Functional linear regression models : application to high-throughput plant phenotyping functional data

Tito Manrique Chuquillanqui

▶ **To cite this version:**

Tito Manrique Chuquillanqui. Functional linear regression models : application to high-throughput plant phenotyping functional data. Statistics [math.ST]. Université Montpellier, 2016. English. NNT : 2016MONTT264 . tel-01809004v2

HAL Id: tel-01809004
https://theses.hal.science/tel-01809004v2

Submitted on 6 Jun 2018

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

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

Collège Doctoral Languedoc-Roussillon

THÈSE

Pour obtenir le grade de Docteur

Délivré par l'**Université de Montpellier**

Préparée au sein de l'école doctorale **I2S***
Et des unités de recherche **MISTEA et IMAG**

Spécialité: **Biostatistique**

Présentée par **M. Tito MANRIQUE**

Functional Linear Regression Models. Application to High-throughput Plant Phenotyping Functional Data.

Soutenue le 19/12/2016 devant le jury composé de :

M. Christophe Crambes	MdC	Univ. de Montpellier	Co-directeur de thèse
Mme. Nadine Hilgert	DR	INRA-Montpellier	Directeur de thèse
M. Alois Kneip	PR	University of Bonn	Rapporteur
Mme. Claire Lacour	MdC	Univ. Paris-Sud Orsay	Examinateur
M. André Mas	PR	Univ. de Montpellier	Examinateur
M. Yves Rozenholc	PR	Univ. Paris Descartes	Rapporteur
M. Nicolas Verzelen	CR	INRA-Montpellier	Invité

Jury présidé par M. André MAS

INRA SCIENCE & IMPACT

Numev

* **I2S :** École doctorale Information Structures Systèmes

To my loving parents Tito Gregorio and Ana Gloria.

Abstract

Functional data analysis (FDA) is a statistical branch that is increasingly being used in many applied scientific fields such as biological experimentation, finance, physics, etc. A reason for this is the use of new data collection technologies that increase the number of observations during a time interval.

Functional datasets are realization samples of some random functions which are measurable functions defined on some probability space with values in an infinite dimensional functional space.

There are many questions that FDA studies, among which functional linear regression is one of the most studied, both in applications and in methodological development.

The objective of this thesis is the study of functional linear regression models when both the covariate X and the response Y are random functions and both of them are time-dependent. In particular we want to address the question of how the history of a random function X influences the current value of another random function Y at any given time t.

In order to do this we are mainly interested in three models: the functional concurrent model (FCCM), the functional convolution model (FCVM) and the historical functional linear model. In particular for the FCVM and FCCM we have proposed estimators which are consistent, robust and which are faster to compute compared to others already proposed in the literature.

Our estimation method in the FCCM extends the Ridge Regression method developed in the classical linear case to the functional data framework. We prove the probability convergence of this estimator, obtain a rate of convergence and develop an optimal selection procedure of the regularization parameter.

The FCVM allows to study the influence of the history of X on Y in a simple way through the convolution. In this case we use the continuous Fourier transform operator to define an estimator of the functional coefficient. This operator transforms the convolution model into a FCCM associated in the frequency domain. The consistency and rate of convergence of the estimator are derived from the FCCM.

The FCVM can be generalized to the historical functional linear model, which is itself a particular case of the fully functional linear model. Thanks to this we have used the

Karhunen–Loève estimator of the historical kernel. The related question about the estimation of the covariance operator of the noise in the fully functional linear model is also treated.

Finally we use all the aforementioned models to study the interaction between Vapour Pressure Deficit (VPD) and Leaf Elongation Rate (LER) curves. This kind of data is obtained with high-throughput plant phenotyping platform and is well suited to be studied with FDA methods.

Keywords : *Functional regression models*, *Functional data*, *Convolution Model*, *Concurrent Model*, *Historical Model*.

Résumé

L'Analyse des Données Fonctionnelles (ADF) est une branche de la statistique qui est de plus en plus utilisée dans de nombreux domaines scientifiques appliqués tels que l'expérimentation biologique, la finance, la physique, etc. Une raison à cela est l'utilisation des nouvelles technologies de collecte de données qui augmentent le nombre d'observations dans un intervalle de temps.

Les jeux de données fonctionnelles sont des échantillons de réalisations de fonctions aléatoires qui sont des fonctions mesurables définies sur un espace de probabilité à valeurs dans un espace fonctionnel de dimension infinie.

Parmi les nombreuses questions étudiées par l'ADF, la régression linéaire fonctionnelle est l'une des plus étudiées, aussi bien dans les applications que dans le développement méthodologique.

L'objectif de cette thèse est l'étude de modèles de régression linéaire fonctionnels lorsque la covariable X et la réponse Y sont des fonctions aléatoires et les deux dépendent du temps. En particulier, nous abordons la question de l'influence de l'histoire d'une fonction aléatoire X sur la valeur actuelle d'une autre fonction aléatoire Y à un instant donné t.

Pour ce faire, nous sommes surtout intéressés par trois modèles: le modèle fonctionnel de concurrence (Functional Concurrent Model: FCCM), le modèle fonctionnel de convolution (Functional Convolution Model: FCVM) et le modèle linéaire fonctionnel historique. En particulier pour le FCVM et FCCM nous avons proposé des estimateurs qui sont consistants, robustes et plus rapides à calculer par rapport à d'autres estimateurs déjà proposés dans la littérature.

Notre méthode d'estimation dans le FCCM étend la méthode de régression Ridge développée dans le cas linéaire classique au cadre de données fonctionnelles. Nous avons montré la convergence en probabilité de cet estimateur, obtenu une vitesse de convergence et développé une méthode de choix optimal du paramètre de régularisation.

Le FCVM permet d'étudier l'influence de l'histoire de X sur Y d'une manière simple par la convolution. Dans ce cas, nous utilisons la transformée de Fourier continue pour définir un estimateur du coefficient fonctionnel. Cet opérateur transforme le modèle de convolution en un FCCM associé dans le domaine des fréquences. La consistance et la vitesse de convergence de l'estimateur sont obtenues à partir du FCCM.

Le FCVM peut être généralisé au modèle linéaire fonctionnel historique, qui est lui-même un cas particulier du modèle linéaire entièrement fonctionnel. Grâce à cela, nous avons utilisé

l'estimateur de Karhunen-Loève du noyau historique. La question connexe de l'estimation de l'opérateur de covariance du bruit dans le modèle linéaire entièrement fonctionnel est également traitée.

Finalement nous utilisons tous les modèles mentionnés ci-dessus pour étudier l'interaction entre le déficit de pression de vapeur (Vapour Pressure Deficit: VPD) et vitesse d'élongation foliaire (Leaf Elongation Rate: LER) courbes. Ce type de données est obtenu avec phénotypage végétal haut débit. L'étude est bien adaptée aux méthodes de l'ADF.

Mots-clefs : *Données fonctionnelles*, *Régression linéaire fonctionnelle*, *Modèle de convolution*, *Modèle de concurrence*, *Modèle historique*.

Acknowledgements

First and foremost I wish to thank my advisors Christophe Crambes and Nadine Hilgert for the patient guidance, kindness, motivation and encouragement throughout my PhD studies. They introduced me to the interesting world of functional data analysis. Their knowledge and experience helped me a lot in understanding this branch of statistics.

I would like to thank Professor Alois Kneip and Professor Yves Rozenholc for their agreement to be the referees of this thesis and for their participation as members of the jury.

I would also like to thank all the members of my thesis committee for guiding me through all these years. Thank you André Mas, Claire Lacour and Nicolas Verzelen for all your good and insightful advices.

My sincere appreciation of the hospitality of all the members of the UMR MISTEA, in particular Pascal Neveu and Alain Rappaport. I'm sure that MISTEA is one of the best working places at Montpellier. Thank you for the friendly atmosphere and encouragement during these years: Nicolas Sutton-Charani, Véronique Sals-Vettorel, Maria Trouche, Alexandre Mairin, Nicolas Verzelen, Brigitte Charnomordic, Christophe Abraham, Meïli Baragatti, Patrice Loisel, Martine Marco, Anne Tireau, Gabrielle Weinrott, Paul-Marie Grollemund, Arnaud Charleroy, Hazaël Jones and Céline Casenave.

Besides I want to express my gratitude to Eladio Ocaña and Félix Escalante of the "Instituto de Matemáticas y Ciencias Afines" (IMCA) in Peru. Their endeavor to create one of the best scientific institutions in my country is quite inspiring and praiseworthy.

I gratefully acknowledge the financial support of INRA and the Labex NUMEV (convention ANR-10-LABX-20) for funding my PhD thesis (under project 2013-1-007).

During all these years in Europe I met many fantastic and inspiring people in many countries. I only want to thank all of them for bringing joy, passion, hope and happiness wherever they are. I've certainly learnt from all of you that life is an amazing journey.

Last but certainly not least I'm deeply indebted to my family: my brothers Ivan and Kevin, my sister Liz, my lovely nephews Ikahel and Alejandro, and above all to my mother Ana Gloria for her immense love and patience and to my father Tito Gregorio who will always inspire me to look for Wisdom and to do great things for the benefit of mankind.

Contents

List of Figures

List of Tables

Résumé Etendu

En biologie, les infrastructures expérimentales (ex. plates-formes de phénotypage, bio-procédés) disposent de nouveaux moyens techniques qui génèrent de grandes quantités de données, de qualité hétérogène, acquises à différentes échelles et dans le temps, de nature et de type variés. Valoriser et exploiter ces masses de données est un défi important pour produire de nouvelles connaissances, voir par exemple Ullah and Finch (2013) ou Wang et al. (2016). Il faut donc développer des méthodes et fournir des outils dédiés à un traitement systématique de ces données, avec une prise en compte adaptée de leur dimension temporelle.

En statistique, les analyses multivariées ont montré leur limite, voir Bickel and Levina (2004), Şentürk and Müller (2010), Hsing and Eubank (2015, p. 1), Ramsay and Silverman (2005, Ch 1)... Les raisons sont par exemple : i) le nombre de pas d'observation p est plus grand que le nombre de réalisations n (cf. Hsing and Eubank (2015, p. 2)); ii) les réalisations ne sont pas observées sur les mêmes grilles de temps (cf. Şentürk and Müller (2010)); iii) il y a de fortes corrélations temporelles au sein des réalisations (cf. Ferraty and Vieu (2006, p. 7)) et iv) la régularité et les dérivées des fonctions aléatoires observées jouent un rôle important dans l'étude des données (cf. Mas and Pumo (2009)).

Une façon plus adaptée de traiter ce type de données est de les considérer comme des "réalisations de processus stochastiques à temps continu" (Hsing and Eubank (2015, Ch 1), Bosq (2000, Ch 1)). Cela permet d'introduire les notions de données fonctionnelles et fonctions aléatoires. Les données fonctionnelles sont des échantillons de réalisations de fonctions aléatoires. Une fonction aléatoire représente une évolution, discrète ou à temps continu, d'une variable aléatoire. D'un point de vue mathématique, les fonctions aléatoires sont des fonctions mesurables définies sur un espace de probabilité avec des valeurs dans un espace de dimension infini (Ferraty and Vieu (2006, Ch 1)).

Dans beaucoup d'applications, les fonctions observées sont univariées, dépendant du temps. Elles peuvent dépendre d'un paramètre de type différent, comme par exemple une longueur d'onde dans le cas de jeux de données de spectrométrie. Pour ces fonctions univariées, on parle aussi de "données courbe" (voir Gasser and Kneip (1995)). Les fonctions

peuvent être aussi multivariées, dépendant du temps, d'une longueur d'onde, de l'espace ou autres, voir (Morris (2015), Wang et al. (2016, p. 1)).

Le modèle de régression linéaire fonctionnelle est l'un des sujets les plus traités en analyse de données fonctionnelles, que ce soit dans les applications ou pour les développements méthodologiques (Morris (2015, p. 3)). C'est un bon moyen pour étudier la relation entre fonctions aléatoires dans des domaines variés, voir par exemple Ullah and Finch (2013) et Wang et al. (2016).

L'objectif de la thèse est d'étudier les modèles de régression linéaire pour données fonctionnelles quand le régresseur X (la covariable, l'entrée) et la réponse Y (la sortie) sont tous deux des fonctions aléatoires dépendant du temps (on parle aussi de données "courbes" ou de données "longitudinales"). En particulier nous souhaitons répondre à la question de comment les valeurs passées du régresseur X influencent la valeur courante de la réponse Y à chaque pas de temps t. Dans ce sens, nous proposerons des méthodes d'estimation dans les modèles qui ont de bonnes propriétés (consistance, robustesse) et qui sont rapides à calculer par rapport à d'autres méthodes déjà développées dans la littérature. Dans ce cadre, nous nous intéressons principalement à deux modèles : le modèle linéaire fonctionnel historique et le modèle de convolution fonctionnel (FCVM) que nous introduisons à présent.

Le modèle linéaire fonctionnel historique, introduit par Malfait and Ramsay (2003), est de la forme

$$Y(t) = \int_0^t \mathscr{K}_{hist}(s,t)X(s)\,ds + \varepsilon(t), \tag{1}$$

où $s,t \geq 0$, $\mathscr{K}_{hist}(s,t)$ est la fonction coefficient de régression historique et ε est un bruit aléatoire fonctionnel avec $\mathbb{E}[\varepsilon] = 0$. Ce modèle est un cas particulier du "modèle de régression complètement fonctionnel" (voir par exemple Horváth and Kokoszka (2012, p. 130) et Ramsay and Silverman (2005, Ch 16)), où X et Y sont reliés par un opérateur noyau plus général. Ce dernier s'écrit comme suit :

$$Y(t) = \int_I \mathscr{K}(s,t)X(s)\,ds + \varepsilon(t), \tag{2}$$

où $s,t \in \mathbb{R}$, $I \subseteq \mathbb{R}$ et $\mathscr{K}(\cdot,\cdot)$ est le noyau intégrable. Cependant l'interprétation et l'estimation sont facilitées quand le noyau $\mathscr{K}$ a la forme de $\mathscr{K}_{hist}$ car cette fonction est définie sur le domaine triangulaire plus simple où $s < t$.

Un moyen encore plus simple d'étudier l'influence du passé de X sur la valeur courante de Y se fait au travers du modèle de convolution fonctionnel (FCVM) défini ci-dessous :

$$Y(t) = \int_0^t \theta(s)X(t-s)\,ds + \varepsilon(t), \tag{3}$$

où $t \geq 0$, θ est le coefficient fonctionnel inconnu à estimer. Dans ce modèle, θ est une fonction qui dépend de s uniquement et pas du pas de temps courant t comme le fait $\mathscr{K}_{hist}$. Toutes ces fonctions sont considérées comme nulles pour $t < 0$, ce qui peut s'interpréter comme le fait que 0 est le point de départ des mesures.

Au delà de ces deux modèles, historique et de convolution, nous avons aussi étudié le modèle concurrent fonctionnel (FCCM), défini par Ramsay and Silverman (2005, Ch 14) comme suit :

$$Y(t) = \beta(t)X(t) + \varepsilon(t), \tag{4}$$

où $t \in \mathbb{R}$ et β est le coefficient fonctionnel inconnu à estimer. De même que le modèle de régression complètement fonctionnel déjà mentionné plus haut, c'est un modèle majeur de régression qui traite le cas des variables réponses fonctionnelles, voir Wang et al. (2016, p. 272)). Son importance est soulignée dans Ramsay and Silverman (2005, p. 220).

Notre intérêt pour le FCCM vient du fait que le FCVM et le FCCM sont reliés grâce à la Transformée de Fourier Continue ; chaque fonction θ, associée à un modèle FCVM et définie sur le domaine temporel, a une transformée de Fourier β dans le domaine des fréquences qui est l'élément fonctionnel inconnu d'un modèle FCCM. Cela établit, sous certaines conditions, l'équivalence entre ces deux modèles (voir la section 1.3). Le modèle linéaire fonctionnel historique est aussi relié au modèle FCCM quand par exemple le noyau historique est exprimé comme le produit de deux fonctions univariées de la façon suivante : $\mathscr{K}(s,t) = \beta(t)\gamma(s)$ (Şentürk and Müller (2010), Kim et al. (2011)).

Dans la suite de ce résumé nous introduirons le cadre théorique, les notations principales et les définitions utilisés tout au long de la thèse. Nous donnerons tout d'abord des aspects généraux sur l'analyse des données fonctionnelles. Nous ferons un compte rendu de la littérature sur les modèles de régression pour données fonctionnelles à réponses fonctionnelles. Nous développerons ensuite notre procédure d'estimation de la fonction θ dans le FCVM, ainsi que son implémentation numérique. Nous décrirons brièvement nos résultats, aussi bien théoriques que pratiques (obtenus en simulation et sur données réelles).

Analyse de données fonctionnelles

Quelques exemples de jeux de données

Voici deux exemples pour donner une idée intuitive de ce que l'on appelle des données "courbes" (voir Wang et al. (2016, p. 258)), type de données faciles à visualiser et que nous étudierons tout au long de ce document.

Données de spectrométrie de masse Cet exemple vient de Koomen et al. (2005) et est commenté dans Morris (2015). Ce jeu de données contient les spectres de masse d'une étude sur le cancer du pancréas menée à l'université du Texas. Du sérum a été prélevé du sang de 139 patients atteints de cancer et de 117 personnes en bonne santé. Alors, des spectres protéomiques $X_i(t)$ ont été obtenus avec un instrument de spectrométrie de masse pour chaque individu $i = 1, \cdots, 256$. La grille des observations discrètes (masse moléculaire par unité de charge m/z) pour chaque courbe est de taille $T = 12096$. Un sous ensemble de ces données est montré à la figure 1.

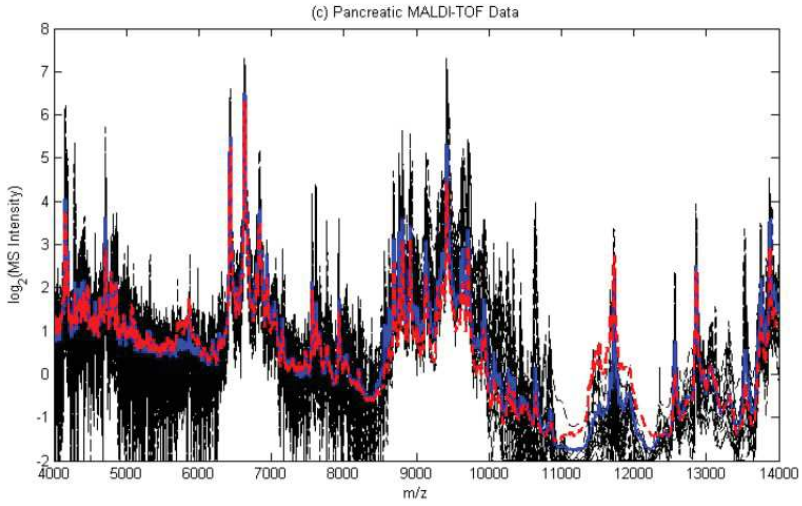

Fig. 1 Log des intensités spectrales issu des données de spectrométrie de masse. Les lignes noires sont les spectres tracés de 20 patients atteint de cancer du pancréas (solide) et de 20 patients contrôle (en pointillés), avec les spectres moyens pour le groupe avec cancer (rouge) et pour le groupe contrôle (bleu).

Données de phénotypage de plantes : Le deuxième exemple est un jeu de données de phénotypage haut-débit de plantes obtenu dans le projet PHENOME. A la figure 2 nous montrons 13 courbes de demande évaporative (Vapor Pressure Deficit - VPD) et de taux d'élongation foliaire (Leaf Elongation Rate - LER). Une courbe contient des données acquises tous les quarts d'heure pendant 24 heures, soit 96 points. VPD et LER sont deux courbes aléatoires qui ont une relation de dépendance de type entrée/sortie. La demande évaporative VPD influence le taux d'élongation foliaire LER.

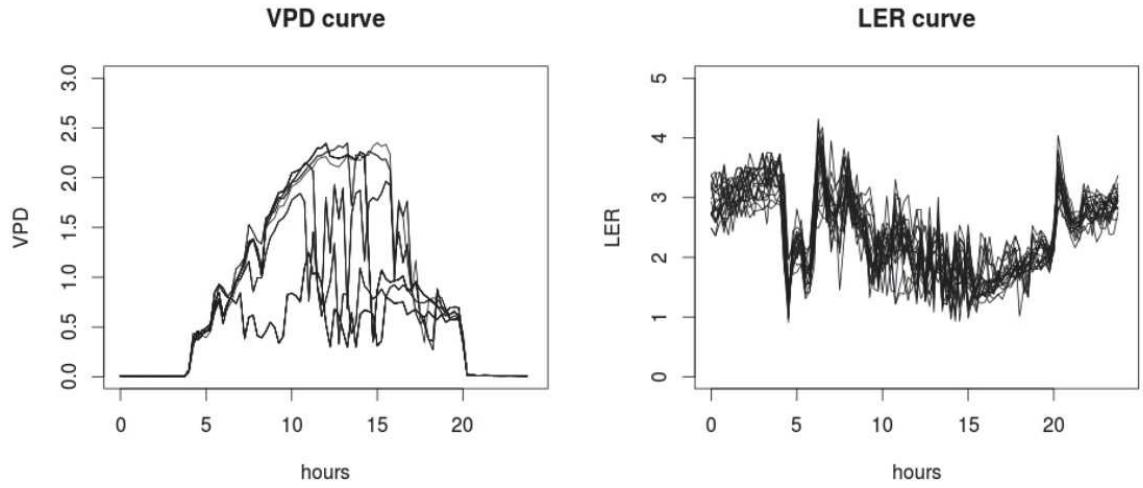

Fig. 2 Exemple de 13 couples de courbes de VPD et LER observées sur 96 pas de temps tout au long d'une journée.

Fonctions aléatoires

Les fonctions aléatoires sont une extension naturelle des variables aléatoires à valeurs réelles et prennent leurs valeurs dans des espaces fonctionnels, plutôt que dans $\mathbb{R}$. Plus généralement Ferraty and Vieu (2006, p. 6) définissent une fonction aléatoire comme une fonction mesurable d'un espace de probabilité $(\Omega, \mathscr{A}, P)$ dans un espace fonctionnel de dimension infini E. Il est courant de considérer E avec sa σ-algèbre de Borel généré par ses ensembles ouverts. Cet espace fonctionnel peut être un espace de Hilbert (voir Horváth and Kokoszka (2012, Ch 2)), un espace de Banach (voir Ledoux and Talagrand (1991, Ch 2)),un espace d'applications (voir Bosq (2000, p. 16)), etc.

Dans cette thèse nous nous intéressons au cas où les fonctions de E sont univariées, par exemple fonction du temps ou d'une fréquence. Comme nous l'avons signalé plus haut, on parle de données courbe ou de données longitudinales pour les données observées à partir de ces fonctions. C'est par exemple le cas quand $E := L^2(I)$, l'ensemble des fonctions Lebesgue de carré intégrable définies sur un intervalle $I \subset \mathbb{R}$. Les méthodes qui traitent ce type de données sont étudiées dans le livre de Ramsay and Silverman (2005) classiquement cité ou dans le livre plus récent de Hsing and Eubank (2015).

Dans ce qui suit nous résumons les principaux éléments concernant les opérateurs espérance et covariance dans les espaces de Hilbert et de Banach (voir Hsing and Eubank (2015) pour plus de détails).

Opérateur Espérance : Soit B un espace de Banach séparable et B^* son espace dual. Une fonction aléatoire à valeurs dans B $X : (\Omega, \mathscr{A}, P) \to B$ est dite **faiblement intégrable** si et seulement si i) la fonction composée $f^*(X)$ est intégrable pour tout $f^* \in B^*$ et ii) il existe un élément de B, noté $\mathbb{E}[X]$, tel que pour tout $f^* \in B^*$

$$\mathbb{E}[f^*(X)] = f^*(\mathbb{E}[X]).$$

L'élément $\mathbb{E}[X]$ est désigné sous le nom de **intégrale faible** de X. Une fonction aléatoire X à valeurs dans B sera définie comme **fortement intégrable** (ou intégrable) si $\mathbb{E}[\|X\|_B] < \infty$, où $\|\cdot\|_B$ est la norme de B. Chaque fois que X est intégrable (fortement), il est courant de noter $\mathbb{E}[X]$ ou $\int X dP$ de manière équivalente.

Considérons la relation suivante d'équivalence entre fonctions aléatoires X et Y à valeurs dans B : $X \sim Y$ si et seulement si $X = Y$ presque sûrement (**a.s.**). En utilisant les classes d'équivalence correspondantes, nous définissons l'espace $L_B^1(P)$ des classes d'équivalence des fonctions intégrables aléatoires à valeurs dans B. Si l'on définit la norme $\|X\|_{L_B^1} := \mathbb{E}\|X\|_B$, alors $L_B^1(P)$ est un espace de Banach. De manière analogue pour $p \in]1, \infty[$, nous définissons les espaces $L_B^p(P)$ des classes de fonctions aléatoires à valeurs dans B telles que $\mathbb{E}[\|X\|_B^p] < \infty$. Similairement à $L_B^1(P)$, cet espace devient un Banach si on définit la norme $\|X\|_{L_B^p} := [\mathbb{E}\|X\|_B^p]^{1/p}$.

Opérateur Covariance : Pour une fonction aléatoire $X \in L_B^2(P)$, avec $\mathbb{E}[X] = 0$, nous définissons son opérateur de covariance comme l'opérateur linéaire borné suivant :

$$\begin{array}{rccl} C_X : & B^* & \to & B \\ & f^* & \mapsto & \mathbb{E}[f^*(X)X]. \end{array}$$

Dans le cas où $\mathbb{E}[X] \neq 0$, nous définissons $C_X := C_{X-\mathbb{E}[X]}$. Il est possible de généraliser cette définition pour définir l'**opérateur de covariance croisée** entre deux fonctions aléatoires. Pour ce faire, considérons $X \in L_{B_1}^2(P)$ et $Y \in L_{B_2}^2(P)$, où B_1 et B_2 sont des espaces de Banach séparables tels que $\mathbb{E}[X] = 0$ et $\mathbb{E}[Y] = 0$. Les opérateurs de covariance croisée de X et Y sont les opérateurs linéaires bornés suivants:

$$\begin{array}{rccl} C_{X,Y} : & B_1^* & \to & B_2 \\ & f^* & \mapsto & \mathbb{E}[f^*(X)Y] \end{array} \quad \text{et} \quad \begin{array}{rccl} C_{Y,X} : & B_2^* & \to & B_1 \\ & g^* & \mapsto & \mathbb{E}[g^*(X)Y]. \end{array}$$

Dans le cas particulier d'un espace de Hilbert séparable H, doté du produit scalaire $\langle \cdot, \cdot \rangle$, la définition de l'espérance est similaire à celle définie pour les espaces de Banach. En outre, la définition de l'opérateur de covariance sera plus simple, grâce au théorème de

représentation de Riez pour l'espace dual H^*. De cette façon, soit X une fonction aléatoire à valeurs dans H telle que $\mathbb{E}[\|X\|_H^2] < \infty$ et $\mathbb{E}[X] = 0$. Alors, l'opérateur de covariance de X est le suivant

$$\begin{array}{rccl} C_X: & H & \to & H \\ & x & \mapsto & \mathbb{E}[\langle x, X\rangle X]. \end{array}$$

Il est connu que cet opérateur est symétrique, positif, nucléaire (voir Bosq (2000, p. 34) pour plus de détails). De même, l'opérateur de covariance croisée est défini comme suit :

$$C_{X,Y}(x) = \mathbb{E}[\langle X, x\rangle Y],$$

pour tout $x \in H$. Cet opérateur est aussi nucléaire (Bosq (2000, p. 34)).

Décomposition spectrale des opérateurs : Soit X une fonction aléatoire à valeurs dans H telle que $\mathbb{E}[\|X\|_H^2] < \infty$ et $\mathbb{E}[X] = 0$. La décomposition spectrale de X existe et est la suivante :

$$C_X(x) = \sum_{j=1}^{\infty} \lambda_j \langle x, v_j\rangle v_j$$

où les $(v_j)_{j\geq 1}$ sont les fonctions propres de C_X et forment une base orthonormale de H, les $(\lambda_j)_{j\geq 1}$ sont les valeurs propres et satisfont :

$$\sum_{j=1}^{\infty} |\lambda_j| = \mathbb{E}[\|X\|_H^2] < \infty,$$

puisque $\lambda_j = \langle C_X(v_j), v_j\rangle = \mathbb{E}(\langle X, v_j\rangle^2)$.

Cette décomposition spectrale sert à projeter les fonctions aléatoires dans un sous-espace de dimension finie engendré par les K premières fonctions propres. Elle intervient dans de nombreuses méthodes bien connues d'analyses pour données fonctionnelles, comme l'analyse en composantes principales fonctionnelle (voir, (Ramsay and Silverman, 2005, Ch 8)) ou les méthodes d'estimation pour le modèle de régression linéaire fonctionnel de type Karhunen-Loeve (voir Crambes and Mas (2013)).

Analyses de données fonctionnelles (ADF) et analyses statistiques multivariées

Dans ce paragraphe, nous abordons la question de l'inadéquation de certaines méthodes multivariées pour traiter des données fonctionnelles. Comme nous l'avons mentionné plus tôt, Wang et al. (2016, p. 1) définit l'ADF comme " l'analyse et la théorie des données

qui sont sous la forme de fonctions, d'images et de formes, ou d'objets plus généraux ". Quant à elle, l'analyse multivariée porte sur la compréhension et l'analyse de l'interaction de plusieurs variables statistiques qui sont en général mesurées simultanément (Johnson and Wichern, 2007, p. 1).

Hsing and Eubank (2015, p. 1) et Ramsay and Silverman (2005, Ch 1), parmi d'autres références, ont montré que certaines méthodes multivariées classiques sont inadaptées pour le traitement des données fonctionnelles. Il y a quatre raisons principales à cela :

1. Une exigence pour appliquer des méthodes multivariées aux données fonctionnelles est que la grille d'observations doit être fixe et la même pour toutes les réalisations. Ce n'est pas requis par l'ADF qui peut être appliquée à d'autres cas, par exemple lorsque les temps d'observation sont aléatoires et indépendants parmi les réalisations (voir par exemple Şentürk and Müller (2010, p. 1257), Yao et al. (2005b, p. 578)).

2. Le nombre de pas de temps d'observation p est plus grand que le nombre de réalisations n (Hsing and Eubank (2015, p. 2)). L'inférence statistique sous cette condition est impossible avec les méthodes classiques de l'analyse multivariée. Les méthodes statistiques pour données à grande dimension (Bühlmann and van de Geer (2011, Ch 1)) et l'ADF sont adaptées à ce type de données. L'une des raisons pour lesquelles le cas $(p >> n)$ est problématique pour l'analyse multivariée est que les opérateurs de covariance sont non-inversibles (mal-conditionnés). Cela rend difficile la résolution des systèmes linéaires dans les modèles de régression, qui sont largement utilisés en analyse multivariée.

3. Les corrélations élevées des mesures successives des variables, lorsqu'elles sont observées sur des pas de temps proches, impliquent un problème mal conditionné qui n'est pas adapté aux méthodes multivariées, car il rend difficile la résolution des systèmes linéaires (voir Yao et al. (2005b), Ferraty and Vieu (2006, p. 7)).

4. Enfin, dans le cas où la régularité et les dérivés des fonctions aléatoires jouent un rôle majeur pour étudier les données (voir par exemple Mas and Pumo (2009), Ramsay and Silverman (2005, Ch 17)), il est nécessaire de considérer la nature fonctionnelle des données, et cela ne peut pas être accompli avec l'approche multivariée.

Modèles de régression linéaire fonctionnelle avec réponse fonctionnelle

Nous donnons dans cette section une courte bibliographie des modèles utilisés pour étudier la dépendance entre deux fonctions aléatoires. Nous considérons ici des fonctions aléatoires définies dans l'espace de Hilbert $L^2(I)$, c'est-à-dire l'espace des fonctions de carré intégrable au sens de Lebesgue sur l'intervalle $I \subseteq \mathbb{R}$. Soient X et Y deux fonctions aléatoires, X la variable explicative, Y la variable réponse. Un modèle naturel qui lie X et Y est le suivant

$$Y = \Psi(X) + \varepsilon,$$

où Ψ est un opérateur fonctionnel et ε est un bruit. Dans ce modèle Ψ résume la façon dont X agit sur Y. Donc estimer Ψ est essentiel pour comprendre cette relation. Se posent alors les questions suivantes : comment définir un estimateur de Ψ à partir d'un échantillon $(X_i, Y_i)_{i=1,\cdots,n}$ de n réalisations i.i.d. de X et Y, et quelles sont les propriétés de cet estimateur (consistance et vitesse de convergence).

Il existe deux approches principales pour répondre à ces questions. D'abord l'approche paramétrique fonctionnelle nécessite que Ψ appartienne à un sous-ensemble particulier $\mathscr{S}$ de l'ensemble des opérateurs linéaires continus $\mathscr{C}$ sur $L^2(I)$, qui peut être indexé par un nombre fini de certains opérateurs linéaires continus fixés sur $L^2(I)$ (Ferraty and Vieu (2006, p. 8)). Un exemple est celui de la régression linéaire fonctionnelle avec sortie fonctionnelle définie dans (2), où l'opérateur linéaire continu est uniquement paramétré avec un seul opérateur noyau intégral.

La seconde approche est l'approche non-paramétrique fonctionnelle. L'ensemble $\mathscr{S}$ n'a pas à être indexé par un ensemble fini d'opérateurs linéaires continus Ferraty and Vieu (2006, p. 8). Les méthodes non-paramétriques fonctionnelles tiennent compte de la régularité des éléments de $\mathscr{S}$, par exemple les modèles additifs fonctionnels étudiés par Müller and Yao (2012), les espaces de Hilbert à noyau reproduisant (voir Lian (2007), Kadri et al. (2010)) ou les méthodes à noyau Ferraty and Vieu (2006).

Tout au long de cette thèse, nous nous intéressons à l'approche paramétrique fonctionnelle, plus précisément au modèle de régression linéaire fonctionnelle avec réponse fonctionnelle. Dans ce qui suit, nous donnons une étude plus détaillée de ce modèle.

Deux modèles majeurs

Dans un article récent, Wang et al. (2016) propose de diviser la classe des modèles de régression linéaire fonctionelle à sortie fonctionnelle en deux grandes catégories : le modèle

concurrent fonctionnel (FCCM) et le modèle de régression entièrement fonctionnel (voir Horváth and Kokoszka (2012, p. 130)). Nous discutons brièvement des deux.

Modèle concurrent fonctionnel : Son équation est la suivante :

$$Y(t) = \beta_0(t) + \beta_1(t)X(t) + \varepsilon(t), \tag{5}$$

où $t \in \mathbb{R}$, $\beta_0(t)$ et $\beta_1(t)$ sont les coefficients fonctionnels du modèle à estimer.

Des modèles étroitement liés ont déjà été discutés par plusieurs auteurs. Par exemple dans West et al. (1985), les auteurs définissent un modèle similaire appelé « modèle linéaire dynamique généralisé » et ils étudient le modèle d'un point de vue bayésien. Hastie and Tibshirani (1993) proposent le 'Varying Coefficient Model'. Ce modèle a la forme

$$\eta = \beta_0(R_0) + X_1\beta_1(R_1) + \cdots X_p\beta_p(R_p), \tag{6}$$

où η est un paramètre qui détermine la distribution de la variable aléatoire Y, $X_1, \cdots, X_p$ et $R_1, \cdots, R_p$ sont des prédicteurs, $\beta_1, \cdots, \beta_p$ sont des fonctions à estimer. Dans le cas le plus simple d'un modèle Gaussien, l'équation (6) prend la forme

$$Y = \beta_0(R_0) + X_1\beta_1(R_1) + \cdots X_p\beta_p(R_p) + \varepsilon,$$

où $\mathbb{E}[\varepsilon] = 0$ et $var[\varepsilon] = \sigma^2$. Dans le cas où toutes les covariables $R_1, \cdots, R_p$ sont la variable de temps t et que les covariables $X_1, \cdots, X_p$ dépendent du temps, le modèle prend alors la forme du modèle concurrent fonctionnel, c'est à dire :

$$Y(t) = \beta_0(t) + X_1(t)\beta_1(t) + \cdots X_p(t)\beta_p(t) + \varepsilon.$$

Beaucoup d'auteurs ont étudié l'estimation des fonctions inconnues β_i. Par exemple Wu et al. (1998) proposent une approche basée sur les noyaux quand il y a k covariables fonctionnelles. Dreesman and Tutz (2001) et Cai et al. (2000) proposent une estimation de type maximum de vraisemblance locale. Zhang and Lee (2000), Fan et al. (2003) et Zhang et al. (2002) proposent des approches par lissage polynomial local. Fan and Zhang (2000) proposent une approche en deux étapes : réaliser tout d'abord des régressions de type moindres carrés ordinaires de manière ponctuelle puis faire un lissage de ces estimateurs grossiers. Huang et al. (2004) proposent d'estimer les fonctions β_i par des B-splines et des méthodes de moindres carrés. Une étude plus poussée du "Varying Coefficient Model" multivarié est présentée dans Zhu et al. (2014).

Nous voulons souligner que, bien que le "varying coefficient model" soit lié au FCCM, il a été conçu à l'origine pour le cas où t est une variable aléatoire et a été traité avec des méthodes d'analyse multivariée. Ainsi, les premiers travaux sur l'estimation ne tiennent pas compte de la nature fonctionnelle des données, comme l'a remarqué Ramsay and Silverman (2005, p. 259). En revanche, Şentürk and Müller (2010) proposent une méthode d'estimation qui est plus proche de l'approche pour données fonctionnelles.

Modèle de régression entièrement fonctionnel Nous rappelons la forme de ce modèle :

$$Y(t) = \int_I \mathscr{K}(s,t)X(s)\,ds + \varepsilon(t).$$

Ce modèle a été popularisé par le travail fondateur de Ramsay and Dalzell (1991). L'estimation du noyau $\mathscr{K}$ est un problème inverse et nécessite une certaine régularisation pour que l'inversion soit possible. Ramsay and Dalzell (1991, p. 552) ont proposé une méthode des moindres carrés pénalisés pour le faire. James (2002) a proposé une méthode pénalisée avec des splines. Cuevas et al. (2002) étudient cette régularisation pour un design fixe. Dans He et al. (2000), les auteurs définissent l' «équation normale fonctionnelle» qui généralise les équations normales multivariées et ils prouvent qu'il existe une solution unique sous certaines conditions. Ils utilisent la décomposition de Karhunen-Loève pour estimer $\mathscr{K}$. Chiou et al. (2004) donnent un bon résumé de cette approche. Cette idée a été revisitée dans Yao et al. (2005a) où les auteurs ont traité des données longitudinales parcimonieuses avec des temps d'observation irréguliers et aléatoires. De plus, un estimateur fondé sur des ondelettes est proposé dans Aguilera et al. (2008), un estimateur basé sur des splines est proposé dans Antoch et al. (2010), et un estimateur de type Karhunen-Loève dans Crambes and Mas (2013).

Dès que X et Y sont dépendants du temps, dans l'équation (2), on voit que des valeurs futures de X sont utilisées pour expliquer des valeurs passées de Y. Ce constat est contre-intuitif. Pour cette raison, Malfait and Ramsay (2003) s'intéressent au cas particulier où $Y(t)$, la valeur de Y au temps t, dépend du passé $\{X(s) : 0 \leq s \leq t\}$. Cela conduit au modèle linéaire fonctionnel historique. Ce modèle particulier est discuté dans la sous-section qui suit.

Modèle de régression linéaire fonctionnelle historique HFLM

Ce modèle, proposé par Malfait and Ramsay (2003), a été défini à l'équation (1) que nous rappelons :

$$Y(t) = \int_0^t \mathscr{K}_{hist}(s,t)X(s)\,ds + \varepsilon(t),$$

pour tout $t > 0$. Nous allons maintenant résumer l'approche de Malfait and Ramsay (2003) pour estimer $\mathscr{K}_{hist}$. Les auteurs proposent d'utiliser une base d'éléments finis $\phi_k(s,t)$. Cette base est faite de fonctions linéaires par morceaux définies sur une grille fixe et adaptée de points. Les auteurs utilisent une approximation $\mathscr{K}_{hist}(s,t) \approx \sum_{k=1}^{K} b_k\,\phi_k(s,t)$ pour transformer le problème comme suit :

$$Y_i(t) = \sum_{k=1}^{K} b_k\,\psi_{ik}(s,t) + \varepsilon_i(t),$$

où $\psi_{ik}(s,t) := \int_0^t X_i(s)\,\phi_k(s,t)\,ds$.

Soient $\mathbf{y}(t)$ et $\mathbf{e}(t)$ les vecteurs de longueur N contenant les valeurs $Y_i(t)$ et $\varepsilon_i(t)$, respectivement. Soit aussi $\Psi(t)$ la matrice $N \times K$ contenant les valeurs $\psi_{ik}(t)$. On note $\mathbf{b}$ le vecteur des coefficients $(b_1,\cdots,b_K)'$. Alors la forme matricielle de l'équation (1) est pour tout $t \in [0,\infty[$:

$$\mathbf{y}(t) = \Psi(t)\,\mathbf{b} + \mathbf{e}(t).$$

Cela mène aux équations normales $\left(\int_0^T \Psi(t)'\Psi(t)dt\right)\mathbf{b} = \int_0^T \Phi(t)'\mathbf{y}(t)dt$. Finalement les auteurs approchent et résolvent cette équation avec un modèle linéaire multivarié.

Avec une approche similaire, Harezlak et al. (2007) ont considéré la même représentation de la fonction $\mathscr{K}_{hist}(s,t)$ à travers la base de fonctions $\phi_k(s,t)$. Dans ce cas, les auteurs ont exploré deux techniques de régularisation différentes qui utilisent une troncature de la base, des pénalités de rugosité, et des pénalités de parcimonie. La première pénalise les valeurs absolues des différences de fonction de base de coefficients (approche LASSO) et la seconde pénalise les carrés de ces différences (méthodologie des splines pénalisées). Enfin, les auteurs ont évalué la qualité de l'estimation avec une extension du critère d'information d'Akaike.

Dans Kim et al. (2011), les auteurs s'intéressent à un type particulier de modèle historique. Ils considèrent les courbes X_i et Y_i comme des données longitudinales parcimonieuses, et la fonction $\mathscr{K}_{hist}(s,t)$ décomposable comme suit : $\mathscr{K}_{hist}(s,t) = \sum_{k=1}^{K} b_k(t)\phi_k(s)$, où les $b_k(t)$ sont les coefficients fonctionnels à estimer et les $\phi_k(s)$ sont des fonctions de base prédéterminées, par exemple des B-splines.

Dans ce cas, le modèle (1) devient un " varying coefficient model" ou un un modèle fonctionnel concurrent (FCCM) après intégration. La procédure d'estimation utilise la décomposition en fonctions de base de Y et X. Ces décompositions sont nécessaires pour

résoudre une équation normale particulière qui utilise les auto-covariances de $X(s)$ et $Y(t)$ et la covariance croisée de $(X(s), Y(t))$.

Un modèle similaire est étudié dans Şentürk and Müller (2010). Les auteurs considèrent la modification suivante de (1)

$$Y(t) = \beta_0(t) + \int_0^{\Delta} \beta_1(t)\gamma(s)X(s)\,ds + \varepsilon(t),$$

pour $t \in [\Delta, T]$ avec $\Delta > 0$ et un choix adapté de $T > 0$. De même que pour Kim et al. (2011), ce cas particulier est fortement lié au FCCM. Il est clair que, après avoir estimé la fonction γ, le problème devient celui de l'estimation dans un FCCM. Dans Şentürk and Müller (2010), une nouvelle méthode pour estimer les fonctions β_0 et β_1 est développée, qui tient compte de la nature fonctionnelle des données. Toutefois, l'idée essentielle pour estimer les fonctions inconnues (γ, β_0 et β_1) est d'utiliser une idée équivalente à l' «équation normale fonctionnelle» : relier l'auto-covariance de X avec la covariance croisée de X et Y, puis d'inverser l'auto-covariance de X pour obtenir la fonction inconnue.

Finalement il est possible de considérer une structure encore plus spécifique de la fonction $\mathscr{K}_{hist}(s,t)$ qui ne dépend pas de la valeur courante t, mais uniquement de s. L'intégrale prend la forme d'une convolution. Pour ce faire, nous supposons qu'il existe une fonction θ telle que $\mathscr{K}_{hist}(s,t) = \theta(t-s)$ et nous appliquons un changement de variables pour obtenir le modèle de convolution fonctionnel ((3), objet du paragraphe suivant.

Modèle de convolution fonctionnel (FCVM)

L'un des principaux objectifs de cette thèse est d'étudier l'influence du passé de la covariable fonctionnelle X sur la valeur actuelle de la réponse $Y(t)$. Une façon de modéliser cette relation de dépendance est à travers le FCVM. Dans ce paragraphe, nous présentons l'origine de ce modèle et dans la section suivante, nous discutons des méthodes d'estimation de la fonction coefficient θ et la place du FCVM entre autres modèles dans la littérature.

Rappelons la forme du modèle FCVM (3), dérivé de Malfait and Ramsay (2003),

$$Y(t) = \int_0^t \theta(s)X(t-s)ds + \varepsilon(t),$$

pour tout $t \geq 0$. Nous nous intéressons à l'estimation de la fonction θ à partir d'un échantillon i.i.d. $(X_i, Y_i)_{i\in\{1,\cdots,n\}}$ des fonctions aléatoires X et Y.

Le FCVM est une extension fonctionnelle des modèles à retards échelonnés en séries temporelles (Greene (2003, Ch 19)). Le modèle de régression dynamique de forme générale

$$y_t = \alpha + \sum_{i=0}^{\infty} \beta_i x_{t-i} + \varepsilon_t.$$

pour tout $t \geq 0$, en est un exemple. Si nous supposons que $x_i = 0$ pour tout $i < 0$ (c'est à dire qu'il y a un point de départ), alors la somme est finie. C'est une discrétisation du FCVM. Les applications de ce modèle sont commentées dans Greene (2003, Ch 19).

La question de l'estimation de θ est centrale dans l'étude du FCVM. C'est le sujet de la section suivante. Nous détaillerons plus précisément la connexion entre le FCVM et le FCCM obtenue avec l'utilisation de la transformée de Fourier continue.

Estimation de θ dans le FCVM

Le FCVM a quatre caractéristiques principales : i) la covariable (entrée) et la réponse (sortie) sont des fonctions aléatoires, ii) la convolution est non périodique (i.e. nous ne considérons pas les fonctions périodiques), iii) la taille de l'échantillon est $n > 1$, en outre, nous sommes intéressés par le comportement asymptotique (n $\to \infty$) et enfin iv) le bruit est fonctionnel. A notre connaissance, il y a peu de documents qui étudient un modèle avec de telles caractéristiques. Cependant, il existe de nombreux modèles qui sont proches de FCVM. Dans ce qui suit, nous explorons certains d'entre eux.

Asencio et al. (2014) étudient un problème lié, dans lequel ils considèrent plus de fonctions covariables (prédicteurs). L'estimation de θ est faite en projetant les fonctions dans une base spline de dimension finie et en utilisant une approche de moindres carrés ordinaires pénalisés pour estimer les coefficients dans cette base. Une autre approche consiste à utiliser le modèle linéaire fonctionnel historique (Malfait and Ramsay (2003)) pour estimer θ, en tenant compte de la forme particulière que la fonction noyau $\mathscr{K}_{hist}$ doit avoir dans ce cas particulier. De la même manière, nous pouvons utiliser l'approche de Harezlak et al. (2007), ou même dans un cas plus restreint les approches de Kim et al. (2011) et Şentürk and Müller (2010). Le FCVM peut aussi être considéré comme un cas particulier du modèle proposé par Kim et al. (2011), quand $K = 1$, et $\phi_K = \theta$ dans $\mathscr{K}_{hist}(s,t) = \sum_{k=1}^{K} b_k(t)\phi_k(s)$.

Şentürk and Müller (2010, p. 1259) proposent une méthode pour estimer θ quand le FCVM a la forme restreinte suivante :

$$Y(t) = \int_0^{\Delta} \theta(s) X(t-s) ds + \varepsilon(t),$$

où $\Delta > 0$ est une valeur fixe. L'estimation de θ dans ce cas est faite en utilisant la décomposition de Karhunen-Loève de l'opérateur de covariance de la fonction aléatoire $Z_t(s) := X(t-s)$, où t est fixé et $s \in [0,\Delta]$. Les auteurs expriment θ dans la base de fonctions propres de cet opérateur, puis estiment les coefficients avec une procédure de moindres carrés ordinaires. Cela produit un estimateur pour chaque pas de temps t. Ils considèrent une grille de temps d'observation, puis ils prennent la moyenne de tous ces estimateurs. Cette approche est similaire à celle de Kim et al. (2011).

A notre connaissance, seuls les articles mentionnées ci-dessus ont abordé l'étude de l'estimation de θ en tenant compte des quatre caractéristiques du FCVM mentionnées plus tôt. L'approche que nous développons dans le chapitre 3 est une nouvelle façon de répondre à cette question. Nous n'utilisons pas de projection dans une base de fonctions de dimension finie. De plus, nous étudions les propriétés asymptotiques de l'estimateur, ce qui n'est fait dans aucune des approches précédentes.

Estimateur par déconvolution de Fourier fonctionnelle (FFDE)

Nous définissons l'estimateur par déconvolution de Fourier fonctionnelle en trois étapes. i) D'abord, nous utilisons la transformée de Fourier continue ($\mathscr{F}$) pour transformer la convolution dans le domaine temporel en une multiplication dans le domaine des fréquences, voir (4). ii) Une fois dans le domaine des fréquences, nous estimons β avec **l'estimateur de régression Ridge fonctionnelle (FRRE)** défini dans Manrique et al. (2016) (voir Chapitre 2), qui est une extension de la méthode de régularisation Ridge (Hoerl (1962)) introduite pour traiter des problèmes mal posés dans la régression linéaire classique. iii) La dernière étape consiste à utiliser la transformée de Fourier continue inverse pour estimer θ. Cette définition est formalisée mathématiquement comme suit.

Soit $(X_i, Y_i)_{i=1,\cdots,n}$ un échantillon i.i.d issu du FCVM (3).

Etape i) Nous utilisons la transformée de Fourier continue ($\mathscr{F}$) définie par

$$\mathscr{F}(f)(\xi) = \int_{t=-\infty}^{+\infty} f(t) e^{-2\pi i t \xi} \, dt,$$

où $\xi \in \mathbb{R}$ et $f \in L^2$. Cet opérateur est utilisé pour transformer le FCVM (3) défini dans le domaine temporel en un modèle équivalent dans le domaine des fréquences :

$$\mathscr{Y}(\xi) = \beta(\xi)\, \mathscr{X}(\xi) + \varepsilon(\xi), \tag{7}$$

où $\xi \in \mathbb{R}$, $\beta := \mathscr{F}(\theta)$ est le coefficient fonctionnel à estimer. $\mathscr{X} := \mathscr{F}(X)$ et $\mathscr{Y} := \mathscr{F}(Y)$ sont les transformées de Fourier de X et Y. Enfin, $\varepsilon := \mathscr{F}(\varepsilon)$ est un bruit additif fonctionnel.

Le modèle (7) dans le domaine des fréquences est un modèle de convolution fonctionnel FCCM de type (4). Clairement l'estimation de β impliquera l'estimation de θ grâce à la transformée de Fourier inverse $\mathscr{F}^{-1}$.

Etape ii) L'estimateur de régression Ridge fonctionnelle (FRRE) de β dans le FCCM (4) ou (7) est défini comme suit :

$$\hat{\beta}_n := \frac{\frac{1}{n}\sum_{i=1}^{n} \mathscr{Y}_i \mathscr{X}_i^*}{\frac{1}{n}\sum_{i=1}^{n} |\mathscr{X}_i|^2 + \frac{\lambda_n}{n}}, \tag{8}$$

où l'exposant $*$ indique le conjugué complexe et λ_n est un paramètre de régularisation positif.

Nous avons choisi un estimateur Ridge de β dans (8), car ainsi il est naturel d'utiliser la transformée de Fourier inverse ($\mathscr{F}^{-1}$) pour estimer θ. Les propriétés de consistance en norme L^2 de l'estimateur de β sont aussi conservées pour l'estimateur de θ. Nous bénéficions de plus de l'efficacité de calcul de l'algorithme de la transformée de Fourier rapide.

Comme nous l'avons vu auparavant, l'idée de transformer le modèle linéaire fonctionnel historique en un FCCM a déjà été proposée par Kim et al. (2011) et d'une manière différente par Şentürk and Müller (2010). Dans ces deux articles, les auteurs ont utilisé des structures spéciales pour la fonction noyau $\mathscr{K}_{hist}$. Ces structures leur permettent de transformer le modèle historique en FCCM. Dans notre cas, nous utilisons une approche différente. Nous n'imposons pas une structure particulière à la fonction du noyau. Nous transformons le modèle FCVM dans le domaine temporel en son équivalent dans le domaine fréquentiel. En conséquence, cela ouvre la possibilité d'utiliser également d'autres méthodes d'estimation de β dans le FCCM afin d'estimer θ dans le FCVM.

Etape iii) L'estimateur par déconvolution de Fourier fonctionnelle (FFDE) de θ dans (3) est défini par

$$\hat{\theta}_n := \mathscr{F}^{-1}(\hat{\beta}_n). \tag{9}$$

Notons que l'estimateur $\hat{\theta}_n$ (FFDE) est à valeurs dans $\mathbb{R}$ et appartient à $L^2(\mathbb{R},\mathbb{R})$ (voir Chapitre 3. Une autre hypothèse importante est que le FFDE se décompose comme suit :

$$\hat{\theta}_n = \theta - \frac{\lambda_n}{n}\mathscr{F}^{-1}\left(\frac{\mathscr{F}(\theta)}{\frac{1}{n}\sum_{i=1}^{n} |\mathscr{F}(X_i)|^2 + \frac{\lambda_n}{n}}\right) + \mathscr{F}^{-1}\left(\frac{\frac{1}{n}\sum_{j=1}^{n} \mathscr{F}(\varepsilon_j)\overline{\mathscr{F}(X_j)}}{\frac{1}{n}\sum_{i=1}^{n} |\mathscr{F}(X_i)|^2 + \frac{\lambda_n}{n}}\right). \tag{10}$$

L'étude de cette décomposition nous permettra de démontrer la consistance de cet estimateur. Notez l'importance de l'équivalence entre le FCVM et le FCCM, en raison de l'utilisation de deux représentations équivalentes de la même information (dans le domaine temporel et le domaine fréquentiel) obtenue grâce à la transformée de Fourier continue.

L'estimateur par déconvolution de Fourier fonctionnelle (FFDE) de θ dans le FCVM est étudié dans le chapitre 3. Nous avons choisi de proposer un tel estimateur pour tirer parti de l'équivalence des modèles en temps et en fréquence, et des propriétés mathématiques de la transformée de Fourier continue. Les avantages de cet estimateur sont à la fois théoriques et pratiques : théoriques, car nous développons une approche construite avec des fonctions aléatoires et des espaces fonctionnels, et pratiques parce que pour implémenter cette méthode, nous utilisons la transformée de Fourier discrète et l'algorithme de transformée de Fourier rapide (FFT) qui améliore la vitesse de calcul des estimateurs de façon significative par rapport à d'autres estimateurs possibles. Nous décrivons dans la suite d'autres estimateurs possibles adaptés de la littérature.

Les méthodes de déconvolution dans la littérature

Considérons à présent d'autres modèles indirectement liés au FCVM. De cela, nous serons en mesure d'adapter certaines techniques pour estimer la fonction θ.

Nous commençons avec le modèle de déconvolution multicanal (voir par exemple De Canditiis and Pensky (2006), Pensky et al. (2010) et Kulik et al. (2015)). Ce modèle est considéré par les méthodes de traitement du signal. De même que pour le FCVM, l'entrée et la sortie sont fonctionnelles (signaux, données courbes), il y a beaucoup de réalisations ($n > 1$, multicanaux) et le bruit est fonctionnel. Mais la différence avec le FCVM est que les auteurs étudient le cas périodique (les signaux sont périodiques, ainsi que que la convolution). En outre, les auteurs ne traitent pas du comportement asymptotique des estimateurs.

Le problème de déconvolution multicanal est une façon de généraliser le problème de la déconvolution en traitement du signal (voir, par exemple Johnstone et al. (2004), Brown and Hwang (2012), Gonzalez and Eddins (2009)). Les auteurs utilisent la convolution (périodique ou non) pour modéliser comment une fonction réponse h transforme un signal g (inconnu) à travers l'équation suivante

$$f(t) = \int_D h(s)g(t-s)ds + \varepsilon(t),$$

où D est le domaine d'intégration ($[0,T]$ dans la cas périodique pour un T fixé, et $[0,t]$ ou $\mathbb{R}$ dans le cas non périodique), f est le signal observé et ε le bruit. Il y a plusieurs méthodes

pour estimer g étant données les fonctions h et f, par exemple la méthode de déconvolution paramétrique de Wiener (Gonzalez and Eddins (2009, Ch 5)).

Si on interprète f comme Y, h comme X et g comme θ, alors on peut appliquer ces méthodes pour estimer θ. Nous remarquons que, bien que ce problème d'estimation est relié au FCVM, il ne traite que du cas $n = 1$. Il n'y a pas d'étude du comportement asymptotique des estimateurs proposés.

De façon similaire, les méthodes de déconvolution en statistique non paramétrique (voir Meister (2009),Johannes et al. (2009)) traitent du cas $n = 1$ et ne considèrent pas les bruits fonctionnels. Le but ici est d'estimer la densité de probabilité d'une variable aléatoire réelle X à partir de l'observation d'une autre variable aléatoire réelle Y telle que $Y = X + Z$, la densité de probabilité de Z étant connue. Pour résoudre ce problème, les auteurs utilisent le fait que la densité de probabilité de la somme de deux variables aléatoires est la convolution de leurs densités respectives. Il pourrait être possible d'adapter ces techniques pour estimer θ dans le FCVM, mais nous pensons que l'estimation serait pire que celle des méthodes de traitement du signal, parce que dans le premier cas le bruit fonctionnel est pas considéré.

En outre, grâce à une approximation numérique de la convolution comme un opérateur matriciel, l'estimation dans le FCVM devient un problème linéaire inverse pour chaque couple (X_i, Y_i). Dans ce cas, pour chaque $i \in \{1, \cdots, n\}$, nous pouvons estimer θ avec des techniques comme la régularisation de Tikhonov, la méthode de décomposition en valeurs singulières, ou des méthodes basées sur des ondelettes (voir par exemple Tikhonov and Arsenin (1977), O'Sullivan (1986), Donoho (1995), Abramovich and Silverman (1998)). Notez encore une fois que ces méthodes ne traitent que le cas $n = 1$. Les propriétés asymptotiques ne sont pas étudiées.

Enfin, une autre méthode apparentée est la déconvolution de Laplace introduite par Comte et al. (2016). Cette méthode traite également du cas $n = 1$. Les auteurs considèrent à la fois la convolution non périodique, comme dans le FCVM, et un bruit fonctionnel.

Dans le chapitre 3 nous avons adapté la déconvolution paramétrique de Wiener, la méthode de décomposition en valeurs singulières, la régularisation de Tikhonov et la déconvolution de Laplace pour estimer θ dans le FCVM.

Contribution de la thèse

Dans cette thèse, nous cherchons à savoir comment le passé du régresseur fonctionnel X influe sur la valeur actuelle de la fonction de réponse Y dans les modèles de régression linéaires.

La thèse, écrite en anglais, est divisée en six chapitres. Le chapitre 1 est une introduction générale, plus exhaustive que ce résumé étendu en français. Les contributions principales de cette thèse sont détaillées dans les chapitres 2 à 4, où nous étudions respectivement le modèle fonctionnel concurrent (chapitre 2), le modèle fonctionnel de convolution (chapitre 3) et le modèle entièrement fonctionnel (chapitre 4). Une illustration sur un jeu de données réelles est faite au chapitre 5. Enfin nous présentons au chapitre 6 les conclusions et les perspectives de cette thèse.

Voici un court résumé de chacun de ces chapitres.

Chapitre 1

Nous y donnons les principales idées sur les modèles de régression linéaire fonctionnelle avec réponse fonctionnelle, ainsi que les définitions et l'arrière-plan théorique utilisé dans les chapitres suivants. Les étapes de l'implémentation numérique des estimateurs sont également détaillées.

Chapitre 2

Dans ce chapitre, nous proposons une approche fonctionnelle pour estimer la fonction inconnue dans le modèle fonctionnel concurrent (FCCM). Cette approche est une généralisation aux données fonctionnelles de la méthode de régression Ridge classique. L'estimateur que nous construisons est ainsi nommé l'estimateur de régression Ridge fonctionnelle (FRRE).

L'importance du modèle FCCM a été mise en évidence dans certains articles et livres, parce que c'est un modèle général auquel tous les modèles linéaires fonctionnels peuvent être réduits (voir par exemple Ramsay and Silverman (2005), Morris (2015), Wang et al. (2016)).

Nous avons prouvé la consistance du FRRE pour la norme L^2, et obtenu sa vitesse de convergence sur l'ensemble des réels, et non pas seulement sur les compacts. Nous avons également fourni une procédure de sélection du paramètre optimal de régularisation λ_n par validation croisée prédictive et par validation croisée généralisée. Les simulations ont montré de bonnes propriétés du FRRE, même sous un très faible rapport signal-bruit. Compte tenu de sa définition simple, le FRRE est plus rapide à calculer que d'autres estimateurs pour le modèle FCCM trouvés dans la littérature, comme celui proposé par Şentürk and Müller (2010).

La définition de cet estimateur le rend apte à être utilisé dans une étape de la procédure d'estimation dans le modèle de convolution fonctionnel, sujet au cœur du Chapitre 3.

Le chapitre 2 est un article que nous avons soumis à Electronic Journal of Statistics.

Chapitre 3

Dans ce chapitre, nous étudions l'estimateur par déconvolution de Fourier fonctionnelle (FFDE) du coefficient fonctionnel dans le modèle (FCVM). Pour ce faire, nous avons mis au point une nouvelle approche qui utilise la dualité des domaines temporel et fréquentiel à travers la transformée de Fourier continue.

Grâce à cette dualité nous associons les modèles FCCM et FCVM et nous pouvons utiliser l'estimateur de régression Ridge fonctionnelle dans le domaine fréquentiel pour définir le FFDE. Cela nous a permis de démontrer la consistance du FFDE pour la norme L^2 et d'obtenir une vitesse de convergence sur l'ensemble des réels. Nous avons également fourni une procédure de sélection du paramètre optimal de régularisation λ_n par validation croisée prédictive avec exclusion.

Nous avons défini d'autres estimateurs pour le FCVM, que nous avons adaptés de différentes méthodes trouvées dans la littérature sur le " problème de déconvolution ". Nous avons ainsi comparé les performances du FFDE avec ces estimateurs. Les simulations ont montré la robustesse, la précision et le temps de calcul rapide du FFDE par rapport aux autres. Le calcul du FFDE est rapide car nous utilisons la transformée de Fourier discrète dans la mise en œuvre numérique. Ceci est une propriété très utile du FFDE.

Ce chapitre est un article bientôt prêt à être soumis.

Chapitre 4

Dans ce chapitre, nous proposons deux estimateurs de l'opérateur de covariance du bruit (Γ_ε) dans la régression linéaire fonctionnelle lorsque la réponse et la covariable sont fonctionnelles, voir le modèle complètement fonctionnel (2). Nous avons étudié les propriétés asymptotiques de ces estimateurs et leur comportement en simulations.

Plus particulièrement, nous avons estimé la trace de l'opérateur de covariance du bruit ($\sigma_\varepsilon^2 = tr(\Gamma_\varepsilon)$). L'estimation de σ_ε^2 rendra possible la construction de tests d'hypothèses dans le cadre du modèle complètement fonctionnel. De plus, σ_ε^2 est impliqué dans la majoration de l'erreur quadratique de prédiction, qui sert à déterminer la vitesse de convergence (Crambes and Mas (2013)). Ainsi avoir un estimateur de σ_ε^2 renseignera sur la qualité de prédiction dans le modèle complètement fonctionnel.

Ce chapitre est un article que nous avons publié dans Statistics and Probability Letters (Volume 113, June 2016, Pages 7–15).

Chapitre 5

Ce chapitre est une illustration de l'implémentation des résultats présentés au chapitre 3. Nous avons utilisé le modèle FCVM (3) et le modèle linéaire fonctionnel historique pour étudier comment la demande évaporative (VPD) influence la vitesse d'élongation foliaire (LER) de plants de maïs. Les données sont des données réelles obtenues dans des plates-formes de phénotypage haut-débit de plantes, lors de deux expériences menées en 2014, **T72A** et **T73A**. Pour les deux expériences, le modèle FCVM est trop simple pour apporter de la connaissance sur l'interaction entre VPD et LER. En revanche, le modèle fonctionnel historique est plus utile pour comprendre cette interaction, car c'est un modèle plus riche.

Pour estimer le noyau historique $\mathscr{K}_{hist}$, nous avons proposé deux estimateurs : l'estimateur de Karhunen-Loève restreint et l'estimateur fonctionnel de Tikhonov. Des deux estimateurs, celui de Tikhonov montre des résultats plus cohérents pour les deux expériences.

Chapter 1

General Introduction

Contents

Thanks to new data collection technologies it is possible to increase the number of observations during a time interval to measure how some quantitative variables dynamically evolve for each of many experimentation subjects (longitudinal data, curve data, time series, etc.). This can be found for instance in biological experimentation, finance, physics, etc. Valorize and exploit these masses of data is a current challenge which will be helpful for these scientific fields (see, e.g., Ullah and Finch (2013) and Wang et al. (2016)).

When analyzing these data some classical methods from multivariate statistical analysis have been shown to be unsuitable (see, e.g., Bickel and Levina (2004), Şentürk and Müller (2010), Hsing and Eubank (2015, p. 1), Ramsay and Silverman (2005, Ch 1), etc). In this sense, some reasons for this inadequacy are for instance: i) the number of observation times p is bigger than the number of realizations n (see, e.g.,Hsing and Eubank (2015, p. 2)); ii) the grid of observation times could slightly differ from one realization to another (see, e.g., Şentürk and Müller (2010)); iii) high correlations between close observation times (see, e.g., Ferraty and Vieu (2006, p. 7)) and iv) the fact that the smoothness and derivatives of the random functions play a major role to study the data (see, e.g., Mas and Pumo (2009)).

A more suitable way to deal with these kind of data is to consider them as "realizations from continuous time stochastic processes" (Hsing and Eubank (2015, Ch 1), Bosq (2000, Ch 1)). This leads to the introduction of more appropriate definitions such as functional data (datasets) and random functions. Functional datasets are realization samples of some random functions. A random function is a random phenomena which has functions as realizations. From a mathematical viewpoint random functions are "measurable functions defined on some probability space with values in an infinite dimensional functional space" (Ferraty and Vieu (2006, Ch 1)). Thorough expositions of these concepts can be found in Section 1.1.

In many applications these functions are time-dependent but they could depend on another unidimensional variable, for instance frequency in the case of spectrometric datasets. In any case whenever these functions are univariate they are also called curve data (see, e.g., Gasser and Kneip (1995)). More generally these functions can be multivariate, for instance when they depend on frequency, time, space or other variables (Morris (2015), Wang et al. (2016, p. 1)). Examples of this more complex kind of datasets are brain and neuroimaging data.

There are many questions that FDA studies, among which functional linear regression is one of the most studied, both in applications and in methodological development (Morris (2015, p. 3)). This is due to the fact that regression models are a good way to study the interrelationship of random functions in diverse fields (see, e.g., Ullah and Finch (2013) and Wang et al. (2016)).

The objective of this thesis is the study of functional linear regression models when both the regressor X (covariate, input) and the response Y (output) are random functions and

time-dependent (i.e. curve data, longitudinal data). In particular we want to address the question of how the history of a random function X (regressor) influences the current value of another random function Y (response) at any given time t. Along with this objective, we want to propose estimation methods which have good properties (consistency, robustness) and which are faster to compute compared to others already proposed in the literature. In order to do this we are mainly interested in two models: the historical functional linear model and the functional convolution model (FCVM) which are introduced in what follows.

The historical functional linear model, introduced by Malfait and Ramsay (2003), has the form

$$Y(t) = \int_0^t \mathscr{K}_{hist}(s,t)X(s)\,ds + \varepsilon(t), \tag{1.1}$$

where $s,t \geq 0$, $\mathscr{K}_{hist}(s,t)$ is the history regression coefficient function and ε is some functional random noise with $\mathbb{E}[\varepsilon] = 0$. This model is a special case of the fully functional regression model (see, e.g., Horváth and Kokoszka (2012, p. 130) and Ramsay and Silverman (2005, Ch 16)), where X and Y are related through a more general kernel operator. This latter can be written as follows

$$Y(t) = \int_I \mathscr{K}(s,t)X(s)\,ds + \varepsilon(t), \tag{1.2}$$

where $s,t \in I \subseteq \mathbb{R}$ and $\mathscr{K}(\cdot,\cdot)$ is the integrable kernel. However the interpretation and the estimation are more easily done when the kernel $\mathscr{K}$ has the simpler form of $\mathscr{K}_{hist}$ because this function is defined over a simpler domain, namely the triangular domain where $s < t$.

An even simplified way to study the influence of the history of X over the current value of Y is through the functional convolution model (FCVM) which is defined next

$$Y(t) = \int_0^t \theta(s)X(t-s)\,ds + \varepsilon(t), \tag{1.3}$$

where $t \geq 0$ and θ is the functional coefficient to be estimated. In this model θ is a function which only depends on s and not on the current time t as $\mathscr{K}_{hist}$ does. Note that all these functions are considered to be equal to zero for all $t < 0$, which is interpreted as the fact that zero is the starting point of the measurements.

Besides these two models (the historical and the FCVM) we have also studied the functional concurrent model (FCCM) defined in Ramsay and Silverman (2005, Ch 14) as follows

$$Y(t) = \beta(t)X(t) + \varepsilon(t), \tag{1.4}$$

where $t \in \mathbb{R}$ and β is the functional coefficient to be estimated. This is one of the two major kinds of functional regression models which deal with functional responses (the other one is the fully functional regression model defined in equation (1.2), see Wang et al. (2016, p. 272)) and its importance was already remarked in Ramsay and Silverman (2005, p. 220).

Our interest in the FCCM arises from the fact that the FCVM and the FCCM are related in a deeper way through the Continuous Fourier Transform, which associates to each convolution estimation problem of θ in the time domain the estimation of β in the frequency domain. This establishes, under some particular conditions, the equivalence between both models (see Section 1.3). Furthermore the historical functional linear model is also related to the FCCM when for instance the historical kernel can be expressed as the product of two univariate functions in the following way, $\mathscr{K}_{hist}(s,t) = \beta(t)\gamma(s)$ (Şentürk and Müller (2010), Kim et al. (2011)).

In the following pages of this chapter we will introduce the main notations, definitions and the theoretical framework which will be used throughout this thesis. Section 1.1 is about general aspects about FDA. In Section 1.2 we review the literature about functional regression models with functional response. In particular we study the major categories of these models, the historical functional linear regression model and the functional convolution model. Then we focus our attention on the estimation of θ in the FCVM in Section 1.3. The numerical implementation of the functional Fourier deconvolution estimator is addressed in Section 1.4. Finally in the Section 1.5 we discuss the contributions of this thesis in theory and practice.

1.1 Functional Data Analysis

1.1.1 Examples of Functional Data Sets

Let us start with three examples of functional datasets. Through these examples we want to give an intuitive idea of what functional data (dataset) refer to. Afterwards we will discuss more theoretical aspects of functional data and random functions. Note that all these examples show a particular kind of functional data referred to as curve data (Wang et al. (2016, p. 258)). We have chosen examples with curve data because they are easier to visualize and because this is the kind of functional data we are concerned with in this thesis.

Mass Spectrometric Data Set : The first example comes from Koomen et al. (2005) and is commented in Morris (2015). The dataset contains mass spectra from a study about pancreatic cancer performed at the University of Texas. To obtain these data blood serum

was taken from 139 pancreatic cancer patients and 117 healthy controls. Then with a mass spectrometry instrument it was obtained the proteomic spectra $X_i(t)$ for each individual $i = 1, \cdots, 256$. The grid of discrete observations (molecular mass per unit charge m/z) for each curve has size $T = 12096$. A subset of this dataset is shown in Figure 1.1.

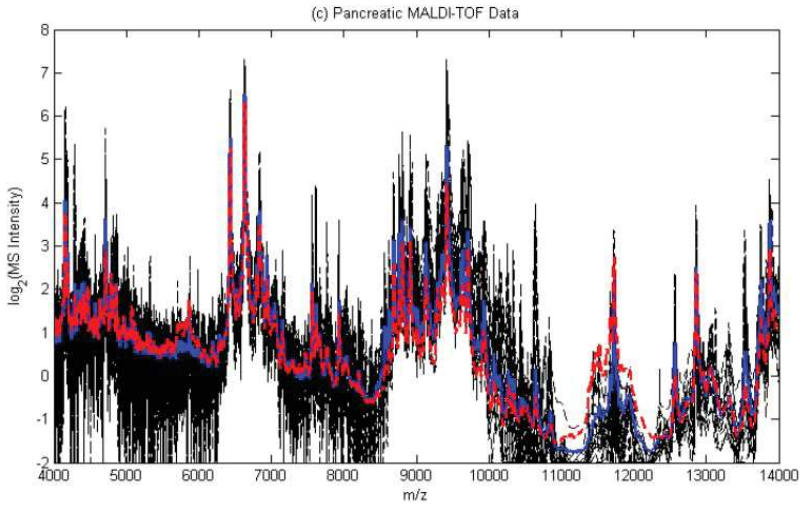

Fig. 1.1 Log spectral intensities from the mass spectrometry data set. Black lines are plotted spectra from 20 pancreatic cancer (solid) and 20 control (dashed) patients, with mean spectra for pancreatic cancer (red) and control (blue).

Oil data set : The second example comes from Ramsay et al. (2009, Ch 1). This is a case in which two random functions arise as input/output pairs and there is a dependency relationship. This dataset has been collected from an oil refinery in Texas. It is shown in Figure 1.2. It represents two variables measured over time in a grid of 193 points. The first variable is the amount of petroleum product at tray level 47 in a distillation column in an oil refinery and the second is the flow of a vapor into the tray (reflux flow). In this case it is known that the amount of petroleum product reacts to the change in the flow of a vapor into the tray. Some functional linear regression models are useful to characterize this dependency. Here both the regressor (input, predictor) and the response (output) are functions.

VPD and LER data set : Finally the last example of functional dataset is about high-throughput plant phenotyping data which was obtained in the project PHENOME. In Figure 1.3 we show 13 pairs of Vapor Pressure Deficit (VPD) and the Leaf Elongation Rate (LER) curves. All these curves have been measured 96 times during one day (one observation every 15 minutes). Again here we have two random functions that arise as input/output pairs with a dependency relationship. It is known that the VPD influences the LER.

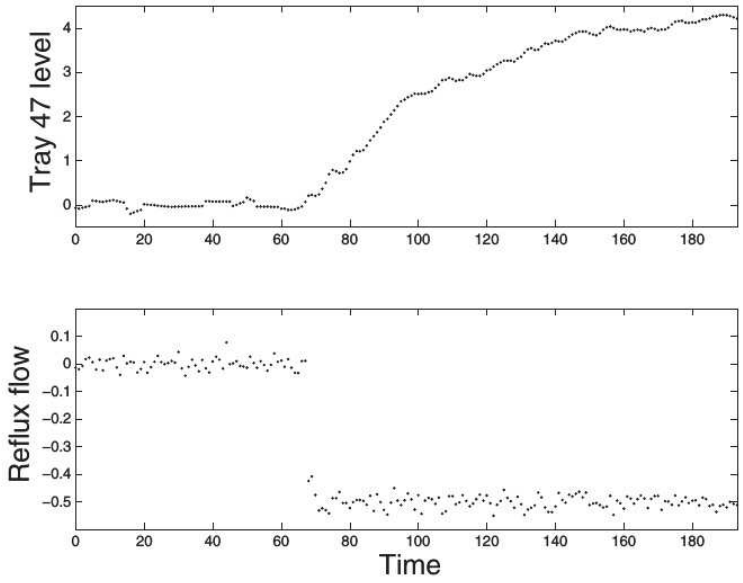

Fig. 1.2 The top panel shows 193 measurements of the amount of petroleum product at tray level 47 in a distillation column of an oil refinery. The bottom panel shows the flow of a vapor into that tray during the experiment.

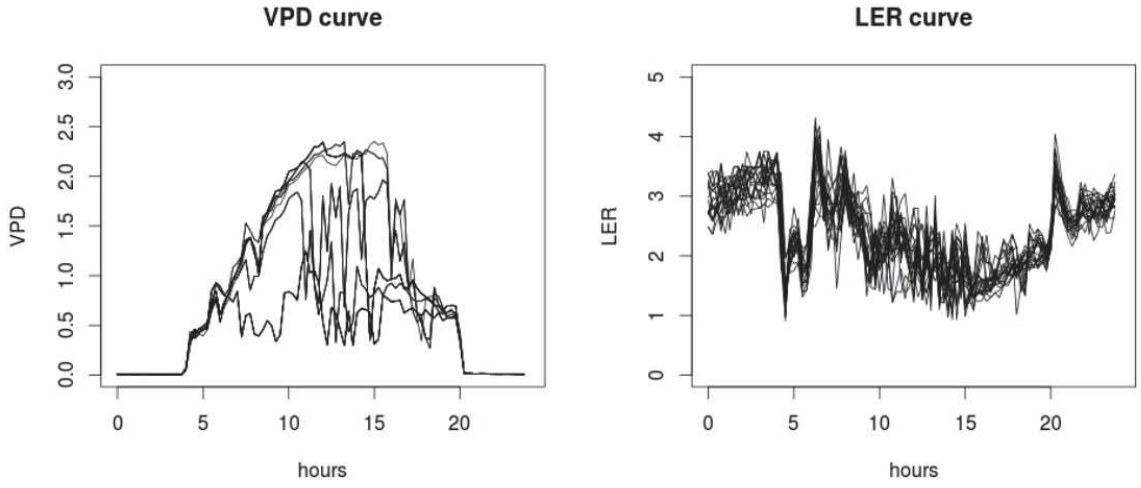

Fig. 1.3 Example of 13 pairs of VPD and LER curves observed 96 times during one day.

In these three examples, each observed curve is a realization of some underlying random function. Thus a functional dataset is a set of many realizations of some random functions. For instance, let X and Y be two random functions. Then the sample $(X_i, Y_i)_{i=1,\cdots,n}$, where each couple (X_i, Y_i) represents two functions (curves) observed over the interval domain $I \subseteq \mathbb{R}$, is called a functional dataset. In practice these curves have been observed in a finite grid of points. A more general and rigorous theoretical definition of a random function is given in the following subsection 1.1.2.

1.1.2 Random Functions

We start with the definition of random functions. They are a natural extension of real valued random variables, which take values in functional spaces and not in $\mathbb{R}$. Generally speaking Ferraty and Vieu (2006, p. 6) define a random function as a measurable function from a probability space $(\Omega, \mathscr{A}, P)$ into a infinite dimensional functional space E. It is common to use E with its Borel σ-algebra generated by its open sets. This functional space may be a Hilbert space (see e.g. Horváth and Kokoszka (2012, Ch 2)), a Banach space (see e.g. Ledoux and Talagrand (1991, Ch 2)), a space of mappings (see e.g. Bosq (2000, p. 16)), etc.

In this thesis we are interested in the case where the functions of E have only one variable (univariate), for instance time or frequency. As stated before, this sort of data are referred to as **curve data** or sometimes longitudinal data. For example this is the case when $E := L^2(I)$, the set of Lebesgue square integrable functions defined over an interval $I \subset \mathbb{R}$. The study of the methods that deal with this kind of data is the main subject of the classical monograph Ramsay and Silverman (2005) or of the recent one Hsing and Eubank (2015).

In what follows we summarize important facts about the definition of the expectation and the covariance operator in Banach spaces and Hilbert spaces (see Hsing and Eubank (2015) for more details).

Expectation : Let B be a separable Banach space and B^* its dual space. A B-valued random function $X : (\Omega, \mathscr{A}, P) \to B$ is said to be **weakly integrable** if and only if i) the composed function $f^*(X)$ is integrable for all $f^* \in B^*$ and ii) there exists an element of B, denoted $\mathbb{E}[X]$, such that for all $f^* \in B^*$

$$\mathbb{E}[f^*(X)] = f^*(\mathbb{E}[X]).$$

The element $\mathbb{E}[X]$ is referred to as the **weak integral** of X. Additionally a B-valued random function X will be defined as **strongly integrable** (or integrable) if $\mathbb{E}[\|X\|_B] < \infty$, where

$\|\cdot\|_B$ is the norm of B. Whenever X is integrable (strongly) it is common to denote $\mathbb{E}[X]$ with $\int X dP$ equivalently.

Let us consider the following equivalence relation among B-random functions X and Y, $X \sim Y$ if and only if $X = Y$ almost surely (**a.s.**). Using the corresponding equivalence classes we define the space $L_B^1(P)$ of equivalence classes of integrable B-random functions. If we define the norm $\|X\|_{L_B^1} := \mathbb{E}\|X\|_B$, $L_B^1(P)$ becomes a Banach space. Analogously for $p \in]1,\infty[$, we define the spaces $L_B^p(P)$ of the classes of B-random functions X such that $\mathbb{E}[\|X\|_B^p] < \infty$. Similarly to $L_B^1(P)$ this space turns to be a Banach space if we define the norm $\|X\|_{L_B^p} := [\mathbb{E}\|X\|_B^p]^{1/p}$.

Covariance Operator: For a random function $X \in L_B^2(P)$, with $\mathbb{E}[X] = 0$, we define its covariance operator as the following bounded linear operator,

$$\begin{array}{rccl} C_X: & B^* & \to & B \\ & f^* & \mapsto & \mathbb{E}[f^*(X)X]. \end{array}$$

In the case where $\mathbb{E}[X] \neq 0$, we define $C_X := C_{X-\mathbb{E}[X]}$. It is possible to generalize this definition to define the **cross-covariance operator** between two random functions. In order to do this let us consider $X \in L_{B_1}^2(P)$ and $Y \in L_{B_2}^2(P)$, where B_1 and B_2 are separable Banach spaces such that $\mathbb{E}[X] = 0$ and $\mathbb{E}[Y] = 0$. The cross-covariance operators of X and Y are the following bounded linear operators:

$$\begin{array}{rccl} C_{X,Y}: & B_1^* & \to & B_2 \\ & f^* & \mapsto & \mathbb{E}[f^*(X)Y] \end{array} \quad \text{and} \quad \begin{array}{rccl} C_{Y,X}: & B_2^* & \to & B_1 \\ & g^* & \mapsto & \mathbb{E}[g^*(X)Y]. \end{array}$$

In the particular case of a separable Hilbert Space H, with inner product $\langle\cdot,\cdot\rangle$, the definition of the expectation is similar to the one defined for Banach spaces. Moreover the definition of the covariance operator will be simpler, because of the Riez representation theorem of the dual space H^*. In this way let X be a H-random function such that $\mathbb{E}[\|X\|_H^2] < \infty$ and $\mathbb{E}[X] = 0$. Then the covariance operator of X is the following

$$\begin{array}{rccl} C_X: & H & \to & H \\ & x & \mapsto & \mathbb{E}[\langle x, X\rangle X]. \end{array}$$

It is known that this operator is symmetric, positive and nuclear (see Bosq (2000, p. 34) for more details). Similarly the cross-covariance operator is defined as follows

$$C_{X,Y}(x) = \mathbb{E}[\langle X, x\rangle Y],$$

for every $x \in H$. This operator is also nuclear (Bosq (2000, p. 34)).

Spectral Decomposition of the Operators : A quite useful fact about the covariance operator C_X, when X is a H-random function such that $\mathbb{E}[\|X\|_H^2] < \infty$ and $\mathbb{E}[X] = 0$, is that there exists the following spectral decomposition

$$C_X(x) = \sum_{j=1}^{\infty} \lambda_j \langle x, v_j \rangle v_j,$$

where the set $(v_j)_{j\geq 1}$ is an orthonormal basis of H called the eigenfunctions of C_X, and $(\lambda_j)_{j\geq 1}$ are the eigenvalues of C_X which satisfy

$$\sum_{j=1}^{\infty} |\lambda_j| = \mathbb{E}[\|X\|_H^2] < \infty,$$

since $\lambda_j = \langle C_X(v_j), v_j \rangle = \mathbb{E}(\langle X, v_j \rangle^2)$.

This spectral decomposition is used to project the random functions into a finite dimensional subspace generated by the first K eigenfunctions in many well-known methods of Functional Analysis such as the Functional Principal Components Analysis (see, e.g., (Ramsay and Silverman, 2005, Ch 8)) and Karhunen-Loève (see, e.g., Crambes and Mas (2013)) estimation methods for the functional linear regression model.

Sequence of Random Functions in Hilbert Spaces

In a similar way as for real valued random variables the Strong Law of Large Numbers (SLLN) apply for independent and identically distributed (i.i.d) sequences of H-random functions (see Bosq (2000)).

Theorem 1. *Let $(X_i)_{i\geq 1}$ be a sequence of i.i.d. H-random functions with expectation $\mathbb{E}[X] \in H$. Then*

$$\frac{\sum_{i=1}^{n} X_i}{n} \quad \xrightarrow{a.s.} \quad \mathbb{E}[X],$$

when $n \to \infty$.

This theorem is also true for i.i.d. B-random functions with B a separable Banach space. A generalization of the Central Limit Theorem (CLT) for Hilbert spaces was also obtained by Varadhan (1962).

Theorem 2. *Let $(X_i)_{i\geq 1}$ be a sequence of i.i.d. H-random functions, where H is separable. If $\mathbb{E}[X] = m \in H$, $\mathbb{E}\|X\|_H^2 < \infty$ and its covariance operator is C_X. Then*

$$\frac{\sum_{i=1}^{n} X_i - m}{\sqrt{n}} \quad \xrightarrow{d} \quad N,$$

when $n \to \infty$, where $N \sim \mathcal{N}(0, C_X)$ is a Gaussian process with mean zero and covariance operator C_X.

In general the CLT does not hold in Banach spaces. But under stronger conditions it is possible to obtain it. In particular Ledoux and Talagrand (1991, Ch 9 and 10) show that the geometry of the space (type and cotype structure) is intrinsically linked to the CLT property. The authors give the characterization of the CLT in separable Banach spaces through the small ball criterion (Ledoux and Talagrand (1991, Section 10.3)).

1.1.3 FDA and Multivariate Statistical Analysis

In this subsection we address the question of the inadequacy of some methods of the Multivariate (Statistical) Analysis to deal with functional data. As we mentioned earlier Wang et al. (2016, p. 1) define FDA as "the analysis and theory of data that are in the form of functions, images and shapes, or more general objects", whereas Multivariate Analysis is concerned in understanding and analyzing the interaction of many statistical outcome variables which in general are measured simultaneously (Johnson and Wichern, 2007, p. 1).

Hsing and Eubank (2015, p. 1) and Ramsay and Silverman (2005, Ch 1) among others have shown that some classical methods from Multivariate Analysis (MVA) are unsuitable to deal with functional data. There are four main reasons for this inadequacy:

1. One requirement to apply MVA methods to functional data is that the grid of observations must be fixed and the same for all the realizations. This is not required by FDA and thus it could be applied to more cases, for instance when the observation times are random and independent among the realizations (see e.g. Şentürk and Müller (2010, p. 1257), Yao et al. (2005b, p. 578)).

2. The number of observation times p is in general bigger than the number of realizations n (Hsing and Eubank (2015, p. 2)). The study of statistical inference under this condition is not possible with classical methods from multivariate analysis. High-dimensional statistics (Bühlmann and van de Geer (2011, Ch 1)) and FDA are suitable for this type of data. One of the reasons why $p >> n$ is problematic for multivariate analysis is the fact that it makes the covariance operators to be non-invertible (ill-conditioning). This in turn makes difficult to solve linear systems in regression models which are widely used in MVA.

3. High correlations of the measured variables when observed in close observation times (see e.g. Yao et al. (2005b), Ferraty and Vieu (2006, p. 7)) is an ill-conditioned problem and then is not suitable for MVA methods because it makes difficult to solve linear systems.

4. Finally in the case where the smoothness and derivatives of the random functions play a major role to study the data (see e.g. Mas and Pumo (2009), Ramsay and Silverman (2005, Ch 17)), it is necessary to consider the functional nature of the data, and this cannot be accomplished with MVA approach.

1.2 Functional Linear Regression Models with Functional Response

The aim of this section is to present a succinct review of some models that are used to study the dependency relationship between two random functions. Here we consider random functions defined in the Hilbert space $L^2(I)$, i.e. the space of Lebesgue square integrable functions defined on the interval $I \subseteq \mathbb{R}$. Let X and Y be two random functions and consider X as the regressor (predictor, input, explanatory) and Y as the response (output, dependent). A natural model that relates X and Y is

$$Y = \Psi(X) + \varepsilon,$$

where Ψ is a functional operator and ε is a noise random variable.

In this model Ψ summarizes the way how X acts upon Y. Thus estimating Ψ is a key element in order to understand this relationship. The estimation question is stated as follows: how to define an estimator from a sample of n i.i.d. realizations of X and Y, $(X_i, Y_i)_{i=1,\cdots,n}$, to estimate Ψ and how is the asymptotic behavior of this estimator (consistency and rate of convergence).

There are two main approaches to deal with this problem. First the functional parametric approach requires Ψ to belong to a particular subset $\mathscr{S}$ of the continuous linear operators $\mathscr{C}$ on $L^2(I)$, which can be indexed by a finite number of some fixed continuous linear operators on $L^2(I)$ (Ferraty and Vieu (2006, p. 8)). An example of this case is the functional linear regression model with functional response defined with equation (1.2), where the continuous linear operator is parametrized solely with one kernel integral operator.

The second approach is the non-parametric functional. In this approach the set $\mathscr{S}$ is not required to be indexed by a finite set of continuous linear operators (Ferraty and Vieu

(2006, p. 8)). The only constraint is the regularity of the elements of $\mathscr{S}$. Some ways to accomplish this are through the functional additive models studied by Müller and Yao (2012), the Reproducing kernel Hilbert spaces (see e.g. Lian (2007), Kadri et al. (2010)) and kernel methods (Ferraty and Vieu (2006)).

Throughout this thesis we are interested in the functional parametric approach, more specifically in the functional linear regression model with Functional Response. In what follows we give a more detailed study of this model.

1.2.1 Two Major Models

The goal of this section is to review the literature about the Functional Linear Regression Models with Functional Response, that is when both the covariate (input) and the response (output) are functional. In a recent article, Wang et al. (2016) propose to divide this class of models into two major categories. The functional concurrent model (FCCM) and the fully functional regression model (see Horváth and Kokoszka (2012, p. 130)). We briefly discuss both of them.

Functional Concurrent Model : The first one is referred to as the functional concurrent model and its equation was given in (1.4), let us recall it

$$Y(t) = \beta(t)\, X(t) + \varepsilon(t),$$

where $t \in \mathbb{R}$. Here $\beta_0(t)$ and $\beta_1(t)$ are the functional coefficients of the model to be estimated.

Some related models have already been discussed by several authors. For instance West et al. (1985) defined a similar model called 'dynamic generalized linear model' and they study the model from a Bayesian point of view. Hastie and Tibshirani (1993) proposed the 'varying-coefficients model': In this model Y is supposed to be a random variable whose distribution depends on a parameter η, of the form

$$\eta = \beta_0(R_0) + X_1\beta_1(R_1) + \cdots X_p\beta_p(R_p), \tag{1.5}$$

where $X_1, \cdots, X_p$ and $R_1, \cdots, R_p$ are the predictors. Here $\beta_1, \cdots, \beta_p$ are the functions to be estimated.

In the simplest case of the Gaussian model, $\eta = \mathbb{E}[Y]$ and Y is normally distributed with mean η and equation (1.5) takes the form

$$Y = \beta_0(R_0) + X_1\beta_1(R_1) + \cdots X_p\beta_p(R_p) + \varepsilon,$$

where $\mathbb{E}[\varepsilon] = 0$ and $var[\varepsilon] = \sigma^2$. In the case where all the covariates $R_1, \cdots, R_p$ are the time variable and when the covariates $X_1, \cdots, X_p$ are time-dependent the model can take the form of the functional concurrent model, that is

$$Y(t) = \beta_0(t) + X_1(t)\beta_1(t) + \cdots X_p(t)\beta_p(t) + \varepsilon.$$

Afterwards many people studied this model trying to estimate the unknown smooth regression functions β_i. For instance Wu et al. (1998) proposes a kernel based method when there are k functional covariates. Dreesman and Tutz (2001) and Cai et al. (2000) propose a local maximum likelihood type of estimation. Zhang and Lee (2000); Fan et al. (2003); and Zhang et al. (2002) propose local polynomial smoothing methods. Fan and Zhang (2000) propose a two-step approach, first performing pointwise Ordinary Least Squares regressions and then smoothing these raw estimators. Huang et al. (2004) propose the use of B-splines to represent the functions β_i and to use the least square method to perform the estimation. A good review about these methods can be found in Fan and Zhang (2008). A further development of the multivariate varying-coefficient model is presented in Zhu et al. (2014).

We want to highlight that although the varying coefficient model is related to the FCCM it was originally designed for the case where t is a random variable and was treated with multivariate analysis methods. Thus the first works about estimation did not take into account the functional nature of the data, as noticed by Ramsay and Silverman (2005, p. 259). In contrast, Şentürk and Müller (2010) propose an estimation method which is closer to the functional data approach.

Fully Functional Regression Model : This model has been defined in the equation (1.2) and is recalled here :

$$Y(t) = \int_I \mathscr{K}(s,t)X(s)\,ds + \varepsilon(t).$$

This model have been first studied by Ramsay and Dalzell (1991). Estimation of the kernel $\mathscr{K}$ is an inverse problem and requires some sort of regularization to achieve inversion. Ramsay and Dalzell (1991, p. 552) proposed a penalized least square method to do this. James (2002) proposed penalized polynomial splines. Cuevas et al. (2002) studied this regularization when the design is fixed. He et al. (2000) defined the 'Functional Normal Equation' which generalizes the multivariate normal equations and they proved that there exists a unique solution under certain conditions. They used the Karhunen-Loève decomposition to estimate $\mathscr{K}$. Chiou et al. (2004) gave a good summary of this approach. This idea has been revisited in Yao et al. (2005a) where they dealt with sparse longitudinal data and when the observation

times are irregular and random. Additionally Aguilera et al. (2008) proposed a wavelet based estimator. Antoch et al. (2010) studied spline estimator. Finally Crambes and Mas (2013) have studied a Karhunen-Loève type estimator.

Whenever Y and X are time-dependent, in the equation (1.2) future values of X are used to explain past values of Y. This fact is counterintuitive and should not be used. For this reason Malfait and Ramsay (2003) are concerned with the particular case where $Y(t)$, the value of Y at time t, depends on the history $\{X(s) : 0 \leq s \leq t\}$. This leads to the historical functional linear model defined in equation (1.1). This particular type of model will be discussed in the following subsection.

1.2.2 Historical Functional Linear Regression Model

As mentioned earlier the historical functional linear model has been proposed in Malfait and Ramsay (2003). It was defined in equation (1.1), namely

$$Y(t) = \int_0^t \mathscr{K}_{hist}(s,t)X(s)\,ds + \varepsilon(t),$$

for all $t \in [0,T]$ and $T \in \mathbb{R}$. Now we want to summarize the approach of Malfait and Ramsay (2003) to estimate $\mathscr{K}_{hist}$. The authors propose to use a finite element basis $\phi_k(s,t)$. This basis is made of piecewise linear functions defined over a fixed and suitable grid of points. Then the authors use the approximation $\mathscr{K}_{hist}(s,t) \approx \sum_{k=1}^{K} b_k\,\phi_k(s,t)$ to transform the problem as follows

$$Y_i(t) = \sum_{k=1}^{K} b_k\,\psi_{ik}(s,t) + \varepsilon_i(t),$$

where $\psi_{ik}(s,t) := \int_0^t X_i(s)\,\phi_k(s,t)\,ds$.

Let $\mathbf{y}(t)$ and $\mathbf{e}(t)$ be the vectors of length N containing the values $Y_i(t)$ and $\varepsilon_i(t)$, respectively. Let also $\Psi(t)$ be the $N \times K$ matrix containing values of $\psi_{ik}(t)$ and let the coefficient vector $\mathbf{b}$ be $(b_1,\cdots,b_K)'$. Then we have the matrix expression of (1.1) for each $t \in [0,T]$,

$$\mathbf{y}(t) = \Psi(t)\,\mathbf{b} + \mathbf{e}(t).$$

This leads to the normal equations $\left(\int_0^T \Psi(t)'\Psi(t)dt\right)\mathbf{b} = \int_0^T \Phi(t)'\mathbf{y}(t)dt$. Finally they approach and solve this equation with a multivariate linear model with penalization.

Following a similar approach, Harezlak et al. (2007) considered the same representation of the function $\mathscr{K}_{hist}(s,t)$ through the basis functions $\phi_k(s,t)$. But in this case the authors explored two different regularization techniques which use basis truncation, roughness

penalties, and sparsity penalties. The first one penalizes the absolute values of the basis function coefficient differences (LASSO approach) and the second one penalizes the squares of these differences (penalized spline methodology). Finally they assessed the estimation quality with an extension of the Akaike Information Criterion.

Kim et al. (2011) are interested in a particular type of the historical functional linear model. They consider the curves X_i and Y_i as being sparse longitudinal data, and the function $\mathscr{K}_{hist}(s,t)$ to be decomposable as follows, $\mathscr{K}_{hist}(s,t) = \sum_{k=1}^{K} d_k(t)\phi_k(s)$, where $d_k(t)$ are the functional coefficients to be estimated and $\phi_k(s)$ are predetermined basis functions, for instance a B-spline basis.

In this case the model (1.1) becomes a varying-coefficients model or functional concurrent model (FCCM) after integration. The estimation procedure uses the functional principal components decompositions for both Y and X. These decompositions are needed to solve the normal equation which uses the auto-covariances of $X(s)$ and $Y(t)$ and the cross-covariance of $(X(s), Y(t))$.

Şentürk and Müller (2010) are interested in a similar model. The authors consider the following modification of (1.1)

$$Y(t) = \beta_0(t) + \beta_1(t) \int_0^{\Delta} \gamma(s)X(s)\,ds + \varepsilon(t),$$

for $t \in [\Delta, T]$ with $\Delta > 0$ and a suitable $T > 0$. Similarly to Kim et al. (2011) this particular case is strongly related to the FCCM. It is clear that after estimating the function γ, the problem becomes a FCCM. Şentürk and Müller (2010) develop a new method to estimate the functions β_0 and β_1, taking into account the functional nature of the data. However the essential idea to estimate the unknown functions here (γ, β_0 and β_1) is to use an equivalent idea to the 'functional normal equation', that is to relate the auto-covariance of X with the cross-covariance of X and Y, and then to invert the auto-covariance of X to obtain the unknown function.

Finally it is possible to consider a more specific structure for the function $\mathscr{K}_{hist}(s,t)$ which does not depend on the current value t but only on s and the integral takes the form of a convolution. In order to do this first we suppose there is a function θ such that $\mathscr{K}_{hist}(s,t) = \theta(t-s)$ and then we apply a change of variables to obtain the functional convolution model (1.3). The study of this model is the goal of the following subsection.

1.2.3 Functional Convolution Model (FCVM)

One of the main goals of this thesis was to study the influence of the history of the functional covariate X on the current value of the functional response $Y(t)$. One way to model this

dependency relationship is through the FCVM. In this subsection we present the origin of this model and in the following section we discuss many estimation methods of the functional coefficient θ and the place of the FCVM among other models in the literature.

Let us recall the equation (1.3) which defines the FCVM, that is

$$Y(t) = \int_0^t \theta(s) X(t-s) ds + \varepsilon(t),$$

for all $t \geq 0$. This model was derived from Malfait and Ramsay (2003). We are interested in the estimation of θ for a given i.i.d. sample $(X_i, Y_i)_{i \in \{1,\cdots,n\}}$ of the random functions X and Y.

The FCVM is a functional extension of distributed lag models in time series (Greene (2003, Ch 19)). One of these models is the dynamic regression model, which has the following general form

$$y_t = \alpha + \sum_{i=0}^{\infty} \beta_i x_{t-i} + \varepsilon_t.$$

for all $t \geq 0$. In this model if we suppose that $x_i = 0$ for all $i < 0$ (i.e. there is a starting point) then the sum becomes finite and this is clearly the discretization of the FCVM. Applications of this model are commented in Greene (2003, Ch 19).

The question of the estimation of θ is central in our study of the FCVM. We address this question in the next section. Besides we comment on a deeper connection between the FCVM and the FCCM obtained through the use of the Continuous Fourier Transform.

1.3 Estimation of θ in the FCVM

There are four main characteristics of the FCVM : i) the covariate (input) and the response (output) are random functions, ii) the convolution is non-periodic (i.e. we do not consider periodic functions), iii) the sample size is $n > 1$, moreover we are interested in the asymptotic behavior ($n \to \infty$) and finally iv) the noise is functional. To the best of our knowledge there are few papers studying this model satisfying these four characteristics. However there are many models which are close to FCVM. In what follows we explore some of them.

Asencio et al. (2014) study a related problem, in which they consider more covariate functions (predictors). The estimation of θ is done by projecting the functions into finite-dimensional spline basis and using a penalized ordinary least square approach to estimate the coefficients in this basis. Another approach is to use the Historical Functional Linear Model (Malfait and Ramsay (2003)) to estimate θ, by taking into account the special shape that the kernel function $\mathscr{K}_{hist}$ shall have in this particular case. In the same way we can use

the approach of Harezlak et al. (2007), or even in a more restricted case the approaches of Kim et al. (2011) and Şentürk and Müller (2010). Note that the FCVM could be seen as a particular case of the model proposed by Kim et al. (2011), namely when $K = 1$, and $\phi_K = \theta$ in $\mathscr{K}_{hist}(s,t) = \sum_{k=1}^{K} b_k(t)\phi_k(s)$.

Şentürk and Müller (2010, p. 1259) proposed a method to estimate θ when the FCVM has the following more restricted form

$$Y(t) = \int_0^{\Delta} \theta(s)X(t-s)ds + \varepsilon(t),$$

where $\Delta > 0$ is a fixed value. The estimation of θ in this case is made by using the Karhunen-Loève decomposition of the covariance operator of the random function $Z_t(s) := X(t-s)$, where t is fixed and $s \in [0,\Delta]$. They express θ in the basis of eigenfunctions of this operator and then they estimate the coefficients with an ordinary least squares procedure. This produces one estimator for each time t. They consider a grid of observation times and then they take the mean of all these estimators. This approach is similar to that of Kim et al. (2011).

To the best of our knowledge only these papers have addressed the study of the estimation of θ under the four characteristics of the FCVM mentioned earlier. The approach we propose in Chapter 3 is a new way to answer this question. We do not use projection into finite dimensional basis nor kernel estimation with finite elements methods. Moreover we study the asymptotic properties of the estimator, which has not been done either in the approaches previously mentioned.

1.3.1 Functional Fourier Deconvolution Estimator (FFDE)

We propose the Functional Fourier Deconvolution Estimator (FFDE) which is defined in three steps. i) First we use the Continuous Fourier Transform ($\mathscr{F}$) to transform the convolution in the time domain into a multiplication in the frequency domain (see (3.2)). ii) Once in the frequency domain, we estimate β with the **Functional Ridge Regression Estimator (FRRE)** defined in Manrique et al. (2016) (see Ch 2), which is an extension of the Ridge Regularization method (Hoerl (1962)) that deals with ill-posed problems in the classical linear regression. iii) The last step consists in using the Inverse Continuous Fourier Transform to estimate θ. This definition is formalized mathematically as follows.

Let $(X_i, Y_i)_{i=1,\cdots,n}$ be an i.i.d sample following the FCVM (1.3).

Step i) We use the Continuous Fourier Transform ($\mathscr{F}$) defined as follows

$$\mathscr{F}(f)(\xi) = \int_{t=-\infty}^{+\infty} f(t) e^{-2\pi i t \xi}\, dt,$$

where $\xi \in \mathbb{R}$ and $f \in L^2$. This operator is used to transform the FCVM (1.3) which is defined in the time domain into its equivalent in the frequency domain. Thus equation (1.3) becomes

$$\mathscr{Y}(\xi) = \beta(\xi)\, \mathscr{X}(\xi) + \varepsilon(\xi), \tag{1.6}$$

where $\xi \in \mathbb{R}$, $\beta := \mathscr{F}(\theta)$ is the functional coefficient to be estimated. $\mathscr{X} := \mathscr{F}(X)$ and $\mathscr{Y} := \mathscr{F}(Y)$ are Fourier transforms of X and Y. Lastly $\varepsilon := \mathscr{F}(\varepsilon)$ is an additive functional noise.

The equivalent problem (eq. (1.6)) in the frequency domain is a particular case of the FCCM (eq. (1.4)). Clearly the estimation of β implies the estimation of θ through $\mathscr{F}^{-1}$.

Step ii) The functional Ridge regression estimator (FRRE) of β in the FCCM (1.4) or in (1.6) is defined as follows

$$\hat{\beta}_n := \frac{\frac{1}{n}\sum_{i=1}^{n} \mathscr{Y}_i\, \mathscr{X}_i^*}{\frac{1}{n}\sum_{i=1}^{n} |\mathscr{X}_i|^2 \,+\, \frac{\lambda_n}{n}}, \tag{1.7}$$

where the exponent $*$ stands for the complex conjugate and λ_n is a positive regularization parameter.

We have defined the functional Ridge regression estimator of β, see (1.7), because with this estimator it is natural to use the inverse Fourier transform ($\mathscr{F}^{-1}$) to estimate θ and to prove the consistency property under the L^2-norm. Besides this we wanted to use the computation efficiency of the Fast Fourier Transform Algorithm.

As we saw before the idea of transforming the historical functional linear model into a FCCM was already proposed by Kim et al. (2011) and in a different way by Şentürk and Müller (2010). In both papers the authors used special structures for the kernel function $\mathscr{K}_{hist}$. These structures allow them to transform the historical model into the FCCM. In our case we use a different approach. We do not impose a particular structure to the kernel function, but we transform the whole model FCVM in the time domain into its equivalent in the frequency domain. As a consequence, it opens the possibility to also use other estimation methods of β in the FCCM (see Subsection 1.2.1) in order to estimate θ in the FCVM.

Step iii) The **FFDE** of θ in (1.3) is defined by

$$\hat{\theta}_n := \mathscr{F}^{-1}(\hat{\beta}_n). \tag{1.8}$$

Note that the estimator $\hat{\theta}_n$ (FFDE) is real valued and belongs to $L^2(\mathbb{R},\mathbb{R})$ (see Chapter 3). Another important property is that the FFDE can be decomposed as follows

$$\hat{\theta}_n = \theta - \frac{\lambda_n}{n}\mathscr{F}^{-1}\left(\frac{\mathscr{F}(\theta)}{\frac{1}{n}\sum_{i=1}^{n}|\mathscr{F}(X_i)|^2 + \frac{\lambda_n}{n}}\right) + \mathscr{F}^{-1}\left(\frac{\frac{1}{n}\sum_{j=1}^{n}\mathscr{F}(\varepsilon_j)\overline{\mathscr{F}(X_j)}}{\frac{1}{n}\sum_{i=1}^{n}|\mathscr{F}(X_i)|^2 + \frac{\lambda_n}{n}}\right). \quad (1.9)$$

The study of this decomposition will allow us to prove the consistency of this estimator. Note the importance of the equivalence between the FCVM and the FCCM, because of the use of two equivalent representations of the same information (time domain and frequency domain) obtained thanks to the Continuous Fourier Transform.

The functional Fourier deconvolution estimator (FFDE) of θ in the FCVM is further studied in Chapter 3. The aim of proposing such an estimator was to take advantage of the equivalence between the time and frequency domains as well as of the mathematical properties of the Continuous Fourier Transform. The advantages of this estimator are both theoretical and practical: theoretical because we develop an approach which uses primarily the fact that we work with random functions and functional spaces, and practical because for implementing this method we use the Fast Fourier Transform (FFT) algorithm which increases the computation speed of the estimators in a significant way over other possible estimators. We describe in the following subsection other possible estimators adapted from the literature.

1.3.2 Deconvolution Methods in the Literature

Next let us consider other models which are indirectly related to the FCVM. From this we will be able to adapt some techniques to estimate θ in the FCVM.

We start with the multichannel deconvolution problem (see e.g. De Canditiis and Pensky (2006), Pensky et al. (2010) and Kulik et al. (2015)). This problem belongs to Signal Processing methods. Similarly to the FCVM, here the input and output are functionals (i.e. signals, curve data), there are many realizations ($n > 1$, multichannel) and the noise is functional. But the difference with the FCVM is that they study the periodic case (the signals are periodical and so is the convolution). Besides the authors do not deal with the asymptotic behavior of the estimator.

The multichannel deconvolution problem is one way to generalize the deconvolution problem in Signal Processing (see e.g Johnstone et al. (2004), Brown and Hwang (2012), Gonzalez and Eddins (2009)). In this problem they use the convolution (periodic or not) to model how an impulse response function h transforms an original signal g (unknown)

through the following equation

$$f(t) = \int_D h(s)g(t-s)ds + \varepsilon(t),$$

where D is the domain of integration ($[0,T]$ in the periodic case for a fixed period T, and $[0,t]$ or $\mathbb{R}$ in the non-periodic one), f is the observed signal and ε the noise. There are several methods to estimate g given the functions h and f, for instance the Parametric Wiener Deconvolution (Gonzalez and Eddins (2009, Ch 5)).

It is clear that if we take one couple (X_i, Y_i) and we interpret f as Y, h as X and g as θ we can apply these methods to estimate θ (apply the deconvolution to Y to obtain θ in (1.3)). At this point we notice that although this problem is related to the FCVM it only deals with the case $n = 1$, and so there is no study of the asymptotic behavior of the estimator.

In a similar way the Deconvolution Problem in Non-parametric statistics (see e.g. Meister (2009), Johannes et al. (2009)) deals with the case $n = 1$ and does not consider a functional noise. The goal here is to estimate the probability density function (pdf) of a real random variable X from the observation of another real random variable Y such that $Y = X + Z$, the pdf of Z being known. To solve this problem they use the fact that the pdf of the sum of two random variables is the convolution of their respective pdf. It might be possible to adapt these techniques to estimate θ in the FCVM, but we think that the estimation would be worse than the one with signal processing methods, because in the former case the functional noise is not considered.

Also, through a numerical approximation of the convolution as a matrix operator, the FCVM becomes a Linear Inverse Problem for each couple (X_i, Y_i). In this case for each $i \in \{1, \cdots, n\}$, we can estimate θ with some of the techniques to solve the linear inverse problem, such as the Tikhonov regularization, the singular value decomposition method, or wavelet based methods (see, e.g., Tikhonov and Arsenin (1977), O'Sullivan (1986), Donoho (1995), Abramovich and Silverman (1998)). Note again that these methods only deal with the case $n = 1$. They do not study the asymptotic case.

Finally another related method is the Laplace deconvolution introduced by Comte et al. (2016). This method also deals with the case $n = 1$. But the authors consider both the non-periodic convolution, as in the FCVM, and a functional noise.

In Chapter 3 we have adapted the parametric Wiener deconvolution, the singular value decomposition method, the Tikhonov regularization and the Laplace deconvolution to estimate θ in the FCVM.

The Section 1.4 deals with the numerical implementation of the FFDE.

1.4 Numerical Implementation of the Functional Fourier Deconvolution Estimator

In this section we discuss how we estimate θ in the FCVM in practice. In particular we describe the necessity to rethink the FCVM in a finite discrete way, and to use the Discrete Fourier Transform as the discrete equivalent of the Continuous Fourier Transform in this new context. We start by describing the discretization of the convolution. To do this properly we start with some definitions.

Throughout this section we use Δ as the discretization step between two observation times (for instance $\Delta = 0.01$). The observation times are defined for every $j \in \mathbb{Z}$ as $t_j := j * \Delta$ and thus they define the grid G_Δ over $\mathbb{R}$. We use a fix grid in this section. With this grid we transform each function $f : \mathbb{R} \to \mathbb{C}$ to a vector $f^d \in \mathbb{C}^{\mathbb{Z}}$ infinite dimensional, with elements $f_j^d := f(t_j) \in \mathbb{C}$. In what follows the superscript ^{d} will denote this discretization.

In this section all the functions will have compact support. Otherwise we should compute convolution of infinite vectors which cannot be done in practice. For simplicity we consider all the functions defined over a compact interval $[0,T]$ with T large enough. Thus we will consider $f^d = f_0^d, \cdots, f_{q-1}^d \in \mathbb{C}^q$, where $q-1 = \max\{j \in \mathbb{N} \,|\, t_j \in [0,T]\}$.

Let RM (rectangular method) be the operator which associates to an integral over $\mathbb{R}$, its numerical approximation by the rectangular method over the grid of points we have already defined. So for a given integral $J = \int_{\mathbb{R}} f(s)ds = \int_0^T f(s)ds$, $RM(J) := \Delta \sum_{j=0}^{q-1} f(t_j) = \Delta \sum_{j=0}^{q-1} f_j^d$.

Understanding how to compute numerically the convolution of two functions is a key element to implement the estimator developed for the FCVM.

We start our discussion by describing the discretization of the convolution of two functions with support included on $[0,T]$,

$$f * g(t) := \int_{-\infty}^{+\infty} f(s)g(t-s)ds = \int_0^T f(s)g(t-s)ds.$$

Approximating this convolution with the Rectangular Method we obtain for every $j \in \mathbb{N}$,

$$RM(f*g)(t_j) = \sum_{l=0}^{q-1} f(t_l)\, g(t_{j-l})\, \Delta = \Delta \sum_{l=0}^{q-1} f_l^d g_{j-l}^d. \tag{1.10}$$

The last sum in equation (3.21) is the convolution between vectors. Thus we can rewrite this equation as follows

$$RM(f*g)(t_j) = \Delta\, (f^d * g^d)_j.$$

for $j \in [0,\cdots,2p-2]$ and where $(f^d * g^d)_j := \sum_{l=0}^{q-1} f_l^d g_{j-l}^d$. Besides note that for $j \notin [0,\cdots,2p-2]$ we have $RM(f*g)(t_j) = 0$ since f and g have compact support.

Additionally we can compute the vector $((f^d * g^d)_0,\cdots,(f^d * g^d)_{2q-2})$ using matrices as follows

$$\Big((f^d * g^d)(0),\cdots,(f^d * g^d)(2q-2)\Big)^T = MC_G\,(f_0^d,\cdots,f_{q-1}^d)^T, \tag{1.11}$$

where MC_G is the matrix associated to the convolution discretized over the grid G, defined as follows

$$MC_G := \begin{pmatrix} g_0^d & 0 & 0 & 0 & \cdots & 0 \\ g_1^d & g_0^d & 0 & 0 & \cdots & 0 \\ g_2^d & g_1^d & g_0^d & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \cdots & \vdots \\ g_{q-2}^d & \cdots & \cdots & g_1^d & g_0^d & 0 \\ g_{q-1}^d & g_{q-2}^d & \cdots & \cdots & \cdots & g_0^d \\ 0 & g_{q-1}^d & g_{q-2}^d & \cdots & g_2^d & g_1^d \\ 0 & 0 & g_{q-1}^d & \cdots & \cdots & g_2^d \\ \vdots & \vdots & \vdots & \ddots & \cdots & \vdots \\ 0 & 0 & \cdots & 0 & g_{q-1}^d & g_{q-2}^d \\ 0 & 0 & 0 & \cdots & 0 & g_{q-1}^d \end{pmatrix} \in \mathbb{R}^{(2q-1)\times q}.$$

Remark : From this we note that the convolution could have a larger support. This arises because an important property of the convolution is that $supp(f*g) \subset \overline{supp(f)+supp(g)}$ (Brezis (2010, p. 106)). Thus in our case $supp(f*g) \subset [0,2T]$. However afterwards we will take T large enough to contain even the convolution. In this way, every time we will consider the convolution of two functions f and g we suppose $\overline{supp(f)+supp(g)} \subset [0,T]$. In this case the number of discretization points q will be defined as before, namely $q-1 = \max\{j \in \mathbb{N} \,|\, t_j \in [0,T]\}$ but now for all $j \geq q$, $(f^d * g^d)_j = 0$. Besides the matrix representation of the convolution through MC_G will still be correct.

In the following subsection we explore the parallel between the continuous convolution of two functions and the convolution of two vectors with respect to the whole model FCVM.

1.4.1 The Discretization of the FCVM and the FFDE

We have defined the functional Fourier deconvolution estimator of θ in the FCVM using the continuous Fourier transform and its inverse (equations (1.7) and (1.8)). Given that both operators are integral operators, we need to use some kind of numerical approach to compute them. The goal of this subsection is to show that the proper way for doing this is by using a

discrete model which behaves like the FCVM. This model will be based on the convolution of finite dimensional vectors. It will be studied through the discrete Fourier transform and its inverse instead of their continuous counterparts.

First let us show that it is not practical to compute the Functional Fourier Deconvolution estimator by direct approximation of the Continuous Fourier Transform and its inverse. This is not possible because these two operators are integrals defined over the whole $\mathbb{R}$. To see why this is a problem let us consider a function $f \in L^2$ with compact support. Then although it is possible to use the Rectangular Method to compute $\mathscr{F}(f)(\xi)$ for every value ξ, we cannot ensure that $\mathscr{F}(f)$ has compact support ((Kammler, 2008, p. 130)). This implies that we need to know the values of $\mathscr{F}(f)$ for all the infinite values of the grid G_Δ to approximate the $\mathscr{F}^{-1}$, which is impossible in practice. Note that even if $\mathscr{F}(f)$ has a compact support we cannot know how large it is and in this case we will need to compute $\mathscr{F}(f)$ over too many points of the grid which again makes the approximation unpractical.

Instead of using the direct approximation of the Continuous Fourier Transform and its inverse, another approach is to propose a finite discretized version of the FCVM, which reflects the main characteristics of the FCVM. In order to achieve this, note two important things: i) the convolution of two functions can be approached by as the convolution of two vectors and ii) the convolution of two vectors is transformed into a multiplication with the Discrete Fourier Transform ((Kammler, 2008, p. 102), Oppenheim and Schafer (2011, p. 60)).

Here we use the definition of the Discrete Fourier Transform found in Kammler (2008, p. 291) or in Bloomfield (2004, p. 41), defined for vectors of $\mathbb{C}^q$ as follows

$$\begin{array}{rccl} \mathscr{F}_d\colon & \mathbb{C}^q & \to & \mathbb{C}^q \\ & f := (f_0, \cdots, f_{q-1}) & \mapsto & (\mathscr{F}_d(f)(0), \cdots, \mathscr{F}_d(f)(q-1)), \end{array}$$

where for every $l = 0, \cdots, q-1$,

$$\mathscr{F}_d(f)(l) := \frac{1}{q} \sum_{r=0}^{q-1} f_r \omega^{rl} \in \mathbb{C}. \tag{1.12}$$

with $\omega := e^{-2\pi i/q}$. If we define the matrix

$$\Omega_q := \begin{pmatrix} 1 & 1 & 1 & \cdots & 1 \\ 1 & (\omega^1)^1 & (\omega^1)^2 & \cdots & (\omega^1)^{(q-1)} \\ 1 & (\omega^2)^1 & (\omega^2)^2 & \cdots & (\omega^2)^{(q-1)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & (\omega^{(q-1)})^1 & (\omega^{(q-1)})^2 & \cdots & (\omega^{(q-1)})^{(q-1)} \end{pmatrix} \tag{1.13}$$

we can write

$$\mathscr{F}_d(f) = \frac{1}{q}\Omega_k \, f \ \in \mathbb{C}^q. \tag{1.14}$$

Furthermore from this definition we can deduce

$$\mathscr{F}_d^{-1} = \Omega_q^*, \tag{1.15}$$

where Ω_q^* is the conjugate transpose of Ω_q.

Remark: We can see that the definition of $\mathscr{F}_d$ depends on the number q, which is the length of the vector. In this way when we apply $\mathscr{F}_d$ to a vector of size p we need to redefine the matrix Ω_p by using $\omega := e^{-2\pi i/p}$.

Finite Discrete version of the FCVM Let us take T large enough such that $[0,T]$ contains $\overline{supp(X)+supp(\theta)}$. Thus the supports of θ, X and Y are also contained in $[0,T]$ (Brezis (2010, p. 106)). Let us define $q-1 = \max\{j \in \mathbb{N} \,|\, t_j \in [0,T]\}$. Now take the discretization of the each function X_i and Y_i of the sample $(X_i,Y_i)_{i=1,\cdots,n}$ over the grid $[t_0,\cdots,t_{q-1}]$, so all these functions will become vectors in $\mathbb{R}^q \subset \mathbb{C}^q$, that is $X_i^d, Y_i^d \in \mathbb{C}^q$ for every $i = 1,\cdots,n$.

Given that the matrix Ω_q has the property of transforming finite convolutions into multiplications, we can use the three steps method as the one used to define the estimator $\hat{\theta}_n$ for the continuous case, namely i) transform the problem with the matrix Ω_q from the time-domain to the frequency one, ii) use the ridge estimator in this domain, and iii) finally come back with the inverse of Ω_q.

The comparison between the continuous and the discrete cases is done next. Note that in the discrete case the multiplication and the division is done the element by element between vectors of same length. Furthermore, $*^d$ is discrete convolution, Δ is the step of discretization and we use $P_q : \mathbb{R}^{2q-1} \to \mathbb{R}^q$, the projection into the first q components, to have vectors of the same length.

CONTINUOUS	DISCRETE
Data and conditions: $\theta \in L^2([0,T])$. For $i = 1,\cdots,n$, $X_i, Y_i, \varepsilon_i \in L^2([0,T])$, $$Y_i = \theta * X_i + \varepsilon_i.$$	**Data and conditions:** $\theta^d \in \mathbb{R}^q$. For $i = 1,\cdots,n$, $X_i^d, Y_i^d, \varepsilon_i^d \in \mathbb{R}^q$, $$Y_i^d = \Delta P_q(\theta^d *^d X_i^d) + \varepsilon_i^d.$$
Estimation steps:	**Estimation steps:**
1. For $i = 1,\cdots,n$, $$\mathscr{F}(Y_i) = \mathscr{F}(\theta)\mathscr{F}(X_i) + \mathscr{F}(\varepsilon_i).$$	1. For $i = 1,\cdots,n$, $$\Omega_q(Y_i^d) = \Delta\Omega_q^d(\theta^d)\cdot\Omega_q(X_i^d) + \Omega_q(\varepsilon_i^d).$$
2. $$\mathscr{F}\hat{(\theta)}_n := \frac{\sum_{i=1}^n \mathscr{F}(Y_i)\overline{\mathscr{F}(X_i)}}{\sum_{i=1}^n \lvert\mathscr{F}(X_i)\rvert^2 + \lambda_n}$$	2. $$\Omega_q\hat{(\theta^d)}_n := \frac{1}{\Delta}\frac{\sum_{i=1}^n \Omega_q(Y_i^d)\overline{\Omega_q(X_i^d)}}{\sum_{i=1}^n \lvert\Omega_q(X_i^d)\rvert^2 + \vec{\lambda}_n},$$ where $\vec{\lambda}_n := (\lambda_n,\cdots,\lambda_n) \in \mathbb{R}^q$.
3. $$\hat{\theta}_n := \mathscr{F}^{-1}(\mathscr{F}\hat{(\theta)}_n)$$	3. $$\hat{\theta}_n^d := \Omega_q^{-1}(\Omega_q\hat{(\theta^d)}_n).$$

From this comparison we can define the numerical estimator of θ over the grid $[t_0,\cdots,t_{q-1}]$ as follows

$$\hat{\theta}_n^d := \frac{1}{\Delta}\Omega_q^{-1}\left[\frac{\sum_{i=1}^n \Omega_q Y_i^d \cdot \overline{\Omega_q X_i^d}}{\sum_{i=1}^n \lvert\Omega_q X_i^d\rvert^2 + \vec{\lambda}_n}\right]. \tag{1.16}$$

1.4.2 Compact Supports and Grid of Observations

From now on we will compute $\hat{\theta}_n$ numerically with equation (3.27). The important question we want to address here is how large the grid of observation points should be to properly estimate θ? In this regard understanding the relationship between the supports of X and θ and the one of their convolution (Y) is an essential element to answer this question. We know that (Brezis (2010, p. 106)),

$$supp(Y) = supp(\theta * X) \subset \overline{supp(X) + supp(\theta)}.$$

Then as mentioned before whenever our grid of observations contains the interval $[0,T]$ and $[0,T]$ contains $\overline{supp(X)+supp(\theta)}$ we will be able to estimate θ over its whole compact support.

The problem arises from the fact that we do not know θ and as a consequence neither $supp(\theta)$ nor $\overline{supp(X)+supp(\theta)}$. Then how big T should be in order to estimate θ correctly?

There are several cases to consider. First let us suppose that the grid of observations covers $[0,T_1]$ and $supp(X), supp(Y) \subset [0,T_1]$ then we can choose $T > T_1$ big enough and estimate θ over $[0,T]$. To see this more clearly let us say that the grid of observations over $[0,T_1]$ is $t_0,\cdots,t_{q_1}$ and over $[0,T]$ is $t_0,\cdots,t_q$, with $q > q_1$. Given that we have only observed the curves over $[0,T_1]$ we only know the vectors $(X_i^d, Y_i^d)_{i=1,\cdots,n} \subset \mathbb{R}^{q_1}$. Then the only thing we need to do before applying equation (3.27) properly is to redefine the vectors X_i^d and Y_i^d by adding zeros such that they will belong to $\mathbb{R}^q$, for instance

$$X_i^d := (X_i^d, 0, \cdots, 0) \in \mathbb{R}^q.$$

This procedure is known as **zero padding** the signal (Gonzalez and Eddins (2009, p. 111)). In this case equation (3.27) is well defined and we will compute θ over $[0,T]$. Note also that $supp(\theta)$ could be bigger than $[0,T]$ but the estimation of θ over $[0,T]$ is still correct.

Secondly we have the case where the grid of observations covers $[0,T_1]$ and we know $supp(X) \subset [0,T_1]$ and $supp(Y) \setminus [0,T_1] \neq \emptyset$. Under these hypotheses we cannot add more zeros to the vectors Y_i^d because if we did it would imply that Y has zero values outside $[0,T_1]$ which contradicts $supp(Y) \setminus [0,T_1] \neq \emptyset$. Thus we cannot apply the property of Ω_q to transform the convolution into a multiplication correctly. This is one restriction to the correct application of the FCVM.

Finally if the grid of observations covers $[0,T_1]$, $supp(X) \setminus [0,T_1] \neq \emptyset$ and $supp(Y) \setminus [0,T_1] \neq \emptyset$ we have the same phenomenon, that is we cannot add more zeros to the vectors X_i^d and Y_i^d to belong to $\mathbb{R}^q$. Thus it is not possible to transform the convolution into a multiplication because q_1 is not big enough. Note that Ω_{q_1} is quite different from Ω_q (see definition 3.24) and the property of transforming the convolution into a multiplication of two vectors only holds when Ω_q is applied to the entire convolution of both vectors, that is q is big enough to contain the convolution.

In any case in order to estimate θ with the functional Fourier deconvolution estimator, the grid of observations should cover $supp(X)$ and $supp(Y)$. This is an important restriction of this estimator.

FFT Algorithm and fast computing : One of the main advantages of the Functional Fourier Deconvolution estimator is that it is calculated very fast. This is due to the fact

that it uses the Fast Fourier Transform to compute the Discrete Fourier Transform. It is known that this algorithm computes the Discrete Fourier Transform of an n-dimensional signal in $O(n\log(n))$ time. The publication of the Cooley-Tukey Fast Fourier transform (FFT) algorithm in 1965 (Cooley and Tukey (1965)) revolutionized the area of digital signal processing because it reduced the order of complexity of the Fourier transform and of the convolution from n^2 to $n\log(n)$, where n is the problem size. Then over the last years new algorithms have improved the performance of the Cooley-Tukey algorithm under some conditions (split-radix FFT, Winograd FFT, etc). Among the recent improvements we highlight the Nearly Optimal Sparse Fourier Transform (Hassanieh et al. (2012)).

1.5 Contribution of this thesis

In this thesis, we want to know how the history of the functional regressor X influences the current value of the functional response Y in functional linear regression models.

This thesis is divided in 6 chapters. We present in Chapter 1 a general introduction of the theoretical background used in the following chapters. The theoretical and practical contributions of this thesis are from Chapter 2 to Chapter 4. In these chapters we studied the functional concurrent model (Chapter 2), the functional convolution model (Chapter 3) and the fully functional model (Chapter 4). An illustration on real datasets is given in Chapter 5. Finally we present in Chapter 6 the conclusions and perspectives of this thesis.

A more detailed review of the contributions is given below.

1.5.1 Chapter 2

In this chapter we propose a functional approach to estimate the unknown function in the Functional Concurrent Model (FCCM). This method is a generalization of the classic Ridge regression method to the functional data framework. For this reason we named this new estimator the *Functional Ridge Regression Estimator* (FRRE).

We proved the consistency of the FREE for the L^2-norm, and obtained a rate of convergence over the whole real line, and not only on compact sets. We also provided a selection procedure of the optimal regularization parameter λ_n through the Leave-One-Out Predictive Cross-Validation and the General Cross-Validation. The whole estimation procedure has been experienced on simulation trials, which showed good properties of the FRRE under very low Signal-to-Noise ratio. Thanks to its simpler definition, the FRRE is faster to compute than other estimators in the FCCM, such as the one proposed in Şentürk and Müller (2010).

Finally the definition of the FRRE is suitable to be used as a step of the estimation procedure in the Functional Convolution Model, which is the focus of the Chapter 3.

This chapter is an article we have submitted to the Electronic Journal of Statistics.

1.5.2 Chapter 3

In this chapter we propose the *Functional Fourier Deconvolution Estimator* (FFDE) of the functional coefficient in the Functional Convolution Model (FCVM). To do this we implemented a new approach which uses the duality between the time domain and frequency domain spaces through the continuous Fourier transform.

Thanks to this duality we associate the FCCM to the FCVM and we can use the Functional Ridge Regression Estimator in the frequency domain to define the FFDE. This fact allowed us to prove the consistency of the FFDE for the L^2-norm, and obtained a rate of convergence over the whole real line. We also provided a selection procedure of the optimal regularization parameter λ_n through the Leave-One-Out Predictive Cross-Validation.

We have defined other estimators for the FCVM, which we adapted from different methods found in the literature about the "deconvolution problem". Then we compared the performance of the FFDE with these alternative estimators. The simulations have shown the robustness, the accuracy and the fast computation time of the FFDE compared to the others. The reason why the FFDE is calculated very fast is that we use the Discrete Fourier Transform for its numerical implementation. This is a very useful property of the FFDE.

This chapter is an article will be submitted to the Electronic Journal of Statistics.

1.5.3 Chapter 4

In this chapter we have proposed two estimators of the covariance operator of the noise (Γ_ε) in functional linear regression when both the response and the covariate are functional (see the fully functional model (1.2)). We studied the asymptotic properties of these estimators and their behavior on simulations.

More particularly we have estimated the trace of the covariance operator of the noise ($\sigma_\varepsilon^2 = tr(\Gamma_\varepsilon)$). The estimation of σ_ε^2 would make possible the construction of hypothesis testing in connection with fully functional model. Furthermore σ_ε^2 is involved in the square prediction error bound that participates to determine the convergence rate (Crambes and Mas (2013)). Thus an estimator of σ_ε^2 will provide details on the prediction quality in the fully functional model.

This chapter is an article published in Statistics and Probability Letters (Volume 113, June 2016, Pages 7–15)

1.5.4 Chapter 5

This chapter is an illustration of the implementation of the results presented in Chapter 3. We have used the FCVM (1.3) and the historical functional linear model (1.1) to study how the Vapour Pressure Deficit (VPD) influences Leaf Elongation Rate (LER) curves obtained on high-throughput plant phenotyping platforms, from experiments carried out in 2014, named here as **T72A** and **T73A**.

In both experiments the FCVM is too simple to bring light about the interaction of the VPD and LER. On contrast the historical functional model is more helpful to understand this interaction because it is more complex.

To estimate the historical kernel $\mathscr{K}_{hist}$ we have proposed two estimators: first the Karhunen-Loève estimator satisfying the historical restriction and secondly a Tikhonov-type functional estimator. Among these two estimators the latter shows more consistent results across both experiments.

Chapter 2

Ridge Regression for the Functional Concurrent Model

Contents

Abstract: The aim of this paper is to propose an estimator of the unknown function in the Functional Concurrent Model (FCM). This is a general model to which all functional linear models can be reduced. We follow a strictly functional approach and extend the Ridge Regression method developed in the classical linear case to the functional data framework. We prove the probability convergence of this estimator and obtain a rate of convergence. Then we study the problem of the optimal selection of the regularization parameter λ_n. Afterwards we present some simulations that show the accuracy of this estimator in fitting the unknown function, despite a very low signal-to-noise ratio.
Keywords and phrases: Ridge regression, functional data, concurrent model, varying coefficient model.

2.1 Introduction

Functional Data Analysis (FDA) proposes very good tools to handle data that are functions of some covariate (e.g. time, when dealing with longitudinal data), see Hsing and Eubank (2015) or Horváth and Kokoszka (2012). These tools allow for better modelling of complex relationships than classical multivariate data analysis, as noticed by Ramsay and Silverman (2005, Ch. 1), Yao et al. (2005a,b), among others.

There are several models in FDA for studying the relationship between two variables. In particular in this paper we are interested in the Functional Concurrent Model (FCM) which is defined as follows

$$Y(t) = \beta(t)X(t) + \varepsilon(t), \tag{2.1}$$

where $t \in \mathbb{R}$, β is the unknown function to be estimated, X, Y are random functions and ε is a noise random function. As stated by Ramsay and Silverman (2005, p. 220), all functional linear models can be reduced to this form. This model is also related to the functional varying coefficient model (VCM) and has been studied for example by Wu et al. (1998) or more recently by Şentürk and Müller (2010).

Another practical advantage of model (2.1) is that it allows to simplify the study of the following convolution model

$$W(s) = \int_{-\infty}^{+\infty} \theta(u)Z(s-u)du + \eta(s) \tag{2.2}$$

through the Fourier transform $\mathscr{F}$ with $Y = \mathscr{F}(W)$, $\beta = \mathscr{F}(\theta)$, $X = \mathscr{F}(Z)$ and $\varepsilon = \mathscr{F}(\eta)$.

As far as we know, despite the abundant literature related to FCM or functional VCM, there is hardly any paper providing a strictly functional approach (i.e. with random functions defined inside normed functional spaces and studying the convergence with their own norms). As noticed by Ramsay and Silverman (2005, p. 259), all these methods come from a multivariate data analysis approach rather than from a functional one. For some applications, for example when the observations are highly auto-correlated, taking this functional nature into account may be decisive. If not, multivariate approaches may cause a loss of information because, as noticed by Şentürk and Müller (2010, p. 1257),

they “do not take full advantage of the functional nature of the underlying data". In practice this loss of information may reduce the accuracy of estimation and prediction. To circumvent this problem, Şentürk and Müller (2010) propose a three-step functional approach based on smoothing and least square estimation. However, the convergence results obtained on compact sets do not allow to study specific models like (2.2), for which convergence on the whole real line is required.

The objective of the present paper is to propose an estimator of the function β in the FCM (2.1) for which such asymptotic results hold. Our estimation approach is based on the Ridge Regression method developed in the classical linear case, see Hoerl (1962). We extended it to the functional data framework of model (2.1). First we establish the consistency of the estimator and get a rate of convergence. Then we propose a method for selecting the regularization parameter. Finally we present some simulation trials which show the accuracy of the estimator in fitting the unknown function β, despite a very low signal-to-noise ratio (SNR).

All the proofs are postponed to Section 2.7.

2.2 Model and Estimator

To remain as general as possible, all considered functions are complex valued functions. Before studying the FCM let us define some useful notations.

We define $L^2(\mathbb{R},\mathbb{C}) = L^2$ the set of square integrable complex valued functions, with the L^2-norm $\|f\|_{L^2} := \left[\int_{\mathbb{R}} |f(x)|^2 dx\right]^{1/2}$, and given a subset $K \subset \mathbb{R}$, $\|f\|_{L^2(K)} := \left[\int_K |f(x)|^2 dx\right]^{1/2}$, where $|\cdot|$ denotes the complex modulus.

The theoretical results given in the next sections are proved on the whole real line. For this reason, we need to restrict the study to the set of functions that vanish at infinity. Let $C_0(\mathbb{R},\mathbb{C}) = C_0$ be the space of complex valued continuous functions, which satisfies : for all $\zeta > 0$ there exists a $R > 0$ such that for all $|t| > R$, $|f(t)| < \zeta$. We use the supremum norm $\|f\|_{C_0} := \sup_{x\in\mathbb{R}} |f(x)|$. In particular for a subset $K \subset \mathbb{R}$, $\|f\|_{C_0(K)} := \sup_{x\in K} |f(x)|$.

Finally throughout this paper the support of a continuous function $f : \mathbb{R} \to \mathbb{C}$ is the set $supp(f) := \{t \in \mathbb{R} : |f(t)| \neq 0\}$. This set is open because f is continuous. Besides we define the boundary of a set S, as $\partial(S) := \overline{S} \setminus int(S)$, where $\overline{S}$ is the closure of S and $int(S)$ is its interior.

2.2.1 General Hypotheses of the FCM

The space C_0 is too large. For instance, its geometry does not allow for the application of the Central Limit Theorem (CLT) under the general hypothesis of the existence of the covariance operator, that is $\mathbb{E}(\|X\|^2_{C_0}) < \infty$ (see Ledoux and Talagrand (1991, Ch 10)). To circumvent this difficulty, we consider functions that belong to the space $C_0 \cap L^2$. Here are general hypotheses that will be used all along the paper:

($HA1_{FCM}$) X, ε are independent $C_0 \cap L^2$ valued random functions, such that $\mathbb{E}(\varepsilon) = \mathbb{E}(X) = 0$,

($HA2_{FCM}$) $\beta \in C_0 \cap L^2$,

($HA3_{FCM}$) $\mathbb{E}(\|\varepsilon\|^2_{C_0}), \mathbb{E}(\|X\|^2_{C_0}), \mathbb{E}(\|\varepsilon\|^2_{L^2})$ and $\mathbb{E}(\|X\|^2_{L^2})$ are all finite.

2.2.2 Functional Ridge Regression Estimator (FRRE)

The definition of the estimator of β is inspired by the estimator introduced by Hoerl (1962) used in the Ridge Regularization method that deal with ill-posed problems in the classical linear regression.

Let $(X_i, Y_i)_{i=1,\cdots,n}$ be an i.i.d sample of FCM (2.1) and a regularization parameter $\lambda_n > 0$. We define the estimator of β as follows

$$\hat{\beta}_n := \frac{\frac{1}{n}\sum_{i=1}^n Y_i X_i^*}{\frac{1}{n}\sum_{i=1}^n |X_i|^2 + \frac{\lambda_n}{n}}, \tag{2.3}$$

where the exponent $*$ stands for the complex conjugate. In the classical linear regression case, Hoerl and Kennard (1970, p. 62) proved that there is always a regularization parameter for which the ridge estimator is better than the Ordinary Linear Squares (OLS) estimator. Huh and Olkin (1995) made a study of some asymptotic properties of the ridge estimator in this context.

2.3 Asymptotic Properties of the FRRE

From the definition (2.3), it is easy to show that the FRRE $\hat{\beta}_n$ decomposes as follows:

$$\hat{\beta}_n = \beta - \frac{\lambda_n}{n}\left[\frac{\beta}{\frac{1}{n}\sum_{i=1}^n |X_i|^2 + \frac{\lambda_n}{n}}\right] + \frac{\frac{1}{n}\sum_{i=1}^n \varepsilon_i X_i^*}{\frac{1}{n}\sum_{i=1}^n |X_i|^2 + \frac{\lambda_n}{n}}. \tag{2.4}$$

The main results of this paper are the probability convergence of the FRRE and the rate of convergence

$$\|\hat{\beta}_n - \beta\|_{L^2} = O_P\left(\max\left[\frac{\lambda_n}{n}, \frac{\sqrt{n}}{\lambda_n}\right]\right),$$

under very large conditions.

2.3.1 Consistency of the Estimator

Theorem 3. *Let us consider the FCM with the general hypotheses ($HA1_{FCM}$), ($HA2_{FCM}$) and ($HA3_{FCM}$). Let $(X_i, Y_i)_{i\geq 1}$ be i.i.d. realizations. We suppose moreover that*

(A1) $\overline{supp(|\beta|)} \subseteq \overline{supp(\mathbb{E}[|X|])}$,

(A2) $(\lambda_n)_{n\geq 1} \subset \mathbb{R}^+$ *is such that* $\frac{\lambda_n}{n} \to 0$ *and* $\frac{\sqrt{n}}{\lambda_n} \to 0$ *as* $n \to +\infty$.

Then

$$\lim_{n\to+\infty} \|\hat{\beta}_n - \beta\|_{L^2} = 0 \quad \textit{in probability.} \tag{2.5}$$

Let us make some comments about the hypotheses.

Remark. *Hypothesis (A2) is classic in the context of ridge regression. Hypothesis (A1) specifies that it is not possible to estimate β outside the support of the modulus of X. From model (2.1), it is clear that β cannot be estimated in the intervals where the function X is zero, as proved in the following proposition:*

Proposition 4. *Let $(X_i, Y_i)_{i=1,\cdots,n}$ be an i.i.d. sample of FCM in $C_0 \cap L^2$ which satisfies hypothesis $(A2)$ and*

(nA1) There exists $t_0 \in \overline{supp(|\beta|)}$ and $\delta > 0$ such that $\mathbb{E}[\|X\|_{C_0([t_0-\delta,t_0+\delta])}] = 0$.

Then there exists a constant $C > 0$ such that almost surely

$$\|\hat{\beta}_n - \beta\|_{L^2} \geq C. \tag{2.6}$$

Proof. For all the independent realizations of X, we have $\mathbb{E}[\|X_n\|_{C_0([t_0-\delta,t_0+\delta])}] = 0$. Then for all $n \in \mathbb{N}$, the function X_n restricted to the interval $[t_0-\delta, t_0+\delta]$ is equal to zero almost surely. Thus over this interval $\hat{\beta}_n = 0$ (a.s.). If we define $C := \|\beta\|_{L^2([t_0-\delta,t_0+\delta])}$ we obtain

$$\|\hat{\beta}_n - \beta\|_{L^2} \geq \|\hat{\beta}_n - \beta\|_{L^2([t_0-\delta,t_0+\delta])} = C \quad (a.s.).$$

□

Hypothesis $(nA1)$ is stronger than the negation of $(A1)$. It provides that there exists some t_0 in $\overline{supp(|\beta|)}$, such that X is zero almost surely in a neighborhood of t_0.

The geometry of L^2 helps a lot in the proof of Theorem 3. By paying attention to the geometry of L^p spaces, it is also possible to generalize this result for those spaces.

2.3.2 Rate of Convergence

To obtain a rate of convergence, we need to control the shapes of the functions β and $\mathbb{E}[|X|]$ on the borders of the support of $\mathbb{E}[|X|]$. Theorem 5 handles the general case where $|\beta|/\mathbb{E}[|X|^2]$ goes to infinity over the points of the set $C_{\beta,\partial X} := supp(|\beta|) \cap \partial(supp(\mathbb{E}[|X|]))$.

Theorem 5. *Let us consider the FCM with the general hypotheses ($HA1_{FCM}$), ($HA2_{FCM}$) and ($HA3_{FCM}$). We assume additionally that $(A1)$ holds, together with :*

(A3) $\mathbb{E}[\||X|^2\|_{L^2}^2] < \infty$.

(A4) $\left\| \frac{|\beta|}{\mathbb{E}[|X|^2]} 1_{\overline{supp(\beta)} \setminus \partial(supp(\mathbb{E}[|X|]))} \right\|_{L^2} < +\infty$.

(A5) There exist positive real numbers $\alpha > 0$, $M_0, M_1, M_2 > 0$ *such that*

(a) For every $p \in C_{\beta,\partial X}$, *there exists an open neighborhood* $J_p \subset supp(|\beta|)$ *such that*

$$\mathbb{E}[|X|^2(t)] \geq |t-p|^\alpha,$$

for every $t \in J_p$ *and*

$$\left\| \frac{1}{\mathbb{E}[|X|^2]} \right\|_{L^2(J_p \setminus \{p\})} \leq M_0,$$

(b) $\sum_{p \in C_{\beta,\partial X}} \|\beta\|^2_{C_0(J_p)} < M_1$,

(c) $\frac{|\beta|}{\mathbb{E}[|X|^2]} 1_{supp(|\beta|) \setminus J} < M_2$, *where* $J = \bigcup_{p \in C_{\beta,\partial X}} J_p$.

(A6) For $n \geq 1$,

$$\lambda_n := n^{1 - \frac{1}{4\alpha+2}},$$

where $\alpha > 0$ *comes from the hypothesis* $(A5)$.

Then

$$\|\hat{\beta}_n - \beta\|_{L^2} = O_P\left(n^{-\gamma}\right), \tag{2.7}$$

where $\gamma := \min\left[\frac{1}{2(2\alpha+1)}, \frac{1}{2} - \frac{1}{2(2\alpha+1)}\right]$.

The following corollary specifies the rate of convergence for $\alpha = 1/2$.

Corollary 6. *Under the hypotheses of Theorem 5,* $n^{-\gamma} = \max\left[\frac{\lambda_n}{n}, \frac{\sqrt{n}}{\lambda_n}\right]$ *and in particular if* $\alpha = 1/2$

$$\|\hat{\beta}_n - \beta\|_{L^2} = O_P\left(\frac{1}{n^{1/4}}\right).$$

Remark. *Hypothesis* $(A3)$ *is classic and allows to apply the CLT on the denominator of* $\hat{\beta}_n$. *Hypothesis* $(A4)$ *is needed because the second term in (2.4), namely* $\left[\frac{\beta}{\frac{1}{n}\sum_{i=1}^n |X_i|^2 + \frac{\lambda_n}{n}}\right]$, *can naturally be* L^2*-bounded under this condition.*

Next $(A5a)$ *requires that around the points* $p \in supp(\beta) \cap \partial(supp(\mathbb{E}[|X|]))$ *the function* $\mathbb{E}[|X|^2]$ *goes to zero slower than a polynomial of degree* α, *which implies that the second term in (2.4) behaves like* $\frac{\beta}{\mathbb{E}[|X|^2]}$ *and determines the rate of convergence.*

Parts (b) *and* (c) *of* $(A5)$ *help us controlling the tails of* β *and* $|X|$ *around infinity. They are useful only when* $card(C_{\beta,\partial X}) = +\infty$. *Note that the set* $C_{\beta,\partial X}$ *is always countable (see the proof of Theorem 5).*

Finally hypothesis $(A6)$ *replaces* $(A2)$ *in Theorem 3, as the rate of convergence strongly depends on the behavior of* $\frac{\beta}{\mathbb{E}[|X|^2]}$ *around the points of* $C_{\beta,\partial X}$*, which depends on* α*. We can see that* $(A6)$ *always implies* $(A2)$*.*

It is possible to get the same convergence results as that of Theorem 5 under assumptions easier to verify, in particular when $C_{\beta,\partial X} = \emptyset$, which is a stronger assumption than hypothesis $(A4bis)$ in Corollary 7.

Corollary 7. *Under hypotheses* $(A1)$*,* $(A2)$ *and* $(A3)$ *and if additionally we assume*

(A4bis) $\frac{|\beta|}{\mathbb{E}[|X|^2]} 1_{supp(|\beta|)} \in L^2 \cap L^\infty$,

then

$$\|\hat{\beta}_n - \beta\|_{L^2} = O_P\left(\max\left[\frac{\lambda_n}{n}, \frac{\sqrt{n}}{\lambda_n}\right]\right). \tag{2.8}$$

Hypothesis $(A4bis)$ is a reformulation of $(A4)$ and part (c) of $(A5)$. It is required to control the second term of (2.4) and the decreasing rate of β with respect to $\mathbb{E}[|X|^2]$ around infinity (tails control).

Finally, Theorem 8 deals with the convergence rate on compact subsets of the support of $\mathbb{E}[|X|^2]$.

Theorem 8. *Under hypotheses* $(A1)$*,* $(A2)$ *and* $(A3)$*, for every compact subset* $K \subset supp(\mathbb{E}[|X|^2])$*, we have*

$$\|\hat{\beta}_n - \beta\|_{L^2(K)} = O_P\left(\max\left[\frac{\lambda_n}{n}, \frac{\sqrt{n}}{\lambda_n}\right]\right). \tag{2.9}$$

Moreover if the support of β is compact, we deduce the following corollary.

Corollary 9. *Under the hypotheses* $(A1)$*,* $(A2)$ *and* $(A3)$*, if* $\overline{supp(\beta)}$ *is compact and is a subset of* $supp(\mathbb{E}[|X|])$*, then*

$$\|\hat{\beta}_n - \beta\|_{L^2} = O_P\left(\max\left[\frac{\lambda_n}{n}, \frac{\sqrt{n}}{\lambda_n}\right]\right).$$

2.4 Selection of the Regularization Parameter

2.4.1 Predictive Cross-Validation (PCV) and Generalized Cross-Validation (GCV)

This section is devoted to developing a selection procedure of the regularization parameter λ_n for a given sample $(X_i, Y_i)_{i\in\{1,\cdots,n\}}$. To solve this problem we chose the Predictive Cross-Validation (PCV) criterion. Its definition, see for instance Febrero-Bande and Oviedo de la Fuente (2012, p. 17) or Hall and Hosseini-Nasab (2006, p. 117), is the following

$$PCV(\lambda_n) := \frac{1}{n}\sum_{i=1}^{n} \|Y_i - \hat{\beta}_n^{(-i)} X_i\|_{L^2}^2,$$

where $\hat{\beta}_n^{(-i)}$ is computed with the sample $(X_j, Y_j)_{j\in\{1,\cdots,i-1,i+1,\cdots,n\}}$. The selection method consists in choosing the value λ_n which minimizes the function $PCV(\cdot)$.

In this subsection we give results that allow for computing faster the PCV by processing only one regression, instead of n. These results use similar ideas as in Green and Silverman (1994, pp. 31-33) about the smoothing parameter selection for smoothing splines.

Proposition 10. *We have*

$$PCV(\lambda_n) = \frac{1}{n}\sum_{i=1}^{n}\left\|\frac{Y_i - \hat{\beta}_n X_i}{1 - A_{i,i}}\right\|_{L^2}^2, \tag{2.10}$$

where $A_{i,i} \in L^2$ is defined as follows $A_{i,i} := |X_i|^2/(\sum_{j=1}^n |X_j|^2 + \lambda_n)$.

This last proposition allows to write the PCV without excluding the ith observation. We then introduce the following Generalized Cross-Validation (GCV), computationally faster than the PCV:

$$GCV(\lambda_n) := \frac{1}{n}\sum_{i=1}^{n}\left\|\frac{Y_i - \hat{\beta}_n X_i}{1 - A}\right\|_{L^2}^2,$$

where $A \in L^2$ is $A := (\frac{1}{n}\sum_{i=1}^n |X_i|^2)/(\sum_{j=1}^n |X_j|^2 + \lambda_n)$.

Remark: From the definition of A, we have that, for every $t \in \mathbb{R}$, $0 \le A(t) \le 1/n$, then $1 \le \frac{1}{1-A(t)} \le \frac{n}{n-1}$, which yields that the GCV criterion is bounded as follows:

$$\frac{1}{n}\sum_{i=1}^{n}\left\|Y_i - \hat{\beta}_n X_i\right\|_{L^2}^2 \le GCV(\lambda_n) \le \frac{1}{n-1}\sum_{i=1}^{n}\left\|Y_i - \hat{\beta}_n X_i\right\|_{L^2}^2.$$

This last inequality gives thus quickly an idea of the GCV values.

2.4.2 Regularization function Parameter

As we are working with functional data, another possibility is to use a time-dependent function $\Lambda_n(t)$ in the estimator defined in (2.3), instead of a constant number λ_n. We shall optimize, for each time t, the choice of $\Lambda_n(t)$. To that aim, we have to compute the PCV for each time $t \in \mathbb{R}$,

$$PCV(\Lambda_n(t)) := \frac{1}{n}\sum_{i=1}^{n}|Y_i(t) - \hat{\beta}_n^{(-i)}(t)\,X_i(t)|^2,$$

where $\hat{\beta}_n^{(-i)}(t)$ is computed with the sample $(X_j(t), Y_j(t))_{j\in\{1,\cdots,n\}\setminus\{i\}}$.

As above, we obtain a simpler formula for $PCV(\Lambda_n(t))$ (see next proposition bellow), which yields a faster computation.

Proposition 11. *We have*

$$PCV(\Lambda_n(t)) = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{Y_i(t)-\hat{\beta}_n(t)X_i(t)}{1-A_{i,i}(t)}\right|^2, \tag{2.11}$$

where $A_{i,i}(t) := \frac{|X_i(t)|^2}{\sum_{j=1}^{n}|X_j(t)|^2+\lambda_n(t)}$.

This criterion is discussed in the next section about simulation studies. Its performance is evaluated and compared to that of $GCV(\lambda_n)$.

2.5 Simulation study

The simulation study follows model (2.1) with an intercept term:

$$Y_i(t) = \beta_0(t) + \beta_1(t)X_i(t) + \varepsilon_i(t),\ \forall i = 1,\ldots,n,\ \forall t \in [0,T]. \tag{2.12}$$

We evaluate our estimation procedure in the case of a low Signal-to-Noise-Ratio (SNR). Both approaches using λ_n and $\Lambda_n(t)$ are compared.

We first give the estimation procedure adapted to model (2.12) and we introduce three criteria to measure the estimation error when estimating β_0 and β_1.

2.5.1 Estimation procedure and evaluation criteria

We compute the mean curve $\bar{X} := \frac{1}{n}\sum_{i=1}^{n}X_i$ and $\bar{Y} := \frac{1}{n}\sum_{i=1}^{n}Y_i$. Thus we use the FRRE to compute the estimators $\hat{\beta}_1$ and $\hat{\beta}_0$, as follows

$$\begin{aligned}\hat{\beta}_1 &:= \frac{\sum_{i=1}^{n}(Y_i-\bar{Y})(X_i-\bar{X})^*}{\sum_{i=1}^{n}|(X_i-\bar{X})|^2 + \lambda_n}, \\ \hat{\beta}_0 &:= \bar{Y}-\hat{\beta}_1\bar{X}.\end{aligned} \tag{2.13}$$

We use 500 Monte Carlo runs to evaluate the mean absolute deviation error (MADE), the weighted average squared error (WASE) and the unweighted average squared error (UASE), defined in the same way as in Şentürk and Müller (2010, p. 1261),

$$\begin{aligned}MADE &:= \frac{1}{2T}\left[\frac{\int_0^T|\beta_0(t)-\hat{\beta}_0(t)|dt}{range(\beta_0)} + \frac{\int_0^T|\beta_1(t)-\hat{\beta}_1(t)|dt}{range(\beta_1)}\right], \\ WASE &:= \frac{1}{2T}\left[\frac{\int_0^T|\beta_0(t)-\hat{\beta}_0(t)|^2dt}{range^2(\beta_0)} + \frac{\int_0^T|\beta_1(t)-\hat{\beta}_1(t)|^2dt}{range^2(\beta_1)}\right], \\ UASE &:= \frac{1}{2T}\left[\int_0^T|\beta_0(t)-\hat{\beta}_0(t)|^2dt + \int_0^T|\beta_1(t)-\hat{\beta}_1(t)|^2dt\right],\end{aligned}$$

where $range(\beta_r)$ is the range of the function β_r.

2.5.2 Setting

The data were simulated on the interval $[0,T]$ ($T=24$), discretized over $p=100$ equispaced observation times. More precisely for $j \in [1,100] \cap \mathbb{N}$, $t_j := j*24/101$. The simulated input curves X_i, for $i=1,\ldots,n$ with $n=150$, were generated with mean function $\mu_X(t) = t + \sin(t)$ and covariance function constructed from the 10 first eigenfunctions of the Wiener Process with its correspondent eigenvalues. Accordingly, for $0 \leq t \leq 24$, we have $X_i(t) = \mu_X(t) + \sum_{j=1}^{10} \rho_j \xi_{ij} \phi_j(t)$, where for $j \geq 1$, $\phi_j(t) = \sqrt{2}\sin((j-1/2)\pi t)$, $\rho_j = 1/((j-1/2)\pi)$ and the ξ_{ij} were generated from $N(0,1)$.

The functions β_0 and β_1 are respectively:

$$\beta_0(t) = \frac{2}{18^2}(t-6)^2 + 1 \qquad \text{and} \qquad \beta_1(t) = \begin{cases} \frac{-2}{16}(t-6)^2 + 2 & \text{if } t \in [2,10], \\ \frac{-2}{16}(t-18)^2 + 2 & \text{if } t \in [14,22], \\ 0 & \text{otherwise.} \end{cases}$$

The noise ε_i was defined as follows: $\varepsilon_i(t) = c_\varepsilon \sum_{j=11}^{20} \rho_j \xi_{ij} \phi_j(t)$, where c_ε is a constant such that the signal-to-noise ratio (SNR) is equal to 2, where $SNR := (tr(Cov(X)))/(tr(Cov(\varepsilon))$, with $Cov(X) := \mathbb{E}(< X, . > X - < \mathbb{E}(X), . > \mathbb{E}(X))$, $Cov(\varepsilon) := \mathbb{E}(< \varepsilon, . > \varepsilon)$ and tr is the trace of an operator.

The general hypotheses ($HA1_{FCM}$) - ($HA3_{FCM}$) are satisfied. The regularization parameter λ_n is optimized over the interval $[0,10]$.

2.5.3 Results

The simulation results are presented in Figures 2.1 and 2.2, and Table 2.1. The performance of the estimators with both regularization parameter λ and regularization curve Λ are illustrated.

Table 2.1 Means and standard deviations of the evaluation criteria MADE, WASE and UASE in the cases of optimal regularization parameter and curve.

λ	Constant λ_{150}		Curve Λ_{150}	
stats	mean	sd	mean	sd
MADE	0.05421	0.01551	0.04719	0.01383
WASE	0.01907	0.01788	0.01366	0.01454
UASE	0.07235	0.06786	0.05185	0.05517

We can see that, even in conditions of high noise ($SNR=2$), the estimation is really good. This shows the robustness of the FRRE. Moreover, the FRRE $\hat{\beta}_0^{(2)}$ and $\hat{\beta}_1^{(2)}$ computed with an optimal regularization curve Λ_{150} give in average better estimations of the functions β_0 and β_1. In this simulation setting, the mean of the optimal regularization curve Λ_n is almost constant (equal to 0.015 where $\beta_1 \neq 0$) with constant value close to the mean optimal regularization parameter λ_n (0.0156).

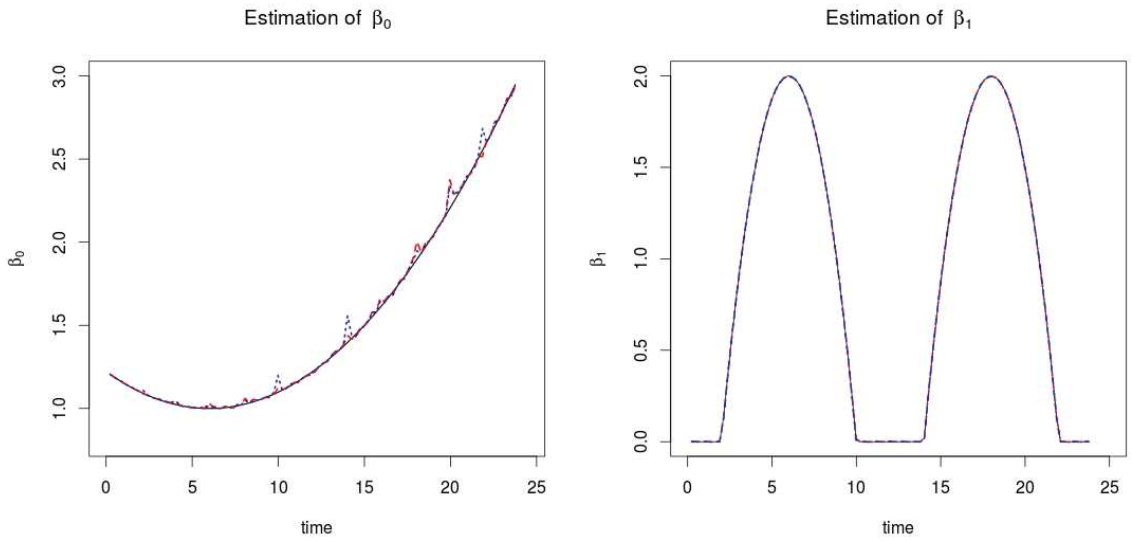

Fig. 2.1 The true functions β_0 and β_1 (solid) compared to the cross-sectional mean curves of the FRRE $\hat{\beta}_0^{(1)}$ and $\hat{\beta}_1^{(1)}$ (red dashed) computed with the optimal regularization parameter λ_{150}, and to the cross-sectional mean curves of the FRRE $\hat{\beta}_0^{(2)}$ and $\hat{\beta}_1^{(2)}$ (blue dotted) computed with an optimal regularization curve Λ_{150}.

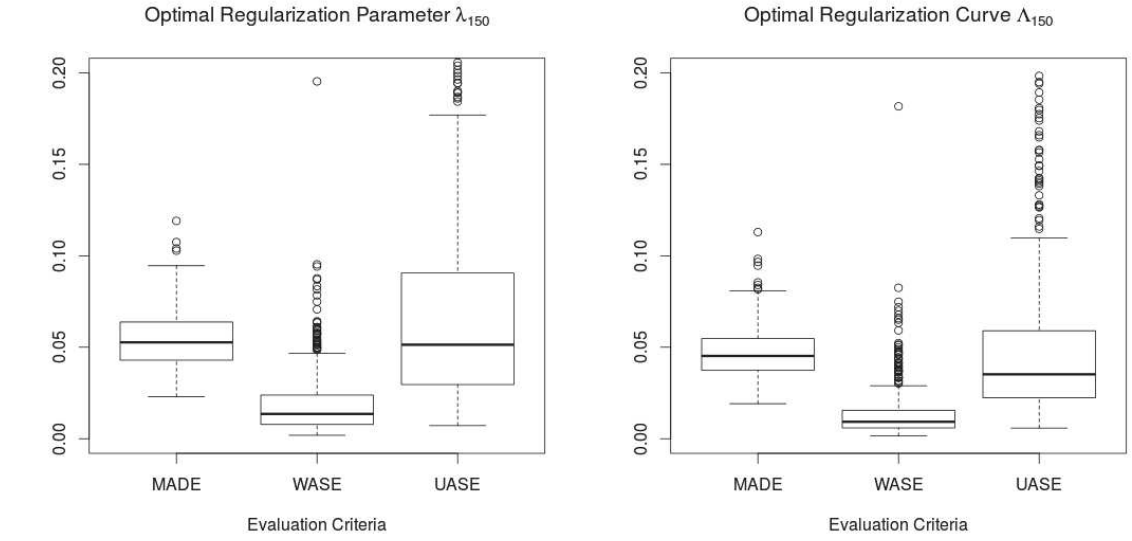

Fig. 2.2 Distribution of the evaluation criteria MADE, WASE and UASE in the cases of an optimal regularization parameter (left panel) and of an optimal regularization curve (right panel).

2.6 Conclusions

In this paper we have generalized the Ridge Regression method to estimate the unknown function of the FCM with the FRRE. We proved its consistency for the L^2-norm, and obtained its rate of convergence over the whole real line, and not only on compact sets. This strong result opens new perspectives for studying models related to the FCM, like the convolution model (2.2).

We also provided a selection procedure of the optimal regularization parameter λ_n through PCV. The simulations showed good properties of the FRRE under very low SNR.

This work shows some similarities with Şentürk and Müller (2010), where the model studied is close to the FCM (2.1) with i) an intercept term, ii) data X and Y observed up to additive noise and iii) a sparse random design. The estimation of the unknown function is done after three preliminary steps, first a smoothing step, then the computation of the raw covariances, and finally the smoothing of them. This three-step procedure requires the choice of several smoothing parameters, while the FRRE only requires the choice of one. In addition, the computation of the raw covariances is done over a square domain, whereas the FRRE directly calculates them over its diagonal. For these reasons, the FRRE is simpler to compute.

2.7 Main Proofs

Before proving Theorem 3, let us first introduce a useful technical lemma. Here we will denote $\varphi := \mathbb{E}[|X|^2] \in C_0$.

Lemma 12. *Under hypotheses* $(A1)$ *and* $(A2)$ *of Theorem 3, if there exists a sequence of functions* $(f_n)_{n\geq 1} \subset C_0$ *such that* $\|f_n - \varphi\|_{C_0} \to 0$*, then there exists*

1. *a sequence* $(C_j)_{j\geq 1}$ *of subsets of* $\mathbb{R}$ *such that*

$$m\left(\limsup_{j\to+\infty} C_j\right) = m\left(\cap_{J\geq 1}[\cup_{j=1}^{J} C_j]\right) = 0,$$

   where m is the Lebesgue measure,

2. *a strictly increasing sequence of natural numbers* $(N_j)_{j\geq 1} \subset \mathbb{N}$ *and a sequence of real numbers* $(d_n)_{n\geq 1} \subset \mathbb{R}$*, with* $\lim_{n\to+\infty} d_n = 0$*,*

such that for every $j \geq 1$ *and* $n \in \{N_j, \cdots, N_{j+1}\}$*,*

$$\left|\frac{\lambda_n}{n}\right| \left\| \frac{\beta}{f_n + \frac{\lambda_n}{n}} \right\|_{C_0(\mathbb{R}\setminus C_j)} \leq d_n. \tag{2.14}$$

Proof of Lemma 12. To start the proof, we notice that $supp(\varphi) = supp(\mathbb{E}[|X|])$, hence $\overline{supp(|\beta|)} \subseteq \overline{supp(\varphi)}$ by hypothesis $(A1)$.

We define the sequence $\alpha_r := \sqrt{\frac{\lambda_r}{r}}$ which is decreasing to 0, and the sets $K_r^{\varphi} := \varphi^{-1}([\alpha_r, +\infty[)$ and $K_q^{\beta} := |\beta|^{-1}([1/q, +\infty[)$ for $r, q \in \mathbb{N}^+$. All these sets are compacts and cover the supports of φ and β respectively, that is $\cup_{r=1}^{\infty} \uparrow K_r^{\varphi} = supp(\varphi)$ and $\cup_{q=1}^{\infty} \uparrow K_q^{\beta} = supp(\beta)$.

Without loss of generality, we can suppose that there exists some $Q_1 \in \mathbb{N}$ such that $K_{Q_1}^{\beta} \neq \emptyset$ (otherwise $\beta \equiv 0$). Then we redefine for all $q \in \mathbb{N}$, $K_q^{\beta} := K_{Q_1+q}^{\beta}$.

Let us take a sequence δ_s decreasing to 0 and define for all $s \in \mathbb{N}$,

$$D_s := B_{\delta_s}(\partial supp(\varphi)) = \cup_{a \in \partial supp(\varphi)} B_{\delta_s}(a) \quad \text{and} \quad C_s := K_s^{\beta} \cap D_s,$$

with $B_{\delta_s}(a) :=]a - \delta_s, a + \delta_s[$. Clearly

$$K_1^{\beta} \setminus C_1 \subset int(supp(\varphi)) = supp(\varphi) = \cup_{r=1}^{\infty} K_r^{\varphi}.$$

since the supports of continuous functions are open.

Thus, from the definition of K_r^{φ} and the fact that α_r goes to zero, there exists $r_1 \in \mathbb{N}$ such that for all $r \geq r_1$, $K_1^{\beta} \setminus C_1 \subset K_r^{\varphi}$.

Moreover, from $(A2)$ there exists $\tilde{r}_1 > r_1$ such that, for all $r \geq \tilde{r}_1$,

$$\max_{r \geq \tilde{r}_1} \frac{\lambda_r}{r} \leq \frac{\lambda_{r_1}}{r_1}.$$

Considering $K_1^{\beta} \setminus C_1$, from the definition of $K_{r_1}^{\varphi}$ and the uniform convergence of $(f_n)_{n \geq 1}$ towards φ, we deduce that there exists $N_1 > \tilde{r}_1$ such that for all $n \geq N_1$ and $t \in K_{r_1}^{\varphi}$,

$$\frac{3}{4}\, \alpha_{r_1} \leq f_n(t) + \frac{\lambda_n}{n}.$$

Thus for all n such that $n \geq N_1$,

$$|\frac{\lambda_n}{n}| \left\| \frac{\beta}{f_n + \frac{\lambda_n}{n}} \right\|_{C_0(K_{r_1}^{\varphi})} \leq |\frac{\lambda_n}{n}| \frac{4}{3\alpha_{r_1}} \|\beta\|_{C_0(\mathbb{R})} \leq \left(\max_{s \geq \tilde{r}_1} [|\frac{\lambda_s}{s}|] \right) \frac{4}{3\alpha_{r_1}} \|\beta\|_{C_0(\mathbb{R})}.$$

In particular we can deduce, for all $n \geq N_1 > r_1$,

$$|\frac{\lambda_n}{n}| \left\| \frac{\beta}{f_n + \frac{\lambda_n}{n}} \right\|_{C_0(K_1^{\beta} \setminus C_1)} \leq |\frac{\lambda_{r_1}}{r_1}| \frac{4}{3\alpha_{r_1}} \|\beta\|_{C_0(\mathbb{R})} \leq \sqrt{\frac{\lambda_{r_1}}{r_1}} \frac{4}{3} \|\beta\|_{C_0(\mathbb{R})}.$$

because of the definition of α_{r_1}.

Similarly

$$K_2^{\beta} \setminus C_2 \subset int(supp(\varphi)),$$

and there exists $r_2 > r_1$ such that for all $r \geq r_2$, $K_2^\beta \setminus C_{\delta_2} \subset K_r^\varphi$. From $(A2)$ there exists $\tilde{r}_2 > r_2$ such that $\max_{r \geq \tilde{r}_2} \frac{\lambda_r}{r} \leq \frac{\lambda_{r_2}}{r_2}$.

Again, given the definition of $K_{r_2}^\varphi$ and the uniform convergence of $(f_n)_{n_geq1}$ towards φ, we deduce that there exists $N_2 > \tilde{r}_2$ such that for all $n \geq N_2$ and $t \in K_{r_2}^\varphi$,

$$\frac{3}{4}\,\alpha_{r_2} \leq f_n(t) + \frac{\lambda_n}{n}.$$

This yields that, for all n such that $n \geq N_2 > r_2$,

$$\left|\frac{\lambda_n}{n}\right| \left\| \frac{\beta}{f_n + \frac{\lambda_n}{n}} \right\|_{C_0(K_2^\beta \setminus C_2)} \leq \sqrt{\frac{\lambda_{r_2}}{r_2}} \frac{4}{3} \|\beta\|_{C_0(\mathbb{R})}.$$

We continue this way to build three strictly increasing sequences $r_j \uparrow \infty$, $\tilde{r}_j \uparrow \infty$ and $N_j \uparrow \infty$ such that for all $j \in \mathbb{N}$,

1. $N_j > \tilde{r}_j > r_j$,
2. $\forall r \geq r_j, \quad K_j^\beta \setminus C_j \subset K_r^\varphi$,
3. $\max_{r \geq \tilde{r}_j} [\frac{\lambda_r}{r}] \leq \frac{\lambda_{r_j}}{r_j}$,
4. $\forall n \geq N_j, |\frac{\lambda_n}{n}| \left\| \frac{\beta}{f_n + \frac{\lambda_n}{n}} \right\|_{C_0(K_j^\beta \setminus C_j)} \leq \sqrt{\frac{\lambda_{r_j}}{r_j}} \frac{4}{3} \|\beta\|_{C_0(\mathbb{R})}$.

Let n be an integer greater than N_1. Then there exists an integer j such that n belongs to the set $\{N_j, N_j+1, \cdots, N_{j+1}-1\}$. The following sequence (d_n) is then defined as follows:

$$d_n := \max \left\{ \frac{4}{3} \sqrt{\frac{\lambda_{r_j}}{r_j}} \|\beta\|_{C_0(\mathbb{R})},\ \frac{1}{j} \right\}. \tag{2.15}$$

It is easy to see that this sequence goes to zero and from (2.15) we conclude that for all $n \in \{N_j, N_j+1, \cdots, N_{j+1}-1\}$,

$$|\frac{\lambda_n}{n}| \left\| \frac{\beta}{f_n + \frac{\lambda_n}{n}} \right\|_{C_0(\mathbb{R} \setminus C_j)} \leq d_n, \tag{2.16}$$

because of $\mathbb{R} \setminus C_{\delta_j} = [K_j^\beta \setminus C_{\delta_j}] \cap [(K_j^\beta)^c \setminus C_{\delta_j}]$ and the definition of K_j^β (outside K_j^β, β is bounded by $1/j$). □

Proof of Theorem 3. From the decomposition (2.4), we obtain

$$\|\hat{\beta}_n - \beta\|_{L^2} \leq \left|\frac{\lambda_n}{n}\right| \left\| \frac{\beta}{\frac{1}{n}\sum_{i=1}^n |X_i|^2 + \frac{\lambda_n}{n}} \right\|_{L^2} + \quad \left\| \frac{\frac{1}{n}\sum_{i=1}^n \varepsilon_i X_i^*}{\frac{1}{n}\sum_{i=1}^n |X_i|^2 + \frac{\lambda_n}{n}} \right\|_{L^2}.$$

Let us start by showing that

$$\left\| \frac{\frac{1}{n}\sum_{i=1}^{n} \varepsilon_i X_i^*}{\frac{1}{n}\sum_{i=1}^{n} |X_i|^2 + \frac{\lambda_n}{n}} \right\|_{L^2} = O_P\left(\frac{\sqrt{n}}{\lambda_n}\right). \tag{2.17}$$

First we have

$$\mathbb{E}[\|\varepsilon X^*\|_{L^2}^2] \leq \mathbb{E}[\|\varepsilon\|_{C_0}^2]\, \mathbb{E}[\|X\|_{L^2}^2] < +\infty,$$

because of $(HA1_{FCM})$ and $(HA3_{FCM})$.

Now due to the moment monotonicity $\mathbb{E}[\|\varepsilon \overline{X}\|_{L^2}] < +\infty$, $\varepsilon \overline{X}$ is strongly integrable with the L^2-norm, so there exists a function $\mathbb{E}[\varepsilon \overline{X}] \in L^2$ which is the zero function because of $(HA1_{FCM})$. We conclude that

$$\mathbb{E}[\varepsilon \overline{X}] = 0 \qquad \text{and} \qquad \mathbb{E}[\|\varepsilon \overline{X}\|_{L^2}^2] < +\infty,$$

which, from the CLT in L^2 (see Theorem 2.7 in Bosq (2000, p. 51) and Ledoux and Talagrand (1991, p. 276) for the rate of convergence), yields to

$$\left\| \frac{1}{n}\sum_{i=1}^{n} \varepsilon_i X_i^* \right\|_{L^2} = O_P\left(\frac{1}{\sqrt{n}}\right).$$

Finally (2.17) is obtained from the fact that

$$\left\| \frac{\frac{1}{n}\sum_{i=1}^{n} \varepsilon_i X_i^*}{\frac{1}{n}\sum_{i=1}^{n} |W_i|^2 + \frac{\lambda_n}{n}} \right\|_{L^2} \leq \left| \frac{n}{\lambda_n} \right| \left\| \frac{1}{n}\sum_{i=1}^{n} \varepsilon_i X_i^* \right\|_{L^2} = O_P\left(\frac{\sqrt{n}}{\lambda_n}\right).$$

As $\frac{\sqrt{n}}{\lambda_n} \to 0$ by $(A3)$, we obtain the probability convergence of this part.

To conclude the proof, it is enough to show that

$$\left| \frac{\lambda_n}{n} \right| \left\| \frac{\beta}{\frac{1}{n}\sum_{i=1}^{n} |X_i|^2 + \frac{\lambda_n}{n}} \right\|_{L^2} \xrightarrow{a.s.} 0. \tag{2.18}$$

To that purpose, we use the fact that

$$\left\| \frac{1}{n}\sum_{i=1}^{n} |X_i|^2 - \mathbb{E}[|X|^2] \right\|_{C_0} \xrightarrow{a.s.} 0,$$

which can be obtained by applying the Strong Law of Large Numbers (SLLN) (see Bosq (2000, p. 47)) to the random function $|X|^2$. Notice here that $\mathbb{E}[|X|^2] \in C_0$.

Now for $\mathscr{S} := \{\omega \in \Omega : \|\frac{1}{n}\sum_i^n |X(\omega)|^2 - \varphi\|_{C_0} \to 0\}$, $P(\mathscr{S}) = 1$. Let us take an arbitrary and fixed value $\omega \in \mathscr{S}$. Then for $n \geq 1$ we define the sequence of functions $f_n := \frac{1}{n}\sum_{i=1}^n |X_i(\omega)|^2$. Clearly this sequence belongs to C_0 and $\|f_n - \varphi\|_{C_0} \to 0$. Thus we can use Lemma 12 which implies that there exists a sequence of subsets of $\mathbb{R}$, $(C_j)_{j\geq 1}$, a strictly increasing sequence of natural numbers

$(N_j)_{j\geq 1} \subset \mathbb{N}$ and a sequence of real numbers $(d_n)_{n\geq 1} \subset \mathbb{R}$ converging to zero, such that inequality (2.14) holds.

At this point we define for $n \geq N_1$, $R_n := \frac{1}{d_n} \to \infty$ and the intervals $\bar{I}_n := [-R_n, +R_n]$. For $n \in \{N_j, N_j+1, \cdots, N_{j+1}-1\}$, by the triangular inequality and inequality (2.14),

$$\begin{aligned} |\tfrac{\lambda_n}{n}| \left\| \frac{\beta}{f_n + \frac{\lambda_n}{n}} \right\|_{L^2(\mathbb{R})} &\leq |\tfrac{\lambda_n}{n}| \left\| \frac{\beta}{f_n + \frac{\lambda_n}{n}} \right\|_{L^2(\bar{I}_n \cap C_j)} + |\tfrac{\lambda_n}{n}| \left\| \frac{\beta}{f_n + \frac{\lambda_n}{n}} \right\|_{L^2(\bar{I}_n \cap C_j^c)} + \\ &\quad + \quad \|\beta\|_{L^2(\bar{I}_n^c)} \\ &\leq \|\beta\|_{L^2(C_j)} + |\tfrac{\lambda_n}{n}| \left\| \frac{\beta}{f_n + \frac{\lambda_n}{n}} \right\|_{C_0(\mathbb{R} \setminus C_j)} \sqrt{2\, R_n} \; + \; \|\beta\|_{L^2(\bar{I}_n^c)}. \end{aligned}$$

In this way we obtain for every $n \in \{N_j, N_j+1, \cdots, N_{j+1}-1\}$,

$$|\frac{\lambda_n}{n}| \left\| \frac{\beta}{f_n + \frac{\lambda_n}{n}} \right\|_{L^2(\mathbb{R})} \leq \|\beta\|_{L^2(C_j)} + d_n \sqrt{\frac{2}{d_n}} + \; \|\beta\|_{L^2(\bar{I}_n^c)}.$$

Thus

$$L := \lim_{n\to\infty} |\frac{\lambda_n}{n}| \left\| \frac{\beta}{f_n + \frac{\lambda_n}{n}} \right\|_{L^2(\mathbb{R})} \leq \lim_{j\to\infty} \|\beta \cdot 1_{C_j}\|_{L^2(\mathbb{R})}.$$

Finally the sequence of functions $|\beta \cdot 1_{C_j}|$ is bounded by β and is pointwise convergent to zero almost everywhere because $\{t \in \mathbb{R} : \beta \cdot 1_{C_j}(t) \to 0\}^c \subset \cap_{l=1}^{\infty} \cup_{s\geq l} C_s \subset \cap_{l=1}^{\infty} \cup_{s\geq l} D_s \subset \cap_{l=1}^{\infty} D_l \subset \partial supp(\varphi)$ which is countable then with measure zero.

By the dominated convergence theorem, $\lim_{j\to\infty} \|\beta \cdot 1_{C_j}\|_{L^2} = 0$. Thus $L = 0$ and so (2.18) is proved because ω is an arbitrary element of $\mathscr{S}$ and $P(\mathscr{S}) = 1$. □

Proof of Theorem 5. We use (2.4) and the triangle inequality to obtain

$$\|\hat{\beta}_n - \beta\|_{L^2} \leq \left|\frac{\lambda_n}{n}\right| \left\| \frac{\beta}{\frac{1}{n}\sum_{i=1}^n |X_i|^2 + \frac{\lambda_n}{n}} \right\|_{L^2(supp(|\beta|))} + \left\| \frac{\frac{1}{n}\sum_{i=1}^n \varepsilon_i X_i^*}{\frac{1}{n}\sum_{i=1}^n |X_i|^2 + \frac{\lambda_n}{n}} \right\|_{L^2}.$$

The proof of

$$\left\| \frac{\frac{1}{n}\sum_{i=1}^n \varepsilon_i X_i^*}{\frac{1}{n}\sum_{i=1}^n |X_i|^2 + \frac{\lambda_n}{n}} \right\|_{L^1} = O_P\left(\frac{\sqrt{n}}{\lambda_n}\right)$$

is the same as in Theorem 3.

Hence, to finish the proof of Theorem 5, we have to show that

$$\left\| \frac{\beta}{\frac{1}{n}\sum_{i=1}^n |X_i|^2 + \frac{\lambda_n}{n}} \right\|^2_{L^2(supp(|\beta|)\setminus J)} + \left\| \frac{\beta}{\frac{1}{n}\sum_{i=1}^n |X_i|^2 + \frac{\lambda_n}{n}} \right\|^2_{L^2(J)} = O_P(1), \tag{2.19}$$

which will lead to

$$\|\hat{\beta}_n - \beta\|_{L^2} = \left|\frac{\lambda_n}{n}\right| O_P(1) + O_P\left(\frac{\sqrt{n}}{\lambda_n}\right) = O_P\left(n^{-\gamma}\right).$$

The proof of (2.19) is based on the two following lemmas.

□

Lemma 13. *Under the assumptions of Theorem 5, we have*

$$\left\| \frac{\beta}{\frac{1}{n}\sum_{i=1}^n |X_i|^2 + \frac{\lambda_n}{n}} \right\|^2_{L^2(supp(|\beta|)\setminus J)} = O_P(1).$$

Proof of Lemma 13. Throughout the proof, we use the following notations to simplify the writing. For all $n \geq 1$, $\bar{\lambda}_n := \frac{\lambda_n}{n}$, $S_n := \sum_{i=1}^n |X_i|^2$, $\bar{S}_n := \frac{S_n}{n}$, $A_n := |\beta|/(\bar{S}_n + \bar{\lambda}_n)$. The support of function $\varphi := \mathbb{E}[|X|^2]$ is $supp(\varphi) = supp(\mathbb{E}[|X|])$, so that $C_{\beta,\partial X} = supp(|\beta|) \setminus \partial(supp(\varphi))$. Finally, the set $C := supp(|\beta|) \setminus J$ satisfies $C \subset supp(\varphi)$.

Let us define for $j \geq 1$, $r_j := \|\varphi\|_{C_0}/2^j$, $r_0 := \|\varphi\|_{C_0} + 1$, the compact sets $K_0 := \emptyset$, $K_j := \varphi^{-1}([r_j, \infty[)$, and $D_j := K_j \setminus K_{j-1}$. So we have $\cup_{j\geq 1} \uparrow K_j = supp(\varphi)$ and we can cover $C = \cup_{j\geq 1}(C \cap D_j)$.

We obtain

$$\begin{aligned}
\|A_n\|^2_{L^2(C)} &= \textstyle\sum_{j\geq 1} \|A_n 1_{\bar{S}_n \in [0, r_j/2]}\|^2_{L^2(C\cap D_j)} + \sum_{j\geq 1} \|A_n 1_{\bar{S}_n > r_j/2}\|^2_{L^2(C\cap D_j)} \\
&\leq \textstyle\frac{1}{\bar{\lambda}_n^2} \sum_{j\geq 1} \|\beta\|^2_{C_0(C\cap D_j)} m(\bar{S}_n \in [0, r_j/2] \cap C \cap D_j) + \\
&\qquad + \sum_{j\geq 1} \frac{2^2}{r_j^2} \frac{r_{j-1}^2}{r_{j-1}^2} \|\beta\|^2_{L^2(C\cap D_j)}.
\end{aligned}$$

Now for each $j \geq 1$, $\frac{r_{j-1}}{r_j} \leq \frac{r_0}{r_1}$ and in the set $C \cap D_j$, $\frac{\beta}{r_{j-1}} < \frac{\beta}{\varphi} \leq \frac{\beta}{r_j}$. Then $\|\beta\|_{C_0} \leq M_2 r_{j-1}$ because of part (c) of $(A5)$. Thus

$$\begin{aligned}
\|A_n\|^2_{L^2(C)} &\leq \textstyle\frac{1}{\bar{\lambda}_n^2} M_2^2 (\frac{r_0}{r_1})^2 \sum_{j\geq 1} r_{j-1}^2 \, m(\bar{S}_n \in [0, r_j/2] \cap C \cap D_j) + \\
&\qquad + \quad 4(\tfrac{r_0}{r_1})^2 \textstyle\sum_{j\geq 1} \|\frac{\beta}{\varphi}\|^2_{L^2(C\cap D_j)}.
\end{aligned}$$

Moreover

$$\begin{aligned}
\textstyle\sum_{j\geq 1} \frac{r_j^2}{4} m(\bar{S}_n \in [0, r_j/2] \cap C \cap D_j) &\leq \textstyle\sum_{j\geq 1} \|(\varphi - \bar{S}_n) 1_{\bar{S}_n \in [0, r_j/2]}\|^2_{L^2(C\cap D_j)} \\
&\leq \|\varphi - \bar{S}_n\|^2_{L^2(C)}.
\end{aligned}$$

Now we can bound A_n

$$\begin{aligned}
\|A_n\|^2_{L^2(C)} &\leq \textstyle\frac{1}{\bar{\lambda}_n^2} M_2^2 (\frac{r_0}{r_1})^2 \times 4 \|\varphi - \bar{S}_n\|^2_{L^2(C)} + 4(\frac{r_0}{r_1})^2 \|\frac{\beta}{\varphi}\|^2_{L^2(C)} \\
&= 4M_2^2 (\tfrac{r_0}{r_1})^2 O_P((\tfrac{\sqrt{n}}{\lambda_n})^2) + 4(\tfrac{r_0}{r_1})^2 \|\tfrac{\beta}{\varphi}\|^2_{L^2(C)} = O_P(1).
\end{aligned}$$

□

Lemma 14. *Under the assumptions of Theorem 5, we have*

$$\left\| \frac{\beta}{\frac{1}{n}\sum_{i=1}^{n}|X_i|^2 + \frac{\lambda_n}{n}} \right\|^2_{L^2(J)} = O_P(1).$$

Proof of Lemma 14. We start the proof by considering the set $C_{\beta,\partial X}$. As $supp(\varphi)$ is an open set in $\mathbb{R}$, it is an union of open intervals. Because of this, $\partial(supp(\varphi))$ is countable. Besides, by hypothesis $(A5)$, for every $p \in C_{\beta,\partial X}$, there is an open neighborhood J_p, in which (a) holds. Thus for all $p \in C_{\beta,\partial X}$, $J_p \cap \partial(supp(\varphi)) = \{p\}$. These intervals J_p are countable and pairwise disjoint.

Now we suppose that $card(C_{\beta,\partial X}) = +\infty$ (the case where this set is finite is similar). We denote its elements as p_v, with $v \geq 1$. So J is the union of disjoints intervals $J = \cup_{v\geq 1} J_v$, where $J_v := J_{p_v}$, and part (b) of $(A5)$ can be written as $\sum_{v\geq 1} \|\beta\|^2_{C_0(J_v)} < M_1$.

For $n \geq 1$, let us define $\xi_n := \bar{\lambda}_n^{2\alpha}$. Clearly from $(A6)$, $\xi_n \downarrow 0$. We define for $l \geq 1$, the compact sets $K_0^\xi := \emptyset$, $K_l^\xi := \varphi^{-1}([\xi_l, \infty[)$, and $D_l^\xi := K_l^\xi \setminus K_{l-1}^\xi$. So we have $\cup \uparrow K_l^\xi = supp(\varphi)$ and we can cover $J_v \setminus \{p_v\} = \cup_{j\geq 1}(J_v \cap D_j^\xi)$ for each fixed $v \geq 1$. Moreover in D_l^ξ, $\frac{1}{\xi_{l-1}} < \frac{1}{\varphi} \leq \frac{1}{\xi_l}$.

Let us take a fixed $v \geq 1$. Given the fact that ξ_l is strictly decreasing to zero, by hypothesis $(A6)$, there exists a unique number $N_v \in \mathbb{N}$ such that

$$\xi_{N_v} < \max_{t\in\partial(J_v)} |t - p_v|^\alpha \leq \xi_{N_v - 1}.$$

Then for every $n \geq N_v$,

$$\begin{aligned}
\|A_n\|^2_{L^2(J_v)} &= \sum_{l=N_v}^{n} \|A_n\|^2_{L^2(J_v\cap D_l^\xi)} + \|A_n\|^2_{L^2(J_v\setminus K_n^\xi)} \\
&= \sum_{l=N_v}^{n} \|A_n 1_{\bar{S}_n\in[0,\xi_l/2]}\|^2_{L^2(J_v\cap D_l^\xi)} + \\
&\quad + \sum_{l=N_v}^{n} \|A_n 1_{\bar{S}_n\geq\xi_l/2}\|^2_{L^2(J_v\cap D_l^\xi)} \quad + \quad \|A_n\|^2_{L^2(J_v\setminus K_n^\xi)} \\
&\leq \|\beta\|^2_{C_0(J_v)} \left[\frac{1}{\bar{\lambda}_n^2} \sum_{l=N_v}^{n} m(\bar{S}_n \in [0,\xi_l/2] \cap J_v \cap D_l^\xi)\right] \\
&\quad + \quad \|\beta\|^2_{C_0(J_v)} \left[\sum_{l=N_v}^{n} \frac{4}{\xi_l^2} m(J_v \cap D_l^\xi) + \quad \frac{1}{\bar{\lambda}_n^2} m(J_v \setminus K_n^\xi)\right].
\end{aligned}$$

Using the inequality

$$\frac{\xi_n^2}{4} \sum_{l=N_v}^{n} m(\bar{S}_n \in [0,\xi_l/2] \cap J_v \cap D_l^\xi) \leq \|\varphi - \bar{S}_n\|_{L^2(J_v)},$$

we obtain

$$\|A_n\|^2_{L^2(J_v)} \leq \|\beta\|^2_{C_0(J_v)} \left[\frac{1}{\bar{\lambda}_n^2} \frac{4}{\rho_n^2} \|\varphi - \bar{S}_n\|_{L^2(J_v)} + 4 \sum_{l=N_v}^{n} \frac{\xi_{l-1}^2}{\xi_l^2} \frac{m(J_v \cap D_l^{\xi})}{\xi_{l-1}^2} + \right.$$

$$\left. + \frac{1}{\bar{\lambda}_n^2} m(J_v \setminus K_n^{\xi}) \right].$$

Because of $(A6)$, there exists $M_3 > 0$ such that for $l \geq 1$, $|\frac{\lambda_{l-1}}{\lambda_l}| \leq M_3$. Thus for $n \geq N_v$,

$$\|A_n\|^2_{L^2(J_v)} \leq \|\beta\|^2_{C_0(J_v)} \left[\frac{4}{\bar{\lambda}_n^{2+4\alpha}} \|\varphi - \bar{S}_n\|_{L^2(J_v)} + \right.$$

$$\left. + 4M_3^2 \|\frac{1}{\varphi}\|^2_{L^2(J_v \cap K_n^{\xi})} + \frac{1}{\bar{\lambda}_n^2} m(J_v \setminus K_n^{\xi}) \right].$$

Moreover, if $t \in J_v \setminus K_n^{\xi}$, $0 \leq \varphi(t) < \xi_n$ hence $|t - p_v|^{\alpha} \leq \varphi(t) < \xi_n$ and in particular $J_v \setminus K_n^{\xi} \subset [p_v - \xi_n^{1/\alpha}, p_v + \xi_n^{1/\alpha}]$. In this way we can prove that for $n \geq N_v$, $m(J_v \setminus K_n^{\xi}) \leq 2\xi_n^{1/\alpha} \leq 2\bar{\lambda}_n^2$. We obtain from this that for every $n \in \{1, \cdots, N_v - 1\}$,

$$\|A_n\|^2_{L^2(J_v)} \leq \frac{1}{\bar{\lambda}_n^2} \|\beta\|^2_{L^2(J_v)},$$

and for $n \geq N_v$,

$$\|A_n\|^2_{L^2(J_v)} \leq 4\|\beta\|^2_{C_0(J_v)} \left[n\|\bar{S}_n - \varphi\|^2_{L^2(J_v)} + M_3^2 \|\frac{1}{\varphi}\|^2_{L^2(J_v)} + 1/2 \right].$$

To finish the proof of this lemma, we bound the sequence $\|A_n\|^2_{L^2(J)} = \sum_{v \geq 1} \|A_n\|^2_{L^2(J_v)}$. In order to do this we define for each $n \geq 1$, the set $C_n := \{n \geq 1 : n \in [1, \cdots, N_v - 1]\}$. We obtain

$$\begin{aligned} \|A_n\|^2_{L^2(J)} &\leq \frac{1}{\bar{\lambda}_n^2} \|\beta\|^2_{L^2(\cup_{v \in C_n} J_v)} + \\ &+ \left(4 \sum_{v \geq 1} \|\beta\|^2_{C_0(J_v)} \right) \left[n\|\bar{S}_n - \varphi\|^2_{L^2(J)} + M_3^2 M_0^2 + 1/2 \right] \\ &\leq \frac{1}{\bar{\lambda}_n^2} \|\beta\|^2_{L^2(\cup_{v \in C_n} J_v)} + 4M_1 \left[O_P(1) + M_3^2 M_0^2 + 1/2 \right]. \end{aligned}$$

For each $n \geq 1$, $v \in C_n$ then $n < N_v$, hence $\xi_n \geq \max_{t \in \partial J_v} (t - p_v)^{\alpha}$, from what we deduce that $m(J_v) \leq 2\,\xi_n^{1/\alpha}$. We obtain for $n \geq 1$

$$\|\beta\|^2_{L^2(\cup_{v \in C_n} J_v)} \leq 2\xi_n^{1/\alpha} \sum_{v \in C_n} \|\beta\|^2_{C_0(J_v)} \leq 2\xi_n^{1/\alpha} \left[\sum_{v \geq 1} \|\beta\|^2_{C_0(J_v)} \right] = 2\xi_n^{1/\alpha} \left[M_1/4 \right],$$

and thus for $n \geq 1$,

$$\begin{aligned} \|A_n\|^2_{L^2(J)} &\leq \tfrac{1}{\bar{\lambda}_n^2} 2\xi_n^{1/\alpha} \tfrac{M_1}{4} + 4M_1 \left[O_P(1) + M_3^2 M_0^2 + 1/2\right] \\ &\leq \tfrac{M_1}{2} + 4M_1 O_P(1) + 4M_1 \left[M_3^2 M_0^2 + 1/2\right] = O_P(1). \end{aligned}$$

□

Proof of Corollary 7. It is a particular case of Theorem 5. First, $(A4bis)$ implies that, for all $t \in supp(\beta)$, $|\beta(t)|/\varphi(t)$ is finite. Thus $supp(\beta) \subset supp(\varphi)$ and $supp(\beta) \cap \partial(supp(\varphi)) = \emptyset$. Because of this, parts (a) and (b) of hypothesis $(A5)$ are true by default.

Moreover, $(A4bis)$ implies $(A4)$, and if we have $J := \emptyset$, $supp(\beta) \cap \partial(supp(\varphi)) = \emptyset$ implies part (c) of $(A5)$. Finally, equation (2.19) in the proof of Theorem 5 is replaced by

$$\left\| \frac{\beta}{\frac{1}{n}\sum_{i=1}^n |X_i|^2 + \frac{\lambda_n}{n}} \right\|^2_{L^2(supp(|\beta|))} = O_P(1),$$

which is proved with the same technique. □

Proof of Theorem 8. We start with the decomposition

$$\|\hat{\beta}_n - \beta\|_{L^2(K)} = \left|\frac{\lambda_n}{n}\right| \left\| \frac{\beta}{\frac{1}{n}\sum_{i=1}^n |X_i|^2 + \frac{\lambda_n}{n}} \right\|_{L^2(K)} + \left\| \frac{\frac{1}{n}\sum_{i=1}^n \varepsilon_i X_i^*}{\frac{1}{n}\sum_{i=1}^n |X_i|^2 + \frac{\lambda_n}{n}} \right\|_{L^2(K)}.$$

The proof of $\left\| \frac{\frac{1}{n}\sum_{i=1}^n \varepsilon_i X_i^*}{\frac{1}{n}\sum_{i=1}^n |X_i|^2 + \frac{\lambda_n}{n}} \right\|_{L^2(K)} = O_P(\frac{\sqrt{n}}{\lambda_n})$ is the same as in Theorem 3. We finish the proof of the theorem by showing

$$\left\| \frac{\beta}{\frac{1}{n}\sum_{i=1}^n |X_i|^2 + \frac{\lambda_n}{n}} \right\|_{L^2(K)} = O_P(1). \tag{2.20}$$

Given that $K \subset supp(\varphi)$, there exists a positive number $s_1 > 0$ such that $K \subset K_{s_1}^{\varphi}$, where $K_{s_1}^{\varphi} := \varphi^{-1}([s_1, \infty[)$ is a compact in $\mathbb{R}$. We define $s := s_1/2$. We have for every $n \in \mathbb{N}$,

$$\left\| \frac{\beta}{\bar{S}_n + \bar{\lambda}_n} \right\|_{L^2(K)} \leq \left\| \frac{\beta}{\bar{S}_n + \bar{\lambda}_n} 1_{\bar{S}_n \in [0,s]} \right\|_{L^2(K)} + \left\| \frac{\beta}{\bar{S}_n + \bar{\lambda}_n} 1_{\bar{S}_n \in [s,\infty[} \right\|_{L^2(K)}.$$

Clearly, the first part above is bounded by

$$\left\| \frac{\beta}{\bar{S}_n + \bar{\lambda}_n} 1_{\bar{S}_n \in [s,\infty[} \right\|_{L^2(K)} \leq \frac{1}{s} \|\beta\|_{L^2(K)} = O_P(1).$$

To bound the other part we have

$$\left\| \frac{\beta}{\bar{S}_n + \bar{\lambda}_n} 1_{\bar{S}_n \in [0,s]} \right\|_{L^2(K)} \leq \frac{1}{\bar{\lambda}_n} \left\| \beta \, 1_{\bar{S}_n \in [0,s]} \right\|_{L^2(K)} \leq \frac{\|\beta\|_{C_0}}{\bar{\lambda}_n} \sqrt{m(K \cap \bar{S}_n \in [0,s])}.$$

Moreover, thanks to hypothesis $(A3)$, we have $\|\bar{S}_n - \varphi\|_{L^2(K)} = O_P(\frac{1}{\sqrt{n}})$. This inequality, together with the fact that $|\bar{S}_n - \varphi| > s$ whenever $\bar{S}_n \in [0,s]$, allows us to obtain

$$\begin{aligned} \|\bar{S}_n - \varphi\|_{L^2(K)} &\geq \|(\bar{S}_n - \varphi) 1_{\bar{S}_n \in [0,s]}\|_{L^2(K)} \geq \sqrt{\textstyle\int_K |s|^2 1_{\bar{S}_n \in [0,s]} dm} \\ &\geq |s| \sqrt{m(K \cap \bar{S}_n \in [0,s])}. \end{aligned}$$

In this way, $\sqrt{m(K \cap \bar{S}_n \in [0,s])} = O_P(\frac{1}{\sqrt{n}})$ and as a consequence

$$\left\| \frac{\beta}{\bar{S}_n + \bar{\lambda}_n} 1_{\bar{S}_n \in [0,s]} \right\|_{L^2(K)} \leq \frac{\|\beta\|_{C_0}}{\bar{\lambda}_n} \, O_P(\frac{1}{\sqrt{n}}) = O_P(\frac{\sqrt{n}}{\lambda_n}),$$

which finishes to prove (2.20). □

Proof of Proposition 10. We only need to prove that for every $i \in \{1, \cdots, n\}$,

$$Y_i - \hat{\beta}_n^{(-i)} X_i = \frac{Y_i - \hat{\beta}_n X_i}{1 - A_{i,i}}. \tag{2.21}$$

Let us take an arbitrary $i \in \{1, \cdots, n\}$. We define for each $j \in \{1, \cdots, n\}$,

$$\tilde{Y}_j := \begin{cases} Y_j & \text{if } j \neq i, \\ \hat{\beta}_n^{(-i)} X_j & \text{otherwise.} \end{cases}$$

Because $\hat{\beta}_n^{(-i)} = \frac{\sum_{l \neq i}^n Y_l \overline{X}_l}{\sum_{l \neq i}^n |X_l|^2 + \lambda_n}$ by definition, we have

$$\begin{aligned} \frac{\sum_{l=1}^n \tilde{Y}_l \overline{X}_l}{S_n + \lambda_n} &= \frac{\sum_{l \neq i}^n Y_l \overline{X}_l}{S_n + \lambda_n} + \frac{\hat{\beta}_n^{(-i)} |X_i|^2}{S_n + \lambda_n} \\ &= \hat{\beta}_n^{(-i)} \left[\frac{\sum_{l \neq i}^n |X_l|^2 + \lambda_n}{S_n + \lambda_n} + \frac{|X_i|^2}{S_n + \lambda_n} \right] = \hat{\beta}_n^{(-i)}. \end{aligned}$$

Then

$$\hat{\beta}_n X_i - \hat{\beta}_n^{(-i)} X_i = \frac{\sum_{l=1}^n Y_l \overline{X}_l - \sum_{l=1}^n \tilde{Y}_l \overline{X}_l}{S_n + \lambda_n} X_i = \frac{Y_i - \hat{\beta}_n^{(-i)} X_i}{S_n + \lambda_n} |X_i|^2,$$

from what we obtain

$$1 - \frac{Y_i - \hat{\beta}_n X_i}{Y_i - \hat{\beta}_n^{(-i)} X_i} = \frac{\hat{\beta}_n X_i - \hat{\beta}_n^{(-i)} X_i}{Y_i - \hat{\beta}_n^{(-i)} X_i} = \frac{|X_i|^2}{S_n + \lambda_n} = A_{i,i},$$

which implies (2.21). □

Proof of Proposition 11. It is similar to that of Proposition 10. □

Chapter 3

Estimation for the Functional Convolution Model

Contents

Abstract: The aim of this paper is to propose an estimator of the unknown function in the Functional Convolution Model (FCVM), which models the relationship between a functional covariate $X(t)$ and a functional response $Y(t)$ through the following equation $Y(t) = \int_0^t \theta(s)X(t-s)ds + \varepsilon(t)$, where θ is the function to be estimated and ε is an additional functional noise. In this way we can study the influence of the history of X on $Y(t)$. We use the continuous Fourier transform ($\mathscr{F}$) and its inverse ($\mathscr{F}^{-1}$) to define an estimator of θ. First we transform the problem into an equivalent one in the frequency domain, where we use the Functional Ridge Regression Estimator (FREE) to estimate $\mathscr{F}(\theta)$. Then we use $\mathscr{F}^{-1}$ to estimate θ in the time domain. We establish consistency properties of the proposed estimator. Afterwards we present some simulations to compare the performance of this estimator with others already developed in the literature.
Keywords and phrases: Convolution model; Concurrent model; Functional data; Historical functional linear model.

3.1 Introduction

Functional Data Analysis (FDA) deals among other things with longitudinal data (e.g. random curves) for which the number of observation times (p) is much bigger than the sample size (n). In this situations, it has been showed that multivariate data analysis methods could fail (see Hsing and Eubank (2015)) and it is more appropriate to use FDA methods that take into account the functional nature of the data. In particular to study the interaction of two such random curves we can consider some Functional linear regression models. In this paper we are interested in the Functional Convolution Model (FCVM)

$$Y(t) = \int_0^t \theta(s)X(t-s)ds + \varepsilon(t), \tag{3.1}$$

where $t \geq 0$, θ is the function to be estimated, X, Y are random functions and ε is a noise random function. All these functions are considered to be equal to zero for all $t < 0$. Integral in model (3.1) is called a convolution.

We can use this model to study the influence of the history of X on $Y(t)$. It was derived from Malfait and Ramsay (2003) by requiring that the function θ remains the same for each time t (i.e it only depends on s). In this paper we study how to estimate θ with a given i.i.d. sample $(X_i, Y_i)_{i\in\{1,\cdots,n\}}$ of the random functions X and Y.

Some related models have been treated in the literature. On the one hand Asencio et al. (2014) studies a related problem, in which they consider more covariate functions. The estimation of θ is done by projecting the functions into finite-dimensional spline subspaces. On the other hand Malfait and Ramsay (2003) consider the case where θ in the model (3.1) depends also on t (i.e. $\theta(s,t)$), thus the integral becomes a kernel operator and they use a Finite Elements method to estimate it over an appropriate domain. Both papers are the most relevant among those in the literature which study the estimation of θ with an i.i.d. sample $(X_i, Y_i)_{i\in\{1,\cdots,n\}}$ and where X and Y are related through the

convolution over $[0,\infty]$. Our approach is a new way to answer this question by using the equivalent representation of the model in the frequency domain as we explain later.

In the literature the FCVM is related to the multichannel deconvolution problem (see e.g. De Canditiis and Pensky (2006), Pensky et al. (2010) and Kulik et al. (2015)) which treats the periodic case with $n > 1$ (multichannel). On the other hand all the following methods deal with the case $n = 1$: The Deconvolution Problem in Non-parametric statistics (see e.g. Meister (2009), Johannes et al. (2009), Comte et al. (2016)), in Signal Processing (see e.g. Johnstone et al. (2004), Brown and Hwang (2012)) or in Linear Inverse Problems (see e.g. Abramovich and Silverman (1998)) and the Laplace Deconvolution (see e.g. Comte et al. (2016)). In this paper we are interested to study the estimation in the non periodic case with $n > 1$ and the asymptotic behavior when n goes to infinity.

We develop another approach to estimate θ. The main idea is to use the Continuous Fourier Transform (CFT) to transform the problem into its equivalent in the frequency domain. This yields to a particular case of the Functional Concurrent Model (FCCM) (Ramsay and Silverman (2005, Ch 14)). The equation of the associated FCCM is

$$\mathscr{Y}(\xi) = \beta(\xi)\,\mathscr{X}(\xi) + \varepsilon(\xi), \tag{3.2}$$

where $\xi \in \mathbb{R}$, β is an unknown function to be estimated, $\mathscr{X} := \mathscr{F}(X)$, $\mathscr{Y} := \mathscr{F}(Y)$ and $\varepsilon := \mathscr{F}(\varepsilon)$ are Fourier transforms of X and Y, and ε respectively.

Once in the frequency domain we use an estimator of β for the FCCM, and then come back to the time domain through the Inverse Continuous Fourier Transform (ICFT). We called the estimator defined in this way the Functional Fourier Deconvolution Estimator (**FFDE**).

The estimation of β have already been discussed by several authors in the context of some related models to the FCCM. For instance Hastie and Tibshirani (1993) has proposed a generalization of FCCM called 'varying coefficient model'. Until now many methods have been developed to estimate the unknown smooth regression function β, for instance by local maximum likelihood estimation (see e.g. Dreesman and Tutz (2001)), or by local polynomial smoothing (see e.g. Fan et al. (2003)). These methods use techniques from multivariate data analysis, as noticed by Ramsay and Silverman (2005, p. 259). In our case we do not use these methods because we want to estimate $\beta = \mathscr{F}(\theta)$ over the whole frequency domain (i.e. for all $\xi \in \mathbb{R}$). To do this we use the Functional Ridge Regression Estimator of β defined in Manrique et al. (2016).

In this paper we prove the consistency of the **FFDE** and obtain a rate of convergence. We present some simulations to compare the performance of this estimator with others we have adapted from different methods found in the literature. These simulations show the robustness, the accuracy and the fast computation time of our estimator compared to the others.

All the proofs are postponed to Appendix 3.B, where we use the main theorems of Manrique et al. (2016) (Appendix 3.A).

3.2 Model and Estimator

Let us first define some useful notations. We define $L^1(\mathbb{R},\mathbb{C})$ the space of integrable complex valued functions, with the L^1-norm $\|f\|_{L^1} := [\int_{\mathbb{R}} |f(x)|dx]$ where $|\cdot|$ denotes the complex modulus. $L^1(\mathbb{R},\mathbb{R})$ is the subspace of integrable real valued functions. In a similar way $L^2(\mathbb{R},\mathbb{C})$ is the space of square integrable complex valued functions, with the L^2-norm $\|f\|_{L^2} := \left[\int_{\mathbb{R}} |f(x)|^2 dx\right]^{1/2}$, and $L^2(\mathbb{R},\mathbb{R})$ is the subspace of square integrable real valued functions. Given a subset $K \subset \mathbb{R}$, $\|f\|_{L^2(K)} := \left[\int_K |f(x)|^2 dx\right]^{1/2}$.

Let $C_0(\mathbb{R},\mathbb{C}) = C_0$ be the space of complex valued continuous functions f that vanish at infinity, that is, for all $\zeta > 0$ there exists a $R > 0$ such that for all $|t| > R$, $|f(t)| < \zeta$. We use the supremum norm $\|f\|_{C_0} := \sup_{x\in\mathbb{R}} |f(x)|$. For a subset $K \subset \mathbb{R}$, this norm is $\|f\|_{C_0(K)} := \sup_{x\in K} |f(x)|$.

The Continuous Fourier Transform (CFT) is denoted by $\mathscr{F}$ and its inverse (ICFT) by $\mathscr{F}^{-1}$. Lastly throughout this paper the support of a continuous function $f : \mathbb{R} \to \mathbb{C}$ is the set $supp(f) := \{t \in \mathbb{R} : |f(t)| \neq 0\}$. This set is open because f is continuous. Besides we define the boundary of a set S, as $\partial(S) := \overline{S} \setminus int(S)$, where $\overline{S}$ is the closure of S and $int(S)$ is its interior.

3.2.1 General Hypotheses of the FCVM

We present the general hypotheses that will be used along the paper.

$(HA1_{FCVM})$	X, ε are independent $L^1(\mathbb{R},\mathbb{R}) \cap L^2(\mathbb{R},\mathbb{R})$ valued random functions such that $\mathbb{E}(\varepsilon) = 0$ and for every $t < 0$, we have $\varepsilon(t) = X(t) = 0$.
$(HA2_{FCVM})$	$\theta \in L^2(\mathbb{R},\mathbb{R})$ and for every $t < 0$, $\theta(t) = 0$.
$(HA3_{FCVM})$	The expectations $\mathbb{E}(\|\varepsilon\|^2_{L_1}), \mathbb{E}(\|X\|^2_{L_1}), \mathbb{E}(\|\varepsilon\|^2_{L^2})$ and $\mathbb{E}(\|X\|^2_{L^2})$ are all finite.

Under the assumption that for every $t < 0$, $\theta(t) = X(t) = 0$, obtained from $(HA1_{FCVM})$ and $(HA2_{FCVM})$, it is possible to write the integral in model (3.1) over the whole real line, that is $Y(t) = \int_{-\infty}^{+\infty} \theta(s)X(t-s)ds + \varepsilon(t)$. It allows to use the CFT to transform the convolution into a multiplication.

3.2.2 Functional Fourier Deconvolution Estimator (FFDE)

To define the **FFDE** we proceed in three steps. i) First we use the CFT to transform the convolution in the time domain into a simple multiplication in the frequency domain (see (3.2)). ii) Once in the frequency domain, we estimate β with the **Functional Ridge Regression Estimator (FRRE)** defined in Manrique et al. (2016), which is an extension of the Ridge Regularization method (Hoerl

(1962)) that deals with ill-posed problems in the classical linear regression. iii) The last step consists in using the ICFT to estimate θ.

Let $(X_i, Y_i)_{i=1,\cdots,n}$ be an i.i.d sample of FCVM (3.1) and λ_n be a positive regularization parameter. The FRRE of β in the FCCM (3.2) is defined as follows

$$\hat{\beta}_n := \frac{\frac{1}{n}\sum_{i=1}^{n} \mathscr{Y}_i \, \mathscr{X}_i^*}{\frac{1}{n}\sum_{i=1}^{n} |\mathscr{X}_i|^2 + \frac{\lambda_n}{n}}, \tag{3.3}$$

where the exponent $*$ stands for the complex conjugate, $\mathscr{Y}_i = \mathscr{F}(Y_i)$ and $\mathscr{X}_i = \mathscr{F}(X_i)$. Finally the **FFDE** of θ in (3.1) is defined by

$$\hat{\theta}_n := \mathscr{F}^{-1}(\hat{\beta}_n). \tag{3.4}$$

3.3 Asymptotic Properties of the FFDE

The main result of this paper is the probability convergence of the **FFDE** with a rate of convergence

$$\|\hat{\theta}_n - \theta\|_{L^2} = O_P\left(\max\left[\frac{\lambda_n}{n}, \frac{\sqrt{n}}{\lambda_n}\right]\right),$$

under large conditions.

The CFT is an isometry in the L^2-space. Thus the study of the asymptotic behavior of $\|\hat{\theta}_n - \theta\|_{L^2}$ is equivalent to that of $\|\hat{\beta}_n - \beta\|_{L^2}$. We use this important fact to show the consistency of the **FFDE**. In addition, it is useful to notice that $\hat{\beta}_n - \beta$ can be decomposed as follows

$$\hat{\beta}_n - \beta = -\frac{\lambda_n}{n}\left[\frac{\beta}{\frac{1}{n}\sum_{i=1}^{n} |\mathscr{X}_i|^2 + \frac{\lambda_n}{n}}\right] + \frac{\frac{1}{n}\sum_{i=1}^{n} \varepsilon_i \mathscr{X}_i^*}{\frac{1}{n}\sum_{i=1}^{n} |\mathscr{X}_i|^2 + \frac{\lambda_n}{n}}. \tag{3.5}$$

Equation (3.5) shows the relationship between the frequencies of θ and X (β and $\mathscr{X}$ respectively) through the ratio $\beta/|\mathscr{X}|^2$. This ratio plays a central role in the asymptotic behavior of the **FFDE**. The consistency is obtained in Theorem 15 under assumption $(A1)$ related to this relationship.

3.3.1 Consistency of the Estimator

Theorem 15. *Let us consider the FCVM with the general hypotheses ($HA1_{FCVM}$), ($HA2_{FCVM}$) and ($HA3_{FCVM}$). Let $(X_i, Y_i)_{i\geq 1}$ be i.i.d. realizations of (X,Y) according to this model. Additionally we suppose that*

(A1) $\overline{supp(|\mathscr{F}(\theta)|)} \subseteq \overline{supp(\mathbb{E}[|\mathscr{F}(X)|])}$,

(A2) $(\lambda_n)_{n\geq 1} \subset \mathbb{R}^+$ *is such that* $\frac{\lambda_n}{n} \to 0$ *and* $\frac{\sqrt{n}}{\lambda_n} \to 0$ *as* $n \to +\infty$.

Then

$$\lim_{n\to+\infty} \|\hat{\theta}_n - \theta\|_{L^2} = 0 \quad \textit{in probability.} \tag{3.6}$$

Remark. *Hypothesis (A2) is classic when studying asymptotic behavior. Hypothesis (A1) specifies that it is not possible to estimate $\mathscr{F}(\theta)$ outside the support of $\mathbb{E}[|\mathscr{F}(X)|]$. In the intervals where $\mathbb{E}[|\mathscr{X}|] = 0$, model (3.2) reduces to $\mathscr{Y} = \varepsilon$ and no estimation of β is possible.*

A practical interpretation of Hypothesis *(A1)* is that the **FFDE** will only converge toward θ if all the non zero frequencies of θ belong to the set of the non-zero frequencies of $\mathbb{E}[X]$, that is, there are enough non-zero frequencies of $\mathbb{E}[X]$ to "save" the information about the non-zero frequencies of θ. This allows to reconstruct θ through the ICFT.

3.3.2 Rate of Convergence

To obtain the rate of convergence we need to assume a stronger hypothesis about the relationship between the frequencies of θ and X. Hypothesis *(A4)* in Theorem 16 is one way to do this.

Theorem 16. *Let us consider the FCVM with the general hypotheses ($HA1_{FCVM}$), ($HA2_{FCVM}$) and ($HA3_{FCVM}$). We assume additionally that:*

(A3) $\mathbb{E}[\||X|^2\|_{L^2}^2] < \infty$,

(A4) $\frac{|\mathscr{F}(\theta)|}{\mathbb{E}[|\mathscr{F}(X)|^2]} 1_{supp(|\mathscr{F}(\theta)|)} \in L^2 \cap L^\infty$,

then

$$\|\hat{\theta}_n - \theta\|_{L^2} = O_P\left(\max\left[\frac{\lambda_n}{n}, \frac{\sqrt{n}}{\lambda_n}\right]\right). \tag{3.7}$$

Remark. *Hypothesis $(A3)$ is classic and allows to apply the CLT on the denominator of the decomposition of $\hat{\beta}_n - \beta$ in (3.5). Hypothesis $(A4)$ implies $(A1)$ and is required to control the first term of (3.5) and the decreasing rate of $\mathscr{F}(\theta)$ with respect to $\mathbb{E}[|\mathscr{F}(X)|^2]$ when frequencies go to infinity (tails control).*

More precisely $(A4)$ implies the strict inclusion, $supp(\mathscr{F}(\theta)) \subset supp(\mathbb{E}[|\mathscr{F}(X)|])$. This condition might be too strong when for instance $supp(\mathbb{E}[|\mathscr{F}(X)|]) = \mathbb{R} \setminus S$, where S is a non-dense and countable set and $supp(\mathscr{F}(\theta)) \cap S \neq \emptyset$. In Theorem 23, given in Appendix 3.C, we weaken Hypothesis $(A4)$ to also obtain the rate of convergence in this case. Nevertheless the hypotheses which replace $(A4)$, namely $(A4bis)$ and $(A5)$ are more difficult to interpret.

The ratio $\frac{|\mathscr{F}(\theta)|}{\mathbb{E}[|\mathscr{F}(X)|^2]}$ in Hypothesis *(A4)*, is a way to measure the regularity of $\mathscr{F}(\theta)$ with respect to that of $\mathbb{E}[|\mathscr{F}(X)|^2]$. The fact that this ratio belongs to $L^2 \cap L^\infty$ can be interpreted as the fact that θ is required to be more regular than $\mathbb{E}[X]$. For example it implies that $\mathscr{F}(\theta)$ decreases to 0 at infinity faster than $\mathbb{E}[|\mathscr{F}(X)|^2]$. This kind of regularity assumption is commonly used in Functional Linear Regression problems, where the unknown function of the regression model is more regular than the covariate X (see e.g. Cardot et al. (1999) and Cardot et al. (2003)).

Finally Theorem 17 deals with the convergence rate on compact subsets of the support of $\mathbb{E}[|\mathscr{F}(X)|^2]$ without using the Hypothesis *(A4)*. This theorem is useful when the support of $\mathscr{F}(\theta)$ is compact.

Theorem 17. *Under hypotheses* $(A1)$ *and* $(A3)$*, for every compact subset* $K \subset supp(\mathbb{E}[|\mathscr{F}(X)|])$*, we have*

$$\|\hat{\theta}_n - \theta\|_{L^2(K)} = O_P\left(\max\left[\frac{\lambda_n}{n}, \frac{\sqrt{n}}{\lambda_n}\right]\right). \tag{3.8}$$

In particular if $\overline{supp(\mathscr{F}(\theta))}$ *is compact and is a subset of* $supp(\mathbb{E}[|\mathscr{F}(X)|])$*, then*

$$\|\hat{\theta}_n - \theta\|_{L^2} = O_P\left(\max\left[\frac{\lambda_n}{n}, \frac{\sqrt{n}}{\lambda_n}\right]\right).$$

3.4 Selection of the Regularization Parameter

In this section we introduce a selection procedure of the regularization parameter λ_n for a given sample $(X_i, Y_i)_{i\in\{1,\cdots,n\}}$. We chose the Leave-one-out Predictive Cross-Validation (LOOPCV) criterion. Its definition can be found in Febrero-Bande and Oviedo de la Fuente (2012, p. 17) or Hall and Hosseini-Nasab (2006, p. 117) and is the following

$$LOOPCV(\lambda_n) := \frac{1}{n}\sum_{i=1}^{n} \|Y_i - Conv(\hat{\theta}_n^{(-i)}, X_i)\|_{L^2}^2,$$

where $\hat{\theta}_n^{(-i)}$ is computed with the sample $(X_j, Y_j)_{j\in\{1,\cdots,i-1,i+1,\cdots,n\}}$ and for $t \geq 0$, $Conv(\hat{\theta}_n^{(-i)}, X_i)(t) := \int_0^t \hat{\theta}_n^{(-i)}(s)\, X_i(t-s)\, ds$ denotes the convolution.

The selection method consists in choosing the value λ_n which minimizes the function $LOOPCV$. Proposition 18 shows a way to compute the LOOPCV with only one regression instead of n while working directly in the frequency domain. The proof is based on the fact that the CFT is an isometry and on similar ideas as in Green and Silverman (1994, pp. 31-33) about the smoothing parameter selection for smoothing splines.

Proposition 18. *We have*

$$LOOPCV(\lambda_n) = \frac{1}{n}\sum_{i=1}^{n} \left\| \frac{\mathscr{Y}_i - \hat{\beta}_n \mathscr{X}_i}{1 - A_{i,i}} \right\|_{L^2}^2, \tag{3.9}$$

where $A_{i,i} \in L^2$ *is defined as follows* $A_{i,i} := |\mathscr{X}|_i^2 / (\sum_{j=1}^{n} |\mathscr{X}_j|^2 + \lambda_n)$ *and* $\hat{\beta}_n$*,* $\mathscr{X}_i$ *and* $\mathscr{Y}_i$ *are defined as in (3.3).*

3.5 Simulation study

The simulation study follows model (3.1) when the supports of all the involved functions are included in the interval $[0, T]$, for some fixed value $T > 0$. We want to illustrate the performance of the **FFDE**

and compare it with that of other existing estimators which we adapted for the FCVM (3.1). We present these competing techniques in Subsection 3.5.1.

We have chosen three different simulation settings in Subsection 3.5.2 to do this comparison. Each one uses different functions θ and X in order to see the strengths and weaknesses of the **FFDE** respect to the others. In Subsection 3.5.3 we present the simulation results and describe the performance of the estimators. We finish this section with a further discussion about the behavior of the **FFDE** and its expected performance.

The numerical implementation of the FFDE is postponed to Appendix 3.D.1.

3.5.1 Competing techniques

To the best of our knowledge there are few estimation techniques which could be adapted to estimate θ in the context of the FCVM, that is when i) the input and output are random functions, ii) the convolution is non-periodic, iii) the sample size is $n > 1$ and iv) the noise is functional. In what follows we describe how we have adapted these techniques.

Parametric Wiener Deconvolution (ParWD): This method belongs to the family of signal processing methods (see Gonzalez and Eddins (2009, Ch 5)). For each realization (X_i, Y_i), X_i is understood as the impulse response and Y_i as the observed signal. Then we use the **ParWD** to estimate the 'unknown signal' θ for each couple (X_i, Y_i) with $i \in \{1, \cdots, n\}$, as follows

$$\hat{\theta}_{wie,i} := \mathscr{F}^{-1}\left(\frac{\mathscr{F}(Y_i)\,\mathscr{F}(X_i)^*}{|\mathscr{F}(X_i)|^2 + \alpha}\right). \tag{3.10}$$

In this way we obtain n estimators of θ. Their mean will be the final estimator of θ in (3.1), that is $\hat{\theta}_{ParWD} := \frac{1}{n}\sum_{i=1}^{n} \hat{\theta}_{wie,i}$. Here we need to calibrate the parameter α, which is a constant number. We use the LOOPCV criteria to choose it. Note that the parameter α replaces the Noise-to-Signal power ratio $\mathbb{E}[|\mathscr{F}(\varepsilon)|^2]/|\mathscr{F}(\theta)|^2$ of the original Wiener deconvolution method (see Gonzalez and Eddins (2009, p. 241)).

Comparing definition (3.10) with the **FFDE** (3.4) we see that both are related; whereas for Wiener we start computing an estimator for each realization (X_i, Y_i) and then taking the mean, Fourier starts computing the mean in the frequency domain to estimate $\mathscr{F}(\theta)$ and then uses the IFCT to come back to the time domain. Due to this fact we will see that both estimators have a similar behavior.

Ill-posed linear inverse problems: The convolution can be understood as a special kind of matrix multiplication and consequently the deconvolution as a matrix inversion problem. We consider two well-known techniques to solve this matrix inversion problem, the **Singular Value Decomposition (SVD)** and the **Tikhonov regularization (Tik)** (see Tikhonov and Arsenin (1977)).

Given that the solution is highly sensitive to the noise (because this matrix inversion is an Hadamard ill-posed problem, see Tikhonov and Arsenin (1977, Ch 1)), we want to get rid of this

noise as much as possible before inverting the matrix. In this way we start by calculating the mean of all realizations and obtain for $t \in [0,T]$,

$$\bar{Y}(t) = \int_0^t \theta(s)\bar{X}(t-s)ds + \bar{\varepsilon}(t), \tag{3.11}$$

where $\bar{X} := \frac{1}{n}\sum_{i=1}^{n} X_i$ and $\bar{Y}$ and $\bar{\varepsilon}$ are defined in a similar way. Now the noise $\bar{\varepsilon}(t)$ is close to zero. Note that this procedure contrasts with the **ParWD** method, where the mean is computed in a second step because the **ParWD** was devised to deconvolve one signal at a time taking already the noise into account.

Next we compute the numerical matrix approximation of this integral equation by using the rectangular method over a uniform grid of observation times $t_0, \cdots, t_{p-1} \in [0,T]$. We obtain

$$\vec{Y} = M_X \vec{\theta} + \vec{\varepsilon},$$

where $\vec{Y} := (Y(t_0), \cdots, Y(t_{p-1}))'$, $\vec{\theta} := (\theta(t_0), \cdots, \theta(t_{p-1}))'$, $\vec{\varepsilon} := (\varepsilon(t_0), \cdots, \varepsilon(t_{p-1}))'$ and M_X is the corresponding lower triangular matrix which approximates the convolution (3.11) on these time steps, namely

$$M_X := \begin{pmatrix} \bar{X}(t_0) & 0 & 0 & \cdots & 0 \\ \bar{X}(t_1) & \bar{X}(t_0) & 0 & \cdots & 0 \\ \bar{X}(t_2) & \bar{X}(t_0) & \bar{X}(t_0) & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \bar{X}(t_{p-1}) & \bar{X}(t_{p-2}) & \bar{X}(t_{p-3}) & \cdots & \bar{X}(t_0) \end{pmatrix}.$$

We consider the SVD of M_X, that is $M_X = USV'$ where S is a diagonal matrix with the singular values of M_X (which are the square roots of the eigenvalues of $M_X'M_X$) and U and V are orthogonal matrices.

The Tikhonov estimator is defined as

$$\hat{\theta}_{Tik} := VS(S^2 + \rho I)^{-1} U'\vec{Y},$$

where ρ is a regularization parameter.

The SVD estimator is defined as

$$\hat{\theta}_{SVD} := VS_k^{+} U'\vec{Y},$$

where S_k is a diagonal matrix with the same first non-zero k diagonal elements as S and zero elsewhere, and S_k^{+} is the pseudo-inverse of S_k, which is obtained by replacing the non-zero elements of the diagonal of S_k by their reciprocals and then transposing the resulting matrix. Here the dimension k is the regularization parameter.

To calibrate the parameters of both estimators we do not use the LOOPCV but the 10-fold Predictive Cross Validation (see Seni and Elder (2010, Ch 3)) to avoid redundancy in calculations due to the use of the mean before inverting M_X in the first step of these two methods.

Laplace estimator (Lap): We use the adapted version of the Laplace estimator introduced by Comte et al. (2016), denoted here as $\hat{\theta}_{Lap}$. We start by calculating the mean of all realizations to eliminate the noise as much as possible since this estimator is designed to solve the problem when $n = 1$ (one couple of X and Y). Thus we obtain for $j = 1, \cdots, p-1$,

$$\bar{Y}(t_j) = \int_0^{t_i} \theta(s)\bar{X}(t_j - s)ds + \bar{\varepsilon}(t_j) \tag{3.12}$$

where $t_0, \cdots, t_{p-1} \in \mathbb{R}$ are the observation times.

In Comte et al. (2016) this equation is interpreted as a discrete noisy version of the linear Volterra equation of the first kind, where the goal is to estimate θ. More precisely the authors use a model where $\bar{\varepsilon}(t_i)$ are i.i.d sub-Gaussian random variables such that $\mathbb{E}[\bar{\varepsilon}(t_i)] = 0$ and $\mathbb{E}[|\bar{\varepsilon}(t_i)|^2] = \sigma^2$.

To estimate θ, the authors use the Laguerre functions, defined for $k \in \mathbb{N}$, $t \geq 0$ and some fixed $a > 0$ as follows

$$\phi_k(t) := \sqrt{2a}e^{-at}\left(\sum_{j=0}^{k}(-1)^j\binom{k}{j}\frac{t^j}{j!}\right).$$

First they use these functions as an orthonormal basis of $L_2(\mathbb{R}_+, \mathbb{R})$ to transform equation (3.12) into an infinite system of linear equations with coefficients obtained from the expansion in the Laguerre basis. They chose the Laguerre functions because the convolution of a couple of these functions is easy to obtain, and satisfies that for $k, l \geq 0$,

$$\int_0^t \phi_k(s)\phi_l(t-s)ds = (2a)^{-1/2}[\phi_{k+l}(t) - \phi_{k+l+1}(t)].$$

Thanks to this fact the latter system is simplified and becomes an infinite lower triangular system of linear equations. Next they solve the finite subsystem of the first M linear equations to compute the estimators of the first M coefficients of θ on the Laguerre basis. The numerical computation of their estimator is done with the **R** package **LaplaceDeconv**. In order to avoid numerical instability, due to the computation of Laguerre functions in **R**, we resize the curves from $[0, T]$ to the interval $[0, 10]$ (stretching the curves) but keeping the SNR equal to 10. In this way to estimate θ we use the initial curves X_i and Y_i stretched to $[0, 10]$ together with the noise with standard deviation equal to σ/n. After computing the Laplace estimator with this data we multiply this one by $10/T$ to resize it. Notice that the true value of σ is necessary to compute $\hat{\theta}_{lap}$ both theoretically (the authors use it to penalize the estimator during the calibration of parameters) and numerically.

Remark : In practice after computing all the estimators defined in this section and the FFDE we have used the spline smoothing method to smooth all of them. This step improves their estimation performance.

3.5.2 Settings

We compared these estimation procedures in three different simulation settings. The goal is to compare how well the **FFDE** estimates θ with respect to the performance of the others. In the first setting the X variable is such that $\mathbb{E}[X] = 0$ which is a situation where the estimation is more difficult, in particular for **SVD** and **Tik**, because they need to invert the associated matrix M_X (see definition of the **SVD** estimator). The second setting uses $\mathbb{E}[X] \neq 0$ and here the inversion of M_X is numerically more stable. In this setting the shape of θ has some periodicity, thus one goal is to asses how well the methods can estimate this periodicity and another is to experience **FFDE** under favorable conditions for **SVD** and **Tik**. The last setting uses θ and X which are well represented with the Laguerre functions. This is a favorable condition for the Laplace estimator (**Lap**). We want to see how the others perform under this condition.

Let us detail each setting. For settings 1 and 2 the data were simulated on the interval $[0,1]$ ($T = 1$), discretized over $p = 100$ equispaced observation times, $t_j := j/100$, with $j = 0, \cdots, 99$. Whereas for Setting 3 the interval is $[0,8]$ ($T = 8$), with $p = 100$ equispaced observation times $t_j := 8j/100$, for $j = 0, \cdots, 99$.

In the Table 3.1 we describe the curves X_i and the functions θ for each setting. In that table BB_i stands for the Brownian Bridge on the interval $[0, 0.5]$ with the process pinned at the origin at both $t = 0$ and $t = 0.5$, for every $i = 1, \cdots, n$. On the other hand for settings 1 and 2 we use $\mathbf{1}_{[0,0.5]}$, the indicator function of the interval $[0, 0.5]$, because we want the support of Y to be $[0,1]$ given that $supp(Y) = supp(X) + supp(\theta)$. In contrast to those settings, in setting 3 the $supp(Y)$ is bigger than $[0,8]$, however the estimation with **FFDE** is still possible due to the fact that the values of $Y(t)$ for $t > 8$ are relatively small compared to the values for $t \in [0,8]$. Note that in general the condition $supp(X) + supp(\theta) = supp(Y) \subseteq [0,T]$ is necessary to compute numerically the **FFDE**. Indeed, in this case the CFT can properly transform the convolution between X and θ into a multiplication in the frequency domain.

For all these settings the noise ε is the White Gaussian Noise defined with a standard deviation σ (σ is constant and for every $t \in [0,T]$, $\sigma^2 = \mathbb{E}[|\varepsilon(t)|^2]$) chosen for each setting such that the Signal-to-Noise-Ratio (SNR) is equal to 10 (interpreted as 10% of noise). Here the SNR is defined as $SNR := \mathbb{E}[\|\theta * X\|^2_{L^2}]/\sigma^2$. Note also that for each setting we have numerically verified that the general hypotheses ($HA1_{FCVM}$) - ($HA3_{FCVM}$) are satisfied.

We evaluate our estimation procedure for sample of sizes $n = 70$ and $n = 400$. Additionally we use the two following criteria to measure the estimation error.

Setting	Curves X_i	Function θ
1	$BB_i(t)\mathbf{1}_{[0,0.5]}(t)$	$(1-4t^2)\mathbf{1}_{[0,0.5]}(t)$
2	$[\frac{1}{2}-8\,(t-\frac{1}{4})^2+\frac{1}{4}\,BB_i(t)]\;\mathbf{1}_{[0,0.5]}(t)$	$[\frac{1}{4}\,\sin(6\pi t)+\frac{3}{4}-\frac{3}{2}\,(t)]\;\mathbf{1}_{[0,0.5]}(t)$
3	$20t^2e^{-3t}+\frac{1}{2}\,BB_i(t/8)\,\mathbf{1}_{[0,4]}(t)$	$(2t+1)e^{-2t}$

Table 3.1 Curves X_i and functions θ for each simulation setting.

Evaluation criteria: We use 100 Monte Carlo runs to evaluate for each simulated sample the mean absolute deviation error (MADE) and the weighted average squared error (WASE) as defined in Şentürk and Müller (2010, p. 1261),

$$MADE := \frac{1}{T}\left[\frac{\int_0^T |\theta(t)-\hat{\theta}(t)|dt}{range(\theta)}\right], \qquad WASE := \frac{1}{T}\left[\frac{\int_0^T |\theta(t)-\hat{\theta}(t)|^2dt}{range^2(\theta)}\right],$$

where $range(\theta)$ is the range of the function θ.

3.5.3 Simulation Results

All the computations have been implemented in **R** on a 2.9 GHz x 4 Intel Core i7-3520M processor, with a 4000KB cache size and 8GB total physical memory. Thanks to Proposition 18, it is possible to compute the **FFDE** with optimized parameter quickly. For the other estimators we have optimized numerically their respective parameters. The computation times are shown in Table 3.2, where we see that **FFDE** outperforms the others.

Setting	size	FFDE	ParWD	SVD	Tik	Lap
1	n=70	**0.15203**	10.09670	9.58702	3.73524	3.07057
	n=400	**0.70470**	225.3133	14.46868	6.10703	4.59315
2	n=70	**0.23510**	105.0812	8.82455	3.72090	4.85827
	n=400	**0.74490**	218.0906	11.1811	5.32442	4.91993
3	n=70	**0.24029**	7.78316	8.97592	3.20933	5.53914
	n=400	**0.71429**	173.6131	12.33914	4.96589	6.47220

Table 3.2 Computation time (in seconds) of the estimators for a given sample and setting.

Now we discuss the estimation performance for each setting separately because they have been chosen to assess various properties of the **FFDE** under different situations.

Setting 1 : First Figure 3.1 shows the true function θ and the cross-sectional mean curves of its five estimators computed from $N = 100$ simulations. The best estimators are **FFDE** and **ParWD**, both of them are close to each other. Note that **FFDE** have difficulty to estimate θ close to the borders. **SVD** and **Tik** are wavy, whereas **Lap** estimates poorly the quadratic part of θ over the subinterval $[0, 0.3]$. Finally all the estimators except **Lap** show an improvement when the sample size increases to $n = 400$, in particular **FFDE** improves considerably.

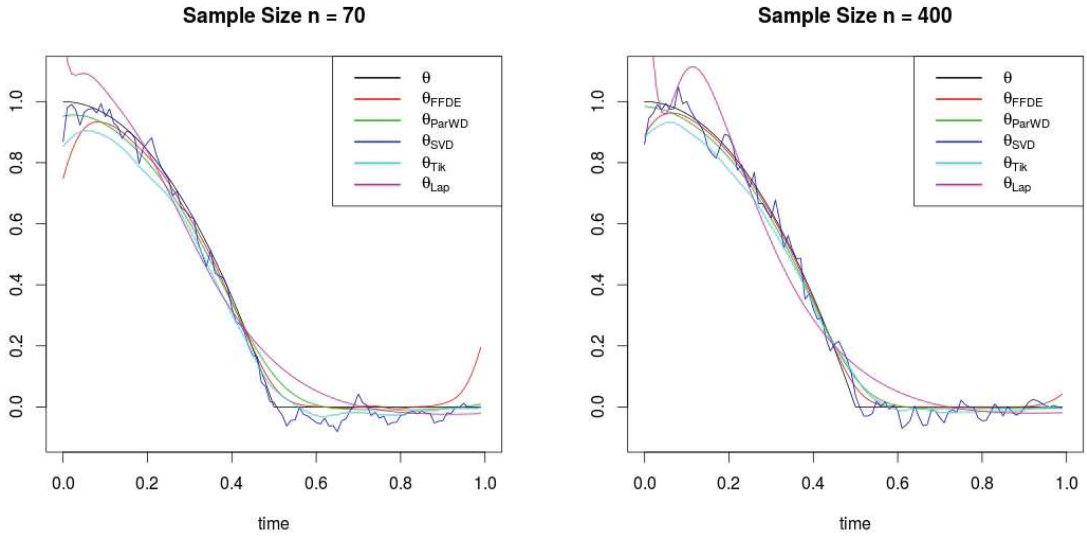

Fig. 3.1 The true function θ (black) compared to the cross-sectional mean curves of the five estimators.

More specifically we can see in Table 3.3 and in the box plots in Figure 3.2 that **FFDE** and **ParWD** are the best estimators, whereas **SVD** is the worst of all of them. When the sample size increases to $n = 400$, **FFDE** is the one which has improved the most.

In this setting **FFDE** and **ParWD** handle well the case where $\mathbb{E}[X] = 0$, because they use the Fast Fourier Algorithm (FFT) to directly deconvolve the convolution of X and θ, whereas **SVD** and **Tik** perform badly because they cannot properly invert the matrix M_X used in their definitions. Besides note that **Lap** does not improve the estimation because we apply it to the mean equation (3.12), which is almost the same when $n = 70$ and $n = 400$, this fact will also be true for Settings 2 and 3. Finally although **SVD** and **Tik** use the mean equation (3.12), they slightly improve due to the strong dependency of the inversion of M_X on the noise.

Setting 2 : Figure 3.3 shows the true function θ and the cross-sectional mean curves of the five estimators. The best estimators are **SVD** and **Tik**, both of them behave similarly. **FFDE** gives a better

	MADE	WASE
n = 70	mean (sd)	mean (sd)
FFDE	**0.04120 (0.00895)**	**0.00400 (0.00212)**
ParWD	**0.03020 (0.00657)**	**0.00157 (0.00062)**
SVD	0.16240 (0.15467)	0.08356 (0.16906)
Tik	0.08797 (0.04836)	0.01573 (0.01764)
Lap	0.16427 (0.11468)	0.10468 (0.30549)
n = 400	mean (sd)	mean (sd)
FFDE	**0.01273 (0.00151)**	**0.00044 (0.00013)**
ParWD	0.02010 (0.00342)	0.00076 (0.00024)
SVD	0.16313 (0.18566)	0.10284 (0.27954)
Tik	0.07641 (0.03702)	0.01120 (0.01170)
Lap	0.18968 (0.13129)	0.15172 (0.28369)

Table 3.3 Mean and standard deviation (sd) of the two criteria, computed from $N = 100$ simulations with sample sizes $n = 70$ and $n = 400$.

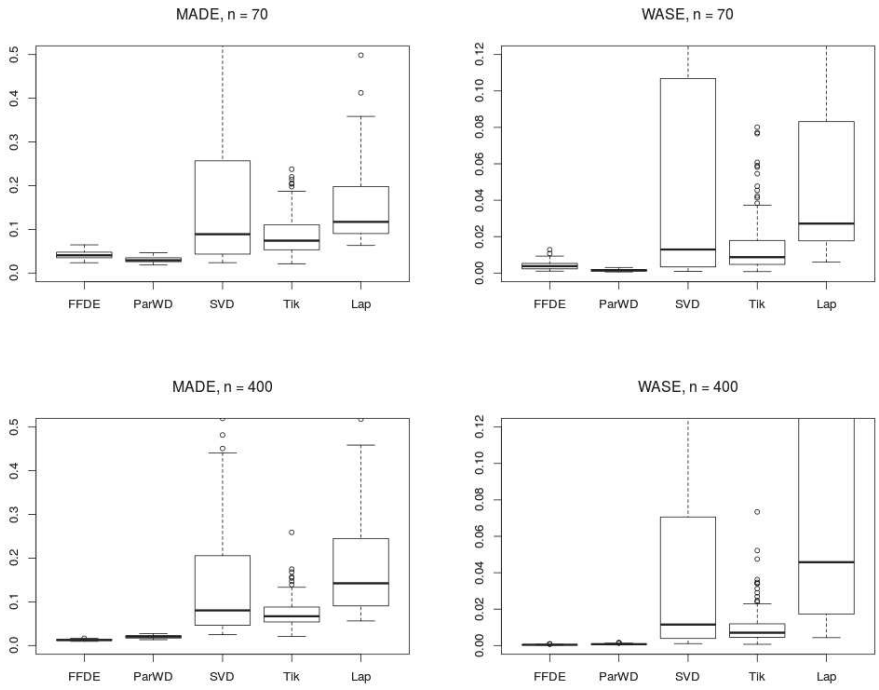

Fig. 3.2 Boxplots of the two criteria over $N = 100$ simulations with sample sizes $n = 70$ and $n = 400$.

estimation than **ParWD**. Note that **FFDE** again have problems to estimate θ close to the borders. On the other hand **Lap** cannot estimate the 'periodic' shape of the curve on the interval $[0.2, 0.7]$. There is a slight improvement on the estimators when the sample size increases to $n = 400$.

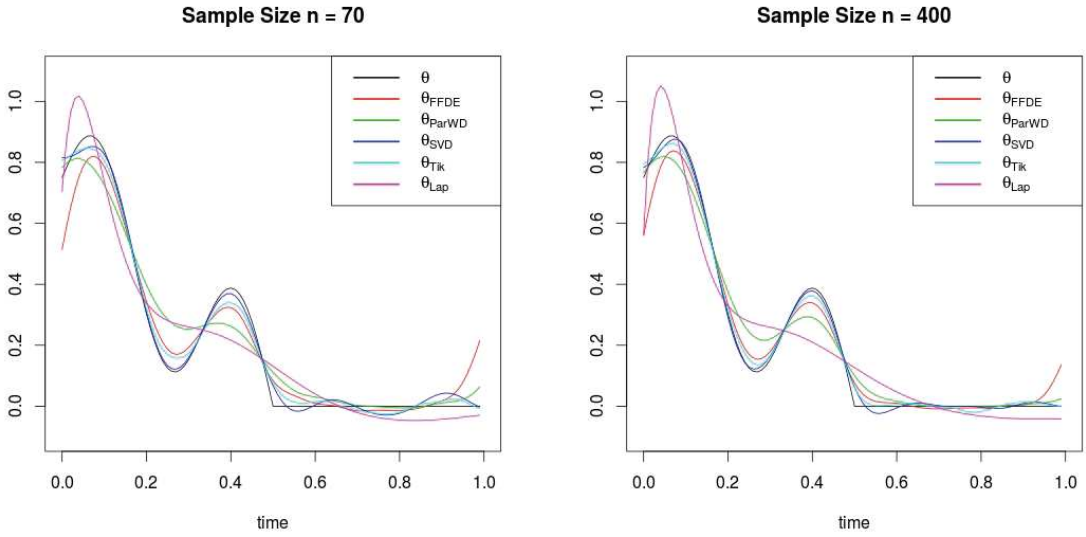

Fig. 3.3 Function θ (black) and the cross-sectional mean curves of the five estimators.

Table 3.4 and the box plots in Figure 3.4 show that **SVD** outperforms the others, but all of them except **Lap** are almost as good. In particular **FFDE** and **Tik** give estimations close to **SVD**. On the other hand **ParWD** is the most scattered one although it roughly behaves like **FFDE**. When the sample size increases to $n = 400$ there is an improvement, **SVD** being the one which improves the most. However **FFDE** and **Tik** remain quite close to **SVD**. **ParWD** is the most scattered estimator.

In this setting **SVD** and **Tik** perform better than the other ones, because this time the matrix M_X is not close to zero and is easier to invert. However **FFDE** and **ParWD** are quite good, this shows that the use of FFT by **FFDE** and **ParWD** is stable in both cases whether $\mathbb{E}[X] = 0$ or not. Furthermore when the sample size increases, **FFDE** is almost as good as **SVD**.

Setting 3: Figure 3.5 shows the function θ and the cross-sectional mean curves of the five estimators. In contrast to Settings 1 and 2, the best estimator is **Lap**, whereas the others perform quite similarly. Again **FFDE** has difficulties to estimate θ close to the borders. Finally all the estimators except **Lap** improve when the sample size increases to $n = 400$. Moreover all of them become better than **Lap**, and **SVD** gives the best estimation.

In Table 3.5 and in the boxplots of Figure 3.6 we can see that **Lap** outperforms the others when $n = 70$. The others give equivalent estimations. **FFDE** has a bigger variation for the WASE criteria. In the case where the sample size is $n = 400$ we obtain an improvement in the estimation, **SVD** being the one improving the most.

	MADE	WASE
n = 70	mean (sd)	mean (sd)
FFDE	0.05913 (0.01074)	0.00670 (0.00245)
ParWD	0.07282 (0.01770)	0.00987 (0.00430)
SVD	**0.04960 (0.01512)**	**0.00402 (0.00284)**
Tik	0.05112 (0.01142)	0.00426 (0.00214)
Lap	0.09178 (0.01830)	0.01472 (0.00616)
n = 400	mean (sd)	mean (sd)
FFDE	0.03754 (0.00636)	0.00283 (0.00108)
ParWD	0.05365 (0.01923)	0.00579 (0.00410)
SVD	**0.02498 (0.00936)**	**0.00010 (0.00100)**
Tik	0.03125 (0.00656)	0.00157 (0.00068)
Lap	0.08690 (0.01047)	0.01257 (0.00324)

Table 3.4 Mean and standard deviation (sd) of the two criteria, computed from $N = 100$ simulations with sample sizes $n = 70$ and $n = 400$.

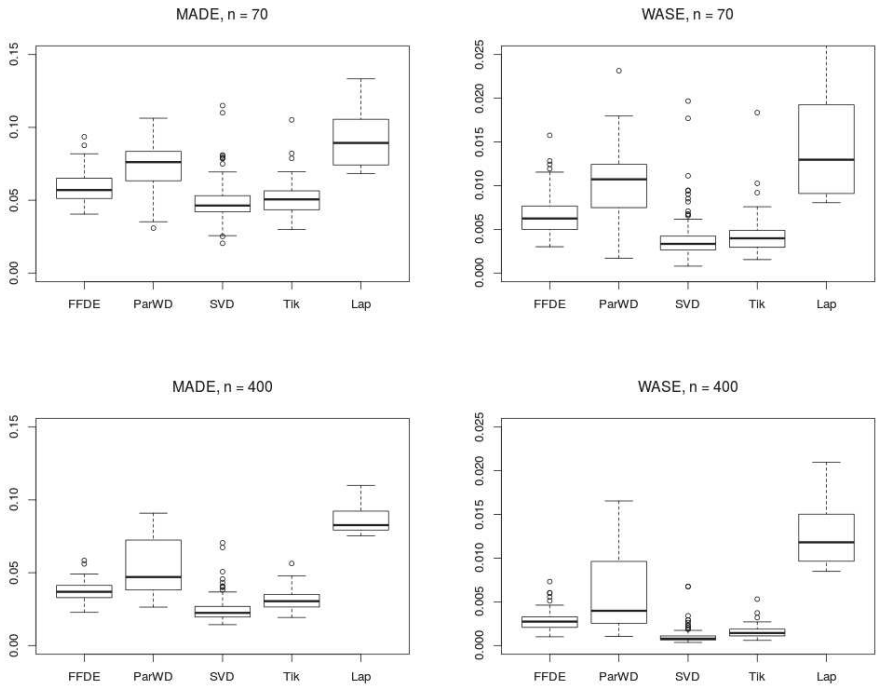

Fig. 3.4 Boxplots of the two criteria over $N = 100$ simulations with sample sizes $n = 70$ and $n = 400$.

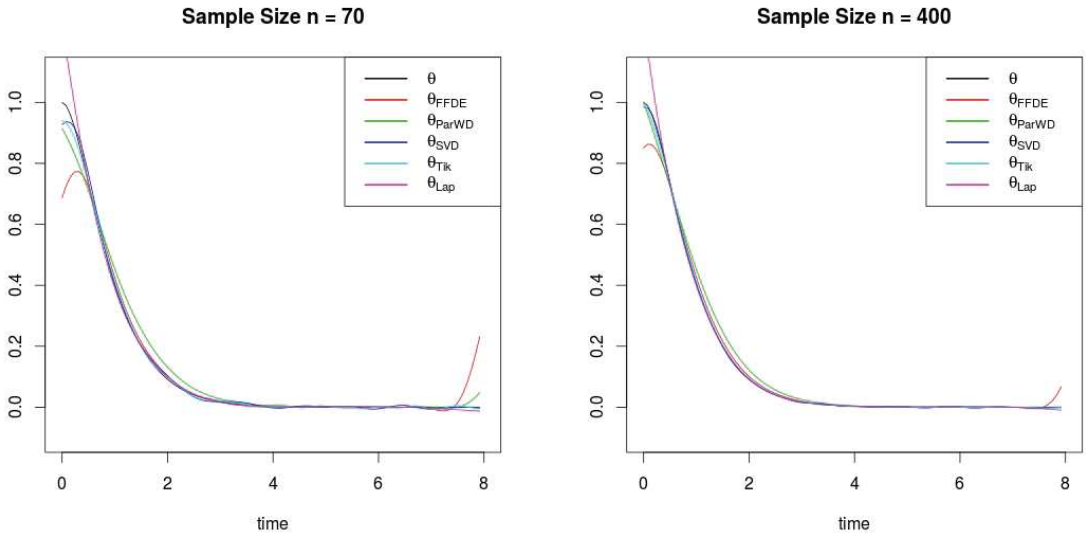

Fig. 3.5 The function θ (black) and the cross-sectional mean curves of the five estimators.

	MADE	WASE
n = 70	mean (sd)	mean (sd)
FFDE	0.00401 (0.00092)	0.00045 (2e-04)
ParWD	0.00303 (0.00114)	0.00021 (0.00021)
SVD	0.00336 (0.0015)	0.00017 (0.00017)
Tik	0.00387 (0.00083)	2e-04 (8e-05)
Lap	**0.00168 (0.00104)**	**0.00012 (0.00013)**
n = 400	mean (sd)	mean (sd)
FFDE	0.00134(0.00017)	7e-05(3e-05)
ParWD	0.00176(0.00038)	7e-05(3e-05)
SVD	**0.00111(0.00056)**	**2e-05(3e-05)**
Tik	0.00095(0.00018)	1e-05(1e-05)
Lap	0.00143(0.00071)	9e-05(6e-05)

Table 3.5 Mean and standard deviation (sd) of the two criteria, computed from $N = 100$ simulations with sample size $n = 70$ and $n = 400$.

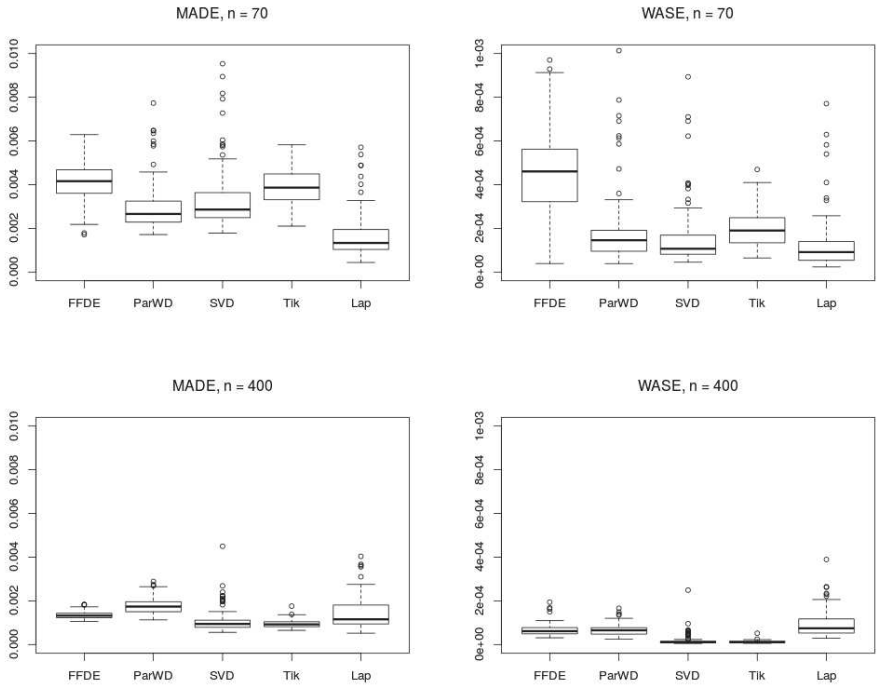

Fig. 3.6 Boxplots of the two criteria over $N = 100$ simulations with sample sizes $n = 70$ and $n = 400$.

In this setting **Lap** performs better than the others because both X and θ are functions well represented with the Laguerre functions. However all the other estimators show a great improvement when $n = 400$. This shows that **SVD** and **Tik** give good estimations as long as $\mathbb{E}[X] \neq 0$. Finally **FFDE** is almost as good as **SVD**.

3.5.4 A further discussion about FFDE

In each of the three settings we have seen that the **FFDE** performed well with very fast computation time and convergence towards θ as the sample size increases. It gives a good estimation in these three settings, even in the disadvantageous case where $\mathbb{E}[X] = 0$ and thus the noise plays a major role.

We note an edge effect for small sample sizes that decreases as n goes to infinity. This effect comes from the second component of the decomposition of $\hat{\theta}_n$ derived from (3.5), namely

$$\mathscr{F}^{-1}(\Psi_n) := -\frac{\lambda_n}{n}\,\mathscr{F}^{-1}\left[\frac{\beta}{\frac{1}{n}\sum_{i=1}^{n}|\mathscr{X}_i|^2 + \frac{\lambda_n}{n}}\right].$$

In Figure 3.7 we see the $\mathscr{F}^{-1}(\Psi_n)$ components for each of the three settings. One of the reasons of this shape is that $\Phi := \mathbb{E}[|\mathscr{X}|^2]$ (denominator) is highly concentrated on the borders. This is shown in Figure 3.8, where for each setting we approximate Φ by the empirical mean with $n = 7000$. Note

that all these functions are positive over the whole interval despite what might be assumed from Figure 3.8.

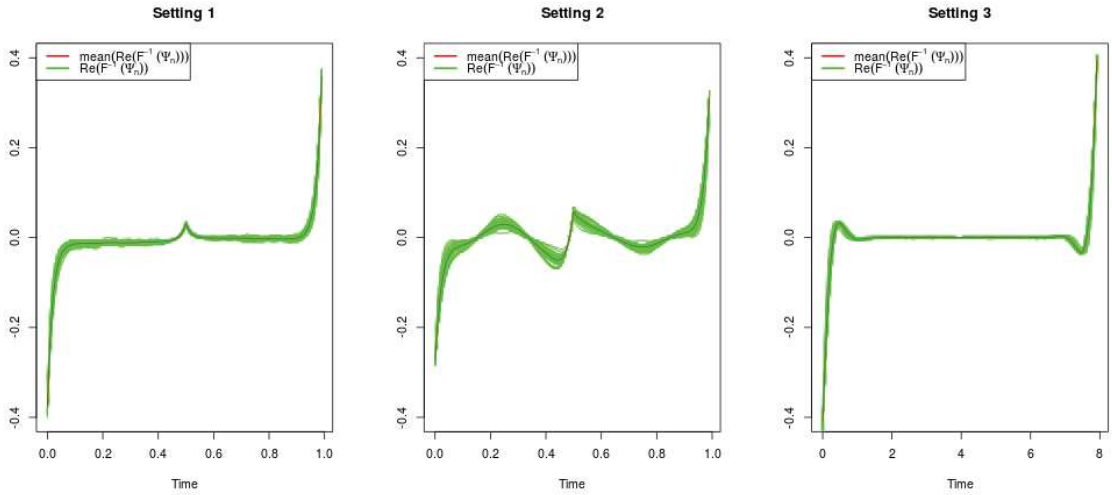

Fig. 3.7 The real part of the function $\mathscr{F}^{-1}(\Psi_n)$ (the imaginary part is equal to constant zero) for setting 1 to 3. In green 50 examples of $\mathscr{F}^{-1}(\Psi_n)$ computed for samples of size $n = 70$. In red the cross-sectional mean in each case.

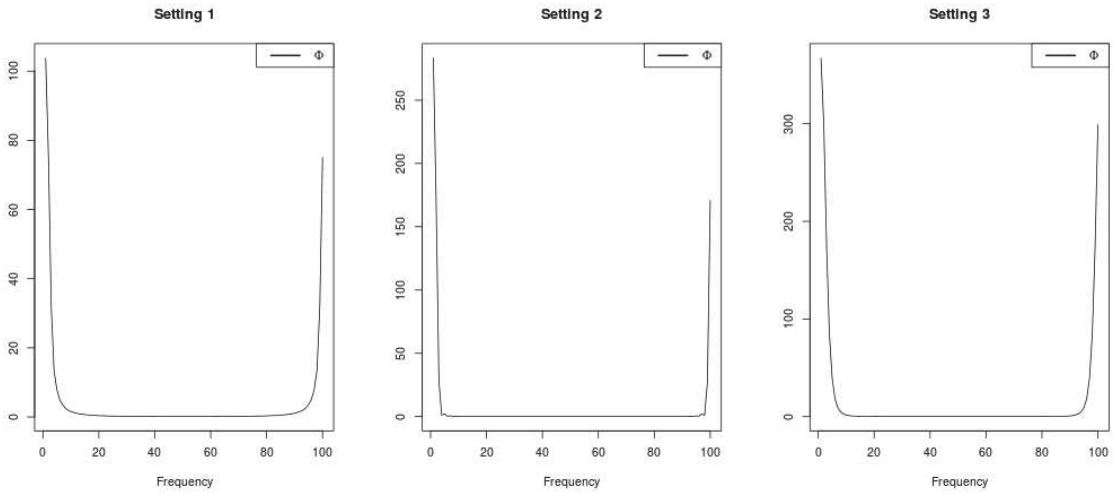

Fig. 3.8 The plots of the function Φ for setting 1 to 3.

From these reasons the difference $\hat{\theta}_n - \theta$ will have higher values close to the borders (edge effect) since $\hat{\theta}_n - \theta \approx \mathscr{F}^{-1}(\Psi_n)$. Note that when n increases we have $\lambda_n/n \to 0$ and thus $\Psi_n \to 0$, so the edge effect will decrease. This fact is observed in the simulation studies when we increase the sample size to $n = 400$.

Whenever $\mathbb{E}[|\mathscr{X}|^2]$ has higher values close to the borders the edge effect should be expected. In that case we propose the practical solution of using the estimation over an appropriate interval before the borders to extrapolate the estimation in the borders. The results of this method are shown in Figure 3.9. We took 10% before the borders (last 5% in each side) to extrapolate over them, we did this for each one of the 100 realizations. The cross-sectional mean of the FFDE estimator before and after removing the edge effect are in green and in red respectively.

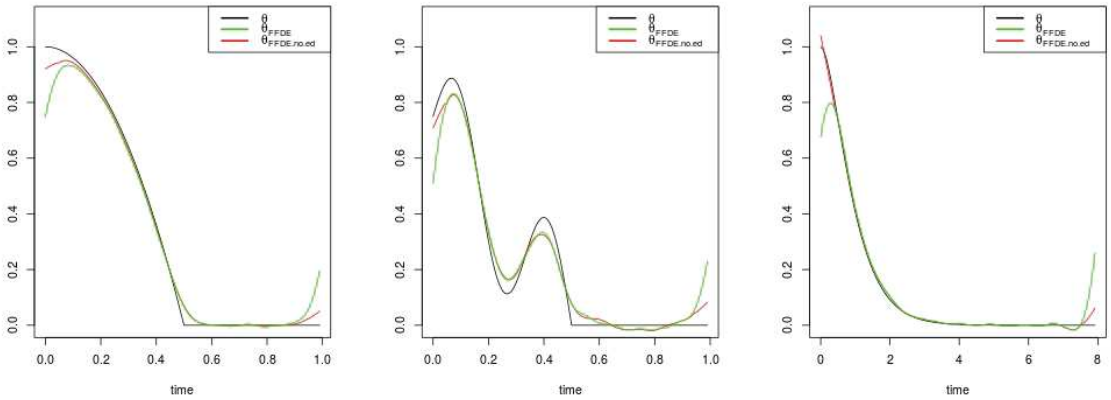

Fig. 3.9 Estimators of θ for each setting. The cross-sectional mean of the FFDE estimator before and after removing the edge effect are the curves in green and in red respectively.

The boxplots of the MADE and WASE criteria before (FFDE) and after removing the edge effect (FFDE.no.ed) are shown in Figure 3.10. We see there that a major improvement in the estimation is done in the setting 1, whereas in settings 2 and 3 this improvement is small and WASE is changing the most.

3.6 Conclusions

In this paper we have defined the **FFDE** for the FCVM. We proved its consistency for the L^2-norm and obtained a rate of convergence. We also provided a selection procedure of the regularization parameter λ_n through the LOOPCV criterion. The simulations showed the robustness of the **FFDE** despite some irregularities on the borders (edge effect). This effect can be reduced by using the estimation over an appropriate interval before these borders.

Compared to other estimation methods adapted from the literature, **FFDE** is almost as good as the best estimator in all the three settings and always with the fastest computation time.

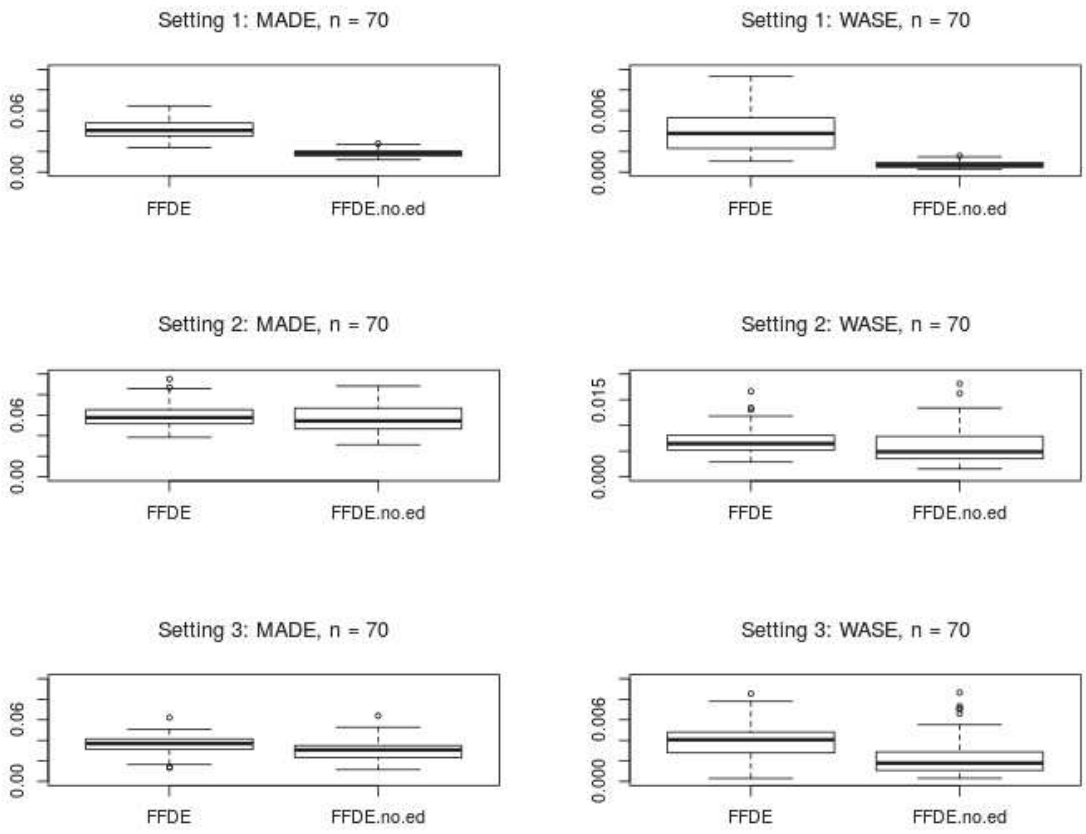

Fig. 3.10 Boxplots of MADE and WASE criteria before (FFDE) and after removing the edge effect (FFDE.no.ed) respectively.

3.7 Acknowledgments

The authors would like to thank the Labex NUMEV (convention ANR-10-LABX-20) for partly funding the PhD thesis of Tito Manrique (under project 2013-1-007).

Appendix

3.A Main Theorems of Manrique et al. (2016)

The general hypotheses used in Manrique et al. (2016) and the results are rewritten with the notation of the associated concurrent model (3.2) to avoid confusion. The general hypotheses are:

- $(HA1_{FCM})$ $\mathscr{X}, \varepsilon$ are independent $C_0 \cap L^2$ valued random functions, such that $\mathbb{E}(\varepsilon) = \mathbb{E}(\mathscr{X}) = 0$,
- $(HA2_{FCM})$ $\beta \in C_0 \cap L^2$,
- $(HA3_{FCM})$ $\mathbb{E}(\|\varepsilon\|^2_{C_0}), \mathbb{E}(\|\mathscr{X}\|^2_{C_0}), \mathbb{E}(\|\varepsilon\|^2_{L^2})$ and $\mathbb{E}(\|\mathscr{X}\|^2_{L^2})$ are all finite.

The main results of Manrique et al. (2016) used in this paper are presented next.

Theorem 19 (Theorem 3.1 in Manrique et al. (2016))**.** *Let us consider the FCM with the general hypotheses ($HA1_{FCM}$), ($HA2_{FCM}$) and ($HA3_{FCM}$). Let $(\mathscr{X}_i, \mathscr{Y}_i)_{i\geq 1}$ be i.i.d. realizations. We suppose moreover that*

(A1) $\overline{supp(|\beta|)} \subseteq \overline{supp(\mathbb{E}[|\mathscr{X}|])}$,

(A2) $(\lambda_n)_{n\geq 1} \subset \mathbb{R}^+$ *is such that* $\frac{\lambda_n}{n} \to 0$ *and* $\frac{\sqrt{n}}{\lambda_n} \to 0$ *as* $n \to +\infty$.

Then

$$\lim_{n\to+\infty} \|\hat{\beta}_n - \beta\|_{L^2} = 0 \quad \textit{in probability.} \tag{3.13}$$

Corollary 20 (Corollary 3.7)**.** *Under hypotheses* $(A1)$, $(A2)$ *and if additionally we assume*

(A3) $\mathbb{E}[\||\mathscr{X}|^2\|^2_{L^2}] < \infty$,

(A4bis) $\frac{|\beta|}{\mathbb{E}[|\mathscr{X}|^2]} 1_{supp(|\beta|)} \in L^2 \cap L^\infty$,

then

$$\|\hat{\beta}_n - \beta\|_{L^2} = O_P\left(\max\left[\frac{\lambda_n}{n}, \frac{\sqrt{n}}{\lambda_n}\right]\right). \tag{3.14}$$

Theorem 21 (Theorem 3.8)**.** *Under hypotheses* $(A1)$, $(A2)$ *and* $(A3)$, *for every compact subset* $K \subset supp(\mathbb{E}[|\mathscr{X}|^2])$, *we have*

$$\|\hat{\beta}_n - \beta\|_{L^2(K)} = O_P\left(\max\left[\frac{\lambda_n}{n}, \frac{\sqrt{n}}{\lambda_n}\right]\right). \tag{3.15}$$

Proposition 22 (Proposition 4.1)**.** *We have*

$$PCV(\lambda_n) = \frac{1}{n}\sum_{i=1}^{n}\left\| \frac{\mathscr{Y}_i - \hat{\beta}_n\,\mathscr{X}_i}{1-A_{i,i}} \right\|_{L^2}^2, \tag{3.16}$$

where $A_{i,i} \in L^2$ is defined as follows $A_{i,i} := |\mathscr{X}_i|^2/(\sum_{j=1}^{n}|\mathscr{X}_j|^2 + \lambda_n)$.

3.B Proofs

Throughout these proofs we use the notation of the associated functional concurrent model (3.2).

Proof of Theorem 15. We use a modified version of Theorem 3.1 of Manrique et al. (2016) to prove Theorem 15 in this paper. In order to do this let us recall the three general hypotheses used to prove Theorem 3.1 of Manrique et al. (2016) rewritten with the notations of (3.2).

- $(HA1_{FCM})$ $\mathscr{X}, \varepsilon$ are independent $C_0 \cap L^2$ valued random functions, such that $\mathbb{E}(\varepsilon) = \mathbb{E}(\mathscr{X}) = 0$,
- $(HA2_{FCM})$ $\beta \in C_0 \cap L^2$,
- $(HA3_{FCM})$ $\mathbb{E}(\|\varepsilon\|_{C_0}^2), \mathbb{E}(\|\mathscr{X}\|_{C_0}^2)$, $\mathbb{E}(\|\varepsilon\|_{L^2}^2)$ and $\mathbb{E}(\|\mathscr{X}\|_{L^2}^2)$ are all finite.

Given that we are interested in a more general version of Theorem 3.1, we will change $(HA1_{FCM})$ for $(HA1bis_{FCM})$ defined as follows

- $(HA1bis_{FCM})$ $\mathscr{X}, \varepsilon$ are independent $C_0 \cap L^2$ valued random functions, such that $\mathbb{E}(\varepsilon) = 0$.

Our goal is to prove that $(HA1bis_{FCM})$, $(HA2_{FCM})$ and $(HA3_{FCM})$ are implied by the general hypotheses of the FCVM (see subsection 3.2.1), and then to prove a generalization of Theorem 3.1 of Manrique et al. (2016) with $(HA1bis_{FCM})$ instead of $(HA1_{FCM})$.

First we show that the hypotheses $(HA1bis_{FCM})$ and $(HA2_{FCM})$ are satisfied. Given that $\theta \in L^1$, then $\beta \in C_0(\mathbb{R},\mathbb{C})$ (see Pinsky (2002, Ch. 2)). Moreover since $\mathscr{F}$ is an isometry in L^2 we obtain $\beta \in L^2(\mathbb{R},\mathbb{C})$. Thus hypothesis $(HA2_{FCM})$ holds. In a similar way we prove that $\mathscr{X}, \varepsilon \in C_0(\mathbb{R},\mathbb{C}) \cap L^2(\mathbb{R},\mathbb{C})$. The linearity of $\mathscr{F}$ implies $\mathbb{E}[\mathscr{F}(\varepsilon)] = 0$ so $(HA1bis_{FCM})$ holds too.

We use the contraction property of $\mathscr{F}$, namely $\|\mathscr{F}(f)\|_{C_0} \leq \|f\|_{L^1}$ (see Pinsky (2002, Ch. 2)) and again the fact that $\mathscr{F}$ is an isometry to prove that $(HA3_{FCM})$ holds.

Next we outline the proof of a generalization of Theorem 3.1 (see Theorem 19 in Appendix 3.A), in which we use $(HA1bis_{FCM})$ instead of $(HA1_{FCM})$. First we need to prove

$$\left\| \frac{\frac{1}{n}\sum_{i=1}^{n}\varepsilon_i \mathscr{X}_i^*}{\frac{1}{n}\sum_{i=1}^{n}|\mathscr{X}_i|^2 + \frac{\lambda_n}{n}} \right\|_{L^2} = O_P\left(\frac{\sqrt{n}}{\lambda_n}\right), \tag{3.17}$$

which helps us to bound the second term of $\|\hat{\beta}_n - \beta\|_{L^2}$ in the decomposition (3.5).

Let us prove 3.17. We have

$$\mathbb{E}[\|\varepsilon\,\mathscr{X}^*\|_{L^2}^2] \leq \mathbb{E}[\|\varepsilon\|_{C_0}^2]\,\mathbb{E}[\|\mathscr{X}\|_{L^2}^2] < \infty,$$

because of $(HA1bis_{FCM})$ and $(HA3_{FCM})$.

Now due to the moment monotonicity $\mathbb{E}[\|\varepsilon\,\mathscr{X}^*\|_{L^2}] < \infty$, $\varepsilon\,\mathscr{X}^*$ is strongly integrable with the L^2-norm, so there exists the expectation $\mathbb{E}[\varepsilon\,\mathscr{X}^*] \in L^2$ which is the zero function because $\mathbb{E}[\varepsilon] = 0$. We conclude that $\mathbb{E}[\varepsilon\,\mathscr{X}^*] = 0$ and $\mathbb{E}[\|\varepsilon\,\mathscr{X}^*\|_{L^2}^2] < \infty$ which, from the CLT in L^2 (see Theorem 2.7 in Bosq (2000, p. 51) and Ledoux and Talagrand (1991, p. 276) for the rate of convergence), yields to

$$\left\| \frac{1}{n}\sum_{i=1}^{n} \varepsilon_i \mathscr{X}_i^* \right\|_{L^2} = O_P\left(\frac{1}{\sqrt{n}}\right).$$

Finally (3.17) is obtained from the fact that

$$\left\| \frac{\frac{1}{n}\sum_{i=1}^{n} \varepsilon_i X_i^*}{\frac{1}{n}\sum_{i=1}^{n} |\mathscr{X}_i|^2 + \frac{\lambda_n}{n}} \right\|_{L^2} \leq \left| \frac{n}{\lambda_n} \right| \left\| \frac{1}{n}\sum_{i=1}^{n} \varepsilon_i \mathscr{X}_i^* \right\|_{L^2} = O_P\left(\frac{\sqrt{n}}{\lambda_n}\right).$$

Notice that hypotheses $(A1)$ and $(A2)$ of Theorem 3.1 of Manrique et al. (2016) (Theorem 19 in Appendix 3.A) are implied by hypotheses $(A1)$ and $(A2)$ of Theorem 15, and the normed functions in (3.17) converge in probability to zero.

Finally with the same argument as in the proof of Theorem 3.1 of Manrique et al. (2016) (Theorem 19 in Appendix 3.A) it is possible to prove

$$\left| \frac{\lambda_n}{n} \right| \left\| \frac{\beta}{\frac{1}{n}\sum_{i=1}^{n} |X_i|^2 + \frac{\lambda_n}{n}} \right\|_{L^2} \xrightarrow{a.s.} 0, \tag{3.18}$$

and thus the triangular inequality applied to the decomposition (3.5) implies that $\|\hat{\beta}_n - \beta\|_{L^2}$ goes to zero in probability. □

Proof of Theorem 16. This is a direct consequence of Corollary 3.7 of Manrique et al. (2016) because hypotheses $(A3)$ and $(A4bis)$ of this corollary are consequences of $(A3)$ and $(A4)$ in Theorem 16. □

Proof of Theorem 17. We start with the triangle inequality applied to (3.5) but restricted to the compact subset K,

$$\|\hat{\beta}_n - \beta\|_{L^2(K)} \leq \left| \frac{\lambda_n}{n} \right| \left\| \frac{\beta}{\frac{1}{n}\sum_{i=1}^{n} |\mathscr{X}_i|^2 + \frac{\lambda_n}{n}} \right\|_{L^2(K)} + \left\| \frac{\frac{1}{n}\sum_{i=1}^{n} \varepsilon_i \mathscr{X}_i^*}{\frac{1}{n}\sum_{i=1}^{n} |\mathscr{X}_i|^2 + \frac{\lambda_n}{n}} \right\|_{L^2(K)}.$$

The proof of $\left\| \frac{\frac{1}{n}\sum_{i=1}^n \varepsilon_i \mathscr{X}_i^*}{\frac{1}{n}\sum_{i=1}^n |\mathscr{X}_i|^2 + \frac{\lambda_n}{n}} \right\|_{L^2(K)} = O_P(\frac{\sqrt{n}}{\lambda_n})$ is the same as in Theorem 15. To finish the proof of this theorem we prove

$$\left\| \frac{\beta}{\frac{1}{n}\sum_{i=1}^n |\mathscr{X}_i|^2 + \frac{\lambda_n}{n}} \right\|_{L^2(K)} = O_P(1), \tag{3.19}$$

which is done with the same method used in Theorem 3.8 of Manrique et al. (2016). □

Proof of Proposition 18. This is a direct consequence of Proposition 4.1 of Manrique et al. (2016). □

3.C Generalization of Theorem 16

Theorem 23. *For the FCVM which satisfies the general hypotheses* $(HA1_{FCVM})$, $(HA2_{FCVM})$ *and* $(HA3_{FCVM})$, *hypotheses* $(A1)$ *in Theorem 15 and* $(A3)$ *in Theorem 16, we additionally assume*

$(A4bis)$ $\left\| \frac{|\mathscr{F}(\theta)|}{\mathbb{E}[|\mathscr{F}(X)|^2]} 1_{\overline{supp(\mathscr{F}(\theta))} \setminus \partial(supp(\mathbb{E}[|\mathscr{F}(X)|]))} \right\|_{L^2} < \infty,$

(A5) *There exist positive real numbers* $\alpha > 0$, $M_0, M_1, M_2 > 0$ *such that*

(a) *For every* $p \in C_{\theta,\partial X}$, *with* $C_{\theta,\partial X} := supp(|\mathscr{F}(\theta)|) \cap \partial(supp(\mathbb{E}[|\mathscr{F}(X)|]))$, *there exists an open interval neighborhood* $J_p \subset supp(|\mathscr{F}(\theta)|)$ *such that first*

$$\mathbb{E}[|\mathscr{F}(X)|^2(\xi)] \geq |\xi - p|^\alpha,$$

for every $\xi \in J_p$ *and secondly*

$$\left\| \frac{1}{\mathbb{E}[|\mathscr{F}(X)|^2]} \right\|_{L^2(J_p \setminus \{p\})} \leq M_0,$$

(b) $\sum_{p \in C_{\theta,\partial X}} \|\beta\|^2_{C_0(J_p)} < M_1$,

(c) $\frac{|\mathscr{F}(\theta)|}{\mathbb{E}[|\mathscr{F}(X)|^2]} 1_{supp(|\mathscr{F}(\theta)|) \setminus J} < M_2$, *where* $J = \bigcup_{p \in C_{\theta,\partial X}} J_p$,

(A6) *For* $n \geq 1$,

$$\lambda_n := n^{1 - \frac{1}{4\alpha+2}}.$$

Then

$$\|\hat{\theta}_n - \theta\|_{L^2} = O_P\left(n^{-\gamma}\right), \tag{3.20}$$

where $\gamma := \min\left[\frac{1}{2(2\alpha+1)}, \frac{1}{2} - \frac{1}{2(2\alpha+1)}\right]$.

Proof. As in the proof of Theorem 15, it is easy to show that $\mathscr{X}$, ε, β and $\mathscr{Y}$ satisfy all the hypotheses of Theorem 3.4 of Manrique et al. (2016), then $\|\hat{\beta}_n - \beta\|_{L^2} = O_P(n^{-\gamma})$. The isometry property of the CFT ends the proof. □

3.D Numerical Implementation of the FFDE

In this appendix we discuss how we estimate θ in the FCVM in practice. In particular we describe the necessity to rethink the FCVM in a finite discrete way, and to use the Discrete Fourier Transform as the discrete equivalent of the Continuous Fourier Transform in this new context. We start by describing the discretization of the convolution. To do this properly we start with some definitions.

Throughout this appendix we use Δ as the discretization step between two observation times (for instance $\Delta = 0.01$). The observation times are defined for every $j \in \mathbb{Z}$ as $t_j := j * \Delta$ and thus they define the grid G_Δ over $\mathbb{R}$. We use a fix grid in this appendix. With this grid we transform each function $f : \mathbb{R} \to \mathbb{C}$ to a vector $f^d \in \mathbb{C}^{\mathbb{Z}}$ infinite dimensional, with elements $f^d_j := f(t_j) \in \mathbb{C}$. In what follows the superscript ^{d} will denote this discretization.

Besides here all the functions will have compact support. Otherwise we should compute the convolution of infinite vectors which cannot be done in practice. For simplicity we consider all the functions defined over a compact interval $[0,T]$ with T large enough. Thus we will consider $f^d = (f^d_0, \cdots, f^d_{q-1}) \in \mathbb{C}^q$, where $q-1 = \max\{j \in \mathbb{N} \,|\, t_j \in [0,T]\}$.

Let RM (rectangular method) be the operator which associates to an integral over $\mathbb{R}$, its numerical approximation by the rectangular method over the grid of points we have already defined. Thus for a given integral $J = \int_{\mathbb{R}} f(s)ds = \int_0^T f(s)ds$ we associate $RM(J) := \Delta \sum_{j=0}^{q-1} f(t_j) = \Delta \sum_{j=0}^{q-1} f^d_j$.

Understanding how to compute numerically the convolution of two functions is a key element to implement the estimator developed for the FCVM.

We start our discussion by describing the discretization of the convolution of two functions with support included on $[0,T]$,

$$f * g(t) := \int_{-\infty}^{+\infty} f(s)g(t-s)ds = \int_0^T f(s)g(t-s)ds.$$

Approximating this convolution with the rectangular method we obtain for every $j \in \mathbb{N}$,

$$RM(f*g)(t_j) = \sum_{l=0}^{q-1} f(t_l)\, g(t_{j-l})\, \Delta = \Delta \sum_{l=0}^{q-1} f^d_l\, g^d_{j-l}. \tag{3.21}$$

The last sum in equation (3.21) is the convolution between vectors. Thus we can rewrite this equation as follows

$$RM(f*g)(t_j) = \Delta\, (f^d * g^d)_j.$$

for $j \in [0, \cdots, 2p-2]$ and where $(f^d * g^d)_j := \sum_{l=0}^{q-1} f^d_l g^d_{j-l}$. Besides note that for $j \notin [0, \cdots, 2p-2]$ we have $RM(f*g)(t_j) = 0$ since f and g have compact support.

Additionally we can compute the vector $((f^d * g^d)_0, \cdots, (f^d * g^d)_{2q-2})$ using matrices as follows

$$\Big((f^d * g^d)(0), \cdots, (f^d * g^d)(2q-2)\Big)^T = MC_G \, (f_0^d, \cdots, f_{q-1}^d)^T, \tag{3.22}$$

where MC_G is the matrix associated to the convolution discretized over the grid G, defined as follows

$$MC_G := \begin{pmatrix} g_0^d & 0 & 0 & 0 & \cdots & 0 \\ g_1^d & g_0^d & 0 & 0 & \cdots & 0 \\ g_2^d & g_1^d & g_0^d & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \cdots & \vdots \\ g_{q-2}^d & \cdots & \cdots & g_1^d & g_0^d & 0 \\ g_{q-1}^d & g_{q-2}^d & \cdots & \cdots & \cdots & g_0^d \\ 0 & g_{q-1}^d & g_{q-2}^d & \cdots & g_2^d & g_1^d \\ 0 & 0 & g_{q-1}^d & \cdots & \cdots & g_2^d \\ \vdots & \vdots & \vdots & \ddots & \cdots & \vdots \\ 0 & 0 & \cdots & 0 & g_{q-1}^d & g_{q-2}^d \\ 0 & 0 & 0 & \cdots & 0 & g_{q-1}^d \end{pmatrix} \in \mathbb{R}^{(2q-1)\times q}.$$

Remark : From this fact we note that the convolution could have a larger support. This arises because an important property of the convolution is that $supp(f*g) \subset \overline{supp(f)+supp(g)}$ (Brezis (2010, p. 106)). Thus in our case $supp(f*g) \subset [0, 2T]$. However afterwards we will take T large enough to contain even the convolution. In this way, every time we will consider the convolution of two functions f and g we suppose $\overline{supp(f)+supp(g)} \subset [0,T]$. In this case the number of discretization points q will be defined as before, namely $q-1 = \max\{j \in \mathbb{N} \,|\, t_j \in [0,T]\}$ but now for all $j \geq q$, $(f^d * g^d)_j = 0$. Besides the matrix representation of the convolution through MC_G will still be correct.

In the following subsection we explore the parallel between the continuous convolution of two functions and the convolution of two vectors with respect to the whole model FCVM.

3.D.1 The Discretization of the FCVM and the FFDE

We have defined the functional Fourier deconvolution estimator of θ in the FCVM using the continuous Fourier transform and its inverse (equations (3.3) and (3.4)). Given that both operators are integral operators, we need to use some kind of numerical approach to compute them. The goal of this subsection is to show that the proper way for doing this is by using a discrete model which behaves like the FCVM. This model will be based on the convolution of finite dimensional vectors. It will be studied through the discrete Fourier transform and its inverse instead of their continuous counterparts.

First let us show that it is not practical to compute the functional Fourier deconvolution estimator by direct approximation of the continuous Fourier transform and its inverse. This is not possible because these two operators are integrals defined over the whole $\mathbb{R}$. To see why this is a problem let us

consider a function $f \in L^2$ with compact support. Then although it is possible to use the Rectangular Method to compute $\mathscr{F}(f)(\xi)$ for every value ξ, we cannot ensure that $\mathscr{F}(f)$ has compact support ((Kammler, 2008, p. 130)). This implies that we need to know the values of $\mathscr{F}(f)$ for all the infinite values of the grid G_Δ to approximate the $\mathscr{F}^{-1}$, which is impossible in practice. Note that even if $\mathscr{F}(f)$ has a compact support we cannot know how large it is and in this case we will need to compute $\mathscr{F}(f)$ over too many points of the grid which again makes the approximation unpractical.

Instead of using the direct approximation of the continuous Fourier transform and its inverse, another approach is to propose a finite discretized version of the FCVM, which reflects the main characteristics of the FCVM. In order to achieve this, note two important things: i) the convolution of two functions can be approached by as the convolution of two vectors and ii) the convolution of two vectors is transformed into a multiplication with the discrete Fourier transform ((Kammler, 2008, p. 102), Oppenheim and Schafer (2011, p. 60)).

Here we use the definition of the discrete Fourier transform found in Kammler (2008, p. 291) or in Bloomfield (2004, p. 41), defined for vectors of $\mathbb{C}^q$ as follows

$$\begin{array}{rccl} \mathscr{F}_d: & \mathbb{C}^q & \to & \mathbb{C}^q \\ & f := (f_0, \cdots, f_{q-1}) & \mapsto & (\mathscr{F}_d(f)(0), \cdots, \mathscr{F}_d(f)(q-1)), \end{array}$$

where for every $l = 0, \cdots, q-1$,

$$\mathscr{F}_d(f)(l) := \frac{1}{q} \sum_{r=0}^{q-1} f_r \omega^{rl} \in \mathbb{C}. \tag{3.23}$$

with $\omega := e^{-2\pi i/q}$. If we define the matrix

$$\Omega_q := \begin{pmatrix} 1 & 1 & 1 & \cdots & 1 \\ 1 & (\omega^1)^1 & (\omega^1)^2 & \cdots & (\omega^1)^{(q-1)} \\ 1 & (\omega^2)^1 & (\omega^2)^2 & \cdots & (\omega^2)^{(q-1)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & (\omega^{(q-1)})^1 & (\omega^{(q-1)})^2 & \cdots & (\omega^{(q-1)})^{(q-1)} \end{pmatrix} \tag{3.24}$$

we can write

$$\mathscr{F}_d(f) = \frac{1}{q} \Omega_k \, f \, \in \mathbb{C}^q. \tag{3.25}$$

Furthermore from this definition we can deduce

$$\mathscr{F}_d^{-1} = \Omega_q^*, \tag{3.26}$$

where Ω_q^* is the conjugate transpose of Ω_q.

Remark: We can see that the definition of $\mathscr{F}_d$ depends on the number q, which is the length of the vector. In this way when we apply $\mathscr{F}_d$ to a vector of size p we need to redefine the matrix Ω_p by using $\omega := e^{-2\pi i/p}$.

Finite Discrete version of the FCVM Let us take T large enough such that $[0,T]$ contains $\overline{supp(X)+supp(\theta)}$. Thus the supports of θ, X and Y are also contained in $[0,T]$ (Brezis (2010, p. 106)). Let us define $q-1 = \max\{j \in \mathbb{N} \,|\, t_j \in [0,T]\}$. Now take the discretization of each function X_i and Y_i of the sample $(X_i,Y_i)_{i=1,\cdots,n}$ over the grid $[t_0,\cdots,t_{q-1}]$, so all these functions will become vectors in $\mathbb{R}^q \subset \mathbb{C}^q$, that is $X_i^d, Y_i^d \in \mathbb{C}^q$ for every $i = 1,\cdots,n$.

Given that the matrix Ω_q has the property of transforming finite convolutions into multiplications, we can use the three steps method as the one used to define the estimator $\hat{\theta}_n$ for the continuous case, namely i) transform the problem with the matrix Ω_q from the time-domain to the frequency one, ii) use the ridge estimator in this domain, and iii) finally come back with the inverse of Ω_q.

The comparison between the continuous and the discrete cases is done next. Note that in the discrete case the multiplication and the division is done the element by element between vectors of same length. Furthermore, $*^d$ is discrete convolution, Δ is the step of discretization and we use $P_q : \mathbb{R}^{2q-1} \to \mathbb{R}^q$, the projection into the first q components, to have vectors of the same length.

CONTINUOUS

Data and conditions: $\theta \in L^2([0,T])$. For $i = 1,\cdots,n$, $X_i, Y_i, \varepsilon_i \in L^2([0,T])$,

$$Y_i = \theta * X_i + \varepsilon_i.$$

Estimation steps:

1. For $i = 1,\cdots,n$,

$$\mathscr{F}(Y_i) = \mathscr{F}(\theta)\mathscr{F}(X_i) + \mathscr{F}(\varepsilon_i).$$

2.

$$\mathscr{F}\hat{(\theta)}_n := \frac{\sum_{i=1}^n \mathscr{F}(Y_i)\overline{\mathscr{F}(X_i)}}{\sum_{i=1}^n |\mathscr{F}(X_i)|^2 + \lambda_n}$$

3.

$$\hat{\theta}_n := \mathscr{F}^{-1}(\mathscr{F}\hat{(\theta)}_n)$$

DISCRETE

Data and conditions: $\theta^d \in \mathbb{R}^q$. For $i = 1,\cdots,n$, $X_i^d, Y_i^d, \varepsilon_i^d \in \mathbb{R}^q$,

$$Y_i^d = \Delta P_q(\theta^d *^d X_i^d) + \varepsilon_i^d.$$

Estimation steps:

1. For $i = 1,\cdots,n$,

$$\Omega_q(Y_i^d) = \Delta\Omega_q^d(\theta^d)\cdot\Omega_q(X_i^d) + \Omega_q(\varepsilon_i^d).$$

2.

$$\Omega_q\hat{(\theta^d)}_n := \frac{1}{\Delta}\frac{\sum_{i=1}^n \Omega_q(Y_i^d)\overline{\Omega_q(X_i^d)}}{\sum_{i=1}^n |\Omega_q(X_i^d)|^2 + \vec{\lambda}_n},$$

where $\vec{\lambda}_n := (\lambda_n,\cdots,\lambda_n) \in \mathbb{R}^q$.

3.

$$\hat{\theta}_n^d := \Omega_q^{-1}(\Omega_q\hat{(\theta^d)}_n).$$

From this comparison we can define the numerical estimator of θ over the grid $[t_0, \cdots, t_{q-1}]$ as follows

$$\hat{\theta}_n^d := \frac{1}{\Delta}\Omega_q^{-1}\left[\frac{\sum_{i=1}^n \Omega_q Y_i^d \cdot \overline{\Omega_q X_i^d}}{\sum_{i=1}^n |\Omega_q X_i^d|^2 + \vec{\lambda}_n}\right]. \tag{3.27}$$

3.D.2 Compact Supports and Grid of Observations

From now on we will compute $\hat{\theta}_n$ numerically with equation (3.27). The important question we want to address here is how large the grid of observation points should be to properly estimate θ? In this regard understanding the relationship between the supports of X and θ and the one of their convolution (Y) is an essential element to answer this question. We know that (Brezis (2010, p. 106)),

$$supp(Y) = supp(\theta * X) \subset \overline{supp(X) + supp(\theta)}.$$

Then as mentioned before whenever our grid of observations contains the interval $[0,T]$ and $[0,T]$ contains $\overline{supp(X) + supp(\theta)}$ we will be able to estimate θ over its whole compact support.

The problem arises from the fact that we do not know θ and as a consequence neither $supp(\theta)$ nor $\overline{supp(X) + supp(\theta)}$. Then how big T should be in order to estimate θ correctly?

There are several cases to consider. First let us suppose that the grid of observations covers $[0,T_1]$ and $supp(X), supp(Y) \subset [0,T_1]$ then we can choose $T > T_1$ big enough and estimate θ over $[0,T]$. To see this more clearly let us say that the grid of observations over $[0,T_1]$ is $t_0, \cdots, t_{q_1}$ and over $[0,T]$ is $t_0, \cdots, t_q$, with $q > q_1$. Given that we have only observed the curves over $[0,T_1]$ we only know the vectors $(X_i^d, Y_i^d)_{i=1,\cdots,n} \subset \mathbb{R}^{q_1}$. Then the only thing we need to do before applying equation (3.27) properly is to redefine the vectors X_i^d and Y_i^d by adding zeros such that they will belong to $\mathbb{R}^q$, for instance

$$X_i^d := (X_i^d, 0, \cdots, 0) \in \mathbb{R}^q.$$

This procedure is known as **zero padding** the signal (Gonzalez and Eddins (2009, p. 111)). In this case equation (3.27) is well defined and we will compute θ over $[0,T]$. Note also that $supp(\theta)$ could be bigger than $[0,T]$ but the estimation of θ over $[0,T]$ is still correct.

Secondly we have the case where the grid of observations covers $[0,T_1]$ and we know $supp(X) \subset [0,T_1]$ and $supp(Y) \setminus [0,T_1] \neq \emptyset$. Under these hypotheses we cannot add more zeros to the vectors Y_i^d because if we did it would imply that Y has zero values outside $[0,T_1]$ which contradicts $supp(Y) \setminus [0,T_1] \neq \emptyset$. Thus we cannot apply the property of Ω_q to transform the convolution into a multiplication correctly. This is one restriction to the correct application of the FCVM.

Finally if the grid of observations covers $[0,T_1]$, $supp(X) \setminus [0,T_1] \neq \emptyset$ and $supp(Y) \setminus [0,T_1] \neq \emptyset$ we have the same phenomenon, that is we cannot add more zeros to the vectors X_i^d and Y_i^d to belong to $\mathbb{R}^q$. Thus it is not possible to transform the convolution into a multiplication because q_1 is not big enough. Note that Ω_{q_1} is quite different from Ω_q (see definition 3.24) and the property of transforming

the convolution into a multiplication of two vectors only holds when Ω_q is applied to the entire convolution of both vectors, that is q is big enough to contain the convolution.

In any case in order to estimate θ with the functional Fourier deconvolution estimator, the grid of observations should cover $supp(X)$ and $supp(Y)$. This is an important restriction of this estimator.

FFT Algorithm and fast computing : One of the main advantages of the functional Fourier deconvolution estimator is that it is calculated very fast. This is due to the fact that it uses the Fast Fourier Transform (FFT) to compute the discrete Fourier transform. It is known that this algorithm computes the discrete Fourier transform of an n-dimensional signal in $O(n\log(n))$ time. The publication of the Cooley-Tukey FFT algorithm in 1965 (Cooley and Tukey (1965)) revolutionized the area of digital signal processing because it reduced the order of complexity of the Fourier transform and of the convolution from n^2 to $n\log(n)$, where n is the problem size. Then over the last years new algorithms have improved the performance of the Cooley-Tukey algorithm under some conditions (split-radix FFT, Winograd FFT, etc). Among the recent improvements we highlight the Nearly Optimal Sparse Fourier Transform (Hassanieh et al. (2012)).

Chapter 4

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

Contents

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

4.1 Introduction

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

$$Y(t) = \int_0^1 \mathscr{S}(t,s)X(s)ds + \varepsilon(t). \tag{4.1}$$

Here $\mathscr{S}(.,.)$ is an unknown integrable kernel: $\int_0^1 \int_0^1 |\mathscr{S}(t,s)|dtds < \infty$, to be estimated. ε is a noise random variable, independent of X. The functional variables X, Y and ε are random functions taking values on the interval $I = [0,1]$ of $\mathbb{R}$. Considering this particular interval is equivalent to considering any other interval $[a,b]$ in what follows. For the sake of clarity, we assume moreover that the random functions X and ε are centered. The case of non centered X and Y functions can be equivalently studied by adding an additive non random intercept function in model (4.1).

In all the sequel we consider a sample $(X_i,Y_i)_{i=1,\ldots,n}$ of independent and identically distributed observations, following (4.1) and taking values in the same Hilbert space $H = \mathbb{L}^2([0,1])$, the space of all real valued square integrable functions defined on $[0,1]$. The objective of this paper is to estimate the unknown noise covariance operator Γ_ε of ε and its trace $\sigma_\varepsilon^2 := tr(\Gamma_\varepsilon)$ from these data sets. The estimation of the noise covariance operator Γ_ε is well known in the context of multivariate multiple regression models, see for example Johnson and Wichern Johnson and Wichern (2007, section 7.7). The question is a little more tricky in the context of functional data. Answering it will then make possible the construction of hypothesis testing in connection with model (4.1).

Functional data analysis has given rise to many theoretical results applied in various domains (economics, biology, finance, etc). The monograph by Ramsay & Silverman Ramsay and Silverman (2005) is a major reference that gives an overview on the subject and highlights the drawbacks of considering a multivariate point of view. Novel asymptotic developments and illustrations on simulated and real data sets are also provided in Horváth & Kokoszka Horváth and Kokoszka (2012). We follow here the approach of Crambes & Mas Crambes and Mas (2013) that studied the prediction in the model (4.1) revisited as:

$$Y = SX + \varepsilon, \tag{4.2}$$

where $S : H \to H$ is a general linear integral operator defined by $S(f)(t) = \int_0^1 \mathscr{S}(t,s)f(s)ds$ for any function f in H. The authors showed that the trace σ_ε^2 is an important constant involved in the square

prediction error bound that participate to determine the convergence rate. The estimation of σ_ε^2 will thus provide details on the prediction quality in model (4.1).

In this context of functional linear regression, it is well known that the covariance operator of X cannot be inverted directly (see Cardot *et al.* Cardot et al. (1999)), thus a regularization is needed. In Crambes and Mas (2013), it is based on the Karhunen-Loève expansion and the functional principal component analysis of the (X_i). This approach is also often used in functional linear models with scalar output, see for example Cardot et al. (1999).

The construction of the estimator $\hat{S}$ is introduced in Section 4.2. Section 4.3 is devoted to the estimation of Γ_ε and its trace. Two types of estimators are given. Convergence properties are established and discussed. The proofs are postponed in Section 4.5. The results are illustrated on simulation trials in Section 4.4.

4.2 Estimation of S

4.2.1 Preliminaries

We denote respectively $<.,.>_H$ and $\|.\|_H$ the inner product and the corresponding norm in the Hilbert space H. We shall recall that $<f,g>_H = \int_0^1 f(t)g(t)dt$, for all functions f and g in $\mathbb{L}^2([0,1])$. In contrast, $<.,.>_n$ and $\|.\|_n$ stand for the inner product and the Euclidean norm in $\mathbb{R}^n$. The tensor product is denoted $\otimes$ and defined by $f \otimes g =< g,. >_H f$ for any functions $f,g \in H$.

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

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

4.2.2 Spectral decomposition of Γ

It is well known that Γ is a symmetric, positive trace-class operator, and thus diagonalizable in an orthonormal basis (see for instance Hsing and Eubank (2015)). Let $(\lambda_j)_{j\geq 1}$ be its non-increasing sequence of eigenvalues, and $(v_j)_{j\geq 1}$ the corresponding eigenfunctions in H. Then Γ decomposes as follows:

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

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

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

4.2.3 Construction of the estimator of S

We start from the moment equation

$$\Delta = S\Gamma. \tag{4.3}$$

On the sub-space $\langle v_1, \cdots, v_k \rangle$, the operator Γ is invertible, more precisely $(\Gamma\Pi_k)^{-1} = \sum_{j=1}^{k} \lambda_j^{-1} v_j \otimes v_j$. As a consequence, with equation (4.3) and the fact that $\Pi_k \Gamma \Pi_k = \Gamma\Pi_k$ we get, on the sub-space $\langle v_1, \cdots, v_k \rangle$, $\Delta\Pi_k = (S\Pi_k)(\Gamma\Pi_k)$. We deduce that $S\Pi_k = \Delta\Pi_k (\Gamma\Pi_k)^{-1}$.

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

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

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

$$\hat{S}_{k_n} = \hat{\Delta}_n \hat{\Gamma}_{k_n}^{+}. \tag{4.4}$$

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

$$\hat{\mathscr{S}}_{k_n}(t,s) = \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{k_n} \left[\left(\frac{1}{\hat{\lambda}_j} \int_0^1 X_i(r) \hat{v}_j(r) dr \right) Y_i(t) \hat{v}_j(s) \right]. \tag{4.5}$$

4.3 Estimation of Γ_ε and its trace

4.3.1 The plug-in estimator

The plug-in estimator of Γ_ε is given by

$$\hat{\Gamma}_{\varepsilon,n} := \frac{1}{n-k_n} \sum_{i=1}^{n} (Y_i - \hat{S}_{k_n} X_i) \otimes (Y_i - \hat{S}_{k_n} X_i) = \frac{1}{n-k_n} \sum_{i=1}^{n} \hat{\varepsilon}_i \otimes \hat{\varepsilon}_i. \tag{4.6}$$

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

Theorem 24. *Let $(X_i, Y_i)_{i=1,\dots,n}$ be a sample of i.i.d. observations following model (4.1). Let $k_n < n$ be an integer. We have*

$$\mathbb{E}[\hat{\Gamma}_{\varepsilon,n}] = \Gamma_\varepsilon + \left(\frac{n}{n-k_n}\right) S\,\mathbb{E}\left(\sum_{i=k_n+1}^{n} \hat{\lambda}_i \hat{v}_i \otimes \hat{v}_i\right) S'. \tag{4.7}$$

The proof of Theorem 24 is postponed in Section 4.5.1. As $\hat{\Gamma}_{k_n} = \sum_{i=1}^{n} \hat{\lambda}_i \hat{v}_i \otimes \hat{v}_i$ and $\hat{\Pi}_{(k_n+1):n} := \sum_{i=k_n+1}^{n} \hat{v}_i \otimes \hat{v}_i$, we deduce the following result:

Corollary 25. *We have*

$$\mathbb{E}[\hat{\Gamma}_{\varepsilon,n}] = \Gamma_\varepsilon + \left(\frac{n}{n-k_n}\right) S\,\mathbb{E}\left(\hat{\Pi}_{(k_n+1):n}\hat{\Gamma}_n\right) S', \tag{4.8}$$

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

Under some additional assumptions, we prove that the plug-in estimator (4.6) of Γ_ε is asymptotically unbiased. Let us consider the following assumptions:

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

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

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

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

Our main result is then the following.

Theorem 26. *Under assumptions (A.1)-(A.4), if $(k_n)_{n\geq 1}$ is a sequence such that $\lim_{n\to+\infty} k_n = +\infty$ and $\lim_{n\to+\infty} k_n/n = 0$, we have*

$$\lim_{n\to+\infty} \left\|\mathbb{E}\left(\hat{\Gamma}_{\varepsilon,n}\right) - \Gamma_\varepsilon\right\|_{\mathcal{N}} = 0. \tag{4.9}$$

The proof is postponed in Section 4.5.2. From the definition of the nuclear norm, we immediately get the following corollary:

Corollary 27. *Under the assumptions of Theorem 26, we have*

$$\lim_{n\to+\infty} \mathbb{E}\left[tr\left(\hat{\Gamma}_{\varepsilon,n}\right)\right] = tr\left(\Gamma_\varepsilon\right). \tag{4.10}$$

4.3.2 Other estimation of Γ_ε

Without loss of generality, we assume in this section that n is a multiple of 3. In formula (4.8), the bias of the plug-in estimator is related to $S\,\mathbb{E}\left(\hat{\Pi}_{(k_n+1):n}\hat{\Gamma}_n\right) S'$. Another way of estimating Γ_ε is thus to subtract an estimator of the bias to the plug-in estimator $\hat{\Gamma}_{\varepsilon,n}$. To achieve this, we split the

n-sample into three sub-samples with size $m = n/3$ to keep good theoretical properties thanks to the independence of the sub-samples. As a consequence, we define

$$\check{B}_n := \hat{S}^{[2]}_{2k_m} \left(\hat{\Pi}^{[1]}_{(k_m+1):m} \hat{\Gamma}^{[1]}_m \right) \left(\hat{S}^{[3]}_{2k_m} \right)', \tag{4.11}$$

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

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

$$\check{\Gamma}_{\varepsilon,n} := \hat{\Gamma}^{[1]}_{\varepsilon,m} - \frac{m}{m-k_m} \check{B}_n. \tag{4.12}$$

The following result is established.

Theorem 28. *Under the assumptions of Theorem 26, we have*

$$\lim_{n \to +\infty} \left\| \mathbb{E}\left(\check{\Gamma}_{\varepsilon,n} \right) - \Gamma_\varepsilon \right\|_{\mathscr{N}} = 0. \tag{4.13}$$

The above result can also be written using the trace.

Corollary 29. *Under the assumptions of Theorem 26, we have*

$$\lim_{n \to +\infty} \mathbb{E}\left[tr\left(\check{\Gamma}_{\varepsilon,n} \right) \right] = tr\left(\Gamma_\varepsilon \right). \tag{4.14}$$

4.3.3 Comments on both estimators

Subtracting an estimator of the bias to the plug-in estimator $\hat{\Gamma}_{\varepsilon,n}$ does not provide an unbiased estimator of $\Gamma_{\varepsilon,n}$. The situation is completely different to that of multivariate multiple regression models, see Johnson and Wichern (2007), where an unbiased estimator of the noise covariance is easily produced.

Both estimators $\hat{\Gamma}_{\varepsilon,n}$ and $\check{\Gamma}_{\varepsilon,n}$ are consistent. We can see from the proofs of Theorems 26 and 28 that $\left\| \mathbb{E}\left(\hat{\Gamma}_{\varepsilon,n} \right) - \Gamma_\varepsilon \right\|_{\mathscr{N}} \leq \frac{n}{n-k_n} \left\| S \right\|_{\mathscr{N}} \left\| S' \right\|_\infty \mathbb{E}\left| \hat{\lambda}_{k_n+1} \right|$ and

$$\left\| \mathbb{E}\left(\check{\Gamma}_{\varepsilon,n} \right) - \Gamma_\varepsilon \right\|_{\mathscr{N}} \leq 2 \frac{n}{n-3k_m} \left\| S \right\|_{\mathscr{N}} \left\| S' \right\|_\infty \mathbb{E}\left| \hat{\lambda}_{k_m+1} \right|.$$

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

4.3.4 Cross validation and Generalized cross validation

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

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

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

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

Proposition 30. *We denote $\boldsymbol{X}$ the matrix with size $n \times k_n$ with general term $\langle X_i, \hat{v}_j \rangle_H$ for $i = 1,\dots,n$ and $j = 1,\dots,k_n$, and $\boldsymbol{H} = \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'$. Then*

$$Y_i - \hat{Y}_i^{[-i]} = \frac{Y_i - \hat{Y}_i}{1 - \boldsymbol{H}_{ii}}, \tag{4.15}$$

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

This proposition allows to write the expression $Y_i - \hat{Y}_i^{[-i]}$ without excluding the ith observation, and allows to get the generalized cross validation criterion, which is computationally faster than the cross validation criterion (see for example Wahba (1990)). The term $\mathbf{H}_{ii}$ can be replaced by the mean $tr(\mathbf{H})/n$. Then, after noticing that $tr(\mathbf{H}) = tr(\mathbf{I}_{k_n}) = k_n$, where $\mathbf{I}_{k_n}$ is the identity matrix with size k_n, we get

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

4.4 Simulations

4.4.1 Setting

The variable X is simulated as a standard Brownian motion on $[0,1]$, with its Karhunen Loève expansion, given by Ash & Gardner Ash and Gardner (1975)

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

where the $v_j(t) := \sqrt{2}\sin((j-1/2)\pi t)$ and $\lambda_j = \frac{1}{(j-0.5)^2\pi^2}$ are the the eigenfunctions and the eigenvalues of the covariance operator of X. In practice, $X(t)$ has been simulated using a truncated version

with 1 000 eigenfunctions. The considered observation times are $[\frac{1}{1000}, \frac{2}{1000}, \cdots, \frac{1000}{1000}]$. We simulate a sample with sizes $n = 300$ and $n = 1\,500$.

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

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

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

4.4.2 Three estimators

We consider three different estimators of the trace of the covariance operator of ε: *(i)* the plug-in estimator given in (4.6), *(ii)* the corrected estimator given in (4.11) and (4.12), and *(iii)* the estimator $\hat{\mathring{\Gamma}}_{\varepsilon,n} := \hat{\Gamma}_{\varepsilon,n} - \left(\frac{n}{n-k_n}\right)\left[\hat{S}_{n,2k_n}(\hat{\Pi}_{(k_n+1):n}\hat{\Gamma}_n)(\hat{S}_{n,2k_n})'\right]$. The third estimator uses the whole sample when trying to remove the bias term, so it is not possible to obtain an immediate consistency result for this estimator because we do not have anymore the independence between the terms $\hat{S}_{n,2k_n}$ and $\hat{\Pi}_{(k_n+1):n}\hat{\Gamma}_n$, but we can see its practical behavior.

4.4.3 Results

We present in table 4.1 (simulation 1) and table 4.2 (simulation 2) the mean values of the trace obtained for the three estimators on $N = 100$ simulations, as well as the *CV* and *GCV* criteria. The criteria have a convex form, that allows to choose a value for k.

In simulation 1, the true value of k is known ($k = 20$) and the values chosen by *CV* and *GCV* are $k = 22$ or $k = 24$. For these values of k, the best estimator is $tr(\check{\Gamma}_{\varepsilon,n})$ for $n = 300$ and $n = 1\,500$. The overestimation of $tr(\hat{\Gamma}_{\varepsilon,n})$ seems to be well corrected by $tr(\check{\Gamma}_{\varepsilon,n})$, even if the usefulness of this bias removal cannot be theoretically proved. On this simulation, the estimator $tr(\hat{\mathring{\Gamma}}_{\varepsilon,n})$ does not behave better than the others, especially for small sample sizes.

In simulation 2, the true value of k is unknown and the value chosen by *CV* and *GCV* is $k = 4$ (for $n = 300$) or $k = 6$ (for $n = 1\,500$). The estimator $tr(\hat{\mathring{\Gamma}}_{\varepsilon,n})$ is slightly better than the two others for $n = 300$. For $n = 1500$, $tr(\check{\Gamma}_{\varepsilon,n})$ is slightly better.

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

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

Table 4.1 *CV* and *GCV* criteria for different values of k and mean values for the estimators of $Tr(\Gamma_\varepsilon)$ (simulation 1 with $n = 300$ and $n = 1500$). All values are given up to a factor of 10^{-3} (the standard deviation is given in brackets up to a factor of 10^{-4}).

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

Table 4.2 *CV* and *GCV* criteria for different values of k and mean values for the estimators of $Tr(\Gamma_\varepsilon)$ (simulation 2 with $n = 300$ and $n = 1500$). All values are given up to a factor of 10^{-3} (the standard deviation is given in brackets up to a factor of 10^{-4}).

4.5 Proofs

4.5.1 Proof of Theorem 24

We begin with preliminary lemmas.

Lemma 31. $\hat{S}_{k_n} = S\hat{\Pi}_{k_n} + \frac{1}{n}\left[\sum_{i=1}^n \varepsilon_i \otimes (\hat{\Gamma}^+_{k_n} X_i)\right]$.

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

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

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

Lemma 32. *We have*

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

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

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

Let us remark that $A = \mathbf{X}\Lambda^{-1}\mathbf{X}'$, where $\mathbf{X}$ is introduced in Proposition 30 and Λ is the diagonal matrix $\Lambda := diag(\hat{\lambda}_1,\cdots,\hat{\lambda}_{k_n})$. We obtain

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

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

Now, coming back to the proof of Theorem 24, we can write

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

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

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

We start with P_I. Using Lemma 31, we have, for $i = 1,\ldots,n$,

$$(S-\hat{S}_{k_n})X_i = S(I-\hat{\Pi}_{k_n})X_i - \frac{1}{n}\sum_{j=1}^n \langle \hat{\Gamma}_{k_n}^+ X_j, X_i\rangle_H \varepsilon_j, \tag{4.16}$$

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

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

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

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

and Lemma 32 gives $\mathbb{E}[P_I^{(4)}] = \frac{k_n}{n-k_n}\Gamma_\varepsilon$. So, we have shown that

$$\mathbb{E}[P_I] = \frac{n}{n-k_n} S\,\mathbb{E}\left[\sum_{i=k_n+1}^n \hat{\lambda}_i \hat{v}_i \otimes \hat{v}_i\right] S' + \frac{k_n}{n-k_n}\Gamma_\varepsilon. \tag{4.17}$$

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

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

By the independence between X and ε, the result of Lemma 32, and a similar computation for P_{III}, we obtain

$$\mathbb{E}[P_{II}] = \mathbb{E}[P_{III}] = -\frac{k_n}{n-k_n}\Gamma_\varepsilon. \tag{4.18}$$

Finally, coming back to the computation of $\mathbb{E}(\hat{\Gamma}_{\varepsilon,n})$, Theorem 24 is a direct consequence of (4.17) and (4.18). □

4.5.2 Proof of Theorem 26

The proof is based on the two following lemmas.

Lemma 33. *Under the assumptions of Theorem 26, we have*

$$\lim_{n\to+\infty} \mathbb{E}\left|\hat{\lambda}_{k_n}\right| = 0. \tag{4.19}$$

Proof: We have $\left(\mathbb{E}\left|\hat{\lambda}_{k_n}\right|\right)^2 \leq 2\lambda_{k_n}^2 + 2\mathbb{E}\left|\hat{\lambda}_{k_n} - \lambda_{k_n}\right|^2$.
From Lemma 2.2 and Theorem 2.5 in Horváth and Kokoszka (2012) with assumption (A.2), we obtain

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

which concludes the proof of the lemma. □

Lemma 34. *Under the assumptions of Theorem 26, we have*

$$\left\|S\mathbb{E}\left(\hat{\Pi}_{(k_n+1):n}\hat{\Gamma}_n\right)S'\right\|_{\mathscr{N}} \leq \|S\|_{\mathscr{N}}\left\|S'\right\|_\infty \mathbb{E}\left|\hat{\lambda}_{k_n+1}\right|. \tag{4.20}$$

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

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

which yields (4.20) as the norm $\|.\|_\infty$ corresponds to the largest eigenvalue of the operator. □

Theorem 26 is proved by combining Corollary 25 with Lemmas 33 and 34, and taking assumption (A.1) into account.

4.5.3 Proof of Theorem 28

We begin with the following lemmas.

Lemma 35. *Under the assumptions of Theorem 26, we have*

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

Proof: We first note that

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

As X and ε are independent, we get that $\mathbb{E}\left(\hat{S}_{2k_m}\right) = S\mathbb{E}\left(\hat{\Pi}_{2k_m}\right)$, which, combined with Corollary 25 and the fact that the three sub-samples are independent, ends the proof. □

Lemma 36. *Under the assumptions of Theorem 26, we have*

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

Proof: The proof is based on the same ideas as that used for proving Lemma 34. We remind that the infinite norm of projection operators are equal to one. □

The proof of Theorem 28 is now a simple combination of Lemmas 33, 34, 35 and 36 and using the triangle inequality.

4.5.4 Proof of Proposition 30

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

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

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

$$\hat{\alpha}(t) = \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{Y}(t) = \left(\langle \hat{\mathscr{S}}_{k_n}(t,.), \hat{v}_1 \rangle_H, \ldots, \langle \hat{\mathscr{S}}_{k_n}(t,.), \hat{v}_{k_n} \rangle_H\right)', \tag{4.21}$$

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

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

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

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

Acknowledgements: We are grateful to both anonymous referees for valuable comments that helped us to improve the paper.

Chapter 5

Modelling of High-throughput Plant Phenotyping with the FCVM

Contents

The purpose of this chapter is to illustrate the implementation of the FCVM on a real dataset acquired in plant science experiments. The dataset consists in curves of Vapour Pressure Deficit (VPD) and Leaf Elongation Rate (LER) obtained on two high-throughput plant phenotyping platforms. The Vapour Pressure Deficit (VPD) is the difference (deficit) between the amount of moisture in the air and how much moisture the air can hold when it is saturated. In addition, the Leaf Elongation Rate (LER) is an important variable that characterize the growth of a plant.

The history of the VPD influences the LER curve. This can be modeled through the historical functional linear model (1.1) or the FCVM (1.3). The objective of this chapter is to understand better how the VPD influences the LER.

5.1 Datasets

5.1.1 Dataset T72A

In this dataset the VPD and LER of 18 plants were measured every 15 minutes from Day 159 to Day 168 of the year 2014 (June and July). This gives 96 observation times per day.

There were two platforms for this experiment: a growth chamber and a greenhouse. In the growth chamber the VPD is repeated, whereas in the greenhouse the VPD is not stable and changes all along the day and among days (sunny or cloudy days). The VPD curves depends on the environment and then they are the same for plants in the same platform in each day. This implies collinearity among these input curves.

For each day the first measurement of a plant could be 7:15am or 0:00am. This depends on whether the plant has been moved from the greenhouse to the growth chamber or vice versa at the previous day. For this reason there are missing values for some plants and some days. In total there is around 12% of missing data. Moreover some plants have not been studied during certain days due to the difference in development speed and phenological stages among plants.

We have extracted the curves which do not have zero values and have at most 5 NA's (missing observations). We used the **R** function **approx** to reconstruct these curves. We kept only the LER curves that have values to ensure that the plant were not stressed.

R Data-frames : The dataset **T72A** contains other information (variables) than the VPD and LER measures. Besides as mentioned before there are missing data. For this reason we have extracted two datasets (**R** data-frames), each of which contain the name of the plants, the dates and either the VPD or LER curves respectively.

It is numerically more stable to apply the deconvolution methods to curves which starts with its support (non-zero part). That is why the VPD and LER curves start at 4am in the morning.

Each of these data-frames has 35 rows and 98 columns. The two first columns contain the name of the plant and the date. The remaining 96 columns represent the variable measured at the 96 observation times starting at 4am until 4am the next day. In Figure 5.1 we plot the VPD and LER curves of these two data-frames from the experiment T72A.

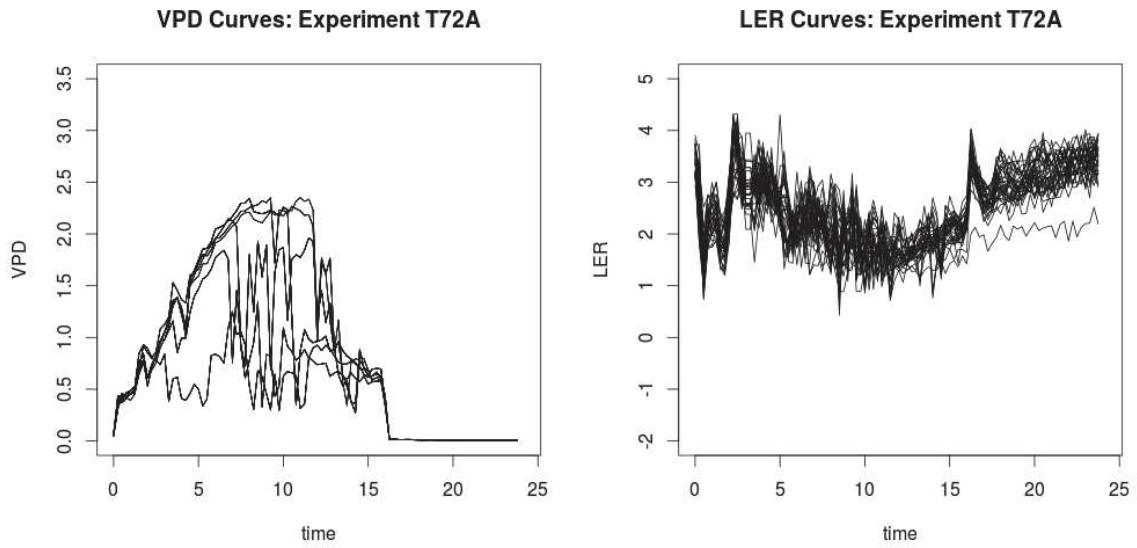

Fig. 5.1 VPD and LER curves from the experiment T72A.

5.1.2 Dataset T73A

In this dataset the VPD and LER of 108 plants where measured every 15 minutes (96 observation times per day). But in this case there are three subsets of 36 plants which have been sown in different dates. The whole experiment took place between the Day 322 to the Day 350 of the year 2014 (November and December).

The conditions of this experiment are similar to those of **T72A**. There were two experimental platforms: a growth chamber and a greenhouse. There are around 15% of missing data. There are collinearity among some of the VPD curves.

Again we have extracted the curves which do not have zero values and have at most have 5 NAs (missing observations). We used the **R** function **approx** to reconstruct these curves. But in contrast with **T72A** the LER curves do not have values higher than 3 in this experiment which implies that the plant were stressed.

R Data-frames : In the same way as **T72A** we have extracted two datasets (**R** data-frames). Each of these datasets has 380 rows and 98 columns. The two first columns contain the name of the plant and the date. The remaining 96 columns represent the variable measured at the 96 observation times starting at 4:30am until 4:30am the next day. In Figure 5.2 we plot the VPD and LER curves of these two data-frames from the experiment **T73A**.

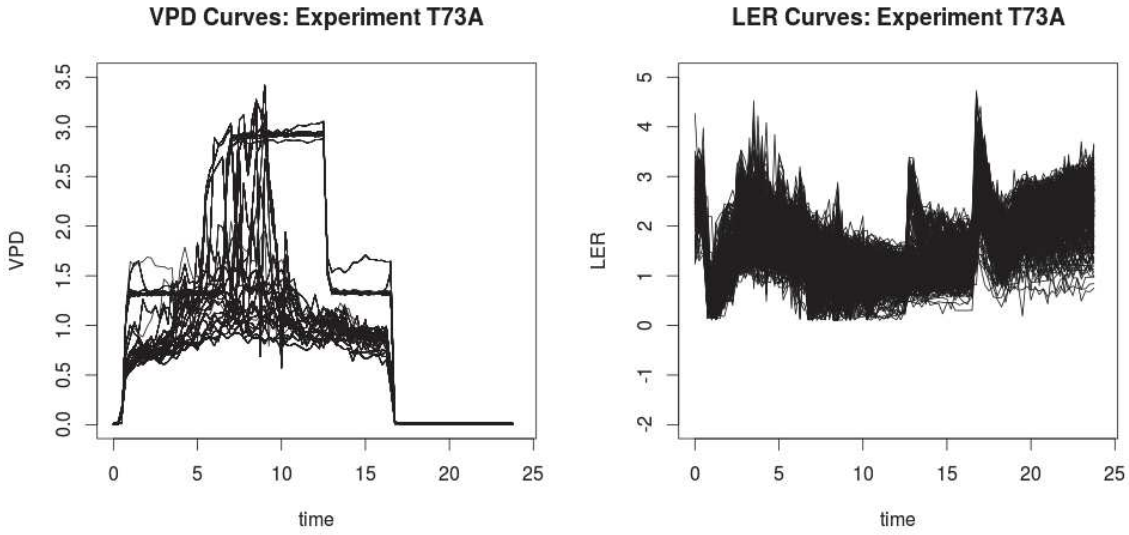

Fig. 5.2 VPD and LER curves from the experiment T73A.

5.2 Functional Convolution Model

In this section we add a functional intercept μ to the model (1.3) to have a larger set of estimators of θ. Then the new FCVM has the form

$$Y(t) = \mu(t) + \int_0^t \theta(s)X(t-s)\,ds + \varepsilon(t). \tag{5.1}$$

Next we describe how to estimate μ and θ in this new situation.

The Estimators : From equation (5.1) it is easy to see that

$$\mathbb{E}[Y] = \mu + \theta * \mathbb{E}[X],$$

where $*$ is the convolution. So if we center the data X and Y we obtain

$$Y - \mathbb{E}[Y] = \theta * (X - \mathbb{E}[X]) + \varepsilon. \tag{5.2}$$

Thus we can use the centered curves $(X_i - \mathbb{E}[X], Y_i - \mathbb{E}[Y])_{i=1,\cdots,n}$ to estimate θ with the Functional Fourier Deconvolution Estimator (FFDE), the Parametric Wiener estimator (ParWD), the adapted Singular Value Decomposition (SVD), the adapted Tikhonov estimator (Tik) and the Laplace estimator (Lap) (see Chapter 3) and then estimate μ through

$$\hat{\mu}_n := \bar{Y}_n - \hat{\theta}_n * \bar{X}_n, \tag{5.3}$$

where $\bar{X}_n$ and $\bar{Y}_n$ are the empirical estimators of the mean functions.

5.2.1 Estimation with Experiment T72A

The results of the estimation of θ and μ are shown in Figure 5.3. We can see three subgroups of estimators. First the Fourier (FFDE) and the Wiener (ParWD) approaches are similar, both of them are monotone decreasing functions. Secondly the SVD and Tikhonov (Tik) approaches are related to each other, and both of them differ of the first two estimators. Lastly the Laplace estimator is quite different from the other ones.

The difference among these subgroups is due to the use of the different methods to compute the estimators: Fourier and Wiener use the discrete Fourier transform, SVD and Tikhonov use the pseudo-inverse of the matrix associated to the convolution, and Laplace uses the Laguerre functions to project the convolution onto a finite dimensional subspace.

All the aforementioned estimators except Laplace use optimized parameters of regression. In the case of the Fourier and Wiener we use the Leave-one-out predictive cross-validation (LOOPCV). The optimal parameters for these are $\lambda_n = 0$ and $\alpha = 0.04465$ respectively (see subsection 3.5.1 in Chapter 3). For the SVD and Tikhonov we use the k-fold predictive cross-validation with $k = 5$ to obtain the optimal parameters $d = 2$ (dimension of inversion for the SVD) and $\rho = 10000$ respectively.

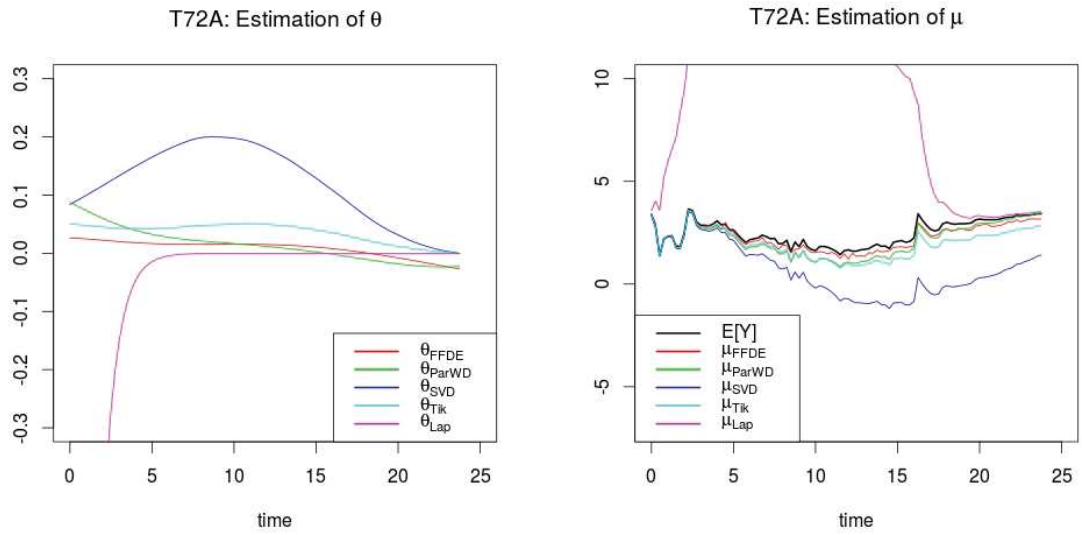

Fig. 5.3 Estimation of θ and μ.

The residuals for each estimators are shown in Figure 5.4. We see that the prediction of Y_i in each model does not improve that much over the empirical mean estimator of $\mathbb{E}[Y]$ (plot (a) in Figure 5.4). In particular the SVD and the Tikhonov methods give worse predictions than Fourier and Wiener. Moreover, Laplace cannot predict the Y_i curves.

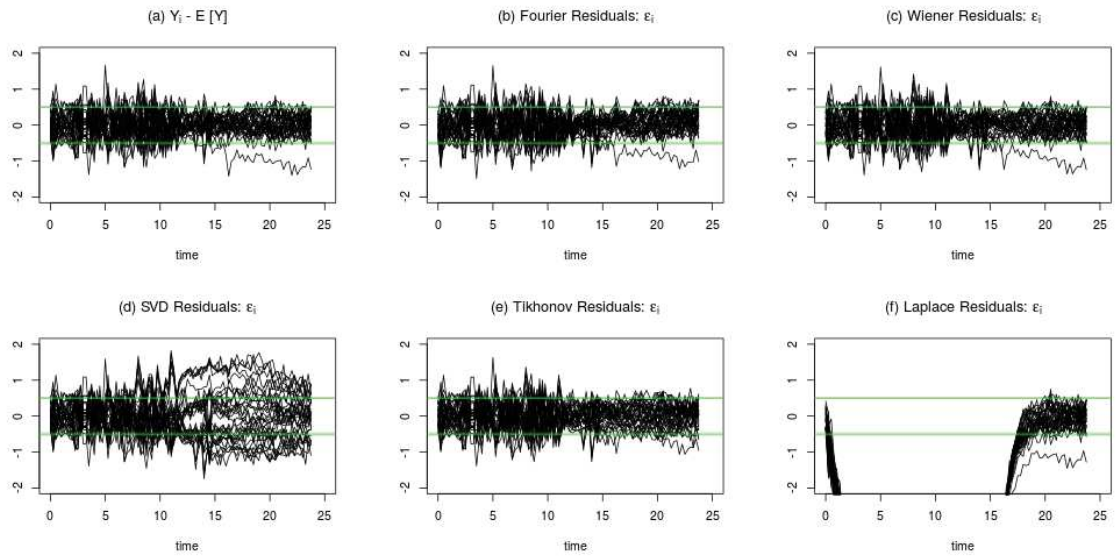

Fig. 5.4 Residuals of the estimators in the FCVM. (a) Residuals of the empirical mean estimator ($Y_i - \bar{Y}_n$). (b) Residuals of the Fourier estimator (FFDE). (c) Residuals of Wiener (ParWD). (d) Residuals of SVD. (e) Residuals of Tikhonov (Tik). (f) Residuals of Laplace (Lap). In all the pictures we plot green lines (constant values -0.5 and 0.5 respectively) to help the comparison.

5.2.2 Estimation with Experiment T73A

The results of the estimation of θ and μ are shown in Figure 5.5. In a similar way to the results with the experiment **T72A** we see that there are three subgroups among these estimators: first Fourier (FFDE) and Wiener (ParWD), secondly SVD and Tikhonov (Tik) and lastly Laplace. This is due to the use of the different methods to compute the estimators as we commented in the experiment **T72A**.

In contrast to the Fourier and Wiener estimators for the experiment **T72A** shown in Figure 5.3 we see here that these estimators have a more complex shape, whereas the SVD and Tikhonov are similar to the previous ones.

The optimized parameters of regression for Fourier and Wiener are $\lambda_n = 88.71029$ and $\alpha = 0.03373$ respectively (see subsection 3.5.1 Chapter 3). And for the SVD and Tikhonov these parameters are $d = 2$ and $\rho = 10000$ respectively.

The residuals for each estimators are shown in Figure 5.6. Again the prediction of Y_i of these methods does not outperform the empirical mean estimator of $\mathbb{E}[Y]$ (plot (a) in Figure 5.6). In particular the SVD and the Tikhonov methods give worse Estimation than Fourier and Wiener. Furthermore, Laplace cannot predict the Y_i curves.

The results in both experiments show that the use of the FCVM does not improve the prediction over the empirical mean estimator of $\mathbb{E}[Y]$. This suggests that a more complex model better explains

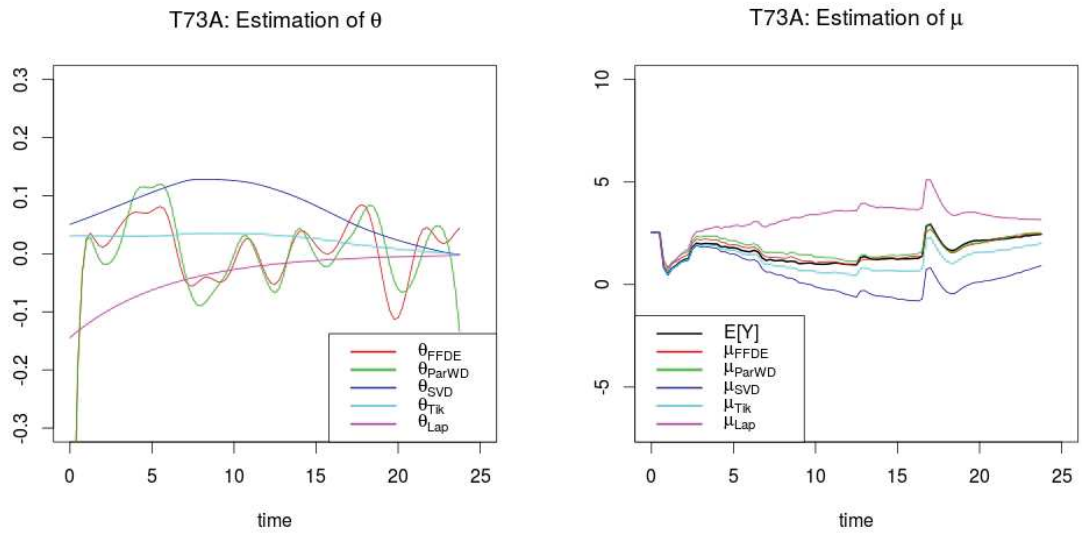

Fig. 5.5 Estimation of θ and μ.

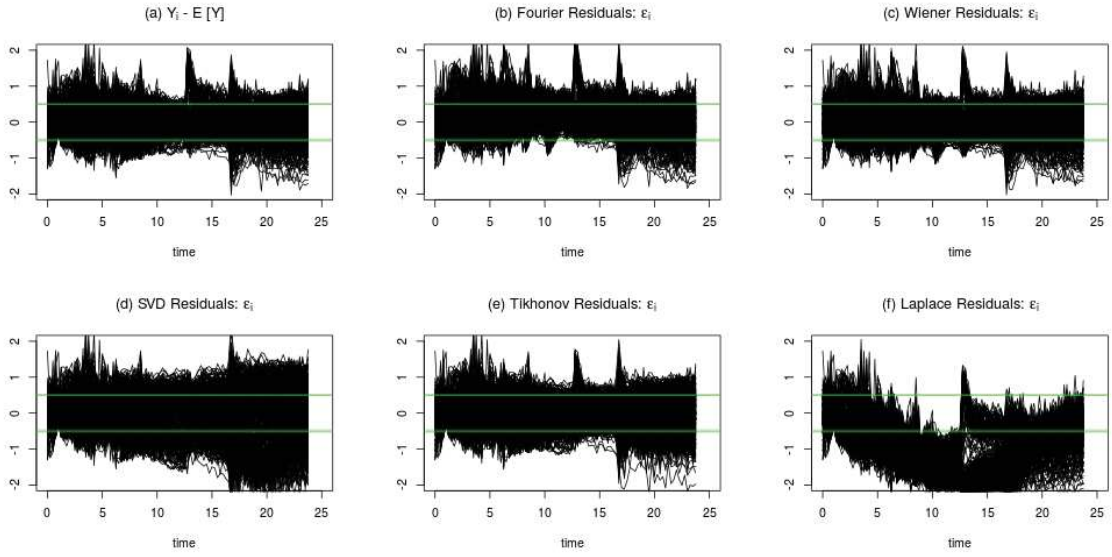

Fig. 5.6 Residuals of the estimators in the FCVM. (a) Residuals of the empirical mean estimator ($Y_i - \bar{Y}_n$). (b) Residuals of the Fourier estimator (FFDE). (c) Residuals of Wiener (ParWD). (d) Residuals of SVD. (e) Residuals of Tikhonov (Tik). (f) Residuals of Laplace (Lap). In all the pictures we plot green lines (constant values -0.5 and 0.5 respectively) to help the comparison.

how the VPD influences the LER. For this reason we use the historical functional linear model in the following section.

5.3 Historical Functional Linear Model

Estimators: Again we add a functional intercept μ to model (1.1) to have a larger set of estimators of the kernel $\mathscr{K}_{hist}$ and to have a similar model to the FCVM with intercept (5.1). Then the new historical model has the form

$$Y(t) = \mu(t) + \int_0^t \mathscr{K}_{hist}(s,t)X(s)\,ds + \varepsilon(t). \tag{5.4}$$

We estimate μ in a similar way to equation (5.3), that is, we use the centered curves to estimate $\mathscr{K}_{hist}$ and then we use the empirical means to estimate μ through

$$\hat{\mu}_n(t) := \bar{Y}_n(t) - \int_0^t \hat{\mathscr{K}}_{hist}(s,t)\bar{X}_n(s)\,ds.$$

The estimation of $\mathscr{K}_{hist}$ is done with two estimators: the Karhunen-Loève estimator (subsection 4.2.3 in Ch 4) and the Tikhonov functional estimator defined below.

Tikhonov Functional Estimator: This estimator is a variation of the Karhunen-Loève one. To define it we use the same elements used in the definition of the Karhunen-Loève estimator (see subsection 4.2.3 in Ch 4). In particular we use the moment equation (4.3). But instead of taking the first k_n dimensions to compute the generalized inverse $\hat{\Gamma}_{k_n}^{+}$ of the covariance operator, we use a positive number $\rho > 0$ which will be the Tikhonov (ridge) regularization parameter. With this value we define the Tikhonov generalized inverse as follows

$$\hat{\Gamma}_{\rho}^{+} := \sum_{j=1}^{n} \frac{\hat{\lambda}_j}{\hat{\lambda}_j^2 + \rho}\, \hat{v}_j \otimes \hat{v}_j,$$

and the Tikhonov Functional Estimator as

$$\hat{S}_{\rho} = \hat{\Delta}_n \hat{\Gamma}_{\rho}^{+}. \tag{5.5}$$

5.3.1 Estimation with Experiment T72A

The top view (level plot) of the Karhunen-Loève and Tikhonov functional estimators of the historical kernel ($\mathscr{K}_{hist}$) are shown in Figure 5.7. We can see that they have both a similar structure, in particular the sub-matrix around the ordered pair $(40,40)$.

To optimize the parameters of regression for these estimators we have used the generalized cross-validation (see subsection 4.3.4 Chapter 4) for the Karhunen-Loève estimator and the k-fold

predictive cross-validation with $k = 5$ for the Tikhonov estimator. The optimal parameters are $k_n = 5$ for Karhunen-Loève and $\rho = 0.001046277$ for Tikhonov.

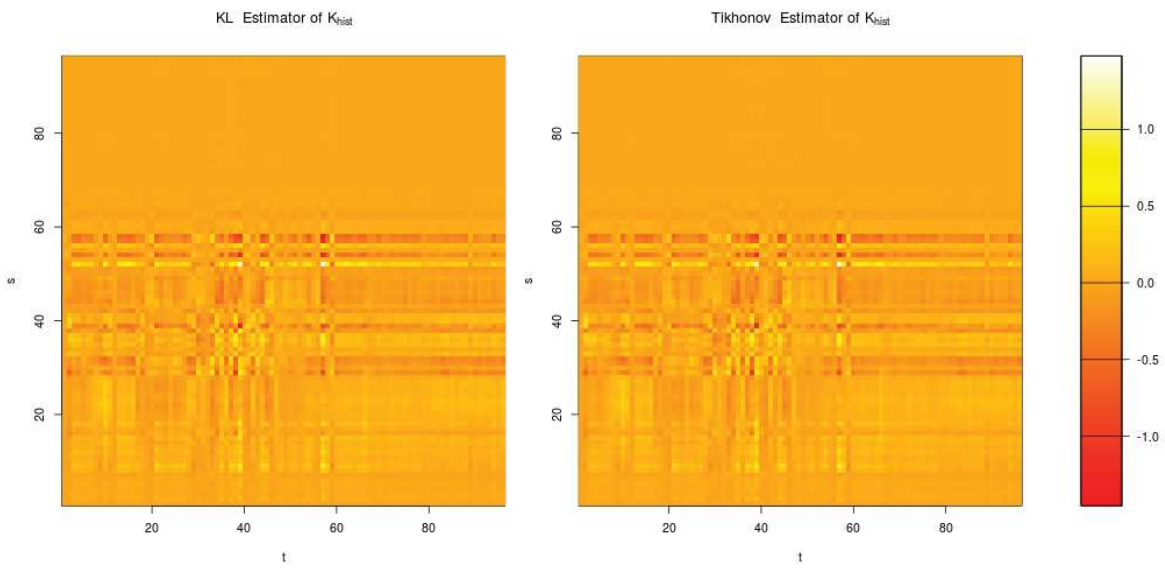

Fig. 5.7 Karhunen-Loève and Tikhonov functional estimators of the historical kernel ($\mathcal{K}_{hist}$).

The estimators of the functional intercept (μ) are shown in Figure 5.8. Both are quite similar which is consistent with the similarity of the kernel estimators. Besides the residuals of each estimation method are shown in Figure 5.9. In that figure we see that the prediction when using this estimators improves over the FCVM (smaller residuals).

5.3.2 Estimation with Experiment T73A

The top view (level plot) of the Karhunen-Loève and Tikhonov functional estimators of the historical kernel ($\mathcal{K}_{hist}$) are shown in Figure 5.10. Again both of them have a similar structure, in particular the diagonal shape for the sub-matrix of the first 60 rows and 60 columns.

We use the same methods to optimize the parameters of regression used for the experiment **T72A**. The optimal parameters now are $k_n = 16$ for Karhunen-Loève and $\rho = 0.5892068$ for Tikhonov.

The estimators of the functional intercept (μ) are shown in Figure 5.11. We find again that both are similar. Additionally the residuals of each estimation method are shown in Figure 5.12. Again there is an slight improvement of the prediction of the Y_i curves over the FCVM.

In both experiments we have improved the quality of prediction and thus the understanding of the interaction between VPD and LER. Nevertheless we need to deal more carefully with some features of the data. In particular the problem of the collinearity among the VPD curves should be addressed.

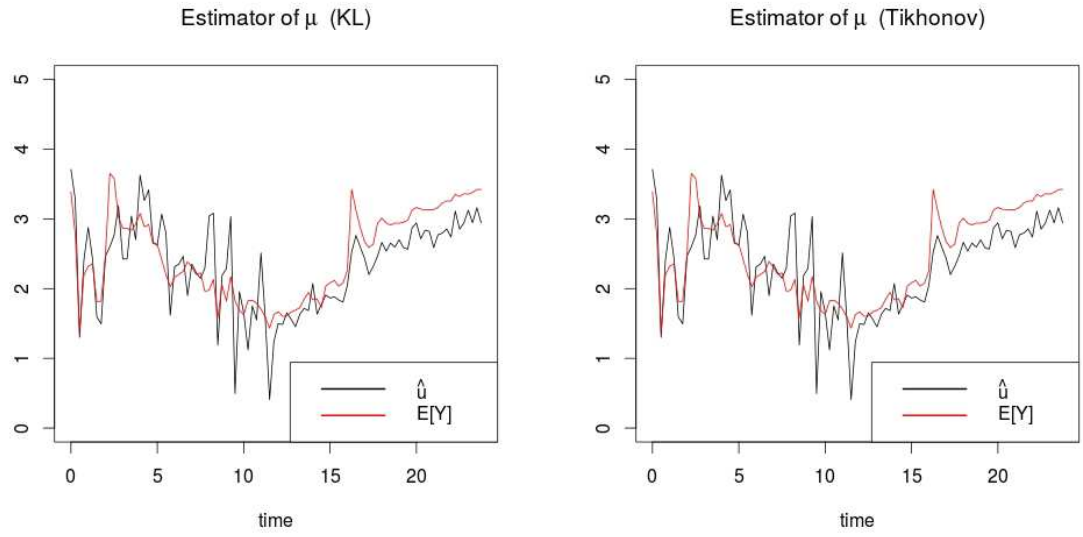

Fig. 5.8 Estimators of μ when the Karhunen-Loève and Tikhonov estimators are use to estimate $\mathscr{K}_{hist}$ in equation (5.4).

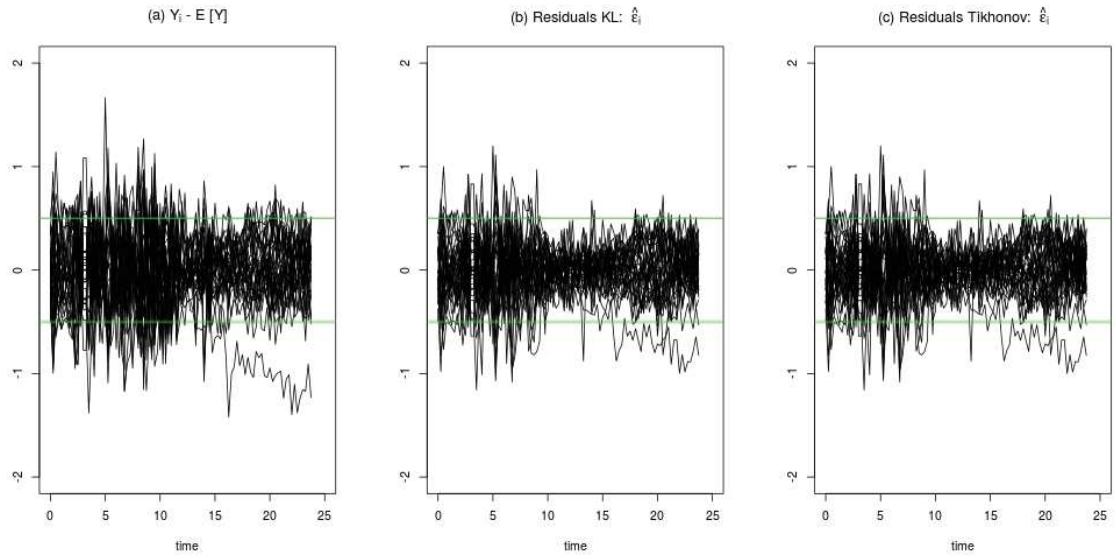

Fig. 5.9 Residuals of the estimators. Left, residuals of the empirical mean estimator ($Y_i - \bar{Y}_n$). Center, residuals of the Karhunen-Loève estimator. Right, residuals of the Tikhonov functional estimator. In all the pictures we plot green lines (constant values -0.5 and 0.5 respectively) to help the comparison.

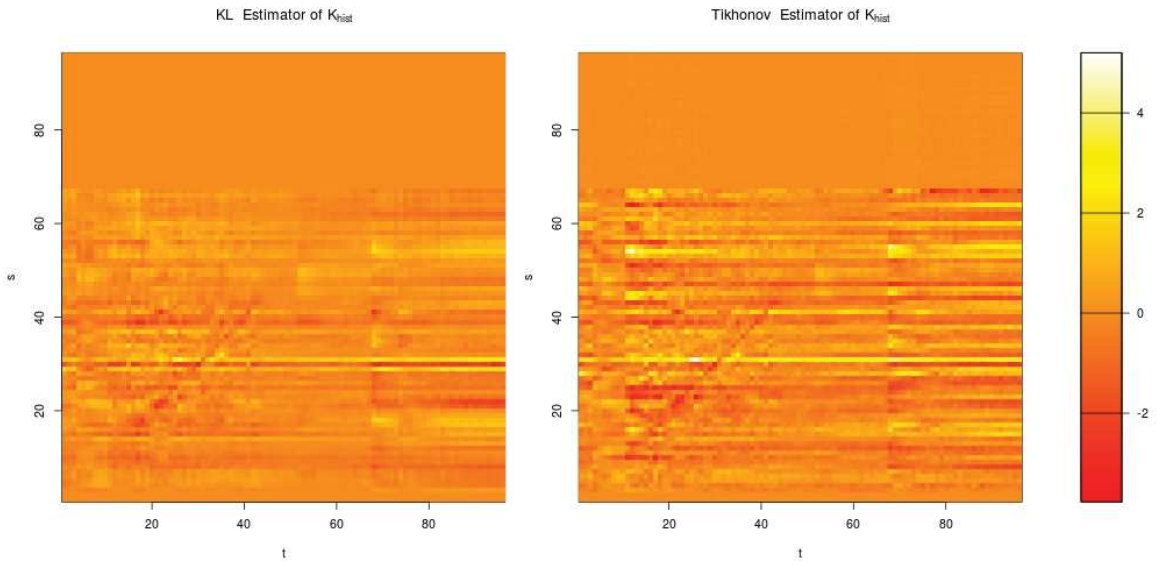

Fig. 5.10 Karhunen-Loève and Tikhonov functional estimators of the historical kernel (K_{hist}).

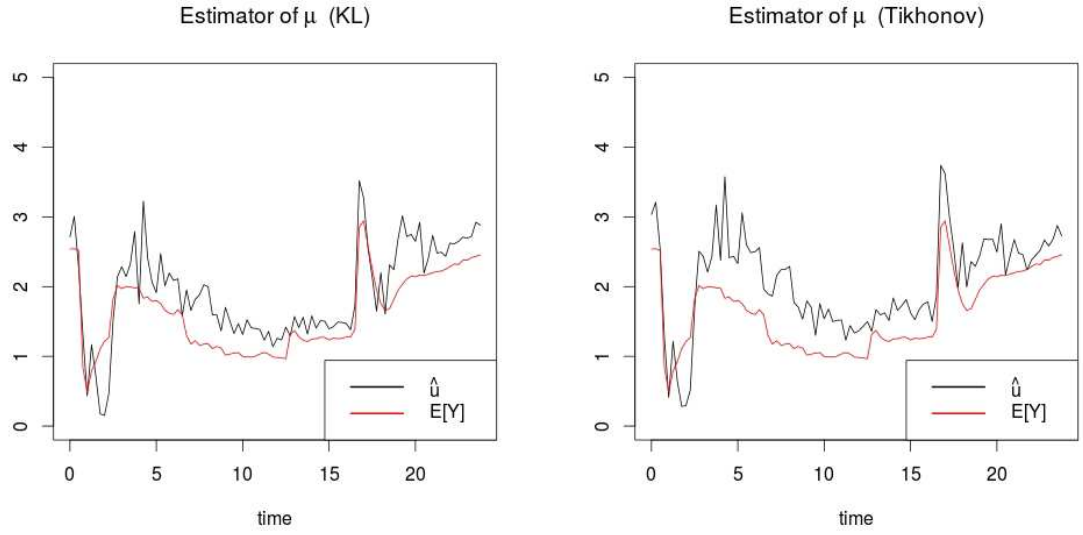

Fig. 5.11 Estimators of μ when the Karhunen-Loève and Tikhonov estimators are used to estimate $\mathscr{K}_{hist}$ in equation (5.4).

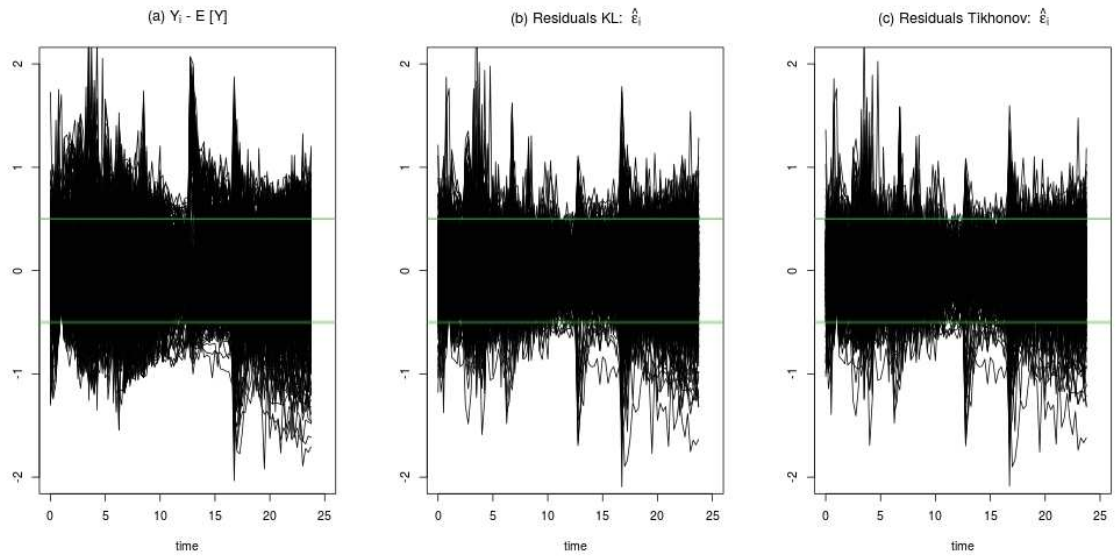

Fig. 5.12 Residuals of the estimators. Left, residuals of the empirical mean estimator ($Y_i - \bar{Y}_n$). Center, residuals of the Karhunen-Loève estimator. Right, residuals of the Tikhonov functional estimator.In all the pictures we plot green lines (constant values -0.5 and 0.5 respectively) to help the comparison.

The objective of the following section is to deal with this question and the necessary restriction on the estimators to follow the historical restriction: “the future does not influence the past”.

5.4 Collinearity and Historical Restriction

Collinearity: In both experiments (**T72A** and **T73A**), the VPD curves are repeated many times. In order to avoid collinearity and identifiability issues we have extracted different VPD curves. After this we have 10 VPD and LER curves for the experiment **T72A** and 40 for **T73A**. These curves have been reconstructed with the **R** function **approx** (linear method) and then saved into the **R** data-frames. We show these curves in Figure 5.13.

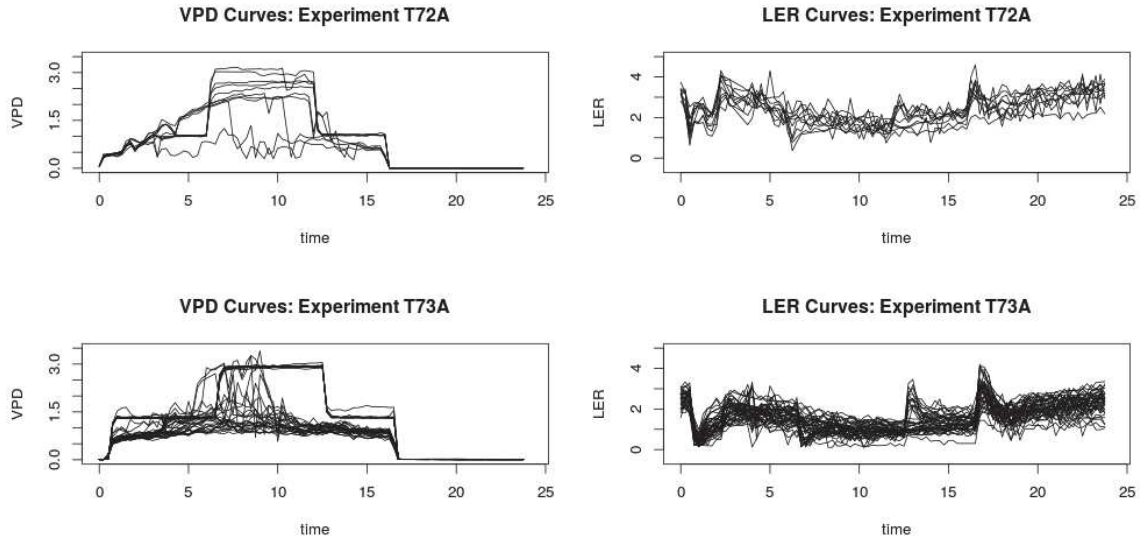

Fig. 5.13 VPD and LER curves from the experiments **T72A** and **T73A** which are not collinear.

Historical Restriction: By this restriction we mean that "the future does not influence the past". To implement this in the kernel estimation methods we must project the estimators onto the subspace where $\mathscr{K}_{hist}$ in equation (5.4) satisfies $\mathscr{K}_{hist}(s,t) = 0$ for all $s > t$.

The results of the estimation are shown in the following two subsections.

5.4.1 Estimation with Experiment T72A

The Karhunen-Loève and Tikhonov estimators of $\mathscr{K}_{hist}$ and their corresponding functional intercepts μ are shown in Figure 5.14. Both use the same calibration of parameters as the one used in section 5.3, namely the generalized cross-validation and the k-fold predictive cross-validation. The optimal parameters were: $d = 24$ and $\rho = 4.947984e-05$ for Karhunen-Loève and Tikhonov respectively.

We can see that both estimators of $\mathscr{K}_{hist}$ are similar. Each of these estimators have some rows with almost constant values (s fixed). This can be interpreted as that the influence of VPD at time

s_1 over LER at each time $t > s_1$ remains almost the same (constant). Additionally note that the μ estimators are too wavy which makes harder the interpretation of the results.

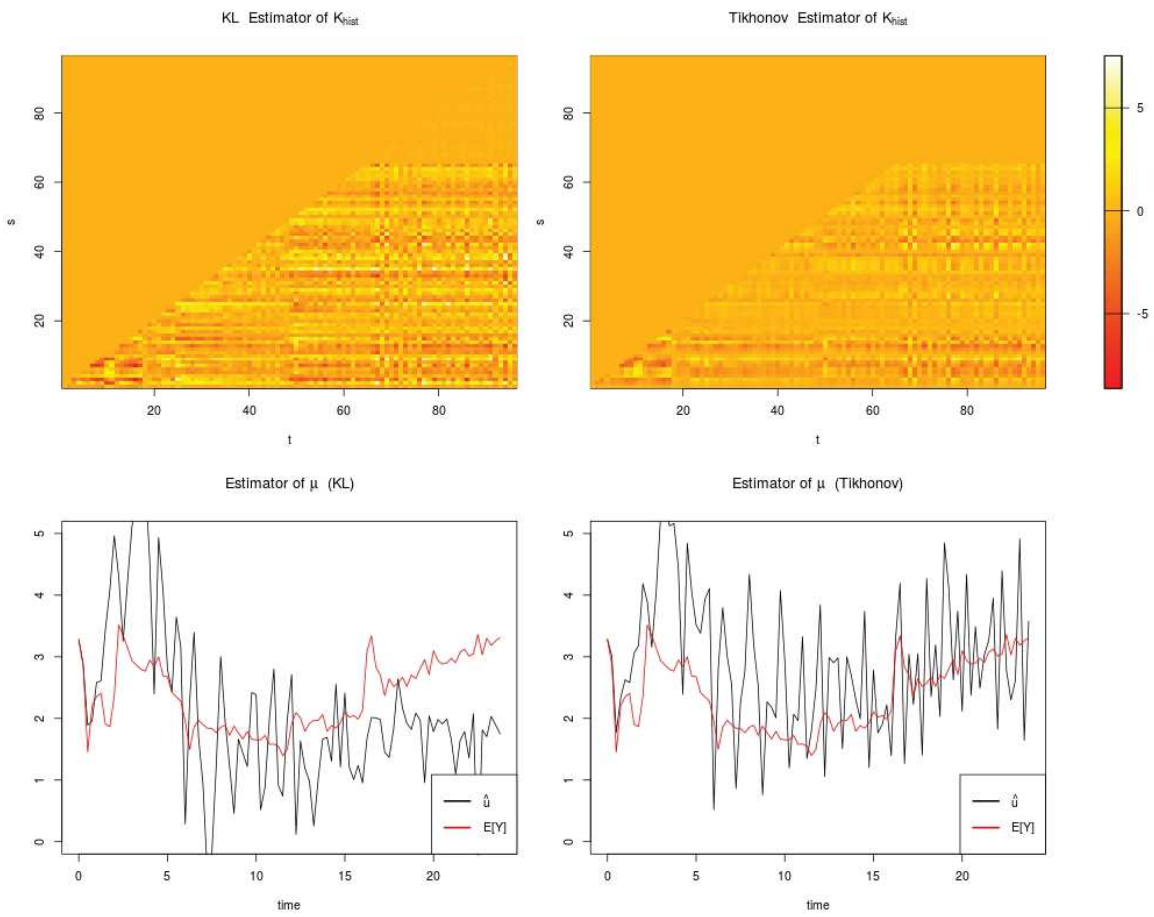

Fig. 5.14 Top left and right: Karhunen-Loève and Tikhonov functional estimators of the historical kernel (K_{hist}) for the experiment **T72A**. These two estimators satisfy the historical restriction. Bottom left and right: Estimators of μ when the Karhunen-Loève and Tikhonov estimators are used to estimate $\mathscr{K}_{hist}$ in equation (5.4).

Finally the residuals are shown in Figure 5.15. We see there that the prediction of Y_i improves greatly after 15 hours. This improvement is due to the non-collinearity of the VPD curves and the invertibility of the covariance matrix. To see this clearly note that the prediction starts to be 'perfect' precisely when the support of VPD ends.

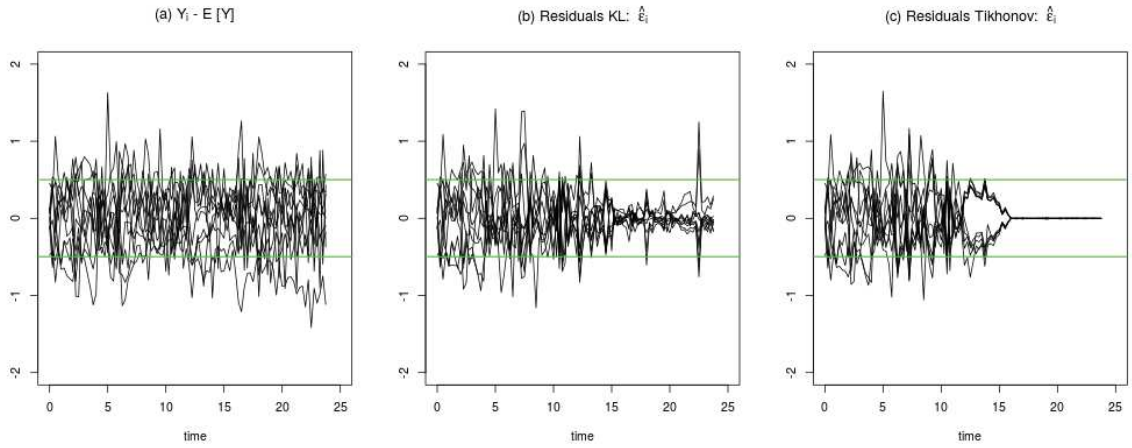

Fig. 5.15 Residuals of the estimators for the experiment **T72A**. Left, residuals of the empirical mean estimator ($Y_i - \bar{Y}_n$). Center, residuals of the Karhunen-Loève estimator. Right, residuals of the Tikhonov functional estimator. In all the pictures we plot green lines (constant values -0.5 and 0.5 respectively) to help the comparison.

5.4.2 Estimation with Experiment T73A

The Karhunen-Loève and Tikhonov estimators of $\mathscr{K}_{hist}$ and their corresponding functional intercepts μ are shown in Figure 5.16. The optimal parameters in this case are $d = 3$ and $\rho = 0.005880569$ for Karhunen-Loève and Tikhonov respectively.

In this case both estimators of $\mathscr{K}_{hist}$ differ a lot, the Karhunen-Loève estimator being close to zero compared to Tikhonov. Nevertheless this difference is due to a numerical instability in the computation of the generalized inverse $\hat{\Gamma}_\rho^+$ of the covariance operator (see equation 5.5). In this way when ρ increase to $\rho = 10$ we obtain similar matrices and again with the same structure.

The Tikhonov estimator still contains parallel rows (s fixed) and is similar to the estimator for the experiment **T72A**. Additionally note that the μ estimators are less wavy than those for **T72A**.

Finally the residuals are shown in Figure 5.17. In this case, although the prediction of Y_i improves over the mean empirical estimator, this improvement is not as important as for the experiment **T72A**.

Conclusions: The historical functional model seems to predict better the LER curves than the FCVM. For this reason it could be more useful to understand how the VPD influences the LER. The estimators of the historical kernel $\mathscr{K}_{hist}$ in both experiments have a similar structure. In particular we note the almost constant rows in each of them. This may suggests that the effect of the VPD on the LER remains almost constant over time. Finally in order to have a better assessment of this result, it would be interesting to compare it with functional non-parametric estimation methods.

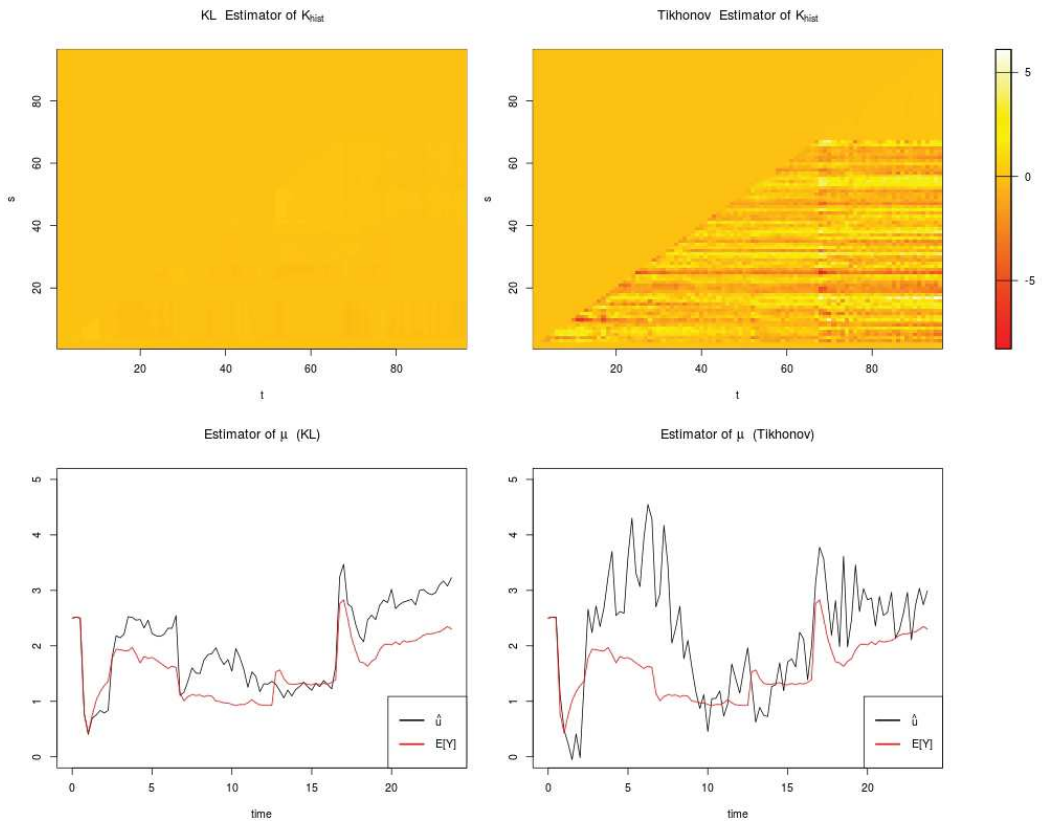

Fig. 5.16 Top left and right: Karhunen-Loève and Tikhonov functional estimators of the historical kernel (K_{hist}) for the experiment **T73A**. These two estimators satisfy the historical restriction. Bottom left and right: Estimators of μ when the Karhunen-Loève and Tikhonov estimators are used to estimate $\mathscr{K}_{hist}$ in equation (5.4).

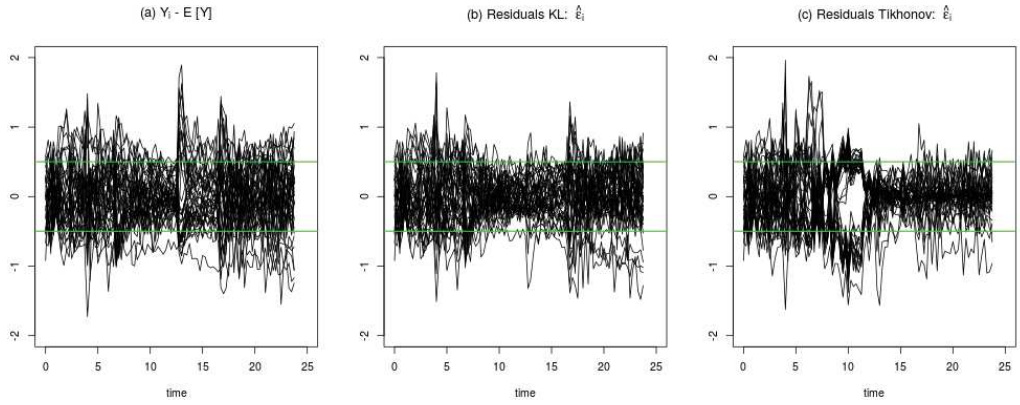

Fig. 5.17 Residuals of the estimators for the experiment **T73A**. Left, residuals of the empirical mean estimator ($Y_i - \bar{Y}_n$). Center, residuals of the Karhunen-Loève estimator. Right, residuals of the Tikhonov functional estimator. In all the pictures we plot green lines (constant values -0.5 and 0.5 respectively) to help the comparison.

Chapter 6

Conclusions and Perspectives

6.1 General Conclusions

This thesis has contributed to the study of how the history of the functional regressor X influences the current value of the functional response Y in functional linear regression models with functional response. In this regard, we have studied the theoretical and practical questions about the estimation for the following models:

1. The Functional Concurrent Model (FCCM), where only the instantaneous action is considered (Chapter 2).

2. The Functional Convolution Model (FCVM), where a fixed historical functional coefficient is used (Chapter 3).

3. The fully functional model, where we were interested in the estimation of the noise covariance operator (Chapter 4).

For the FCVM and the FCCM, the consistency and a rate of convergence were obtained, along with the numerical study of the robustness of the estimators. Additionally the shorter computation time of both estimators compared to others from the literature has also been shown.

Finally in Chapter 5 we apply these models and also the historical functional model to study how the Vapour Pressure Deficit (VPD) influences the Leaf Elongation Rate (LER) with a real dataset. This is a starting point for future research.

6.2 Perspectives

There are still many questions to be studied in future research. Here we outline some of them.

- The optimal rate of convergence of the functional Ridge regression estimator (2.3) and the functional Fourier deconvolution (3.4) are still unknown. One way to deal with this question is by considering estimators with other types of penalization like thresholding. This could give better theoretical properties but maybe with numerical instabilities.

- We can use the FCVM or the historical functional model in the context of a functional ANCOVA model where a qualitative is introduced a genotype factor for example.

  In this way, for instance the FCVM (1.3) will generalize as follows. For $t \in [0,\infty[$, $j \in \{1,\cdots,J\}$ and $k \in \{1,\cdots,n_j\}$ (replications)

  $$Y_{jk}(t) = \mu_j(t) \; + \; \int_0^t \theta_j(s) X_{jk}(t-s)ds \; + \; \varepsilon_{jk}(t).$$

  Potentially these functional ANCOVA models will be useful to differentiate and compare the VPD and LER interaction among different genotypes.

- The introduction of more functional covariates which have an instantaneous or historical influence over the response variable is an important generalization of the models studied in this thesis. For instance the following model: for $i \in \{1,\cdots,n\}$ and $t \in [0,T]$,

  $$Y_i(t) = \mu(t) + \beta(t)X_{1,i} + \int_0^t \mathscr{K}_{hist}(t,s)X_{2,i} + \varepsilon_i(t),$$

  where X_1 and X_2 are two functional covariates which influence Y in a different way.

- The historical functional model applied to the VPD and LER interaction has shown that the estimator of the historical kernel ($\mathscr{K}_{hist}$) has a structure that might be interpreted such that the influence of VPD at time s_1 over LER at each time $t > s_1$ remains almost the same (rows with almost constant values). This interpretation might be useful but it would be interesting to compare this result with functional non-parametric estimation methods (Ferraty and Vieu (2006)) to better understand this structure.

References

Abramovich, F. u. and Silverman, B. (1998). Wavelet decomposition approaches to statistical inverse problems. *Biometrika*, 85(1):115–129.

Aguilera, A., Ocaña, F., and Valderrama, M. (2008). Estimation of functional regression models for functional responses by wavelet approximation. In *Functional and Operatorial Statistics*, pages 15–21. Springer.

Antoch, J., Prchal, L., Rosaria De Rosa, M., and Sarda, P. (2010). Electricity consumption prediction with functional linear regression using spline estimators. *Journal of Applied Statistics*, 37(12):2027–2041.

Asencio, M., Hooker, G., and Gao, H. O. (2014). Functional convolution models. *Statistical Modelling*, page 1471082X13508262.

Ash, R. and Gardner, M. (1975). *Topics in Stochastic Processes: By Robert B. Ash and Melvin F. Gardner*. Probability and mathematical statistics. Academic Press.

Bickel, P. J. and Levina, E. (2004). Some theory for fisher's linear discriminant function,'naive bayes', and some alternatives when there are many more variables than observations. *Bernoulli*, pages 989–1010.

Bloomfield, P. (2004). *Fourier Analysis of Time Series: An Introduction*. Wiley Series in Probability and Statistics. Wiley.

Bosq, D. (2000). *Linear Processes in Function Spaces: Theory and Applications*, volume 149 of *Lectures Notes in Statistics*. Springer-Verlag, New York.

Brezis, H. (2010). *Functional analysis, Sobolev spaces and partial differential equations*. Springer Science & Business Media.

Brown, R. and Hwang, P. (2012). *Introduction to Random Signals and Applied Kalman Filtering with MATLAB Exercises*. John Wiley & Sons., fourth edition.

Bühlmann, P. and van de Geer, S. (2011). *Statistics for High-Dimensional Data: Methods, Theory and Applications*. Springer Series in Statistics. Springer Berlin Heidelberg.

Cai, Z., Fan, J., and Li, R. (2000). Efficient estimation and inferences for varying-coefficient models. *Journal of the American Statistical Association*, 95(451):888–902.

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. (2003). Spline estimators for the functional linear model. *Statistica Sinica*, pages 571–591.

Chiou, J.-M., Müller, H.-G., and Wang, J.-L. (2004). Functional response models. *Statistica Sinica*, pages 675–693.

Comte, F., Cuenod, C.-A., Pensky, M., and Rozenholc, Y. (2016). Laplace deconvolution on the basis of time domain data and its application to dynamic contrast enhanced imaging. *arXiv preprint arXiv:1405.7107*.

Cooley, J. W. and Tukey, J. W. (1965). An algorithm for the machine calculation of complex fourier series. *Mathematics of computation*, 19(90):297–301.

Crambes, C. and Mas, A. (2013). Asymptotics of prediction in functional linear regression with functional outputs. *Bernoulli*, 19(5B):2627–2651.

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.

De Canditiis, D. and Pensky, M. (2006). Simultaneous wavelet deconvolution in periodic setting. *Scandinavian Journal of Statistics*, 33(2):293–306.

Donoho, D. L. (1995). Nonlinear solution of linear inverse problems by wavelet–vaguelette decomposition. *Applied and computational harmonic analysis*, 2(2):101–126.

Dreesman, J. M. and Tutz, G. (2001). Non-stationary conditional models for spatial data based on varying coefficients. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 50(1):1–15.

Fan, J., Yao, Q., and Cai, Z. (2003). Adaptive varying-coefficient linear models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 65(1):57–80.

Fan, J. and Zhang, J.-T. (2000). Two-step estimation of functional linear models with applications to longitudinal data. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 62(2):303–322.

Fan, J. and Zhang, W. (2008). Statistical methods with varying coefficient models. *Statistics and its Interface*, 1(1):179.

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

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

Gasser, T. and Kneip, A. (1995). Searching for structure in curve samples. *Journal of the american statistical association*, 90(432):1179–1188.

Gonzalez, R.C., W. R. and Eddins, S. (2009). *Digital Image Processing Using MATLAB*. Gatesmark Publishing, United States., second edition.

Green, P. J. and Silverman, B. W. (1994). *Nonparametric regression and generalized linear models: a roughness penalty approach*. Chapman & Hall / CRC Press.

Greene, W. H. (2003). *Econometric analysis, 5th*. Prentice Hall, Ed.. Upper Saddle River, NJ, sixth edition.

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.

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.

Hassanieh, H., Indyk, P., Katabi, D., and Price, E. (2012). Nearly optimal sparse fourier transform. In *Proceedings of the forty-fourth annual ACM symposium on Theory of computing*, pages 563–578. ACM.

Hastie, T. and Tibshirani, R. (1993). Varying-coefficient models. *Journal of the Royal Statistical Society. Series B (Methodological)*, 55(4):757–796.

He, G., Müller, H., and Wang, J. (2000). Extending correlation and regression from multivariate to functional data. *Asymptotics in statistics and probability*, pages 197–210.

Hoerl, A. E. (1962). Application of ridge analysis to regression problems. *Chemical Engineering Progress*, 58(3):54–59.

Hoerl, A. E. and Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12(1):55–67.

Horváth, L. and Kokoszka, P. (2012). *Inference for Functional Data with Applications*, volume 200 of *Springer Series in Statistics*. Springer, New York.

Hsing, T. and Eubank, R. (2015). *Theoretical Foundations of Functional Data Analysis, with an Introduction to Linear Operators*. Wiley Series in Probability and Statistics. John Wiley & Sons, Ltd., Chichester.

Huang, J. Z., Wu, C. O., and Zhou, L. (2004). Polynomial spline estimation and inference for varying coefficient models with longitudinal data. *Statistica Sinica*, pages 763–788.

Huh, M.-H. and Olkin, I. (1995). Asymptotic aspects of ordinary ridge regression. *American Journal of Mathematical and Management Sciences*, 15(3-4):239–254.

James, G. M. (2002). Generalized linear models with functional predictors. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64(3):411–432.

Johannes, J. et al. (2009). Deconvolution with unknown error distribution. *The Annals of Statistics*, 37(5A):2301–2323.

Johnson, R. and Wichern, D. (2007). *Applied Multivariate Statistical Analysis*. Applied Multivariate Statistical Analysis. Pearson Prentice Hall.

Johnstone, I. M., Kerkyacharian, G., Picard, D., and Raimondo, M. (2004). Wavelet deconvolution in a periodic setting. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 66(3):547–573.

Kadri, H., Duflos, E., Preux, P., Canu, S., Davy, M., et al. (2010). Nonlinear functional regression: a functional rkhs approach. In *AISTATS*, volume 10, pages 111–125.

Kammler, D. (2008). *A First Course in Fourier Analysis*. Cambridge University Press.

Kim, K., Şentürk, D., and Li, R. (2011). Recent history functional linear models for sparse longitudinal data. *Journal of statistical planning and inference*, 141(4):1554–1566.

Kulik, R., Sapatinas, T., and Wishart, J. R. (2015). Multichannel deconvolution with long range dependence: Upper bounds on the lp-risk. *Applied and Computational Harmonic Analysis*, 38(3):357–384.

Ledoux, M. and Talagrand, M. (1991). *Probability in Banach Spaces, Isoperimetry and Processes*, volume 23 of *A Series of Modern Surveys in Mathematics Series*. Springer-Verlag, Berlin.

Lian, H. (2007). Nonlinear functional models for functional responses in reproducing kernel hilbert spaces. *Canadian Journal of Statistics*, 35(4):597–606.

Malfait, N. and Ramsay, J. O. (2003). The historical functional linear model. *Canadian Journal of Statistics*, 31(2):115–128.

Manrique, T., Crambes, C., and Hilgert, N. (2016). Ridge regression for the functional concurrent model. *arXiv preprint arXiv:7777.7777*.

Mas, A. and Pumo, B. (2009). Functional linear regression with derivatives. *Journal of Nonparametric Statistics*, 21(1):19–40.

Meister, A. (2009). *Deconvolution Problems in Nonparametric Statistics*, volume 193 of *Lecture Notes in Statistics*. Springer Science & Business Media.

Morris, J. S. (2015). Functional regression. *Annual Review of Statistics and its Applications Vol. 2*.

Müller, H.-G. and Yao, F. (2012). Functional additive models. *Journal of the American Statistical Association*.

Oppenheim, A. and Schafer, R. (2011). *Discrete-Time Signal Processing*. Pearson Education.

O'Sullivan, F. (1986). A statistical perspective on ill-posed inverse problems. *Statistical science*, pages 502–518.

Pensky, M., Sapatinas, T., et al. (2010). On convergence rates equivalency and sampling strategies in functional deconvolution models. *The Annals of Statistics*, 38(3):1793–1844.

Pinsky, M. (2002). *Introduction to Fourier Analysis and Wavelets*. Graduate studies in mathematics. American Mathematical Society.

Ramsay, J., Hooker, G., and Graves, S. (2009). *Functional Data Analysis with R and MATLAB*. Use R! Springer New York.

Ramsay, J. O. and Dalzell, C. (1991). Some tools for functional data analysis. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 539–572.

Ramsay, J. O. and Silverman, B. W. (2005). *Functional data analysis*. Springer, New York, second edition.

Seni, G. and Elder, J. (2010). *Ensemble Methods in Data Mining: Improving Accuracy Through Combining Predictions*. Synthesis lectures on data mining and knowledge discovery. Morgan & Claypool Publishers.

Şentürk, D. and Müller, H.-G. (2010). Functional varying coefficient models for longitudinal data. *Journal of the American Statistical Association*, 105(491):1256–1264.

Tikhonov, A. and Arsenin, V. (1977). *Solutions of ill-posed problems*. Scripta series in mathematics. Winston.

Ullah, S. and Finch, C. F. (2013). Applications of functional data analysis: A systematic review. *BMC medical research methodology*, 13(1):1.

Wahba, G. (1990). *Spline Models for Observational Data*. CBMS-NSF Regional Conference Series in Applied Mathematics. Society for Industrial and Applied Mathematics (SIAM, 3600 Market Street, Floor 6, Philadelphia, PA 19104).

Wang, J.-L., Chiou, J.-M., and Müller, H.-G. (2016). Functional data analysis. *Annual Review of Statistics and Its Application*, 3(1):257–295.

West, M., Harrison, P. J., and Migon, H. S. (1985). Dynamic generalized linear models and bayesian forecasting. *Journal of the American Statistical Association*, 80(389):73–83.

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.

Yao, F., Müller, H.-G., and Jane-Ling, W. (2005a). Functional linear regression analysis for longitudinal data. *The Annals of Statistics*, 33(6):2873–2903.

Yao, F., Müller, H.-G., and Wang, J.-L. (2005b). Functional data analysis for sparse longitudinal data. *Journal of the American Statistical Association*, 100(470):577–590.

Zhang, W. and Lee, S.-Y. (2000). Variable bandwidth selection in varying-coefficient models. *Journal of Multivariate Analysis*, 74(1):116–134.

Zhang, W., Lee, S.-Y., and Song, X. (2002). Local polynomial fitting in semivarying coefficient model. *Journal of Multivariate Analysis*, 82(1):166–188.

Zhu, H., Fan, J., and Kong, L. (2014). Spatially varying coefficient model for neuroimaging data with jump discontinuities. *Journal of the American Statistical Association*, 109(507):1084–1098.

Functional Linear Regression Models. Application to High-throughput Plant Phenotyping Functional Data.

Functional data analysis (FDA) is a statistical branch that is increasingly being used in many applied scientific fields such as biological experimentation, finance, physics, etc. A reason for this is the use of new data collection technologies that increase the number of observations during a time interval. Functional datasets are realization samples of some random functions which are measurable functions defined on some probability space with values in an infinite dimensional functional space. There are many questions that FDA studies, among which functional linear regression is one of the most studied, both in applications and in methodological development.

The objective of this thesis is the study of functional linear regression models when both the covariate X and the response Y are random functions and both of them are time-dependent. In particular we want to address the question of how the history of a random function X influences the current value of another random function Y at any given time t. In order to do this we are mainly interested in three models: the functional concurrent model (FCCM), the functional convolution model (FCVM) and the historical functional linear model. In particular for the FCVM and FCCM we have proposed estimators which are consistent, robust and which are faster to compute compared to others already proposed in the literature. Our estimation method in the FCCM extends the Ridge Regression method developed in the classical linear case to the functional data framework. We prove the probability convergence of this estimator, obtain a rate of convergence and develop an optimal selection procedure of the regularization parameter. The FCVM allows to study the influence of the history of X on Y in a simple way through the convolution. In this case we use the continuous Fourier transform operator to define an estimator of the functional coefficient. This operator transforms the convolution model into a FCCM associated in the frequency domain. The consistency and rate of convergence of the estimator are derived from the FCCM. The FCVM can be generalized to the historical functional linear model, which is itself a particular case of the fully functional linear model. Thanks to this we have used the Karhunen–Loève estimator of the historical kernel. The related question about the estimation of the covariance operator of the noise in the fully functional linear model is also treated. Finally we use all the aforementioned models to study the interaction between Vapour Pressure Deficit (VPD) and Leaf Elongation Rate (LER) curves. This kind of data is obtained with high-throughput plant phenotyping platform and is well suited to be studied with FDA methods.

Keywords : *Functional regression models*, *Functional data*, *Convolution Model*, *Concurrent Model*, *Historical Model*.