ESTIMATES AND BOOTSTRAP CALIBRATION FOR FUNCTIONAL REGRESSION WITH SCALAR RESPONSE

Adela Martínez Calvo

Departamento de Estatística e Investigación Operativa

Universidade de Santiago de Compostela

Don Wenceslao González Manteiga, catedrático do Departamento de Estatística e Investigación Operativa da Universidade de Santiago de Compostela, Don Frédéric Ferraty, catedrático do Institut de Mathématiques de Toulouse da Univesité Paul Sabatier – Toulouse III, e Don Philippe Vieu, catedrático do Institut de Mathématiques de Toulouse da Univesité Paul Sabatier – Toulouse III, informan que a memoria titulada

ESTIMATES AND BOOTSTRAP CALIBRATION
FOR FUNCTIONAL REGRESSION WITH SCALAR RESPONSE

foi realizada baixo a súa dirección por Dona Adela Martínez Calvo, estimando que a interesada se atopa en condicións de optar ao grao de Doutor, polo que solicitan que sexa admitida a trámite para a súa lectura e defensa pública.

En Santiago de Compostela, a 21 de xaneiro de 2013.

Os directores:

Prof. Dr. Wenceslao González Manteiga

Prof. Dr. Frédéric Ferraty

Prof. Dr. Philippe Vieu

A doutoranda:

Adela Martínez Calvo

Contents

Preface

Nowadays the progress of computational tools (both memory and capacity increasing) allows to create, store and work with large databases. In many cases, the dataset is made up of observations from a finite dimensional distribution, measured over a period of time or recorded at different spatial locations. When the temporal or spatial grid is fine enough, the sample can be considered as an observation of a random variable on a certain functional space. Analysing this kind of data with standard multivariate methods and ignoring its functional feature may fail dramatically (curse of dimensionality, collinearity, valuable information loss, etc.). In these cases, specific statistical techniques are required in order to manage, leak and draw relevant underlying information.

This fact has turned *Functional Data Analysis* (FDA) into one of the most active statistical fields in recent years. From the seminal works in the eighties and the nineties (Grenander, 1981; Dauxois et al., 1982; Ramsay, 1982; Bosq, 1991), FDA gave rise to several books (Bosq, 2000; Ramsay and Silverman, 1997, 2002, 2005; Ferraty and Vieu, 2006b; Ferraty and Romain, 2011; Horváth and Kokoszka, 2012), special issues in high impact factor journals (Davidian et al., 2004; González-Manteiga and Vieu, 2007; Valderrama, 2007; Ferraty, 2010) and international workshops devoted to both methodological and applied developments for functional data (Dabo-Niang and Ferraty, 2008; Ferraty, 2011). Furthermore, some contributions focused on the state of the art of FDA were published during the last decade (Rice, 2004; Müller, 2005; González-Manteiga and Vieu, 2011; Delsol et al., 2011a; Cuevas, 2012).

It must be emphasized that functional data come up in a natural way in most scientific fields: econometrics (e.g., daily stock returns or electricity production/demand curves), engineering sciences (e.g., satellite imagery, topographic maps or image recognition), environmetrics and climatology (e.g., meteorological measurements or fluvial flows curves), medicine (e.g., growth curves or genetic data), chemometrics (e.g., spectrometric data), etc. This fact generated a variety of applied FDA works in the literature, such as Ramsay and Silverman (2002), and Ferraty and Vieu (2006a).

Due to the novelty of FDA, there is a wide range of research lines which could be explored: from the extension to the functional context of well–known multidimensional methods, to the creation of new statistical techniques devoted to specific functional data issues. Nevertheless, in spite of the functional nature of the data, the pursued aims are essentially the same as the usual stated objectives for a multivariate dataset analysis. Consequently, the developed methodology has been intended to satisfy similar needs:

- Preprocess the data: registration and feature alignment, smoothing techniques, etc.
- Depict and explore the data, highlighting their most important features: measures of centrality and dispersion, detection of outliers, *Functional Principal Component Analysis* (FPCA), etc.
- Functional data classification.
- Build models to explain the relationship between functional variables: parametric and nonparametric regression models.
- Functional statistical inference: confidence intervals, hypotheses testing, etc.

This manuscript mainly deals with the last two items. As far as the fourth item is concerned, the work has been focused on the functional linear model with scalar response (although some contributions

to nonparametric regression are also included in the last chapter), whereas regarding the fifth item a bootstrap procedure has been developed, which allows to build confidence intervals and calibrate hypotheses tests related to the functional linear model.

The thesis has been structured as follows.

Chapter 1. Introduction to FDA. In this chapter, some basic concepts on FDA are defined. The notion of functional data is introduced, and some examples are given, which will be suitable to illustrate the methods proposed in the next chapters. A general background is also presented: preprocessing techniques, functional descriptive statistics, and some key exploratory methods in FDA. Furthermore, this chapter is used to fix the notation and give a brief summary of the state of the art on statistical methods for functional data.

Chapter 2. Functional regression models. Chapter 2 is devoted to functional regression models. At first, a general review of functional regression is presented, and then the efforts are concentrated on models with scalar response and functional covariate. There are two main approaches to discuss this subject: the parametric and the nonparametric approaches. As regards the parametric approach, the two most popular estimators are introduced: estimators based on basis systems and FPCA–type estimators. As far as the nonparametric approach is concerned, the functional version of the multivariate kernel–type estimator is also analysed in this chapter.

Chapter 3. Presmoothing in functional linear regression. The chapter is focused on the functional linear model with scalar response, and explanatory variable valued in a functional space. As it was discussed in Chapter 2, FPCA has been used to estimate the model functional parameter in recent statistical literature. A modification of this approach by using presmoothing techniques is proposed in Chapter 3: either presmoothing via covariance structure or presmoothing via response variable. For these new estimators, consistency is stated and efficiency by comparison with the standard FPCA estimator is analysed from a theoretical point of view. The effectiveness of the proposed presmoothed estimators relative to the standard FPCA estimator and the penalized B–splines estimator is also tested by means of simulation studies and real data applications.

Chapter 4. Bootstrap in functional linear regression. Weak convergence for a wide class of FPCA–type estimators has been obtained for the functional linear model with scalar response in the literature. Consequently, asymptotic confidence intervals for the linear regression operator can be computed from this theoretical result. Chapter 4 presents an alternative approach to this issue which allows obtaining pointwise confidence intervals by means of a bootstrap procedure. Specifically, algorithms for naive and wild bootstrap are developed and asymptotic validity of both bootstrap methods is proven. In addition, a simulation study compares the practical performance of asymptotic and bootstrap confidence intervals in terms of length and coverage rates for two linear regression operators and several sample sizes.

Chapter 5. Testing in functional linear regression. The first part of this chapter is devoted to introduce a consistent bootstrap method to calibrate the distribution of test statistics for assesing the lack of dependence in the functional linear model with scalar response. The asymptotic theory related to this bootstrap approach is also developed. The second part of Chapter 5 presents a bootstrap method for checking the equality of two linear models. For this methodology, a study of its main asymptotic properties is done in order to show its consistency and correctness. Finally, a simulation study and a real data example illustrate the performance of the proposed bootstrap techniques in practice.

Chapter 6. Thresholding in nonparametric functional regression. This chapter presents an exploratory tool focused on the detection of underlying complex structures in the nonparametric

regression model with scalar response and functional covariate. The proposed methodology analyses the existence of hidden patterns via a threshold procedure. A cross–validation criterion which allows to estimate the parameters involved in the threshold model is also introduced, and the usefulness of its graphical representation is studied. Furthermore, a simulation study and applications to real datasets show the effectiveness of the threshold approach from a practical point of view.

Conclusions and further research. An overview of the main conclusions of the thesis and some open questions which require further research are included in this section.

Summary. This section presents a summary of this manuscript in English language.

Resumo en galego. A summary of this manuscript in Galician language is enclosed in the final part of this dissertation.

As well as the theoretical developments of the methodology presented in Chapter 3, Chapter 4, Chapter 5 and Chapter 6, it is important to highlight that all the proposed methods were implemented and applied to both simulated and real datasets. For this purpose, the statistical software *R* was chosen (see further details in R Development Core Team, 2010, or `http://www.r-project.org`), and *R* routines were developed for each one of the new techniques compiled in this document. In order to build them, some existing routines were taking as a starting point. For instance, functions related with the paper by Cardot et al. (2003c) (available at `http://www.math.univ-toulouse.fr/staph`, section *"Softwares on line"*) were considered for the model parameter estimation in the functional linear model with scalar response, whereas *R* routines corresponding to nonparametric methods discussed in the book by Ferraty and Vieu (2006b) (available at `http://www.math.univ-toulouse.fr/staph/npfda`) served as a basis for nonparametric regression contributions. Furthermore, some specific *R* packages devoted to FDA, such as *fda* or *fda.usc* (see Ramsay et al., 2011, and Febrero-Bande and Oviedo de la Fuente, 2011, respectively), were also consulted.

Acknowledgements. *I would like to thank my advisors, Prof. Wenceslao González–Manteiga, Prof. Frédéric Ferraty, and Prof. Philippe Vieu, for their fruitful comments and their support during this years. I also wish to thank Prof. Gil González–Rodríguez, who has allowed me to enrich the theoretical contents of Chapter 5 by including some theoretical results which he developed for our paper González-Manteiga et al. (2012).*

This work has been supported by Ministry of Science and Innovation (FPI grant BES–2006–13389; national projects MTM2005–00820 and MTM2008–03010); and by Consellería de Innovación e Industria (regional project PGIDIT07PXIB207031PR) and Consellería de Economía e Industria (regional project 10MDS207015PR), Xunta de Galicia.

Chapter 1

Introduction to FDA

The basic concepts on FDA are presented in this chapter. Some essential definitions related to functional data and functional spaces are introduced, jointly with the real datasets which will be used to show the practical behaviour of some techniques introduced in the forthcoming chapters. Next, a summarized review of some of the tools developed to preprocess and explore functional data is presented. Special attention is paid to FPCA since it will be the basis of the functional linear regression estimators proposed in Chapter 3. In order to elaborate this chapter, the books by Ramsay and Silverman (2005) and Ferraty and Vieu (2006b) have been used as main references. Furthermore, overviews by González-Manteiga and Vieu (2011) and Cuevas (2012) have been really helpful in order to keep the list of references up to date.

An important part of the state of the art of the techniques compiled in this chapter was published in Delsol et al. (2011a).

1.1 Functional data

1.1.1 What are functional data?

In several fields of the applied sciences, the development of new technologies which enables the continuous time monitoring of many measures makes the objects of study to be curves instead of scalars or vectors. This is what happens when one has to manage daily temperature datasets (Ramsay and Silverman, 2005), spectrometric curves (Ferraty and Vieu, 2006b), concentrations of atmospheric pollutants (Fernández de Castro et al., 2005; Febrero-Bande et al., 2007), and many other measurements. In these contexts, the use of functional methods is essential. The functional approach allows to study and interpret the global behaviour of the stochastic process, something impossible to achieve when the data are analysed as realizations of a random vector. However, the discrete nature of the real datasets must not be forgotten: although the grid where observations are taken becomes finer and finer, the discretization effect cannot be ignored.

The main idea in FDA is to suppose that a random variable X is observed in a discrete grid $\{t_l\}_{l=1}^{L}$ with each $t_l \in T$. If the time instants t_l are close enough, one can assume that $\{X(t_l)\}_{l=1}^{L}$ is an observation of a functional random variable

$$X = \{X(t); t \in T\}.$$

Note that T is often a real interval such as $T = [T_0, T_1] \subset \mathbb{R}$, and each observation can be considered as a curve. Nevertheless, FDA covers many other situations where the space T is a more general functional space, including multidimensional ones. Next, the functional variable and functional dataset definitions introduced by Ferraty and Vieu (2006b), jointly with the functional random sample definition, are recalled in order to fix ideas.

Definition 1.1.1. A random variable X is called a *functional random variable* if it takes values in an infinite dimensional space or functional space $\mathcal{S}$. An observation x of X is called a *functional datum*.

Given a probability space $(\Omega, \mathcal{A}, \mathbb{P})$, a functional random variable X is indeed a measurable mapping from Ω to an infinite dimensional space or functional space $\mathcal{S}$. In this case, it will be said that X is a functional random variable valued in $\mathcal{S}$, or X is a $\mathcal{S}$–valued functional random variable.

Definition 1.1.2. Given a functional random variable X and $n \in \mathbb{N}^*$, a *functional random sample* of X of lenght n is a set $\{X_i\}_{i=1}^n$ of independent and identically distributed (i.i.d.) functional random variables with the same distribution as X. An observation $\{x_i\}_{i=1}^n$ of $\{X_i\}_{i=1}^n$ is called a *functional dataset*.

The general definitions given before include both the most common available functional datasets as curves or surfaces, and more general functional variables valued in metric or semi–metric spaces.

Sometimes, one faces high–dimensional observations which cannot be correctly analysed by means of the classical multivariate tools, due to the high correlation of their components, but which do not seem to be functional by nature at first sight. To solve this issue, Chen et al. (2011) proposed *stringing* methods in order to map this high–dimensional data into an infinite dimensional functional space in such a way that they can be treated using FDA methodology.

There are important links between FDA and Longitudinal Data Analysis. Although some FDA techniques can be used in longitudinal context, this cannot be done in general since longitudinal data are notable for the sparseness of the discretization grid. Some recent contributions discussed and compared longitudinal and functional approaches (see, for instance, James (2002); James and Sugar (2003); Davidian et al. (2004); Zhao et al. (2004); Müller (2005); Hall et al. (2006); Yao (2007); Müller and Yang (2010); James (2011)).

1.1.2 Examples

In this subsection, some examples of functional data are introduced, which will be analysed by means of different statistical methods throughout the manuscript. The first example corresponds to a simulated dataset and the next three to real data. Furthermore, a brief state of the art on applied contributions involving real functional data can be found in the next subsection.

Example 1. Brownian motion

A simulated dataset is the first example of functional data included in this section: a Brownian motion. A *Brownian motion* defined on $T = [0, 1]$ is a random Gaussian element B such that

$$\mathbb{E}(B(t)) = 0 \quad \text{and} \quad \text{Cov}(B(s), B(t)) = \min(s, t), \quad \forall s, t \in T.$$

A sample of 50 observations of a Brownian motion is depicted in Figure 1.1 (see page 3). The curves are discretized on a grid of 100 equidistant points.

Example 2. Canadian weather data

The second example comes from the Canadian weather data. This dataset contains averaged daily temperature and precipitation at 35 different locations in Canada from 1960 to 1994. The Canadian weather data illustrated many functional methods in the book by Ramsay and Silverman (2005) (see also the webpage `http://www.psych.mcgill.ca/misc/fda`), and it is available in the *R* package *fda* (see Ramsay et al., 2011). Figure 1.2 (see page 3) shows the 35 daily temperature curves at each weather station, being

$X_i(t)$: daily averaged temperature in the weather station i in the day t, with $t \in \{1, \ldots, 365\}$.

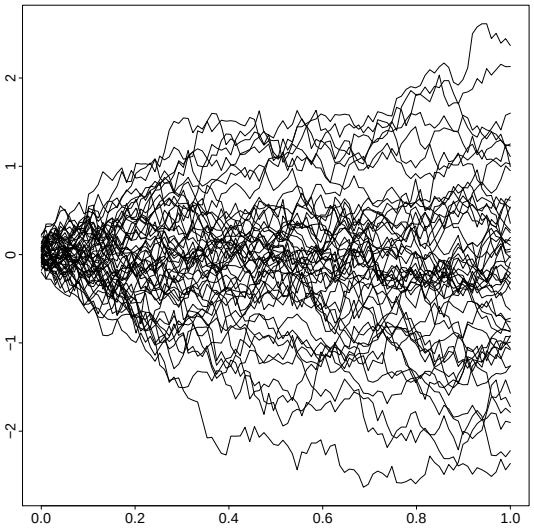

Figure 1.1: Brownian motion. Sample of 50 simulated observations.

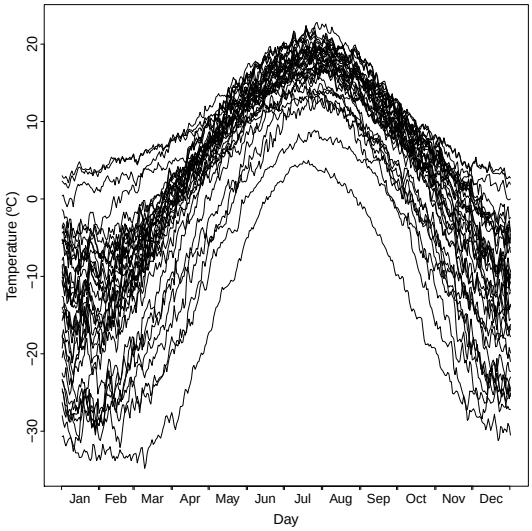

Figure 1.2: Canadian weather data. Sample of 35 daily temperature curves.

Example 3. Spectrometric data

The third example corresponds to spectrometric curves: the pork data. The spectrometric dataset is a part of a sample which can be downloaded from `http://lib.stat.cmu.edu/datasets/tecator`, and it was also analysed from a functional point of view in the literature (Ferraty and Vieu, 2006b; see also the companion website `http://www.math.univ-toulouse.fr/staph/npfda`). The dataset concerns a sample of 215 pieces of finely chopped meat. For each unit, a spectrometric curve is observed which corresponds to the absorbance[1] at 100 wavelengths. The spectrometric dataset was recorded on a Tecator Infratec Food and Feed Analyzer. The left panel of Figure 1.3 (see page 4) collects these 215 curves, given by

$$X_i(t): \text{ absorbance of the piece } i \text{ of meat at } t \text{ nm, with } t \in \{850, \ldots, 1050\}$$

(note that nm is the symbol for nanometre). In the chemometric community, it is well–known that derivatives of spectra are more informative than the original ones. This is the reason why right panel of Figure 1.3 collects the second derivatives of the spectrometric curves which will play an important role later on.

Example 4. Atmospheric pollution data

Finally, an environmental example have been selected: air pollution data. The data is a time series Z corresponding to the concentration of hourly averaged NO_x measured in the neighbourhood of a power station belongs to ENDESA, located in As Pontes in the Northwest of Spain. The NO_x level was measured each minute from 2007 to 2009. The time series Z has been divided in various paths corresponding to 4 hour periods (that is, the curves are discretized at 240 points). Thus,

$$X_i(t) = Z(t + (i-1) \times 240): \text{ hourly averaged } NO_x \text{ in the minute } t \text{ (period } i\text{), with } t \in \{1, \ldots, 240\}.$$

Figure 1.4 (see page 4) presents 100 of these curves. Note that, though the NO_x values are low in general, during unfavourable meteorological conditions the NO_x levels may quickly rise and cause an

[1]The absorbance is defined as $A_w = -\log_{10}(I(w)/I_0(w))$, where $I(w)$ is the intensity of light at the wavelength w that has passed through the sample, whereas $I_0(w)$ is the intensity of the light before it enters the sample.

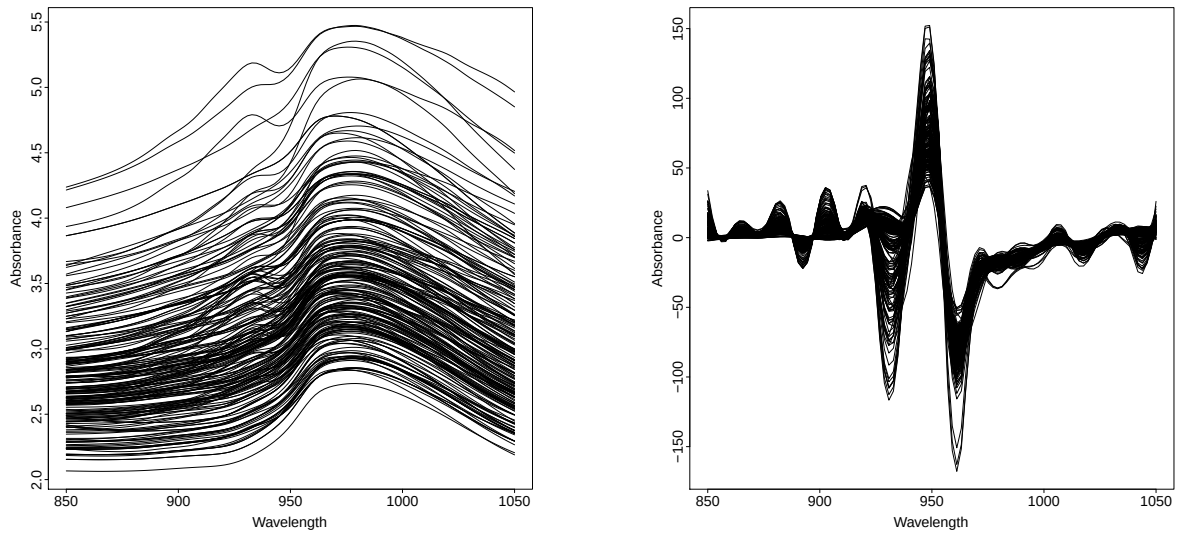

Figure 1.3: Spectrometric data. Sample of 215 spectrometric curves (left panel) and their second derivatives (right panel).

air pollution episode[2]. This fact explains why the sample consists of several almost constant curves and very few increasing/decreasing curves which correspond to air pollution episodes.

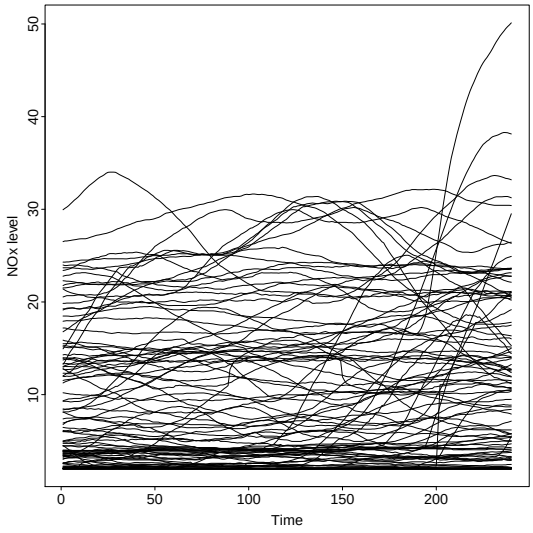

Figure 1.4: Atmospheric pollution data. Sample of 100 hourly averaged NO_x levels.

1.1.3 Other functional datasets

Many contributions devoted to FDA contain applied issues, for example, Ramsay and Silverman (2002) or Ferraty and Vieu (2006a). These studies concern functional data coming from a variety of scientific fields. Next, a selected collection of these applied contributions are summarized.

[2]An *air pollution episode* is a period of abnormally high concentration of air pollutants, often due to low winds and temperature inversion, that can cause illness and death.

a) Climatology and environmetrics

Recent environmental and climatological problems have required the use of FDA techniques (see, for instance, Escabias et al., 2005; Ramsay and Silverman, 2005). Many of the applied contributions to climatology are related with El Niño phenomenon (Besse et al., 2000; Valderrama et al., 2002; Antoniadis and Sapatinas, 2003; Ferraty et al., 2005), or other climatic datasets (Vidakovic, 2001; Ramsay and Silverman, 2002; Hall and Vial, 2006b).

As far as environmetrics applications are concerned, air pollution studies (Damon and Guillas, 2002; Fernández de Castro et al., 2005; Aneiros-Pérez et al., 2004; Cardot and Sarda, 2006; Cardot et al., 2007b; Febrero-Bande et al., 2007; Meiring, 2007; Fernández de Castro and González-Manteiga, 2008; Ferraty and Vieu, 2009), papers on atmospheric radioactivity (Cardot et al., 2007d; Hlubinka and Prchal, 2007), and water quality control contributions (Henderson, 2006; Nerini and Ghattas, 2007) can be found in the literature.

b) Chemometrics

Spectrometry has focused the efforts of the statistical FDA community due to the intrinsic functional nature of the spectrometric curves. From the early papers (Leurgans et al., 1993) to the most recent works (Leardi, 2003; Amato et al., 2006; Ferraty et al., 2010a), spectrometric data have been the object of study of many authors, specially the pork dataset presented before (see, amongst others, Borggaard and Thodberg, 1992; Ferraty and Vieu, 2002; Ferré and Yao, 2005; Amato et al., 2006; Ferraty and Vieu, 2006b; Ferraty et al., 2006; Aneiros-Pérez and Vieu, 2006; Ferraty et al., 2007a; Mas and Pumo, 2007; Ferraty et al., 2007b; Ferraty and Vieu, 2009; Burba et al., 2009).

c) Engineering

Functional dataset comes easily from engineering frameworks as the satellite imagery or signal recognition. In the literature, FDA has been applied to satellite data (Vidakovic, 2001; Cardot et al., 2003a; Cardot and Sarda, 2006), radar curves (Dabo-Niang et al., 2007) or sound signal recognition datasets (Lucero, 1999; Hall et al., 2001; Hastie et al., 2001; Ferraty and Vieu, 2003, 2006b).

d) Econometrics

Functional data appears quite often in the economics, where continuous time series can be split and analysed as curves. Examples can be found in Kneip and Utikal (2001); Ramsay and Silverman (2002); Ferraty et al. (2002); Laukaitis and Rackauskas (2002); Kawasaki and Ando (2004); Reddy and Dass (2006); Jank and Shmueli (2006); Wang et al. (2008); Laukaitis (2008); Liu and Müller (2009); Benko et al. (2009); Müller et al. (2011).

e) Biometrics

One of the most important scientific fields involving FDA is medicine. Many medical datasets were analysed from a functional viewpoint: growth curves (Ramsay et al., 1995; Gasser et al., 1998; James et al., 2000; James and Sugar, 2003; Ramsay and Silverman, 2005; Liu and Yang, 2009), human motion and perception (Ramsay et al., 1996; Ramsay and Silverman, 2002; Spitzner et al., 2003; Ormoneit et al., 2005; López-Pintado and Romo, 2007; Antoniadis and Sapatinas, 2007), placebo effects (Tarpey et al., 2003), mortality (Hyndman and Ullah, 2007; Chiou and Müller, 2009), genetics (Parker and Wen, 2009), cancerology (Ramsay and Silverman, 2005; Cao and Ramsay, 2007; Erbas et al., 2007), cardiology (Clot, 2002; Ratcliffe et al., 2002a,b; Cuevas et al., 2004; Harezlak et al., 2007), neurology (Epifanio and Ventura-Campos, 2011; Goldsmith et al., 2011, 2012) or ophthalmology (Locantore et al., 1999).

With regard to biostatistics, FDA methods have successfully been applied to many contexts as the handbook by Härdle et al. (2007) captured. Among the applications one can emphasize works connected with animal biological problems (Chiou et al., 2003a,b; Müller and Stadtmüller, 2005; Chiou

and Müller, 2007), and contributions on genetics (Araki et al., 2004; Leng and Müller, 2005; Opgen-Rhein and Strimmer, 2006).

f) Further applications

There are many other sciences where FDA methods find feasible applications: geology (Manté et al., 2007), oceanology (Nerini and Ghattas, 2007), demography (Hyndman and Ullah, 2007), graphology (Hastie et al., 1995; Ramsay, 2000a,b), etc.

Another source of functional datasets arises from the standard multivariate methods. There are certain interesting curves which are estimated when one works with a multivariate sample, for example the density function. These kinds of curves can be treated as functional data, and analysed by means of FDA methodology. Some contributions can be found in the literature for density functions (Kneip and Utikal, 2001; Ramsay and Silverman, 2002; Nerini and Ghattas, 2007; Delicado, 2007), distribution functions (Manté et al., 2007), regression functions (Härdle and Marron, 1990; Heckman and Zamar, 2000), or other probabilistic functional characteristics (Rossi et al., 2002).

1.2 Functional space

Let $\{X_i\}_{i=1}^n$ be a functional random sample, that is, n i.i.d. functional random variables with the same distribution as a functional variable X valued in an abstract space $\mathcal{S}$. In many common situations, functional data are curves, so $\mathcal{S} = L^2([0,1])$ is quite often used (Crambes et al., 2009). However, more general spaces, as Hilbert spaces ($\mathcal{S} = (\mathcal{H}, \langle\cdot,\cdot\rangle)$) or Banach spaces ($\mathcal{S} = (\mathcal{B}, \|\cdot\|)$), are considered sometimes in order to solve certain technical problems (Bosq, 2000). In other cases, a broader framework is considered: a functional space $\mathcal{S}$ endowed with a semi–metric $d(\cdot,\cdot)$ (Ferraty and Vieu, 2006b).

Throughout this manuscript, it has been assumed that the space where functional variables take values is a real separable Hilbert space $\mathcal{H}$, although sometimes the functional space has been restricted to the well–known L^2–space to illustrate certain methods. On the other hand, the need of measures to determine the closeness of functional observations has motivated the use of semi–metrics in FDA. These are the reasons why this section has been devoted to the introduction of some concepts connected with both Hilbert spaces and semi–metrics.

1.2.1 The Hilbert space $\mathcal{H}$

First of all, some basic concepts, which are necessary to introduce the definition of Hilbert space, are included in order to clarify the kind of features and properties of the functional space.

Definition 1.2.1. A *field* F is a set with two binary operations[3], usually called *addition* ($+_F : F \times F \to F$) and *multiplication* ($\cdot_F : F \times F \to F$), that satisfies the following axioms for all $a, b, c \in F$:

(i) Associativity of $+_F$ and $\cdot_F$: $(a +_F b) +_F c = a +_F (b +_F c)$, and $(a \cdot_F b) \cdot_F c = a \cdot_F (b \cdot_F c)$.

(ii) Commutativity of $+_F$ and $\cdot_F$: $a +_F b = b +_F a$, and $a \cdot_F b = b \cdot_F a$.

(iii) Identity element of $+_F$ and $\cdot_F$: $\exists\, 0_F \in F$ such that $a +_F 0_F = a = 0_F +_F a$, $\forall a \in F$, and $\exists\, 1_F \in F$ such that $a \cdot_F 1_F = a = 1_F \cdot_F a$, $\forall a \in F$.

(iv) Inverse element of $+_F$ and $\cdot_F$: $\forall a \in F$, $\exists -a \in F$ such that $a +_F (-a) = 0 = (-a) +_F a$. In addition, if $a \neq 0$, then $\exists a^{-1} \in F$ such that $a \cdot_F a^{-1} = 1 = a^{-1} \cdot_F a$.

(v) Distributivity of $\cdot_F$ with respect to $+_F$ and distributivity of $+_F$ with respect to $\cdot_F$: $a \cdot_F (b +_F c) = a \cdot_F b +_F a \cdot_F c$, and $(a +_F b) \cdot_F c = a \cdot_F c +_F b \cdot_F c$.

The elements of a field F are called *scalars*.

---

[3] A *binary operation* on a non–empty set A is a map $f : A \times A \to A$ such that f is defined for every pair of elements

For instance, the rational numbers ($\mathbb{Q}$), the real numbers ($\mathbb{R}$) or the complex numbers ($\mathbb{C}$) are fields.

Remark 1.2.2. To simplify the notation, and whenever there is no possible confusion, $a +_F b$, $a \cdot_F b$, 0_F and 1_F will be denoted by $a + b$, ab, 0 and 1, respectively.

Definition 1.2.3. A *vector space V over a field F* is a set with two operations, usually called *addition* ($+_V : V \times V \to V$) and *scalar multiplication* ($\cdot_V : F \times V \to V$), that satisfies the following axioms for all $a, b \in F$ and $x, y, z \in V$:

(i) Associativity of $+_V$: $(x +_V y) +_V z = x +_V (y +_V z)$.

(ii) Commutativity of $+_V$: $x +_V y = y +_V x$.

(iii) Identity element of $+_V$ and $\cdot_V$: $\exists 0_V \in V$ such that $y +_V 0_V = y = 0_V +_V y$, $\forall y \in V$; if $1_F \in F$ denotes the multiplicative identity element in F, then $1_F \cdot_V y = y$, $\forall y \in V$.

(iv) Inverse element of $+_V$: $\forall y \in V$, $\exists -y \in V$ such that $y +_V (-y) = 0 = (-y) +_V y$.

(v) Distributivity of $\cdot_V$ with respect to $+_V$, and $\cdot_V$ with respect to $+_F$: $a \cdot_V (x +_V y) = a \cdot_V x +_V a \cdot_V y$, and $(a +_F b) \cdot_V y = a \cdot_V y +_V b \cdot_V y$.

(vi) $a \cdot_V (b \cdot_V y) = (a \cdot_F b) \cdot_V y$.

The elements of a vector space V are called *vectors*.

The most commonly used vector spaces are those over $\mathbb{R}$ (real vector spaces) or $\mathbb{C}$ (complex vector spaces). Examples of vector spaces include the n–dimensional Euclidean spaces $\mathbb{R}^n$ and many functional spaces, such as spaces of continuous functions, spaces of measurable functions or spaces of summable functions.

Remark 1.2.4. Again, the notation will be simplified denoting $x +_V y$, $a \cdot_V y$ and 0_V by $x + v$, ay and 0, respectively, when there is no possible confusion.

Definition 1.2.5. Let V be a vector space and let F be either the field of real numbers $\mathbb{R}$ or the field of complex numbers $\mathbb{C}$. An *inner product* is a map $\langle \cdot, \cdot \rangle : V \times V \to F$ that verifies the following properties for all $x, y, z \in V$ and for all $a \in F$:

(i) Conjugate symmetry[4]: $\langle x, y \rangle = \overline{\langle y, x \rangle}$.

(ii) Linearity in the first argument: $\langle x + y, z \rangle = \langle x, z \rangle + \langle y, z \rangle$, and $\langle ax, y \rangle = a \langle x, y \rangle$.

(iii) Positive definiteness: $\langle x, x \rangle \geq 0$ with equality only for $x = 0$.

A vector space V over the field F together with an inner product is called an *inner product space*.

Inner product spaces are the Euclidean space $\mathbb{R}^n$ with the inner product $\langle x, y \rangle = \sum_{i=1}^{n} x_i y_i$, or the space of continuous $\mathbb{C}$–valued functions on the interval $[a, b]$ with the inner product $\langle f, g \rangle = \int_a^b f(t)\overline{g(t)}dt$.

Remark 1.2.6. For $F = \mathbb{R}$, the axioms of conjugate symmetry and linearity in the definition of inner product are reduced to symmetry and bilinearity. Therefore, the inner product is a positive definite symmetric bilinear form. On the other hand, if $F = \mathbb{C}$, the inner product is a positive definite Hermitian form.

---

in A, and f uniquely essociates each pair of elements in A to some element of A.

[4]The *conjugate* of the complex number $z = a + ib$, where a and b are real numbers, is $\overline{z} = a - ib$.

Definition 1.2.7. Let V be a vector space and let F be either the field of real numbers $\mathbb{R}$ or the field of complex numbers $\mathbb{C}$. A *norm* is a map $\|\cdot\| : V \to F$ that verifies the following properties for all $x, y \in V$ and for all $a \in F$:

(i) Triangle inequality: $\|x + y\| \leq \|x\| + \|y\|$.

(ii) Positive homogeneity: $\|ax\| = |a| \, \|x\|$.

(iii) Positive definiteness: $\|x\| \geq 0$, and $\|x\| = 0$ for $x = 0$ only.

If $\|\cdot\|$ verifies the previous properties except non–degeneracy, that is, $\|x\| = 0$ does not preclude that $x \neq 0$, then $\|\cdot\|$ is called *semi–norm*.
Furthermore, a vector space V over the field F together with a norm (respectively, semi–norm) is called a *normed space* (respectively, *semi–norm space*).

Given an inner product space V, a norm can be generated by its inner product $\langle \cdot, \cdot \rangle$ as follows

$$\|x\| = \langle x, x \rangle^{1/2}, \quad \forall x \in V.$$

Thus, V is a normed space with this *induced norm*. Consequently, the examples of inner product spaces commented before are also examples of norm spaces when the norm induced by the inner product is considered.

Definition 1.2.8. A *metric* on a set Z is a map $d(\cdot, \cdot) : Z \times Z \to \mathbb{R}$ that verifies the following properties for all $x, y, z \in Z$:

(i) Triangle inequality: $d(x, y) \leq d(x, z) + d(z, y)$.

(ii) Symmetry: $d(x, y) = d(y, x)$.

(iii) Positive definiteness: $d(x, y) \geq 0$, and $d(x, y) = 0$ for $x = y$ only.

If $d(\cdot, \cdot)$ verifies the previous properties except non–degeneracy, that is, $d(x, y) = 0$ does not preclude that $x \neq y$, then $d(\cdot, \cdot)$ is called *semi–metric*.
Furthermore, a set Z provided with a metric (respectively, semi–metric) is called a *metric space* (respectively, *semi–metric space*).

Remark 1.2.9. The concept of semi–metric included in the previous definition is often referred to as *pseudometric* in the mathematical literature. Nevertheless, the term semi–metric has been chosen in this thesis because of its relation to semi–norm notion and because of that it was used in the FDA monograph by Ferraty and Vieu (2006b).

Definition 1.2.10. A *complete metric space* is a metric space in which every Cauchy sequence[5] is convergent.

Definition 1.2.11. A *Hilbert space* $\mathcal{H}$ is a vector space over $\mathbb{R}$ or $\mathbb{C}$ with an inner product $\langle \cdot, \cdot \rangle$ such that the induced norm defined by $\|x\| = \langle x, x \rangle^{1/2}$, $\forall x \in \mathcal{H}$, turns $\mathcal{H}$ into a complete metric space.

For instance, the complex space ℓ^2, which consists of all infinite sequences of complex numbers $\{z_n\}_{n=1}^{\infty}$

[5] Given a metric space (Z, d), a sequence $\{z_i\}_{i \in \mathbb{N}^*}$ is called *Cauchy sequence* if for any $\varepsilon > 0$ there is a positive integer N_0 such that $d(z_m, z_n) < \varepsilon$, for all $m, n > N_0$.

such that $\sum_{n=1}^{\infty}|z_n|^2$ converges, with the inner product $\langle\{w_n\}_{n=1}^{\infty},\{z_n\}_{n=1}^{\infty}\rangle=\sum_{n=1}^{\infty}w_n\overline{z_n}$ is a Hilbert space. Another example of Hilbert space is the space $L^2([0,1])$ of square–integrable functions with respect to the Lebesgue measure on the unit interval. Then, the inner product of $f,g\in L^2([0,1])$ is defined by $\langle f,g\rangle=\int_0^1 f(t)\overline{g(t)}dt$ in the complex case, and $\langle f,g\rangle=\int_0^1 f(t)g(t)dt$ in the real case.

Definition 1.2.12. A *separable space* is a topological[6] space such that it has a countable[7] dense[8] subset.

From now on, the functional space where X is valued will be a real separable Hilbert space denoted by $(\mathcal{H},\langle\cdot,\cdot\rangle)$, that is, $\mathcal{H}$ is a vector space with an inner product $\langle\cdot,\cdot\rangle:\mathcal{H}\times\mathcal{H}\to\mathbb{R}$ such that is also a complete metric space with respect to the induced norm.

Remark 1.2.13. A Hilbert space is separable if and only if it has a countable orthonormal basis. Hence, there exists a countable sequence $\{e_j\}_{j=1}^{\infty}$ of mutually orthonormal elements of $\mathcal{H}$, that is, $\langle e_{j_1},e_{j_2}\rangle=\delta_{j_1,j_2}$ with $\delta_{j_1,j_2}=0$ if $j_1\neq j_2$ and $\delta_{j_1,j_2}=1$ if $j_1=j_2$, such that span the space. Thus,

$$x=\sum_{j=1}^{\infty}\langle x,e_j\rangle e_j,\quad\forall x\in\mathcal{H},$$

and the induced norm can be given by $\|x\|^2=\sum_{j=1}^{\infty}|\langle x,e_j\rangle|^2$ (*Parseval's identity*).

1.2.2 Associated spaces and tensor products

Given a real separable Hilbert space $(\mathcal{H},\langle\cdot,\cdot\rangle)$, recall that the induced norm has been denoted by $\|\cdot\|=\langle\cdot,\cdot\rangle^{1/2}$. Now, other associated spaces and norms of interest (the space of Hilbert–Schmidt operators and the dual space) and certain tensor notation are going to be presented.

a) The space of Hilbert–Schmidt operators

Let $\mathcal{HS}$ be the *space of Hilbert–Schmidt operators* defined on $\mathcal{H}$ given by

$$\mathcal{HS}=\{U:\mathcal{H}\to\mathcal{H}\text{ such that }U\text{ is a bounded operator, and }\sum_{j=1}^{\infty}\|Ue_j\|^2<\infty,\text{ for all }\{e_j\}_{j=1}^{\infty}\text{ orthonormal basis of }\mathcal{H}\}.$$

For all $U\in\mathcal{HS}$, one can consider the usual Hilbert–Schmidt norm given by

$$\|U\|_{\mathcal{HS}}=\left(\sum_{j=1}^{\infty}\|Ue_j\|^2\right)^{1/2},$$

or the uniform norm

$$\|U\|_{\infty}=\sup_{\|x\|=1}\|Ux\|.$$

Remark 1.2.14. These two norms satisfy the next inequality: $\|U\|_{\infty}\leq\|U\|_{\mathcal{HS}}$ for all $U\in\mathcal{HS}$.

---

[6] A *topological space* is a set Z together with a collection C of subsets of Z (called open sets) such that: the empty set $\emptyset$ is in C, Z is in C, the intersection of a finite number of sets in C is in C, and the union of an arbitrary number of sets in C is also in C.

[7] A *countable set* is a set of the same cardinality as the set of natural numbers $\mathbb{N}$.

[8] A subset W of a topological space Z is *dense* if its closure satisfies $\overline{W}=Z$.

b) The dual space $\mathcal{H}'$

The *dual space* of $\mathcal{H}$ is the space of all continuous linear functions from the space $\mathcal{H}$ into $\mathbb{R}$

$$\mathcal{H}' = \{T : \mathcal{H} \to \mathbb{R} \text{ such that } T \text{ is continuous and linear}\}.$$

The dual space is provided with a natural norm defined by

$$\|T\|_{\mathcal{H}'} = \sup_{\|x\|=1} |Tx|,$$

which verifies that for all $\{e_j\}_{j=1}^{\infty}$ orthonormal basis of $\mathcal{H}$

$$\|T\|_{\mathcal{H}'} = \left(\sum_{j=1}^{\infty} (Te_j)^2\right)^{1/2}.$$

Remark 1.2.15. For all $T \in \mathcal{H}'$ and $U \in \mathcal{HS}$, the following inequality holds $\|TU\|_{\mathcal{H}'} \leq \|T\|_{\mathcal{H}'}\|U\|_{\infty}$.

c) Tensor notation

To finish this section, some tensor product notation that will be used later is introduced. Given $x, y, z \in \mathcal{H}$, the tensor product $x \otimes_{\mathcal{H}} y$ is defined as the following operator

$$\begin{array}{rccl} x \otimes_{\mathcal{H}} y : & \mathcal{H} & \to & \mathcal{H} \\ & z & \to & x \otimes_{\mathcal{H}} y(z) = \langle x, z\rangle y. \end{array}$$

Besides, for $x, z \in \mathcal{H}$ and $y \in \mathbb{R}$, the tensor product $x \otimes_{\mathcal{H}'} y$ is defined as

$$\begin{array}{rccl} x \otimes_{\mathcal{H}'} y : & \mathcal{H} & \to & \mathbb{R} \\ & z & \to & x \otimes_{\mathcal{H}'} y(z) = \langle x, z\rangle y. \end{array}$$

These kinds of tensor products will be useful to define certain functional operators in future sections.

1.2.3 Measuring distances: semi–metrics

Sooner or later, there are two questions that appear when FDA comes to work on. The first difficulty is to decide how to measure the distance between functional elements of $\mathcal{H}$. The second question is related to one of the greatest worries in finite dimensional spaces called the *curse of dimensionality*: the larger the working space dimension is, the larger the sparseness of sample data is. This crucial point for finite dimensional case is even more critical when working with infinite dimensional spaces.

The classic measures of the closeness of two elements in $\mathbb{R}^p$ are norms, for example, the Euclidean norm. The choice of which norm to use is not an issue because of the equivalence among norms in the finite dimensional Euclidean spaces. Nevertheless, in the infinite dimensional spaces the norms are not equivalent so the selection of a preliminary norm turns into a priority problem. Ferraty and Vieu (2006b) considered the notion of norm to be too restrictive in FDA context, and they suggested the use of semi–metrics (recall Definition 1.2.8, page 8), which seem to be more suitable for this type of data (see Geenens (2011) for an analysis of the usefulness of semi–metrics in order to avoid the curse of dimensionality).

In the Hilbert space case, a semi–metric $d(\cdot, \cdot)$ can be easily defined: the metric induced by $\langle \cdot, \cdot \rangle$ is also a semi–metric. However, it could be considered any other semi–metric, not necessarily derived from the inner product of $\mathcal{H}$. Next, some examples of semi–metrics are collected. They were proposed by Ferraty and Vieu (2006b) for the particular case in which the functional data are curves, being $\mathcal{H}$ the classical L^2–space and, consequently, $\langle x, y\rangle = \int_0^1 x(t)y(t)dt$ for all $x, y \in L^2[0, 1]$. Among the proposed semi–metrics, both the semi–metrics based on functional principal components analysis (FPCA) and the semi–metrics based on the multivariate partial least squares regression

(MPLSR) are suitable for rough datasets, whereas the semi–metrics based on derivatives are well adapted for smooth curves. One must choose the semi–metric that has a better behaviour with regard to the data. All the R routines which compute these semi–metrics are available in the website `http://www.math.univ-toulouse.fr/staph/npfda`.

In the examples below, $\{X_i\}_{i=1}^n$ is a sample of i.i.d. curves as the $L^2[0,1]$–valued random functional variable X, and x is a fixed element of $L^2[0,1]$.

Example 1. Semi–metrics based on FPCA

Principal components analysis is a very useful tool for dimension reduction in the multivariate context. In the functional case, FPCA can be used to reduce the dimension and subsequently calculate the distances between the projected data. Due to the importance of FPCA techniques, Section 1.4.3, *"a) Functional principal component analysis (FPCA)"*, in this chapter describes FPCA in great detail (see page 25). Here the required concepts in order to build the proposed semi–metric are presented briefly, using notation which will be rigorously introduced in the above–mentioned section.

Let $\{(\lambda_j, v_j)\}_{j=1}^\infty$ (with $\lambda_1 \geq \lambda_2, \ldots$) be the eigenvalues and eigenfunctions of the second moment operator Γ (for more details, see Section 1.4.1, *"b) Measures of dispersion"*, page 22). The orthonormal basis $\{v_j\}_{j=1}^\infty$ allows to write X as $X = \sum_{j=1}^\infty (\int_0^1 X(t)v_j(t)dt)v_j$, and build truncated expansions as follows

$$X^{[k]} = \sum_{j=1}^{k} \left(\int_0^1 X(t)v_j(t)dt \right) v_j,$$

where the parameter k indicates the "resolution level" which has been applied. Note that $X^{[k]}$ minimizes $\mathbb{E}(\int_0^1 (X(t) - P_k X(t))^2 dt)$ over all the projections P_k of X into k–dimensional spaces. Hence, the following family of semi–metrics based on the L^2–norm is proposed

$$d_k^{PCA}(X_i, x) = \sqrt{\sum_{j=1}^{k} \left(\int_0^1 (X_i(t) - x(t))v_j(t)dt \right)^2}.$$

In practice, the eigenfunctions of Γ are unknown. Therefore, they are replaced by the eigenfunctions $\{\hat{v}_j\}_{j=1}^\infty$ of the empirical second order operator Γ_n (see again Section 1.4.1, *"b) Measures of dispersion"*, page 22). Furthermore, the curves are usually observed in a discrete grid $\{t_l\}_{l=1}^L$ and the integral involved in the semi–metric should be replaced by an approximation. Hence, $d_k^{PCA}(X_i, x)$ will be approximated, if the grid is fine enough, by

$$\tilde{d}_k^{PCA}(X_i, x) = \sqrt{\sum_{j=1}^{k} \left(\sum_{l=1}^{L} w_l (X_i(t_l) - x(t_l))\hat{v}_j(t_l)dt \right)^2},$$

where $\{w_l\}_{l=1}^L$ are quadrature weights for the approximate integration.

Remark 1.2.16. This parameterized family of semi–metrics does not require that functional data verify smoothness conditions. Then these semi–metrics can be applied to quite rough data. However, there are two implicit assumptions: the data must be balanced[9], and the grid of measurements must be sufficiently fine.

Example 2. Semi–metrics based on MPLSR

Sometimes, two variables are simultaneously observed: a response variable and an ifunctional covariate. This fact enables the construction of a family of semi–metrics by means of multivariate regression techniques. The *Multivariate Partial Least Squares Regression* (MPLSR) is a statistical method used when the regression model consists of a multivariate response and a multivariate predictor. The key

[9] A functional dataset is *balanced* if all curves are measured at the same points $\{t_l\}_{l=1}^L$.

of the MPLSR is to obtain a simultaneous decomposition of both the independent and dependent variables in components such that the covariance between the two sets of variables is maximized.

The number of components depends on a certain parameter, called number of factors, which plays a similar role to the dimension k in the PCA: a large number of factors increases the variability, whereas a small number of factors reduces the accuracy. In spite of the similarities, it must be reminded that the PLS components have been designed to explain both the predictor and the response, whereas the PCA components only incorporate information of the independent variables.

Denoting by $\{\hat{v}_j^q\}_{j=1}^p$ the vectors obtained from MPLSR with q the number of factors and p the number of scalar responses, the following family of semi–metrics can be defined

$$\tilde{d}_q^{PLS}(X_i, x) = \sqrt{\sum_{j=1}^{p} \left(\sum_{l=1}^{L} w_j (X_i(t_l) - x(t_l)) \hat{v}_j^q(t_l) \right)^2},$$

where $\{w_l\}_{l=1}^L$ are quadrature weights for the approximate integration.

Remark 1.2.17. These semi–metrics can be applied to balanced data without smoothness restrictions when the grid is fine. Furthermore, note that a multivariate response ($p > 1$) is often an adequate choice, since the semi–metric measures the closeness between curves in terms of a unique direction when $p = 1$, and valuable information contained in the functional dataset may be lost with this drastic dimension reduction.

Example 3. Semi–metrics based on derivatives

When the functional data fulfil some regularity conditions, one can think of measuring the distance between two observations using the distance between their derivatives of a certain order. Hence, the following family of semi–metrics is proposed

$$d_q^{deriv}(X_i, x) = \sqrt{\int_0^1 \left(X_i^{(q)}(t) - x^{(q)}(t) \right)^2 dt},$$

where the superscript (q) denotes the qth derivative. In particular, d_0^{deriv} is the L^2–norm.

In order to avoid the numerical instability of the derivatives computations, the curves are approximated using B–splines. The procedure entails fixing the B–splines basis $\{B_1, \ldots, B_{n_B}\}$ and solving

$$(\hat{\beta}_{i1}, \ldots, \hat{\beta}_{in_B}) = \arg \inf_{(\alpha_1, \ldots, \alpha_{n_B}) \in \mathbb{R}^{n_B}} \sum_{l=1}^{L} \left(X_i(t_l) - \sum_{b=1}^{n_B} \alpha_b B_b(t_l) \right)^2,$$

$$(\hat{\beta}_{01}, \ldots, \hat{\beta}_{0n_B}) = \arg \inf_{(\alpha_1, \ldots, \alpha_{n_B}) \in \mathbb{R}^{n_B}} \sum_{l=1}^{L} \left(x(t_l) - \sum_{b=1}^{n_B} \alpha_b B_b(t_l) \right)^2.$$

In this situation, the observations can be approximated by $\hat{X}_i = \sum_{b=1}^{n_B} \hat{\beta}_{ib} B_b$ and $\hat{x} = \sum_{b=1}^{n_B} \hat{\beta}_{0b} B_b$, respectively, and their derivatives by $\hat{X}_i^{(q)} = \sum_{b=1}^{n_B} \hat{\beta}_{ib} B_b^{(q)}$ and $\hat{x}^{(q)} = \sum_{b=1}^{n_B} \hat{\beta}_{0b} B_b^{(q)}$. Hence, the semi–metric is approximated by

$$\tilde{d}_q^{deriv}(X_i, x) = \sqrt{\int_0^1 \left(\hat{X}_i^{(q)}(t) - \hat{x}^{(q)}(t) \right)^2 dt},$$

where the integral can be computed using, for instance, the Gauss method.

Remark 1.2.18. The semi–metrics based on derivatives should be applied to quite smooth curves. Nevertheless, they have the advantage of tolerating different grids of discretization for each curve since the original unbalanced observations can be replaced by B–splines approximations, which can be computed in a same grid for all the curves.

1.3 Preprocessing functional data

From a practical viewpoint, one can only observe discretized versions of the functional datasets. Hence, before processing the data one usually need to use some smoothing procedure in order to obtain continuous functional data. Another preprocessing which is normally required is the curve registration. These techniques try to suppress some artificial effects in the data as shifts, alignment problems or unbalanced samples.

Next, a brief summary of the main methods devoted to solve this kind of problems when the functional data are curves, that is, when the functional space is $L^2[0,1]$, is presented, although some of them can be extended and applied to other types of functional data.

1.3.1 Smoothing functional data

As it was said before, a functional dataset consists of discretized observations in practice

$$\{X_i(t_l), l = 1, \ldots, L\}_{i=1}^n,$$

being $\{t_l\}_{l=1}^L$ a grid of the unit interval $T = [0,1]$. How can raw discrete data be turned into smooth functional data? This step is given by means of standard nonparametric smoothing procedures, for example, kernel smoothers, basis representations (Fourier basis, splines, wavelets, etc.) or more sophisticated methods which include a priori information on the shape of the functional data. In general, the application of this type of techniques is based on some smoothness conditions (for example, the existence of one or more derivatives).

This presmoothing treatment plays an important role in the presence of missing data, when the discretization grid is not fine enough or when the sample is unbalanced. Moreover, presmoothing is a key tool when measurement errors are involved in the discretized observations, that is, when one has not obtained the original functional data but

$$\{X_i(t_l) + \epsilon_{il}, l = 1, \ldots, L\}_{i=1}^n,$$

being ϵ_{il} the measurement and/or recording error of the data i at the time t_l (see, for instance, Hitchcock et al., 2006; Crambes, 2007; Zhang and Chen, 2007). Next, a general description of two popular smoothing methods (linear smoothing and smoothing by basis functions) is presented. Some notions of smoothing with roughness penalty have also been included.

In the following subsections, $x \in L^2[0,1]$ is a fixed arbitrary curve, and

$$z_l = x(t_l) + \epsilon_l, \text{ for each } l \in \{1, \ldots, L\} \tag{1.1}$$

denote the observed values. Furthermore, bold letter and symbols will be used to denote vectors and matrices.

a) Linear smoothing

An important issue is how to recover the function x starting from the observed L–vector $\mathbf{Z}$ associated to x given by $\mathbf{Z} = (z_1, \ldots, z_L)^t$, where the superscript t denotes the *transposed* vector or matrix. This reconstruction can be done by means of linear smoothers. A *linear smoother* allows to estimate $x(t_{l_0})$ using the following linear combination of the discrete observations

$$\hat{x}(t_{l_0}) = \sum_{l=1}^{L} s_{l_0}(t_l) z_l,$$

where $s_{l_0}(t_l)$ weights the observation z_l for fitting the value of x at the time t_{l_0}. Let $\mathbf{S}$ be the $L \times L$–matrix given by $(\mathbf{S})_{l_1,l_2} = s_{l_1}(t_{l_2})$. Then, one can rewrite the previous linear combination in matrix terms

$$\hat{x}(\mathbf{t}) = \mathbf{SZ},$$

where $\hat{x}(\mathbf{t}) = (\hat{x}(t_1), \ldots, \hat{x}(t_L))^t$.

It seems reasonable to think that the value $x(t)$ must be mostly influenced by the observed z_j corresponding to t_j near t. This assumption leads to consider local weighting functions. One of the simplest linear smoother based on this located weighting principle is the kernel estimator defined by

$$\hat{x}(t) = \sum_{l=1}^{L} s_{l,h}(t) z_l,$$

with $s_{l,h}(t) = K(h^{-1}(t_l - t))$, where $K(\cdot)$ is a *kernel function* and h is called the *smoothing parameter* or *bandwidth*. In the fixed design context (i.e., $\{t_l\}_{l=1}^{L}$ are fixed), another possibility is to use the estimator based on the Gasser and Müller' weights

$$s_{l,h}(t) = \frac{1}{h} \int_{\bar{t}_{l-1}}^{\bar{t}_l} K(h^{-1}(u - t)) du,$$

with $\bar{t}_l = (t_{l+1} + t_l)/2$, $1 < l < L$, $\bar{t}_0 = t_1$ y $\bar{t}_L = t_L$. As far as the random design is concerned (i.e., $\{t_l\}_{l=1}^{L}$ are random), popular kernel estimators are the Nadaraya–Watson estimator, which considers

$$s_{l,h}(t) = \frac{K(h^{-1}(t_l - t))}{\sum_{r=1}^{L} K(h^{-1}(t_r - t))},$$

or the local linear kernel estimator, for which it can be found that the weights are the following

$$s_{l,h}(t) = \frac{K(h^{-1}(t_l - t)) \left(1 - D_1 D_2^{-1}(t_l - t)\right)}{\sum_{r=1}^{L} K(h^{-1}(t_r - t)) \left(1 - D_1 D_2^{-1}(t_r - t)\right)},$$

where $D_1 = \sum_{l=1}^{L} (t_l - t) K(h^{-1}(t_l - t))$ and $D_2 = \sum_{l=1}^{L} (t_l - t)^2 K(h^{-1}(t_l - t))$.

The kernel function $K(\cdot)$ is often a symmetric smooth function, with compact support in $[-1, 1]$ or with a fast decreasing for arguments u such that $|u| \geq 1$. In particular, the following kernel functions are commonly used

Uniform:	$K(u) = 0.5\mathbb{I}_{\{\lvert u\rvert \leq 1\}}$
Triangle:	$K(u) = (1 - \lvert u\rvert)\mathbb{I}_{\{\lvert u\rvert \leq 1\}}$
Quadratic:	$K(u) = 0.75(1 - u^2)\mathbb{I}_{\{\lvert u\rvert \leq 1\}}$
Gaussian:	$K(u) = (2\pi)^{-1/2} \exp(-0.5u^2)$

where $\mathbb{I}$ is the indicator function (see Figure 1.5, page 15). Nevertheless, as it happens in the multivariate case, it seems that the selection of the kernel function is not a key issue, since there are no noticeable changes in the smoothing results when different kernels are used.

The smoothing parameter h regulates the degree of concentration in the vicinity of t. The choice of a suitable h is crucial since the bandwidth determines the bias–variance trade–off: a small h implies small bias and large variance; whereas a large bandwidth reduces the variance but increases the bias. The bandwidth selection is still an open question in the FDA literature: although there are some data–driven techniques (for example, cross–validation methods), none of them have universal validity.

Leaving out the bandwidth selection question, kernel estimators are easy to implement and have nice asymptotic properties. As a drawback of using the Nadaraya–Watson estimator, note that it behaves badly near the limits of the data support (boundary effect). On the contrary, the local linear estimator does not present this boundary effect.

b) Smoothing by basis functions

The second smoothing approach that has been considered is the approximation of x by basis functions. Ramsay and Silverman (2005) defined a basis function system as a set of known functions $\{\phi_j\}_{j=1}^{\infty}$ such that they are mathematical independent of each other, and any function can be approximated arbitrarily well by taking a linear combination of a sufficiently large number J of the functions $\{\phi_j\}_{j=1}^{\infty}$.

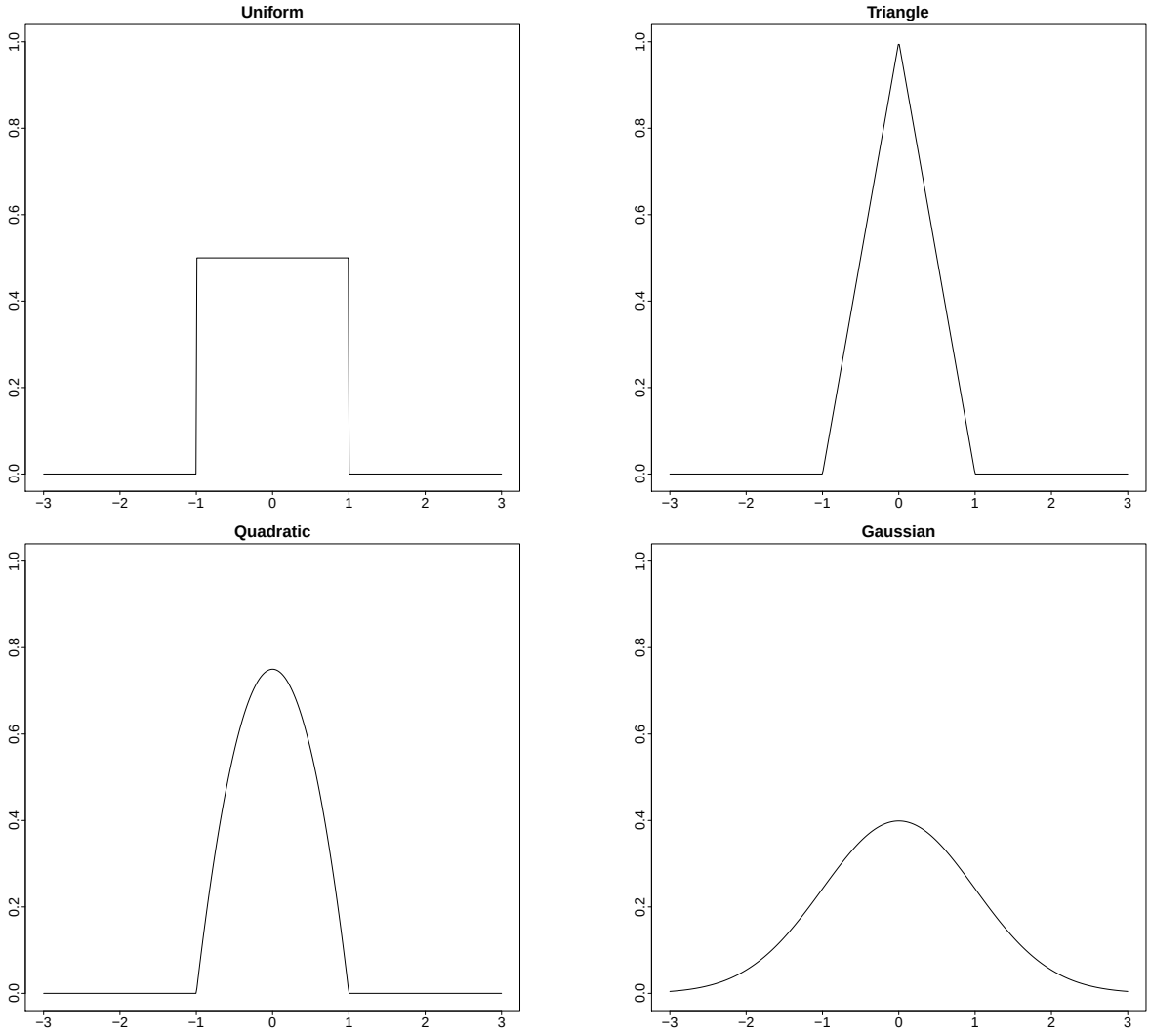

Figure 1.5: Usual symmetric kernels: uniform (top left panel), triangle (top right panel), quadratic (bottom left panel) and Gaussian (bottom right panel).

Hence, a basis system allows to move the main information of x from the functional space to the J–dimensional space spanned by $\{\phi_j\}_{j=1}^J$, since any x can be approximated by the following truncated linear expansion

$$x^{[J]}(t) = \sum_{j=1}^{J} c_j \phi_j(t), \tag{1.2}$$

in other words

$$x^{[J]} = \mathbf{\Phi}^t \mathbf{C}$$

being $\mathbf{C} = (c_1, \ldots, c_J)^t$ the J–vector of the coefficients, and $\mathbf{\Phi} = (\phi_1, \ldots, \phi_J)^t$ the functional J–vector of the basis functions.

Remark 1.3.1. If the qth derivatives of x and the basis functions $\{\phi_j\}_{j=1}^J$ exist for $q = 1, \ldots, Q$, then the qth derivative of x can be approximated by

$$x^{[J](q)}(t) = \sum_{j=1}^{J} c_j \phi_j^{(q)}(t), \text{ for each } q = 1, \ldots, Q,$$

or equivalently, $x^{[J](q)} = (\mathbf{\Phi}^{(q)})^t\mathbf{C}$ with $\mathbf{\Phi}^{(q)} = (\phi_1^{(q)}, \ldots, \phi_J^{(q)})^t$.

At this point, there are three key questions: how to select the basis system, how to estimate the vector of the coefficients $\mathbf{C}$, and how to choose J.

Question 1. How to select the basis system? The choice of an appropriate basis system for the observed data is the first crucial decision that one has to make. There is not any method which allows to select the basis system in an automatic data–driven way, so one must choose it according to the data features: Fourier basis for periodic data, B–splines basis for non–periodic data, wavelets for data with discontinuities or rapid changes in behavior, etc. Below some basis systems which are commonly used in the literature have been summarized (see further details in Ramsay and Silverman, 2005).

(i) Fourier basis system. The Fourier basis with period P is a periodic basis defined as

$$\phi_0(t) = 1, \quad \phi_{2j-1}(t) = \sin(j\omega t), \quad \phi_{2j}(t) = \cos(j\omega t),$$

with $j \in \{1, 2, \ldots\}$ and $\omega = 2\pi/P$. The Fourier basis is orthonormal when the values $\{t_l\}_{l=1}^L$ are equidistant on T, the period coincides with the length of T, and the basis functions are divided by certain constants.
This basis is a very attractive option when the data have a periodic nature. Furthermore, the expansions for the derivatives of x can be easily computed since the derivatives of sines and cosines are well–known. From a computational viewpoint, the Fast Fourier Transform provide a fast and efficient method in order to calculate the Fourier coefficients if the sample size is a power of 2 and $\{t_l\}_{l=1}^L$ are equally spaced. As a disadvantage, it has to be mentioned the bad behaviour of the Fourier approximation when there are discontinuities in x or in its first derivatives.

(ii) Spline basis system. First of all, it must be defined what a spline is. To construct a spline, the interval T has to be divided into S subintervals separated by $\{\tau_s\}_{s=1}^{S-1}$. These values, that define the subintervals, are called *breakpoints* or *knots*. For a fixed q, a *spline* of order q is a polynomial of order q (degree[10] $q-1$) over each subinterval, on condition that each pair of consecutive polynomials, and their derivatives up to order $q-2$, joins up smoothly at the knots. The increment of the number of knots has a direct impact on the flexibility of the spline, which produces an increasingly better approximation to the data. With respect to the knots distribution, one should try to select them so that the following requirements are fulfilled:

(a) each subinterval contains at least one value of the grid $\{t_l\}_{l=1}^L$,

(b) there are more knots in the areas where the curve exhibits abrupt changes, and less knots where its shape is smoother.

Once the spline notion has been stated, the next step is to build a system of basis functions $\{\phi_j\}_{j=1}^{\infty}$ such that satisfies the next conditions:

(1) each ϕ_j is a spline function with order q and knots $\{\tau_s\}_{s=1}^{S-1}$,

(2) any linear combination of basis elements is a spline function,

(3) any spline function with order q and knots $\{\tau_s\}_{s=1}^{S-1}$ can be expressed as a linear combination of $\{\phi_j\}_{j=1}^{\infty}$.

Among all the spline basis systems, one of the most popular is the B–splines basis, though other options are also considered in literature as natural splines or M–splines (see a general review in de Boor, 2001).
In general, the spline basis is chosen to approximate non–periodic functions because it requires a low computational cost and it is extremely flexible and adaptable for almost any type of data.

[10]The *degree* of a polynomial is the highest power defining the polynomial. The *order* of a polynomial is the number of coefficients defining the polynomial, which is one more than its degree.

(iii) Wavelets. Selected a function ϕ called *mother wavelet*, a basis for all square–integrable functions on $(-\infty, +\infty)$ can be built considering, for integers j and k, all translations and dilations of the form

$$\phi_{jk}(t) = 2^{j/2}\phi(2^j t - k).$$

The basis is orthogonal if a suitable mother wavelet is chosen. Moreover, some properties of ϕ are inherited by ϕ_{jk}, for example, if the mother wavelet is compact supported, the support of all basis will also be compact. More details about wavelets can be found in Daubechies (1992). This type of basis fits efficiently discontinuities and sudden increasing or decreasing movements. Furthermore, when x is measured at 2^M regularly spaced points, the coefficients can be calculated fast by means of a Discrete Wavelet Transform.

(iv) Other basis systems. The advantages of the previous basis turn them into the most widely used basis. Nevertheless, other simpler basis systems can produce good results in certain specific situations. For instance, exponential basis ($\phi_j(t) = e^{\lambda_j t}$, where all λ_j are different), power basis ($\phi_j(t) = t^{\lambda_j}$), or polynomial basis ($\phi_j(t) = (t-\omega)^j$, with $j = 0, \ldots, J$) can be adequate systems when the analytic expressions of the curves are linear combinations of exponential terms, power terms or polynomial terms, respectively.

Question 2. How to estimate the coefficients? Once a suitable basis has been selected, one can build the corresponding expansion (1.2) for x (see page 15). Hence, the issue is how to estimate the coefficients c_j starting from the observed pairs $\{(t_l, z_l)\}_{l=1}^L$ (see (1.1), page 13). Three common procedures for estimating the J–vector $\mathbf{C} = (c_1, \ldots, c_J)^t$ are the ordinary or unweighted least squares estimator, the weighted least squares estimator, and the localized least squares estimator. Given that the first approach can be seen as a particular case of the second one, the weighted least squares estimator and the localized least squares estimator are described in detail below, whereas the ordinary least squares estimator is commented as a noteworthy subcase of the weighted least squares estimator.

(i) Weighted least squares estimator. The coefficients c_j can be obtained as the solution of the following weighted least squares problem

$$\min_{c_1,\ldots,c_J} \sum_{l_1=1}^{L}\sum_{l_2=1}^{L} w_{l_1 l_2}\left(z_{l_1} - \sum_{j=1}^{J} c_j\phi_j(t_{l_1})\right)\left(z_{l_2} - \sum_{j=1}^{J} c_j\phi_j(t_{l_2})\right),$$

or in matrix notation

$$\min_{\mathbf{C}} (\mathbf{Z} - \mathbf{\Phi C})^t \mathbf{W} (\mathbf{Z} - \mathbf{\Phi C}), \tag{1.3}$$

being $\mathbf{Z} = (z_1, \ldots, z_L)^t$, $\mathbf{\Phi}$ the $L \times J$–matrix given by $(\mathbf{\Phi})_{l,j} = \phi_j(t_l)$ and $\mathbf{W}$ a symmetric positive definite $L \times L$–matrix. When the variance–covariance matrix $\mathbf{\Sigma}_\epsilon$ for the measurement errors $\{\epsilon_l\}_{l=1}^L$ is known (recall (1.1), page 13), $\mathbf{W} = \mathbf{\Sigma}_\epsilon^{-1}$ is selected; otherwise, one estimates $\hat{\mathbf{\Sigma}}_\epsilon$ (for example, using the sample covariance matrix for the residuals), and takes $\mathbf{W} = \hat{\mathbf{\Sigma}}_\epsilon^{-1}$. In order to simplify the choice of the weighting matrix, it is often assumed that errors are uncorrelated, so $\mathbf{W}$ is a diagonal matrix, and the minimization problem turns into

$$\min_{c_1,\ldots,c_J} \sum_{l=1}^{L} w_{ll}\left(z_l - \sum_{j=1}^{J} c_j\phi_j(t_l)\right)^2. \tag{1.4}$$

In any case, the solution of the weighted least squares problem can be found, whenever suitable smoothness conditions on the basis hold, taking the first derivative in (1.3), and make it equal to zero: $2\mathbf{\Phi}^t\mathbf{W}\mathbf{\Phi}\mathbf{C} - 2\mathbf{\Phi}^t\mathbf{W}\mathbf{Z} = 0$. Hence, one obtains the solution

$$\hat{\mathbf{C}} = (\mathbf{\Phi}^t\mathbf{W}\mathbf{\Phi})^{-1}\mathbf{\Phi}^t\mathbf{W}\mathbf{Z},$$

and, consequently, the fitted values are given by $\hat{\mathbf{Z}} = \mathbf{\Phi}\hat{\mathbf{C}} = \mathbf{\Phi}(\mathbf{\Phi}^t\mathbf{W}\mathbf{\Phi})^{-1}\mathbf{\Phi}^t\mathbf{W}\mathbf{Z}$.

Remark 1.3.2. The weighted least squares estimator can be interpreted as a linear smoother, where the matrix $\mathbf{S}$ (usually called *hat matrix*) is

$$\mathbf{S} = \mathbf{\Phi}(\mathbf{\Phi}^t\mathbf{W}\mathbf{\Phi})^{-1}\mathbf{\Phi}^t\mathbf{W}.$$

(see Section 1.3.1, *"a) Linear smoothing"*, page 13). In this context, the following orthogonal property is satisfied $(\mathbf{Z}-\hat{\mathbf{Z}})^t\mathbf{W}\hat{\mathbf{Z}} = 0$.

Remark 1.3.3. In the simplest case, corresponding to i.i.d. errors $\{\epsilon_l\}_{l=1}^L$ with $\mathbb{E}(\epsilon_l) = 0$ and $\mathrm{Var}(\epsilon_l) = \sigma_\epsilon^2$, $\forall l \in \{1,\ldots,L\}$, the variance–covariance matrix $\mathbf{\Sigma}_\epsilon$ is equal to $\sigma_\epsilon^2 1_{L\times L}$, with $1_{L\times L}$ the $L\times L$–identity matrix (a $L\times L$–matrix with ones in the main diagonal and zeros elsewhere). Then, the least squares problem (1.3) turns into $\min_{\mathbf{C}}(\mathbf{Z}-\mathbf{\Phi C})^t(\mathbf{Z}-\mathbf{\Phi C})$, and the weighted least squares estimator is, in fact, the ordinary least squares estimator. In particular, the coefficients are computed by means of $\hat{\mathbf{C}} = (\mathbf{\Phi}^t\mathbf{\Phi})^{-1}\mathbf{\Phi}^t\mathbf{Z}$, and the fitted values by $\hat{\mathbf{Z}} = \mathbf{\Phi}(\mathbf{\Phi}^t\mathbf{\Phi})^{-1}\mathbf{\Phi}^t\mathbf{Z}$. Note that the ordinary least squares criterion is inadequate when the errors are correlated or heterocedastic. In these situations, the weighted least squares fitting should be used.

(ii) Localized least squares estimator. In order to estimate x at a fixed time t, local weighting functions are often required. The weighted least squares criterion defined in (1.4) can be modified in order to include local weights as follows

$$\min_{c_1,\ldots,c_J} \sum_{l=1}^{L} w_l(t)\left(z_l - \sum_{j=1}^{J} c_j\phi_j(t_l)\right)^2,$$

or equivalently,

$$\min_{\mathbf{C}}(\mathbf{Z}-\mathbf{\Phi C})^t\mathbf{W}(t)(\mathbf{Z}-\mathbf{\Phi C}),$$

with $w_l(t) = K(h^{-1}(t_l - t))$, for $K(\cdot)$ a kernel function and h a strictly positive bandwidth. Futhermore, $\mathbf{W}(t)$ is a diagonal $L\times L$–matrix which contains all the weight functions, that is,

$$\mathbf{W}(t) = \mathrm{diag}(w_1(t),\ldots,w_L(t)).$$

In this situation, one needs to solve the equation $2\mathbf{\Phi}^t\mathbf{W}(t)\mathbf{\Phi C} - 2\mathbf{\Phi}^t\mathbf{W}(t)\mathbf{Z} = 0$. Hence, the estimated vector of coefficients, which depends on t, is computed as

$$\hat{\mathbf{C}}(t) = (\mathbf{\Phi}^t\mathbf{W}(t)\mathbf{\Phi})^{-1}\mathbf{\Phi}^t\mathbf{W}\mathbf{Z},$$

and $\hat{x}(t) = \mathbf{\Phi}\hat{\mathbf{C}}(t) = \mathbf{\Phi}(\mathbf{\Phi}^t\mathbf{W}(t)\mathbf{\Phi})^{-1}\mathbf{\Phi}^t\mathbf{W}(t)\mathbf{Z}$. This approach allows to capture local features of the data suitably, although the need to repeat all the calculations for each new time t may increase the computational cost considerably.

Remark 1.3.4. As in the previous case, the localized least squares estimator can be seen as a linear smoother with local hat matrix

$$\mathbf{S}(t) = \mathbf{\Phi}(\mathbf{\Phi}^t\mathbf{W}(t)\mathbf{\Phi})^{-1}\mathbf{\Phi}^t\mathbf{W}(t).$$

Question 3. How to choose the number of basis functions? The parameter J allows to adjust the "smoothness" of the data. A large J fits better the data, but noise and artificial variability may be introduced in the fitting. On the other hand, if one chooses a small J, perhaps notable data features get lost. Thus, the larger J, the larger the variance; whereas, the smaller J, the larger the bias. In order to balance this trade–off, the key is to find a value for J such that an error criterium is minimized: either a local error criterium, such as the *Mean Square Error* (MSE)

$$\mathrm{MSE}(\hat{x}(t)) = \mathbb{E}((\hat{x}(t)-x(t))^2) = \mathrm{Bias}^2(\hat{x}(t)) + \mathrm{Var}(\hat{x}(t)),$$

or a global error criterium, such as the *Mean Integrated Square Error* (MISE)

$$\text{MISE}(\hat{x}) = \int_0^1 \text{MSE}(\hat{x}(t)).$$

In practice, there are algorithms in the multivariate literature that select the order of the expansion J which can be adapted to FDA field, for example, methods based on stepwise variable selection or variable–pruning procedures.

c) Smoothing with a roughness penalty

The smoothing with roughness penalty is a powerful tool when it comes to approximate discrete data by functional data. Besides the standard advantages of any smoothing method, this technique is especially beneficial when the estimation of derivatives is involved. The roughness penalty approach is based on the minimization of a criterion that ensures certain regularity conditions for the fitted values. Therefore, the goal is to find a curve that gives a good fit to the data, but restricting the search to "smooth" curves. This kind of constraint allows to control simultaneously both bias and variance.

The first issue that arises is how to quantify the roughness of a curve x. The most widespread measures of the roughness are based on the derivatives of x. If the qth derivative of x exists, a natural penalty is

$$\text{PEN}_q(x) = \int_0^1 \left(x^{(q)}(s)\right)^2 ds,$$

where $x^{(q)}$ denotes the qth derivative of x. In particular, the integrated squared second derivative

$$\text{PEN}_2(x) = \int_0^1 \left(x^{(2)}(s)\right)^2 ds$$

is a common choice in many practical situations.

Once the penalty on the roughest curves has been chosen, the next step is the introduction of the penalty into the optimization problem. The aim is to find a curve x that solves the penalized problem

$$\min_x \big((\mathbf{Z} - x(\mathbf{t}))^t \mathbf{W}(\mathbf{Z} - x(\mathbf{t})) + \rho \text{PEN}_q(x)\big),$$

where $\mathbf{Z} = (z_1, \ldots, z_L)^t$ is the observed L–vector, $x(\mathbf{t}) = (x(t_1), \ldots, x(t_L))^t$ is the L–vector of the values of x in the discretization grid, $\mathbf{W}$ is the weighting $L \times L$–matrix, and ρ is a smoothing parameter. The smoothing parameter determines if the priority is to suitable fit the data (smaller values for ρ) or to reduce the variability of the fitted curve (larger values for ρ). In practice, a cross–validation method or a generalized cross–validation method can be used to select an adequate value of ρ (see, for instance, Ramsay and Silverman, 2005, Chapter 5).

Coming back to the basis expansion that were exposed in previous sections (see Remark 1.3.1, page 15), the roughness penalty can be adapted for each fixed x as follows

$$\text{PEN}_q(x) = \int_0^1 \left((\mathbf{\Phi}^{(q)}(s))^t \mathbf{C}\right)^2 ds = \mathbf{C}^t \mathbf{R} \mathbf{C},$$

where $\mathbf{\Phi}^{(q)}(s)$ is the J–vector defined by $\mathbf{\Phi}^{(q)}(s) = (\phi_1^{(q)}(s), \ldots, \phi_J^{(q)}(s))^t$, and $\mathbf{R}$ is the $J \times J$–matrix given by $\mathbf{R} = \int_0^1 \mathbf{\Phi}^{(q)}(s)(\mathbf{\Phi}^{(q)}(s))^t ds$. Then, the penalized least squares problem can be expressed as

$$\min_{\mathbf{C}} \big((\mathbf{Z} - \mathbf{\Phi C})^t \mathbf{W}(\mathbf{Z} - \mathbf{\Phi C}) + \rho \mathbf{C}^t \mathbf{R} \mathbf{C}\big), \tag{1.5}$$

being $\mathbf{\Phi}$ the $L \times J$–matrix given by $(\mathbf{\Phi})_{l,j} = \phi_j(t_l)$. Note that the previous minimization problem is equal to the weighted least squares problem in (1.3) when $\rho = 0$. The solution of the minimization problem (1.5) verifies $2\mathbf{\Phi}^t \mathbf{W} \mathbf{\Phi C} - 2\mathbf{\Phi}^t \mathbf{W Z} + 2\rho \mathbf{R C} = 0$. Hence, the penalized estimator of the vector of coefficients is

$$\hat{\mathbf{C}} = (\mathbf{\Phi}^t \mathbf{W} \mathbf{\Phi} + \rho \mathbf{R})^{-1} \mathbf{\Phi}^t \mathbf{W Z},$$

and the fitted values are $\hat{\mathbf{Z}} = \mathbf{\Phi}\hat{\mathbf{C}} = \mathbf{\Phi}(\mathbf{\Phi}^t \mathbf{W} \mathbf{\Phi} + \rho \mathbf{R})^{-1} \mathbf{\Phi}^t \mathbf{W Z}$.

Remark 1.3.5. The smoothing with roughness penalty is a linear smoothing method too. In this case, the hat matrix is defined as

$$\mathbf{S}_\rho = \mathbf{\Phi}\hat{\mathbf{C}} = \mathbf{\Phi}(\mathbf{\Phi}^t\mathbf{W}\mathbf{\Phi} + \rho\mathbf{R})^{-1}\mathbf{\Phi}^t\mathbf{W},$$

and depends on the smoothing parameter ρ. Note that this hat matrix is a sub–projection operator (it is not a projection because $\mathbf{S}_\rho\mathbf{S}_\rho \neq \mathbf{S}_\rho$).

1.3.2 Registering functional data

Registration is a problem of critical importance for functional data (see Ramsay and Silverman (2005, Chapter 7) or Ramsay (2011)). This preprocessing may be needed when the variation in functional observations involves amplitude and/or phase. Next the most common situations and the procedures used to solve these issues are briefly described.

a) Amplitude variation

The vertical variation or amplitude variation corresponds to the familiar vertical shift: X_{i_1} and X_{i_2} may differ at points $\{t_l\}_{l=1}^L$ at which they are compared, but they exhibit the same global shape features. The amplitude effects are often removed centring and/or rescaling the data.

A classical example of amplitude variation is the spectrometric data, since there is a vertical shift in the curves which is called calibration effect in chemometrics (see left panel of Figure 1.3, page 4). The amplitude variation is an artificial effect that has no link with the studied chemical structure, hence the need to suppress it. In this case, one way for removing the shift is to consider the derivatives of the curves (see right panel of Figure 1.3, page 4)

b) Phase variation

Sometimes, X_{i_1} and X_{i_2} cannot be compared at the same times $\{t_l\}_{l=1}^L$ because they have different behaviours, whereas the comparison can be done if the time scale is previously transformed. This situation happens when the functional data exhibit a phase variation. The most simple phase variation corresponds to a horizontal shift. This often happens when one is interested on a segment of the functional data which is arbitrary located inside the complete recorded observation. The aim is to build a new sample

$$\tilde{X}_i(t) = X_i(t + \delta_i), \quad i = 1, \ldots, n,$$

where δ_i is a shift parameter which enables the functional data alignment. In order to estimate appropriate shift parameters, one has to define a criterion which indicates when two observations are registered. One possibility is to identify a feature or landmark that is a characteristic of the functional data associated with a specific argument value (maxima, minima, zero crossing, etc.). Then, each functional data is shifted so that the selected feature occurs at a fixed point.

Depending on the landmarks, more complicated transformations of the argument t can be required. For example, if F features are selected, for each X_i the arguments $\{t_{if}\}_{f=1}^F$ associated with each landmark have to be identified. Next, one has to look for n transformations $\{w_i\}_{i=1}^n$ such that the registered observations

$$\tilde{X}_i(t) = X_i(w_i(t)), \quad i = l, \ldots, n$$

have similar argument values for the chosen landmarks. In the engineering literature, the transformations $\{w_i\}_{i=1}^n$ are called *time–warping* functions.

The estimation of some common features of the functional data has generated many contributions (Kneip and Gasser, 1988; Kneip and Engel, 1995; Rønn, 2001; Gamboa et al., 2007). Furthermore, both landmark registration and warping techniques have been studied in depth in the literature. For the former, one can look up Kneip and Gasser (1992), Gasser and Kneip (1995), Wang and Gasser (1998), and Liu and Yang (2009). For the latter, some interesting references are Wang and Gasser (1997), Ramsay and Li (1998), Wang and Gasser (1999), and Kneip et al. (2000).

1.4 Exploring functional data

Many methods have been developed, or simply adapted from the existing multivariate methods, for analysing a functional sample. Nevertheless, the impossibility of defining a density notion causes serious analytic difficulties when one explores functional data. The origin of this drawback is the lack of a natural functional measure which carries out the same role than the Lebesgue measure does in the multivariate context. Contributions on the functional density problem can be found in Jacob and Oliveira (1995), Dabo-Niang (2004), Delaigle and Hall (2010) or Dabo-Niang et al. (2010).

Here, a review of the main tools devoted to the exploration of the structure of a functional dataset is presented: measures of position and dispersion, functional data classification and spectral analysis. In this section, $\{X_i\}_{i=1}^n$ is a sample of i.i.d. functional variables as a random functional variable X valued in $\mathcal{H}$, being $(\mathcal{H}, \langle\cdot,\cdot\rangle)$ a real separable Hilbert space. Nevertheless, sometimes the usual functional space $L^2[0,1]$ has been considered in order to introduce some concepts.

1.4.1 Descriptive statistics

a) Measures of position

Mean. In classical real data analysis, the centrality measure par excellence is the mean. Hence, the main issue is to give a notion of mean for functional datasets. When the functional space is $L^2[0,1]$, the first attempt is to define the *functional mean* at each fixed time in $T = [0,1]$ as

$$\mathbb{E}(X)(t) = \mathbb{E}(X(t)), \quad \forall t \in T,$$

and calculate a pointwise average across the sampled elements, that is,

$$\overline{X}(t) = \frac{1}{n}\sum_{i=1}^{n} X_i(t). \tag{1.6}$$

This naive approach has been refined by means of smoothing procedures (Rice and Silverman, 1991; Gervini, 2006; Li and Hsing, 2010; Cai and Yuan, 2011; Bunea et al., 2011), and robust methods such as the trimmed mean where only a certain percentage of the central data are used (Cuesta-Albertos and Fraiman, 2006). Other recent alternative is to replace the traditional cross–sectional mean by the manifold mean such as Chen and Müller (2012) have proposed. Furthermore, inference based on the mean has recently been developed by Horváth et al. (2012) in the functional time series context.
However, all these proposals do not yield a central measure from a purely functional viewpoint. A common way to do this is based on functional depth ideas (Fraiman and Muniz, 2001; Cuevas et al., 2006; Febrero-Bande et al., 2007; López-Pintado and Jornsten, 2007; López-Pintado and Romo, 2007, 2009).

Median and quantiles. An alternative central measure is the *functional median*. Along the years, various median definitions have been introduced for variables valued in infinite dimensional spaces (Kemperman, 1987; Vardi and Zhang, 2000; Cadre, 2001; Ferraty and Vieu, 2006b; Gervini, 2008; Chaouch and Goga, 2012). Moreover, the functional depth can also be used to define and estimate other median notions (Fraiman and Muniz, 2001; Cuevas et al., 2006; López-Pintado and Romo, 2006, 2007; López-Pintado and Jornsten, 2007). Specifically, a functional depth provides a criterion to order a sample of curves from the center–outward, that is, from the deepest curves (curves which attain the maximum value of the functional depth) to the most extreme ones (curves which attain the minimum value of the functional depth). Hence, the functional median will be the deepest curve for a certain functional depth.
Apart from the functional median, *functional quantiles* can be also introduced for functional datasets by means of the functional depth concept and the order that is induced by this depth. More details can be found in Fraiman and Muniz (2001); López-Pintado and Romo (2006, 2007); López-Pintado and Jornsten (2007); Cuevas et al. (2007); Cheng and de Gooijer (2007); Chaouch (2008). A new projection–based definition of quantiles for infinite dimensional Hilbert spaces can be found in Fraiman and Pateiro-López (2012).

Mode. The *functional mode* is another centrality measure, which can be defined and estimated following different approaches as can be seen in Gasser et al. (1998), Hall and Heckman (2002), Cuevas et al. (2006), Dabo-Niang et al. (2006, 2007), Cuevas et al. (2007) or Dabo-Niang et al. (2010). For instance, the concept of functional mode introduced by Cuevas et al. (2006) lies in selecting the curve most "densely surrounded" of the functional dataset $\{x_i\}_{i=1}^n$, in particular, the curve which solves the maximization problem $\max_{i\in\{1,\ldots,n\}} \sum_{j=1}^n K(h^{-1}\|x_j - x_i\|)$, where $K(\cdot)$ is a kernel function and h a bandwidth.

b) Measures of dispersion

For this section, it is necessary to recall the tensor notation that was introduced in Section 1.2.2, *"c) Tensor notation"* (see page 10), since it will be used.

Covariance. To introduce the covariance concept, it is neccesary to assume that the $\mathcal{H}$–valued variable X satisfies $\mathbb{E}(\|X\|^2) < \infty$. In the particular case $\mathcal{H} = L^2[0,1]$, the dependence of records along time can be summarized by means of the *covariance function* given by

$$\mathcal{K}_X(t_1,t_2) = \mathbb{E}((X(t_1) - \mu_X(t_1))(X(t_2) - \mu_X(t_2))) = \mathbb{E}(X(t_1)X(t_2)) - \mu_X(t_1)\mu_X(t_2), \quad \forall t_1,t_2 \in T,$$

where $\mu_X \in \mathcal{H}$ denotes the expected value of X, that is, $\mu_X = \mathbb{E}(X)$. In practice, the covariance function is estimated using its empirical counterpart

$$\begin{aligned}\hat{\mathcal{K}}_X(t_1,t_2) &= \frac{1}{n}\sum_{i=1}^n (X_i(t_1) - \overline{X}(t_1))(X_i(t_2) - \overline{X}(t_2)) \\ &= \frac{1}{n}\sum_{i=1}^n (X_i(t_1)X_i(t_2)) - \overline{X}(t_1)\overline{X}(t_2), \quad \forall t_1,t_2 \in T,\end{aligned}$$

with $\overline{X}$ defined as (1.6) (see page 21).
In the most general case, i.e., when $(\mathcal{H}, \langle\cdot,\cdot\rangle)$ is a real separable Hilbert space, the *covariance operator* Γ_X can be defined as $\Gamma_X = \mathbb{E}((X - \mu_X) \otimes_{\mathcal{H}} (X - \mu_X))$, and therefore

$$\begin{array}{rccl}\Gamma_X: & \mathcal{H} & \to & \mathcal{H} \\ & x & \to & \Gamma_X x = \mathbb{E}(\langle X - \mu_X, x\rangle(X - \mu_X)) = \mathbb{E}(\langle X, x\rangle X) - \langle \mu_X, x\rangle \mu_X.\end{array}$$

Note that Γ_X is just the functional version of the standard multivariate variance–covariance matrix. The covariance operator is a nuclear[11], self–adjoint[12] and positive[13] operator (Dauxois et al., 1982). Its eigenvalues will be denoted by $\{\lambda_j\}_{j=1}^\infty$ (such that $\lambda_1 \geq \lambda_2 \geq \ldots$), and the associated orthonormal eigenfunctions by $\{v_j\}_{j=1}^\infty$. Given a sample $\{X_i\}_{i=1}^n$, the covariance operator Γ_X can be estimated by its empirical counterpart $\Gamma_{X,n} = n^{-1}\sum_{i=1}^n (X_i - \overline{X}) \otimes_{\mathcal{H}} (X_i - \overline{X})$ defined as

$$\begin{array}{rccl}\Gamma_{X,n}: & \mathcal{H} & \to & \mathcal{H} \\ & x & \to & \Gamma_{X,n}x = \frac{1}{n}\sum_{i=1}^n \langle (X_i - \overline{X}), x\rangle (X_i - \overline{X}) = \frac{1}{n}\sum_{i=1}^n \langle X_i, x\rangle X_i - \langle \overline{X}, x\rangle \overline{X}.\end{array}$$

Furthermore, the eigenvalues and eigenfunctions of $\Gamma_{X,n}$ will be denoted by $\{(\hat{\lambda}_j, \hat{v}_j)\}_{j=1}^\infty$, respectively (being $\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \ldots \geq \hat{\lambda}_n \geq 0 = \hat{\lambda}_{n+1} = \ldots$).

[11] An operator U on a Hilbert space $\mathcal{H}$, given by $U : \mathcal{H} \to \mathcal{H}$, is called *nuclear* if it can be written as $U = \sum_{j=1}^\infty u_j f_j \otimes_{\mathcal{H}} g_j$, where $\{f_j\}_{j=1}^\infty$ and $\{g_j\}_{j=1}^\infty$ are orthonormal sets of $\mathcal{H}$, and $\{u_j\}_{j=1}^\infty$ is a set of real numbers such that $\sum_{j=1}^\infty u_j < \infty$.

[12] A *self-adjoint* operator is a linear operator A defined on a linear everywhere–dense set $D(A)$ in a Hilbert space $\mathcal{H}$ such that it coincides with its adjoint operator A^*, that is, such that $D(A) = D(A^*)$ and $\langle Ax, y\rangle = \langle x, Ay\rangle$ for all $x, y \in D(A)$.

[13] A *positive* operator on a Hilbert space $\mathcal{H}$ is a linear operator A such that $\langle Ax, x\rangle \geq 0$.

Remark 1.4.1. The covariance operator will play a key role in FPCA exposed in Section 1.4.3, *"a) Functional principal component analysis (FPCA)"* (see page 25). This fact motivated papers devoted to estimate the covariance operator (Rice and Silverman, 1991; Lee et al., 2002; Gervini, 2006), and some benchmark of its components (Hall and Vial, 2006a).

Remark 1.4.2. Considering the case $\mathcal{H} = L^2([0,1])$, one has

$$(\Gamma_X x)(t) = \mathbb{E}\left(\left(\int_0^1 (X(s) - \mu_X(s))x(s)ds\right)(X(t) - \mu_X(t))\right) = \int_0^1 \mathcal{K}_X(s,t)x(s)ds,$$

that is, the covariance function $\mathcal{K}_X$ is the kernel of the covariance operator Γ_X. Analogously, $\hat{\mathcal{K}}_X$ is the kernel of $\Gamma_{X,n}$

$$(\Gamma_{X,n} x)(t) = \frac{1}{n}\sum_{i=1}^n \left(\left(\int_0^1 (X_i(s) - \overline{X}(s))x(s)ds\right)(X_i(t) - \overline{X}(t))\right) = \int_0^1 \hat{\mathcal{K}}_X(s,t)x(s)ds.$$

Remark 1.4.3. When X is a centred variable, i.e., $\mu_X = 0$, the covariance operator coincides with the second moment operator and becomes in $\Gamma_X = \mathbb{E}(X \otimes_{\mathcal{H}} X)$, thus being $\Gamma_X x = \mathbb{E}(\langle X, x\rangle X)$ for all $x \in \mathcal{H}$. Consequently, the covariance operator can be estimated by the empirical second moment operator $\Gamma_{X,n} = n^{-1}\sum_{i=1}^n X_i \otimes_{\mathcal{H}} X_i$, that is, $\Gamma_{X,n}x = n^{-1}\sum_{i=1}^n \langle X_i, x\rangle X_i$ for all $x \in \mathcal{H}$.

Remark 1.4.4. Whenever there is no possible confusion, Γ_X and $\Gamma_{X,n}$ will be denoted by Γ and Γ_n, respectively, in order to simplify the notation.

Cross–covariance. Let X and Y be two $\mathcal{H}$–valued variables such that $\mathbb{E}(\|X\|^2) < \infty$ and $\mathbb{E}(\|Y\|^2) < \infty$. If one want to analyse the degree of dependence between them when the functional space is $\mathcal{H} = L^2[0,1]$, the *cross–covariance function* defined as

$$\mathcal{K}_{X,Y}(t_1,t_2) = \mathbb{E}((X(t_1) - \mu_X(t_1))(Y(t_2) - \mu_Y(t_2))) = \mathbb{E}(X(t_1)Y(t_2)) - \mu_X(t_1)\mu_Y(t_2), \quad \forall t_1, t_2 \in T,$$

can be used, being μ_X and μ_Y the expected values of X and Y, respectively (i.e., $\mu_X = \mathbb{E}(X)$ and $\mu_Y = \mathbb{E}(Y)$). Starting from a sample $\{(X_i, Y_i)\}_{i=1}^n$ of i.i.d. functional variables distributed as (X, Y), the cross–covariance function can be estimated by

$$\begin{aligned}\hat{\mathcal{K}}_{X,Y}(t_1,t_2) &= \frac{1}{n}\sum_{i=1}^n (X_i(t_1) - \overline{X}(t_1))(Y_i(t_2) - \overline{Y}(t_2)) \\ &= \frac{1}{n}\sum_{i=1}^n (X_i(t_1)Y_i(t_2)) - \overline{X}(t_1)\overline{Y}(t_2), \quad \forall t_1, t_2 \in T,\end{aligned}$$

with $\overline{X}$ and $\overline{Y}$ defined as (1.6) (see page 21).

If $\mathcal{H}$ is a general real separable Hilbert space, besides of the cross–covariance function, one can also compute the *cross–covariance operator* defined by $\Delta_{X,Y} = \mathbb{E}((X - \mu_X) \otimes_{\mathcal{H}} (Y - \mu_Y))$, that is,

$$\begin{array}{rccl}\Delta_{X,Y}: & \mathcal{H} & \to & \mathcal{H} \\ & x & \to & \Delta_{X,Y}x = \mathbb{E}(\langle X - \mu_X, x\rangle(Y - \mu_Y)) = \mathbb{E}(\langle X, x\rangle Y) - \langle \mu_X, x\rangle\mu_Y.\end{array}$$

From a sample $\{(X_i, Y_i)\}_{i=1}^n$ of (X, Y), the cross–covariance operator can be estimated by means of the empirical operator $\Delta_{X,Y,n} = n^{-1}\sum_{i=1}^n (X_i - \overline{X}) \otimes_{\mathcal{H}} (Y_i - \overline{Y})$ given by

$$\begin{array}{rccl}\Delta_{X,Y,n}: & \mathcal{H} & \to & \mathcal{H} \\ & x & \to & \Delta_{X,Y,n}x = \frac{1}{n}\sum_{i=1}^n \langle X_i - \overline{X}, x\rangle(Y_i - \overline{Y}) = \frac{1}{n}\sum_{i=1}^n \langle X_i, x\rangle Y_i - \langle \overline{X}, x\rangle\overline{Y}.\end{array}$$

An especially interesting situation happens when Y is a real random variable. Then, if $\mathcal{H} = L^2([0,1])$, the cross–covariance function and its empirical estimator turn into

$$\begin{aligned}\mathcal{K}_{X,Y}(t) &= \mathbb{E}((X(t) - \mu_X(t))(Y - \mu_Y)) = \mathbb{E}(X(t)Y) - \mu_X(t)\mu_Y, \quad \forall t \in T, \\ \hat{\mathcal{K}}_{X,Y}(t) &= \frac{1}{n}\sum_{i=1}^n (X_i(t) - \overline{X}(t))(Y_i - \overline{Y}) = \frac{1}{n}\sum_{i=1}^n (X_i(t)Y_i) - \overline{X}(t)\overline{Y}, \quad \forall t \in T.\end{aligned}$$

As far as a general real separable Hilbert space $\mathcal{H}$ is concerned, the cross–covariance operator $\Delta_{X,Y}$ turns into $\Delta_{X,Y} = \mathbb{E}((X-\mu_X)\otimes_{\mathcal{H}'}(Y-\mu_Y))$, defined as

$$\begin{array}{rccl} \Delta_{X,Y}: & \mathcal{H} & \to & \mathbb{R} \\ & x & \to & \Delta_{X,Y}x = \mathbb{E}(\langle X-\mu_X, x\rangle(Y-\mu_Y)) = \mathbb{E}(\langle X, x\rangle Y) - \langle \mu_X, x\rangle \mu_Y. \end{array}$$

In this case, the cross–covariance operator can be estimated using its empirical counterpart $\Delta_{X,Y,n} = n^{-1}\sum_{i=1}^{n}(X_i-\overline{X})\otimes_{\mathcal{H}'}(Y_i-\overline{Y})$, that is,

$$\begin{array}{rccl} \Delta_{X,Y,n}: & \mathcal{H} & \to & \mathbb{R} \\ & x & \to & \Delta_{X,Y,n}x = \frac{1}{n}\sum_{i=1}^{n}\langle X_i-\overline{X}, x\rangle(Y_i-\overline{Y}) = \frac{1}{n}\sum_{i=1}^{n}\langle X_i, x\rangle Y_i - \langle \overline{X}, x\rangle \overline{Y}. \end{array}$$

Remark 1.4.5. If $\mathcal{H} = L^2([0,1])$, it is easy to show that

$$(\Delta_{X,Y}x)(t) = \mathbb{E}\left(\left(\int_0^1 (X(s)-\mu_X(s))x(s)ds\right)(Y(t)-\mu_Y(t))\right) = \int_0^1 \mathcal{K}_{X,Y}(s,t)x(s)ds,$$

$$(\Delta_{X,Y,n}x)(t) = \frac{1}{n}\sum_{i=1}^{n}\left(\left(\int_0^1 (X_i(s)-\overline{X}(s))x(s)ds\right)(Y_i(t)-\overline{Y}(t))\right) = \int_0^1 \hat{\mathcal{K}}_{X,Y}(s,t)x(s)ds.$$

Hence, $\mathcal{K}_{X,Y}$ and $\hat{\mathcal{K}}_{X,Y}$ are the kernel functions of $\Delta_{X,Y}$ and $\Delta_{X,Y,n}$, respectively.

Remark 1.4.6. If both X and Y have zero–mean then $\Delta_{X,Y} = \mathbb{E}(X\otimes_{\mathcal{H}}Y)$ with $\Delta_{X,Y}x = \mathbb{E}(\langle X, x\rangle Y)$ for all $x\in\mathcal{H}$. Therefore, the cross–covariance operator can be estimated by $\Delta_{X,Y,n} = n^{-1}\sum_{i=1}^{n} X_i\otimes_{\mathcal{H}} Y_i$ being $\Delta_{X,Y,n}x = n^{-1}\sum_{i=1}^{n}\langle X_i, x\rangle Y_i$ for all $x\in\mathcal{H}$. Furthermore, when $Y\in\mathbb{R}$, the cross–covariance operator and its empirical estimator turn into $\Delta_{X,Y} = \mathbb{E}(X\otimes_{\mathcal{H}'}Y)$ and $\Delta_{X,Y,n} = n^{-1}\sum_{i=1}^{n} X_i\otimes_{\mathcal{H}'} Y_i$.

Remark 1.4.7. In order to abbreviate the notation, the cross–covariance operator and its empirical counterpart will be denoted by Δ and Δ_n when confusion is not possible.

1.4.2 Functional data classification

Other types of exploratory tools are designed to gauge the existence of clusters in the functional random sample $\{X_i\}_{i=1}^n$, and to identify them when they exist. The identification process, which sets the different clusters, is often based on any of the distribution parameters that were presented in the previous section as the mean or the mode. Within this scope, a general classification algorithm usable with any centrality measure was proposed by Ferraty and Vieu (2006b), functional classification focused on modal curve was presented by Dabo-Niang et al. (2006, 2007), and depth–based functional classification tools were presented by López-Pintado and Romo (2006), Cuevas et al. (2007), Cuesta-Albertos and Nieto-Reyes (2010), Li et al. (2012), and Sguera et al. (2012). On the other hand, some contributions were devoted to the adaptation of well–known multivariate k–means procedures to functional classification (see, for instance, Tarpey and Kinateder, 2003; Abraham et al., 2003; Mizuta, 2004; Cuesta-Albertos and Fraiman, 2007; Tarpey, 2007). There were other papers which analysed the functional classification for sparsely observed curves (James and Sugar, 2003; James, 2011), or which proposed alternatives approaches, for example, methods based on neural networks (Rossi et al., 2004), on regression trees (Nerini and Ghattas, 2007), on regression techniques as PLS approach (Preda and Saporta, 2005a; Preda et al., 2007; Delaigle and Hall, 2012), or on similarity measurements such as a rank correlation (Heckman and Zamar, 2000). Furthermore, a recent survey on supervised and unsupervised classification with functional data can be found in Baíllo et al. (2011).

1.4.3 Spectral analysis

From the seminal papers by Rao (1958) and Tucker (1958), factorial analysis has turned into a key tool for exploring functional variables. The underlying idea of any factorial analysis is the spectral decomposition of a functional operator, which is chosen depending on the proposed problem and the considered methodology. Next, a review of the main contributions related to factorial analysis is presented. It is focused on FPCA which will be recalled later to define functional linear regression estimators (see Section 2.3.2 in Chapter 2, page 35, and Chapter 3).

a) Functional principal component analysis (FPCA)

The analysis of the covariance structure of a functional dataset can report valuable information on the features and the behaviour of the data. Since a direct study of the covariance operator is not feasible, FPCA is a fundamental tool for summarizing and visualizing the underlying framework of this operator in a clear and easy way. One of the advantages of FPCA is that the functional data are projected on the space spanned by the eigenelements of the covariance operator. Therefore, a reduction of the problem dimension is possible if one restricts the projection to only the first eigenfunctions. Next, it will be exposed briefly how FPCA works (for more details, see Ramsay and Silverman, 2005, Chapter 8 or Hall, 2011).

Definition of FPCA. Let X be a zero–mean functional variable valued in $\mathcal{H}$. FPCA can be defined by means of the following stepwise procedure:

Step 1. Find the element $\xi_1 \in \mathcal{H}$ that solves

$$\max_{\|\xi_1\|=1} \mathbb{E}(\langle \xi_1, X\rangle^2) = \max_{\|\xi_1\|=1} \mathbb{E}(f_1^2),$$

being $f_1 = \langle \xi_1, X\rangle$ the first principal component score, which gives the projection of X on the direction ξ_1.

Step 2. For $j \in \{2, 3, \ldots\}$, compute $\xi_j \in \mathcal{H}$ such that ξ_j is solution of the next optimization problem

$$\max_{\|\xi_j\|=1} \mathbb{E}(\langle \xi_j, X\rangle^2) = \max_{\|\xi_j\|=1} \mathbb{E}(f_j^2)$$

subject to the $j-1$ constraints

$$\langle \xi_k, \xi_j\rangle = 0, \text{ for all } 1 \leq k < j.$$

In this case, $f_j = \langle \xi_j, X\rangle$ is the jth principal component score, and corresponds to the projection of X on the direction ξ_j.

The directions $\{\xi_j\}_{j=1}^{\infty}$ obtained in the course of the procedure will form an orthonormal basis of $\mathcal{H}$. On each step, the goal is to determine the main source of residual variation in X. In fact, the percentage of variability explained for the first J components can be determined using the next expression

$$\frac{\sum_{j=1}^{J} \mathbb{E}(f_j^2)}{\sum_{j=1}^{\infty} \mathbb{E}(f_j^2)} \times 100. \tag{1.7}$$

Eigenanalysis. Another characterization of FPCA can be derived from the spectral decomposition of the covariance operator Γ (see Section 1.4.1, *"b) Measures of dispersion"*, page 22). The key involves rewriting the maximization problem as follows

$$\begin{cases} \max_{\|\xi_1\|=1} \langle \xi_1, \Gamma\xi_1\rangle, \\ \max_{\|\xi_j\|=1} \langle \xi_j, \Gamma\xi_j\rangle, \text{ subject to } \langle \xi_k, \xi_j\rangle = 0, \text{ for } 1 \leq k < j. \end{cases} \tag{1.8}$$

The previous optimization problem can be solved by considering the eigenfunction problem

$$\Gamma v = \lambda v,$$

that is, the principal components are precisely the orthonormal eigenfunctions $\{v_j\}_{j=1}^{\infty}$ of the covariance operator, arranged from largest eigenvalue to smallest one. Furthermore, the associated eigenvalues $\{\lambda_j\}_{j=1}^{\infty}$ represent the variation in X explained by each component v_j, since $\langle v_j, \Gamma v_j\rangle = \lambda_j$. Hence, the variability (1.7) explained by $\{v_j\}_{j=1}^{J}$ can also be expressed as

$$\frac{\sum_{j=1}^{J} \lambda_j}{\sum_{j=1}^{\infty} \lambda_j} \times 100.$$

Estimation. Let $\{X_i\}_{i=1}^n$ be a centred sample of X. The estimation of FPCA can be done following the next steps:

Step 1. Look for the element $\xi_1 \in \mathcal{H}$ that solves

$$\max_{\|\xi_1\|=1} \frac{1}{n}\sum_{i=1}^n \langle \xi_1, X_i\rangle^2 = \max_{\|\xi_1\|=1} \frac{1}{n}\sum_{i=1}^n f_{i1}^2,$$

where $f_{i1} = \langle \xi_1, X_i\rangle$ for $i \in \{1,\ldots,n\}$. The values $\{f_{i1}\}_{i=1}^n$ are called the first principal component scores.

Step 2. For $j \in \{2,3,\ldots\}$, find $\xi_j \in \mathcal{H}$ that solves

$$\max_{\|\xi_j\|=1} \frac{1}{n}\sum_{i=1}^n \langle \xi_j, X_i\rangle^2 = \max_{\|\xi_j\|=1} \frac{1}{n}\sum_{i=1}^n f_{ij}^2$$

subject to the $j-1$ constraints

$$\langle \xi_k, \xi_j\rangle = 0, \text{ for all } 1 \leq k < j,$$

being $\{f_{ij}\}_{i=1}^n = \{\langle \xi_j, X_i\rangle\}_{i=1}^n$ the jth principal component scores.

The previous procedure is equivalent to find the solution of the eigenequation

$$\Gamma_n v = \lambda v$$

(see Section 1.4.1, *"b) Measures of dispersion"*, page 22). The eigenfunctions $\{\hat{v}_j\}_{j=1}^\infty$ of Γ_n are estimators of the functional principal components of X, and the empirical eigenvalues $\{\hat{\lambda}_j\}_{j=1}^\infty$ of Γ_n estimate the variability of X explained by each component: $\langle \hat{v}_j, \Gamma_n \hat{v}_j\rangle = \hat{\lambda}_j$, for all $j \in \{1,2,\ldots\}$.

Remark 1.4.8. As it happens in the multivariate case, the variation defined by $n^{-1}\sum_{i=1}^n f_{ij}^2$ is reduced step by step, so not all the principal components are computed in practice, but only the first ones. Usually, one obtains the functional principal components, one after another, until a certain percentage P of explained variability is reached, that is, one selects J such that

$$\frac{\sum_{j=1}^J (\frac{1}{n}\sum_{i=1}^n f_{ij}^2)}{\sum_{j=1}^\infty (\frac{1}{n}\sum_{i=1}^n f_{ij}^2)} \times 100 > P,$$

being this expression the empirical version of (1.7). Furthermore, the condition can also be expressed in terms of the empirical eigenelements $\{\hat{\lambda}_j, \hat{v}_j\}_{j=1}^\infty$ as follows

$$\frac{\sum_{j=1}^J \hat{\lambda}_j}{\sum_{j=1}^\infty \hat{\lambda}_j} \times 100 > P.$$

It is well–known that $\{\hat{v}_j\}_{j=1}^\infty$ is an orthonormal basis of $\mathcal{H}$. Remark 1.2.13 (see page 9) implies that $\|X_i\|^2 = \sum_{j=1}^\infty \langle X_i, \hat{v}_j\rangle^2$, so $n^{-1}\sum_{i=1}^n \|X_i\|^2 = \sum_{j=1}^\infty \hat{\lambda}_j$. Hence the denominator in the previous expression could be computed using the empirical estimation of $\mathbb{E}(\|X\|^2)$.

Although only a few statisticians paid attention to FPCA at first (Rao, 1958; Deville, 1974; Dauxois et al., 1982; Ramsay, 1982), the number of contributions devoted to this topic increased considerably from nineties to nowadays (see a general review in Ramsay and Silverman, 2005). Existing literature includes papers concerned with the estimation of the covariance operator (Rice and Silverman, 1991; Lee et al., 2002; Gervini, 2006; Hall and Vial, 2006a), works focused on theoretical aspects (Hall et al., 2006; Hall and Hosseini-Nasab, 2006, 2009) or studies related to computational issues (Ocaña et al., 2007).

On the other hand, interesting contributions have analysed the importance of the scalar product involved in the projection step (Silverman, 1996; Ocaña et al., 1999; Locantore et al., 1999) or the effect of introducing smoothing techniques in the standard FPCA such as splines ideas (Ramsay and Dalzell, 1991; Pezzulli and Silverman, 1993; Silverman, 1996; Cardot, 2000; James et al., 2000; Yao and Lee, 2006; Zhou et al., 2008), local linear smoothing (Li and Hsing, 2010) or kernel–type approaches (Boente and Fraiman, 2000). In particular, the methods presented in two of these papers (specifically, Pezzulli and Silverman, 1993, and Silverman, 1996), which can be seen as a combination of the standard FPCA and smoothing techniques, are going to be described briefly. These two approaches will be recalled in Section 3.4 in Chapter 3 (see page 60).

(i) Pezzulli and Silverman' smoothed FPCA. A roughness penalty was introduced in the optimization problem (1.8) by Pezzulli and Silverman (1993), who considered the following stepwise maximization process

$$\begin{cases} \max_{\|\xi_1\|=1} \left(\langle \xi_1, \Gamma\xi_1\rangle - \alpha\langle \xi_1, Q\xi_1\rangle\right), \\ \max_{\|\xi_j\|=1} \left(\langle \xi_j, \Gamma\xi_j\rangle - \alpha\langle \xi_j, Q\xi_j\rangle\right), \text{ subject to } \langle \xi_k, \xi_j\rangle = 0, \text{ for } 1 \leq k < j, \end{cases}$$

where α is a positive smoothing parameter, and Q is a symmetric nonnegative operator that quantifies the *smoothness* of solutions (for instance, an usual choice for Q is the fourth derivative operator).
In practice, one needs to find the pairs $\{(\hat{\lambda}_j^{\alpha,1}, \hat{v}_j^{\alpha,1})\}_{j=1}^{\infty}$, which solve the generalized eigenproblem

$$(\Gamma_n - \alpha Q)v = \lambda v.$$

Each eigenfunction $\hat{v}_j^{\alpha,1}$ estimates the jth functional principal component, and its contribution to the explained variability is determined by the associated eigenvalue $\hat{\lambda}_j^{\alpha,1}$.

(ii) Silverman's smoothed FPCA. In the smoothed FPCA proposed by Silverman (1996), the roughness penalty is not inserted into the maximization problem, but into the inner product definition. For a given smoothing parameter α, a new inner product is defined as

$$\langle x, y\rangle_\alpha = \langle x, y\rangle + \alpha\langle x, Qy\rangle,$$

being Q a symmetric nonnegative operator that controls the smoothness restrictions. This inner product allows to build an induced norm by means of $\|x\|_\alpha = \langle x, x\rangle_\alpha^{1/2}$.
Once the smoothed inner product is stated, Silverman (1996) proposed to solve

$$\begin{cases} \max_{\|\xi_1\|_\alpha=1} \langle \xi_1, \Gamma\xi_1\rangle, \\ \max_{\|\xi_j\|_\alpha=1} \langle \xi_j, \Gamma\xi_j\rangle, \text{ subject to } \langle \xi_k, \xi_j\rangle_\alpha = 0, \text{ for } 1 \leq k < j. \end{cases}$$

It is important to highlight that the smoothness requirements only affect the orthonormality constraints. In this situation, the estimation procedure leads to the solution $\{(\hat{\lambda}_j^{\alpha,2}, \hat{v}_j^{\alpha,2})\}_{j=1}^{\infty}$ of the eigenproblem

$$\Gamma_n v = \lambda(1_{\mathcal{H}} + \alpha Q)v,$$

with $1_{\mathcal{H}}$ the identity operator on $\mathcal{H}$, i.e., $1_{\mathcal{H}}x = x$ for all $x \in \mathcal{H}$. Hence, the obtained eigenelements allow to estimate the functional principal components, and the variability of X which is explained using each of them.

Finally, note that FPCA has been successfully extended to a very wide range of situations: time series (Bosq, 1991; Aguilera et al., 1997, 1999; Bosq, 2000), bi–functional data (Spitzner et al., 2003), longitudinal data (James, 2002; Yao et al., 2005a), or conditional frameworks (Cardot, 2007).

Remark 1.4.9. Alternatives to functional principal components can be found in Park et al. (2009), who tried to construct more informative structural components, or in Chen and Müller (2012), who proposed the use of functional manifold components.

b) Functional canonical analysis

When the sample involves two functional variables (X, Y) and the goal is to determine the relationship between them, the canonical analysis is a reasonable statistical methodology to be borne in mind. First contribution on functional canonical analysis were presented by Dauxois and Pousse (1975), and more recent advances can be found in He et al. (2000, 2003, 2004); Dauxois et al. (2004); Gou and Fyfe (2004), Ramsay and Silverman (2005, Chapter 11), He et al. (2010).

c) Functional linear discriminant analysis

Functional linear discriminant analysis is also concerned with a pair of variables (X, Y) but, unlike canonical analysis, Y must be valued in a finite set in this case. Linear discriminant analysis has been combined with other kind of techniques as penalization procedures, dimensionality reduction or partial least squares approaches. Some interesting contributions in this field are the following: Hastie et al. (1995); Ramsay and Silverman (1997); James and Hastie (2001); Ramsay and Silverman (2005); Preda et al. (2007); Shin (2008); and Delaigle and Hall (2012).

Chapter 2

Functional regression models

After introducing the functional regression concept, this chapter focuses on the functional regression model for scalar response. Two approaches to this problem are presented in this chapter: functional linear regression (parametric approach) and kernel–type regression (nonparametric approach). For the former estimators based on basis expansions (Ramsay and Silverman, 2005) and FPCA–type estimators (Cardot et al., 1999, 2003c, 2007c; Ferraty et al., 2012a) are described, whereas an extension of the multivariate Nadaraya–Watson estimator (Ferraty and Vieu, 2004, 2006b; Ferraty et al., 2007a) is considered for the latter. A final appendix compiles proofs of results and technical lemmas required for the chapter. The state of the art of the functional regression approaches included in the first sections of the chapter was partly updated by means of the contributions by González-Manteiga and Vieu (2011) and Cuevas (2012).

The conditional errors for the standard FPCA estimator (see Theorem 2.3.14, page 39) were published in the paper by Ferraty et al. (2012a).

2.1 What does functional regression mean?

One is often interested in studying the behaviour of a random variable Y (response or dependent variable) taking into account that part of its variability could be explained by means of a random variable X (covariate or independent variable). In this case, one considers the following *regression model*

$$Y = m(X) + \epsilon,$$

where $m(\cdot)$ is a *regression function*, and ϵ is a random variable called *error*. Table 2.1 compiles the four possible situations that one can find according to the nature of the response and the covariate, that is, depending on whether the concerned variables are valued in a finite dimensional space (for example, $\mathbb{R}^p$ with $p \geq 1$) or they belong to a general functional space (for example, the Hilbert space $\mathcal{H}$). Case A is the classic multivariate regression model, which has often been the subject of study in the statistical literature. The remaining cases correspond to *functional regression models* since at least one of the involved variables is a functional random variable.

	$Y \in \mathbb{R}^p$	$Y \in \mathcal{H}$
$X \in \mathbb{R}^p$	Case A. $m : \mathbb{R}^p \to \mathbb{R}^p$	Case B. $m : \mathbb{R}^p \to \mathcal{H}$
$X \in \mathcal{H}$	Case C. $m : \mathcal{H} \to \mathbb{R}^p$	Case D. $m : \mathcal{H} \to \mathcal{H}$

Table 2.1: Regression models depending on the nature of X and Y.

Case B. For this case, when the dependent variable is functional and the independent variable is multivariate, one can see some interesting contributions in the following works: Faraway (1997); Brumback and Rice (1998); Yu and Lambert (1999); Guo (2002); Chiou et al. (2003a,b, 2004); Ramsay and Silverman (2005, Chapter 13); Cardot et al. (2007d); Hlubinka and Prchal (2007); Nerini and Ghattas (2007); Zhang and Chen (2007); Bücher et al. (2011); and Zhang (2011).

Case C. The prediction of scalar response ($p = 1$) from functional covariate was discussed in recent papers both from a parametric (Ramsay and Silverman, 2005) and from a nonparametric approaches (Ferraty and Vieu, 2006b). For instance, a comparison of two parametric estimators and a nonparametric one can be found in Baíllo (2009). Since the following sections are especially devoted to Case C, references related to this regression model can be seen in detail below.

Case D. First advances for dealing with this case, where both X and Y are functional variables, can be found in Ramsay and Dalzell (1991) and Hastie and Tibshirani (1993). There are some more recent contributions focused on the linear case (Hoover et al., 1998; Wu et al., 1998; Cuevas et al., 2002; Malfait and Ramsay, 2003; Fan and Zhang, 2004; Kneip et al., 2004; Ramsay and Silverman, 2005, Chapter 14 and 16; Müller and Zhang, 2005; Yao et al., 2005b; Prchal and Sarda, 2007; Kokoszka et al., 2008; Antoch et al., 2008; Wu et al., 2010; Antoch et al., 2010; Müller et al., 2011; Kim et al., 2011; Horváth and Reeder, 2012). For nonparametric estimator, see for instance the kernel regression studied by Ferraty et al. (2011a) and Ferraty et al. (2012d).

An important functional setting deals with the multilevel functional data and the functional ANOVA. Some advances in this field can be found in Spitzner et al. (2003); Morris et al. (2003); Abramovich et al. (2004); Cuevas et al. (2004); Abramovich and Angelini (2006); Morris and Carroll (2006); Antoniadis and Sapatinas (2007); Zhang and Chen (2007); Schott (2007); Di et al. (2009); Cuesta-Albertos and Febrero-Bande (2010); Staicu et al. (2010); González-Rodríguez et al. (2012).

Discrimination of functional data can be analysed as a particular regression case, where the response variable is valued in a finite set and denotes the group each observation belongs to. This question can be discussed both from a linear viewpoint (see Section 1.4.3, *"c) Functional linear discriminant analysis"*, in Chapter 1, page 28), and from a nonparametric viewpoint (see Ferraty and Vieu, 2004 or Ferraty and Vieu, 2006b, Chapter 8). For the latter, some authors developed specific tools based on projection methods (Hall et al., 2001), on kernel methods (Ferraty and Vieu, 2003), on k–NN methods (Biau et al., 2005; Ferraty and Vieu, 2006b; Cérou and Guyader, 2006; Abraham et al., 2006), on sliced inverse regression (Ferré and Yao, 2005), on SVM ideas (Rossi and Villa, 2006), or on PLS approaches (Costanzo et al., 2006; Preda et al., 2007).

Another special situation that requires a distinct treatment is the functional time series analysis. In this case, the functional covariate is the past continuous path of the process, and the response is a characteristic of the future of the process. This problem needs a suitable regression approach in order to take into account the dependence between the covariate and the response (see, for instance, Bosq, 2000, Ferraty and Vieu, 2006b, Part IV, Bosq and Blanke, 2007, Delsol, 2011, or Mas and Pumo, 2011). Some contributions which analysed this issue when the response variable is scalar are the following ones: Ferraty et al. (2002); Ait-Saïdi et al. (2005); Ferraty et al. (2005); Masry (2005); Ezzahrioui and Ould-Saïd (2008b); Aneiros-Pérez and Vieu (2008); Delsol (2009); and Aneiros-Pérez et al. (2011).

If the response variable is functional, then autoregressive models are the key to predict the future path (Bosq, 1991, 2000). This approach resulted in theoretical advances (Merlevède et al., 1997; Pumo, 1998; Guillas, 2001, 2002; Mas, 2002; Mourid, 2002; Bosq, 2002, 2003; Antoniadis and Sapatinas, 2003; Mas and Menneteau, 2003; Damon and Guillas, 2005; Menneteau, 2005; Mas, 2007b; Bosq, 2007; Mas and Pumo, 2007; Ruiz-Medina et al., 2007; Horváth et al., 2010; Ruiz-Medina and Salmerón, 2010; Ruiz-Medina, 2011, 2012b; Horváth and Kokoszka, 2012, Chapter 13), and applied contributions (Besse and Cardot, 1996; Besse et al., 2000; Bouzas et al., 2002; Valderrama et al., 2002; Damon and Guillas, 2002; Valderrama et al., 2003; Bouzas et al., 2006; Laukaitis, 2008; Antoch et al., 2010; Ruiz-Medina and Espejo, 2012). However, to describe nonlinear situations other kind of models are required, such

as autoregressive conditional heteroscedastic models (ARCH models) which are commonly used in econometrics. A functional version of ARCH model can be found in Hörmann et al. (2012).

Furthermore, some recent papers tried to generalize time dependence to spatial dependence (Delicado et al., 2010; Guillas and Lai, 2010; Giraldo et al., 2010a,b) or to define new moment–based notions of dependence (Hörmann and Kokoszka, 2010). A general overview of the main contributions (and also future research lines) to spatial functional statistics can be found in Ruiz-Medina (2012a).

Remark 2.1.1. In order to reduce the dimension of the functional covariates in the regression context, the variable selection by means of techniques such as the LASSO method can be an interesting option. Some recent contributions related with this research line are Kneip and Sarda (2011) and Lee and Park (2012).

From now on, the chapter is focused on the Case C with $p = 1$, which corresponds to the regression model with functional covariate and scalar response.

2.2 Functional regression for scalar response

If the functional regression model with scalar response is considered (i.e., Case C with $p = 1$ in Table 2.1, page 29), one could assume that the regression operator $m : \mathcal{H} \to \mathbb{R}$ belongs to a specific parametric family, one could just impose certain smoothness conditions on $m(\cdot)$, or one could even combine these two approaches. To fix ideas, the well–known parametric, nonparametric and semiparametric notions in the multivariate context will be extended to the functional regression field, such as it was proposed in Ferraty and Vieu (2006b, 2011b).

Definition 2.2.1. A *model* for the estimation of the regression operator $m(\cdot)$ consists in introducing some constraint of the form $m \in C$. The model is called *parametric* when C is indexed by N_C elements of $\mathcal{H}$ ($N_C < +\infty$), that is,

$$C = \{C_{(\theta_1,\ldots,\theta_{N_C})} | (\theta_1, \ldots, \theta_{N_C}) \in \mathcal{H} \times \overset{N_C}{\ldots} \times \mathcal{H}\}.$$

The model is called *nonparametric* when C cannot be indexed by a finite number of parameters. The model is called *semiparametric* when it consists of both parametric and nonparametric components.

A clear advantage of parametric models is the simplification of the estimation process, since the estimation of the functional operator is turned into the estimation of a finite number of parameters, which can be easily interpreted. Nevertheless, this approach may provide misleading estimates if the regression operator does not belong to the selected parametric family. Hence, nonparametric estimation is a better option when one has no previous information about $m(\cdot)$ despite the increase of computational complexity. The trade–off between parametric and nonparametric models is balanced by means of semiparametric models, which try to combine the simplicity and interpretability of parametric approach and nonparametric flexibility. Next, a brief summary of references related with these three families of regression models is presented.

Parametric models. The most usual parametric model is the functional linear model that was studied at first by Ramsay and Dalzell (1991) and Hastie and Mallows (1993). Later, several research lines were developed in order to estimate the functional parameter involved in the model. Among the published contributions, the methods based on smoothing splines (Ramsay and Silverman, 2005; Cardot et al., 2003c, 2007a; Crambes et al., 2009), on FPCA (Cardot et al., 1999; Cai and Hall, 2006; Hall and Hosseini-Nasab, 2006; Crambes, 2007), on presmoothing ideas (Martínez-Calvo, 2008; Ferraty et al., 2012a), or on PLS techniques (Preda and Saporta, 2005a,b; Escabias et al., 2007; Reiss and Ogden, 2007) should be mentioned. For each estimator, an outstanding question is to establish the rates of convergence. Some efforts focused on this aim can be found in Cai and Hall (2006), Hall and Horowitz (2007), Cardot et al. (2007c) or Crambes et al. (2009). See also some advances on point

impact linear regression in McKeague and Sen (2010) and some advances on detection of influential observations in Febrero-Bande et al. (2010). Furthermore, an interesting issue was studied by Wang et al. (2012): how to transform (X, Y) when the linear assumption is not satisfied in order to obtain linearly related transformed data. Other problems on functional linear regression which also were the subject of recent research can be found in Ramsay and Silverman (2005), Besse et al. (2005), and Cardot and Sarda (2006, 2011).
Besides the linear regression, there are more complex regression models that were analysed recently, such as the generalized functional linear model (James, 2002; Cardot and Sarda, 2005; Müller and Stadtmüller, 2005; Li et al., 2010), or the special case of the functional logistic regression (Ratcliffe et al., 2002b; Escabias et al., 2004, 2005; Aguilera et al., 2006; Lindquist and McKeague, 2009), and the functional multilogit model (Besse et al., 2005).

Nonparametric models. The use of nonparametric models in the functional context started with the monograph by Ferraty and Vieu (2006b). The most popular nonparametric methods are based on the classical Nadaraya–Watson estimator, and were developed by Ferraty and Vieu (2002, 2004, 2006b, 2011a). Other works contributed complementary theoretical results (Masry, 2005; Delsol, 2007; Ferraty et al., 2007a; Delsol, 2009; Aspirot et al., 2009; Ferraty et al., 2010b,c), applied issues linked with the smoothing parameter selection (Benhenni et al., 2007; Rachdi and Vieu, 2007), or bootstrap techniques for the nonparametric functional regression context (Ferraty et al., 2010c, 2012d).
Apart from the standard kernel estimator, alternative nonparametric estimators were proposed, such as a k–NN estimator (Burba et al., 2009; Biau et al., 2010) and a local linear estimator (Boj et al., 2008; Baíllo and Grané, 2008, 2009; Barrientos-Marin et al., 2010; Boj et al., 2010). Other approaches which can be found in the literature are nonparametric methods based on neural networks (Rossi et al., 2005; Rossi and Conan-Guez, 2005, 2006), and methods based on reproducing kernel Hilbert spaces ideas (Preda, 2007).

Semiparametric models. Recently, the semiparametric models have started emerging in the functional regression context. As mentioned before, these kinds of models consist of a parametric term and a nonparametric one, which allow balancing the advantages and disadvantages of both components. Although semiparametric regression is still starting out, some contributions were already presented in order to give functional versions of some classic semiparametric models, such as additive models (Aneiros-Pérez et al., 2004; Ferraty and Vieu, 2009), single index models (Ait-Saïdi et al., 2005; Amato et al., 2006; Ait-Saïdi et al., 2008), partial linear models (Aneiros-Pérez and Vieu, 2006, 2008; Shin, 2009; Lian, 2011; Vilar et al., 2012; Aneiros-Pérez and Vieu, 2012), varying–coefficient functional models (Cai, 2011), and sliced inverse regression (Ferré and Yao, 2005; Ferré and Villa, 2006).
In all the previous references, the proposed techniques are based on conditional expectation, that is, the authors considered the regression operator given by $m(X) = \mathbb{E}(Y|X)$. Nevertheless, other characteristics of the conditional distribution can be studied, for instance, the conditional quantiles (Cardot et al., 2005; Ferraty et al., 2005, 2006; Cardot et al., 2007b; Ezzahrioui and Ould-Saïd, 2008b) or the conditional mode (Ferraty et al., 2006; Ezzahrioui and Ould-Saïd, 2008a). There were also some contributions to functional conditional hazard estimation (Ferraty et al., 2008; Quintela del Río, 2008), and robust regression estimation (Crambes et al., 2008).

The following sections are devoted to functional linear regression and functional kernel regression. For the former, estimators based on basis expansions and estimators based on FPCA are going to be analysed. For the latter, an adaptation of the multivariate Nadaraya–Watson estimator is considered.

2.3 Functional linear regression for scalar response

Recall that $(\mathcal{H}, \langle\cdot,\cdot\rangle)$ is a real separable Hilbert space and $\|\cdot\|$ denotes its induced norm. Then, the *functional linear model with scalar response* can be expressed as

$$Y = \langle\theta, X\rangle + \epsilon, \tag{2.1}$$

where Y is a real random variable, $m(\cdot) = \langle \theta, \cdot \rangle$ is a linear regression operator with $\theta \in \mathcal{H}$ and $\|\theta\|^2 < \infty$, X is a zero–mean random variable valued in $\mathcal{H}$ such that $\mathbb{E}(\|X\|^2) < \infty$, and ϵ is a real random variable satisfying that $\mathbb{E}(\epsilon) = 0$, $\mathrm{Var}(\epsilon) = \sigma^2 < \infty$, and $\mathbb{E}(\epsilon X) = 0$.

Remark 2.3.1. If the variables X and Y are not centred, it can be defined

$$\tilde{Y} = Y - \mathbb{E}(Y) \quad \text{and} \quad \tilde{X} = X - \mathbb{E}(X).$$

It can be deduced that the regression model $\tilde{Y} = \langle \theta, \tilde{X} \rangle + \epsilon$ is equivalent to

$$Y = \theta_0 + \langle \theta, X \rangle + \epsilon,$$

with $\theta_0 = \mathbb{E}(Y) - \langle \theta, \mathbb{E}(X) \rangle$. Hence, one can just focus on the model (2.1) with zero–mean variables, without loss of generality.

Remark 2.3.2. Although the broad framework of Hilbert spaces has been fixed, functional data are often valued in well–known common spaces. For example, the space $\mathcal{H} = L^2([0,1])$ is usually considered when the functional data are curves, that is, when $X = \{X(t), t \in [0,1]\}$. In this situation, the general regression model (2.1) becomes

$$Y = \int_0^1 \theta(t) X(t) dt + \epsilon,$$

when one takes the inner product $\langle x, y \rangle = \int_0^1 x(t) y(t) dt$ for all $x, y \in L^2[0,1]$.

The problem that comes up at this point is how to estimate the model parameter θ by means of an i.i.d. sample $\{(X_i, Y_i)\}_{i=1}^n$ from (X, Y). In practice, one cannot measure the continuous functional data $\{X_i\}_{i=1}^n$, but a dataset which consists of their discretized versions. Therefore, the sample is none other than

$$\{(X_i(t_1), \ldots, X_i(t_l), \ldots, X_i(t_L), Y_i)\}_{i=1}^n,$$

being $\{t_l\}_{l=1}^L$ the discrete grid where each X_i is recorded. As a first naive attempt to estimate θ, one could forget the functional nature of X, and consider each discretized X_i as a multidimensional covariate. Hence the problem comes down to estimating the ordinary multivariate linear regression

$$Y_i = \sum_{l=1}^{L} \theta(t_l) X_i(t_l) + \epsilon_i, \quad \text{for } i = 1, \ldots, n.$$

This simple approach has clear drawbacks, such as the need to solve a system of n equations with L unknowns, being almost always $L >> n$, and the existence of infinite solutions which fit the observed data perfectly.

In order to obtain an identifiable solution which can be interpreted, one could use some method of regularization (see preprocessing techniques in Section 1.3 in Chapter 1, page 13). On the other hand, one of the most widely used techniques for estimating θ are the methods based on basis systems as Fourier series, wavelets or splines (see a general review in Ramsay and Silverman, 2005, and see Cardot et al., 2003c or Crambes et al., 2009 for the popular splines approach). The main ideas linked with this methodology has been summarized in Section 2.3.1. Another widespread methodology is based on FPCA, and it has been developed and analysed during the last years (Cardot et al., 1999, 2003c; Cai and Hall, 2006; Hall and Hosseini-Nasab, 2006; Hall and Horowitz, 2007; Cardot et al., 2007c). This alternative estimator is presented in Section 2.3.2.

2.3.1 Estimators based on basis expansions

Let $\{\phi_j\}_j^J$ be a basis function system in the functional space $L^2[0,1]$. As explained in Section 1.3.1, *"b) Smoothing by basis functions"*, in Chapter 1 (see page 14), a basis system allows to project the main information of the data onto the finite dimensional space spanned by the first basis elements. Therefore, the model parameter θ can be expressed as a linear combination of the basis elements

$$\theta^{[J_\theta]} = \sum_{j=1}^{J_\theta} b_j \phi_j = \mathbf{\Phi}^t \mathbf{B},$$

with $\mathbf{B} = (b_1, \ldots, b_{J_\theta})^t$ the J_θ–vector of the coefficients, and $\mathbf{\Phi} = (\phi_1, \ldots, \phi_{J_\theta})^t$ the functional J_θ–vector of the basis functions. The choice of J_θ is a key issue, since one needs a value large enough to avoid the loss of information, but small enough to make the interpretation of the model parameter easier.

One can repeat the procedure for the functional covariates by means of another basis function system $\{\psi_j\}_j^\infty$ (with $\{\psi_j\}_{j=1}^\infty$ not necessarily equal to $\{\phi_j\}_{j=1}^\infty$). Then, one gets

$$X_i^{[J_X]} = \sum_{j=1}^{J_X} c_{ij}\psi_j = \mathbf{\Psi}^t\mathbf{C}_i, \quad \text{for } i = 1, \ldots, n,$$

where $\mathbf{C}_i = (c_{i1}, \ldots, c_{iJ_X})^t$ is the J_X–vector of the coefficients for X_i, and $\mathbf{\Psi} = (\psi_1, \ldots, \psi_{J_X})^t$ is the functional J_X–vector of the basis functions.

Consequently, the regression model (2.1) (see page 32) could be approximated by

$$Y_i = \langle \theta^{[J_\theta]}, X_i^{[J_X]} \rangle + \epsilon_i = \sum_{j_1=1}^{J_\theta}\sum_{j_2=1}^{J_X} b_{j_1}c_{ij_2}\langle \phi_{j_1}, \psi_{j_2}\rangle + \epsilon_i = \mathbf{C}_i^t\mathbf{M}_{\Psi\Phi}\mathbf{B} + \epsilon_i, \quad \text{for } i = 1, \ldots, n$$

where $\mathbf{M}_{\Psi\Phi}$ is the $J_X \times J_\theta$–matrix given by $(\mathbf{M}_{\Psi\Phi})_{j_2,j_1} = \langle \psi_{j_2}, \phi_{j_1}\rangle$. Moreover, if $\mathbf{C}$ is the $n \times J_X$–matrix whose rows are the vectors $\mathbf{C}_i^t$, one can consider

$$\mathbf{Y} = \mathbf{C}\mathbf{M}_{\Psi\Phi}\mathbf{B} + \mathbf{e} = \mathbf{\Xi}\mathbf{B} + \mathbf{e},$$

where $\mathbf{Y}$ is the n–vector of the observed responses, $\mathbf{\Xi}$ is the $n \times J_\theta$–matrix defined by $\mathbf{\Xi} = \mathbf{C}\mathbf{M}_{\Psi\Phi}$, and $\mathbf{e}$ is the n–vector of the errors.

Next, two methods for estimating $\mathbf{B}$ based on the minimization of the ordinary residual sum of squares, and based on the minimization of a penalized residual sum of squares, respectively, are presented.

a) Least squares estimator

The K_θ–vector of coefficients $\mathbf{B}$ can be estimated optimizing the standard least squares problem

$$\min_{\mathbf{B}} (\mathbf{Y} - \mathbf{\Xi}\mathbf{B})^t(\mathbf{Y} - \mathbf{\Xi}\mathbf{B}).$$

It can be shown that the solution of this problem verifies the equation $2\mathbf{\Xi}^t\mathbf{\Xi}\mathbf{B} - 2\mathbf{\Xi}^t\mathbf{Y} = 0$. Hence, $\mathbf{B}$ is estimated by

$$\hat{\mathbf{B}} = (\mathbf{\Xi}^t\mathbf{\Xi})^{-1}\mathbf{\Xi}^t\mathbf{Y},$$

and the derived predictions $\hat{\mathbf{Y}} = \mathbf{\Xi}\hat{\mathbf{B}} = \mathbf{\Xi}(\mathbf{\Xi}^t\mathbf{\Xi})^{-1}\mathbf{\Xi}^t\mathbf{Y}$.

b) Penalized least squares estimator

The introduction of roughness penalties in the ordinary least squares problem allows to fix regularity conditions for the model parameter, and avoid excessive local fluctuation in the estimation. The trick is to consider the following penalized residual sum of squares

$$\min_{\mathbf{B}} \big((\mathbf{Y} - \mathbf{\Xi}\mathbf{B})^t(\mathbf{Y} - \mathbf{\Xi}\mathbf{B}) + \rho \mathrm{PEN}_q(\theta)\big),$$

being ρ the smoothing parameter, and $\mathrm{PEN}_q(\cdot)$ a roughness penalty, for example, any of the penalties considered in Section 1.3.1, *"c) Smoothing with a roughness penalty"*, in Chapter 1 (see page 19).

When the qth derivative of θ exists, $\mathrm{PEN}_q(\theta) = \int_0^1 (\theta^{(q)}(s))^2 ds$ can be selected, and the penalty can be expressed as

$$\mathrm{PEN}_q(\theta) = \int_0^1 \left((\mathbf{\Phi}^{(q)}(s))^t\mathbf{B}\right)^2 ds = \mathbf{B}^t\mathbf{R}\mathbf{B},$$

where $\mathbf{\Phi}^{(q)}(s) = (\phi_1^{(q)}(s), \ldots, \phi_{J_\theta}^{(q)}(s))^t$ is the J_θ–vector of the derivatives and $\mathbf{R}$ is the $J_\theta \times J_\theta$–matrix given by $\mathbf{R} = \int_0^1 \mathbf{\Phi}^{(q)}(s)(\mathbf{\Phi}^{(q)}(s))^t ds$. Consequently, the penalized optimization problem can be rewritten as follows

$$\min_{\mathbf{B}} \big((\mathbf{Y} - \mathbf{\Xi B})^t(\mathbf{Y} - \mathbf{\Xi B}) + \rho \mathbf{B}^t \mathbf{R} \mathbf{B}\big).$$

To solve this problem, one has to find the vector $\mathbf{B}$ which satisfies $2\mathbf{\Xi}^t\mathbf{\Xi B} - 2\mathbf{\Xi}^t\mathbf{Y} + 2\rho\mathbf{R}\mathbf{B} = 0$. Hence, the estimated parameter is

$$\hat{\mathbf{B}} = (\mathbf{\Xi}^t\mathbf{\Xi} + \rho\mathbf{R})^{-1}\mathbf{\Xi}^t\mathbf{Y},$$

and the fitted responses are $\hat{\mathbf{Y}} = \mathbf{\Xi}\hat{\mathbf{B}} = \mathbf{\Xi}(\mathbf{\Xi}^t\mathbf{\Xi} + \rho\mathbf{R})^{-1}\mathbf{\Xi}^t\mathbf{Y}$.

Remark 2.3.3. The balance between bias and variance is controlled by the smoothing parameter. The choice of ρ can be done in a subjective way or by data–driven selectors, such as a cross–validation method (see further details in Ramsay and Silverman, 2005, Chapter 15).

Example. The penalized B–splines estimator. The penalized B–splines estimator is an estimator based on basis expansions. Let $\{B_{k,j}, j = 1, \ldots, k+q\}$ be the normalized B–splines basis of the space S_{qk} of splines defined on $[0,1]$ with degree q and $(k-1)$ equispaced interior knots. The *penalized B–splines estimator* is defined as

$$\hat{\theta}_{PS} = \sum_{j=1}^{q+k} \hat{b}_j B_{k,j}, \tag{2.2}$$

where $\hat{\mathbf{b}} = (\hat{b}_1, \ldots, \hat{b}_{q+k})$ is the solution of the minimization problem

$$\min_{\mathbf{b}} \left(\frac{1}{n} \sum_{i=1}^{n} \left(Y_i - \sum_{j=1}^{q+k} \langle b_j B_{k,j}, X_i \rangle \right)^2 + \rho \|(\mathbf{B}_k^{(r)})^t \mathbf{b}\|^2 \right), \tag{2.3}$$

with ρ the smoothing parameter, and $\mathbf{B}_k^{(r)}$ the functional vector of derivatives of order r of the B–splines (see, for instance, Cardot et al., 2003c).

Remark 2.3.4. Let $\mathbf{\Xi}$ be the $n \times (q+k)$–matrix defined by $(\mathbf{\Xi})_{l,j} = \langle B_{k,j}, X_l \rangle$, let $\mathbf{R}$ be the $(q+k) \times (q+k)$–matrix defined by $(\mathbf{R})_{l,j} = \langle B_{k,l}^{(q)}, B_{k,j}^{(q)} \rangle$. Then the penalized B–splines estimator is an example of penalized least squares estimator, since the minimization problem (2.3) is equivalent to

$$\min_{\mathbf{b}} \Big((\mathbf{Y} - \mathbf{\Xi b})^t (\mathbf{Y} - \mathbf{\Xi b}) + \frac{\rho}{n} \mathbf{b}^t \mathbf{R} \mathbf{b} \Big).$$

2.3.2 Estimators based on FPCA

In this section, some recent results on FPCA–type estimators for the functional linear model (2.1) (see page 32) are briefly recalled. Firstly, the construction of this kind of estimators following the papers by Cardot et al. (1999, 2003c), and the most general approach by Cardot et al. (2007c), is presented. Then, theoretical results related to consistency, conditional errors and some asymptotics are compiled.

Let $(\mathcal{H}, \langle \cdot, \cdot \rangle)$ be a real separable Hilbert space. Recall that $\{(\lambda_j, v_j)\}_{j=1}^{\infty}$ denote the eigenelements of the second moment operator $\Gamma = \mathbb{E}(X \otimes_{\mathcal{H}} X)$, and $\Delta = \mathbb{E}(X \otimes_{\mathcal{H}'} Y)$ is the cross second moment operator. Furthermore, $\{(\hat{\lambda}_j, \hat{v}_j)\}_{j=1}^{\infty}$ are the eigenvalues and eigenfunctions of $\Gamma_n = n^{-1} \sum_{i=1}^n X_i \otimes_{\mathcal{H}} X_i$, and the empirical version of Δ is denoted by $\Delta_n = n^{-1} \sum_{i=1}^n X_i \otimes_{\mathcal{H}'} Y_i$. See further details in Section 1.4.1, *"b) Measures of dispersion"*, in Chapter 1 (see page 22).

a) Definition of standard FPCA estimator

In order to estimate the model parameter, Cardot et al. (2003c) studied the optimization problem

$$\min_{\beta \in \mathcal{H}} \mathbb{E}\left((Y - \langle \beta, X \rangle)^2\right).$$

When

(C.2.1) $\mathrm{Ker}(\Gamma) = \{0\}$, and $\sum_{j=1}^{\infty} (\Delta v_j/\lambda_j)^2 < +\infty$,

where $\mathrm{Ker}(\Gamma)$ denotes the *kernel* or *null space* of Γ defined as $\mathrm{Ker}(\Gamma) = \{x \in \mathcal{H} \,|\, \Gamma x = 0\}$, the parameter θ is the unique solution to this minimization problem, and it satisfies the equation $\Delta x = \langle \theta, \Gamma x \rangle$ for all $x \in \mathcal{H}$. This fact allows to express the model parameter as

$$\theta = \sum_{j=1}^{\infty} \frac{\Delta v_j}{\lambda_j} v_j.$$

Remark 2.3.5. The application of Remark 1.2.13 in Chapter 1 (see page 9), taking as orthonormal basis $\{v_j\}_j$, leads to $\theta = \sum_{j=1}^{\infty} \langle \theta, v_j \rangle v_j$. Moreover, the equality $\Delta x = \langle \theta, \Gamma x \rangle$, for all $x \in \mathcal{H}$, ensures that $\Delta v_j = \lambda_j \langle \theta, v_j \rangle$ for all j. Thus, $\theta = \sum_{j=1}^{\infty} \lambda_j^{-1} \Delta v_j v_j$.

There is no bounded inverse for Γ, so Cardot et al. (1999) projected the data on the subspace spanned by the first k_n eigenfunctions of Γ_n. Given $\{(X_i, Y_i)\}_{i=1}^n$ a sample of i.i.d. random variables drawn from (X, Y), they proposed to estimate θ by

$$\hat{\theta}_{k_n} = \sum_{j=1}^{k_n} \frac{\Delta_n \hat{v}_j}{\hat{\lambda}_j} \hat{v}_j, \tag{2.4}$$

where $\{k_n\}_{n=1}^{\infty}$ is a sequence of positive integers such that $k_n \to +\infty$, $k_n \leq n$, and $\hat{\lambda}_{k_n} > 0$.

Remark 2.3.6. Note that (2.4) is the truncated version of $\hat{\theta} = \sum_{j=1}^{\infty} \hat{\lambda}_j^{-1} \Delta_n \hat{v}_j \hat{v}_j$, which satisfies

$$\Delta_n x = \langle \hat{\theta}, \Gamma_n x \rangle, \quad \forall x \in \mathrm{Im}(\Gamma_n),$$

where $\mathrm{Im}(\Gamma_n)$ denotes the *image* of the operator Γ_n, i.e., $\mathrm{Im}(\Gamma_n) = \{x' \in \mathcal{H} \,|\, x' = \Gamma_n x,\, x \in \mathcal{H}\}$.

Alternative construction. The estimator $\hat{\theta}_{k_n}$ can be obtained by projecting the functional observations onto a finite subspace of $\mathcal{H}$, and applying the same arguments used in Faraldo-Roca and González-Manteiga (1987) in order to reduce the conditional mean square error of the standard least squares estimator as follows. Let $\{e_j\}_{j=1}^{\infty}$ be an orthonormal basis of $\mathcal{H}$, and fix $k_n < n$. For all $x \in \mathcal{H}$, the corresponding boldfaced letter $\mathbf{x}$ denotes the k_n–vector given by $\mathbf{x} = (\langle x, e_1 \rangle, \ldots, \langle x, e_{k_n} \rangle)^t$. Following the steps given by Cristóbal-Cristóbal et al. (1987), consider the generalized optimization problem

$$\min_{\mathbf{b}} \mathbb{E}_{\mu_n} \left((\hat{m}_{k_n}(\mathbf{X}) - \mathbf{X}^t \mathbf{b})^2 \right), \quad \text{with} \quad \mathbf{b} = (\langle \beta, e_1 \rangle, \ldots, \langle \beta, e_{k_n} \rangle)^t, \quad \forall \beta \in \mathcal{H}, \tag{2.5}$$

where

- $\hat{m}_{k_n}(\mathbf{x}) = \sum_{i=1}^n Y_i \delta(\mathbf{x}, \mathbf{X}_i) / \sum_{i=1}^n \delta(\mathbf{x}, \mathbf{X}_i)$ is a nonparametric estimator of the regression function $m_{k_n}(\mathbf{x}) = \mathbb{E}(Y | \mathbf{X} = \mathbf{x})$, and
- $\mu_n(\mathbf{x}) = \int_{-\infty}^{\mathbf{x}} \hat{f}_n(\mathbf{t}) d\mathbf{t}$ is a weighting function where $\hat{f}_n$ is a nonparametric estimator of the density f of $\mathbf{X}$ defined as $\hat{f}_n(\mathbf{x}) = n^{-1} \sum_{i=1}^n \delta(\mathbf{x}, \mathbf{X}_i)$,

being $\delta(\cdot, \cdot)$ a measurable function from $\mathbb{R}^{k_n} \times \mathbb{R}^{k_n}$ into $\mathbb{R}$. Solving (2.5) for the special case

$$\delta(\mathbf{u}, \mathbf{w}) = \begin{cases} 1, & \text{if } \mathbf{u} = \mathbf{w}, \\ 0, & \text{if } \mathbf{u} \neq \mathbf{w}, \end{cases}$$

one gets that $\beta_0 \in \mathcal{H}$ is a solution of (2.5) if and only if $\mathbf{b}_0 = (\langle \beta_0, e_1 \rangle, \ldots, \langle \beta_0, e_{k_n} \rangle)^t$ satisfies the k_n–dimensional normal equation $\mathbf{\Delta}_n = \mathbf{\Gamma}_n \mathbf{b}_0$ where

$$\mathbf{\Gamma}_n = \frac{1}{n} \sum_{i=1}^n \mathbf{X}_i \mathbf{X}_i^t \quad \text{and} \quad \mathbf{\Delta}_n = \frac{1}{n} \sum_{i=1}^n \mathbf{X}_i Y_i.$$

This amounts to k_n equations

$$\Delta_n e_l = \sum_{j=1}^{k_n} \langle \Gamma_n e_l, e_j \rangle \langle \beta_0, e_j \rangle, \quad \text{for all } l \in \{1, \ldots, k_n\}.$$

Taking $e_l = \hat{v}_l$, one gets $\langle \beta_0, \hat{v}_l \rangle = \hat{\lambda}_l^{-1} \Delta_n \hat{v}_l$ for all $l \in \{1, \ldots, k_n\}$, and the estimator (2.4) appears as the projection of β_0 onto the subspace of $\mathcal{H}$ spanned by $\{\hat{v}_j\}_{j=1}^{k_n}$ (i.e., $\hat{\theta}_{k_n} = \sum_{j=1}^{k_n} \langle \beta_0, \hat{v}_j \rangle \hat{v}_j$).

Remark 2.3.7. The standard FPCA estimator $\hat{\theta}_{k_n}$ is the solution of (2.5) for a basic nonparametric estimator (i.e., $\hat{m}_{k_n}(\mathbf{X}_i) = Y_i$). A natural extension of $\hat{\theta}_{k_n}$ consists in investigating on the solution of (2.5) for a general nonparametric estimator, which is equivalent to presmooth the responses before estimating the functional parameter θ (see more details in Chapter 3).

b) Definition of general class of FPCA–type estimators

Following the standard FPCA approach, Cardot et al. (2007c) analysed a large class of FPCA–type estimators, where (2.4) is included as particular case. They assumed that $\lambda_1 > \lambda_2 > \ldots > 0$, where the multiplicity of each λ_j is one, and they solved the original problem of ill–conditioned inverse of Γ by means of a regularization procedure for which it is necessary to introduce some notation. They define

$$\delta_j = \begin{cases} \lambda_1 - \lambda_2, & \text{if } j = 1, \\ \min(\lambda_{j-1} - \lambda_j, \lambda_j - \lambda_{j+1}), & \text{if } j \neq 1, \end{cases}$$

fix a strictly positive sequence $c = c_n$ such that $c \to 0$ and $c < \lambda_1$, and set

$$k_n^c = \sup \{j : \lambda_j + \delta_j/2 \geq c\}.$$

They also consider a sequence $\{f_n^c : [c, +\infty) \to \mathbb{R}\}_n$ of positive functions such that

(C.2.2) f_n^c is decreasing on $[c, \lambda_1 + \delta_1]$,

(C.2.3) $(f_n^c)'(x)$ exists for all $x \in [c, +\infty)$, and

(C.2.4) $\sup_{x \geq c} |x f_n^c(x) - 1| = o(n^{-1/2})$.

Then, the proposed estimator of the model parameter θ is

$$\hat{\theta}_c = \sum_{j=1}^{n} f_n^c(\hat{\lambda}_j) \Delta_n \hat{v}_j \hat{v}_j. \tag{2.6}$$

Remark 2.3.8. Regarding the support of f_n^c, one gets

$$\hat{\theta}_c = \sum_{j=1}^{J_n^c} f_n^c(\hat{\lambda}_j) \Delta_n \hat{v}_j \hat{v}_j,$$

where $J_n^c = \sup \{j : \hat{\lambda}_j \geq c\}$. Obviously, J_n^c is a random sequence that can be different from k_n^c. However, looking at proof of Lemma 5 by Cardot et al. (2007c), it can be deduced that, if $(k_n^c)^2 \log k_n^c / \sqrt{n} \to 0$, $\mathbb{P}(J_n^c \neq k_n^c) \to 0$ fast enough to consider only the case $J_n^c = k_n^c$ (rate $n^{-1/2}$).

Example 1. From the previous remark, it is easy to deduce that the standard FPCA estimator (2.4) is asymptotically equivalent to (2.6) when

$$f_n(x) = x^{-1} \mathbb{I}_{\{x \geq c\}}.$$

Example 2. The ridge–type FPCA estimator proposed by Martínez-Calvo (2008) and Ferraty et al. (2012a) also belongs to this large family of estimators based on FPCA. In this case, one has to choose

$$f_n(x) = (x + \alpha_n)^{-1}\mathbb{I}_{\{x \geq c\}},$$

for α_n a sequence of positive parameters.

c) Consistency

In this section, Theorem 3.2 by Cardot et al. (1999) is reproduced. This result ensures the almost surely convergence of the standard FPCA estimator $\hat{\theta}_{k_n}$. First of all, the almost surely convergence and its associated "big O" and "little o" notation are introduced.

Definition 2.3.9. Let $\{Z_n\}_{n\in\mathbb{N}}$ be a sequence of real random variables and let Z be a real random variable, all of them defined on the same probability space $(\Omega, \mathcal{A}, \mathbb{P})$. $\{Z_n\}_{n\in\mathbb{N}}$ converges *almost surely* ($a.s.$ or $a.s.-\mathbb{P}$) to Z, that is,

$$\lim_{n\to\infty} Z_n = Z \quad a.s., \quad \text{or equivalently,} \quad Z_n \to Z \quad a.s.,$$

if and only if

$$\mathbb{P}\left(\omega \in \Omega : \lim_{n\to\infty} Z_n(\omega) = Z(\omega)\right) = 1.$$

Definition 2.3.10. Let $\{Z_n\}_{n\in\mathbb{N}}$ be a sequence of real random variables and let Z be a real random variable, all of them defined on the same probability space $(\Omega, \mathcal{A}, \mathbb{P})$. Let $\{u_n\}_{n\in\mathbb{N}}$ be a deterministic sequence of positive real numbers. The *rate of almost surely convergence* of $\{Z_n\}_{n\in\mathbb{N}}$ to Z is of order u_n, that is,

$$Z_n - Z = O_{a.s.}(u_n),$$

if and only if

$$\mathbb{P}\left(\omega \in \Omega : \exists c < \infty, \exists N, \forall n > N, |Z_n(\omega) - Z(\omega)| \leq c u_n\right) = 1.$$

Furthermore,

$$Z_n - Z = o_{a.s.}(u_n) \quad \text{if and only if} \quad (u_n)^{-1}(Z_n - Z) \to 0 \quad a.s.$$

On the other hand, recall that $\|\cdot\|_{\mathcal{H}'}$ denotes the norm of the dual space $\mathcal{H}'$ (see Section 1.2.2, *"b) The dual space $\mathcal{H}'$ "*, in Chapter 1, page 10). It is also necessary to introduce the following notation $a_j = 2\sqrt{2}/\delta_j$, that is,

$$a_j = \begin{cases} 2\sqrt{2}/(\lambda_1 - \lambda_2), & \text{if } j = 1, \\ 2\sqrt{2}/\min(\lambda_{j-1} - \lambda_j, \lambda_j - \lambda_{j+1}), & \text{if } j \neq 1, \end{cases}$$

and $m(\cdot) = \langle \theta, \cdot \rangle$ and $\hat{m}_{k_n}(\cdot) = \langle \hat{\theta}_{k_n}, \cdot \rangle$. In addition, the following assumptions are required:

(C.2.5) $\lambda_1 > \lambda_2 > \ldots > 0$, and $\hat{\lambda}_1 > \hat{\lambda}_2 > \ldots > \hat{\lambda}_{k_n} > 0$ $a.s.$

(C.2.6) $\|X\| \leq c_1$ $a.s.$

(C.2.7) $\exists c_2 > 0, \forall l \geq 1, \mathbb{E}(|\epsilon|^l) < l! c_2 < +\infty$,

(C.2.8) $n\lambda_{k_n}^4 / \log n \to +\infty$, and $n\lambda_{k_n}^2 / ((\sum_{j=1}^{k_n} a_j)^2 \log n) \to +\infty$.

Theorem 2.3.11 (Cardot et al., 1999)**.** *If* **(C.2.1)** *and* **(C.2.5)**–**(C.2.8)** *are satisfied, then*

$$\|\hat{m}_{k_n} - m\|_{\mathcal{H}'} \to 0 \quad a.s.$$

Remark 2.3.12. Actually, Cardot et al. (1999) obtained the previous result assuming the hypothesis

(C.2.7') $|\epsilon| \leq c_2$ *a.s.*

instead of **(C.2.7)**, but it can be shown that consistency holds despite this modification. The key point is to replace the second part of Lemma 5.3 by Cardot et al. (1999) by Lemma 2.5.1 (see page 47). This assumption has been altered in order to extend the original result by Cardot et al. (1999) to a larger class of errors, which includes, for instance, non–bounded errors such as the Gaussian ones.

Remark 2.3.13. If $k_n = o(\log n)$, **(C.2.8)** is satisfied when λ_j are geometrically or algebraically decreasing, that is, when

$$\lambda_j = c_1 c_2^j, \quad \text{with } c_1 > 0 \text{ and } 0 < c_2 < 1,$$

or

$$\lambda_j = c_1 j^{-c_3}, \quad \text{with } c_1 > 0 \text{ and } c_3 > 1.$$

d) Conditional errors

The next theorem gives the conditional mean square prediction error for a new response

$$Y_{n+1} = \langle \theta, X_{n+1} \rangle + \epsilon_{n+1},$$

and the conditional mean square estimation error. For any $n \in \mathbb{N}^*$, denote $\mathcal{X}^n = \{X_1, \ldots, X_n\}$, and let $\mathbb{E}_{\mathcal{X}^n}(\cdot)$ be the expectation conditionally on $\mathcal{X}^n$. Furthermore, let $\hat{R}_{k_n}$ be the term defined as

$$\hat{R}_{k_n} = \sum_{j > k_n} \langle \theta, \hat{v}_j \rangle \hat{v}_j. \tag{2.7}$$

Theorem 2.3.14 (Ferraty et al., 2012a)**.** *For the standard FPCA estimator* (2.4)*, it holds that*

$$\mathbb{E}_{\mathcal{X}^{n+1}}(Y_{n+1} - \langle \hat{\theta}_{k_n}, X_{n+1} \rangle)^2 = \sigma^2 + \frac{\sigma^2}{n} \sum_{j=1}^{k_n} \frac{\langle X_{n+1}, \hat{v}_j \rangle^2}{\hat{\lambda}_j} + \langle X_{n+1}, \hat{R}_{k_n} \rangle^2,$$

$$\mathbb{E}_{\mathcal{X}^n}(\|\theta - \hat{\theta}_{k_n}\|^2) = \frac{\sigma^2}{n} \sum_{j=1}^{k_n} \frac{1}{\hat{\lambda}_j} + \|\hat{R}_{k_n}\|^2,$$

where $\hat{R}_{k_n}$ *is defined in* (2.7).

The proof of the previous result can be found in the appendix of the chapter (see Section 2.5.2, page 48).

Remark 2.3.15. Theorem 2.3.14 can be derived directly from Lemma 2.5.2 (see page 49), which allows to obtain the conditional errors for a general type of estimators in a simple way. The conditional estimation error of $\hat{\theta}_{k_n}$ given in Theorem 2.3.14 was already studied in literature. In fact, Theorem 5 in Hall and Hosseini-Nasab (2006) gives conditions in order to find that

$$\mathbb{E}(\|\theta - \hat{\theta}_{k_n}\|^2) \sim \frac{\sigma^2}{n} \sum_{j=1}^{k_n} \frac{1}{\lambda_j} + \|R_{k_n}\|^2,$$

where $R_{k_n} = \sum_{j > k_n} \langle \theta, v_j \rangle v_j$, and $W_n \sim Z_n$ means that the ratio of W_n and Z_n converges to 1 when $n \to +\infty$.

e) Asymptotic normality

In this section, the general FPCA–type estimator $\hat{\theta}_c$ defined by (2.6) is considered. Cardot et al. (2007c) proposed in their work to derive a Central Limit Theorem (CLT) in the functional linear regression model with scalar response. In particular, they gave the following weak convergence results for the prediction at a fixed $x \in \mathcal{H}$, being the following assumptions required:

(C.2.9) $\sum_{j=1}^{+\infty} |\langle \theta, v_j \rangle| < +\infty$,

(C.2.10) $\exists \lambda$ a convex positive function such that $\lambda_j = \lambda(j)$, for j large,

(C.2.11) $\mathbb{E}(\|X\|^4) < \infty$, and $\sup_j (\mathbb{E}(\langle X, v_j \rangle^4)/\lambda_j^2) < +\infty$,

(C.2.12) $(k_n^c)^3 (\log(k_n^c))^2 / (t_{n,x}^c \sqrt{n}) \to 0$ with $t_{n,x}^c = \sqrt{\sum_{j=1}^{k_n^c} \lambda_j (f_n^c(\lambda_j))^2 \langle x, v_j \rangle^2}$,

(C.2.13) $\sum_{j=1}^{+\infty} (\langle x, v_j \rangle^2 / \lambda_j) < +\infty$.

In addition, recall the definition of convergence in distribution.

Definition 2.3.16. Let $\{Z_n\}_{n\in\mathbb{N}}$ be a sequence of real random variables with distribution F_n and let Z be a real random variable with distribution F. $\{Z_n\}_{n\in\mathbb{N}}$ converges *in distribution* or *weakly* (d or w) to Z, that is,

$$Z_n \xrightarrow{d} Z, \quad \text{or equivalently,} \quad Z_n \xrightarrow{w} Z,$$

if and only if

$$\lim_{n\to\infty} F_n(x) = F(x),$$

for every $x \in \mathbb{R}$ at which F is continuous.

Theorem 2.3.17 (Cardot et al., 2007c)**.** *Given a value x in $\mathcal{H}$, if* **(C.2.1)***,* **(C.2.2)**–**(C.2.4)***, and* **(C.2.9)**–**(C.2.13)** *hold, then*

$$\frac{\sqrt{n}}{t_{n,x}^c \sigma} (\langle \hat{\theta}_c, x \rangle - \langle \hat{\Pi}_{k_n^c} \theta, x \rangle) \xrightarrow{w} \mathcal{N}(0,1),$$

being $\hat{\Pi}_{k_n^c}$ the projector onto the subspace spanned by the first k_n^c eigenfunctions of Γ_n, and $\mathcal{N}(0,1)$ the distribution of a Gaussian random variable with mean equals to 0 and variance equals to 1.

Corollary 2.3.18 (Cardot et al., 2007c)**.** *Theorem 2.3.17 still holds if $t_{n,x}^c$ (defined in assumption* **(C.2.12)***) is replaced by its empirical counterpart*

$$\hat{t}_{n,x}^c = \sqrt{\sum_{j=1}^{k_n^c} \hat{\lambda}_j (f_n^c(\hat{\lambda}_j))^2 \langle x, \hat{v}_j \rangle^2},$$

and σ is replaced by a consistent estimate $\hat{\sigma}$.

Remark 2.3.19. Note that the bias term is random. Cardot et al. (2007c) remarked that very specific assumptions on θ or $\{\lambda_j\}_j$ are required in order to replace the bias by a non–random quantity.

Remark 2.3.20. Confidence intervals for prediction can be derived from the CLT in Corollary 2.3.18. Let $\alpha \in [0,1]$ be a certain confidence level, and let z_α be the quantile of order α from a $\mathcal{N}(0,1)$ distribution. If it is assumed that θ (or x) is very well approximated by its projection $\hat{\Pi}_{k_n^c}\theta$ (or $\hat{\Pi}_{k_n^c}x$), then Corollary 2.3.18 allows to evaluate asymptotic confidence intervals for $\langle\theta, x\rangle$ by computing

$$\mathrm{CI}_{x,\alpha}^{asy} = \left[\langle\hat{\theta}_c, x\rangle - \hat{t}_{n,x}^c\hat{\sigma}n^{-1/2}z_{1-\alpha/2}, \langle\hat{\theta}_c, x\rangle + \hat{t}_{n,x}^c\hat{\sigma}n^{-1/2}z_{1-\alpha/2}\right], \tag{2.8}$$

for which $\mathbb{P}(m(x) \in \mathrm{CI}_{x,\alpha}^{asy}) \approx 1-\alpha$.

2.4 Functional nonparametric regression for scalar response

This section is focused on the kernel–type estimators for the functional nonparametric model expressed as

$$Y = m(X) + \epsilon, \tag{2.9}$$

where Y is a real random variable, $m(\cdot)$ is a functional regression operator which satisfies some smoothness restrictions, X is a zero–mean random variable valued in an abstract space $\mathcal{S}$, and ϵ is a real random variable satisfying that $\mathbb{E}(\epsilon) = 0$, $\mathrm{Var}(\epsilon) = \sigma^2 < \infty$, and $\mathbb{E}(\epsilon X) = 0$.

The most general formulation assumes that the functional space $\mathcal{S}$ is a semi–metric space endowed with a semi–metric $d(\cdot,\cdot)$, so this abstract approach has been almost always kept in the next subsections except for certain results which require specifically normed spaces. In order to preserve the notation used up to now, the semi–metric space $(\mathcal{H}, d(\cdot,\cdot))$, where $\mathcal{H}$ is still a real separable Hilbert space, and $d(\cdot,\cdot)$ is the semi–metric induced by the inner product, i.e., $d(x,y) = \|x-y\| = \langle x-y, x-y\rangle^{1/2}$ (in fact, $d(\cdot,\cdot)$ is a metric with this construction), is going to be selected. However, note that any other semi–metric space could be considered (see some examples of semi–metrics in Section 1.2.3 in Chapter 1, page 10).

Among the different existing nonparametric methods, the section is focused on the kernel–type estimators based on conditional expectation introduced by Ferraty and Vieu (2006b), and studied by Ferraty and Vieu (2004) and Ferraty et al. (2007a), although other nonparametric techniques based on conditional median or conditional mode were also developed in the literature (see, for instance, Ferraty and Vieu, 2006b). Next, some outstanding advances related with this functional kernel methodology have been summed up.

2.4.1 Kernel–type estimators

Given the functional model (2.9), the goal is to estimate the regression operator $m(\cdot)$ defined by

$$\begin{array}{rccl} m: & \mathcal{H} & \to & \mathbb{R} \\ & x & \to & m(x) = \mathbb{E}(Y|X=x). \end{array}$$

As mentioned before, the regression operator does not belong to any parametric family in this case, but $m(\cdot)$ satisfies certain regularity assumptions. Usually, one considers continuity–type conditions

(C.2.14) $m \in C_{\mathcal{H},0} = \{f : \mathcal{H} \to \mathbb{R} \text{ such that } \lim_{d(x,x')\to 0} f(x') = f(x)\}$,

or Lipschitz–type conditions, that is, $\exists\beta > 0$ such that

(C.2.15) $m \in Lip_{\mathcal{H},\beta} = \{f : \mathcal{H} \to \mathbb{R} \text{ such that } \exists c > 0, |f(x) - f(x')| < cd(x,x')^\beta\}$.

In the following subsections, the functional kernel estimator is presented, together with some interesting results. In some of them, the *small ball probability function* $\varphi_x(\cdot)$ will play a decisive role. Let $B(x,h)$ be a ball centred at $x \in \mathcal{H}$ with radius h, that is, $B(x,h) = \{x' \in \mathcal{H} \,|\, d(x,x') \leq h\}$. Then $\varphi_x(h)$ will denote the probability of the ball $B(x,h)$, i.e.,

$$\varphi_x(h) = \mathbb{P}(X \in B(x,h)).$$

More details about the small ball probability notion and examples of what $\varphi_x(h)$ is some specific cases can be found in Li and Shao (2001), and Ferraty and Vieu (2006b, 2011a).

a) Definition of kernel estimator

Given $x \in \mathcal{H}$ and a sample $\{(X_i, Y_i)\}_{i=1}^n$ drawn from the pair (X, Y), Ferraty and Vieu (2006b) proposed the following estimator for the regression operator

$$\hat{m}_h(x) = \frac{\sum_{i=1}^n Y_i K(h^{-1} d(x, X_i))}{\sum_{i=1}^n K(h^{-1} d(x, X_i))}, \tag{2.10}$$

where $K(\cdot)$ is an asymmetric kernel, and $h = h(n)$ is a strictly positive real bandwidth. The kernel estimator (2.10) is the natural adaptation of the Nadaraya–Watson estimator to the functional context, where the ordinary multivariate norm was replaced by the functional semi–metric $d(\cdot, \cdot)$ which measures the distances between functional observations. Given that $d(x, y) \geq 0$ for all $x, y \in \mathcal{H}$, the support of $K(\cdot)$ should be positive. Hence, asymmetric kernels are the most adequate choice: for instance, the asymetric versions of the kernel functions introduced in Figure 1.5 (see Chapter 1, page 15), which are plotted in Figure 2.1.

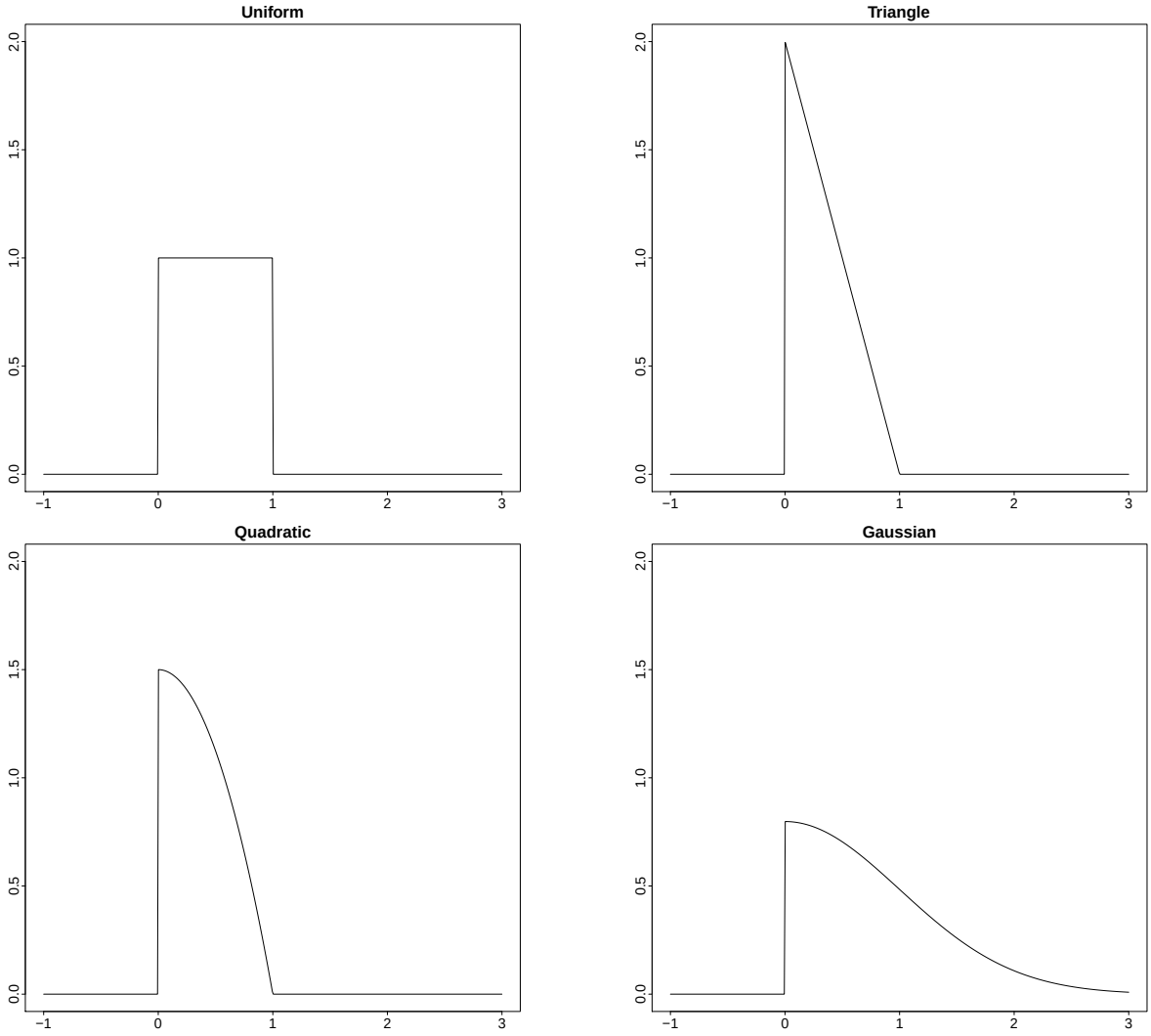

Figure 2.1: Usual asymmetric kernels: asymmetric uniform (top left panel), asymmetric triangle (top right panel), asymmetric quadratic (bottom left panel) and asymmetric Gaussian (bottom right panel).

This kernel estimator is actually a weighted average of $\{Y_i\}_{i=1}^n$. This issue can be easily seen if one

considers the weights

$$w_{ih}(x) = \frac{K(h^{-1}d(x, X_i))}{\sum_{i'=1}^{n} K(h^{-1}d(x, X_{i'}))},$$

which satisfy $\sum_{i=1}^{n} w_{ih}(x) = 1$. Hence, the estimator (2.10) can be expressed as the weighted average of the observed responses, that is,

$$\hat{m}_h(x) = \sum_{i=1}^{n} w_{ih}(x) Y_i.$$

It is clear that the kernel estimator depends on the choice of the kernel function $K(\cdot)$, and the smoothing parameter h. That is why some notes on $K(\cdot)$ and h selection are presented below.

Kernel functions. The kernel $K(\cdot)$ is usually an asymmetric version of kernel functions exposed in Section 1.3.1, *"a) Linear smoothing"*, in Chapter 1 (see page 13). In general, the most common kernels belong to one of the following families (see Ferraty and Vieu, 2006b).

Definition 2.4.1. A function $K : \mathbb{R} \to [0, +\infty)$ is called a *kernel of type I* if $\int K(u)du = 1$, and if there exist $c_1, c_2 \in \mathbb{R}$ such that $0 < c_1 < c_2 < +\infty$ and

$$c_1 \mathbb{I}_{\{u \in [0,1]\}} \leq K(u) \leq c_2 \mathbb{I}_{\{u \in [0,1]\}},$$

being $\mathbb{I}$ the indicator function.

Definition 2.4.2. A function $K : \mathbb{R} \to [0, +\infty)$ is called a *kernel of type II* if $\int K(u)du = 1$, its support is $[0, 1]$, and its first derivative K' exists on $[0, 1]$ and satisfies

$$c_1 \leq K' \leq c_2,$$

for $c_1, c_2 \in \mathbb{R}$ such that $-\infty < c_1 < c_2 < 0$.

The main discontinuous asymmetric kernels belong to the first family (e.g., asymmetric uniform), whereas the second family contains the usual asymmetric continuous ones (e.g., asymmetric triangle or asymmetric quadratic). There is a close link between these kernel families and the small ball probability φ_x exactly as the next lemmas state.

Lemma 2.4.3 (Ferraty and Vieu, 2006b)**.** *Let $K(\cdot)$ be a kernel of type I. Then there exist $c_1, c_2 \in \mathbb{R}$ with $0 \leq c_1, c_2 < +\infty$ such that*

$$c_1 \varphi_x(h) \leq \mathbb{E}(K(h^{-1}d(x, X))) \leq c_2 \varphi_x(h).$$

The previous result can be extended to kernel of type II. For this purpose, it is necessary that the next assumption

(C.2.16) $\exists c > 0, \exists \varepsilon_0, \forall \varepsilon < \varepsilon_0, \int_0^{\varepsilon} \varphi_x(u)du > c\varepsilon\varphi_x(\varepsilon)$

is satisfied by the small ball probability φ_x.

Lemma 2.4.4 (Ferraty and Vieu, 2006b)**.** *Let $K(\cdot)$ be a kernel of type II. If $\varphi_x(\cdot)$ verifies* **(C.2.16)** *then there exist $c_1, c_2 \in \mathbb{R}$ with $0 \leq c_1, c_2 < +\infty$ such that, for h small enough,*

$$c_1 \varphi_x(h) \leq \mathbb{E}(K(h^{-1}d(x, X))) \leq c_2 \varphi_x(h).$$

Bandwidth. The bandwidth h acts as a smoothing parameter. For example, if $K(\cdot)$ is a positive kernel supported on $[0,1]$, the estimator (2.10) only takes into account the responses Y_i associated to curves X_i such that $d(x,X_i) \leq h$, since $K(h^{-1}d(x,X_i)) = 0$ when the distance between x and X_i is larger than h. Therefore, the larger h is, the larger the number of response values which are involved in the estimation procedure is and the smoother estimator is obtained. Consequently, if h is too small, the corresponding estimator will probably be too rough and sensitive to small alterations of the responses. This is a constant issue shared by all the nonparametric methods: the search of the right balance between bias and variance.

Remark 2.4.5. From a practical point of view, it seems that the choice of $K(\cdot)$ has no significant effect on the obtained results, whereas the choice of the smoothing parameter is crucial. Some data–driven methods were proposed in the literature in order to estimate h. For instance, one can use a cross–validation method, and estimates h by means of the solution of the minimization problem

$$\min_h \mathrm{CV}(h) = \min_h \sum_{j=1}^{n} (Y_j - \hat{m}_h^{(-j)}(X_j)),$$

where

$$\hat{m}_h^{(-j)}(x) = \frac{\sum_{i\neq j} Y_i K(h^{-1}d(x,X_i))}{\sum_{i\neq j} K(h^{-1}d(x,X_i))}.$$

b) Consistency

Ferraty and Vieu (2006b) showed in their monograph the almost complete convergence of the kernel estimator (in fact, they calculated the rate of convergence). First of all, recall what almost complete convergence means.

Definition 2.4.6. Let $\{Z_n\}_{n\in\mathbb{N}}$ be a sequence of real random variables, and let Z be a real random variable. $\{Z_n\}_{n\in\mathbb{N}}$ converges *almost completely* (*a.co.*) to Z, that is,

$$\lim_{n\to\infty} Z_n = Z \quad a.co., \quad \text{or equivalently,} \quad Z_n \to Z \quad a.co.,$$

if and only if

$$\forall \varepsilon > 0, \quad \sum_{n\in\mathbb{N}} \mathbb{P}(|Z_n - Z| > \varepsilon) < \infty.$$

Definition 2.4.7. Let $\{Z_n\}_{n\in\mathbb{N}}$ be a sequence of real random variables, let Z be a real random variable, and let $\{u_n\}_{n\in\mathbb{N}}$ be a deterministic sequence of positive real numbers. The *rate of almost complete convergence* of $\{Z_n\}_{n\in\mathbb{N}}$ to Z is of order u_n, that is,

$$Z_n - Z = O_{a.co.}(u_n),$$

if and only if

$$\exists \varepsilon > 0, \quad \sum_{n\in\mathbb{N}} \mathbb{P}(|Z_n - Z| > \varepsilon u_n) < \infty.$$

Furthermore,

$$Z_n - Z = o_{a.co.}(u_n) \quad \text{if and only if} \quad (u_n)^{-1}(Z_n - Z) \to 0 \quad a.co.$$

Once the almost complete convergence has been introduced, one is in a position to expose the main results presented by Ferraty and Vieu (2006b). The following hypotheses are required to obtain the results below:

(C.2.17) $\forall \varepsilon > 0, \quad \varphi_x(\varepsilon) > 0$,

(C.2.18) $h = h(n)$ is a positive sequence such that $h \to 0$, and $\log n/(n\varphi_x(h)) \to 0$ when $n \to \infty$,

(C.2.19) $K(\cdot)$ is a kernel of type I, or $K(\cdot)$ is a kernel of type II such that **(C.2.16)** holds,

(C.2.20) $\forall m \geq 2, \quad \mathbb{E}(|Y^m| \, | X = x) = \sigma_m(x) < \infty$ with $\sigma_m(\cdot)$ continuous at x.

Theorem 2.4.8 (Ferraty and Vieu, 2006b)**.** *Under* **(C.2.14)** *and* **(C.2.17)**–**(C.2.20)**, *it holds that*

$$\hat{m}_h(x) \to m(x) \quad a.co.$$

Theorem 2.4.9 (Ferraty and Vieu, 2006b)**.** *Under* **(C.2.15)** *and* **(C.2.17)**–**(C.2.20)**, *it holds that*

$$\hat{m}_h(x) - m(x) = O(h^\beta) + O_{a.co.}\left(\sqrt{\frac{\log n}{n\varphi_x(h)}}\right).$$

Remark 2.4.10. The assumption **(C.2.17)** is the functional adaptation of a usual condition in multivariate context: the density of the predictor variable is strictly positive. As far as **(C.2.20)** is concerned, this hypothesis allows to deal with non–bounded response variables.

An extended study of the rate of convergence of the estimator (2.10) was developed by Ferraty and Vieu (2004) (for a more general regression framework, see Ferraty et al., 2010b). One of their corollaries states the uniform consistency if the next assumptions hold:

(C.2.21) $\mathcal{C}$ can be expressed as $\mathcal{C} \subset \bigcup_{k=1}^{\eta} B(t_k, l)$, for any finite set $\{t_k\}_{k=1}^{\eta} \subset \mathcal{H}$ and any $l > 0$. Moreover, $\exists \gamma_1, \gamma_2 > 0$ such that $\eta l_1^\gamma = \gamma_2$,

(C.2.22) $K(\cdot)$ is a Lipschitz continuous kernel of type II,

(C.2.23) there is a function ϕ such that $\int_0^1 \phi(ht)dt/\phi(h) > c > 0$, and verifying

$$\exists c_1, c_2 \in (0, +\infty), \quad \forall x \in \mathcal{C}, \quad 0 < c_1\phi(h) \leq \varphi_x(h) \leq c_2\phi(h).$$

Theorem 2.4.11 (Ferraty and Vieu, 2004)**.** *Let $\mathcal{C}$ be a compact subset of $\mathcal{H}$. Under* **(C.2.15)**, **(C.2.18)** *and* **(C.2.20)**, *if* **(C.2.21)**–**(C.2.23)** *are satisfied, then it holds that*

$$\sup_{x \in \mathcal{C}} |\hat{m}_h(x) - m(x)| = O(h^\beta) + O\left(\sqrt{\frac{\log n}{n\phi(h)}}\right) \quad a.s.$$

Remark 2.4.12. Although the topological assumption **(C.2.21)** seems very restrictive for a start, a wide class of projection–based semi–metric spaces satisfies this condition (see, for instance, Ferraty and Vieu, 2008, Proposition 3.1).

c) Bias and variance

Ferraty et al. (2007a) computed the expressions of the bias and the variance for the kernel estimator when the functional space is a Banach space, and the considered semi–metric is $d(x, y) = \|x - y\|$. To

achieve this goal, some notation must be introduced. Let ψ_x be the function defined as

$$\psi_x(s) = \mathbb{E}\left[(m(X) - m(x))|d(x, X) = s\right], \quad \forall s \in \mathbb{R},$$

and let $\tau_{x,h}$ be the function given by

$$\tau_{x,h}(s) = \frac{\varphi_x(hs)}{\varphi_x(h)} = \mathbb{P}(d(x, X) \leq hs | d(x, X) \leq h), \quad \forall s \in [0, 1].$$

In addition, the following conditions are required:

(C.2.24) $\psi'_x(0)$ exists,

(C.2.25) $K(\cdot)$ is supported on $[0, 1]$, it has continuous derivative on $[0, 1)$, $K'(s) \leq 0$ and $K(1) > 0$,

(C.2.26) $h \to 0$, $\varphi_x(0) = 0$, and $n\varphi_x(h) \to +\infty$,

(C.2.27) $\forall s \in [0, 1]$, $\tau_{x,h}(s) \to \tau_{x,0}(s)$ when $h \to 0$.

Theorem 2.4.13 (Ferraty et al., 2007a)**.** *When* **(C.2.14)** *and* **(C.2.24)**–**(C.2.27)** *are satisfied, it holds that*

$$\mathbb{E}(\hat{m}_h(x)) = m(x) + \psi'_x(0)\frac{M_{x,0}}{M_{x,1}}h + O\left(\frac{1}{n\varphi_x(h)}\right) + o(h),$$
$$Var(\hat{m}_h(x)) = \frac{\sigma^2}{n\varphi_x(h)}\frac{M_{x,2}}{M_{x,1}^2} + o\left(\frac{1}{n\varphi_x(h)}\right),$$

where $M_{x,0} = K(1) - \int_0^1 (sK(s))'\tau_{x,0}(s)ds$, $M_{x,1} = K(1) - \int_0^1 K'(s)\tau_{x,0}(s)ds$ *and* $M_{x,2} = K^2(1) - \int_0^1 (K^2)'(s)\tau_{x,0}(s)ds$.

Remark 2.4.14. Once again, the smoothing parameter plays a key role in the bias–variance trade–off: large values of h reduce the variance (increase the bias), whereas small values of the bandwidth raise the variance (decrease the bias). Nevertheless, the previous theorem gives expressions that depend on unknown quantities, so these theoretical results cannot be used to find a value of h which balances bias and variance in practice. To avoid this drawback, Ferraty et al. (2007a) proposed a bandwidth selector based on a wild bootstrap method.

d) Asymptotic normality

The asymptotic distribution of the kernel estimator was also obtained by Ferraty et al. (2007a) in the following theorem, where the same notation introduced in the previous section is used, and the next condition is necessary:

(C.2.28) $\psi'_x(0) \neq 0$, and $M_{x,0} > 0$.

Theorem 2.4.15 (Ferraty et al., 2007a)**.** *When* **(C.2.14)**, **(C.2.24)**–**(C.2.27)**, *and* **(C.2.28)** *are satisfied, it holds that*

$$\sqrt{n\hat{\varphi}_x(h)}(\hat{m}_h(x) - m(x) - \psi'_x(0)\frac{M_{x,0}}{M_{x,1}}h)\frac{M_{x,1}}{\sigma\sqrt{M_{x,2}}} \xrightarrow{w} \mathcal{N}(0, 1),$$

being $\hat{\varphi}_x$ *the empirical counterpart of* $\varphi_x(h)$*, i.e.,* $\hat{\varphi}_x(h) = \#(i : d(x, X_i) \leq h)/n$.

Besides, if the hypothesis

(C.2.29) $h\sqrt{n\varphi_x(h)} \to 0$

is verified, a simpler result can be derived from the previous theorem by means of the cancellation of the bias term.

Corollary 2.4.16 (Ferraty et al., 2007a)**.** *When* **(C.2.14)**, **(C.2.24)**–**(C.2.28)**, *and* **(C.2.29)** *are satisfied, it holds that*

$$\sqrt{n\hat{\varphi}_x(h)}(\hat{m}_h(x) - m(x))\frac{M_{x,1}}{\sigma\sqrt{M_{x,2}}} \xrightarrow{w} \mathcal{N}(0,1),$$

being $\hat{\varphi}_x$ *the empirical counterpart of* $\varphi_x(h)$*, i.e.,* $\hat{\varphi}_x(h) = \#(i : d(x, X_i) \leq h)/n$*.*

2.5 Appendix Chapter 2

In this appendix Lemma 2.5.1 has been compiled, which allow to replace the hypothesis **(C.2.7')**, proposed by Cardot et al. (1999) in their result for consistency of $\hat{m}_{k_n}$, by **(C.2.7)** (see Remark 2.3.12, page 39). This modification makes the extension of the original result by Cardot et al. (1999) to a larger class of errors.

Furthermore, the proof of Theorem 2.3.14 is also presented in this appendix, jointly with the technical Lemma 2.5.2 which is required to show it.

2.5.1 Formulation and proof of Lemma 2.5.1

Lemma 2.5.1. *Under* **(C.2.6)** *and* **(C.2.7)**, *it holds that*

$$\mathbb{P}(\|\Delta_n - \Delta\|_{\mathcal{H}'} > \xi) \leq 2\exp\left(-\frac{\xi^2 n}{2\,c_3'(c_3' + c_4'\xi)}\right),$$

where c_3' *and* c_4' *are positive constants.*

Proof. This lemma adapts the second part of Lemma 5.3 in Cardot et al. (1999) to the weaker assumption **(C.2.7)**.

Take

$$W_i = X_i \otimes_{\mathcal{H}'} Y_i - \mathbb{E}(X \otimes_{\mathcal{H}'} Y), \quad i = 1, \ldots, n,$$

with $\mathbb{E}(W_i) = 0$. Note that $W_i = X_i \otimes_{\mathcal{H}'} m(X_i) - \mathbb{E}(X \otimes_{\mathcal{H}'} m(X)) + X_i \otimes_{\mathcal{H}'} \epsilon_i$, and consequently

$$\|W_i\|_{\mathcal{H}'} \leq \|m\|_{\mathcal{H}'}(\|X_i\| + \mathbb{E}\|X_i\|) + |\epsilon_i|\,\|X_i\|.$$

As a by–product, $\forall l \geq 2$, one has

$$\|W_i\|_{\mathcal{H}'}^l \leq \sum_{k=0}^{l} C_l^k \|m\|_{\mathcal{H}'}^k (\|X_i\| + \mathbb{E}\|X_i\|)^k |\epsilon_i|^{l-k} \|X_i\|^{l-k}.$$

From this inequality, **(C.2.6)**, **(C.2.7)**, and $\|m\|_{\mathcal{H}'} < c_0 < +\infty$ imply

$$\mathbb{E}(\|W_i\|_{\mathcal{H}'}^l) \leq l! \sum_{k=0}^{l} C_l^k c_0^k (2c_1)^k c_2 c_1^{l-k} = l!\, c_2 \sum_{k=0}^{l} C_l^k (2c_0c_1)^k c_1^{l-k} = l!\, c_2 (c_1 + 2c_0c_1)^l.$$

Then $\mathbb{E}(\|W_i\|_{\mathcal{H}'}^l) \leq (l!/2) b_i^2 c^{l-2}$, where $b_i = \sqrt{2c_2}(c_1 + 2c_0c_1)$ and $c = c_1 + 2c_0c_1$. One can then apply the Yurinskii's exponential inequality (see Yurinskii, 1976) to obtain

$$\mathbb{P}\left(\left\|\sum_{i=1}^n W_i\right\|_{\mathcal{H}'} > xB_n\right) \leq 2\exp\left(-\frac{x^2}{2(1+1.62xc/B_n)}\right),$$

where $B_n = (\sum_{i=1}^n b_i^2)^{1/2} = \sqrt{n}\, c_3'$ with $c_3' = \sqrt{2c_2}(c_1 + 2c_0c_1)$. Hence,

$$\mathbb{P}(\|\Delta_n - \Delta\|_{\mathcal{H}'} > \xi) = \mathbb{P}\left(\left\|\sum_{i=1}^n W_i\right\|_{\mathcal{H}'} > \frac{\sqrt{n}\xi}{c_3'} B_n\right) \leq 2\exp\left(-\frac{\xi^2 n}{2\, c_3'(c_3' + c_4'\xi)}\right),$$

where c_4' is a positive finite constant. ■

2.5.2 Proof of Theorem 2.3.14

Theorem 2.3.14 is easily shown using Lemma 2.5.2 (see page 49) with $\gamma_j = \hat{\lambda}_j^{-1}$ and $w_j = \hat{v}_j$. In this case,

$$R_{k_n}^{(\gamma,w)} = \theta - \sum_{j=1}^{k_n} \hat{\lambda}_j^{-1} \langle \Gamma_n \hat{v}_j, \theta\rangle \hat{v}_j = \hat{R}_{k_n},$$

where $\hat{R}_{k_n}$ is defined in (2.7) (see page 39). Therefore, applying Lemma 2.5.2, one has

$$\begin{aligned}\mathbb{E}_{\mathcal{X}^{n+1}}(Y_{n+1} - \langle \hat{\theta}_{k_n}, X_{n+1}\rangle)^2 &= \sigma^2 + \frac{\sigma^2}{n}\sum_{j_1=1}^{k_n}\sum_{j_2=1}^{k_n} \frac{\langle \Gamma_n \hat{v}_{j_1}, \hat{v}_{j_2}\rangle}{\hat{\lambda}_{j_1}\hat{\lambda}_{j_2}} \langle X_{n+1}, \hat{v}_{j_1}\rangle\langle X_{n+1}, \hat{v}_{j_2}\rangle + \langle X_{n+1}, \hat{R}_{k_n}\rangle^2 \\ &= \sigma^2 + \frac{\sigma^2}{n}\sum_{j=1}^{k_n} \frac{\langle X_{n+1}, \hat{v}_j\rangle^2}{\hat{\lambda}_j} + \langle X_{n+1}, \hat{R}_{k_n}\rangle^2\end{aligned}$$

for the conditional prediction error, and one gets

$$\mathbb{E}_{\mathcal{X}^n}(\|\theta - \hat{\theta}_{k_n}\|^2) = \frac{\sigma^2}{n}\sum_{j_1=1}^{k_n}\sum_{j_2=1}^{k_n} \frac{\langle \Gamma_n \hat{v}_{j_1}, \hat{v}_{j_2}\rangle}{\hat{\lambda}_{j_1}\hat{\lambda}_{j_2}} \langle \hat{v}_{j_1}, \hat{v}_{j_2}\rangle + \|\hat{R}_{k_n}\|^2 = \frac{\sigma^2}{n}\sum_{j=1}^{k_n} \frac{1}{\hat{\lambda}_j} + \|\hat{R}_{k_n}\|^2$$

for the conditional estimation error.

2.5.3 Formulation and proof of Lemma 2.5.2

Next, a general result providing expressions for the conditional errors of a wide class of estimators will be presented. The high degree of generality of this lemma has an advantage: it allows deriving the conditional errors of the standard FPCA estimator and other FPCA–type estimators as almost direct corollaries.

Lemma 2.5.2. *Considering the regression model* (2.1) *(see page 32), let* $\hat{\theta}$ *be an estimator for* θ *such that*

$$\hat{\theta} = \sum_{j=1}^{k_n} \gamma_j \Delta_n w_j w_j,$$

where $\{(\gamma_j, w_j)\}_j \subset \mathbb{R} \times \mathcal{H}$ *only depend on* $\mathcal{X}^n = \{X_1, \dots, X_n\}$*. Then, it holds that*

$$\mathbb{E}_{\mathcal{X}^{n+1}}(Y_{n+1} - \langle \hat{\theta}, X_{n+1} \rangle)^2 = \sigma^2 + \frac{\sigma^2}{n} \sum_{j_1=1}^{k_n} \sum_{j_2=1}^{k_n} \gamma_{j_1} \gamma_{j_2} \langle \Gamma_n w_{j_1}, w_{j_2} \rangle \langle X_{n+1}, w_{j_1} \rangle \langle X_{n+1}, w_{j_2} \rangle + \langle X_{n+1}, R_{k_n}^{(\gamma,w)} \rangle^2,$$

and

$$\mathbb{E}_{\mathcal{X}^n}(\|\theta - \hat{\theta}\|^2) = \frac{\sigma^2}{n} \sum_{j_1=1}^{k_n} \sum_{j_2=1}^{k_n} \gamma_{j_1} \gamma_{j_2} \langle \Gamma_n w_{j_1}, w_{j_2} \rangle \langle w_{j_1}, w_{j_2} \rangle + \|R_{k_n}^{(\gamma,w)}\|^2,$$

where $R_{k_n}^{(\gamma,w)} = \theta - \sum_{j=1}^{k_n} \gamma_j \langle \Gamma_n w_j, \theta \rangle w_j$*.*

Proof. Observe that $\Delta_n x = \langle \Gamma_n x, \theta \rangle + \Delta_n^\epsilon x$ for all $x \in \mathcal{H}$, with $\Delta_n^\epsilon = n^{-1} \sum_{i=1}^n X_i \otimes_{\mathcal{H}'} \epsilon_i$. Consequently, the difference $\theta - \hat{\theta}$ can be expressed as

$$\theta - \hat{\theta} = R_{k_n}^{(\gamma,w)} - \sum_{j=1}^{k_n} \gamma_j \Delta_n^\epsilon w_j w_j, \quad \text{with} \quad R_{k_n}^{(\gamma,w)} = \theta - \sum_{j=1}^{k_n} \gamma_j \langle \Gamma_n w_j, \theta \rangle w_j. \tag{2.11}$$

Then the regression model (2.1), (2.11), and conditions on ϵ imply that

$$\begin{aligned}
\mathbb{E}_{\mathcal{X}^{n+1}}(Y_{n+1} - \langle \hat{\theta}, X_{n+1} \rangle)^2 &= \mathbb{E}_{\mathcal{X}^{n+1}} \left(\epsilon_{n+1} + \left\langle X_{n+1}, R_{k_n}^{(\gamma,w)} - \sum_{j=1}^{k_n} \gamma_j \Delta_n^\epsilon w_j w_j \right\rangle \right)^2 \\
&= \sigma^2 + \sum_{j_1=1}^{k_n} \sum_{j_2=1}^{k_n} \gamma_{j_1} \gamma_{j_2} \mathbb{E}_{\mathcal{X}^n}(\Delta_n^\epsilon w_{j_1} \Delta_n^\epsilon w_{j_2}) \langle X_{n+1}, w_{j_1} \rangle \langle X_{n+1}, w_{j_2} \rangle \\
&\quad - 2 \langle X_{n+1}, R_{k_n}^{(\gamma,w)} \rangle \left\langle X_{n+1}, \sum_{j=1}^{k_n} \gamma_j \mathbb{E}_{\mathcal{X}^n}(\Delta_n^\epsilon w_j) w_j \right\rangle + \langle X_{n+1}, R_{k_n}^{(\gamma,w)} \rangle^2.
\end{aligned}$$

Now, note that

$$\mathbb{E}_{\mathcal{X}^n}(\Delta_n^\epsilon w_j) = 0, \quad \mathbb{E}_{\mathcal{X}^n}(\Delta_n^\epsilon w_{j_1} \Delta_n^\epsilon w_{j_2}) = \frac{\sigma^2}{n} \langle \Gamma_n w_{j_1}, w_{j_2} \rangle. \tag{2.12}$$

Therefore, using (2.12), one has

$$\mathbb{E}_{\mathcal{X}^{n+1}}(Y_{n+1} - \langle \hat{\theta}, X_{n+1} \rangle)^2 = \sigma^2 + \frac{\sigma^2}{n} \sum_{j_1=1}^{k_n} \sum_{j_2=1}^{k_n} \gamma_{j_1} \gamma_{j_2} \langle \Gamma_n w_{j_1}, w_{j_2} \rangle \langle X_{n+1}, w_{j_1} \rangle \langle X_{n+1}, w_{j_2} \rangle + \langle X_{n+1}, R_{k_n}^{(\gamma,w)} \rangle^2.$$

Analogously, for the conditional estimation error, (2.11) gives

$$\mathbb{E}_{\mathcal{X}^n}(\|\theta - \hat{\theta}\|^2) = \sum_{j_1=1}^{k_n} \sum_{j_2=1}^{k_n} \gamma_{j_1} \gamma_{j_2} \mathbb{E}_{\mathcal{X}^n}(\Delta_n^\epsilon w_{j_1} \Delta_n^\epsilon w_{j_2}) \langle w_{j_1}, w_{j_2} \rangle - 2 \left\langle R_{k_n}^{(\gamma,w)}, \sum_{j=1}^{k_n} \gamma_j \mathbb{E}_{\mathcal{X}^n}(\Delta_n^\epsilon w_j) w_j \right\rangle + \|R_{k_n}^{(\gamma,w)}\|^2,$$

and, applying (2.12) again,

$$\mathbb{E}_{\mathcal{X}^n}(\|\theta - \hat{\theta}\|^2) = \frac{\sigma^2}{n} \sum_{j_1=1}^{k_n} \sum_{j_2=1}^{k_n} \gamma_{j_1}\gamma_{j_2}\langle \Gamma_n w_{j_1}, w_{j_2}\rangle\langle w_{j_1}, w_{j_2}\rangle + \|R_{k_n}^{(\gamma,w)}\|^2.$$

■

Chapter 3

Presmoothing in functional linear regression

In this chapter, the functional linear model with scalar response and explanatory variable valued in a functional space is considered. In recent statistical literature, FPCA has been used to estimate the model functional parameter. This approach can be modified by using presmoothing techniques: either presmoothing via covariance structure or presmoothing via response variable. For these new estimators, consistency is stated and efficiency by comparison with the standard FPCA estimator is studied from a theoretical point of view. Furthermore, the finite sample performance of the proposed presmoothed estimators is also analysed by means of a simulation study and three real data applications. Finally, the proofs of the main results in the chapter and some technical lemmas are gathered together in the appendix.

The first tools for developing the methodology which is compiled in this chapter were presented in Martínez-Calvo (2008). Later, the complete research on the approach based on presmoothing via covariance structure was published in Ferraty et al. (2012a)

3.1 Why introduce presmoothing techniques?

In the previous chapter, it was showed how classical multivariate methods have been adapted to the functional context where the response Y and/or the explanatory variable X are valued in a functional space. Particularly, the functional linear model with scalar response was specified for being the subject of several studies in the recent literature, as it is in this chapter. Recall that, given a real separable Hilbert space $(\mathcal{H}, \langle\cdot,\cdot\rangle)$ ($\|\cdot\|$ denotes the induced norm), the functional linear model with scalar response was introduced in (2.1) (see Chapter 2, page 32) as

$$Y = m(X) + \epsilon = \langle\theta, X\rangle + \epsilon,$$

where Y is a real random variable, $m(\cdot) = \langle\theta,\cdot\rangle$ is a linear regression operator such that $\theta \in \mathcal{H}$ and $\|\theta\|^2 < \infty$, X is a zero–mean random variable valued in $\mathcal{H}$ satisfying that $\mathbb{E}(\|X\|^2) < \infty$, and ϵ is a real random variable such that $\mathbb{E}(\epsilon) = 0$, $\mathrm{Var}(\epsilon) = \sigma^2$, and $\mathbb{E}(\epsilon X) = 0$. Remark 2.3.1 (see Chapter 2, page 33) showed that there is no loss of generality in the consideration of zero–mean variables in the model, whereas Remark 2.3.2 (see Chapter 2, page 33) provides the particular expression of the model when $\mathcal{H} = L^2([0,1])$.

As it was commented in Chapter 2, estimators for the functional linear model with scalar response can be obtained based on FPCA (Cardot et al., 1999, 2003c; Cai and Hall, 2006; Hall and Hosseini-Nasab, 2006; Hall and Horowitz, 2007; Cardot et al., 2007c). Since the FPCA estimator is reintroduced in this chapter, recall that when **(C.2.1)** holds (see Chapter 2, page 36), the model parameter can be

expressed as $\theta = \sum_{j=1}^{\infty} \lambda_j^{-1} \Delta v_j v_j$. Therefore, given $\{(X_i, Y_i)\}_{i=1}^n$ a random sample of i.i.d. variables drawn from (X, Y), the standard FPCA estimator was defined in (2.4) (see Chapter 2, page 36) by

$$\hat{\theta}_{k_n} = \sum_{j=1}^{k_n} \frac{\Delta_n \hat{v}_j}{\hat{\lambda}_j} \hat{v}_j,$$

where $\{(\hat{\lambda}_j, \hat{v}_j)\}_{j=1}^{\infty}$ are the eigenvalues and eigenfunctions of the empirical second moment operator $\Gamma_n = n^{-1} \sum_{i=1}^n X_i \otimes_{\mathcal{H}} X_i$, $\Delta_n = n^{-1} \sum_{i=1}^n X_i \otimes_{\mathcal{H}'} Y_i$ is the cross second moment operator, and $\{k_n\}_{n=1}^{\infty}$ is a sequence of positive integers such that $k_n \to +\infty$, $k_n \leq n$, and $\hat{\lambda}_{k_n} > 0$. Further details related to this estimator can be found in Section 2.3.2 (see Chapter 2, page 35), for instance, the normal equation that $\hat{\theta}_{k_n}$ satisfies (see Remark 2.3.6 in Chapter 2, page 36), the alternative construction of $\hat{\theta}_{k_n}$ as the projection of the model parameter onto the subspace spanned by $\hat{v}_1, \ldots, \hat{v}_{k_n}$ (see Chapter 2, page 36), or the expressions of the conditional errors for $\hat{\theta}_{k_n}$ (see Theorem 2.3.14 in Chapter 2, page 39). In this chapter, the FPCA estimator is revisited in order to improve its behaviour in terms of conditional mean square errors by introducing presmoothing techniques.

The choice of presmoothing methods was decided on in light of Faraldo-Roca and González-Manteiga (1987) and Cristóbal-Cristóbal et al. (1987). These authors proposed the application of the least squares principle on the pairs $(X_i, \hat{m}_n(X_i))$ instead of (X_i, Y_i), where $\hat{m}_n$ is a nonparametric kernel–type estimator of the regression function $m(\cdot)$. This alteration to the minimization problem produced efficient estimators that reduced the mean square error of the classical least squares estimators for the linear regression model, except in the compact support case. Later, Janssen et al. (2001) showed that the inefficiency problem in the compact support case could be rectified using *boundary kernels*. A similar procedure was developed by Akritas (1996) to fit polynomial regression models to data with incomplete observations. Since then, the presmoothing methods have been successfully applied in areas such as model selection procedures (see Aerts et al., 2010, who smoothed the response data prior to model selection by Akaike's Information Criterion), and censored/truncated survival data analysis (see Cao-Abad et al., 2005; Jácome and Iglesias-Pérez, 2008; Jácome et al., 2008, who replaced censoring indicator variables by values of a nonparametric regression estimator).

In FDA, presmoothing processes are usually included as preliminary steps in such a way that the observations are replaced by their smoothed approximations. For instance, Hitchcock et al. (2006) examined the effect of this substitution on estimating the dissimilarities among elements in the dataset. Another way to use presmoothing methods has been considered by Zhang and Chen (2007) when dealing with the regression model $Y_i(t) = X_i^t \theta(t) + V_i(t) + \epsilon_i(t)$, $i = 1, \ldots, n$, where the covariate X_i is multidimensional and independent of t (note that X_i^t denotes the transpose of the vector X_i), the process V_i represents the ith individual variation, and Y_i is the ith response process. These authors proposed an estimator of θ based on a local polynomial kernel reconstruction of $f_i(t) = X_i^t \theta(t) + V_i(t)$.

In this chapter, presmoothing is not used as a preprocessing tool, but rather as a way to build a new efficient FPCA–type estimator, that reduces the conditional mean square errors of the standard FPCA one, following the ideas of Faraldo-Roca and González-Manteiga (1987) in the real case. The key here is similar to the one that motivates multivariate ridge regression: circumvent the problem of an ill–conditioned covariance operator by means of an artificial perturbation of its eigenvalues. To see the usefulness of the proposed approach, here is an example that shows the instability of the FPCA estimator when the eigenvalues are close to 0; it will be analysed in detail in the simulation study (see Section 3.5).

Example. Consider $\mathcal{H} = L^2([0,1])$ and the linear regression model $Y = \int_0^1 \theta(t) X(t) dt + \epsilon$ (see Remark 2.3.2 in Chapter 2, page 33) where the explanatory curves are

$$X(t) = a_1 \sqrt{2} \sin(\pi t) + a_2 \sqrt{2} \cos(\pi t) + a_3 \sqrt{2} \sin(2\pi t) + a_4 \sqrt{2} \cos(2\pi t), \quad t \in [0,1]$$

with $a_l \sim \mathcal{U}(-1/3^{l-1}, 1/3^{l-1})$ for all $l \in \{1, \ldots, 4\}$, the model parameter is

$$\theta(t) = 2\sqrt{2} \cos(2\pi t), \quad t \in [0,1]$$

and $\epsilon \sim \mathcal{N}(0, \sigma^2)$ with $\sigma = 0.2\sqrt{\mathbb{E}(\langle X, \theta\rangle^2)}$. The calculation of the standard FPCA estimator involves the eigenelements of the second moment operator of X (see (2.4) in Chapter 2, page 36) and, in this example, it can be shown that only the first four eigenvalues are different from zero. To analyse the effect of null eigenvalues, 200 samples of 100 observations are simulated, and the mean square prediction and estimation errors (see (3.4), page 63) when $k_n \in \{1, \ldots, 8\}$ eigenelements of the second order operator are involved in the FPCA estimator are computed. Figure 3.1 presents the obtained results (see black solid line). The errors when the eigenvalues are slightly perturbed, adding $\alpha = 10^{-5}$ to them (see grey dashed line in Figure 3.1) are also calculated. It can be seen that the presence of null eigenvalues ($k_n > 4$), which hardly affects prediction error, considerably increases the estimation error of the FPCA estimator. Moreover, it seems that perturbation of eigenvalues allows to keep small estimation errors, even when null eigenvalues are involved. Thus, all these reasonings led to developing a FPCA–type estimator based on a presmoothing method that avoids the inconvenience of ill–conditioning.

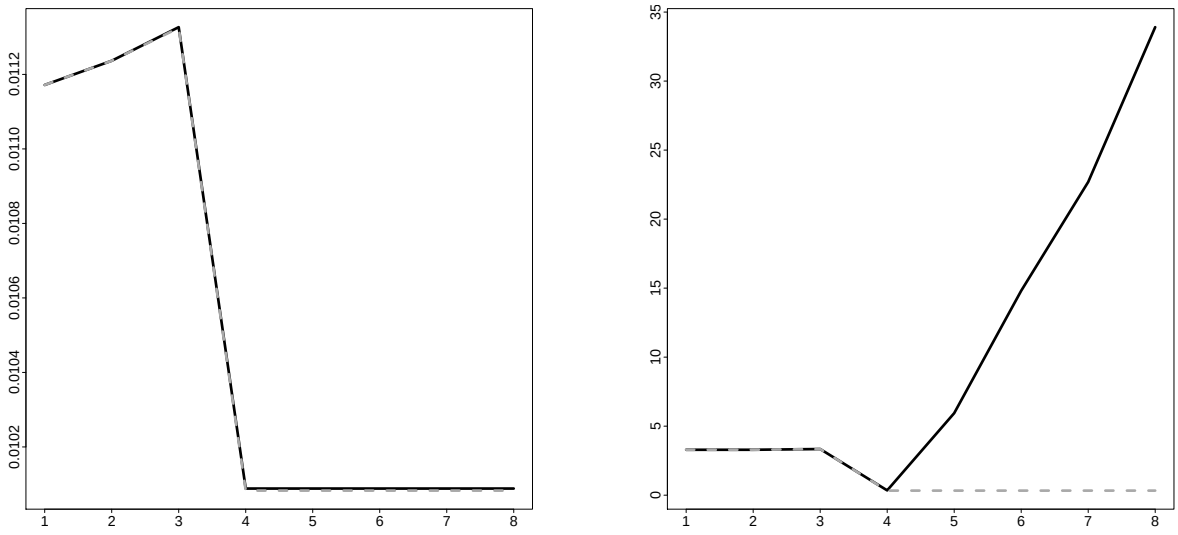

Figure 3.1: Example. Mean square prediction error (left panel) and estimated mean square estimation error (right panel) for the standard FPCA estimator (solid black line) and for the FPCA–type estimator with perturbed eigenvalues (dashed grey line).

In order to achieve this goal, a new estimator based on FPCA and presmoothing via covariance structure is proposed in Section 3.2. The section contains theoretical results (consistency and mean square error expressions) that allow to compare it with the standard FPCA estimator. Section 3.3 presents an alternative approach based on presmoothing via response variable (also including consistency and mean square error expressions), whereas Section 3.4 compiles some heuristics on other presmoothing methods. Section 3.5 includes a simulation study to analyse the behaviour of the different proposals from a practical point of view, and Section 3.6 is devoted to data applications. Some final comments can be found in Section 3.7, and an appendix compiles technical lemmas and the proofs of the main results in Section 3.8.

3.2 Presmoothing via covariance structure

3.2.1 Definition of estimator

By Remark 2.3.6 (see Chapter 2, page 36), the standard FPCA estimator $\hat{\theta}_{k_n}$ can be derived by solving the equation

$$\Delta_n x = \langle \beta, \Gamma_n x \rangle, \quad \forall x \in \mathrm{Im}(\Gamma_n),$$

and truncating the expansion of the solution to the first k_n components. The key of the proposed method is to perturb the previous normal equation, and find a function $\beta \in \mathcal{H}$ such that

$$\Delta_n x = \langle \beta, (\Gamma_n + \alpha_n 1_{\mathcal{H}})x \rangle, \quad \forall x \in \mathrm{Im}(\Gamma_n),$$

where α_n is a positive real sequence verifying that $\alpha_n \to 0$ when $n \to \infty$, and $1_{\mathcal{H}}$ is the identity operator in $\mathcal{H}$, i.e., $1_{\mathcal{H}} x = x$ for all $x \in \mathcal{H}$. From this equation, one can derive the next estimator for θ

$$\hat{\theta}_{k_n}^{\alpha_n} = \sum_{j=1}^{k_n} \frac{\Delta_n \hat{v}_j}{\hat{\lambda}_j + \alpha_n} \hat{v}_j, \tag{3.1}$$

which has the same structure as the standard FPCA estimator (2.4) (see Chapter 2, page 36), being the main difference the presence of perturbed empirical eigenvalues in the denominator.

Remark 3.2.1. Note that (3.1) can be seen as a functional version of the estimator for the ordinary multivariate ridge regression with penalization term equals to α_n times the usual norm of the model parameter.

Alternative construction. In order to build $\hat{\theta}_{k_n}^{\alpha_n}$, a procedure similar to the one proposed for $\hat{\theta}_{k_n}$ in Section 2.3.2 (see Chapter 2, page 35) can be followed. The optimization problem (2.5), that is,

$$\min_{\mathbf{b}} \mathbb{E}_{\mu_n}[(\hat{m}_{k_n}(\mathbf{X}) - \mathbf{X}^t \mathbf{b})^2], \quad \text{with} \quad \mathbf{b} = (\langle \beta, e_1 \rangle, \ldots, \langle \beta, e_{k_n} \rangle)^t, \forall \beta \in \mathcal{H},$$

can be solved with

- $\hat{m}_{k_n}(\mathbf{x}) = \sum_{i=1}^n Y_i \delta(\mathbf{x}, \mathbf{X}_i) / \sum_{i=1}^n \delta(\mathbf{x}, \mathbf{X}_i)$ is a nonparametric estimator of the regression function $m_{k_n}(\mathbf{x}) = \mathbb{E}(Y|\mathbf{X} = \mathbf{x})$, and
- $\mu_n(\mathbf{x}) = \int_{-\infty}^{\mathbf{x}} \hat{f}_n(\mathbf{t}) d\mathbf{t}$ is a weighting function where $\hat{f}_n$ is a nonparametric estimator of the density of $\mathbf{X}$, namely f, defined as $\hat{f}_n(\mathbf{x}) = n^{-1} \sum_{i=1}^n \delta(\mathbf{x}, \mathbf{X}_i)$,

using

$$\delta(\mathbf{u}, \mathbf{w}) = {h_n}^{-k_n} K^*(h_n^{-1}(\mathbf{u} - \mathbf{w})),$$

where $K^*(\mathbf{x}) = \prod_{j=1}^{k_n} K(x_j)$ for all $\mathbf{x} = (x_1, \ldots, x_{k_n})^t \in \mathbb{R}^{k_n}$, with $K : \mathbb{R} \to \mathbb{R}$ a symmetric positive kernel such that $\int K(z)dz = 1$, $\int zK(z)dz = 0$ and $\int z^2 K(z)dz = c(K) < \infty$, and h_n a strictly positive sequence of bandwidths (note that h_n usually satisfies $h_n \to 0$ and $nh_n \to \infty$). In this situation, $\beta_0 \in \mathcal{H}$ is a solution of (2.5) if and only if $\mathbf{b}_0 = (\langle \beta_0, e_1 \rangle, \ldots, \langle \beta_0, e_{k_n} \rangle)^t$ satisfies the associated k_n–dimensional normal equation $\mathbf{\Delta}_n = (\mathbf{\Gamma}_n + h_n^2 c(K) 1_{k_n \times k_n}) \mathbf{b}_0$, with $\mathbf{\Delta}_n = n^{-1} \sum_{i=1}^n \mathbf{X}_i Y_i$, $\mathbf{\Gamma}_n = n^{-1} \sum_{i=1}^n \mathbf{X}_i \mathbf{X}_i^t$, and $1_{k_n \times k_n}$ the $k_n \times k_n$–identity matrix (a $k_n \times k_n$–matrix with ones on the main diagonal and zeros elsewhere). This equation can be expressed as

$$\Delta_n e_l = \sum_{j=1}^{k_n} \langle \Gamma_n e_l, e_j \rangle \langle \beta_0, e_j \rangle + h_n^2 c(K) \langle \beta_0, e_l \rangle, \quad \text{for all } l \in \{1, \ldots, k_n\}.$$

By choosing $e_l = \hat{v}_l$, it holds that $\langle \beta_0, \hat{v}_l \rangle = (\Delta_n \hat{v}_l)/(\hat{\lambda}_l + h_n^2 c(K))$ and (3.1) can be derived with $\alpha_n = h_n^2 c(K)$.

3.2.2 Consistency

In order to state convergence results for the presmoothed FPCA estimator $\hat{\theta}_{k_n}^{\alpha_n}$, recall that $\| \cdot \|_{\mathcal{H}'}$ denotes the norm of the dual space $\mathcal{H}'$, and $\| \cdot \|_\infty$ denotes the uniform norm of the space of Hilbert–Schmidt operators defined on $\mathcal{H}$ (see Section 1.2.2, *"a) The space of Hilbert–Schmidt operators"* and *"b) The dual space $\mathcal{H}'$"*, in Chapter 1, page 9).

The following notation is also used in the next theorem: $m(\cdot) = \langle \theta, \cdot \rangle$ and $\hat{m}_{k_n}^{\alpha_n}(\cdot) = \langle \hat{\theta}_{k_n}^{\alpha_n}, \cdot \rangle$. Apart from the previously introduced assumptions for proving the consistency of the standard FPCA estimator (see further details in Section 2.3.2, *"c) Consistency"*, in Chapter 2, page 38), and extra assumption is required in this case:

(C.3.1) $\lambda_{k_n}^2/\alpha_n \to \infty$.

Theorem 3.2.2 (Ferraty et al., 2012a)**.** *Under the assumptions of Theorem 2.3.11 (see Chapter 2, page 39), if* **(C.3.1)** *is also satisfied, it holds that*

$$\|\hat{m}_{k_n}^{\alpha_n} - m\|_{\mathcal{H}'} \to 0 \quad a.s.$$

The proof of previous theorem can be found in the appendix of the chapter (see Section 3.8.1, page 71).

3.2.3 Conditional errors

Next, the conditional mean square prediction error and the conditional mean square estimation error for the presmoothed FPCA estimator $\hat{\theta}_{k_n}^{\alpha_n}$ are presented. In the theoretical results below, $Y_{n+1} = \langle \theta, X_{n+1}\rangle + \epsilon_{n+1}$ is a new response and $\mathbb{E}_{\mathcal{X}^n}(\cdot)$ is the expectation conditionally on $\mathcal{X}^n = \{X_1, \ldots, X_n\}$ for $n \in \mathbb{N}^*$. In addition, the next conditions will be necessary:

(C.3.2) $\hat{\lambda}_{k_n}/\alpha_n \to \infty$ *a.s.*

(C.3.3) $n\alpha_n \to 0$.

On the other hand, recall the $o_{a.s.}$ notation introduced in Definition 2.3.10 (see Chapter 2, page 38), and recall that $\hat{R}_{k_n}$ was defined in (2.7) (see Chapter 2, page 39) as

$$\hat{R}_{k_n} = \sum_{j>k_n} \langle \theta, \hat{v}_j\rangle \hat{v}_j.$$

Theorem 3.2.3 (Ferraty et al., 2012a)**.** *For the presmoothed FPCA estimator* (3.1)*, if* **(C.3.2)** *is satisfied, it holds that*

$$\begin{aligned}
&\mathbb{E}_{\mathcal{X}^{n+1}}(Y_{n+1} - \langle \hat{\theta}_{k_n}^{\alpha_n}, X_{n+1}\rangle)^2 - \mathbb{E}_{\mathcal{X}^{n+1}}(Y_{n+1} - \langle \hat{\theta}_{k_n}, X_{n+1}\rangle)^2 \\
&= \left(-2\alpha_n \frac{\sigma^2}{n} \sum_{j=1}^{k_n} \frac{\langle X_{n+1}, \hat{v}_j\rangle^2}{\hat{\lambda}_j^2} + 2\alpha_n \langle X_{n+1}, \hat{R}_{k_n}\rangle \langle X_{n+1}, \hat{T}_{k_n}\rangle + \alpha_n^2 \langle X_{n+1}, \hat{T}_{k_n}\rangle^2\right)(1 + o_{a.s.}(1)), \\
&\mathbb{E}_{\mathcal{X}^n}(\|\theta - \hat{\theta}_{k_n}^{\alpha_n}\|^2) - \mathbb{E}_{\mathcal{X}^n}(\|\theta - \hat{\theta}_{k_n}\|^2) = \left(-2\alpha_n \frac{\sigma^2}{n} \sum_{j=1}^{k_n} \frac{1}{\hat{\lambda}_j^2} + \alpha_n^2 \|\hat{T}_{k_n}\|^2\right)(1 + o_{a.s.}(1)),
\end{aligned}$$

where $\hat{T}_{k_n} = \sum_{j=1}^{k_n} \hat{\lambda}_j^{-1} \langle \theta, \hat{v}_j\rangle \hat{v}_j$*, and* $\hat{R}_{k_n}$ *is defined in* (2.7) *(see Chapter 2, page 39).*

Corollary 3.2.4 (Ferraty et al., 2012a)**.** *Under the assumptions of Theorem 3.2.3, if* **(C.3.3)** *is satisfied, it holds that*

$$\begin{aligned}
&\mathbb{E}_{\mathcal{X}^{n+1}}(Y_{n+1} - \langle \hat{\theta}_{k_n}^{\alpha_n}, X_{n+1}\rangle)^2 - \mathbb{E}_{\mathcal{X}^{n+1}}(Y_{n+1} - \langle \hat{\theta}_{k_n}, X_{n+1}\rangle)^2 \\
&= \left(-2\alpha_n \frac{\sigma^2}{n} \sum_{j=1}^{k_n} \frac{\langle X_{n+1}, \hat{v}_j\rangle^2}{\hat{\lambda}_j^2} + 2\alpha_n \langle X_{n+1}, \hat{R}_{k_n}\rangle \langle X_{n+1}, \hat{T}_{k_n}\rangle\right)(1 + o_{a.s.}(1)), \\
&\mathbb{E}_{\mathcal{X}^n}(\|\theta - \hat{\theta}_{k_n}^{\alpha_n}\|^2) - \mathbb{E}_{\mathcal{X}^n}(\|\theta - \hat{\theta}_{k_n}\|^2) = \left(-2\alpha_n \frac{\sigma^2}{n} \sum_{j=1}^{k_n} \frac{1}{\hat{\lambda}_j^2}\right)(1 + o_{a.s.}(1)).
\end{aligned}$$

The proof of the previous results can be found in the appendix of this chapter (see Section 3.8.3, page 74, and Section 3.8.4, page 75). They are derived from the general Lemma 2.5.2 (see Chapter 2, page 49).

Remark 3.2.5. Corollary 3.2.4 shows that the bias term $\hat{R}_{k_n}$ plays a fundamental role in the conditional mean square error for prediction. In fact, the presmoothed estimator gives better or worse results than the standard FPCA estimator $\hat{\theta}_{k_n}$ depending on the order of $\langle X_{n+1}, \hat{R}_{k_n}\rangle$. Thus if $\langle X_{n+1}, \hat{R}_{k_n}\rangle(\alpha_n\langle X_{n+1}, \hat{T}_{k_n}\rangle)^{-1} = o_{a.s.}(1)$, then

$$\mathbb{E}_{\mathcal{X}^{n+1}}(Y_{n+1} - \langle\hat{\theta}_{k_n}^{\alpha_n}, X_{n+1}\rangle)^2 - \mathbb{E}_{\mathcal{X}^{n+1}}(Y_{n+1} - \langle\hat{\theta}_{k_n}, X_{n+1}\rangle)^2 = \left(-2\alpha_n\frac{\sigma^2}{n}\sum_{j=1}^{k_n}\frac{\langle X_{n+1}, \hat{v}_j\rangle^2}{\hat{\lambda}_j^2} + \alpha_n^2\langle X_{n+1}, \hat{T}_{k_n}\rangle^2\right)(1 + o_{a.s.}(1)).$$

With $f(\alpha_n) = -2\alpha_n n^{-1}\sigma^2\sum_{j=1}^{k_n}\hat{\lambda}_j^{-2}\langle X_{n+1}, \hat{v}_j\rangle^2 + \alpha_n^2\langle X_{n+1}, \hat{T}_{k_n}\rangle^2$, take $\alpha_n^{opt,1} = \arg\min_\alpha f(\alpha)$. It can be shown that $\alpha_n^{opt,1} = n^{-1}\sigma^2\sum_{j=1}^{k_n}\hat{\lambda}_j^{-2}\langle X_{n+1}, \hat{v}_j\rangle^2\langle X_{n+1}, \hat{T}_{k_n}\rangle^{-2}$ and, in this case

$$\mathbb{E}_{\mathcal{X}^{n+1}}(Y_{n+1} - \langle\hat{\theta}_{k_n}^{\alpha_n^{opt,1}}, X_{n+1}\rangle)^2 - \mathbb{E}_{\mathcal{X}^{n+1}}(Y_{n+1} - \langle\hat{\theta}_{k_n}, X_{n+1}\rangle)^2 = \left(-\frac{\sigma^4}{n^2}\left(\sum_{j=1}^{k_n}\frac{\langle X_{n+1}, \hat{v}_j\rangle^2}{\hat{\lambda}_j^2}\right)^2\langle X_{n+1}, \hat{T}_{k_n}\rangle^{-2}\right)(1 + o_{a.s.}(1)).$$

Hence, there is a decrease in the prediction error using $\hat{\theta}_{k_n}^{\alpha_n^{opt,1}}$, and this reduction is more important when σ^2 is large and/or sample size n is small.

Remark 3.2.6. Under the assumptions of Corollary 3.2.4, one has that $\hat{\theta}_{k_n}^{\alpha_n}$ improves against $\hat{\theta}_{k_n}$ in terms of the conditional mean square error for estimation, when above all, σ^2 is large and/or n is small. Moreover, one can look for the value of α_n that minimizes $g(\alpha_n) = -2\alpha_n n^{-1}\sigma^2\sum_{j=1}^{k_n}\hat{\lambda}_j^{-2} + \alpha_n^2\|\hat{T}_{k_n}\|^2$. This is $\alpha_n^{opt,2} = n^{-1}\sigma^2(\sum_{j=1}^{k_n}\hat{\lambda}_j^{-2})\|\hat{T}_{k_n}\|^{-2}$, for which one gets

$$\mathbb{E}_{\mathcal{X}^n}(\|\theta - \hat{\theta}_{k_n}^{\alpha_n^{opt,2}}\|^2) - \mathbb{E}_{\mathcal{X}^n}(\|\theta - \hat{\theta}_{k_n}\|^2) = \left(-\frac{\sigma^4}{n^2}\left(\sum_{j=1}^{k_n}\frac{1}{\hat{\lambda}_j^2}\right)^2\|\hat{T}_{k_n}\|^{-2}\right)(1 + o_{a.s.}(1)).$$

Previous remarks ensure second order efficiency, that is, $\hat{\theta}_{k_n}^{\alpha_n}$ performs better than the standard FPCA estimator in small samples, and the same as $\hat{\theta}_{k_n}$ in large ones. Second order efficiency was already achieved for linear regression estimators based on presmoothing in the real case (Faraldo-Roca and González-Manteiga, 1987; Janssen et al., 2001). Presmoothing techniques also allowed this kind of gain in efficiency in other contexts (see the presmoothed Nelson–Aalen estimator versus the classical Nelson–Aalen estimator in Cao-Abad et al., 2005).

3.3 Presmoothing via response variable

3.3.1 Definition of estimator

Another possible way to combine presmoothing techniques and FPCA estimators is to solve the following normal equation

$$\Delta_n^{h_n}x = \langle\beta, \Gamma_n x\rangle, \quad \forall x \in \mathrm{Im}(\Gamma_n),$$

where $\Delta_n^{h_n} = n^{-1}\sum_{i=1}^n X_i \otimes_{\mathcal{H}'} \hat{m}_{h_n}(X_i)$, with $\hat{m}_{h_n}(\cdot)$ the nonparametric kernel estimator introduced by Ferraty and Vieu (2006b), and defined in (2.10) (see Chapter 2, page 42) as

$$\hat{m}_{h_n}(x) = \frac{\sum_{i=1}^n Y_i K(h_n^{-1}d(x, X_i))}{\sum_{i=1}^n K(h_n^{-1}d(x, X_i))},$$

being now $K(\cdot)$ an asymmetric kernel and h_n a strictly positive real bandwidth such that $h_n \to 0$ when $n \to +\infty$. This equation leads to the following presmoothed estimator for θ

$$\hat{\theta}_{k_n}^{h_n} = \sum_{j=1}^{k_n} \frac{\Delta_n^{h_n} \hat{v}_j}{\hat{\lambda}_j} \hat{v}_j. \tag{3.2}$$

Alternative construction. As it was done for $\hat{\theta}_{k_n}$ and $\hat{\theta}_{k_n}^{\alpha_n}$, the estimator $\hat{\theta}_{k_n}^{h_n}$ can also be obtained by means of the multivariate optimization problem (2.5) (see Chapter 2, page 36) given by

$$\min_{\mathbf{b}} \mathbb{E}_{\mu_n}[(\hat{m}_{k_n}(\mathbf{X}) - \mathbf{X}^t\mathbf{b})^2], \quad \text{with} \quad \mathbf{b} = (\langle \beta, e_1 \rangle, \ldots, \langle \beta, e_{k_n} \rangle)^t, \forall \beta \in \mathcal{H},$$

where

- $\hat{m}_{k_n}(\mathbf{x}) = \sum_{i=1}^n Y_i \delta(\mathbf{x}, \mathbf{X}_i) / \sum_{i=1}^n \delta(\mathbf{x}, \mathbf{X}_i)$ is a nonparametric estimator of the regression function $m_{k_n}(\mathbf{x}) = \mathbb{E}(Y|\mathbf{X} = \mathbf{x})$, whereas, in this case,
- $\mu_n(\mathbf{x}) = n^{-1} \sum_{i=1}^n \mathbb{I}_{\{\mathbf{X}_i \in (-\infty, \mathbf{x}]\}}$ is the weighting function, being $\mathbb{I}$ is the indicator function.

Furthermore, it can be selected again

$$\delta(\mathbf{u}, \mathbf{w}) = h_n^{-k_n} K^*(h_n^{-1}(\mathbf{u} - \mathbf{w})),$$

with $K^*(\mathbf{x}) = \prod_{j=1}^{k_n} K(x_j)$, being $K(\cdot)$ a symmetric positive kernel which verifies that $\int K(z)dz = 1$, $\int zK(z)dz = 0$ and $\int z^2 K(z)dz = c(K) < \infty$, and h_n a strictly positive sequence of bandwidths such that $h_n \to 0$ and $nh_n \to \infty$. Consequently, $\beta_0 \in \mathcal{H}$ is a solution of (2.5) if and only if $\mathbf{b}_0 = (\langle \beta_0, e_1 \rangle, \ldots, \langle \beta_0, e_{k_n} \rangle)^t$ satisfies the associated k_n–dimensional normal equation $\mathbf{\Delta}_n^{h_n} = \mathbf{\Gamma}_n \mathbf{b}_0$, where

$$\mathbf{\Delta}_n^{h_n} = n^{-1} \sum_{i=1}^n \mathbf{X}_i \hat{m}_{k_n}(\mathbf{X}_i) \quad \text{and} \quad \mathbf{\Gamma}_n = n^{-1} \sum_{i=1}^n \mathbf{X}_i \mathbf{X}_i^t.$$

This fact ensures that

$$\Delta_n^{h_n} e_l = \sum_{j=1}^{k_n} \langle \Gamma_n e_l, e_j \rangle \langle \beta_0, e_j \rangle, \quad \text{for all } l \in \{1, \ldots, k_n\}.$$

If $e_l = \hat{v}_l$, one gets $\langle \beta_0, \hat{v}_l \rangle = (\Delta_n^{h_n} \hat{v}_l)/\hat{\lambda}_l$, being the associated estimator the one presented in (3.2).

k–NN approach. Although all the theoretical advances below involve $\hat{\theta}_{k_n}^{h_n}$, which is based on the kernel estimator $\hat{m}_{h_n}$, a different approach in the simulation studies has been used (see Section 3.5, page 63): a *k–nearest neighbours estimator* (k–NN). The key idea of the k–NN procedure is to replace the global bandwidth h_n by a local bandwidth $h_k(x)$ which ensures that, for any x, a fixed number k of observations are taken into account to calculate the value of the estimator, that is,

$$\hat{m}_{\text{k–NN}}(x) = \frac{\sum_{i=1}^n Y_i K(h_k(x)^{-1} d(x, X_i))}{\sum_{i=1}^n K(h_k(x)^{-1} d(x, X_i))},$$

where $h_k(x)$ is a bandwidth such that there are exactly k elements in the sample which verify that $d(x, X_i) < h_k(x)$. Hence, the next estimator for θ can be defined

$$\hat{\theta}_{k_n}^{\text{k–NN}} = \sum_{j=1}^{k_n} \frac{\Delta_n^{\text{k–NN}} \hat{v}_j}{\hat{\lambda}_j} \hat{v}_j, \tag{3.3}$$

with $\Delta_n^{\text{k–NN}} = n^{-1} \sum_{i=1}^n X_i \otimes_{\mathcal{H}'} \hat{m}_{\text{k–NN}}(X_i)$.

Remark 3.3.1. The estimator $\hat{\theta}_{k_n}^{h_n}$ have been considered in order to state convergence and other theoretical results, due to the easiness of calculations when the bandwidth is global and independent of x. Nevertheless, some previous simulations revealed that $\hat{\theta}_{k_n}^{\text{k–NN}}$ provides more reduced errors than $\hat{\theta}_{k_n}^{h_n}$, as can be seen in the following example. This is the reason why the results for $\hat{\theta}_{k_n}^{\text{k–NN}}$ are included in the simulation studies (see Section 3.5, page 63) instead of the results obtained by $\hat{\theta}_{k_n}^{h_n}$.

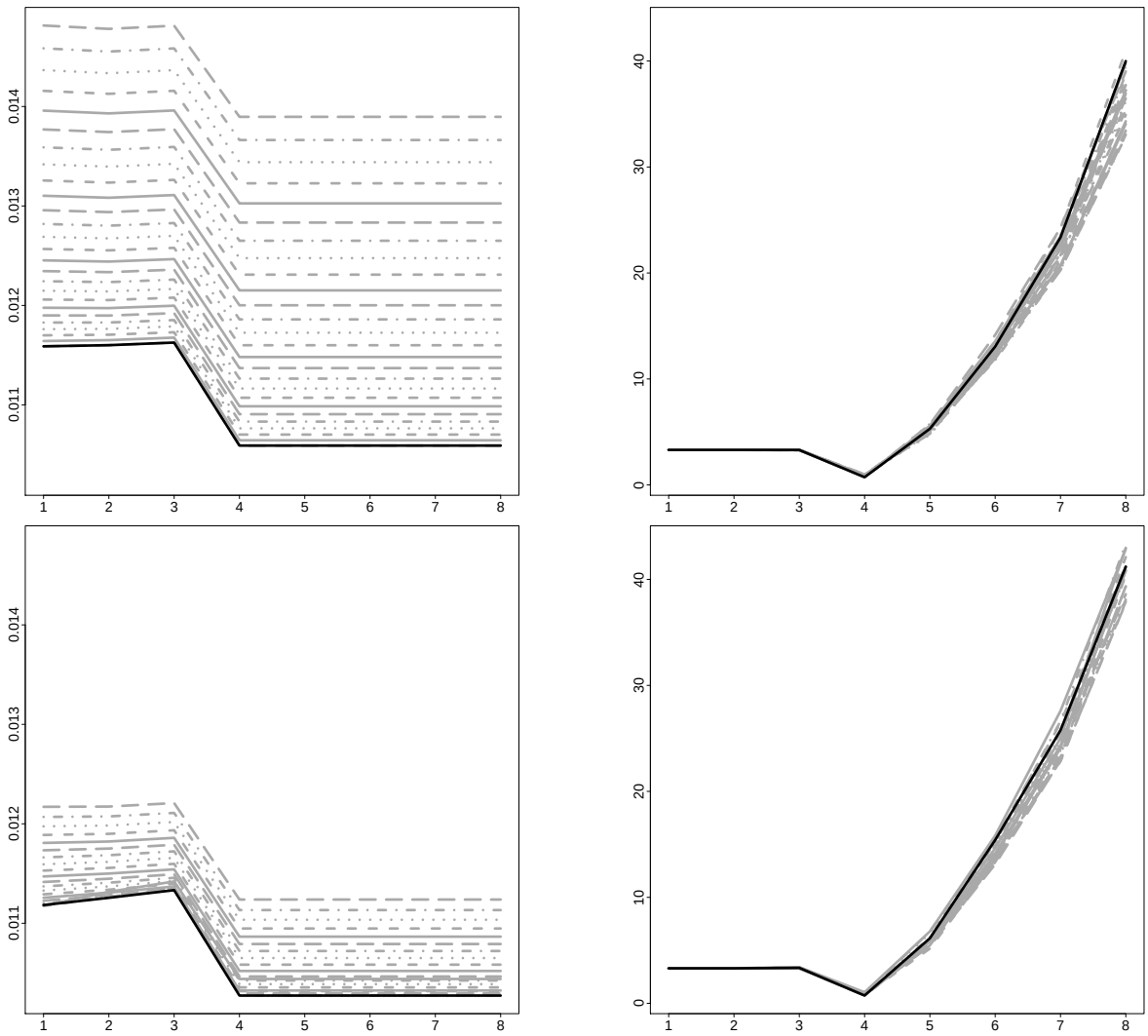

Figure 3.2: Example. Mean square prediction error (left panels) and estimated mean square estimation error (right panels) for $\hat{\theta}_{k_n}^{h_n}$ (first row) and for $\hat{\theta}_{k_n}^{\text{k–NN}}$ (second row). Each grey line corresponds to a different value of the smoothing parameter (i.e., h_n for $\hat{\theta}_{k_n}^{h_n}$ and $neig$ for $\hat{\theta}_{k_n}^{\text{k–NN}}$), whereas solid black line corresponds to curve which gives the minimum value for each error.

Example. Recall the example introduced at the beginning of this chapter in which

$$X(t) = a_1\sqrt{2}\sin(\pi t) + a_2\sqrt{2}\cos(\pi t) + a_3\sqrt{2}\sin(2\pi t) + a_4\sqrt{2}\cos(2\pi t), \quad t \in [0,1]$$

being $a_l \sim \mathcal{U}(-1/3^{l-1}, 1/3^{l-1})$ for all $l \in \{1,\ldots,4\}$, the model parameter is $\theta(t) = 2\sqrt{2}\cos(2\pi t)$ and $\epsilon \sim \mathcal{N}(0,\sigma^2)$ with $\sigma = 0.2\sqrt{\mathbb{E}(\langle X,\theta\rangle^2)}$. From the model $Y = \int_0^1 \theta(t)X(t)dt + \epsilon$, 200 samples of 100 observations are generated, and $\hat{\theta}_{k_n}^{h_n}$ and $\hat{\theta}_{k_n}^{\text{k–NN}}$ are computed for $k_n \in \{1,\ldots,8\}$, $neig \in \{2,\ldots,21\}$, and h_n valued in a grid in $[0.75, 2]$. For both $\hat{m}_{h_n}$ and $\hat{m}_{\text{k–NN}}$, the quadratic kernel and the semi–metric $d(\cdot,\cdot)$ based on the first derivative are selected. For each sample, the mean square prediction error and the mean square estimation error (see (3.4), page 63) are obtained, and the mean of these errors over the 200 samples is plotted in Figure 3.2, where each curve corresponds to a different value of h_n (respectively, $neig$). Furthermore, the curves where the minimum value is reached are highlighted by means of a solid black line. It seems that both estimators give similar results in terms of the estimation error, whereas the prediction errors of the k–NN version are generally smaller than errors of $\hat{\theta}_{k_n}^{h_n}$. On

the other hand, both estimators are suffering the effect on estimation error of the null eigenvalues when $k_n > 4$, unlike the first presmoothing proposal $\hat{\theta}_{k_n}^{\alpha_n}$ (see Figure 3.1, page 53).

3.3.2 Consistency

The following theorem can be derived for $m(\cdot) = \langle \theta, \cdot \rangle$ and $\hat{m}_{k_n}^{h_n}(\cdot) = \langle \hat{\theta}_{k_n}^{h_n}, \cdot \rangle$ using the same notation as in the previous section and the next conditions

(C.3.4) $\mathbb{P}(X \in \mathcal{C}) = 1$ being $\mathcal{C}$ a compact subset of $\mathcal{H}$ satisfying **(C.2.21)** (see hypotheses of Theorem 2.4.11 in Chapter 2, page 45),

(C.3.5) $\lambda_{k_n}/h_n^{\beta} \to +\infty$ and $\lambda_{k_n}\sqrt{n\phi(h_n)/\log n} \to +\infty$ (being $\phi(\cdot)$ defined in **(C.2.23)** in Chapter 2, page 45).

Theorem 3.3.2. *Under the assumptions of Theorem 2.3.11 (see Chapter 2, page 39) and Theorem 2.4.11 (see Chapter 2, page 45), if* **(C.3.4)**–**(C.3.5)** *are also satisfied, it holds that*

$$\|\hat{m}_{k_n}^{h_n} - m\|_{\mathcal{H}'} \to 0 \quad a.s.$$

The proof of Theorem 3.3.2 can be found in the appendix of the chapter (see Section 3.8.5, page 76).

3.3.3 Conditional errors

The following theoretical advances introduce the expressions of the conditional errors of prediction and estimation for the presmoothed FPCA estimator $\hat{\theta}_{k_n}^{h_n}$.

Theorem 3.3.3. *For the presmoothed FPCA estimator* (3.2)*, it holds that*

$$\begin{aligned}
&\mathbb{E}_{\mathcal{X}^{n+1}}(Y_{n+1} - \langle \hat{\theta}_{k_n}^{h_n}, X_{n+1}\rangle)^2 - \mathbb{E}_{\mathcal{X}^{n+1}}(Y_{n+1} - \langle \hat{\theta}_{k_n}, X_{n+1}\rangle)^2 \\
&= \sum_{j_1=1}^{k_n}\sum_{j_2=1}^{k_n} \frac{U_n^{h_n}(\hat{v}_{j_1})U_n^{h_n}(\hat{v}_{j_2})}{\hat{\lambda}_{j_1}\hat{\lambda}_{j_2}}\langle X_{n+1}, \hat{v}_{j_1}\rangle\langle X_{n+1}, \hat{v}_{j_2}\rangle \\
&\quad + \sigma^2 \sum_{j_1=1}^{k_n}\sum_{j_2=1}^{k_n} \frac{W_n^{h_n}(\hat{v}_{j_1}, \hat{v}_{j_2})}{\hat{\lambda}_{j_1}\hat{\lambda}_{j_2}}\langle X_{n+1}, \hat{v}_{j_1}\rangle\langle X_{n+1}, \hat{v}_{j_2}\rangle - \frac{\sigma^2}{n}\sum_{j=1}^{k_n}\frac{\langle X_{n+1}, \hat{v}_j\rangle^2}{\hat{\lambda}_j} \\
&\quad - 2\langle X_{n+1}, \hat{R}_{k_n}\rangle\langle X_{n+1}, \sum_{j=1}^{k_n}\frac{U_n^{h_n}(\hat{v}_j)}{\hat{\lambda}_j}\hat{v}_j\rangle,
\end{aligned}$$

$$\mathbb{E}_{\mathcal{X}^n}(\|\theta - \hat{\theta}_{k_n}^{h_n}\|^2) - \mathbb{E}_{\mathcal{X}^n}(\|\theta - \hat{\theta}_{k_n}\|^2) = \sum_{j=1}^{k_n}\frac{(U_n^{h_n}(\hat{v}_j))^2}{\hat{\lambda}_j^2} + \sigma^2\sum_{j=1}^{k_n}\frac{W_n^{h_n}(\hat{v}_j, \hat{v}_j)}{\hat{\lambda}_j^2} - \frac{\sigma^2}{n}\sum_{j=1}^{k_n}\frac{1}{\hat{\lambda}_j},$$

where

$$U_n^{h_n}(x) = \frac{1}{n}\sum_{i=1}^{n}\langle X_i, x\rangle\left(\sum_{l=1}^{n} w_{l,h_n}(X_i)(m(X_l) - m(X_i))\right), \quad \forall x \in \mathcal{H},$$

and

$$W_n^{h_n}(x, y) = \frac{1}{n^2}\sum_{i_1=1}^{n}\sum_{i_2=1}^{n}\langle X_{i_1}, x\rangle\langle X_{i_2}, y\rangle\left(\sum_{l=1}^{n} w_{l,h_n}(X_{i_1})w_{l,h_n}(X_{i_2})\right), \quad \forall x, y \in \mathcal{H},$$

being $w_{l,h_n}(x) = K(h_n^{-1}d(x, X_l))/\sum_{l'=1}^{n} K(h_n^{-1}d(x, X_{l'}))$ *the weights of the kernel estimator* $\hat{m}_{h_n}$*, and* $\hat{R}_{k_n}$ *is defined in* (2.7) *(see Chapter 2, page 39).*

Corollary 3.3.4. *For the presmoothed FPCA estimator* (3.2)*, if* $w_{l,h_n}(X_i) = 1$ *when* $i = l$*, and* $w_{l,h_n}(X_i) = 0$ *when* $i \neq l$*,* $\forall i = 1, \ldots, n$*, then*

$$\mathbb{E}_{\mathcal{X}^{n+1}}(Y_{n+1} - \langle \hat{\theta}_{k_n}^{h_n}, X_{n+1} \rangle)^2 = \mathbb{E}_{\mathcal{X}^{n+1}}(Y_{n+1} - \langle \hat{\theta}_{k_n}, X_{n+1} \rangle)^2,$$
$$\mathbb{E}_{\mathcal{X}^n}(\|\theta - \hat{\theta}_{k_n}^{h_n}\|^2) = \mathbb{E}_{\mathcal{X}^n}(\|\theta - \hat{\theta}_{k_n}\|^2).$$

Corollary 3.3.5. *For the presmoothed FPCA estimator* (3.2)*, if* $w_{l,h_n}(X_i) = n^{-1}$*,* $\forall i = 1, \ldots, n$*, then*

$$\begin{aligned}
&\mathbb{E}_{\mathcal{X}^{n+1}}(Y_{n+1} - \langle \hat{\theta}_{k_n}^{h_n}, X_{n+1} \rangle)^2 - \mathbb{E}_{\mathcal{X}^{n+1}}(Y_{n+1} - \langle \hat{\theta}_{k_n}, X_{n+1} \rangle)^2 \\
&= \frac{\sigma^2}{n} \left(\langle X_{n+1}, \hat{M}_{k_n} \rangle^2 - \sum_{j=1}^{k_n} \frac{\langle X_{n+1}, \hat{v}_j \rangle^2}{\hat{\lambda}_j} \right) + \Big(\langle \theta, \overline{X} \rangle \langle X_{n+1}, \hat{M}_{k_n} \rangle - \langle X_{n+1}, \theta \rangle \Big)^2 - \langle X_{n+1}, \hat{R}_{k_n} \rangle^2, \\
&\mathbb{E}_{\mathcal{X}^n}(\|\theta - \hat{\theta}_{k_n}^{h_n}\|^2) - \mathbb{E}_{\mathcal{X}^n}(\|\theta - \hat{\theta}_{k_n}\|^2) = \frac{\sigma^2}{n} \left(\|\hat{M}_{k_n}\|^2 - \sum_{j=1}^{k_n} \frac{1}{\hat{\lambda}_j} \right) + \Big(\langle \theta, \overline{X} \rangle \|\hat{M}_{k_n}\| - \|\theta\| \Big)^2 \\
&\quad + 2 \langle \theta, \overline{X} \rangle \Big(\|\hat{M}_{k_n}\| \, \|\theta\| - \langle \hat{M}_{k_n}, \theta \rangle \Big) - \|\hat{R}_{k_n}\|^2,
\end{aligned}$$

where $\hat{M}_{k_n} = \sum_{j=1}^{k_n} \hat{\lambda}_j^{-1} \langle \overline{X}, \hat{v}_j \rangle \hat{v}_j$*, and* $\hat{R}_{k_n}$ *is defined in* (2.7) *(see Chapter 2, page 39).*

The proof of these results can be found in the appendix of the chapter (see Section 3.8.8, page 78, Section 3.8.10, page 82, and Section 3.8.11, page 82).

Remark 3.3.6. It is not easy to compare the conditional errors of $\hat{\theta}_{k_n}^{h_n}$ and $\hat{\theta}_{k_n}$ by means of Theorem 3.3.3, apart from the extreme cases considered in Corollary 3.3.4 and Corollary 3.3.5. Hence, each specific choice of w_{l,h_n} should be analysed individually by means of Theorem 3.3.3 in order to state either the efficiency or inefficiency of the estimator $\hat{\theta}_{k_n}^{h_n}$ with regard to the standard estimator $\hat{\theta}_{k_n}$ from a theoretical point of view.

3.4 Heuristics on alternative presmoothing

In Section 3.2, the presmoothed estimator $\hat{\theta}_{k_n}^{\alpha_n}$ was proposed, which is based on a perturbation of the covariance structure. In view of the good theoretical properties of this estimator, it has been examined what happens when one replaces in $\hat{\theta}_{k_n}$ the standard FPCA components by smoothing FPCA components, such as those proposed by Pezzulli and Silverman (1993) or Silverman (1996). Further details about Pezulli and Silverman' smoothed FPCA and Silverman's smoothed FPCA can be found in Section 1.4.3, *"a) Functional principal component analysis (FPCA)"* (see Chapter 1, page 25).

3.4.1 Using Pezzulli and Silverman' presmoothed FPCA

Definition of estimator

Recall that Pezzulli and Silverman (1993) proposed to find the eigenelements $\{(\hat{\lambda}_j^{\alpha_n,1}, \hat{v}_j^{\alpha_n,1})\}_{j=1}^{\infty}$, which solve the generalized eigenproblem

$$(\Gamma_n - \alpha_n Q)v = \lambda v,$$

where α_n is the positive smoothing parameter, and Q is a symmetric nonnegative operator. Hence, a new estimator can be defined as follows

$$\hat{\theta}_{k_n}^{PS,\,\alpha_n} = \sum_{j=1}^{k_n} \frac{\Delta_n \hat{v}_j^{\alpha_n,1}}{\hat{\lambda}_j^{\alpha_n,1}} \hat{v}_j^{\alpha_n,1}.$$

Conditional errors

In order to determine the effect that smoothing has on the estimation of the eigenelements of Γ, i.e., $\{(\lambda_j, v_j)\}_{j=1}^{\infty}$, some heuristic calculations have been carried out using the standard technique of asymptotic expansions, similarly to Pezzulli and Silverman (1993). For this purpose, it is necessary to define

$$\Pi_j = \sum_{k \neq j} (\hat{\lambda}_j - \hat{\lambda}_k)^{-1} P_k, \quad \rho_j = \langle \hat{v}_j, Q\hat{v}_j \rangle, \quad \text{and} \quad \rho_{j_1,j_2} = \langle \hat{v}_{j_1}, Q\hat{v}_{j_2} \rangle,$$

where P_k is the projection onto the subspace spanned by the empirical eigenfunction $\hat{v}_k$. Substituting the expansions

$$\begin{cases} \hat{v}_j^{\alpha_n,1} = \hat{v}_j + \alpha_n \hat{v}_j^{(1)} + \alpha_n^2 \hat{v}_j^{(2)} + \ldots, \\ \hat{\lambda}_j^{\alpha_n,1} = \hat{\lambda}_j + \alpha_n \hat{\lambda}_j^{(1)} + \alpha_n^2 \hat{\lambda}_j^{(2)} + \ldots, \end{cases}$$

in $(\Gamma_n - \alpha_n Q)v = \lambda v$, matching the coefficients of α_n and applying the normalization conditions, one can derive the values of $\{(\hat{\lambda}_j^{(l)}, \hat{v}_j^{(l)})\}_l$. In this case, for instance, one gets $\hat{v}_j^{(1)} = -\Pi_j Q \hat{v}_j$ and $\hat{\lambda}_j^{(1)} = -\rho_j$. This kind of expansions allows to obtain the following theorem for $\hat{\theta}_{k_n}^{PS,\,\alpha_n}$.

Theorem 3.4.1. *For the estimator $\hat{\theta}_{k_n}^{PS,\,\alpha_n}$, if* **(C.3.2)** *(see page 55) is satisfied, it holds that*

$$\begin{aligned}
&\mathbb{E}_{\mathcal{X}^{n+1}}(Y_{n+1} - \langle \hat{\theta}_{k_n}^{PS,\,\alpha_n}, X_{n+1} \rangle)^2 - \mathbb{E}_{\mathcal{X}^{n+1}}(Y_{n+1} - \langle \hat{\theta}_{k_n}, X_{n+1} \rangle)^2 \\
&= \Bigg(-2\alpha_n \frac{\sigma^2}{n} \sum_{j=1}^{k_n} \frac{\langle X_{n+1}, \hat{v}_j \rangle \langle X_{n+1}, \Pi_j Q \hat{v}_j \rangle}{\hat{\lambda}_j} + \alpha_n \frac{\sigma^2}{n} \sum_{j=1}^{k_n} \rho_j \frac{\langle X_{n+1}, \hat{v}_j \rangle^2}{\hat{\lambda}_j^2} \\
&\quad + \alpha_n \frac{\sigma^2}{n} \sum_{j_1=1}^{k_n} \sum_{j_2=1}^{k_n} \frac{\rho_{j_1,j_2}}{\hat{\lambda}_{j_1} \hat{\lambda}_{j_2}} \langle X_{n+1}, \hat{v}_{j_1} \rangle \langle X_{n+1}, \hat{v}_{j_2} \rangle + 2\alpha_n \langle X_{n+1}, \hat{R}_{k_n} \rangle \langle X_{n+1}, \hat{C}_{k_n}^{PS} \rangle \\
&\quad + \alpha_n^2 \langle X_{n+1}, \hat{C}_{k_n}^{PS} \rangle^2 \Bigg) (1 + o_{a.s.}(1)), \\
&\mathbb{E}_{\mathcal{X}^n}(\|\theta - \hat{\theta}_{k_n}^{PS,\,\alpha_n}\|^2) - \mathbb{E}_{\mathcal{X}^n}(\|\theta - \hat{\theta}_{k_n}\|^2) = \left(2\alpha_n \frac{\sigma^2}{n} \sum_{j=1}^{k_n} \frac{\rho_j}{\hat{\lambda}_j^2} + 2\alpha_n \langle \hat{R}_{k_n}, \hat{C}_{k_n}^{PS} \rangle + \alpha_n^2 \|\hat{C}_{k_n}^{PS}\|^2 \right) \\
&\quad \cdot (1 + o_{a.s.}(1)),
\end{aligned}$$

where $\hat{C}_{k_n}^{PS} = \sum_{j=1}^{k_n} (\langle \theta, \hat{v}_j \rangle \Pi_j Q \hat{v}_j + \langle \theta, \Pi_j Q \hat{v}_j \rangle \hat{v}_j - \hat{\lambda}_j^{-1} \langle \theta, Q\hat{v}_j \rangle \hat{v}_j)$, and $\hat{R}_{k_n}$ is defined in (2.7) *(see Chapter 2, page 39).*

The previous result is proved in the appendix of this chapter (see Section 3.8.12, page 83).

On account of the complexity of the errors expressions in Theorem 3.4.1, one can analyse what happens for the particular case of $Q = I$. Note that $\{(\hat{\lambda}_j^{\alpha_n,1}, \hat{v}_j^{\alpha_n,1})\}_j = \{(\hat{\lambda}_j - \alpha_n, \hat{v}_j)\}_j$ in this case. Hence, the estimator becomes

$$\hat{\theta}_{k_n}^{PS,\,\alpha_n} = \sum_{j=1}^{k_n} \frac{\Delta_n \hat{v}_j}{\hat{\lambda}_j - \alpha_n} \hat{v}_j,$$

and it is easy to deduce the following corollary.

Corollary 3.4.2. *Under assumptions of Theorem 3.4.1, if $Q = I$, it holds that*

$$\mathbb{E}_{\mathcal{X}^{n+1}}(Y_{n+1} - \langle \hat{\theta}_{k_n}^{PS,\,\alpha_n}, X_{n+1}\rangle)^2 - \mathbb{E}_{\mathcal{X}^{n+1}}(Y_{n+1} - \langle \hat{\theta}_{k_n}, X_{n+1}\rangle)^2$$
$$= \left(2\alpha_n \frac{\sigma^2}{n}\sum_{j=1}^{k_n}\frac{\langle X_{n+1}, \hat{v}_j\rangle^2}{\hat{\lambda}_j^2} - 2\alpha_n\langle X_{n+1}, \hat{R}_{k_n}\rangle\langle X_{n+1}, \hat{T}_{k_n}\rangle + \alpha_n^2\langle X_{n+1}, \hat{T}_{k_n}\rangle^2\right)(1 + o_{a.s.}(1)),$$

$$\mathbb{E}_{\mathcal{X}^n}(\|\theta - \hat{\theta}_{k_n}^{PS,\,\alpha_n}\|^2) - \mathbb{E}_{\mathcal{X}^n}(\|\theta - \hat{\theta}_{k_n}\|^2) = \left(2\alpha_n\frac{\sigma^2}{n}\sum_{j=1}^{k_n}\frac{1}{\hat{\lambda}_j^2} + \alpha_n^2\|\hat{T}_{k_n}\|^2\right)(1 + o_{a.s.}(1)),$$

where $\hat{T}_{k_n}$ is defined as in Theorem 3.2.3 (see page 55).

The proof of Corollary 3.4.2 can be found in Section 3.8.13, page 85.

Remark 3.4.3. Corollary 3.4.2 shows that $\hat{\theta}_{k_n}^{PS,\,\alpha_n}$ generates larger conditional estimation errors than the standard FPCA estimator, whereas conclusions related with the prediction error are not clear, depending on the bias term $\hat{R}_{k_n}$ and on $\hat{T}_{k_n}$. On the other hand, note that these conditional errors have the same structure than the computed errors for $\hat{\theta}_{k_n}^{\alpha_n}$ (see Theorem 3.2.3, page 55), which is not surprising in view of the similarities between $\hat{\theta}_{k_n}^{\alpha_n}$ and $\hat{\theta}_{k_n}^{PS,\,\alpha_n}$ with $Q = I$. In fact,

$$\mathbb{E}_{\mathcal{X}^{n+1}}(Y_{n+1} - \langle \hat{\theta}_{k_n}^{PS,\,\alpha_n}, X_{n+1}\rangle)^2 - \mathbb{E}_{\mathcal{X}^{n+1}}(Y_{n+1} - \langle \hat{\theta}_{k_n}^{\alpha_n}, X_{n+1}\rangle)^2$$
$$= \left(4\alpha_n\frac{\sigma^2}{n}\sum_{j=1}^{k_n}\frac{\langle X_{n+1}, \hat{v}_j\rangle^2}{\hat{\lambda}_j^2} - 4\alpha_n\langle X_{n+1}, \hat{R}_{k_n}\rangle\langle X_{n+1}, \hat{T}_{k_n}\rangle\right)(1 + o_{a.s.}(1)),$$

$$\mathbb{E}_{\mathcal{X}^n}(\|\theta - \hat{\theta}_{k_n}^{PS,\,\alpha_n}\|^2) - \mathbb{E}_{\mathcal{X}^n}(\|\theta - \hat{\theta}_{k_n}^{\alpha_n}\|^2) = \left(4\alpha_n\frac{\sigma^2}{n}\sum_{j=1}^{k_n}\frac{1}{\hat{\lambda}_j^2}\right)(1 + o_{a.s.}(1)).$$

3.4.2 Using Silverman's presmoothed FPCA

Definition of estimator

If the smoothed FPCA proposed by Silverman (1996) is considered, one needs to find the pairs $\{(\hat{\lambda}_j^{\alpha_n,2}, \hat{v}_j^{\alpha_n,2})\}_{j=1}^{\infty}$ that solve the eigenproblem

$$\Gamma_n v = \lambda(I + \alpha_n Q)v,$$

where the orthonormality constraints are based on the next inner product

$$\langle x, y\rangle_{\alpha_n} = \langle x, y\rangle + \alpha_n\langle x, Qy\rangle,$$

being α_n the smoothing parameter and Q a symmetric nonnegative operator. A new estimator based on this eigenelements can be defined as follows

$$\hat{\theta}_{k_n}^{S,\,\alpha_n} = \sum_{j=1}^{k_n}\frac{\Delta_n\hat{v}_j^{\alpha_n,2}}{\hat{\lambda}_j^{\alpha_n,2}}\hat{v}_j^{\alpha_n,2}.$$

Conditional errors

The asymptotic expansions for the eigenelements $\{(\hat{\lambda}_j^{\alpha_n,2}, \hat{v}_j^{\alpha_n,2})\}_{j=1}^{\infty}$ can be calculated following the same procedure as in Pezzulli and Silverman' case. Specifically, one gets $\hat{v}_j^{(1)} = -((\rho_j/2) + \hat{\lambda}_j\Pi_j Q)\hat{v}_j$ and $\hat{\lambda}_j^{(1)} = -\rho_j\hat{\lambda}_j$, and the following result can be stated.

Theorem 3.4.4. *For the estimator $\hat{\theta}_{k_n}^{S,\,\alpha_n}$, if* **(C.3.2)** *(see page 55) is satisfied, it holds that*

$$\begin{aligned}
&\mathbb{E}_{\mathcal{X}^{n+1}}(Y_{n+1} - \langle \hat{\theta}_{k_n}^{S,\,\alpha_n}, X_{n+1}\rangle)^2 - \mathbb{E}_{\mathcal{X}^{n+1}}(Y_{n+1} - \langle \hat{\theta}_{k_n}, X_{n+1}\rangle)^2 \\
&= \left(-2\alpha_n \frac{\sigma^2}{n} \sum_{j=1}^{k_n} \langle X_{n+1}, \hat{v}_j\rangle \langle X_{n+1}, \Pi_j Q \hat{v}_j\rangle + 2\alpha_n \langle X_{n+1}, \hat{R}_{k_n}\rangle \langle X_{n+1}, \hat{C}_{k_n}^S\rangle + \alpha_n^2 \langle X_{n+1}, \hat{C}_{k_n}^S\rangle^2 \right) \\
&\quad \cdot (1 + o_{a.s.}(1)), \\
&\mathbb{E}_{\mathcal{X}^n}(\|\theta - \hat{\theta}_{k_n}^{S,\,\alpha_n}\|^2) - \mathbb{E}_{\mathcal{X}^n}(\|\theta - \hat{\theta}_{k_n}\|^2) = \left(2\alpha_n \langle \hat{R}_{k_n}, \hat{C}_{k_n}^S\rangle + \alpha_n^2 \|\hat{C}_{k_n}^S\|^2 \right)(1 + o_{a.s.}(1)),
\end{aligned}$$

where $\hat{C}_{k_n}^S = \sum_{j=1}^{k_n} (\rho_j \langle \theta, \hat{v}_j\rangle \hat{v}_j + \hat{\lambda}_j (\langle \theta, \hat{v}_j\rangle \Pi_j Q \hat{v}_j + \langle \theta, \Pi_j Q \hat{v}_j\rangle \hat{v}_j) - \langle \theta, Q\hat{v}_j\rangle \hat{v}_j)$, and $\hat{R}_{k_n}$ is defined in (2.7) *(see Chapter 2, page 39).*

Section 3.8.14 presents the proof of the previous result (see page 85).

Remark 3.4.5. The expressions for the conditional errors in Theorem 3.4.4 depend on many terms which make their interpretation cumbersome, apart from the particular case $Q = I$. In this specific situation, $\{(\hat{\lambda}_j^{\alpha_n,2}, \hat{v}_j^{\alpha_n,2})\}_j = \{((1+\alpha_n)^{-1}\hat{\lambda}_j, (1+\alpha_n)^{-1/2}\hat{v}_j)\}_j$, and consequently $\hat{\theta}_{k_n}^{S,\,\alpha_n} = \hat{\theta}_{k_n}$.

These heuristics on alternative presmoothing approaches show how the general Lemma 2.5.2 (see Chapter 2, page 49) is applied to many situations in order to compute the conditional errors of estimation and prediction. However, sometimes the expressions are too complex to compare them with the errors of other estimators (recall Theorem 3.4.1 or Theorem 3.4.4), except for the simplest situations (e.g., Corollary 3.4.2 or previous remark).

3.5 Simulation study

This section is devoted to the presentation of two simulation studies for the regression model presented in Remark 2.3.2 (see Chapter 2, page 33), that is, for $\mathcal{H} = L^2([0,1])$ and

$$Y = \int_0^1 \theta(t) X(t) dt + \epsilon.$$

The first case (Case A) is characterized by the existence of null eigenvalues: $\lambda_j > 0$ if $j \in \{1, \ldots, 4\}$, and $\lambda_j = 0$ otherwise (it corresponds to the motivation example introduced at the beginning of this chapter). In the second case (Case B), the eigenvalues of the covariance operator decrease quickly, but all of them are different from zero.

In both studies, 200 samples were simulated, each containing $2\,n$ observations drawn from the previous functional linear model with $\epsilon \sim \mathcal{N}(0, \sigma^2)$ and signal–to–noise ratio $r = \sigma / \sqrt{\mathbb{E}\langle X, \theta\rangle^2}$. Each sample is split up into two subsamples of size n: $\{(X_i, Y_i)\}_{i=1}^n$ is used to estimate the parameter θ (*learning sample*), and $\{(X_i, Y_i)\}_{i=n+1}^{2n}$ is used to test the obtained results (*test sample*). Different values for the signal–to–noise ratio r ($r = 0.02, 0.2, 0.5, 1, 2$), for the sample size n ($n = 25, 50, 100, 200$), and for different percentages of outliers ($out = 0\%, 10\%, 20\%$), constructed by changing the original model error by $\epsilon^* \sim \mathcal{N}(7, \sigma^2)$, were fixed.

For each case, the standard estimator $\hat{\theta}_{k_n}$ (see (2.4) in Chapter 2, page 36) and the presmoothed estimators $\hat{\theta}_{k_n}^{\alpha_n}$ (see (3.1), page 54) and $\hat{\theta}_{k_n}^{\text{k–NN}}$ (see (3.3), page 57) proposed in this chapter were calculated. The penalized B–splines estimator $\hat{\theta}_{PS}$ (see (2.2) in Chapter 2, page 35) were also computed.

To compare the behaviour of these estimators, the following risk functions were used

$$R(Y) = \frac{1}{n} \sum_{i=n+1}^{2n} (Y_i - \hat{Y}_i)^2, \quad \text{and} \quad R(\theta) = \int (\theta(t) - \hat{\theta}(t))^2 dt, \tag{3.4}$$

the mean square prediction error and the mean square estimation error of θ, respectively, being $\hat{\theta} \in \{\hat{\theta}_{PS}, \hat{\theta}_{k_n}, \hat{\theta}^{\alpha_n}_{k_n}, \hat{\theta}^{\text{k–NN}}_{k_n}\}$. To calculate the first, for each simulated sample, each estimator $\hat{\theta}$ was built from the learning sample $\{(X_i, Y_i)\}_{i=1}^{n}$. The test sample $\{(X_i, Y_i)\}_{i=n+1}^{2n}$ produces the prediction $\hat{Y}_i = \langle \hat{\theta}, X_i \rangle$ for each $i \in \{n+1, \ldots, 2n\}$, and the corresponding value of $R(Y)$.

Before presenting the simulation results, some issues on the discretization effect and the parameter selection must be commented.

How to work with discrete data? Since curves are not recorded continuously, X and θ were discretized on $p = 100$ equispaced design points $(t_1, \ldots, t_{100})$ so that integrals involved had to be approximated. Quadrature weights of p^{-1} were used, i.e.,

$$\int_0^1 x(t)dt \approx p^{-1} \sum_{l=1}^{p} x(t_l),$$

though more complex integral approximations are available, such as the Gauss method. As a result, the eigenelements of Γ_n were approximated by those of the $p \times p$–matrix $\mathbf{\Gamma}_n = (\gamma_{l_1,l_2})_{l_1=1,\ldots,p;l_2=1,\ldots,p}$ with $\gamma_{l_1,l_2} = (np)^{-1} \sum_{i=1}^{n} X_i(t_{l_1}) X_i(t_{l_2})$, and Δ_n was approximated by the p–vector $\mathbf{\Delta}_n = (\delta_l)_{l=1,\ldots,p}$ where $\delta_l = (np)^{-1} \sum_{i=1}^{n} X_i(t_l) Y_i$. The risk functions (3.4) also depend on integral calculations and they were approximated by $R(Y) \approx n^{-1} \sum_{i=1}^{n} (Y_i - p^{-1} \sum_{l=1}^{p} \theta(t_l) X_i(t_l))^2$ and $R(\theta) \approx p^{-1} \sum_{l=1}^{p} (\theta(t_l) - \hat{\theta}(t_l))^2$.

Parameter selection. The estimators of θ which were computed require that some parameters are fixed/estimated previously by the user.

$\hat{\theta}_{PS}$: The parameters $k = 20$, $q = 4$ and $r = 2$ were fixed, that is, the normalized B–splines basis with degree 4, with 19 equispaced interior knots, and a roughness penalty based on the second derivative was considered. Moreover, a *generalized cross–validation method* (GCV) was used to select ρ. That way one looked for the value of the smoothing parameter which minimized the GCV criterion

$$\text{GCV}(\rho) = \frac{n}{(n - tr(\mathbf{S}_\rho))^2} \sum_{i=1}^{n} (Y_i - \langle \hat{\theta}_{PS}, X_i \rangle)^2,$$

being $\mathbf{S}_\rho$ the hat matrix which relates the fitted values $\hat{Y}_i = \langle \hat{\theta}_{PS}, X_i \rangle$ to the observed values Y_i.

$\hat{\theta}_{k_n}$: The number of principal components involved was also computed by means of a GCV method. Hence, k_n was selected minimizing the following GCV criterion

$$\text{GCV}(k_n) = \frac{n}{(n - tr(\mathbf{S}_{k_n}))^2} \sum_{i=1}^{n} (Y_i - \langle \hat{\theta}_{k_n}, X_i \rangle)^2,$$

where $\mathbf{S}_{k_n}$ is the hat matrix associated to $\hat{\theta}_{k_n}$. In this case, it can be shown that $tr(\mathbf{S}_{k_n}) = k_n$.

$\hat{\theta}^{\alpha_n}_{k_n}$: Both k_n and α_n were chosen in order to minimize the criterion

$$\text{GCV}(k_n, \alpha_n) = \frac{n}{(n - tr(\mathbf{S}_{k_n,\alpha_n}))^2} \sum_{i=1}^{n} (Y_i - \langle \hat{\theta}^{\alpha_n}_{k_n}, X_i \rangle)^2,$$

being $\mathbf{S}_{k_n,\alpha_n}$ the hat matrix which corresponds to $\hat{\theta}^{\alpha_n}_{k_n}$. Some calculations allow to express its trace as $tr(\mathbf{S}_{k_n,\alpha_n}) = \sum_{j=1}^{k_n} (\hat{\lambda}_j + \alpha_n)^{-1} \hat{\lambda}_j$.

$\hat{\theta}^{\text{k–NN}}_{k_n}$: For this estimator, k_n and the number of neighbours $neig$ minimized the next GCV criterion

$$\text{GCV}(k_n, neig) = \frac{n}{(n - tr(\mathbf{S}_{k_n,neig}))^2} \sum_{i=1}^{n} (Y_i - \langle \hat{\theta}^{\text{k–NN}}_{k_n}, X_i \rangle)^2,$$

with $\mathbf{S}_{k_n,neig}$ the hat matrix for $\hat{\theta}_{k_n}^{\text{k–NN}}$. It can be also seen that the trace of $\mathbf{S}_{k_n,neig}$ is given by

$$tr(\mathbf{S}_{k_n,neig}) = \frac{1}{n}\sum_{j=1}^{k_n}\sum_{i=1}^{n}\frac{\langle X_i,\hat{v}_j\rangle}{\hat{\lambda}_j}\sum_{l=1}^{n} w_{l,h_{neig}}(X_i)\langle X_l,\hat{v}_j\rangle,$$

being $w_{l,h_{neig}}(\cdot)$ the weights of the k–NN estimator $\hat{m}_{\text{k–NN}}$.

Furthermore, for $\hat{\theta}_{k_n}$, $\hat{\theta}_{k_n}^{\alpha_n}$ and $\hat{\theta}_{k_n}^{\text{k–NN}}$, the "optimal" parameters which minimize the risk functions $R(Y)$ and $R(\theta)$ were also calculated. These two alternative methods are going to be denoted as $opt_{R(Y)}$ and $opt_{R(\theta)}$, respectively.

3.5.1 Case A. Existence of null eigenvalues

Consider the curves

$$X(t) = a_1\sqrt{2}\sin(\pi t) + a_2\sqrt{2}\cos(\pi t) + a_3\sqrt{2}\sin(2\pi t) + a_4\sqrt{2}\cos(2\pi t), \quad \forall t\in[0,1],$$

with $a_l \sim \mathcal{U}(-1/3^{l-1}, 1/3^{l-1})$ for all $l\in\{1,\ldots,4\}$, and

$$\theta(t) = 2\sqrt{2}\cos(2\pi t), \quad \forall t\in[0,1].$$

It can be shown that the eigenvalues of Γ are

$$\lambda_j = \text{Var}(a_j) = 1/3^{2j-1} \text{ for } j\in\{1,\ldots,4\}, \quad \text{and} \quad \lambda_j = 0 \text{ for } j>4.$$

Moreover, the model parameter is twice the fourth eigenfunction of the second moment operator. This is then a favourable case for the FPCA–type estimators if one selects an adequate k_n.

The median of the risk functions for the different values of n, r and *out* are compiled in Table 3.1, Table 3.2, Table 3.3, and Table 3.4 (see pages 66–69). From them, one can see that both $\hat{\theta}_{k_n}$ and $\hat{\theta}_{k_n}^{\alpha_n}$ give better estimations for θ than the penalized B–splines estimator, as expected given the connection between the model parameter and the fourth functional principal component. Furthermore, $\hat{\theta}_{k_n}^{\text{k–NN}}$ improves the results of $\hat{\theta}_{PS}$ in general, although it seems to have a worse performance when the sample size is small, the percentage of outliers is large and the GCV criterion is used to select the involved parameters (see $R(\theta)$ for GCV method in Table 3.1 and Table 3.2). As for $R(Y)$, the four estimators have values of the same order.

When comparing the GCV method with the optimal selections, it can be seen that errors when the GCV choice is used are similar to the $opt_{R(Y)}$ ones in terms of prediction error and generally higher than $opt_{R(\theta)}$ errors when the estimation error is concerned. The results are reasonable since GCV techniques minimize criteria strongly linked with prediction error.

Focusing on the FPCA–type estimators, the results show that $\hat{\theta}_{k_n}^{\alpha_n}$ obtains smaller errors than $\hat{\theta}_{k_n}$ in most cases. Nevertheless, note that the improvement is almost negligible for $R(Y)$, whereas it can be very significant for $R(\theta)$ when the signal–to–noise ratio is high, when the sample size is small, or when there is an important presence of outliers. These results coincide with the theoretical conclusions exposed in Remark 3.2.5 (see page 56) and Remark 3.2.6 (see page 56). On the other hand, $\hat{\theta}_{k_n}^{\text{k–NN}}$ gives larger errors than the standard FPCA estimator in general, except if $opt_{R(Y)}$ is the choice for selecting the involved parameters.

3.5.2 Case B. Non existence of null eigenvalues

The functional linear model with X a Brownian motion (see Section 1.1.2 in Chapter 1, page 2), and

$$\theta(t) = \log(15t^2+10) + \cos(4\pi t), \quad \forall t\in[0,1],$$

was simulated. It is well–known that the eigenelements of the second moment operator of a Brownian motion are

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

			GCV				$opt_{R(Y)}$			$opt_{R(\theta)}$		
error	r	*out*	$\hat{\theta}_{PS}$	$\hat{\theta}_{k_n}$	$\hat{\theta}^{\alpha_n}_{k_n}$	$\hat{\theta}^{\text{k–NN}}_{k_n}$	$\hat{\theta}_{k_n}$	$\hat{\theta}^{\alpha_n}_{k_n}$	$\hat{\theta}^{\text{k–NN}}_{k_n}$	$\hat{\theta}_{k_n}$	$\hat{\theta}^{\alpha_n}_{k_n}$	$\hat{\theta}^{\text{k–NN}}_{k_n}$
$R(Y)$	0.02	0%	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
		10%	4.25	4.32	4.13	4.08	4.08	3.88	3.86	4.15	3.97	3.96
		20%	10.48	10.55	10.02	10.24	10.13	9.70	9.69	10.39	9.85	9.94
	0.2	0%	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
		10%	4.11	4.14	4.07	4.07	4.01	3.84	3.86	4.10	3.98	3.98
		20%	10.58	10.63	10.19	10.24	10.21	9.70	9.63	10.40	9.89	9.88
	0.5	0%	0.07	0.07	0.06	0.06	0.06	0.06	0.06	0.07	0.06	0.06
		10%	4.32	4.36	4.11	4.14	4.10	3.86	3.90	4.22	4.01	4.04
		20%	10.69	10.87	10.21	10.45	10.33	9.75	9.75	10.43	9.98	10.08
	1	0%	0.25	0.25	0.25	0.24	0.24	0.22	0.22	0.25	0.24	0.24
		10%	4.47	4.50	4.32	4.37	4.35	4.12	4.10	4.39	4.26	4.25
		20%	10.91	11.01	10.44	10.55	10.51	9.99	9.90	10.71	10.20	10.22
	2	0%	1.02	1.05	1.00	0.98	0.97	0.89	0.89	0.98	0.97	0.93
		10%	5.12	5.16	4.94	4.96	4.96	4.71	4.68	5.05	4.87	4.87
		20%	11.78	11.77	11.10	11.35	11.33	10.54	10.48	11.52	10.79	11.02
$R(\theta)$	0.02	0%	0.17	0.01	0.01	5.25	1.16	0.01	4.45	0.01	0.00	0.41
		10%	12.62	9.02	3.89	16.55	11.16	3.97	15.93	6.12	3.40	3.59
		20%	19.39	11.95	3.98	116.45	18.32	4.17	20.11	7.92	3.49	4.37
	0.2	0%	1.37	3.32	2.32	3.35	3.33	1.51	3.34	1.03	0.41	0.51
		10%	12.40	7.82	3.91	18.34	11.99	4.15	24.74	5.96	3.45	3.60
		20%	25.28	16.52	4.06	65.36	29.25	4.25	25.93	11.00	3.52	4.25
	0.5	0%	5.00	3.41	3.37	3.39	3.40	3.33	3.37	3.33	1.55	0.73
		10%	13.86	8.71	3.87	14.90	10.29	4.00	9.02	6.31	3.41	3.59
		20%	24.47	16.72	4.02	118.13	22.87	4.24	22.34	9.68	3.55	4.58
	1	0%	5.44	3.58	3.45	3.61	3.73	3.42	3.78	3.46	3.09	1.42
		10%	13.72	8.57	3.96	15.89	10.89	4.01	7.99	6.62	3.45	3.67
		20%	24.26	13.95	4.04	84.46	23.84	4.24	19.57	10.51	3.48	4.42
	2	0%	6.58	4.40	3.61	4.94	4.60	3.58	4.02	3.84	3.36	3.14
		10%	11.52	7.65	3.90	15.17	11.19	4.00	14.23	5.58	3.41	3.65
		20%	23.37	14.74	4.10	70.42	25.79	4.15	15.03	11.03	3.52	4.92

Table 3.1: Case A. Median of $R(Y)$ and $R(\theta)$ for sample size $n = 25$.

with all eigenvalues of Γ strictly positive.

From Table 3.5, Table 3.6, Table 3.7, and Table 3.8 (see pages 70–73), one sees that $\hat{\theta}^{\alpha_n}_{k_n}$ gives better estimates for θ than $\hat{\theta}_{k_n}$, according with the theoretical results. Again, simulations confirm the effect of the sample size and noise on the expected reduction for estimation of θ (see Remark 3.2.6, page 56): more improvement when n is small and when the "noise level" is large. As happens in Case A, $\hat{\theta}_{k_n}$ and $\hat{\theta}^{\alpha_n}_{k_n}$ give similar results in terms of $R(Y)$. Comparing the presmoothed estimator $\hat{\theta}^{\alpha_n}_{k_n}$ with $\hat{\theta}_{PS}$, the behaviour is similar as far as prediction error is concerned, $\hat{\theta}^{\alpha_n}_{k_n}$ gives smaller estimation errors than the penalized B–splines estimator when n is small or the noise is large.

With regard to $\hat{\theta}^{\text{k–NN}}_{k_n}$ the conclusions are similar to those derived from the Case A: $\hat{\theta}^{\text{k–NN}}_{k_n}$ does not reduce the prediction errors obtained by $\hat{\theta}_{k_n}$, and its estimation error increases abruptly when the GCV method is considered and there are outliers in the sample.

These simulations suggest that the presmoothing estimator $\hat{\theta}^{\alpha_n}_{k_n}$ improves the standard FPCA linear estimator, and especially when the sample size is small, whereas $\hat{\theta}^{\text{k–NN}}_{k_n}$ do not significantly reduce the conditional errors of $\hat{\theta}_{k_n}$. The choice of the parameters of the proposed presmoothed estimators is of course a key point. A general practical guideline is to choose these parameters by cross–validation techniques. Even if the results above might point to a new way for selecting these parameters that could give even more improvement (see results for the "optimal" choice with respect $R(\theta)$), CV gives good data–driven results, except for $\hat{\theta}^{\text{k–NN}}_{k_n}$.

			GCV				$opt_{R(Y)}$			$opt_{R(\theta)}$		
error	r	*out*	$\hat{\theta}_{PS}$	$\hat{\theta}_{k_n}$	$\hat{\theta}^{\alpha_n}_{k_n}$	$\hat{\theta}^{\text{k–NN}}_{k_n}$	$\hat{\theta}_{k_n}$	$\hat{\theta}^{\alpha_n}_{k_n}$	$\hat{\theta}^{\text{k–NN}}_{k_n}$	$\hat{\theta}_{k_n}$	$\hat{\theta}^{\alpha_n}_{k_n}$	$\hat{\theta}^{\text{k–NN}}_{k_n}$
$R(Y)$	0.02	0%	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
		10%	5.03	5.05	5.00	4.99	4.97	4.85	4.87	4.98	4.92	4.94
		20%	10.13	10.15	9.92	10.06	9.92	9.71	9.75	10.00	9.81	9.86
	0.2	0%	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
		10%	5.05	5.09	5.03	5.01	5.00	4.86	4.90	5.03	4.94	4.95
		20%	10.12	10.19	10.01	10.07	9.97	9.75	9.78	10.03	9.87	9.93
	0.5	0%	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06
		10%	5.13	5.14	5.07	5.11	5.09	4.95	4.98	5.12	5.03	5.06
		20%	10.27	10.27	10.08	10.16	10.12	9.86	9.90	10.18	9.93	10.03
	1	0%	0.25	0.25	0.25	0.24	0.24	0.23	0.24	0.25	0.24	0.24
		10%	5.40	5.41	5.29	5.32	5.28	5.10	5.13	5.33	5.16	5.21
		20%	10.47	10.48	10.32	10.33	10.35	10.01	10.05	10.40	10.13	10.27
	2	0%	0.97	0.98	0.97	0.96	0.95	0.93	0.92	0.97	0.95	0.94
		10%	6.18	6.22	6.17	6.14	6.10	5.88	5.94	6.13	6.07	6.03
		20%	11.14	11.24	10.96	11.01	11.01	10.70	10.70	11.10	10.79	10.87
$R(\theta)$	0.02	0%	0.16	0.00	0.00	8.11	1.69	0.00	7.61	0.00	0.00	0.38
		10%	7.85	5.31	3.68	12.75	6.12	3.79	7.16	4.56	3.35	3.62
		20%	13.95	10.32	3.95	40.66	11.93	4.13	9.98	6.83	3.44	4.23
	0.2	0%	0.57	0.58	0.58	3.92	3.30	0.57	3.75	0.50	0.20	0.31
		10%	8.99	6.12	3.85	13.76	8.27	3.94	7.06	5.40	3.40	3.81
		20%	13.63	8.90	3.94	54.58	12.64	4.12	13.17	6.24	3.44	4.20
	0.5	0%	4.17	3.38	3.34	3.37	3.35	3.30	3.34	2.87	1.03	0.58
		10%	8.88	5.40	3.82	13.25	9.64	3.82	7.54	4.69	3.36	3.71
		20%	12.63	8.76	3.95	34.59	12.49	4.11	13.49	6.55	3.41	4.28
	1	0%	5.06	3.47	3.39	3.43	3.53	3.38	3.60	3.39	2.65	1.00
		10%	10.02	6.54	3.89	9.87	9.07	3.91	9.37	5.11	3.40	3.86
		20%	12.65	8.39	3.93	18.03	13.90	4.01	18.03	6.55	3.47	4.40
	2	0%	5.45	3.77	3.50	3.74	4.27	3.54	3.89	3.63	3.32	2.67
		10%	9.38	5.97	3.86	14.52	9.35	3.89	9.48	4.96	3.39	3.71
		20%	14.49	8.16	3.94	37.83	12.62	4.06	12.09	6.64	3.45	4.20

Table 3.2: Case A. Median of $R(Y)$ and $R(\theta)$ for sample size $n = 50$.

3.6 Real data application

In order to demonstrate the improved performance of the proposed methodology with respect to the FPCA estimator in applications, three functional datasets with different size n were chosen: Canadian weather data ($n = 35$), spectrometric data ($n = 215$), and atmospheric pollution data ($n = 1,000$). The aim is to see how the sample size affects to the improvement of $\hat{\theta}^{\alpha}_{k_n}$ with respect to the standard FPCA estimator, and how the behaviour of $\hat{\theta}^{\text{k–NN}}_{k_n}$ is data applications.

Since the variables involved are not centred, the regression model with non–zero intercept given by Remark 2.3.1 (see Chapter 2, page 33) was used, and the next steps were followed.

Step 1. Calculate the sample mean of curves ($\overline{X}$) and scalar responses ($\overline{Y}$).

Step 2. Split the sample into a learning sample $\{(X_i, Y_i)\}_{i\in I_{LS}}$ and a test sample $\{(X_i, Y_i)\}_{i\in I_{TS}}$.

Step 3. Use the centred learning sample $\{(X_i - \overline{X}, Y_i - \overline{Y})\}_{i\in I_{LS}}$ to build $\hat{\theta}$ ($\hat{\theta}_{PS}$, $\hat{\theta}_{k_n}$, $\hat{\theta}^{\alpha_n}_{k_n}$ or $\hat{\theta}^{\text{k–NN}}_{k_n}$), and estimate the intercept term by $\hat{\theta}_0 = \overline{Y} - \langle\hat{\theta}, \overline{X}\rangle$.

Step 4. Compute the predicted responses for the test sample, $\hat{Y}_i = \hat{\theta}_0 + \langle\hat{\theta}, X_i\rangle$, $\forall i \in I_{TS}$.

Step 5. Obtain $R(Y) = \frac{1}{\#I_{TS}} \sum_{i\in I_{TS}} (Y_i - \hat{Y}_i)^2$, where $\#I_{TS}$ denotes the test sample size.

To avoid the effect of sample selection, this procedure was iterated 100 times, and the mean, the median, and the standard deviation of $R(Y)$ over these replications were calculated.

The general guidelines of the simulations study to build the different estimates of θ were followed. In particular, the involved parameters were selected by GCV (because a prediction objective exists in the real data application).

			GCV				$opt_{R(Y)}$			$opt_{R(\theta)}$		
error	r	*out*	$\hat{\theta}_{PS}$	$\hat{\theta}_{k_n}$	$\hat{\theta}^{\alpha_n}_{k_n}$	$\hat{\theta}^{\text{k–NN}}_{k_n}$	$\hat{\theta}_{k_n}$	$\hat{\theta}^{\alpha_n}_{k_n}$	$\hat{\theta}^{\text{k–NN}}_{k_n}$	$\hat{\theta}_{k_n}$	$\hat{\theta}^{\alpha_n}_{k_n}$	$\hat{\theta}^{\text{k–NN}}_{k_n}$
$R(Y)$	0.02	0%	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
		10%	4.99	4.99	4.96	4.98	4.94	4.88	4.91	4.97	4.93	4.95
		20%	9.97	9.98	9.92	9.96	9.89	9.78	9.80	9.95	9.86	9.89
	0.2	0%	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
		10%	5.00	5.01	4.97	4.99	4.96	4.90	4.91	4.99	4.93	4.96
		20%	9.97	10.00	9.94	9.98	9.91	9.78	9.82	9.95	9.85	9.90
	0.5	0%	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06
		10%	5.08	5.10	5.04	5.07	5.02	4.93	4.97	5.03	4.98	5.03
		20%	10.05	10.07	10.00	10.02	9.96	9.82	9.86	10.00	9.90	9.95
	1	0%	0.25	0.25	0.24	0.25	0.24	0.24	0.24	0.24	0.24	0.25
		10%	5.24	5.24	5.21	5.24	5.18	5.12	5.14	5.23	5.18	5.19
		20%	10.21	10.27	10.17	10.23	10.16	10.00	10.06	10.19	10.09	10.15
	2	0%	0.97	0.97	0.97	0.97	0.96	0.94	0.95	0.96	0.96	0.97
		10%	5.92	5.93	5.91	5.92	5.87	5.82	5.79	5.89	5.86	5.84
		20%	10.96	10.98	10.85	10.92	10.85	10.68	10.73	10.92	10.80	10.82
$R(\theta)$	0.02	0%	0.16	0.00	0.00	7.64	2.78	0.00	7.69	0.00	0.00	0.37
		10%	7.03	4.74	3.73	5.66	6.84	3.75	6.32	4.27	3.37	3.69
		20%	8.23	5.56	3.89	8.31	6.96	3.97	8.13	4.53	3.41	3.86
	0.2	0%	0.40	0.23	0.27	5.19	1.94	0.19	4.36	0.23	0.10	0.50
		10%	7.22	4.54	3.85	5.14	5.87	3.72	6.26	4.07	3.34	3.51
		20%	8.86	5.84	3.87	8.41	7.94	3.79	7.04	4.71	3.38	3.81
	0.5	0%	1.76	3.33	3.32	3.35	3.33	3.17	3.33	1.48	0.60	0.61
		10%	7.63	4.93	3.65	6.40	5.57	3.66	6.05	4.19	3.36	3.62
		20%	9.04	6.38	3.85	14.14	8.71	4.01	11.36	4.89	3.38	4.01
	1	0%	4.94	3.41	3.37	3.39	3.40	3.34	3.39	3.34	1.52	0.82
		10%	6.94	4.63	3.74	5.09	6.24	3.70	6.00	4.07	3.36	3.52
		20%	8.29	5.67	3.82	9.70	7.11	3.76	10.29	4.60	3.37	3.77
	2	0%	5.23	3.58	3.43	3.69	3.70	3.41	3.60	3.45	2.97	2.98
		10%	6.99	4.81	3.72	6.35	5.93	3.68	6.13	4.14	3.36	3.56
		20%	9.40	6.47	3.87	8.66	8.48	3.88	8.83	5.00	3.38	3.89

Table 3.3: Case A. Median of $R(Y)$ and $R(\theta)$ for sample size $n = 100$.

3.6.1 Canadian weather data

Firstly, the Canadian weather data is considered (for further details, see Section 1.1.2 in Chapter 1, page 2). Here Y is the logarithm of total annual precipitation at each weather station, and X is the daily temperature curve (see Ramsay and Silverman, 2005, Chapter 15, for a Fourier basis approach to this case). The original sample was split into two subsamples: a learning sample (25 stations) and a testing one (10 stations). For $\hat{\theta}^{\text{k–NN}}_{k_n}$, the quadratic kernel, and the semi–metric based on the first derivatives were selected. Table 3.9 (see page 73) shows the mean, median, and standard deviation of $R(Y)$. Note that both $\hat{\theta}^{\alpha_n}_{k_n}$ and $\hat{\theta}^{\text{k–NN}}_{k_n}$ yield to a reduction of the error of the standard FPCA estimator and the penalized B–splines estimator.

3.6.2 Spectrometric data

For the second illustration, spectrometric data was selected (see Section 1.1.2 in Chapter 1, page 2). The dataset contains 215 spectrometric curves obtained from pieces of finely chopped meat, and a scalar value corresponding to the fat content of each of them. The second derivatives of the spectrometric curve for each meat piece were taken as the functional variable X, and its fat content as the response Y (Ferraty and Vieu, 2006b also chose the second derivatives in order to get better predictive results in the functional nonparametric regression context). The sample was split into a learning sample of 160 pieces and a test sample of 55 pieces. For $\hat{\theta}^{\text{k–NN}}_{k_n}$, the quadratic kernel, and the semi–metric based on FPCA with $q = 2$ were selected. Table 3.10 (see page 73) contains the results for this example. In this case, the best results are produced by $\hat{\theta}^{\alpha_n}_{k_n}$.

			GCV				$opt_{R(Y)}$			$opt_{R(\theta)}$		
error	r	*out*	$\hat{\theta}_{PS}$	$\hat{\theta}_{k_n}$	$\hat{\theta}^{\alpha_n}_{k_n}$	$\hat{\theta}^{\text{k–NN}}_{k_n}$	$\hat{\theta}_{k_n}$	$\hat{\theta}^{\alpha_n}_{k_n}$	$\hat{\theta}^{\text{k–NN}}_{k_n}$	$\hat{\theta}_{k_n}$	$\hat{\theta}^{\alpha_n}_{k_n}$	$\hat{\theta}^{\text{k–NN}}_{k_n}$
$R(Y)$	0.02	0%	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
		10%	4.95	4.95	4.94	4.94	4.92	4.90	4.91	4.93	4.92	4.93
		20%	9.88	9.90	9.86	9.90	9.85	9.79	9.80	9.86	9.82	9.84
	0.2	0%	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
		10%	4.96	4.96	4.95	4.95	4.95	4.92	4.92	4.96	4.94	4.95
		20%	9.88	9.89	9.86	9.88	9.84	9.79	9.81	9.86	9.84	9.86
	0.5	0%	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06
		10%	5.00	5.01	5.00	5.01	4.99	4.95	4.96	5.00	4.98	4.99
		20%	9.96	9.98	9.95	9.98	9.93	9.86	9.89	9.94	9.90	9.92
	1	0%	0.24	0.24	0.24	0.24	0.23	0.23	0.23	0.24	0.24	0.24
		10%	5.24	5.25	5.22	5.24	5.20	5.14	5.18	5.22	5.18	5.22
		20%	10.11	10.14	10.10	10.11	10.07	10.00	10.03	10.10	10.05	10.08
	2	0%	0.96	0.96	0.96	0.96	0.95	0.94	0.94	0.96	0.96	0.95
		10%	5.92	5.94	5.92	5.92	5.91	5.85	5.83	5.91	5.88	5.86
		20%	10.90	10.92	10.84	10.90	10.83	10.70	10.77	10.86	10.77	10.82
$R(\theta)$	0.02	0%	0.16	0.00	0.00	8.49	2.44	0.00	8.41	0.00	0.00	0.35
		10%	6.14	3.92	3.55	4.12	4.44	3.55	4.87	3.66	3.33	3.43
		20%	7.38	4.55	3.70	5.57	5.04	3.65	5.01	4.12	3.35	3.60
	0.2	0%	0.27	0.12	0.13	6.87	1.27	0.11	4.59	0.12	0.05	0.36
		10%	5.73	3.78	3.52	3.90	4.38	3.61	4.67	3.67	3.33	3.41
		20%	6.91	4.47	3.64	5.10	5.92	3.70	6.44	4.00	3.34	3.64
	0.5	0%	1.02	1.93	1.48	3.34	3.33	0.95	3.35	0.82	0.31	0.35
		10%	5.63	4.00	3.56	4.18	4.36	3.54	4.75	3.69	3.33	3.43
		20%	7.06	4.41	3.69	4.65	5.04	3.71	5.58	4.00	3.37	3.56
	1	0%	4.14	3.36	3.34	3.35	3.35	3.32	3.35	2.76	1.17	0.82
		10%	5.84	3.94	3.56	4.28	5.15	3.60	4.88	3.69	3.33	3.49
		20%	7.80	4.87	3.74	6.22	6.95	3.72	8.32	3.99	3.35	3.65
	2	0%	5.04	3.48	3.40	3.46	3.51	3.38	3.39	3.38	2.57	1.76
		10%	6.24	4.31	3.58	4.69	4.52	3.61	4.68	3.76	3.33	3.47
		20%	7.97	4.92	3.72	5.78	5.73	3.69	7.31	4.28	3.38	3.67

Table 3.4: Case A. Median of $R(Y)$ and $R(\theta)$ for sample size $n = 200$.

3.6.3 Atmospheric pollution data

Finally, an environmental example have been considered: the atmospheric pollution data (see Section 1.1.2 in Chapter 1, page 2). The data here correspond to hourly averaged NO_x concentrations measured in the neighbourhood of a power station of ENDESA, located in As Pontes in the Northwest of Spain, from 2007 to 2009. The aim is to forecast NO_x with half an hour horizon to allow the power plant staff to preclude NO_x concentrations reaching the limits fixed by the current environmental legislation (see a functional kernel and a linear autoregressive approach to this problem with SO_2 levels in Fernández de Castro et al., 2005). Each curve X was built with 240 consecutive minute–by–minute values of hourly averaged NO_x concentration[1], and the NO_x value half an hour ahead was taken as response Y. As in previous examples, different pairs of learning/testing samples were considered, consisted of 750 and 250 observations, respectively. For $\hat{\theta}^{\text{k–NN}}_{k_n}$, the quadratic kernel, and the semi–metric based on the L^2–norm were selected. Table 3.11 (see page 73) compiles the main results. For this dataset, the smallest values of $R(Y)$ correspond to $\hat{\theta}_{PS}$. Among the FPCA–type estimators, $\hat{\theta}^{\alpha_n}_{k_n}$ improves slightly the results of the standard estimator, whereas $\hat{\theta}^{\text{k–NN}}_{k_n}$ gives the largest errors.

Remark 3.6.1. In Corollary 3.2.4 (see page 55), the second order efficiency of the estimator $\hat{\theta}^{\alpha}_{k_n}$ in terms of the conditional estimation error was stated. The existence of an inverse relation between the improvement of $\hat{\theta}^{\alpha}_{k_n}$ with respect to $\hat{\theta}_{k_n}$, and the sample size n was also indicated. This fact is reflected in the real data applications: the gain of the presmoothed estimator decreases from small to large sample size (see Table 3.9, Table 3.10 and Table 3.11, page 73).

[1]Each value of this hourly averaged NO_x concentration is computed as the average of 60 minute–by–minute NO_x values, which correspond to the values at the previous 59 time points and at the current time point.

			GCV				$opt_{R(Y)}$			$opt_{R(\theta)}$		
error	r	*out*	$\hat{\theta}_{PS}$	$\hat{\theta}_{k_n}$	$\hat{\theta}^{\alpha_n}_{k_n}$	$\hat{\theta}^{\text{k–NN}}_{k_n}$	$\hat{\theta}_{k_n}$	$\hat{\theta}^{\alpha_n}_{k_n}$	$\hat{\theta}^{\text{k–NN}}_{k_n}$	$\hat{\theta}_{k_n}$	$\hat{\theta}^{\alpha_n}_{k_n}$	$\hat{\theta}^{\text{k–NN}}_{k_n}$
$R(Y)$	0.02	0%	0.00	0.00	0.00	0.05	0.00	0.00	0.05	0.00	0.00	0.07
		10%	4.26	4.31	4.26	4.32	4.04	3.84	3.80	4.17	4.03	4.09
		20%	10.97	11.07	10.70	11.04	10.13	9.59	9.51	10.36	9.94	10.01
	0.2	0%	0.12	0.12	0.12	0.16	0.10	0.10	0.15	0.11	0.11	0.18
		10%	4.38	4.37	4.34	4.42	4.09	3.91	3.91	4.22	4.11	4.14
		20%	10.93	10.94	10.77	11.05	10.05	9.56	9.54	10.38	10.20	10.25
	0.5	0%	0.71	0.72	0.71	0.78	0.65	0.62	0.69	0.70	0.68	0.76
		10%	5.26	5.22	5.10	5.11	4.68	4.52	4.51	4.93	4.74	4.83
		20%	11.42	11.49	11.21	11.40	10.72	9.96	9.93	11.00	10.67	10.60
	1	0%	2.75	2.82	2.71	2.68	2.52	2.35	2.37	2.62	2.52	2.56
		10%	7.31	7.30	6.93	6.92	6.44	6.04	6.09	6.66	6.38	6.51
		20%	13.96	14.18	13.59	13.44	12.70	12.06	11.94	13.20	12.71	12.69
	2	0%	11.12	11.23	10.73	10.89	10.45	9.93	9.60	10.70	10.26	10.30
		10%	16.60	16.85	16.04	16.14	15.19	13.69	13.86	15.76	14.83	14.97
		20%	22.70	23.19	21.85	21.75	20.37	19.28	18.79	21.02	20.31	20.50
$R(\theta)$	0.02	0%	0.06	0.47	0.45	1.39	0.43	0.43	1.45	0.42	0.40	1.31
		10%	10.27	6.82	4.01	14.75	9.58	3.12	12.59	3.73	1.67	1.86
		20%	17.43	11.98	6.29	128.15	18.73	4.11	24.99	6.96	1.79	2.42
	0.2	0%	0.72	1.40	1.36	1.68	1.41	1.36	1.80	1.16	0.96	1.38
		10%	10.07	6.56	4.07	20.24	7.47	3.22	13.20	3.78	1.70	1.86
		20%	19.90	17.26	6.55	163.58	25.91	7.01	28.16	5.98	2.02	2.42
	0.5	0%	1.45	1.95	2.03	2.47	2.65	1.89	2.88	1.53	1.21	1.42
		10%	9.23	6.87	4.08	16.81	12.16	3.56	15.62	3.60	1.65	1.93
		20%	20.72	14.72	6.29	142.17	19.25	5.14	25.76	7.75	2.07	2.50
	1	0%	5.50	4.08	3.38	5.86	6.35	2.66	11.10	2.39	1.47	1.69
		10%	13.43	10.46	5.06	36.89	11.20	3.50	21.48	4.95	1.67	1.90
		20%	24.08	14.42	6.56	83.28	22.13	4.99	17.68	8.15	2.17	3.00
	2	0%	14.97	9.47	4.48	28.14	19.50	5.46	22.56	5.52	1.84	2.34
		10%	33.05	19.14	6.61	60.22	44.83	6.56	32.91	8.86	2.18	2.52
		20%	43.98	27.31	6.53	141.35	36.98	6.54	41.44	13.36	2.50	2.84

Table 3.5: Case B. Median of $R(Y)$ and $R(\theta)$ for sample size $n = 25$.

3.7 Final conclusions

Throughout this chapter, new FPCA–type estimators for the linear model parameter θ based on different presmoothing techniques have been introduced.

First of all, $\hat{\theta}^{\alpha_n}_{k_n}$ has been proposed, which can be seen as an extension of the ordinary multivariate ridge regression estimator to general Hilbert spaces: one slightly perturbs the eigenvalues of the second moment operator in order to avoid ill–conditioned problems. It has been shown that the presmoothed estimator preserves the consistency properties of the standard FPCA estimator, and expressions for conditional mean square errors for prediction and estimation have been obtained. A remark highlights the effect of the bias term in the conditional prediction error: one can only obtain clear efficiency when the bias is negligible. As far as the conditional estimation error is concerned, one is able to get improvement over the FPCA estimate, especially if the model noise is large and/or the sample size is small. Then, other presmoothing approaches as $\hat{\theta}^{h_n}_{k_n}$, or the estimators based on presmoothed FPCA analysis by Pezzulli and Silverman (1993) and Silverman (1996), have been introduced. The consistency of $\hat{\theta}^{h_n}_{k_n}$, and the conditional error expressions for all of them, have been obtained.

The effectiveness of the presmoothed estimators $\hat{\theta}^{\alpha_n}_{k_n}$ and $\hat{\theta}^{\text{k–NN}}_{k_n}$ relative to the standard FPCA estimator and the penalized B–splines estimator have been tested by means of simulation studies and data applications. In terms of conditional estimation error, $\hat{\theta}^{\alpha_n}_{k_n}$ gives better results than the FPCA estimator, and it is a serious rival to the spline estimator when the sample size is small or the noise level is large. Nevertheless, second order efficiency only generates a clear improvement for small sample size, whereas for n large enough $\hat{\theta}_{k_n}$ and $\hat{\theta}^{\alpha}_{k_n}$ have similar behaviour. On the other hand, the efficiency of $\hat{\theta}^{\text{k–NN}}_{k_n}$ could not be stated. Furthermore, from a practical point of view, it has been seen that this

			GCV				$opt_{R(Y)}$			$opt_{R(\theta)}$		
error	r	*out*	$\hat{\theta}_{PS}$	$\hat{\theta}_{k_n}$	$\hat{\theta}^{\alpha_n}_{k_n}$	$\hat{\theta}^{\text{k–NN}}_{k_n}$	$\hat{\theta}_{k_n}$	$\hat{\theta}^{\alpha_n}_{k_n}$	$\hat{\theta}^{\text{k–NN}}_{k_n}$	$\hat{\theta}_{k_n}$	$\hat{\theta}^{\alpha_n}_{k_n}$	$\hat{\theta}^{\text{k–NN}}_{k_n}$
$R(Y)$	0.02	0%	0.00	0.00	0.00	0.02	0.00	0.00	0.02	0.00	0.00	0.03
		10%	5.17	5.21	5.15	5.26	4.98	4.88	4.92	5.06	5.01	5.05
		20%	10.22	10.28	10.26	10.47	9.93	9.67	9.71	10.06	9.87	9.94
	0.2	0%	0.11	0.11	0.11	0.13	0.10	0.10	0.12	0.11	0.11	0.14
		10%	5.32	5.28	5.27	5.31	5.10	4.98	4.99	5.20	5.10	5.13
		20%	10.35	10.33	10.23	10.57	10.05	9.80	9.81	10.16	10.07	10.08
	0.5	0%	0.66	0.69	0.67	0.67	0.64	0.62	0.63	0.65	0.65	0.67
		10%	5.83	5.89	5.83	5.95	5.65	5.56	5.61	5.74	5.74	5.78
		20%	11.07	11.29	10.96	11.25	10.75	10.41	10.40	10.86	10.63	10.68
	1	0%	2.69	2.63	2.59	2.66	2.56	2.45	2.49	2.58	2.55	2.56
		10%	7.79	7.79	7.67	7.67	7.41	7.23	7.13	7.59	7.48	7.36
		20%	13.61	13.77	13.51	13.73	13.02	12.63	12.65	13.30	13.00	13.35
	2	0%	11.09	10.99	10.70	10.76	10.27	10.07	10.30	10.51	10.41	10.54
		10%	16.32	16.34	15.71	15.76	15.66	14.98	14.93	15.87	15.33	15.31
		20%	21.57	21.54	21.08	21.46	20.29	19.67	19.87	20.65	20.08	20.40
$R(\theta)$	0.02	0%	0.04	0.33	0.33	1.24	0.32	0.31	1.28	0.31	0.29	1.14
		10%	5.14	4.36	3.34	20.58	5.34	3.00	7.43	2.53	1.52	1.88
		20%	8.89	7.39	3.84	71.15	14.95	3.95	12.54	4.05	1.76	2.50
	0.2	0%	0.64	1.28	1.26	1.37	1.23	1.18	1.39	1.03	0.76	1.17
		10%	5.41	4.11	3.30	11.11	6.77	2.58	7.57	2.74	1.45	1.64
		20%	9.38	6.04	3.63	55.30	7.16	3.53	17.77	3.73	1.69	1.99
	0.5	0%	0.97	1.64	1.65	1.95	1.60	1.43	2.10	1.35	1.16	1.28
		10%	6.34	4.85	3.70	15.70	5.35	2.44	6.36	2.81	1.46	1.77
		20%	10.21	10.26	5.51	75.29	9.80	3.55	14.70	4.34	1.75	2.45
	1	0%	2.68	3.26	2.92	4.18	3.51	2.20	4.06	1.94	1.37	1.49
		10%	6.60	5.48	3.38	23.25	6.73	3.04	11.19	3.01	1.61	1.76
		20%	12.60	8.40	4.17	74.19	15.33	4.21	12.97	4.21	1.68	2.52
	2	0%	9.65	6.28	3.64	14.97	9.41	3.96	16.15	3.69	1.63	1.97
		10%	9.11	7.06	4.20	23.10	12.08	3.62	26.68	4.68	1.70	2.32
		20%	18.37	12.81	6.36	82.13	17.30	5.38	21.93	7.31	2.18	3.06

Table 3.6: Case B. Median of $R(Y)$ and $R(\theta)$ for sample size $n = 50$.

estimator reduces the conditional error of $\hat{\theta}_{k_n}$ in some cases, whereas it gives larger errors than the FPCA estimator in other ones.

3.8 Appendix Chapter 3

In this appendix, the proofs of all the theorems and corollaries (in order of appearance) introduced throughout the chapter have been compiled, jointly with the technical lemmas required to prove them.

3.8.1 Proof of Theorem 3.2.2

This proof is similar to that of Theorem 3.2 in Cardot et al. (1999).

First, take $m_{k_n} = \Delta\Pi_{k_n}(\Pi_{k_n}\Gamma\Pi_{k_n})^{-1}$, where Π_{k_n} is the orthogonal projection onto the space spanned by the first k_n eigenfunctions of Γ. It is clear that

$$\|m - \hat{m}^{\alpha_n}_{k_n}\|_{\mathcal{H}'} \leq \|m - m_{k_n}\|_{\mathcal{H}'} + \|m_{k_n} - \hat{m}^{\alpha_n}_{k_n}\|_{\mathcal{H}'}.$$

Cardot et al. (1999) showed that $\|m - m_{k_n}\|_{\mathcal{H}'} \to 0$, so one just needs to show that $\|m_{k_n} - \hat{m}^{\alpha_n}_{k_n}\|_{\mathcal{H}'} \to 0$.

Let $E_n = \{\lambda_{k_n}/2 < \hat{\lambda}_{k_n} < 3\lambda_{k_n}/2\}$. In E_n, Lemma 3.8.1 (see page 74) ensures that

$$\|m_{k_n} - \hat{m}^{\alpha_n}_{k_n}\|_{\mathcal{H}'} \leq \delta_n\|\Delta\|_{\mathcal{H}'}\|\Gamma - \Gamma_n\|_\infty + 2\frac{\|\Delta - \Delta_n\|_{\mathcal{H}'}}{\lambda_{k_n}} + 2\alpha_n\frac{\|\Delta\|_{\mathcal{H}'}}{\lambda^2_{k_n}},$$

			GCV				$opt_{R(Y)}$			$opt_{R(\theta)}$		
error	r	*out*	$\hat{\theta}_{PS}$	$\hat{\theta}_{k_n}$	$\hat{\theta}^{\alpha_n}_{k_n}$	$\hat{\theta}^{\text{k–NN}}_{k_n}$	$\hat{\theta}_{k_n}$	$\hat{\theta}^{\alpha_n}_{k_n}$	$\hat{\theta}^{\text{k–NN}}_{k_n}$	$\hat{\theta}_{k_n}$	$\hat{\theta}^{\alpha_n}_{k_n}$	$\hat{\theta}^{\text{k–NN}}_{k_n}$
$R(Y)$	0.02	0%	0.00	0.00	0.00	0.02	0.00	0.00	0.01	0.00	0.00	0.02
		10%	5.01	5.05	5.00	5.02	4.94	4.89	4.89	4.98	4.96	4.96
		20%	9.98	9.99	9.93	10.09	9.82	9.74	9.75	9.93	9.86	9.87
	0.2	0%	0.11	0.11	0.11	0.12	0.11	0.11	0.12	0.11	0.11	0.13
		10%	5.15	5.17	5.14	5.19	5.06	5.00	5.00	5.09	5.09	5.10
		20%	10.16	10.20	10.14	10.31	10.02	9.91	9.96	10.12	10.06	10.07
	0.5	0%	0.66	0.68	0.67	0.69	0.66	0.65	0.66	0.66	0.67	0.68
		10%	5.69	5.71	5.65	5.67	5.61	5.52	5.55	5.65	5.62	5.62
		20%	10.68	10.68	10.64	10.74	10.51	10.36	10.38	10.58	10.49	10.49
	1	0%	2.57	2.59	2.57	2.58	2.52	2.48	2.52	2.57	2.52	2.55
		10%	7.66	7.70	7.67	7.66	7.57	7.45	7.42	7.63	7.56	7.55
		20%	13.00	13.01	12.86	13.07	12.55	12.44	12.56	12.75	12.67	12.72
	2	0%	10.70	10.72	10.54	10.57	10.37	10.26	10.26	10.56	10.42	10.51
		10%	15.68	15.80	15.51	15.41	15.28	14.88	14.99	15.51	15.14	15.37
		20%	20.52	20.90	20.60	20.57	20.21	19.89	19.83	20.36	20.19	20.10
$R(\theta)$	0.02	0%	0.02	0.25	0.24	1.10	0.24	0.24	1.12	0.23	0.23	0.96
		10%	2.40	2.82	2.35	5.86	3.35	2.03	4.21	1.77	1.31	1.51
		20%	4.47	3.90	2.93	30.94	5.38	2.86	6.85	2.72	1.44	1.77
	0.2	0%	0.60	1.23	1.13	1.24	1.15	1.01	1.22	0.86	0.65	1.01
		10%	2.43	2.63	2.43	6.47	3.45	2.11	4.03	1.85	1.31	1.43
		20%	4.78	4.50	3.06	32.77	4.96	2.81	4.98	2.43	1.57	1.87
	0.5	0%	0.79	1.46	1.48	1.52	1.43	1.31	1.48	1.23	1.04	1.14
		10%	3.07	3.10	2.74	6.14	3.91	2.25	4.84	2.01	1.32	1.40
		20%	3.68	3.10	2.56	40.80	4.56	2.57	7.37	2.24	1.39	1.73
	1	0%	1.51	1.80	1.83	2.69	3.15	1.94	2.62	1.56	1.25	1.37
		10%	3.96	3.42	2.68	5.35	4.65	2.42	5.50	2.36	1.53	1.62
		20%	3.86	3.97	2.64	30.56	5.61	3.12	8.28	2.50	1.58	1.89
	2	0%	4.42	3.60	2.77	13.52	5.58	2.83	6.18	2.39	1.44	1.77
		10%	6.06	5.09	3.21	15.05	9.50	3.35	9.36	3.09	1.53	1.90
		20%	9.16	7.64	4.09	30.66	10.02	3.53	9.57	4.17	1.68	2.22

Table 3.7: Case B. Median of $R(Y)$ and $R(\theta)$ for sample size $n = 100$.

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

$$\mathbb{P}(\|m_{k_n} - \hat{m}_{k_n}^{\alpha_n}\|_{\mathcal{H}'} > \eta) \leq \mathbb{P}\left(\|\Gamma - \Gamma_n\|_\infty > \frac{\eta}{3\delta_n\|\Delta\|_{\mathcal{H}'}}\right) + \mathbb{P}\left(\|\Delta - \Delta_n\|_{\mathcal{H}'} > \frac{\lambda_{k_n}\eta}{6}\right) + \mathbb{I}_{\{\alpha_n > \eta\lambda_{k_n}^2/(6\|\Delta\|_{\mathcal{H}'})\}} + \mathbb{P}(\overline{E}_n). \tag{3.5}$$

It can be shown that

$$\mathbb{P}(\overline{E}_n) \leq \mathbb{P}(\|\Gamma - \Gamma_n\|_\infty > \lambda_{k_n}/2) \leq 2\exp\left(-Cn\lambda_{k_n}^2\right), \tag{3.6}$$

where C is a positive constant independent of n, and the last inequality is derived from Lemma 5.3 in Cardot et al. (1999). Furthermore,

$$\mathbb{P}\left(\|\Gamma - \Gamma_n\|_\infty > \frac{\eta}{3\delta_n\|\Delta\|_{\mathcal{H}'}}\right) \leq 2\exp\left(-\frac{A_\eta n}{\delta_n^2}\right), \tag{3.7}$$

$$\mathbb{P}\left(\|\Delta - \Delta_n\|_{\mathcal{H}'} > \frac{\lambda_{k_n}\eta}{6}\right) \leq 2\exp\left(-B_\eta n\lambda_{k_n}^2\right) \tag{3.8}$$

where A_η and B_η are positive constants independent of n. These two inequalities are obtained using Lemma 5.3 in Cardot et al. (1999), and Lemma 2.5.1 (see Chapter 2, page 47). Using (3.6), (3.7), and (3.8) in (3.5), one gets

$$\mathbb{P}(\|m_{k_n} - \hat{m}_{k_n}^{\alpha_n}\|_{\mathcal{H}'} > \eta) \leq 2\exp\left(-\frac{A_\eta n}{\delta_n^2}\right) + 2\exp\left(-B_\eta n\lambda_{k_n}^2\right) + \mathbb{I}_{\{\alpha_n > \eta\lambda_{k_n}^2/(6\|\Delta\|_{\mathcal{H}'})\}} + 2\exp\left(-Cn\lambda_{k_n}^2\right).$$

			GCV				$opt_{R(Y)}$			$opt_{R(\theta)}$		
error	r	*out*	$\hat{\theta}_{PS}$	$\hat{\theta}_{k_n}$	$\hat{\theta}^{\alpha_n}_{k_n}$	$\hat{\theta}^{\text{k–NN}}_{k_n}$	$\hat{\theta}_{k_n}$	$\hat{\theta}^{\alpha_n}_{k_n}$	$\hat{\theta}^{\text{k–NN}}_{k_n}$	$\hat{\theta}_{k_n}$	$\hat{\theta}^{\alpha_n}_{k_n}$	$\hat{\theta}^{\text{k–NN}}_{k_n}$
$R(Y)$	0.02	0%	0.00	0.00	0.00	0.01	0.00	0.00	0.01	0.00	0.00	0.01
		10%	4.95	4.97	4.94	4.97	4.91	4.89	4.89	4.94	4.93	4.93
		20%	9.91	9.91	9.90	9.95	9.83	9.79	9.79	9.87	9.85	9.85
	0.2	0%	0.11	0.11	0.11	0.12	0.11	0.11	0.11	0.11	0.11	0.12
		10%	5.08	5.08	5.07	5.09	5.03	5.00	5.01	5.06	5.04	5.05
		20%	10.00	10.02	9.99	10.04	9.92	9.86	9.86	9.96	9.95	9.94
	0.5	0%	0.66	0.66	0.66	0.67	0.65	0.65	0.65	0.66	0.65	0.66
		10%	5.68	5.71	5.68	5.68	5.62	5.57	5.59	5.65	5.63	5.61
		20%	10.63	10.63	10.60	10.66	10.51	10.42	10.47	10.56	10.51	10.54
	1	0%	2.61	2.64	2.63	2.62	2.59	2.58	2.57	2.61	2.60	2.60
		20%	7.65	7.65	7.62	7.66	7.53	7.48	7.53	7.56	7.54	7.60
		10%	12.44	12.47	12.42	12.45	12.35	12.28	12.27	12.40	12.37	12.35
	2	0%	10.53	10.50	10.54	10.54	10.44	10.39	10.39	10.49	10.43	10.46
		10%	15.41	15.47	15.37	15.43	15.24	15.18	15.19	15.28	15.32	15.34
		20%	20.59	20.57	20.59	20.50	20.23	20.08	20.05	20.32	20.23	20.23
$R(\theta)$	0.02	0%	0.01	0.22	0.21	0.95	0.21	0.21	1.00	0.21	0.21	0.88
		10%	1.85	1.98	2.04	4.09	2.17	1.71	2.57	1.49	1.22	1.31
		20%	1.91	2.21	2.17	11.11	3.04	1.93	5.03	1.81	1.29	1.51
	0.2	0%	0.53	1.02	0.79	1.15	0.88	0.75	1.15	0.60	0.49	0.91
		10%	1.51	2.04	1.92	2.99	2.17	1.65	2.39	1.56	1.21	1.36
		20%	2.43	2.55	2.12	15.12	3.43	1.93	3.61	1.86	1.28	1.40
	0.5	0%	0.68	1.34	1.32	1.36	1.28	1.22	1.28	1.15	0.89	1.10
		10%	2.08	2.23	2.31	3.73	2.43	1.75	2.56	1.57	1.24	1.33
		20%	2.46	2.46	2.26	11.24	2.99	1.94	4.12	1.71	1.26	1.48
	1	0%	1.23	1.65	1.59	1.76	1.69	1.46	2.03	1.33	1.14	1.22
		10%	2.36	2.23	2.39	4.88	2.73	1.86	3.61	1.63	1.27	1.35
		20%	2.86	2.68	2.58	8.20	4.05	2.22	4.32	2.09	1.38	1.47
	2	0%	2.43	2.44	2.32	3.71	3.57	2.19	3.78	1.81	1.29	1.46
		10%	3.75	3.43	2.80	6.60	4.64	2.84	6.45	2.19	1.49	1.66
		20%	4.85	3.85	3.21	17.51	5.72	2.48	5.96	2.47	1.44	1.76

Table 3.8: Case B. Median of $R(Y)$ and $R(\theta)$ for sample size $n = 200$.

	GCV			
error	$\hat{\theta}_{PS}$	$\hat{\theta}_{k_n}$	$\hat{\theta}^{\alpha_n}_{k_n}$	$\hat{\theta}^{\text{k–NN}}_{k_n}$
mean($R(Y)$)	0.03698	0.03566	0.03293	0.02601
median($R(Y)$)	0.03626	0.03466	0.03230	0.02585
sd($R(Y)$)	0.01772	0.01579	0.01436	0.01483

Table 3.9: Canadian weather data. Mean, median and standard deviation of $R(Y)$.

	GCV			
error	$\hat{\theta}_{PS}$	$\hat{\theta}_{k_n}$	$\hat{\theta}^{\alpha_n}_{k_n}$	$\hat{\theta}^{\text{k–NN}}_{k_n}$
mean($R(Y)$)	6.656	6.926	6.604	8.072
median($R(Y)$)	6.313	6.524	6.088	7.598
sd($R(Y)$)	1.935	2.153	1.939	2.157

Table 3.10: Spectrometric data. Mean, median and standard deviation of $R(Y)$.

	GCV			
error	$\hat{\theta}_{PS}$	$\hat{\theta}_{k_n}$	$\hat{\theta}^{\alpha_n}_{k_n}$	$\hat{\theta}^{\text{k–NN}}_{k_n}$
mean($R(Y)$)	2.865	3.238	3.207	19.664
median($R(Y)$)	2.780	3.074	3.062	19.202
sd($R(Y)$)	0.966	0.866	0.833	5.221

Table 3.11: Atmospheric pollution data. Mean, median and standard deviation of $R(Y)$.

Cardot et al. (1999) showed $\exp(-A_\eta n/\delta_n^2)$, $\exp(-B_\eta n\lambda_{k_n}^2)$ and $\exp(-Cn\lambda_{k_n}^2)$ are general terms of convergent series under hypothesis **(C.2.8)**. On the other hand, under **(C.3.1)**, there is an n_0 such that $\alpha_n < \eta\lambda_{k_n}^2/(6\|\Delta\|_{\mathcal{H}'})$ for all $n > n_0$. Therefore, $\sum_{n\in\mathbb{N}^*}\mathbb{P}(\|m_{k_n} - \hat{m}_{k_n}^{\alpha_n}\|_{\mathcal{H}'} > \eta) < \infty$, and Borel–Cantelli Lemma gives $\|m_{k_n} - \hat{m}_{k_n}^{\alpha_n}\|_{\mathcal{H}'} \to 0$ *a.s.*

3.8.2 Formulation and proof of Lemma 3.8.1

Lemma 3.8.1. *With* $\gamma_n = \|\Delta\|_{\mathcal{H}'}\{1/(\lambda_{k_n}\hat{\lambda}_{k_n}) + 2(1/\lambda_{k_n} + 1/\hat{\lambda}_{k_n})\sum_{j=1}^{k_n} a_j\}$,

$$\|m_{k_n} - \hat{m}_{k_n}^{\alpha_n}\|_{\mathcal{H}'} \leq \gamma_n\|\Gamma - \Gamma_n\|_\infty + \frac{\|\Delta - \Delta_n\|_{\mathcal{H}'}}{\hat{\lambda}_{k_n}} + \alpha_n\frac{\|\Delta\|_{\mathcal{H}'}}{\lambda_{k_n}\hat{\lambda}_{k_n}}.$$

Proof. Take $\tilde{\Gamma}_{k_n} = \sum_{j=1}^{k_n}\lambda_j\hat{v}_j \otimes_{\mathcal{H}} \hat{v}_j$. The first step is to write

$$\begin{aligned}\|m_{k_n} - \hat{m}_{k_n}^{\alpha_n}\|_{\mathcal{H}'} \leq{}& \|\Delta\Pi_{k_n}\|_{\mathcal{H}'}\|(\Pi_{k_n}\Gamma\Pi_{k_n})^{-1} - \tilde{\Gamma}_{k_n}^{-1}\|_\infty + \|\Delta\Pi_{k_n}\|_{\mathcal{H}'}\|\tilde{\Gamma}_{k_n}^{-1} - (\hat{\Pi}_{k_n}(\Gamma_n + \alpha_n I)\hat{\Pi}_{k_n})^{-1}\|_\infty \\ &+ \|\Delta\Pi_{k_n} - \Delta_n\hat{\Pi}_{k_n}\|_{\mathcal{H}'}\|(\hat{\Pi}_{k_n}(\Gamma_n + \alpha_n I)\hat{\Pi}_{k_n})^{-1}\|_\infty.\end{aligned} \tag{3.9}$$

From (11) and (14) in the proof of Lemma 5.1 in Cardot et al. (1999),

$$\|\Delta\Pi_{k_n}\|_{\mathcal{H}'}\|(\Pi_{k_n}\Gamma\Pi_{k_n})^{-1} - \tilde{\Gamma}_{k_n}^{-1}\|_\infty \leq 2\frac{\|\Delta\|_{\mathcal{H}'}}{\lambda_{k_n}}\|\Gamma - \Gamma_n\|_\infty\sum_{j=1}^{k_n} a_j, \tag{3.10}$$

$$\|\Delta\Pi_{k_n} - \Delta_n\hat{\Pi}_{k_n}\|_{\mathcal{H}'} \leq 2\|\Delta\|_{\mathcal{H}'}\|\Gamma - \Gamma_n\|_\infty\sum_{j=1}^{k_n} a_j + \|\Delta - \Delta_n\|_{\mathcal{H}'}. \tag{3.11}$$

Moreover, with arguments as in the proof of Lemma 5.1 in Cardot et al. (1999), one has

$$\|\Delta\Pi_{k_n}\|_{\mathcal{H}'}\|\tilde{\Gamma}_{k_n}^{-1} - (\hat{\Pi}_{k_n}(\Gamma_n + \alpha_n I)\hat{\Pi}_{k_n})^{-1}\|_\infty \leq \frac{\|\Delta\|_{\mathcal{H}'}}{\lambda_{k_n}\hat{\lambda}_{k_n}}(\|\Gamma - \Gamma_n\|_\infty + \alpha_n), \tag{3.12}$$

$$\|(\hat{\Pi}_{k_n}(\Gamma_n + \alpha_n I)\hat{\Pi}_{k_n})^{-1}\|_\infty \leq \frac{1}{\hat{\lambda}_{k_n}}. \tag{3.13}$$

Hence, using (3.10), (3.11), (3.12), and (3.13) in (3.9), one gets

$$\begin{aligned}\|m_{k_n} - \hat{m}_{k_n}^{\alpha_n}\|_{\mathcal{H}'} \leq{}& \|\Delta\|_{\mathcal{H}'}\left(\frac{1}{\lambda_{k_n}\hat{\lambda}_{k_n}} + 2\left(\frac{1}{\lambda_{k_n}} + \frac{1}{\hat{\lambda}_{k_n}}\right)\sum_{j=1}^{k_n} a_j\right)\|\Gamma - \Gamma_n\|_\infty + \frac{\|\Delta - \Delta_n\|_{\mathcal{H}'}}{\hat{\lambda}_{k_n}} \\ &+ \alpha_n\frac{\|\Delta\|_{\mathcal{H}'}}{\lambda_{k_n}\hat{\lambda}_{k_n}}.\end{aligned}$$

■

3.8.3 Proof of Theorem 3.2.3

Consider Lemma 2.5.2 (see Chapter 2, page 49), with $\gamma_j = (\hat{\lambda}_j + \alpha_n)^{-1}$ and $w_j = \hat{v}_j$. Then,

$$R_{k_n}^{(\gamma,w)} = \theta - \sum_{j=1}^{k_n}\frac{\hat{\lambda}_j}{\hat{\lambda}_j + \alpha_n}\langle\hat{v}_j, \theta\rangle\hat{v}_j = \hat{R}_{k_n} + \sum_{j=1}^{k_n}\frac{\alpha_n}{\hat{\lambda}_j + \alpha_n}\langle\hat{v}_j, \theta\rangle\hat{v}_j,$$

with $\hat{R}_{k_n}$ defined in (2.7) (see Chapter 2, page 39). One then obtains for the conditional prediction error

$$\begin{aligned}\mathbb{E}_{\mathcal{X}^{n+1}}(Y_{n+1} - \langle \hat{\theta}_{k_n}^{\alpha_n}, X_{n+1}\rangle)^2 = \sigma^2 + \frac{\sigma^2}{n}\sum_{j=1}^{k_n}\frac{\hat{\lambda}_j}{(\hat{\lambda}_j+\alpha_n)^2}\langle X_{n+1},\hat{v}_j\rangle^2 \\ + \langle X_{n+1}, \hat{R}_{k_n} + \sum_{j=1}^{k_n}\frac{\alpha_n}{\hat{\lambda}_j+\alpha_n}\langle \hat{v}_j,\theta\rangle\hat{v}_j\rangle^2.\end{aligned} \tag{3.14}$$

Some calculations and **(C.3.2)** allow to obtain

$$\frac{\hat{\lambda}_j}{(\hat{\lambda}_j+\alpha_n)^2} = \frac{1}{\hat{\lambda}_j} - 2\frac{\alpha_n}{\hat{\lambda}_j^2} + \frac{\alpha_n^2}{\hat{\lambda}_j(\hat{\lambda}_j+\alpha_n)^2}\left(3+2\frac{\alpha_n}{\hat{\lambda}_j}\right) = \frac{1}{\hat{\lambda}_j} - 2\frac{\alpha_n}{\hat{\lambda}_j^2}(1+o_{a.s.}(1)), \tag{3.15}$$

$$\frac{\alpha_n}{\hat{\lambda}_j+\alpha_n} = \frac{\alpha_n}{\hat{\lambda}_j}\left(1-\frac{\alpha_n}{\hat{\lambda}_j+\alpha_n}\right) = \frac{\alpha_n}{\hat{\lambda}_j}(1+o_{a.s.}(1)). \tag{3.16}$$

Using (3.15) and (3.16) in (3.14), one gets

$$\begin{aligned}\mathbb{E}_{\mathcal{X}^{n+1}}(Y_{n+1} - \langle \hat{\theta}_{k_n}^{\alpha_n}, X_{n+1}\rangle)^2 = \sigma^2 + \frac{\sigma^2}{n}\sum_{j=1}^{k_n}\frac{\langle X_{n+1},\hat{v}_j\rangle^2}{\hat{\lambda}_j} - 2\alpha_n\frac{\sigma^2}{n}\sum_{j=1}^{k_n}\frac{\langle X_{n+1},\hat{v}_j\rangle^2}{\hat{\lambda}_j^2}(1+o_{a.s.}(1)) \\ + \langle X_{n+1},\hat{R}_{k_n}\rangle^2 + 2\alpha_n\langle X_{n+1},\hat{R}_{k_n}\rangle\langle X_{n+1},\hat{T}_{k_n}\rangle(1+o_{a.s.}(1)) + \alpha_n^2\langle X_{n+1},\hat{T}_{k_n}\rangle^2(1+o_{a.s.}(1)),\end{aligned}$$

where $\hat{T}_{k_n} = \sum_{j=1}^{k_n}\hat{\lambda}_j^{-1}\langle\theta,\hat{v}_j\rangle\hat{v}_j$. Comparing this expression with the conditional prediction error for $\hat{\theta}_{k_n}$ given in Theorem 2.3.14 (see Chapter 2, page 39), one gets

$$\begin{aligned}\mathbb{E}_{\mathcal{X}^{n+1}}(Y_{n+1} - \langle \hat{\theta}_{k_n}^{\alpha_n}, X_{n+1}\rangle)^2 - \mathbb{E}_{\mathcal{X}^{n+1}}(Y_{n+1} - \langle \hat{\theta}_{k_n}, X_{n+1}\rangle)^2 = -2\alpha_n\frac{\sigma^2}{n}\sum_{j=1}^{k_n}\frac{\langle X_{n+1},\hat{v}_j\rangle^2}{\hat{\lambda}_j^2}(1+o_{a.s.}(1)) \\ + 2\alpha_n\langle X_{n+1},\hat{R}_{k_n}\rangle\langle X_{n+1},\hat{T}_{k_n}\rangle(1+o_{a.s.}(1)) + \alpha_n^2\langle X_{n+1},\hat{T}_{k_n}\rangle^2(1+o_{a.s.}(1)).\end{aligned}$$

On the other hand, for the conditional estimation error Lemma 2.5.2 (see Chapter 2, page 49) implies

$$\mathbb{E}_{\mathcal{X}^n}(\|\theta - \hat{\theta}_{k_n}^{\alpha_n}\|^2) = \frac{\sigma^2}{n}\sum_{j=1}^{k_n}\frac{\hat{\lambda}_j}{(\hat{\lambda}_j+\alpha_n)^2} + \|\hat{R}_{k_n}\|^2 + \left\|\sum_{j=1}^{k_n}\frac{\alpha_n}{\hat{\lambda}_j+\alpha_n}\langle\hat{v}_j,\theta\rangle\hat{v}_j\right\|^2. \tag{3.17}$$

Using (3.15) and (3.16) in (3.17), one has

$$\mathbb{E}_{\mathcal{X}^n}(\|\theta - \hat{\theta}_{k_n}^{\alpha_n}\|^2) = \frac{\sigma^2}{n}\sum_{j=1}^{k_n}\frac{1}{\hat{\lambda}_j} - 2\alpha_n\frac{\sigma^2}{n}\sum_{j=1}^{k_n}\frac{1}{\hat{\lambda}_j^2}(1+o_{a.s.}(1)) + \|\hat{R}_{k_n}\|^2 + \alpha_n^2\|\hat{T}_{k_n}\|^2(1+o_{a.s.}(1)),$$

with $\hat{T}_{k_n}$ defined as before. Bearing in mind Theorem 2.3.14 (see Chapter 2, page 39),

$$\mathbb{E}_{\mathcal{X}^n}(\|\theta - \hat{\theta}_{k_n}^{\alpha_n}\|^2) - \mathbb{E}_{\mathcal{X}^n}(\|\theta - \hat{\theta}_{k_n}\|^2) = -2\alpha_n\frac{\sigma^2}{n}\sum_{j=1}^{k_n}\frac{1}{\hat{\lambda}_j^2}(1+o_{a.s.}(1)) + \alpha_n^2\|\hat{T}_{k_n}\|^2(1+o_{a.s.}(1)).$$

3.8.4 Proof of Corollary 3.2.4

Note that

$$\begin{aligned}\langle X_{n+1},\hat{T}_{k_n}\rangle^2 = \left(\sum_{j=1}^{k_n}\frac{\langle\theta,\hat{v}_j\rangle}{\hat{\lambda}_j}\langle X_{n+1},\hat{v}_j\rangle\right)^2 \le \left(\sum_{j=1}^{k_n}\langle\theta,\hat{v}_j\rangle^2\right)\left(\sum_{j=1}^{k_n}\frac{\langle X_{n+1},\hat{v}_j\rangle^2}{\hat{\lambda}_j^2}\right) \\ \le \|\theta\|^2\sum_{j=1}^{k_n}\frac{\langle X_{n+1},\hat{v}_j\rangle^2}{\hat{\lambda}_j^2},\end{aligned}$$

and so $\alpha_n\langle X_{n+1},\hat{T}_{k_n}\rangle^2(n^{-1}\sum_{j=1}^{k_n}\hat{\lambda}_j^{-2}\langle X_{n+1},\hat{v}_j\rangle^2)^{-1}\leq n\alpha_n\|\theta\|^2$. Hence, using Theorem 3.2.3 and **(C.3.3)**, one has

$$\mathbb{E}_{\mathcal{X}^{n+1}}(Y_{n+1}-\langle\hat{\theta}_{k_n}^{\alpha_n},X_{n+1}\rangle)^2-\mathbb{E}_{\mathcal{X}^{n+1}}(Y_{n+1}-\langle\hat{\theta}_{k_n},X_{n+1}\rangle)^2=-2\alpha_n\frac{\sigma^2}{n}\sum_{j=1}^{k_n}\frac{\langle X_{n+1},\hat{v}_j\rangle^2}{\hat{\lambda}_j^2}(1+o_{a.s.}(1))$$
$$+2\alpha_n\langle X_{n+1},\hat{R}_{k_n}\rangle\langle X_{n+1},\hat{T}_{k_n}\rangle(1+o_{a.s.}(1)).$$

Furthermore,

$$\|\hat{T}_{k_n}\|^2=\sum_{j=1}^{k_n}\frac{\langle\theta,\hat{v}_j\rangle^2}{\hat{\lambda}_j^2}\leq\|\theta\|^2\sum_{j=1}^{k_n}\frac{1}{\hat{\lambda}_j^2}.$$

Then $\alpha_n\|\hat{T}_{k_n}\|^2(n^{-1}\sum_{j=1}^{k_n}\hat{\lambda}_j^{-2})\leq n\alpha_n\|\theta\|^2$. Applying this inequality and **(C.3.3)** to Theorem 3.2.3, one gets

$$\mathbb{E}_{\mathcal{X}^n}(\|\theta-\hat{\theta}_{k_n}^{\alpha_n}\|^2)-\mathbb{E}_{\mathcal{X}^n}(\|\theta-\hat{\theta}_{k_n}\|^2)=-2\alpha_n\frac{\sigma^2}{n}\sum_{j=1}^{k_n}\frac{1}{\hat{\lambda}_j^2}(1+o_{a.s.}(1)).$$

3.8.5 Proof of Theorem 3.3.2

The notation and the structure of this proof are the same as that of Theorem 3.2.2.

One gets

$$\|m-\hat{m}_{k_n}^{h_n}\|_{\mathcal{H}'}\leq\|m-m_{k_n}\|_{\mathcal{H}'}+\|m_{k_n}-\hat{m}_{k_n}^{h_n}\|_{\mathcal{H}'},$$

where $m_{k_n}=\Delta\Pi_{k_n}(\Pi_{k_n}\Gamma\Pi_{k_n})^{-1}$ being Π_{k_n} the orthogonal projection onto the space spanned by the first k_n eigenfunctions of Γ. Thus one only has to show that $\|m_{k_n}-\hat{m}_{k_n}^{h_n}\|_{\mathcal{H}'}\to 0$, since Cardot et al. (1999) showed that $\|m-m_{k_n}\|_{\mathcal{H}'}\to 0$.

Applying Lemma 3.8.2 (see page 77), one gets in E_n

$$\|m_{k_n}-\hat{m}_{k_n}^{h_n}\|_{\mathcal{H}'}\leq\delta_n\|\Delta\|_{\mathcal{H}'}\|\Gamma-\Gamma_n\|_\infty+2\frac{\|\Delta-\Delta_n^{h_n}\|_{\mathcal{H}'}}{\lambda_{k_n}},$$

being $E_n=\{\lambda_{k_n}/2<\hat{\lambda}_{k_n}<3\lambda_{k_n}/2\}$, and $\delta_n=2/\lambda_{k_n}^2+6\sum_{j=1}^{k_n}a_j/\lambda_{k_n}$. Consequently,

$$\mathbb{P}(\|m_{k_n}-\hat{m}_{k_n}^{h_n}\|_{\mathcal{H}'}>\eta)\leq\mathbb{P}\left(\|\Gamma-\Gamma_n\|_\infty>\frac{\eta}{2\delta_n\|\Delta\|_{\mathcal{H}'}}\right)+\mathbb{P}\left(\|\Delta-\Delta_n^{h_n}\|_{\mathcal{H}'}>\frac{\lambda_{k_n}\eta}{4}\right)+\mathbb{P}(\overline{E}_n).\quad(3.18)$$

Using (3.6), (3.7), and Lemma 3.8.3 (see page 77) in (3.18), one has

$$\mathbb{P}(\|m_{k_n}-\hat{m}_{k_n}^{h_n}\|_{\mathcal{H}'}>\eta)\leq 2\exp(-A_\eta n/\delta_n^2)+4\exp(-B_\eta n\lambda_{k_n}^2)+2\exp(-Cn\lambda_{k_n}^2)$$
$$+\mathbb{P}\left(\sup_{x\in\mathcal{C}}|m(x)-\hat{m}_{h_n}(x)|>C_\eta\lambda_{k_n}\right).$$

where C, A_η, B_η and C_η are positive constants independent of n. Cardot et al. (1999) showed $\exp(-A_\eta n/\delta_n^2)$, $\exp(-B_\eta n\lambda_{k_n}^2)$ and $\exp(-Cn\lambda_{k_n}^2)$ are general terms of convergent series under hypothesis **(C.2.8)**. On the other hand, Theorem 2.4.11 (see Chapter 2, page 45) and **(C.3.5)** ensure that

$$\mathbb{P}\left(\sup_{x\in\mathcal{C}}|m(x)-\hat{m}_{h_n}(x)|>C_\eta\lambda_{k_n}\right)=0,$$

for n large enough. As a result, one obtains that $\sum_{n\in\mathbb{N}^*}\mathbb{P}(\|m_{k_n}-\hat{m}_{k_n}^{h_n}\|_{\mathcal{H}'}>\eta)<\infty$ and, applying Borel–Cantelli Lemma, one has $\|m_{k_n}-\hat{m}_{k_n}^{h_n}\|_{\mathcal{H}'}\to 0\quad a.s.$

3.8.6 Formulation and proof of Lemma 3.8.2

Lemma 3.8.2. *With* $\gamma_n = \|\Delta\|_{\mathcal{H}'}\{1/(\lambda_{k_n}\hat{\lambda}_{k_n}) + 2(1/\lambda_{k_n} + 1/\hat{\lambda}_{k_n})\sum_{j=1}^{k_n} a_j\}$,

$$\|m_{k_n} - \hat{m}_{k_n}^{h_n}\|_{\mathcal{H}'} \leq \gamma_n \|\Gamma - \Gamma_n\|_\infty + \frac{\|\Delta - \Delta_n^{h_n}\|_{\mathcal{H}'}}{\hat{\lambda}_{k_n}}.$$

Proof. It suffices to follow the proof of Lemma 5.1 in Cardot et al. (1999), replacing Δ_n with $\Delta_n^{h_n}$. ■

3.8.7 Formulation and proof of Lemma 3.8.3

Lemma 3.8.3. *Under* **(C.2.6)**, **(C.2.7)**, *and* **(C.3.4)** *(see Chapter 2, page 38 and page 59), it holds that*

$$\mathbb{P}(\|\Delta_n^{h_n} - \Delta\|_{\mathcal{H}'} > \xi) \leq 4\exp\left(-\frac{\xi^2 n}{2\,c_5(c_5 + c_6\xi)}\right) + \mathbb{P}\left(\sup_{x\in\mathcal{C}}|m(x) - \hat{m}_{h_n}(x)| > \frac{\xi}{c_7}\right),$$

where c_5, c_6 *and* c_7 *are positive constants.*

Proof. This lemma adapts the part (b) of Lemma 5.3 in Cardot et al. (1999) to the weaker assumption **(C.2.7)** and the operator $\Delta_n^{h_n}$.

Remark that

$$\|\Delta_n^{h_n} - \Delta\|_{\mathcal{H}'} \leq \|\Delta_n^{h_n} - \Delta_n\|_{\mathcal{H}'} + \|\Delta_n - \Delta\|_{\mathcal{H}'},$$

where, by **(C.2.6)**,

$$\|\Delta_n^{h_n} - \Delta_n\|_{\mathcal{H}'} = \left\|\frac{1}{n}\sum_{i=1}^n X_i \otimes_{\mathcal{H}'} (Y_i - \hat{m}_{h_n}(X_i))\right\|_{\mathcal{H}'} \leq \frac{1}{n}\left\|\sum_{i=1}^n X_i \otimes_{\mathcal{H}'} \epsilon_i\right\|_{\mathcal{H}'} + c_1 \sup_{x\in\mathcal{C}}|m(x) - \hat{m}_{h_n}(x)|.$$

Therefore, taking $U_i = X_i \otimes_{\mathcal{H}'} \epsilon_i$ for $i = 1, \ldots, n$, one has

$$\begin{aligned}\mathbb{P}(\|\Delta_n^{h_n} - \Delta\|_{\mathcal{H}'} > \xi) \leq\ & \mathbb{P}\left(\frac{1}{n}\|\sum_{i=1}^n U_i\|_{\mathcal{H}'} > \frac{\xi}{3}\right) + \mathbb{P}\left(\sup_{x\in\mathcal{C}}|m(x) - \hat{m}_{h_n}(x)| > \frac{\xi}{3c_1}\right) \\ & + \mathbb{P}\left(\|\Delta_n - \Delta\|_{\mathcal{H}'} > \frac{\xi}{3}\right).\end{aligned} \tag{3.19}$$

Note that $\mathbb{E}(U_i) = 0$, and $\|U_i\|_{\mathcal{H}'} \leq |\epsilon_i|\,\|X_i\|$. As a by–product, $\forall l \geq 2$, it holds $\|U_i\|_{\mathcal{H}'}^l \leq |\epsilon_i|^l \|X_i\|^l$. From this inequality, **(C.2.6)**, and **(C.2.7)**, one gets $\mathbb{E}(\|U_i\|_{\mathcal{H}'}^l) \leq l! c_2 c_1^l$. Then

$$\mathbb{E}(\|U_i\|_{\mathcal{H}'}^l) \leq \frac{l!}{2} b_i^2 c^{l-2}, \quad \forall l \geq 2$$

where $b_i = \sqrt{2c_2c_1^2}$ and $c = c_1$. Therefore, the hypotheses of the Yurinskii exponential inequality (Yurinskii, 1976) hold, and one gets

$$\mathbb{P}\left(\left\|\sum_{i=1}^n U_i\right\|_{\mathcal{H}'} > xB_n\right) \leq 2\exp\left(-\frac{x^2}{2(1 + 1.62xc/B_n)}\right),$$

where $B_n = (\sum_{i=1}^n b_i^2)^{1/2} = \sqrt{n}\,c_3$ with $c_3 = \sqrt{2c_2}c_1$. Hence,

$$\mathbb{P}\left(\frac{1}{n}\|\sum_{i=1}^n U_i\|_{\mathcal{H}'} > \frac{\xi}{3}\right) = \mathbb{P}\left(\left\|\sum_{i=1}^n U_i\right\|_{\mathcal{H}'} > \frac{\sqrt{n}\xi}{3c_3}B_n\right) \leq 2\exp\left(-\frac{\xi^2 n}{6\,c_3(3c_3 + c_4\xi)}\right), \tag{3.20}$$

where c_4 is a positive finite constant. Therefore, using (3.20) and Lemma 2.5.1 (see Chapter 2, page 47) in (3.19), one gets

$$\mathbb{P}(\|\Delta_n^{h_n} - \Delta\|_{\mathcal{H}'} > \xi) \leq 4\exp\left(-\frac{\xi^2 n}{2\,c_5(c_5 + c_6\xi)}\right) + \mathbb{P}\left(\sup_{x\in\mathcal{C}} |m(x) - \hat{m}_{h_n}(x)| > \frac{\xi}{c_7}\right),$$

with c_5, c_6 and c_7 positive finite constants. ■

3.8.8 Proof of Theorem 3.3.3

Consider Lemma 3.8.4 (see page 79), with $\gamma_j = \hat{\lambda}_j^{-1}$ and $\omega_j = \hat{v}_j$. In this case,

$$R_{k_n}^{(\gamma,\omega)} = \theta - \sum_{j=1}^{k_n} \hat{\lambda}_j^{-1} \langle \Gamma_n \hat{v}_j, \theta\rangle \hat{v}_j = \hat{R}_{k_n}$$

with $\hat{R}_{k_n}$ defined in (2.7) (see Chapter 2, page 39). One then obtains for the conditional prediction error,

$$\begin{aligned}\mathbb{E}_{\mathcal{X}^{n+1}}(Y_{n+1} - \langle \hat{\theta}_{k_n}^{h_n}, X_{n+1}\rangle)^2 = \sigma^2 &+ \sum_{j_1=1}^{k_n}\sum_{j_2=1}^{k_n} \frac{U_n^{h_n}(\hat{v}_{j_1})U_n^{h_n}(\hat{v}_{j_2})}{\hat{\lambda}_{j_1}\hat{\lambda}_{j_2}} \langle X_{n+1}, \hat{v}_{j_1}\rangle\langle X_{n+1}, \hat{v}_{j_2}\rangle \\ &+ \sigma^2 \sum_{j_1=1}^{k_n}\sum_{j_2=1}^{k_n} \frac{W_n^{h_n}(\hat{v}_{j_1}, \hat{v}_{j_2})}{\hat{\lambda}_{j_1}\hat{\lambda}_{j_2}} \langle X_{n+1}, \hat{v}_{j_1}\rangle\langle X_{n+1}, \hat{v}_{j_2}\rangle \\ &- 2\langle X_{n+1}, \hat{R}_{k_n}\rangle \left\langle X_{n+1}, \sum_{j=1}^{k_n} \frac{U_n^{h_n}(\hat{v}_j)}{\hat{\lambda}_j}\hat{v}_j \right\rangle + \langle X_{n+1}, \hat{R}_{k_n}\rangle^2,\end{aligned}$$

and for the conditional estimation error,

$$\mathbb{E}_{\mathcal{X}^n}(\|\theta - \hat{\theta}_{k_n}^{h_n}\|^2) = \sum_{j=1}^{k_n} \frac{(U_n^{h_n}(\hat{v}_j))^2}{\hat{\lambda}_j^2} + \sigma^2 \sum_{j=1}^{k_n} \frac{W_n^{h_n}(\hat{v}_j, \hat{v}_j)}{\hat{\lambda}_j^2} + \|\hat{R}_{k_n}\|^2,$$

where

$$U_n^{h_n}(x) = \frac{1}{n}\sum_{i=1}^{n} \langle X_i, x\rangle \left(\sum_{l=1}^{n} w_{l,h_n}(X_i)(m(X_l) - m(X_i))\right), \quad \forall x \in \mathcal{H},$$

and

$$W_n^{h_n}(x,y) = \frac{1}{n^2}\sum_{i_1=1}^{n}\sum_{i_2=1}^{n} \langle X_{i_1}, x\rangle\langle X_{i_2}, y\rangle \left(\sum_{l=1}^{n} w_{l,h_n}(X_{i_1})w_{l,h_n}(X_{i_2})\right), \quad \forall x, y \in \mathcal{H},$$

being $w_{l,h_n}(x) = K(h_n^{-1}d(x, X_l))/\sum_{l'=1}^{n} K(h_n^{-1}d(x, X_{l'}))$ the weights of the kernel estimator $\hat{m}_{h_n}$. Hence the proof is finished using Theorem 2.3.14 (see Chapter 2, page 39).

3.8.9 Formulation and proof of Lemma 3.8.4

Lemma 3.8.4. *Considering the regression model* (2.1) *(see Chapter 2, page 32), let* $\hat{\theta}$ *be an estimator for* θ *such that*

$$\hat{\theta} = \sum_{j=1}^{k_n} \gamma_j \Delta_n^{h_n} \omega_j \omega_j,$$

where $\{(\gamma_j, \omega_j)\}_j \subset \mathbb{R} \times \mathcal{H}$ *only depend on* $\mathcal{X}^n = \{X_1, \ldots, X_n\}$*. Then, it holds that*

$$\begin{aligned}\mathbb{E}_{\mathcal{X}^{n+1}}(Y_{n+1} - \langle \hat{\theta}, X_{n+1}\rangle)^2 &= \sigma^2 + \sum_{j_1=1}^{k_n}\sum_{j_2=1}^{k_n} \gamma_{j_1}\gamma_{j_2} U_n^{h_n}(\omega_{j_1}) U_n^{h_n}(\omega_{j_2}) \langle X_{n+1}, \omega_{j_1}\rangle \langle X_{n+1}, \omega_{j_2}\rangle \\ &+ \sigma^2 \sum_{j_1=1}^{k_n}\sum_{j_2=1}^{k_n} \gamma_{j_1}\gamma_{j_2} W_n^{h_n}(\omega_{j_1}, \omega_{j_2}) \langle X_{n+1}, \omega_{j_1}\rangle \langle X_{n+1}, \omega_{j_2}\rangle \\ &- 2\langle X_{n+1}, R_{k_n}^{(\gamma,\omega)}\rangle \left\langle X_{n+1}, \sum_{j=1}^{k_n} \gamma_j U_n^{h_n}(\omega_j)\omega_j \right\rangle + \langle X_{n+1}, R_{k_n}^{(\gamma,\omega)}\rangle^2,\end{aligned}$$

and

$$\begin{aligned}\mathbb{E}_{\mathcal{X}^n}(\|\theta - \hat{\theta}\|^2) &= \sum_{j_1=1}^{k_n}\sum_{j_2=1}^{k_n} \gamma_{j_1}\gamma_{j_2} U_n^{h_n}(\omega_{j_1}) U_n^{h_n}(\omega_{j_2}) \langle \omega_{j_1}, \omega_{j_2}\rangle \\ &+ \sigma^2 \sum_{j_1=1}^{k_n}\sum_{j_2=1}^{k_n} \gamma_{j_1}\gamma_{j_2} W_n^{h_n}(\omega_{j_1}, \omega_{j_2}) \langle \omega_{j_1}, \omega_{j_2}\rangle - 2\left\langle R_{k_n}^{(\gamma,\omega)}, \sum_{j=1}^{k_n} \gamma_j U_n^{h_n}(\omega_j)\omega_j \right\rangle + \|R_{k_n}^{(\gamma,\omega)}\|^2,\end{aligned}$$

where

$$U_n^{h_n}(x) = \frac{1}{n}\sum_{i=1}^{n} \langle X_i, x\rangle \left(\sum_{l=1}^{n} w_{l,h_n}(X_i)(m(X_l) - m(X_i)) \right), \quad \forall x \in \mathcal{H},$$

and

$$W_n^{h_n}(x, y) = \frac{1}{n^2}\sum_{i_1=1}^{n}\sum_{i_2=1}^{n} \langle X_{i_1}, x\rangle \langle X_{i_2}, y\rangle \left(\sum_{l=1}^{n} w_{l,h_n}(X_{i_1}) w_{l,h_n}(X_{i_2}) \right), \quad \forall x, y \in \mathcal{H},$$

being $w_{l,h_n}(x) = K(h_n^{-1} d(x, X_l))/\sum_{l'=1}^{n} K(h_n^{-1} d(x, X_{l'}))$ *the weights of the kernel estimator* $\hat{m}_{h_n}$*, and* $R_{k_n}^{(\gamma,\omega)} = \theta - \sum_{j=1}^{k_n} \gamma_j \langle \Gamma_n \omega_j, \theta\rangle \omega_j$*.*

Proof. This proof is similar to that of Lemma 2.5.2 (see Chapter 2, page 49).

Note that $\Delta_n^{h_n} x = \langle \Gamma_n x, \theta\rangle + \Delta_n^{\epsilon} x + Z_n^{h_n} x$ for all $x \in \mathcal{H}$, where $\Delta_n^{\epsilon} = n^{-1}\sum_{i=1}^{n} X_i \otimes_{\mathcal{H}'} \epsilon_i$ and

$$Z_n^{h_n} = \frac{1}{n}\sum_{i=1}^{n} X_i \otimes_{\mathcal{H}'} \left(\sum_{l=1}^{n} w_{l,h_n}(X_i)(Y_l - Y_i) \right),$$

being w_{l,h_n} the weights of the kernel estimator $\hat{m}_{h_n}$, i.e.,

$$w_{l,h_n}(x) = \frac{K(h_n^{-1} d(x, X_l))}{\sum_{l'=1}^{n} K(h_n^{-1} d(x, X_{l'}))}.$$

Hence, the difference $\theta - \hat{\theta}$ can be expressed as

$$\theta - \hat{\theta} = R_{k_n}^{(\gamma,\omega)} - \sum_{j=1}^{k_n} \gamma_j \Delta_n^{\epsilon} \omega_j \omega_j - \sum_{j=1}^{k_n} \gamma_j Z_n^{h_n} \omega_j \omega_j, \quad \text{with} \quad R_{k_n}^{(\gamma,\omega)} = \theta - \sum_{j=1}^{k_n} \gamma_j \langle \Gamma_n \omega_j, \theta\rangle \omega_j. \tag{3.21}$$

Then the regression model (2.1) (see Chapter 2, page 32), (3.21), and conditions on ϵ imply that

$$\begin{aligned}
\mathbb{E}_{\mathcal{X}^{n+1}}(Y_{n+1} - \langle\hat{\theta}, X_{n+1}\rangle)^2 &= \mathbb{E}_{\mathcal{X}^{n+1}}\left(\epsilon_{n+1} + \left\langle X_{n+1}, R_{k_n}^{(\gamma,\omega)} - \sum_{j=1}^{k_n}\gamma_j\Delta_n^{\epsilon}\omega_j\omega_j - \sum_{j=1}^{k_n}\gamma_j Z_n^{h_n}\omega_j\omega_j\right\rangle\right)^2 \\
&= \sigma^2 + \sum_{j_1=1}^{k_n}\sum_{j_2=1}^{k_n}\gamma_{j_1}\gamma_{j_2}\mathbb{E}_{\mathcal{X}^n}(\Delta_n^{\epsilon}\omega_{j_1}\Delta_n^{\epsilon}\omega_{j_2})\langle X_{n+1},\omega_{j_1}\rangle\langle X_{n+1},\omega_{j_2}\rangle \\
&\quad + \sum_{j_1=1}^{k_n}\sum_{j_2=1}^{k_n}\gamma_{j_1}\gamma_{j_2}\mathbb{E}_{\mathcal{X}^n}(Z_n^{h_n}\omega_{j_1}Z_n^{h_n}\omega_{j_2})\langle X_{n+1},\omega_{j_1}\rangle\langle X_{n+1},\omega_{j_2}\rangle \\
&\quad + 2\sum_{j_1=1}^{k_n}\sum_{j_2=1}^{k_n}\gamma_{j_1}\gamma_{j_2}\mathbb{E}_{\mathcal{X}^n}(\Delta_n^{\epsilon}\omega_{j_1}Z_n^{h_n}\omega_{j_2})\langle X_{n+1},\omega_{j_1}\rangle\langle X_{n+1},\omega_{j_2}\rangle \\
&\quad - 2\langle X_{n+1}, R_{k_n}^{(\gamma,\omega)}\rangle\left\langle X_{n+1}, \sum_{j=1}^{k_n}\gamma_j\mathbb{E}_{\mathcal{X}^n}(\Delta_n^{\epsilon}\omega_j)\omega_j\right\rangle \\
&\quad - 2\langle X_{n+1}, R_{k_n}^{(\gamma,\omega)}\rangle\left\langle X_{n+1}, \sum_{j=1}^{k_n}\gamma_j\mathbb{E}_{\mathcal{X}^n}(Z_n^{h_n}\omega_j)\omega_j\right\rangle + \langle X_{n+1}, R_{k_n}^{(\gamma,\omega)}\rangle^2.
\end{aligned} \tag{3.22}$$

It can be shown that

$$\mathbb{E}_{\mathcal{X}^n}(\Delta_n^{\epsilon}\omega_j) = \frac{1}{n}\sum_{i=1}^{n}\langle X_i,\omega_j\rangle\mathbb{E}_{\mathcal{X}^n}(\epsilon_i) = 0, \tag{3.23}$$

$$\mathbb{E}_{\mathcal{X}^n}(\Delta_n^{\epsilon}\omega_{j_1}\Delta_n^{\epsilon}\omega_{j_2}) = \frac{1}{n^2}\sum_{i_1=1}^{n}\sum_{i_2=1}^{n}\langle X_{i_1},\omega_{j_1}\rangle\langle X_{i_2},\omega_{j_2}\rangle\mathbb{E}_{\mathcal{X}^n}(\epsilon_{i_1}\epsilon_{i_2}) = \frac{\sigma^2}{n}\langle\Gamma_n\omega_{j_1},\omega_{j_2}\rangle, \tag{3.24}$$

and

$$\begin{aligned}
\mathbb{E}_{\mathcal{X}^n}(\Delta_n^{\epsilon}\omega_{j_1}Z_n^{h_n}\omega_{j_2}) &= \frac{1}{n^2}\sum_{i_1=1}^{n}\sum_{i_2=1}^{n}\langle X_{i_1},\omega_{j_1}\rangle\langle X_{i_2},\omega_{j_2}\rangle\left(\sum_{l=1}^{n}w_{l,h_n}(X_{i_2})\mathbb{E}_{\mathcal{X}^n}(\epsilon_{i_1}(Y_l - Y_{i_2}))\right) \\
&= \frac{1}{n^2}\sum_{i_1=1}^{n}\sum_{i_2=1}^{n}\langle X_{i_1},\omega_{j_1}\rangle\langle X_{i_2},\omega_{j_2}\rangle\left(\sum_{l=1}^{n}w_{l,h_n}(X_{i_2})\mathbb{E}_{\mathcal{X}^n}(\epsilon_{i_1}(\epsilon_l - \epsilon_{i_2}))\right) \\
&= \frac{\sigma^2}{n^2}\sum_{i_1=1}^{n}\sum_{i_2=1}^{n}\langle X_{i_1},\omega_{j_1}\rangle\langle X_{i_2},\omega_{j_2}\rangle w_{i_1,h_n}(X_{i_2}) - \frac{\sigma^2}{n}\langle\Gamma_n\omega_{j_1},\omega_{j_2}\rangle.
\end{aligned} \tag{3.25}$$

Furthermore, if $U_n^{h_n}$ and $W_n^{h_n}$ are defined as

$$U_n^{h_n}(x) = \frac{1}{n}\sum_{i=1}^{n}\langle X_i, x\rangle\left(\sum_{l=1}^{n}w_{l,h_n}(X_i)(m(X_l) - m(X_i))\right), \quad \forall x\in\mathcal{H},$$

and

$$W_n^{h_n}(x,y) = \frac{1}{n^2}\sum_{i_1=1}^{n}\sum_{i_2=1}^{n}\langle X_{i_1},x\rangle\langle X_{i_2},y\rangle\left(\sum_{l=1}^{n}w_{l,h_n}(X_{i_1})w_{l,h_n}(X_{i_2})\right), \quad \forall x,y\in\mathcal{H},$$

respectively, then one gets

$$\mathbb{E}_{\mathcal{X}^n}(Z_n^{h_n}\omega_j) = \frac{1}{n}\sum_{i=1}^{n}\langle X_i,\omega_j\rangle\left(\sum_{l=1}^{n}w_{l,h_n}(X_i)\mathbb{E}_{\mathcal{X}^n}(Y_l - Y_i)\right) = U_n^{h_n}(\omega_j), \tag{3.26}$$

and

$$
\begin{aligned}
&\mathbb{E}_{\mathcal{X}^n}(Z_n^{h_n}\omega_{j_1}Z_n^{h_n}\omega_{j_2})\\
&=\frac{1}{n^2}\sum_{i_1=1}^{n}\sum_{i_2=1}^{n}\langle X_{i_1},\omega_{j_1}\rangle\langle X_{i_2},\omega_{j_2}\rangle\left(\sum_{l_1=1}^{n}\sum_{l_2=1}^{n}w_{l_1,h_n}(X_{i_1})w_{l_2,h_n}(X_{i_2})\mathbb{E}_{\mathcal{X}^n}((Y_{l_1}-Y_{i_1})(Y_{l_2}-Y_{i_2}))\right)\\
&=U_n^{h_n}(\omega_{j_1})U_n^{h_n}(\omega_{j_2})\\
&\quad+\frac{1}{n^2}\sum_{i_1=1}^{n}\sum_{i_2=1}^{n}\langle X_{i_1},\omega_{j_1}\rangle\langle X_{i_2},\omega_{j_2}\rangle\left(\sum_{l_1=1}^{n}\sum_{l_2=1}^{n}w_{l_1,h_n}(X_{i_1})w_{l_2,h_n}(X_{i_2})\mathbb{E}_{\mathcal{X}^n}((\epsilon_{l_1}-\epsilon_{i_1})(\epsilon_{l_2}-\epsilon_{i_2}))\right)\\
&=U_n^{h_n}(\omega_{j_1})U_n^{h_n}(\omega_{j_2})+\frac{1}{n^2}\sum_{i_1=1}^{n}\sum_{i_2=1}^{n}\langle X_{i_1},\omega_{j_1}\rangle\langle X_{i_2},\omega_{j_2}\rangle\left(\sum_{l=1}^{n}w_{l,h_n}(X_{i_1})w_{l,h_n}(X_{i_2})\mathbb{E}_{\mathcal{X}^n}(\epsilon_l^2)\right)\\
&\quad-\frac{2}{n^2}\sum_{i_1=1}^{n}\sum_{i_2=1}^{n}\langle X_{i_1},\omega_{j_1}\rangle\langle X_{i_2},\omega_{j_2}\rangle\left(\sum_{l=1}^{n}w_{i_2,h_n}(X_{i_1})w_{l,h_n}(X_{i_2})\mathbb{E}_{\mathcal{X}^n}(\epsilon_{i_2}^2)\right)\\
&\quad+\frac{1}{n^2}\sum_{i=1}^{n}\langle X_i,\omega_{j_1}\rangle\langle X_i,\omega_{j_2}\rangle\left(\sum_{l_1=1}^{n}\sum_{l_2=1}^{n}w_{l_1,h_n}(X_i)w_{l_2,h_n}(X_i)\mathbb{E}_{\mathcal{X}^n}(\epsilon_i^2)\right)\\
&=U_n^{h_n}(\omega_{j_1})U_n^{h_n}(\omega_{j_2})+\sigma^2W_n^{h_n}(\omega_{j_1},\omega_{j_2})-2\frac{\sigma^2}{n^2}\sum_{i_1=1}^{n}\sum_{i_2=1}^{n}\langle X_{i_1},\omega_{j_1}\rangle\langle X_{i_2},\omega_{j_2}\rangle w_{i_2,h_n}(X_{i_1})\\
&\quad+\frac{\sigma^2}{n}\langle\Gamma_n\omega_{j_1},\omega_{j_2}\rangle.
\end{aligned}
\tag{3.27}
$$

Consequently, using (3.23), (3.24), (3.25), (3.26), and (3.27) in (3.22) one has

$$
\begin{aligned}
\mathbb{E}_{\mathcal{X}^{n+1}}(Y_{n+1}-\langle\hat{\theta},X_{n+1}\rangle)^2=\sigma^2&+\sum_{j_1=1}^{k_n}\sum_{j_2=1}^{k_n}\gamma_{j_1}\gamma_{j_2}U_n^{h_n}(\omega_{j_1})U_n^{h_n}(\omega_{j_2})\langle X_{n+1},\omega_{j_1}\rangle\langle X_{n+1},\omega_{j_2}\rangle\\
&+\sigma^2\sum_{j_1=1}^{k_n}\sum_{j_2=1}^{k_n}\gamma_{j_1}\gamma_{j_2}W_n^{h_n}(\omega_{j_1},\omega_{j_2})\langle X_{n+1},\omega_{j_1}\rangle\langle X_{n+1},\omega_{j_2}\rangle\\
&-2\langle X_{n+1},R_{k_n}^{(\gamma,\omega)}\rangle\left\langle X_{n+1},\sum_{j=1}^{k_n}\gamma_jU_n^{h_n}(\omega_j)\omega_j\right\rangle+\langle X_{n+1},R_{k_n}^{(\gamma,\omega)}\rangle^2.
\end{aligned}
$$

For the conditional estimation error, (3.21) gives

$$
\begin{aligned}
&\mathbb{E}_{\mathcal{X}^n}(\|\theta-\hat{\theta}\|^2)\\
&=\mathbb{E}_{\mathcal{X}^n}\left(\left\langle R_{k_n}^{(\gamma,\omega)}-\sum_{j=1}^{k_n}\gamma_j\Delta_n^{\epsilon}\omega_j\omega_j-\sum_{j=1}^{k_n}\gamma_jZ_n^{h_n}\omega_j\omega_j,R_{k_n}^{(\gamma,\omega)}-\sum_{j=1}^{k_n}\gamma_j\Delta_n^{\epsilon}\omega_j\omega_j-\sum_{j=1}^{k_n}\gamma_jZ_n^{h_n}\omega_j\omega_j\right\rangle\right)\\
&=\sum_{j_1=1}^{k_n}\sum_{j_2=1}^{k_n}\gamma_{j_1}\gamma_{j_2}\mathbb{E}_{\mathcal{X}^n}(\Delta_n^{\epsilon}\omega_{j_1}\Delta_n^{\epsilon}\omega_{j_2})\langle\omega_{j_1},\omega_{j_2}\rangle+\sum_{j_1=1}^{k_n}\sum_{j_2=1}^{k_n}\gamma_{j_1}\gamma_{j_2}\mathbb{E}_{\mathcal{X}^n}(Z_n^{h_n}\omega_{j_1}Z_n^{h_n}\omega_{j_2})\langle\omega_{j_1},\omega_{j_2}\rangle\\
&\quad+2\sum_{j_1=1}^{k_n}\sum_{j_2=1}^{k_n}\gamma_{j_1}\gamma_{j_2}\mathbb{E}_{\mathcal{X}^n}(\Delta_n^{\epsilon}\omega_{j_1}Z_n^{h_n}\omega_{j_2})\langle\omega_{j_1},\omega_{j_2}\rangle-2\left\langle R_{k_n}^{(\gamma,\omega)},\sum_{j=1}^{k_n}\gamma_j\mathbb{E}_{\mathcal{X}^n}(\Delta_n^{\epsilon}\omega_j)\omega_j\right\rangle\\
&\quad-2\left\langle R_{k_n}^{(\gamma,\omega)},\sum_{j=1}^{k_n}\gamma_j\mathbb{E}_{\mathcal{X}^n}(Z_n^{h_n}\omega_j)\omega_j\right\rangle+\|R_{k_n}^{(\gamma,\omega)}\|^2.
\end{aligned}
\tag{3.28}
$$

As a result, using (3.23), (3.24), (3.25), (3.26), and (3.27) in (3.28), one has

$$\mathbb{E}_{\mathcal{X}^n}(\|\theta-\hat{\theta}\|^2)=\sum_{j_1=1}^{k_n}\sum_{j_2=1}^{k_n}\gamma_{j_1}\gamma_{j_2}U_n^{h_n}(\omega_{j_1})U_n^{h_n}(\omega_{j_2})\langle\omega_{j_1},\omega_{j_2}\rangle$$
$$+\sigma^2\sum_{j_1=1}^{k_n}\sum_{j_2=1}^{k_n}\gamma_{j_1}\gamma_{j_2}W_n^{h_n}(\omega_{j_1},\omega_{j_2})\langle\omega_{j_1},\omega_{j_2}\rangle-2\left\langle R_{k_n}^{(\gamma,\omega)},\sum_{j=1}^{k_n}\gamma_j U_n^{h_n}(\omega_j)\omega_j\right\rangle+\|R_{k_n}^{(\gamma,\omega)}\|^2.$$

■

3.8.10 Proof of Corollary 3.3.4

Note that if $w_{l,h_n}(X_i)=1$ when $i=l$, and $w_{l,h_n}(X_i)=0$ when $i\neq l$, then $U_n^{h_n}(x)=0$ for all $x\in\mathcal{H}$, and $W_n^{h_n}(x,y)=n^{-1}\langle\Gamma_n x,y\rangle$ for all $x,y\in\mathcal{H}$. Thus, the proof is finished by Theorem 3.3.3, given that

$$\mathbb{E}_{\mathcal{X}^{n+1}}(Y_{n+1}-\langle\hat{\theta}_{k_n}^{h_n},X_{n+1}\rangle)^2=\sigma^2+\frac{\sigma^2}{n}\sum_{j=1}^{k_n}\frac{\langle X_{n+1},\hat{v}_j\rangle^2}{\hat{\lambda}_j}+\langle X_{n+1},\hat{R}_{k_n}\rangle^2,$$

$$\mathbb{E}_{\mathcal{X}^n}(\|\theta-\hat{\theta}_{k_n}^{h_n}\|^2)=\frac{\sigma^2}{n}\sum_{j=1}^{k_n}\frac{1}{\hat{\lambda}_j}+\|\hat{R}_{k_n}\|^2.$$

3.8.11 Proof of Corollary 3.3.5

Note that if $w_{l,h_n}(X_i)=n^{-1}$ for all $i=1,\ldots,n$, then

$$U_n^{h_n}(x)=\frac{1}{n}\sum_{i=1}^{n}\langle X_i,x\rangle\left(\langle\theta,\overline{X}\rangle-\langle\theta,X_i\rangle\right)=\langle\theta,\overline{X}\rangle\langle\overline{X},x\rangle-\langle\Gamma_n x,\theta\rangle,\quad\forall x\in\mathcal{H},$$

$$W_n^{h_n}(x,y)=\frac{1}{n^2}\sum_{i_1=1}^{n}\sum_{i_2=1}^{n}\langle X_{i_1},x\rangle\langle X_{i_2},y\rangle\left(\sum_{l=1}^{n}\frac{1}{n^2}\right)=\frac{1}{n}\langle\overline{X},x\rangle\langle\overline{X},y\rangle,\quad\forall x,y\in\mathcal{H},$$

where $\overline{X}=n^{-1}\sum_{i=1}^{n}X_i$. Consequently, Theorem 3.3.3 gives

$$\mathbb{E}_{\mathcal{X}^{n+1}}(Y_{n+1}-\langle\hat{\theta}_{k_n}^{h_n},X_{n+1}\rangle)^2-\mathbb{E}_{\mathcal{X}^{n+1}}(Y_{n+1}-\langle\hat{\theta}_{k_n},X_{n+1}\rangle)^2$$
$$=\left(\sum_{j=1}^{k_n}\frac{(\langle\theta,\overline{X}\rangle\langle\overline{X},\hat{v}_j\rangle-\hat{\lambda}_j\langle\hat{v}_j,\theta\rangle)}{\hat{\lambda}_j}\langle X_{n+1},\hat{v}_j\rangle\right)^2+\frac{\sigma^2}{n}\sum_{j=1}^{k_n}\frac{\langle\overline{X},\hat{v}_j\rangle}{\hat{\lambda}_j}\langle X_{n+1},\hat{v}_j\rangle-\frac{\sigma^2}{n}\sum_{j=1}^{k_n}\frac{\langle X_{n+1},\hat{v}_j\rangle^2}{\hat{\lambda}_j}$$
$$-2\langle X_{n+1},\hat{R}_{k_n}\rangle\left\langle X_{n+1},\sum_{j=1}^{k_n}\frac{\langle\theta,\overline{X}\rangle\langle\overline{X},\hat{v}_j\rangle-\hat{\lambda}_j\langle\hat{v}_j,\theta\rangle}{\hat{\lambda}_j}\hat{v}_j\right\rangle$$
$$=\left(\langle\theta,\overline{X}\rangle\langle X_{n+1},\hat{M}_{k_n}\rangle-\langle X_{n+1},\theta-\hat{R}_{k_n}\rangle\right)^2+\frac{\sigma^2}{n}\left(\langle X_{n+1},\hat{M}_{k_n}\rangle^2-\sum_{j=1}^{k_n}\frac{\langle X_{n+1},\hat{v}_j\rangle^2}{\hat{\lambda}_j}\right)$$
$$-2\langle\theta,\overline{X}\rangle\langle X_{n+1},\hat{R}_{k_n}\rangle\langle X_{n+1},\hat{M}_{k_n}\rangle+2\langle X_{n+1},\hat{R}_{k_n}\rangle\langle X_{n+1},\theta-\hat{R}_{k_n}\rangle$$
$$=\frac{\sigma^2}{n}\left(\langle X_{n+1},\hat{M}_{k_n}\rangle^2-\sum_{j=1}^{k_n}\frac{\langle X_{n+1},\hat{v}_j\rangle^2}{\hat{\lambda}_j}\right)+\left(\langle\theta,\overline{X}\rangle\langle X_{n+1},\hat{M}_{k_n}\rangle-\langle X_{n+1},\theta\rangle\right)^2$$
$$+2\left(\langle\theta,\overline{X}\rangle\langle X_{n+1},\hat{M}_{k_n}\rangle-\langle X_{n+1},\theta\rangle\right)\langle X_{n+1},\hat{R}_{k_n}\rangle+\langle X_{n+1},\hat{R}_{k_n}\rangle^2$$
$$-2\langle\theta,\overline{X}\rangle\langle X_{n+1},\hat{R}_{k_n}\rangle\langle X_{n+1},\hat{M}_{k_n}\rangle+2\langle X_{n+1},\hat{R}_{k_n}\rangle\langle X_{n+1},\theta\rangle-2\langle X_{n+1},\hat{R}_{k_n}\rangle^2$$
$$=\frac{\sigma^2}{n}\left(\langle X_{n+1},\hat{M}_{k_n}\rangle^2-\sum_{j=1}^{k_n}\frac{\langle X_{n+1},\hat{v}_j\rangle^2}{\hat{\lambda}_j}\right)+\left(\langle\theta,\overline{X}\rangle\langle X_{n+1},\hat{M}_{k_n}\rangle-\langle X_{n+1},\theta\rangle\right)^2-\langle X_{n+1},\hat{R}_{k_n}\rangle^2,$$

and

$$\mathbb{E}_{\mathcal{X}^n}(\|\theta-\hat{\theta}_{k_n}^{h_n}\|^2)-\mathbb{E}_{\mathcal{X}^n}(\|\theta-\hat{\theta}_{k_n}\|^2)=\sum_{j=1}^{k_n}\frac{(\langle\theta,\overline{X}\rangle\langle\overline{X},\hat{v}_j\rangle-\hat{\lambda}_j\langle\hat{v}_j,\theta\rangle)^2}{\hat{\lambda}_j^2}+\frac{\sigma^2}{n}\sum_{j=1}^{k_n}\frac{\langle\overline{X},\hat{v}_j\rangle^2}{\hat{\lambda}_j^2}-\frac{\sigma^2}{n}\sum_{j=1}^{k_n}\frac{1}{\hat{\lambda}_j}$$

$$=\frac{\sigma^2}{n}\left(\sum_{j=1}^{k_n}\frac{\langle\overline{X},\hat{v}_j\rangle^2}{\hat{\lambda}_j^2}-\sum_{j=1}^{k_n}\frac{1}{\hat{\lambda}_j}\right)+\langle\theta,\overline{X}\rangle^2\sum_{j=1}^{k_n}\frac{\langle\overline{X},\hat{v}_j\rangle^2}{\hat{\lambda}_j^2}-2\langle\theta,\overline{X}\rangle\left\langle\sum_{j=1}^{k_n}\frac{\langle\overline{X},\hat{v}_j\rangle}{\hat{\lambda}_j}\hat{v}_j,\theta\right\rangle+\sum_{j=1}^{k_n}\langle\theta,\hat{v}_j\rangle^2$$

$$=\frac{\sigma^2}{n}\left(\|\hat{M}_{k_n}\|^2-\sum_{j=1}^{k_n}\frac{1}{\hat{\lambda}_j}\right)+\langle\theta,\overline{X}\rangle^2\|\hat{M}_{k_n}\|^2-2\langle\theta,\overline{X}\rangle\langle\hat{M}_{k_n},\theta\rangle+\|\theta\|^2-\|\hat{R}_{k_n}\|^2$$

$$=\frac{\sigma^2}{n}\left(\|\hat{M}_{k_n}\|^2-\sum_{j=1}^{k_n}\frac{1}{\hat{\lambda}_j}\right)+\Big(\langle\theta,\overline{X}\rangle\|\hat{M}_{k_n}\|-\|\theta\|\Big)^2+2\langle\theta,\overline{X}\rangle\Big(\|\hat{M}_{k_n}\|\,\|\theta\|-\langle\hat{M}_{k_n},\theta\rangle\Big)-\|\hat{R}_{k_n}\|^2,$$

with $\hat{M}_{k_n}=\sum_{j=1}^{k_n}\hat{\lambda}_j^{-1}\langle\overline{X},\hat{v}_j\rangle\hat{v}_j$.

3.8.12 Proof of Theorem 3.4.1

Consider Lemma 2.5.2 (see Chapter 2, page 49), with $\gamma_j=(\hat{\lambda}_j^{\alpha_n,1})^{-1}$ and $w_j=\hat{v}_j^{\alpha_n,1}$. Therefore, the conditional prediction error is given by

$$\begin{aligned}\mathbb{E}_{\mathcal{X}^{n+1}}(Y_{n+1}-\langle\hat{\theta}_{k_n}^{PS,\,\alpha_n},X_{n+1}\rangle)^2&=\sigma^2\\&+\frac{\sigma^2}{n}\sum_{j_1=1}^{k_n}\sum_{j_2=1}^{k_n}(\hat{\lambda}_{j_1}^{\alpha_n,1}\hat{\lambda}_{j_2}^{\alpha_n,1})^{-1}\langle\Gamma_n\hat{v}_{j_1}^{\alpha_n,1},\hat{v}_{j_2}^{\alpha_n,1}\rangle\langle X_{n+1},\hat{v}_{j_1}^{\alpha_n,1}\rangle\langle X_{n+1},\hat{v}_{j_2}^{\alpha_n,1}\rangle+\langle X_{n+1},R_{k_n}^{(\gamma,w)}\rangle^2,\end{aligned}\tag{3.29}$$

with $R_{k_n}^{(\gamma,w)}=\theta-\sum_{j=1}^{k_n}(\hat{\lambda}_j^{\alpha_n,1})^{-1}\langle\Gamma_n\hat{v}_j^{\alpha_n,1},\theta\rangle\hat{v}_j^{\alpha_n,1}$. Recall that $\{\hat{\lambda}_j^{\alpha_n,1},\hat{v}_j^{\alpha_n,1}\}_j$ satisfy

$$\Gamma_n\hat{v}_j^{\alpha_n,1}=\hat{\lambda}_j^{\alpha_n,1}\hat{v}_j^{\alpha_n,1}+\alpha_nQ\hat{v}_j^{\alpha_n,1}\quad\text{and}\quad\langle\hat{v}_{j_1}^{\alpha_n,1},\hat{v}_{j_2}^{\alpha_n,1}\rangle=\delta_{j_1j_2},\tag{3.30}$$

where $\delta_{j_1j_2}=1$ if $j_1=j_2$, and 0 otherwise. Hence,

$$R_{k_n}^{(\gamma,w)}=\theta-\sum_{j=1}^{k_n}\langle\theta,\hat{v}_j^{\alpha_n,1}\rangle\hat{v}_j^{\alpha_n,1}-\alpha_n\sum_{j=1}^{k_n}\frac{\langle\theta,Q\hat{v}_j^{\alpha_n,1}\rangle}{\hat{\lambda}_j^{\alpha_n,1}}\hat{v}_j^{\alpha_n,1}.\tag{3.31}$$

Using (3.30) and (3.31) in (3.29), one gets

$$\begin{aligned}\mathbb{E}_{\mathcal{X}^{n+1}}(Y_{n+1}-\langle\hat{\theta}_{k_n}^{PS,\,\alpha_n},X_{n+1}\rangle)^2&=\sigma^2\\&+\frac{\sigma^2}{n}\sum_{j=1}^{k_n}\frac{\langle X_{n+1},\hat{v}_j^{\alpha_n,1}\rangle^2}{\hat{\lambda}_j^{\alpha_n,1}}+\alpha_n\frac{\sigma^2}{n}\sum_{j_1=1}^{k_n}\sum_{j_2=1}^{k_n}\frac{\rho_{j_1,j_2}^{\alpha_n,1}}{\hat{\lambda}_{j_1}^{\alpha_n,1}\hat{\lambda}_{j_2}^{\alpha_n,1}}\langle X_{n+1},\hat{v}_{j_1}^{\alpha_n,1}\rangle\langle X_{n+1},\hat{v}_{j_2}^{\alpha_n,1}\rangle\\&+\left\langle X_{n+1},\theta-\sum_{j=1}^{k_n}\langle\theta,\hat{v}_j^{\alpha_n,1}\rangle\hat{v}_j^{\alpha_n,1}-\alpha_n\sum_{j=1}^{k_n}\frac{\langle\theta,Q\hat{v}_j^{\alpha_n,1}\rangle}{\hat{\lambda}_j^{\alpha_n,1}}\hat{v}_j^{\alpha_n,1}\right\rangle^2,\end{aligned}\tag{3.32}$$

where $\rho_{j_1,j_2}^{\alpha_n,1}=\langle\hat{v}_{j_1}^{\alpha_n,1},Q\hat{v}_{j_2}^{\alpha_n,1}\rangle$. On the other hand, due to the asymptotic expansions of $\hat{\lambda}_j^{\alpha_n,1}$ and $\hat{v}_j^{\alpha_n,1}$, ones has

$$\frac{1}{\hat{\lambda}_j^{\alpha_n,1}}=\frac{1}{\hat{\lambda}_j}+\alpha_n\frac{\rho_j}{\hat{\lambda}_j^2}(1+o_{a.s.}(1))\quad\text{and}\quad\hat{v}_j^{\alpha_n,1}=\hat{v}_j-\alpha_n\Pi_jQ\hat{v}_j+o_{a.s.}(\alpha_n).\tag{3.33}$$

Thus, using (3.33) in (3.32),

$$\begin{aligned}
\mathbb{E}_{\mathcal{X}^{n+1}}(Y_{n+1} - \langle \hat{\theta}_{k_n}^{PS,\,\alpha_n}, X_{n+1}\rangle)^2 &= \sigma^2 + \frac{\sigma^2}{n}\sum_{j=1}^{k_n}\frac{\langle X_{n+1}, \hat{v}_j\rangle^2}{\hat{\lambda}_j} \\
&- 2\alpha_n\frac{\sigma^2}{n}\sum_{j=1}^{k_n}\frac{\langle X_{n+1}, \hat{v}_j\rangle\langle X_{n+1}, \Pi_j Q\hat{v}_j\rangle}{\hat{\lambda}_j}(1+o_{a.s.}(1)) + \alpha_n\frac{\sigma^2}{n}\sum_{j=1}^{k_n}\rho_j\frac{\langle X_{n+1}, \hat{v}_j\rangle^2}{\hat{\lambda}_j^2}(1+o_{a.s.}(1)) \\
&+ \alpha_n\frac{\sigma^2}{n}\sum_{j_1=1}^{k_n}\sum_{j_2=1}^{k_n}\frac{\rho_{j_1,j_2}}{\hat{\lambda}_{j_1}\hat{\lambda}_{j_2}}\langle X_{n+1}, \hat{v}_{j_1}\rangle\langle X_{n+1}, \hat{v}_{j_2}\rangle(1+o_{a.s.}(1)) + \langle X_{n+1}, \hat{R}_{k_n}\rangle^2 \\
&+ 2\alpha_n\langle X_{n+1}, \hat{R}_{k_n}\rangle\langle X_{n+1}, \hat{C}_{k_n}^{PS}\rangle(1+o_{a.s.}(1)) + \alpha_n^2\langle X_{n+1}, \hat{C}_{k_n}^{PS}\rangle^2(1+o_{a.s.}(1)),
\end{aligned}$$

with $\hat{C}_{k_n}^{PS} = \sum_{j=1}^{k_n}(\langle\theta, \hat{v}_j\rangle\Pi_j Q\hat{v}_j + \langle\theta, \Pi_j Q\hat{v}_j\rangle\hat{v}_j - \hat{\lambda}_j^{-1}\langle\theta, Q\hat{v}_j\rangle\hat{v}_j)$, and by Theorem 2.3.14 (see Chapter 2, page 39)

$$\begin{aligned}
&\mathbb{E}_{\mathcal{X}^{n+1}}(Y_{n+1} - \langle \hat{\theta}_{k_n}^{PS,\,\alpha_n}, X_{n+1}\rangle)^2 - \mathbb{E}_{\mathcal{X}^{n+1}}(Y_{n+1} - \langle \hat{\theta}_{k_n}, X_{n+1}\rangle)^2 \\
&= -2\alpha_n\frac{\sigma^2}{n}\sum_{j=1}^{k_n}\frac{\langle X_{n+1}, \hat{v}_j\rangle\langle X_{n+1}, \Pi_j Q\hat{v}_j\rangle}{\hat{\lambda}_j}(1+o_{a.s.}(1)) + \alpha_n\frac{\sigma^2}{n}\sum_{j=1}^{k_n}\rho_j\frac{\langle X_{n+1}, \hat{v}_j\rangle^2}{\hat{\lambda}_j^2}(1+o_{a.s.}(1)) \\
&+ \alpha_n\frac{\sigma^2}{n}\sum_{j_1=1}^{k_n}\sum_{j_2=1}^{k_n}\frac{\rho_{j_1,j_2}}{\hat{\lambda}_{j_1}\hat{\lambda}_{j_2}}\langle X_{n+1}, \hat{v}_{j_1}\rangle\langle X_{n+1}, \hat{v}_{j_2}\rangle(1+o_{a.s.}(1)) \\
&+ 2\alpha_n\langle X_{n+1}, \hat{R}_{k_n}\rangle\langle X_{n+1}, \hat{C}_{k_n}^{PS}\rangle(1+o_{a.s.}(1)) + \alpha_n^2\langle X_{n+1}, \hat{C}_{k_n}^{PS}\rangle^2(1+o_{a.s.}(1)).
\end{aligned}$$

For the conditional estimation error, one gets from Lemma 2.5.2 (see Chapter 2, page 49)

$$\mathbb{E}_{\mathcal{X}^n}(\|\theta - \hat{\theta}_{k_n}^{PS,\,\alpha_n}\|^2) = \frac{\sigma^2}{n}\sum_{j_1=1}^{k_n}\sum_{j_2=1}^{k_n}(\hat{\lambda}_{j_1}^{\alpha_n,1}\hat{\lambda}_{j_2}^{\alpha_n,1})^{-1}\langle\Gamma_n\hat{v}_{j_1}^{\alpha_n,1}, \hat{v}_{j_2}^{\alpha_n,1}\rangle\langle\hat{v}_{j_1}^{\alpha_n,1}, \hat{v}_{j_2}^{\alpha_n,1}\rangle + \|R_{k_n}^{(\gamma,w)}\|^2. \quad (3.34)$$

Using (3.30) and (3.31) in (3.34),

$$\begin{aligned}
\mathbb{E}_{\mathcal{X}^n}(\|\theta - \hat{\theta}_{k_n}^{PS,\,\alpha_n}\|^2) &= \frac{\sigma^2}{n}\sum_{j=1}^{k_n}\frac{1}{\hat{\lambda}_j^{\alpha_n,1}} + \alpha_n\frac{\sigma^2}{n}\sum_{j=1}^{k_n}\frac{\rho_j^{\alpha_n,1}}{(\hat{\lambda}_j^{\alpha_n,1})^2} \\
&+ \left\|\theta - \sum_{j=1}^{k_n}\langle\theta, \hat{v}_j^{\alpha_n,1}\rangle\hat{v}_j^{\alpha_n,1} - \alpha_n\sum_{j=1}^{k_n}\frac{\langle\theta, Q\hat{v}_j^{\alpha_n,1}\rangle}{\hat{\lambda}_j^{\alpha_n,1}}\hat{v}_j^{\alpha_n,1}\right\|^2,
\end{aligned} \quad (3.35)$$

where $\rho_j^{\alpha_n,1} = \langle\hat{v}_j^{\alpha_n,1}, Q\hat{v}_j^{\alpha_n,1}\rangle$. As a result, replacing (3.33) in (3.35), it can be shown that

$$\mathbb{E}_{\mathcal{X}^n}(\|\theta - \hat{\theta}_{k_n}^{PS,\,\alpha_n}\|^2) = \frac{\sigma^2}{n}\sum_{j=1}^{k_n}\frac{1}{\hat{\lambda}_j} + 2\alpha_n\frac{\sigma^2}{n}\sum_{j=1}^{k_n}\frac{\rho_j}{\hat{\lambda}_j^2}(1+o_{a.s.}(1)) + \|\hat{R}_{k_n} + \alpha_n\hat{C}_{k_n}^{PS}\|^2(1+o_{a.s.}(1)),$$

and by Theorem 2.3.14 (see Chapter 2, page 39)

$$\begin{aligned}
\mathbb{E}_{\mathcal{X}^n}(\|\theta - \hat{\theta}_{k_n}^{PS,\,\alpha_n}\|^2) - \mathbb{E}_{\mathcal{X}^n}(\|\theta - \hat{\theta}_{k_n}\|^2) = &\left(2\alpha_n\frac{\sigma^2}{n}\sum_{j=1}^{k_n}\frac{\rho_j}{\hat{\lambda}_j^2} + 2\alpha_n\langle\hat{R}_{k_n}, \hat{C}_{k_n}^{PS}\rangle + \alpha_n^2\|\hat{C}_{k_n}^{PS}\|^2\right) \\
&\cdot(1+o_{a.s.}(1)).
\end{aligned}$$

3.8.13 Proof of Corollary 3.4.2

If $Q = I$, $\rho_j = 1$, $\rho_{j_1,j_2} = \delta_{j_1,j_2}$, and $\Pi_j Q\hat{v}_j = 0$. Thus, Theorem 3.4.1 implies

$$\begin{aligned}
&\mathbb{E}_{\mathcal{X}^{n+1}}(Y_{n+1} - \langle\hat{\theta}_{k_n}^{PS,\,\alpha_n}, X_{n+1}\rangle)^2 - \mathbb{E}_{\mathcal{X}^{n+1}}(Y_{n+1} - \langle\hat{\theta}_{k_n}, X_{n+1}\rangle)^2 \\
&= \left(2\alpha_n\frac{\sigma^2}{n}\sum_{j=1}^{k_n}\frac{\langle X_{n+1},\hat{v}_j\rangle^2}{\hat{\lambda}_j^2} - 2\alpha_n\langle X_{n+1},\hat{R}_{k_n}\rangle\langle X_{n+1},\hat{T}_{k_n}\rangle + \alpha_n^2\langle X_{n+1},\hat{T}_{k_n}\rangle^2\right)(1+o_{a.s.}(1)), \\
&\mathbb{E}_{\mathcal{X}^n}(\|\theta - \hat{\theta}_{k_n}^{PS,\,\alpha_n}\|^2) - \mathbb{E}_{\mathcal{X}^n}(\|\theta - \hat{\theta}_{k_n}\|^2) = \left(2\alpha_n\frac{\sigma^2}{n}\sum_{j=1}^{k_n}\frac{1}{\hat{\lambda}_j^2} + \alpha_n^2\|\hat{T}_{k_n}\|^2\right)(1+o_{a.s.}(1)),
\end{aligned}$$

where $\hat{T}_{k_n}$ is defined as in Theorem 3.2.3 (see page 55).

3.8.14 Proof of Theorem 3.4.4

In this case, Lemma 2.5.2 (see Chapter 2, page 49) is going to be applied with $\gamma_j = (\hat{\lambda}_j^{\alpha_n,2})^{-1}$ and $w_j = \hat{v}_j^{\alpha_n,2}$. Then, the conditional prediction error is

$$\begin{aligned}
&\mathbb{E}_{\mathcal{X}^{n+1}}(Y_{n+1} - \langle\hat{\theta}_{k_n}^{S,\,\alpha_n}, X_{n+1}\rangle)^2 = \sigma^2 \\
&\quad + \frac{\sigma^2}{n}\sum_{j_1=1}^{k_n}\sum_{j_2=1}^{k_n}(\hat{\lambda}_{j_1}^{\alpha_n,2}\hat{\lambda}_{j_2}^{\alpha_n,2})^{-1}\langle\Gamma_n\hat{v}_{j_1}^{\alpha_n,2},\hat{v}_{j_2}^{\alpha_n,2}\rangle\langle X_{n+1},\hat{v}_{j_1}^{\alpha_n,2}\rangle\langle X_{n+1},\hat{v}_{j_2}^{\alpha_n,2}\rangle + \langle X_{n+1},R_{k_n}^{(\gamma,w)}\rangle^2,
\end{aligned} \tag{3.36}$$

where $R_{k_n}^{(\gamma,w)} = \theta - \sum_{j=1}^{k_n}(\hat{\lambda}_j^{\alpha_n,2})^{-1}\langle\Gamma_n\hat{v}_j^{\alpha_n,2},\theta\rangle\hat{v}_j^{\alpha_n,2}$. Note that $\{\hat{\lambda}_j^{\alpha_n,2},\hat{v}_j^{\alpha_n,2}\}_j$ satisfy

$$\Gamma_n\hat{v}_j^{\alpha_n,2} = \hat{\lambda}_j^{\alpha_n,2}(\hat{v}_j^{\alpha_n,2} + \alpha_n Q\hat{v}_j^{\alpha_n,2}), \tag{3.37}$$

whereas the orthonormality conditions are based on the penalized inner product $\langle\cdot,\cdot\rangle_{\alpha_n}$, i.e.,

$$\langle\hat{v}_{j_1}^{\alpha_n,2},\hat{v}_{j_2}^{\alpha_n,2}\rangle_{\alpha_n} = \langle\hat{v}_{j_1}^{\alpha_n,2},\hat{v}_{j_2}^{\alpha_n,2}\rangle + \alpha_n\langle\hat{v}_{j_1}^{\alpha_n,2},Q\hat{v}_{j_2}^{\alpha_n,2}\rangle = \langle\hat{v}_{j_1}^{\alpha_n,2},\hat{v}_{j_2}^{\alpha_n,2}\rangle + \alpha_n\rho_{j_1,j_2}^{\alpha_n,2} = \delta_{j_1j_2}, \tag{3.38}$$

where $\rho_{j_1,j_2}^{\alpha_n,2} = \langle\hat{v}_{j_1}^{\alpha_n,2},Q\hat{v}_{j_2}^{\alpha_n,2}\rangle$ (with $\rho_j^{\alpha_n,2} = \langle\hat{v}_j^{\alpha_n,2},Q\hat{v}_j^{\alpha_n,2}\rangle$), and $\delta_{j_1j_2} = 1$ if $j_1 = j_2$, and 0 otherwise. Therefore,

$$R_{k_n}^{(\gamma,w)} = \theta - \sum_{j=1}^{k_n}\langle\theta,\hat{v}_j^{\alpha_n,2}\rangle\hat{v}_j^{\alpha_n,2} - \alpha_n\sum_{j=1}^{k_n}\langle\theta,Q\hat{v}_j^{\alpha_n,2}\rangle\hat{v}_j^{\alpha_n,2}. \tag{3.39}$$

Using (3.37), (3.38) and (3.39) in (3.36), one has

$$\begin{aligned}
&\mathbb{E}_{\mathcal{X}^{n+1}}(Y_{n+1} - \langle\hat{\theta}_{k_n}^{S,\,\alpha_n}, X_{n+1}\rangle)^2 \\
&= \sigma^2 + \frac{\sigma^2}{n}\sum_{j=1}^{k_n}\frac{\langle X_{n+1},\hat{v}_j^{\alpha_n,2}\rangle^2}{\hat{\lambda}_j^{\alpha_n,2}} + \left\langle X_{n+1},\theta - \sum_{j=1}^{k_n}\langle\theta,\hat{v}_j^{\alpha_n,2}\rangle\hat{v}_j^{\alpha_n,2} - \alpha_n\sum_{j=1}^{k_n}\langle\theta,Q\hat{v}_j^{\alpha_n,2}\rangle\hat{v}_j^{\alpha_n,2}\right\rangle^2.
\end{aligned} \tag{3.40}$$

In this case, the asymptotic expansions of the eigenelements imply

$$\frac{1}{\hat{\lambda}_j^{\alpha_n,2}} = \frac{1}{\hat{\lambda}_j} + \alpha_n\frac{\rho_j}{\hat{\lambda}_j}(1+o_{a.s.}(1)) \quad\text{and}\quad \hat{v}_j^{\alpha_n,2} = \hat{v}_j - \alpha_n\left(\frac{\rho_j}{2} + \hat{\lambda}_j\Pi_j Q\right)\hat{v}_j + o_{a.s.}(\alpha_n). \tag{3.41}$$

As a result, using (3.41) in (3.40), one gets for the conditional prediction error

$$\begin{aligned}
\mathbb{E}_{\mathcal{X}^{n+1}}(Y_{n+1} - \langle\hat{\theta}_{k_n}^{S,\,\alpha_n}, X_{n+1}\rangle)^2 &= \sigma^2 + \frac{\sigma^2}{n}\sum_{j=1}^{k_n}\frac{\langle X_{n+1},\hat{v}_j\rangle^2}{\hat{\lambda}_j} \\
&- 2\alpha_n\frac{\sigma^2}{n}\sum_{j=1}^{k_n}\langle X_{n+1},\hat{v}_j\rangle\langle X_{n+1},\Pi_j Q\hat{v}_j\rangle(1+o_{a.s.}(1)) + \langle X_{n+1},\hat{R}_{k_n}\rangle^2 \\
&+ 2\alpha_n\langle X_{n+1},\hat{R}_{k_n}\rangle\langle X_{n+1},\hat{C}_{k_n}^S\rangle(1+o_{a.s.}(1)) + \alpha_n^2\langle X_{n+1},\hat{C}_{k_n}^S\rangle^2(1+o_{a.s.}(1)),
\end{aligned}$$

where $\hat{C}^S_{k_n} = \sum_{j=1}^{k_n} (\rho_j \langle \theta, \hat{v}_j \rangle \hat{v}_j + \hat{\lambda}_j (\langle \theta, \hat{v}_j \rangle \Pi_j Q \hat{v}_j + \langle \theta, \Pi_j Q \hat{v}_j \rangle \hat{v}_j) - \langle \theta, Q\hat{v}_j \rangle \hat{v}_j)$, and by the error expressions obtained in Theorem 2.3.14 (see Chapter 2, page 39)

$$\begin{aligned}
&\mathbb{E}_{\mathcal{X}^{n+1}}(Y_{n+1} - \langle \hat{\theta}^{S,\,\alpha_n}_{k_n}, X_{n+1} \rangle)^2 - \mathbb{E}_{\mathcal{X}^{n+1}}(Y_{n+1} - \langle \hat{\theta}_{k_n}, X_{n+1} \rangle)^2 \\
&= -2\alpha_n \frac{\sigma^2}{n} \sum_{j=1}^{k_n} \langle X_{n+1}, \hat{v}_j \rangle \langle X_{n+1}, \Pi_j Q \hat{v}_j \rangle (1 + o_{a.s.}(1)) + 2\alpha_n \langle X_{n+1}, \hat{R}_{k_n} \rangle \langle X_{n+1}, \hat{C}^S_{k_n} \rangle (1 + o_{a.s.}(1)) \\
&\quad + \alpha_n^2 \langle X_{n+1}, \hat{C}^S_{k_n} \rangle^2 (1 + o_{a.s.}(1)).
\end{aligned}$$

As far as the conditional estimation error is concerned, Lemma 2.5.2 (see Chapter 2, page 49) ensures

$$\mathbb{E}_{\mathcal{X}^n}(\|\theta - \hat{\theta}^{S,\,\alpha_n}_{k_n}\|^2) = \frac{\sigma^2}{n} \sum_{j_1=1}^{k_n} \sum_{j_2=1}^{k_n} (\hat{\lambda}_{j_1}^{\alpha_n,2} \hat{\lambda}_{j_2}^{\alpha_n,2})^{-1} \langle \Gamma_n \hat{v}_{j_1}^{\alpha_n,2}, \hat{v}_{j_2}^{\alpha_n,2} \rangle \langle \hat{v}_{j_1}^{\alpha_n,2}, \hat{v}_{j_2}^{\alpha_n,2} \rangle + \|R^{(\gamma,w)}_{k_n}\|^2. \quad (3.42)$$

Using (3.37), (3.38) and (3.39) in (3.42), one has

$$\begin{aligned}
\mathbb{E}_{\mathcal{X}^n}(\|\theta - \hat{\theta}^{S,\,\alpha_n}_{k_n}\|^2) = \frac{\sigma^2}{n} \sum_{j=1}^{k_n} \frac{1}{\hat{\lambda}_j^{\alpha_n,2}} - \alpha_n \frac{\sigma^2}{n} \sum_{j=1}^{k_n} \frac{\rho_j^{\alpha_n,2}}{\hat{\lambda}_j^{\alpha_n,2}} \\
+ \left\| \theta - \sum_{j=1}^{k_n} \langle \theta, \hat{v}_j^{\alpha_n,2} \rangle \hat{v}_j^{\alpha_n,2} - \alpha_n \sum_{j=1}^{k_n} \langle \theta, Q\hat{v}_j^{\alpha_n,2} \rangle \hat{v}_j^{\alpha_n,2} \right\|^2.
\end{aligned} \quad (3.43)$$

Then, using (3.41) in (3.43),

$$\mathbb{E}_{\mathcal{X}^n}(\|\theta - \hat{\theta}^{S,\,\alpha_n}_{k_n}\|^2) = \frac{\sigma^2}{n} \sum_{j=1}^{k_n} \frac{1}{\hat{\lambda}_j} + \|\hat{R}_{k_n}\|^2 + 2\alpha_n \langle \hat{R}_{k_n}, \hat{C}^S_{k_n} \rangle (1 + o_{a.s.}(1)) + \alpha_n^2 \|\hat{C}^S_{k_n}\|^2 (1 + o_{a.s.}(1)),$$

and by Theorem 2.3.14 (see Chapter 2, page 39),

$$\mathbb{E}_{\mathcal{X}^n}(\|\theta - \hat{\theta}^{S,\,\alpha_n}_{k_n}\|^2) - \mathbb{E}_{\mathcal{X}^n}(\|\theta - \hat{\theta}_{k_n}\|^2) = 2\alpha_n \langle \hat{R}_{k_n}, \hat{C}^S_{k_n} \rangle (1 + o_{a.s.}(1)) + \alpha_n^2 \|\hat{C}^S_{k_n}\|^2 (1 + o_{a.s.}(1)).$$

Chapter 4

Bootstrap in functional linear regression

Dealing with the functional linear model with functional explanatory variable and scalar response, and as commented previously, one of the most popular methods for parameter model estimation is based on FPCA. Weak convergence for a wide class of FPCA–type estimators has recently been proved and, as a result, asymptotic confidence sets can be obtained. In this chapter, an alternative approach in order to compute pointwise confidence intervals by means of a bootstrap procedure is proposed, obtaining also its asymptotic validity in the sense specified in Theorem 4.3.6 (i.e., the conditional distribution of the estimator can be approximated by the bootstrap distribution). In addition, a simulation study allows to compare the practical performance of asymptotic and bootstrap confidence intervals in terms of coverage rates for different sample sizes.

The methodology presented in this chapter was firstly introduced by González-Manteiga and Martínez-Calvo (2010) and it gave rise to the contribution by González-Manteiga and Martínez-Calvo (2011).

4.1 How to build confidence intervals?

Current technology collects data in such a fine grid that recorded measurements can be seen as observations of variables valued in functional spaces. This fact has aroused great interest in developing techniques that are focused on functional data sets, and many authors have made an effort to adapt the existing multivariate regression methods to the functional regression model with scalar response. As mentioned in previous chapters, regarding *parametric* regression, the most extensively studied model is the functional linear model with scalar response given by (2.1) (see Chapter 2, page 32), that is,

$$Y = \langle \theta, X \rangle + \epsilon,$$

where $m(\cdot) = \langle \theta, \cdot \rangle : \mathcal{H} \to \mathbb{R}$ is a linear regression operator such that $(\mathcal{H}, \langle \cdot, \cdot \rangle)$ is a real separable Hilbert space and $\theta \in \mathcal{H}$ satisfies $\|\theta\|^2 < \infty$ (being $\| \cdot \| = \langle \cdot, \cdot \rangle^{1/2}$), X is a zero–mean random variable valued in $\mathcal{H}$, and Y and ϵ are real random variables with the latter verifying that $\mathbb{E}(\epsilon) = 0$, $\mathbb{E}(\epsilon^2) = \sigma^2 < +\infty$, and $\mathbb{E}(\epsilon X) = 0$. In addition, $\mathbb{E}(\|X\|^4) < \infty$ is required throughout this chapter.

Among the different techniques for estimating θ, methods based on FPCA are quite popular (Cardot et al., 1999, 2003c; Cai and Hall, 2006; Hall and Hosseini-Nasab, 2006; Hall and Horowitz, 2007; Cardot et al., 2007c), and this kind of estimators have been considered in this chapter. However, the aim is not to estimate the regression function $m(\cdot)$ but to obtain pointwise confidence intervals for a certain confidence level $\alpha \in (0, 1)$, that is, $\mathrm{CI}_{x,\alpha} \subset \mathbb{R}$ such that $\mathbb{P}(m(x) \in \mathrm{CI}_{x,\alpha}) = 1 - \alpha$ for a fixed $x \in \mathcal{H}$.

In order to compute such intervals, asymptotic and bootstrap approaches have been widely used in the multivariate regression context. For example, it is well–known the asymptotic normality of the stan-

dard least squares estimator for the linear model, or the asymptotic normality of the Nadaraya–Watson estimator for more general regression functions in a nonparametric setting. Hence, approximated confidence intervals can be built using normal quantiles. As far as bootstrap is concerned, its introduction by Efron (1979) resulted in a new distribution approximation applicable to a large number of situations (Bickel and Freedman, 1981; Singh, 1981; Parr, 1985). In particular, bootstrap validity was obtained for linear and nonparametric regression models by Freedman (1981) and Cao-Abad (1991), respectively.

Nowadays, the adaptation of these procedures to the functional context has been initiated for parametric and nonparametric estimators of the regression operator $m(\cdot)$. In this sense, Cardot et al. (2007c) proved weak convergence for a large class of FPCA–type estimators and Ferraty et al. (2007a) obtained normality results for the nonparametric estimator proposed by Ferraty and Vieu (2006b). Moreover, Ferraty et al. (2010c) and Ferraty et al. (2012d) showed the validity of the bootstrap in nonparametric functional regression with scalar response and functional response, respectively. In this chapter, a bootstrap procedure for the functional linear model with scalar response has been proposed, and its asymptotic validity has been analysed.

Although this chapter is focused on the use of bootstrap techniques in the functional context, some authors studied this issue in other contexts, for instance, in functional estimation (Cuevas et al., 2006). An updated state of the art of methodological and practical developments for resampling methods for functional data (including bootstrap) can be found in McMurry and Politis (2011). Furthermore, contributions such as Giné and Zinn (1990), Dudley (1990), Sheehy and Wellner (1992), Politis and Romano (1994) or van der Vaart and Wellner (1996) provided theoretical tools which allow to obtain validity results for applications of bootstrap to FDA.

Coming back to the problem formulated in this chapter, obtaining pointwise confidence intervals for the linear regression operator $m(\cdot) = \langle \theta, \cdot \rangle$, Section 4.2 is devoted to recall some notation about the general FPCA–type estimator introduced in (2.6) (see Chapter 2, page 37) and present the asymptotic confidence intervals that can be derived from its weak convergence (see Theorem 2.3.17 in Chapter 2, page 40). In Section 4.3, a naive and a wild bootstrap procedures are described (they are similar to the ones considered by Ferraty et al., 2010c in the nonparametric case), and its asymptotic validity is shown. A simulation study is compiled in Section 4.4 in order to compare asymptotic and bootstrap intervals. Finally, some conclusions can be found in Section 4.5, whereas appendix collects the proof of the main result of the chapter and some necessary technical lemmas (see Section 4.6).

4.2 Asymptotic confidence intervals for linear regression

First of all, recall the general class of FPCA–type estimators introduced in Section 2.3.2, *"b) Definition of general class of FPCA–type estimators"*, in Chapter 2 (see page 37). Let $\{(X_i, Y_i)\}_{i=1}^n$ be a sample of i.i.d. random variables drawn from (X, Y). Assuming that the eigenvalues of Γ verify that $\lambda_1 > \lambda_2 > \ldots > 0$, with the multiplicity of each λ_j equals to one, and assuming that **(C.2.1)** holds (see Chapter 2, page 36), Cardot et al. (2007c) developed asymptotic theory for a large class of FPCA–type estimators given by (2.6) (see Chapter 2, page 37), that is,

$$\hat{\theta}_c = \sum_{j=1}^{n} f_n^c(\hat{\lambda}_j) \Delta_n \hat{v}_j \hat{v}_j,$$

where $c = c_n$ is a strictly positive sequence such that $c \to 0$ and $c < \lambda_1$, $\{f_n^c : [c, +\infty) \to \mathbb{R}\}_n$ is a sequence of positive functions such that **(C.2.2)**–**(C.2.4)** are satisfied (see Chapter 2, page 37), whereas, as usual, $\Delta_n = n^{-1} \sum_{i=1}^n X_i \otimes_{\mathcal{H}'} Y_i$, and $\{(\hat{\lambda}_j, \hat{v}_j)\}_{j=1}^{\infty}$ are the eigenvalues and the eigenfunctions of $\Gamma_n = n^{-1} \sum_{i=1}^n X_i \otimes_{\mathcal{H}} X_i$.

Furthermore, Remark 2.3.8 (see Chapter 2, page 37) highlighted that if

$$k_n^c = \sup \{j : \lambda_j + \delta_j/2 \geq c\}$$

verifies that $(k_n^c)^2 \log k_n^c / \sqrt{n} \to 0$, then $\hat{\theta}_c \approx \sum_{j=1}^{k_n^c} f_n^c(\hat{\lambda}_j) \Delta_n \hat{v}_j \hat{v}_j$. Hence, the standard FPCA estimator $\hat{\theta}_{k_n}$ (see (2.4) in Chapter 2, page 36) is asymptotically equivalent to $\hat{\theta}_c$ when $f_n(x) = x^{-1} \mathbb{I}_{\{x \geq c\}}$,

whereas the presmoothed FPCA estimator $\hat{\theta}_{k_n}^{\alpha_n}$ (see (3.1) in Chapter 3, page 54) is asymptotically equivalent to $\hat{\theta}_c$ when $f_n(x) = (x + \alpha_n)^{-1}\mathbb{I}_{\{x \geq c\}}$.

For $\hat{\theta}_c$, Cardot et al. (2007c) derived confidence sets for prediction by means of a CLT for the weak topology of the functional space $\mathcal{H}$ (see Theorem 2.3.17 in Chapter 2, page 40). In particular, given $x \in \mathcal{H}$, Corollary 2.3.18 (see Chapter 2, page 40) ensures that

$$\frac{\sqrt{n}}{\hat{t}_{n,x}^c \hat{\sigma}}(\hat{m}_c(x) - \langle \hat{\Pi}_{k_n^c}\theta, x\rangle) \xrightarrow{w} \mathcal{N}(0,1),$$

being $\hat{t}_{n,x}^c = \sqrt{\sum_{j=1}^{k_n^c} \hat{\lambda}_j (f_n^c(\hat{\lambda}_j))^2 \langle x, \hat{v}_j\rangle^2}$, $\hat{\sigma}^2$ a consistent estimate of σ^2, $\hat{m}_c(\cdot) = \langle \hat{\theta}_c, \cdot\rangle$, and $\hat{\Pi}_{k_n^c}$ the projector onto the subspace spanned by the first k_n^c eigenfunctions of Γ_n.

Thus, the approximated asymptotic confidence intervals for $m(\cdot)$ introduced in (2.8) (see Chapter 2, page 41) can be derived as follows for a fixed confidence level $\alpha \in (0,1)$. Let z_α be the quantile of order α from a $\mathcal{N}(0,1)$ distribution, and suppose that θ (or x) can be "well" approximated by its projection $\hat{\Pi}_{k_n^c}\theta$ (or $\hat{\Pi}_{k_n^c}x$). Then, Corollary 2.3.18 (see Chapter 2, page 40) ensures that

$$\mathrm{CI}_{x,\alpha}^{asy} = \left[\hat{m}_c(x) - z_{1-\alpha/2}\frac{\hat{t}_{n,x}^c\hat{\sigma}}{\sqrt{n}}, \hat{m}_c(x) + z_{1-\alpha/2}\frac{\hat{t}_{n,x}^c\hat{\sigma}}{\sqrt{n}}\right]$$

satisfies that $\mathbb{P}(m(x) \in \mathrm{CI}_{x,\alpha}^{asy}) \approx 1 - \alpha$.

Remark 4.2.1. Broadly speaking, the construction of $\mathrm{CI}_{x,\alpha}^{asy}$ is implicitly based on the replacement of θ (or x) by $\hat{\Pi}_{k_n^c}\theta$ (or $\hat{\Pi}_{k_n^c}x$) in Corollary 2.3.18 (see Chapter 2, page 40). However, Cardot et al. (2007c) indicated in their Remark 5 that this substitution requires very restrictive conditions either on θ or on the eigenvalues $\{\lambda_j\}_{j=1}^{\infty}$. Despite this fact, in practice, intervals $\mathrm{CI}_{x,\alpha}^{asy}$ will be built, but taking into account that they can be decentred due to this replacement.

4.3 Bootstrap confidence intervals for linear regression

The introduction of bootstrap techniques in this chapter has as its objective to build pointwise confidence intervals for the regression operator, which are able to compete with the asymptotic approach presented in Section 4.2. In Section 4.3.1, the bootstrap procedure that will be considered in this chapter is introduced, and Section 4.3.2 collects the main theorem that ensures the asymptotic validity of the proposed bootstrap method.

4.3.1 Naive and wild bootstrap

The multivariate *naive* and *wild* bootstrap procedures have been adapted to the functional context, in the same way as Ferraty et al. (2010c) did for the nonparametric case. Algorithms for resampling proceeds as follows.

Algorithm 4.3.1 (Naive bootstrap)**.**

Step 1. *Construct a pilot estimator for θ: $\hat{\theta}_d = \sum_{j=1}^n f_n^d(\hat{\lambda}_j)\Delta_n\hat{v}_j\hat{v}_j$. Obtain the residuals $\hat{\epsilon}_i = Y_i - \langle\hat{\theta}_d, X_i\rangle$ for all $i = 1,\ldots,n$.*

Step 2. *Draw $\hat{\epsilon}_1^*,\ldots,\hat{\epsilon}_n^*$ i.i.d. random variables from the cumulative distribution of $(\hat{\epsilon}_i - \bar{\hat{\epsilon}})_{i=1}^n$, where $\bar{\hat{\epsilon}} = n^{-1}\sum_{i=1}^n \hat{\epsilon}_i$.*

Step 3. *Define $Y_i^* = \langle\hat{\theta}_d, X_i\rangle + \hat{\epsilon}_i^*$, for all $i = 1,\ldots,n$.*

Step 4. *Use the bootstrap sample $\{(X_i, Y_i^*)\}_{i=1}^n$ in order to obtain the following estimator of θ: $\hat{\theta}_{c,d}^* = \sum_{j=1}^n f_n^c(\hat{\lambda}_j)\Delta_n^*\hat{v}_j\hat{v}_j$, where $\Delta_n^* = n^{-1}\sum_{i=1}^n X_i \otimes_{\mathcal{H}'} Y_i^*$.*

Step 5. *Repeat Step 2–Step 4 a large number of times $n_{boot} \in \mathbb{N}$ in order to obtain a sequence of values $\{\hat{\theta}^{*,l}_{c,d}\}_{l=1}^{n_{boot}}$.*

Algorithm 4.3.2 (Wild bootstrap)**.**

Step 1. *Construct a pilot estimator for θ: $\hat{\theta}_d = \sum_{j=1}^{n} f_n^d(\hat{\lambda}_j)\Delta_n \hat{v}_j \hat{v}_j$. Obtain the residuals $\hat{\epsilon}_i = Y_i - \langle \hat{\theta}_d, X_i \rangle$ for all $i = 1, \ldots, n$.*

Step 2. *Define $\hat{\epsilon}_i^* = \hat{\epsilon}_i V_i$ for $i = 1, \ldots, n$, being $V_1, \ldots, V_n$ i.i.d. random variables independent of the data $\{(X_i, Y_i)\}_{i=1}^n$, such that $\mathbb{E}(V_1) = 0$ and $\mathbb{E}(V_1^2) = 1$.*

Step 3. *Define $Y_i^* = \langle \hat{\theta}_d, X_i \rangle + \hat{\epsilon}_i^*$, for all $i = 1, \ldots, n$.*

Step 4. *Use the bootstrap sample $\{(X_i, Y_i^*)\}_{i=1}^n$ in order to obtain the following estimator of θ: $\hat{\theta}^*_{c,d} = \sum_{j=1}^{n} f_n^c(\hat{\lambda}_j)\Delta_n^* \hat{v}_j \hat{v}_j$, where $\Delta_n^* = n^{-1}\sum_{i=1}^{n} X_i \otimes_{\mathcal{H}'} Y_i^*$.*

Step 5. *Repeat Step 2–Step 4 a large number of times $n_{boot} \in \mathbb{N}$ in order to obtain a sequence of values $\{\hat{\theta}^{*,l}_{c,d}\}_{l=1}^{n_{boot}}$.*

As it can be seen in the previous algorithm, two sequences must be fixed: d for the pilot estimator $\hat{\theta}_d$ and c for $\hat{\theta}^*_{c,d}$. In fact, the following assumption will be required to guarantee the asymptotic validity of the bootstrap method

(C.4.1) $c = c_n$ and $d = d_n$ are strictly positive sequence such that $c, d \to 0$ and $c, d < \lambda_1$. Moreover, $c \leq d$ for all n.

Remark 4.3.3. Note that under **(C.4.1)**, that is, if $c \leq d$, the number of principal components used for constructing $\hat{\theta}^*_{c,d}$ is larger than the number of components used for $\hat{\theta}_d$. Therefore, in some way, one should *oversmooth* when the pilot estimator is calculated. In this sense, the need to choose the sequences d and c creates a similar problem to the selection of bandwidth in the nonparametric regression: b for the pilot estimator and h for the bootstrap one. However, the choice of optimal *smoothing* parameters is still an open question in both this parametric context and nonparametric one: see Mammen (2000) or González-Manteiga et al. (2004) for the multivariate nonparametric regression, and Ferraty et al. (2010c) for the functional nonparametric regression.

4.3.2 Asymptotic validity of the bootstrap

From now on, x will denote a fixed element of $\mathcal{H}$. Furthermore, recall the convergence in probability definition and its associated "big O" and "little o" notation.

Definition 4.3.4. Let $\{Z_n\}_{n\in\mathbb{N}}$ be a sequence of real random variables and let Z be a real random variable. $\{Z_n\}_{n\in\mathbb{N}}$ converges in *probability* ($\mathbb{P}$) to Z, that is,

$$\text{plim}_{n\to\infty} Z_n = Z, \quad \text{or equivalently,} \quad Z_n \xrightarrow{\mathbb{P}} Z,$$

if and only if

$$\forall \varepsilon > 0, \quad \lim_{n\to\infty} \mathbb{P}(|Z_n - Z| > \varepsilon) = 0.$$

Definition 4.3.5. Let $\{Z_n\}_{n\in\mathbb{N}}$ be a sequence of real random variables, let Z be a real random variable, and let $\{u_n\}_{n\in\mathbb{N}}$ be a deterministic sequence of positive real numbers. The *rate of convergence in probability* of $\{Z_n\}_{n\in\mathbb{N}}$ to Z is of order u_n, that is,

$$Z_n - Z = O_{\mathbb{P}}(u_n),$$

if and only if

$$\lim_{p} \overline{\lim}_{n\to\infty} \mathbb{P}(|Z_n - Z| < pu_n) = 1.$$

Furthermore,

$$Z_n - Z = o_{\mathbb{P}}(u_n) \quad \text{if and only if} \quad (u_n)^{-1}(Z_n - Z) \xrightarrow{\mathbb{P}} 0.$$

In order to obtain the main result in this chapter, it is necessary that the hypotheses required by Cardot et al. (2007c) in their weak convergence results for the prediction at a given value hold (see Theorem 2.3.17 in Chapter 2, page 40). This fact will be essential to show the asymptotic validity of the bootstrap procedure in the next theorem, where $\mathbb{P}_{\mathcal{X}^n\mathcal{Y}^n}$ denotes the probability conditionally on $(\mathcal{X}^n, \mathcal{Y}^n) = \{(X_1, Y_1), \ldots, (X_n, Y_n)\}$, and $\mathbb{P}_{\mathcal{X}^n}$ denotes the probability conditionally on $\mathcal{X}^n = \{X_1, \ldots, X_n\}$. In addition, the following notation is also used: $\hat{m}^*_{c,d}(\cdot) = \langle \hat{\theta}^*_{c,d}, \cdot\rangle$, $\hat{m}_d(\cdot) = \langle \hat{\theta}_d, \cdot\rangle$, and $\hat{m}_c(\cdot) = \langle \hat{\theta}_c, \cdot\rangle$.

Theorem 4.3.6. *Under the assumptions of Theorem 2.3.17 (see Chapter 2, page 40), if* **(C.4.1)** *is satisfied, it holds that, for both the naive and the wild bootstrap,*

$$\sup_{y\in\mathbb{R}} |\mathbb{P}_{\mathcal{X}^n\mathcal{Y}^n}(\sqrt{n}(\hat{m}^*_{c,d}(x) - \hat{m}_d(x)) \le y) - \mathbb{P}_{\mathcal{X}^n}(\sqrt{n}(\hat{m}_c(x) - \langle \hat{\Pi}_{k_n^c}\theta, x\rangle) \le y)| \xrightarrow{\mathbb{P}} 0,$$

where $\hat{\Pi}_{k_n^c}$ *is the projector onto the subspace spanned by the first* k_n^c *eigenfunctions of* Γ_n*.*

The proof of the previous theorem can be found in Section 4.6.1 (see page 102).

Let $\alpha \in (0,1)$ be a confidence level. Assuming that θ (or x) can be well approximated by its projection $\hat{\Pi}_{k_n^c}\theta$ (or $\hat{\Pi}_{k_n^c}x$), Theorem 4.3.6 ensures that the distribution of the true error $(\hat{m}_c(x) - m(x))$ can be approximated by the distribution of the bootstrap error $(\hat{m}^*_{c,d}(x) - \hat{m}_d(x))$. Hence, the pointwise α–quantile $q_\alpha(x)$ of the distribution of the true error can be approximated by the corresponding pointwise bootstrap α–quantile $q^*_\alpha(x)$. Therefore, one can approximate the $(1-\alpha)$–confidence interval for $m(x)$ by means of

$$\mathrm{CI}^*_{x,\alpha} = \left[\hat{m}_c(x) - q^*_{1-\alpha/2}(x), \hat{m}_c(x) - q^*_{\alpha/2}(x)\right], \tag{4.1}$$

since $\mathbb{P}(m(x) \in \mathrm{CI}^*_{x,\alpha}) \approx 1 - \alpha$.

Remark 4.3.7. As highlighted in Remark 4.2.1 for the asymptotic intervals $\mathrm{CI}^{asy}_{x,\alpha}$, the replacement of θ (or x) by its projection onto the subspace spanned by the first k_n eigenfunctions of Γ_n needs certain specific hypotheses on θ or on the eigenvalues of Γ which are hardly ever fulfilled. Hence, a decentring effect could also appear when the bootstrap confidence intervals (4.1) are built in practice.

4.4 Simulation study

For the simulation study, $\mathcal{H} = L^2([0,1])$ was selected, with $\langle x, y\rangle = \int_0^1 x(t)y(t)dt$ for all $x, y \in L^2([0,1])$.

The study required the simulation of 500 samples, each one consisted of n observations ($n = 50, 100, 200, 500$) from the model $Y = \langle \theta, X\rangle + \epsilon = \int_0^1 \theta(t)X(t)dt + \epsilon$, where X is a Brownian motion and $\epsilon \sim \mathcal{N}(0, \sigma^2)$ with signal–to–noise ratio $r = \sigma/\sqrt{\mathbb{E}(\langle X, \theta\rangle^2)} = 0.2$. For the parameter θ, the

following functions were considered

$$\theta_1(t) = \sin(4\pi t) \quad \text{and} \quad \theta_2(t) = |t - 0.5|, \qquad \forall t \in [0, 1].$$

Six deterministic curves x

$$x_1(t) = \sin(\pi t/2), \quad x_2(t) = \sin(3\pi t/2), \quad x_3(t) = t,$$
$$x_4(t) = t^2, \quad x_5(t) = 2|t - 0.5| \quad \text{and} \quad x_6(t) = 2\mathbb{I}_{\{t>0.5\}}, \qquad \forall t \in [0, 1],$$

were also fixed, for which pointwise asymptotic and bootstrap prediction intervals, their empirical coverage rate and length were obtained. Note that all the curves (X, θ and x) were discretized to $p = 100$ equispaced design points in the interval $[0, 1]$, and quadrature weights of p^{-1} were used to approximate integrals which had to be computed. Further details about this approximation procedure can be found in Section 3.5, *"How to work with discrete data?"* (see Chapter 3, page 63).

In order to illustrate the behaviour of confidence intervals, two confidence levels ($\alpha = 0.05, 0.10$) were considered and the standard FPCA estimator $\hat{\theta}_{k_n}$ (see (2.4) in Chapter 2, page 36) was chosen for estimating the model parameter θ. For this estimator, k_n was selected by means of GCV method (see Section 3.5, *"Parameter selection"*, in Chapter 3, page 63), denoting this value by $\hat{k}_n^{\text{GCV}}$.

Since the confidence intervals presented in this chapter are based on asymptotic and bootstrap approximations of the pointwise α–quantile $q_\alpha(x)$ of the distribution of the true error $(\langle\hat{\theta}_{k_n}, x\rangle - \langle\theta, x\rangle)$, it was decided to estimate these quantiles by means of a Monte–Carlo procedure. In this way, the following confidence intervals for $m(x)$ were computed

$$\text{CI}_{x,\alpha}^{mc} = \left[\langle\hat{\theta}_{k_n}, x\rangle - q_{1-\alpha/2}^{mc}(x), \langle\hat{\theta}_{k_n}, x\rangle - q_{\alpha/2}^{mc}(x)\right], \tag{4.2}$$

where $q_\alpha^{mc}(x)$ is the estimated α–quantile of the distribution of the true error obtained from $1,000$ Monte–Carlo iterations, that is, from $\{\langle\hat{\theta}_{k_n}^{mc,l} - \theta, x\rangle\}_{l=1}^{1,000}$.

The asymptotic confidence intervals $\text{CI}_{x,\alpha}^{asy}$ defined in (2.8) (see Chapter 2, page 41) were also obtained by means of

$$\text{CI}_{x,\alpha}^{asy} = \left[\langle\hat{\theta}_{k_n}, x\rangle - z_{1-\alpha/2}\frac{\hat{t}_{n,x}^{k_n}\hat{\sigma}}{\sqrt{n}}, \langle\hat{\theta}_{k_n}, x\rangle + z_{1-\alpha/2}\frac{\hat{t}_{n,x}^{k_n}\hat{\sigma}}{\sqrt{n}}\right], \tag{4.3}$$

where $\hat{t}_{n,x}^{k_n} = \sqrt{\sum_{j=1}^{k_n}\hat{\lambda}_j^{-1}\langle x, \hat{v}_j\rangle^2}$ and the true variance σ^2 was estimated by the residual mean of squares for the $\hat{k}_n^{\text{GCV}}$ value.

The previous confidence intervals were compared with the bootstrap intervals introduced in (4.1) (see page 91) and defined as

$$\text{CI}_{x,\alpha}^{*} = \left[\langle\hat{\theta}_{k_n}, x\rangle - q_{1-\alpha/2}^{*}(x), \langle\hat{\theta}_{k_n}, x\rangle - q_{\alpha/2}^{*}(x)\right], \tag{4.4}$$

where the pointwise bootstrap quantiles were computed from $1,000$ bootstrap iterations of the wild bootstrap procedure presented in Algorithm 4.3.2 (see page 90), that is, from $\{\langle\hat{\theta}_{k_n^c,k_n^d}^{*,l} - \hat{\theta}_{k_n^d}^{l}, x\rangle\}_{l=1}^{1,000}$. For the first step in Algorithm 4.3.2, the construction of the pilot estimator of θ, ten different pilot values for k_n^d were considered in order to analyse its effect in the results: $k_n^d \in \{\hat{k}_n^{\text{GCV}} - 6, \ldots, \hat{k}_n^{\text{GCV}} + 3\}$ (recall that $\hat{k}_n^{\text{GCV}}$ is the number of principal components selected by GCV for the original sample). For the second step, $\{V_1, \ldots, V_n\}$ were drawn from the following sum of two Dirac distributions: $0.1(5 + \sqrt{5})\delta_{(1-\sqrt{5})/2} + 0.1(5 - \sqrt{5})\delta_{(1+\sqrt{5})/2}$ (i.e., $\mathbb{P}(V_i = (1 - \sqrt{5})/2) = 0.1(5 + \sqrt{5})$ and $\mathbb{P}(V_i = (1 + \sqrt{5})/2) = 0.1(5 - \sqrt{5})$ for all $i = 1 \ldots, n$). Finally, GCV was used again to select an adequate k_n^c for $\hat{\theta}_{k_n^c,k_n^d}^{*,l}$ in the fourth step of the algorithm.

Note that several types of bootstrap confidence intervals have been proposed in the literature. Hall (1988) examined and compared seven bootstrap methods, among them the *studentized* bootstrap. In the simulation study, studentized versions of the Monte–Carlo intervals and bootstrap intervals were

also computed in order to compare them with (4.2) and (4.4) as follows. The studentized Monte–Carlo intervals were defined by

$$\mathrm{CI}_{x,\alpha}^{mc,s} = \left[\langle \hat{\theta}_{k_n}, x \rangle - q_{1-\alpha/2}^{mc,s}(x) \frac{\hat{t}_{n,x}^{k_n} \hat{\sigma}}{\sqrt{n}}, \langle \hat{\theta}_{k_n}, x \rangle - q_{\alpha/2}^{mc,s}(x) \frac{\hat{t}_{n,x}^{k_n} \hat{\sigma}}{\sqrt{n}} \right], \tag{4.5}$$

where $q_\alpha^{mc,s}(x)$ is the α–quantile estimated from $\{\sqrt{n}\langle \hat{\theta}_{k_n}^{mc,l} - \theta, x\rangle / (\hat{t}_{n,x}^{mc,l,k_n} \hat{\sigma}^{mc,l})\}_{l=1}^{1,000}$. Analogously, the studentized bootstrap intervals were built as

$$\mathrm{CI}_{x,\alpha}^{*,s} = \left[\langle \hat{\theta}_{k_n}, x \rangle - q_{1-\alpha/2}^{*,s}(x) \frac{\hat{t}_{n,x}^{k_n} \hat{\sigma}}{\sqrt{n}}, \langle \hat{\theta}_{k_n}, x \rangle - q_{\alpha/2}^{*,s}(x) \frac{\hat{t}_{n,x}^{k_n} \hat{\sigma}}{\sqrt{n}} \right], \tag{4.6}$$

where $q_\alpha^{*,s}(x)$ were computed from $\{\sqrt{n}\langle \hat{\theta}_{k_n^c, k_n^d}^{*,l} - \hat{\theta}_{k_n^d}^{l}, x\rangle / (\hat{t}_{n,x}^{*,l,k_n^c} \hat{\sigma}^{*,l})\}_{l=1}^{1,000}$.

Therefore, the simulation study compared the confidence intervals for $m(x)$ obtained by Monte–Carlo (see (4.2) and (4.5)), by asymptotic approximations (see (4.3)), and by bootstrap methodology (see (4.4) and (4.6)). Table 4.1, Table 4.2, Table 4.3, and Table 4.4 (see pages 94–97) contain the empirical coverage rate and the mean length of the intervals for the different sample sizes when the model parameter is θ_1, whereas Table 4.5, Table 4.6, Table 4.7, and Table 4.8 (see pages 98–101) show the same information for the second case in which the model parameter is θ_2. In order to clarify tables, the mean length of the intervals was multiplied by 10^2 and appears in brackets after the empirical coverage rate.

From these results, one can deduce that with a correct pilot k_n selection, bootstrap intervals have empirical coverage rates nearer to α than asymptotic intervals. However, these *optimal* bootstrap intervals tend to be larger than the asymptotic ones.

In general, the asymptotic approach tends to give larger coverage rates and shorter intervals than the *optimal* bootstrap intervals. Furthermore, asymptotic intervals look like being decentred. It has been mentioned before that using both asymptotic and bootstrap confidence intervals could cause a decentring effect due to the substitution of θ (or x) by its projection onto the subspace spanned by the first k_n eigenfunctions of Γ_n (see Remark 4.2.1, page 89, and Remark 4.3.7, page 91). However, it seems that the bootstrap intervals can prevent this effect with an adequate pilot selection and a length increment.

As commented in the construction of the confidence intervals were built, it was assumed that either θ or x is well approximated by its projection on the subspace spanned by the first k_n eigenfunctions of Γ_n. In the simulation, θ_1 is a smooth function whereas θ_2 does not have a continuous derivative. Besides, the deterministic curves x are ordered from greatest to least according to their *smoothness* level. This fact justify that both asymptotic and bootstrap intervals give worse results for θ_2 than for θ_1, and better results for $x_1, \ldots, x_4$ than for x_5 and x_6. Note that results are really misleading when $\theta = \theta_2$ and $x = x_5$; in this case, the empirical coverage rates are quite far from the theoretical ones. On the other hand, the sample size affects asymptotic intervals seriously: their empirical coverage is far from the nominal one when $n = 50$. Nevertheless, bootstrap intervals behave properly for all sample sizes.

For bootstrap procedures, it is not easy to deduce from simulations which is the best choice of the pilot parameter k_n^d. Comparing the results of $\mathrm{CI}_{x,\alpha}^{*}$ and $\mathrm{CI}_{x,\alpha}^{*,s}$, it can be found that the k_n^d selection has less effect in the empirical coverage for the *studentized* intervals than for $\mathrm{CI}_{x,\alpha}^{*}$. For the latter, the most adequate k_n^d seems to be smaller than the value obtained by GCV and it tends to increase in accordance with the smoothness of x when a moderate sample size is considered. Nevertheless, this reasoning is not clear in large sample size cases.

To sum up, bootstrap intervals can be an interesting alternative to the asymptotic confidence intervals, specially for small sample sizes. However, the profit of bootstrap procedure is subject to a correct choice of pilot k_n^d, being this choice more critical for $\mathrm{CI}_{x,\alpha}^{*}$ than for $\mathrm{CI}_{x,\alpha}^{*,s}$. Furthermore, the adequate selection of k_n^d seems to be influenced by the smoothness of θ and x, and by the sample size n. Undoubtedly, this is an open question which requires further research.

α	CI	x_1	x_2	x_3	x_4	x_5	x_6
5%	$\mathrm{CI}^{mc}_{x,\alpha}$	4.6 (1.35)	6.6 (4.09)	6.0 (1.21)	8.0 (1.37)	6.4 (5.53)	6.2 (7.88)
	$\mathrm{CI}^{mc,s}_{x,\alpha}$	4.0 (1.40)	4.6 (4.18)	4.6 (1.25)	6.6 (1.42)	5.6 (5.65)	6.2 (8.05)
	$\mathrm{CI}^{asy}_{x,\alpha}$	8.8 (1.15)	9.0 (3.44)	10.6 (1.02)	13.2 (1.15)	19.8 (3.83)	23.0 (5.46)
	$\mathrm{CI}^{*}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}} + 3$	11.2 (1.13)	10.4 (3.37)	11.2 (1.01)	12.8 (1.14)	14.0 (4.61)	15.6 (6.47)
	$\mathrm{CI}^{*}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}} + 2$	10.6 (1.14)	10.8 (3.38)	11.4 (1.01)	13.4 (1.14)	14.4 (4.53)	17.0 (6.33)
	$\mathrm{CI}^{*}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}} + 1$	10.4 (1.15)	10.4 (3.41)	12.0 (1.02)	13.2 (1.14)	15.6 (4.45)	19.6 (6.27)
	$\mathrm{CI}^{*}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}$	10.6 (1.15)	11.6 (3.43)	11.6 (1.02)	13.6 (1.15)	14.4 (4.41)	18.8 (6.23)
	$\mathrm{CI}^{*}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}} - 1$	6.4 (1.36)	8.8 (4.04)	8.0 (1.21)	10.2 (1.37)	11.2 (4.97)	15.2 (7.11)
	$\mathrm{CI}^{*}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}} - 2$	5.4 (1.67)	5.4 (4.99)	5.8 (1.48)	7.4 (1.67)	7.6 (5.95)	10.8 (8.69)
	$\mathrm{CI}^{*}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}} - 3$	4.4 (2.11)	3.2 (6.33)	4.6 (1.88)	5.8 (2.11)	6.4 (7.33)	9.8 (10.97)
	$\mathrm{CI}^{*}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}} - 4$	3.2 (2.62)	2.2 (7.74)	3.8 (2.32)	4.2 (2.59)	5.0 (8.75)	7.2 (13.59)
	$\mathrm{CI}^{*}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}} - 5$	2.2 (2.96)	1.8 (8.80)	2.8 (2.63)	2.4 (2.92)	4.2 (9.69)	5.4 (15.63)
	$\mathrm{CI}^{*}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}} - 6$	1.4 (3.21)	0.4 (9.56)	2.0 (2.86)	2.4 (3.19)	3.2 (10.29)	3.4 (17.17)
	$\mathrm{CI}^{*,s}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}} + 3$	6.2 (1.35)	6.2 (4.00)	7.4 (1.19)	8.8 (1.34)	11.2 (4.98)	14.0 (7.07)
	$\mathrm{CI}^{*,s}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}} + 2$	6.8 (1.33)	6.8 (3.96)	6.8 (1.18)	9.2 (1.33)	11.0 (4.94)	14.2 (6.97)
	$\mathrm{CI}^{*,s}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}} + 1$	6.0 (1.33)	6.4 (3.94)	7.2 (1.18)	10.2 (1.32)	12.6 (4.86)	16.6 (6.91)
	$\mathrm{CI}^{*,s}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}$	6.0 (1.32)	7.0 (3.93)	7.4 (1.18)	9.8 (1.31)	13.0 (4.83)	16.0 (6.87)
	$\mathrm{CI}^{*,s}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}} - 1$	5.8 (1.32)	6.4 (3.92)	7.4 (1.18)	9.4 (1.33)	12.4 (4.86)	14.6 (6.99)
	$\mathrm{CI}^{*,s}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}} - 2$	6.0 (1.32)	6.4 (3.93)	7.6 (1.18)	8.8 (1.34)	9.4 (5.11)	13.0 (7.17)
	$\mathrm{CI}^{*,s}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}} - 3$	5.8 (1.32)	5.8 (3.95)	7.2 (1.18)	9.6 (1.35)	9.4 (5.33)	12.0 (7.45)
	$\mathrm{CI}^{*,s}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}} - 4$	6.8 (1.32)	5.8 (3.93)	6.8 (1.19)	8.2 (1.36)	8.4 (5.62)	9.4 (7.84)
	$\mathrm{CI}^{*,s}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}} - 5$	7.0 (1.31)	7.0 (3.93)	7.8 (1.18)	10.2 (1.36)	8.8 (5.65)	7.2 (8.09)
	$\mathrm{CI}^{*,s}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}} - 6$	6.2 (1.31)	6.0 (3.91)	7.8 (1.18)	9.4 (1.37)	8.4 (5.51)	7.2 (8.33)
10%	$\mathrm{CI}^{mc}_{x,\alpha}$	10.2 (1.13)	10.2 (3.40)	9.8 (1.01)	12.8 (1.15)	11.8 (4.60)	10.6 (6.53)
	$\mathrm{CI}^{mc,s}_{x,\alpha}$	9.2 (1.16)	8.8 (3.48)	10.2 (1.04)	13.0 (1.18)	11.2 (4.69)	12.0 (6.68)
	$\mathrm{CI}^{asy}_{x,\alpha}$	17.4 (0.97)	15.2 (2.89)	18.0 (0.86)	19.2 (0.96)	26.0 (3.21)	29.8 (4.58)
	$\mathrm{CI}^{*}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}} + 3$	16.2 (0.96)	18.8 (2.86)	18.2 (0.85)	17.4 (0.97)	19.2 (3.85)	24.6 (5.40)
	$\mathrm{CI}^{*}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}} + 2$	17.2 (0.96)	18.0 (2.87)	18.2 (0.86)	19.6 (0.97)	21.0 (3.77)	26.8 (5.29)
	$\mathrm{CI}^{*}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}} + 1$	17.2 (0.97)	18.0 (2.88)	18.8 (0.86)	19.4 (0.97)	20.6 (3.70)	26.2 (5.21)
	$\mathrm{CI}^{*}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}$	17.4 (0.97)	17.6 (2.89)	18.4 (0.86)	19.2 (0.97)	21.8 (3.66)	27.6 (5.16)
	$\mathrm{CI}^{*}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}} - 1$	12.6 (1.15)	12.0 (3.42)	13.8 (1.03)	14.4 (1.16)	18.6 (4.08)	20.8 (5.87)
	$\mathrm{CI}^{*}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}} - 2$	10.4 (1.41)	10.8 (4.22)	10.0 (1.26)	12.4 (1.41)	14.8 (4.86)	18.2 (7.13)
	$\mathrm{CI}^{*}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}} - 3$	6.6 (1.78)	5.8 (5.35)	6.6 (1.59)	8.0 (1.78)	10.8 (5.92)	13.6 (8.93)
	$\mathrm{CI}^{*}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}} - 4$	5.6 (2.21)	4.6 (6.55)	5.4 (1.96)	5.6 (2.18)	8.0 (7.03)	10.0 (10.97)
	$\mathrm{CI}^{*}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}} - 5$	3.8 (2.51)	2.6 (7.44)	4.0 (2.22)	4.8 (2.46)	7.0 (7.71)	7.2 (12.58)
	$\mathrm{CI}^{*}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}} - 6$	3.0 (2.72)	1.8 (8.09)	3.0 (2.41)	3.6 (2.68)	5.4 (8.14)	5.2 (13.75)
	$\mathrm{CI}^{*,s}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}} + 3$	11.2 (1.13)	11.6 (3.35)	11.8 (1.00)	13.8 (1.13)	15.6 (4.19)	21.0 (5.92)
	$\mathrm{CI}^{*,s}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}} + 2$	10.8 (1.12)	11.4 (3.32)	13.2 (0.99)	14.2 (1.11)	17.0 (4.13)	21.4 (5.83)
	$\mathrm{CI}^{*,s}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}} + 1$	11.4 (1.12)	12.2 (3.30)	13.4 (0.99)	14.2 (1.11)	18.8 (4.06)	22.4 (5.77)
	$\mathrm{CI}^{*,s}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}$	11.8 (1.11)	12.4 (3.30)	13.2 (0.98)	14.4 (1.10)	18.6 (4.03)	23.0 (5.74)
	$\mathrm{CI}^{*,s}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}} - 1$	11.6 (1.11)	12.2 (3.28)	12.2 (0.99)	15.0 (1.12)	17.0 (4.06)	20.6 (5.83)
	$\mathrm{CI}^{*,s}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}} - 2$	12.6 (1.10)	11.6 (3.30)	12.4 (0.99)	14.6 (1.12)	16.6 (4.26)	19.8 (5.98)
	$\mathrm{CI}^{*,s}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}} - 3$	11.6 (1.11)	10.2 (3.31)	12.6 (0.99)	13.6 (1.13)	14.0 (4.43)	18.2 (6.20)
	$\mathrm{CI}^{*,s}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}} - 4$	12.0 (1.11)	11.8 (3.29)	12.2 (1.00)	15.0 (1.14)	15.0 (4.69)	15.2 (6.50)
	$\mathrm{CI}^{*,s}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}} - 5$	13.4 (1.10)	11.0 (3.28)	13.0 (0.99)	14.2 (1.14)	14.0 (4.71)	14.0 (6.70)
	$\mathrm{CI}^{*,s}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}} - 6$	12.8 (1.10)	10.6 (3.28)	13.2 (0.99)	14.6 (1.15)	12.6 (4.64)	13.6 (6.90)

Table 4.1: Empirical coverage and mean length ($\times 10^2$) of confidence intervals for the model parameter θ_1 and sample size $n = 50$, in brackets.

α	CI	x_1	x_2	x_3	x_4	x_5	x_6
5%	$\mathrm{CI}^{mc}_{x,\alpha}$	4.8 (0.89)	6.4 (2.68)	4.8 (0.80)	4.8 (0.91)	5.4 (3.85)	3.8 (5.43)
	$\mathrm{CI}^{mc,s}_{x,\alpha}$	4.4 (0.90)	5.0 (2.69)	4.4 (0.81)	4.4 (0.92)	6.8 (3.79)	3.8 (5.42)
	$\mathrm{CI}^{asy}_{x,\alpha}$	6.0 (0.83)	6.8 (2.47)	6.0 (0.74)	6.6 (0.83)	14.6 (2.89)	14.2 (4.13)
	$\mathrm{CI}^{*}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}+3$	6.8 (0.82)	7.2 (2.44)	7.8 (0.73)	8.0 (0.83)	7.2 (3.52)	8.2 (4.95)
	$\mathrm{CI}^{*}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}+2$	7.0 (0.82)	7.8 (2.44)	7.6 (0.73)	8.0 (0.84)	8.4 (3.42)	7.4 (4.84)
	$\mathrm{CI}^{*}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}+1$	7.4 (0.83)	7.2 (2.44)	8.2 (0.74)	7.8 (0.83)	9.2 (3.37)	8.6 (4.77)
	$\mathrm{CI}^{*}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}$	7.2 (0.83)	7.6 (2.44)	7.8 (0.74)	8.0 (0.84)	9.0 (3.32)	9.4 (4.72)
	$\mathrm{CI}^{*}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}-1$	6.0 (0.90)	6.8 (2.66)	6.2 (0.80)	6.2 (0.91)	8.2 (3.46)	8.8 (4.93)
	$\mathrm{CI}^{*}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}-2$	4.2 (1.09)	5.0 (3.20)	4.2 (0.97)	4.8 (1.09)	7.4 (4.03)	7.8 (5.82)
	$\mathrm{CI}^{*}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}-3$	1.8 (1.37)	3.0 (4.08)	2.8 (1.22)	3.6 (1.38)	5.8 (5.01)	5.4 (7.41)
	$\mathrm{CI}^{*}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}-4$	2.2 (1.69)	2.6 (5.04)	1.6 (1.50)	2.4 (1.69)	4.4 (5.96)	4.4 (9.31)
	$\mathrm{CI}^{*}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}-5$	1.4 (1.97)	2.2 (5.87)	1.2 (1.75)	1.4 (1.96)	3.4 (6.69)	3.0 (10.94)
	$\mathrm{CI}^{*}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}-6$	0.8 (2.18)	1.0 (6.56)	0.6 (1.95)	0.2 (2.19)	2.4 (7.30)	2.2 (12.30)
	$\mathrm{CI}^{*,s}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}+3$	5.0 (0.89)	5.6 (2.63)	5.2 (0.79)	6.0 (0.90)	7.4 (3.49)	7.8 (4.97)
	$\mathrm{CI}^{*,s}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}+2$	5.2 (0.88)	6.2 (2.62)	5.4 (0.79)	6.4 (0.89)	8.6 (3.45)	7.4 (4.90)
	$\mathrm{CI}^{*,s}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}+1$	5.8 (0.88)	6.2 (2.61)	5.4 (0.78)	6.0 (0.89)	9.6 (3.42)	8.8 (4.87)
	$\mathrm{CI}^{*,s}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}$	5.8 (0.88)	6.0 (2.61)	5.4 (0.78)	6.4 (0.89)	9.2 (3.40)	10.6 (4.84)
	$\mathrm{CI}^{*,s}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}-1$	5.0 (0.88)	6.2 (2.58)	6.0 (0.78)	6.0 (0.89)	9.4 (3.42)	9.8 (4.85)
	$\mathrm{CI}^{*,s}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}-2$	5.8 (0.87)	6.2 (2.59)	5.0 (0.78)	6.6 (0.89)	8.6 (3.49)	9.6 (4.94)
	$\mathrm{CI}^{*,s}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}-3$	5.6 (0.87)	6.2 (2.60)	5.8 (0.78)	6.0 (0.89)	8.0 (3.75)	7.6 (5.11)
	$\mathrm{CI}^{*,s}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}-4$	5.6 (0.87)	6.0 (2.61)	5.6 (0.78)	6.2 (0.90)	7.2 (3.76)	5.8 (5.35)
	$\mathrm{CI}^{*,s}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}-5$	5.6 (0.87)	6.4 (2.61)	5.0 (0.79)	5.6 (0.91)	7.4 (3.85)	3.8 (5.59)
	$\mathrm{CI}^{*,s}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}-6$	5.8 (0.87)	5.6 (2.62)	5.4 (0.79)	5.2 (0.92)	7.2 (3.84)	3.2 (5.82)
10%	$\mathrm{CI}^{mc}_{x,\alpha}$	9.2 (0.75)	9.0 (2.24)	9.0 (0.67)	9.8 (0.77)	10.8 (3.20)	9.0 (4.51)
	$\mathrm{CI}^{mc,s}_{x,\alpha}$	10.2 (0.75)	8.8 (2.26)	9.6 (0.68)	8.8 (0.77)	10.8 (3.17)	9.4 (4.54)
	$\mathrm{CI}^{asy}_{x,\alpha}$	13.4 (0.69)	12.8 (2.08)	13.0 (0.62)	15.0 (0.70)	22.0 (2.43)	22.6 (3.47)
	$\mathrm{CI}^{*}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}+3$	13.6 (0.69)	13.6 (2.06)	14.6 (0.62)	15.2 (0.70)	12.4 (2.94)	14.4 (4.13)
	$\mathrm{CI}^{*}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}+2$	14.2 (0.70)	13.4 (2.06)	14.4 (0.62)	15.2 (0.70)	14.2 (2.85)	16.0 (4.04)
	$\mathrm{CI}^{*}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}+1$	14.6 (0.70)	14.0 (2.06)	14.8 (0.62)	15.8 (0.70)	16.4 (2.80)	18.2 (3.96)
	$\mathrm{CI}^{*}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}$	13.8 (0.70)	14.0 (2.06)	14.8 (0.62)	15.8 (0.70)	17.0 (2.76)	18.2 (3.91)
	$\mathrm{CI}^{*}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}-1$	10.8 (0.76)	12.2 (2.25)	11.8 (0.68)	12.0 (0.76)	16.4 (2.86)	17.4 (4.06)
	$\mathrm{CI}^{*}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}-2$	8.6 (0.92)	10.0 (2.70)	8.4 (0.82)	9.0 (0.92)	13.6 (3.31)	14.0 (4.78)
	$\mathrm{CI}^{*}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}-3$	6.8 (1.16)	5.8 (3.45)	5.8 (1.03)	6.8 (1.16)	10.6 (4.09)	10.2 (6.04)
	$\mathrm{CI}^{*}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}-4$	5.4 (1.43)	4.4 (4.25)	4.2 (1.27)	5.2 (1.42)	8.6 (4.82)	7.4 (7.50)
	$\mathrm{CI}^{*}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}-5$	3.6 (1.66)	3.2 (4.96)	3.2 (1.47)	3.6 (1.65)	5.6 (5.38)	4.8 (8.76)
	$\mathrm{CI}^{*}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}-6$	2.0 (1.84)	1.6 (5.54)	2.0 (1.64)	1.8 (1.84)	3.8 (5.82)	3.6 (9.83)
	$\mathrm{CI}^{*,s}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}+3$	11.2 (0.75)	10.6 (2.21)	12.2 (0.67)	12.2 (0.75)	13.2 (2.95)	14.6 (4.17)
	$\mathrm{CI}^{*,s}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}+2$	12.0 (0.74)	11.0 (2.20)	11.8 (0.66)	11.8 (0.75)	14.2 (2.91)	16.4 (4.12)
	$\mathrm{CI}^{*,s}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}+1$	11.6 (0.74)	11.2 (2.20)	11.8 (0.66)	12.8 (0.75)	15.0 (2.88)	16.6 (4.09)
	$\mathrm{CI}^{*,s}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}$	11.0 (0.74)	11.0 (2.19)	11.8 (0.66)	12.6 (0.75)	16.6 (2.85)	18.6 (4.06)
	$\mathrm{CI}^{*,s}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}-1$	11.4 (0.74)	11.0 (2.18)	12.0 (0.66)	12.2 (0.75)	16.6 (2.88)	17.2 (4.08)
	$\mathrm{CI}^{*,s}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}-2$	11.0 (0.73)	11.0 (2.18)	12.2 (0.66)	12.4 (0.75)	15.0 (2.93)	13.6 (4.15)
	$\mathrm{CI}^{*,s}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}-3$	11.6 (0.73)	10.8 (2.18)	12.0 (0.66)	12.4 (0.75)	13.8 (3.13)	11.6 (4.29)
	$\mathrm{CI}^{*,s}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}-4$	11.4 (0.74)	10.6 (2.19)	11.2 (0.66)	12.4 (0.76)	13.2 (3.16)	10.0 (4.49)
	$\mathrm{CI}^{*,s}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}-5$	11.2 (0.73)	10.8 (2.20)	12.4 (0.66)	12.4 (0.77)	11.6 (3.25)	8.0 (4.69)
	$\mathrm{CI}^{*,s}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}-6$	12.2 (0.73)	10.2 (2.20)	13.0 (0.66)	10.8 (0.78)	12.6 (3.25)	6.6 (4.87)

Table 4.2: Empirical coverage and mean length ($\times 10^2$) of confidence intervals for the model parameter θ_1 and sample size $n = 100$, in brackets.

α	CI	x_1	x_2	x_3	x_4	x_5	x_6
5%	$\mathrm{CI}^{mc}_{x,\alpha}$	4.2 (0.61)	4.8 (1.83)	4.4 (0.55)	3.4 (0.63)	4.2 (2.73)	4.6 (3.83)
	$\mathrm{CI}^{mc,s}_{x,\alpha}$	4.2 (0.61)	4.8 (1.84)	4.2 (0.55)	4.8 (0.63)	5.4 (2.70)	4.4 (3.81)
	$\mathrm{CI}^{asy}_{x,\alpha}$	5.0 (0.59)	5.6 (1.77)	5.6 (0.53)	5.4 (0.60)	10.6 (2.20)	9.2 (3.10)
	$\mathrm{CI}^{*}_{x,\alpha},\hat{k}_n^{\mathrm{GCV}}+3$	5.2 (0.58)	6.6 (1.75)	6.0 (0.52)	5.0 (0.60)	6.0 (2.63)	5.8 (3.70)
	$\mathrm{CI}^{*}_{x,\alpha},\hat{k}_n^{\mathrm{GCV}}+2$	5.2 (0.58)	6.4 (1.75)	6.0 (0.52)	5.0 (0.60)	6.8 (2.58)	6.0 (3.61)
	$\mathrm{CI}^{*}_{x,\alpha},\hat{k}_n^{\mathrm{GCV}}+1$	5.2 (0.58)	6.4 (1.75)	6.0 (0.52)	5.8 (0.60)	8.0 (2.52)	6.6 (3.54)
	$\mathrm{CI}^{*}_{x,\alpha},\hat{k}_n^{\mathrm{GCV}}$	5.4 (0.58)	6.2 (1.76)	6.4 (0.52)	4.8 (0.60)	7.8 (2.49)	7.0 (3.50)
	$\mathrm{CI}^{*}_{x,\alpha},\hat{k}_n^{\mathrm{GCV}}-1$	4.4 (0.60)	5.8 (1.82)	5.8 (0.54)	5.2 (0.62)	8.4 (2.47)	7.6 (3.46)
	$\mathrm{CI}^{*}_{x,\alpha},\hat{k}_n^{\mathrm{GCV}}-2$	4.0 (0.67)	4.4 (2.01)	5.0 (0.60)	4.2 (0.68)	8.0 (2.58)	8.0 (3.68)
	$\mathrm{CI}^{*}_{x,\alpha},\hat{k}_n^{\mathrm{GCV}}-3$	3.2 (0.83)	3.2 (2.49)	3.6 (0.74)	3.4 (0.84)	6.6 (3.12)	7.4 (4.58)
	$\mathrm{CI}^{*}_{x,\alpha},\hat{k}_n^{\mathrm{GCV}}-4$	2.4 (1.04)	1.4 (3.14)	3.4 (0.93)	2.4 (1.05)	6.2 (3.80)	6.4 (5.83)
	$\mathrm{CI}^{*}_{x,\alpha},\hat{k}_n^{\mathrm{GCV}}-5$	2.2 (1.28)	0.6 (3.86)	2.2 (1.14)	1.6 (1.29)	4.6 (4.55)	4.6 (7.31)
	$\mathrm{CI}^{*}_{x,\alpha},\hat{k}_n^{\mathrm{GCV}}-6$	1.6 (1.46)	0.6 (4.39)	1.6 (1.30)	0.8 (1.47)	4.2 (5.01)	4.2 (8.44)
	$\mathrm{CI}^{*,s}_{x,\alpha},\hat{k}_n^{\mathrm{GCV}}+3$	4.2 (0.60)	5.6 (1.81)	5.4 (0.54)	5.2 (0.62)	8.2 (2.57)	6.6 (3.61)
	$\mathrm{CI}^{*,s}_{x,\alpha},\hat{k}_n^{\mathrm{GCV}}+2$	4.6 (0.60)	5.4 (1.81)	6.0 (0.54)	4.8 (0.62)	8.0 (2.55)	6.2 (3.58)
	$\mathrm{CI}^{*,s}_{x,\alpha},\hat{k}_n^{\mathrm{GCV}}+1$	4.6 (0.60)	5.8 (1.82)	5.6 (0.54)	4.8 (0.62)	8.4 (2.53)	7.2 (3.53)
	$\mathrm{CI}^{*,s}_{x,\alpha},\hat{k}_n^{\mathrm{GCV}}$	4.4 (0.60)	6.0 (1.81)	6.0 (0.54)	5.0 (0.62)	8.2 (2.51)	7.0 (3.52)
	$\mathrm{CI}^{*,s}_{x,\alpha},\hat{k}_n^{\mathrm{GCV}}-1$	4.4 (0.60)	6.0 (1.81)	5.2 (0.54)	5.0 (0.62)	8.8 (2.52)	6.6 (3.52)
	$\mathrm{CI}^{*,s}_{x,\alpha},\hat{k}_n^{\mathrm{GCV}}-2$	4.8 (0.60)	5.6 (1.81)	5.6 (0.54)	4.8 (0.62)	9.2 (2.49)	6.4 (3.53)
	$\mathrm{CI}^{*,s}_{x,\alpha},\hat{k}_n^{\mathrm{GCV}}-3$	5.0 (0.60)	5.8 (1.81)	6.0 (0.54)	4.8 (0.62)	7.6 (2.59)	6.4 (3.63)
	$\mathrm{CI}^{*,s}_{x,\alpha},\hat{k}_n^{\mathrm{GCV}}-4$	4.2 (0.60)	5.6 (1.80)	5.6 (0.55)	5.4 (0.63)	7.8 (2.67)	6.2 (3.78)
	$\mathrm{CI}^{*,s}_{x,\alpha},\hat{k}_n^{\mathrm{GCV}}-5$	5.0 (0.60)	5.2 (1.81)	5.6 (0.54)	5.2 (0.63)	6.0 (2.81)	5.2 (3.98)
	$\mathrm{CI}^{*,s}_{x,\alpha},\hat{k}_n^{\mathrm{GCV}}-6$	4.0 (0.60)	5.2 (1.81)	5.8 (0.55)	5.0 (0.64)	6.4 (2.73)	4.4 (4.16)
10%	$\mathrm{CI}^{mc}_{x,\alpha}$	8.0 (0.51)	9.4 (1.53)	8.4 (0.46)	10.4 (0.53)	9.4 (2.27)	9.6 (3.19)
	$\mathrm{CI}^{mc,s}_{x,\alpha}$	8.6 (0.51)	9.0 (1.55)	8.8 (0.46)	9.0 (0.53)	9.0 (2.27)	8.6 (3.20)
	$\mathrm{CI}^{asy}_{x,\alpha}$	9.4 (0.49)	10.6 (1.49)	10.4 (0.44)	11.4 (0.50)	19.2 (1.84)	15.0 (2.60)
	$\mathrm{CI}^{*}_{x,\alpha},\hat{k}_n^{\mathrm{GCV}}+3$	9.6 (0.49)	10.6 (1.48)	10.8 (0.44)	10.0 (0.51)	11.2 (2.20)	9.6 (3.09)
	$\mathrm{CI}^{*}_{x,\alpha},\hat{k}_n^{\mathrm{GCV}}+2$	9.4 (0.49)	10.8 (1.47)	11.4 (0.44)	10.8 (0.51)	13.0 (2.15)	10.6 (3.01)
	$\mathrm{CI}^{*}_{x,\alpha},\hat{k}_n^{\mathrm{GCV}}+1$	9.2 (0.49)	11.4 (1.48)	11.0 (0.44)	10.2 (0.51)	12.4 (2.10)	12.2 (2.95)
	$\mathrm{CI}^{*}_{x,\alpha},\hat{k}_n^{\mathrm{GCV}}$	10.0 (0.49)	10.0 (1.48)	10.8 (0.44)	10.6 (0.50)	14.2 (2.07)	12.0 (2.91)
	$\mathrm{CI}^{*}_{x,\alpha},\hat{k}_n^{\mathrm{GCV}}-1$	9.0 (0.51)	9.2 (1.53)	10.0 (0.46)	9.8 (0.52)	15.8 (2.04)	13.0 (2.87)
	$\mathrm{CI}^{*}_{x,\alpha},\hat{k}_n^{\mathrm{GCV}}-2$	8.0 (0.56)	9.2 (1.69)	8.6 (0.50)	8.4 (0.57)	14.0 (2.12)	12.4 (3.04)
	$\mathrm{CI}^{*}_{x,\alpha},\hat{k}_n^{\mathrm{GCV}}-3$	6.0 (0.70)	7.2 (2.10)	7.2 (0.63)	6.0 (0.71)	10.8 (2.55)	10.8 (3.74)
	$\mathrm{CI}^{*}_{x,\alpha},\hat{k}_n^{\mathrm{GCV}}-4$	5.2 (0.88)	3.2 (2.64)	5.2 (0.78)	4.2 (0.89)	10.0 (3.09)	8.0 (4.72)
	$\mathrm{CI}^{*}_{x,\alpha},\hat{k}_n^{\mathrm{GCV}}-5$	3.6 (1.08)	2.2 (3.25)	4.2 (0.96)	3.2 (1.09)	7.8 (3.67)	6.6 (5.86)
	$\mathrm{CI}^{*}_{x,\alpha},\hat{k}_n^{\mathrm{GCV}}-6$	2.4 (1.23)	1.8 (3.70)	3.4 (1.10)	1.8 (1.23)	6.4 (4.01)	5.4 (6.75)
	$\mathrm{CI}^{*,s}_{x,\alpha},\hat{k}_n^{\mathrm{GCV}}+3$	8.0 (0.51)	9.0 (1.53)	9.4 (0.46)	9.4 (0.52)	12.6 (2.16)	10.8 (3.04)
	$\mathrm{CI}^{*,s}_{x,\alpha},\hat{k}_n^{\mathrm{GCV}}+2$	8.8 (0.51)	9.2 (1.53)	9.4 (0.46)	10.0 (0.52)	13.8 (2.14)	10.4 (3.01)
	$\mathrm{CI}^{*,s}_{x,\alpha},\hat{k}_n^{\mathrm{GCV}}+1$	9.0 (0.51)	8.6 (1.53)	9.4 (0.46)	9.2 (0.52)	14.4 (2.12)	13.8 (2.98)
	$\mathrm{CI}^{*,s}_{x,\alpha},\hat{k}_n^{\mathrm{GCV}}$	8.8 (0.51)	9.2 (1.52)	9.6 (0.45)	9.8 (0.52)	15.4 (2.11)	12.8 (2.95)
	$\mathrm{CI}^{*,s}_{x,\alpha},\hat{k}_n^{\mathrm{GCV}}-1$	9.0 (0.51)	9.2 (1.53)	9.6 (0.45)	9.6 (0.52)	13.0 (2.12)	12.6 (2.96)
	$\mathrm{CI}^{*,s}_{x,\alpha},\hat{k}_n^{\mathrm{GCV}}-2$	9.2 (0.50)	9.8 (1.52)	11.0 (0.45)	9.8 (0.52)	14.4 (2.10)	11.6 (2.98)
	$\mathrm{CI}^{*,s}_{x,\alpha},\hat{k}_n^{\mathrm{GCV}}-3$	9.8 (0.51)	9.2 (1.52)	9.4 (0.46)	9.0 (0.53)	12.2 (2.16)	10.6 (3.05)
	$\mathrm{CI}^{*,s}_{x,\alpha},\hat{k}_n^{\mathrm{GCV}}-4$	9.6 (0.51)	9.8 (1.52)	10.0 (0.46)	9.8 (0.53)	10.4 (2.24)	9.0 (3.19)
	$\mathrm{CI}^{*,s}_{x,\alpha},\hat{k}_n^{\mathrm{GCV}}-5$	9.2 (0.51)	9.8 (1.52)	10.2 (0.46)	10.0 (0.53)	10.4 (2.33)	7.6 (3.35)
	$\mathrm{CI}^{*,s}_{x,\alpha},\hat{k}_n^{\mathrm{GCV}}-6$	10.0 (0.51)	10.2 (1.52)	10.2 (0.46)	9.4 (0.54)	9.4 (2.31)	7.0 (3.50)

Table 4.3: Empirical coverage and mean length ($\times 10^2$) of confidence intervals for the model parameter θ_1 and sample size $n = 200$, in brackets.

α	CI	x_1	x_2	x_3	x_4	x_5	x_6
5%	$\mathrm{CI}^{mc}_{x,\alpha}$	5.4 (0.38)	4.4 (1.13)	5.0 (0.34)	5.2 (0.39)	3.8 (1.75)	5.8 (2.47)
	$\mathrm{CI}^{mc,s}_{x,\alpha}$	5.4 (0.38)	3.8 (1.13)	4.6 (0.34)	6.0 (0.39)	5.4 (1.74)	5.2 (2.44)
	$\mathrm{CI}^{asy}_{x,\alpha}$	5.8 (0.37)	4.4 (1.12)	6.4 (0.33)	6.8 (0.38)	12.8 (1.49)	10.2 (2.09)
	$\mathrm{CI}^{*}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}+3$	6.6 (0.37)	5.6 (1.11)	5.6 (0.33)	6.8 (0.38)	7.4 (1.76)	6.0 (2.47)
	$\mathrm{CI}^{*}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}+2$	6.4 (0.37)	5.2 (1.11)	5.6 (0.33)	6.8 (0.38)	8.4 (1.72)	6.4 (2.42)
	$\mathrm{CI}^{*}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}+1$	6.0 (0.37)	5.6 (1.11)	5.2 (0.33)	6.6 (0.38)	7.8 (1.69)	6.0 (2.38)
	$\mathrm{CI}^{*}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}$	6.2 (0.37)	4.6 (1.11)	6.0 (0.33)	6.6 (0.38)	9.2 (1.67)	6.6 (2.35)
	$\mathrm{CI}^{*}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}-1$	6.6 (0.37)	4.4 (1.12)	5.4 (0.34)	6.4 (0.38)	11.0 (1.61)	7.2 (2.26)
	$\mathrm{CI}^{*}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}-2$	6.2 (0.39)	4.2 (1.15)	5.4 (0.35)	6.2 (0.39)	11.4 (1.57)	8.6 (2.22)
	$\mathrm{CI}^{*}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}-3$	5.0 (0.42)	4.0 (1.27)	4.0 (0.38)	4.6 (0.43)	9.0 (1.67)	8.4 (2.38)
	$\mathrm{CI}^{*}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}-4$	4.4 (0.53)	3.6 (1.60)	3.2 (0.48)	3.8 (0.54)	8.0 (2.03)	7.2 (2.98)
	$\mathrm{CI}^{*}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}-5$	2.8 (0.68)	2.8 (2.05)	2.8 (0.61)	2.2 (0.69)	5.6 (2.58)	6.2 (3.89)
	$\mathrm{CI}^{*}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}-6$	2.0 (0.81)	2.2 (2.42)	1.4 (0.72)	1.4 (0.82)	5.2 (2.88)	5.6 (4.68)
	$\mathrm{CI}^{*,s}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}+3$	6.0 (0.38)	5.2 (1.12)	5.0 (0.34)	6.4 (0.39)	8.8 (1.70)	7.0 (2.39)
	$\mathrm{CI}^{*,s}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}+2$	6.0 (0.38)	5.2 (1.12)	4.6 (0.34)	6.4 (0.39)	9.4 (1.69)	7.6 (2.38)
	$\mathrm{CI}^{*,s}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}+1$	6.6 (0.38)	4.6 (1.13)	5.6 (0.34)	6.2 (0.39)	9.2 (1.68)	6.8 (2.35)
	$\mathrm{CI}^{*,s}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}$	6.2 (0.38)	4.6 (1.12)	5.4 (0.34)	6.4 (0.39)	10.2 (1.67)	7.6 (2.34)
	$\mathrm{CI}^{*,s}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}-1$	6.2 (0.38)	5.0 (1.12)	5.4 (0.34)	7.0 (0.39)	9.2 (1.67)	6.8 (2.34)
	$\mathrm{CI}^{*,s}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}-2$	6.2 (0.38)	4.6 (1.12)	5.2 (0.34)	7.4 (0.39)	10.4 (1.66)	6.6 (2.33)
	$\mathrm{CI}^{*,s}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}-3$	6.4 (0.37)	5.0 (1.12)	5.0 (0.34)	6.8 (0.39)	8.8 (1.67)	5.8 (2.35)
	$\mathrm{CI}^{*,s}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}-4$	7.2 (0.37)	4.6 (1.12)	6.4 (0.34)	6.0 (0.39)	8.8 (1.70)	5.2 (2.42)
	$\mathrm{CI}^{*,s}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}-5$	6.0 (0.37)	4.8 (1.12)	5.2 (0.34)	5.6 (0.39)	8.6 (1.76)	4.6 (2.52)
	$\mathrm{CI}^{*,s}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}-6$	5.6 (0.37)	4.2 (1.12)	5.4 (0.34)	4.8 (0.40)	7.6 (1.78)	4.4 (2.65)
10%	$\mathrm{CI}^{mc}_{x,\alpha}$	11.4 (0.32)	9.8 (0.95)	11.0 (0.29)	10.6 (0.33)	9.8 (1.46)	11.0 (2.06)
	$\mathrm{CI}^{mc,s}_{x,\alpha}$	10.8 (0.32)	10.6 (0.95)	10.2 (0.29)	10.2 (0.33)	10.8 (1.46)	11.4 (2.06)
	$\mathrm{CI}^{asy}_{x,\alpha}$	11.2 (0.31)	10.0 (0.94)	12.0 (0.28)	11.4 (0.32)	18.2 (1.25)	17.8 (1.76)
	$\mathrm{CI}^{*}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}+3$	11.4 (0.31)	10.2 (0.93)	11.8 (0.28)	11.4 (0.32)	12.6 (1.47)	10.8 (2.07)
	$\mathrm{CI}^{*}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}+2$	11.6 (0.31)	10.4 (0.93)	11.2 (0.28)	11.0 (0.32)	14.0 (1.44)	11.6 (2.02)
	$\mathrm{CI}^{*}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}+1$	11.0 (0.31)	10.2 (0.93)	12.0 (0.28)	12.2 (0.32)	14.2 (1.41)	12.6 (1.98)
	$\mathrm{CI}^{*}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}$	11.4 (0.31)	10.6 (0.93)	10.8 (0.28)	11.0 (0.32)	14.8 (1.39)	13.4 (1.96)
	$\mathrm{CI}^{*}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}-1$	11.2 (0.31)	10.8 (0.94)	11.6 (0.28)	11.4 (0.32)	17.0 (1.33)	15.4 (1.88)
	$\mathrm{CI}^{*}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}-2$	10.6 (0.32)	9.2 (0.97)	10.8 (0.29)	10.2 (0.33)	18.0 (1.29)	16.2 (1.84)
	$\mathrm{CI}^{*}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}-3$	9.4 (0.36)	8.8 (1.07)	9.0 (0.32)	8.8 (0.36)	17.4 (1.37)	15.8 (1.96)
	$\mathrm{CI}^{*}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}-4$	7.8 (0.45)	6.6 (1.35)	7.4 (0.40)	6.6 (0.46)	13.4 (1.66)	12.4 (2.43)
	$\mathrm{CI}^{*}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}-5$	5.4 (0.57)	5.6 (1.73)	5.4 (0.51)	4.6 (0.58)	10.8 (2.09)	9.2 (3.14)
	$\mathrm{CI}^{*}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}-6$	3.6 (0.68)	4.2 (2.04)	4.0 (0.61)	3.8 (0.69)	8.6 (2.32)	7.4 (3.75)
	$\mathrm{CI}^{*,s}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}+3$	11.4 (0.32)	9.6 (0.95)	11.8 (0.28)	11.2 (0.33)	13.8 (1.43)	12.6 (2.01)
	$\mathrm{CI}^{*,s}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}+2$	11.0 (0.32)	9.8 (0.95)	11.2 (0.28)	11.4 (0.32)	15.4 (1.42)	13.0 (1.99)
	$\mathrm{CI}^{*,s}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}+1$	11.0 (0.32)	9.2 (0.95)	11.2 (0.28)	10.8 (0.32)	14.6 (1.41)	14.4 (1.98)
	$\mathrm{CI}^{*,s}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}$	11.6 (0.32)	9.6 (0.95)	11.6 (0.28)	11.2 (0.32)	15.0 (1.40)	13.8 (1.97)
	$\mathrm{CI}^{*,s}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}-1$	11.6 (0.32)	9.8 (0.95)	12.0 (0.28)	11.6 (0.32)	15.8 (1.40)	14.6 (1.97)
	$\mathrm{CI}^{*,s}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}-2$	10.6 (0.32)	9.6 (0.95)	11.4 (0.28)	11.2 (0.32)	14.4 (1.39)	13.2 (1.96)
	$\mathrm{CI}^{*,s}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}-3$	11.2 (0.31)	10.4 (0.94)	12.2 (0.28)	11.8 (0.33)	14.6 (1.40)	12.6 (1.97)
	$\mathrm{CI}^{*,s}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}-4$	10.8 (0.32)	10.0 (0.94)	11.0 (0.28)	11.4 (0.33)	14.2 (1.43)	11.2 (2.03)
	$\mathrm{CI}^{*,s}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}-5$	11.4 (0.31)	9.6 (0.94)	10.8 (0.28)	11.4 (0.33)	13.2 (1.48)	9.4 (2.12)
	$\mathrm{CI}^{*,s}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}-6$	11.6 (0.31)	9.2 (0.95)	11.6 (0.29)	11.2 (0.33)	11.8 (1.51)	7.6 (2.23)

Table 4.4: Empirical coverage and mean length ($\times 10^2$) of confidence intervals for the model parameter θ_1 and sample size $n = 500$, in brackets.

α	CI	x_1	x_2	x_3	x_4	x_5	x_6
5%	$\mathrm{CI}^{mc}_{x,\alpha}$	5.6 (1.84)	5.2 (5.64)	4.6 (1.64)	5.2 (1.85)	5.4 (7.27)	5.8 (8.58)
	$\mathrm{CI}^{mc,s}_{x,\alpha}$	4.8 (1.91)	4.6 (5.71)	5.4 (1.69)	6.0 (1.87)	5.2 (8.45)	4.4 (8.35)
	$\mathrm{CI}^{asy}_{x,\alpha}$	9.0 (1.66)	8.4 (4.95)	9.0 (1.46)	10.2 (1.59)	48.8 (4.42)	15.0 (6.18)
	$\mathrm{CI}^{*}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}+3$	10.8 (1.65)	9.6 (4.88)	8.4 (1.47)	10.4 (1.65)	28.6 (6.11)	8.4 (8.39)
	$\mathrm{CI}^{*}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}+2$	10.6 (1.66)	10.2 (4.90)	8.6 (1.47)	10.4 (1.65)	30.6 (5.96)	7.8 (8.21)
	$\mathrm{CI}^{*}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}+1$	10.2 (1.66)	9.4 (4.93)	9.2 (1.47)	10.0 (1.65)	33.6 (5.80)	9.0 (8.03)
	$\mathrm{CI}^{*}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}$	10.6 (1.66)	8.8 (4.93)	8.8 (1.47)	10.6 (1.64)	34.2 (5.76)	8.6 (7.95)
	$\mathrm{CI}^{*}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}-1$	8.0 (1.84)	6.8 (5.42)	6.8 (1.63)	7.2 (1.80)	29.6 (6.27)	7.8 (8.28)
	$\mathrm{CI}^{*}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}-2$	6.6 (1.94)	5.8 (5.78)	5.8 (1.72)	6.6 (1.90)	26.6 (6.74)	7.4 (8.43)
	$\mathrm{CI}^{*}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}-3$	5.6 (2.00)	5.8 (5.96)	4.6 (1.77)	5.8 (1.95)	24.6 (7.01)	6.8 (8.50)
	$\mathrm{CI}^{*}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}-4$	5.0 (2.04)	4.8 (6.05)	5.0 (1.80)	5.6 (1.99)	25.0 (7.16)	5.8 (8.56)
	$\mathrm{CI}^{*}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}-5$	4.4 (2.07)	4.4 (6.10)	4.4 (1.83)	5.4 (2.01)	24.8 (7.30)	6.2 (8.61)
	$\mathrm{CI}^{*}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}-6$	3.8 (2.08)	4.4 (6.15)	4.6 (1.84)	4.6 (2.03)	24.6 (7.35)	6.2 (8.59)
	$\mathrm{CI}^{*,s}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}+3$	6.8 (1.86)	6.0 (5.53)	5.8 (1.65)	7.4 (1.83)	25.4 (6.84)	6.4 (8.32)
	$\mathrm{CI}^{*,s}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}+2$	6.4 (1.86)	4.8 (5.50)	5.4 (1.64)	7.6 (1.82)	26.0 (6.77)	6.8 (8.25)
	$\mathrm{CI}^{*,s}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}+1$	6.6 (1.85)	6.6 (5.49)	6.6 (1.63)	7.6 (1.81)	26.2 (6.68)	8.0 (8.13)
	$\mathrm{CI}^{*,s}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}$	7.2 (1.84)	5.8 (5.48)	6.6 (1.62)	8.0 (1.80)	28.2 (6.62)	9.4 (8.05)
	$\mathrm{CI}^{*,s}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}-1$	9.0 (1.83)	6.8 (5.44)	6.8 (1.62)	8.0 (1.81)	25.6 (6.79)	7.8 (8.05)
	$\mathrm{CI}^{*,s}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}-2$	7.8 (1.82)	6.6 (5.44)	6.0 (1.62)	7.4 (1.81)	26.4 (6.79)	6.8 (7.99)
	$\mathrm{CI}^{*,s}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}-3$	8.4 (1.82)	6.2 (5.45)	6.0 (1.61)	8.0 (1.81)	24.2 (6.92)	6.6 (7.92)
	$\mathrm{CI}^{*,s}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}-4$	8.2 (1.82)	6.8 (5.45)	6.8 (1.62)	7.8 (1.82)	23.2 (6.83)	7.4 (7.87)
	$\mathrm{CI}^{*,s}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}-5$	8.4 (1.81)	6.2 (5.44)	6.2 (1.62)	7.2 (1.82)	23.6 (6.84)	7.0 (7.84)
	$\mathrm{CI}^{*,s}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}-6$	7.6 (1.80)	6.0 (5.43)	6.8 (1.61)	8.0 (1.81)	24.8 (6.78)	7.2 (7.80)
10%	$\mathrm{CI}^{mc}_{x,\alpha}$	11.2 (1.55)	10.4 (4.68)	10.0 (1.38)	11.0 (1.54)	10.4 (5.95)	12.0 (6.97)
	$\mathrm{CI}^{mc,s}_{x,\alpha}$	10.8 (1.59)	8.6 (4.75)	10.4 (1.40)	11.4 (1.56)	10.0 (6.94)	11.6 (6.98)
	$\mathrm{CI}^{asy}_{x,\alpha}$	16.6 (1.39)	14.2 (4.16)	15.8 (1.22)	17.4 (1.33)	56.2 (3.71)	25.2 (5.19)
	$\mathrm{CI}^{*}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}+3$	17.6 (1.39)	15.6 (4.13)	14.8 (1.24)	15.6 (1.40)	39.6 (5.06)	13.8 (6.92)
	$\mathrm{CI}^{*}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}+2$	17.8 (1.40)	15.6 (4.14)	15.0 (1.24)	15.0 (1.39)	41.8 (4.89)	14.0 (6.75)
	$\mathrm{CI}^{*}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}+1$	18.0 (1.40)	16.6 (4.17)	14.8 (1.24)	15.2 (1.39)	45.4 (4.73)	16.6 (6.56)
	$\mathrm{CI}^{*}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}$	17.6 (1.41)	16.6 (4.18)	15.6 (1.24)	15.4 (1.38)	45.6 (4.68)	18.0 (6.44)
	$\mathrm{CI}^{*}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}-1$	14.6 (1.56)	13.2 (4.59)	13.6 (1.38)	13.0 (1.52)	41.0 (5.07)	14.8 (6.61)
	$\mathrm{CI}^{*}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}-2$	12.4 (1.65)	10.2 (4.88)	11.6 (1.45)	12.6 (1.60)	39.0 (5.40)	13.8 (6.66)
	$\mathrm{CI}^{*}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}-3$	11.4 (1.70)	8.4 (5.03)	9.6 (1.49)	11.6 (1.64)	37.6 (5.59)	14.2 (6.67)
	$\mathrm{CI}^{*}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}-4$	9.6 (1.73)	8.2 (5.12)	10.2 (1.52)	10.6 (1.67)	36.8 (5.70)	13.0 (6.68)
	$\mathrm{CI}^{*}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}-5$	8.6 (1.75)	7.8 (5.17)	8.6 (1.54)	10.0 (1.69)	35.8 (5.79)	12.8 (6.72)
	$\mathrm{CI}^{*}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}-6$	8.2 (1.76)	7.6 (5.20)	8.2 (1.55)	9.4 (1.70)	36.2 (5.87)	13.2 (6.72)
	$\mathrm{CI}^{*,s}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}+3$	11.4 (1.57)	10.8 (4.64)	10.8 (1.39)	11.4 (1.54)	35.0 (5.48)	13.0 (6.97)
	$\mathrm{CI}^{*,s}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}+2$	13.6 (1.56)	11.2 (4.62)	11.0 (1.38)	11.4 (1.52)	36.6 (5.41)	13.8 (6.92)
	$\mathrm{CI}^{*,s}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}+1$	13.2 (1.55)	11.6 (4.61)	11.8 (1.37)	12.4 (1.52)	40.0 (5.32)	15.8 (6.78)
	$\mathrm{CI}^{*,s}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}$	14.8 (1.54)	12.2 (4.59)	11.4 (1.36)	13.0 (1.51)	40.0 (5.28)	16.6 (6.72)
	$\mathrm{CI}^{*,s}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}-1$	14.4 (1.54)	11.2 (4.57)	12.4 (1.36)	12.4 (1.52)	35.4 (5.53)	15.6 (6.75)
	$\mathrm{CI}^{*,s}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}-2$	14.4 (1.53)	12.8 (4.57)	11.4 (1.36)	11.4 (1.52)	33.8 (5.62)	14.0 (6.69)
	$\mathrm{CI}^{*,s}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}-3$	14.2 (1.53)	11.0 (4.58)	12.8 (1.36)	12.4 (1.52)	32.4 (5.69)	14.8 (6.63)
	$\mathrm{CI}^{*,s}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}-4$	14.2 (1.52)	11.4 (4.57)	12.4 (1.36)	12.8 (1.53)	31.2 (5.68)	15.0 (6.58)
	$\mathrm{CI}^{*,s}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}-5$	14.2 (1.52)	11.4 (4.57)	12.6 (1.36)	12.8 (1.53)	31.2 (5.73)	14.4 (6.58)
	$\mathrm{CI}^{*,s}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}-6$	14.8 (1.51)	11.0 (4.55)	13.0 (1.35)	13.2 (1.52)	31.8 (5.73)	14.8 (6.54)

Table 4.5: Empirical coverage and mean length ($\times 10^2$) of confidence intervals for the model parameter θ_2 and sample size $n = 50$, in brackets.

α	CI	x_1	x_2	x_3	x_4	x_5	x_6
5%	$\mathrm{CI}^{mc}_{x,\alpha}$	7.0 (1.26)	5.0 (3.83)	5.8 (1.13)	4.0 (1.28)	3.8 (5.55)	3.8 (6.32)
	$\mathrm{CI}^{mc,s}_{x,\alpha}$	6.6 (1.27)	4.4 (3.82)	5.6 (1.13)	4.8 (1.27)	4.6 (6.34)	3.8 (6.08)
	$\mathrm{CI}^{asy}_{x,\alpha}$	8.2 (1.19)	5.6 (3.56)	7.6 (1.05)	7.0 (1.16)	53.2 (3.46)	11.0 (4.77)
	$\mathrm{CI}^{*}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}+3$	9.0 (1.18)	6.0 (3.56)	6.8 (1.05)	6.2 (1.20)	34.6 (4.67)	4.6 (6.30)
	$\mathrm{CI}^{*}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}+2$	8.0 (1.18)	6.4 (3.56)	6.4 (1.06)	6.4 (1.20)	37.2 (4.50)	5.0 (6.10)
	$\mathrm{CI}^{*}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}+1$	9.4 (1.18)	6.4 (3.57)	7.0 (1.06)	6.4 (1.20)	42.0 (4.35)	6.0 (5.95)
	$\mathrm{CI}^{*}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}$	8.6 (1.19)	6.4 (3.57)	7.0 (1.06)	6.2 (1.20)	42.2 (4.28)	6.0 (5.85)
	$\mathrm{CI}^{*}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}-1$	6.6 (1.26)	5.8 (3.78)	6.0 (1.12)	5.4 (1.25)	43.2 (4.37)	4.8 (5.81)
	$\mathrm{CI}^{*}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}-2$	5.8 (1.34)	4.2 (4.00)	5.2 (1.18)	4.4 (1.32)	37.4 (4.66)	4.8 (5.91)
	$\mathrm{CI}^{*}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}-3$	4.8 (1.40)	4.4 (4.17)	4.2 (1.24)	3.4 (1.37)	35.0 (4.90)	5.8 (6.02)
	$\mathrm{CI}^{*}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}-4$	4.4 (1.43)	4.0 (4.26)	3.6 (1.26)	3.8 (1.39)	33.8 (5.04)	5.4 (6.01)
	$\mathrm{CI}^{*}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}-5$	3.8 (1.46)	3.4 (4.34)	3.8 (1.28)	3.4 (1.41)	33.8 (5.16)	5.0 (6.05)
	$\mathrm{CI}^{*}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}-6$	3.8 (1.46)	3.2 (4.38)	4.0 (1.29)	4.0 (1.43)	32.6 (5.20)	5.6 (6.08)
	$\mathrm{CI}^{*,s}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}+3$	7.2 (1.25)	5.2 (3.77)	5.8 (1.12)	5.8 (1.25)	32.8 (4.74)	6.6 (5.95)
	$\mathrm{CI}^{*,s}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}+2$	7.2 (1.25)	5.4 (3.75)	6.2 (1.11)	5.4 (1.25)	35.2 (4.63)	5.4 (5.90)
	$\mathrm{CI}^{*,s}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}+1$	7.2 (1.25)	5.4 (3.75)	6.8 (1.11)	6.0 (1.25)	37.8 (4.53)	6.8 (5.85)
	$\mathrm{CI}^{*,s}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}$	7.8 (1.25)	5.4 (3.74)	6.8 (1.11)	6.2 (1.24)	39.2 (4.50)	8.4 (5.79)
	$\mathrm{CI}^{*,s}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}-1$	8.2 (1.24)	5.2 (3.73)	6.4 (1.10)	5.4 (1.24)	41.8 (4.49)	6.0 (5.82)
	$\mathrm{CI}^{*,s}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}-2$	7.4 (1.23)	5.2 (3.72)	6.0 (1.10)	6.2 (1.25)	34.8 (4.67)	6.0 (5.82)
	$\mathrm{CI}^{*,s}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}-3$	8.0 (1.23)	5.2 (3.70)	5.8 (1.11)	5.4 (1.25)	33.8 (4.77)	5.2 (5.79)
	$\mathrm{CI}^{*,s}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}-4$	8.2 (1.23)	5.8 (3.71)	6.8 (1.10)	5.6 (1.25)	33.2 (4.82)	5.6 (5.79)
	$\mathrm{CI}^{*,s}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}-5$	8.2 (1.23)	5.4 (3.70)	6.4 (1.10)	5.4 (1.25)	33.2 (4.90)	5.4 (5.78)
	$\mathrm{CI}^{*,s}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}-6$	7.0 (1.23)	5.2 (3.71)	6.0 (1.10)	5.6 (1.25)	32.0 (4.94)	5.6 (5.78)
10%	$\mathrm{CI}^{mc}_{x,\alpha}$	11.6 (1.06)	8.2 (3.20)	9.4 (0.94)	9.6 (1.07)	8.2 (4.59)	9.0 (5.17)
	$\mathrm{CI}^{mc,s}_{x,\alpha}$	11.4 (1.06)	8.4 (3.20)	9.6 (0.95)	8.0 (1.06)	8.0 (5.28)	8.8 (5.12)
	$\mathrm{CI}^{asy}_{x,\alpha}$	13.8 (1.00)	10.4 (2.99)	11.0 (0.88)	10.6 (0.97)	60.2 (2.91)	17.8 (4.00)
	$\mathrm{CI}^{*}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}+3$	13.0 (0.99)	9.8 (3.01)	11.8 (0.89)	10.0 (1.01)	44.6 (3.89)	9.4 (5.21)
	$\mathrm{CI}^{*}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}+2$	14.2 (1.00)	10.4 (3.01)	11.4 (0.89)	10.2 (1.01)	48.8 (3.72)	9.6 (5.05)
	$\mathrm{CI}^{*}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}+1$	13.6 (1.00)	10.2 (3.01)	10.6 (0.89)	10.6 (1.01)	50.0 (3.58)	11.4 (4.90)
	$\mathrm{CI}^{*}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}$	14.0 (1.00)	10.6 (3.01)	11.4 (0.89)	10.8 (1.01)	51.4 (3.52)	12.8 (4.80)
	$\mathrm{CI}^{*}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}-1$	11.0 (1.06)	9.2 (3.19)	9.8 (0.94)	9.6 (1.05)	53.0 (3.53)	11.6 (4.71)
	$\mathrm{CI}^{*}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}-2$	9.6 (1.13)	8.6 (3.38)	8.8 (1.00)	8.0 (1.11)	52.6 (3.75)	12.2 (4.73)
	$\mathrm{CI}^{*}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}-3$	9.2 (1.18)	6.2 (3.52)	7.2 (1.04)	6.4 (1.15)	47.6 (3.93)	11.8 (4.77)
	$\mathrm{CI}^{*}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}-4$	8.8 (1.21)	6.4 (3.59)	7.4 (1.06)	6.0 (1.17)	47.6 (4.03)	12.2 (4.76)
	$\mathrm{CI}^{*}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}-5$	8.0 (1.23)	5.8 (3.66)	6.8 (1.08)	6.4 (1.19)	47.6 (4.09)	12.4 (4.77)
	$\mathrm{CI}^{*}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}-6$	8.4 (1.24)	6.2 (3.69)	7.0 (1.09)	6.0 (1.20)	47.6 (4.14)	11.4 (4.78)
	$\mathrm{CI}^{*,s}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}+3$	12.4 (1.05)	8.2 (3.16)	10.2 (0.94)	8.6 (1.05)	42.0 (3.93)	11.4 (5.01)
	$\mathrm{CI}^{*,s}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}+2$	12.4 (1.05)	8.8 (3.16)	10.0 (0.93)	8.4 (1.05)	44.6 (3.82)	10.8 (4.97)
	$\mathrm{CI}^{*,s}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}+1$	12.0 (1.05)	9.6 (3.16)	9.4 (0.93)	9.2 (1.05)	46.0 (3.72)	12.6 (4.91)
	$\mathrm{CI}^{*,s}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}$	12.4 (1.05)	8.8 (3.15)	9.8 (0.93)	9.0 (1.05)	48.2 (3.70)	12.6 (4.85)
	$\mathrm{CI}^{*,s}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}-1$	12.2 (1.04)	9.4 (3.13)	9.8 (0.93)	9.0 (1.05)	50.0 (3.77)	11.2 (4.90)
	$\mathrm{CI}^{*,s}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}-2$	12.4 (1.04)	10.0 (3.13)	9.6 (0.93)	8.8 (1.05)	45.8 (3.94)	11.2 (4.90)
	$\mathrm{CI}^{*,s}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}-3$	13.2 (1.04)	9.2 (3.12)	10.2 (0.93)	8.0 (1.06)	43.4 (4.04)	11.2 (4.89)
	$\mathrm{CI}^{*,s}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}-4$	12.8 (1.04)	9.0 (3.12)	10.4 (0.93)	8.6 (1.05)	42.2 (4.10)	11.8 (4.89)
	$\mathrm{CI}^{*,s}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}-5$	13.0 (1.04)	9.0 (3.12)	11.2 (0.93)	8.6 (1.06)	42.0 (4.17)	11.8 (4.87)
	$\mathrm{CI}^{*,s}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}-6$	13.4 (1.04)	10.6 (3.12)	11.0 (0.93)	8.4 (1.05)	43.2 (4.20)	12.0 (4.87)

Table 4.6: Empirical coverage and mean length ($\times 10^2$) of confidence intervals for the model parameter θ_2 and sample size $n = 100$, in brackets.

α	CI	x_1	x_2	x_3	x_4	x_5	x_6
5%	$\mathrm{CI}^{mc}_{x,\alpha}$	3.6 (0.87)	3.4 (2.64)	4.4 (0.78)	5.2 (0.89)	6.4 (4.25)	5.2 (4.69)
	$\mathrm{CI}^{mc,s}_{x,\alpha}$	5.0 (0.87)	3.6 (2.62)	4.8 (0.78)	6.2 (0.88)	6.8 (4.89)	4.4 (4.57)
	$\mathrm{CI}^{asy}_{x,\alpha}$	5.6 (0.85)	3.8 (2.54)	6.0 (0.75)	7.4 (0.84)	49.6 (2.74)	10.2 (3.78)
	$\mathrm{CI}^{*}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}+3$	5.4 (0.84)	4.2 (2.51)	6.2 (0.75)	7.0 (0.85)	31.4 (3.66)	4.6 (4.83)
	$\mathrm{CI}^{*}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}+2$	5.0 (0.85)	4.8 (2.52)	6.0 (0.75)	7.8 (0.85)	36.2 (3.52)	6.0 (4.70)
	$\mathrm{CI}^{*}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}+1$	5.2 (0.85)	4.8 (2.52)	6.2 (0.76)	7.6 (0.85)	37.8 (3.42)	6.0 (4.59)
	$\mathrm{CI}^{*}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}$	5.4 (0.85)	3.6 (2.52)	6.0 (0.75)	7.6 (0.85)	39.2 (3.36)	6.2 (4.53)
	$\mathrm{CI}^{*}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}-1$	5.4 (0.86)	3.6 (2.57)	6.0 (0.77)	6.4 (0.86)	41.4 (3.20)	6.4 (4.32)
	$\mathrm{CI}^{*}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}-2$	4.8 (0.91)	3.2 (2.72)	5.2 (0.81)	6.2 (0.90)	37.4 (3.32)	7.4 (4.30)
	$\mathrm{CI}^{*}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}-3$	3.6 (0.96)	2.0 (2.85)	3.8 (0.85)	5.2 (0.94)	35.4 (3.49)	7.2 (4.33)
	$\mathrm{CI}^{*}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}-4$	3.4 (0.98)	2.0 (2.92)	3.8 (0.86)	4.4 (0.95)	34.8 (3.54)	8.6 (4.34)
	$\mathrm{CI}^{*}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}-5$	3.4 (1.00)	2.0 (2.98)	3.2 (0.88)	3.4 (0.98)	32.6 (3.64)	7.0 (4.35)
	$\mathrm{CI}^{*}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}-6$	2.8 (1.02)	1.8 (3.04)	2.8 (0.90)	2.6 (0.99)	33.6 (3.72)	7.2 (4.36)
	$\mathrm{CI}^{*,s}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}+3$	5.2 (0.87)	3.6 (2.59)	5.0 (0.77)	6.2 (0.87)	29.2 (3.76)	6.0 (4.60)
	$\mathrm{CI}^{*,s}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}+2$	5.0 (0.87)	3.6 (2.59)	5.2 (0.77)	6.4 (0.87)	34.0 (3.67)	7.0 (4.54)
	$\mathrm{CI}^{*,s}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}+1$	4.8 (0.87)	3.6 (2.59)	5.8 (0.77)	5.8 (0.87)	36.8 (3.59)	8.0 (4.52)
	$\mathrm{CI}^{*,s}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}$	5.6 (0.87)	4.2 (2.59)	5.6 (0.77)	6.8 (0.87)	37.4 (3.56)	7.6 (4.48)
	$\mathrm{CI}^{*,s}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}-1$	4.8 (0.86)	3.6 (2.58)	5.6 (0.77)	6.6 (0.87)	40.4 (3.46)	6.8 (4.52)
	$\mathrm{CI}^{*,s}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}-2$	5.6 (0.87)	3.6 (2.58)	6.0 (0.77)	6.8 (0.87)	39.0 (3.52)	6.4 (4.50)
	$\mathrm{CI}^{*,s}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}-3$	6.2 (0.86)	3.4 (2.57)	5.6 (0.77)	5.4 (0.88)	35.8 (3.63)	6.2 (4.52)
	$\mathrm{CI}^{*,s}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}-4$	5.4 (0.86)	3.4 (2.58)	5.6 (0.77)	6.0 (0.88)	33.8 (3.65)	6.4 (4.53)
	$\mathrm{CI}^{*,s}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}-5$	6.0 (0.86)	3.2 (2.58)	5.4 (0.77)	6.4 (0.89)	32.4 (3.72)	5.6 (4.52)
	$\mathrm{CI}^{*,s}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}-6$	5.4 (0.86)	3.8 (2.57)	5.2 (0.77)	6.2 (0.89)	33.2 (3.77)	5.2 (4.50)
10%	$\mathrm{CI}^{mc}_{x,\alpha}$	9.0 (0.74)	6.4 (2.21)	10.0 (0.66)	10.0 (0.75)	11.2 (3.54)	10.2 (3.86)
	$\mathrm{CI}^{mc,s}_{x,\alpha}$	8.8 (0.73)	8.0 (2.20)	9.4 (0.66)	10.8 (0.74)	11.8 (4.12)	10.4 (3.84)
	$\mathrm{CI}^{asy}_{x,\alpha}$	11.0 (0.71)	9.4 (2.13)	11.4 (0.63)	12.0 (0.71)	57.6 (2.30)	16.0 (3.17)
	$\mathrm{CI}^{*}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}+3$	11.4 (0.71)	9.4 (2.12)	11.0 (0.63)	12.4 (0.72)	40.2 (3.06)	9.8 (4.02)
	$\mathrm{CI}^{*}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}+2$	10.8 (0.71)	9.4 (2.13)	11.6 (0.63)	12.8 (0.72)	43.8 (2.92)	10.4 (3.90)
	$\mathrm{CI}^{*}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}+1$	11.6 (0.71)	9.6 (2.12)	11.0 (0.63)	12.8 (0.72)	45.4 (2.84)	11.0 (3.80)
	$\mathrm{CI}^{*}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}$	11.2 (0.71)	9.0 (2.12)	11.4 (0.63)	13.4 (0.71)	47.8 (2.78)	12.6 (3.73)
	$\mathrm{CI}^{*}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}-1$	10.4 (0.73)	9.4 (2.16)	10.6 (0.65)	12.6 (0.72)	51.6 (2.62)	13.8 (3.54)
	$\mathrm{CI}^{*}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}-2$	9.0 (0.77)	7.0 (2.29)	9.6 (0.68)	10.6 (0.75)	51.0 (2.69)	14.2 (3.49)
	$\mathrm{CI}^{*}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}-3$	7.6 (0.81)	6.0 (2.40)	8.4 (0.71)	8.8 (0.79)	48.2 (2.81)	13.8 (3.48)
	$\mathrm{CI}^{*}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}-4$	6.8 (0.82)	4.6 (2.46)	7.0 (0.73)	8.4 (0.80)	49.6 (2.85)	15.0 (3.46)
	$\mathrm{CI}^{*}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}-5$	6.0 (0.84)	4.4 (2.51)	6.2 (0.74)	9.0 (0.82)	46.6 (2.91)	14.6 (3.45)
	$\mathrm{CI}^{*}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}-6$	5.4 (0.86)	4.0 (2.56)	5.8 (0.76)	8.0 (0.83)	44.8 (2.96)	13.8 (3.44)
	$\mathrm{CI}^{*,s}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}+3$	10.4 (0.73)	8.4 (2.18)	9.8 (0.65)	11.6 (0.73)	38.6 (3.12)	12.4 (3.87)
	$\mathrm{CI}^{*,s}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}+2$	10.2 (0.73)	9.0 (2.18)	10.2 (0.65)	12.4 (0.73)	42.4 (3.03)	11.2 (3.82)
	$\mathrm{CI}^{*,s}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}+1$	10.6 (0.73)	9.0 (2.18)	10.6 (0.65)	12.4 (0.73)	45.2 (2.98)	13.4 (3.80)
	$\mathrm{CI}^{*,s}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}$	10.0 (0.73)	8.8 (2.18)	11.0 (0.65)	13.0 (0.73)	47.2 (2.94)	14.2 (3.76)
	$\mathrm{CI}^{*,s}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}-1$	10.8 (0.73)	9.4 (2.17)	11.2 (0.65)	12.6 (0.73)	47.4 (2.90)	12.0 (3.79)
	$\mathrm{CI}^{*,s}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}-2$	10.2 (0.73)	8.8 (2.17)	11.6 (0.65)	11.6 (0.73)	45.4 (2.96)	12.4 (3.79)
	$\mathrm{CI}^{*,s}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}-3$	11.0 (0.73)	8.6 (2.17)	10.6 (0.65)	11.2 (0.74)	43.0 (3.06)	11.2 (3.80)
	$\mathrm{CI}^{*,s}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}-4$	9.4 (0.73)	8.4 (2.17)	11.2 (0.65)	11.4 (0.74)	40.2 (3.10)	11.6 (3.81)
	$\mathrm{CI}^{*,s}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}-5$	10.8 (0.73)	8.6 (2.17)	10.6 (0.65)	11.6 (0.74)	40.4 (3.16)	10.8 (3.81)
	$\mathrm{CI}^{*,s}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}-6$	10.6 (0.73)	8.6 (2.17)	10.8 (0.65)	10.8 (0.75)	39.6 (3.21)	10.4 (3.80)

Table 4.7: Empirical coverage and mean length ($\times 10^2$) of confidence intervals for the model parameter θ_2 and sample size $n = 200$, in brackets.

α	CI	x_1	x_2	x_3	x_4	x_5	x_6
5%	$\mathrm{CI}^{mc}_{x,\alpha}$	5.8 (0.55)	4.8 (1.64)	4.4 (0.49)	4.2 (0.56)	6.0 (3.01)	5.0 (3.22)
	$\mathrm{CI}^{mc,s}_{x,\alpha}$	5.6 (0.55)	4.2 (1.64)	4.6 (0.49)	4.4 (0.56)	5.6 (3.52)	5.8 (3.09)
	$\mathrm{CI}^{asy}_{x,\alpha}$	6.4 (0.54)	4.4 (1.62)	5.6 (0.48)	4.6 (0.55)	56.2 (1.92)	11.2 (2.65)
	$\mathrm{CI}^{*}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}+3$	7.2 (0.54)	4.6 (1.61)	5.4 (0.48)	4.8 (0.55)	37.4 (2.53)	5.0 (3.30)
	$\mathrm{CI}^{*}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}+2$	6.8 (0.54)	4.8 (1.62)	5.6 (0.48)	4.6 (0.55)	41.6 (2.45)	4.8 (3.22)
	$\mathrm{CI}^{*}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}+1$	6.6 (0.54)	4.8 (1.61)	6.0 (0.48)	4.6 (0.55)	43.2 (2.37)	6.2 (3.13)
	$\mathrm{CI}^{*}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}$	6.8 (0.54)	4.8 (1.61)	5.8 (0.48)	4.6 (0.55)	43.6 (2.34)	6.4 (3.09)
	$\mathrm{CI}^{*}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}-1$	6.2 (0.54)	4.2 (1.62)	5.4 (0.49)	4.8 (0.55)	49.4 (2.16)	8.0 (2.92)
	$\mathrm{CI}^{*}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}-2$	6.2 (0.55)	4.4 (1.64)	5.6 (0.49)	5.0 (0.55)	55.0 (2.07)	8.6 (2.81)
	$\mathrm{CI}^{*}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}-3$	5.0 (0.57)	3.2 (1.71)	3.8 (0.51)	4.0 (0.57)	54.0 (2.10)	11.6 (2.76)
	$\mathrm{CI}^{*}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}-4$	5.2 (0.59)	3.2 (1.76)	4.4 (0.52)	4.0 (0.58)	53.4 (2.12)	11.4 (2.70)
	$\mathrm{CI}^{*}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}-5$	4.4 (0.62)	2.2 (1.84)	3.2 (0.55)	3.8 (0.61)	48.2 (2.23)	9.6 (2.73)
	$\mathrm{CI}^{*}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}-6$	3.6 (0.64)	1.8 (1.91)	2.2 (0.56)	2.8 (0.63)	44.4 (2.33)	9.8 (2.76)
	$\mathrm{CI}^{*,s}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}+3$	7.2 (0.55)	4.2 (1.64)	4.8 (0.49)	4.2 (0.56)	33.8 (2.59)	8.0 (3.11)
	$\mathrm{CI}^{*,s}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}+2$	5.8 (0.54)	4.2 (1.63)	5.8 (0.49)	5.0 (0.56)	38.0 (2.54)	7.2 (3.09)
	$\mathrm{CI}^{*,s}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}+1$	6.6 (0.55)	4.8 (1.63)	5.2 (0.49)	5.0 (0.56)	38.6 (2.48)	7.6 (3.07)
	$\mathrm{CI}^{*,s}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}$	6.6 (0.55)	4.4 (1.63)	4.6 (0.49)	5.0 (0.56)	40.2 (2.46)	7.6 (3.06)
	$\mathrm{CI}^{*,s}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}-1$	6.2 (0.55)	4.4 (1.63)	5.8 (0.49)	4.6 (0.56)	44.4 (2.36)	8.0 (3.05)
	$\mathrm{CI}^{*,s}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}-2$	5.8 (0.54)	5.0 (1.63)	5.0 (0.49)	4.8 (0.56)	47.6 (2.31)	6.4 (3.03)
	$\mathrm{CI}^{*,s}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}-3$	6.8 (0.54)	4.2 (1.63)	5.8 (0.49)	4.6 (0.56)	48.0 (2.34)	5.8 (3.03)
	$\mathrm{CI}^{*,s}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}-4$	6.4 (0.54)	4.8 (1.63)	5.6 (0.49)	4.6 (0.56)	46.0 (2.38)	6.6 (3.04)
	$\mathrm{CI}^{*,s}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}-5$	6.6 (0.54)	4.8 (1.62)	5.2 (0.49)	5.0 (0.57)	40.2 (2.48)	5.8 (3.07)
	$\mathrm{CI}^{*,s}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}-6$	6.8 (0.54)	4.6 (1.63)	5.2 (0.49)	4.4 (0.57)	38.4 (2.56)	6.0 (3.10)
10%	$\mathrm{CI}^{mc}_{x,\alpha}$	10.6 (0.46)	11.0 (1.38)	10.4 (0.41)	8.2 (0.47)	10.4 (2.53)	11.2 (2.67)
	$\mathrm{CI}^{mc,s}_{x,\alpha}$	11.2 (0.46)	10.8 (1.38)	11.4 (0.41)	8.8 (0.47)	10.8 (3.00)	11.6 (2.61)
	$\mathrm{CI}^{asy}_{x,\alpha}$	11.4 (0.45)	11.2 (1.36)	11.4 (0.41)	9.8 (0.46)	63.8 (1.61)	17.6 (2.23)
	$\mathrm{CI}^{*}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}+3$	11.0 (0.45)	11.6 (1.36)	11.6 (0.41)	8.8 (0.47)	47.4 (2.12)	10.4 (2.75)
	$\mathrm{CI}^{*}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}+2$	11.6 (0.45)	11.4 (1.36)	11.0 (0.41)	9.2 (0.46)	49.0 (2.05)	11.6 (2.68)
	$\mathrm{CI}^{*}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}+1$	11.4 (0.45)	12.0 (1.36)	11.6 (0.41)	9.4 (0.46)	50.2 (1.97)	12.0 (2.60)
	$\mathrm{CI}^{*}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}$	12.4 (0.45)	11.4 (1.36)	11.2 (0.41)	9.0 (0.46)	52.0 (1.94)	14.2 (2.56)
	$\mathrm{CI}^{*}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}-1$	11.6 (0.46)	11.4 (1.37)	10.4 (0.41)	9.2 (0.46)	58.8 (1.78)	15.2 (2.41)
	$\mathrm{CI}^{*}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}-2$	11.6 (0.46)	11.2 (1.38)	10.8 (0.41)	9.6 (0.47)	61.4 (1.69)	16.0 (2.30)
	$\mathrm{CI}^{*}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}-3$	9.4 (0.48)	7.4 (1.44)	9.4 (0.43)	7.8 (0.48)	64.2 (1.70)	17.8 (2.23)
	$\mathrm{CI}^{*}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}-4$	8.6 (0.49)	6.6 (1.48)	9.4 (0.44)	7.4 (0.49)	64.8 (1.71)	18.2 (2.17)
	$\mathrm{CI}^{*}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}-5$	7.4 (0.52)	5.0 (1.55)	7.6 (0.46)	6.4 (0.51)	63.6 (1.79)	18.2 (2.18)
	$\mathrm{CI}^{*}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}-6$	6.4 (0.54)	5.2 (1.61)	6.0 (0.47)	4.8 (0.53)	60.8 (1.87)	18.2 (2.19)
	$\mathrm{CI}^{*,s}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}+3$	11.0 (0.46)	11.2 (1.37)	11.4 (0.41)	9.6 (0.47)	43.8 (2.14)	12.6 (2.62)
	$\mathrm{CI}^{*,s}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}+2$	12.0 (0.46)	10.8 (1.37)	10.6 (0.41)	9.4 (0.47)	46.6 (2.10)	12.6 (2.60)
	$\mathrm{CI}^{*,s}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}+1$	11.2 (0.46)	10.6 (1.37)	11.0 (0.41)	9.4 (0.47)	47.2 (2.05)	14.0 (2.58)
	$\mathrm{CI}^{*,s}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}$	11.0 (0.46)	10.8 (1.37)	11.2 (0.41)	9.0 (0.47)	47.6 (2.03)	14.6 (2.57)
	$\mathrm{CI}^{*,s}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}-1$	11.6 (0.46)	11.8 (1.37)	10.8 (0.41)	8.0 (0.47)	51.6 (1.97)	14.2 (2.57)
	$\mathrm{CI}^{*,s}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}-2$	11.2 (0.46)	11.2 (1.37)	10.6 (0.41)	8.4 (0.47)	55.6 (1.93)	13.0 (2.55)
	$\mathrm{CI}^{*,s}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}-3$	12.0 (0.46)	11.0 (1.37)	12.0 (0.41)	8.8 (0.47)	56.4 (1.97)	13.8 (2.55)
	$\mathrm{CI}^{*,s}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}-4$	11.6 (0.46)	10.0 (1.37)	11.6 (0.41)	8.6 (0.47)	56.4 (2.01)	13.8 (2.56)
	$\mathrm{CI}^{*,s}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}-5$	11.2 (0.46)	11.2 (1.37)	11.0 (0.41)	7.6 (0.48)	50.6 (2.10)	13.6 (2.59)
	$\mathrm{CI}^{*,s}_{x,\alpha}, \hat{k}_n^{\mathrm{GCV}}-6$	11.2 (0.46)	12.0 (1.37)	11.2 (0.41)	7.8 (0.48)	48.2 (2.18)	13.0 (2.61)

Table 4.8: Empirical coverage and mean length ($\times 10^2$) of confidence intervals for the model parameter θ_2 and sample size $n = 500$, in brackets.

4.5 Final conclusions

In this chapter, naive and wild bootstrap techniques have been proposed in order to get pointwise confidence intervals for the regression operator $m(x)$ in the functional linear model with scalar response. For these two bootstrap approaches, algorithms have been presented (see Algorithm 4.3.1 and Algorithm 4.3.2, pages 89–90) and their asymptotic validity has been proved (see Theorem 4.3.6, page 91).

From a practical point of view, a simulation study has been carried out, allowing to confirm the good behaviour of bootstrap confidence intervals. Moreover, it was observed that empirical coverage rates of these intervals are closer to nominal α than the coverage rates of confidence intervals based on asymptotic normality results when an *optimal* value for the pilot parameter k_n^d is considered, specially for small sample sizes. However, further research is required in order to find a data–driven method which allows selecting this *optimal* value.

4.6 Appendix Chapter 4

This appendix includes the proof of Theorem 4.3.6, the main result of the chapter, together with the technical lemmas which are necessary to prove it.

4.6.1 Proof of Theorem 4.3.6

In order to prove Theorem 4.3.6, the proof of Theorem 3.2 in Ferraty et al. (2010c) will be mimicked.

Firstly, it is defined

$$\begin{aligned}\phi_{c,d}(y) &= \Phi\left(\frac{y-\sqrt{n}(\mathbb{E}_{\mathcal{X}^n\mathcal{Y}^n}(\hat{m}^*_{c,d}(x))-\hat{m}_d(x))}{\sqrt{n\mathrm{Var}_{\mathcal{X}^n\mathcal{Y}^n}(\hat{m}^*_{c,d}(x))}}\right),\\ \phi_c(y) &= \Phi\left(\frac{y-\sqrt{n}(\mathbb{E}_{\mathcal{X}^n}(\hat{m}_c(x))-\langle\hat{\Pi}_{k_n^c}\theta,x\rangle)}{\sqrt{n\mathrm{Var}_{\mathcal{X}^n}(\hat{m}_c(x))}}\right),\end{aligned} \tag{4.7}$$

where Φ is the distribution function of the standard normal distribution $\mathcal{N}(0,1)$, $\mathbb{E}_{\mathcal{X}^n\mathcal{Y}^n}$ and $\mathrm{Var}_{\mathcal{X}^n\mathcal{Y}^n}$ denote the expectation and the variance conditionally on $(\mathcal{X}^n,\mathcal{Y}^n)=\{(X_1,Y_1),\ldots,(X_n,Y_n)\}$, and $\mathbb{E}_{\mathcal{X}^n}$ and $\mathrm{Var}_{\mathcal{X}^n}$ denote the expectation and the variance conditionally on $\mathcal{X}^n=\{X_1,\ldots,X_n\}$. Then, one can write

$$\begin{aligned}&\mathbb{P}_{\mathcal{X}^n\mathcal{Y}^n}(\sqrt{n}(\hat{m}^*_{c,d}(x)-\hat{m}_d(x))\le y)-\mathbb{P}_{\mathcal{X}^n}(\sqrt{n}(\hat{m}_c(x)-\langle\hat{\Pi}_{k_n^c}\theta,x\rangle)\le y)\\ &=(\mathbb{P}_{\mathcal{X}^n\mathcal{Y}^n}(\sqrt{n}(\hat{m}^*_{c,d}(x)-\hat{m}_d(x))\le y)-\phi_{c,d}(y))+(\phi_{c,d}(y)-\phi_c(y))\\ &\quad+(\phi_c(y)-\mathbb{P}_{\mathcal{X}^n}(\sqrt{n}(\hat{m}_c(x)-\langle\hat{\Pi}_{k_n^c}\theta,x\rangle)\le y))\\ &=T_1(y)+T_2(y)+T_3(y).\end{aligned}$$

Lemma 4.6.1 ensures $T_1(y)\to 0$ *a.s.* and $T_3(y)\to 0$ *a.s.* for a fixed y, and the uniform convergence can be obtained using the continuity of Φ and Polya's Theorem. Finally, Lemma 4.6.3 (see page 104) allows to obtain $\sup_y|T_2(y)|\xrightarrow{\mathbb{P}}0$.

4.6.2 Formulation and proof of Lemma 4.6.1

Lemma 4.6.1. *When the assumptions of Theorem 4.3.6 (see page 91) hold, then*

$$\frac{\hat{m}_c(x) - \mathbb{E}_{\mathcal{X}^n}(\hat{m}_c(x))}{\sqrt{Var_{\mathcal{X}^n}(\hat{m}_c(x))}} \xrightarrow{w} \mathcal{N}(0,1) \quad \textit{and} \quad \frac{\hat{m}^*_{c,d}(x) - \mathbb{E}_{\mathcal{X}^n\mathcal{Y}^n}(\hat{m}^*_{c,d}(x))}{\sqrt{Var_{\mathcal{X}^n\mathcal{Y}^n}(\hat{m}^*_{c,d}(x))}} \xrightarrow{w} \mathcal{N}(0,1),$$

where $\mathbb{E}_{\mathcal{X}^n}$ and $Var_{\mathcal{X}^n}$ denote the expectation and the variance conditionally on $\mathcal{X}^n = \{X_1, \ldots, X_n\}$, whereas $\mathbb{E}_{\mathcal{X}^n\mathcal{Y}^n}$ and $Var_{\mathcal{X}^n\mathcal{Y}^n}$ denote the expectation and the variance conditionally on $(\mathcal{X}^n, \mathcal{Y}^n) = \{(X_1, Y_1), \ldots, (X_n, Y_n)\}$.

Proof. The first statement is shown by Cardot et al. (2007c) (see Theorem 2.3.17 in Chapter 2, page 40). The second one follows after considering Lemma 4.6.2 and checking the next Liapunov's condition

$$\frac{\sum_{i=1}^n \mathbb{E}_{\mathcal{X}^n\mathcal{Y}^n}|(n^{-1}\sum_{j=1}^n f_n^c(\hat{\lambda}_j)\langle X_i, \hat{v}_j\rangle\langle x, \hat{v}_j\rangle)\epsilon_i^*|^3}{(n^{-1}\sigma^2(\hat{t}^c_{n,x})^2)^{3/2}} \xrightarrow{\mathbb{P}} 0.$$

It can be shown that the numerator is $O_{\mathbb{P}}(n^{-2})$. Moreover, using **(C.2.13)** and $[\hat{t}^c_{n,x}]^2/[t^c_{n,x}]^2 \xrightarrow{\mathbb{P}} 1$ (see proof of Corollary 2 by Cardot et al., 2007c), one has that the denominator is $O_{\mathbb{P}}(n^{-3/2})$. Therefore, Liapunov's condition is verified. ■

4.6.3 Formulation and proof of Lemma 4.6.2

Lemma 4.6.2. *When the assumptions of Theorem 4.3.6 (see page 91) hold, then*

$$\begin{aligned} &\mathbb{E}_{\mathcal{X}^n}(\hat{m}_c(x)) = \langle \hat{\Pi}_{k_n^c}\theta, x\rangle + o_{\mathbb{P}}(n^{-1/2}), \quad && Var_{\mathcal{X}^n}(\hat{m}_c(x)) = \tfrac{\sigma^2}{n}(\hat{t}^c_{n,x})^2 + o_{\mathbb{P}}(n^{-3/2}), \\ &\mathbb{E}_{\mathcal{X}^n\mathcal{Y}^n}(\hat{m}^*_{c,d}(x)) = \hat{m}_d(x) + o_{\mathbb{P}}(n^{-1/2}), \quad && Var_{\mathcal{X}^n\mathcal{Y}^n}(\hat{m}^*_{c,d}(x)) = \tfrac{\sigma^2}{n}(\hat{t}^c_{n,x})^2 + o_{\mathbb{P}}(n^{-1}), \end{aligned}$$

where $\hat{\Pi}_{k_n^c}$ is the projector onto the subspace spanned by the first k_n^c eigenfunctions of Γ_n, and $\hat{t}^c_{n,x} = \sqrt{\sum_{j=1}^{k_n^c} \hat{\lambda}_j (f_n^c(\hat{\lambda}_j))^2\langle x, \hat{v}_j\rangle^2}$.

Proof. Rewriting $\hat{m}_c(x) = n^{-1}\sum_{i=1}^n (\sum_{j=1}^n f_n^c(\hat{\lambda}_j)\langle X_i, \hat{v}_j\rangle\langle x, \hat{v}_j\rangle)Y_i$, then

$$\begin{aligned} \mathbb{E}_{\mathcal{X}^n}(\hat{m}_c(x)) &= \frac{1}{n}\sum_{i=1}^n \left(\sum_{j=1}^n f_n^c(\hat{\lambda}_j)\langle X_i, \hat{v}_j\rangle\langle x, \hat{v}_j\rangle\right)\langle \theta, X_i\rangle = \sum_{j=1}^n \hat{\lambda}_j f_n^c(\hat{\lambda}_j)\langle \theta, \hat{v}_j\rangle\langle x, \hat{v}_j\rangle \\ &= \langle \hat{\Pi}_{k_n^c}\theta, x\rangle + o_{\mathbb{P}}(n^{-1/2}), \end{aligned}$$

where the last equality follows from assumption **(C.2.4)** (see Chapter 2, page 37) and Remark 2.3.8 (see Chapter 2, page 37). For the variance, using Remark 2.3.8 again, one gets

$$\begin{aligned} \mathrm{Var}_{\mathcal{X}^n}(\hat{m}_c(x)) &= \frac{\sigma^2}{n^2}\sum_{i=1}^n \left(\sum_{j=1}^n f_n^c(\hat{\lambda}_j)\langle X_i, \hat{v}_j\rangle\langle x, \hat{v}_j\rangle\right)^2 \\ &= \frac{\sigma^2}{n}\sum_{j_1=1}^n\sum_{j_2=1}^n f_n^c(\hat{\lambda}_{j_1}) f_n^c(\hat{\lambda}_{j_2})\langle \Gamma_n \hat{v}_{j_1}, \hat{v}_{j_2}\rangle\langle x, \hat{v}_{j_1}\rangle\langle x, \hat{v}_{j_1}\rangle = \frac{\sigma^2}{n}\sum_{j=1}^n \hat{\lambda}_j (f_n^c(\hat{\lambda}_j))^2\langle x, \hat{v}_j\rangle^2 \\ &= \frac{\sigma^2}{n}(\hat{t}^c_{n,x})^2 + o_{\mathbb{P}}(n^{-3/2}). \end{aligned}$$

On the other hand, $\hat{m}^*_{c,d}(x) = n^{-1}\sum_{i=1}^n (\sum_{j=1}^n f_n^c(\hat{\lambda}_j)\langle X_i, \hat{v}_j\rangle\langle x, \hat{v}_j\rangle)Y_i^*$, so one can reproduce the reasoning which was done for $\hat{m}_c$ and obtain

$$\begin{aligned}\mathbb{E}_{\mathcal{X}^n\mathcal{Y}^n}(\hat{m}^*_{c,d}(x)) &= \frac{1}{n}\sum_{i=1}^n\left(\sum_{j=1}^n f_n^c(\hat{\lambda}_j)\langle X_i, \hat{v}_j\rangle\langle x, \hat{v}_j\rangle\right)\langle\hat{\theta}_d, X_i\rangle = \sum_{j=1}^n \hat{\lambda}_j f_n^c(\hat{\lambda}_j)\langle\hat{\theta}_d, \hat{v}_j\rangle\langle x, \hat{v}_j\rangle \\ &= \langle\hat{\Pi}_{k_n^c}\hat{\theta}_d, x\rangle + o_{\mathbb{P}}(n^{-1/2}) = \hat{m}_d(x) + o_{\mathbb{P}}(n^{-1/2}),\end{aligned}$$

where the last equality comes from **(C.4.1)**. Moreover, for the naive bootstrap one has that

$$\begin{aligned}\mathrm{Var}_{\mathcal{X}^n\mathcal{Y}^n}(\hat{m}^*_{c,d}(x)) &= \frac{1}{n^2}\sum_{i=1}^n\left(\sum_{j=1}^n f_n^c(\hat{\lambda}_j)\langle X_i, \hat{v}_j\rangle\langle x, \hat{v}_j\rangle\right)^2\left(\frac{1}{n}\sum_{i=1}^n(\hat{\epsilon}_i - \bar{\hat{\epsilon}})^2\right) \\ &= \frac{\sigma^2 + o_{\mathbb{P}}(1)}{n}\sum_{j=1}^n \hat{\lambda}_j(f_n^c(\hat{\lambda}_j))^2\langle x, \hat{v}_j\rangle^2 = \frac{\sigma^2}{n}(\hat{t}^c_{n,x})^2 + o_{\mathbb{P}}(n^{-1}),\end{aligned}$$

and similar calculations allow to reach the same result when one considers the wild bootstrap procedure. ■

4.6.4 Formulation and proof of Lemma 4.6.3

Lemma 4.6.3. *When the assumptions of Theorem 4.3.6 (see page 91) hold, then*

$$\sup_{y\in\mathbb{R}}|\phi_{c,d}(y) - \phi_c(y)| \xrightarrow{\mathbb{P}} 0,$$

with $\phi_{c,d}$ and ϕ_c defined in (4.7) *(see page 102).*

Proof. First of all, it is necessary to prove that for any $a \in \mathbb{R}$ and $b > 0$

$$\sup_{z\in\mathbb{R}}|\Phi(a + bz) - \Phi(z)| \leq |a| + \max(b, b^{-1}) - 1. \tag{4.8}$$

In order to show (4.8), let z be a fixed value $z \in \mathbb{R}$. One has that

$$|\Phi(a + bz) - \Phi(z)| \leq |\Phi(a + bz) - \Phi(bz)| + |\Phi(bz) - \Phi(z)|. \tag{4.9}$$

It can be found that

$$|\Phi(a + bz) - \Phi(bz)| \leq |a|. \tag{4.10}$$

On the other hand, $|\Phi(bz) - \Phi(z)| \leq |b - 1|\max(b^{-1}, 1)$. Therefore,

$$|\Phi(bz) - \Phi(z)| \leq \max(b, b^{-1}) - 1. \tag{4.11}$$

Then, one gets the inequality (4.8) replacing (4.10) and (4.11) in (4.9).

Consider $a_0 = -(\mathbb{E}_{\mathcal{X}^n\mathcal{Y}^n}(\hat{m}^*_{c,d}(x)) - \hat{m}_d(x) - \mathbb{E}_{\mathcal{X}^n}(\hat{m}_c(x)) + \langle\hat{\Pi}_{k_n^c}\theta, x\rangle)/\sqrt{\mathrm{Var}_{\mathcal{X}^n\mathcal{Y}^n}(\hat{m}^*_{c,d}(x))}$, $b_0 = \sqrt{\mathrm{Var}_{\mathcal{X}^n}(\hat{m}_c(x))/\mathrm{Var}_{\mathcal{X}^n\mathcal{Y}^n}(\hat{m}^*_{c,d}(x))}$, and $z_0 = (y - \sqrt{n}(\mathbb{E}_{\mathcal{X}^n}(\hat{m}_c(x)) - \langle\hat{\Pi}_{k_n^c}\theta, x\rangle))/\sqrt{n\mathrm{Var}_{\mathcal{X}^n}(\hat{m}_c(x))}$. Substituting these values in (4.8), one has

$$\sup_{y\in\mathbb{R}}|\phi_{c,d}(y) - \phi_c(y)| = \sup_{z_0\in\mathbb{R}}|\Phi(a_0 + b_0z_0) - \Phi(z_0)| \leq |a_0| + \max(b_0, b_0^{-1}) - 1.$$

Thus, the convergence follows from Lemma 4.6.2 (see page 103). ■

Chapter 5

Testing in functional linear regression

As indicated before, functional data have been the subject of many research works over the last years, being functional regression one of the most discussed issues. In this chapter, the functional linear model with scalar response is considered but, unlike Chapter 3 and Chapter 4, including an intercept term, that is, $Y = \langle \theta, X \rangle + b + \epsilon$, where Y and ϵ are real random variables, X is a random variable valued in a separable Hilbert space $(\mathcal{H}, \langle \cdot, \cdot \rangle)$, and the model parameters b and θ belong to $\mathbb{R}$ and $\mathcal{H}$, respectively. In this context, a consistent bootstrap method to calibrate the distribution of test statistics for assesing $H_0 : \theta = 0$ versus $H_1 : \theta \neq 0$ (i.e., for testing the lack of dependence) is developed, and the related asymptotic theory is presented. Next, two linear models, $Y_1 = \langle \theta_1, X_1 \rangle + b_1 + \epsilon_1$ and $Y_2 = \langle \theta_2, X_2 \rangle + b_2 + \epsilon_2$, satisfying that X_1 and X_2 have the same covariance operator, and ϵ_1 and ϵ_2 have the same variance, are taken. Then, a bootstrap method for checking the equality of the two linear models, i.e., for testing $H_0 : \theta_1 = \theta_2$ versus $H_1 : \theta_1 \neq \theta_2$, is introduced, and its asymptotic properties are studied. Finally, a simulation study and a real data example illustrate the performance of the proposed bootstrap techniques in practice.

The applications of bootstrap calibration for testing in functional linear regression compiled in this chapter were firstly introduced in González-Manteiga and Martínez-Calvo (2010). Furthermore, the asymptotic theory for the first testing problem, i.e., the lack of dependence test, was developed in González-Manteiga et al. (2012)[1].

5.1 Testing in FDA

Although the literature on model construction and estimation methods for FDA has increased considerably during the last years, there are still few contributions on testing procedures. This fact makes this field an important challenge for the statistical community nowadays. Among the testing issues which researchers already started studying can be highlighted curve comparison problems (Cuevas et al., 2004; Delicado, 2007; Hall and Van Keilegom, 2007; Ferraty et al., 2007b; Zhang and Chen, 2007; Berkes et al., 2009; Bugni et al., 2009; Cuesta-Albertos and Febrero-Bande, 2010; Zhang, 2011; Horváth et al., 2012; González-Rodríguez et al., 2012), hypothesis testing on the structure of the distribution of functional data (Viele, 2001; James and Sood, 2006; Hall and Vial, 2006b; Mas, 2007a), goodness–of–fit tests for parametric distributions (Cuesta-Albertos et al., 2007), or testing a specific form of the conditional density function (Ferraty et al., 2012c).

[1] In particular, Theorem 5.2.3, Corollary 5.2.4, Theorem 5.2.5, Theorem 5.2.7, Theorem 5.2.9, Lemma 5.7.1 and Lemma 5.7.2 were developed by Prof. Gil González–Rodríguez, who has allowed to include them in this thesis in order to enrich the theoretical contents of this chapter.

Other testing problems analysed in the literature are testing procedures in regression models with a functional component. In this context, tests for no effect of regression or no effect of any of the terms included in the model (Cardot et al., 2003b, 2004; Gadiaga and Ignaccolo, 2005; Müller and Stadtmüller, 2005; Antoniadis and Sapatinas, 2007; Cardot et al., 2007d; Kokoszka et al., 2008; Delsol et al., 2011b; Horváth and Reeder, 2011; Hilgert et al., 2012; Aneiros-Pérez and Vieu, 2012), as well as tests for a specific parametric form of the regression operator (Cardot et al., 2007d; Bücher et al., 2011; Delsol et al., 2011b; García-Portugués et al., 2012) can be emphasized. Other contributions focused on testing the lack of fit by examining the residuals of the regression model (Chiou and Müller, 2007; Gabrys et al., 2010; Patilea et al., 2012). Moreover, tests for dimension reduction (Delsol et al., 2011b), for detecting change–points in the regression operator (Horváth and Reeder, 2012), for choosing among two nested linear models (Shen and Faraway, 2004), or for testing equality of two linear models (Horváth et al., 2009) were also developed.

Many of these testing methods require the use of resampling techniques, such as bootstrap, in order to be applied in practice since either their asymptotic theory has not been developed, or it is difficult to compute, or it does not exhibit an acceptable behaviour for small sample size cases. In this regard, the application of bootstrap to the functional field has been successfully initiated (see a general overview of resampling methods for functional data in McMurry and Politis (2011)). For instance, Politis and Romano (1994) derived theoretical results that support the asymptotic validity of a stationary bootstrap method for a broad class of estimators. Cuevas et al. (2006) proposed bootstrap confidence bands for several functional estimators such as the sample and the trimmed functional means. In the regression field, Ferraty et al. (2010c), and González-Manteiga and Martínez-Calvo (2011) showed the validity of the bootstrap in the estimation of nonparametric functional regression and functional linear model, respectively, for scalar response, whereas the asymptotic validity of a componentwise bootstrap procedure was proved by Ferraty et al. (2012d) when a nonparametric regression is considered and both the response and the regressor are functional. On the other hand, as remarked before, bootstrap techniques can be also very helpful for testing purposes, since they may be used to approximate the distribution of the statistic under the null hypothesis. For example, Cuevas et al. (2004) developed a sort of parametric bootstrap to obtain quantiles for a functional ANOVA test, and González-Rodríguez et al. (2012) proved the validity of a residual bootstrap in that context. Hall and Vial (2006b) and, more recently, Bathia et al. (2010) studied the finite dimensionality of functional data using a bootstrap approximation for independent and dependent samples, respectively. Furthermore, in the nonparametric regression model with functional response, Bücher et al. (2011) proposed tests for the hypothesis that the regression function presents a specific parametric form, and constructed their bootstrap versions, obtaining tests with an accurate approximation of the nominal level for small samples.

This chapter is devoted to two testing problems in the functional linear model with scalar response: testing the lack of dependence and testing the equality of two models. Bootstrap methods for these two issues are introduced in Section 5.2 and Section 5.3, respectively. In both cases, their empirical behaviour is analysed by means of a simulation study in Section 5.4, and applications to real datasets in Section 5.5. Then, a brief final discussion and a technical appendix with the proofs of the results presented in the chapter can be found in Section 5.6 and Section 5.7.

5.2 Test for lack of dependence

This section is going to be focused on the functional linear regression model with scalar response that is described below. Let $(\mathcal{H}, \langle\cdot,\cdot\rangle)$ be a separable Hilbert space, and let $\|\cdot\|$ be the norm associated with its inner product. Moreover, let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and let (X, Y) be a measurable mapping from Ω to $\mathcal{H} \times \mathbb{R}$, that is, X is an $\mathcal{H}$–valued random element whereas Y is a real random variable, such that (X, Y) verifies the following linear model with scalar response,

$$Y = \langle \theta, X \rangle + b + \epsilon, \tag{5.1}$$

where $\theta \in \mathcal{H}$ is a fixed functional model parameter, $b \in \mathbb{R}$ is the intercept term, and ϵ is a real random variable such that $\mathbb{E}(\epsilon) = 0$, $\mathbb{E}(\epsilon^2) = \sigma^2 < \infty$, and $\mathbb{E}(\epsilon X) = 0$.

The main aim of this section is to develop a consistent general bootstrap resampling approach to calibrate the distribution of statistics for testing the significance of the relationship between X and Y, that is, for testing $H_0 : \theta = 0$ versus $H_1 : \theta \neq 0$, on the basis of a simple random sample $\{(X_i, Y_i)\}_{i=1}^n$ drawn from (X, Y). Bootstrap techniques will become an alternative useful tool when the asymptotics of test statistics are unknown or when they are not accurate enough for small sample sizes.

Testing the lack of dependence between X and Y has stirred up a great interest during the last years due to its practical applications in the functional context. For instance, Kokoszka et al. (2008) proposed a test for lack of dependence in the functional linear model with functional response which was applied to magnetometer curves consisting of minute–by–minute records of the horizontal intensity of the magnetic field measured at observatories located at different latitude. The aim was to analyse if the high–latitude records had a linear effect on the mid–latitude or low–latitude records. On the other hand, Cardot et al. (2007d) presented a statistical procedure to check if a real–valued covariate has an effect on a functional response in a nonparametric regression context, using this methodology for a study of atmospheric radiation. In this case, the dataset were radiation profiles curves measured at a random time and the authors tested if the radiation profiles changed along the time.

Some contributions have studied similar tests in other functional regression contexts. For instance, in the case of scalar response and functional predictor, Müller and Stadtmüller (2005) analysed the generalized functional linear regression model and tested whether the predictor function has any influence on the outcome; Horváth and Reeder (2011) studied a quadratic functional regression model and tested the significance of the nonlinear term in the model; Aneiros-Pérez and Vieu (2012) assessed the linear component of a functional partially linear model; and Gadiaga and Ignaccolo (2005) and Delsol et al. (2011b) checked if the explanatory variable has an effect on the response using nonparametric methods. Moreover, Antoniadis and Sapatinas (2007) provided procedures for testing whether certain fixed–effects functional components or the random–effects functional components are equal to zero in a general functional mixed–effects model with functional response.

Regarding the regression model (5.1), testing the significance of the relationship between the functional covariate and the scalar response has been the subject of recent works, and asymptotic approaches for this problem can be found in Cardot et al. (2003b, 2004), Kokoszka et al. (2008) or Hilgert et al. (2012) (Kokoszka et al. (2008) studied indeed the more general situation of functional response). The methods presented in these papers are mainly based on the calibration of the statistics distribution by using asymptotic distribution approximations. In contrast, a consistent bootstrap calibration is proposed in order to approximate the statistics distribution under H_0. For that purpose, some notation and basic concepts about the regression model (5.1), the asymptotic theory for the testing procedure, and the consistency of the bootstrap techniques that are proposed, are introduced in Section 5.2.1. In Section 5.2.2, the bootstrap calibration is presented as an alternative to the asymptotic theory previously exposed. Later, in Section 5.4.1 and Section 5.5.1, a simulation study and a real data application allow to show the performance of the bootstrap methodology in comparison with the asymptotic approach.

5.2.1 Asymptotic theory for testing and bootstrap procedures

Theoretical background

Riesz Representation Theorem ensures that the functional linear model with scalar response can be handled theoretically within the considered framework. Specifically, let $\mathcal{H}$ be the separable Hilbert space of square Lebesgue integrable functions on a given compact set $C \subset \mathbb{R}$, denoted by $\mathcal{L}^2(C, \lambda)$, with the usual inner product and the associated norm $\|\cdot\|$. The functional linear model with scalar response between a random function X and a real random variable Y is defined as

$$Y = \Psi(X) + \epsilon, \tag{5.2}$$

where Ψ is a continuous linear operator, i.e., $\Psi \in \mathcal{H}'$ being $\mathcal{H}'$ the dual space of $\mathcal{H}$ with associated norm $\|\cdot\|_{\mathcal{H}'}$ (see Section 1.2.2, *"b) The dual space $\mathcal{H}'$"*, in Chapter 1, page 10), and ϵ is a real centred random variable with finite variance and independent of X. In virtue of Riesz Representation Theorem, $\mathcal{H}$ and

$\mathcal{H}'$ are isometrically identified, in such a way that for any $\Psi \in \mathcal{H}'$ there exists a unique $\theta \in \mathcal{H}$ such that $\|\theta\| = \|\Psi\|_{\mathcal{H}'}$ and $\Psi(x) = \langle \theta, x \rangle$ for all $x \in \mathcal{H}$. Consequently, the model presented in equation (5.2) is just a particular case of the one considered in (5.1).

Previous works regarding functional linear model assumed that $b = 0$ (see Cardot et al. 2003b, and Kokoszka et al. 2008). Of course, the intercept term can be embedded in the variable counterpart of the model, as in the multivariate case, as follows. Let $\mathcal{H}_e$ be the product space $\mathcal{H} \times \mathbb{R}$ with the corresponding inner product $\langle \cdot, \cdot \rangle_e$. Defining $\tilde{X} = (X, 1)$ and $\tilde{\theta} = (\theta, b) \in \mathcal{H}_e$, the model considered in (5.1) can be rewritten as $Y = \langle \tilde{\theta}, \tilde{X} \rangle_e + \epsilon$ and, consequently, $\tilde{X}$ cannot be assumed to be centred. Nevertheless, in the context of the linear independence test, the aim is to check if $\theta = 0$ or not, which is not equivalent to checking whether $\tilde{\theta} = 0$ or not. In addition, the intercept term b cannot be assumed to be equal to 0 in practice. Because of all of this, the intercept term b has been written explicitly in this section in order to avoid any kind of confusion.

In the same way, the random element X is assumed to be centred in the above–mentioned papers. Although, the asymptotic distribution of the proposed statistics does not change in many cases if $\{X_i\}_{i=1}^n$ is replaced by the dependent sample $\{X_i - \overline{X}\}_{i=1}^n$, where $\overline{X} = n^{-1} \sum_{i=1}^n X_i$, the situation could be quite different when the bootstrap version of the statistics are considered. In fact, as it will be shown afterwards, different bootstrap statistics could be considered when this replacement is done. Hence, for the developments in this section, it will not be assumed that the predictor variable X is centred.

Linear independence test

Assuming that the linear model defined in (5.1) holds (see page 106), where the regression operator $m(\cdot)$ is given by $m(x) = \mathbb{E}(Y|X = x) = \langle \theta, x \rangle + b$, for all $x \in \mathcal{H}$, the goal is to develop correct and consistent bootstrap techniques for testing

$$\begin{cases} H_0: & \theta = 0 \\ H_1: & \theta \neq 0 \end{cases} \tag{5.3}$$

on the basis of a random sample $\{(X_i, Y_i)\}_{i=1}^n$ of i.i.d. random elements with the same distribution as (X, Y). In other words, the objective is to check whether X and Y are linearly independent (H_0) or not (H_1).

Recall the definitions of the covariance and cross–covariance operators (see Section 1.4.1, *"b) Measures of dispersion"*, in Chapter 1, page 22), and the normal equation that link them when the linear model holds. Given a generic $\mathcal{H}$–valued random element H such that $\mathbb{E}(\|H\|^2) < \infty$, its associated covariance operator Γ_H is defined as

$$\Gamma_H = \mathbb{E}((H - \mu_H) \otimes_{\mathcal{H}} (H - \mu_H)) = \mathbb{E}(H \otimes_{\mathcal{H}} H) - \mu_H \otimes_{\mathcal{H}} \mu_H,$$

where $\mu_H \in \mathcal{H}$ denotes the expected value of H. From now on, it will be assumed that $\mathbb{E}(\|X\|^2) < \infty$, and thus, as a consequence of Hölder's inequality, $\mathbb{E}(Y^2) < \infty$. Whenever there is no possible confusion, Γ_X will be abbreviated as Γ. It is well–known that Γ is a nuclear and self–adjoint operator and, in particular, a compact operator of trace class. Therefore, in virtue of the Spectral Decomposition Theorem, there is an orthonormal basis of H, $\{v_j\}_{j=1}^\infty$, consisting of the eigenfunctions of Γ with associated eigenvalues $\{\lambda_j\}_{j=1}^\infty$ (i.e., $\Gamma v_j = \lambda_j v_j$ for all j). As usual, the eigenvalues are assumed to be arranged in decreasing order ($\lambda_1 \geq \lambda_2 \geq \ldots$). Furthermore, these eigenvalues are non–negative since the operator Γ is symmetric and non–negative definite. On the other hand, the cross–covariance operator between X and Y is given by

$$\Delta_{X,Y} = \mathbb{E}((X - \mu_X) \otimes_{\mathcal{H}'} (Y - \mu_Y)) = \mathbb{E}(X \otimes_{\mathcal{H}'} Y) - \mu_X \otimes_{\mathcal{H}'} \mu_Y,$$

where $\mu_X \in \mathcal{H}$ and $\mu_Y \in \mathbb{R}$ denote the expected value of X and Y, respectively. Again, $\Delta_{X,Y}$ will be abbreviated as Δ in order to simplify the notation, when confusion is not possible. Of course, $\Delta \in \mathcal{H}'$ and the following relation between the covariance and cross–covariance operators and the regression parameter θ is satisfied

$$\Delta x = \langle \theta, \Gamma x \rangle, \quad \forall x \in \mathcal{H}. \tag{5.4}$$

The Hilbert space $\mathcal{H}$ can be expressed as the direct sum of two orthogonal subspaces induced by the self–adjoint operator Γ: the *kernel* or *null space* of Γ defined as $\mathrm{Ker}(\Gamma) = \{x \in \mathcal{H}/\Gamma x = 0\}$, and the closure of the *image* or *range* of Γ, $\overline{\mathrm{Im}(\Gamma)}$, where $\mathrm{Im}(\Gamma) = \{x' \in \mathcal{H} \,|\, x' = \Gamma x,\, x \in \mathcal{H}\}$. Then, θ is determined uniquely by $\theta = \theta_1 + \theta_2$, with $\theta_1 \in \mathrm{Ker}(\Gamma)$ and $\theta_2 \in \overline{\mathrm{Im}(\Gamma)}$. It is easy to check that $\mathrm{Var}(\langle \theta_1, X \rangle) = 0$, since $\theta_1 \in \mathrm{Ker}(\Gamma)$. Consequently, the model introduced in (5.1) can be expressed as

$$Y = \langle \theta_2, X \rangle + \langle \theta_1, \mu_X \rangle + b + \epsilon.$$

Hence, there is an identifiability problem, since it is not possible to distinguish between the term $\langle \theta_1, \mu_X \rangle$ and the intercept term b and, as a result, it is not feasible to check whether $\theta_1 = 0$ or not. Taking this fact into account, the hypothesis test will be restricted to check

$$\begin{cases} H_0: & \theta_2 = 0 \\ H_1: & \theta_2 \neq 0 \end{cases} \tag{5.5}$$

on the basis of the available sample information. Note that, in this case, according to the relation between the operators Δ, Γ and the regression parameter θ shown in (5.4), $\theta_2 = 0$ if, and only if, $\Delta x = 0$ for all $x \in \mathcal{H}$. Hence, the hypothesis test in (5.5) is equivalent to

$$\begin{cases} H_0: & \|\Delta\|_{\mathcal{H}'} = 0 \\ H_1: & \|\Delta\|_{\mathcal{H}'} \neq 0. \end{cases} \tag{5.6}$$

Remark 5.2.1. Recall that μ_X is assumed to be equal 0 in previous works, revised at the beginning of Section 5.2. Therefore, the preceding reasoning leads to the fact that θ_1 cannot be estimated based on the information provided by X (see, for instance, Cardot et al. (2003b)). Consequently, the hypothesis testing is also restricted to (5.6). In addition, Cardot et al. (2003b) assumed for technical reasons that $\overline{\mathrm{Im}(\Gamma)}$ is an infinite dimensional space. On the contrary, this restriction is not imposed in the study developed here.

Remark 5.2.2. Note that another usual assumption is that the intercept term vanishes. Although this is not common in most of situations, it should be noted that if $b = 0$ and X is not assumed to be centred as in this section, then an interesting possibility appears: to check whether $\theta_1 = 0$ or not by testing the nullity of the intercept term of the model, and thus to check the original hypothesis in (5.3). This open problem cannot be solved with the methodology described in the current section, or with the techniques developed in the existing literature, since the idea is based on testing (5.6), which is equivalent to the restricted test (5.5) but not to the unrestricted one in (5.3).

Testing procedure and asymptotic theory

According to the connection between $\|\cdot\|_{\mathcal{H}'}$ and $\|\cdot\|$, the dual norm of $\Delta \in \mathcal{H}'$ can be expressed equivalently in terms of the $\mathcal{H}$–valued random element

$$T^{Ind} = \mathbb{E}((X - \mu_X)(Y - \mu_Y))$$

as follows

$$\|\Delta\|_{\mathcal{H}'} = \|\mathbb{E}((X - \mu_X) \otimes_{\mathcal{H}'} (Y - \mu_Y))\|_{\mathcal{H}'} = \|\mathbb{E}((X - \mu_X)(Y - \mu_Y))\| = \|T^{Ind}\|.$$

Therefore, given an i.i.d. sample $\{(X_i, Y_i)\}_{i=1}^n$ drawn from (X, Y), $\|T^{Ind}\|$ can be estimated in a natural way by means of its empirical counterpart $\|T_n^{Ind}\|$, being T_n^{Ind} the $\mathcal{H}$–valued random element defined as

$$T_n^{Ind} = \frac{1}{n} \sum_{i=1}^{n} (X_i - \overline{X})(Y_i - \overline{Y}),$$

where $\overline{X}$ and $\overline{Y}$ denote the corresponding sample means, i.e., $\overline{X} = n^{-1} \sum_{i=1}^n X_i$ and $\overline{Y} = n^{-1} \sum_{i=1}^n Y_i$. The next theorem establishes some basic properties of T_n^{Ind} when the assumption

(C.5.1) $\mathbb{E}(\|X\|^4) < \infty$

is satisfied.

Theorem 5.2.3 (González-Manteiga et al., 2012)**.** *Assuming that* (5.1) *and* **(C.5.1)** *hold, then*

(i) $\mathbb{E}(T_n^{Ind}) = n^{-1}(n-1)T^{Ind}$.

(ii) T_n^{Ind} *converges a.s.* $-\mathbb{P}$ *to* T^{Ind} *as* $n \to \infty$.

(iii) $\sqrt{n}(T_n^{Ind} - T^{Ind})$ *converges in law, as* $n \to \infty$*, to a centred Gaussian element* Z *in* $\mathcal{H}$ *with covariance operator* Γ_Z *given by* $\Gamma_Z = \sigma^2\Gamma + \mathbb{E}(\langle\theta, X-\mu_X\rangle^2(X-\mu_X)\otimes_{\mathcal{H}}(X-\mu_X))$.

The proof of Theorem 5.2.3 can be found in the appendix of the chapter (see Section 5.7.1, page 134).

In order to simplify the notation, from now on, given any $\mathcal{H}$–valued random element H such that $\mathbb{E}(\|H\|^2) < \infty$, Z_H will denote a centred Gaussian element in $\mathcal{H}$ with covariance operator Γ_H.

Corollary 5.2.4 (González-Manteiga et al., 2012)**.** *Under the assumptions of Theorem 5.2.3, if the null hypothesis in* (5.6) *is satisfied (i.e.,* $\|\Delta\|_{\mathcal{H}'} = 0$*), then* $\sqrt{n}\,T_n^{Ind}$ *converges in law to* $Z_{(X-\mu_X)^\epsilon}$ *with covariance operator* $\sigma^2\Gamma$*, and consequently,* $\|\sqrt{n}\,T_n^{Ind}\|$ *converges in law to* $\|Z_{(X-\mu_X)^\epsilon}\|$.

The proof of the previous corollary has been included in Section 5.7.2 (see page 134).

In contrast to Theorem 1 in Cardot et al. (2003b), the result in Corollary 5.2.4 is directly established on the Hilbert space $\mathcal{H}$ instead of on its dual space $\mathcal{H}'$. In addition, no assumptions of centred random elements X or null intercept term are necessary. However, these two assumptions could be easily removed in that paper in order to establish a dual result of Corollary 5.2.4. Furthermore, in view of Corollary 5.2.4, the asymptotic null distribution of $\|\sqrt{n}\,T_n^{Ind}\|$ is not explicitly known. This is the reason why no further research on how to use in practice this statistic (or its dual one) for checking if θ_2 equals 0 was carried out in Cardot et al. (2003b). Indeed, an alternative statistic that is used in the simulation section for comparative purposes is considered. Nevertheless, it is still possible to use $\|\sqrt{n}\,T_n^{Ind}\|$ as a core statistic in order to solve this test in practice by means of bootstrap techniques.

Next, a natural way of using the asymptotic result of Corollary 5.2.4 for addressing the testing problem under study is presented. Consider a consistent (at least under H_0) estimator $\hat{\sigma}^2$ of σ^2 (for instance, the sample variance of Y, or perhaps the estimator introduced by Cardot et al. (2003b), provided that its theoretical behaviour is analysed). Then, according to Slutsky's Theorem $\|\sqrt{n}\,T_n^{Ind}\|/\hat{\sigma}$ converges in law under H_0 to the norm of Z_X. Since its covariance operator Γ is unknown, it can be approximated by the empirical operator Γ_n. Thus, $\|Z_X\|$ can be approximated by $\|Z_n\|$, being Z_n a centred Gaussian element in $\mathcal{H}$ with covariance operator Γ_n. The distribution of $\|Z_n\|$ is still difficult to compute directly, however one can use the CLT and approximate its distribution by means of Monte–Carlo method, that is, by the distribution of $\|(n_{mc})^{-1}\sum_{i=1}^{n_{mc}}(X_i^* - \overline{X})\|$ for a large value of n_{mc}, being $\{X_i^*\}_{i=1}^{n_{mc}}$ i.i.d. random elements chosen at random from the fixed population $(X_1, \ldots, X_n)$. Obviously, this method is a precursor of the bootstrap procedures.

Local alternatives. In order to complete the asymptotic study of the statistic $\|\sqrt{n}\,T_n^{Ind}\|$, its behaviour under local alternatives is studied. For this purpose, let $\theta \in \mathcal{H}$ be such that $\|\theta_2\| > 0$, and given $\delta_n > 0$ take the modified random sample

$$Y_i^n = \left\langle \frac{\delta_n}{\sqrt{n}}\theta, X_i \right\rangle + b + \epsilon_i, \quad i = 1, \ldots, n.$$

In this situation, the null hypothesis is not verified. Nevertheless, if the condition

(C.5.2) $\delta_n \to \infty$ and $\delta_n/\sqrt{n} \to 0$ as $n \to \infty$,

is satisfied, then $\|\delta_n\theta/\sqrt{n}\| \to 0$, that is, H_0 is approached at rate $\delta_n/\sqrt{n}$. Under these conditions, the following theorem that establishes the behaviour of the statistic under the considered local alternatives can be easily deduced.

Theorem 5.2.5 (González-Manteiga et al., 2012)**.** *Under the assumptions of Theorem 5.2.3, and with the above notation, if* **(C.5.2)** *holds, then*

$$\mathbb{P}\left(\left\|\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})(Y_i^n-\overline{Y}^n)\right)\right\|\leq t\right)\to 0,\quad \forall t\in\mathbb{R}$$

as $n\to\infty$.

The proof of Theorem 5.2.5 is compiled in Section 5.7.3 (see page 134).

Bootstrap procedures and consistency

The difficulty in using the previously proposed statistic to solve the hypothesis test by means of asymptotic procedures suggests the development of appropriated bootstrap techniques. The asymptotic consistency of a bootstrap approach is guaranteed if the associated bootstrap statistic converges in law to a non–degenerated distribution irrespectively of H_0 being satisfied or not. In addition, in order to ensure its asymptotic correctness, this limit distribution must coincide with the asymptotic distribution of the testing statistic provided that H_0 holds.

For the first issue, recall that, by item 3 in Theorem 5.2.3,

$$\sqrt{n}(T_n^{Ind}-T^{Ind})=\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n}((X_i-\overline{X})(Y_i-\overline{Y})-\mathbb{E}((X-\mu_X)(Y-\mu_Y)))\right) \tag{5.7}$$

converges in law to $Z_{(X-\mu_X)(Y-\mu_Y)}$, regardless of whether H_0 holds or not. For the second one, the asymptotic limit established in Corollary 5.2.4 plays a fundamental role for defining appropriate bootstrap statistics: if H_0 is satisfied then (5.7) converges in law to $Z_{(X-\mu_X)\epsilon}$ with covariance operator $\sigma^2\Gamma$. Thus, (5.7) is an interesting statistic in order to be mimicked by a bootstrap one, and the consistency and correctness of this bootstrap statistic will be ensured if it converges to the above–mentioned limit distributions.

On the other hand, by the Strong Law of Large Numbers (SLLN), $n^{-1}\sum_{i=1}^{n}(Y_i-\overline{Y})^2$ converges $a.s.-\mathbb{P}$ to $\sigma_Y=\mathrm{Var}(Y)=\sigma^2+\mathbb{E}(\langle\theta,X-\mu_X\rangle^2)$ as $n\to\infty$. Furthermore, $n^{-1/2}\sum_{i=1}^{n}(X_i-\mu_X)$ converges in law to Z_X by the CLT for separable Hilbert–valued random elements. Therefore, the Slutsky's Theorem ensures that

$$\left(\frac{1}{n}\sum_{i=1}^{n}(Y_i-\overline{Y})^2\right)\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n}(X_i-\mu_X)\right) \tag{5.8}$$

converges in law to $(\sigma^2+\mathbb{E}(\langle\theta,X-\mu_X\rangle^2))Z_X$, whose covariance operator is $(\sigma^2+\mathbb{E}(\langle\theta,X-\mu_X\rangle^2)\Gamma$. In particular, when H_0 is satisfied, this operator reduces again to $\sigma^2\Gamma$. Therefore, another possibility consists in mimicking this second statistic by means of a bootstrap one, improving the approximation suggested in the previous section. Moreover, the left term in the product in (5.8) could be substituted by any other consistent estimator of σ^2 under H_0, which converges to a finite constant if H_0 does not hold. Anyway, this second approach could lead to worse results than (5.7) under the null hypothesis, because the possible dependence between X and ϵ is lost (since the resampling would focus only on the X information).

Two possibilities for mimicking the above–mentioned statistics are going to be explored, namely a *naive* paired bootstrap and a *wild* bootstrap approach.

Naive bootstrap. Let $\{(X_i^*, Y_i^*)\}_{i=1}^n$ be a collection of i.i.d. random elements drawn at random from $\{(X_i, Y_i)\}_{i=1}^n$. Then, the following *naive* paired bootstrap statistic can be defined

$$T_n^{Ind,N*} = \frac{1}{n}\sum_{i=1}^{n}((X_i^* - \overline{X}^*)(Y_i^* - \overline{Y}^*) - (X_i - \overline{X})(Y_i - \overline{Y})).$$

In addition, defining $\hat{\sigma}^2 = n^{-1}\sum_{i=1}^n (Y_i - \overline{Y})^2$ and

$$\hat{\sigma}^{*2} = \frac{1}{n}\sum_{i=1}^{n}(Y_i^* - \overline{Y}^*)^2,$$

that is, the empirical estimator of σ_Y^2 under H_0 and its corresponding bootstrap version, then the naive bootstrap approach is described in the following algorithm.

Algorithm 5.2.6 (Naive bootstrap)**.**

Step 1. *Compute the value of the statistic T_n^{Ind} (or the value of the statistic $T_n^{Ind}/\hat{\sigma}$).*

Step 2. *Draw $\{(X_i^*, Y_i^*)\}_{i=1}^n$, a sequence of i.i.d. random elements chosen at random from the initial sample $\{(X_i, Y_i)\}_{i=1}^n$, and compute $a_n = \|T_n^{Ind,N*}\|$ (or compute $b_n = \|T_n^{Ind,N*}\|/\hat{\sigma}^*$).*

Step 3. *Repeat Step 2 a large number of times $n_{boot} \in \mathbb{N}$ in order to obtain a sequence of values $\{a_n^l\}_{l=1}^{n_{boot}}$ (or a sequence of values $\{b_n^l\}_{l=1}^{n_{boot}}$).*

Step 4. *Approximate the p–value of the test by the proportion of values in $\{a_n^l\}_{l=1}^{n_{boot}}$ greater than or equal to $\|T_n^{Ind}\|$ (or by the proportion of values in $\{b_n^l\}_{l=1}^{n_{boot}}$ greater than or equal to $\|T_n^{Ind}\|/\hat{\sigma}$).*

The asymptotic behaviour of the naive bootstrap statistic will be obtained as a consequence of results on bootstrapping general empirical measures by Giné and Zinn (1990) (see Lemma 5.7.1, page 135). Hence, the following theorem establishes the asymptotic consistency and correctness of the naive bootstrap approach.

Theorem 5.2.7 (González-Manteiga et al., 2012)**.** *Under the assumptions of Theorem 5.2.3, it holds that $\sqrt{n}\,T_n^{Ind,N*}$ converges in law to $Z_{(X-\mu_X)(Y-\mu_Y)}$ a.s.$-\mathbb{P}$. In addition, $\hat{\sigma}^{*2}$ converges in probability to $\sigma_Y^2 = \sigma^2 + \mathbb{E}(\langle\theta, X - \mu_X\rangle^2)$ a.s.$-\mathbb{P}$.*

The proof of Theorem 5.2.7 can be found in Section 5.7.4 (see page 134).

Wild bootstrap. Analogously, let $\{\epsilon_i^*\}_{i=1}^n$ be i.i.d. centred real random variables, independent of $\{(X_i, Y_i)\}_{i=1}^n$, satisfying that $\mathbb{E}((\epsilon_i^*)^2) = 1$ and $\int_0^\infty \mathbb{P}(|\epsilon_1^*| > t)^{1/2}dt < \infty$. To guarantee this last assumption, it is enough that $\mathbb{E}((\epsilon_i^*)^d) < \infty$ for certain $d > 2$. Let $T_n^{Ind,W*}$ be the following *wild* bootstrap statistic

$$T_n^{Ind,W*} = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})\epsilon_i^*,$$

and let $\hat{\sigma}$ and $\hat{\sigma}^*$ be computed as in the naive bootstrap case. Hence, the proposed wild bootstrap approach can be applied by means of the following algorithm.

Algorithm 5.2.8 (Wild bootstrap)**.**

Step 1. *Compute the value of the statistic T_n^{Ind} (or the value of the statistic $T_n^{Ind}/\hat{\sigma}$).*

Step 2. *Draw $\{\epsilon_i^*\}_{i=1}^n$ a sequence of i.i.d. random elements drawn from a real random variable ϵ^* independent of $\{(X_i, Y_i)\}_{i=1}^n$, which satisfies $\mathbb{E}(\epsilon^*) = 0$, $\mathbb{E}((\epsilon^*)^2) = 1$ and $\int_0^\infty(\mathbb{P}(|\epsilon^*| > t)^{1/2}) < \infty$, and compute $a_n = \|T_n^{Ind,W*}\|$ (or compute $b_n = \|T_n^{Ind,W*}\|/\hat{\sigma}^*$).*

Step 3. *Repeat Step 2 a large number of times $n_{boot} \in \mathbb{N}$ in order to obtain a sequence of values $\{a_n^l\}_{l=1}^{n_{boot}}$ (or a sequence of values $\{b_n^l\}_{l=1}^{n_{boot}}$).*

Step 4. *Approximate the p–value of the test by the proportion of values in $\{a_n^l\}_{l=1}^{n_{boot}}$ greater than or equal to $\|T_n^{Ind}\|$ (or by the proportion of values in $\{b_n^l\}_{l=1}^{n_{boot}}$ greater than or equal to $\|T_n^{Ind}\|/\hat{\sigma}$).*

In order to analyse the asymptotic behaviour of the wild bootstrap statistic, Lemma 5.7.2 (see page 136) will be fundamental. As a result of this lemma, the asymptotic consistency and correctness of the wild bootstrap approach is guaranteed by the following theorem.

Theorem 5.2.9 (González-Manteiga et al., 2012)**.** *Under the assumptions of Theorem 5.2.3, it holds that $\sqrt{n}\, T_n^{Ind,W*}$ converges in law to $Z_{(X-\mu_X)(Y-\mu_Y)}$ a.s. $-\,\mathbb{P}$.*

The proof of the previous theorem can be found in the appendix of the chapter (see Section 5.7.6, page 136).

5.2.2 Bootstrap calibration vs. asymptotic theory

For simplicity, suppose from now on that $b = 0$ and $\mu_X = 0$ (consequently, $\mu_Y = 0$) in (5.1) (see page 106), that is, assume that the regression model is given by $Y = \langle \theta, X \rangle + \epsilon$ where both Y and X are zero–mean random elements valued in $\mathbb{R}$ and $\mathcal{H}$, respectively. In this situation, different statistics can be used for testing the lack of dependence between X and Y. Bearing in mind (5.5) (see page 109), one may think about using an estimator of θ in order to test the null hypothesis. In an alternative way, expression (5.6) (see page 109) can be a motivation for a different class of statistics based on the estimation of $\|\Delta\|_{\mathcal{H}'}$.

An asymptotic distribution free statistic based on the latter approach was given by Cardot et al. (2003b). They proposed as test statistic

$$T_{1,n}^{Ind} = \frac{1}{\sqrt{k_n}} \left(\frac{1}{\hat{\sigma}^2} \left\| \sqrt{n} \Delta_n \hat{A}_n \right\|_{\mathcal{H}'}^2 - k_n \right),$$

where $\{k_n\}_{n=1}^{\infty}$ is a sequence of positive integers such that $k_n \to +\infty$ and $k_n \leq n$, $\hat{\sigma}^2$ is a consistent estimator of σ^2 under H_0, $\Delta_n = n^{-1} \sum_{i=1}^n X_i \otimes_{\mathcal{H}'} Y_i$, and $\hat{A}_n = \sum_{j=1}^{k_n} \hat{\lambda}_j^{-1/2} \hat{v}_j \otimes_{\mathcal{H}} \hat{v}_j$ with $\{(\hat{\lambda}_j, \hat{v}_j)\}_{j=1}^{\infty}$ the eigenelements of $\Gamma_n = n^{-1} \sum_{i=1}^n X_i \otimes_{\mathcal{H}} X_i$. Therefore, note that, in fact,

$$T_{1,n}^{Ind} = \frac{1}{\sqrt{k_n}} \left(\frac{n}{\hat{\sigma}^2} \sum_{j=1}^{k_n} \frac{(\Delta_n \hat{v}_j)^2}{\hat{\lambda}_j} - k_n \right). \tag{5.9}$$

Cardot et al. (2003b) showed that, under certain conditions on k_n and $\hat{\sigma}^2$, $T_{1,n}^{Ind}$ converges in distribution, under H_0 in (5.6), to a centred Gaussian variable with variance equal to 2. Hence, H_0 is rejected if $|T_{1,n}^{Ind}| > \sqrt{2} z_{1-\alpha/2}$, where z_α is the α–quantile of a $\mathcal{N}(0,1)$; otherwise, there is no evidence to reject the null hypothesis. Besides, Cardot et al. (2003b) also proposed another calibration of the statistic distribution based on a permutation mechanism.

As it was already commented, another possibility is to try to test (5.5), or equivalently, to test whether $\|\theta\|^2 = 0$ or not. For this purpose, recall some issues related with the standard FPCA estimator introduced in Section 2.3.2, *"a) Definition of standard FPCA estimator"*, in Chapter 2 (see page 35). Assuming that **(C.2.1)** holds (see Chapter 2, page 36), then $\theta = \sum_{j=1}^{\infty} \lambda_j^{-1} \Delta v_j v_j$ and, consequently, $\|\theta\|^2 = \sum_{j=1}^{\infty} (\lambda_j^{-1} \Delta v_j)^2$. A natural estimator for θ is the FPCA estimator based on the first k_n functional principal components (2.4) (see Chapter 2, page 36) given by $\hat{\theta}_{k_n} = \sum_{j=1}^{k_n} \hat{\lambda}_j^{-1} \Delta_n \hat{v}_j \hat{v}_j$, where $\Delta_n = n^{-1} \sum_{i=1}^n X_i \otimes_{\mathcal{H}'} Y_i$, $\{(\hat{\lambda}_j, \hat{v}_j)\}_{j=1}^{\infty}$ are the eigenelements of $\Gamma_n = n^{-1} \sum_{i=1}^n X_i \otimes_{\mathcal{H}} X_i$, and $\{k_n\}_{n=1}^{\infty}$ is a sequence of positive integers such that $k_n \to +\infty$, $k_n \leq n$, and $\hat{\lambda}_{k_n} > 0$. Since

$\|\hat{\theta}_{k_n}\|^2 = \sum_{j=1}^{k_n} (\hat{\lambda}_j^{-1} \Delta_n \hat{v}_j)^2$ is an estimator of $\|\theta\|^2$, in order to test the lack of dependence between X and Y one may consider the statistic

$$T_{2,n}^{Ind} = \sum_{j=1}^{k_n} \left(\frac{\Delta_n \hat{v}_j}{\hat{\lambda}_j} \right)^2. \tag{5.10}$$

Unfortunately, the limit distribution of (5.10) is currently unknown, so further research about this statistic is necessary in order to obtain it. Meanwhile, its asymptotic distribution can be calibrated using the bootstrap approach proposed below.

Finally, another statistic is the one based on ideas proposed throughout Section 5.2.1, that is,

$$T_{3,n}^{Ind} = \left\| \frac{1}{n} \sum_{i=1}^{n} (X_i - \overline{X})(Y_i - \overline{Y}) \right\|, \tag{5.11}$$

which will be denoted by *F–test* from now on, since it is the natural generalization of the well–known F–test in the finite dimensional context. Another possibility is to consider the *studentized* version of (5.11) defined by

$$T_{3s,n}^{Ind} = \frac{1}{\hat{\sigma}} \left\| \frac{1}{n} \sum_{i=1}^{n} (X_i - \overline{X})(Y_i - \overline{Y}) \right\|, \tag{5.12}$$

where $\hat{\sigma}^2$ is a consistent estimator of σ^2 (at least under H_0), for instance, the empirical estimation of σ^2 (recall that studentized versions were also computed for confidence intervals; see (4.5) and (4.6) in Chapter 4, page 93).

In general, for the statistics such as (5.9), (5.10), (5.11) and (5.12), the calibration of the distribution can be obtained by bootstrap. For instance, for (5.11) and (5.12), the following bootstrap statistics can be considered

$$T_{3,n}^{Ind,*} = \left\| \frac{1}{n} \sum_{i=1}^{n} (X_i - \overline{X})(Y_i - \overline{Y})\epsilon_i^* \right\|, \tag{5.13}$$

$$T_{3s,n}^{Ind,*} = \frac{1}{\hat{\sigma}^*} \left\| \frac{1}{n} \sum_{i=1}^{n} (X_i - \overline{X})(Y_i - \overline{Y})\epsilon_i^* \right\|, \tag{5.14}$$

where $\{\epsilon_i^*\}_{i=1}^n$ and $\hat{\sigma}^*$ are built following the wild bootstrap approach described in Algorithm 5.2.8 (see page 112). Furthermore, in the previous section, both naive and wild bootstrap were shown to be consistent for the F–test. This fact guarantees from a theoretical point of view that the distribution of $T_{3,n}^{Ind}$ and $T_{3s,n}^{Ind}$ can be approximated by the corresponding bootstrap distribution of (5.13) and (5.14), and H_0 can be rejected when the approximated p–value of the statistic is smaller than α.

A similar bootstrap calibration is proposed for the tests based on $T_{1,n}^{Ind}$ and $T_{2,n}^{Ind}$. Although the consistency of the bootstrap procedure in these cases has not been proved in this chapter. For (5.9), two bootstrap statistics are built

$$T_{1,n}^{Ind,*(a)} = \frac{1}{\sqrt{k_n}} \left(\frac{n}{\hat{\sigma}^{*2}} \sum_{j=1}^{k_n} \frac{(\Delta_n^* \hat{v}_j)^2}{\hat{\lambda}_j} - k_n \right), \tag{5.15}$$

$$T_{1,n}^{Ind,*(b)} = \frac{1}{\sqrt{k_n}} \left(\frac{n}{\hat{\sigma}^{2}} \sum_{j=1}^{k_n} \frac{(\Delta_n^* \hat{v}_j)^2}{\hat{\lambda}_j} - k_n \right). \tag{5.16}$$

The difference between the two proposed bootstrap approximations is that in the first one the estimation of σ^2, denoted by $\hat{\sigma}^{*2}$, is computed using the bootstrap sample generated in each iteration. On the other hand, for (5.10), the proposed bootstrap statistic is

$$T_{2,n}^{Ind,*} = \sum_{j=1}^{k_n} \left(\frac{\Delta_n^* \hat{v}_j}{\hat{\lambda}_j} \right)^2. \tag{5.17}$$

In order to obtain the bootstrap distribution of (5.15), (5.16) and (5.17), the following wild bootstrap algorithm can be used.

Algorithm 5.2.10 (Wild bootstrap)**.**

Step 1. *Compute the value of the statistic $T_{1,n}^{Ind}$ (or the value of the statistic $T_{2,n}^{Ind}$).*

Step 2. *Draw $\{\epsilon_i^*\}_{i=1}^n$ a sequence of i.i.d. random elements drawn from ϵ^*, which satisfies $\mathbb{E}(\epsilon^*) = 0$, $\mathbb{E}((\epsilon^*)^2) = 1$ and $\int_0^\infty (\mathbb{P}(|\epsilon^*| > t)^{1/2}) < \infty$, and define $Y_i^* = Y_i \epsilon_i^*$ for all $i = 1, \ldots, n$.*

Step 3. *Build $\Delta_n^* = n^{-1} \sum_{i=1}^n X_i \otimes_{\mathcal{H}'} Y_i^*$, and compute $a_n = |T_{1,n}^{Ind,*}|$ (or compute $b_n = |T_{2,n}^{Ind,*}|$).*

Step 4. *Repeat Steps 2 and 3 a large number of times $n_{boot} \in \mathbb{N}$ in order to obtain a sequence of values $\{a_n^l\}_{l=1}^{n_{boot}}$ (or a sequence of values $\{b_n^l\}_{l=1}^{n_{boot}}$).*

Step 5. *Approximate the p–value of the test by the proportion of values in $\{a_n^l\}_{l=1}^{n_{boot}}$ greater than or equal to $|T_{1,n}^{Ind}|$ (or by the proportion of values in $\{b_n^l\}_{l=1}^{n_{boot}}$ greater than or equal to $|T_{2,n}^{Ind}|$).*

5.3 Test for equality of linear models

As stated in the previous section, let $(\mathcal{H}, \langle \cdot, \cdot \rangle)$ be a separable Hilbert space (being $\| \cdot \|$ its associated norm), and let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space. In this case, let (X_1, Y_1) and (X_2, Y_2) be two measurable mappings from Ω to $\mathcal{H} \times \mathbb{R}$ (i.e., X_1 and X_2 are $\mathcal{H}$–valued random elements and Y_1 and Y_2 are real random variables) such that both of them satisfy the functional linear model with scalar response as follows

$$\begin{cases} Y_1 = \langle \theta_1, X_1 \rangle + b_1 + \epsilon_1, \\ Y_2 = \langle \theta_2, X_2 \rangle + b_2 + \epsilon_2, \end{cases} \tag{5.18}$$

where $\theta_1, \theta_2 \in \mathcal{H}$ are the fixed functional model parameters, $b_1, b_2 \in \mathbb{R}$ are the intercept terms, and ϵ_1 and ϵ_2 are real random variables such that $\mathbb{E}(\epsilon_1) = \mathbb{E}(\epsilon_2) = 0$, $\mathbb{E}(\epsilon_1^2) = \sigma_1^2 < \infty$, $\mathbb{E}(\epsilon_2^2) = \sigma_2^2 < \infty$, and $\mathbb{E}(\epsilon_1 X_1) = \mathbb{E}(\epsilon_2 X_2) = 0$.

This section is focused on the introduction of a bootstrap calibration procedure to approximate the distribution of statistics for testing the equality of two functional linear models with scalar response, that is, for testing $H_0 : \theta_1 = \theta_2$ versus $H_1 : \theta_1 \neq \theta_2$. Simple random samples $\{(X_{1,i}, Y_{1,i})\}_{i=1}^{n_1}$ and $\{(X_{2,i}, Y_{2,i})\}_{i=1}^{n_2}$ drawn from (X_1, Y_1) and (X_2, Y_2), which are assumed to be independent, will be used in order to achieve this objective. Note that bootstrap methods can be an especially interesting option when the asymptotic distribution of a test statistic is hard to compute or it has not an appropriate behaviour for small sample sizes.

The problem of checking the equality of two functional linear models has barely been studied in FDA literature. In this sense, the paper by Horváth et al. (2009) is the most noteworthy contribution to this issue. The authors compared two functional linear models in which explanatory variables are curves and responses can be either scalars or curves, and they tested the null hypothesis that the two regression operators are the same. The proposed test statistics have asymptotic chi–squared distribution. Hence, they used this distribution in practice, when they applied their methodology to the two real data applications considered in their work.

In this section, a test statistics for assesing the equality of the linear models in (5.18) is proposed and a bootstrap algorithm is introduced to approximate its distribution. Section 5.3.1 compiles theoretical background for the regression model (5.18), describes the testing procedure and its asymptotic properties, and presents the bootstrap techniques and their associated consistency and correctness results. Next, Section 5.3.2 is devoted to compare the bootstrap calibration and the asymptotic one. Recall that a simulation study and a real data application related to this testing problem can be found in Section 5.4.2 and Section 5.5.2, respectively.

5.3.1 Asymptotic theory for testing and bootstrap procedures

Theoretical background

First of all, as it was done for the test of lack of dependence, the special case where $\mathcal{H} = \mathcal{L}^2(C, \lambda)$, i.e., the separable Hilbert space of square Lebesgue integrable functions on a given compact set $C \subset \mathbb{R}$ with the usual inner product and norm, will be analysed from a theoretical viewpoint. In this situation, the linear dependence of the scalar responses Y_1 and Y_2 on the functional random predictors X_1 and X_2 is modelled by

$$\begin{cases} Y_1 = \Psi_1(X_1) + \epsilon_1, \\ Y_2 = \Psi_2(X_2) + \epsilon_2, \end{cases} \tag{5.19}$$

where $\Psi_1, \Psi_2 \in \mathcal{H}'$ are continuous linear operators (for further information about the dual space $\mathcal{H}'$ and its associated norm, see Section 1.2.2, *"b) The dual space $\mathcal{H}'$ "*, in Chapter 1, page 10), and ϵ_1 and ϵ_2 are zero–mean real random variables with finite variance and independent of X_1 and X_2, respectively. Riesz Representation Theorem states that $\mathcal{H}$ and $\mathcal{H}'$ are isometrically identified, and there exist unique $\theta_1, \theta_2 \in \mathcal{H}$ satisfying that $\|\theta_1\| = \|\Psi_1\|_{\mathcal{H}'}$, $\|\theta_2\| = \|\Psi_2\|_{\mathcal{H}'}$, and $\Psi_1(x) = \langle \theta_1, x \rangle$ and $\Psi_2(x) = \langle \theta_2, x \rangle$ for all $x \in \mathcal{H}$. This fact ensures that (5.19) can be seen as a particular case of (5.18).

Horváth et al. (2009) assumed that the intercept terms b_1 and b_2 are equal to zero. The intercept terms in (5.18) can be embedded in the variable counterpart of the models if one defines $\tilde{X}_1 = (X_1, 1)$, $\tilde{X}_2 = (X_2, 1)$, $\tilde{\theta}_1 = (\theta_1, b_1) \in \mathcal{H}_e$ and $\tilde{\theta}_2 = (\theta_2, b_2) \in \mathcal{H}_e$, where $\mathcal{H}_e$ is the product space $\mathcal{H} \times \mathbb{R}$ with the corresponding inner product $\langle \cdot, \cdot \rangle_e$. Hence, the two models in (5.18) can be expressed as $Y_1 = \langle \tilde{\theta}_1, \tilde{X}_1 \rangle_e + \epsilon_1$ and $Y_2 = \langle \tilde{\theta}_2, \tilde{X}_2 \rangle_e + \epsilon_2$, respectively. In this case, neither $\tilde{X}_1$ nor $\tilde{X}_2$ can be assumed to be zero–mean random elements. Given that testing $\theta_1 = \theta_2$ is not equivalent to testing $\tilde{\theta}_1 = \tilde{\theta}_2$, since the intercept terms cannot be assumed to be zero in practice, both b_1 and b_2 has been written explicitly in the following. Furthermore, Horváth et al. (2009) also supposed that X_1 and X_2 are mean zero random elements. Given that this condition could be too restrictive in some practical cases, the predictor variables are not assumed to be centred in the following theoretical results.

Equality of two linear models test

Given the two linear models defined in (5.18) (see page 115), the aim is to build a correct and consistent bootstrap method that allows testing

$$\begin{cases} H_0: & \theta_1 = \theta_2 \\ H_1: & \theta_1 \neq \theta_2 \end{cases} \tag{5.20}$$

by means of two independent random samples, denoted by $\{(X_{1,i}, Y_{1,i})\}_{i=1}^{n_1}$ and $\{(X_{2,i}, Y_{2,i})\}_{i=1}^{n_2}$, of i.i.d. random elements drawn from (X_1, Y_1) and (X_2, Y_2), respectively. In order to obtain the theoretical results compiled below, it will be assumed that the two samples have the same size, i.e.,

(C.5.3) $n_1 = n_2 = n$, with $n \to \infty$.

Remark 5.3.1. Assumption **(C.5.3)**, which is a severe restriction, has been introduced in order to simplify the notation and the calculations involved in the theoretical results of this section. However, the developments in the proofs seem to indicate that **(C.5.3)** could be replaced by assuming that there exists a constant $0 < c < \infty$ such that $n_1/n_2 \to c$ when $n_1, n_2 \to \infty$. Nevertheless, additional messy notation and tedious work will be surely required for this. Due to this fact, **(C.5.3)** has been considered in this chapter, although alternative (and less restrictive) assumptions will be studied in detail in future research.

From now on, assume that $\mathbb{E}(\|X_1\|^2) < \infty$ and $\mathbb{E}(\|X_2\|^2) < \infty$. Consequently, $\mathbb{E}(Y_1^2) < \infty$ and $\mathbb{E}(Y_2^2) < \infty$ by the Hölder's inequality. Furthermore, suppose that the covariance operator of X_1 and the covariance operator X_2 are equal, and the errors ϵ_1 and ϵ_2 have equal variances, that is,

(C.5.4) $\mathbb{E}((X_1 - \mu_{X_1}) \otimes_{\mathcal{H}} (X_1 - \mu_{X_1})) = \mathbb{E}((X_2 - \mu_{X_2}) \otimes_{\mathcal{H}} (X_2 - \mu_{X_2})) = \Gamma_X$, and $\sigma_1^2 = \sigma_2^2 = \sigma^2$,

where $\mu_{X_1} \in \mathcal{H}$ and $\mu_{X_2} \in \mathcal{H}$ denote the expected values of X_1 and X_2, respectively. Whenever there is no possible confusion, Γ_X will be abbreviated as Γ. As usual, $\{(\lambda_j, v_j)\}_{j=1}^{\infty}$ will denote the eigenvalues and eigenfunctions of Γ, where the eigenvalues are assumed to be arranged in decreasing order ($\lambda_1 \geq \lambda_2 \geq \ldots$). Moreover, the cross–covariance operator between X_1 and Y_1, and the cross–covariance operator between X_2 and Y_2 are given by

$$\Delta_{X_1,Y_1} = \mathbb{E}((X_1 - \mu_{X_1}) \otimes_{\mathcal{H}'} (Y_1 - \mu_{Y_1})) \quad \text{and} \quad \Delta_{X_2,Y_2} = \mathbb{E}((X_2 - \mu_{X_2}) \otimes_{\mathcal{H}'} (Y_2 - \mu_{Y_2}))$$

where $\mu_{X_1}, \mu_{X_2} \in \mathcal{H}$ denote the expected value of X_1 and X_2, and $\mu_{Y_1}, \mu_{Y_2} \in \mathbb{R}$ denote the expected value of Y_1 and Y_2. The cross–covariance operators Δ_{X_1,Y_1} and Δ_{X_2,Y_2} will be denoted by Δ_1 and Δ_2, respectively, in order to simplify the notation. It can be shown that $\Delta_1, \Delta_2 \in \mathcal{H}'$ and, in addition, $\Delta_1 x = \langle \theta_1, \Gamma x \rangle$ and $\Delta_2 x = \langle \theta_2, \Gamma x \rangle$ for all $x \in \mathcal{H}$. Thus, the following relation between the covariance operator, the cross–covariance operators and the regression parameters is satisfied

$$(\Delta_1 - \Delta_2)x = \langle \theta_1 - \theta_2, \Gamma x \rangle, \quad \forall x \in \mathcal{H}. \tag{5.21}$$

As previously commented in this chapter, the Hilbert space $\mathcal{H}$ can be seen as the direct sum of two orthogonal subspaces induced by the self–adjoint operator Γ: the kernel of Γ denoted by $\text{Ker}(\Gamma)$, and the closure of the image of Γ denoted by $\overline{\text{Im}(\Gamma)}$. Therefore, θ_1 and θ_2 are determined uniquely by $\theta_1 = \theta_{1,1} + \theta_{1,2}$ and $\theta_2 = \theta_{2,1} + \theta_{2,2}$, where $\theta_{1,1}, \theta_{2,1} \in \text{Ker}(\Gamma)$ and $\theta_{1,2}, \theta_{2,2} \in \overline{\text{Im}(\Gamma)}$. It can be shown that $\text{Var}(\langle \theta_{1,1}, X_1 \rangle) = \text{Var}(\langle \theta_{2,1}, X_2 \rangle) = 0$, and as a result (5.18) can be rewritten as

$$\begin{cases} Y_1 = \langle \theta_{1,2}, X_1 \rangle + \langle \theta_{1,1}, \mu_{X_1} \rangle + b_1 + \epsilon_1, \\ Y_2 = \langle \theta_{2,2}, X_2 \rangle + \langle \theta_{2,1}, \mu_{X_2} \rangle + b_2 + \epsilon_2. \end{cases}$$

Consequently, it will be unfeasible to test whether $\theta_{1,1} = \theta_{2,1}$ or not, since $\langle \theta_{1,1}, \mu_{X_1} \rangle$ is indistinguishable from the intercept term b_1 and, analogously, $\langle \theta_{2,1}, \mu_{X_2} \rangle$ is indistinguishable from b_2. Consequently, the hypothesis test will be restricted to check

$$\begin{cases} H_0 : & \theta_{1,2} = \theta_{2,2} \\ H_1 : & \theta_{1,2} \neq \theta_{2,2}. \end{cases} \tag{5.22}$$

Furthermore, due to (5.21), one gets that $\theta_{1,2} = \theta_{2,2}$ if, and only if, $\Delta_1 x = \Delta_2 x$ for all $x \in \mathcal{H}$. Hence, the hypothesis test in (5.22) is equivalent to

$$\begin{cases} H_0 : & \|\Delta_1 - \Delta_2\|_{\mathcal{H}'} = 0 \\ H_1 : & \|\Delta_1 - \Delta_2\|_{\mathcal{H}'} \neq 0. \end{cases} \tag{5.23}$$

Remark 5.3.2. Horváth et al. (2009) assumed that $\mu_{X_1} = \mu_{X_2} = 0$. This assumption jointly with the previous reasoning imply that neither $\theta_{1,1}$ nor $\theta_{2,1}$ can be estimated based on the information provided by X_1 and X_2, so the hypothesis testing is also restricted to (5.22) (or, equivalently, to (5.23)).

Remark 5.3.3. If $b_1 = b_2 = 0$ and both μ_{X_1} and μ_{X_2} are different from zero, one could test if $\theta_{1,1} = \theta_{2,1}$ by checking whether the intercept terms in the models are equal or not. However, this is still an open problem in the literature, which will not be solved in this section since it has been focused on the restricted test (5.22) and not on the unrestricted test (5.20).

Testing procedure and asymptotic theory

It is straightforward to show that $\|\Delta_1 - \Delta_2\|_{\mathcal{H}'}$ can be expressed depending on the $\mathcal{H}$–valued random element

$$T^{Eq} = \mathbb{E}((X_1 - \mu_{X_1})(Y_1 - \mu_{Y_1}) - (X_2 - \mu_{X_2})(Y_2 - \mu_{Y_2}))$$

as follows

$$\begin{aligned} \|\Delta_1 - \Delta_2\|_{\mathcal{H}'} &= \|\mathbb{E}((X_1 - \mu_{X_1}) \otimes_{\mathcal{H}'} (Y_1 - \mu_{Y_1}) - (X_2 - \mu_{X_2}) \otimes_{\mathcal{H}'} (Y_2 - \mu_{Y_2}))\|_{\mathcal{H}'} \\ &= \|\mathbb{E}((X_1 - \mu_{X_1})(Y_1 - \mu_{Y_1}) - (X_2 - \mu_{X_2})(Y_2 - \mu_{Y_2}))\| = \|T^{Eq}\|. \end{aligned}$$

Then, given two independent random samples $\{(X_{1,i}, Y_{1,i})\}_{i=1}^n$ and $\{(X_{2,i}, Y_{2,i})\}_{i=1}^n$ of i.i.d. random elements drawn from (X_1, Y_1) and (X_2, Y_2), respectively, $\|T^{Eq}\|$ can be estimated by means of its empirical counterpart $\|T_n^{Eq}\|$, where T_n^{Eq} is the following $\mathcal{H}$–valued random element

$$T_n^{Eq} = \frac{1}{n}\sum_{i=1}^{n}((X_{1,i} - \overline{X}_1)(Y_{1,i} - \overline{Y}_1) - (X_{2,i} - \overline{X}_2)(Y_{2,i} - \overline{Y}_2)),$$

with $\overline{X}_1 = n^{-1}\sum_{i=1}^n X_{1,i}$, $\overline{X}_2 = n^{-1}\sum_{i=1}^n X_{2,i}$, $\overline{Y}_1 = n^{-1}\sum_{i=1}^n Y_{1,i}$, and $\overline{Y}_2 = n^{-1}\sum_{i=1}^n Y_{2,i}$. Some properties of T_n^{Eq} are stated in the next theorem and corollary, being the next assumption necessary:

(C.5.5) $\mathbb{E}(\|X_1\|^4) < \infty$ and $\mathbb{E}(\|X_2\|^4) < \infty$.

Furthermore, recall that, given an $\mathcal{H}$–valued random element H such that $\mathbb{E}(\|H\|^2) < \infty$, Z_H denotes a centred Gaussian element in $\mathcal{H}$ with covariance operator Γ_H.

Theorem 5.3.4. *Assuming that* (5.18)*,* **(C.5.3)***,* **(C.5.4)** *and* **(C.5.5)** *hold, then*

(i) $\mathbb{E}(T_n^{Eq}) = n^{-1}(n-1)T^{Eq}$.

(ii) T_n^{Eq} *converges a.s.* $-\mathbb{P}$ *to* T^{Eq} *as* $n \to \infty$.

(iii) $\sqrt{n}(T_n^{Eq} - T^{Eq})$ *converges in law, as* $n \to \infty$*, to* $Z_{(X_1-\mu_{X_1})(Y_1-\mu_{Y_1})-(X_2-\mu_{X_2})(Y_2-\mu_{Y_2})}$ *with covariance operator* $2\sigma^2\Gamma + \mathbb{E}(\langle\theta_1, X_1-\mu_{X_1}\rangle^2(X_1-\mu_{X_1})\otimes_{\mathcal{H}}(X_1-\mu_{X_1})) + \mathbb{E}(\langle\theta_2, X_2-\mu_{X_2}\rangle^2(X_2-\mu_{X_2})\otimes_{\mathcal{H}}(X_2-\mu_{X_2}))$.

Corollary 5.3.5. *Under the assumptions of Theorem 5.3.4, if the null hypothesis in* (5.23) *is satisfied (i.e.,* $\|\Delta_1 - \Delta_2\|_{\mathcal{H}'} = 0$*), then* $\sqrt{n}\,T_n^{Eq}$ *converges in law to* $Z_{(X_1-\mu_{X_1})(Y_1-\mu_{Y_1})-(X_2-\mu_{X_2})(Y_2-\mu_{Y_2})}$ *with covariance operator* $2\sigma^2\Gamma + \mathbb{E}(\langle\theta_1, X_1-\mu_{X_1}\rangle^2(X_1-\mu_{X_1})\otimes_{\mathcal{H}}(X_1-\mu_{X_1})) + \mathbb{E}(\langle\theta_1, X_2-\mu_{X_2}\rangle^2(X_2-\mu_{X_2})\otimes_{\mathcal{H}}(X_2-\mu_{X_2}))$*, and consequently,* $\|\sqrt{n}\,T_n^{Eq}\|$ *converges in law to* $\|Z_{(X_1-\mu_{X_1})(Y_1-\mu_{Y_1})-(X_2-\mu_{X_2})(Y_2-\mu_{Y_2})}\|$.

The proofs of the previous results can be found in the appendix of the chapter (see Section 5.7.8, page 137, and Section 5.7.9, page 137, respectively).

Although the asymptotic null distribution of $\|\sqrt{n}\,T_n^{Eq}\|$ cannot be explicitly calculated from Corollary 5.3.5, this statistic could be used in practice approximating its distribution by means of bootstrap methods. Nevertheless, an alternative statistic will be considered in the simulation study for comparative purposes.

Local alternatives. To finish the asymptotic analysis of the statistic $\|\sqrt{n}\,T_n^{Eq}\|$, it is necessary to study its behaviour under local alternatives. For this purpose, it will be assumed that $\theta_1, \theta_2 \in \mathcal{H}$ satisfy that $\|\theta_{1,2}\| > 0$ and $\theta_2 = (1 - \delta_n/\sqrt{n})\theta_1$, where δ_n is a positive sequence such that **(C.5.2)** holds (see page 110). Then, the modified random sample

$$\begin{cases} Y_{1,i} = \langle\theta_1, X_{1,i}\rangle + b_1 + \epsilon_{1,i}, & \forall i \in \{1,\ldots,n\}, \\ Y_{2,i}^n = \langle(1 - \delta_n/\sqrt{n})\theta_1, X_{2,i}\rangle + b_2 + \epsilon_{2,i}, & \forall i \in \{1,\ldots,n\}, \end{cases}$$

can be considered. Obviously, H_0 is not verified in this case. However, $\|\theta_1 - (1 - \delta_n/\sqrt{n})\theta_1\| = \delta_n\|\theta_1\|/\sqrt{n} \to 0$, so the null hypothesis is approached at rate $\delta_n/\sqrt{n}$. The behaviour of the proposed statistic under this kind of local alternatives is stated in the following theorem.

Theorem 5.3.6. *Under the assumptions of Theorem 5.3.4, and with the above notation, if* **(C.5.2)** *holds, then*

$$\mathbb{P}\left(\left\|\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n}((X_{1,i}-\overline{X}_1)(Y_{1,i}-\overline{Y}_1)-(X_{2,i}-\overline{X}_2)(Y_{2,i}^n-\overline{Y}_2^n))\right)\right\|\leq t\right)\to 0,\quad \forall t\in\mathbb{R}$$

as $n\to\infty$.

Its proof is collected in Section 5.7.10 (see page 138).

Bootstrap procedures and consistency

As an alternative to using the asymptotic distributions introduced in the previous section, bootstrap procedures are proposed below in order to check the null hypothesis in (5.23).

First of all, let $\sqrt{n}(T_n^{Eq}-T^{Eq})$ be the statistic given by

$$\begin{aligned}\sqrt{n}(T_n^{Eq}-T^{Eq})=\sqrt{n}\Bigg(&\frac{1}{n}\sum_{i=1}^{n}((X_{1,i}-\overline{X}_1)(Y_{1,i}-\overline{Y}_1)-(X_{2,i}-\overline{X}_2)(Y_{2,i}-\overline{Y}_2)\\&-\mathbb{E}((X_1-\mu_{X_1})(Y_1-\mu_{Y_1})-(X_2-\mu_{X_2})(Y_2-\mu_{Y_2})))\Bigg).\end{aligned}\tag{5.24}$$

By Theorem 5.3.4, the previous statistic converges in law to a centred Gaussian element with covariance operator $2\sigma^2\Gamma+\mathbb{E}(\langle\theta_1,X_1-\mu_{X_1}\rangle^2(X_1-\mu_{X_1})\otimes_{\mathcal{H}}(X_1-\mu_{X_1}))+\mathbb{E}(\langle\theta_2,X_2-\mu_{X_2}\rangle^2(X_2-\mu_{X_2})\otimes_{\mathcal{H}}(X_2-\mu_{X_2}))$, regardless H_0 holds or not. In addition, if H_0 is satisfied, Corollary 5.3.5 ensures that (5.24) converges in law to a centred Gaussian element with covariance operator $2\sigma^2\Gamma+\mathbb{E}(\langle\theta_1,X_1-\mu_{X_1}\rangle^2(X_1-\mu_{X_1})\otimes_{\mathcal{H}}(X_1-\mu_{X_1}))+\mathbb{E}(\langle\theta_1,X_2-\mu_{X_2}\rangle^2(X_2-\mu_{X_2})\otimes_{\mathcal{H}}(X_2-\mu_{X_2}))$.

Consequently, this statistic seems to be a good bet for being mimicked by a bootstrap one. The consistency and correctness of the bootstrap statistic will be guaranteed whenever its limit distribution irrespectively of H_0 and its limit distribution under H_0 coincide with the above–mentioned asymptotic distributions. Next, a naive paired bootstrap and a wild bootstrap are proposed and their asymptotic properties are studied.

Naive bootstrap. Let $\{(X_{1,i}^*,Y_{1,i}^*)\}_{i=1}^n$ and $\{(X_{2,i}^*,Y_{2,i})^*\}_{i=1}^n$ two independent collections of i.i.d. random elements drawn at random from (X_1,Y_1) and (X_2,Y_2), respectively. Then, the next *naive* paired bootstrap statistic can be considered

$$\begin{aligned}T_n^{Eq,N*}=\frac{1}{n}\sum_{i=1}^{n}&((X_{1,i}^*-\overline{X}_1^*)(Y_{1,i}^*-\overline{Y}_1^*)-(X_{2,i}^*-\overline{X}_2^*)(Y_{2,i}^*-\overline{Y}_2^*)\\&-(X_{1,i}-\overline{X}_1)(Y_{1,i}-\overline{Y}_1)+(X_{2,i}-\overline{X}_2)(Y_{2,i}-\overline{Y}_2)).\end{aligned}$$

The following algorithm describes in detail the naive bootstrap approach.

Algorithm 5.3.7 (Naive bootstrap)**.**

Step 1. *Compute the value of the statistic* T_n^{Eq}.

Step 2. *Draw* $\{(X_{1,i}^*,Y_{1,i}^*)\}_{i=1}^n$ *and* $\{(X_{2,i}^*,Y_{2,i})^*\}_{i=1}^n$*, two independent sequences of i.i.d. random elements chose at random from the initial samples* (X_1,Y_1) *and* (X_2,Y_2)*, respectively, and compute* $a_n=\|T_n^{Eq,N*}\|$.

Step 3. *Repeat Step 2 a large number of times* $n_{boot}\in\mathbb{N}$ *in order to obtain a sequence of values* $\{a_n^l\}_{l=1}^{n_{boot}}$.

Step 4. *Approximate the p–value of the test by the proportion of values in* $\{a_n^l\}_{l=1}^{n_{boot}}$ *greater than or equal to* $\|T_n^{Eq}\|$.

The asymptotic consistency and correctness of the naive bootstrap approach is stated in the following theorem.

Theorem 5.3.8. *Under the assumptions of Theorem 5.3.4, it holds that* $\sqrt{n}\,T_n^{Eq,N*}$ *converges in law to* $Z_{(X_1-\mu_{X_1})(Y_1-\mu_{Y_1})-(X_2-\mu_{X_2})(Y_2-\mu_{Y_2})}$ *a.s.* $-\,\mathbb{P}$.

The proof of Theorem 5.3.8 can be found in Section 5.7.11 (see page 138).

Wild bootstrap. Let $\{\epsilon_i^*\}_{i=1}^n$ be i.i.d. centred real random variables, independent of $\{(X_{1,i}, Y_{1,i})\}_{i=1}^n$ and $\{(X_{2,i}, Y_{2,i})\}_{i=1}^n$, such that $\mathbb{E}((\epsilon_i^*)^2) = 1$ and $\int_0^\infty \mathbb{P}(|\epsilon_1^*| > t)^{1/2} dt < \infty$ (for instance, $\mathbb{E}((\epsilon_i^*)^d) < \infty$ for certain $d > 2$ ensures the second condition). Then, a *wild* bootstrap statistic can be defined as follows

$$T_n^{Eq,W*} = \frac{1}{n}\sum_{i=1}^{n}((X_{1,i} - \overline{X}_1)(Y_{1,i} - \overline{Y}_1) - (X_{2,i} - \overline{X}_2)(Y_{2,i} - \overline{Y}_2))\epsilon_i^*.$$

Thus, the wild bootstrap approach can be applied by means of the following algorithm.

Algorithm 5.3.9 (Wild bootstrap)**.**

Step 1. *Compute the value of the statistic* T_n^{Eq}.

Step 2. *Draw* $\{\epsilon_i^*\}_{i=1}^n$ *a sequence of i.i.d. random elements drawn from a real random variable* ϵ^* *independent of* $\{(X_{1,i}, Y_{1,i})\}_{i=1}^n$ *and* $\{(X_{2,i}, Y_{2,i})\}_{i=1}^n$*, which satisfies* $\mathbb{E}(\epsilon^*) = 0$*,* $\mathbb{E}((\epsilon^*)^2) = 1$ *and* $\int_0^\infty (\mathbb{P}(|\epsilon^*| > t)^{1/2}) < \infty$*, and compute* $a_n = \|T_n^{Eq,W*}\|$.

Step 3. *Repeat Step 2 a large number of times* $n_{boot} \in \mathbb{N}$ *in order to obtain a sequence of values* $\{a_n^l\}_{l=1}^{n_{boot}}$.

Step 4. *Approximate the p–value of the test by the proportion of values in* $\{a_n^l\}_{l=1}^{n_{boot}}$ *greater than or equal to* $\|T_n^{Eq}\|$.

The asymptotic behaviour of the wild bootstrap statistic is analysed in the following theorem.

Theorem 5.3.10. *Under the assumptions of Theorem 5.3.4, it holds that* $\sqrt{n}\,T_n^{Eq,W*}$ *converges in law to* $Z_{(X_1-\mu_{X_1})(Y_1-\mu_{Y_1})-(X_2-\mu_{X_2})(Y_2-\mu_{Y_2})}$ *a.s.* $-\,\mathbb{P}$.

See the proof of the previous theorem in Section 5.7.12, page 139.

5.3.2 Bootstrap calibration vs. asymptotic theory

From now on, it is assumed that the model (5.18) (see page 115) holds with $b_1 = b_2 = 0$ and $\mu_{X_1} = \mu_{X_2} = 0$ (thus, $\mu_{Y_1} = \mu_{Y_2} = 0$). Then, different test statistics can be considered in order to check if the two linear models are equal or not.

First of all, note that Horváth et al. (2009) developed several methods of comparing two linear models with functional explanatory variables and scalar responses, based on test statistics with chi–squared asymptotic distribution, when $\mathcal{H} = L^2([0,1])$. This methodology can be applied to the case analysed here if **(C.5.3)**, **(C.5.4)** and **(C.5.5)** hold. Furthermore, it has to be assumed that the errors satisfy that $\mathbb{E}(\epsilon_1^4) < \infty$ and $\mathbb{E}(\epsilon_2^4) < \infty$, and the first k eigenvalues of the common covariance operator Γ are nonzero and distinct, that is, $\lambda_1 > \lambda_2 > \ldots > \lambda_k > 0$. In this situation, let $\{(\hat{\lambda}_j, \hat{v}_j)\}_{j=1}^\infty$ be the eigenelements of the empirical covariance operator Γ_n computed with respect to the pooled sample, i.e., $\Gamma_n = (2n)^{-1}\sum_{i=1}^n (X_{1,i} \otimes_{\mathcal{H}} X_{1,i} + X_{2,i} \otimes_{\mathcal{H}} X_{2,i})$. Given the set of empirical eigenfunctions $\{\hat{v}_j\}_{j=1}^\infty$,

which is an orthonormal basis of $L^2([0,1])$, let $\mathbf{X}_1$ and $\mathbf{X}_2$ be the $n \times k$–matrices given by $(\mathbf{X}_1)_{i,j} = \langle X_{1,i}, \hat{v}_j \rangle$ and $(\mathbf{X}_2)_{i,j} = \langle X_{2,i}, \hat{v}_j \rangle$. In addition, let $\mathbf{Y}_1$ and $\mathbf{Y}_2$ be the n–vectors given by $\mathbf{Y}_1 = (Y_{1,1}, \ldots, Y_{1,n})^t$ and $\mathbf{Y}_2 = (Y_{2,1}, \ldots, Y_{2,n})^t$, and let $\hat{\mathbf{\Sigma}}_k$ be the $k \times k$–matrix given by $(\hat{\mathbf{\Sigma}}_k)_{j,j} = \hat{\lambda}_j^{-1}$ for all $j \in \{1, \ldots, k\}$, and $(\hat{\mathbf{\Sigma}}_k)_{j_1,j_2} = 0$ for all $j_1, j_2 \in \{1, \ldots, k\}$ such that $j_1 \neq j_2$. Then, the reasoning in Horváth et al. (2009) led to the following test statistic

$$T_{1,n}^{Eq} = \frac{n}{2\hat{\sigma}^2}(\hat{\mathbf{C}}_1 - \hat{\mathbf{C}}_2)^t \text{diag}(\hat{\lambda}_1, \ldots, \hat{\lambda}_k)(\hat{\mathbf{C}}_1 - \hat{\mathbf{C}}_2), \tag{5.25}$$

where $\hat{\mathbf{C}}_1 = (\mathbf{X}_1^t\mathbf{X}_1)^{-1}\mathbf{X}_1^t\mathbf{Y}_1$ and $\hat{\mathbf{C}}_2 = (\mathbf{X}_2^t\mathbf{X}_2)^{-1}\mathbf{X}_2^t\mathbf{Y}_2$, and $\hat{\sigma}^2$ are residual standard deviations from the estimated regression models computed with respect to the two samples together, i.e., $\hat{\sigma}^2 = (2n-k)^{-1}\sum_{i=1}^n ((Y_{1,i} - \hat{Y}_{1,i})^2 + (Y_{2,i} - \hat{Y}_{2,i})^2)$. Horváth et al. (2009) showed that the distribution of (5.25) can be approximated by the chi–squared distribution with k degrees of freedom. Thus, H_0 is rejected if $T_{1,n}^{Eq} > q_{1-\alpha}$ being q_α the α–quantile of a χ_k^2. Finally, note that test statistics proposed by Horváth et al. (2009) can be applied to check $H_0 : \|\theta_1 - \theta_2\| = 0$ versus $H_1 : \|\theta_1 - \theta_2\| \neq 0$ in more general situations that the specific case which is studied in this chapter such as: when the covariance operators of X_1 and X_2 are not equal, when σ_1 and σ_2 are different, when the two samples have different sizes of roughly the same order (i.e., $n_1 \neq n_2$ such that $n_1/n_2 \to c$ as $n_1, n_2 \to \infty$, with $0 < c < \infty$), or when the response variables Y_1 and Y_2 are functional.

Another choice to check whether $\|\theta_1 - \theta_2\| = 0$ or not is based on the standard FPCA estimator introduced in Section 2.3.2, *"a) Definition of standard FPCA estimator"*, in Chapter 2 (see page 35). Recall that, under **(C.2.1)** (see Chapter 2, page 36), the model parameters θ_1 and θ_2 can be expressed as $\theta_1 = \sum_{j=1}^\infty \lambda_j^{-1}\Delta_1 v_j v_j$ and $\theta_2 = \sum_{j=1}^\infty \lambda_j^{-1}\Delta_2 v_j v_j$, respectively, whenever **(C.5.4)** holds. Therefore, $\|\theta_1 - \theta_2\|^2 = \|\sum_{j=1}^\infty \lambda_j^{-1}(\Delta_1 v_j - \Delta_2 v_j)v_j\|^2 = \sum_{j=1}^\infty \lambda_j^{-2}(\Delta_1 v_j - \Delta_2 v_j)^2$. The model parameters can be estimated by the standard FPCA estimator (see (2.4) in Chapter 2, page 36), so the norm of their difference can be approximated by $\|\hat{\theta}_{1,k_n} - \hat{\theta}_{2,k_n}\|^2 = \|\sum_{j=1}^{k_n} \hat{\lambda}_j^{-1}(\Delta_{1,n}\hat{v}_j - \Delta_{2,n}\hat{v}_j)\hat{v}_j\|^2 = \sum_{j=1}^{k_n} \hat{\lambda}_j^{-2}(\Delta_{1,n}\hat{v}_j - \Delta_{2,n}\hat{v}_j)^2$, where $\Delta_{1,n} = n^{-1}\sum_{i=1}^n X_{1,i} \otimes_{\mathcal{H}'} Y_{1,i}$, $\Delta_{2,n} = n^{-1}\sum_{i=1}^n X_{2,i} \otimes_{\mathcal{H}'} Y_{2,i}$, $\{(\hat{\lambda}_j, \hat{v}_j)\}_{j=1}^\infty$ are the eigenelements of $\Gamma_n = (2n)^{-1}\sum_{i=1}^n (X_{1,i} \otimes_{\mathcal{H}} X_{1,i} + X_{2,i} \otimes_{\mathcal{H}} X_{2,i})$, and $\{k_n\}_{n=1}^\infty$ is a sequence of positive integers such that $k_n \to +\infty$, $k_n \leq n$, and $\hat{\lambda}_{k_n} > 0$. Hence, the next test statistic can be considered

$$T_{2,n}^{Eq} = \sum_{j=1}^{k_n} \left(\frac{\Delta_{1,n}\hat{v}_j - \Delta_{2,n}\hat{v}_j}{\hat{\lambda}_j} \right)^2. \tag{5.26}$$

Since the asymptotic distribution of (5.26) is not known, a bootstrap approach is required to use this statistic in practice.

Finally, the developments presented in Section 5.3.1 lead to propose as test statistic

$$T_{3,n}^{Eq} = \left\| \frac{1}{n}\sum_{i=1}^n ((X_{1,i} - \overline{X}_1)(Y_{1,i} - \overline{Y}_1) - (X_{2,i} - \overline{X}_2)(Y_{2,i} - \overline{Y}_2)) \right\|. \tag{5.27}$$

The calibration of the distribution of (5.25), (5.26) and (5.27) under H_0 can be done by means of bootstrap techniques. For (5.27), the bootstrap statistics defined as

$$T_{3,n}^{Eq,*} = \left\| \frac{1}{n}\sum_{i=1}^n ((X_{1,i} - \overline{X}_1)(Y_{1,i} - \overline{Y}_1) - (X_{2,i} - \overline{X}_2)(Y_{2,i} - \overline{Y}_2))\epsilon_i^* \right\| \tag{5.28}$$

can be considered. Hence, the p–values of $T_{3,n}^{Eq}$ can be approximated using the wild bootstrap approach described in Algorithm 5.3.9 (see page 120), which is consistent based on the theoretical results introduced in the previous section. Therefore, H_0 is rejected when the approximated p–value of (5.27) is smaller than α.

For the statistics $T_{1,n}^{Eq}$ and $T_{2,n}^{Eq}$ another wild bootstrap procedure is proposed, although its consistency and correctness has not been proved in this chapter. The bootstrap statistics for (5.25) are

given by the following expressions

$$T_{1,n}^{Eq,*(a)} = \frac{n}{2\hat{\sigma}^{*2}}(\hat{\mathbf{C}}_1^* - \hat{\mathbf{C}}_2^*)^t \mathrm{diag}(\hat{\lambda}_1, \ldots, \hat{\lambda}_k)(\hat{\mathbf{C}}_1^* - \hat{\mathbf{C}}_2^*), \tag{5.29}$$

$$T_{1,n}^{Eq,*(b)} = \frac{n}{2\hat{\sigma}^{2}}(\hat{\mathbf{C}}_1^* - \hat{\mathbf{C}}_2^*)^t \mathrm{diag}(\hat{\lambda}_1, \ldots, \hat{\lambda}_k)(\hat{\mathbf{C}}_1^* - \hat{\mathbf{C}}_2^*). \tag{5.30}$$

The only difference between (5.29) and (5.30) is the estimation of σ^2 for each iteration: it computed from each bootstrap sample in the first one, whereas it is obtained from the original samples in the second one. Regarding to (5.26), the proposed bootstrap statistic is

$$T_{2,n}^{Eq,*} = \sum_{j=1}^{k_n} \left(\frac{\Delta_{1,n}^* \hat{v}_j - \Delta_{2,n}^* \hat{v}_j}{\hat{\lambda}_j} \right)^2. \tag{5.31}$$

The bootstrap distributions of (5.29), (5.30) and (5.31) are obtained by means of the following wild bootstrap algorithm. For the pilot estimation of the common model parameter $\theta = \theta_1 = \theta_2$, under H_0, the standard FPCA estimator is considered (see (2.4) in Chapter 2, page 36).

Algorithm 5.3.11 (Wild bootstrap)**.**

Step 1. *Compute the value of the statistic $T_{1,n}^{Eq}$ (or the value of the statistic $T_{2,n}^{Eq}$).*

Step 2. *Construct a pilot estimator of the common parameter θ: $\hat{\theta}_{k_n^{pilot}} = \sum_{j=1}^{k_n^{pilot}} \hat{\lambda}_j^{-1} \Delta_n(\hat{v}_j)\hat{v}_j$, where $\{(\hat{\lambda}_j, \hat{v}_j)\}_{j=1}^{\infty}$ are the eigenelements of $\Gamma_n = (2n)^{-1}\sum_{i=1}^n (X_{1,i} \otimes_{\mathcal{H}} X_{1,i} + X_{2,i} \otimes_{\mathcal{H}} X_{2,i})$, and $\Delta_n = (2n)^{-1}\sum_{i=1}^n (X_{1,i} \otimes_{\mathcal{H}'} Y_{1,i} + X_{2,i} \otimes_{\mathcal{H}'} Y_{2,i})$. Obtain the residuals for each sample: $\hat{\epsilon}_{1,i} = Y_{1,i} - \langle \hat{\theta}_{k_n}, X_{1,i} \rangle$ and $\hat{\epsilon}_{2,i} = Y_{2,i} - \langle \hat{\theta}_{k_n}, X_{2,i} \rangle$ for all $i = 1, \ldots, n$.*

Step 3. *Draw $\{\epsilon_{1,i}^*\}_{i=1}^n$ and $\{\epsilon_{2,i}^*\}_{i=1}^n$ two independent sequences of i.i.d. random elements drawn from ϵ^*, which satisfies $\mathbb{E}(\epsilon^*) = 0$, $\mathbb{E}((\epsilon^*)^2) = 1$ and $\int_0^\infty (\mathbb{P}(|\epsilon^*| > t)^{1/2}) < \infty$, and define $Y_{1,i}^* = \langle \hat{\theta}_{k_n}, X_{1,i} \rangle + \hat{\epsilon}_{1,i}\epsilon_{1,i}^*$ and $Y_{2,i}^* = \langle \hat{\theta}_{k_n}, X_{2,i} \rangle + \hat{\epsilon}_{2,i}\epsilon_{2,i}^*$ for all $i = 1, \ldots, n$. Consequently, define $\mathbf{Y}_1^* = (Y_{1,1}^*, \ldots, Y_{1,n}^*)^t$ and $\mathbf{Y}_2^* = (Y_{2,1}^*, \ldots, Y_{2,n}^*)^t$.*

Step 4. *Build $\hat{\mathbf{C}}_1^* = (\mathbf{X}_1^t\mathbf{X}_1)^{-1}\mathbf{X}_1^t\mathbf{Y}_1^*$ and $\hat{\mathbf{C}}_2^* = (\mathbf{X}_2^t\mathbf{X}_2)^{-1}\mathbf{X}_2^t\mathbf{Y}_2^*$ (or $\Delta_{1,n}^* = n^{-1}\sum_{i=1}^n X_{1,i} \otimes_{\mathcal{H}'} Y_{1,i}^*$ and $\Delta_{2,n}^* = n^{-1}\sum_{i=1}^n X_{2,i} \otimes_{\mathcal{H}'} Y_{2,i}^*$), and compute $a_n = |T_{1,n}^{Eq,*}|$ (or compute $b_n = |T_{2,n}^{Eq,*}|$).*

Step 5. *Repeat Steps 3 and 4 a large number of times $n_{boot} \in \mathbb{N}$ in order to obtain a sequence of values $\{a_n^l\}_{l=1}^{n_{boot}}$ (or a sequence of values $\{b_n^l\}_{l=1}^{n_{boot}}$).*

Step 6. *Approximate the p–value of the test by the proportion of values in $\{a_n^l\}_{l=1}^{n_{boot}}$ greater than or equal to $|T_{1,n}^{Eq}|$ (or by the proportion of values in $\{b_n^l\}_{l=1}^{n_{boot}}$ greater than or equal to $|T_{2,n}^{Eq}|$).*

5.4 Simulation study

In this section a simulation study illustrates the performance of the asymptotic approach and the bootstrap calibration for testing the lack of dependence (see Section 5.4.1), and the equality of two linear models (see Section 5.4.2). In both cases, $\mathcal{H} = L^2[0,1]$ was selected, with its usual inner product, i.e., $\langle x, y \rangle = \int_0^1 x(t)y(t)dt$ for all $x, y \in L^2[0,1]$.

5.4.1 Testing the lack of dependence

For this first issue, the considered regression model was (5.1) (see page 106) with $b = 0$, $\mu_X = 0$ and $\mu_Y = 0$, that is,

$$Y = \int_0^1 \theta(t)X(t)dt + \epsilon,$$

where X is a centred random element valued in $L^2[0,1]$, Y and ϵ are centred random variables valued in $\mathbb{R}$, and $\theta \in L^2[0,1]$ is the fixed model parameter. Then, $ns = 500$ samples were simulated, each one consisted of n observations ($n = 50, 100$) from this functional linear model, being X a Brownian motion and $\epsilon \sim \mathcal{N}(0, \sigma^2)$ with signal–to–noise ratio $r = \sigma / \sqrt{\mathbb{E}(\langle \theta, X \rangle^2)}$. Under H_0,

$$\theta_{H_0}(t) = 0, \quad \forall t \in [0,1]$$

was taken as the parameter of the linear model, whereas under H_1, the selected parameter was

$$\theta_{H_1}(t) = \sin(2\pi t^3)^3, \quad \forall t \in [0,1].$$

Furthermore, $\sigma = 1$ was chosen under H_0, while in the alternative H_1 $\sigma = r\sqrt{\mathbb{E}(\langle \theta, X \rangle^2)}$ with three different values for the signal–to–noise ratio ($r = 0.5, 1, 2$) was considered. Remark that both X and θ were discretized to $p = 100$ equidistant design points in $[0,1]$. As done in previous chapters, quadrature weights of p^{-1} were used to approximate integrals involved in the calculations (see more details in Section 3.5, *"How to work with discrete data?"*, in Chapter 3, page 63).

The statistical tests which were introduced in Section 5.2.2 (see page 113) were considered, that is, $T_{1,n}^{Ind}$, $T_{2,n}^{Ind}$, $T_{3,n}^{Ind}$ and $T_{3s,n}^{Ind}$ defined as (5.9), (5.10), (5.11) and (5.12), respectively. For $T_{1,n}^{Ind}$, three distribution approximations were considered: its asymptotic distribution (that is, $\mathcal{N}(0,2)$), and the two calibrations based on the bootstrap statistics $T_{1,n}^{Ind,*(a)}$ and $T_{1,n}^{Ind,*(b)}$, given by (5.15) and (5.16). For $T_{2,n}^{Ind}$, $T_{3,n}^{Ind}$ and $T_{3s,n}^{Ind}$ only the bootstrap approaches based on the statistics $T_{2,n}^{Ind,*}$, $T_{3,n}^{Ind,*}$ and $T_{3s,n}^{Ind,*}$ (see (5.17), (5.13) and (5.14), respectively) were computed. In most of these statistics, an estimation of σ (or σ^*) is involved. It was decided to use the following estimate based on the residual sum of squares, which was proposed by Cardot et al. (2003b),

$$\hat{\sigma}^2 = \frac{1}{n - tr(\mathbf{S}_\rho)} \sum_{i=1}^{n} (Y_i - \mathbf{S}_\rho Y_i)^2,$$

where $\mathbf{S}_\rho$ is the hat matrix for the penalized B–splines estimator, i.e., $\mathbf{S}_\rho Y_i = \langle \hat{\theta}_{PS}, X_i \rangle$ (recall (2.2) in Chapter 2, page 35), with the next choices: B–splines with degree 4 and 20 equispaced knots, second derivatives for the penalty, and parameter ρ selected by GCV (see Section 3.5, *"Parameter estimation"*, in Chapter 3, page 63). Bedides, $\hat{\sigma}^{*2}$ is computed analogously using the bootstrap sample generated in each iteration.

In order to calibrate the bootstrap distributions, the wild bootstrap algorithm introduced in Section 5.2.2 (see Algorithm 5.2.10, page 115) was used for $T_{1,n}^{Ind,*(a)}$, $T_{1,n}^{Ind,*(b)}$ and $T_{2,n}^{Ind,*}$, whereas the wild bootstrap algorithm introduced in Section 5.2.1 (see Algorithm 5.2.8, page 112) was computed for $T_{3,n}^{Ind,*}$ and $T_{3s,n}^{Ind,*}$. In all these cases, $n_{boot} = 1,000$ bootstrap iterations were computed to approximate the corresponding p–values. Furthermore, $\{\epsilon_i^*\}_{i=1}^n$ were drawn from the following sum of two Dirac distributions: $0.1(5+\sqrt{5})\delta_{(1-\sqrt{5})/2} + 0.1(5-\sqrt{5})\delta_{(1+\sqrt{5})/2}$ (i.e., $\mathbb{P}(\epsilon_i^* = (1-\sqrt{5})/2) = 0.1(5+\sqrt{5})$ and $\mathbb{P}(\epsilon_i^* = (1+\sqrt{5})/2) = 0.1(5-\sqrt{5})$ for all $i = 1 \ldots, n$).

Since the significance level α and the parameter k_n involved in $T_{1,n}^{Ind}$ and $T_{2,n}^{Ind}$ (and in their bootstrap versions) must be fixed to run the procedure, the study was done for four significance levels ($\alpha = 0.2, 0.1, 0.05, 0.01$) and for different numbers of principal components ($k_n = 1, \ldots, 20$). Nevertheless, in order to simplify the presentation, the information collected in the following tables corresponds to only three of the values of k_n which were analyed ($k_n = 5, 10, 20$), whereas the results depicted in the figures correspond to only two of the values of α ($\alpha = 0.1, 0.05$).

Table 5.1 (see page 124) displays the empirical size (i.e., the percentage of rejections under H_0) of the test statistics obtained in the simulation study and Figure 5.1 (see page 124) shows the effect of the parameter k_n for the testing procedures based on $T_{1,n}^{Ind}$ and $T_{2,n}^{Ind}$. For $T_{1,n}^{Ind}$, it can be highlighted that bootstrap approaches present empirical sizes closer to the nominal level α than the asymptotic approximation for $T_{1,n}^{Ind}$, mainly when k_n is small and $\alpha = 0.2$. If one compares the performance of the two bootstrap procedures proposed, it seems that if σ^2 is bootstrapped ($T_{1,n}^{Ind,*(a)}$) the results are better than if the same estimation of the variance is considered in all the bootstrap replications

($T_{1,n}^{Ind,*(b)}$) above all when k_n is large. As far as $T_{2,n}^{Ind}$ is concerned, the estimated levels are near to the nominal ones, being the largest k_n the cases in which they are farther from the theoretical α. Finally, it must be remarked that the F–test and its studentized versions also get good results in terms of test levels.

			$T_{1,n}^{Ind}$									$T_{2,n}^{Ind}$			$T_{3,n}^{Ind}$	$T_{3s,n}^{Ind}$
			$\mathcal{N}(0,2)$			$T_{1,n}^{Ind,*(a)}$			$T_{1,n}^{Ind,*(b)}$			$T_{2,n}^{Ind,*}$			$T_{3,n}^{Ind,*}$	$T_{3s,n}^{Ind,*}$
n	α	k_n	5	10	20	5	10	20	5	10	20	5	10	20		
50	20%		19.4	17.6	16.0	21.4	21.6	20.0	21.6	19.0	15.2	19.8	20.8	18.4	21.6	20.8
	10%		10.8	10.4	8.2	9.0	10.8	10.6	8.0	7.2	3.2	8.6	7.2	7.2	11.8	11.2
	5%		8.2	7.0	4.4	5.0	4.0	4.6	5.0	2.4	0.0	4.0	3.2	3.0	6.0	6.2
	1%		4.8	4.2	2.2	1.2	0.4	0.0	0.6	0.0	0.0	0.2	0.6	0.4	0.6	1.2
100	20%		15.0	19.4	20.0	20.8	21.0	19.0	21.0	20.8	18.0	21.4	19.4	17.6	21.6	21.2
	10%		8.6	9.6	9.0	11.8	10.8	10.4	10.4	9.6	6.2	9.8	8.8	7.0	11.6	11.8
	5%		5.6	5.2	4.0	4.4	4.6	3.6	3.6	3.4	2.2	4.6	5.2	2.8	5.6	5.6
	1%		2.6	2.4	1.2	1.4	1.2	0.8	1.2	0.6	0.2	1.0	0.6	0.8	0.4	0.4

Table 5.1: Testing the lack of dependence. Estimated levels for $T_{1,n}^{Ind}$ (using the asymptotic distribution $\mathcal{N}(0,2)$ and the bootstrap distributions of $T_{1,n}^{Ind,*(a)}$ and $T_{1,n}^{Ind,*(b)}$), $T_{2,n}^{Ind}$ (using the bootstrap distribution of $T_{2,n}^{Ind,*}$), $T_{3,n}^{Ind}$ (using the bootstrap distribution of $T_{3,n}^{Ind,*}$), and $T_{3s,n}^{Ind}$ (using the bootstrap distribution of $T_{3s,n}^{Ind,*}$).

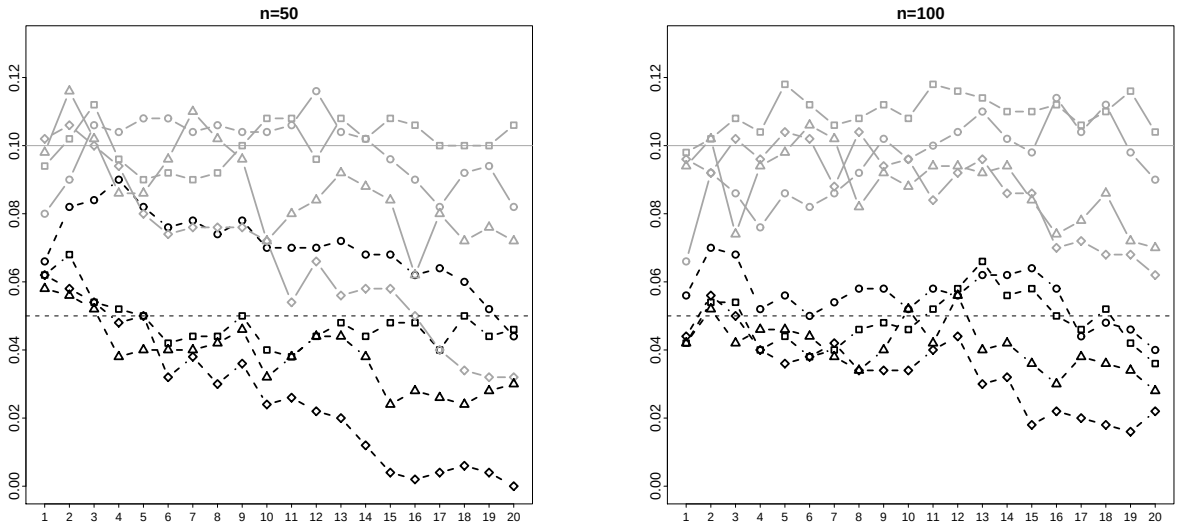

Figure 5.1: Testing the lack of dependence. Estimated levels when $n = 50$ (left panel) and $n = 100$ (right panel) for $T_{1,n}^{Ind}$ (circle for the asymptotic distribution $\mathcal{N}(0,2)$; square for the bootstrap distributions of $T_{1,n}^{Ind,*(a)}$; diamond for the bootstrap distributions of $T_{1,n}^{Ind,*(b)}$), and $T_{2,n}^{Ind}$ (triangle for the bootstrap distribution of $T_{2,n}^{Ind,*}$). Significance level: $\alpha = 0.1$ (solid grey lines) and $\alpha = 0.05$ (dashed black lines).

On the other hand, Table 5.2 (see page 125) and Figure 5.2 (see page 126) show the empirical power obtained with the different procedures for each considered signal–to–noise ratio r. In terms of power, when $r = 0.5$ the results for all the methods are similar, except for $T_{2,n}^{Ind}$ for which the empirical power decreases drastically, above all when k_n increases (this effect is also observed for $r = 1$ and $r = 2$). This fact seems to be due to the construction of $T_{2,n}^{Ind}$ since this test statistic is the only one which does not involve the estimation of σ^2. In addition, the power of $T_{1,n}^{Ind}$ also falls abruptly when $T_{1,n}^{Ind,*(b)}$ is considered, n and α are small and k_n is very large.

A similar situation can be observed when $r = 1$ and $r = 2$. For $r = 2$, it can be seen that the empirical power is smaller for all the methods in general, being obtained an important loss of power when the sample is small ($n = 50$), and k_n increases and/or α decreases. Furthermore, in this case, the empirical power relies heavily on the selected k_n value (see Figure 5.2, page 126). Hence, the advantage of using $T_{3,n}^{Ind}$ or $T_{3s,n}^{Ind}$ is that they do not require the selection of k_n or any other parameter, and they are competitive in terms of power. Nevertheless, it also seems that an adequate k_n selection can make $T_{1,n}^{Ind}$ obtain better empirical power than $T_{3,n}^{Ind}$ or $T_{3s,n}^{Ind}$ in some cases.

				$T_{1,n}^{Ind}$									$T_{2,n}^{Ind}$			$T_{3,n}^{Ind}$	$T_{3s,n}^{Ind}$
				$\mathcal{N}(0,2)$			$T_{1,n}^{Ind,*(a)}$			$T_{1,n}^{Ind,*(b)}$			$T_{2,n}^{Ind,*}$			$T_{3,n}^{Ind,*}$	$T_{3s,n}^{Ind,*}$
r	n	α	k_n	5	10	20	5	10	20	5	10	20	5	10	20		
0.5	50	20%		100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	88.8	0.0	0.0	100.0	100.0
		10%		100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	60.8	0.0	0.0	100.0	100.0
		5%		100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	99.0	32.2	0.0	0.0	100.0	100.0
		1%		100.0	100.0	100.0	100.0	100.0	100.0	100.0	99.4	51.4	3.4	0.0	0.0	99.4	100.0
	100	20%		100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	1.0	0.0	100.0	100.0
		10%		100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	0.0	0.0	100.0	100.0
		5%		100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	98.4	0.0	0.0	100.0	100.0
		1%		100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	70.0	0.0	0.0	100.0	100.0
1	50	20%		100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	98.2	66.6	3.6	0.2	100.0	100.0
		10%		100.0	100.0	100.0	100.0	100.0	99.8	100.0	99.8	89.6	33.6	0.8	0.0	100.0	100.0
		5%		100.0	100.0	99.8	100.0	100.0	99.6	100.0	99.0	59.6	16.6	0.2	0.0	99.2	99.2
		1%		100.0	100.0	99.6	99.6	97.6	94.6	95.2	67.6	2.6	2.2	0.0	0.0	87.8	92.4
	100	20%		100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	97.0	7.8	0.0	100.0	100.0
		10%		100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	86.4	2.2	0.0	100.0	100.0
		5%		100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	67.8	1.0	0.0	100.0	100.0
		1%		100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	99.8	21.6	0.2	0.0	100.0	100.0
2	50	20%		85.4	75.6	66.8	89.0	81.2	77.2	89.0	76.8	51.4	34.0	11.8	7.2	90.4	89.8
		10%		80.0	68.6	56.4	79.4	68.6	59.4	76.4	57.4	20.2	16.6	4.0	2.4	79.0	79.0
		5%		74.4	62.2	48.4	67.4	51.6	43.6	60.8	37.8	6.2	10.4	1.0	0.4	67.8	67.2
		1%		67.4	51.4	35.6	40.0	26.4	20.2	25.4	6.0	0.0	0.8	0.0	0.0	34.4	39.0
	100	20%		99.8	98.8	94.6	100.0	99.8	98.0	100.0	99.2	94.2	60.0	14.6	7.6	99.8	99.8
		10%		99.6	96.6	91.2	99.6	97.2	93.6	99.6	96.0	82.4	34.2	6.2	2.0	97.8	97.4
		5%		99.6	95.6	85.8	97.8	94.0	85.8	97.2	90.4	64.6	18.0	2.8	0.4	94.4	94.4
		1%		97.6	91.4	75.4	88.2	76.4	64.0	85.2	63.4	26.2	2.2	0.8	0.0	79.2	82.4

Table 5.2: Testing the lack of dependence. Empirical power for $T_{1,n}^{Ind}$ (using the asymptotic distribution $\mathcal{N}(0,2)$ and the bootstrap distributions of $T_{1,n}^{Ind,*(a)}$ and $T_{1,n}^{Ind,*(b)}$), $T_{2,n}^{Ind}$ (using the bootstrap distribution of $T_{2,n}^{Ind,*}$), $T_{3,n}^{Ind}$ (using the bootstrap distribution of $T_{3,n}^{Ind,*}$), and $T_{3s,n}^{Ind}$ (using the bootstrap distribution of $T_{3s,n}^{Ind,*}$).

5.4.2 Testing the equality of linear models

To show the behaviour of the proposed procedures for checking the equality of the two linear models in (5.18) (see page 115), the simplified case in which $b_1 = b_2 = 0$, $\mu_{X_1} = \mu_{X_2} = 0$ and $\mu_{Y_1} = \mu_{Y_2} = 0$ is going to be considered. Hence, (5.18) turns into

$$\begin{cases} Y_1 = \int_0^1 \theta_1(t)X_1(t)dt + \epsilon_1, \\ Y_2 = \int_0^1 \theta_2(t)X_2(t)dt + \epsilon_2, \end{cases}$$

where X_1 and X_2 are centred random elements valued in $L^2[0,1]$, Y_1, Y_2, ϵ_1 and ϵ_2 are centred random variables valued in $\mathbb{R}$, and $\theta_1, \theta_2 \in L^2[0,1]$ are the fixed model parameters. Hence, $ns = 500$ samples of n observations ($n = 50, 100$) from each of these two functional linear model were simulated, where X_1 and X_2 were Brownian motions and $\epsilon_1, \epsilon_2 \sim \mathcal{N}(0, \sigma^2)$ with signal–to–noise ratio $r = \sigma/\sqrt{\mathbb{E}(\langle\theta, X\rangle^2)} = 0.2$. Under H_0, the two model parameters coincide

$$\theta_{1,H_0}(t) = \theta_{2,H_0}(t) = 2\sin(0.5\pi t) + 4\sin(1.5\pi t) + 5\sin(2.5\pi t), \quad \forall t \in [0,1]$$

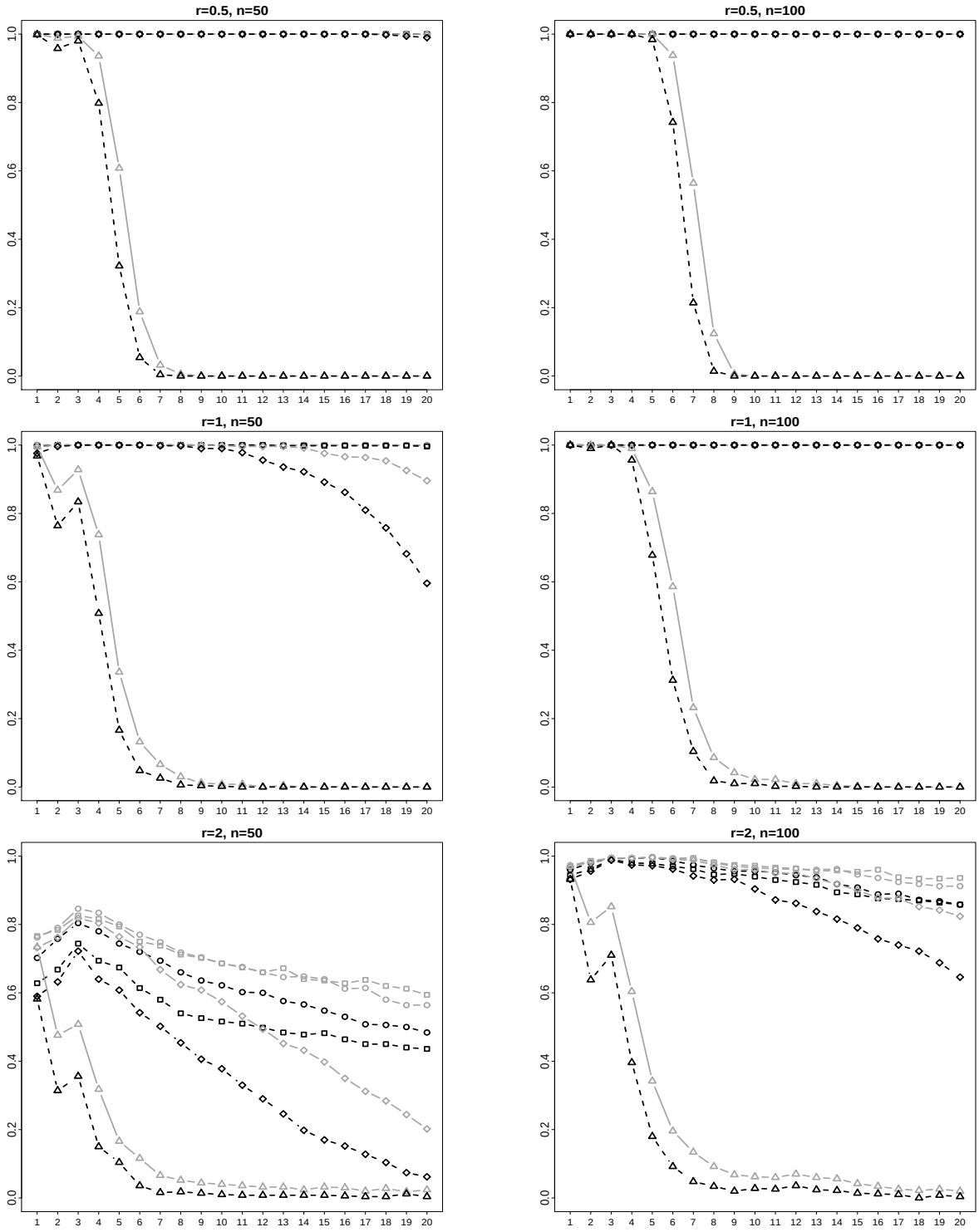

Figure 5.2: Testing the lack of dependence. Empirical power when $n = 50$ (left panels) and $n = 100$ (right panels) for $T_{1,n}^{Ind}$ (circle for the asymptotic distribution $\mathcal{N}(0, 2)$; square for the bootstrap distributions of $T_{1,n}^{Ind,*(a)}$; diamond for the bootstrap distributions of $T_{1,n}^{Ind,*(b)}$), and $T_{2,n}^{Ind}$ (triangle for the bootstrap distribution of $T_{2,n}^{Ind,*}$). Significance level: $\alpha = 0.1$ (solid grey lines) and $\alpha = 0.05$ (dashed black lines). Rows correspond to the three different values for the signal–to–noise ratio ($r = 0.5, 1, 2$).

whereas under H_1 the parameters were defined as

$$\begin{cases} \theta_{1,H_1}(t) = 2\sin(0.5\pi t) + 4\sin(1.5\pi t) + 5\sin(2.5\pi t), & \forall t \in [0,1], \\ \theta_{2,H_1}(t) = 2(2\sin(0.5\pi t) + 4\sin(1.5\pi t) + 5\sin(2.5\pi t)), & \forall t \in [0,1]. \end{cases}$$

As was commented in Section 5.4.1, the functional covariates and the model parameters were discretized to $p = 100$ equidistant design points in $[0,1]$, using quadrature weights of p^{-1} in order to approximate integrals which appear in the different computations (further details in Section 3.5, *"How to work with discrete data?"*, in Chapter 3, page 63).

For this simulation study, the statistical tests presented in Section 5.3.2 (see page 120) were computed: $T_{1,n}^{Eq}$, $T_{2,n}^{Eq}$ and $T_{3,n}^{Eq}$ (see (5.25), (5.26) and (5.27), respectively). For $T_{1,n}^{Eq}$, three different approximations of its distribution were considered: the asymptotic approach (that is, χ_k^2), and the two calibrations based on the bootstrap statistics $T_{1,n}^{Eq,*(a)}$ and $T_{1,n}^{Eq,*(b)}$, given by (5.29) and (5.30). For $T_{2,n}^{Eq}$ and $T_{3,n}^{Eq}$, only the approaches based on the bootstrap statistics $T_{2,n}^{Eq,*}$ and $T_{3,n}^{Eq,*}$ were computed (see (5.31) and (5.28)). In $T_{1,n}^{Eq}$ and its bootstrap versions, an estimation of σ (or σ^*) is required. The following estimate based on the residual sum of squares for the standard FPCA estimator was chosen

$$\hat{\sigma}^2 = \frac{1}{n - tr(\mathbf{S}_{k_n})} \sum_{i=1}^{n} (Y_i - \mathbf{S}_{k_n} Y_i)^2,$$

where $\mathbf{S}_{k_n}$ is the hat matrix for the standard FPCA estimator, i.e., $\mathbf{S}_{k_n} Y_i = \langle \hat{\theta}_{k_n}, X_i \rangle$ with $tr(\mathbf{S}_{k_n}) = k_n$ (recall (2.4) in Chapter 2, page 36), being k_n selected by GCV (see Section 3.5, *"Parameter estimation"*, in Chapter 3, page 63).

The wild bootstrap algorithm introduced in Section 5.3.2 (see Algorithm 5.3.11, page 122) was considered for calibrating the bootstrap distribution of $T_{1,n}^{Eq,*(a)}$, $T_{1,n}^{Eq,*(b)}$ and $T_{2,n}^{Eq,*}$. For the pilot FPCA estimator the value of k_n^{pilot} was obtained by GCV using the initial samples. On the other hand, the wild bootstrap algorithm introduced in Section 5.3.1 (see Algorithm 5.3.9, page 120) was computed in order to calibrate the bootstrap distribution of $T_{3,n}^{Eq,*}$. As in the previous simulation study, $n_{boot} = 1,000$ bootstrap iterations were done to approximate the p–values of the bootstrap test statistics, and the real random variable ϵ^* was built as the following sum of two Dirac distributions: $0.1(5+\sqrt{5})\delta_{(1-\sqrt{5})/2} + 0.1(5-\sqrt{5})\delta_{(1+\sqrt{5})/2}$.

Finally, note that the simulation study was done for four significance levels ($\alpha = 0.2, 0.1, 0.05, 0.01$) and for different numbers of principal components ($k = 1, \ldots, 10$ and $k_n = 1, \ldots, 10$) involved in $T_{1,n}^{Eq}$, $T_{2,n}^{Eq}$ and their associated bootstrap statistics (althought the simulation study was computed for $k, k_n \in \{1, \ldots, 20\}$, the results for $k, k_n \in \{11, \ldots, 20\}$ were omitted given that the estimated levels and the empirical power for these values were similar to those obtained for $k, k_n = 10$). In order to simplify the presentation, the results for only two of the values of α ($\alpha = 0.1, 0.05$) are shown in the following figures, whereas three of the values of k_n ($k/k_n = 1, 5, 10$) were selected for the next tables.

The comparison of the estimated levels for test statistics can be found in Table 5.3 (see page 128). In addition, Figure 5.3 (see page 128) illustrates the effect of the k/k_n parameter for the testing procedures based on $T_{1,n}^{Eq}$ and $T_{2,n}^{Eq}$. Note that bootstrap calibrations for $T_{1,n}^{Eq}$ have empirical sizes closer to α than the χ_k^2 approximation when k is large and n is small. If one compares the behaviour of the two bootstrap approaches, it can be seen that the estimated levels for $T_{1,n}^{Eq,*(a)}$ (σ^2 bootstrapped for each iteration) are nearer to the nominal than the estimated levels for $T_{1,n}^{Eq,*(b)}$ (same initial estimate $\hat{\sigma}^2$ for all the replications) when k is large and $n = 50$, above all for $\alpha = 0.2$ and $\alpha = 0.1$. As far as $T_{2,n}^{Eq}$ is concerned, the estimated levels are also quite near to the nominal ones, except for the largest k_n values. Regarding to $T_{3,n}^{Eq}$, its behaviour is also adequate in most of cases, except when n and α are both small.

The empirical power of the different proposed methods is shown in Table 5.4 (see page 129). Furthermore, Figure 5.4 (see page 129) allows visualizing the influence of the parameter k/k_n on the empirical power of $T_{1,n}^{Eq}$ and $T_{2,n}^{Eq}$. The highest powers correspond to procedures based on $T_{1,n}^{Eq}$, for

			$T_{1,n}^{Eq}$									$T_{2,n}^{Eq}$			$T_{3,n}^{Eq}$
			χ_k^2			$T_{1,n}^{Eq,*(a)}$			$T_{1,n}^{Eq,*(b)}$			$T_{2,n}^{Eq,*}$			$T_{3,n}^{Eq,*}$
n	α	k/k_n	1	5	10	1	5	10	1	5	10	1	5	10	
50	20%		16.6	22.2	28.4	19.2	19.2	19.6	19.8	21.4	23.0	18.8	19.4	16.4	19.0
	10%		8.0	10.4	17.0	9.4	9.2	10.8	9.4	10.0	13.4	10.4	8.2	8.6	7.4
	5%		4.2	5.6	9.6	4.6	4.4	5.6	3.8	5.4	5.8	4.0	5.0	3.8	1.6
	1%		0.6	1.0	2.4	0.6	0.8	1.4	0.8	0.8	1.0	0.6	1.0	0.6	0.4
100	20%		18.6	21.4	23.2	19.2	20.0	20.4	19.6	21.6	21.6	19.0	17.6	16.8	17.6
	10%		7.2	9.4	11.8	10.6	9.6	9.2	10.4	10.2	9.4	9.8	8.8	8.0	8.4
	5%		4.6	4.8	6.0	4.4	5.0	5.0	4.6	5.2	5.0	4.2	5.2	4.4	4.8
	1%		0.8	1.4	0.8	1.0	1.2	0.8	1.2	1.2	0.8	0.6	0.4	1.2	1.2

Table 5.3: Testing the equality of linear models. Estimated levels for $T_{1,n}^{Eq}$ (using the asymptotic distribution χ_k^2 and the bootstrap distributions of $T_{1,n}^{Eq,*(a)}$ and $T_{1,n}^{Eq,*(b)}$), $T_{2,n}^{Eq}$ (using the bootstrap distribution of $T_{2,n}^{Eq,*}$), and $T_{3,n}^{Eq}$ (using the bootstrap distribution of $T_{3,n}^{Eq,*}$).

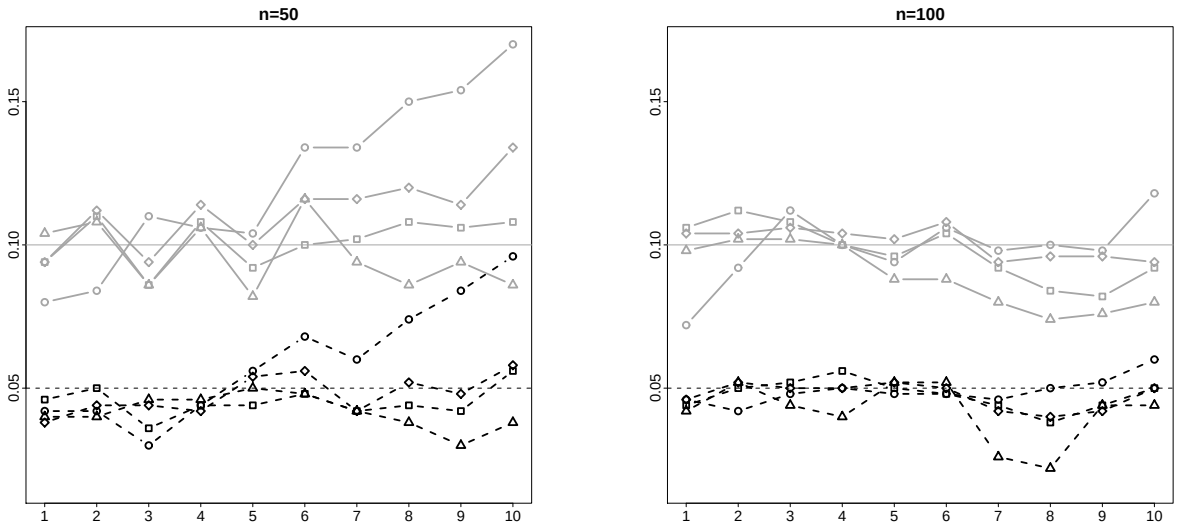

Figure 5.3: Testing the equality of linear models. Estimated levels when $n = 50$ (left panel) and $n = 100$ (right panel) for $T_{1,n}^{Eq}$ (circle for the asymptotic distribution χ_k^2; square for the bootstrap distributions of $T_{1,n}^{Eq,*(a)}$; diamond for the bootstrap distributions of $T_{1,n}^{Eq,*(b)}$), and $T_{2,n}^{Eq}$ (triangle for the bootstrap distribution of $T_{2,n}^{Eq,*}$). Significance level: $\alpha = 0.1$ (solid grey lines) and $\alpha = 0.05$ (dashed black lines).

which the three approximations of the test distribution give similar results: the maximum power is reached for all k, apart from $k = 1$. On the other hand, the empirical power of $T_{2,n}^{Eq}$ is clearly smaller than that obtained by means of $T_{1,n}^{Eq}$ and, in addition, it decreases drastically when k_n increases. The results for $T_{3,n}^{Eq}$ are not satisfactory either, since the empirical power is small, above all when $n = 50$, and $\alpha = 0.05$ or $\alpha = 0.01$. Since neither $T_{2,n}^{Eq}$ nor $T_{3,n}^{Eq}$ involve the estimation of σ^2, a possible future research line could be to *studentize* these statistics, and analyse if their studentized versions improve the present results in terms of empirical power.

			$T_{1,n}^{Eq}$									$T_{2,n}^{Eq}$			$T_{3,n}^{Eq}$
			χ_k^2			$T_{1,n}^{Eq,*(a)}$			$T_{1,n}^{Eq,*(b)}$			$T_{2,n}^{Eq,*}$			$T_{3,n}^{Eq,*}$
n	α	k/k_n	1	5	10	1	5	10	1	5	10	1	5	10	
50	20%		99.4	100.0	100.0	94.6	100.0	100.0	94.8	100.0	100.0	80.4	46.2	3.4	76.0
	10%		96.8	100.0	100.0	93.0	100.0	100.0	93.0	100.0	100.0	77.6	33.6	0.8	62.0
	5%		95.0	100.0	100.0	90.6	100.0	100.0	91.4	100.0	100.0	76.8	23.4	0.0	48.0
	1%		85.6	100.0	100.0	88.2	100.0	100.0	89.4	100.0	100.0	74.0	8.4	0.0	22.0
100	20%		100.0	100.0	100.0	99.6	100.0	100.0	99.6	100.0	100.0	91.2	74.4	5.8	97.4
	10%		100.0	100.0	100.0	99.4	100.0	100.0	99.2	100.0	100.0	90.4	63.4	0.4	91.2
	5%		100.0	100.0	100.0	99.2	100.0	100.0	99.2	100.0	100.0	89.4	52.0	0.0	83.8
	1%		99.8	100.0	100.0	98.0	100.0	100.0	98.2	100.0	100.0	86.6	35.2	0.0	58.0

Table 5.4: Testing the equality of linear models. Empirical power for $T_{1,n}^{Eq}$ (using the asymptotic distribution χ_k^2 and the bootstrap distributions of $T_{1,n}^{Eq,*(a)}$ and $T_{1,n}^{Eq,*(b)}$), $T_{2,n}^{Eq}$ (using the bootstrap distribution of $T_{2,n}^{Eq,*}$), and $T_{3,n}^{Eq}$ (using the bootstrap distribution of $T_{3,n}^{Eq,*}$).

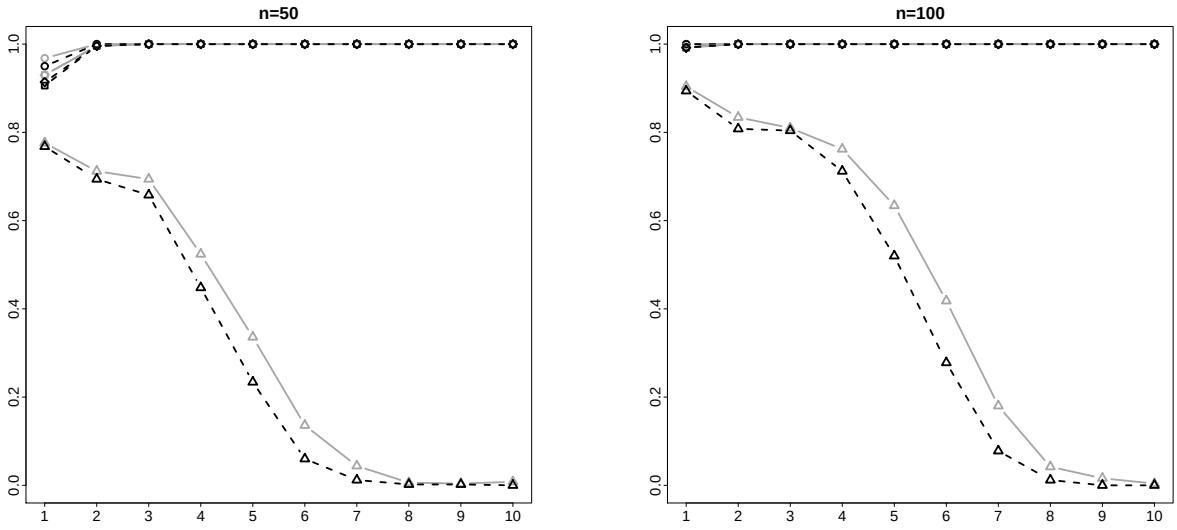

Figure 5.4: Testing the equality of linear models. Empirical power when $n = 50$ (left panel) and $n = 100$ (right panel) for $T_{1,n}^{Eq}$ (circle for the asymptotic distribution χ_k^2; square for the bootstrap distributions of $T_{1,n}^{Eq,*(a)}$; diamond for the bootstrap distributions of $T_{1,n}^{Eq,*(b)}$), and $T_{2,n}^{Eq}$ (triangle for the bootstrap distribution of $T_{2,n}^{Eq,*}$). Significance level: $\alpha = 0.1$ (solid grey lines) and $\alpha = 0.05$ (dashed black lines).

5.5 Real data application

For the data application, the environmental example introduced in the first chapter has been considered: the atmospheric pollution data (see Section 1.1.2 in Chapter 1, page 2). Recall that this dataset contains concentrations of hourly averaged NO_x (measured in $\mu g/m^3$) in the neighbourhood of a power station belonging to ENDESA, located in As Pontes in the Northwest of Spain (from Jan 2007 to Dec 2009). Next, the tests for lack of dependence and for equality of linear models previously studied are applied to this real data.

5.5.1 Testing the lack of dependence

During unfavourable meteorological conditions, NO_x levels can quickly rise and cause an air pollution episode. Due to this fact, an important issue is to forecast NO_x levels with half an hour horizon in

order to allow the power plant staff to avoid NO_x concentrations reaching the limit values fixed by the current environmental legislation. Therefore, it is necessary to estimate properly the regression model which defines the relationship between the observed NO_x concentration in the last minutes (X) and the NO_x concentration with half an hour horizon (Y). For this purpose, a first step could be to determine if there exists a linear dependence between X and Y by means of the testing procedures presented in Section 5.2.

A sample of size $n = 300$ was built, where each curve X corresponds to 240 consecutive minute–by–minute values of hourly averaged NO_x concentration, and the response Y corresponds to the hourly averaged NO_x value half an hour ahead. Taking $\alpha = 0.05$, the tests for lack of dependence reject the null hypothesis in all cases (thus, there is a linear relationship between the variables), except $T_{2,n}^{Ind}$ when k_n is large (see Table 5.5 and Figure 5.5). Nevertheless, as it was commented in the simulation study, this test statistic does not take into account the variance term and its power is clearly lower than the power of the other tests. As a result, it seems that the relationship between the last 240 observations of the NO_x concentration and the NO_x value half an hour ahead could be modelled by means of a functional linear regression with scalar response. Indeed, the estimation of this linear model was done in Section 3.6 in Chapter 3 (see Table 3.11, page 73).

	$T_{1,n}^{Ind}$									$T_{2,n}^{Ind}$			$T_{3,n}^{Ind}$	$T_{3s,n}^{Ind}$
	$\mathcal{N}(0,2)$			$T_{1,n}^{Ind,*(a)}$			$T_{1,n}^{Ind,*(b)}$			$T_{2,n}^{Ind,*}$			$T_{3,n}^{Ind,*}$	$T_{3s,n}^{Ind,*}$
k_n	1	5	10	1	5	10	1	5	10	1	5	10		
	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.003	0.006	0.000	0.000

Table 5.5: Atmospheric pollution data. Testing the lack of dependence. P–values for $T_{1,n}^{Ind}$ (using the asymptotic distribution $\mathcal{N}(0,2)$ and the bootstrap distributions of $T_{1,n}^{Ind,*(a)}$ and $T_{1,n}^{Ind,*(b)}$), $T_{2,n}^{Ind}$ (using the bootstrap distribution of $T_{2,n}^{Ind,*}$), $T_{3,n}^{Ind}$ (using the bootstrap distribution of $T_{3,n}^{Ind,*}$), and $T_{3s,n}^{Ind}$ (using the bootstrap distribution of $T_{3s,n}^{Ind,*}$).

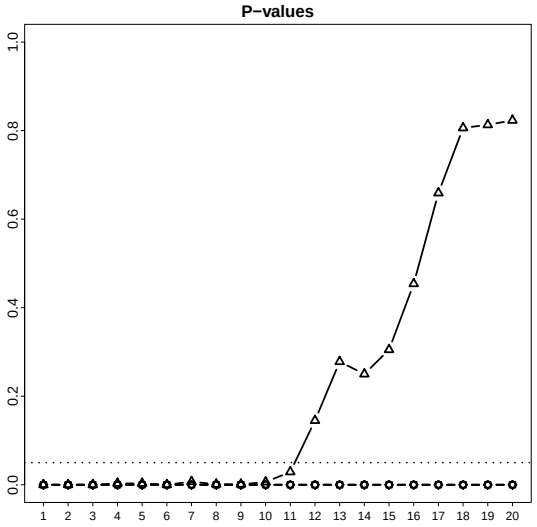

Figure 5.5: Atmospheric pollution data. Testing the lack of dependence. P–values for $T_{1,n}^{Ind}$ (circle for the asymptotic distribution $\mathcal{N}(0,2)$; square for the bootstrap distributions of $T_{1,n}^{Ind,*(a)}$; diamond for the bootstrap distributions of $T_{1,n}^{Ind,*(b)}$), and $T_{2,n}^{Ind}$ (triangle for the bootstrap distribution of $T_{2,n}^{Ind,*}$). Horizontal dotted line corresponds to the significance level $\alpha = 0.5$.

5.5.2 Testing the equality of linear models

The observed NO_x levels are low (and practically constant) in general. However, under certain meteorological conditions, these concentrations can quickly rise and, thus, cause an air pollution episode. Hence, one could think that the linear regression model which links the curves of hourly averaged NO_x concentrations (X) and the hourly averaged NO_x levels half an hour ahead (Y) is different when the covariates in the sample corresponds to low–valued curves, medium–valued curves or high–valued curves.

In order to check this assumption, three samples of size $n = 100$ were considered, where the curves X consist of 240 consecutive minute–by–minute values of hourly averaged NO_x concentration, and the responses Y are the associated hourly averaged NO_x values half an hour ahead. The first sample, denoted by bin 1, corresponds to curves X such that the last measured NO_x level belongs to the interval $[0\,\mu g/m^3, 10\,\mu g/m^3)$, i.e., $X(240) \in [0\,\mu g/m^3, 10\,\mu g/m^3)$. The second sample, or bin 2, consists of pairs (X, Y) satisfying that $X(240) \in [10\,\mu g/m^3, 20\,\mu g/m^3)$. Finally, the third sample is called bin 3 and their covariates verifying that $X(240)$ is equal or higher to $20\,\mu g/m^3$.

Table 5.6 and Figure 5.6 (see page 132) show the obtained results when the equality of models among the different bins is tested using the techniques introduced in Section 5.3. Note that $T_{1,n}^{Eq}$ with the asymptotic approach rejects the equality of models in all cases irrespectively of k (except if bin 2 and bin 3 are compared and $k = 1$), whereas $T_{2,n}^{Eq}$ never rejects H_0. On the other hand the behaviour of $T_{1,n}^{Eq}$ for bootstrap calibrations is similar for both $T_{1,n}^{Eq,*(a)}$ and $T_{1,n}^{Eq,*(b)}$: H_0 cannot be rejected when bin 1 and bin 2 are compared; H_0 is rejected when bin 1 and bin 3 are compared for moderate k values (and not rejected for very small or large k); and H_0 cannot be rejected when bin 2 and bin 3 are compared, except if the parameter k is very large. Finally, $T_{3,n}^{Eq}$, which does not depend on any parameter, does not reject the null hypothesis only for the first studied case (bin 1 vs. bin 2).

	$T_{1,n}^{Eq}$									$T_{2,n}^{Eq}$			$T_{3,n}^{Eq}$
	χ_k^2			$T_{1,n}^{Eq,*(a)}$			$T_{1,n}^{Eq,*(b)}$			$T_{2,n}^{Eq,*}$			$T_{3,n}^{Eq,*}$
k/k_n	1	5	10	1	5	10	1	5	10	1	5	10	
bin 1 vs. bin 2	0.000	0.000	0.002	0.706	0.060	0.946	0.698	0.056	0.941	0.472	0.306	0.437	0.701
bin 1 vs. bin 3	0.000	0.000	0.000	0.165	0.012	0.013	0.160	0.008	0.002	0.626	0.336	0.823	0.038
bin 2 vs. bin 3	0.890	0.023	0.000	0.776	0.902	0.331	0.789	0.932	0.315	0.546	0.895	0.720	0.035

Table 5.6: Atmospheric pollution data. Testing the equality of linear models. P–values for $T_{1,n}^{Eq}$ (using the asymptotic distribution χ_k^2 and the bootstrap distributions of $T_{1,n}^{Eq,*(a)}$ and $T_{1,n}^{Eq,*(b)}$), $T_{2,n}^{Eq}$ (using the bootstrap distribution of $T_{2,n}^{Eq,*}$), and $T_{3,n}^{Eq}$ (using the bootstrap distribution of $T_{3,n}^{Eq,*}$) when bin 1 is compared with bin 2, bin 1 is compared with bin 3, and bin 2 is compared with bin 3.

Given that this is a multiple testing problem, corrected p–values were also computed. Two adjustement methods were considered: Bonferroni's method and Benjamini and Hochberg' method (see Benjamini and Hochberg, 1995). Both techniques were applied using the *R* routine *p.adjust* available in the package *stats* (see R Development Core Team, 2010). The obtained results are shown in Table 5.7 and Figure 5.7 (see page 132) for the Bonferroni's method, and Table 5.8 and Figure 5.8 (see page 133) for Benjamini and Hochberg' method. Note that similar conclusions can be derived using the corrected p–values to those obtained using the original ones.

5.6 Final conclusions

The proposed bootstrap methods are competitive alternatives to tests based on asymptotic distributions, and they often give test sizes closer to the nominal ones. In terms of power, the statistic tests which include a consistent estimation of the error variance σ^2 obtain higher empirical power than the test statistics which do not take it into account. Furthermore, in all the statistics involving a k/k_n

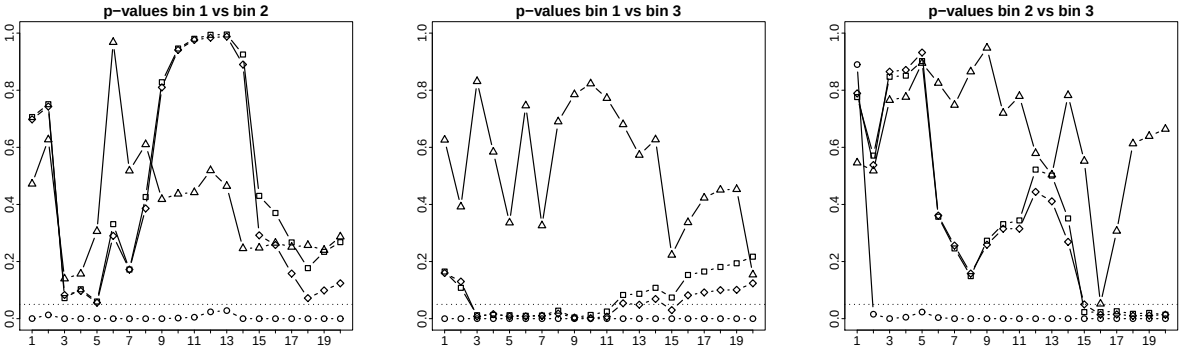

Figure 5.6: Atmospheric pollution data. Testing the equality of linear models. P–values for $T_{1,n}^{Eq}$ (circle for the asymptotic distribution χ_k^2; square for the bootstrap distributions of $T_{1,n}^{Eq,*(a)}$; diamond for the bootstrap distributions of $T_{1,n}^{Eq,*(b)}$), and $T_{2,n}^{Eq}$ (triangle for the bootstrap distribution of $T_{2,n}^{Eq,*}$) when bin 1 is compared with bin 2 (left panel), bin 1 is compared with bin 3 (centre panel), and bin 2 is compared with bin 3 (right panel). Horizontal dotted line corresponds to the significance level $\alpha = 0.5$.

	$T_{1,n}^{Eq}$									$T_{2,n}^{Eq}$			$T_{3,n}^{Eq}$
	χ_k^2			$T_{1,n}^{Eq,*(a)}$			$T_{1,n}^{Eq,*(b)}$			$T_{2,n}^{Eq,*}$			$T_{3,n}^{Eq,*}$
k/k_n	1	5	10	1	5	10	1	5	10	1	5	10	
bin 1 vs. bin 2	0.001	0.000	0.006	1.000	0.180	1.000	1.000	0.168	1.000	1.000	0.918	1.000	1.000
bin 1 vs. bin 3	0.000	0.000	0.000	0.495	0.036	0.039	0.480	0.024	0.006	1.000	1.000	1.000	0.114
bin 2 vs. bin 3	1.000	0.070	0.000	1.000	1.000	0.993	1.000	1.000	0.945	1.000	1.000	1.000	0.105

Table 5.7: Atmospheric pollution data. Testing the equality of linear models. P–values adjusted using Bonferroni correction for $T_{1,n}^{Eq}$ (using the asymptotic distribution χ_k^2 and the bootstrap distributions of $T_{1,n}^{Eq,*(a)}$ and $T_{1,n}^{Eq,*(b)}$), $T_{2,n}^{Eq}$ (using the bootstrap distribution of $T_{2,n}^{Eq,*}$), and $T_{3,n}^{Eq}$ (using the bootstrap distribution of $T_{3,n}^{Eq,*}$) when bin 1 is compared with bin 2, bin 1 is compared with bin 3, and bin 2 is compared with bin 3.

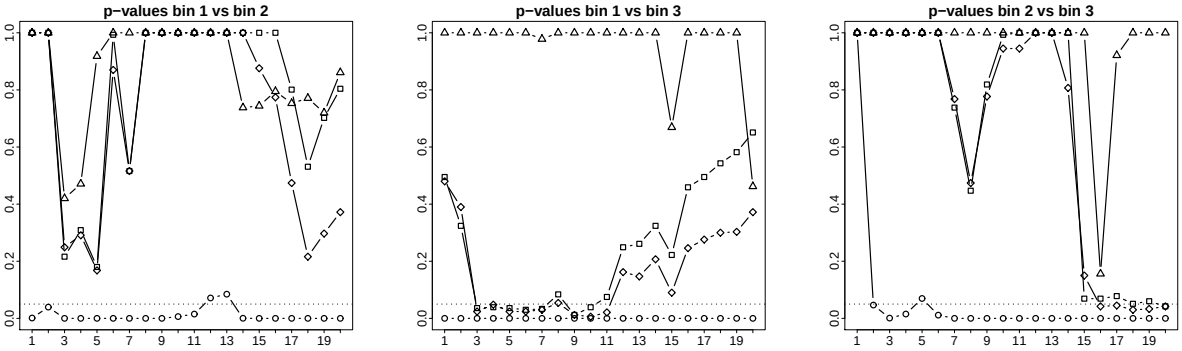

Figure 5.7: Atmospheric pollution data. Testing the equality of linear models. P–values adjusted using Bonferroni correction for $T_{1,n}^{Eq}$ (circle for the asymptotic distribution χ_k^2; square for the bootstrap distributions of $T_{1,n}^{Eq,*(a)}$; diamond for the bootstrap distributions of $T_{1,n}^{Eq,*(b)}$), and $T_{2,n}^{Eq}$ (triangle for the bootstrap distribution of $T_{2,n}^{Eq,*}$) when bin 1 is compared with bin 2 (left panel), bin 1 is compared with bin 3 (centre panel), and bin 2 is compared with bin 3 (right panel). Horizontal dotted line corresponds to the significance level $\alpha = 0.5$.

	$T_{1,n}^{Eq}$									$T_{2,n}^{Eq}$			$T_{3,n}^{Eq}$
	χ_k^2			$T_{1,n}^{Eq,*(a)}$			$T_{1,n}^{Eq,*(b)}$			$T_{2,n}^{Eq,*}$			$T_{3,n}^{Eq,*}$
k/k_n	1	5	10	1	5	10	1	5	10	1	5	10	
bin 1 vs. bin 2	0.001	0.000	0.002	0.776	0.090	0.946	0.789	0.084	0.941	0.626	0.504	0.823	0.701
bin 1 vs. bin 3	0.000	0.000	0.000	0.495	0.036	0.039	0.480	0.024	0.006	0.626	0.504	0.823	0.057
bin 2 vs. bin 3	0.890	0.023	0.000	0.776	0.902	0.497	0.789	0.932	0.473	0.626	0.895	0.823	0.057

Table 5.8: Atmospheric pollution data. Testing the equality of linear models. P–values adjusted using Benjamini and Hochberg correction for $T_{1,n}^{Eq}$ (using the asymptotic distribution χ_k^2 and the bootstrap distributions of $T_{1,n}^{Eq,*(a)}$ and $T_{1,n}^{Eq,*(b)}$), $T_{2,n}^{Eq}$ (using the bootstrap distribution of $T_{2,n}^{Eq,*}$), and $T_{3,n}^{Eq}$ (using the bootstrap distribution of $T_{3,n}^{Eq,*}$) when bin 1 is compared with bin 2, bin 1 is compared with bin 3, and bin 2 is compared with bin 3.

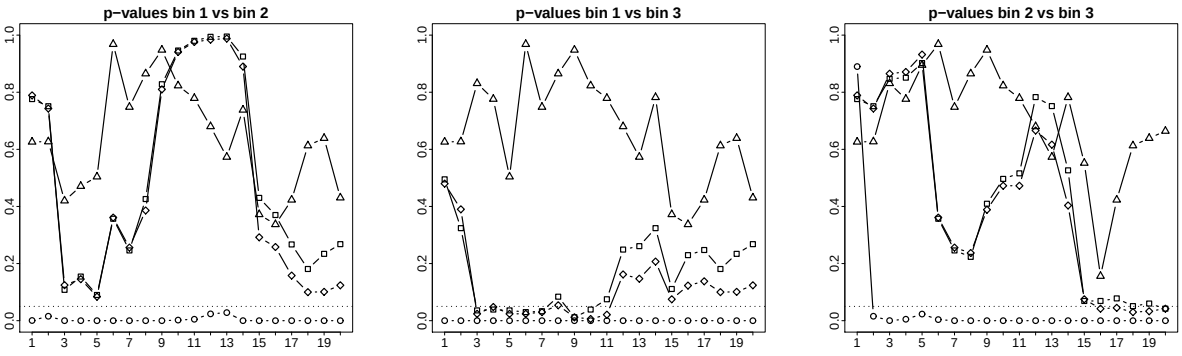

Figure 5.8: Atmospheric pollution data. Testing the equality of linear models. P–values adjusted using Benjamini–Hochberg correction for $T_{1,n}^{Eq}$ (circle for the asymptotic distribution χ_k^2; square for the bootstrap distributions of $T_{1,n}^{Eq,*(a)}$; diamond for the bootstrap distributions of $T_{1,n}^{Eq,*(b)}$), and $T_{2,n}^{Eq}$ (triangle for the bootstrap distribution of $T_{2,n}^{Eq,*}$) when bin 1 is compared with bin 2 (left panel), bin 1 is compared with bin 3 (centre panel), and bin 2 is compared with bin 3 (right panel). Horizontal dotted line corresponds to the significance level $\alpha = 0.5$.

parameter, a suitable choice of k/k_n seems to be a quite important point and it is currently an open question.

Besides of the optimal k/k_n selection, other issues related to these hypotheses tests require further research, such as their extension to functional linear models with functional response. In addition, for the test of lack of dependence, it would be interesting to combine it with the functional ANOVA test (see Cuevas et al. (2004), and González-Rodríguez et al. (2012)) in order to develop an ANCOVA test in this context. On the other hand, for the test of equality, the extension of the results to other test statistics proposed by Horváth et al. (2009), which are applicable to more general situations (different covariance structures, different sample sizes,...), is still a pending issue.

5.7 Appendix Chapter 5

This section contains the proofs of the main results of the chapter and some necessary technical lemmas.

5.7.1 Proof of Theorem 5.2.3

Since T_n^{Ind} can be equivalently expressed as

$$\begin{aligned} T_n^{Ind} &= \frac{1}{n}\sum_{i=1}^{n}(X_i-\mu_X)(Y_i-\mu_Y) - (\overline{X}-\mu_X)(\overline{Y}-\mu_Y) \\ &= \frac{n-1}{n^2}\sum_{i=1}^{n}(X_i-\mu_X)(Y_i-\mu_Y) - \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j\neq i}(X_i-\mu_X)(Y_j-\mu_Y), \end{aligned}$$

it is straightforward to check item 1. Furthermore, the $a.s.-\mathbb{P}$ convergence is a direct application of the SLLN for separable Hilbert–valued random elements.

On the other hand, given that $\mathbb{E}(\|(X-\mu_X)(Y-\mu_Y)\|^2)<\infty$, the convergence in law can be deduced by applying the CLT for separable Hilbert–valued random elements (see, for instance, Laha and Rohatgi (1979)) together with Slutsky's Theorem. The concrete expression of the operator Γ_Z can be obtained by simple computations as follows

$$\begin{aligned} \Gamma_Z &= \Gamma_{(X-\mu_X)(Y-\mu_Y)} = \Gamma_{(X-\mu_X)(\epsilon+\langle\theta,X-\mu_X\rangle)} = \Gamma_{(X-\mu_X)\epsilon} + \Gamma_{(X-\mu_X)\langle\theta,X-\mu_X\rangle} \\ &= \mathbb{E}(\epsilon^2(X-\mu_X)\otimes_{\mathcal{H}}(X-\mu_X)) + \mathbb{E}(\langle\theta,X-\mu_X\rangle^2(X-\mu_X)\otimes_{\mathcal{H}}(X-\mu_X)) \\ &= \sigma^2\Gamma + \mathbb{E}(\langle\theta,X-\mu_X\rangle^2(X-\mu_X)\otimes_{\mathcal{H}}(X-\mu_X)). \end{aligned}$$

5.7.2 Proof of Corollary 5.2.4

Corollary 5.2.4 can be derived from item 3 in Theorem 5.2.3 because, under the null hypothesis, $\|T^{Ind}\| = \|\Delta\|_{\mathcal{H}'} = 0$, and $Y-\mu_Y=\epsilon$ (hence, $\Gamma_Z = \Gamma_{(X-\mu_X)(Y-\mu_Y)} = \Gamma_{(X-\mu_X)\epsilon} = \sigma^2\Gamma$).

5.7.3 Proof of Theorem 5.2.5

Theorem 5.2.5 is easily proven taking into account that $n^{-1}\sum_{i=1}^{n}(X_i-\overline{X})(Y_i^n-\overline{Y}^n)$ converges $a.s.-\mathbb{P}$ to $\mathbb{E}((X-\mu_X)(Y^n-\mu_{Y^n}))$ when $n\to\infty$, by the SLLN for separable Hilbert–valued random elements, and $\|\mathbb{E}((X-\mu_X)(Y^n-\mu_{Y^n}))\| = \delta_n\|\Gamma\theta\|/\sqrt{n}$.

5.7.4 Proof of Theorem 5.2.7

First of all, note that

$$\begin{aligned} \sqrt{n}\,T_n^{Ind,N*} &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(X_i^*-\mu_X)(Y_i^*-\mu_Y) - \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(X_i-\mu_X)(Y_i-\mu_Y) \\ &\quad - \sqrt{n}(\overline{X}^*-\mu_X)(\overline{Y}^*-\mu_Y) + \sqrt{n}(\overline{X}-\mu_X)(\overline{Y}-\mu_Y) \\ &= \sqrt{n}S_n^* + \frac{1}{\sqrt{n}}\sqrt{n}(\overline{X}^*-\overline{X})\sqrt{n}(\overline{Y}^*-\overline{Y}) - \sqrt{n}(\overline{X}^*-\overline{X})(\overline{Y}^*-\mu_Y) - (\overline{X}^*-\mu_X)\sqrt{n}(\overline{Y}^*-\overline{Y}), \end{aligned} \tag{5.32}$$

where S_n^* is defined as $S_n^* = n^{-1}\sum_{i=1}^{n}(X_i^*-\mu_X)(Y_i^*-\mu_Y) - n^{-1}\sum_{i=1}^{n}(X_i-\mu_X)(Y_i-\mu_Y)$. Given that $\{(X_i^*-\mu_X)(Y_i^*-\mu_Y)\}_{i=1}^n$ are i.i.d. $\mathcal{H}$–valued random elements chosen at random from the *bootstrap population* $\{(X_i-\mu_X)(Y_i-\mu_Y)\}_{i=1}^n$ and $\mathbb{E}(\|(X-\mu_X)(Y-\mu_Y)\|^2)<\infty$, item 1 in Lemma 5.7.1 guarantees that the bootstrap statistic $\sqrt{n}S_n^*$ converges in law to $Z_{(X-\mu_X)(Y-\mu_Y)}$ $a.s.-\mathbb{P}$.

On the other hand, given that $\mathbb{E}(\|X\|^2)<\infty$ and $\mathbb{E}(Y^2)<\infty$, items 1 and 2 in Lemma 5.7.1, together with Slutsky's Theorem, ensure that the last three terms in (5.32) converge in probability to 0 $a.s.-\mathbb{P}$, and consequently the convergence in law stated in the theorem is proven.

Finally, note that $\hat{\sigma}^{*2} = \overline{(Y^*)^2} - (\overline{Y}^*)^2$. Then, the convergence of $\hat{\sigma}^{*2}$ holds in virtue of items 2 and 3 in Lemma 5.7.1, since $\mathbb{E}(Y^2)<\infty$.

5.7.5 Formulation and proof of Lemma 5.7.1

As was indicated previously, the asymptotic behaviour of the naive bootstrap statistic is going to be analysed through some results on bootstrapping general empirical measures presented by Giné and Zinn (1990) (specifically, through their Theorem 2.4). For this reason, it is necessary to introduce some notation used by Giné and Zinn (1990), and check certain conditions required in their paper, in order to obtain the Lemma 5.7.1 below.

It should be noted that the bootstrap results in Giné and Zinn (1990) refer to empirical processes indexed by a class of functions $\mathcal{F}$, and based on a probability measure $\mathbb{P}$, that particularly extend to the bootstrap about the mean in separable Banach (and thus Hilbert) spaces. In order to establish this connection, it is enough to choose

$$\mathcal{F} = \{f \in \mathcal{H}' | \, \|f\|_{\mathcal{H}'} \leq 1\}$$

(see Giné (1997) and Kosorok (2008), for a general overview of indexed empirical processes). Note that $\mathcal{F}$ is *image admissible Suslin*[2] considering the weak topology. This fact ensures that $\mathcal{F}$ satisfies the measurability conditions which are necessary to apply Theorem 2.4 in Giné and Zinn (1990). In addition, let F be defined as $F(x) = \sup_{f\in\mathcal{F}} |f(x)| = \|x\|$ for all $x \in \mathcal{H}$, which satisfies

$$\mathbb{E}(F^2(X)) = \mathbb{E}(\|X\|^2) < \infty.$$

Finally, consider the bounded and linear (so continuous) operator δ from $\mathcal{H}$ to $l^\infty(\mathcal{F})$[3], given by $\delta(x)(f) = \delta_x(f) = f(x)$ for all $x \in \mathcal{H}$ and $f \in \mathcal{F}$, and denote by $\mathrm{Im}(\delta) \subset l^\infty(\mathcal{F})$ its range. Since $\|\delta(x)\|_\infty = \|x\|$ for all $x \in \mathcal{H}$, then there exists $\delta^{-1} : \mathrm{Im}(\delta) \to \mathcal{H}$, such that δ^{-1} is continuous (see, for instance, Lemma 6.16 in Kosorok (2008)). In addition, as $\mathrm{Im}(\delta)$ is closed, Dugundji's Theorem allows to consider a continuous extension $\delta^{-1} : l^\infty(\mathcal{F}) \to \mathcal{H}$ (see, for instance, Theorem 10.9 in Kosorok (2008)). Thus, following the typical empirical process notation, the empirical process $n^{-1/2}\sum_{i=1}^n (\delta_{X_i} - \mathbb{P})$ indexed in $\mathcal{F}$ is directly connected with $n^{-1/2}\sum_{i=1}^n (X_i - \mathbb{E}(X))$ by means of the continuous mapping δ^{-1} and vice versa. As a result, the CLT for separable Hilbert–valued random elements (see, for instance, Laha and Rohatgi (1979)) together with the Continuous Mapping Theorem applied to δ guarantee that $\mathcal{F} \in CLT(\mathbb{P})$[4], which is another condition involved in Theorem 2.4 by Giné and Zinn (1990).

All these considerations lead to the results collected in the following lemma.

Lemma 5.7.1 (González-Manteiga et al., 2012)**.** *Let ξ be a measurable mapping from a probabilistic space denoted by $(\Omega, \mathcal{A}, \mathbb{P})$ to a separable Hilbert space $(\mathcal{H}, \langle\cdot,\cdot\rangle)$ with corresponding norm $\|\cdot\|$ such that $\mathbb{E}(\|\xi\|^2) < \infty$. Let $\{\xi_i\}_{i=1}^n$ be a sequence of i.i.d. random elements with the same distribution as ξ, and let $\{\xi_i^*\}_{i=1}^n$ be i.i.d. from $\{\xi_i\}_{i=1}^n$. Then*

(i) $\sqrt{n}(\overline{\xi}^* - \overline{\xi})$ converges in law to Z_ξ a.s. $-\,\mathbb{P}$

(ii) $\overline{\xi}^*$ converges in probability to $\mathbb{E}(\xi)$ a.s. $-\,\mathbb{P}$

(iii) $\overline{\|\xi^*\|^2}$ converges in probability to $\mathbb{E}(\|\xi\|^2)$ a.s. $-\,\mathbb{P}$

---

[2] A separable measurable space $(Y, \mathcal{S})$ will be called a *Suslin space* iff there is a Polish space X and a Borel measurable map from X onto Y (a *Polish space* is a topological space metrizable as a complete separable metric space).
If $(\Omega, \mathcal{A})$ is a measurable space and $\mathcal{F}$ a set, then a real–valued function $X : (f, \omega) \to X(f, \omega)$ will be called *image admissible Suslin via* $(Y, \mathcal{S}, T)$ iff $(Y, \mathcal{S})$ is a Suslin measurable space, T is a function from Y onto $\mathcal{F}$, and $(y, \omega) \to X(T(y), \omega)$ is jointly measurable on $Y \times \Omega$. Let $\mathcal{F}$ be a set of functions on Ω and $X(f, \omega) \equiv f(\omega)$. $\mathcal{F}$ will be called *image admissible Suslin* if X is image admissible Suslin via some $(Y, \mathcal{S}, T)$ as above.

[3] $l^\infty(\mathcal{F})$ is the space of the bounded functions $\mathcal{F} \to \mathbb{R}$ with the uniform norm topology.

[4] A *Radon measure* γ is a measure defined on the σ–algebra of Borel sets of a topological space X which is locally finite (i.e., for any $x \in X$ there is a neighbourhood which has finite measure) and inner regular (i.e., $\gamma(B) = \sup\{\gamma(K) : K \subset B, K \text{ compact}\}$).
A sequence $\{Y_n\}_{n=1}^\infty$ of random elements of $l^\infty(\mathcal{F})$ *converges weakly* in $l^\infty(\mathcal{F})$ if there exists a Radon probability measure γ on $l^\infty(\mathcal{F})$ such that for all $H : l^\infty(\mathcal{F}) \to \mathbb{R}$ bounded and continuous, $\lim_{n\to\infty} \mathbb{E}^* H(Y_n) = \int H d\gamma$.
$\mathcal{F} \in CLT(\mathbb{P})$ if the sequence $\{\sqrt{n}(\mathbb{P}_n - \mathbb{P})(f) : f \in \mathcal{F}\}$ converges weakly in $l^\infty(\mathcal{F})$ to a Radon centred Gaussian probability measure $\gamma_{\mathbb{P}}$ on $l^\infty(\mathcal{F})$, being $\mathbb{P}_n$ the empirical measure given by $\mathbb{P}_n = n^{-1}\sum_{i=1}^n \delta_{X_i}$.

Proof. Taking into account the reasoning commented before the lemma formulation, Theorem 2.4 of Giné and Zinn (1990) ensures that $\sqrt{n}(\mathbb{P}_n^*(\omega) - \mathbb{P}_n(\omega))$ converges in law to a centred Gaussian process on $\mathcal{F}$, $G = \delta(Z_\xi)$, $a.s.-\mathbb{P}$, where $\mathbb{P}_n(\omega) = n^{-1}\sum_{i=1}^n \delta_{\xi_i(\omega)}$, and $\mathbb{P}_n^*(\omega)$ is the empirical measure based on $\{\xi_i^*\}_{i=1}^n$, i.e., $\mathbb{P}_n^*(\omega) = n^{-1}\sum_{i=1}^n \delta_{\xi_i^*(\omega)}$. Consequently, by applying the Continuous Mapping Theorem to δ^{-1}, $\sqrt{n}(\overline{\xi}^* - \overline{\xi})$ converges in law to $Z_\xi = \delta^{-1}(G)$, and item 1 is shown. Note that item 1 is also a direct consequence of Remark 2.5 of Giné and Zinn (1990); nevertheless it was proven based on Theorem 2.4 to illustrate the technique.

Items 2 and 3 can be checked in a similar way by applying Theorem 2.6 of Giné and Zinn (1990). ■

5.7.6 Proof of Theorem 5.2.9

Remark that, for all $\omega \in \Omega$,

$$\begin{aligned}
\sqrt{n}\, T_n^{Ind,W*}(\omega) &= \frac{1}{\sqrt{n}}\sum_{i=1}^n (X_i(\omega)-\mu_X)(Y_i(\omega)-\mu_Y)\epsilon_i^* - \left(\frac{1}{\sqrt{n}}\sum_{i=1}^n (X_i(\omega)-\mu_X)\epsilon_i^*\right)(\overline{Y}(\omega)-\mu_Y) \\
&\quad - (\overline{X}(\omega)-\mu_X)\frac{1}{\sqrt{n}}\sum_{i=1}^n (Y_i(\omega)-\mu_Y)\epsilon_i^* + (\overline{X}(\omega)-\mu_X)(\overline{Y}(\omega)-\mu_Y)\frac{1}{\sqrt{n}}\sum_{i=1}^n \epsilon_i^* \\
&= S_n^* - \left(\frac{1}{\sqrt{n}}\sum_{i=1}^n (X_i(\omega)-\mu_X)\epsilon_i^*\right)(\overline{Y}(\omega)-\mu_Y) \\
&\quad - (\overline{X}(\omega)-\mu_X)\left(\frac{1}{\sqrt{n}}\sum_{i=1}^n (Y_i(\omega)-\mu_Y)\epsilon_i^*\right) + (\overline{X}(\omega)-\mu_X)(\overline{Y}(\omega)-\mu_Y)\left(\frac{1}{\sqrt{n}}\sum_{i=1}^n \epsilon_i^*\right),
\end{aligned} \tag{5.33}$$

where $S_n^* = n^{-1/2}\sum_{i=1}^n (X_i(\omega)-\mu_X)(Y_i(\omega)-\mu_Y)\epsilon_i^*$. According to Lemma 5.7.2, for almost all $\omega \in \Omega$, S_n^* converges in law to $Z_{(X-\mu_X)(Y-\mu_Y)}$, since $\mathbb{E}(\|(X-\mu_X)(Y-\mu_Y)\|^2) < \infty$.

On the other hand, $(\overline{X}(\omega)-\mu_X)$ and $(\overline{Y}(\omega)-\mu_Y)$ converge to 0 by SLLN. This fact jointly with Lemma 5.7.2 and Slutsky's Theorem guarantee the convergence in probability to 0 of the last three summands in (5.33), given that $\mathbb{E}(\|(X-\mu_X)\|^2) < \infty$ and $\mathbb{E}((Y-\mu_Y)^2) < \infty$. Thus, the result is reached by applying again Slutsky's Theorem.

5.7.7 Formulation and proof of Lemma 5.7.2

Lemma 5.7.2 (González-Manteiga et al., 2012)**.** *Let ξ be a measurable mapping from a probabilistic space denoted by $(\Omega, \mathcal{A}, \mathbb{P})$ to a separable Hilbert space $(\mathcal{H}, \langle\cdot,\cdot\rangle)$ with corresponding norm $\|\cdot\|$ such that $\mathbb{E}(\|\xi\|^2) < \infty$. Let $\{\xi_i\}_{i=1}^n$ be a sequence of i.i.d. random elements with the same distribution as ξ, and let $\{W_i\}_{i=1}^n$ be a sequence of i.i.d. real random variables independent of $\{\xi_i\}_{i=1}^n$, with $\mathbb{E}(W_i) = 0$ and $\int_0^\infty (\mathbb{P}(|W_1| > t)^{1/2}) < \infty$. Then the following statements are equivalent*

(i) $\mathbb{E}(\|\xi\|^2) < \infty$ (and, consequently, $\sqrt{n}(\overline{\xi} - \mathbb{E}(\xi))$ converges in law to Z_ξ).

(ii) For almost all $\omega \in \Omega$, $n^{-1/2}\sum_{i=1}^n W_i\xi_i(\omega)$ converges in law to Z_ξ.

Proof. This lemma is a particularization of a result due to Ledoux, Talagrand and Zinn (see Giné and Zinn (1990) and Ledoux and Talagrand (1988)). See also the Multiplier CLT in Kosorok (2008) for the empirical process indexed by a class of measurable functions counterpart. ■

5.7.8 Proof of Theorem 5.3.4

Remark that T_n^{Eq} can be expressed as

$$\begin{aligned}
T_n^{Eq} &= \frac{1}{n}\sum_{i=1}^{n}((X_{1,i}-\mu_{X_1})(Y_{1,i}-\mu_{Y_1})-(X_{2,i}-\mu_{X_2})(Y_{2,i}-\mu_{Y_2})) \\
&\quad -((\overline{X}_1-\mu_{X_1})(\overline{Y}_1-\mu_{Y_1})-(\overline{X}_2-\mu_{X_2})(\overline{Y}_2-\mu_{Y_2})) \\
&= \frac{n-1}{n^2}\sum_{i=1}^{n}((X_{1,i}-\mu_{X_1})(Y_{1,i}-\mu_{Y_1})-(X_{2,i}-\mu_{X_2})(Y_{2,i}-\mu_{Y_2})) \\
&\quad -\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j\neq i}((X_{1,i}-\mu_{X_1})(Y_{1,j}-\mu_{Y_1})-(X_{2,i}-\mu_{X_2})(Y_{2,j}-\mu_{Y_2})),
\end{aligned}$$

so it is straightforward to check item 1. Moreover, by the SLLN for separable Hilbert–valued random elements and the Slutsky's Theorem, the $a.s.-\mathbb{P}$ convergence in item 2 can be stated.

On the other hand, one has that

$$\begin{aligned}
\sqrt{n}(T_n^{Eq}-T^{Eq}) &= \sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n}((X_{1,i}-\mu_{X_1})(Y_{1,i}-\mu_{Y_1})-(X_{2,i}-\mu_{X_2})(Y_{2,i}-\mu_{Y_2}))-T^{Eq}\right) \\
&\quad -\frac{1}{\sqrt{n}}\sqrt{n}(\overline{X}_1-\mu_{X_1})\sqrt{n}(\overline{Y}_1-\mu_{Y_1})+\frac{1}{\sqrt{n}}\sqrt{n}(\overline{X}_2-\mu_{X_2})\sqrt{n}(\overline{Y}_2-\mu_{Y_2}).
\end{aligned}$$

The last two terms in the right side of the previous expression converge $a.s.-\mathbb{P}$ to 0 as $n\to\infty$ by the CLT for separable Hilbert–valued random elements, since $\mathbb{E}(\|X_1\|^2)<\infty$, $\mathbb{E}(\|X_2\|^2)<\infty$, $\mathbb{E}(Y_1^2)<\infty$ and $\mathbb{E}(Y_2^2)<\infty$, and the Slutsky's Theorem. Therefore, due to Slutsky's Theorem, item 3 will be proven if it is showed the convergence in law of

$$\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n}((X_{1,i}-\mu_{X_1})(Y_{1,i}-\mu_{Y_1})-(X_{2,i}-\mu_{X_2})(Y_{2,i}-\mu_{Y_2}))-T^{Eq}\right).$$

Given that $\mathbb{E}(\|(X_1-\mu_{X_1})(Y_1-\mu_{Y_1})-(X_2-\mu_{X_2})(Y_2-\mu_{Y_2})\|^2)<\infty$, this convergence can be deduced by applying again the CLT for separable Hilbert–valued random elements. Finally, taking into account that **(C.5.4)** is satisfied, the concrete expression of the operator Γ_Z can be obtained as follows

$$\begin{aligned}
\Gamma_Z &= \Gamma_{(X_1-\mu_{X_1})(Y_1-\mu_{Y_1})-(X_2-\mu_{X_2})(Y_2-\mu_{Y_2})} = \Gamma_{(X_1-\mu_{X_1})(\epsilon_1+\langle\theta_1,X_1-\mu_{X_1}\rangle)-(X_2-\mu_{X_2})(\epsilon_2+\langle\theta_2,X_2-\mu_{X_2}\rangle)} \\
&= \Gamma_{(X_1-\mu_{X_1})\epsilon_1}+\Gamma_{(X_1-\mu_{X_1})\langle\theta_1,X_1-\mu_{X_1}\rangle}+\Gamma_{(X_2-\mu_{X_2})\epsilon_2}+\Gamma_{(X_2-\mu_{X_2})\langle\theta_2,X_2-\mu_{X_2}\rangle} \\
&= \mathbb{E}(\epsilon_1^2(X_1-\mu_{X_1})\otimes_{\mathcal{H}}(X_1-\mu_{X_1}))+\mathbb{E}(\langle\theta_1,X_1-\mu_{X_1}\rangle^2(X_1-\mu_{X_1})\otimes_{\mathcal{H}}(X_1-\mu_{X_1})) \\
&\quad +\mathbb{E}(\epsilon_2^2(X_2-\mu_{X_2})\otimes_{\mathcal{H}}(X_2-\mu_{X_2}))+\mathbb{E}(\langle\theta_2,X_2-\mu_{X_2}\rangle^2(X_2-\mu_{X_2})\otimes_{\mathcal{H}}(X_2-\mu_{X_2})) \\
&= 2\sigma^2\Gamma+\mathbb{E}(\langle\theta_1,X_1-\mu_{X_1}\rangle^2(X_1-\mu_{X_1})\otimes_{\mathcal{H}}(X_1-\mu_{X_1})) \\
&\quad +\mathbb{E}(\langle\theta_2,X_2-\mu_{X_2}\rangle^2(X_2-\mu_{X_2})\otimes_{\mathcal{H}}(X_2-\mu_{X_2})).
\end{aligned}$$

5.7.9 Proof of Corollary 5.3.5

Corollary 5.3.5 can be derived from item 3 in Theorem 5.3.4 taking into account that, under the null hypothesis, $\|T^{Eq}\|=\|\Delta_1-\Delta_2\|_{\mathcal{H}'}=0$.

5.7.10 Proof of Theorem 5.3.6

Remark that

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n}((X_{1,i}-\overline{X}_1)(Y_{1,i}-\overline{Y}_1)-(X_{2,i}-\overline{X}_2)(Y_{2,i}^n-\overline{Y}_2^n))\\
&=\frac{1}{n}\sum_{i=1}^{n}((X_{1,i}-\mu_{X_1})(Y_{1,i}-\mu_{Y_1})-(X_{2,i}-\mu_{X_2})(Y_{2,i}^n-\mu_{Y_2^n}))-((\overline{X}_1-\mu_{X_1})(\overline{Y}_1-\mu_{Y_1})\\
&\quad-(\overline{X}_2-\mu_{X_2})(\overline{Y}_2^n-\mu_{Y_2^n})).
\end{aligned}$$

One gets that the previous statistic converges $a.s.-\mathbb{P}$ to $\mathbb{E}((X_1-\mu_{X_1})(Y_1-\mu_{Y_1})-(X_2-\mu_{X_2})(Y_2^n-\mu_{Y_2^n}))$, as $n\to\infty$, by the SLLN for separable Hilbert–valued random elements and the Slutsky's Theorem. Hence, Theorem 5.3.6 is easily proven given that, when **(C.5.4)** is satisfied, $\|\mathbb{E}((X_1-\mu_{X_1})(Y_1-\mu_{Y_1})-(X_2-\mu_{X_2})(Y_2^n-\mu_{Y_2^n}))\|=\delta_n\|\Gamma\theta_1\|/\sqrt{n}$.

5.7.11 Proof of Theorem 5.3.8

First of all, note that

$$\begin{aligned}
\sqrt{n}\,T_n^{Eq,N*}&=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}((X_{1,i}^*-\mu_{X_1})(Y_{1,i}^*-\mu_{Y_1})-(X_{2,i}^*-\mu_{X_2})(Y_{2,i}^*-\mu_{Y_2}))\\
&-\frac{1}{\sqrt{n}}\sum_{i=1}^{n}((X_{1,i}-\mu_{X_1})(Y_{1,i}-\mu_{Y_1})-(X_{2,i}-\mu_{X_2})(Y_{2,i}-\mu_{Y_2}))-\sqrt{n}(\overline{X}_1^*-\mu_{X_1})(\overline{Y}_1^*-\mu_{Y_1})\\
&+\sqrt{n}(\overline{X}_2^*-\mu_{X_2})(\overline{Y}_2^*-\mu_{Y_2}))+\sqrt{n}(\overline{X}_1-\mu_{X_1})(\overline{Y}_1-\mu_{Y_1})-\sqrt{n}(\overline{X}_2-\mu_{X_2})(\overline{Y}_2-\mu_{Y_2})\\
&=\sqrt{n}S_n^*+\frac{1}{\sqrt{n}}\sqrt{n}(\overline{X}_1^*-\overline{X}_1)\sqrt{n}(\overline{Y}_1^*-\overline{Y}_1)-\sqrt{n}(\overline{X}_1^*-\overline{X}_1)(\overline{Y}_1^*-\mu_{Y_1})-(\overline{X}_1^*-\mu_{X_1})\sqrt{n}(\overline{Y}_1^*-\overline{Y}_1)\\
&-\frac{1}{\sqrt{n}}\sqrt{n}(\overline{X}_2^*-\overline{X}_2)\sqrt{n}(\overline{Y}_2^*-\overline{Y}_2)+\sqrt{n}(\overline{X}_2^*-\overline{X}_2)(\overline{Y}_2^*-\mu_{Y_2})+(\overline{X}_2^*-\mu_{X_2})\sqrt{n}(\overline{Y}_2^*-\overline{Y}_2),
\end{aligned}\tag{5.34}$$

where S_n^* is defined as

$$\begin{aligned}
S_n^*&=\frac{1}{n}\sum_{i=1}^{n}((X_{1,i}^*-\mu_{X_1})(Y_{1,i}^*-\mu_{Y_1})-(X_{2,i}^*-\mu_{X_2})(Y_{2,i}^*-\mu_{Y_2}))\\
&\quad-\frac{1}{n}\sum_{i=1}^{n}((X_{1,i}-\mu_{X_1})(Y_{1,i}-\mu_{Y_1})-(X_{2,i}-\mu_{X_2})(Y_{2,i}-\mu_{Y_2})).
\end{aligned}$$

Given that $\{(X_{1,i}^*-\mu_{X_1})(Y_{1,i}^*-\mu_{Y_1})-(X_{2,i}^*-\mu_{X_2})(Y_{2,i}^*-\mu_{Y_2})\}_{i=1}^n$ are i.i.d. $\mathcal{H}$–valued random elements chosen at random from the *bootstrap population* $\{(X_{1,i}-\mu_{X_1})(Y_{1,i}-\mu_{Y_1})-(X_{2,i}-\mu_{X_2})(Y_{2,i}-\mu_{Y_2})\}_{i=1}^n$ and $\mathbb{E}(\|(X_1-\mu_{X_1})(Y_1-\mu_{Y_1})-(X_2-\mu_{X_2})(Y_2-\mu_{Y_2})\|^2)<\infty$, the bootstrap statistic $\sqrt{n}S_n^*$ converges in law to $Z_{(X_1-\mu_{X_1})(Y_1-\mu_{Y_1})-(X_2-\mu_{X_2})(Y_2-\mu_{Y_2})}$ $a.s.-\mathbb{P}$ due to item 1 in Lemma 5.7.1 (see page 135).

In addition, the last six terms in (5.34) converge in probability to 0 $a.s.-\mathbb{P}$ by items 1 and 2 in Lemma 5.7.1, together with Slutsky's Theorem, since $\mathbb{E}(\|X_1\|^2)<\infty$, $\mathbb{E}(\|X_2\|^2)<\infty$, $\mathbb{E}(Y_1^2)<\infty$ and $\mathbb{E}(Y_2^2)<\infty$. As a result, the convergence in law stated in the theorem is proven applying again the Slutsky's Theorem.

5.7.12 Proof of Theorem 5.3.10

Remark that, for all $\omega \in \Omega$,

$$\begin{aligned}
\sqrt{n}\, T_n^{Eq,W*}(\omega) &= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} ((X_{1,i}(\omega) - \mu_{X_1})(Y_{1,i}(\omega) - \mu_{Y_1}) - (X_{2,i}(\omega) - \mu_{X_2})(Y_{2,i}(\omega) - \mu_{Y_2}))\epsilon_i^* \\
&- \left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} (X_{1,i}(\omega) - \mu_{X_1})\epsilon_i^* \right) (\overline{Y}_1(\omega) - \mu_{Y_1}) - (\overline{X}_1(\omega) - \mu_{X_1}) \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (Y_{1,i}(\omega) - \mu_{Y_1})\epsilon_i^* \\
&+ (\overline{X}_1(\omega) - \mu_{X_1})(\overline{Y}_1(\omega) - \mu_{Y_1}) \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \epsilon_i^* + \left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} (X_{2,i}(\omega) - \mu_{X_2})\epsilon_i^* \right) (\overline{Y}_2(\omega) - \mu_{Y_2}) \\
&+ (\overline{X}_2(\omega) - \mu_{X_2}) \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (Y_{2,i}(\omega) - \mu_{Y_2})\epsilon_i^* - (\overline{X}_2(\omega) - \mu_{X_2})(\overline{Y}_2(\omega) - \mu_{Y_2}) \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \epsilon_i^* \\
&= S_n^* - \left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} (X_{1,i}(\omega) - \mu_{X_1})\epsilon_i^* \right) (\overline{Y}_1(\omega) - \mu_{Y_1}) - (\overline{X}_1(\omega) - \mu_{X_1}) \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (Y_{1,i}(\omega) - \mu_{Y_1})\epsilon_i^* \\
&+ \left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} (X_{2,i}(\omega) - \mu_{X_2})\epsilon_i^* \right) (\overline{Y}_2(\omega) - \mu_{Y_2}) + (\overline{X}_2(\omega) - \mu_{X_2}) \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (Y_{2,i}(\omega) - \mu_{Y_2})\epsilon_i^* \\
&+ ((\overline{X}_1(\omega) - \mu_{X_1})(\overline{Y}_1(\omega) - \mu_{Y_1}) - (\overline{X}_2(\omega) - \mu_{X_2})(\overline{Y}_2(\omega) - \mu_{Y_2})) \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \epsilon_i^*,
\end{aligned} \tag{5.35}$$

where $S_n^* = n^{-1/2} \sum_{i=1}^{n} ((X_{1,i}(\omega) - \mu_{X_1})(Y_{1,i}(\omega) - \mu_{Y_1}) - (X_{2,i}(\omega) - \mu_{X_2})(Y_{2,i}(\omega) - \mu_{Y_2}))\epsilon_i^*$. Applying Lemma 5.7.2 (see page 136), S_n^* converges in law to $Z_{(X_1-\mu_{X_1})(Y_1-\mu_{Y_1})-(X_2-\mu_{X_2})(Y_2-\mu_{Y_2})}$, for almost all $\omega \in \Omega$, since $\mathbb{E}(\|(X_1 - \mu_{X_1})(Y_1 - \mu_{Y_1}) - (X_2 - \mu_{X_2})(Y_2 - \mu_{Y_2})\|^2) < \infty$.

Furthermore, $(\overline{X}_1(\omega) - \mu_{X_1})$, $(\overline{Y}_1(\omega) - \mu_{Y_1})$, $(\overline{X}_2(\omega) - \mu_{X_2})$ and $(\overline{Y}_2(\omega) - \mu_{Y_2})$ converge to 0 by the SLLN. This fact, Lemma 5.7.2 and Slutsky's Theorem ensure the convergence in probability to 0 of the last six summands in (5.35), given that $\mathbb{E}(\|X_1 - \mu_{X_1}\|^2) < \infty$, $\mathbb{E}((Y_1 - \mu_{Y_1})^2) < \infty$, $\mathbb{E}(\|X_2 - \mu_{X_2}\|^2) < \infty$ and $\mathbb{E}((Y_2 - \mu_{Y_2})^2) < \infty$. Consequently, the theorem is proven by the application of the Slutsky's Theorem.

Chapter 6

Thresholding in nonparametric functional regression

The main goal of this chapter is to introduce an exploratory tool that allows the detection of underlying complex structures in functional data. The chapter is focused on the nonparametric regression model with scalar response Y and functional covariate X, and presents a methodology for the analysis of hidden patterns via a threshold procedure. Furthermore, several threshold functions are proposed, based on X and/or Y, and a cross–validation criterion is introduced in order to estimate the threshold value. A graphical tool is also obtained as a by–product, which can be helpful to decide if there are concealed structures that require a more detailed analysis of the data. The effectiveness of the proposal is shown by means of a simulation study and its application to real datasets.

The threshold approach compiled in this chapter was introduced in Ferraty et al. (2011b, 2012b).

6.1 Why consider a threshold approach?

In previous chapters, the functional regression with scalar response (specifically, the functional linear model) has been analysed from a *parametric* viewpoint. However, the more general nonparametric framework has been considered in this chapter. Recall the functional nonparametric regression model for scalar response defined in (2.9) (see Chapter 2, page 41)

$$Y = m(X) + \epsilon,$$

where Y is a real random variable, X is a random variable valued in a separable Hilbert space $(\mathcal{H}, \langle\cdot,\cdot\rangle)$, $m : \mathcal{H} \to \mathbb{R}$ is the regression operator, and ϵ is a real random variable such that $\mathbb{E}(\epsilon|X) = 0$ and $\mathbb{E}(\epsilon^2|X) = \sigma_\epsilon^2(X) < \infty$. A particularly interesting situation is $\mathcal{H} = L^2[a,b]$ with $a, b \in \mathbb{R}$ (note that X is a curve in this case, and the inner product is given by $\langle x, y\rangle = \int_a^b x(t)y(t)dt$ for all $x, y \in L^2[a,b]$), since most of the available functional datasets are this kind of data. Due to this, the simulation study and the real applications which are included in this chapter correspond to datasets where the covariate is a curve. Nevertheless the notion of functional variable covers other types of data. That is why all theoretical results for the techniques proposed in this chapter were developed for the wider context of a general separable Hilbert space $(\mathcal{H}, \langle\cdot,\cdot\rangle)$. This fact guarantees the application of these theoretical advances to a larger area of functional data, not only curves, but also any other elements belonging to a separable Hilbert space.

Sometimes, one is confronted with complex regression structures, which are unlikely detectable using standard graphical or descriptive techniques, such as the existence of different subsamples of functional covariates or different regression models in the sample. These situations can lead to obtain

biased results or incorrect conclusions if the hidden patterns are not taken into account. The objective of this chapter is to detect this kind of structures by means of an exploratory method which studies the existence of a threshold value in the covariate X or in the real response Y and, when this threshold exists, estimates it by means of a cross–validation procedure. Moreover, the cross–validation criterion can be plotted providing a graphical support when one has to decide if there is any threshold on the dataset or not.

Contributions on thresholding applications in FDA field are currently few. For instance, Bunea et al. (2011) proposed data–driven methods based on thresholded least squares estimators for inference about the mean function of a Gaussian process. Their aim was to use the thresholding approach to adapt estimates to the sparsity of the mean function related to the approximating basis. On the other hand, Chesneau et al. (2012) introduced an adaptative methodology to deal with both density and regression estimation problems based on wavelet thresholding. Nevertheless, as far as the author is concerned, the use of thresholding in the nonparametric regression context in order to find hidden patterns in a dataset has not been previously studied in FDA literature.

Section 6.2 is devoted to introduce the methodology and show some potentially interesting threshold functions. Furthermore, some theoretical properties are also presented. Then, the method is tested with a simulation study and real data applications, being the main results collected in Section 6.3 and Section 6.4. Finally, a brief discussion of the main conclussions of the work can be found in Section 6.5, and proofs and technical lemmas have been compiled in the appendix of the chapter (see Section 6.6).

6.2 Threshold methodology

Firstly, the threshold estimator is introduced in Section 6.2.1. In addition, some families of threshold functions are proposed, and a theoretical result related with the mean square convergence of the threshold estimator is also presented. Next, a cross–validation method which allows for estimating the threshold value, when it exists, is built in Section 6.2.2, and its optimality with respect to MISE criterium is studied. From now on, let $\{(X_i, Y_i)\}_{i=0}^n$ be an of i.i.d. sample of (X, Y).

6.2.1 Regression model and estimate

The key of the procedure is to rewrite the regression operator $m(\cdot)$ given by

$$\begin{array}{rccl} m: & \mathcal{H} & \to & \mathbb{R} \\ & x & \to & m(x) = \mathbb{E}(Y|X = x) \end{array}$$

as a finite sum of operators as follows. First of all, let Ψ be a function such that $\Psi : \mathcal{H} \times \mathbb{R} \to \mathcal{E}$, being $\mathcal{E}$ a beforehand fixed space, and let $\{(\mathcal{E}_1^\upsilon, \ldots, \mathcal{E}_{N_\mathcal{E}}^\upsilon)\}_{\upsilon \in \Upsilon}$ be an indexed family of sets where $N_\mathcal{E}$ is a fixed integer such that $1 < N_\mathcal{E} < \infty$ and, for all $\upsilon \in \Upsilon$,

$$\begin{cases} \mathcal{E}_s^\upsilon \subset \mathcal{E} \quad \forall s \in S, \quad \mathcal{E}_{s_1}^\upsilon \cap \mathcal{E}_{s_2}^\upsilon = \emptyset \quad \forall s_1, s_2 \in S \text{ such that } s_1 \neq s_2, \\ \mathbb{P}(\Psi(X, Y) \in \mathcal{E}_s^\upsilon) > 0 \quad \forall s \in S, \quad \text{and} \quad \mathbb{P}(\Psi(X, Y) \in \bigcup_{s \in S} \mathcal{E}_s^\upsilon) = 1, \end{cases}$$

where $S = \{1, \ldots, N_\mathcal{E}\}$. For each $s \in S$, the next definitions can be introduced $Y_s^\upsilon = Y \mathbb{I}_{\{\Psi(X,Y) \in \mathcal{E}_s^\upsilon\}}$ being $\mathbb{I}$ the indicator function,

$$\begin{array}{rccl} m_s^\upsilon: & \mathcal{H} & \to & \mathbb{R} \\ & x & \to & m_s^\upsilon(x) = \mathbb{E}(Y_s^\upsilon|X = x), \end{array}$$

and $\epsilon_s^\upsilon = \epsilon \mathbb{I}_{\{\Psi(X,Y) \in \mathcal{E}_s^\upsilon\}}$. Thus, one gets that $Y = \sum_{s \in S} Y_s^\upsilon$, $m(x) = \sum_{s \in S} m_s^\upsilon(x)$, and $\epsilon = \sum_{s \in S} \epsilon_s^\upsilon$. Consequently, the regression model (2.9) (see Chapter 2, page 41) can be expressed as

$$\sum_{s \in S} Y_s^\upsilon = \sum_{s \in S} m_s^\upsilon(X) + \sum_{s \in S} \epsilon_s^\upsilon.$$

Once the regression operator $m(\cdot)$ is written as the sum of the operators m_s^υ, each component m_s^υ can be estimated separately, using the sample $\{(X_i, Y_{i,s}^\upsilon)\}_{i=1}^n$, where $Y_{i,s}^\upsilon = Y_i \mathbb{I}_{\{\Psi(X_i,Y_i)\in\mathcal{E}_s^\upsilon\}}$ for $i = 1, \ldots, n$. Thus, an indexed family of estimators can be built by means of

$$\hat{m}^\upsilon(x) = \sum_{s\in S} \hat{m}_s^\upsilon(x), \quad \forall \upsilon \in \Upsilon.$$

Specifically, for each $s \in S$, the kernel–type estimator (2.10) (see Chapter 2, page 42) has been taken with the semi–metric induced by the inner product of $\mathcal{H}$ as the semi–metric $d(\cdot,\cdot)$, that is,

$$\hat{m}_s^\upsilon(x) = \frac{\sum_{i=1}^n Y_{i,s}^\upsilon K(h_s^{-1}\|X_i - x\|)}{\sum_{i=1}^n K(h_s^{-1}\|X_i - x\|)},$$

where $\|\cdot\| = \langle\cdot,\cdot\rangle^{1/2}$ is the induced norm of $\mathcal{H}$, $K(\cdot)$ is a kernel function, and $h_s = h_s(n)$ is a sequence of strictly positive real bandwidths such that $h_s \in H_n \subset \mathbb{R}^+$ for all $s \in S$. In order to simplify the notation, let $\tilde{K}_{s,i}$ be $\tilde{K}_{s,i}(x) = K(h_s^{-1}\|X_i - x\|)$, and introduce the following terms

$$\hat{m}_{s,N}^\upsilon(x) = \frac{1}{n}\sum_{i=1}^n \frac{Y_{i,s}^\upsilon \tilde{K}_{s,i}(x)}{\mathbb{E}(\tilde{K}_{s,0}(x))} \quad \text{and} \quad \hat{m}_{s,D}^\upsilon(x) = \frac{1}{n}\sum_{i=1}^n \frac{\tilde{K}_{s,i}(x)}{\mathbb{E}(\tilde{K}_{s,0}(x))}.$$

Hence, the estimator $\hat{m}^\upsilon$ previously introduced can be expressed as

$$\hat{m}^\upsilon(x) = \sum_{s\in S} \frac{\sum_{i=1}^n Y_{i,s}^\upsilon \tilde{K}_{s,i}(x)}{\sum_{i=1}^n \tilde{K}_{s,i}(x)} = \sum_{s\in S} \frac{\hat{m}_{s,N}^\upsilon(x)}{\hat{m}_{s,D}^\upsilon(x)}. \tag{6.1}$$

Remark 6.2.1. When the same bandwidth is selected for all the components $\hat{m}_s^\upsilon$ in the estimator $\hat{m}^\upsilon$, that is, when exists $h \in H_n$ such that $h_s = h$ for all $s \in S$, the proposed estimator (6.1) is just the standard kernel–type estimator (2.10) (see Chapter 2, page 42) with the semi–metric induced by the inner product of $\mathcal{H}$, i.e., $d(x,y) = \|x - y\| = \langle x, y\rangle^{1/2}$ for all $x, y \in \mathcal{H}$, given by

$$\hat{m}_h(x) = \frac{\sum_{i=1}^n Y_i K(h^{-1}\|X_i - x\|)}{\sum_{i=1}^n K(h^{-1}\|X_i - x\|)}, \tag{6.2}$$

which has been studied in detail during the last few years (see, for instance, Ferraty and Vieu, 2006b; Ferraty et al., 2007a, 2010b).

Examples of threshold functions

The proposed method requires the specification of a threshold function in advance, and this choice should be done as far as possible in accordance with the pattern to be possibly detected in the dataset. Next, some useful threshold functions are presented, which can be taken into consideration when $\mathcal{H} = L^2[a,b]$, $\mathcal{E} = \mathbb{R}$, $\Upsilon \subset \mathbb{R}$, $N_\mathcal{E} = 2$ and the indexed family of pairs $\{(\mathcal{E}_1^\upsilon, \mathcal{E}_2^\upsilon)\}_{\upsilon\in\Upsilon}$ is given by $\mathcal{E}_1^\upsilon = (-\infty, \upsilon]$ and $\mathcal{E}_2^\upsilon = (\upsilon, +\infty)$ for each $\upsilon \in \Upsilon$.

Threshold on Y. When one suspects that there is a threshold connected to the response, functions which only depend on Y, i.e., $\Psi : L^2[a,b] \times \mathbb{R} \to \mathbb{R}$ such that $\Psi(x,y) = f(y)$ with $f : \mathbb{R} \to \mathbb{R}$, can be considered. The most adequate function, according to the kind of structure one wants to detect, should be selected: $\Psi(x,y) = |y|$, $\Psi(x,y) = \log(y)$, $\Psi(x,y) = \exp(y)$, $\Psi(x,y) = \cos(y)$,...

Threshold on X. If the threshold is only related to the covariate, then threshold functions only depending on X, such as $\Psi : L^2[a,b] \times \mathbb{R} \to \mathbb{R}$ with $\Psi(x,y) = f(x)$ being $f : L^2[a,b] \to \mathbb{R}$, can be a good choice. For instance, the following alternatives could be used:

- From the induced norm of $L^2[a,b]$ defined as $\|x\| = (\int_a^b x^2(t)dt)^{1/2}$ for all $x \in L^2[a,b]$, a family of semi–norms can be constructed by means of $\Psi(x,y) = \|x^{(q)}\|$, where $x^{(q)}$ is the qth derivative of the curve x. Note that if $q = 0$, this threshold function is just the norm of $L^2[a,b]$.

- Selecting $\{e_j\}_{j=1}^{+\infty}$ an orthonormal basis of $L^2[a,b]$ (according to the kind of data, it could be B–splines basis, Fourier basis, principal components basis,...), and projecting the data on their first J elements, one can define $\Psi(x,y) = \|x_J\|_J$, where $x_J = (\langle x, e_1\rangle, \ldots, \langle x, e_J\rangle)^t$ and $\|\cdot\|_J$ is a norm in $\mathbb{R}^J$ (e.g., $\|x_J\|_J = \sqrt{x_J^t M x_J}$ being M a fixed $J \times J$–matrix).
- When there is one interesting direction $e \in L^2[a,b]$, one can consider $\Psi(x,y) = |\langle x, e\rangle|$.
- In other cases, the dataset nature can lead to choose another type of functions, such as $\Psi(x,y) = \max_{t\in[0,1]} |x(t)|$, $\Psi(x,y) = \int_0^1 x(t)dt$ or $\Psi(x,y) = \|x - x_0\|$ for a fixed $x_0 \in L^2[a,b]$.

Obviously, more complicated Ψ functions can be selected. For example, threshold functions which depend simultaneously on X and Y, or related with the projection of the curve onto several directions (e.g., $\Psi(x,y) = (|\langle x, e_1\rangle|, \ldots, |\langle x, e_J\rangle|)$) can be considered. One can also construct threshold functions based on measures of *concentration* of the curves such as the small ball probability, that is, $\Psi(x,y) = \varphi_x(h) = \mathbb{P}(X \in B(x,h))$ (in this case, an estimation for φ_x and for h should be previously obtained), or on any other measure related with X and/or Y. However, one must bear in mind that these options will probably imply an increment of the computational cost of the estimation process.

A particular scenario

Suppose that

(C.6.1) there exists a compact set $\mathcal{C}$ of $\mathcal{H}$ such that $\mathbb{P}(X \in \mathcal{C}) = 1$.

Moreover, assume that $\mathcal{E}$ is a metric space, provided with a metric $\rho_{\mathcal{E}}(\cdot,\cdot)$, and Ψ is a threshold function only related to the covariate X such that

(C.6.2) $\Psi : \mathcal{H} \times \mathbb{R} \to \mathcal{E}$ with $\Psi(x,y) = \tilde{\Psi}(x)$ for all $(x,y) \in \mathcal{H} \times \mathbb{R}$, where $\tilde{\Psi} : \mathcal{H} \to \mathcal{E}$ is continuous on $\mathcal{C}$.

Note that when the threshold only depends on X

$$Y_s^{\upsilon} = Y\mathbb{I}_{\{\tilde{\Psi}(X)\in\mathcal{E}_s^{\upsilon}\}} \quad \text{and} \quad m_s^{\upsilon}(X) = \mathbb{E}(Y_s^{\upsilon}|X) = \mathbb{E}(Y|X)\mathbb{I}_{\{\tilde{\Psi}(X)\in\mathcal{E}_s^{\upsilon}\}} = m(X)\mathbb{I}_{\{\tilde{\Psi}(X)\in\mathcal{E}_s^{\upsilon}\}}. \tag{6.3}$$

Given that $\tilde{\Psi}$ is continuous on a compact set $\mathcal{C}$, $\tilde{\Psi}$ is uniformly continuous on $\mathcal{C}$ by Heine–Cantor theorem. Thus, for any $\varepsilon > 0$, there is a $\delta > 0$ such that

$$\forall x_1, x_2 \in \mathcal{C} \text{ satisfying } \|x_1 - x_2\| < \delta, \text{ the inequality } \rho_{\mathcal{E}}(\tilde{\Psi}(x_1), \tilde{\Psi}(x_2)) < \varepsilon \text{ holds.} \tag{6.4}$$

On the other hand, let $\{(\mathcal{E}_1^{\upsilon}, \ldots, \mathcal{E}_{N_{\mathcal{E}}}^{\upsilon})\}_{\upsilon\in\Upsilon}$ be an indexed family of sets such that $N_{\mathcal{E}}$ is finite and

(C.6.3) $$\begin{cases} \mathcal{E}_s^{\upsilon} \subset \mathcal{E} \quad \forall s \in S, \quad \mathcal{E}_{s_1}^{\upsilon} \cap \mathcal{E}_{s_2}^{\upsilon} = \emptyset \quad \forall s_1, s_2 \in S \text{ such that } s_1 \neq s_2, \\ \mathbb{P}(\tilde{\Psi}(X) \in \mathcal{E}_s^{\upsilon}) \geq c_0 > 0 \quad \forall s \in S, \quad \text{and} \quad \mathbb{P}(\tilde{\Psi}(X) \in \bigcup_{s\in S} \mathcal{E}_s^{\upsilon}) = 1, \end{cases}$$

which also satisfies the following condition

(C.6.4) there exists $D > 0$ such that

$$D = \min_{s_1,s_2\in S, s_1\neq s_2} \rho_{\mathcal{E}}(\mathcal{E}_{s_1}^{\upsilon}, \mathcal{E}_{s_2}^{\upsilon}), \quad \text{where} \quad \rho_{\mathcal{E}}(\mathcal{E}_{s_1}^{\upsilon}, \mathcal{E}_{s_2}^{\upsilon}) = \inf_{e_1\in\mathcal{E}_{s_1}^{\upsilon}, e_2\in\mathcal{E}_{s_2}^{\upsilon}} \rho_{\mathcal{E}}(e_1, e_2).$$

It is important to highlight that, since $\tilde{\Psi}$ is uniformly continuous (see (6.4)) and $D > 0$, there exists $\delta_D > 0$ such that for all $x_1, x_2 \in \mathcal{C}$ verifying $\|x_1 - x_2\| < \delta_D$, $\rho_{\mathcal{E}}(\tilde{\Psi}(x_1), \tilde{\Psi}(x_2)) < D$, and as result, $\tilde{\Psi}(x_1)$ and $\tilde{\Psi}(x_2)$ belongs to the same subset of $\{(\mathcal{E}_1^{\upsilon}, \ldots, \mathcal{E}_{N_{\mathcal{E}}}^{\upsilon})\}_{\upsilon\in\Upsilon}$. Hence,

$$\forall x_1, x_2 \in \mathcal{C} \text{ satisfying } \|x_1 - x_2\| < \delta_D, \quad \mathbb{I}_{\{\tilde{\Psi}(x_1)\in\mathcal{E}_s^{\upsilon}\}} = \mathbb{I}_{\{\tilde{\Psi}(x_2)\in\mathcal{E}_s^{\upsilon}\}}, \quad \forall s \in S. \tag{6.5}$$

Therefore, in the situation that has just been described, if the bandwidth h_s satisfies $h_s < \delta_D$, (6.3) and (6.5) imply that $\hat{m}_s^{\upsilon}(x)$ is indeed

$$\hat{m}_s^{\upsilon}(x) = \frac{\sum_{i=1}^n Y_i \mathbb{I}_{\{\tilde{\Psi}(X_i)\in\mathcal{E}_s^{\upsilon}\}} \tilde{K}_{s,i}(x)}{\sum_{i=1}^n \tilde{K}_{s,i}(x)} = \frac{\sum_{i=1}^n Y_i \tilde{K}_{s,i}(x)}{\sum_{i=1}^n \tilde{K}_{s,i}(x)} \mathbb{I}_{\{\tilde{\Psi}(x)\in\mathcal{E}_s^{\upsilon}\}} = \hat{m}_{h_s}(x)\mathbb{I}_{\{\tilde{\Psi}(x)\in\mathcal{E}_s^{\upsilon}\}}, \tag{6.6}$$

where $\hat{m}_{h_s}$ is the standard kernel estimator with bandwidth h_s (see (6.2), page 143). In addition, since there is only a $s_x \in S$ such that $\tilde{\Psi}(x) \in \mathcal{E}_{s_x}^{\upsilon}$, $\hat{m}_{s_x}^{\upsilon}(x)$ is the only non–null component of $\hat{m}^{\upsilon}(x)$ and, consequently, $\hat{m}^{\upsilon}(x) = \sum_{s\in S} \hat{m}_s^{\upsilon}(x) = \hat{m}_{s_x}^{\upsilon}(x) = \hat{m}_{h_{s_x}}(x)$.

Remark 6.2.2. In a certain sense, $\hat{m}^{\upsilon}(x)$ can be seen as a kernel–type estimator with "local" bandwith which depends on the value $\tilde{\Psi}(x)$: for all $x \in \mathcal{C}$ such that $\tilde{\Psi}(x) \in \mathcal{E}_1^{\upsilon}$, the kernel estimator is computed using the bandwidth h_1; for all $x \in \mathcal{C}$ such that $\tilde{\Psi}(x) \in \mathcal{E}_2^{\upsilon}$, the kernel estimator is computed using the bandwidth h_2,...

Mean square convergence

In the following, let $x \in \mathcal{C}$ be a fixed element, such that s_x denotes the only $s_x \in S$ such that $\tilde{\Psi}(x) \in \mathcal{E}_{s_x}^{\upsilon}$. Expectation and variance of each component $\hat{m}_s^{\upsilon}$ of the proposed threshold estimator $\hat{m}^{\upsilon}$ (see (6.1), page 143) are given in this section for the scenario described above. Hence, the expectation and the variance of $\hat{m}^{\upsilon}$ can be obtained as a corollary.

Next, before formulating the theoretical results of this section, recall the definition of ψ_x and $\tau_{x,h}$, which were introduced by Ferraty et al. (2007a) in order to compute the bias and variance expressions of the standard kernel estimator (see Section 2.4.1, *"c) Bias and variance"*, in Chapter 2, page 45),

$$\psi_x(t) = \mathbb{E}\left((m(X) - m(x)) \,|\, \|X - x\| = t\right), \quad \forall t \in \mathbb{R},$$

and

$$\tau_{x,h}(t) = \frac{\varphi_x(ht)}{\varphi_x(h)}, \quad \forall t \in [0,1],$$

where φ_x is the small ball probability given by $\varphi_x(h) = \mathbb{P}(X \in B(x,h)) = \mathbb{P}(\|X - x\| \leq h)$. Moreover, the following assumptions are also required for stating the results below:

(C.6.5) $\varphi_x(0) = 0$, and $\tau_{x,h}(t) \to \tau_{x,0}(t)$ as $h \to 0$ for all $t \in [0,1]$.

(C.6.6) $m(\cdot)$ and $\sigma_\epsilon^2(\cdot)$ are continuous in a neighbourhood of x.

(C.6.7) $\psi_x'(0)$ exists.

(C.6.8) For all $s \in S$, the sequence of bandwiths h_s verifies that $\lim_{n\to+\infty} h_s = 0$, $\lim_{n\to+\infty} n\varphi_x(h_s) = +\infty$, and $h_s < \delta_D$ (with δ_D defined as in (6.5)).

(C.6.9) $K(\cdot)$ is a kernel supported on $[0,1]$ with a continuous derivative on $[0,1)$ such that $K(1) > 0$ and $K'(t) < 0$.

Theorem 6.2.3. *Under* **(C.6.1)**–**(C.6.4)**, *if* **(C.6.5)**–**(C.6.9)** *are satisfied, then for all* $s \in S$

$$\mathbb{E}(\hat{m}_s^{\upsilon}(x)) = \left(m(x) + \psi_x'(0)\frac{M_{x,0}}{M_{x,1}}h_s + O\left(\frac{1}{n\varphi_x(h_s)}\right) + o(h_s)\right)\mathbb{I}_{\{\tilde{\Psi}(x)\in\mathcal{E}_s^{\upsilon}\}},$$

$$Var(\hat{m}_s^{\upsilon}(x)) = \left(\sigma_\epsilon^2(x)\frac{M_{x,2}}{M_{x,1}^2}\frac{1}{n\varphi_x(h_s)} + o\left(\frac{1}{n\varphi_x(h_s)}\right)\right)\mathbb{I}_{\{\tilde{\Psi}(x)\in\mathcal{E}_s^{\upsilon}\}},$$

$$Cov(\hat{m}_{s_1}^{\upsilon}(x), \hat{m}_{s_2}^{\upsilon}(x)) = 0,$$

with $M_{x,0} = K(1) - \int_0^1 (tK(t))'\tau_{x,0}(t)dt$, $M_{x,1} = K(1) - \int_0^1 K'(t)\tau_{x,0}(t)dt$ *and* $M_{x,2} = K^2(1) - \int_0^1 (K^2)'(t)\tau_{x,0}(t)dt$.

Corollary 6.2.4. *Under the assumptions of Theorem 6.2.3, it holds that*

$$\mathbb{E}(\hat{m}^{\upsilon}(x)) = m(x) + \psi_x'(0)\frac{M_{x,0}}{M_{x,1}}\sum_{s\in S} h_s \mathbb{I}_{\{\tilde{\Psi}(x)\in\mathcal{E}_s^{\upsilon}\}} + \sum_{s\in S}\left(O\left(\frac{1}{n\varphi_x(h_s)}\right) + o(h_s)\right)\mathbb{I}_{\{\tilde{\Psi}(x)\in\mathcal{E}_s^{\upsilon}\}},$$

$$Var(\hat{m}^{\upsilon}(x)) = \sigma_{\epsilon}^2(x)\frac{M_{x,2}}{M_{x,1}^2}\sum_{s\in S}\frac{1}{n\varphi_x(h_s)}\mathbb{I}_{\{\tilde{\Psi}(x)\in\mathcal{E}_s^{\upsilon}\}} + \sum_{s\in S}\left(o\left(\frac{1}{n\varphi_x(h_s)}\right)\right)\mathbb{I}_{\{\tilde{\Psi}(x)\in\mathcal{E}_s^{\upsilon}\}},$$

with $M_{x,0} = K(1) - \int_0^1 (tK(t))'\tau_{x,0}(t)dt$, $M_{x,1} = K(1) - \int_0^1 K'(t)\tau_{x,0}(t)dt$ *and* $M_{x,2} = K^2(1) - \int_0^1 (K^2)'(t)\tau_{x,0}(t)dt$. *In particular, if* s_x *denotes the only* $s_x \in S$ *such that* $\tilde{\Psi}(x) \in \mathcal{E}_{s_x}^{\upsilon}$, *then*

$$\mathbb{E}(\hat{m}^{\upsilon}(x)) = m(x) + \psi_x'(0)\frac{M_{x,0}}{M_{x,1}}h_{s_x} + O\left(\frac{1}{n\varphi_x(h_{s_x})}\right) + o(h_{s_x}),$$

$$Var(\hat{m}^{\upsilon}(x)) = \sigma_{\epsilon}^2(x)\frac{M_{x,2}}{M_{x,1}^2}\frac{1}{n\varphi_x(h_{s_x})} + o\left(\frac{1}{n\varphi_x(h_{s_x})}\right).$$

The proofs of the previous results can be found in the appendix of the chapter (see Section 6.6.1 and Section 6.6.2, respectively).

Remark 6.2.5. Theorem 6.2.3 is a direct consequence of Theorem 1 in Ferraty et al. (2007a) (see Theorem 2.4.13 in Chapter 2, page 46), since $\hat{m}_s^{\upsilon}(x)$ can be expressed as (6.6) (see page 145).

6.2.2 Cross–validation criterion

Several parameters need to be estimated in $\hat{m}^{\upsilon}$ (see (6.1), page 143): the threshold υ and the $N_{\mathcal{E}}$ bandwidths $\{h_s\}_{s\in S}$. From now on, the following notation is going to be used: $(\{h_s\}_{s\in S}) \equiv (h_1, \ldots, h_{N_{\mathcal{E}}})$, $(\upsilon, \{h_s\}_{s\in S}) \equiv (\upsilon, h_1, \ldots, h_{N_{\mathcal{E}}})$, $H_n^{N_{\mathcal{E}}} \equiv H_n \times \overset{N_{\mathcal{E}}}{\cdots} \times H_n$, and $\Upsilon \times H_n^{N_{\mathcal{E}}} \equiv \Upsilon \times H_n \times \overset{N_{\mathcal{E}}}{\cdots} \times H_n$.

To obtain adequate values for $(\upsilon, \{h_s\}_{s\in S})$, it is proposed to use one of the most widespread techniques in the literature: a cross–validation method (see, for instance, Härdle and Marron, 1985). In this case, the aim is to find $(\upsilon, \{h_s\}_{s\in S}) \in \Upsilon \times H_n^{N_{\mathcal{E}}}$ minimizing the following cross–validation criterion

$$\mathrm{CV}(\upsilon, \{h_s\}_{s\in S}) = \frac{1}{n}\sum_{j=1}^{n}(Y_j - \hat{m}^{\upsilon,(-j)}(X_j))^2, \tag{6.7}$$

where $\hat{m}^{\upsilon,(-j)}(x) = \sum_{s\in S}\hat{m}_s^{\upsilon,(-j)}(x)$ being

$$\hat{m}_s^{\upsilon,(-j)}(x) = \frac{\sum_{i\neq j} Y_{i,s}^{\upsilon}\tilde{K}_{s,i}(x)}{\sum_{i\neq j}\tilde{K}_{s,i}(x)}.$$

Moreover, adopting the following notation

$$\hat{m}_{s,N}^{\upsilon,(-j)}(x) = \frac{1}{n-1}\sum_{i\neq j}\frac{Y_{i,s}^{\upsilon}\tilde{K}_{s,i}(x)}{\mathbb{E}(\tilde{K}_{s,0}(x))} \quad \text{and} \quad \hat{m}_{s,D}^{\upsilon,(-j)}(x) = \frac{1}{n-1}\sum_{i\neq j}\frac{\tilde{K}_{s,i}(x)}{\mathbb{E}(\tilde{K}_{s,0}(x))},$$

then $\hat{m}^{\upsilon,(-j)}$ can be built by means of the next expression

$$\hat{m}^{\upsilon,(-j)}(x) = \sum_{s\in S}\frac{\sum_{i\neq j} Y_{i,s}^{\upsilon}\tilde{K}_{s,i}(x)}{\sum_{i\neq j}\tilde{K}_{s,i}(x)} = \sum_{s\in S}\frac{\hat{m}_{s,N}^{\upsilon,(-j)}(x)}{\hat{m}_{s,D}^{\upsilon,(-j)}(x)}.$$

Minimizing the CV criterion (6.7), the model parameters $(\upsilon, \{h_s\}_{s\in S})$ in (6.1) (see page 143) will be estimated by

$$(\upsilon_{\mathrm{CV}}, \{h_{s,\mathrm{CV}}\}_{s\in S}) = \arg\min_{(\upsilon,\{h_s\}_{s\in S})\in\Upsilon\times H_n^{N_{\mathcal{E}}}} \mathrm{CV}(\upsilon, \{h_s\}_{s\in S}).$$

The theoretical results below are focused on showing the optimality of this data–driven selection regarding to the mean integrated squared error given by

$$\mathrm{MISE}(\upsilon, \{h_s\}_{s\in S}) = \mathbb{E}((m(X) - \hat{m}^{\upsilon}(X))^2), \tag{6.8}$$

which depends on the unknown regression operator and it is incalculable in practice. For that purpose, it is necessary to introduce the definition of the Kolmogorov's ζ–entropy and certain assumptions. Given a subset $\mathcal{S} \subset \mathcal{H}$ and $\zeta > 0$, the Kolmogorov's ζ–entropy of $\mathcal{S}$ is defined as $H_{\mathcal{S}}(\zeta) = \log N_{\zeta}(\mathcal{S})$, where $N_{\zeta}(\mathcal{S})$ is the minimal number of open balls in $\mathcal{H}$ of radius ζ such that $\mathcal{S}$ is covered, that is,

$$N_{\zeta}(\mathcal{S}) = \min\{N \in \mathbb{N} : \exists (x_1, \ldots, x_N) \in \mathcal{H} \times \overset{N}{\ldots} \times \mathcal{H} \text{ such that } \mathcal{S} \subset \bigcup_{k=1}^{N} B(x_k, \zeta)\},$$

with $B(x_k, \zeta) = \{x \in \mathcal{H} : \|x - x_k\| \leq \zeta\}$. Besides, the conditions which are required are the following:

(C.6.10) $\varphi_x(0) = 0$ and $\tau_{x,h}(t) \to \tau_{x,0}(t)$ as $h \to 0$ for all $t \in [0,1]$, for all $x \in \mathcal{C}$. Furthermore, for $h > 0$, $0 < c_1\phi(h) \leq \varphi_x(h) \leq c_2\phi(h) < +\infty$ for all $x \in \mathcal{C}$, being $c_1, c_2 > 0$ and ϕ a bijective increasing function satisfying that $\exists c_3 > 0$ and $\exists h_0 > 0$ such that $\phi'(h) < c_3$ for all $h < h_0$.

(C.6.11) For all $s \in S$, there exist $c_4, c_5 > 0$ such that $c_4\, n^{-\nu_2} < \phi(h_s) < c_5\, n^{-\nu_1}$, with $0 < \nu_1 < \nu_2 < 1$ (thus, $\lim_{n\to+\infty} n\phi(h_s) = +\infty$).

(C.6.12) There exist $c_6 > 0$ and $\beta > 0$ such that $|m(x_1) - m(x_2)| \leq c_6\|x_1 - x_2\|^{\beta}$, for all $x_1, x_2 \in \mathcal{C}$.

(C.6.13) For all $p \geq 1$, $\mathbb{E}(|Y|^p\,|X = x) \leq c_7 < +\infty$ for all $x \in \mathcal{C}$.

(C.6.14) There exists $c_8 > 0$ such that $\mathbb{E}(Y^2|X = x) = \sigma(x) \geq c_8$, with σ continuous on $\mathcal{C}$.

(C.6.15) $K(\cdot)$ is an asymmetric, bounded and Lipschitz kernel supported on $[0,1]$ with a continuous derivative on $[0,1)$ such that $K(1) > 0$ and $K'(t) < 0$ for all $t \in [0,1)$.

(C.6.16) For all $s \in S$, the sequence of bandwiths h_s verifies that $\lim_{n\to+\infty} h_s = 0$ and $h_s < \delta_D$ (with δ_D defined as in (6.5)).

(C.6.17) For n large enough, $(\log n)^2/(n\phi(h_s)) < H_{\mathcal{C}}(\log n/n) < (n\phi(h_s))/\log n$, for all $s \in S$ (consequently, $\lim_{n\to+\infty} H_{\mathcal{C}}(\log n/n)/(n\phi(h_s)) = 0$ and $\lim_{n\to+\infty} \log n/(n\phi(h_s)) = 0$). Furthermore, the Kolmogorov's entropy of $\mathcal{C}$ verifies for some $c_9 > 1$ that

$$\sum_{n=1}^{+\infty} \exp\left\{(1 - c_9)H_{\mathcal{C}}\left(\frac{\log n}{n}\right)\right\} < +\infty.$$

(C.6.18) $card(\Upsilon \times H_n^{N_\varepsilon}) = n^{\alpha}$ with $\alpha > 0$.

The first optimality result ensures that $(\upsilon_{\mathrm{CV}}, \{h_{s,\mathrm{CV}}\}_{s\in S})$ approximates the optimal choice in terms of MISE criterion, that is, $(\upsilon_{\mathrm{CV}}, \{h_{s,\mathrm{CV}}\}_{s\in S})$ approximates

$$(\upsilon^*, \{h_s^*\}_{s\in S}) = \arg \min_{(\upsilon, \{h_s\}_{s\in S}) \in \Upsilon \times H_n^{N_\varepsilon}} \mathrm{MISE}(\upsilon, \{h_s\}_{s\in S}).$$

Theorem 6.2.6. *Under* **(C.6.1)**–**(C.6.4)** *and* **(C.6.10)**–**(C.6.18)**, *it holds that*

$$\frac{MISE(\upsilon^*, \{h_s^*\}_{s\in S})}{MISE(\upsilon_{CV}, \{h_{s,CV}\}_{s\in S})} \to 1 \quad a.s.$$

The proof of Theorem 6.2.6 can be found in Section 6.6.3.

Remark 6.2.7. Ait-Saïdi et al. (2008) showed the asymptotic optimality of the cross–validation techniques in the single–functional index model. The procedure and reasonings which they proposed in their paper were mimicked in the proof of the technical lemmas which are necessary to get Theorem 6.2.6 and obtain the theoretical properties of the cross–validation method.

On the other hand, for $\upsilon \in \Upsilon$, one can define

$$(\{h_{s,\mathrm{CV}}(\upsilon)\}_{s\in S}) = \arg \min_{(\{h_s\}_{s\in S})\in H_n^{N_\mathcal{E}}} \mathrm{CV}(\upsilon, \{h_s\}_{s\in S})$$

and

$$(\{h_s^*(\upsilon)\}_{s\in S}) = \arg \min_{(\{h_s\}_{s\in S})\in H_n^{N_\mathcal{E}}} \mathrm{MISE}(\upsilon, \{h_s\}_{s\in S}).$$

Then, Theorem 6.2.8 indicates that $(\{h_s^*(\upsilon)\}_{s\in S})$ can be approximated by $(\{h_{s,\mathrm{CV}}(\upsilon)\}_{s\in S})$, whereas Theorem 6.2.9 shows that both CV and MISE criteria have similar shape.

Theorem 6.2.8. *Under hypotheses of Theorem 6.2.6, it holds that*

$$\frac{MISE(\upsilon, \{h_s^*(\upsilon)\}_{s\in S})}{MISE(\upsilon, \{h_{s,CV}(\upsilon)\}_{s\in S})} \to 1 \quad a.s.$$

for each $\upsilon \in \Upsilon$.

Theorem 6.2.9. *Under hypotheses of Theorem 6.2.6, it holds that*

$$\sup_{\upsilon\in\Upsilon} \left| \frac{CV(\upsilon, \{h_{s,CV}(\upsilon)\}_{s\in S}) - MISE(\upsilon, \{h_s^*(\upsilon)\}_{s\in S}) - \hat{\sigma}_\epsilon^2}{MISE(\upsilon, \{h_s^*(\upsilon)\}_{s\in S})} \right| \to 0 \quad a.s.$$

where $\hat{\sigma}_\epsilon^2 = n^{-1}\sum_{j=1}^n \epsilon_j^2$.

Section 6.6.9 and Section 6.6.10 compile the proofs of Theorem 6.2.8 and Theorem 6.2.9, respectively.

Thanks to the previous results, the behaviour of the MISE criterion can be deduced, which can not be obtained from a practical point of view since it depends on the unknown regression operator, by means of the analysis of the CV criterion that can be derived from the data. In addition, by selecting a grid of possible υ values, and plotting $\mathrm{CV}(\upsilon, \{h_{s,\mathrm{CV}}(\upsilon)\}_{s\in S})$, it will be obtained an almost constant curve when there is no threshold in the data, and a convex curve with minimum in υ_0 when a theshold exists for $\upsilon = \upsilon_0$. As a result, depicting the CV criterion as a function of υ yields graphical tool for analysing the existence of threshold in the data, and estimate its value when it exists. For more details, see the simulation study in the next section, where some examples of CV graphics are presented and interpreted in this way.

6.3 Simulation study

The practical behaviour of the proposed methodology is analysed in this section by means of a simulation study. The case $\mathcal{H} = L^2[0,\pi]$ with $\|\cdot\|$ denoting the standard L^2–norm given by $\|x\| =$

$(\int_0^\pi x^2(t)dt)^{1/2}$ for all $x \in L^2[0,\pi]$ was considered. For each of the three cases analysed in the simulation study, $ns = 200$ samples of size $n = 200$ drawn from

$$\begin{cases} Y_i = m_1(X_i) + \epsilon_i & \text{for } i = 1, \ldots, n/2, \\ Y_i = m_2(X_i) + \epsilon_i & \text{for } i = n/2+1, \ldots, n, \end{cases} \tag{6.9}$$

were generated, where $\epsilon_i \sim \mathcal{N}(0, \sigma_\epsilon)$ with $\sigma_\epsilon = 0.01$ for all $i = 1, \ldots, n$. The covariates X_i were simulated as

$$X_i(t) = a_i\sqrt{2/\pi}\cos(2t), \quad \forall t \in [0,\pi] \tag{6.10}$$

being $a_i \sim \mathcal{U}(u_1, u_2)$ with $u_1, u_2 \in \mathbb{R}$. The impossibility of dealing with the continuous curves in practice led to discretize the curves $\{X_i\}_{i=1}^n$ in an equidistant grid formed by $p = 100$ values in $[0,\pi]$. In order to calculate the norm of a curve, the approximation of the involved integral was done using quadrature weights of $p^{-1}\pi$ (see Section 3.5, *"How to work with discrete data?"*, in Chapter 3, page 63).

The standard nonparametric estimator $\hat{m}_h$ given by (6.2) (see page 143), and the estimator $\hat{m}^\upsilon$ based on the proposed procedure defined in (6.1) (see page 143) were calculated. In both cases, the quadratic kernel was used and the parameters were selected minimizing the corresponding CV criteria. For $\hat{m}^\upsilon$, the same setting as the one proposed for examples in Section 6.2.1 was fixed: $\Upsilon \subset \mathbb{R}$, $N_{\mathcal{E}} = 2$ and the indexed family of pairs $\{(\mathcal{E}_1^\upsilon, \mathcal{E}_2^\upsilon)\}_{\upsilon\in\Upsilon}$ defined as $\mathcal{E}_1^\upsilon = (-\infty, \upsilon]$ and $\mathcal{E}_2^\upsilon = (\upsilon, +\infty)$ for each $\upsilon \in \Upsilon$. In this situation, the next threshold functions were considered

$$\Psi_1(X_i, Y_i) = Y_i \quad \text{and} \quad \Psi_2(X_i, Y_i) = \|X_i\| = |a_i|,$$

where the first one only depends on the response value and the second one only depends on the norm of the covariate. In order to distinguish which threshold function is being considered in each moment, a new superscript is used in such a way that the estimator based on Ψ_1 was denoted by $\hat{m}^{\upsilon,1}$, whereas the estimator based on Ψ_2 by $\hat{m}^{\upsilon,2}$.

When the selected threshold function is Ψ_2, another estimator for the regression operator was also built as follows. Fixing the threshold value $\hat{\upsilon}$ estimated during the calculation of $\hat{m}^{\upsilon,2}$, it is defined $\hat{I}_s = \{i \in \{1, \ldots, n\} | \Psi_2(X_i, Y_i) \in \mathcal{E}_s^{\hat{\upsilon}}\}$ for all $s \in S$. Hence, for a new observation (X_{n+1}, Y_{n+1}), one can obtain $\hat{s} \in S$ such that $\Psi_2(X_{n+1}, Y_{n+1}) \in \mathcal{E}_{\hat{s}}^{\hat{\upsilon}}$, and one can predict the response value Y_{n+1} by $\hat{Y}_{n+1} = \hat{m}_{\hat{s}}^{\hat{\upsilon},2}(X_{n+1})$ where $\hat{m}_{\hat{s}}^{\hat{\upsilon},2}$ is defined as

$$\hat{m}_{\hat{s}}^{\hat{\upsilon},2}(X_{n+1}) = \frac{\sum_{i\in\hat{I}_{\hat{s}}} Y_i K(\tilde{h}_{\hat{s}}^{-1}\|X_i - X_{n+1}\|)}{\sum_{i\in\hat{I}_{\hat{s}}} K(\tilde{h}_{\hat{s}}^{-1}\|X_i - X_{n+1}\|)}. \tag{6.11}$$

That is, the original sample is split into two subsamples according to $\hat{\upsilon}$, and the standard nonparametric estimator is computed independently for each one. Then, being (X_{n+1}, Y_{n+1}) a new observation, it is assigned to the subsample $\hat{s}$ (with $\hat{s}$ the unique subsample such that $\Psi_2(X_{n+1}, Y_{n+1}) \in \mathcal{E}_{\hat{s}}^{\hat{\upsilon}}$), and the estimator which corresponds to this subsample is used.

For each iteration in the simulation study, a learning sample $\{(X_i, Y_i)\}_{i=1}^n$ and a testing sample $\{(X_i, Y_i)\}_{i=n+1}^{2n}$ were generated to assess the performance of all these estimators. The different estimators $\tilde{m} \in \{\hat{m}_h, \hat{m}^{\upsilon,1}, \hat{m}^{\upsilon,2}, \hat{m}_{\hat{s}}^{\hat{\upsilon},2}\}$ were obtained with the learning sample, and then the testing sample was used to calculate the mean square prediction error

$$\text{MSPE} = \frac{1}{n}\sum_{i=n+1}^{2n} (Y_i - \tilde{m}(X_i))^2. \tag{6.12}$$

MSPE was obtained for each replicate, but only summary measures will be shown. Specifically, mean, median and standard deviation of the prediction error are reported.

Remark 6.3.1. The *R* routine *funopare.kernel.cv* developed by Ferraty and Vieu (2006b) (and available in `http://www.math.univ-toulouse.fr/staph/npfda`) was used to calculate the standard kernel estimator $\hat{m}_h$. Furthermore, *funopare.kernel.cv* was adapted to calculate the estimators introduced in this chapter. On the other hand, as previously commented, the estimators were computed using a global

bandwidth for both standard estimator and the proposed estimators. However, a simulation study was also run using a k–nearest neighbours method (local bandwidths) for all estimators, obtaining similar results.

Next, the main results of the three cases considered in the simulation study are presented. Each case begins with a brief description of the dataset, and then the CV graphics and the MSPE summary are shown and analysed.

6.3.1 Case A. Threshold on Y (same curves in both subsamples)

Model. For this case, it is assumed that the functional covariates X belong to the family (6.10) being $a_i \sim \mathcal{U}(0,1)$ for all $i = 1, \ldots, n$, and the model (6.9) holds with the following regression submodels

$$\begin{cases} m_1(X_i) = \|X_i\| = a_i & \text{for } i = 1, \ldots, n/2, \\ m_2(X_i) = \|X_i\| + b = a_i + b & \text{for } i = n/2+1, \ldots, n, \end{cases}$$

where $b \in \{0, 1, 2\}$. Hence, one has that $\Psi_1(X_i, Y_i) \approx m_1(X_i) \in [0, 1]$ for $1 \leq i \leq n/2$, and $\Psi_1(X_i, Y_i) \approx m_2(X_i) \in [b, b+1]$ for $n/2 < i \leq n$. Consequently, when $b = 0$, there is no threshold in the response; when $b = 1$, one could take $\upsilon \approx 1$; and when $b = 2$, there is a *gap* in the responses and any $\upsilon \in [1, 2]$ could be detected as possible threshold. On the other hand, when Ψ_2 is considered, one should not find any threshold since the same curves are simulated for the two subsamples.

CV criteria. $\mathrm{CV}(\upsilon, \{h_{s,\mathrm{CV}}(\upsilon)\}_{s \in S})$ was computed for υ in a fixed grid of $\Upsilon = [0, b+1]$ for Ψ_1 and $\Upsilon = [0, 1]$ for Ψ_2. The obtained CV criteria for the three subcases according to b (columns) and the two threshold functions Ψ_1 and Ψ_2 (rows) are plotted in Figure 6.1 (see page 151). The black solid line is the pointwise mean of all CV criteria, and its minimum is indicated by means of a black point. As it was expected, the first column shows clearly that there is no threshold neither in the response nor in the curves. For Ψ_1, both the threshold value 1 (when $b = 1$) and the threshold interval $[1, 2]$ (when $b = 2$) are detected by the graphical tool. One has also observed that for Ψ_2 the CV curves only are almost constant in the first subcase (when $b = 0$), although in the others there is no threshold either. This fact could be due to the strong link between the regression operators (m_1 and m_2) and the threshold function (Ψ_2).

Mean square prediction error (MSPE) summary. Table 6.1 (see page 151) presents the mean, the median and the standard deviation of the mean square prediction error for the three subcases. The row named "%" correspond to the percentage of cases such that $\mathrm{MSPE}(\hat{m}_h) > \mathrm{MSPE}(\tilde{m})$ (and $\mathrm{MSPE}(\hat{m}_h) = \mathrm{MSPE}(\tilde{m})$, in brackets) for each $\tilde{m} \in \{\hat{m}^{\upsilon,1}, \hat{m}^{\upsilon,2}, \hat{m}^{\hat{\upsilon},2}_{\hat{s}}\}$. Results in Table 6.1 indicate that there is no significant improvement in terms of MSPE when the threshold estimators are used. Furthermore, the results are better for Ψ_1 than for Ψ_2, as it was expected since the simulated threshold is related to the values of the response variable Y. Looking at the column which corresponds to $\hat{m}^{\upsilon,1}$, one also notices that, when the gap in Y increases (i.e., b increases), the percentage of cases for which the error of $\hat{m}^{\upsilon,1}$ is smaller than the error of $\hat{m}_h$ also increases.

6.3.2 Case B. Threshold on Y and X (equally concentrated curves)

Model. In this case, the curves were constructed as (6.10) with $a_i \sim \mathcal{U}(0,1)$ for $1 \leq i \leq n/2$, and $a_i \sim \mathcal{U}(b, b+1)$ for $n/2 < i \leq n$, where $b \in \{0, 1/2, 1\}$. The simulated submodels in (6.9) were

$$\begin{cases} m_1(X_i) = \max_{t \in [0,\pi]} |X_i(t)| = \sqrt{2/\pi}\, a_i & \text{for } i = 1, \ldots, n/2, \\ m_2(X_i) = \|X_i\| = a_i & \text{for } i = n/2+1, \ldots, n. \end{cases}$$

Therefore, one gets that $\Psi_1(X_i, Y_i) \approx m_1(X_i) \in [0, \sqrt{2/\pi}]$ and $\Psi_2(X_i, Y_i) \in [0, 1]$ for $1 \leq i \leq n/2$, whereas $\Psi_1(X_i, Y_i) \approx m_2(X_i) \in [b, b+1]$ and $\Psi_2(X_i, Y_i) \in [b, b+1]$ for $n/2 < i \leq n$. Hence, there is no threshold on X or Y when $b = 0$, while threshold can be detected using any of the two proposed

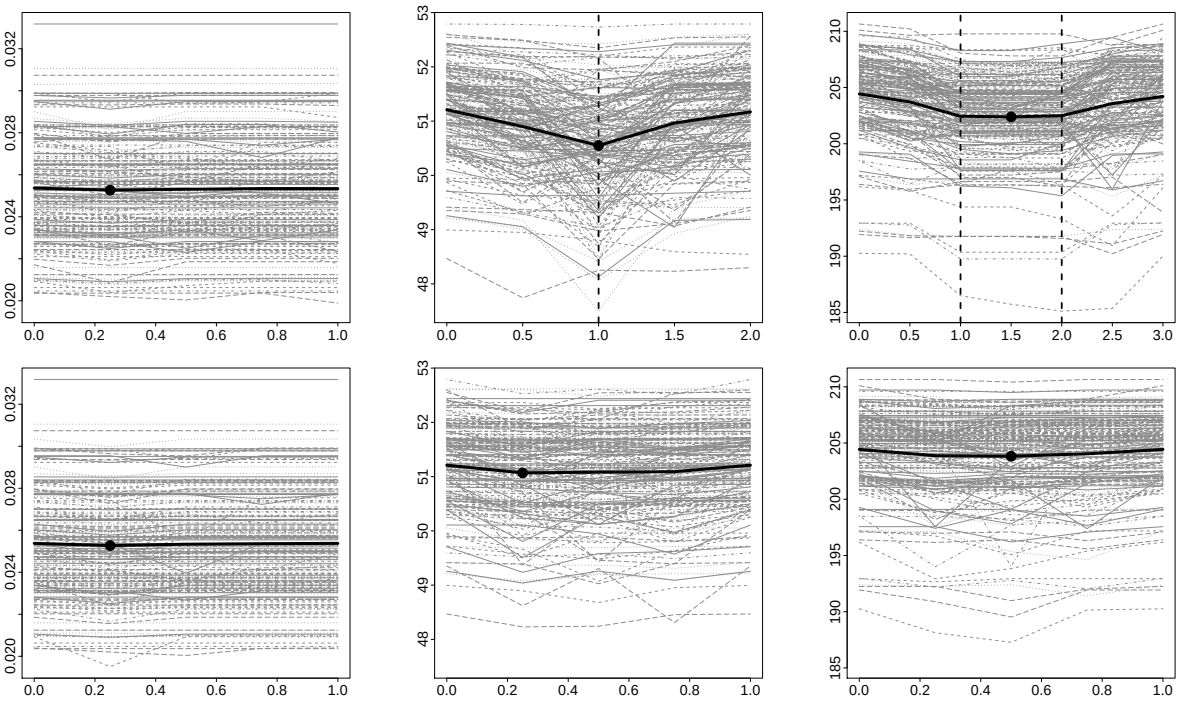

Figure 6.1: Case A. CV criteria for each replication (grey curves) and pointwise mean of all the CV criteria (black solid curve) with its minimum (black point). First row corresponds to $\Psi_1(X_i, Y_i) = Y_i$ and second one corresponds to $\Psi_2(X_i, Y_i) = \|X_i\|$. Each column corresponds to the different values of b ($b \in \{0, 1, 2\}$). Vertical dashed lines indicate the true threshold value/interval (when it exists).

b	MSPE	$\hat{m}_h$	$\hat{m}^{\upsilon,1}$	$\hat{m}^{\upsilon,2}$	$\hat{m}^{\hat{\upsilon},2}_{\hat{s}}$
0	mean	0.00013	0.00013	0.00013	0.00013
	median	0.00013	0.00013	0.00013	0.00013
	sd	0.00002	0.00002	0.00002	0.00002
	%	—	28%(20%)	26%(32%)	26%(20%)
1	mean	0.25550	0.25796	0.26087	0.26072
	median	0.25474	0.25653	0.25879	0.25860
	sd	0.00668	0.00892	0.01146	0.00980
	%	—	36%(0%)	22%(0%)	9%(11%)
2	mean	1.02278	1.02976	1.05131	1.04693
	median	1.01606	1.02133	1.03812	1.03548
	sd	0.03433	0.04287	0.05110	0.05131
	%	—	41%(0%)	15%(0%)	6%(18%)

Table 6.1: Case A. Mean, median and standard deviation of MSPE. % is the percentage of cases such that $\text{MSPE}(\hat{m}_h) > \text{MSPE}(\tilde{m})$ ($\text{MSPE}(\hat{m}_h) = \text{MSPE}(\tilde{m})$, in brackets) for $\tilde{m} \in \{\hat{m}^{\upsilon,1}, \hat{m}^{\upsilon,2}, \hat{m}^{\hat{\upsilon},2}_{\hat{s}}\}$.

threshold functions when $b = 1$ (any $\upsilon \in [\sqrt{2/\pi}, 1]$ for Ψ_1 and $\upsilon \approx 1$ for Ψ_2). It must be emphasized that the case $b = 1/2$ represents an intermediate situation, since the images of Ψ_1 and Ψ_2 for each subsample are overlapped.

CV criteria. The cross–validation criteria $\text{CV}(\upsilon, \{h_{s,\text{CV}}(\upsilon)\}_{s \in S})$, for υ in a fixed grid of $\Upsilon = [0, b{+}1]$, are plotted in Figure 6.2 (see page 152), where rows corresponds to the two threshold functions and columns to each value for b. Again, the black solid line is the pointwise mean of all CV criteria, and a black point indicates its minimum. The first column ($b = 0$) shows that there is no threshold neither

for Ψ_1 nor for Ψ_2. In the second column ($b = 1/2$), although there is no clear threshold, both Ψ_1 and Ψ_2 have overlapped images for the two subsamples. This is the reason why the CV curves seem to detect something in $[1/2, \sqrt{2/\pi}]$ for Ψ_1 and in $[1/2, 1]$ for Ψ_2. Finally, both Ψ_1 and Ψ_2 estimate correctly their respective threshold values for $b = 1$ (third column).

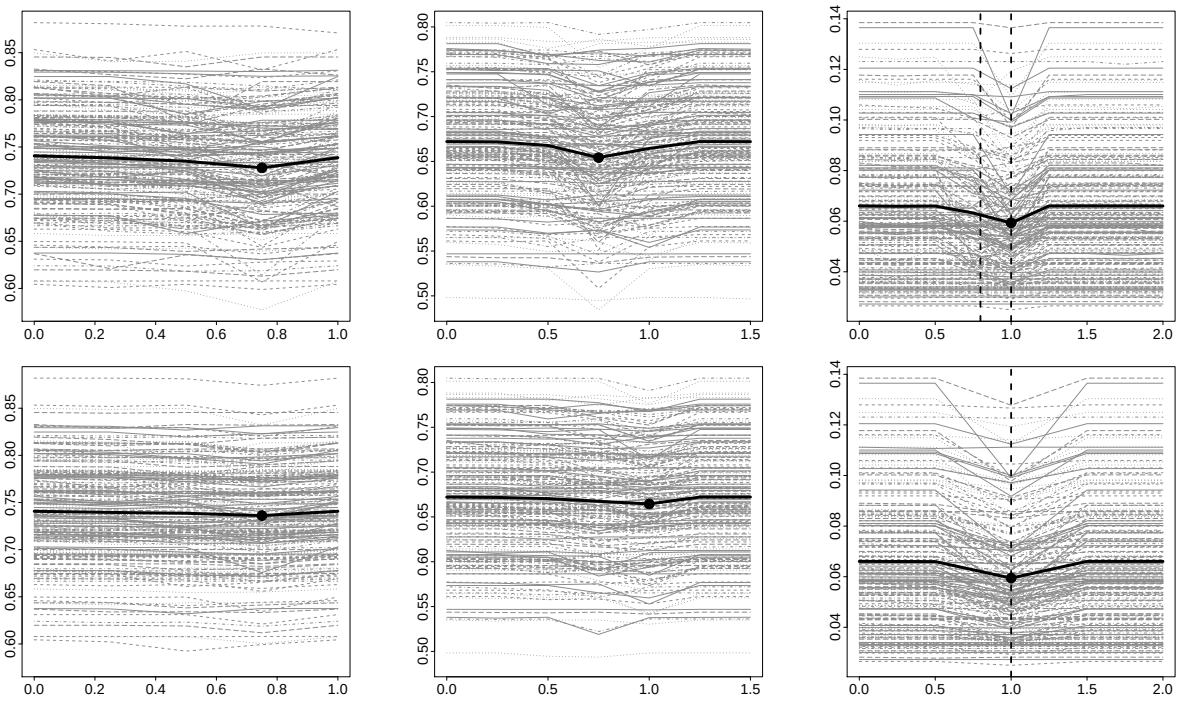

Figure 6.2: Case B. CV criteria for each replication (grey curves) and pointwise mean of all the CV criteria (black solid curve) with its minimum (black point). First row corresponds to $\Psi_1(X_i, Y_i) = Y_i$ and second one corresponds to $\Psi_2(X_i, Y_i) = \|X_i\|$. Each column corresponds to the different values of b ($b \in \{0, 1/2, 1\}$). Vertical dashed lines indicate the true threshold value/interval (when it exists).

Mean square prediction error (MSPE) summary. The mean, the median and the standard deviation of MSPE for the three subcases is shown in Table 6.2 (see page 153). Furthermore, the row "%" compiles the percentages of cases such that $\text{MSPE}(\hat{m}_h) > \text{MSPE}(\tilde{m})$ (and $\text{MSPE}(\hat{m}_h) = \text{MSPE}(\tilde{m})$, in brackets) for each computed estimator. From Table 6.2 it can be deduced that there is no improvement in terms of the mean square prediction error when there is no threshold ($b = 0$) and $\hat{m}^{\upsilon,1}$, $\hat{m}^{\upsilon,2}$ and $\hat{m}_{\hat{s}}^{\hat{\upsilon},2}$ are considered. Nevertheless, when the threshold exists ($b = 1$), $\hat{m}_{\hat{s}}^{\hat{\upsilon},2}$ produces smaller MSPE values than the standard estimator $\hat{m}_h$, so one can reduce the prediction error if the sample is split into two subsamples and one works with each one separately.

6.3.3 Case C. Threshold on Y and X (differently concentrated curves)

Model. Let X be as (6.10) being $a_i \sim \mathcal{U}(0, 1)$ for $1 \leq i \leq n/2$ and $a_i \sim \mathcal{U}(1, b + 1)$ for $n/2 < i \leq n$, with $b \in \{1, 1/2, 1/4\}$. Moreover, the same submodels as in Case B are considered in this case, that is,

$$\begin{cases} m_1(X_i) = \max_{t\in[0,\pi]} |X_i(t)| = \sqrt{2/\pi}a_i & \text{for } i = 1, \ldots, n/2, \\ m_2(X_i) = \|X_i\| = a_i & \text{for } i = n/2 + 1, \ldots, n. \end{cases}$$

Consequently, $\Psi_1(X_i, Y_i) \approx m_1(X_i) \in [0, \sqrt{2/\pi}]$ for $1 \leq i \leq n/2$, and $\Psi_1(X_i, Y_i) \approx m_2(X_i) \in [1, b + 1]$ for $n/2 < i \leq n$, so any value in the interval $[\sqrt{2/\pi}, 1]$ can be taken as threshold in this case. On

b	MSPE	$\hat{m}_h$	$\hat{m}^{\upsilon,1}$	$\hat{m}^{\upsilon,2}$	$\hat{m}_{\hat{s}}^{\hat{\upsilon},2}$
0	mean	0.00372	0.00382	0.00378	0.00378
	median	0.00368	0.00378	0.00375	0.00375
	sd	0.00030	0.00037	0.00032	0.00033
	%	—	26%(0%)	28%(0%)	28%(4%)
1/2	mean	0.00339	0.00347	0.00346	0.00338
	median	0.00340	0.00346	0.00345	0.00336
	sd	0.00033	0.00037	0.00041	0.00036
	%	—	33%(0%)	38%(0%)	56%(0%)
1	mean	0.00035	0.00034	0.00034	0.00019
	median	0.00033	0.00032	0.00033	0.00016
	sd	0.00013	0.00014	0.00014	0.00011
	%	—	64%(4%)	58%(8%)	91%(6%)

Table 6.2: Case B. Mean, median and standard deviation of MSPE. % is the percentage of cases such that $\text{MSPE}(\hat{m}_h) > \text{MSPE}(\tilde{m})$ ($\text{MSPE}(\hat{m}_h) = \text{MSPE}(\tilde{m})$, in brackets) for $\tilde{m} \in \{\hat{m}^{\upsilon,1}, \hat{m}^{\upsilon,2}, \hat{m}_{\hat{s}}^{\hat{\upsilon},2}\}$.

the other hand, $\Psi_2(X_i, Y_i) \in [0, 1]$ for all $1 \leq i \leq n/2$, and $\Psi_2(X_i, Y_i) \in [1, b+1]$ otherwise. Hence, $\upsilon \approx 1$ is the threshold value when Ψ_2 is used. Note that, in this case, b is not directly linked with the existence of threshold but with the concentration of curves which belong to the second subsample.

CV criteria. The curves $\text{CV}(\upsilon, \{h_{s,\text{CV}}(\upsilon)\}_{s\in S})$ were obtained for υ in a fixed grid of $\Upsilon = [0, b+1]$, and they are depicted in Figure 6.3. As it was expected, it seems that threshold can be detected using both Ψ_1 and Ψ_2 in all the subcases.

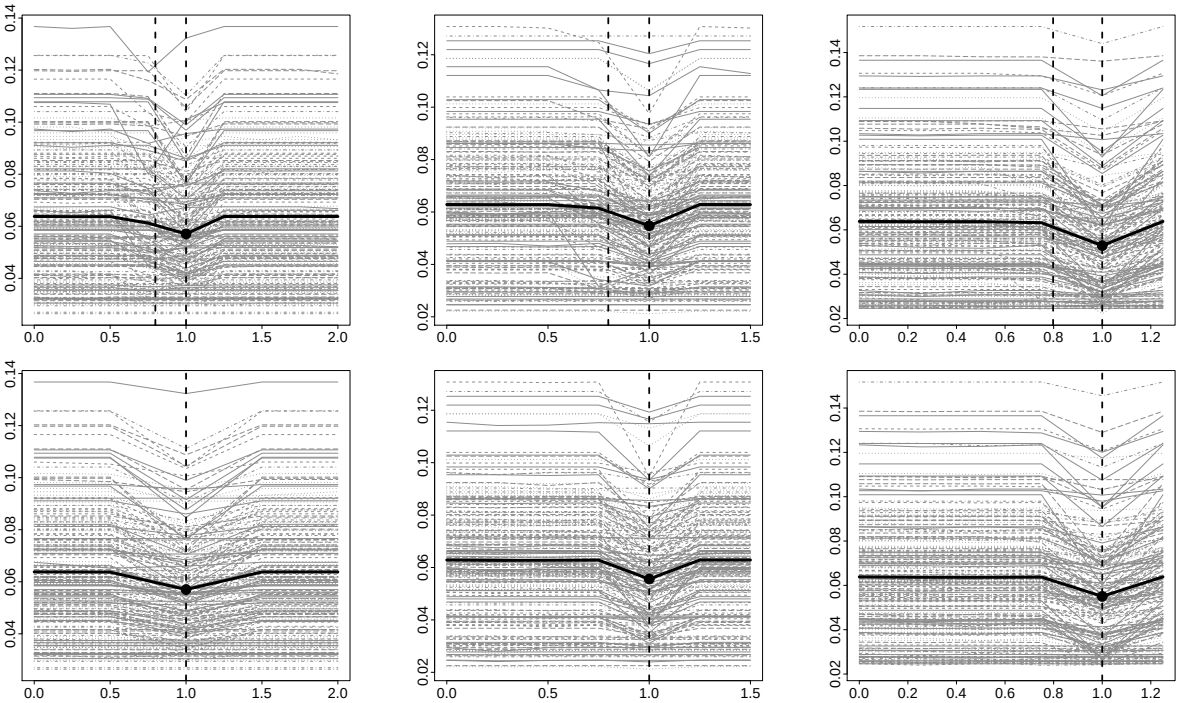

Figure 6.3: Case C. CV criteria for each replication (grey curves) and pointwise mean of all the CV criteria (black solid curve) with its minimum (black point). First row corresponds to $\Psi_1(X_i, Y_i) = Y_i$ and second one corresponds to $\Psi_2(X_i, Y_i) = \|X_i\|$. Each column corresponds to the different values of b ($b \in \{1, 1/2, 1/4\}$). Vertical dashed lines indicate the true threshold value/interval (when it exists).

Mean square prediction error (MSPE) summary. Furthermore, as it was done in the previous scenarios, Table 6.3 compiles the main results related to the MSPE, and the percentage of cases such that $\text{MSPE}(\hat{m}_h) > \text{MSPE}(\tilde{m})$ (and $\text{MSPE}(\hat{m}_h) = \text{MSPE}(\tilde{m})$, in brackets) for each $\tilde{m} \in \{\hat{m}^{\upsilon,1}, \hat{m}^{\upsilon,2}, \hat{m}_{\hat{s}}^{\hat{\upsilon},2}\}$. Table 6.3 shows that $\hat{m}^{\upsilon,1}$ and $\hat{m}^{\upsilon,2}$ give similar MSPE values to the standard estimator $\hat{m}_h$, whereas $\hat{m}_{\hat{s}}^{\hat{\upsilon},2}$ can noticeably reduce the prediction error since the two different subsamples are detected and analysed separately.

b	MSPE	$\hat{m}_h$	$\hat{m}^{\upsilon,1}$	$\hat{m}^{\upsilon,2}$	$\hat{m}_{\hat{s}}^{\hat{\upsilon},2}$
0	mean	0.00033	0.00033	0.00033	0.00019
	median	0.00032	0.00030	0.00031	0.00016
	sd	0.00012	0.00012	0.00014	0.00010
	%	—	63%(4%)	54%(11%)	92%(5%)
1/2	mean	0.00033	0.00035	0.00035	0.00019
	median	0.00030	0.00028	0.00028	0.00014
	sd	0.00018	0.00030	0.00027	0.00017
	%	—	66%(4%)	60%(4%)	92%(2%)
1	mean	0.00036	0.00037	0.00041	0.00016
	median	0.00030	0.00030	0.00030	0.00014
	sd	0.00020	0.00024	0.00042	0.00009
	%	—	55%(2%)	50%(2%)	92%(0%)

Table 6.3: Case C. Mean, median and standard deviation of MSPE. % is the percentage of cases such that $\text{MSPE}(\hat{m}_h) > \text{MSPE}(\tilde{m})$ ($\text{MSPE}(\hat{m}_h) = \text{MSPE}(\tilde{m})$, in brackets) for $\tilde{m} \in \{\hat{m}^{\upsilon,1}, \hat{m}^{\upsilon,2}, \hat{m}_{\hat{s}}^{\hat{\upsilon},2}\}$.

6.4 Real data application

In this section, the results from the application of the proposed methodology to three different datasets will be presented: Canadian weather data, spectrometric data, and atmospheric pollution data. It is well–known that sometimes the derivatives are more informative than the original curves themselves. This is the reason why the original functional curves of the dataset (X), their first derivatives (X'), or their second derivatives (X'') were taken as functional covariates in the functional regression model.

The standard nonparametric estimator $\hat{m}_h$ (see (6.2), page 143) and the proposed estimator $\hat{m}^{\upsilon}$ (see (6.1), page 143) were calculated, taking the quadratic kernel and choosing the involved parameters by means of the CV method. For $\hat{m}^{\upsilon}$, $\Upsilon \subset \mathbb{R}$, $N_{\mathcal{E}} = 2$, $\{(\mathcal{E}_1^{\upsilon}, \mathcal{E}_2^{\upsilon})\}_{\upsilon \in \Upsilon}$ defined as $\mathcal{E}_1^{\upsilon} = (-\infty, \upsilon]$ and $\mathcal{E}_2^{\upsilon} = (\upsilon, +\infty)$ for each $\upsilon \in \Upsilon$ were considered, and the following threshold functions[1]

$$\Psi_1(X_i, Y_i) = Y_i, \quad \Psi_2(X_i, Y_i) = \|X_i\|, \quad \text{and} \quad \Psi_{3k}(X_i, Y_i) = |\langle X_i, \hat{v}_k \rangle| \text{ for } k \in \{1, 2\},$$

where $\hat{v}_k$ is the empirical estimator of the kth functional principal component of X, i.e., $\hat{v}_k$ is the kth eigenfunction of $\Gamma_n = n^{-1} \sum_{i=1}^n (X_i - \overline{X}) \otimes (X_i - \overline{X})$ with $\overline{X} = n^{-1} \sum_{i=1}^n X_i$. Mantaining the notation introduced in the simulation study, a new superscript indicates which threshold function is taken in each case: $\hat{m}^{\upsilon,1}$ is based on Ψ_1, $\hat{m}^{\upsilon,2}$ is based on Ψ_2, and $\hat{m}^{\upsilon,3k}$ is based on Ψ_{3k} for $k \in \{1, 2\}$. In addition, $\hat{m}_{\hat{s}}^{\hat{\upsilon},2}$ and $\hat{m}_{\hat{s}}^{\hat{\upsilon},3k}$ for $k \in \{1, 2\}$ were also computed since the threshold functions chosen in these three cases only depend on the functional covariate X (see (6.11), page 149).

Furthermore, the dataset was randomly split into a learning sample used to build the estimators, and a testing sample which allows computing the MSPE (see (6.12), page 149). This partition procedure was iterated 200 times in order to avoid the effect of the subsampling construction.

[1] In the simulation study the threshold function $\Psi_{31}(X_i, Y_i) = |\langle X_i, \hat{v}_k \rangle|$ was also computed. Nevertheless, it can be shown that $\Psi_{31}(X_i, Y_i) \approx |a_i|$ in that scenario (recall that $\Psi_2(X_i, Y_i) = |a_i|$ too), since the curves were simulated from (6.10). Hence, the results using Ψ_{31} were similar to those which were obtained using Ψ_2, and this is the reason why only Ψ_1 and Ψ_2 were mentioned in the simulation study.

6.4.1 Canadian weather data

The first example is the Canadian weather data (see Section 1.1.2 in Chapter 1, page 2). This dataset contains daily temperature and precipitation at $n = 35$ different locations in Canada averaged over 1960 to 1994. For the weather station i, $X_i(t)$ is the daily averaged temperature in the day t, with $t \in \{1, \ldots, 365\}$, and Y_i is the logarithm of total annual precipitation in the station. The dataset was split at random into a learning sample of 25 stations and a testing sample of 10 stations.

CV criteria. Figure 6.4 (see page 156) shows the curves $\mathrm{CV}(\upsilon, \{h_{s,\mathrm{CV}}(\upsilon)\}_{s\in S})$ for each of the four threshold functions (rows), and each of the three possible covariates (columns): X, X' and X''. It seems that a threshold is clearly detected when, for instance, Ψ_1 is considered, and the covariate is the first derivative of the original temperature curves (X'). Since certain underlying structure was found, the median of the estimated thresholds ($\mathrm{median}(\{\hat{\upsilon}_l\}_{l=1}^{ns=200}) \approx 2.89$) when the first derivatives are involved in the regression model as covariate, and the threshold function is Ψ_1, was considered and the sample was split using this value into two subsamples (see Figure 6.5, page 157). It seems that the first subsample consists of the weather stations with higher differences between summer and winter temperatures, whereas the second one corresponds to weather stations with lower temperature variation over the year. On the other hand, the Canadian weather stations are classified into four climatic zones: atlantic, pacific, continental, and arctic. In Figure 6.6 (see page 157), the 35 stations were drawn by means of their coordinates (latitude–longitude), with different symbols to identify these four climatic zones and different tones (light and dark) to distinguish the two detected subsamples. As a conclusion, it can be said that the threshold procedure seems to highlight the differences between pacific–atlantic weather stations and arctic–continental ones.

Mean square prediction error (MSPE) summary. With regard to the MSPE, its mean, median and standard deviation for the different cases are compiled in Table 6.4. In general, it is observed that the threshold estimators do not reduce the prediction errors obtained by the standard nonparametric estimator $\hat{m}_h$. Moreover, the prediction error significantly decreases when the first derivatives are used in all cases.

	MSPE	$\hat{m}_h$	$\hat{m}^{\upsilon,1}$	$\hat{m}^{\upsilon,2}$	$\hat{m}^{\hat{\upsilon},2}_{\hat{s}}$	$\hat{m}^{\upsilon,31}$	$\hat{m}^{\hat{\upsilon},31}_{\hat{s}}$	$\hat{m}^{\upsilon,32}$	$\hat{m}^{\hat{\upsilon},32}_{\hat{s}}$
X	mean	0.051	0.052	0.052	0.053	0.060	0.048	0.055	0.054
	median	0.049	0.046	0.046	0.049	0.055	0.045	0.047	0.050
	sd	0.019	0.027	0.023	0.020	0.025	0.021	0.026	0.025
	%	—	55%(0%)	56%(0%)	47%(0%)	26%(0%)	66%(0%)	46%(0%)	44%(0%)
X'	mean	0.029	0.032	0.035	0.031	0.041	0.031	0.042	0.038
	median	0.026	0.029	0.033	0.028	0.035	0.028	0.032	0.026
	sd	0.016	0.021	0.020	0.015	0.037	0.016	0.040	0.032
	%	—	51%(0%)	36%(0%)	42%(0%)	25%(0%)	42%(0%)	39%(0%)	46%(0%)
X''	mean	0.041	0.044	0.045	0.045	0.050	0.049	0.050	0.056
	median	0.041	0.042	0.040	0.037	0.048	0.043	0.044	0.048
	sd	0.022	0.024	0.028	0.033	0.029	0.035	0.030	0.035
	%	—	50%(0%)	46%(0%)	52%(0%)	28%(2%)	39%(1%)	28%(1%)	22%(0%)

Table 6.4: Canadian weather data. Mean, median and standard deviation of MSPE. % is the percentage of cases such that $\mathrm{MSPE}(\hat{m}_h) > \mathrm{MSPE}(\tilde{m})$ ($\mathrm{MSPE}(\hat{m}_h) = \mathrm{MSPE}(\tilde{m})$, in brackets) for $\tilde{m} \in \{\hat{m}^{\upsilon,1}, \hat{m}^{\upsilon,2}, \hat{m}^{\hat{\upsilon},2}_{\hat{s}}, \hat{m}^{\upsilon,31}, \hat{m}^{\hat{\upsilon},31}_{\hat{s}}, \hat{m}^{\upsilon,32}, \hat{m}^{\hat{\upsilon},32}_{\hat{s}}\}$.

6.4.2 Spectrometric data

The spectrometric data is a part of a pork dataset which has been analysed from a functional point of view in the recent literature (see Section 1.1.2 in Chapter 1, page 2). The data concerns a sample of $n = 215$ pieces of finely chopped meat. For each piece, one observes a spectrometric curve which corresponds to the absorbance at 100 wavelengths, and its fat content. Hence, $X_i(t)$ is the absorbance

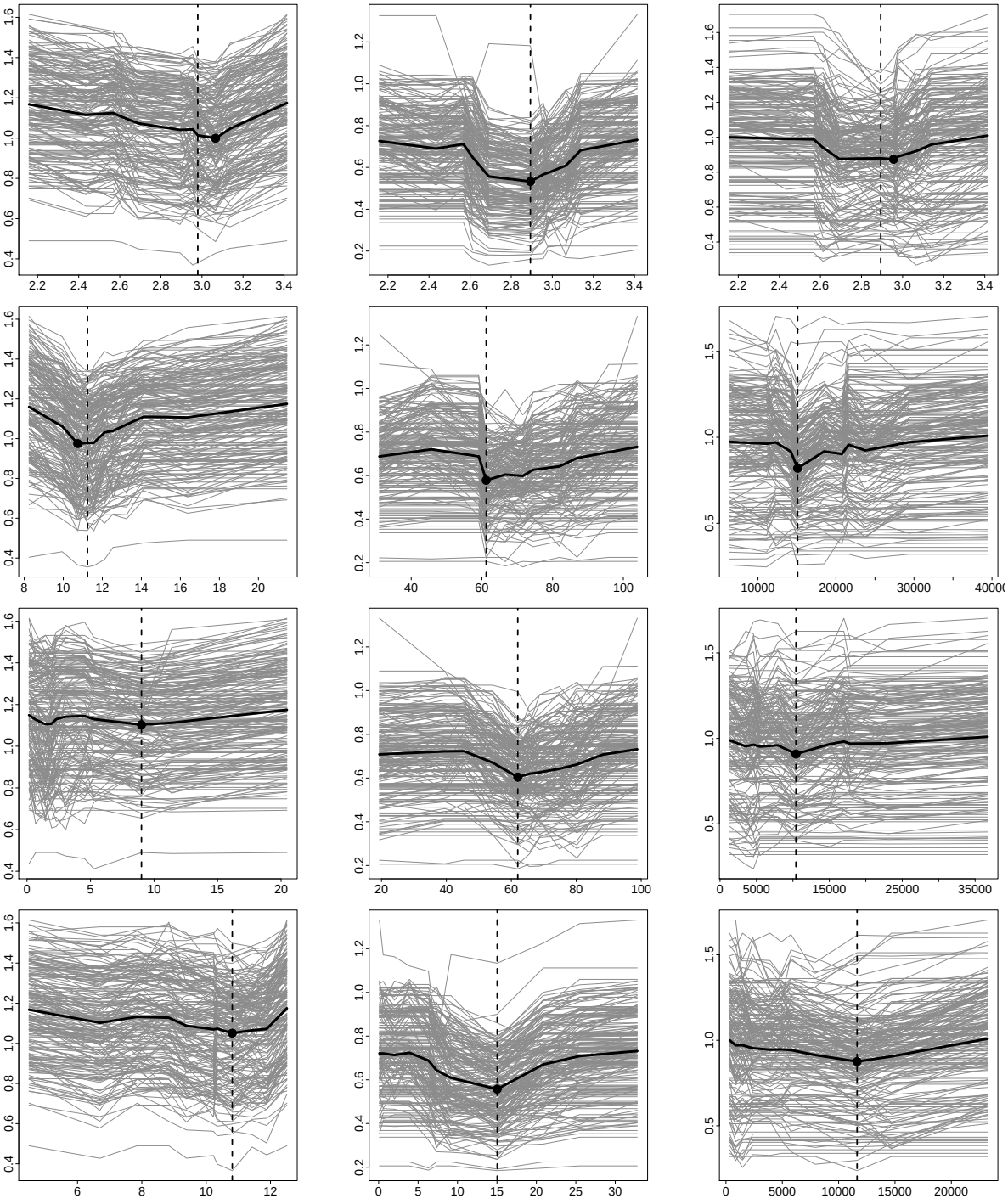

Figure 6.4: Canadian weather data. CV criteria for each replication (grey curves) and pointwise mean of all the CV criteria (black solid curve) with its minimum (black point). Rows correspond to the four threshold functions: $\Psi_1(X_i, Y_i) = Y_i$, $\Psi_2(X_i, Y_i) = \|X_i\|$, $\Psi_{3k}(X_i, Y_i) = |\langle X_i, \hat{v}_k \rangle|$ for $k \in \{1, 2\}$. Columns correspond to the three covariates: X (first column), X' (second column) and X'' (third column). Vertical dashed line corresponds to the median of the estimated threshold values.

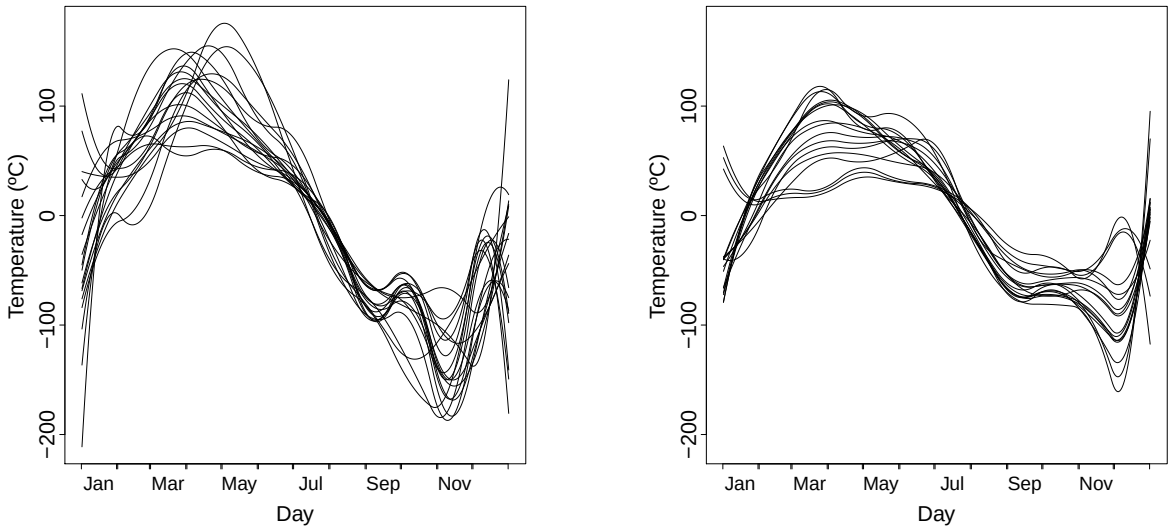

Figure 6.5: Canadian weather data. Subsamples of original temperature curves defined using $\Psi_1(X_i, Y_i) = Y_i$ as threshold function: the first subsample corresponds to curves X_i' such that $Y_i \leq 2.89$ (left panel), whereas the second one corresponds to those curves such that $Y_i > 2.89$ (right panel).

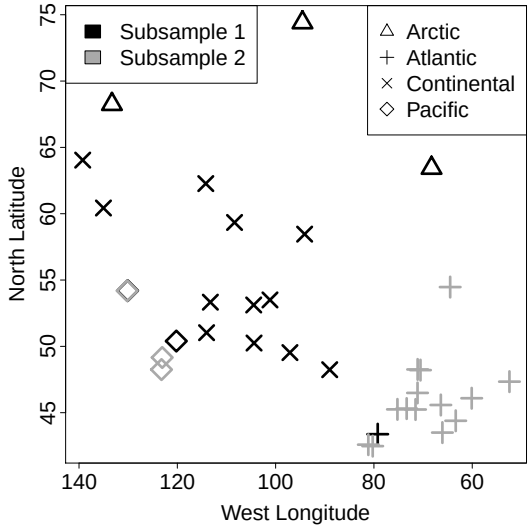

Figure 6.6: Canadian weather data. Subsamples of original temperature curves defined using $\Psi_1(X_i, Y_i) = Y_i$ as threshold function, plotted using the geografical coordinates of the weather stations where the measurements were recorded.

of the ith piece of meat at the wavelength t, where $t \in \{850, \ldots, 1050\}$, and Y_i is the corresponding fat content. In this case, learning samples of 160 pieces of meat and testing samples of 55 pieces of meat were considered.

CV criteria. The cross–validation curves $\mathrm{CV}(\upsilon, \{h_{s,\mathrm{CV}}(\upsilon)\}_{s\in S})$ which were obtained are plotted in Figure 6.7 (see page 158). In this figure, rows correspond to each threshold function, and columns indicate if one has used the original curves (first column), their first derivatives (second column) or their second derivatives (third column). It can be seen that the existence of a threshold is detected when the covariates are the second derivatives of the spectrometric curves for all the considered threshold

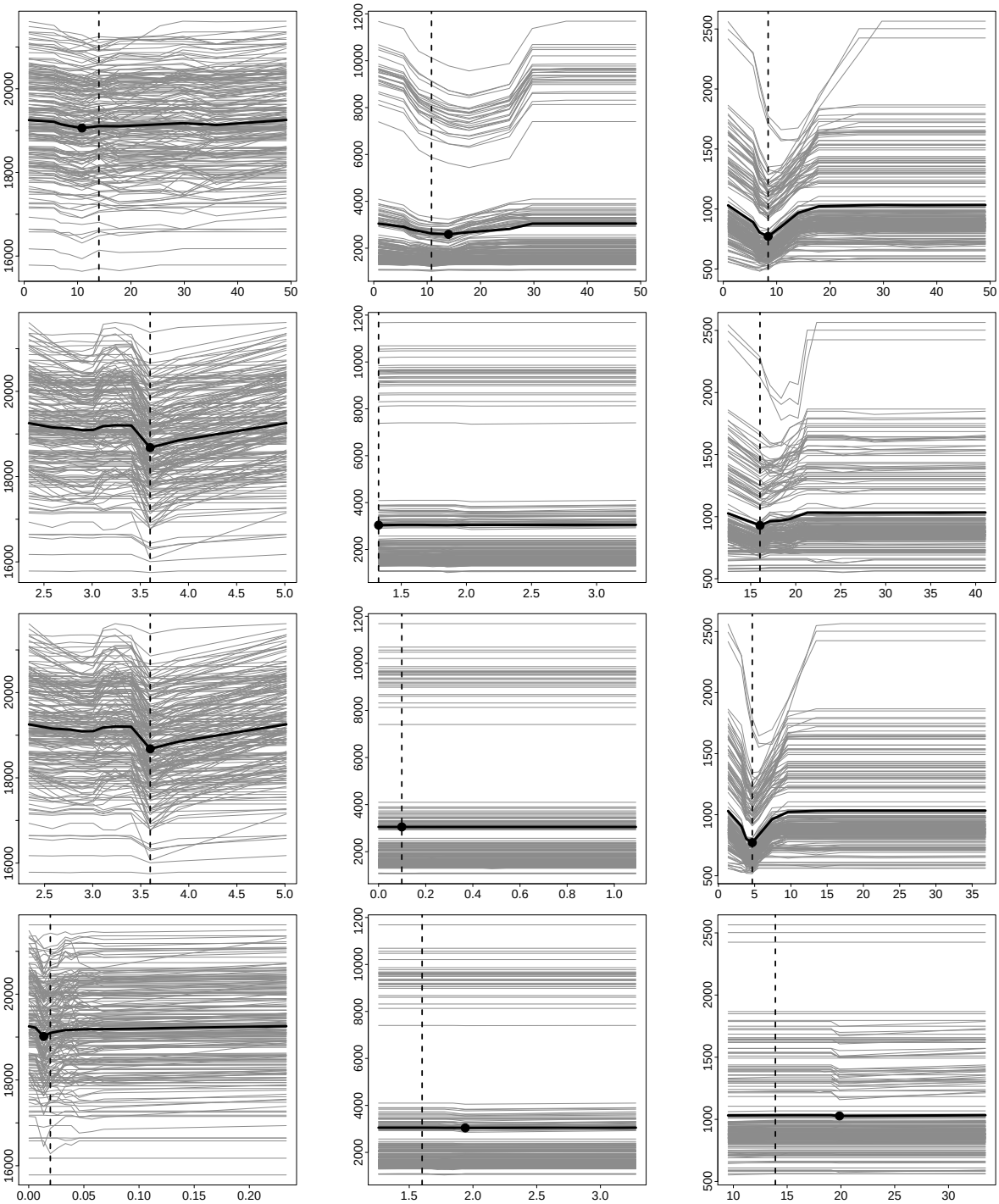

Figure 6.7: Spectrometric data. CV criteria for each replication (grey curves) and pointwise mean of all the CV criteria (black solid curve) with its minimum (black point). Rows correspond to the four threshold functions: $\Psi_1(X_i, Y_i) = Y_i$, $\Psi_2(X_i, Y_i) = \|X_i\|$, $\Psi_{3k}(X_i, Y_i) = |\langle X_i, \hat{v}_k \rangle|$ for $k \in \{1, 2\}$. Columns correspond to the three covariates: X (first column), X' (second column) and X'' (third column). Vertical dashed line corresponds to the median of the estimated threshold values.

functions, except for Ψ_{32}. Furthermore, when the original curves are taken as covariates, it seems that a threshold is also detected for Ψ_2 and Ψ_{31}. Taking into account the median of the threshold values computed over replications when, for instance, X'' and Ψ_{31} are considered (median$(\{\hat{v}_l\}_{l=1}^{ns=200}) \approx 4.67$), one can divide the data into two groups: second derivatives belonging to the first group are drawn in the left panel of Figure 6.8, whereas second derivatives belonging to the second one are drawn in the right panel. It seems that the most noticeable difference between the two groups is the presence/absence of a big valley in the wavelengths around 920–940 nm.

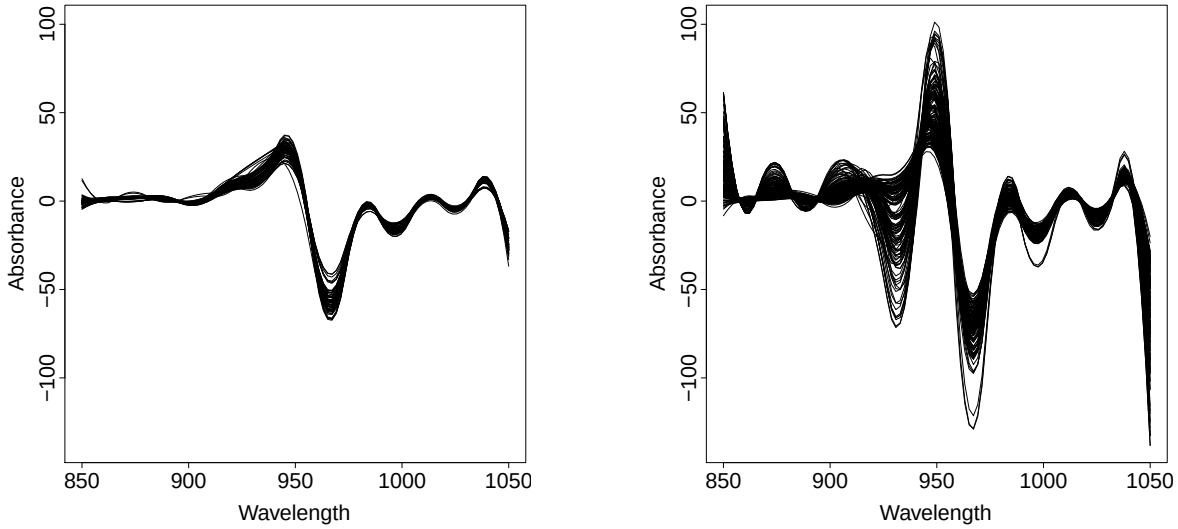

Figure 6.8: Spectrometric data. Subsamples of second derivatives of spectrometric curves defined using $\Psi_{31}(X_i'', Y_i) = |\langle X_i'', \hat{v}_1\rangle|$ as threshold function: the first subsample corresponds to curves X_i'' such that $|\langle X_i'', \hat{v}_1\rangle| \leq 4.67$ (left panel), whereas the second one corresponds to those curves such that $|\langle X_i'', \hat{v}_1\rangle| > 4.67$ (right panel).

Mean square prediction error (MSPE) summary. Table 6.5 (see page 160) compiles the mean, the median and the standard deviation of MSPE for the different threshold functions, estimators and covariates. Note that the smallest MSPE values are obtained when the second derivatives of the spectrometric curves are introduced in the regression model. This fact is not surprising because it is known in the chemometric community that the derivatives of spectra are more informative than the original spectra. On the other hand, note that the proposed theshold estimators produce smaller prediction errors than $\hat{m}_h$ in many of the analysed cases.

6.4.3 Atmospheric pollution data

Finally, an environmental example have been selected: an air pollution data (see Section 1.1.2 in Chapter 1, page 2). The data is a time series corresponding to the concentration of hourly averaged NO_x in the neighbourhood of a power station property of ENDESA, located in As Pontes in the Northwest of Spain. The NO_x level was measured each minute from 2007 to 2009. The time series has been divided in various paths corresponding to 4 hour periods (i.e., curves are discretized at 240 points). Among these paths, $n = 1,000$ of them were selected, in such a way that $X_i(t)$ is the hourly averaged NO_x measurement for the minute t during the period i, with $t \in \{1, \ldots, 240\}$, and the scalar response Y_i is the value of the hourly averaged NO_x level 30 minutes ahead. From this dataset, learning samples and testing samples of 750 and 250 observations, respectively, were obtained.

CV criteria. Figure 6.9 (see page 161) presents the CV curves defined as $\mathrm{CV}(\upsilon, \{h_{s,\mathrm{CV}}(\upsilon)\}_{s\in S})$ for the four threshold functions (rows) and the three possible inputs (columns). Examining the CV

	MSPE	$\hat{m}_h$	$\hat{m}^{\upsilon,1}$	$\hat{m}^{\upsilon,2}$	$\hat{m}^{\hat{\upsilon},2}_{\hat{s}}$	$\hat{m}^{\upsilon,31}$	$\hat{m}^{\hat{\upsilon},31}_{\hat{s}}$	$\hat{m}^{\upsilon,32}$	$\hat{m}^{\hat{\upsilon},32}_{\hat{s}}$
X	mean	123.30	124.78	122.77	85.08	122.72	85.19	123.42	128.21
	median	120.51	121.93	120.64	73.99	120.63	73.60	121.35	127.68
	sd	20.46	21.58	19.23	29.92	19.22	30.32	20.03	19.70
	%	—	43%(0%)	50%(0%)	95%(0%)	50%(0%)	95%(0%)	40%(2%)	36%(1%)
X'	mean	20.63	17.70	20.70	17.93	20.65	21.41	20.63	19.93
	median	12.61	11.26	12.71	10.99	12.61	13.17	12.76	12.94
	sd	18.23	15.47	18.17	16.72	18.21	17.63	18.17	16.34
	%	—	86%(0%)	62%(4%)	56%(3%)	26%(37%)	48%(9%)	33%(32%)	38%(8%)
X''	mean	6.87	5.17	6.18	6.06	5.18	4.26	6.94	6.12
	median	6.05	4.61	5.57	5.58	4.64	4.05	6.09	5.75
	sd	2.50	2.01	2.17	2.16	1.99	1.32	2.51	2.18
	%	—	100%(0%)	90%(0%)	82%(0%)	100%(0%)	97%(0%)	28%(20%)	49%(4%)

Table 6.5: Spectrometric data. Mean, median and standard deviation of MSPE. % is the percentage of cases such that MSPE($\hat{m}_h$) > MSPE($\tilde{m}$) (MSPE($\hat{m}_h$) = MSPE($\tilde{m}$), in brackets) for $\tilde{m} \in \{\hat{m}^{\upsilon,1}, \hat{m}^{\upsilon,2}, \hat{m}^{\hat{\upsilon},2}_{\hat{s}}, \hat{m}^{\upsilon,31}, \hat{m}^{\hat{\upsilon},31}_{\hat{s}}, \hat{m}^{\upsilon,32}, \hat{m}^{\hat{\upsilon},32}_{\hat{s}}\}$.

criteria, it seems that all the curves are almost constant when X' and X'' are considered, whereas a convex shape appears when the original curves are selected, specially for Ψ_1, Ψ_2 and Ψ_{31}. Using the median of the estimated threhold values when Ψ_2 and the original NO_x curves are considered (median($\{\hat{\upsilon}_l\}_{l=1}^{ns=200}$) ≈ 5.36), the functional dataset can be split into two groups (100 curves of each group are plotted in Figure 6.10, page 162). Hence, it seems that the threshold procedure separates curves corresponding to the lowest NO_x levels from the rest of curves in which air pollution episodes are included.

Mean square prediction error (MSPE) summary. The mean, the median and the standard deviation of MSPE are shown in Table 6.6 for the different analysed cases. Note that the covariates with the highest predictive impact are the original NO_x curves, giving the threshold estimators smaller errors than $\hat{m}_h$ in most of cases.

	MSPE	$\hat{m}_h$	$\hat{m}^{\upsilon,1}$	$\hat{m}^{\upsilon,2}$	$\hat{m}^{\hat{\upsilon},2}_{\hat{s}}$	$\hat{m}^{\upsilon,31}$	$\hat{m}^{\hat{\upsilon},31}_{\hat{s}}$	$\hat{m}^{\upsilon,32}$	$\hat{m}^{\hat{\upsilon},32}_{\hat{s}}$
X	mean	26.88	25.52	25.52	22.79	25.59	23.42	26.38	25.41
	median	25.87	24.42	24.46	21.73	24.61	22.66	25.20	24.50
	sd	6.36	6.25	6.23	5.74	6.23	5.78	6.33	6.00
	%	—	100%(0%)	98%(0%)	100%(0%)	98%(0%)	98%(0%)	100%(0%)	94%(0%)
X'	mean	80.25	80.30	80.27	64.97	80.26	66.29	80.27	65.36
	median	80.07	80.08	79.95	64.65	80.13	66.17	79.96	65.08
	sd	7.25	7.25	7.26	7.39	7.27	6.36	7.26	5.99
	%	—	57%(0%)	66%(0%)	100%(0%)	68%(0%)	100%(0%)	62%(0%)	100%(0%)
X''	mean	80.34	80.34	80.34	73.74	80.34	75.57	80.35	73.80
	median	79.60	79.63	79.61	73.92	79.61	75.38	79.61	73.95
	sd	7.09	7.09	7.09	8.03	7.09	8.06	7.09	8.04
	%	—	48%(0%)	52%(2%)	93%(0%)	50%(1%)	94%(0%)	44%(0%)	94%(0%)

Table 6.6: Atmospheric pollution data. Mean, median and standard deviation of MSPE. % is the percentage of cases such that MSPE($\hat{m}_h$) > MSPE($\tilde{m}$) (MSPE($\hat{m}_h$) = MSPE($\tilde{m}$), in brackets) for $\tilde{m} \in \{\hat{m}^{\upsilon,1}, \hat{m}^{\upsilon,2}, \hat{m}^{\hat{\upsilon},2}_{\hat{s}}, \hat{m}^{\upsilon,31}, \hat{m}^{\hat{\upsilon},31}_{\hat{s}}, \hat{m}^{\upsilon,32}, \hat{m}^{\hat{\upsilon},32}_{\hat{s}}\}$.

6.5 Final conclusions

In this chapter, a graphical tool that detects threshold structures in the context of the nonparametric functional model with scalar response has been presented. This methodology allows to detect hidden patterns related to both the functional covariate X and the scalar response Y. For that purpose, an

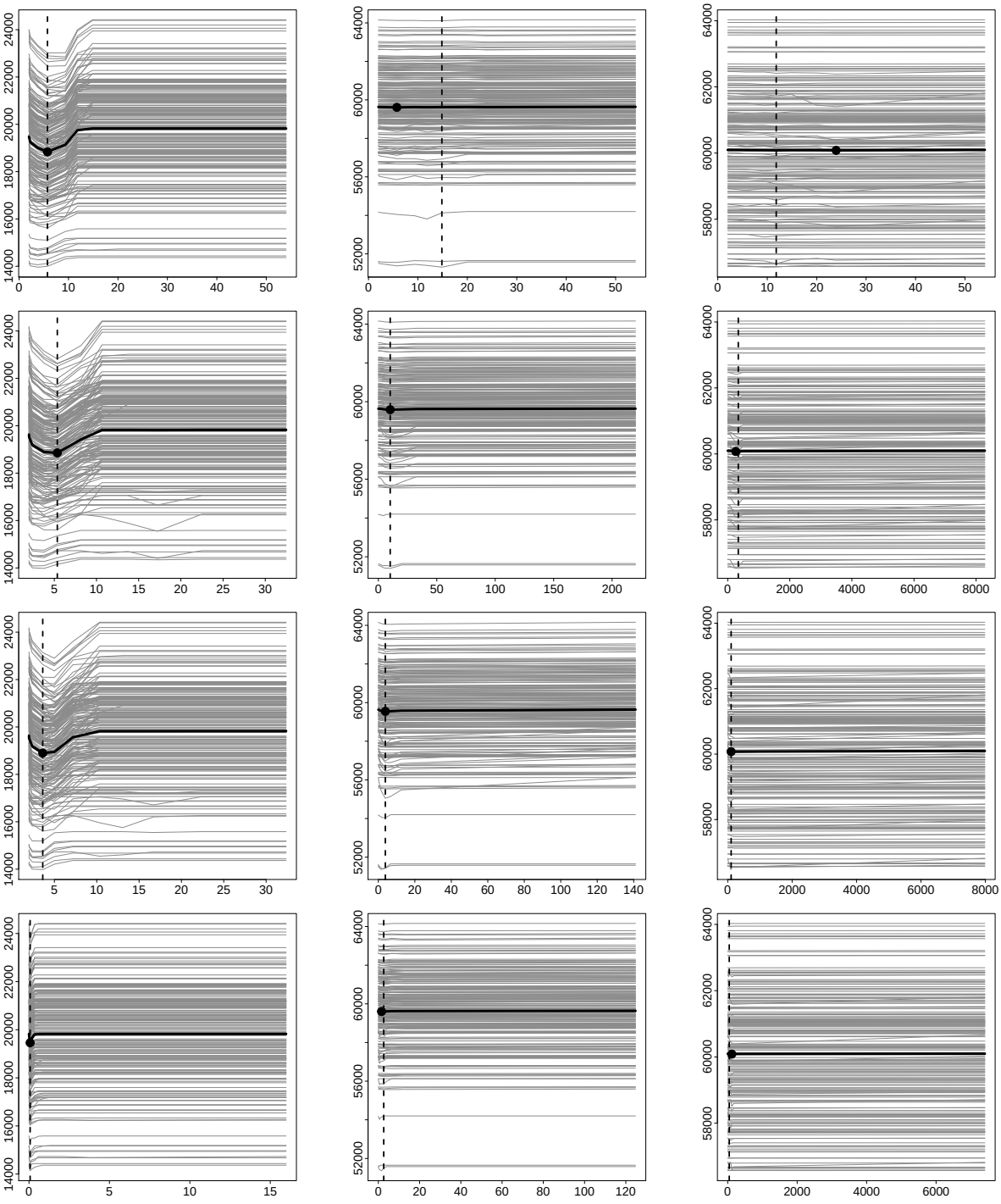

Figure 6.9: Atmospheric pollution data. CV criteria for each replication (grey curves) and pointwise mean of all the CV criteria (black solid curve) with its minimum (black point). Rows correspond to the four threshold functions: $\Psi_1(X_i, Y_i) = Y_i$, $\Psi_2(X_i, Y_i) = \|X_i\|$, $\Psi_{3k}(X_i, Y_i) = |\langle X_i, \hat{v}_k \rangle|$ for $k \in \{1, 2\}$. Columns correspond to the three covariates: X (first column), X' (second column) and X'' (third column). Vertical dashed line corresponds to the median of the estimated threshold values.

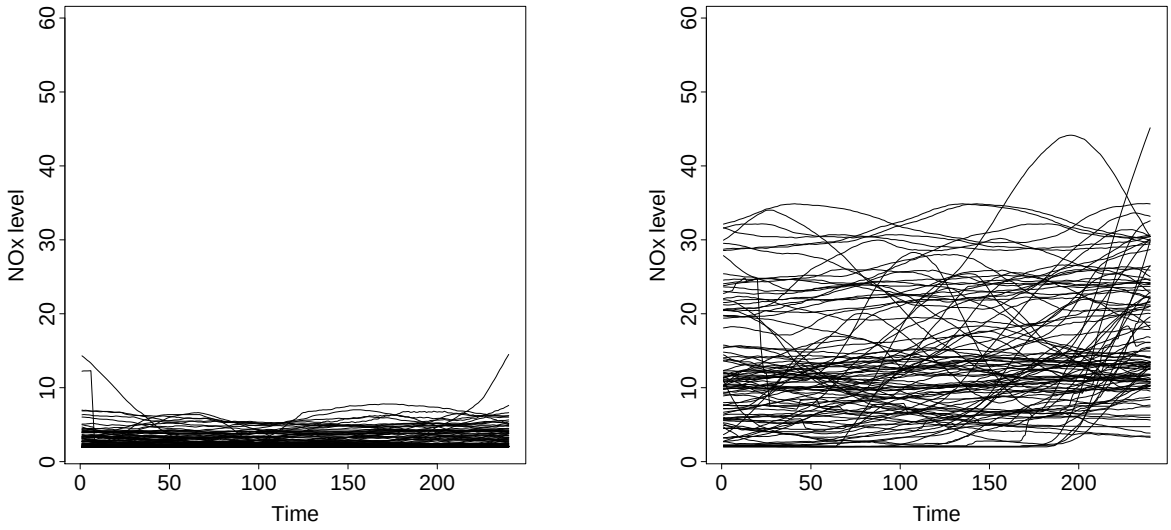

Figure 6.10: Atmospheric pollution data. Subsamples of original NO_x curves defined using $\Psi_2(X_i, Y_i) = \|X_i\|$ as threshold function: the first subsample corresponds to curves X_i such that $\|X_i\| \leq 5.36$ (100 of them are plotted in left panel), whereas the second one corresponds to those curves such that $\|X_i\| > 5.36$ (100 of them are plotted in right panel).

adequate threshold function must be chosen by the user according to the structure one wants to detect. Nevertheless, how to select this threshold function in a data–driven and efficient way is still an open question that will require further research.

When threshold exists, it have been proposed to estimate its value by means of a cross–validation method. Its optimality with respect to MISE criterium was derived in a particular scenario, so further research is necessary in order to extend this type of theoretical results to other situations, for instance, threshold functions which depends on the response Y. On the other hand, the graphical representation of the cross–validation criterion is the basis of a useful graphical tool.

In terms of the mean square prediction error, both the proposed estimators and the standard nonparametric estimator obtain similar results. However it can be observed in the simulation study that, in some cases, the mean square prediction error can be reduced if each subsample detected by the threshold technique is studied separately. Furthermore, in view of the data applications, one can conclude that the proposed methodology allows to detect some kind of hidden structures which are present in the data, although the effectiveness of the procedure depends on the choice of the threshold function Ψ. In the Canadian weather example, a geographical pattern appeared when one looked for threshold in the response. For the spectrometric data, good predictive results were generated when the sample was split according to the behaviour of the second derivatives in certain influential wavelength interval. Finally, a weak structure came up in the atmospheric pollution dataset (only if the original curves were considered), which seemed to be linked with the NO_x concentration level of each curve.

6.6 Appendix Chapter 6

This section is an appendix devoted to compile the proofs of the main results introduced in the chapter. Furthermore, the technical lemmas used to develope these proofs are also included. In order to simplify the presentation, the proofs of theorems and the main lemmas required to show them have been compiled at first, whereas the remainder auxiliary lemmas have been gathered in Section 6.6.11.

In order to simplify the notation, in the following C will denote a generic positive constant which may take on different values even in the same formula.

6.6.1 Proof of Theorem 6.2.3

As commented in (6.6) (see page 145), $\hat{m}_s^{\upsilon}(x) = \hat{m}_{h_s}(x)\mathbb{I}_{\{\tilde{\Psi}(x)\in\mathcal{E}_s^{\upsilon}\}}$, where $\hat{m}_{h_s}$ is the standard kernel estimator (6.2) (see page 143). The required assumptions in this theorem guarantee that Theorem 1 by Ferraty et al. (2007a) (see Theorem 2.4.13 in Chapter 2, page 46) can be applied, which stated the following asymptotics for the standard kernel estimator

$$\mathbb{E}(\hat{m}_{h_s}(x)) = m(x) + \psi_x'(0)\frac{M_{x,0}}{M_{x,1}}h_s + O\left(\frac{1}{n\varphi_x(h_s)}\right) + o(h_s), \tag{6.13}$$

and

$$\mathrm{Var}(\hat{m}_{h_s}(x)) = \sigma_\epsilon^2(x)\frac{M_{x,2}}{M_{x,1}^2}\frac{1}{n\varphi_x(h_s)} + o\left(\frac{1}{n\varphi_x(h_s)}\right), \tag{6.14}$$

for all $s \in S$, with $M_{x,0} = K(1) - \int_0^1 (tK(t))'\tau_{x,0}(t)dt$, $M_{x,1} = K(1) - \int_0^1 K'(t)\tau_{x,0}(t)dt$ and $M_{x,2} = K^2(1) - \int_0^1 (K^2)'(t)\tau_{x,0}(t)dt$. Therefore, the expectation of $\hat{m}_s^{\upsilon}(x)$ comes from (6.13) as follows

$$\begin{aligned}\mathbb{E}(\hat{m}_s^{\upsilon}(x)) &= \mathbb{E}\left(\hat{m}_{h_s}(x)\mathbb{I}_{\{\tilde{\Psi}(x)\in\mathcal{E}_s^{\upsilon}\}}\right) = \mathbb{E}(\hat{m}_{h_s}(x))\mathbb{I}_{\{\tilde{\Psi}(x)\in\mathcal{E}_s^{\upsilon}\}}\\ &= \left(m(x) + \psi_x'(0)\frac{M_{x,0}}{M_{x,1}}h_s + O\left(\frac{1}{n\varphi_x(h_s)}\right) + o(h_s)\right)\mathbb{I}_{\{\tilde{\Psi}(x)\in\mathcal{E}_s^{\upsilon}\}},\end{aligned}$$

whereas the expression for variance is obtained using (6.14)

$$\begin{aligned}\mathrm{Var}(\hat{m}_s^{\upsilon}(x)) &= \mathrm{Var}(\hat{m}_{h_s}(x)\mathbb{I}_{\{\tilde{\Psi}(x)\in\mathcal{E}_s^{\upsilon}\}}) = \mathrm{Var}(\hat{m}_{h_s}(x))\left(\mathbb{I}_{\{\tilde{\Psi}(x)\in\mathcal{E}_s^{\upsilon}\}}\right)^2\\ &= \left(\sigma_\epsilon^2(x)\frac{M_{x,2}}{M_{x,1}^2}\frac{1}{n\varphi_x(h_s)} + o\left(\frac{1}{n\varphi_x(h_s)}\right)\right)\mathbb{I}_{\{\tilde{\Psi}(x)\in\mathcal{E}_s^{\upsilon}\}}.\end{aligned}$$

Furthermore, for all $s_1, s_2 \in S$ such that $s_1 \neq s_2$, one has

$$\begin{aligned}&\mathrm{Cov}(\hat{m}_{s_1}^{\upsilon}(x), \hat{m}_{s_2}^{\upsilon}(x)) = \mathrm{Cov}\left(\hat{m}_{h_{s_1}}(x)\mathbb{I}_{\{\tilde{\Psi}(x)\in\mathcal{E}_{s_1}^{\upsilon}\}}, \hat{m}_{h_{s_2}}(x)\mathbb{I}_{\{\tilde{\Psi}(x)\in\mathcal{E}_{s_2}^{\upsilon}\}}\right)\\ &= \mathrm{Cov}(\hat{m}_{h_{s_1}}(x), \hat{m}_{h_{s_2}}(x))\mathbb{I}_{\{\tilde{\Psi}(x)\in\mathcal{E}_{s_1}^{\upsilon}\}}\mathbb{I}_{\{\tilde{\Psi}(x)\in\mathcal{E}_{s_2}^{\upsilon}\}} = \mathrm{Cov}(\hat{m}_{h_{s_1}}(x), \hat{m}_{h_{s_2}}(x))\mathbb{I}_{\{\tilde{\Psi}(x)\in\mathcal{E}_{s_1}^{\upsilon}\cap\mathcal{E}_{s_2}^{\upsilon}\}} = 0,\end{aligned}$$

since $\mathcal{E}_{s_1}^{\upsilon} \cap \mathcal{E}_{s_2}^{\upsilon} = \emptyset$ for all $s_1 \neq s_2$ due to **(C.6.3)**.

6.6.2 Proof of Corollary 6.2.4

Given that $\hat{m}^{\upsilon}(x) = \sum_{s\in S}\hat{m}_s^{\upsilon}(x)$, one has that $\mathbb{E}(\hat{m}^{\upsilon}(x)) = \sum_{s\in S}\mathbb{E}(\hat{m}_s^{\upsilon}(x))$ and

$$\mathrm{Var}(\hat{m}^{\upsilon}(x)) = \sum_{s\in S}\mathrm{Var}(\hat{m}_s^{\upsilon}(x)) + \sum_{s_1\in S}\sum_{s_1\neq s_2}\mathrm{Cov}(\hat{m}_{s_1}^{\upsilon}(x), \hat{m}_{s_2}^{\upsilon}(x)).$$

Hence, applying Theorem 6.2.3 (see page 145), for the expectation of the threshold estimator one has that

$$\begin{aligned}\mathbb{E}(\hat{m}^{\upsilon}(x)) &= \sum_{s\in S}\mathbb{E}(\hat{m}_s^{\upsilon}(x)) = \sum_{s\in S}\left(m(x) + \psi_x'(0)\frac{M_{x,0}}{M_{x,1}}h_s + O\left(\frac{1}{n\varphi_x(h_s)}\right) + o(h_s)\right)\mathbb{I}_{\{\tilde{\Psi}(x)\in\mathcal{E}_s^{\upsilon}\}}\\ &= m(x) + \psi_x'(0)\frac{M_{x,0}}{M_{x,1}}\sum_{s\in S}h_s\mathbb{I}_{\{\tilde{\Psi}(x)\in\mathcal{E}_s^{\upsilon}\}} + \sum_{s\in S}\left(O\left(\frac{1}{n\varphi_x(h_s)}\right) + o(h_s)\right)\mathbb{I}_{\{\tilde{\Psi}(x)\in\mathcal{E}_s^{\upsilon}\}},\end{aligned}$$

and for its variance one gets

$$\begin{aligned}\mathrm{Var}(\hat{m}^{\upsilon}(x)) &= \sum_{s\in S}\mathrm{Var}(\hat{m}_s^{\upsilon}(x)) + \sum_{s_1\in S}\sum_{s_1\neq s_2}\mathrm{Cov}(\hat{m}_{s_1}^{\upsilon}(x),\hat{m}_{s_2}^{\upsilon}(x))\\ &= \sum_{s\in S}\left(\sigma_\epsilon^2(x)\frac{M_{x,2}}{M_{x,1}^2}\frac{1}{n\varphi_x(h_s)} + o\left(\frac{1}{n\varphi_x(h_s)}\right)\right)\mathbb{I}_{\{\tilde{\Psi}(x)\in\mathcal{E}_s^{\upsilon}\}}\\ &= \sigma_\epsilon^2(x)\frac{M_{x,2}}{M_{x,1}^2}\sum_{s\in S}\frac{1}{n\varphi_x(h_s)}\mathbb{I}_{\{\tilde{\Psi}(x)\in\mathcal{E}_s^{\upsilon}\}} + \sum_{s\in S}\left(o\left(\frac{1}{n\varphi_x(h_s)}\right)\right)\mathbb{I}_{\{\tilde{\Psi}(x)\in\mathcal{E}_s^{\upsilon}\}}.\end{aligned}$$

Furthermore, if s_x denotes the only $s_x \in S$ such that $\tilde{\Psi}(x) \in \mathcal{E}_{s_x}^{\upsilon}$, then

$$\mathbb{E}(\hat{m}^{\upsilon}(x)) = m(x) + \psi_x'(0)\frac{M_{x,0}}{M_{x,1}}h_{s_x} + O\left(\frac{1}{n\varphi_x(h_{s_x})}\right) + o(h_{s_x}),$$

and

$$\mathrm{Var}(\hat{m}^{\upsilon}(x)) = \sigma_\epsilon^2(x)\frac{M_{x,2}}{M_{x,1}^2}\frac{1}{n\varphi_x(h_{s_x})} + o\left(\frac{1}{n\varphi_x(h_{s_x})}\right).$$

6.6.3 Proof of Theorem 6.2.6

The theorem will be proven by showing that

$$\left|\frac{\mathrm{MISE}(\upsilon_{\mathrm{CV}},\{h_{s,\mathrm{CV}}\}_{s\in S}) - \mathrm{MISE}(\upsilon^*,\{h_s^*\}_{s\in S})}{\mathrm{MISE}(\upsilon_{\mathrm{CV}},\{h_{s,\mathrm{CV}}\}_{s\in S})}\right| \to 0 \quad a.s.$$

Let $\hat{\sigma}_\epsilon^2$ be defined as $\hat{\sigma}_\epsilon^2 = n^{-1}\sum_{j=1}^n \epsilon_j^2$, and note that $\mathrm{MISE}(\upsilon_{\mathrm{CV}},\{h_{s,\mathrm{CV}}\}_{s\in S}) \geq \mathrm{MISE}(\upsilon^*,\{h_s^*\}_{s\in S})$ and $\mathrm{CV}(\upsilon^*,\{h_s^*\}_{s\in S}) \geq \mathrm{CV}(\upsilon_{\mathrm{CV}},\{h_{s,\mathrm{CV}}\}_{s\in S})$. Thus, one has that

$$\begin{aligned}|\mathrm{MISE}(\upsilon_{\mathrm{CV}},\{h_{s,\mathrm{CV}}\}_{s\in S}) - \mathrm{MISE}(\upsilon^*,\{h_s^*\}_{s\in S})| &\leq |-\mathrm{CV}(\upsilon_{\mathrm{CV}},\{h_{s,\mathrm{CV}}\}_{s\in S})\\ &+ \mathrm{MISE}(\upsilon_{\mathrm{CV}},\{h_{s,\mathrm{CV}}\}_{s\in S}) + \hat{\sigma}_\epsilon^2 + \mathrm{CV}(\upsilon^*,\{h_s^*\}_{s\in S}) - \mathrm{MISE}(\upsilon^*,\{h_s^*\}_{s\in S}) - \hat{\sigma}_\epsilon^2|.\end{aligned}$$

As a result, one gets

$$\begin{aligned}&\left|\frac{\mathrm{MISE}(\upsilon_{\mathrm{CV}},\{h_{s,\mathrm{CV}}\}_{s\in S}) - \mathrm{MISE}(\upsilon^*,\{h_s^*\}_{s\in S})}{\mathrm{MISE}(\upsilon_{\mathrm{CV}},\{h_{s,\mathrm{CV}}\}_{s\in S})}\right|\\ &\leq \left|\frac{\mathrm{CV}(\upsilon_{\mathrm{CV}},\{h_{s,\mathrm{CV}}\}_{s\in S}) - \mathrm{MISE}(\upsilon_{\mathrm{CV}},\{h_{s,\mathrm{CV}}\}_{s\in S}) - \hat{\sigma}_\epsilon^2}{\mathrm{MISE}(\upsilon_{\mathrm{CV}},\{h_{s,\mathrm{CV}}\}_{s\in S})}\right|\\ &\quad + \left|\frac{\mathrm{CV}(\upsilon^*,\{h_s^*\}_{s\in S}) - \mathrm{MISE}(\upsilon^*,\{h_s^*\}_{s\in S}) - \hat{\sigma}_\epsilon^2}{\mathrm{MISE}(\upsilon^*,\{h_s^*\}_{s\in S})}\right|\left|\frac{\mathrm{MISE}(\upsilon^*,\{h_s^*\}_{s\in S})}{\mathrm{MISE}(\upsilon_{\mathrm{CV}},\{h_{s,\mathrm{CV}}\}_{s\in S})}\right|\\ &\leq 2\sup_{(\upsilon,\{h_s\}_{s\in S})\in\Upsilon\times H_n^{N_\mathcal{E}}}\left|\frac{\mathrm{CV}(\upsilon,\{h_s\}_{s\in S}) - \mathrm{MISE}(\upsilon,\{h_s\}_{s\in S}) - \hat{\sigma}_\epsilon^2}{\mathrm{MISE}(\upsilon,\{h_s\}_{s\in S})}\right|,\end{aligned}$$

where the last inequality is true since $\mathrm{MISE}(\upsilon^*,\{h_s^*\}_{s\in S}) \leq \mathrm{MISE}(\upsilon_{\mathrm{CV}},\{h_{s,\mathrm{CV}}\}_{s\in S})$. Hence, the convergence is deduced from Lemma 6.6.1.

6.6.4 Formulation and proof of Lemma 6.6.1

The main aim of the next lemma is to allow to show the optimality of the CV procedure (see (6.7), page 146) with respect to the MISE criterion (see (6.8), page 147). For proving this lemma, one needs to introduce some other quadratic distances such as the average squared error

$$\mathrm{ASE}(\upsilon,\{h_s\}_{s\in S}) = \frac{1}{n}\sum_{j=1}^n (m(X_j) - \hat{m}^{\upsilon}(X_j))^2 \tag{6.15}$$

and the following two terms

$$\widetilde{\mathrm{ASE}}(\upsilon, \{h_s\}_{s\in S}) = \frac{1}{n}\sum_{j=1}^{n}(m(X_j) - \hat{m}^{\upsilon,(-j)}(X_j))^2, \quad \text{and} \tag{6.16}$$

$$\mathrm{CT}(\upsilon, \{h_s\}_{s\in S}) = \frac{1}{n}\sum_{j=1}^{n}\epsilon_j(\hat{m}^{\upsilon,(-j)}(X_j) - m(X_j)). \tag{6.17}$$

Lemma 6.6.1. *Under hypotheses of Theorem 6.2.6,*

$$\sup_{(\upsilon,\{h_s\}_{s\in S})\in\Upsilon\times H_n^{N_\mathcal{E}}}\left|\frac{CV(\upsilon,\{h_s\}_{s\in S}) - MISE(\upsilon,\{h_s\}_{s\in S}) - \hat{\sigma}_\epsilon^2}{MISE(\upsilon,\{h_s\}_{s\in S})}\right| \to 0 \quad a.s.$$

where $\hat{\sigma}_\epsilon^2 = n^{-1}\sum_{j=1}^{n}\epsilon_j^2$.

Proof. First of all, note that the CV criterion (see (6.7), page 146) can be expressed as

$$\begin{aligned}
&\mathrm{CV}(\upsilon, \{h_s\}_{s\in S}) = \frac{1}{n}\sum_{j=1}^{n}(Y_j - \hat{m}^{\upsilon,(-j)}(X_j))^2 = \frac{1}{n}\sum_{j=1}^{n}((m(X_j) - \hat{m}^{\upsilon,(-j)}(X_j)) + \epsilon_j)^2 \\
&= \frac{1}{n}\sum_{j=1}^{n}(m(X_j) - \hat{m}^{\upsilon,(-j)}(X_j))^2 + \frac{2}{n}\sum_{j=1}^{n}(m(X_j) - \hat{m}^{\upsilon,(-j)}(X_j))\epsilon_j + \frac{1}{n}\sum_{j=1}^{n}\epsilon_j^2 \\
&= \widetilde{\mathrm{ASE}}(\upsilon, \{h_s\}_{s\in S}) - 2\mathrm{CT}(\upsilon, \{h_s\}_{s\in S}) + \hat{\sigma}_\epsilon^2 \\
&= \widetilde{\mathrm{ASE}}(\upsilon, \{h_s\}_{s\in S}) - \mathrm{ASE}(\upsilon, \{h_s\}_{s\in S}) + \mathrm{ASE}(\upsilon, \{h_s\}_{s\in S}) - 2\mathrm{CT}(\upsilon, \{h_s\}_{s\in S}) + \hat{\sigma}_\epsilon^2.
\end{aligned}$$

Therefore, one has

$$\begin{aligned}
|\mathrm{CV}(\upsilon, \{h_s\}_{s\in S}) - \mathrm{MISE}(\upsilon, \{h_s\}_{s\in S}) - \hat{\sigma}_\epsilon^2| &\le |\widetilde{\mathrm{ASE}}(\upsilon, \{h_s\}_{s\in S}) - \mathrm{ASE}(\upsilon, \{h_s\}_{s\in S})| \\
&+ |\mathrm{ASE}(\upsilon, \{h_s\}_{s\in S}) - \mathrm{MISE}(\upsilon, \{h_s\}_{s\in S})| + 2|\mathrm{CT}(\upsilon, \{h_s\}_{s\in S})|.
\end{aligned}$$

Taking into account this fact, one gets

$$\begin{aligned}
&\sup_{(\upsilon,\{h_s\}_{s\in S})\in\Upsilon\times H_n^{N_\mathcal{E}}}\left|\frac{\mathrm{CV}(\upsilon,\{h_s\}_{s\in S}) - \mathrm{MISE}(\upsilon,\{h_s\}_{s\in S}) - \hat{\sigma}_\epsilon^2}{\mathrm{MISE}(\upsilon,\{h_s\}_{s\in S})}\right| \\
&\le \sup_{(\upsilon,\{h_s\}_{s\in S})\in\Upsilon\times H_n^{N_\mathcal{E}}}\left|\frac{\widetilde{\mathrm{ASE}}(\upsilon,\{h_s\}_{s\in S}) - \mathrm{ASE}(\upsilon,\{h_s\}_{s\in S})}{\mathrm{MISE}(\upsilon,\{h_s\}_{s\in S})}\right| \\
&\quad + \sup_{(\upsilon,\{h_s\}_{s\in S})\in\Upsilon\times H_n^{N_\mathcal{E}}}\left|\frac{\mathrm{ASE}(\upsilon,\{h_s\}_{s\in S}) - \mathrm{MISE}(\upsilon,\{h_s\}_{s\in S})}{\mathrm{MISE}(\upsilon,\{h_s\}_{s\in S})}\right| \\
&\quad + 2\sup_{(\upsilon,\{h_s\}_{s\in S})\in\Upsilon\times H_n^{N_\mathcal{E}}}\left|\frac{\mathrm{CT}(\upsilon,\{h_s\}_{s\in S})}{\mathrm{MISE}(\upsilon,\{h_s\}_{s\in S})}\right|.
\end{aligned}$$

Then, the theorem is proven due to Lemma 6.6.2, Lemma 6.6.4 (see page 168) and Lemma 6.6.5 (see page 170). ■

6.6.5 Formulation and proof of Lemma 6.6.2

Recall that C will denote a generic positive constant which may take on different values even in the same formula.

Lemma 6.6.2. *Under hypotheses of Theorem 6.2.6,*

$$\sup_{(\upsilon,\{h_s\}_{s\in S})\in\Upsilon\times H_n^{N_\mathcal{E}}}\left|\frac{\widetilde{ASE}(\upsilon,\{h_s\}_{s\in S})-ASE(\upsilon,\{h_s\}_{s\in S})}{MISE(\upsilon,\{h_s\}_{s\in S})}\right|\to 0\quad a.s.$$

where $\widetilde{ASE}(\upsilon,\{h_s\}_{s\in S})$ *and* $ASE(\upsilon,\{h_s\}_{s\in S})$ *are defined in* (6.16) *(see page 165) and* (6.15) *(see page 164), respectively.*

Proof. This proof is analogous to the proof of Lemma 3 by Ait-Saïdi et al. (2008). By Lemma 6.6.6 (see page 173), the following expression for $\mathrm{ASE}(\upsilon,\{h_s\}_{s\in S})$ can be obtained

$$\begin{aligned}
\mathrm{ASE}(\upsilon,\{h_s\}_{s\in S}) &= \frac{1}{n}\sum_{j=1}^{n}(m(X_j)-\hat{m}^{\upsilon}(X_j))^2=\frac{1}{n}\sum_{j=1}^{n}\left(\sum_{s\in S}(m_s^{\upsilon}(X_j)-\hat{m}_s^{\upsilon}(X_j))\right)^2\\
&=\frac{1}{n}\sum_{j=1}^{n}\left(\sum_{s\in S}(\hat{m}_{s,D}^{\upsilon}(X_j)(m_s^{\upsilon}(X_j)-\hat{m}_s^{\upsilon}(X_j))+(1-\hat{m}_{s,D}^{\upsilon}(X_j))(m_s^{\upsilon}(X_j)-\hat{m}_s^{\upsilon}(X_j)))\right)^2\\
&=\frac{1}{n}\sum_{j=1}^{n}\left(\sum_{s\in S}(\hat{m}_{s,D}^{\upsilon}(X_j)m_s^{\upsilon}(X_j)-\hat{m}_{s,N}^{\upsilon}(X_j))\right)^2+o_{a.co.}(\mathrm{ASE}(\upsilon,\{h_s\}_{s\in S}))\\
&=\mathrm{ASE}^*(\upsilon,\{h_s\}_{s\in S})+o_{a.co.}(\mathrm{ASE}(\upsilon,\{h_s\}_{s\in S})),
\end{aligned}$$

where $\mathrm{ASE}^*(\upsilon,\{h_s\}_{s\in S})=n^{-1}\sum_{j=1}^{n}(\sum_{s\in S}(\hat{m}_{s,D}^{\upsilon}(X_j)m_s^{\upsilon}(X_j)-\hat{m}_{s,N}^{\upsilon}(X_j)))^2$. Analogously, it can be seen that

$$\widetilde{\mathrm{ASE}}(\upsilon,\{h_s\}_{s\in S})=\widetilde{\mathrm{ASE}}^*(\upsilon,\{h_s\}_{s\in S})+o_{a.co.}(\widetilde{\mathrm{ASE}}(\upsilon,\{h_s\}_{s\in S})),$$

with $\widetilde{\mathrm{ASE}}^*(\upsilon,\{h_s\}_{s\in S})=n^{-1}\sum_{j=1}^{n}(\sum_{s\in S}(\hat{m}_{s,D}^{\upsilon,(-j)}(X_j)m_s^{\upsilon}(X_j)-\hat{m}_{s,N}^{\upsilon,(-j)}(X_j)))^2$. In order to finish the proof of the lemma, the equivalence between $\mathrm{ASE}^*(\upsilon,\{h_s\}_{s\in S})$ and $\widetilde{\mathrm{ASE}}^*(\upsilon,\{h_s\}_{s\in S})$ can be found by means of a similar procedure to that given by Härdle and Marron (1985) as follows.

First of all, note that

$$\hat{m}_{s,N}^{\upsilon,(-j)}(X_j)=\frac{1}{n-1}\sum_{i\neq j}\frac{Y_{i,s}^{\upsilon}\tilde{K}_{s,i}(X_j)}{\mathbb{E}(\tilde{K}_{s,0}(X_j))}=\frac{n}{n-1}\hat{m}_{s,N}^{\upsilon}(X_j)-\frac{1}{n-1}\frac{Y_{j,s}^{\upsilon}K(0)}{\mathbb{E}(\tilde{K}_{s,0}(X_j))},$$

$$\hat{m}_{s,D}^{\upsilon,(-j)}(X_j)=\frac{1}{n-1}\sum_{i\neq j}\frac{\tilde{K}_{s,i}(X_j)}{\mathbb{E}(\tilde{K}_{s,0}(X_j))}=\frac{n}{n-1}\hat{m}_{s,D}^{\upsilon}(X_j)-\frac{1}{n-1}\frac{K(0)}{\mathbb{E}(\tilde{K}_{s,0}(X_j))}.$$

Hence, one has that

$$\begin{aligned}
\widetilde{\mathrm{ASE}}^*(\upsilon,\{h_s\}_{s\in S}) &= \frac{1}{n}\sum_{j=1}^{n}\left(\sum_{s\in S}(\hat{m}_{s,D}^{\upsilon,(-j)}(X_j)m_s^{\upsilon}(X_j)-\hat{m}_{s,N}^{\upsilon,(-j)}(X_j))\right)^2\\
&=\frac{1}{n}\sum_{j=1}^{n}\left(\frac{n}{n-1}\sum_{s\in S}(\hat{m}_{s,D}^{\upsilon}(X_j)m_s^{\upsilon}(X_j)-\hat{m}_{s,N}^{\upsilon}(X_j))+\frac{K(0)}{n-1}\sum_{s\in S}\frac{Y_{j,s}^{\upsilon}-m_s^{\upsilon}(X_j)}{\mathbb{E}(\tilde{K}_{s,0}(X_j))}\right)^2\\
&=\frac{n^2}{(n-1)^2}\mathrm{ASE}^*(\upsilon,\{h_s\}_{s\in S})+2\frac{n}{(n-1)^2}K(0)\frac{1}{n}\sum_{j=1}^{n}\left(\sum_{s\in S}(\hat{m}_{s,D}^{\upsilon}(X_j)m_s^{\upsilon}(X_j)-\hat{m}_{s,N}^{\upsilon}(X_j))\right)\\
&\quad\cdot\left(\sum_{s\in S}\frac{Y_{j,s}^{\upsilon}-m_s^{\upsilon}(X_j)}{\mathbb{E}(\tilde{K}_{s,0}(X_j))}\right)+\frac{1}{(n-1)^2}K^2(0)\frac{1}{n}\sum_{j=1}^{n}\left(\sum_{s\in S}\frac{Y_{j,s}^{\upsilon}-m_s^{\upsilon}(X_j)}{\mathbb{E}(\tilde{K}_{s,0}(X_j))}\right)^2,
\end{aligned}$$

and, consequently,

$$
\begin{aligned}
|\widetilde{\mathrm{ASE}}^{*}(\upsilon,\{h_s\}_{s\in S})-\mathrm{ASE}^{*}(\upsilon,\{h_s\}_{s\in S})| &\leq \frac{2n-1}{(n-1)^2}\mathrm{ASE}^{*}(\upsilon,\{h_s\}_{s\in S}) \\
&+2\frac{n}{(n-1)^2}K(0)\left|\frac{1}{n}\sum_{j=1}^{n}\left(\sum_{s\in S}(\hat{m}^{\upsilon}_{s,D}(X_j)m^{\upsilon}_{s}(X_j)-\hat{m}^{\upsilon}_{s,N}(X_j))\right)\left(\sum_{s\in S}\frac{Y^{\upsilon}_{j,s}-m^{\upsilon}_{s}(X_j)}{\mathbb{E}(\tilde{K}_{s,0}(X_j))}\right)\right| \\
&+\frac{1}{(n-1)^2}K^2(0)\frac{1}{n}\sum_{j=1}^{n}\left(\sum_{s\in S}\frac{Y^{\upsilon}_{j,s}-m^{\upsilon}_{s}(X_j)}{\mathbb{E}(\tilde{K}_{s,0}(X_j))}\right)^2.
\end{aligned}
\tag{6.18}
$$

Using the Cauchy–Schwarz inequality, it can be found that

$$
\begin{aligned}
&\left|\frac{1}{n}\sum_{j=1}^{n}\left(\sum_{s\in S}(\hat{m}^{\upsilon}_{s,D}(X_j)m^{\upsilon}_{s}(X_j)-\hat{m}^{\upsilon}_{s,N}(X_j))\right)\left(\sum_{s\in S}\frac{Y^{\upsilon}_{j,s}-m^{\upsilon}_{s}(X_j)}{\mathbb{E}(\tilde{K}_{s,0}(X_j))}\right)\right| \\
&\leq (\mathrm{ASE}^{*}(\upsilon,\{h_s\}_{s\in S}))^{1/2}\left(\frac{1}{n}\sum_{j=1}^{n}\left(\sum_{s\in S}\frac{Y^{\upsilon}_{j,s}-m^{\upsilon}_{s}(X_j)}{\mathbb{E}(\tilde{K}_{s,0}(X_j))}\right)^2\right)^{1/2}.
\end{aligned}
\tag{6.19}
$$

In addition, the SLLN ensures that

$$
\frac{1}{n}\sum_{j=1}^{n}\left(\sum_{s\in S}\frac{Y^{\upsilon}_{j,s}-m^{\upsilon}_{s}(X_j)}{\mathbb{E}(\tilde{K}_{s,0}(X_j))}\right)^2\to\mathbb{E}\left(\left(\sum_{s\in S}\frac{Y^{\upsilon}_{s}-m^{\upsilon}_{s}(X)}{\mathbb{E}(\tilde{K}_{s,0}(X))}\right)^2\right)\quad a.s., \tag{6.20}
$$

whereas Lemma 6.6.7 (see page 174) states that

$$
\mathbb{E}\left(\left(\sum_{s\in S}\frac{Y^{\upsilon}_{s}-m^{\upsilon}_{s}(X)}{\mathbb{E}(\tilde{K}_{s,0}(X))}\right)^2\right)\leq C\left(\sum_{s\in S}\frac{1}{\phi(h_s)}\right)^2. \tag{6.21}
$$

Therefore, (6.18), (6.19), (6.20) and (6.21) ensure that

$$
\begin{aligned}
|\widetilde{\mathrm{ASE}}^{*}(\upsilon,\{h_s\}_{s\in S})-\mathrm{ASE}^{*}(\upsilon,\{h_s\}_{s\in S})| &\leq \frac{2n-1}{(n-1)^2}\mathrm{ASE}^{*}(\upsilon,\{h_s\}_{s\in S}) \\
&+2C^{1/2}\frac{n}{(n-1)^2}K(0)(\mathrm{ASE}^{*}(\upsilon,\{h_s\}_{s\in S}))^{1/2}\left(\sum_{s\in S}\frac{1}{\phi(h_s)}\right)(1+o_{a.s.}(1)) \\
&+C\frac{1}{(n-1)^2}K^2(0)\left(\sum_{s\in S}\frac{1}{\phi(h_s)}\right)^2(1+o_{a.s.}(1)).
\end{aligned}
$$

By the previous expression and Lemma 6.6.3, it can be found that

$$
\sup_{(\upsilon,\{h_s\}_{s\in S})\in\Upsilon\times H_n^{N_{\mathcal{E}}}}\left|\frac{\widetilde{\mathrm{ASE}}^{*}(\upsilon,\{h_s\}_{s\in S})-\mathrm{ASE}^{*}(\upsilon,\{h_s\}_{s\in S})}{\mathrm{MISE}(\upsilon,\{h_s\}_{s\in S})}\right|\to 0\quad a.s.
$$

which leads to Lemma 6.6.2. ■

6.6.6 Formulation and proof of Lemma 6.6.3

Recall that C will denote a generic positive constant which may take on different values even in the same formula.

Lemma 6.6.3. *Under hypotheses of Theorem 6.2.6,*

$$\mathit{MISE}(\upsilon, \{h_s\}_{s\in S}) \geq c_{10} \sum_{s\in S} \frac{1}{n\phi(h_s)}.$$

Proof. This proof is analogous to the proof of Lemma 1 by Ait-Saïdi et al. (2008). For each $(\upsilon, \{h_s\}_{s\in S}) \in \Upsilon \times H_n^{N_\varepsilon}$, one gets

$$\begin{aligned}\mathrm{MISE}(\upsilon, \{h_s\}_{s\in S}) &= \mathbb{E}((m(X) - \hat{m}^\upsilon(X))^2) = \mathbb{E}(\mathbb{E}((m(X) - \hat{m}^\upsilon(X))^2|X)) \\ &= \mathbb{E}((m(X) - \mathbb{E}(\hat{m}^\upsilon(X)|X))^2) + \mathbb{E}(\mathrm{Var}(\hat{m}^\upsilon(X)|X)) \geq \mathbb{E}(\mathrm{Var}(\hat{m}^\upsilon(X)|X)).\end{aligned}$$

Note that hypotheses of Theorem 6.2.6 (see page 148) ensure that Corollary 6.2.4 (see page 146) can be applied. Thus, Corollary 6.2.4 and assumptions **(C.6.3)**, **(C.6.10)**, **(C.6.14)** and **(C.6.15)** lead to

$$\mathbb{E}(\mathrm{Var}(\hat{m}^\upsilon(X)|X)) \geq C \sum_{s\in S} \frac{1}{n\phi(h_s)} \mathbb{E}\left(\mathbb{I}_{\{\tilde{\Psi}(X)\in\mathcal{E}_s^\upsilon\}}\right) = C \sum_{s\in S} \frac{1}{n\phi(h_s)} \mathbb{P}(\tilde{\Psi}(X) \in \mathcal{E}_s^\upsilon) \geq C \sum_{s\in S} \frac{1}{n\phi(h_s)},$$

which completes the proof of Lemma 6.6.3. ■

6.6.7 Formulation and proof of Lemma 6.6.4

Lemma 6.6.4. *Under hypotheses of Theorem 6.2.6,*

$$\sup_{(\upsilon,\{h_s\}_{s\in S})\in\Upsilon\times H_n^{N_\mathcal{E}}} \left| \frac{\mathit{ASE}(\upsilon, \{h_s\}_{s\in S}) - \mathit{MISE}(\upsilon, \{h_s\}_{s\in S})}{\mathit{MISE}(\upsilon, \{h_s\}_{s\in S})} \right| \to 0 \quad a.s.$$

where $\mathit{ASE}(\upsilon, \{h_s\}_{s\in S})$ *is defined in* (6.15) *(see page 164).*

Proof. This lemma is analogous to Lemma 2 by Ait-Saïdi et al. (2008). Recall that it was shown in the proof of Lemma 6.6.2 (see page 166) that

$$\mathrm{ASE}(\upsilon, \{h_s\}_{s\in S}) = \mathrm{ASE}^*(\upsilon, \{h_s\}_{s\in S}) + o_{a.co.}(\mathrm{ASE}(\upsilon, \{h_s\}_{s\in S})),$$

with $\mathrm{ASE}^*(\upsilon, \{h_s\}_{s\in S}) = n^{-1}\sum_{j=1}^n (\sum_{s\in S} (\hat{m}_{s,D}^\upsilon(X_j) m_s^\upsilon(X_j) - \hat{m}_{s,N}^\upsilon(X_j)))^2$. Similar calculations and Lemma 6.6.6 (see page 173) allow to obtain the following expression for $\mathrm{MISE}(\upsilon, \{h_s\}_{s\in S})$

$$\begin{aligned}\mathrm{MISE}(\upsilon, \{h_s\}_{s\in S}) &= \mathbb{E}((m(X) - \hat{m}^\upsilon(X))^2) = \mathbb{E}\left(\left(\sum_{s\in S} (m_s^\upsilon(X) - \hat{m}_s^\upsilon(X))\right)^2\right) \\ &= \mathbb{E}\left(\left(\sum_{s\in S} (\hat{m}_{s,D}^\upsilon(X)(m_s^\upsilon(X) - \hat{m}_s^\upsilon(X)) + (1 - \hat{m}_{s,D}^\upsilon(X))(m_s^\upsilon(X) - \hat{m}_s^\upsilon(X)))\right)^2\right) \\ &= \mathbb{E}\left(\left(\sum_{s\in S} (\hat{m}_{s,D}^\upsilon(X) m_s^\upsilon(X) - \hat{m}_{s,N}^\upsilon(X))\right)^2\right) + o_{a.co.}(\mathrm{MISE}(\upsilon, \{h_s\}_{s\in S})) \\ &= \mathrm{MISE}^*(\upsilon, \{h_s\}_{s\in S}) + o_{a.co.}(\mathrm{MISE}(\upsilon, \{h_s\}_{s\in S})),\end{aligned}$$

where $\mathrm{MISE}^*(\upsilon, \{h_s\}_{s\in S}) = \mathbb{E}\left(\left(\sum_{s\in S} (\hat{m}_{s,D}^\upsilon(X) m_s^\upsilon(X) - \hat{m}_{s,N}^\upsilon(X))\right)^2\right)$. Hence, the lemma can be proven by showing the equivalence between $\mathrm{ASE}^*(\upsilon, \{h_s\}_{s\in S})$ and $\mathrm{MISE}^*(\upsilon, \{h_s\}_{s\in S})$. Specifically, it

is enough to show that

$$\sup_{(\upsilon,\{h_s\}_{s\in S})\in\Upsilon\times H_n^{N_\mathcal{E}}}\left|\frac{\mathrm{ASE}^*(\upsilon,\{h_s\}_{s\in S})-\mathrm{MISE}^*(\upsilon,\{h_s\}_{s\in S})}{\mathrm{MISE}^*(\upsilon,\{h_s\}_{s\in S})}\right|\to 0\quad a.s. \tag{6.22}$$

For this purpose, first of all, assume that $s_0\in S$ is selected, and $\upsilon\in\Upsilon$ and $h_s\in H_n$ for all $s\in S$ with $s\neq s_0$ are fixed. In addition, consider $\lambda\in\Lambda=\{1/\phi(h_{s_0})+\sum_{s\neq s_0}1/\phi(h_s):h_{s_0}\in H_n\}$. Then, a delta sequence estimator $\hat{g}_\lambda:\mathcal{H}\to\mathbb{R}$ can be defined as follows

$$\hat{g}_\lambda(x)=\frac{1}{n}\sum_{i=1}^{n}\delta_\lambda(x,X_i,Y_i),$$

where

$$\delta_\lambda(x,X_i,Y_i)=\frac{(m_{s_0}^\upsilon(x)-Y_{i,s_0}^\upsilon)K(f(\lambda)^{-1}\|X_i-x\|)}{\mathbb{E}(K(f(\lambda)^{-1}\|X_0-x\|))}+\sum_{s\neq s_0}\frac{(m_s^\upsilon(x)-Y_{i,s}^\upsilon)\tilde{K}_{s,i}(x)}{\mathbb{E}(\tilde{K}_{s,0}(x))},$$

with $f(\lambda)=\phi^{-1}((\lambda-(\sum_{s\neq s_0}1/\phi(h_s)))^{-1})$, being ϕ^{-1} the inverse function of ϕ (note that **(C.6.10)** ensures that ϕ is a bijective function, so there is a unique inverse function ϕ^{-1} which is also a bijection). Given that $f(\lambda)=h_{s_0}$, in fact, $\delta_\lambda(x,X_i,Y_i)$ is

$$\delta_\lambda(x,X_i,Y_i)=\sum_{s\in S}\frac{(m_s^\upsilon(x)-Y_{i,s}^\upsilon)\tilde{K}_{s,i}(x)}{\mathbb{E}(\tilde{K}_{s,0}(x))}. \tag{6.23}$$

In this situation, $\hat{g}_\lambda(x)=\sum_{s\in S}(\hat{m}_{s,D}^\upsilon(x)m_s^\upsilon(x)-\hat{m}_{s,N}^\upsilon(x))$, and it may be considered that $\hat{g}_\lambda$ estimates the operator $g:\mathcal{H}\to\mathbb{R}$ defined as $g(x)=0$ for all $x\in\mathcal{H}$. Besides, computing the mean integrated squared error and the average squared error for $\hat{g}_\lambda$ (denoted by $\mathrm{MISE}_{\hat{g}}(\lambda)$ and $\mathrm{ASE}_{\hat{g}}(\lambda)$, respectively), one has

$$\mathrm{MISE}_{\hat{g}}(\lambda)=\mathbb{E}((g(X)-\hat{g}_\lambda(X))^2)=\mathbb{E}((\hat{g}_\lambda(X))^2)=\mathrm{MISE}^*(\upsilon,\{h_s\}_{s\in S}), \tag{6.24}$$

and

$$\mathrm{ASE}_{\hat{g}}(\lambda)=\frac{1}{n}\sum_{j=1}^{n}(g(X_j)-\hat{g}_\lambda(X_j))^2=\frac{1}{n}\sum_{j=1}^{n}(\hat{g}_\lambda(X_j))^2=\mathrm{ASE}^*(\upsilon,\{h_s\}_{s\in S}). \tag{6.25}$$

On the other hand, **(C.6.11)**, **(C.6.18)** and Lemma 6.6.9 (see page 177) indicate that the assumptions of the theoretical results for delta sequence estimators in Marron and Härdle (1986) hold. Thus, applying Theorem 2 by Marron and Härdle (1986), one gets

$$\sup_{\lambda\in\Lambda}\left|\frac{\mathrm{ASE}_{\hat{g}}(\lambda)-\mathrm{MISE}_{\hat{g}}(\lambda)}{\mathrm{MISE}_{\hat{g}}(\lambda)}\right|\to 0\quad a.s. \tag{6.26}$$

Taking into account (6.24) and (6.25), it can be seen

$$\left|\frac{\mathrm{ASE}^*(\upsilon,\{h_s\}_{s\in S})-\mathrm{MISE}^*(\upsilon,\{h_s\}_{s\in S})}{\mathrm{MISE}^*(\upsilon,\{h_s\}_{s\in S})}\right|\leq\sup_{\lambda\in\Lambda}\left|\frac{\mathrm{ASE}_{\hat{g}}(\lambda)-\mathrm{MISE}_{\hat{g}}(\lambda)}{\mathrm{MISE}_{\hat{g}}(\lambda)}\right|.$$

Hence, (6.22) is verified due to (6.26), and consequently the proof of the lemma is complete. ■

6.6.8 Formulation and proof of Lemma 6.6.5

Recall that C will denote a generic positive constant which may take on different values even in the same formula.

Lemma 6.6.5. *Under hypotheses of Theorem 6.2.6,*

$$\sup_{(\upsilon,\{h_s\}_{s\in S})\in\Upsilon\times H_n^{N_\mathcal{E}}}\left|\frac{CT(\upsilon,\{h_s\}_{s\in S})}{MISE(\upsilon,\{h_s\}_{s\in S})}\right|\to 0 \quad a.s.$$

where $CT(\upsilon,\{h_s\}_{s\in S})$ is defined in (6.17) *(see page 165).*

Proof. The proof of this lemma can be obtained following the proof of Lemma 4 in Ait-Saïdi et al. (2008), which in turn is based on the ideas proposed by Härdle and Marron (1985), as follows. By the second statement in Lemma 6.6.6 (see page 173), one can see that

$$\begin{aligned}
\mathrm{CT}(\upsilon,\{h_s\}_{s\in S}) &= \frac{1}{n}\sum_{j=1}^{n}\epsilon_j(\hat{m}^{\upsilon,(-j)}(X_j)-m(X_j)) = \frac{1}{n}\sum_{j=1}^{n}\epsilon_j\left(\sum_{s\in S}(\hat{m}_s^{\upsilon,(-j)}(X_j)-m_s^{\upsilon}(X_j))\right)\\
&= \frac{1}{n}\sum_{j=1}^{n}\epsilon_j\left(\sum_{s\in S}(\hat{m}_{s,D}^{\upsilon,(-j)}(X_j)(\hat{m}_s^{\upsilon,(-j)}(X_j)-m_s^{\upsilon}(X_j)))\right)\\
&\quad+\frac{1}{n}\sum_{j=1}^{n}\epsilon_j\left(\sum_{s\in S}((1-\hat{m}_{s,D}^{\upsilon,(-j)}(X_j))(\hat{m}_s^{\upsilon,(-j)}(X_j)-m_s^{\upsilon}(X_j)))\right)\\
&= \frac{1}{n}\sum_{j=1}^{n}\epsilon_j\left(\sum_{s\in S}(\hat{m}_{s,N}^{\upsilon,(-j)}(X_j)-\hat{m}_{s,D}^{\upsilon,(-j)}(X_j)m_s^{\upsilon}(X_j))\right)+o_{a.co}(\mathrm{CT}(\upsilon,\{h_s\}_{s\in S}))\\
&= \mathrm{CT}^*(\upsilon,\{h_s\}_{s\in S})+o_{a.co}(\mathrm{CT}(\upsilon,\{h_s\}_{s\in S})),
\end{aligned}$$

where $\mathrm{CT}^*(\upsilon,\{h_s\}_{s\in S}) = n^{-1}\sum_{j=1}^{n}\epsilon_j\left(\sum_{s\in S}(\hat{m}_{s,N}^{\upsilon,(-j)}(X_j)-\hat{m}_{s,D}^{\upsilon,(-j)}(X_j)m_s^{\upsilon}(X_j))\right)$. This fact and Lemma 6.6.3 (see page 168) allow to deduce that it is enough to show

$$\sup_{(\upsilon,\{h_s\}_{s\in S})\in\Upsilon\times H_n^{N_\mathcal{E}}}\left|\left(\sum_{s\in S}\frac{1}{n\phi(h_s)}\right)^{-1}\mathrm{CT}^*(\upsilon,\{h_s\}_{s\in S})\right|\to 0 \quad a.s.$$

in order to prove the lemma. In addition, note that if $i\neq j$ and $\|X_i-X_j\|\leq h_s$, (6.3) (see page 144) and (6.5) (see page 144) imply that

$$(Y_{i,s}^{\upsilon}-m_s^{\upsilon}(X_j)) = (Y_{i,s}^{\upsilon}-m_s^{\upsilon}(X_i))+(m_s^{\upsilon}(X_i)-m_s^{\upsilon}(X_j)) = (\epsilon_i+(m(X_i)-m(X_j)))\,\mathbb{I}_{\{\tilde{\Psi}(X_j)\in\mathcal{E}_s^{\upsilon}\}}.$$

Therefore,

$$\begin{aligned}
|\mathrm{CT}^*(\upsilon,\{h_s\}_{s\in S})| &= \left|\frac{1}{n}\sum_{j=1}^{n}\epsilon_j\left(\sum_{s\in S}(\hat{m}_{s,N}^{\upsilon,(-j)}(X_j)-\hat{m}_{s,D}^{\upsilon,(-j)}(X_j)m_s^{\upsilon}(X_j))\right)\right|\\
&= \left|\sum_{s\in S}\left(\frac{1}{n(n-1)}\sum_{j=1}^{n}\sum_{i\neq j}\frac{\epsilon_j(Y_{i,s}^{\upsilon}-m_s^{\upsilon}(X_j))\tilde{K}_{s,i}(X_j)}{\mathbb{E}(\tilde{K}_{s,0}(X_j))}\right)\right|\\
&\leq \sum_{s\in S}\left(\frac{1}{n(n-1)}\sum_{j=1}^{n}\sum_{i\neq j}\frac{|\epsilon_j||\epsilon_i|\mathbb{I}_{\{\tilde{\Psi}(X_i)\in\mathcal{E}_s^{\upsilon}\}}\tilde{K}_{s,i}(X_j)}{\mathbb{E}(\tilde{K}_{s,0}(X_j))}\right)\\
&\quad+\sum_{s\in S}\left(\frac{1}{n(n-1)}\sum_{j=1}^{n}\sum_{i\neq j}\frac{|\epsilon_j||m(X_i)-m(X_j)|\mathbb{I}_{\{\tilde{\Psi}(X_j)\in\mathcal{E}_s^{\upsilon}\}}\tilde{K}_{s,i}(X_j)}{\mathbb{E}(\tilde{K}_{s,0}(X_j))}\right)\\
&\leq \sum_{s\in S}\left(\frac{1}{n(n-1)}\sum_{j=1}^{n}\sum_{i\neq j}U_{i,j,s}^{\upsilon}+\frac{1}{n(n-1)}\sum_{j=1}^{n}\sum_{i\neq j}V_{i,j,s}^{\upsilon}\right),
\end{aligned}$$

where

$$U^{\upsilon}_{i,j,s} = \frac{|\epsilon_j||\epsilon_i|\tilde{K}_{s,i}(X_j)}{\mathbb{E}(\tilde{K}_{s,0}(X_j))} \quad \text{and} \quad V^{\upsilon}_{i,j,s} = \frac{|\epsilon_j||m(X_i)-m(X_j)|\tilde{K}_{s,i}(X_j)}{\mathbb{E}(\tilde{K}_{s,0}(X_j)),}$$

for $i, j \in \{1, \ldots, n\}$ such that $i \neq j$. Furthermore, note that

$$\left|\left(\sum_{s\in S}\frac{1}{n\phi(h_s)}\right)^{-1} \mathrm{CT}^*(\upsilon, \{h_s\}_{s\in S})\right| \leq \left|\left(\frac{1}{n\phi(h_s)}\right)^{-1} \mathrm{CT}^*(\upsilon, \{h_s\}_{s\in S})\right|$$
$$\leq \sum_{s\in S}\left(\left|\left(\frac{1}{n\phi(h_s)}\right)^{-1}\frac{1}{n(n-1)}\sum_{j=1}^{n}\sum_{i\neq j}U^{\upsilon}_{i,j,s}\right| + \left|\left(\frac{1}{n\phi(h_s)}\right)^{-1}\frac{1}{n(n-1)}\sum_{j=1}^{n}\sum_{i\neq j}V^{\upsilon}_{i,j,s}\right|\right).$$

Consequently, the lemma will be established as soon as, for each $s \in S$, one states that

$$\sup_{(\upsilon,h_s)\in\Upsilon\times H_n}\left|\left(\frac{1}{n\phi(h_s)}\right)^{-1}\frac{1}{n(n-1)}\sum_{j=1}^{n}\sum_{i\neq j}U^{\upsilon}_{i,j,s}\right| \to 0 \quad a.s. \tag{6.27}$$

and

$$\sup_{(\upsilon,h_s)\in\Upsilon\times H_n^{N\varepsilon}}\left|\left(\frac{1}{n\phi(h_s)}\right)^{-1}\frac{1}{n(n-1)}\sum_{j=1}^{n}\sum_{i\neq j}V^{\upsilon}_{i,j,s}\right| \to 0 \quad a.s. \tag{6.28}$$

In order to prove (6.27), note that by **(C.6.18)** and Chebyshev's inequality, given $\eta > 0$ and for all $p = 1, 2, \ldots$, one has

$$\mathbb{P}\left(\sup_{(\upsilon,h_s)\in\Upsilon\times H_n}\left|\left(\frac{1}{n\phi(h_s)}\right)^{-1}\frac{1}{n(n-1)}\sum_{j=1}^{n}\sum_{i\neq j}U^{\upsilon}_{i,j,s}\right| > \eta\right)$$
$$\leq \eta^{-2p} card(\Upsilon\times H_n)\sup_{(\upsilon,h_s)\in\Upsilon\times H_n}\mathbb{E}\left(\left(\left(\frac{1}{n\phi(h_s)}\right)^{-1}\frac{1}{n(n-1)}\sum_{j=1}^{n}\sum_{i\neq j}U^{\upsilon}_{i,j,s}\right)^{2p}\right)$$
$$\leq \eta^{-2p}n^{\alpha}\sup_{(\upsilon,h_s)\in\Upsilon\times H_n}\left\{\frac{\phi(h_s)^{2p}}{(n-1)^{2p}}\mathbb{E}\left(\left(\sum_{j=1}^{n}\sum_{i\neq j}U^{\upsilon}_{i,j,s}\right)^{2p}\right)\right\}.$$

Hence, it is enough to show that, for p large enough,

$$\sum_{n=1}^{\infty}n^{\alpha}\sup_{(\upsilon,h_s)\in\Upsilon\times H_n}\left\{\frac{\phi(h_s)^{2p}}{(n-1)^{2p}}\mathbb{E}\left(\left(\sum_{j=1}^{n}\sum_{i\neq j}U^{\upsilon}_{i,j,s}\right)^{2p}\right)\right\} < \infty, \tag{6.29}$$

to prove (6.27) due to Borel–Cantelli Lemma. Analogously, it can be found that (6.28) can be verified by showing that

$$\sum_{n=1}^{\infty}n^{\alpha}\sup_{(\upsilon,h_s)\in\Upsilon\times H_n}\left\{\frac{\phi(h_s)^{2p}}{(n-1)^{2p}}\mathbb{E}\left(\left(\sum_{j=1}^{n}\sum_{i\neq j}V^{\upsilon}_{i,j,s}\right)^{2p}\right)\right\} < \infty. \tag{6.30}$$

To obtain (6.29), note that using Lemma 6.6.8 (see page 176) it can be seen that

$$\mathbb{E}\left(\left(\sum_{j=1}^{n}\sum_{i\neq j}U^{\upsilon}_{i,j,s}\right)^{2p}\right) = \sum_{I_{2p}}\mathbb{E}\left(\prod_{l=1}^{2p}U^{\upsilon}_{i_l,j_l,s}\right) \leq C\phi(h_s)^{-2p}\sum_{q=2}^{4p}\sum_{J_q}\mathbb{E}\left(\prod_{l=1}^{q}|\epsilon_{r_l}|^{a_l}\tilde{K}^{b_l}_{s,u_l}(X_{w_l})\right),$$

where $I_{2p} = \{(i_1, \ldots, i_{2p}, j_1, \ldots, j_{2p}) \in \{1, \ldots, n\}^{2p}$, such that $i_1 \neq j_1, \ldots, i_{2p} \neq j_{2p}\}$, $J_q \subset I_{2p}$ is the subset which contains the elements of I_{2p} with only q different integers, and $\sum_{l=1}^{q} a_l = 4p$ $(a_l \geq 1)$ and $\sum_{l=1}^{q} b_l = 2p$. It can be shown that

$$\begin{aligned}\mathbb{E}\left(\prod_{l=1}^{q} |\epsilon_{r_l}|^{a_l} \tilde{K}_{s,u_l}^{b_l}(X_{w_l})\right) &= \mathbb{E}\left(\mathbb{E}\left(\prod_{l=1}^{q} |\epsilon_{r_l}|^{a_l} \tilde{K}_{s,u_l}^{b_l}(X_{w_l}) | X_{r_1}, \ldots, X_{r_q}\right)\right) \\ &= \mathbb{E}\left(\prod_{l=1}^{q} \mathbb{E}(|\epsilon_{r_l}|^{a_l} | X_{r_l}) \prod_{l=1}^{q} \tilde{K}_{s,u_l}^{b_l}(X_{w_l})\right).\end{aligned}$$

This last quantity vanishes when $q > 2p$. Using this fact and Lemma 6.6.8 (see page 176), and taking into account that there are $q/2$ separated pairs with q different integers, one can obtain

$$\mathbb{E}\left(\left(\sum_{j=1}^{n}\sum_{i\neq j} U_{i,j,s}^{\upsilon}\right)^{2p}\right) \leq C\phi(h_s)^{-2p} \sum_{q=2}^{2p} n^q \phi(h_s)^{q/2} \leq C n^{2p} \phi(h_s)^{-p},$$

where the last inequality is due to **(C.6.11)**. Therefore, using **(C.6.11)** again, one has

$$\sum_{n=1}^{\infty} n^{\alpha} \sup_{(\upsilon,h_s)\in\Upsilon\times H_n} \left\{\frac{\phi(h_s)^{2p}}{(n-1)^{2p}} \mathbb{E}\left(\left(\sum_{j=1}^{n}\sum_{i\neq j} U_{i,j,s}^{\upsilon}\right)^{2p}\right)\right\} \leq C \sum_{n=1}^{\infty} n^{\alpha-\nu_1 p},$$

so (6.29) holds for p large enough, and consequently (6.27) is proven.

Analogously, it can be checked (6.30) as follows in order to show the convergence for the term related to $V_{i,j,s}^{\upsilon}$. Firstly, Lemma 6.6.8 (see page 176) and **(C.6.12)** can be used to get

$$\mathbb{E}\left(\left(\sum_{j=1}^{n}\sum_{i\neq j} V_{i,j,s}^{\upsilon}\right)^{2p}\right) = \sum_{I_{2p}} \mathbb{E}\left(\prod_{l=1}^{2p} V_{i_l,j_l,s}^{\upsilon}\right) \leq C\phi(h_s)^{-2p} \sum_{q=2}^{4p}\sum_{J_q} \mathbb{E}\left(\prod_{l=1}^{q} |\epsilon_{r_l}|^{a_l} \tilde{K}_{s,u_l}^{b_l}(X_{w_l})\right),$$

with I_{2p} and J_q defined as above, and $\sum_{l=1}^{q} a_l = \sum_{l=1}^{q} b_l = \sum_{l=1}^{q} c_l = 2p$. Then, similar arguments to those used for the case $U_{i,j,s}^{\upsilon}$ above lead to

$$\sum_{n=1}^{\infty} n^{\alpha} \sup_{(\upsilon,h_s)\in\Upsilon\times H_n} \left\{\frac{\phi(h_s)^{2p}}{(n-1)^{2p}} \mathbb{E}\left(\left(\sum_{j=1}^{n}\sum_{i\neq j} V_{i,j,s}^{\upsilon}\right)^{2p}\right)\right\} \leq C \sum_{n=1}^{\infty} n^{\alpha-\nu_1 p}.$$

Consequently, one gets (6.30) for p large enough, and thus (6.28) is shown. ■

6.6.9 Proof of Theorem 6.2.8

The result stated in Theorem 6.2.8 can be proven following step by step the proof of Theorem 6.2.6 as follows. Given that $\mathrm{MISE}(\upsilon, \{h_{s,\mathrm{CV}}(\upsilon)\}_{s\in S}) \geq \mathrm{MISE}(\upsilon, \{h_s^*(\upsilon)\}_{s\in S})$ and $\mathrm{CV}(\upsilon, \{h_s^*(\upsilon)\}_{s\in S}) \geq \mathrm{CV}(\upsilon, \{h_{s,\mathrm{CV}}(\upsilon)\}_{s\in S})$, it can be shown that

$$\begin{aligned}|\mathrm{MISE}(\upsilon, \{h_{s,\mathrm{CV}}(\upsilon)\}_{s\in S}) - \mathrm{MISE}(\upsilon, \{h_s^*(\upsilon)\}_{s\in S})| &\leq |-\mathrm{CV}(\upsilon, \{h_{s,\mathrm{CV}}(\upsilon)\}_{s\in S}) \\ &+ \mathrm{MISE}(\upsilon, \{h_{s,\mathrm{CV}}(\upsilon)\}_{s\in S}) + \hat{\sigma}_{\epsilon}^2 + \mathrm{CV}(\upsilon, \{h_s^*(\upsilon)\}_{s\in S}) - \mathrm{MISE}(\upsilon, \{h_s^*(\upsilon)\}_{s\in S}) - \hat{\sigma}_{\epsilon}^2|,\end{aligned}$$

with $\hat{\sigma}_\epsilon^2 = n^{-1}\sum_{j=1}^n \epsilon_j^2$ as usual. Then

$$\begin{aligned}
&\left|\frac{\text{MISE}(\upsilon, \{h_{s,\text{CV}}(\upsilon)\}_{s\in S}) - \text{MISE}(\upsilon, \{h_s^*(\upsilon)\}_{s\in S})}{\text{MISE}(\upsilon, \{h_{s,\text{CV}}(\upsilon)\}_{s\in S})}\right| \\
&\leq \left|\frac{\text{CV}(\upsilon, \{h_{s,\text{CV}}(\upsilon)\}_{s\in S}) - \text{MISE}(\upsilon, \{h_{s,\text{CV}}(\upsilon)\}_{s\in S}) - \hat{\sigma}_\epsilon^2}{\text{MISE}(\upsilon, \{h_{s,\text{CV}}(\upsilon)\}_{s\in S})}\right| \\
&\quad + \left|\frac{\text{CV}(\upsilon, \{h_s^*(\upsilon)\}_{s\in S}) - \text{MISE}(\upsilon, \{h_s^*(\upsilon)\}_{s\in S}) - \hat{\sigma}_\epsilon^2}{\text{MISE}(\upsilon, \{h_s^*(\upsilon)\}_{s\in S})}\right|\left|\frac{\text{MISE}(\upsilon, \{h_s^*(\upsilon)\}_{s\in S})}{\text{MISE}(\upsilon, \{h_{s,\text{CV}}(\upsilon)\}_{s\in S})}\right| \\
&\leq 2 \sup_{(\upsilon, \{h_s\}_{s\in S})\in \Upsilon\times H_n^{N_\mathcal{E}}} \left|\frac{\text{CV}(\upsilon, \{h_s\}_{s\in S}) - \text{MISE}(\upsilon, \{h_s\}_{s\in S}) - \hat{\sigma}_\epsilon^2}{\text{MISE}(\upsilon, \{h_s\}_{s\in S})}\right|.
\end{aligned}$$

Consequently, the theorem is proven due to Lemma 6.6.1 (see page 165).

6.6.10 Proof of Theorem 6.2.9

In this case, one has that

$$\begin{aligned}
&|\text{CV}(\upsilon, \{h_{s,\text{CV}}(\upsilon)\}_{s\in S}) - \text{MISE}(\upsilon, \{h_s^*(\upsilon)\}_{s\in S}) - \hat{\sigma}_\epsilon^2| = |\text{MISE}(\upsilon, \{h_{s,\text{CV}}(\upsilon)\}_{s\in S}) \\
&\quad - \text{MISE}(\upsilon, \{h_s^*(\upsilon)\}_{s\in S}) + \text{CV}(\upsilon, \{h_{s,\text{CV}}(\upsilon)\}_{s\in S}) - \text{MISE}(\upsilon, \{h_{s,\text{CV}}(\upsilon)\}_{s\in S}) - \hat{\sigma}_\epsilon^2|.
\end{aligned}$$

Therefore,

$$\begin{aligned}
&\left|\frac{\text{CV}(\upsilon, \{h_{s,\text{CV}}(\upsilon)\}_{s\in S}) - \text{MISE}(\upsilon, \{h_s^*(\upsilon)\}_{s\in S}) - \hat{\sigma}_\epsilon^2}{\text{MISE}(\upsilon, \{h_s^*(\upsilon)\}_{s\in S})}\right| \leq \left|\frac{\text{MISE}(\upsilon, \{h_{s,\text{CV}}(\upsilon)\}_{s\in S})}{\text{MISE}(\upsilon, \{h_s^*(\upsilon)\}_{s\in S})} - 1\right| \\
&\quad + \left|\frac{\text{CV}(\upsilon, \{h_{s,\text{CV}}(\upsilon)\}_{s\in S}) - \text{MISE}(\upsilon, \{h_{s,\text{CV}}(\upsilon)\}_{s\in S}) - \hat{\sigma}_\epsilon^2}{\text{MISE}(\upsilon, \{h_{s,\text{CV}}(\upsilon)\}_{s\in S})}\right|\left|\frac{\text{MISE}(\upsilon, \{h_{s,\text{CV}}(\upsilon)\}_{s\in S})}{\text{MISE}(\upsilon, \{h_s^*(\upsilon)\}_{s\in S})}\right| \\
&\leq \left|\frac{\text{MISE}(\upsilon, \{h_{s,\text{CV}}(\upsilon)\}_{s\in S})}{\text{MISE}(\upsilon, \{h_s^*(\upsilon)\}_{s\in S})} - 1\right| \\
&\quad + \sup_{(\upsilon, \{h_s\}_{s\in S})\in \Upsilon\times H_n^{N_\mathcal{E}}} \left|\frac{\text{CV}(\upsilon, \{h_s\}_{s\in S}) - \text{MISE}(\upsilon, \{h_s\}_{s\in S}) - \hat{\sigma}_\epsilon^2}{\text{MISE}(\upsilon, \{h_s\}_{s\in S})}\right|\left|\frac{\text{MISE}(\upsilon, \{h_{s,\text{CV}}(\upsilon)\}_{s\in S})}{\text{MISE}(\upsilon, \{h_s^*(\upsilon)\}_{s\in S})}\right|.
\end{aligned}$$

Hence, Theorem 6.2.8 (see page 148) and Lemma 6.6.1 (see page 165) allow to finish the proof.

6.6.11 Auxiliary technical lemmas

Formulation and proof of Lemma 6.6.6

For the lemma below, recall the definition of the almost completely convergence introduced in Definition 2.4.6 and Definition 2.4.7 (see Chapter 2, page 44).

Lemma 6.6.6. *Under* **(C.6.1)**, **(C.6.10)**, **(C.6.15)** *and* **(C.6.17)**, *it holds that*

$$\sup_{x\in\mathcal{C}} |\hat{m}_{s,D}^{\upsilon}(x) - 1| = O_{a.co.}\left(\sqrt{\frac{H_\mathcal{C}(\log n/n)}{n\phi(h_s)}}\right), \quad \forall s \in S,$$

and

$$\sup_{x\in\mathcal{C}} |\hat{m}_{s,D}^{\upsilon,(-j)}(x) - 1| = O_{a.co.}\left(\sqrt{\frac{H_\mathcal{C}(\log n/n)}{n\phi(h_s)}}\right), \quad \forall s \in S.$$

Proof. Ferraty et al. (2010b) studied rates of uniform consistency for a generalized nonparametric regression context in terms of almost completely convergence. In particular, the first statement in the lemma corresponds to Lemma 8 in Ferraty et al. (2010b). Furthermore, the result still hold if $\hat{m}_{s,D}^{\upsilon}$ is replaced with $\hat{m}_{s,D}^{\upsilon,(-j)}$, since the second statement can be seen as a corollary of the first one. ■

Formulation and proof of Lemma 6.6.7

Recall that C will denote a generic positive constant which may take on different values even in the same formula.

Lemma 6.6.7. *Under hypotheses of Theorem 6.2.6,*

(i) for $p = 1, 2, \ldots$, there exists $c_{11,p} > 0$ such that

$$\mathbb{E}\left(\left|\sum_{s\in S}\frac{Y_s^v - m_s^v(X)}{\mathbb{E}(\tilde{K}_{s,0}(X))}\right|^p\right) \leq c_{11,p}\left(\sum_{s\in S}\frac{1}{\phi(h_s)}\right)^p.$$

(ii) for $p = 1, 2, \ldots$, there exists $c_{12,p} > 0$ such that

$$\mathbb{E}\left(\left|\sum_{s\in S}\frac{(Y_{i,s}^v - m_s^v(X_i))\tilde{K}_{s,i}(X_i)}{\mathbb{E}(\tilde{K}_{s,0}(X_i))}\right|^p |X_i\right) \leq c_{12,p}\left(\sum_{s\in S}\frac{1}{\phi(h_s)}\right)^p \quad a.s.$$

(iii) for $p = 1, 2, \ldots$, there exists $c_{13,p} > 0$ such that for all $i \neq j$

$$\mathbb{E}\left(\left|\sum_{s\in S}\frac{(Y_{j,s}^v - m_s^v(X_i))\tilde{K}_{s,j}(X_i)}{\mathbb{E}(\tilde{K}_{s,0}(X_i))}\right|^p |X_i\right) \leq c_{13,p}\left(\sum_{s\in S}\frac{1}{\phi(h_s)}\right)^{p-1} \quad a.s.$$

(iv) there exists $c_{14} > 0$ such that for all $i \neq j$

$$\mathbb{E}\left(\left(\sum_{s\in S}\frac{(Y_{j,s}^v - m_s^v(X_i))\tilde{K}_{s,j}(X_i)}{\mathbb{E}(\tilde{K}_{s,0}(X_i))}\right)^2 |X_i\right) \geq c_{14}\sum_{s\in S}\frac{\mathbb{I}_{\{\tilde{\Psi}(X_i)\in\mathcal{E}_s^v\}}}{\phi(h_s)} \quad a.s.$$

Proof.

Proof of item *(i)*. First of all, note that (6.3) (see page 144) ensures that

$$Y_s^v - m_s^v(X) = Y\mathbb{I}_{\{\tilde{\Psi}(X)\in\mathcal{E}_s^v\}} - m(X)\mathbb{I}_{\{\tilde{\Psi}(X)\in\mathcal{E}_s^v\}} = (Y - m(X))\mathbb{I}_{\{\tilde{\Psi}(X)\in\mathcal{E}_s^v\}} = \epsilon\mathbb{I}_{\{\tilde{\Psi}(X)\in\mathcal{E}_s^v\}}. \quad (6.31)$$

Using (6.31), and the fact that $\mathcal{E}_{s_1}^v \cap \mathcal{E}_{s_2}^v = \emptyset$ for all $s_1 \neq s_2$, one gets

$$\mathbb{E}\left(\left|\sum_{s\in S}\frac{Y_s^v - m_s^v(X)}{\mathbb{E}(\tilde{K}_{s,0}(X))}\right|^p\right) = \mathbb{E}\left(|\epsilon|^p\left|\sum_{s\in S}\frac{\mathbb{I}_{\{\tilde{\Psi}(X)\in\mathcal{E}_s^v\}}}{\mathbb{E}(\tilde{K}_{s,0}(X))}\right|^p\right) = \mathbb{E}\left(\mathbb{E}(|\epsilon|^p|X)\sum_{s\in S}\frac{\mathbb{I}_{\{\tilde{\Psi}(X)\in\mathcal{E}_s^v\}}}{(\mathbb{E}(\mathbb{E}(\tilde{K}_{s,0}(X)|X))^p}\right).$$

Then, by assumption **(C.6.13)** and Lemma 6.6.8 (see page 176),

$$\begin{aligned}\mathbb{E}\left(\left|\sum_{s\in S}\frac{Y_s^v - m_s^v(X)}{\mathbb{E}(\tilde{K}_{s,0}(X))}\right|^p\right) &\leq C\mathbb{E}\left(\sum_{s\in S}\frac{\mathbb{I}_{\{\tilde{\Psi}(X)\in\mathcal{E}_s^v\}}}{(\phi(h_s))^p}\right) = C\sum_{s\in S}\frac{\mathbb{E}(\mathbb{I}_{\{\tilde{\Psi}(X)\in\mathcal{E}_s^v\}})}{(\phi(h_s))^p} = C\sum_{s\in S}\frac{\mathbb{P}(\tilde{\Psi}(X)\in\mathcal{E}_s^v)}{(\phi(h_s))^p}\\ &\leq C\sum_{s\in S}\frac{1}{(\phi(h_s))^p} \leq C\left(\sum_{s\in S}\frac{1}{\phi(h_s)}\right)^p.\end{aligned}$$

Proof of item *(ii)*. By (6.3) (see page 144), **(C.6.13)** and the fact that $\mathcal{E}_{s_1}^{\upsilon} \cap \mathcal{E}_{s_2}^{\upsilon} = \emptyset$ for all $s_1 \neq s_2$, one has

$$\mathbb{E}\left(\left|\sum_{s\in S}\frac{(Y_{i,s}^{\upsilon}-m_s^{\upsilon}(X_i))\tilde{K}_{s,i}(X_i)}{\mathbb{E}(\tilde{K}_{s,0}(X_i))}\right|^p|X_i\right)=\mathbb{E}\left(|\epsilon_i|^p\left(\sum_{s\in S}\frac{\mathbb{I}_{\{\tilde{\Psi}(X_i)\in\mathcal{E}_s^{\upsilon}\}}K(0)}{\mathbb{E}(\tilde{K}_{s,0}(X_i))}\right)^p|X_i\right)$$
$$\leq C\mathbb{E}(|\epsilon_i|^p|X_i)\left(\sum_{s\in S}\frac{\mathbb{I}_{\{\tilde{\Psi}(X_i)\in\mathcal{E}_s^{\upsilon}\}}}{\mathbb{E}(\tilde{K}_{s,0}(X_i))}\right)^p\leq C\sum_{s\in S}\frac{1}{(\mathbb{E}(\tilde{K}_{s,0}(X_i)))^p}=C\sum_{s\in S}\frac{1}{(\mathbb{E}(\mathbb{E}(\tilde{K}_{s,0}(X_i)|X_i)))^p}.$$

Thus, Lemma 6.6.8 (see page 176) leads to

$$\mathbb{E}\left(\left|\sum_{s\in S}\frac{(Y_{i,s}^{\upsilon}-m_s^{\upsilon}(X_i))\tilde{K}_{s,i}(X_i)}{\mathbb{E}(\tilde{K}_{s,0}(X_i))}\right|^p|X_i\right)\leq C\sum_{s\in S}\frac{1}{(\phi(h_s))^p}\leq C\left(\sum_{s\in S}\frac{1}{\phi(h_s)}\right)^p.$$

Proof of item *(iii)*. If $i \neq j$ and $\|X_i - X_j\| \leq h_s$, then (6.3) (see page 144), (6.5) (see page 144), (6.31) and assumption **(C.6.12)** lead to

$$\begin{aligned}|Y_{j,s}^{\upsilon}-m_s^{\upsilon}(X_i)|&\leq|Y_{j,s}^{\upsilon}-m_s^{\upsilon}(X_j)|+|m_s^{\upsilon}(X_j)-m_s^{\upsilon}(X_i)|\\&=|\epsilon_j|\mathbb{I}_{\{\tilde{\Psi}(X_j)\in\mathcal{E}_s^{\upsilon}\}}+|m(X_j)-m(X_i)|\mathbb{I}_{\{\tilde{\Psi}(X_j)\in\mathcal{E}_s^{\upsilon}\}}\leq(|\epsilon_j|+Ch_s^{\beta})\mathbb{I}_{\{\tilde{\Psi}(X_j)\in\mathcal{E}_s^{\upsilon}\}}.\end{aligned}\tag{6.32}$$

Using (6.32), **(C.6.13)**, **(C.6.16)** and the fact that $\mathcal{E}_{s_1}^{\upsilon} \cap \mathcal{E}_{s_2}^{\upsilon} = \emptyset$ for all $s_1 \neq s_2$, it can be shown that

$$\mathbb{E}\left(\left|\sum_{s\in S}\frac{(Y_{j,s}^{\upsilon}-m_s^{\upsilon}(X_i))\tilde{K}_{s,j}(X_i)}{\mathbb{E}(\tilde{K}_{s,0}(X_i))}\right|^p|X_i\right)\leq\mathbb{E}\left((|\epsilon_j|+C)^p\left(\sum_{s\in S}\frac{\mathbb{I}_{\{\tilde{\Psi}(X_j)\in\mathcal{E}_s^{\upsilon}\}}\tilde{K}_{s,j}(X_i)}{\mathbb{E}(\tilde{K}_{s,0}(X_i))}\right)^p|X_i\right)$$
$$\leq\mathbb{E}\left(\mathbb{E}((|\epsilon_j|+C)^p|X_j)\left(\sum_{s\in S}\frac{\mathbb{I}_{\{\tilde{\Psi}(X_j)\in\mathcal{E}_s^{\upsilon}\}}\tilde{K}_{s,j}(X_i)}{\mathbb{E}(\tilde{K}_{s,0}(X_i))}\right)^p|X_i\right)$$
$$\leq C\mathbb{E}\left(\left(\sum_{s\in S}\frac{\mathbb{I}_{\{\tilde{\Psi}(X_j)\in\mathcal{E}_s^{\upsilon}\}}\tilde{K}_{s,j}(X_i)}{\mathbb{E}(\tilde{K}_{s,0}(X_i))}\right)^p|X_i\right)\leq C\mathbb{E}\left(\sum_{s\in S}\frac{\tilde{K}_{s,j}^p(X_i)}{(\mathbb{E}(\tilde{K}_{s,0}(X_i)))^p}|X_i\right)$$
$$=C\sum_{s\in S}\frac{\mathbb{E}(\tilde{K}_{s,j}^p(X_i)|X_i)}{(\mathbb{E}(\tilde{K}_{s,0}(X_i)))^p}=C\sum_{s\in S}\frac{\mathbb{E}(\tilde{K}_{s,j}^p(X_i)|X_i)}{(\mathbb{E}(\mathbb{E}(\tilde{K}_{s,0}(X_i)|X_i)))^p}.$$

Hence, using Lemma 6.6.8 (see page 176),

$$\mathbb{E}\left(\left|\sum_{s\in S}\frac{(Y_{j,s}^{\upsilon}-m_s^{\upsilon}(X_i))\tilde{K}_{s,j}(X_i)}{\mathbb{E}(\tilde{K}_{s,0}(X_i))}\right|^p|X_i\right)\leq C\sum_{s\in S}\frac{\phi(h_s)}{(\phi(h_s))^p}=C\sum_{s\in S}\frac{1}{(\phi(h_s))^{p-1}}\leq C\left(\sum_{s\in S}\frac{1}{\phi(h_s)}\right)^{p-1}.$$

Proof of item *(iv)*. Using the reasonings presented in (6.32), (6.5) (see page 144), the fact that $\mathcal{E}_{s_1}^{\upsilon} \cap \mathcal{E}_{s_2}^{\upsilon} = \emptyset$ for all $s_1 \neq s_2$, and **(C.6.14)**, then

$$\begin{aligned}
&\mathbb{E}\left(\left(\sum_{s\in S}\frac{(Y_{j,s}^{\upsilon}-m_s^{\upsilon}(X_i))\tilde{K}_{s,j}(X_i)}{\mathbb{E}(\tilde{K}_{s,0}(X_i))}\right)^2|X_i\right)\\
&=\mathbb{E}\left(\sum_{s\in S}\frac{(\epsilon_j+(m(X_j)-m(X_i)))^2\mathbb{I}_{\{\tilde{\Psi}(X_i)\in\mathcal{E}_s^{\upsilon}\}}\tilde{K}_{s,j}^2(X_i)}{(\mathbb{E}(\tilde{K}_{s,0}(X_i)))^2}|X_i\right)\\
&=\sum_{s\in S}\frac{\mathbb{E}(\mathbb{E}((\epsilon_j+(m(X_j)-m(X_i)))^2|X_j)\tilde{K}_{s,j}^2(X_i)|X_i)\mathbb{I}_{\{\tilde{\Psi}(X_i)\in\mathcal{E}_s^{\upsilon}\}}}{(\mathbb{E}(\tilde{K}_{s,0}(X_i)))^2}\\
&=\sum_{s\in S}\frac{\mathbb{E}((\mathbb{E}(\epsilon_j^2|X_j)+(m(X_j)-m(X_i))^2)\tilde{K}_{s,j}^2(X_i)|X_i)\mathbb{I}_{\{\tilde{\Psi}(X_i)\in\mathcal{E}_s^{\upsilon}\}}}{(\mathbb{E}(\tilde{K}_{s,0}(X_i)))^2}\\
&\geq C\sum_{s\in S}\frac{\mathbb{E}(\tilde{K}_{s,j}^2(X_i)|X_i)\mathbb{I}_{\{\tilde{\Psi}(X_i)\in\mathcal{E}_s^{\upsilon}\}}}{(\mathbb{E}(\mathbb{E}(\tilde{K}_{s,0}(X_i)|X_i)))^2}.
\end{aligned}$$

Therefore, using Lemma 6.6.8

$$\mathbb{E}\left(\left(\sum_{s\in S}\frac{(Y_{j,s}^{\upsilon}-m_s^{\upsilon}(X_i))\tilde{K}_{s,j}(X_i)}{\mathbb{E}(\tilde{K}_{s,0}(X_i))}\right)^2|X_i\right)\geq C\sum_{s\in S}\frac{\mathbb{I}_{\{\tilde{\Psi}(X_i)\in\mathcal{E}_s^{\upsilon}\}}}{\phi(h_s)}.$$

■

Formulation and proof of Lemma 6.6.8

Lemma 6.6.8. *Under* **(C.6.1)**, **(C.6.10)** *and* **(C.6.15)**, *for all* $\gamma > 0$ *and for all* $i \neq j$, *there exist* $c_{15,\gamma}, c_{16,\gamma} > 0$ *such that*

$$c_{15,\gamma}\phi(h_s) \leq \mathbb{E}(\tilde{K}_{s,j}^{\gamma}(X_i)|X_i) \leq c_{16,\gamma}\phi(h_s) \quad a.s., \quad \forall s \in S.$$

Proof. Note that **(C.6.15)** ensures that there exist $c, c' > 0$ such that

$$c\mathbb{I}_{\{t\in[0,1]\}} \leq K(t) \leq c'\mathbb{I}_{\{t\in[0,1]\}}, \quad \forall t \in [0,1]. \tag{6.33}$$

For all $\gamma > 0$, using (6.33) with $t = h_s^{-1}\|X_j - X_i\|$, one gets

$$c^{\gamma}\mathbb{I}_{\{\|X_j-X_i\|\leq h_s\}} \leq \tilde{K}_{s,j}^{\gamma}(X_i) \leq (c')^{\gamma}\mathbb{I}_{\{\|X_j-X_i\|\leq h_s\}}.$$

Therefore, applying the conditional expectation, one has

$$c^{\gamma}\mathbb{P}(\|X_j - X_i\| \leq h_s|X_i) \leq \mathbb{E}(\tilde{K}_{s,j}^{\gamma}(X_i)|X_i) \leq (c')^{\gamma}\mathbb{P}(\|X_j - X_i\| \leq h_s|X_i),$$

and, consequently,

$$c^{\gamma}c_1\phi(h_s) \leq \mathbb{E}(\tilde{K}_{s,j}^{\gamma}(X_i)|X_i) \leq (c')^{\gamma}c_2\phi(h_s),$$

since **(C.6.10)** holds. Hence, the proof is finished by taking $c_{15,\gamma} = c^{\gamma}c_1$ and $c_{16,\gamma} = (c')^{\gamma}c_2$. ■

Formulation and proof of Lemma 6.6.9

Recall that C will denote a generic positive constant which may take on different values even in the same formula.

Lemma 6.6.9. *Under hypotheses of Theorem 6.2.6,*

(i) $\forall p = 1, 2, \ldots, \forall q = 2, \ldots, 2p,$

$$\left| \mathbb{E}\left(\prod_{i=1}^{q} \prod_{j=1}^{q} (\delta_\lambda(X_i, X_j, Y_j))^{a_{ij}} \right) \right| \leq c_{17,p} \left(\sum_{s \in S} \frac{1}{\phi(h_s)} \right)^{p-q/2},$$

where $a_{ij} \in \{0, \ldots, p\}$, $\sum_{i=1}^{q} \sum_{j=1}^{q} a_{ij} = p$, *and, for each* $i \in \{1, \ldots, q\}$, *there exists* $j \neq i$ *such that either* $a_{ij} \neq 0$ *or* $a_{ji} \neq 0$,

(ii) $\left| \mathbb{E}\left(\left(\mathbb{E}\left(\delta_\lambda(X_3, X_1, Y_1) \delta_\lambda(X_3, X_2, Y_2) | X_1, X_2 \right) \right)^2 \right) \right| \leq c_{18} \sum_{s \in S} \frac{1}{\phi(h_s)}$,

(iii) $\left| \mathbb{E}\left(\delta_\lambda(X_3, X_1, Y_1) \delta_\lambda(X_3, X_2, Y_2) \right) \right| \leq c_{19}$,

(iv) $\mathbb{E}\left(\left(\delta_\lambda(X_1, X_2, Y_2) \right)^2 \right) \geq c_{20} \sum_{s \in S} \frac{1}{\phi(h_s)}$,

(v) $\forall p = 1, 2, \ldots, \mathbb{E}\left(\left(\mathbb{E}\left(\delta_\lambda(X_1, X_2, Y_2) | X_1 \right) \right)^{2p} \right) \leq c_{21,p}$,

(vi) $\forall p = 1, 2, \ldots, \mathbb{E}\left(\left(\delta_\lambda(X_1, X_1, Y_1) \right)^{2p} \right) \leq c_{22,p} \left(\sum_{s \in S} \frac{1}{\phi(h_s)} \right)^{2p}$,

where the operator δ_λ *is defined in* (6.23) *(see page 169).*

Proof. This lemma is analogous to Lemma 6 by Ait-Saïdi et al. (2008).

Proof of item *(i)*. By Jensen's inequality, (6.32) (see page 175), **(C.6.13)** and the fact that $\mathcal{E}_{s_1}^{v} \cap \mathcal{E}_{s_2}^{v} = \emptyset$ for all $s_1 \neq s_2$, it can be seen that

$$\begin{aligned}
&\left| \mathbb{E}\left(\prod_{i=1}^{q} \prod_{j=1}^{q} (\delta_\lambda(X_i, X_j, Y_j))^{a_{ij}} \right) \right| \leq \mathbb{E}\left(\prod_{i=1}^{q} \prod_{j=1}^{q} |\delta_\lambda(X_i, X_j, Y_j)|^{a_{ij}} \right) \\
&\leq \mathbb{E}\left(\prod_{i=1}^{q} \prod_{j=1}^{q} \left(\sum_{s \in S} \frac{|m_s^v(X_i) - Y_{j,s}^v| \tilde{K}_{s,j}(X_i)}{\mathbb{E}(\tilde{K}_{s,0}(X_i))} \right)^{a_{ij}} \right) \\
&\leq \mathbb{E}\left(\prod_{i=1}^{q} \prod_{j=1}^{q} \left(\sum_{s \in S} \frac{(|\epsilon_j| + C) \mathbb{I}_{\{\tilde{\Psi}(X_j) \in \mathcal{E}_s^v\}} \tilde{K}_{s,j}(X_i)}{\mathbb{E}(\tilde{K}_{s,0}(X_i))} \right)^{a_{ij}} \right) \\
&\leq C\mathbb{E}\left(\prod_{i=1}^{q} \prod_{j=1}^{q} \left(\sum_{s \in S} \frac{\mathbb{I}_{\{\tilde{\Psi}(X_j) \in \mathcal{E}_s^v\}} \tilde{K}_{s,j}(X_i)}{\mathbb{E}(\tilde{K}_{s,0}(X_i))} \right)^{a_{ij}} \right) \leq C\mathbb{E}\left(\prod_{i=1}^{q} \prod_{j=1}^{q} \left(\sum_{s \in S} \frac{\mathbb{I}_{\{\tilde{\Psi}(X_j) \in \mathcal{E}_s^v\}} \tilde{K}_{s,j}^{a_{ij}}(X_i)}{(\mathbb{E}(\tilde{K}_{s,0}(X_i)))^{a_{ij}}} \right) \right) \\
&\leq C\mathbb{E}\left(\sum_{s \in S} \left(\prod_{i=1}^{q} \prod_{j=1}^{q} \frac{\mathbb{I}_{\{\tilde{\Psi}(X_j) \in \mathcal{E}_s^v\}} \tilde{K}_{s,j}^{a_{ij}}(X_i)}{(\mathbb{E}(\tilde{K}_{s,0}(X_i)))^{a_{ij}}} \right) \right) \leq C \sum_{s \in S} \mathbb{E}\left(\prod_{i=1}^{q} \prod_{j=1}^{q} \frac{\tilde{K}_{s,j}^{a_{ij}}(X_i)}{(\mathbb{E}(\mathbb{E}(\tilde{K}_{s,0}(X_i)) | X_i))^{a_{ij}}} \right).
\end{aligned} \tag{6.34}$$

On the other hand, Lemma 6.6.8 (see page 176) guarantees that

$$\mathbb{E}\left(\prod_{i=1}^{q} \prod_{j=1}^{q} \frac{\tilde{K}_{s,j}^{a_{ij}}(X_i)}{(\mathbb{E}(\tilde{K}_{s,0}(X_i)))^{a_{ij}}} \right) \leq C \frac{1}{(\phi(h_s))^p} \mathbb{E}\left(\prod_{i=1}^{q} \prod_{j=1}^{q} \tilde{K}_{s,j}^{a_{ij}}(X_i) \right).$$

Besides, the restrictions on the definition of the pairs (i,j) and a_{ij} imply that there are $q/2$ separated pairs (i,j) with $a_{ij}\neq 0$. This fact and Lemma 6.6.8 (see page 176) allow to deduce

$$\mathbb{E}\left(\prod_{i=1}^{q}\prod_{j=1}^{q}\frac{\tilde{K}_{s,j}^{a_{ij}}(X_i)}{(\mathbb{E}(\tilde{K}_{s,0}(X_i)))^{a_{ij}}}\right)\leq C\frac{1}{(\phi(h_s))^{p}}(\phi(h_s))^{q/2}=C\frac{1}{(\phi(h_s))^{p-q/2}}. \tag{6.35}$$

Consequently, by (6.34) and (6.35), one has,

$$\left|\mathbb{E}\left(\prod_{i=1}^{q}\prod_{j=1}^{q}(\delta_\lambda(X_i,X_j,Y_j))^{a_{ij}}\right)\right|\leq C\sum_{s\in S}\frac{1}{(\phi(h_s))^{p-q/2}}\leq C\left(\sum_{s\in S}\frac{1}{\phi(h_s)}\right)^{p-q/2}.$$

Proof of item *(ii)*. It can be shown that

$$\begin{aligned}&\left|\mathbb{E}\left(\left(\mathbb{E}\left(\delta_\lambda(X_3,X_1,Y_1)\delta_\lambda(X_3,X_2,Y_2)|X_1,X_2\right)\right)^2\right)\right|\\&\leq\mathbb{E}\left(\mathbb{E}\left(|\delta_\lambda(X_3,X_1,Y_1)\delta_\lambda(X_3,X_2,Y_2)\delta_\lambda(X_4,X_1,Y_1)\delta_\lambda(X_4,X_2,Y_2)||X_1,X_2\right)\right)\\&=\mathbb{E}\left(|\delta_\lambda(X_3,X_1,Y_1)\delta_\lambda(X_3,X_2,Y_2)\delta_\lambda(X_4,X_1,Y_1)\delta_\lambda(X_4,X_2,Y_2)|\right).\end{aligned}$$

Hence, due to (6.32) (see page 175), the fact that $\mathcal{E}_{s_1}^{v}\cap\mathcal{E}_{s_2}^{v}=\emptyset$ for all $s_1\neq s_2$, and **(C.6.13)**, it can be found that

$$\begin{aligned}&\left|\mathbb{E}\left(\left(\mathbb{E}\left(\delta_\lambda(X_3,X_1,Y_1)\delta_\lambda(X_3,X_2,Y_2)|X_1,X_2\right)\right)^2\right)\right|\\&\leq\mathbb{E}\Bigg(\sum_{s_1\in S}\sum_{s_2\in S}\sum_{s_3\in S}\sum_{s_4\in S}\Bigg(|m_{s_1}^{v}(X_3)-Y_{1,s_1}^{v}||m_{s_2}^{v}(X_3)-Y_{2,s_2}^{v}||m_{s_3}^{v}(X_4)-Y_{1,s_3}^{v}||m_{s_4}^{v}(X_4)-Y_{2,s_4}^{v}|\\&\qquad\cdot\frac{\tilde{K}_{s_1,1}(X_3)\tilde{K}_{s_2,2}(X_3)\tilde{K}_{s_3,1}(X_4)\tilde{K}_{s_4,2}(X_4)}{\mathbb{E}(\tilde{K}_{s_1,0}(X_3))\mathbb{E}(\tilde{K}_{s_2,0}(X_3))\mathbb{E}(\tilde{K}_{s_3,0}(X_4))\mathbb{E}(\tilde{K}_{s_4,0}(X_4))}\Bigg)\Bigg)\\&\leq\mathbb{E}\Bigg(\sum_{s_1\in S}\sum_{s_2\in S}\sum_{s_3\in S}\sum_{s_4\in S}\Bigg((|\epsilon_1|+C)^2(|\epsilon_2|+C)^2\mathbb{I}_{\{\tilde{\Psi}(X_3)\in\mathcal{E}_{s_1}^{v}\}}\mathbb{I}_{\{\tilde{\Psi}(X_3)\in\mathcal{E}_{s_2}^{v}\}}\mathbb{I}_{\{\tilde{\Psi}(X_4)\in\mathcal{E}_{s_3}^{v}\}}\\&\qquad\cdot\mathbb{I}_{\{\tilde{\Psi}(X_4)\in\mathcal{E}_{s_4}^{v}\}}\frac{\tilde{K}_{s_1,1}(X_3)\tilde{K}_{s_2,2}(X_3)\tilde{K}_{s_3,1}(X_4)\tilde{K}_{s_4,2}(X_4)}{\mathbb{E}(\tilde{K}_{s_1,0}(X_3))\mathbb{E}(\tilde{K}_{s_2,0}(X_3))\mathbb{E}(\tilde{K}_{s_3,0}(X_4))\mathbb{E}(\tilde{K}_{s_4,0}(X_4))}\Bigg)\Bigg)\\&\leq\mathbb{E}\Bigg(\mathbb{E}((|\epsilon_1|+C)^2(|\epsilon_2|+C)^2|X_1,X_2,X_3,X_4)\sum_{s_1\in S}\sum_{s_3\in S}\Bigg(\mathbb{I}_{\{\tilde{\Psi}(X_3)\in\mathcal{E}_{s_1}^{v}\}}\mathbb{I}_{\{\tilde{\Psi}(X_4)\in\mathcal{E}_{s_3}^{v}\}}\\&\qquad\cdot\frac{\tilde{K}_{s_1,1}(X_3)\tilde{K}_{s_1,2}(X_3)\tilde{K}_{s_3,1}(X_4)\tilde{K}_{s_3,2}(X_4)}{(\mathbb{E}(\tilde{K}_{s_1,0}(X_3)))^2(\mathbb{E}(\tilde{K}_{s_3,0}(X_4)))^2}\Bigg)\Bigg)\\&\leq C\mathbb{E}\left(\sum_{s_1\in S}\sum_{s_3\in S}\left(\mathbb{I}_{\{\tilde{\Psi}(X_3)\in\mathcal{E}_{s_1}^{v}\}}\mathbb{I}_{\{\tilde{\Psi}(X_4)\in\mathcal{E}_{s_3}^{v}\}}\frac{\tilde{K}_{s_1,1}(X_3)\tilde{K}_{s_1,2}(X_3)\tilde{K}_{s_3,1}(X_4)\tilde{K}_{s_3,2}(X_4)}{(\mathbb{E}(\tilde{K}_{s_1,0}(X_3)))^2(\mathbb{E}(\tilde{K}_{s_3,0}(X_4)))^2}\right)\right).\end{aligned}$$

Given that $\tilde{K}_{s_3,2}(X_4) \leq C$ for all $s \in S$ (since **(C.6.15)** holds), $\tilde{K}_{s,i}(X_j) = \tilde{K}_{s,j}(X_i)$, the indicator functions are bounded, and Lemma 6.6.8 (see page 176) can be applied, one gets

$$\begin{aligned}
&\left|\mathbb{E}\left(\left(\mathbb{E}\left(\delta_\lambda(X_3,X_1,Y_1)\delta_\lambda(X_3,X_2,Y_2)|X_1,X_2\right)\right)^2\right)\right| \\
&\leq C\sum_{s_1\in S}\sum_{s_3\in S}\frac{\mathbb{E}(\tilde{K}_{s_1,1}(X_3)\tilde{K}_{s_1,2}(X_3)\tilde{K}_{s_3,1}(X_4))}{(\mathbb{E}(\mathbb{E}(\tilde{K}_{s_1,0}(X_3)|X_3)))^2(\mathbb{E}(\mathbb{E}(\tilde{K}_{s_3,0}(X_4)|X_4)))^2} \\
&\leq C\sum_{s_1\in S}\sum_{s_3\in S}\frac{\mathbb{E}(\tilde{K}_{s_1,1}(X_3)\mathbb{E}(\tilde{K}_{s_1,2}(X_3)|X_1,X_3,X_4)\tilde{K}_{s_3,1}(X_4))}{(\phi(h_{s_1}))^2(\phi(h_{s_3}))^2} \\
&\leq C\sum_{s_1\in S}\sum_{s_3\in S}\frac{\mathbb{E}(\tilde{K}_{s_1,1}(X_3)\mathbb{E}(\tilde{K}_{s_3,1}(X_4)|X_1,X_3))}{\phi(h_{s_1})(\phi(h_{s_3}))^2} \leq C\sum_{s_1\in S}\sum_{s_3\in S}\frac{\mathbb{E}(\tilde{K}_{s_1,1}(X_3))}{\phi(h_{s_1})\phi(h_{s_3})} \leq C\sum_{s\in S}\frac{1}{\phi(h_s)}.
\end{aligned}$$

Proof of item *(iii)*. Regarding to this item, Jensen's inequality, (6.32) (see page 175), (6.5) (see page 144), the fact that $\mathcal{E}^v_{s_1}\cap\mathcal{E}^v_{s_2}=\emptyset$ for all $s_1\neq s_2$, and **(C.6.13)** imply

$$\begin{aligned}
&|\mathbb{E}\left(\delta_\lambda(X_3,X_1,Y_1)\delta_\lambda(X_3,X_2,Y_2)\right)| \leq \mathbb{E}\left(|\delta_\lambda(X_3,X_1,Y_1)\delta_\lambda(X_3,X_2,Y_2)|\right) \\
&\leq \mathbb{E}\left(\sum_{s_1\in S}\sum_{s_2\in S}\frac{|m^v_{s_1}(X_3)-Y^v_{1,s_1}||m^v_{s_2}(X_3)-Y^v_{2,s_2}|\tilde{K}_{s_1,1}(X_3)\tilde{K}_{s_2,2}(X_3)}{\mathbb{E}(\tilde{K}_{s_1,0}(X_3))\mathbb{E}(\tilde{K}_{s_2,0}(X_3))}\right) \\
&\leq \mathbb{E}\left(\sum_{s\in S}\frac{(|\epsilon_1|+C)(|\epsilon_2|+C)\mathbb{I}_{\{\tilde{\Psi}(X_3)\in\mathcal{E}^v_s\}}\tilde{K}_{s,1}(X_3)\tilde{K}_{s,2}(X_3)}{(\mathbb{E}(\tilde{K}_{s,0}(X_3)))^2}\right) \\
&= \sum_{s\in S}\frac{\mathbb{E}(\mathbb{E}((|\epsilon_1|+C)(|\epsilon_2|+C)|X_1,X_2,X_3)\mathbb{I}_{\{\tilde{\Psi}(X_3)\in\mathcal{E}^v_s\}}\tilde{K}_{s,1}(X_3)\tilde{K}_{s,2}(X_3))}{(\mathbb{E}(\tilde{K}_{s,0}(X_3)))^2} \\
&\leq C\sum_{s\in S}\frac{\mathbb{E}(\mathbb{I}_{\{\tilde{\Psi}(X_3)\in\mathcal{E}^v_s\}}\tilde{K}_{s,1}(X_3)\tilde{K}_{s,2}(X_3))}{(\mathbb{E}(\tilde{K}_{s,0}(X_3)))^2} = C\sum_{s\in S}\frac{\mathbb{E}(\mathbb{I}_{\{\tilde{\Psi}(X_3)\in\mathcal{E}^v_s\}}\mathbb{E}(\tilde{K}_{s,1}(X_3)|X_3)\mathbb{E}(\tilde{K}_{s,2}(X_3)|X_3))}{(\mathbb{E}(\mathbb{E}(\tilde{K}_{s,0}(X_3)|X_3)))^2}.
\end{aligned}$$

Hence, the application of Lemma 6.6.8 (see page 176) leads to

$$|\mathbb{E}\left(\delta_\lambda(X_3,X_1,Y_1)\delta_\lambda(X_3,X_2,Y_2)\right)| \leq C\sum_{s\in S}\mathbb{P}(\tilde{\Psi}(X_3)\in\mathcal{E}^v_s) \leq CN_{\mathcal{E}} \leq C.$$

Proof of item *(iv)*. This item comes from Lemma 6.6.7 (see page 174) and **(C.6.3)** as follows

$$\begin{aligned}
&\mathbb{E}\left(\left(\delta_\lambda(X_1,X_2,Y_2)\right)^2\right) = \mathbb{E}\left(\left(\sum_{s\in S}\frac{(m^v_s(X_1)-Y^v_{2,s})\tilde{K}_{s,2}(X_1)}{\mathbb{E}(\tilde{K}_{s,0}(X_1))}\right)^2\right) \\
&= \mathbb{E}\left(\mathbb{E}\left(\left(\sum_{s\in S}\frac{(m^v_s(X_1)-Y^v_{2,s})\tilde{K}_{s,2}(X_1)}{\mathbb{E}(\tilde{K}_{s,0}(X_1))}\right)^2|X_1\right)\right) \geq C\sum_{s\in S}\frac{\mathbb{P}(\tilde{\Psi}(X_1)\in\mathcal{E}^v_s)}{\phi(h_s)} \geq C\sum_{s\in S}\frac{1}{\phi(h_s)}.
\end{aligned}$$

Proof of item *(v)*. By Jensen's inequality and Lemma 6.6.7 (see page 174),

$$|\mathbb{E}\left(\delta_\lambda(X_1,X_2,Y_2)|X_1\right)| \leq \mathbb{E}\left(|\delta_\lambda(X_1,X_2,Y_2)|\,|X_1\right) = \mathbb{E}\left(\left|\sum_{s\in S}\frac{(m^v_s(X_1)-Y^v_{2,s})\tilde{K}_{s,2}(X_1)}{\mathbb{E}(\tilde{K}_{s,0}(X_1))}\right||X_1\right) \leq C.$$

Thus, $\mathbb{E}\left(\left(\mathbb{E}\left(\delta_\lambda(X_1,X_2,Y_2)|X_1\right)\right)^{2p}\right) \leq C$.

Proof of item *(vi)*. This item is a direct consequence of Lemma 6.6.7 (see page 174) given that

$$\begin{aligned}\mathbb{E}\left((\delta_\lambda(X_1,X_1,Y_1))^{2p}\right) &= \mathbb{E}\left(\left(\sum_{s\in S}\frac{(m_s^v(X_1)-Y_{1,s}^v)K(0)}{\mathbb{E}(\tilde{K}_{s,0}(X_1))}\right)^{2p}\right)\\ &\leq C\mathbb{E}\left(\left(\sum_{s\in S}\frac{m_s^v(X_1)-Y_{1,s}^v}{\mathbb{E}(\tilde{K}_{s,0}(X_1))}\right)^{2p}\right)\leq C\left(\sum_{s\in S}\frac{1}{\phi(h_s)}\right)^{2p}.\end{aligned}$$

■

Conclusions and further research

The FDA problems which have been analysed throughout the present manuscript have been focused on the regression model with functional covariate and scalar response. The three main aims of the thesis have been the following:

- The proposal of new estimators for functional linear regression model with scalar response, based on FPCA and presmoothing techniques, which exhibit a good behaviour in terms of consistency and efficiency.
- The development of a bootstrap approach for functional linear regression model that allows obtaining pointwise confidence intervals for the FPCA-type estimates, and to test different hypotheses related to the model parameter, such as the lack of dependence and the equality of linear models.
- The construction of an exploratory tool for nonparametric functional regression model which analyses the existence of hidden patterns in functional data via a thresholding procedure.

The two first items were studied in a parametric regression context, specifically, the functional linear model, whereas the third one was developed for a nonparametric functional model. Next, final conclusions, and some open questions, related with this three research issues are summarized.

Presmoothing in functional linear regression

In Chapter 3, four different FPCA–type estimators for the linear model parameter θ were introduced, all of them based on presmoothing techniques:

(i) Presmoothing via covariance structure: $\hat{\theta}_{k_n}^{\alpha_n} = \sum_{j=1}^{k_n} \frac{\Delta_n \hat{v}_j}{\hat{\lambda}_j + \alpha_n} \hat{v}_j$.

(ii) Presmoothing via response variable: $\hat{\theta}_{k_n}^{h_n} = \sum_{j=1}^{k_n} \frac{\Delta_n^{h_n} \hat{v}_j}{\hat{\lambda}_j} \hat{v}_j$.

(iii) Using Pezzulli and Silverman' presmoothed FPCA: $\hat{\theta}_{k_n}^{PS,\,\alpha_n} = \sum_{j=1}^{k_n} \frac{\Delta_n \hat{v}_j^{\alpha_n,1}}{\hat{\lambda}_j^{\alpha_n,1}} \hat{v}_j^{\alpha_n,1}$.

(iv) Using Silverman's presmoothed FPCA: $\hat{\theta}_{k_n}^{S,\,\alpha_n} = \sum_{j=1}^{k_n} \frac{\Delta_n \hat{v}_j^{\alpha_n,2}}{\hat{\lambda}_j^{\alpha_n,2}} \hat{v}_j^{\alpha_n,2}$.

The first proposal, $\hat{\theta}_{k_n}^{\alpha_n}$, can be seen as an extension of the ordinary multivariate ridge regression estimator to general Hilbert spaces. The key idea is to avoid ill–conditioned problems by perturbing slightly the eigenvalues of the second moment operator. The consistency and expressions for conditional mean square errors for prediction and estimation were obtained for $\hat{\theta}_{k_n}^{\alpha_n}$. Using the conditional estimation error, it can be seen, from a theoretical point of view, that this presmoothed estimate gets improvement over the FPCA estimate, especially when the model noise is large and/or the sample size is small.

As far as $\hat{\theta}_{k_n}^{h_n}$ is concerned, its consistency was stated and the expressions for conditional mean square errors for prediction and estimation were computed. In the simulation study and the real data applications, $\hat{\theta}_{k_n}^{h_n}$ was replaced by

$$\hat{\theta}_{k_n}^{\text{k–NN}} = \sum_{j=1}^{k_n} \frac{\Delta_n^{\text{k–NN}} \hat{v}_j}{\hat{\lambda}_j} \hat{v}_j,$$

because $\hat{\theta}_{k_n}^{\text{k–NN}}$ provided more reduced errors than $\hat{\theta}_{k_n}^{h_n}$ in some previous simulations. Due to this fact, it could be interesting to find the conditional error expressions for $\hat{\theta}_{k_n}^{\text{k–NN}}$ in order to compare them from a theoretical point of view. On the other hand, the efficiency of $\hat{\theta}_{k_n}^{\text{k–NN}}$ could not be stated, so this is an open question too.

Regarding to $\hat{\theta}_{k_n}^{PS,\,\alpha_n}$ and $\hat{\theta}_{k_n}^{S,\,\alpha_n}$, only their conditional error expressions were computed, so the study of their consistency properties is a pending issue. Furthermore, these estimators were ommited in the simulation study. As future work, these estimators could be included in a numerical study in order to compare them with the other proposed presmoothed estimators and check if they are competitive alternatives.

Bootstrap techniques in functional linear regression

Naive and wild bootstrap algorithms were used to build pointwise confidence intervals for the regression operator $m(\cdot)$ in Chapter 4 as follows

$$\text{CI}_{x,\alpha}^{*} = \left[\langle \hat{\theta}_{k_n}, x \rangle - q_{1-\alpha/2}^{*}(x), \langle \hat{\theta}_{k_n}, x \rangle - q_{\alpha/2}^{*}(x) \right].$$

Apart from proving the asymptotic validity of this bootstrap approach from a theoretical point of view, the simulation study allowed to confirm the good behaviour of bootstrap confidence intervals with respect to the asymptotic confidence intervals

$$\text{CI}_{x,\alpha}^{asy} = \left[\langle \hat{\theta}_{k_n}, x \rangle - z_{1-\alpha/2} \frac{\hat{t}_{n,x}^{k_n} \hat{\sigma}}{\sqrt{n}}, \langle \hat{\theta}_{k_n}, x \rangle + z_{1-\alpha/2} \frac{\hat{t}_{n,x}^{k_n} \hat{\sigma}}{\sqrt{n}} \right],$$

which are based on weak convergence results. In particular, it was noted that an adequate selection of the pilot parameter k_n^d involved in the bootstrap procedure makes empirical coverage rates of bootstrap intervals be closer to nominal level α than the coverage rates of asymptotic confidence intervals. Unfortunately, a data–driven methodology which automatically selects this *optimal* value of k_n^d is not yet available, so further research is required in order to solve this issue.

Chapter 5 was devoted to introduce bootstrap algorithms in order to calibrate the distribution of test statistics for testing the lack of dependence ($H_0 : \theta = 0$ versus $H_1 : \theta \neq 0$) or the equality of two linear models ($H_0 : \theta_1 = \theta_2$ versus $H_1 : \theta_1 \neq \theta_2$). The chapter only presented the asymptotic theory for some of the proposed bootstrap methods, so a future research line could be to try to develope the asymptotic theory for the remaining bootstrap proposals.

From a practical point of view, the simulation study showed that bootstrap methods are competitive alternatives to tests based on asymptotic distributions, since they often give test sizes closer to the nominal ones. In addition, it must be recalled that the test statistics which included a consistent estimation of the error variance σ^2 obtained higher empirical power than the test statistics which did not take it into account. Due to this fact, it could be interesting to study in future works what happens if test statistics which does not include $\hat{\sigma}^2$ are studentized, that is, to analyse if their studentized versions improve the obtained results in the simulation study in terms of empirical power.

It is important to highlight that for all the studied statistic test in which a k/k_n parameter is involved, the selection of this parameter seems to be a key issue, and it is another research line that should be taken into account in the future. Besides of the optimal k/k_n selection, other issues related to these hypotheses tests require further research, such as their extension to functional linear models

with functional response. In addition, for the test of lack of dependence, it would be interesting to combine it with the functional ANOVA test (see Cuevas et al. (2004), and González-Rodríguez et al. (2012)) in order to develop an ANCOVA test in this context, whereas, for the test of equality, the extension of the results to other test statistics proposed by Horváth et al. (2009), which are applicable to more general situations (such as different covariance structures, different sample sizes,...), is also a pending problem.

Thresholding in nonparametric functional regression

A graphical tool that detects threshold structures in the context of the functional linear model with scalar response was presented in this chapter. This methodology allows to find hidden patterns related to both the functional covariate X and the scalar response Y. For this purpose, an adequate threshold function must be chosen by the user according to the structure one wants to detect. Indeed, the real data applications showed that the methodology allows to detect some kind of hidden structures, although the effectiveness of the procedure depends on the choice of the threshold function. Nevertheless, how to select this threshold function in a data–driven and efficient way is still an open question that will require further research.

A cross–validation method was proposed in order to estimate the threshold value, when it exists, and its optimality with respect to MISE was studied for a particular scenario. However, it would be interesting to develop more general optimality results, which may include other cases, for instance, threshold functions which depends on the response Y.

In the simulation study, it was found that the threshold estimators and the standard nonparametric estimator obtain similar results in terms of the mean square prediction error, whereas the mean square prediction error can be reduced if each subsample detected by the threshold technique is studied separately. As future work, it could be interesting to try to prove theoretically this performance which was observed in practice.

Finally, recall that, as was commented in the preface, all the new techniques introduced througthout this manuscript were implemented using the statistical software *R* (see further details in R Development Core Team, 2010, or `http://www.r-project.org`). Given that the use of this free software is very widespread among the statistical community, the developed routines will be available online in order to allow any other researcher interested in FDA to use them.

Summary

Nowadays the progress of computational tools (both memory and capacity increasing) allows to create, store and work with large databases. In many cases, the dataset is made up of observations from a finite dimensional distribution, measured over a period of time or recorded at different spatial locations. When the temporal or spatial grid is fine enough, the sample can be considered as an observation of a random variable on a certain functional space. Analysing this kind of data with standard multivariate methods and ignoring its functional feature may fail dramatically (curse of dimensionality, collinearity, valuable information loss, etc.). In these cases, specific statistical techniques are required in order to manage, leak and draw relevant underlying information.

This fact has turned *Functional Data Analysis* (FDA) into one of the most active statistical fields in recent years. From the seminal works in the eighties and the nineties (Grenander, 1981; Dauxois et al., 1982; Ramsay, 1982; Bosq, 1991), FDA gave rise to several books (Bosq, 2000; Ramsay and Silverman, 1997, 2002, 2005; Ferraty and Vieu, 2006b; Ferraty and Romain, 2011; Horváth and Kokoszka, 2012), special issues in high impact factor journals (Davidian et al., 2004; González-Manteiga and Vieu, 2007; Valderrama, 2007; Ferraty, 2010) and international workshops devoted to both methodological and applied developments for functional data (Dabo-Niang and Ferraty, 2008; Ferraty, 2011). Furthermore, some contributions focused on the state of the art of FDA were published during the last years (Rice, 2004; Müller, 2005; González-Manteiga and Vieu, 2011; Delsol et al., 2011a; Cuevas, 2012).

It must be emphasized that functional data come up in a natural way in most scientific fields: econometrics (e.g., daily stock returns or electricity production/demand curves), engineering sciences (e.g., satellite imagery, topographic maps or image recognition), environmetrics and climatology (e.g., meteorological measurements or fluvial flows curves), medicine (e.g., growth curves or genetic data), chemometrics (e.g., spectrometric data), etc. This fact generated a variety of applied FDA works in the literature, such as Ramsay and Silverman (2002), and Ferraty and Vieu (2006a).

Due to the novelty of FDA, there is a wide range of research lines which could be explored: from the extension to the functional context of well–known multidimensional methods, to the creation of new statistical techniques devoted to specific functional data issues. Nevertheless, in spite of the functional nature of the data, the pursued aims are essentially the same as the usual stated objectives for a multivariate dataset analysis. Consequently, the developed methodology has been intended to satisfy similar needs:

- Preprocess the data: registration and feature alignment, smoothing techniques, etc.
- Depict and explore the data, highlighting their most important features: measures of centrality and dispersion, detection of outliers, *Functional Principal Component Analysis* (FPCA), etc.
- Functional data classification.
- Build models to explain the relationship between functional variables: parametric and nonparametric regression models.
- Functional statistical inference: confidence intervals, hypotheses testing, etc.

This manuscript mainly deals with the last two items. As far as the fourth item is concerned, the work has been focused on the functional linear model with scalar response (although some contributions

to nonparametric regression are also included in the last chapter), whereas regarding the fifth item a bootstrap procedure has been developed, which allows to build confidence intervals and calibrate hypotheses tests related to the linear model.

The thesis has been structured in six chapters. For each of them, a brief summary is given below. Then some computational issues and an acknowledgement section are included.

Chapter 1. Introduction to FDA

The first chapter of this thesis is used to fix the notation and give a brief summary of the state of the art on statistical methods for functional data. The main idea in FDA is to suppose that a random variable X is observed in a discrete grid $\{t_l\}_{l=1}^{L}$ with each $t_l \in T$. If the time instants t_l are close enough, one can assume that $\{X(t_l)\}_{l=1}^{L}$ is an observation of a functional random variable

$$X = \{X(t); t \in T\}.$$

Note that T is often a real interval such as $T = [T_0, T_1] \subset \mathbb{R}$, and each observation can be considered as a curve. Nevertheless, FDA covers many other situations where the space T is a more general functional space, as it can be deduced from the following definitions.

Definition 1. A random variable X is called a *functional random variable* if it takes values in an infinite dimensional space or functional space $\mathcal{S}$. An observation x of X is called a *functional datum*.

Definition 2. Given a functional random variable X and $n \in \mathbb{N}^*$, a *functional random sample* of X of lenght n is a set $\{X_i\}_{i=1}^{n}$ of independent and identically distributed (i.i.d.) functional random variables with the same distribution as X. An observation $\{x_i\}_{i=1}^{n}$ of $\{X_i\}_{i=1}^{n}$ is called a *functional dataset*.

Once it has been defined what functional data are, some examples are given in the chapter, which will illustrate the methods proposed in the next chapters.

Throughout the manuscript, it has been assumed that the space where functional variables take values is a real separable Hilbert space $\mathcal{H}$ (although sometimes the functional space has been restricted to the well–known L^2–space to illustrate certain methods). This motivates the introduction of some associated spaces, such as the space of Hilbert–Schmidt operators and the dual space $\mathcal{H}'$, some tensor notation (that is, operators $\otimes_{\mathcal{H}}$ and $\otimes_{\mathcal{H}'}$) and semi–metrics, which are very useful in order to determine the closeness of functional observations.

Then, a general background of existing FDA tools is presented: preprocessing techniques (smothing and registration methods), functional descriptive statistics (measures of position and dispersion), and some key exploratory methods (for instance, *Functional Principal Component Analysis* (FPCA), which will be recalled later to define functional linear regression estimators).

Chapter 2. Functional regression models

This chapter is devoted to functional regression models. At first, a general review of functional regression is presented, and then the efforts are concentrated on models with scalar response and functional covariate. There are two main approaches to discuss this subject: the parametric and the nonparametric approaches. As regards the parametric approach, the most usual parametric model is the *functional linear model with scalar response* given by

$$Y = \langle \theta, X \rangle + \epsilon,$$

where Y is a real random variable, $m(\cdot) = \langle \theta, \cdot \rangle$ is a linear regression operator with $\theta \in \mathcal{H}$ and $\|\theta\|^2 < \infty$, X is a zero–mean random variable valued in $\mathcal{H}$ such that $\mathbb{E}(\|X\|^2) < \infty$, and ϵ is a real random variable satisfying that $\mathbb{E}(\epsilon) = 0$, $\mathrm{Var}(\epsilon) = \sigma^2 < \infty$, and $\mathbb{E}(\epsilon X) = 0$. The two most popular estimators in this situation are introduced in this chapter: estimators based on basis systems, such as the *penalized B–splines estimator*

$$\hat{\theta}_{PS} = \sum_{j=1}^{q+k} \hat{b}_j B_{k,j}$$

(see, for instance, Cardot et al., 2003c), and FPCA–type estimators, such as the standard FPCA estimator

$$\hat{\theta}_{k_n} = \sum_{j=1}^{k_n} \frac{\Delta_n \hat{v}_j}{\hat{\lambda}_j} \hat{v}_j$$

(see Cardot et al., 1999, 2003c, 2007c).

As far as the nonparametric approach is concerned, the functional nonparametric model can be expressed as

$$Y = m(X) + \epsilon,$$

where Y is a real random variable, $m(\cdot)$ is a functional regression operator which satisfies some smoothness restrictions, X is a zero–mean random variable valued in an abstract space $\mathcal{S}$ endowed with a semi–metric $d(\cdot, \cdot)$, and ϵ is a real random variable satisfying that $\mathbb{E}(\epsilon) = 0$, $\mathrm{Var}(\epsilon) = \sigma^2 < \infty$, and $\mathbb{E}(\epsilon X) = 0$. Among the different existing nonparametric methods, the functional version of the multivariate kernel–type estimator defined as

$$\hat{m}_h(x) = \frac{\sum_{i=1}^n Y_i K(h^{-1} d(x, X_i))}{\sum_{i=1}^n K(h^{-1} d(x, X_i))},$$

is analysed in this chapter (see Ferraty and Vieu, 2004, 2006b; Ferraty et al., 2007a).

Chapter 3. Presmoothing in functional linear regression

The chapter is focused on the functional linear model with scalar response, and explanatory variable valued in a functional space. As it was discussed in previous chapter, FPCA has been used to estimate the model functional parameter in recent statistical literature. A modification of this approach by using presmoothing techniques is proposed in this chapter: either presmoothing via covariance structure or presmoothing via response variable. Specifically, four different FPCA–type estimators for the linear model parameter θ were introduced, all of them based on presmoothing techniques:

(i) Presmoothing via covariance structure: $\hat{\theta}_{k_n}^{\alpha_n} = \sum_{j=1}^{k_n} \frac{\Delta_n \hat{v}_j}{\hat{\lambda}_j + \alpha_n} \hat{v}_j$.

(ii) Presmoothing via response variable: $\hat{\theta}_{k_n}^{h_n} = \sum_{j=1}^{k_n} \frac{\Delta_n^{h_n} \hat{v}_j}{\hat{\lambda}_j} \hat{v}_j$.

(iii) Using Pezzulli and Silverman' presmoothed FPCA (Pezzulli and Silverman, 1993):

$$\hat{\theta}_{k_n}^{PS,\,\alpha_n} = \sum_{j=1}^{k_n} \frac{\Delta_n \hat{v}_j^{\alpha_n,1}}{\hat{\lambda}_j^{\alpha_n,1}} \hat{v}_j^{\alpha_n,1}.$$

(iv) Using Silverman's presmoothed FPCA (Silverman, 1996): $\hat{\theta}_{k_n}^{S,\,\alpha_n} = \sum_{j=1}^{k_n} \frac{\Delta_n \hat{v}_j^{\alpha_n,2}}{\hat{\lambda}_j^{\alpha_n,2}} \hat{v}_j^{\alpha_n,2}$.

The first proposal, $\hat{\theta}_{k_n}^{\alpha_n}$, can be seen as an extension of the ordinary multivariate ridge regression estimator to general Hilbert spaces. The key idea is to avoid ill–conditioned problems by perturbing slightly the eigenvalues of the second moment operator. The consistency and expressions for conditional mean square errors for prediction and estimation were obtained for $\hat{\theta}_{k_n}^{\alpha_n}$. Using the conditional

estimation error, it can be seen, from a theoretical point of view, that this presmoothed estimate gets improvement over the FPCA estimate, especially when the model noise is large and/or the sample size is small. As far as $\hat{\theta}_{k_n}^{h_n}$ is concerned, its consistency was stated and the expressions for conditional mean square errors for prediction and estimation were computed. Regarding to $\hat{\theta}_{k_n}^{PS,\,\alpha_n}$ and $\hat{\theta}_{k_n}^{S,\,\alpha_n}$, only their conditional error expressions were computed using some heuristic calculations by means of the standard technique of asymptotic expansions.

The effectiveness of the proposed estimators based on presmoothing via covariance structure and presmoothing via response variable relative to the standard FPCA estimator and the penalized B–splines estimator is also tested by means of simulation studies and real data applications. In both cases, $\hat{\theta}_{k_n}^{h_n}$ was replaced by

$$\hat{\theta}_{k_n}^{\text{k–NN}} = \sum_{j=1}^{k_n} \frac{\Delta_n^{\text{k–NN}} \hat{v}_j}{\hat{\lambda}_j} \hat{v}_j,$$

given that $\hat{\theta}_{k_n}^{\text{k–NN}}$ provided more reduced errors than $\hat{\theta}_{k_n}^{h_n}$ in some previous simulations. The simulations suggest that the presmoothing estimator $\hat{\theta}_{k_n}^{\alpha_n}$ improves the standard FPCA linear estimator, especially when the sample size is small, whereas $\hat{\theta}_{k_n}^{\text{k–NN}}$ do not significantly reduce the conditional errors of $\hat{\theta}_{k_n}$. The choice of the parameters of the proposed presmoothed estimators is of course a key point. A general practical guideline is to choose these parameters by cross–validation techniques.

Chapter 4. Bootstrap in functional linear regression

Dealing with the functional linear model with functional explanatory variable and scalar response, and as commented previously, one of the most popular methods for parameter model estimation is based on FPCA. Weak convergence for a wide class of FPCA–type estimators has recently been proved and, as a result, the next asymptotic confidence intervals for the linear regression operator $m(\cdot)$ can be obtained for a fixed confidence level $\alpha \in (0,1)$

$$CI_{x,\alpha}^{asy} = \left[\langle \hat{\theta}_{k_n}, x \rangle - z_{1-\alpha/2} \frac{\hat{t}_{n,x}^{k_n} \hat{\sigma}}{\sqrt{n}}, \langle \hat{\theta}_{k_n}, x \rangle + z_{1-\alpha/2} \frac{\hat{t}_{n,x}^{k_n} \hat{\sigma}}{\sqrt{n}} \right],$$

where z_α is the quantile of order α from a $\mathcal{N}(0,1)$ distribution. In this chapter, an alternative approach in order to compute pointwise confidence intervals by means of a bootstrap procedure is proposed, obtaining also its asymptotic validity (that is, the conditional distribution of the estimator can be approximated by the bootstrap distribution). In particular, algorithms for naive and wild bootstrap are developed and the bootstrap intervals can be constructed as

$$CI_{x,\alpha}^{*} = \left[\langle \hat{\theta}_{k_n}, x \rangle - q_{1-\alpha/2}^{*}(x), \langle \hat{\theta}_{k_n}, x \rangle - q_{\alpha/2}^{*}(x) \right],$$

being $q_\alpha^*(x)$ the pointwise bootstrap α–quantile.

A simulation study compares the practical performance of asymptotic and bootstrap confidence intervals in terms of length and coverage rates for two linear regression operators and several sample sizes. This numerical study allowed confirm the good behaviour of bootstrap confidence intervals with respect to the asymptotic confidence intervals. In particular, it was noted that an adequate selection of the pilot parameter k_n^d involved in the bootstrap procedure makes empirical coverage rates of bootstrap intervals be closer to nominal level α than the coverage rates of asymptotic confidence intervals. Unfortunately, a data–driven methodology which automatically selects this *optimal* value of k_n^d is not yet available, so further research is required in order to solve this issue.

Chapter 5. Testing in functional linear regression

In this chapter, the functional linear model with scalar response is considered but including an intercept term, that is,

$$Y = \langle \theta, X \rangle + b + \epsilon,$$

where Y and ϵ are real random variables, X is a random variable valued in a separable Hilbert space $(\mathcal{H}, \langle \cdot, \cdot \rangle)$, and the model parameters b and θ belong to $\mathbb{R}$ and $\mathcal{H}$, respectively. In this context, a consistent bootstrap method to calibrate the distribution of test statistics for assesing $H_0 : \theta = 0$ versus $H_1 : \theta \neq 0$ (i.e., for testing the lack of dependence) is developed, and the related asymptotic theory is presented.

Next, two linear models,

$$Y_1 = \langle \theta_1, X_1 \rangle + b_1 + \epsilon_1 \quad \text{and} \quad Y_2 = \langle \theta_2, X_2 \rangle + b_2 + \epsilon_2,$$

satisfying that X_1 and X_2 have the same covariance operator, and ϵ_1 and ϵ_2 have the same variance, are taken. A bootstrap method for checking the equality of the two linear models, i.e., for testing $H_0 : \theta_1 = \theta_2$ versus $H_1 : \theta_1 \neq \theta_2$, is introduced, and a study of its main asymptotic properties is done in order to show its consistency and correctness.

From a practical point of view, the simulation study showed that bootstrap methods are competitive alternatives to tests based on asymptotic distributions, since they often give test sizes closer to the nominal ones. In addition, it must be recalled that the statistic tests which included a consistent estimation of the error variance σ^2 obtained higher empirical power than the test statistics which did not take it into account. Due to this fact, it could be interesting to study in future works what happens if statistic tests which does not include $\hat{\sigma^2}$ are studentized, that is, to analyse if their studentized versions improve the obtained results in the simulation study in terms of empirical power. Finally, a real data example also illustrates the performance of the proposed bootstrap techniques in practice.

It is important to highlight that for all the studied statistic test in which a k/k_n parameter is involved, the selection of this parameter seems to be a key issue. It is another research line that should be taken into account in the future. Besides the optimal k/k_n selection, other issues related to these hypotheses tests require further research, such as their extension to functional linear models with functional response.

Chapter 6. Thresholding in nonparametric functional regression

This chapter presents an exploratory tool focused on the detection of underlying complex structures in the nonparametric regression model with scalar response and functional covariate. The proposed methodology analyses the existence of hidden patterns related to the functional covariate X and/or the scalar response Y via a threshold procedure. For this purpose, an adequate threshold function must be chosen by the user according to the structure one wants to detect. A cross–validation criterion which allows to estimate the parameters involved in the threshold model is also introduced, and the usefulness of its graphical representation is studied.

A simulation study and applications to real datasets show the effectiveness of the threshold approach from a practical point of view. In the simulation study, it was found that the threshold estimators and the standard nonparametric estimator obtain similar results in terms of the mean square prediction error, whereas the mean square estimation error can be reduced if each subsample detected by the threshold technique is studied separately. In addition, the real data applications showed that the methodology allows to detect some kind of hidden structures, although the effectiveness of the procedure depends on the choice of the threshold function. Nevertheless, how to select this threshold function in a data–driven and efficient way is still an open question that will require further research.

Some computational issues. As well as the theoretical developments of the methodology presented throughout the manuscript, it is important to highlight that all the proposed methods were implemented and applied to both simulated and real datasets. For this purpose, the statistical free software

R was chosen (see further details in R Development Core Team, 2010, or `http://www.r-project.org`), and R routines were developed for each one of the new techniques compiled in this document. Given that the use of this free software is very widespread among the statistical community, the developed routines will be available online in order to allow any other researcher interested in FDA to use them. It must be noted that, in order to implement the new techniques introduced in this thesis, some existing routines were taking as a starting point. For instance, functions related with the paper by Cardot et al. (2003c) (available at `http://www.math.univ-toulouse.fr/staph`, section *"Softwares on line"*) were considered for the model parameter estimation in the functional linear model with scalar response, whereas R routines corresponding to nonparametric methods discussed in the book by Ferraty and Vieu (2006b) (available at `http://www.math.univ-toulouse.fr/staph/npfda`) served as a basis for nonparametric regression contributions. Furthermore, some specific R packages devoted to FDA, such as *fda* or *fda.usc* (see Ramsay et al., 2011, and Febrero-Bande and Oviedo de la Fuente, 2011, respectively), were also consulted.

Acknowledgements. *I would like to thank my advisors, Prof. Wenceslao González–Manteiga, Prof. Frédéric Ferraty, and Prof. Philippe Vieu, for their fruitful comments and their support during this years. I also wish to thank Prof. Gil González–Rodríguez, who has allowed me to enrich the theoretical contents of Chapter 5 by including some theoretical results which he developed for our paper González-Manteiga et al. (2012).*

This work has been supported by Ministry of Science and Innovation (FPI grant BES–2006–13389; national projects MTM2005–00820 and MTM2008–03010); and by Consellería de Innovación e Industria (regional project PGIDIT07PXIB207031PR) and Consellería de Economía e Industria (regional project 10MDS207015PR), Xunta de Galicia.

Resumo en galego

Hoxe en día, o progreso das ferramentas computacionais (tanto o aumento de capacidade como o de memoria) permite crear, almacenar e traballar con grandes bases de datos. En moitos casos, os conxuntos de datos que se manexan están compostos por observacións xeradas dunha distribución finito dimensional, aínda que medidas ao longo dun período de tempo ou rexistradas en diferentes localizacións espaciais. Cando ese *grid* temporal ou espacial é suficientemente fino, a mostra pode chegar a ser considerada como unha observación dunha variable aleatoria que toma valores nun espazo funcional. Analizar este tipo de datos cos métodos multivariantes estándar e ignorar a súa natureza funcional pode fallar estrepitosamente debido, por exemplo, á maldición da dimensionalidade, á existencia de colinearidade, ou á perda de información valiosa relativa á estrutura funcional subxacente. Nestes casos, requírese o emprego de técnicas estatísticas específicas, adaptadas a esta tipoloxía de datos, para manexar, filtrar e obter toda a información relevante contida nos datos.

Este feito converteu á Análise de Datos Funcionais (FDA de aquí en diante, pola terminoloxía inglesa *Functional Data Analysis*) nun dos campos estatísticos máis activos durante estes últimos anos. Desde os primeiras traballos iniciais dedicados a esta temática nos anos oitenta e noventa (Grenander, 1981; Dauxois et al., 1982; Ramsay, 1982; Bosq, 1991), a FDA deu lugar a varios libros (Bosq, 2000; Ramsay and Silverman, 1997, 2002, 2005; Ferraty and Vieu, 2006b; Ferraty and Romain, 2011; Horváth and Kokoszka, 2012), números especiais en revistas de alto factor de impacto (Davidian et al., 2004; González-Manteiga and Vieu, 2007; Valderrama, 2007; Ferraty, 2010) e *workshops* internacionais adicados tanto a desenvolvementos metodolóxicos como a desenvolvementos aplicados para datos funcionais (Dabo-Niang and Ferraty, 2008; Ferraty, 2011). Ademais, algunhas contribucións centradas en revisións do estado da arte da FDA foron publicadas durante os últimos anos (Rice, 2004; Müller, 2005; González-Manteiga and Vieu, 2011; Delsol et al., 2011a; Cuevas, 2012).

Cómpre resaltar que os datos funcionais xorden de forma natural na maior parte dos campos científicos: econometría (por exemplo, valoración de activos financeiros ou curvas de produción e demanda de electricidade), ciencias e enxeñaría (por exemplo, imaxes por satélite, planos topográficos ou técnicas de recoñecemento de imaxe), medio ambiente e climatoloxía (por exemplo, medicións meteorolóxicas ou curvas de fluxos fluviais), medicina (por exemplo, curvas de crecemento ou datos xenéticos), química (por exemplo, datos espectrométricos), etc. Debido a isto, xeráronse numerosos traballos aplicados relativos á FDA na literatura estatística, como as monografías de Ramsay and Silverman (2002) e Ferraty and Vieu (2006a).

Debido á novidade da FDA, aínda hai un longo listado de liñas de investigación que non foron iniciadas ata o momento e que serían susceptibles de ser exploradas nos próximos anos: desde a extensión ao contexto funcional de métodos multidimensionais ben coñecidos e estudados na literatura existente, á creación de novas técnicas estatísticas focalizadas, e especialmente adaptadas, ao tratamento de datos funcionais e aos seus problemas específicos. Non obstante, a pesar da natureza funcional dos conxuntos de datos que se manexan, os obxectivos que se marcan na FDA son esencialmente os mesmos que os que aparecen habitualmente cando un analiza conxuntos de datos multivariantes. Por conseguinte, a metodoloxía desenvolvida a día de hoxe no eido da FDA foi aparecendo co gallo de satisfacer necesidades similares ás existentes no ámbito da estatística multivariante:

- Preprocesado de datos: problemas de rexistro e aliñamento de características dos datos, técnicas de suavizado, etc.

- Análise descritiva e exploratoria dos datos, subliñando as súas características máis importantes: medidas de centralidade e dispersión, detección de valores atípicos ou *outliers*, Análise de Compoñentes Principais Funcionais (FPCA de aquí en diante, pola terminoloxía inglesa *Functional Principal Component Analysis*), etc.
- Clasificación de datos funcionais.
- Construción de modelos que expliquen a relación entre variables funcionais: modelos de regresión paramétricos e non–paramétricos.
- Inferencia estatística funcional: intervalos de confianza, contrastes de hipóteses, etc.

Este manuscrito trata principalmente dos últimos dous puntos do listado anterior. Polo que concirne ao cuarto punto, o traballo concentrouse no modelo lineal funcional con resposta escalar (aínda que no último capítulo da tese se inclúen tamén algunhas contribucións á regresión non–paramétrica funcional), mentres que no relacionado co quinto punto do listado desenvolveuse un procedemento *bootstrap* que permite tanto a construción de intervalos de confianza como a calibración de contrastes de hipóteses relativas ao modelo lineal funcional con resposta escalar.

A tese foi estruturada en seis capítulos diferentes. Para cada un deles, exponse a continuación un breve resumo do seu contido. Despois de ditos resumos, inclúense algunhas cuestións computacionais relativas á implementación das novas metodoloxías propostas neste documento, xunto cun apartado de agradecementos.

Capítulo 1. Introdución á FDA

O primeiro capítulo desta tese é empregado para fixar a notación e dar un breve resumo xeral do estado da arte sobre os métodos estatísticos para datos funcionais. A idea clave na FDA consiste en supoñer que unha variable aleatoria X é observada nun *grid* discreto $\{t_l\}_{l=1}^L$ onde cada $t_l \in T$. Se os instantes de tempo t_l están próximos abondo, un podería asumir que $\{X(t_l)\}_{l=1}^L$ é unha única observación dunha variable aleatoria funcional

$$X = \{X(t); t \in T\}.$$

É importante salientar que T é a miúdo un intervalo real tal que $T = [T_0, T_1] \subset \mathbb{R}$, polo que cada observación pode ser considerada como se fose unha curva. Non obstante, a FDA cobre moitas outras situacións onde o espazo T é un espazo funcional máis xeral, tal e como se pode deducir a partir das seguintes definicións.

Definición 1. Dise que unha variable aleatoria X é unha *variable aleatoria funcional* se toma valores nun espazo dimensional infinito ou nun espazo funcional, denotado por $\mathcal{S}$. Unha observación x de X dise que é un *dato funcional*.

Definición 2. Dada unha variable aleatoria funcional X e un valor $n \in \mathbb{N}^*$, unha *mostra aleatoria funcional* de X de lonxitude n é un conxunto $\{X_i\}_{i=1}^n$ de variables aleatorias funcionais independentes e identicamente distribuídas (i.i.d. de aquí en diante) coa mesma distribución que X. Unha observación $\{x_i\}_{i=1}^n$ de $\{X_i\}_{i=1}^n$ dise que é un *conxunto de datos funcional*.

Unha vez que se ten definido o que son os datos funcionais, danse no capítulo algúns exemplos reais deste tipo de datos, os cales serán moi útiles á hora de ilustrar os métodos propostos nos seguintes capítulos.

Ao longo do manuscrito, suponse habitualmente que o espazo onde as variables funcionais toman valores é un espazo de Hilbert real e separable denotado por $\mathcal{H}$ (aínda que ás veces o espazo funcional se restrinxe ao ben coñecido espazo L^2 para poder ilustrar con maior claridade certos métodos). Isto

motiva a introdución dalgúns espazos asociados, como é o espazo dos operadores de Hilbert–Schmidt ou o espazo dual $\mathcal{H}'$, de certas notacións relativas aos produtos tensorias (é dicir, os operadores $\otimes_{\mathcal{H}}$ e $\otimes_{\mathcal{H}'}$) e da noción de semi–métrica, que é unha ferramenta verdadeiramente útil á hora de determinar a proximidade entre dúas observacións de tipo funcional.

Tras estas definicións e conceptos relacionados cos espazos de Hilbert, preséntase unha revisión xeral das ferramentas da FDA existentes hoxe en día: técnicas para o preprocesado de datos (métodos de suavizado e métodos de rexistro), estatística descritiva funcional (medidas de posición e dispersión), e algúns métodos exploratorios clave (como por exemplo, a FPCA, a cal será recordada máis adiante para poder definir certo tipo de estimadores para o modelo de regresión lineal funcional).

Capítulo 2. Modelos de regresión funcional

Este capítulo está dedicado aos modelos de regresión funcionais. Ao principio do mesmo, preséntase unha revisión xeral dos modelos de regresión funcional, para logo, durante o resto do capítulo, concentrarse nos modelos de regresión con resposta escalar e covariable funcional. Hai dous enfoques principais na literatura da FDA á hora de discutir este tema: o enfoque paramétrico e o enfoque non–paramétrico. En canto ao enfoque paramétrico, o modelo paramétrico máis usual é o *modelo lineal funcional con resposta escalar* dado por

$$Y = \langle \theta, X \rangle + \epsilon,$$

onde Y é unha variable aleatoria real, $m(\cdot) = \langle \theta, \cdot \rangle$ é un operador de regresión lineal con $\theta \in \mathcal{H}$ e $\|\theta\|^2 < \infty$, X é unha variable aleatoria de media cero que toma valores no espazo $\mathcal{H}$ tal que $\mathbb{E}(\|X\|^2) < \infty$, e ϵ é unha variable aleatoria real que verifica as seguintes condicións: $\mathbb{E}(\epsilon) = 0$, $\mathrm{Var}(\epsilon) = \sigma^2 < \infty$, e $\mathbb{E}(\epsilon X) = 0$. As dúas estimacións máis populares nesta situación que se acaba de describir introdúcense neste capítulo: os estimadores baseados en sistemas de bases, como o estimador *B–splines* penalizado

$$\hat{\theta}_{PS} = \sum_{j=1}^{q+k} \hat{b}_j B_{k,j}$$

(ver, por exemplo, Cardot et al., 2003c), e os estimadores tipo FPCA, como o estimador FPCA estándar

$$\hat{\theta}_{k_n} = \sum_{j=1}^{k_n} \frac{\Delta_n \hat{v}_j}{\hat{\lambda}_j} \hat{v}_j$$

(ver Cardot et al., 1999, 2003c, 2007c).

Polo que concirne ao enfoque non–paramétrico do problema, o modelo non–paramétrico funcional pode ser expresado como

$$Y = m(X) + \epsilon,$$

onde Y é unha variable aleatoria real, $m(\cdot)$ é un operador de regresión funcional que satisfai algunhas restricións de suavidade, X é unha variable aleatoria con media cero que toma valores nun espazo abstracto $\mathcal{S}$ dotado dunha semi–métrica $d(\cdot, \cdot)$, e ϵ é unha variable aleatoria real que satisfai as seguintes condicións $\mathbb{E}(\epsilon) = 0$, $\mathrm{Var}(\epsilon) = \sigma^2 < \infty$, e $\mathbb{E}(\epsilon X) = 0$. Entre os diferentes métodos non–paramétricos que existen na actualidade, a versión funcional do estimador tipo núcleo multivariante definido como

$$\hat{m}_h(x) = \frac{\sum_{i=1}^n Y_i K(h^{-1} d(x, X_i))}{\sum_{i=1}^n K(h^{-1} d(x, X_i))},$$

é analizado neste capítulo (ver Ferraty and Vieu, 2004, 2006b; Ferraty et al., 2007a).

Capítulo 3. Presuavización na regresión lineal funcional

Este capítulo focalízase no modelo lineal funcional con resposta escalar e variable explicativa que toma valores nun espazo funcional. Tal e como se discutiu no capítulo anterior, a FPCA foi empregada para estimar o parámetro funcional do modelo na literatura estatística recente. Unha modificación deste enfoque utilizando técnicas presuavizado é xustamente o que se propón no presente capítulo: ou ben mediante a presuavización da estrutura de covarianza da variable funcional, ou ben mediante a presuavización da variable resposta escalar. Concretamente, son introducidos catro estimadores tipo FPCA diferentes para o parámetro θ do modelo lineal, todos eles baseados en técnicas de presuavizado:

(i) Presuavización da estrutura de covarianza: $\hat{\theta}_{k_n}^{\alpha_n} = \sum_{j=1}^{k_n} \frac{\Delta_n \hat{v}_j}{\hat{\lambda}_j + \alpha_n} \hat{v}_j$.

(ii) Presuavización da variable resposta: $\hat{\theta}_{k_n}^{h_n} = \sum_{j=1}^{k_n} \frac{\Delta_n^{h_n} \hat{v}_j}{\hat{\lambda}_j} \hat{v}_j$.

(iii) Uso da FPCA presuavizada proposta por Pezzulli and Silverman (Pezzulli and Silverman, 1993):

$$\hat{\theta}_{k_n}^{PS,\,\alpha_n} = \sum_{j=1}^{k_n} \frac{\Delta_n \hat{v}_j^{\alpha_n,1}}{\hat{\lambda}_j^{\alpha_n,1}} \hat{v}_j^{\alpha_n,1}.$$

(iv) Uso da FPCA presuavizada proposta por Silverman (Silverman, 1996):

$$\hat{\theta}_{k_n}^{S,\,\alpha_n} = \sum_{j=1}^{k_n} \frac{\Delta_n \hat{v}_j^{\alpha_n,2}}{\hat{\lambda}_j^{\alpha_n,2}} \hat{v}_j^{\alpha_n,2}.$$

A primeira proposta, $\hat{\theta}_{k_n}^{\alpha_n}$, pode ser vista como unha extensión do estimador de regresión ordinario *ridge* multivariante a espazos máis xerais, neste caso aos espazos de Hilbert. A idea clave é evitar os problemas de mal–condicionamento perturbando lixeiramente os autovalores do operador momento de segunda orde. A consistencia e as expresións para os erros cadráticos medios condicionais para a predición e para a estimación son obtidos para $\hat{\theta}_{k_n}^{\alpha_n}$. Empregando o erro de estimación condicional, pode verse, desde un punto de vista teórico, que este estimador presuavizado consegue mellorar os resultados do estimador FPCA estándar, especialmente cando o ruído do modelo é grande e/ou o tamaño de mostra é pequeno. Polo que concirne ao estimador $\hat{\theta}_{k_n}^{h_n}$, a súa consistencia foi establecida e as expresións para os seus erros cadráticos medios condicionais para a predición e a estimación foron calculadas. En canto aos estimadores presuavizados $\hat{\theta}_{k_n}^{PS,\,\alpha_n}$ e $\hat{\theta}_{k_n}^{S,\,\alpha_n}$, soamente se calcularon as súas expresións para os erros condicionais empregando para iso algúns cálculos heurísticos baseados nas técnicas estándar de expansións asintóticas.

A efectividade dos estimadores propostos baseados na presuavización da estrutura de covarianza e na presuavización da variable resposta con respecto ao estimador FPCA estándar e o estimador B–splines penalizado tamén se comprobou mediante un estudo de simulación e a aplicación destas metodoloxías a conxuntos de datos reais. En ambos os dous casos, $\hat{\theta}_{k_n}^{h_n}$ foi substituído por

$$\hat{\theta}_{k_n}^{\text{k–NN}} = \sum_{j=1}^{k_n} \frac{\Delta_n^{\text{k–NN}} \hat{v}_j}{\hat{\lambda}_j} \hat{v}_j,$$

dado que $\hat{\theta}_{k_n}^{\text{k–NN}}$ proporcionou erros máis reducidos que os de $\hat{\theta}_{k_n}^{h_n}$ nalgunhas simulacións previas. As simulacións suxiren que o estimador presuavizado $\hat{\theta}_{k_n}^{\alpha_n}$ mellora os resultados obtidos polo estimador FPCA estándar, especialmente cando o tamaño de mostra é pequeno, mentres que o estimador $\hat{\theta}_{k_n}^{\text{k–NN}}$ non reduce significativamente os erros condicionais de $\hat{\theta}_{k_n}$. A selección dos parámetros involucrados nos estimadores presuavizados propostos é, por suposto, un punto clave. Unha pauta práctica xeral é elixir estes parámetros mediante técnicas de validación cruzada.

Capítulo 4. *Bootstrap* na regresión funcional lineal

Para tratar o modelo lineal funcional con variable independente funcional e variable resposta escalar, tal e como foi comentado anteriormente, un dos métodos máis populares para a estimación do parámetro do modelo está baseado na FPCA. A converxencia débil para unha clase ampla de estimadores tipo FPCA foi recentemente demostrada e, como resultado, os seguintes intervalos de confianza asintóticos para o operador de regresión lineal $m(\cdot)$ poden ser obtidos para un nivel de confianza fixado $\alpha \in (0,1)$

$$\mathrm{CI}^{asy}_{x,\alpha} = \left[\langle \hat{\theta}_{k_n}, x \rangle - z_{1-\alpha/2} \frac{\hat{t}^{k_n}_{n,x} \hat{\sigma}}{\sqrt{n}}, \langle \hat{\theta}_{k_n}, x \rangle + z_{1-\alpha/2} \frac{\hat{t}^{k_n}_{n,x} \hat{\sigma}}{\sqrt{n}} \right],$$

onde z_α é o cuantil de orde α dunha distribución $\mathcal{N}(0,1)$. Neste capítulo, proponse unha aproximación alternativa ao problema que permite calcular intervalos de confianza punto a punto mediante un procedemento *bootstrap*, obténdose tamén a súa validez asintótica (validez asintótica no sentido de que a distribución condicional do estimador pode ser aproximada pola distribución *bootstrap*). En particular, algoritmos para o *bootstrap naive* e o *wild bootstrap* foron desenvolvidos, polo que os intervalos *bootstrap* poden ser construídos como

$$\mathrm{CI}^{*}_{x,\alpha} = \left[\langle \hat{\theta}_{k_n}, x \rangle - q^{*}_{1-\alpha/2}(x), \langle \hat{\theta}_{k_n}, x \rangle - q^{*}_{\alpha/2}(x) \right],$$

sendo $q^{*}_{\alpha}(x)$ o α–cuantil *bootstrap* punto a punto.

Un estudo de simulación compara o rendemento práctico dos intervalos de confianza asintóticos e os intervalos de confianza *bootstrap* en termos de lonxitude e cobertura para dous operadores de regresión lineais e varios tamaños de mostra. Este estudo numérico permitiu confirmar o bo comportamento dos intervalos de confianza *boostrap* con respecto aos intervalos de confianza asintóticos. En particular, observouse que unha selección axeitada do parámetro piloto k^d_n implicado no procedemento *bootstrap* fai que as coberturas empíricas dos intervalos *bootstrap* estean máis próximos ao nivel nominal α que as coberturas dos intervalos de confianza asintóticos. Desafortunadamente, non existe ningunha metodoloxía polo de agora que permita seleccionar automaticamente este valor "óptimo" de k^d_n, así que vai ser precisa máis investigación neste tema para poder resolver esta cuestión.

Capítulo 5. Contrastes de hipóteses na regresión funcional lineal

O modelo lineal funcional con resposta escalar é o modelo considerado neste capítulo, pero desta vez incluíndo un intercepto non nulo no modelo, isto é,

$$Y = \langle \theta, X \rangle + b + \epsilon,$$

onde Y e ϵ son variables aleatorias reais, X é unha variable aleatoria que toma valores nun espazo de Hilbert separable $(\mathcal{H}, \langle \cdot, \cdot \rangle)$, e os parámetros do modelo b e θ pertencen a $\mathbb{R}$ e $\mathcal{H}$, respectivamente. Neste contexto, desenvólvese un método *bootstrap* consistente para calibrar a distribución do test estatístico para contrastar a hipótese nula $H_0 : \theta = 0$ fronte a hipótese alternativa $H_1 : \theta \neq 0$ (isto é, para contrastar a falta de dependencia). Ademais, a teoría asintótica relacionada con esta metodoloxía *bootstrap* tamén se presenta neste capítulo.

A continuación, considéranse dous modelos lineais,

$$Y_1 = \langle \theta_1, X_1 \rangle + b_1 + \epsilon_1 \quad \text{and} \quad Y_2 = \langle \theta_2, X_2 \rangle + b_2 + \epsilon_2,$$

que verifiquen que X_1 e X_2 teñen o mesmo operador de covarianza, e ϵ_1 e ϵ_2 teñen a mesma varianza. No capítulo introdúcese un método *bootstrap* para comprobar a igualdade destes dous modelos lineais, isto é, para contrastar $H_0 : \theta_1 = \theta_2$ contra $H_1 : \theta_1 \neq \theta_2$. Tamén se desenvolveu un estudo das principais propiedades asintóticas de do método *bootstrap* proposto, estudo que permite afirmar que é consistente e correcto.

Desde un punto de vista práctico, o estudo de simulación demostrou que os métodos *bootstrap* son alternativas competitivas aos tests baseados en distribucións asintóticas, posto que a miúdo dan proporcións de rexeitamentos baixo a hipótese nula máis próximas aos nominais. Ademais, debe recordarse que os tests estatísticos que incluíron unha estimación consistente da varianza de erro σ^2 obtiveron potencias empíricas máis elevadas que os test estatísticos que non a tiña en conta. Debido a este feito, podería ser interesante estudar no futuro que sucede se os tests estatísticos que non inclúen a estimación da varianza $\hat{\sigma}^2$ son *estudentizados*, isto é, analizar se as súas versións *estudentizadas* poden mellorar os resultados obtidos no estudo de simulación en termos de potencia empírica. Finalmente, un exemplo de datos reais ilustra tamén o comportamento das técnicas *bootstrap* propostas na práctica.

É importante subliñar que para todos os tests estatísticos estudados nos cales un parámetro k/k_n está implicado, a selección de dito parámetro parece ser unha cuestión clave. Esta é outra liña de investigación que debería ser tida en conta no futuro. Ademais da selección do k/k_n óptimo, hai tamén outras cuestións relacionadas cos tests de hipóteses que aínda requiren máis investigación e desenvolvemento, como a súa extensión a modelos lineais funcionais onde a resposta sexa tamén funcional.

Capítulo 6. *Thresholding* na regresión non–paramétrica funcional

Este último capítulo presenta unha ferramenta exploratoria focalizada na detección de estruturas complexas subxacentes no modelo de regresión non–paramétrico con resposta escalar e covariable funcional. A metodoloxía proposta analiza a existencia de modelos ocultos relacionados coa covariable funcional X e/ou a resposta escalar Y mediante un procedemento *threshold*. Para iso, unha función *threshold* axeitada debe ser elixida polo usuario segundo a estrutura que se queira detectar. Unha criterio de validación cruzada, que permite estimar os parámetros implicados no modelo *threshold*, tamén se desenvolve, e a utilidade da representación gráfica de dito criterio de validación cruzada é estudada no capítulo.

Un estudo de simulación e varias aplicacións a conxuntos de datos reais amosan a efectividade do enfoque *threshold* desde un punto de vista práctico. No estudo de simulación, observouse que os estimadores *threshold* e o estimador non–paramétrico estándar obteñen resultados similares en termos do erro cadrático medio de predición, mentres que o erro cadrático medio de estimación pode ser reducido se cada submostra detectada pola técnica *threshold* é estudada separadamente. Ademais, as aplicacións a datos reais demostraron que a metodoloxía permite detectar algúns tipos de estruturas ocultas, aínda que a efectividade do procedemento depende da selección da función *threshold*. Lamentablemente, como seleccionar esta función *threshold* de forma automática e eficaz é aínda unha pregunta aberta que require máis traballo e investigación no futuro.

Algunhas cuestións computacionais. Ademais dos desenvolvementos teóricos da metodoloxía presentada ao longo do manuscrito, é importante subliñar que todos os novos métodos propostos foron implementados e aplicados a conxuntos de datos tanto simulados coma reais. Para iso, o software estatístico *R* foi o elixido (ver R Development Core Team, 2010, ou `http://www.r-project.org`), e varias rutinas para *R* foron desenvolvidas para poder implementar cada unha das novas técnicas recompiladas neste documento. Dado que o uso de *R* está moi estendido entre a comunidade estatística, é de esperar que as rutinas que se desenvolveron, e que estarán dispoñibles online, poidan ser facilmente utilizadas por calquera outro investigador interesado na FDA.

É importante destacar tamén que, para implementar as novas técnicas introducidas nesta tese, algunhas rutinas existentes foron tomadas como punto de partida. Por exemplo, as funcións relativas ao artigo de Cardot et al. (2003c) (dispoñibles en `http://www.math.univ-toulouse.fr/staph`, na sección *"Softwares on line"*) foron consideradas para a estimación do parámetro do modelo no caso do modelo lineal funcional con resposta escalar, mentres que as rutinas de *R* correspondentes aos métodos de estatística non–paramétricos discutidos no libro de Ferraty and Vieu (2006b) (dispoñible en `http://www.math.univ-toulouse.fr/staph/npfda`) foron tidas en conta para as contribucións en regresión non–paramétrica. Ademais, algúns paquetes específico de *R* dedicados a FDA, como *fda* ou

fda.usc (ver Ramsay et al., 2011, e Febrero-Bande and Oviedo de la Fuente, 2011, respectivamente), serviron de referencia.

Agradecementos. *Quero agradecerlle aos profesores Wenceslao González–Manteiga, Frédéric Ferraty e Philippe Vieu o seu apoio e dedicación durante a realización desta tese. Tamén quero darlle as grazas ao profesor Gil González–Rodríguez, que me permitiu enriquecer os contidos teóricos do Capítulo 5 incluíndo algúns resultados que desenvolveu para o artigo González-Manteiga et al. (2012).*

Este traballo foi financiado polo Ministerio de Ciencia e Innovación (mediante a bolsa FPI BES–2006–13389 e os proxectos nacionais MTM2005–00820 e MTM2008–03010) e pola Xunta de Galicia (mediante os proxectos autonómicos PGIDIT07PXIB207031PR da Consellería de Innovación e Industria, e 10MDS207015PR da Consellería de Economía e Industria).

Esta tese está dedicada a meus pais e avós, a Laura, a Fran, a Ana, a Iria, a María I. e a María S., pola súa paciencia e apoio incondicional durante todos estes anos. Grazas a vós, á vosa comprensión e aos vosos ánimos, puiden levar a cabo este traballo.

Quero darlle ás grazas aos membros da Sección de Medio Ambiente da Central Térmica de As Pontes pola amizade que me amosaron durante o ano de bolseira que pasei con eles, e ao grupo de traballo STAPH de Toulouse pola boa acollida que me ofreceron durante as miñas estadías en Francia. Moitas grazas tamén a todos os compañeiros cos que compartín seminarios dos venres, charlas do SII e/ou veciñanza na sala de bolseiros, aos membros do departamento de Estatística e I.O. da USC (especialmente, aos que me acompañaron na miña iniciación no mundo da docencia universitaria e da investigación), e a todos aqueles cos que traballei durante a miña etapa de técnico do Nodo CESGA.

Bibliography

Abraham, C., Biau, G., and Cadre, B. (2006). On the Kernel Rule for Function Classification. *Annals of the Institute of Statistical Mathematics*, 58(3):619–633.

Abraham, C., Cornillon, P. A., Matzner-Løber, E., and Molinari, N. (2003). Unsupervised Curve Clustering using B–Splines. *Scandinavian Journal of Statistics*, 30(3):581–595.

Abramovich, F. and Angelini, C. (2006). Testing in mixed–effects FANOVA models. *Journal of Statistical Planning and Inference*, 136(12):4326–4348.

Abramovich, F., Antoniadis, A., Sapatinas, T., and Vidakovic, B. (2004). Optimal testing in a fixed–effects functional analysis of variance model. *International Journal of Wavelets, Multiresolution and Information Processing*, 2(4):323–349.

Aerts, M., Hens, N., and Simonoff, J. S. (2010). Model selection in regression based on pre–smoothing. *Journal of Applied Statistics*, 37(9):1455–1472.

Aguilera, A. M., Escabias, M., and Valderrama, M. J. (2006). Using principal components for estimating logistic regression with high–dimensional multicollinear data. *Computational Statistics & Data Analysis*, 50(8):1905–1924.

Aguilera, A. M., Ocaña, F. A., and Valderrama, M. J. (1997). Estimated principal component prediction model for continuous–time stochastic processes. *Applied Stochastic Models and Data Analysis*, 13(2):61–72.

Aguilera, A. M., Ocaña, F. A., and Valderrama, M. J. (1999). Forecasting time series by functional PCA. Discussion of several weighted approaches. *Computational Statistics*, 14(3):443–467.

Ait-Saïdi, A., Ferraty, F., and Kassa, R. (2005). Single functional index model for a time series. *Revue Roumaine de Mathématiques Pures et Appliquées*, 50(4):321–330.

Ait-Saïdi, A., Ferraty, F., Kassa, R., and Vieu, P. (2008). Cross–validated estimations in the single–functional index model. *Statistics: A Journal of Theoretical and Applied Statistics*, 42(6):475–494.

Akritas, M. G. (1996). On the use of nonparametric regression techniques for fitting parametric regression models. *Biometrics*, 52(4):1342–1362.

Amato, U., Antoniadis, A., and de Feis, I. (2006). Dimension reduction in functional regression with applications. *Computational Statistics & Data Analysis*, 50(9):2422–2446.

Aneiros-Pérez, G., Cao, R., and Vilar-Fernández, J. M. (2011). Functional methods for time series prediction: a nonparametric approach. *Journal of Forecasting*, 30(4):377–392.

Aneiros-Pérez, G., Cardot, H., Estévez-Pérez, G., and Vieu, P. (2004). Maximum ozone concentration forecasting by functional non–parametric approaches. *Environmetrics*, 15(7):675–685.

Aneiros-Pérez, G. and Vieu, P. (2006). Semi–functional partial linear regression. *Statistics & Probability Letters*, 76(11):1102–1110.

Aneiros-Pérez, G. and Vieu, P. (2008). Nonparametric time series prediction: A semi–functional partial linear modeling. *Journal of Multivariate Analysis*, 99(5):834–857.

Aneiros-Pérez, G. and Vieu, P. (2012). Testing linearity in semi–parametric functional data analysis. *Computational Statistics*. To appear.

Antoch, J., Prchal, L., Rosa, M. R., and Sarda, P. (2008). Functional Linear Regression with Functional Response: Application to Prediction of Electricity Consumption. In Dabo-Niang, S. and Ferraty, F., editors, *Functional and Operatorial Statistics*, Contributions to Statistics, pages 23–29, Heidelberg. Physica–Verlag.

Antoch, J., Prchal, L., Rosa, M. R., and Sarda, P. (2010). Electricity consumption prediction with functional linear regression using spline estimators. *Journal of Applied Statistics*, 37(12):2027–2041.

Antoniadis, A. and Sapatinas, T. (2003). Wavelet methods for continuous–time prediction using Hilbert–valued autoregressive processes. *Journal of Multivariate Analysis*, 87(1):133–158.

Antoniadis, A. and Sapatinas, T. (2007). Estimation and inference in functional mixed–effects models. *Computational Statistics & Data Analysis*, 51(10):4793–4813.

Araki, Y., Konishi, S., and Imoto, S. (2004). Functional discriminant analysis for microarray gene expression data via radial basis function networks. In Antoch, J., editor, *COMPSTAT 2004 – Proceedings in Computational Statistics. 16th Symposium Held in Prague, Czech Republic, 2004*, pages 589–596, Heidelberg. Physica–Verlag.

Aspirot, L., Bertin, K., and Perera, G. (2009). Asymptotic normality of the Nadaraya–Watson estimator for nonstationary functional data and applications to telecommunications. *Journal of Nonparametric Statistics*, 21(5):535–551.

Baíllo, A. (2009). A note on functional linear regression. *Journal of Statistical Computation and Simulation*, 79(5):657–669.

Baíllo, A., Cuevas, A., and Fraiman, R. (2011). Classification methods for functional data. In Ferraty, F. and Romain, Y., editors, *The Oxford Handbook of Functional Data Analysis*, Oxford Handbooks in Mathematics, pages 259–297. Oxford University Press, Oxford.

Baíllo, A. and Grané, A. (2008). Local linear regression for functional predictor and scalar response. In Dabo-Niang, S. and Ferraty, F., editors, *Functional and Operatorial Statistics*, Contributions to Statistics, pages 47–51, Heidelberg. Physica–Verlag.

Baíllo, A. and Grané, A. (2009). Local linear regression for functional predictor and scalar response. *Journal of Multivariate Analysis*, 100(1):102–111.

Barrientos-Marin, J., Ferraty, F., and Vieu, P. (2010). Locally modelled regression and functional data. *Journal of Nonparametric Statistics*, 22(5):617–632.

Bathia, N., Yao, Q., and Ziegelmann, F. (2010). Identifying the finite dimensionality of curve time series. *Annals of Statistics*, 38(6):3352–3386.

Benhenni, K., Ferraty, F., Rachdi, M., and Vieu, P. (2007). Local smoothing regression with functional data. *Computational Statistics*, 22(3):353–369.

Benjamini, Y. and Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 57(1):289–300.

Benko, M., Härdle, W., and Kneip, A. (2009). Common Functional Principal Components. *Annals of Statistics*, 37(1):1–34.

Berkes, I., Gabrys, R., Horváth, L., and Kokoszka, P. (2009). Detecting changes in the mean of functional observations. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 71(5):927–946.

Besse, P. C. and Cardot, H. (1996). Approximation spline de la prévision d'un processus fonctionnel autorégressif d'ordre 1. *Canadian Journal of Statistics*, 24(4):467–487.

Besse, P. C., Cardot, H., Faivre, R., and Goulard, M. (2005). Statistical modelling of functional data. *Applied Stochastic Models in Business and Industry*, 21(2):165–173.

Besse, P. C., Cardot, H., and Stephenson, D. B. (2000). Autoregressive Forecasting of Some Functional Climatic Variations. *Scandinavian Journal of Statistics*, 27(4):673–687.

Biau, G., Bunea, F., and Wegkamp, M. H. (2005). Functional classification in Hilbert spaces. *IEEE Transactions on Information Theory*, 51(6):2163–2172.

Biau, G., Cérou, F., and Guyader, A. (2010). Rates of convergence of the functional k–nearest neighbor estimate. *IEEE Transactions on Information Theory*, 56(4):2034–2040.

Bickel, P. J. and Freedman, D. A. (1981). Some asymptotic theory for the bootstrap. *Annals of Statistics*, 9(6):1196–1217.

Boente, G. and Fraiman, R. (2000). Kernel–based functional principal components. *Statistics & Probability Letters*, 48(4):335–345.

Boj, E., Delicado, P., and Fortiana, J. (2008). Local linear functional regression based on weighted distance–based regression. In Dabo-Niang, S. and Ferraty, F., editors, *Functional and Operatorial Statistics*, Contributions to Statistics, pages 57–64, Heidelberg. Physica–Verlag.

Boj, E., Delicado, P., and Fortiana, J. (2010). Distance–based local linear regression for functional predictors. *Computational Statistics & Data Analysis*, 54(2):429–437.

Borggaard, C. and Thodberg, H. H. (1992). Optimal minimal neural interpretation of spectra. *Analytical Chemistry*, 64(5):545–551.

Bosq, D. (1991). Modelization, nonparametric estimation and prediction for continuous time processes. In Roussas, G., editor, *Nonparametric Functional Estimation and Related Topics. Proceedings of the NATO Advanced Study Institute, Spetses, Greece, July 29–August 10, 1990*, volume 335 of *NATO Science Series C: Mathematical and Physical Sciences*, pages 509–529, Dordrecht, The Netherlands. Kluwer Academic Publishers/Springer.

Bosq, D. (2000). *Linear Processes in Function Spaces: Theory and Applications*, volume 149 of *Lecture Notes in Statistics*. Springer, New York.

Bosq, D. (2002). Estimation of Mean and Covariance Operator of Autoregressive Processes in Banach Spaces. *Statistical Inference for Stochastic Processes*, 5(3):287–306.

Bosq, D. (2003). Berry–Esseen inequality for linear processes in Hilbert spaces. *Statistics & Probability Letters*, 63(3):243–247.

Bosq, D. (2007). General linear processes in Hilbert spaces and prediction. *Journal of Statistical Planning and Inference*, 137(3):879–894.

Bosq, D. and Blanke, D. (2007). *Inference and Prediction in Large Dimensions*. Wiley Series in Probability and Statistics. Wiley, West Sussex, England.

Bouzas, P. R., Aguilera, A. M., and Ruiz-Fuentes, N. (2002). Functional Principal Component Modelling of the Intensity of a Doubly Stochastic Poisson Process. In Härdle, W. and Rönz, B., editors, *Compstat. Proceedings in Computational Statistics: 15th Symposium Held in Berlin, Germany, 2002*, pages 373–376, Heidelberg. Physica–Verlag.

Bouzas, P. R., Valderrama, M. J., Aguilera, A. M., and Ruiz-Fuentes, N. (2006). Modelling the mean of a doubly stochastic Poisson process by functional data analysis. *Computational Statistics & Data Analysis*, 50(10):2655–2667.

Brumback, B. and Rice, J. (1998). Smoothing Spline Models for the Analysis of Nested and Crossed Samples of Curves: Rejoinder. *Journal of the American Statistical Association*, 93(443):961–976.

Bücher, A., Dette, H., and Wieczorek, G. (2011). Testing model assumptions in functional regression models. *Journal of Multivariate Analysis*, 102(10):1472–1488.

Bugni, F. A., Hall, P., Horowitz, J. L., and Neumann, G. R. (2009). Goodness–of–fit tests for functional data. *Econometrics Journal*, 12(s1):S1–S18.

Bunea, F., Ivanescu, A. E., and Wegkamp, M. H. (2011). Adaptive inference for the mean of a Gaussian process in functional data. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 73(4):531–558.

Burba, F., Ferraty, F., and Vieu, P. (2009). k–Nearest Neighbour method in functional nonparametric regression. *Journal of Nonparametric Statistics*, 21(4):453–469.

Cadre, B. (2001). Convergent estimators for the L1–median of banach valued random variable. *Statistics: A Journal of Theoretical and Applied Statistics*, 35(4):509–521.

Cai, T. T. and Hall, P. (2006). Prediction in functional linear regression. *Annals of Statistics*, 34(5):2159–2179.

Cai, T. T. and Yuan, M. (2011). Optimal estimation of the mean function based on discretely sampled functional data: Phase transition. *Annals of Statistics*, 39(5):2330–2355.

Cai, Z. (2011). Functional coefficient models for economics and financial data. In Ferraty, F. and Romain, Y., editors, *The Oxford Handbook of Functional Data Analysis*, Oxford Handbooks in Mathematics, pages 166–186. Oxford University Press, Oxford.

Cao, J. and Ramsay, J. O. (2007). Parameter cascades and profiling in functional data analysis. *Computational Statistics*, 22(3):335–351.

Cao-Abad, R. (1991). Rate of convergence for the wild bootstrap in nonparametric regression. *Annals of Statistics*, 19(4):2226–2231.

Cao-Abad, R., López de Ullibarri, I., Janssen, P., and Veraverbeke, N. (2005). Presmoothed Kaplan–Meier and Nelson–Aalen estimators. *Journal of Nonparametric Statistics*, 17(1):31–56.

Cardot, H. (2000). Nonparametric estimation of smoothed principal components analysis of sampled noisy functions. *Journal of Nonparametric Statistics*, 12(4):503–538.

Cardot, H. (2007). Conditional functional principal components analysis. *Scandinavian Journal of Statistics*, 34(2):317–335.

Cardot, H., Crambes, C., Kneip, A., and Sarda, P. (2007a). Smoothing splines estimators in functional linear regression with errors–in–variables. *Computational Statistics & Data Analysis*, 51(10):4832–4848.

Cardot, H., Crambes, C., and Sarda, P. (2005). Quantile regression when the covariates are functions. *Journal of Nonparametric Statistics*, 17(7):841–856.

Cardot, H., Crambes, C., and Sarda, P. (2007b). Ozone Pollution Forecasting using Conditional Mean and Conditional Quantiles with Functional Covariates. In Hardle, W., Mori, Y., and Vieu, P., editors, *Statistical Methods for Biostatistics and Related Fields*, pages 221–244. Physica–Verlag, New York.

Cardot, H., Faivre, R., and Goulard, M. (2003a). Functional approaches for predicting land use with the temporal evolution of coarse resolution remote sensing data. *Journal of Applied Statistics*, 30(10):1185–1199.

Cardot, H., Ferraty, F., Mas, A., and Sarda, P. (2003b). Testing hypotheses in the functional linear model. *Scandinavian Journal of Statistics*, 30(1):241–255.

Cardot, H., Ferraty, F., and Sarda, P. (1999). Functional Linear Model. *Statistics & Probability Letters*, 45(1):11–22.

Cardot, H., Ferraty, F., and Sarda, P. (2003c). Spline estimators for the functional linear model. *Statistica Sinica*, 13(3):571–591.

Cardot, H., Goia, A., and Sarda, P. (2004). Testing for No Effect in Functional Linear Regression Models, Some Computational Approaches. *Communications in Statistics – Simulation and Computation*, 33(1):179–199.

Cardot, H., Mas, A., and Sarda, P. (2007c). CLT in functional linear regression models. *Probability Theory and Related Fields*, 138(3–4):325–361.

Cardot, H., Prchal, L., and Sarda, P. (2007d). No effect and lack–of–fit permutation tests for functional regression. *Computational Statistics*, 22(3):371–390.

Cardot, H. and Sarda, P. (2005). Estimation in generalized linear models for functional data via penalized likelihood. *Journal of Multivariate Analysis*, 92(1):24–41.

Cardot, H. and Sarda, P. (2006). Linear Regression Models for Functional Data. In Sperlich, S., Härdle, W., and Aydinli, G., editors, *The Art of Semiparametrics*, Contributions to Statistics, pages 49–66. Physica–Verlag, Heidelberg.

Cardot, H. and Sarda, P. (2011). Functional linear regression. In Ferraty, F. and Romain, Y., editors, *The Oxford Handbook of Functional Data Analysis*, Oxford Handbooks in Mathematics, pages 21–46. Oxford University Press, Oxford.

Cérou, F. and Guyader, A. (2006). Nearest neighbor classification in infinite dimension. *ESAIM: Probability and Statistics*, 10:340–355.

Chaouch, M. (2008). *Contribution to the nonparametric geometric quantiles estimation and functional data analysis*. PhD thesis, University of Bourgogne, France.

Chaouch, M. and Goga, C. (2012). Using Complex Surveys to Estimate the L1–Median of a Functional Variable: Application to Electricity Load Curves. *International Statistical Review*, 80(1):40–59.

Chen, D. and Müller, H. G. (2012). Nonlinear manifold representations for functional data. *Annals of Statistics*, 40(1):1–29.

Chen, K., Chen, K., Müller, H. G., and Wang, J. L. (2011). Stringing High–Dimensional Data for Functional Analysis. *Journal of the American Statistical Association*, 106(493):275–284.

Cheng, Y. and de Gooijer, J. G. (2007). On the uth geometric conditional quantile. *Journal of Statistical Planning and Inference*, 137(6):1914–1930.

Chesneau, C., Kachour, M., and Maillot, B. (2012). Nonparametric Estimation for Functional Data by Wavelet Thresholding. Available at `http://hal.archives-ouvertes.fr/hal-00634800`.

Chiou, J. M. and Müller, H. G. (2007). Diagnostics for functional regression via residual processes. *Computational Statistics & Data Analysis*, 51(10):4849–4863.

Chiou, J. M. and Müller, H. G. (2009). Modeling Hazard Rates as Functional Data for the Analysis of Cohort Lifetables and Mortality Forecasting. *Journal of the American Statistical Association*, 104(486):572–585.

Chiou, J. M., Müller, H. G., and Wang, J. L. (2003a). Functional quasi–likelihood regression models with smooth random effects. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 65(2):405–423.

Chiou, J. M., Müller, H. G., and Wang, J. L. (2004). Functional response models. *Statistica Sinica*, 14(3):675–693.

Chiou, J. M., Müller, H. G., Wang, J. L., and Carey, J. R. (2003b). A functional multiplicative effects model for longitudinal data, with application to reproductive histories of female medflies. *Statistica Sinica*, 13(4):1119–1133.

Clot, D. (2002). Using functional PCA for cardiac motion exploration. In *Second IEEE International Conference on Data Mining (ICDM'02)*, pages 91–98, Los Alamitos, CA, USA. IEEE Computer Society.

Costanzo, G. D., Preda, C., and Saporta, G. (2006). Anticipated prediction in discriminant analysis on functional data for binary response. In Rizzi, A. and Vichi, M., editors, *COMPSTAT 2006 – Proceedings in Computational Statistics. 17th Symposium Held in Rome, Italy, 2006*, pages 821–828, Heidelberg. Physica–Verlag.

Crambes, C. (2007). Régression fonctionnelle sur composantes principales pour variable explicative bruité. *Comptes Rendus Mathematique*, 345(9):519–522.

Crambes, C., Delsol, L., and Laksaci, A. (2008). Robust nonparametric estimation for functional data. *Journal of Nonparametric Statistics*, 20(7):573–598.

Crambes, C., Kneip, A., and Sarda, P. (2009). Smoothing splines estimators for functional linear regression. *Annals of Statistics*, 37(1):35–72.

Cristóbal-Cristóbal, J. A., Faraldo-Roca, P., and González-Manteiga, W. (1987). A class of linear regression parameter estimators constructed by nonparametric estimation. *Annals of Statistics*, 15(2):603–609.

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

Cuesta-Albertos, J. A. and Febrero-Bande, M. (2010). A simple multiway ANOVA for functional data. *TEST*, 19(3):537–557.

Cuesta-Albertos, J. A. and Fraiman, R. (2006). Impartial trimmed means for functional data. In Liu, R. Y., Serfling, R., and Souvaine, D. L., editors, *Data Depth: Robust Multivariate Analysis, Computational Geometry and Applications*, volume 72 of *DIMACS: Series in Discrete Mathematics and Theoretical Computer Science*, pages 121–146, USA. American Mathematical Society.

Cuesta-Albertos, J. A. and Fraiman, R. (2007). Impartial trimmed k–means for functional data. *Computational Statistics & Data Analysis*, 51(10):4864–4877.

Cuesta-Albertos, J. A. and Nieto-Reyes, A. (2010). Functional Classification and the Random Tukey Depth. In Borgelt, C., Rodríguez, G. G., Trutschnig, W., Lubiano, M. A., Gil, M. A., Grzegorzewski, P., and Hryniewicz, O., editors, *Combining Soft Computing and Statistical Methods in Data Analysis*, volume 77 of *Advances in Intelligent and Soft Computing*, pages 123–130, Heidelberg. Springer.

Cuevas, A. (2012). A partial overview of the theory of statistics with functional data. Preprint.

Cuevas, A., Febrero, M., and Fraiman, R. (2002). Linear Functional Regression: The Case of Fixed Design and Functional Response. *Canadian Journal of Statistics*, 30(2):285–300.

Cuevas, A., Febrero, M., and Fraiman, R. (2004). An anova test for functional data. *Computational Statistics & Data Analysis*, 47(1):111–122.

Cuevas, A., Febrero, M., and Fraiman, R. (2006). On the use of the bootstrap for estimating functions with functional data. *Computational Statistics & Data Analysis*, 51(2):1063–1074.

Cuevas, A., Febrero, M., and Fraiman, R. (2007). Robust estimation and classification for functional data via projection–based depth notions. *Computational Statistics*, 22(3):481–496.

Dabo-Niang, S. (2004). Density estimation by orthogonal series in an infinite dimensional space: application to processes of diffusion type I. *Journal of Nonparametric Statistics*, 16(1–2):171–186.

Dabo-Niang, S. and Ferraty, F., editors (2008). *Functional and Operatorial Statistics*. Contributions to Statistics. Physica–Verlag, Heidelberg.

Dabo-Niang, S., Ferraty, F., and Vieu, P. (2006). Mode estimation for functional random variable and its application for curves classification. *Far East Journal of Theoretical Statistics*, 18(1):93–119.

Dabo-Niang, S., Ferraty, F., and Vieu, P. (2007). On the using of modal curves for radar waveforms classification. *Computational Statistics & Data Analysis*, 51(10):4878–4890.

Dabo-Niang, S., Yao, A. F., Pischedda, L., Cuny, P., and Gilbert, F. (2010). Spatial mode estimation for functional random fields with application to bioturbation problem. *Stochastic Environmental Research and Risk Assessment*, 24(4):487–497.

Damon, J. and Guillas, S. (2002). The inclusion of exogenous variables in functional autoregressive ozone forecasting. *Environmetrics*, 13(7):759–774.

Damon, J. and Guillas, S. (2005). Estimation and Simulation of Autoregressive Hilbertian Processes with Exogenous Variables. *Statistical Inference for Stochastic Processes*, 8(2):185–204.

Daubechies, I. (1992). *Ten Lectures on Wavelets*, volume 61 of *CBMS–NSF Series in Applied Mathematics*. Society for Industrial and Applied Mathematics, Philadelphia.

Dauxois, J., Nkiet, G. M., and Romain, Y. (2004). Canonical analysis relative to a closed subspace. *Linear Algebra and its Applications*, 388:119–145.

Dauxois, J. and Pousse, A. (1975). Une extension de l'analyse canonique. Quelques applications. *Annales de l'Institut Henri Poincare (B) Probability and Statistics*, 11(4):355–379.

Dauxois, J., Pousse, A., and Romain, Y. (1982). Asymptotic theory for the principal component analysis of a vector random function: Some applications to statistical inference. *Journal of Multivariate Analysis*, 12(1):136–154.

Davidian, M., Lin, X., and Wang, J. L. (2004). Introduction to the Emerging Issues in Longitudinal and Functional Data Analysis (with discussion). *Statistica Sinica*, 14(3):613–629.

de Boor, C. (2001). *A Practical Guide to Splines*, volume 27 of *Applied Mathematical Sciences*. Springer, New York, revised edition.

Delaigle, A. and Hall, P. (2010). Defining probability density for a distribution of random functions. *Annals of Statistics*, 38(2):1171–1193.

Delaigle, A. and Hall, P. (2012). Achieving near perfect classification for functional data. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 74(2):267–286.

Delicado, P. (2007). Functional k–sample problem when data are density functions. *Computational Statistics*, 22(3):391–410.

Delicado, P., Giraldo, R., Comas, C., and Mateu, J. (2010). Statistics for spatial functional data: some recent contributions. *Environmetrics*, 21(3–4):224–239.

Delsol, L. (2007). Régression non–paramétrique fonctionnelle : Expressions asymptotiques des moments. *Annales de l'Institut de Statistique de l'Université de Paris*, LI(3):43–67.

Delsol, L. (2009). Advances on asymptotic normality in non–parametric functional time series analysis. *Statistics: A Journal of Theoretical and Applied Statistics*, 43(1):13–33.

Delsol, L. (2011). Nonparametric methods for alpha–mixing functional data. In Ferraty, F. and Romain, Y., editors, *The Oxford Handbook of Functional Data Analysis*, Oxford Handbooks in Mathematics, pages 130–165. Oxford University Press, Oxford.

Delsol, L., Ferraty, F., and Martínez Calvo, A. (2011a). Functional Data Analysis: An Interdisciplinary Statistical Topic. In Gettler Summa, M., Bottou, L., Goldfarb, B., Murtagh, F., Pardoux, C., and Touati, M., editors, *Statistical Learning and Data Science*, Computer Science & Data Analysis, pages 189–195. Chapman and Hall/CRC, Boca Raton, USA.

Delsol, L., Ferraty, F., and Vieu, P. (2011b). Structural test in regression on functional variables. *Journal of Multivariate Analysis*, 102(3):422–447.

Deville, J. C. (1974). Méthodes statistiques et numériques de l'analyse harmonique. *Annales de l'Institute National de la Statistique et des Études Économiques*, 15:3–97.

Di, C. Z., Crainiceanu, C. M., Caffo, B. S., and Punjabi, N. M. (2009). Multilevel functional principal component analysis. *Annals of Applied Statistics*, 3(1):458–488.

Dudley, R. M. (1990). Nonlinear functionals of empirical measures and the bootstrap. In Eberlein, E., Kuelbs, J., and Marcus, M. B., editors, *Probability in Banach spaces, 7 (Oberwolfach, 1988)*, volume 21 of *Progress in Probability Series*, pages 63–82. Birkhäuser, Boston.

Efron, B. (1979). Bootstrap methods: another look at the jackknife. *Annals of Statistics*, 7(1):1–26.

Epifanio, I. and Ventura-Campos, N. (2011). Functional data analysis in shape analysis. *Computational Statistics & Data Analysis*, 55(9):2758–2773.

Erbas, B., Hyndman, R. J., and Gertig, D. M. (2007). Forecasting age–specific breast cancer mortality using functional data models. *Statistics in Medicine*, 26(2):458–470.

Escabias, M., Aguilera, A. M., and Valderrama, M. J. (2004). Principal component estimation of functional logistic regression: discussion of two different approaches. *Journal of Nonparametric Statistics*, 16(3–4):365–384.

Escabias, M., Aguilera, A. M., and Valderrama, M. J. (2005). Modeling environmental data by functional principal component logistic regression. *Environmetrics*, 16(1):95–107.

Escabias, M., Aguilera, A. M., and Valderrama, M. J. (2007). Functional PLS logit regression model. *Computational Statistics & Data Analysis*, 51(10):4891–4902.

Ezzahrioui, M. and Ould-Saïd, E. (2008a). Asymptotic normality of a nonparametric estimator of the conditional mode function for functional data. *Journal of Nonparametric Statistics*, 20(1):3–18.

Ezzahrioui, M. and Ould-Saïd, E. (2008b). Asymptotic Results of a Nonparametric Conditional Quantile Estimator for Functional Time Series. *Communications in Statistics – Theory and Methods*, 37(17):2735–2759.

Fan, J. and Zhang, J. (2004). Sieve empirical likelihood ratio tests for nonparametric functions. *Annals of Statistics*, 32(5):1858–1907.

Faraldo-Roca, P. and González-Manteiga, W. (1987). On efficiency of a new class of linear regression estimates obtained by preliminary non–parametric estimation. In Puri, M. L., Perez-Vilaplana, J., and Wertz, W., editors, *New Perspectives in Theoretical and Applied Statistics*, pages 229–242, New York. Wiley.

Faraway, J. J. (1997). Regression analysis for a functional response. *Technometrics*, 39(3):254–261.

Febrero-Bande, M., Galeano, P., and González-Manteiga, W. (2007). A Functional Analysis of NOx levels: location and scale estimation and outlier detection. *Computational Statistics*, 22(3):411–427.

Febrero-Bande, M., Galeano, P., and González-Manteiga, W. (2010). Measures of influence for the functional linear model with scalar response. *Journal of Multivariate Analysis*, 101(2):327–339.

Febrero-Bande, M. and Oviedo de la Fuente, M. (2011). *fda.usc: Functional Data Analysis and Utilities for Statistical Computing (fda.usc).* R package version 0.9.4 available at `http://CRAN.R-project.org/package=fda.usc`.

Fernández de Castro, B. and González-Manteiga, W. (2008). Boosting for real and functional samples: an application to an environmental problem. *Stochastic Environmental Research and Risk Assessment*, 22(1):27–37.

Fernández de Castro, B., Guillas, S., and González-Manteiga, W. (2005). Functional Samples and Bootstrap for Predicting Sulfur Dioxide Levels. *Technometrics*, 47(2):212–222.

Ferraty, F. (2010). High–dimensional data: a fascinating statistical challenge. *Journal of Multivariate Analysis*, 101(2):305–306. Statistical Methods and Problems in Infinite–dimensional Spaces, 1st International Workshop on Functional and Operatorial Statistics (IWFOS'2008).

Ferraty, F., editor (2011). *Recent Advances in Functional Data Analysis and Related Topics.* Contributions to Statistics. Physica–Verlag, Heidelberg.

Ferraty, F., Goia, A., and Vieu, P. (2002). Functional nonparametric model for time series: a fractal approach for dimension reduction. *TEST*, 11(2):317–344.

Ferraty, F., González-Manteiga, W., Martínez-Calvo, A., and Vieu, P. (2012a). Presmoothing in functional linear regression. *Statistica Sinica*, 22(1):69–94.

Ferraty, F., Hall, P., and Vieu, P. (2010a). Most–predictive design points for functional data predictors. *Biometrika*, 97(4):807–824.

Ferraty, F., Laksaci, A., Tadj, A., and Vieu, P. (2010b). Rate of uniform consistency for nonparametric estimates with functional variables. *Journal of Statistical Planning and Inference*, 140:335–352.

Ferraty, F., Laksaci, A., Tadj, A., and Vieu, P. (2011a). Kernel regression with functional response. *Electronic Journal of Statistics*, 5:159–171.

Ferraty, F., Laksaci, A., and Vieu, P. (2006). Estimating Some Characteristics of the Conditional Distribution in Nonparametric Functional Models. *Statistical Inference for Stochastic Processes*, 9(1):47–76.

Ferraty, F., Martínez-Calvo, A., and Vieu, P. (2011b). Thresholding in Nonparametric Functional Regression with Scalar Response. In Ferraty, F., editor, *Recent Advances in Functional Data Analysis and Related Topics*, Contributions to Statistics, pages 103–109. Physica–Verlag, Heidelberg.

Ferraty, F., Martínez-Calvo, A., and Vieu, P. (2012b). Thresholding in Nonparametric Functional Regression: an Exploratory Tool. Preprint.

Ferraty, F., Mas, A., and Vieu, P. (2007a). Nonparametric regression on functional data: inference and practical aspects. *Australian & New Zealand Journal of Statistics*, 49(3):267–286.

Ferraty, F., Quintela del Río, A., and Vieu, P. (2012c). Specification test for conditional distribution with functional data. *Econometric Theory*, 28:363–386.

Ferraty, F., Rabhi, A., and Vieu, P. (2005). Conditional Quantiles for Dependent Functional Data with Application to the Climatic El Niño Phenomenon. *Sankhyā: The Indian Journal of Statistics*, 67(2):378–398.

Ferraty, F., Rabhi, A., and Vieu, P. (2008). Estimation non–paramétrique de la fonction de hasard avec variable explicative fonctionnelle. *Revue Roumaine de Mathématiques Pures et Appliquées*, 53(1):1–18.

Ferraty, F. and Romain, Y., editors (2011). *The Oxford Handbook of Functional Data Analysis*. Oxford Handbooks in Mathematics. Oxford University Press, Oxford.

Ferraty, F., Van Keilegom, I., and Vieu, P. (2010c). On the validity of the bootstrap in non–parametric functional regression. *Scandinavian Journal of Statistics*, 37(2):286–306.

Ferraty, F., Van Keilegom, I., and Vieu, P. (2012d). Regression when both response and predictor are functions. *Journal of Multivariate Analysis*, 109:10–28.

Ferraty, F. and Vieu, P. (2002). The functional nonparametric model and application to spectrometric data. *Computational Statistics*, 17(4):545–564.

Ferraty, F. and Vieu, P. (2003). Curves discrimination: a nonparametric functional approach. *Computational Statistics & Data Analysis*, 44(1–2):161–173.

Ferraty, F. and Vieu, P. (2004). Nonparametric models for functional data, with application in regression, time series prediction and curve discrimination. *Journal of Nonparametric Statistics*, 16(1):111–125.

Ferraty, F. and Vieu, P. (2006a). Functional Nonparametric Statistics in Action. In Sperlich, S., Härdle, W., and Aydinli, G., editors, *The Art of Semiparametrics*, Contributions to Statistics, pages 112–129. Physica–Verlag, Heidelberg.

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

Ferraty, F. and Vieu, P. (2008). Erratum: "Nonparametric models for functional data, with application in regression, time series prediction and curve discrimination". *Journal of Nonparametric Statistics*, 20(2):187–189.

Ferraty, F. and Vieu, P. (2009). Additive prediction and boosting for functional data. *Computational Statistics & Data Analysis*, 53(4):1400–1413.

Ferraty, F. and Vieu, P. (2011a). Kernel regression estimation for functional data. In Ferraty, F. and Romain, Y., editors, *The Oxford Handbook of Functional Data Analysis*, Oxford Handbooks in Mathematics, pages 72–129. Oxford University Press, Oxford.

Ferraty, F. and Vieu, P. (2011b). Unifying presentation for functional regression modelling. In Ferraty, F. and Romain, Y., editors, *The Oxford Handbook of Functional Data Analysis*, Oxford Handbooks in Mathematics, pages 3–20. Oxford University Press, Oxford.

Ferraty, F., Vieu, P., and Viguier-Pla, S. (2007b). Factor–based comparison of groups of curves. *Computational Statistics & Data Analysis*, 51(10):4903–4910.

Ferré, L. and Villa, N. (2006). Multilayer Perceptron with Functional Inputs: an Inverse Regression Approach. *Scandinavian Journal of Statistics*, 33(4):807–823.

Ferré, L. and Yao, A. F. (2005). Smoothed functional inverse regression. *Statistica Sinica*, 15(3):665–683.

Fraiman, R. and Muniz, G. (2001). Trimmed means for functional data. *TEST*, 10(2):419–440.

Fraiman, R. and Pateiro-López, B. (2012). Quantiles for finite and infinite dimensional data. *Journal of Multivariate Analysis*, 108:1–14.

Freedman, D. A. (1981). Bootstrapping regression models. *Annals of Statistics*, 9(6):1218–1228.

Gabrys, R., Horváth, L., and Kokoszka, P. (2010). Tests for Error Correlation in the Functional Linear Model. *Journal of the American Statistical Association*, 105(491):1113–1125.

Gadiaga, D. and Ignaccolo, R. (2005). Test of no–effect hypothesis by nonparametric regression. *Afrika Statistika*, 1(1):67–76.

Gamboa, F., Loubes, J. M., and Maza, E. (2007). Semi–parametric estimation of shifts. *Electronic Journal of Statistics*, 1:616–640.

García-Portugués, E., González-Manteiga, W., and Febrero-Bande, M. (2012). A goodness–of–fit test for the functional linear model with scalar response. *ArXiv e–prints*. Available at `http://arxiv.org/abs/1205.6167`.

Gasser, T., Hall, P., and Presnell, B. (1998). Nonparametric estimation of the mode of a distribution of random curves. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 60(4):681–691.

Gasser, T. and Kneip, A. (1995). Searching for Structure in Curve Sample. *Journal of the American Statistical Association*, 90(432):1179–1188.

Geenens, G. (2011). Curse of dimensionality and related issues in nonparametric functional regression. *Statistics Surveys*, 5:30–43.

Gervini, D. (2006). Free–Knot Spline Smoothing for Functional Data. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 68(4):671–687.

Gervini, D. (2008). Robust functional estimation using the median and spherical principal components. *Biometrika*, 95(3):587–600.

Giné, E. (1997). Lectures on some aspects of the bootstrap. In Pierre, B., editor, *Lectures on Probability Theory and Statistics (Saint–Flour, 1996)*, volume 1665 of *Lecture Notes in Mathematics*, pages 37–151. Springer, Berlin.

Giné, E. and Zinn, J. (1990). Bootstrapping General Empirical Measures. *Annals of Probability*, 18(2):851–869.

Giraldo, R., Delicado, P., and Mateu, J. (2010a). Continuous Time–Varying Kriging for Spatial Prediction of Functional Data: An Environmental Application. *Journal of Agricultural, Biological, and Environmental Statistics*, 15(1):66–82.

Giraldo, R., Delicado, P., and Mateu, J. (2010b). Ordinary kriging for function–valued spatial data. *Environmental and Ecological Statistics*, pages 1–16.

Goldsmith, J., Bobb, J., Crainiceanu, C. M., Caffo, B., and Reich, D. (2011). Penalized Functional Regression. *Journal of Computational and Graphical Statistics*, 20(4):830–851.

Goldsmith, J., Crainiceanu, C. M., Caffo, B., and Reich, D. (2012). Longitudinal penalized functional regression for cognitive outcomes on neuronal tract measurements. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 61(3):453–469.

González-Manteiga, W., González-Rodríguez, G., Martínez-Calvo, A., and García-Portugués, E. (2012). Bootstrap independence test for functional linear models. *ArXiv e–prints*. Available at `http://arxiv.org/abs/1210.1072`.

González-Manteiga, W. and Martínez-Calvo, A. (2010). Bootstrap Calibration in Functional Linear Regression Models with Applications. In Lechevallier, Y. and Saporta, G., editors, *Proceedings of COMPSTAT'2010. 19th International Conference on Computational Statistics. Paris, France, August 22–27, 2010*, pages 199–207, Heidelberg. Physica–Verlag.

González-Manteiga, W. and Martínez-Calvo, A. (2011). Bootstrap in functional linear regression. *Journal of Statistical Planning and Inference*, 141(1):453–461.

González-Manteiga, W., Martínez-Miranda, M. D., and Pérez-González, A. (2004). The choice of smoothing parameter in nonparametric regression through Wild Bootstrap. *Computational Statistics & Data Analysis*, 47(3):487–515.

González-Manteiga, W. and Vieu, P. (2007). Statistics for Functional Data. *Computational Statistics & Data Analysis*, 51(10):4788–4792.

González-Manteiga, W. and Vieu, P. (2011). Methodological Richness of Functional Data Analysis. In Gettler Summa, M., Bottou, L., Goldfarb, B., Murtagh, F., Pardoux, C., and Touati, M., editors, *Statistical Learning and Data Science*, Computer Science & Data Analysis, pages 197–203. Chapman and Hall/CRC, Boca Raton, USA.

González-Rodríguez, G., Colubi, A., and Gil, M. A. (2012). Fuzzy data treated as functional data: A one–way ANOVA test approach. *Computational Statistics & Data Analysis*, 56(4):943–955.

Gou, Z. and Fyfe, C. (2004). A canonical correlation neural network for multicollinearity and functional data. *Neural Networks*, 17(2):285–293.

Grenander, U. (1981). *Abstract inference*. Probability & Mathematical Statistics. John Wiley & Sons, New York.

Guillas, S. (2001). Rates of convergence of autocorrelation estimates for autoregressive Hilbertian processes. *Statistics & Probability Letters*, 55(3):281–291.

Guillas, S. (2002). Doubly Stochastic Hilbertian Processes. *Journal of Applied Probability*, 39(3):566–580.

Guillas, S. and Lai, M. J. (2010). Bivariate splines for spatial functional regression models. *Journal of Nonparametric Statistics*, 22(4):477–497.

Guo, W. (2002). Functional Mixed Effects Models. *Biometrics*, 58(1):121–128.

Hall, P. (1988). Theoretical comparison of bootstrap confidence intervals. *Annals of Statistics*, 16(3):927–953.

Hall, P. (2011). Functional principal component analysis. In Ferraty, F. and Romain, Y., editors, *The Oxford Handbook of Functional Data Analysis*, Oxford Handbooks in Mathematics, pages 210–234. Oxford University Press, Oxford.

Hall, P. and Heckman, N. E. (2002). Estimating and Depicting the Structure of a Distribution of Random Functions. *Biometrika*, 89(1):145–158.

Hall, P. and Horowitz, J. L. (2007). Methodology and convergence rates for functional linear regression. *Annals of Statistics*, 35(1):70–91.

Hall, P. and Hosseini-Nasab, M. (2006). On properties of functional principal components analysis. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 68(1):109–126.

Hall, P. and Hosseini-Nasab, M. (2009). Theory for high–order bounds in functional principal components analysis. *Mathematical proceedings of the Cambridge Philosophical Society*, 146(1):225–256.

Hall, P., Müller, H. G., and Wang, J. L. (2006). Properties of principal component methods for functional and longitudinal data analysis. *Annals of Statistics*, 34(3):1493–1517.

Hall, P., Poskitt, D. S., and Presnell, B. (2001). A Functional Data–Analytic Approach to Signal Discrimination. *Technometrics*, 43(1):1–9.

Hall, P. and Van Keilegom, I. (2007). Two–sample tests in functional data analysis starting from discrete data. *Statistica Sinica*, 17(4):1511–1531.

Hall, P. and Vial, C. (2006a). Assessing extrema of empirical principal component functions. *Annals of Statistics*, 34(3):1518–1544.

Hall, P. and Vial, C. (2006b). Assessing the finite dimensionality of functional data. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 68(4):689–705.

Härdle, W. and Marron, J. S. (1985). Optimal bandwidth selection in nonparametric regression function estimation. *Annals of Statistics*, 13(4):1465–1481.

Härdle, W. and Marron, J. S. (1990). Semiparametric comparison of regression curves. *Annals of Statistics*, 18(1):63–89.

Härdle, W., Mori, Y., and Vieu, P., editors (2007). *Statistical Methods for Biostatistics and Related Fields*. Springer, Heidelberg.

Harezlak, J., Coull, B. A., Laird, N. M., Magari, S. R., and Christiani, D. C. (2007). Penalized solutions to functional regression problems. *Computational Statistics & Data Analysis*, 51(10):4911–4925.

Hastie, T., Buja, A., and Tibshirani, R. (1995). Penalized Discriminant Analysis. *Annals of Statistics*, 23(1):73–102.

Hastie, T. and Mallows, C. (1993). A Statistical View of Some Chemometrics Regression Tools: Discussion. *Technometrics*, 35(2):140–143.

Hastie, T. and Tibshirani, R. (1993). Varying–coefficient models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 55(4):757–796.

Hastie, T., Tibshirani, R., and Friedman, J. H. (2001). *The Elements of Statistical Learning. Data Mining, Inference, and Prediction*. Springer Series in Statistics. Springer, New York.

He, G., Müller, H. G., and Wang, J. L. (2000). Extending correlation and regression from multivariate to functional data. In Puri, M. L., editor, *Asymptotics in statistics and probability*, pages 301–315. VSP International Science Publishers, Zeist, The Neatherlands.

He, G., Müller, H. G., and Wang, J. L. (2003). Functional canonical analysis for square integrable stochastic processes. *Journal of Multivariate Analysis*, 85(1):54–77.

He, G., Müller, H. G., and Wang, J. L. (2004). Methods of canonical analysis for functional data. *Journal of Statistical Planning and Inference*, 122(1–2):141–159.

He, G., Müller, H. G., Wang, J. L., and Yang, W. (2010). Functional linear regression via canonical analysis. *Bernoulli*, 16(3):705–729.

Heckman, N. E. and Zamar, R. H. (2000). Comparing the shapes of regression functions. *Biometrika*, 87(1):135–144.

Henderson, B. (2006). Exploring between site differences in water quality trends: a functional data analysis approach. *Environmetrics*, 17(1):65–80.

Hilgert, N., Mas, A., and Verzelen, N. (2012). Minimax adaptive tests for the Functional Linear model. *ArXiv e–prints*. Available at `http://arxiv.org/abs/1206.1194`.

Hitchcock, D. B., Casella, G., and Booth, J. G. (2006). Improved estimation of dissimilarities by presmoothing functional data. *Journal of the American Statistical Association*, 101(473):211–222.

Hlubinka, D. and Prchal, L. (2007). Changes in atmospheric radiation from the statistical point of view. *Computational Statistics & Data Analysis*, 51(10):4926–4941.

Hoover, D. R., Rice, J. A., Wu, C. O., and Yang, L. P. (1998). Nonparametric Smoothing Estimates of Time–Varying Coefficient Models with Longitudinal Data. *Biometrika*, 85(4):809–822.

Hörmann, S., Horváth, L., and Reeder, R. (2012). A Functional Version of the ARCH Model. *Econometric Theory*, page To appear.

Hörmann, S. and Kokoszka, P. (2010). Weakly dependent functional data. *Annals of Statistics*, 38(3):1845–1884.

Horváth, L., Hušková, M., and Kokoszka, P. (2010). Testing the stability of the functional autoregressive process. *Journal of Multivariate Analysis*, 101(2):352–367.

Horváth, L. and Kokoszka, P. (2012). *Inference for Functional Data with Applications*, volume 200 of *Springer Series in Statistics*. Springer, New York.

Horváth, L., Kokoszka, P., and Reeder, R. (2012). Estimation of the mean of functional time series and a two–sample problem. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 75(1):103–122.

Horváth, L., Kokoszka, P., and Reimherr, M. (2009). Two sample inference in functional linear models. *Canadian Journal of Statistics*, 37(4):571–591.

Horváth, L. and Reeder, R. (2011). A test of significance in functional quadratic regression. *ArXiv e–prints*. Available at `http://arxiv.org/abs/1105.0014`.

Horváth, L. and Reeder, R. (2012). Detecting changes in functional linear models. *Journal of Multivariate Analysis*, 111:310–334.

Hyndman, R. J. and Ullah, M. S. (2007). Robust forecasting of mortality and fertility rates: A functional data approach. *Computational Statistics & Data Analysis*, 51(10):4942–4956.

Jacob, P. and Oliveira, P. E. (1995). A General Approach To Nonparametric Histogram Estimation. *Statistics: A Journal of Theoretical and Applied Statistics*, 27(1):73–92.

Jácome, M., Gijbels, I., and Cao, R. (2008). Comparison of presmoothing methods in kernel density estimation under censoring. *Computational Statistics*, 23(3):381–406.

Jácome, M. A. and Iglesias-Pérez, M. C. (2008). Presmoothed estimation with left–truncated and right–censored data. *Comunications in Statistics – Theory and Methods*, 37(18):2964–2983.

James, G. (2011). Sparse functional data analysis. In Ferraty, F. and Romain, Y., editors, *The Oxford Handbook of Functional Data Analysis*, Oxford Handbooks in Mathematics, pages 298–324. Oxford University Press, Oxford.

James, G. M. (2002). Generalized Linear Models with Functional Predictors. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64(3):411–432.

James, G. M. and Hastie, T. J. (2001). Functional Linear Discriminant Analysis for Irregularly Sampled Curves. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63(3):533–550.

James, G. M., Hastie, T. J., and Sugar, C. A. (2000). Principal component models for sparse functional data. *Biometrika*, 87(3):587–602.

James, G. M. and Sood, A. (2006). Performing hypothesis tests on the shape of functional data. *Computational Statistics & Data Analysis*, 50(7):1774–1792.

James, G. M. and Sugar, C. A. (2003). Clustering for Sparsely Sampled Functional Data. *Journal of the American Statistical Association*, 98(462):397–408.

Jank, W. and Shmueli, G. (2006). Functional Data Analysis in Electronic Commerce Research. *Statistical Science*, 21(2):155–166.

Janssen, P., Swanepoel, J., and Veraverbeke, N. (2001). Efficiency of linear regression estimators based on presmoothing. *Communications in Statistics – Theory and Methods*, 30(10):2079–2097.

Kawasaki, Y. and Ando, T. (2004). Functional data analysis of the dynamics of yield curves. In Antoch, J., editor, *COMPSTAT 2004 – Proceedings in Computational Statistics. 16th Symposium Held in Prague, Czech Republic, 2004*, pages 1285–1292, Heidelberg. Physica–Verlag.

Kemperman, J. H. B. (1987). The median of a finite measure on a Banach space. In Dodge, Y., editor, *Statistical Data Analysis Based on the L1–Norm and Related Methods*, pages 217–230, Amsterdam. North–Holland.

Kim, K., Sentürk, D., and Li, R. (2011). Recent history functional linear models for sparse longitudinal data. *Journal of Statistical Planning and Inference*, 141(4):1554–1566.

Kneip, A. and Engel, J. (1995). Model Estimation in Nonlinear Regression Under Shape Invariance. *Annals of Statistics*, 23(2):551–570.

Kneip, A. and Gasser, T. (1988). Convergence and Consistency Results for Self–Modeling Nonlinear Regression. *Annals of Statistics*, 16(1):82–112.

Kneip, A. and Gasser, T. (1992). Statistical Tools to Analyze Data Representing a Sample of Curves. *Annals of Statistics*, 20(3):1266–1305.

Kneip, A., Li, X., MacGibbon, K. B., and Ramsay, J. O. (2000). Curve Registration by Local Regression. *Canadian Journal of Statistics*, 28(1):19–29.

Kneip, A. and Sarda, P. (2011). Factor models and variable selection in high–dimensional regression analysis. *Annals of Statistics*, 39(5):2410–2447.

Kneip, A., Sickles, R. C., and Song, W. (2004). Functional data analysis and mixed effect models. In Antoch, J., editor, *COMPSTAT 2004 – Proceedings in Computational Statistics. 16th Symposium Held in Prague, Czech Republic, 2004*, pages 315–326, Heidelberg. Physica–Verlag.

Kneip, A. and Utikal, K. J. (2001). Inference for Density Families Using Functional Principal Component Analysis. *Journal of the American Statistical Association*, 96(454):519–542.

Kokoszka, P., Maslova, I., Sojka, J., and Zhu, L. (2008). Testing for lack of dependence in the functional linear model. *Canadian Journal of Statistics*, 36(2):207–222.

Kosorok, M. R. (2008). *Introduction to Empirical Processes and Semiparametric Inference*. Springer Series in Statistics. Springer, New York.

Laha, R. G. and Rohatgi, V. K. (1979). *Probability Theory*. Wiley Series in Probability and Statistics. Wiley, New York.

Laukaitis, A. (2008). Functional data analysis for cash flow and transactions intensity continuous–time prediction using Hilbert–valued autoregressive processes. *European Journal of Operational Research*, 185(3):1607–1614.

Laukaitis, A. and Rackauskas, A. (2002). Functional data analysis of payment systems. *Nonlinear Analysis: Modelling and Control*, 7(2):53–68.

Leardi, R., editor (2003). *Nature–inspired methods in chemometrics: genetic algorithms and artificial neural networks*. Data Handling in Science and Technology. Elsevier Science, New York.

Ledoux, M. and Talagrand, M. (1988). Un critère sur les petites boules dans le théorème limite central. *Probab. Theory Related Fields*, 77(1):29–47.

Lee, E. R. and Park, B. U. (2012). Sparse estimation in functional linear regression. *Journal of Multivariate Analysis*, 105(1):1–17.

Lee, S. Y., Zhang, W., and Song, X. Y. (2002). Estimating the covariance function with functional data. *British Journal of Mathematical and Statistical Psychology*, 55(2):247–261.

Leng, X. and Müller, H. G. (2005). Classification using functional data analysis for temporal gene expression data. *Bioinformatics*, 22(1):68–76.

Leurgans, S. E., Moyeed, R. A., and Silverman, B. W. (1993). Canonical Correlation Analysis when the Data are Curves. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 55(3):725–740.

Li, J., Cuesta-Albertos, J. A., and Liu, R. Y. (2012). DD–Classifier: Nonparametric Classification Procedure Based on DD–Plot. *Journal of the American Statistical Association*, 107(498):737–753.

Li, W. V. and Shao, Q. M. (2001). Gaussian processes: Inequalities, small ball probabilities and applications. In Shanbhag, D. and Rao, C., editors, *Stochastic Processes: Theory and Methods*, volume 19 of *Handbook of Statistics*, pages 533–597. Elsevier, New York.

Li, Y. and Hsing, T. (2010). Uniform convergence rates for nonparametric regression and principal component analysis in functional/longitudinal data. *Annals of Statistics*, 38(6):3321–3351.

Li, Y., Wang, N., and Carrol, R. J. (2010). Generalized Functional Linear Models With Semiparametric Single–Index Interactions. *Journal of the American Statistical Association*, 105(490):621–633.

Lian, H. (2011). Functional partial linear model. *Journal of Nonparametric Statistics*, 23(1):115–128.

Lindquist, M. A. and McKeague, I. W. (2009). Logistic Regression With Brownian–Like Predictors. *Journal of the American Statistical Association*, 104(488):1575–1585.

Liu, B. and Müller, H. G. (2009). Estimating Derivatives for Samples of Sparsely Observed Functions, With Application to Online Auction Dynamics. *Journal of the American Statistical Association*, 104(486):704–717.

Liu, X. and Yang, M. C. K. (2009). Simultaneous curve registration and clustering for functional data. *Computational Statistics & Data Analysis*, 53(4):1361–1376.

Locantore, N., Marron, J. S., Simpson, D. G., Tripoli, N., Zhang, J. T., and Cohen, K. L. (1999). Robust principal component analysis for functional data. *TEST*, 8(1):1–73.

López-Pintado, S. and Jornsten, R. (2007). Functional analysis via extensions of the band depth. In Liu, R., Strawderman, W., and Zhang, C. H., editors, *Complex Datasets and Inverse Problems: Tomography, Networks and Beyond*, volume 54 of *IMS Lecture Notes–Monograph Series*, pages 103–120. Institute of Mathematical Statistics, Beachwood, Ohio, USA.

López-Pintado, S. and Romo, J. (2006). Depth–based classification for functional data. In Liu, R. Y., Serfling, R., and Souvaine, D. L., editors, *Data Depth: Robust Multivariate Analysis, Computational Geometry and Applications*, volume 72 of *DIMACS: Series in Discrete Mathematics and Theoretical Computer Science*, pages 103–119, USA. American Mathematical Society.

López-Pintado, S. and Romo, J. (2007). Depth–based inference for functional data. *Computational Statistics & Data Analysis*, 51(10):4957–4968.

López-Pintado, S. and Romo, J. (2009). On the Concept of Depth for Functional Data. *Journal of the American Statistical Association*, 104(486):718–734.

Lucero, J. C. (1999). Computation of the harmonics–to–noise ratio of a voice signal using a functional data analysis algorithm. *Journal of Sound and Vibration*, 222(3):512–520.

Malfait, N. and Ramsay, J. O. (2003). The Historical Functional Linear Model. *Canadian Journal of Statistics*, 31(2):115–128.

Mammen, E. (2000). Resampling Methods for Nonparametric Regression. In Schimek, M. G., editor, *Smoothing and Regression: Approaches, Computation, and Application*, pages 425–450, New York. Wiley.

Manté, C., Yao, A. F., and Degiovanni, C. (2007). Principal component analysis of measures, with special emphasis on grain–size curves. *Computational Statistics & Data Analysis*, 51(10):4969–4983.

Marron, J. S. and Härdle, W. (1986). Random Approximations to some measures of accuracy in nonparametric curve estimation. *Journal of Multivariate Analysis*, 20(1):91–113.

Martínez-Calvo, A. (2008). Presmoothing in functional linear regression. In Dabo-Niang, S. and Ferraty, F., editors, *Functional and Operatorial Statistics*, Contributions to Statistics, pages 223–229, Heidelberg. Physica–Verlag.

Mas, A. (2002). Weak convergence for the covariance operators of a Hilbertian linear process. *Stochastic Processes and their Applications*, 99(1):117–135.

Mas, A. (2007a). Testing for the Mean of Random Curves: A Penalization Approach. *Statistical Inference for Stochastic Processes*, 10(2):147–163.

Mas, A. (2007b). Weak convergence in the functional autoregressive model. *Journal of Multivariate Analysis*, 98(6):1231–1261.

Mas, A. and Menneteau, L. (2003). Large and moderate deviations for infinite–dimensional autoregressive processes. *Journal of Multivariate Analysis*, 87(2):241–260.

Mas, A. and Pumo, B. (2007). The ARHD model. *Journal of Statistical Planning and Inference*, 137(2):538–553.

Mas, A. and Pumo, B. (2011). Linear processes for functional data. In Ferraty, F. and Romain, Y., editors, *The Oxford Handbook of Functional Data Analysis*, Oxford Handbooks in Mathematics, pages 47–71. Oxford University Press, Oxford.

Masry, E. (2005). Nonparametric regression estimation for dependent functional data: asymptotic normality. *Stochastic Processes and their Applications*, 115(1):155–177.

McKeague, I. W. and Sen, B. (2010). Fractals with point impact in functional linear regression. *Annals of Statistics*, 38(4):2559–2586.

McMurry, T. and Politis, D. (2011). Resampling methods for functional data. In Ferraty, F. and Romain, Y., editors, *The Oxford Handbook of Functional Data Analysis*, Oxford Handbooks in Mathematics, pages 189–209. Oxford University Press, Oxford.

Meiring, W. (2007). Oscillations and Time Trends in Stratospheric Ozone Levels. *Journal of the American Statistical Association*, 102(479):788–802.

Menneteau, L. (2005). Some laws of the iterated logarithm in Hilbertian autoregressive models. *Journal of Multivariate Analysis*, 92(2):405–425.

Merlevède, F., Peligrad, M., and Utev, S. (1997). Sharp Conditions for the CLT of Linear Processes in a Hilbert Space. *Journal of Theoretical Probability*, 10(3):681–693.

Mizuta, M. (2004). Clustering methods for functional data: k–means, single linkage and moving clustering. In Antoch, J., editor, *COMPSTAT 2004 – Proceedings in Computational Statistics. 16th Symposium Held in Prague, Czech Republic, 2004*, pages 1479–1486, Heidelberg. Physica–Verlag.

Morris, J. S. and Carroll, R. J. (2006). Wavelet–based functional mixed models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 68(2):179–199.

Morris, J. S., Vannucci, M., J., B. P., and Carroll, R. J. (2003). Wavelet–Based Nonparametric Modeling of Hierarchical Functions in Colon Carcinogenesis. *Journal of the American Statistical Association*, 98(463):573–583.

Mourid, T. (2002). Estimation and Prediction of Functional Autoregressive Processes. *Statistics: A Journal of Theoretical and Applied Statistics*, 36(2):125–138.

Müller, H. G. (2005). Functional Modelling and Classification of Longitudinal Data. *Scandinavian Journal of Statistics*, 32(2):223–240.

Müller, H. G., Sen, R., and Stadtmüller, U. (2011). Functional data analysis for volatility. *Journal of Econometrics*, 165(2):233–245.

Müller, H. G. and Stadtmüller, U. (2005). Generalized functional linear models. *Annals of Statistics*, 33(2):774–805.

Müller, H. G. and Yang, W. (2010). Dynamic relations for sparsely sampled Gaussian processes. *TEST*, 19(1):1–29.

Müller, H. G. and Zhang, Y. (2005). Time–Varying Functional Regression for Predicting Remaining Lifetime Distributions from Longitudinal Trajectories. *Biometrics*, 61(4):1064–1075.

Nerini, D. and Ghattas, B. (2007). Classifying densities using functional regression trees: Applications in oceanology. *Computational Statistics & Data Analysis*, 51(10):4984–4993.

Ocaña, F. A., Aguilera, A. M., and Escabias, M. (2007). Computational considerations in functional principal component analysis. *Computational Statistics*, 22(3):449–465.

Ocaña, F. A., Aguilera, A. M., and Valderrama, M. J. (1999). Functional Principal Components Analysis by Choice of Norm. *Journal of Multivariate Analysis*, 71(2):262–276.

Opgen-Rhein, R. and Strimmer, K. (2006). Infering gene dependency networks from genomic longitudinal data: a functional data approach. *REVSTAT–Statistical Journal*, 4(1):53–65.

Ormoneit, D., Black, M. J., Hastie, T., and Kjellström, H. (2005). Representing cyclic human motion using functional analysis. *Image and Vision Computing*, 23(14):1264–1276.

Park, J., Gasser, T., and Rousson, V. (2009). Structural components in functional data. *Computational Statistics & Data Analysis*, 53(9):3452–3465.

Parker, B. and Wen, J. (2009). Predicting microRNA targets in time–series microarray experiments via functional data analysis. *BMC Bioinformatics*, 10(Suppl 1):S32.

Parr, W. C. (1985). The bootstrap: Some large sample theory and connections with robustness. *Statistics & Probability Letters*, 3(2):97–100.

Patilea, V., Sánchez-Sellero, C., and Saumard, M. (2012). Projection–based nonparametric goodness–of–fit testing with functional covariates. *ArXiv e–prints*. Available at `http://arxiv.org/abs/1205.5578`.

Pezzulli, S. and Silverman, B. W. (1993). Some properties of smoothed principal components analysis for functional data. *Computational Statistics*, 8:1–16.

Politis, D. N. and Romano, J. P. (1994). Limit theorems for weakly dependent Hilbert space valued random variables with application to the stationary bootstrap. *Statistica Sinica*, 4(2):461–467.

Prchal, L. and Sarda, P. (2007). Spline estimator for the functional linear regression with functional response. Preprint.

Preda, C. (2007). Regression models for functional data by reproducing kernel Hilbert spaces methods. *Journal of Statistical Planning and Inference*, 137(3):829–840.

Preda, C. and Saporta, G. (2005a). Clusterwise PLS regression on a stochastic process. *Computational Statistics & Data Analysis*, 49(1):99–108.

Preda, C. and Saporta, G. (2005b). PLS regression on a stochastic process. *Computational Statistics & Data Analysis*, 48(1):149–158.

Preda, C., Saporta, G., and Lévéder, C. (2007). PLS classification of functional data. *Computational Statistics*, 22(2):223–235.

Pumo, B. (1998). Prediction of Continuous Time Processes by C[0,1]–Valued Autoregressive Process. *Statistical Inference for Stochastic Processes*, 1(3):297–309.

Quintela del Río, A. (2008). Hazard function given a functional variable: Non–parametric estimation under strong mixing conditions. *Journal of Nonparametric Statistics*, 20(5):413–430.

R Development Core Team (2010). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. Available at `http://www.R-project.org`.

Rachdi, M. and Vieu, P. (2007). Nonparametric regression for functional data: Automatic smoothing parameter selectiona. *Journal of Statistical Planning and Inference*, 137(9):2784–2801.

Ramsay, J. (2011). Curve registration. In Ferraty, F. and Romain, Y., editors, *The Oxford Handbook of Functional Data Analysis*, Oxford Handbooks in Mathematics, pages 235–258. Oxford University Press, Oxford.

Ramsay, J. O. (1982). When the data are functions. *Psychometrika*, 47(4):379–396.

Ramsay, J. O. (2000a). Differential Equation Models for Statistical Functions. *Canadian Journal of Statistics*, 28(2):225–240.

Ramsay, J. O. (2000b). Functional Components of Variation in Handwriting. *Journal of the American Statistical Association*, 95(449):9–15.

Ramsay, J. O., Bock, R. D., and Gasser, T. (1995). Comparison of height acceleration curves in the Fels, Zurich, and Berkeley growth data. *Annals of Human Biology*, 22(5):413–426.

Ramsay, J. O. and Dalzell, C. J. (1991). Some Tools for Functional Data Analysis. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 53(3):539–572.

Ramsay, J. O. and Li, X. (1998). Curve registration. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 60(2):351–363.

Ramsay, J. O., Munhall, K. G., Gracco, V. L., and Ostry, D. J. (1996). Functional data analyses of lip motion. *Journal of the Acoustical Society of America*, 99(6):3718–3727.

Ramsay, J. O. and Silverman, B. W. (1997). *Functional Data Analysis*. Springer Series in Statistics. Springer, New York.

Ramsay, J. O. and Silverman, B. W. (2002). *Applied Functional Data Analysis. Methods and Case Studies*. Springer Series in Statistics. Springer, New York.

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

Ramsay, J. O., Wickham, H., Graves, S., and Hooker, G. (2011). *fda: Functional Data Analysis*. R package version 2.2.6 available at `http://CRAN.R-project.org/package=fda`.

Rao, C. R. (1958). Some Statistical Methods for Comparison of Growth Curves. *Biometrics*, 14(1):1–17.

Ratcliffe, S. J., Leader, L. R., and Heller, G. Z. (2002a). Functional data analysis with application to periodically stimulated foetal heart rate data. I: Functional regression. *Statistics in Medicine*, 21(8):1103–1114.

Ratcliffe, S. J., Leader, L. R., and Heller, G. Z. (2002b). Functional data analysis with application to periodically stimulated foetal heart rate data. II: Functional logistic regression. *Statistics in Medicine*, 21(8):1115–1127.

Reddy, S. K. and Dass, M. (2006). Modeling On–Line Art Auction Dynamics Using Functional Data Analysis. *Statistical Science*, 21(2):179–193.

Reiss, P. T. and Ogden, R. T. (2007). Functional Principal Component Regression and Functional Partial Least Squares. *Journal of the American Statistical Association*, 102(479):984–996.

Rice, J. A. (2004). Functional and longitudinal data analysis: perspectives on smoothing. *Statistica Sinica*, 14(3):631–647.

Rice, J. A. and Silverman, B. W. (1991). Estimating the Mean and Covariance Structure Nonparametrically When the Data are Curves. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 53(1):223–243.

Rønn, B. B. (2001). Nonparametric Maximum Likelihood Estimation for Shifted Curves. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63(2):243–259.

Rossi, F. and Conan-Guez, B. (2005). Functional Multi–Layer Perceptron: a Nonlinear Tool for Functional Data Analysis. *Neural Networks*, 18(1):45–60.

Rossi, F. and Conan-Guez, B. (2006). Theoretical Properties of Projection Based Multilayer Perceptrons with Functional Inputs. *Neural Processing Letters*, 23(1):55–70.

Rossi, F., Conan-Guez, B., and El Golli, A. (2004). Clustering functional data with the SOM algorithm. In *Proceedings of XIIth European Symposium on Artificial Neural Networks (ESANN 2004)*, pages 305–312, Bruges, Belgium.

Rossi, F., Delannay, N., Conan-Guez, B., and Verleysen, M. (2005). Representation of functional data in neural networks. *Neurocomputing*, 64:183–210. Trends in Neurocomputing: 12th European Symposium on Artificial Neural Networks.

Rossi, F. and Villa, N. (2006). Support vector machine for functional data classification. *Neurocomputing*, 69(7–9):730–742. New Issues in Neurocomputing: 13th European Symposium on Artificial Neural Networks.

Rossi, N., Wang, X., and Ramsay, J. O. (2002). Nonparametric Item Response Function Estimates with the EM Algorithm. *Journal of Educational and Behavioral Statistics*, 27(3):291–317.

Ruiz-Medina, M. D. (2011). Spatial autoregressive and moving average Hilbertian processes. *Journal of Multivariate Analysis*, 102(2):292–305.

Ruiz-Medina, M. D. (2012a). New challenges in spatial and spatiotemporal functional statistics for high-dimensional data. *Spatial Statistics*, 1:82–91.

Ruiz-Medina, M. D. (2012b). Spatial functional prediction from spatial autoregressive Hilbertian processes. *Environmetrics*, 23(1):119–128.

Ruiz-Medina, M. D. and Espejo, R. M. (2012). Spatial autoregressive functional plug–in prediction of ocean surface temperature. *Stochastic Environmental Research and Risk Assessment*, 26(3):335–344.

Ruiz-Medina, M. D. and Salmerón, R. (2010). Functional maximum-likelihood estimation of ARH(p) models. *Stochastic Environmental Research and Risk Assessment*, 24(1):131–146.

Ruiz-Medina, M. D., Salmerón, R., and Angulo, J. M. (2007). Kalman filtering from POP–based diagonalization of ARH(1). *Computational Statistics & Data Analysis*, 51(10):4994–5008.

Schott, J. R. (2007). Some high–dimensional tests for a one–way MANOVA. *Journal of Multivariate Analysis*, 98(9):1825–1839.

Sguera, C., Galeano, P., and Lillo, R. E. (2012). Spatial depth–based classification for functional data. UC3M Working papers. Statistics and Econometrics, 12–06, Universidad Carlos III de Madrid.

Sheehy, A. and Wellner, J. A. (1992). Uniform Donsker Classes of Functions. *Annals of Probability*, 20(4):1983–2030.

Shen, Q. and Faraway, J. (2004). An F test for linear models with functional responses. *Statistica Sinica*, 14(4):1239–1257.

Shin, H. (2008). An extension of Fisher's discriminant analysis for stochastic processes. *Journal of Multivariate Analysis*, 99(6):1191–1216.

Shin, H. (2009). Partial functional linear regression. *Journal of Statistical Planning and Inference*, 139(10):3405–3418.

Silverman, B. W. (1996). Smoothed functional principal components analysis by choice of norm. *Annals of Statistics*, 24(1):1–24.

Singh, K. (1981). On the asymptotic accuracy of Efron's bootstrap. *Annals of Statistics*, 9(6):1187–1195.

Spitzner, D. J., Marron, J. S., and Essick, G. K. (2003). Mixed–Model Functional ANOVA for Studying Human Tactile Perception. *Journal of the American Statistical Association*, 98(462):263–272.

Staicu, A. M., Crainiceanu, C. M., and Carroll, R. J. (2010). Fast methods for spatially correlated multilevel functional data. *Biostatistics*, 11(2):177–194.

Tarpey, T. (2007). Linear Transformations and the k–Means Clustering Algorithm: Applications to Clustering Curves. *The American Statistician*, 61(1):34–40.

Tarpey, T. and Kinateder, K. K. J. (2003). Clustering Functional Data. *Journal of Classification*, 20(1):93–114.

Tarpey, T., Petkova, E., and Ogden, R. T. (2003). Profiling Placebo Responders by Self–Consistent Partitioning of Functional Data. *Journal of the American Statistical Association*, 98(464):850–858.

Tucker, L. (1958). Determination of parameters of a functional relation by factor analysis. *Psychometrika*, 23(1):19–23.

Valderrama, M. J. (2007). An overview to modelling functional data. *Computational Statistics*, 22(3):331–334.

Valderrama, M. J., Ocaña, F. A., and Aguilera, A. M. (2002). Forecasting PC–ARIMA Models for Functional Data. In Härdle, W. and Rönz, B., editors, *Compstat. Proceedings in Computational Statistics. 15th Symposium Held in Berlin, Germany, 2002*, pages 25–36, Heidelberg. Physica–Verlag.

Valderrama, M. J., Ortega-Moreno, M., González, P., and Aguilera, A. M. (2003). Derivation of a State–Space Model by Functional Data Analysis. *Computational Statistics*, 18(4):533–546.

van der Vaart, A. and Wellner, J. (1996). *Weak Convergence and Empirical Processes*. Springer Series in Statistics. Springer, New York.

Vardi, Y. and Zhang, C. H. (2000). The multivariate L1–median and associated data depth. *Proceedings of the National Academy of Sciences*, 97(4):1423–1426.

Vidakovic, B. (2001). Wavelet–based functional data analysis: theory, applications and ramifications. F3399. In Kobayashi, T., editor, *Proceedings of the Third Pacific Symposium on Flow Visualisation and Image Processing(PSFVIP–3)*.

Viele, K. (2001). Evaluating Fit in Functional Data Analysis Using Model Embeddings. *Canadian Journal of Statistics*, 29(1):51–66.

Vilar, J. M., Cao, R., and Aneiros, G. (2012). Forecasting next–day electricity demand and price using nonparametric functional methods. *International Journal of Electrical Power & Energy Systems*, 39(1):48–55.

Wang, G., Lin, N., and Zhang, B. (2012). Functional linear regression after spline transformation. *Computational Statistics & Data Analysis*, 56(3):587–601.

Wang, K. and Gasser, T. (1997). Alignment of curves by dynamic time warping. *Annals of Statistics*, 25(3):1251–1276.

Wang, K. and Gasser, T. (1998). Asymptotic and bootstrap confidence bounds for the structural average of curves. *Annals of Statistics*, 26(3):972–991.

Wang, K. and Gasser, T. (1999). Synchronizing sample curves nonparametrically. *Annals of Statistics*, 27(2):439–460.

Wang, S., Jank, W., and Shmueli, G. (2008). Explaining and Forecasting Online Auction Prices and Their Dynamics Using Functional Data Analysis. *Journal of Business & Economic Statistics*, 26(2):144–160.

Wu, C. O., Chiang, C. T., and Hoover, D. R. (1998). Asymptotic Confidence Regions for Kernel Smoothing of a Varying–Coefficient Model with Longitudinal Data. *Journal of the American Statistical Association*, 93(444):1388–1402.

Wu, Y., Fan, J., and Müller, H. G. (2010). Varying–coefficient functional linear regression. *Bernoulli*, 16(3):730–758.

Yao, F. (2007). Asymptotic distributions of nonparametric regression estimators for longitudinal or functional data. *Journal of Multivariate Analysis*, 98(1):40–56.

Yao, F. and Lee, T. C. M. (2006). Penalized spline models for functional principal component analysis. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 68(1):3–25.

Yao, F., Müller, H. G., and Wang, J. L. (2005a). Functional Data Analysis for Sparse Longitudinal Data. *Journal of the American Statistical Association*, 100(470):577–590.

Yao, F., Müller, H. G., and Wang, J. L. (2005b). Functional linear regression analysis for longitudinal data. *Annals of Statistics*, 33(6):2873–2903.

Yu, Y. and Lambert, D. (1999). Fitting Trees to Functional Data, with an Application to Time–of–Day Patterns. *Journal of Computational and Graphical Statistics*, 8(4):749–762.

Yurinskii, V. V. (1976). Exponential Inequalities for Sums of Random Vectors. *Journal of Multivariate Analysis*, 6(4):473–499.

Zhang, J. T. (2011). Statistical inferences for linear models with functional responses. *Statistica Sinica*, 21(3):1431–1451.

Zhang, J. T. and Chen, J. (2007). Statistical inferences for functional data. *Annals of Statistics*, 35(3):1052–1079.

Zhao, X., Marron, J. S., and Wells, M. T. (2004). The functional data analysis view of longitudinal data. *Statistica Sinica*, 14(3):789–808.

Zhou, L., Huang, J. Z., and Carroll, R. J. (2008). Joint modelling of paired sparse functional data using principal components. *Biometrika*, 95(3):601–619.

List of Figures

List of Tables