Statistical Inference for Stochastic Processes (2005) 8: 185–204

Estimation and Simulation of Autoregressive Hilbertian Processes with Exogenous Variables

JULIEN DAMON[1] AND SERGE GUILLAS[2,*]
[1] *Université Paris 6 (Pierre et Marie Curie) and Médiamétrie, FRANCE*
[2] *Université Paris 6 (Pierre et Marie Curie) and Ecole des mines de Douai, FRANCE*

(Received in final form November 14, 2001; Accepted January 30, 2004)

Abstract. We present the autoregressive Hilbertian with exogenous variables model (ARHX) which intends to take into account the dependence structure of random curves viewed as H-valued random variables, where H is a Hilbert space of functions, under the influence of explanatory variables. Limit theorems and consistent estimators are derived from an autoregressive representation. A simulation study illustrates the accuracy of the estimation by making a comparison on forecasts with other functional models.

Key words. autoregressive processes, exogenous variables, functional data, forecasting, simulation, ARHX.

1. Introduction

Autoregressive Hilbertian processes (ARH) and autoregressive Hilbertian processes with exogenous variables (ARHX) can be viewed as extensions of real valued discrete time standard autoregressive processes (AR) and autoregressive processes with exogenous variables (ARX). The popular ARX model already showed better descriptive and forecasting properties for phenomena influenced by explanatory variables than the classical AR model. Several applications were carried out in various fields including pollution assessment [2, 26] or engineering [32, 35]. ARX or autoregressive moving-average with exogenous variables (ARMAX) models have been investigated by [1, 6, 11, 15, 17, 18, 20, 25, 27, 33] among others. Bosq and Shen [10], Chen and Shen [13], Cai and Masry [12] removed the linearity specifications, using nonparametric techniques such as kernel estimation, method of sieves or local polynomial fitting.

As noticed by Bondon [7], if one wants to forecast more than one step ahead, the 'plug-in' method may show some defectiveness. Indeed, replacing the first future values of a time series by predicted ones may affect badly the forecasted ones afterwards because errors are made in the substitution. Moreover, when discrete time series are made of observations of a continuous-time process, it is appealing to use this feature. In that case, one

*Author for correspondence: Current address: University of Chicago, CISES, 5734S. Ellis Avenue, Chicago, IL 60637, USA. Phone: (773) 834 9929; Fax: (773) 834 1045; E-mail: guillas@uchicago.edu

approach is then to make a division of $\mathbb{R}$ into intervals of length δ, and view our discrete time series $(x_i, i \in \mathbb{Z})$ as observations of only a few values of functions $(X_k, k \in \mathbb{Z})$ whose argument is in $[0, \delta]$, in the following sense: $X_k(t) = x_{k\delta+t}$. Predictions are then naturally δ-step ahead ones and a great part of the dynamic is taken into consideration.

Bosq [8] was the first to propose to regard the link between those functions as an autoregressive dependence whose autocorrelation is a linear operator – allowing a possibly nonlinear behavior of the underlying discrete time series – on an appropriate function space, making thus autoregressive Hilbert or Banach valued time series. As a matter of fact, this theory belongs to 'functional data analysis', as presented in [31] or [30] except that independence is not assumed. Nevertheless the estimation of the covariance structure is also a efficient tool. Benyelles and Mourid [3], Mas [21], Merlevède [22], Mourid [23] and Pumo [28,29], worked on estimation procedures, asymptotic normality, estimation of δ, prediction and simulation as well as on extensions of the ARH model. Bosq [9] is an up-to-date reference for a statement on linear processes taking their values in function spaces. Note that Besse and Cardot [4] and Besse et al. [5] introduced the use of smoothing splines in that framework and illustrated the quality of their predictions in comparison with the seasonal autoregressive integrated moving-average (SARIMA) parametric model on real data sets such as traffic or climatic variations. One of their findings is that for some real life forecasting problems, functional time series improves upon SARIMA models. Indeed, using the great bulk of information entails better predictions, especially when the stochastic process holds some functional properties such as some particular smoothness – a Sobolev space is then a natural framework – or a typical variation. Bosq [9] also points out that the aforementioned functional discretization of a continuous-time process with seasonality yields a stationary Hilbertian process, when the discretization is carried out with respect to the period δ.

The simulation of such processes is really a hard task due to the infinite dimensional distributions involved in the model. However, Besse and Cardot [4] and Pumo [28] managed to make simulations by means of the Karhunen–Loève expansion or Brownian motion simulation.

The purpose of this paper is to present the ARHX model. Thus, we generalize both parametric ARX models and nonparametric ARH models. In Section 2, the law of large numbers and the central-limit theorem are discussed and the estimation procedure is shown to be consistent. The key tool is the autoregressive representation of such processes, and the proofs are stated in Appendix A. In Section 3, a general simulation method is provided. It extends [4] and [28] since we embrace the non-Gaussian case and we do not assume that the functional correlation operator is diagonal. We illustrate the efficiency of the estimation procedure by a Monte-Carlo study. A few examples are given, and comparisons are made on forecasts with some other functional procedures

(ARH, functional kernel). Our model presented the best predictive skills on those simulations. Note that Damon and Guillas [14] applied the ARHX modeling to forecast ground level ozone, and showed that on that particular real data set, the ARHX predictions were more accurate than those given by the ARH or the functional kernel.

2. Autoregressive Representation

Let H be a real and separable Hilbert space. Let $\rho, a_1, \dots, a_q$ be bounded linear operators on H. Let $(\varepsilon_n)_{n\in Z}$ be a strong Hilbertian white noise (SWN) i.e. a sequence of i.i.d. H-valued random variables satisfying $E\varepsilon_n = 0$, $0 < E\|\varepsilon_n\|^2 = \sigma^2 < \infty, n \in Z$. We consider the following autoregressive Hilbertian with exogenous variables of order one model, denoted by ARHX(1):

$$X_n = \rho(X_{n-1}) + a_1(Z_{n,1}) + \cdots + a_q(Z_{n,q}) + \varepsilon_n, \quad n \in Z, \tag{1}$$

where $Z_{n,1}, \dots, Z_{n,q}$ are q autoregressive of order one exogenous variables associated respectively with operators $u_1, \dots, u_q$ and strong white noises $(\eta_{n,1}), \dots, (\eta_{n,q})$, i.e. $Z_{n,i} = u_i(Z_{n-1,i}) + \eta_{n,i}$. We suppose that the noises (ε_n), $(\eta_{n,1}), \dots, (\eta_{n,q})$ are independent and that there is a n such that $\|u_i^n\|_{\mathcal{L}} < 1$. This way, for all $i = 1, \dots, q$, the whole processes (ε_n) and $(Z_{n,i})$ are independent since $Z_{n,i}$ can be expressed as a infinite moving average of the $\eta_{n-p,1}$, $p = 0, \dots, \infty$.

Consider the Cartesian product H^{q+1}, separable Hilbert space equipped with the scalar product $\langle (x_1, \dots, x_{q+1}), (y_1, \dots, y_{q+1}) \rangle_{q+1} = \sum \langle x_i, y_i \rangle$. The spaces of bounded linear operators and Hilbert–Schmidt operators on H^{q+1} will be denoted respectively by $\mathcal{L}_{q+1}$ and $\mathcal{S}_{q+1}$.

Let us denote

$$T_n = \begin{pmatrix} X_n \\ Z_{n+1,1} \\ \vdots \\ Z_{n+1,q} \end{pmatrix}, \quad \varepsilon'_n = \begin{pmatrix} \varepsilon_n \\ \eta_{n,1} \\ \vdots \\ \eta_{n,p} \end{pmatrix} \quad \text{and} \quad \rho' = \begin{pmatrix} \rho & a_1 & \cdots & \cdots & a_q \\ 0 & u_1 & 0 & \cdots & 0 \\ 0 & 0 & u_2 & 0 & \vdots \\ \vdots & \vdots & & \ddots & 0 \\ 0 & 0 & 0 & 0 & u_q \end{pmatrix}.$$

When (X_n) is an ARHX(1) defined by Equation (1), then (T_n) is a H^{q+1}-valued ARH(1) process, as we see in the following autoregressive representation

$$T_n = \rho'(T_{n-1}) + \varepsilon'_n, \quad n \in Z. \tag{2}$$

Remark 2.1. A natural extension would be to consider a higher order of autoregression. The model is then written

$$X_n = \rho_1(X_{n-1}) + \cdots + \rho_p(X_{n-p}) + a_1(Z_{n,1}) + \cdots + a_q(Z_{n,q}) + \varepsilon_n, \quad n \in Z. \tag{3}$$

Accordingly, the representation is now built with

$$T_n = \begin{pmatrix} X_n \\ \vdots \\ X_{n-p+1} \\ Z_{n+1,1} \\ \vdots \\ Z_{n+1,q} \end{pmatrix}, \quad \varepsilon'_n = \begin{pmatrix} \varepsilon_n \\ 0 \\ \vdots \\ 0 \\ \eta_{n,1} \\ \vdots \\ \eta_{n,p} \end{pmatrix}, \quad \rho' = \begin{pmatrix} \rho_1 & \cdots & \cdots & \rho_p & a_1 & \cdots & \cdots & a_q \\ I & 0 & \cdots & \cdots & \cdots & \cdots & \cdots & 0 \\ 0 & I & 0 & \cdots & \cdots & \cdots & \cdots & \vdots \\ 0 & \cdots & I & 0 & \cdots & \cdots & \cdots & 0 \\ 0 & \cdots & \cdots & 0 & u_1 & 0 & \cdots & 0 \\ 0 & \cdots & \cdots & \cdots & 0 & u_2 & 0 & \vdots \\ \vdots & \cdots & \cdots & \cdots & \vdots & 0 & \ddots & 0 \\ 0 & \cdots & \cdots & \cdots & 0 & 0 & 0 & u_q \end{pmatrix}.$$

Remark 2.2. We can also consider the case where there are some kind of 'feedback' properties, i.e. when there occurs causality in two directions. For example with $q = 1$, if $Z_n = b(X_n) + u(Z_{n-1}) + \eta_n$, with b a bounded operator on H, we write

$$T_n = \rho'(T_{n-1}) + \varepsilon'_n \quad \text{with} \quad \rho' = \begin{pmatrix} \rho & a \\ b & u \end{pmatrix}.$$

The results presented below do not apply to that case, but the estimation procedure is relatively robust with respect to the hypothesis $b = 0$.

In the rest of the paper, we assume that the following condition (C) holds.

$$\exists j_0, \left\|\rho'^{j_0}\right\|_{\mathcal{L}_{q+1}} < 1. \tag{C}$$

Remark 2.3. (C) may hold with $\|\rho'\| = 1$. For instance, with one exogenous variable we have

$$\rho' = \begin{pmatrix} \rho & a \\ 0 & u \end{pmatrix} \quad \text{and} \quad \rho'^2 = \begin{pmatrix} \rho^2 & \rho a + au \\ 0 & u^2 \end{pmatrix}.$$

So $\left\|\rho'^2\right\|$ may be strictly less than 1 with appropriate ρ, a, u (e.g. $\rho = a = u, \rho^2 = 0$).

Denote by P_i the projection operators $(x_1, \ldots, x_{q+1}) \mapsto x_i$.

PROPOSITION 2.1 Equation (1) has a unique stationary solution given by

$$X_n = \sum_{j=0}^{\infty} (P_1 \rho'^j)(\varepsilon'_{n-j}), \quad n \in \mathbb{Z}.$$

The series converges a.s. and in $L^2_H(\Omega, \mathcal{A}, \mathcal{P})$.

The covariance operator $C^{X,Y}$ of two H-valued r.v. X and Y is

$$C^{X,Y}(x) := E[\langle X, x \rangle Y], \quad x \in H,$$

and C^X stands for $C^{X,X}$. The autocovariance of a stationary process (X_n) is the sequence $(C_h, h \in \mathbb{Z})$ of operators defined by $C_h = C^{X_0, X_h}, h \in \mathbb{Z}$. Note that C_0 is generally denoted by C or C^X. Those operators are involved in the estimation procedure of ρ (see Section 2.2). We can easily show that $C_{-h} = C_h^*$, and the following relations between the C_h for $h \in \mathbb{Z}$.

PROPOSITION 2.2 *If* (X_n) *is an ARHX(1) process following (1), then*

$$C_h = \rho C_{h-1} + \sum_{j=1}^{q} a_j C^{X_0, Z_{h,j}}, \quad h \geq 1; \quad C_0 = \rho C_{-1} + \sum_{j=1}^{q} a_j C^{X_0, Z_{0,j}} + C_\varepsilon.$$

2.1. LIMIT THEOREMS

Let us denote $S_n = X_1 + \cdots + X_n$. From the results in [9, Chapter. 3], we derive easily the following result.

THEOREM 2.1 *Let* (X_n) *be an ARHX(1) process. Then*

(i)

$$\frac{n^{1/4}}{(\ln n)^{\beta}} \frac{S_n}{n} \underset{n\to\infty}{\longrightarrow} 0 \quad \text{a.s.}, \quad \beta > 1/2.$$

(ii)

$$nE\left\|\frac{S_n}{n}\right\|^2 \underset{n\to\infty}{\longrightarrow} \sum_{h=-\infty}^{+\infty} E\langle X_0, X_h\rangle.$$

(iii) If $E\left(e^{\gamma\|\varepsilon_0'\|_{q+1}^2}\right) < \infty$ *for some* $\gamma > 0$,

$$\left\|\frac{S_n}{n}\right\| = \mathrm{O}\left(\left(\frac{\ln n}{n}\right)^{1/2}\right) \text{a.s.}$$

We now examine the central-limit Theorem. I_{q+1} stands for the identity operator on H^{q+1}.

THEOREM 2.2 (CLT) *Let* (X_n) *be an ARHX(1) process. Then, if* $I_{q+1} - \rho'$ *is invertible in* H^{q+1}

$$\frac{S_n}{\sqrt{n}} \underset{n\to\infty}{\overset{\mathcal{D}}{\longrightarrow}} \mathcal{N}(0, \Gamma),$$

with

$$\Gamma = P_1 (I_{q+1} - \rho')^{-1} C_{\varepsilon'} (I_{q+1} - \rho'^{*})^{-1} P_1.$$

Moreover, if conditions (i) and (ii) of Lemma A.2 in Appendix A hold, then

$$\Gamma = (I - \rho)^{-1} C_\varepsilon (I - \rho^*)^{-1} + \sum_{i=1}^{q} a_i (I - \rho)^{-1} (I - u_i)^{-1} C_\varepsilon \left(I - u_i^*\right)^{-1} (I - \rho^*)^{-1} a_i^*.$$

2.2. AUTOCOVARIANCE ESTIMATION

We suppose that

$$E\|T_0\|_{q+1}^4 < \infty. \tag{4}$$

Recall that for any u, v in H, $u \otimes v$ is the operator such that for all x in H, $u \otimes v(x) = \langle u, x\rangle v$. Now we define the following empirical covariance operators (respective estimators of $C^T, C^X = C, C^{Z_j}, C^{X,Z_j^+}$), where $Z_j^+ = \left(Z_{i+1,j}\right)_{i\in\mathbb{Z}}$:

$$C_n^T = \frac{1}{n}\sum T_i \otimes T_i, \quad C_n = C_n^X = \frac{1}{n}\sum X_i \otimes X_i,$$
$$C_n^{Z_j} = \frac{1}{n}\sum Z_{i,j} \otimes Z_{i,j}, \quad C_n^{XZ_j^+} = \frac{1}{n}\sum X_i \otimes Z_{i+1,j}.$$

We set $V_i(.) = \langle T_i, .\rangle_{q+1} T_i - C^T$. V_i is a $\mathcal{S}_{q+1}$-valued process. Using [9, Chapter 4] we get the following results about the rates of convergence of (C_n), denoting by $(\alpha_k, k \geq 1)$ the sequence of strong mixing coefficients of (T_n) and $\left(\lambda_j^V, j \geq 1\right)$ the sequence of eigenvalues of the covariance operator of V_0.

PROPOSITION 2.3 Under (4), we have

(i)

$$\limsup_{n\to\infty} nE\|C_n - C\|_{\mathcal{S}}^2 \leq \sum_{h=-\infty}^{+\infty} E\langle V_0, V_h\rangle_{\mathcal{S}_{q+1}},$$

(ii)

$$n^{1/4}(\log n)^{-\beta}\|C_n - C\|_{\mathcal{S}} \to 0 \quad \text{a.s.,} \quad \beta > 1/2.$$

(iii) If $\|T_0\|$ is bounded,

$$\|C_n - C\|_{\mathcal{S}} = \mathrm{O}\left(\left(\frac{\log n}{n}\right)^{1/2}\right)\text{a.s.}$$

(iv) If $E\left(e^{\gamma\|T_0\|^2}\right) < \infty$ for some $\gamma > 0$, and if there exist $a > 0$ and $r \in]0, 1[$ such that $\alpha_k \leq ar^k$, $k \geq 1$, $\lambda_j^V \leq ar^j$, $j \geq 1$,

then

$$\|C_n - C\|_{\mathcal{S}} = \mathrm{O}\left(\frac{(\log n)^{5/2}}{n^{1/2}}\right)\text{a.s.}$$

Remark 2.4. We could have chosen to consider the following representation of the ARHX(1) process with q exogenous variables, say $U_n = \rho''(U_{n-1}) + \varepsilon_n''$ with

$$U_n = \begin{pmatrix} X_n \\ X_{n-1} \\ Z_{n+1,1} \\ \vdots \\ Z_{n+1,q} \end{pmatrix}, \quad \varepsilon_n'' = \begin{pmatrix} \varepsilon_n \\ 0 \\ \eta_{n,1} \\ \vdots \\ \eta_{n,p} \end{pmatrix} \quad \text{and} \quad \rho'' = \begin{pmatrix} \rho & 0 & a_1 & \cdots & \cdots & a_q \\ I & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & u_1 & 0 & \cdots & 0 \\ 0 & 0 & 0 & u_2 & 0 & \vdots \\ \vdots & \vdots & \vdots & & \ddots & 0 \\ 0 & 0 & 0 & 0 & 0 & u_q \end{pmatrix}.$$

This way, we would have obtained similar results for the cross-covariance estimator $D_n = \frac{1}{n-1}\sum_{i=1}^{n-1} X_i \otimes X_{i+1}$ of $D = C_1 = C^{X_0,X_1}$.

2.3. ESTIMATION

We now make use the autoregressive representation in order to construct theoretical estimators of the functional parameters of model (1). Indeed, we can consider the estimator ρ'_n of ρ' in the simple ARH context.

Denote by (v_j), and $\left(\lambda_j^T\right)$ the eigenelements of C^T. Denote by $\tilde{\pi}^{k_n}$ the orthogonal projection over the subspace $_{k_n}$ spanned by the k_n eigenvectors $v_{1,n}, ..., v_{k_n,n}$ of C_n^T associated with the k_n greatest eigenvalues $\lambda_{1,n}^T, \ldots, \lambda_{k_n,n}^T$ ((k_n) is a sequence of integers such that $k_n \leq n$ for all $n \geq 1$ and $k_n \to \infty$). We make the following assumptions:

- $E\|T_0\|^4 < \infty$,
- $\lambda_1^T > \lambda_2^T > \cdots > \lambda_j^T > \cdots > 0$,
- $\lambda_{k_n,n}^T > 0$, $n \geq 1$ (a.s.).

We define the following estimators of the covariance and of its inverse

$$\tilde{C}_n^T = \tilde{\pi}^{k_n} C_n^T = \sum_{j=1}^{k_n} \lambda_{j,n}^T v_{j,n} \otimes v_{j,n} \quad \text{and} \quad \left(\tilde{C}_n^T\right)^{-1} = \sum_{j=1}^{k_n} \left(\lambda_{j,n}^T\right)^{-1} v_{j,n} \otimes v_{j,n}.$$

The estimator of ρ' is then $\rho'_n = \tilde{\pi}^{k_n} D_n^T \left(\tilde{C}_n^T\right)^{-1} \tilde{\pi}^{k_n}$. Define

$$\rho_n = P_1 \rho'_n P_1; \quad a_{i,n} = P_1 \rho'_n P_{i+1}, \ \ i = 1, ..., q; \quad u_{i,n} = P_{i+1} \rho'_n P_{i+1}, \ \ i = 1, ..., q.$$

THEOREM 2.3 *If ρ' is an Hilbert–Schmidt operator, and [9, Equation (8.65)] holds for the eigenvalues of C^T then*

$$\|\rho_n - \rho\|_{\mathcal{L}} \underset{n\to\infty}{\to} 0 \ \text{a.s.}, \quad \|a_{i,n} - a_i\|_{\mathcal{L}} \underset{n\to\infty}{\to} 0 \ \text{a.s.}, \quad \|u_{i,n} - u_i\|_{\mathcal{L}} \underset{n\to\infty}{\to} 0 \quad \text{a.s.}$$

Guillas [16] obtained rates of convergence in the L^2 sense for the estimation of the autocorrelation operator ρ in the ARH model. We can apply those results to our context using the same arguments as previously. Whether the eigenvectors of C^T are known or unknown, we get then respectively up to $n^{-1/3}$- and up to $n^{-1/4}$-rates of convergence depending of the previous knowledge of the decreasing speed of the eigenvalues of C^T.

3. A Simulation Study

To be able to simulate ARHX processes we may simply simulate an adequate ARH process in H^{q+1}, based on relation (2). Pumo [28] used the Karhunen–Loève theorem to simulate both Wiener processes and continuous ARH. Working with Wiener processes, he constructed the orthonormal basis associated with the Karhunen–Loève decomposition and decided to simulate continuous ARH by using a simple ρ operator in this basis. As ARH processes do not suppose especially Wiener noises, we choose to generalize this simulation to a larger class of processes by employing strong white noises.

We developed one specific library in the statistical software R (see [19]), which might be soon available at the CRAN (http://cran.r-project.org).

3.1. SIMULATION PRINCIPLE

Since it is more rational to work with finite dimensional processes, a natural way to simulate an ARH process is to consider a m-dimensional subspace of H, denoted by H_m, and to create (X_n) as an ARH process in H_m. We project (X_n) in the chosen discretization scheme to obtain our observations. We work with a number of discretization points d inferior to m in order to reason as if H_m approximates H. The larger m is, the closer we are to a functional modeling. Numerical applications were carried out with $m = 100$ and $d = 10$ discretization points. As we expected, to obtain optimal k_n smaller than 5 in the estimation, 10 discretization points were sufficient enough to measure the influence of k_n on our estimates.

Concretely, we simulated ARHX processes (X_n) with one exogenous variable (Z_n) ($q = 1$, dropping consequently in this section the i index, i.e. $Z_n = u(Z_{n-1}) + \eta_n$) as ARH processes T in H_m^2, using the representation (2). To do so, we needed to choose a basis of H_m in which are expressed ρ' and the simulated strong white noise $\left(\varepsilon_n'\right)$. The way to obtain those elements is exposed in the sequel, and the simulation is then straightforward.

The main advantage of this approach, in comparison to [4] and [28] is that we are able to simulate ARH processes with a non Gaussian noise or a nondiagonal correlation operator ρ. Figure 1 displays two examples of simulated processes: the first one with a uniform noise over $[-\sqrt{3}, \sqrt{3}]$ and a diagonal correlation operator ρ, and the second one with a nondiagonal ρ in the basis given by the eigenvectors of the covariance operator with a Gaussian noise. On the first hand, when ρ is diagonal, the effect is somehow reproductive as shown in Figure 1(a). On the other hand, when ρ is non diagonal, we may observe as in Figure 1(b) an asymmetric effect, especially on peaks. In this latter case, the curves are more difficult to forecast because of cross-correlations. Consequently, the method introduced in this paper is able to cover a broader variety of continuous-time processes.

– The basis can be constructed from a arbitrary noncollinear family of vectors, using the Gram–Schmidt orthonormalization. This technique enables us to control the shape of the first eigenvectors of the covariance operator. We noticed that the choice of the basis does not really influence the quality of the estimation. Accordingly, we opted for sinusoidal functions as first vectors of the basis.
– ρ' is written

$$\rho' = \begin{pmatrix} \rho & a \\ 0 & u \end{pmatrix} \tag{5}$$

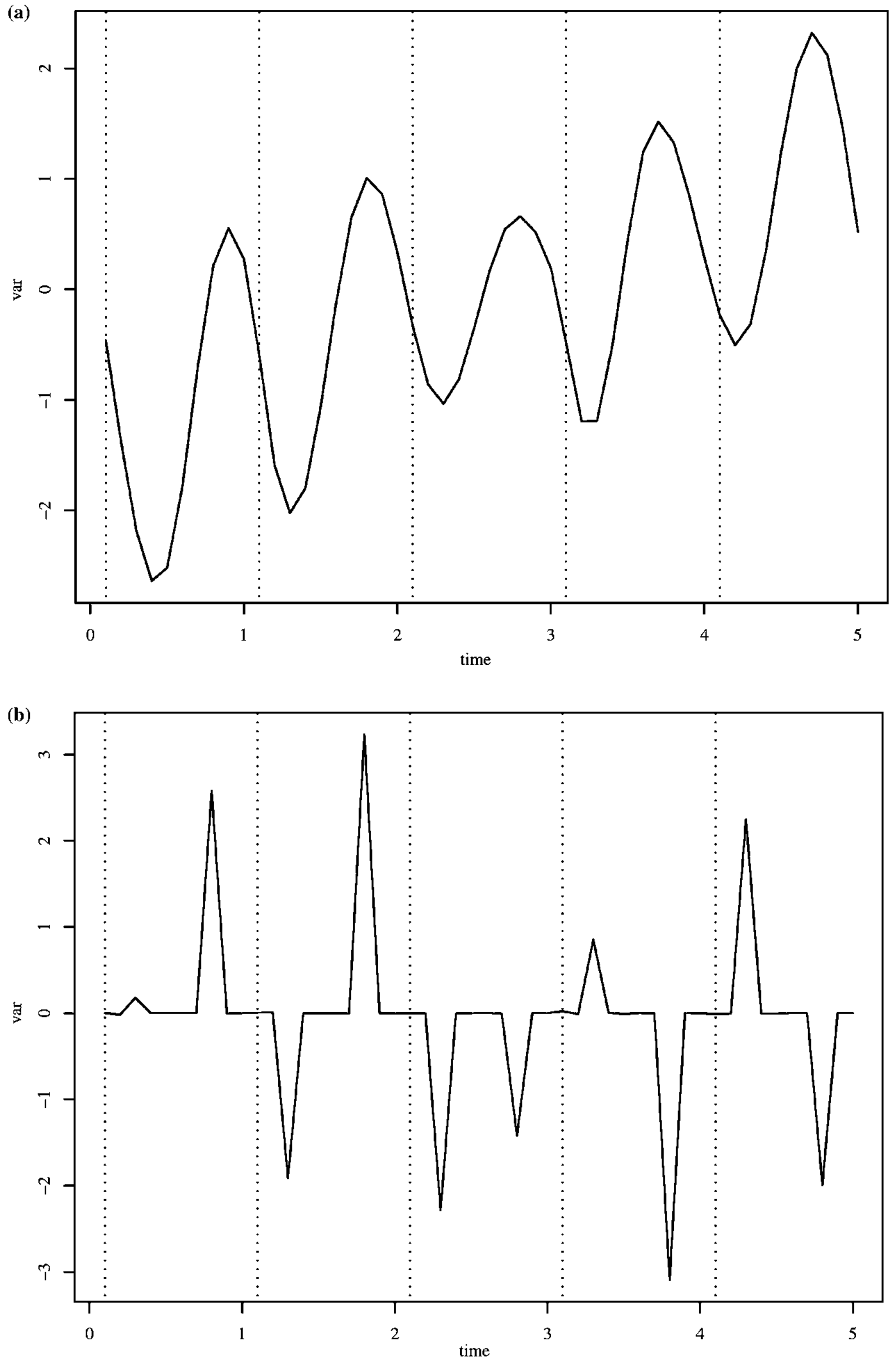

Figure 1. Two simulated sample paths of an ARH(1) process: (a) with a Uniform($[-\sqrt{3}, \sqrt{3}]$) noise, and (b) with a nondiagonal correlation operator.

in the induced basis of H_m^2. In the following, the numerical expressions of ρ' will not be given in this basis, but in the basis constructed with the first eigenvectors of C^T.
- The relation [9, 3.13] $C^\varepsilon = C^X - \rho C^X \rho^*$, is useful for simulating the noise adapted to a given covariance operator C^X. Pumo [28] simulated an ARH model, using this formula when all the operators are diagonals. We also apply this result to (T_n) in the nondiagonal case with the covariance operator

$$C^T = \begin{pmatrix} C^X & C^{XZ^+} \\ C^{Z^+X} & C^Z \end{pmatrix}.$$

The inner components of C^T verify the following relation

$$C^{X,Z^+} = uC^{X,Z^+}\rho^* + uC^Z a^*. \tag{6}$$

Indeed, for all x in H,

$$\begin{aligned} C^{X,Z^+}(x) &= E[\langle X_n, x\rangle Z_{n+1}] = E[\langle \rho(X_{n-1}) + a(Z_n) + \varepsilon_n, x\rangle (u(Z_n) + \eta_n)] \\ &= E[\langle \rho(X_{n-1}), x\rangle u(Z_n)] + E[\langle a(Z_n), x\rangle u(Z_n)] \\ &= E[\langle X_{n-1}, \rho^*(x)\rangle u(Z_n)] + E[\langle Z_n, a^*(x)\rangle u(Z_n)]. \end{aligned}$$

With this knowledge, we proceed in four steps:

1. choose the parameters ρ, a, u, C^X and C^{X,Z^+}, or more precisely their matrix representation in H_m,
2. compute from (6) the matrix representation of the operator C^Z, and therefore obtain C^T knowing that $C^{Z^+,X} = \left(C^{X,Z^+}\right)^*$,
3. determine $C^{\varepsilon'}$ using the relation $C^{\varepsilon'} = C^T - \rho' C^T \rho'^*$.
4. Simulate the noise $\left(\varepsilon'_n\right)$ from an i.i.d. sequence (ζ_n) of random variables in H_m^2 using the relation $\varepsilon'_n = \left(C^{\varepsilon'}\right)^{\frac{1}{2}}\zeta_n$, $n \geq 1$.

As one may notice, those different computations are not always possible due to non invertible matrices and the fact that the final matrix associated with $C^{\varepsilon'}$ needs to be positive definite.

3.2. MODELS

In Section 3.3, we compared the forecasting properties of the ARHX model to the ones issued from the ARH and the functional kernel models which both do not include the exogenous variable. Moreover, we distinguished two different implementations of the ARHX estimation. Those models are presented below, except the aforementioned ARH.

3.2.1. *Functional kernel model*

One nonparametric way to deal with the conditional expectation $\rho(x) = E[X_i|X_{i-1} = x]$, where (X_i) is a H-valued process, is to consider a predictor inspired by the classical kernel regression, as in [24] and [34]. It is a general estimation procedure because linearity of ρ is not assumed. Ref. [5] already considered the following functional kernel estimator of ρ:

$$\hat{\rho}_{h_n}(x) = \frac{\sum_{i=1}^{n-1} X_{i+1} \cdot K\left(\frac{\|X_i - x\|_H}{h_n}\right)}{\sum_{i=1}^{n-1} K\left(\frac{\|X_i - x\|_H}{h_n}\right)}, \quad x \in H, \tag{7}$$

where h_n is the bandwidth, $\|.\|_H$ is the norm of H (H is chosen to be $L^2[0, d]$ in our simulation), and K is the usual Gaussian kernel. Hence we get the predicted value of X_{n+1} given by $\hat{X}_{n+1} = \hat{\rho}_{h_n}(X_n)$ where h_n is obtained by cross-validation.

3.2.2. *Autoregressive Hilbertian process with exogenous variables*

Two approaches are at least possible, in order to estimate an ARHX process. The first one is simply to apply the theoretical techniques exposed previously. This model will be denoted by the acronym ARHX(a).

The second – denoted by the acronym ARHX(b) – deals with situations where the endogenous process hides the exogenous influence. An empirical answer to this problem is then to impose to the eigenvectors forming the basis of the autocovariance operator C^T to be built with eigenvectors of the autocovariance operators C^X, C^{Z_1}, ..., C^{Z_n} and naturally blocks of 0. This might seem very restrictive, but this construction can induce a better behavior than the usual one in some situations as shown in Section 3.3. The cross-validation procedure is leading to multiple selection of dimension reduction for each of the variables and is therefore more complex. Applying this method to an real forecasting problem leads to a model choice criterion by using a cross validation technique (see [14]). Indeed, if a variable is not worth introducing, its associated dimension will be 0.

3.3. RESULTS

Let $X_{i,j}$ be the *ith* coordinate of the *jth* observation of (X_n) and $\hat{X}_{i,j}$ the prediction of $X_{i,j}$. In order to compare the curves, we compute for integers $p = 1, 2$ respectively the following empirical L^p-errors and the L^∞-errors on a sample of n observations based on the discretization scheme

$$\|\hat{X} - X\|_{L^p} = \frac{1}{n}\sum_{j=1}^{n}\left(\frac{1}{d}\sum_{i=0}^{d-1}\|\hat{X}_{i,j} - X_{i,j}\|^p\right)^{1/p}, \quad p = 1, 2, \tag{8}$$

$$\|\hat{X} - X\|_{L^\infty} = \frac{1}{n}\sum_{j=1}^{n} \sup_{i=0,\ldots,d-1} |\hat{X}_{i,j} - X_{i,j}|.$$

In the numerical applications, we used the method described in Section 3.1 with a Gaussian sequence (ζ_n) and diverse sets of coefficients. For each simulation, we did not choose all the matrix coefficients but only those in the first block. The other coefficients were null except on the diagonal of ρ, u , C^X and C^{X,Z^+} where we put a decreasing sequence of numbers, e.g. $\lambda_i := \lambda i^{-2}$, $\lambda > 0$, λ smaller than the other coefficients.

3.3.1. *Efficiency of estimation*

A simple Monte-Carlo study showed us that the estimation of the parameters is efficient for relatively small size of samples. For instance, we considered ARHX processes in H_{100} estimated by the ARHX(a) model. We used a training sample size varying from 10 to 200, by steps of size 10, on which we estimated the models. Ten complementary observations were used to benchmark on. We simulated 200 processes for each value of the sample size and computed the mean of the 200 L^2-errors between ARH-X(a) forecasts and observations on the benchmark set.

In Figure 2, we see that L^2-errors become almost constant from a sample size of approximately 50. In the following, we systematically simulated processes with a sample size of 400 to ensure a good estimation.

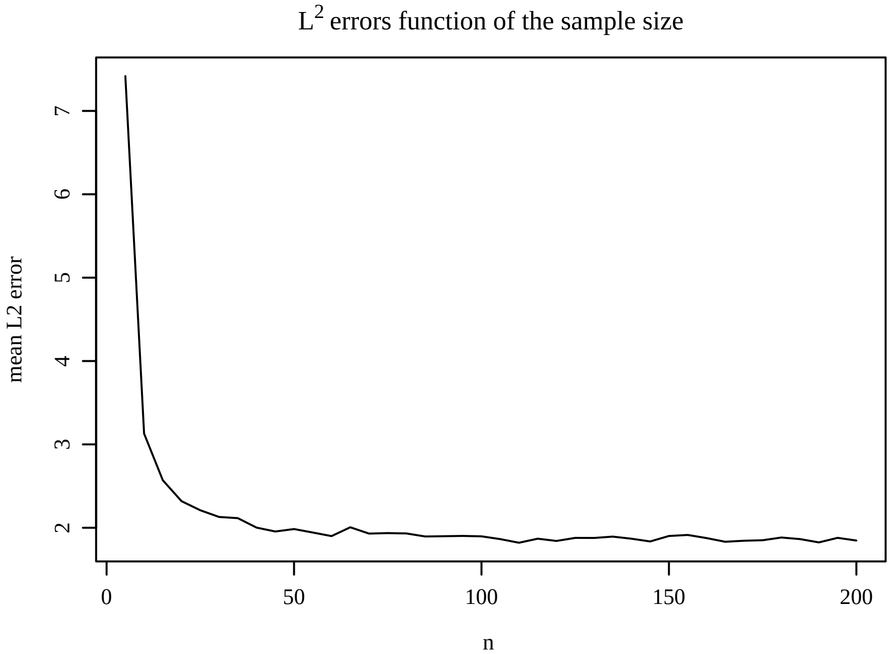

Figure 2. Monte Carlo estimation of the L^2 error, 200 replications for each sample size

Table I. Example of cross-validation for the *ARHX* model

k_n	L^1	L^2	L^∞	k_n	L^1	L^2	L^∞
1	0.518	0.604	0.94	11	0.405	0.467	0.709
2	0.47	0.542	0.834	12	0.405	0.466	0.708
3	0.401	0.462	0.703	13	0.406	0.467	0.709
4	0.401	0.462	0.7	14	0.408	0.47	0.713
5	0.402	0.463	0.702	15	0.409	0.47	0.713
6	0.404	0.465	0.706	16	0.415	0.477	0.724
7	0.405	0.466	0.707	17	0.412	0.474	0.718
8	0.405	0.466	0.707	18	0.412	0.475	0.72
9	0.405	0.467	0.708	19	0.412	0.475	0.72
10	0.406	0.468	0.71	20	0.412	0.475	0.72

3.3.2. *Parameters estimation: an example*

Once the size of sample was determined by the precedent study, we needed to validate that the quality of estimation is good with a sample size of 400. Therefore, we considered the following simulation of an ARHX process.

To make things clearer, we only give in the following lines the parameters involved in Equation (2) associated with our model. Their calculation is straightforward using formulae (5) and (6), starting from a set of coefficients for the matrices of ρ, u , C^X and C^{X,Z^+}.

The first four eigenvalues of the matrix associated with the covariance operator C^T are denoted λ' in a vectorial form, and the matrix expression of ρ' in the basis of the four eigenvectors associated with λ' is denoted by R'. The optimal value of k_n was obtained using cross-validation over the 80 last observations (the fifth of the sample size). As it appears in Table I, the minimum is obtained for $k_n = 4$, according to the theoretical structure of our simulated model. Notice that if we over-estimate this dimension, the loss of precision is moderate. The estimation performed with $k_n = 4$ yields the following estimates $\widehat{\lambda'}$ and $\widehat{R'}$ of λ' and R' respectively. The coefficients are well approximated, especially for the first components which contain the great bulk of the information.

$$\lambda' = \begin{pmatrix} 0.4150 \\ 0.1890 \\ 0.1150 \\ 0.0090 \end{pmatrix}, \quad R' = \begin{pmatrix} 0.672 & -0.134 & 0 & 0 \\ 0.366 & 0.228 & 0 & 0 \\ 0 & 0 & 0.9 & 0 \\ 0 & 0 & 0 & 0.34 \end{pmatrix},$$

$$\widehat{\lambda'} = \begin{pmatrix} 0.3800 \\ 0.1871 \\ 0.1197 \\ 0.0094 \end{pmatrix}, \quad \widehat{R'} = \begin{pmatrix} 0.615 & 0.196 & -0.053 & -0.046 \\ -0.404 & 0.317 & -0.009 & -0.131 \\ -0.001 & 0.067 & 0.904 & -0.033 \\ 0.005 & -0.013 & -0.001 & 0.342 \end{pmatrix}.$$

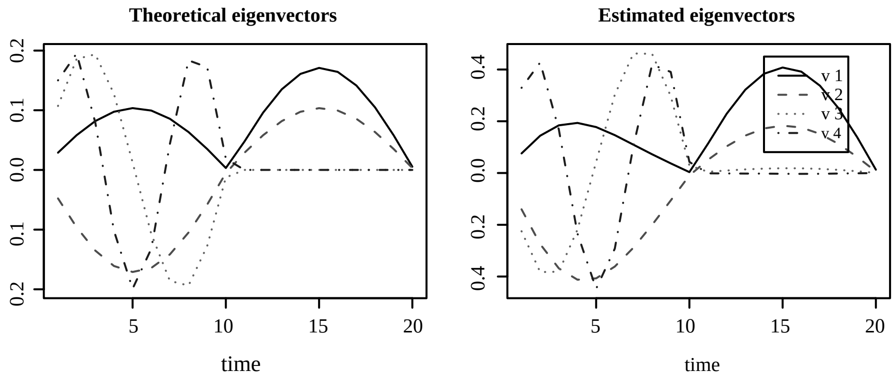

Figure 3. Estimated versus theoretical eigenvectors of C^T.

Moreover, we provided in Figure 3 the theoretical and estimated eigenvectors of the matrix associated with the covariance operator C^T. The similarity between the two sets of curves is very good. Note that we actually estimate the eigenvector space, and therefore on Figure 3, $-v_3$ and not v_3 is estimated.

3.3.3. *Comparison*

Finally, Monte-Carlo simulations show us how the inclusion of exogenous variables in functional time series can improve the predictions. We provide two sets of parameters which will be referred as S1 and S2. As we explained in Section 3.1, a simulation is characterized by the choice of a basis and the coefficients of the matrix associated with ρ, a, u, C^X and C^{X,Z^+}. We precise here those latter parameters as the choice of the basis is not crucial. The notation diag is used to refer to the diagonal elements of a diagonal matrix. The influence of the exogenous variable on the endogenous one is stronger when the ratio of the norms of C^{X,Z^+} and C^X is larger.

S1. A model with a strong influence of the exogenous variable:

$$\rho = \text{diag}(0.6, 0.5, \lambda_1, ...), \quad u = \text{diag}(0.1, 0.1, 0.7, 0.7, \lambda_1, ...),$$
$$C^X = \text{diag}(30, 30, \lambda_1, ...),$$
$$a = \text{diag}(0.8, 0.8, 0.1, 0.1, 0, ...), \quad C^{X,Z^+} = \text{diag}(2, 2, 10, 10, 0, ...).$$

In its two first directions, the exogenous variable exert a strong influence on the endogenous variable and its autocorrelation is weak. In the third and fourth directions, the exogenous variable exert a weak influence on the endogenous variable and its autocorrelation is strong.

Table II. Mean of L^2-errors computed on 10,000 *ARHX*(1) replications

Parameters	ARH	ARHX(a)	ARHX(b)	Kernel	Persistence
S1	5.616	2.532	2.513	5.831	6.204
S2	1.586	1.395	1.396	1.680	1.787

S2. A model with a little influence of the exogenous variable:

$$\rho = \mathrm{diag}(0.3, 0.7, 0.2, \lambda_1, ...), \quad u = \mathrm{diag}(0.7, 0.4, \lambda_1, ...),$$

$$C^X = \mathrm{diag}(3, 2, 1, \lambda_1, ...),$$

$$a = \begin{pmatrix} 0.6 & 0.2 & 0 & 0 \\ 0.1 & 0.5 & 0 & 0 \\ 0 & 0 & \lambda_1 & 0 \\ 0 & 0 & 0 & \ddots \end{pmatrix}, \quad C^{X,Z^+} = \begin{pmatrix} 0.5 & 0.3 & 0 & 0 \\ 0.2 & 0.5 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \ddots \end{pmatrix}.$$

Here, the exogenous variable combines in its two first directions a strong influence on the endogenous variable with a strong autocorrelation.

Ten thousand replications for each simulation set were computed. For each replication, we used the 320 first observations to calibrate the various models (using cross-validation for certain parameters) and the 80 other to compute the forecasts and contrast the errors.

Table II completes the validation of our methodology. One may notice, in addition to the ARHX supremacy, a better behavior of the ARH model comparing to the functional kernel. This might be explained by the fact that ARHX models may be interpreted as ARH models with a non independent noise or by the better predictive skills of ARH models over functional kernel in common situations, see e.g. [5].

Finally, we observe a good fitting of the exogenous influence by ARHX models, with a distinction between the (a) and (b) versions. This latter model is a little bit more accurate when the exogenous variable is hidden by the endogenous one (S1 case), because the estimator focuses more on the eigenvectors of C^Z.

Acknowledgement

We are grateful to T. Mourid for helpful discussions.

Appendix A. Proofs

Proof of Proposition 2.1. First, we know that Equation (2) has a unique stationary solution $T_n = \sum_{j\geq 0} \rho'^j(\varepsilon'_{n-j}), n \in \mathbb{Z}$, applying [9, Theorem 3.1] to the case $H = H^{q+1}$ where the series converges a.s. and in $L^2_H(\Omega, \mathcal{A}, \mathcal{P})$. It suffices to remark for existence that $X_n = P_1 T_n$. For uniqueness, let us suppose that $X_{n,1}$ is another stationary solution of Equation (1). Then we can define $T_{n,1} = (X_{n,1}, Z_{n+1})'$, which is a stationary solution of Equation (2), and by uniqueness of this solution, $T_{n,1} = T_n$ a.s., so $X_{n,1} = X_n$ a.s.

Proof of Proposition 2.2. For any integer h we have for all x in H

$$\begin{aligned} C_{-h}(x) &= E[\langle X_h, x\rangle X_0] \\ &= E[\langle \rho(X_{h-1}), x\rangle X_0] + E\left[\left\langle \sum_{j=1}^{q} a_j(Z_{h,j}), x \right\rangle X_0\right] + E[\langle \varepsilon_h, x\rangle X_0] \\ &= E[\langle X_{h-1}, \rho^*(x)\rangle X_0] + \sum_{j=1}^{q} E\Big[\Big\langle Z_{h,j}, a_j^*(x)\Big\rangle X_0\Big] + E[\langle \varepsilon_h, x\rangle X_0]. \end{aligned}$$

So

$$C_{-h} = C_{-h+1}\rho^* + \sum_{j=1}^{q} C^{Z_{h,j},X_0} a_j^* + \delta_{h0} C^{\varepsilon}, \text{ where } \delta_{h0} = \begin{cases} 0 & \text{if } h = 0, \\ 1 & \text{otherwise.} \end{cases}$$

Indeed, by Proposition 2.1 $X_n = \sum_{j=0}^{\infty} (P_1 \rho'^j)(\varepsilon'_{n-j})$, and because of the independence of the noises (ε_n) and $(\eta_{n,i})$ we can claim that $E[\langle \varepsilon_h, .\rangle X_0] = \delta_{h0} C^{\varepsilon}$. Thus

$$C_h = \rho C_{h-1} + \sum_{j=1}^{q} a_j C^{X_0, Z_{h,j}}, \quad h \geq 1; \quad C_0 = \rho C_{-1} + \sum_{j=1}^{q} a_j C^{X_0, Z_{0,j}} + C^{\varepsilon}, \quad h = 0.$$

LEMMA A.1. *Let (X_n) be an ARHX(1) defined by Equation (1) then*

$$|E\langle X_0, X_h\rangle| \leq \|\rho'^h\|_{\mathcal{L}_{q+1}} (E\|X_0\|^2 + \sum_{i=1}^{q} E\|Z_{0,i}\|^2) + \sum_{i=1}^{q} \|u_i^h\|_{\mathcal{L}} E\|Z_{0,i}\|^2, \quad h \geq 1.$$

Proof of Lemma A.1. T_n is an $ARH^{q+1}(1)$, so by [9, Lemma 3.2]

$$\left| E\langle T_0, T_h\rangle_{q+1} \right| \leq \|\rho'^h\|_{\mathcal{L}_{q+1}} E\|T_0\|_{q+1}^2, \quad h \geq 1,$$

i.e.

$$\left| E\langle X_0, X_h\rangle + \sum_{i=1}^{q} E\langle Z_{0,i}, Z_{h,i}\rangle \right| \leq \|\rho'^h\|_{\mathcal{L}_{q+1}} \left(E\|X_0\|^2 + \sum_{i=1}^{q} E\|Z_{0,i}\|^2 \right), \quad h \geq 1.$$

But

$$\begin{aligned} |E\langle X_0, X_h\rangle| &= \left| E\langle X_0, X_h\rangle + \sum_{i=1}^{q} E\langle Z_{0,i}, Z_{h,i}\rangle - \sum_{i=1}^{q} E\langle Z_{0,i}, Z_{h,i}\rangle \right| \\ &\leq \|\rho'^h\|_{\mathcal{L}_{q+1}} \left(E\|X_0\|^2 + \sum_{i=1}^{q} E\|Z_{0,i}\|^2 \right) + |E\langle Z_{0,i}, Z_{h,i}\rangle| \\ &\leq \|\rho'^h\|_{\mathcal{L}_{q+1}} (E\|X_0\|^2 + \sum_{i=1}^{q} E\|Z_{0,i}\|^2) + \sum_{i=1}^{q} \|u_i^h\|_{\mathcal{L}} E\|Z_{0,i}\|^2, \end{aligned}$$

since by [9, Lemma 3.2], for all i

$$\left|E\left\langle Z_{0,i}, Z_{h,i}\right\rangle\right| \leq \left\|u^h\right\|_{\mathcal{L}} E\left\|Z_{0,i}\right\|^2, \quad h \geq 1.$$

Proof of Theorem 2.1. (i) The proof relies upon Lemma A.1 and is similar to [9, Theorem 3.7].

(ii) By [9, Theorem. 3.8],

$$nE\left\|\frac{T_1+\cdots+T_n}{n}\right\|_{q+1}^2 \underset{n\to\infty}{\to} \sum_{h=-\infty}^{+\infty} E\langle T_0, T_h\rangle_{q+1}$$

and we have

$$\begin{aligned} nE\left\|\frac{T_1+\cdots+T_n}{n}\right\|_{q+1}^2 &= \frac{1}{n}\sum_{1\leq i,j\leq n} E\left\langle T_i, T_j\right\rangle_{q+1} \\ &= nE\left\|\frac{X_1+\cdots+X_n}{n}\right\|^2 + \sum_{i=1}^{q} nE\left\|\frac{Z_{1,i}+\cdots+Z_{n,i}}{n}\right\|^2, \end{aligned}$$

but for all i

$$nE\left\|\frac{Z_{1,i}+\cdots+Z_{n,i}}{n}\right\|^2 \underset{n\to\infty}{\to} \sum_{h=-\infty}^{+\infty} E\left\langle Z_{0,i}, Z_{h,i}\right\rangle$$

by [9, Theorem. 3.8]. Accordingly

$$\begin{aligned} nE\left\|\frac{X_1+\cdots+X_n}{n}\right\|^2 &\underset{n\to\infty}{\to} \left(\sum_{h=-\infty}^{+\infty} E\langle T_0, T_h\rangle_{q+1} - \sum_{i=1}^{q}\sum_{h=-\infty}^{+\infty} E\left\langle Z_{0,i}, Z_{h,i}\right\rangle\right) \\ &= \sum_{h=-\infty}^{+\infty} E\langle X_0, X_h\rangle. \end{aligned}$$

(iii) The result is based on [9, Cor. 3.2] and the relation

$$\left\|\frac{T_1+\cdots+T_n}{n}\right\|_{q+1} = \left\|\frac{X_1+\cdots+X_n}{n}\right\| + \sum_{i=1}^{q}\left\|\frac{Z_{1,i}+\cdots+Z_{n,i}}{n}\right\|.$$

LEMMA A.2. *Suppose that*

(i) $\rho, a_1, ..., a_q, u_1, ..., u_q$ *commute*

(ii) $I-\rho, I-u_1, ..., I-u_q$ *are invertible,*

then $I_{q+1}-\rho'$ *is invertible in* H^{q+1} *and*

$$\left(I_{q+1}-\rho'\right)^{-1} = \begin{pmatrix} (I-\rho)^{-1} & a_1(I-\rho)^{-1}(I-u_1)^{-1} & \cdots & \cdots & a_q(I-\rho)^{-1}(I-u_q)^{-1} \\ 0 & (I-u_1)^{-1} & 0 & \cdots & 0 \\ \vdots & 0 & \ddots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & 0 \\ 0 & 0 & \cdots & 0 & (I-u_q)^{-1} \end{pmatrix}.$$

Proof of Lemma A.2. Consider the case $q = 1$ for simplicity. We calculate the following product

$$\begin{pmatrix} I-\rho & -a_1 \\ 0 & I-u_1 \end{pmatrix}\begin{pmatrix} (I-\rho)^{-1} & a_1(I-\rho)^{-1}(I-u_1)^{-1} \\ 0 & (I-u_1)^{-1} \end{pmatrix}$$
$$= \begin{pmatrix} (I-\rho)(I-\rho)^{-1} & a_1(I-u_1)^{-1} - a_1(I-u_1)^{-1} \\ 0 & (I-u_1)(I-u_1)^{-1} \end{pmatrix} = \begin{pmatrix} I & 0 \\ 0 & I \end{pmatrix}.$$

Proof of Theorem 2.2.

$$\frac{T_1 + \cdots + T_n}{\sqrt{n}} \underset{n\to\infty}{\overset{\mathcal{D}}{\to}} \mathcal{N}(0, \Gamma'),$$

where

$$\Gamma' = (I_{q+1} - \rho')^{-1} C_{\varepsilon'} (I_{q+1} - \rho'^*)^{-1}$$

because (T_n) is an ARH(1) such that (C) holds. By continuity of weak convergence with respect to continuous transformation,

$$\frac{S_n}{\sqrt{n}} = P_1\left(\frac{T_1 + \cdots + T_n}{\sqrt{n}}\right) \underset{n\to\infty}{\overset{\mathcal{D}}{\to}} \mathcal{N}(0, P_1\Gamma' P_1).$$

Since the noises are independent, we may write

$$C^{\varepsilon'} = \begin{pmatrix} C^{\varepsilon} & \cdots & \cdots & 0 \\ \vdots & C^{\eta_1} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & C^{\eta_q} \end{pmatrix}.$$

We are interested in the first block of $\Gamma' = (I_{q+1} - \rho')^{-1} C^{\varepsilon'} (I_{q+1} - \rho'^*)^{-1}$. But a general matrix product ABC when B is diagonal leads to $(ABC)_{11} = \sum_{k=1}^{q+1} A_{1k} B_{kk} C_{k1}$. In our situation, $B_{22} = 0$, and when $\rho, a_1, ..., a_q, u_1, ..., u_q$ commute then by the previous lemma we get (knowing that $C_{k1} = A_{1k}^*$)

$$\Gamma = (I-\rho)^{-1} C^{\varepsilon} (I-\rho^*)^{-1} + \sum_{i=1}^{q} a_i (I-\rho)^{-1} (I-u_i)^{-1} C^{\varepsilon} (I-u_i^*)^{-1} (I-\rho^*)^{-1} a_i^*.$$

Proof of Proposition 2.3. For point (i), we write C^T and C_n^T in a block form

$$C^T = \begin{pmatrix} C^X & C^{X,Z_1^+} & \cdots & C^{X,Z_q^+} \\ C^{Z_1^+,X} & C^{Z_1} & & \vdots \\ \vdots & \vdots & \ddots & \\ C^{Z_q^+,X} & \cdots & & C^{Z_q} \end{pmatrix}, \quad C_n^T = \begin{pmatrix} C_n^X & C_n^{X,Z_1^+} & \cdots & C_n^{X,Z_q^+} \\ C_n^{Z_1^+,X} & C_n^{Z_1} & & \vdots \\ \vdots & \vdots & \ddots & \\ C_n^{Z_q^+,X} & \cdots & & C_n^{Z_q} \end{pmatrix}.$$

Thus,

$$\left\|C_n^T - C^T\right\|_{\mathcal{S}_{q+1}}^2 \geq \|C_n - C\|_{\mathcal{S}}^2 + \sum_{j=1}^{q}\left\|C_n^{Z_j} - C^{Z_j}\right\|_{\mathcal{S}}^2 + \sum_{j=1}^{q}\left\|C_n^{X,Z_j^+} - C^{X,Z_j^+}\right\|_{\mathcal{S}}^2. \tag{9}$$

But by [9, Theorem 4.1],

$$\limsup_{n\to\infty} nE\left\|C_n^T - C^T\right\|_{\mathcal{S}_{q+1}}^2 = \sum_{h=-\infty}^{+\infty} E\langle V_0, V_h\rangle_{\mathcal{S}_{q+1}}.$$

This point entails easily part (*i*) of the proposition. For (*ii*) and (*iii*), we use (9) and the results in [9, Chapter 4].

Proof of Theorem 2.3. It suffices to remark that for all i and j

$$\left\|P_i(\rho_n' - \rho')P_j\right\|_{\mathcal{L}} \leq \|P_i\|_{\mathcal{L}}\left\|\rho_n' - \rho'\right\|_{\mathcal{L}}\left\|P_j\right\|_{\mathcal{L}} \leq \left\|\rho_n' - \rho'\right\|_{\mathcal{L}}$$

and then apply [9, Theorem. 8.7], knowing that

$$\rho_n - \rho = P_1(\rho_n' - \rho')P_1,$$
$$a_{i,n} - a_i = P_1(\rho_n' - \rho')P_{i+1}; u_{i,n} - u_i = P_{i+1}(\rho_n' - \rho')P_{i+1},\ i = 1, ..., q.$$

References

1. Baillie, R. T.: Asymptotic prediction mean squared error for vector autoregressive models, *Biometrika* **66** (3) (1979), 675–678.
2. Bauer, G., Deistler, M. and Scherrer W.: Time series models for short term forecasting of ozone in the eastern part of Austria, *Environmetrics* **12** (2001), 117–130.
3. Benyelles, B. and Mourid T.: Estimation de la période d'un processus à temps continu à représentation autorégressive, *C. R. Acad. Sci. Paris Sér. I Math.* **333** (2001), 245–248.
4. Besse, P. and Cardot H.: Approximation spline de la prévision d'un processus fonctionnel autorégressif d'ordre 1, *Revue Canadienne de la Statistique/Canadian Journal of Statistics* **24** (1996), 467–487.
5. Besse, P., Cardot, H. and Stephenson D.: Autoregressive forecasting of some functional climatic variations, *Scand. J. Statist.* **27** (4) (2000), 673–687.
6. Bierens, H. J.: Least squares estimation of linear and nonlinear ARMAX models under data heterogeneity. *Ann. Économ. Statist.* **20-21** (1990), 143–169.
7. Bondon, P.: Recursive relations for multistep prediction of a stationary time series, *J. Time Ser. Anal.* **22** (4) (2001), 399–410.
8. Bosq, D.: Nonparametric prediction for unbounded almost stationary processes, In: *Nonparametric functional estimation and related topics (Spetses, 1990)*. Kluwer Acad. Publ., Dordrecht: (1991), pp. 389–403.
9. Bosq, D.: Linear Processes in Function Spaces: Theory and Applications, Lecture Notes in Statistics, Vol. 149, New York: Springer-Verlag, 2000.
10. Bosq, D. and Shen, J.: Estimation of an autoregressive semiparametric model with exogenous variables, *J. Statist. Plann. Inference* **68** (1) (1998), 105–127.
11. Boutahar, M.: Strong Consistency of least squares estimates in general $\mathrm{ARX}_d(p, s)$ system, *Stochast. Stochast. Rep.* **38** (3) (1992), 175–184.
12. Cai, Z. and Masry, E.: Nonparametric estimation of additive nonlinear ARX time series: local linear fitting and projections, *Econom. Theory* **16** (4) (2000), 465–501.

13. Chen, X. and Shen, X.: Sieve extremum estimates for weakly dependent data, *Econometrica* **66**(2) (1998), 289–314.
14. Damon, J. and Guillas, S.: The inclusion of exogenous variables in functional autoregressive ozone forecasting, *Environmetrics* **13** (2002), 759–774.
15. Duflo, M.: *Random Iterative Models*, Springer-Verlag, Berlin, 1997.
16. Guillas, S.: Rates of convergence of autocorrelation estimates for autoregressive Hilbertian processes, 2001. *Statistics and Probability Letters*, **55** (2003), 281–291.
17. Hannan, E. J. and Deistler, M.: The Statistical Theory of Linear Systems. John Wiley & Sons Inc, New York, 1988.
18. Hoque, A. and Peters, T. A.: Finite sample analysis of the ARMAX models, *Sankhyā Ser.* B **48** (2), (1986), 266–283.
19. Ihaka, R. and Gentleman, R.: R: a language for data analysis and graphics, *J. Graphical Comput. Statist.* **5** (1996), 299–314.
20. Liu, S. I.: Bayesian multiperiod forecasts for ARX models, *Ann. Inst. Statist. Math.* **47** (2) (1995), 211–224.
21. Mas, A.: Normalité asymptotique de l'estimateur empirique de l'opérateur d'autocorrélation d'un processus ARH(1), *C. R. Acad. Sci. Paris Sér. I Math.* **329** (10) (1999), 899–902.
22. Merlevède, F.: Résultats de convergence presque sûre pour l'estimation et la prévision des processus linéaires hilbertiens, *C. R. Acad. Sci. Paris Sér. I Math.* **324**(5) (1997), 573–576.
23. Mourid, T.: Contribution à la statistique des processus autorégressifs à temps continu, D.Sc. thesis, Université Paris 6, 1995.
24. Nadaraja, E. A.: On a regression estimate (Russian), *Teor. Verojatnast. i Primenen.* **9**, (1964), 157–159.
25. Penm, J. H. W., Penm, J. H. and Terrell R. D.: The recursive fitting of subset VARX models, *J. Time Ser. Anal.* **14** (6) (1993), 603–619.
26. Pitard, A. and Viel, J.: A model selection tool in multi-pollutant time series: the granger-causality diagnosis, *Environmetrics* **10** (1999), 53–65.
27. Poskitt, D. S. and Tremayne A. R.: Testing misspecification in vector time series models with exogenous variables, *J. Roy. Statist. Soc. Ser.* B **46** (2) (1984), 304–315.
28. Pumo, B.: Estimation et Prévision de Processus Autorégressifs Fonctionnels. Applications Aux Processus À Temps Continu, Ph.D. thesis, Université Paris 6, 1992.
29. Pumo, B.: Prediction of continuous time processes by C[0,1]-valued autoregressive process, *Statis. Inference Stochast. Process.* **1** (3) (1998), 297–309.
30. Ramsay, J. and Silverman B.: *Functional Data Analysis*, Springer-Verlag, 1997.
31. Rice, J. A. and Silverman B. W.: Estimating the mean and covariance structure nonparametrically when the data are curves, *J. Roy. Statist. Soc. Ser.* B **53** (1), (1991), 233–242.
32. Riveraa, D. E. and Gaikwada S. V.: Digital PID Controller Design Using ARX Estimation, *Comput. Chem. Engi.* **20** (1996), 1317–1334.
33. Spliid, H.: A fast estimation method for the vector autoregressive moving average model with exogenous variables, *J. Amer. Statist. Assoc.* **78** (384) (1983), 843–849.
34. Watson, G. S.: Smooth regression analysis, *Sankhya Ser.* A **26** (1964), 359–372.
35. Yoshida, H. and Kumar, S.: Development of ARX model based off-line FDD technique for energy efficient buildings, *Renew. Energy* **22** (2001), 53–59.