



**HAL**  
open science

## Local polynomial estimation of regression operators from functional data with correlated errors

Karim Benhenni, Ali Hajj Hassan, Yingcai Su

► **To cite this version:**

Karim Benhenni, Ali Hajj Hassan, Yingcai Su. Local polynomial estimation of regression operators from functional data with correlated errors. *Journal of Multivariate Analysis*, 2019, 170, pp.80-94. 10.1016/j.jmva.2018.10.008 . hal-01927109

**HAL Id: hal-01927109**

**<https://hal.science/hal-01927109v1>**

Submitted on 21 Oct 2021

**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.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

# Local polynomial estimation of regression operators from functional data with correlated errors

Karim Benhenni<sup>a</sup>, Ali Hajj Hassan<sup>a</sup>, Yingcai Su<sup>b</sup>

<sup>a</sup>*Université Grenoble Alpes, Laboratoire Jean Kuntzmann, UMR 5224, CNRS, France*

<sup>b</sup>*Department of Mathematics, Missouri State University, USA*

---

## Abstract

This article considers the problem of nonparametric estimation of the regression operator  $r$  in a functional regression model  $Y = r(x) + \varepsilon$  with a scalar response  $Y$ , a functional explanatory variable  $x$ , and a second order stationary error process  $\varepsilon$ . We construct a local polynomial estimator of  $r$  together with its Fréchet derivatives from functional data with correlated errors. The convergence in mean squared error of the constructed estimator is studied for both short and long range dependent error processes. Simulation studies on the performance of the proposed estimator are conducted, and applications to independent data and El Niño time series data are given.

*Keywords:* El Niño time series, Fréchet derivatives, Functional fixed-design data, Local polynomial estimation, Nonparametric regression operator, Short and long memory processes.

---

## 1. Introduction

Data in the form of a collection of continuous function curves become increasingly common in practice. Accordingly, the interest in functional data analysis (FDA) has been growing steadily over the past few decades. The books [14, 25, 38, 39], and the overview paper [18] provide a general coverage of FDA theory and methodology. More recent contributions to this field can be found in the book [2]. FDA-based statistical techniques find applications in a wide range of areas such as environmental, medical, meteorological sciences and statistical control processes [40]. Interestingly, FDA techniques were even applied to problems of radon entry rate estimation [15] and identification of abnormal nitrogen oxides emission levels [21].

In the finite-dimensional framework, there is a vast literature on local polynomial regression estimation under the assumption of independent random errors. In this regard, asymptotic expressions for the bias and variance of the corresponding estimators can be found in [20] and [41]. In the case of regression estimation with dependent errors, local polynomial estimation is studied under mixing conditions in [34], association in [33], and (stationary) short range dependence in [26] and [36]. For more general stationary error processes (correlated errors) and local polynomial estimators based on repeated independent observations, the rates of convergence of mean squared errors (MSE) and asymptotic distributions were obtained in [7].

In the infinite-dimensional framework, in the case of dependent data and random functional designs, consistent estimators of regression operators with scalar responses under certain mixing conditions were obtained in [24]. There are many practical situations in which functional designs are fixed and the errors are correlated. For functional fixed designs, the asymptotic properties of the Nadaraya–Watson kernel estimator in terms of MSE consistency and asymptotic normality were obtained in [9]. Local linear regression for functional data and scalar response with independent errors was introduced in [5] and [13]. The MSE consistency of the local linear regression estimator when the covariate takes values in a Hilbert space, namely  $L_2[0, 1]$ , was derived in [5]. An easier computational expression for the local estimator was constructed and its almost complete convergence, along with the corresponding rates, was shown in [6]. A more general approach for constructing the local linear estimator was taken in [12], where a bound on the MSE was also given. Other related problems such as conditional quantile regression or conditional distribution estimation based on local linear methods can be found in [37] and [43].

Here, we extend these results to the local polynomial estimation framework and study the effect of short and long memory error processes, as defined in [11], on the rate of convergence of the MSE. More precisely, we construct a

truncated local polynomial estimator of the regression operator together with its Fréchet derivatives based on functional data with correlated errors, and then establish the asymptotic behavior, including the rate of convergence, of the corresponding MSE. The procedure proposed in this article for the estimation of Fréchet derivatives of the regression operator is nonparametric. To the best of our knowledge, the only earlier work involving estimation of functional derivatives is due to Hall et al. [29]. In the construction and analysis of the proposed estimators, they used the first order derivative of the regression operator taken at the orthonormal eigenfunctions of the estimated covariance kernel.

The nonparametric  $k$ -nearest-neighbor ( $k$ NN) smoothing technique was used by Kara et al. [31] in the estimation of a regression operator for iid samples. In their approach, the value of the parameter  $k$  is fixed but the smoothing parameter, the bandwidth, is data dependent. It is of interest to see how the  $k$ NN method works out in the setting of dependent samples. A nonparametric approach with data-driven smoothing parameters to the regression operator estimation was also investigated in [19] for iid errors. It seems worthwhile to investigate possible extensions of the approach of this article to a semi-parametric set-up, functional single index model as investigated in [1] and [28], and also to multiple index functional regression models as in [17].

In nonparametric time series modeling, to overcome the difficulty of using an overwhelmingly high number of past observations, Aneiros and Vieu [3, 4] proposed a functional approach, namely the semi-functional partial linear regression (SFPLR) model. The work reported in this article may be extended to their semi-parametric situations, when the response, explanatory variables, dependence of data, and covariates are properly identified or introduced. That being said, SFPLR modeling is interesting in its own right and certainly a promising approach for the myriad of real life functional time series data. The results of this article are potentially useful for linear and nonlinear inverse problems in geophysical sciences [35] as well. In our discussion and analysis, short and long memory dependent errors are assumed. It is worth noting that the idea of functional canonical correlation coefficients as in [30] provides another way to characterize the dependence structure between and within functional samples. All of these possible extensions or further studies are beyond the scope of this article and thus will not be discussed further. However, readers are referred to [27] and [32] for recent surveys on the topic of the nonparametric FDA.

The organization of this paper is as follows. In Section 2, we state the functional regression model and define the aforementioned nonparametric estimator. In Section 3, we set up the appropriate assumptions and give the MSE convergence of the local estimator of the regression operator and its Fréchet derivatives along with the corresponding rates. Sections 4 and 5 are devoted to various simulation studies for both short and long memory error processes and also to an application to spectrometric data and El Niño time series data. Finally, a summary and some further comments are given in Section 6, and the proofs of the main results are presented in the Appendix. Moreover, the codes and data for reproducing the simulations and data analysis are included in the Online Supplement.

## 2. Estimation of the regression operator and its Fréchet derivatives

We consider a regression model with regression operator  $r$  defined, for all  $i \in \{1, \dots, n\}$ , by

$$Y_i = r(x_i) + \varepsilon_i,$$

where  $Y_1, \dots, Y_n$  are observations of the response variable  $Y$ ,  $x_1, \dots, x_n$  the explanatory data of functional fixed-design type, and  $\varepsilon_1, \dots, \varepsilon_n$  correlated random errors. The response  $Y$  is a real-valued random variable such that  $E|Y| < \infty$ . The explanatory variable  $x$  is assumed to belong to a metric space  $(H, d)$  equipped with the metric  $d$ . The random error  $\varepsilon$  is assumed to be a centered second order stationary process with autocovariance function defined, for all  $i, j \in \mathbb{Z}$ , by  $E(\varepsilon_i \varepsilon_j) = \rho_\varepsilon(i - j)$ , and unit variance, viz.  $E(\varepsilon_i^2) = \rho_\varepsilon(0) = 1$ .

In this section, a local polynomial estimator of the regression operator  $r$ , together with its Fréchet derivatives, is proposed. In the next section, we investigate its asymptotic performance for both short and long range dependence (or memory) error processes. For this, we recall briefly the definitions of short and long memory processes.

For some stochastic processes, such as the popular autoregressive and moving average models, the autocovariance function decays exponentially and is summable, viz.

$$\sum_{\tau=-\infty}^{+\infty} |\rho_\varepsilon(\tau)| < \infty.$$

These random processes are called short range dependent (SRD) or short range correlated stochastic processes. Some standard time series models, such as the popular Fractional ARIMA processes or Fractional Difference processes, assume the non-summability of the autocovariance sequence  $\{\rho_\varepsilon(i) : i \in \mathbb{Z}\}$ , which captures the intuition underlying long range dependence (LRD) or long range memory, viz.

$$\sum_{\tau=-\infty}^{+\infty} |\rho_\varepsilon(\tau)| = \infty.$$

For instance, a stationary process  $\{\varepsilon_t : t \in \mathbb{Z}\}$  is a long memory process if for some real  $\gamma \in (0, 1]$ , and some positive constant  $C$ , its autocovariance function satisfies, as  $j \rightarrow \infty$ ,

$$|\rho_\varepsilon(j)| \sim C |j|^{-\gamma}; \quad (1)$$

see [11] for more details.

Let  $U$  be an open subset of  $H$ . If  $f : U \rightarrow R$  is  $p + 1$  times differentiable on  $U$ , the Taylor expansion of  $f$  around  $x \in U$  is

$$f(y) = \sum_{j=0}^p \frac{1}{j!} D^j f(x) d^j(x, y) + R_p(y),$$

where  $D^j f(x)$  denote the  $j$ th Fréchet derivative of  $f$  at  $x$ , with  $x$  and  $y$  being points of  $U$  and  $d$  the metric of  $H$ . The remainder is given by

$$R_p(y) = \frac{1}{p!} \int_0^1 D^{p+1} f\{x + td(x, y)\} d^{p+1}(x, y) (1-t)^p dt = o\{d^p(x, y)\}$$

as  $d(x, y) \rightarrow 0$ ; see [16] and [42] for more details.

When the observations  $Y_1, \dots, Y_n$  are possibly dependent, our goal is to estimate the regression operator  $r$  and its Fréchet derivatives  $r_j = D^j r$ . To this end, we introduce a functional truncated version of the local polynomial kernel estimate. The estimators of  $r_0(x), \dots, r_p(x)$  based on the observations  $Y_1, \dots, Y_n$  and the functional fixed design  $x_1, \dots, x_n$  are obtained for  $x \in U$ , by minimizing the criterion

$$\sum_{i=1}^n \left\{ Y_i - \sum_{j=0}^p \beta_j(x) d^j(x_i, x) \right\}^2 K_h\{d(x_i, x)\}$$

with  $K_h(\cdot) = h^{-1} K(\cdot/h)$  and  $\beta_j(x) = r_j(x)/j!$ , for all  $j \in \{0, \dots, p\}$ , where the kernel  $K$  is a real-valued function defined on  $\mathbb{R}_+$  with a compact support and  $h = h(n)$  is the bandwidth such that  $h \in \mathbb{R}_+$  and  $h \rightarrow 0$  as  $n \rightarrow \infty$ .

Let  $\mathbf{Y}_n = (Y_1, \dots, Y_n)^\top$  and denote the canonical basis of  $\mathbb{R}^{p+1}$  by  $(\mathbf{e}_k)_{k=0}^p$ , i.e., the  $(k+1)$ st component of the column vector  $\mathbf{e}_k$  is 1 and the rest of the components are zeros. Define the  $n \times (p+1)$  matrix  $\mathbf{X}_n$  by

$$\mathbf{X}_n = \begin{pmatrix} 1 & d^1(x_1, x) & \dots & d^p(x_1, x) \\ \vdots & \vdots & \dots & \vdots \\ 1 & d^1(x_n, x) & \dots & d^p(x_n, x) \end{pmatrix}$$

and the diagonal matrix  $\mathbf{W}_n = \text{diag}[K_h\{d(x_1, x)\}, \dots, K_h\{d(x_n, x)\}]$ . Then the estimators  $\widehat{r}_j(x)$  are given, for all  $j \in \{0, \dots, p\}$ , by

$$\widehat{r}_j(x) = j! \mathbf{e}_j^\top \widehat{\boldsymbol{\beta}}_n(x), \quad (2)$$

where  $\widehat{\boldsymbol{\beta}}_n(x) = (\mathbf{X}_n^\top \mathbf{W}_n \mathbf{X}_n)^{-1} \mathbf{X}_n^\top \mathbf{W}_n \mathbf{Y}_n$ .

The truncated functional Nadaraya–Watson estimator of the regression operator  $r$  can be obtained by taking  $j = 0$  and  $p = 0$ , viz.

$$\widehat{r}_{NW}(x) = \sum_{x_i \in B(x, h)} W_i(x) Y_i$$

where  $B(x, h) = \{y : d(x, y) \leq h\}$  is the ball with center  $x$  and radius  $h$  in the metric space  $(H, d)$ , and  $W_i(x) = \Delta_i(x) / \sum_{x_{i'} \in B(x, h)} \Delta_{i'}(x)$  with  $\Delta_i(x) = K_h\{d(x, x_i)\}$ .

### 3. Asymptotic performance

#### 3.1. Notations and assumptions

The following notations and assumptions are required for the statements of the main results. The kernel  $K$  is a function with compact support on  $[0, 1]$  and has a continuous derivative in  $(0, 1)$  such that

$$K(1) > 0 \quad \text{and} \quad K'(s) \leq 0. \quad (3)$$

Thus the kernel  $K$  is a decreasing function that gives more weight to the argument close to zero.

For all  $j \in \{0, \dots, p\}$  and a bandwidth  $h > 0$ , let

$$\varphi_j(h) = \frac{1}{n} \sum_{i=1}^n \frac{d^j(x_i, x)}{h^j} \mathbf{1}_{B(x, h)}(x_i)$$

where  $\mathbf{1}_D(x)$  stands for the indicator function of a set  $D$  in  $H$ , and for all  $k, \ell \in \{0, \dots, p\}$  and  $s, s' \in (0, 1]$ ,

$$\mu_{k, \ell}(hs, hs') = \frac{1}{n^2} \sum_{\substack{1 \leq i, j \leq n \\ i \neq j}} \frac{d^k(x_i, x)}{(hs)^k} \frac{d^\ell(x_j, x)}{(hs')^\ell} \rho_\epsilon(i - j) \mathbf{1}_{B(x, hs)}(x_i) \mathbf{1}_{B(x, hs')}(x_j).$$

In the functional random design, the small ball probabilities are determined by the distribution function  $F_x(t) = \Pr\{d(X, x) \leq t\}$ , and  $E\{d^j(X, x)\}$  represents the  $j$ th moment of the random variable  $d(X, x)$ . In the functional fixed design, this is equivalent to the empirical  $j$ th moment frequency  $\varphi_j(t)$ . Likewise,  $\mu_{k, \ell}(t, s)$  represents the empirical  $(k, \ell)$ th joint weighted moment frequency. Then we assume that  $\varphi_j$  and  $\mu_{k, \ell}$  satisfy a certain regular variation condition at the origin, i.e., as  $h \rightarrow 0$ ,

$$\tau_j^h(s) = \varphi_j(hs) / \varphi_j(h) \rightarrow \tau_j(s) \quad (4)$$

and

$$\psi_{k, \ell}^h(s, s') = \mu_{k, \ell}(hs, hs') / \mu_{k, \ell}(h, h) \rightarrow \psi_{k, \ell}(s, s'). \quad (5)$$

When Fréchet derivatives of the regression operator are estimated, assumptions (4) and (5) are natural extensions of the assumptions considered in [9] and [22], which correspond to the case  $j = 0$  of this article.

Finally, we list the following constants that are related to  $K$ , the  $\tau_j$ s, and  $\psi$ s; they will be used in the next section. For  $\nu \in \{1, 2\}$ ,

$$\begin{aligned} C_j^\nu &= K^\nu(1) - \int_0^1 \tau_j(s) s^j (K^\nu)'(s) ds, \\ C_{k, \ell}^{1,1} &= K^2(1) - K(1) \int_0^1 \psi_{k, \ell}(s, 1) K'(s) ds - K(1) \int_0^1 \psi_{k, \ell}(1, s) K'(s) ds + \int_0^1 \int_0^1 \psi_{k, \ell}(s, t) K'(s) K'(t) ds dt. \end{aligned} \quad (6)$$

#### 3.2. Asymptotic results

For the asymptotic expressions of the bias and the variance of the local polynomial estimators, we have the following results.

**Theorem 1.** *Assume that the regression operator  $r$  is  $p+1$  times continuously differentiable around  $x$ . If assumptions (3) and (4) are satisfied, then for all  $j \in \{0, \dots, p\}$ , we have*

$$E\{\hat{r}_j(x)\} - r_j(x) = j! \frac{r_{p+1}(x)}{(p+1)!} h^{p+1-j} \mathbf{e}_j^\top \Sigma_h^{-1} \tilde{\mu}_h \{1 + o(1)\},$$

where  $\Sigma_h = (\varphi_{k+\ell}(h) C_{k, \ell}^1)_{k, \ell \in \{0, \dots, p\}}$  is a  $(p+1) \times (p+1)$  matrix,  $\tilde{\mu}_h = (\varphi_{p+1}(h) C_{p+1}^1, \dots, \varphi_{2p+1}(h) C_{2p+1}^1)^\top$  is a column vector of size  $p+1$ , and the constants  $C_j^1$  are as in (6).

**Theorem 2.** Assume that the regression operator  $r$  is  $p + 1$  times continuously differentiable around  $x$ . If assumptions (3), (4), and (5) are satisfied, then for all  $j \in \{0, \dots, p\}$ , we have

$$\text{var}\{\widehat{r}_j(x)\} = \frac{1}{n} \mathbf{e}_j^\top \mathbf{H}^{-1} \boldsymbol{\Sigma}_h^{-1} (\boldsymbol{\Sigma}_h^{(2)} + n \boldsymbol{\Sigma}_h^{(1,1)}) \boldsymbol{\Sigma}_h^{-1} \mathbf{H}^{-1} \mathbf{e}_j \{1 + o(1)\},$$

where  $\mathbf{H} = \text{diag}(1, h, \dots, h^p)$ , the matrices  $\boldsymbol{\Sigma}_h^{(2)} = (\varphi_{k+\ell}(h) C_{k+\ell}^2)_{k,\ell \in \{0, \dots, p\}}$  and  $\boldsymbol{\Sigma}_h^{(1,1)} = (\mu_{k,\ell}(h, h) C_{k,\ell}^{1,1})_{k,\ell \in \{0, \dots, p\}}$  are of size  $(p + 1) \times (p + 1)$ , and the constants  $C_j^2$  and  $C_{k,\ell}^{1,1}$  are as in (6).

We can obtain simpler expressions of the asymptotic bias and variance terms under some assumptions on the empirical frequencies.

**Corollary 1.** Assume that there exist some positive constants  $\alpha_0, \dots, \alpha_p$  and  $\gamma_{k,\ell}$ , for all  $k, \ell \in \{0, \dots, p\}$  such that, as  $h \rightarrow 0$ ,  $\varphi_j(h)/\varphi_0(h) \rightarrow \alpha_j$  and  $\mu_{k,\ell}(h, h)/\mu_{0,0}(h, h) \rightarrow \gamma_{k,\ell}$ , then under assumptions (3), (4), and (5), we have

$$\mathbb{E}\{\widehat{r}_j(x)\} - r_j(x) = j! \frac{r_{p+1}(x)}{(p+1)!} h^{p+1-j} \mathbf{e}_j^\top \mathbf{M}^{-1} \boldsymbol{\mu} \{1 + o(1)\}$$

and

$$\text{var}\{\widehat{r}_j(x)\} = \frac{1}{nh^{2j}\varphi_0(h)} \mathbf{e}_j^\top \mathbf{M}^{-1} \mathbf{M}^{(2)} \mathbf{M}^{-1} \mathbf{e}_j \{1 + o(1)\} + \frac{\mu_{0,0}(h, h)}{h^{2j}\varphi_0^2(h)} \mathbf{e}_j^\top \mathbf{M}^{-1} \mathbf{M}^{(1,1)} \mathbf{M}^{-1} \mathbf{e}_j \{1 + o(1)\},$$

where  $\boldsymbol{\mu} = (\alpha_{p+1} C_{p+1}^1, \dots, \alpha_{2p+1} C_{2p+1}^1)^\top$  is a column vector of size  $p + 1$ ,  $\mathbf{M} = (\alpha_{k+\ell} C_{k+\ell}^1)_{k,\ell \in \{0, \dots, p\}}$ ,  $\mathbf{M}^{(2)} = (\alpha_{k+\ell} C_{k+\ell}^2)_{k,\ell \in \{0, \dots, p\}}$ , and  $\mathbf{M}^{(1,1)} = (\gamma_{k,\ell} C_{k,\ell}^{1,1})_{k,\ell \in \{0, \dots, p\}}$  are matrices of size  $(p + 1) \times (p + 1)$ .

**Remark 1.** The asymptotic MSE of the Nadaraya–Watson estimator  $\widehat{r}_{NW}$  can be deduced from Theorems 1 and 2 by taking  $j = 0$  and  $p = 0$  (local constant)

$$\text{MSE}(\widehat{r}_{NW}) = O(h^2) + \frac{1}{n\varphi_0(h)} \frac{C_0^2}{(C_0^1)^2} + \frac{\mu_{0,0}(h, h)}{\varphi_0^2(h)} \frac{C_{0,0}^{1,1}}{(C_0^1)^2} \{1 + o(1)\},$$

where  $\varphi_0(h)$  is the empirical frequency of curves belonging to the closed ball of center  $x$  and radius  $h$ . In this case, the dominant term in the variance is of order  $O\{\mu_{0,0}(h)/\varphi_0^2(h)\}$  for a long memory error process, and  $O[1/\{n\varphi_0(h)\}]$  for a short memory error process.

We also deduce from Theorems 1, 2 and Corollary 1, the corresponding rates of convergence of the local estimator of the regression operator and its Fréchet derivatives when the errors are iid.

**Corollary 2.** If  $Y_1, \dots, Y_n$  (or  $\varepsilon_1, \dots, \varepsilon_n$ ) are independent and identically distributed, then under assumptions (3), (4), and (5), and if  $nh^{2j}\varphi_0(h) \rightarrow \infty$  as  $n \rightarrow \infty$ , we have

$$\text{MSE}\{\widehat{r}_j(x)\} = O(h^{2(p+1-j)}) + \frac{1}{nh^{2j}\varphi_0(h)} \mathbf{e}_j^\top \mathbf{M}^{-1} \mathbf{M}^{(2)} \mathbf{M}^{-1} \mathbf{e}_j \{1 + o(1)\}.$$

In the case of long and short memory error processes, we deduce the following rates of convergence of the local estimator of  $r$  and its Fréchet derivatives.

**Corollary 3.** In the case of a long memory, under Assumptions (1), (3), (4), and (5) we have that if  $\{n\varphi_0(h)\}^\gamma h^{2j} \rightarrow \infty$  as  $n \rightarrow \infty$ , then

$$\text{MSE}\{\widehat{r}_j(x)\} = O(h^{2(p+1-j)}) + O\left[\frac{1}{\{n\varphi_0(h)\}^\gamma h^{2j}}\right].$$

**Corollary 4.** In the case of a short memory, under Assumptions (3), (4), and (5) we have that if  $n\varphi_0(h)h^{2j} \rightarrow \infty$  as  $n \rightarrow \infty$ , then

$$\text{MSE}\{\widehat{r}_j(x)\} = O(h^{2(p+1-j)}) + O\left\{\frac{1}{n\varphi_0(h)h^{2j}}\right\}.$$

It should be noted that, as in the finite-dimensional framework, the performance in terms of the rate of convergence of the MSE declines when higher Fréchet derivatives of the regression operator are estimated. The bias term is of smaller order for higher-order polynomial approximation. The rate of convergence of the variance does not depend on the order  $p$  and hence can only be affected by the order of approximation through the asymptotic constant.

It can also be observed from these results that if the local polynomials are used for estimating any order Fréchet derivative  $r_1, \dots, r_p$ , the rate of convergence coming from the variance component is lower in the case of the long memory error process than in the case of the short memory error process.

#### 4. Simulation studies

This section gives some simulations which show the advantage of using the local polynomial estimates of the regression operator  $r$  in the case of the functional fixed design model with correlated errors. To this end, we simulate two types of smooth curves: parabolic curves and sinusoidal curves. A sample of 100 curves of each type is plotted in Figure 1.

For both cases, we generate different curves of sample size  $n = 200$ ; each curve is discretized with 100 equidistant values in  $[0, 1]$ . The response variable is obtained from the model defined, for all  $k \in \{1, \dots, n\}$ , by

$$y_k = r(x_k) + \epsilon_k.$$

Recall that the error process  $\epsilon$  is a second order stationary process. As a short memory process, we consider an Ornstein–Uhlenbeck process (OU) which satisfies the stochastic differential equation

$$d\epsilon_t = (\theta_1 - \theta_2 \epsilon_t) dt + \theta_3 dW_t,$$

where  $\theta_1$ ,  $\theta_2$  and  $\theta_3$  are three deterministic parameters and  $W_t$  denotes the Wiener process. Here, we take  $\theta_1 = 0$ ,  $\theta_2 = 0.75$  and  $\theta_3 = 0.5$ . For a long memory error process, we use an autoregressive fractionally integrated moving average ARFIMA( $p, d, q$ ) with parameters  $p = q = 0$  and fractional differencing parameter  $d = 0.3$ .

The standard deviation  $\sigma_\epsilon$  of the error process is controlled by considering various values of the Signal-to-Noise Ratio (SNR), namely: 2, 3, 5, 10 and 20, where

$$\text{SNR} = \frac{\sigma_r^2}{\sigma_\epsilon^2} = \frac{1}{n\sigma_\epsilon^2} \sum_{k=1}^n \{r(x_k) - \overline{r(x)}\}^2$$

and  $\overline{r(x)}$  denotes the average of the  $r(x_k)$ s.

Our main goal here is to compare the performance of the following estimators of the regression operator  $r$ :

- a) the Nadaraya–Watson (NW) estimator obtained for  $j = 0$  and  $p = 0$ ;
- b) the linear estimator obtained for  $j = 0$  and  $p = 1$ ;
- c) the quadratic estimator obtained for  $j = 0$  and  $p = 2$ .

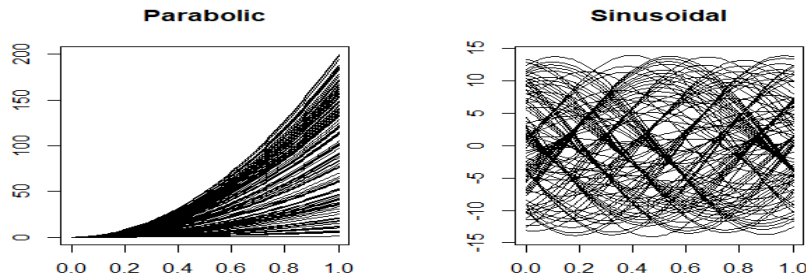


Fig. 1. A sample of 100 curves of parabolic and sinusoidal curves.

The three estimators are constructed from the Epanechnikov kernel. The explanatory variable  $x$  belongs to the metric space  $L^2 [0, 1]$  equipped with the metric  $d$ : for any two curves  $x_k$  and  $x_{k'}$ ,

$$d(x_k, x_{k'}) = \sqrt{\int_0^1 \{x_k(t) - x_{k'}(t)\}^2 dt}.$$

The selection of a proper smoothing parameter (bandwidth) is a crucial step for the estimation of the regression operator. We use a local bandwidth choice which is known to outperform the global one; see [8]. The method for selecting the bandwidth  $h$  at the curve  $x$  consists of choosing  $\widehat{h}_x$  to minimize, over a particular subset  $H_n$  of smoothing parameters to be specified, the local cross-validation criterion

$$LCV_x(h) = \frac{1}{n} \sum_{i=1}^n \{Y_i - \widehat{r}_h^i(x_i)\}^2 W_{n,x}(x_i),$$

where  $\widehat{r}_h^i$  is the leave-one-curve-out estimate of  $r$ . The locally selected bandwidth is

$$\widehat{h}_x = \arg \min_{h \in H_n} LCV_x(h).$$

We take, for all  $i \in \{1, \dots, n\}$ , the local weights

$$W_{n,x}(x_i) = \begin{cases} 1 & \text{if } d(x_i, x) < h, \\ 0 & \text{otherwise.} \end{cases}$$

In practice, one replaces the minimization of  $LCV_x$  over the infinite set  $(0, \infty)$  (i.e.,  $h > 0$ ) with a minimization problem over a finite set. To this end, we consider the local bandwidth  $h_k(x)$  such that the ball centered at  $x$  of radius  $h_k(x)$  contains exactly  $k$  neighbors (i.e.,  $k$  curves around  $x$ ). As a direct consequence, we have for all  $x \in H$ ,  $W_{1,x}(x_i) + \dots + W_{n,x}(x_i) = k$  when  $h = h_k(x)$ . In this way, we can look for the optimal local bandwidth among those which consider respectively one neighbor, two neighbors, etc. The optimal retained number  $k_{\text{opt}}$  of neighbors allows to build the optimal local bandwidths  $h_{k_{\text{opt}}}(x)$  at any  $x \in H$ . Finally, the performance of an estimator  $\hat{r}$  is evaluated by computing the relative mean squared error, viz.

$$\text{RMSE}(\hat{r}) = \frac{\sum_{k=1}^n \{y_k - \hat{r}(x_k)\}^2}{\sum_{k=1}^n (y_k - \bar{y})^2}.$$

#### 4.1. Parabolic curves and linear operator

We first generate 200 parabolic smooth curves by setting, for all  $k \in \{1, \dots, 200\}$  and  $t \in [0, 1]$ ,  $x_k(t) = kt^2$ . We take the following linear regression operator

$$r(x) = \int_0^1 x(t) dt.$$

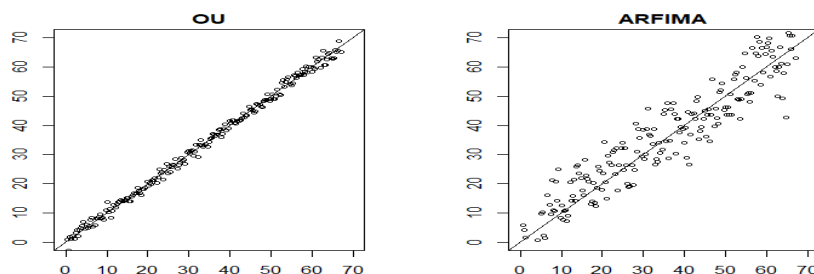
Figure 2 gives the plot of the  $y_k$ s versus the  $r(x_k)$ s for the two error processes when  $\text{SNR} = 10$ . The simulation results are summarized in Table 1. The values of RMSE of each estimator increase when the SNR decreases. It appears that our estimators (linear and quadratic) outperforms the Nadaraya–Watson estimator regardless of the type of error process and the value of SNR that are being used. For instance, when  $\text{SNR} = 10$ , there is a reduction of the RMSE, 26% for OU and 33% for ARFIMA (respectively 56% for OU and 54% for ARFIMA), of the local linear estimator (respectively the local quadratic estimator) over the NW estimator. Lastly, Figures 3–4 give a graphical comparison by displaying the estimated responses in terms of the observed responses for the two types of error processes when  $\text{SNR} = 10$ .



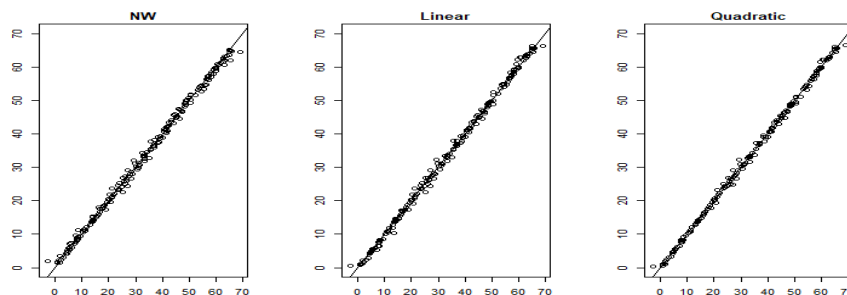
**Table 1**

Parabolic curves: RMSE of each estimator in terms of error processes, ARFIMA and OU, for different values of SNR.

Error Process	SNR	NW	Linear	Quadratic
OU	20	0.00162	0.00114	0.00063
	10	0.00310	0.00230	0.00136
	5	0.00604	0.00471	0.00294
	3	0.00998	0.00797	0.00523
	2	0.01488	0.01213	0.00814
ARFIMA	20	0.02805	0.01926	0.01310
	10	0.05581	0.03729	0.02569
	5	0.10551	0.06778	0.04731
	3	0.15977	0.09971	0.07024
	2	0.21188	0.13002	0.09211



**Fig. 2.** Parabolic curves:  $y_k$ s versus  $r(x_k)$ s with SNR = 10.



**Fig. 3.** Parabolic curves: SNR = 10, OU error process:  $y_k$ s versus their estimations.

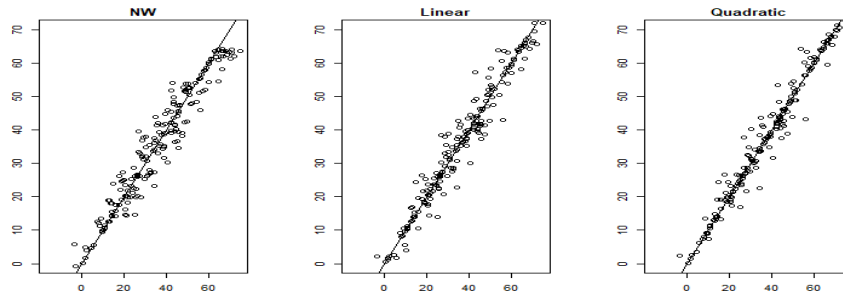
#### 4.2. Sinusoidal curves and nonlinear operator

The second simulation study consists in creating a set of 200 sinusoidal smooth curves by setting, for all  $k \in \{1, \dots, 200\}$  and  $t \in [0, 1]$ ,  $x_k(t) = \sqrt{k} \sin\{4(k - t)\}$ . We take the following nonlinear regression operator:

$$r(x) = \int_0^1 \frac{dt}{1 + x^2(t)}.$$

Figure 5 gives the plot of the  $y_k$ s versus the  $r(x_k)$ s for the two error processes.

Once again, the simulation results summarized in Table 2 show the superiority of the local linear and quadratic estimators over the Nadaraya–Watson estimator. For instance, when SNR = 10 there is a significant reduction of the



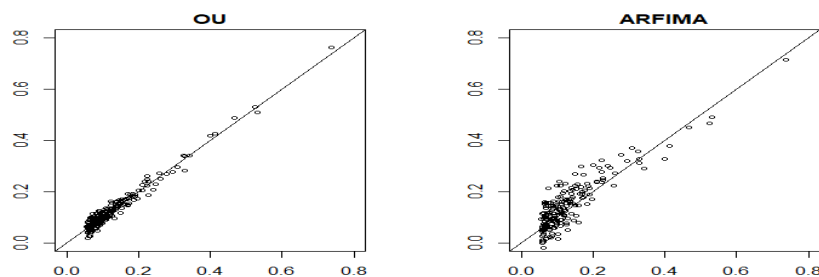
**Fig. 4.** Parabolic curves: SNR = 10, ARFIMA error process:  $y_k$ s versus their estimations.

RMSE, 81% for OU and 51% for ARFIMA (respectively 90% for OU and 57% for ARFIMA), of the local linear estimator (respectively the local quadratic estimator) over the NW estimator. As in the parabolic curves situation, the performance of the estimators is affected by the nature of the error process. To illustrate this point, we displayed in Figures 6–7 the observed responses in terms of their estimations when SNR = 10. Moreover, it seems from these figures that the linear and the quadratic estimators are more robust with respect to some high values (outliers) than the classical NW estimator. Finally, The shape of the curves has an impact on the quality of the estimation. The estimates obtained are less efficient than those of parabolic curves.

**Table 2**

Sinusoidal curves: RMSE of each estimator in terms of error processes, ARFIMA and OU, for different values of SNR.

Error Process	SNR	NW	Linear	Quadratic
OU	20	0.10374	0.01430	0.01016
	10	0.11012	0.02055	0.01147
	5	0.12091	0.03200	0.01763
	3	0.13366	0.04515	0.02892
	2	0.14819	0.06169	0.04375
ARFIMA	20	0.12086	0.04064	0.03324
	10	0.15040	0.07399	0.06397
	5	0.20712	0.13691	0.11923
	3	0.27356	0.20925	0.18117
	2	0.34200	0.28051	0.24283



**Fig. 5.** Sinusoidal curves:  $y_k$ s versus  $r(x_k)$ s with SNR = 10.

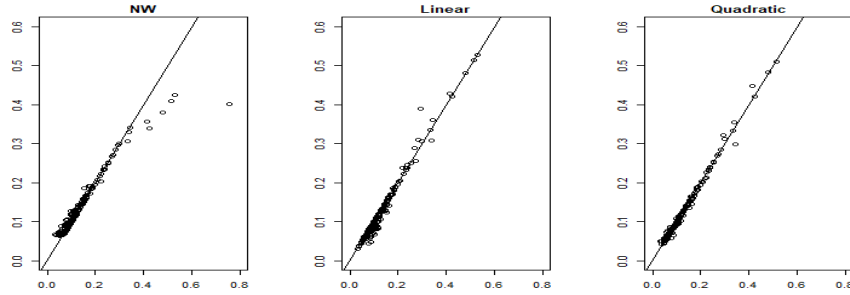


Fig. 6. Sinusoidal curves: SNR = 10, OU error process:  $y_k$ s versus their estimations.

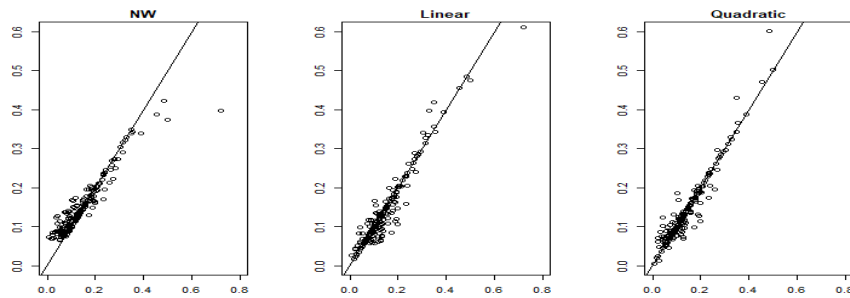


Fig. 7. Sinusoidal curves: SNR = 10, ARFIMA error process:  $y_k$ s versus their estimations.

## 5. Application to real data

This section aims to confirm the superiority of the local linear and quadratic estimators over the (local constant) Nadaraya–Watson one using two well-known functional data sets: the spectrometric curves (case of independent errors) and El Niño time series (case of dependent errors).

### 5.1. Spectrometric data

For the  $i$ th unit among 215 pieces of finely chopped meat, we observe one spectrometric curve  $x_i$  which corresponds to the absorbency measured at 100 wavelengths, and its fat content  $y_i$  is obtained by an analytical chemical process.

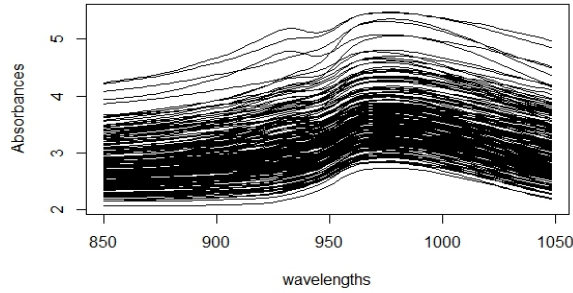
We also compare the performance of the three estimators by computing the relative mean squared errors of the 215 estimated values  $\hat{y}_i$ . The corresponding results are shown in Table 3. Furthermore, we compare the estimation performance of the three estimators constructed from the semi-metric based on the second derivative because of the smoothness of the spectrometric curves; see Figure 8.

When the  $L_2$  metric is used, there is an important gain of 74% (respectively 68%) in RMSE of linear estimator (respectively quadratic estimator) compared to the NW one. The use of the semi-metric based on the second derivative

**Table 3**

Spectrometric data: RMSE of each estimator with  $L_2$  metric and semi-metric based on second derivative.

Distance	NW	Linear	Quadratic
$L_2$	0.24886	0.06371	0.07866
Second derivative	0.01292	0.00659	0.00635

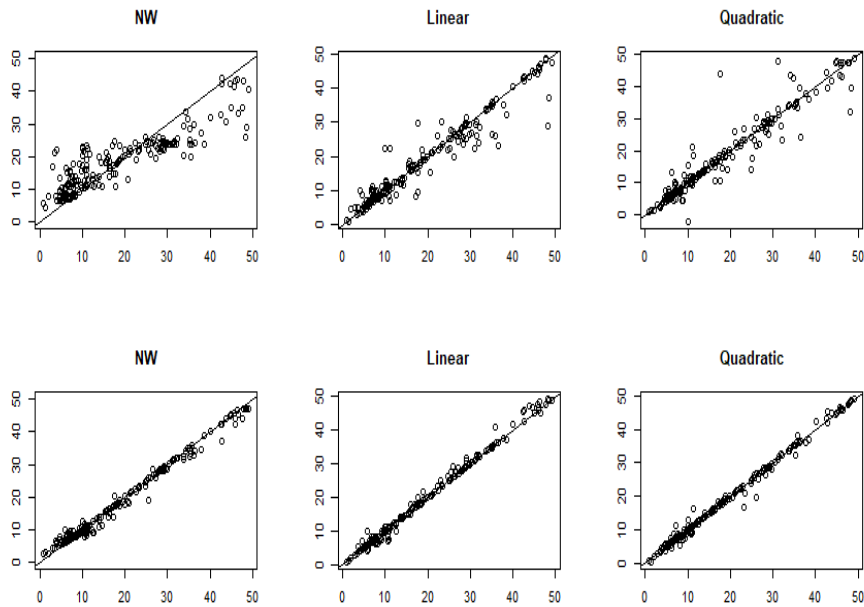


**Fig. 8.** The spectrometric curves.

significantly improves the quality of the estimation of each estimator, but the linear and quadratic estimators outperform the NW estimator. Moreover, it is seen that for this data set, there is no need to use a higher-order approximation since the local linear estimator already has a very good performance. All these facts are illustrated in Figure 9, where the estimated fat values of each estimator are plotted versus the observed ones with respectively  $L_2$  metric and second derivative semi-metric.

### 5.2. El Niño data

The second application concerns the El Niño time series which gives monthly sea surface temperatures over 54 years, from June, 1950 to May, 2004. This series was considered by [23] for dependent functional data. The data are recorded as a sequence of 648 real values as follows.



**Fig. 9.** Spectrometric data: Estimated versus observed fat values for the three estimators with  $L_2$  metric (top panels) and second derivative semi-metric (bottom panels).

Let  $Z(1), \dots, Z(648)$  be the El Niño time series. Then as in [23], we can construct subsequences by setting  $z_i(t) = Z\{12(i-1) + t\}$  for all  $t \in \{1, \dots, 12\}$  and  $i \in \{1, \dots, 54\}$ . Because the climatic phenomenon is changing continuously over time, it is natural to consider each annual curve as a continuous curve  $\{z_i(t) : t \in [0; 12]\}$  observed at 12 discretized points. Therefore, this time series can be viewed as a sample of 54 dependent functional data.

In order to study the performance of our estimators we will ignore the 54th year and will predict it from the 53 previous ones. That is, for all  $i \in \{1, \dots, 52\}$  and any fixed  $\delta \in \{1, \dots, 12\}$ , we take  $Y_i(\delta) = z_{i+1}(\delta)$ . In this way, we build a sample of 52 dependent pairs  $(Y_1(\delta), z_1), \dots, (Y_{52}(\delta), z_{52})$  where the  $Y_i(\delta)$ s are real-valued random variables and the  $z_i$ s are functional data.

Our main goal here is to predict, for each of the 12 months, the sea surface temperature of the 54th year by means of the Niño data of the 53 previous years. The corresponding estimates based on  $L^2$  metric are numerically illustrated by computing the mean squared error

$$\sum_{\delta=1}^{12} \{\hat{y}_{53}(\delta) - y_{53}\}^2$$

over the 12 estimated temperatures. They are as follows:

NW	Linear	Quadratic
0.233	0.617	2.624

While the performance of the local linear estimate is good, there is no real improvement of these local polynomial estimates over the classical estimates for prediction purposes. Moreover, it is seen for this data set that the higher-order approximation does not bring a major improvement over the local constant or linear estimates. This may be due to the fact that the bandwidths used in the three local estimates, local constant ( $p = 0$ ), linear ( $p = 1$ ) and quadratic ( $p = 2$ ) are very different and this has a significant effect on the bias and the variance. Besides, the local polynomial estimate depends very much on the inverse of a matrix and this may be subject to high numerical instability when the sample size  $n$  and the order  $p$  increase.

Other semi-metrics could be used, such as the one based on principal component analysis. However, the main goal here is not really the choice of the semi-metric but a comparison in terms of performance of the three estimators corresponding to  $p \in \{0, 1, 2\}$ .

## 6. Further comments

We established some asymptotic results for estimating the regression operator and its Fréchet derivatives from a polynomial approximation of any order in a Banach space equipped with a metric. We showed how these asymptotic performances are affected by the order of the local polynomial estimation. When the curves live in a semi-metric space with a semi-metric based on the derivatives of the curves, namely for every integer  $\ell \geq 2$ ,

$$\sqrt{\int \{x_k^{(\ell)}(t) - x_{k'}^{(\ell)}(t)\}^2 dt},$$

we conducted simulations studies with  $\ell = 2$  using the same type of curves and regression operators and arrived at the same conclusions as in the metric space considered in Section 4 ( $\ell = 0$ ). However, the theoretical results in this work cannot be generalized to a semi-metric space because the polynomial expansion of the regression operator is not available. The case with random designs and correlated errors involves more technical difficulties and should be treated separately in possible further work. For instance, we can refer to [10] when the Nadaraya–Watson estimator of  $r$  is used. Extensions of this work to semi-parametric functional models could also be considered in future research.

## Appendix

In the following proofs, we will use the fact that from assumptions (3), (4) and (5), for any  $\nu \in \{1, 2\}$  and as  $h \rightarrow 0$ ,

$$C_j^\nu(h) = K^\nu(1) - \int_0^1 \tau_j^h(s) s^j (K^\nu)'(s) ds \rightarrow C_j^\nu$$

and

$$C_{k,\ell}^{1,1}(h) = K^2(1) - K(1) \int_0^1 \psi_{k,\ell}(hs, h) K'(s) ds \\ - K(1) \int_0^1 \psi_{k,\ell}(h, hs) K'(s) ds + \int_0^1 \int_0^1 \psi_{k,\ell}(hs, ht) K'(s) K'(t) ds dt \rightarrow C_{k,\ell}^{1,1},$$

where  $C_j^v$  and  $C_{k,\ell}^{1,1}$  are as in (6).

**Proof of Theorem 1.** Write  $\mathbf{r}_n = (r(x_1), \dots, r(x_n))^\top$  and define the  $(p+1) \times (p+1)$  matrix  $\mathbf{S}_n = \mathbf{X}_n^\top \mathbf{W}_n \mathbf{X}_n$  with  $(k, \ell)$ th element given by

$$s_{k+\ell} = \sum_{i=1}^n d^{k+\ell}(x_i, x) K_h\{d(x_i, x)\}.$$

It follows from the definition of the estimator (2) that  $E\{\widehat{\boldsymbol{\beta}}_n(x)\} = \mathbf{S}_n^{-1} \mathbf{X}_n^\top \mathbf{W}_n \mathbf{r}_n$ . The Taylor expansion of  $r(x_i)$  up to order  $p+1$  yields

$$r(x_i) = r(x) + d(x_i, x) r_1(x) + \dots + \frac{d^{p+1}(x_i, x)}{(p+1)!} r_{p+1}(x) + o\{d^{p+1}(x_i, x)\}$$

and thus

$$\mathbf{r}_n = \mathbf{X}_n \boldsymbol{\beta}(x) + \{\beta_{p+1} + o(1)\} \begin{pmatrix} d^{p+1}(x_1, x) \\ \vdots \\ d^{p+1}(x_n, x) \end{pmatrix}.$$

As a result, the bias in the estimation of  $\boldsymbol{\beta}(x)$  is

$$E\{\widehat{\boldsymbol{\beta}}_n(x)\} - \boldsymbol{\beta}(x) = \{1 + o(1)\} \beta_{p+1} \mathbf{S}_n^{-1} \mathbf{c}_n,$$

where  $\mathbf{c}_n = (s_{p+1}, \dots, s_{2p+1})^\top$ . We approximate the elements of the matrix  $\mathbf{S}_n$  as follows. For each  $i \in D_x(h)$ , we have

$$K_h\{d(x_i, x)\} = K(1) \mathbf{1}_{B(x,h)}(x_i) - \int_0^1 \mathbf{1}_{B(x,hs)}(x_i) K'(s) ds.$$

Then

$$s_{k+\ell} = K(1) n h^{k+\ell} \varphi_{k+\ell}(h) - n h^{k+\ell} \int_0^1 \varphi_{k+\ell}(hs) s^{k+\ell} K'(s) ds = n h^{k+\ell} \varphi_{k+\ell}(h) \left\{ K(1) - \int_0^1 \tau_{k+\ell}^h(s) s^{k+\ell} K'(s) ds \right\}.$$

From assumption (4) with  $\nu = 1$ , we then have, for all  $k, \ell \in \{0, \dots, p\}$ ,

$$s_{k+\ell} = n h^{k+\ell} \varphi_{k+\ell}(h) C_{k+\ell}^1(h) \{1 + o(1)\}.$$

It follows that  $\mathbf{S}_n = n\{1 + o(1)\} \mathbf{H} \boldsymbol{\Sigma}_h \mathbf{H}$ . Now,

$$\mathbf{c}_n = \begin{pmatrix} s_{p+1} \\ \vdots \\ s_{2p+1} \end{pmatrix} = \begin{pmatrix} n h^{p+1} \varphi_{p+1}(h) C_{p+1}^1(h) \{1 + o(1)\} \\ \vdots \\ n h^{2p+1} \varphi_{2p+1}(h) C_{2p+1}^1(h) \{1 + o(1)\} \end{pmatrix} = n h^{p+1} \mathbf{H} \tilde{\boldsymbol{\mu}}_h \{1 + o(1)\},$$

so that  $\mathbf{S}_n^{-1} \mathbf{c}_n = h^{p+1} \mathbf{H}^{-1} \boldsymbol{\Sigma}_h^{-1} \tilde{\boldsymbol{\mu}}_h \{1 + o(1)\}$  and hence

$$E\{\widehat{\boldsymbol{\beta}}_n(x)\} - \boldsymbol{\beta}(x) = \beta_{p+1} h^{p+1} \mathbf{H}^{-1} \boldsymbol{\Sigma}_h^{-1} \tilde{\boldsymbol{\mu}}_h \{1 + o(1)\},$$

from which the asymptotic bias expression of  $\hat{r}_j(x)$ , for all  $j \in \{0, \dots, p\}$ , in Theorem 1 is obtained.  $\square$

**Proof of Theorem 2.** Define the  $n \times n$  matrix  $\mathbf{V}_n = \{\rho_\epsilon(i - j)\}$  and the  $(p + 1) \times (p + 1)$  matrix  $\mathbf{S}_n^* = \mathbf{X}_n^\top \mathbf{W}_n \mathbf{V}_n \mathbf{W}_n \mathbf{X}_n$ . Noting that  $\text{var}(\mathbf{Y}_n) = \mathbf{V}_n$ , we can see from the definition of the estimator (2) that

$$\text{var}\{\widehat{\boldsymbol{\beta}}_n(x)\} = \mathbf{S}_n^{-1} \mathbf{S}_n^* \mathbf{S}_n^{-1},$$

where the  $(k, \ell)$ th element of the matrix  $\mathbf{S}_n^*$  is

$$\begin{aligned} s_{k,\ell}^* &= \sum_{i=1}^n \sum_{j=1}^n d^k(x_i, x) d^\ell(x_j, x) \Delta_i(x) \Delta_j(x) \rho_\epsilon(i - j) \\ &= \sum_{i=1}^n d^{k+\ell}(x_i, x) \Delta_i^2(x) \rho_\epsilon(0) + \sum_{\substack{1 \leq i, j \leq n \\ i \neq j}} d^k(x_i, x) d^\ell(x_j, x) \Delta_i(x) \Delta_j(x) \rho_\epsilon(i - j). \end{aligned}$$

For the first term, following the same steps as for deriving the asymptotic approximation of the elements  $s_{k+\ell}$  of the matrix  $\mathbf{S}_n$ , we obtain

$$\sum_{i=1}^n d^{k+\ell}(x_i, x) \Delta_i^2(x) \rho_\epsilon(0) = nh^{k+\ell} \varphi_{k+\ell}(h) C_{k+\ell}^2(h) \rho_\epsilon(0) \{1 + o(1)\}.$$

For the second term, let  $D_x(h) = \{i : d(x, x_i) \leq h\} = \{i : x_i \in B(x, h)\}$ . Then using the assumption (5), we deduce that

$$\begin{aligned} \sum_{\substack{1 \leq i, j \leq n \\ i \neq j}} d^k(x_i, x) d^\ell(x_j, x) \Delta_i(x) \Delta_j(x) \rho_\epsilon(i - j) &= K^2(1) \sum_{\substack{i, j \in D_x(h) \\ i \neq j}} d^k(x_i, x) d^\ell(x_j, x) \rho_\epsilon(i - j) \\ &\quad - K(1) \int_0^1 ds K'(s) \sum_{\substack{i \in D_x(h) \cap D_x(hs) \\ j \in D_x(h) \\ i \neq j}} d^k(x_i, x) d^\ell(x_j, x) \rho_\epsilon(i - j) \\ &\quad - K(1) \int_0^1 ds K'(s) \sum_{\substack{j \in D_x(h) \cap D_x(hs) \\ i \in D_x(h) \\ i \neq j}} d^k(x_i, x) d^\ell(x_j, x) \rho_\epsilon(i - j) \\ &\quad + \int_0^1 \int_0^1 ds ds' K'(s) K'(s') \sum_{\substack{i, j \in D_x(hs) \\ i \neq j}} d^k(x_i, x) d^\ell(x_j, x) \rho_\epsilon(i - j). \end{aligned}$$

Given that  $D_x(hs) \subset D_x(h)$  for all  $s \in [0, 1]$ , we have  $D_x(h) \cap D_x(hs) = D_x(hs)$ , it follows that

$$\begin{aligned} \sum_{\substack{1 \leq i, j \leq n \\ i \neq j}} d^k(x_i, x) d^\ell(x_j, x) \Delta_i(x) \Delta_j(x) \rho_\epsilon(i - j) &= K^2(1) n^2 h^{k+\ell} \mu_{k,\ell}(h, h) - K(1) n^2 h^{k+\ell} \int_0^1 \mu_{k,\ell}(hs, h) K'(s) ds \\ &\quad - K(1) n^2 h^{k+\ell} \int_0^1 \mu_{k,\ell}(h, hs) K'(s) ds + n^2 h^{k+\ell} \int_0^1 \int_0^1 \mu_{k,\ell}(hs, ht) K'(s) K'(t) ds dt \\ &= n^2 h^{k+\ell} \mu_{k,\ell}(h, h) \{K^2(1) - K(1) \int_0^1 \psi_{k,\ell}(s, 1) K'(s) ds \\ &\quad - K(1) \int_0^1 \psi_{k,\ell}(1, s) K'(s) ds + \int_0^1 \int_0^1 \psi_{k,\ell}(s, t) K'(s) K'(t) ds dt\} \{1 + o(1)\} \\ &= n^2 h^{k+\ell} \mu_{k,\ell}(h, h) C_{k,\ell}^{1,1}(h) \{1 + o(1)\}. \end{aligned}$$

Therefore, combining the two asymptotic expressions, we have

$$s_{k,\ell}^* = nh^{k+\ell} \varphi_{k+\ell}(h) C_{k+\ell}^2(h) \rho_\epsilon(0) \{1 + o(1)\} + n^2 h^{k+\ell} \mu_{k,\ell}(h, h) C_{k,\ell}^{1,1}(h) \{1 + o(1)\}.$$

It follows that

$$\mathbf{S}_n^* = n \mathbf{H} \boldsymbol{\Sigma}_h^{(2)} \mathbf{H} \rho_\epsilon(0) \{1 + o(1)\} + n^2 \mathbf{H} \boldsymbol{\Sigma}_h^{(1,1)} \mathbf{H} \{1 + o(1)\}$$

and

$$\begin{aligned}
\text{var}\{\hat{\beta}(x)\} &= \mathbf{S}_n^{-1} \mathbf{S}_n^* \mathbf{S}_n^{-1} \\
&= \frac{1}{n} \mathbf{H}^{-1} \Sigma_h^{-1} \mathbf{H}^{-1} \left\{ n \mathbf{H} \Sigma_h^{(2)} \mathbf{H} \rho_\epsilon(0) + n^2 \mathbf{H} \Sigma_h^{(1,1)} \mathbf{H} \right\} \frac{1}{n} \mathbf{H}^{-1} \Sigma_h^{-1} \mathbf{H}^{-1} \{1 + o(1)\} \\
&= \frac{1}{n} \mathbf{H}^{-1} \Sigma_h^{-1} \left\{ \Sigma_h^{(2)} \rho_\epsilon(0) + n \Sigma_h^{(1,1)} \right\} \Sigma_h^{-1} \mathbf{H}^{-1} \{1 + o(1)\}.
\end{aligned}$$

This concludes the argument.  $\square$

**Proof of Corollary 3.** Let  $n_x(h) = \text{card}\{i : x_i \in B(x, h)\} = \text{card}\{D_x(h)\}$ , so that  $\varphi_0(h) = n_x(h)/n$ . Then under assumptions (1), (3), (4), and (5), we have, for some constant  $C$ ,

$$\begin{aligned}
|\mu_{0,0}(h, h)| &= \frac{1}{n^2} \sum_{\substack{1 \leq i, j \leq n \\ i \neq j}} |\rho_\epsilon(i - j)| \mathbf{1}_{B(x, h)}(x_i) \mathbf{1}_{B(x, h)}(x_j) \\
&\leq \frac{1}{n^2} \sum_{\substack{i, j \in D_x(h) \\ i \neq j}} C |i - j|^{-\gamma} \leq \frac{2C}{n^2} \sum_{\substack{i, j \in D_x(h) \\ i < j}} |i - j|^{-\gamma} \leq \frac{2C}{n^2} \sum_{k=1}^{n_x(h)} \{n_x(h) - k\} k^{-\gamma} \\
&\leq \frac{2C n_x(h)^{2-\gamma}}{n^2} \times \frac{1}{n_x(h)} \sum_{k=1}^{n_x(h)} \left\{1 - \frac{k}{n_x(h)}\right\} \left\{\frac{k}{n_x(h)}\right\}^{-\gamma} \\
&\asymp \frac{n_x(h)^{2-\gamma}}{n^2} \int_0^1 (1-u) u^{-\gamma} du \asymp \frac{\varphi_0(h)^2}{\{n\varphi_0(h)\}^\gamma}
\end{aligned}$$

where the notation  $a_n \asymp b_n$  is used to indicate that  $a_n = O(b_n)$  or  $b_n = O(a_n)$  as  $n \rightarrow \infty$ . Under the assumptions of Corollary 1, we deduce that

$$\text{var}\{\widehat{r}_j(x)\} = O\left\{\frac{1}{nh^{2j}\varphi_0(h)}\right\} + O\left\{\frac{\mu_{0,0}(h)}{h^{2j}\varphi_0^2(h)}\right\} = O\left\{\frac{1}{nh^{2j}\varphi_0(h)}\right\} + O\left[\frac{1}{h^{2j}\{n\varphi_0(h)\}^\gamma}\right] = O\left[\frac{1}{h^{2j}\{n\varphi_0(h)\}^\gamma}\right]$$

since  $n\varphi_0(h) \rightarrow \infty$  as  $n \rightarrow \infty$  and  $\gamma \in (0, 1)$ .  $\square$

**Proof of Corollary 4.** From assumptions (3), (4) and (5) and using the same notations as in the previous proof, we have, for some constant  $C$ ,

$$|\mu_{0,0}(h, h)| = \frac{1}{n^2} \sum_{\substack{1 \leq i, j \leq n \\ i \neq j}} |\rho_\epsilon(i - j)| \mathbf{1}_{B(x, h)}(x_i) \mathbf{1}_{B(x, h)}(x_j) \leq \frac{2}{n^2} \sum_{\substack{i, j \in D_x(h) \\ i \leq j}} |\rho_\epsilon(i - j)| \leq \frac{2}{n^2} \sum_{k=1}^{n_x(h)} \{n_x(h) - k\} |\rho_\epsilon(k)| \leq \frac{n_x(h)}{n^2} \sum_{k=1}^n |\rho_\epsilon(k)|.$$

Given that the error process is of short memory, we have  $\sum_{k=1}^n |\rho_\epsilon(k)| = O(1)$ . It then follows that

$$\mu_{0,0}(h, h) = O\{n_x(h)/n^2\} = O\{n^{-1}\varphi_0(h)\}.$$

Therefore,

$$\text{var}\{\widehat{r}_j(x)\} = O\left\{\frac{1}{nh^{2j}\varphi_0(h)}\right\},$$

which concludes the argument.  $\square$

## Acknowledgments

We would like to thank three anonymous reviewers, Prof. Philippe Vieu (the Managing Editor), and Prof. Christian Genest (the Editor-in-Chief) for their constructive comments and valuable suggestions which have led to an improved version of this article.



## References

- [1] A. Ait-Saïdi, F. Ferraty, R. Kassa, P. Vieu, Cross-validated estimation in the single-function index model, *Statistics* 42 (2008) 475–494.
- [2] G. Aneiros, E. Bongiorno, R. Cao, P. Vieu, *Contributions to Statistics: Functional Statistics and Related Fields*, Springer, New York, 2017.
- [3] G. Aneiros, P. Vieu, Nonparametric time series prediction: A semi-functional partial linear modeling, *J. Multivariate Anal.* 99 (2008) 834–857.
- [4] G. Aneiros, P. Vieu, Partial linear modeling with multi-functional covariates, *Comput. Statist.* 30 (2015) 647–671.
- [5] A. Baillo, A. Grané, Locally modeled regression and functional data, *J. Multivariate Anal.* 100 (2009) 102–111.
- [6] J. Barrientos-Marin, F. Ferraty, P. Vieu, Local linear regression for functional predictor and scalar response, *J. Nonparametr. Statist.* 22 (2010) 617–632.
- [7] K. Benhenni, D. Degras, Local polynomial estimation of the mean function and its derivatives based on functional data and regular designs, *ESAIM: PS* 18 (2014) 881–899.
- [8] K. Benhenni, F. Ferraty, M. Rachdi, P. Vieu, Local smoothing regression with functional data, *Comput. Statist.* 22 (2007) 353–369.
- [9] K. Benhenni, S. Hedli-Griche, M. Rachdi, Estimation of the regression operator from functional fixed-design with correlated errors, *J. Multivariate Anal.* 101 (2010) 476–490.
- [10] K. Benhenni, S. Hedli-Griche, M. Rachdi, Regression models with correlated errors based on functional random design, *Test* 26 (2017) 1–21.
- [11] J. Beran, *Statistics for Long-Memory Processes*, Chapman & Hall, New York, 1994.
- [12] A. Berline, A. Elamine, A. Mas, Local linear regression for functional data, *Ann. Inst. Stat. Math.* 63 (2011) 1047–1075.
- [13] E. Boj, D. Delicado, J. Fortiana, Distance-based local linear regression for functional predictor, *Comput. Statist. Data Anal.* 54 (2004) 429–437.
- [14] D. Bosq, *Linear Processes in Function Spaces: Theory and Application*. Springer, New York, 2000.
- [15] M. Brabec, K. Jilek, Simplified radon entry rate estimation methodology from tracer and radon concentrations based on functional data analysis approach, In: *Horizons in World Physics*, A. Reimer (Ed.), Nova Science Publishers, 2011, pp. 187–205
- [16] H. Cartan, *Cours de calcul différentiel*, Éditions Hermann, Paris, 2007.
- [17] D. Chen, P. Hall, H.-G. Müller, Single and multiple index function regression models with nonparametric link, *Ann. Statist.* 39 (2011) 1720–1747.
- [18] A. Cuevas, A partial overview of the theory of statistics with functional data, *J. Stat. Plann. Inf.* 147 (2014) 1–23.
- [19] G. Chagny, A. Roche, Adaptive estimation in the functional nonparametric regression model, *J. Multivariate Anal.* 146 (2016) 105–118.
- [20] J. Fan, I. Gijbels, *Local Polynomial Modeling and Its Applications*, Chapman & Hall, London, 1996.
- [21] M. Febrero, P. Galeano, W. González-Manteiga, Outlier detection in functional data by depth measures with application to identify abnormal  $\text{NO}_x$  levels, *Environmetrics* 19 (2008) 331–345.
- [22] F. Ferraty, A. Mas, P. Vieu, Nonparametric regression on functional data: Inference and practical aspects, *Aust. N. Z. J. Stat.* 49 (2007) 267–286.
- [23] F. Ferraty, A. Rabhi, P. Vieu, Conditional quantiles for dependent functional data with application to the climatic El Niño phenomenon, *Sankhyā* 67 (2005) 378–398.
- [24] F. Ferraty, P. Vieu, Nonparametric analysis for functional data, with application in regression method, *J. Nonparametr. Statist.* 16 (2004) 111–125.
- [25] F. Ferraty, P. Vieu, *Nonparametric Analysis for Functional Data: Theory and Practice*, Springer, New York, 2006.
- [26] M. Francisco-Fernández, J. Vilar-Fernández, Local polynomial regression estimation with correlated errors, *Comm. Statist. Theory Methods* 30 (2001) 1271–1293.
- [27] G. Geenens, A Nonparametric Functional Method for Signature Recognition, In: *Recent Advances in Functional Data Analysis and Related Topics*, F. Ferraty (Ed.), 2001, pp. 141–147.
- [28] A. Goia, P. Vieu, A partitioned single functional index model, *Comput. Statist.* 30 (2015) 673–692.
- [29] P. Hall, H.-G. Müller, F. Yao, Estimation of functional derivatives, *Ann. Statist.* 37 (2009) 3307–3329.
- [30] G. He, H.-G. Müller, J.L. Wang, Functional canonical analysis for square integrable stochastic processes, *J. Multivariate Anal.* 85 (2003) 54–77.
- [31] L. Kara, A. Laksaci, M. Rachdi, P. Vieu, Data-driven kNN estimation in nonparametric functional data analysis, *J. Multivariate Anal.* 153 (2017) 176–1898.
- [32] N. Ling, P. Vieu, Nonparametric modeling for functional data: Selected survey and tracks for future, *Statistics: A Journal of Theoretical and Applied Statistics* (2018) 934–949.
- [33] E. Masry, Local polynomial regression estimation with correlated errors, *J. Multivariate Anal.* 86 (2003) 330–359.
- [34] E. Masry, J. Fan, Local polynomial estimation of regression functions for mixing processes, *J. Multivariate Anal.* 86 (1997) 330–359.
- [35] P.R. McGillivray, D.W. Oldenburg, Methods for calculating Fréchet derivatives and sensitivities for the non-linear inverse problem: A comparative study, *Geophysical Prospecting* 38 (1990) 499–524.
- [36] A. Pérez-González, J. Vilar-Fernández, W. González-Manteiga, Asymptotic properties of local polynomial regression with missing data and correlated errors, *Ann. Inst. Statist. Math.* 61 (2009) 85–109.
- [37] M. Rachdi, A. Laksaci, J. Demongeot, A. Abdali, F. Madani, Theoretical and practical aspects of the quadratic error in the local linear estimation of the conditional density for functional data, *Comput. Statist. Data Anal.* 73 (2014) 53–68.
- [38] J.O. Ramsay, B.W. Silverman, *Applied Functional Data Analysis: Methods and Case Studies*, Springer-Verlag, New York, 2002.
- [39] J.O. Ramsay, B.W. Silverman, *Functional Data Analysis*, Springer, New York, 2005.
- [40] J. Sancho, J. Pastor, J. Martínez, M. García, Variability analysis by statistical control process and functional data analysis: Case of study applied to power system harmonics assessment, *Key Engineering Materials* 615 (2014) 118–123.
- [41] M. Wand, M. Jones, *Kernel Smoothing*, Chapman & Hall, London, 1995.
- [42] E. Zeidler, *Nonlinear Functional Analysis and Its Applications I: Fixed-Point Theorems*, Springer, New York, 1995.
- [43] K. Zoulikha, A. Laksaci, Functional quantile regression: Local linear modelisation, *Contributions to Statistics: Functional Statistics and Related Fields* (2017) 155–160.