



HAL
open science

Estimation of the noise covariance operator in functional linear regression with functional outputs

Christophe Crambes, Nadine Hilgert, Tito Manrique Chuquillanqui

► To cite this version:

Christophe Crambes, Nadine Hilgert, Tito Manrique Chuquillanqui. Estimation of the noise covariance operator in functional linear regression with functional outputs. *Statistics and Probability Letters*, 2016, 113, pp.7-15. 10.1016/j.spl.2016.02.006 . hal-01331242

HAL Id: hal-01331242

<https://hal.science/hal-01331242>

Submitted on 27 May 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License

Estimation of the noise covariance operator in functional linear regression with functional outputs

Crambes Christophe^a, Hilgert Nadine^b, Manrique Tito^{a,b}

^aIMAG, Université de Montpellier, Place Eugène Bataillon, 34095 Montpellier Cedex 5, France

^bMISTEA, INRA Montpellier, 2 Place Pierre Viala, 34060 Montpellier Cedex 2, France

Abstract

This work deals with the estimation of the noise in functional linear regression when both the response and the covariate are functional. Namely, we propose two estimators of the covariance operator of the noise. We give some asymptotic properties of these estimators, and we study their behaviour on simulations.

Keywords: functional linear regression, functional response, noise covariance operator

1. Introduction

We consider the following functional linear regression model where the functional output $Y(\cdot)$ is related to a random function $X(\cdot)$ through

$$Y(t) = \int_0^1 \mathcal{S}(t, s) X(s) ds + \varepsilon(t). \quad (1)$$

Here $\mathcal{S}(\cdot, \cdot)$ is an unknown integrable kernel: $\int_0^1 \int_0^1 |\mathcal{S}(t, s)| dt ds < \infty$, to be estimated. ε is a noise random variable, independent of X . The functional variables X , Y and ε are random functions taking values on the interval $I = [0, 1]$ of \mathbb{R} . Considering this particular interval is equivalent to considering any other interval $[a, b]$ in what follows. For the sake of clarity, we assume moreover that the random functions X and ε are centered. The case of non centered X and Y functions can be equivalently studied by adding an additive non random intercept function in model (1).

In all the sequel we consider a sample $(X_i, Y_i)_{i=1, \dots, n}$ of independent and identically distributed observations, following (1) and taking values in the

same Hilbert space $H = \mathbb{L}^2([0, 1])$, the space of all real valued square integrable functions defined on $[0, 1]$. The objective of this paper is to estimate the unknown noise covariance operator Γ_ε of ε and its trace $\sigma_\varepsilon^2 := \text{tr}(\Gamma_\varepsilon)$ from these data sets. The estimation of the noise covariance operator Γ_ε is well known in the context of multivariate multiple regression models, see for example Johnson and Wichern [1, section 7.7]. The question is a little more tricky in the context of functional data. Answering it will then make possible the construction of hypothesis testing in connection with model (1).

Functional data analysis has given rise to many theoretical results applied in various domains (economics, biology, finance, etc...). The monograph by Ramsay & Silverman [2] is a major reference that gives an overview on the subject and highlights the drawbacks of considering a multivariate point of view. Novel asymptotic developments and illustrations on simulated and real data sets are also provided in Horváth & Kokoszka [3]. We follow here the approach of Crambes & Mas [4] that studied the prediction in the model (1) revisited as:

$$Y = S X + \varepsilon, \quad (2)$$

where $S : H \rightarrow H$ is a general linear integral operator defined by $S(f)(t) = \int_0^1 \mathcal{S}(t, s) f(s) ds$ for any function f in H . The authors showed that the trace σ_ε^2 is an important constant involved in the square prediction error bound that participate to determine the convergence rate. The estimation of σ_ε^2 will thus provide details on the prediction quality in model (1).

In this context of functional linear regression, it is well known that the covariance operator of X cannot be inverted directly (see Cardot *et al.* [5]), thus a regularization is needed. In [4], it is based on the Karhunen-Loève expansion and the functional principal component analysis of the (X_i) . This approach is also often used in functional linear models with scalar output, see for example [5].

The construction of the estimator \hat{S} is introduced in Section 2. Section 3 is devoted to the estimation of Γ_ε and its trace. Two types of estimators are given. Convergence properties are established and discussed. The proofs are postponed in Section 5. The results are illustrated on simulation trials in Section 4.

2. Estimation of S

2.1. Preliminaries

We denote respectively $\langle \cdot, \cdot \rangle_H$ and $\|\cdot\|_H$ the inner product and the corresponding norm in the Hilbert space H . We shall recall that $\langle f, g \rangle_H =$

$\int_0^1 f(t)g(t)dt$, for all functions f and g in $\mathbb{L}^2([0, 1])$. In contrast, $\langle \cdot, \cdot \rangle_n$ and $\|\cdot\|_n$ stand for the inner product and the Euclidean norm in \mathbb{R}^n . The tensor product is denoted \otimes and defined by $f \otimes g = \langle g, \cdot \rangle_H f$ for any functions $f, g \in H$.

We assume that X and ε have a second moment, that is: $\mathbb{E}[\|X\|_H^2] < \infty$ and $\mathbb{E}[\|\varepsilon\|_H^2] < \infty$. The covariance operator of X is the linear operator defined on H as follows: $\Gamma := \mathbb{E}[X \otimes X]$. The cross covariance operator of X and Y is defined as $\Delta := \mathbb{E}[Y \otimes X]$. The empirical counterparts of these operators are: $\hat{\Gamma}_n := \frac{1}{n} \sum_{i=1}^n X_i \otimes X_i$ and $\hat{\Delta}_n := \frac{1}{n} \sum_{i=1}^n Y_i \otimes X_i$.

An objective of the paper is to study the trace σ_ε^2 . We thus introduce the nuclear norm defined by $\|A\|_{\mathcal{N}} = \sum_{j=1}^{+\infty} |\mu_j|$, for any operator A such that $\sum_{j=1}^{+\infty} |\mu_j| < +\infty$ where $(\mu_j)_{j \geq 1}$ is the sequence of the eigenvalues of A . We denote $\|\cdot\|_\infty$ the operator norm defined by $\|A\|_\infty = \sup_{\|u\|=1} \|Au\|$.

2.2. Spectral decomposition of Γ

It is well known that Γ is a symmetric, positive trace-class operator, and thus diagonalizable in an orthonormal basis (see for instance [6]). Let $(\lambda_j)_{j \geq 1}$ be its non-increasing sequence of eigenvalues, and $(v_j)_{j \geq 1}$ the corresponding eigenfunctions in H . Then Γ decomposes as follows:

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

For any integer k , we define $\Pi_k := \sum_{j=1}^k v_j \otimes v_j$ the projection operator on the sub-space $\langle v_1, \dots, v_k \rangle$. By projecting Γ on this sub-space, we get :

$$\Gamma|_{\langle v_1, \dots, v_k \rangle} := \Gamma \Pi_k = \sum_{j=1}^k \lambda_j v_j \otimes v_j.$$

2.3. Construction of the estimator of S

We start from the moment equation

$$\Delta = S \Gamma. \quad (3)$$

On the sub-space $\langle v_1, \dots, v_k \rangle$, the operator Γ is invertible, more precisely $(\Gamma \Pi_k)^{-1} = \sum_{j=1}^k \lambda_j^{-1} v_j \otimes v_j$. As a consequence, with equation (3) and the fact that $\Pi_k \Gamma \Pi_k = \Gamma \Pi_k$ we get, on the sub-space $\langle v_1, \dots, v_k \rangle$, $\Delta \Pi_k = (S \Pi_k) (\Gamma \Pi_k)$. We deduce that $S \Pi_k = \Delta \Pi_k (\Gamma \Pi_k)^{-1}$.

Now, taking $k = k_n$, denoting $\hat{\Pi}_{k_n} := \sum_{j=1}^{k_n} \hat{v}_j \otimes \hat{v}_j$ and the generalized inverse $\hat{\Gamma}_{k_n}^+ := (\hat{\Gamma}_n \hat{\Pi}_{k_n})^{-1}$, we are able to define the estimator of S . We have

$$\hat{\Gamma}_n = \sum_{j=1}^{\infty} \hat{\lambda}_j \hat{v}_j \otimes \hat{v}_j = \sum_{j=1}^n \hat{\lambda}_j \hat{v}_j \otimes \hat{v}_j,$$

with eigenvalues $\hat{\lambda}_1 \geq \dots \geq \hat{\lambda}_n \geq 0 = \hat{\lambda}_{n+1} = \hat{\lambda}_{n+2} = \dots \in \mathbb{R}^1$ and orthonormal eigenfunctions $\hat{v}_1, \hat{v}_2, \dots \in H$. By taking $\hat{\lambda}_{k_n} > 0$, with $k_n < n$, we define the operator $\hat{\Gamma}_{k_n} = \sum_{j=1}^{k_n} \hat{\lambda}_j \hat{v}_j \otimes \hat{v}_j$ and we get $\hat{\Gamma}_{k_n}^+ = \sum_{j=1}^{k_n} (\hat{\lambda}_j)^{-1} \hat{v}_j \otimes \hat{v}_j$. Hence we define **the estimator of S** as follows

$$\hat{S}_{k_n} = \hat{\Delta}_n \hat{\Gamma}_{k_n}^+. \quad (4)$$

Finally, the associated kernel of \hat{S}_{k_n} , estimating S , is

$$\hat{S}_{k_n}(t, s) = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^{k_n} \left[\left(\frac{1}{\hat{\lambda}_j} \int_0^1 X_i(r) \hat{v}_j(r) dr \right) Y_i(t) \hat{v}_j(s) \right]. \quad (5)$$

3. Estimation of Γ_ε and its trace

3.1. The plug-in estimator

The plug-in estimator of Γ_ε is given by

$$\hat{\Gamma}_{\varepsilon, n} := \frac{1}{n - k_n} \sum_{i=1}^n (Y_i - \hat{S}_{k_n} X_i) \otimes (Y_i - \hat{S}_{k_n} X_i) = \frac{1}{n - k_n} \sum_{i=1}^n \hat{\varepsilon}_i \otimes \hat{\varepsilon}_i. \quad (6)$$

This estimator is biased, for a fixed n , as stated in the next theorem:

Theorem 3.1. *Let $(X_i, Y_i)_{i=1, \dots, n}$ be a sample of i.i.d. observations following model (1). Let $k_n < n$ be an integer. We have*

$$\mathbb{E}[\hat{\Gamma}_{\varepsilon, n}] = \Gamma_\varepsilon + \left(\frac{n}{n - k_n} \right) S \mathbb{E} \left(\sum_{i=k_n+1}^n \hat{\lambda}_i \hat{v}_i \otimes \hat{v}_i \right) S'. \quad (7)$$

The proof of Theorem 3.1 is postponed in Section 5.1. As $\hat{\Gamma}_{k_n} = \sum_{i=1}^n \hat{\lambda}_i \hat{v}_i \otimes \hat{v}_i$ and $\hat{\Pi}_{(k_n+1):n} := \sum_{i=k_n+1}^n \hat{v}_i \otimes \hat{v}_i$, we deduce the following result:

Corollary 3.1. *We have*

$$\mathbb{E}[\hat{\Gamma}_{\varepsilon, n}] = \Gamma_\varepsilon + \left(\frac{n}{n - k_n} \right) S \mathbb{E} \left(\hat{\Pi}_{(k_n+1):n} \hat{\Gamma}_n \right) S', \quad (8)$$

where $\hat{\Pi}_{(k_n+1):n}$ is the projection on the sub-space $\langle \hat{v}_{k_n+1}, \dots, \hat{v}_n \rangle$.

Under some additional assumptions, we prove that the plug-in estimator (6) of Γ_ε is asymptotically unbiased. Let us consider the following assumptions:

(A.1) The operator S is a nuclear operator, in other words $\|S\|_{\mathcal{N}} < +\infty$.

(A.2) The variable X satisfies $\mathbb{E}\|X\|^4 < +\infty$.

(A.3) We have almost surely $\hat{\lambda}_1 > \hat{\lambda}_2 > \dots > \hat{\lambda}_{k_n} > 0$.

(A.4) We have $\lambda_1 > \lambda_2 > \dots > 0$.

Our main result is then the following.

Theorem 3.2. *Under assumptions (A.1)-(A.4), if $(k_n)_{n \geq 1}$ is a sequence such that $\lim_{n \rightarrow +\infty} k_n = +\infty$ and $\lim_{n \rightarrow +\infty} k_n/n = 0$, we have*

$$\lim_{n \rightarrow +\infty} \left\| \mathbb{E} \left(\hat{\Gamma}_{\varepsilon, n} \right) - \Gamma_\varepsilon \right\|_{\mathcal{N}} = 0. \quad (9)$$

The proof is postponed in Section 5.2. From the definition of the nuclear norm, we immediately get the following corollary:

Corollary 3.2. *Under the assumptions of Theorem 3.2, we have*

$$\lim_{n \rightarrow +\infty} \mathbb{E} \left[\text{tr} \left(\hat{\Gamma}_{\varepsilon, n} \right) \right] = \text{tr} \left(\Gamma_\varepsilon \right). \quad (10)$$

3.2. Other estimation of Γ_ε

Without loss of generality, we assume in this section that n is a multiple of 3. In formula (8), the bias of the plug-in estimator is related to $S \mathbb{E} \left(\hat{\Pi}_{(k_n+1):n} \hat{\Gamma}_n \right) S'$. Another way of estimating Γ_ε is thus to subtract an estimator of the bias to the plug-in estimator $\hat{\Gamma}_{\varepsilon, n}$. To achieve this, we split the n -sample into three sub-samples with size $m = n/3$ to keep good theoretical properties thanks to the independence of the sub-samples. As a consequence, we define

$$\check{B}_n := \hat{S}_{2k_m}^{[2]} \left(\hat{\Pi}_{(k_m+1):m}^{[1]} \hat{\Gamma}_m^{[1]} \right) \left(\hat{S}_{2k_m}^{[3]} \right)', \quad (11)$$

where the quantities with superscripts [1], [2] and [3] are respectively estimated with the first, second and third part of the sample. We use $2k_m$ eigenvalues (where $k_m \leq n/2$) in the estimation of S with the second and third sub-sample in order to avoid orthogonality between $\hat{S}_{2k_m}^{[2]}$, $\hat{S}_{2k_m}^{[3]}$ and $\hat{\Pi}_{(k_m+1):m}^{[1]} \hat{\Gamma}_m^{[1]}$.

We are now in a position to define another estimator of Γ_ε :

$$\check{\Gamma}_{\varepsilon, n} := \hat{\Gamma}_{\varepsilon, m}^{[1]} - \frac{m}{m - k_m} \check{B}_n. \quad (12)$$

The following result is established.

Theorem 3.3. *Under the assumptions of Theorem 3.2, we have*

$$\lim_{n \rightarrow +\infty} \|\mathbb{E}(\check{\Gamma}_{\varepsilon,n}) - \Gamma_{\varepsilon}\|_{\mathcal{N}} = 0. \quad (13)$$

The above result can also be written using the trace.

Corollary 3.3. *Under the assumptions of Theorem 3.2, we have*

$$\lim_{n \rightarrow +\infty} \mathbb{E}[\text{tr}(\check{\Gamma}_{\varepsilon,n})] = \text{tr}(\Gamma_{\varepsilon}). \quad (14)$$

3.3. Comments on both estimators

Subtracting an estimator of the bias to the plug-in estimator $\hat{\Gamma}_{\varepsilon,n}$ does not provide an unbiased estimator of $\Gamma_{\varepsilon,n}$. The situation is completely different to that of multivariate multiple regression models, see [1], where an unbiased estimator of the noise covariance is easily produced.

Both estimators $\hat{\Gamma}_{\varepsilon,n}$ and $\check{\Gamma}_{\varepsilon,n}$ are consistent. We can see from the proofs of Theorems 3.2 and 3.3 that $\|\mathbb{E}(\hat{\Gamma}_{\varepsilon,n}) - \Gamma_{\varepsilon}\|_{\mathcal{N}} \leq \frac{n}{n-k_n} \|S\|_{\mathcal{N}} \|S'\|_{\infty} \mathbb{E}|\hat{\lambda}_{k_n+1}|$, and $\|\mathbb{E}(\check{\Gamma}_{\varepsilon,n}) - \Gamma_{\varepsilon}\|_{\mathcal{N}} \leq 2 \frac{n}{n-3k_m} \|S\|_{\mathcal{N}} \|S'\|_{\infty} \mathbb{E}|\hat{\lambda}_{k_m+1}|$.

Number 2 in the estimation bound of $\check{\Gamma}_{\varepsilon,n}$ is due to the use of the triangle inequality. In this way, we cannot prove that subtracting the bias may improve the estimation of Γ_{ε} , nor of its trace. We will study the behavior of both estimators by simulations in the next section.

3.4. Cross validation and Generalized cross validation

Whatever the estimator, we have to choose a dimension k_n of principal components in order to compute the estimator. We chose to select it with cross validation and generalized cross validation. First, we define the usual cross validation criterion (in the framework of functional response)

$$CV(k_n) = \frac{1}{n} \sum_{i=1}^n \left\| Y_i - \hat{Y}_i^{[-i]} \right\|_H^2,$$

where $\hat{Y}_i^{[-i]}$ is the predicted value of Y_i using the whole sample except the i th observation, namely $\hat{Y}_i^{[-i]} = \hat{S}_{k_n}^{[-i]} X_i$, where $\hat{S}_{k_n}^{[-i]}$ is the estimator of the operator S using the whole sample except the i th observation. Note that the criterion is based on the residuals.

The following property allows to introduce the generalized cross validation criterion.

Proposition 3.1. We denote \mathbf{X} the matrix with size $n \times k_n$ with general term $\langle X_i, \hat{v}_j \rangle_H$ for $i = 1, \dots, n$ and $j = 1, \dots, k_n$, and $\mathbf{H} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$. Then

$$Y_i - \hat{Y}_i^{[-i]} = \frac{Y_i - \hat{Y}_i}{1 - \mathbf{H}_{ii}}, \quad (15)$$

where \hat{Y}_i is the predicted value of Y_i using the whole sample and \mathbf{H}_{ii} is the i th diagonal term of the matrix \mathbf{H} .

This proposition allows to write the expression $Y_i - \hat{Y}_i^{[-i]}$ without excluding the i th observation, and allows to get the generalized cross validation criterion, which is computationally faster than the cross validation criterion (see for example [7]). The term \mathbf{H}_{ii} can be replaced by the mean $\text{tr}(\mathbf{H})/n$. Then, after noticing that $\text{tr}(\mathbf{H}) = \text{tr}(\mathbf{I}_{k_n}) = k_n$, where \mathbf{I}_{k_n} is the identity matrix with size k_n , we get

$$GCV(k_n) = \frac{n}{(n - k_n)^2} \sum_{i=1}^n \left\| Y_i - \hat{Y}_i \right\|_H^2.$$

4. Simulations

4.1. Setting

The variable X is simulated as a standard Brownian motion on $[0, 1]$, with its Karhunen Loève expansion, given by Ash & Gardner [8]

$$X(t) = \sum_{j=1}^{\infty} \xi_j \sqrt{\lambda_j} v_j(t), \quad t \in [0, 1],$$

where the $v_j(t) := \sqrt{2} \sin((j - 1/2)\pi t)$ and $\lambda_j = \frac{1}{(j - 0.5)^2 \pi^2}$ are the the eigenfunctions and the eigenvalues of the covariance operator of X . In practice, $X(t)$ has been simulated using a truncated version with 1 000 eigenfunctions. The considered observation times are $[\frac{1}{1000}, \frac{2}{1000}, \dots, \frac{1000}{1000}]$. We simulate a sample with sizes $n = 300$ and $n = 1\,500$.

We simulate the noise ε as a Standard Brownian motion multiplied by 0.1 (ratio noise-signal = 10%). Thus the trace of the covariance operator of ε will be $\text{tr}(\Gamma_\varepsilon) = 0.005$.

Simulation 1. The operator S is $S = \Pi_{20} := \sum_{j=1}^{20} v_j \otimes v_j$, where $v_j(t) := \sqrt{2} \sin((j - 1/2)\pi t)$ are the the eigenfunctions of the covariance operator X .

Simulation 2. The operator S is the integral operator defined by $SX = \int_0^1 \mathcal{S}(t, s)X(s)ds$, where the kernel of S is $\mathcal{S}(t, s) = t^2 + s^2$.

4.2. Three estimators

We consider three different estimators of the trace of the covariance operator of ε : (i) the plug-in estimator given in (6), (ii) the corrected estimator given in (11) and (12), and (iii) the estimator $\hat{\Gamma}_{\varepsilon, n} := \hat{\Gamma}_{\varepsilon, n} - \left(\frac{n}{n-k_n}\right) \left[\hat{S}_{n, 2k_n}(\hat{\Pi}_{(k_n+1):n}\hat{\Gamma}_n)(\hat{S}_{n, 2k_n})'\right]$. The third estimator uses the whole sample when trying to remove the bias term, so it is not possible to obtain an immediate consistency result for this estimator because we do not have anymore the independence between the terms $\hat{S}_{n, 2k_n}$ and $\hat{\Pi}_{(k_n+1):n}\hat{\Gamma}_n$, but we can see its practical behaviour.

4.3. Results

We present in table 1 (simulation 1) and table 2 (simulation 2) the mean values of the trace obtained for the three estimators on $N = 100$ simulations, as well as the *CV* and *GCV* criteria. The criteria have a convex form, that allows to choose a value for k .

n	k	$CV(k)$	$GCV(k)$	$tr(\hat{\Gamma}_{\varepsilon, n})$	$tr(\check{\Gamma}_{\varepsilon, n})$	$tr(\hat{\check{\Gamma}}_{\varepsilon, n})$
n=300	16	6.67 (3.5)	6.67 (3.5)	6.32 (3.3)	5.29 (7.1)	4.79 (3.3)
	18	6.04 (3.5)	6.04 (3.5)	5.67 (3.3)	5.13 (7.2)	4.74 (3.4)
	20	5.66 (3.7)	5.66 (3.7)	5.28 (3.4)	5.06 (7.1)	4.7 (3.4)
	22	5.5698 (3.7)	5.57 (3.6)	5.16 (3.4)	5.04 (7.1)	4.67 (3.4)
	24	5.57 (3.7)	5.568 (3.7)	5.12 (3.4)	5.02 (7.1)	4.63 (3.4)
	26	5.59 (3.8)	5.59 (3.8)	5.11 (3.4)	5.02 (7.2)	4.58 (3.4)
n=1500	18	5.67 (1.7)	5.67 (1.7)	5.6 (1.7)	5.03 (2.5)	4.97 (1.7)
	20	5.15 (1.7)	5.15 (1.7)	5.08 (1.7)	5.01 (2.5)	4.96 (1.7)
	22	5.12 (1.7)	5.12 (1.7)	5.04 (1.7)	5.01 (2.6)	4.95 (1.7)
	24	5.11 (1.7)	5.11 (1.7)	5.04 (1.7)	5.01 (2.6)	4.95 (1.7)
	26	5.12 (1.7)	5.12 (1.7)	5.03 (1.7)	5 (2.6)	4.94 (1.7)
	28	5.13 (1.7)	5.13 (1.7)	5.03 (1.7)	5 (2.6)	4.93 (1.7)

Table 1: *CV* and *GCV* criteria for different values of k and mean values for the estimators of $Tr(\Gamma_\varepsilon)$ (simulation 1 with $n = 300$ and $n = 1500$). All values are given up to a factor of 10^{-3} (the standard deviation is given in brackets up to a factor of 10^{-4}).

In simulation 1, the true value of k is known ($k = 20$) and the values chosen by *CV* and *GCV* are $k = 22$ or $k = 24$. For these values of k , the best estimator is $tr(\check{\Gamma}_{\varepsilon, n})$ for $n = 300$ and $n = 1500$. The overestimation

of $tr(\hat{\Gamma}_{\varepsilon,n})$ seems to be well corrected by $tr(\check{\Gamma}_{\varepsilon,n})$, even if the usefulness of this bias removal cannot be theoretically proved. On this simulation, the estimator $tr(\hat{\Gamma}_{\varepsilon,n})$ does not behave better than the others, especially for small sample sizes.

n	k	$CV(k)$	$GCV(k)$	$tr(\hat{\Gamma}_{\varepsilon,n})$	$tr(\check{\Gamma}_{\varepsilon,n})$	$tr(\hat{\Gamma}_{\varepsilon,n})$
n=300	2	5.37 (3.6)	5.37 (3.6)	5.34 (3.6)	5.03 (6.4)	5.07 (3.2)
	4	5.17 (3.3)	5.17 (3.3)	5.11 (3.2)	5.02 (6.4)	5 (3.1)
	6	5.18 (3.2)	5.18 (3.2)	5.08 (3.2)	5 (6.5)	4.96 (3.2)
	8	5.21 (3.2)	5.21 (3.2)	5.07 (3.2)	5 (6.4)	4.93 (3.2)
	10	5.25 (3.3)	5.25 (3.3)	5.07 (3.2)	5 (6.6)	4.89 (3.2)
n=1500	2	5.28 (1.7)	5.28 (1.7)	5.28 (1.7)	5.04 (2.8)	5.05 (1.7)
	4	5.07 (1.7)	5.07 (1.7)	5.05 (1.7)	5.01 (2.6)	5.02 (1.7)
	6	5.05 (1.7)	5.05 (1.7)	5.03 (1.7)	5 (2.6)	5.01 (1.7)
	8	5.06 (1.7)	5.06 (1.7)	5.03 (1.7)	5 (2.5)	5 (1.7)
	10	5.06 (1.7)	5.06 (1.7)	5.03 (1.7)	5 (2.5)	4.99 (1.7)

Table 2: CV and GCV criteria for different values of k and mean values for the estimators of $Tr(\Gamma_\varepsilon)$ (simulation 2 with $n = 300$ and $n = 1500$). All values are given up to a factor of 10^{-3} (the standard deviation is given in brackets up to a factor of 10^{-4}).

In simulation 2, the true value of k is unknown and the value chosen by CV and GCV is $k = 4$ (for $n = 300$) or $k = 6$ (for $n = 1500$). The estimator $tr(\hat{\Gamma}_{\varepsilon,n})$ is slightly better than the two others for $n = 300$. For $n = 1500$, $tr(\hat{\Gamma}_{\varepsilon,n})$ is slightly better.

On both simulations, $tr(\hat{\Gamma}_{\varepsilon,n})$ and $tr(\check{\Gamma}_{\varepsilon,n})$ show a good estimation accuracy and are quite equivalent. From a practical point of view, $tr(\hat{\Gamma}_{\varepsilon,n})$ may be preferred as it is easy to implement. The bias removal of $tr(\check{\Gamma}_{\varepsilon,n})$ will give a more precise estimation.

5. Proofs

5.1. Proof of Theorem 3.1

We begin with preliminary lemmas.

Lemma 5.1. $\hat{S}_{k_n} = S \hat{\Pi}_{k_n} + \frac{1}{n} \left[\sum_{i=1}^n \varepsilon_i \otimes (\hat{\Gamma}_{k_n}^+ X_i) \right]$.

Proof: From the definition of the estimator $\hat{S}_{k_n} := \hat{\Delta}_n \hat{\Gamma}_{k_n}^+$, we get

$$\hat{S}_{k_n} = \left[\frac{1}{n} \sum_{i=1}^n Y_i \otimes X_i \right] \hat{\Gamma}_{k_n}^+ = \left\{ S \left[\frac{1}{n} \sum_{i=1}^n X_i \otimes X_i \right] + \frac{1}{n} \sum_{i=1}^n \varepsilon_i \otimes X_i \right\} \hat{\Gamma}_{k_n}^+,$$

and the result comes from the fact that $\hat{\Gamma}_n \hat{\Gamma}_{k_n}^+ = \hat{\Pi}_{k_n}$. \square

Lemma 5.2. *We have*

$$\sum_{i=1}^n \sum_{j=1}^n \langle \hat{\Gamma}_{k_n}^+ X_j, X_i \rangle_H^2 = n^2 k_n \quad \text{and} \quad \sum_{i=1}^n \langle \hat{\Gamma}_{k_n}^+ X_i, X_i \rangle_H = n k_n.$$

Proof: We denote A the $n \times n$ matrix defined, for $r, s \in 1, \dots, n$, by

$$A_{(r,s)} := \langle \hat{\Gamma}_{k_n}^+ X_r, X_s \rangle_H = \sum_{l=1}^{k_n} \hat{\lambda}_l^{-1} \langle X_r, \hat{v}_l \rangle_H \langle X_s, \hat{v}_l \rangle_H.$$

Let us remark that $A = \mathbf{X} \Lambda^{-1} \mathbf{X}'$, where \mathbf{X} is introduced in Proposition 3.1 and Λ is the diagonal matrix $\Lambda := \text{diag}(\hat{\lambda}_1, \dots, \hat{\lambda}_{k_n})$. We obtain

$$\sum_{i,j=1}^n \langle \hat{\Gamma}_{k_n}^+ X_j, X_i \rangle_H^2 = \text{tr}(A'A) = \text{tr}[\mathbf{X} \Lambda^{-1} (nI_n) \mathbf{X}'] = n \text{tr}[(\mathbf{X}'\mathbf{X}) \Lambda^{-1}] = n^2 k_n.$$

The second part of the lemma can be obtained in a similar way. \square

Now, coming back to the proof of Theorem 3.1, we can write

$$\hat{\Gamma}_{\varepsilon,n} = \frac{1}{n-k_n} \sum_{i=1}^n [(S - \hat{S}_{k_n})(X_i) + \varepsilon_i] \otimes [(S - \hat{S}_{k_n})(X_i) + \varepsilon_i],$$

hence we have $\mathbb{E}(\hat{\Gamma}_{\varepsilon,n}) = \mathbb{E}[P_I + P_{II} + P_{III} + P_{IV}]$ with

$$\begin{aligned} P_I &:= \frac{1}{n-k_n} \sum_{i=1}^n (S - \hat{S}_{k_n})(X_i) \otimes (S - \hat{S}_{k_n})(X_i), \\ P_{II} &:= \frac{1}{n-k_n} \sum_{i=1}^n (S - \hat{S}_{k_n})(X_i) \otimes \varepsilon_i, \\ P_{III} &:= \frac{1}{n-k_n} \sum_{i=1}^n \varepsilon_i \otimes (S - \hat{S}_{k_n})(X_i), \\ P_{IV} &:= \frac{1}{n-k_n} \sum_{i=1}^n \varepsilon_i \otimes \varepsilon_i. \end{aligned}$$

We start with P_I . Using Lemma 5.1, we have, for $i = 1, \dots, n$,

$$(S - \hat{S}_{k_n}) X_i = S(I - \hat{\Pi}_{k_n}) X_i - \frac{1}{n} \sum_{j=1}^n \langle \hat{\Gamma}_{k_n}^+ X_j, X_i \rangle_H \varepsilon_j, \quad (16)$$

and we can decompose $P_I = P_I^{(1)} + P_I^{(2)} + P_I^{(3)} + P_I^{(4)}$, where

$$\begin{aligned} P_I^{(1)} &= \frac{1}{n-k_n} \sum_{i=1}^n S(I - \hat{\Pi}_{k_n}) X_i \otimes S(I - \hat{\Pi}_{k_n}) X_i, \\ P_I^{(2)} &= \frac{1}{n-k_n} \sum_{i=1}^n [-\frac{1}{n} \sum_{j=1}^n \langle \hat{\Gamma}_{k_n}^+ X_j, X_i \rangle_H \varepsilon_j] \otimes S(I - \hat{\Pi}_{k_n}) X_i, \\ P_I^{(3)} &= \frac{1}{n-k_n} \sum_{i=1}^n S(I - \hat{\Pi}_{k_n}) X_i \otimes [-\frac{1}{n} \sum_{j=1}^n \langle \hat{\Gamma}_{k_n}^+ X_j, X_i \rangle_H \varepsilon_j], \\ P_I^{(4)} &= \frac{1}{n-k_n} \sum_{i=1}^n [-\frac{1}{n} \sum_{j=1}^n \langle \hat{\Gamma}_{k_n}^+ X_j, X_i \rangle_H \varepsilon_j] \otimes [-\frac{1}{n} \sum_{j=1}^n \langle \hat{\Gamma}_{k_n}^+ X_j, X_i \rangle_H \varepsilon_j]. \end{aligned}$$

First we have $P_I^{(1)} = \frac{n}{n-k_n} S \left[\sum_{i=k_n+1}^n \hat{\lambda}_i \hat{v}_i \otimes \hat{v}_i \right] S'$. From the independence between X and ε , we have $\mathbb{E}[P_I^{(2)}] = \mathbb{E}[P_I^{(3)}] = 0$. Finally, we get

$$P_I^{(4)} = \frac{1}{n-k_n} \sum_{i=1}^n \left[\frac{1}{n^2} \sum_{j,l=1}^n \langle \hat{\Gamma}_{k_n}^+ X_j, X_i \rangle_H \langle \hat{\Gamma}_{k_n}^+ X_l, X_i \rangle_H \varepsilon_j \otimes \varepsilon_l \right],$$

$$\begin{aligned} \text{hence } \mathbb{E}[P_I^{(4)}] &= \frac{1}{n-k_n} \sum_{i=1}^n \left[\frac{1}{n^2} \sum_{j=1}^n \mathbb{E}[\langle \hat{\Gamma}_{k_n}^+ X_j, X_i \rangle_H^2] \mathbb{E}(\varepsilon_j \otimes \varepsilon_j) \right] \\ &= \frac{1}{n^2(n-k_n)} \mathbb{E} \left[\sum_{i=1}^n \sum_{j=1}^n \langle \hat{\Gamma}_{k_n}^+ X_j, X_i \rangle_H^2 \right] \Gamma_\varepsilon, \end{aligned}$$

and Lemma 5.2 gives $\mathbb{E}[P_I^{(4)}] = \frac{k_n}{n-k_n} \Gamma_\varepsilon$. So, we have shown that

$$\mathbb{E}[P_I] = \frac{n}{n-k_n} S \mathbb{E} \left[\sum_{i=k_n+1}^n \hat{\lambda}_i \hat{v}_i \otimes \hat{v}_i \right] S' + \frac{k_n}{n-k_n} \Gamma_\varepsilon. \quad (17)$$

Now, we decompose P_{II} in the following way

$$P_{II} = \frac{1}{n-k_n} \sum_{i=1}^n [S(I - \hat{\Pi}_k)(X_i) \otimes \varepsilon_i + \frac{1}{n-k_n} \sum_{i=1}^n \left[-\frac{1}{n} \sum_{j=1}^n \langle \Gamma_{k_n}^+ X_j, X_i \rangle_H \varepsilon_j \otimes \varepsilon_i \right]].$$

By the independence between X and ε , the result of Lemma 5.2, and a similar computation for P_{III} , we obtain

$$\mathbb{E}[P_{II}] = \mathbb{E}[P_{III}] = -\frac{k_n}{n-k_n} \Gamma_\varepsilon. \quad (18)$$

Finally, coming back to the computation of $\mathbb{E}(\hat{\Gamma}_{\varepsilon,n})$, Theorem 3.1 is a direct consequence of (17) and (18). \square

5.2. Proof of Theorem 3.2

The proof is based on the two following lemmas.

Lemma 5.3. *Under the assumptions of Theorem 3.2, we have*

$$\lim_{n \rightarrow +\infty} \mathbb{E} \left| \hat{\lambda}_{k_n} \right| = 0. \quad (19)$$

Proof: We have $\left(\mathbb{E} \left| \hat{\lambda}_{k_n} \right| \right)^2 \leq 2\lambda_{k_n}^2 + 2\mathbb{E} \left| \hat{\lambda}_{k_n} - \lambda_{k_n} \right|^2$.

From Lemma 2.2 and Theorem 2.5 in [3] with assumption (A.2), we obtain

$$\left(\mathbb{E} \left| \hat{\lambda}_{k_n} \right| \right)^2 \leq 2\lambda_{k_n}^2 + 2 \left\| \hat{\Gamma}_n - \Gamma \right\|_\infty^2 \leq 2\lambda_{k_n}^2 + \frac{2}{n} \mathbb{E} \|X\|^4,$$

which concludes the proof of the lemma. \square

Lemma 5.4. *Under the assumptions of Theorem 3.2, we have*

$$\left\| S \mathbb{E} \left(\hat{\Pi}_{(k_n+1):n} \hat{\Gamma}_n \right) S' \right\|_{\mathcal{N}} \leq \|S\|_{\mathcal{N}} \|S'\|_{\infty} \mathbb{E} \left| \hat{\lambda}_{k_n+1} \right|. \quad (20)$$

Proof: Immediate properties of norms $\|\cdot\|_{\infty}$ and $\|\cdot\|_{\mathcal{N}}$ give

$$\left\| S \mathbb{E} \left(\hat{\Pi}_{(k_n+1):n} \hat{\Gamma}_n \right) S' \right\|_{\mathcal{N}} \leq \|S\|_{\mathcal{N}} \|S'\|_{\infty} \left\| \mathbb{E} \left(\hat{\Pi}_{(k_n+1):n} \hat{\Gamma}_n \right) \right\|_{\infty},$$

which yields (20) as the norm $\|\cdot\|_{\infty}$ corresponds to the largest eigenvalue of the operator. \square

Theorem 3.2 is proved by combining Corollary 3.1 with Lemmas 5.3 and 5.4, and taking assumption (A.1) into account.

5.3. Proof of Theorem 3.3

We begin with the following lemmas.

Lemma 5.5. *Under the assumptions of Theorem 3.2, we have*

$$\begin{aligned} \mathbb{E} \left(\check{\Gamma}_{\varepsilon,n} \right) &= \Gamma_{\varepsilon} + \left(\frac{m}{m - k_m} \right) \left[S \mathbb{E} \left(\hat{\Pi}_{(k_m+1):m}^{[1]} \hat{\Gamma}_m^{[1]} \right) S' \right] \\ &\quad - \left(\frac{m}{m - k_m} \right) \left[S \mathbb{E} \left(\hat{\Pi}_{2k_m}^{[2]} \right) \mathbb{E} \left(\hat{\Pi}_{(k_m+1):m}^{[1]} \hat{\Gamma}_m^{[1]} \right) \mathbb{E} \left(\hat{\Pi}_{2k_m}^{[3]} \right)' S' \right]. \end{aligned}$$

Proof: We first note that

$$\begin{aligned} \hat{S}_{2k_m} &= \hat{\Delta}_m \hat{\Gamma}_{2k_m}^+ = \left[S \left(\frac{1}{m} \sum_{i=1}^m X_i \otimes X_i \right) + \frac{1}{m} \sum_{i=1}^m \varepsilon_i \otimes X_i \right] \Gamma_{2k_m}^+ \\ &= S \hat{\Pi}_{2k_m} + \frac{1}{m} \sum_{i=1}^m \varepsilon_i \otimes \Gamma_{2k_m}^+ X_i. \end{aligned}$$

As X and ε are independent, we get that $\mathbb{E} \left(\hat{S}_{2k_m} \right) = S \mathbb{E} \left(\hat{\Pi}_{2k_m} \right)$, which, combined with Corollary 3.1 and the fact that the three sub-samples are independent, ends the proof. \square

Lemma 5.6. *Under the assumptions of Theorem 3.2, we have*

$$\left\| S \mathbb{E} \left(\hat{\Pi}_{2k_m}^{[2]} \right) \mathbb{E} \left(\hat{\Pi}_{(k_m+1):m}^{[1]} \hat{\Gamma}_m^{[1]} \right) \mathbb{E} \left(\hat{\Pi}_{2k_m}^{[3]} \right)' S' \right\|_{\mathcal{N}} \leq \|S\|_{\mathcal{N}} \|S'\|_{\infty} \mathbb{E} \left| \hat{\lambda}_{k_m+1} \right|.$$

Proof: The proof is based on the same ideas as that used for proving Lemma 5.4. We remind that the infinite norm of projection operators are equal to one. \square

The proof of Theorem 3.3 is now a simple combination of Lemmas 5.3, 5.4, 5.5 and 5.6 and using the triangle inequality.

5.4. Proof of Proposition 3.1

We consider the model $Y_i(t) = \sum_{j=1}^{k_n} \langle X_i, \hat{v}_j \rangle_H \alpha_j(t) + \eta_i(t)$, for $i = 1, \dots, n$ and for all t . Here $\eta_i(t) = \varepsilon_i(t) + \sum_{j=k_n+1}^{\infty} \langle X_i, \hat{v}_j \rangle_H \alpha_j(t)$. Writing this model in a matrix form, we have

$$\mathbf{Y}(t) = \mathbf{X}\boldsymbol{\alpha}(t) + \boldsymbol{\eta},$$

where \mathbf{Y} and $\boldsymbol{\eta}$ are the vectors with size n and respective general terms Y_i and η_i and $\boldsymbol{\alpha}$ is the vector with size k_n and general term α_j . We can easily see that the associated mean square estimator is

$$\hat{\boldsymbol{\alpha}}(t) = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{Y}(t) = \left(\langle \hat{\mathcal{S}}_{k_n}(t, \cdot), \hat{v}_1 \rangle_H, \dots, \langle \hat{\mathcal{S}}_{k_n}(t, \cdot), \hat{v}_{k_n} \rangle_H \right)', \quad (21)$$

where $\hat{\mathcal{S}}_{k_n}(t, s)$ is the estimator of \mathcal{S} . Now, denoting \mathbf{Y}^* the vector with size n such that $Y_r^* = Y_r$ for $r \neq i$, $Y_i^* = \hat{Y}_i^{[-i]}$, $\mathbf{X}^{[-i]}$ the matrix \mathbf{X} without the i^{th} row and $\hat{\boldsymbol{\alpha}}^{[-i]}(t)$ the estimator of $\boldsymbol{\alpha}(t)$ using the whole sample except the i^{th} observation, we have, for any vector $\mathbf{a} = (a_1, \dots, a_{k_n})'$ of functions of H and for any t

$$\|\mathbf{Y}^*(t) - \mathbf{X}\mathbf{a}(t)\|_n \geq \left\| \mathbf{Y}^{[-i]}(t) - \mathbf{X}^{[-i]}\hat{\boldsymbol{\alpha}}^{[-i]}(t) \right\|_{n-1} \geq \left\| \mathbf{Y}^*(t) - \mathbf{X}\hat{\boldsymbol{\alpha}}^{[-i]}(t) \right\|_n.$$

The fact that $(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{Y}^*(t)$ minimizes $\|\mathbf{Y}^*(t) - \mathbf{X}\mathbf{a}(t)\|_n$ leads to $\hat{\boldsymbol{\alpha}}^{[-i]}(t) = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{Y}^*(t)$, hence $\mathbf{X}\hat{\boldsymbol{\alpha}}^{[-i]}(t) = \mathbf{H}\mathbf{Y}^*(t)$. The end of the proof comes from

$$Y_i - \hat{Y}_i^{[-i]} = Y_i - (\mathbf{H}\mathbf{Y}^*)_i = Y_i - \sum_{\substack{r=1 \\ r \neq i}}^n \mathbf{H}_{ir} Y_r - \mathbf{H}_{ii} \hat{Y}_i^{[-i]} = Y_i - \hat{Y}_i + \mathbf{H}_{ii} (Y_i - \hat{Y}_i^{[-i]}).$$

Acknowledgements: The authors are grateful to both anonymous referees for valuable comments that helped us to improve the paper. They also would like to thank the Labex NUMEV (convention ANR-10-LABX-20) for partly funding the PhD thesis of Tito Manrique (under project 2013-1-007).

- [1] R. A. Johnson, D. W. Wichern, Applied multivariate statistical analysis, sixth Edition, Prentice Hall Englewood Cliffs, NJ, 2007.
- [2] J. O. Ramsay, B. W. Silverman, Functional data analysis, 2nd Edition, Springer Series in Statistics, Springer, New York, 2005.

- [3] L. Horváth, P. Kokoszka, Inference for functional data with applications., New York, NY: Springer, 2012. doi:10.1007/978-1-4614-3655-3.
- [4] C. Crambes, A. Mas, Asymptotics of prediction in functional linear regression with functional outputs., *Bernoulli* 19 (5B) (2013) 2627–2651. doi:10.3150/12-BEJ469.
- [5] H. Cardot, F. Ferraty, P. Sarda, Functional linear model, *Statistics & Probability Letters* 45 (1) (1999) 11 – 22.
- [6] T. Hsing, R. Eubank, Theoretical foundations of functional data analysis with an introduction to linear operators., Wiley, Chichester, 2015.
- [7] G. Wahba, Spline models for observational data., SIAM, Philadelphia, 1991.
- [8] R. B. Ash, M. F. Gardner, Topics in stochastic processes, Academic Press [Harcourt Brace Jovanovich Publishers], New York, 1975, probability and Mathematical Statistics, Vol. 27.