JMVA-15-27

Efficiency in Multivariate Functional Nonparametric Models with Autoregressive Errors

S. Dabo-Niang[a,c], S. Guillas[b], C. Ternynck[a,*]

[a] *Laboratory LEM, University of Lille, Villeneuve-d'Ascq, France.*
[b] *University College London, London, United Kingdom.*
[c] *Modal team INRIA, Lille, France.*

Abstract

In this paper, we introduce a new procedure for the estimation in the nonlinear functional regression model where the explanatory variable takes values in an abstract function space and the residual process is autocorrelated. Moreover, we consider the case where the response variable takes its values in $\mathbb{R}^d$, $d \geq 1$. The procedure consists in a pre-whitening transformation of the dependent variable based on the estimated autocorrelation. We establish both consistency and asymptotic normality of the regression function estimate. For kernel methods encountered in the literature, the correlation structure is commonly ignored (the so-called "working independence estimator"); we show here that there is a strong benefit in taking into account the autocorrelation in the error process. We also find that the improvement in efficiency can be large in our functional setting, up to 25% in the presence of high autocorrelation levels. We discover that the additional step of iterating the fitting process actually deteriorates the estimation. We illustrate the skills of the methods on simulations as well as on application on ozone levels over the US.

Keywords: Autoregressive process, Functional data, Kernel regression, Pre-whitening, Time Series

1. Introduction

The use of functional random variables is spreading in statistical analyses due to the availability of high frequency data and of new mathematical strategies to deal with such statistical objects. The field is known as Functional Data Analysis (FDA). Applications of FDA are growing across fields as diverse as energy studies (Antoniadis et al., 2014), linguistics (Aston et al., 2010), atmospheric chemistry (Park et al., 2013), and human vision (Ogden and Greene, 2010). The functional variables are mainly curves, but surfaces and manifolds are nowadays considered (e.g. Guillas and Lai, 2010; Sangalli et al., 2013). For an introduction to this field as well as illustrations and applications, see

*Corresponding author
Email addresses: sophie.dabo@univ-lille3.fr (S. Dabo-Niang), s.guillas@ucl.ac.uk (S. Guillas), ternynck.camille@gmail.com (C. Ternynck)

Ramsay and Silverman (2005). Besides, Ferraty and Vieu (2006) present nonparametric methods, suited to such functional regression, with a more mathematical flavor. More recently, Cuevas (2014) provides an updated survey of the state of the art in FDA theory.

Among the nonparametric functional regression methods, the kernel estimator is often used to estimate the regression operator. It yields almost sure consistency in the case of an independent sample (Ferraty and Vieu, 2002) or an $\alpha-$mixing sample (Ferraty et al., 2002a,b), but also asymptotic normality in the independent case (Ferraty et al., 2007) with exact computation of all the constants for its precise use in practice. Masry (2005) established the asymptotic normality of the nonparametric regression estimator for strongly mixing processes albeit with abstract expressions of the constants so this is more challenging to use in practice. Delsol (2007, 2009) generalized the results of Ferraty et al. (2007) to the case of an α-mixing dataset.

In this paper, we consider the regression of a multivariate random variable onto a functional random variable. The estimation of the regression function is tackled by means of a nonparametric kernel approach. The existing kernel regression estimators dealing with functional explanatory variables are for scalar response; we have not found existing research on functional nonparametric modeling for multivariate response. With multivariate explanatory variables and a multivariate response, Xiang et al. (2013) proposed a kernel estimate of the regression function. Our regression model below is an extension of (Xiang et al., 2013):

$$\mathbf{Y}_t = \mathbf{m}(X_t) + \mathbf{u}_t, \qquad t = 1, \ldots, T, \tag{1}$$

where $\mathbf{Y}_t = (Y_{t,1}, \ldots, Y_{t,d})' \in \mathbb{R}^d$, $\mathbf{m}(X_t) = (m_1(X_t), \ldots, m_d(X_t))'$, the explanatory variable is functional (that is, X_t takes values in some possibly infinite-dimensional space), $\mathbf{u}_t = (u_{t,1}, \ldots, u_{t,d})'$. Moreover, the stationary residual process $\mathbf{u}_t$ is autocorrelated and independent of X_t. We do not necessarily assume that $(X_t, \mathbf{Y}_t)_t$ is strictly stationary, second order stationarity suffices.

Although, for the kernel methods proposed in the literature, it is generally better to ignore the correlation structure entirely (the so-called "working independence estimator", e.g. Ruckstuhl et al. (2000), Lin and Carroll (2000)), we show here that taking into account the autocorrelation of the error process helps improve the estimation of the regression function.

We extend the kernel-based procedure proposed by Xiao et al. (2003) for estimating $m(x)$ in the time series regression model for multivariate explanatory variables x to a functional setting. Xiao et al. (2003) showed that their procedure is more efficient than the conventional local polynomial method. The main idea is to transform the original regression model, so that this transformed regression has a residual term that is uncorrelated. This transformation depends on the function $\mathbf{m}(\cdot)$ and on the parameters of the autoregressive representation of $\mathbf{u}$, since the regression function is nonlinear. The error correlation structure is assumed to have an autoregressive representation. Firstly, the parameters of the autoregressive representation are estimated. In a second

step, a transformation $\widehat{\underline{\mathbf{Y}}}_t$ of the dependent variable $\mathbf{Y}_t$ is constructed by plugging in the estimated autocorrelation parameter. Finally, the estimation of $\mathbf{m}$ is carried out on this filtered series $\widehat{\underline{\mathbf{Y}}}_t$.

The remainder of the paper is organized as follows. In Section 2, we introduce the estimation method as well as the assumptions. We then provide asymptotic results for the estimator proposed. Section 3 is devoted to a simulation case study and an illustration of our method for ozone levels over the US. The conclusion is done in Section 4 while the proofs are given in the Appendix.

2. Assumptions and main results

Suppose that we have a sample $\{(X_1, \mathbf{Y}_1), \ldots, (X_T, \mathbf{Y}_T)\}$, where X_t, $t = 1, \ldots, T$, is a random variable taking its values in a semi-metric space $(\mathcal{C}, d)$ of infinite dimension and $\mathbf{Y}_t \in \mathbb{R}^d$ is the response from the nonparametric regression (1). We assume that the residual process $\mathbf{u}_t \in \mathbb{R}^d$ is stationary, has mean $\mathbf{0}$ with cross-covariance (auto-covariance in the univariate case) $\gamma_{\mathbf{u}}$ and has the following invertible linear process representation (with bounded coefficients):

$$\mathbf{u}_t = \sum_{k=0}^{\infty} \Psi_k \mathbf{e}_{t-k} = \Psi(L)\mathbf{e}_t \tag{2}$$

where $\Psi_0 = I$ (identity matrix), $\Psi(L) = \sum_{k=0}^{\infty} \Psi_k L^k$ is a $d \times d$ matrix in the lag operator L ($L^k(\mathbf{e}_t) = \mathbf{e}_{t-k}$), the (i,j)th element of $\Psi(L)$ is $\psi_{ij}(L) = \sum_{k=0}^{\infty} c_{ij}(k)L^k$, the $\mathbf{e}_t \in \mathbb{R}^d$ form a white noise process with mean $\mathbb{E}(\mathbf{e}_t) = \mathbf{0}$, $\mathbb{E}(\mathbf{e}_t\mathbf{e}_t') = \Sigma_{\mathbf{e}}$ is a positive definite matrix, $\mathbb{E}(\mathbf{e}_t\mathbf{e}_{t+k}) = \mathbf{0}$ for $k \neq 0$ and $\mathbb{E}[|e_{t,j}|] < \infty$, $\forall j = 1, \ldots, d$.

Let $\Psi(L)^{-1} = \Pi(L) = I - \sum_{k=1}^{\infty} \Pi_k L^k$ with $\Pi_0 = I$, or as done for Ψ let the (i,j)th element of $\Pi(L)$ be $\pi_{ij}(L) = \sum_{k=0}^{\infty} a_{ij}(k)L^k$. So we have, the infinite autoregressive representation

$$\Pi(L)\mathbf{u}_t = \mathbf{e}_t. \tag{3}$$

Note that stationary, causal and invertible vector ARMA processes $\mathbf{u}_t - \sum_{k=1}^{p} \Phi_k \mathbf{u}_{t-k} = \mathbf{e}_{t-k} - \sum_{k=1}^{q} \Theta_k \mathbf{e}_{t-k}$ can be represented as in (2)-(3) if all roots of $det\{\Phi(L)\}$ and $det\{\Theta(L)\}$ are all greater than one in absolute value.

Here, we consider a truncated version of $\Pi(L)$ at order Q, that is $\Pi(L) = I - \sum_{k=1}^{Q} \Pi_k L^k$, where the truncation parameter Q is large enough. Applying $\Pi(L)$ to the regression in Equation (1), we obtain $\Pi(L)\mathbf{Y}_t = \Pi(L)\mathbf{m}(X_t) + \mathbf{e}_t$. Then let the regression model $\underline{\mathbf{Y}}_t = \mathbf{m}(X_t) + \mathbf{e}_t$, with $\underline{\mathbf{Y}}_t = \mathbf{Y}_t - \sum_{k=1}^{Q} \Pi_k L^k(\mathbf{Y}_t - \mathbf{m}(X_t))$, so the error term in this transformed model is now uncorrelated. The matrix of coefficients $\{\Psi_k\}_{k=0}^{\infty}$ and the regression function $\mathbf{m}(\cdot)$ are unknown, except for the fact that $\mathbf{m}(\cdot)$ is a smooth function. If $\underline{\mathbf{Y}}_t$ were known then a nonparametric kernel regression of $\underline{\mathbf{Y}}_t$ on X_t would be more efficient than the conventional kernel estimation. In this work, we employ a Nadaraya-Watson estimator as introduced in Ferraty and Vieu (2004), Masry (2005), Dabo-Niang and Rhomari (2009) where for any sample $\{V_t, X_t\}$, the estimation of the regression of $V_t \in \mathbb{R}$ onto $X_t \in (\mathcal{C}, d(\cdot,\cdot))$ is

$$\frac{\sum_{t=1}^{T} V_t K\left(\frac{d(x,X_t)}{h}\right)}{\sum_{s=1}^{T} K\left(\frac{d(x,X_s)}{h}\right)}, \qquad x \in \mathcal{C}$$

where $K(\cdot)$ is a function over $[0,+\infty[$ called kernel, $h>0$ is the bandwidth parameter and $d(\cdot,\cdot)$ is a semi-metric. For $x \in \mathcal{C}$ fixed, let $\widehat{m}_j(x)$ be the corresponding estimator with $V_t = Y_{t,j}$ and let $\overline{m}_j(x)$ be the corresponding estimator when $V_t = \underline{Y}_{t,j}$. Let K_0 and K_1 be two kernels over $[0;+\infty[$, h_0 and h_1 the two corresponding bandwidths. Consider the estimator $\overline{\mathbf{m}}(x) = (\overline{m}_1(x), \ldots, \overline{m}_d(x))'$ where

$$\overline{m}_j(x) = \frac{\frac{1}{T\mathbb{E}\left[K_0\left(\frac{d(x,X_1)}{h_0}\right)\right]} \sum_{t=1}^{T} \underline{Y}_{t,j} K_0\left(\frac{d(x,X_t)}{h_0}\right)}{\frac{1}{T\mathbb{E}\left[K_0\left(\frac{d(x,X_1)}{h_0}\right)\right]} \sum_{s=1}^{T} K_0\left(\frac{d(x,X_s)}{h_0}\right)} = \frac{\overline{m}_{2,j}(x)}{\overline{m}_1(x)},\ j = 1, \ldots, d$$

In practice, the matrix of coefficients $\{\Pi_k\}_{k=1}^{Q}$ is unknown and therefore $\underline{\mathbf{Y}}_t$ is not computable, so the regression $\underline{\mathbf{Y}}_t = \mathbf{m}(X_t) + \mathbf{e}_t$ and $\overline{\mathbf{m}}(x)$ are unworkable. A feasible estimator is obtained by replacing $\underline{\mathbf{Y}}_t$ by its approximation based on estimates of $\{\Pi_k\}_{k=1}^{Q}$. The proposed estimation procedure extends Xiao et al. (2003).

1. For $j = 1, \ldots, d$: obtain a preliminary consistent estimate of $\mathbf{m}$ by the regression of $\mathbf{Y}_t$ on X_t with corresponding kernel K_0 and bandwidth h_0 assuming i.i.d. errors. Denote the preliminary estimate as $\widehat{\mathbf{m}}(X_t)$ and calculate the estimated residuals $\widehat{\mathbf{u}}_t = \mathbf{Y}_t - \widehat{\mathbf{m}}(X_t)$, $\widehat{\mathbf{u}}_t = (u_{t,1}, \ldots, u_{t,d})'$, $\widehat{u}_{t,j} = Y_{t,j} - \widehat{m}_j(X_t)$.
2. Conduct an estimation of the $VAR(Q)$ matrix coefficients in the autoregression of $\widehat{\mathbf{u}}_t$: $\widehat{\mathbf{u}}_t = \widehat{\Pi}_1 \widehat{\mathbf{u}}_{t-1} + \cdots + \widehat{\Pi}_Q \widehat{\mathbf{u}}_{t-Q} + \mathbf{e}_t$, where $\mathbf{e}_t = (e_{t,1}, \ldots, e_{t,d})'$ is a vector of i.i.d. noise.
3. Construct an approximation of $\underline{\mathbf{Y}}_t$, $t = 2, \ldots, T$ that is $\widehat{\underline{\mathbf{Y}}}_t = \mathbf{Y}_t - \widehat{\Pi}_1 \left(\mathbf{Y}_{t-1} - \widehat{\mathbf{m}}(X_{t-1})\right) - \cdots - \widehat{\Pi}_Q \left(\mathbf{Y}_{t-Q} - \widehat{\mathbf{m}}(X_{t-Q})\right)$. The proposed estimator of $\mathbf{m}(x)$ is then obtained from the regression of $\widehat{\underline{\mathbf{Y}}}_t$ on X_t with corresponding kernel K_1 and bandwidth h_1, resulting in the estimator $\widetilde{\mathbf{m}}(x) = (\widetilde{m}_1(x), \ldots, \widetilde{m}_d(x))'$:

$$\widetilde{m}_j(x) = \frac{\frac{1}{T\mathbb{E}\left[K_1\left(\frac{d(x,X_1)}{h_1}\right)\right]} \sum_{t=2}^{T} \widehat{\underline{Y}}_{t,j} K_1\left(\frac{d(x,X_1)}{h_1}\right)}{\frac{1}{T\mathbb{E}\left[K_1\left(\frac{d(x,X_1)}{h_1}\right)\right]} \sum_{s=2}^{T} K_1\left(\frac{d(x,X_s)}{h_1}\right)}$$

For simplicity, let $Q = 1$ in the following. The proofs and results remain similar in the general case $Q > 1$. Note that one can iterate this process, in case the initial estimate of the autocorrelation

is not accurate enough as the bias in this estimate will propagate to the filtered series and hence to the estimation of $\mathbf{m}(x)$. In our numerical studies, we present both the initial estimate and the estimate based on an additional iteration of the steps above.

Let us now explain in details the theoretical set-up that enables us to prove the asymptotic results in our paper. We first assume that the error process $\{\mathbf{u}_t\}$ is independent of the process $\{X_t\}$ and that $\mathbb{E}[\mathbf{e}_t|X_t] = 0$. We consider that the processes $\{X_t, \mathbf{Y}_t\}$ are α-mixing, the most general case of weakly dependent variables. Let $\mathcal{F}_a^b$ be the σ-algebra of events generated by the random variables $\{X_t,\ \mathbf{Y}_t\}_{t=a}^b$ and set (Rosenblatt (1956))

$$\sup_{A\in\mathcal{F}_{-\infty}^0, B\in\mathcal{F}_k^\infty} |\mathbb{P}(A\cap B) - \mathbb{P}(A)\mathbb{P}(B)| = \alpha(k) \underset{k\to\infty}{\longrightarrow} 0.$$

Let $|\cdot|$ denote the L_1-norm when it is applied to a vector; $|\mathbf{y}| = \sum_{j=1}^d |y_j|$, $\mathbf{y} = (y_1, \ldots, y_d)'$ and the usual matrix norm when applied to a matrix.

Our assumptions are listed below:

H1 *(Smoothness)*

(1) $m_j(\cdot)$ is a bounded Lipschitz function: $|m_j(u) - m_j(v)| \le c_3 d(u,v)^\beta$ for all $u, v \in (\mathcal{C}, d)$ for some $\beta > 0$.

(2) Let $\mathbf{G}_2(u) = Var[\mathbf{Y}_t|X_t = u]$, $u \in (\mathcal{C}, d)$, the variance matrix of $\mathbf{Y}_t$ given $X_t = u$. $\mathbf{G}_2(u)$ is independent of t and is continuous in some neighborhood of x

$$\sup_{\{u:d(x,u)\le h\}} |\mathbf{G}_2(u) - \mathbf{G}_2(x)| = o(1) \quad \text{as } h \to 0$$

Assume $\mathbb{E}|\mathbf{Y}_t|^\nu < \infty$ and $\mathbb{E}[|\mathbf{e}_t|^\nu] < \infty$, for some $\nu > 2$. Assume

$$\mathbf{G}_\nu(u) = \mathbb{E}[|\mathbf{Y}_t - \mathbf{m}(x)|^\nu | X_t = u], u \in (\mathcal{C}, d)$$

is continuous in some neighborhood of x.

(3) Define

$$\mathbf{G}(u, v; x) = \mathbb{E}[(\mathbf{Y}_t - \mathbf{m}(x))(\mathbf{Y}_s - \mathbf{m}(x))|X_t = u, X_s = v], \quad t \neq s \text{ and } u, v \in (\mathcal{C}, d)$$

Assume that $\mathbf{G}(u, v; x)$ does not depend on t, s and is continuous in some neighborhood of (x, x).

H2 *(Kernel)* The kernels K_i, $i = 0$ or 1, are symmetric nonnegative bounded kernels with compact support $[0, 1]$ satisfying

(1) $\int K_i(u)du = 1$ and $c_{1,i}\mathbf{1}_{[0,1]} < K_i < c_{2,i}\mathbf{1}_{[0,1]}$, $c_{1,i}$ and $c_{2,i}$ are two finite constants.

(2) For $j = 1, 2$, we have $I_j(h) \to C_j$ as $h \to 0$, for some positive constant C_j, with

$$I_j(h) = \frac{1}{\phi(h)/h} \int_0^1 K_i^j(u)\phi'(hv)dv \qquad \text{where } \phi(\cdot) \text{ is defined below.}$$

Let $\mathcal{B}(x,h)$ be a ball centered at $x \in (\mathcal{C}, d)$ with radius h and let f_k, $k = 1, 2$ and 3, be finite nonnegative functionals. Finally, we introduce the following notations, where $F_x^t(h)$ corresponds to the well-known notion of small ball probabilities (see e.g. Dabo-Niang (2004), Ferraty and Vieu (2006)):

$$\begin{aligned}
F_x^t(h) &= \mathbb{P}[X_t \in \mathcal{B}(x,h)] &&:= \mathbb{P}[d(X_t,x) \le h] \\
F_{x,x}^{s,t}(h) &= \mathbb{P}[(X_t,X_s) \in \mathcal{B}(x,h) \times \mathcal{B}(x,h)] &&:= \mathbb{P}[d(X_t,x) \le h, d(X_s,x) \le h] \\
F_{x,y}^{s,t}(h) &= \mathbb{P}[(X_t,X_s) \in \mathcal{B}(x,h) \times \mathcal{B}(y,h)] &&:= \mathbb{P}[d(X_t,x) \le h, d(X_s,y) \le h]
\end{aligned}$$

H3 *(Distributions)*

(1) $F_x^t(h) = \phi(h) f_1(x)$ as $h \to 0$, where $\phi(0) = 0$ and $\phi(h)$ is absolutely continuous in a neighborhood of the origin and $f_1(X_t)$ is uniformly bounded and bounded away from zero.

(2) $\sup_{t \neq s} F_{x,x}^{s,t}(h) \le \psi_1(h) f_2(x)$ as $h \to 0$, where $\psi_1(h) \to 0$ as $h \to 0$ and $f_2(X_t) < \infty$ is uniformly bounded and bounded away from zero.
Assume that the ratio $\psi_1(h)/\phi^2(h)$ is bounded. It is also supposed that $\exists \zeta_1 \in (0,1)$, $0 < F_{x,x}(h) = O(\phi(h)^{1+\zeta_1})$.

(3) $\sup_{t \neq s} F_{x,y}^{s,t}(h) \le \psi_2(h) f_3(x,y)$ as $h \to 0$, where $\psi_2(h) \to 0$ as $h \to 0$ and $f_3(X_t, X_s) < \infty$ is uniformly bounded and bounded away from zero.
Assume that the ratio $\psi_2(h)/\phi^2(h)$ is bounded.

H4 *(Mixing)* $\displaystyle\sum_{l=1}^{\infty} l^{\delta} [\alpha(l)]^{1-2/\nu} < \infty$
for some $\nu > 2$ and $\delta > 1 - 2/\nu$. Note that ν is the order of the moment in **H1**(2).

H5 Let $h_i \to 0$, $h_0/h_1 \to 0$ and $\frac{\log T}{T^{1/2}\phi(h_0)} \to 0$ as $T \to \infty$. Let $\{v_T\}$ be a sequence of positive integers satisfying $v_T \to \infty$ such that $v_T = o((T\phi(h_0))^{1/2})$ and $(T/\phi(h_0))^{1/2}\alpha(v_T) \to 0$, $Th_0^{2\beta} \to 0$ as $T \to \infty$.

Remark 1.

- *Hypothesis **H1**(1) is a mild smoothness assumption for kernel functions in nonparametric estimation whereas hypotheses **H1**(2) and **H1**(3) are continuity assumptions on certain second-order moments.*

- *Hypothesis **H2**(1) on the kernel is standard. From hypothesis **H2**(2), if the kernel K_i satisfies $0 < c_1 \leq K_i(t) \leq c_2 < \infty$, then $c_1 \leq I_j(h) \leq c_2$. In fact, this assumption yields an expression of the asymptotic variance (rather than upper and lower bounds).*

- *Hypotheses of type **H3** were proposed in Masry (2005) and have been motivated by the work of Gasser et al. (1998). These hypotheses are linked to the volume of an n-ball. When $X \in \mathbb{R}^d$, $f_1(x)$ refers to the probability density of the random variable X and $\phi(h)$ is the volume of the unit ball in $\mathbb{R}^d$. Assumptions **H3**(2) and **H3**(3) concern the behavior of joint distribution.*

- *Hypothesis **H4** is a standard assumption on the decay of the strongly mixing coefficient $\alpha(l)$ and hypothesis **H5** concerns the rate of the decay of the mixing coefficient.*

Let $\Delta_t^{(i)}(x) = K_i\left(\frac{d(x, X_t)}{h_i}\right)$, $\mathbf{Z}_t^{(i)}(x) = [\underline{\mathbf{Y}}_t - \widetilde{\mathbf{m}}(x)]\Delta_t^{(i)}(x) - \mathbb{E}\left[(\underline{\mathbf{Y}}_t - \widetilde{\mathbf{m}}(x))\Delta_t^{(i)}(x)\right]$, for $i = 0, 1$ (see below). In the following, $\xrightarrow{d}$ denotes the convergence in distribution. The following theorem gives the asymptotic normality of the estimator $\widetilde{\mathbf{m}}(x)$ based on the transformation of the dependent variable.

Theorem 1. *Under assumptions **H1**-**H5**, we have*

$$(T\phi(h_1))^{1/2}[\widetilde{\mathbf{m}}(x) - \mathbf{m}(x) - \mathbf{B}_T(x)] \xrightarrow{d} \mathcal{N}_d(0, \Sigma_x)$$

with $\mathbf{B}_T(x) = \mathbb{E}[\widetilde{\mathbf{m}}(x)] - \mathbf{m}(x)$, $\Sigma_x = \frac{C_2\mathbf{G}_2(x)}{C_1^2 f_1(x)} = \lim_{T\to\infty} \frac{\phi(h_1)Var(\mathbf{Z}_T^{(1)}(x))}{\mathbb{E}^2(\Delta_1^{(1)}(x))}$ is the $(d \times d)$ asymptotic covariance matrix, $x \in (\mathcal{C}, d)$ whenever $f_1(x) > 0$.

The following theorem gives a consistency result of the estimator $\widetilde{\mathbf{m}}(x)$.

Theorem 2. *Under assumptions **H1**-**H5**,*

$$\lim_{T\to\infty} \widetilde{\mathbf{m}}(x) = \mathbf{m}(x) \quad \textit{in probability.}$$

Remark 2. *One can establish a convergence in probability of $\widetilde{\mathbf{m}}(x)$ with rate (for instance assuming for simplicity the boundedness of the response, even though the bound can be arbitrarily large) and state that:*

$$|\widetilde{\mathbf{m}}(x) - \mathbf{m}(x)| = O_p(h_0^{\beta}) + o_p\left(\sqrt{\frac{1}{T\phi(h_1)}}\right)$$

under conditions of Theorem 2.

The proofs of these theorems are postponed in the Appendix section.

3. Numerical results

3.1. Simulation study

We investigate the proposed estimator on simulated data, considering first the univariate case ($d = 1$). The functional observations X_t (with $t = 1, \ldots, T$) are defined by $X_t(w) = 1 + 10e_{0,t} +$

$3e_{1,t}w^2 + 4e_{2,t}(1-w)^3$, $w \in [0,1]$ where $e_{0,t}$, $e_{1,t}$ and $e_{2,t}$ are i.i.d. $\mathcal{N}(0,1)$. We take $m(X_t(w)) = \sqrt{|0.5\int_0^1 X_t(w)dw|}$. The error process u_t is an $AR(1)$ process, that is $u_t = \epsilon_t + \rho\epsilon_{t-1}$ where ϵ_t are i.i.d. $\mathcal{N}(0, \sigma = 1)$. Various values of ρ are considered. The number of replications is 200. We report the relative efficiency (denoted as RE) calculated as the ratio of squared errors. Table 1 describes summary statistics of the relative efficiency for $T = 200$ whereas Table 2 gives the average of the relative efficiency for different values of T. $RE1$ reports the relative efficiency of the proposed efficient estimator $\widetilde{m}(x)$ over the conventional estimator $\widehat{m}(x)$, and $RE2$ concerns the relative efficiency of the iterated estimator, over $\widehat{m}(x)$. We did not implement the efficient estimator of Xiao et al. (2003) as we only consider here for simplicity the case of one lag, but the efficient estimator could be used in our context with larger lags than one. Instead here we report results about the iterated estimates. The semi-metric $d(\cdot,\cdot)$ for computing proximities between curves X_t plays a major role and depends on the specified statistical problem and dataset. After trying some semi-metrics which can select most of the pertinent information of the curves, we choose $d(\cdot,\cdot)$ inside the family of principal component semi-metrics (see Ferraty and Vieu (2006)) which is defined by $d_q^{PCA}(X_t, X) = \sqrt{\sum_{k=1}^{q}\left(\int[X_t(w) - X(w)]v_k(w)dw\right)^2}$ where $v_1, v_2, \ldots$ are the orthonormal eigenfunctions of the covariance operator and q is a tuning parameter. We have considered different values for this parameter q and the value of 1 is better suited to this simulation study. This number of principal components allows to explain around 98% of variation of the curves. This choice for the semi-metric based on principal components is well adapted to the considered polynomial functions X we deal with and for which it is important to take into account large variations of the data. Regarding the implementation of the estimators, we use the quadratic kernel (Epanechnikov) (defined by $K(x) = \frac{3}{4}\left(1-x^2\right)\mathbf{1}_{[-1;+1]}(x)$). Another choice to make is the bandwidth parameters. It is well known that the performance of the kernel estimate depends on the choice of the window parameter. The bound in Remark 2 allows us to choose the window parameters that minimize this bound. This choice of the bandwidths leads to be optimal in the finite dimensional case. In practice, a useful bandwidth choice method is cross-validation as follows. For instance, cross-validation ideas are encountered in the finite dimensional setting (e.g. Hardle and Marron (1985), Hart and Vieu (1990)) as well as on the infinite dimensional one (e.g. Rachdi and Vieu (2007), Benhenni et al. (2007)).

1. We consider the preliminary estimate $\widehat{m}$ of m by the regression of Y_t on X_t with quadratic kernel K, the semi-metric d_1^{PCA} and data driven bandwidth h_0^{opt} assuming i.i.d. errors (see Ferraty and Vieu (2006) for more details):

$$h_0^{opt} = arg\min_h \sum_{t=1}^{T}(Y_t - \widehat{m}_{-t}(X_t))^2$$

where

$$\widehat{m}_{-t}(x) = \frac{\sum_{u=1,u\neq t}^{T} Y_u K\left(\frac{d(x,X_u)}{h_0}\right)}{\sum_{s=1,s\neq t}^{T} K\left(\frac{d(x,X_s)}{h_0}\right)}$$

We calculate the estimated residuals $\widehat{u}_t = Y_t - \widehat{m}(X_t)$.

2. We conduct an estimation of the AR(1) coefficients in the autoregression of $\widehat{u}_t$: $\widehat{u}_t = \widehat{a}_1 \widehat{u}_{t-1} + \eta$, as in Section 2. We construct $\widehat{\underline{Y}}_t = Y_t - \widehat{a}_1 \left(Y_{t-1} - \widehat{m}(X_{t-1})\right)$, $t = 2, \ldots, T$ and the estimate $\widetilde{m}(x)$ from the regression of $\widehat{\underline{Y}}_t$ on X_t with quadratic kernel K, the semi-metric d_1^{PCA} and optimal data driven bandwidth h_1^{opt} in the same way as above, replacing $\widehat{m}(x)$ by $\widetilde{m}(x)$ resulting in:

$$h_1^{opt} = arg \min_h \sum_{t=2}^{T} (Y_t - \widetilde{m}_{-t}(X_t))^2.$$

The results in Table 1 show that there is great variability in the improvements across replications. The inter-quartile ranges of the relative efficiencies are nevertheless tight: typically within $0.1 - 0.2$, except when the improvements are large (e.g. for $\rho = 0.9$). The mean improvements for the estimator is always smaller than 1 (i.e. showing a positive impact of our method) except when $\rho \leq 0.1$, a very small level of autocorrelation, where the values are 1.004 and 1.007, still very close to 1. The iterated estimator is much less efficient than the initial estimator. It seems that the additional steps are adding several layers of noise in the procedure and therefore degrade the estimation. Table 2 allows us to see the effect of sample size on the mean relative efficiency. It seems that such benefit is stronger whenever the autocorrelation is higher (as expected to be able to capture it properly). Moreover, for $\rho = 0.9$, 0.6 and 0.25, three iterations have been implemented and the relative efficiency is still worst than at the first iteration and worst than at the second iteration. The results concerning these three iterations are given in Table 3.

Table 1: Elementary statistics of the relative efficiency for 2 iterations with $T = 200$

ρ	RE	Min	Q_1	Med	$Mean$	Q_3	Max
0.99	1	0.110	0.917	0.978	0.904	0.998	1.128
	2	0.116	0.928	0.982	0.941	1.033	1.564
0.95	1	0.068	0.702	0.883	0.813	0.966	1.111
	2	0.132	0.781	0.948	0.915	1.061	1.798
0.90	1	0.116	0.551	0.776	**0.732**	0.925	1.596
	2	0.120	0.737	0.897	0.925	1.093	2.766
0.80	1	0.251	0.669	0.782	0.797	0.941	1.568
	2	0.303	0.814	0.989	1.037	1.219	2.950
0.60	1	0.242	0.734	0.871	0.884	1.010	1.929
	2	0.271	0.921	1.080	1.144	1.277	3.270
0.50	1	0.393	0.804	0.909	0.921	1.049	1.463
	2	0.396	0.967	1.092	1.127	1.243	2.209
0.25	1	0.666	0.943	1.002	0.997	1.044	1.620
	2	0.580	0.971	1.037	1.055	1.113	1.777
0.10	1	0.822	0.977	1.001	1.007	1.028	1.436
	2	0.791	0.981	1.010	1.029	1.063	1.459
0.00	1	0.469	0.982	0.998	1.004	1.025	1.421
	2	0.466	0.972	1.004	1.008	1.033	1.555

Table 2: Mean of the relative efficiency for $T = 100$, 200 and 500 considering 2 iterations

T	100		200		500	
ρ	$RE1$	$RE2$	$RE1$	$RE2$	$RE1$	$RE2$
0.99	0.921	0.969	0.904	0.941	0.853	0.867
0.95	0.845	0.950	0.813	0.915	0.768	0.971
0.90	0.825	1.015	0.732	0.925	0.733	0.897
0.80	0.851	1.080	0.797	1.037	0.763	1.009
0.60	0.908	1.410	0.884	1.144	0.872	1.115
0.50	0.969	1.204	0.921	1.127	0.925	1.138
0.25	0.997	1.056	0.997	1.055	0.998	1.071
0.10	1.013	1.034	1.007	1.029	1.001	1.021
0.00	1.024	1.030	1.004	1.008	1.009	1.013

Table 3: Elementary statistics of the relative efficiency for 3 iterations with $T = 200$

ρ	RE	Min	Q_1	Med	$Mean$	Q_3	Max
0.90	1	0.116	0.551	0.777	0.731	0.924	1.590
	2	0.119	0.736	0.897	0.925	1.095	2.793
	3	0.151	0.755	0.967	1.005	1.212	2.708
0.60	1	0.241	0.736	0.873	0.884	1.010	1.929
	2	0.271	0.921	1.079	1.145	1.277	3.264
	3	0.536	0.997	1.182	1.233	1.390	4.019
0.25	1	0.666	0.943	1.001	0.996	1.042	1.619
	2	0.581	0.971	1.037	1.054	1.114	1.767
	3	0.591	0.980	1.049	1.074	1.135	1.697

The efficiencies for functional data seem better than for univariate time series (Xiao et al. (2003)), although Xiao et al. (2003) considered an ARMA(1,1) case - and an AR(2) pre-whitening - in their simulations that is more challenging (but in dimension one, not in infinite dimension as here). Indeed in Xiao et al. (2003), the relative improvement was never below 0.85. Here, we can reach average reductions below 0.75 for high correlation and long enough time series to capture this high level of correlation accurately. According to Ferraty and Vieu (2006), the curse of dimensionality, a well-known concept in nonparametric inference, does not affect functional data with high correlation. This, combined with an appropriate choice of the semi-metric, can explain the fact that the efficiencies seem better in functional context than univariate one. One illustration is given in Figure 1: for one replication, considering $T = 200$ and a value of $\rho = 0.9$. The black curve displays the true function $m(X_t)$, the blue curve corresponds to the standard kernel estimation whereas the red and green curves correspond to the proposed estimator with one or two iterations respectively. Note that in this case the common estimate of the curve is far from the true curve. On the contrary, the curves obtained considering our methodology not only have the same shape as the true curve but are very close to the truth. In this case, the information of the autocorrelation function of the error process clearly improves the quality of the regression estimation. However, when the autoregressive parameter is small, as expected, our methodology does not improve the results obtained through the standard kernel procedure that does not account for correlation in the errors. For example, Figure 2 shows the curves obtained considering $\rho = 0.25$ for one replication. We cannot see large differences between the displayed curves. The three estimation curves are close to the curve representing the true function.There is asymmetry in our regression fit, as we can notice for instance that extremely low values are not well captured overall but the largest values

of the response are, and much more so by our method, especially in the case of high autocorrelation as shown in Figure 2 compare to Figure 1.

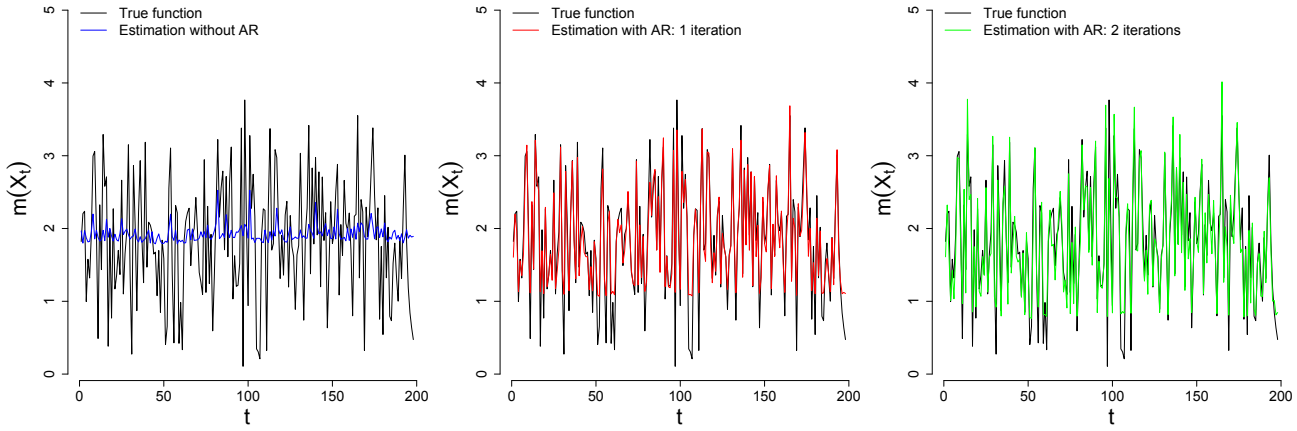

Figure 1: Estimates for $T = 200$ and $\rho = 0.9$. Black curve: true function $m(X_t)$, blue curve: standard kernel estimation, red and green curves: proposed estimator with one or two iterations respectively.

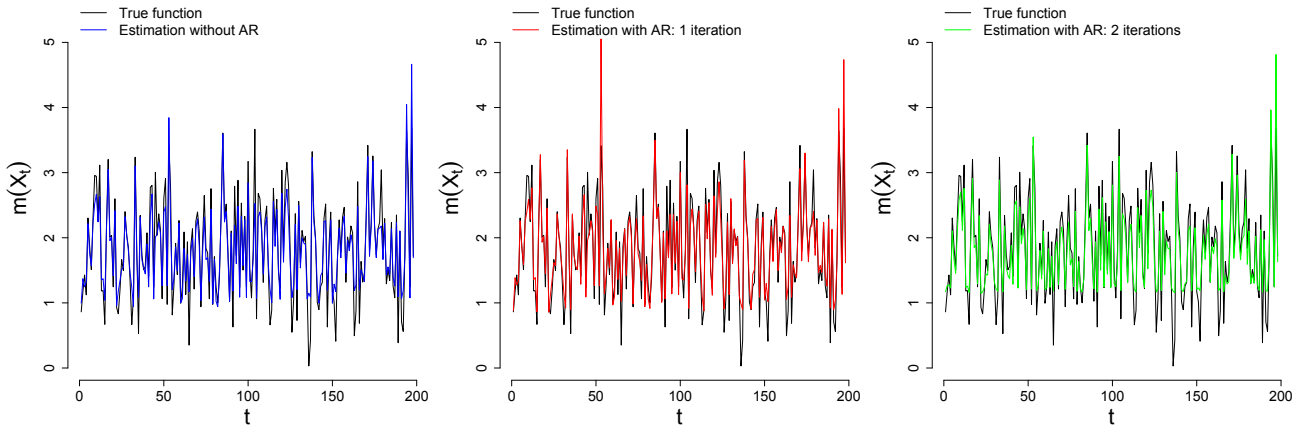

Figure 2: Estimates for $T = 200$ and $\rho = 0.25$. Black curve: true function $m(X_t)$, blue curve: standard kernel estimation, red and green curves: proposed estimator with one or two iterations respectively.

Some results are also obtained in environments with various levels of noise. Different values of the parameter σ from $\epsilon_t \sim \mathcal{N}(0, \sigma)$ have been tested. More precisely, different values of the signal-to-noise ratio (snr) have been considered where snr is defined as $Var(m(X_t))/\sigma^2$. The applied values are $snr = 1$, 0.75, 0.5, 0.25 and 0.05. The obtained results are displayed in Table 4. We notice that when σ is higher, the relative efficiency is worse but for high autocorrelation our method still provides some improvements. For example, considering $\rho = 0.9$, the averaged $RE1$ is 0.722 (resp. 0.748) with $\sigma = 0.77$ (resp. $\sigma = 3.42$).

We now present a challenging simulation in order to explore the robustness of our method in a

Table 4: Elementary statistics of the relative efficiency for $T = 200$ for different levels of signal-to-noise ratio considering 2 iterations

snr	RE	ρ	σ	Min	Q_1	Med	$Mean$	Q_3	Max	ρ	σ	Min	Q_1	Med	$Mean$	Q_3	Max
1.00	1	0.99	0.76	0.124	0.899	0.976	0.896	0.998	1.156	0.50	0.75	0.408	0.832	0.926	0.931	1.045	1.357
	2			0.156	0.921	0.981	0.938	1.032	1.608			0.494	0.990	1.122	1.175	1.310	3.215
0.75	1		0.87	0.116	0.905	0.977	0.899	0.998	1.156		0.87	0.398	0.821	0.921	0.929	1.046	1.499
	2			0.130	0.923	0.981	0.939	1.032	1.608			0.428	0.987	1.115	1.158	1.272	3.215
0.50	1		1.07	0.105	0.914	0.979	0.902	0.998	1.156		1.06	0.398	0.820	0.920	0.925	1.046	1.499
	2			0.105	0.926	0.983	0.940	1.032	1.608			0.362	0.975	1.108	1.144	1.259	3.215
0.25	1		1.51	0.091	0.921	0.981	0.907	0.100	1.156		1.50	0.329	0.813	0.920	0.925	1.047	1.707
	2			0.080	0.930	0.984	0.943	1.033	1.608			0.297	0.957	1.095	1.133	1.248	3.215
0.05	1		3.38	0.072	0.929	0.983	0.915	1.002	1.168		3.35	0.225	0.799	0.919	0.924	1.047	2.284
	2			0.053	0.930	0.985	0.949	1.034	1.826			0.297	0.938	1.081	1.119	1.244	3.504
1.00	1	0.95	0.82	0.073	0.699	0.869	0.809	0.967	1.222	0.25	0.83	0.716	0.950	1.002	0.995	1.042	1.538
	2			0.165	0.770	0.952	0.925	1.073	1.700			0.482	0.974	1.041	1.066	1.134	1.811
0.75	1		0.94	0.069	0.697	0.873	0.809	0.965	1.222		0.96	0.716	0.946	1.001	0.994	1.042	1.582
	2			0.141	0.766	0.950	0.920	1.071	1.756			0.482	0.973	1.037	1.064	1.132	1.945
0.50	1		1.15	0.056	0.704	0.873	0.811	0.967	1.222		1.18	0.716	0.947	1.001	0.995	1.041	1.651
	2			0.117	0.776	0.949	0.917	1.067	1.828			0.482	0.971	1.033	1.057	1.122	1.945
0.25	1		1.63	0.047	0.713	0.886	0.815	0.972	1.222		1.67	0.472	0.946	1.000	0.997	1.043	2.070
	2			0.093	0.769	0.949	0.915	1.069	1.828			0.482	0.968	1.033	1.052	1.121	1.945
0.05	1		3.65	0.038	0.734	0.908	0.829	0.982	1.946		3.73	0.232	0.942	0.999	1.002	1.046	2.452
	2			0.031	0.779	0.954	0.922	1.074	2.339			0.315	0.962	1.029	1.051	1.117	2.528
1.00	1	0.90	0.77	0.124	0.528	0.753	**0.722**	0.912	1.526	0.10	0.73	0.828	0.984	1.003	1.011	1.027	1.894
	2			0.146	0.768	0.927	0.966	1.129	3.603			0.787	0.985	1.011	1.036	1.067	1.927
0.75	1		0.88	0.122	0.534	0.760	**0.723**	0.913	1.571		0.84	0.756	0.981	1.003	1.009	1.027	1.894
	2			0.129	0.760	0.907	0.950	1.117	3.603			0.752	0.983	1.012	1.036	1.065	1.927
0.50	1		1.08	0.109	0.546	0.764	**0.726**	0.921	1.611		1.03	0.755	0.980	1.003	1.009	1.028	1.894
	2			0.123	0.753	0.899	0.937	1.100	3.603			0.752	0.983	1.011	1.035	1.065	1.927
0.25	1		1.53	0.091	0.550	0.772	**0.733**	0.928	1.611		1.46	0.755	0.977	1.002	1.008	1.029	1.894
	2			0.105	0.737	0.895	0.923	1.091	3.603			0.653	0.981	1.010	1.033	1.062	1.927
0.05	1		3.42	0.067	0.565	0.796	**0.748**	0.945	1.940		3.26	0.167	0.975	1.002	1.008	1.029	1.936
	2			0.105	0.737	0.898	0.920	1.094	3.603			0.599	0.979	1.010	1.032	1.063	2.135
1.00	1	0.80	0.79	0.223	0.665	0.818	0.807	0.961	1.536	0.00	0.82	0.362	0.985	0.999	1.000	1.019	1.262
	2			0.301	0.863	1.045	1.109	1.299	3.989			0.356	0.976	1.004	1.008	1.030	1.400
0.75	1		0.92	0.223	0.663	0.802	0.802	0.952	1.560		0.94	0.362	0.985	0.998	1.000	1.020	1.389
	2			0.292	0.846	1.025	1.079	1.267	3.989			0.356	0.973	1.003	1.008	1.031	1.523
0.50	1		1.12	0.223	0.663	0.797	0.801	0.948	1.566		1.16	0.362	0.983	0.998	1.000	1.019	1.389
	2			0.283	0.833	1.009	1.055	1.245	3.989			0.356	0.972	1.002	1.006	1.031	1.523
0.25	1		1.59	0.221	0.658	0.794	0.797	0.945	1.606		1.64	0.362	0.979	0.998	1.000	1.019	1.559
	2			0.283	0.802	0.990	1.034	1.219	3.989			0.356	0.972	1.002	1.006	1.032	1.523
0.05	1		3.55	0.189	0.662	0.813	0.806	0.958	1.828		3.66	0.362	0.978	0.998	1.003	1.022	2.777
	2			0.183	0.796	0.981	1.028	1.196	3.989			0.356	0.972	1.002	1.006	1.035	1.648
1.00	1	0.60	0.76	0.270	0.760	0.874	0.893	1.013	2.236								
	2			0.352	0.981	1.104	1.197	1.306	3.436								
0.75	1		0.88	0.250	0.757	0.876	0.889	1.011	2.236								
	2			0.296	0.959	1.100	1.177	1.296	3.436								
0.50	1		1.08	0.250	0.753	0.879	0.892	1.013	2.236								
	2			0.296	0.941	1.091	1.164	1.285	3.436								
0.25	1		1.53	0.250	0.743	0.875	0.892	1.015	2.236								
	2			0.296	0.926	1.074	1.147	1.265	3.571								
0.05	1		3.41	0.246	0.741	0.875	0.898	1.019	3.087								
	2			0.276	0.906	1.062	1.137	1.256	5.764								

more complex setting. The set-up is multivariate with bidimensional response (Y_1, Y_2). First, the functions X_t are still stationary (to satisfy the assumptions) but are now considered dependent. Their dependence is introduced as an extension of the simple simulations with similar functional observations as before: the X_t (with $t = 1, \ldots, T$) are defined by $X_t(w) = 1 + 10e_{0,t} + 3e_{1,t}w^2 + 4e_{2,t}(1-w)^3$, $w \in [0,1]$ where $e_{0,t}$ and $e_{1,t}$ are i.i.d. $\mathcal{N}(0,1)$, but now $e_{2,t}$ is an autoregressive process $e_{2,t} = 0.9e_{2,t} + \epsilon_{t-1}^{(2)}$ where $\epsilon_t^{(2)}$ are i.i.d. $\mathcal{N}(0,1)$. The two response functions are also more complex, with a sinusoid integrated against the functional explanatory variable to provide further

nonlinear effects. We now take

$$m_1(X_t(w)) = m_2(X_t(w)) = 2\sqrt{\left|2\int_0^1 X_t(w)sin(\pi w)dw\right|}.$$

The error process $\mathbf{u}_t$ is a $VAR(1)$ process, that is $\mathbf{u}_t$: $\mathbf{u}_t = \Pi_1\mathbf{u}_{t-1} + \mathbf{e}_t$, where $\mathbf{e}_t$ are i.i.d. $\mathcal{N}(0, \Sigma)$:

$$\Pi_1 = \begin{pmatrix} a & -0.3 \\ 0.1 & a/2 \end{pmatrix} \quad \Sigma = \begin{pmatrix} b & 0.8 \\ 0.8 & b/4 \end{pmatrix}$$

Various values of a and b are considered. The number of replications is 100. Table 5 presents the relative efficiencies with respect to a and b. The results show that the relative improvement can be as large as in the univariate case for more complex functions (and now including dependence in the explanatory variables). The case of a low autocorrelation level (0.25) demonstrates that the method still provides good results, equivalent to not taking into account the autocorrelation. When the autocorrelation is larger (e.g. 0.8-0.95), the results are typically best, as in the univariate case. For very high autocorrelation levels (e.g. 0.99), the relative efficiencies deteriorate slightly, due to the length of the series, as in the univariate case. The noise level seems to have a positive effect from 2 till 16 and then can possibly damage the efficiencies, especially for m_1. Indeed, the noise needs to be large enough for the autocorrelation structure to be identified, and thus benefit the estimation, but not too high when it starts harming the estimation procedure. Nevertheless, even with a noise level 2-3 times higher than the optimal one, the efficiencies are relatively close to the optimal ones. Overall, the range of noise levels is large (from 1 to 24 times more), and the method shows resilience across the range.

3.2. Real data application

Here, we illustrate our methodology for the ozone concentration forecasting problem and compare our predictions with the ones obtained using the classical kernel regression model for functional data. The goal is to forecast ground-level ozone concentrations using observations from monitoring stations within the south-eastern US region, over a span of 3 months in the summer of 2005. These forecasts may contribute to better public health: for example, hourly forecasts made one day ahead of this harmful pollutant allow people avoid outdoor activities likely to damage their health.

We are given the ozone concentration for different stations for every hour from June 2 to August 31, 2005 (that is 91 days). Since some of the stations had missing values, we use linear interpolation to estimate the missing values. We are interested in 1-day ahead ozone forecasting (specifically, r-hours ahead ozone forecasting, for $r = 1, \ldots, 24$ that is, from 12am to 11pm). We have implemented the univariate and multivariate versions of our procedure, e.g. modeling the error term with an $AR(1)$ and with a $VAR(1)$. We denote the ozone concentration at time t by $Z(t)$ where t refers to the day and the hour of observation. We suppose that $Z(t)$ is observed for $t \in [1; 2160)$ (24

Table 5: Mean of relative efficiencies, bivariate model, 100 replications

m_1					
a \ b	2	4	16	32	48
0.25	1.00	0.99	1.00	0.99	0.97
0.80	0.90	0.84	0.77	0.83	0.76
0.90	0.83	0.79	0.74	0.78	0.79
0.95	0.83	0.76	0.71	0.77	0.75
0.99	0.82	0.78	0.77	0.79	0.77

m_2					
a \ b	2	4	16	32	48
0.25	0.99	0.98	0.93	0.88	0.89
0.80	0.94	0.89	0.79	0.75	0.73
0.90	0.91	0.84	0.73	0.73	0.68
0.95	0.88	0.81	0.69	0.69	0.72
0.99	0.86	0.82	0.74	0.75	0.74

hours $\times$ 90 days) and we are interested in predicting $Z(2160+r)$ for $r = 1, 2, \ldots, 24$. In order to apply the functional methodology, we cut the original time series into a set of daily functional data. Here, we have decided to predict future ozone concentration by using the concentration data for the whole last day (24 hours). In order to illustrate our purpose, we will not use the 91th day and we will predict it by means of the data corresponding to the 90 previous ones. Then, as presented in Ferraty and Vieu (2006), for fixed r, the data will be reorganized into a functional explanatory sample $\{X_i, i = 1, \ldots, 89\}$ which will be loaded in the following 89×24 matrix:

$Z(1)$	$Z(2)$	$\cdots$	$Z(24)$
$Z(25)$	$Z(26)$	$\cdots$	$Z(48)$
$\vdots$			
$Z(2113)$	$Z(2114)$	$\cdots$	$Z(2136)$

and (for the univariate modeling) a response real sample $\{Y_i^{(r)}, i = 1, \ldots, 89\}$, which will be loaded in the following 89-dimensional vector:

$Z(24+r)$	$Z(48+r)$	$\cdots$	$Z(2136+r)$

For a fixed horizon r, we will predict the value of $Z(2160+r)$. In the following, the predictions have been achieved for any value of $r = 1, 2, \ldots, 24$. Note that in our procedure several parameters need to be selected. For the kernel, we use the quadratic one. Cross-validation methods, expressed in terms of k-nearest neighbours, are used for (local) smoothing parameter selection, see Chapter 7 in Ferraty and Vieu (2006). Moreover, we use a semi-metric based on the first functional principal components of the data curves. For each considered situation, as in the previous simulation study, we have tested different values for the tuning parameter q. The following reported results come

from a suitable choice of q, different according to the case. Finally, we proceed in the following way:

1. Select the horizon prediction r and organize the data as explained previously;
2. Predict $Y_{90}^{(r)}$, at fixed horizon r, by the classical kernel regression approach with a local choice of the number k of neighbors:

$$\widehat{Y}_{90}^{(r)} = \widehat{r}(X_{90}) = \frac{\sum_{i=1}^{89} Y_i^{(r)} K\left(\frac{d(X_i,X_{90})}{h_{k_{opt}(X_{i_0})}}\right)}{\sum_{i=1}^{89} K\left(\frac{d(X_i,X_{90})}{h_{k_{opt}(X_{i_0})}}\right)}$$

where $i_0 = \arg\min_{i=1,\ldots,89} d(X_{90}, X_i)$ and $h_{k_{opt}(X_{i_0})}$ is the bandwidth corresponding to the optimal number of neighbors at X_{i_0} obtained by

$$k_{opt}(X_{i_0}) = \arg\min_k \left| Y_{i_0} - \frac{\sum_{i=1, i\neq i_0}^{n} Y_i K\left(\frac{d(X_i,X_{i_0})}{h_{k(X_{i_0})}}\right)}{\sum_{i=1, i\neq i_0}^{n} K\left(\frac{d(X_i,X_{i_0})}{h_{k(X_{i_0})}}\right)} \right|$$

3. At step 2., during the learning step, the 89 response variables are estimated, denoted $\widehat{Y}_t$, $t = 1, \ldots, 89$. Then, we construct the residual terms $\{\widehat{u}_t\}$, where for $t = 1, \ldots, 89$, $\widehat{u}_t = Y_t - \widehat{Y}_t$. We estimate the $AR(1)$ coefficient, a_1, in the autoregression of $\widehat{u}_t$.
4. Construct $\widehat{\underline{Y}}_t$, $t = 2, \ldots, 89$, as explained in Section 2, as an alternative to Step 2.

$$\widetilde{Y}_{90}^{(r)} = \widetilde{r}(X_{90}) = \frac{\sum_{i=2}^{89} \widehat{\underline{Y}}_i^{(r)} K\left(\frac{d(X_i,X_{90})}{h_{k_{opt}(X_{i_0})}}\right)}{\sum_{i=2}^{89} K\left(\frac{d(X_i,X_{90})}{h_{k_{opt}(X_{i_0})}}\right)}$$

We apply the previous procedure on Station 20 to predict ozone on August 31st, the 91st day. The series of observations are represented in the left panel of Figure 3. On the right panel of this figure, the daily curves are plotted in grey and the black curve represents the curve we want to forecast. The results obtained at Step 2. (by the classical kernel method) are displayed in blue whereas those of Step 4. considering an $AR(1)$ (from our procedure presented in Section 2) are displayed in red. From this figure, one can observe that our method improves upon the results obtained with the classical method, in particular for the second half of the day. In fact, the estimated coefficients in the autoregression of $\widehat{u}_i$ are higher for that part of the day, see Table 6. The figures of ACFs (autocorrelation functions) and PACFs (partial autocorrelation functions) show that there is insignificant autocorrelation in the errors, except from 11am ($r = 12$) to 16pm ($r = 17$) where an $AR(1)$ would be the best fit. Hence, using our approach for these hours is justified. We even use our approach for all the hours of the day since there is only a tiny loss by doing so when there is very little autocorrelation. We notice in Figure 3 that the forecasts using our method are

much closer to observations over the afternoon (11am till 16am) compared to the i.i.d. regression, when the benefit of our approach is indeed maximized. Figure 4 displays autocorrelation and partial autocorrelation functions for some horizons r ($r = 6$, 12, 17 and 21). The bold character in Table 6 is used to emphasize horizons for which the corresponding autocorrelation function justifies an $AR(1)$. In addition, we compute the mean squared errors (MSE) to compare the results obtained by the different methods. For Station 20, the MSE from the classical approach is 159.91 whereas with our approach it is reduced to 126.3. The corresponding relative efficiency is 0.79. Again we note that the fact of taking into account the autocorrelation in the error process allows to improves ozone forecasting.

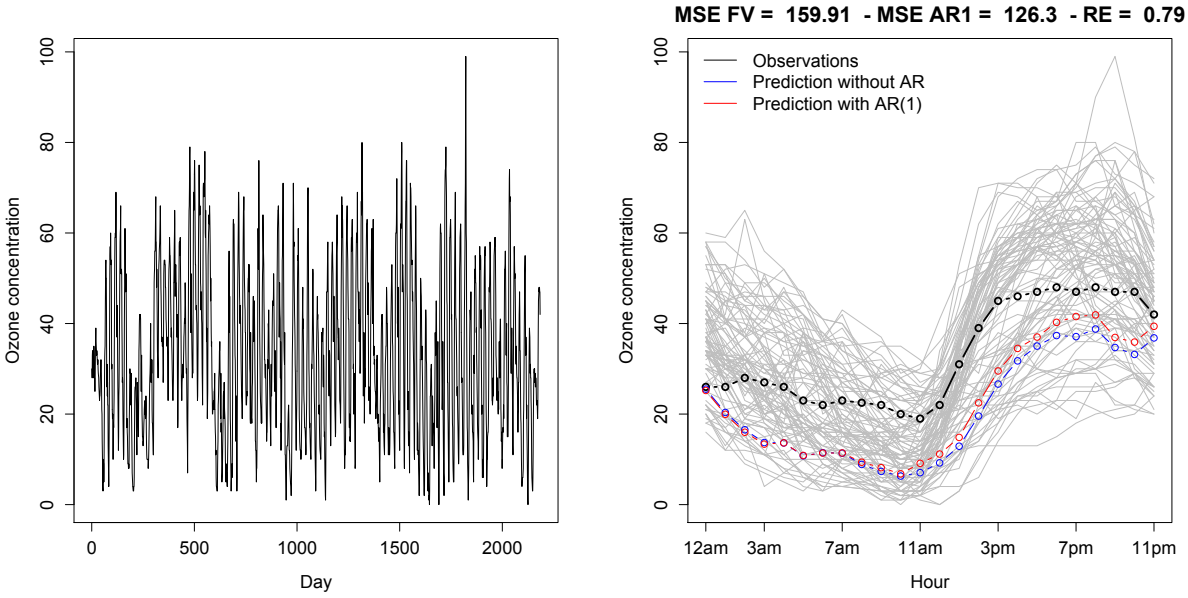

Figure 3: Ozone concentrations and predictions considering an $AR(1)$ modeling of the error term at Station 20 Left: all the series. Right: predictions.

Table 6: Station 20: for horizon predictions r, estimated autoregressive coefficients $\widehat{a}_1$

r	1	2	3	4	5	6	7	8	9	10	11	12
$\widehat{a}_1$	−0.06	−0.13	−0.13	−0.14	0.01	0.02	−0.02	0.01	0.12	0.18	0.12	**0.30**
r	13	14	15	16	17	18	19	20	21	22	23	24
$\widehat{a}_1$	**0.33**	**0.22**	**0.30**	**0.27**	**0.25**	0.19	0.18	0.16	0.08	0.12	0.16	0.17

Now, we present results obtained considering the multivariate extension. Instead of fixing the horizon r and repeating the procedure for all the horizons to study, we predict all the period in one implementation. We construct the 89 vectors $\mathbf{Y}_t$ in the following way. For $t = 1, \ldots, 89$,

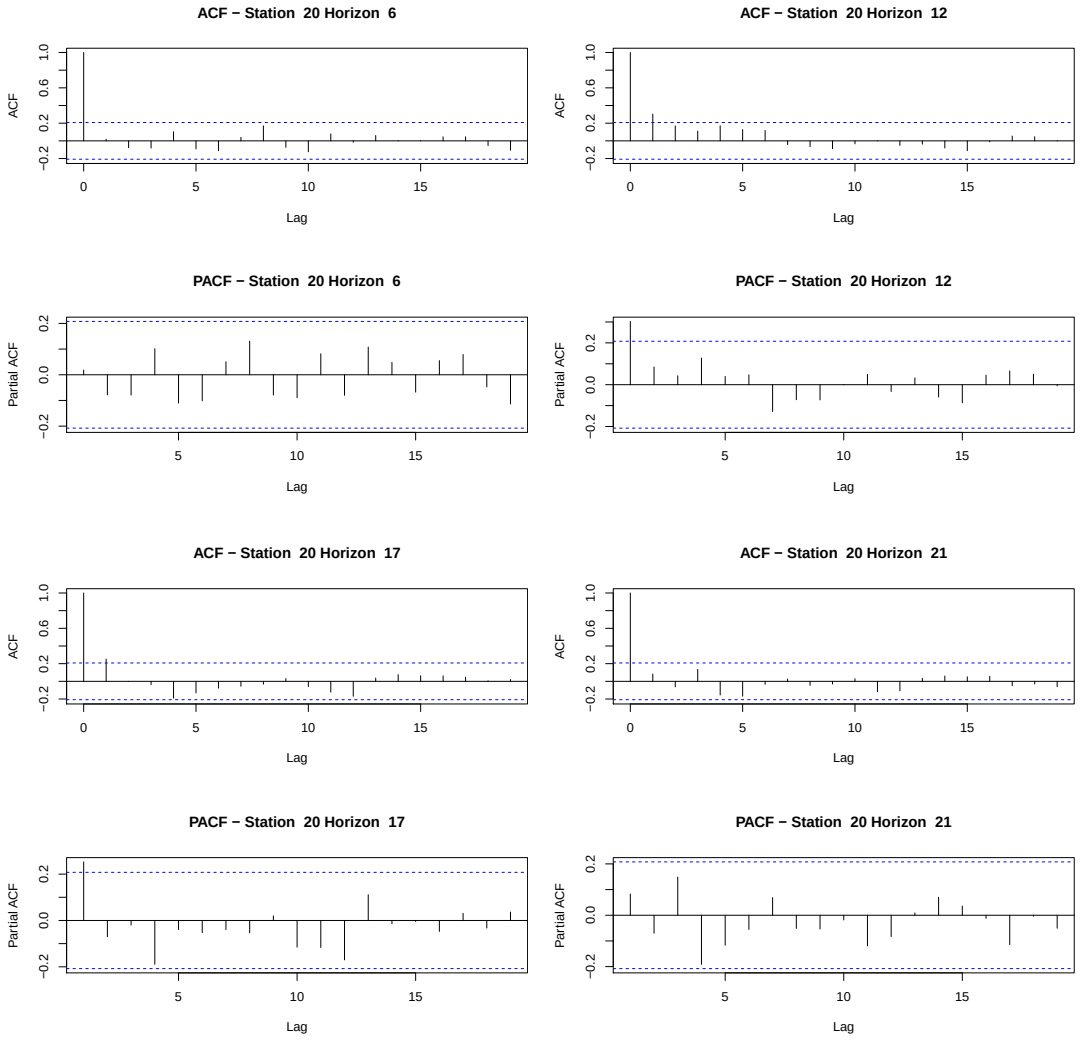

Figure 4: Autocorrelation and partial autocorrelation functions (acf and pacf) at Station 20, for $r = 6$, 12, 17, 21.

$$\begin{aligned}\mathbf{Y}_t &= (Y_{t,1}, \ldots, Y_{t,24})' \\ &= \Big(Z(24 \times t + 1) \quad Z(24 \times t + 2) \quad \cdots \quad Z(24 \times t + 24)\Big)'\end{aligned}$$

The estimation procedure is similar to the univariate case, the main change appears in the error modeling. In this illustration, we estimate the $VAR(1)$ coefficients of $\widehat{\mathbf{u}}_t$. We proceed in the following way:

1. For $j = 1, \ldots, 24$, predict $Y_{90,j}$ by the classical kernel regression approach with a local choice

of the number k of neighbors:

$$\widehat{Y}_{90,j} = \widehat{r}(X_{90}) = \frac{\sum_{i=1}^{89} Y_{i,j} K\left(\frac{d(X_i,X_{90})}{h_{k_{opt}(X_{i_0})}}\right)}{\sum_{i=1}^{89} K\left(\frac{d(X_i,X_{90})}{h_{k_{opt}(X_{i_0})}}\right)}$$

where $i_0 = \arg\min_{i=1,\ldots,89} d(X_{90}, X_i)$ and $h_{k_{opt}(X_{i_0})}$ is the bandwidth corresponding to the optimal number of neighbors at X_{i_0}.

2. At step 1., during the learning step, the 89 vectors are estimated, denoted $\widehat{\mathbf{Y}}_t$, $t = 1, \ldots, 89$. Then, we construct the residual terms $\{\widehat{\mathbf{u}}_t\}$, where for $t = 1, \ldots, 89$, $\widehat{\mathbf{u}}_i = \mathbf{Y}_t - \widehat{\mathbf{Y}}_t$. We estimate the $VAR(1)$ coefficients, Π_1, in the autoregression of $\widehat{\mathbf{u}}_t$.
3. Construct $\underline{\widehat{\mathbf{Y}}}_t$, $t = 2, \ldots, 89$, as explained in Section 2, as an alternative to Step 1, we have

$$\widetilde{Y}_{90,j} = \widetilde{r}(X_{90}) = \frac{\sum_{i=2}^{89} \underline{\widehat{Y}}_{i,j} K\left(\frac{d(X_i,X_{90})}{h_{k_{opt}(X_{i_0})}}\right)}{\sum_{i=2}^{89} K\left(\frac{d(X_i,X_{90})}{h_{k_{opt}(X_{i_0})}}\right)}$$

As it is done previously, we apply this multivariate procedure on Station 20 to predict ozone on August 31st, the 91st day. On Figure 5, the daily curves are plotted in grey and the black curve represents the curve we want to forecast. The results obtained at Step 1 (by the classical kernel method) are displayed in blue whereas those of Step 3 considering a $VAR(1)$ (from our procedure presented in Section 2) are displayed in red. Note that the results obtained with the classical method are similar to the univariate case. From this figure, one can observe that our method again improves upon the results obtained with the classical method. The MSE is 85.25 and the corresponding relative efficiency is 0.53 which is much better than using the univariate modeling of the error term.

To conclude, the case study shows that using our methodology can improve the ozone concentration predictions obtained with the classical kernel regression estimate. Neighboring stations should contain additional information on each other. A joint modeling is possible, as shown in Ettinger et al. (2012). Time series of random surfaces at one particular hour of ozone are used in that paper to predict a particular hour at one location the day after. It was shown there that this approach of borrowing strength across space compares favorably to using the whole series of times (i.e. a curve) over the previous day to predict the future values a particular hour at one location the day after.

4. Conclusion

We have developed a two-stage procedure in order to estimate a nonlinear functional regression where the explanatory variable is functional and the residual process is stationary and autocor-

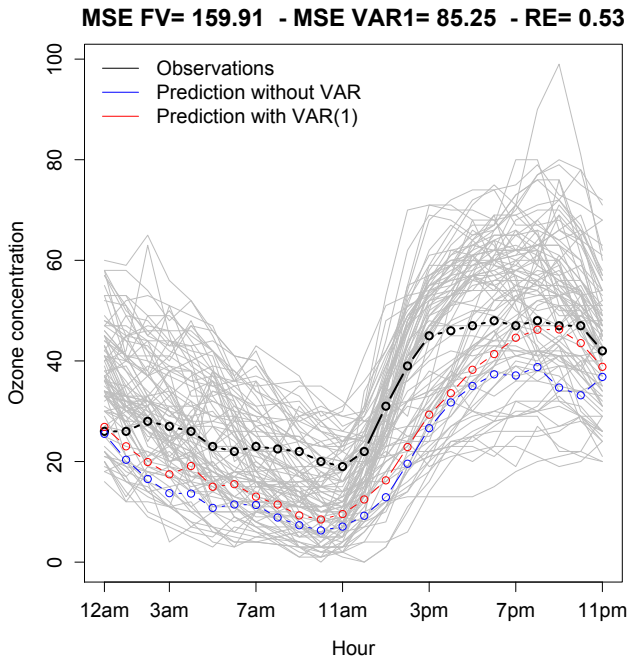

Figure 5: Ozone concentrations and predictions at Station 20 considering a $VAR(1)$ modeling of the error term

related. We have considered the case where the response variable is valued in $\mathbb{R}^d$, $d \geq 1$. This constitutes, to our knowledge a first step in the research on functional nonparametric modeling with a multivariate response. We have used the information of the autocorrelation function of the error process to improve the kernel-based estimation of the regression function. The asymptotic normality of our estimator is proved under some conditions. Some numerical results from a simulation case study and an application on real data illustrate the benefit of using this approach. Our methodology improves the standard kernel estimator in presence of highly autocorrelated data.

Potential improvements relate to the optimal implementation of our method. Indeed, the numerical illustrations indicate that there is a "sweet spot" where the number of time points provide enough information relative to the autocorrelation level to allow an optimal reduction of the prediction error. Another aspect is that for small autocorrelation levels, our approach deteriorates slightly the prediction errors compared to the use of independence-based kernel methods. An improved method should account for that fact and revert back to the basic independence-based kernel methods in these regimes.

Besides, in the literature on regression estimation with functional predictors, nonparameric methods with a Nadaraya-Watson-type estimator are successfully used. However, some authors proposed instead to use semi-parametric models which are shown to be interesting alternatives. For example, we could deal with functional partial linear model (see e.g. Aneiros-Pérez and Vieu (2008), Dabo-Niang and Guillas (2010)), functional index model (see e.g. Chen et al. (2011)),

or with functional projection pursuit regression (see e.g. Ferraty et al. (2013)). These kind of alternatives to fully non-parametric models are beyond the scope of this work but will be the aim of futur investigations. In this further work, our procedure taking into account the autocorrelation in the error term could be adapted in the more general framework of functional index models (Chen et al. (2011)). More precisely, we could place a linear model inside a link function. We could first estimate the parameters of the linear model using least squares methodology. Secondly, the link function could be estimated non parametrically including our two-stage procedure.

Finally, our approach could be extended to other time series of functional data. Aue et al. (2015) recently provided a dimension reduction technique with functional principal components (FPC) analysis that enables the use of vector-valued time series of FPC scores. However, this model did not allow of autocorrelation in the residuals that can still be present as we show in our ozone application above. A combination of the two approaches and that of Aue et al. (2014), that provided fully functional regression models allowing for autocorrelated errors, would have the potential to further improve the quality of predictions.

5. Appendix: Proofs of Theorems

We use C to signify a generic positive constant whose exact value may vary from case to case.

Lemma 1. *Under assumptions **H1**-**H5** (or more precisely **H1**, **H2**, **H3**(1), **H3**(2), **H4**, **H5**)*

$$(T\phi(h_0))^{1/2}\left[\overline{\mathbf{m}}(x) - \mathbf{m}(x) - \mathbf{B}_T^*(x)\right] \xrightarrow{d} \mathcal{N}_d(0, \Sigma_x)$$

with $\Sigma_x = \frac{C_2\mathbf{G}_2(x)}{C_1^2 f_1(x)}$ *is the* $(d \times d)$ *asymptotic covariance matrix with elements,* $x \in (\mathcal{C}, d)$ *whenever* $f_1(x) > 0$, *and* $\mathbf{B}_T^*(x) = \mathbb{E}[\overline{\mathbf{m}}(x)] - \mathbf{m}(x)$.

Lemma 1 comes from Theorem 5 in Masry (2005). The functions $\mathbf{G}_2(x)$, $f_1(x)$ and $\phi(h_0)$ are given in assumptions **H1** and **H3**.

Remark 3. *As stated in Masry (2005), if in addition to the assumptions of Lemma 1 we have* $T\phi(h_0)h_0^{2\beta} \to 0$ *(it is a stronger assumption on the bandwidth parameter) then one can remove the bias term from Lemma 1 that is* $(T\phi(h_0))^{1/2}\left[\overline{\mathbf{m}}(x) - \mathbf{m}(x)\right] \xrightarrow{d} \mathcal{N}_d(0, \Sigma_x)$.

5.1. Proof of Lemma 1

The proof follows work of Masry (2005). We make use of the Cramér-Wold device (see, e.g. Van der Vaart (1998), p.16), according to which it is sufficient to prove that for all $\mathbf{v} \in \mathbb{R}^d$, $\|\mathbf{v}\| \neq 0$, we have

$$(T\phi(h_0))^{1/2}\left[\mathbf{v}'\left(\overline{\mathbf{m}}(x) - \mathbf{m}(x) - \mathbf{B}_T^*(x)\right)\right] \xrightarrow{d} \mathcal{N}_d(0, \mathbf{v}'\Sigma_x\mathbf{v})$$

Note that $\underline{\mathbf{Y}}_t = \mathbf{m}(X_t) + \mathbf{e}_t$; $\mathbf{v}'\left(\overline{\mathbf{m}}(x) - \mathbf{m}(x)\right) = (\overline{m}_{\mathbf{v}}(x) - m_{\mathbf{v}}(x))$, where

$$\overline{m}_{\mathbf{v}}(x) = \frac{\dfrac{1}{T\mathbb{E}\left[K_0\left(\frac{d(x,X_1)}{h_0}\right)\right]}\sum_{t=1}^{T}\mathbf{v}'\underline{\mathbf{Y}}_t K_0\left(\dfrac{d(x,X_t)}{h_0}\right)}{\dfrac{1}{T\mathbb{E}\left[K_0\left(\frac{d(x,X_1)}{h_0}\right)\right]}\sum_{s=1}^{T}K_0\left(\dfrac{d(x,X_s)}{h_0}\right)} = \frac{\overline{m}_{2\mathbf{v}}(x)}{\overline{m}_1(x)}, \quad m_{\mathbf{v}}(x) = \mathbf{v}'m(x).$$

Then, we obtain $\overline{m}_{\mathbf{v}}(x) = m_{\mathbf{v}}(x) + B_T^{\mathbf{v}}(x) + V_T^{\mathbf{v}}(x)$ where $B_T^{\mathbf{v}}(x)$ is the bias term and $V_T^{\mathbf{v}}(x)$ is the variance effect defined by

$$B_T^{\mathbf{v}}(x) = \frac{\mathbb{E}[\overline{m}_{2\mathbf{v}}(x)] - m_{\mathbf{v}}(x)\mathbb{E}[\overline{m}_1(x)]}{\mathbb{E}[\overline{m}_1(x)]} \qquad V_T^{\mathbf{v}}(x) = \frac{Q_T^{\mathbf{v}}(x) - B_T^{\mathbf{v}}(x)(\overline{m}_1(x) - \mathbb{E}[\overline{m}_1(x)])}{\overline{m}_1(x)}$$

with $Q_T^{\mathbf{v}}(x) = (\overline{m}_{2\mathbf{v}}(x) - \mathbb{E}[\overline{m}_{2\mathbf{v}}(x)]) - m_{\mathbf{v}}(x)(\overline{m}_1(x) - \mathbb{E}[\overline{m}_1(x)])$. By the result of Masry (2005), $B_T^{\mathbf{v}}(x) = o(h_0^{\beta})$ and using same lines as in the proof of Theorem 5 in Masry (2005), we have

$$(T\phi(h_0))^{1/2}\left[\mathbf{v}'\left(\overline{\mathbf{m}}(x) - \mathbf{m}(x) - \mathbf{B}_T^*(x)\right)\right] \xrightarrow{d} \mathcal{N}_d(0, \mathbf{v}'\Sigma_x\mathbf{v}).$$

5.2. Proof of Theorem 1

We have $\widetilde{\mathbf{m}}(x) = \dfrac{\frac{1}{T\mathbb{E}\left[\Delta_1^{(1)}(x)\right]}\sum_{t=1}^{T}\widehat{\underline{\mathbf{Y}}}_t\Delta_t^{(1)}(x)}{\frac{1}{T\mathbb{E}\left[\Delta_1^{(1)}(x)\right]}\sum_{t=1}^{T}\Delta_t^{(1)}(x)} = \dfrac{\widetilde{\mathbf{m}}_2(x)}{\widetilde{m}_1(x)}, \quad \widetilde{\mathbf{m}}_2(x) = (\widetilde{m}_{1,2}(x), \ldots, \widetilde{m}_{d,2}(x))'$

where

$$\widetilde{m}_{j,2}(x) = \frac{1}{T\mathbb{E}\left[\Delta_1^{(1)}(x)\right]}\sum_{t=1}^{T}\widehat{\underline{Y}}_{t,j}\Delta_t^{(1)}(x),\ j = 1, \ldots, d$$

$$\widetilde{m}_1(x) = \frac{1}{T\mathbb{E}\left[\Delta_1^{(1)}(x)\right]}\sum_{t=1}^{T}\Delta_t^{(1)}(x)$$

$$\sum_{t=1}^{T}\widehat{\underline{Y}}_{t,j}\Delta_t^{(1)}(x) = \sum_{t=1}^{T}\underline{Y}_{t,j}\Delta_t^{(1)}(x) - \sum_{t=1}^{T}\left[(\widehat{\Pi}_1 - \Pi_1)u_{t-1}\right]_j\Delta_t^{(1)}(x) + \sum_{t=1}^{T}\left[\Pi_1\left(\widehat{\mathbf{m}}(X_{t-1}) - \mathbf{m}(X_{t-1})\right)\right]_j\Delta_t^{(1)}(x)$$
$$+ \sum_{t=1}^{T}\left[\left(\widehat{\Pi}_1 - \Pi_1\right)\left(\widehat{\mathbf{m}}(X_{t-1}) - \mathbf{m}(X_{t-1})\right)\right]_j\Delta_t^{(1)}(x)$$

Then $\widetilde{m}_{j,2}(x) = \overline{m}_{j,2}(x)^\sharp - \widetilde{m}_{j,21}(x) + \widetilde{m}_{j,22}(x) + \widetilde{m}_{j,23}(x)$ with

$$\overline{m}_{j,2}(x)^\sharp = \frac{1}{T\mathbb{E}\left[\Delta_t^{(1)}(x)\right]} \sum_{t=1}^{T} \underline{Y}_{t,j} \Delta_t^{(1)}(x)$$

$$\widetilde{m}_{j,21}(x) = \frac{1}{T\mathbb{E}\left[\Delta_t^{(1)}(x)\right]} \sum_{t=1}^{T} [(\widehat{\Pi}_1 - \Pi_1) u_{t-1}]_j \Delta_t^{(1)}(x)$$

$$\widetilde{m}_{j,22}(x) = \frac{1}{T\mathbb{E}\left[\Delta_t^{(1)}(x)\right]} \sum_{t=1}^{T} \left[\Pi_1 \left(\widehat{\mathbf{m}}(X_{t-1}) - \mathbf{m}(X_{t-1})\right)\right]_j \Delta_t^{(1)}(x)$$

$$\widetilde{m}_{j,23}(x) = \frac{1}{T\mathbb{E}\left[\Delta_t^{(1)}(x)\right]} \sum_{t=1}^{T} \left[(\widehat{\Pi}_1 - \Pi_1) \left(\widehat{\mathbf{m}}(X_{t-1}) - \mathbf{m}(X_{t-1})\right)\right]_j \Delta_t^{(1)}(x)$$

Note that $\overline{m}_{j,2}(x)^\sharp = \overline{m}_{j,2}(x)$ and $\widetilde{m}_1(x) = \overline{m}_1(x)$ with K_0 and h_0 replaced by K_1 and h_1 respectively. Since $\widetilde{m}_j(x) = \dfrac{\widetilde{m}_{j,2}(x)}{\widetilde{m}_1(x)}$ we have $\widetilde{\mathbf{m}}(x) = \overline{\mathbf{m}}(x) - Q_{T_1} + Q_{T_2} + Q_{T_3}$ with $Q_{T_l} = (Q_{1,T_l}, \ldots, Q_{d,T_l})$, with $Q_{j,T_l} = \frac{\widetilde{m}_{j,2l}(x)}{\widetilde{m}_1(x)}$, for $l = 1, 2, 3$, $j = 1, \ldots, d$. We analyze the asymptotic properties of $Q_{j,T_l}, l = 1, 2, 3$, in Lemmas 2, 3 and 4, which are key results for proof of this theorem since Lemma 1 proves the asymptotic normality of $\overline{\mathbf{m}}(x)$.

Lemma 2. *Under assumptions **H1**-**H5**, for $j = 1, \ldots, d$, we have $Q_{j,T_1} = o_p\left(\sqrt{\frac{1}{T\phi(h_1)}}\right)$.*

Lemma 3. *Under assumptions **H1**-**H4**, for $j = 1, \ldots, d$, we have $Q_{j,T_2} = O_p\left(h_0^\beta\right) + o_p\left(\sqrt{\frac{1}{T\phi(h_1)}}\right)$.*

Lemma 4. *Under assumptions **H1**-**H5**, for $j = 1, \ldots, d$, we have $Q_{j,T_3} = O_p\left(h_0^\beta\right) + o_p\left(\sqrt{\frac{1}{T\phi(h_1)}}\right)$.*

The proofs of Lemmas 2, 3 and 4 are given in the supplementary file.

Proof of Theorem 2

The proof is similar to that of Theorem 1 since $\widetilde{m}_j(x) = \overline{m}_j(x) - Q_{j,T_1} + Q_{j,T_2} + Q_{j,T_3}$ with $Q_{j,T_l} = \frac{\widetilde{m}_{j,2l}(x)}{\widetilde{m}_1(x)}$, for $l = 1, 2, 3$, $j = 1, \ldots, d$. It then follows directly from Corollary 1 of Masry (2005), Lemmas 2, 3 and 4, which are key results for proof of this theorem.

Aneiros-Pérez, G., Vieu, P., 2008. Nonparametric time series prediction: A semi-functional partial linear modeling. Journal of Multivariate Analysis 99, 834–857.

Antoniadis, A., Brossat, X., Cugliari, J., Poggi, J.M., 2014. Une approche fonctionnelle pour la prévision non-paramétrique de la consommation d'électricité. Journal de la Société Française de Statistique 155, 202–219.

Aston, J.A., Chiou, J.M., Evans, J.P., 2010. Linguistic pitch analysis using functional principal component mixed effect models. J. R. Stat. Soc. Ser. C. Appl. Stat. 59, 297–317.

Aue, A., Hormann, S., Horvàth, L., Huvskovà, M., 2014. Dependent functional linear models with applications to monitoring structural change. Statistica Sinica 24, 1043–1073.

Aue, A., Norinho, D.D., Hormann, S., 2015. On the prediction of stationary functional time series. Journal of the American Statistical Association 110.

Benhenni, K., Ferraty, F., Rachdi, M., Vieu, P., 2007. Local smoothing regression with functional data. Computational Statistics 22, 353–369.

Chen, D., Hall, P., Müller, H.G., et al., 2011. Single and multiple index functional regression models with nonparametric link. The Annals of Statistics 39, 1720–1747.
Cuevas, A., 2014. A partial overview of the theory of statistics with functional data. J. Statist. Plann. Inference 147, 1–23.
Dabo-Niang, S., 2004. Kernel density estimator in an infinite-dimensional space with a rate of convergence in the case of diffusion process. Appl. Math. Lett. 17, 381–386.
Dabo-Niang, S., Guillas, S., 2010. Functional semiparametric partially linear model with autoregressive errors. Journal of Multivariate Analysis 101, 307–315.
Dabo-Niang, S., Rhomari, N., 2009. Kernel regression estimation in a banach space. J. Statist. Plann. Inference 139, 1421–1434.
Delsol, L., 2007. CLT and L_q errors in nonparametric functional regression. C. R. Math. Acad. Sci. 345, 411–414.
Delsol, L., 2009. Advances on asymptotic normality in non-parametric functional time series analysis. Statistics 43, 13–33.
Ettinger, B., Guillas, S., Lai, M.J., 2012. Bivariate splines for ozone concentration forecasting. Environmetrics 23, 317–328.
Ferraty, F., Goia, A., Salinelli, E., Vieu, P., 2013. Functional projection pursuit regression. Test 22, 293–320.
Ferraty, F., Goia, A., Vieu, P., 2002a. Functional nonparametric model for time series : a fractal approach for dimension reduction. Test 11, 317–344.
Ferraty, F., Goia, A., Vieu, P., 2002b. Régression non-paramétrique pour des variables aléatoires fonctionnelles mélangeantes. C. R. Math. Acad. Sci. Paris 334, 217–220.
Ferraty, F., Mas, A., Vieu, P., 2007. Advances on nonparametric regression for functionnal data. Aust. N. Z. J. Stat. 49, 267–286.
Ferraty, F., Vieu, P., 2002. The functional nonparametric model and application to spectrometric data. Comput. Statist. 17, 545–564.
Ferraty, F., Vieu, P., 2004. Nonparametric models for functional data, with application in regression, time series prediction and curve discrimination. J. Nonparametr. Stat. 16, 111–125.
Ferraty, F., Vieu, P., 2006. Nonparametric functional data analysis: theory and practice. Springer.
Gasser, T., Hall, P., Presnell, B., 1998. Nonparametric estimation of the mode of a distribution of random curves. J. R. Stat. Soc. Ser. B Stat. Methodol. 60, 681–691.
Guillas, S., Lai, M.J., 2010. Bivariate splines for spatial functional regression models. J. Nonparametr. Stat. 22, 477–497.
Hardle, W., Marron, J.S., 1985. Optimal bandwidth selection in nonparametric regression function estimation. The Annals of Statistics , 1465–1481.
Hart, J.D., Vieu, P., 1990. Data-driven bandwidth choice for density estimation based on dependent data. The Annals of Statistics , 873–890.
Lin, X., Carroll, R.J., 2000. Nonparametric function estimation for clustered data when the predictor is measured without/with error. Journal of the American statistical Association 95, 520–534.
Masry, E., 2005. Nonparametric regression estimation for dependent functional data: asymptotic normality. Stochastic Process. Appl. 115, 155–177.
Ogden, R.T., Greene, E., 2010. Wavelet modeling of functional random effects with application to human vision data. J. Statist. Plann. Inference 140, 3797–3808.
Park, A., Guillas, S., Petropavlovskikh, I., 2013. Trends in stratospheric ozone profiles using functional mixed models. Atmospheric Chemistry and Physics 13, 11473–11501.
Rachdi, M., Vieu, P., 2007. Nonparametric regression for functional data: Automatic smoothing parameter selection. Journal of Statistical Planning and Inference 137, 2784–2801.
Ramsay, J., Silverman, B., 2005. Functional data analysis. 2nd ed., Springer.
Rosenblatt, M., 1956. A central limit theorem and a strong mixing condition. Proceedings of the National Academy of Sciences of the United States of America 42, 43.
Ruckstuhl, A., Welsh, A., Carroll, R.J., 2000. Nonparametric function estimation of the relationship between two repeatedly measured variables. Statist. Sinica 10, 51–71.
Sangalli, L.M., Ramsay, J.O., Ramsay, T.O., 2013. Spatial spline regression models. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 75, 681–703.
Van der Vaart, A., 1998. Asymptotic Statistics. volume 3. Cambridge University Press.
Xiang, D., Qiu, P., Pu, X., 2013. Nonparametric regression analysis of multivariate longitudinal data. Statistica Sinica 23, 769–789.
Xiao, Z., Linton, O., Carroll, R., Mammen, E., 2003. More efficient local polynomial estimation in nonparametric

regression with autocorrelated errors. J. Amer. Statist. Assoc. 98, 980–992.