ELSEVIER

Statistics & Probability Letters 45 (1999) 11–22

STATISTICS & PROBABILITY LETTERS

www.elsevier.nl/locate/stapro

Functional linear model

Hervé Cardot[a], Frédéric Ferraty[b], Pascal Sarda[b,*]

[a] *Unité Biométrie et Intelligence Artificielle, INRA, Toulouse, BP 27, 31326 Castanet-Tolosan Cédex, France*

[b] *Laboratoire de Statistique et Probabilités, UMR CNRS C5583, Université Paul Sabatier, 118, route de Narbonne, 31062 Toulouse Cedex, France*

Received June 1998; received in revised form December 1998

Abstract

In this paper, we study a regression model in which explanatory variables are sampling points of a continuous-time process. We propose an estimator of regression by means of a Functional Principal Component Analysis analogous to the one introduced by Bosq [(1991) NATO, ASI Series, pp. 509–529] in the case of Hilbertian AR processes. Both convergence in probability and almost sure convergence of this estimator are stated. © 1999 Elsevier Science B.V. All rights reserved

Keywords: Functional linear model; Functional data analysis; Hilbert spaces; Convergence

1. Introduction

Classical regression models, such as generalized linear models, may be inadequate in some statistical studies: it is the case when explanatory variables are digitized points of a curve. Examples can be found in different fields of application such as chemometrics (Frank and Friedman, 1993), linguistic (Hastie et al., 1995) and many other areas (see Hastie and Mallows, 1993; Ramsay and Silverman, 1997, among others).

In this context, Frank and Friedman (1993) describe and compare different estimation procedures – Partial Least Squares, Ridge Regression and Principal Component Regression – which take into account both the number of explanatory variables (which may exceed the sample size) and the high correlation between these variables. On the other hand, several authors (see below) have developped models which allow to describe the "functional" nature of explanatory variables.

Formally, the above situation can be described through the following *functional linear model*. Let Y be a real random variable (r.r.v.) and $X = (X(t), t \in [0, 1])$ be a continuous-time process defined on the same space

* Corresponding author.

E-mail address: sarda@cict.fr (P. Sarda)

PII: S0167-7152(99)00036-X

$(\Omega, \mathscr{A}, P)$. Assuming that $\mathbb{E}(\int_0^1 X^2(t)\,\mathrm{d}t) < \infty$, the dependence between X and Y is expressed as

$$Y = \int_0^1 \psi(t)X(t)\,\mathrm{d}t + \varepsilon, \tag{1}$$

where ψ is a square integrable function defined on $[0,1]$ and ε is an r.r.v. independent of X with zero mean and variance equal to σ^2.

Hastie and Mallows (1993) introduce an estimator of function ψ based on the minimization of a cubic spline criterion and Marx and Eilers (1996) use a smooth basis of B-splines and then introduce a difference penalty in the log-likelihood in order to derive a P-splines estimator of ψ.

Alternatively, model (1) can be generalized to the case where X is a random variable valued in a real separable Hilbert space H and the relation between X and Y can now be written as

$$Y = \Psi(X) + \varepsilon, \tag{2}$$

where Ψ is an element of H', and H' is the space of $\mathbb{R}$-valued continuous linear operators defined on H.

Following ideas from Bosq (1991) in the case of ARH processes, we propose in Section 2 below, an estimator of the operator Ψ. This estimator is based on the spectral analysis of the empirical second moment operator of X, which is then inverted in the space spanned by k_n eigenvectors associated with the k_n greatest eigenvalues. The main results are stated in Section 3, that is convergence in probability and almost sure convergence for this estimator. Computational aspects for the method are discussed in Section 4 through a simulation study. A sketch of the proofs are given in Section 5 (detailed proofs may be found in Cardot et al., 1998).

2. Definition of estimator

The inner product and norm in H are, respectively, denoted by $\langle .,. \rangle_H$ and $\| \,.\, \|_H$ and the usual norm $\| \,.\, \|_{H'}$ in H' is defined as

$$\forall T \in H', \ \|T\|_{H'} = \sup_{\|x\|_H = 1} |Tx|,$$

and satisfies

$$\forall T \in H', \ \|T\|_{H'} = \left(\sum_{i \in \mathbb{N}} (Te_i)^2 \right)^{1/2},$$

where $(e_i)_{i\in\mathbb{N}}$ is an orthonormal basis in H. Assuming that the Hilbertian variable X satisfies

$$\mathbb{E}[\|X\|_H^2] = \int_\Omega \|X(\omega)\|_H^2 \ \mathrm{d}P(\omega) < +\infty,$$

we define (cf. Grenander, 1963), from Riesz's Theorem, the *second moment operator* Γ of X by

$$\Gamma(x) = \mathbb{E}(X \otimes_H X(x)) = \mathbb{E}(\langle X, x\rangle_H X), \quad \forall x \in H.$$

The operator Γ is nuclear (and therefore is an Hilbert–Schmidt operator), self-adjoint and positive. Let us define Δ as the *cross second moment operator* between X and Y

$$\Delta(x) = \mathbb{E}(X \otimes_{H'} Y(x)) = \mathbb{E}(\langle X, x\rangle_H Y), \quad \forall x \in H.$$

It is easy to see that we have by relation (2)

$$\Delta = \Psi\Gamma. \tag{3}$$

In general, the inverse of Γ does not exist and even if it does, since Γ is nuclear, its inverse is not bounded when H is a set with infinite dimension. Direct estimation of Γ^{-1} is then problematic. Nevertheless, in order to estimate Ψ, one can think of projecting observations in a subspace of H with finite dimension (depending on n).

Let (X_i, Y_i), $i = 1, \ldots, n$, be a sample from (X, Y). Empirical versions of operators Γ and Δ are defined by

$$\Gamma_n = \frac{1}{n} \sum_{i=1}^{n} X_i \otimes_H X_i, \tag{4}$$

$$\Delta_n = \frac{1}{n} \sum_{i=1}^{n} X_i \otimes_{H'} Y_i. \tag{5}$$

Let us note $\hat{\lambda}_1 \geqslant \hat{\lambda}_2 \geqslant \cdots \geqslant \hat{\lambda}_n \geqslant 0 = \hat{\lambda}_{n+1} = \ldots$, the eigenvalues of Γ_n and $\hat{V}_1, \hat{V}_2, \ldots$, orthonormal eigenvectors associated with these eigenvalues. Let $(k_n)_{n \in \mathbb{N}^*}$ be a sequence of positive integers such that $\lim_{n\to\infty} k_n = +\infty$ with $k_n \leqslant n$ and $\hat{H}_{k_n}$ be the space spanned by $\{\hat{V}_j,\ j = 1, \ldots, k_n\}$. In H, let $\hat{\Pi}_{k_n}$ be the orthogonal projection on this subspace

$$\hat{\Pi}_{k_n} = \sum_{j=1}^{k_n} \hat{V}_j \otimes_H \hat{V}_j.$$

If we suppose $\hat{\lambda}_{k_n} > 0$, we define an estimator $\hat{\Psi}_{k_n}$ of Ψ as

$$\hat{\Psi}_{k_n} = \Delta_n \hat{\Pi}_{k_n} (\hat{\Pi}_{k_n} \Gamma_n \hat{\Pi}_{k_n})^{-1}. \tag{6}$$

Remark. Projecting observations onto the space $\hat{H}_{k_n}$ spanned by eigenvectors of Γ_n leads us to an "optimal" linear representation of X_i with respect to the explained variance (see Dauxois et al., 1982).

3. Main results

In order to state the main results of the paper, let us introduce the following condition:

(H$_0$) $\quad \hat{\lambda}_1 > \hat{\lambda}_2 > \cdots > \hat{\lambda}_{k_n} > 0$ a.s.,

which insures almost surely that $\hat{\Pi}_{k_n} \Gamma_n \hat{\Pi}_{k_n}$ is regular and its eigenvectors are identifiable. Let us note $(\lambda_j)_{j \in \mathbb{N}^*}$ the sequence of decreasing eigenvalues of Γ and let us define

$$a_j = \frac{2\sqrt{2}}{\lambda_1 - \lambda_2} \quad \text{if } j = 1,$$

$$a_j = \frac{2\sqrt{2}}{\min(\lambda_{j-1} - \lambda_j, \lambda_j - \lambda_{j+1})} \quad \text{if } j \neq 1.$$

Theorem 3.1. *Suppose that* (H$_0$) *and the following hypotheses are satisfied*:

(H$_1$) $\quad \lambda_1 > \lambda_2 > \cdots > 0,$

(H$_2$) $\quad \mathbb{E}\|X\|_H^4 < +\infty,$

(H$_3$) $\quad \lim_{n\to+\infty} n\lambda_{k_n}^4 = +\infty,$

$$\lim_{n\to+\infty} \frac{n\lambda_{k_n}^2}{(\sum_{j=1}^{k_n} a_j)^2} = +\infty.$$

Then

$$\| \hat{\Psi}_{k_n} - \Psi \|_{H'} \underset{n\to+\infty}{\longrightarrow} 0 \quad \textit{in probability.}$$

Theorem 3.2. *Suppose that* (H_0), (H_1) *and the following hypotheses are satisfied*:

$$(H_4) \qquad \|X\|_H \leqslant c_1 \quad a.s.,$$

$$(H_5) \qquad |\varepsilon| \leqslant c_2 \quad a.s.,$$

$$(H_6) \qquad \lim_{n\to+\infty} \frac{n\lambda_{k_n}^4}{\log n} = +\infty,$$

$$\lim_{n\to+\infty} \frac{n\lambda_{k_n}^2}{(\sum_{j=1}^{k_n} a_j)^2 \log n} = +\infty.$$

Then

$$\|\hat{\Psi}_{k_n} - \Psi\|_{H'} \underset{n\to+\infty}{\longrightarrow} 0, \quad a.s.$$

Remark. If we have $k_n = o(\log n)$ and

$$\lambda_j = ar^j \text{ with } a > 0,\ r \in]0,1[$$

or

$$\lambda_j = aj^{-\gamma} \text{ with } a > 0,\ \gamma > 1,$$

then (H_3) and (H_6) are satisfied.

In other words, if the eigenvalues of Γ are geometrically or exponentially decreasing, we may have Theorems 3.1 and 3.2 satisfied so long as the sequence k_n tends slowly enough to infinity. The same kind of hypotheses are also introduced in Bosq (1991) or in Cardot (1998) and they allow them to obtain rates of convergence.

4. A simulation study

We have simulated samples (X_i, Y_i), $i = 1, \ldots, n$, from model (1) in which $X(t)$ is a Brownian motion defined on $[0,1]$, ε is normal with mean 0 and variance 0.2 $\operatorname{var}(\Psi(X))$. The Hilbert space H is $L^2[0,1]$ and the eigenelements of the covariance operator of X are known to be (see Ash and Gardner, 1975):

$$\lambda_j = \frac{1}{(j-0.5)^2\pi^2}, \qquad V_j(t) = \sqrt{2}\sin\{(j-0.5)\pi t\}, \quad t \in [0,1],\ j = 1,2,\ldots$$

In that case, assumptions (H_3) (respectively (H_6)) on the sequence of eigenvalues in Theorem 3.1 (respectively Theorem 3.2) are fulfilled provided that the dimension k_n tends slowly enough to infinity, i.e. satisfying the constraint $k_n = o(\log n)$ (see the remark at the end of Section 3).

We made simulations for two different functions ψ:

- $\psi_1(t) = \sin(\pi t/2) + 0.5\sin(3\pi t/2) + 0.25\sin(5\pi t/2)$.
- $\psi_2(t) = \sin(4\pi t)$.

In the first case, the function ψ_1 is a linear combination of eigenfunctions associated with the three greatest eigenvalues of Γ, so that the best dimension k_n should be 3. We tried several sample sizes in each case:

Table 1
Quadratic error for the estimate of ψ_1

k_n	$n = 50$	$n = 200$	$n = 1000$
2	3.33	6.58	2.99
3	5.46	3.93	**1.76**
4	11.27	3.93	1.79
5	10.5	3.92	1.96
6	74.3	3.97	3.72

Table 2
Quadratic error for the estimate of ψ_2

k_n	$n = 50$	$n = 200$	$n = 1000$
4	81.8	40.8	15.76
5	9.9	18.14	1.9
6	13.6	6.02	1.02
7	15	5.58	0.92
8	19.2	4.38	0.49
9	51.6	7.64	0.54
10	53	11.62	1.71

$n = 50, 200, 1000$. To deal practically with the Brownian random functions $X_i(t)$, their sample paths were discretized by 100 points equispaced in $[0, 1]$.

The aim of our study is to look at the best dimension k_n for the estimation procedure and so we have considered the following error criterion:

$$R(\hat{\psi}_{k_n}) = \frac{\int_0^1 (\psi(t) - \hat{\psi}_{k_n}(t))^2 \, \mathrm{d}t}{\int_0^1 \psi^2(t) \, \mathrm{d}t}.$$

Table 1 (resp. Table 2) gives the quadratic errors of estimators of ψ_1 (resp. ψ_2) for each sample size and different dimensions k_n. In each case, one can notice that $R(\hat{\psi}_{k_n})$ looks like a convex function of dimension k_n and a too large k_n gives bad estimates of ψ by increasing the variance of the estimate. Also, it appears that, for the first example, the best dimension selected for the estimation procedure is reasonably close to the theoretical "optimal" dimension. This last point illustrates the good behaviour of our estimator. In real life study, this quadratic criterion error cannot be computed and on the other hand, it is clear from the above simulation that the quality of the estimator depends considerably on the choice of the dimension value, k_n. A data-driven selection method such as penalized cross validation may be used in that case (see Vieu, 1995, which uses such a criterion for the order choice in nonlinear autoregressive models).

We have drawn the estimates $\hat{\psi}_{k_n}$ of function ψ_1. Fig. 1 shows the good performance of the estimation procedure for reasonably large sample size. For smaller sample size the estimator shows rough features even if the general form of the function is restituted. We think that this aspect of the estimator could be corrected by the introduction of a preliminary smoothing procedure such as in Besse and Cardot (1996). We will investigate this topic in a further study.

5. Proof of theorems

Let $(V_j)_{j\in\mathbb{N}^*}$ be a sequence of orthonormal eigenvectors associated with $(\lambda_j)_{j\in\mathbb{N}^*}$ and let us define in H the operator Ψ_{k_n} as the "theoretical" version of $\hat{\Psi}_{k_n}$

$$\Psi_{k_n} = \Delta \Pi_{k_n} (\Pi_{k_n} \Gamma \Pi_{k_n})^{-1}, \tag{7}$$

where Π_{k_n} is the orthogonal projection onto the space H_{k_n} spanned by $V_1, \ldots, V_{k_n}$. First of all, let us remark that

$$\|\Psi - \hat{\Psi}_{k_n}\|_{H'} \leqslant \|\Psi - \Psi_{k_n}\|_{H'} + \|\Psi_{k_n} - \hat{\Psi}_{k_n}\|_{H'}. \tag{8}$$

We have for the first term on the right side of inequality (8)

$$\|\Psi - \Psi_{k_n}\|_{H'}^2 = \sum_{j=1}^{\infty} |(\Psi - \Psi_{k_n})(V_j')|^2 = \sum_{j>k_n} |\Psi(V_j')|^2,$$

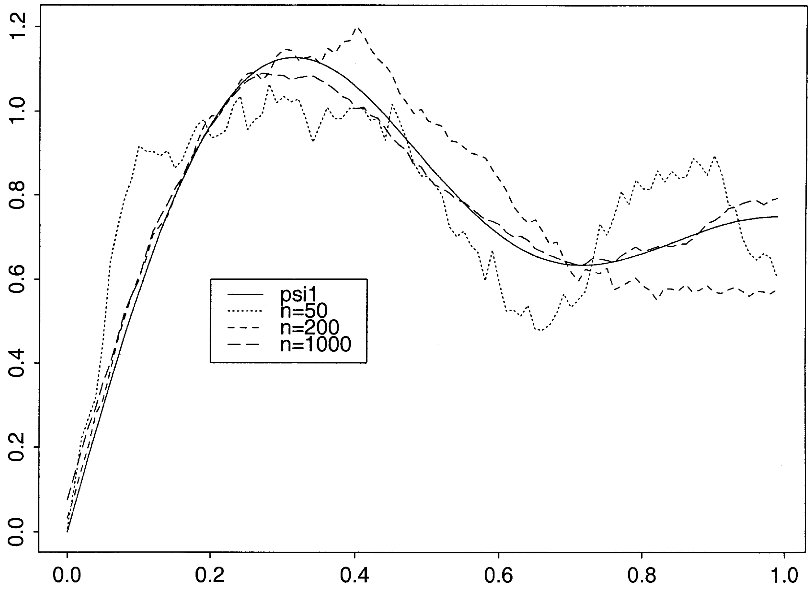

Fig. 1.

where

$$V_j' = (\mathrm{sign}\langle \hat{V}_j, V_j\rangle_H)V_j, \quad j \geqslant 1.$$

Since $\Psi \in H'$, we get

$$\|\Psi - \Psi_{k_n}\|_{H'} \underset{n\to+\infty}{\longrightarrow} 0. \tag{9}$$

We derive in the following lemma an upper bound for $\|\Psi_{k_n} - \hat{\Psi}_{k_n}\|_{H'}$. Bosq (1991) proves the analogous of this lemma for an ARH(1). Let $\mathscr{H}$ be the space of Hilbert–Schmidt operators defined on H. We consider in $\mathscr{H}$ the usual Hilbert–Schmidt norm defined as

$$\|U\|_{\mathscr{H}} = \left(\sum_{i\in\mathbb{N}} \|Ue_i\|_H^2\right)^{1/2},$$

or the uniform norm defined as

$$\|U\|_\infty = \sup_{\|x\|_H} \|Ux\|_H \ (\leqslant \ \| \ U \ \|_{\mathscr{H}}).$$

Lemma 5.1.

$$\|\Psi_{k_n} - \hat{\Psi}_{k_n}\|_{H'} \leqslant \gamma_n \|\Gamma - \Gamma_n\|_\infty + \frac{1}{\hat{\lambda}_{k_n}}\|\Delta - \Delta_n\|_{H'},$$

where

$$\gamma_n = \|\Delta\|_{H'}\left\{\frac{1}{\lambda_{k_n}\hat{\lambda}_{k_n}} + 2\left(\frac{1}{\lambda_{k_n}} + \frac{1}{\hat{\lambda}_{k_n}}\right)\sum_{j=1}^{k_n} a_j\right\}.$$

Proof. We first define the following operator $\widetilde{\Gamma}_{k_n}$ in H

$$\widetilde{\Gamma}_{k_n} = \sum_{j=1}^{k_n} \lambda_j \hat{V}_j \otimes_H \hat{V}_j.$$

We have

$$\|\Psi_{k_n} - \hat{\Psi}_{k_n}\|_{H'} \leqslant \|\Delta\Pi_{k_n}\|_{H'} \|(\Pi_{k_n}\Gamma\Pi_{k_n})^{-1} - \widetilde{\Gamma}_{k_n}^{-1}\|_\infty + \|\Delta\Pi_{k_n}\widetilde{\Gamma}_{k_n}^{-1} - \hat{\Psi}_{k_n}\|_{H'}. \tag{10}$$

Since $\|\Pi_{k_n}\|_\infty = 1$, and then using Lemma 3.1 in Bosq (1991), we have

$$\begin{aligned}\|\Delta\Pi_{k_n}\|_{H'} \|(\Pi_{k_n}\Gamma\Pi_{k_n})^{-1} - \widetilde{\Gamma}_{k_n}^{-1}\|_\infty &\leqslant \frac{2\|\Delta\|_{H'}}{\lambda_{k_n}} \sum_{j=1}^{k_n} \|V_j' - \hat{V}_j\|_H \\ &\leqslant \frac{2\|\Delta\|_{H'}}{\lambda_{k_n}} \|\Gamma - \Gamma_n\|_\infty \sum_{j=1}^{k_n} a_j. \end{aligned} \tag{11}$$

The second term in (10) gives us

$$\begin{aligned}\|\Delta\Pi_{k_n}\widetilde{\Gamma}_{k_n}^{-1} - \hat{\Psi}_{k_n}\|_\infty \leqslant{}& \|\Delta\Pi_{k_n}\|_{H'} \|\widetilde{\Gamma}_{k_n}^{-1} - (\hat{\Pi}_{k_n}\Gamma_n\hat{\Pi}_{k_n})^{-1}\|_\infty \\ &+ \|\Delta_n\hat{\Pi}_{k_n} - \Delta\Pi_{k_n}\|_{H'} \|(\hat{\Pi}_{k_n}\Gamma_n\hat{\Pi}_{k_n})^{-1}\|_\infty.\end{aligned}$$

Then, by Lemma 3.1 in Bosq (1991) and since the functions $\hat{V}_j$ are orthonormal and $\|\hat{\Pi}_{k_n}\|_\infty = 1$, we find

$$\begin{aligned}(\|\widetilde{\Gamma}_{k_n}^{-1} - (\hat{\Pi}_{k_n}\Gamma_n\hat{\Pi}_{k_n})^{-1}\|_\infty)^2 &= \left(\sup_{\|x\|_H=1} \left\| \sum_{j=1}^{k_n} \left(\frac{1}{\lambda_j} - \frac{1}{\hat{\lambda}_j}\right) \hat{V}_j \otimes_H \hat{V}_j(x) \right\|_H \right)^2 \\ &\leqslant \sup_{\|x\|_H=1} \sum_{j=1}^{k_n} \frac{(\lambda_j - \hat{\lambda}_j)^2}{(\lambda_j\hat{\lambda}_j)^2} \|\hat{V}_j \otimes_H \hat{V}_j(x)\|_H^2 \\ &= \frac{\|\Gamma - \Gamma_n\|_\infty^2}{(\lambda_{k_n}\hat{\lambda}_{k_n})^2}.\end{aligned} \tag{12}$$

It is easy to see that

$$\|(\hat{\Pi}_{k_n}\Gamma_n\hat{\Pi}_{k_n})^{-1}\|_\infty = \hat{\lambda}_{k_n}^{-1}. \tag{13}$$

Finally, with the same arguments as above, we have

$$\|\Delta_n\hat{\Pi}_{k_n} - \Delta\Pi_{k_n}\|_{H'} \leqslant 2\|\Delta\|_{H'} \|\Gamma - \Gamma_n\|_\infty \sum_{j=1}^{k_n} a_j + \|\Delta - \Delta_n\|_{H'}. \tag{14}$$

Using (11)–(14) in (10) gives us the lemma. □

5.1. Convergence in probability

The following lemma gives us the mean square convergence for operators Γ_n and Δ_n.

Lemma 5.2. *If X satisfied* (H_2) *then*

$$\mathbb{E}\|\Gamma - \Gamma_n\|_\infty^2 \leqslant \frac{\mathbb{E}\|X\|_H^4}{n}, \tag{15}$$

$$\mathbb{E}\|\Delta - \Delta_n\|_{H'}^2 \leqslant \frac{\|\Psi\|_{H'}^2 \mathbb{E}\|X\|_H^4}{n} + \frac{\sigma^2}{n}\mathbb{E}\|X\|_H^2. \tag{16}$$

Proof. The proof for (15) is similar to the analogous for the real case. Since $\mathbb{E}\Delta_n = \Delta$, we have

$$\mathbb{E}\|\Delta - \Delta_n\|_{H'}^2 = \mathbb{E}\|\Delta_n\|_{H'}^2 - \|\Delta\|_{H'}^2, \tag{17}$$

and with the independence of the X_i's we get

$$\mathbb{E}\|\Delta_n\|_{H'}^2 = \frac{1}{n}\sum_{j\in\mathbb{N}} \mathbb{E}(\langle X, e_j\rangle_H Y)^2 + \|\Delta\|_{H'}^2 - \frac{1}{n}\|\Delta\|_{H'}^2. \tag{18}$$

Now, we can write

$$\frac{1}{n}\sum_{j\in\mathbb{N}} \mathbb{E}(\langle X, e_j\rangle_H Y)^2 \leqslant \frac{\|\Psi\|_{H'}^2 \mathbb{E}\|X\|_H^4 + \sigma^2\mathbb{E}\|X\|_H^2}{n}. \tag{19}$$

The result is a consequence of (18) and (19). □

Let us now consider the following event:

$$E_n = \left\{ \frac{\lambda_{k_n}}{2} < \hat{\lambda}_{k_n} < \frac{3\lambda_{k_n}}{2} \right\}. \tag{20}$$

In E_n, we have with Lemma 5.1

$$\|\Psi_{k_n} - \hat{\Psi}_{k_n}\|_{H'} \leqslant \delta_n \|\Delta\|_{H'} \|\Gamma - \Gamma_n\|_\infty + \frac{2}{\lambda_{k_n}}\|\Delta - \Delta_n\|_{H'},$$

where

$$\delta_n = \frac{2}{\lambda_{k_n}^2} + \frac{6}{\lambda_{k_n}}\sum_{j=1}^{k_n} a_j.$$

It follows that

$$P(\|\Psi_{k_n} - \hat{\Psi}_{k_n}\|_{H'} > \eta, E_n) \leqslant P\left(\|\Gamma - \Gamma_n\|_\infty > \frac{\eta}{2\delta_n\|\Delta\|_{H'}} \right) + P\left(\|\Delta - \Delta_n\|_{H'} > \frac{\lambda_{k_n}\eta}{4} \right). \tag{21}$$

Then, we have

$$P(\|\Psi_{k_n} - \hat{\Psi}_{k_n}\|_{H'} > \eta, E_n) \leqslant \frac{4\delta_n^2\|\Delta\|_{H'}^2}{\eta^2}\mathbb{E}\|\Gamma - \Gamma_n\|_\infty^2 + \frac{16}{\lambda_{k_n}^2\eta^2}\mathbb{E}\|\Delta - \Delta_n\|_{H'}^2. \tag{22}$$

Otherwise, we get

$$P(\overline{E}_n) \leqslant P\left(\|\Gamma - \Gamma_n\|_\infty > \frac{\lambda_{k_n}}{2} \right)$$

$$\leqslant \frac{4}{\lambda_{k_n}^2} \mathbb{E}\|\Gamma - \Gamma_n\|_\infty^2. \tag{23}$$

We get with (22), (23) and Lemma 5.2

$$P(\|\Psi_{k_n} - \hat{\Psi}_{k_n}\|_{H'} > \eta) \leqslant \frac{4\|\Delta\|_{H'}^2 \mathbb{E}\|X\|_H^4}{\eta^2} \cdot \frac{\delta_n^2}{n}$$

$$+ \frac{16}{\eta^2}(\|\Psi\|_{H'}^2 \mathbb{E}\|X\|_H^4 + \sigma^2 \mathbb{E}\|X\|_H^2) \cdot \frac{1}{n\lambda_{k_n}^2} + 4\mathbb{E}\|X\|_H^4 \cdot \frac{1}{n\lambda_{k_n}^2}. \tag{24}$$

It suffices to use (9), (24), (H_2) and (H_3) to get the proof of Theorem 3.1.

5.2. *Almost sure convergence*

(1) We give in the following lemma, bounds for $P(\|\Gamma_n - \Gamma\|_\infty > \xi)$ and $P(\|\Delta_n - \Delta\|_{H'} > \xi)$.

Lemma 5.3. *Under* (H_4) *and* (H_5) *we have*
(a)

$$P(\|\Gamma_n - \Gamma\|_\infty > \xi) \leqslant 2\exp\left\{ -\frac{\xi^2 n}{2c_3(c_3 + c_4\xi)} \right\},$$

(b)

$$P(\|\Delta_n - \Delta\|_{H'} > \xi) \leqslant 2\exp\left\{ -\frac{\xi^2 n}{2c_1c_2(c_1c_2 + c_4\xi)} \right\},$$

where c_1 and c_2 are defined in (H_4) *and* (H_5) *and where c_3 and c_4 are positive constants.*

Proof. The lemma is a consequence of corollary from Yurinskiĭ (Yurinskiĭ, 1976, p. 491). Let us define for $1 \leqslant i \leqslant n$

$$Z_i = X_i \otimes_H X_i - \Gamma.$$

It is obvious that $\mathbb{E}(Z_i) = 0$. The hypothesis (H_4) implies that

$$\|Z_i\|_{\mathscr{H}} \leqslant \|X_i\|_H^2 + \mathbb{E}(\|X_i\|_H^2)$$

$$\leqslant 2c_1^2 \quad \text{a.s.}$$

This last inequality implies with $c_3 = 2c_1^2$ that

$$\mathbb{E}(\|Z_i\|_{\mathscr{H}}^m) \leqslant c_3^m \leqslant \frac{m!}{2} b_i^2 c^{m-2}, \quad \forall m \geqslant 2,$$

with $b_i = c_3$. Let $B_n^2 = \sum_{i=1}^n b_i^2 = nc_3^2$; applying now Yurinskiĭ's result to $(Z_i)_{i=1,\dots,n}$, we get part (a) of lemma since for $\xi > 0$

$$\begin{aligned} P(\|\Gamma_n - \Gamma\|_\infty > \xi) &\leqslant P(\|\Gamma_n - \Gamma\|_{\mathscr{H}} > \xi) \\ &= P\left(\left\|\sum_{i=1}^n Z_i\right\|_{\mathscr{H}} > \frac{\xi\sqrt{n}}{c_3} B_n\right) \\ &\leqslant 2\exp\left\{-\frac{\xi^2 n}{2c_3^2\left(1 + c_4\frac{\xi}{c_3}\right)}\right\}. \end{aligned}$$

Part (b) can be shown in the same way using Yurinskiĭ's corollary for the sequence $(U_i)_{i=1,\dots,n}$ defined as

$$U_i = X_i \otimes_{H'} \varepsilon_i, \quad i = 1, \dots, n. \qquad \square$$

(2) Lemma (5.3) allows us to write

$$P\left(\|\Gamma_n - \Gamma\|_\infty > \frac{\zeta}{2\delta_n\|\Delta\|_{H'}}\right) \leqslant 2\exp\left\{-\frac{\zeta^2}{8\|\Delta\|_{H'}^2 c_3(c_3 + \frac{c_4\zeta}{2\delta_n\|\Delta\|_{H'}})}.\frac{n}{\delta_n^2}\right\}.$$

Then, we get

$$P\left(\|\Gamma_n - \Gamma\|_\infty > \frac{\zeta}{2\delta_n\|\Delta\|_{H'}}\right) \leqslant 2\exp\left\{-A_\zeta\frac{n}{\delta_n^2}\right\}, \tag{25}$$

where A_ζ is a positive constant independent of n since

$$\forall n \in \mathbb{N}^*, \quad \delta_n > \frac{1}{\lambda_1^2}.$$

Analogously, we have

$$P\left(\|\Delta_n - \Delta\|_{H'} > \frac{\lambda_{k_n}\zeta}{4}\right) \leqslant 2\exp\{-B_\zeta n\lambda_{k_n}^2\}, \tag{26}$$

where B_ζ is a positive constant independent of n. Moreover with E_n defined as in (20) we have

$$P(\overline{E_n}) \leqslant 2\exp\left\{-\frac{n\lambda_{k_n}^2}{4c_3(2c_3 + c_4\lambda_{k_n})}\right\}.$$

This implies that

$$P(\overline{E_n}) \leqslant 2\exp\{-Cn\lambda_{k_n}^2\}, \tag{27}$$

where C is a positive constant independent of n. Finally, with (25)–(27) and using decomposition (21), we obtain the following inequality:

$$P(\|\Psi_{k_n} - \hat{\Psi}_{k_n}\|_{H'} > \zeta) \leqslant 2\left\{ \underbrace{\exp\left\{-A_\zeta \frac{n}{\delta_n^2}\right\}}_{u_n} + \underbrace{\exp\{-B_\zeta n\lambda_{k_n}^2\}}_{v_n} + \underbrace{\exp\{-Cn\lambda_{k_n}^2\}}_{w_n} \right\}.$$

It suffices now to show that u_n, v_n and w_n are general terms of a convergent series. Let us remark that

$$-\frac{\log u_n}{\log n} = \frac{A_\zeta n}{\delta_n^2 \log n}.$$

Now,

$$\frac{\delta_n^2 \log n}{n} = 4\frac{\log n}{n\lambda_{k_n}^4} + 36\frac{(\sum_{j=1}^{k_n} a_j)^2 \log n}{n\lambda_{k_n}^2} + 24\frac{(\sum_{j=1}^{k_n} a_j)\log n}{n\lambda_{k_n}^3},$$

and with (H_6), we get

$$\lim_{n\to+\infty} \frac{\log n}{n\lambda_{k_n}^4} = 0 \quad \text{and} \quad \lim_{n\to+\infty} \frac{(\sum_{j=1}^{k_n} a_j)^2 \log n}{n\lambda_{k_n}^2} = 0,$$

which implies that

$$\lim_{n\to+\infty} \frac{(\sum_{j=1}^{k_n} a_j)\log n}{n\lambda_{k_n}^3} = 0,$$

and also the convergence of series u_n since

$$\lim_{n\to+\infty} -\frac{\log u_n}{\log n} = +\infty.$$

The result for sequences v_n and w_n are obtained in the same way. We then get

$$\sum_{n\in\mathbb{N}^*} P(\|\Psi_{k_n} - \hat{\Psi}_{k_n}\|_{H'} > \zeta) < +\infty,$$

which gives us with Borel–Cantelli Lemma

$$\|\Psi_{k_n} - \hat{\Psi}_{k_n}\|_{H'} \underset{n\to+\infty}{\longrightarrow} 0, \quad \text{a.s.} \tag{28}$$

The proof of Theorem 3.2 is complete with (8), (9) and (28). □

References

Ash, R.B., Gardner, M.F., 1975. Topics in Stochastic Processes. Academic Press, New York.

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

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.

Cardot, H., 1998. Convergence du lissage spline de la prévision des processus autorégressifs fonctionnels. C.R. Acad. Sci. Paris, Sér. I, t. 326, 755–758.

Cardot, H., Ferraty, F., Sarda, P., 1998. Modèle linéaire fonctionnel. Publ. Lab. Statist. Probab. 04:98, Toulouse, France.

Dauxois, J., Pousse, A., Romain, Y., 1982. Asymptotic theory for the principal component analysis of a random vector function: some applications to statistical inference. J. Multivariate Anal. 12, 136–154.

Frank, I.E., Friedman, J.H., 1993. A statistical view of some chemometrics regression tools. Technometrics 35, 109–148.

Grenander, U., 1963. Probabilities on Algebraic Structures. Almqvist & Wiksell, Stockholm.

Hastie, T., Buja, A., Tibshirani, R., 1995. Penalized discriminant analysis. Ann. Statist. 23, 73–102.

Hastie, T., Mallows, C., 1993. A discussion of “A Statistical View of Some Chemometrics Regression Tools” by I.E. Frank and J.H. Friedman. Technometrics 35, 140–143.

Marx, B.D., Eilers, P.H., 1996. Generalized linear regression on sampled signals with penalized likelihood. In: Forcina, A., Marchetti, G.M., Hatzinger, R., Galmacci, G. (Eds.), Statistical Modelling, Proceedings of the Eleventh International Workshop on Statistical Modelling, Orvietto.

Ramsay, J.O., Silverman, B.W., 1997. Functional Data Analysis. Springer, Berlin.

Vieu, P., 1995. Order choice in nonlinear autoregressive models. Statistics 26, 307–328.

Yurinskiĭ, V.V., 1976. Exponential inequalities for sums of random vectors. J. Multivariate Anal. 6, 473–499.