

# Asymptotic properties of a componentwise ARH(1) plug-in predictor

Javier Álvarez-Liébana<sup>1</sup>, Denis Bosq<sup>2</sup> and M. Dolores Ruiz-Medina<sup>1</sup>

2nd June 2021

<sup>1</sup> Department of Statistics and O. R., University of Granada, Spain. <sup>2</sup> LSTA, Université Pierre et Marie Curie–Paris 6, Paris, France.  
*E-mail: javialvaliebana@ugr.es*

## Summary

This paper presents new results on prediction of linear processes in function spaces. The autoregressive Hilbertian process framework of order one (ARH(1) process framework) is adopted. A componentwise estimator of the autocorrelation operator is formulated, from the moment-based estimation of its diagonal coefficients, with respect to the orthogonal eigenvectors of the auto-covariance operator, which are assumed to be known. Mean-square convergence to the theoretical autocorrelation operator, in the space of Hilbert-Schmidt operators, is proved. Consistency then follows in that space. For the associated ARH(1) plug-in predictor, mean absolute convergence to the corresponding conditional expectation, in the considered Hilbert space, is obtained. Hence, consistency in that space also holds. A simulation study is undertaken to illustrate the finite-large sample behavior of the formulated componentwise estimator and predictor. The performance of the presented approach is compared with alternative approaches in the previous and current ARH(1) framework literature, including the case of unknown eigenvectors.

**Journal of Multivariate Analysis**, 155, pp. 12-34. DOI: [doi.org/10.1016/j.jmva.2016.11.009](https://doi.org/10.1016/j.jmva.2016.11.009)

*Key words:* ARH(1) processes; consistency; functional prediction; mean absolute and quadratic convergence.

## 1 Introduction

In the last few decades, an extensive literature on statistical inference from functional random variables has emerged. This work was motivated in part by the statistical analysis of high-dimensional data, as well as data of a continuous (infinite-dimensional) nature; see, e.g., Bosq [2000, 2007]; Dedecker and Merlevède [2003]; Ferraty and Vieu [2006]; Merlevède [1996b, 1997]; Ramsay and Silverman [2005]; Ruiz-Medina [2012]. New developments in functional data analysis are described, e.g., in Bongiorno et al. [2014];

Cuevas [2014]; Horváth and Kokoszka [2012]; Hsing and Eubank [2015], and in a recent Special Issue of this journal Goia and Vieu [2016].

These references include a nice summary on the statistics theory for functional data, contemplating covariance operator theory and eigenfunction expansion, perturbation theory, smoothing and regularization, probability measures on a Hilbert spaces, functional principal component analysis, functional counterparts of the multivariate canonical correlation analysis, the two sample problem and the change point problem, functional linear models, functional test for independence, functional time series theory, spatially distributed curves, software packages and numerical implementation of the statistical procedures discussed, among other topics.

The special case of functional regression models, in which the predictor is a random function and the response is scalar, has been particularly well studied. Various specifications of the functional regression parameter arise in fields such as biology, climatology, chemometrics, and economics. To avoid the computational (high-dimensional) limitations of the nonparametric approach, several parametric and semi-parametric methods have been proposed; see, e.g., Ferraty et al. [2012] and the references therein. In Ferraty et al. [2012], a combination of a spline approximation and the one-dimensional Nadaraya-Watson approach was proposed to avoid high dimensionality issues. Generalizations to the case of more regressors (all functional, or both functional and real) were also addressed in the nonparametric, semi-parametric, and parametric frameworks; for an overview, see Aneiros-Pérez and Vieu [2006]; Febrero-Bande and González-Manteiga [2013]; Ferraty and Vieu [2009].

In the nonparametric regression framework, the case where the covariates and the response are functional was considered by Ferraty et al. [2012], where a functional version of the Nadaraya-Watson estimator was proposed for the estimation of the regression operator and shown to be point-wise asymptotically normal. Resampling techniques were used to overcome the difficulties arising in the estimation of the asymptotic bias and variance. Semi-functional partial linear regression, introduced in Aneiros-Pérez and Vieu [2008], allows the prediction of a real-valued random variable from a set of real-valued explanatory variables, and a time-dependent functional explanatory variable. Motivated by genetic and environmental applications, a semi-parametric maximum likelihood method for the estimation of odds ratio association parameters was developed by Chen et al. [2012] in a high-dimensional data context.

In the autoregressive Hilbertian time series framework, several estimation and prediction procedures have been proposed and studied. Mas [1999] established, under suitable conditions, the asymptotic normal distribution of the formulated estimator of the autocorrelation operator, based on projection into the theoretical eigenvectors. In Bosq [2000]; Bosq and Blanke [2007], the problem of prediction of linear processes in function spaces was addressed. In particular, sufficient conditions for the consistency of the

empirical autocovariance and cross-covariance operators were obtained. The asymptotic normal distribution of the empirical autocovariance operator was also derived. Moreover, the asymptotic properties of the empirical eigenvalues and eigenvectors were analysed.

Guillas [2001] established the efficiency of a componentwise estimator of the autocorrelation operator, based on projection into the empirical eigenvector system of the autocovariance operator. Consistency, in the space of bounded linear operators, of the formulated estimator of the autocorrelation operator, and of its associated ARH(1) plug-in predictor was later proved by Mas [2004]. He derived sufficient conditions for the weak convergence of the ARH(1) plug-in predictor to a Hilbert-valued Gaussian random variable (see Mas [2007]). Simultaneously, Mas and Menneteau [2003a] obtained high deflection results or large and moderate deviations for infinite-dimensional autoregressive processes. Furthermore, the law of the iterated logarithm for the covariance operator estimator was formulated by Menneteau [2005].

The main properties of the class of autoregressive Hilbertian processes with random coefficients were investigated by Mourid [2004]. Kargin and Onatski [2008] gave interesting extensions of the autoregressive Hilbertian framework, based on the spectral decomposition of the autocorrelation operator, and not of the autocovariance operator. The first generalization on autoregressive processes of order greater than one was proposed by Mourid [1993], in order to improve prediction. ARHX(1) models; i.e., autoregressive Hilbertian processes with exogenous variables were studied by Damon and Guillas [2002, 2005]. In Guillas [2000, 2001] a doubly stochastic formulation of the autoregressive Hilbertian process was investigated. The ARHD model was introduced by Marion and Pumo [2004], taking into account the regularity of trajectories through the derivatives. The conditional autoregressive Hilbertian process (CARH process) was considered by Cugliari [2011], developing parallel projection estimation methods to predict such processes. In the Banach-valued context, we refer to the papers by Bensmain and Mourid [2001]; Dehling and Sharipov [2005]; Pumo [1992, 1998], among others.

In this paper, we assume that the autocorrelation operator belongs to the Hilbert-Schmidt class, and admits a diagonal spectral decomposition in terms of the orthogonal eigenvector system of the autocovariance operator. Such is the case, e.g., of an autocorrelation operator defined as a continuous function of the autocovariance operator. A componentwise estimator of the autocorrelation operator is then constructed in terms of the eigenvectors of the autocovariance operator, which are assumed to be known. This occurs when the random initial condition is defined as the solution, in the mean-square sense, of a stochastic differential equation driven by white noise. Beyond this case, the sparse representation and whitening properties of wavelet bases can be exploited to obtain a diagonal representation of the autocovariance and cross-covariance operators, in terms of a common and known wavelet basis. Unconditional bases, like wavelet bases, also allow the diagonal spectral series representation of the distributional kernels of Calderón-Zygmund operators.

Under the assumptions stated in [Appendices 2–4](#), we establish the convergence in the  $\mathcal{L}^2$ -sense of a componentwise estimator of the autocorrelation operator in the space of Hilbert–Schmidt operators  $\mathcal{S}(H)$ , i.e.,  $\mathcal{L}_{\mathcal{S}(H)}^2(\Omega, \mathcal{A}, \mathcal{P})$ , is derived. Consistency then follows in  $\mathcal{S}(H)$ . Under the same conditions, consistency in  $H$  of the associated ARH(1) plug-in predictor is obtained, from its convergence in the  $\mathcal{L}^1$ -sense in the Hilbert space  $H$ , i.e., in the space  $\mathcal{L}_H^1(\Omega, \mathcal{A}, \mathcal{P})$ . The Gaussian framework is analysed in [Appendix 4](#) and illustrated in [Appendix 5](#), where examples show the behaviour of the proposed componentwise autocorrelation operator estimator, and associated predictor, for large sample sizes. We also present there a comparative study with alternative ARH(1) prediction techniques, including componentwise parameter estimation of the autocorrelation operator, from known and unknown eigenvectors, as well as kernel (nonparametric) functional estimation, and penalized, spline and wavelet, estimation. Final comments on the application of the proposed approach from real data are provided in [Appendix 6](#).

## 2 Preliminaries

This section contains the preliminary definitions and lemmas that will be used to derive the main results of this paper. In the following,  $H$  denotes a real separable Hilbert space. Recall that, from Bosq [2000], a zero-mean ARH(1) process  $X = \{X_n, n \in \mathbb{Z}\}$  satisfies, for all  $n \in \mathbb{Z}$ , the equation

$$X_n = \rho(X_{n-1}) + \varepsilon_n, \quad (1)$$

where  $\rho$  denotes the autocorrelation operator of the process  $X$ , which belongs to the space  $\mathcal{L}(H)$  of bounded linear operators, such that  $\|\rho^k\|_{\mathcal{L}(H)} < 1$ , for all integers  $k \geq k_0$  beyond a certain  $k_0 \geq 1$ , with  $\|\cdot\|_{\mathcal{L}(H)}$  denoting the norm in the space  $\mathcal{L}(H)$ . The Hilbert-valued innovation process  $\varepsilon = \{\varepsilon_n, n \in \mathbb{Z}\}$  is assumed to be a strong-white noise which is uncorrelated with the random initial condition. That is,  $\varepsilon$  is a Hilbert-valued zero-mean stationary process, with independent and identically distributed components in time, with  $\sigma_\varepsilon^2 = \mathbb{E}\{\|\varepsilon_n\|_H^2\} < \infty$ , for all  $n \in \mathbb{Z}$ . We restrict our attention here to the case where  $\rho$  is such that

$$\|\rho\|_{\mathcal{L}(H)} < 1.$$

The following assumptions are made.

**Assumption A1.** The autocovariance operator

$$C = \mathbb{E}\{X_n \otimes X_n\} = \mathbb{E}\{X_0 \otimes X_0\}, \quad n \in \mathbb{Z},$$

is a positive, self-adjoint and trace operator. As a result, it admits the following diagonal spectral representation

$$C = \sum_{j=1}^{\infty} C_j \phi_j \otimes \phi_j,$$

in terms of an orthonormal system  $\{\phi_j, j \geq 1\}$  of eigenvectors which are known. Here,

$$C_1 \geq C_2 \geq \dots \geq C_j \geq \dots > 0$$

denote the real positive eigenvalues of  $C$  arranged in decreasing order of magnitude and

$$\sum_{j=1}^{\infty} C_j < \infty.$$

**Assumption A2.** The autocorrelation operator  $\rho$  is a self-adjoint and Hilbert–Schmidt operator, admitting the diagonal spectral decomposition

$$\rho = \sum_{j=1}^{\infty} \rho_j \phi_j \otimes \phi_j, \quad \sum_{j=1}^{\infty} \rho_j^2 < \infty,$$

where  $\{\rho_j, j \geq 1\}$  is the system of eigenvalues of the autocorrelation operator  $\rho$ , with respect to the orthonormal system of eigenvectors  $\{\phi_j, j \geq 1\}$  of the autocovariance operator  $C$ .

Note that, under **Assumption A2**,

$$\|\rho\|_{\mathcal{L}(H)} = \sup_{j \geq 1} |\rho_j| < 1.$$

*Remark 1* **Assumption A2** holds, in particular, when operator  $\rho$  is defined as a continuous function of operator  $C$  (see [Dautray and Lions, 1990, pp. 119–140] and **Remark 4**).

In the following, for any  $n \in \mathbb{Z}$ , let

$$D = \mathbb{E} \{X_n \otimes X_{n+1}\} = \mathbb{E} \{X_0 \otimes X_1\}$$

be the cross-covariance operator of the ARH(1) process  $X$ .

*Remark 2* Under **Assumptions A1–A2**, it follows from equation (1) that

$$C_\varepsilon = C_\rho C \rho = \sum_{j=1}^{\infty} C_j (1 - \rho_j^2) \phi_j \otimes \phi_j = \sum_{j=1}^{\infty} \sigma_j^2 \phi_j \otimes \phi_j.$$

By projecting equation (1) into the orthonormal system  $\{\phi_j, j \geq 1\}$ , we also have, for each  $j \geq 1$  and all  $n \in \mathbb{Z}$ , the AR(1) equation

$$X_{n,j} = \rho_j X_{n-1,j} + \varepsilon_{n,j}, \quad n \in \mathbb{Z}, \quad (2)$$

where  $X_{n,j} = \langle X_n, \phi_j \rangle_H$  and  $\varepsilon_{n,j} = \langle \varepsilon_n, \phi_j \rangle_H$ , for all  $n \in \mathbb{Z}$ . From equation (2), we have, for each  $j \geq 1$  and all  $n \in \mathbb{Z}$ ,

$$\begin{aligned} \rho_j &= \rho(\phi_j)(\phi_j) = \langle \phi_j, DC^{-1}(\phi_j) \rangle_H = \langle D(\phi_j), \phi_j \rangle_H \langle C^{-1}(\phi_j), \phi_j \rangle_H \\ &= \frac{\mathbb{E}\{X_{n,j}X_{n-1,j}\}}{\mathbb{E}\{X_{n-1,j}^2\}} = \frac{D_j}{C_j}, \quad n \in \mathbb{Z}, \end{aligned} \quad (3)$$

where

$$D_j = \langle D(\phi_j), \phi_j \rangle_H = \mathbb{E}\{X_{n,j}X_{n-1,j}\}, \quad C_j^{-1} = [\mathbb{E}\{X_{n-1,j}^2\}]^{-1}, \quad X_{n,j} = \langle X_n, \phi_j \rangle_H,$$

given that, for all  $j \geq 1$ ,

$$D = \sum_{j=1}^{\infty} D_j \phi_j \otimes \phi_j, \quad D_j = \rho_j C_j, \quad j \geq 1. \quad (4)$$

Let us now consider the Banach space  $L_{\mathcal{H}}^2(\Omega, \mathcal{A}, \mathcal{P})$  of the equivalence classes of  $\mathcal{L}_{\mathcal{H}}^2(\Omega, \mathcal{A}, \mathcal{P})$ , the space of zero-mean second-order Hilbert-valued random variables ( $\mathcal{H}$ -valued random variables) with finite seminorm given by

$$\|Z\|_{\mathcal{L}_{\mathcal{H}}^2(\Omega, \mathcal{A}, \mathcal{P})} = \sqrt{\mathbb{E}\{\|Z\|_{\mathcal{H}}^2\}}, \quad \forall Z \in \mathcal{L}_{\mathcal{H}}^2(\Omega, \mathcal{A}, \mathcal{P}).$$

That is, for  $Z, Y \in \mathcal{L}_{\mathcal{H}}^2(\Omega, \mathcal{A}, \mathcal{P})$ ,  $Z$  and  $Y$  belong to the same equivalence class if and only if

$$\mathbb{E}\{\|Z - Y\|_{\mathcal{H}}\} = 0.$$

The convergence in the seminorm of  $\mathcal{L}_{\mathcal{S}(H)}^2(\Omega, \mathcal{A}, \mathcal{P})$  will be considered in [Proposition 1](#), where  $\mathcal{H} = \mathcal{S}(H)$  denotes the Hilbert space of Hilbert-Schmidt operators on a Hilbert space  $H$ .

For each  $n \in \mathbb{Z}$ , let us consider the following biorthogonal representation of the functional value  $X_n$  of the ARH(1) process  $X = \{X_n, n \in \mathbb{Z}\}$ , and of the functional value  $\varepsilon_n$  of its innovation process:

$$X_n = \sum_{j=1}^{\infty} \sqrt{C_j} \frac{\langle X_n, \phi_j \rangle_H}{\sqrt{C_j}} \phi_j = \sum_{j=1}^{\infty} \sqrt{C_j} \eta_j(n) \phi_j, \quad (5)$$

$$\varepsilon_n = \sum_{j=1}^{\infty} \sigma_j \frac{\langle \varepsilon_n, \phi_j \rangle_H}{\sigma_j} \phi_j = \sum_{j=1}^{\infty} \sigma_j \tilde{\eta}_j(n) \phi_j, \quad (6)$$

where

$$\eta_j(n) = \frac{\langle X_n, \phi_j \rangle_H}{\sqrt{C_j}} = \frac{X_{n,j}}{\sqrt{C_j}}, \quad \tilde{\eta}_j(n) = \frac{\langle \varepsilon_n, \phi_j \rangle_H}{\sigma_j} = \frac{\varepsilon_{n,j}}{\sigma_j}, \quad n \in \mathbb{Z}, \quad j \geq 1.$$

Here, under **Assumptions A1–A2**, for  $C_\varepsilon = \mathbb{E} \{ \varepsilon_n \otimes \varepsilon_n \} = \mathbb{E} \{ \varepsilon_0 \otimes \varepsilon_0 \}$ ,  $n \in \mathbb{Z}$ ,

$$C_\varepsilon(\phi_j) = \sigma_j^2 \phi_j, \quad j \geq 1,$$

where, as before,  $\{\phi_j, j \geq 1\}$  denotes the system of eigenvectors of the autocovariance operator  $C$ , and

$$\sum_{j=1}^{\infty} \sigma_j^2 = \sigma_\varepsilon^2 = \mathbb{E} \{ \|\varepsilon_n\|_H^2 \},$$

for all  $n \in \mathbb{Z}$ .

The following lemma provides the convergence, in the seminorm of  $\mathcal{L}_H^2(\Omega, \mathcal{A}, \mathcal{P})$ , of the series expansions (5)–(6).

**Lemma 1** *Let  $X = \{X_n, n \in \mathbb{Z}\}$  be a zero-mean ARH(1) process. Under **Assumptions A1–A2**, for any  $n \in \mathbb{Z}$ , the following limit holds*

$$\lim_{M \rightarrow \infty} \mathbb{E} \left\{ \left\| X_n - \hat{X}_{n,M} \right\|_H^2 \right\} = 0,$$

where  $\hat{X}_{n,M} = \sum_{j=1}^M \sqrt{C_j} \eta_j(n) \phi_j$ . Furthermore,

$$\lim_{M \rightarrow \infty} \left\| \mathbb{E} \left\{ \left( X_n - \hat{X}_{n,M} \right) \otimes \left( X_n - \hat{X}_{n,M} \right) \right\} \right\|_{\mathcal{S}(H)}^2 = 0.$$

*Similar assertions hold for the biorthogonal series representation*

$$\varepsilon_n = \sum_{j=1}^{\infty} \sigma_j \frac{\langle \varepsilon_n, \phi_j \rangle_H}{\sigma_j} \phi_j = \sum_{j=1}^{\infty} \sigma_j \tilde{\eta}_j(n) \phi_j.$$

*Proof.*

Under **Assumption A1**, from the trace property of  $C$ , the sequence

$$\left\{ \widehat{X}_{n,M} = \sum_{j=1}^M \sqrt{C_j} \eta_j(n) \phi_j, \quad M \geq 1 \right\}$$

satisfies, for  $M$  sufficiently large, and  $L > 0$ , arbitrary,

$$\begin{aligned} \|\widehat{X}_{n,M+L} - \widehat{X}_{n,M}\|_{\mathcal{L}_H^2(\Omega, \mathcal{A}, P)}^2 &= \mathbb{E} \left\{ \|\widehat{X}_{n,M+L} - \widehat{X}_{n,M}\|_H^2 \right\} \\ &= \sum_{j=M+1}^{M+L} \sum_{k=M+1}^{M+L} \sqrt{C_j} \sqrt{C_k} \mathbb{E} \{ \eta_j(n) \eta_k(n) \} \langle \phi_j, \phi_k \rangle_H \\ &= \sum_{j=M+1}^{M+L} C_j \rightarrow 0, \quad \text{when } M \rightarrow \infty, \end{aligned} \quad (7)$$

since, under **Assumption A1**,  $\sum_{j=1}^{\infty} C_j < \infty$ . Hence,  $\left\{ \sum_{j=1}^M C_j, \quad M \geq 1 \right\}$  is a Cauchy sequence. Thus,

$$\lim_{M \rightarrow \infty} \sum_{j=M+1}^{M+L} C_j = 0,$$

for  $L > 0$  arbitrary. From equation (7),

$$\left\{ \widehat{X}_{n,M} = \sum_{j=1}^M \sqrt{C_j} \eta_j(n) \phi_j, \quad M \geq 1 \right\}$$

is also a Cauchy sequence in  $\mathcal{L}_H^2(\Omega, \mathcal{A}, P)$ . Thus, the sequence  $\{\widehat{X}_{n,M}, M \geq 1\}$  has finite limit in  $\mathcal{L}_H^2(\Omega, \mathcal{A}, P)$ , for all  $n \in \mathbb{Z}$ .

Furthermore,

$$\begin{aligned} \lim_{M \rightarrow \infty} \mathbb{E} \left\{ \left\| X_n - \widehat{X}_{n,M} \right\|_H^2 \right\} &= \mathbb{E} \left\{ \|X_n\|_H^2 \right\} + \lim_{M \rightarrow \infty} \sum_{j=1}^M \sum_{h=1}^M \sqrt{C_j} \sqrt{C_h} \mathbb{E} \{ \eta_j(n) \eta_h(n) \} \langle \phi_j, \phi_h \rangle_H \\ &\quad - 2 \lim_{M \rightarrow \infty} \sum_{j=1}^M \sqrt{C_j} \mathbb{E} \{ \langle X_n, \eta_j(n) \phi_j \rangle_H \} = \sigma_X^2 \\ &\quad - \lim_{M \rightarrow \infty} \sum_{j=1}^M C_j = 0. \end{aligned} \quad (8)$$

In the derivation of the identities in (7)–(8), we have applied that, for every  $j, h \geq 1$ ,

$$\begin{aligned}
C(\phi_j) &= C_j \phi_j, & \sigma_X^2 &= \mathbb{E} \{ \|X_n\|_H^2 \} = \sum_{j=1}^{\infty} C_j < +\infty, & \langle \phi_j, \phi_h \rangle_H &= \delta_{j,h}, \\
\mathbb{E} \{ \eta_j(n) \eta_h(n) \} &= \delta_{j,h}, & \mathbb{E} \{ \langle X_n, \eta_j(n) \phi_j \rangle_H \} &= \sqrt{C_j}.
\end{aligned} \tag{9}$$

Moreover, from identities in (9),

$$\begin{aligned}
& \left\| \mathbb{E} \left\{ \left( X_n - \lim_{M \rightarrow \infty} \widehat{X}_{n,M} \right) \otimes \left( X_n - \lim_{M \rightarrow \infty} \widehat{X}_{n,M} \right) \right\} \right\|_{\mathcal{S}(H)}^2 \\
&= \left\| \mathbb{E} \{ X_n \otimes X_n \} + \lim_{M \rightarrow \infty} \sum_{j=1}^M \sum_{h=1}^M \sqrt{C_j} \sqrt{C_h} \phi_j \otimes \phi_h \mathbb{E} \{ \eta_j(n) \eta_h(n) \} \right. \\
&\quad \left. - 2 \lim_{M \rightarrow \infty} \sum_{j=1}^M \mathbb{E} \left\{ X_n \otimes \sqrt{C_j} \eta_j(n) \phi_j \right\} \right\|_{\mathcal{S}(H)}^2 \\
&= \left\| \mathbb{E} \{ X_n \otimes X_n \} + \lim_{M \rightarrow \infty} \left[ \sum_{j=1}^M C_j \phi_j \otimes \phi_j - 2 \sum_{j=1}^M C_j \phi_j \otimes \phi_j \right] \right\|_{\mathcal{S}(H)}^2 \\
&= \left\| \mathbb{E} \{ X_n \otimes X_n \} - \lim_{M \rightarrow \infty} \sum_{j=1}^M C_j \phi_j \otimes \phi_j \right\|_{\mathcal{S}(H)}^2 = 0.
\end{aligned} \tag{10}$$

In a similar way, we can derive the convergence to  $\varepsilon_n$ , in  $\mathcal{L}_H^2(\Omega, \mathcal{A}, \mathcal{P})$ , of the series  $\sum_{j=1}^{\infty} \sigma_j \tilde{\eta}_j(n) \phi_j$ , for every  $n \in \mathbb{Z}$ , since  $\varepsilon$  is assumed to be strong-white noise, and hence, its covariance operator  $C_\varepsilon$  is in the trace class. We can also obtain an analogous to equation (10). ■

In equations (5)–(6), for every  $n \in \mathbb{Z}$ ,

$$\begin{aligned}
\mathbb{E} \{ \eta_j(n) \} &= 0, & \mathbb{E} \{ \eta_j(n) \eta_h(n) \} &= \delta_{j,h}, & j, h &\geq 1, & n \in \mathbb{Z}, \\
\mathbb{E} \{ \tilde{\eta}_j(n) \} &= 0, & \mathbb{E} \{ \tilde{\eta}_j(n) \tilde{\eta}_h(n) \} &= \delta_{j,h}, & j, h &\geq 1, & n \in \mathbb{Z}.
\end{aligned} \tag{11}$$

Note that, from **Assumption A2** for each  $j \geq 1$ ,  $\{X_{n,j}, n \in \mathbb{Z}\}$  in equation (2) defines a stationary and invertible AR(1) process. In addition, from equations (5) and (9), for every  $n \in \mathbb{Z}$ ,

and  $j, p \geq 1$ ,

$$\begin{aligned}
X_n &= \sum_{j=1}^{\infty} X_{n,j} \phi_j, \\
\mathbb{E} \{X_{n,j} X_{n,p}\} &= \sum_{k=0}^{\infty} \sum_{h=0}^{\infty} \rho_j^k \rho_p^h \mathbb{E} \{\varepsilon_{n-k,j} \varepsilon_{n-h,p}\} = \delta_{j,p} \sum_{k=0}^{\infty} \rho_j^{2k} \sigma_j^2 = \delta_{j,p} \frac{\sigma_j^2}{1 - \rho_j^2}, \\
\mathbb{E} \left\{ \|X_n\|_H^2 \right\} &= \sum_{j=1}^{\infty} \mathbb{E} \{X_{n,j}^2\} = \sum_{j=1}^{\infty} \langle C(\phi_j), \phi_j \rangle_H = \sum_{j=1}^{\infty} C_j = \sigma_X^2 < \infty,
\end{aligned} \tag{12}$$

which implies that

$$C_j = \frac{\sigma_j^2}{1 - \rho_j^2}, \quad j \geq 1.$$

In particular, we obtain, for each  $j \geq 1$ , and for every  $n \in \mathbb{Z}$ ,

$$\begin{aligned}
\mathbb{E} \{ \eta_j(n) \eta_j(n+1) \} &= \mathbb{E} \left\{ \frac{X_{n,j}}{\sqrt{C_j}} \frac{X_{n+1,j}}{\sqrt{C_j}} \right\} = \frac{\mathbb{E} \{ X_{n,j} X_{n+1,j} \}}{C_j} \\
&= \frac{\sum_{k=0}^{\infty} \sum_{h=0}^{\infty} \rho_j^{k+h} \mathbb{E} \{ \varepsilon_{n-k,j} \varepsilon_{n+1-h,j} \}}{C_j} \\
&= \frac{\sum_{k=0}^{\infty} \rho_j^{2k+1} \sigma_j^2}{C_j} = \frac{\sigma_j^2}{C_j} \frac{\rho_j}{1 - \rho_j^2} = \rho_j.
\end{aligned} \tag{13}$$

*Remark 3* From equation (2) and [Lemma 1](#), keeping in mind that

$$C_j = \frac{\sigma_j^2}{1 - \rho_j^2}, \quad j \geq 1,$$

the following invertible and stationary AR(1) process can be defined:

$$\eta_j(n) = \rho_j \eta_j(n-1) + \sqrt{1 - \rho_j^2} \tilde{\eta}_j(n), \quad 0 < \rho_j^2 \leq \rho_j < 1, \tag{14}$$

where, for each  $j \geq 1$ ,  $\{\eta_j(n), n \in \mathbb{Z}\}$  and  $\{\tilde{\eta}_j(n), n \in \mathbb{Z}\}$  are respectively introduced in equations (5)-(6). In the following, for each  $j \geq 1$ , we assume that

$$\mathbb{E} \left\{ (\tilde{\eta}_j(n))^4 \right\} < \infty, \quad n \in \mathbb{Z},$$

to ensure ergodicity for all second-order moments, in the mean-square sense; see, e.g., [Hamilton, 1994, pp. 192–193].

Furthermore,

$$\begin{aligned}
D = \mathbb{E}\{X_n \otimes X_{n+1}\} &= \sum_{j=1}^{\infty} \sum_{p=1}^{\infty} \mathbb{E}\{\langle X_n, \phi_j \rangle_H \langle X_{n+1}, \phi_p \rangle_H\} \phi_j \otimes \phi_p \\
&= \sum_{j=1}^{\infty} \sum_{p=1}^{\infty} \sqrt{C_j} \sqrt{C_p} \frac{\mathbb{E}\{\langle X_n, \phi_j \rangle_H \langle X_{n+1}, \phi_p \rangle_H\}}{\sqrt{C_j} \sqrt{C_p}} \phi_j \otimes \phi_p \\
&= \sum_{j=1}^{\infty} \sum_{p=1}^{\infty} \sqrt{C_j} \sqrt{C_p} \mathbb{E}\{\eta_j(n) \eta_p(n+1)\} \phi_j \otimes \phi_p.
\end{aligned}$$

*Remark 4* In particular, **Assumption A2** holds if the following orthogonality condition is satisfied, for all  $n \in \mathbb{Z}$  and  $j, p \geq 1$ ,

$$\mathbb{E}\{\eta_j(n) \eta_p(n+1)\} = \delta_{j,p},$$

where  $\delta_{j,p}$  denotes the Kronecker Delta function. In practice, unconditional bases, e.g., wavelet bases, lead to a sparse representation for functional data; see, e.g., Nason [2008]; Ogden [1997]; Vidakovic [1998] for statistically-oriented treatments. Wavelet bases are also designed for sparse representation of kernels defining integral operators, in  $L^2$  spaces with respect to a suitable measure (see Mallat [2009]). The Discrete Wavelet Transform (DWT) approximately decorrelates or whitens data (see Vidakovic [1998]). In particular, operators  $C$  and  $D$  could admit an almost diagonal representation with respect to the self-tensorial tensorial product of a suitable wavelet basis.

### 3 Estimation and prediction results

A componentwise estimator of the autocorrelation operator and of the associated ARH(1) plug-in predictor are formulated in this section. Their convergence to the corresponding theoretical functional values are derived in the spaces  $\mathcal{L}_{\mathcal{S}(H)}^2(\Omega, \mathcal{A}, \mathcal{P})$  and  $\mathcal{L}_H(\Omega, \mathcal{A}, \mathcal{P})$ , respectively. Their consistency in the spaces  $\mathcal{S}(H)$  and  $H$  then follows.

From equation (3), for each  $j \geq 1$ , and for a given sample size  $n$ , one can consider the usual respective moment-based estimators  $\widehat{D}_{n,j}$  and  $\widehat{C}_{n,j}$  of  $D_j$  and  $C_j$ , in the AR(1) framework, given by

$$\widehat{D}_{n,j} = \frac{1}{n-1} \sum_{i=0}^{n-2} X_{i,j} X_{i+1,j}, \quad \widehat{C}_{n,j} = \frac{1}{n} \sum_{i=0}^{n-1} X_{i,j}^2.$$

The following truncated componentwise estimator of  $\rho$  is then formulated:

$$\widehat{\rho}_{k_n} = \sum_{j=1}^{k_n} \widehat{\rho}_{n,j} \phi_j \otimes \phi_j, \quad (15)$$

where, for each  $j \geq 1$ ,

$$\widehat{\rho}_{n,j} = \frac{\widehat{D}_{n,j}}{\widehat{C}_{n,j}} = \frac{\frac{1}{n-1} \sum_{i=0}^{n-2} X_{i,j} X_{i+1,j}}{\frac{1}{n} \sum_{i=0}^{n-1} X_{i,j}^2} = \frac{n}{n-1} \frac{\sum_{i=0}^{n-2} X_{i,j} X_{i+1,j}}{\sum_{i=0}^{n-1} X_{i,j}^2}. \quad (16)$$

Here, the truncation parameter indicates that we have considered the first  $k_n$  eigenvectors associated with the first  $k_n$  eigenvalues, arranged in decreasing order of their modulus magnitude. Furthermore,  $k_n$  is such that

$$\lim_{n \rightarrow \infty} k_n = \infty, \quad \frac{k_n}{n} < 1, \quad n \geq 2. \quad (17)$$

The following additional condition will be assumed on  $k_n$  for the derivation of the subsequent results:

**Assumption A3.** The truncation parameter  $k_n$  in (15) is such that

$$\lim_{n \rightarrow \infty} C_{k_n} \sqrt{n} = \infty.$$

*Remark 5* **Assumption A3** has also been considered in [Bosq, 2000, p. 217], to ensure weak consistency of the proposed estimator of  $\rho$ , as well as, in [Mas, 1999, Proposition 4], in the derivation of asymptotic normality.

From **Remark 3**, for each  $j \geq 1$ ,  $\eta_j = \{\eta_j(n), n \in \mathbb{Z}\}$  in equation (14) defines a stationary and invertible AR(1) process, ergodic in the mean-square sense; see, e.g., Bartlett [1946]. Therefore, in view of equations (11) and (13), for each  $j \geq 1$ , there exist two positive constants  $K_{j,1}$  and  $K_{j,2}$  such that the following identities hold:

$$\lim_{n \rightarrow \infty} \frac{\mathbb{E} \left\{ \left[ 1 - \frac{1}{n} \sum_{i=0}^{n-1} \eta_j^2(i) \right]^2 \right\}}{\frac{1}{n}} = K_{j,1}, \quad (18)$$

$$\lim_{n \rightarrow \infty} \frac{\mathbb{E} \left\{ \left[ \rho_j - \frac{1}{n-1} \sum_{i=0}^{n-2} \eta_j(i) \eta_j(i+1) \right]^2 \right\}}{\frac{1}{n}} = K_{j,2}. \quad (19)$$

Equations (18)-(19) imply, for  $n$  sufficiently large,

$$\text{Var} \left\{ \frac{1}{n} \sum_{i=0}^{n-1} \eta_j^2(i) \right\} \leq \frac{\tilde{K}_{j,1}}{n}, \quad (20)$$

$$\text{Var} \left\{ \frac{1}{n-1} \sum_{i=0}^{n-2} \eta_j(i) \eta_j(i+1) \right\} \leq \frac{\tilde{K}_{j,2}}{n}, \quad (21)$$

for certain positive constants  $\tilde{K}_{j,1}$  and  $\tilde{K}_{j,2}$ , for each  $j \geq 1$ . Equivalently, for  $n$  sufficiently large,

$$\mathbb{E} \left\{ \left( 1 - \frac{1}{n} \sum_{i=0}^{n-1} \eta_j^2(i) \right)^2 \right\} \leq \frac{\tilde{K}_{j,1}}{n}, \quad (22)$$

$$\mathbb{E} \left\{ \left( \rho_j - \frac{1}{n-1} \sum_{i=0}^{n-1} \eta_j(i) \eta_j(i+1) \right)^2 \right\} \leq \frac{\tilde{K}_{j,2}}{n}, \quad (23)$$

The following assumption is now considered.

**Assumption A4.** We assume that

$$S = \sup_{j \geq 1} (\tilde{K}_{j,1} + \tilde{K}_{j,2}) < \infty.$$

*Remark 6* From equation (16), applying the Cauchy-Schwarz's inequality, we obtain, for each  $j \geq 1$ ,

$$\begin{aligned} |\hat{\rho}_{n,j}| &= \frac{n}{n-1} \left| \frac{\sum_{i=0}^{n-2} X_{i,j} X_{i+1,j}}{\sum_{i=0}^{n-1} X_{i,j}^2} \right| \leq \frac{n}{n-1} \frac{\sqrt{\sum_{i=0}^{n-2} X_{i,j}^2 \sum_{i=0}^{n-2} X_{i+1,j}^2}}{\sum_{i=0}^{n-1} X_{i,j}^2} \\ &\leq \frac{n}{n-1} \sqrt{\frac{\sum_{i=0}^{n-2} X_{i+1,j}^2}{\sum_{i=0}^{n-1} X_{i,j}^2}} \leq \frac{n}{n-1} \text{ a.s.} \end{aligned} \quad (24)$$

### 3.1 Convergence in $\mathcal{L}_{\mathcal{S}(H)}^2(\Omega, \mathcal{A}, \mathcal{P})$

Next, the convergence of  $\widehat{\rho}_{k_n}$  to  $\rho$ , in the space  $\mathcal{L}_{\mathcal{S}(H)}^2(\Omega, \mathcal{A}, \mathcal{P})$ , is derived under the setting of conditions formulated in the previous sections.

**Proposition 1** *Let  $X = \{X_n, n \in \mathbb{Z}\}$  be a zero-mean standard ARH(1) process. Under **Assumptions A1–A4**, the following limit holds:*

$$\lim_{n \rightarrow \infty} \|\rho - \widehat{\rho}_{k_n}\|_{\mathcal{L}_{\mathcal{S}(H)}^2(\Omega, \mathcal{A}, \mathcal{P})}^2 = 0. \quad (25)$$

Specifically,

$$\|\rho - \widehat{\rho}_{k_n}\|_{\mathcal{L}_{\mathcal{S}(H)}^2(\Omega, \mathcal{A}, \mathcal{P})}^2 \leq g(n), \quad \text{with } g(n) = \mathcal{O}\left(\frac{1}{C_{k_n}^2 n}\right), \quad n \rightarrow \infty. \quad (26)$$

*Remark 7* [Bosq, 2000, Corollary 4.3] can be applied to obtain weak convergence results, in terms of weak expectation, using the empirical eigenvectors. See definition of weak expectation at the beginning of [Bosq, 2000, Section 1.3, p. 27]).

*Proof.* For each  $j \geq 1$ , the following almost surely inequality is satisfied:

$$\begin{aligned} |\rho_j - \widehat{\rho}_{n,j}| &= \left| \frac{D_j}{C_j} - \frac{\widehat{D}_{n,j}}{\widehat{C}_{n,j}} \right| = \left| \frac{D_j \widehat{C}_{n,j} - \widehat{D}_{n,j} C_j}{C_j \widehat{C}_{n,j}} \right| \\ &= \left| \frac{D_j \widehat{C}_{n,j} - \widehat{D}_{n,j} C_j + \widehat{C}_{n,j} \widehat{D}_{n,j} - \widehat{C}_{n,j} \widehat{D}_{n,j}}{C_j \widehat{C}_{n,j}} \right| \\ &= \left| \frac{D_j - \widehat{D}_{n,j}}{C_j} + \frac{\widehat{C}_{n,j} - C_j}{C_j} \frac{\widehat{D}_{n,j}}{\widehat{C}_{n,j}} \right| \leq \frac{1}{C_j} \left( |\widehat{\rho}_{n,j}| |C_j - \widehat{C}_{n,j}| + |D_j - \widehat{D}_{n,j}| \right). \end{aligned}$$

Thus, under **Assumptions A1–A2**, from equation (24), for each  $j \geq 1$ ,

$$\begin{aligned} (\rho_j - \widehat{\rho}_{n,j})^2 &\leq \frac{1}{C_j^2} \left( |\widehat{\rho}_{n,j}| |C_j - \widehat{C}_{n,j}| + |D_j - \widehat{D}_{n,j}| \right)^2 \\ &\leq \frac{2}{C_j^2} \left( (\widehat{\rho}_{n,j})^2 (C_j - \widehat{C}_{n,j})^2 + (D_j - \widehat{D}_{n,j})^2 \right) \\ &\leq \frac{2}{C_j^2} \left( \left( \frac{n}{n-1} \right)^2 (C_j - \widehat{C}_{n,j})^2 + (D_j - \widehat{D}_{n,j})^2 \right) \text{ a.s.,} \end{aligned}$$

which implies, for each  $j \geq 1$ ,

$$\mathbb{E} \left\{ (\rho_j - \widehat{\rho}_{n,j})^2 \right\} \leq \frac{2}{C_j^2} \left( \left( \frac{n}{n-1} \right)^2 \mathbb{E} \left\{ (C_j - \widehat{C}_{n,j})^2 \right\} + \mathbb{E} \left\{ (D_j - \widehat{D}_{n,j})^2 \right\} \right). \quad (27)$$

Under **Assumption A2**, from equations (15) and (27),

$$\begin{aligned} \|\rho - \widehat{\rho}_{k_n}\|_{\mathcal{L}_{\mathcal{S}(H)}^2(\Omega, \mathcal{A}, \mathcal{P})}^2 &= \mathbb{E} \left\{ \|\rho - \widehat{\rho}_{k_n}\|_{\mathcal{S}(H)}^2 \right\} = \sum_{j=1}^{k_n} \mathbb{E} \left\{ (\rho_j - \widehat{\rho}_{n,j})^2 \right\} + \sum_{j=k_n+1}^{\infty} \mathbb{E} \left\{ \rho_j^2 \right\} \\ &\leq \sum_{j=1}^{k_n} \frac{2}{C_j^2} \left( \left( \frac{n}{n-1} \right)^2 \mathbb{E} \left\{ (C_j - \widehat{C}_{n,j})^2 \right\} \right. \\ &\quad \left. + \mathbb{E} \left\{ (D_j - \widehat{D}_{n,j})^2 \right\} \right) + \sum_{j=k_n+1}^{\infty} \rho_j^2 \\ &\leq \frac{2}{C_{k_n}^2} \sum_{j=1}^{k_n} \left( \frac{n}{n-1} \right)^2 \left( \mathbb{E} \left\{ (C_j - \widehat{C}_{n,j})^2 \right\} \right. \\ &\quad \left. + \mathbb{E} \left\{ (D_j - \widehat{D}_{n,j})^2 \right\} \right) + \sum_{j=k_n+1}^{\infty} \rho_j^2 \\ &\leq \frac{2 \left( \frac{n}{n-1} \right)^2}{C_{k_n}^2} \sum_{j=1}^{k_n} \left( \mathbb{E} \left\{ (C_j - \widehat{C}_{n,j})^2 \right\} + \mathbb{E} \left\{ (D_j - \widehat{D}_{n,j})^2 \right\} \right) \\ &\quad + \sum_{j=k_n+1}^{\infty} \rho_j^2. \end{aligned} \quad (28)$$

Furthermore, from (5) and (16), for each  $j \geq 1$ ,

$$\widehat{C}_{n,j} = \frac{1}{n} \sum_{i=0}^{n-1} X_{i,j}^2 = \frac{1}{n} \sum_{i=0}^{n-1} C_j \eta_j^2(i), \quad (29)$$

$$\widehat{D}_{n,j} = \frac{1}{n-1} \sum_{i=0}^{n-2} X_{i,j} X_{i+1,j} = \frac{1}{n-1} \sum_{i=0}^{n-2} C_j \eta_j(i) \eta_j(i+1), \quad (30)$$

where, considering equation (4),

$$D_j = \mathbb{E} \{ X_{n,j} X_{n+1,j} \} = C_j \mathbb{E} \{ \eta_j(n) \eta_j(n+1) \} = C_j \rho_j, \quad (31)$$

for each  $j \geq 1$ . Equations (28)–(31) then lead to

$$\begin{aligned} \|\rho - \widehat{\rho}_{k_n}\|_{\mathcal{L}_{\mathcal{S}(H)}^2(\Omega, \mathcal{A}, \mathcal{P})}^2 &\leq \frac{2\left(\frac{n}{n-1}\right)^2}{C_{k_n}^2} \sum_{j=1}^{k_n} C_j^2 \left( \mathbb{E} \left\{ \left( 1 - \frac{1}{n} \sum_{i=0}^{n-1} \eta_j^2(i) \right)^2 \right\} \right. \\ &\quad \left. + \mathbb{E} \left\{ \left( \rho_j - \frac{1}{n-1} \sum_{i=0}^{n-2} \eta_j(i+1)\eta_j(i) \right)^2 \right\} \right) \\ &\quad + \sum_{j=k_n+1}^{\infty} \rho_j^2. \end{aligned}$$

For each  $j \geq 1$ , and for  $n$  sufficiently large, considering equations (22)–(23), under **Assumption A4**,

$$\begin{aligned} \mathbb{E} \left\{ \|\rho - \widehat{\rho}_{k_n}\|_{\mathcal{S}(H)}^2 \right\} &\leq \frac{2\left(\frac{n}{n-1}\right)^2}{C_{k_n}^2} \sum_{j=1}^{k_n} C_j^2 \left( \frac{\widetilde{K}_{j,1} + \widetilde{K}_{j,2}}{n} \right) + \sum_{j=k_n+1}^{\infty} \rho_j^2 \\ &\leq \frac{2S\left(\frac{n}{n-1}\right)^2}{C_{k_n}^2 n} \sum_{j=1}^{k_n} C_j^2 + \sum_{j=k_n+1}^{\infty} \rho_j^2. \end{aligned} \tag{32}$$

From the trace property of operator  $C$ ,

$$\lim_{n \rightarrow \infty} \sum_{j=1}^{k_n} C_j^2 = \sum_{j=1}^{\infty} C_j^2 < \infty, \tag{33}$$

and from the Hilbert–Schmidt property of  $\rho$ ,

$$\lim_{n \rightarrow \infty} \sum_{j=k_n+1}^{\infty} \rho_j^2 = 0. \tag{34}$$

Thus, in view of equations (32)–(34),

$$\|\rho - \widehat{\rho}_{k_n}\|_{\mathcal{L}_{\mathcal{S}(H)}^2(\Omega, \mathcal{A}, \mathcal{P})}^2 = \mathbb{E} \left\{ \|\rho - \widehat{\rho}_{k_n}\|_{\mathcal{S}(H)}^2 \right\} \leq g(n) = \mathcal{O} \left( \frac{1}{C_{k_n}^2 n} \right), \quad n \rightarrow \infty, \tag{35}$$

where

$$g(n) = \frac{2S\left(\frac{n}{n-1}\right)^2}{C_{k_n}^2 n} \sum_{j=1}^{k_n} C_j^2 + \sum_{j=k_n+1}^{\infty} \rho_j^2. \tag{36}$$

Under **Assumption A3**, equations (35)–(36) imply

$$\lim_{n \rightarrow \infty} \|\rho - \widehat{\rho}_{k_n}\|_{\mathcal{L}_{\mathcal{S}(H)}^2(\Omega, \mathcal{A}, \mathcal{P})}^2 = 0,$$

as we wanted to prove. ■

Note that consistency of  $\widehat{\rho}_{k_n}$  in the space  $\mathcal{S}(H)$  directly follows from equation (25) in [Proposition 1](#).

**Corollary 1** *Let  $X = \{X_n, n \in \mathbb{Z}\}$  be a zero-mean standard ARH(1) process. Under [Assumptions A1–A4](#), as long as  $n \rightarrow \infty$ ,*

$$\|\rho - \widehat{\rho}_{k_n}\|_{\mathcal{S}(H)} \xrightarrow{p} 0,$$

where, as usual,  $\xrightarrow{p}$  denotes the convergence in probability.

### 3.2 Consistency of the ARH(1) plug-in predictor.

Let us consider  $\mathcal{L}(H)$  the space of bounded linear operators on  $H$ , with the norm

$$\|\mathcal{A}\|_{\mathcal{L}(H)} = \sup_{x \in H} \frac{\|\mathcal{A}(x)\|_H}{\|x\|_H},$$

for every  $\mathcal{A} \in \mathcal{L}(H)$ . In particular, for each  $x \in H$ ,

$$\|\mathcal{A}(x)\|_H \leq \|\mathcal{A}\|_{\mathcal{L}(H)} \|x\|_H. \quad (37)$$

In the following, we denote by

$$\widehat{X}_n = \widehat{\rho}_{k_n}(X_{n-1}) \quad (38)$$

as usual, the ARH(1) plug-in predictor of  $X_n$ , as an estimator of the conditional expectation  $E\{X_n|X_{n-1}\} = \rho(X_{n-1})$ . The following proposition provides the consistency of  $\widehat{X}_n = \widehat{\rho}_{k_n}(X_{n-1})$  in  $H$ .

**Proposition 2** *Let  $X = \{X_n, n \in \mathbb{Z}\}$  be a zero-mean standard ARH(1) process. Under [Assumptions A1–A4](#),*

$$\lim_{n \rightarrow \infty} E\{\|(\rho - \widehat{\rho}_{k_n})(X_{n-1})\|_H\} = 0.$$

Specifically,

$$E\{\|(\rho - \widehat{\rho}_{k_n})(X_{n-1})\|_H\} \leq h(n), \quad h(n) = \mathcal{O}\left(\frac{1}{C_{k_n}\sqrt{n}}\right), \quad n \rightarrow \infty.$$

In particular,

$$\|(\rho - \widehat{\rho}_{k_n})(X_{n-1})\|_H \xrightarrow{p} 0,$$

where, as usual,  $\xrightarrow{p}$  denotes the convergence in probability.

*Proof.*

From (37) and [Proposition 1](#), for  $n$  sufficiently large, the following almost surely inequality holds:

$$\left\| \rho(X_{n-1}) - \widehat{X}_n \right\|_H \leq \|\rho - \widehat{\rho}_{k_n}\|_{\mathcal{L}(H)} \|X_{n-1}\|_H,$$

where, as given in equation (38),  $\widehat{X}_n = \widehat{\rho}_{k_n}(X_{n-1})$ . Thus,

$$\mathbb{E} \left\{ \left\| \rho(X_{n-1}) - \widehat{X}_n \right\|_H \right\} \leq \mathbb{E} \left\{ \|\rho - \widehat{\rho}_{k_n}\|_{\mathcal{L}(H)} \|X_{n-1}\|_H \right\}. \quad (39)$$

From the Cauchy-Schwarz's inequality, keeping in mind that, for a Hilbert-Schmidt operator  $\mathcal{K}$ , it always holds that  $\|\mathcal{K}\|_{\mathcal{L}(H)} \leq \|\mathcal{K}\|_{\mathcal{S}(H)}$ , we have from equation (39),

$$\begin{aligned} \mathbb{E} \left\{ \left\| X_n - \widehat{X}_n \right\|_H \right\} &\leq \sqrt{\mathbb{E} \left\{ \|\rho - \widehat{\rho}_{k_n}\|_{\mathcal{L}(H)}^2 \right\}} \sqrt{\mathbb{E} \left\{ \|X_{n-1}\|_H^2 \right\}} \\ &\leq \sqrt{\mathbb{E} \left\{ \|\rho - \widehat{\rho}_{k_n}\|_{\mathcal{S}(H)}^2 \right\}} \sqrt{\mathbb{E} \left\{ \|X_{n-1}\|_H^2 \right\}} \\ &= \sqrt{\mathbb{E} \left\{ \|\rho - \widehat{\rho}_{k_n}\|_{\mathcal{S}(H)}^2 \right\}} \sigma_X, \end{aligned} \quad (40)$$

where, as before,  $\sigma_X^2 = \mathbb{E} \left\{ \|X_{n-1}\|_H^2 \right\} = \sum_{j=1}^{\infty} C_j < \infty$ ,  $n \in \mathbb{Z}$  (see equation (9)).

Since from [Proposition 1](#) (see equation (26)),

$$\|\rho - \widehat{\rho}_{k_n}\|_{\mathcal{L}_{\mathcal{S}(H)}^2(\Omega, \mathcal{A}, \mathcal{P})} \leq g(n), \quad \text{with } g(n) = \mathcal{O} \left( \frac{1}{C_{k_n}^2 n} \right), \quad n \rightarrow \infty,$$

from equation (40), we obtain,

$$\mathbb{E} \left\{ \|(\rho - \widehat{\rho}_{k_n})(X_{n-1})\|_H \right\} \leq h(n),$$

where  $h(n) = \sigma_X \sqrt{g(n)}$ , with  $g(n)$  being given in (36). In particular, under [Assumption A3](#),

$$\lim_{n \rightarrow \infty} \mathbb{E} \left\{ \|(\rho - \widehat{\rho}_{k_n})(X_{n-1})\|_H \right\} = 0,$$

which implies that

$$\|(\rho - \widehat{\rho}_{k_n})(X_{n-1})\|_H = \left\| \rho(X_{n-1}) - \widehat{X}_n \right\|_H \xrightarrow{p} 0, \quad n \rightarrow \infty.$$

■

## 4 The Gaussian case

In this section, we prove that, in the Gaussian ARH(1) context, **Assumptions A1–A2** and **A4** also hold. From equation (11), for  $n \geq 1$ ,

$$\mathbb{E} \left\{ \frac{\sum_{i=0}^{n-1} \eta_j^2(i)}{n} \right\} = 1.$$

Furthermore, for each  $j \geq 1$  and  $n \geq 2$ , the  $n \times 1$  random vector  $\boldsymbol{\eta}_j^T = (\eta_j(0), \dots, \eta_j(n-1))$  follows a Multivariate Normal distribution with null mean vector, and covariance matrix

$$\boldsymbol{\Sigma} = \begin{pmatrix} 1 & \rho_j & 0 & \dots & \dots & 0 \\ \rho_j & 1 & \rho_j & 0 & \dots & 0 \\ 0 & \rho_j & 1 & \rho_j & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \rho_j & 1 \end{pmatrix}_{n \times n}. \quad (41)$$

It is well-known (see, for example, Gurland [1956]) that the variance of a quadratic form defined from a multivariate Gaussian vector  $\mathbf{y} \sim N(\boldsymbol{\mu}, \boldsymbol{\Lambda})$ , and a symmetric matrix  $\mathbf{Q}$  is given by:

$$\text{Var} \{ \mathbf{y}^T \mathbf{Q} \mathbf{y} \} = 2\text{Tr}(\mathbf{Q} \boldsymbol{\Lambda} \mathbf{Q} \boldsymbol{\Lambda}) + 4\boldsymbol{\mu}^T \mathbf{Q} \boldsymbol{\Lambda} \mathbf{Q} \boldsymbol{\mu}. \quad (42)$$

For each  $j \geq 1$ , applying equation (42), with  $\mathbf{y} = \boldsymbol{\eta}_j$ ,  $\boldsymbol{\Lambda} = \boldsymbol{\Sigma}$  in (41), and  $\mathbf{Q} = \mathbf{I}d_n$ , the  $n \times n$  identity matrix, keeping in mind  $\mathbb{E} \{ \eta_j(i) \eta_j(i+1) \} = \rho_j$ , for every  $i \in \mathbb{Z}$ ,

$$\text{Var} \{ \boldsymbol{\eta}_j^T \mathbf{I}d_n \boldsymbol{\eta}_j \} = \text{Var} \left\{ \sum_{i=0}^{n-1} \eta_j^2(i) \right\} = 2\text{Tr}(\boldsymbol{\Sigma} \boldsymbol{\Sigma}) = 2(n + 2(n-1)\rho_j^2). \quad (43)$$

Furthermore, from equation (43), for each  $j \geq 1$ ,

$$\text{Var} \left\{ \frac{\sum_{i=0}^{n-1} \eta_j^2(i)}{n} \right\} = \frac{2}{n^2} (n + 2(n-1)\rho_j^2) = \frac{2}{n} + 4 \left( \frac{1}{n} - \frac{1}{n^2} \right) \rho_j^2. \quad (44)$$

We then obtain, from equation (44),

$$\begin{aligned}
\lim_{n \rightarrow \infty} \text{Var} \left\{ \frac{\sum_{i=0}^{n-1} \eta_j^2(i)}{n} \right\} &= \lim_{n \rightarrow \infty} \text{E} \left\{ \left( 1 - \frac{\sum_{i=0}^{n-1} \eta_j^2(i)}{n} \right)^2 \right\} \\
&= \lim_{n \rightarrow \infty} \frac{2}{n} + 4 \left( \frac{1}{n} - \frac{1}{n^2} \right) \rho_j^2 = 0.
\end{aligned} \tag{45}$$

Equation (45) leads to

$$\lim_{n \rightarrow \infty} \frac{\text{Var} \left\{ \frac{\sum_{i=0}^{n-1} \eta_j^2(i)}{n} \right\}}{\frac{1}{n}} = 2 + 4\rho_j^2.$$

Hence, for each  $j \geq 1$ ,  $K_{j,1}$  in equation (18) is given by

$$K_{j,1} = 2 + 4\rho_j^2,$$

and, from equation (44),

$$\text{Var} \left\{ \frac{\sum_{i=0}^{n-1} \eta_j^2(i)}{n} \right\} \leq 2 + 4 \left( \frac{1}{n} - \frac{1}{n^2} \right) \rho_j^2 \leq 2 + 4\rho_j^2 \leq 6.$$

Thus, for every  $j \geq 1$ ,  $\tilde{K}_{j,1}$  in equation (20) satisfies

$$\tilde{K}_{j,1} \leq 6.$$

*Remark 8* Note that, from [Lemma 1](#), for each  $j \geq 1$  and  $i \in \mathbb{Z}$ ,

$$\text{E} \{ \tilde{\eta}_j^4(i) \} = 3.$$

Thus, the assumption considered in [Remark 3](#) holds, and for each  $j \geq 1$ , the  $AR(1)$  process  $\eta_j = \{\eta_j(n), n \in \mathbb{Z}\}$  is ergodic for all second-order moments, in the mean-square sense; see [[Hamilton, 1994](#), pp. 192–193].

For  $n \geq 2$ , and for each  $j \geq 1$ , we are now going to compute  $K_{j,2}$  in (19). The  $(n-1) \times 1$  random

vectors

$$\boldsymbol{\eta}_j^* = (\eta_j(0), \dots, \eta_j(n-2))^T, \quad \boldsymbol{\eta}_j^{**} = (\eta_j(1), \dots, \eta_j(n-1))^T$$

are multivariate Normal distributed, with null mean vector, and covariance matrix

$$\tilde{\boldsymbol{\Sigma}} = \begin{pmatrix} 1 & \rho_j & 0 & \dots & \dots & 0 \\ \rho_j & 1 & \rho_j & 0 & \dots & 0 \\ 0 & \rho_j & 1 & \rho_j & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \dots & 0 & 0 & \rho_j & 1 \end{pmatrix}_{(n-1) \times (n-1)}. \quad (46)$$

From equation (13), for each  $j \geq 1$ ,

$$\mathbb{E} \left\{ \sum_{i=0}^{n-2} \eta_j(i) \eta_j(i+1) \right\} = \sum_{i=0}^{n-2} \rho_j = (n-1) \rho_j = \text{Tr} \left( \mathbb{E} \{ \boldsymbol{\eta}_j^* [\boldsymbol{\eta}_j^{**}]^T \} \right), \quad (47)$$

where

$$\mathbb{E} \{ \boldsymbol{\eta}_j^* [\boldsymbol{\eta}_j^{**}]^T \} = \mathbb{E} \{ \boldsymbol{\eta}_j^* \otimes \boldsymbol{\eta}_j^{**} \} = \rho_j \mathbf{I}_{n-1}, \quad (48)$$

with, as before,  $\mathbf{I}_{n-1}$  denoting the  $(n-1) \times (n-1)$  identity matrix.

However, the variance of

$$\sum_{i=0}^{n-2} \eta_j(i) \eta_j(i+1)$$

depends greatly on the distribution of  $\boldsymbol{\eta}_j^*$  and  $\boldsymbol{\eta}_j^{**}$ . In the Gaussian case, keeping in mind that

$$\boldsymbol{\eta}_j^* = (\eta_j(0), \dots, \eta_j(n-2))^T, \quad \boldsymbol{\eta}_j^{**} = (\eta_j(1), \dots, \eta_j(n-1))^T$$

are zero-mean multivariate Normal distributed vectors with covariance matrix  $\tilde{\boldsymbol{\Sigma}}$  given in (46), and having cross-covariance matrix in (48), we can compute the variance of  $\sum_{i=0}^{n-2} \eta_j(i) \eta_j(i+1)$ , from (47)–(48), as follows. First,

$$\begin{aligned} \text{Var} \{ [\boldsymbol{\eta}_j^*]^T \mathbf{I}_{n-1} \boldsymbol{\eta}_j^{**} \} &= \mathbb{E} \{ [\boldsymbol{\eta}_j^*]^T \mathbf{I}_{n-1} \boldsymbol{\eta}_j^{**} [\boldsymbol{\eta}_j^*]^T \mathbf{I}_{n-1} \boldsymbol{\eta}_j^{**} \} \\ &\quad - \left( \mathbb{E} \{ [\boldsymbol{\eta}_j^*]^T \mathbf{I}_{n-1} \boldsymbol{\eta}_j^{**} \} \right)^2. \end{aligned}$$

This can be rewritten as

$$\sum_{i=0}^{n-2} \sum_{p=0}^{n-2} \mathbb{E} \{ \eta_j(i) \eta_j(i+1) \eta_j(p) \eta_j(p+1) \} - \left( \mathbb{E} \{ [\boldsymbol{\eta}_j^*]^T \mathbf{I}_{n-1} \boldsymbol{\eta}_j^{**} \} \right)^2,$$

which is equal to

$$\begin{aligned}
& \sum_{i=0}^{n-2} \mathbb{E} \{ \eta_j(i) \eta_j(i+1) \} \sum_{p=0}^{n-2} \mathbb{E} \{ \eta_j(p) \eta_j(p+1) \} + \sum_{i=0}^{n-2} \sum_{p=0}^{n-2} \mathbb{E} \{ \eta_j(i) \eta_j(p) \} \mathbb{E} \{ \eta_j(i+1) \eta_j(p+1) \} \\
& + \sum_{i=0}^{n-2} \sum_{p=0}^{n-2} \mathbb{E} \{ \eta_j(i) \eta_j(p+1) \} \mathbb{E} \{ \eta_j(i+1) \eta_j(p) \} \\
& - \left( \mathbb{E} \{ [\boldsymbol{\eta}_j^*]^T \mathbf{I} \mathbf{d}_{n-1} \boldsymbol{\eta}_j^{**} \} \right)^2.
\end{aligned}$$

This then reduces to

$$\begin{aligned}
& [\text{Tr}(\mathbb{E} \{ \boldsymbol{\eta}_j^* \otimes \boldsymbol{\eta}_j^{**} \})]^2 + \text{Tr}(\tilde{\boldsymbol{\Sigma}} \tilde{\boldsymbol{\Sigma}}) \\
& + \text{Tr}(\mathbb{E} \{ \boldsymbol{\eta}_j^* \otimes \boldsymbol{\eta}_j^{**} \} [\mathbb{E} \{ \boldsymbol{\eta}_j^* \otimes \boldsymbol{\eta}_j^{**} \}]^T) - [\text{Tr}(\mathbb{E} \{ \boldsymbol{\eta}_j^* \otimes \boldsymbol{\eta}_j^{**} \})]^2,
\end{aligned} \tag{49}$$

which is the same as

$$\begin{aligned}
& \text{Tr}(\tilde{\boldsymbol{\Sigma}} \tilde{\boldsymbol{\Sigma}}) + \text{Tr}(\mathbb{E} \{ \boldsymbol{\eta}_j^* \otimes \boldsymbol{\eta}_j^{**} \} [\mathbb{E} \{ \boldsymbol{\eta}_j^* \otimes \boldsymbol{\eta}_j^{**} \}]^T) \\
& = (n-1) + 2(n-2)\rho_j^2 + (n-1)\rho_j^2,
\end{aligned}$$

where, from (48),

$$\mathbb{E} \{ \boldsymbol{\eta}_j^* \otimes \boldsymbol{\eta}_j^{**} \} [\mathbb{E} \{ \boldsymbol{\eta}_j^* \otimes \boldsymbol{\eta}_j^{**} \}]^T = \begin{pmatrix} \rho_j^2 & 0 & \dots & \dots & 0 \\ 0 & \rho_j^2 & 0 & \dots & 0 \\ \vdots & \ddots & \ddots & \vdots & \vdots \\ 0 & \dots & \ddots & \ddots & \rho_j^2 \end{pmatrix} = \rho_j^2 \mathbf{I} \mathbf{d}_{n-1}.$$

From (49),

$$\text{Var} \left\{ \frac{\sum_{i=0}^{n-2} \eta_j(i) \eta_j(i+1)}{n-1} \right\} = \frac{(n-1) + 2(n-2)\rho_j^2 + (n-1)\rho_j^2}{(n-1)^2}. \tag{50}$$

Therefore, for each  $j \geq 1$ ,

$$\lim_{n \rightarrow \infty} n \text{Var} \left\{ \frac{\sum_{i=0}^{n-2} \eta_j(i) \eta_j(i+1)}{n-1} \right\} = 1 + 3\rho_j^2.$$

Thus, for each  $j \geq 1$ ,  $K_{j,2}$  in (19) is given by  $K_{j,2} = 1 + 3\rho_j^2$ . From equation (50),

$$\text{Var} \left\{ \frac{\sum_{i=0}^{n-2} \eta_j(i) \eta_j(i+1)}{n-1} \right\} \leq 1 + 3\rho_j^2 \leq 4.$$

Hence, for every  $j \geq 1$ ,  $\tilde{K}_{j,2}$  in equation (21) satisfies

$$\tilde{K}_{j,2} \leq 4.$$

Therefore, the constant  $S$  in [Assumption A4](#) is such that  $S \leq 6 + 4 = 10$ .

## 5 Simulation study

A simulation study is undertaken to illustrate the behaviour of the formulated componentwise estimator of the autocorrelation operator, and of its associated ARH(1) plug-in predictor for large sample sizes. The results are reported in [Appendix 5.1](#). In [Appendix 5.2](#), a comparative study is developed, from the implementation of the ARH(1) plug-in prediction techniques proposed in Antoniadis and Sapatinas [2003]; Besse et al. [2000]; Bosq [2000]; Guillas [2001]. In the subsequent sections, we restrict our attention to the Gaussian case

### 5.1 Behaviour of $\hat{\rho}$ and $\hat{X}_n$ for large sample sizes

Let  $(-\Delta)_{(a,b)}$  be the Dirichlet negative Laplacian operator on  $(a, b)$  given by

$$\begin{aligned} (-\Delta)_{(a,b)}(f)(x) &= -\frac{d^2}{dx^2}f(x), \quad x \in (a, b) \subset \mathbb{R}, \\ f(a) &= f(b) = 0. \end{aligned}$$

The eigenvectors  $\{\phi_j, j \geq 1\}$  and eigenvalues  $\{\lambda_j((-\Delta)_{(a,b)}), j \geq 1\}$  of  $(-\Delta)_{(a,b)}$  satisfy, for each  $j \geq 1$  and for each  $x \in (a, b)$ ,

$$(-\Delta)_{(a,b)}\phi_j(x) = \lambda_j((-\Delta)_{(a,b)})\phi_j(x), \quad \phi_j(a) = \phi_j(b) = 0. \quad (51)$$

For each  $j \geq 1$  and  $x \in [a, b]$ , the solution to equation (51) is given by (see [Grebekov and Nguyen,

2013, p. 6]):

$$\phi_j(x) = \sqrt{\frac{2}{b-a}} \sin\left(\frac{\pi j x}{b-a}\right), \quad \forall x \in [a, b], \quad \lambda_j((-\Delta)_{(a,b)}) = \frac{\pi^2 j^2}{(b-a)^2}. \quad (52)$$

We consider here the operator  $C$  defined as

$$C = ((-\Delta)_{(a,b)})^{-2(1-\gamma_1)}, \quad \gamma_1 \in (0, 1/2).$$

From [Dautray and Lions, 1990, pp. 119–140], the eigenvectors of  $C$  coincide with the eigenvectors of  $(-\Delta)_{(a,b)}$ , and its eigenvalues  $\{C_j, j \geq 1\}$  are given by:

$$C_j = [\lambda_j((-\Delta)_{(a,b)})]^{-2(1-\gamma_1)} = \left[ \frac{\pi^2 j^2}{(b-a)^2} \right]^{-2(1-\gamma_1)}. \quad (53)$$

Additionally, considering

$$\rho = \left[ \frac{(-\Delta)_{(a,b)}}{\lambda_1((-\Delta)_{(a,b)}) - \epsilon} \right]^{-(1-\gamma_2)}, \quad \gamma_2 \in (0, 1/2),$$

for certain positive constant  $\epsilon < \lambda_1((-\Delta)_{(a,b)})$  close to zero,  $\rho$  is a positive self-adjoint Hilbert-Schmidt operator, whose eigenvectors coincide with the eigenvectors of  $(-\Delta)_{(a,b)}$ , and whose eigenvalues  $\{\rho_j, j \geq 1\}$  are such that  $\rho_j < 1$ , for every  $j \geq 1$ , and

$$\rho_j^2 = \left[ \frac{\lambda_j((-\Delta)_{(a,b)})}{\lambda_1((-\Delta)_{(a,b)}) - \epsilon} \right]^{-2(1-\gamma_2)}, \quad \rho_j^2 \in (0, 1), \quad \gamma_2 \in (0, 1/2), \quad (54)$$

where, as before,  $\{\lambda_j((-\Delta)_{(a,b)}), j \geq 1\}$  are given in equation (52).

From (12), the eigenvalues  $\{\sigma_j^2, j \geq 1\}$  of  $C_\epsilon$  are then defined, for each  $j \geq 1$ , as

$$\sigma_j^2 = C_j (1 - \rho_j^2) = [\lambda_j((-\Delta)_{(a,b)})]^{-2(1-\gamma_1)} - \frac{[\lambda_j((-\Delta)_{(a,b)})]^{-2(2-\gamma_1-\gamma_2)}}{[\lambda_1((-\Delta)_{(a,b)}) - \epsilon]^{-2(1-\gamma_2)}}.$$

Note that  $C_\epsilon$  is in the trace class, since the trace property of  $C$ , and the fact that  $\rho_j^2 < 1$ , for every  $j \geq 1$ , implies

$$\sum_{j=1}^{\infty} \sigma_j^2 = \sum_{j=1}^{\infty} C_j (1 - \rho_j^2) < \sum_{j=1}^{\infty} C_j < \infty.$$

For this particular example of operator  $C$ , we have considered truncation parameter  $k_n$  of the form

$$k_n = n^{1/\alpha}, \quad (55)$$

for a suitable  $\alpha > 0$ , which, in particular, allows verification of (17). From equation (53), one has, for  $\gamma_1 \in (0, 1/2)$ ,

$$\sqrt{n}C_{k_n} = \sqrt{n} [\lambda_{k_n} (-\Delta_{(a,b)})]^{-2(1-\delta_1)} = \sqrt{n} \left( \frac{\pi k_n}{b-a} \right)^{-4(1-\delta_1)}, \quad \delta_1 > 1.$$

From equation (55), **Assumption A3** is then satisfied if

$$1/2 - \frac{4(1-\gamma_1)}{\alpha} > 0, \quad \text{i.e., if } \alpha > 8(1-\gamma_1) > 4. \quad (56)$$

since  $\gamma_1 \in (0, 1/2)$ . Fix  $\gamma_1 = 0.4$  and  $\gamma_2 = 9/20$ . Then, from equation (56),  $\alpha > 48/10$ . In particular, the values  $\alpha_1 = 5$  and  $\alpha_2 = 6$  have been tested, in Table 1 below, for  $H = L^2((a, b))$ , and  $(a, b) = (0, 4)$ , where  $L^2((a, b))$  denotes the space of square integrable functions on  $(a, b)$ .

The computed empirical truncated functional mean square error  $\text{EMSE}_{\widehat{\rho}_{k_n}}$  of the estimator  $\widehat{\rho}_{k_n}$  of  $\rho$ , for a sample size  $n$ , is given by:

$$\text{EMSE}_{\widehat{\rho}_{k_n}} = \frac{1}{N} \sum_{w=1}^N \sum_{j=1}^{k_n} (\rho_j - \widehat{\rho}_{n,j}^w)^2, \quad (57)$$

$$\widehat{\rho}_{n,j}^w = \frac{\widehat{D}_{n,j}^w}{\widehat{C}_{n,j}^w} = \frac{\frac{1}{n-1} \sum_{i=0}^{n-2} X_{i,j}^w X_{i+1,j}^w}{\frac{1}{n} \sum_{i=0}^{n-1} (X_{i,j}^w)^2}, \quad (58)$$

where  $N$  denotes the number of simulations, and for each  $j = 1, \dots, k_n$ ,  $\widehat{\rho}_{n,j}^w$  represents the estimator of  $\rho_j$ , based on the  $w$ -th generation of the values  $X_{0,j}^w, \dots, X_{n-1,j}^w$ , with  $X_{i,j}^w = \langle X_i^w, \phi_j \rangle_H$ , for  $w = 1, \dots, 700$ , and  $i = 0, \dots, n-1$ .

For the plug-in predictor  $\widehat{X}_n = \widehat{\rho}_{k_n}(X_{n-1})$ , we compute the empirical version  $\text{UB(EMAE)}_{\widehat{X}_n^{k_n}}$  of the derived upper bound (40), which, for each  $n \in \mathbb{Z}$ , is given by

$$\text{UB(EMAE)}_{\widehat{X}_n^{k_n}} = \sqrt{\frac{1}{N} \sum_{w=1}^N \sum_{j=1}^{k_n} (\rho_j - \widehat{\rho}_{n,j}^w)^2 \text{E} \left\{ \|\widehat{X}_{n-1}^w\|_H^2 \right\}}. \quad (59)$$

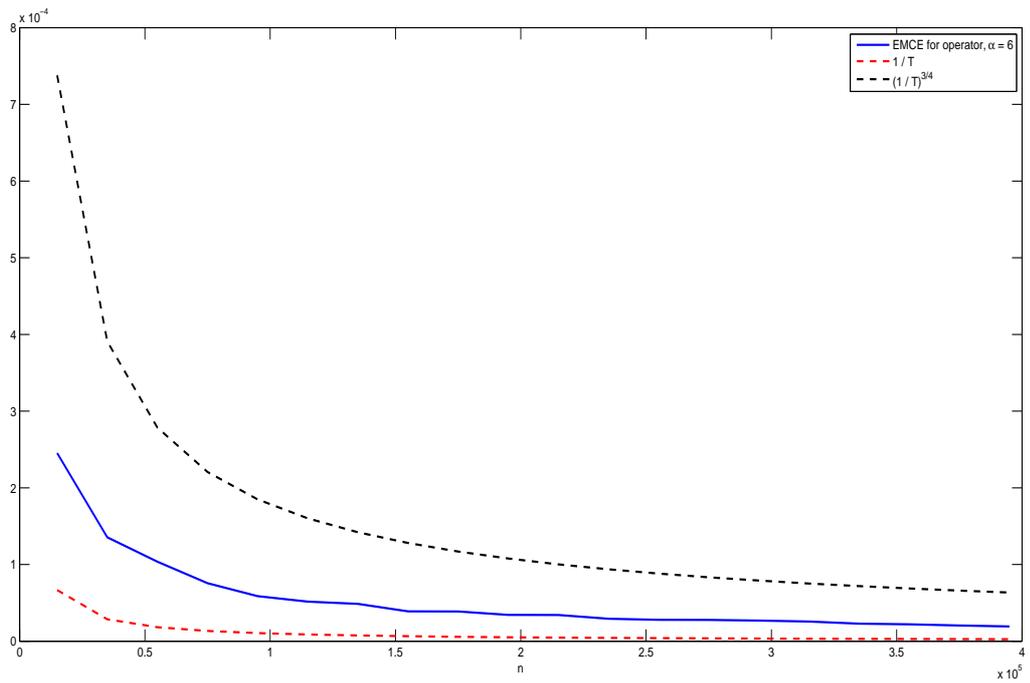
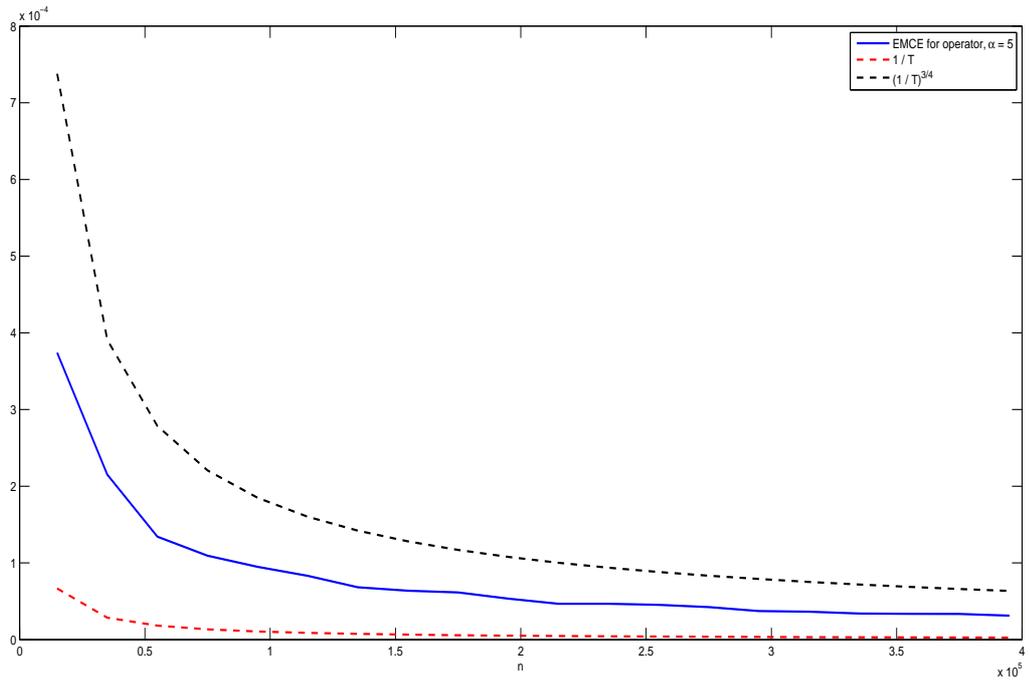
From  $N = 700$  realizations, for each one of the elements of the sequence of sample sizes

$$\{n_t, t = 1, \dots, 20\} = \{15000 + 20000(t-1), t = 1, \dots, 20\},$$

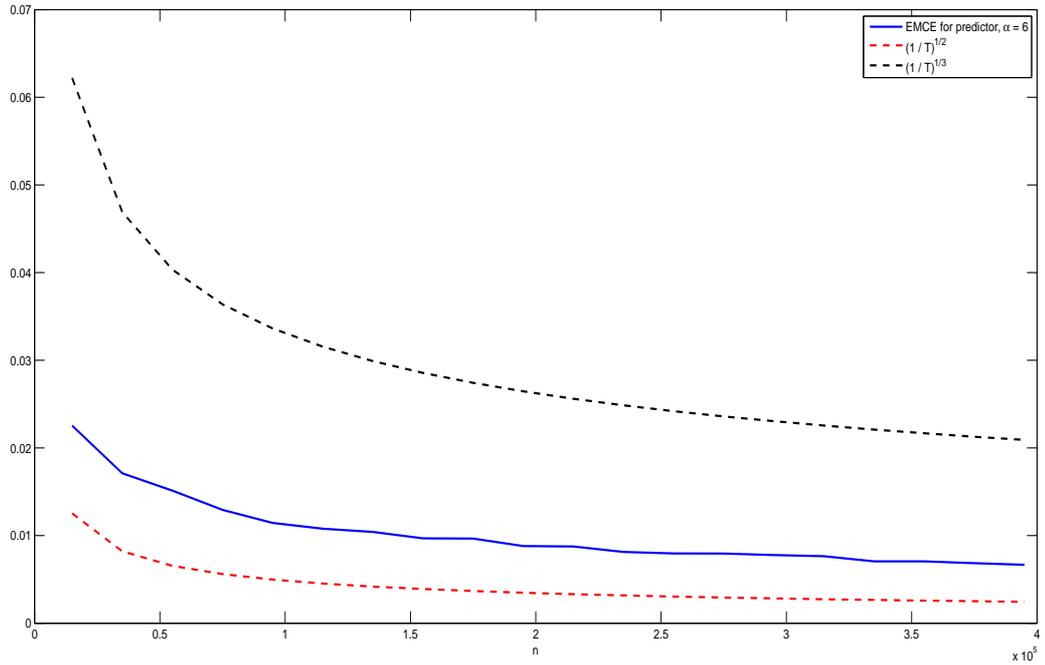
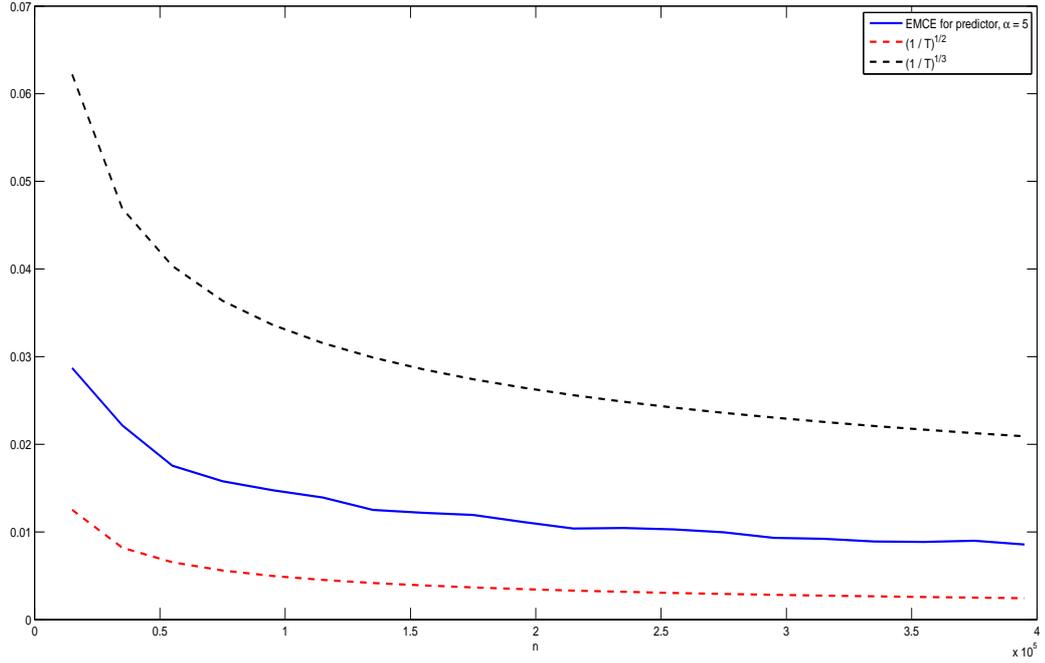
the  $\text{EMSE}_{\widehat{\rho}_{k_n}}$  and  $\text{UB(EMAE)}_{\widehat{X}_n^{k_n}}$  values, for  $\alpha = 5$  and  $\alpha = 6$ , are displayed in Table 1, where the abbreviated notations  $\text{MSE}_{\widehat{\rho}_{k_n,1}}$ , for  $\text{EMSE}_{\widehat{\rho}_{k_n}}$ , and  $\text{UB}_{\widehat{X}_n^{k_n,1}}$ , for  $\text{UB(EMAE)}_{\widehat{X}_n^{k_n}}$ , are used (see also Figures 1–2).

**Table 1:**  $\text{EMSE}_{\hat{\rho}_{k_n}}$  (here,  $\text{MSE}_{\hat{\rho}_{k_{n,i}}}$ ), and  $\text{UB}(\text{EMAE})_{\hat{X}_n^{k_n}}$  (here,  $\text{UB}_{\hat{X}_n^{k_{n,i}}}$ ) values, in (57)–(59), based on  $N = 700$  simulations, for  $\gamma_1 = 0.4$  and  $\gamma_2 = 9/20$ , considering the sample sizes  $\{n_t = 15000 + 20000(t - 1), t = 1, \dots, 20\}$  and the corresponding  $k_{n,1}$  and  $k_{n,2}$  values, for  $\alpha_1 = 5$  and  $\alpha_2 = 6$ .

$n$	$k_{n,1}$	$\text{MSE}_{\hat{\rho}_{k_{n,1}}}$	$\text{UB}_{\hat{X}_n^{k_{n,1}}}$	$k_{n,2}$	$\text{MSE}_{\hat{\rho}_{k_{n,2}}}$	$\text{UB}_{\hat{X}_n^{k_{n,2}}}$
$n_1 = 15000$	6	$3.74 (10)^{-4}$	$2.87 (10)^{-2}$	4	$2.45 (10)^{-4}$	$2.25 (10)^{-2}$
$n_2 = 35000$	8	$2.15 (10)^{-4}$	$2.21 (10)^{-2}$	5	$1.35 (10)^{-4}$	$1.71 (10)^{-2}$
$n_3 = 55000$	8	$1.34 (10)^{-4}$	$1.75 (10)^{-2}$	6	$1.03 (10)^{-4}$	$1.51 (10)^{-2}$
$n_4 = 75000$	9	$1.09 (10)^{-4}$	$1.57 (10)^{-2}$	6	$7.55 (10)^{-5}$	$1.29 (10)^{-2}$
$n_5 = 95000$	9	$9.48 (10)^{-5}$	$1.47 (10)^{-2}$	6	$5.86 (10)^{-5}$	$1.14 (10)^{-2}$
$n_6 = 115000$	10	$8.31 (10)^{-5}$	$1.39 (10)^{-2}$	6	$5.16 (10)^{-5}$	$1.07 (10)^{-2}$
$n_7 = 135000$	10	$6.81 (10)^{-5}$	$1.25 (10)^{-2}$	7	$4.86 (10)^{-5}$	$1.04 (10)^{-2}$
$n_8 = 155000$	10	$6.37 (10)^{-5}$	$1.21 (10)^{-2}$	7	$3.88 (10)^{-5}$	$9.66 (10)^{-3}$
$n_9 = 175000$	11	$6.14 (10)^{-5}$	$1.19 (10)^{-2}$	7	$3.87 (10)^{-5}$	$9.65 (10)^{-3}$
$n_{10} = 195000$	11	$5.34 (10)^{-5}$	$1.11 (10)^{-2}$	7	$3.42 (10)^{-5}$	$8.79 (10)^{-3}$
$n_{11} = 215000$	11	$4.67 (10)^{-5}$	$1.03 (10)^{-2}$	7	$3.40 (10)^{-5}$	$8.74 (10)^{-3}$
$n_{12} = 235000$	11	$4.66 (10)^{-5}$	$1.03 (10)^{-2}$	7	$2.92 (10)^{-5}$	$8.12 (10)^{-3}$
$n_{13} = 255000$	12	$4.53 (10)^{-5}$	$1.02 (10)^{-2}$	7	$2.77 (10)^{-5}$	$7.95 (10)^{-3}$
$n_{14} = 275000$	12	$4.24 (10)^{-5}$	$9.95 (10)^{-3}$	8	$2.77 (10)^{-5}$	$7.94 (10)^{-3}$
$n_{15} = 295000$	12	$3.72 (10)^{-5}$	$9.32 (10)^{-3}$	8	$2.67 (10)^{-5}$	$7.76 (10)^{-3}$
$n_{16} = 315000$	12	$3.62 (10)^{-5}$	$9.21 (10)^{-3}$	8	$2.55 (10)^{-5}$	$7.64 (10)^{-3}$
$n_{17} = 335000$	12	$3.39 (10)^{-5}$	$8.91 (10)^{-3}$	8	$2.28 (10)^{-5}$	$7.04 (10)^{-3}$
$n_{18} = 355000$	12	$3.34 (10)^{-5}$	$8.86 (10)^{-3}$	8	$2.20 (10)^{-5}$	$7.04 (10)^{-3}$
$n_{19} = 375000$	13	$3.34 (10)^{-5}$	$8.86 (10)^{-3}$	8	$2.04 (10)^{-5}$	$6.84 (10)^{-3}$
$n_{20} = 395000$	13	$3.12 (10)^{-5}$	$8.56 (10)^{-3}$	8	$1.92 (10)^{-5}$	$6.65 (10)^{-3}$



**Figure 1:**  $EMSE_{\hat{\rho}_{k_n}}$  values (blue line), in (57)–(58), based on  $N = 700$  simulations, for  $\gamma_1 = 0.4$  and  $\gamma_2 = 9/20$ , considering the sample sizes  $\{n_t = 15000 + 20000(t - 1), t = 1, \dots, 20\}$  and the corresponding  $k_{n,1}$  and  $k_{n,2}$  values, for  $\alpha_1 = 5$  (left-hand side) and  $\alpha_2 = 6$  (right-hand side), against curves  $(1/n_t)^{3/4}$  (black dot line) and  $1/n_t$  (red dot line).



**Figure 2:**  $UB(EMAE)_{\widehat{X}_n^{k_n}}$  values (blue line), in (59), based on  $N = 700$  simulations, for  $\gamma_1 = 0.4$  and  $\gamma_2 = 9/20$ , considering the sample sizes  $\{n_t = 15000 + 20000(t - 1), t = 1, \dots, 20\}$  and the corresponding  $k_{n,1}$  and  $k_{n,2}$  values, for  $\alpha_1 = 5$  (left-hand side) and  $\alpha_2 = 6$  (right-hand side), against curves  $(1/n_t)^{1/2}$  (red dot line) and  $(1/n_t)^{1/3}$  (black dot line).

In this paper, a one-parameter model of  $k_n$  is selected depending on parameter  $\alpha$ . In [Guillas, 2001, Example 2], in the same spirit, for an equivalent spectral class of operators  $C$ , a three-parameter model

is established for  $k_n$  to ensure convergence in quadratic mean in the space  $\mathcal{L}(H)$  of the componentwise estimator of  $\rho$  constructed from the known eigenvectors of  $C$ . The numerical results displayed in Table 1 and Figures 1–2 illustrate the fact that the proposed componentwise estimator  $\widehat{\rho}_{k_n}$  presents a speed of convergence to  $\rho$ , in quadratic mean in  $S(H)$ , faster than  $n^{-1/3}$ , which corresponds to the optimal case for the componentwise estimator of  $\rho$  proposed in Guillas [2001], in the case of known eigenvectors of  $C$ ; see, in particular, [Guillas, 2001, Theorem 1, Remark 2 and Example 2]. For larger values of the parameters  $\gamma_1$  than 2.4, and  $\alpha$  than 6, a faster velocity of convergence of  $\widehat{\rho}_{k_n}$  to  $\rho$ , in quadratic mean in the space  $S(H)$ , will be obtained. However, larger sample sizes are required for larger values of  $\alpha$ , in order to estimate a given number of coefficients of  $\rho$ . A more detailed discussion about comparison of the rates of convergence of the ARH(1) plug-in predictors proposed in Antoniadis and Sapatinas [2003]; Besse et al. [2000]; Bosq [2000]; Guillas [2001] can be found in the next section.

## 5.2 A comparative study

In this section, the performance of our approach is compared with those ones given in Antoniadis and Sapatinas [2003]; Besse et al. [2000]; Bosq [2000]; Guillas [2001], including the case of unknown eigenvectors of  $C$ . In the last case, our approach and the approaches presented in Bosq [2000]; Guillas [2001] are implemented in terms of the empirical eigenvectors.

### 5.2.1 Theoretical–eigenvector–based componentwise estimators

Let us first compare the performance of our ARH(1) plug-in predictor, defined in (38), and the ones formulated in Bosq [2000]; Guillas [2001], in terms of the theoretical eigenvectors  $\{\phi_j, j \geq 1\}$  of  $C$ . Note that, in this first part of our comparative study, we consider the previous generated Gaussian ARH(1) process, with autocovariance and autocorrelation operators defined from equations (53) and (54), for different rates of convergence to zero of parameters  $C_j$  and  $\rho_j^2$ ,  $j \geq 1$ , with both sequences being summable sequences. Since we restrict our attention to the Gaussian case, conditions  $A_1$ ,  $B_1$  and  $C_1$ , formulated in [Bosq, 2000, pp. 211–212] are satisfied by the generated ARH(1) process. Similarly, Conditions  $H_1$ – $H_3$  in [Guillas, 2001, p. 283] are satisfied as well.

In [Bosq, 2000, Section 8.2] the following estimator of  $\rho$  is proposed

$$\widehat{\rho}_n(x) = \left( \Pi^{k_n} D_n \widehat{C}_n^{-1} \Pi^{k_n} \right) (x) = \sum_{l=1}^{k_n} \widehat{\rho}_{n,l}(x) \phi_l, \quad x \in H, \quad (60)$$

$$\widehat{\rho}_{n,l}(x) = \frac{1}{n-1} \sum_{i=0}^{n-2} \sum_{j=1}^{k_n} \frac{1}{\widehat{C}_{n,j}} \langle \phi_j, x \rangle_H X_{i,j} X_{i+1,l}, \quad (61)$$

in the finite dimensional subspace

$$H_{k_n} = \text{span}(\phi_1, \dots, \phi_{k_n})$$

of  $H$ , where  $\Pi^{k_n}$  is the orthogonal projector over  $H_{k_n}$ , and, as before,  $X_{i,j} = \langle X_i, \phi_j \rangle_H$ , for  $j \geq 1$ .

A modified estimator of  $\rho$  is studied in [Guillas, 2001, Section 2], given by

$$\widehat{\rho}_{n,a}(x) = \left( \Pi^{k_n} D_n \widehat{C}_{n,a}^{-1} \Pi^{k_n} \right) (x) = \sum_{l=1}^{k_n} \widehat{\rho}_{n,a,l}(x) \phi_l, \quad x \in H, \quad (62)$$

$$\widehat{\rho}_{n,a,l}(x) = \frac{1}{n-1} \sum_{i=1}^{n-1} \sum_{j=1}^{k_n} \frac{1}{\max(\widehat{C}_{n,j}, a_n)} \langle \phi_j, x \rangle_H X_{i,j} X_{i+1,l}, \quad (63)$$

where

$$\widehat{C}_{n,a}^{-1}(x) = \sum_{j=1}^{k_n} \frac{1}{\max(\widehat{C}_{n,j}, a_n)} \langle \phi_j, x \rangle_H \phi_j \text{ a.s.}$$

Here,  $\{a_n, n \in \mathbb{N}\}$  is such that (see [Guillas, 2001, Theorem 1])

$$\alpha \frac{C_{k_n}^\gamma}{n^\varepsilon} \leq a_n \leq \beta \lambda_{k_n}, \quad \alpha > 0, \quad 0 < \beta < 1, \quad \varepsilon < 1/2, \quad \gamma \geq 1.$$

Tables 2–3 display the truncated, for two different  $k_n$  rules, empirical values of  $E\{\|\rho(X_{n-1}) - \widehat{\rho}_{k_n}(X_{n-1})\|_H\}$ , based on  $N = 700$  generations of each one of the functional samples considered with sizes  $n_t = 15000 + 20000(t-1)$ ,  $t = 1, \dots, 20$ , when

$$C_j = b_C j^{-\delta_1}, \quad b_C > 0, \quad \rho_j^2 = b_\rho j^{-\delta_2}, \quad b_\rho > 0.$$

Specifically,  $\widehat{\rho}_{k_n}$  is computed from equations (15)–(16) (see third column),  $\widehat{\rho}_{k_n} = \widehat{\rho}_n$ , with  $\widehat{\rho}_n$  being given in equations (60)–(61) (see fourth column), and  $\widehat{\rho}_{n,a} = \widehat{\rho}_{n,a}$ , with  $\widehat{\rho}_{n,a}$  being defined in (62)–(63) (see fifth column).

In Table 2,  $\delta_1 = 2.4$ ,  $\delta_2 = 1.1$ , and  $k_n = \lceil n^{1/\alpha} \rceil$ , for  $\alpha = 6$ , according to our [Assumption A3](#), which is also considered in [Bosq, 2000, p. 217] to ensure weak consistency of the proposed estimator of  $\rho$ . In Table 3, the same empirical values are displayed for  $\delta_1 = \frac{61}{60}$ ,  $\delta_2 = 1.1$ , and  $k_n$  is selected according to [Guillas, 2001, Example 2]. Thus, in Table 3,

$$k_n = \lceil n^{\frac{1-2\epsilon}{\delta_1(4+2\gamma)}} \rceil, \quad \gamma \geq 1, \quad \epsilon < 1/2. \quad (64)$$

In particular we have chosen  $\gamma = 2$ , and  $\epsilon = 0.04\delta_1$ . Note that, from [Guillas, 2001, Theorem 1 and Remark 1], for the choice made of  $k_n$  in Table 3, convergence to  $\rho$ , in quadratic mean in the space  $\mathcal{L}(H)$ , holds for  $\widehat{\rho}_{n,a}$  given in (62)–(63).

**Table 2:** Truncated empirical values of  $E\|\rho(X_{n-1}) - \hat{\rho}_{k_n}(X_{n-1})\|_H$ , for  $\hat{\rho}_{k_n}$  given in equations (15)-(16) (third column), in equations (60)-(61) (fourth column), and in equations (62)-(63) (fifth column), based on  $N = 700$  simulations, for  $\delta_1 = 2.4$  and  $\delta_2 = 1.1$ , considering the sample sizes  $\{n_t = 15000 + 20000(t - 1), t = 1, \dots, 20\}$  and the corresponding  $k_n = \lceil n^{1/\alpha} \rceil$  values, for  $\alpha = 6$ .

$n$	$k_n$	Our Approach	Bosq (2000)	Guillas (2001)
$n_1 = 15000$	4	$2.25 (10)^{-2}$	$2.57 (10)^{-2}$	$2.36 (10)^{-2}$
$n_2 = 35000$	5	$1.71 (10)^{-2}$	$1.72 (10)^{-2}$	$1.84 (10)^{-2}$
$n_3 = 55000$	6	$1.51 (10)^{-2}$	$1.65 (10)^{-2}$	$1.53 (10)^{-2}$
$n_4 = 75000$	6	$1.29 (10)^{-2}$	$1.46 (10)^{-2}$	$1.37 (10)^{-2}$
$n_5 = 95000$	6	$1.14 (10)^{-2}$	$1.20 (10)^{-2}$	$1.16 (10)^{-2}$
$n_6 = 115000$	6	$1.07 (10)^{-2}$	$1.10 (10)^{-2}$	$1.11 (10)^{-2}$
$n_7 = 135000$	7	$1.04 (10)^{-2}$	$1.06 (10)^{-2}$	$1.07 (10)^{-2}$
$n_8 = 155000$	7	$9.66 (10)^{-3}$	$9.91 (10)^{-3}$	$1.01 (10)^{-2}$
$n_9 = 175000$	7	$9.65 (10)^{-3}$	$9.79 (10)^{-3}$	$9.68 (10)^{-3}$
$n_{10} = 195000$	7	$8.79 (10)^{-3}$	$9.12 (10)^{-3}$	$8.93 (10)^{-3}$
$n_{11} = 215000$	7	$8.74 (10)^{-3}$	$8.79 (10)^{-3}$	$8.83 (10)^{-3}$
$n_{12} = 235000$	7	$8.12 (10)^{-3}$	$8.69 (10)^{-3}$	$8.75 (10)^{-3}$
$n_{13} = 255000$	7	$7.95 (10)^{-3}$	$8.53 (10)^{-3}$	$8.73 (10)^{-3}$
$n_{14} = 275000$	8	$7.94 (10)^{-3}$	$8.52 (10)^{-3}$	$8.58 (10)^{-3}$
$n_{15} = 295000$	8	$7.76 (10)^{-3}$	$8.49 (10)^{-3}$	$8.36 (10)^{-3}$
$n_{16} = 315000$	8	$7.64 (10)^{-3}$	$7.88 (10)^{-3}$	$8.13 (10)^{-3}$
$n_{17} = 335000$	8	$7.04 (10)^{-3}$	$7.24 (10)^{-3}$	$7.59 (10)^{-3}$
$n_{18} = 355000$	8	$7.04 (10)^{-3}$	$7.23 (10)^{-3}$	$6.92 (10)^{-3}$
$n_{19} = 375000$	8	$6.84 (10)^{-3}$	$6.89 (10)^{-3}$	$6.90 (10)^{-3}$
$n_{20} = 395000$	8	$6.65 (10)^{-3}$	$6.67 (10)^{-3}$	$6.85 (10)^{-3}$

**Table 3:** Truncated empirical values of  $E\|\rho(X_{n-1}) - \widehat{\rho}_{k_n}(X_{n-1})\|_H$ , for  $\widehat{\rho}_{k_n}$  given in equations (15)–(16) (third column), in equations (60)–(61) (fourth column), and in equations (62)–(63) (fifth column), based on  $N = 700$  simulations, for  $\delta_1 = \frac{61}{60}$  and  $\delta_2 = 1.1$ , considering the sample sizes  $\{n_t = 15000 + 20000(t - 1), t = 1, \dots, 20\}$  and the corresponding  $k_n$  given in (64).

$n$	$k_n$	Our Approach	Bosq (2000)	Guillas (2001)
$n_1 = 15000$	2	$9.91 (10)^{-3}$	$1.39 (10)^{-2}$	$1.26 (10)^{-2}$
$n_2 = 35000$	3	$8.78 (10)^{-3}$	$1.34 (10)^{-2}$	$1.24 (10)^{-2}$
$n_3 = 55000$	3	$7.89 (10)^{-3}$	$1.15 (10)^{-2}$	$1.14 (10)^{-2}$
$n_4 = 75000$	3	$6.49 (10)^{-3}$	$1.01 (10)^{-2}$	$8.58 (10)^{-3}$
$n_5 = 95000$	3	$6.36 (10)^{-3}$	$9.09 (10)^{-3}$	$8.29 (10)^{-3}$
$n_6 = 115000$	3	$6.14 (10)^{-3}$	$7.65 (10)^{-3}$	$7.26 (10)^{-3}$
$n_7 = 135000$	3	$5.91 (10)^{-3}$	$7.03 (10)^{-3}$	$6.69 (10)^{-3}$
$n_8 = 155000$	3	$5.73 (10)^{-3}$	$6.77 (10)^{-3}$	$6.54 (10)^{-3}$
$n_9 = 175000$	3	$5.44 (10)^{-3}$	$6.74 (10)^{-3}$	$6.16 (10)^{-3}$
$n_{10} = 195000$	3	$5.10 (10)^{-3}$	$6.69 (10)^{-3}$	$5.97 (10)^{-3}$
$n_{11} = 215000$	4	$5.01 (10)^{-3}$	$6.48 (10)^{-3}$	$5.94 (10)^{-3}$
$n_{12} = 235000$	4	$4.85 (10)^{-3}$	$6.45 (10)^{-3}$	$5.83 (10)^{-3}$
$n_{13} = 255000$	4	$4.17 (10)^{-3}$	$6.17 (10)^{-3}$	$5.68 (10)^{-3}$
$n_{14} = 275000$	4	$4.64 (10)^{-3}$	$5.99 (10)^{-3}$	$5.60 (10)^{-3}$
$n_{15} = 295000$	4	$4.55 (10)^{-3}$	$5.94 (10)^{-3}$	$5.58 (10)^{-3}$
$n_{16} = 315000$	4	$4.48 (10)^{-3}$	$5.69 (10)^{-3}$	$5.50 (10)^{-3}$
$n_{17} = 335000$	4	$4.38 (10)^{-3}$	$5.58 (10)^{-3}$	$5.44 (10)^{-3}$
$n_{18} = 355000$	4	$4.16 (10)^{-3}$	$5.45 (10)^{-3}$	$5.42 (10)^{-3}$
$n_{19} = 375000$	4	$3.91 (10)^{-3}$	$5.34 (10)^{-3}$	$5.32 (10)^{-3}$
$n_{20} = 395000$	4	$3.86 (10)^{-3}$	$5.29 (10)^{-3}$	$5.26 (10)^{-3}$

One can observe in Table 2 a similar performance of the three methods compared with the truncation order  $k_n$  satisfying [Assumption A3](#), with slightly worse results being obtained from the estimator defined in (62)–(63), specially, for the sample size  $n_8 = 155000$ . Furthermore, in Table 3, a better performance of our approach is observed for the smallest sample sizes (from  $n_1 = 15000$  until  $n_4 = 75000$ ). For the remaining largest sample sizes, only slight differences are observed, with, again, a better performance of our approach, very close to the other two approaches presented in Bosq [2000]; Guillas [2001].

### 5.2.2 Empirical-eigenvector-based componentwise estimators

In this section, we address the case where  $\{\phi_j, j \geq 1\}$  are unknown, as is often the case in practice. Specifically, for a given sample size  $n$ , let  $\{\phi_{n,j}, j \geq 1\}$  be the empirical counterpart of the theoretical

eigenvectors  $\{\phi_j, j \geq 1\}$ , satisfying, for every  $j \geq 1$ ,

$$C_n(\phi_{n,j}) = \frac{1}{n} \sum_{i=0}^{n-1} \langle X_i, \phi_{n,j} \rangle_H X_i = C_{n,j} \phi_{n,j},$$

where  $\{C_{n,j}, j \geq 1\}$  denotes the system of eigenvalues associated with the system of empirical eigenvectors

$\{\phi_{n,j}, j \geq 1\}$ . We then consider the following estimators for comparison purposes

$$\tilde{\rho}_{n,j} = \frac{\frac{1}{n-1} \sum_{i=0}^{n-2} \tilde{X}_{i,j} \tilde{X}_{i+1,j}}{\frac{1}{n} \sum_{i=0}^{n-1} (\tilde{X}_{i,j})^2}, \quad \tilde{\rho}_{k_n} = \sum_{j=1}^{k_n} \tilde{\rho}_{n,j} \phi_{n,j} \otimes \phi_{n,j}, \quad (65)$$

$$\tilde{\rho}_n(x) = \left( \tilde{\Pi}^{k_n} D_n C_n^{-1} \tilde{\Pi}^{k_n} \right) (x) = \sum_{l=1}^{k_n} \tilde{\rho}_{n,l}(x) \phi_{n,l}, \quad x \in H,$$

$$\tilde{\rho}_{n,l}(x) = \frac{1}{n-1} \sum_{i=0}^{n-2} \sum_{j=1}^{k_n} \frac{1}{C_{n,j}} \langle \phi_{n,j}, x \rangle_H \tilde{X}_{i,j} \tilde{X}_{i+1,l}, \quad (66)$$

$$\tilde{\rho}_{n,a}(x) = \left( \tilde{\Pi}^{k_n} D_n C_{n,a}^{-1} \tilde{\Pi}^{k_n} \right) (x) = \sum_{l=1}^{k_n} \tilde{\rho}_{n,a,l}(x) \phi_{n,l}, \quad x \in H,$$

$$\tilde{\rho}_{n,a,l}(x) = \frac{1}{n-1} \sum_{i=0}^{n-2} \sum_{j=1}^{k_n} \frac{1}{\max(C_{n,j}, a_n)} \langle \phi_{n,j}, x \rangle_H \tilde{X}_{i,j} \tilde{X}_{i+1,l}, \quad (67)$$

where, for  $i \in \mathbb{Z}$ , and  $j \geq 1$ ,  $\tilde{X}_{i,j} = \langle X_i, \phi_{n,j} \rangle_H$ ,  $\tilde{\Pi}^{k_n}$  denotes the orthogonal projector into the space

$$\tilde{H}_{k_n} = \text{span}(\phi_{n,1}, \dots, \phi_{n,k_n}).$$

The Gaussian ARH(1) process is generated under **Assumptions A1–A2**, as well as  $C'_1$  in [Bosq, 2000, p. 218]. Note that conditions  $A_1$  and  $B'_1$  in Bosq [2000] already hold. Moreover, as given in [Bosq, 2000, Theorem 8.8 and Example 8.6], for

$$C_j = b_C j^{-\delta_1}, \quad b_C > 0, \quad \delta_1 > 0,$$

with, in particular,  $\delta_1 = 2.4$ , and for

$$\rho_j = b_\rho j^{-\delta_2}, \quad b_\rho > 0,$$

with  $\delta_2 = 1.1$ , the estimator  $\tilde{\rho}_n$  converges almost surely to  $\rho$  under the condition

$$\frac{nC_{kn}^2}{\ln(n) \left( \sum_{j=1}^{k_n} b_j \right)^2} \rightarrow \infty,$$

where

$$b_1 = 2\sqrt{2}(C_1 - C_2)^{-1}, \quad b_j = 2\sqrt{2} \max \{ (C_{j-1} - C_j)^{-1}, (C_j - C_{j+1})^{-1} \}, \quad j \geq 2.$$

In Table 4,  $k_n = \lceil \ln(n) \rceil$  has been tested; see [Bosq, 2000, Example 8.6].

**Table 4:** Truncated empirical values of  $E \{ \|\rho(X_{n-1}) - \tilde{\rho}_{k_n}(X_{n-1})\|_H \}$ , for  $\tilde{\rho}_{k_n} = \tilde{\rho}_{k_n}$  given in equation (65) (third column),  $\tilde{\rho}_{k_n} = \tilde{\rho}_n$  defined in equation (66) (fourth column) and  $\tilde{\rho}_{k_n} = \tilde{\rho}_{n,a}$  defined in equation (67) (fifth column), based on  $N = 700$  simulations, for  $\delta_1 = 2.4$  and  $\delta_2 = 1.1$ , considering the sample sizes  $\{n_t = 15000 + 20000(t - 1), t = 1, \dots, 20\}$  and  $k_n = \lceil \ln(n) \rceil$ .

$n$	$k_n$	Our approach	Bosq (2000)	Guillas (2001)
$n_1 = 15000$	9	$8.42 (10)^{-2}$	1.061	1.035
$n_2 = 35000$	10	$5.51 (10)^{-2}$	1.019	1.005
$n_3 = 55000$	10	$4.75 (10)^{-2}$	1.017	0.999
$n_4 = 75000$	11	$4.43 (10)^{-2}$	1.015	0.995
$n_5 = 95000$	11	$3.68 (10)^{-2}$	1.013	0.988
$n_6 = 115000$	11	$3.51 (10)^{-2}$	1.011	0.963
$n_7 = 135000$	11	$3.23 (10)^{-2}$	1.008	0.925
$n_8 = 155000$	11	$2.95 (10)^{-2}$	1.007	0.912
$n_9 = 175000$	12	$2.94 (10)^{-2}$	1.006	0.911
$n_{10} = 195000$	12	$2.80 (10)^{-2}$	0.995	0.891
$n_{11} = 215000$	12	$2.71 (10)^{-2}$	0.902	0.862
$n_{12} = 235000$	12	$2.59 (10)^{-2}$	0.890	0.820
$n_{13} = 255000$	12	$2.58 (10)^{-2}$	0.878	0.800
$n_{14} = 275000$	12	$2.35 (10)^{-2}$	0.872	0.783
$n_{15} = 295000$	12	$2.28 (10)^{-2}$	0.860	0.778
$n_{16} = 315000$	12	$2.27 (10)^{-2}$	0.842	0.747
$n_{17} = 335000$	12	$2.16 (10)^{-2}$	0.822	0.714
$n_{18} = 355000$	12	$2.14 (10)^{-2}$	0.800	0.707
$n_{19} = 375000$	12	$2.09 (10)^{-2}$	0.778	0.687
$n_{20} = 395000$	12	$2.06 (10)^{-2}$	0.769	0.662

A better performance of our estimator (65) in comparison with estimator (66), formulated in Bosq

[2000], and estimator (67), formulated in [Guillas, 2001, Example 4 and Remark 4], is observed in Table 4. Note that, in particular, in [Guillas, 2001, Example 4 and Remark 4], smaller values of  $k_n$  than  $\ln(n)$  are required for a given sample size  $n$ , to ensure convergence in quadratic mean, and, in particular, weak-consistency. However, considering a smaller discretization step size  $\Delta t = 0.015$  than in Table 4, where  $\Delta t = 0.08$ , and for  $k_n = \lceil n^{1/6} \rceil$ , (i.e.,  $\alpha = 6$ ), we obtain in Table 5, for the same parameter values  $\delta_1 = 2.4$  and  $\delta_2 = 1.1$ , better results than in Table 4, since a smaller number of coefficients of  $\rho$  (parameters) to be estimated is considered in Table 5, from a richer sample information (coming from the smaller discretization step size considered). One can also observe in Table 5 a similar performance of the three approaches studied. In Table 6, the value  $k_n = \lceil e'n^{1/(8\delta_1+2)} \rceil$ , with  $e' = \frac{17}{10}$  proposed in [Guillas, 2001, Example 4 and Remark 4] is considered to compute the truncated empirical values of  $E\{\|\rho(X_{n-1}) - \tilde{\rho}_{k_n}(X_{n-1})\|_H\}$ , for  $\tilde{\rho}_{k_n}$  defined in equation (65) (third column), for  $\tilde{\rho}_{k_n} = \tilde{\rho}_n$  given in equation (66) (fourth column), and for  $\tilde{\rho}_{k_n} = \tilde{\rho}_{n,a}$  in equation (67) (fifth column). A similar performance of the three approaches is observed, with the exception of  $n_{20} = 395000$ , where the approach presented in Guillas [2001] displays a slightly better performance

**Table 5:** Truncated empirical values of  $E \{ \|\rho(X_{n-1}) - \tilde{\rho}_{k_n}(X_{n-1})\|_H \}$ , for  $\tilde{\rho}_{k_n}$  defined in equation (65) (third column), for  $\tilde{\rho}_{k_n} = \tilde{\rho}_n$  given in equation (66) (fourth column), and for  $\tilde{\rho}_{k_n} = \tilde{\rho}_{n,a}$  in equation (67) (fifth column), based on  $N = 200$  (due to high-dimensionality) simulations, for  $\delta_1 = 2.4$  and  $\delta_2 = 1.1$ , considering the sample sizes  $\{n_t = 15000 + 20000(t - 1), t = 1, \dots, 20\}$  and  $k_n = \lceil n^{1/6} \rceil$ .

$n$	$k_n$	Our approach	Bosq (2000)	Guillas (2001)
$n_1 = 15000$	4	$9.88 (10)^{-2}$	$9.25 (10)^{-2}$	0.106
$n_2 = 35000$	5	$9.52 (10)^{-2}$	$9.07 (10)^{-2}$	$9.86 (10)^{-2}$
$n_3 = 55000$	6	$9.12 (10)^{-2}$	$8.92 (10)^{-2}$	$9.39 (10)^{-2}$
$n_4 = 75000$	6	$8.48 (10)^{-2}$	$8.64 (10)^{-2}$	$8.98 (10)^{-2}$
$n_5 = 95000$	6	$7.61 (10)^{-2}$	$8.30 (10)^{-2}$	$8.46 (10)^{-2}$
$n_6 = 115000$	6	$7.05 (10)^{-2}$	$7.96 (10)^{-2}$	$8.04 (10)^{-2}$
$n_7 = 135000$	7	$6.99 (10)^{-2}$	$7.84 (10)^{-2}$	$7.82 (10)^{-2}$
$n_8 = 155000$	7	$6.70 (10)^{-2}$	$7.45 (10)^{-2}$	$7.40 (10)^{-2}$
$n_9 = 175000$	7	$6.49 (10)^{-2}$	$7.03 (10)^{-2}$	$7.07 (10)^{-2}$
$n_{10} = 195000$	7	$5.88 (10)^{-2}$	$6.74 (10)^{-2}$	$6.80 (10)^{-2}$
$n_{11} = 215000$	7	$5.63 (10)^{-2}$	$6.46 (10)^{-2}$	$6.57 (10)^{-2}$
$n_{12} = 235000$	7	$5.30 (10)^{-2}$	$6.28 (10)^{-2}$	$6.37 (10)^{-2}$
$n_{13} = 255000$	7	$5.05 (10)^{-2}$	$6.19 (10)^{-2}$	$6.24 (10)^{-2}$
$n_{14} = 275000$	8	$4.88 (10)^{-2}$	$5.99 (10)^{-2}$	$6.15 (10)^{-2}$
$n_{15} = 295000$	8	$4.58 (10)^{-2}$	$5.74 (10)^{-2}$	$6.04 (10)^{-2}$
$n_{16} = 315000$	8	$4.24 (10)^{-2}$	$5.52 (10)^{-2}$	$5.93 (10)^{-2}$
$n_{17} = 335000$	8	$3.86 (10)^{-2}$	$5.24 (10)^{-2}$	$5.70 (10)^{-2}$
$n_{18} = 355000$	8	$3.70 (10)^{-2}$	$5.02 (10)^{-2}$	$5.53 (10)^{-2}$
$n_{19} = 375000$	8	$3.55 (10)^{-2}$	$4.88 (10)^{-2}$	$5.36 (10)^{-2}$
$n_{20} = 395000$	8	$3.46 (10)^{-2}$	$4.70 (10)^{-2}$	$5.23 (10)^{-2}$

**Table 6:** Truncated empirical values of  $E \{ \|\rho(X_{n-1}) - \tilde{\rho}_{k_n}(X_{n-1})\|_H \}$ , for  $\tilde{\rho}_{k_n}$  defined in equation (65) (third column), for  $\tilde{\rho}_{k_n} = \tilde{\rho}_n$  given in equation (66) (fourth column), and for  $\tilde{\rho}_{k_n} = \tilde{\rho}_{n,a}$  in equation (67) (fifth column), based on  $N = 200$  (due to high-dimensionality) simulations, for  $\delta_1 = 2.4$  and  $\delta_2 = 1.1$ , considering the sample sizes  $\{n_t = 15000 + 20000(t - 1), t = 1, \dots, 20\}$  and  $k_n = \lceil e' n^{1/(8\delta_1+2)} \rceil$ ,  $e' = \frac{17}{10}$ .

$n$	$k_n$	Our approach	Bosq (2000)	Guillas (2001)
$n_1 = 15000$	2	$6.78 (10)^{-2}$	$8.77 (10)^{-2}$	$6.64 (10)^{-2}$
$n_2 = 35000$	2	$6.72 (10)^{-2}$	$8.61 (10)^{-2}$	$6.30 (10)^{-2}$
$n_3 = 55000$	2	$6.46 (10)^{-2}$	$8.48 (10)^{-2}$	$6.17 (10)^{-2}$
$n_4 = 75000$	2	$6.24 (10)^{-2}$	$8.20 (10)^{-2}$	$5.76 (10)^{-2}$
$n_5 = 95000$	2	$5.42 (10)^{-2}$	$7.84 (10)^{-2}$	$5.03 (10)^{-2}$
$n_6 = 115000$	2	$4.84 (10)^{-2}$	$7.34 (10)^{-2}$	$4.56 (10)^{-2}$
$n_7 = 135000$	2	$4.27 (10)^{-2}$	$6.95 (10)^{-2}$	$3.94 (10)^{-2}$
$n_8 = 155000$	2	$3.64 (10)^{-2}$	$6.60 (10)^{-2}$	$3.65 (10)^{-2}$
$n_9 = 175000$	3	$3.51 (10)^{-2}$	$6.52 (10)^{-2}$	$3.42 (10)^{-2}$
$n_{10} = 195000$	3	$3.38 (10)^{-2}$	$6.16 (10)^{-2}$	$3.24 (10)^{-2}$
$n_{11} = 215000$	3	$3.16 (10)^{-2}$	$5.78 (10)^{-2}$	$2.85 (10)^{-2}$
$n_{12} = 235000$	3	$2.98 (10)^{-2}$	$5.53 (10)^{-2}$	$2.60 (10)^{-2}$
$n_{13} = 255000$	3	$2.83 (10)^{-2}$	$5.15 (10)^{-2}$	$2.34 (10)^{-2}$
$n_{14} = 275000$	3	$2.50 (10)^{-2}$	$4.85 (10)^{-2}$	$2.05 (10)^{-2}$
$n_{15} = 295000$	3	$2.23 (10)^{-2}$	$4.46 (10)^{-2}$	$1.83 (10)^{-2}$
$n_{16} = 315000$	3	$2.15 (10)^{-2}$	$4.30 (10)^{-2}$	$1.58 (10)^{-2}$
$n_{17} = 335000$	3	$2.06 (10)^{-2}$	$4.14 (10)^{-2}$	$1.40 (10)^{-2}$
$n_{18} = 355000$	3	$1.98 (10)^{-2}$	$3.95 (10)^{-2}$	$1.24 (10)^{-2}$
$n_{19} = 375000$	3	$1.89 (10)^{-2}$	$3.77 (10)^{-2}$	$1.05 (10)^{-2}$
$n_{20} = 395000$	3	$1.82 (10)^{-2}$	$3.70 (10)^{-2}$	$9.93 (10)^{-3}$

### 5.2.3 Kernel-based nonparametric and penalized estimation

In practice, curves are observed in discrete times, and should be approximated by smooth functions. In Besse et al. [2000], the following optimization problem is considered:

$$\hat{X}_i = \operatorname{argmin} \left\| L\hat{X}_i \right\|_{L^2}^2, \quad \hat{X}_i(t_j) = X_i(t_j), \quad j = 1, \dots, p, \quad i = 0, \dots, n-1, \quad (68)$$

where  $L$  is a linear differential operator of order  $d$ . Our interpolation is computed by Matlab *smoothingspline* method. Non-linear kernel regression is then considered, in terms of the smoothed functional

data, solution to (68), as follows:

$$\widehat{X}_n^{h_n} = \widehat{\rho}_{h_n}(X_{n-1}), \quad \widehat{\rho}_{h_n}(x) = \frac{\sum_{i=0}^{n-2} \widehat{X}_{i+1} K\left(\frac{\|\widehat{X}_i - x\|_{L^2}^2}{h_n}\right)}{\sum_{i=0}^{n-2} K\left(\frac{\|\widehat{X}_i - x\|_{L^2}^2}{h_n}\right)},$$

where  $K$  is the usual Gaussian kernel, and

$$\|\widehat{X}_i - x\|_{L^2}^2 = \int (\widehat{X}_i(t) - x(t))^2 dt, \quad i = 0, \dots, n-2.$$

Alternatively, in Besse et al. [2000], prediction, in the context of functional autoregressive processes (FAR(1) processes), under the linear assumption on  $\rho$ , which is considered to be a compact operator, with  $\|\rho\| < 1$ , is also studied, from smooth data  $\widehat{X}_1, \dots, \widehat{X}_n$ , solving the optimization problem

$$\min_{\widehat{X}_i \in H_q} \frac{1}{n} \sum_{i=0}^{n-1} \left( \frac{1}{p} \sum_{j=1}^p \left( X_i(t_j) - \widehat{X}_i^{q,l}(t_j) \right)^2 + l \left\| D^2 \widehat{X}_i^{q,l} \right\|_{L^2}^2 \right), \quad (69)$$

where  $l$  is the smoothing parameter,  $H_q$  is the  $q$ -dimensional functional subspace spanned by the leading eigenvectors of the autocovariance operator  $C$  associated with its largest eigenvalues. Thus, smoothness and rank constraint are considered in the computation of the solution to the optimization problem (69).

Such a solution is obtained by means of functional PCA.

The following regularized empirical estimators of  $C$  and  $D$  are then considered, with inversion of  $C$  in the subspace  $H_q$ :

$$\widehat{C}_{q,l} = \frac{1}{n} \sum_{i=0}^{n-1} \widehat{X}_i \otimes \widehat{X}_i, \quad \widehat{D}_{q,l} = \frac{1}{n-1} \sum_{i=0}^{n-2} \widehat{X}_i \otimes \widehat{X}_{i+1}.$$

Thus, the regularized estimator of  $\rho$  is given by

$$\widehat{\rho}_{q,l} = \widehat{D}_{q,l} \widehat{C}_{q,l}^{-1},$$

and the predictor

$$\widehat{X}_n^{q,l} = \widehat{\rho}_{q,l} X_{n-1}.$$

Due to computational cost limitations, in Table 7, the following statistics are evaluated to compare the performance of the two above-referred prediction methodologies:

$$EMAE_{\widehat{X}_n}^{h_n} = \frac{1}{p} \sum_{j=1}^p \left( X_n(t_j) - \widehat{X}_n^{h_n}(t_j) \right)^2, \quad (70)$$

$$EMAE_{\hat{X}_n}^{q,l} = \frac{1}{p} \sum_{j=1}^p \left( X_n(t_j) - \hat{X}_n^{q,l}(t_j) \right)^2. \quad (71)$$

**Table 7:**  $EMAE_{\hat{X}_n}^{h_{n,i}}$ ,  $i = 1, 2$ , and  $EMAE_{\hat{X}_n}^{q,l}$  values (see (70) and (71), respectively), with  $q = 7$ , based on  $N = 200$  simulations, for  $\delta_1 = 2.4$  and  $\delta_2 = 1.1$ , considering now the sample sizes  $\{n_t = 750 + 500(t - 1), t = 1, \dots, 13\}$   $h_{n,1} = 0.1$  and  $h_{n,2} = 0.3$ .

$n$	$EMAE_{\hat{X}_n}^{h_{n,1}}$	$EMAE_{\hat{X}_n}^{h_{n,2}}$	$EMAE_{\hat{X}_n}^{q,l}$
$n_1 = 750$	$8.57 (10)^{-2}$	$8.85 (10)^{-2}$	$8.99 (10)^{-2}$
$n_2 = 1250$	$7.67 (10)^{-2}$	$8.43 (10)^{-2}$	$8.69 (10)^{-2}$
$n_3 = 1750$	$7.15 (10)^{-2}$	$7.12 (10)^{-2}$	$8.05 (10)^{-2}$
$n_4 = 2250$	$7.09 (10)^{-2}$	$6.87 (10)^{-2}$	$7.59 (10)^{-2}$
$n_5 = 2750$	$6.87 (10)^{-2}$	$6.67 (10)^{-2}$	$7.31 (10)^{-2}$
$n_6 = 3250$	$6.52 (10)^{-2}$	$5.92 (10)^{-2}$	$7.28 (10)^{-2}$
$n_7 = 3750$	$6.20 (10)^{-2}$	$5.56 (10)^{-2}$	$7.13 (10)^{-2}$
$n_8 = 4250$	$6.06 (10)^{-2}$	$5.32 (10)^{-2}$	$7.06 (10)^{-2}$
$n_9 = 4750$	$5.67 (10)^{-2}$	$5.25 (10)^{-2}$	$6.47 (10)^{-2}$
$n_{10} = 5250$	$5.24 (10)^{-2}$	$5.12 (10)^{-2}$	$6.08 (10)^{-2}$
$n_{11} = 5750$	$5.01 (10)^{-2}$	$4.82 (10)^{-2}$	$5.75 (10)^{-2}$
$n_{12} = 6250$	$4.90 (10)^{-2}$	$4.49 (10)^{-2}$	$5.33 (10)^{-2}$
$n_{13} = 6750$	$4.87 (10)^{-2}$	$3.87 (10)^{-2}$	$4.97 (10)^{-2}$

It can be observed a similar performance of the kernel-based and penalized FAR(1) predictors, from smooth functional data, which is also comparable, considering one realization, to the performance obtained in Table 6, from the empirical eigenvectors.

#### 5.2.4 Wavelet-based prediction for ARH(1) processes

The approach presented in Antoniadis and Sapatinas [2003] is now studied. Specifically, wavelet-based regularization is applied to obtain smooth estimates of the sample paths. The projection onto the space  $V_J$ , generated by translations of the scaling function  $\phi_{Jk}$ ,  $k = 0, \dots, 2^J - 1$ , at level  $J$ , associated with a multiresolution analysis of  $H$ , is first considered. For a given primary resolution level  $j_0$ , with  $j_0 < J$ , the following wavelet decomposition at  $J - j_0$  resolution levels can be computed for any projected curve  $\Phi_{V_J} X_i$ , in the space  $V_J$ , for  $i = 0, \dots, n - 1$ :

$$\Phi_{V_J} X_i = \sum_{k=0}^{2^{j_0}-1} c_{j_0 k}^i \phi_{j_0 k} + \sum_{j=j_0}^{J-1} \sum_{k=0}^{2^j-1} d_{j k}^i \psi_{j k},$$

$$c_{j_0 k}^i = \langle \Phi_{V_J} X_i, \phi_{j_0 k} \rangle_H, \quad d_{j k}^i = \langle \Phi_{V_J} X_i, \psi_{j k} \rangle_H.$$

For  $i = 0, \dots, n-1$ , the following variational problem is solved to obtain the smooth estimate of the curve  $X_i$ :

$$\inf_{f^i \in H} \left\{ \|\Phi_{V_J} X_i - f^i\|_{L^2}^2 + \lambda \|\Phi_{V_{j_0}^\perp} f\|^2; f \in H \right\}, \quad (72)$$

where  $\Phi_{V_{j_0}^\perp}$  denotes the orthogonal projection operator of  $H$  onto the orthogonal complement of  $V_{j_0}$ , and for  $i = 0, 1 \dots n-1$ ,

$$f^i = \sum_{k=0}^{2^{j_0}-1} \alpha_{j_0 k}^i \phi_{j_0 k} + \sum_{j=j_0}^{\infty} \sum_{k=0}^{2^j-1} \beta_{j k}^i \psi_{j k}.$$

Using the equivalent sequence of norms of fractional Sobolev spaces of order  $s$  with  $s > 1/2$ , on a suitable interval (in our case,  $s = \delta_1$ ), the minimization of (72) is equivalent to the optimization problem, for  $i = 0, \dots, n-1$ ,

$$\sum_{k=0}^{2^{j_0}-1} (\alpha_{j_0 k}^i - c_{j_0 k}^i)^2 + \sum_{j=j_0}^{J-1} \sum_{k=0}^{2^j-1} (d_{j k}^i - \beta_{j k}^i)^2 + \sum_{j=j_0}^{\infty} \sum_{k=0}^{2^j-1} \lambda 2^{j s} [\beta_{j k}^i]^2. \quad (73)$$

The solution to (73) is given by, for  $i = 0, \dots, n-1$ ,

$$\begin{aligned} \widehat{\alpha}_{j_0 k}^i &= c_{j_0 k}^i, \quad k = 0, 1, \dots, 2^{j_0} - 1, \\ \widehat{\beta}_{j_0 k}^i &= \frac{d_{j k}^i}{(1 + \lambda 2^{2s j})}, \quad j = j_0, \dots, J-1, \quad k = 0, 1, \dots, 2^j - 1, \\ \widehat{\beta}_{j_0 k}^i &= 0, \quad j \geq J, \quad k = 0, 1, \dots, 2^j - 1. \end{aligned}$$

In particular, in the subsequent computations, we have considered the following value of the smoothing parameter  $\lambda$  (see Angelini et al. [2003]):

$$\widehat{\lambda}^M = \frac{\left( \sum_{j=1}^M \sigma_j^2 \right) \left( \sum_{j=1}^M C_j \right)}{n}.$$

The following smoothed data are then computed

$$\widetilde{X}_{i, \widehat{\lambda}^M} = \sum_{k=0}^{2^{j_0}-1} \widehat{\alpha}_{j_0 k}^i \phi_{j_0 k} + \sum_{j=j_0}^{J-1} \sum_{k=0}^{2^j-1} \widehat{\beta}_{j_0 k}^i \psi_{j k},$$

removing the trend

$$\widetilde{a}_{n, \widehat{\lambda}^M} = \frac{1}{n} \sum_{i=0}^{n-1} \widetilde{X}_{i, \widehat{\lambda}^M}$$

to obtain

$$\widetilde{Y}_{i, \widehat{\lambda}^M} = \widetilde{X}_{i, \widehat{\lambda}^M} - \widetilde{a}_{n, \widehat{\lambda}^M}, \quad i = 0, \dots, n-1,$$

for the computation of

$$\begin{aligned}\tilde{\rho}_{n,\hat{\lambda}^M}(x) &= \left( \tilde{\Pi}_{\hat{\lambda}^M}^{k_n} \tilde{D}_{n,\hat{\lambda}^M} \tilde{C}_{n,\hat{\lambda}^M}^{-1} \tilde{\Pi}_{\hat{\lambda}^M}^{k_n} \right) (x) = \sum_{l=1}^{k_n} \tilde{\rho}_{n,\hat{\lambda}^M,l}(x) \tilde{\phi}_l^M, \quad x \in H, \\ \tilde{\rho}_{n,\hat{\lambda}^M,l}(x) &= \sum_{j=1}^{k_n} \frac{1}{n-1} \sum_{i=0}^{n-2} \frac{1}{\tilde{C}_{n,\hat{\lambda}^M,j}} \langle \tilde{\phi}_j^M, x \rangle_H \tilde{Y}_{i,\hat{\lambda}^M,j} \tilde{Y}_{i+1,\hat{\lambda}^M,l},\end{aligned}$$

for  $x \in H$  and

$$\tilde{C}_{n,\hat{\lambda}^M} = \frac{1}{n} \sum_{i=0}^{n-1} \tilde{Y}_{i,\hat{\lambda}^M} \otimes \tilde{Y}_{i,\hat{\lambda}^M},$$

where

$$\tilde{Y}_{i,\hat{\lambda}^M,j} = \langle \tilde{Y}_{i,\hat{\lambda}^M}, \hat{\phi}_{j,\hat{\lambda}^M} \rangle,$$

and

$$\tilde{C}_{n,\hat{\lambda}^M,j} = \langle \tilde{C}_{n,\hat{\lambda}^M}, \hat{\phi}_{j,\hat{\lambda}^M} \rangle,$$

for every  $j \geq 1$ . Table 8 displays the empirical truncated approximation of the expectation  $\mathbb{E} \{ \|\tilde{\rho}_{k_n}(X_{n-1}) - \rho(X_{n-1})\|_H \}$  and  $\mathbb{E} \{ \|\tilde{\rho}_{n,\hat{\lambda}^M}(X_{n-1}) - \rho(X_{n-1})\|_H \}$ , respectively obtained applying our approach, and the approach in Antoniadis and Sapatinas [2003], in the estimation of the autocorrelation operator  $\rho$ . Here, we have tested  $k_{n_i} = \lceil n^{1/\alpha_i} \rceil$ ,  $i = 1, 2$ , with  $\alpha_1 = 6$ , according to **Assumption A3**, and  $\alpha_2 > 4\delta_1$ , according to

$$H_4 : nC_{k_n}^4 \rightarrow \infty$$

in [Antoniadis and Sapatinas, 2003, p. 149]. In particular, we have considered  $\delta_1 = 2.4$ , and  $\alpha_2 = 10$ . From the results displayed in Table 8, one can observe a similar performance for the two truncation rules implemented, and approaches compared, for the small sample sizes tested. A similar accuracy is also displayed by the approaches presented in Besse et al. [2000], for such small sample sizes (see Table 7).

**Table 8:** Truncated empirical values of  $E\{\|\rho(X_{n-1}) - \tilde{\rho}_{k_n}(X_{n-1})\|_H\}$ , with  $\tilde{\rho}_{k_n}$  defined in equation (65), and of  $E\{\|\tilde{\rho}_{n,\hat{\lambda}_M}(X_{n-1}) - \rho(X_{n-1})\|_H\}$ , based on  $N = 200$  simulations, for  $\delta_1 = 2.4$  and  $\delta_2 = 1.1$ , considering the sample sizes  $\{n_t = 750 + 500(t - 1), t = 1, \dots, 13\}$ , using  $\hat{\lambda}_M$ ,  $M = 50$ , and the corresponding  $k_{n,i} = \lceil n^{1/\alpha_i} \rceil$ , for  $\alpha_1 = 6$  and  $\alpha_2 = 10$ . Here, O.A. means *Our Approach*.

$n$	$k_{n,1}$	O.A.	Antoniadis and Sapatinas [2003]	$k_{n,2}$	O.A.	Antoniadis and Sapatinas [2003]
750	3	0.070	0.091	1	0.064	0.059
1250	3	0.055	0.087	2	0.051	0.043
1750	3	0.047	0.080	2	0.045	0.039
2250	3	0.041	0.079	2	0.041	0.038
2750	3	0.037	0.073	2	0.036	0.035
3250	3	0.034	0.072	2	0.033	0.031
3750	3	0.033	0.068	2	0.033	0.029
4250	4	0.033	0.067	2	0.031	0.029
4750	4	0.032	0.066	2	0.031	0.026
5250	4	0.031	0.064	2	0.028	0.023
5750	4	0.030	0.060	2	0.020	0.019
6250	4	0.028	0.058	2	0.017	0.015
6750	4	0.028	0.056	2	0.019	0.014

## 6 Final comments

As noted before, in this paper, the eigenvectors of  $C$  are considered to be known in the derivation of the results on consistency. This assumption is satisfied, e.g., when the random initial condition is given as the solution, in the mean-square sense, of a stochastic differential equation driven by white noise (e.g., the Wiener measure), since the eigenvectors of the differential operator involved in that equation coincide with the eigenvectors of the autocovariance operator of the ARH(1) process. In the case where the eigenvectors of the autocovariance operator are unknown, the numerical results displayed in Tables 4–6 illustrate the fact that our approach displays, in terms of the empirical eigenvectors, very similar prediction results to those obtained with the implementation of the componentwise estimators proposed in Bosq [2000]; Guillas [2001], with a better performance of our approach in the more unfavorable case, corresponding to a large discretization step size, and truncation order (see Table 4 computed for  $k_n = \lceil \ln(n) \rceil$ ).

Regarding [Assumption A2](#), [Remark 1](#) provides an example where this assumption is satisfied. However, our approach can still be applied in a wider range of situations. Wavelet bases are well suited for sparse representation of functions; recent work has considered combining them with sparsity-inducing

penalties, both for semiparametric regression (see, e.g., Wand and Ormerod [2011]), and for regression with functional or kernel predictors (see Wand and Ormerod [2011]; Zhao et al. [2015, 2012], among others). The latter papers focused on  $\ell_1$  penalization, also known as the lasso (see Tibshirani [1996]), in the wavelet domain. Alternatives to the lasso include the SCAD penalty by Fan and Li [2001], and the adaptive lasso by Zou [2006]. The  $\ell_1$  penalty in the elastic net criterion has the effect of shrinking small coefficients to zero. This can be interpreted as imposing a prior that favors a sparse estimate. The above mentioned smoothing techniques, based on wavelets, can be applied to obtain a smooth sparse approximation  $\widehat{X}_1, \dots, \widehat{X}_n$  of the functional data  $X_1, \dots, X_n$ , whose empirical auto-covariance operator

$$\widehat{C}_n = \frac{1}{n} \sum_{i=0}^{n-1} \widehat{X}_i \otimes \widehat{X}_i$$

and cross-covariance operator

$$\widehat{D}_n = \frac{1}{n-1} \sum_{i=0}^{n-2} \widehat{X}_i \otimes \widehat{X}_{i+1}$$

admits a diagonal representation in terms of wavelets.

In the literature, shrinkage approaches for estimating a high-dimensional covariance matrix are employed to circumvent the limitations of the sample covariance matrix. In particular, a new family of nonparametric Stein-type shrinkage covariance estimators is proposed in Touloumis [2015] (see also references therein), whose members are written as a convex linear combination of the sample covariance matrix and of a predefined invertible diagonal target matrix. These results can be applied to our framework, considering the shrinkage estimators of the autocovariance and cross-covariance operators, with respect to a common suitable wavelet basis, which can lead to an empirical diagonal representation of both operators.

In the Supplementary Material provided (see [Appendix 7](#)), a numerical example is provided to illustrate the performance of our approach, in the case of a pseudo-diagonal autocorrelation operator.

## 7 Supplementary Material: non-diagonal autocorrelation operator

This Section provides as a numerical example where the methodology proposed in such paper still works beyond the considered [Assumption A2](#). In particular, this section illustrates the performance of the proposed estimation methodology, when [Assumption A2](#) is not satisfied, but  $\rho$  is close to be diagonal in some sense. The numerical results obtained are compared to those ones derived from the computation of the ARH(1) predictors, based on the componentwise estimators proposed in Bosq [2000];

Guillas [2001] where this diagonal assumption is not required. The Gaussian ARH(1) process generated has autocorrelation operator  $\rho$  with coefficients  $\rho_{j,h}$  with respect to the basis  $\{\phi_j \otimes \phi_h, j, h \geq 1\}$ , given by

$$\rho_{j,j}^2 = \left( \frac{\lambda_j \left( (-\Delta)_{(a,b)} \right)}{\lambda_1 \left( (-\Delta)_{(a,b)} \right) - \epsilon} \right)^{-\delta_2}, \quad (74)$$

in the diagonal, and outside of the diagonal

$$\rho_{j,j+a}^2 = \frac{0.01}{5a^2}, \quad a = 1, \dots, 5, \quad \rho_{j+a,j}^2 = \frac{0.02}{5a^2}, \quad a = 1, \dots, 5, \quad (75)$$

where  $\rho_{j,j+a}^2 = \rho_{j+a,j}^2 = 0$  when  $a \geq 6$ . The coefficients of the autocovariance operator  $C_\epsilon$  of the innovation process  $\epsilon$ , with respect to the mentioned basis  $\{\phi_j \otimes \phi_h, j, h \geq 1\}$ , are given by

$$\sigma_{j,j}^2 = C_j (1 - \rho_{j,j}^2)$$

in the diagonal, and outside of the diagonal by

$$\sigma_{j,j+a}^2 = \frac{0.015}{5a^2}, \quad a = 1, 2, 3, 4, 5, \quad \sigma_{j+a,j}^2 = \frac{0.01}{5a^2}, \quad a = 1, 2, 3, 4, 5, \quad (76)$$

where  $\sigma_{j,j+a}^2 = \sigma_{j+a,j}^2 = 0$  when  $a \geq 6$ . Table 9 below displays the empirical truncated values of  $E \left\{ \left\| \rho(X_{n-1}) - \widehat{\rho}_{k_n}^{ND}(X_{n-1}) \right\|_H \right\}$  based on  $N = 200$  simulations of each one of the 20 functional samples considered, with sizes  $\{n_t = 15000 + 20000(t-1), t = 1, \dots, 20\}$ , for the corresponding  $k_n$  values obtained, in this case, by the rule  $k_n = \lceil n^{1/\alpha} \rceil$ , with  $\alpha = 6$ . We have considered parameter  $\delta_1 = 2.4$  in the definition of the eigenvalues of  $C$ ; but, in this case, as noted before, operators  $\rho$  and  $C_\epsilon$  are non-diagonal (see equations 75–76). The estimators of  $\rho$  and the associated plug-in predictors are computed, for the three approaches compared, under the assumption that the eigenvectors of  $C$  are known.

As expected, in Table 9, an outperformance of the approaches in Bosq [2000]; Guillas [2001] is observed in comparison with our methodology. However, for large sample sizes, the ARH(1) prediction methodology proposed here still can be applied with an order of magnitude of  $10^{-2}$  for the empirical errors associated with  $\widehat{\rho}_{k_n}$  given in equation 65. Thus, in the pseudodiagonal autocorrelation operator case, in some sense, our approach could still be considered. As referred in our paper, an example is given in the case where the autocovariance and autocorrelation operators admit a sparse representation in terms of a suitable orthonormal wavelet basis (see, for instance, Angelini et al. [2003]; Antoniadis and Sapatinas [2003]).

**Table 9:** Truncated empirical values of  $E \{ \|\rho(X_{n-1}) - \hat{\rho}_{k_n}^{ND}(X_{n-1})\|_H \}$ , for  $\hat{\rho}_{k_n}^{ND}$  given in equations (15)–(16) (third column), in equations (60)–(61) (fourth column), and in equations (62)–(63) (fifth column), from the non-diagonal data generated by equations (74)–(76), based on  $N = 200$  (due to high-dimensionality) simulations, for  $\delta_1 = 2.4$  and  $\delta_2 = 1.1$ , considering the sample sizes  $\{n_t = 15000 + 20000(t - 1), t = 1, \dots, 20\}$  and the corresponding  $k_n = \lceil n^{1/\alpha} \rceil$ ,  $\alpha = 6$  values. The eigenvectors  $\{\phi_j, j \geq 1\}$  are assumed to be known.

$n$	$k_n$	Our approach	Bosq (2000)	Guillas (2001)
$n_1 = 15000$	4	0.581	$8.94 (10)^{-2}$	0.1055
$n_2 = 35000$	5	0.560	$7.05 (10)^{-2}$	$9.49 (10)^{-2}$
$n_3 = 55000$	6	0.548	$6.67 (10)^{-2}$	$9.14 (10)^{-2}$
$n_4 = 75000$	6	0.532	$6.24 (10)^{-2}$	$8.85 (10)^{-2}$
$n_5 = 95000$	6	0.512	$5.89 (10)^{-2}$	$8.47 (10)^{-2}$
$n_6 = 115000$	6	0.498	$5.62 (10)^{-2}$	$8.04 (10)^{-2}$
$n_7 = 135000$	7	0.495	$5.57 (10)^{-2}$	$7.66 (10)^{-2}$
$n_8 = 155000$	7	0.481	$5.28 (10)^{-2}$	$7.24 (10)^{-2}$
$n_9 = 175000$	7	0.474	$5.01 (10)^{-2}$	$6.78 (10)^{-2}$
$n_{10} = 195000$	7	0.461	$4.90 (10)^{-2}$	$6.30 (10)^{-2}$
$n_{11} = 215000$	7	0.442	$4.69 (10)^{-2}$	$6.07 (10)^{-2}$
$n_{12} = 235000$	7	0.425	$4.45 (10)^{-2}$	$5.82 (10)^{-2}$
$n_{13} = 255000$	7	0.411	$4.25 (10)^{-2}$	$5.54 (10)^{-2}$
$n_{14} = 275000$	8	0.408	$4.14 (10)^{-2}$	$5.16 (10)^{-2}$
$n_{15} = 295000$	8	0.381	$4.09 (10)^{-2}$	$4.81 (10)^{-2}$
$n_{16} = 315000$	8	0.360	$3.85 (10)^{-2}$	$4.53 (10)^{-2}$
$n_{17} = 335000$	8	0.349	$3.56 (10)^{-2}$	$4.29 (10)^{-2}$
$n_{18} = 355000$	8	0.330	$3.29 (10)^{-2}$	$3.98 (10)^{-2}$
$n_{19} = 375000$	8	0.320	$2.90 (10)^{-2}$	$3.75 (10)^{-2}$
$n_{20} = 395000$	8	0.318	$2.62 (10)^{-2}$	$3.44 (10)^{-2}$

## Acknowledgments

This work has been supported in part by project MTM2015-71839-P (co-funded by Feder funds), of the DGI, MINECO, Spain.

## References

[Aneiros-Pérez and Vieu 2006] ANEIROS-PÉREZ, G. ; VIEU, P.: Semi-functional partial linear regression. *Statist. Probab. Lett.* 76 (2006), pp. 1102–1110. – DOI: [doi.org/10.1016/j.spl.2005.12.007](https://doi.org/10.1016/j.spl.2005.12.007)

- [Aneiros-Pérez and Vieu 2008] ANEIROS-PÉREZ, G. ; VIEU, P.: Nonparametric time series prediction: a semifunctional partial linear modeling. *J. Multivariate Anal.* 99 (2008), pp. 834–857. – DOI: [doi.org/10.1016/j.jmva.2007.04.010](https://doi.org/10.1016/j.jmva.2007.04.010)
- [Angelini et al. 2003] ANGELINI, C. ; CANDITIIS, D. D. ; LEBLANC, F.: Wavelet regression estimation in nonparametric mixed effect models. *J. Multivariate Anal.* 85 (2003), pp. 267–291. – DOI: [doi.org/10.1016/S0047-259X\(02\)00055-6](https://doi.org/10.1016/S0047-259X(02)00055-6)
- [Antoniadis and Sapatinas 2003] ANTONIADIS, A. ; SAPATINAS, T.: Wavelet methods for continuous-time prediction using Hilbert-valued autoregressive processes. *J. Multivariate Anal.* 87 (2003), pp. 133–158. – DOI: [doi.org/10.1016/S0047-259X\(03\)00028-9](https://doi.org/10.1016/S0047-259X(03)00028-9)
- [Bartlett 1946] BARTLETT, M. S.: On the theoretical specification and sampling properties of autocorrelated time series. *Supplement to J. Roy. Stat. Soc.* 8 (1946), pp. 27–41. – URL <http://www.jstor.org/stable/2983611>
- [Bensmain and Mourid 2001] BENSMAIN, N. ; MOURID, T.: Estimateur "sieve" de l'opérateur d'un processus ARH(1). *C. R. Acad. Sci. Paris Sér. I Math.* 332 (2001), pp. 1015–1018. – DOI: [doi.org/10.1016/S0764-4442\(01\)01954-1](https://doi.org/10.1016/S0764-4442(01)01954-1)
- [Besse et al. 2000] BESSE, P. C. ; CARDOT, H. ; STEPHENSON, D. B.: Autoregressive forecasting of some functional climatic variations. *Scand. J. Statist.* 27 (2000), pp. 673–687. – DOI: [doi.org/10.1111/1467-9469.00215](https://doi.org/10.1111/1467-9469.00215)
- [Bongiorno et al. 2014] BONGIORNO, G. ; GOIA, A. ; SALINELLI, E. ; VIEU, P.: *Contributions in infinite-dimensional statistics and related topics*. In: Soc. Editrice Esculapio, Bologna, 2014. – ISBN 9788874887637
- [Bosq 2000] BOSQ, D.: *Linear Processes in Function Spaces*. Springer, New York, 2000. – ISBN 9781461211549
- [Bosq 2007] BOSQ, D.: General linear processes in Hilbert spaces and prediction. *J. Statist. Plann. Inference* 137 (2007), pp. 879–894. – DOI: [doi.org/10.1016/j.jspi.2006.06.014](https://doi.org/10.1016/j.jspi.2006.06.014)
- [Bosq and Blanke 2007] BOSQ, D. ; BLANKE, D.: *Inference and predictions in large dimensions*. Wiley, 2007. – ISBN 9780470017616
- [Chen et al. 2012] CHEN, J. ; LIN, D. ; HOCHNER, H.: Semiparametric maximum likelihood methods for analyzing genetic and environmental effects with case-control motherchild pair data. *Biometrics* 68 (2012), pp. 869–877. – DOI: [doi.org/10.1111/j.1541-0420.2011.01728.x](https://doi.org/10.1111/j.1541-0420.2011.01728.x)

- [Cuevas 2014] CUEVAS, A.: A partial overview of the theory of statistics with functional data. *J. Statist. Plann. Inference* 147 (2014), pp. 1–23. – DOI: [doi.org/10.1016/j.jspi.2013.04.002](https://doi.org/10.1016/j.jspi.2013.04.002)
- [Cugliari 2011] CUGLIARI, J.: *Prévision non paramétrique de processus à valeurs fonctionnelles. Application à la consommation d'électricité*, University of Paris-Sud 11, PhD. Thesis, 2011. – URL <https://tel.archives-ouvertes.fr/tel-00647334>
- [Damon and Guillas 2002] DAMON, J. ; GUILLAS, S.: The inclusion of exogenous variables in functional autoregressive ozone forecasting. *Environmetrics* 13 (2002), pp. 759–774. – DOI: [doi.org/10.1002/env.527](https://doi.org/10.1002/env.527)
- [Damon and Guillas 2005] DAMON, J. ; GUILLAS, S.: Estimation and simulation of autoregressive Hilbertian processes with exogenous variables. *Stat. Inference Stoch. Process.* 8 (2005), pp. 185–204. – DOI: [doi.org/10.1007/s11203-004-1031-6](https://doi.org/10.1007/s11203-004-1031-6)
- [Dautray and Lions 1990] DAUTRAY, R. ; LIONS, J.-L.: *Mathematical Analysis and Numerical Methods for Science and Technology Volume 3: Spectral Theory and Applications*. Springer, New York, 1990. – ISBN 9783642615290
- [Dedecker and Merlevède 2003] DEDECKER, J. ; MERLEVÈDE, F.: The conditional central limit theorem in Hilbert spaces. *Stochastic Process. Appl.* 108 (2003), pp. 229–262. – DOI: [doi.org/10.1016/j.spa.2003.07.004](https://doi.org/10.1016/j.spa.2003.07.004)
- [Dehling and Sharipov 2005] DEHLING, H. ; SHARIPOV, O. S.: Estimation of mean and covariance operator for Banach space valued autoregressive processes with dependent innovations. *Stat. Inference Stoch. Process.* 8 (2005), pp. 137–149. – DOI: [doi.org/10.1007/s11203-003-0382-8](https://doi.org/10.1007/s11203-003-0382-8)
- [Fan and Li 2001] FAN, J. ; LI, R.: Variable selection via nonconcave penalized likelihood and its oracle properties. *J. Amer. Statist. Assoc.* 96 (2001), pp. 1348–1360. – URL <http://www.jstor.org/stable/3085904>
- [Febrero-Bande and González-Manteiga 2013] FEBRERO-BANDE, M. ; GONZÁLEZ-MANTEIGA, W.: Generalized additive models for functional data. *Test* 22 (2013), pp. 278–292. – DOI: [doi.org/10.1007/s11749-012-0308-0](https://doi.org/10.1007/s11749-012-0308-0)
- [Ferraty et al. 2012] FERRATY, F. ; KEILEGOM, I. V. ; VIEU, P.: Regression when both response and predictor are functions. *J. Multivariate Anal.* 109 (2012), pp. 10–28. – DOI: [doi.org/10.1016/j.jmva.2012.02.008](https://doi.org/10.1016/j.jmva.2012.02.008)
- [Ferraty and Vieu 2006] FERRATY, F. ; VIEU, P.: *Nonparametric functional data analysis: theory and practice*. Springer, 2006. – ISBN 9780387303697

- [Ferraty and Vieu 2009] FERRATY, F. ; VIEU, P.: Additive prediction and boosting for functional data. *Comput. Statist. Data Anal.* 53 (2009), pp. 1400–1413. – DOI: [doi.org/10.1016/j.csda.2008.11.023](https://doi.org/10.1016/j.csda.2008.11.023)
- [Goia and Vieu 2016] GOIA, A. ; VIEU, P.: An introduction to recent advances in high/infinite dimensional statistics. *J. Multivariate Anal.* 146 (2016), pp. 1–6. – DOI: [doi.org/10.1016/j.jmva.2015.12.001](https://doi.org/10.1016/j.jmva.2015.12.001)
- [Grebencov and Nguyen 2013] GREBENKOV, D. S. ; NGUYEN, B. T.: Geometrical structure of Laplacian eigenfunctions. *SIAM Rev.* 55 (2013), pp. 601–667. – DOI: [doi.org/10.1137/120880173](https://doi.org/10.1137/120880173)
- [Guillas 2000] GUILLAS, S.: Non-causalité et discrétisation fonctionnelle, théorèmes limites pour un processus ARHX(1). *C. R. Acad. Sci. Paris Sér. I* 331 (2000), pp. 91–94
- [Guillas 2001] GUILLAS, S.: Rates of convergence of autocorrelation estimates for autoregressive Hilbertian processes. *Statist. Probab. Lett.* 55 (2001), pp. 281–291. – DOI: [doi.org/10.1016/S0167-7152\(01\)00151-1](https://doi.org/10.1016/S0167-7152(01)00151-1)
- [Gurland 1956] GURLAND, J.: Quadratic forms in normally distributed random variables. *Sankhya A* 17 (1956), pp. 37–50. – URL <http://www.jstor.org/stable/25048285>
- [Hamilton 1994] HAMILTON, J. D.: *Time Series Analysis*. Princeton University Press, Princeton, 1994. – ISBN 9780691042893
- [Horváth and Kokoszka 2012] HORVÁTH, L. ; KOKOSZKA, P.: *Inference for functional data with applications*. Springer, New York, 2012. – ISBN 9781461436553
- [Hsing and Eubank 2015] HSING, T. ; EUBANK, R.: *Theoretical Foundations of Functional Data Analysis, with an Introduction to Linear Operators*. Wiley, 2015. – ISBN 9781118762547
- [Kargin and Onatski 2008] KARGIN, V. ; ONATSKI, A.: Curve forecasting by functional autoregression. *J. Multivariate Anal.* 99 (2008), pp. 2508–2526. – DOI: [doi.org/10.1016/j.jmva.2008.03.001](https://doi.org/10.1016/j.jmva.2008.03.001)
- [Mallat 2009] MALLAT, S. G.: *A Wavelet Tour of Signal Processing: The Sparse Way, 3rd ed.* Academic Press, Burlington, MA, 2009. – ISBN 9780123743701
- [Marion and Pumo 2004] MARION, J. M. ; PUMO, B.: Comparison of ARH(1) and ARHD(1) models on physiological data. *Ann. I.S.U.P.* 48 (2004), pp. 29–38. – URL [https://www.researchgate.net/publication/288849889\\_Comparaison\\_des\\_modeles\\_ARH1\\_et\\_ARHD1\\_sur\\_des\\_dc](https://www.researchgate.net/publication/288849889_Comparaison_des_modeles_ARH1_et_ARHD1_sur_des_dc)
- [Mas 1999] 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 (1999), pp. 899–902. – DOI: [doi.org/10.1016/S0764-4442\(00\)87496-0](https://doi.org/10.1016/S0764-4442(00)87496-0)

- [Mas 2004] MAS, A.: Consistance du prédicteur dans le modèle ARH(1): le cas compact. *Ann. I.S.U.P.* 48 (2004), pp. 39–48. – URL <http://www.math.univ-montp2.fr/~mas/JIIsup2.pdf>
- [Mas 2007] MAS, A.: Weak-convergence in the functional autoregressive model. *J. Multivariate Anal.* 98 (2007), pp. 1231–1261. – DOI: [doi.org/10.1016/j.jmva.2006.05.010](https://doi.org/10.1016/j.jmva.2006.05.010)
- [Mas and Menneteau 2003a] MAS, A. ; MENNETEAU, L.: Large and moderate deviations for infinite dimensional autoregressive processes. *J. Multivariate Anal.* 87 (2003a), pp. 241–260. – DOI: [doi.org/10.1016/S0047-259X\(03\)00053-8](https://doi.org/10.1016/S0047-259X(03)00053-8)
- [Menneteau 2005] MENNETEAU, L.: Some laws of the iterated logarithm in Hilbertian autoregressive models. *J. Multivariate Anal.* 92 (2005), pp. 405–425. – DOI: [doi.org/10.1016/j.jmva.2003.07.001](https://doi.org/10.1016/j.jmva.2003.07.001)
- [Merlevède 1996b] MERLEVÈDE, F.: Lois des grands nombres et loi du logarithme itéré compacte pour des processus linéaires à valeurs dans un espace de Banach de type 2. *C. R. Acad. Sci. Paris Sér. I Math* 323 (1996b), pp. 521–524
- [Merlevède 1997] MERLEVÈDE, F.: Résultats de convergence presque sûre pour l’estimation et la prévision des processus linéaires Hilbertiens. *C. R. Acad. Sci. Paris Sér. I* 324 (1997), pp. 573–576
- [Mourid 1993] MOURID, T.: Processus autorégressifs Banachiques d’ordre supérieur. *C. R. Acad. Sci. Paris Sér. I Math.* 317 (1993), pp. 1167–1172
- [Mourid 2004] MOURID, T.: Processus autorégressifs Hilbertiens à coefficients aléatoires. *Ann. I. S. U. P.* 48 (2004), pp. 79–85. – URL [http://www.lsta.lab.upmc.fr/modules/resources/download/labsta/Annales\\_ISUP/Couv\\_ISUP\\_48-3.pdf](http://www.lsta.lab.upmc.fr/modules/resources/download/labsta/Annales_ISUP/Couv_ISUP_48-3.pdf)
- [Nason 2008] NASON, G. P.: *Wavelet Methods in Statistics with R*. Springer, Berlin, 2008. – ISBN 9780387759616
- [Ogden 1997] OGDEN, R. T.: *Essential Wavelets for Statistical Applications and Data Analysis*. Birkhauser, Boston, MA, 1997. – ISBN 9781461207092
- [Pumo 1992] PUMO, B.: *Estimation et prévision de processus autorégressifs fonctionnels*. Paris, Université de Paris 6, PhD. Thesis, 1992
- [Pumo 1998] PUMO, B.: Prediction of continuous time processes by  $C_{[0,1]}$ -valued autoregressive process. *Stat. Inference Stoch. Process.* 1 (1998), pp. 297–309
- [Ramsay and Silverman 2005] RAMSAY, J. O. ; SILVERMAN, B. W.: *Functional data analysis, 2nd ed.* Springer, New York, 2005. – ISBN 978038740080

- [Ruiz-Medina 2012] RUIZ-MEDINA, M. D.: Spatial functional prediction from spatial autoregressive Hilbertian processes. *Environmetrics* 23 (2012), pp. 119–128. – DOI: [doi.org/10.1002/env.1143](https://doi.org/10.1002/env.1143)
- [Tibshirani 1996] TIBSHIRANI, R. J.: Regression shrinkage and selection via the LASSO. *J. R. Stat. Soc. Ser. B Stat. Methodol.* 58 (1996), pp. 267–288. – URL <http://www.jstor.org/stable/2346178>
- [Touloumis 2015] TOULOUMIS, A.: Nonparametric Stein-type shrinkage covariance matrix estimators in high-dimensional settings. *Comput. Statist. Data Anal.* 83 (2015), pp. 251–261. – DOI: [doi.org/10.1016/j.csda.2014.10.018](https://doi.org/10.1016/j.csda.2014.10.018)
- [Vidakovic 1998] VIDAKOVIC, B.: Nonlinear wavelet shrinkage with Bayes rules and Bayes factors. *J. Amer. Statist. Assoc.* 93 (1998), pp. 173–179. – DOI: [doi.org/10.2307/2669614](https://doi.org/10.2307/2669614)
- [Wand and Ormerod 2011] WAND, M. P. ; ORMEROD, J. T.: Penalized wavelets: embedding wavelets into semiparametric regression. *Electron. J. Stat.* 5 (2011), pp. 1654–1717. – DOI: [doi.org/10.1214/11-EJS652](https://doi.org/10.1214/11-EJS652)
- [Zhao et al. 2015] ZHAO, Y. ; CHEN, H. ; OGDEN, R. T.: Wavelet-based weighted LASSO and screening approaches in functional linear regression. *J. Comput. Graph. Statist.* 24 (2015), pp. 655–675. – DOI: [doi.org/10.1080/10618600.2014.925458](https://doi.org/10.1080/10618600.2014.925458)
- [Zhao et al. 2012] ZHAO, Y. ; OGDEN, R. T. ; REISS, P. T.: Wavelet-based LASSO in functional linear regression. *J. Comput. Graph. Statist.* 21 (2012), pp. 600–617. – DOI: [doi.org/10.1080/10618600.2012.679241](https://doi.org/10.1080/10618600.2012.679241)
- [Zou 2006] ZOU, H.: The adaptive LASSO and its oracle properties. *J. Amer. Statist. Assoc.* 101 (2006), pp. 1418–1429. – DOI: [doi.org/10.1198/016214506000000735](https://doi.org/10.1198/016214506000000735)