Journal of Multivariate Analysis 151 (2016) 54–68

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

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

Best estimation of functional linear models

Giacomo Aletti [a,b], Caterina May [c,*], Chiara Tommasi [b]

[a] ADAMSS Center, Milan, Italy
[b] Università degli Studi di Milano, Milan, Italy
[c] Università del Piemonte Orientale, Novara, Italy

ARTICLE INFO

Article history:
Received 16 July 2015
Available online 29 July 2016

AMS subject classifications:
primary 62H12
secondary 47B06
62M15
62M99

Keywords:
Functional data analysis
Sobolev spaces
Linear models
Repeated measurements
Gauss–Markov theorem
Riesz representation theorem
Best linear unbiased estimator

ABSTRACT

Observations that are realizations of some continuous process are frequently found in science, engineering, economics, and other fields. In this paper, we consider linear models with possible random effects and where the responses are random functions in a suitable Sobolev space. In particular, the processes cannot be observed directly. By using smoothing procedures on the original data, both the response curves and their derivatives can be reconstructed, both as an ensemble and separately. From these reconstructed functions, one representative sample is obtained to estimate the vector of functional parameters. A simulation study shows the benefits of this approach over the common method of using information either on curves or derivatives. The main theoretical result is a strong functional version of the Gauss–Markov theorem. This ensures that the proposed functional estimator is more efficient than the best linear unbiased estimator (BLUE) based only on curves or derivatives.

1. Introduction

Observations which are realizations of some continuous process are ubiquitous in many fields like science, engineering, economics and other fields. For this reason, the interest in statistical modelling of functional data is increasing, with applications in many areas. Reference monographs on functional data analysis are, for instance, the books of Ramsay and Silverman [11] and Horváth and Kokoszka [7], and the book of Ferraty and Vieu [5] for the non-parametric approach. They cover topics like data representation, smoothing and registration; regression models; classification, discrimination and principal component analysis; derivatives and principal differential analysis; and many other.

Regression models with functional variables can cover different situations; for example, functional responses, or functional predictors, or both. In this paper, linear models with functional response and multivariate (or univariate) regressor are examined. We consider the case of repeated measurements, where the theoretical results remain valid in the standard case. The focus of the work is to find the best estimation of the functional coefficients of the regressors.

The use of derivatives is very important in exploratory analysis of functional data; as well as for inference and prediction methodologies. High quality derivative information may be determined, for instance, by reconstructing the functions with spline smoothing procedures. Recent developments in the estimation of derivatives are contained in Sangalli et al. [12] and

* Corresponding author.
E-mail addresses: giacomo.aletti@unimi.it (G. Aletti), caterina.may@uniupo.it (C. May), chiara.tommasi@unimi.it (C. Tommasi).

http://dx.doi.org/10.1016/j.jmva.2016.07.005

in Pigoli and Sangalli [10]. See also Baraldo et al. [3], who have obtained derivatives in the context of survival analysis, and Hall et al. [6] who have estimated derivatives in a non-parametric model.

Curves and derivatives are reconstructed from a set of observed values. The reason for this is that the response processes cannot be observed directly. In the literature, the usual space for functional data is L^2 and the observed values are used to reconstruct either curve functions or derivatives.

The most common method to reconstruct derivatives is to build the sample of functions by using a smoothing procedure on the data, and then to differentiate these curve functions. However, the sample of functions and the sample of derivatives may be obtained separately. For instance, different smoothing techniques may be used to obtain the functions and the derivatives. Another possibility is when two sets of data are available, which are suitable to estimate functions and derivatives, respectively.

Some examples of curve and derivative data are: studying how the velocity of a car on a particular street is influenced by some covariates, the velocity measured by a police radar; GPS-tracked position estimation. In chemical experiments, data on reaction velocity and concentration may be collected separately. The novelty of the present work is that both information on curves and derivatives (that are not obtained by differentiation of the curves themselves) are used to estimate the functional coefficients.

The heuristic justification for this choice is that the data may provide different information about curve functions and their derivatives; it is therefore always recommended to use all available information. In fact, in this paper we prove that if we take into consideration both information about curves and their derivatives, we obtain the best linear unbiased estimates for the functional coefficients. Therefore, the common method of using information on either curve functions or their associated derivatives provides always a less efficient estimate (see Theorem 3 and Remark 2). For this reason, our theoretical results may have a relevant impact in practice.

Analogous to the Riesz Representation Theorem, we can find a representative function in H^1 which incorporates the information provided by a curve function and a derivative (which belong to L^2). Hence, from the two samples of reconstructed functions and their associated derivatives, only one representative sample is obtained and we use this representative sample to estimate the functional parameters. Once this method is given, the consequential theoretical results may appear as a straightforward extension of the well-known classical ones; their proof, however, requires much more technical effort and is not a straightforward extension.

The OLS estimator (based on both curves and derivatives through their Riesz representatives in H^1) is provided and some practical considerations are drawn. In general, the OLS estimator is not a BLUE because of the possible correlation between curves and derivatives. Therefore, a different representation of the data is provided (which takes into account this correlation). The resulting version of the Gauss–Markov theorem is proven in the proper infinite-dimensional space (H^1), showing that our sample of representative functions carries all the relevant information on the parameters. We propose an unbiased estimator which is linear with respect to the new sample of representatives and which minimizes a suitable covariance matrix (called global variance). This estimator is denoted H^1-functional SBLUE.

A simulation study numerically demonstrates the superiority of the H^1-functional SBLUE with respect to both the OLS estimators which are based only on curves or derivatives. This suggests that both sources of information should be used jointly, when available. A rough way of considering information on both curves and derivatives is to make a convex combination of the two OLS estimators. Simulation results show that the H^1-functional SBLUE is more efficient, as expected.

Let us finally remark that the results in this paper provide a strong theoretical foundation to generalize the theory of optimal design of experiments when functional observations occur (see Aletti et al. [1,2]).

The paper is organized as follows. Section 2 describes the model and proposes the OLS estimator obtained from the Riesz representation of the data. Section 3 explains some considerations which are fundamental from a practical point of view. Section 4 presents the construction of the H^1-functional SBLUE. Finally, Section 5 is devoted to the simulation study. Section 6 is a summary together with some final remarks. Some additional results and the proofs of theorems are deferred to Appendix A.1.

2. Model description and Riesz representation

Let us first consider a regression model where the response y is a random function that depends linearly on a known variable $\mathbf{x}$, which is a vector (or scalar) through a functional coefficient, that needs to be estimated. In particular, we assume that there are n units (subjects or clusters), and $r \geq 1$ observations per unit at a condition $\mathbf{x}_i$ ($i = 1, \dots, n$). Note that $\mathbf{x}_1, \dots \mathbf{x}_n$ are not necessarily different. In the context of repeated measurements, we consider the following random effect model:

$$y_{ij}(t) = \mathbf{f}(\mathbf{x}_i)^\top \boldsymbol{\beta}(t) + \alpha_i(t) + \varepsilon_{ij}(t) \quad i = 1, \dots, n;\ j = 1, \dots, r, \tag{1}$$

where: t belongs to a compact set $\tau \subseteq \mathbb{R}$; $y_{ij}(t)$ denotes the response curve of the jth observation at the ith experiment; $\mathbf{f}(\mathbf{x}_i)$ is a p-dimensional vector of known functions; $\boldsymbol{\beta}(t)$ is an unknown p-dimensional functional vector; $\alpha_i(t)$ is a zero-mean process which denotes the random effect due to the ith experiment and takes into account the correlation among the r repetitions; $\varepsilon_{ij}(t)$ is a zero-mean error process. Let us note that we are interested in precise estimation of the fixed effects $\boldsymbol{\beta}(t)$; herein the random effects are nuisance parameters.

An example of model (1) can be found in Shen and Faraway [13], where an ergonomic problem is considered (in this case there are n clusters of observations for the same individual); if $r = 1$ this model reduces to the functional response model described, for instance, in Horváth and Kokoszka [7].

In a real world setting, the functions $y_{ij}(t)$ are not directly observed. By using a smoothing procedure on the original data, the investigator can reconstruct both the functions and their first derivatives, obtaining $y_{ij}^{(f)}(t)$ and $y_{ij}^{(d)}(t)$, respectively. Hence we can assume that the model for the reconstructed functional data is

$$\begin{cases} y_{ij}^{(f)}(t) = \mathbf{f}(\mathbf{x}_i)^\top \boldsymbol{\beta}(t) + \alpha_i^{(f)}(t) + \varepsilon_{ij}^{(f)}(t) \\ y_{ij}^{(d)}(t) = \mathbf{f}(\mathbf{x}_i)^\top \boldsymbol{\beta}'(t) + \alpha_i^{(d)}(t) + \varepsilon_{ij}^{(d)}(t) \end{cases} \quad i = 1, \dots, n;\ j = 1, \dots, r, \tag{2}$$

where

1. the n couples $(\alpha_i^{(f)}(t), \alpha_i^{(d)}(t))$ are independent and identically distributed bivariate vectors of zero-mean processes such that $\mathrm{E}(\|\alpha_i^{(f)}(t)\|^2_{L^2(\tau)} + \|\alpha_i^{(d)}(t)\|^2_{L^2(\tau)}) < \infty$, that is, $(\alpha_i^{(f)}(t), \alpha_i^{(d)}(t)) \in L^2(\Omega; \mathcal{L}^2)$, where $\mathcal{L}^2 = L^2(\tau) \times L^2(\tau)$;
2. the $n \times r$ couples $(\varepsilon_{ij}^{(f)}(t), \varepsilon_{ij}^{(d)}(t))$ are independent and identically distributed bivariate vectors of zero mean processes, with $\mathrm{E}(\|\varepsilon_{ij}^{(f)}(t)\|^2_{L^2} + \|\varepsilon_{ij}^{(d)}(t)\|^2_{L^2}) < \infty$.

As a consequence of the above assumptions: the data $y_{ij}^{(f)}(t)$ and $y_{ij}^{(d)}(t)$ can be correlated; the couples $(y_{ij}^{(f)}(t), y_{ij}^{(d)}(t))$ and $(y_{kl}^{(f)}(t), y_{kl}^{(d)}(t))$ are independent whenever $i \neq k$. The possible correlation between $(y_{ij}^{(f)}(t), y_{ij}^{(d)}(t))$ and $(y_{il}^{(f)}(t), y_{il}^{(d)}(t))$ is due to the common random effect $(\alpha_i^{(f)}(t), \alpha_i^{(d)}(t))$.

Note that the investigator might reconstruct each function $y_{ij}^{(f)}(t)$ and its derivative $y_{ij}^{(d)}(t)$ separately. In this case, the right-hand term of the second equation in (2) is not the derivative of the right-hand term of the first equation. The particular case when $y_{ij}^{(d)}(t)$ is obtained by differentiation $y_{ij}^{(f)}(t)$ is the most simple situation in model (2).

Let $\mathbf{B}(t)$ be an estimator of $\boldsymbol{\beta}(t)$, formed by p random functions in the Sobolev space H^1. Recall that a function $g(t)$ is in H^1 if $g(t)$ and its derivative $g'(t)$ belong to L^2. Moreover, H^1 is a Hilbert space with inner product

$$\begin{aligned} \langle g_1(t), g_2(t)\rangle_{H^1} &= \langle g_1(t), g_2(t)\rangle_{L^2} + \langle g_1'(t), g_2'(t)\rangle_{L^2} \\ &= \langle (g_1(t), g_1'(t)), (g_2(t), g_2'(t))\rangle_{\mathcal{L}^2} \\ &= \int g_1(t)g_2(t)dt + \int g_1'(t)g_2'(t)dt, \quad g_1(t), g_2(t) \in H^1. \end{aligned} \tag{3}$$

Definition 1. We define the H^1-*global covariance matrix* $\Sigma_{\mathbf{B}}$ of an unbiased estimator $\mathbf{B}(t)$ as the $p \times p$ matrix whose (l_1, l_2)th element is

$$\mathrm{E}\langle B_{l_1}(t) - \beta_{l_1}(t), B_{l_2}(t) - \beta_{l_2}(t)\rangle_{H^1}. \tag{4}$$

This global notion of covariance has been used also in Menafoglio et al. [8, Definition 2], in the context of predicting georeferenced functional data. The authors have found a BLUE estimator for the drift of their underlying process, which can be seen as an example of the results provided in this paper.

Given a pair $(y^{(f)}(t), y^{(d)}(t)) \in L^2 \times L^2$, a linear continuous operator on H^1 may be defined as follows

$$\phi(h) = \langle y^{(f)}, h\rangle_{L^2} + \langle y^{(d)}, h'\rangle_{L^2} = \langle (y^{(f)}, y^{(d)}),\ (h, h')\rangle_{\mathcal{L}^2}, \quad \forall h \in H^1.$$

From the Riesz representation theorem, there exists a unique $\tilde{y} \in H^1$ such that

$$\langle \tilde{y}, h\rangle_{H^1} = \langle y^{(f)}, h\rangle_{L^2} + \langle y^{(d)}, h'\rangle_{L^2}, \quad \forall h \in H^1. \tag{5}$$

Definition 2. The unique element $\tilde{y} \in H^1$ defined in (5) is called the *Riesz representative* of the couple $(y^{(f)}(t), y^{(d)}(t)) \in \mathcal{L}^2$.

This definition will be useful to provide a nice expression for the functional OLS estimator $\widehat{\boldsymbol{\beta}}(t)$. Actually the Riesz representative synthesizes, in some sense, in H^1 the information of both $y^{(f)}(t)$ and $y^{(d)}(t)$.

Note that, since

$$\langle (y^{(f)}, y^{(d)}) - (\tilde{y}, \tilde{y}'),\ (h, h')\rangle_{\mathcal{L}^2} = 0, \quad \forall h \in H^1$$

the Riesz representative $(\tilde{y}, \tilde{y}')$ may be seen as the projection of $(y^{(f)}, y^{(d)}) \in \mathcal{L}^2$ onto the immersion of H^1 in $\mathcal{L}^2$, a linear closed subspace.

The *functional OLS estimator* for the model (2) is

$$\begin{aligned}\widehat{\boldsymbol{\beta}}(t) &= \arg\min_{\boldsymbol{\beta}(t)}\Big(\sum_{j=1}^{r}\sum_{i=1}^{n}\|y_{ij}^{(f)}(t)-\mathbf{f}(\mathbf{x}_i)^\top\boldsymbol{\beta}(t)\|_{L^2}^2+\sum_{j=1}^{r}\sum_{i=1}^{n}\|y_{ij}^{(d)}(t)-\mathbf{f}(\mathbf{x}_i)^\top\boldsymbol{\beta}'(t)\|_{L^2}^2\Big)\\ &= \arg\min_{\boldsymbol{\beta}(t)}\sum_{j=1}^{r}\sum_{i=1}^{n}\Big(\|y_{ij}^{(f)}(t)-\mathbf{f}(\mathbf{x}_i)^\top\boldsymbol{\beta}(t)\|_{L^2}^2+\|y_{ij}^{(d)}(t)-\mathbf{f}(\mathbf{x}_i)^\top\boldsymbol{\beta}'(t)\|_{L^2}^2\Big)\end{aligned}$$

The quantity

$$\|y_{ij}^{(f)}(t)-\mathbf{f}(\mathbf{x}_i)^\top\boldsymbol{\beta}(t)\|_{L^2}^2+\|y_{ij}^{(d)}(t)-\mathbf{f}(\mathbf{x}_i)^\top\boldsymbol{\beta}'(t)\|_{L^2}^2$$

resembles

$$\|y_{ij}(t)-\mathbf{f}(\mathbf{x}_i)^\top\boldsymbol{\beta}(t)\|_{H^1}^2,$$

because $y_{ij}^{(f)}(t)$ and $y_{ij}^{(d)}(t)$ reconstruct $y_{ij}(t)$ and its derivative function, respectively. The functional OLS estimator $\widehat{\boldsymbol{\beta}}(t)$ minimizes, in this sense, the sum of the H^1-norm of the unobservable residuals $y_{ij}(t)-\mathbf{f}(\mathbf{x}_i)^\top\boldsymbol{\beta}(t)$.

Theorem 1. *Given the model in (2),*

(a) *the functional OLS estimator $\widehat{\boldsymbol{\beta}}(t)$ can be computed by*

$$\widehat{\boldsymbol{\beta}}(t)=(F^\top F)^{-1}F^\top\bar{\mathbf{y}}(t), \tag{6}$$

where $\bar{\mathbf{y}}(t)=(\bar{y}_1(t),\ldots,\bar{y}_n(t))^\top$ is a vector, whose component ith is the mean of the Riesz representatives of the replications:

$$\bar{y}_i(t)=\frac{\sum_{j=1}^{r}\tilde{y}_{ij}(t)}{r},$$

and $F=[\mathbf{f}(\mathbf{x}_1),\ldots,\mathbf{f}(\mathbf{x}_n)]^\top$ is the $n\times p$ design matrix.

(b) *The estimator $\widehat{\boldsymbol{\beta}}(t)$ is unbiased and its global covariance matrix is $\sigma^2(F^\top F)^{-1}$, where $\sigma^2=\mathrm{E}(\|\bar{y}_i(t)-\mathbf{f}(\mathbf{x}_i)^\top\boldsymbol{\beta}(t)\|_{H^1}^2)$.*

Remark 1. Previous results may be generalized to other Sobolev spaces. The extension to H^m, $m\geq 2$, is straightforward. Moreover, in a Bayesian context, the investigator might have a different *a priori* consideration of $y_{ij}^{(f)}(t)$ and $y_{ij}^{(d)}(t)$. Thus, different weights may be used for curves and derivatives, and the inner product given in (3) may be extended to

$$\langle g_1(t),g_2(t)\rangle_H=\lambda\int_\tau g_1(t)g_2(t)dt+(1-\lambda)\int_\tau g_1'(t)g_2'(t)dt,\quad \lambda\in[0,1].$$

Let $\widehat{\boldsymbol{\beta}}_\lambda(t)$ be the OLS estimator obtained by using this last inner product. Note that, for $\lambda=1/2$, we obtain $\widehat{\boldsymbol{\beta}}_{1/2}(t)=\widehat{\boldsymbol{\beta}}(t)$ defined in Theorem 1. The behaviour of the $\widehat{\boldsymbol{\beta}}_\lambda(t)$ is explored in Section 5 for different choices of λ.

3. Practical considerations

In a real world context, we work with a finite dimensional subspace $\mathcal{S}$ of H^1. Let $S=\{w_1(t),\ldots,w_N(t)\}$ be a base of $\mathcal{S}$. Without loss of generality, we may assume that $\langle w_h(t),w_k(t)\rangle_{H^1}=\delta_h^k$, where

$$\delta_h^k=\begin{cases}1 & \text{if } h=k;\\ 0 & \text{if } h\neq k;\end{cases}$$

is the Kronecker delta symbol, since a Gram–Schmidt orthonormalization procedure may be always applied. More precisely, given any base $\tilde{S}=\{\tilde{w}_1(t),\ldots,\tilde{w}_N(t)\}$ in H_1, the corresponding orthonormal base is given by:

for $k=1$, define $w_1(t)=\tilde{w}_1(t)/\|\tilde{w}_1(t)\|_{H^1}$,

for $k\geq 2$, let $\hat{w}_k(t)=\tilde{w}_k(t)-\sum_{h=1}^{n-1}\langle\tilde{w}_k(t),w_h(t)\rangle_{H^1}w_h(t)$, and $w_k(t)=\hat{w}_k(t)/\|\hat{w}_k(t)\|_{H^1}$.

With this orthonormalized base, the projection $\tilde{y}(t)_{\mathcal{S}}$ on $\mathcal{S}$ of the Riesz representative $\tilde{y}(t)$ of the couple $(y^{(f)}(t),y^{(d)}(t))$ is given by

$$\begin{aligned}\tilde{y}(t)_{\mathcal{S}} &= \sum_{k=1}^{N}\langle\tilde{y}(t),w_k(t)\rangle_{H^1}\cdot w_k(t)\\ &= \sum_{k=1}^{N}\Big(\langle y^{(f)}(t),w_k(t)\rangle_{L^2}+\langle y^{(d)}(t),w_k'(t)\rangle_{L^2}\Big)w_k(t),\end{aligned} \tag{7}$$

where the last equality comes from the definition (5) of the Riesz representative. Now, if $\mathbf{m}_l = (m_{l,1}, \ldots, m_{l,n})^\top$ is the lth row of $(F^\top F)^{-1}F^\top$, then

$$\begin{aligned}\langle \hat{\boldsymbol\beta}_l(t), w_k(t)\rangle_{H^1} &= \sum_{i=1}^n \langle m_{l,i}\bar{y}_i(t), w_k(t)\rangle_{H^1} \\ &= \sum_{i=1}^n m_{l,i}\langle \bar{y}_i(t), w_k(t)\rangle_{H^1}, \quad \text{for any } k = 1, \ldots, N, \\ \hat{\boldsymbol\beta}_l(t)_{\mathcal{S}} &= \mathbf{m}_l^\top \bar{\mathbf{y}}(t)_{\mathcal{S}},\end{aligned}$$

hence $\widehat{\boldsymbol\beta}(t)_{\mathcal{S}} = (F^\top F)^{-1}F^\top \bar{\mathbf{y}}(t)_{\mathcal{S}}$.

Let us note that, even if the Riesz representative (5) is implicitly defined, its projection on $\mathcal{S}$ can be easily computed by (7). From a practical point of view, the statistician can work with the data $(y_{ij}^{(f)}(t), y_{ij}^{(d)}(t))$ projected on a finite linear subspace $\mathcal{S}$ and the corresponding OLS estimator $\widehat{\boldsymbol\beta}(t)_{\mathcal{S}}$ is the projection on $\mathcal{S}$ of the OLS estimator $\widehat{\boldsymbol\beta}(t)$ given in Section 2.

It is straightforward to prove that the estimator (6) becomes

$$\widehat{\boldsymbol\beta}(t) = (F^\top F)^{-1}F^\top \mathbf{y}^{(f)}(t),$$

in two cases: when we do not take into consideration $y^{(d)}$, or when $y^{(d)} = (y^{(f)})'$. Up to our knowledge, this is the most common situation considered in the literature (see Ramsay and Silverman [11, Chapt. 13]). However, from the simulation study of Section 5, the OLS estimator $\widehat{\boldsymbol\beta}$ is less efficient when it is based only on $y^{(f)}$.

4. Strong H^1-BLUE in functional linear models

Let $\mathbf{B}(t) = \mathbf{C}(\mathbf{y}^{(f)}(t), \mathbf{y}^{(d)}(t))$, where $\mathbf{C} : \mathcal{R} \subseteq (\mathcal{L}^2)^{nr} \to (H^1)^p$ is a linear closed operator; in this case $\mathbf{B}(t)$ is called a *linear estimator*. The domain of C, denoted by $\mathcal{R}$, will be defined in (18). Theorem 2 will ensure that the dataset $(\mathbf{y}^{(f)}(t), \mathbf{y}^{(d)}(t))$ is contained in $\mathcal{R}$.

Definition 3. Analogous to classical settings, we define the H^1-functional best linear unbiased estimator (H^1-BLUE) as the estimator with minimal (in the sense of Loewner Partial Order[1]) H^1-global covariance matrix (4), in the class of the linear unbiased estimators $\mathbf{B}(t)$ of $\boldsymbol\beta(t)$.

From the definition of Loewner Partial Order, a H^1-BLUE minimizes the quantity

$$\mathrm{E}\Big(\Big\langle \sum_{i=1}^p \alpha_i\big(B_i(t) - \beta_i(t)\big), \sum_{i=1}^p \alpha_i\big(B_i(t) - \beta_i(t)\big)\Big\rangle_{H^1}\Big)$$

for any choice of $(\alpha_1, \ldots, \alpha_p)$, in the class of the linear unbiased estimators $\mathbf{B}(t)$ of $\boldsymbol\beta(t)$. In other words, the H^1-BLUE minimizes the H^1-global variance of any linear combination of its components. A stronger request is the following.

Definition 4. We define the H^1-strong functional best linear unbiased estimator (H^1-SBLUE) as the estimator with minimal global variance,

$$\mathrm{E}\Big(\Big\langle \mathrm{O}(\mathbf{B}(t) - \boldsymbol\beta(t)), \mathrm{O}(\mathbf{B}(t) - \boldsymbol\beta(t))\Big\rangle_{H^1}\Big)$$

for any choice of a (sufficiently regular) continuous linear operator $\mathrm{O} : (H^1)^p \to H^1$, in the class of the linear unbiased estimators $\mathbf{B}(t)$ of $\boldsymbol\beta(t)$.

4.1. $H^1_{\mathbf{R}}$-representation on the Hilbert space $\mathcal{L}^2_{\mathbf{R}}$

Recall that, for any given (i, j), the couple $(\alpha_i^{(f)}(t)+\varepsilon_{ij}^{(f)}(t), \alpha_i^{(d)}(t)+\varepsilon_{ij}^{(d)}(t))$ is a process with values in $\mathcal{L}^2 = L^2(\tau)\times L^2(\tau)$. Let $\mathbf{R}(s,t) = \sum_k \lambda_k \boldsymbol\Psi_k(s)\boldsymbol\Psi_k(t)^\top$ be the spectral representation of the covariance matrix of the process

$$\mathbf{e}_i^\top(t) = (e_i^{(f)}(t), e_i^{(d)}(t)) = \frac{1}{r}\sum_{j=1}^r (\alpha_i^{(f)}(t) + \varepsilon_{ij}^{(f)}(t), \alpha_i^{(d)}(t) + \varepsilon_{ij}^{(d)}(t)), \quad i = 1, \ldots, n \tag{8}$$

which means $\lambda_k > 0$, $\sum_k \lambda_k < \infty$ and the sequence $\{\boldsymbol\Psi_k(t), k = 1, 2, \ldots\}$ are orthonormal bivariate vectors in $\mathcal{L}^2$. Without loss of generality assume that the $\mathcal{L}^2$-closure of the linear span of $\{\boldsymbol\Psi_k(t), k = 1, 2, \ldots\}$ includes H^1 (see Remark 3):

[1] Given two symmetric matrices A and B, $A \geq B$ in Loewner Partial Order if $A - B$ is positive definite.

$\overline{\mathcal{L}^2 \cap \text{span}\{\boldsymbol{\Psi}_k(t), k = 1, 2, \ldots\}} \supseteq H^1$. Note that $\mathbf{R}(s, t)$, the covariance matrix of the process $\mathbf{e}_i(t)$, does not depend on i. From Karhunen–Loève Theorem (see, e.g., Perrin et al. [9]), there exists an array of zero-mean unit variance random variables $\{e_{i,k}; i = 1, \ldots, n; k = 1, 2, \ldots\}$ such that

$$\mathbf{e}_i(t) = \sum_k \sqrt{\lambda_k} e_{i,k} \boldsymbol{\Psi}_k(t). \tag{9}$$

The linearity of the covariance operator with respect to the first process, together with the symmetry in j given in the hypothesis (1) and (2), ensures that

$$\mathrm{E}\Big(\big(\alpha_i^{(f)}(s) + \varepsilon_{ij}^{(f)}(s),\ \alpha_i^{(d)}(s) + \varepsilon_{ij}^{(d)}(s)\big)^\top \cdot \mathbf{e}_i^\top(t)\Big) = \mathbf{R}(s, t) = \sum_k \lambda_k \boldsymbol{\Psi}_k(s) \boldsymbol{\Psi}_k(t)^\top. \tag{10}$$

Now, for $i = 1, \ldots, n; j = 1, \ldots, r; k = 1, 2, \ldots$, let

$$X_{ij,k} = \Big\langle \boldsymbol{\Psi}_k, \big(\alpha_i^{(f)} + \varepsilon_{ij}^{(f)},\ \alpha_i^{(d)} + \varepsilon_{ij}^{(d)}\big)^\top \Big\rangle_{\mathcal{L}^2},$$

and hence

$$\big(\alpha_i^{(f)}(s) + \varepsilon_{ij}^{(f)}(s),\ \alpha_i^{(d)}(s) + \varepsilon_{ij}^{(d)}(s)\big)^\top = \sum_k X_{ij,k} \boldsymbol{\Psi}_k(s), \qquad \frac{1}{r}\sum_{j=1}^r X_{ij,k} = \sqrt{\lambda_k} e_{i,k}.$$

The independence assumptions in the hypothesis (1) and (2) ensures that the joint law of the processes $(\alpha_{i_1}^{(f)} + \varepsilon_{i_1 j}^{(f)},\ \alpha_{i_1}^{(d)} + \varepsilon_{i_1 j}^{(d)})$ and $\mathbf{e}_{i_2}$ does not depend on j, hence

$$\mathrm{E}(X_{i_1 1,k_1} \sqrt{\lambda_{k_2}} e_{i_2,k_2}) = \mathrm{E}(X_{i_1 2,k_1} \sqrt{\lambda_{k_2}} e_{i_2,k_2}) = \cdots = \mathrm{E}(X_{i_1 r,k_1} \sqrt{\lambda_{k_2}} e_{i_2,k_2}).$$

From (10), the linearity of the expectation ensures that

$$\delta_{i_1}^{i_2} \delta_{k_1}^{k_2} \lambda_{k_1} = \mathrm{E}(\sqrt{\lambda_{k_1}} e_{i_1,k_1} \sqrt{\lambda_{k_2}} e_{i_2,k_2}) = \sqrt{\lambda_{k_2}} \mathrm{E}(X_{i_1 j,k_1} e_{i_2,k_2}), \quad j = 1, \ldots, r. \tag{11}$$

Let us observe that the elements of $\overline{\mathcal{L}^2 \cap \text{span}\{\boldsymbol{\Psi}_k(t), k = 1, 2, \ldots\}}$ are the functions $\mathbf{a}$ such that $\mathbf{a} = \sum_k \langle \mathbf{a}, \boldsymbol{\Psi}_k \rangle_{\mathcal{L}^2} \cdot \boldsymbol{\Psi}_k$ and $\|\mathbf{a}\|^2_{\mathcal{L}^2} = \sum_k \langle \mathbf{a}, \boldsymbol{\Psi}_k \rangle^2_{\mathcal{L}^2} < \infty$. In the following definition a stronger condition is required.

Definition 5. Given the spectral representation of $\mathbf{R}(s, t)$, let

$$\mathcal{L}^2_{\mathbf{R}} = \Big\{ \mathbf{a} \in \overline{\mathcal{L}^2 \cap \text{span}\{\boldsymbol{\Psi}_k(t), k = 1, 2, \ldots\}} \colon \sum_k \frac{\langle \mathbf{a}, \boldsymbol{\Psi}_k \rangle^2_{\mathcal{L}^2}}{\lambda_k} < \infty \Big\} \tag{12}$$

be a new Hilbert space, with inner product

$$\langle \mathbf{a}, \mathbf{b} \rangle_{\mathcal{L}^2_{\mathbf{R}}} = \sum_k \frac{\langle \mathbf{a}, \boldsymbol{\Psi}_k \rangle_{\mathcal{L}^2} \langle \mathbf{b}, \boldsymbol{\Psi}_k \rangle_{\mathcal{L}^2}}{\lambda_k}. \tag{13}$$

Note that $\|\cdot\|_{\mathcal{L}^2} \leq \|\cdot\|_{\mathcal{L}^2_{\mathbf{R}}} / \max(\lambda_k)$. An orthonormal base for $\mathcal{L}^2_{\mathbf{R}}$ is given by $(\boldsymbol{\Phi}_k)_k$, where $\boldsymbol{\Phi}_k = \sqrt{\lambda_k} \boldsymbol{\Psi}_k$ for any k.

Consider now the following linear closed dense subset of $\mathcal{L}^2_{\mathbf{R}}$:

$$K = \Big\{ \mathbf{b} \in \mathcal{L}^2_{\mathbf{R}} \colon \sum_k \frac{\langle \boldsymbol{\Psi}_k, \mathbf{b} \rangle^2_{\mathcal{L}^2}}{\lambda_k^2} < \infty \Big\}.$$

Observe that $\boldsymbol{\Psi}_k \in K$ for all k. If K^* is the $\mathcal{L}^2_{\mathbf{R}}$-dual space of K, the Gelfand triple $K \subset \mathcal{L}^2_{\mathbf{R}} \subset K^*$ implies that $\overline{\mathcal{L}^2 \cap \text{span}\{\boldsymbol{\Psi}_k(t), k = 1, 2, \ldots\}} \subseteq K^*$.

Analogous to the geometric interpretation of the Riesz representation, we construct the $H^1_{\mathbf{R}}$-*representation* in the following way. For any element $\mathbf{b} \in \mathcal{L}^2_{\mathbf{R}}$, we call $H^1_{\mathbf{R}}$-*representative* its $\mathcal{L}^2_{\mathbf{R}}$-projection on H^1, and we denote it with the symbol $b^{(\mathbf{R})}$. In particular, for any k, let $\psi_k^{(\mathbf{R})}(t)$ be the $H^1_{\mathbf{R}}$-representative of $\boldsymbol{\Psi}_k$, that is, the unique element in $H^1 \cap \mathcal{L}^2_{\mathbf{R}}$ such that

$$\langle (\psi_k^{(\mathbf{R})}, \psi_k^{(\mathbf{R})\prime})^\top, (g, g')^\top \rangle_{\mathcal{L}^2_{\mathbf{R}}} = \langle \boldsymbol{\Psi}_k, (g, g')^\top \rangle_{\mathcal{L}^2_{\mathbf{R}}} = \frac{\langle \boldsymbol{\Psi}_k, (g, g')^\top \rangle_{\mathcal{L}^2}}{\lambda_k}, \quad \forall g \in H^1 \cap \mathcal{L}^2_{\mathbf{R}}.$$

Note that the $H^1_{\mathbf{R}}$-representatives of the orthonormal system $(\mathbf{\Phi}_k)_k$ of $\mathcal{L}^2_{\mathbf{R}}$ are given by $\phi_k^{(\mathbf{R})}(t) = \sqrt{\lambda_k}\psi_k^{(\mathbf{R})}(t)$, where, by definition of projection,

$$\|\phi_k^{(\mathbf{R})}(t)\|_{H^1_{\mathbf{R}}} = \|(\phi_k^{(\mathbf{R})}(t), \phi_k^{(\mathbf{R})'}(t))^\top\|_{\mathcal{L}^2_{\mathbf{R}}} \leq \|\mathbf{\Phi}_k(t)\|_{\mathcal{L}^2_{\mathbf{R}}} = 1. \tag{14}$$

Moreover,

$$\langle(\phi_h^{(\mathbf{R})}, \phi_h^{(\mathbf{R})'})^\top, \mathbf{\Phi}_k\rangle_{\mathcal{L}^2_{\mathbf{R}}} = \langle(\phi_h^{(\mathbf{R})}, \phi_h^{(\mathbf{R})'})^\top, (\phi_k^{(\mathbf{R})}, \phi_k^{(\mathbf{R})'})^\top\rangle_{\mathcal{L}^2_{\mathbf{R}}} = \langle\mathbf{\Phi}_h, (\phi_k^{(\mathbf{R})}, \phi_k^{(\mathbf{R})'})^\top\rangle_{\mathcal{L}^2_{\mathbf{R}}}, \tag{15}$$

and the $H^1_{\mathbf{R}}$-representation of any $\mathbf{b} \in \mathcal{L}^2_{\mathbf{R}}$ can be written as

$$b^{(\mathbf{R})} = \sum_h \langle\mathbf{b}, \mathbf{\Psi}_h\rangle_{\mathcal{L}^2}\psi_h^{(\mathbf{R})} = \sum_h \langle\mathbf{b}, \mathbf{\Phi}_h\rangle_{\mathcal{L}^2_{\mathbf{R}}}\phi_h^{(\mathbf{R})}. \tag{16}$$

When $\mathbf{a} \in \overline{\mathcal{L}^2 \cap \operatorname{span}\{\mathbf{\Psi}_k(t), k = 1, 2, \ldots\}}$, it is again possible to define formally its $H^1_{\mathbf{R}}$-representation in the following way:

$$a^{(\mathbf{R})}(t) = \sum_k \langle\mathbf{a}, \mathbf{\Psi}_k\rangle_{\mathcal{L}^2}\psi_k^{(\mathbf{R})}(t). \tag{17}$$

In this case, if $a^{(\mathbf{R})} \in H^1$, an analogous of the standard projection can be obtained: $(a^{(\mathbf{R})}, a^{(\mathbf{R})'})$ it is the unique element in K^* of the form (a, a') with $a \in H^1$ such that

$$\langle\mathbf{a}, (h, h')^\top\rangle_{\mathcal{L}^2_{\mathbf{R}}} = \langle(a, a')^\top, (h, h')^\top\rangle_{\mathcal{L}^2_{\mathbf{R}}}, \quad \forall(h, h') \in K.$$

It will be useful to observe that, as a consequence, when $\mathbf{a} = (\mathbf{f}(\mathbf{x}_i)^\top\boldsymbol{\beta}, \mathbf{f}(\mathbf{x}_i)^\top\boldsymbol{\beta}')$, then its $H^1_{\mathbf{R}}$-representative is $\mathbf{f}(\mathbf{x}_i)^\top\boldsymbol{\beta}$.

Lemma 1. *Given $\mathbf{e}_i$ as in (8), its $H^1_{\mathbf{R}}$-representative*

$$e_i^{(\mathbf{R})} = \sum_k \sqrt{\lambda_k}e_{i,k}\psi_k^{(\mathbf{R})},$$

belongs to $L^2(\Omega; H^1)$, for any $i = 1, \ldots, n$.

The following theorem is a direct consequence of the previous results.

Theorem 2. *The following equation holds in $L^2(\Omega; H^1)$:*

$$\bar{y}_i^{(\mathbf{R})}(t)(\omega) = \mathbf{f}(\mathbf{x}_i)^\top\boldsymbol{\beta}(t) + e_i^{(\mathbf{R})}(t)(\omega) \quad i = 1, \ldots, n,$$

where each $\bar{y}_i^{(\mathbf{R})}$ is the $H^1_{\mathbf{R}}$-representation of the mean $(\bar{\mathbf{y}}_i^{(f)}(t), \bar{\mathbf{y}}_i^{(d)}(t))$ of the observations given in (A.2). As a consequence, $\bar{y}_i^{(\mathbf{R})}(t)$ belongs to $L^2(\Omega; H^1)$, and hence $\bar{y}_i^{(\mathbf{R})}(\omega) \in H^1$ a.s.

We define

$$\mathcal{R} = \{\mathbf{y} \in \left(\overline{\mathcal{L}^2 \cap \operatorname{span}\{\mathbf{\Psi}_k(t), k = 1, 2, \ldots\}}\right)^{nr} : y_i^{(\mathbf{R})} \in H^1, i = 1, \ldots, n\}. \tag{18}$$

The vector $\bar{\mathbf{y}}^{(\mathbf{R})}(t) = \left(\bar{y}_1^{(\mathbf{R})}, \bar{y}_2^{(\mathbf{R})}, \ldots, \bar{y}_n^{(\mathbf{R})}\right)^\top$ plays the rôle of the Riesz representative of Theorem 1 in the following SBLUE theorem.

Theorem 3. *The functional estimator*

$$\widehat{\boldsymbol{\beta}}^{(\mathbf{R})}(t) = (F^\top F)^{-1}F^\top\bar{\mathbf{y}}^{(\mathbf{R})}(t), \tag{19}$$

for the model (2) is a H^1-functional SBLUE.

Remark 2. From the proof of Theorem 3 (see Appendix A.1) we have that $\widehat{\boldsymbol{\beta}}^{(\mathbf{R})}(t)$ is the best estimator among all the estimators $\mathbf{B}(t) = \mathbf{C}(\mathbf{y}^{(f)}(t), \mathbf{y}^{(d)}(t))$ where $\mathbf{C} : \mathcal{R} \to (H^1)^p$ is any linear closed unbiased operator. Therefore, $\widehat{\boldsymbol{\beta}}^{(\mathbf{R})}(t)$ is also better than the best linear unbiased estimators based only on $\mathbf{y}^{(f)}(t)$ or $\mathbf{y}^{(d)}(t)$, since they are defined by some linear unbiased operator.

Remark 3. The assumption $\overline{\mathcal{L}^2 \cap \operatorname{span}\{\mathbf{\Psi}_k(t), k = 1, 2, \ldots\}} \supseteq H^1$ ensures that the each component of the unknown $\boldsymbol{\beta}(t)$ is in $\operatorname{span}\{\mathbf{\Psi}_k(t), k = 1, 2, \ldots\}$. As a consequence, we have noted that the $H^1_{\mathbf{R}}$-representative of $(\mathbf{f}(\mathbf{x}_i)^\top\boldsymbol{\beta}, \mathbf{f}(\mathbf{x}_i)^\top\boldsymbol{\beta}')$, is $\mathbf{f}(\mathbf{x}_i)^\top\boldsymbol{\beta}$. If this assumption is not true, it may happen that $\beta_l \notin \operatorname{span}\{\mathbf{\Psi}_k(t), k = 1, 2, \ldots\}$ for some $l = 1, \ldots, p$, and then β_l would have a nonzero projection on the orthogonal complement of $\operatorname{span}\{\mathbf{\Psi}_k(t), k = 1, 2, \ldots\}$. Since on the orthogonal complement we do not observe any noise, this means that we would have a deterministic subproblem, that, without loss of generality, we can ignore.

5. Simulations

In this section, it is explored, throughout a simulation study, when it is more convenient to use the whole information on both reconstructed functions and derivatives with respect to the partial use of $y^{(f)}(t)$ (or $y^{(d)}(t)$). The idea is that using the whole information on curves and derivatives is much more convenient as the dependence between $y^{(f)}(t)$ and $y^{(d)}(t)$ is smaller and their spread is more comparable.

In this study, for each scenario listed below, 1000 datasets are simulated from model (2) by a Monte Carlo method, with $n = 18, r = 3, p = 3$,

$$\boldsymbol{\beta}(t) = \begin{pmatrix} \sin(\pi t) + \sin(2\pi t) + \sin(4\pi t) \\ -\sin(\pi t) + \cos(\pi t) - \sin(2\pi t) + \cos(2\pi t) - \sin(4\pi t) + \cos(4\pi t) \\ +\sin(\pi t) + \cos(\pi t) + \sin(2\pi t) + \cos(2\pi t) + \sin(4\pi t) + \cos(4\pi t) \end{pmatrix}, \quad t \in (-1, 1),$$

and

$$F^{\top} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ -1.00 & -0.75 & -0.50 & -0.25 & 0.00 & 0.25 & 0.50 & 0.75 & 1.00 & -1.00 & -0.75 & -0.50 & -0.25 & 0.00 & 0.25 & 0.50 & 0.75 & 1.00 \end{bmatrix}.$$

In what follows, we compare the following different estimators: the SBLUE $\widehat{\boldsymbol{\beta}}^{(\mathbf{R})}(t)$ (see Section 4), the OLS estimators $\widehat{\boldsymbol{\beta}}_{\lambda}(t)$ (see Remark 1), and $\hat{\boldsymbol{\beta}}^{(c)}_{\lambda}(t) = \lambda\hat{\boldsymbol{\beta}}^{(f)}(t) + (1-\lambda)\hat{\boldsymbol{\beta}}^{(d)}(t)$, where $\hat{\boldsymbol{\beta}}^{(f)}(t)$ is the OLS estimator based on $\mathbf{y}^{(f)}(t)$ and $\hat{\boldsymbol{\beta}}^{(d)}(t)$ is the OLS estimator based on $\mathbf{y}^{(d)}(t)$, with $0 \le \lambda \le 1$.

Let us note that $\hat{\boldsymbol{\beta}}^{(c)}_{\lambda}(t)$ is a compound OLS estimator; it is a rough way of taking into account both the sources of information on $\mathbf{y}^{(f)}(t)$ and $\mathbf{y}^{(d)}(t)$. Of course, setting $\lambda = 0$ we ignore completely the information on the functions and $\hat{\boldsymbol{\beta}}^{(c)}_{0}(t) = \hat{\boldsymbol{\beta}}^{(d)}(t) = \widehat{\boldsymbol{\beta}}_0(t)$, vice versa setting $\lambda = 1$ means to ignore the information on the derivatives and thus $\hat{\boldsymbol{\beta}}^{(c)}_{1}(t) = \hat{\boldsymbol{\beta}}^{(f)}(t) = \widehat{\boldsymbol{\beta}}_1(t)$.

All the computations are developed using R package.

In Fig. 1 it is plotted: one dataset of curves and derivatives (black lines); the regression functions $\mathbf{f}(\mathbf{x}_i)^{\top}\boldsymbol{\beta}(t)$ and $\mathbf{f}(\mathbf{x}_i)^{\top}\boldsymbol{\beta}'(t)$ (green lines); the SBLUE predictions $\mathbf{f}(\mathbf{x}_i)^{\top}\widehat{\boldsymbol{\beta}}^{(\mathbf{R})}(t)$ and $\mathbf{f}(\mathbf{x}_i)^{\top}\widehat{\boldsymbol{\beta}}^{(\mathbf{R})\prime}(t)$ (blue lines); the OLS predictions $\mathbf{f}(\mathbf{x}_i)^{\top}\widehat{\boldsymbol{\beta}}_{1/2}(t)$ and $\mathbf{f}(\mathbf{x}_i)^{\top}\widehat{\boldsymbol{\beta}}_{1/2}'(t)$ (red lines).

5.1. Dependence between functions and derivatives

We consider three different scenarios; we generate functional data $y^{(f)}_{ij}(t)$ and $y^{(d)}_{ij}(t)$ such that

1. $(\alpha^{(f)}_i(t), \varepsilon^{(f)}_{ij}(t))$ is independent on $(\alpha^{(d)}_i(t), \varepsilon^{(d)}_{ij}(t))$;
2. $(\alpha^{(f)}_i(t), \varepsilon^{(f)}_{ij}(t))$ and $(\alpha^{(d)}_i(t), \varepsilon^{(d)}_{ij}(t))$ are mildly dependent (the degree of dependence is randomly obtained);
3. $(\alpha^{(f)}_i(t), \varepsilon^{(f)}_{ij}(t))$ and $(\alpha^{(d)}_i(t), \varepsilon^{(d)}_{ij}(t))$ are fully dependent: $(\alpha^{(d)}_i(t), \varepsilon^{(d)}_{ij}(t)) = (\alpha^{(f)\prime}_i(t), \varepsilon^{(f)\prime}_{ij}(t))$, and hence $y^{(d)}_{ij}(t) = y^{(f)\prime}_{ij}(t)$.

The performance of the different estimators is evaluated by comparing the H^1-norm of the p-components of the estimation errors. Fig. 2 depicts the Monte Carlo distribution of the H^1-norm of the first component: $\|\widehat{\boldsymbol{\beta}}_{\lambda,1}(t) - \boldsymbol{\beta}_1(t)\|_{H^1}$ for different values of λ (red box-plot, (6)), $\|\hat{\boldsymbol{\beta}}^{(c)}_{\lambda,1}(t) - \boldsymbol{\beta}_1(t)\|_{H^1}$ for different values of λ (yellow box-plots) and $\|\widehat{\boldsymbol{\beta}}^{(\mathbf{R})}_1(t) - \boldsymbol{\beta}_1(t)\|_{H^1}$ (blue box-plot).

From the comparison of the box-plots corresponding to $\lambda = 0$ and $\lambda = 1$ with the other cases, we observe that it is always more convenient to use the whole information on $y^{(f)}(t)$ and $y^{(d)}(t)$ (this behaviour is more evident in scenario 1). Among the three estimators $\hat{\boldsymbol{\beta}}^{(c)}_{\lambda}(t)$, $\hat{\boldsymbol{\beta}}_{\lambda}(t)$ and $\widehat{\boldsymbol{\beta}}^{(\mathbf{R})}(t)$, the SBLUE is the most precise, as expected. When there is a one-to-one dependence between $y^{(f)}(t)$ and $y^{(d)}(t)$, one source of information is redundant and all the functional estimators coincide (bottom panel of Fig. 2).

5.2. Spread of functions and derivatives

Also in this case, we consider three different scenarios. Let

$$r_{ll} = \frac{\left(\Sigma_{\hat{\boldsymbol{\beta}}^{(f)}}\right)_{ll}}{\left(\Sigma_{\hat{\boldsymbol{\beta}}^{(d)}}\right)_{ll}}, \quad l = 1, \dots, p,$$

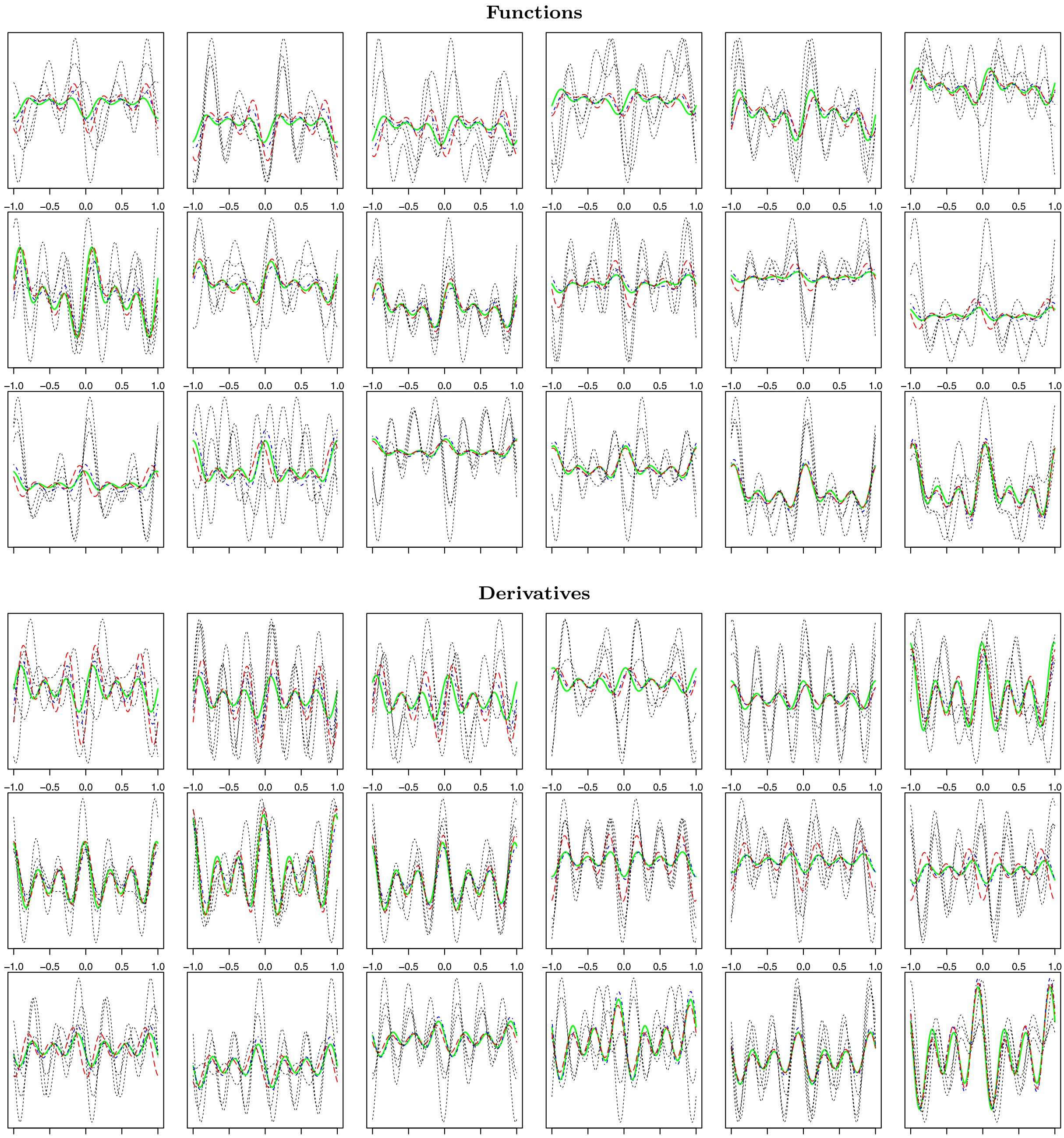

Fig. 1. Simulated data from model (2) and predicted curves. Black lines: simulated data of curves (top panel) and derivatives (bottom panel). In each ith box ($i = 1, \ldots, 18$) the $j = 1, \ldots, 3$ replications are plotted. Blue lines: predictions based on SBLUE estimator. Red lines: predictions based on OLS estimator. Green lines: theoretical curves $\mathbf{f}(\mathbf{x}_i)^\top \boldsymbol{\beta}(t)$ in top panel and $\mathbf{f}(\mathbf{x}_i)^\top \boldsymbol{\beta}'(t)$ in bottom panel. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

where $\Sigma_.$ denotes the H^1-global covariance matrix defined in (4). We generate functional data $y_{ij}^{(f)}(t)$ and $y_{ij}^{(d)}(t)$ with a different spread, such that

1. $r_{ll} \cong 0.25$ (in this sense, $y_{ij}^{(f)}(t)$ is "more concentrate" than $y_{ij}^{(d)}(t)$);
2. $r_{ll} \cong 1$ ($y_{ij}^{(f)}(t)$ and $y_{ij}^{(d)}(t)$ have more or less the same spread);
3. $r_{ll} \cong 4$ ($y_{ij}^{(d)}(t)$ is "more concentrate" than $y_{ij}^{(f)}(t)$).

As before, the performance of the different estimators is evaluated by comparing the H^1-norm of the p-components of the estimation errors. Fig. 3 depicts the Monte Carlo distribution of the H^1-norm of the first component: $\|\hat{\boldsymbol{\beta}}_{\lambda,1}(t) - \boldsymbol{\beta}_1(t)\|_{H^1}$ for

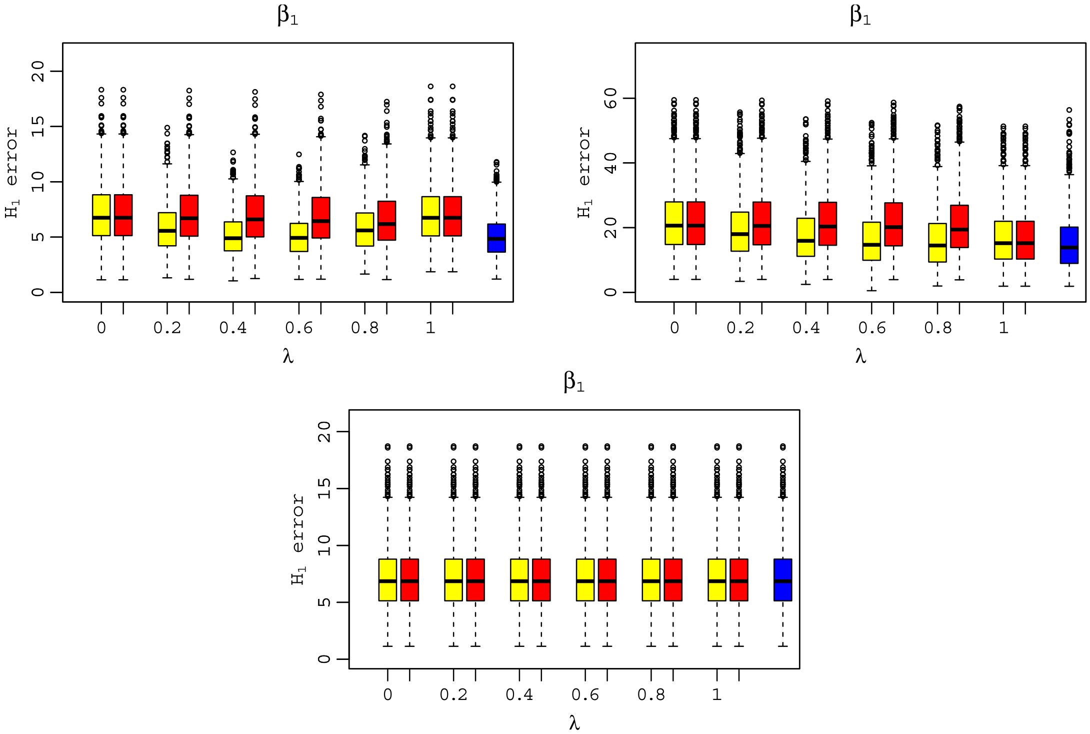

Fig. 2. H^1 norm of the first components estimation errors, for compound OLS estimators (yellow box-plots), OLS estimators (red box-plots), SBLUE estimators (blue box-plots). Top-left panel: scenario 1, independence. Top-right panel: scenario 2, mild dependence. Bottom panel: scenario 3, full dependence. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

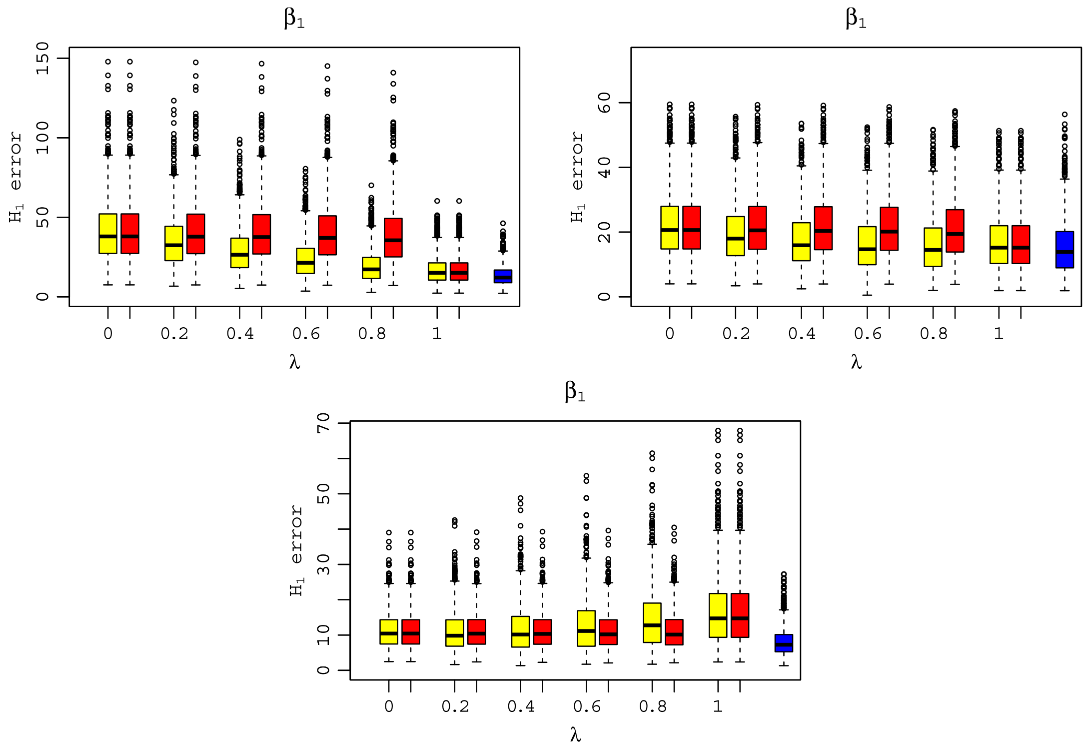

Fig. 3. H^1 norm of the first components estimation errors, for compound OLS estimators (yellow box-plots), OLS estimators (red box-plots), SBLUE estimators (blue box-plots). Top-left panel: scenario 1, independence. Top-right panel: scenario 2, mild dependence. Bottom panel: scenario 3, full dependence. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

different values of λ (red box-plot, (6)), $\|\hat{\boldsymbol{\beta}}^{(c)}_{\lambda,1}(t)-\boldsymbol{\beta}_1(t)\|_{H^1}$ for different values of λ (yellow box-plots) and $\|\widehat{\boldsymbol{\beta}}^{(\mathbf{R})}_1(t)-\boldsymbol{\beta}_1(t)\|_{H^1}$ (blue box-plot).

From the comparison of the box-plots of $\hat{\boldsymbol{\beta}}^{(c)}_{\lambda}(t)$ and $\hat{\boldsymbol{\beta}}_{\lambda}(t)$ corresponding to $\lambda = 0$ and $\lambda = 1$ with the other cases, it seems more convenient to use just the "less spread" information: $y^{(f)}(t)$ in Scenario 1 and $y^{(d)}(t)$ in Scenario 2. Comparing the precision of $\hat{\boldsymbol{\beta}}^{(c)}_{\lambda}(t)$ and $\hat{\boldsymbol{\beta}}_{\lambda}(t)$ with the one of the $\widehat{\boldsymbol{\beta}}^{(\mathbf{R})}(t)$, however, the SBLUE is the most precise, as expected. Hence, we suggest the use of the whole available information through the use of the SBLUE. Of course, when one of the sources of information has a spread near to zero then the most precise estimator is the one that uses just that piece of information and $\widehat{\boldsymbol{\beta}}^{(\mathbf{R})}(t)$ reflects this behaviour.

6. Summary

Functional data are suitably modelled in separable Hilbert spaces (see Horváth and Kokoszka [7] and Bosq [4]) and L^2 is usually sufficient to handle the majority of the techniques proposed in the literature of functional data analysis.

Instead we consider proper Sobolev spaces; since we guess that the data may provide information on both curve functions and their derivatives. The classical theory for linear regression models is extended in this context by means of the sample of Riesz representatives. Roughly speaking, the Riesz representatives are "quantities" which incorporate both functions and their associated derivative information in a non trivial way. A generalization of the Riesz representatives are proposed to take into account the possible correlation between curves and derivatives. These generalized Riesz representatives are called just "representatives".

Using a sample of representatives, we prove a strong, generalized version of the well known Gauss–Markov theorem for functional linear regression models. Despite the complexity of the problem, we obtain an elegant and simple solution through the use of the representatives which belong to a Sobolev space. This result states that the proposed estimator, which takes into account both information about curves and derivatives (throughout the representatives), is much more efficient than the usual OLS estimator based only on one sample of functions (curves or derivatives). The superiority of the proposed estimator is also showed in the simulation study described in Section 5.

Appendix. Proofs

Proof of Theorem 1. Part (a). We consider the sum of square residuals:

$$S\Big(\boldsymbol{\beta}(t)\Big) = \sum_{j=1}^{r}\sum_{i=1}^{n}\Big(\|y^{(f)}_{ij}(t) - \mathbf{f}(\mathbf{x}_i)^\top\boldsymbol{\beta}(t)\|^2_{L^2} + \|y^{(d)}_{ij}(t) - \mathbf{f}(\mathbf{x}_i)^\top\boldsymbol{\beta}'(t)\|^2_{L^2}\Big)$$

$$= \sum_{j=1}^{r}\sum_{i=1}^{n}\Big(\langle y^{(f)}_{ij}(t) - \mathbf{f}(\mathbf{x}_i)^\top\boldsymbol{\beta}(t), y^{(f)}_{ij}(t) - \mathbf{f}(\mathbf{x}_i)^\top\boldsymbol{\beta}(t)\rangle_{L^2} + \langle y^{(d)}_{ij}(t) - \mathbf{f}(\mathbf{x}_i)^\top\boldsymbol{\beta}'(t), y^{(d)}_{ij}(t) - \mathbf{f}(\mathbf{x}_i)^\top\boldsymbol{\beta}'(t)\rangle_{L^2}\Big).$$

The Gâteaux derivative of $S(\cdot)$ at $\boldsymbol{\beta}(t)$ in the direction of $\mathbf{g}(t) \in (H^1)^p$ is

$$\begin{aligned}\lim_{h\to 0}\frac{S(\boldsymbol{\beta}(t)+h\mathbf{g}(t)) - S(\boldsymbol{\beta}(t))}{h} &= 2\Bigg\{\sum_{j=1}^{r}\sum_{i=1}^{n}\Big(\langle y^{(f)}_{ij}(t) - \mathbf{f}(\mathbf{x}_i)^\top\boldsymbol{\beta}(t), \mathbf{f}(\mathbf{x}_i)^\top\mathbf{g}(t)\rangle_{L^2} \\ &\quad + \langle y^{(d)}_{ij}(t) - \mathbf{f}(\mathbf{x}_i)^\top\boldsymbol{\beta}'(t), \mathbf{f}(\mathbf{x}_i)^\top\mathbf{g}'(t)\rangle_{L^2}\Big)\Bigg\} \\ &= 2r\Big(\langle F^\top\bar{\mathbf{y}}^{(f)}(t) - F^\top F\boldsymbol{\beta}(t), \mathbf{g}(t)\rangle_{(L^2)^p} + \langle F^\top\bar{\mathbf{y}}^{(d)}(t) - F^\top F\boldsymbol{\beta}'(t), \mathbf{g}'(t)\rangle_{(L^2)^p}\Big),\end{aligned} \tag{A.1}$$

where $\bar{\mathbf{y}}^{(f)}(t)$ and $\bar{\mathbf{y}}^{(d)}(t)$ are two $n \times 1$ vectors whose ith elements are

$$\bar{\mathbf{y}}^{(f)}_i(t) = \frac{\sum_{j=1}^{r} y^{(f)}_{ij}(t)}{r}, \qquad \bar{\mathbf{y}}^{(d)}_i(t) = \frac{\sum_{j=1}^{r} y^{(d)}_{ij}(t)}{r}. \tag{A.2}$$

Developing the right-hand side of (A.1), we have that the Gâteaux derivative is

$$\begin{aligned}&= 2r\Big\{\Big(\langle F^\top\bar{\mathbf{y}}^{(f)}(t), \mathbf{g}(t)\rangle_{(L^2)^p} + \langle F^\top\bar{\mathbf{y}}^{(d)}(t), \mathbf{g}'(t)\rangle_{(L^2)^p}\Big) - \Big(\langle F^\top F\boldsymbol{\beta}(t), \mathbf{g}(t)\rangle_{(L^2)^p} + \langle F^\top F\boldsymbol{\beta}'(t), \mathbf{g}'(t)\rangle_{(L^2)^p}\Big)\Big\} \\ &= 2r\Big(\langle F^\top\bar{\mathbf{y}}(t), \mathbf{g}(t)\rangle_{(H^1)^p} - \langle F^\top F\boldsymbol{\beta}(t), \mathbf{g}(t)\rangle_{(H^1)^p}\Big),\end{aligned} \tag{A.3}$$

where $\bar{\mathbf{y}}(t)$ is a $n \times 1$ vector whose ith element is the Riesz representative of $\Big(\bar{\mathbf{y}}^{(f)}_i(t), \bar{\mathbf{y}}^{(d)}_i(t)\Big)$.

The Gâteaux derivative (A.3) is equal to 0 for any $\mathbf{g}(t) \in (H^1)^p$ if and only if $\widehat{\boldsymbol{\beta}}(t)$ is given by the following equation:

$$F^\top F\widehat{\boldsymbol{\beta}}(t) = F^\top \bar{\mathbf{y}}(t),$$

which proves the first statement of the theorem.

Part (b) Definition 2 and model (2) imply that, for any $h(t) \in H^1$,

$$\begin{aligned}\big\langle \mathrm{E}(\tilde{y}_{ij}(t)), h(t)\big\rangle_{H^1} &= \mathrm{E}\big(\langle y_{ij}^{(f)}(t), h(t)\rangle_{L^2}\big) + \mathrm{E}\big(\langle y_{ij}^{(d)}(t), h'(t)\rangle_{L^2}\big)\\ &= \big\langle \mathbf{f}(\mathbf{x}_i)^\top \boldsymbol{\beta}(t), h(t)\big\rangle_{H^1},\end{aligned}$$

then $\mathrm{E}(\bar{\mathbf{y}}(t)) = F\boldsymbol{\beta}(t)$, and hence $\widehat{\boldsymbol{\beta}}(t)$ is unbiased. Moreover,

$$\bar{y}_i(t) - \mathbf{f}(\mathbf{x}_i)^\top \boldsymbol{\beta}(t) = \tilde{\alpha}_i(t) + \frac{\sum_{j=1}^{r} \tilde{\varepsilon}_{ij}(t)}{r}, \quad i = 1, \ldots, n \tag{A.4}$$

where $\tilde{\alpha}_i(t)$ and $\tilde{\varepsilon}_{ij}(t)$ denote the Riesz representatives of $(\alpha_i^{(f)}(t), \alpha_i^{(d)}(t))$ and $(\varepsilon_{ij}^{(f)}(t), \varepsilon_{ij}^{(d)}(t))$, respectively. From the hypothesis (1) and (2) in the model (2), the left-hand side quantities in (A.4) are zero-mean i.i.d. processes, for $i = 1, \ldots, n$. Therefore, the global covariance matrix of $\bar{\mathbf{y}}(t)$ is $\sigma^2 I_n$, where $\sigma^2 = \mathrm{E}(\|\bar{y}_i(t) - \mathbf{f}(\mathbf{x}_i)^\top \boldsymbol{\beta}(t)\|_{H^1}^2)$. Hence, the global covariance matrix of $\widehat{\boldsymbol{\beta}}(t)$ is $\Sigma_{\widehat{\beta}} = \sigma^2 (F^\top F)^{-1}$. □

Proof of Lemma 1. We have that

$$\begin{aligned}\mathrm{E}\|e_i^{(\mathbf{R})}\|_{H^1}^2 &= \mathrm{E}\sum_h \Big\langle \boldsymbol{\Psi}_h, \sum_k \sqrt{\lambda_k} e_{i,k} (\psi_k^{(\mathbf{R})}, \psi_k^{(\mathbf{R})'})\Big\rangle_{\mathcal{L}^2}^2\\ &= \sum_h \mathrm{E} \sum_{k_1,k_2} \sqrt{\lambda_{k_1}}\sqrt{\lambda_{k_2}} e_{i,k_1} e_{i,k_2} \langle \boldsymbol{\Psi}_h, (\psi_{k_1}^{(\mathbf{R})}, \psi_{k_1}^{(\mathbf{R})'})\rangle_{\mathcal{L}^2} \langle \boldsymbol{\Psi}_h, (\psi_{k_2}^{(\mathbf{R})}, \psi_{k_2}^{(\mathbf{R})'})\rangle_{\mathcal{L}^2}\\ &= \sum_{k,h} \lambda_k \langle \boldsymbol{\Psi}_h, (\psi_k^{(\mathbf{R})}, \psi_k^{(\mathbf{R})'})\rangle_{\mathcal{L}^2}^2 = \sum_{k,h} \lambda_h \{\langle \boldsymbol{\Phi}_h, (\phi_k^{(\mathbf{R})}, \phi_k^{(\mathbf{R})'})\rangle_{\mathcal{L}^2_{\mathbf{R}}}\}^2.\end{aligned}$$

From (15), the last term is equal to $\sum_{k,h} \lambda_h (\langle (\phi_h^{(\mathbf{R})}, \phi_h^{(\mathbf{R})'}), \boldsymbol{\Phi}_k\rangle_{\mathcal{L}^2_{\mathbf{R}}})^2$. Hence,

$$\begin{aligned}\mathrm{E}\|e_i^{(\mathbf{R})}\|_{H^1}^2 &= \sum_{k,h} \lambda_h \{\langle (\phi_h^{(\mathbf{R})}, \phi_h^{(\mathbf{R})'}), \boldsymbol{\Phi}_k\rangle_{\mathcal{L}^2_{\mathbf{R}}}\}^2 = \sum_h \lambda_h \sum_k \{\langle (\phi_h^{(\mathbf{R})}, \phi_h^{(\mathbf{R})'}), \boldsymbol{\Phi}_k\rangle_{\mathcal{L}^2_{\mathbf{R}}}\}^2\\ &= \sum_h \lambda_h \|(\phi_h^{(\mathbf{R})}, \phi_h^{(\mathbf{R})'})\|_{\mathcal{L}^2_{\mathbf{R}}}^2 \le \sum_h \lambda_h,\end{aligned}$$

where the last inequality follows from (14). Since $\sum_h \lambda_h < \infty$, we get the thesis. □

A.1. Proof of Theorem 3

The estimator $\widehat{\boldsymbol{\beta}}^{(\mathbf{R})}(t)$ is a linear map which associates an element $\widehat{\boldsymbol{\beta}}^{(\mathbf{R})}(t)$ in $(H^1)^p$ to any nr-tuple $(y_{ij}^{(f)}(t), y_{ij}^{(d)}(t))$. In what follows, we show that it is the "best" among all the linear unbiased closed operators $\mathbf{C} : \mathcal{R} \to (H^1)^p$.

The model (2) may be written in the following vectorial form:

$$\begin{cases}\mathbf{y}^{(f)}(t) = (F \otimes \mathbf{1}_r)\boldsymbol{\beta}(t) + (\boldsymbol{\alpha}^{(f)}(t) \otimes \mathbf{1}_r) + \boldsymbol{\varepsilon}^{(f)}(t)\\ \mathbf{y}^{(d)}(t) = (F \otimes \mathbf{1}_r)\boldsymbol{\beta}'(t) + (\boldsymbol{\alpha}^{(d)}(t) \otimes \mathbf{1}_r) + \boldsymbol{\varepsilon}^{(d)}(t)\end{cases} \tag{A.5}$$

where $\mathbf{1}_r$ is the column vector of length r with all components equal to 1.

In general, if

$$\mathbf{y}^{(1)}(t) = \Big(y_{11}^{(1)}(t), \ldots, y_{1r}^{(1)}(t), y_{21}^{(1)}(t), \ldots, y_{2r}^{(1)}(t), \ldots, y_{n1}^{(1)}(t), \ldots, y_{nr}^{(1)}(t)\Big)^\top$$

and

$$\mathbf{y}^{(2)}(t) = \Big(y_{11}^{(2)}(t), \ldots, y_{1r}^{(2)}(t), y_{21}^{(2)}(t), \ldots, y_{2r}^{(2)}(t), \ldots, y_{n1}^{(2)}(t), \ldots, y_{nr}^{(2)}(t)\Big)^\top$$

are two $nr \times 1$ block vectors in $\mathcal{R}$, we may define the following n dimensional vector

$$\bar{\mathbf{y}}^{(1,2)(\mathbf{R})}(t) = \Big(\bar{y}_1^{(1,2)(\mathbf{R})}(t), \ldots, \bar{y}_n^{(1,2)(\mathbf{R})}(t)\Big)^\top, \tag{A.6}$$

where $\bar{y}_i^{(1,2)(\mathbf{R})}(t)$ is the $H^1_{\mathbf{R}}$ representation of

$$\Big(\frac{\sum_{j=1}^{r} y_{ij}^{(1)}(t)}{r}, \frac{\sum_{j=1}^{r} y_{ij}^{(2)}(t)}{r}\Big).$$

Now we can introduce the following linear operator

$$\mathbf{D}\Big(\mathbf{y}^{(1)}(t), \mathbf{y}^{(2)}(t)\Big) = \mathbf{C}\Big(\mathbf{y}^{(1)}(t), \mathbf{y}^{(2)}(t)\Big) - (F^\top F)^{-1} F^\top\, \bar{\mathbf{y}}^{(1,2)(\mathbf{R})}(t). \tag{A.7}$$

Hence,

$$\begin{aligned}\mathbf{D}(\mathbf{y}^{(f)}(t), \mathbf{y}^{(d)}(t)) &= \mathbf{C}(\mathbf{y}^{(f)}(t), \mathbf{y}^{(d)}(t)) - (F^\top F)^{-1} F^\top \bar{\mathbf{y}}^{(\mathbf{R})}(t)\\ &= \mathbf{C}(\mathbf{y}^{(f)}(t), \mathbf{y}^{(d)}(t)) - \widehat{\boldsymbol{\beta}}^{(\mathbf{R})}(t)\end{aligned} \tag{A.8}$$

and

$$\mathbf{C}(\mathbf{y}^{(f)}(t), \mathbf{y}^{(d)}(t)) = \mathbf{D}(\mathbf{y}^{(f)}(t), \mathbf{y}^{(d)}(t)) + \widehat{\boldsymbol{\beta}}^{(\mathbf{R})}(t).$$

The thesis follows immediately if we prove that $\mathrm{O}(\mathbf{D}(\mathbf{y}^{(f)}(t), \mathbf{y}^{(d)}(t)))$ and $\mathrm{O}(\widehat{\boldsymbol{\beta}}^{(\mathbf{R})}(t))$ are uncorrelated.

Since both $\mathbf{B}(t)$ and $\widehat{\boldsymbol{\beta}}^{(\mathbf{R})}(t)$ are unbiased, $\mathrm{E}\big(\mathbf{D}(\mathbf{y}^{(f)}(t), \mathbf{y}^{(d)}(t))\big) = \mathbf{0}$, and thus we have to prove that

$$\mathrm{E}\big\langle\, \mathrm{O}(\mathbf{D}(\mathbf{y}^{(f)}(t), \mathbf{y}^{(d)}(t))),\ \mathrm{O}(\widehat{\boldsymbol{\beta}}^{(\mathbf{R})}(t) - \boldsymbol{\beta}(t))\, \big\rangle_{H^1} = 0, \tag{A.9}$$

for any choice of linear operator $\mathrm{O} : (H^1)^p \to H^1$.

The proof of equality (A.9) is developed in four steps.

First step. *The goal of this step is to prove that* $\mathbf{D}$ *applied to the deterministic part of the model* $\Big((F \otimes \mathbf{1}_r)\boldsymbol{\beta}(t), (F \otimes \mathbf{1}_r)\boldsymbol{\beta}'(t)\Big)$ *is identically null. As a consequence,*

$$\mathbf{D}\Big(\mathbf{y}^{(f)}(t), \mathbf{y}^{(d)}(t)\Big) = \mathbf{D}\Big(\boldsymbol{\alpha}^{(f)}(t) \otimes \mathbf{1}_r + \boldsymbol{\varepsilon}^{(f)}(t), \boldsymbol{\alpha}^{(d)}(t) \otimes \mathbf{1}_r + \boldsymbol{\varepsilon}^{(d)}(t)\Big). \tag{A.10}$$

From the linearity of the closed operator $\mathbf{C}$, and the zero-mean hypothesis (1) and (2), we have that

$$\begin{aligned}\mathrm{E}\Big(\mathbf{C}\Big(\mathbf{y}^{(f)}(t), \mathbf{y}^{(d)}(t)\Big)\Big) &= \mathrm{E}\Big(\mathbf{C}\Big((F \otimes \mathbf{1}_r)\boldsymbol{\beta}(t) + (\boldsymbol{\alpha}^{(f)}(t) \otimes \mathbf{1}_r) + \boldsymbol{\varepsilon}^{(f)}(t),\ (F \otimes \mathbf{1}_r)\boldsymbol{\beta}'(t) + (\boldsymbol{\alpha}^{(d)}(t) \otimes \mathbf{1}_r) + \boldsymbol{\varepsilon}^{(d)}(t)\Big)\Big)\\ &= \mathbf{C}((F \otimes \mathbf{1}_r)\boldsymbol{\beta}(t), (F \otimes \mathbf{1}_r)\boldsymbol{\beta}'(t)).\end{aligned}$$

Since $\mathrm{E}\Big(\mathbf{C}\Big(\mathbf{y}^{(f)}(t), \mathbf{y}^{(d)}(t)\Big)\Big) = \boldsymbol{\beta}(t)$ we have that

$$\mathbf{C}\Big((F \otimes \mathbf{1}_r)\boldsymbol{\beta}(t), (F \otimes \mathbf{1}_r)\boldsymbol{\beta}'(t)\Big) = \boldsymbol{\beta}(t). \tag{A.11}$$

In addition, from the definition (A.6) if

$$\mathbf{y}^{(1)}(t) = F\boldsymbol{\beta}(t) \otimes \mathbf{1}_r \quad \text{and} \quad \mathbf{y}^{(2)}(t) = F\boldsymbol{\beta}'(t) \otimes \mathbf{1}_r$$

then

$$\bar{\mathbf{y}}^{(1,2)(\mathbf{R})}(t) = F\boldsymbol{\beta}(t). \tag{A.12}$$

Combining (A.7), (A.11) and (A.12) gives

$$\mathbf{D}((F \otimes \mathbf{1}_r)\boldsymbol{\beta}(t), (F \otimes \mathbf{1}_r)\boldsymbol{\beta}'(t)) = \mathbf{0}, \tag{A.13}$$

and hence (A.10).

Second step. *Representation of the linear operator* D_l.

For the linearity of the lth component of $\mathbf{D}$ with respect to the bivariate observations $\Big(y_{ij}^{(1)}(t), y_{ij}^{(2)}(t)\Big)$:

$$D_l\Big(\mathbf{y}^{(1)}(t), \mathbf{y}^{(2)}(t)\Big) = \sum_{i,j} D_{l,ij}\Big(y_{ij}^{(1)}(t), y_{ij}^{(2)}(t)\Big), \tag{A.14}$$

where, for any $i = 1, \dots, n$ and $j = 1, \dots, r$, $D_{l,ij}$ is linear. The domain of $D_{l,ij}$ is contained in $L^2(\mathbf{R}^2)$. Let $(\phi_g)_g$ be an orthonormal base of $H^1_{\mathbf{R}}$. We express the linear operator $y = D_{l,ij}(x)$ in terms of the base $(\boldsymbol{\Psi}_k)_k$ for x and $(\phi_g)_g$ for y. In fact, $\mathcal{R} \subseteq (\mathcal{L}^2)^{nr}$ and $y \in H^1 \subseteq K^*$ (see (17)). Accordingly,

$$D_{l,ij}(y_{ij}^{(1)}(t), y_{ij}^{(2)}(t)) = \sum_{k,g} \langle \boldsymbol{\Psi}_k, (y_{ij}^{(1)}(t),\ y_{ij}^{(2)}(t))^\top \rangle_{\mathcal{L}^2}\, d_{l,ij}^{k,g}\, \phi_g(t), \tag{A.15}$$

where

$$d^{k,g}_{l,ij} = \langle D_{l,ij}(\boldsymbol{\Psi}_k)(t), \phi_g(t)\rangle_{H^1_{\mathbf{R}}}.$$

Third step. *Proof of*

$$\sum_k \sum_{i=1}^n \sum_{j=1}^r m_{l_2,i}\, d^{k,g}_{l_1,ij}\langle \boldsymbol{\Psi}_k,\ (h,h')^\top\rangle_{\mathcal{L}^2} = 0, \quad g, l_1, l_2,\ h \in H^1,$$

where $\mathbf{m}_{l_2} = (m_{l_2,1},\dots,m_{l_2,n})^\top$ *is the* l_2*th row of* $(F^\top F)^{-1}F^\top$*. In particular, since* $H^1_{\mathbf{R}} \subseteq H^1$,

$$\sum_{i,j,k}^n m_{l_2,i}\, d^{k,g}_{l_1,ij}\langle \boldsymbol{\Psi}_k,\ m_{l_2,i}(h,h')^\top\rangle_{\mathcal{L}^2} = 0, \quad g, l_1, l_2,\ h \in H^1_{\mathbf{R}}. \tag{A.16}$$

Let $\mathbf{h}^{(l_2)}(t) \in (H^1)^p$ be the null vector except for the l_2th component which is $h(t) \in H^1$, and let $\mathbf{h}(t) = (F^\top F)^{-1}\mathbf{h}^{(l_2)}(t) \in (H^1)^p$. Setting $\boldsymbol{\beta}(t) = \mathbf{h}(t)$ in (A.13),

$$\begin{aligned}
0 &= D_{l_1}((F\otimes \mathbf{1}_r)\mathbf{h}(t), (F\otimes \mathbf{1}_r)\mathbf{h}'(t))\\
&= D_{l_1}((F\mathbf{h}(t))\otimes \mathbf{1}_r, (F\mathbf{h}'(t))\otimes \mathbf{1}_r)\\
&= D_{l_1}(F(F^\top F)^{-1}\mathbf{h}^{(l_2)}(t)\otimes \mathbf{1}_r, F(F^\top F)^{-1}\mathbf{h}^{(l_2)\prime}(t)\otimes \mathbf{1}_r)\\
&= D_{l_1}(h(t)\mathbf{m}_{l_2}\otimes \mathbf{1}_r, h'(t)\mathbf{m}_{l_2}\otimes \mathbf{1}_r)\\
&= \sum_{i=1}^n\sum_{j=1}^r D_{l_1,ij}(h(t)m_{l_2,i},\ h'(t)m_{l_2,i})\\
&= \sum_g \Big\{ \sum_{k,i,j} (\langle \boldsymbol{\Psi}_k, (m_{l_2,i}h,\ m_{l_2,i}h')^\top\rangle_{\mathcal{L}^2})\, d^{k,g}_{l_1,ij}\Big\}\phi_g(t),
\end{aligned} \tag{A.17}$$

where the last equality is due to (A.15).

Fourth step. *Proof of* (A.9):

$$\mathrm{E}\big\langle \mathrm{O}(\mathbf{D}(\mathbf{y}^{(f)}(t), \mathbf{y}^{(d)}(t))),\ \mathrm{O}(\widehat{\boldsymbol{\beta}}^{(\mathbf{R})}(t) - \boldsymbol{\beta}(t))\big\rangle_{H^1} = 0,$$

for any choice of linear operator $\mathrm{O} : (H^1)^p \to H^1$.

From Theorem 2 and from (19), $\widehat{\boldsymbol{\beta}}^{(\mathbf{R})}(t) - \boldsymbol{\beta}(t) = (F^\top F)^{-1}F^\top \mathbf{e}^{(R)}$, and hence

$$\begin{aligned}
\mathrm{E}\big\langle \mathrm{O}(\mathbf{D}(\mathbf{y}^{(f)}(t), \mathbf{y}^{(d)}(t))),\ \mathrm{O}(\widehat{\boldsymbol{\beta}}^{(\mathbf{R})}(t) - \boldsymbol{\beta}(t))\big\rangle_{H^1} &= \mathrm{E}\big\langle \mathrm{O}(\mathbf{D}(\mathbf{y}^{(f)}(t), \mathbf{y}^{(d)}(t))),\ \mathrm{O}((F^\top F)^{-1}F^\top\mathbf{e}^{(R)}(t))\big\rangle_{H^1}\\
&= \mathrm{E}\Big\langle \mathrm{O}(\mathbf{D}(\boldsymbol{\alpha}^{(f)}(t)\otimes \mathbf{1}_r + \boldsymbol{\varepsilon}^{(f)}(t), \boldsymbol{\alpha}^{(d)}(t)\otimes \mathbf{1}_r + \boldsymbol{\varepsilon}^{(d)}(t))),\\
&\qquad \mathrm{O}((F^\top F)^{-1}F^\top \mathbf{e}^{(R)}(t))\Big\rangle_{H^1},
\end{aligned} \tag{A.18}$$

where the last equality is a consequence of (A.10).

Since $x \in (H^1)^p \subseteq (K^*)^p$ (see (17)), we express the linear operator $y = \mathrm{O}(x)$ in terms of the base $(\phi_{g_1}\times\phi_{g_2}\times\cdots\times\phi_{g_p})_{g_1,\dots,g_p}$ for x and $(\zeta_h)_h$ for y, where $(\zeta_h)_h$ is an orthonormal base of H^1. To begin with, from the linearity of the operator $\mathrm{O} : (H^1)^p \to H^1$, we have that

$$\mathrm{O}(b_1(t),\dots,b_p(t)) = \sum_{l=1}^p \mathrm{O}(\underbrace{0,\dots,0}_{l-1 \text{ times}}, b_l(t), \underbrace{0,\dots,0}_{p-l \text{ times}}).$$

Since $b_l(t) = \sum_g \langle b_l(t), \phi_g(t)\rangle_{H^1_{\mathbf{R}}}\,\phi_g(t) = \sum_g b^g_l \phi_g(t)$, where $b^g_l = \langle b_l(t), \phi_g(t)\rangle_{H^1_{\mathbf{R}}}$, we have

$$\mathrm{O}(b_1(t),\dots,b_p(t)) = \sum_{l,g} b^g_l\ \mathrm{O}(\underbrace{0,\dots,0}_{l-1 \text{ times}}, \phi_g(t), \underbrace{0,\dots,0}_{p-l \text{ times}}).$$

Now, setting

$$O^{g,h}_l = \big\langle \mathrm{O}(\underbrace{0,\dots,0}_{l-1 \text{ times}}, \phi_g(t), \underbrace{0,\dots,0}_{p-l \text{ times}}), \zeta_h(t)\big\rangle_{H^1},$$

then we have the representation of O in terms of the required bases:

$$\mathrm{O}(b_1(t),\dots,b_p(t)) = \sum_{l,g,h} b^g_l\ O^{g,h}_l\ \zeta_h(t).$$

Hence, from Eqs. (A.18), (A.14) and (A.15), the thesis (A.9) becomes

$$\mathrm{E}\Big\langle \sum_{l,g,h}\Big\{\sum_{i,j,k}(\langle \boldsymbol{\Psi}_k,\ (\alpha_i^{(f)}(t)+\varepsilon_{ij}^{(f)}(t),\alpha_i^{(d)}(t)+\varepsilon_{ij}^{(d)}(t))^\top\rangle_{\mathcal{L}^2})\, d_{l_1,ij}^{k,g}\Big\} O_l^{g,h}\ \zeta_h(t),$$

$$\sum_{l,g,h}\Big(\big\langle \mathbf{e}^{(R)}(t)^\top \mathbf{m}_l,\ \phi_g(t)\big\rangle_{H^1_{\mathbf{R}}}\Big) O_l^{g,h}\ \zeta_h(t)\Big\rangle_{H^1} = 0.$$

From (11) and (13), since $\langle \zeta_{h_1}, \zeta_{h_2}\rangle_{H^1} = \delta_{h_1}^{h_2}$, the left-hand side of the last equation becomes

$$\begin{aligned}
&\mathrm{E}\Big\langle \sum_{l,g,h} O_l^{g,h}\ \zeta_h(t)\Big\{\sum_{i,j,k}(\langle \boldsymbol{\Psi}_k,\ (\alpha_i^{(f)}(t)+\varepsilon_{ij}^{(f)}(t),\alpha_i^{(d)}(t)+\varepsilon_{ij}^{(d)}(t))^\top\rangle_{\mathcal{L}^2})\, d_{l,ij}^{k,g}\Big\},\\
&\quad \sum_{l,g,h} O_l^{g,h}\ \zeta_h(t)\big\langle \mathbf{e}^{(R)}(t)^\top \mathbf{m}_{l_2},\ \phi_g(t)\big\rangle_{H^1_{\mathbf{R}}}\Big\rangle_{H^1}\\
&= \mathrm{E}\Big(\sum_{l_1,l_2,g_1,g_2,h} O_{l_1}^{g_1,h} O_{l_2}^{g_2,h}\sum_{i_1,j,k_1} X_{i_1j,k_1}\, d_{l_1,i_1j}^{k_1,g_1}\sum_{i_2,k_2}\sqrt{\lambda_{k_2}} e_{i_2,k_2}\ \langle \psi_{k_2}^{(\mathbf{R})}(t),\ \phi_{g_2}(t)\rangle_{H^1_{\mathbf{R}}}\, m_{l_2,i_2}\Big)\\
&= \sum_{l_1,l_2,g_1,g_2,h} O_{l_1}^{g_1,h} O_{l_2}^{g_2,h}\sum_{i_1,i_2,j}\sum_{k_1,k_2}\sqrt{\lambda_{k_2}}\ d_{l_1,i_1j}^{k_1,g_1}\ m_{l_2,i_2}\ \mathrm{E}\big(X_{i_1j,k_1} e_{i_2,k_2}\big)\langle \psi_{k_2}^{(\mathbf{R})}(t),\ \phi_{g_2}(t)\rangle_{H^1_{\mathbf{R}}}\\
&= \sum_{l_1,l_2,g_1,g_2,h} O_{l_1}^{g_1,h} O_{l_2}^{g_2,h}\sum_{i_1,i_2,j}\sum_{k_1,k_2}\delta_{i_1}^{i_2}\delta_{k_1}^{k_2}\lambda_{k_1}\ d_{l_1,i_1j}^{k_1,g_1}\ m_{l_2,i_2}\langle \psi_{k_2}^{(\mathbf{R})}(t),\ \phi_{g_2}(t)\rangle_{H^1_{\mathbf{R}}}\\
&= \sum_{l_1,l_2,g_1,g_2,h} O_{l_1}^{g_1,h} O_{l_2}^{g_2,h}\sum_{i,j}\sum_{k} d_{l_1,ij}^{k,g_1}\ m_{l_2,i}\lambda_k\langle \psi_{k}^{(\mathbf{R})}(t),\ \phi_{g_2}(t)\rangle_{H^1_{\mathbf{R}}}\\
&= \sum_{l_1,l_2,g_1,g_2,h} O_{l_1}^{g_1,h} O_{l_2}^{g_2,h}\sum_{i,j}\sum_{k} d_{l_1,ij}^{k,g_1}\ m_{l_2,i}\big(\lambda_k\langle \boldsymbol{\Psi}_k(t),\ (\phi_{g_2}(t),\phi'_{g_2}(t))^\top\rangle_{\mathcal{L}^2_{\mathbf{R}}}\big)\\
&= \sum_{l_1,l_2,g_1,g_2,h} O_{l_1}^{g_1,h} O_{l_2}^{g_2,h}\sum_{i,j}\sum_{k} d_{l_1,ij}^{k,g_1}\ m_{l_2,i}\langle \boldsymbol{\Psi}_k(t),\ (\phi_{g_2}(t),\phi'_{g_2}(t))^\top\rangle_{\mathcal{L}^2}\\
&= 0,
\end{aligned}$$

the last equality being a consequence of (A.16).

References

[1] G. Aletti, C. May, C. Tommasi, Optimal designs for linear models with functional responses, in: E.G. Bongiorno, E. Salinelli, A. Goia, P. Vieu (Eds.), Contributions in Infinite-Dimensional Statistics and Related Topics, Società Editrice Esculapio, ISBN: 978-8874-887637, 2014, pp. 19–24. http://dx.doi.org/10.15651/978-88-748-8763-7.

[2] G. Aletti, C. May, C. Tommasi, On Applying Optimal Design of Experiments when Functional Observations Occur, Springer International Publishing, Cham, ISBN: 978-3-319-31266-8, 2016, pp. 1–9. http://dx.doi.org/10.1007/978-3-319-31266-8_1.

[3] S. Baraldo, F. Ieva, A.M. Paganoni, V. Vitelli, Outcome prediction for heart failure telemonitoring via generalized linear models with functional covariates, Scand. J. Stat. (ISSN: 0303-6898) 40 (3) (2013) 403–416. http://dx.doi.org/10.1111/j.1467-9469.2012.00818.x.

[4] D. Bosq, Linear Processes in Function Spaces, in: Lecture Notes in Statistics, vol. 149, Springer-Verlag, New York, ISBN: 0-387-95052-4, 2000, http://dx.doi.org/10.1007/978-1-4612-1154-9.

[5] F. Ferraty, P. Vieu, Nonparametric Functional Data Analysis, in: Springer Series in Statistics, Springer, New York, ISBN: 0-387-30369-3, 2006, 978-0387-30369-7.

[6] P. Hall, H.-G. Müller, F. Yao, Estimation of functional derivatives, Ann. Statist. (ISSN: 0090-5364) 37 (6A) (2009) 3307–3329. http://dx.doi.org/10.1214/09-AOS686.

[7] L. Horváth, P. Kokoszka, Inference for Functional Data with Applications, in: Springer Series in Statistics, Springer, New York, ISBN: 978-1-4614-3654-6, 2012, http://dx.doi.org/10.1007/978-1-4614-3655-3.

[8] A. Menafoglio, P. Secchi, M. Dalla Rosa, A universal kriging predictor for spatially dependent functional data of a Hilbert space, Electron. J. Stat. (ISSN: 1935-7524) 7 (2013) 2209–2240. http://dx.doi.org/10.1214/13-EJS843.

[9] G. Perrin, C. Soize, D. Duhamel, C. Funfschilling, Karhunen-Loève expansion revisited for vector-valued random fields: scaling, errors and optimal basis, J. Comput. Phys. (ISSN: 0021-9991) 242 (2013) 607–622. http://dx.doi.org/10.1016/j.jcp.2013.02.036.

[10] D. Pigoli, L. Sangalli, Wavelets in functional data analysis: Estimation of multidimensional curves and their derivatives, Comput. Statist. Data Anal. 56 (6) (2012) 1482–1498.

[11] J.O. Ramsay, B.W. Silverman, Functional Data Analysis, second ed., in: Springer Series in Statistics, Springer, New York, ISBN: 978-0387-40080-8, 2005, 0-387-40080-X.

[12] L. Sangalli, P.b. Secchi, S. Vantini, A. Veneziani, Efficient estimation of three-dimensional curves and their derivatives by free-knot regression splines, applied to the analysis of inner carotid artery centrelines, J. Roy. Statist. Soc. Ser. C 58 (3) (2009) 285–306.

[13] Q. Shen, J. Faraway, An F test for linear models with functional responses, Statist. Sinica (ISSN: 1017-0405) 14 (4) (2004) 1239–1257.