Journal of Multivariate Analysis 155 (2017) 12–34

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

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

J. Álvarez-Liébana [a], D. Bosq [b], M.D. Ruiz-Medina [a,*]

[a] Department of Statistics and Operation Research, Faculty of Sciences, Universidad de Granada, Campus Fuente Nueva s/n, 18071 Granada, Spain

[b] LSTA, Université Pierre et Marie Curie-Paris 6, 4, place Jussieu, 75005, Paris, France

ARTICLE INFO

Article history:
Received 7 March 2016
Available online 1 December 2016

AMS subject classifications:
11C20
11H55
11M50
15A24
15A63
15A69
34L05
60B12
60G10
60G15
60G25
60H25
62M10
62M15
62M20

Keywords:
ARH(1) processes
Consistency
Functional prediction
Mean absolute and quadratic convergence

ABSTRACT

This paper presents new results on the prediction of linear processes in function spaces. The autoregressive Hilbertian process framework of order one (ARH(1) framework) is adopted. A component-wise estimator of the autocorrelation operator is derived from the moment-based estimation of its diagonal coefficients with respect to the orthogonal eigenvectors of the autocovariance operator, which are assumed to be known. Mean-square convergence to the theoretical autocorrelation operator is proved in the space of Hilbert–Schmidt operators. Consistency then follows in that space. Mean absolute convergence, in the underlying Hilbert space, of the ARH(1) plug-in predictor to the conditional expectation is obtained as well. A simulation study is undertaken to illustrate the large-sample behavior of the formulated component-wise estimator and predictor. Additionally, alternative component-wise (with known and unknown eigenvectors), regularized, wavelet-based penalized, and nonparametric kernel estimators of the autocorrelation operator are compared with the one presented here, in terms of prediction.

1. Introduction

In the last few decades, an extensive literature on statistical inference from functional random variables has emerged. This work was motivated in part by the statistical analysis of high-dimensional data, as well as data of a continuous (infinite-dimensional) nature; see, e.g., [9,10,17,23,41,42,49–51]. New developments in functional data analysis are described, e.g., in [8,13,31,32], and in a recent Special Issue of this journal [25].

The special case of functional regression models, in which the predictor is a random function and the response is scalar, has been particularly well studied. Various specifications of the functional regression parameter arise in fields such

* Corresponding author.
E-mail addresses: javialvaliebana@ugr.es (J. Álvarez-Liébana), denis.bosq@upmc.fr (D. Bosq), mruiz@ugr.es (M.D. Ruiz-Medina).

http://dx.doi.org/10.1016/j.jmva.2016.11.009

as biology, climatology, chemometrics, and economics. To avoid the computational (high-dimensional) limitations of the nonparametric approach, several parametric and semi-parametric methods have been proposed; see, e.g., [21] and the references therein.

In Ferraty et al. [21], a combination of a spline approximation and the one-dimensional Nadaraya–Watson approach was proposed to avoid high dimensionality issues. Generalizations to the case of more regressors (all functional, or both functional and real) were also addressed in the nonparametric, semi-parametric, and parametric frameworks; for an overview, see [1,20,24].

In the nonparametric regression framework, the case where the covariates and the response are functional was considered by Ferraty et al. [22], where a functional version of the Nadaraya–Watson estimator was proposed for the estimation of the regression operator and shown to be point-wise asymptotically normal. Resampling techniques were used to overcome the difficulties arising in the estimation of the asymptotic bias and variance.

Semi-functional partial linear regression, introduced in [2], allows the prediction of a real-valued random variable from a set of real-valued explanatory variables, and a time-dependent functional explanatory variable. Motivated by genetic and environmental applications, a semiparametric maximum likelihood method for the estimation of odds ratio association parameters was developed by Chen et al. [12] in a high-dimensional data context.

In the autoregressive Hilbertian time series framework, several estimation and prediction procedures have been proposed and studied. Mas [36] established, under suitable conditions, the asymptotic normal distribution of the formulated estimator of the autocorrelation operator, based on projection into the theoretical eigenvectors. In [9,11], the problem of prediction of linear processes in function spaces was addressed. In particular, sufficient conditions for the consistency of the empirical autocovariance and cross-covariance operators were obtained. The asymptotic normal distribution of the empirical autocovariance operator was also derived. Moreover, the asymptotic properties of the empirical eigenvalues and eigenvectors were analyzed.

Guillas [28] established the efficiency of a component-wise estimator of the autocorrelation operator, based on projection into the empirical eigenvector system of the autocovariance operator. Consistency, in the space of bounded linear operators, of the formulated estimator of the autocorrelation operator, and of its associated ARH(1) plug-in predictor was later proved by Mas [37]. He later derived sufficient conditions for the weak convergence of the ARH(1) plug-in predictor to a Hilbert-valued Gaussian random variable; see [38]. In parallel, Mas obtained high deflection results or large and moderate deviations for infinite-dimensional autoregressive processes [39]. Furthermore, the law of the iterated logarithm for the covariance operator estimator was formulated by Menneteau [40].

The main properties of the class of autoregressive Hilbertian processes with random coefficients were investigated by Mourid [44]. Kargin and Onatski [33] gave interesting extensions of the autoregressive Hilbertian framework, based on the spectral decomposition of the autocorrelation operator, and not of the autocovariance operator. The first generalization on autoregressive processes of order greater than one was proposed by Mourid [43], in order to improve prediction. ARHX(1) models, i.e., autoregressive Hilbertian processes with exogenous variables were studied by Damon and Guillas [15]. In [27,28] a doubly stochastic formulation of the autoregressive Hilbertian process was investigated. The ARHD model was introduced by Marion and Pumo [35], taking into account the regularity of trajectories through the derivatives. The conditional autoregressive Hilbertian process (CARH process) was considered by Cugliari [14], developing parallel projection estimation methods to predict such processes. In the Banach-valued context, we refer to the papers by [6,18,47,48], among others.

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

Under the assumptions stated in Sections 2 and 3, we establish the convergence in the $\mathcal{L}^2$-sense of a component-wise estimator of the autocorrelation operator in the space of Hilbert–Schmidt operators $\mathcal{S}(H)$, i.e., convergence in the space $\mathcal{L}^2_{\mathcal{S}(H)}(\Omega, \mathcal{A}, \mathcal{P})$. Consistency then follows in $\mathcal{S}(H)$. Under the same conditions, consistency in H of the associated ARH(1) plug-in predictor is obtained, from its convergence in the $\mathcal{L}^1$-sense in the Hilbert space H, i.e., in the space $\mathcal{L}^1_H(\Omega, \mathcal{A}, \mathcal{P})$.

The Gaussian framework is analyzed in Section 4 and illustrated in Section 5, where examples show the behavior of the proposed component-wise autocorrelation operator estimator, and associated predictor, for large sample sizes. We also present there a comparative study with alternative ARH(1) prediction techniques, including component-wise parameter estimation of the autocorrelation operator, from known and unknown eigenvectors, as well as kernel (nonparametric) functional estimation, and penalized, spline and wavelet, estimation. Final comments on the application of the proposed approach from real data are provided in Section 6.

2. Preliminaries

This section contains the preliminary definitions and lemmas that will be used to derive the main results of this paper. In the following, H denotes a real separable Hilbert space. Recall from [9] that a zero-mean ARH(1) process $X = \{X_n : n \in \mathbb{Z}\}$ satisfies, for all $n \in \mathbb{Z}$, the equation

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

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

The following assumptions are made.

Assumption A1. The autocovariance operator $C = \mathrm{E}(X_n \otimes X_n) = \mathrm{E}(X_0 \otimes X_0)$, for all $n \in \mathbb{Z}$, is a positive self-adjoint and trace operator. As a result, it admits the following diagonal spectral representation

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

in terms of an orthonormal system $\{\phi_j : j \geq 1\}$ of eigenvectors which are known. Here, $C_1 \geq C_2 \geq \cdots > 0$ denote the real positive eigenvalues of C arranged in decreasing order of magnitude and $\sum_{j=1}^{\infty} C_j < \infty$.

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

$$\rho = \sum_{j=1}^{\infty} \rho_j \phi_j \otimes \phi_j, \qquad \sum_{j=1}^{\infty} \rho_j^2 < \infty, \tag{3}$$

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

Note that, under Assumption A2, $\|\rho\|_{\mathcal{L}(H)} = \sup_{j\geq 1} \rho_j < 1$.

Remark 1. Assumption A2 holds in particular when the operator ρ is defined as a continuous function of operator C; see [16, pp. 119–140]. See also Remark 4.

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

Remark 2. Under Assumptions A1–A2, it follows from Eq. (1) that

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

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

$$X_{n,j} = \rho_j X_{n-1,j} + \varepsilon_{n,j}, \tag{4}$$

where $X_{n,j} = \left\langle X_n, \phi_j \right\rangle_H$ and $\varepsilon_{n,j} = \left\langle \varepsilon_n, \phi_j \right\rangle_H$ for all $n \in \mathbb{Z}$. From Eq. (4), we have, for each $j \geq 1$ and all $n \in \mathbb{Z}$,

$$\begin{aligned} \rho_j &= \rho(\phi_j)(\phi_j) = \left\langle \phi_j, DC^{-1}(\phi_j) \right\rangle_H = \left\langle D(\phi_j), \phi_j \right\rangle_H \left\langle C^{-1}(\phi_j), \phi_j \right\rangle_H \\ &= \frac{\mathrm{E}(X_{n,j} X_{n-1,j})}{\mathrm{E}(X_{n-1,j}^2)} = \frac{D_j}{C_j}, \end{aligned} \tag{5}$$

where $D_j = \left\langle D(\phi_j), \phi_j \right\rangle_H = \mathrm{E}(X_{n,j} X_{n-1,j})$ and $C_j^{-1} = \{\mathrm{E}(X_{n-1,j}^2)\}^{-1}$, given that, for all $j \geq 1$,

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

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

$$\forall Z \in \mathcal{L}^2_{\mathcal{H}}(\Omega, \mathcal{A}, \mathcal{P}), \quad \|Z\|_{\mathcal{L}^2_{\mathcal{H}}(\Omega,\mathcal{A},\mathcal{P})} = \sqrt{\mathrm{E}\left(\|Z\|^2_{\mathcal{H}}\right)}. \tag{7}$$

That is, for $Z, Y \in \mathcal{L}^2_{\mathcal{H}}(\Omega, \mathcal{A}, \mathcal{P})$, Z and Y belong to the same equivalence class if and only if $\mathrm{E}\left(\|Z-Y\|^2_{\mathcal{H}}\right) = 0$. The convergence in the seminorm of $\mathcal{L}^2_{\mathcal{S}(H)}(\Omega, \mathcal{A}, \mathcal{P})$ will be considered in Proposition 1, where $\mathcal{H} = \mathcal{S}(H)$ denotes the Hilbert space of Hilbert–Schmidt operators on a Hilbert space H.

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

$$X_n = \sum_{j=1}^{\infty} \sqrt{C_j}\,\frac{\langle X_n, \phi_j\rangle_H}{\sqrt{C_j}}\phi_j = \sum_{j=1}^{\infty} \sqrt{C_j}\,\frac{X_{n,j}}{\sqrt{C_j}}\phi_j = \sum_{j=1}^{\infty} \sqrt{C_j}\,\eta_j(n)\phi_j, \tag{8}$$

$$\varepsilon_n = \sum_{j=1}^{\infty} \sigma_j \frac{\langle \varepsilon_n, \phi_j\rangle_H}{\sigma_j}\phi_j = \sum_{j=1}^{\infty} \sigma_j \frac{\varepsilon_{n,j}}{\sigma_j}\phi_j = \sum_{j=1}^{\infty} \sigma_j \widetilde{\eta}_j(n)\phi_j, \tag{9}$$

where $\eta_j(n) = \langle X_n, \phi_j\rangle_H / \sqrt{C_j} = X_{n,j}/\sqrt{C_j}$ and $\widetilde{\eta}_j(n) = \langle \varepsilon_n, \phi_j\rangle_H / \sigma_j = \varepsilon_{n,j}/\sigma_j$, for every $n \in \mathbb{Z}$, and for each $j \geq 1$. Here, under Assumptions A1 and A2, for $R_\varepsilon = \mathrm{E}(\varepsilon_n \otimes \varepsilon_n) = \mathrm{E}(\varepsilon_0 \otimes \varepsilon_0)$, $n \in \mathbb{Z}$, one has, for all $j \geq 1$,

$$R_\varepsilon \phi_j = \sigma_j^2 \phi_j,$$

where, as before, $\{\phi_j : j \geq 1\}$ denotes the system of eigenvectors of the autocovariance operator C, and $\sum_{j\geq 1} \sigma_j^2 = \sigma_\varepsilon^2 = \mathrm{E}(\|\varepsilon_n\|_H^2)$, for all $n \in \mathbb{Z}$.

The following lemma provides the convergence, in the seminorm of $\mathcal{L}^2_H(\Omega, \mathcal{A}, \mathcal{P})$, of the series expansions (8) and (9).

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*

$$\lim_{M\to\infty} \mathrm{E}(\|X_n - \widehat{X}_{n,M}\|_H^2) = 0, \tag{10}$$

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

$$\lim_{M\to\infty} \|\mathrm{E}\{(X_n - \widehat{X}_{n,M}) \otimes (X_n - \widehat{X}_{n,M})\}\|^2_{\mathcal{S}(H)} = 0. \tag{11}$$

Similar assertions hold for the biorthogonal series representation

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

Proof. Under Assumption A1, from the trace property of C, the sequence $\widehat{X}_{n,M}$ satisfies, for M sufficiently large, and $L > 0$, arbitrary,

$$\begin{aligned}
\|\widehat{X}_{n,M+L} - \widehat{X}_{n,M}\|^2_{\mathcal{L}^2_H(\Omega,\mathcal{A},P)} &= \mathrm{E}(\|\widehat{X}_{n,M+L} - \widehat{X}_{n,M}\|_H^2) \\
&= \sum_{j=M+1}^{M+L}\sum_{k=M+1}^{M+L} \sqrt{C_j}\sqrt{C_k}\,\mathrm{E}\left\{\eta_j(n)\eta_k(n)\right\}\langle \phi_j, \phi_k\rangle_H \\
&= \sum_{j=M+1}^{M+L} C_j \to 0, \quad M \to \infty,
\end{aligned} \tag{12}$$

since, under Assumption A1, $\sum_{j=1}^{M} C_j$ is a Cauchy sequence, and $\sum_{j=M+1}^{M+L} C_j$ converges to zero when $M \to \infty$, for $L > 0$, arbitrary. From Eq. (12), $\widehat{X}_{n,M}$ is also a Cauchy sequence in $L^2_H(\Omega, \mathcal{A}, P)$ and hence it has finite limit in $L^2_H(\Omega, \mathcal{A}, P)$.

Furthermore,

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

$$= \sigma_X^2 - \lim_{M\to\infty} \sum_{j=1}^{M} C_j = 0. \tag{13}$$

In the derivation of the identities in (12) and (13), we used the fact that, for every $j,\ h \geq 1$,

$$\begin{aligned} &C\phi_j = C_j\phi_j, \qquad \langle \phi_j, \phi_h\rangle_H = \delta_{j,h}, \qquad \sigma_X^2 = \mathrm{E}(\|X_n\|_H^2) = \sum_{j=1}^{\infty} C_j < \infty \\ &\mathrm{E}\left\{\eta_j(n)\eta_h(n)\right\} = \delta_{j,h}, \qquad \mathrm{E}\left\{\langle X_n, \eta_j(n)\phi_j\rangle_H\right\} = \sqrt{C_j}. \end{aligned} \tag{14}$$

Moreover, from identities in (14), we have

$$\begin{aligned} &\left\| \mathrm{E}\left\{ \left(X_n - \lim_{M\to\infty} \widehat{X}_{n,M} \right) \otimes \left(X_n - \lim_{M\to\infty} \widehat{X}_{n,M} \right) \right\} \right\|_{\mathcal{S}(H)}^2 \\ &= \left\| \mathrm{E}(X_n \otimes X_n) + \lim_{M\to\infty} \sum_{j=1}^{M}\sum_{h=1}^{M} \sqrt{C_j}\sqrt{C_h}\phi_j \otimes \phi_h \mathrm{E}\left\{\eta_j(n)\eta_h(n)\right\} - 2\lim_{M\to\infty}\sum_{j=1}^{M} \mathrm{E}\left\{ X_n \otimes \sqrt{C_j}\eta_j(n)\phi_j \right\} \right\|_{\mathcal{S}(H)}^2 \\ &= \left\| \mathrm{E}(X_n \otimes X_n) + \lim_{M\to\infty} \left(\sum_{j=1}^{M} C_j\phi_j \otimes \phi_j - 2\sum_{j=1}^{M} C_j\phi_j \otimes \phi_j \right) \right\|_{\mathcal{S}(H)}^2 \\ &= \left\| \mathrm{E}(X_n \otimes X_n) - \lim_{M\to\infty} \sum_{j=1}^{M} C_j\phi_j \otimes \phi_j \right\|_{\mathcal{S}(H)}^2 = 0. \end{aligned} \tag{15}$$

In a similar way, we can derive the convergence of $\sum_{j=1}^{\infty} \sigma_j \widetilde{\eta}_j(n)\phi_j$ to ε_n, in $\mathcal{L}_H^2(\Omega, \mathcal{A}, \mathcal{P})$, for every $n \in \mathbb{Z}$, since ε is assumed to be strong-white noise, and hence, its covariance operator R_ε is in the trace class. We can also obtain an analog to Eq. (15). □

In Eqs. (8)–(9), for every $n \in \mathbb{Z}$ and $j, h \geq 1$, we have

$$\mathrm{E}\left\{\eta_j(n)\right\} = 0, \qquad \mathrm{E}\left\{\eta_j(n)\eta_h(n)\right\} = \delta_{j,h}, \tag{16}$$

$$\mathrm{E}\left\{\widetilde{\eta}_j(n)\right\} = 0, \qquad \mathrm{E}\left\{\widetilde{\eta}_j(n)\widetilde{\eta}_h(n)\right\} = \delta_{j,h}. \tag{17}$$

Note that, for each $j \geq 1$, $\left\{X_{n,j} : n \in \mathbb{Z}\right\}$ in Eq. (4) defines a stationary and invertible AR(1) process. In addition, from Eqs. (8) and (14), for every $n \in \mathbb{Z}$ and $j, p \geq 1$,

$$\begin{aligned} X_n &= \sum_{j=1}^{\infty} X_{n,j}\phi_j, \\ \mathrm{E}(X_{n,j}X_{n,p}) &= \sum_{k=0}^{\infty}\sum_{h=0}^{\infty} \rho_j^k\rho_p^h \mathrm{E}(\varepsilon_{n-k,j}\varepsilon_{n-h,p}) = \delta_{j,p}\sum_{k=0}^{\infty} \rho_j^{2k}\sigma_j^2 = \delta_{j,p}\frac{\sigma_j^2}{1-\rho_j^2}, \\ \mathrm{E}\left(\|X_n\|_H^2\right) &= \sum_{j=1}^{\infty} \mathrm{E}(X_{n,j}^2) = \sum_{j=1}^{\infty} \langle C\left(\phi_j\right), \phi_j\rangle_H = \sum_{j=1}^{\infty} C_j = \sigma_X^2 < \infty, \end{aligned} \tag{18}$$

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

$$\begin{aligned} \mathrm{E}\left\{\eta_j(n)\eta_j(n+1)\right\} &= \mathrm{E}\left(\frac{X_{n,j}}{\sqrt{C_j}}\frac{X_{n+1,j}}{\sqrt{C_j}}\right) = \frac{\mathrm{E}(X_{n,j}X_{n+1,j})}{C_j} \\ &= \frac{\sum_{k=0}^{\infty}\sum_{h=0}^{\infty} \rho_j^{k+h}\mathrm{E}(\varepsilon_{n-k,j}\varepsilon_{n+1-h,j})}{C_j} \\ &= \frac{\sum_{k=0}^{\infty} \rho_j^{2k+1}\sigma_j^2}{C_j} = \frac{\sigma_j^2}{C_j}\frac{\rho_j}{1-\rho_j^2} = \rho_j. \end{aligned} \tag{19}$$

Remark 3. From Eq. (4) and Lemma 1, keeping in mind that $C_j = \sigma_j^2 / \left(1 - \rho_j^2\right)$, for each $j \geq 1$, the following invertible and stationary AR(1) process can be defined:

$$\eta_j(n) = \rho_j \eta_j(n-1) + \sqrt{1-\rho_j^2}\widetilde{\eta}_j(n), \quad 0 < \rho_j^2 \leq \left|\rho_j\right| < 1, \tag{20}$$

where, for each $j \geq 1$, $\left\{\eta_j(n) : n \in \mathbb{Z}\right\}$ and $\left\{\widetilde{\eta}_j(n) : n \in \mathbb{Z}\right\}$ are respectively introduced in Eqs. (8) and (9). In the following, for each $j \geq 1$, we assume that $\mathrm{E}\left\{\widetilde{\eta}_j(n)^4\right\} < \infty$, for every $n \in \mathbb{Z}$, to ensure ergodicity for all second-order moments, in the mean-square sense; see, e.g., [30, pp. 192–193].

Furthermore,

$$\begin{aligned} D = \mathrm{E}(X_n \otimes X_{n+1}) &= \sum_{j=1}^{\infty}\sum_{p=1}^{\infty} \mathrm{E}\left\{\left\langle X_n, \phi_j\right\rangle_H \left\langle X_{n+1}, \phi_p\right\rangle_H\right\} \phi_j \otimes \phi_p \\ &= \sum_{j=1}^{\infty}\sum_{p=1}^{\infty} \sqrt{C_j C_p}\, \frac{\mathrm{E}\left\{\left\langle X_n, \phi_j\right\rangle_H \left\langle X_{n+1}, \phi_p\right\rangle_H\right\}}{\sqrt{C_j C_p}} \phi_j \otimes \phi_p \\ &= \sum_{j=1}^{\infty}\sum_{p=1}^{\infty} \sqrt{C_j C_p}\, \mathrm{E}\left\{\eta_j(n)\eta_p(n+1)\right\} \phi_j \otimes \phi_p. \end{aligned} \tag{21}$$

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

$$\mathrm{E}\left\{\eta_j(n)\eta_p(n+1)\right\} = \delta_{j,p}, \tag{22}$$

where $\delta_{j,p}$ denotes the Kronecker delta function. In practice, unconditional bases, e.g., wavelet bases, lead to a sparse representation for functional data; see, e.g., [45,46,54], 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 [34]. The Discrete Wavelet Transform (DWT) approximately decorrelates or whitens data; see [54]. In particular, operators C and D could admit an almost diagonal representation with respect to the self-tensorial product of a suitable wavelet basis.

3. Estimation and prediction results

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

From Eq. (5), for each $j \geq 1$, and for a given sample size n, one can consider the usual respective moment-based estimators $\widehat{D}_{n,j}$ and $\widehat{C}_{n,j}$ of D_j and C_j, in the AR(1) framework, given by

$$\widehat{D}_{n,j} = \frac{1}{n-1}\sum_{i=0}^{n-2} X_{i,j}X_{i+1,j}, \tag{23}$$

$$\widehat{C}_{n,j} = \frac{1}{n}\sum_{i=0}^{n-1} X_{i,j}^2. \tag{24}$$

The following truncated component-wise estimator of ρ is then formulated,

$$\widehat{\rho}_{k_n} = \sum_{j=1}^{k_n} \widehat{\rho}_{n,j}\phi_j \otimes \phi_j, \tag{25}$$

where, for each $j \geq 1$,

$$\widehat{\rho}_{n,j} = \frac{\widehat{D}_{n,j}}{\widehat{C}_{n,j}} = \frac{\sum_{i=0}^{n-2} X_{i,j}X_{i+1,j}/(n-1)}{\sum_{i=0}^{n-1} X_{i,j}^2/n} = \frac{n}{n-1}\,\frac{\sum_{i=0}^{n-2} X_{i,j}X_{i+1,j}}{\sum_{i=0}^{n-1} X_{i,j}^2}. \tag{26}$$

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

$$\lim_{n\to\infty} k_n = \infty, \quad k_n/n < 1,\ n \geq 2. \tag{27}$$

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 5. Assumption A3 has also been considered in [9, p. 217] to ensure weak consistency of the proposed estimator of ρ, as well as in [36] (see Proposition 4, p. 902) in the derivation of asymptotic normality.

From Remark 3, for each $j \geq 1$, $\eta_j = \left\{\eta_j(n);\ n \in \mathbb{Z}\right\}$ in Eq. (20) defines a stationary and invertible AR(1) process, ergodic in the mean-square sense; see, e.g., [5]. Therefore, in view of Eqs. (16) and (19), for each $j \geq 1$, there exist two positive constants $K_{j,1}$ and $K_{j,2}$ such that the following identities hold:

$$\lim_{n\to\infty} n\,\mathrm{E}\left[\left\{1 - \sum_{i=0}^{n-1} \eta_j^2(i)/n\right\}^2\right] = K_{j,1}, \tag{28}$$

$$\lim_{n\to\infty} n\,\mathrm{E}\left[\left\{\rho_j - \sum_{i=0}^{n-2} \eta_j(i)\eta_j(i+1)/(n-1)\right\}^2\right] = K_{j,2}. \tag{29}$$

Eqs. (28) and (29) imply, for n sufficiently large,

$$\mathrm{var}\left\{\frac{1}{n}\sum_{i=0}^{n-1} \eta_j^2(i)\right\} \leq \frac{\widetilde{K}_{j,1}}{n}, \tag{30}$$

$$\mathrm{var}\left\{\frac{1}{n-1}\sum_{i=0}^{n-2} \eta_j(i)\eta_j(i+1)\right\} \leq \frac{\widetilde{K}_{j,2}}{n}, \tag{31}$$

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

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

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

The following assumption is now considered.

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

Remark 6. From Eq. (26), applying the Cauchy–Schwarz inequality, we obtain, for each $j \geq 1$,

$$\left|\widehat{\rho}_{n,j}\right| \leq \frac{n}{n-1}\sqrt{\sum_{i=0}^{n-2} X_{i+1,j}^2 \Big/ \sum_{i=0}^{n-1} X_{i,j}^2} \leq \frac{n}{n-1} \text{ a.s.} \tag{34}$$

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

Next, the convergence of $\widehat{\rho}_{k_n}$ to ρ in the space $\mathcal{L}^2_{\mathcal{S}(H)}(\Omega, \mathcal{A}, \mathcal{P})$ 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:*

$$\lim_{n\to\infty} \left\|\rho - \widehat{\rho}_{k_n}\right\|^2_{\mathcal{L}^2_{\mathcal{S}(H)}(\Omega,\mathcal{A},\mathcal{P})} = 0. \tag{35}$$

Specifically,

$$\left\|\rho - \widehat{\rho}_{k_n}\right\|^2_{\mathcal{L}^2_{\mathcal{S}(H)}(\Omega,\mathcal{A},\mathcal{P})} \leq g(n), \quad \text{with } g(n) = \mathcal{O}\left(\frac{1}{C_{k_n}^2 n}\right),\ n \to \infty. \tag{36}$$

Remark 7. Corollary 4.3 in p. 107 of [9] can be applied to obtain weak convergence results, in terms of weak expectation, using the empirical eigenvectors. See the definition of weak expectation at the beginning of Section 1.3 in p. 27 of [9].

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

$$\left|\rho_j - \widehat{\rho}_{n,j}\right| = \left|\frac{D_j}{C_j} - \frac{\widehat{D}_{n,j}}{\widehat{C}_{n,j}}\right| = \left|\frac{D_j - \widehat{D}_{n,j}}{C_j} + \frac{\widehat{C}_{n,j} - C_j}{C_j}\frac{\widehat{D}_{n,j}}{\widehat{C}_{n,j}}\right| \leq \frac{1}{C_j}(|\widehat{\rho}_{n,j}\,||C_j - \widehat{C}_{n,j}| + |D_j - \widehat{D}_{n,j}|). \tag{37}$$

Thus, under Assumptions A1–A2, from Eq. (34), for each $j \geq 1$,

$$\begin{aligned}\left(\rho_j - \widehat{\rho}_{n,j}\right)^2 &\leq \frac{1}{C_j^2}\left(\left|\widehat{\rho}_{n,j}\right|\ |C_j - \widehat{C}_{n,j}| + |D_j - \widehat{D}_{n,j}|\right)^2 \\ &\leq \frac{2}{C_j^2}\left\{\left(\widehat{\rho}_{n,j}\right)^2 (C_j - \widehat{C}_{n,j})^2 + (D_j - \widehat{D}_{n,j})^2\right\} \\ &\leq \frac{2}{C_j^2}\left\{\left(\frac{n}{n-1}\right)^2 (C_j - \widehat{C}_{n,j})^2 + (D_j - \widehat{D}_{n,j})^2\right\} \text{ a.s.,}\end{aligned} \tag{38}$$

which implies

$$\mathrm{E}\left\{\left(\rho_j - \widehat{\rho}_{n,j}\right)^2\right\} \leq \frac{2}{C_j^2}\left[\left(\frac{n}{n-1}\right)^2 \mathrm{E}\{(C_j - \widehat{C}_{n,j})^2\} + \mathrm{E}\{(D_j - \widehat{D}_{n,j})^2\}\right]. \tag{39}$$

Under Assumption A2, from Eqs. (25) and (39),

$$\begin{aligned}\|\rho - \widehat{\rho}_{k_n}\|^2_{\mathcal{L}^2_{\mathcal{S}(H)}(\Omega,\mathcal{A},\mathcal{P})} &= \mathrm{E}\left(\left\|\rho - \widehat{\rho}_{k_n}\right\|^2_{\mathcal{S}(H)}\right) \\ &= \sum_{j=1}^{k_n} \mathrm{E}\left\{\left(\rho_j - \widehat{\rho}_{n,j}\right)^2\right\} + \sum_{j=k_n+1}^{\infty} \mathrm{E}(\rho_j^2) \\ &\leq \sum_{j=1}^{k_n} \frac{2}{C_j^2}\left[\left(\frac{n}{n-1}\right)^2 \mathrm{E}\{(C_j - \widehat{C}_{n,j})^2\} + \mathrm{E}\{(D_j - \widehat{D}_{n,j})^2\}\right] + \sum_{j=k_n+1}^{\infty} \rho_j^2 \\ &\leq \frac{2}{C_{k_n}^2}\sum_{j=1}^{k_n}\left(\frac{n}{n-1}\right)^2\left[\mathrm{E}\{(C_j - \widehat{C}_{n,j})^2\} + \mathrm{E}\{(D_j - \widehat{D}_{n,j})^2\}\right] + \sum_{j=k_n+1}^{\infty} \rho_j^2 \\ &\leq \frac{2\left(\frac{n}{n-1}\right)^2}{C_{k_n}^2}\sum_{j=1}^{k_n}\left[\mathrm{E}\{(C_j - \widehat{C}_{n,j})^2\} + \mathrm{E}\{(D_j - \widehat{D}_{n,j})^2\}\right] + \sum_{j=k_n+1}^{\infty} \rho_j^2.\end{aligned} \tag{40}$$

Furthermore, from Eqs. (8) and (26), for $j \geq 1$,

$$\widehat{C}_{n,j} = \frac{1}{n}\sum_{i=0}^{n-1} X_{i,j}^2 = \frac{1}{n}\sum_{i=0}^{n-1} C_j\eta_j^2(i), \tag{41}$$

$$\widehat{D}_{n,j} = \frac{1}{n-1}\sum_{i=0}^{n-2} X_{i,j}X_{i+1,j} = \frac{1}{n-1}\sum_{i=0}^{n-2} C_j\eta_j(i)\eta_j(i+1), \tag{42}$$

where, considering Eq. (6),

$$D_j = \mathrm{E}\left\{X_{n,j}X_{n+1,j}\right\} = C_j\mathrm{E}\left\{\eta_j(n)\eta_j(n+1)\right\} = C_j\rho_j \tag{43}$$

for all $j \geq 1$ and $n \in \mathbb{Z}$. Eqs. (40)–(43) then lead to

$$\|\rho - \widehat{\rho}_{k_n}\|^2_{\mathcal{L}^2_{\mathcal{S}(H)}(\Omega,\mathcal{A},\mathcal{P})} \leq \frac{2\left(\frac{n}{n-1}\right)^2}{C_{k_n}^2}\sum_{j=1}^{k_n} C_j^2\left[\mathrm{E}\left[\left\{1 - \frac{1}{n}\sum_{i=0}^{n-1}\eta_j^2\,(i)\right\}^2\right]\right.$$

$$+\,\mathrm{E}\left[\left[\left\{\rho_j-\frac{1}{n-1}\sum_{i=0}^{n-2}\eta_j(i+1)\eta_j(i)\right\}^2\right]\right]+\sum_{j=k_n+1}^{\infty}\rho_j^2. \tag{44}$$

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

$$\begin{aligned}\mathrm{E}\left\{\left\|\rho-\widehat{\rho}_{k_n}\right\|_{\mathcal{S}(H)}^2\right\} &\leq \frac{2\left(\frac{n}{n-1}\right)^2}{C_{k_n}^2}\sum_{j=1}^{k_n}C_j^2\left(\frac{\widetilde{K}_{j,1}+\widetilde{K}_{j,2}}{n}\right)+\sum_{j=k_n+1}^{\infty}\rho_j^2\\ &\leq \frac{2S\left(\frac{n}{n-1}\right)^2}{C_{k_n}^2 n}\sum_{j=1}^{k_n}C_j^2+\sum_{j=k_n+1}^{\infty}\rho_j^2.\end{aligned} \tag{45}$$

From the trace property of operator C,

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

and from the Hilbert–Schmidt property of ρ,

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

Thus, in view of Eq. (45), one has, as $n\to\infty$,

$$\left\|\rho-\widehat{\rho}_{k_n}\right\|_{\mathcal{L}^2_{\mathcal{S}(H)}(\Omega,\mathcal{A},\mathcal{P})}^2=\mathrm{E}\left(\left\|\rho-\widehat{\rho}_{k_n}\right\|_{\mathcal{S}(H)}^2\right)\leq g(n)=\mathcal{O}\left(\frac{1}{C_{k_n}^2 n}\right) \tag{46}$$

with

$$g(n)=\frac{2S\left(\frac{n}{n-1}\right)^2}{C_{k_n}^2 n}\sum_{j=1}^{k_n}C_j^2+\sum_{j=k_n+1}^{\infty}\rho_j^2. \tag{47}$$

Under Assumption A3, Eq. (46) implies

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

as we wanted to prove. □

Note that consistency of $\widehat{\rho}_{k_n}$ in the space $\mathcal{S}(H)$ directly follows from Eq. (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 $n\to\infty$,*

$$\left\|\rho-\widehat{\rho}_{k_n}\right\|_{\mathcal{S}(H)}\to^p 0, \tag{48}$$

where $\to^p$ denotes convergence in probability.

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

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

$$\|\mathcal{A}\|_{\mathcal{L}(H)}=\sup_{X\in H}\frac{\|\mathcal{A}(X)\|_H}{\|X\|_H}, \tag{49}$$

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

$$\|\mathcal{A}(X)\|_H\leq\|\mathcal{A}\|_{\mathcal{L}(H)}\|X\|_H. \tag{50}$$

In the following, we denote by

$$\widehat{X}_n=\widehat{\rho}_{k_n}(X_{n-1}) \tag{51}$$

the ARH(1) plug-in predictor of X_n, as an estimator of the conditional expectation $\mathrm{E}(X_n|X_{n-1})=\rho(X_{n-1})$. The following proposition provides the convergence in $\mathcal{L}^1_H(\Omega,\mathcal{A},P)$ of $\widehat{\rho}_{k_n}(X_{n-1})$ to $\rho(X_{n-1})$, and hence, the weak consistency of $\widehat{X}_n=\widehat{\rho}_{k_n}(X_{n-1})$ in H, for the approximation of $\rho(X_{n-1})$.

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

$$\lim_{n\to\infty} \mathrm{E}\left\{\left\|\rho\left(X_{n-1}\right) - \widehat{\rho}_{k_n}\left(X_{n-1}\right)\right\|_H\right\} = 0. \tag{52}$$

Specifically,

$$\mathrm{E}\left\{\left\|\rho\left(X_{n-1}\right) - \widehat{\rho}_{k_n}\left(X_{n-1}\right)\right\|_H\right\} \le h\left(n\right), \quad h\left(n\right) = \mathcal{O}\left(\frac{1}{C_{k_n}\sqrt{n}}\right), \; n \to \infty. \tag{53}$$

In particular,

$$\left\|\rho\left(X_{n-1}\right) - \widehat{\rho}_{k_n}\left(X_{n-1}\right)\right\|_H \to^p 0. \tag{54}$$

Proof. From (50) and Proposition 1, for n sufficiently large, the following almost surely inequality holds:

$$\left\|\rho\left(X_{n-1}\right) - \widehat{\rho}_{k_n}\left(X_{n-1}\right)\right\|_H \le \left\|\rho - \widehat{\rho}_{k_n}\right\|_{\mathcal{L}(H)} \|X_{n-1}\|_H. \tag{55}$$

Thus,

$$\mathrm{E}\left\{\left\|\rho\left(X_{n-1}\right) - \widehat{\rho}_{k_n}\left(X_{n-1}\right)\right\|_H\right\} \le \mathrm{E}\left\{\left\|\rho - \widehat{\rho}_{k_n}\right\|_{\mathcal{L}(H)} \|X_{n-1}\|_H\right\}. \tag{56}$$

From the Cauchy–Schwarz inequality, keeping in mind that, for a Hilbert–Schmidt operator $\mathcal{K}$, one always has $\|\mathcal{K}\|_{\mathcal{L}(H)} \le \|\mathcal{K}\|_{\mathcal{S}(H)}$, we have from Eq. (56),

$$\begin{aligned}\mathrm{E}\left\{\left\|\rho\left(X_{n-1}\right) - \widehat{\rho}_{k_n}\left(X_{n-1}\right)\right\|_H\right\} &\le \sqrt{\mathrm{E}\left(\left\|\rho - \widehat{\rho}_{k_n}\right\|^2_{\mathcal{L}(H)}\right)}\sqrt{\mathrm{E}\left(\|X_{n-1}\|_H^2\right)} \\ &\le \sqrt{\mathrm{E}\left(\left\|\rho - \widehat{\rho}_{k_n}\right\|^2_{\mathcal{S}(H)}\right)}\sqrt{\mathrm{E}\left(\|X_{n-1}\|_H^2\right)} \\ &= \sqrt{\mathrm{E}\left(\left\|\rho - \widehat{\rho}_{k_n}\right\|^2_{\mathcal{S}(H)}\right)}\sigma_X,\end{aligned} \tag{57}$$

where, as before,

$$\sigma_X^2 = \mathrm{E}\left(\|X_{n-1}\|_H^2\right) = \sum_{j=1}^{\infty} C_j < \infty,$$

for each $n \in \mathbb{Z}$; see Eq. (14). From Proposition 1 (see Eq. (36)),

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

then, the following upper bound is obtained in Eq. (57)

$$\mathrm{E}\left\{\left\|\rho\left(X_{n-1}\right) - \widehat{\rho}_{k_n}\left(X_{n-1}\right)\right\|_H\right\} \le h\left(n\right), \tag{58}$$

where $h\left(n\right) = \sigma_X\sqrt{g\left(n\right)}$, with $g\left(n\right)$ being given in (47). In particular, under Assumption A3,

$$\lim_{n\to\infty} \mathrm{E}\left\{\left\|\rho\left(X_{n-1}\right) - \widehat{\rho}_{k_n}\left(X_{n-1}\right)\right\|_H\right\} = 0, \tag{59}$$

which implies that

$$\left\|\rho\left(X_{n-1}\right) - \widehat{\rho}_{k_n}\left(X_{n-1}\right)\right\|_H \to^p 0, \quad n \to \infty. \quad \square \tag{60}$$

4. The Gaussian case

In this section, we prove that, in the Gaussian ARH(1) context, Assumptions A1 and A2 imply that Assumption A4 also holds.

From Eq. (16), for $n \ge 1$,

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

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

$$\Sigma_{n\times n} = \begin{pmatrix} 1 & \rho_j & 0 & \cdots & \cdots & 0 \\ \rho_j & 1 & \rho_j & 0 & \cdots & 0 \\ 0 & \rho_j & 1 & \rho_j & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & 0 & 0 & \rho_j & 1 \end{pmatrix}_{n\times n}. \tag{61}$$

It is well-known (see, e.g., [29]) that the variance of a quadratic form defined from a multivariate Gaussian vector $\mathbf{y} \sim \mathcal{N}(\boldsymbol{\mu}, \Lambda)$, and a symmetric matrix $\mathbf{Q}$ is given by

$$\operatorname{var}\left(\mathbf{y}^\top Q\mathbf{y}\right) = 2\operatorname{trace}\left(Q\Lambda Q\Lambda\right) + 4\boldsymbol{\mu}^\top Q\Lambda Q\boldsymbol{\mu}. \tag{62}$$

For each $j \geq 1$, applying Eq. (62), with $\mathbf{y} = \boldsymbol{\eta}_j$, $\Lambda = \Sigma_{n\times n}$, in Eq. (61), and $Q = I_{n\times n}$, the $n \times n$ identity matrix, keeping in mind $\mathrm{E}\left\{\eta_j(i)\eta_j(i+1)\right\} = \rho_j$, for every $i \in \mathbb{Z}$,

$$\begin{aligned} \operatorname{var}\left(\boldsymbol{\eta}_j^\top I_{n\times n}\boldsymbol{\eta}_j\right) &= \operatorname{var}\left(\sum_{i=0}^{n-1}\eta_j^2(i)\right) \\ &= 2\operatorname{trace}\left(\Sigma_{n\times n}\Sigma_{n\times n}\right) = 2\{n + 2(n-1)\rho_j^2\}. \end{aligned} \tag{63}$$

Furthermore, from Eq. (63), for each $j \geq 1$,

$$\operatorname{var}\left\{\frac{1}{n}\sum_{i=0}^{n-1}\eta_j^2(i)\right\} = \frac{2}{n^2}\left\{n + 2(n-1)\rho_j^2\right\} = \frac{2}{n} + 4\left(\frac{1}{n} - \frac{1}{n^2}\right)\rho_j^2. \tag{64}$$

We then obtain from Eq. (64),

$$\begin{aligned} \lim_{n\to\infty}\operatorname{var}\left\{\frac{1}{n}\sum_{i=0}^{n-1}\eta_j^2(i)\right\} &= \lim_{n\to\infty}\mathrm{E}\left[\left\{1 - \frac{1}{n}\sum_{i=0}^{n-1}\eta_j^2(i)\right\}^2\right] \\ &= \lim_{n\to\infty}\frac{2}{n} + 4\left(\frac{1}{n} - \frac{1}{n^2}\right)\rho_j^2 = 0. \end{aligned} \tag{65}$$

Eq. (65) leads to

$$\lim_{n\to\infty} n\operatorname{var}\left(\frac{\sum_{i=0}^{n-1}\eta_j^2(i)}{n}\right) = 2 + 4\rho_j^2. \tag{66}$$

Hence, for each $j \geq 1$, $K_{j,1}$, in Eq. (28) is given by $K_{j,1} = 2 + 4\rho_j^2$, and, from Eq. (64),

$$\operatorname{var}\left\{\frac{1}{n}\sum_{i=0}^{n-1}\eta_j^2(i)\right\} \leq 2 + 4\left(\frac{1}{n} - \frac{1}{n^2}\right)\rho_j^2 \leq 2 + 4\rho_j^2 \leq 6. \tag{67}$$

Thus, for every $j \geq 1$, $\widetilde{K}_{j,1}$ in Eq. (30) satisfies $\widetilde{K}_{j,1} \leq 6$.

Remark 8. Note that, from Lemma 1, for each $j \geq 1$ and all $i \in \mathbb{Z}$,

$$\mathrm{E}\left\{\widetilde{\eta}_j^4(i)\right\} = 3. \tag{68}$$

Thus, the assumption considered in Remark 3 holds, and for each $j \geq 1$, the AR(1) process $\eta_j = \left\{\eta_j(i) : i \in \mathbb{Z}\right\}$ is ergodic for all second-order moments, in the mean-square sense; see pp. 192–193 of [30].

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 $\boldsymbol{\eta}_j^\star = \big(\eta_j(0), \ldots, \eta_j(n-2)\big)^\top$ and $\boldsymbol{\eta}_j^{\star\star} = \big(\eta_j(1), \ldots, \eta_j(n-1)\big)^\top$ are Multivariate Normal distributed, with null mean vector, and covariance matrix

$$\widetilde{\Sigma}_{(n-1)\times(n-1)} = \begin{pmatrix} 1 & \rho_j & 0 & \cdots & \cdots & 0 \\ \rho_j & 1 & \rho_j & 0 & \cdots & 0 \\ 0 & \rho_j & 1 & \rho_j & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & 0 & 0 & \rho_j & 1 \end{pmatrix}_{(n-1)\times(n-1)}. \tag{69}$$

From Eq. (19), for each $j \geq 1$,

$$\mathrm{E}\left\{\sum_{i=0}^{n-2}\eta_j(i)\eta_j(i+1)\right\} = \sum_{i=0}^{n-2}\rho_j = (n-1)\rho_j = \mathrm{trace}\left[\mathrm{E}\left\{\boldsymbol{\eta}_j^{\star}\left(\boldsymbol{\eta}_j^{\star\star}\right)^{\top}\right\}\right], \tag{70}$$

where

$$\mathrm{E}\left\{\boldsymbol{\eta}_j^{\star}\left(\boldsymbol{\eta}_j^{\star\star}\right)^{\top}\right\} = \mathrm{E}\left\{\boldsymbol{\eta}_j^{\star}\otimes\boldsymbol{\eta}_j^{\star\star}\right\} = \rho_j I_{(n-1)\times(n-1)}, \tag{71}$$

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 $\boldsymbol{\eta}_j^{\star}$ and $\boldsymbol{\eta}_j^{\star\star}$. In the Gaussian case, keeping in mind that $\boldsymbol{\eta}_j^{\star} = \left(\eta_j(0),\ldots,\eta_j(n-2)\right)^{\top}$ and $\boldsymbol{\eta}_j^{\star\star} = \left(\eta_j(1),\ldots,\eta_j(n-1)\right)^{\top}$ are zero-mean multivariate Normal distributed vectors with covariance matrix $\widetilde{\Sigma}_{(n-1)\times(n-1)}$ given in Eq. (69), and having cross-covariance matrix (71), we can compute the variance of $\sum_{i=0}^{n-2}\eta_j(i)\eta_j(i+1)$, from (70) and (71), as follows. First,

$$\mathrm{var}\left\{\left(\boldsymbol{\eta}_j^{\star}\right)^{\top}I_{(n-1)\times(n-1)}\boldsymbol{\eta}_j^{\star\star}\right\} = \mathrm{E}\left\{\left(\boldsymbol{\eta}_j^{\star}\right)^{\top}I_{(n-1)\times(n-1)}\boldsymbol{\eta}_j^{\star\star}\left(\boldsymbol{\eta}_j^{\star}\right)^{\top}I_{(n-1)\times(n-1)}\boldsymbol{\eta}_j^{\star\star}\right\} - \left[\mathrm{E}\left\{\left(\boldsymbol{\eta}_j^{\star}\right)^{\top}I_{(n-1)\times(n-1)}\boldsymbol{\eta}_j^{\star\star}\right\}\right]^2.$$

This can be rewritten as

$$\sum_{i=0}^{n-2}\sum_{p=0}^{n-2}\mathrm{E}\left\{\eta_j(i)\eta_j(i+1)\eta_j(p)\eta_j(p+1)\right\} - \left[\mathrm{E}\left\{\left(\boldsymbol{\eta}_j^{\star}\right)^{\top}I_{(n-1)\times(n-1)}\boldsymbol{\eta}_j^{\star\star}\right\}\right]^2$$

which is equal to

$$\begin{aligned}&\sum_{i=0}^{n-2}\mathrm{E}\left\{\eta_j(i)\eta_j(i+1)\right\}\sum_{p=0}^{n-2}\mathrm{E}\left\{\eta_j(p)\eta_j(p+1)\right\} + \sum_{i=0}^{n-2}\sum_{p=0}^{n-2}\mathrm{E}\left\{\eta_j(i)\eta_j(p)\right\}\mathrm{E}\left\{\eta_j(i+1)\eta_j(p+1)\right\}\\ &\quad + \sum_{i=0}^{n-2}\sum_{p=0}^{n-2}\mathrm{E}\left\{\eta_j(i)\eta_j(p+1)\right\}\mathrm{E}\left\{\eta_j(i+1)\eta_j(p)\right\} - \left[\mathrm{E}\left\{\left(\boldsymbol{\eta}_j^{\star}\right)^{\top}I_{(n-1)\times(n-1)}\boldsymbol{\eta}_j^{\star\star}\right\}\right]^2.\end{aligned}$$

This then reduces to

$$\begin{aligned}&\left[\mathrm{trace}\{\mathrm{E}(\boldsymbol{\eta}_j^{\star}\otimes\boldsymbol{\eta}_j^{\star\star})\}\right]^2 + \mathrm{trace}(\widetilde{\Sigma}_{(n-1)\times(n-1)}\widetilde{\Sigma}_{(n-1)\times(n-1)})\\ &\quad + \mathrm{trace}\left[\mathrm{E}(\boldsymbol{\eta}_j^{\star}\otimes\boldsymbol{\eta}_j^{\star\star})\left\{\mathrm{E}(\boldsymbol{\eta}_j^{\star}\otimes\boldsymbol{\eta}_j^{\star\star})\right\}^{\top}\right] - \left[\mathrm{trace}\left\{\mathrm{E}(\boldsymbol{\eta}_j^{\star}\otimes\boldsymbol{\eta}_j^{\star\star})\right\}\right]^2,\end{aligned}$$

which is the same as

$$\mathrm{trace}(\widetilde{\Sigma}_{(n-1)\times(n-1)}\widetilde{\Sigma}_{(n-1)\times(n-1)}) + \mathrm{trace}[\mathrm{E}(\boldsymbol{\eta}_j^{\star}\otimes\boldsymbol{\eta}_j^{\star\star})\{\mathrm{E}(\boldsymbol{\eta}_j^{\star}\otimes\boldsymbol{\eta}_j^{\star\star})\}^{\top}] = (n-1) + 2(n-2)\rho_j^2 + (n-1)\rho_j^2, \tag{72}$$

where, from Eq. (71),

$$\mathrm{E}(\boldsymbol{\eta}_j^{\star}\otimes\boldsymbol{\eta}_j^{\star\star})\{\mathrm{E}(\boldsymbol{\eta}_j^{\star}\otimes\boldsymbol{\eta}_j^{\star\star})\}^{\top} = \begin{pmatrix}\rho_j^2 & 0 & \cdots & \cdots & 0\\ 0 & \rho_j^2 & 0 & \cdots & 0\\ \vdots & \ddots & \ddots & \vdots & \vdots\\ 0 & \cdots & \ddots & \ddots & \rho_j^2\end{pmatrix} = \rho_j^2 I_{(n-1)\times(n-1)}.$$

From Eq. (72),

$$\mathrm{var}\left\{\frac{1}{n-1}\sum_{i=0}^{n-2}\eta_j(i)\eta_j(i+1)\right\} = \frac{(n-1)+2(n-2)\rho_j^2+(n-1)\rho_j^2}{(n-1)^2}. \tag{73}$$

Therefore, for each $j \geq 1$,

$$\lim_{n\to\infty} n\,\mathrm{var}\left\{\frac{1}{n-1}\sum_{i=0}^{n-2}\eta_j(i)\eta_j(i+1)\right\} = 1+3\rho_j^2. \tag{74}$$

Thus, for each $j \geq 1$, $K_{j,2}$ in (29) is given by $K_{j,2} = 1+3\rho_j^2$. From Eq. (73),

$$\mathrm{var}\left\{\frac{1}{n-1}\sum_{i=0}^{n-2}\eta_j(i)\eta_j(i+1)\right\} \leq 1+3\rho_j^2 \leq 4. \tag{75}$$

Thus, for every $j \geq 1$, $\widetilde{K}_{j,2}$ in Eq. (31) satisfies $\widetilde{K}_{j,2} \leq 4$. Therefore, the constant S in Assumption A4 is such that $S \leq 6+4 = 10$.

5. Simulation study

A simulation study is undertaken to illustrate the behavior of the formulated component-wise estimator of the autocorrelation operator, and of its associated ARH(1) plug-in predictor for large sample sizes. The results are reported in Section 5.1. In Section 5.2, a comparative study is developed, from the implementation of the ARH(1) plug-in prediction techniques proposed in [4,7,9,28]. In the subsequent sections, we restrict our attention to the Gaussian case.

5.1. Behavior of $\widehat{\rho}$ and $\widehat{X}_n$ for large sample sizes

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

$$(-\Delta)_{(a,b)}(f)(x) = -\frac{d^2}{dx^2}f(x), \quad x \in (a,b) \subset \mathbb{R}, \\ f(a) = f(b) = 0. \tag{76}$$

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

$$(-\Delta)_{(a,b)}\phi_j(x) = \lambda_j\left((-\Delta)_{(a,b)}\right)\phi_j(x), \qquad \phi_j(a) = \phi_j(b) = 0. \tag{77}$$

For each $j \geq 1$ and $x \in [a, b]$, the solution to (77) is given by (see [26, p. 6]):

$$\phi_j(x) = \frac{2}{b-a}\sin\left(\frac{\pi j x}{b-a}\right), \qquad \lambda_j\left((-\Delta)_{(a,b)}\right) = \frac{\pi^2 j^2}{(b-a)^2}. \tag{78}$$

We consider here the operator C defined by

$$C = \left((-\Delta)_{(a,b)}\right)^{-2(1-\gamma_1)}, \quad \gamma_1 \in (0, 1/2). \tag{79}$$

From [16, 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

$$C_j = \left(\lambda_j\left((-\Delta)_{(a,b)}\right)\right)^{-2(1-\gamma_1)} = \left\{\frac{\pi^2 j^2}{(b-a)^2}\right\}^{-2(1-\gamma_1)}, \quad \gamma_1 \in (0, 1/2). \tag{80}$$

Additionally, considering

$$\rho = \left\{\frac{(-\Delta)_{(a,b)}}{\lambda_1\left((-\Delta)_{(a,b)}\right) - \epsilon}\right\}^{-(1-\gamma_2)}, \quad \gamma_2 \in (0, 1/2), \tag{81}$$

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

$$\rho_j^2 = \left\{\frac{\lambda_j\left((-\Delta)_{(a,b)}\right)}{\lambda_1\left((-\Delta)_{(a,b)}\right) - \epsilon}\right\}^{-2(1-\gamma_2)}, \quad \rho_j^2 \in (0,1), \ \gamma_2 \in (0, 1/2), \tag{82}$$

where, as before, $\{\lambda_j\left((-\Delta)_{(a,b)}\right) : j \geq 1\}$ are given in Eq. (78).

From (18), the eigenvalues $\{\sigma_j : j \geq 1\}$ of R_ε are defined, for each $j \geq 1$, as

$$\begin{aligned}\sigma_j^2 &= C_j\left(1 - \rho_j^2\right) \\ &= \left\{\lambda_j\left((-\Delta)_{(a,b)}\right)\right\}^{-2(1-\gamma_1)} - \frac{\left\{\lambda_j\left((-\Delta)_{(a,b)}\right)\right\}^{-2(2-\gamma_1-\gamma_2)}}{\left\{\lambda_1\left((-\Delta)_{(a,b)}\right) - \epsilon\right\}^{-2(1-\gamma_2)}}.\end{aligned} \tag{83}$$

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

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

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

$$k_n = n^{1/\alpha}, \tag{84}$$

Table 1
$\mathrm{EMSE}_{\widehat{\rho}_{k_n}}$ (here, $\mathrm{MSE}_{\widehat{\rho}_{k_{n,1}}}$), and $\mathrm{UB}(\mathrm{EMAE}_{\widehat{X}_n^{k_n}})$ (here, $\mathrm{UB}_{\widehat{X}_n^{k_{n,1}}}$) values, in Eqs. (87)–(89), for $\gamma_1 = 4/10$ and $\gamma_2 = 9/20$, considering the sample sizes $n_t = 15\,000 + 20\,000(t-1)$, $t = 1, \ldots, 20$, and the corresponding $k_{n,1}$ and $k_{n,2}$ values, for $\alpha_1 = 5$ and $\alpha_2 = 6$.

n	$k_{n,1}$	$\mathrm{MSE}_{\widehat{\rho}_{k_{n,1}}}$	$\mathrm{UB}_{\widehat{X}_n^{k_{n,1}}}$	$k_{n,2}$	$\mathrm{MSE}_{\widehat{\rho}_{k_{n,2}}}$	$\mathrm{UB}_{\widehat{X}_n^{k_{n,2}}}$
$n_1 = 15\,000$	6	$3.74\,(10)^{-4}$	$2.87\,(10)^{-2}$	5	$2.45\,(10)^{-4}$	$2.25\,(10)^{-2}$
$n_2 = 35\,000$	8	$2.15\,(10)^{-4}$	$2.21\,(10)^{-2}$	5	$1.35\,(10)^{-4}$	$1.71\,(10)^{-2}$
$n_3 = 55\,000$	8	$1.34\,(10)^{-4}$	$1.75\,(10)^{-2}$	6	$1.03\,(10)^{-4}$	$1.51\,(10)^{-2}$
$n_4 = 75\,000$	9	$1.09\,(10)^{-4}$	$1.57\,(10)^{-2}$	6	$7.55\,(10)^{-5}$	$1.29\,(10)^{-2}$
$n_5 = 95\,000$	9	$9.48\,(10)^{-5}$	$1.47\,(10)^{-2}$	6	$5.86\,(10)^{-5}$	$1.14\,(10)^{-2}$
$n_6 = 115\,000$	10	$8.31\,(10)^{-5}$	$1.39\,(10)^{-2}$	6	$5.16\,(10)^{-5}$	$1.07\,(10)^{-2}$
$n_7 = 135\,000$	10	$6.81\,(10)^{-5}$	$1.25\,(10)^{-2}$	7	$4.86\,(10)^{-5}$	$1.04\,(10)^{-2}$
$n_8 = 155\,000$	10	$6.37\,(10)^{-5}$	$1.21\,(10)^{-2}$	7	$3.88\,(10)^{-5}$	$9.66\,(10)^{-3}$
$n_9 = 175\,000$	11	$6.14\,(10)^{-5}$	$1.19\,(10)^{-2}$	7	$3.87\,(10)^{-5}$	$9.65\,(10)^{-3}$
$n_{10} = 195\,000$	11	$5.34\,(10)^{-5}$	$1.11\,(10)^{-2}$	7	$3.42\,(10)^{-5}$	$8.79\,(10)^{-3}$
$n_{11} = 215\,000$	11	$4.67\,(10)^{-5}$	$1.03\,(10)^{-2}$	7	$3.40\,(10)^{-5}$	$8.74\,(10)^{-3}$
$n_{12} = 235\,000$	11	$4.66\,(10)^{-5}$	$1.03\,(10)^{-2}$	7	$2.92\,(10)^{-5}$	$8.12\,(10)^{-3}$
$n_{13} = 255\,000$	12	$4.53\,(10)^{-5}$	$1.02\,(10)^{-2}$	7	$2.77\,(10)^{-5}$	$7.95\,(10)^{-3}$
$n_{14} = 275\,000$	12	$4.24\,(10)^{-5}$	$9.95\,(10)^{-3}$	8	$2.77\,(10)^{-5}$	$7.94\,(10)^{-3}$
$n_{15} = 295\,000$	12	$3.72\,(10)^{-5}$	$9.32\,(10)^{-3}$	8	$2.67\,(10)^{-5}$	$7.76\,(10)^{-3}$
$n_{16} = 315\,000$	12	$3.62\,(10)^{-5}$	$9.21\,(10)^{-3}$	8	$2.55\,(10)^{-5}$	$7.64\,(10)^{-3}$
$n_{17} = 335\,000$	12	$3.39\,(10)^{-5}$	$8.91\,(10)^{-3}$	8	$2.28\,(10)^{-5}$	$7.04\,(10)^{-3}$
$n_{18} = 355\,000$	12	$3.34\,(10)^{-5}$	$8.86\,(10)^{-3}$	8	$2.20\,(10)^{-5}$	$7.04\,(10)^{-3}$
$n_{18} = 375\,000$	13	$3.34\,(10)^{-5}$	$8.86\,(10)^{-3}$	8	$2.04\,(10)^{-5}$	$6.84\,(10)^{-3}$
$n_{20} = 395\,000$	13	$3.12\,(10)^{-5}$	$8.56\,(10)^{-3}$	8	$1.92\,(10)^{-5}$	$6.65\,(10)^{-3}$

for a suitable $\alpha > 0$, which, in particular, allows verification of (27). From Eq. (80), one has, for $\gamma_1 \in (0, 1/2)$,

$$\sqrt{n}\, C_{k_n} = \sqrt{n}\left\{\lambda_{k_n}\left(-\Delta_{(a,b)}\right)\right\}^{-2(1-\gamma_1)} = \sqrt{n}\left(\frac{\pi k_n}{b-a}\right)^{-4(1-\gamma_1)}. \tag{85}$$

From Eq. (84), Assumption A3 is then satisfied if

$$1/2 - \frac{4(1-\gamma_1)}{\alpha} > 0, \quad \text{i.e., if } \alpha > 8(1-\gamma_1) > 4, \tag{86}$$

since $\gamma_1 \in (0, 1/2)$. Fix $\gamma_1 = 4/10$ and $\gamma_2 = 9/20$. Then, from Eq. (86), $\alpha > 48/10$. In particular, the values $\alpha_1 = 5$ and $\alpha_2 = 6$ have been tested, in Table 1, for $H = L^2((a, b))$, and $(a, b) = (0, 4)$, where $L^2((a, b))$ denotes the space of square integrable functions on (a, b).

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

$$\mathrm{EMSE}_{\widehat{\rho}_{k_n}} = \frac{1}{N}\sum_{w=1}^{N}\sum_{j=1}^{k_n}\left(\rho_j - \widehat{\rho}_{n,j}^{w}\right)^2, \tag{87}$$

$$\widehat{\rho}_{n,j}^{w} = \frac{\widehat{D}_{n,j}^{w}}{\widehat{C}_{n,j}^{w}} = \frac{\frac{1}{n-1}\sum_{i=0}^{n-2} X_{i,j}^{w} X_{i+1,j}^{w}}{\frac{1}{n}\sum_{i=0}^{n-1}\left(X_{i,j}^{w}\right)^2}, \tag{88}$$

where N denotes the number of simulations, and for each $j = 1, \ldots, k_n$, $\widehat{\rho}_{n,j}^{w}$ represents the estimator of ρ_j, based on the wth 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 $\widehat{X}_n = \widehat{\rho}_{k_n}(X_{n-1})$, we compute the empirical version $\mathrm{UB}(\mathrm{EMAE}_{\widehat{X}_n^{k_n}})$ of the derived upper bound (57), which, for each $n \in \mathbb{Z}$, is given by

$$\mathrm{UB}(\mathrm{EMAE}_{\widehat{X}_n^{k_n}}) = \sqrt{\frac{1}{N}\sum_{w=1}^{N}\sum_{j=1}^{k_n}\left(\rho_j - \widehat{\rho}_{n,j}^{w}\right)^2 \widehat{\mathrm{E}\left\{\left\|X_{n-1}^{w}\right\|_H^2\right\}}}. \tag{89}$$

From $N = 700$ realizations, for each one of the elements of the sequence of sample sizes $n_t = 15\,000 + 20\,000(t-1)$, $t = 1, \ldots, 20$, the $\mathrm{EMSE}_{\widehat{\rho}_{k_n}}$ and $\mathrm{UB}(\mathrm{EMAE}_{\widehat{X}_n^{k_n}})$ values, for $\alpha = 5$ and $\alpha = 6$, are displayed in Table 1, where the abbreviated notations $\mathrm{MSE}_{\widehat{\rho}_{k_{n,1}}}$, for $\mathrm{EMSE}_{\widehat{\rho}_{k_n}}$, and $\mathrm{UB}_{\widehat{X}_n^{k_{n,1}}}$, for $\mathrm{UB}(\mathrm{EMAE}_{\widehat{X}_n^{k_n}})$, are used; see also Figs. 1 and 2.

In this paper, a one-parameter model of k_n is selected depending on parameter α. In Example 2, in p. 286 in [28], in the same spirit, for an equivalent spectral class of operators C, a three-parameter model is established for k_n to ensure

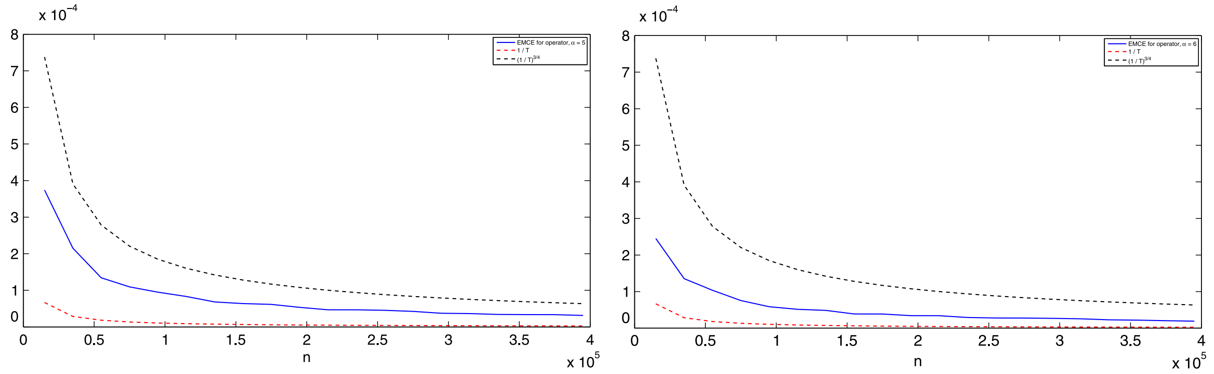

Fig. 1. $\text{EMSE}_{\widehat{\rho}_{k_n}}$ values (blue line), in (87)–(88), for $\gamma_1 = 4/10$ and $\gamma_2 = 9/20$, considering the sample sizes $T = n_t = 15\,000 + 20\,000(t-1)$, $t = 1, \ldots, 20$, and the corresponding $k_{n,1}$ and $k_{n,2}$ values, for $\alpha_1 = 5$ (left-hand side) and $\alpha_2 = 6$ (right-hand side), against $(1/T)^{3/4} = (1/n_t)^{3/4}$ (black dot line) and $1/T = (1/n_t)$ (red dot line). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

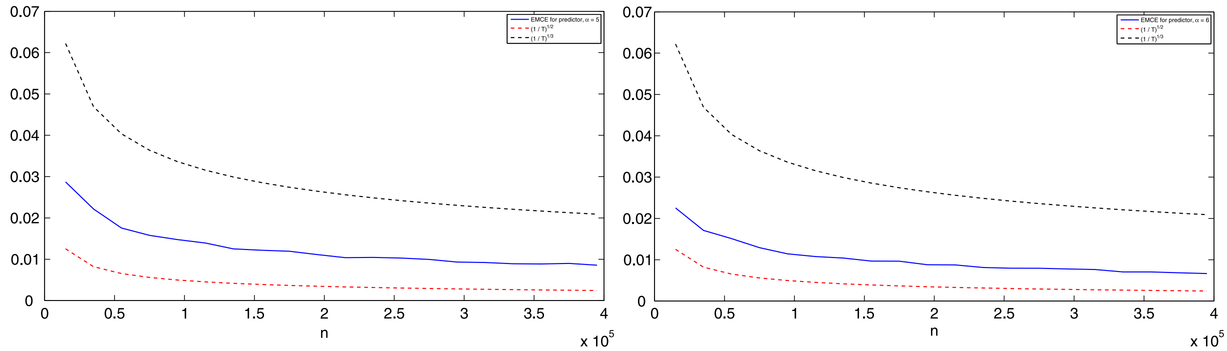

Fig. 2. $\text{UB}(\text{EMAE}_{\widehat{X}_n^{k_n}})$ values (blue line), in (89), for $\gamma_1 = 4/10$ and $\gamma_2 = 9/20$, considering the sample sizes $T = n_t = 15\,000 + 20\,000(t-1)$, $t = 1, \ldots, 20$, and the corresponding $k_{n,1}$ and $k_{n,2}$ values, for $\alpha_1 = 5$ (left-hand side) and $\alpha_2 = 6$ (right-hand side), against $(1/T)^{1/2} = (1/n_t)^{1/2}$ (red dot line) and $(1/T)^{1/3} = (1/n_t)^{1/3}$ (black dot line). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

convergence in quadratic mean in the space $\mathcal{L}(H)$ of the component-wise estimator of ρ constructed from the known eigenvectors of C.

The numerical results displayed in Table 1 and Figs. 1–2 illustrate the fact that the proposed component-wise estimator $\widehat{\rho}_{k_n}$ presents a speed of convergence to ρ, in quadratic mean in $S(H)$, faster than $n^{-1/3}$, which corresponds to the optimal case for the component-wise estimator of ρ proposed in [28], in the case of known eigenvectors of C; see, in particular, Theorem 1, Remark 2 and Example 2 in [28].

For larger values of the parameters γ_1 than 0.4, and α than 6, a faster speed of convergence of $\widehat{\rho}_{k_n}$ to ρ, in quadratic mean in the space $S(H)$, will be obtained. However, larger sample sizes are required for larger values of α, in order to estimate a given number of coefficients of ρ. A more detailed discussion about the comparison of the rates of convergence of the ARH(1) plug-in predictors proposed in [4,7,9,28] can be found in the next section.

5.2. *A comparative study*

In this section, the performance of our approach is compared with those given in [4,7,9,28], including the case of unknown eigenvectors of C. In the last case, our approach and the approaches presented in [9,28] are implemented in terms of the empirical eigenvectors.

5.2.1. *Theoretical-eigenvector-based component-wise estimators*

Let us first compare the performance of our ARH(1) plug-in predictor, defined in (51), and the ones formulated in [9,28], 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 Eqs. (80) and (82), for different rates of convergence to zero of parameters C_j and ρ_j^2, $j \geq 1$, with both sequences being summable sequences. Since we restrict our attention to the Gaussian case, Conditions A_1, B_1 and C_1, formulated in pp. 211–212 in [9], are satisfied by the generated ARH(1) process. Similarly, Conditions H_1–H_3 in p. 283 of [28] are satisfied as well.

Table 2
Truncated empirical values of $\mathrm{E}\left\{\|\rho(X_{n-1}) - \widehat{\rho}_{k_n}(X_{n-1})\|_H\right\}$, for $\widehat{\rho}_{k_n}$ given in Eqs. (25)–(26) (third column), in Eqs. (90)–(91) (fourth column), and in Eqs. (92)–(93) (fifth column), for $\delta_1 = 2.4$ and $\delta_2 = 1.1$, considering the sample sizes $n_t = 15\,000 + 20\,000(t-1)$, $t = 1, \dots, 20$, and the corresponding $k_n = n^{1/\alpha}$ values, for $\alpha = 6$.

n	k_n	Our Approach	Bosq [9]	Guillas [28]
$n_1 = 15\,000$	5	$2.25\,(10)^{-2}$	$2.57\,(10)^{-2}$	$2.36\,(10)^{-2}$
$n_2 = 35\,000$	5	$1.71\,(10)^{-2}$	$1.72\,(10)^{-2}$	$1.84\,(10)^{-2}$
$n_3 = 55\,000$	6	$1.51\,(10)^{-2}$	$1.65\,(10)^{-2}$	$1.53\,(10)^{-2}$
$n_4 = 75\,000$	6	$1.29\,(10)^{-2}$	$1.46\,(10)^{-2}$	$1.37\,(10)^{-2}$
$n_5 = 95\,000$	6	$1.14\,(10)^{-2}$	$1.20\,(10)^{-2}$	$1.16\,(10)^{-2}$
$n_6 = 115\,000$	6	$1.07\,(10)^{-2}$	$1.10\,(10)^{-2}$	$1.11\,(10)^{-2}$
$n_7 = 135\,000$	7	$1.04\,(10)^{-2}$	$1.06\,(10)^{-2}$	$1.07\,(10)^{-2}$
$n_8 = 155\,000$	7	$9.66\,(10)^{-3}$	$9.91\,(10)^{-3}$	$1.01\,(10)^{-2}$
$n_9 = 175\,000$	7	$9.65\,(10)^{-3}$	$9.79\,(10)^{-3}$	$9.68\,(10)^{-3}$
$n_{10} = 195\,000$	7	$8.79\,(10)^{-3}$	$9.12\,(10)^{-3}$	$8.93\,(10)^{-3}$
$n_{11} = 215\,000$	7	$8.74\,(10)^{-3}$	$8.79\,(10)^{-3}$	$8.83\,(10)^{-3}$
$n_{12} = 235\,000$	7	$8.12\,(10)^{-3}$	$8.69\,(10)^{-3}$	$8.75\,(10)^{-3}$
$n_{13} = 255\,000$	7	$7.95\,(10)^{-3}$	$8.53\,(10)^{-3}$	$8.73\,(10)^{-3}$
$n_{14} = 275\,000$	8	$7.94\,(10)^{-3}$	$8.52\,(10)^{-3}$	$8.58\,(10)^{-3}$
$n_{15} = 295\,000$	8	$7.76\,(10)^{-3}$	$8.49\,(10)^{-3}$	$8.36\,(10)^{-3}$
$n_{16} = 315\,000$	8	$7.64\,(10)^{-3}$	$7.88\,(10)^{-3}$	$8.13\,(10)^{-3}$
$n_{17} = 335\,000$	8	$7.04\,(10)^{-3}$	$7.24\,(10)^{-3}$	$7.59\,(10)^{-3}$
$n_{18} = 355\,000$	8	$7.04\,(10)^{-3}$	$7.23\,(10)^{-3}$	$6.92\,(10)^{-3}$
$n_{19} = 375\,000$	8	$6.84\,(10)^{-3}$	$6.89\,(10)^{-3}$	$6.90\,(10)^{-3}$
$n_{20} = 395\,000$	8	$6.65\,(10)^{-3}$	$6.67\,(10)^{-3}$	$6.85\,(10)^{-3}$

In Section 8.2 of [9] the following estimator of ρ is proposed

$$\widehat{\rho}_n(x) = \left(\Pi^{k_n} D_n \widehat{C}_n^{-1} \Pi^{k_n}\right)(x) = \sum_{\ell=1}^{k_n} \widehat{\rho}_{n,\ell}(x)\phi_\ell, \quad x \in H, \tag{90}$$

$$\widehat{\rho}_{n,\ell}(x) = \frac{1}{n-1} \sum_{i=0}^{n-2} \sum_{j=1}^{k_n} \frac{1}{\widehat{C}_{n,j}} \langle \phi_j, x\rangle_H X_{i,j} X_{i+1,\ell}, \tag{91}$$

in the finite-dimensional subspace $H_{k_n} = \operatorname{span}\left(\phi_1, \dots, \phi_{k_n}\right)$ of H, where Π^{k_n} is the orthogonal projector over H_{k_n}, and, as before, $X_{i,j} = \left\langle X_i, \phi_j\right\rangle_H$, with $\widehat{C}_{n,j}$ being defined in (24), for each $j \geq 1$.

A modified estimator of ρ is studied in [28], given by

$$\widehat{\rho}_{n,a}(x) = \left(\Pi^{k_n} D_n \widehat{C}_{n,a}^{-1} \Pi^{k_n}\right)(x) = \sum_{\ell=1}^{k_n} \widehat{\rho}_{n,a,\ell}(x)\phi_\ell, \quad x \in H, \tag{92}$$

$$\widehat{\rho}_{n,a,\ell}(x) = \frac{1}{n-1} \sum_{i=1}^{n-1} \sum_{j=1}^{k_n} \frac{1}{\max(\widehat{C}_{n,j}, a_n)} \langle \phi_j, x\rangle_H X_{i,j} X_{i+1,l}, \tag{93}$$

where $\widehat{C}_{n,a}^{-1}(x) = \sum_{j=1}^{k_n} 1/\max(\widehat{C}_{n,j}, a_n)\langle \phi_j, x\rangle_H \phi_j$ (a.s.). Here, the sequence $\{a_n : n \in \mathbb{N}\}$ is such that (see Theorem 1 in [28])

$$\alpha \frac{C_{k_n}^{\gamma}}{n^{\varepsilon}} \leq a_n \leq \beta \lambda_{k_n}, \quad \alpha > 0,\ 0 < \beta < 1,\ \varepsilon < 1/2,\ \gamma \geq 1. \tag{94}$$

Tables 2 and 3 display the truncated, for two different k_n rules, empirical values of $\mathrm{E}\left\{\|\rho(X_{n-1}) - \widehat{\rho}_{k_n}(X_{n-1})\|_H\right\}$, based on $N = 700$ generations of each one of the functional samples considered with size $n_t = 15\,000 + 20\,000(t-1)$, $t = 1, \dots, 20$, when $C_j = b_C j^{-\delta_1}$, $b_C > 0$, and $\rho_j = b_\rho j^{-\delta_2}$, $b_\rho > 0$. Specifically, $\widehat{\rho}_{k_n}$ is computed from Eqs. (25) and (26) (see third column), $\widehat{\rho}_{k_n} = \widehat{\rho}_n$, with $\widehat{\rho}_n$ being given in Eqs. (90)–(91) (see fourth column), and $\widehat{\rho}_{k_n} = \widehat{\rho}_{n,a}$, with $\widehat{\rho}_{n,a}$ being defined in (92)–(93) (see fifth column).

In Table 2, $\delta_1 = 2.4$, $\delta_2 = 1.1$, and $k_n = n^{1/\alpha}$, for $\alpha = 6$, according to Assumption A3, which is also considered in p. 217 of [9] to ensure weak consistency of the proposed estimator of ρ. In Table 3, the same empirical values are displayed for $\delta_1 = 61/60$, $\delta_2 = 1.1$, and k_n is selected according to Example 2, in p. 286 of [28]. Thus, in Table 3,

$$k_n = n^{\frac{1-2\epsilon}{\delta_1(4+2\gamma)}}, \quad \gamma \geq 1,\ \epsilon < 1/2. \tag{95}$$

In particular, we have chosen $\gamma = 2$, and $\epsilon = 0.04\delta_1$. Note that, from Theorem 1 and Remark 1 in [28], for the choice made of k_n in Table 3, convergence to ρ, in quadratic mean in the space $\mathcal{L}(H)$, holds for $\widehat{\rho}_{n,a}$ given in Eqs. (92)–(93).

Table 3

Truncated empirical values of $\mathrm{E}\left\{\|\rho(X_{n-1})-\widehat{\rho}_{k_n}(X_{n-1})\|_H\right\}$, for $\widehat{\rho}_{k_n}$ given in Eqs. (25)–(26) (third column), in Eqs. (90)–(91) (fourth column), and in Eqs. (92)–(93) (fifth column), for $\delta_1 = 61/60$ and $\delta_2 = 1.1$, considering the sample sizes $n_t = 15\,000+20\,000(t-1)$, $t = 1,\dots,20$, and the corresponding k_n values given in (95).

n	k_n	Our Approach	Bosq [9]	Guillas [28]
$n_1 = 15\,000$	2	$9.91\,(10)^{-3}$	$1.39\,(10)^{-2}$	$1.26\,(10)^{-2}$
$n_2 = 35\,000$	3	$8.78\,(10)^{-3}$	$1.34\,(10)^{-2}$	$1.24\,(10)^{-2}$
$n_3 = 55\,000$	3	$7.89\,(10)^{-3}$	$1.15\,(10)^{-2}$	$1.14\,(10)^{-2}$
$n_4 = 75\,000$	3	$6.49\,(10)^{-3}$	$1.01\,(10)^{-2}$	$8.58\,(10)^{-3}$
$n_5 = 95\,000$	3	$6.36\,(10)^{-3}$	$9.09\,(10)^{-3}$	$8.29\,(10)^{-3}$
$n_6 = 115\,000$	3	$6.14\,(10)^{-3}$	$7.65\,(10)^{-3}$	$7.26\,(10)^{-3}$
$n_7 = 135\,000$	3	$5.91\,(10)^{-3}$	$7.03\,(10)^{-3}$	$6.69\,(10)^{-3}$
$n_8 = 155\,000$	3	$5.73\,(10)^{-3}$	$6.77\,(10)^{-3}$	$6.54\,(10)^{-3}$
$n_9 = 175\,000$	3	$5.44\,(10)^{-3}$	$6.74\,(10)^{-3}$	$6.16\,(10)^{-3}$
$n_{10} = 195\,000$	3	$5.10\,(10)^{-3}$	$6.69\,(10)^{-3}$	$5.97\,(10)^{-3}$
$n_{11} = 215\,000$	4	$5.01\,(10)^{-3}$	$6.48\,(10)^{-3}$	$5.94\,(10)^{-3}$
$n_{12} = 235\,000$	4	$4.85\,(10)^{-3}$	$6.45\,(10)^{-3}$	$5.83\,(10)^{-3}$
$n_{13} = 255\,000$	4	$4.17\,(10)^{-3}$	$6.17\,(10)^{-3}$	$5.68\,(10)^{-3}$
$n_{14} = 275\,000$	4	$4.64\,(10)^{-3}$	$5.99\,(10)^{-3}$	$5.60\,(10)^{-3}$
$n_{15} = 295\,000$	4	$4.55\,(10)^{-3}$	$5.94\,(10)^{-3}$	$5.58\,(10)^{-3}$
$n_{16} = 315\,000$	4	$4.48\,(10)^{-3}$	$5.69\,(10)^{-3}$	$5.50\,(10)^{-3}$
$n_{17} = 335\,000$	4	$4.38\,(10)^{-3}$	$5.58\,(10)^{-3}$	$5.44\,(10)^{-3}$
$n_{18} = 355\,000$	4	$4.16\,(10)^{-3}$	$5.45\,(10)^{-3}$	$5.42\,(10)^{-3}$
$n_{19} = 375\,000$	4	$3.91\,(10)^{-3}$	$5.34\,(10)^{-3}$	$5.32\,(10)^{-3}$
$n_{20} = 395\,000$	4	$3.86\,(10)^{-3}$	$5.29\,(10)^{-3}$	$5.26\,(10)^{-3}$

One can observe in Table 2 a similar performance of the three methods compared with the truncation order k_n satisfying Assumption A3, with slightly worse results being obtained from the estimator defined in Eqs. (92)–(93), especially, for the sample size $n_8 = 155\,000$. Furthermore, in Table 3, a better performance of our approach is observed for the smallest sample sizes (from $n_1 = 15\,000$ to $n_4 = 75\,000$). For the remaining largest sample sizes, only slight differences are observed, with, again, a better performance of our approach, very close to the other two approaches presented in [9,28].

5.2.2. *Empirical-eigenvector-based component-wise estimators*

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

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

where $\left\{C_{n,j} : j \geq 1\right\}$ denotes the system of eigenvalues associated with the system of empirical eigenvectors $\left\{\phi_{n,j} : j \geq 1\right\}$. We then consider the following estimators for comparison purposes

$$\widetilde{\rho}_{n,j} = \frac{\frac{1}{n-1}\sum_{i=0}^{n-2}\widetilde{X}_{i,j}\widetilde{X}_{i+1,j}}{\frac{1}{n}\sum_{i=0}^{n-1}(\widetilde{X}_{i,j})^2}, \qquad \widetilde{\rho}_{k_n} = \sum_{j=1}^{k_n}\widetilde{\rho}_{n,j}\phi_{n,j}\otimes\phi_{n,j} \tag{96}$$

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

$$\widetilde{\rho}_{n,l}(x) = \frac{1}{n-1}\sum_{i=0}^{n-2}\sum_{j=1}^{k_n}\frac{1}{C_{n,j}}\langle\phi_{n,j}, x\rangle_H\widetilde{X}_{i,j}\widetilde{X}_{i+1,l} \tag{97}$$

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

$$\widetilde{\rho}_{n,a,l}(x) = \frac{1}{n-1}\sum_{i=0}^{n-2}\sum_{j=1}^{k_n}\frac{1}{\max\left(C_{n,j}, a_n\right)}\langle\phi_{n,j}, x\rangle_H\widetilde{X}_{i,j}\widetilde{X}_{i+1,l}, \tag{98}$$

where, for $i\in\mathbb{Z}$, and $j \geq 1$, $\widetilde{X}_{i,j} = \left\langle X_i, \phi_{n,j}\right\rangle_H$, $\widetilde{\Pi}^{k_n}$ denotes the orthogonal projector into the space $\widetilde{H}_{k_n} = \mathrm{span}\left(\phi_{n,1}, \dots, \phi_{n,k_n}\right)$.

Table 4
Truncated empirical values of $\mathrm{E}\left\{\|\rho(X_{n-1}) - \widetilde{\rho}_{k_n}(X_{n-1})\|_H\right\}$, for $\widetilde{\rho}_{k_n}$ defined in Eq. (96) (third column), for $\widetilde{\rho}_{k_n} = \widetilde{\rho}_n$ given in Eq. (97) (fourth column), and for $\widetilde{\rho}_{k_n} = \widetilde{\rho}_{n,a}$ in Eq. (98) (fifth column), for $\delta_1 = 2.4$ and $\delta_2 = 1.1$, considering the sample sizes $n_t = 15\,000 + 20\,000(t-1),\ t = 1, \ldots, 20$, and $k_n = \ln(n)$.

n	k_n	Our Approach	Bosq [9]	Guillas [28]
$n_1 = 15\,000$	9	$8.42\,(10)^{-2}$	1.0614	1.0353
$n_2 = 35\,000$	10	$5.51\,(10)^{-2}$	1.0186	1.0052
$n_3 = 55\,000$	10	$4.75\,(10)^{-2}$	1.0174	0.9986
$n_4 = 75\,000$	11	$4.43\,(10)^{-2}$	1.0153	0.9951
$n_5 = 95\,000$	11	$3.68\,(10)^{-2}$	1.0127	0.9883
$n_6 = 115\,000$	11	$3.51\,(10)^{-2}$	1.0113	0.9627
$n_7 = 135\,000$	11	$3.23\,(10)^{-2}$	1.0081	0.9247
$n_8 = 155\,000$	11	$2.95\,(10)^{-2}$	1.0066	0.9119
$n_9 = 175\,000$	12	$2.94\,(10)^{-2}$	1.0057	0.9113
$n_{10} = 195\,000$	12	$2.80\,(10)^{-2}$	0.9948	0.8912
$n_{11} = 215\,000$	12	$2.71\,(10)^{-2}$	0.9017	0.8615
$n_{12} = 235\,000$	12	$2.59\,(10)^{-2}$	0.8896	0.8201
$n_{13} = 255\,000$	12	$2.58\,(10)^{-2}$	0.8783	0.8004
$n_{14} = 275\,000$	12	$2.35\,(10)^{-2}$	0.8719	0.7832
$n_{15} = 295\,000$	12	$2.28\,(10)^{-2}$	0.8602	0.7780
$n_{16} = 315\,000$	12	$2.27\,(10)^{-2}$	0.8424	0.7469
$n_{17} = 335\,000$	12	$2.16\,(10)^{-2}$	0.8217	0.7140
$n_{18} = 355\,000$	12	$2.14\,(10)^{-2}$	0.8001	0.7066
$n_{19} = 375\,000$	12	$2.09\,(10)^{-2}$	0.7778	0.6872
$n_{20} = 395\,000$	12	$2.06\,(10)^{-2}$	0.7693	0.6621

The Gaussian ARH(1) process is generated under Assumptions A1–A2, as well as C_1' in p. 218 in [9]. Note that conditions A_1 and B_1' in [9] already hold. Moreover, as given in Theorem 8.8 and Example 8.6, in p. 221 of [9], for $C_j = b_C j^{-\delta_1}$, ($b_C > 0$, $\delta_1 > 1$), with, in particular, $\delta_1 = 2.4$, and for $\rho_j = b_\rho j^{-\delta_2},\ b_\rho > 0$, with $\delta_2 = 1.1$, the estimator $\widetilde{\rho}_n$ converges almost surely to ρ under the condition

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

where

$$u_j = 2\sqrt{2}\max\left\{(C_{j-1} - C_j)^{-1}, (C_j - C_{j+1})^{-1}\right\}, \quad j \geq 2.$$

In Table 4, $k_n \simeq \ln(n)$ was tested; see Example 8.6, in p. 221 of [9]. A better performance of our estimator (96) in comparison with estimator (97), formulated in [9], and estimator (98), formulated in [28] – see Example 4 and Remark 4, in p. 291 of [28] – is observed in Table 4. Note that, in particular, in Example 4 and Remark 4, in p. 291 of [28], smaller values of k_n than $\ln(n)$ are required for a given sample size n, to ensure convergence in quadratic mean, and, in particular, weak-consistency.

However, considering a smaller discretization step size $\Delta_t = 0.015$ than in Table 4, where $\Delta_t = 0.08$, and for $k_n = n^{1/6}$, (i.e., $\alpha = 6$), we obtain in Table 5, for the same parameter values $\delta_1 = 2.4$ and $\delta_2 = 1.1$, better results than in Table 4, since a smaller number of coefficients of ρ (parameters) must be estimated in Table 5, from a richer sample information (coming from the smaller discretization step size considered). One can also observe in Table 5 a similar performance of the three approaches studied.

In Table 6, the value $k_n = \lceil e' n^{1/(8\delta_1+2)}\rceil$, $e' = 17/10$, proposed in Example 4 and Remark 4, in p. 291 of [28], is considered to compute the truncated empirical values of $\mathrm{E}\left\{\|\rho(X_{n-1}) - \widetilde{\rho}_{k_n}(X_{n-1})\|_H\right\}$, for $\widetilde{\rho}_{k_n}$ defined in Eq. (96) (third column), for $\widetilde{\rho}_{k_n} = \widetilde{\rho}_n$ given in Eq. (97) (fourth column), and for $\widetilde{\rho}_{k_n} = \widetilde{\rho}_{n,a}$ in Eq. (98) (fifth column). A similar performance of the three approaches is observed, with the exception of $n_{20} = 395\,000$, where the approach presented in [28] displays a slightly better performance.

5.2.3. *Kernel-based nonparametric and penalized estimation*

In practice, curves are observed in discrete times, and should be approximated by smooth functions. In [7], the following optimization problem is considered:

$$\widehat{X}_i = \operatorname{argmin}\|L\widehat{X}_i\|_{L^2}^2, \qquad \widehat{X}_i(t_j) = X_i(t_j), \quad j = 1, \ldots, p,\ i = 0, \ldots, n-1, \tag{99}$$

Table 5
Truncated empirical values of $\mathrm{E}\left\{\|\rho(X_{n-1})-\widetilde{\rho}_{k_n}(X_{n-1})\|_H\right\}$, for $\widetilde{\rho}_{k_n}$ defined in Eq. (96) (third column), for $\widetilde{\rho}_{k_n}=\widetilde{\rho}_n$ given in Eq. (97) (fourth column), and for $\widetilde{\rho}_{k_n}=\widetilde{\rho}_{n,a}$ in Eq. (98) (fifth column), for $\delta_1=2.4$ and $\delta_2=1.1$, considering the sample sizes $n_t=15\,000+20\,000(t-1),\ t=1,\dots,20$, and $k_n=n^{1/6}$.

n	k_n	Our Approach	Bosq [9]	Guillas [28]
$n_1=15\,000$	4	$9.88\,(10)^{-2}$	$9.25\,(10)^{-2}$	0.1059
$n_2=35\,000$	5	$9.52\,(10)^{-2}$	$9.07\,(10)^{-2}$	$9.86\,(10)^{-2}$
$n_3=55\,000$	6	$9.12\,(10)^{-2}$	$8.92\,(10)^{-2}$	$9.39\,(10)^{-2}$
$n_4=75\,000$	6	$8.48\,(10)^{-2}$	$8.64\,(10)^{-2}$	$8.98\,(10)^{-2}$
$n_5=95\,000$	6	$7.61\,(10)^{-2}$	$8.30\,(10)^{-2}$	$8.46\,(10)^{-2}$
$n_6=115\,000$	6	$7.05\,(10)^{-2}$	$7.96\,(10)^{-2}$	$8.04\,(10)^{-2}$
$n_7=135\,000$	7	$6.99\,(10)^{-2}$	$7.84\,(10)^{-2}$	$7.82\,(10)^{-2}$
$n_8=155\,000$	7	$6.70\,(10)^{-2}$	$7.45\,(10)^{-2}$	$7.40\,(10)^{-2}$
$n_9=175\,000$	7	$6.49\,(10)^{-2}$	$7.03\,(10)^{-2}$	$7.07\,(10)^{-2}$
$n_{10}=195\,000$	7	$5.88\,(10)^{-2}$	$6.74\,(10)^{-2}$	$6.80\,(10)^{-2}$
$n_{11}=215\,000$	7	$5.63\,(10)^{-2}$	$6.46\,(10)^{-2}$	$6.57\,(10)^{-2}$
$n_{12}=235\,000$	7	$5.30\,(10)^{-2}$	$6.28\,(10)^{-2}$	$6.37\,(10)^{-2}$
$n_{13}=255\,000$	7	$5.05\,(10)^{-2}$	$6.19\,(10)^{-2}$	$6.24\,(10)^{-2}$
$n_{14}=275\,000$	8	$4.88\,(10)^{-2}$	$5.99\,(10)^{-2}$	$6.15\,(10)^{-2}$
$n_{15}=295\,000$	8	$4.58\,(10)^{-2}$	$5.74\,(10)^{-2}$	$6.04\,(10)^{-2}$
$n_{16}=315\,000$	8	$4.24\,(10)^{-2}$	$5.52\,(10)^{-2}$	$5.93\,(10)^{-2}$
$n_{17}=335\,000$	8	$3.86\,(10)^{-2}$	$5.24\,(10)^{-2}$	$5.70\,(10)^{-2}$
$n_{18}=355\,000$	8	$3.70\,(10)^{-2}$	$5.02\,(10)^{-2}$	$5.53\,(10)^{-2}$
$n_{19}=375\,000$	8	$3.55\,(10)^{-2}$	$4.88\,(10)^{-2}$	$5.36\,(10)^{-2}$
$n_{20}=395\,000$	8	$3.46\,(10)^{-2}$	$4.70\,(10)^{-2}$	$5.23\,(10)^{-2}$

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 (99), as follows:

$$\widehat{X}_n^{h_n}=\widehat{\rho}_{h_n}(X_{n-1}),\qquad \widehat{\rho}_{h_n}(x)=\frac{\sum_{i=0}^{n-2}\widehat{X}_{i+1}K\left(\frac{\|\widehat{X}_i-x\|_{L^2}^2}{h_n}\right)}{\sum_{i=0}^{n-2}K\left(\frac{\|\widehat{X}_i-x\|_{L^2}^2}{h_n}\right)},\tag{100}$$

where K is the usual Gaussian kernel, and $\|\widehat{X}_i-x\|_{L^2}^2=\int\{\widehat{X}_i(t)-x(t)\}^2dt$, for $i=0,\dots,n-2$.

Alternatively, in [7], prediction, in the context of Functional Autoregressive processes (FAR(1) processes), under the linear assumption on ρ, which is considered to be a compact operator, with $\|\rho\|<1$, is also studied, from smooth data $\widehat{X}_1,\dots,\widehat{X}_n$, solving the optimization problem

$$\min_{\widehat{X}_i\in H_q}\frac{1}{n}\sum_{i=0}^{n-1}\left[\frac{1}{p}\sum_{j=1}^{p}\left\{X_i(t_j)-\widehat{X}_i^{q,\ell}(t_j)\right\}^2+\ell\left\|D^2\widehat{X}_i^{q,\ell}\right\|_{L^2}^2\right],\tag{101}$$

where ℓ is the smoothing parameter, and H_q is the q-dimensional functional subspace spanned by the leading eigenvectors of the autocovariance operator C associated with its largest eigenvalues. Thus, smoothness and rank constraint are considered in the computation of the solution to the optimization problem (101). Such a solution is obtained by means of functional PCA.

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

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

Thus, the regularized estimator of ρ is given by $\widehat{\rho}_{q,l}=\widehat{D}_{q,l}\widehat{C}_{q,l}^{-1}$, and the predictor $\widehat{X}_n^{q,l}=\widehat{\rho}_{q,l}(X_{n-1})$. Due to computational cost limitations, in Table 7, the following statistics are evaluated to compare the performance of the two above-referred prediction methodologies:

$$\mathrm{EMAE}_{\widehat{X}_n}^{h_n}=\frac{1}{p}\sum_{j=1}^{p}\{X_n(t_j)-\widehat{X}_n^{h_n}(t_j)\}^2\tag{102}$$

$$\mathrm{EMAE}_{\widehat{X}_n}^{q,l}=\frac{1}{p}\sum_{j=1}^{p}\{X_n(t_j)-\widehat{X}_n^{q,l}(t_j)\}^2.\tag{103}$$

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.

Table 6
Truncated empirical values of $\mathrm{E}\left\{\|\rho(X_{n-1})-\widetilde{\rho}_{k_n}(X_{n-1})\|_H\right\}$, for $\widetilde{\rho}_{k_n}$ defined in Eq. (96) (third column), for $\widetilde{\rho}_{k_n}=\widetilde{\rho}_n$ given in Eq. (97) (fourth column), and for $\widetilde{\rho}_{k_n}=\widetilde{\rho}_{n,a}$ in Eq. (98) (fifth column), for $\delta_1 = 2.4$ and $\delta_2 = 1.1$, considering the sample sizes $n_t = 15\,000 + 20\,000(t-1),\ t = 1,\dots,20$, and $k_n = \lceil e' n^{1/(8\delta_1+2)}\rceil,\ e' = 17/10$.

n	k_n	Our Approach	Bosq [9]	Guillas [28]
$n_1 = 15\,000$	2	$6.78\,(10)^{-2}$	$8.77\,(10)^{-2}$	$6.64\,(10)^{-2}$
$n_2 = 35\,000$	2	$6.72\,(10)^{-2}$	$8.61\,(10)^{-2}$	$6.30\,(10)^{-2}$
$n_3 = 55\,000$	2	$6.46\,(10)^{-2}$	$8.48\,(10)^{-2}$	$6.17\,(10)^{-2}$
$n_4 = 75\,000$	2	$6.24\,(10)^{-2}$	$8.20\,(10)^{-2}$	$5.76\,(10)^{-2}$
$n_5 = 95\,000$	2	$5.42\,(10)^{-2}$	$7.84\,(10)^{-2}$	$5.03\,(10)^{-2}$
$n_6 = 115\,000$	2	$4.84\,(10)^{-2}$	$7.34\,(10)^{-2}$	$4.56\,(10)^{-2}$
$n_7 = 135\,000$	2	$4.27\,(10)^{-2}$	$6.95\,(10)^{-2}$	$3.94\,(10)^{-2}$
$n_8 = 155\,000$	2	$3.64\,(10)^{-2}$	$6.60\,(10)^{-2}$	$3.65\,(10)^{-2}$
$n_9 = 175\,000$	3	$3.51\,(10)^{-2}$	$6.52\,(10)^{-2}$	$3.42\,(10)^{-2}$
$n_{10} = 195\,000$	3	$3.38\,(10)^{-2}$	$6.16\,(10)^{-2}$	$3.24\,(10)^{-2}$
$n_{11} = 215\,000$	3	$3.16\,(10)^{-2}$	$5.78\,(10)^{-2}$	$2.85\,(10)^{-2}$
$n_{12} = 235\,000$	3	$2.98\,(10)^{-2}$	$5.53\,(10)^{-2}$	$2.60\,(10)^{-2}$
$n_{13} = 255\,000$	3	$2.83\,(10)^{-2}$	$5.15\,(10)^{-2}$	$2.34\,(10)^{-2}$
$n_{14} = 275\,000$	3	$2.50\,(10)^{-2}$	$4.85\,(10)^{-2}$	$2.05\,(10)^{-2}$
$n_{15} = 295\,000$	3	$2.23\,(10)^{-2}$	$4.46\,(10)^{-2}$	$1.83\,(10)^{-2}$
$n_{16} = 315\,000$	3	$2.15\,(10)^{-2}$	$4.30\,(10)^{-2}$	$1.58\,(10)^{-2}$
$n_{17} = 335\,000$	3	$2.06\,(10)^{-2}$	$4.14\,(10)^{-2}$	$1.40\,(10)^{-2}$
$n_{18} = 355\,000$	3	$1.98\,(10)^{-2}$	$3.95\,(10)^{-2}$	$1.24\,(10)^{-2}$
$n_{19} = 375\,000$	3	$1.89\,(10)^{-2}$	$3.77\,(10)^{-2}$	$1.05\,(10)^{-2}$
$n_{20} = 395\,000$	3	$1.82\,(10)^{-2}$	$3.70\,(10)^{-2}$	$9.93\,(10)^{-3}$

Table 7
$\mathrm{EMAE}^{h_{n,i}}_{\widehat{X}_n}$, $i = 1, 2$, and $\mathrm{EMAE}^{q,l}_{\widehat{X}_n}$ values (see Eqs. (102) and (103), respectively), with $q = 7$, for $\delta_1 = 2.4$ and $\delta_2 = 1.1$, considering now the sample sizes $n_t = 750 + 500(t-1),\ t = 1,\dots,13$, $h_{n,1} = 0.1$ and $h_{n,2} = 0.3$.

n	k_n	$\mathrm{EMAE}^{h_{n,1}}_{\widehat{X}_n}$	$\mathrm{EMAE}^{h_{n,2}}_{\widehat{X}_n}$	$\mathrm{EMAE}^{q,l}_{\widehat{X}_n}$
$n_1 = 750$	3	0.0857	0.0885	0.0899
$n_2 = 1250$	3	0.0767	0.0843	0.0869
$n_3 = 1750$	3	0.0715	0.0712	0.0805
$n_4 = 2250$	3	0.0709	0.0687	0.0759
$n_5 = 2750$	3	0.0687	0.0667	0.0731
$n_6 = 3250$	3	0.0652	0.0592	0.0728
$n_7 = 3750$	3	0.0620	0.0556	0.0713
$n_8 = 4250$	4	0.0606	0.0532	0.0706
$n_9 = 4750$	4	0.0567	0.0525	0.0647
$n_{10} = 5250$	4	0.0524	0.0512	0.0608
$n_{11} = 5750$	4	0.0501	0.0482	0.0575
$n_{12} = 6250$	4	0.0490	0.0449	0.0533
$n_{13} = 6750$	4	0.0487	0.0387	0.0497

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 $\phi_{Jk},\ k = 0,\dots,2^J-1$, at level J, associated with a multiresolution analysis of H, is first considered. For a given primary resolution level j_0, with $j_0 < J$, the following wavelet decomposition at $J-j_0$ resolution levels can be computed for any projected curve $\Phi_{V_J}X_i$, in the space V_J, for $i = 0,\dots,n-1$:

$$\Phi_{V_J}X_i = \sum_{k=0}^{2^{j_0}-1} c^i_{j_0k}\phi_{j_0k} + \sum_{j=j_0}^{J-1}\sum_{k=0}^{2^j-1} d^i_{jk}\psi_{jk}$$

$$c^i_{j_0k} = \langle \Phi_{V_J}X_i, \phi_{j_0k}\rangle_H, \qquad d^i_{jk} = \langle \Phi_{V_J}X_i, \psi_{jk}\rangle_H. \tag{104}$$

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

$$\inf_{f^i\in H}\left(\left\|\Phi_{V_J}X_i - f^i\right\|^2_{L^2} + \lambda\left\|\Phi_{V^{\perp}_{j_0}}f^i\right\|^2 : f^i \in H\right), \tag{105}$$

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

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

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

$$\sum_{k=0}^{2^{j_0}-1} (\alpha_{j_0 k}^i - c_{j_0 k}^i)^2 + \sum_{j=j_0}^{J-1} \sum_{k=0}^{2^j-1} (d_{jk}^i - \beta_{jk}^i)^2 + \sum_{j=j_0}^{\infty} \sum_{k=0}^{2^j-1} \lambda 2^{js} (\beta_{jk}^i)^2. \tag{106}$$

The solution to (106) is given, for $i = 0, \ldots, n-1$, by

$$\widehat{\alpha_{j_0 k}^i} = c_{j_0 k}^i, \quad k = 0, \ldots, 2^{j_0} - 1 \tag{107}$$

$$\widehat{\beta_{jk}^i} = \frac{d_{jk}^i}{(1 + \lambda 2^{2sj})}, \quad k = 0, \ldots, 2^j - 1,\ j = j_0, \ldots, J-1 \tag{108}$$

$$\widehat{\beta_{jk}^i} = 0, \quad k = 0, \ldots, 2^j - 1,\ j \geq J. \tag{109}$$

In particular, in the subsequent computations, we have considered the smoothing parameter

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

see [3]. The following smoothed data are then computed

$$\widetilde{X}_{i,\widehat{\lambda}^M} = \sum_{k=0}^{2^{j_0}-1} \widehat{\alpha_{j_0 k}^i} \phi_{j_0 k} + \sum_{j=j_0}^{J-1} \sum_{k=0}^{2^j-1} \widehat{\beta_{jk}^i} \psi_{jk}, \tag{110}$$

removing the trend $\widetilde{a}_{n,\widehat{\lambda}^M} = \sum_{i=0}^{n-1} \widetilde{X}_{i,\widehat{\lambda}^M}/n$ to obtain $\widetilde{Y}_{i,\widehat{\lambda}^M} = \widetilde{X}_{i,\widehat{\lambda}^M} - \widetilde{a}_{n,\widehat{\lambda}^M}$, $i = 0, \ldots, n-1$, for the computation of

$$\widetilde{\rho}_{n,\widehat{\lambda}^M}(x) = \left(\widetilde{\Pi}_{\widehat{\lambda}^M}^{k_n} \widetilde{D}_{n,\widehat{\lambda}^M} \widetilde{C}_{n,\widehat{\lambda}^M}^{-1} \widetilde{\Pi}_{\widehat{\lambda}^M}^{k_n} \right)(x) = \sum_{\ell=1}^{k_n} \widetilde{\rho}_{n,\widehat{\lambda}^M,\ell}(x) \widetilde{\phi}_\ell^M,$$

$$\widetilde{\rho}_{n,\widehat{\lambda}^M,\ell}(x) = \sum_{j=1}^{k_n} \frac{1}{n-1} \sum_{i=0}^{n-2} \frac{1}{\widetilde{C}_{n,\widehat{\lambda}^M,j}} \langle \widetilde{\phi}_j^M, x \rangle_H \widetilde{Y}_{i,\widehat{\lambda}^M,j} \widetilde{Y}_{i+1,\widehat{\lambda}^M,\ell},$$

for $x \in H$, and

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

where

$$\widetilde{Y}_{i,\widehat{\lambda}^M,j} = \left\langle \widetilde{Y}_{i,\widehat{\lambda}^M}, \widehat{\phi}_{j,\widehat{\lambda}^M} \right\rangle, \quad \text{and} \quad \widetilde{C}_{n,\widehat{\lambda}^M,j} = \left\langle \widetilde{C}_{n,\widehat{\lambda}^M} \widehat{\phi}_{j,\widehat{\lambda}^M}, \widehat{\phi}_{j,\widehat{\lambda}^M} \right\rangle,$$

for every $j \geq 1$. Table 8 displays the empirical truncated approximations of

$$\mathrm{E}\left\{ \|\widehat{\rho}_{k_n}(X_{n-1}) - \rho(X_{n-1})\|_H \right\} \quad \text{and} \quad \mathrm{E}\{\|\widetilde{\rho}_{n,\widehat{\lambda}^M}(X_{n-1}) - \rho(X_{n-1})\|_H\},$$

respectively obtained by applying our approach, and the approach in [4], in the estimation of the autocorrelation operator ρ. Here, we have tested $k_{n_i} = n^{1/\alpha_i}$, $i = 1, 2$, with $\alpha_1 = 6$, according to Assumption A3, and $\alpha_2 > 4\delta_1$, according to $H_4 : nC_{k_n}^4 \to \infty$ in p. 149 of [4]. In particular, we have considered $\delta_1 = 2.4$, and $\alpha_2 = 10$.

From the results displayed in Table 8, one can observe a similar performance for the two truncation rules implemented, and approaches compared, for the small sample sizes tested. A similar performance is also displayed by the approaches presented in [7], for such small sample sizes; see Table 7.

Table 8
Truncated empirical values of $E\{\|\rho(X_{n-1}) - \widetilde{\rho}_{k_n}(X_{n-1})\|_H\}$, with $\widetilde{\rho}_{k_n}$ defined in Eq. (96), and of $E\{\|\widetilde{\rho}_{n,\widehat{\lambda}^M}(X_{n-1}) - \rho(X_{n-1})\|_H\}$, 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 $\widehat{\lambda}_M,\ M = 50$, and the corresponding $k_{n,i} = n^{1/\alpha_i}$, for $\alpha_1 = 6,\ \alpha_2 = 10$. Here, O.A. means *Our Approach* and [4] means *The approach presented in* [4].

n	$k_{n,1}$	O.A.	[4]	$k_{n,2}$	O.A.	[4]
$n_1 = 750$	3	0.0702	0.0911	1	0.0636	0.0589
$n_2 = 1250$	3	0.0550	0.0873	2	0.0509	0.0429
$n_3 = 1750$	3	0.0473	0.0803	2	0.0455	0.0394
$n_4 = 2250$	3	0.0414	0.0795	2	0.0409	0.0377
$n_5 = 2750$	3	0.0365	0.0734	2	0.0355	0.0349
$n_6 = 3250$	3	0.0343	0.0719	2	0.0333	0.0307
$n_7 = 3750$	3	0.0330	0.0675	2	0.0325	0.0293
$n_8 = 4250$	4	0.0328	0.0672	2	0.0313	0.0286
$n_9 = 4750$	4	0.0317	0.0664	2	0.0309	0.0256
$n_{10} = 5250$	4	0.0309	0.0636	2	0.0276	0.0229
$n_{11} = 5750$	4	0.0298	0.0598	2	0.0203	0.0196
$n_{12} = 6250$	4	0.0283	0.0583	2	0.0166	0.0153
$n_{13} = 6750$	4	0.0276	0.0555	2	0.0148	0.0137

6. Final comments

As noted before, in this paper, the eigenvectors of C are considered to be known in the derivation of the results on consistency. This assumption is satisfied, e.g., when the random initial condition is given as the solution, in the mean-square sense, of a stochastic differential equation driven by white noise (e.g., the Wiener measure), since the eigenvectors of the differential operator involved in that equation coincide with the eigenvectors of the autocovariance operator of the ARH(1) process.

In the case where the eigenvectors of the autocovariance operator are unknown, the numerical results displayed in Section 5.2.2 (see Tables 4–6) illustrate the fact that our approach displays, in terms of the empirical eigenvectors, very similar prediction results to those obtained with the implementation of the component-wise estimators proposed in [9,28], with a better performance of our approach in the more unfavorable case, corresponding to a large discretization step size, and truncation order; see Table 4 computed for $k_n = \ln(n)$.

Regarding Assumption A2, Remark 1 provides an example where this assumption is satisfied. However, our approach can still be applied in a wider range of situations. Wavelet bases are well suited for sparse representation of functions; recent work has considered combining them with sparsity-inducing penalties, both for semiparametric regression (see, e.g., [55]), and for regression with functional or kernel predictors; see [55–57], among others. The latter papers focused on ℓ_1 penalization, also known as the lasso [52], in the wavelet domain. Alternatives to the lasso include the SCAD penalty [19], and the adaptive lasso [58]. 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 approximation $\widehat{X}_1, \ldots \widehat{X}_n$ of the functional data $X_1, \ldots, X_n$, whose empirical autocovariance operator, and cross-covariance operator, respectively defined by $\widehat{C}_n = \frac{1}{n}\sum_{i=0}^{n-1} \widehat{X}_i \otimes \widehat{X}_i$, and $\widehat{D}_n = \frac{1}{n-1}\sum_{i=0}^{n-2} \widehat{X}_i \otimes \widehat{X}_{i+1}$, admit a diagonal representation in terms of wavelets. In the literature, shrinkage approaches for estimating a high-dimensional covariance matrix are employed to circumvent the limitations of the sample covariance matrix. In particular, a new family of nonparametric Stein-type shrinkage covariance estimators is proposed in [53] (see also references therein), whose members are written as a convex linear combination of the sample covariance matrix and of a predefined invertible diagonal target matrix. These results can be applied to our framework, considering the shrinkage estimators of the autocovariance, C, and cross-covariance, D, operators, with respect to a common suitable wavelet basis, which can lead to an empirical diagonal representation of both operators.

In the supplementary material (see Appendix A), a numerical example is provided to illustrate the performance of our approach, in the case of a pseudo-diagonal autocorrelation operator.

Acknowledgments

The authors are grateful to the Editor, Associate Editor, and referees for comments which helped improve the readability of this paper.

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

Appendix A. Supplementary material

Supplementary material related to this article can be found online at http://dx.doi.org/10.1016/j.jmva.2016.11.009.

References

[1] G. Aneiros-Pérez, P. Vieu, Semi-functional partial linear regression, Statist. Probab. Lett. 76 (2006) 1102–1110.
[2] G. Aneiros-Pérez, P. Vieu, Nonparametric time series prediction: a semi-functional partial linear modeling, J. Multivariate Anal. 99 (2008) 834–857.
[3] C. Angelini, D. De Canditiis, F. Leblanc, Wavelet regression estimation in nonparametric mixed effect models, J. Multivariate Anal. 85 (2003) 267–291.
[4] A. Antoniadis, T. Sapatinas, Wavelet methods for continuous-time prediction using Hilbert-valued autoregressive processes, J. Multivariate Anal. 87 (2003) 133–158.
[5] M.S. Bartlett, On the theoretical specification and sampling properties of autocorrelated time series, J. Roy. Stat. Soc. 8 (1946) 27–41. Supplement to.
[6] N. Bensmain, T. Mourid, Estimateur "sieve" de l'opérateur d'un processus ARH(1), C. R. Acad. Sci., Paris I 332 (2001) 1015–1018.
[7] P.C. Besse, H. Cardot, D.B. Stephenson, Autoregressive forecasting of some functional climatic variations, Scand. J. Stat. 27 (2000) 673–687.
[8] E.G. Bongiorno, A. Goia, E. Salinelli, P. Vieu, Contributions in Infinite-Dimensional Statistics and Related Topics, Soc. Editrice Esculapio, Bologna, 2014.
[9] D. Bosq, Linear Processes in Function Spaces, Springer, New York, 2000.
[10] D. Bosq, General linear processes in Hilbert spaces and prediction, J. Statist. Plann. Inference 137 (2007) 879–894.
[11] D. Bosq, D. Blanke, Inference and Predictions in Large Dimensions, Wiley, Chichester, 2007.
[12] J. Chen, D. Lin, H. Hochner, Semiparametric maximum likelihood methods for analyzing genetic and environmental effects with case-control mother-child pair data, Biometrics 68 (2012) 869–877.
[13] A. Cuevas, A partial overview of the theory of statistics with functional data, J. Statist. Plann. Inference 147 (2014) 1–23.
[14] J. Cugliari, Prévision non paramétrique de processus à valeurs fonctionnelles. Application à la consommation d'électricité (Ph.D. thesis), University of Paris-Sud 11, 2011.
[15] J. Damon, S. Guillas, Estimation and simulation of autoregressive processes with exogenous variables, Statist. Inference 8 (2005) 185–204.
[16] R. Dautray, J.-L. Lions, Mathematical Analysis and Numerical Methods for Science and Technology Volume 3: Spectral Theory and Applications, Springer, New York, 1990.
[17] J. Dedecker, F. Merlevede, The conditional central limit theorem in Hilbert spaces, Stochastic Process. Appl. 108 (2003) 229–262.
[18] H. Dehling, O.S. Sharipov, Estimation of mean and covariance operator for Banach space valued autoregressive processes with dependent innovations, Stat. Inference Stoch. Process. 8 (2005) 137–149.
[19] J. Fan, R. Li, Variable selection via nonconcave penalized likelihood and its oracle properties, J. Amer. Statist. Assoc. 96 (2001) 1348–1360.
[20] M. Febrero-Bande, W. Gonzalez-Manteiga, Generalized additive models for functional data, TEST 22 (2013) 278–292.
[21] F. Ferraty, A. Goia, E. Salinelli, P. Vieu, Functional projection pursuit regression, TEST 22 (2013) 293–320.
[22] F. Ferraty, I.V. Keilegom, P. Vieu, Regression when both response and predictor are functions, J. Multivariate Anal. 109 (2012) 10–28.
[23] F. Ferraty, P. Vieu, Nonparametric Functional Data Analysis, Springer, New York, 2006.
[24] F. Ferraty, P. Vieu, Additive prediction and boosting for functional data, Comput. Statist. Data Anal. 53 (2009) 1400–1413.
[25] A. Goia, P. Vieu, An introduction to recent advances in high/infinite dimensional statistics, J. Multivariate Anal. 146 (2016) 1–6.
[26] D.S. Grebenkov, B.T. Nguyen, Geometrical structure of Laplacian eigenfunctions, SIAM Rev. 55 (2013) 601–667.
[27] S. Guillas, Noncausality and functional discretization, limit theorems for an ARHX(1) process, C. R. Acad. Sci., Paris I 331 (2000) 91–94.
[28] S. Guillas, Rates of convergence of autocorrelation estimates for autoregressive Hilbertian processes, Statist. Probab. Lett. 55 (2001) 281–291.
[29] J. Gurland, Quadratic forms in normally distributed random variables, Sankhyā 17 (1956) 37–50.
[30] J.D. Hamilton, Time Series Analysis, Princeton University Press, Princenton, 1994.
[31] L. Horváth, P. Kokoszka, Inference for Functional Data with Applications, Springer, New York, 2012.
[32] T. Hsing, R. Eubank, Theoretical Foundations of Functional Data Analysis, with an Introduction to Linear Operators, Wiley, 2015.
[33] V. Kargin, A. Onatski, Curve forecasting by functional autoregression, J. Multivariate Anal. 99 (2008) 2508–2526.
[34] S. Mallat, A Wavelet Tour of Signal Processing: The Sparse Way, third ed., Academic Press, Burlington, MA, 2009.
[35] J.-M. Marion, B. Pumo, Comparison of ARH(1) and ARHD(1) models on physiological data, Ann. Inst. Stat. Univ. Paris 48 (2004) 29–38.
[36] A. Mas, Normalité asymptotique de l'estimateur empirique de l'opérateur d'autocorrélation d'un processus ARH(1), C. R. Acad. Sci., Paris I 329 (1999) 899–902.
[37] A. Mas, Consistance du prédicteur dans le modèle ARH(1): le cas compact, Ann. Inst. Stat. Univ. Paris 48 (2004) 39–48.
[38] A. Mas, Weak-convergence in the functional autoregressive model, J. Multivariate Anal. 98 (2007) 1231–1261.
[39] A. Mas, L. Menneteau, Large and moderate deviations for infinite dimensional autoregressive processes, J. Multivariate Anal. 87 (2003) 241–260.
[40] L. Menneteau, Some laws of the iterated logarithm in Hilbertian autoregressive models, J. Multivariate Anal. 92 (2005) 405–425.
[41] F. Merlevede, 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 I 323 (1996) 521–524.
[42] F. Merlevede, Résultats de convergence presque sûre pour l'estimation et la prévision des processus linéaires Hilbertiens, C. R. Acad. Sci., Paris I 324 (1997) 573–576.
[43] T. Mourid, Processus autorégressifs Banachiques d'ordre supérieur, C. R. Acad. Sci., Paris I 317 (1993) 1167–1172.
[44] T. Mourid, Processus autorégressifs Hilbertiens à coefficients aléatoires, Ann. Inst. Stat. Univ. Paris 48 (2004) 79–85.
[45] G.P. Nason, Wavelet Methods in Statistics with R, Springer, Berlin, 2008.
[46] R.T. Ogden, Essential Wavelets for Statistical Applications and Data Analysis, Birkhäuser, Boston, MA, 1997.
[47] B. Pumo, Estimation et prévision de processus autorégressifs fonctionnels (Ph.D. thesis), Université de Paris 6, 1992.
[48] B. Pumo, Prediction of continuous time processes by $C_{[0,1]}$-valued autoregressive process, Stat. Inference Stoch. Process. 1 (1998) 297–309.
[49] J.O. Ramsay, B.W. Silverman, Functional Data Analysis, second ed., Springer, New York, 2005.
[50] M.D. Ruiz-Medina, Spatial autoregressive and moving average Hilbertian processes, J. Multivariate Anal. 102 (2011) 292–305.
[51] M.D. Ruiz-Medina, Spatial functional prediction from spatial autoregressive Hilbertian processes, Environmetrics 23 (2012) 119–128.
[52] R.J. Tibshirani, Regression shrinkage and selection via the LASSO, J. R. Stat. Soc. Ser. B Stat. Methodol. 58 (1996) 267–288.
[53] A. Touloumis, Nonparametric stein-type shrinkage covariance matrix estimators in high-dimensional settings, Comput. Statist. Data Anal. 83 (2015) 251–261.
[54] B. Vidakovic, Nonlinear wavelet shrinkage with Bayes rules and Bayes factors, J. Amer. Statist. Assoc. 93 (1998) 173–179.
[55] M.P. Wand, J.T. Ormerod, Penalized wavelets: Embedding wavelets into semiparametric regression, Electron. J. Stat. 5 (2011) 1654–1717.
[56] Y. Zhao, H. Chen, R.T. Ogden, Wavelet-based weighted LASSO and screening approaches in functional linear regression, J. Comput. Graph. Statist. 24 (2015) 655–675.
[57] Y. Zhao, R.T. Ogden, P.T. Reiss, Wavelet-based LASSO in functional linear regression, J. Comput. Graph. Statist. 21 (2012) 600–617.
[58] H. Zou, The adaptive LASSO and its oracle properties, J. Amer. Statist. Assoc. 101 (2006) 1418–1429.