



HAL
open science

Super-optimal rate of convergence in non-parametric estimation for functional valued processes

Christophe Chesneau, Bertrand Maillot

► **To cite this version:**

Christophe Chesneau, Bertrand Maillot. Super-optimal rate of convergence in non-parametric estimation for functional valued processes. 2013. hal-00850123

HAL Id: hal-00850123

<https://hal.science/hal-00850123>

Preprint submitted on 3 Aug 2013

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

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

Super-optimal rate of convergence in non-parametric estimation for functional valued processes.

Christophe Chesneau* and Bertrand Maillot*

Abstract

In this paper, we consider the non-parametric estimation of the generalised regression function for continuous time processes with irregular paths when the regressor takes values in a semi-metric space. We establish the mean-square convergence of our estimator with the same super-optimal rate as when the regressor is real valued.

Keywords: Regression estimation; Functional variables; Infinite dimensional space; Small balls probabilities.

1 Introduction

Since the pioneer works of Nadaraya (1964) and Watson (1964), the nonparametric estimation of the regression function has been very widely studied for real and vectorial regressors (see, for example, Rosenblatt (1969), Stone (1982), Collomb et Härdle (1986), Krzyżak et Pawlak (1987), Roussas (1990) and Bosq (1993)) and more recently, the case when the regressor takes values in a semi-metric space of infinite dimension has been addressed (*e.g.* Ferraty et Vieu (2004), Masry (2005), Ferraty *et al.* (2007), Ferraty et Vieu (2011)). In the regression estimation framework, it is well known that the efficiency of a non parametric 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 space. However, when studying continuous time processes with irregular paths, it has been shown in Bosq (1997) that even when the regressor is \mathbb{R}^d -valued, we can estimate the regression function with the parametric rate of convergence $\mathcal{O}\left(\frac{1}{\sqrt{T}}\right)$. This kind of suroptimal rate of convergence for nonparametric estimators is always obtained under hypotheses on the joint probability density functions of the process which are very similar from those introduced by Castellana et Leadbetter (1986). 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 can not be applied. Under hypotheses on probabilities of small balls, we show that we can reach suroptimal 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. A numerical study can be found in Section 4. The proofs are postponed to Section 5.

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

2 Problem and assumptions

Let $\{X_t, Y_t\}_{t \in [0, \infty)}$ be a 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 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 Jirina (1954, 1959) and Grunig (1966) for conditions giving the existence of the conditional probability). Throughout this paper, \mathcal{C} 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

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

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) There exist three constants $(c_1, C, \eta) \in (0, \infty)^3$, such that, for any $x \in \mathcal{C}$ and any $(u, v) \in \mathcal{B}(x, c_1)^2$, we have

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

(H2) For any $x \in \mathcal{C}$ and any $h > 0$, set $\mathcal{B}(x, h) := \{y \in \mathcal{C}; d(y, x) \leq h\}$. There exist a function ϕ

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

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

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

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

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

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

(H4) There exists a constant $R > 0$ such that

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

Comments: (H1) is a very classical Hölderian condition on the true regression function. The assumption on small balls probabilities given in (H2)-(i) is widely used in non-parametric estimation for functional data (see, *e.g.*, the monograph by Ferraty et Vieu (2006)). In a functional framework, this condition can be satisfied only locally since if \mathcal{H} is infinite dimensional, this hypothesis implies that we cannot find any open ball included in \mathcal{C} . Assumptions (H2)-(ii) and (H3) are an adaptation to infinite dimensional processes of the conditions on the density function introduced in Castellana et Leadbetter (1986) for real valued processes. Finally, it is much less restrictive to impose (H4) than supposing that $\Psi(Y_t)$ is bounded (see examples in Bosq (1998) p.131).

3 Estimator and result

We define the generalized regression function estimate by

$$\hat{r}_T(x) := \begin{cases} \frac{\int_{t=0}^T \Psi(Y_t) K(h_T^{-1} d(x, X_t)) dt}{\int_{t=0}^T K(h_T^{-1} d(x, X_t)) dt} & \text{if } \int_{t=0}^T K(h_T^{-1} d(x, X_t)) dt \neq 0, \\ \frac{\int_{t=0}^T \Psi(Y_t) dt}{T} & \text{otherwise,} \end{cases} \quad (2)$$

where $K(x) = \mathbb{I}_{[0,1]}(x)$ 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 Bosq (1998) p.130 with the use of the semi-metric d instead of the simple difference used in the real case.

Theorem 3.1 explores the performance of $\hat{r}_T(x)$ in term of mean-square error.

Theorem 3.1 *Suppose that (H1)-(H4) hold. Let r be (1) and \hat{r}_T be (2) defined with $h_T = \mathcal{O}\left(T^{\frac{1}{\eta}}\right)$. Then, for any $x \in \mathcal{C}$, we have*

$$\mathbb{E}(\hat{r}_T(x) - r(x))^2 = \mathcal{O}\left(\frac{1}{T}\right). \quad (3)$$

4 Simulations

We simulated our functional valued process as follows.

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

$$dOU_t = -9(OU_t - 2)dt + 6dW_t,$$

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

Then, denoting the floor function by $\lfloor \cdot \rfloor$, we defined our functional process for any $t \in [0, T]$ setting

$$X_t := (1 + \lfloor OU_t \rfloor - OU_t) P_{num(\lfloor OU_t \rfloor)} + (OU_t - \lfloor OU_t \rfloor) P_{num(\lfloor OU_t \rfloor + 1)},$$

where P_i is the Legendre polynomial of degree n and

$$num(x) := 2 * sign(x) * x - sign(x) * (1 + sign(x))/2.$$

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

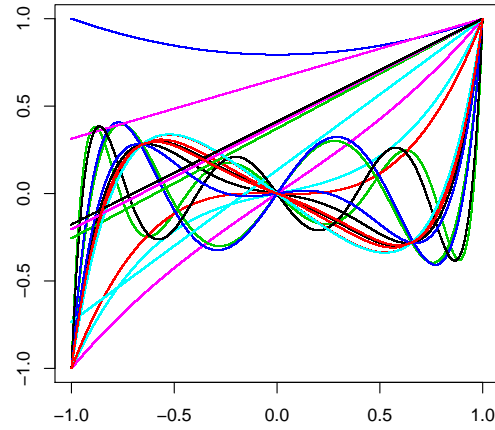
$$\Psi(x) = \int_{u=-1}^1 x(u)(2u + x(u))du$$

and set

$$Y_t = \Psi(X_t) + U_t,$$

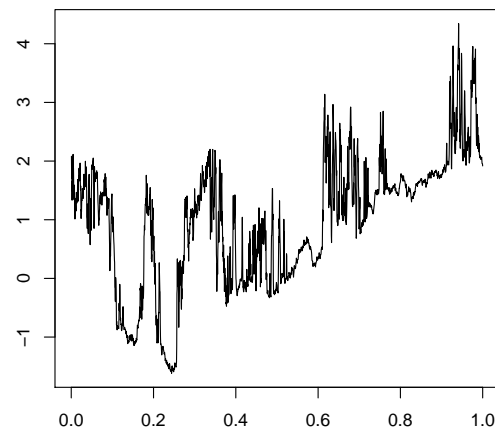
where $U_t = W'_t - W'_{T-1}$ and W'_t is a wiener process independent of X .

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 $\mathcal{C} := (X_{i/2}, i \in 1, 2, \dots, 20)$. We represent theses functions in the following figure.



Panel of 20 curves

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



Path of the process Y on [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, i.e., when we observe (X_t, Y_t) only for $t \in \mathbb{N}$, and use the estimator defined in Ferraty et Vieu (2004) with the indicator function as the kernel. When working with the discrete time process we used the data-driven way of choosing the bandwidth proposed in Benhenni *et al.* (2007). When working with the continuous

time process, i.e. observed on a very thin grid, for $T = 50$, we chose the same bandwidth as the one used for the discret time process and for $T > 50$ who used the bandwidth $h_T = h_{50} \frac{50}{T}$. In the following table, we give the mean square error evaluated on the functions of the panel for different $T=50, 500$ and 2000 .

	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.

On the following graphics, 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.

5 Proofs

5.1 Intermediary results

In the sequel, we use the following notations:

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

$$\hat{r}_{1,T}(x) := \frac{1}{T \mathbb{E}(\Delta_{T,0}(x))} \int_{t=0}^T \Delta_{T,t}(x) dt$$

and

$$\hat{r}_{2,T}(x) := \frac{1}{T \mathbb{E}(\Delta_{T,0}(x))} \int_{t=0}^T \Psi(Y_t) \Delta_{T,t}(x) dt.$$

Lemma 5.1 below studies the behavior of the bias of $\hat{r}_{2,T}$.

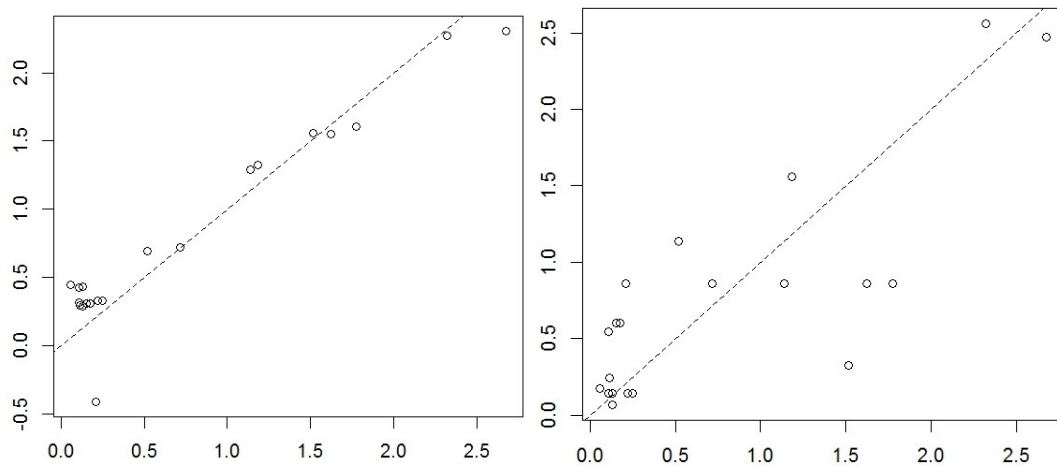
Lemma 5.1 *Under the conditions of Theorem 3.1, we have*

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

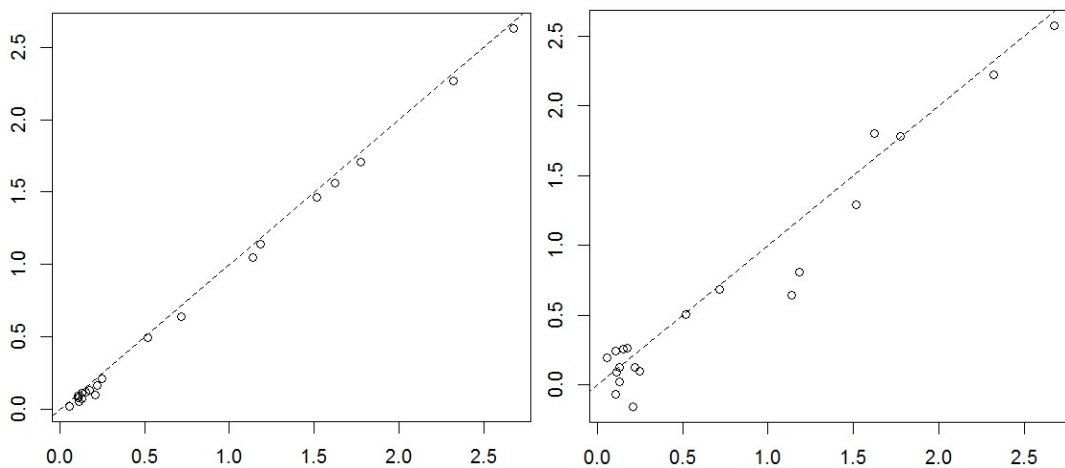
Lemma 5.2 below provides an upper bound for the variances of $\hat{r}_{1,T}$ and $\hat{r}_{2,T}$.

Lemma 5.2 *Under the conditions of Theorem 3.1, we have*

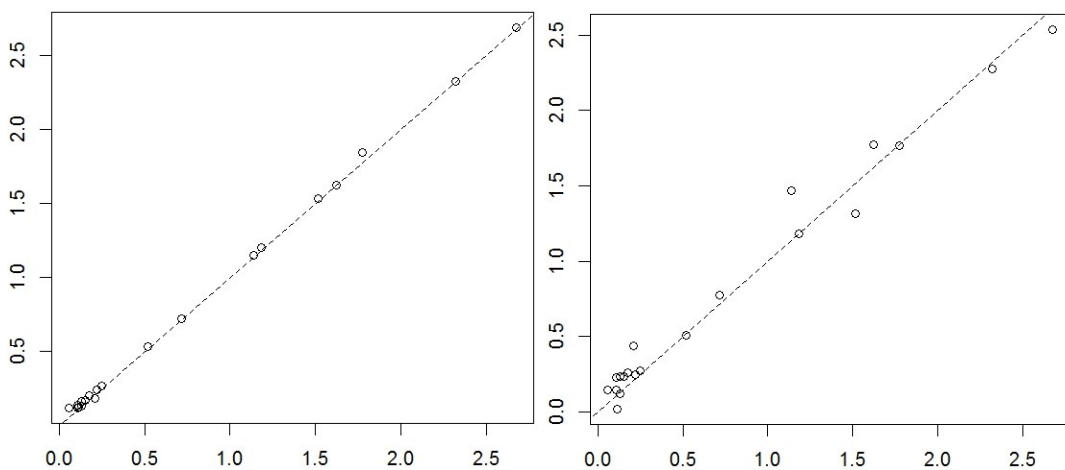
$$\sup_{x \in \mathcal{C}} (\text{Var}(\hat{r}_{2,T}(x)) + \text{Var}(\hat{r}_{1,T}(x))) = \mathcal{O}\left(\frac{1}{T}\right).$$



(a) Outputs for $T=50$



(b) Outputs for $T=500$



(c) Outputs for $T=2000$

Figure 1: 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.

5.2 Proofs of the intermediary results

In the proofs of Lemmas 5.1 and 5.2, for the sake of conciseness, we fix $x \in \mathcal{C}$ and, when no confusion is possible, use the notations $\Psi_t := \Psi(Y_t)$ and $\Delta_{T,t} := \Delta_{T,t}(x)$.

Proof of Lemma 5.1

Observe that

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

Hence

$$E(\widehat{r}_{2,T}(x)) - r(x) = \frac{E\left(r(X_0) \Delta_{T,0}\right)}{E(\Delta_{T,0})} - r(x) = \frac{E\left((r(X_0) - r(x)) \Delta_{T,0}\right)}{E(\Delta_{T,0})}.$$

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 $h^\eta = \mathcal{O}\left(\frac{1}{T}\right)$, there exists a constant C_0 such that

$$|E(\widehat{r}_{2,T}(x)) - r(x)| \leq \frac{E\left(|r(X_0) - r(x)| \Delta_{T,0}\right)}{E(\Delta_{T,0})} \leq C h_T^\eta = \mathcal{O}\left(\frac{1}{T}\right).$$

This ends the proof of Lemma 5.1. \square

Proof of Lemma 5.2 Observe that, by Fubini's Theorem,

$$\text{Var}(\widehat{r}_{2,T}(x)) = \frac{1}{T^2 E(\Delta_{T,0})^2} \int_{t=0}^T \int_{s=0}^T \text{Cov}(\Psi_s \Delta_{T,s}, \Psi_t \Delta_{T,t}) dt ds. \quad (4)$$

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

$$E(\Psi_s \Psi_t | X_s, X_t) = R(X_s) R(X_t) + R(X_s) E(\varepsilon_t | X_s, X_t) + R(X_t) E(\varepsilon_s | X_s, X_t) + E(\varepsilon_s \varepsilon_t | X_s, X_t).$$

Therefore, the covariance term can be expanded as follows:

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

Set

$$d_t := R(X_t) - r(x).$$

We have

$$\begin{aligned}
& \text{Cov}(\Psi_s \Delta_{T,s}, \Psi_t \Delta_{T,t}) \\
&= r(x)^2 \text{E}(\Delta_{T,s} \Delta_{T,t}) + r(x) \left(\text{E}(\Delta_{T,s} \Delta_{T,t} d_t) + \text{E}(\Delta_{T,s} \Delta_{T,t} d_s) \right) \\
&+ \text{E}(\Delta_{T,s} \Delta_{T,t} d_t d_s) + r(x) \text{E} \left(\Delta_{T,s} \Delta_{T,t} \left(\text{E}(\varepsilon_t | X_s, X_t) + \text{E}(\varepsilon_s | X_s, X_t) \right) \right) \\
&+ \text{E} \left(\Delta_{T,s} \Delta_{T,t} \left(d_s \text{E}(\varepsilon_t | X_s, X_t) + d_t \text{E}(\varepsilon_s | X_s, X_t) \right) \right) \\
&+ \text{E} \left(\Delta_{T,s} \Delta_{T,t} \text{E}(\varepsilon_s \varepsilon_t | X_s, X_t) \right) - r(x)^2 \text{E}(\Delta_{T,s})^2 \\
&- \text{E}(\Delta_{T,s} d_s)^2 - 2r(x) \text{E}(\Delta_{T,s} d_s) \text{E}(\Delta_{T,s}) \\
&= r(x)^2 \left(\text{E}(\Delta_{T,s} \Delta_{T,t}) - \text{E}(\Delta_{T,s})^2 \right) - \left(\text{E}(\Delta_{T,s} d_s)^2 + 2r(x) \text{E}(\Delta_{T,s} d_s) \text{E}(\Delta_{T,s}) \right) \\
&+ \text{E}(\Delta_{T,s} \Delta_{T,t} Q),
\end{aligned}$$

with

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

The triangular inequality and Jensen's inequality yield

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

where

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

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

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

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

$$M \leq \left(2|r(x)| C h_T^\eta + C^2 h_T^{2\eta} \right) \text{E}(\Delta_{T,s})^2 \leq \left(2r(x) C h_T^\eta + C^2 h_T^{2\eta} \right) \beta_2^2 \phi(h_T)^2.$$

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

$$\Delta_{T,s} \Delta_{T,t} |Q| \leq \Delta_{T,s} \Delta_{T,t} \left(2|r(x)| C h_T^\eta + C^2 h_T^{2\eta} + (2(|r(x)| + C h_T^\eta) + 1) g_1(|s-t|) \right).$$

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

$$\mathbf{E}(\Delta_{T,s}\Delta_{T,t}) \leq |\text{Cov}(\Delta_{T,s}, \Delta_{T,t})| + \mathbf{E}(\Delta_{T,s})^2 \leq (\beta_2^2 + g_0(|s-t|))\phi(h_T)^2.$$

Hence

$$N \leq \left(2|r(x)|Ch_T^\eta + C^2h_T^{2\eta} + (2(|r(x)| + Ch_T^\eta) + 1)g_1(|s-t|)\right)(\beta_2^2 + g_0(|s-t|))\phi(h_T)^2.$$

Therefore, setting

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

the obtained upper bounds for L , M and N yield

$$\left| \text{Cov}(\Psi_s\Delta_{T,s}, \Psi_t\Delta_{T,t}) \right| \leq G_T(|s-t|). \quad (5)$$

Final bound. Combining (4) and (5), and using (H2)-(i), we have

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

Since g_0 and g_1 are integrable and $h^\eta = \mathcal{O}\left(\frac{1}{T}\right)$, there exists a constant C_0 such that

$$\text{Var}(\widehat{r}_{2,T}(x)) \leq \frac{C_0}{T}.$$

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

$$\text{Var}(\widehat{r}_{1,T}(x)) \leq \frac{C_1}{T}.$$

This last inequality concludes the proof of Lemma 5.2. \square

Proof of Theorem 3.1 Note that, when $\widehat{r}_{1,T}(x) \neq 0$,

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

Therefore we can write

$$\begin{aligned} \widehat{r}_T(x) - r(x) &= (\widehat{r}_T(x)(1 - \widehat{r}_{1,T}(x))) + (\widehat{r}_{2,T}(x) - \mathbf{E}(\widehat{r}_{2,T}(x))) + (\mathbf{E}(\widehat{r}_{2,T}(x)) - r(x)). \end{aligned}$$

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

$$\mathbb{E}(\widehat{r}_T(x) - r(x))^2 \leq 3(U + V + W),$$

where

$$U = \mathbb{E} \left(\widehat{r}_T(x)(1 - \widehat{r}_{1,T}(x)) \right)^2, \quad V = \mathbb{E} \left(\widehat{r}_{2,T}(x) - \mathbb{E}(\widehat{r}_{2,T}(x)) \right)^2$$

and

$$W = \left(\mathbb{E}(\widehat{r}_{2,T}(x)) - r(x) \right)^2.$$

Upper bound for V. Lemma 5.2 yields

$$V = \text{Var}(\widehat{r}_{2,T}(x)) = \mathcal{O}\left(\frac{1}{T}\right).$$

Upper bound for W. Lemma 5.1 yields

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

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

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

Let \mathcal{T}_T be the sigma-algebra generated by $\{X_t, t \in [0, T]\}$. Using (H4) and Lemma 5.2, we get

$$\begin{aligned} U &= \mathbb{E} \left(\mathbb{E}(\widehat{r}_T^2(x)(1 - \widehat{r}_{1,T}(x))^2 | \mathcal{T}_T) \right) = \mathbb{E} \left(\mathbb{E}(\widehat{r}_T^2(x) | \mathcal{T}_T)(1 - \widehat{r}_{1,T}(x))^2 \right) \\ &= \mathbb{E} \left(\mathbb{E} \left(\left(\int_{t=0}^T Z_t \Psi(Y_t) dt \right)^2 | \mathcal{T}_T \right) (1 - \widehat{r}_{1,T}(x))^2 \right) \\ &= \mathbb{E} \left(\int_{(s,t) \in [0,T]^2} Z_t Z_s \mathbb{E}(\Psi(Y_t)\Psi(Y_s) | \mathcal{T}_T) ds dt (1 - \widehat{r}_{1,T}(x))^2 \right) \\ &\leq \mathbb{E} \left(\int_{(s,t) \in [0,T]^2} Z_t Z_s R ds dt (1 - \widehat{r}_{1,T}(x))^2 \right) \\ &\leq R \mathbb{E} \left((1 - \widehat{r}_{1,T}(x))^2 \right) = R \text{Var}(\widehat{r}_{1,T}(x)) = \mathcal{O}\left(\frac{1}{T}\right). \end{aligned}$$

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

$$\mathbb{E}(\widehat{r}_T(x) - r(x))^2 = \mathcal{O}\left(\frac{1}{T}\right).$$

Theorem 3.1 is proved. \square

References

- Benhenni, K., Ferraty, F., Rachdi, M., et Vieu, P. (2007). Local smoothing regression with functional data. *Comput. Statist.*, **22**(3), 353–369.
- Bosq, D. (1993). Vitesses optimales et superoptimales des estimateurs fonctionnels pour les processus à temps continu. *C. R. Acad. Sci. Paris Sér. I Math.*, **317**(11), 1075–1078.
- Bosq, D. (1997). Parametric rates of nonparametric estimators and predictors for continuous time processes. *Ann. Statist.*, **25**(3), 982–1000.
- Bosq, D. (1998). *Nonparametric statistics for stochastic processes*, volume 110 of *Lecture Notes in Statistics*. Springer-Verlag, New York, second edition. Estimation and prediction.
- Castellana, J. V. et Leadbetter, M. R. (1986). On smoothed probability density estimation for stationary processes. *Stochastic Process. Appl.*, **21**(2), 179–193.
- Collomb, G. et Härdle, W. (1986). Strong uniform convergence rates in robust nonparametric time series analysis and prediction: kernel regression estimation from dependent observations. *Stochastic Process. Appl.*, **23**(1), 77–89.
- Ferraty, F. et Vieu, P. (2004). Nonparametric models for functional data, with application in regression, time-series prediction and curve discrimination. *J. Nonparametr. Stat.*, **16**(1-2), 111–125. The International Conference on Recent Trends and Directions in Nonparametric Statistics.
- Ferraty, F. et Vieu, P. (2006). *Nonparametric functional data analysis*. Springer Series in Statistics. Springer, New York. Theory and practice.
- Ferraty, F. et Vieu, P. (2011). Kernel regression estimation for functional data. In *The Oxford handbook of functional data analysis*, pages 72–129. Oxford Univ. Press, Oxford.
- Ferraty, F., Mas, A., et Vieu, P. (2007). Nonparametric regression on functional data: inference and practical aspects. *Aust. N. Z. J. Stat.*, **49**(3), 267–286.
- Grunig, R. (1966). Probabilités conditionnelles régulières sur des tribus de type non dénombrable. *Ann. Inst. H. Poincaré Probab. Statist.*, **2**(3), 227–229.
- Jirina, M. (1954). Conditional probabilities on σ algebras with countable basis. *Cz. Math. J.*, pages 372–380.
- Jirina, M. (1959). On regular conditional probabilities. *Cz. Math. J.*, pages 773–795.
- Krzyżak, A. et Pawlak, M. (1987). The pointwise rate of convergence of the kernel regression estimate. *J. Statist. Plann. Inference*, **16**(2), 159–166.
- Masry, E. (2005). Nonparametric regression estimation for dependent functional data: asymptotic normality. *Stochastic Process. Appl.*, **115**(1), 155–177.
- Nadaraya, E. (1964). On estimating regression. *Theor. Prob. Appl.*, **9**, 141–142.

- Rosenblatt, M. (1969). Conditional probability density and regression estimators. In *Multivariate Analysis, II (Proc. Second Internat. Sympos., Dayton, Ohio, 1968)*, pages 25–31. Academic Press, New York.
- Roussas, G. G. (1990). Nonparametric regression estimation under mixing conditions. *Stochastic Process. Appl.*, **36**(1), 107–116.
- Stone, C. J. (1982). Optimal global rates of convergence for nonparametric regression. *Ann. Statist.*, **10**(4), 1040–1053.
- Watson, G. (1964). Smooth regression analysis. *Sankhya, Ser. A*, **26**, 359–372.