Conditional Autoregressive Hilbertian processes

Jairo Cugliari

Conditional Autoregressive Hilbertian processes

Jairo Cugliari

INRIA Research Team Select, Université Paris Sud Bât. 425, 91405 Orsay Cedex, France

Abstract

When considering the problem of forecasting a continuous-time stochastic process over an entire time-interval in terms of its recent past, the notion of Autoregressive Hilbert space processes (ARH) arises. This model can be seen as a generalization of the classical autoregressive processes to Hilbert space valued random variables. Its estimation presents several challenges that were addressed by many authors in recent years.

In this paper, we propose an extension based on this model by introducing a conditioning process on the ARH. In this way, we are aiming a double objective. First, the intrinsic linearity of ARH is overwhelm. Second, we allow the introduction of exogenous covariates on this function-valued time series model.

We begin defining a new kind of processes that we call Conditional ARH. We then propose estimators for the infinite dimensional parameters associated to such processes. Using two classes of predictors defined within the ARH framework, we extend these to our case. Consistency results are provided as well as a real data application related to electricity load forecasting.

Keywords: Functional Data, Nonparametric, Forecasting, Exogenous covariate
2000 MSC: 62G08, 62M10

1. Introduction

We consider a function-valued process $Z = (Z_k, k \in \mathbb{Z})$ where for each k, Z_k is a random element taking his values in some functional space F. A popular choice is to set $F = H$ a real separable Hilbert space because of the rich geometric properties of Hilbert spaces. As for classical time series, an important task is the problem of obtaining some information about the future value Z_{n+1} from the observed discrete sequence $Z_1, \ldots, Z_n$. Then, the best predictor (in the quadratic mean loss function sense) of the future observation Z_{n+1} is its conditional expectation given the past

$$\tilde{Z}_{n+1} = \mathbb{E}(Z_{n+1}|Z_n, \ldots, Z_1), \tag{1}$$

which may depend on the unknown distribution of Z.

One important case arises when one assumes that Z is a strictly stationary zero-mean Autoregressive Hilbertian process of order 1 ARH(1), introduced by Bosq [1] and defined by

$$Z_{k+1} = \rho Z_k + \epsilon_k, \qquad k \in \mathbb{Z}, \tag{2}$$

with ρ a bounded linear operator over H and $\epsilon = (\epsilon_k, k \in \mathbb{Z})$ a strong H-valued white noise. For this process, the best predictor of Z_{n+1} given the past observations is $\tilde{Z}_{n+1} = \rho Z_n$. Notice that ρ is

---

usually unknown. Two forecasting strategies can be followed here. The first one is to first estimate ρ and then apply it to the last observation Z_n to obtain a prediction of Z_{n+1} (see Bosq [1], Besse and Cardot [2], Pumo [3]). Alternatively, one may directly predict $\tilde{Z}_{n+1}$ by estimating the relevant elements of the range of ρ^* (see Antoniadis and Sapatinas [4]). Adopting this last strategy and using some wavelet decomposition the later authors obtain considerable better prediction results. The choice of a wavelet basis is guided by the good approximation properties they have to represent quite irregular trajectories of Z. Kargin and Onatski [5] also use the second strategy but propose to use a data-dependent basis adapted to the prediction task.

While ARH processes are a natural generalization of the well known autoregressive processes in Euclidean spaces, the infinite dimension of the space H produces new challenges for their estimation and prediction (see Mas and Pumo [6] for a recent review on this topic). A second issue is the study of some of the extensions developed on the scalar case to the Hilbertian framework, like for instance higher order ARH processes studied in Pumo [7].

We are interested in another extension taking into account some exogenous information modeled by the influence of covariates in the model given by equation (2). We may cite Mas and Pumo [8] that uses the derivative of Z_k as a covariate, or Damon and Guillas [9] that introduces a function-valued covariate also following an ARH process. In both these works, the covariates are introduced as additive terms in the equation (2).

Alternatively, one may introduce exogenous information through the linear operator ρ. Like in the scalar case, one may consider a more general case where the parameter ρ depends on some covariate. For such cases, the exogenous information may be incorporated in a non-additive manner. Guillas [10] propose to model Z by a doubly stochastic Hilbert process defined by

$$Z_k = \rho_{V_k}(Z_{k-1}) + \epsilon_k, \qquad k \in \mathbb{Z}, \tag{3}$$

where $V = (V_k, k \in \mathbb{Z})$ is a sequence of independent identically distributed Bernoulli variables. The intuition behind the model is that there exists two regimes expressed through two different operators, ρ_0 and ρ_1. At each instant k, one of the regimes is randomly chosen as the result of the drawn of the associate Bernoulli variable V_k. The resulting process admits to have one of the regimes to be explosive if it is not visited too often. In such a case, equation (3) has a unique stationary solution.

In this paper, we introduce the *Conditional Autoregressive Hilbertian process* (CARH), constructed such that conditionally on an exogenous covariate V, the process Z follows an ARH process. While CARH definition is similar to equation (3), it differs mainly in two ways. The first one is related to the nature of the process V which we assume to be a multivariate random process with some continuous distribution. Second, we propose predictors that will accomplish the prediction task using the exogenous information. Indeed, the exogenous information of the actual regime is used to found similar local situations on the observed past.

The paper is structured as follows. In Section 2 we introduce the main definitions and we present the model. Linear operators on Hilbert spaces are intensively used through out the article. On Appendix A we recall some important facts on this topic that we use on the article. We also propose estimators for the unknown parameters as well as two classes of predictors. The main results about the convergence of the estimators and predictors are shown in Section 3 postponing the proofs until the Appendix B. Finally, Section 4 contains a real data application of CARH processes which illustrates empirically the performance of the predictors.

2

2. Conditional ARH process: CARH.

After some notations, we define the CARH process and we propose estimators for the associated parameters operators. Then, we follow prediction strategies similar to those adopted in previous studies for ARH processes, to obtain classes of predictors for the CARH process.

2.1. Preliminaries

All variables are defined on the same probability space $(\Omega, \mathcal{F}, \mathbb{P})$. We consider a sequence $Z = (Z_k, k \in \mathbb{Z})$ of Hilbert space valued random variables, i.e. each random variable Z_k is a measurable map from the probability space in an real separable Hilbert space H endowed with its Borel σ-field, $\mathcal{H}$. The space H is equipped with the scalar product $< ., . >_H$ and the induced norm $\|.\|_H$). We also consider a sequence of real $\mathbb{R}^d$−valued random variables $V = (V_k, k \in \mathbb{Z})$. Both sequences Z and V are assumed to be stationary. We will focus on the behaviour of Z conditionally on V. We will further assume that Z is strongly integrable.

The conditional expectation is characterized by the conditional distribution of Z given V, i.e. by the conditional probability $\mathbb{P}_{Z|V}$ on $\mathcal{H}$. In order to ensure that this conditional probability is properly defined as a measure (in the sense that it represents a regular version of the conditional probability), it is assumed that a transition probability exists that associates to each $v \in \mathbb{R}^d$ a probability measure $\mathbb{P}^v$ on $(H, \mathcal{H})$ such that

$$\mathbb{P}^v_{Z|V}(A) = \mathbb{P}^v(A), \qquad \text{for every } A \in \mathcal{H}, v \in \mathbb{R}^d.$$

We call $\mathbb{P}^v$ the sampling measure and denote $\mathbb{E}^v$ the induced expectation. We restrict our attention to functions defined over a real compact interval $\mathcal{T}$ and we assume hereafter T to be $[0, 1]$ without loss of generality. More precisely, we set H to be the subspace of continuous functions on the space of classes of 4-th order $\mathbb{P}$−integrable functions.

2.2. The model

A sequence $(Z, V) = \{(Z_k, V_k), k \in \mathbb{Z}\}$ of $H \times \mathbb{R}^d$-valued random variables is a Conditional Autoregressive Hilbertian process (CARH) of order 1 if it is stationary and such that, for each k

$$Z_k = a + \rho_{V_k}(Z_{k-1} - a) + \epsilon_k, \tag{4}$$

where the conditional mean function $a_v = \mathbb{E}^v[Z_0|V], v \in \mathbb{R}^d$, is the conditional expectation (on V) of the process, $\epsilon = (\epsilon_k, k \in \mathbb{Z})$ is an H−valued white noise and $(\rho_{V_k}, k \in \mathbb{Z})$ is a sequence of random operators such that, conditionally on V, ρ_V is a linear compact operator on H (see Appendix A). Additionally, V and ϵ are independent process. Using the following assumptions we prove the existence and uniqueness of the CARH processes (see Appendix B for the proof).

Assumptions 2.1. Assume that:

1. There exists a map $v \mapsto P^v$ that assigns a probability measure on $(H, \mathcal{H})$ to each value v in the support of V.
2. $\sup_n \|\rho_{V_n}\|_{\mathcal{L}} = M_\rho < 1$.

Theorem 2.2. *Under Assumptions 2.1, equation* (4) *defines a* CARH *process with an unique stationary solution given by*

$$Z_k = a + \sum_{j=0}^{\infty} \left(\prod_{p=0}^{j-1} \rho_{V_{k-p}} \right) (\epsilon_{k-j}),$$

with the convention $\prod_{p=0}^{j-1} \rho_{V_{k-p}} = Id$ *(the identity operator) for* $j = 0$.

3

The first condition in Assumption 2.1 has already been discussed. The second one ensures the contraction of the conditional autoregressive operator through the supremum norm of linear operators (see Appendix A).

2.3. Associated operators

Hereafter we make the additional assumption that $\mathbb{E}^v[\|Z\|_H^4|V] < \infty$. Let us note H^* the topological dual of H, i.e. the space of bounded linear functionals on H. We introduce two linear operators mapping from H^* to H associated to the CARH process. Thanks to the Riesz representation, H^* the topological dual of H can be identified with H, and the operators may be defined as follows:

$$z \in H \mapsto \Gamma_v z = \mathbb{E}^v[((Z_0 - a) \otimes (Z_0 - a))(z)|V] \qquad \text{and}$$
$$z \in H \mapsto \Delta_v z = \mathbb{E}^v[((Z_0 - a) \otimes (Z_1 - a))(z)|V],$$

that we call *conditional (on V) covariance* and *cross covariance operators* respectively. We have used the tensor product notation $(u \otimes v)(z) = <u, z>_H v$ for $u, v, z \in H$.

For each $v \in \mathbb{R}^d$, both Γ_v and Δ_v are trace-class and hence Hilbert-Schmidt. In addition, Γ_v is positive definite and self adjoint. Then, we may write down the spectral decomposition of Γ_v as

$$\Gamma_v = \sum_{j \in \mathbb{N}} \lambda_{v,j}(e_{v,j} \otimes e_{v,h})$$

where $(\lambda_{v,j}, e_{v,j})_{j \in \mathbb{N}}$ are the eigen-elements of Γ_v. The eigenvalues may be arranged to form a non-negative decreasing sequence of numbers tending towards zero.

As a direct consequence of the choice made for H, the operators have associated kernels γ_v and δ_v defined over $L_2([0,1]^2)$ such that

$$\Gamma_v(z)(t) = \int_0^1 \gamma_v(s,t)z(s)ds,$$
$$\Delta_v(z)(t) = \int_0^1 \delta_v(s,t)z(s)ds, \qquad t \in [0,1], v \in \mathbb{R}^d, z \in H,$$

with $\gamma_v(.,.)$ a continuous, symmetric and positive kernel and $\delta_v(.,.)$ a continuous kernel. The kernels turn to be the *conditional covariance function* $\gamma_v(s,t) = \mathbb{E}^v[(Z_0(s) - a(s))(Z_0(t) - a(t))|V]$, and the *one-step-ahead conditional cross covariance function* $\delta_v(s,t) = \mathbb{E}^v[(Z_0(s) - a(s))(Z_1(t) - a(t))|V]$, $(s,t) \in [0,1]^2, v \in \mathbb{R}^d$.

A Yule-Walker like relation links the operators Δ_v, Γ_v and ρ_v. For each $v \in \mathbb{R}^d$ we have

$$\Delta_v = \rho_v \Gamma_v. \tag{5}$$

Using the property of the adjoint and the symmetry of Γ_v, we obtain from (5) the following key relation for the estimation of ρ_v (see Section 2.5),

$$\Delta_v^* = \Gamma_v \rho_v^*. \tag{6}$$

2.4. Estimation of a, Γ_v, Δ_v.

The parameters can be estimated from data. We call $\{(Z_1, V_1), \ldots, (Z_n, V_n)\}$ the observed data supposed to come from a CARH process. We use nonparametric Nadaraya-Watson like estimators to estimate the infinite-dimensional parameters a_v, Γ_v and Δ_v. This is a popular choice when the the parameters are defined through conditional expectations.

2.4.1. Estimation of a_v.

We estimate the conditional mean function of the process $a_v(t) = \mathbb{E}^v[Z_0(t)|V]$ for all $t \in [0, 1]$ using the observations $\{(Z_1, V_1), \ldots, (Z_n, V_n)\}$. In order to properly define the framework, let us introduce some quantities. For some fixed $t \in [0, 1]$, set $Y = Z_0(t)$ and $Y_i = Z_i(t), i = 1, \ldots, n$. Let us assume that the distribution of V admits a density f with respect to the Lebesgue measure. We define for $v \in \mathbb{R}^d$

$$g_v(t) = \mathbb{E}^v[Z_0(t)f(V)|V],$$

and provided that $f(v) > 0$ we rewrite the parameter as the regression of Y against V,

$$a_v(t) = \mathbb{E}^v[Y|V]$$
$$= \frac{g_v(t)}{f(v)}.$$

When $f(v) = 0$ we set $a_v(t) = \mathbb{E}[Y]$.

The introduced quantities can be estimated by Nadaraya-Watson kernel based estimators. In our case, we use the following estimators for f and g_v respectively,

$$\widehat{f}_n(v) = \frac{1}{nh_a^d} \sum_{i=1}^{n} K(h_a^{-1}(V_i - v)) \qquad \text{and} \tag{7}$$

$$\widehat{g}_{v,n}(t) = \frac{1}{nh_a^d} \sum_{i=1}^{n} K(h_a^{-1}(V_i - v))Y_i, \tag{8}$$

where $K : \mathbb{R}^d \mapsto \mathbb{R}$ is a unitary square-integrable $d-$dimensional kernel and the bandwidth $h_a = (h_{a,n})_{n \in \mathbb{N}}$ is a decreasing sequence of positive numbers tending to 0 called the bandwidth. The estimator of $a_v(t)$ is then given by

$$\widehat{a}_{v,n}(t) = \frac{\widehat{g}_{v,n}(t)}{\widehat{f}_n(v)},$$

which can be written as $\widehat{a}_{v,n} = \sum_{i=1}^{n} w_{n,i}(v, h_a)Y_i$ which is a weighted mean of the observed values with weights given by

$$w_{n,i}(v, h) = \frac{K(h^{-1}(V_i - v))}{\sum_{i=1}^{n} K(h^{-1}(V_i - v))}. \tag{9}$$

2.4.2. Estimation of Γ_v.

For the estimation of Γ_v we proceed in an analogous way. Without loss of generality, we assume that Z is centered. First, for $(s, t) \in [0, 1]^2$ fixed, consider the real valued variables $Y = Z_0(s)Z_0(t)$ and the observations $Y_i = Z_i(s)Z_i(t)$ with $i = 1, \ldots, n$. Now redefine the auxiliary quantity g_v using the new definition of Y and Y_i, $i = 1, \ldots, n$. Set $g_v(t) = \mathbb{E}^v[Z_0(s)Z_0(t)|V]$ and write the parameter again as the regression of Y against V. Then, with a similar reasoning it follows that the estimator of the kernel of γ_v at (s, t) is

$$\widehat{\gamma}_{v,n}(s, t) = \sum_{i=1}^{n} w_{n,i}(v, h_\gamma)Z_i(s)Z_i(t),$$

with weights given by (9). Moreover, on the general case of a not necessarily centered process the estimator of Γ_v can be written as

$$\widehat{\Gamma}_{v,n} = \sum_{i=1}^{n} w_{n,i}(v, h_\gamma)(Z_i - \widehat{a}_{v,n}) \otimes (Z_i - \widehat{a}_{v,n}). \tag{10}$$

2.4.3. Estimation of the conditional cross covariance operator Δ_v.

Again, the estimation of the operator is done through the estimation of its kernel, which is in this case the conditional cross covariance function δ_v. We work first with the centered process. Fix $(s,t) \in [0,1]^2$ and again redefine $Y = Z_0(s)Z_1(t)$ and the observations $Y_i = Z_{i-1}(s)Z_i(t)$ for $i = 2, \ldots, n$. Define f and g_v and their estimators of the same form as (7) and (8) respectively using the bandwidth h_δ and the new variables Y and Y_i. The resulting estimator of $\delta_v(s,t)$ is

$$\widehat{\delta}_{v,n}(s,t) = \sum_{i=2}^{n} w_{n,i}(v, h_\delta) Z_{i-1}(s) Z_i(t).$$

We can now plug-in the estimated kernel on the operator which yields the estimator of Δ_v. We write it for the general case of a non centered process as,

$$\hat{\Delta}_{n,v} = \sum_{i=2}^{n} w_{n,i}(v, h_\delta)(Z_{i-1} - \hat{a}_n(v)) \otimes (Z_i - \hat{a}_n(v)),$$

where the weights are given by equation (9).

Remark

If the denominator on equation (9) defining the weights is equal to zero, i.e. $\widehat{f}_n(v) = 0$, then one usually sets the weights to $w_{n,i}(v, h_a) = n^{-1}$ or $w_{n,i}(v) = 0$ for all $i = 1, \ldots, n$ in order to define the estimator for all $v \in \mathbb{R}^d$. The weights are more important for those segments Z_i with closer value of V_i to the target v. The bandwidth plays a key role, tuning the proximity of the scatter of $\mathbb{R}^d$ to v via the scaling of the kernel function. Large values of h lead to weights $w_{n,i}$ that are not negligible for an important number of observations. Conversely, small values result in only few observations having a significant impact on the estimator. This produces the common trade-off between bias and variance of kernel regression estimators.

2.5. Estimation of ρ_v.

The intrinsic infinite dimension of the space makes difficult the estimation of the operator ρ_v. If H is finite-dimensional, the equation (5) provides a natural way of estimating ρ_v. One may plug-in the empirical counterparts of the covariance operators and solve the equation in ρ_v. However, when H has infinite dimension, Γ_v is not invertible anymore. To well identify ρ_v from (5) the eigenvalues of Γ_v need to be strictly positive. An analogous assumption is to ask the kernel of Γ_v to be null (see Mas and Pumo [6]). In this case, a linear measurable mapping Γ_v^{-1} can be defined as $\Gamma_v^{-1} = \sum_{j \in \mathbb{N}} \lambda_{v,j}^{-1}(e_{v,j} \otimes e_{v,j})$ with domain

$$\mathcal{D}_{\Gamma_v^{-1}} = \left\{ z = \sum_{j \in \mathbb{N}} <e_{v,j}, z> e_{v,j} \in H : \sum_{j \in \mathbb{N}} \left(\frac{<e_{v,j}, z>_H}{\lambda_{v,j}} \right)^2 < \infty \right\},$$

that is a dense subset of H. It turns to be an unbounded operator and in consequence continuous nowhere. Hence, there is no hope to obtain any theoretical asymptotic result. However, from (5) we obtain that

$$\rho_v^\diamond = \Delta_v \Gamma_v^{-1},$$

where $\rho_v^\diamond$ is the conditional autoregression operator ρ_v restricted to $\mathcal{D}_{\Gamma_v^{-1}}$ as a consequence of $\Gamma_v \Gamma_v^{-1} = I_{\mathcal{D}_{\Gamma_v^{-1}}}$. On the other hand, since the adjoint of a linear operator in H with a dense

domain is closed (*closed graph theorem*, see for example Kato [11, Theorem 5.20]) and since the range of the adjoint of the cross-covariance operator, Δ_v^*, is included in $\mathcal{D}_{\Gamma_v^{-1}}$ we can deduce from (6) that over $\mathcal{D}_{\Gamma_v^{-1}}$,

$$\rho_v^* = \Gamma_v^{-1}\Delta_v^*.$$

As pointed out by Mas [12] one can use classical results on linear operators to extend by continuity the definition of $\rho_v^\diamond$ to H, in order to obtain

$$\rho_v = \mathrm{Ext}(\rho_v^\diamond) = (\Gamma_v^{-1}\Delta_v^*)^* = (\Delta_v\Gamma_v^{-1})^{**}.$$

Therefore one may focus on the estimation of ρ_v^* because of the theoretical properties are applicable to ρ_v through the composition of ρ^* by the adjoint operator.

We can now propose two classes of estimators for ρ_v^* (see Mas [12] for analogy with the estimators on the ARH setting). The first one, the class of *projection estimators*, projects the function space valued observations on an appropriate subspace H_{v,k_n} of finite dimension $k_n = k_{v,n}$. Let Π_{v,k_n} be the projector operator over H_{v,k_n}. Then one inverts the linear operator defined by the random matrix $\Pi_{v,k_n}\Gamma_{v,n}\Pi_{v,kn}$ and completes with the null operator on the orthogonal subspace. For example, the space H_{v,k_n} may be set equal to the one generated by the first k_n eigenfunctions of Γ_v. Then, the subspace H_{v,k_n} is estimated by $\widehat{H}_{v,k_n}$, the linear span of the first k_n empirical eigenfunctions. By this way, if P_{v,k_n} is the projection operator on $\widehat{H}_{v,k_n}$, the estimator of ρ_v^* can be written as

$$\widehat{\rho}_{v,n,k_n}^* = (P_{v,k_n}\widehat{\Gamma}_{v,n}P_{v,k_n})^{-1}\widehat{\Delta}_{v,n}^*P_{v,k_n}. \tag{11}$$

The estimation solution by projection over a finite dimensional space is equivalent to approximate Γ_v^{-1} by a linear operator with additional regularity $\Gamma_v^\dagger$ defined as

$$\Gamma_v^\dagger = \sum_{j=1}^{k_n} b(\lambda_{v,j})(e_{v,j}\otimes e_{v,j}),$$

where $(k_n)_n$ is an increasing sequence of integers tending to infinity and b is some smooth function converging point-wise to $x \mapsto 1/x$. Indeed, $\Gamma_v^\dagger \to \Gamma_v^{-1}$ when $k_n \to \infty$. The choice of taking $b(x) = 1/x$ yields, for a finite k_n, to set $\Gamma_v^\dagger$ equal to a spectral cut of Γ_v^{-1}. However, this choice is not unique. Mas [12] considers a family of functions $b_{p,\alpha} : \mathbb{R}^+ \mapsto \mathbb{R}^+$ with $p \in \mathbb{N}$ such that

$$b_{p,\alpha}(x) = \frac{x^p}{(x+\alpha_n)^{p+1}},$$

with α_n a strictly positive sequence that tends to 0 as $n \to +\infty$. With this, the second class of estimators for ρ^*, the *resolvent class*, is defined as

$$\widehat{\rho}_{v,n,p,\alpha}^* = b_{p,\alpha}(\widehat{\Gamma}_{v,n})\widehat{\Delta}_{v,n}^*, \tag{12}$$

where we write $b_{p,\alpha}(\widehat{\Gamma}_{v,n}) = (\widehat{\Gamma}_{v,n} + \alpha_n I)^{-(p+1)}$ with $p \geq 0$, $\alpha_n \geq 0$, $n \geq 0$. Then, the operator $b_{p,\alpha}(\widehat{\Gamma}_{v,n})$ can be associated to a regularized approximation of Γ_v^{-1} (see Antoniadis and Sapatinas [4] for a discussion on this topic applied to the ARH estimation).

Finally, both classes of estimators allow one to predict the future value Z_{n+1} from the observations by first estimating the autocorrelation operator ρ_v^* and then applying it to the last available observation Z_n.

3. Main results

In this section we announce the main theoretical results that justify the choices made on the estimators presented in the previous section.

Neither $(V_k, k \in \mathbb{Z})$ nor $(Z_k, k \in \mathbb{Z})$ are assumed to have independent components. We deal with their dependence through a strong mixing hypothesis, that is, we assume each sequence to be asymptotically independent by controlling the decay of the dependence. Many contexts of mixing exist in the literature. In general one relies upon a measure of the decay of a dependence of two observations as a function of their time gap. We use the 2-α-mixing setting, a slightly weaker setting than the α-mixing one (see Bosq and Blanke [13]). Let $X = (X_k, k \in \mathbb{Z})$ be a stationary random process and consider the $\sigma-$algebras $\sigma(X_0)$ and $\sigma(X_k)$ and the 2-α-mixing coefficients are defined as

$$\alpha_X^{(2)}(k) = \sup_{B \in \sigma(X_0); C \in \sigma(X_k)} |P(B \cap C) - P(B)P(C)|.$$

When $\lim_{k \to \infty} \alpha_X^{(2)}(k) = 0$ we say that X is 2-α-mixing. If the mixing coefficients have a geometrical decay, then the corresponding process is called geometrically mixing (GSM).

3.1. Convergence of the mean function estimator a_v.

We first prove the pointwise convergence, i.e. for a fixed $t \in [0, 1]$, using the additional Assumptions 3.1. A uniform convergence is obtained by assuming the last two conditions of Assumptions 3.1 to hold uniformly on $[0, 1]$. See Appendix B for proofs together with explicit constants (depending on v) for the convergence rate.

Assumptions 3.1. Assume that:

i. V admits a probability density function f and for each $s \neq t$, (V_s, V_t) has a density f_{V_s, V_t} such that $\sup_{|s-t|>1} \|G_{s,t}\|_\infty < \infty$ where $G_{s,t} = f_{V_s, V_t} - f \otimes f$.

ii. Both $(Z_k, k \in \mathbb{Z})$ and $(V_k, k \in \mathbb{Z})$ are strong mixing processes with geometrically decaying coefficients $\alpha^{(2)}(k) = \beta_0 e^{-\beta_1 k}$ for some $\beta_0, \beta 1 > 0$ and $k \geq 1$.

iii. $\|Z_k\|_H = M_Z < \infty$, $\forall k$.

iv. The kernel K is a bounded symmetric density satisfying

    1. $\lim_{v \to \infty} \|v\|_{\mathbb{R}^d}^d K(v) = 0$,
    2. $\int_{\mathbb{R}^d} \|v\|_{\mathbb{R}^d}^3 K(v) dv < \infty$,
    3. $\int_{\mathbb{R}^d} |v_i||v_j| K(v) dv < \infty$ for $i, j = 1, \ldots, d$.

v. The maps $v \mapsto f(v)$ and $v \mapsto g_v(t), t \in [0, 1]$ belongs to $C_d^2(b)$ the space of twice continuously differentiable functions z defined on $\mathbb{R}^d$ and such that

$$\left\| \frac{\partial^2 z}{\partial v_i \partial v_j} \right\|_\infty \leq b.$$

vi. $\mathbb{E}^v[Z_0^2(t)|V]f(v), t \in [0, 1]$ is strictly positive, continuous and bounded at $v \in \mathbb{R}^d$.

Proposition 3.2. *Under Assumptions 2.1 and 3.1, for a bandwidth verifying* $h_{a,n} = c_n \left(\frac{\ln n}{n}\right)^{1/(d+4)}$, $c_n \to c > 0$, *when* $n \to \infty$, *we have*

1.

$$\widehat{f}_n(v) - f(v) = \mathcal{O}\left(\left(\frac{\ln n}{n}\right)^{\frac{2}{4+d}}\right) \qquad a.s., \tag{13}$$

2.

$$\widehat{a}_{v,n}(t) - a_v(t) = \mathcal{O}\left(\left(\frac{\ln n}{n}\right)^{\frac{2}{4+d}}\right) \qquad a.s.$$

Let us comment the assumptions for this result. The density condition 3.1(**i**) may be droped if one uses a more general framework like in Dabo-Niang and Rhomari [14] where no density assumption is done and the observations are independent. However, similar results for dependent data are not available yet. The hypothesis concerning the decay of the mixing coefficients allows us to control the variance of the estimators. We impose some weak conditions on the kernel K that are usual in nonparametric estimation. All symmetric kernels defined over a compact support verify the hypothesis, but also more general ones like the Gaussian kernel. Conditions **v** and **vi** are used to control the bias terms of the estimators that is purely analytical.

The convergence rates obtained in Proposition 3.2 are the usual ones. They rapidly degrade with the raise of the dimension of $\mathbb{R}^d$, the space where V lives, as the consequence of the *curse of dimensionality*. In one hand, the first result is well know on the estimation of a multidimensional density functions, even for dependent data. We include it for sake of comprehension. Note that only the observations of $V_1, \ldots, V_n$ are used to estimate $f(v)$ the density of V at $v \in \mathbb{R}^d$. This result is true for each $t \in [0,1]$. On the other hand, the consistency of $\widehat{a}_{v,n}(t)$ is only valid for some fixed value $t \in [0,1]$. However, we can obtain a version of this result that holds true uniformly on $[0,1]$ (conditionally on V).

Proposition 3.3. *Under Assumptions 2.1, 3.1(**i-iv**) and if 3.1(**v-vi**) hold true for all $t \in [0,1]$, a bandwidth verifying $h_{a,n} = c_n \left(\frac{\ln n}{n}\right)^{1/(d+4)}$, $c_n \to c > 0$, when $n \to \infty$, yields*

$$\|\widehat{a}_n(v,.) - a(v,.)\|_H = \mathcal{O}\left(\left(\frac{\ln n}{n}\right)^{\frac{2}{4+d}}\right) \qquad a.s.$$

3.2. Convergence of $\widehat{\Gamma}_{v,n}$ and $\widehat{\Delta}_{v,n}$.

Similarly to the convergence of the conditional mean function, we first prove the pointwise convergence of $\widehat{\gamma}_{v,n}(s,t)$ and $\widehat{\delta}_{v,n}(s,t)$, and then extend the result to the uniform convergence of these kernels over $[0,1]^2$. Then, the consistency of the operators follows. In addition, we obtain the consistency for the estimators of the spectral elements of Γ_v.

Proposition 3.4. *Under Assumptions 2.1 and 3.1, and if $\mathbb{E}[\|\|Z\|_H^4|V] < \infty$ then for a bandwidth verifying $h_{\gamma,n} = c_n \left(\frac{\ln n}{n}\right)^{1/(d+4)}$, $c_n \to c > 0$, when $n \to \infty$, we have*

$$\widehat{\gamma}_{v,n} - \gamma_v = \mathcal{O}\left(\left(\frac{\ln n}{n}\right)^{\frac{2}{4+d}}\right) \qquad a.s..$$

Again, the result is valid uniformly for $(t,s) \in [0,1]^2$. Through the equivalence between Hilbert-Schmidt norm and the integral operator norm (on $L_2([0,1]^2)$) one has,

$$\|\widehat{\Gamma}_{v,n} - \Gamma_v\|_{\mathcal{K}_2} = \|\widehat{\gamma}_{v,n}(.,.) - \gamma_v(.,.)\|_{L^2([0,1]^2)}^2$$

$$= \int_0^1 \int_0^1 (\widehat{\gamma}_{v,n}(s,t) - \gamma_v(s,t))^2 ds dt$$

9

and thus the strong consistency of $\widehat{\Gamma}_{v,n}$ follows.

Proposition 3.5. *Under Assumptions 2.1, 3.1(**i-iv**) and if 3.1(**v-vi**) hold true for all $t \in [0,1]$, and $\mathbb{E}[\|Z\|_H^4|V] < \infty$, then a bandwidth verifying $h_{\gamma,n} = c_n \left(\frac{\ln n}{n}\right)^{1/(d+4)}$, with $c_n \to c > 0$, when $n \to \infty$, yields*

$$\|\widehat{\Gamma}_{v,n} - \Gamma_v\|_{\mathcal{K}_2} = \mathcal{O}\left(\left(\frac{\ln n}{n}\right)^{\frac{2}{4+d}}\right) \qquad a.s..$$

Now, one may use the consistency properties of the empirical eigenvalues $\widehat{\lambda}_{v,j,n}$ as estimators of the true ones $\lambda_{v,j}, j \geq 1$, obtained by Bosq [15] in the dependent case. Also a result concerning the convergence of the empirical conditional eigenfunctions $\widehat{e}_{v,j,n}$ is provided. See Mas and Menneteau [16] for a general transfer approach of limit theorem properties and modes of convergence from the estimator of a covariance operator to the estimators of its eigenvalues.

Corollary 3.6. *Under the conditions of Proposition 3.5, we have*

1.

$$\sup_{j \geq 1} |\widehat{\lambda}_{v,j,n} - \lambda_{v,j}| = \mathcal{O}\left(\left(\frac{\ln n}{n}\right)^{\frac{2}{4+d}}\right) \qquad a.s.$$

2.

$$\|\widehat{e}'_{v,j,n} - e_{v,j}\|_H = \xi_{v,j} \mathcal{O}\left(\left(\frac{\ln n}{n}\right)^{\frac{2}{4+d}}\right) \qquad a.s.$$

where $\widehat{e}'_{v,j,n} = <\widehat{e}_{v,j,n}, e_{v,j)} >_H \widehat{e}_{v,j,n}$ and $\xi_{v,1} = 2\sqrt{2}/(\lambda_{v,1}-\lambda_{v,2})$, $\xi_{v,j} = 2\sqrt{2}/\min(\lambda_{v,j-1}-\lambda_{v,j}, \lambda_{v,j}-\lambda_{v,j-1})$ for $j \geq 2$.

Note that the conditional eigenfunctions are estimated up to their sign. This causes problems both in practice and in theory. The estimated object is the eigen-space generated by the associated eigenfunction and not its direction.

Finally, using similar arguments we obtain the convergence of the conditional cross-covariance operator.

Proposition 3.7. *Under Assumptions 2.1, 3.1(**i-iv**) and if 3.1(**v-vi**) hold true for all $t \in [0,1]$, and $\mathbb{E}[\|Z\|_H^4|V] < \infty$, then for a bandwidth verifying $h_n = c_n \left(\frac{\ln n}{n}\right)^{1/(d+4)}$ with $c_n \to c > 0$, when $n \to \infty$, yields*

$$\|\widehat{\Delta}_{v,n} - \Delta_v\|_{\mathcal{K}_2} = \mathcal{O}\left(\left(\frac{\ln n}{n}\right)^{\frac{2}{4+d}}\right) \qquad a.s..$$

3.3. Convergence of the predictors

The two proposed classes of estimators for ρ^* can be use to predict Z_{n+1} by applying them to the last observed function Z_n. However, since Z_n was used on the construction of the estimator and the process has a memory length of 1, a better approach is to study the prediction error on the next element of the sequence. We introduce a final set of assumptions needed to shown the convergence in probability that we denote $\xrightarrow{P}$.

Assumptions 3.8.

1. $\mathbb{E}[\|Z\|_H^4|V] < \infty$.

2. Γ_v is one-to-one.
3. $\mathbb{P}(\liminf \mathcal{E}_n) = 1$, where $\mathcal{E}_n = \{\omega \in \Omega : \dim(\mathrm{Rg}(P_v^{k_n} \widehat{\Gamma}_{v,n} P_v^{k_n})) = k_n\}$, with $\mathrm{Rg}(A)$ denoting the range of the operator A.
4. $n\lambda_{k_n}^4(v) \to \infty$ and $(1/n)\sum_{j=1}^{k_n} \xi_k(v)/\lambda_k^2(v) \to 0$, as $n \to \infty$.

A strong finite fourth conditional moment of Z was used for the definition of Γ_v and Δ_v. The second condition in 3.8 is necessary to uniquely define the conditional autoregression operator ρ_v. The third one is necessary to guarantee that the random operator $P_v^{k_n} \widehat{\Gamma}_{v,n} P_v^{k_n}$ is almost sure invertible. Controlling the decay of the eigenvalues of the conditional covariance operator is used for the consistency of the projection class operator (see Corollary 3.6 for the definition of ξ). Alternatively, one may set $\Lambda_v(k) = \lambda_k(v)$ where $\Lambda_v : \mathbb{R} \to \mathbb{R}$ is a convex function (see Mas [17]).

Theorem 3.9. *If Assumptions 2.1, 3.1 hold true $\forall t \in [0,1]$ and 3.8, if $\lambda_k(v) = c_0 c_1^k, c_0 > 0, c_1 \in (0,1)$ and if $k_n = o(\ln n)$ as $n \to \infty$, then*

$$\|\widehat{\rho}_{v,n,k_n}^*(Z_{n+1}) - \rho^*(Z_{n+1})\|_H \xrightarrow{P} 0$$

Theorem 3.10. *If Assumptions 2.1, 3.1 hold true $\forall t \in [0,1]$ and 3.8(**i-ii**), and if $b_n \to 0, b_n^{p+2}\sqrt{n} \to \infty$ for some $p \geq 0$ as $n \to \infty$, then*

$$\|\widehat{\rho}_{v,n,p,\alpha}^*(Z_{n+1}) - \rho^*(Z_{n+1})\|_H \xrightarrow{P} 0$$

4. Empirical study

We apply the CARH process model to predict the electricity daily load curve for the french producer EDF (*Électricité de France*). Our aim is to introduce the temperature information as an exogenous covariate on a functional prediction model using CARH processes. The electricity demand is highly sensitive to meteorological conditions. In particular, changes in temperature during winter have a high impact on the French national demand. This relationship is not linear and depends on the hour of the day, the day of the week and the month of the cold season. Moreover, it is unknown in which way the temperature should be coded in order to extract the relevant information for a prediction model. More details on this dataset are given in Antoniadis et al. [18].

We compare in terms of prediction error, the AutoRegressive Hilbertian model (ARH) and the Conditional AutoRegressive model (CARH). The data we use are the electricity load for the first three months of 2009 (where the load is very sensitive to temperature changes) recorded at a 30 minutes resolution and an estimate of the national temperature computed by EDF recorded each hour. The function-valued process Z is the sequence of daily loads of the national grid. As the calendar has a very important effect on the electricity demand, we work only with one day-type, namely the weekdays from Mondays to Friday excluding holidays. The covariate V is constructed as an univariate summary of the daily temperature profile. Concretely, we compute the variation coefficient of the temperature records for each day. The total number of observations is 41, where we use the first 33 (approximately 80%) for calibration of the model and the last 8 to measure the prediction quality of the calibrated model.

For both models we use projection type estimators (see Equation (11)). Using the calibration dataset we estimate the parameter k_n, that is the dimension of the projection space for both models. In addition, we estimate the bandwidth parameters for the CARH model. The results of the parameters' estimation is summarised in Table 1. We also compute the in-sample estimation

	ARH	CARH
Dimension (k_n)	2	5
Estimation error	1616	929
Prediction error	1522	1265

Table 1: Values of the estimated parameters for both ARH and CARH prediction models. The estimated set of bandwidths is $h_a = 1.21 \times 10^{-1}, h_\gamma = \times 10^{-4}, h_\delta = 3.95 \times 10^{-1}$.

error as the prediction error obtained using the set of parameters that minimise the root mean square error (RMSE) on the training dataset. The CARH model seems to obtain a better fit on the calibration set since it presents a smaller estimation error.
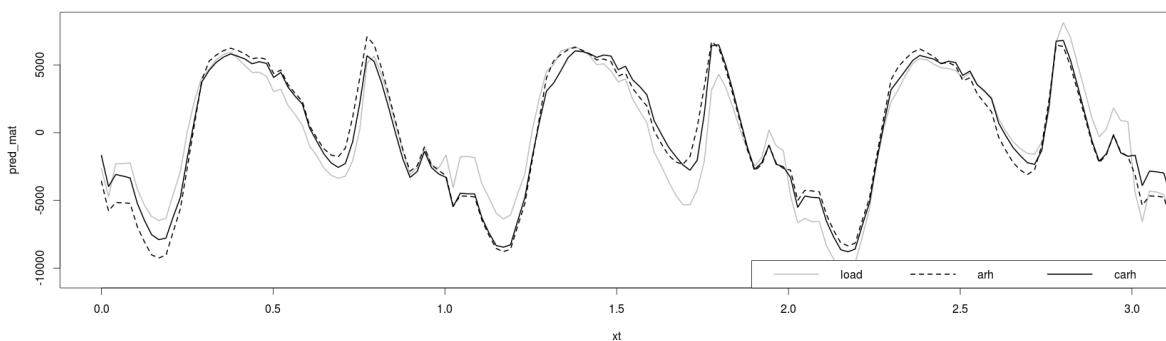

Figure 1: Three days of the electricty demand (solid gray) with the one-day ahead predictions using ARH (dashed) and CARH (solid black) models.

In order to estimate the prediction error we use the test dataset. We compute the error as the RMSE. Again the CARH model presents a smaller error than the ARH model. On Figure 1 we present three days of the electricity demand as well as their predictions using the ARH and CARH models. The effect of the covariate seems to be expressed locally in some parts of the day. Effectively, it corresponds to the daytime demand which seems to be reasonable because the effect of the temperature on the electricity demand is higher during day hours than night hours.

Appendix A. Linear operators in Hilbert spaces

We recall here some relevant facts about linear operators on Hilbert space (see Kato [11, Chap. 5] for details).

We note H^* the topological dual of H, i.e. the space of bounded linear functionals on H. Thanks to the Riesz representation H^* can be identified with H. We note $\mathcal{L}$ the space of bounded linear operators from H to H equipped with the uniform norm

$$\|\rho\|_{\mathcal{L}} = \sup_{\|z\|_H \leq 1} \|\rho(z)\|_H, \qquad \rho \in \mathcal{L}, z \in H.$$

This space seems to be a too large space, so one usually consider the subspace of compact operators $\mathcal{K}$ that is easier to deal with (see Mas [17]). For instance, if the operator ρ is compact then it admits a unique spectral decomposition, i.e. for two bases $(\phi_j)_{j \in \mathbb{N}}$ and $(\psi_j)_{j \in \mathbb{N}}$ and a sequence of

numbers $(\lambda_j)_{j \in \mathbb{N}}$ that we can choose to be non-negative (choosing the sign of ψ_j) we have

$$\rho = \sum_{j \in \mathbb{N}} \lambda_j \psi_j \otimes \phi_j,$$

where we use the tensor product notation $(u \otimes v)(z) = <u, x>_H v$ for any elements $z, u, v \in H$. We say that a operator ρ is self-adjoint if $<\rho u, v>_H = <u, \rho v>_H$ for all $u, v \in H$. If ρ is symmetric the decomposition becomes $\rho = \sum_{j \in \mathbb{N}} \lambda_j \phi_j \otimes \phi_j$ with eigen-elements $(\lambda_j, \phi_j)_{j \in \mathbb{N}}$. If ρ is not self-adjoint, we call ρ^* its adjoint. Finally we say that ρ is positive-definite if it satisfies $<\rho z, z>_H \geq 0$ for all $z \in H$. Two subspaces of $\mathcal{K}$ will be of our interest: the space of Hilbert-Schmidt operators $\mathcal{K}_2$ and the space of trace class (or nuclear) operators $\mathcal{K}_1$ defined respectively as

$$\mathcal{K}_2 = \{A \in \mathcal{K} : \sum_{j \in \mathbb{N}} \lambda_j^2 < \infty\}, \qquad \mathcal{K}_1 = \{A \in \mathcal{K} : \sum_{j \in \mathbb{N}} |\lambda_j| < \infty\}.$$

The Hilbert-Schmidt operators form a separable Hilbert space with inner product $<\rho, \tau>_{\mathcal{K}_2} = \sum_{j \in \mathbb{N}} <\rho \psi_j, \tau \psi_j>$ with $(\psi_j)_j$ an orthonormal basis and $\rho, \tau \in \mathcal{K}_2$ (the product does not depends on the choice of the basis, see Kato [11, p. 262]). The associated norm yields from $\|\rho\|_{\mathcal{K}_2}^2 = \sum_{j \in \mathbb{N}} \|\rho \psi_j\|_H^2 = \sum_{j \in \mathbb{N}} \lambda_j^2$. On the other hand the space of trace-class operator endowed with the norm $\|.\|_{\mathcal{K}_1}$ defined as $\|\rho\|_{\mathcal{K}_1} = \sum_j |\lambda_j|$ is a separable Banach space. Finally, from the continuity of the inclusions $\mathcal{K}_1 \subset \mathcal{K}_2 \subset \mathcal{K} \subset \mathcal{L}$ we have that

$$\|.\|_{\mathcal{K}_1} \geq \|.\|_{\mathcal{K}_2} \geq \|.\|_{\mathcal{L}}.$$

Appendix B. Sketch of proofs.

Proof of Theorem 2.2.

We mimic the proof of Theorem 1 in Guillas [10]. To prove the existence, Let

$$\eta_m^{m'} = \mathbb{E} \left\| \sum_{j=m}^{m'} \left(\prod_{p=0}^{j-1} \rho_{V_{n-p}} \right) (\epsilon_{n-j}) \right\|_H^2$$

$$= \sum_{j=m}^{m'} \mathbb{E} \left\| \left(\prod_{p=0}^{j-1} \rho_{V_{n-p}} \right) (\epsilon_{n-j}) \right\|_H^2$$

$$\leq \sum_{j=m}^{m'} \mathbb{E} \left[\left\| \prod_{p=0}^{j-1} \rho_{V_{n-p}} \right\|_{\mathcal{L}}^2 \|\epsilon_{n-j}\|_H^2 \right]$$

$$\leq \sum_{j=m}^{m'} \mathbb{E} \left[\prod_{p=0}^{j-1} \|\rho_{V_{n-p}}\|_{\mathcal{L}}^2 \right] \underbrace{\mathbb{E} \|\epsilon_{n-j}\|_H^2}_{=\sigma^2}$$

where we used the independence between V and ϵ gives

$$\mathbb{E} \left\langle \left(\prod_{p=0}^{j-1} \rho_{V_{n-p}} \right) (\epsilon_{n-j}), \left(\prod_{p=0}^{j'-1} \rho_{V_{n-p}} \right) (\epsilon_{n-j'}) \right\rangle = 0, \qquad \text{for } j \neq j'.$$

13

Finally, we obtain

$$\eta_m^{m'} \leq \sigma^2 \, \mathbb{E} \left[\prod_{p=0}^{j-1} \| \rho_{V_{n-p}} \|_{\mathcal{L}}^2 \right] \leq \sigma M_\rho^{2j}.$$

We have that the upper bound is the general term of a convergent series. For m, m' tending to infinity, $\eta_m^{m'}$ tend to zero and the Cauchy criterion gives the mean square convergence of the solution.

Now, consider the stationary process $W_n = a + \sum_{j=0}^{\infty} \left(\prod_{p=0}^{j-1} \rho_{V_{n-p}} \right) (\epsilon_{n-j})$. From the almost surely boundedness of ρ_{V_n} we have that it is indeed a solution of the CARH process:

$$(W_n - a) - \rho_{V_n}(W_{n-1} - a) =$$

$$= \sum_{j=0}^{\infty} \left(\prod_{p=0}^{j-1} \rho_{V_{n-p}} \right) \epsilon_{n-j} - \sum_{j=0}^{\infty} \rho_{V_n} \left(\prod_{p=0}^{j-1} \rho_{V_{n-1-p}} \right) \epsilon_{n-1-j}$$

$$= \sum_{j=0}^{\infty} \left(\prod_{p=0}^{j-1} \rho_{V_{n-p}} \right) \epsilon_{n-j} - \sum_{j=0}^{\infty} \left(\prod_{p=0}^{j} \rho_{V_{n-p}} \right) \epsilon_{n-1-j}$$

$$= \sum_{j=0}^{\infty} \left(\prod_{p=0}^{j-1} \rho_{V_{n-p}} \right) \epsilon_{n-j} - \sum_{j'=1}^{\infty} \left(\prod_{p=0}^{j'-1} \rho_{V_{n-p}} \right) \epsilon_{n-j'}$$

$$= \epsilon_n.$$

Proof of Proposition 3.2

The proof is based on the classical decomposition in terms of bias and variance of the estimators. The bias term is purely analytical. The variance term is composed by the variance and covariance of the estimator's terms. The dependency of the data is controlled by means of the following exponential inequality (a proof can be founded in Bosq and Blanke [13, p. 140]).

Lemma Appendix B.1. *Let $W = (W_t)$ be a zero-mean real valued stationary process with $\sup_{1 \leq t \leq n} \|W_t\|_\infty = M < \infty$, $(M > 0)$. Then for $q \in [1, n/2]$, $\kappa > 0$, $\epsilon > 0$, $p = n/(2q)$,*

$$\mathbb{P} \left(\left| \sum_{i=1}^{n} W_i \right| > n\epsilon \right) < \frac{8M}{\epsilon \kappa} (1 + \kappa) \alpha_X \left(\left[\frac{n}{2q} \right] \right) +$$

$$4 \exp \left(- \frac{n^2 \epsilon^2 / q}{8(1 + \kappa) \sigma(q) + \frac{2M}{3}(1 + \kappa) n^2 q^{-2} \epsilon} \right), \quad \text{(B.1)}$$

with $\sigma(q)$ an intricate quantity involving the pairwise covariances of W. We will only need a bound of $\sigma(q)$ that in the stationary case turns out to be

$$\sigma(q) < ([p] + 2)(\mathrm{Var}(W_0) + 2 \sum_{l=1}^{[p]+1} |\mathrm{Cov}(W_0, W_l)|). \quad \text{(B.2)}$$

Proof of 1. One has

$$\mathbb{E} \, \widehat{f}_n(v) - f(v) = \int_{\mathbb{R}^d} K(u)(f(v - h_n u) - f(v)) du$$

14

Using Taylor formula and the symmetry of K one gets

$$\mathbb{E}\,\widehat{f}_n(v) - f(v) = \frac{h_n^2}{2}\int_{\mathbb{R}^d} K(u)\left(\sum_{i,j=1}^{d} u_i u_j \frac{\partial^2 f}{\partial v_i \partial v_j}(v - \theta h_n u)\right)du$$

where $0 < \theta < 1$. Finally, Lebesgue dominated convergence theorem gives

$$h_n^{-2}|\,\mathbb{E}\,\widehat{f}_n(v) - f(v)| \to b_2(v) = 1/2\left(\sum_{i,j=1}^{d} \frac{\partial^2 f}{\partial v_i \partial v_j}(v)\int_{\mathbb{R}^d} u_i u_j K(u)du\right) \qquad \text{(B.3)}$$

We use (B.1) to deal with the variance term $\widehat{f}_n(v) - \mathbb{E}\,\widehat{f}_n(v)$. Define $W_i = K_h(v - V_i) - \mathbb{E}\,K_h(v - V_i)$, with $K_h(.) = K(./h)$. Then, $M = 2h_n^{-d}\|K\|_\infty$. Let us choose $q_n = \frac{n}{2p_0 \ln n}$ for some $p_0 > 0$. Which yields on a logarithmic order for $p_n = \frac{1}{p_0 \ln n}$. This choices and the boundeness of f and $G_{s,t}$ entail on B.2,

$$\sigma(q_n) < (p_n + 2)\,\mathrm{Var}(K_h(v - V_1)) + (p_n + 2)^2 \sup_{|s-t|>1} \|G_{s,t}\|_\infty$$

$$< p_n h_n^{-d}\|K\|_2^2 f(v)(1 + o(1)).$$

Now take $\epsilon = \eta\sqrt{\frac{\ln n}{nh_n^d}}$, for some $\eta > 0$, then

$$\mathbb{P}\left(n^{-1}\left|\sum_{i=1}^{n} W_i\right| > \eta\sqrt{\frac{\ln n}{nh_n^d}}\right) < \frac{8\beta_0 c^{d/2}}{\eta\kappa}(1 + \kappa)\|K\|_\infty \frac{n^{\frac{2+d}{4+d}} - \beta_1 p_0}{(\ln n)^{\frac{2+d}{4+d}}}$$

$$+ 4\exp\left(-\frac{\eta^2 \ln n}{4(1 + \kappa)^2\|K\|_2^2 f(v)(1 + o(1))}\right)$$

If we take $\eta > 2(1 + \kappa)\|K\|_2\sqrt{f(v)}$ and $p_0 > 2\beta$, then where both terms are $o(n^{-\lambda})$, for some $\lambda > 0$, in which case

$$\sum_n \mathbb{P}\left\{\left(\frac{n}{\ln n}\right)^{\frac{2}{4+d}}\left|n^{-1}\sum_{i=1}^{n} W_i\right| > \eta c_n^{-d/2}\right\} < \infty.$$

So Borel-Cantelli lemma implies $\limsup_{n\to+\infty}\left(\frac{n}{\ln n}\right)^{\frac{2}{4+d}}|\widehat{f}_n(v) - \mathbb{E}\,\widehat{f}_n(v)| \leq 2c_n^{-d/2}(1+\kappa)\|K\|_2\sqrt{f(v)}$ almost surely for all $\kappa > 0$. We have finally

$$\limsup_{n\to+\infty}\left(\frac{n}{\ln n}\right)^{\frac{2}{4+d}}|\widehat{f}_n(v) - f(v)| \leq 2c_n^{-d/2}\|K\|_2\sqrt{f(v)} + c^2|b_2(v)|,$$

which gives (13).

Proof of 2. We use the following decomposition, omitting the argument v,

$$\widehat{a}_n - a = \frac{\widehat{g}_{v,n} - a\widehat{f}_n}{\widehat{f}_n}.$$

From (13) we have for the denominator that $\widehat{f}_n \to f(x) \neq 0$ almost surely. We work out the numerator through the following decomposition between variance and bias terms. Let $\psi_n = (n/\ln n)^{2/(4+d)}$, then one has

$$\psi_n|\widehat{g}_{v,n} - a\widehat{f}_n| \leq \underbrace{\psi_n|\widehat{g}_{v,n} - a\widehat{f}_n - \mathbb{E}(\widehat{g}_{v,n} - a\widehat{f}_n)|}_{:=A_n} + \underbrace{\psi_n|\,\mathbb{E}(\widehat{g}_{v,n} - a\widehat{f}_n)|}_{:=B_n}.$$

We first study A_n using as before the exponential type inequality (B.1) with the redefined random variables
$$W_i = K_h(v - V_i)(Y_i - a_v) - \mathbb{E}\left(K_h(v - V_i)(Y - a_v)\right)$$
with the precedent choices of q_n and p_n. First, one has $|W_i| \leq 2h_n^{-d}\|K\|_\infty(1 + o(1))$. Next, using Bochner lemma (Bosq and Blanke [13, p. 135]) we obtain

$$h_n^d \operatorname{Var}(W_1) \leq h_n^{-d} \mathbb{E}[K_h^2(v - V_1)(Y_1 - a_v)^2] \to f(v)\|K\|_2^2\Sigma(v)$$

where $\Sigma(v) = (\mathbb{E}^v[Y_0^2|V] - a_v)$ is the conditional variance parameter. The logarithmic order of p_n and the control on F gives $\sigma^2(q_n) \leq p_n h_n^{-d} f(v)2\Sigma(v)\|K\|_2^2(1 + o(1))$. As before, taking $p_0 > 2/\beta_1$ and for a large enough η, Borel-Cantelli lemma entails

$$\limsup_{n\to\infty} A_n \leq 2c^{-d/2}\sqrt{\Sigma(v)f(v)} \qquad \text{a.s.}$$

For the bias term we write

$$\mathbb{E}(\widehat{g}_{v,n}(v) - a(v)\widehat{f}_n(v)) = h_n^{-d} \int_{\mathbb{R}^d} K_{h_n}(v - t)(g(t) - f(t)a(v))dt.$$

Then, we use the Taylor formula to expand $g(t) - f(t)a(v)$ and Assumptions 3.1(**iii-iv**) to obtain

$$\psi_n|B_n| \to b_a(v) = \frac{1}{2}\left|\sum_{i,j=1}^d \left\{\frac{\partial^2 g}{\partial v_i\partial v_j}(v) - a(v)\frac{\partial^2 f}{\partial v_i\partial v_j}(v)\right\}\int u_i u_j K(u)du\right|$$

Finally, putting all the elements together one obtains,

$$\limsup_{n\to\infty}\left(\frac{n}{\ln n}\right)^{\frac{2}{4+d}}|\widehat{a}_n(v) - a(v)| \leq 2c^{-d/2}\|K\|_2\sqrt{f(v)\Sigma(v)} + c^2\frac{|b_a(v)|}{f(v)} \qquad (\text{B.4})$$

from with the result is derived.

Proof of Proposition 3.3.

The only terms on equation B.4 that depends on the value fixed for t are the conditional variance parameter Σ and the bias b_a. With the new hypothesis holding uniformly, for each $v \in \mathbb{R}^d$, $\Sigma(v,t)$ and $b_a(v,t)$ are bounded uniformly on $[0,1]$. Then, recalling that

$$\|\widehat{a}_n(v,.) - a(v,.)\|_H^2 = \int_0^1 (\widehat{a}_n(v,t) - a(v,t))^2 dt,$$

we obtain the derived result.

Proof of Proposition 3.4.

The proof follows the same lines that those used to show Proposition 3.2(2). In particular,

$$\widehat{r}_n - r = \frac{\widehat{g}_{v,n} - r\widehat{f}_n}{\widehat{f}_n}.$$

gives the decomposition between variance and bias terms,

$$\psi_n|\widehat{g}_{v,n} - r\widehat{f}_n| \leq \underbrace{\psi_n|\widehat{g}_{v,n} - r\widehat{f}_n - \mathbb{E}(\widehat{g}_{v,n} - r\widehat{f}_n)|}_{:=A_n} + \underbrace{\psi_n|\mathbb{E}(\widehat{g}_{v,n} - r\widehat{f}_n)|}_{:=B_n}.$$

16

Which yields on
$$\limsup_{n\to\infty} A_n \leq 2c^{-d/2}\sqrt{\Sigma(v)f(v)} \qquad \text{a.s.}$$
where, by the redefinition of Y, $\Sigma(v) = \mathbb{E}^v[(Z_0(s)Z_0(t))^2|V] - r(v,s,t)$.

Again using Taylor formula to expand $g(t) - f(t)r(v)$ and the precedent Assumptions we obtain

$$\psi_n|B_n| \to b_r(v) = \frac{1}{2}\left|\sum_{i,j=1}^{d}\left\{\frac{\partial^2 g}{\partial v_i \partial v_j}(v) - r(v)\frac{\partial^2 f}{\partial v_i \partial v_j}(v)\right\}\int u_i u_j K(u)du\right|$$

Finally, resembling the terms we get the equivalent of Equation (B.4) with the redefined Σ and the bias b_r, from with the result is derived.

Proof of Proposition 3.5.

First, consider the following decomposition
$$\widehat{\Gamma}_{v,n} = \widehat{R}_n(v) - \widetilde{a}_n(v) \otimes \widehat{a}_n(v) - \widehat{a}_n(v) \otimes \widetilde{a}_n(v) + \widehat{a}_n(v) \otimes \widehat{a}_n(v),$$
where $\widehat{R}_n(v) = \sum_{i=1}^{n} w_{n,i}(v,h_\gamma)Z_i \otimes Z_i$ is the empirical counterpart of the second order moment operator $R(v) = \mathbb{E}^v[Z_0 \otimes Z_0|V]$, and $\widetilde{a}_n(v) = \sum_{i=1}^{n} w_{n,i}(v,h_\gamma)Z_i$. Second, we obtain that

$$\Gamma_v - \widehat{\Gamma}_{v,n} = R(v) - \widehat{R}_n(v) - a(v) \otimes a(v) + \widetilde{a}_n(v) \otimes \widehat{a}_n(v) + \widehat{a}_n(v) \otimes \widetilde{a}_n(v) - \widehat{a}_n(v) \otimes \widehat{a}_n(v).$$

Hence, we can control the estimation error regrouping the terms of the above decomposition (we drop the argument v),

$$\|\Gamma_v - \widehat{\Gamma}_{v,n}\|_{\mathcal{K}_2} = \|R - \widehat{R}_n\|_{\mathcal{K}_2} + \|\widetilde{a}_n \otimes \widehat{a}_n - a \otimes a\|_{\mathcal{K}_2} + \|\widehat{a}_n \otimes (\widetilde{a}_n - \widehat{a}_n)\|_{\mathcal{K}_2}. \tag{B.5}$$

From Propositions 3.4 and 3.3 it follows that

$$\|R - \widehat{R}_n\|_{\mathcal{K}_2} = \mathcal{O}\left(\left(\frac{n}{\ln n}\right)^{\frac{2}{4+d}}\right) \qquad \text{a.s.}$$

The second term of the left hand side of equation (B.5) is equal to

$$\|\widetilde{a}_n \otimes (\widehat{a}_n - a) + (\widetilde{a}_n - a) \otimes a\|_{\mathcal{K}_2} \leq \|\widetilde{a}_n\|_H\|\widehat{a}_n - a\|_{\mathcal{K}_2} + \|\widetilde{a}_n - a\|_{\mathcal{K}_2}\|a\|_H.$$

Since both $\|a\|_H$ and $\|\widetilde{a}_n\|_H$ are bounded and using Proposition 3.3 successively for $\widetilde{a}_n$ and $\widehat{a}_n$ with their respective sequences of bandwidths $h_{\gamma,n}$ and $h_{a,n}$, we obtain that

$$\|\widetilde{a}_n \otimes \widehat{a}_n - a \otimes a\|_{\mathcal{K}_2} = \mathcal{O}\left(\left(\frac{n}{\ln n}\right)^{\frac{2}{4+d}}\right) \qquad \text{a.s.}$$

With a similar reasoning, the same kind of result is obtained for the third term in (B.5). Putting the result for the three terms together conclude the proof.

Proof of Corollary 3.6.

First item is a direct consequence of the following property on eigenvalues of compact linear operators Bosq [15, p. 104],

$$\sup_{j\geq 1}|\lambda_j(v) - \hat{\lambda}_{j,n}(v)| \leq \|\Gamma_v - \widehat{\Gamma}_{v,n}\|_{\mathcal{L}},$$

and the asymptotic result obtained for $\|\Gamma_v - \widehat{\Gamma}_{v,n}\|_{\mathcal{K}_2}$.

For the second item, Bosq (2000, Lemma 4.3) shows that, for each $j \geq 1$,

$$\|e_j(v) - e'_{j,n}(v)\|_H \leq \xi_j\|\Gamma_v - \widehat{\Gamma}_{v,n}\|_{\mathcal{L}}.$$

Again, the rates of convergence follows from Proposition 3.5.

17

Proof of Proposition 3.7.

The proof follows the same guidelines that those of Proposition 3.5, replacing $\hat{R}(v)$ and $R(v)$ by $\hat{R}_1(v) = \sum_{i=1}^{n-1} w_{n,i}(v, h_\gamma) Z_i(s) Z_{i+1}(t)$ and $R_1(v) = \mathbb{E}^v[Z_0(s)Z_1(t)|V]$ respectively. Then, a decomposition like B.5 and the same kind of observations done for that proof entails the result.

Proof of Theorem 3.9.

The proof follows along the same lines of Proposition 4.6 in Bosq [1] by using Propositions 3.3, 3.5, 3.7 and Corollary 3.6.

Proof of Theorem 3.10.

The proof follows along the same lines of Proposition 3 in [12, Chapter 3] by using Propositions 3.3, 3.5 and 3.7.

References

[1] D. Bosq, Modelization, nonparametric estimation and prediction for continuous time processes, in: G. Roussas (Ed.), Nonparametric functional estimation and related topics, NATO ASI Series, 1991, pp. 509–529.

[2] P. Besse, H. Cardot, Approximation spline de la prévision d'un processus fonctionnel autorégressif d'ordre 1, Canadian Journal of Statistics 24 (1996) 467–487.

[3] B. Pumo, Prediction of continuous time processes by c [0, 1]-valued autoregressive process, Statistical Inference for Stochastic Processes 1 (1998) 297–309.

[4] A. Antoniadis, T. Sapatinas, Wavelet methods for continuous-time prediction using Hilbert-valued autoregressive processes, Journal of Multivariate Analysis 87 (2003) 133–158.

[5] V. Kargin, A. Onatski, Curve forecasting by functional autoregression, Journal of Multivariate Analysis 99 (2008) 2508–2526.

[6] A. Mas, B. Pumo, Linear processes for functional data, in: F. Ferraty, Y. Romain (Eds.), The Oxford Handbook of Functional Data Analysis, Oxford Handbooks in Mathematics, Oxford University Press, 2011, pp. 47–71.

[7] B. Pumo, Estimation et prévision de processus autorégressifs fonctionnels, Ph.D. thesis, University of Paris 6, 1992.

[8] A. Mas, B. Pumo, The ARHD process, J. of Statistical Planning and Inference 137 (2007) 538–553.

[9] J. Damon, S. Guillas, The inclusion of exogenous variables in functional autoregressive ozone forecasting, Environmetrics 13 (2002) 759–774.

[10] S. Guillas, Doubly stochastic Hilbertian processes, Journal of Applied Probability 39 (2002) 566–580.

[11] T. Kato, Perturbation Theory for Linear Operators, Springer-Verlag, Berlin, 1976.

[12] A. Mas, Estimation d'opérateurs de corrélation de processus fonctionnels: lois limites, tests, déviations modérées, Ph.D. thesis, Université Paris 6, 2000.

[13] D. Bosq, D. Blanke, Inference and Prediction in Large Dimensions, Wiley series in probability and statistics, John Wiley & Sons, Ltd., 2007.

[14] S. Dabo-Niang, N. Rhomari, Kernel regression estimation in a banach space, Journal of Statistical Planning and Inference 139 (2009) 1421–1434.

[15] D. Bosq, Linear processes in function spaces: Theory and applications, Springer-Verlag, New York, 2000.

[16] A. Mas, L. Menneteau, Perturbation approach applied to the asymptotic study of random operators., Progress in Probability 55 (2003) 127–133.

[17] A. Mas, Weak convergence in the functional autoregressive model, Journal of Multivariate Analysis 98 (2007) 1231–1261.

[18] A. Antoniadis, X. Brossat, J. Cugliari, J.-M. Poggi, Clustering functional data with wavelets, International Journal of Wavelets, Multiresolution and Information Processing accepted for publication (2013).