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

Javier Álvarez-Liébana¹, Denis Bosq² and M. Dolores Ruiz-Medina¹

2nd April 2018

¹ Department of Statistics and O. R., University of Granada, Spain.
² 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.

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, motivated, in part, by the statistical analysis of high-dimensional data, as well as data with a continuous (infinite-dimensional) nature (see [11, 12, 19, 25, 44, 45, 52, 53], among others). A selected set of new developments and branches of the functional statistical theory can be found in [10]. Particularly, in the current literature, we refer to the recent books [34] and [35], surveys [15], and Special Issue [29], and the references therein. 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.

In the functional regression model context, we particularly refer to the case where the predictor is a random function, and the response is scalar, since this topic has been widely developed. Different specifications of the regression operator arise, in several applied fields, such as chemometrics, biology, economics and climatology. An active research area in the literature is devoted to estimate the regression operator. To avoid the computational (high-dimensionality) limitations of the nonparametric approach, several parametric and semi-parametric methods have been proposed (see, for example, [23], and the references therein). In particular, in [23], to avoid the problem of high dimensionality, a combination of a spline approximation and the one-dimensional Nadaraya-Watson approach is proposed. The generalizations in the case of more regressors (all functional, or both functional and real) have been addressed in the nonparametric, semi-parametric, and parametric frameworks (see for an overview [1], [22], and [26]). For functional response and covariate, in the nonparametric regression framework, we refer to [24], where a functional version of the Nadaraya-Watson estimator is proposed for the estimation of the regression operator. Its pointwise asymptotic normality is derived as well. Resampling ideas are applied to overcome the difficulties arising in the estimation of the asymptotic bias and variance. Semifunctional partial linear regression, introduced in [2], allows the prediction of a real-valued random variable from a set of real explanatory variables, and a time-dependent functional explanatory variable. The asymptotic, and practical properties in real-data applications, of the estimators derived are studied as well. Motivated by genetic and environmental applications, in [14], a semi-parametric maximum likelihood method for the estimation of odds ratio association parameters is also developed in a high dimensional data framework.

In the autoregressive Hilbertian time series framework, several estimation and prediction procedures have been proposed with the derivation of the corresponding asymptotic theory. In [39], the weak-convergence to the normal distribution of the estimator of the autocorrelation operator studied, based on projection into the theoretical eigenvectors, is established, under suitable conditions. In [11] and [13], the problem of prediction of linear processes in function spaces is addressed. In particular, sufficient conditions for the consistency of the empirical auto- and cross-covariance operators are obtained. The asymptotic normal distribution of the empirical auto-covariance operator is also derived. Projection into the empirical eigenvectors is considered, in the formulation of a componentwise estimator of the autocorrelation operator, in absence of information about the theoretical ones. Asymptotic properties of the empirical eigenvalues and eigenvectors are also analyzed. In [31], the efficiency of a componentwise estimator of the autocorrelation operator, based on projection into the empirical eigenvector system of
the auto-covariance operator, is obtained. In [40], 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 is proved. Sufficient conditions for the weak-convergence of the ARH(1) plug-in predictor to a Hilbert-valued Gaussian random variable are derived in [41]. High deflection results or large and moderate deviations for infinite-dimensional autoregressive processes are derived in [42]. The law of the iterated logarithm for the covariance operator estimator is obtained in [43]. The main properties for the class of autoregressive Hilbertian processes with random coefficients are investigated in [47]. Interesting extensions of the autoregressive Hilbertian framework can be found in [36], who offers a new estimate, based on the spectral decomposition of the autocorrelation operator, and not of the auto-covariance operator. The first generalization on autoregressive processes of order greater than one is proposed in [46], in order to improve prediction. ARHX (1) models, i.e., autoregressive Hilbertian processes with exogenous variables are formulated in [17]. In [30] and [31] a doubly stochastic formulation of the autoregressive Hilbertian process is studied. The ARHD model is introduced in [38], taking into account the regularity of trajectories through the derivatives. Recently, [16] introduced the conditional autoregressive Hilbertian processes (CARH processes), as a new class of processes, and developed parallel projection estimation methods to predict such processes. In the Banach-valued context, we refer to the papers by [6, 20, 50, 51], 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 auto-covariance operator. That is the case, for example, of an autocorrelation operator defined as a continuous function of the autocovariance operator (see Remark 1 below). A componentwise estimator of the autocorrelation operator is formulated in terms of such an eigenvector system. In the derivation of the results presented in this paper we assume that the eigenvectors \( \phi_j, j \geq 1 \) of the autocovariance operator \( C \) are known. That is the case, for example, of defining our random initial condition as the solution, in the mean-square sense, of a stochastic differential equation driven by white noise (e.g., the Wiener measure). However, 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 (see Remark 3 below), in terms of a common and known wavelet basis. Smoothing functional data with a suitable penalization norm in terms of wavelets also could lead to an empirical regularized diagonal approximation of the autocorrelation operator in terms of wavelets. Finally, we refer to shrinkage estimation of the autocovariance and cross covariance operators in terms of a suitable common wavelet basis, allowing a diagonal empirical approximation of both operators, in terms of such a basis (see also Section 6).

Under the setting of assumptions established in this paper (see Section 2), convergence to the auto-
correlation operator, in \( L^2 \)-sense in the space of Hilbert-Schmidt operators \( S(H) \), i.e., convergence in the space \( L^2_{S(H)}(\Omega, \mathcal{A}, \mathcal{P}) \), is derived, for the formulated componentwise estimator. Consistency then follows in \( S(H) \) (see Section 3.1). Under the same setting of conditions, consistency in \( H \) of the associated ARH(1) plug-in predictor is obtained as well, from its convergence in the \( L^1 \)-sense in the Hilbert space \( H \), i.e., in the space \( L^1_H(\Omega, \mathcal{A}, \mathcal{P}) \) (see Section 3.2). The Gaussian framework is analyzed in Section 4. In that framework, the numerical examples studied in Section 5 illustrate the behavior of the formulated componentwise autocorrelation operator estimator and associated predictor for large sample sizes. A comparative study with alternative ARH(1) prediction techniques (even in a non-diagonal scenario), based on componentwise parameter estimators of the autocorrelation operator (including the case of empirical eigenvectors), as well as based on kernel (nonparametric) functional estimators, and penalized, spline and wavelet, estimators, is also performed. Final conclusions and some guidelines for the application of the proposed approach from real-data are provided in Section 6.

2 Preliminaries.

The preliminary definitions and lemmas, that will be applied in the derivation of the main results of this paper, in the context of ARH(1) processes, are now provided. In the following, let us denote by \( H \) a real separable Hilbert space. Recall that a zero-mean ARH(1) process \( X = (X_n, n \in \mathbb{Z}) \) satisfies the equation (see [11], among others)

\[
X_n = \rho (X_{n-1}) + \varepsilon_n, \quad n \in \mathbb{Z},
\]

where \( \rho \) denotes the autocorrelation operator of process \( X \), which belongs to the space \( \mathcal{L}(H) \) of bounded linear operators, such that \( \| \rho^k \|_{\mathcal{L}(H)} < 1 \), for \( k \geq k_0 \), and for 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 considered to be strong-white noise, i.e., \( \varepsilon \) is a Hilbert-valued zero-mean stationary process, with independent and identically distributed components in time, and with \( \sigma^2_\varepsilon = E \| \varepsilon_n \|^2_H < \infty \), 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 will be made:

**Assumption A1.** The autocovariance operator \( C = E[X_n \otimes X_n] = E[X_0 \otimes X_0] \), for every \( n \in \mathbb{Z} \), is a positive self-adjoint and trace operator. It then admits the following
diagonal spectral representation, in terms of its eigenvectors \((\phi_j, j \geq 1)\)

\[
C = \sum_{j=1}^{\infty} C_j \phi_j \otimes \phi_j, \tag{2}
\]

where \(\sum_{j=1}^{\infty} C_j < \infty\), with \((C_j, j \geq 1)\) denoting the system of real positive eigenvalues of \(C\), arranged in decreasing order of their magnitudes \(C_1 \geq C_2 \geq \cdots \geq C_j \geq C_{j+1} \geq \cdots > 0\).

**Assumption A2.** The autocorrelation operator \(\rho\) is a self-adjoint and Hilbert-Schmidt operator, admitting the following diagonal spectral decomposition:

\[
\rho = \sum_{j=1}^{\infty} \rho_j \phi_j \otimes \phi_j, \quad \sum_{j=1}^{\infty} \rho_j^2 < \infty, \tag{3}
\]

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

**Remark 1** **Assumption A2** holds, in particular, when operator \(\rho\) is defined as a continuous function of operator \(C\) (see [18], pp. 119-140). See also Remark 3 below.

In the following, for any \(n \in \mathbb{Z}\), let \(D = E[X_n \otimes X_{n+1}] = E[X_0 \otimes X_1]\) be the cross-covariance operator of the ARH(1) process \(X = (X_n, n \in \mathbb{Z})\).

Under **Assumptions A1-A2**, by projection of equation (1) into the orthonormal system \((\phi_j, j \geq 1)\), we obtain, for each \(j \geq 1\), the following AR(1) equation:

\[
X_{n,j} = \rho_j X_{n-1,j} + \varepsilon_{n,j}, \quad n \in \mathbb{Z}, \tag{4}
\]

where \(X_{n,j} = \langle X_n, \phi_j \rangle_H\) and \(\varepsilon_{n,j} = \langle \varepsilon_n, \phi_j \rangle_H\), for all \(n \in \mathbb{Z}\).

Under **Assumptions A1-A2**, from equation (4), for each \(j \geq 1\),

\[
\rho_j = \rho(\phi_j)(\phi_j) = \langle \phi_j, DC^{-1}(\phi_j) \rangle_H
= \langle D(\phi_j), \phi_j \rangle_H \langle C^{-1}(\phi_j), \phi_j \rangle_H
= \frac{E[X_{n,j} X_{n-1,j}]}{E[X_{n-1,j}^2]} = \frac{D_j}{C_j}, \quad \forall n \in \mathbb{Z}, \tag{5}
\]
where \( D_j = \langle D(\phi_j), \phi_j \rangle_H = E[X_{n,j}X_{n-1,j}], \) \( C_j^{-1} = [E[X^2_{n-1,j}]]^{-1} \) and \( X_{n,j} = \langle X_n, \phi_j \rangle_H, \ j \geq 1, \) since

\[
D = \sum_{j=1}^{\infty} D_j \phi_j \otimes \phi_j, \quad D_j = \rho_j C_j, \quad j \geq 1. \tag{6}
\]

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

\[
\|Z\|_{L^2_H(\Omega, A, \mathcal{P})} = \sqrt{E\|Z\|^2_H}, \quad \forall Z \in L^2_H(\Omega, A, \mathcal{P}). \tag{7}
\]

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

\[
E\|Z - Y\|_H = 0.
\]

In the next section, we will consider, in particular, \( H = S(H) \), the Hilbert space of Hilbert-Schmidt operators on a Hilbert space \( H \), i.e., we will consider, the space \( L^2_S(H) (\Omega, A, \mathcal{P}) \) with the seminorm

\[
\|Y\|_{L^2_S(H) (\Omega, A, \mathcal{P})}^2 = E\|Y\|^2_S(H), \quad \forall Y \in L^2_S(H) (\Omega, A, \mathcal{P}). \tag{8}
\]

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

\[
X_n = \sum_{j=1}^{\infty} \sqrt{C_j} \frac{\langle X_n, \phi_j \rangle_H}{\sqrt{C_j}} \phi_j = \sum_{j=1}^{\infty} \sqrt{C_j} \eta_j(n) \phi_j, \tag{9}
\]

\[
\varepsilon_n = \sum_{j=1}^{\infty} \sigma_j \frac{\langle \varepsilon_n, \phi_j \rangle_H}{\sigma_j} \phi_j = \sum_{j=1}^{\infty} \sigma_j \tilde{\eta}_j(n) \phi_j, \tag{10}
\]

where \( \eta_j(n) = \frac{\langle X_n, \phi_j \rangle_H}{\sqrt{C_j}} = \frac{X_n \phi_j}{\sqrt{C_j}} \) and \( \tilde{\eta}_j(n) = \frac{\langle \varepsilon_n, \phi_j \rangle_H}{\sigma_j} = \frac{\varepsilon_n \phi_j}{\sigma_j} \), for every \( n \in \mathbb{Z} \), and for each \( j \geq 1 \). Here, under Assumptions A1-A2, for \( R_\varepsilon = E[\varepsilon_n \otimes \varepsilon_n] = E[\varepsilon_0 \otimes \varepsilon_0], \ n \in \mathbb{Z}, \)

\[
R_\varepsilon \phi_j = \sigma_j^2 \phi_j, \quad j \geq 1,
\]

where, as before, \( (\phi_j, \ j \geq 1) \) denotes the system of eigenvectors of the autocovariance operator \( C \), and

\[
\sum_{j \geq 1} \sigma_j^2 = \sigma_0^2 = E\|\varepsilon_n\|_H^2, \quad \text{for all} \ n \in \mathbb{Z}.
\]

The following lemma provides the convergence, in the seminorm of \( L^2_H(\Omega, A, \mathcal{P}) \), of the series expan-
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

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

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

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

Similar assertions hold for the biorthogonal series representation

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

Proof.

Under Assumption A1, from the trace property of $C$, the sequence $\left( \hat{X}_{n,M} = \sum_{j=1}^{M} \sqrt{C_j} \eta_j(n) \phi_j, \ M \geq 1 \right)$ satisfies, for $M$ sufficiently large, and $L > 0$, arbitrary,

$$\|\hat{X}_{n,M+L} - \hat{X}_{n,M}\|_{L^2_H(\Omega, \mathcal{A}, P)}^2 = E \|\hat{X}_{n,M+L} - \hat{X}_{n,M}\|_H^2$$

$$= \sum_{j=M+1}^{M+L} \sum_{k=M+1}^{M+L} \sqrt{C_j} \sqrt{C_k} E[\eta_j(n) \eta_k(n)] \langle \phi_j, \phi_k \rangle_H$$

$$= \sum_{j=M+1}^{M+L} C_j \to 0, \ \text{when} \ M \to \infty,$$

since, under Assumption A1, $\sum_{j=1}^{\infty} C_j < \infty$, hence, $\left( \sum_{j=1}^{M} C_j, \ M \geq 1 \right)$ is a Cauchy sequence. Thus, $\sum_{j=M+1}^{M+L} C_j$ converges to zero when $M \to \infty$, for $L > 0$, arbitrary. From equation (13), $\left( \hat{X}_{n,M} = \sum_{j=1}^{M} \sqrt{C_j} \eta_j(n) \phi_j, \ M \geq 1 \right)$ is also a Cauchy sequence in $L^2_H(\Omega, \mathcal{A}, P)$. Thus, the sequence $\left( \hat{X}_{n,M}, \ M \geq 1 \right)$ has finite limit in $L^2_H(\Omega, \mathcal{A}, P)$, for all $n \in \mathbb{Z}$. 

\section*{7}
Furthermore,

\[
\lim_{M \to \infty} E \left\| X_n - \hat{X}_{n,M} \right\|^2_H = E \left\| X_n \right\|^2_H \\
+ \lim_{M \to \infty} \sum_{j=1}^{M} \sum_{h=1}^{M} \sqrt{C_j} \sqrt{C_h} E \left[ \eta_j(n) \eta_h(n) \right] \langle \phi_j, \phi_h \rangle_H \\
- 2 \lim_{M \to \infty} \sum_{j=1}^{M} \sqrt{C_j} E \left[ \langle X_n, \eta_j(n) \phi_j \rangle_H \right] = \sigma_X^2 - \lim_{M \to \infty} \sum_{j=1}^{M} C_j = 0.
\]

(14)

In the derivation of the identities in (13) and (14), we have applied that, for every \( j, h \geq 1 \),

\[
C \phi_j = C_j \phi_j, \quad \sigma_X^2 = E \| X_n \|^2_H = \sum_{j=1}^{\infty} C_j < +\infty, \\
E \left[ \eta_j(n) \eta_h(n) \right] = \delta_{j,h}, \quad \langle \phi_j, \phi_h \rangle_H = \delta_{j,h}, \quad E \left[ \langle X_n, \eta_j(n) \phi_j \rangle_H \right] = \sqrt{C_j}.
\]

(15)

Moreover, from identities in (15),

\[
\left\| E \left[ \left( X_n - \lim_{M \to \infty} \hat{X}_{n,M} \right) \otimes \left( X_n - \lim_{M \to \infty} \hat{X}_{n,M} \right) \right] \right\|^2_{S(H)} \\
\leq \left\| E[ X_n \otimes X_n] + \lim_{M \to \infty} \sum_{j=1}^{M} \sum_{h=1}^{M} \sqrt{C_j} \sqrt{C_h} \phi_j \otimes \phi_h E[ \eta_j(n) \eta_h(n) \right] \\
- 2 \lim_{M \to \infty} \sum_{j=1}^{M} E[ X_n \otimes \sqrt{C_j} \eta_j(n) \phi_j] \right\|^2_{S(H)} \\
= \left\| E[ X_n \otimes X_n] + \lim_{M \to \infty} \left[ \sum_{j=1}^{M} C_j \phi_j \otimes \phi_j - 2 \sum_{j=1}^{M} C_j \phi_j \otimes \phi_j \right] \right\|^2_{S(H)} \\
= \left\| E[ X_n \otimes X_n] + \lim_{M \to \infty} \sum_{j=1}^{M} C_j \phi_j \otimes \phi_j \right\|^2_{S(H)} = 0.
\]

(16)

In a similar way, we can derive convergence to \( \varepsilon_n \), in \( L^2_H(\Omega, \mathcal{A}, \mathbb{P}) \), of the series \( \sum_{j=1}^{\infty} \sigma_j \eta_j(n) \phi_j \), for every \( n \in \mathbb{Z} \), since \( \varepsilon \) is assumed to be strong-white noise, and hence, its covariance operator \( R_\varepsilon \) is in the trace class. We can also obtain an analogous to equation (16).
In equations (9)-(10), for every \( n \in \mathbb{Z} \),

\[
E[\eta_j(n)] = 0, \quad E[\eta_j(n)\eta_h(n)] = \delta_{j,h}, \quad j, h \geq 1, \quad n \in \mathbb{Z} \quad (17)
\]

\[
E[\overline{\eta}_j(n)] = 0, \quad E[\overline{\eta}_j(n)\overline{\eta}_h(n)] = \delta_{j,h}, \quad j, h \geq 1, \quad n \in \mathbb{Z} \quad (18)
\]

Note that, from Assumption A2 for each \( j \geq 1 \), \((X_{n,j}, n \in \mathbb{Z})\) in equation (4) defines a stationary and invertible AR(1) process. In addition, from equations (9) and (15), for every \( n \in \mathbb{Z} \), and \( j, p \geq 1 \),

\[
X_n = \sum_{j=1}^{\infty} X_{n,j}\phi_j,
\]

\[
E[X_{n,j}X_{n,p}] = \sum_{k=0}^{\infty} \sum_{h=0}^{\infty} \rho_j^k\rho_p^h E[\varepsilon_{n-k,j}\varepsilon_{n-h,p}] = \delta_{j,p} \sum_{k=0}^{\infty} \rho_j^{2k}\sigma_j^2 = \delta_{j,p} \frac{\sigma_j^2}{1-\rho_j^2},
\]

\[
E\|X_n\|_H^2 = \sum_{j=1}^{\infty} E[X_{n,j}^2] = \sum_{j=1}^{\infty} (C(\phi_j), \phi_j)_H = \sum_{j=1}^{\infty} C_j = \sigma_X^2 < \infty,
\]

which implies that \( C_j = \frac{\sigma_j^2}{1-\rho_j^2} \), for each \( j \geq 1 \). In particular, we obtain, for each \( j \geq 1 \), and for every \( n \in \mathbb{Z} \),

\[
E[\eta_j(n)\eta_j(n+1)] = E\left[\frac{X_{n,j}X_{n+1,j}}{\sqrt{C_j}}\right] = \frac{E[X_{n,j}X_{n+1,j}]}{C_j}
\]

\[
= \sum_{k=0}^{\infty} \sum_{h=0}^{\infty} \rho_j^{k+h} E[\varepsilon_{n-k,j}\varepsilon_{n+1-h,j}] / C_j
\]

\[
= \sum_{k=0}^{\infty} \rho_j^{2k+1}\sigma_j^2 / C_j = \sigma_j^2 / C_j (1-\rho_j^2) = \rho_j, \quad n \in \mathbb{Z}.
\]

Remark 2 From equation (4) and Lemma 1, keeping in mind that \( C_j = \frac{\sigma_j^2}{1-\rho_j^2} \), for each \( j \geq 1 \), the following invertible and stationary AR(1) process can be defined as

\[
\eta_j(n) = \rho_j \eta_j(n-1) + \sqrt{1-\rho_j^2}\overline{\eta}_j(n), \quad 0 < \rho_j^2 \leq \rho_j < 1,
\]

where, for each \( j \geq 1 \), \((\eta_j(n), n \in \mathbb{Z})\) and \((\overline{\eta}_j(n), n \in \mathbb{Z})\) are respectively introduced in equations (9)-(10). In the following, for each \( j \geq 1 \), we assume that \( E[\overline{\eta}_j(n)]^4 < \infty \), for every \( n \in \mathbb{Z} \), to ensure ergodicity for all second-order moments, in the mean-square sense (see, for example, [33], pp. 192-193).
Furthermore,
\[ D = E[X_n \otimes X_{n+1}] = \sum_{j=1}^{\infty} \sum_{p=1}^{\infty} E \left[ \langle X_n, \phi_j \rangle_H \langle X_{n+1}, \phi_p \rangle_H \right] \phi_j \otimes \phi_p \]
\[ = \sum_{j=1}^{\infty} \sum_{p=1}^{\infty} \sqrt{C_j} \sqrt{C_p} E \left[ \langle X_n, \phi_j \rangle_H \langle X_{n+1}, \phi_p \rangle_H \right] \phi_j \otimes \phi_p \]
\[ = \sum_{j=1}^{\infty} \sum_{p=1}^{\infty} \sqrt{C_j} \sqrt{C_p} E \left[ \eta_j(n) \eta_p(n+1) \right] \phi_j \otimes \phi_p. \]  
(22)

Remark 3 In particular, Assumption A2 holds if the following orthogonality condition is satisfied
\[ E \left[ \eta_j(n) \eta_p(n+1) \right] = \delta_{j,p}, \quad j, p \geq 1, \quad n \in \mathbb{Z}, \]  
(23)

where \( \delta_{j,p} \) denotes the Kronecker Delta function. In practice, unconditional bases, like wavelet bases, lead to a sparse representation for functional data (see, for example, [48, 49, 56] 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 [37]). The Discrete Wavelet Transform (DWT) approximately decorrelates or whitens data (see [56]). 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 is formulated in this section. Their convergence to the corresponding theoretical functional values, in the spaces \( \mathcal{L}^2_{S(H)}(\Omega, \mathcal{A}, P) \) and \( \mathcal{L}_H(\Omega, \mathcal{A}, P) \), is respectively derived as well. Their consistency in the spaces \( S(H) \) and \( H \) then follows.

From equation (5), for each \( j \geq 1 \), and for a given sample size \( n \geq 1 \), one can consider the usual respective moment-based estimators \( \hat{D}_{n,j} \) and \( \hat{C}_{n,j} \) of \( D_j \) and \( C_j \), in the AR(1) framework, given by
\[ \hat{D}_{n,j} = \frac{1}{n-1} \sum_{i=0}^{n-2} X_{i,j} X_{i+1,j}, \quad \hat{C}_{n,j} = \frac{1}{n} \sum_{i=0}^{n-1} X_{i,j}^2. \]  
(24)
The following truncated componentwise estimator of $\rho$ is then formulated:

\[
\hat{\rho}_n = \sum_{j=1}^{k_n} \hat{\rho}_{n,j} \phi_j \otimes \phi_j,
\]

where, for each $j \geq 1$,

\[
\hat{\rho}_{n,j} = \frac{\hat{D}_{n,j}}{\hat{C}_{n,j}} = \frac{\frac{1}{n-1} \sum_{i=0}^{n-2} X_{i,j} X_{i+1,j}}{\frac{1}{n} \sum_{i=0}^{n-1} X_{i,j}^2} = \frac{n}{n-1} \sum_{i=0}^{n-2} X_{i,j} X_{i+1,j}.
\]

Here, the truncation parameter $k_n$ 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

\[
\lim_{n \to \infty} k_n = \infty, \quad \frac{k_n}{n} < 1, \quad n > 1.
\]

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 (25) is such that

\[
\lim_{n \to \infty} C_{k_n} \sqrt{n} = \infty.
\]

**Remark 4** Assumption A3 has also been considered in p. 217 of [11], to ensure weak consistency of the proposed estimator of $\rho$, as well as, in [39] (see Proposition 4, p. 902), in the derivation of asymptotic normality.

From Remark 2, for each $j \geq 1$, $\eta_j = (\eta_j(n), n \in \mathbb{Z})$ in equation (21) defines a stationary and invertible AR(1) process, ergodic in the mean-square sense (see, for example, [5]). Therefore, in view of equations (17) and (20), for any $j \geq 1$, there exist two positive constants $K_{j,1}$ and $K_{j,2}$ such that the following identities hold:

\[
\lim_{n \to \infty} \frac{E \left[ 1 - \frac{1}{n} \sum_{i=0}^{n-1} \eta_j^2(i) \right]^2}{K_{j,1}} = K_{j,2}.
\]

\[
\lim_{n \to \infty} \frac{E \left[ \rho_j - \frac{1}{n} \sum_{i=0}^{n-2} \eta_j(i) \eta_j(i+1) \right]^2}{K_{j,2}} = K_{j,2}.
\]
Equations (28)-(29) imply, for $n$ sufficiently large,

$$\text{Var} \left( \frac{1}{n} \sum_{i=0}^{n-1} \eta_j^2(i) \right) \leq \frac{\tilde{K}_{j,1}}{n},$$  \hspace{1cm} (30)

$$\text{Var} \left( \frac{1}{n-1} \sum_{i=0}^{n-2} \eta_j(i) \eta_j(i+1) \right) \leq \frac{\tilde{K}_{j,2}}{n},$$  \hspace{1cm} (31)

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

$$E \left[ \left( 1 - \frac{1}{n} \sum_{i=0}^{n-1} \eta_j^2(i) \right)^2 \right] \leq \frac{\tilde{K}_{j,1}}{n},$$  \hspace{1cm} (32)

$$E \left[ \left( \rho_j - \frac{1}{n-1} \sum_{i=0}^{n-1} \eta_j(i) \eta_j(i+1) \right)^2 \right] \leq \frac{\tilde{K}_{j,2}}{n}.$$  \hspace{1cm} (33)

The following assumption is now considered.

**Assumption A4.** $S = \sup_{j \geq 1} (\tilde{K}_{j,1} + \tilde{K}_{j,2}) < \infty$.

**Remark 5** From equation (26), applying Cauchy–Schwarz's inequality, we obtain, for each $j \geq 1$,

$$|\hat{\rho}_{n,j}| = \frac{n}{n-1} \left| \sum_{i=0}^{n-2} X_{i,j} X_{i+1,j} \right| \leq \frac{n}{n-1} \sqrt{\frac{\sum_{i=0}^{n-2} X_{i,j}^2 \sum_{i=0}^{n-2} X_{i+1,j}^2}{\sum_{i=0}^{n-1} X_{i,j}^2}} \leq \frac{n}{n-1} \text{ a.s.}$$  \hspace{1cm} (34)

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

In the following proposition, convergence of $\hat{\rho}_{k_n}$ to $\rho$, in the space $L^2_{S(H)}(\Omega, \mathcal{A}, \mathcal{P})$, 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:

$$\lim_{n \to \infty} \| \rho - \hat{\rho}_{k_n} \|_{L^2_{S(H)}(\Omega, \mathcal{A}, \mathcal{P})}^2 = 0.$$  \hspace{1cm} (35)
Specifically,
\[ \| \rho - \hat{\rho}_n \|^2_{L_2(H^2(\Omega,A,P))} \leq g(n), \quad \text{with} \quad g(n) = O \left( \frac{1}{C_{\rho n}} \right), \quad n \to \infty. \] (36)

Remark 6 Corollary 4.3, in [11], p. 107, 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 Section 1.3, in p. 27, in [11]).

Proof. For each \( j \geq 1 \), the following almost surely inequality is satisfied:
\[
| \rho_j - \hat{\rho}_{n,j} | = \frac{|D_j - \hat{D}_{n,j}C_j|}{C_j \hat{C}_{n,j}} = \frac{|D_j \hat{C}_{n,j} - \hat{D}_{n,j}C_j| + |\hat{D}_{n,j} \hat{C}_{n,j} - \hat{D}_{n,j} C_j|}{C_j \hat{C}_{n,j}} \\
\leq \frac{1}{C_j} \left( |\hat{\rho}_{n,j}| \left| C_j - \hat{C}_{n,j} \right| + \left| D_j - \hat{D}_{n,j} \right| \right). \tag{37}
\]

Thus, under Assumptions A1–A2, from equation (34), for each \( j \geq 1 \),
\[
(\rho_j - \hat{\rho}_{n,j})^2 \leq \frac{1}{C_j^2} \left( |\hat{\rho}_{n,j}| \left| C_j - \hat{C}_{n,j} \right| + \left| D_j - \hat{D}_{n,j} \right| \right)^2 \\
\leq \frac{2}{C_j^2} \left( (\hat{\rho}_{n,j})^2 \left( C_j - \hat{C}_{n,j} \right)^2 + \left( D_j - \hat{D}_{n,j} \right)^2 \right) \\
\leq \frac{2}{C_j^2} \left( \left( \frac{n}{n-1} \right)^2 \left( C_j - \hat{C}_{n,j} \right)^2 + \left( D_j - \hat{D}_{n,j} \right)^2 \right) \ a.s., \tag{38}
\]

which implies, for each \( j \geq 1 \),
\[
E[ (\rho_j - \hat{\rho}_{n,j})^2 ] \leq \frac{2}{C_j^2} \left( \left( \frac{n}{n-1} \right)^2 E \left[ (C_j - \hat{C}_{n,j})^2 \right] + E \left[ (D_j - \hat{D}_{n,j})^2 \right] \right). \tag{39}
\]

Under Assumption A2, from equations (25) and (39),
\[ \| \rho - \tilde{\rho}_k \|_{L^2(\Omega, A, P)}^2 = E \| \rho - \tilde{\rho}_k \|_{S(H)}^2 \geq \sum_{j=1}^{k_n} E \left[ (\rho_j - \tilde{\rho}_{n,j})^2 \right] + \sum_{j=k_n+1}^{\infty} \rho_j^2 \]

\[ \leq \sum_{j=1}^{k_n} \frac{2}{C_j} \left( \frac{n}{n-1} \right)^2 E \left[ (C_j - \tilde{C}_{n,j})^2 \right] + E \left[ (D_j - \tilde{D}_{n,j})^2 \right] + \sum_{j=k_n+1}^{\infty} \rho_j^2 \]

\[ \leq \frac{2}{C_k^2} \sum_{j=1}^{k_n} \left( \frac{n}{n-1} \right)^2 E \left[ (C_j - \tilde{C}_{n,j})^2 \right] + E \left[ (D_j - \tilde{D}_{n,j})^2 \right] + \sum_{j=k_n+1}^{\infty} \rho_j^2 \]

\[ \leq 2 \left( \frac{n}{n-1} \right)^2 \sum_{j=1}^{k_n} \left( E \left[ (C_j - \tilde{C}_{n,j})^2 \right] + E \left[ (D_j - \tilde{D}_{n,j})^2 \right] \right) + \sum_{j=k_n+1}^{\infty} \rho_j^2. \]  

(40)

Furthermore, from (9) and (26), for \( j \geq 1 \),

\[ \tilde{C}_{n,j} = \frac{1}{n} \sum_{i=0}^{n-1} X_{i,j}^2 = \frac{1}{n} \sum_{i=0}^{n-1} C_j \eta_j^2(i), \]  

(41)

\[ \tilde{D}_{n,j} = \frac{1}{n-1} \sum_{i=0}^{n-2} X_{i,j} X_{i+1,j} = \frac{1}{n-1} \sum_{i=0}^{n-2} C_j \eta_j(i) \eta_j(i+1), \]  

(42)

where, considering equation (6),

\[ D_j = E [X_{n,j} X_{n+1,j}] = C_j E [\eta_j(n) \eta_j(n+1)] = C_j \rho_j, \]  

(43)

for each \( j \geq 1 \). Equations (40)–(43) then lead to

\[ \| \rho - \tilde{\rho}_k \|_{L^2(\Omega, A, P)}^2 \leq \frac{2}{C_k^2} \left( \frac{n}{n-1} \right)^2 \sum_{j=1}^{k_n} C_j^2 \left[ E \left[ \left( \frac{1}{n} \sum_{i=0}^{n-1} \eta_j^2(i) \right)^2 \right] \right. \]

\[ + E \left[ \left( \rho_j - \frac{1}{n-1} \sum_{i=0}^{n-2} \eta_j(i+1) \eta_j(i) \right)^2 \right] \]

\[ + \sum_{j=k_n+1}^{\infty} \rho_j^2. \]  

(44)
For each \( j \geq 1 \), and for \( n \) sufficiently large, considering equations (32)–(33), under Assumption A4

\[
E \| \rho - \hat{\rho}_{k_n} \|^2_{S(H)} \leq 2 \left( \frac{n}{n-1} \right)^2 \frac{k_n}{C_{k_n}^2} \sum_{j=1}^{k_n} C_j^2 \left( \frac{\tilde{K}_{j,1} + \tilde{K}_{j,2}}{n} \right) + \sum_{j=k_n+1}^{\infty} \rho_j^2 \leq 2S \left( \frac{n}{n-1} \right)^2 \frac{k_n}{C_{k_n}^2} \sum_{j=1}^{k_n} C_j^2 + \sum_{j=k_n+1}^{\infty} \rho_j^2.
\]

(45)

From the trace property of operator \( C \),

\[
\lim_{n \to \infty} \frac{k_n}{C_{k_n}^2} \sum_{j=1}^{k_n} C_j^2 = \sum_{j=1}^{\infty} C_j^2 < \infty,
\]

(46)

and from the Hilbert-Schmidt property of \( \rho \),

\[
\lim_{n \to \infty} \sum_{j=k_n+1}^{\infty} \rho_j^2 = 0.
\]

(47)

Thus, in view of equations (45)–(47),

\[
\| \rho - \hat{\rho}_{k_n} \|^2_{L^2_{S(H)}(\Omega,A,P)} = E \| \rho - \hat{\rho}_{k_n} \|^2_{S(H)} \leq g(n) = O \left( \frac{1}{C_{k_n}^2 n} \right), \ n \to \infty,
\]

(48)

where

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

(49)

Under Assumption A3, equations (48) and (49) imply

\[
\lim_{n \to \infty} \| \rho - \hat{\rho}_{k_n} \|^2_{L^2_{S(H)}(\Omega,A,P)} = 0,
\]

as we wanted to prove. \( \square \)

Note that consistency of \( \hat{\rho}_{k_n} \) in the space \( S(H) \) directly follows from equation (35) 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 \( n \to \infty \),

\[
\| \rho - \hat{\rho}_{k_n} \|_{S(H)} \to^p 0,
\]

(50)

where, as usual, \( \to^p \) denotes the convergence in probability.
3.2 Consistency of the ARH(1) plug-in predictor.

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

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

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

$$\|A(X)\|_H \leq \|A\|_{\mathcal{L}(H)} \|X\|_H.$$  \hspace{1cm} (52)

In the following, we denote by

$$\hat{X}_n = \hat{\rho}_{kn} (X_{n-1})$$

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 $\hat{X}_n = \hat{\rho}_{kn} (X_{n-1})$ in $H$.

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

$$\lim_{n \to \infty} E \| (\rho - \hat{\rho}_{kn}) (X_{n-1}) \|_H = 0.$$ \hspace{1cm} (54)

Specifically,

$$E \| (\rho - \hat{\rho}_{kn}) (X_{n-1}) \|_H \leq h (n), \quad h (n) = O \left( \frac{1}{C_{kn} \sqrt{n}} \right), \quad n \to \infty.$$ \hspace{1cm} (55)

In particular,

$$\| (\rho - \hat{\rho}_{kn}) (X_{n-1}) \|_H \to^p 0,$$ \hspace{1cm} (56)

where, as usual, $\to^p$ denotes the convergence in probability.

**Proof.**

From (52) and Proposition 1, for $n$ sufficiently large, the following almost surely inequality holds:

$$\| \rho (X_{n-1}) - \hat{X}_n \|_H \leq \| \rho - \hat{\rho}_{kn} \|_{\mathcal{L}(H)} \| X_{n-1} \|_H,$$ \hspace{1cm} (57)

where, as given in equation (53), $\hat{X}_n = \hat{\rho}_{kn} (X_{n-1})$. Thus,

$$E \| \rho (X_{n-1}) - \hat{X}_n \|_H \leq E \left[ \| \rho - \hat{\rho}_{kn} \|_{\mathcal{L}(H)} \| X_{n-1} \|_H \right].$$ \hspace{1cm} (58)

From Cauchy-Schwarz's inequality, keeping in mind that, for a Hilbert-Schmidt operator $K$, it always
holds that $\|K\|_{L(H)} \leq \|K\|_{S(H)}$, we have from equation (58),

$$E\left\| X_n - \hat{X}_n \right\|_H \leq \sqrt{E\left\| \rho - \hat{\rho}_{k_n} \right\|_{L(H)}^2 \sqrt{E\left\| X_{n-1} \right\|_H^2}} \leq \sqrt{E\left\| \rho - \hat{\rho}_{k_n} \right\|_{S(H)}^2 \sqrt{E\left\| X_{n-1} \right\|_H^2}} = \sqrt{E\left\| \rho - \hat{\rho}_{k_n} \right\|_{S(H)}^2 \sigma_X},$$

(59)

where, as before, $\sigma_X^2 = E\left\| X_n - 1 \right\|_H^2 = \infty \sum_{j=1}^{\infty} C_j < \infty$, for each $n \in \mathbb{Z}$ (see equation (15)).

Since from Proposition 1 (see equation (36)),

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

from equation (59), we obtain,

$$E \|(\rho - \hat{\rho}_{k_n})(X_{n-1})\|_H \leq h(n),$$

(60)

where $h(n) = \sigma_X \sqrt{g(n)}$, with $g(n)$ being given in (49). In particular, under Assumption A3,

$$\lim_{n \to \infty} E \|(\rho - \hat{\rho}_{k_n})(X_{n-1})\|_H = 0,$$

(61)

which implies that

$$\|(\rho - \hat{\rho}_{k_n})(X_{n-1})\|_H = \left\| \rho(X_{n-1}) - \hat{X}_n \right\|_H \to^p 0, \quad n \to \infty,$$

(62)

\[\blacksquare\]

4 The Gaussian case.

In this section, we restrict our attention to the Gaussian ARH(1) context. In that context, we prove that, under Assumptions A1-A2 and Assumption A4 holds. From equation (17),

$$E \left[ \sum_{i=0}^{n-1} \eta_j^2(i) \right] = 1, \quad n \geq 1.$$

(63)

We now compute the variance of $\sum_{i=0}^{n-1} \eta_j^2(i)$, applying the theory of quadratic forms for Gaussian vectors with correlated components. Specifically, for each $j \geq 1$, and $n \geq 2$, the $n \times 1$ random vector $\eta_j^T = (\eta_j(0), \ldots, \eta_j(n-1))$ is Multivariate Normal distributed, with null mean vector, and covariance
matrix
\[
\Sigma_{n \times n} = \begin{pmatrix}
1 & \rho_j & 0 & \ldots & 0 \\
\rho_j & 1 & \rho_j & 0 & \ldots \\
0 & \rho_j & 1 & \rho_j & \ldots \\
\vdots & \vdots & \vdots & \vdots & \ddots \\
0 & \ldots & 0 & \rho_j & 1
\end{pmatrix}_{n \times n},
\] (64)

where, under **Assumptions A1–A2**, we have applied (20), i.e.,
\[E[\eta_j(i)\eta_j(i+1)] = \rho_j,\] for every \(i \in \mathbb{Z}\), and \(j \geq 1\).

It is well-known (see, for example, [32]) that the variance of a quadratic form defined from a multivariate Gaussian vector \(\mathbf{y} \sim N(\boldsymbol{\mu}, \Lambda)\), and a symmetric matrix \(Q\) is given by:
\[
\text{Var}[\mathbf{y}^TQ\mathbf{y}] = 2\text{tr}[Q\Lambda Q] + 4\mathbf{\mu}^TQ\Lambda\mathbf{\mu},
\] (65)

where \(\text{tr}\) denotes the trace.

For each \(j \geq 1\), applying equation (65), with \(\mathbf{y} = \eta_j\), \(\Lambda = \Sigma_{n \times n}\), in (64), and \(Q = I_{n \times n}\), the \(n \times n\) identity matrix, we obtain
\[
\text{Var}[\eta_j^T I_{n \times n} \eta_j] = \text{Var}\left[\sum_{i=0}^{n-1} \eta_j^2(i)\right] = 2\text{tr}[\Sigma_{n \times n} \Sigma_{n \times n}] = 2 \left(n + 2(n-1)\rho_j^2\right).
\] (66)

Furthermore, from equation (66), for each \(j \geq 1\),
\[
\text{Var}\left[\sum_{i=0}^{n-1} \eta_j^2(i) \right] = \frac{2}{n^2} \left(n + 2(n-1)\rho_j^2\right) = \frac{2}{n} + 4 \left(\frac{1}{n} - \frac{1}{n^2}\right) \rho_j^2.
\] (67)

From equation (67), we then obtain
\[
\lim_{n \to \infty} \text{Var}\left[\sum_{i=0}^{n-1} \eta_j^2(i) \right] = \lim_{n \to \infty} E\left[\left(\frac{\sum_{i=0}^{n-1} \eta_j^2(i)}{n}\right)^2\right] = \lim_{n \to \infty} \frac{2}{n} + 4 \left(\frac{1}{n} - \frac{1}{n^2}\right) \rho_j^2 = 0.
\] (68)
Equation (68) leads to

$$\lim_{n \to \infty} \frac{\text{Var} \left[ \sum_{i=0}^{n-1} \eta_i^2(i) \right]}{n} = 2 + 4 \rho_j^2.$$  \hspace{1cm} (69)

Hence, for each $j \geq 1$, $K_{j,1}$ in equation (28) is given by

$$K_{j,1} = 2 + 4 \rho_j^2,$$ \hspace{1cm} (70)

and, from equation (67), $\text{Var} \left[ \sum_{i=0}^{n-1} \eta_i^2(i) \right] \leq 2 + 4 \left( \frac{1}{n} - \frac{1}{n^2} \right) \rho_j^2 \leq 2 + 4 \rho_j^2 \leq 6$. Thus, for every $j \geq 1$, $\tilde{K}_{j,1}$ in equation (30) satisfies

$$\tilde{K}_{j,1} \leq 6.$$ \hspace{1cm} (71)

**Remark 7** Note that, from Lemma 1, for each $j \geq 1$,

$$E \left[ \bar{\eta}_i^1(i) \right] = 3, \quad \forall i \in \mathbb{Z}.$$ \hspace{1cm} (72)

Thus, the assumption considered in Remark 2 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 pp. 192–193 of [33]).

For $n \geq 2$, and for each $j \geq 1$, we are now going to compute $K_{j,2}$ in (29). The $(n-1) \times 1$ random vectors $\eta_j^* = (\eta_j(0), \ldots, \eta_j(n-2))^T$ and $\eta_j^{**} = (\eta_j(1), \ldots, \eta_j(n-1))^T$ are Multivariate Normal distributed, with null mean vector, and covariance matrix

$$\tilde{\Sigma}_{(n-1) \times (n-1)} = \begin{pmatrix}
1 & \rho_j & 0 & \ldots & \ldots & 0 \\
\rho_j & 1 & \rho_j & 0 & \ldots & 0 \\
0 & \rho_j & 1 & \rho_j & \ldots & 0 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
0 & \ldots & 0 & 0 & \rho_j & 1 \\
\end{pmatrix}_{(n-1) \times (n-1)}. \hspace{1cm} (73)

From equation (20), for each $j \geq 1$,
\[ E \left[ \sum_{i=0}^{n-2} \eta_j(i)\eta_j(i+1) \right] = \sum_{i=0}^{n-2} \rho_j = (n-1)\rho_j = \text{tr} \left( E \left[ \eta_j^* \eta_j^{**} \right] \right), \quad (74) \]

where

\[ E \left[ \eta_j^* \eta_j^{**} \right] = E \left[ \eta_j^* \otimes \eta_j^{**} \right] = \rho_j I_{(n-1) \times (n-1)}, \quad (75) \]

with, as before, \( I_{(n-1) \times (n-1)} \) 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 \( \eta_j^* \) and \( \eta_j^{**} \). In the Gaussian case, keeping in mind that \( \eta_j^* = (\eta_j(0), \ldots, \eta_j(n-2))^T \) and \( \eta_j^{**} = (\eta_j(1), \ldots, \eta_j(n-1))^T \) are zero-mean multivariate Normal distributed vectors with covariance matrix \( \Sigma_{(n-1) \times (n-1)} \) given in (73), and having cross-covariance matrix (75), we can compute the variance of \( \sum_{i=0}^{n-2} \eta_j(i)\eta_j(i+1) \), from (74)-(75), as follows:

\[ \text{Var} \left[ \eta_j^* I_{(n-1) \times (n-1)} \eta_j^{**} \right] = E \left[ \eta_j^* I_{(n-1) \times (n-1)} \eta_j^{**} \eta_j^* I_{(n-1) \times (n-1)} \eta_j^{**} \right] \]

\[ - (E \left[ \eta_j^* I_{(n-1) \times (n-1)} \eta_j^{**} \right])^2 \]

\[ = \sum_{i=0}^{n-2} \sum_{p=0}^{n-2} E [\eta_j(i)\eta_j(i+1)\eta_j(p)\eta_j(p+1)] - (E \left[ \eta_j^* I_{(n-1) \times (n-1)} \eta_j^{**} \right])^2 \]

\[ = \sum_{i=0}^{n-2} E [\eta_j(i)\eta_j(i+1)] \sum_{p=0}^{n-2} E [\eta_j(p)\eta_j(p+1)] \]

\[ + \sum_{i=0}^{n-2} \sum_{p=0}^{n-2} E [\eta_j(i)\eta_j(p)] E [\eta_j(i+1)\eta_j(p+1)] \]

\[ + \sum_{i=0}^{n-2} \sum_{p=0}^{n-2} E [\eta_j(i)\eta_j(p+1)] E [\eta_j(i+1)\eta_j(p)] - (E \left[ \eta_j^* I_{(n-1) \times (n-1)} \eta_j^{**} \right])^2 \]

\[ = [\text{tr} \left( E[\eta_j^* \otimes \eta_j^{**}] \right)]^2 + \text{tr} \left( \Sigma_{(n-1) \times (n-1)} \Sigma_{(n-1) \times (n-1)} \right) \]

\[ + \text{tr} \left( E[\eta_j^* \otimes \eta_j^{**}] E[\eta_j^* \otimes \eta_j^{**}]^T \right) - [\text{tr} \left( E[\eta_j^* \otimes \eta_j^{**}] \right)]^2 \]

\[ = \text{tr} \left( \Sigma_{(n-1) \times (n-1)} \Sigma_{(n-1) \times (n-1)} \right) + \text{tr} \left( E[\eta_j^* \otimes \eta_j^{**}] E[\eta_j^* \otimes \eta_j^{**}] \right) \]

\[ = (n-1) + 2(n-2)\rho_j^2 + (n-1)\rho_j^2, \quad (76) \]

where, from (75),

\[ E[\eta_j^* \otimes \eta_j^{**}] E[\eta_j^* \otimes \eta_j^{**}]^T = \begin{pmatrix} \rho_j^2 & 0 & \ldots & 0 \\ 0 & \rho_j^2 & 0 & \ldots \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \ldots & \ldots & \rho_j^2 \end{pmatrix} = \rho_j^2 I_{(n-1) \times (n-1)}. \]

From (76),
Var \[ \sum_{i=0}^{n-2} \eta_j(i)\eta_j(i+1) \] \( \overline{n-1} \) 
\[ \frac{(n-1) + 2(n-2)\rho_j^2 + (n-1)\rho_j^2}{(n-1)^2} \] \( \text{(77)} \)

Therefore, for each \( j \geq 1 \),

\[ \lim_{n \to \infty} \text{Var} \left[ \sum_{i=0}^{n-2} \eta_j(i)\eta_j(i+1) \right] \overline{1/n} = 1 + 3\rho_j^2. \] \( \text{(78)} \)

Thus, for each \( j \geq 1 \), \( K_{j,2} \) in (29) is given by \( K_{j,2} = 1 + 3\rho_j^2 \). From equation (77), Var \[ \sum_{i=0}^{n-2} \eta_j(i)\eta_j(i+1) \] \( \overline{n-1} \) \( \leq 1 + 3\rho_j^2 \leq 4 \). Thus, for every \( j \geq 1 \), \( \tilde{K}_{j,2} \) in equation (31) satisfies

\[ \tilde{K}_{j,2} \leq 4. \] \( \text{(79)} \)

From equations (71) and (79), the constant \( S \) in Assumption A4 is such that \( S \leq 6 + 4 = 10 \).

5 Simulations.

A simulation study is undertaken to illustrate the behavior of the formulated componentwise estimator of the autocorrelation operator, and of its associated ARH(1) plug-in predictor for large sample sizes, in Section 5.1. By implementation of alternative ARH(1) plug-in prediction techniques, from the previous and current literature, a comparative study is also developed in Section 5.2. In the subsequent sections, we restrict our attention to the Gaussian ARH(1) framework.

5.1 Behavior of \( \hat{\rho} \) and \( \hat{X}_n \) for large sample sizes

In this subsection, the empirical functional mean-squared errors, associated with the proposed componentwise estimator of the autocorrelation operator, and associated ARH(1) plug-in predictor are evaluated from repeated generations of a sequence of samples with increasing, sufficiently large, sizes. In the simulation of that samples, from an ARH(1) process, operators \( \rho \), \( C \) and \( R_\varepsilon \) are defined within the family of fractional elliptic polynomials, with constant coefficients, of the Dirichlet negative Laplacian operator on an interval.

Let us first consider the Dirichlet negative Laplacian operator on an interval \((a, b), (-\Delta)_{(a,b)}\), given by
\(-\Delta_{(a,b)} (f)(x) = \frac{d^2}{dx^2} f(x), \quad x \in (a,b) \subset \mathbb{R},
\)
\[ f(a) = f(b) = 0. \quad \text{(80)} \]

The eigenvectors \((\phi_j, \ j \geq 1)\), and eigenvalues \((\lambda_j ((-\Delta)_{(a,b)}), \ j \geq 1)\) of \((-\Delta)_{(a,b)}\) satisfy, for each \(j \geq 1\),
\[ (-\Delta)_{(a,b)} \phi_j (x) = \lambda_j ((-\Delta)_{(a,b)}) \phi_j (x), \quad x \in (a,b),
\]
\[ \phi_j (a) = \phi_j (b) = 0. \quad \text{(81)} \]

For each \(j \geq 1\), the solution to equation (81) is given by (see [27], p. 6):
\[ \phi_j (x) = \frac{2}{b-a} \sin \left( \frac{\pi j x}{b-a} \right), \quad \forall x \in [a,b], \quad j \geq 1, \quad \text{(82)} \]
\[ \lambda_j ((-\Delta)_{(a,b)}) = \frac{\pi^2 j^2}{(b-a)^2}, \quad j \geq 1. \quad \text{(83)} \]

We consider here operator \(C\) defined as (see Remark 1)
\[ C = ((-\Delta)_{(a,b)})^{-\delta_1/2}, \quad \delta_1 > 1. \quad \text{(84)} \]

From [18], 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:
\[ C_j = [\lambda_j ((-\Delta)_{(a,b)})]^{-\delta_1/2} \left[ \frac{\pi^2 j^2}{(b-a)^2} \right]^{-\delta_2/2} = a_j^{-\delta_2}, \quad a = \left( \frac{\pi}{b-a} \right)^{-\delta_1}, \quad \delta_1 > 1. \quad \text{(85)} \]

Additionally, considering
\[ \rho = \left[ \frac{(\frac{\pi}{b-a})^{-\delta_1}}{\lambda_1 ((-\Delta)_{(a,b)}) - \epsilon} \right]^{-\delta_2/2}, \quad 1 < \delta_2 < 2, \quad \text{(86)} \]

for certain positive constant \(\epsilon < \lambda_1 ((-\Delta)_{(a,b)})\) close to zero, \(\rho\) is a positive self-adjoint Hilbert-Schmidt operator (with a non extremely fast decay rate), 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
\[ \rho_j^2 = \left( \frac{\lambda_j ((-\Delta)_{(a,b)})}{\lambda_1 ((-\Delta)_{(a,b)}) - \epsilon} \right)^{-\delta_2}, \quad \rho_j^2 \in (0,1), \quad 1 < \delta_2 < 2, \quad j \geq 1. \quad \text{(87)} \]
where, as before,

\[ 0 < \lambda_1 (-(\Delta)_{(a,b)}) \leq \lambda_2 (-(\Delta)_{(a,b)}) \leq \ldots \leq \lambda_j (-(\Delta)_{(a,b)}) \leq \ldots \]

From (19), the eigenvalues \( \sigma_j^2, \ j \geq 1 \) of \( R_\varepsilon \) are then defined, for each \( j \geq 1 \), as

\[ \sigma_j^2 = C_j (1 - \rho_j^2) = [\lambda_j (-(\Delta)_{(a,b)})]^{-\delta_1/2} - \frac{\lambda_j (-(\Delta)_{(a,b)}) - \varepsilon}{(\lambda_1 (-(\Delta)_{(a,b)}) - \varepsilon)^{-\delta_2}}. \]  

(88)

Note that \( R_\varepsilon \) is in the trace class, since the trace property of \( C \), and the fact that \( \rho_j < 1 \), 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

\[ k_n = n^{1/\alpha}, \]  

(89)

for a suitable \( \alpha > 0 \), which, in particular, allows verification of (27). From equation (85),

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

(90)

From equation (89), Assumption A3 is then satisfied if

\[ 1/2 - \frac{\delta_1}{\alpha} > 0, \quad \text{i.e., if} \quad \alpha > 2\delta_1 > 2, \quad \text{since} \quad \delta_1 > 1. \]  

(91)

Let us fix \( \delta_1 = 2.4 \) and \( \delta_2 = 1.1 \). Then, from equation (91), \( \alpha > 48/10 \). 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 \( (a,b) \).

Generations have been performed under Assumptions A1–A4, for our particular choice of \( C \) and selection of \( k_n \) in a diagonal scenario. The computed empirical truncated functional mean square error \( \text{EMSE}_{\hat{\rho}_{k_n}} \) of the estimator \( \hat{\rho}_{k_n} \) of \( \rho \), with \( n \) being, as before, the sample size, is given by:

\[ \text{EMSE}_{\hat{\rho}_{k_n}} = \frac{1}{N} \sum_{w=1}^{N} \sum_{j=1}^{k_n} (\rho_j - \hat{\rho}_{w,j})^2, \]  

(92)

\[ \hat{\rho}_{w,j} = \frac{\hat{D}_{n,j} w}{C_{n,j} w} = \frac{1}{n} \sum_{i=0}^{n-2} X_{i,j} X_{i+1,j}, \]  

(93)

\[ \frac{1}{n} \sum_{i=0}^{n-1} (X_{i,j})^2, \]  

23
where \( N \) denotes the number of simulations, and for each \( j = 1, \ldots, k_n \), \( \hat{\rho}_{n,j} \) represents the estimator of \( \rho_j \), based on the \( w^{th} \) generation of the values \( X_{0,j}^w, \ldots, X_{n-1,j}^w \), with \( X_{i,j}^w = \langle X_i^w, \phi_j \rangle_H \), for \( w = 1, \ldots, 700 \), and \( i = 0, \ldots, n - 1 \). For the plug-in predictor \( \hat{X}_n = \hat{\rho}_{n, (X_{n-1})} \), we compute the empirical version \( UB(EMA\hat{E}_{\hat{X}_n}) \) of the derived upper bound \((59)\), which, for each \( n \in \mathbb{Z} \), is given by

\[
UB(EMA\hat{E}_{\hat{X}_n}) = \sqrt{\frac{1}{N} \sum_{w=1}^{N} \left( \frac{1}{k_n} \sum_{j=1}^{k_n} \left( \rho_j - \hat{\rho}_{n,j}^w \right)^2 E \| X_{n-1}^w \|_H^2 \right) \hat{E}_{\hat{X}_n}} = \sqrt{\frac{1}{N} \sum_{w=1}^{N} \sum_{j=1}^{k_n} \left( \rho_j - \hat{\rho}_{n,j}^w \right)^2 \hat{\sigma}_X^2} = \sqrt{EMSE_{\hat{\rho}_{n,1}} \hat{\sigma}_X},
\]

where \( \hat{\sigma}_X^2 = \sum_{j=1}^{k_n} C_j \).

From \( N = 700 \) realizations, for each one of the values of the sample sizes \((n_t, t = 1, \ldots, 20) = (15000 + 20000(t-1), \ t = 1, \ldots, 20)\), the \( EMSE_{\hat{\rho}_{n_t}} \) and \( UB(EMA\hat{E}_{\hat{X}_{n_t}}) \) values, for \( \alpha = 5 \) and \( \alpha = 6 \), are displayed in Table 1, where the abbreviated notations \( MSE_{\hat{\rho}_{n_t,1}} \), for \( EMSE_{\hat{\rho}_{n_t}} \), and \( UB(EMA\hat{E}_{\hat{X}_{n_t,1}}) \), for \( UB(EMA\hat{E}_{\hat{X}_{n_t}}) \), are used (see also Figures 1-2).
| $n$  | $k_{n,1}$ | $MSE_{\hat{\rho}_{n,1}}$ | UB$_{X_{n,k_{n,1}}}$ | $k_{n,2}$ | $MSE_{\hat{\rho}_{n,2}}$ | UB$_{X_{n,k_{n,2}}}$ |
|------|-----------|-----------------|-----------------|-----------|-----------------|-----------------|
| $n_1 = 15000$ | 6 | $3.74 \times 10^{-4}$ | $2.87 \times 10^{-2}$ | 4 | $2.45 \times 10^{-4}$ | $2.25 \times 10^{-2}$ |
| $n_2 = 35000$ | 8 | $2.15 \times 10^{-4}$ | $2.21 \times 10^{-2}$ | 5 | $1.35 \times 10^{-4}$ | $1.71 \times 10^{-2}$ |
| $n_3 = 55000$ | 8 | $1.34 \times 10^{-4}$ | $1.75 \times 10^{-2}$ | 6 | $1.03 \times 10^{-4}$ | $1.51 \times 10^{-2}$ |
| $n_4 = 75000$ | 9 | $1.09 \times 10^{-4}$ | $1.57 \times 10^{-2}$ | 6 | $7.55 \times 10^{-5}$ | $1.29 \times 10^{-2}$ |
| $n_5 = 95000$ | 9 | $9.48 \times 10^{-5}$ | $1.47 \times 10^{-2}$ | 6 | $5.86 \times 10^{-5}$ | $1.14 \times 10^{-2}$ |
| $n_6 = 115000$ | 10 | $8.31 \times 10^{-5}$ | $1.39 \times 10^{-2}$ | 6 | $5.16 \times 10^{-5}$ | $1.07 \times 10^{-2}$ |
| $n_7 = 135000$ | 10 | $6.81 \times 10^{-5}$ | $1.25 \times 10^{-2}$ | 7 | $4.86 \times 10^{-5}$ | $1.04 \times 10^{-2}$ |
| $n_8 = 155000$ | 10 | $6.37 \times 10^{-5}$ | $1.21 \times 10^{-2}$ | 7 | $3.88 \times 10^{-5}$ | $9.66 \times 10^{-3}$ |
| $n_9 = 175000$ | 11 | $6.14 \times 10^{-5}$ | $1.19 \times 10^{-2}$ | 7 | $3.87 \times 10^{-5}$ | $9.65 \times 10^{-3}$ |
| $n_{10} = 195000$ | 11 | $5.34 \times 10^{-5}$ | $1.11 \times 10^{-2}$ | 7 | $3.42 \times 10^{-5}$ | $8.79 \times 10^{-3}$ |
| $n_{11} = 215000$ | 11 | $4.67 \times 10^{-5}$ | $1.03 \times 10^{-2}$ | 7 | $3.40 \times 10^{-5}$ | $8.74 \times 10^{-3}$ |
| $n_{12} = 235000$ | 11 | $4.66 \times 10^{-5}$ | $1.03 \times 10^{-2}$ | 7 | $2.92 \times 10^{-5}$ | $8.12 \times 10^{-3}$ |
| $n_{13} = 255000$ | 12 | $4.53 \times 10^{-5}$ | $1.02 \times 10^{-2}$ | 7 | $2.77 \times 10^{-5}$ | $7.95 \times 10^{-3}$ |
| $n_{14} = 275000$ | 12 | $4.24 \times 10^{-5}$ | $9.95 \times 10^{-3}$ | 8 | $2.77 \times 10^{-5}$ | $7.94 \times 10^{-3}$ |
| $n_{15} = 295000$ | 12 | $3.72 \times 10^{-5}$ | $9.32 \times 10^{-3}$ | 8 | $2.67 \times 10^{-5}$ | $7.76 \times 10^{-3}$ |
| $n_{16} = 315000$ | 12 | $3.62 \times 10^{-5}$ | $9.21 \times 10^{-3}$ | 8 | $2.55 \times 10^{-5}$ | $7.64 \times 10^{-3}$ |
| $n_{17} = 335000$ | 12 | $3.39 \times 10^{-5}$ | $8.91 \times 10^{-3}$ | 8 | $2.28 \times 10^{-5}$ | $7.04 \times 10^{-3}$ |
| $n_{18} = 355000$ | 12 | $3.34 \times 10^{-5}$ | $8.86 \times 10^{-3}$ | 8 | $2.20 \times 10^{-5}$ | $7.04 \times 10^{-3}$ |
| $n_{19} = 375000$ | 13 | $3.34 \times 10^{-5}$ | $8.86 \times 10^{-3}$ | 8 | $2.04 \times 10^{-5}$ | $6.84 \times 10^{-3}$ |
| $n_{20} = 395000$ | 13 | $3.12 \times 10^{-5}$ | $8.56 \times 10^{-3}$ | 8 | $1.92 \times 10^{-5}$ | $6.65 \times 10^{-3}$ |

Table 1: $EMSE_{\hat{\rho}_{n}}$ (here, $MSE_{\hat{\rho}_{n,1}}$), and UB(EMAE$_{X_{n}}$) (here, UB$_{X_{n,k_{n,1}}}$) values, in (92)-(94), based on $N = 700$ simulations, for $\delta_1 = 2.4$ and $\delta_2 = 1.1$, considering the sample sizes $n_t = 15000 + 20000(t-1)$, $t = 1, \ldots, 20$, and the corresponding $k_{n,1}$ and $k_{n,2}$ values, for $\alpha_1 = 5$ and $\alpha_2 = 6$. 

25
In the example considered, a one-parameter model of \( k_n \) is selected depending on parameter \( \alpha \). In Example 2, in p. 286 in [31], with 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 \( L(H) \) 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 \( \hat{\rho}_{k_n} \) presents a velocity 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 [31], in the case of known eigenvectors of \( C \) (see, in particular, Theorem 1, Remark 2 and Example 2 in [31]). For larger values of the parameters \( \delta_1 \) than 2.4, and \( \alpha \) than 6, a faster velocity of convergence of \( \hat{\rho}_{k_n} \) 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 [4, 9, 11] and [31] can be found in the next subsection.

5.2 A comparative study

In this section, the performance of our approach is compared with those ones given in [4, 9, 11, 31], including the case of unknown eigenvectors of \( C \). In the last case, our approach and the approaches presented in [11] and [31] are implemented in terms of the empirical eigenvectors (see section 5.2.2).

5.2.1 Theoretical-eigenvector-based componentwise estimators

Let us first compare the performance of our ARH(1) plug-in predictor, defined in (53), and the ones formulated in [11] and [31], 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 (85) and (87), 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 A1, B1 and C1, formulated in [11], pp. 211–212, are satisfied by the generated ARH(1) process. Similarly, Conditions H1–H3 in p. 283 of [31] are satisfied by our generated Gaussian ARH(1) process. The estimators of \( \rho \) from [11] and [31] we are going to test here are given as follows. In Section 8.2 of [11] the following estimator of \( \rho \) is proposed

\[
\hat{\rho}_n(x) = \left( \Pi^{k_n} D_n \hat{C}_n^{-1} \Pi^{k_n} \right)(x) = \sum_{l=1}^{k_n} \hat{\rho}_{n,l}(x) \phi_l, \quad x \in H, \tag{95}
\]

in the finite dimensional subspace \( H_{k_n} = sp(\phi_1, \ldots, \phi_{k_n}) \) 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 [31], Section 2, given by

\[
\hat{\rho}_{n,a}(x) = \left( \Pi^{k_n} D_n \hat{C}_{n,a}^{-1} \Pi^{k_n} \right)(x) = \sum_{l=1}^{k_n} \hat{\rho}_{n,a,l}(x) \phi_l, \quad x \in H, \tag{97}
\]

\[
\hat{\rho}_{n,a,l}(x) = \frac{1}{n-1} \sum_{i=0}^{n-2} \sum_{j=1}^{k_n} \frac{1}{\hat{C}_{n,j}} \langle \phi_j, x \rangle_X X_{i,j} X_{i+1,l}, \tag{98}
\]

where \( \hat{C}_{n,a}^{-1} = \sum_{j=1}^{k_n} \frac{1}{\max(\hat{C}_{n,j}, a_n)} \langle \phi_j, x \rangle_{H} \phi_j \) (a.s.). Here, \( (a_n, \ n \in \mathbb{N}) \) is such that (see Theorem 1
in [31])

\[
\alpha C_{k_n}^{\gamma} \leq a_n \leq \beta \lambda_{k_n}, \quad \alpha > 0, \ 0 < \beta < 1, \ \varepsilon < 1/2, \ \gamma \geq 1. \tag{99}
\]

Tables 2-3 below display results for \( \hat{\rho}_{k_n} = \hat{\rho}_{k_n} \), where \( \hat{\rho}_{k_n} \) is given in equations (25)-(26) (see third column), \( \hat{\rho}_{n} = \hat{\rho}_{n} \), where \( \hat{\rho}_{n} \) is given in equations (95)-(96) (see fourth column), and \( \hat{\rho}_{k_n} = \hat{\rho}_{n,a} \), where \( \hat{\rho}_{n,a} \) is defined in (97)-(98) (see fifth column). Specifically, Tables 2-3 below provide the truncated, for two different \( k_n \) rules, empirical values of 

\( E \| \rho (X_{n-1}) - \hat{\rho}_{k_n}(X_{n-1}) \|_H \), based on \( N = 700 \) generations of each one of the functional samples considered with size \( n_t = 15000 + 20000(t - 1), \ t = 1, \ldots, 20 \), when \( C_j \) and \( \rho_j^2, \ j \geq 1 \), are defined as done above in equations (85) and (87). Specifically, in Table 2, the following parameter values have been considered: \( \delta_1 = 2.4 \ \delta_2 = 1.1 \), and \( k_n = n^{1/\alpha} \), for \( \alpha = 6 \), according to our Assumption A3, (which is also considered in p. 217 of [11], to ensure weak consistency of the proposed estimator of \( \rho \)). While, 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 Example 2, in p. 286 in [31]. Thus, in Table 3,

\[
k_n = n^{\frac{1+\gamma}{2(\varepsilon+2)}}, \quad \gamma \geq 1, \ \varepsilon < 1/2. \tag{100}
\]

In particular we have chosen \( \gamma = 2 \), and \( \varepsilon = 0.04 \delta_1 \). Note that, from Theorem 1 and Remark 1 in [31], for the choice made of \( k_n \) in Table 3, convergence to \( \rho \), in quadratic mean in the space \( L(H) \), holds for \( \hat{\rho}_{n,a} \) given in (98).
Our Approach
Bosq (2000)
Guillas (2001)

| 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 2: Truncated empirical values of \( E[\|\rho(X_{n-1}) - \hat{\rho}_k(X_{n-1})\|_H, \text{for } \hat{\rho}_k \text{ given in equations } (25)-(26) \text{ (third column), in equations } (95)-(96) \text{ (fourth column), and in equations } (97)-(98) \text{ (fifth column), based on } N = 700 \text{ simulations, for } \delta_1 = 2.4 \text{ and } \delta_2 = 1.1, \text{ considering the sample sizes } n_t = 15000 + 20000(t - 1), \text{ for } t = 1, \ldots, 20, \text{ and the corresponding } k_n = n^{1/\alpha} \text{ values, for } \alpha = 6 \text{ (from Assumption A3), assuming that } (\phi_j, j \geq 1) \text{ are known.} \)
Our Approach

| $n$ | $k_n$ | 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 | $3.86 (10)^{-3}$ | $5.29 (10)^{-3}$ | $5.26 (10)^{-3}$ |

Table 3: 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 (25)–(26) (third column), in equations (95)–(96) (fourth column), and in equations (97)–(98) (fifth column), based on $N = 700$ simulations, for $\delta_1 = \frac{64}{200}$ and $\delta_2 = 1.1$, considering the sample sizes $n_t = 15000 + 20000(t - 1)$, $t = 1, \ldots, 20$, and the corresponding $k_n$ given in (100). It can be observed in Table 2 a similar performance of the three methods compared with the truncation order $k_n$ satisfying our Assumption A3, with slightly worse results being obtained from the estimator defined in equations (97)–(98), 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 $n_4 = 75000$). While, for the remaining largest sample sizes, slightly differences are observed, with a better performance of our approach, very close to the other two approaches presented in [11] and [31].
5.2.2 Empirical-eigenvector-based componentwise estimators

In this section, we address the case where \((\phi_j, \ j \geq 1)\) are unknown, as usually it occurs in real-data problems. Specifically, an empirical version of the estimators of \(\rho\) formulated in equations (25)–(26), (95)–(96) and (97)–(98), consisting of replacing the theoretical eigenvectors \((\phi_j, \ j \geq 1)\) by their empirical (random) counterparts \((\phi_{n,j}, \ j \geq 1)\), the eigenvectors of \(C_n = \frac{1}{n} \sum_{i=0}^{n-1} X_i \otimes X_i, \ n \geq 1\), is now studied. Specifically, taking \((C_{n,j}, \ j \geq 1)\) as the eigenvalues of \(C_n\), we consider

\[
\tilde{\rho}_{n,j} = \frac{\frac{1}{n-1} \sum_{i=0}^{n-2} \tilde{X}_{i,j} \tilde{X}_{i+1,j}}{\frac{1}{n} \sum_{i=0}^{n-1} \left(\tilde{X}_{i,j}\right)^2}, \quad \tilde{\rho}_{k,n} = \sum_{j=1}^{k_n} \tilde{\rho}_{n,j} \phi_{n,j} \otimes \phi_{n,j}, \quad (101)
\]

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

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

\[
\tilde{\rho}_{n,a,l}(x) = \frac{1}{n-1} \sum_{i=0}^{n-2} \sum_{j=1}^{k_n} \frac{1}{\max(C_{n,j}, a_n)} \langle \phi_{n,j}, x \rangle_H \tilde{X}_{i,j} \tilde{X}_{i+1,l}, \quad (102)
\]

where, for \(i \in \mathbb{Z}\), and \(j \geq 1\), \(\tilde{X}_{i,j} = \langle X_i, \phi_{n,j} \rangle_H\), \(\tilde{H}^{k_n}\) denotes the orthogonal projector into the space \(\tilde{H}_{k_n} = sp(\phi_{n,1}, \ldots, \phi_{n,k_n})\). The Gaussian ARH(1) process is generated under Assumptions A1–A2, as well as \(C'_1\) in p. 218 in [11]. Note that conditions A1 and \(B'_1\) in [11] already hold. Moreover, as given in Theorem 8.8 and Example 8.6, in p. 221 in [11], for \(C_j\), \(\rho^2_j\), \(j \geq 1\), defined as above in equations (85) and (87), with, in particular, \(\delta_1 = 2.4\), and \(\delta_2 = 1.1\), the estimator \(\tilde{\rho}_n\) converges almost surely to \(\rho\) under the condition

\[
\frac{n C_{kn}^2}{\log(n) \left(\sum_{j=1}^{k_n} b_j\right)^2} \rightarrow \infty,
\]

where \(b_1 = 2\sqrt{2} (C_1 - C_2)^{-1}\) and \(b_j = 2\sqrt{2} \max \left[ (C_{j-1} - C_j)^{-1}, (C_j - C_{j+1})^{-1}\right]\), for \(j \geq 2\).

Particularly, in Table 4 the truncation criteria from Assumption A3 is used, and in Table 4, \(k_n \simeq \log(n)\) has been tested (see Example 8.6, in [11], p. 221).
A better performance of our estimator (101) in comparison with estimator (102), formulated in [11], and estimator (103), formulated in [31] (see Example 4 and Remark 4, in p. 291 in [31]) is observed below in Table 4. Note that, in particular, in Example 4 and Remark 4, in [31], p. 291, smaller values of \( k_n \) than \( \log(n) \) 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, applying truncation at \( k_n = n^{1/6} \), (i.e., \( \alpha = 6 \)), we obtain (see 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. In particular, it can be observed in Table 5 below, a similar performance of the three

| \( n \)  | \( k_n \) | Our approach | Bosq (2000) | Guillas (2001) |
|--------|----------|--------------|-------------|----------------|
| \( n_1 = 15000 \) | 9        | 8.42 (10)^{-2} | 1.0614      | 1.0353         |
| \( n_2 = 35000 \) | 10       | 5.51 (10)^{-2} | 1.0186      | 1.0052         |
| \( n_3 = 55000 \) | 10       | 4.75 (10)^{-2} | 1.0174      | 0.9986         |
| \( n_4 = 75000 \) | 11       | 4.43 (10)^{-2} | 1.0153      | 0.9951         |
| \( n_5 = 95000 \) | 11       | 3.68 (10)^{-2} | 1.0127      | 0.9883         |
| \( n_6 = 115000 \) | 11      | 3.51 (10)^{-2} | 1.0113      | 0.9627         |
| \( n_7 = 135000 \) | 11      | 3.23 (10)^{-2} | 1.0081      | 0.9247         |
| \( n_8 = 155000 \) | 11      | 2.95 (10)^{-2} | 1.0066      | 0.9119         |
| \( n_9 = 175000 \) | 12      | 2.94 (10)^{-2} | 1.0057      | 0.9113         |
| \( n_{10} = 195000 \) | 12   | 2.80 (10)^{-2} | 0.9948      | 0.8912         |
| \( n_{11} = 215000 \) | 12   | 2.71 (10)^{-2} | 0.9017      | 0.8615         |
| \( n_{12} = 235000 \) | 12   | 2.59 (10)^{-2} | 0.8896      | 0.8201         |
| \( n_{13} = 255000 \) | 12   | 2.58 (10)^{-2} | 0.8783      | 0.8004         |
| \( n_{14} = 275000 \) | 12   | 2.35 (10)^{-2} | 0.8719      | 0.7832         |
| \( n_{15} = 295000 \) | 12   | 2.28 (10)^{-2} | 0.8602      | 0.7780         |
| \( n_{16} = 315000 \) | 12   | 2.27 (10)^{-2} | 0.8424      | 0.7469         |
| \( n_{17} = 335000 \) | 12   | 2.16 (10)^{-2} | 0.8217      | 0.7140         |
| \( n_{18} = 355000 \) | 12   | 2.14 (10)^{-2} | 0.8001      | 0.7066         |
| \( n_{19} = 375000 \) | 12   | 2.09 (10)^{-2} | 0.7778      | 0.6872         |
| \( n_{20} = 395000 \) | 12   | 2.06 (10)^{-2} | 0.7693      | 0.6621         |

Table 4: Truncated empirical values of \( E \| \rho(X_{n-1}) - \tilde{\rho}_{h_t}(X_{n-1}) \|_{H_t} \) for \( \tilde{\rho}_k = \rho_{h_t} \), given in equation (101) (third column), \( \rho_{h_t} = \tilde{\rho}_n \) defined in equation (102) (fourth column) and \( \rho_{n,a} = \tilde{\rho}_n \) defined in equation (103) (fifth column), based on \( N = 700 \) simulations, for \( \delta_1 = 2.4 \) and \( \delta_2 = 1.1 \), considering the sample sizes \( n_t = 15000 + 20000(t - 1), t = 1, \ldots, 20 \), and \( k_n = \log(n) \). The time discretization step is assumed as \( h_t = 0.08 \).
approaches studied, with a worse performance of the approach presented in [31], for \( n_1 = 15000 \), since a too large number of coefficients of \( \rho \) to be estimated is required, for the smallest sample size. To illustrate this fact, in Table 6, the value \( k_n = \lceil e' n^{1/(8\delta_1^2)} \rceil \), \( e' = \frac{4}{99} \) proposed, in Example 4 and Remark 4, in [31], p. 291, 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 (101) (third column), for \( \tilde{\rho}_{k_n} = \tilde{\rho}_n \) given in equation (102) (fourth column), and for \( \tilde{\rho}_{k_n} = \tilde{\rho}_{n,a} \) in equation (103) (fifth column). With this truncation rule, requiring a smaller number of coefficients of \( \rho \) to be estimated for a given sample size, a better performance of the estimator proposed in [31] is obtained for the largest sample size tested \( n_{20} = 395000 \), with a similar performance of the three approaches tested, for the rest of sample sizes considered, with such a truncation rule.

| \( n \) | \( k_n \) | Our approach | Bosq (2000) | Guillas (2001) |
|---|---|---|---|---|
| 15000 | 4 | 9.88 (10)^{-2} | 9.25 (10)^{-2} | 0.1059 |
| 35000 | 5 | 9.52 (10)^{-2} | 9.07 (10)^{-2} | 0.986 (10)^{-2} |
| 55000 | 6 | 9.12 (10)^{-2} | 8.92 (10)^{-2} | 0.939 (10)^{-2} |
| 75000 | 6 | 8.48 (10)^{-2} | 8.64 (10)^{-2} | 0.98 (10)^{-2} |
| 95000 | 6 | 7.61 (10)^{-2} | 8.30 (10)^{-2} | 0.846 (10)^{-2} |
| 115000 | 6 | 7.05 (10)^{-2} | 7.96 (10)^{-2} | 0.804 (10)^{-2} |
| 135000 | 7 | 6.99 (10)^{-2} | 7.84 (10)^{-2} | 0.782 (10)^{-2} |
| 155000 | 7 | 6.70 (10)^{-2} | 7.45 (10)^{-2} | 0.740 (10)^{-2} |
| 175000 | 7 | 6.49 (10)^{-2} | 7.03 (10)^{-2} | 0.707 (10)^{-2} |
| 195000 | 7 | 5.88 (10)^{-2} | 6.74 (10)^{-2} | 0.680 (10)^{-2} |
| 215000 | 7 | 5.63 (10)^{-2} | 6.46 (10)^{-2} | 0.657 (10)^{-2} |
| 235000 | 7 | 5.30 (10)^{-2} | 6.28 (10)^{-2} | 0.637 (10)^{-2} |
| 255000 | 7 | 5.05 (10)^{-2} | 6.19 (10)^{-2} | 0.624 (10)^{-2} |
| 275000 | 8 | 4.88 (10)^{-2} | 5.99 (10)^{-2} | 0.615 (10)^{-2} |
| 295000 | 8 | 4.58 (10)^{-2} | 5.74 (10)^{-2} | 0.604 (10)^{-2} |
| 315000 | 8 | 4.24 (10)^{-2} | 5.52 (10)^{-2} | 0.593 (10)^{-2} |
| 335000 | 8 | 3.86 (10)^{-2} | 5.24 (10)^{-2} | 0.570 (10)^{-2} |
| 355000 | 8 | 3.70 (10)^{-2} | 5.02 (10)^{-2} | 0.553 (10)^{-2} |
| 375000 | 8 | 3.55 (10)^{-2} | 4.88 (10)^{-2} | 0.536 (10)^{-2} |
| 395000 | 8 | 3.46 (10)^{-2} | 4.70 (10)^{-2} | 0.523 (10)^{-2} |

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 (101) (third column), for \( \tilde{\rho}_{k_n} = \tilde{\rho}_n \) given in equation (102) (fourth column), and for \( \tilde{\rho}_{k_n} = \tilde{\rho}_{n,a} \) in equation (103) (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 \( n_t = 15000 + 20000(t-1) \), \( t = 1, \ldots, 20 \), and \( k_n = n^{1/6} \). The time discretization step size is now \( h_2 = 0.015 \).
Our approach considers the sample sizes $n_i$ and the empirical values of $E\| \rho(\hat{X}_{n-1}) - \tilde{\rho}_{k_n}(X_{n-1}) \|_H$, for $\tilde{\rho}_{k_n}$ defined in equation (101) (third column), for $\tilde{\rho}_{k_n} = \tilde{\rho}_n$ given in equation (102) (fourth column), and for $\tilde{\rho}_{k_n} = \tilde{\rho}_{n,a}$ in equation (103) (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 $n_i = 15000 + 20000(t - 1)$, $t = 1, \ldots, 20$, and $k_n = \left\lceil c' n^{1/(6\delta + 2)} \right\rceil$, $c' = \frac{15}{10}$. The time discretization step size is now $h_t = 0.015$.

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

Table 6: Truncated empirical values of $E\| \rho(\hat{X}_{n-1}) - \tilde{\rho}_{k_n}(X_{n-1}) \|_H$, for $\tilde{\rho}_{k_n}$ defined in equation (101) (third column), for $\tilde{\rho}_{k_n} = \tilde{\rho}_n$ given in equation (102) (fourth column), and for $\tilde{\rho}_{k_n} = \tilde{\rho}_{n,a}$ in equation (103) (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 $n_i = 15000 + 20000(t - 1)$, $t = 1, \ldots, 20$, and $k_n = \left\lceil c' n^{1/(6\delta + 2)} \right\rceil$, $c' = \frac{15}{10}$. The time discretization step size is now $h_t = 0.015$.

5.2.3 Kernel-based non-parametric and penalized estimation

In practice, curves are observed in discrete times, and should be approximated by smooth functions. In [9], the following optimization problem is considered:

$$\hat{X}_i = \arg\min_{\hat{X}_i} \left\| L\hat{X}_i \right\|_{L_2}, \quad \hat{X}_i(t_j) = X_i(t_j), \quad j = 1, \ldots, p, \quad i = 0, \ldots, n - 1,$$

(104)

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 (104), as follows:

\[
\hat{X}^h_n = \rho_n(X_{n-1}), \quad \rho_n(x) = \frac{\sum_{i=0}^{n-2} \hat{X}_{i+1} K \left( \frac{\|\hat{X}_i - x\|_2}{h_n} \right)}{\sum_{i=0}^{n-2} K \left( \frac{\|\hat{X}_i - x\|_2}{h_n} \right)},
\]

(105)

where \(K\) is the usual Gaussian kernel, and

\[
\|\hat{X}_i - x\|_2^2 = \int (\hat{X}_i(t) - x(t))^2 dt, \quad i = 0, \ldots, n - 2.
\]

Alternatively, in [9], 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 \(\hat{X}_1, \ldots, \hat{X}_n\), solving the optimization problem

\[
\min_{\hat{X}_i \in H_q} \frac{1}{n} \sum_{i=0}^{n-1} \left( \frac{1}{p} \sum_{j=1}^{p} \left( X_i(t_j) - \hat{X}_i^{q,l}(t_j) \right)^2 + l \left\| D^2 \hat{X}_i^{q,l} \right\|_2^2 \right),
\]

(106)

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 constrain is considered in the computation of the solution to the optimization problem (106). 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 \(H_q\):

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

Thus, the regularized estimator of \(\rho\) is given by

\[
\hat{\rho}_{q,l} = \hat{D}_{q,l} \hat{C}_{q,l}^{-1},
\]

and the predictor

\[
\hat{X}_n^{q,l} = \hat{\rho}_{q,l} X_{n-1}.
\]

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

\[
EMAE_{X_n}^{h_n} = \frac{1}{p} \sum_{j=1}^{p} \left( X_n(t_j) - \hat{X}_n^h(t_j) \right)^2, \quad (107)
\]
\[
EMAE_{\hat{X}_n}^{q,l} = \frac{1}{p} \sum_{j=1}^{p} \left( X_n(t_j) - \hat{X}_n^{q,l}(t_j) \right)^2 .
\] (108)

| \(n\) | \(EMAE_{\hat{X}_n}^{h_{n,1}}\) | \(EMAE_{\hat{X}_n}^{h_{n,2}}\) | \(EMAE_{\hat{X}_n}^{q,l}\) |
|-----|----------------|----------------|----------------|
| \(n_1 = 750\) | 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} |

Table 7: \(EMAE_{\hat{X}_n}^{h_{n,i}}\), \(i = 1, 2\), and \(EMAE_{\hat{X}_n}^{q,l}\) values (see (107) and (108), respectively), with \(q = 7\), based on \(N = 200\) simulations, for \(\delta_1 = 2.4\) and \(\delta_2 = 1.1\), considering now the sample sizes \(n_i = 750 + 500(t - 1), t = 1, \ldots, 13, h_{n,1} = 0.1\) and \(h_{n,2} = 0.3\)

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 [4] 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 \(\phi_{Jk}, k = 0, \ldots, 2^J - 1\), 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 \(i = 0, \ldots, n - 1\):

\[
\Phi_{V_J}X_i = \sum_{k=0}^{2^{j_0}-1} c_{j_0k}^i \phi_{j_0k} + \sum_{j=j_0}^{J-1} \sum_{k=0}^{2^j-1} d_{jk}^i \psi_{jk},
\]

\[
c_{j_0k}^i = \langle \Phi_{V_J}X_i, \phi_{j_0k} \rangle_H, \quad d_{jk}^i = \langle \Phi_{V_J}X_i, \psi_{jk} \rangle_H.
\] (109)
For \( i = 0, \ldots, n - 1 \), the following variational problem is solved to obtain the smooth estimate of the curve \( X_i \):

\[
\inf_{f^i \in H} \left\{ \| \Phi V_j X_i - f^i \|_{L^2}^2 + \lambda \left\| \Phi V_{i,0} f^i \right\|_2^2 : f \in H \right\},
\]

(110)

where \( \Phi V_{i,0} \) denotes the orthogonal projection operator of \( H \) onto the orthogonal complement of \( V_{j,0} \), and for \( i = 0, 1 \ldots n - 1 \),

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

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 (110) is equivalent to the optimization problem, for \( i = 0, \ldots, n - 1 \),

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

(111)

The solution to (111) is given by, for \( i = 0, \ldots, n - 1 \),

\[
\begin{align*}
\alpha_{j_0 k}^i &= c_{j_0 k}^i, \quad k = 0, 1, \ldots, 2^{j_0} - 1, \\
\beta_{j_0 k}^i &= \frac{d_{j k}^i}{(1 + \lambda 2^{j s})}, \quad j = j_0, \ldots, J - 1, \quad k = 0, 1, \ldots, 2^j - 1, \\
\beta_{j_0 k}^i &= 0, \quad j \geq J, \quad k = 0, 1, \ldots, 2^j - 1.
\end{align*}
\]

(112) (113) (114)

In particular, in the subsequent computations, we have considered the following value of the smoothing parameter \( \lambda \) (see [3]):

\[
\hat{\lambda}^M = \left( \frac{\sum_{j=1}^{M} C_j}{n} \right)^{\frac{1}{3}},
\]

The following smoothed data are then computed

\[
\tilde{X}_{i,\hat{\lambda}^M} = \sum_{k=0}^{2^{j_0}-1} \alpha_{j_0 k}^i \phi_{j_0 k} + \sum_{j=j_0}^{\infty} \sum_{k=0}^{2^j-1} \beta_{j k}^i \psi_{j k},
\]

(115)

removing the trend \( \tilde{a}_{n,\hat{\lambda}^M} = \frac{1}{n} \sum_{i=0}^{n-1} \tilde{X}_{i,\hat{\lambda}^M} \) to obtain

\[
\tilde{Y}_{i,\hat{\lambda}^M} = \tilde{X}_{i,\hat{\lambda}^M} - \tilde{a}_{n,\hat{\lambda}^M}, \quad i = 0, \ldots, n - 1,
\]

37
for the computation of

$$\tilde{\rho}_{n,\lambda M}(x) = \left(\tilde{H}_n^{\lambda M}, \tilde{D}_{n,\lambda M} \tilde{C}_{n,\lambda M}^{-1} \tilde{H}_n^{\lambda M}\right)(x) = \sum_{l=1}^{k_n} \tilde{\rho}_{n,\lambda M,l}(x) \tilde{\phi}_l^M, \ x \in H,$$

$$\tilde{\rho}_{n,\lambda M,l}(x) = \frac{1}{n} \sum_{j=1}^{k_n} \frac{1}{n-1} \sum_{i=0}^{n-2} \frac{1}{C_{n,\lambda M,l}} \tilde{\phi}_i^M \langle x \rangle_H \tilde{Y}_{i,\lambda M,j} \tilde{Y}_{i+1,\lambda M,l},$$

$$\tilde{C}_{n,\lambda M} = \frac{1}{n} \sum_{i=0}^{n-1} \tilde{Y}_{i,\lambda M} \otimes \tilde{Y}_{i,\lambda M}.$$ (116)

$$\tilde{C}_{n,\lambda M,j} = \langle \tilde{Y}_{i,\lambda M}, \tilde{\phi}_j \rangle. (117)$$

where \(\tilde{Y}_{i,\lambda M,j} = \langle \tilde{Y}_{i,\lambda M}, \tilde{\phi}_j \rangle\), and \(\tilde{C}_{n,\lambda M,j} = \langle \tilde{C}_{n,\lambda M} \tilde{\phi}_j \tilde{\mu} \rangle\). Table 8 displays the empirical truncated approximation of \(E\|\tilde{\rho}_{n,\lambda M}(X_n-1) - \rho(X_n-1)\|_H\), applying our approach and the above-described approach presented in [4], for \(k_n = n^{1/\alpha_i}, \ i = 1, 2\), with \(\alpha_1 = 6\), according to our Assumption A3, and \(\alpha_2 > 4\delta_1\), according to \(H_4: nC_n^4 \rightarrow \infty\) in p. 149 in [4]. In particular, we have considered \(\delta_1 = 2.4\), and \(\alpha_2 = 10\).

From the results displayed in Table 8, it can be observed a similar performance for the two truncation rules implemented, and approaches compared, for the small sample sizes tested. A similar performance is also displayed by the approaches presented in [9], for such small sample sizes (see Table 7).

| \(n\)   | \(k_{n,1}\) | O.A.     | \(k_{n,2}\) | O.A.     |
|-------|-----------|---------|-----------|---------|
| \(n_1\) = 750 | 3         | 0.0702  | 1         | 0.0636  |
| \(n_2\) = 1250 | 3         | 0.0550  | 2         | 0.0509  |
| \(n_3\) = 1750 | 3         | 0.0473  | 2         | 0.0455  |
| \(n_4\) = 2250 | 3         | 0.0414  | 2         | 0.0409  |
| \(n_5\) = 2750 | 3         | 0.0365  | 2         | 0.0355  |
| \(n_6\) = 3250 | 3         | 0.0343  | 2         | 0.0333  |
| \(n_7\) = 3750 | 3         | 0.0330  | 2         | 0.0325  |
| \(n_8\) = 4250 | 4         | 0.0328  | 2         | 0.0313  |
| \(n_9\) = 4750 | 4         | 0.0317  | 2         | 0.0309  |
| \(n_{10}\) = 5250 | 4         | 0.0309  | 2         | 0.0276  |
| \(n_{11}\) = 5750 | 4         | 0.0298  | 2         | 0.0203  |
| \(n_{12}\) = 6250 | 4         | 0.0283  | 2         | 0.0166  |
| \(n_{13}\) = 6750 | 4         | 0.0276  | 2         | 0.0148  |

Table 8: Truncated empirical values of \(E\|\rho(X_n-1) - \tilde{\rho}_{n,\lambda M}(X_n-1)\|_H\), with \(\tilde{\rho}_{n,\lambda M}\) defined in equation (101), and of \(E\|\tilde{\rho}_{n,\lambda M}(X_n-1) - \rho(X_n-1)\|_H\), based on \(N = 200\) simulations, for \(\delta_1 = 2.4\) and \(\delta_2 = 1.1\), considering the sample sizes \(n_t = 750 + 500(t-1), t = 1, \ldots, 13\), using \(\lambda_M, M = 50\), and the corresponding \(k_{n,1} = n^{1/\alpha_i}\), for \(\alpha_1 = 6\) and \(\alpha_2 = 10\). Here, O.A. means Our Approach and [4] means The approach presented in [4].
5.2.5 Non-diagonal autocorrelation operator

The methodology proposed can still be applied in the case of a non-diagonal autocorrelation operator $\rho$, with respect to the autocovariance eigenvector system $(C_j, j \geq 1)$, with coefficients close to zero outside of the main diagonal. This fact is illustrated below in Table 9, where 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

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

(118)

$$
\rho^2_{j,j+a} = \frac{0.01}{5a^2}, \ a = 1, 2, 3, 4, 5, \quad \rho^2_{j+1,j} = \frac{0.02}{5a^2}, \ a = 1, 2, 3, 4, 5,
$$

(119)

where $\rho^2_{j,h} = 0$ when $h \neq j - 1, j, j + 1$, and $\rho^2_{j,j+a} = \rho^2_{j+1,j} = 0$ when $a \geq 6$. The coefficients of the auto-covariance operator $R$ of the innovation process $\varepsilon$, with respect to the mentioned basis $(\phi_j \otimes \phi_h, j,h \geq 1)$, are defined as follows:

$$
\sigma^2_{j,j} = C_j \left( 1 - \rho^2_{j,j} \right),
$$

(120)

$$
\sigma^2_{j,j+a} = \frac{0.015}{5a^2}, \ a = 1, 2, 3, 4, 5, \quad \sigma^2_{j+1,j} = \frac{0.02}{5a^2}, \ a = 1, 2, 3, 4, 5,
$$

(121)

where $\sigma^2_{j,h} = 0$ when $h \neq j - 1, j, j + 1$, and $\sigma^2_{j,j+a} = \sigma^2_{j+1,j} = 0$ when $a \geq 6$.

In Table 9, we have considered $k_n = \lceil n^{1/\alpha} \rceil$, with $\alpha = 6$, $\delta_1 = 2.4$, $\delta_2 = 1.1$ and time discretization step $h_t = 0.015$. As expected, it can be observed a better performance of the approaches presented in [11] and [31] against our approach, in terms of the theoretical eigenvectors, for this non-diagonal autocorrelation operator case. Although the performance of our approach is not too bad for the announced case where the coefficients of $\rho$, with respect to $(\phi_j \otimes \phi_h, j,h \geq 1)$, $j \neq h$, are close to zero. Hence, in the general non-diagonal autocorrelation operator case, our approach can be applied, when smoothing of data has been previously implemented, considering, for example, suitable penalized wavelet nonparametric regression (see, for instance, [3] and [4]), leading to a sparse, almost diagonal, representation of $\hat{C}_n$ and $\hat{D}_n$, in terms of suitable wavelet bases (see comments given in the next section).
Our approach

| $n_1 = 15000$ | 4 | 0.5812 | 8.94($10)^{-2}$ | 0.1055 |
| $n_2 = 35000$ | 5 | 0.5604 | 7.05($10)^{-2}$ | 9.49($10)^{-2}$ |
| $n_3 = 55000$ | 6 | 0.5480 | 6.67($10)^{-2}$ | 9.14($10)^{-2}$ |
| $n_4 = 75000$ | 6 | 0.5322 | 6.24($10)^{-2}$ | 8.85($10)^{-2}$ |
| $n_5 = 95000$ | 6 | 0.5115 | 5.89($10)^{-2}$ | 8.47($10)^{-2}$ |
| $n_6 = 115000$ | 6 | 0.4975 | 5.62($10)^{-2}$ | 8.04($10)^{-2}$ |
| $n_7 = 135000$ | 7 | 0.4946 | 5.57($10)^{-2}$ | 7.66($10)^{-2}$ |
| $n_8 = 155000$ | 7 | 0.4810 | 5.28($10)^{-2}$ | 7.24($10)^{-2}$ |
| $n_9 = 175000$ | 7 | 0.4735 | 5.01($10)^{-2}$ | 6.78($10)^{-2}$ |
| $n_{10} = 195000$ | 7 | 0.4608 | 4.90($10)^{-2}$ | 6.30($10)^{-2}$ |
| $n_{11} = 215000$ | 7 | 0.4424 | 4.69($10)^{-2}$ | 6.07($10)^{-2}$ |
| $n_{12} = 235000$ | 7 | 0.4250 | 4.45($10)^{-2}$ | 5.82($10)^{-2}$ |
| $n_{13} = 255000$ | 7 | 0.4106 | 4.25($10)^{-2}$ | 5.54($10)^{-2}$ |
| $n_{14} = 275000$ | 8 | 0.4080 | 4.14($10)^{-2}$ | 5.16($10)^{-2}$ |
| $n_{15} = 295000$ | 8 | 0.3808 | 4.09($10)^{-2}$ | 4.81($10)^{-2}$ |
| $n_{16} = 315000$ | 8 | 0.3604 | 3.85($10)^{-2}$ | 4.53($10)^{-2}$ |
| $n_{17} = 335000$ | 8 | 0.3489 | 3.56($10)^{-2}$ | 4.29($10)^{-2}$ |
| $n_{18} = 355000$ | 8 | 0.3302 | 3.29($10)^{-2}$ | 3.98($10)^{-2}$ |
| $n_{19} = 375000$ | 8 | 0.3204 | 2.90($10)^{-2}$ | 3.75($10)^{-2}$ |
| $n_{20} = 395000$ | 8 | 0.3177 | 2.62($10)^{-2}$ | 3.44($10)^{-2}$ |

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 (25)-(26) (third column), in equations (95)-(96) (fourth column), and in equations (97)-(98) (fifth column), from the non-diagonal data generated by equations (118)-(121), based on $N = 200$ (due to high-dimensionality) simulations, for $\delta_1 = 2.4$ and $\delta_2 = 1.1$, considering the sample sizes $n_t = 15000 + 20000(t-1)$, $t = 1, \ldots, 20$, and the corresponding $k_n = \lceil n^{1/\alpha} \rceil$, $\alpha = 6$ values (truncation criteria from Assumption A3), assuming that $(\phi_j, j \geq 1)$ are known. The time discretization step is assumed as $h_t = 0.015$.

6 Final comments.

The present paper proves that, under Assumptions A1–A4, the formulated componentwise estimator $\hat{\rho}_{k_n}$ converges to $\rho$, in the space $L^2_{S_{ii}(H)}(\Omega, A, P)$. Its associated ARH(1) plug-in predictor converges, in the space $L^2_{H}(\Omega, A, P)$, to the predictor $E[X_n | X_{n-1}] = \rho(X_{n-1})$. Therefore, both, componentwise functional parameter estimator and plug-in predictor, are consistent.

In the numerical examples considered, for the special family of covariance operators studied, and for
the particular selection of $k_n$ made, the empirical truncated functional mean-square errors of $\hat{\rho}_n$ display a rate of convergence to zero faster than $(1/n)^{3/4}$, and slower than $(1/n)$, for the large sample sizes studied. Furthermore, the empirical truncated functional mean absolute errors of $\hat{X}_n$ present a decay velocity faster than $(1/n)^{3/8}$ (or, as displayed, faster than $(1/n)^{3/9}$), and slower than $(1/n)^{1/2}$, for the large sample sizes studied.

In practice, our approach can be implemented when the eigenvectors of the autocovariance operator $C$ are known. In the situation pointed out in Remark 1, where $\rho$ and $C$ share the same eigenvector system, we refer to the case of physical phenomena, whose dynamics is described from a random initial condition satisfying a given stochastic differential equation, in the mean-square sense, with driven process defined by a white noise process in time (e.g., Gaussian white noise process). In this case, the system of eigenvectors of the auto-covariance operator is known, since coincide with the eigenvectors of the differential operator defining such a stochastic differential equation. Otherwise, as illustrated in the simulation study undertaken, they must be approximated by some empirical eigenvectors (e.g., the eigenvectors of the empirical autocovariance operator). We comment here some alternatives to obtain a suitable basis of empirical eigenvectors satisfying Assumptions A1-A2. Since wavelet bases are well suited for sparse representation of functions, recent work has considered combining them with sparsity-inducing penalties, both for semiparametric regression ([57]), and for regression with functional or kernel predictors (see [57], [58] and [59], among others). The latter papers focused on $l_1$ penalization, also known as the lasso (see [54]), in the wavelet domain. Alternatives to the lasso include the SCAD penalty (see [21]), and the adaptive lasso (see, for example, [60]). The $l_1$ 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 representation $\tilde{X}_1, \ldots, \tilde{X}_n$ of the functional data $X_1, \ldots, X_n$, whose empirical auto-covariance operator

\[
\hat{C}_n = \frac{1}{n} \sum_{i=0}^{n-1} \tilde{X}_i \otimes \tilde{X}_i
\]

and cross-covariance operator

\[
\hat{D}_n = \frac{1}{n-1} \sum_{i=0}^{n-2} \tilde{X}_i \otimes \tilde{X}_{i+1}
\]

admits a diagonal representation in terms of wavelets.

On the other hand, estimating a covariance matrix is an important task in applications where the number of variables is larger than the number of observations. In the literature, shrinkage approaches for estimating a high-dimensional covariance matrix are employed to circumvent the limitations of the sample covariance matrix. Particularly, a new family of nonparametric Stein-type shrinkage covariance estimators is proposed in [55] (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 estimator of the auto-covariance $C$ and cross-covariance operator $D$, with respect to a common suitable wavelet function basis, which can lead to a empirical diagonal representation of both operators, in terms of such wavelet basis.

Acknowledgments

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

References

[1] G. Aneiros-Pérez and P. Vieu, Semi-functional partial linear regression, *Stat. Probab. Letters* 76, 2006, pp. 1102–1110.

[2] G. Aneiros-Pérez and P. Vieu, Nonparametric time series prediction: A semi-functional partial linear modeling, *J. Multivariate Anal.* 99, 2008, pp. 834–857.

[3] C. Angelini, D. De Canditiis and F. Leblanc, Wavelet regression estimation in nonparametric mixed effect models, *J. Multivariate Anal.* 85, 2003, pp. 267–291.

[4] A. Antoniadis and T. Sapatinas, Wavelet methods for continuous-time prediction using Hilbert-valued autoregressive processes, *J. Multivariate Anal.* 87, 2003, pp. 133–158.

[5] M. S. Bartlett, On the theoretical specification and sampling properties of autocorrelated time series, Supplement to the *Journal of the Royal Statistical Society* 8, 1946, pp. 27–41.

[6] N. Bensmain and T. Mourid, Estimateur “sieve” de l’opérateur d’un processus ARH(1), *C. R. Acad. Sci. Paris Sér. I Math.* 332, 2001, pp. 1015–1018.

[7] P. C. Besse and H. Cardot, Approximation spline de la prévision d’un processus fonctionnel autorégressif d’ordre 1, *Canad. J. Statist.* 24, 1996, pp. 467–487.

[8] P. C. Besse, H. Cardot and F. Ferraty, Simultaneous non-parametric regressions of unbalanced longitudinal data, *Comp. Stat. Data Anal.* 24, 1997, pp. 255–270.

[9] P. C. Besse, H. Cardot and D. B. Stephenson, Autoregressive Forecasting of Some Functional Climatic Variations, *Scand. J. Statist.* 27, 2000, pp. 673–687.

[10] E. G. Bongiorno, A. Goia, E. Salinelli and P. Vieu, *Contributions in infinite-dimensional statistics and related topics*, Società Editrice Esculapio, Bologna 2014.
[11] D. Bosq, *Linear Processes in Function Spaces*, Springer-Verlag, New York 2000.

[12] D. Bosq, General linear processes in Hilbert spaces and prediction, *J. Stat. Plann. Inference* 137, 2007, pp. 879–894.

[13] D. Bosq and D. Blanke, *Inference and Predictions in Large Dimensions*, John Wiley, Chichester 2007.

[14] J. Chen, D. Lin and H. Hochner, Semiparametric maximum likelihood methods for analyzing genetic and environmental effects with case-control motherchild pair data, *Biometrics* 68, 2012, pp. 869–877.

[15] A. Cuevas, A partial overview of the theory of statistics with functional data, *J. Stat. Plann. Infer.* 147, 2014, pp. 1–23.

[16] J. Cugliari, *Prévision non paramétrique de processus à valeurs fonctionnelles. Application à la consommation d’électricité*, PhD thesis, University of Paris-Sud 11, Paris 2011.

[17] J. Damon and S. Guillas, Estimation and simulation of autoregressive processes with exogenous variables, *Stat. Inference* 8, 2005, pp. 185–204.

[18] R. Dautray and J. -L. Lions, *Mathematical Analysis and Numerical Methods for Science and Technology Volume 3: Spectral Theory and Applications*, Springer, New York 1990.

[19] J. Dedecker and F. Merlevede, The conditional central limit theorem in Hilbert spaces, *Stochastic processes and their applications* 108, 2003, pp. 229–262.

[20] H. Dehling and O. S. Sharipov, Estimation of mean and covariance operator for Banach space valued autoregressive processes with dependent innovations, *Stat. Inference Stoch. Process.* 8, 2005, pp. 137–149.

[21] J. Fan and R. Li, Variable selection via nonconcave penalized likelihood and its oracle properties, *J. Amer. Statist. Assoc.* 96, 2001, pp. 1348–1360.

[22] M. Febrero-Bande and W. Gonzalez-Manteiga, Generalized additive models for functional data, *Test* 22, 2013, pp. 278–292.

[23] F. Ferraty, A. Goia, E. Salinelli and P. Vieu, Functional projection pursuit regression, *Test* 22, 2013, pp. 293–320.

[24] F. Ferraty, I. V. Keilegom and P. Vieu, Regression when both response and predictor are functions, *J. Multivariate Anal.* 109, 2012, pp. 10–28.

[25] F. Ferraty and P. Vieu, *Nonparametric functional data analysis. Springer Series in Statistics*, Springer, New York 2006.
[26] F. Ferraty and P. Vieu, Additive prediction and boosting for functional data, *Comp. Stat. Data Anal.* 53, 2009, pp. 1400–1413.

[27] D. S. Grebenkov and B. T. Nguyen, Geometrical structure of Laplacian eigenfunctions, *SIAM Review* 55, 2013, pp. 601–667.

[28] U. Grenander, *Abstract Inference*, Wiley, New York 1981.

[29] A. Goia and P. Vieu, An introduction to recent advances in high/infinite dimensional statistics, *J. Multivariate Anal.* 146, 2016, pp. 1–6.

[30] S. Guillas, Noncausality and functional discretization, limit theorems for an ARHX(1) process, *C. R. Acad. Sci. Paris Sér. I Math.* 331, 2000, pp. 91–94.

[31] S. Guillas, Rates of convergence of autocorrelation estimates for autoregressive Hilbertian processes, *Stat. Prob. Lett.* 55, 2001, pp. 281–291.

[32] J. Gurland, Quadratic forms in normally distributed random variables, *Sankhya* 17, 1956, pp. 37–50.

[33] J. D. Hamilton, *Time Series Analysis*, Princeton University Press, Princeton 1994.

[34] L. Horváth and P. Kokoszka, *Inference for Functional Data with Applications*, in: *Springer Series in Statistics*, Springer, New York 2012.

[35] T. Hsing and R. Eubank, *Theoretical Foundations of Functional Data Analysis, with an Introduction to Linear Operators*, Wiley Series in Probability and Statistics, John Wiley & Sons, United Kingdom 2015.

[36] V. Kargin and A. Onatski, Curve forecasting by functional autoregression, *J. Multivariate Anal.* 99, 2008, pp. 2508–2526.

[37] S. Mallat, *A Wavelet Tour of Signal Processing: The Sparse Way*, 3rd ed. Academic Press, Burlington, MA 2009.

[38] J. -M. Marion and B. Pumo, Comparison of ARH(1) and ARHD(1) models on physiological data, *Ann. I.S.U.P.* 48, 2004, pp. 29–38.

[39] A. Mas, 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.

[40] A. Mas, Consistance du prédicteur dans le modèle ARH(1): le cas compact, *Ann. I.S.U.P.* 48, 2004, pp. 39–48.
[41] A. Mas, Weak-convergence in the functional autoregressive model, *J. Multivariate Anal.* 98, 2007, pp. 1231–1261.

[42] A. Mas and L. Menneteau, Large and moderate deviations for infinite dimensional autoregressive processes, *J. Multivariate Anal.* 87, 2003, pp. 241–260.

[43] L. Menneteau, Some laws of the iterated logarithm in Hilbertian autoregressive models, *J. Multivariate Anal.* 92, 2005, pp. 405–425.

[44] F. Merlevède, 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, 1996, pp. 521–524.

[45] F. Merlevède, Résultats de convergence presque sûre pour l’estimation et la prévision des processus linéaires hilbertiens, *C. R. Acad. Sci. Paris Sér. I Math.* 324, 1997, pp. 573–576.

[46] T. Mourid, Processus autorégressifs banachiques d’ordre supérieur, *C. R. Acad. Sci. Paris Sér. I Math.* 317, 1993, pp. 1167–1172.

[47] T. Mourid, Processus autorégressifs hilbertiens à coefficients aléatoires, *Ann. I.S.U.P.* 48, 2004, pp. 79–85.

[48] G. P. Nason, *Wavelet methods in Statistics with R*, Springer 2008.

[49] R. T. Ogden, *Essential Wavelets for Statistical Applications and Data Analysis*, Birkhauser, Boston, MA 1997.

[50] B. Pumo, *Estimation et prévision de processus autorégressifs fonctionnels*, PhD thesis, University of Paris 6, Paris 1992.

[51] B. Pumo, Prediction of continuous time processes by $C_{[0,1]}$-valued autoregressive process, *Stat. Inference Stoch. Process.* 1, 1998, pp. 297–309.

[52] J. O. Ramsay and B. W. Silverman, *Functional data analysis*, second ed. Springer Series in Statistics. Springer, New York 2005.

[53] M. D. Ruiz-Medina, Spatial functional prediction from spatial autorregresive Hilbertian processes, *Environmetrics* 23, 2012, pp. 119–128.

[54] R. Tibshirani, Regression shrinkage and selection via the lasso, *J. Roy. Statist. Soc. Ser. B*, 58, pp. 267–288.

[55] A. Touloumis, Nonparametric Stein-type Shrinkage Covariance Matrix Estimators in High-Dimensional Settings, *Comp. Stat. Data Anal.* 83, 2015, pp. 251–261.
[56] B. Vidakovic, Nonlinear wavelet shrinkage with Bayes rules and Bayes factors, *J. Amer. Statist. Assoc.* 93, 1998, pp. 173–179.

[57] M. P. Wand and J. T. Ormerod, Penalized wavelets: Embedding wavelets into semiparametric regression, *Electron. J. Stat.* 5, 2011, pp. 1654–1717.

[58] Y. Zhao, H. Chen and R. T. Ogden, Wavelet-based weighted LASSO and screening approaches in functional linear regression, *J. Comput. Graph. Statist.*, 2015, to appear.

[59] Y. Zhao, R. T. Ogden and P. T. Reiss, Wavelet-based lasso in functional linear regression, *J. Comput. Graph. Statist.* 21, 2012, pp. 600–617.

[60] H. Zou, The adaptive lasso and its oracle properties, *J. Amer. Statist. Assoc.* 101, 2006, pp. 1418–1429.