

Research Article

Superoptimal Rate of Convergence in Nonparametric Estimation for Functional Valued Processes

Christophe Chesneau and Bertrand Maillot

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

Correspondence should be addressed to Christophe Chesneau; christophe.chesneau@gmail.com

Received 5 March 2014; Revised 12 May 2014; Accepted 4 June 2014; Published 16 September 2014

Academic Editor: Armelle Guillou

Copyright © 2014 C. Chesneau and B. Maillot. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

We consider the nonparametric estimation of the generalised regression function for continuous time processes with irregular paths when the regressor takes values in a semimetric space. We establish the mean-square convergence of our estimator with the same superoptimal rate as when the regressor is real valued.

1. Introduction

Since the pioneer works of [1, 2], the nonparametric estimation of the regression function has been very widely studied for real and vectorial regressors (see, e.g., [3–8]) and, more recently, the case when the regressor takes values in a semimetric space of infinite dimension has been addressed. Interest in this type of explanatory variables has increased quickly since the foundational work of Ramsay and Silverman (1997), who proposed efficient methods for linear modelling (see [9] for a reissue of this work or [10, 11] for other developments on this topic). Later, fully nonparametric methods have been proposed (e.g., [12–15]) but the increased generality comes at a price in terms of convergence rate: in the regression estimation framework, it is well known that the efficiency of a nonparametric estimator decreases quickly when the dimension of the regressor grows. This problem, known as the “curse of dimensionality,” is due to the sparsity of data in high dimensional spaces. However, when studying continuous time processes with irregular paths, it has been shown in [16] that even when the regressor is \mathbb{R}^d -valued, we can estimate the regression function with the parametric rate of convergence $\mathcal{O}(1/\sqrt{T})$. This kind of superoptimal rate of convergence for nonparametric estimators is always obtained under hypotheses on the joint probability density functions of the process which are very similar to those introduced by [17]. Since there is no equivalent of the Lebesgue measure on an

infinite-dimensional Hilbert space, the definition of a density is less natural in the infinite-dimensional framework and the classical techniques cannot be applied. Under hypotheses about probabilities of small balls, we show that we can reach superoptimal rates of convergence for nonparametric estimation of the regression function when the regressor takes values in an infinite-dimensional space.

Notations and assumptions are presented in Section 2. Section 3 introduces our estimator and the main result. We comment on hypotheses and results and give some examples of processes fulfilling our hypotheses in Section 4. A numerical study can be found in Section 5. The proofs are postponed to Section 6.

2. Problem and Assumptions

Let $\{X_t, Y_t\}_{t \in [0, \infty)}$ be a measurable continuous time process defined on a probability space (Ω, \mathcal{F}, P) and observed for $t \in [0, T]$, where Y_t is real valued and X_t takes values in a semimetric vectorial space \mathcal{H} equipped with the semimetric $d(\cdot, \cdot)$. 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 [18–20] for conditions giving the existence of the conditional probability). Throughout this paper, \mathcal{E} denotes a compact set of \mathcal{H} . Let

Ψ be a real valued Borel function defined on \mathbb{R} and consider the generalized regression function

$$r(x) := E(\Psi(Y_0) | X_0 = x), \quad x \in \mathcal{E}. \quad (1)$$

We aim to estimate r from $\{X_t, Y_t\}_{t \in [0, T]}$.

We gather hereafter the assumptions that are needed to establish our result.

(H1) For any $x \in \mathcal{X}$ and any $h > 0$, set $\mathcal{B}(x, h) := \{y \in \mathcal{E}; d(y, x) \leq h\}$. There exist three constants $(c_1, C, \eta) \in (0, \infty)^2 \times]0, 1]$ such that, for any $x \in \mathcal{E}$ and any $(u, v) \in \mathcal{B}(x, c_1)^2$, we have

$$|r(u) - r(v)| \leq Cd(u, v)^\eta. \quad (2)$$

(H2) There exist

(i) a function ϕ and three constants $(\beta_1, \beta_2, c_2) \in [0, \infty)^3$ such that, for any $x \in \mathcal{E}$ and any $h \in (0, c_2]$, we have

$$0 < \beta_1 \phi(h) \leq P(X_0 \in \mathcal{B}(x, h)) \leq \beta_2 \phi(h), \quad (3)$$

(ii) a constant $c_3 > 0$ and a function g_0 integrable on $(0, \infty)$ such that, for any $x \in \mathcal{E}$, any $s > t \geq 0$, and any $h \in (0, c_3]$, we have

$$\left| P\left((X_t, X_s) \in \mathcal{B}(x, h)^2\right) - P(X_t \in \mathcal{B}(x, h))^2 \right| \leq g_0(s - t) \phi(h)^2. \quad (4)$$

(H3) For any $t \geq 0$, we set $\varepsilon_t := \Psi(Y_t) - E(\Psi(Y_t) | X_t)$. There exists an integrable bounded function g_1 on $[0, \infty)$ such that, for any $(s, t) \in [0, \infty)^2$, we have

$$\max\{|E(\varepsilon_s | X_s, X_t)|, |E(\varepsilon_s \varepsilon_t | X_s, X_t)|\} \leq g_1(|s - t|). \quad (5)$$

(H4) Let \mathcal{F}_T be the sigma-algebra generated by $\{X_t, t \in [0, T]\}$. There exists a constant $R > 0$, not depending on T , such that

$$\sup_{t \in [0, T]} E(\Psi(Y_t)^2 | \mathcal{F}_T) < R. \quad (6)$$

3. Estimator and Result

We define the generalized regression function estimate by

$$\hat{r}_T(x) := \begin{cases} \frac{\int_{t=0}^T \Psi(Y_t) K(h_T^{-1}d(x, X_t)) dt}{\int_{t=0}^T K(h_T^{-1}d(x, X_t)) dt} \\ \text{if } \int_{t=0}^T K(h_T^{-1}d(x, X_t)) dt \neq 0, \\ \frac{\int_{t=0}^T \Psi(Y_t) dt}{T} \quad \text{otherwise,} \end{cases} \quad (7)$$

where $K(x) = \mathbb{1}_{[0,1]}(x)$ is the indicator function on $[0, 1]$ and h_T is a bandwidth decreasing to 0 when $T \rightarrow \infty$. Remark that this estimator is the same as the one defined in [21, page 130] with the use of the semimetric d instead of the simple difference used in the real case.

Theorem 1 explores the performance of $\hat{r}_T(x)$ in terms of mean-square error.

Theorem 1. *Suppose that (H1)–(H4) hold. Let r be (1) and \hat{r}_T be (7) defined with $h_T = \mathcal{O}(T^{-1/\eta})$. Then, one has*

$$\sup_{x \in \mathcal{E}} E(\hat{r}_T(x) - r(x))^2 = \mathcal{O}\left(\frac{1}{T}\right). \quad (8)$$

We can compare this rate of convergence with the one obtained for discrete time processes in [14], which is, with our notations,

$$(\hat{r}_n(x) - r(x))^2 = \mathcal{O}\left(h_n^2 + \frac{1}{n\phi(h_n)}\right). \quad (9)$$

Remark that, with infinite-dimensional variables, $\phi(h)$ can decrease to zero, when h tends to zero, at an exponential rate so that h_n have to tend to zero at a logarithmic rate.

4. Comments and Examples

(H1) is a very classical Hölderian condition on the true regression function, but, in the infinite-dimensional framework, this condition depends on the semimetric used.

The assumption on small balls probabilities given in (H2)-(i) is widely used in nonparametric estimation for functional data (see, e.g., the monograph [22]). However, we want to point out the fact that if we define equivalence classes using the semidistance d , we can construct a quotient space on which d is a distance and if this quotient space is infinite-dimensional, then this condition can be satisfied only very locally in that for any point x of our compact \mathcal{E} , we can find, for any $\varepsilon > 0$, a point y and a positive number $h < \varepsilon$ such that $d(x, y) < \varepsilon$ and $P(X_0 \in \mathcal{B}(y, h)) \leq \beta_1 \phi(h)$: in that case, we could not extend our hypothesis to every point in an open ball (see [23] for a result on the consequences of a similar hypothesis on every point in a ball).

The most specific and restrictive assumption is (H2)-(ii), which is an adaptation to infinite-dimensional processes of the conditions on the density function introduced in [17] for real valued processes and transposed in [21, pages 135-136] to the estimation of the regression function with a vectorial regressor. Note that when $\mathcal{X} = \mathbb{R}^d$ and $\phi(h) = h^d$, the rate of convergence obtained in Theorem 5.3 in [21, page 136] is the same as the one we obtain here, and the condition I2 used implies (H2)-(ii). On the other hand, processes can meet (H2)-(ii) and infringe the condition in [21], especially when the vectorial process X_t does not admit a density. For real valued processes, a slightly different version of the Castellana and Leadbetter hypothesis on the joint density is given in [24] where it is shown that this hypothesis is satisfied for a wide class of diffusion processes, including the Ornstein-Uhlenbeck Process: these processes are also examples of

the range of applications of our result. Real continuous-time fractional ARMA processes studied in [25] are given as examples in [26]. Depending on the choice of the impulse response functions, a vector composed of such d processes can fulfil (H2)-(ii) for any d : using the notations of [25], if $((X_{1,T}), \dots, (X_{d,T}))$ are d independent processes complying with conditions of Proposition 4 in [25] with $D > -1/2 - 1/d$ and $a > 0$, then the vectorial process $((X_{1,T}), \dots, (X_{d,T}))$ meets (H2)(ii). For processes valued in infinite-dimensional spaces, we can also give the example of hidden processes: let (Z_t) be a nonobserved process valued in \mathbb{R}^d , for which conditions of Theorem 5.3 in [21, page 136] hold for every x in a compact \mathcal{A} , let Γ be an unknown function from \mathbb{R}^d to a space \mathcal{H} (that can be infinite-dimensional) equipped with a semimetric d , and let $(X_t) = (\Gamma(Z_t))$ be an observed process. If there exist two positive constants (a, b) such that for any $(x, y) \in \mathbb{R}^d \times \mathbb{R}^d$, $a\|x - y\| \leq d(\Gamma(x), \Gamma(y)) \leq b\|x - y\|$, then (X_t) fulfills (H2) with $\phi(h) = h^d$ and $\mathcal{C} = \Psi(\mathcal{A})$. Note that even if $\mathcal{H} = \mathbb{R}^{d'}$ with $d' > d$, (X_t) does not satisfy the assumptions usually imposed to vectorial processes to obtain a superoptimal rate.

There are two conditions in (H3). The condition $|E(\varepsilon_s | X_s, X_t)| \leq g_1(|s - t|)$ is less restrictive than imposing that the regressor and the noise are independent. $|E(\varepsilon_s \varepsilon_t | X_s, X_t)| \leq g_1(|s - t|)$ is a weak condition on the decay of dependence as the distance between observations increases, and (ε_t) may not be α -mixing. Note that we do not impose to (ε_t) to be an irregular path process.

At last, it is much less restrictive to impose (H4) than to suppose that $\Psi(Y_t)$ is bounded. In particular, this assumption allows us to consider the model

$$\Psi(Y_t) = r(X_t) + \varepsilon_t, \tag{10}$$

where r is a bounded function, (ε_t) is a square integrable process, and ε_t and (X_t) are independent.

On a given space, we can define many semidistances and hypotheses (H1)-(H2,) as well as the estimator itself, depending largely on the choice of this semidistance: the importance of this choice is widely discussed in [22] and a method to choose the semimetric for independent variables is proposed in [27], but this method does not ensure that (H1) holds. Actually, we can obtain a semimetric d such that $d(x, y) = 0 \not\Rightarrow r(x) = r(y)$. It would be of interest to develop a data driven method adapted to continuous time processes to select the semimetric.

In the statement of our theorem, we impose that $h_T = \mathcal{T}^{-1/\eta}$ where η is an unknown parameter so that the adaptation to continuous time processes of the method developed in [28] to choose the bandwidth would be interesting but is not in theory necessary in our framework. In point of fact, and it is what was very surprising when Castellana and Leadbetter first obtained a superoptimal rate of convergence, the bound for the variance of the estimator does not depend on h_t and we can choose $h_T = T^{-\log(T)}$ which will always satisfy $h_T = \mathcal{T}^{-1/\eta}$ for T large enough: even if this choice has no reason to be optimal, it leads to the claimed superoptimal rate of convergence.

Recently, results have been obtained in the case where the response Y is valued in a Banach space, which can be infinite-dimensional (see [29, 30]). Note that until Ψ is a real valued Borelian function, there is no need to change our proofs to obtain our result if Y is valued in a Banach space. However, in the case where $\Psi(Y)$ is a Banach valued variable, we could not easily adapt our proofs and obtaining a superoptimal rate would involve very different techniques; it would be an interesting extension for further works.

5. Simulations

We chose $\mathbb{L}^2([-1, 1])$ endowed with its natural norm as the functional space and simulated our process as follows.

At first we simulated an Ornstein-Uhlenbeck process solution of the stochastic differential equation

$$dOU_t = -9(OU_t - 2)dt + 6dW_t, \tag{11}$$

where W_t denotes a Wiener process. Here, we took $dt = 0.0005$.

Denoting the floor function by $\lfloor \cdot \rfloor$, let Γ be the function from \mathbb{R} to $\mathbb{L}^2([-1, 1])$ defined by

$$\Gamma(x) := (1 + \lfloor x \rfloor - x) P_{\text{num}(\lfloor x \rfloor)} + (x - \lfloor x \rfloor) P_{\text{num}(\lfloor x \rfloor + 1)} \quad \forall x \in \mathbb{R}, \tag{12}$$

where P_i is the Legendre polynomial of degree i and $\text{num}(x) := 1 + 2 \times \text{sign}(x) \times x - \text{sign}(x) \times (1 + \text{sign}(x))/2$. Then we define our functional process for any $t \in [0, T]$ setting

$$X_t := \Gamma(OU_t). \tag{13}$$

For any square integrable function x on $[-1, 1]$, we define the function

$$r(x) := \int_{u=-1}^1 x(u) (2u + x(u)) du \tag{14}$$

and set

$$Y_t = r(X_t) + U_t, \tag{15}$$

where $U_t = W'_t - W'_{T-1}$ and W'_t is a Wiener process independent of X .

In order to obtain a panel of 20 points (in $\mathbb{L}^2([-1, 1])$) where we can evaluate the regression function, we did a first simulation with $T = 10$ and set $\mathcal{C} := (X_{i/2}, i \in 1, 2, \dots, 20)$. Once obtained, \mathcal{C} is considered as a deterministic set. We represent these functions in Figure 1.

Remark. We check here that the simulated processes fulfil our hypotheses.

At first, denoting by Id the identity function on $[-1, 1]$, for any $(x, y) \in \mathbb{L}^2([-1, 1]) \times \mathbb{L}^2([-1, 1])$, we have

$$\begin{aligned} |r(x) - r(y)| &= \left| \|x\|^2 - \|y\|^2 + 2 \langle x - y, \text{Id} \rangle \right| \\ &\leq (\|x + y\| + 2) \|x - y\| \end{aligned} \tag{16}$$

and r satisfies (H1) with $\eta = 1$.

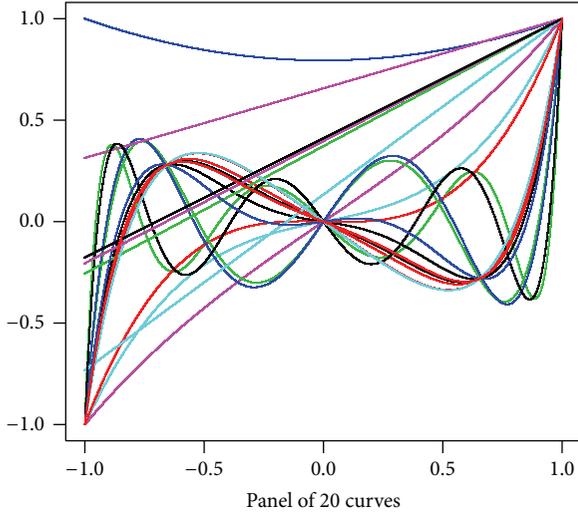


FIGURE 1

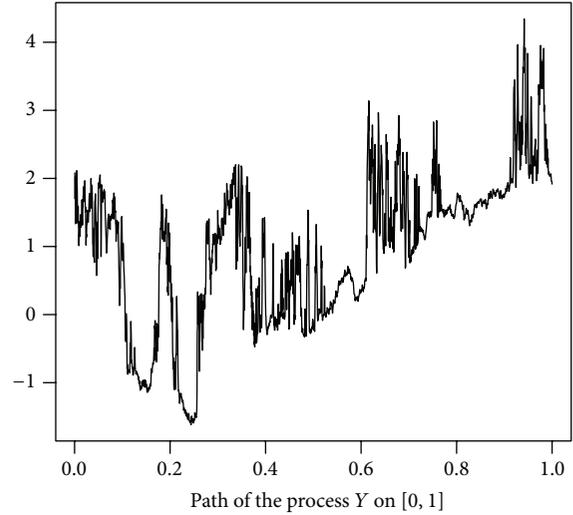


FIGURE 2

The Ornstein-Uhlenbeck process satisfies the part of Condition I2 on the regressor’s density in [21, page 136]. Moreover, Γ is a bijection from \mathbb{R} to $\text{Im}(\Gamma)$, and it can be shown that, for some constant C , there exist $0 < a < b$ such that for any $0 < h < C$ and any $x \in \Gamma^{-1}(\mathcal{C})$, the two following implications are correct:

$$\begin{aligned} (z \in \mathcal{B}(x, ah)) &\implies (\Gamma(z) \in \mathcal{B}(\Gamma(x), h)), \\ (\Gamma(z) \in \mathcal{B}(x, h)) &\implies (z \in \mathcal{B}(x, bh)), \end{aligned} \tag{17}$$

which implies that (H2)(i)-(ii) are fulfilled when taking $\phi(h) = h$.

Since (ε_t) and (X_t) are independent and $\text{Cov}(\varepsilon_t, \varepsilon_s) = 0$ if $|t - s| > 1$, (H3) is satisfied.

Finally, the model used in the simulation corresponds to the choice of the identity function for Ψ in (1), where (Y_t) is an unbounded process and $r(\cdot)$ is not a bounded function. However, $r(\cdot)$ is bounded on $\text{Im}(\Gamma)$ and so (H4) is fulfilled.

We simulated the paths of the process $(X_t, Y_t)_{t \in [0, T]}$ for different values of T . Figure 2 represents the path of the process (Y_t) for $t \in [0, 1]$.

We estimated the regression function at each point in \mathcal{C} , for different values of T , and compared our results to those obtained when studying a discrete time functional process, that is, when we observe (X_t, Y_t) only for $t \in \mathbb{N}$, and we use the estimator defined in [12] with the indicator function as the kernel: it corresponds to an infinite-dimensional version of Nadaraya-Watson estimator with a uniform kernel. When working with the discrete time process we used the data-driven way of choosing the bandwidth proposed in [28]. When working with the continuous time process that is observed on a very thin grid, for $T = 50$, we chose the same bandwidth as the one used for the discrete time process and, for $T > 50$, we supposed r to be Lipschitz (i.e., $\eta = 1$, which is the case here) and used the bandwidth $h_T = h_{50}(50/T)$. In Table 1, we give the mean square error evaluated on the functions of the panel for different $T = 50, 500$, and 2000.

TABLE 1

	Continuous time process	Discrete time process
$T = 50$	0.056623	0.231032
$T = 500$	0.003235	0.037855
$T = 2000$	0.000698	0.0155137

We can see that, for $T = 50$, we already have a smaller mean square error with the estimator using the continuous time process, and when T increase, the mean square error seems to decrease much more quickly when working with the continuous time process. However, the continuous time approach takes much more time and much more memory; we had to split the calculation into several parts and delete intermediate calculations to avoid saturating memory.

In Figures 3 and 4, we have in abscissa the value of the real regression function applied to each function of our panel and in ordinate the estimated value of the regression function. We represent on the left the results for the continuous time estimator and on the right the results for the discrete time estimator.

6. Proofs

6.1. *Intermediary Results.* In the sequel, we use the following notations:

$$\begin{aligned} \Delta_{T,t}(x) &= K(h_T^{-1}d(x, X_t)), \\ \hat{r}_{1,T}(x) &:= \frac{1}{TE(\Delta_{T,0}(x))} \int_{t=0}^T \Delta_{T,t}(x) dt, \\ \hat{r}_{2,T}(x) &:= \frac{1}{TE(\Delta_{T,0}(x))} \int_{t=0}^T \Psi(Y_t) \Delta_{T,t}(x) dt. \end{aligned} \tag{18}$$

Lemma 2 below studies the behavior of the bias of $\hat{r}_{2,T}$.

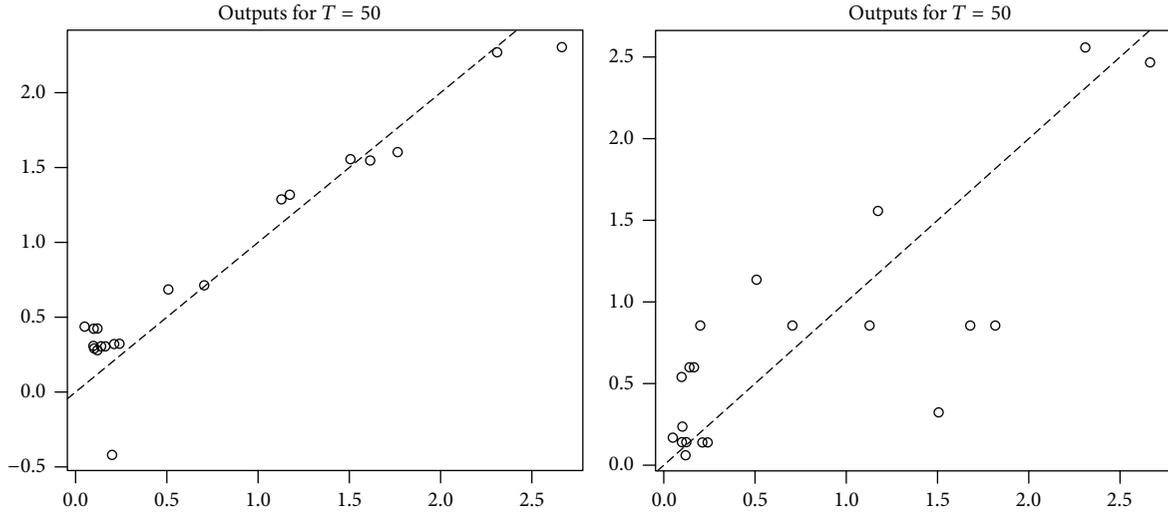


FIGURE 3: Continuous time estimator (left) and discrete time estimator (right); in abscissa the value of the real regression function applied to each function of our panel and in ordinate the estimated value of the regression function.

Lemma 2. Under the conditions of Theorem 1, one has

$$\sup_{x \in \mathcal{E}} |E(\hat{r}_{2,T}(x)) - r(x)| = \mathcal{O}\left(\frac{1}{T}\right). \quad (19)$$

Lemma 3 below provides an upper bound for the variances of $\hat{r}_{1,T}$ and $\hat{r}_{2,T}$.

Lemma 3. Under the conditions of Theorem 1, one has

$$\sup_{x \in \mathcal{E}} (\text{Var}(\hat{r}_{2,T}(x)) + \text{Var}(\hat{r}_{1,T}(x))) = \mathcal{O}\left(\frac{1}{T}\right). \quad (20)$$

6.2. *Proofs of the Intermediary Results.* For the sake of conciseness, when no confusion is possible, we use the notations $\Psi_t := \Psi(Y_t)$ and $\Delta_{T,t} := \Delta_{T,t}(x)$.

Proof of Lemma 2. Observe that, for any $x \in \mathcal{E}$,

$$\begin{aligned} E(\hat{r}_{2,T}(x)) &= \frac{1}{TE(\Delta_{T,0})} \int_{t=0}^T E(\Psi_t \Delta_{T,t}) dt \\ &= \frac{E(\Psi_0 \Delta_{T,0})}{E(\Delta_{T,0})} = \frac{E(E(r(X_0) + \varepsilon_0 | X_0) \Delta_{T,0})}{E(\Delta_{T,0})} \\ &= \frac{E(r(X_0) \Delta_{T,0})}{E(\Delta_{T,0})}. \end{aligned} \quad (21)$$

Hence,

$$\begin{aligned} E(\hat{r}_{2,T}(x)) - r(x) &= \frac{E(r(X_0) \Delta_{T,0})}{E(\Delta_{T,0})} - r(x) \\ &= \frac{E((r(X_0) - r(x)) \Delta_{T,0})}{E(\Delta_{T,0})}. \end{aligned} \quad (22)$$

Owing to (H1), we have $|r(X_0) - r(x)| \Delta_{T,0} \leq \Delta_{T,0} \sup_{u \in \mathcal{B}(x, h_T)} |r(u) - r(x)| \leq C \Delta_{T,0} h_T^\eta$. Therefore, by Jensen's inequality and $h^\eta = \mathcal{O}(1/T)$, we have

$$\begin{aligned} \sup_{x \in \mathcal{E}} |E(\hat{r}_{2,T}(x)) - r(x)| \\ \leq \sup_{x \in \mathcal{E}} \frac{E(|r(X_0) - r(x)| \Delta_{T,0})}{E(\Delta_{T,0})} \leq C h_T^\eta = \mathcal{O}\left(\frac{1}{T}\right). \end{aligned} \quad (23)$$

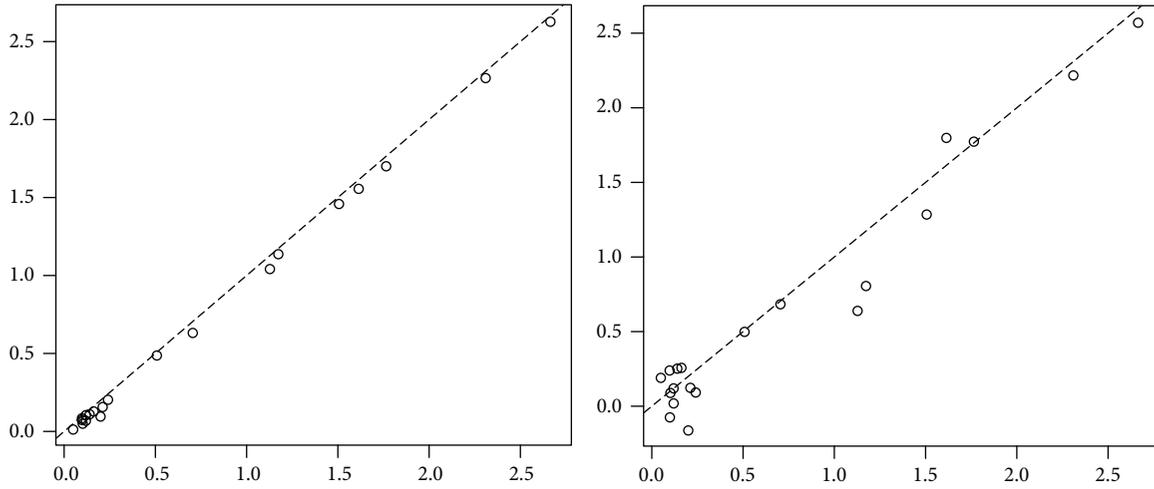
This ends the proof of Lemma 2. \square

Proof of Lemma 3. For any $x \in \mathcal{E}$, by Fubini's Theorem, we have

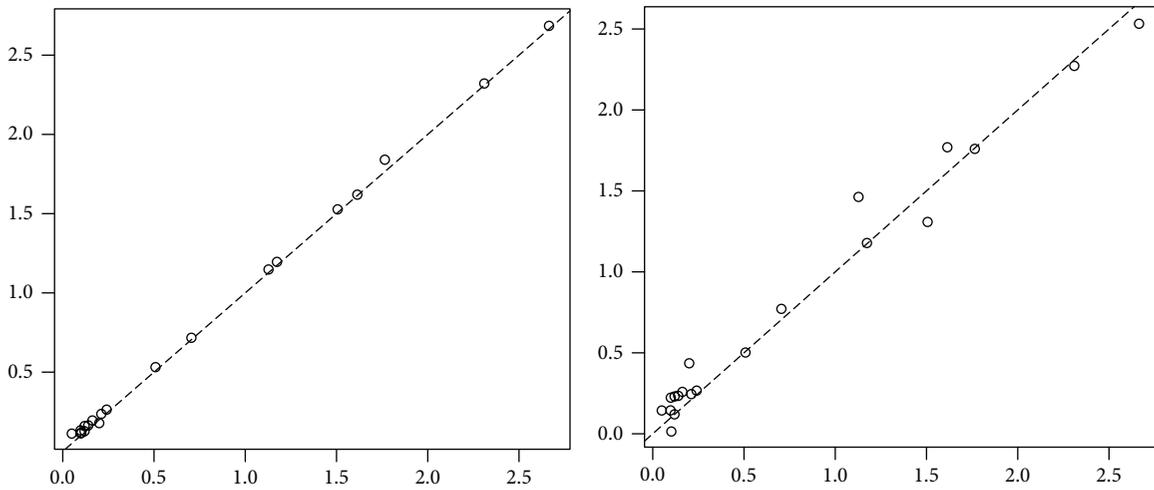
$$\begin{aligned} \text{Var}(\hat{r}_{2,T}(x)) \\ = \frac{1}{T^2 E(\Delta_{T,0})^2} \int_{t=0}^T \int_{s=0}^T \text{Cov}(\Psi_s \Delta_{T,s}, \Psi_t \Delta_{T,t}) dt ds. \end{aligned} \quad (24)$$

Upper Bound of the Covariance Term. In order to simplify the notations, we set $R(X_t) := E(\Psi_t | X_t)$ and $\varepsilon_t := \Psi_t - R_t$. Note that

$$\begin{aligned} E(\Psi_s \Psi_t | X_s, X_t) &= R(X_s) R(X_t) + R(X_s) E(\varepsilon_t | X_s, X_t) \\ &\quad + R(X_t) E(\varepsilon_s | X_s, X_t) \\ &\quad + E(\varepsilon_s \varepsilon_t | X_s, X_t). \end{aligned} \quad (25)$$



(a) Outputs for $T = 500$



(b) Outputs for $T = 2000$

FIGURE 4: Continuous time estimator (left) and discrete time estimator (right); in abscissa the value of the real regression function applied to each function of our panel and in ordinate the estimated value of the regression function.

Therefore, the covariance term can be expended as follows:

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

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

(26)

Set

$$d_t := R(X_t) - r(x). \tag{27}$$

We have

$$\begin{aligned}
 \text{Cov}(\Psi_s \Delta_{T,s}, \Psi_t \Delta_{T,t}) &= r(x)^2 E(\Delta_{T,s} \Delta_{T,t}) \\
 &\quad + r(x) (E(\Delta_{T,s} \Delta_{T,t} d_t) \\
 &\quad \quad + E(\Delta_{T,s} \Delta_{T,t} d_s))
 \end{aligned}$$

$$\begin{aligned}
 &+ E(\Delta_{T,s}\Delta_{T,t}d_t d_s) + r(x) \\
 &\times E(\Delta_{T,s}\Delta_{T,t}(E(\varepsilon_t | X_s, X_t) \\
 &\quad + E(\varepsilon_s | X_s, X_t))) \\
 &+ E(\Delta_{T,s}\Delta_{T,t} \\
 &\quad \times (d_s E(\varepsilon_t | X_s, X_t) \\
 &\quad + d_t E(\varepsilon_s | X_s, X_t))) \\
 &+ E(\Delta_{T,s}\Delta_{T,t}E(\varepsilon_s \varepsilon_t | X_s, X_t)) \\
 &- r(x)^2 E(\Delta_{T,s})^2 - E(\Delta_{T,s}d_s)^2 \\
 &- 2r(x)E(\Delta_{T,s}d_s)E(\Delta_{T,s}) \\
 = &r(x)^2 (E(\Delta_{T,s}\Delta_{T,t}) - E(\Delta_{T,s})^2) \\
 &- (E(\Delta_{T,s}d_s)^2 \\
 &\quad + 2r(x)E(\Delta_{T,s}d_s)E(\Delta_{T,s})) \\
 &+ E(\Delta_{T,s}\Delta_{T,t}Q), \tag{28}
 \end{aligned}$$

with

$$\begin{aligned}
 Q = &d_s E(\varepsilon_t | X_s, X_t) + d_t E(\varepsilon_s | X_s, X_t) \\
 &+ d_s d_t + E(\varepsilon_s \varepsilon_t | X_s, X_t) \\
 &+ r(x)(d_s + d_t + E(\varepsilon_t | X_s, X_t) + E(\varepsilon_s | X_s, X_t)). \tag{29}
 \end{aligned}$$

The triangular inequality and Jensen's inequality yield

$$|\text{Cov}(\Psi_s \Delta_{T,s}, \Psi_t \Delta_{T,t})| \leq L + M + N, \tag{30}$$

where

$$\begin{aligned}
 L = &r(x)^2 |E(\Delta_{T,s}\Delta_{T,t}) - E(\Delta_{T,s})^2|, \\
 M = &E(\Delta_{T,s} |d_s|)^2 + 2|r(x)|E(\Delta_{T,s} |d_s|)E(\Delta_{T,s}), \tag{31} \\
 N = &E(\Delta_{T,s}\Delta_{T,t} |Q|).
 \end{aligned}$$

Upper Bound for L. Using (H2)-(ii), we have

$$L \leq r(x)^2 g_0(|s-t|)\phi(h_T)^2. \tag{32}$$

Upper Bound for M. Owing to (H1), we have $\Delta_{T,s}|d_s| \leq \Delta_{T,s} \sup_{u \in \mathcal{B}(x, h_T)} |r(u) - r(x)| \leq C \Delta_{T,s} h_T^\eta$. It follows from this inequality and (H2)-(i) that

$$\begin{aligned}
 M \leq &(2|r(x)|Ch_T^\eta + C^2 h_T^{2\eta})E(\Delta_{T,s})^2 \\
 \leq &(2r(x)Ch_T^\eta + C^2 h_T^{2\eta})\beta_2^2 \phi(h_T)^2. \tag{33}
 \end{aligned}$$

Upper Bound for N. By similar techniques to those in the bound for M and (H3), we obtain

$$\begin{aligned}
 \Delta_{T,s}\Delta_{T,t} |Q| \leq &\Delta_{T,s}\Delta_{T,t} \\
 &\times (2|r(x)|Ch_T^\eta + C^2 h_T^{2\eta} \\
 &\quad + (2(|r(x)| + Ch_T^\eta) + 1)g_1(|s-t|)). \tag{34}
 \end{aligned}$$

On the other hand, by (H2)-(ii),

$$\begin{aligned}
 E(\Delta_{T,s}\Delta_{T,t}) \leq &|\text{Cov}(\Delta_{T,s}, \Delta_{T,t})| + E(\Delta_{T,s})^2 \\
 \leq &(\beta_2^2 + g_0(|s-t|))\phi(h_T)^2. \tag{35}
 \end{aligned}$$

Hence,

$$\begin{aligned}
 N \leq &(2|r(x)|Ch_T^\eta + C^2 h_T^{2\eta} \\
 &\quad + (2(|r(x)| + Ch_T^\eta) + 1)g_1(|s-t|)) \\
 &\quad \times (\beta_2^2 + g_0(|s-t|))\phi(h_T)^2. \tag{36}
 \end{aligned}$$

Therefore, setting

$$\begin{aligned}
 G_T(y) := &r(x)^2 g_0(y)\phi(h_T)^2 \\
 &+ (2|r(x)|Ch_T^\eta + C^2 h_T^{2\eta} \\
 &\quad + (2(|r(x)| + Ch_T^\eta) + 1)g_1(y)) \\
 &\quad \times (\beta_2^2 + g_0(y))\phi(h_T)^2 \\
 &+ (2r(x)Ch_T^\eta + C^2 h_T^{2\eta})\beta_2^2 \phi(h_T)^2, \tag{37}
 \end{aligned}$$

the obtained upper bounds for L, M, and N yield

$$\begin{aligned}
 |\text{Cov}(\Psi_s \Delta_{T,s}, \Psi_t \Delta_{T,t})| \\
 \leq G_T(|s-t|). \tag{38}
 \end{aligned}$$

Final Bound. Combining (24) and (38) and using (H2)-(i), we have

$$\begin{aligned}
 \text{Var}(\tilde{r}_{2,T}(x)) \leq &\frac{2}{T^2 E(\Delta_{T,0})^2} \int_{t=0}^T \int_{s=t}^T G_T(s-t) dt ds \\
 \leq &\frac{2}{TE(\Delta_{T,0})^2} \int_{y=0}^T G_T(y) dy \tag{39} \\
 \leq &\frac{2}{T\beta_1^2 \phi(h_T)^2} \int_{y=0}^T G_T(y) dy.
 \end{aligned}$$

Since g_0 and g_1 are integrable and r is bounded on \mathcal{C} and $h^\eta = \mathcal{O}(1/T)$, there exists a constant C_0 such that

$$\sup_{x \in \mathcal{C}} \text{Var}(\tilde{r}_{2,T}(x)) \leq \frac{C_0}{T}. \tag{40}$$

The special choice of $\Psi : (x) \mapsto 1$ leads us to

$$\sup_{x \in \mathcal{E}} \text{Var}(\hat{r}_{1,T}(x)) \leq \frac{C_1}{T}. \tag{41}$$

This last inequality concludes the proof of Lemma 3. \square

Proof of Theorem 1. We can write

$$\begin{aligned} \hat{r}_T(x) - r(x) &= (\hat{r}_T(x) - \hat{r}_{2,T}(x)) \\ &\quad + (\hat{r}_{2,T}(x) - E(\hat{r}_{2,T}(x))) \\ &\quad + (E(\hat{r}_{2,T}(x)) - r(x)). \end{aligned} \tag{42}$$

The elementary inequality: $(a + b + c)^2 \leq 3(a^2 + b^2 + c^2)$, $(a, b, c) \in (0, \infty)^3$, yields

$$\sup_{x \in \mathcal{E}} E(\hat{r}_T(x) - r(x))^2 \leq 3(U + V + W), \tag{43}$$

where

$$\begin{aligned} U &= \sup_{x \in \mathcal{E}} E(\hat{r}_T(x) - \hat{r}_{2,T}(x))^2, \\ V &= \sup_{x \in \mathcal{E}} E(\hat{r}_{2,T}(x) - E(\hat{r}_{2,T}(x)))^2, \\ W &= \sup_{x \in \mathcal{E}} (E(\hat{r}_{2,T}(x)) - r(x))^2. \end{aligned} \tag{44}$$

Upper Bound for V. Lemma 3 yields

$$V = \sup_{x \in \mathcal{E}} \text{Var}(\hat{r}_{2,T}(x)) = \mathcal{O}\left(\frac{1}{T}\right). \tag{45}$$

Upper Bound for W. Lemma 2 yields

$$W = \mathcal{O}\left(\frac{1}{T^2}\right). \tag{46}$$

Upper Bound for U. We define, for any $t \in [0, T]$, the quantity:

$$Z_t := \begin{cases} \frac{K(h_T^{-1}d(x, X_t))}{\int_{t=0}^T K(h_T^{-1}d(x, X_t)) dt} \\ \text{if } \int_{t=0}^T K(h_T^{-1}d(x, X_t)) dt \neq 0, \\ \frac{1}{T} \text{ otherwise.} \end{cases} \tag{47}$$

Note that, when $\hat{r}_{1,T}(x) \neq 0$,

$$\hat{r}_{2,T}(x) = \hat{r}_T(x) \times \hat{r}_{1,T}(x), \tag{48}$$

so that

$$\begin{aligned} U &\leq \sup_{x \in \mathcal{E}} E(\hat{r}_T(x) (1 - \hat{r}_{1,T}(x)))^2 \\ &\quad + E\left(\mathbb{1}_{\{0\}}(\hat{r}_{1,T}(x)) \int_0^T \frac{\Psi(Y_t)}{T} dt\right). \end{aligned} \tag{49}$$

Using (H4) and Lemma 3, we get

$$\begin{aligned} &\sup_{x \in \mathcal{E}} E(\hat{r}_T(x) (1 - \hat{r}_{1,T}(x)))^2 \\ &= \sup_{x \in \mathcal{E}} E\left(E(\hat{r}_T^2(x) (1 - \hat{r}_{1,T}(x))^2 \mid \mathcal{F}_T)\right) \\ &= \sup_{x \in \mathcal{E}} E\left(E(\hat{r}_T^2(x) \mid \mathcal{F}_T)\right. \\ &\quad \left. \times (1 - \hat{r}_{1,T}(x))^2\right) \\ &= \sup_{x \in \mathcal{E}} E\left(E\left(\left(\int_{t=0}^T Z_t \Psi(Y_t) dt\right)^2 \mid \mathcal{F}_T\right)\right. \\ &\quad \left. \times (1 - \hat{r}_{1,T}(x))^2\right) \\ &= \sup_{x \in \mathcal{E}} E\left(\int_{(s,t) \in [0,T]^2} Z_t Z_s E(\Psi(Y_t) \Psi(Y_s) \mid \mathcal{F}_T) ds dt\right. \\ &\quad \left. \times (1 - \hat{r}_{1,T}(x))^2\right) \\ &\leq \sup_{x \in \mathcal{E}} E\left(\int_{(s,t) \in [0,T]^2} Z_t Z_s R ds dt\right. \\ &\quad \left. \times (1 - \hat{r}_{1,T}(x))^2\right) \\ &\leq \sup_{x \in \mathcal{E}} RE\left((1 - \hat{r}_{1,T}(x))^2\right) \\ &= \sup_{x \in \mathcal{E}} R \text{Var}(\hat{r}_{1,T}(x)) = \mathcal{O}\left(\frac{1}{T}\right). \end{aligned} \tag{50}$$

Similarly, (H4), Lemma 3, and Chebyshev's inequality lead to

$$\begin{aligned} &\sup_{x \in \mathcal{E}} E\left(\mathbb{1}_{\{0\}}(\hat{r}_{1,T}(x)) \int_0^T \frac{\Psi(Y_t)}{T} dt\right) \\ &\leq \sup_{x \in \mathcal{E}} RE(\mathbb{1}_{\{0\}}(\hat{r}_{1,T}(x))) \\ &\leq \sup_{x \in \mathcal{E}} RP(|\hat{r}_{1,T}(x) - 1| \geq 1) \\ &\leq \frac{R}{T}. \end{aligned} \tag{51}$$

We finally obtain

$$U = \mathcal{O}\left(\frac{1}{T}\right). \tag{52}$$

Putting the obtained upper bounds for U , V , and W together, we have

$$\sup_{x \in \mathcal{E}} E(\hat{r}_T(x) - r(x))^2 = \mathcal{O}\left(\frac{1}{T}\right). \tag{53}$$

Theorem 1 is proved. \square

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

The authors wish to thank the Editor and two anonymous referees for their constructive suggestions which led to some improvements in some earlier versions of the paper.

References

- [1] E. Nadaraya, "On estimating regression," *Theory of Probability and Its Applications*, vol. 9, pp. 141–142, 1964.
- [2] G. Watson, "Smooth regression analysis," *Sankhyā*, vol. 26, no. 4, pp. 359–372, 1964.
- [3] M. Rosenblatt, "Conditional probability density and regression estimators," in *Multivariate Analysis II (Proceedings of the 2nd International Symposium, Dayton, Ohio, 1968)*, pp. 25–31, Academic Press, New York, NY, USA, 1969.
- [4] C. J. Stone, "Optimal global rates of convergence for nonparametric regression," *The Annals of Statistics*, vol. 10, no. 4, pp. 1040–1053, 1982.
- [5] G. Collomb and W. Härdle, "Strong uniform convergence rates in robust nonparametric time series analysis and prediction: kernel regression estimation from dependent observations," *Stochastic Processes and their Applications*, vol. 23, no. 1, pp. 77–89, 1986.
- [6] A. Krzyzak and M. Pawlak, "The pointwise rate of convergence of the kernel regression estimate," *Journal of Statistical Planning and Inference*, vol. 16, no. 2, pp. 159–166, 1987.
- [7] G. G. Roussas, "Nonparametric regression estimation under mixing conditions," *Stochastic Processes and Their Applications*, vol. 36, no. 1, pp. 107–116, 1990.
- [8] D. Bosq, "Vitesses optimales et superoptimales des estimateurs fonctionnels pour les processus à temps continu," *Comptes Rendus de l'Académie des Sciences I: Mathematics*, vol. 317, no. 11, pp. 1075–1078, 1993.
- [9] J. O. Ramsay and B. W. Silverman, *Functional Data Analysis*, Springer Series in Statistics, Springer, New York, NY, USA, 2nd edition, 2005.
- [10] J. O. Ramsay and B. W. Silverman, *Applied Functional Data Analysis: Methods and Case Studies*, Springer Series in Statistics, Springer, New York, NY, USA, 2002.
- [11] L. Horváth and P. Kokoszka, *Inference for Functional Data with Applications*, Springer Series in Statistics, Springer, New York, NY, USA, 2012.
- [12] F. Ferraty and P. Vieu, "Nonparametric models for functional data, with application in regression, time-series prediction and curve discrimination," *Journal of Nonparametric Statistics*, vol. 16, no. 1-2, pp. 111–125, 2004, The International Conference on Recent Trends and Directions in Nonparametric Statistics.
- [13] E. Masry, "Nonparametric regression estimation for dependent functional data: asymptotic normality," *Stochastic Processes and Their Applications*, vol. 115, no. 1, pp. 155–177, 2005.
- [14] F. Ferraty, A. Mas, and P. Vieu, "Nonparametric regression on functional data: inference and practical aspects," *Australian & New Zealand Journal of Statistics*, vol. 49, no. 3, pp. 267–286, 2007.
- [15] F. Ferraty and P. Vieu, "Kernel regression estimation for functional data," in *The Oxford Handbook of Functional Data Analysis*, pp. 72–129, Oxford University Press, Oxford, UK, 2011.
- [16] D. Bosq, "Parametric rates of nonparametric estimators and predictors for continuous time processes," *The Annals of Statistics*, vol. 25, no. 3, pp. 982–1000, 1997.
- [17] J. V. Castellana and M. R. Leadbetter, "On smoothed probability density estimation for stationary processes," *Stochastic Processes and Their Applications*, vol. 21, no. 2, pp. 179–193, 1986.
- [18] M. Jirina, "Conditional probabilities on strictly separable σ -algebras," *Czechoslovak Mathematical Journal*, vol. 4, no. 79, pp. 372–380, 1954.
- [19] M. Jirina, "On regular conditional probabilities," *Czechoslovak Mathematical Journal*, vol. 9, no. 84, pp. 445–451, 1959.
- [20] R. Grunig, "Probabilités conditionnelles régulières sur des tribus de type non dénombrable," *Annales de l'institut Henri Poincaré B: Probabilités et Statistiques*, vol. 2, no. 3, pp. 227–229, 1966.
- [21] D. Bosq, *Nonparametric Statistics for Stochastic Processes: Estimation and Prediction*, vol. 110 of *Lecture Notes in Statistics*, Springer, New York, NY, USA, 2nd edition, 1998.
- [22] F. Ferraty and P. Vieu, *Nonparametric Functional Data Analysis: Theory and Practice*, Springer Series in Statistics, Springer, New York, NY, USA, 2006.
- [23] J. Azaïs and J.-C. Fort, "Remark on the finite-dimensional character of certain results of functional statistics," *Comptes Rendus Mathématique*, vol. 351, no. 3-4, pp. 139–141, 2013.
- [24] F. Leblanc, "Density estimation for a class of continuous time processes," *Mathematical Methods of Statistics*, vol. 6, no. 2, pp. 171–199, 1997.
- [25] M.-C. Viano, C. Deniau, and G. Oppenheim, "Continuous-time fractional ARMA processes," *Statistics and Probability Letters*, vol. 21, no. 4, pp. 323–336, 1994.
- [26] D. Blanke, "Sample paths adaptive density estimation," *Mathematical Methods of Statistics*, vol. 13, no. 2, pp. 123–152, 2004.
- [27] C. Timmermans, L. Delsol, and R. von Sachs, "Using Bagidis in nonparametric functional data analysis: predicting from curves with sharp local features," *Journal of Multivariate Analysis*, vol. 115, pp. 421–444, 2013.
- [28] K. Benhenni, F. Ferraty, M. Rachdi, and P. Vieu, "Local smoothing regression with functional data," *Computational Statistics*, vol. 22, no. 3, pp. 353–369, 2007.
- [29] F. Ferraty, A. Laksaci, A. Tadj, and P. Vieu, "Kernel regression with functional response," *Electronic Journal of Statistics*, vol. 5, pp. 159–171, 2011.
- [30] F. Ferraty, I. van Keilegom, and P. Vieu, "Regression when both response and predictor are functions," *Journal of Multivariate Analysis*, vol. 109, pp. 10–28, 2012.



Hindawi

Submit your manuscripts at
<http://www.hindawi.com>

