uc3m | Universidad Carlos III de Madrid

MSc in Big Data Analytics (2017–2018)

A goodness-of-fit test for the functional linear model with functional response

by

Gonzalo Álvarez Pérez

Madrid, September 13th, 2018

Supervisor:
Eduardo García Portugués

Acknowledgements

I would like to express my gratitude to Eduardo García Portugués for his supervision, patience, and confidence on me. I am also grateful to all the people that have contributed to make a great experience out of this year and, especially, to my parents, this accomplishment would not have been possible without them.

Abstract

The field of functional data analysis has seen a rapid development over the last two decades due to the technological advances that have enabled to collect statistical information over extremely fine grids. In the search of new ways to exploit these rich datasets, novel statistical methods have been developed. This work proposes a goodness-of-fit test for the null hypothesis of a functional linear model with functional response based on the random projections paradigm. The test is a generalization of a previous goodness-of-fit test, constructed for the functional linear model with scalar response. The test statistic is simple to compute applying geometrical and matrix reasonings, and the calibration of the test on its distribution is studied by means of a wild bootstrap on the residuals, for both expansions on Fourier and functional principal components bases.

Keywords: Functional data; functional linear model; goodness-of-fit; random projections; Cramér–von Mises statistic; Fourier basis; functional principal components; wild bootstrap.

Notation and acronyms

Notation

All along this work we deal with functions, vectors, and scalars. It is therefore needed to introduce some notation to establish the difference between these objects:

- Spaces and fields will be denoted with the typefont[1] $\mathbb{X}$. Random functions embedded in a Hilbert space will be denoted by uppercase calligraphic symbols $\mathcal{X}$, whereas their observations with lowercase calligraphic characters $\mathcal{x}$, in analogy with the notation employed for random variables. The projection of these functions on a truncated basis of k elements is denoted by $\mathcal{X}^{(k)}$ and $\mathcal{x}^{(k)}$, respectively. Moreover, operators acting on this kind of functions will be denoted also by uppercase calligraphic symbols, *e.g.* $\mathcal{F}(\mathcal{X})$.

- Matrices will be denoted by uppercase bold letters $\mathbf{X}$, and vectors by bold lowercase $\mathbf{x}$. Their projections on k-truncated bases will be $\mathbf{X}_k$ and $\mathbf{x}_k$. It should be pointed out that the possible confusion between matrix $\mathbf{X}$ and random vector $\mathbf{X}$ does not take place in the text. Scalars will be denoted by regular typefonts x.

Acronyms

CvM: Cramér–von Mises
ECDF: Empirical Cumulative Distribution Function
FDA: Functional Data Analysis
FLM: Functional Linear Model
FLMFR: Functional Linear Model with Functional Response
FLMSR: Functional Linear Model with Scalar Response
FPC: Functional Principal Component

[1] This typefont will be also used in two special cases with no risk of confusion: to denote the expectancy $\mathbb{E}$ of a random variable, and to denote probabilities $\mathbb{P}$.

FPCA: Functional Principal Components Analysis
KDE: Kernel Density Estimate
KS: Kolmogorov–Smirnov
PC: Principal Component
PCA: Principal Component Analysis
PCvM: Projected Cramér–von Mises
PKS: Projected Kolmogorov–Smirnov
RMPP: Residual Marked empirical Process based on Projections
SVD: Singular Value Decomposition

Contents

1 Introduction

Functional data analysis has become very popular during the last decades due to the technological progress in monitoring devices, electronic equipment, computational tools and memory capacity, that allow to observe phenomena in a more accurate way by producing statistical information sampled over finer and finer grids. This has led to an increasing availability of data for continuous processes, usually time-dependent data —as stock prices in finances, temperature evolution in meteorology, and path trajectories for objects in movement in kinematics—, but not necessarily —as wavelength-dependent absorption spectroscopies in physics—. More importantly, functional data analysis is a very attractive field of research, since it broadens many specific topics in statistics, linear algebra, mathematical analysis and programming, such as functional analysis, statistical inference, modeling, resampling, stochastic processes, etc. In particular, this work employs many of these ideas, starting from the very abstract conception of a goodness-of-fit test to its code development in `R`. Some of the most known references in the field of functional data analysis are Ramsay and Silverman (2005), Ferraty and Vieu (2006), Ferraty and Romain (2011), and Kokoszka and Reimherr (2017).

The functional data is commonly related to a univariate functional variable. In such cases, it may be useful to determine the relation of the variables by means of a regression model, that can help to predict the functional output $\mathcal{Y}$ from the functional input $\mathcal{X}$. In the context of regression models with functional covariate and functional response:

$$\mathcal{Y} = \mathcal{M}(\mathcal{X}) + \mathcal{E}.$$

The simplest and most known parametric model is the functional linear model with functional response:

$$\mathcal{Y} = \mathcal{M}_{\mathcal{B}}(\mathcal{X}) + \mathcal{E},$$

where $\mathcal{E}$ is a random function accounting for the error. The model can be thought as a generalization of the multivariate regression case. The main difference is that the regression coefficient is now an unknown operator $\mathcal{M}_{\mathcal{B}}$ that belongs to the class of linear and bounded operators between Hilbert spaces. Therefore, it satisfies the Riesz representation theorem, which enables us to assume $\mathcal{M}_{\mathcal{B}} \in \mathfrak{B}$, being $\mathbb{H}_1$ and $\mathbb{H}_2$ Hilbert spaces and:

$$\mathfrak{B} := \{\mathcal{X} \in \mathbb{H}_1 \mapsto \mathcal{Y} = \langle\langle \mathcal{X}, \mathcal{B}\rangle\rangle \in \mathbb{H}_2 : \mathcal{B} \in \mathbb{H}_1 \otimes \mathbb{H}_2\},$$

where $\langle\langle \mathcal{X}, \mathcal{B}\rangle\rangle$ is defined by:

$$(\mathcal{X}, \mathcal{B}) \in \mathbb{H}_1 \times (\mathbb{H}_1 \otimes \mathbb{H}_2) \mapsto \langle\langle \mathcal{X}, \mathcal{B}\rangle\rangle = \langle \mathcal{X}(\cdot), \mathcal{B}(\cdot, \star)\rangle \in \mathbb{H}_2.$$

In such a way that:

$$\mathcal{Y} = \mathcal{M}_{\mathcal{B}}(\mathcal{X}) + \mathcal{E} = \langle\langle \mathcal{X}, \mathcal{B}\rangle\rangle + \mathcal{E},$$

We propose a goodness-of-fit test for the the null hypothesis of the functional linear model with functional response:

$$H_0 : \mathcal{M} \in \mathfrak{B} \qquad \text{or} \qquad H_0 : \mathcal{M} = \mathcal{M}_{\mathcal{B}} \text{ for some } \mathcal{B} \in \mathbb{H}_1 \otimes \mathbb{H}_2\},$$

which is a generalization of the goodness-of-fit test for the functional linear model with scalar response, already proposed by García-Portugués et al. (2014). There, a fruitful methodology to study functional data is employed: the use of random projections —usually suitable when treating high-dimensional data— as a way to overcome the curse of dimensionality. The aim is to characterize the behavior of an infinite-dimensional functional process by means of the behavior of the one-dimensional inner products of the functional process with suitable random functions, allowing to benefit from the numerous procedures that are available in the one-dimensional case. Instead of testing a given null hypothesis in the functional space, we test the transformation of this hypothesis on a one-dimensional randomly chosen projection. The paradigm of random projections has already been applied for the goodness-of-fit tests for parametric families of functional distributions by Cuesta-Albertos et al. (2007), including goodness-of-fit tests for gaussianity and for the Black-Scholes model. Moreover, the test statistic is of a Cramer–von Mises type and is based on a generalization of a previous test designed by Escanciano (2006) for the case of a regression model with multivariate covariates. It is easy to compute using geometrical and matrix arguments. Moreover, The calibration of the test its distribution is studied applying a resampling procedure based on a wild bootstrap on the residuals.

This work is organized as follows. Some background on functional data, such as Hilbert spaces and basis expansions —Fourier and functional principal components—, is introduced on Chapter 2. The functional linear model and its estimations are discussed in Chapter 3. Chapter 4 introduces the goodness-of-fit test, including the random projections paradigm and the theoretical arguments of the test, jointly with the bootstrap calibration procedure. The finite sample properties of the test are illustrated by a simulation study in Chapter 5. Conclusions and possible extensions of the work are outlined in Chapter 6. Finally, the contributed code of the project is detailed in Appendix A.

2 Functional data

In this work we aim on proposing a goodness-of-fit test for the null hypothesis of the Functional Linear Model (FLM) with functional response (FLMFR). This chapter is intended to give the reader some basic notions on functional data.

2.1 Basic notions

One of the inherent and crucial problems when managing functional data is the choice of a suitable functional space, being metric, Banach, and Hilbert spaces the most common elections. These spaces are endowed with increasing richer structure, since the mechanisms available in the former are included in the latter. Particularly, metric spaces allow to measure distances between functions. Furthermore, in Banach spaces functions can be measured and Cauchy sequences are convergent. Finally, Hilbert spaces possess an inner product, which enables to consider functional bases. We elaborate these concepts as follows:

Definition 1.1. A *metric space* is an ordered pair $(\mathbb{V}, d)$ where $\mathbb{V}$ is a set and d is a function $d\colon \mathbb{V}\times\mathbb{V} \to \mathbb{R}$ such that for any $x, y, z \in \mathbb{V}$:

1. $d(x, y) = 0 \Leftrightarrow x = y$,
2. $d(x, y) = d(y, x)$,
3. $d(x, z) \leq d(x, y) + d(y, z)$.

In such a case d is said to be a *metric* on $\mathbb{V}$.

Definition 1.2. An *inner product space* is a real or complex vector space $\mathbb{V}$ over the field $\mathbb{F}$ together with a map $\langle\cdot,\cdot\rangle : \mathbb{V} \times \mathbb{V} \to \mathbb{F}$ that satisfies for all vectors $x, y, z \in \mathbb{V}$ and all scalars $a \in \mathbb{F}$:

1. $\langle x, y\rangle = \overline{\langle y, x\rangle}$
2. $\langle ax, y\rangle = a\langle x, y\rangle$ and $\langle x + y, z\rangle = \langle x, z\rangle + \langle y, z\rangle$
3. $\langle x, x\rangle \geq 0$, with $\langle x, x\rangle = 0 \Leftrightarrow x = 0$.

In such a case $\langle \cdot, \cdot \rangle$ is said to be an *inner product*.

Definition 1.3. A *Hilbert space* $\mathbb{H}$ is an inner product space that is also a complete metric space with respect to the distance function induced by the inner product, *i.e.*, every Cauchy sequence in $\mathbb{H}$ converges in $\mathbb{H}$.

A Hilbert space is a natural generalization of the finite-dimensional Euclidean space, that is achieved via an inner product. The inner product is the crucial structure, since it allows to project the elements of the Hilbert space onto another elements, thus enabling to build bases and norms. While there exist many types of metrics and norm spaces, the L^p spaces are among the most common. The $L^p[a, b]$ space, $a, b \in \mathbb{R}$, $1 \leq p < \infty$, is defined as the set of all functions $\mathcal{F} : [a, b] \to \mathbb{R}$ such that their norm $||\mathcal{F}||_p$ is finite, where

$$||\mathcal{F}||_p = \left(\int_a^b |\mathcal{F}(t)|_p dt \right)^{\frac{1}{p}}.$$

The election of an arbitrary interval $[a, b]$ is done just to fix the integration limits, since any interval can be considered without conceptual modifications. The space L^2 is the only one having an associated inner product $\langle \cdot, \cdot \rangle$, such that $||\mathcal{F}||_p = \langle \mathcal{F}, \mathcal{F} \rangle^{1/2}$. For two functions $\mathcal{F}, \mathcal{G} \in L^2[a, b]$, their inner product is defined as

$$\langle \mathcal{F}, \mathcal{G} \rangle = \int_a^b \mathcal{F}(t)\mathcal{G}(t)dt.$$

In what follows we will consider as our working space the Hilbert space $\mathbb{H} = L^2[a, b]$. However, any other Hilbert space could be employed without any conceptual change. The inner product allows for a basis representation of the elements of $\mathbb{H}$ and, given a functional basis $\{\Psi_j\}_{j=1}^{\infty}$ of $\mathbb{H}$, then any function $\mathcal{X}$ in $\mathbb{H}$ can be expressed by the linear combination:

$$\mathcal{X} = \sum_{j=1}^{\infty} x_j \Psi_j,$$

where $x_j = \langle \mathcal{X}, \Psi_j \rangle$, $j \geq 1$. A basis is said to be orthogonal if $\langle \Psi_i, \Psi_j \rangle = 0$, $i \neq j$ and orthonormal if, in addition, $\langle \Psi_j, \Psi_j \rangle = 1$, $j \geq 1$. Typical examples of basis of $\mathbb{H}$ are: the collection of monomials that are used to construct power series, $\{1, t, t^2, t^3, ..., t^k, ...\}$; the B-splines basis (see *e.g.* de Boor (2001)); or the Fourier series system, a deterministic set of basis elements $\{1, \sin(2\pi jx), \cos(2\pi jx)\}_{j=1}^{\infty}$ that does not depend on the data. Fourier expansions present excellent computational properties if the observations are equally spaced and are the natural election to deal with periodic datasets, such as the weather cycle. Nonetheless, if the the observations are not periodic, it can be problematical to use a Fourier basis expansion, since periodicity is enforced in the representation. Moreover, the use of Fourier expansions can lead to spurious signals near the boundaries (border effects) or sharp transitions (ringing artifacts in signal processing). These issues can be addressed by considering data-driven bases, *e.g.* Functional Principal Components (FPCs), which will be treated in detail in Sections 2.2 and 2.3.

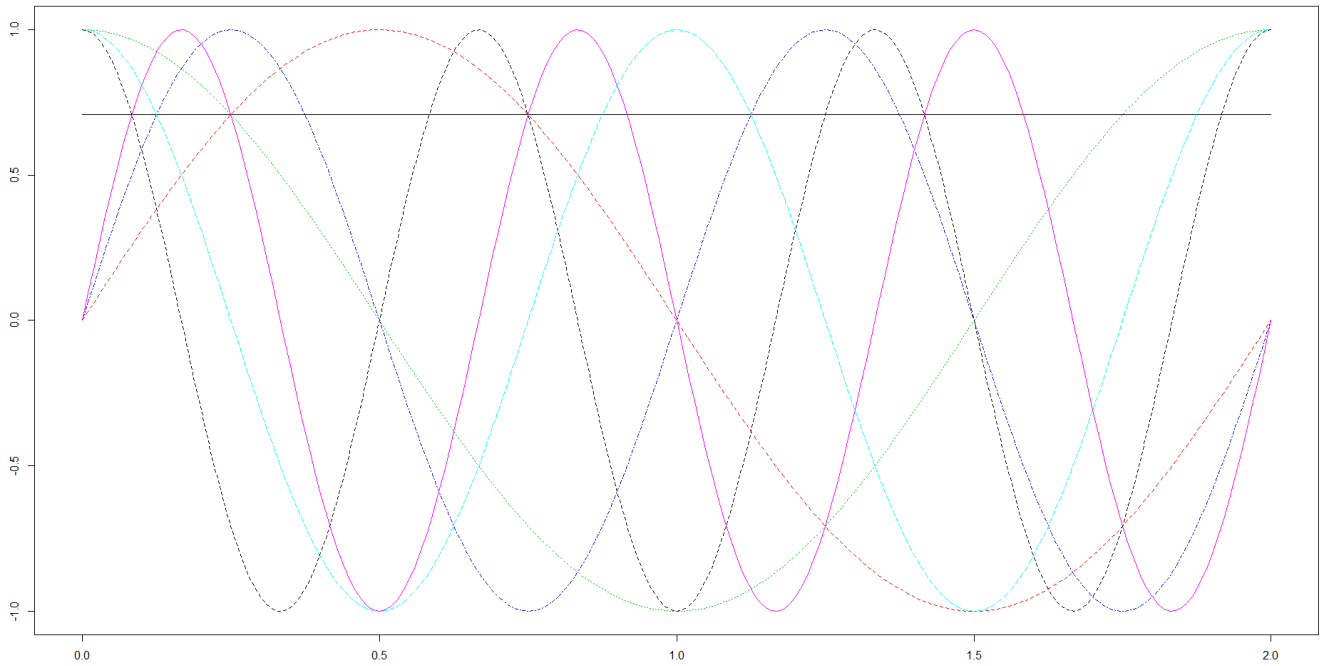

Figure 2.1: A Fourier basis with 7 elements in $L^2[0,2]$.

For the development of the test statistic, we will also need to introduce a p-truncated basis $\{\Psi_j\}_{j=1}^p$, which corresponds to the first p elements of the infinite basis $\{\Psi_j\}_{j=1}^\infty$. The representation of $\mathcal{X}$ in this truncated basis is denoted by

$$\mathcal{X}^{(p)} = \sum_{j=1}^{p} x_j \Psi_j.$$

We will denote by $\mathbf{x}$ and by $\mathbf{x}_p$ the vector of coefficients of $\mathcal{X}$ in the original and in the p-truncated basis, respectively.

In order to manage functional random projections we define the functional analogue of the euclidean p-sphere[1] $\mathbb{S}^p = \{\mathbf{x} \in \mathbb{R}^p : ||\mathbf{x}||_{\mathbb{R}^p} = 1\}$, *i.e.*, the *functional sphere of* $\mathbb{H}$, defined as $\mathbb{S}_{\mathbb{H}} = \{\mathcal{F} \in \mathbb{H} : ||\mathcal{F}||_{\mathbb{H}} = 1\}$, and the *functional sphere of dimension* p, which is the set of functions of $\mathbb{H}$ that, expressed in the p-truncated basis, have unit norm: $\mathbb{S}^p_{\mathbb{H}} = \left\{\mathcal{F} = \sum_{j=1}^p x_j \Psi_j \in \mathbb{H} : ||\mathcal{F}||_{\mathbb{H}} = 1\right\}$.

The relationship between $\mathbb{S}^p$ and $\mathbb{S}^p_{\mathbb{H}}$ is particularly interesting to develop the test. It is given in García-Portugués et al. (2014) and we adapt it here to our conventions for completeness, since it will be used for the development of the test statistic. Let $\mathbf{\Psi} = (\langle \Psi_i, \Psi_j \rangle)_{ij}$ be the matrix of inner products of the p-truncated basis, $\mathbb{S}^p_{\mathbf{\Psi}} = \left\{\mathbf{x} \in \mathbb{R}^p : \mathbf{x}^T \mathbf{\Psi} \mathbf{x} = 1\right\}$, the p-ellipsoid generated by this matrix and the Cholesky decomposition $\mathbf{R}^T\mathbf{R}$ of $\mathbf{\Psi}$ (a semi-positive matrix). First of all, we have the trivial isomorphism that maps elements of $\mathbb{S}^p_{\mathbb{H}}$ to elements of $\mathbb{S}^p_{\mathbf{\Psi}}$ by means of the functional coefficients: $\mathcal{S} : \mathcal{F} = \sum_{j=1}^p x_j \Psi_j \in \mathbb{S}^p_{\mathbb{H}} \mapsto \mathcal{S}(\mathcal{F}) = \mathbf{x} \in \mathbb{S}^p_{\mathbf{\Psi}}$. Recall that functions $\mathcal{S}$ and $\mathcal{S}^{-1}$ are well defined because

[1] Observe that we are denoting the sphere of dimension $p-1$ in $\mathbb{R}^{p-1}$ by $\mathbb{S}^p$ and not by $\mathbb{S}^{p-1}$, the standard convention in mathematics. The reason is that $\mathbb{S}^p$ offers a more immediate connection with the dimension of the p-truncated basis.

$$||\mathcal{F}||^2_{\mathbb{H}} = \left\langle \sum_{j=1}^{p} x_j \Psi_j, \sum_{j=1}^{p} x_j \Psi_j \right\rangle = \mathbf{x}^T \mathbf{\Psi} \mathbf{x}.$$

We must consider also a linear transformation from $\mathbb{S}^p$ to $\mathbb{S}^p_{\mathbf{\Psi}}$, which is given by $\mathcal{R} : \mathbf{x} \in \mathbb{S}^p \mapsto \mathcal{R}(\mathbf{x}) = \mathbf{R}^{-1}\mathbf{x} \in \mathbb{S}^p_{\mathbf{\Psi}}$ and whose Jacobian is $|\mathbf{R}|^{-1}$, the determinant of the matrix $\mathbf{R}^{-1}$.

Using these two transformations, the integration of a functional operator $\mathcal{T}$ with respect to a functional covariate $\mathcal{D}^{(p)}$ in $\mathbb{S}^p_{\mathbb{H}}$ can be reduced to a real integration on the p-sphere:

$$\int_{\mathbb{S}^p_{\mathbb{H}}} \mathcal{T}\left(\mathcal{D}^{(p)}\right) d\mathcal{D}^{(p)} = \int_{\mathbb{S}^p_{\mathbf{\Psi}}} \mathcal{T}\left(\sum_{j=1}^{p} d_j \Psi_j\right) d\mathbf{d}_p = \int_{\mathbb{S}^p} |\mathbf{R}|^{-1} \mathcal{T}\left(\sum_{j=1}^{p} (\mathbf{R}^{-1} g)_j \Psi_j\right) d\mathbf{d}_p \tag{2.1}$$

In the case where the basis is orthonormal, $\mathbf{\Psi}$ and $\mathbf{R}$ are the identity matrix of order p. Then the coefficients of $\mathcal{D}^{(p)} \in \mathbb{S}^p_{\mathbb{H}}$ in the basis $\{\Psi_j\}_{j=1}^p$ belong to $\mathbb{S}^p$ without any transformation.

2.2 Functional principal components

Functional Principal Components Analysis (FPCA) provides a data-driven basis[2]. There are several approaches to compute the PCs bases in the practice —see Ramsay and Silverman (2005) (pages 162–166) and Jolliffe (2002) (pages 318–320). In this work, we will focus in two of them: the discretization of the observed functions and the use of more general numerical quadrature arguments, as the Simpson's rule. The latter will be discussed in 2.3.

Texts on multivariate data analysis tend to define PCA as the task of finding the eigenvalues and eigenvectors of the covariance matrix. The simplest technique is to discretize the functions[3] $\mathcal{X}_i$ in a fine grid of $N_{\mathcal{X}}$ equally spaced values s_j, yielding an $n \times N_{\mathcal{X}}$ data matrix $\mathbf{X}$. By denoting the covariance matrix[4] $\mathbf{V} = n^{-1}\mathbf{X}^T\mathbf{X}$, the eigenanalysis to be done is:

$$\mathbf{V}\mathbf{u} = \lambda \mathbf{u} \tag{2.2}$$

for n-vectors $\mathbf{u}$.

Note that we might have $N_{\mathcal{X}} \gg n$. Rather than working with the $N_{\mathcal{X}} \times N_{\mathcal{X}}$ matrix $\mathbf{V}$, one possible way to solve the eigenequation (2.2) is to find the Singular Value Decomposition (SVD) $\mathbf{UDW}$ of $\mathbf{X}$. The variance matrix satisfies $n\mathbf{V} = \mathbf{WD}^2\mathbf{W}$, and thus the non-zero eigenvalues of $\mathbf{V}$ are the squares of the singular values of $\mathbf{X}$. The corresponding eigenvectors are given by columns of $\mathbf{U}$. Actually, employing the SVD to perform PCA is more convenient from the numerical point of view than forming the covariance matrix, since the formation of $\mathbf{X}^T\mathbf{X}$ can cause loss of precision. This is detailed in books on PCA, (see *e.g.*

[2]To be precise, the population version of FPCs is a deterministic basis formed by the eigenfunctions which diagonalize the covariance operator. The estimation of such eigenfunctions in terms of the the empirical sample covariance operator is indeed the estimation of the FPCs, which are the ones forming a data-driven basis. This occurs even for traditional PCs, not necessarily considered in the functional framework.

[3]Here we detail the FPCA only for the functional covariate $\mathcal{X}$ for simplicity. The procedure is analogous for the response $\mathcal{Y}$.

[4]Some authors may prefer to use $n-1$ instead of n to define the variance-covariance matrix.

Jolliffe (2002)). An example of a matrix that has a stable SVD, but forming its $\mathbf{X}^T\mathbf{X}$ can be disastrous, is given by the Läuchli matrix:

$$\begin{pmatrix} 1 & 0 & 0 & \epsilon \\ 1 & 0 & \epsilon & 0 \\ 1 & \epsilon & 0 & 0 \end{pmatrix},$$

where ϵ is a tiny number.

2.3 Numerical quadrature methods

The simple approach outlined in Section 2.2 does not usually yield the best results. Therefore, we apply more sophisticated numerical quadrature methods instead, which are based on numerical integration schemes.

The sample covariance between $\mathcal{X}(s)$ and $\mathcal{X}(t)$ can be defined as $\mathcal{V}(s,t) = n^{-1}\sum_{i=1}^{n}\mathcal{X}_i(s)\mathcal{X}_i(t)$. The FPCs emerge as the solutions to an eigenequation involving this covariance function. Specifically, the FPC eigenfunctions $\mathcal{P}_j(s)$ are found, each of them satisfying:

$$\int \mathcal{V}(s,t)\mathcal{P}(t)dt = \rho\mathcal{P}(s) \tag{2.3}$$

for a suitable eigenvalue ρ. The left-hand side of (2.3) is an integral transform of the weight function $\mathcal{P}$, the covariance operator $\tilde{\mathcal{V}}$, which is given by:

$$\tilde{\mathcal{V}}(\mathcal{P}) := \int \mathcal{V}(\cdot,t)\mathcal{P}(t)dt. \tag{2.4}$$

Consequently, we may express the eigenequation directly as:

$$\tilde{\mathcal{V}}(\mathcal{P}) = \rho\mathcal{P}, \tag{2.5}$$

where $\mathcal{P}$ is now an eigenfunction rather than an eigenvector. The eigenequation (2.3) involves the integral $\int \mathcal{V}(s,t)\mathcal{P}(t)dt$. The most common schemes for numerical integration approximate this kind of integrals by a sum of discrete values of the form:

$$\int \mathcal{F}(s)ds = \sum_{i=1}^{n} w_i\mathcal{F}(s_i) \tag{2.6}$$

The direct discretization of (2.3) without applying any numerical integration scheme is a fairly crude special case, which yields the less accurate approximation given in Section 2.2. For the time being, we restrict our attention to linear quadrature schemes of the form (2.6), which applied the left-hand side of (2.3) yield:

$$\int \mathcal{V}(s,t)\mathcal{P}(t)dt = \sum_{i=1}^{n} w_i \cdot \mathcal{V}.(s,t_i)\mathcal{P}(t_i) \tag{2.7}$$

Hence:

$$\sum_{i=1}^{n} w_i \cdot \mathcal{V}(s,t_i)\mathcal{P}(t_i) = \rho\mathcal{P}(s). \tag{2.8}$$

And now this equation can be discretized in s, yielding:

$$\sum_{i=1}^{n}\sum_{j=1}^{m} w_i \cdot w_j \cdot \mathcal{V}(s_j,t_i)\mathcal{P}(t_i) = \sum_{j=1}^{m} w_j \cdot \rho\mathcal{P}(s_j). \tag{2.9}$$

In the simplest case, the weights[5] w_i and w_j are given by trapezoidal rules. However, it should be pointed out that (2.9) defines a quite general scheme, as the grids in which s and t are discretized do not need to be equispaced. There are basically three aspects of the approximation that can be adjusted: the quadrature points s_j; the number of quadrature points $N_{\mathcal{X}}$, and the quadrature weights w_j, attached to each function value in the sum.

A simple example is the trapezoidal rule, in which the interval of integration is divided into $N_{\mathcal{X}} - 1$ equal intervals, each of width h. The s_j are the boundaries of the interval with s_1 and s_n denoting the lower and upper limits of integration, respectively. The approximation is:

$$\int \mathcal{F}(s)ds \approx h\left[\frac{\mathcal{F}(s_1)}{2} + \sum_{j=2}^{N_{\mathcal{X}}-1}\mathcal{F}(s_j) + \frac{\mathcal{F}(s_{N_{\mathcal{X}}})}{2}\right].$$

Note that the weights w_j are $h/2, h, ..., h, h/2$ and that accuracy is simply controlled by the choice of $N_{\mathcal{X}}$. The trapezoidal rule has some important advantages: the original raw data are often collected for equally spaced argument values, the weights are trivial, and although the accuracy of the method is modest relative to other more sophisticated schemes, it is often sufficient for the objectives at hand. Another alternative is the Simpson's rule, given by:

$$\int \mathcal{F}(s)ds \approx \frac{h}{3}\sum_{j=1}^{N_{\mathcal{X}}/2}\Big[\mathcal{F}(s_{2j-2}) + 4\mathcal{F}(s_{2j-1}) + \mathcal{F}(s_{2j})\Big],$$

which is based on a quadratic interpolation between the quadrature points, rather than a linear one as in the trapezoidal rule, and is exact for polynomials up to and including degree 3 —see Press et al. (1992), page 126.

A very useful method is provided by the trapezoidal rule for unequal spacing:

$$\int \mathcal{F}(s)ds \approx \frac{1}{2}\left\{(s_1 - s_0)\,\mathcal{F}(s_0) + \sum_{i=2}^{N_{\mathcal{X}}}(s_i - s_{i-2})\,\mathcal{F}(s_{i-1}) + (s_{N_{\mathcal{X}}} - s_{N_{\mathcal{X}}-1})\,\mathcal{F}(s_{N_{\mathcal{X}}})\right\},$$

which can be applied to non-uniform grids, constituting an interesting extension of the present work.

Applying any of the previous quadrature schemes to the operator $\tilde{\mathcal{V}}$ in equation (2.4), yields the discrete approximation:

[5] In the practice all the numerical factors in (2.9) can be embedded into the eigenvalue ρ.

$$\tilde{\mathcal{V}}(\mathcal{P}) \approx \mathbf{VWp},$$

where $\mathbf{V}$ contains the covariance function evaluated at the quadrature points $\mathcal{V}(s_i, s_j)$; $\mathbf{p}$ contains the values $\mathcal{P}(s_i)$; and $\mathbf{W}$ is a diagonal matrix containing the quadrature weights w_i. The approximately equivalent matrix eigenanalysis problem is then $\mathbf{VWp} = \rho\mathbf{p}$.

However, most quadrature schemes use positive weights. Therefore, we can rewrite the approximate eigenequation $\mathbf{VWp} = \rho\mathbf{p}$ in a more standard way, analogous to the computations carried out for the simplest case outlined in Section 2.2 —when we are not using numerical integration techniques and use the SVD to find the PCs:

$$\mathbf{W}^{1/2}\mathbf{VW}^{1/2}\mathbf{u} = \rho\mathbf{u},$$

where $\mathbf{u} := \mathbf{W}^{1/2}\mathbf{p}$ and $\mathbf{u}^T\mathbf{u} = 1$. Then we proceed as follows:

1. Fix n, the weights w_i, and the grid values s_i.
2. Find the eigenvalues ρ_m and the eigenvectors $\mathbf{u}_m$ of $\mathbf{W}^{1/2}\mathbf{VW}^{1/2}$.
3. Compute $\mathbf{p}_m = \mathbf{W}^{1/2}\mathbf{u}_m$.
4. If required, apply interpolation schemes to transform each vector $\mathbf{p}_m$ into a function $\mathcal{P}_m$.

3

The functional linear model with functional response

Consider the context of regression models with functional covariate and functional response:

$$\mathcal{Y} = \mathcal{M}(\mathcal{X}) + \mathcal{E}. \tag{3.1}$$

Given the random variables $\mathcal{X}, \mathcal{Y}$ in the Hilbert spaces $\mathbb{H}_1$ and $\mathbb{H}_2$, respectively (that is, random functions), the regression function $\mathcal{M} : \mathbb{H}_1 \to \mathbb{H}_2$ is defined as $\mathcal{M}(\mathcal{X}) = \mathbb{E}[\mathcal{Y}|\mathcal{X} = x]$. We consider $\mathbb{H}_1 = L^2[a_s, b_s]$ and $\mathbb{H}_2 = L^2[a_t, b_t]$, so that (3.1) becomes:

$$\mathcal{Y}(t) = \mathcal{M}(\mathcal{X}(s))(t) + \mathcal{E}(t),\ s \in [a_s, b_s],\ t \in [a_t, b_t]. \tag{3.2}$$

The simplest and best-known parametric model of the form (3.2) is the FLMFR, where $\mathcal{M}(\mathcal{X}) = \mathcal{M}_{\mathcal{B}}(\mathcal{X}) = \mathbb{E}[\mathcal{Y}|\mathcal{X} = x]$, with:

$$\mathcal{M}_{\mathcal{B}}(\mathcal{X}(s)) = \langle\langle \mathcal{X}(s), \mathcal{B}(s,t)\rangle\rangle = \int_{a_s}^{b_s} \mathcal{X}(s)\mathcal{B}(s,t)ds,\ s \in [a_s, b_s],\ t \in [a_t, b_t],$$

where $\mathcal{M}_{\mathcal{B}}$ is a linear mapping between $\mathbb{H}_1$ and $\mathbb{H}_2$, that is, $\mathcal{M}_{\mathcal{B}}(a\mathcal{X} + b\mathcal{Y}) = a\mathcal{M}_{\mathcal{B}}(\mathcal{X}) + b\mathcal{M}_{\mathcal{B}}(\mathcal{Y})$. The product $\langle\langle \mathcal{X}, \mathcal{B}\rangle\rangle$ is defined by:

$$(\mathcal{X}, \mathcal{B}) \in \mathbb{H}_1 \times (\mathbb{H}_1 \otimes \mathbb{H}_2) \mapsto \langle\langle \mathcal{X}, \mathcal{B}\rangle\rangle = \langle \mathcal{X}(\cdot), \mathcal{B}(\cdot, \star)\rangle \in \mathbb{H}_2.$$

Therefore, $\mathcal{M}_{\mathcal{B}}$ belongs to the class of linear and bounded[1] operators between Hilbert spaces, *i.e.*, $\mathcal{M}_{\mathcal{B}} \in \mathfrak{B}$, with:

$$\mathfrak{B} := \{\mathcal{X}(s) \in \mathbb{H}_1 \mapsto \mathcal{Y}(t) = \langle\langle \mathcal{X}(s), \mathcal{B}(s,t)\rangle\rangle \in \mathbb{H}_2 : \mathcal{B}(s,t) \in \mathbb{H}_1 \otimes \mathbb{H}_2\},\ s \in [a_s, b_s],\ t \in [a_t, b_t]. \tag{3.3}$$

This representation is motivated by the fact that we consider that $\mathcal{M}_{\mathcal{B}}$ is a bounded linear function, and therefore the Riesz representation theorem is satisfied, providing the ground for our description of

[1] This is a regularity condition imposed so that $\mathcal{M}_{\mathcal{B}}$ satisfies the Riesz representation theorem, crucial for its representation as $\langle\langle \mathcal{X}, \mathcal{B}\rangle\rangle$.

$\mathcal{M}_{\mathcal{B}}$ by $\langle\langle \mathcal{X}, \mathcal{B} \rangle\rangle$. Therefore, we assume the representation of the class $\mathfrak{B}$ for every bounded linear function between $\mathbb{H}_1$ and $\mathbb{H}_2$, thus resulting (3.2) into:

$$\mathcal{Y}(t) = \mathcal{M}_{\mathcal{B}}(\mathcal{X}(s))(t) + \mathcal{E}(t) = \int_{a_s}^{b_s} \mathcal{X}(s)\mathcal{B}(s,t)ds + \mathcal{E}(t),\ s \in [a_s, b_s],\ t \in [a_t, b_t]. \tag{3.4}$$

We also assume, without loss of generality, that our variables are already centered, that is, $\mathbb{E}[\mathcal{Y}] = \mathbb{E}[\mathcal{X}] = 0$, in order to avoid an intercept in the model[2]. The dependencies on s and t will be omitted from now on for economy, unless otherwise stated in cases where it can lead to confusion. As the different functions in (3.4) belong to a Hilbert space, they can be represented in a certain basis. In this work, we will use the only orthonormal bases in order to simplify the treatment[3], given by $\{\Psi_i\}_{i=1}^{\infty}$ and $\{\Phi_j\}_{j=1}^{\infty}$ for representing $\mathcal{X}$ and $\mathcal{Y}$ respectively, in such a way that $\mathcal{X} = \sum_{i=1}^{\infty} x_i \Psi_i$ and $\mathcal{Y} = \sum_{j=1}^{\infty} y_j \Phi_j$.

Furthermore[4], $\mathcal{B} = \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} b_{ij}(\Psi_i \otimes \Phi_j)$. The tensor product $\otimes$ is defined by $\langle\langle \mathcal{F}, \Psi_i \otimes \Phi_j \rangle\rangle = \langle \mathcal{F}, \Psi_i \rangle \Phi_j$, being $\mathcal{F}$ a function in $\mathbb{H}_1$. Therefore:

$$\langle\langle \mathcal{X}, \mathcal{B} \rangle\rangle = \left\langle \left\langle \sum_{k=1}^{\infty} x_k \Psi_k, \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} b_{ij}\Psi_i \otimes \Phi_j \right\rangle \right\rangle = \sum_{i=1}^{\infty}\sum_{j=1}^{\infty}\sum_{k=1}^{\infty} b_{ij}x_k \langle\langle \Psi_k, \Psi_i \otimes \Phi_j \rangle\rangle . \tag{3.5}$$

Bearing in mind the definition of the tensor product and that the bases of principal components are orthogonal and have been normalized, *i.e.*, they are orthogonal and thus $\langle \Psi_i, \Psi_k \rangle = \delta_{ik}$, we have that $\langle\langle \Psi_k, \Psi_i \otimes \Phi_j \rangle\rangle = \langle \Psi_k, \Psi_i \rangle \Phi_j = \delta_{ik}\Phi_j$. Using this, equation (3.5) becomes:

$$\langle\langle \mathcal{X}, \mathcal{B} \rangle\rangle = \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} b_{ij}x_i\Phi_j = \sum_{j=1}^{\infty}\left(\sum_{i=1}^{\infty} b_{ij}x_i\right)\Phi_j. \tag{3.6}$$

Comparing this with the basis expansion of $\mathcal{Y}$, *i.e.*, $\mathcal{Y} = \sum_{j=1}^{\infty} y_j\Phi_j$, we arrive to the following multivariate model:

$$y_j = \sum_{i=1}^{\infty} b_{ij}x_i,\ \forall j = 1, ..., \infty. \tag{3.7}$$

In matrix notation:

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \end{pmatrix} = \begin{pmatrix} b_{11} & b_{12} & \cdots \\ b_{21} & b_{22} & \cdots \\ \vdots & \vdots & \ddots \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \end{pmatrix}. \tag{3.8}$$

In order to solve this in the practice, a finite number of basis elements must be chosen, that don't need to be the same for both bases. We will denote by $p_{\mathcal{X}}$ the dimension of the truncated basis in the

[2] If they are not centered, it suffices to apply this model to the centered variables. It is easy to prove that this is equivalent to include an intercept in the regression of the non-centered variables.

[3] In particular, two different alternatives for representing $\mathcal{X}$ and $\mathcal{Y}$ will be considered: Fourier basis expansions and FPCs.

[4] If $\mathbb{H}_1$ and $\mathbb{H}_2$ have orthonormal bases $\{\Phi_k\}$ and $\{\Psi_l\}$, respectively, then $\{\Phi_k \otimes \Psi_l\}$ is an orthonormal basis for $\mathbb{H}_1 \otimes \mathbb{H}_2$. In particular, the dimension of the Hilbert tensor product is the product (as cardinal numbers) of the Hilbert dimensions.

functional covariate space and by $p_{\mathcal{Y}}$ the size of the truncated basis in the response space, that is, $\{\Psi_i\}_{i=1}^{p_{\mathcal{X}}}$ and $\{\Phi_j\}_{j=1}^{p_{\mathcal{Y}}}$. With this, (3.4) transforms into:

$$\mathcal{Y}^{(p_{\mathcal{Y}})} = \langle\langle \mathcal{X}^{(p_{\mathcal{X}})}, \mathcal{B}^{(p_{\mathcal{X}},p_{\mathcal{Y}})}\rangle\rangle + \mathcal{E}^{(p_{\mathcal{Y}})} = \int_{a_s}^{b_s} \mathcal{X}^{(p_{\mathcal{X}})}\mathcal{B}^{(p_{\mathcal{X}},p_{\mathcal{Y}})}ds + \mathcal{E}^{(p_{\mathcal{Y}})},\ s \in [a_s, b_s],\ t \in [a_t, b_t], \tag{3.9}$$

and (3.7) into the multivariate model:

$$y_j = \sum_{i=1}^{p_{\mathcal{X}}} b_{ij}x_i,\ \forall j = 1, ..., p_{\mathcal{Y}} \tag{3.10}$$

This is:

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_{p_{\mathcal{Y}}} \end{pmatrix} = \begin{pmatrix} b_{11} & b_{12} & \cdots & b_{1p_{\mathcal{Y}}} \\ b_{21} & b_{22} & \cdots & b_{2p_{\mathcal{Y}}} \\ \vdots & \vdots & \ddots & \vdots \\ b_{p_{\mathcal{X}}1} & b_{p_{\mathcal{X}}2} & \cdots & b_{p_{\mathcal{X}}p_{\mathcal{Y}}} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_{p_{\mathcal{X}}} \end{pmatrix}. \tag{3.11}$$

Furthermore, note that the Functional Linear Model with Scalar Response (FLMSR):

$$Y = \langle\langle \mathcal{X}(s), \mathcal{B}(s)\rangle\rangle + \varepsilon = \langle \mathcal{X}(s), \mathcal{B}(s)\rangle + \varepsilon = \int_{a_s}^{b_s} \mathcal{X}(s)\mathcal{B}(s)ds + \varepsilon,\ s \in [a_s, b_s], \tag{3.12}$$

arises as a particular case of the previous model by setting $k_{\mathcal{Y}} = 1$. Now, there is only one basis element Φ_j, which is equal to 1 and, therefore, the expansion coefficients are directly given by the data in Y and the multivariate model (3.10) transforms into an univariate model.

3.1 Estimation of the model

Given a sample $(\mathcal{X}_1, \mathcal{Y}_1), ..., (\mathcal{X}_n, \mathcal{Y}_n)$:

$$\mathcal{X} = \begin{pmatrix} \mathcal{X}_1(s_1) & \cdots & \mathcal{X}_1(s_{N_{\mathcal{X}}}) \\ \vdots & \ddots & \vdots \\ \mathcal{X}_N(s_1) & \cdots & \mathcal{X}_N(s_{N_{\mathcal{X}}}) \end{pmatrix} \quad \text{and} \quad \mathcal{Y} = \begin{pmatrix} \mathcal{Y}_1(t_1) & \cdots & \mathcal{Y}_1(t_{n_{\mathcal{Y}}}) \\ \vdots & \ddots & \vdots \\ \mathcal{Y}_N(t_1) & \cdots & \mathcal{Y}_N(t_{n_{\mathcal{Y}}}) \end{pmatrix}. \tag{3.13}$$

the estimation of the functional parameter can be done by minimizing the Residual Sum of Squares (RSS):

$$\hat{\mathcal{B}} = \arg\min_{\mathcal{B}\in\mathbb{H}_1\otimes\mathbb{H}_2} \sum_{i=1}^{n} \left(\mathcal{Y}_i - \mathcal{M}_{\mathcal{B}}(\mathcal{X}_i)\right)^2.$$

A possible method to search for the parameter that minimizes the RSS is representing the functional data and the functional parameter in the truncated functional bases $\{\Psi_i\}_{i=1}^{p_{\mathcal{X}}}$ and $\{\Phi_j\}_{j=1}^{p_{\mathcal{Y}}}$, respectively:

$$\mathcal{X}_i = \sum_{i=1}^{p_{\mathcal{X}}} x_{ij}\Psi_i, \qquad \mathcal{Y}_i = \sum_{J=1}^{p_{\mathcal{Y}}} y_{ij}\Phi_i$$

Using the orthogonal projection matrix $\mathbf{P}_\Omega = \left(\mathbf{X}_{p\mathcal{X}}^T \mathbf{X}_{p\mathcal{X}}\right)^{-1} \mathbf{X}_{p\mathcal{X}}^T$ on the columns of $\mathbf{X}^{p\mathcal{X}}$, we have:

$$\hat{\mathbf{B}}_{p\mathcal{X}p\mathcal{Y}} = \left(\mathbf{X}_{p\mathcal{X}}^T \mathbf{X}_{p\mathcal{X}}\right)^{-1} \mathbf{X}_{p\mathcal{X}}^T \mathbf{Y}_{p\mathcal{Y}} \tag{3.14}$$

and

$$\hat{\mathbf{Y}}_{p\mathcal{Y}} = \mathbf{P}_\Omega \mathbf{Y}_{p\mathcal{Y}} = \mathbf{X}_{p\mathcal{X}} \left(\mathbf{X}_{p\mathcal{X}}^T \mathbf{X}_{p\mathcal{X}}\right)^{-1} \mathbf{X}_{p\mathcal{X}}^T \mathbf{Y}_{p\mathcal{Y}} = \mathbf{X}_{p\mathcal{X}} \hat{\mathbf{B}}_{p\mathcal{X}p\mathcal{Y}}, \tag{3.15}$$

Note that we might well have more discretization points than observations. In such a case, it is enforced to use a Moore–Penrose pseudoinverse to compute $\left(\mathbf{X}_{p\mathcal{X}}^T \mathbf{X}_{p\mathcal{X}}\right)^{-1}$ (see the developed code of the work in Appendix A).. Then, the estimation problem can be the solved as a general multivariate regression model *e.g.* with least squares.

3.2 Numerical examples

An illustrative example of the methodology introduced so far is presented for clarity in this section.

A simulation of several processes is performed, and the estimation of the $\mathcal{B}$-surface is done with both Fourier basis expansions and functional principal component analysis. Figure 3.1 shows a simulation of $n = 250$ realizations, in which $\mathcal{X}$ is a Brownian motion, with $\mathcal{X} \in L^2[0, 2]$

$$\mathcal{B}(s,t) = \frac{5}{2} \frac{\cos(2\pi t s)}{1 + (s - 0.5)^2}, \tag{3.16}$$

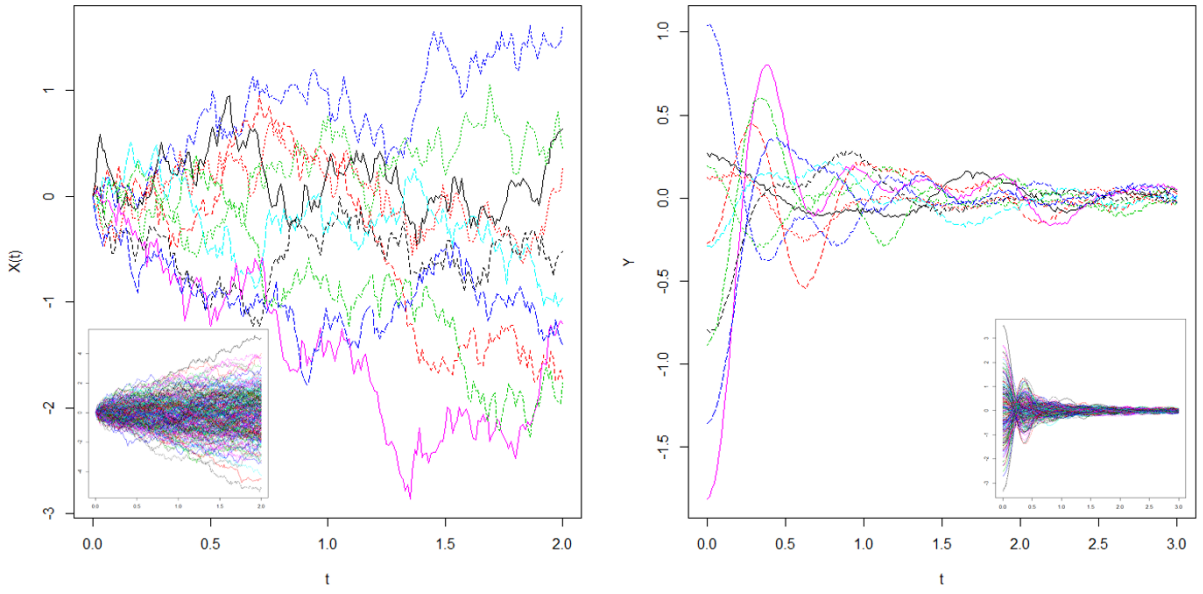

Figure 3.1: Simulation of a functional linear model ($n = 250$) under the null hypothesis, being $\mathcal{X}$ a Brownian motion, $\mathcal{B}(s,t)$ given by (3.16), $\mathcal{E}$ a white noise process, *i.e.* $\mathcal{E}(t) \sim N(0, \sigma^2)$ and $\mathcal{Y} = \langle\langle \mathcal{X}, \mathcal{B} \rangle\rangle + \mathcal{E}$. The grids are equispaced for both $\mathcal{X}$ and $\mathcal{Y}$, with 201 points on each. Only the first realizations are plotted for clearness, the complete simulations are shown in the miniatures.

$\mathcal{E}$ is a white noise process, *i.e.* $\mathcal{E}(t) \sim N(0, \sigma^2)$, independently with $\sigma = 0.1$, and $\mathcal{Y} = \langle\langle \mathcal{X}, \mathcal{B}\rangle\rangle + \mathcal{E}$, with $\mathcal{Y} \in L^2[0, 3]$. The grids are equispaced for both $\mathcal{X}$ and $\mathcal{Y}$, with 201 points on each. The processes are shown in Figure 3.1. For the sake of clearness, we only show the first 10 observations. The whole processes are shown in miniatures.

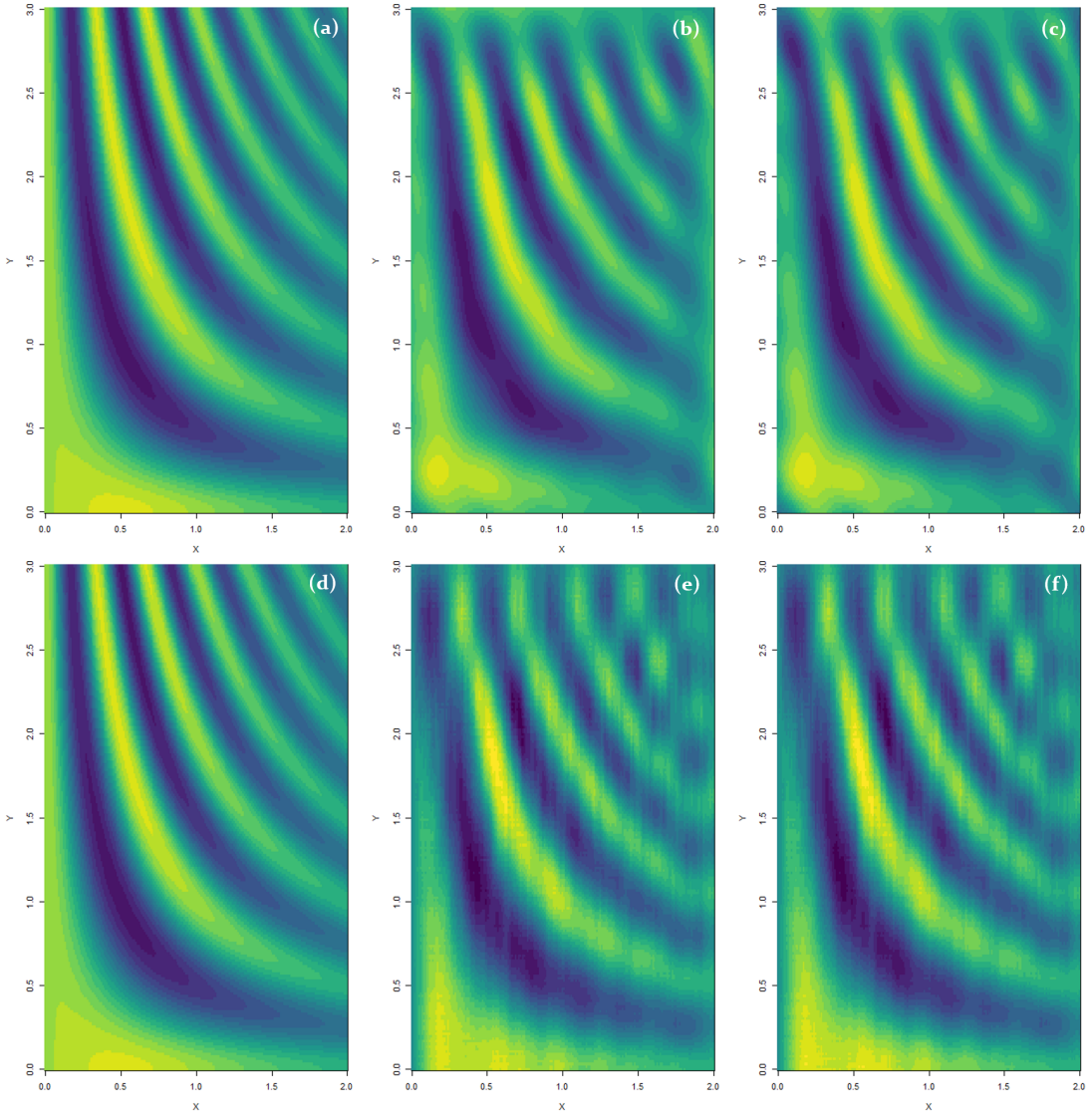

Figure 3.2: Estimation of the $\mathcal{B}$-surface using Fourier basis expansions and FPCs, with 11 basis elements for both $\mathcal{X}$ and $\mathcal{Y}$ on each. The theoretical surfaces are plotted in **(a)** and **(d)** for comparison. **(b)** and **(e)** show this surface projected onto the Fourier and FPCs bases, respectively. **(c)** and **(f)** show the estimations on such bases.

The estimations of the surface are shown in Figure 3.2 for Fourier basis expansions (upper row) and FPCA (lower row), with 11 basis elements for both $\mathcal{X}$ and $\mathcal{Y}$. The theoretical surface is plotted in Figures 3.2(a) and 3.2(d) for comparison. Figures 3.2(b) and 3.2(e) show this surface projected onto the Fourier and FPCA bases, thus $\mathcal{B}$ represented on the basis used for the estimation. This gives the best

representation we can achieve with these truncated bases (the representation in the infinite basis is perfect). The typical border effects due to the use of a Fourier representation are present already here. Figures 3.2(c) and 3.2(f) show the estimations for $\hat{\mathcal{B}}$. The explained variances with FPCA have been computed from the eigenvalues resulting from the SVD decomposition, and are 60% and 73% respectively. This suggests to increase the number of FPCs for the representation of $\mathcal{X}$ if a better estimation of $\mathcal{B}$ is needed.

4 The goodness-of-fit test

The goodness-of-fit test is presented in this chapter, which is divided into three sections. The first deals with the random projections paradigm; the second establishes the theoretical fundamentals of the test; the third section outlines the implementation of the test statistic, by means of some geometrical and matrix arguments. The last one details the bootstrap resampling procedure on the residuals of the estimation for the calibration of the test statistic.

4.1 Random projections

Random projections are very suitable when dealing with high-dimensional data, since they offer an alternative to overcome the curse of the dimensionality. The main idea behind is to reduce the dimension and characterize the distribution of the multidimensional data by the distribution of randomly projected data.

In the goodness-of-fit field, this is specially useful, since the higher the model dimension, the less efficient and powerful the test strategies become. This technique has already been used to develop a goodness-of-fit test for multivariate regression models based on random projections in Escanciano (2006) and has been generalized to the FLMSR by García-Portugués et al. (2014). The procedure considered there is quite common within FDA: instead of testing a given null hypothesis in the functional space, the transformation of this hypothesis on a one-dimensional randomly chosen projection is tested, by using random projections arising from considering the inner product of the functional variables $\mathcal{X}$ and $\mathcal{Y}$ with a suitable family of random directions in $\mathbb{H}_1$ and $\mathbb{H}_2$, respectively. This allows to benefit from the numerous procedures that are available in the one-dimensional case.

A very interesting result on projections was provided by Patilea et al. (2012). Here the authors state a characterization of the conditional expectation of a scalar variable Y with respect to a functional variable $\mathcal{X}$ given in terms of the conditional expectation of Y with respect to the projected $\mathcal{X}$. The result is provided in the following lemma:

Lemma 4.1 Patilea et al. (2012). Let Y be a random variable and $\mathcal{X}$ a functional random variable in the functional space $\mathbb{H}$. The following statements are equivalent:

I $\mathbb{E}[Y|\mathcal{X}=x]=0$, for almost every (a.e.) $x\in\mathbb{H}$.

II $\mathbb{E}[Y|\,\langle\mathcal{X},\mathcal{D}\rangle=u]=0$, for a.e. $u\in\mathbb{R}$ and $\forall\mathcal{D}\in\mathbb{S}_{\mathbb{H}}$.

III $\mathbb{E}[Y|\,\langle\mathcal{X},\mathcal{D}\rangle=u]=0$, for a.e. $u\in\mathbb{R}$ and $\forall\mathcal{D}\in\mathbb{S}^p_{\mathbb{H}}$, $\forall p\geq 1$.

4.2 Theoretical arguments.

Let $\mathcal{X}$ and $\mathcal{Y}$ be functional random variables in the Hilbert space $\mathbb{H}$ and consider the context of regression models with functional covariate and functional response:

$$\mathcal{Y}=\mathcal{M}(\mathcal{X})+\mathcal{E}.$$

Much of the existing literature is concerned with parametric modeling, where m is assumed to belong to a given parametric family, this is $\mathcal{M}\in\mathcal{M}_\theta=\{m_\theta:\theta\in\Theta\}$, for a certain parameter θ. Therefore, one considers the test of this parametric regression model, in which the null hypothesis is given by $H_0: m\in\mathcal{M}_\theta=\{m_\theta:\theta\in\Theta\}$, against a general alternative $H_1: m\notin\mathcal{M}_\theta$.

The parametric model we are interested in is the FLM with functional response, of the form (3.2), where $\mathcal{M}(\mathcal{X})=\mathbb{E}[\mathcal{Y}|\mathcal{X}]=\mathcal{M}_{\mathcal{B}}(\mathcal{X})$, which is equivalent to say that the regression function of $\mathcal{Y}$ on $\mathcal{X}$, $\mathcal{M}$, belongs to the family $\mathfrak{B}$ given in (3.3). This is, given a random sample $\{(\mathcal{X}_i,\mathcal{Y}_i)\}_{i=1}^n$, we are interested in checking if a functional linear model is suitable to explain the relation between the functional covariate and the functional response, *i.e.*, test for the composite hypothesis:

$$H_0:\mathcal{M}\in\mathfrak{B}\qquad\text{or}\qquad H_0:\mathcal{M}=\mathcal{M}_{\mathcal{B}}\text{ for some }\mathcal{B}\in\mathbb{H}_1\otimes\mathbb{H}_2,\tag{4.1}$$

versus a general alternative of the form $H_1:\mathcal{M}\notin\mathfrak{B}$, or $H_1:\mathcal{M}\neq\mathcal{M}_{\mathcal{B}}$ for some $\mathcal{B}\in\mathbb{H}_1\otimes\mathbb{H}_2$.

The pillar of the goodness-of-fit tests we present is the a.s. characterization of the null hypothesis, re-expressed as $H_0:\mathbb{E}[\mathcal{Y}-\mathcal{M}_{\mathcal{B}}(\mathcal{X})|\mathcal{X}]=0$ for some $\mathcal{B}\in\mathbb{H}_1\otimes\mathbb{H}_2$, by means of the associated projected hypothesis on $\mathcal{D}_{\mathcal{X}}\in\mathbb{H}_1$ and $\mathcal{D}_{\mathcal{Y}}\in\mathbb{H}_2$, defined as $H_0^{\mathcal{D}_{\mathcal{X}},\mathcal{D}_{\mathcal{Y}}}:\mathbb{E}[\langle\mathcal{Y}-\mathcal{M}_{\mathcal{B}}(\mathcal{X}),\mathcal{D}_{\mathcal{Y}}\rangle\,|\,\langle\mathcal{X},\mathcal{D}_{\mathcal{X}}\rangle]=0$. In the following, we identify $\mathcal{Y}-\mathcal{M}_{\mathcal{B}}(\mathcal{X})$ by $\mathcal{Y}$ for the sake of simplicity in notation.

The key point to test the null hypothesis H_0 is the following conjecture, which gives the characterization of H_0 in terms of the random projections of $\mathcal{X}$ and $\mathcal{Y}$.

Conjecture 4.3. Let $\mathcal{B}$ be an element of $\mathbb{H}_1\otimes\mathbb{H}_2$. The following statements are equivalent:

I $\mathcal{M}(\mathcal{X})=\langle\langle\mathcal{X},\mathcal{B}\rangle\rangle,\forall\mathcal{X}\in\mathbb{H}_1$.

II $\mathbb{E}[\mathcal{Y}-\langle\langle\mathcal{X},\mathcal{B}\rangle\rangle|\mathcal{X}=x]=0$, for a.e $\mathcal{X}\in\mathbb{H}_1$.

III $\mathbb{E}[\mathcal{Y}-\langle\langle\mathcal{X},\mathcal{B}\rangle\rangle|\,\langle\mathcal{X},\mathcal{D}_{\mathcal{X}}\rangle=u]=0$, for a.e. $u\in\mathbb{R}$ and $\forall\mathcal{D}_{\mathcal{X}}\in\mathbb{S}_{\mathbb{H}_1}$.

IV $\mathbb{E}[\mathcal{Y}-\langle\langle\mathcal{X},\mathcal{B}\rangle\rangle|\,\langle\mathcal{X},\mathcal{D}_{\mathcal{X}}\rangle=u]=0$, for a.e. $u\in\mathbb{R}$ and $\forall\mathcal{D}_{\mathcal{X}}\in\mathbb{S}^{p_{\mathcal{X}}}_{\mathbb{H}_1}$, $\forall p_{\mathcal{X}}\geq 1$.

V $\mathbb{E}[\langle\mathcal{Y}-\langle\langle\mathcal{X},\mathcal{B}\rangle\rangle,\mathcal{D}_{\mathcal{Y}}\rangle\,|\,\langle\mathcal{X},\mathcal{D}_{\mathcal{X}}\rangle=u]=0$, for a.e. $u\in\mathbb{R}$ and $\forall\mathcal{D}_{\mathcal{X}}\in\mathbb{S}_{\mathbb{H}_1}$ and $\forall\mathcal{D}_{\mathcal{Y}}\in\mathbb{S}_{\mathbb{H}_2}$.

VI $\mathbb{E}[\langle \mathcal{Y} - \langle\langle \mathcal{X}, \mathcal{B}\rangle\rangle, \mathcal{D}_{\mathcal{Y}}\rangle \,|\, \langle \mathcal{X}, \mathcal{D}_{\mathcal{X}}\rangle = u] = 0$, for a.e. $u \in \mathbb{R}$ and $\forall \mathcal{D}_{\mathcal{X}} \in \mathbb{S}_{\mathbb{H}_1}^{p_{\mathcal{X}}}$, $\forall p_{\mathcal{X}} \geq 1$, and $\forall \mathcal{D}_{\mathcal{Y}} \in \mathbb{S}_{\mathbb{H}_2}^{p_{\mathcal{Y}}}$, $\forall p_{\mathcal{Y}} \geq 1$.

Comment. Implication I $\Leftrightarrow$ II is trivial; II $\Leftrightarrow$ III $\Leftrightarrow$ IV could be done applying Lemma 4.1 to each point s; III $\Rightarrow$ V and IV $\Rightarrow$ VI are clear due to the linearity of the inner product $\langle \cdot, \cdot \rangle$ and the conditional expectation; III $\Rightarrow$ V and IV $\Rightarrow$ VI are trivial; III $\Leftarrow$ V and IV $\Leftarrow$ VI should be proved, but they would be easy, since if $\langle \mathcal{A}, \mathcal{X}\rangle = 0$ for all $\mathcal{A}$ in $\mathbb{S}_{\mathbb{H}}$ (even $\mathcal{A} = \mathcal{X}/||\mathcal{X}||$), then $\mathcal{X}$ must be 0. Obviously, the proof of Conjecture 4.3 requires more detail and rigor, since the outlined comment does not grant the validity of the statements. However, it is useful to provide a motivation of the test —not the characterization of the null hypothesis H_0, that would arise from implications V $\Rightarrow$ III and VI $\Rightarrow$ IV.

The idea is therefore to consider two kinds of projections: one for the regressor and another one for the response functions, denoted by $\mathcal{D}_{\mathcal{X}}$ and $\mathcal{D}_{\mathcal{Y}}$, respectively. Then H_0 should be characterized by the null value of the moment $\mathbb{E}[\langle \mathcal{Y} - \langle\langle \mathcal{X}, \mathcal{B}\rangle\rangle, \mathcal{D}_{\mathcal{Y}}\rangle \,|\, \langle \mathcal{X}, \mathcal{D}_{\mathcal{X}}\rangle = u] = 0$, for a.e. $u \in \mathbb{R}$ and $\mathcal{D}_{\mathcal{X}}, \mathcal{D}_{\mathcal{Y}} \in \mathbb{S}_{\mathbb{H}}$. The deviation from H_0 can be measured by the empirical process arising from the estimation of this moment:

$$R_n(\mathcal{D}_{\mathcal{X}}, \mathcal{D}_{\mathcal{Y}}, u) := \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left\langle \hat{\mathcal{E}}_i, \mathcal{D}_{\mathcal{Y}} \right\rangle \mathbb{I}_{\{\langle \mathcal{X}, \mathcal{D}_{\mathcal{X}}\rangle \leq u\}}, \tag{4.2}$$

that will be named as the Residual Marked empirical Process based on Projections (RMPP). The marks of (4.2) are given by the projected residuals $\{\langle \hat{\mathcal{E}}_i, \mathcal{D}_{\mathcal{Y}}\rangle\}_{i=1}^n = \{\langle \mathcal{Y}_i - \langle \mathcal{X}_i, \hat{\mathcal{B}}\rangle, \mathcal{D}_{\mathcal{Y}}\rangle\}_{i=1}^n$, and the jumps by the projected functional regressor in the direction $\mathcal{D}_{\mathcal{X}}$, this is, $\{\langle \mathcal{X}_i, \mathcal{D}_{\mathcal{X}}\rangle\}_{i=1}^n$. Note that the RMPP only depends on the residuals of the model and thus can be easily extended to other regression models, provided there exist estimation methods for them.

To measure the distance of the empirical process (4.2) from zero, two possibilities are the classical Cramér-von Mises (CvM) and Kolmogorov-Smirnov (KS) norms, adapted to the projected space $\Pi = \mathbb{S}_{\mathbb{H}_1} \times \mathbb{S}_{\mathbb{H}_2} \times \mathbb{R}$, yielding the Projected Cramér-von Mises (PCvM) and Projected Kolmogorov-Smirnov (PKS) norms:

$$\begin{aligned} \text{PCvM}_n &= \int_{\Pi} R_n(\mathcal{D}_{\mathcal{X}}, \mathcal{D}_{\mathcal{Y}}, u)^2 F_{n,\mathcal{D}_{\mathcal{X}}}(du) \omega_{\mathcal{X}}(d\mathcal{D}_{\mathcal{X}}) \omega_{\mathcal{Y}}(d\mathcal{D}_{\mathcal{Y}}), \\ \text{PKS}_n &= \sup_{(\mathcal{D}_{\mathcal{X}}, \mathcal{D}_{\mathcal{Y}}, u) \in \Pi} |R_n(\mathcal{D}_{\mathcal{X}}, \mathcal{D}_{\mathcal{Y}}, u)|, \end{aligned} \tag{4.3}$$

where $F_{n,\mathcal{D}_{\mathcal{X}}}$ is the Empirical Cumulative Distribution Function (ECDF) of the projected functional data in the direction $\mathcal{D}_{\mathcal{X}}$ (*i.e.* the ECDF of the data $\{\langle \mathcal{X}_i, \mathcal{D}_{\mathcal{X}}\rangle\}_{i=1}^n$) and $\omega_{\mathcal{X}}$ and $\omega_{\mathcal{Y}}$ represent suitable measures on $\mathbb{S}_{\mathbb{H}_1}$ and $\mathbb{S}_{\mathbb{H}_2}$, respectively. Unfortunately, the infinite dimensions of the spaces $\mathbb{S}_{\mathbb{H}_1}$ and $\mathbb{S}_{\mathbb{H}_2}$ make infeasible to compute the functionals (4.3) and some kind of discretization is needed. A solution to this problem is to consider the properties of the Hilbert spaces and basis representations.

Let us introduce some required notation. Let $\{\Psi_i\}_{i=1}^{\infty}$ and $\{\Phi_j\}_{j=1}^{\infty}$ be bases of $\mathbb{H}_1$ and $\mathbb{H}_2$, respectively and consider the $p_{\mathcal{X}}$-truncated and $p_{\mathcal{Y}}$-truncated bases $\{\Psi_i\}_{i=1}^{p_{\mathcal{X}}}$ and $\{\Phi_j\}_{j=1}^{p_{\mathcal{Y}}}$, with matrix of inner products $\mathbf{\Psi}$ and $\mathbf{\Phi}$, respectively. We denote by $\mathcal{X}_i^{(p_{\mathcal{X}})}$ and $\mathcal{D}_{\mathcal{X}}^{(p_{\mathcal{X}})}$ to the representation of the functions $\mathcal{X}_i$ and $\mathcal{D}_{\mathcal{X}}$ in the $p_{\mathcal{X}}$-truncated basis, with matrices of coefficients $\mathbf{X}_{i,p_{\mathcal{X}}}$ and $\mathbf{d}_{p_{\mathcal{X}}}$, respectively, and for $i = 1, ..., n$.

Analogously, we use the notations $\mathcal{Y}_i^{(p_\mathcal{Y})}$, $\mathcal{D}_\mathcal{Y}^{(p_\mathcal{Y})}$, $\mathbf{Y}_{i,p_\mathcal{Y}}$ and $\mathbf{d}_{p_\mathcal{Y}}$, for $i = 1, ..., n$. Since $\{\Psi_i\}_{i=1}^\infty$ and $\{\Phi_j\}_{j=1}^\infty$ are arbitrary bases (and may not be orthonormal), we have that:

$$\left\langle \mathcal{X}_i^{(p_\mathcal{X})}, \mathcal{D}_\mathcal{X}^{(p_\mathcal{X})} \right\rangle = \mathbf{X}_{i,p_\mathcal{X}}^T \mathbf{\Psi} \mathbf{d}_{\mathcal{X},p_\mathcal{X}},$$

and

$$\left\langle \hat{\mathcal{E}}_i^{(p_\mathcal{Y})}, \mathcal{D}_\mathcal{Y}^{(p_\mathcal{Y})} \right\rangle = \hat{\mathbf{E}}_{i,p_\mathcal{Y}}^T \mathbf{\Phi} \mathbf{d}_{\mathcal{Y},p_\mathcal{Y}},$$

where $\hat{\mathbf{E}}_{i,p_\mathcal{Y}}^T$ is the matrix of coefficients of $\hat{\mathcal{E}}_i^{(p_\mathcal{Y})}$ in the $p_\mathcal{Y}$-truncated basis $\{\Phi_j\}_{j=1}^{p_\mathcal{Y}}$.

By analogy with the previously defined $F_{n,\mathcal{D}_\mathcal{X}}$, we denote by $F_{n,\mathcal{D}_\mathcal{X}^{(p_\mathcal{X})}}$ the ECDF of the projected functional data expressed in the $p_\mathcal{X}$-truncated basis, both for the projector and for the functional data. Then, the RMPP can be expressed in terms of a $p_\mathcal{X}$ and a $p_\mathcal{Y}$-truncated basis, yielding:

$$R_{n,p_\mathcal{X},p_\mathcal{Y}}\left(\mathcal{D}_\mathcal{X}^{(p_\mathcal{X})}, \mathcal{D}_\mathcal{Y}^{(p_\mathcal{Y})}, u\right) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \left\langle \hat{\mathcal{E}}_i^{(p_\mathcal{Y})}, \mathcal{D}_\mathcal{Y}^{(p_\mathcal{Y})} \right\rangle \mathbb{I}_{\{\langle \mathcal{X}, \mathcal{D}_\mathcal{X} \rangle \leq u\}},$$

which, applying basis expansions, transforms into:

$$R_{n,p_\mathcal{X},p_\mathcal{Y}}(\mathbf{d}_{\mathcal{X},p_\mathcal{X}}, \mathbf{d}_{\mathcal{Y},p_\mathcal{Y}}, u) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \left(\hat{\mathbf{E}}_{i,p_\mathcal{Y}}^T \mathbf{\Phi} \mathbf{d}_{\mathcal{Y},p_\mathcal{Y}} \mathbb{I}_{\left\{\mathbf{X}_{i,p_\mathcal{X}}^T \mathbf{\Psi} \mathbf{d}_{\mathcal{X},p_\mathcal{X}} \leq u\right\}} \right),$$

where $\hat{\mathbf{B}}_{p_\mathcal{X} p_\mathcal{Y}}$ represents the coefficients of $\hat{\mathcal{B}}$ in the $(p_\mathcal{X} \times p_\mathcal{Y})$-truncated basis $\{\Psi_i \otimes \Phi_j\}$, $i = 1, ...\ p_\mathcal{X}$, $j = 1, ...,\ p_\mathcal{Y}$.

Bearing in mind this, our test statistic proposal is a modified version of the PCvM statistic in (4.3) that results from expressing all the functions in truncated bases of $\mathbb{H}_1$ and $\mathbb{H}_2$:

$$\text{PCvM}_{n,p_\mathcal{X} p_\mathcal{Y}} = \int_{\mathbb{S}_{\mathbb{H}_1}^{p_\mathcal{X}} \times \mathbb{S}_{\mathbb{H}_2}^{p_\mathcal{Y}} \times \mathbb{R}} R_{n,p_\mathcal{X},p_\mathcal{Y}}\left(\mathcal{D}_\mathcal{X}^{(p_\mathcal{X})}, \mathcal{D}_\mathcal{Y}^{(p_\mathcal{Y})}, u\right)^2 F_{n,\mathcal{D}_\mathcal{X}^{(p_\mathcal{X})}}(du) \omega_\mathcal{X}(d\mathcal{D}_\mathcal{X}^{(p_\mathcal{X})}) \omega_\mathcal{Y}(d\mathcal{D}_\mathcal{Y}^{(p_\mathcal{Y})}). \tag{4.4}$$

The PCvM statistic presents important computational advantages and can be adapted to the given framework of Escanciano (2006) for the finite dimensional case. The most important advantage of this statistic is that we can derive an explicit expression where there is no need to compute the RMPP for different projections[1].

Using that the integration in the p-sphere of $\mathbb{H}$ can be expressed as the integration in the p-sphere of

[1] This property that does not hold for the KS statistic.

$\mathbb{R}^p$ via the transformations defined in Section 2.1, we have:

$$\begin{aligned}
\text{PCvM}_{n,p_\mathcal{X}p_\mathcal{Y}} &= \int_{\mathbb{S}^{p_\mathcal{X}}_{\mathbb{H}_1}\times\mathbb{S}^{p_\mathcal{Y}}_{\mathbb{H}_2}\times\mathbb{R}} R_{n,p_\mathcal{X},p_\mathcal{Y}}\left(\mathcal{D}^{(p_\mathcal{X})}_\mathcal{X},\mathcal{D}^{(p_\mathcal{Y})}_\mathcal{Y},u\right)^2 F_{n,\mathcal{D}^{(p_\mathcal{X})}_\mathcal{X}}(du)\omega_\mathcal{X}(d\mathcal{D}^{(p_\mathcal{X})}_\mathcal{X})\omega_\mathcal{Y}(d\mathcal{D}^{(p_\mathcal{Y})}_\mathcal{Y}) \\
&= \int_{\mathbb{S}^{p_\mathcal{X}}_{\mathbb{H}_1}\times\mathbb{S}^{p_\mathcal{Y}}_{\mathbb{H}_2}\times\mathbb{R}} R_{n,p_\mathcal{X},p_\mathcal{Y}}(\mathbf{d}_{p_\mathcal{X}},\mathbf{d}_{p_\mathcal{Y}},u)^2 F_{n,\mathbf{d}_{\mathcal{X},p_\mathcal{X}}}(du)\omega_\mathcal{X}(d\mathbf{d}_{\mathcal{X},p_\mathcal{X}})\omega_\mathcal{Y}(d\mathbf{d}_{\mathcal{Y},p_\mathcal{Y}}) \\
&= \int_{\mathbb{S}^{p_\mathcal{X}}\times\mathbb{S}^{,p_\mathcal{Y}}\times\mathbb{R}} |\mathbf{R}_\mathcal{X}|^{-1}|\mathbf{R}_\mathcal{Y}|^{-1} R_{n,p_\mathcal{X},p_\mathcal{Y}}(\mathbf{R}_\mathcal{X}^{-1}\mathbf{d}_{p_\mathcal{X}},\mathbf{R}_\mathcal{Y}^{-1}\mathbf{d}_{p_\mathcal{Y}},u)^2 \\
&\quad\cdot F_{n,\mathbf{R}_\mathcal{X}^{-1}\mathbf{d}_{\mathcal{X},p_\mathcal{X}}}(du)\omega_\mathcal{X}(d\mathbf{d}_{\mathcal{X},p_\mathcal{X}})\omega_\mathcal{Y}(d\mathbf{d}_{\mathcal{Y},p_\mathcal{Y}}) \\
&= \int_{\mathbb{S}^{p_\mathcal{X}}\times\mathbb{S}^{,p_\mathcal{Y}}\times\mathbb{R}} |\mathbf{R}_\mathcal{X}|^{-1}|\mathbf{R}_\mathcal{Y}|^{-1}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^n \hat{\mathbf{E}}^T_{i,p_\mathcal{Y}}\mathbf{d}_{\mathcal{Y},p_\mathcal{Y}}\mathbb{I}_{\left\{\mathbf{X}^T_{i,p_\mathcal{X}}\mathbf{d}_{\mathcal{X},p_\mathcal{X}}\leq u\right\}}\right)^2 \\
&\quad\cdot F_{n,\mathbf{R}_\mathcal{X}^{-1}\mathbf{d}_{\mathcal{X},p_\mathcal{X}}}(du)\omega_\mathcal{X}(d\mathbf{d}_{\mathcal{X},p_\mathcal{X}})\omega_\mathcal{Y}(d\mathbf{d}_{\mathcal{Y},p_\mathcal{Y}}),
\end{aligned} \tag{4.5}$$

where $\omega_\mathcal{X}(d\mathbf{d}_{\mathcal{X},p_\mathcal{X}})$ and $\omega_\mathcal{Y}(d\mathbf{d}_{\mathcal{Y},p_\mathcal{Y}})$ now represent measures in the finite $p_\mathcal{X}-$ and $p_\mathcal{Y}$-spheres $\mathbb{S}^p$ that, for simplicity purposes, are taken as the uniform distribution on $\mathbb{S}^p$. Essentially, what it is done is to treat the functional process as a $(p_\mathcal{X}\times p_\mathcal{Y})$-multivariate process, expressing the functions $\mathcal{X}$, $\mathcal{Y}$, and the surface $\mathcal{B}$ in bases of $p_\mathcal{X}$, $p_\mathcal{Y}$, and $(p_\mathcal{X}\times p_\mathcal{Y})$ elements, respectively.

4.3 Implementation

Following the steps of Escanciano (2006), it is possible to derive a simpler expression for (4.5). Using the definition of the RMPP in a truncated bases, the fact that $F_{n,\mathbf{R}_\mathcal{X}^{-1}\mathbf{d}_{\mathcal{X},p_\mathcal{X}}}$ is the ECDF of $\left\{\mathbf{X}^T_{i,p_\mathcal{X}}\mathbf{\Psi}\mathbf{R}^{-1}\mathbf{d}_{p_\mathcal{X}}\right\}_{i=1}^n = \left\{\mathbf{X}^T_{i,p_\mathcal{X}}\mathbf{R}^T\mathbf{d}_{p_\mathcal{X}}\right\}_{i=1}^n$ and some simple algebra, we have:

$$\begin{aligned}
\text{PCvM}_{n,p_\mathcal{X}p_\mathcal{Y}} &= \int_{\mathbb{S}^{p_\mathcal{X}}\times\mathbb{S}^{p_\mathcal{Y}}\times\mathbb{R}} |\mathbf{R}_\mathcal{X}|^{-1}|\mathbf{R}_\mathcal{Y}|^{-1} R_{n,p_\mathcal{X},p_\mathcal{Y}}(\mathbf{R}_\mathcal{X}^{-1}\mathbf{d}_{p_\mathcal{X}},\mathbf{R}_\mathcal{Y}^{-1}\mathbf{d}_{p_\mathcal{Y}},u)^2 \\
&\quad\cdot F_{n,\mathbf{R}_\mathcal{X}^{-1}\mathbf{d}_{\mathcal{X},p_\mathcal{X}}}(du)\omega_\mathcal{X}(d\mathbf{d}_{\mathcal{X},p_\mathcal{X}})\omega_\mathcal{Y}(d\mathbf{d}_{\mathcal{Y},p_\mathcal{Y}}) \\
&= \int_{\mathbb{S}^{p_\mathcal{X}}\times\mathbb{S}^{p_\mathcal{Y}}\times\mathbb{R}} |\mathbf{R}_\mathcal{X}|^{-1}|\mathbf{R}_\mathcal{Y}|^{-1}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^n \hat{\mathbf{E}}^T_{i,p_\mathcal{Y}}\mathbf{d}_{\mathcal{Y},p_\mathcal{Y}}\mathbb{I}_{\left\{\mathbf{X}^T_{i,p_\mathcal{X}}\mathbf{d}_{\mathcal{X},p_\mathcal{X}}\leq u\right\}}\right)^2 \\
&\quad\cdot F_{n,\mathbf{R}_\mathcal{X}^{-1}\mathbf{d}_{\mathcal{X},p_\mathcal{X}}}(du)\omega_\mathcal{X}(d\mathbf{d}_{\mathcal{X},p_\mathcal{X}})\omega_\mathcal{Y}(d\mathbf{d}_{\mathcal{Y},p_\mathcal{Y}}) \\
&= \frac{1}{n^2}\sum_{i=1}^n\sum_{j=1}^n\sum_{r=1}^n A_{ijr}|\mathbf{R}_\mathcal{Y}|^{-1}\int_{\mathbb{S}^{p_\mathcal{Y}}} \hat{\mathbf{E}}^T_{i,p_\mathcal{Y}}\mathbf{d}_{\mathcal{Y},p_\mathcal{Y}}\hat{\mathbf{E}}^T_{j,p_\mathcal{Y}}\mathbf{d}_{\mathcal{Y},p_\mathcal{Y}}d\mathbf{d}_{\mathcal{Y},p_\mathcal{Y}}.
\end{aligned} \tag{4.6}$$

The integral of the last term can be computed using some integration techniques on the $p_\mathcal{Y}$-sphere, yielding

$$\int_{\mathbb{S}^{p_\mathcal{Y}}} \hat{\mathbf{E}}^T_{i,p_\mathcal{Y}}\mathbf{d}_{\mathcal{Y},p_\mathcal{Y}}\hat{\mathbf{E}}^T_{j,p_\mathcal{Y}}\mathbf{d}_{\mathcal{Y},p_\mathcal{Y}}d\mathbf{d}_{\mathcal{Y},p_\mathcal{Y}} = \frac{\pi^{\frac{p_\mathcal{Y}}{2}-1}}{\Gamma\left(\frac{p_\mathcal{Y}}{2}+1\right)p_\mathcal{Y}}\hat{\mathbf{E}}^T_{i,p_\mathcal{Y}}\hat{\mathbf{E}}_{j,p_\mathcal{Y}}.$$

So that:

$$\text{PCvM}_{n,p_{\mathcal{X}},p_{\mathcal{Y}}} = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{r=1}^{n} A_{ijr}|\mathbf{R}_{\mathcal{Y}}|^{-1}\frac{\pi^{\frac{p_{\mathcal{Y}}}{2}-1}}{\Gamma\left(\frac{p_{\mathcal{Y}}}{2}+1\right)p_{\mathcal{Y}}}\hat{\mathbf{E}}_{i,p_{\mathcal{Y}}}^{T}\hat{\mathbf{E}}_{j,p_{\mathcal{Y}}}.$$

The terms A_{ijr} are the same as the ones given in García-Portugués et al. (2014). For the sake of completeness, they will be reproduced here, but we claim no credit for it. These terms represent the integrals:

$$\begin{aligned} A_{ijr} &:= \int_{\mathbb{S}^{p_{\mathcal{X}}}} |\mathbf{R}|^{-1}\mathbb{I}_{\left\{\mathbf{X}_{i,p_{\mathcal{X}}}^{T}\mathbf{R}^{T}\mathbf{d}_{p_{\mathcal{X}}}\leq\mathbf{X}_{r,p_{\mathcal{X}}}^{T}\mathbf{R}^{T}\mathbf{d}_{p_{\mathcal{X}}}\right\}}\mathbb{I}_{\left\{\mathbf{X}_{j,p_{\mathcal{X}}}^{T}\mathbf{R}^{T}\mathbf{d}_{p_{\mathcal{X}}}\leq\mathbf{X}_{r,p_{\mathcal{X}}}^{T}\mathbf{R}^{T}\mathbf{d}_{p_{\mathcal{X}}}\right\}}d\mathbf{d}_{p_{\mathcal{X}}} \\ &= \int_{\mathbb{S}^{p}} |\mathbf{R}|^{-1}\mathbb{I}_{\left\{(\mathbf{R}\mathbf{X}_{i,p_{\mathcal{X}}}-\mathbf{R}\mathbf{X}_{r,p_{\mathcal{X}}})^{T}\mathbf{d}_{p_{\mathcal{X}}}\leq 0,\ (\mathbf{R}\mathbf{X}_{j,p_{\mathcal{X}}}-\mathbf{R}\mathbf{X}_{r,p_{\mathcal{X}}})^{T}\mathbf{d}_{p_{\mathcal{X}}}\leq 0\right\}}d\mathbf{d}_{p_{\mathcal{X}}} \\ &= |\mathbf{R}|^{-1}\int_{S_{ijr}} d\mathbf{d}_{p_{\mathcal{X}}}, \end{aligned} \tag{4.7}$$

where $Sijr = \left\{\boldsymbol{\xi}\in\mathbb{S}^{p_{\mathcal{X}}} : \frac{\pi}{2}\leq\measuredangle(\mathbf{X}_{i,p_{\mathcal{X}}}-\mathbf{X}_{r,p_{\mathcal{X}}},\boldsymbol{\xi})\leq\frac{3\pi}{2}, \frac{\pi}{2}\leq\measuredangle(\mathbf{X}_{j,p_{\mathcal{X}}}-\mathbf{X}_{r,p_{\mathcal{X}}},\boldsymbol{\xi})\leq\frac{3\pi}{2}\right\}$ and $\measuredangle(\mathbf{a},\mathbf{b})$ represents the angle between vectors $\mathbf{a}$ and $\mathbf{b}$. To simplify notation, we denote $\mathbf{X}'_{k,p_{\mathcal{X}}} = \mathbf{R}\mathbf{X}_{k,p_{\mathcal{X}}}$ ($\mathbf{X}'_{k,p_{\mathcal{X}}} = \mathbf{X}_{k,p_{\mathcal{X}}}$ if the basis is orthonormal) for $k = 1,...,n$. Depending on $\mathbf{X}'_{i,p_{\mathcal{X}}},\mathbf{X}'_{j,p_{\mathcal{X}}},\mathbf{X}'_{r,p_{\mathcal{X}}}$, the region S_{ijr} can be the whole sphere $\mathbb{S}^{p_{\mathcal{X}}}$ ($\mathbf{X}'_{i,p_{\mathcal{X}}} = \mathbf{X}'_{j,p_{\mathcal{X}}} = \mathbf{X}'_{r,p_{\mathcal{X}}}$), a hemisphere of $\mathbb{S}^{p_{\mathcal{X}}}$ ($\mathbf{X}'_{i,p_{\mathcal{X}}} = \mathbf{X}'_{j,p_{\mathcal{X}}},\mathbf{X}'_{j,p_{\mathcal{X}}} = \mathbf{X}'_{r,p_{\mathcal{X}}}$ or $\mathbf{X}'_{i,p_{\mathcal{X}}} = \mathbf{X}'_{r,p_{\mathcal{X}}}$) or a spherical wedge —shown in Figure 4.1— of width angle given by:

$$\left|\pi - \arccos\left(\frac{(\mathbf{X}'_{i,p_{\mathcal{X}}}-\mathbf{X}'_{r,p_{\mathcal{X}}})^{T}(\mathbf{X}'_{j,p_{\mathcal{X}}}=\mathbf{X}'_{r,p_{\mathcal{X}}})}{||\mathbf{X}'_{i,p_{\mathcal{X}}}-\mathbf{X}'_{r,p_{\mathcal{X}}})||\cdot||\mathbf{X}'_{j,p_{\mathcal{X}}}=\mathbf{X}'_{r,p_{\mathcal{X}}}||}\right)\right|. \tag{4.8}$$

Thus A_{ijr} is the product of the surface area of a spherical wedge of angle $A_{ijr}^{(0)}$ times $|\mathbf{R}|^{-1}$, and is given by:

$$A_{ijr} = A_{ijr}^{(0)}\frac{\pi^{\frac{p}{2}-1}}{\Gamma\left(\frac{p}{2}\right)}|\mathbf{R}|^{-1}, \qquad A_{ijr}^{(0)} = \begin{cases} 2\pi, & \text{if } \mathbf{X}'_{i,p_{\mathcal{X}}} = \mathbf{X}'_{j,p_{\mathcal{X}}} = \mathbf{X}'_{r,p_{\mathcal{X}}}, \\ \pi, & \text{if } \mathbf{X}'_{i,p_{\mathcal{X}}} = \mathbf{X}'_{j,p_{\mathcal{X}}}, \mathbf{X}'_{i,p_{\mathcal{X}}} = \mathbf{X}'_{r,p_{\mathcal{X}}}, \text{ or } \mathbf{X}'_{j,p_{\mathcal{X}}} = \mathbf{X}'_{r,p_{\mathcal{X}}}, \\ (4.8), & \text{otherwise.} \end{cases}$$

Joining these terms results the closed and easily computable expression of the statistic

$$\begin{aligned} \text{PCvM}_{n,p_{\mathcal{X}},p_{\mathcal{Y}}} &= \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{r=1}^{n} A_{ijr}|\mathbf{R}_{\mathcal{Y}}|^{-1}\frac{\pi^{\frac{p_{\mathcal{X}}}{2}-1}}{\Gamma\left(\frac{p_{\mathcal{X}}}{2}+1\right)p_{\mathcal{X}}}\hat{\mathbf{E}}_{i,p_{\mathcal{X}}}^{T}\hat{\mathbf{E}}_{j,p_{\mathcal{X}}} \\ &= \frac{|\mathbf{R}_{\mathcal{Y}}|^{-1}}{n^2}\frac{\pi^{\frac{p_{\mathcal{X}}}{2}-1}}{\Gamma\left(\frac{p_{\mathcal{X}}}{2}+1\right)p_{\mathcal{X}}}\sum_{i=1}^{n}\sum_{j=1}^{n}(\mathbf{A}_{\bullet})_{ij}\,\hat{\mathbf{E}}_{i,p_{\mathcal{X}}}^{T}\hat{\mathbf{E}}_{j,p_{\mathcal{X}}} \\ &= \frac{|\mathbf{R}_{\mathcal{Y}}|^{-1}}{n^2}\frac{\pi^{\frac{p_{\mathcal{Y}}}{2}-1}}{\Gamma\left(\frac{p_{\mathcal{Y}}}{2}+1\right)p_{\mathcal{Y}}}\text{Tr}\left[\hat{\mathbf{E}}_{p_{\mathcal{Y}}}^{T}\left(\mathbf{A}_{\bullet}\right)\hat{\mathbf{E}}_{p_{\mathcal{Y}}}\right], \end{aligned} \tag{4.9}$$

where $\mathbf{A}_{\bullet} = \sum_{r=1}^{n} A_{ijr}$ is a $n\times n$ matrix and $\text{Tr}(\mathbf{C})$ denotes the trace of the matrix $\mathbf{C}$. As we are considering orthonormal bases along this work, $|\mathbf{R}_{\mathcal{Y}}| = 1$.

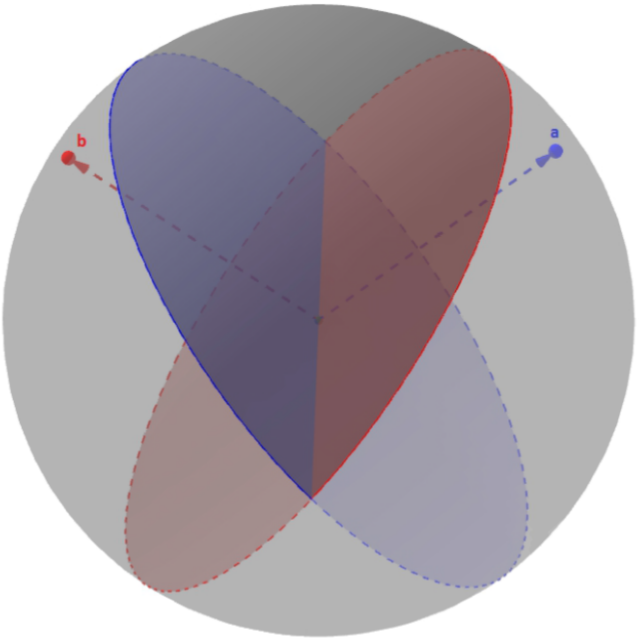

Figure 4.1: Spherical wedge $S_{\mathbf{a},\mathbf{b}} = \left\{\boldsymbol{\xi} \in \mathbb{S}^p : \frac{\pi}{2} \leq \measuredangle(\boldsymbol{\xi}, \mathbf{a}) \leq \frac{3\pi}{2}, \frac{\pi}{2} \leq \measuredangle(\boldsymbol{\xi}, \mathbf{b}) \leq \frac{3\pi}{2}\right\}$ defined by points $\mathbf{a}$ and $\mathbf{b}$ in $\mathbb{S}^2$. Extracted from García-Portugués et al. (2014).

The statistic has been implemented in `R`, using some functions from Febrero-Bande and Oviedo de la Fuente (2012). Let us remark that to speed up the computation of the test statistic, the critical parts of the code are implemented in `FORTRAN`, *e.g.*, the matrix $\mathbf{A}_{\bullet}$.

4.4 Bootstrap resampling

To calibrate the distribution of the statistic $\text{PCvM}_{n,p_{\mathcal{X}},p_{\mathcal{Y}}}$ under the null hypothesis we apply a wild bootstrap on the residuals. This bootstrap methodology is consistent in the finite dimensional case, as it was shown by Stute (1997), and is suitable for situations with potential heterocedasticity, which are common in FDA. The resampling procedure can be done either by perturbing the residuals in the functional space or their components when expressed in a certain basis. The former approach requires to compute basis expansions on each iteration and find the components of the functional processes in an adequate basis to perform the estimation. Therefore, we will use the latter strategy: to perturb the components of the residuals on a fixed basis. Given the initial processes $\mathcal{X}$ and $\mathcal{Y}$, the first step is to find their basis expansions on a truncated basis $\mathcal{X}^{(p_{\mathcal{X}})}$ and $\mathcal{Y}^{(p_{\mathcal{Y}})}$, given by $\mathbf{X}_{p_{\mathcal{X}}}$ and $\mathbf{Y}_{p_{\mathcal{Y}}}$ and compute the estimation of the coefficients of $\mathcal{B}^{(p_{\mathcal{X}},p_{\mathcal{Y}})}$ in the tensor product basis, *i.e.*, $\hat{\mathbf{B}}_{p_{\mathcal{X}}p_{\mathcal{Y}}}$. The resampling process is the following:

1. Estimate the residuals: $\hat{\mathbf{E}}_{i,p_{\mathcal{Y}}} = \mathbf{Y}_{i,p_{\mathcal{Y}}} - \mathbf{X}_{i,p_{\mathcal{X}}}\hat{\mathbf{B}}_{p_{\mathcal{X}}p_{\mathcal{Y}}}$, $i = 1, ..., n$.

2. Draw independent random variables $V_1^*, ..., V_n^*$ satisfying

$$\mathbb{E}^*[V_i^*] = 0 \text{ and } \mathbb{E}^*[V_i^{*2}] = 1$$

For example, if V^* is a discrete random variable such that:

$$\mathbb{P}\left\{V^+ = \frac{1-\sqrt{5}}{2}\right\} = \frac{5+\sqrt{5}}{10} \quad \text{and} \quad \mathbb{P}\left\{V^+ = \frac{1+\sqrt{5}}{2}\right\} = \frac{5-\sqrt{5}}{10}$$

we have the *golden section bootstrap*.

3. Generate the bootstrap residuals: $\mathbf{E}^*_{i,p_\mathcal{Y}} = \hat{\mathbf{E}}_{i,p_\mathcal{Y}} V_i^*$, $i = 1, ..., n$.
4. Estimate $\hat{\mathbf{B}}^*_{p_\mathcal{X} p_\mathcal{Y}}$ from the sample $\{(\mathbf{X}_{i,p_\mathcal{X}}, \mathbf{Y}^*_{i,p_\mathcal{Y}})\}_{i=1}^n$, by setting $\mathbf{Y}^*_{i,p_\mathcal{Y}} = \mathbf{X}_{i,p_\mathcal{X}} \hat{\mathbf{B}}_{p_\mathcal{X} p_\mathcal{Y}} + \mathbf{E}^*_{i,p_\mathcal{Y}}$, $i = 1, ..., n$.
5. Estimated the bootstrap residuals $\hat{\mathbf{E}}^*_{i,p_\mathcal{Y}} = \mathbf{Y}^*_{i,p_\mathcal{Y}} - \mathbf{X}_{i,p_\mathcal{X}} \hat{\mathbf{B}}^*_{p_\mathcal{X} p_\mathcal{Y}}$, $i = 1, ..., n$.

Figure 4.2 illustrates this algorithm. The upper row shows the different responses obtained along the procedure, whereas the lower contains the residuals. Figures 4.2(a) and 4.2(e) show the simulated response $\mathcal{Y}_i$, $i = 1, ..., n$ and noise $\mathcal{E}_i$, $i = 1, ..., n$, respectively. The corresponding estimations $\hat{\mathcal{Y}}_i$ and $\hat{\mathcal{E}}_i$, $i = 1, ..., n$ are plotted in Figures 4.2(b) and 4.2(f). Figure 4.2(g) contains the perturbed residuals $\mathcal{E}^*_i$, $i = 1, ..., n$, which yield the functional response $\mathcal{Y}^*_i$, $i = 1, ..., n$, shown in Figure 4.2(c). The corresponding estimations $\hat{\mathcal{Y}}^*_i$ and $\hat{\mathcal{E}}^*_i$, $i = 1, ..., n$ are plotted in Figures 4.2(d) and 4.2(h), respectively. Note how the typical border effects arising when using Fourier basis expansions are present in the resampling process.

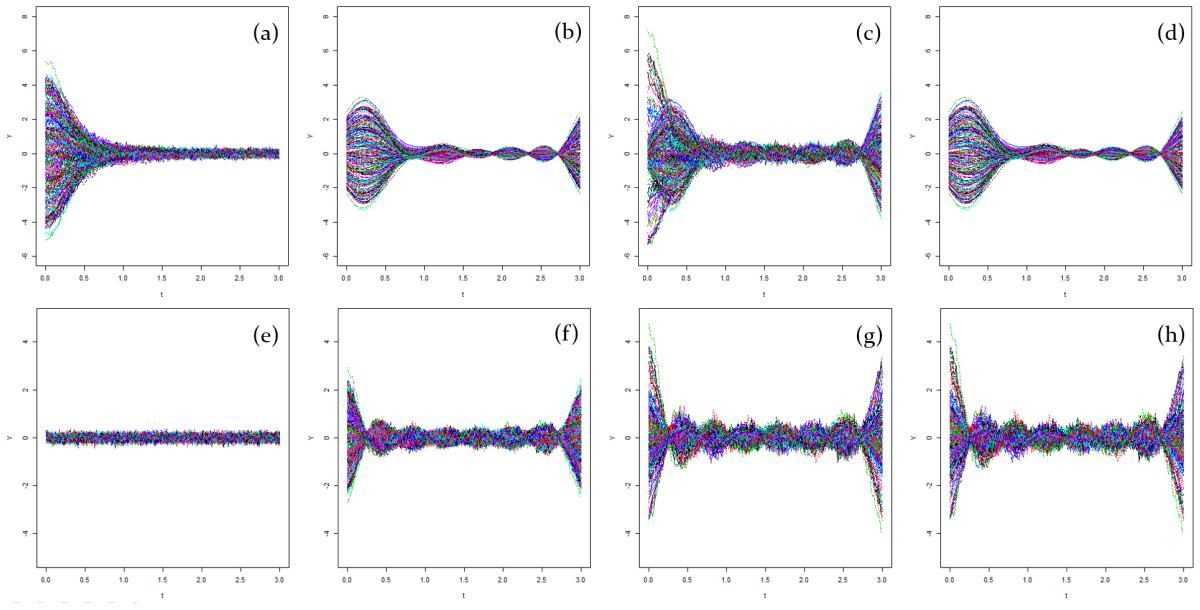

Figure 4.2: The upper row shows the different responses obtained along the procedure, whereas the lower contains the residuals. **(a)** Simulated response $\mathcal{Y}$, **(b)** Estimated response $\hat{\mathcal{Y}}$, **(c)** Perturbed response $\mathcal{Y}^*$, **(d)** Estimation of the perturbed response $\hat{\mathcal{Y}}^*$, **(d)** Simulated residuals $\mathcal{E}$, **(e)** Estimated residuals $\hat{\mathcal{E}}$, **(f)** Perturbed residuals $\mathcal{E}^*$, **(g)** Estimation of the perturbed residuals $\hat{\mathcal{E}}^*$

Then, the procedure to calibrate the test is the following. In Step 1 we compute the test statistic with the residuals under H_0 using the implementation of Section 4.3. Then repeat Steps 2–5 for $b = 1, ..., B$, computing each time the bootstrap statistic

$$\mathrm{PCvM}^b_{n,p_\mathcal{X},p_\mathcal{Y}} = \frac{1}{n^2}\frac{\pi^{\frac{p_\mathcal{Y}}{2}-1}}{\Gamma\left(\frac{p_\mathcal{Y}}{2}+1\right)p_\mathcal{Y}}\mathrm{Tr}\left[\hat{\mathbf{E}}^{*,T}_{p_\mathcal{Y}}\left(\mathbf{A}_\bullet\right)\hat{\mathbf{E}}^{*}_{p_\mathcal{Y}}\right].$$

Finally, the estimation of the p-value of the test is done by means of a Monte Carlo procedure: p-value $\approx \#\{\mathrm{PCvM}^b_{n,p_\mathcal{X},p_\mathcal{Y}} \leq \mathrm{PCvM}_{n,p_\mathcal{X},p_\mathcal{Y}}\}/B$. For computational efficiency, it is important to note that we do not have to compute again the matrix $\mathbf{A}_\bullet$ in the bootstrap replicates.

A very interesting fact of the FLM is that Step 5 can be easily performed using the properties of the estimation of $\hat{\mathbf{B}}_{p_\mathcal{X}p_\mathcal{Y}}$. From (3.14), the matrix vector of coefficients of $\hat{\mathcal{B}}^{(p_\mathcal{X}p_\mathcal{Y})}$ is estimated throughout $\hat{\mathbf{B}}_{p_\mathcal{X}p_\mathcal{Y}} = \left(\mathbf{X}^T_{p_\mathcal{X}}\mathbf{X}_{p_\mathcal{X}}\right)^{-1}\mathbf{X}^T_{p_\mathcal{X}}\mathbf{Y}_{p_\mathcal{Y}}$. Then, the estimated bootstrap residuals in a certain basis representation can be obtained as:

$$\hat{\mathbf{E}}^{*} = \mathbf{Y}^*_{p_\mathcal{Y}} - \hat{\mathbf{Y}}^*_{p_\mathcal{Y}} = \mathbf{Y}^*_{p_\mathcal{Y}} - \mathbf{P}_\Omega\mathbf{Y}^*_{p_\mathcal{Y}} = \left(\mathbb{1}_{p_\mathcal{X}\times p_\mathcal{X}} - \mathbf{X}_{p_\mathcal{X}}\left(\mathbf{X}^T_{p_\mathcal{X}}\mathbf{X}_{p_\mathcal{X}}\right)^{-1}\mathbf{X}^T_{p_\mathcal{X}}\right)\mathbf{Y}^*_{p_\mathcal{Y}},$$

where $\mathbf{Y}^*_{p_\mathcal{Y}}$ is the vector of bootstrap responses given by Step 4 and $\mathbb{I}_{p_\mathcal{X}\times p_\mathcal{X}}$ is the identity matrix of order $p_\mathcal{X} \times p_\mathcal{Y}$. The projection matrix $\mathbb{I}_{p_\mathcal{X}p_\mathcal{Y}} - \mathbf{X}_{p_\mathcal{X}}\left(\mathbf{X}^T_{p_\mathcal{X}}\mathbf{X}_{p_\mathcal{X}}\right)^{-1}\mathbf{X}^T_{p_\mathcal{X}}$ remains the same for all the bootstrap replicates, so it can be stored without the need of computing it again. Obtaining the residuals in this way implies a significative computational saving.

5 Simulation study

To illustrate the finite sample properties of the proposed test, a simulation study is carried out. First of all, we describe briefly the simulation setting.

5.1 Simulation setting

The different data generating processes used in the simulation study are encoded as follows. For the k-th simulation scenario S_k, with $k = 1, 2$, the deviation from H_0 is measured by a deviation coefficient δ_d, with $\delta_0 = 0$ and $\delta_d > 0$ for $d = 1, 2, 3$. Then, under $H_{k,d}$, we denote data generation by

$$\mathcal{Y} = \langle\langle \mathcal{X}_k, \mathcal{B}_k \rangle\rangle + \delta_{k,d}\Delta(\mathcal{X}_k) + \mathcal{E},$$

where the deviations from the linear model are constructed by including the nonlinear term. The error $\mathcal{E}$ is a white noise process, *i.e.* $\mathcal{E}(t) \sim N(0, \sigma^2)$, with[1] $\sigma = 0.1$, $\Delta(\mathcal{X}) := \exp(\mathcal{X})$. The two functional processes $\mathcal{X}_k$ considered in this simulation study, both discretized in 201 equidistant points, are the following:

- **BM**. Brownian motion, whose eigenfunctions are $\psi_j(t) := \sqrt{2}\sin\left(\left(j - \frac{1}{2}\right)\pi t\right)$, $j \geq 1$. We will consider this process in $[0, 2]$.
- **OU**. Ornstein-Uhlenbeck process $\{\mathcal{X}_t\}$, defined by $d\mathcal{X}_t = \alpha(\mu - \mathcal{X}_t)dt + \sigma d\mathcal{W}(t)$, where $\mathcal{W}(t)$ is a Wiener process, μ is the mean and α and σ are positive parameters. We will consider this process in $[0, 2]$, with $\alpha = 1/3$, $\mu = 0$, $\sigma = 1$, and $\mathcal{X}(0) \sim \mathcal{N}\left(0, \frac{\sigma^2}{2\alpha}\right)$.

The linear operators $\mathcal{B}_k(s, t)$, $k = 1, 2$ considered in this simulation study are given by:

$$\mathcal{B}_1(s, t) = \frac{\cos(2\pi t s)}{1 + (s - 0.5)^2} \qquad \text{and} \qquad \mathcal{B}_2(s, t) = \frac{\tanh(1 + s + t^3)}{0.25 + t}(\pi s)^2$$

Figure 5.1 shows $n = 100$ realizations of the processes in Scenarios 1 and 2. These data have been simulated using functions from Febrero-Bande and Oviedo de la Fuente (2012). The upper row shows Scenario 1 and the lower Scenario 2. The covariate processes $\mathcal{X}$ are plotted in the column on the left and

[1] A more appropriate choice for σ can be the following: choose such that, under H_0, $R^2 = 0.95$.

the figures in the middle shows the corresponding observations of $\mathcal{Y}$, without any perturbation. The right column shows the perturbed $\mathcal{Y}$ for $\delta = 0.1$. The perturbations considered in the study are much smaller than these, and are not visible to the naked eye (see Table 5.1).

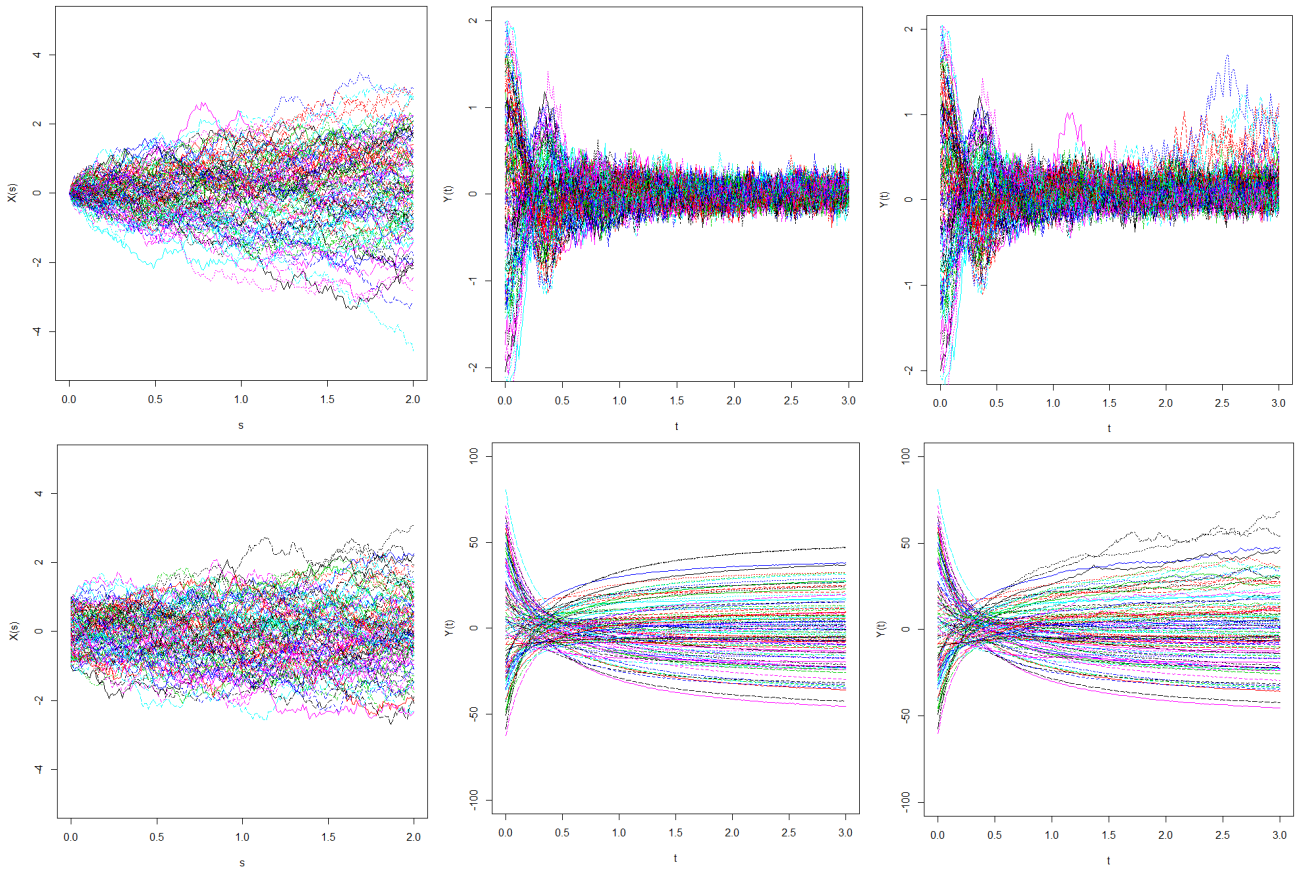

Figure 5.1: Examples of the simulations considered for the simulation study. The upper row shows Scenario 1. In the figure on the left $n = 100$ realizations of $\mathcal{X}$, a Brownian motion are plotted. The figure in the middle shows the corresponding observations of $\mathcal{Y}$, without any perturbation. The figure on the right shows the same process, perturbed by $\delta = 0.1$. The perturbations considered in the study are much smaller and are not visible to the naked eye, meaning that the test is quite powerful. The lower row shows Scenario 2. From left to right: $n = 100$ realizations of an Ornstein–Uhlenbeck process, the response processes without perturbation and the perturbed response, for $\delta = 0.1$. Again, the perturbations considered for the simulation are much smaller.

In the following, the number of bootstrap replicates considered will be $B = 500$ and the number of Monte Carlo replicates for determining the empirical sizes and powers will be $M = 500$. The sample sizes will be $n = 100$ and $n = 200$. The intervals for $\mathcal{X}$ and $\mathcal{Y}$ are $[0, 2]$ and $[0, 3]$, respectively. The detailed description of the simulation scenarios is given in Table 5.1.

Scenario	Process $\mathcal{X}$	$\mathcal{B}(s,t)$	Deviations	Basis expansion
S_1	BM	$\mathcal{B}_1$	$\delta_{1,1} = 0.001$, $\delta_{1,2} = 0.002$, $\delta_{1,3} = 0.005$	Fourier
S_2	OU	$\mathcal{B}_2$	$\delta_{2,1} = 0.005$, $\delta_{2,2} = 0.007$, $\delta_{2,3} = 0.010$	FPCs

Table 5.1: Simulation scenarios and deviations from the null hypothesis.

5.2 Simulation results

The empirical powers of the PCvM test are studied under the null hypothesis and for the three deviations from the null

$$H_{k,d} : \mathcal{Y} = \langle\langle \mathcal{X}_k, \mathcal{B}_k \rangle\rangle + \delta_{k,d}\Delta(\mathcal{X}_k) + \mathcal{E},$$

for Fourier basis expansions and FPCs —both of them with $p_{\mathcal{X}} = p_{\mathcal{Y}} = 3$, 5 and 7 basis elements— and for two sample sizes: $n = 100$ and $n = 200$.

Figure 5.2 shows the Kernel Density Estimate (KDEs) of the PCvM statistic —black solid line— and its bootstrap replicates —red lines—, computed with Fourier basis expansions for $n = 100$ and $p_{\mathcal{X}} = p_{\mathcal{Y}} = 5$ as an example. Note that the disagreement between the KDEs of the PCvM statistic and its bootstrap replicates grows with the deviations δ, as it is expected. On the other hand, Figure 5.3 collects the histograms of the Monte Carlo replicated p-values for this case. The histogram on the left, corresponding to $\delta = 0$ is approximately flat, as it is expected, since the distribution of the p-values is expected to be uniform under the null hypothesis. The other histograms show that the bigger the perturbation δ, the more the histogram deviates from the uniformity.

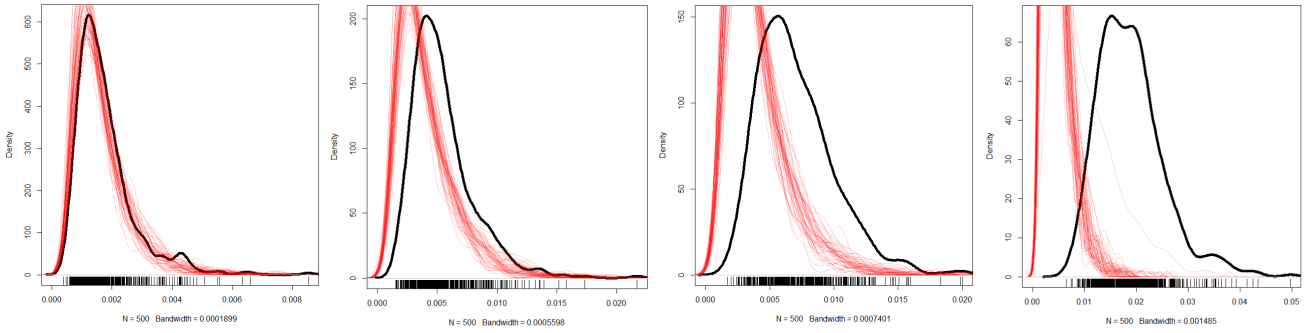

Figure 5.2: KDEs of the PCvM statistic —black solid line— and its bootstrap replicates —red lines—, computed with Fourier basis expansions for $n = 100$ and $p_{\mathcal{X}} = p_{\mathcal{Y}} = 5$, with deviations $\delta_{1,1} = 0.001$, $\delta_{1,2} = 0.002$, $\delta_{1,3} = 0.005$.

The results of the test are collected in Tables 5.2 and 5.3, with Fourier basis expansions and FPCs respectively. All of the tests seem to calibrate well the significance level, $\alpha = 0.05$, except in the case $p_{\mathcal{X}} = p_{\mathcal{Y}} = 7$. Actually, in the case of FPCs the calibration of the test is a little worse, probably due to the fact that it is a data-driven basis and we are not taking this into account in the bootstrap resampling.

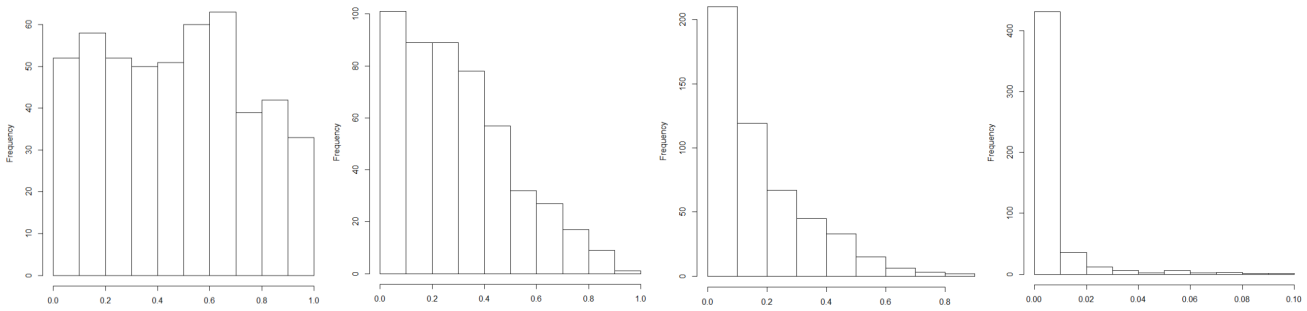

Figure 5.3: Histograms of the Monte Carlo replicated p-values computed with Fourier basis expansions for $n = 100$ and $p_{\mathcal{X}} = p_{\mathcal{Y}} = 5$, with deviations $\delta_{1,1} = 0.001$, $\delta_{1,2} = 0.002$, $\delta_{1,3} = 0.005$.

This might be addressed by exploring different bootstrap procedures, *e.g.* imposing the perturbations on the functional residuals and not on their projections onto the basis. Furthermore, the power of the test increases with increasing number of basis elements, specially in the case of FPCs.

	$n = 100$			$n = 200$		
Model	$p_{\mathcal{X}} = p_{\mathcal{Y}} = 3$	$p_{\mathcal{X}} = p_{\mathcal{Y}} = 5$	$p_{\mathcal{X}} = p_{\mathcal{Y}} = 7$	$p_{\mathcal{X}} = p_{\mathcal{Y}} = 3$	$p_{\mathcal{X}} = p_{\mathcal{Y}} = 5$	$p_{\mathcal{X}} = p_{\mathcal{Y}} = 7$
$H_{1,0}$	0.058	0.046	0.072	0.054	0.052	0.062
$H_{1,1}$	0.130	0.118	0.116	0.252	0.222	0.224
$H_{1,2}$	0.228	0.236	0.372	0.648	0.632	0.726
$H_{1,3}$	0.939	0.974	0.990	1.000	1.000	1.000

Table 5.2: Calibration and empirical power of the goodness-of-fit test for scenario S_1 under a significance level $\alpha = 0.05$.

	$n = 100$			$n = 200$		
Model	$p_{\mathcal{X}} = p_{\mathcal{Y}} = 3$	$p_{\mathcal{X}} = p_{\mathcal{Y}} = 5$	$p_{\mathcal{X}} = p_{\mathcal{Y}} = 7$	$p_{\mathcal{X}} = p_{\mathcal{Y}} = 3$	$p_{\mathcal{X}} = p_{\mathcal{Y}} = 5$	$p_{\mathcal{X}} = p_{\mathcal{Y}} = 7$
$H_{2,0}$	0.058	0.050	0.066	0.062	0.052	0.074
$H_{2,1}$	0.106	0.128	0.250	0.198	0.230	0.588
$H_{2,2}$	0.092	0.174	0.446	0.226	0.288	0.816
$H_{2,3}$	0.120	0.278	0.714	0.290	0.556	0.942

Table 5.3: Calibration and empirical power of the goodness-of-fit test for scenario S_2 under a significance level $\alpha = 0.05$.

Figures 5.4 and 5.5 show the KDEs of the PCvM statistic and its bootstrap replicates computed with FPCs for $n = 200$ and $p_{\mathcal{X}} = p_{\mathcal{Y}} = 5$. Again, the disagreement between the kernel density estimates of the PCvM statistic and its bootstrap replicates grows with the deviations, as it is expected.

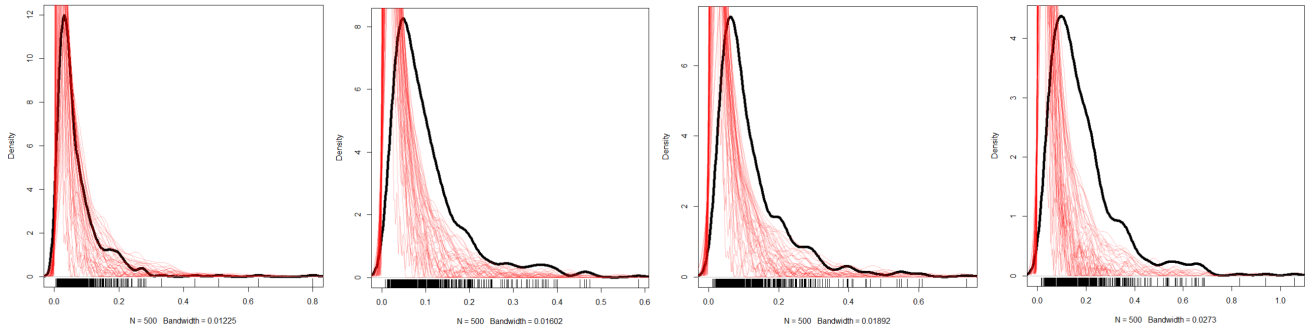

Figure 5.4: KDEs of the PCvM statistic —black solid line— and its bootstrap replicates —red lines— , computed with FPCs for $n = 200$ and $p_{\mathcal{X}} = p_{\mathcal{Y}} = 5$, with deviations $\delta_{1,1} = 0.002$, $\delta_{1,2} = 0.005$, $\delta_{1,3} = 0.010$.

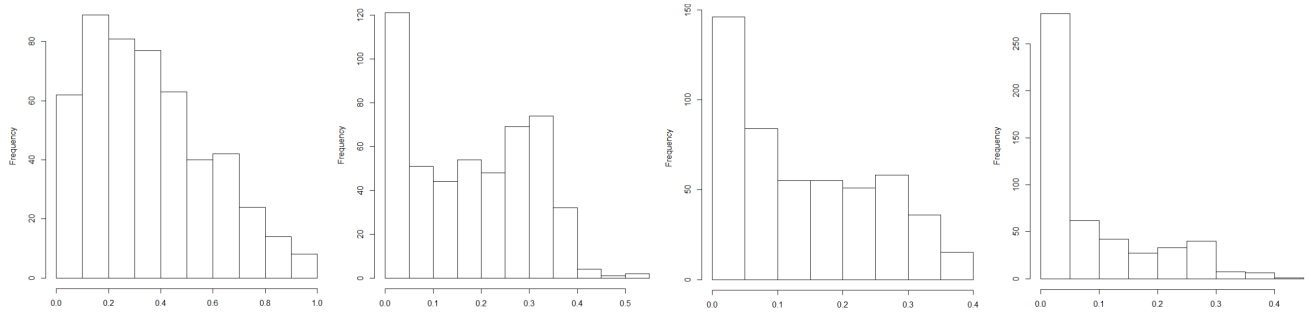

Figure 5.5: Histograms of the Monte Carlo replicated p-values computed with FPCs for $n = 200$ and $p_{\mathcal{X}} = p_{\mathcal{Y}} = 5$, with deviations $\delta_{1,1} = 0.002$, $\delta_{1,2} = 0.005$, $\delta_{1,3} = 0.010$.

6 Conclusions and outlook

A brand-new goodness-of-fit test for the null hypothesis of the functional linear model with functional response based on random projections has been presented. The test is constructed adapting the propose of García-Portugués et al. (2014) to the functional-response framework, with two kinds of basis representations: a deterministic one (Fourier basis expansions) and a data-driven one (FPCs). The test is calibrated on its distribution by means of a wild bootstrap on the residuals expressed in these bases. The simulation study shows that the test behaves well in the practice, as it respects the significance level and has good power. However, it is possible to extend the work in several directions:

1. One of the most immediate future extensions is a more thorough simulation study, by:
   (a) Applying the test to data which are not equispaced, using the numerical schemes for non-equispaced data discussed in Section 2.3.
   (b) Implementing some hyperparameter tunning of the number of basis elements chosen on the projections, such as cross-validation, AIC, or BIC criteria.
   (c) Making a more appropriate choice of the standard deviation of the noise, *e.g.* setting it to meet a given condition, such as a fixed R^2, for what it is necessary to investigate the literature on this R^2 coefficient for the FLMFR.
2. Explore some penalty estimation of $\mathcal{B}$, *e.g.*, making use of shrinkage regression models like ridge or lasso regression, in order to introduce sparsity and penalize the redundant components. This might be also a solution for the hyperparameter tunning of the number of basis elements chosen on the projections, outlined in point 1(b) of these conclusions.
3. Explore more bootstrap procedures: the calibration and power of the test might be improved by imposing the perturbations on the functional residuals and not on their projections onto a chosen basis. This might be specially important for the FPCs expression, as it was noted in Chapter 5.
4. Apply the test to some real dataset.

Some of this issues might be addressed by the author and the supervisor of this work in a subsequent paper.

A

Developed code

The code developed in this thesis, intended to implement a goodness-of-fit test for regression on Fourier basis expansions and FPCs for the FLMFR, is available in the GitHub repository `flm.fr` at:

https://github.com/gonzaloalper/flm.fr

There are several scripts of interest:

- `flm.fourier.R`: example of regression on Fourier basis expansions for the FLM with functional response.
- `fregre.pc.ex.R`: example of regression on FPCs for the FLM with functional response.
- `wb_fourier.R`: implements the wild bootstrap and the simulation study on the projected residuals in a Fourier basis.
- `wb_fpca.R`: implements the wild bootstrap and the simulation study on the projected residuals in a FPCs basis.

The "R" directory contains exclusively `.R` functions, such as:

- `flm_test.R`: performs the goodness-of-fit test.
- `fourier_expansion.R`: computes the projection of a given functional variable onto a Fourier basis.
- `fpc.R`: PCA for functional data.
- `integrateSimp1D.R`: implements the Simpson's rule in one dimension for equispaced data. Extracted from fda.usc
- `integrateSimp2D.R`: implements the Simpson's rule in two dimensions for equispaced data. Extracted from fda.usc

- `linear_model.R`: genertes a linear/non-linear model from provided X, the surface, the noise and the deviation are given.
- `PCvM_statistic.R`: implementation of the test statistic.
- `pseudoinverse`: computes a pseudo-inverse by means of a singular value decompsoition (SVD), needed for the estimatin of the model when $N >> n$.
- `trap1D_unequal.R`: implements the trapezoidal rule in one dimensions for non-equispaced data. Useful for extensions of the work

It should be pointed out that some of these functions and scripts make wide use of the code developed by Febrero-Bande and Oviedo de la Fuente (2012).

Bibliography

Cuesta-Albertos, J., del Barrio, E., Fraiman, R., and Matrán, C. (2007). The random projection method in goodness of fit for functional data. *Computational Statistics and Data Analysis*, 51(10):4814–4831.

de Boor, C. (2001). *A practical guide to splines*. Applied Mathematical Sciences, Springer-Verlag, New York, revised edition.

Escanciano, J. C. (2006). A consistent diagnostic test for regression models using projections. *Econometric Theory*, 22(6):1030–1051.

Febrero-Bande, M. and Oviedo de la Fuente, M. (2012). Statistical computing in functional data analysis: The R package fda.usc. *Journal of Statistical Software*, 51(4):1–28.

Ferraty, F. and Romain, Y. (2011). *The Oxford Handbook of Functional Data Analysis*. Oxford University Press Inc., New York.

Ferraty, F. and Vieu, P. (2006). *Nonparametric Functional Data Analysis: Theory and Practice*. Springer Series in Statistics. Springer, New York.

García-Portugués, E., González-Manteiga, W., and Febrero-Bande, M. (2014). A goodness-of-fit test for the functional linear model with scalar response. *Journal of Computational and Graphical Statistics*, 23(3):761–778.

Jolliffe, I. (2002). *Principal Component Analysis*. Springer Series in Statistics. Springer, New York.

Kokoszka, P. and Reimherr, M. (2017). *Introduction to Functional Data Analysis*. Texts in Statistical Science. CRC Press.

Patilea, V., Sanchez-Sellero, C., and Saumard, M. (2012). Projection-based nonparametric testing for functional covariate effect. *https://arxiv.org/abs/1205.5578*.

Press, W. H., Teukolsky, S. A., Vetterling, W. T., and Flannery, B. P. (1992). *Numerical Recipes in Fortran 77. The Art of Scientific Computing*. Volume 1 of Fortran Numerical Recipes. Cambridge University Press., second edition.

Ramsay, J. O. and Silverman, B. W. (2005). *Functional Data Analysis*. Springer Series in Statistics. Springer, New York, second edition.

Stute, W. (1997). Nonparametric model checks for regression. *Annals of Statistics*, 25(2):613–641.