, \Psi_s\Delta_{T,s})| \,\mathrm{d}t\,\mathrm{d}s \\ =: & \, W_1 + W_2 + W_3. \end{aligned}$$

Considering the first term, we can write

$$W_1 \leq \int_{\Gamma\cap\{|t-s|<\delta\}} \mathrm{Var}(\Psi_t\Delta_{T,t}) \,\mathrm{d}t\,\mathrm{d}s \leq \int_{\Gamma\cap\{|t-s|<\delta\}} E(\Psi_t\Delta_{T,t})^2 \,\mathrm{d}t\,\mathrm{d}s.$$

Using conditions (3) and (D.1) (a), we have

$$E(\Psi_t\Delta_{T,t})^2 = E(E(\Psi_0^2|X_0)\Delta_{T,i}^2) \leq \beta_2 M_1 \|K\|_\infty^2 \varphi(h_T).$$

Thus,

$$W_1 \leq 2\delta T\varphi(h_T)\beta_2 M_1 \|K\|_\infty^2 .$$

Considering the term W_2, we can write

$$|W_2| \leq \int_{\Gamma\cap\{\delta\leq|t-s|\leq v_T\}} (|E(\Psi_s\Psi_t\Delta_{T,s}\Delta_{T,t})| + (E(\Psi_0\Delta_{T,0}))^2) \,\mathrm{d}s\,\mathrm{d}t.$$

Then conditions (3) and (D.1) (a) imply that

$$|E(\Psi_0\Delta_{T,0})| = |E(E(\Psi_0|X_0)\Delta_{T,0})| \leq (M_1+1)E(\Delta_{T,0}) \leq (M_1+1)\beta_2\|K\|_\infty\varphi(h_T).$$

Moreover, for T large enough and for $|t-s| \geq \delta$, we have

$$\begin{aligned} E(\Psi_s\Psi_t\Delta_{T,s}\Delta_{T,t}) &= E(E(\Psi_s\Psi_t|X_s, X_t)\Delta_{T,s}\Delta_{T,t}) \leq M_1 E(\Delta_{T,s}\Delta_{T,t}) \\ &\leq \beta_3 M_1 \|K\|_\infty^2 \varphi(h_T)^2. \end{aligned}$$

Therefore, for some constant $C_2 > 0$ independent of x, we obtain

$$W_2 \leq C_2 T\varphi(h_T).$$

Finally, by Davydov's inequality, see (1.10) in [2], and the assumptions (D.1) (e) and (H) (b), it follows that

$$\begin{aligned}
W_3 &\leq 2\frac{p2^{(p-2)/p}\|\Psi_0\Delta_{T,0}\|_p^2}{(p-2)}\int_{\Gamma\cap\{v_T<|t-s|\}}\alpha(|t-s|)^{(p-2)/p}\,\mathrm{d}t\,\mathrm{d}s\\
&\leq 2\frac{cp2^{(p-2)/p}\|\Psi_0\|_p^2\|K\|_\infty^2}{(p-2)}\int_{\Gamma\cap\{v_n<|t-s|\}}|s-t|^{-a(p-2)/p}\,\mathrm{d}t\,\mathrm{d}s\\
&\leq 4\frac{cp^2 2^{(p-2)/p}\|\Psi_0\|_p^2\|K\|_\infty^2 T v_T^{(p-a(p-2))/p}}{(a(p-2)-p)(p-2)}\\
&\leq 4\frac{cp^2 2^{(p-2)/p}\|\Psi_0\|_p^2\|K\|_\infty^2 T\varphi(h_T)}{(a(p-2)-p)(p-2)}.
\end{aligned}$$

Thus, there exists a constant $C_3>0$ such that

$$s_T \leq C_3 T\varphi(h_T), \tag{16}$$

ending the proof Lemma 5.2. □

Proof of Lemma 5.3: In order to simplify the notations, we set $R(X_t) := E\big(\Psi_t|X_t\big)$ and $r(x) := r(x,y)$. Observe that, by Fubini's theorem,

$$\mathrm{Var}(\widehat{r}_{2,T}(x)) = \frac{1}{T^2E(\Delta_0)^2}\int_{t=0}^{T}\int_{s=0}^{T}\mathrm{Cov}(\Psi_s\Delta_{T,s}(x),\Psi_t\Delta_{T,t}(x))\,\mathrm{d}t\,\mathrm{d}s.$$

The covariance term can be expended as follows:

$$\begin{aligned}
\mathrm{Cov}(\Psi_s\Delta_{T,s}(x),\Psi_t\Delta_{T,t}(x)) &= E(\Delta_{T,s}(x)\Delta_{T,t}(x)R(X_s)R(X_t))\\
&\quad+E\big(\Delta_{T,s}(x)\Delta_{T,t}(x)(R(X_s)E(\varepsilon_t|X_s,X_t)+R(X_t)E(\varepsilon_s|X_s,X_t))\big)\\
&\quad+E(\Delta_{T,s}(x)\Delta_{T,t}(x)E(\varepsilon_s\varepsilon_t|X_s,X_t))-E\big(\Delta_{T,s}(x)(R(X_s))\big)^2.
\end{aligned}$$

Taking $d_t = R(X_t)-r(x)$, we have

$$\begin{aligned}
\mathrm{Cov}(\Psi_s\Delta_{T,s}(x),\Psi_t\Delta_{T,t}(x)) &= E(\Delta_{T,s}(x)\Delta_{T,t}(x)r(x)r(x))\\
&\quad+E(\Delta_{T,s}(x)\Delta_{T,t}(x)r(x)d_t)+E(\Delta_{T,s}(x)\Delta_{T,t}(x)r(x)d_s)\\
&\quad+E(\Delta_{T,s}(x)\Delta_{T,t}(x)d_td_s)\\
&\quad+E\big(\Delta_{T,s}(x)\Delta_{T,t}(x)(r(x)E(\varepsilon_t|X_s,X_t)+r(x)E(\varepsilon_s|X_s,X_t))\big)\\
&\quad+E\big(\Delta_{T,s}(x)\Delta_{T,t}(x)(d_sE(\varepsilon_t|X_s,X_t)+d_tE(\varepsilon_s|X_s,X_t))\big)\\
&\quad+E(\Delta_{T,s}(x)\Delta_{T,t}(x)E(\varepsilon_s\varepsilon_t|X_s,X_t))-E\big(\Delta_{T,s}(x)(r(x))\big)^2\\
&\quad-E(\Delta_{T,s}(x)d_s)^2-2r(x)E(\Delta_{T,s}(x)d_s)E(\Delta_{T,s}(x)).
\end{aligned}$$

From conditions (1) and (D.3) (b), we obtain

$$\begin{aligned}|\mathrm{Cov}(\Psi_s\Delta_{T,s}(x),\Psi_t\Delta_{T,t}(x))| &\leq r(x)^2|E(\Delta_{T,s}(x)\Delta_{T,t}(x))-E(\Delta_{T,s}(x))^2|\\ &\quad+(2|r(x)|Ch^\eta+C^2h^{2\eta}+(2(|r(x)|+Ch^\eta)+1)g_1(|s-t|))\\ &\quad\times E(\Delta_{T,s}(x)\Delta_{T,t}(x))\\ &\quad+(2r(x)Ch^\eta+C^2h^{2\eta})E(\Delta_{T,s}(x))^2.\end{aligned}$$

Thus, making use of hypothesis (D.3) (a), we have

$$\begin{aligned}|\mathrm{Cov}(\Psi_s\Delta_{T,s}(x),\Psi_t\Delta_{T,t}(x))| &\leq r(x)^2g_0(|s-t|)\varphi(h)^2\\ &\quad+(2|r(x)|Ch^\eta+C^2h^{2\eta}+(2(|r(x)|+Ch^\eta)+1)g_1(|s-t|))\\ &\quad\times(\beta_2+g_0(|s-t|))\varphi(h)^2\\ &\quad+(2r(x)Ch^\eta+C^2h^{2\eta})\beta_2^2\varphi(h)^2=:G_T(|s-t|).\end{aligned}$$

This allows us to write

$$\begin{aligned}\mathrm{Var}(\widehat{r}_{2,T}(x)) &\leq \frac{2}{T^2E(\Delta_0)^2}\int_{t=0}^T\int_{s=t}^T G_T(s-t)\,\mathrm{d}t\,\mathrm{d}s\\ &\leq \frac{2}{TE(\Delta_0)^2}\int_{s=0}^T G_T(s)\,\mathrm{d}s.\end{aligned}$$

So, since g_0 and g_1 are integrable and $h^\eta=o(1/T)$, there exists a constant M_0 such that

$$\mathrm{Var}(\widehat{r}_{2,T}(x))\leq\frac{M_0}{T}.$$

This last inequality concludes the proof of Lemma 5.3. □

Proof of Theorem 3.1: We will establish the proof under condition D.1 (c) ($\star$), and only make a remark about the case when D.1 (c) ($\star\star$) is satisfied.

We will first show that

$$\limsup_{T\to\infty}\sup_{y\in\mathcal{S}}\sup_{x\in\mathcal{C}}\frac{\int_{t=0}^T|\Psi(y,Y_t)\Delta_{T,t}-E(\Psi(y,Y_t)\Delta_{T,t})|\,\mathrm{d}t}{V_T}\leq L'\qquad\text{a.s.,}$$

where $V_T:=\sqrt{T\varphi(h_T)\ln(T)}$.

Making use of the condition (H) (a), we have

$$\left|\frac{\partial\Delta_{T,t}}{\partial T}\right|=\left|\frac{h_T'}{h_T^2}d(x,X_t)K'(h_T^{-1}d(x,X_t))\right|\leq\frac{h_T'}{h_T}\|K'\|_\infty=\mathcal{O}\Big(\frac{1}{T}\Big)\tag{17}$$

which subsequently implies that there exists a constant $C_4 > 0$ such that for any $n \in \mathbb{N}^*$, any $x \in \mathcal{C}$ and any $T' \in [n, n+1[$

$$|\Delta_{T',t}(x) - \Delta_{n,t}(x))| \leq \frac{C_4}{n}. \tag{18}$$

So, for any positive constant A and for $n \in \mathbb{N}$ large enough, we have, by condition (D.1) (a),

$$\sup_{y\in\mathcal{S}} \sup_{x\in\mathcal{C}} \sup_{T'\in[n,n+1[} \frac{n}{V_n} |E(\Psi(y,Y_t)\Delta_{n,t}(x)) - E(\Psi(y,Y_t)\Delta_{T',t}(x))| < \frac{A}{4}.$$

Thus, since V_T is an increasing function of T, we can write

$$\begin{aligned}
&P\biggl(\sup_{y\in\mathcal{S}} \sup_{x\in\mathcal{C}} \sup_{T'\in[n,n+1[} \biggl|\frac{1}{V_{T'}} \int_0^{T'} \Psi(y,Y_t)\Delta_{T',t}(x) - E(\Psi(y,Y_t)\Delta_{T',t}(x))\,\mathrm{d}t\biggr| > A\biggr) \\
&\leq P\biggl(\sup_{y\in\mathcal{S}} \sup_{x\in\mathcal{C}} \sup_{T'\in[n,n+1[} \frac{1}{V_n} \int_0^{n} |\Psi(y,Y_t)(\Delta_{T',t}(x) - \Delta_{n,t}(x))|\,\mathrm{d}t > \frac{A}{4}\biggr) \\
&\quad + P\biggl(\sup_{y\in\mathcal{S}} \sup_{x\in\mathcal{C}} \biggl|\frac{1}{V_n} \int_0^{n} \Psi(y,Y_t)\Delta_{n,t}(x) - E(\Psi(y,Y_t)\Delta_{n,t}(x))\,\mathrm{d}t\biggr| > \frac{A}{4}\biggr) \\
&\quad + P\biggl(\sup_{y\in\mathcal{S}} \sup_{x\in\mathcal{C}} \sup_{T'\in[n,n+1[} \frac{1}{V_n} |\Psi(y,Y_t)\Delta_{T',t}(x) - E(\Psi(y,Y_t)\Delta_{T',t}(x))| > \frac{A}{4}\biggr) \\
&= A_1 + A_2 + A_3.
\end{aligned}$$

The relation (18) yields the following upper bound for A_1

$$A_1 \leq P\biggl(\frac{C_4}{nV_n} \int_0^n \sup_{y\in\mathcal{S}} |\Psi(y,Y_t)|\,\mathrm{d}t > \frac{A}{4}\biggr).$$

From Davydov's inequality, under the conditions (D.1), there exists a constant $C_5 > 0$ such that

$$\begin{aligned}
&\mathrm{Var}\biggl(\int_0^n \sup_{y\in\mathcal{S}} |\Psi(y,Y_t)|\,\mathrm{d}\biggr) \\
&\quad \leq 2\frac{p}{p-2} 2^{(p-2)/p} \Bigl\|\sup_{y\in\mathcal{S}} \Psi(y,Y_t)\Bigr\|_p^2 \int_{[0,n]^2} \min\Bigl((c|s-t|)^{-a(p-2)/p}, \frac{1}{4}\Bigr)\,\mathrm{d}s\,\mathrm{d}t \\
&\quad \leq C_5 n.
\end{aligned}$$

That is for n large enough, we have

$$\frac{C_4 E\bigl(\int_0^n \sup_{y\in\mathcal{S}} |\Psi(y,Y_t)|\,\mathrm{d}t\bigr)}{nV_n} \leq \frac{A}{8}.$$

So, using the Bienaymé–Tchebychev inequality we obtain

$$A_1 \le \frac{64 C_5 C_4^2}{V_n^2 n A^2}. \tag{19}$$

From the condition (2), we can cover the compact $\mathcal{C}$ with $L_n = (n/h_n)^{d_1}$ balls of center x_k, $k \in [1, L_n]$, and radius $\nu_n \le \big(C_1^{1/d_1} h_n/n\big)$. We easily observe that the relation (18) implies for n large enough that

$$\sup_{y\in\mathcal{S}} \sup_{k\in[1,L_n]} \sup_{x\in\mathcal{B}(x_k,\nu_n)} \left| \frac{1}{V_n} \int_0^n E\big(\Psi(y, Y_t)(\Delta_{n,t}(x_k) - \Delta_{n,t}(x))\big)\, \mathrm{d}t \right| < \frac{A}{12}.$$

So, we can write

$$\begin{aligned} A_2 \le{} & P\bigg(\sup_{y\in\mathcal{S}} \sup_{k\in[1,L_n]} \sup_{x\in\mathcal{B}(x_k,L_n)} \left| \frac{1}{V_n} \int_0^n \Psi(y, Y_t)(\Delta_{n,t}(x_k) - \Delta_{n,t}(x))\, \mathrm{d}t \right| > \frac{A}{12} \bigg) \\ & + P\bigg(\sup_{y\in\mathcal{S}} \sup_{k\in[1,L_n]} \left| \frac{1}{V_n} \int_0^n \Psi(y, Y_t)\Delta_{n,t}(x_k) \right. \\ & \left. - E(\Psi(y, Y_t)\Delta_{n,t}(x_k))\, \mathrm{d}t \right| > \frac{A}{12} \bigg) \\ =:{} & B_1 + B_2. \end{aligned}$$

From the condition (K), it is easily seen for some $m > 0$ that the kernel K is a m-Lipschitz function. Therefore, since $|\Delta_{n,t}(x_k) - \Delta_{n,t}(x)| \le m\nu_n/h_n$ for any $x \in \mathcal{B}(x_k, \nu_n)$, following the calculations made for A_1, we obtain

$$B_1 = \mathcal{O}\Big(\frac{1}{V_n^2 n}\Big). \tag{20}$$

We can cover $\mathcal{S}$ with $L'_n = n^{p_1}$ intervals of center $y_{k'}$, $k' \in [1, L'_n]$ and length $2l \le C_6/L'_n$ where C_6 is independent of n. Moreover, condition (D.1) (c) implies that for a large enough n

$$\begin{aligned} \sup_{k'\in[1,L'_n]} \sup_{k\in[1,L_n]} \sup_{y\in[y_{k'}\pm l]} & \left| \frac{1}{V_n} \int_0^n E\big(\Delta_{n,t}(x_k)(\Psi(y, Y_t) \right. \\ & \left. - \Psi(y_{k'}, Y_t))\big)\, \mathrm{d}t \right| < \frac{A}{12}. \end{aligned} \tag{21}$$

So, we can decompose the term B_2 as follows:

$$
\begin{aligned}
B_2 &\le P\Big(\sup_{k'\in[1,L'_n]} \sup_{k\in[1,L_n]} \sup_{y\in[y_{k'}\pm l]} \Big| \frac{1}{V_n} \int_0^n \Delta_{n,t}(x_k)(\Psi(y,Y_t) \\
&\qquad - \Psi(y_{k'},Y_t))\,\mathrm{d}t \Big| > \frac{A}{12} \Big) \\
&\qquad + P\Big(\sup_{k'\in[1,L'_n]} \sup_{k\in[1,L_n]} \Big| \frac{1}{V_n} \int_0^n \Psi(y_{k'},Y_t)\Delta_{n,t}(x_k) \\
&\qquad - E(\Psi(y_{k'},Y_t)\Delta_{n,t}(x_k))\,\mathrm{d}t \Big| > \frac{A}{12} \Big) \\
&=: B'_1 + B'_2.
\end{aligned}
$$

Taking the first term, we have

$$
B'_1 \le \sum_{k'=1}^{L'_n} P\Big(\frac{\|K\|_\infty}{V_n} \int_0^n \sup_{y\in[y_{k'}\pm l]} |\Psi(y,Y_t) - \Psi(y_{k'},Y_t)|\,\mathrm{d}t > \frac{A}{12} \Big).
$$

As for the term A_1, in the setting of conditions (D.1) (c) and (D.1) (e), we make use of Davydov's and Bienaymé–Tchebychev's inequalities to obtain

$$
B'_1 \le \frac{C_7 n^{1-p_1(2\eta_1-1)}}{V_n^2}, \tag{22}
$$

where C_7 is a positive constant.

To treat the term B'_2, we introduce the following constant

$$
C_8 = 8\frac{p}{2p-1}(2^a c)^{(p-1)/(a+p)},
$$

where c is defined in (D.1) (e). First, remark that the kernel K is bounded, so condition (D.1) (b) implies that there exists a constant $Q_2 > 0$ such that for any $\lambda > 0$,

$$
P\Big(\Big| \frac{\int_0^n \Psi(y_{k'},Y_t)\Delta_{n,t}(x_k) - E\big(\Psi(y_{k'},Y_t)\Delta_{n,t}(x_k)\big)\,\mathrm{d}t}{Q_2} \Big| > \lambda \Big) \le \lambda^{-p}.
$$

So, a simple discretization of our estimator allows us to apply the Fuk–Nagaev inequality, see [24, formula (6.19 a)], to obtain for any $y \in \mathcal{S}$, any $x \in \mathcal{C}$ and any $z \ge 1$, that

$$
\begin{aligned}
&P\Big(\Big| \frac{1}{V_n} \int_0^n \Psi(y_{k'},Y_t)\Delta_{n,t}(x_k) - E(\Psi(y_{k'},Y_t)\Delta_{n,t}(x_k))\,\mathrm{d}t \Big| > \frac{A}{12} \Big) \\
&\qquad \le C_8 \Big(\Big(1 + \frac{A^2 V_n^2}{48^2 z s_n Q_2^2}\Big)^{-z/2} + \frac{n}{z}\Big(\frac{48 z Q_2}{A V_n}\Big)^{(a+1)p/(a+p)} \Big).
\end{aligned} \tag{23}
$$

In view of (16), under condition (H) (b), choosing A large enough and $z = \ln(n)$, there exists $C_9 > 0$ such that for some $\xi' > 1$

$$B_2' \leq \frac{C_9}{n\ln(n)^{\xi'}}. \tag{24}$$

Thus, from (20), (22), (24) for A large enough there exists $C_{10} > 0$ such that

$$A_2 \leq \frac{C_{10}}{n\ln(n)^{\xi'}}. \tag{25}$$

Finally, the hypothesis (D.1) (b) leads to

$$A_3 = \mathcal{O}(V_n^{-p}). \tag{26}$$

So from (19), (25) and (26), using the Borel–Cantelli lemma, we have for a constant $L' > 0$

$$\limsup_{T\to\infty} \sup_{y\in\mathcal{S}} \sup_{x\in\mathcal{C}} \sqrt{\frac{T\varphi(h_T)}{\ln(T)}} |\widehat{r}_{2,T}(x,y) - E(\widehat{r}_{2,T}(x,y))| \leq L' \qquad \text{a.s.}$$

Since $E\widehat{r}_{1,T}(x) = 1$, we have for some $L'' > 0$

$$\limsup_{T\to\infty} \sup_{y\in\mathcal{S}} \sup_{x\in\mathcal{C}} \sqrt{\frac{T\varphi(h_T)}{\ln(T)}} |\widehat{r}_{1,T}(x) - 1| \leq L'' \qquad \text{a.s.} \tag{27}$$

The statement (27) implies that, almost surely, there exists $T_0 \in \mathbb{R}_+$ such that for all $T > T_0$

$$\widehat{r}_T(x,y) = \frac{\widehat{r}_{2,T}(x,y)}{\widehat{r}_{1,T}(x)}.$$

Then, the use of the following decomposition concludes the proof of Theorem 3.1:

$$\begin{aligned} \frac{\widehat{r}_{2,T}(x,y)}{\widehat{r}_{1,T}(x)} - r(x,y) = {} & \frac{r(x,y)}{\widehat{r}_{1,T}(x)}\big(1 - \widehat{r}_{1,T}(x)\big) \\ & + \frac{1}{\widehat{r}_{1,T}(x)}\Big(\big(\widehat{r}_{2,T}(x,y) - E(\widehat{r}_{2,T}(x,y))\big) \\ & - \big(r(x,y) - E(\widehat{r}_{2,T}(x,y))\big)\Big). \end{aligned}$$

□

Remark 5.1. Under D.1 (c) (⋆⋆), this proof is valid except the calculation of a bound for B_2'. We just have to remark that under D.1 (c) (⋆⋆), the term B_2' can be bounded by making use of inequalities (21) and (24).

PROOF OF COROLLARY 3.1: Note that Corollary 3.1 follows as a direct application of Theorem 3.1. It suffices to take $\Psi(y,Y) = 1\!\!1_{[-\infty,y]}(Y)$. □

Proof of Corollary 3.2: From Corollary 3.1 and Hypothesis (D.2), almost surely, we have for T large enough

$$\sup_{x\in\mathcal{C}} \widehat{F}_T(a_1|x) - \theta_1(x) < 0, \qquad \sup_{x\in\mathcal{C}} \big(\theta_2(x) - \widehat{F}_T\big(\widehat{u_{b_1}(x)}|x\big)\big) < 0.$$

So, it follows that almost surely for T large enough,

$$\widehat{u_\theta(x)} \in \mathcal{S} \qquad \text{for any } (\theta, x) \in [\theta_1, \theta_2] \times \mathcal{C}.$$

Set $W_T = L_1(h_T^\eta + \sqrt{\ln(T)/T\varphi(h_T)})$. By Corollary 3.1, we obtain that, almost surely for T large enough,

$$\sup_{x\in\mathcal{C}} \sup_{\theta\in[\theta_1,\theta_2]} \big|\widehat{F}_T\big(\widehat{u_\theta(x)}|x\big) - F\big(\widehat{u_\theta(x)}|x\big)\big| \le W_T \tag{28}$$

and

$$\sup_{x\in\mathcal{C}} \sup_{\theta\in[\theta_1,\theta_2]} \lim_{u\to\widehat{u_\theta(x)}, u<\widehat{u_\theta(x)}} \big|\widehat{F}_T(u|x) - F(u|x)\big| \le W_T.$$

But, in view of (7), we always have

$$\theta \in \Big[\lim_{u\to\widehat{u_\theta(x)}, u<\widehat{u_\theta(x)}} \widehat{F}_T(u), \widehat{F}_T\big(\widehat{u_\theta(x)}\big)\Big].$$

Therefore, since F is continuous, almost surely we have for T large enough

$$\sup_{x\in\mathcal{C}} \sup_{\theta\in[\theta_1,\theta_2]} \big|\widehat{F}_T\big(\widehat{u_\theta(x)}|x\big) - \theta\big| \le 2W_T. \tag{29}$$

Condition (D.2) implies that for any $x \in \mathcal{C}$ and any $\theta \in [\theta_1, \theta_2]$ we have

$$\big|F(u_\theta(x)|x) - F\big(\widehat{u_\theta(x)}|x\big)\big| > \min\big(C', C'\big|u_\theta(x) - \widehat{u_\theta(x)}\big|^{\eta_2}\big). \tag{30}$$

Making use of the following decomposition

$$\begin{aligned}\big|F(u_\theta(x)|x) - F\big(\widehat{u_\theta(x)}|x\big)\big| \le{}& \big|F(u_\theta(x)|x) - \widehat{F}_T\big(\widehat{u_\theta(x)}|x\big)\big| \\ &+ \big|\widehat{F}_T\big(\widehat{u_\theta(x)}|x\big) - F\big(\widehat{u_\theta(x)}|x\big)\big|\end{aligned}$$

and the statements (28) and (29), it is easily seen that almost surely we have for T large enough

$$\sup_{x\in\mathcal{C}} \sup_{\theta\in[\theta_1,\theta_2]} \big|F(u_\theta(x)|x) - F\big(\widehat{u_\theta(x)}|x\big)\big| \le 3W_T.$$

This last inequality, in view of (30), achieves the proof. □

Proof of Theorem 3.2: Using the decomposition (28), in view of Lemma 5.1 and Lemma 5.3, Theorem 3.2 follows by Bienaymé–Tchebychev's inequality. □

PROOF OF COROLLARY 3.3: The proof is a straightforward consequence of Theorem 3.2. □

PROOF OF COROLLARY 3.4: Fix $x \in \mathcal{C}$, $\theta \in [\theta_1, \theta_2]$, $\varepsilon > 0$ and set $V_T := \sqrt{\ln_m(T)/T}$, $u' := u_\theta(x) + (\varepsilon/2)V_T^{1/\eta_2}$, and $u'' := u_\theta(x) + \varepsilon V_T^{1/\eta_2}$. We begin the proof by establishing the upper bound that is

$$\lim_{T\to\infty} P\big(\widehat{u_\theta(x)} - u_\theta(x) > \varepsilon V_T^{1/\eta_2}\big) = 0. \tag{31}$$

In this respect, observe that

$$\begin{aligned}
P\big(\widehat{u_\theta(x)} - u_\theta(x) > \varepsilon V_T^{1/\eta_2}\big) &= P\Big(\widehat{u_\theta(x)} - \frac{\varepsilon}{2} V_T^{1/\eta_2} > u'\Big)\\
&\le P\Big(\widehat{F}_T\big(\widehat{u_\theta(x)} - \frac{\varepsilon}{2}V_T^{1/\eta_2}|x\big) - \widehat{F}_T(u'|x) \ge 0\Big)\\
&\le P\Big(\widehat{F}_T\big(\widehat{u_\theta(x)} - \frac{\varepsilon}{2}V_T^{1/\eta_2}|x\big) - F(u'|x) \ge -C'\Big(\frac{\varepsilon}{2}\Big)^{\eta_2} V_T\Big)\\
&\quad + P\Big(F(u'|x) - \widehat{F}_T(u'|x) \ge C'\Big(\frac{\varepsilon}{2}\Big)^{\eta_2} V_T\Big).
\end{aligned}$$

Since $\widehat{F}_T\big(\widehat{u_\theta(x)} - (\varepsilon/2)V_T^{1/\eta_2}|x\big) < \theta = F(u_\theta(x)|x)$, from the condition (D.2) we obtain for T large enough

$$P\Big(\widehat{F}_T\big(\widehat{u_\theta(x)} - \frac{\varepsilon}{2}V_T^{1/\eta_2}|x\big) - F(u'|x) \ge -C'\Big(\frac{\varepsilon}{2}\Big)^{\eta_2} V_T\Big) = 0.$$

Moreover, Corollary 3.3 leads to the following result

$$\lim_{T\to\infty} P\Big(F(u'|x) - \widehat{F}_T(u'|x) \ge C'\Big(\frac{\varepsilon}{2}\Big)^{\eta_2} V_T\Big) = 0,$$

which proves the statement (31). It is easily seen that the same arguments lead to the following lower bound

$$\lim_{T\to\infty} P\big(\widehat{u_\theta(x)} - u_\theta(x) < -\varepsilon V_T^{1/\eta_2}\big) = 0.$$

In view of (31), this last result concludes the proof of Corollary 3.4. □

6. Summary

In this paper, we determined the convergence rates associated with the generalized regression function through the use of kernel estimation techniques. Several specific results follow, including the convergence of the conditional distribution function and the conditional quantiles. Discussions on their applications to functional parametric rate, time series prediction and autoregressive process are provided. In particular, some important results of [15] are extended and full proofs are given. We hope that the current theoretical work will serve as a basis for applied purposes in the field.

Acknowledgement. The authors are very grateful to the reviewer for very constructive comments which helped improve the first version of the manuscript.

References

[1] Bosq D., *Vitesses optimales et superoptimales des estimateurs fonctionnels pour les processus à temps continu*, C. R. Acad. Sci. Paris Sér. I Math. **317** (1993), no. 11, 1075–1078 (French).

[2] Bosq D., *Nonparametric Statistics for Stochastic Processes*, Estimation and Prediction, Lecture Notes in Statistics, 110, Springer, New York, 1998.

[3] Bosq D., *Linear Processes in Function Spaces*, Theory and Applications, Lecture Notes in Statistics, 149, Springer, New York, 2000.

[4] Bradley R. C., *Basic properties of strong mixing conditions. A survey and some open questions*, Probab. Surv. **2** (2005), 107–144.

[5] Cardot H., Ferraty F., Sarda P., *Functional linear model*, Statist. Probab. Lett. **45** (1999), no. 1, 11–22.

[6] Cardot H., Ferraty F., Sarda P., *Spline estimators for the functional linear model*, Statist. Sinica **13** (2003), no. 3, 571–591.

[7] Castellana J. V., Leadbetter M. R., *On smoothed probability density estimation for stationary processes*, Stochastic Process. Appl. **21** (1986), no. 2, 179–193.

[8] 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. **23** (1986), no. 1, 77–89.

[9] Dabo-Niang S., *Density estimation in a separable metric space*, Ann. I.S.U.P. **47** (2003), no. 1–2, 3–21.

[10] Dabo-Niang S., *Density estimation by orthogonal series in an infinite dimensional space: application to processes of diffusion type I*, The International Conf. on Recent Trends and Directions in Nonparametric Statistics, J. Nonparametr. Stat. **16** (2004), no. 1–2, 171–186.

[11] Demongeot J., Laksaci A., Madani F., Rachdi M., *A fast functional locally modeled conditional density and mode for functional time-series*, Recent Advances in Functional Data Analysis and Related Topics, Contrib. Statist., Physica Verlag, Springer, Heidelberg, 2011, 85–90.

[12] Ferraty F., Goia A., Vieu P., *Functional nonparametric model for time series: a fractal approach for dimension reduction*, Test **11** (2002), no. 2, 317–344.

[13] Ferraty F., Laksaci A., Tadj A., Vieu P., *Kernel regression with functional response*, Electron. J. Stat. **5** (2011), 159–171.

[14] Ferraty F., Rabhi A., Vieu P, *Conditional quantiles for dependent functional data with application to the climatic El Niño phenomenon*, Sankhyā **67** (2005), no. 2, 378–398.

[15] Ferraty F., Vieu P., *Nonparametric models for functional data, with application in regression, time-series prediction and curve discrimination*, The International Conf. on Recent Trends and Directions in Nonparametric Statistics, J. Nonparametr. Stat. **16** (2004), no. 1–2, 111–125.

[16] Ferraty F., Vieu P., *Nonparametric Functional Data Analysis*, Theory and Practice, Springer Series in Statistics, Springer, New York, 2006.

[17] Frank I. E., Friedman J. H., *A statistical view of some chemometric regression tools*, Technometrics **35** (1993), no. 2, 109–135.

[18] Grunig R., *Probabilités conditionnelles régulières sur des tribus de type non dénombrable*, Ann. Inst. H. Poincaré Sect. B (N.S.) **2** (1965/1966), no. 3, 227–229 (French).

[19] Jiřina M., *Conditional probabilities on strictly separable σ-algebras*, Czechoslovak. Math. J. **4(79)** (1954), 372–380.

[20] Jiřina M., *On regular conditional probabilities*, Czechoslovak. Math. J. **9(84)** (1959), 445–451.

[21] Krzyżak A., Pawlak M., *The pointwise rate of convergence of the kernel regression estimate*, J. Statist. Plann. Inference **16** (1987), no. 2, 159–166.

[22] Masry E., *Nonparametric regression estimation for dependent functional data: asymptotic normality*, Stochastic Process. Appl. **115** (2005), no. 1, 155–177.

[23] Ramsay J. O., Silverman B. W., *Functional Data Analysis*, Springer Series in Statistics, Springer, New York, 2005.

[24] Rio E., *Théorie asymptotique des processus aléatoires faiblement dépendants*, Mathématiques & Applications, 31, Springer, Berlin, 2000 (French).

[25] Rosenblatt M., *Conditional probability density and regression estimators*, in Multivariate Analysis II, Proc. Second Internat. Sympos., Dayton, Ohio, 1968, Academic Press, New York, 1969, pages 25–31.

[26] Roussas G. G., *Nonparametric regression estimation under mixing conditions*, Stochastic Process. Appl. **36** (1990), no. 1, 107–116.

[27] Stone C. J., *Optimal global rates of convergence for nonparametric regression*, Ann. Statist. **10** (1982), no. 4, 1040–1053.

[28] Watson G. S., *Smooth regression analysis*, Sankhyā Ser. A **26** (1964), 359–372.

B. Maillot, Ch. Chesneau:

Université de Caen Normandie, LMNO, Campus II, Science 3, 14032, Caen, France

E-mail: bertrand.maillot@unicaen.fr

E-mail: christophe.chesneau@gmail.com

(*Received* June 29, 2020*, revised* January 24, 2021)