Statistics & Probability Letters 55 (2001) 281–291

STATISTICS & PROBABILITY LETTERS

www.elsevier.com/locate/stapro

Rates of convergence of autocorrelation estimates for autoregressive Hilbertian processes

Serge Guillas

Université Paris VI and Ecole des Mines de Douai, Laboratoire de Statistique Théorique et Appliquée, 175 rue du Chevaleret, 75013 Paris, France

Received March 2001; received in revised form May 2001

Abstract

We show consistency in the mean integrated quadratic sense of an estimator of the autocorrelation operator ρ in the autoregressive Hilbertian of order one model. Two main cases are considered, and we obtain upper bounds for the corresponding rates. © 2001 Elsevier Science B.V. All rights reserved

MSC: 62G05; 62M10

Keywords: Hilbert space; Autoregressive processes; Estimation; Rate of convergence; Functional data

1. Introduction

Let H be a real and separable Hilbert space with norm $\|\cdot\|$. Let ρ be a bounded operator on H. We suppose that $\sum_{n=0}^{\infty}\|\rho^n\|_{\mathscr{L}}<\infty$, where $\|\cdot\|_{\mathscr{L}}$ is the linear norm of operators in H. Let (ε_n) be a strong Hilbertian white noise (SWN), that is a sequence of i.i.d. random variables with values in H satisfying

$$\forall n\in\mathbb{Z},\quad E\varepsilon_n=0,\quad 0<E\|\varepsilon_n\|^2=\sigma^2<\infty.$$

We will consider in this paper the autoregressive Hilbertian of order one model, denoted by ARH(1). It is the unique stationary solution of the equation

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

See Bosq (2000) for an extensive study of the ARH(1) model.

Such Hilbertian processes can theoretically and practically handle situations where continuous-time processes are involved. Precisely, if $(x_t, t\in\mathbb{R})$ is a continuous-time process with continuous paths, then

$$X_k(t)=x_{k\delta+t},\quad 0\leqslant t\leqslant\delta,\ \ k\in\mathbb{Z}$$

E-mail address: guillas@ccr.jussieu.fr (S. Guillas).

PII: S0167-7152(01)00151-1

is a discrete $L^2([0,\delta])$-valued process. Various applications have already been made. For example, Cavallini et al. (1994) made forecasts of electricity consumption; and, by means of smoothing splines, Besse and Cardot (1996) predicted traffic while Besse et al. (2000) made forecasts of the climatic variation called el niño.

Several extensions of the ARH(1) model have been made. We may mention ARH(p) models (see Mourid, 1995), and ARH(1) models with exogenous variables (see Guillas, 2000). Moreover, Cardot et al. (1999) studied a regression model with similar techniques.

Let us denote by C and D, respectively, the covariance and cross-covariance operator of the stationary process X:

$$C(x)=E[\langle X_0,x\rangle X_0],\quad D(x)=E[\langle X_0,x\rangle X_1].$$

It can easily be shown that C is a symmetric positive and compact operator. Defining for all elements u,v in H the operator $u\otimes v$ by

$$u\otimes v(x)=\langle u,x\rangle v, x\in H,$$

we then obtain the decomposition in a complete orthonormal basis (v_j) of H:

$$C=\sum_{j=1}^{\infty}\lambda_j v_j\otimes v_j,$$

where (λ_j) is a decreasing sequence of positive numbers such that

$$\sum_{j=1}^{\infty}\lambda_j=E\|X_0\|^2<\infty.$$

The estimation of ρ is a rather intricate problem. Indeed, classical techniques such as maximum likelihood or least squares are not accurate in this Hilbertian context. Bosq (1991) proposed an alternative technique, which may be seen as a generalization of principal components regression in function spaces. It works as follows: estimate the eigenvectors (v_j) and the eigenvalues (λ_j) of the covariance operator and try to use the relation $D=\rho C$ in order to get ρ. C_n and D_n are the following respective unbiased estimators of C and D:

$$C_n=\frac{1}{n}\sum_{i=1}^{n}X_i\otimes X_i,\quad D_n=\frac{1}{n-1}\sum_{i=1}^{n-1}X_i\otimes X_{i+1}$$

and we denote by (v_{jn}) and (λ_{jn}) the empirical eigenelements of C_n. We would like to define an estimator of ρ as $\rho_n=D_nC_n^{-1}$, but C_n is not invertible in general, so we have to make a projection on the space H_{k_n} spanned by the k_n first eigenvectors of C_n, obtaining this way an invertible operator in H_{k_n}. Naturally, the choice of k_n may not be easy and is usually done empirically or by a cross-validation procedure. In this paper, we will give some ideas about this choice in relatively precise situations.

Bosq (2000) showed almost sure consistency of ρ_n. Mas (1999) obtained results about limit in distribution of ρ_n. The purpose of this paper is to establish consistency of a slight modification of ρ_n in the L^2 mode, that is to say by considering $E\|\hat{\rho}_{n,a}-\rho\|^2_{\mathscr{L}}$, and to obtain rates of convergence when the eigenvectors are known and when they are not.

While in the finite-dimensional case this rate of convergence may reach a $1/n$-rate when the eigenvalues of C are bounded by below (see Bosq, 2000, Section 8.1) we will find in the infinite-dimensional case where the eigenvectors are known a $n^{-1/3}$-rate, and in the general case a $n^{-1/4}$-rate.

In both cases, we will assume the existence of a sequence (a_n) satisfying

$$\exists 0<\beta<1,\quad 0<a_n\leqslant\beta\lambda_{k_n},\quad n\in\mathbb{N}.$$

The sequence (a_n) plays the role of a regulation parameter which allows to control better the inverse of the covariance operator. This idea is similar to the ridge regression technique in linear regression—see Hoerl and

Kennard (1970)—that is, add a multiple of the identity to the covariance matrix in order to stabilize regression estimates.

We also make use of the following assumptions:

(H_1) X is a ARH(1) such that $E\|X_0\|^4 < \infty$.

(H_2) For all j, $\lambda_j > 0$.

(H_3) For all j, $\lambda_j > \lambda_{j+1}$.

2. Known eigenvectors

The case considered here is the case where the eigenvectors v_j of C are known.

Consider the following unbiased estimators of the (λ_j):

$$\hat{\lambda}_{j,n} = \frac{1}{n}\sum_{i=1}^{n} \langle X_i, v_j\rangle^2.$$

For consistency of the $\hat{\lambda}_{j,n}$, see Bosq (2000).

Consider now the following estimators of C:

$$\hat{C}_n = \sum_{j=1}^{k_n} \hat{\lambda}_{jn} v_j \otimes v_j, \quad \hat{C}_{n,a} = \sum_{j=1}^{k_n} \max(\hat{\lambda}_{jn}, a_n) v_j \otimes v_j.$$

Let us define

$$\hat{\rho}_{n,a} = \pi^{k_n} D_n \hat{C}_{n,a}^{-1} \pi^{k_n},$$

where π^{k_n} denotes the orthogonal projector over H_{k_n}, and

$$\hat{C}_{n,a}^{-1} = \sum_{j=1}^{k_n} [\max(\hat{\lambda}_{jn}, a_n)]^{-1} v_j \otimes v_j.$$

Our goal is to find an upper bound for $E\|\hat{\rho}_{n,a} - \rho\|^2_{\mathscr{L}}$.

Lemma 1. *Under* (H_1) *and* (H_2),

$$E\|\hat{\rho}_{n,a} - \rho\|^2_{\mathscr{L}} \leqslant \frac{c_0}{na_n^2} + \frac{c_1}{n\lambda_{k_n}^4} + \frac{c_2}{n\lambda_{k_n}^2 a_n^2} + 2\lambda_{k_n+1}^2. \tag{2}$$

Proof. First, write the decomposition

$$\hat{\rho}_{n,a} - \rho = (\pi^{k_n} D_n \hat{C}_{n,a}^{-1} \pi^{k_n} - \pi^{k_n}\rho\pi^{k_n}) + (\pi^{k_n}\rho\pi^{k_n} - \rho).$$

Observe now that

$$\pi^{k_n}\rho\pi^{k_n}(x) = \pi^{k_n} D \sum_{j=1}^{k_n} \lambda_j^{-1} \langle v_j, x\rangle v_j$$

and let us set

$$C_{\pi^{k_n}} = \sum_{j=1}^{k_n} \lambda_j v_j \otimes v_j,$$
$$C_{\pi^{k_n}}^{-1} = \sum_{j=1}^{k_n} \lambda_j^{-1} v_j \otimes v_j.$$

Accordingly,

$$\hat{\rho}_{n,a} - \rho = (\pi^{k_n} D_n \hat{C}_{n,a}^{-1} \pi^{k_n} - \pi^{k_n} D C_{\pi^{k_n}}^{-1} \pi^{k_n}) + (\pi^{k_n} \rho \pi^{k_n} - \rho). \tag{3}$$

The first term may be written as

$$\begin{aligned}\pi^{k_n} D_n \hat{C}_{n,a}^{-1} \pi^{k_n} - \pi^{k_n} D C_{\pi^{k_n}}^{-1} \pi^{k_n} &= \pi^{k_n} (D_n \hat{C}_{n,a}^{-1} - D C_{\pi^{k_n}}^{-1}) \pi^{k_n} \\ &= \pi^{k_n} [(D_n - D) \hat{C}_{n,a}^{-1} + D(\hat{C}_{n,a}^{-1} - C_{\pi^{k_n}}^{-1})] \pi^{k_n} \\ &= \pi^{k_n} [(D_n - D) \hat{C}_{n,a}^{-1} - D \hat{C}_{n,a}^{-1} (\hat{C}_{n,a} - C_{\pi^{k_n}}) C_{\pi^{k_n}}^{-1}] \pi^{k_n},\end{aligned}$$

hence

$$\begin{aligned}&\|\pi^{k_n} D_n \hat{C}_{n,a}^{-1} \pi^{k_n} - \pi^{k_n} D C_{\pi^{k_n}}^{-1} \pi^{k_n}\|_{\mathscr{L}}^2 \\ &\quad \leqslant 2\|D_n - D\|_{\mathscr{L}}^2 \|\hat{C}_{n,a}^{-1}\|_{\mathscr{L}}^2 + 2\|D\|_{\mathscr{L}}^2 \|\hat{C}_{n,a}^{-1}\|_{\mathscr{L}}^2 \|C_{\pi^{k_n}}^{-1}\|_{\mathscr{L}}^2 \|\hat{C}_{n,a} - C_{\pi^{k_n}}\|_{\mathscr{L}}^2 \\ &\quad \leqslant 2a_n^{-2} \|D_n - D\|_{\mathscr{L}}^2 + 2a_n^{-2} \|D\|_{\mathscr{L}}^2 \|C_{\pi^{k_n}}^{-1}\|_{\mathscr{L}}^2 \|\hat{C}_{n,a} - C_{\pi^{k_n}}\|_{\mathscr{L}}^2,\end{aligned}$$

because

$$\|\hat{C}_{n,a}^{-1}\|_{\mathscr{L}} \leqslant a_n^{-1}.$$

Thus, by (3),

$$E\|\hat{\rho}_{n,a} - \rho\|_{\mathscr{L}}^2 \leqslant 2[2a_n^{-2} E\|D_n - D\|_{\mathscr{L}}^2 + 2a_n^{-2} \|D\|_{\mathscr{L}}^2 \|C_{\pi^{k_n}}^{-1}\|_{\mathscr{L}}^2 E\|\hat{C}_{n,a} - C_{\pi^{k_n}}\|_{\mathscr{L}}^2] + 2E\|\rho \pi^{k_n} - \rho\|_{\mathscr{L}}^2. \tag{4}$$

The second term of the right-hand side is easily bounded from above by $2\lambda_{k_n+1}^2$. For the first term (Bosq, 2000, Theorem 4.8) gives

$$E\|D_n - D\|_{\mathscr{L}}^2 = \mathrm{O}\left(\frac{1}{n}\right)$$

and clearly

$$\|C_{\pi^{k_n}}^{-1}\|_{\mathscr{L}}^2 = \frac{1}{\lambda_{k_n}^2}.$$

Moreover,

$$\begin{aligned}E\|\hat{C}_{n,a} - C_{\pi^{k_n}}\|_{\mathscr{L}}^2 \leqslant\ & 2E(\|\hat{C}_{n,a} - \hat{C}_n\|_{\mathscr{L}}^2 \mathbf{1}_{\hat{C}_{n,a} \neq \hat{C}_n}) + 2E(\|\hat{C}_n - C_{\pi^{k_n}}\|_{\mathscr{L}}^2 \mathbf{1}_{\hat{C}_{n,a} \neq \hat{C}_n}) \\ & + E(\|\hat{C}_{n,a} - C_{\pi^{k_n}}\|_{\mathscr{L}}^2 \mathbf{1}_{\hat{C}_{n,a} = \hat{C}_n}).\end{aligned}$$

Now, we find an upper bound to $P(\hat{C}_{n,a} \neq \hat{C}_n)$, knowing that the sequence $(\hat{\lambda}_{jn})$ is not decreasing with respect to j. Observe that

$$P(\hat{C}_{n,a} \neq \hat{C}_n) = P\left(a_n > \min_{j=1,\dots,k_n} \hat{\lambda}_{jn}\right).$$

Let us define the discrete random variable $I_{k_n} = \arg\min\{\hat{\lambda}_{jn},\ j = 1,\ldots,k_n\}$. We then obtain

$$
\begin{aligned}
P(\hat{C}_{n,a} \neq \hat{C}_n) &= P(a_n > \hat{\lambda}_{I_{k_n}n}) \\
&= P(\hat{\lambda}_{I_{k_n},n} - \lambda_{I_{k_n}} < a_n - \lambda_{I_{k_n}}) \\
&\leqslant P(|\hat{\lambda}_{I_{k_n},n} - \lambda_{I_{k_n}}| \geqslant (1-\beta)\lambda_{I_{k_n}}) \\
&\leqslant P(|\hat{\lambda}_{I_{k_n},n} - \lambda_{I_{k_n}}| \geqslant (1-\beta)\lambda_{k_n}),
\end{aligned}
$$

so

$$
\begin{aligned}
P(\hat{C}_{n,a} \neq \hat{C}_n) &\leqslant P\left(\sup_{j=1,\ldots,k_n} |\hat{\lambda}_{jn} - \lambda_j| \geqslant (1-\beta)\lambda_{k_n}\right) \\
&\leqslant P(\|\hat{C}_n - C\|_{\mathscr{L}} \geqslant (1-\beta)\lambda_{k_n}) \\
&\leqslant \frac{K}{n(1-\beta)^2\lambda_{k_n}^2}
\end{aligned}
$$

with a constant $K > 0$, applying the Chebychev inequality since

$$E\|\hat{C}_n - C\|_{\mathscr{L}}^2 \leqslant 2E\|\hat{C}_n - C_n\|_{\mathscr{L}}^2 + 2E\|C_n - C\|_{\mathscr{L}}^2$$

and

$$
\begin{aligned}
E\|\hat{C}_n - C_n\|_{\mathscr{L}}^2 &= E\left\|\sum_{j=1}^{k_n}[\hat{\lambda}_{jn} - \lambda_{jn}]v_j \otimes v_j\right\|_{\mathscr{L}}^2 \\
&\leqslant E\sup_{j=1,\ldots,k_n}\left|\hat{\lambda}_{jn} - \lambda_{jn}\right|^2 \leqslant 2E\sup_{j=1,\ldots,k_n}\left|\hat{\lambda}_{jn} - \lambda_j\right|^2 + 2E\sup_{j=1,\ldots,k_n}|\lambda_{jn} - \lambda_j|^2 \\
&= \mathrm{O}\left(\frac{1}{n}\right)
\end{aligned}
$$

by Bosq (2000, Theorem 4.4, Corollary 4.5) so by Bosq (2000, Theorem 4.1)

$$E\|\hat{C}_n - C\|_{\mathscr{L}}^2 = \mathrm{O}\left(\frac{1}{n}\right).$$

Note that

$$\|\hat{C}_{n,a} - \hat{C}_n\|_{\mathscr{L}}^2 = \left\|\sum_{j=1}^{k_n}[\max(\hat{\lambda}_{jn}, a_n) - \hat{\lambda}_{jn}]v_j \otimes v_j\right\|_{\mathscr{L}}^2 \leqslant a_n^2$$

and that

$$
\begin{aligned}
E\|\hat{C}_n - C_{\pi^{k_n}}\|_{\mathscr{L}}^2 &= E\left\|\sum_{j=1}^{k_n}[\hat{\lambda}_{jn} - \lambda_j]v_j \otimes v_j\right\|_{\mathscr{L}}^2 \leqslant E\sup_{j=1,\ldots,k_n}|\hat{\lambda}_{jn} - \lambda_j|^2 \\
&\leqslant E\|C_n - C\|_{\mathscr{L}}^2 = \mathrm{O}\left(\frac{1}{n}\right),
\end{aligned}
$$

Therefore we get, by (4)

$$E\|\hat{\rho}_{n,a}-\rho\|^2_{\mathscr{L}} \leqslant \frac{c_0}{na_n^2}+\frac{c_1}{n\lambda_{k_n}^4}+\frac{c_2}{n\lambda_{k_n}^2 a_n^2}+2\lambda_{k_n+1}^2. \qquad \square$$

Theorem 1. *Suppose that* (H_1) *and* (H_2) *hold, and that there exist* $\alpha>0$, $0<\beta<1$, $\varepsilon<1/2$ *and* $\gamma\geqslant 1$ *such that*

$$\alpha\frac{\lambda_{k_n}^{\gamma}}{n^{\varepsilon}} \leqslant a_n \leqslant \beta\lambda_{k_n},$$

then

$$E\|\hat{\rho}_{n,a}-\rho\|^2_{\mathscr{L}} = \mathrm{O}\left(\frac{1}{n^{(1-2\varepsilon)}\lambda_{k_n}^{2(1+\gamma)}}\right)+\mathrm{O}(\lambda_{k_n}^2).$$

Proof. It is an easy consequence of (2), using the inequalities $\alpha\lambda_{k_n}^{\gamma}/n^{\varepsilon}\leqslant a_n$ and $\lambda_{k_n+1}\leqslant\lambda_{k_n}$. $\square$

Remark 1. The optimal choice of λ_{k_n} is such that

$$\lambda_{k_n}^2=\frac{c}{n^{(1-2\varepsilon)}\lambda_{k_n}^{2+2\gamma}} \quad \text{i.e.} \quad \lambda_{k_n}^{4+2\gamma}=\frac{c}{n^{(1-2\varepsilon)}}, \quad c>0. \tag{5}$$

The rate of convergence in quadratic mean is then of order

$$\lambda_{k_n}^2 \asymp n^{-(1-2\varepsilon)/(\gamma+2)}.$$

Remark 2. When $\varepsilon=0$ and in the most favorable case where $\gamma=1$, the rate of convergence is of order $n^{-1/3}$.

Example 1. If $\lambda_j=ar^j$, where $a>0$ and $0<r<1$, by (5), we get

$$r^{(4+2\gamma)k_n}=\frac{d}{n^{(1-2\varepsilon)}}, \quad d>0,$$

which yields

$$k_n=\left\lfloor\frac{\ln d-(1-2\varepsilon)\ln n}{(4+2\gamma)\ln r}\right\rfloor.$$

Example 2. If $\lambda_j=aj^{-\delta}$ where $a>0$ and $\delta>1$, by (9), we get

$$k_n=\lfloor en^{(1-2\varepsilon)/(4+2\gamma)\delta}\rfloor, \quad e>0.$$

3. General case

We consider here the empirical eigenelements of C given by

$$C_n(v_{jn})=\lambda_{jn}v_{jn},$$

where $\lambda_{1n}\geqslant\ldots\geqslant\lambda_{nn}\geqslant 0=\lambda_{n+1,n}=\lambda_{n+2,n}=\cdots$, and (v_{jn}) constitutes an orthonormal system of H. We denote $\tilde{H}_{k_n}$ the space spanned by $v_{1n},\ldots,v_{k_n,n}$. We assume in this section that each eigensubspace associated to the eigenvectors λ_j is one dimensional. Consider the following empirical eigenvectors for identifiability reasons:

$$v'_{jn} = \operatorname{sgn}\langle v_{jn}, v_j\rangle v_j, \quad j \geqslant 1.$$

Consider the following estimators of C:

$$\tilde{C}_n = \sum_{j=1}^{k_n} \lambda_{jn} v_{jn} \otimes v_{jn}, \quad \tilde{C}_{n,a} = \sum_{j=1}^{k_n} \max(\lambda_{jn}, a_n) v_{jn} \otimes v_{jn}.$$

Let us set

$$\tilde{\rho}_{n,a} = \tilde{\pi}^{k_n} D_n \tilde{C}_{n,a}^{-1} \tilde{\pi}^{k_n},$$

where $\tilde{\pi}^{k_n}$ denotes the orthogonal projector over $\tilde{H}_{k_n}$, and

$$\tilde{C}_{n,a}^{-1} = \sum_{j=1}^{k_n} [\max(\lambda_{jn}, a_n)]^{-1} v_{jn} \otimes v_{jn}.$$

We will show analogous results as in the previous section, using only slightly different techniques. We will use in the sequel the following numbers defined under (H_3):

$$\Lambda_{k_n} = \sup_{j=1,\dots,k_n} \frac{1}{\lambda_j - \lambda_{j+1}},$$

cf. Bosq (2000, p. 107). Now we can give an upper bound for $E\|\tilde{\rho}_{n,a} - \rho\|^2_{\mathscr{L}}$.

Lemma 2. *Under* (H_1)–(H_3)

$$E\|\tilde{\rho}_{n,a} - \rho\|^2_{\mathscr{L}} \leqslant \frac{c'_0}{na_n^2} + \frac{c'_1}{n\lambda_{k_n}^4} + \frac{c'_2 \Lambda_{k_n}^2}{n\lambda_{k_n}^2 a_n^2} + \frac{c'_3 \Lambda_{k_n}^2}{n\lambda_{k_n}^2} + 2\lambda_{k_n+1}^2. \tag{6}$$

Proof. First, let us denote

$$C_{\pi^{k_n}} = \sum_{j=1}^{k_n} \lambda_j v_j \otimes v_j,$$

$$C_{\pi^{k_n}}^{-1} = \sum_{j=1}^{k_n} \lambda_j^{-1} v_j \otimes v_j$$

and write

$$\begin{aligned}\tilde{\rho}_{n,a} - \rho = {} & (\tilde{\pi}^{k_n} D_n \tilde{C}_{n,a}^{-1} \tilde{\pi}^{k_n} - \tilde{\pi}^{k_n} D C_{\pi^{k_n}}^{-1} \tilde{\pi}^{k_n}) + (\tilde{\pi}^{k_n} D C_{\pi^{k_n}}^{-1} \tilde{\pi}^{k_n} - \pi^{k_n} D C_{\pi^{k_n}}^{-1} \pi^{k_n}) \\ & + (\pi^{k_n} D C_{\pi^{k_n}}^{-1} \pi^{k_n} - \rho). \end{aligned} \tag{7}$$

The first term may be written as

$$\begin{aligned}\tilde{\pi}^{k_n} D_n \tilde{C}_{n,a}^{-1} \tilde{\pi}^{k_n} - \tilde{\pi}^{k_n} D C_{\pi^{k_n}}^{-1} \tilde{\pi}^{k_n} &= \tilde{\pi}^{k_n} (D_n \tilde{C}_{n,a}^{-1} - D C_{\pi^{k_n}}^{-1}) \tilde{\pi}^{k_n} \\ &= \tilde{\pi}^{k_n} [(D_n - D) \tilde{C}_{n,a}^{-1} + D(\tilde{C}_{n,a}^{-1} - C_{\pi^{k_n}}^{-1})] \tilde{\pi}^{k_n} \\ &= \tilde{\pi}^{k_n} [(D_n - D) \tilde{C}_{n,a}^{-1} - D \tilde{C}_{n,a}^{-1} (\tilde{C}_{n,a} - C_{\pi^{k_n}}) C_{\pi^{k_n}}^{-1}] \tilde{\pi}^{k_n},\end{aligned}$$

hence

$$\begin{aligned}&\|\tilde{\pi}^{k_n} D_n \tilde{C}_{n,a}^{-1} \tilde{\pi}^{k_n} - \tilde{\pi}^{k_n} D C_{\pi^{k_n}}^{-1} \tilde{\pi}^{k_n}\|_{\mathscr{L}}^2 \\ &\quad \leqslant 2\|D_n - D\|_{\mathscr{L}}^2 \|\tilde{C}_{n,a}^{-1}\|_{\mathscr{L}}^2 + 2\|D\|_{\mathscr{L}}^2 \|\tilde{C}_{n,a}^{-1}\|_{\mathscr{L}}^2 \|C_{\pi^{k_n}}^{-1}\|_{\mathscr{L}}^2 \|\tilde{C}_{n,a} - C_{\pi^{k_n}}\|_{\mathscr{L}}^2 \\ &\quad \leqslant 2a_n^{-2}\|D_n - D\|_{\mathscr{L}}^2 + 2a_n^{-2}\|D\|_{\mathscr{L}}^2 \|C_{\pi^{k_n}}^{-1}\|_{\mathscr{L}}^2 \|\tilde{C}_{n,a} - C_{\pi^{k_n}}\|_{\mathscr{L}}^2.\end{aligned}$$

Thus, by (7),

$$\begin{aligned}E\|\tilde{\rho}_{n,a} - \rho\|_{\mathscr{L}}^2 &\leqslant 2[2a_n^{-2} E\|D_n - D\|_{\mathscr{L}}^2 + 2a_n^{-2}\|D\|_{\mathscr{L}}^2 \|C_{\pi^{k_n}}^{-1}\|_{\mathscr{L}}^2 E\|\tilde{C}_{n,a} - C_{\pi^{k_n}}\|_{\mathscr{L}}^2] \\ &\quad + 2E\|\tilde{\pi}^{k_n} D C_{\pi^{k_n}}^{-1} \tilde{\pi}^{k_n} - \pi^{k_n} D C_{\pi^{k_n}}^{-1} \pi^{k_n}\|_{\mathscr{L}}^2 + 2E\|\pi^{k_n} D C_{\pi^{k_n}}^{-1} \pi^{k_n} - \rho\|_{\mathscr{L}}^2. \qquad (8)\end{aligned}$$

The third term of the right-hand side is easily bounded from above by $2\lambda_{k_n+1}^2$.

Let us now focus on the first term. Bosq (2000, Theorem 4.8) gives

$$E\|D_n - D\|_{\mathscr{L}}^2 = \mathrm{O}\left(\frac{1}{n}\right),$$

and clearly

$$\|C_{\pi^{k_n}}^{-1}\|_{\mathscr{L}}^2 = \frac{1}{\lambda_{k_n}^2}.$$

Moreover,

$$\begin{aligned}E\|\tilde{C}_{n,a} - C_{\pi^{k_n}}\|_{\mathscr{L}}^2 &\leqslant 2E(\|\tilde{C}_{n,a} - \tilde{C}_n\|_{\mathscr{L}}^2 \mathbf{1}_{\tilde{C}_{n,a} \neq \tilde{C}_n}) + 2E(\|\tilde{C}_n - C_{\pi^{k_n}}\|_{\mathscr{L}}^2 \mathbf{1}_{\tilde{C}_{n,a} \neq \tilde{C}_n}) \\ &\quad + E(\|\tilde{C}_{n,a} - C_{\pi^{k_n}}\|_{\mathscr{L}}^2 \mathbf{1}_{\tilde{C}_{n,a} = \tilde{C}_n}).\end{aligned}$$

Now, we find an upper bound for $P(\tilde{C}_{n,a} \neq \tilde{C}_n)$. Fortunately, the sequence (λ_{jn}) is decreasing with respect to j. Therefore,

$$P(\tilde{C}_{n,a} \neq \tilde{C}_n) = P\left(a_n > \min_{j=1,\dots,k_n} \lambda_{jn}\right),$$

$$\begin{aligned}P(\tilde{C}_{n,a} \neq \tilde{C}_n) &= P(a_n > \lambda_{k_n,n}) \\ &= P(\lambda_{k_n,n} - \lambda_{k_n} < a_n - \lambda_{k_n}) \\ &\leqslant P(|\lambda_{k_n,n} - \lambda_{k_n}| \geqslant (1-\beta)\lambda_{k_n}) \\ &\leqslant P\left(\sup_{j=1,\dots,k_n} |\lambda_{jn} - \lambda_j| \geqslant (1-\beta)\lambda_{k_n}\right) \\ &\leqslant P(\|C_n - C\|_{\mathscr{L}} \geqslant (1-\beta)\lambda_{k_n}) \\ &\leqslant \frac{K}{n(1-\beta)^2 \lambda_{k_n}^2}\end{aligned}$$

with a constant $K > 0$, applying the Chebychev inequality and knowing that

$$E\|C_n - C\|^2_{\mathscr{L}} = \mathrm{O}\left(\frac{1}{n}\right).$$

Note that

$$\|\tilde{C}_{n,a} - \tilde{C}_n\|^2_{\mathscr{L}} \leqslant a_n^2$$

and that

$$\begin{aligned} E\|\tilde{C}_n - C_{\pi^{k_n}}\|^2_{\mathscr{L}} &= E\left\|\sum_{j=1}^{k_n} \lambda_{jn} v_{jn} \otimes v_{jn} - \sum_{j=1}^{k_n} \lambda_j v_j \otimes v_j\right\|^2_{\mathscr{L}} \\ &\leqslant 2E\left\|\sum_{j=1}^{k_n} \lambda_{jn} v_{jn} \otimes v_{jn} - \sum_{j=1}^{k_n} \lambda_j v_{jn} \otimes v_{jn}\right\|^2_{\mathscr{L}} + 2E\left\|\sum_{j=1}^{k_n} \lambda_j v_{jn} \otimes v_{jn} - \sum_{j=1}^{k_n} \lambda_j v_j \otimes v_j\right\|^2_{\mathscr{L}} \\ &\leqslant 2E\left\|\sum_{j=1}^{k_n} (\lambda_{jn} - \lambda_j) v_{jn} \otimes v_{jn}\right\|^2_{\mathscr{L}} + 2E\left\|\sum_{j=1}^{k_n} \lambda_j (v_{jn} \otimes v_{jn} - v_j \otimes v_j)\right\|^2_{\mathscr{L}}. \end{aligned}$$

But

$$E\left\|\sum_{j=1}^{k_n} (\lambda_{jn} - \lambda_j) v_{jn} \otimes v_{jn}\right\|^2_{\mathscr{L}} \leqslant E \sup_{j \geqslant 1} |\lambda_{jn} - \lambda_j|^2 \leqslant E\|C_n - C\|^2_{\mathscr{L}} = \mathrm{O}\left(\frac{1}{n}\right),$$

by Bosq (2000, Theorem 4.1), and

$$v_{jn} \otimes v_{jn} - v_j \otimes v_j = v_{jn} \otimes v_{jn} - v'_{jn} \otimes v'_{jn} = (v_{jn} - v'_{jn}) \otimes v_{jn} + v'_{jn} \otimes (v_{jn} - v'_{jn}),$$

so

$$\begin{aligned} E\left\|\sum_{j=1}^{k_n} \lambda_j (v_{jn} \otimes v_{jn} - v_j \otimes v_j)\right\|^2_{\mathscr{L}} &\leqslant 2E\left\|\sum_{j=1}^{k_n} \lambda_j (v_{jn} - v'_{jn}) \otimes v_{jn}\right\|^2_{\mathscr{L}} + 2E\left\|\sum_{j=1}^{k_n} \lambda_j v'_{jn} \otimes (v_{jn} - v'_{jn})\right\|^2_{\mathscr{L}} \\ &\leqslant 4 \sup_{j=1,\dots,k_n} |\lambda_j|^2 E\|v_{jn} - v'_{jn}\|^2_{\mathscr{L}} \leqslant 4|\lambda_1|^2 E \sup_{j=1,\dots,k_n} \|v_{jn} - v'_{jn}\|^2_{\mathscr{L}} \\ &\leqslant 4\,|\lambda_1|^2 \frac{8\Lambda^2_{k_n}}{n} nE\|C_n - C\|^2_{\mathscr{L}}, \end{aligned}$$

by Bosq (2000, Lemma 4.3). Accordingly by Bosq (2000, Theorem 4.1),

$$E\left\|\sum_{j=1}^{k_n} \lambda_j (v_{jn} \otimes v_{jn} - v_j \otimes v_j)\right\|^2_{\mathscr{L}} = \mathrm{O}\left(\frac{\Lambda^2_{k_n}}{n}\right).$$

For the second term of (8), we write

$$
\begin{aligned}
E\|\tilde{\pi}^{k_n} DC_{\pi^{k_n}}^{-1} \tilde{\pi}^{k_n} - \pi^{k_n} DC_{\pi^{k_n}}^{-1} \pi^{k_n}\|_{\mathscr{L}}^2 &\leqslant 2E\|\tilde{\pi}^{k_n} DC_{\pi^{k_n}}^{-1} \tilde{\pi}^{k_n} - \tilde{\pi}^{k_n} DC_{\pi^{k_n}}^{-1} \pi^{k_n}\|_{\mathscr{L}}^2 \\
&\quad + 2E\|\tilde{\pi}^{k_n} DC_{\pi^{k_n}}^{-1} \pi^{k_n} - \pi^{k_n} DC_{\pi^{k_n}}^{-1} \pi^{k_n}\|_{\mathscr{L}}^2 \\
&\leqslant 2E\|\tilde{\pi}^{k_n} DC_{\pi^{k_n}}^{-1} (\tilde{\pi}^{k_n} - \pi^{k_n})\|_{\mathscr{L}}^2 \\
&\quad + 2E\|(\tilde{\pi}^{k_n} - \pi^{k_n}) DC_{\pi^{k_n}}^{-1} \pi^{k_n}\|_{\mathscr{L}}^2 \\
&\leqslant \frac{L}{\lambda_{k_n}^2} E\|\tilde{\pi}^{k_n} - \pi^{k_n}\|_{\mathscr{L}}^2
\end{aligned}
$$

with a constant $L > 0$. Finally, notice that by a similar calculus as previously,

$$
E\|\tilde{\pi}^{k_n} - \pi^{k_n}\|_{\mathscr{L}}^2 = E\left\|\sum_{j=1}^{k_n} v_{jn} \otimes v_{jn} - v'_{jn} \otimes v'_{jn}\right\|_{\mathscr{L}}^2 = \mathrm{O}\left(\frac{\Lambda_{k_n}^2}{n}\right).
$$

Consequently, (8) entails

$$
E\|\tilde{\rho}_{n,a} - \rho\|_{\mathscr{L}}^2 \leqslant \frac{c'_0}{n a_n^2} + \frac{c'_1}{n \lambda_{k_n}^4} + \frac{c'_2 \Lambda_{k_n}^2}{n \lambda_{k_n}^2 a_n^2} + \frac{c'_3 \Lambda_{k_n}^2}{n \lambda_{k_n}^2} + 2\lambda_{k_n+1}^2
$$

with positive constants c'_i . □

Theorem 2. *Suppose that* (H_1)–(H_3) *hold, and that there exist* $\alpha > 0$, $0 < \beta < 1$, $\varepsilon < 1/2$ *and* $\gamma \geqslant 1$ *such that*

$$
\alpha \frac{\lambda_{k_n}^{\gamma}}{n^{\varepsilon}} \leqslant a_n \leqslant \beta \lambda_{k_n},
$$

then

$$
E\|\tilde{\rho}_{n,a} - \rho\|_{\mathscr{L}}^2 = \mathrm{O}\left(\frac{\Lambda_{k_n}^2}{n^{(1-2\varepsilon)} \lambda_{k_n}^{2(1+\gamma)}}\right) + \mathrm{O}(\lambda_{k_n}^2).
$$

Proof. It is an easy consequence of (6), using the inequalities $\alpha \lambda_{k_n}^{\gamma}/n^{\varepsilon} \leqslant a_n$ and $\lambda_{k_n+1} \leqslant \lambda_{k_n}$. □

Remark 3. The optimal choice of λ_{k_n} is such that

$$
\lambda_{k_n}^2 = \frac{c' \Lambda_{k_n}^2}{n^{(1-2\varepsilon)} \lambda_{k_n}^{2+2\gamma}} \quad \text{i.e.} \quad \lambda_{k_n}^{4+2\gamma} = \frac{c' \Lambda_{k_n}^2}{n^{(1-2\varepsilon)}}, \quad c' > 0. \tag{9}
$$

The rate of convergence in quadratic mean is then of order

$$
\lambda_{k_n}^2 \asymp \left(\frac{\Lambda_{k_n}^2}{n^{(1-2\varepsilon)}}\right)^{1/(2+\gamma)}.
$$

Example 3. If $\lambda_j = a r^j$, where $a > 0$ and $0 < r < 1$, by (9), we get

$$
r^{(6+2\gamma)k_n} = \frac{d'}{n^{(1-2\varepsilon)}}, \quad d' > 0
$$

which yields

$$k_n = \left\lfloor \frac{\ln d' - (1-2\varepsilon)\ln n}{(6+2\gamma)\ln r} \right\rfloor.$$

The rate of convergence in quadratic mean is then of order

$$r^{-2(1-2\varepsilon)\ln n/(6+2\gamma)\ln r} = n^{-(1-2\varepsilon)/(\gamma+3)}.$$

Example 4. If $\lambda_j = aj^{-\delta}$, where $a > 0$ and $\delta > 1$, a few calculations yield

$$\Lambda_{k_n}^2 \approx M k_n^{2(\delta+1)}, \quad M \geqslant 0,$$

and by (9), we get

$$k_n = \lfloor e' n^{(1-2\varepsilon)/[2\delta(\gamma+3)+2]} \rfloor, \quad e' > 0.$$

The rate of convergence in quadratic mean is then of order $k_n^{-2\delta}$, i.e.

$$n^{-\delta(1-2\varepsilon)/[\delta(\gamma+3)+1]}.$$

Remark 4. When $\varepsilon = 0$ and in the most favorable case where $\gamma = 1$, the rate of convergence in Example 4 is of order $n^{-\delta/(4\delta+1)}$ and therefore asymptotically of order $n^{-1/4}$ as $\delta \to \infty$, which is the rate of convergence in Example 3.

References

Bosq, D., 1991. Modelization, non-parametric estimation and prediction for continuous time processes. In: Roussas, G. (Ed.), Nonparametric Functional Estimation and Related Topics, NATO, ASI Series, pp. 509–529.

Bosq, D., 2000. Linear Processes in Function Spaces, Lecture Notes in Statistics. Springer, Berlin.

Besse, P., Cardot, H., 1996. Approximation spline de la prévision d'un processus fonctionnel autorégressif d'ordre 1. Rev. Canad. Statist./Canad. J. Statist. 24, 467–487.

Besse, P., Cardot, H., Stephenson, D., 2000. Autoregressive forecasting of some functional climatic variations. Scand. J. Statist. 27 (4), 673–687.

Cardot, H., Ferraty, F., Sarda, P., 1999. Functional linear model. Statist. Probab. Lett. 45, 11–22.

Cavallini, A., Montanari, G.C., Loggini, M., Lessi, O., Cacciari, M., 1994. Nonparametric prediction of harmonic levels in electrical network Proceed. IEEE ICHPS 6, Bologna.

Guillas, S., 2000. Non-causalité et discrétisation fonctionnelle, théorèmes limites pour un processus ARHX(1). C. R. Acad. Sci. Paris, Ser. I t. 331, 91–94.

Hoerl, A.E., Kennard, R.W., 1970. Ridge regression: biased estimation for nonorthogonal problems. Technometrics 8, 27–51.

Mas, A., 1999. Normalité asymptotique de l'estimateur empirique de l'opérateur d'autocorrélation d'un processus ARH(1). C. R. Acad. Sci. Paris, Ser. I t. 329, 899–902.

Mourid, T., 1995. Contribution à la statistique des processus autorégressifs à temps continu. Thesis, Université Paris VI.