THÈSE

présentée en vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ PAUL SABATIER TOULOUSE III

Discipline : Mathématiques
Spécialité : Statistique

par

Christophe Crambes

Modèles de régression linéaire pour variables explicatives fonctionnelles

---

Directeurs de thèse : Hervé Cardot et Pascal Sarda

Soutenue le 23 novembre 2006 devant le jury composé de Messieurs :

Benoît Cadre	Université Montpellier II	Rapporteur
Hervé Cardot	CESAER - ENESAD INRA Dijon	Directeur
Antonio Cuevas	Universidad Autónoma de Madrid	Rapporteur
Frédéric Ferraty	Université Paul Sabatier	Examinateur
Alois Kneip	Universität Bonn	Examinateur
Pascal Sarda	Université Paul Sabatier	Directeur

Laboratoire de Statistique et Probabilités
UMR CNRS 5583, Université Paul Sabatier, Toulouse III

Mémoire de thèse de doctorat

Modèles de régression linéaire pour variables explicatives fonctionnelles

Christophe Crambes

REMERCIEMENTS

Je voudrais tout d'abord remercier Pascal SARDA et Hervé CARDOT pour avoir accepté d'encadrer ma thèse. Je tiens à les remercier de m'avoir accordé leur confiance depuis l'année de mon DESS, où ils ont encadré mon stage de fin d'année, et m'ont alors encouragé à poursuivre en DEA et en thèse. Ils ont toujours fait preuve d'une très grande disponibilité à mon égard et je réalise aujourd'hui à quel point travailler avec eux a été enrichissant.

Je tiens ensuite à remercier Benoît CADRE et Antonio CUEVAS pour avoir accepté d'être les rapporteurs de cette thèse. Je suis très flatté de l'intérêt qu'ils ont porté à ce travail. Leur relecture attentive du manuscript ainsi que leurs remarques pertinentes ont contribué à améliorer la version finale de ce document.

Je suis très heureux qu'Alois KNEIP ait accepté de faire partie de mon jury. Les deux séjours au cours desquels il m'a accueilli à Mayence puis à Bonn m'ont énormément apporté aussi bien sur le plan mathématique que sur le plan humain, et les travaux que nous avons en cours vont nous permettre de poursuivre notre collaboration au-delà de cette thèse.

Je souhaite également remercier Frédéric FERRATY de faire partie de mon jury. Il m'a toujours apporté de bons conseils par rapport à mon travail, et j'ai toujours pu trouver la porte de son bureau ouverte chaque fois que j'en ai eu besoin.

Je voudrais à présent remercier les professeurs du Laboratoire de Statistique et Probabilités que j'ai pu cotoyer pendant ces trois années de thèse,

notamment les membres du groupe travail STAPH : Philippe Vieu, Yves Romain, Alain Boudou, Sylvie Viguier, et Luboš que j'ai plaisir à voir chaque fois qu'il revient à Toulouse. Je remercie également Fabrice Gamboa pour la confiance qu'il m'a accordée pendant mon année de DEA, je garde un très bon souvenir de son enseignement. Je souhaite également remercier Anne Ruiz-Gazen, que je connais mieux depuis la 1ère Rencontre des Jeunes Statisticiens à Aussois. Elle s'est montrée très disponible pour répondre à mes questions et discuter avec elle m'a permis d'envisager de nouvelles pistes de recherche.

Je tiens aussi à remercier particulièrement Françoise Michel pour sa disponibilité, sa bonne humeur et son efficacité pour prendre en charge les problèmes administratifs que l'on rencontre au quotidien.

Ces trois années de thèse m'ont également permis de rencontrer des doctorants avec qui je passe de très bons moments. Les doctorants arrivés l'an dernier, Maxime, Laurent, Florent et Amélie, ont apporté leur bonne humeur pendant la pause quizz de midi. Mes remerciements vont aussi aux doctorants arrivés en thèse la même année que moi ou l'année suivante, qui vont me laisser de très bons souvenirs : Delphine (avec qui ça a été un plaisir de partager le bureau ces deux dernières années), Marielle, Agnès, Solenn, Myriam et Diana. Je ne saurais oublier les doctorants qui m'ont accueilli à mon arrivée en thèse, et tous les bons moments qu'on a passé : Renaud, Clément, Cécile, Yan, Élie, Abdelaâti, Nicolas et Jean-Pierre. Je souhaite aussi remercier Sébastien, à qui j'ai posé d'innombrables questions sur LaTeX, sur **R**, et je retiens avant tout sa disponibilité et sa bonne humeur. Enfin, je connais Lionel depuis le DEA et on partage le même bureau depuis le début de notre thèse, et je tiens à lui dire quel plaisir j'ai eu de pouvoir faire ma thèse en même temps que lui, pour tous les bons moments passés pendant ces années.

Enfin, je voudrais remercier ma famille, plus particulièrement mes parents qui m'ont toujours soutenu dans les études et qui m'ont permis de les réaliser dans les meilleures conditions possibles. Je remercie aussi ma sœur Magali, ainsi que Marc, Julie et Anthony pour tous les moments qu'on passe ensemble chaque fois que je reviens à Perpignan. Enfin, pour tout ce qu'elle m'apporte, je remercie Marine.

TABLE DES MATIÈRES

INTRODUCTION

La statistique fonctionnelle a connu un très important développement ces dernières années. Cette branche de la statistique vise a étudier des données qui, de part leur structure et le fait qu'elles soient collectées sur des grilles très fines, peuvent être assimilées à des courbes ou à des surfaces, par exemple fonctions du temps ou de l'espace. Le besoin de considérer ce type de données, maintenant couramment rencontré sous le nom de *données fonctionnelles* dans la littérature, est avant tout un besoin pratique. Compte tenu des capacités actuelles des appareils de mesure et de stockage informatique, les situations pouvant fournir de telles données sont multiples et issues de domaines variés : on peut imaginer par exemple des courbes de croissance, de température, des images observées par satellite, ... Donner une liste exhaustive des situations où de telles données sont rencontrées n'est pas envisageable, mais des exemples précis de données fonctionnelles seront abordés dans cette thèse.

Cependant, au-delà de cet aspect pratique, il est nécessaire de donner un cadre théorique pour l'étude de ces données. Bien que la statistique fonctionnelle ait les mêmes objectifs que les autres branches de la statistique (analyse de données, inférence, ...), les données ont cette particularité de prendre leurs valeurs dans des espaces de fonctions, et les méthodes usuelles de la statistique multivariée sont ici mises en défaut. Par exemple, considérons que l'on dispose des observations de n courbes en p points de discrétisation, ces courbes étant utilisées comme prédicteur d'une autre variable. Si on regroupe ces données notées x_{ij} (pour i allant de 1 à n et j allant de 1 à p) sous forme d'une matrice de taille $n \times p$,

$$\mathbf{X} = \begin{pmatrix} x_{11} & \dots & x_{1p} \\ \vdots & & \vdots \\ x_{n1} & \dots & x_{np} \end{pmatrix},$$

la méthode des moindres carrés ordinaires, très courante en statistique multivariée, peut donner de très mauvais résultats dans cette situation, puisque cette méthode amène à l'inversion de la matrice $\mathbf{X}^\tau\mathbf{X}$ qui peut se révéler difficile voire impossible pour deux raisons. La première est que p est généralement grand (on peut même avoir $p > n$, et ainsi une matrice $\mathbf{X}^\tau\mathbf{X}$ non inversible). La seconde raison est qu'il y a de fortes chances d'avoir une colinéarité importante entre les p prédicteurs du fait qu'ils sont les points de mesure d'une même fonction. Pour contourner ce problème, des solutions ont été envisagées, les plus courantes étant

- la "ridge regression" (introduite initialement par Hoerl et Kennard, 1980), qui consiste à ajouter un terme de pénalisation dans le critère des moindres carrés. Cela amène à inverser la matrice $(\mathbf{X}^\tau \mathbf{X} + \lambda \mathbf{I}_p)$ (avec λ réel strictement positif et $\mathbf{I}_p$ matrice identité de taille p) au lieu de $\mathbf{X}^\tau \mathbf{X}$,
- la régression sur composantes principales, qui consiste à réduire la dimension p en utilisant les k premières composantes principales issues de l'analyse en composantes principales du tableau $\mathbf{X}$ (avec k entier non nul "convenablement" choisi),
- la régression "partial least squares" (voir Helland, 1990), qui est une méthode algorithmique basée à chaque étape sur la régression par moindres carrés ordinaires sur les résidus de l'étape précédente.

Ces différentes méthodes sont étudiées et comparées dans un article de Frank et Friedman (1993), en vue de leur application dans le domaine de la chimiométrie. Comme souligné par Hastie et Mallows (1993) dans leur discussion de cet article, l'approche qui consiste à voir une courbe uniquement à travers un vecteur de points de mesure est réductrice, ne serait-ce que par le fait que les points de mesure doivent être les mêmes pour chaque courbe observée, ce qui n'est pas forcément le cas en pratique. Cette approche conduit également au problème que l'on perd la structure de courbe si on utilise uniquement les mesures de la courbe en certains points. C'est pourquoi il paraît préférable de traiter les données en tenant compte de leur nature fonctionnelle.

Les tout premiers travaux dans lesquels on retrouve cette idée de données fonctionnelles sont finalement relativement "anciens". Rao (1958) et Tucker (1958) envisagent ainsi l'analyse en composantes principales et l'analyse factorielle pour des données fonctionnelles et considèrent même explicitement les données fonctionnelles comme un type particulier de données. Par la suite, Ramsay (1982) dégage la notion de données fonctionnelles et soulève la question de l'adaptation des méthodes de la statistique multivariée à ce cadre fonctionnel. À partir de là, les travaux pour explorer la statistique fonctionnelle commencent à se multiplier, pour finalement aboutir aujourd'hui à des ouvrages faisant référence en la matière, comme par exemple les monographies de Ramsay et Silverman (2002 et 2005).

Les travaux réalisés constituent ainsi à l'heure actuelle une littérature très dense, que ce soit sur un plan théorique ou appliqué. D'un point de vue théorique, la notion de *variable aléatoire fonctionnelle* (c'est-à-dire une variable aléatoire à valeurs dans un espace de fonctions) est apparue, ce qui a nécessité d'expliciter des notions simples pour une telle variable, comme par exemple l'espérance ou l'opérateur de covariance. Des techniques déjà existantes en statistique multivariée (comme par exemple l'analyse en composantes principales) peuvent alors se développer dans ce cadre fonctionnel, utilisant notamment les connaissances en théorie des opérateurs dans les espaces de Hilbert (voir par exemple Dunford et Schwarz, 1963, Gohberg et Krein, 1971). Parmi les points de départ de cette généralisation au cadre fonctionnel des méthodes de la statistique multivariée, Deville (1974) introduit une analyse en composantes principales de courbes, tandis que la thèse de Dauxois et Pousse (1976) va au-delà de l'analyse en composantes principales dans un espace de Hilbert, s'attachant à traiter un certain nombre de méthodes regroupées sous le nom d'analyses factorielles, dont l'analyse en composantes principales fait partie. L'article de Dauxois, Pousse et Romain (1982) aborde davantage les aspects asymptotiques du problème, fournissant notamment des résultats de convergence pour l'opérateur de covariance empirique.

Dans cette thèse, on propose d'apporter une contribution à l'étude des données fonctionnelles dans le contexte où la variable fonctionnelle sert à expliquer un phénomène représenté par une autre variable. Le problème qui va nous intéresser est celui de la régression dans le cas où la variable explicative est fonctionnelle. C'est un sujet sur lequel la littérature est très conséquente. D'un point de vue très général, ce modèle de régression fonctionnelle peut s'écrire

$$Y_i = r\left(X_i\right) + \epsilon_i, \qquad i = 1, \ldots, n, \tag{1}$$

où, pour tout $i = 1, \ldots, n$,

- X_i appartient à un espace de Hilbert H dont le produit scalaire et la semi-norme associée seront notés respectivement $\langle ., . \rangle$ et $\|.\|_H$. Notons que X_i pourra être aléatoire ou non, suivant que l'on considère un modèle à plan aléatoire ou à plan fixe,
- Y_i est une variable aléatoire réelle (variable d'intérêt),
- ϵ_i est une variable aléatoire d'erreur,
- les variables aléatoires considérées sont toutes définies sur un même espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$.

Le but est alors d'estimer l'opérateur $r : H \longrightarrow \mathbb{R}$ inconnu, sur la base des données $(X_i, Y_i)_{i=1,\ldots,n}$.

Dans cette thèse, on va s'intéresser à un modèle un peu plus particulier que le modèle précédent (1). Il s'agit du *modèle linéaire fonctionnel*, qui s'écrit sous la forme

$$Y_i = \langle \alpha, X_i \rangle + \epsilon_i, \qquad i = 1, \ldots, n, \tag{2}$$

où le but est d'estimer $\alpha \in H$ inconnu, sur la base des données $(X_i, Y_i)_{i=1,\ldots,n}$. Les hypothèses faites sur $\epsilon_1, \ldots, \epsilon_n$ seront détaillées le moment venu. Nous verrons en effet que ces hypothèses diffèreront suivant les situations envisagées. D'abord décrit par Ramsay et Dalzell (1991), ce modèle est toujours l'objet de travaux récents, comme en témoignent par exemple les articles de Cardot, Ferraty et Sarda (1999, 2003) étudiant le cas d'une variable réponse réelle et donnant une méthode d'estimation de α à l'aide de ce qu'ils introduisent comme la régression sur composantes principales fonctionnelle, ou encore à l'aide de fonctions splines. Des vitesses de convergence sont également obtenues pour les estimateurs qu'ils construisent. Dans cette thèse, c'est à ce modèle (2) que l'on va s'intéresser. C'est un modèle très populaire en analyse de données fonctionnelles, il est important de noter toutefois que d'autres modèles qui lui sont liés (par exemple des extensions de ce modèle linéaire fonctionnel (2)) sont également sujets d'études récentes. Là encore, il semble impossible de lister les modèles existants relatifs à (1) et (2). On va cependant donner un tour d'horizon des modèles les plus fréquemment rencontrés.

- Il est possible de considérer que la variable d'intérêt est elle aussi de nature fonctionnelle, comme la variable explicative. Cuevas, Febrero et Fraiman (2002) ou encore Chiou, Müller et Wang (2004) se sont ainsi intéressés à ce modèle linéaire fonctionnel avec une variable réponse fonctionnelle, c'est-à-dire lorsque Y_i (pour $i = 1, \ldots, n$) appartient aussi à un espace de Hilbert. Cuevas, Febrero et Fraiman (2002) travaillent dans le cadre d'un modèle à plan fixe (c'est-à-dire que $X_1, \ldots, X_n$ sont non aléatoires), alors que Chiou, Müller et Wang (2004) considèrent $X_1, \ldots, X_n$ aléatoires, basant leur méthode d'estimation sur les décompositions de Karhunen-Loève des X_i et Y_i, $i = 1, \ldots, n$. L'article plus bibliographique de Müller (2005) passe en revue diverses méthodes d'estimation pour des modèles linéaires fonctionnels avec une variable d'intérêt réelle ou fonctionnelle et une variable explicative multidimensionnelle ou fonctionnelle. Il étend aussi ces méthodes, considérant notamment des variables explicatives dont les points de mesure peuvent être irrégulièrement espacés et en faible nombre, situation étudiée dans les articles de Yao, Müller et Wang (2005a, 2005b).

- Une autre extension possible du modèle linéaire fonctionnel est le modèle linéaire fonctionnel généralisé, qui est la version fonctionnelle du modèle linéaire généralisé introduit par Wedderburn (1974) puis repris dans un ouvrage par McCullagh et Nelder (1989). Ce modèle linéaire fonctionnel généralisé a été notamment étudié par Cardot et Sarda (2005) ainsi que Müller et Stadtmüller (2005). Dans ce modèle, on suppose que la loi conditionnelle de Y_i sachant $X_i = x$ appartient à la famille exponentielle. Par exemple, cela permet de traiter le cas particulier important de la régression fonctionnelle binomiale, où on a $Y_i \in \{0, 1\}$ pour $i = 1, \ldots, n$ (voir Müller et Stadtmüller, 2005).
- Un autre modèle qui connaît une grande popularité récente pour des raisons pratiques (et qui est en fait un cas particulier du modèle linéaire fonctionnel généralisé cité ci-dessus) est la version fonctionnelle de la classification, développé, entre autres, dans les travaux de Berlinet, Biau et Rouvière (2005). Dans ce modèle de classification, la variable réponse Y_i est cette fois un label associé à la courbe X_i. Le but est de construire une règle de classification de façon à pouvoir attribuer un label à une nouvelle observation. Berlinet, Biau et Rouvière (2005) proposent ainsi une construction de règle de classification basée sur une décomposition de $X_1, \ldots, X_n$ dans une base d'ondelettes. Ils montrent aussi une certaine forme d'optimalité pour cette règle, prouvant qu'asymptotiquement elle prédit aussi bien que la meilleure règle possible, la règle de Bayes (voir à ce sujet Devroye, Györfi et Lugosi, 1996). Müller et Stadtmüller (2005) ont également abordé ce problème en interprétant le problème de classification (à deux labels) avec variable explicative fonctionnelle comme un cas particulier du modèle linéaire fonctionnel généralisé, avec une variable réponse binaire.
- Enfin, une dernière approche importante consiste à revenir au modèle (1) et d'estimer directement l'opérateur r de façon nonparamétrique. Cette approche a été développée par Ferraty et Vieu (2002, 2003) qui donnent un estimateur à noyau de l'opérateur r et obtiennent là aussi des vitesses de convergence pour cet estimateur. On reviendra un peu plus loin sur cette approche. Ces techniques nonparamétriques avec variables fonctionnelles sont récentes et les principaux travaux peuvent être trouvés dans la monographie de Ferraty et Vieu (2006).

Comme cela a déjà été souligné, l'étude de ces divers modèles est motivée au départ par des problèmes pratiques. La variété des domaines dans lesquels les outils de la statistique fonctionnelle interviennent est considérable. La monographie de Ramsay et Silverman (2002) est à elle seule une mine de situations concrètes de données fonctionnelles et de méthodes différentes, allant par exemple de l'étude de la forme d'os déterrés par des archéologues à l'étude de l'enregistrement de l'activité du cerveau lorsqu'on fait prononcer à un individu une syllabe (par électromyographie, c'est-à-dire en enregistrant avec des électrodes placées à la surface de la peau l'activité électrique provoquée par le mouvement des lèvres et en mesurant l'accélération des lèvres). Il est important de noter que, parmi tous ces cas concrets, se pose souvent le problème du traitement préalable des données. Ramsay et Silverman (2002) mettent notamment en évidence, sur des données relatives à des courbes de croissance, le fait que ces courbes montrent deux types de variabilité, l'amplitude (qui se rapporte à des variations en taille pour des caractéristiques particulières comme le pic de croissance de la puberté) et la phase (qui se rapporte à des variations dans le temps de caractéristiques particulières). Un prétraitement des données doit viser à éliminer la phase de façon à pouvoir concentrer l'étude sur la variation en amplitude. C'est l'objet de travaux tels que les articles de Ramsay et Li (1998), ainsi que Kneip, Li, Mac Gibbon et Ramsay (2000). Sans vouloir en faire une liste exhaustive, on souhaite citer quelques domaines dans lesquels apparaissent les données fonctionnelles, pour donner une idée du type de problèmes que la statistique fonctionnelle permet de résoudre.

- En biologie, on trouve en premier lieu le travail précurseur de Rao (1958) concernant une étude de courbes de croissance. Plus récemment, un autre exemple est l'étude des variations de l'angle du genou durant la marche (voir Ramsay et Silverman, 2002). Concernant la biologie animale, des études de la ponte de mouches méditerranéennes ont été faites par plusieurs auteurs (Chiou, Müller, Wang et Carey, 2003, Chiou, Müller et Wang, 2003, Cardot, 2006). Les données consistent en des courbes donnant pour chaque mouche la quantité d'œufs pondus en fonction du temps.
- La chimiométrie fait aussi partie des champs d'étude propices à l'utilisation de méthodes de la statistique fonctionnelle. Parmi les travaux existants sur le sujet, on peut citer Frank et Friedman (1993) dont on a déjà parlé en début d'introduction, ainsi que Hastie et Mallows (1993) qui ont commenté l'article de Frank et Friedman (1993) en apportant un exemple de courbes mesurant la log-intensité d'un rayon laser réfracté en fonction de l'angle de réfraction. Plus récemment, Ferraty et Vieu (2002) se sont intéressés à l'étude de la contenance de graisse de morceaux de viande (variable d'intérêt) étant données les courbes d'absorbsions de longueurs d'ondes infra-rouge de ces morceaux de viande (variable explicative).

- Des applications liées à l'environnement ont été notamment étudiées par Aneiros-Perez, Cardot, Estevez-Perez et Vieu (2004) qui ont travaillé sur un problème de prévision de pollution. Ces données consistent en des mesures de pics de pollution par l'ozone chaque jour (variable d'intérêt) étant données des courbes de polluants ainsi que de courbes météorologiques de la veille (variables explicatives). Ces données seront également utilisées dans la partie appliquée de cette thèse (partie IV), et seront alors explicitées en détail à ce moment-là.
- La climatologie est un domaine où les données fonctionnelles apparaissent naturellement. Une étude du phénomène El Niño (courant chaud de l'océan Pacifique) a ainsi été réalisée par Besse, Cardot et Stephenson (2000). Dans ce travail, les données consistent en des mesures de la température de ce courant en fonction du temps, et la prédiction est faite en utilisant un modèle autorégressif fonctionnel (voir à ce sujet Bosq, 2000).
- En linguistique, des travaux ont également été réalisés, notamment concernant la reconnaissance vocale. On peut citer par exemple Hastie, Buja et Tibshirani (1995), Berlinet, Biau et Rouvière (2005) ou encore Ferraty et Vieu (2003). Ces travaux sont fortement liés aux méthodes de classification lorsque la variable explicative est une courbe. Brièvement, les données sont des courbes correspondant à des enregistrements de phonèmes prononcés par différents individus. On associe un label à chaque phonème (variable d'intérêt) et le but est d'établir une classification de ces courbes en utilisant comme variable explicative la courbe enregistrée.
- Dans le domaine de la graphologie, l'apport des techniques de la statistique fonctionnelle a là aussi trouvé une application. Les travaux sur ce problème sont par exemple ceux de Hastie, Buja et Tibshirani (1995) et Ramsay (2000). Ce dernier modélise par exemple la position du stylo (abscisses et ordonnées en fonction du temps) à l'aide d'équations différentielles.
- Les applications à l'économie sont aussi relativement nombreuses. Des travaux ont notamment été effectués par Kneip et Utikal (2001), et récemment par Benko, Härdle et Kneip (2005), basés notamment sur une analyse en composantes principales fonctionnelle. Cette méthode d'estimation sera analysée lorsqu'on l'utilisera (voir partie III), même si on peut déjà souligner que l'idée de base est, lors de l'estimation de l'opérateur de covariance, d'estimer des produits scalaires entre les courbes observées au lieu d'estimer des courbes elles-mêmes.

Ce rapide tour d'horizon donne une idée de la diversité des approches permettant d'étudier les modèles (1) et (2). Dans cette thèse, on propose d'apporter des contributions à cette étude. Les différentes approches qui seront considérées, toujours reliées au modèle (2), auront ainsi pour but de proposer une estimation du paramètre fonctionnel α. Parmi les différentes méthodes possibles, l'une d'entre elles consiste à faire des hypothèses de régularité sur α (ces hypothèses seront analysées en détail dans la suite pour chaque approche envisagée). Il est alors fréquent d'estimer α par projection sur un espace de fonctions régulières dont on connaît une base. Dans cette thèse, on a considéré pour chaque approche un espace de fonctions splines, même si plusieurs autres bases (comme par exemple les bases de Fourier, les bases d'ondelettes, ...) peuvent aussi être envisagées. Depuis leur introduction (voir notamment de Boor, 1978, Schumaker, 1981, un peu plus récemment Dierckx, 1993), les splines connaissent une grande popularité, notamment grâce à une mise en œuvre pratique relativement simple. Pour situer brièvement le contexte, une spline polynômiale (univariée) sur l'intervalle $[0, 1]$ (on choisit cet intervalle pour simplifier) est une fonction polynômiale s de degré q par morceaux (avec $q \in \mathbb{N}$) définie à l'aide de $k-1$ points $x_1, \ldots, x_{k-1}$ (avec $k \in \mathbb{N}, k \geq 2$), appelés nœuds, formant une subdivision de l'intervalle $[0, 1]$,

$$0 < x_1 < \ldots < x_{k-1} < 1,$$

et s vérifiant les propriétés suivantes :

- s est un polynôme de degré q sur chaque sous-intervalle $[0, x_1]$, $[x_1, x_2]$, $\ldots$, $[x_{k-2}, x_{k-1}]$, $[x_{k-1}, 1]$,
- s est une fonction de classe $\mathcal{C}^{q-1}$ sur l'intervalle $[0, 1]$ (par convention, s est une fonction en escalier lorsque $q = 0$).

Plus précisément, étant donné un degré $q \in \mathbb{N}$, l'ensemble $\mathcal{S}_q\,(x_1, \ldots, x_{k-1})$ des fonctions splines ayant pour nœuds $x_1, \ldots, x_{k-1}$ et de degré q est l'ensemble des fonctions s qui s'écrivent

$$s(t) = \sum_{j=0}^{q} \theta_j t^j + \sum_{j=1}^{k-1} \delta_j \,(t - x_j)_+^{r-1},$$

où $\theta_0, \ldots, \theta_q, \delta_1, \ldots, \delta_{k-1} \in \mathbb{R}$ et, pour tout entier $j \geq 1$,

$$u_+^j = \begin{cases} u^j \text{ si } u \geq 0, \\ 0 \text{ si } u < 0. \end{cases}$$

On montre alors que $\mathcal{S}_q\left(x_1, \ldots, x_{k-1}\right)$ est un espace vectoriel de dimension $k+q$. En prenant comme nœuds les points de mesure des observations, on parle de splines de lissage. Ces splines particulières ont notamment été étudiées par Eubank (1988). Dans le cas où les nœuds sont en d'autres points, on parle de splines de régression. Ces deux types de fonctions splines seront utilisées dans la thèse. On va juste préciser quelques notations et propriétés pour ces deux types de fonctions splines.

- Pour les splines de régression, on se donne un entier $k \geq 2$ et un degré $q \in \mathbb{N}$ de polynôme, et on considère l'espace des fonctions splines de degré q avec $k-1$ nœuds intérieurs sur $[0,1]$. On prendra ces nœuds équidistants pour simplifier. Cet espace de fonctions splines est un espace vectoriel de dimension $k+q$. Une base de cet espace est l'ensemble des fonctions B-splines normalisées (voir par exemple de Boor, 1978), que l'on notera dans toute la suite
$$\mathbf{B}_{k,q} = \left(B_1, \ldots, B_{k+q}\right)^\tau .$$
Ainsi, une fonction spline s'écrit comme une combinaison linéaire de ces fonctions de base, de la forme $\mathbf{B}_{k,q}^\tau \boldsymbol{\theta}$ avec $\boldsymbol{\theta} \in \mathbb{R}^{k+q}$.
- Pour les splines de lissage, on prend comme nœuds les points $t_1, \ldots, t_p$ auxquels sont mesurées les courbes $X_1, \ldots, X_n$ et comme degré de polynôme $q = 2m-1$ avec m entier strictement positif. On suppose de plus que ces splines sont des polynômes de degré $m-1$ sur les intervalles $[0, t_1]$ et $[t_p, 1]$. On montre alors (voir Eubank, 1988) que cet ensemble de fonctions splines (appelé espace des fonctions splines naturelles) est un espace vectoriel de dimension p, dont une base sera notée dans toute la suite
$$\mathbf{b}(t) = \left(b_1(t), \ldots, b_p(t)\right)^\tau .$$
Une propriété importante de ces splines naturelles est qu'il existe une correspondance bijective entre cet espace et $\mathbb{R}^p$ de la façon suivante. Pour tout vecteur $\boldsymbol{w} = \left(w_1, \ldots, w_p\right)^\tau \in \mathbb{R}^p$, il existe une unique fonction spline naturelle, appelée spline d'interpolation associée à $\boldsymbol{w}$ et notée $s_{\boldsymbol{w}}$, telle que, pour tout $j = 1, \ldots, p$,
$$s_{\boldsymbol{w}}(t_j) = w_j.$$
En notant $\mathbf{B}$ la matrice de taille $p \times p$ ayant pour éléments $b_i(t_j)$ pour i et j allant de 1 à p, on montre que $s_{\boldsymbol{w}}$ est donnée par

$$s_{\boldsymbol{w}}(t) = \mathbf{b}(t)^\tau \left(\mathbf{B}^\tau \mathbf{B}\right)^{-1} \mathbf{B}^\tau \boldsymbol{w}. \tag{3}$$

De plus, une propriété importante d'une telle spline d'interpolation est qu'elle vérifie

$$\int_0^1 s_{\boldsymbol{w}}^{(m)}(t)^2 dt \leq \int_0^1 f^{(m)}(t)^2 dt, \tag{4}$$

pour toute autre fonction f avec $f^{(m)} \in L^2([0,1])$ et $f(t_j) = w_j$ pour tout $j = 1, \ldots, p$. Notons enfin une autre propriété importante des fonctions splines naturelles. Étant données des observations $(x_1, y_1), \ldots, (x_n, y_n)$, on montre (voir encore Eubank, 1998) que le problème de minimisation

$$\min_{f \in L^2([0,1])} \left\{ \frac{1}{n} \sum_{i=1}^{n} (y_i - f(x_i))^2 + \rho \int_0^1 f^{(m)}(t) dt \right\},$$

avec $\rho > 0$, admet une unique solution qui est une spline naturelle de degré $2m-1$ avec comme nœuds intérieurs $x_1, \ldots, x_n$ (spline de lissage). Pour l'estimation de la moyenne conditionnelle, ce type de problème avec pénalisation est préconisé par exemple par Eubank (1988), Wahba (1990) ou encore Green et Silverman (1994). Cela permet de chercher une fonction f dont on contrôle le lissage au moyen de la pénalisation sur la norme L^2 de sa dérivée d'ordre m.

Cette première partie d'introduction présentait le cadre de travail général de cette thèse. On va dans la suite de cette introduction donner une présentation des différentes contributions apportées à l'étude du modèle (2).

Partie I : Estimation de quantiles conditionnels

La première contribution apportée à l'étude du modèle (2) est l'objet de la partie I. On propose de considérer le modèle (2) d'un point de vue de la régression sur quantiles, pour proposer une alternative à la régression usuelle sur la moyenne. Concernant la statistique multivariée, la régression sur quantiles est apparue dans les années 70. Bien que la régression sur la moyenne dispose de propriétés qui en font un modèle très populaire (calculs explicites aisés, propriétés d'optimalité en cas d'erreurs gaussiennes), Mosteller et Tukey (1977) mettent en évidence le fait qu'elle ne peut donner qu'une vision incomplète des données. Ils lancent alors l'idée de la régression sur quantiles, reprise ensuite par Koenker et Bassett (1978). La monographie de Koenker (2005) donne actuellement une vision d'ensemble sur pratiquement 30 ans de travaux concernant la régression sur quantiles dans le cadre d'une variable explicative multivariée. Outre le fait que cette alternative à la régression sur la moyenne permette de donner une meilleure idée de la distribution des données (car calculer un quantile d'un certain ordre pour une loi de probabilité revient finalement à inverser la fonction de répartition de cette loi), elle offre également d'autres avantages, comme par exemple le fait de permettre la construction d'intervalles de prédiction, ou encore de disposer d'une certaine forme de robustesse. Concernant ce dernier point, comme souligné par Koenker (2005), la moyenne conditionnelle possède des propriétés d'optimalité lorsque les erreurs sont gaussiennes. Si ce n'est pas le cas (notamment lorsqu'on est en présence de données aberrantes), la performance par exemple de la médiane peut être supérieure à celle de la moyenne : la médiane présente une certaine forme de robustesse vis-à-vis des données aberrantes. D'un point de vue appliqué, l'utilisation des quantiles est présente dans des domaines aussi variés que l'agronomie (pour estimer des seuils de rendement), la médecine (voir par exemple l'article de Lejeune et Sarda, 1988, concernant les courbes de croissance) ou en fiabilité (toujours concernant l'estimation de seuils). Récemment, des travaux se sont intéressés à l'estimation de quantiles (non conditionnels) pour des variables aléatoires multivariées ou à valeurs dans un espace de Banach, introduisant notamment la notion de boule médiane. Il s'agit par exemple des articles de Averous et Meste (1997) puis de Cadre (2001). Au niveau de l'estimation de la médiane conditionnelle pour variable d'intérêt multivariée, Berlinet, Cadre et Gannoun (2001) ont proposé une méthode d'estimation nonparamétrique basée sur une estimation à noyau de la fonction de répartition (multivariée) conditionnelle. Cette situation (variable d'intérêt multivariée ou à valeurs dans un espace de Banach) ne sera pas envisagée ici, puisqu'on considèrera uniquement une variable d'intérêt réelle, seule la variable explicative pouvant prendre ses valeurs dans un espace de fonctions.

Dans le cadre d'une variable explicative univariée ou multivariée, outre les travaux initiaux de Koenker et Bassett (1978), l'estimation de quantiles conditionnels donne lieu à une littérature abondante. En notant toujours $Y_1, \dots, Y_n$ les observations (réelles) de la variable d'intérêt et $\mathbf{X}_1, \dots, \mathbf{X}_n$ les observations (multivariées) de la variable explicative, étant donné un réel $\alpha \in]0, 1[$, on note $g_\alpha(\boldsymbol{x})$ le quantile conditionnel sachant $\mathbf{X}_i = \boldsymbol{x}$ (avec $\boldsymbol{x} \in \mathbb{R}^p$) défini par

$$\mathbb{P}(Y_i \leq g_\alpha(\mathbf{X}_i) | \mathbf{X}_i = \boldsymbol{x}) = \alpha,$$

où $\mathbb{P}(.|\mathbf{X}_i = \boldsymbol{x})$ désigne la loi conditionnelle de Y_i sachant $\mathbf{X}_i = \boldsymbol{x}$. La fonction g_α, définie de $\mathbb{R}^p$ dans $\mathbb{R}$ est appelée fonction quantile conditionnel d'ordre α. L'article bibliographique de Poiraud-Casanova et Thomas-Agnan (1998) fait un large tour d'horizon des méthodes d'estimation de cette fonction quantile conditionnel. Ces méthodes peuvent être réparties en deux classes.

- La première de ces classes utilise le fait déjà souligné que, pour calculer un quantile d'un certain ordre pour une loi de probabilité, on peut revenir à l'inversion de la fonction de répartition (conditionnelle) de cette loi. Le tout est alors de donner une méthode d'estimation de cette fonction de répartition. Cette idée est utilisée par exemple dans un article de Bhattacharya et Gangopadhyay (1990) qui proposent une estimation à noyau et par la méthode des plus proches voisins de la fonction de répartition. Les vitesses nonparamétriques usuelles sont également obtenues pour les estimateurs construits. Avec cette même idée, Ducharme, Gannoun, Guertin et Jéquier (1995) donnent un estimateur à noyau de la fonction de répartition conditionnelle, puis en l'inversant, obtiennent la normalité asymptotique de l'estimateur des quantiles conditionnels.
- La seconde classe d'estimation de quantiles conditionnels regroupe les méthodes plus directes, basées sur la recherche de quantiles conditionnels comme solutions d'un problème de minimisation. Bassett et Koenker (1978) étudient ainsi un estimateur de la médiane conditionnelle en minimisant un critère du type moindres valeurs absolues, et prouvent sa consistance et sa normalité asymptotique. D'un point de vue numérique, des méthodes algorithmiques de résolution de tels problèmes de minimisation (dont la solution n'est pas explicite) peuvent être trouvées dans Koenker (2005), comme par exemple l'utilisation de la méthode du simplexe. Plus généralement, ceci peut être étendu à n'importe quel quantile. On cherche ainsi un estimateur de g_α parmi une certaine classe de fonctions r_α minimisant une quantité du type

$$\frac{1}{n}\sum_{i=1}^{n} l_\alpha\left(Y_i - r_\alpha\left(\mathbf{X}_i\right)\right), \tag{5}$$

où la fonction de perte l_α (qui remplace la fonction carré par rapport à un problème de minimisation classique de type moindres carrés), baptisée “check function” par Koenker (2005), est définie par

$$l_\alpha\left(u\right) = |u| + \left(2\alpha - 1\right)u.$$

> On retombe sur l'estimation de la médiane par minimisation d'un critère du type moindres valeurs absolues lorsque $\alpha = 1/2$. Plusieurs types d'estimateurs sont envisageables, ayant en commun le fait de minimiser cette quantité (5). Par exemple, He et Shi (1994) proposent un estimateur de g_α basé sur des splines de régression et obtiennent des vitesses de convergence usuelles en statistique nonparamétrique. On reviendra sur cet estimateur au cours de cette partie I de la thèse. On rencontre d'autres estimateurs dans la littérature, comme par exemple l'estimateur de Koenker, Ng et Portnoy (1994) basé sur des splines de lissage, ou encore l'estimateur proposé par Lejeune et Sarda (1988) utilisant la méthode de régression polynômiale locale. Enfin, Fan, Hu et Truong (1994) donnent quant à eux un estimateur à noyau de g_α.

La partie I de cette thèse propose donc de généraliser la notion de régression sur quantiles au cas où la variable explicative est fonctionnelle. On considère ainsi des observations $X_1, \ldots, X_n$ appartenant à l'espace de Hilbert $L^2([0,1])$ des fonctions définies de $[0,1]$ dans $\mathbb{R}$ de carré intégrable, muni de son produit scalaire usuel défini par

$$\langle f, g \rangle = \int_0^1 f(t)g(t)dt,$$

pour toutes fonctions f et g de $L^2([0,1])$, et de norme associée $\|.\|_{L^2}$. En pratique, les courbes observées sont de carré intégrable dans une très grande majorité de cas, le fait de se restreindre à des fonctions définies sur $[0,1]$ ne fait pas perdre en généralité, puisqu'on peut toujours s'y ramener (pour des fonctions définies au départ sur un intervalle $[a,b]$) à l'aide de la transformation affine

$$x \longmapsto \frac{x-a}{b-a}.$$

Dans ce contexte, on cherche alors r_α dans une certaine classe d'opérateurs qui minimise

$$\frac{1}{n}\sum_{i=1}^{n} l_\alpha \left(Y_i - r_\alpha\left(X_i\right)\right). \tag{6}$$

On se place ensuite dans le cadre d'un modèle linéaire, c'est-à-dire que l'on suppose que $r_\alpha(X_i)$ s'écrit $\langle \Psi_\alpha, X_i \rangle$ pour tout $i = 1, \ldots, n$ avec $\Psi_\alpha \in L^2([0,1])$. L'objectif de cette partie est de proposer un estimateur Ψ_α à l'aide de splines de régression, solution d'une version pénalisée du problème de minimisation (6). Cette approche s'inspire ainsi, dans le cas réel, des travaux de He et Shi (1994), ou encore de Koenker, Ng et Portnoy (1994) du point de vue de l'introduction d'une pénalisation. On verra que, dans le cadre fonctionnel, l'introduction d'une pénalisation est importante pour assurer l'existence d'un estimateur solution du problème de minimisation (on peut voir aussi les travaux de Cardot, Ferraty et Sarda, 2003, concernant l'introduction d'une telle pénalisation dans le cadre de l'estimation de la moyenne conditionnelle). Le comportement asymptotique de cet estimateur sera ensuite étudié et on donnera une borne supérieure pour la vitesse de convergence, relativement à une norme particulière, la norme induite par l'opérateur de covariance de X_i.

Partie II : Estimateur par splines de lissage dans le modèle linéaire fonctionnel

Dans la partie précédente, l'estimation de quantiles conditionnels a été vue comme une alternative possible à l'estimation de la moyenne conditionnelle. Cependant, cette dernière reste la plus populaire concernant l'étude des modèles (1) et (2). On propose dans cette partie une nouvelle méthode d'estimation de la moyenne conditionnelle. Comme cela a déjà été remarqué, les modèles (1) et (2) sont concernés par de nombreux travaux. La procédure d'estimation proposée dans cette partie, basée sur des techniques de splines de lissage, offre un complément aux techniques déjà existantes.

- Par exemple, Ferraty et Vieu (2002, 2006) considèrent le modèle (1) et se placent d'un point de vue nonparamétrique, en estimant directement l'opérateur r. Ils donnent une méthode d'estimation à noyau (en adaptant l'estimateur de Nadaraya-Watson à ce cadre fonctionnel) et fournissent des résultats de convergence pour cet estimateur. En adoptant un point de vue nonparamétrique, ils font simplement des hypothèses sur r du type "r continu" ou "r höldérien" et obtiennent leurs résultats en introduisant la dimension fractale de la loi de probabilité de $X_1, \ldots, X_n$, en d'autres termes en contrôlant les probabilités du type $\mathbb{P}(X_i \in \mathcal{B}(x, \delta))$ lorsque δ tend vers zéro, où $\mathcal{B}(x, \delta)$ désigne la boule de centre x et de rayon δ pour la semi-norme $\|.\|_H$ de H. Cette approche nonparamétrique ne sera pas envisagée dans cette thèse. On considèrera directement le modèle (2) et on proposera une méthode d'estimation de α basé sur la résolution d'un problème de minimisation de type moindres carrés pénalisés.

- Ce type d'approche d'estimation de α se retrouve par exemple dans les travaux de Goutis (1998) qui donne une méthode d'estimation de α, utilisant cependant le produit scalaire défini comme l'intégrale du produit des dérivées secondes des fonctions au lieu du produit scalaire usuel de L^2. Cardot, Ferraty et Sarda (1999, 2003) donnent quant à eux deux méthodes d'estimation de α, leur implémentation ainsi que des résultats de convergence pour les estimateurs construits. La première méthode d'estimation introduit la régression sur composantes principales fonctionnelle, basée sur la diagonalisation de l'opérateur de covariance de X_i et généralisant ainsi la méthode de régression sur composantes principales dans le cas multivarié. La seconde méthode d'estimation de α utilise les splines de régression (en choisissant $k-1$ nœuds équirépartis dans l'intervalle $[0,1]$). Cet estimateur est obtenu comme solution d'un problème de minimisation de type moindres carrés pénalisés. On recherche alors un estimateur spline $\widehat{\alpha}$ de α, s'écrivant ainsi $\widehat{\alpha} = \mathbf{B}_{k,q}^{\tau}\widehat{\boldsymbol{\theta}}$ avec $\widehat{\boldsymbol{\theta}} \in \mathbb{R}^{k+q}$ solution du problème de minimisation

$$\min_{\boldsymbol{\theta}\in\mathbb{R}^{k+q}} \left\{ \frac{1}{n}\sum_{i=1}^{n} \left(Y_i - \langle \mathbf{B}_{k,q}^{\tau}\boldsymbol{\theta}, X_i\rangle\right)^2 + \rho \left\| \left(\mathbf{B}_{k,q}^{\tau}\boldsymbol{\theta}\right)^{(m)} \right\|_{L^2}^2 \right\}. \tag{7}$$

  Ce problème de minimisation est consitué de deux termes, le premier étant un terme de résidus de type moindres carrés classique et le second étant un terme de régularisation avec un paramètre de lissage (ρ) et la norme au carré de la dérivée d'ordre m de la fonction spline que l'on recherche. Comme souligné par Cardot, Ferraty et Sarda (2003), ce terme de régularisation permet d'assurer l'existence et la consistance de l'estimateur spline construit. Il permet aussi en pratique de contrôler le lissage de cet estimateur. Il est important de noter que le problème de minimisation (7) admet une solution explicite, ce qui permet une mise en œuvre pratique relativement simple.

L'approche que l'on présente dans cette partie II peut être davantage rapprochée des travaux de Cardot, Ferraty et Sarda (2003). En ce qui concerne notre travail, on a considéré des splines de lissage (c'est-à-dire avec des nœuds placés aux points de mesure des courbes $X_1, \ldots, X_n$ observées). Pour utiliser ces splines de lissage, on introduit les points de mesure des courbes. On suppose ainsi que $X_1, \ldots, X_n$ sont à valeurs dans $L^2([0,1])$, espace des fonctions définies de $[0,1]$ dans $\mathbb{R}$, de carré intégrable, muni de son produit scalaire $\langle .,. \rangle$ et de sa norme $\|.\|_{L^2}$ usuels. Pour simplifier, on supposera que toutes les courbes sont observées en des points de discrétisation $t_1 < \ldots < t_p$, les mêmes pour toutes les courbes, et équirépartis, c'est-à-dire que $t_j - t_{j-1} = 1/p$ pour tout $j = 2, \ldots, p$. Le modèle (2) sera alors approximé par un modèle linéaire fonctionnel discret. On associe ensuite à ce modèle un problème de minimisation de type moindres carrés pénalisés, dont la solution (estimation de α) s'écrit de façon explicite comme une fonction spline. On établit dans cette partie II un résultat de convergence sur l'estimation de α. On peut même améliorer les vitesses obtenues en posant des hypothèses plus fortes sur la régularité des courbes $X_1, \ldots, X_n$. L'analyse de ces vitesses sera détaillée au cours de cette partie II.

Partie III : Modèle linéaire fonctionnel lorsque la variable explicative est bruitée

Jusqu'à présent, au vu de l'écriture des modèles (1) et (2), on a toujours implicitement fait l'hypothèse que les courbes $X_1, \ldots, X_n$ sont observées sans erreur. Cette hypothèse peut se révéler assez peu réaliste en pratique puisque de nombreuses erreurs (comme entre autres des erreurs de mesure) peuvent empêcher de connaître les courbes $X_1, \ldots, X_n$ exactement. Il semble alors plus réaliste de considérer que la variable explicative réellement disponible est une variable W_i (pour $i = 1, \ldots, n$) telle qu'en chaque point de mesure t_j (pour $j = 1, \ldots, p$), on a

$$W_i(t_j) = X_i(t_j) + \delta_{ij}, \tag{8}$$

où $(\delta_{ij})_{i=1,\ldots,n, j=1,\ldots,p}$ est une suite de variables aléatoires indépendantes et identiquement distribuées représentant les erreurs (de mesures, ...) faites en chaque point $t_1, \ldots, t_p$, et telles que $\mathbb{E}(\delta_{ij}) = 0$ et $\mathbb{E}(\delta_{ij}^2) = \sigma_\delta^2$ pour tout $i = 1, \ldots, n$ et pour tout $j = 1, \ldots, p$.

Ce modèle avec des erreurs dans les variables explicatives a été l'objet de nombreuses études dans le cadre multivarié (c'est-à-dire lorsque $\mathbf{X}_1, \ldots, \mathbf{X}_n$ sont des éléments de $\mathbb{R}^p$). Par exemple, Fuller (1987) donne pour ce modèle bruité une méthode par maximum de vraisemblance. Des résultats asymptotiques sont également donnés par Gleser (1981). Une méthode numérique importante, connue sous le nom de moindres carrés orthogonaux, a été notamment présentée par Golub et Van Loan (1980), puis reprise et développée dans un ouvrage de Van Huffel et Vandewalle (1991). L'idée de départ de cette méthode consiste à rajouter dans le problème de minimisation des moindres carrés la quantité

$$\frac{1}{n}\sum_{i=1}^{n}\|\mathbf{W}_i - \mathbf{X}_i\|^2 ,$$

où $\mathbf{W}_i$ et $\mathbf{X}_i$ (pour $i = 1, \ldots, n$) désignent les vecteurs de taille p de termes généraux respectifs $W_i(t_j)$ et $X_i(t_j)$, pour tout $j = 1, \ldots, p$, et $\|.\|$ désigne la norme vectorielle euclidienne usuelle (ici dans $\mathbb{R}^p$). On détermine alors (voir Golub et Van Loan, 1980) la solution du problème de minimisation, construisant ainsi un estimateur consistant de la moyenne conditionnelle.

Dans notre cadre fonctionnel, le cas de variables explicatives bruitées a déjà été considéré. Les méthodes envisagées sont généralement basées sur un débruitage de chaque courbe par des techniques de lissage (voir par exemple Chiou, Muller et Wang, 2003, Cardot, 2006). Dans cette partie, on propose de généraliser la méthode des moindres carrés orthogonaux à ce cadre fonctionnel, fournissant ainsi une méthode “globale” de débruitage, et non plus courbe par courbe. Après avoir expliqué le fonctionnement et le principe de résolution de cette méthode des moindres carrés orthogonaux dans le cas multivarié, l'objet de cette partie III sera de donner sa généralisation au cas d'une variable explicative fonctionnelle, et de s'intéresser au comportement asymptotique de l'estimateur construit par la méthode des moindres carrés orthogonaux. La méthode sera envisagée à la fois pour les splines de lissage (c'est-à-dire dans le même contexte que dans la partie II) et pour les splines de régression (c'est-à-dire dans le même contexte que dans les travaux de Cardot, Ferraty et Sarda, 2003).

Concernant ce problème de variable explicative bruitée, une autre piste a commencée a être envisagée au cours de cette thèse. En revenant à une méthode de débruitage courbe par courbe, elle consiste à considérer un lissage de chaque courbe bruitée (par exemple un lissage à noyau), puis de faire une régression sur composantes principales fonctionnelle à l'aide de la technique utilisée par Kneip et Utikal (2001) ou encore par Benko, Härdle et Kneip (2005). Comme cela a déjà été signalé, dans ces articles, lors de l'estimation de l'opérateur de covariance, on estime des produits scalaires entre les courbes observées plutôt que les courbes elles-mêmes. Cette méthode d'estimation a commencé à donner des résultats encourageants tant au niveau pratique sur des simulations que théorique avec la recherche de vitesses de convergence pour l'estimateur de α. Ces premiers résultats seront également présentés dans cette partie III.

Partie IV : Simulations et application à la prévision de pics de pollution

Dans cette dernière partie, on se propose de mettre en œuvre les différents types d'estimateurs étudiés sur des jeux de données. Cette partie appliquée propose d'abord une étude sur des données simulées (permettant de juger la qualités des techniques d'estimation). Concernant l'estimation spline de quantiles conditionnels présentée dans la partie I, le problème de minimisation que l'on résoud n'ayant pas de solution explicite, on utilisera un algorithme de type moindres carrés itérés pondérés (voir par exemple Ruppert et Caroll, 1988, également Lejeune et Sarda, 1988), permettant d'obtenir une méthode numérique de construction de l'estimateur. On y présentera de plus, concernant l'estimation de la moyenne conditionnelle et de quantiles conditionnels, une façon de travailler avec plusieurs variables explicatives, au moyen d'un modèle additif et l'introduction d'un algorithme de type backfitting (voir Hastie et Tibshirani, 1990), dont le principe sera donné dans cette partie IV. On étudie dans un deuxième temps des données réelles dans le but d'apporter des réponses à un problème de prévision de pics de pollution dans la région de Toulouse (France). Ce thème de recherche, très important du point de vue de la protection de l'environnement, est l'objet de nombreuses études. Par exemple, Ghattas (1999) propose une méthode de prévision de pics de pollution par l'ozone à l'aide d'arbres de régression (voir à ce sujet Breiman, Friedman, Olshen et Stone, 1984), utilisant un jeu de données mesurées dans la région de Marseille (France). Damon et Guillas (2002) ont un point de vue davantage fonctionnel (observation de courbes d'ozone) et basent leur méthode de prévision de pics de pollution sur un modèle autorégressif hilbertien (voir Bosq, 2000).

Dans cette partie, le but est d'utiliser les méthodes introduites dans chacune des parties I, II et III de cette thèse sur un jeu de données réel pour

- construire une prévision du pic de pollution du lendemain à l'aide de la moyenne, de la médiane,
- construire des intervalles de prédiction du pic de pollution du lendemain,

- tenir compte des erreurs de mesure éventuelles sur la ou les variables explicatives.

Ces données ont été fournies par l'ORAMIP (Observatoire Régional de l'Air en Midi-Pyrénées). Elles ont déjà été l'objet d'une précédente étude à l'aide d'une méthode d'estimation nonparamétrique à noyau par Aneiros-Perez, Cardot, Estevez-Perez et Vieu (2004). Il s'agit de l'estimateur nonparamétrique préalablement introduit par Ferraty et Vieu (2002). Aneiros-Perez, Cardot, Estevez-Perez et Vieu (2004) utilisent de plus le même type d'algorithme backfitting que celui présenté dans cette thèse pour étudier le cas de plusieurs variables explicatives.

Sans rentrer pour le moment dans les détails (les données seront présentées dans la partie IV), ces données consistent en des mesures horaires de polluants (comme par exemple l'ozone, qui sera le polluant auquel on s'intéresse, ou encore le monoxyde d'azote), ainsi que de variables météorologiques (par exemple la vitesse du vent). Une première partie du travail consistera en une étude descriptive de ces données (qui ont nécessité un premier traitement, en raison de données manquantes). Puis dans un deuxième temps, on présentera des méthodes de prévision de pics de pollution (par la moyenne conditionnelle, la médiane conditionnelle, par intervalles de prédiction) en utilisant les estimateurs splines considérés dans cette thèse (l'estimateur spline de quantiles conditionels présenté dans la partie I, l'estimateur de la moyenne conditionnelle par splines de lissage présenté dans la partie II, celui de la partie III tenant compte des erreurs de mesure, ainsi que l'estimateur de la moyenne conditionnelle par splines de régression de Cardot, Ferraty et Sarda, 2003).

PARTIE I

ESTIMATION SPLINE DE QUANTILES CONDITIONNELS POUR VARIABLE EXPLICATIVE FONCTIONNELLE

I.1. PRÉSENTATION DE L'ESTIMATEUR

Dans ce chapitre de présentation, on donne le principe de la construction de l'estimateur de quantiles conditionnels pour variables explicatives fonctionnelles, ainsi que des propriétés asymptotiques concernant cet estimateur. Ce travail a fait l'objet d'une note aux *Comptes Rendus de l'Académie des Sciences* (voir Cardot, Crambes et Sarda, 2004a) et d'un article paru dans *Journal of Nonparametric Statistics* (voir Cardot, Crambes et Sarda, 2005). Cet article est donné au chapitre I.2.

On reprend les notations introduites précédemment, et on considère dans cette partie que l'on se place dans un cadre de modèle à plan aléatoire. Ainsi, $X_1, \ldots, X_n$ sont des variables aléatoires indépendantes, identiquement distribuées (de même loi qu'une variable aléatoire X), à valeurs dans $L^2([0,1])$, muni de son produit scalaire usuel et de sa norme associée. Pour simplifier, mais sans perdre en généralité, on suppose que les variables $X_1, \ldots, X_n$ sont centrées, c'est-à-dire que $\mathbb{E}(X) = 0$. Les variables aléatoires $Y_1, \ldots, Y_n$ sont à valeurs dans $\mathbb{R}$, de même loi qu'une variable aléatoire Y. On supposera également qu'il y a indépendance entre les couples (X_1, Y_1), ..., (X_n, Y_n). Comme cela a été souligné dans l'introduction de cette thèse, lorsqu'on se place du point de vue de la régression sur quantiles, étant donné un nombre réel $\alpha \in]0,1[$, on cherche r_α qui minimise le critère (5). En faisant l'hypothèse d'un modèle linéaire, on écrit $r_\alpha(X_i) = \langle \Psi_\alpha, X_i \rangle$ pour tout $i = 1, \ldots, n$ avec $\Psi_\alpha \in L^2([0,1])$. En faisant sur Ψ_α des hypothèses de régularité qui seront détaillées plus loin, on souhaite proposer ici un estimateur spline de Ψ_α, en utilisant les splines de régression. De façon analogue à ce qui est fait par Cardot, Ferraty et Sarda (2003) dans le cadre de l'estimation de la moyenne conditionnelle, on estime alors Ψ_α comme une combinaison linéaire des fonctions de base des fonctions B-splines. Pour estimer le vecteur $\widehat{\boldsymbol{\theta}}$ des coefficients de cette combinaison linéaire, il est naturel de revenir au critère (5) à minimiser. Cependant, en raison de la décroissance vers zéro des valeurs propres de l'opérateur de covariance associé à X (voir à ce sujet Dauxois, Pousse et Romain, 1982) , on va considérer un problème de minimisation pénalisé. Cette approche par pénalisation est préconisée comme on l'a vu en introduction notamment par Wahba (1990) ainsi que Green et Silverman (1994) dans le cadre univarié et l'estimation de la moyenne conditionnelle par splines de lissage. Toujours dans le cadre univarié, Koenker, Hu et Portnoy (1994) ont également une approche par pénalisation pour l'estimation de quantiles conditionnels par splines de lissage, à la différence que leur pénalisation porte sur la norme L^1 de la dérivée d'ordre m de la fonction cherchée (avec $m \in \mathbb{N}$). Pour des raisons de facilité de mise en œuvre en pratique, on considèrera ici une pénalisation portant sur la norme L^2 de la dérivée d'ordre m de la fonction cherchée. Cette approche est utilisée par Cardot, Ferraty et Sarda (2003) dans le cadre de l'estimation spline de la moyenne conditionnelle dans le modèle linéaire fonctionnel. Cette pénalisation va permettre de contrôler le degré de lissage de la solution, par l'intermédiaire d'un paramètre de lissage ρ. Elle permet de faire un compromis entre l'ajustement au données et le lissage de l'estimateur. Notons que le problème de minimisation que l'on considèrera aura une solution $\widehat{\Psi}_\alpha$ que l'on ne peut pas écrire explicitement, contrairement au problème de minimisation (7) dans le cadre de l'estimation de la moyenne conditionnelle. Ceci est dû au fait que la fonction objectif l_α n'est pas dérivable en zéro. Néanmoins, une méthode algorithmique de résolution de ce problème sera présentée dans la partie IV de cette thèse.

On présente maintenant le résultat de convergence de cet estimateur spline. Dorénavant, on fait dépendre le nombre de nœuds k et le paramètre de lissage ρ de n ($k = k_n$ tend vers l'infini et $\rho = \rho_n$ tend vers zéro lorsque n tend vers l'infini). Comme dans les travaux de Cardot, Ferraty et Sarda (2003), l'existence et le comportement asymptotique de cet estimateur sont liés à l'inversion de la matrice

$$\widehat{\mathbf{C}}_{\rho_n} = \frac{1}{n}\mathbf{A}^\tau\mathbf{A} + \rho_n\mathbf{G}_{k_n},$$

où $\mathbf{A}$ est la matrice de taille $n \times (k+q)$ et de terme général $\langle X_i, B_j\rangle$ pour $i = 1, \ldots, n$ et $j = 1, \ldots, k+q$. Ainsi, pour pouvoir inverser cette matrice $\widehat{\mathbf{C}}_{\rho_n}$, il faut contrôler le comportement de ses plus petites valeurs propres. Plus précisément, la plus petite valeur propre de $\widehat{\mathbf{C}}_{\rho_n}$, notée $\lambda_{\min}(\widehat{\mathbf{C}}_{\rho_n})$, tend vers zéro lorsque n tend vers l'infini, la vitesse de convergence de l'estimateur va donc dépendre de la vitesse de convergence vers zéro de $\lambda_{\min}(\widehat{\mathbf{C}}_{\rho_n})$. On introduit donc une suite $(\eta_n)_{n\in\mathbb{N}}$ telle que l'espace Ω_n défini par

$$\Omega_n = \left\{\omega \in \Omega/\lambda_{\min}(\widehat{\mathbf{C}}_{\rho_n}) > c\eta_n\right\}, \tag{I.1}$$

soit de probabilité tendant vers 1 lorsque n tend vers l'infini (avec c constante). Cardot, Ferraty et Sarda (2003) montrent qu'une telle suite $(\eta_n)_{n\in\mathbb{N}}$ existe et qu'on a même

$$\lambda_{\min}(\widehat{\mathbf{C}}_{\rho_n}) \geq c\eta_n + o_{\mathbb{P}}\left((k_n^2 n^{1-\delta})^{-1/2}\right), \tag{I.2}$$

avec $\delta \in]0,1[$ et

$$\eta_n = \frac{\rho_n}{k_n}.$$

Les hypothèses permettant d'établir le résultat de convergence de notre estimateur $\widehat{\Psi}_\alpha$ sont classiques en statistique fonctionnelle (voir notamment Cardot, Ferraty et Sarda, 2003, ainsi que Bosq, 2000, pour des hypothèses analogues). On suppose que la variable X est presque sûrement bornée dans L^2 et que la dérivée d'ordre p' de la fonction Ψ_α est ν-höldérienne (et on pose $p = p' + \nu$), cette hypothèse de régularité sur la fonction Ψ_α permettant essentiellement d'utiliser une approximation spline de Ψ_α par un résultat dû à de Boor (1978). On suppose également que les valeurs propres de l'opérateur de covariance associé à X a ses valeurs propres strictement positives. Enfin, une dernière hypothèse technique concerne la densité conditionnelle de $\epsilon = Y - \langle \Psi_\alpha, X\rangle$ sachant $X = x$: cette densité sera supposée continue et bornée inférieurement en zéro par une constante strictement positive, uniformément par rapport à $x \in L^2([0,1])$. Cette dernière hypothèse permet notamment d'assurer l'unicité du quantile conditionnel d'ordre α.

Sous ces hypothèses, on donne alors une borne pour la vitesse de convergence de $\widehat{\Psi}_\alpha$ vis-à-vis de la semi-norme induite par l'opérateur de covariance associé à X (voir le théorème I.1 de Cardot, Crambes et Sarda, 2005, donné dans le chapitre I.2). Comme on le verra alors, cette vitesse est un

$$O_{\mathbb{P}}\left(\frac{1}{k_n^{2p}}+\frac{1}{n\eta_n}+\frac{\rho_n^2}{k_n\eta_n}+\rho_n k_n^{2(m-p)}\right).$$

Un corollaire immédiat de ce résultat est obtenu en prenant en particulier $\eta_n = \rho_n/k_n$, comme dans Cardot, Ferraty et Sarda (2003). Puis, en optimisant cette vitesse par choix de k_n et ρ_n, on obtient une vitesse en $O_{\mathbb{P}}\left(n^{-2p/(4p+1)}\right)$. On retrouve ainsi la vitesse obtenue par Cardot, Ferraty et Sarda (2003) dans le cadre de l'estimation spline de la moyenne conditionnelle lorsque la variable explicative est fonctionnelle.

I.2. QUANTILE REGRESSION WHEN THE COVARIATES ARE FUNCTIONS

I.4.1. Introduction

Because of the increasing performances of measurement apparatus and computers, many data are collected and saved on thinner and thinner time scales or spatial grids (temperature curves, spectrometric curves, satellite images, ...). So, we are led to process data comparable to curves or more generally to functions of continuous variables (time, space). These data are called *functional data* in the literature (see Ramsay and Silverman, 2002). Thus, there is a need to develop statistical procedures as well as theory for this kind of data and actually many recent works study models taking into account the functional nature of the data.

Mainly in a formal way, the oldest works in that direction intended to give a mathematical framework based on the theory of linear operators in Hilbert spaces (see Deville, 1974, Dauxois and Pousse, 1976). After that and in an other direction, practical aspects of extensions of descriptive statistical methods like for example Principal Component Analysis have been considered (see Besse and Ramsay, 1986). The monographs by Ramsay and Silverman (1997, 2002) are important contributions in this area.

As pointed out by Ramsay and Silverman (1997), "*the goals of functional data analysis are essentially the same as those of other branches of Statistics*" : one of this goal is the explanation of variations of a dependent variable Y (response) by using information from an independent functional variable X (explanatory variable). In many applications, the response is a scalar : see Frank and Friedman (1993), Ramsay and Silverman (1997), ... Traditionally, one deals, for such a problem, with estimating the regression on the mean *i.e.* the minimizer among some class of functionals r of

$$\mathbb{E}\left[(Y - r(X))^2\right].$$

As when X is a vector of real numbers, the two main approaches are linear (see Ramsay and Dalzell, 1991, for the functional linear model) or purely nonparametric (see Ferraty and Vieu, 2002, which adapt kernel estimation to the functional setting). It is also known that estimating the regression on the median or more generally on quantiles has some interest. The problem is then to estimate the minimizer among g_α of

$$\mathbb{E}\left[l_\alpha\left(Y - g_\alpha(X)\right)\right], \tag{I.3}$$

where $l_\alpha(u) = |u| + (2\alpha - 1)u$. The value $\alpha = 1/2$ corresponds to the conditional median whereas values $\alpha \in]0,1[$ correspond to conditional quantiles of order α. The advantage of estimating conditional quantiles may be found in many applications such as in agronomy (estimation of yield thresholds), in medicine or in reliability. Besides robust aspects of the median, it may also help to derive some kind of confidence prediction intervals based on quantiles.

In our work, we assume that the conditional quantile of order α can be written as

$$g_\alpha(X) = \langle \Psi_\alpha, X \rangle, \tag{I.4}$$

where $\langle ., . \rangle$ is a functional inner product and the parameter of the model Ψ_α is a function to be estimated. This is the equivalent of the linear model for regression quantiles studied by Koenker and Bassett (1978) where the inner product is the Euclidean one and the parameter is a vector of scalars. We choose to estimate the function Ψ_α by a "direct" method : writing our estimator as a linear combination of B-splines, it minimizes the empirical version of expectation (I.3) with the addition of a penalty term proportional to the square norm of a given order derivative of the spline. The penalization term allows on one side to control the regularity of the estimator and on the other side to get consistency.

Unlike for the square function, minimization of function l_α does not lead to an explicit expression of the estimator. While computation of the estimator can be resolved by using traditional algorithms (for instance based on Iteratively Weighted Least Squares), the convexity of l_α allows theoretical developments.

In section 2, we define more precisely the framework of our study and the spline estimator of the functional parameter Ψ_α. Section 3 is devoted to the asymptotic behaviour of our estimator : we study L^2 convergence and derive an upper bound for the rate of convergence. Comments on the model and on the optimality of the rate of convergence are given in section 4. Finally, the proofs are gathered in section 5.

I.4.2. Construction of the estimator

In this work, the data consist of an i.i.d. sample of pairs $(X_i, Y_i)_{i=1,\dots,n}$ drawn from a population distribution (X, Y). We consider explanatory variables X_i which are square integrable (random) functions defined on $[0,1]$, *i.e.* are elements of the space $L^2([0,1])$ so that $X_i = (X_i(t), t \in [0,1])$. The response Y_i is a scalar belonging to $\mathbb{R}$. Assume that H, the range of X, is a closed subspace of $L^2([0,1])$. For Y having a finite expectation, $\mathbb{E}(|Y|) < +\infty$, and for $\alpha \in]0,1[$, the *conditional α-quantile* functional g_α of Y given X is a functional defined on H minimizing (I.3).

Our aim is to generalize the linear model introduced by Koenker and Bassett (1978). In our setting, it consists in assuming that g_α is a linear and continuous functional defined on H and then it follows that $g_\alpha(X)$ can be written as in (I.4). Taking the usual inner product in $L^2([0,1])$, we can write

$$g_\alpha(X) = \langle \Psi_\alpha, X \rangle = \int_0^1 \Psi_\alpha(t) X(t)\, dt,$$

where Ψ_α is the functional coefficient in H to be estimated, the order α being fixed. From now on we consider, for simplicity, that the random variables X_i are centered, that is to say $\mathbb{E}(X_i(t)) = 0$, for t a. e.

When X is multivariate, Bassett and Koenker (1978) study the *least absolute error (LAE)* estimator for the conditional median, which can be extended to any quantile replacing the absolute value by the convex function l_α in the criterion to be minimized (see Koenker and Bassett, 1978). In our case where we have to estimate a function belonging to an infinite dimensional space, we are looking at an estimator in the form of an expansion in some basis of B-splines functions and then minimizing a similar criterion with however the addition of a penalty term.

Before describing in details the estimation procedure, let us note that estimation of conditional quantiles has received a special attention in the multivariate case. As said before, linear modelling has been mainly investigated by Bassett and Koenker (1978). For nonparametric models, we may distinguish two different approaches : "indirect" estimators which are based on a preliminary estimation of the conditional cumulative distribution function (cdf) and "direct" estimators which are based on the minimizing the empirical version of criterion (I.3). In the class of "indirect" estimators, Bhattacharya and Gangopadhyay (1990) study a kernel estimator of the conditional cdf, and estimation of the quantile is achieved by inverting this estimated cdf. In the class of "direct" estimators, kernel estimators based on local fit have been proposed (see Tsybakov, 1986, Lejeune and Sarda, 1988 or Fan, Hu and Truong, 1994) ; in a similar approach, He and Shi (1994) and Koenker, Ng and Portnoy (1994) propose a spline estimator. Although our setting is quite different, we adapt in our proofs below some arguments of the work by He and Shi (1994).

In nonparametric estimation, it is usual to assume that the function to be estimated is sufficiently smooth so that it can be expended in some basis : the degree of smoothness is quantified by the number of derivatives and a lipschitz condition for the derivative of greatest order (see condition (H.2) below). It is also quite usual to approximate such kind of functions by means of regression splines (see de Boor, 1978, for a guide for splines). For this, we have to select a degree q in $\mathbb{N}$ and a subdivision of $[0, 1]$ defining the position of the knots. Although it is not necessary, we take equispaced knots so that only the number of the knots has to be selected : for k in $\mathbb{N}^\star$, we consider $k-1$ knots that define a subdivision of the interval $[0, 1]$ into k sub-intervals. For asymptotic theory, the degree q is fixed but the number of sub-intervals k depends on the sample size n, $k = k_n$. It is well-known that a spline function is a piecewise polynomial : we consider here piecewise polynomials of degree q on each sub-interval, and $(q-1)$ times differentiable on $[0, 1]$. This space of spline functions is a vectorial space of dimension $k+q$. A basis of this vectorial space is the set of the so-called normalized B-spline functions, that we note by $\mathbf{B}_{k,q} = (B_1, \ldots, B_{k+q})^\tau$.

Then, we estimate Ψ_α by a linear combination of functions B_l. This leads us to find a vector $\widehat{\boldsymbol{\theta}} = (\widehat{\theta}_1, \ldots, \widehat{\theta}_{k+q})^\tau$ in $\mathbb{R}^{k+q}$ such that

$$\widehat{\Psi}_\alpha = \sum_{l=1}^{k+q} \widehat{\theta}_l B_l = \mathbf{B}_{k,q}^\tau \widehat{\boldsymbol{\theta}}. \tag{I.5}$$

It is then natural to look for $\widehat{\Psi}_\alpha$ as the minimizer of the empirical version of (I.3) among functional g_α of the form (I.4) with functions Ψ_α belonging to the space of spline functions defined above. We will however consider a penalized criterion as we will see now. In our setting, the pseudo-design matrix $\mathbf{A}$ is the matrix of dimension $n \times (k+q)$ and elements $\langle X_i, B_j\rangle$ for $i = 1, \ldots, n$ and $j = 1, \ldots, k+q$. Even if we do not have an explicit expression for a solution to the minimization problem, it is known that the solution would depend on the properties of the inverse of the matrix $\frac{1}{n}\mathbf{A}^\tau\mathbf{A}$ which is the $(k+q) \times (k+q)$ matrix with general term $\langle \Gamma_{X,n}(B_j), B_l\rangle$, where $\Gamma_{X,n}$ is the empirical version of the covariance operator Γ_X of X defined for all u in $L^2([0,1])$ by

$$\Gamma_X u = \mathbb{E}\left(\langle X, u\rangle X\right). \tag{I.6}$$

We know that Γ_X is a nuclear operator (see Dauxois, Pousse and Romain, 1982), consequently no bounded inverse exists for this operator. Moreover, as a consequence of the first monotonicity principle (see theorem 7.1 in Weinberger, 1974), the restriction of this operator to the space of spline functions has smaller eigenvalues than Γ_X. Finally, it appears to be impossible to control the speed of convergence to zero of the smallest eigenvalue of $\frac{1}{n}\mathbf{A}^\tau\mathbf{A}$ (when n tends to infinity) : in that sense, we are faced with an inversion problem that can be qualified as ill-conditioned. A way to circumvent this problem is to introduce a penalization term in the minimization criterion (see Ramsay and Silverman, 1997, or Cardot, Ferraty and Sarda, 2003, for a similar approach in the functional linear model). Thus, the main role of the penalization is to control the inversion of the matrix linked to the solution of the problem and it consists in restricting the space of solutions. The penalization introduced below will have another effect since we also want to control the smoothness of our estimator. For this reason, and following several authors (see references above), we choose a penalization which allows to control the norm of the derivative of order $m > 0$ of any linear combination of B-spline functions, so that it can be expressed matricially. Denoting by $(\mathbf{B}_{k,q}^\tau\boldsymbol{\theta})^{(m)}$ the m-th derivative of the spline function $\mathbf{B}_{k,q}^{\tau}\boldsymbol{\theta}$, we have

$$\left\|(\mathbf{B}_{k,q}^{\tau}\boldsymbol{\theta})^{(m)}\right\|_{L^2}^2 = \boldsymbol{\theta}^{\tau}\mathbf{G}_k\boldsymbol{\theta}, \quad \forall\boldsymbol{\theta} \in \mathbb{R}^{k+q},$$

where $\mathbf{G}_k$ is the $(k+q)\times(k+q)$ matrix with general term $\langle B_j^{(m)}, B_l^{(m)}\rangle$.

Then, the vector $\widehat{\boldsymbol{\theta}}$ in (I.5) is chosen as the solution of the following minimization problem

$$\min_{\boldsymbol{\theta}\in\mathbb{R}^{k+q}} \left\{ \frac{1}{n}\sum_{i=1}^{n} l_\alpha(Y_i - \langle \mathbf{B}_{k,q}^{\tau}\boldsymbol{\theta}, X_i\rangle) + \rho \left\|(\mathbf{B}_{k,q}^{\tau}\boldsymbol{\theta})^{(m)}\right\|_{L^2}^2 \right\}, \tag{I.7}$$

where ρ is the penalization parameter. In the next section, we present a convergence result of the solution of (I.7). Note that the role of the penalization also clearly appears in this result.

I.4.3. Convergence result

We present in this section the main result on the convergence of our estimator, when n goes to infinity ($k = k_n \to +\infty$, $\rho = \rho_n \to 0$). The behaviour of our estimator is linked to a penalized version of the matrix $\widehat{\mathbf{C}} = \frac{1}{n}\mathbf{A}^{\tau}\mathbf{A}$. More precisely, adopting the same notations as in Cardot, Ferraty and Sarda (2003), the existence and convergence of our estimator depend on the inverse of the matrix $\widehat{\mathbf{C}}_{\rho_n} = \widehat{\mathbf{C}} + \rho_n\mathbf{G}_{k_n}$. Under the hypotheses of theorem I.1 below, the smallest eigenvalue of $\widehat{\mathbf{C}}_{\rho_n}$, noted $\lambda_{\min}(\widehat{\mathbf{C}}_{\rho_n})$, tends to zero as the sample size n tends to infinity. As the rate of convergence of $\widehat{\Psi}_\alpha$ depends on the speed of convergence of $\lambda_{\min}(\widehat{\mathbf{C}}_{\rho_n})$ to zero, we introduce a sequence $(\eta_n)_{n\in\mathbb{N}}$ such that the set Ω_n defined by

$$\Omega_n = \left\{\omega/\lambda_{\min}(\widehat{\mathbf{C}}_{\rho_n}) > c\eta_n\right\}, \tag{I.8}$$

has probability which goes to 1 when n goes to infinity. Cardot, Ferraty and Sarda (2003) have shown that such a sequence exists in the sense that under hypotheses of theorem I.1, there exists a strictly positive sequence $(\eta_n)_{n\in\mathbb{N}}$ tending to zero as n tends to infinity and such that

$$\lambda_{\min}(\widehat{\mathbf{C}}_{\rho_n}) \geq c\eta_n + o_{\mathbb{P}}\left((k_n^2 n^{1-\delta})^{-1/2}\right), \tag{I.9}$$

with $\delta \in]0,1[$.

To prove the convergence result of the estimator $\widehat{\Psi}_\alpha$, we assume that the following hypotheses are satisfied.

$(A.1)$ $\|X\|_{L^2} \leq C_0 < +\infty, \quad a.s.$

$(A.2)$ The function Ψ_α is supposed to have a p'-th derivative $\Psi_\alpha^{(p')}$ such that

$$\left|\Psi_\alpha^{(p')}(t) - \Psi_\alpha^{(p')}(s)\right| \leq C_1|t-s|^\nu, \quad s,t \in [0;1],$$

where $C_1 > 0$ and $\nu \in [0,1]$. In what follows, we set $p = p' + \nu$ and we suppose that $q \geq p \geq m$.

$(A.3)$ The eigenvalues of Γ_X (defined in (I.6)) are strictly positive.

$(A.4)$ For $x \in H$, the random variable ϵ defined by $\epsilon = Y - \langle \Psi_\alpha, X\rangle$ has conditional density function f_x given $X = x$, continuous and bounded below by a strictly positive constant at 0, uniformly for $x \in H$.

We derive in theorem I.1 below an upper bound for the rate of convergence with respect to some kind of L^2-norm. Indeed, the operator Γ_X is strictly non-negative, so we can associate it a semi-norm noted $\|.\|_{\Gamma_X}$ and defined by $\|u\|_{\Gamma_X}^2 = \langle \Gamma_X u, u\rangle$. Then, we have the following result.

Theorem I.1. — *Under hypotheses* $(A.1) - (A.4)$, *if we also suppose that there exists* β, γ *in* $]0,1[$ *such that* $k_n \sim n^\beta$, $\rho_n \sim n^{-\gamma}$ *and* $\eta_n \sim n^{-\beta-(1-\delta)/2}$ *(where* δ *is defined in relation (I.9)), then*

(i) $\widehat{\Psi}_\alpha$ *exists and is unique except on a set whose probability goes to zero as* n *goes to infinity,*

(ii) $$\left\|\widehat{\Psi}_\alpha - \Psi_\alpha\right\|^2_{\Gamma_X} = O_{\mathbb{P}}\left(\frac{1}{k_n^{2p}} + \frac{1}{n\eta_n} + \frac{\rho_n^2}{k_n\eta_n} + \rho_n k_n^{2(m-p)}\right).$$

I.4.4. Some comments

(i) Hypotheses $(A.1)$ and $(A.3)$ are quite usual in the functional setting : see for instance Bosq (2000) or Cardot, Ferraty and Sarda (2003). Hypothesis $(A.4)$ implies uniqueness of the conditional quantile of order α.

(ii) Some arguments in the proof of theorem I.1 are inspired from the demonstration of He and Shi (1994) within the framework of real covariates. Moreover, some results from Cardot, Ferraty and Sarda (2003) are also useful, mainly to deal with the penalization term as pointed out above. Note that it is assumed in the model of He and Shi (1994) that the error term is independent of X : condition $(A.4)$ allows us to deal with a more general setting, as in Koenker and Bassett (1978).

(iii) It is possible to choose particular values for β and γ to optimize the upper bound for the rate of convergence in theorem I.1. In particular, we remark the importance to control the speed of convergence to zero of the smallest eigenvalue of $\widehat{\mathbf{C}}_{\rho_n}$ by η_n. For example, Cardot, Ferraty and Sarda (2003) have shown that, under hypotheses of theorem I.1, relation (I.9) is true with $\eta_n = \rho_n/k_n$. This gives us

$$\left\|\widehat{\Psi}_\alpha - \Psi_\alpha\right\|^2_{\Gamma_X} = O_{\mathbb{P}}\left(\frac{1}{k_n^{2p}} + \frac{k_n}{n\rho_n} + \rho_n + \rho_n k_n^{2(m-p)}\right).$$

A corollary is obtained if we take $k_n \sim n^{1/(4p+1)}$ and $\rho_n \sim n^{-2p/(4p+1)}$; then we get

$$\left\|\widehat{\Psi}_\alpha - \Psi_\alpha\right\|^2_{\Gamma_X} = O_{\mathbb{P}}\left(n^{-2p/(4p+1)}\right).$$

We can imagine that, with stronger hypotheses on the random function X, we can find a sequence η_n greater than ρ_n/k_n, that will improve the convergence speed of the estimator. As a matter of fact, the rate derived in theorem I.1 does not imply the rate obtained by Stone (1982), that is to say a rate of order $n^{-2p/(2p+1)}$. Indeed, suppose that $1/k_n^{2p}$, $1/(n\eta_n)$ and $\rho_n^2/(k_n\eta_n)$ are all of order $n^{-2p/(2p+1)}$. This would imply that $k_n \sim n^{1/(2p+1)}$ and $\eta_n \sim n^{-1/(2p+1)}$, which contradicts the condition $\eta_n \sim n^{-\beta-(1-\delta)/2}$. Nevertheless, it is possible to obtain a speed of order $n^{-2p/(2p+1)+\kappa}$. This leads to $k_n \sim n^{1/(2p+1)-\kappa/(2p)}$ and $\eta_n \sim n^{-1/(2p+1)-\kappa}$. Then, the condition $\eta_n \sim n^{-\beta-(1-\delta)/2}$ implies $\kappa = p(1-\delta)/(2p+1)$. So finally, we get $k_n \sim n^{(1+\delta)/2(2p+1)}$, $\rho_n \sim n^{(-4p-1+\delta)/4(2p+1)}$ and $\eta_n \sim n^{(-p-1+p\delta)/(2p+1)}$. The convergence result would be then

$$\left\| \widehat{\Psi}_\alpha - \Psi_\alpha \right\|_{\Gamma_X}^2 = O_{\mathbb{P}}\left(n^{-p(1+\delta)/(2p+1)} \right).$$

A final remark is that the last term $\rho_n k_n^{2(m-p)}$ of the speed in theorem I.1 is not always negligible compared to the other terms. However, it will be the case if we suppose that $m \leq p/(1+\delta) + (1-\delta)/4(1+\delta)$.

(iv) This quantile estimator is quite useful in practice, specially for forecasting purpose (by conditional median or inter-quantiles intervals). From a computational point of view, several algorithms may be used : we have implemented in the R language an algorithm based on the Iterated Reweighted Least Square (IRLS). Note that even for real data cases, the curves are always observed in some discretization points, the regression splines is easy to implement by approximating inner products with quadrature rules. The IRLS algorithm (see Ruppert and Carroll, 1988, Lejeune and Sarda, 1988) allows to build conditional quantiles spline estimators and gives satisfactory forecast results. This algorithm has been used in particular on the "ORAMIP" ("Observatoire Régional de l'Air en Midi-Pyrénées") data to forecast pollution in the city of Toulouse (France) : the results of this practical study are described in Cardot, Crambes and Sarda (2004b). We are interested in predicting the ozone concentration one day ahead, knowing the ozone curve (concentration along time) the day before. In that special case, conditional quantiles were also useful to predict an ozone threshold such that the probability to exceed this threshold is a given risk $1-\alpha$. In other words, it comes back to give an estimation of the α-quantile maximum ozone knowing the ozone curve the day before.

I.4.5. Proof of the convergence result

The proof of the result is based on the same kind of decomposition of $\widehat{\Psi}_\alpha - \Psi_\alpha$ as the one used by He and Shi (1994). The main difference comes from the fact that our design matrix is ill-conditioned, which led us to add the penalization term treated using some arguments from Cardot, Ferraty and Sarda (2003).

Hypothesis $(A.2)$ implies (see de Boor, 1978) that there exists a spline function $\Psi_\alpha^\star = \mathbf{B}_{k_n,q}^\tau \boldsymbol{\theta}^\star$, called *spline approximation* of Ψ_α, such that

$$\sup_{t\in[0,1]} |\Psi_\alpha^\star(t) - \Psi_\alpha(t)| \leq \frac{C_2}{k_n^p}. \tag{I.10}$$

In what follows, we set $R_i = \langle \Psi_\alpha^\star - \Psi_\alpha, X_i \rangle$. We deduce from (I.10) and from hypothesis $(A.1)$ that there exists a positive constant C_3 such that

$$\max_{i=1,\dots,n} |R_i| \leq \frac{C_3}{k_n^p}, \quad a.s. \tag{I.11}$$

The operator $\Gamma_{X,n}$ allows to define the empirical version of the L^2 norm by $\|u\|_{\Gamma_{X,n}}^2 = \langle \Gamma_{X,n} u, u \rangle$. At first, we show the result (ii) of theorem I.1 for the penalized empirical L^2 norm. Writing $\widehat{\Psi}_\alpha - \Psi_\alpha = (\widehat{\Psi}_\alpha - \Psi_\alpha^\star) + (\Psi_\alpha^\star - \Psi_\alpha)$, we get

$$\begin{aligned} & \left\| \widehat{\Psi}_\alpha - \Psi_\alpha \right\|_{\Gamma_{X,n}}^2 + \rho_n \left\| (\widehat{\Psi}_\alpha - \Psi_\alpha)^{(m)} \right\|_{L^2}^2 \\ \leq \; & \frac{2}{n} \sum_{i=1}^n \langle \widehat{\Psi}_\alpha - \Psi_\alpha^\star, X_i \rangle^2 + \frac{2}{n} \sum_{i=1}^n \langle \Psi_\alpha^\star - \Psi_\alpha, X_i \rangle^2 \\ & + 2\rho_n \left\| (\widehat{\Psi}_\alpha - \Psi_\alpha^\star)^{(m)} \right\|_{L^2}^2 + 2\rho_n \left\| (\Psi_\alpha^\star - \Psi_\alpha)^{(m)} \right\|_{L^2}^2 . \end{aligned}$$

Now, using again hypothesis $(A.1)$, we get almost surely and for all $i = 1, \dots, n$, the inequality $\langle \Psi_\alpha^\star - \Psi_\alpha, X_i \rangle^2 \leq C_0^2 C_2^2 / k_n^{2p}$. Moreover, lemma 8 of

Stone (1985) gives us the existence of a positive constant C_4 that satisfies $\left\|(\Psi_\alpha - \Psi_\alpha^\star)^{(m)}\right\|_{L^2}^2 \leq C_4 k_n^{2(m-p)}$. So we deduce

$$
\begin{aligned}
&\left\|\widehat{\Psi}_\alpha - \Psi_\alpha\right\|_{\Gamma_{X,n}}^2 + \rho_n \left\|(\widehat{\Psi}_\alpha - \Psi_\alpha)^{(m)}\right\|_{L^2}^2 \\
\leq\ & \frac{2}{n}\sum_{i=1}^n \langle \widehat{\Psi}_\alpha - \Psi_\alpha^\star, X_i\rangle^2 + 2\rho_n \left\|(\widehat{\Psi}_\alpha - \Psi_\alpha^\star)^{(m)}\right\|_{L^2}^2 \\
& + \frac{2C_0^2 C_2^2}{k_n^{2p}} + 2C_4 \rho_n k_n^{2(m-p)}, \quad a.s.
\end{aligned}
\tag{I.12}
$$

Our goal is now to compare our estimator $\widehat{\Psi}_\alpha$ with the spline approximation $\Psi_\alpha^\star$. For that, we adopt the following transformation $\boldsymbol{\theta} = \widehat{\mathbf{C}}_{\rho_n}^{-1/2}\boldsymbol{\beta} + \boldsymbol{\theta}^\star$. Then, we define on the set Ω_n

$$
\begin{aligned}
f_i(\boldsymbol{\beta}) \ =\ & l_\alpha\left[Y_i - \langle \mathbf{B}_{k_n,q}^\tau\left(\widehat{\mathbf{C}}_{\rho_n}^{-1/2}\boldsymbol{\beta} + \boldsymbol{\theta}^\star\right), X_i\rangle\right] \\
& + \rho_n \left\|\left[\mathbf{B}_{k_n,q}^\tau\left(\widehat{\mathbf{C}}_{\rho_n}^{-1/2}\boldsymbol{\beta} + \boldsymbol{\theta}^\star\right)\right]^{(m)}\right\|_{L^2}^2 .
\end{aligned}
$$

We notice that minimizing $\sum_{i=1}^n f_i(\boldsymbol{\beta})$ comes back to the minimization of the criterion (I.7). We are interested by the behaviour of the function f_i around zero : $f_i(\mathbf{0})$ is the value of our loss criterion when $\boldsymbol{\theta} = \boldsymbol{\theta}^\star$. Let us also notice that the inverse of the matrix $\widehat{\mathbf{C}}_{\rho_n}$ appears in the definition of f_i. This inverse exists on the set Ω_n defined by (I.8), and which probability goes to 1 as n goes to infinity. Lemma I.1 below allows us to get the results (i) and (ii) of theorem I.1 for the penalized empirical L^2 norm.

Lemma I.1. — *Under the hypotheses of theorem I.1, for all $\epsilon > 0$, there exists L (sufficiently large) and $(\delta_n)_{n\in\mathbb{N}}$ with $\delta_n = \sqrt{1/(n\eta_n) + \rho_n^2/(k_n\eta_n)}$ such that, for n large enough*

$$
\mathbb{P}\left[\inf_{\|\boldsymbol{\beta}\|=L\delta_n}\sum_{i=1}^n f_i(\boldsymbol{\beta}) > \sum_{i=1}^n f_i(\mathbf{0})\right] > 1-\epsilon.
$$

We use convexity arguments to prove the result (i). The existence of the solution of the minimization problem (I.7) is guaranteed since the function to be minimized is convex, if we keep in mind that

$$\rho \boldsymbol{\theta}^{\tau} \mathbf{G}_k \boldsymbol{\theta} = \rho \left\| \left(\mathbf{B}_{k,q}^{\tau} \boldsymbol{\theta} \right)^{(m)} \right\|_{L^2}^2 \geq 0.$$

Using the convexity of f_i, the result of lemma I.1 means that for all $\epsilon > 0$ there exists L such that, for n large enough (as $L\delta_n$ goes to zero), we can not find more than one minimum for the function $\sum_{i=1}^{n} f_i$ with probability $1 - \epsilon$.

As we use the one-to-one transformation $\boldsymbol{\theta} = \widehat{\mathbf{C}}_{\rho_n}^{-1/2} \boldsymbol{\beta} + \boldsymbol{\theta}^{\star}$ on the set Ω_n, we deduce the existence and the uniqueness of the solution of (I.7) on a subset of Ω_n whose probability goes to one as n goes to infinity, which proves point (i) of theorem I.1.

Now, let ϵ be strictly positive ; using lemma I.1 and the convexity of function f_i, there exists L such that, for n large enough

$$\mathbb{P} \left[\inf_{\|\boldsymbol{\beta}\| \geq L\delta_n} \sum_{i=1}^{n} f_i(\boldsymbol{\beta}) > \sum_{i=1}^{n} f_i(\mathbf{0}) \right] > 1 - \epsilon. \tag{I.13}$$

On the other hand, using the definition of f_i and the minimization criterion (I.7), we have

$$\begin{aligned} & \frac{1}{n} \sum_{i=1}^{n} f_i \left(\widehat{\mathbf{C}}_{\rho_n}^{1/2} \widehat{\boldsymbol{\theta}} - \widehat{\mathbf{C}}_{\rho_n}^{1/2} \boldsymbol{\theta}^{\star} \right) \\ = & \inf_{\boldsymbol{\theta} \in \mathbb{R}^{k_n + q}} \left[\frac{1}{n} \sum_{i=1}^{n} l_{\alpha} \left(Y_i - \langle \mathbf{B}_{k_n,q}^{\tau} \boldsymbol{\theta}, X_i \rangle \right) + \rho_n \left\| \left(\mathbf{B}_{k_n,q}^{\tau} \boldsymbol{\theta} \right)^{(m)} \right\|_{L^2}^2 \right], \end{aligned}$$

so we finally get

$$\frac{1}{n}\sum_{i=1}^{n} f_i\left(\widehat{\mathbf{C}}_{\rho_n}^{1/2}\widehat{\boldsymbol{\theta}} - \widehat{\mathbf{C}}_{\rho_n}^{1/2}\boldsymbol{\theta}^{\star}\right) \leq \frac{1}{n}\sum_{i=1}^{n} f_i(\mathbf{0}).$$

Then, combining this with equation (I.13), we obtain

$$\text{(I.14)} \qquad \mathbb{P}\left[\inf_{\|\boldsymbol{\beta}\|\geq L\delta_n}\sum_{i=1}^{n} f_i(\boldsymbol{\beta}) > \sum_{i=1}^{n} f_i\left(\widehat{\mathbf{C}}_{\rho_n}^{1/2}\widehat{\boldsymbol{\theta}} - \widehat{\mathbf{C}}_{\rho_n}^{1/2}\boldsymbol{\theta}^{\star}\right)\right] > 1-\epsilon.$$

Now, using the definition of $\widehat{\mathbf{C}}_{\rho_n}$, we have

$$\begin{aligned}
& \mathbb{P}\left[\frac{1}{n}\sum_{i=1}^{n}\langle\widehat{\Psi}_\alpha - \Psi^{\star}_\alpha, X_i\rangle^2 + \rho_n\left\|(\widehat{\Psi}_\alpha - \Psi^{\star}_\alpha)^{(m)}\right\|_{L^2}^2 \leq L^2\delta_n^2\right] \\
= \; & 1-\mathbb{P}\left[\left\|\widehat{\mathbf{C}}_{\rho_n}^{1/2}(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}^{\star})\right\| > L\delta_n\right] \\
\geq \; & \mathbb{P}\left[\inf_{\|\boldsymbol{\beta}\|\geq L\delta_n}\sum_{i=1}^{n} f_i(\boldsymbol{\beta}) > \sum_{i=1}^{n} f_i\left(\widehat{\mathbf{C}}_{\rho_n}^{1/2}\widehat{\boldsymbol{\theta}} - \widehat{\mathbf{C}}_{\rho_n}^{1/2}\boldsymbol{\theta}^{\star}\right)\right].
\end{aligned}$$

With relation (I.14), this last probability is greater than $1-\epsilon$, so we obtain

$$\frac{1}{n}\sum_{i=1}^{n}\langle\widehat{\Psi}_\alpha - \Psi^{\star}_\alpha, X_i\rangle^2 + \rho_n\left\|(\widehat{\Psi}_\alpha - \Psi^{\star}_\alpha)^{(m)}\right\|_{L^2}^2 = O_{\mathbb{P}}\left(\delta_n^2\right) = O_{\mathbb{P}}\left(\frac{1}{n\eta_n} + \frac{\rho_n^2}{k_n\eta_n}\right).$$

This last result, combined with inequality (I.13) finally gives us the equivalent of result (ii) for the penalized empirical L^2 norm. Point (ii) (with the norm $\|.\|_{\Gamma_X}$) then follows from lemma I.2 below, and achieves the proof of theorem I.1 (ii).

Lemma I.2. — *Let f and g be two functions supposed to be m times differentiable and such that*

$$\|f-g\|_{\Gamma_{X,n}}^2 + \rho_n\|\left(f-g\right)^{(m)}\|_{L^2}^2 = O_{\mathbb{P}}(u_n),$$

with u_n going to zero when n goes to infinity. Under hypotheses $(A.1)$ and $(A.3)$ and if moreover $\|g\|_{L^2}$ and $\|g^{(m)}\|_{L^2}$ are supposed to be bounded, we have

$$\|f - g\|_{\Gamma_X}^2 = O_{\mathbb{P}}(u_n).$$

Proof of lemma I.1. — This proof is based on three preliminary lemmas. We denote by T_n the set of the random variables $(X_1, \ldots, X_n)$. Under hypotheses of theorem I.1, we have the following results.

Lemma I.3. — *There exists a constant C_5 such that, on the set Ω_n defined by (I.8), we have*

$$\max_{i=1,\ldots,n} \left| \langle \mathbf{B}_{k_n,q}^{\tau} \widehat{\mathbf{C}}_{\rho_n}^{-1/2} \boldsymbol{\beta}, X_i \rangle \right| \leq \frac{C_5 \, \|\boldsymbol{\beta}\|}{\sqrt{k_n \eta_n}}, \quad a.s.$$

Lemma I.4. — *For all $\epsilon > 0$ and for any sequence (L_n) such that $L_n \leq \sqrt{n k_n \eta_n \delta_n^2}$, we have*

$$\lim_{n \to +\infty} \mathbb{P} \left[\sup_{\|\boldsymbol{\beta}\|=1} \left| \sum_{i=1}^{n} \left(f_i(L_n \delta_n \boldsymbol{\beta}) - f_i(\mathbf{0}) - \mathbb{E}\left[f_i(L_n \delta_n \boldsymbol{\beta}) - f_i(\mathbf{0}) | T_n \right] \right) \right| > \epsilon \delta_n^2 n \right] = 0.$$

Lemma I.5. — *For all $\epsilon > 0$, there exists $L = L_\epsilon$ (sufficiently large) such that*

$$\mathbb{P} \left[\inf_{\|\boldsymbol{\beta}\|=1} \sum_{i=1}^{n} \mathbb{E}\left[f_i(L \delta_n \boldsymbol{\beta}) - f_i(\mathbf{0}) | T_n \right] > \delta_n^2 n \right] > 1 - \epsilon.$$

These three lemmas allow us to prove lemma I.1. Indeed, let L be a strictly positive real number ; we denote

$$A_n = \sum_{i=1}^{n} \left(f_i(L \delta_n \boldsymbol{\beta}) - f_i(\mathbf{0}) \right),$$

and

$$B_n = \sum_{i=1}^{n} \mathbb{E}\left[f_i(L\delta_n\boldsymbol{\beta}) - f_i(\mathbf{0})|T_n\right].$$

Using lemmas I.4 and I.5, given $\epsilon > 0$, we can find $L = L_\epsilon$ such that, for n large enough, $\mathbb{P}\left(\inf_{\|\boldsymbol{\beta}\|=1} B_n > \delta_n^2 n\right) > 1-\epsilon$ and $\sup_{\|\boldsymbol{\beta}\|=1} |A_n - B_n| = o_{\mathbb{P}}(\delta_n^2 n)$. Then, we deduce

$$\mathbb{P}\left[\inf_{\|\boldsymbol{\beta}\|=1} \sum_{i=1}^{n} f_i(L\delta_n\boldsymbol{\beta}) - \sum_{i=1}^{n} f_i(\mathbf{0}) > 0\right] > 1-\epsilon,$$

which achieves the proof of lemma I.1.

Proof of lemma I.3. — Using lemma 6.2 of Cardot, Ferraty and Sarda (2003), we have

$$\lambda_{\min}(\widehat{\mathbf{C}}_{\rho_n}) \geq C_5'\eta_n + o_{\mathbb{P}}((k_n^2 n^{1-\delta})^{-1/2}).$$

Noticing that $\left|\langle \mathbf{B}_{k_n,q}^\tau \widehat{\mathbf{C}}_{\rho_n}^{-1/2}\boldsymbol{\beta}, X_i\rangle\right|^2 \leq \langle \mathbf{B}_{k_n,q}^\tau, X_i\rangle \widehat{\mathbf{C}}_{\rho_n}^{-1} \langle \mathbf{B}_{k_n,q}, X_i\rangle \, \|\boldsymbol{\beta}\|^2$, we deduce that

$$\begin{aligned}
&\left|\langle \mathbf{B}_{k_n,q}^\tau \widehat{\mathbf{C}}_{\rho_n}^{-1/2}\boldsymbol{\beta}, X_i\rangle\right|^2 \\
\leq\ & \langle \mathbf{B}_{k_n,q}^\tau, X_i\rangle \langle \mathbf{B}_{k_n,q}, X_i\rangle \, \|\boldsymbol{\beta}\|^2 \left[\frac{1}{C_5'\eta_n} + o_{\mathbb{P}}\left(\left(k_n^2 n^{1-\delta}\right)^{-1/2}\right)\right],
\end{aligned}$$

using the fact that $\left\|\widehat{\mathbf{C}}_{\rho_n}^{-1}\right\| = \lambda_{\max}(\widehat{\mathbf{C}}_{\rho_n}^{-1}) = 1/\lambda_{\min}(\widehat{\mathbf{C}}_{\rho_n})$. Then, noticing that

$$\langle \mathbf{B}_{k_n,q}^\tau, X_i\rangle \langle \mathbf{B}_{k_n,q}, X_i\rangle = \sum_{j=1}^{k_n+q} \langle B_j, X_i\rangle^2 = O\left(\frac{1}{k_n}\right),$$

this gives us $\left|\langle \mathbf{B}_{k_n,q}^{\tau} \widehat{\mathbf{C}}_{\rho_n}^{-1/2} \boldsymbol{\beta}, X_i\rangle\right|^2 \leq C_5'' \left\|\boldsymbol{\beta}\right\|^2 / (k_n\eta_n) + o_{\mathbb{P}}\left(n^{(\delta-1)/2}\right)$ almost surely, and achieves the proof of lemma I.3.

Proof of lemma I.4. — Considering the definition of functions f_i and l_α, we have

$$\begin{aligned}
&\sup_{\|\boldsymbol{\beta}\|\leq 1} \left|\sum_{i=1}^{n} \Big(f_i(L\delta_n\boldsymbol{\beta}) - f_i(\mathbf{0}) - \mathbb{E}\left[f_i(L\delta_n\boldsymbol{\beta}) - f_i(\mathbf{0})|T_n\right]\Big)\right| \\
= &\sup_{\|\boldsymbol{\beta}\|\leq 1} \left|\sum_{i=1}^{n} \Big(\Big|\epsilon_i - L\delta_n\langle\, \mathbf{B}_{k_n,q}^{\tau} \widehat{\mathbf{C}}_{\rho_n}^{-1/2} \boldsymbol{\beta}, X_i\rangle - R_i\Big| - |\epsilon_i - R_i|\right. \\
&\left. -\mathbb{E}\Big[\Big|\epsilon_i - L\delta_n\langle \mathbf{B}_{k_n,q}^{\tau} \widehat{\mathbf{C}}_{\rho_n}^{-1/2} \boldsymbol{\beta}, X_i\rangle - R_i\Big| - |\epsilon_i - R_i||T_n\Big]\Big)\right|,
\end{aligned}$$

where $\epsilon_1, \ldots, \epsilon_n$ are n real random variables independent and identically distributed defined by $\epsilon_i = Y_i - \langle \Psi_\alpha, X_i\rangle$ for all $i = 1, \ldots, n$. Let us also denote $\Delta_i(\boldsymbol{\beta}) = \left|\epsilon_i - L\delta_n\langle \mathbf{B}_{k_n,q}^{\tau} \widehat{\mathbf{C}}_{\rho_n}^{-1/2} \boldsymbol{\beta}, X_i\rangle - R_i\right| - |\epsilon_i - R_i|$. To prove lemma I.4, it suffices to show that, for all $\epsilon > 0$, there exists $L = L_n$ such that

$$\lim_{n\to+\infty} \mathbb{P}\left(\sup_{\|\boldsymbol{\beta}\|\leq 1}\left|\sum_{i=1}^{n}\left[\Delta_i(\boldsymbol{\beta}) - \mathbb{E}(\Delta_i(\boldsymbol{\beta})|T_n)\right]\right| > \epsilon\delta_n^2 n\right) = 0.$$

Let ϵ be a real number strictly positive and $\mathcal{C}$ the subset of $\mathbb{R}^{k_n+q}$ defined by $\mathcal{C} = \left\{\boldsymbol{\beta} \in \mathbb{R}^{k_n+q} / \left\|\boldsymbol{\beta}\right\| \leq 1\right\}$. As $\mathcal{C}$ is a compact set, we can cover it with open balls, that is to say $\mathcal{C} = \bigcup_{j=1}^{K_n} \mathcal{C}_j$ with K_n chosen, for all j from 1 to K_n, such that

$$\text{(I.15)} \qquad \text{diam}\,(\mathcal{C}_j) \leq \frac{\epsilon\delta_n\sqrt{k_n\eta_n}}{8C_5 L}.$$

Hence

$$K_n \leq \left(\frac{8C_5 L}{\epsilon \delta_n \sqrt{k_n \eta_n}} \right)^{k_n+q}. \tag{I.16}$$

Now, for $1 \leq j \leq K_n$, let $\boldsymbol{\beta}_j$ be in $\mathcal{C}_j$; using the definition of $\Delta_i(\boldsymbol{\beta})$ and the triangular inequality, we have

$$\begin{aligned}
& \min_{j=1,\ldots,K_n} \left| \sum_{i=1}^n \left[\Delta_i(\boldsymbol{\beta}) - \mathbb{E}(\Delta_i(\boldsymbol{\beta})|T_n)\right] - \left[\Delta_i(\boldsymbol{\beta}_j) - \mathbb{E}(\Delta_i(\boldsymbol{\beta}_j)|T_n)\right] \right| \\
\leq \;& 2L\delta_n \min_{j=1,\ldots,K_n} \sum_{i=1}^n \left| \langle \mathbf{B}_{k_n,q}^\tau \widehat{\mathbf{C}}_{\rho_n}^{-1/2}(\boldsymbol{\beta} - \boldsymbol{\beta}_j), X_i \rangle \right|.
\end{aligned}$$

Then, using lemma I.3, we get

$$\begin{aligned}
& \min_{j=1,\ldots,K_n} \left| \sum_{i=1}^n \left[\Delta_i(\boldsymbol{\beta}) - \mathbb{E}(\Delta_i(\boldsymbol{\beta})|T_n)\right] - \left[\Delta_i(\boldsymbol{\beta}_j) - \mathbb{E}(\Delta_i(\boldsymbol{\beta}_j)|T_n)\right] \right| \\
\leq \;& 2L\delta_n \frac{C_5 n}{\sqrt{k_n \eta_n}} \min_{j=1,\ldots,K_n} \left\| \boldsymbol{\beta} - \boldsymbol{\beta}_j \right\|,
\end{aligned}$$

this last inequality being true only on the set Ω_n defined by (I.8). Moreover, there exists a unique $j_0 \in \{1, \ldots, K_n\}$ such that $\boldsymbol{\beta} \in \mathcal{C}_{j_0}$, which gives us with relation (I.15)

$$\min_{j=1,\ldots,K_n} \left| \sum_{i=1}^n \left[\Delta_i(\boldsymbol{\beta}) - \mathbb{E}(\Delta_i(\boldsymbol{\beta})|T_n)\right] - \left[\Delta_i(\boldsymbol{\beta}_j) - \mathbb{E}(\Delta_i(\boldsymbol{\beta}_j)|T_n)\right] \right| \leq \frac{\epsilon}{4}\delta_n^2 n. \tag{I.17}$$

On the other hand, we have

$$\sup_{\boldsymbol{\beta} \in \mathcal{C}} |\Delta_i(\boldsymbol{\beta})| \leq L\delta_n \sup_{\boldsymbol{\beta} \in \mathcal{C}} |\langle \mathbf{B}_{k_n,q}^\tau \widehat{\mathbf{C}}_{\rho_n}^{-1/2} \boldsymbol{\beta}, X_i \rangle|,$$

and using lemma I.3 again, we get, on Ω_n,

$$\sup_{\boldsymbol{\beta}\in\mathcal{C}} |\Delta_i(\boldsymbol{\beta})| \leq \frac{C_5 L\delta_n}{\sqrt{k_n \eta_n}}. \tag{I.18}$$

Besides, for $\boldsymbol{\beta}$ fixed in $\mathcal{C}$, with the same arguments as before, if we denote by $T^\star$ the set of the random variables $(X_1, \ldots, X_n, \ldots)$, we have

$$\sum_{i=1}^n \mathbb{V}\left(\Delta_i(\boldsymbol{\beta})|T^\star\right) \leq \sum_{i=1}^n L^2\delta_n^2 \mathbb{V}\left(|\langle \mathbf{B}_{k_n,q}^\tau \widehat{\mathbf{C}}_{\rho_n}^{-1/2}\boldsymbol{\beta}, X_i\rangle|^2 |T^\star\right).$$

Then, using the definition of $\widehat{\mathbf{C}}_{\rho_n}$, we remark that

$$\sum_{i=1}^n \left|\langle \mathbf{B}_{k_n,q}^\tau \widehat{\mathbf{C}}_{\rho_n}^{-1/2}\boldsymbol{\beta}, X_i\rangle\right|^2 = n\left\|\boldsymbol{\beta}\right\|^2 - n\rho_n \boldsymbol{\beta}^\tau \widehat{\mathbf{C}}_{\rho_n}^{-1/2}\mathbf{G}_{k_n}\widehat{\mathbf{C}}_{\rho_n}^{-1/2}\boldsymbol{\beta}, \tag{I.19}$$

which gives us

$$\sum_{i=1}^n \mathbb{V}\left(\Delta_i(\boldsymbol{\beta})|T^\star\right) \leq nL^2\delta_n^2. \tag{I.20}$$

We are now able to prove lemma I.4. Using first relation (I.17), we have

$$\begin{aligned}
&\mathbb{P}\left[\left(\sup_{\|\boldsymbol{\beta}\|\leq 1}\left|\sum_{i=1}^n \left[\Delta_i(\boldsymbol{\beta}) - \mathbb{E}\left(\Delta_i(\boldsymbol{\beta})|T_n\right)\right]\right| > \epsilon\delta_n^2 n\right)\cap\Omega_n \middle| T^\star\right]\\
\leq\ &\mathbb{P}\left[\left(\max_{j=1,\ldots,K_n}\left|\sum_{i=1}^n \left[\Delta_i(\boldsymbol{\beta}_j) - \mathbb{E}\left(\Delta_i(\boldsymbol{\beta}_j)|T_n\right)\right]\right| > \frac{\epsilon}{2}\delta_n^2 n\right)\cap\Omega_n \middle| T^\star\right],
\end{aligned}$$

and then

$$\begin{aligned}
&\mathbb{P}\left[\left(\sup_{\|\boldsymbol{\beta}\|\leq 1}\left|\sum_{i=1}^{n}\left[\Delta_i(\boldsymbol{\beta})-\mathbb{E}\left(\Delta_i(\boldsymbol{\beta})|T_n\right)\right]\right|>\epsilon\delta_n^2 n\right)\cap\Omega_n\middle|T^\star\right]\\
\leq\;& K_n\max_{j=1,\ldots,K_n}\mathbb{P}\left[\left(\left|\sum_{i=1}^{n}\left[\Delta_i(\boldsymbol{\beta}_j)-\mathbb{E}\left(\Delta_i(\boldsymbol{\beta}_j)|T_n\right)\right]\right|>\frac{\epsilon}{2}\delta_n^2 n\right)\cap\Omega_n\middle|T^\star\right].
\end{aligned}$$

By inequalities (I.18) and (I.20), we apply Bernstein inequality (see Uspensky, 1937) and inequality (I.16) to obtain

$$\begin{aligned}
&\mathbb{P}\left[\left(\sup_{\|\boldsymbol{\beta}\|\leq 1}\left|\sum_{i=1}^{n}\left[\Delta_i(\boldsymbol{\beta})-\mathbb{E}\left(\Delta_i(\boldsymbol{\beta})|T_n\right)\right]\right|>\epsilon\delta_n^2 n\right)\cap\Omega_n\middle|T^\star\right]\\
\leq\;& 2\exp\left\{\ln\left(\frac{8C_5Ln}{\epsilon\delta_n\sqrt{k_n\eta_n}}\right)^{k_n+q}-\frac{\epsilon^2\delta_n^4n^2/4}{2nL^2\delta_n^2+2C_5L\delta_n\times\epsilon\delta_n^2n/(2\sqrt{k_n\eta_n})}\right\}.
\end{aligned}$$

This bound does not depend on the sample $T^\star=(X_1,\ldots,X_n,\ldots)$, hence, if we take the expectation on both sides of this inequality above, we deduce

$$\begin{aligned}
&\mathbb{P}\left[\left(\sup_{\|\boldsymbol{\beta}\|\leq 1}\sum_{i=1}^{n}\left[\Delta_i(\boldsymbol{\beta})-\mathbb{E}\left(\Delta_i(\boldsymbol{\beta})|T_n\right)\right]>\epsilon\delta_n^2 n\right)\cap\Omega_n\right]\\
\leq\;& 2\exp\left\{-\frac{\epsilon^2\delta_n^2\sqrt{k_n\eta_n}n}{8L^2\sqrt{k_n\eta_n}+4C_5L\delta_n}\right.\\
&\left.\times\left[1-\frac{(k_n+q)(8L^2\sqrt{k_n\eta_n}+4C_5L\delta_n)}{\epsilon^2\delta_n^2\sqrt{k_n\eta_n}n}\ln\left(\frac{8C_5Ln}{\epsilon\delta_n\sqrt{k_n\eta_n}}\right)\right]\right\}.
\end{aligned}$$

If $L=L_n\leq\sqrt{nk_n\eta_n\delta_n^2}$, we have

$$\begin{aligned}
&\frac{\delta_n^2\sqrt{k_n\eta_n}n}{L^2\sqrt{k_n\eta_n}}\geq\frac{1}{k_n\eta_n}\xrightarrow[n\to+\infty]{}+\infty,\\
&\frac{\delta_n^2\sqrt{k_n\eta_n}n}{L\delta_n}\geq\sqrt{n}\xrightarrow[n\to+\infty]{}+\infty,\\
&\frac{k_nL^2\sqrt{k_n\eta_n}}{\delta_n^2\sqrt{k_n\eta_n}n}\leq k_n^2\eta_n\xrightarrow[n\to+\infty]{}0,
\end{aligned}$$

$$\frac{k_n L \delta_n}{\delta_n^2 \sqrt{k_n \eta_n} n} \leq \frac{k_n}{\sqrt{n}} \xrightarrow[n \to +\infty]{} 0.$$

This leads to

$$\lim_{n \to +\infty} \mathbb{P}\left[\left(\sup_{\|\boldsymbol{\beta}\| \leq 1} \left| \sum_{i=1}^{n} \left[\Delta_i(\boldsymbol{\beta}) - \mathbb{E}\left(\Delta_i(\boldsymbol{\beta}) | T_n\right)\right] \right| > \epsilon \delta_n^2 n \right) \cap \Omega_n \right] = 0,$$

and with the fact that Ω_n has probability tending to 1 when n goes to infinity, we finally obtain

$$\lim_{n \to +\infty} \mathbb{P}\left[\sup_{\|\boldsymbol{\beta}\| \leq 1} \left| \sum_{i=1}^{n} \left[\Delta_i(\boldsymbol{\beta}) - \mathbb{E}\left(\Delta_i(\boldsymbol{\beta}) | T_n\right)\right] \right| > \epsilon \delta_n^2 n \right] = 0,$$

which achieves the proof of lemma I.4.

Proof of lemma I.5. — Let a and b be two real numbers. We denote by $F_{i\epsilon}$ the random repartition function of ϵ_i given T_n and by $f_{i\epsilon}$ the random density function of ϵ_i given T_n. As $\mathbb{E}\left(l_\alpha(\epsilon_i + b) | T_n\right) = \int_{\mathbb{R}} l_\alpha(s+b)\, dF_{i\epsilon}(s)$, we obtain, using the definition of l_α,

$$\begin{aligned}
& \mathbb{E}\left(l_\alpha(\epsilon_i + a + b) - l_\alpha(\epsilon_i + b) | T_n\right) \\
= \; & 2\alpha \int_{-a-b}^{+\infty} s \, dF_{i\epsilon}(s) - 2(1-\alpha) \int_{-\infty}^{-a-b} s \, dF_{i\epsilon}(s) \\
& -2\alpha \int_{-b}^{+\infty} s \, dF_{i\epsilon}(s) + 2(1-\alpha) \int_{-\infty}^{-b} s \, dF_{i\epsilon}(s) \\
& +2\alpha(a+b) \int_{-a-b}^{+\infty} dF_{i\epsilon}(s) - 2(1-\alpha)(a+b) \int_{-\infty}^{-a-b} dF_{i\epsilon}(s) \\
& -2\alpha b \int_{-b}^{+\infty} dF_{i\epsilon}(s) + 2(1-\alpha) b \int_{-\infty}^{-b} dF_{i\epsilon}(s),
\end{aligned}$$

what gives us

$$
\begin{aligned}
& \mathbb{E}\left(l_\alpha(\epsilon_i + a + b) - l_\alpha(\epsilon_i + b)|T_n\right) \\
= \quad & 2\int_{-a-b}^{-b} s\; dF_{i\epsilon}(s) + 2\alpha a + 2b\int_{-a-b}^{-b} dF_{i\epsilon}(s) - 2a\int_{-\infty}^{-a-b} dF_{i\epsilon}(s).
\end{aligned}
$$

Then, noticing that $dF_{i\epsilon}(s) = f_{i\epsilon}(s)ds$ and using a Taylor linearization at first order of around 0 (we write $f_{i\epsilon}(s) = f_{i\epsilon}(0) + o(1)$ and $F_{i\epsilon}(-a-b) = F_{i\epsilon}(0) - (a+b)f_{i\epsilon}(0) + o(a+b)$), we finally obtain (with $F_{i\epsilon}(0) = \alpha$)

$$
\mathbb{E}\left(l_\alpha(\epsilon_i + a + b) - l_\alpha(\epsilon_i + b)|T_n\right) = f_{i\epsilon}(0)a^2 + 2f_{i\epsilon}(0)ab + (\frac{a^2}{2} + ab)r_{iab},
$$

with $r_{iab} \longrightarrow 0$ when $a, b \longrightarrow 0$. If we set $L' = \sqrt{2}L$ and $R'_i = \sqrt{2}R_i$, this relation gives us

$$
\begin{aligned}
\text{(I.21)} \quad & \sum_{i=1}^{n} \mathbb{E}\left[l_\alpha\left(\epsilon_i - L\delta_n\langle \mathbf{B}_{k_n,q}^{\tau}\widehat{\mathbf{C}}_{\rho_n}^{-1/2}\boldsymbol{\beta}, X_i\rangle - R_i\right) - l_\alpha\left(\epsilon_i - R_i\right)|T_n\right] \\
= \quad & 2\sum_{i=1}^{n} f_{i\epsilon}(0)\left[L'^2\delta_n^2\langle \mathbf{B}_{k_n,q}^{\tau}\widehat{\mathbf{C}}_{\rho_n}^{-1/2}\boldsymbol{\beta}, X_i\rangle^2 + L'\delta_n\langle \mathbf{B}_{k_n,q}^{\tau}\widehat{\mathbf{C}}_{\rho_n}^{-1/2}\boldsymbol{\beta}, X_i\rangle R'_i\right] \\
& + \sum_{i=1}^{n}\left[L'^2\delta_n^2\langle \mathbf{B}_{k_n,q}^{\tau}\widehat{\mathbf{C}}_{\rho_n}^{-1/2}\boldsymbol{\beta}, X_i\rangle^2 + L'\delta_n\langle \mathbf{B}_{k_n,q}^{\tau}\widehat{\mathbf{C}}_{\rho_n}^{-1/2}\boldsymbol{\beta}, X_i\rangle R'_i\right] r_{i\boldsymbol{\beta}},
\end{aligned}
$$

with $r_{i\boldsymbol{\beta}} \longrightarrow 0$. Considering $\boldsymbol{\beta}$ such that $\|\boldsymbol{\beta}\| = 1$, we have, using relation (I.11)

$$
\begin{aligned}
\text{(I.22)} \quad & L'^2\delta_n^2\langle \mathbf{B}_{k_n,q}^{\tau}\widehat{\mathbf{C}}_{\rho_n}^{-1/2}\boldsymbol{\beta}, X_i\rangle^2 + L'\delta_n\langle \mathbf{B}_{k_n,q}^{\tau}\widehat{\mathbf{C}}_{\rho_n}^{-1/2}\boldsymbol{\beta}, X_i\rangle R'_i \\
\geq \quad & \frac{1}{2}\, L'^2\delta_n^2\langle \mathbf{B}_{k_n,q}^{\tau}\widehat{\mathbf{C}}_{\rho_n}^{-1/2}\boldsymbol{\beta}, X_i\rangle^2 - \frac{C_3^2}{k_n^{2p}}, \quad a.s.
\end{aligned}
$$

Moreover, we set $V_n = \sup_{\|\boldsymbol{\beta}\|=1} \max_{i=1,\dots,n} |r_{i\boldsymbol{\beta}}|$. Using lemma I.3 and relation (I.11), we have

$$L\delta_n \left|\langle \mathbf{B}^{\tau}_{k_n,q}\widehat{\mathbf{C}}^{-1/2}_{\rho_n}\boldsymbol{\beta}, X_i\rangle\right| + |R_i| \leq C_5 \frac{L\delta_n \|\boldsymbol{\beta}\|}{\sqrt{k_n\eta_n}} + \frac{C_3}{k_n^p}.$$

We deduce from this that, for all $\boldsymbol{\beta}$ such that $\|\boldsymbol{\beta}\| = 1$ and for all $i = 1,\dots,n$,

$$\sup_{\|\boldsymbol{\beta}\|=1} \max_{i=1,\dots,n} L\delta_n \left|\langle \mathbf{B}^{\tau}_{k_n,q}\widehat{\mathbf{C}}^{-1/2}_{\rho_n}\boldsymbol{\beta}, X_i\rangle\right| + |R_i| \longrightarrow 0,$$

and $r_{i\boldsymbol{\beta}} \longrightarrow 0$ when $L\delta_n \left|\langle \mathbf{B}^{\tau}_{k_n,q}\widehat{\mathbf{C}}^{-1/2}_{\rho_n}\boldsymbol{\beta}, X_i\rangle\right| + |R_i| \longrightarrow 0$, hence, we can conclude that $\sup_{\|\boldsymbol{\beta}\|=1} \max_{i=1,\dots,n} |r_{i\boldsymbol{\beta}}| \longrightarrow 0$. Then with condition (A.4), we have $\mathbb{1}_{\{V_n<\min_i f_{i\epsilon}(0)/4\}} = \mathbb{1}_{\mathbb{R}}$ for n large enough, and

$$\begin{aligned}
&\left|\left[L'^2\delta_n^2\langle \mathbf{B}^{\tau}_{k_n,q}\widehat{\mathbf{C}}^{-1/2}_{\rho_n}\boldsymbol{\beta}, X_i\rangle^2 + L'\delta_n\langle \mathbf{B}^{\tau}_{k_n,q}\widehat{\mathbf{C}}^{-1/2}_{\rho_n}\boldsymbol{\beta}, X_i\rangle R_i'\right] r_{i\boldsymbol{\beta}}\right| \\
\leq\ & \frac{1}{4}\min_{i=1,\dots,n} f_{i\epsilon}(0)\left|L'^2\delta_n^2\langle \mathbf{B}^{\tau}_{k_n,q}\widehat{\mathbf{C}}^{-1/2}_{\rho_n}\boldsymbol{\beta}, X_i\rangle^2 + L'\delta_n\langle \mathbf{B}^{\tau}_{k_n,q}\widehat{\mathbf{C}}^{-1/2}_{\rho_n}\boldsymbol{\beta}, X_i\rangle R_i'\right| \\
\leq\ & 2\min_{i=1,\dots,n} f_{i\epsilon}(0)\left[\frac{3}{16}\, L'^2\delta_n^2\langle \mathbf{B}^{\tau}_{k_n,q}\widehat{\mathbf{C}}^{-1/2}_{\rho_n}\boldsymbol{\beta}, X_i\rangle^2 + \frac{C_3^2}{8k_n^{2p}}\right].
\end{aligned} \tag{I.23}$$

Using inequalities (I.23) and (I.24), relation (I.22) becomes then

$$\begin{aligned}
&\sum_{i=1}^n \mathbb{E}\left[l_\alpha\left(\epsilon_i - L\delta_n\langle \mathbf{B}^{\tau}_{k_n,q}\widehat{\mathbf{C}}^{-1/2}_{\rho_n}\boldsymbol{\beta}, X_i\rangle - R_i\right) - l_\alpha\left(\epsilon_i - R_i\right)|T_n\right] \\
\geq\ & 2\min_{i=1,\dots,n} f_{i\epsilon}(0)\left[\frac{5}{16}\, L'^2\delta_n^2\sum_{i=1}^n\langle \mathbf{B}^{\tau}_{k_n,q}\widehat{\mathbf{C}}^{-1/2}_{\rho_n}\boldsymbol{\beta}, X_i\rangle^2 - \frac{9}{8}\,\frac{C_3^2 n}{k_n^{2p}}\right].
\end{aligned}$$

Now, we come back to the definition of function f_i to obtain

$$\begin{aligned}
&\frac{1}{\delta_n^2 n} \inf_{\|\boldsymbol{\beta}\|=1} \sum_{i=1}^n \mathbb{E}\left[f_i(L\delta_n\boldsymbol{\beta}) - f_i(\mathbf{0})|T_n\right] \\
\geq\ & 2 \min_{i=1,\ldots,n} f_{i\epsilon}(0) \left[\frac{5L'^2}{16n} \sum_{i=1}^n \langle \mathbf{B}_{k_n,q}^\tau \widehat{\mathbf{C}}_{\rho_n}^{-1/2}\boldsymbol{\beta}, X_i\rangle^2 - \frac{9C_3^2}{8k_n^{2p}\delta_n^2}\right] \\
&+\rho_n L^2 \left\| \left(\mathbf{B}_{k_n,q}^\tau \widehat{\mathbf{C}}_{\rho_n}^{-1/2}\boldsymbol{\beta}\right)^{(m)} \right\|_{L^2}^2 + 2\frac{L\rho_n}{\delta_n} \langle \left(\mathbf{B}_{k_n,q}^\tau \widehat{\mathbf{C}}_{\rho_n}^{-1/2}\boldsymbol{\beta}\right)^{(m)}, \left(\mathbf{B}_{k_n,q}^\tau \boldsymbol{\theta}^\star\right)^{(m)}\rangle.
\end{aligned}$$

Reminding that $L'^2 = 2L^2$ and taking $\xi = \min(\frac{5}{4}\min_{i=1,\ldots,n} f_{i\epsilon}(0), 1)$, we have $\xi > 0$ by hypothesis $(A.4)$ and then

$$\begin{aligned}
&\frac{1}{\delta_n^2 n} \inf_{\|\boldsymbol{\beta}\|=1} \sum_{i=1}^n \mathbb{E}\left[f_i(L\delta_n\boldsymbol{\beta}) - f_i(\mathbf{0})|T_n\right] \\
\geq\ & \xi L^2 \inf_{\|\boldsymbol{\beta}\|=1} \left[\frac{1}{n} \sum_{i=1}^n \langle \mathbf{B}_{k_n,q}^\tau \widehat{\mathbf{C}}_{\rho_n}^{-1/2}\boldsymbol{\beta}, X_i\rangle^2 + \rho_n \left\| \left(\mathbf{B}_{k_n,q}^\tau \widehat{\mathbf{C}}_{\rho_n}^{-1/2}\boldsymbol{\beta}\right)^{(m)} \right\|_{L^2}^2\right] \\
&-\frac{9}{4} \min_{i=1,\ldots,n} f_{i\epsilon}(0) \frac{C_3^2}{k_n^{2p}\delta_n^2} + \frac{2L\rho_n}{\delta_n} \langle \left(\mathbf{B}_{k_n,q}^\tau \widehat{\mathbf{C}}_{\rho_n}^{-1/2}\boldsymbol{\beta}\right)^{(m)}, \left(\mathbf{B}_{k_n,q}^\tau \boldsymbol{\theta}^\star\right)^{(m)}\rangle.
\end{aligned}$$

Using relation (I.19), we get

$$\begin{aligned}
&\frac{1}{\delta_n^2 n} \inf_{\|\boldsymbol{\beta}\|=1} \sum_{i=1}^n \mathbb{E}\left[f_i(L\delta_n\boldsymbol{\beta}) - f_i(\mathbf{0})|T_n\right] \\
\geq\ & \xi L^2 - \frac{9}{4} \min_{i=1,\ldots,n} f_{i\epsilon}(0) \frac{C_3^2}{k_n^{2p}\delta_n^2} + \frac{2L\rho_n}{\delta_n} \langle \left(\mathbf{B}_{k_n,q}^\tau \widehat{\mathbf{C}}_{\rho_n}^{-1/2}\boldsymbol{\beta}\right)^{(m)}, \left(\mathbf{B}_{k_n,q}^\tau \boldsymbol{\theta}^\star\right)^{(m)}\rangle.
\end{aligned}$$

Moreover, for $\|\boldsymbol{\beta}\| = 1$, the infimum of $\langle \left(\mathbf{B}_{k_n,q}^\tau \widehat{\mathbf{C}}_{\rho_n}^{-1/2}\boldsymbol{\beta}\right)^{(m)}, \left(\mathbf{B}_{k_n,q}^\tau \boldsymbol{\theta}^\star\right)^{(m)}\rangle$ is obtained for $\boldsymbol{\beta} = -\widehat{\mathbf{C}}_{\rho_n}^{1/2}\boldsymbol{\theta}^\star / \left\|\widehat{\mathbf{C}}_{\rho_n}^{1/2}\boldsymbol{\theta}^\star\right\|$. Using the fact that the spline approximation has a bounded m-th derivative, we deduce the existence of a constant $C_9 > 0$ such that

$$\inf_{\|\boldsymbol{\beta}\|=1} \langle \left(\mathbf{B}_{k_n,q}^{\tau} \widehat{\mathbf{C}}_{\rho_n}^{-1/2} \boldsymbol{\beta}\right)^{(m)}, \left(\mathbf{B}_{k_n,q}^{\tau} \boldsymbol{\theta}^{\star}\right)^{(m)} \rangle \geq -\frac{C_9}{\sqrt{\eta_n}},$$

hence we obtain

$$\begin{aligned} & \frac{1}{\delta_n^2 n} \inf_{\|\boldsymbol{\beta}\|=1} \sum_{i=1}^{n} \mathbb{E}\left[f_i(L\delta_n \boldsymbol{\beta}) - f_i(\mathbf{0}) | T_n\right] \\ \geq \;& \xi L^2 - \frac{9}{4} \min_{i=1,\dots,n} f_{i\epsilon}(0) \frac{C_3^2}{k_n^{2p} \delta_n^2} - 2C_9 \frac{L\rho_n}{\delta_n \sqrt{\eta_n}}, \end{aligned}$$

that is to say

$$\begin{aligned} & \frac{1}{\delta_n^2 n} \inf_{\|\boldsymbol{\beta}\|=1} \sum_{i=1}^{n} \mathbb{E}\left[f_i(L\delta_n \boldsymbol{\beta}) - f_i(\mathbf{0}) | T_n\right] \\ \geq \;& \xi L^2 \left(1 - \frac{9 \min_{i=1,\dots,n} f_{i\epsilon}(0) C_3^2}{4\xi L^2 k_n^{2p} \delta_n^2} - \frac{2C_9 \rho_n}{\xi L \delta_n \sqrt{\eta_n}}\right). \end{aligned}$$

Noticing that

$$\begin{aligned} &\text{for } \delta_n^2 \sim \frac{1}{n\eta_n}, \text{ we have } \frac{1}{k_n^{2p}\delta_n^2} \sim \frac{n\eta_n}{k_n^{2p}} \xrightarrow[n\to+\infty]{} 0, \\ &\text{for } \delta_n^2 \sim \frac{\rho_n}{k_n\eta_n}, \text{ we have } \frac{\rho_n}{\delta_n\sqrt{\eta_n}} \sim \sqrt{\rho_n k_n} \xrightarrow[n\to+\infty]{} 0, \end{aligned}$$

The last quantity in the inequality above can be made arbitrarily large as n goes to infinity by choosing $L = L_\epsilon$ sufficiently large. This leads to

$$\lim_{n\to+\infty} \mathbb{P}\left(\frac{1}{\delta_n^2 n} \inf_{\|\boldsymbol{\beta}\|=1} \sum_{i=1}^{n} \mathbb{E}\left[f_i(L\delta_n \boldsymbol{\beta}) - f_i(\mathbf{0}) | T_n\right] > 1\right) = 1,$$

which achieves the proof of lemma I.5.

Proof of lemma I.2. — Writing $\Gamma_X = (\Gamma_X - \Gamma_{X,n}) + \Gamma_{X,n}$, we make the following decomposition

$$\|f-g\|_{\Gamma_X}^2 = 2\|\Gamma_X - \Gamma_{X,n}\| \left(\|f\|_{L^2}^2 + \|g\|_{L^2}^2\right) + \|f-g\|_{\Gamma_{X,n}}^2. \tag{I.24}$$

Now, let us decompose f as follows : $f = P + R$ with $P(t) = \sum_{l=0}^{m-1} \frac{t^l}{l!} f^{(l)}(0)$ and $R(t) = \int_0^t \frac{(t-u)^{m-1}}{(m-1)!} f^{(m)}(u)\, du$. P belongs to the space $\mathcal{P}_{m-1}$ of polynomials of degree at most $m-1$, whose dimension is finite and equal to m. Using hypothesis $(A.3)$, there exists a constant $C_6 > 0$ such that we have $\|P\|_{L^2}^2 \leq C_6 \|P\|_{\Gamma_{X,n}}^2$. Then, we can deduce

$$\begin{aligned} \|f\|_{L^2}^2 &\leq 2\|P\|_{L^2}^2 + 2\|R\|_{L^2}^2 \\ &\leq 2C_6\|P\|_{\Gamma_{X,n}}^2 + 2\|R\|_{L^2}^2 \\ &\leq 4C_6\|f\|_{\Gamma_{X,n}}^2 + 4C_6\|\Gamma_{X,n}\|\ \|R\|_{L^2}^2 + 2\|R\|_{L^2}^2. \end{aligned} \tag{I.25}$$

As $\Gamma_{X,n}$ is a bounded operator (by hypothesis $(A.1)$), there exists a constant $C_7 > 0$ such that we have $\|\Gamma_{X,n}\| \leq C_7$. Moreover, under Cauchy-Schwarz inequality, there exists a constant $C_8 > 0$ such that $\|R\|_{L^2}^2 \leq C_8 \|f^{(m)}\|_{L^2}^2$. Relation (I.26) gives $\|f\|_{L^2}^2 \leq 4C_6 \|f\|_{\gamma_{X,n}}^2 + (4C_6C_7 + 2)\, C_8 \|f^{(m)}\|_{L^2}^2$. Then, if we write $f = (f-g) + g$, we finally deduce

$$\begin{aligned} \|f\|_{L^2}^2 \leq\ & 8C_6\|f-g\|_{\Gamma_{X,n}}^2 + (8C_6C_7+4)\, C_8\|(f-g)^{(m)}\|_{L^2}^2 \\ & +8C_6\|\Gamma_{X,n}\|\ \|g\|_{L^2}^2 + (8C_6C_7+4)\, C_8\|g^{(m)}\|_{L^2}^2. \end{aligned} \tag{I.26}$$

We have supposed that $\|g\|_{L^2}$ and $\|g^{(m)}\|_{L^2}$ are bounded, so

$$8C_6\|\Gamma_{X,n}\|\ \|g\|_{L^2}^2 + (8C_6C_7+4)\, C_8\|g^{(m)}\|_{L^2}^2 = O(1),$$

and the hypothesis $\|f-g\|_{\Gamma_{X,n}}^2 + \rho_n\|(f-g)^{(m)}\|_{L^2}^2 = O_{\mathbb{P}}(u_n)$ gives us the bounds $\|f-g\|_{\Gamma_{X,n}}^2 = O_{\mathbb{P}}(u_n)$ and $\|(f-g)^{(m)}\|_{L^2}^2 = O_{\mathbb{P}}(u_n/\rho_n)$. Then, relation (I.27) becomes

$$\|f\|_{L^2}^2 = O_{\mathbb{P}}\Big(1 + \frac{u_n}{\rho_n}\Big). \tag{I.27}$$

Finally, we have $\|\Gamma_X - \Gamma_{X,n}\| = o_{\mathbb{P}}(n^{(\delta-1)/2}) = o_{\mathbb{P}}(\rho_n)$ from lemma 5.3 of Cardot, Ferraty and Sarda (1999). This equality, combined with equations (I.24) and (I.27) gives us $\|f - g\|_{\Gamma_X}^2 = O_{\mathbb{P}}(u_n)$, which is the announced result.

I.3. COMMENTAIRES ET PERSPECTIVES

La vitesse que l'on obtient concernant l'estimation de quantiles conditionnels dans ce cadre fonctionnel est moins bonne que la vitesse nonparamétrique unidimensionnelle usuelle de Stone (1982), qui est en $n^{-2p/(2p+1)}$. Remarquons toutefois que ce résultat de convergence est obtenu sous des hypothèses classiques et relativement faibles. On peut imaginer qu'il est possible de l'atteindre en mettant des hypothèses plus fortes sur X, dans le but par exemple de trouver un η_n plus grand que ρ_n/k_n. Des travaux sont actuellement en cours en collaboration avec Hervé Cardot, Alois Kneip et Pascal Sarda en ce qui concerne l'amélioration de ces vitesses. Ces travaux concernent dans un premier temps l'estimation de la moyenne conditionnelle (voir la partie II de cette thèse), mais on peut raisonnablement espérer obtenir aussi des résultats concernant l'estimation de quantiles conditionnels.

Outre cette amélioration des vitesses de convergence, ce travail sur l'estimation de quantiles conditionnels pour variables explicatives fonctionnelles ouvre plusieurs perspectives. Par exemple, on peut imaginer d'autres méthodes d'estimation dans la catégorie d'estimation de quantiles par des méthodes directes (c'est-à-dire par minimisation d'un critère de type (5)). On peut ainsi envisager par exemple un estimateur en utilisant une autre base que les splines (Fourier, ondelettes, ...). On peut également envisager une méthode d'estimation à noyau, étendant au cas d'une variable explicative fonctionnelle les travaux de Fan, Hu et Truong (1994), adoptant ainsi un point de vue nonparamétrique.

Sur un autre plan, concernant l'hypothèse d'indépendance des X_i, il semble envisageable de pouvoir l'alléger et d'obtenir ainsi des vitesses de convergence pour des données dépendantes, en faisant par exemple des hypothèses de mélange (voir par exemple à ce sujet Rio, 2000). Ferraty, Rabhi et Vieu

(2005) se posent ce problème de l'estimation de quantiles conditionnels pour variables fonctionnelles dépendantes, mais ont un point de vue nonparamétrique et passent par l'inversion de la fonction de répartition conditionnelle pour estimer les quantiles conditionnels.

Enfin, une autre perspective qu'il semble intéressant d'envisager est l'estimation de quantiles conditionnels lorsque non seulement la variable explicative est fonctionnelle mais aussi la variable d'intérêt. On pourrait même envisager dans un premier temps une variable réponse uniquement multivariée. Les travaux déjà cités en introduction de Averous et Meste (1997) et de Cadre (2001) semblent très utiles dans le cadre de cette étude.

PARTIE II

ESTIMATEUR PAR SPLINES DE LISSAGE DANS LE MODÈLE LINÉAIRE FONCTIONNEL

II.1. CONSTRUCTION DE L'ESTIMATEUR

L'objet de cette partie est de construire un estimateur de la moyenne conditionnelle dans le cadre du modèle (2), et d'étudier son comportement asymptotique. Cette étude fait l'objet d'un article à paraître dans *Computational Statistics and Data Analysis* (voir Cardot, Crambes, Kneip et Sarda, 2006). Pour des raisons pratiques, comme cet article étudie aussi le cas où la variable explicative est bruitée, cet article est donné dans la partie III de la thèse (chapitre III.3), puisque cette partie concerne le cas du modèle linéaire fonctionnel lorsque la varaible explicative est bruitée. Dans ce premier chapitre, on donne le principe de construction de l'estimateur. En conservant les notations de l'introduction, on se place dans un modèle à plan fixe ou aléatoire. Le premier cas correspond à des situations où $X_1, \ldots, X_n$ sont des fonctions fixées non aléatoires. On supposera ces courbes dans $L^2([0,1])$. Des exemples de telles situations se trouvent par exemple en chimiométrie où $X_1, \ldots, X_n$ sont des réponses fonctionnelles obtenues sous des conditions expérimentales prédéterminées (voir par exemple à ce sujet Cuevas, Febrero et Fraiman, 2002). Sinon, dans le cas du modèle à plan aléatoire, $X_1, \ldots, X_n$ seront des variables aléatoires à valeurs dans $L^2([0,1])$, de même loi qu'une variable aléatoire X. Sans perte de généralité, on supposera que les variables $X_1, \ldots, X_n$ sont centrées (c'est-à-dire $\mathbb{E}(X) = 0$ dans le modèle à plan aléatoire et $\frac{1}{n}\sum_{i=1}^n X_i = 0$ dans le modèle à plan fixe). Les variables aléatoires $Y_1, \ldots, Y_n$ sont à valeurs dans $\mathbb{R}$, de même loi qu'une variable aléatoire Y. On supposera également qu'il y a indépendance entre les couples (X_1, Y_1), $\ldots$, (X_n, Y_n). Ainsi, les variables aléatoires ϵ_1, $\ldots$, ϵ_n définies dans (2) sont indépendantes identiquement distribuées, de même loi que ϵ, telle que $\mathbb{E}(\epsilon) = 0$. Dans le cas du modèle à plan aléatoire, on suppose de plus que $\mathbb{E}(\|X\|^2) < +\infty$, ce qui assure l'existence de l'opérateur de covariance Γ_X, et que $\mathbb{E}(X\epsilon) = 0$. Pour alléger les preuves, cette dernière hypothèse sera plutôt remplacée par l'hypothèse plus forte que $\mathbb{E}(\epsilon|X) = 0$.

En se plaçant sur l'espace $L^2([0,1])$ muni de son produit scalaire usuel, le modèle (2) s'écrit ainsi, pour tout $i = 1, \ldots, n$,

$$Y_i = \int_0^1 \alpha(t) X_i(t) dt + \epsilon_i, \tag{II.1}$$

le but étant de donner une estimation de α sur la base des observations (X_1, Y_1), $\ldots$, (X_n, Y_n).

En pratique, les courbes n'étant pas entièrement disponibles, on note $t_1 < \ldots < t_p$ les points de mesure de ces courbes que l'on suppose identiques pour toutes les courbes et équidistants entre 0 et 1. Dans cette situation discrétisée, on notera $\langle ., . \rangle_p$ la version discrétisée du produit scalaire de $L^2([0,1])$, défini par

$$\langle f, g \rangle_p = \frac{1}{p} \sum_{j=1}^{p} f(t_j) g(t_j),$$

pour toutes fonctions f et g de $L^2([0,1])$. Cette version discrétisée du produit scalaire usuel de $L^2([0,1])$ en donne une approximation dont la qualité dépend à la fois de la taille de p mais aussi de la régularité des fonctions f et g. La prise en compte de l'erreur due à cette approximation est l'objet de travaux en cours, et ne sera pas présentée ici. Dans cette partie de la thèse, on considèrera donc le modèle (II.1) sous la forme

$$Y_i = \frac{1}{p} \sum_{j=1}^{p} \alpha(t_j) X_i(t_j) + \epsilon_i. \tag{II.2}$$

Dans la suite, on adopte les notations matricielles suivantes. On note $\mathbf{Y} = (Y_1, \ldots, Y_n)^\tau$, $\boldsymbol{\alpha} = (\alpha(t_1), \ldots, \alpha(t_p))^\tau$, $\mathbf{X}_i = (X_i(t_1), \ldots, X_i(t_p))^\tau$ pour tout $i = 1, \ldots, n$ et $\boldsymbol{\epsilon} = (\epsilon_1, \ldots, \epsilon_n)^\tau$. On note de plus $\mathbf{X}$ la matrice de taille $n \times p$ et de terme général $X_i(t_j)$ pour $i = 1, \ldots, n$ et $j = 1, \ldots, p$. En utilisant ces notations matricielles, le modèle (II.2) s'écrit donc

$$\mathbf{Y} = \frac{1}{p}\mathbf{X}\boldsymbol{\alpha} + \boldsymbol{\epsilon}. \tag{II.3}$$

On souhaite maintenant donner une estimation de α basée sur les splines de lissage, plus précisément, on va estimer le vecteur $\boldsymbol{\alpha} \in \mathbb{R}^p$ comme les valeurs aux points de mesure $t_1, \ldots, t_p$ d'une fonction lisse. On va donc supposer que α est une fonction m fois dérivable, avec $m \in \mathbb{N}$. En reprenant les notations sur les splines de lissage vues dans l'introduction de la thèse, l'idée de l'estimation de $\boldsymbol{\alpha}$ à l'aide de splines de lissage amène à chercher $\widehat{\boldsymbol{\alpha}}^*$ solution du problème de minimisation

$$\min_{\boldsymbol{a}\in\mathbb{R}^p} \left\{ \frac{1}{n} \left\| \mathbf{Y} - \frac{1}{p}\mathbf{X}\boldsymbol{a} \right\|^2 + \rho \int_I s_{\boldsymbol{a}}^{(m)}(t)^2 dt \right\}, \tag{II.4}$$

où $\|.\|$ désigne ici la norme euclidienne usuelle de $\mathbb{R}^n$, $s_{\boldsymbol{a}}$ est la spline d'interpolation associée au vecteur $\boldsymbol{a}$, et ρ est un paramètre de lissage permettant à nouveau le compromis entre l'ajustement au données et le lissage de l'estimateur. En utilisant la relation (3), on peut écrire

$$\int_I s_{\boldsymbol{a}}^{(m)}(t)^2 dt = \boldsymbol{a}^\tau \mathbf{A}_m^* \boldsymbol{a},$$

avec

$$\mathbf{A}_m^* = \mathbf{B}\left(\mathbf{B}^\tau\mathbf{B}\right)^{-1} [\int_I \mathbf{b}^{(m)}(t)\mathbf{b}^{(m)}(t)^\tau dt] \left(\mathbf{B}^\tau\mathbf{B}\right)^{-1} \mathbf{B}^\tau.$$

La solution au problème de minimisation (II.4) peut alors s'écrire explicitement par

$$\widehat{\boldsymbol{\alpha}}^* = \frac{1}{np}\left(\frac{1}{np^2}\mathbf{X}^\tau\mathbf{X} + \rho\mathbf{A}_m^*\right)^{-1}\mathbf{X}^\tau\mathbf{Y} = \frac{1}{n}\left(\frac{1}{np}\mathbf{X}^\tau\mathbf{X} + \rho p\mathbf{A}_m^*\right)^{-1}\mathbf{X}^\tau\mathbf{Y}. \tag{II.5}$$

L'étude de cet estimateur $\widehat{\boldsymbol{\alpha}}^*$ dépend alors du comportement de la matrice $\mathbf{A}_m^*$, plus précisément des valeurs propres de la matrice $p\mathbf{A}_m^*$. Par exemple, Utreras (1983) montre que cette matrice $p\mathbf{A}_m^*$ possède m valeurs propres nulles $\mu_{1,p} = \ldots = \mu_{m,p} = 0$, tandis que, lorsque p tend vers l'infini,

$$\sum_{j=m+1}^{p} \frac{1}{\mu_{j,p}} \longrightarrow \sum_{j=m+1}^{\infty} (\pi j)^{-2m}, \tag{II.6}$$

où $0 < \mu_{m+1,p} < \ldots < \mu_{p,p}$ désignent les $p - m$ valeurs propres non nulles de $p\mathbf{A}_m^*$. Comme dans (II.6) la série converge uniquement si $m \neq 0$, on supposera que ceci est vérifié dans la suite.

Le fait que cette matrice $p\mathbf{A}_m^*$ ait m valeurs propres nulles pose problème pour l'inversion de la matrice $\frac{1}{np}\mathbf{X}^\tau\mathbf{X} + \rho p\mathbf{A}_m^*$, donc pour l'existence de l'estimateur. Pour contourner ce problème, on va légèrement modifier l'estimateur $\widehat{\boldsymbol{\alpha}}^*$ en procédant de la façon suivante. On note E_m le sous-espace propre de dimension m associé aux m valeurs propres nulles de $p\mathbf{A}_m^*$, et $\mathbf{P}_m$ la matrice de projection sur ce sous-espace. On définit alors

$$\mathbf{A}_m = \mathbf{P}_m + p\mathbf{A}_m^*.$$

L'estimateur de $\boldsymbol{\alpha}$ sera alors défini par

$$\widehat{\boldsymbol{\alpha}}_{FLS,X} = \frac{1}{np}\left(\frac{1}{np^2}\mathbf{X}^\tau\mathbf{X} + \frac{\rho}{p}\mathbf{A}_m\right)^{-1}\mathbf{X}^\tau\mathbf{Y} = \frac{1}{n}\left(\frac{1}{np}\mathbf{X}^\tau\mathbf{X} + \rho\mathbf{A}_m\right)^{-1}\mathbf{X}^\tau\mathbf{Y}, \tag{II.7}$$

solution du problème de minimisation

$$\min_{\boldsymbol{a}\in\mathbb{R}^p}\left\{\frac{1}{n}\left\|\mathbf{Y} - \frac{1}{p}\mathbf{X}\boldsymbol{a}\right\|^2 + \frac{\rho}{p}\boldsymbol{a}^\tau\mathbf{A}_m\boldsymbol{a}\right\}, \tag{II.8}$$

FLS signifiant "Functional Least Squares". Ainsi, le problème des valeurs propres nulles de $p\mathbf{A}_m^*$ disparaît, la matrice $\mathbf{A}_m$ possédant m valeurs propres égales à 1, les $p-m$ valeurs propres restantes coïncidant avec les valeurs propres $\mu_{m+1,p} < \ldots < \mu_{p,p}$ de $p\mathbf{A}_m^*$. Finalement, l'estimation du paramètre fonctionnel α est définie par

$$\widehat{\alpha}_{FLS,X} = s_{\widehat{\boldsymbol{\alpha}}_{FLS,X}},$$

interpolation spline associée au vecteur $\widehat{\boldsymbol{\alpha}}_{FLS,X}$.

II.2. RÉSULTAT DE CONVERGENCE

On étudie dans ce chapitre le comportement asymptotique de notre estimateur lorsque n et p tendent vers l'infini. Les résultats que l'on présente ici (et leurs preuves) font également partie de l'article de Cardot, Crambes, Kneip et Sarda (2006) donné au chapitre III.3 de la thèse. Cette étude est notamment basée sur le comportement de l'inverse de la matrice $\mathbf{A}_m$. En utilisant la relation (II.6), on a $\mathrm{Tr}\,(\mathbf{A}_m^{-1}) \longrightarrow \sum_{j=m+1}^{\infty}(\pi j)^{-2m} + m =: D_0$ lorsque $p \to \infty$. Ainsi, pour toute constante $D_1 > D_0$, il existe $p_0 \in \mathbb{N}$ tel que, pour tout $p \geq p_0$,

$$\mathrm{Tr}\left(\mathbf{A}_m^{-1}\right) \leq D_1. \tag{II.9}$$

Le comportement asymptotique de l'estimateur $\widehat{\boldsymbol{\alpha}}_{FLS,X}$ de $\boldsymbol{\alpha}$ est étudié relativement à la semi-norme induite par l'opérateur de covariance empirique discrétisé et définie par

$$\|\boldsymbol{u}\|_{\Gamma_{X,n,p}}^2 = \frac{1}{p}\boldsymbol{u}^\tau \left(\frac{1}{np}\mathbf{X}^\tau\mathbf{X}\right)\boldsymbol{u}. \tag{II.10}$$

L'hypothèse de régularité faite sur la fonction α est la suivante. On suppose que α est m fois dérivable et $\alpha^{(m)} \in L^2([0,1])$. Ainsi, on peut définir $D_2 = \int_0^1 \alpha^{(m)}(t)^2 dt$ et $D_3^* = \int_0^1 \alpha(t)^2 dt$. On en déduit que $\frac{1}{p}\boldsymbol{\alpha}^\tau\mathbf{P}_m\boldsymbol{\alpha} \leq \frac{1}{p}\boldsymbol{\alpha}^\tau\boldsymbol{\alpha} \longrightarrow C_3^*$ si $p \to \infty$. Alors, pour toute constante $D_3 > D_3^*$, il existe $p_1 \in \mathbb{N}$ tel que $p_1 \geq p_0$ et $\frac{1}{p}\boldsymbol{\alpha}^\tau\mathbf{P}_m\boldsymbol{\alpha} \leq D_3$ pour tout $p \geq p_1$.

Comme noté auparavant, $X_1, \ldots, X_n$ peuvent être aléatoires ou pas. Dans tous les cas, les espérances du théorème qui suit sont relatives à la loi de probabilité induite par la variable aléatoire ϵ. Lorsque $X_1, \ldots, X_n$ sont aléatoires, ces espérances doivent être vues comme conditionnelles à $X_1, \ldots, X_n$. On donne maintenant le résultat asymptotique concernant $\widehat{\boldsymbol{\alpha}}_{FLS,X}$. Sous l'hypothèse de régularité sur la fonction α faite précédemment ($\alpha^{(m)} \in L^2([0,1])$), avec les définitions de D_1, D_2, D_3, p_1, on obtient pour tout $n \in \mathbb{N}$ et tout $p \geq p_1$

$$\|\mathbb{E}(\widehat{\boldsymbol{\alpha}}_{FLS,X}) - \boldsymbol{\alpha}\|^2_{\Gamma_{X,n,p}} \leq \rho \left(\frac{1}{p} \boldsymbol{\alpha}^\tau \mathbf{P}_m \boldsymbol{\alpha} + D_2 \right) \leq \rho \left(D_3 + D_2 \right),$$

et

$$\frac{1}{p} \mathbb{E} \left(\|\widehat{\boldsymbol{\alpha}}_{FLS,X} - \mathbb{E}(\widehat{\boldsymbol{\alpha}}_{FLS,X})\|^2 \right) \leq \frac{\sigma_\epsilon^2}{n\rho} D_1.$$

Ce résultat (biais et variance) a la conséquence immédiate suivante. Si on suppose de plus que, pour tous n, p (cette inégalité étant presque sûre dans le cas où $X_1, \ldots, X_n$ sont aléatoires),

$$\sup_{i=1,\ldots,n} \sup_{j=1,\ldots,p} |X_i(t_j)| \leq D_4 < +\infty,$$

on a alors, en prenant $\rho \sim n^{-1/2}$ avec $n \to \infty$,

$$\|\widehat{\boldsymbol{\alpha}}_{FLS,X} - \boldsymbol{\alpha}\|^2_{\Gamma_{X,n,p}} = O_P\left(n^{-1/2} \right).$$

II.3. COMMENTAIRES ET PERSPECTIVES

Le résultat précédent donne une vitesse moins bonne que la vitesse nonparamétrique unidimensionnelle usuelle de Stone (1982). Cependant, des travaux en collaboration avec Hervé Cardot, Alois Kneip et Pascal Sarda sont actuellement en cours dans le but d'améliorer cette vitesse, donnant des résultats encourageants. D'autre part, ces travaux prennent aussi en compte le passage du cas discrétisé considéré ici au cas "fonctionnel". Plus précisément, on cherche aussi à établir des résultats de convergence (si possible avec des vitesses aussi bonnes que dans le cas discrétisé) pour $\widehat{\alpha}_{FLS,X} = s_{\widehat{\boldsymbol{\alpha}}_{FLS,X}}$ relativement aux semi-normes $\|.\|_{\Gamma_{X,n}}$ et $\|.\|_{\Gamma_X}$. Ces travaux en cours tiennent également compte de l'approximation du produit scalaire, ce qui n'est pas fait dans cette thèse. On considère en effet ici le modèle s'écrivant $Y_i = \frac{1}{p}\sum_{j=1}^{p} X_i(t_j)\alpha(t_j) + \epsilon_i$ pour $i = 1, \ldots, n$ et non $Y_i = \langle X_i, \alpha \rangle + \epsilon_i$ pour $i = 1, \ldots, n$.

D'autre part, comme cela a été évoqué dans les perspectives d'estimation de quantiles conditionnels pour variables explicatives fonctionnelles, on peut aussi, pour l'estimation de la moyenne conditionnelle envisager la construction d'un estimateur par ondelettes (voir à par exemple ce sujet les livres de Daubechies, 1992 et Cohen, 2003). Les ondelettes ont connu un intérêt grandissant ces dernières années. Elles présentent en effet des avantages par rapport aux bases de Fourier par exemple, comme de permettre la représentation d'un signal à la fois en temps et en échelle.

PARTIE III

MODÈLE LINÉAIRE FONCTIONNEL LORSQUE LA VARIABLE EXPLICATIVE EST BRUITÉE

III.1. MOINDRES CARRÉS ORTHOGONAUX - CAS MULTIVARIÉ

L'objet de ce premier chapitre de la partie III est de donner une description de la méthode des moindres carrés orthogonaux ("Total Least Squares" en anglais, abrégé en TLS) dans le cas où la variable explicative est multivariée, c'est-à-dire élément de $\mathbb{R}^p$, et de donner une méthode algorithmique de résolution de ce problème (voir par exemple Golub et Van Loan, 1980). Ces travaux, bien que connus dans ce cadre multivarié, nous seront très utiles dans notre cadre fonctionnel, c'est pourquoi on les rappelle ici. On considère donc que le modèle s'écrit, pour i allant de 1 à n,

$$\text{(III.1)} \qquad \begin{cases} Y_i = \mathbf{X}_i^\tau \boldsymbol{\alpha} + \epsilon_i, \\ \mathbf{W}_i = \mathbf{X}_i + \boldsymbol{\delta}_i, \end{cases}$$

avec $\mathbf{X}_i = (X_{i1}, \ldots, X_{ip})^\tau$, $\mathbf{W}_i = (W_{i1}, \ldots, W_{ip})^\tau$ et $\boldsymbol{\delta}_i = (\delta_{i1}, \ldots, \delta_{ip})^\tau$ vecteurs de $\mathbb{R}^p$. On doit alors estimer $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_p)^\tau \in \mathbb{R}^p$ sur la base des observations disponibles $(\mathbf{W}_1, Y_1)$, ..., $(\mathbf{W}_n, Y_n)$. Dans toute la suite, on adopte les notations matricielles suivantes : on note $\mathbf{X}$, $\mathbf{W}$ et $\boldsymbol{\delta}$ les matrices de taille $n \times p$ et de termes généraux respectifs X_{ij}, W_{ij} et δ_{ij} pour i allant de 1 à n et j allant de 1 à p, et on note $\mathbf{Y} = (Y_1, \ldots, Y_n)^\tau$ et $\boldsymbol{\epsilon} = (\epsilon_1, \ldots, \epsilon_n)^\tau$. Comme souligné en introduction, l'idée des moindres carrés orthogonaux est de rajouter un terme d'erreur sur la variable explicative dans le problème de minimisation, à savoir

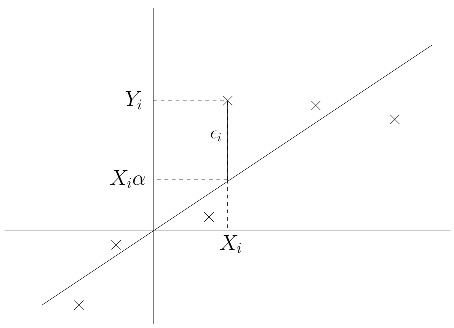
$$\frac{1}{n} \sum_{i=1}^{n} \|\mathbf{W}_i - \mathbf{X}_i\|^2 ,$$

pour finalement déterminer simultanément une estimation de $\boldsymbol{\alpha}$ et $\mathbf{X}_1, \ldots, \mathbf{X}_n$ en résolvant le problème de minimisation

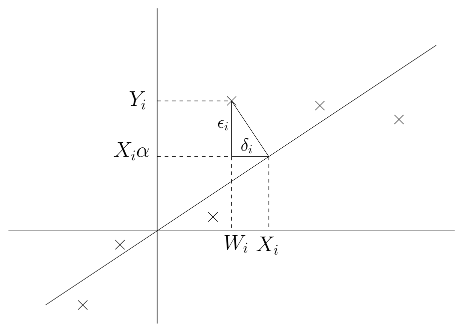
$$\text{(III.2)} \qquad \min_{\boldsymbol{a}\in\mathbb{R}^p, \mathbf{X}_i\in\mathbb{R}^p} \left\{ \frac{1}{n} \sum_{i=1}^{n} \left[(Y_i - \mathbf{X}_i^\tau \boldsymbol{a})^2 + (\mathbf{X}_i - \mathbf{W}_i)^\tau (\mathbf{X}_i - \mathbf{W}_i) \right] \right\}.$$

La représentation graphique dans le cas univarié ($p = 1$) permet de justifier l'appellation de moindres carrés orthogonaux. En effet, lorsqu'on résoud un problème de moindres carrés ordinaires, on cherche la droite qui minimise la somme des carrés des distances "verticales" des points du nuage jusqu'à la droite. En comparaison, lorsqu'on résoud le problème de minimisation (III.2), on cherche en fait la droite qui minimise la somme des carrés des distances "orthogonales" des points du nuage jusqu'à la droite (voir les figures ci-dessous).

MC Ord



MC Orth



La résolution du problème de minimisation (III.2) est basée sur de l'algèbre matricielle, notamment la décomposition en valeurs singulières d'une matrice rectangulaire (un ouvrage de référence en algèbre matricielle est par exemple celui de Golub et Van Loan, 1996). Initialement proposé par Golub et Van Loan (1980), cette méthode des moindres carrés orthogonaux a ensuite été reprise dans un ouvrage de Van Huffel et Vandewalle (1991). Le résultat est le suivant.

Proposition III.1. — *La solution en $\boldsymbol{a} \in \mathbb{R}^p$ au problème de minimisation (III.2), notée $\widehat{\boldsymbol{\alpha}}_{TLS}$, est donnée par*

$$\widehat{\boldsymbol{\alpha}}_{TLS} = \left(\mathbf{W}^\tau \mathbf{W} - \sigma^2_{min} \mathbf{I}_p\right)^{-1} \mathbf{W}^\tau \mathbf{Y}, \tag{III.3}$$

où $\mathbf{I}_p$ désigne la matrice identité de taille p et σ^2_{min} est la plus petite valeur propre non nulle de la matrice $(\mathbf{W}, \mathbf{Y})^\tau (\mathbf{W}, \mathbf{Y})$, où $(\mathbf{W}, \mathbf{Y})$ désigne la matrice obtenue en concaténant les matrices $\mathbf{W}$ et $\mathbf{Y}$.

On remarque cette écriture fait apparaître la solution du problème de minimisation comme une version corrigée de l'estimateur par moindres carrés ordinaires. En effet, il apparaît un terme $-\sigma^2_{\min} \mathbf{I}_p$ qui peut être vu comme un terme de “dérégularisation”. Celui-ci peut s'interpréter comme un terme de correction qui vise à diminuer le biais induit par la présence de la matrice $\mathbf{W}^\tau \mathbf{W}$ dans l'expression de l'estimateur, au lieu de la matrice $\mathbf{X}^\tau \mathbf{X}$ (non disponible) car on observe $\mathbf{W}$ et non $\mathbf{X}$. On donne la preuve de ce résultat, l'idée étant reprise ensuite dans notre cadre de variable explicative fonctionnelle.

Preuve: On introduit la norme de Frobenius d'une matrice $\mathbf{A}$ de terme général a_{ij} ($i = 1, \ldots, n$ et $j = 1, \ldots, p$), notée $\|.\|_F$ et définie par

$$\|\mathbf{A}\|_F^2 = \sum_{i=1}^{n} \sum_{j=1}^{p} a_{ij}^2 = \mathrm{Tr}\left(\mathbf{A}^\tau \mathbf{A}\right).$$

Le problème de minimisation (III.2) s'écrit alors

$$\min_{((\mathbf{W},\mathbf{Y})-(\boldsymbol{\delta},\boldsymbol{\epsilon}))\binom{\alpha}{-1}=0} \|(\boldsymbol{\delta}, \boldsymbol{\epsilon})\|_F^2 ,$$

ce qui revient à résoudre, en notant $\mathbf{A} = (\mathbf{W}, \mathbf{Y})$, $\mathbf{E} = (\boldsymbol{\delta}, \boldsymbol{\epsilon})$ et $\boldsymbol{x} = (\boldsymbol{\alpha}^\tau, -1)^\tau$,

$$\min_{\mathbf{A}\boldsymbol{x}=\mathbf{E}\boldsymbol{x}} \|\mathbf{E}\|_F^2 .$$

En notant $\|.\|$ la norme matricielle euclidienne usuelle, on a ainsi

$$\|\mathbf{E}\|_F^2 \geq \|\mathbf{E}\| = \sup_{\boldsymbol{x}\neq \mathbf{0}} \frac{\boldsymbol{x}^\tau \mathbf{E}^\tau \mathbf{E}\boldsymbol{x}}{\boldsymbol{x}^\tau \boldsymbol{x}} = \sup_{\boldsymbol{x}\neq \mathbf{0}} \frac{\boldsymbol{x}^\tau \mathbf{A}^\tau \mathbf{A}\boldsymbol{x}}{\boldsymbol{x}^\tau \boldsymbol{x}}.$$

Or, on remarque que, si on prend

$$\mathbf{E} = \frac{\mathbf{A}\boldsymbol{x}\boldsymbol{x}^\tau}{\boldsymbol{x}^\tau \boldsymbol{x}},$$

on a bien $\mathbf{A}\boldsymbol{x} = \mathbf{E}\boldsymbol{x}$ et

$$\|\mathbf{E}\|_F^2 = \mathrm{Tr}\left(\mathbf{E}^\tau \mathbf{E}\right) = \frac{\boldsymbol{x}\boldsymbol{x}^\tau \mathbf{A}^\tau \mathbf{A}\boldsymbol{x}\boldsymbol{x}^\tau}{\left(\boldsymbol{x}^\tau \boldsymbol{x}\right)^2} = \frac{\boldsymbol{x}^\tau \mathbf{A}^\tau \mathbf{A}\boldsymbol{x}}{\boldsymbol{x}^\tau \boldsymbol{x}}.$$

Il ne reste alors plus qu'à minimiser $\frac{\boldsymbol{x}^\tau \mathbf{A}^\tau \mathbf{A}\boldsymbol{x}}{\boldsymbol{x}^\tau \boldsymbol{x}}$ en $\boldsymbol{x}$. On considère donc la diagonalisation de $\mathbf{A}^\tau \mathbf{A}$ et on note $\sigma^2_{\min}$ la plus petite valeur propre non nulle, associée au vecteur propre noté $\boldsymbol{v}_{\min}$. Ainsi, la solution au problème de minimisation est obtenue pour $\boldsymbol{x} = k\boldsymbol{v}_{\min}$. La dernière composante (la $(p+1)^{\text{ème}}$) donne la valeur de $k = -1/v_{\min_{p+1}}$. La solution au problème de minimisation (III.2) est donc donnée par

$$\widehat{\boldsymbol{\alpha}}_{TLS} = \left(-\frac{1}{v_{\min_{p+1}}}\right) \begin{pmatrix} v_{\min_1} \\ \vdots \\ v_{\min_p} \end{pmatrix} \quad \text{et} \quad \left(\widehat{\boldsymbol{\delta}}, \widehat{\boldsymbol{\epsilon}}\right) = \mathbf{A}\boldsymbol{v}_{\min}{\boldsymbol{v}_{\min}}^\tau.$$

L'écriture de $\widehat{\boldsymbol{\alpha}}_{TLS}$ s'obtient alors immédiatement de la façon suivante. Comme $\binom{\widehat{\boldsymbol{\alpha}}_{TLS}}{-1}$ est vecteur propre de la matrice $\mathbf{A}^\tau \mathbf{A}$ associé à la valeur propre $\sigma^2_{\min}$, on a donc

$$\begin{pmatrix} \mathbf{W}^\tau \\ \mathbf{Y}^\tau \end{pmatrix} (\mathbf{W}, \mathbf{Y}) \begin{pmatrix} \widehat{\boldsymbol{\alpha}}_{TLS} \\ -1 \end{pmatrix} = \sigma^2_{\min} \begin{pmatrix} \widehat{\boldsymbol{\alpha}}_{TLS} \\ -1 \end{pmatrix},$$

ce qui donne (en considérant les p premières composantes de ce vecteur)

$$\mathbf{W}^\tau \mathbf{W}\widehat{\boldsymbol{\alpha}}_{TLS} - \mathbf{W}^\tau \mathbf{Y} = \sigma^2_{\min}\widehat{\boldsymbol{\alpha}}_{TLS},$$

et cela achève la preuve de la proposition III.1. □

Il est possible, pour faire face à des éventuels problèmes de conditionnement (dû au fait que les valeurs propres de la matrice $\mathbf{W}^\tau\mathbf{W}$ peuvent décroître rapidement vers zéro), de considérer une version régularisée du problème de minimisation (III.2). Ce cas de figure a notamment été envisagé dans un article de Golub, Hansen et O'Leary (1999) qui considèrent le problème de minimisation

(III.4)
$$\min_{\boldsymbol{a}\in\mathbb{R}^p,\mathbf{X}_i\in\mathbb{R}^p}\left\{\frac{1}{n}\sum_{i=1}^n\left[(Y_i-\mathbf{X}_i^\tau\boldsymbol{a})^2+(\mathbf{X}_i-\mathbf{W}_i)(\mathbf{X}_i-\mathbf{W}_i)^\tau\right]+\rho\boldsymbol{a}^\tau\mathbf{L}^\tau\mathbf{L}\boldsymbol{a}\right\},$$

où $\mathbf{L}$ est une matrice de taille $p\times p$ fixée et ρ est un paramètre de régularisation qui permet de contrôler le mauvais conditionnement de la matrice $\mathbf{W}$. Golub, Hansen et O'Leary (1999) montrent alors le résultat suivant, dont la preuve se calque sur celle de la proposition III.1, en incorporant simplement en plus la régularisation (cette preuve ne sera donc pas donnée ici).

Proposition III.2. — *La solution en $\boldsymbol{a}\in\mathbb{R}^p$ au problème de minimisation (III.4), notée $\widehat{\boldsymbol{\alpha}}_{TLS,pen}$, est donnée par*

$$\widehat{\boldsymbol{\alpha}}_{TLS,pen}=\left(\mathbf{W}^\tau\mathbf{W}+\rho\mathbf{L}^\tau\mathbf{L}-\sigma^2_{min,pen}\mathbf{I}_p\right)^{-1}\mathbf{W}^\tau\mathbf{Y}, \tag{III.5}$$

où $\sigma^2_{min,pen}$ est la plus petite valeur propre non nulle de la matrice

$$(\mathbf{W},\mathbf{Y})^\tau(\mathbf{W},\mathbf{Y})+\begin{pmatrix}\mathbf{L}^\tau\mathbf{L} & \mathbf{0}\\ \mathbf{0} & 0\end{pmatrix}.$$

Ce cas faisant intervenir une pénalisation est important dans notre contexte fonctionnel, où, comme cela a été souligné, l'apport d'une pénalisation s'avère fondamental. Ce dernier résultat permettra ainsi la généralisation de la méthode des moindres carrés orthogonaux au cas d'une variable explicative fonctionnelle.

III.2. MOINDRES CARRÉS ORTHOGONAUX - CAS FONCTIONNEL

Dans ce chapitre, le but est de généraliser la méthode des moindres carrés orthogonaux à notre cadre fonctionnel. On souhaite proposer deux estimateurs splines, le premier étant obtenu en modifiant l'estimateur par splines de régression introduit par Cardot, Ferraty et Sarda (1999, 2003), et le second étant basé sur l'estimateur par splines de lissage introduit dans la partie II de la thèse dans le cas où la variable explicative n'était pas bruitée.

On va d'abord s'attacher à donner la méthode de construction de l'estimateur par splines de régression, qui a été étudié le premier chronologiquement (voir Crambes, 2005, pour un premier travail sur ce sujet). Il a été rappelé rapidement dans l'introduction comment l'estimateur de Cardot, Ferraty et Sarda (1999, 2003) est construit. On cherche un estimateur $\widehat{\alpha} = \mathbf{B}_{k,q}^{\tau}\widehat{\boldsymbol{\theta}}$ avec $\widehat{\boldsymbol{\theta}} \in \mathbb{R}^{k+q}$ solution du problème de minimisation (7). Comme souligné précédemment, ce problème de minimisation admet une solution explicite. Celle-ci est donnée par

$$\widehat{\boldsymbol{\theta}}_{FLS,X} = \frac{1}{n}(\frac{1}{n}\mathbf{D}_X^{\tau}\mathbf{D}_X + \rho\mathbf{G}_k)^{-1}\,\mathbf{D}_X^{\tau}\,\mathbf{Y}, \tag{III.6}$$

avec

$$\mathbf{D}_X = \begin{pmatrix} \langle B_1, X_1\rangle & \dots & \langle B_{k+q}, X_1\rangle \\ \vdots & & \vdots \\ \langle B_1, X_n\rangle & \dots & \langle B_{k+q}, X_n\rangle \end{pmatrix},$$

et

$$\mathbf{G}_k = \begin{pmatrix} \langle B_1^{(m)}, B_1^{(m)} \rangle & \dots & \langle B_1^{(m)}, B_{k+q}^{(m)} \rangle \\ \vdots & & \vdots \\ \langle B_{k+q}^{(m)}, B_1^{(m)} \rangle & \dots & \langle B_{k+q}^{(m)}, B_{k+q}^{(m)} \rangle \end{pmatrix}.$$

III.2.1. Construction de l'estimateur (splines de régression)

Ici, les courbes $X_1, \dots, X_n$ ne sont pas connues, les courbes réellement observées $W_1, \dots, W_n$ sont définies par (8). Pour étendre la méthode des moindres carrés orthogonaux à ce contexte, on va donc considérer le problème de minimisation

$$\min_{\boldsymbol{\theta} \in \mathbb{R}^{k+q}, X_i \in L^2(I)} \left\{ \frac{1}{n} \sum_{i=1}^{n} \left[\left(Y_i - \langle \mathbf{B}_{k,q}^{\tau} \boldsymbol{\theta}, X_i \rangle \right)^2 + \left\| \widetilde{\mathbf{X}}_i - \widetilde{\mathbf{W}}_i \right\|^2 \right] + \rho \left\| \left(\mathbf{B}_{k,q}^{\tau} \boldsymbol{\theta} \right)^{(m)} \right\|_{L^2}^2 \right\}, \tag{III.7}$$

où $\widetilde{\mathbf{X}}_i$ et $\widetilde{\mathbf{W}}_i$ sont les versions splines de X_i et W_i. Plus précisément, en notant $\mathbf{X}$ la matrice $n \times p$ de terme général $(X_i(t_j))$ pour $i = 1, \dots, n$ et $j = 1, \dots, p$ et si $\boldsymbol{\beta}$ est la matrice $p \times (k+q)$ de terme général $B_r(t_j)$ pour $j = 1, \dots, p$ et $r = 1, \dots, k+q$, alors $\widetilde{\mathbf{X}}$ est la matrice $n \times (k+q)$ définie par

$$\widetilde{\mathbf{X}} = \mathbf{X}\boldsymbol{\beta},$$

et $\widetilde{\mathbf{X}}_i$ est la $i^{\text{ème}}$ ligne de $\widetilde{\mathbf{X}}$. Maintenant, en utilisant une technique analogue à celle présentée dans le cas multivarié (au chapitre sur les moindres carrés orthogonaux), on montre le résultat suivant.

Proposition III.3. — *La solution en $\boldsymbol{\theta} \in \mathbb{R}^{k+q}$ du problème de minimisation (III.7), notée $\widehat{\boldsymbol{\theta}}_{FTLS}$ (FTLS pour "Functional Total Least Squares") est donnée par*

$$\text{(III.8)} \qquad \widehat{\boldsymbol{\theta}}_{FTLS} = \frac{1}{n}\left(\frac{1}{n}\mathbf{D}_W^\tau \mathbf{D}_W + \rho \mathbf{G}_k - \sigma_{min}^2 \mathbf{B}_k\right)^{-1} \mathbf{D}_W^\tau \mathbf{Y},$$

où les matrices $\mathbf{D}_W$ *et* $\mathbf{B}_k$ *sont définies par*

$$\mathbf{D}_W = \begin{pmatrix} \langle B_1, W_1 \rangle & \dots & \langle B_{k+q}, W_1 \rangle \\ \vdots & & \vdots \\ \langle B_1, W_n \rangle & \dots & \langle B_{k+q}, W_n \rangle \end{pmatrix},$$

$$\mathbf{B}_k = \begin{pmatrix} \langle B_1, B_1 \rangle & \dots & \langle B_{k+q}, B_1 \rangle \\ \vdots & & \vdots \\ \langle B_1, B_{k+q} \rangle & \dots & \langle B_{k+q}, B_{k+q} \rangle \end{pmatrix},$$

et σ_{min}^2 *est la plus petite valeur propre de la matrice*

$$\frac{1}{n}\left(\frac{\mathbf{D}_W}{\sqrt{p}}, \mathbf{Y}\right)^\tau \left(\frac{\mathbf{D}_W}{\sqrt{p}}, \mathbf{Y}\right) + \boldsymbol{\gamma}\left(\boldsymbol{\gamma}^\tau\boldsymbol{\gamma}\right)^{-1}\left(\rho\mathbf{K}_k\right)\left(\boldsymbol{\gamma}^\tau\boldsymbol{\gamma}\right)^{-1}\boldsymbol{\gamma}^\tau,$$

avec $\boldsymbol{\gamma}$ *matrice* $(p+1)\times(k+q+1)$ *donnée par*

$$\boldsymbol{\gamma} = \begin{pmatrix} \frac{\boldsymbol{\beta}}{\sqrt{p}} & \mathbf{0} \\ \mathbf{0} & 0 \end{pmatrix},$$

et $\mathbf{K}_k$ *matrice* $(k+q+1)\times(k+q+1)$ *donnée par*

$$\mathbf{K}_k = \begin{pmatrix} \mathbf{G}_k & \mathbf{0} \\ \mathbf{0} & 0 \end{pmatrix}.$$

La preuve de ce résultat est donnée en annexe (voir partie V.1.). On constate que ce résultat est une analogie directe à la proposition III.2, la matrice identité intervenant dans la "dérégularisation" est remplacée par la matrice $\mathbf{B}_k$.

Le problème qui vient se rajouter ici est que les valeurs propres de la matrice $\frac{1}{n}\left(\frac{\mathbf{D}_W}{\sqrt{p}},\mathbf{Y}\right)^\tau\left(\frac{\mathbf{D}_W}{\sqrt{p}},\mathbf{Y}\right)$ décroissent vers zéro et cela cause des problèmes numériques pour le calcul de σ^2_{min}. Ainsi, on va contourner ce problème par l'intermédiaire d'un résultat qui donne le lien entre $\mathbf{D}_X^\tau\mathbf{D}_X$ (non accessible) et $\mathbf{D}_W^\tau\mathbf{D}_W$, ce qui va permettre de modifier la "dérégularisation". La preuve de ce résultat est également donnée en annexe (voir partie V.1.).

Proposition III.4. — *On fait l'hypothèse suivante.*

$(B.0)$ *Les variables X_i vérifient (p.s.)*

$$\sup_{i=1,\ldots,n}\sup_{t\in[0,1]}|X_i(t)|\leq c_0,$$

où c_0 ne dépend pas de n. On a alors

$$\frac{1}{n}\mathbf{D}_W^\tau\mathbf{D}_W=\frac{1}{n}\mathbf{D}_X^\tau\mathbf{D}_X+\frac{\sigma_\delta^2}{p}\mathbf{B}_k+\mathbf{R}_1, \tag{III.9}$$

où $\mathbf{R}_1$ est une matrice vérifiant

$$\|\mathbf{R}_1\|=O_{\mathbb{P}}\left(\frac{1}{n^{1/2}p^{1/2}k^{1/2}}\right).$$

Enfin, comme σ^2_δ n'est pas toujours connu, on peut l'estimer. On choisit ici de l'estimer nonparamétriquement, en utilisant les travaux de Gasser, Sroka et Jennen-Steinmetz (1986). Comme on est en présence de points de mesure équidistants, cet estimateur est donné par

$$\widehat{\sigma}^2_\delta=\frac{1}{n}\sum_{i=1}^n\frac{1}{6(p-2)}\sum_{j=2}^{p-1}\left[W_i(t_{j-1})-W_i(t_j)+W_i(t_{j+1})-W_i(t_j)\right]^2. \tag{III.10}$$

Finalement, l'estimateur de $\boldsymbol{\theta}$ est donné par

$$\widehat{\boldsymbol{\theta}}_{FTLS} = \frac{1}{n}\left(\frac{1}{n}\mathbf{D}_W^\tau \mathbf{D}_W + \rho \mathbf{G}_k - \frac{\widehat{\sigma}_\delta^2}{p}\mathbf{B}_k\right)^{-1} \mathbf{D}_W^\tau \mathbf{Y}, \tag{III.11}$$

et l'estimateur de α est donné par

$$\widehat{\alpha}_{FTLS,k} = \mathbf{B}_{k,q}^\tau \widehat{\boldsymbol{\theta}}_{FTLS}.$$

III.2.2. Résultat de convergence

Le résultat de convergence à venir donne une borne supérieure pour la vitesse de convergence de $\widehat{\alpha}_{FTLS}$ vers α au sens de la semi-norme induite par l'opérateur de covariance Γ_X. On supposera que $k = k_n \to +\infty$ et $\rho = \rho_n \to 0$ quand $n \to +\infty$. On fait également tendre p vers l'infini. La preuve de ce résultat sera basée sur une décomposition de la forme $\widehat{\alpha}_{FTLS} - \alpha = \widehat{\alpha}_{FTLS} - \widehat{\alpha} + \widehat{\alpha} - \alpha$, où $\widehat{\alpha}$ est l'estimateur de α par splines de régression introduit par Cardot, Ferraty et Sarda (1999, 2003), supposant que les courbes $X_1, \ldots, X_n$ sont directement accessibles et non bruitées. Comme ces derniers ont déjà établi un résultat de convergence concernant $\widehat{\alpha}$, il faudra évaluer l'écart entre cet estimateur et celui par moindres carrés orthogonaux. Ainsi, pour établir notre résultat de convergence, on aura besoin des hypothèses faites par Cardot, Ferraty et Sarda (2003) assurant la convergence de leur estimateur. Ces hypothèses sont les suivantes.

$(B.1)$ La variable X vérifie

$$\|X\| \leq c_1 < +\infty, \quad p.s.$$

$(B.2)$ La fonction α admet une dérivée d'ordre p' et $\alpha^{(p')}$ vérifie

$$\left|\alpha^{(p')}(t) - \alpha^{(p')}(s)\right| \leq c_2 |t - s|^\nu,$$

pour $s, t \in [0, 1]$, où $c_2 > 0$ et $\nu \in [0, 1]$. Dans ce qui suit, on pose $d = p' + \nu$ et on suppose que $q \geq d \geq m$.

$(B.3)$ Les valeurs propres de Γ_X sont strictement positives.

On fait aussi les hypothèses suivantes qui nous permettront de contrôler la vitesse de l'écart entre $\widehat{\alpha}_{FTLS}$ et $\widehat{\alpha}$. Ces hypothèses sont les suivantes.

$(B.0)$ (hypothèse de la proposition III.4) Les variables X_i vérifient (p.s.)

$$\sup_{i=1,\ldots,n}\sup_{t\in[0,1]}|X_i(t)|\leq c_0,$$

où c_0 ne dépend pas de n.

$(B.4)$ Les variables δ_{ij} vérifient

$$\sup_{i=1,\ldots,n}\sup_{j=1,\ldots,p}\mathbb{E}\left(\delta_{ij}^4\right)\leq c_3.$$

où c_3 ne dépend pas de n et de p.

$(B.5)$ Les variables Y_i et δ_{ij} sont indépendantes pour tout $i=1,\ldots,n$ et $j=1,\ldots,p$ et il existe une constante $c_4>0$ indépendante de n telle que $\sup_{i=1,\ldots,n}\mathbb{E}\left(Y_i^2\right)\leq c_4$.

On a alors le résultat suivant.

Proposition III.5. — *Sous les hypothèses qui précèdent, en supposant de plus que $1/p=o(\rho_n/k_n)$, qu'il existe une constante $c_5>0$ indépendante de n telle que $k_n^{1/2}\left\|\left(\langle\Gamma_{X,n}\alpha,B_j\rangle\right)_{j=1,\ldots,k_n+q}\right\|\geq c_5$ et qu'il existe $\beta,\gamma\in]0,1[$ tels que $k_n\sim n^{\beta}$, $\rho_n\sim n^{-(1-\gamma)/2}$, on a*

$$\|\widehat{\alpha}_{FTLS,k_n}-\alpha\|_{\Gamma_X}^2=O_{\mathbb{P}}\left(\frac{1}{k_n^{2d}}+\frac{k_n}{n\rho_n}+\rho_n+\frac{k_n}{np\rho_n}\right).$$

La preuve ne sera pas donnée ici. Elle est simplement basée sur la comparaison entre $\widehat{\alpha}_{FTLS,k_n}$ et $\widehat{\alpha}$ et entre $\widehat{\alpha}$ et α. Un résultat de convergence a déjà été obtenu par Cardot, Ferraty et Sarda (2003) concernant $\|\widehat{\alpha}-\alpha\|_{\Gamma_X}$. Plus précisément, sous les hypothèses $(B.1)-(B.3)$, dès que $1/p=o(\rho_n/k_n)$, on a

$$\|\widehat{\alpha}-\alpha\|_{\Gamma_X}^2=O_{\mathbb{P}}\left(\frac{1}{k_n^{2d}}+\frac{k_n}{n\rho_n}+\rho_n\right).$$

Pour prouver la proposition III.5, on montre finalement que

$$\|\widehat{\alpha}_{FTLS,k_n}-\widehat{\alpha}\|_{\Gamma_X}^2=O_{\mathbb{P}}\left(\frac{k_n}{np\rho_n}\right).$$

ce dernier résultat se prouvant de façon analogue à la preuve du théorème III.2 (voir le chapitre suivant).

III.2.3. Commentaires

Vitesse de convergence. — On remarque que $k_n/(np\rho_n)$ sera négligeable par rapport à $k_n/(n\rho_n)$ si p est assez grand. Cela signifie qu'à partir du moment où le nombre de points de mesure est suffisamment grand, l'effet du bruit est négligeable. Sous cette hypothèse, une vitesse optimale peut être trouvée en choisissant un ρ_n et un k_n particuliers (voir Cardot, Ferraty et Sarda, 2003). En prenant $\rho_n \sim n^{-2d/(4d+1)}$ et $k_n \sim n^{1/(4d+1)}$, on a alors

$$\|\widehat{\alpha}_{FTLS,k_n} - \alpha\|_{\Gamma_X}^2 = O_{\mathbb{P}}\left(n^{-2d/(4d+1)}\right).$$

Effet de la dérégularisation. — Regardons ce qui se passe si on estime $\boldsymbol{\theta}$ sans faire intervenir de dérégularisation, mais en utilisant les courbes disponibles $W_1, \ldots, W_n$. Plus précisément, on a $\widehat{\alpha}_W = \mathbf{B}_{k,q}^{\tau}\widehat{\boldsymbol{\theta}}_W$ avec

$$\widehat{\boldsymbol{\theta}}_W = \frac{1}{n}\left(\frac{1}{n}\mathbf{D}_W^{\tau}\mathbf{D}_W + \rho\mathbf{G}_k\right)^{-1}\mathbf{D}_W^{\tau}\mathbf{Y}.$$

Alors, avec des arguments comparables à ceux utilisés pour la preuve du résultat précédent, on obtient, si p est assez grand, la même vitesse qu'avec l'estimateur par moindres carrés orthogonaux. Dans ce cas, la correction induite par la méthode des moindres carrés orthogonaux ne semble pas avoir un impact fondamental sur la vitesse de convergence. Cependant, on peut voir sur des simulations que l'estimateur par moindres carrés orthogonaux est meilleur (voir le chapitre suivant).

III.2.4. Estimateur par splines de lissage

Jusqu'à présent, on s'est intéressé à l'estimateur par splines de régression. On va maintenant considérer un estimateur par splines de lissage. La construction de cet estimateur est basée sur une correction de l'estimateur par splines de lissages introduit dans la partie II, en s'inspirant directement de la méthode des moindres carrés orthogonaux. La construction de l'estimateur se fait de façon très proche de celle utilisée pour les splines de régression. La dérégularisation diffère, ne faisant plus intervenir la matrice $\mathbf{B}_k$ mais la matrice identité. Cet estimateur est présenté et étudié dans l'article de Cardot, Crambes, Kneip et Sarda (2006), donné dans le chapitre suivant.

III.2.5. Perspectives

Les modèles avec variables bruitées sont certainement amenés à se développer dans le futur, tant ils semblent en adéquation avec la réalité (à partir du moment où on mesure des variables, elles sont nécessairement entachées d'erreurs). Du coup, les perpectives de travaux futurs sur ce modèle défini par (2) et (8) sont multiples.

L'idée la plus naturelle pour supprimer le bruit de la courbe explicative est de la lisser (par exemple par un lissage à noyau). Cette méthode a commencé à être envisagée et des premiers travaux sont en cours. Le début de cette étude est présenté à la partie III.4. de cette thèse. Les premiers résultats semblent encourageants.

Toujours concernant une variable explicative bruitée, il est peut-être envisageable de considérer l'estimation de quantiles conditionnels. La transposition de la méthode des moindres carrés orthogonaux à ce contexte n'est à priori pas évidente, principalement dû au fait que le problème de minimisation relatif à l'estimation de quantiles n'a pas de solution explicite.

Enfin, une autre perspective à envisager dans ce contexte de variable explicative bruitée propose un travail à plus long terme. L'idée serait de considérer un bruit à temps continu. Un problème se pose immédiatement : il n'existe pas de bruit blanc à temps continu. On peut alors envisager un bruit continu $\delta(t)$ qui vérifierait par exemple $Cov\left(\delta(t), \delta(t+h)\right) \rightarrow 0$ lorsque h tend vers zéro.

III.3. FUNCTIONAL LINEAR REGRESSION WITH ERRORS-IN-VARIABLES

III.3.1. Introduction

A very common problem in statistics is to explain the effects of a covariate on a response (variable of interest). While the covariate is usually considered as a vector of scalars, nowadays, in many applications (for instance in climatology, remote sensing, linguistics, ...) the data come from the observation of a continuous phenomenon over time or space : see Ramsay and Silverman (2002) or Ferraty and Vieu (2006) for examples. The increasing performances of measurement instruments permit henceforth to collect these data on dense grids and they can not be considered anymore as variables taking values in $\mathbb{R}^p$. This necessitated to develop for this kind of data *ad hoc* techniques which have been popularized under the name of *functional data analysis* and have been deeply studied these last years (to get a theoretical and practical overview on functional data analysis, we refer to the books of Bosq, 2000, Ramsay and Silverman, 1997, Ramsay and Silverman, 2002 and Ferraty and Vieu, 2006).

Our study takes place in this framework of functional data analysis in the context of regression estimation evocated above. Thus, we consider here the case of a functional covariate while the response is scalar. To be more precise, we first consider observations $(X_i, Y_i)_{i=1,\ldots,n}$, where the X_i's are real functions defined on an interval I of $\mathbb{R}$ with the assumption that it is square integrable over I. As usually assumed in the literature, we then work on the separable real Hilbert space $L^2(I)$ of such functions f defined on I such that $\int_I f(t)^2 dt$ is finite. This space is endowed with its usual inner product $\langle ., . \rangle$ defined by $\langle f, g \rangle = \int_I f(t)g(t)dt$ for $f, g \in L^2(I)$, and the associated norm is noted $\|.\|_{L^2}$. Now, the model we consider to sumarize the link between covariates X_i and responses Y_i is a linear model introduced in Ramsay and Dalzell (1991) and defined by

$$Y_i = \int_I \alpha(t) X_i(t) dt + \epsilon_i, \quad i = 1, \ldots, n, \tag{III.12}$$

where $\alpha \in L^2(I)$ is an unknown functional parameter and $\epsilon_i, i = 1, \ldots, n$ are i.i.d. real random variables satisfying $\mathbb{E}(\epsilon_i) = 0$ and $\mathbb{E}(\epsilon_i^2) = \sigma_\epsilon^2$. The functional parameter α has been estimated in various ways in the literature : see Ramsay and Silverman (1997), Marx and Eilers (1999), Cardot, Ferraty and Sarda (1999) and Cardot, Ferraty and Sarda (2003). Here, our final goal is to deal with the problem of estimating α in the case where $X_i(t)$ is corrupted by some unobservable error.

Before going further, let us note that there can be different ways to generate the curves X_i. One possibility is a fixed design, that is, $X_1, \ldots, X_n$ are fixed, non-random functions. Examples are experiments in chemical or engineering applications, where X_i corresponds to functional responses obtained under various, predetermined experimental conditions (see for instance Cuevas, Febrero and Fraiman, 2002). In other applications one may assume a random design, where $X_1, \ldots, X_n$ are an i.i.d. sample. In any case, $Y_1, \ldots, Y_n$ are independent and the expectations always refer to the probability distribution induced by the random variables $\epsilon_1, \ldots, \epsilon_n$, only. In the case of random design, they thus formally have to be interpreted as conditional expectation given $X_1, \ldots, X_n$. This implies for instance that $\mathbb{E}(\epsilon_i | X_i) = 0$ and $\mathbb{E}(\epsilon_i^2 | X_i) = \sigma_\epsilon^2$.

In what precedes it is implicitly assumed that the curves X_i are observed without error (in model (III.12) all the errors are confined to the variable Y_i by the way of ϵ_i). Unfortunately, this assumption does not seem to be very realistic in practice, and many errors (instrument errors, human errors, ...) prevent to know $X_1, \ldots, X_n$ exactly. Furthermore, it is to be noticed that in practice, the whole curves are not available, so we suppose in the following that the curves are observed in p discretization points $t_1 < \ldots < t_p$ belonging to I, that we will take equispaced in order to simplify. Taking from now on $I = [0, 1]$ in order to simplify the notations, we thus have $t_j - t_{j-1} = \frac{1}{p}$ for all $j = 2, \ldots, p$. Thus, we observe discrete noisy trajectories

$$W_i(t_j) = X_i(t_j) + \delta_{ij}, \; i = 1, \ldots, n, \; j = 1, \ldots, p, \tag{III.13}$$

where $(\delta_{ij})_{i=1,\ldots,n,j=1,\ldots,p}$ is a sequence of independent real random variables, such that, for all $i = 1, \ldots, n$ and $j = 1, \ldots, p$

$$\mathbb{E}(\delta_{ij}) = 0,$$

and

$$\mathbb{E}(\delta_{ij}^2) = \sigma_\delta^2.$$

The noise components δ_{ij} are not discrete realizations of continuous time "random noise" stochastic process and must be interpreted as random measurement errors at the finite discretization points (see *e.g.* Cardot, 2000 and Chiou, Müller and Wang, 2003 for similar points of view).

The problem of the *Errors-in-Variables* linear model has already been studied in many ways in the case where the covariate takes values in $\mathbb{R}$ or $\mathbb{R}^p$, that is to say when it is univariate or multivariate. For instance, the maximum likelihood method has been applied to this context (see Fuller, 1987), and asymptotic results have been obtained (see for example Gleser, 1981). Because this problem is strongly linked to the problem of solving linear systems

$$\mathbf{A}\boldsymbol{x} \approx \mathbf{b},$$

where $\boldsymbol{x} \in \mathbb{R}^p$ is unknown, $\mathbf{b} \in \mathbb{R}^n$ and $\mathbf{A}$ is a matrix of size $n \times p$, some numerical approaches have also been proposed. One of the most famous is the *Total Least Squares* (*TLS*) method (see for example Golub and Van Loan, 1980 or Van Huffel and Vandewalle, 1991).

Now, coming back to model (III.12), very few works have been done in the case of Errors-in-Variables : in a recent work from Chiou, Müller and Wang (2003), a two-step approach is proposed which consists in first smoothing the noisy trajectories in order to get denoised curves and then build functional estimators. The point of view adopted here is quite different and deals with the extension of the TLS approach in the context of the functional linear model.

Let us describe our formal framework for Errors-in-Variables which is inspired from what is done in the literature. We introduce a discretized version of the inner product $\langle ., . \rangle$ will be denoted by $\langle ., . \rangle_p$ and defined for $f, g \in L^2(I)$ by

$$\langle f, g \rangle_p = \frac{1}{p} \sum_{j=1}^{p} f(t_j) g(t_j).$$

This approximation of $\langle ., . \rangle$ by $\langle ., . \rangle_p$ is valid only if p is large enough, so we assume this from now on. In this context of discretized curves, relation (III.12) then writes

$$Y_i = \frac{1}{p} \sum_{j=1}^{p} \alpha(t_j) X_i(t_j) + \epsilon_i, \quad i = 1, \ldots, n. \tag{III.14}$$

Finally the problem is to estimate α using data $(W_i(t_j), Y_i)_{i=1,\ldots,n, j=1,\ldots,p}$ where $W_1(t_j), \ldots, W_n(t_j)$ are noisy observations of $X_1(t_j), \ldots, X_n(t_j)$ for $j = 1, \ldots, p$. The generalization of the TLS method to the case where X_i is a functional random variable is presented. As in the multivariate case, the TLS method consists in a modification of a (penalized) least squares estimator of α for non-noisy observations : see Marx and Eilers (1999) and Cardot, Ferraty and Sarda (2003) for such kind of estimators based on B-splines with two different penalties. Here, we introduce another estimator based on smoothing splines which, as far as we know, has not been studied previously in the literature. Some convergence results are also given in this first section (in the non-noisy case) which serve as a basis for convergence results of the TLS estimator given in the section after. A more detailed study of the asymptotic behavior of the smoothing splines estimator will be the subject of a forthcoming work. Then, the results of convergence for the TLS estimator are commented and we propose some numerical simulations presenting an evaluation of our estimation procedure. Finally, we give the proof of our results.

III.3.2. Estimation of α in the non-noisy case

We adopt the following matrix notations : $\mathbf{Y} = (Y_1, \ldots, Y_n)^\tau$, $\mathbf{X}_i = (X_i(t_1), \ldots, X_i(t_p))^\tau$ for all $i = 1, \ldots, n$, $\boldsymbol{\alpha} = (\alpha(t_1), \ldots, \alpha(t_p))^\tau$ and $\boldsymbol{\epsilon} = (\epsilon_1, \ldots, \epsilon_n)^\tau$. Moreover, we denote by $\mathbf{X}$ the $n \times p$ matrix with general term $X_i(t_j)$ for all

$i = 1, \ldots, n$ and for all $j = 1, \ldots, p$. Using these notations, the model (III.14) then writes

$$\mathbf{Y} = \frac{1}{p}\mathbf{X}\boldsymbol{\alpha} + \boldsymbol{\epsilon}. \tag{III.15}$$

In this section, we first assume that $\boldsymbol{X}$ is observable without errors and our estimation procedure for α is motivated by the popular smoothing splines approach. Then, we want to estimate α as a smooth function, *i.e.* we assume that α is m times differentiable for some fixed $m \in \mathbb{N}^*$.

At first we briefly come back to the smoothing splines procedure in the usual univariate case. For some noisy observations z_i of a smooth function $f(t_i)$ at design points $t_1, \ldots, t_p$, an estimate $\widehat{f}$ is obtained by minimizing $\frac{1}{p}\sum_i (z_i - v(t_i))^2 + \rho \int_I v^{(m)}(t)^2 dt$ for some smoothing parameter $\rho > 0$. Minimization takes place over all functions v in an m-th order Sobolev space, that is $D^m v \in L^2(I)$. It can be shown (for an overview of results in spline theory, consider de Boor, 1978, and Eubank, 1988) that the solution $\widehat{f}$ is in the space $NS^m(t_1, \ldots, t_p)$ of *natural splines* of order $2m$ with knots at $t_1, \ldots, t_p$. This is a p-dimensional linear functions space with $D^m v \in L^2(I)$ for any $v \in NS^m(t_1, \ldots, t_p)$, and there exist basis functions $b_1, \ldots, b_p$ such that $NS^m(t_1, \ldots, t_p) = \left\{\sum_j \theta_j b_j \mid \theta_1, \ldots, \theta_p \in \mathbb{R}\right\}$. Different possible basis functions proposed by various authors are discussed in Eubank (1988). An important property of natural splines is that for any vector $\boldsymbol{w} = (w_1, \ldots, w_p)^\tau \in \mathbb{R}^p$, there exists a unique natural spline interpolant $s_{\boldsymbol{w}}$ with $s_{\boldsymbol{w}}(t_j) = w_j$, $j = 1, \ldots, p$. With $\mathbf{b}(t) = (b_1(t), \ldots, b_p(t))^\tau$ and $\mathbf{B}$ denoting the $p \times p$ matrix with elements $b_i(t_j)$, $s_{\boldsymbol{w}}$ is given by

$$s_{\boldsymbol{w}}(t) = \mathbf{b}(t)^\tau \left(\mathbf{B}^\tau \mathbf{B}\right)^{-1} \mathbf{B}^\tau \boldsymbol{w}. \tag{III.16}$$

Moreover such a spline interpolant satisfies the following fine property

$$\begin{aligned} &\int_I s_{\boldsymbol{w}}^{(m)}(t)^2 dt \leq \int_I f^{(m)}(t)^2 dt \ \text{ for any other function } f \\ &\qquad \text{with } f^{(m)} \in L^2(I) \text{ and } f(t_j) = w_j, j = 1, \ldots, p. \end{aligned} \tag{III.17}$$

The inequality (III.17) implies that the solution $\widehat{f}$ is given by $\widehat{f} = s_{\widehat{\boldsymbol{w}}}$, where $\widehat{\boldsymbol{w}}$ is obtained by minimizing $\frac{1}{p}\sum_i (z_i - w_i)^2 + \rho \int_I s_{\boldsymbol{w}}^{(m)}(t)^2 dt$ over all vectors $\boldsymbol{w} \in \mathbb{R}^p$.

These ideas readily generalize to the problem of estimating $\boldsymbol{\alpha}$ in (III.15) and then the function α. An estimator $\widehat{\boldsymbol{\alpha}}^*_{FLS,X}$ may be obtained by solving the minimization problem

$$\text{(III.18)} \qquad \min_{\boldsymbol{a} \in \mathbb{R}^p} \left\{ \frac{1}{n} \left\| \mathbf{Y} - \frac{1}{p}\mathbf{X}\boldsymbol{a} \right\|^2 + \rho \int_I s_{\boldsymbol{a}}^{(m)}(t)^2 dt \right\},$$

where $\|.\|$ stands for the usual Euclidean norm, and $\rho > 0$ is a smoothing parameter allowing a trade-off between the goodness-of-fit to the data and the smoothness of the fit. By (III.16), we have $\int_I s_{\boldsymbol{a}}^{(m)}(t)^2 dt = \boldsymbol{a}^\tau \mathbf{A}^*_m \boldsymbol{a}$, where $\mathbf{A}^*_m = \mathbf{B}\left(\mathbf{B}^\tau\mathbf{B}\right)^{-1}\left[\int_I \mathbf{b}^{(m)}(t)\mathbf{b}^{(m)}(t)^\tau dt\right]\left(\mathbf{B}^\tau\mathbf{B}\right)^{-1}\mathbf{B}^\tau$ is a $p \times p$ matrix. Therefore, (III.18) can be reformulated in the form

$$\text{(III.19)} \qquad \min_{\boldsymbol{a} \in \mathbb{R}^p} \left\{ \frac{1}{n} \left\| \mathbf{Y} - \frac{1}{p}\mathbf{X}\boldsymbol{a} \right\|^2 + \rho \boldsymbol{a}^\tau \mathbf{A}^*_m \boldsymbol{a} \right\},$$

leading to the solution

$$\widehat{\boldsymbol{\alpha}}^*_{FLS,X} = \frac{1}{np}\left(\frac{1}{np^2}\mathbf{X}^\tau\mathbf{X} + \rho\mathbf{A}^*_m\right)^{-1}\mathbf{X}^\tau\mathbf{Y} = \frac{1}{n}\left(\frac{1}{np}\mathbf{X}^\tau\mathbf{X} + \rho p\mathbf{A}^*_m\right)^{-1}\mathbf{X}^\tau\mathbf{Y}.$$

However, there is a problem with this estimator which is due to the structure of the eigenvalues of $p\mathbf{A}^*_m$. These eigenvalues have been studied by many authors and a discussion of general results is given by Eubank (1988). The most precise results in our context are presented in Utreras (1983). It is shown that this matrix has exactly m zero eigenvalues $\mu_{1,p} = \ldots = \mu_{m,p} = 0$, while as $p \to \infty$,

$$\sum_{j=m+1}^{p} \frac{1}{\mu_{j,p}} \longrightarrow \sum_{j=m+1}^{\infty} (\pi j)^{-2m}, \tag{III.20}$$

where $0 < \mu_{m+1,p} < \ldots < \mu_{p,p}$ denote the $p - m$ non-zero eigenvalues of $p\mathbf{A}_m^*$. The series given in (III.20) converges for $m \neq 0$, so we assume this in the following.

Due to the m zero eigenvalues, existence of $\widehat{\boldsymbol{\alpha}}^*_{FLS,X}$ can only be guaranteed by introducing constraints on the structure of $\mathbf{X}$. This can, however, be avoided by introducing a minor modification of this estimator. The m-dimensional eigenspace corresponding to $\mu_{1,p} = \ldots = \mu_{m,p} = 0$ is the linear vector space generated by all (discretized) polynomials of degree $m-1$, that is, E_m consists of all vectors $\boldsymbol{w} \in \mathbb{R}^p$ with $w_i = \theta_1 + \sum_{j=1}^{m-1} \theta_{j+1} t_i^j$, $i = 1, \ldots, p$, for some coefficients $\theta_1, \ldots, \theta_m$. Let $\mathbf{P}_m$ denote the $p \times p$ projection matrix projecting into the space E_m, and set $\mathbf{A}_m = \mathbf{P}_m + p\mathbf{A}_m^*$. Our final estimator $\widehat{\boldsymbol{\alpha}}_{FLS,X}$ is then defined by

(III.21)

$$\widehat{\boldsymbol{\alpha}}_{FLS,X} = \frac{1}{np}\left(\frac{1}{np^2}\mathbf{X}^\tau\mathbf{X} + \frac{\rho}{p}\mathbf{A}_m\right)^{-1}\mathbf{X}^\tau\mathbf{Y} = \frac{1}{n}\left(\frac{1}{np}\mathbf{X}^\tau\mathbf{X} + \rho\mathbf{A}_m\right)^{-1}\mathbf{X}^\tau\mathbf{Y},$$

and a corresponding estimator of α is provided by $\widehat{\alpha}_{FLS,X} = s_{\widehat{\boldsymbol{\alpha}}_{FLS,X}}$. It is immediately verified that $\widehat{\boldsymbol{\alpha}}_{FLS,X}$ is solution of the modified minimization problem

$$\min_{\boldsymbol{a} \in \mathbb{R}^p} \left\{ \frac{1}{n} \left\| \mathbf{Y} - \frac{1}{p}\mathbf{X}\boldsymbol{a} \right\|^2 + \frac{\rho}{p} \boldsymbol{a}^\tau \mathbf{A}_m \boldsymbol{a} \right\}.$$

By definition, the matrix $\mathbf{A}_m$ possesses m eigenvalues equal to 1, while the remaining $p - m$ eigenvalues coincide with the eigenvalues $\mu_{m+1,p} < \ldots < \mu_{p,p}$ of $p\mathbf{A}_m^*$. Thus, by (III.20), we obtain $\mathrm{Tr}\left(\mathbf{A}_m^{-1}\right) \longrightarrow \sum_{j=m+1}^{\infty} (\pi j)^{-2m} + m =: D_0$ as $p \to \infty$. It follows that for any constant $D_1 > D_0$ there exists a $p_0 \in \mathbb{N}$ such that

$$\mathrm{Tr}\left(\mathbf{A}_m^{-1}\right) \leq D_1, \tag{III.22}$$

for all $p \geq p_0$.

We will now study the behavior of our estimator for large values of n and p. The behavior of our estimator will be evaluated with respect to the semi-norm

$$\|\boldsymbol{u}\|^2_{\Gamma_{X,n,p}} = \frac{1}{p}\boldsymbol{u}^\tau \left(\frac{1}{np}\mathbf{X}^\tau \mathbf{X}\right)\boldsymbol{u}.$$

It is well-known that functional linear regression belongs to the class of ill-posed problems. The semi-norm $\|.\|_{\Gamma_{X,n,p}}$ may be seen as a discretized version of L^2 semi-norms which are usually applied in this context. It is not possible to derive any bound for the bias by using the Euclidean norm. Suppose, for example, that all functions X_i lie in a low dimensional linear function space $\mathcal{X}$. Then any structure of α which is orthogonal to $\mathcal{X}$ cannot be identified from the data.

The regularity assumption that we will do on α follows.

$(H.1)$ For some $m \in \mathbb{N}^*$, α is m times differentiable and $\alpha^{(m)} \in L^2(I)$.

Then, let $D_2 = \int_I \alpha^{(m)}(t)^2 dt$ and $D_3^* = \int_I \alpha(t)^2 dt$. By construction of $\mathbf{P}_m$, $\mathbf{P}_m\boldsymbol{\alpha}$ provides the best approximation (in a least squares sense) of $\boldsymbol{\alpha}$ by (discretized) polynomials of degree m, and $\frac{1}{p}\boldsymbol{\alpha}^\tau\mathbf{P}_m\boldsymbol{\alpha} \leq \frac{1}{p}\boldsymbol{\alpha}^\tau\boldsymbol{\alpha} \longrightarrow D_3^*$ as $p \to \infty$. Let D_3 denote an arbitrary constant with $D_3^* < D_3 < \infty$. There then exists a $p_1 \in \mathbb{N}$ with $p_1 \geq p_0$ such that $\frac{1}{p}\boldsymbol{\alpha}^\tau\mathbf{P}_m\boldsymbol{\alpha} \leq D_3$ for all $p \geq p_1$.

As noticed before, $X_1, \ldots, X_n$ can be either fixed, non-random functions or an i.i.d. sample of random functions. In any case, expected values and variance of $\widehat{\boldsymbol{\alpha}}_{FLS,X}$ as stated in the theorem will refer to the probability distribution induced by the random variable ϵ. In the case of random design, they stand for conditional expectation given $X_1, \ldots, X_n$.

Theorem III.1. — *Under assumption* $(H.1)$ *and the definitions of* D_1, D_2, D_3, p_1, *we obtain for all* $n \in \mathbb{N}$, *all* $p \geq p_1$ *and every matrix* $\mathbf{X} \in \mathbb{R}^n \times \mathbb{R}^p$

$$\text{(III.23)} \qquad \|\mathbb{E}(\widehat{\boldsymbol{\alpha}}_{FLS,X}) - \boldsymbol{\alpha}\|^2_{\Gamma_{X,n,p}} \leq \rho\left(\frac{1}{p}\boldsymbol{\alpha}^\tau\mathbf{P}_m\boldsymbol{\alpha} + D_2\right) \leq \rho\left(D_3 + D_2\right),$$

as well as

$$\text{(III.24)} \qquad \frac{1}{p}\mathbb{E}\left(\|\widehat{\boldsymbol{\alpha}}_{FLS,X} - \mathbb{E}(\widehat{\boldsymbol{\alpha}}_{FLS,X})\|^2\right) \leq \frac{\sigma_\epsilon^2}{n\rho}D_1.$$

Remark When adding some additional constraint like

$$(H.2) \qquad \sup_{i=1,\ldots,n}\sup_{t\in[0,1]}|X_i(t)| \leq D_4 < +\infty,$$

with D_4 independant of n, or when $(H.2)$ is almost surely satisfied in the case of a random design, then we have

$$\mathbb{E}\left(\|\widehat{\boldsymbol{\alpha}}_{FLS,X} - \mathbb{E}(\widehat{\boldsymbol{\alpha}}_{FLS,X})\|^2_{\Gamma_{X,n,p}}\right) \leq \frac{D_4^2}{p}\mathbb{E}\left(\|\widehat{\boldsymbol{\alpha}}_{FLS,X} - \mathbb{E}(\widehat{\boldsymbol{\alpha}}_{FLS,X})\|^2\right),$$

and the theorem implies that

$$\|\widehat{\boldsymbol{\alpha}}_{FLS,X} - \boldsymbol{\alpha}\|^2_{\Gamma_{X,n,p}} = O_{\mathbb{P}}\left(n^{-1/2}\right),$$

if $\rho = \rho_n \sim n^{-1/2}$ as $n \to \infty$. This rate obviously compares favorably to existing rates in the literature.

III.3.3. Total Least Squares method for functional covariates

We address now the estimation of α from noisy covariates. At first, let us describe how the TLS method works in the case of a covariate belonging to $\mathbb{R}^p$. In that case, we have

$$Y_i = \mathbf{X}_i^\tau \boldsymbol{\alpha} + \epsilon_i, \ i = 1, \ldots, n,$$

and

$$\mathbf{W}_i = \mathbf{X}_i + \boldsymbol{\delta}_i, \ i = 1, \ldots, n,$$

where $\boldsymbol{\alpha}$, $\mathbf{X}_i$, $\mathbf{W}_i$ and $\boldsymbol{\delta}_i$ are vectors of $\mathbb{R}^p$. The TLS approach relies on the simultaneous estimation of $\boldsymbol{\alpha}$ and $\mathbf{X}_i$ by considering the minimization problem (see for example Van Huffel and Vandewalle, 1991)

$$\min_{\boldsymbol{a}\in\mathbb{R}^p,\mathbf{X}_i\in\mathbb{R}^p}\left\{\frac{1}{n}\sum_{i=1}^{n}\left[(Y_i-\mathbf{X}_i^\tau\boldsymbol{a})^2+(\mathbf{X}_i-\mathbf{W}_i)^\tau(\mathbf{X}_i-\mathbf{W}_i)\right]\right\}. \tag{III.25}$$

The TLS algorithm solving (III.25) is given in Van Huffel and Vandewalle (1991). In some cases, the singular values of the matrix $\mathbf{W}$ can quickly decrease to zero, and the minimization problem (III.25) is then *ill-conditioned*. A possible way to circumvent this problem is to introduce a regularization in (III.25), and the minimization problem we consider is then

$$\begin{aligned}\min_{\boldsymbol{a}\in\mathbb{R}^p,\mathbf{X}_i\in\mathbb{R}^p}\Bigg\{\frac{1}{n}\sum_{i=1}^{n}&\left[(Y_i-\mathbf{X}_i^\tau\boldsymbol{a})^2+(\mathbf{X}_i-\mathbf{W}_i)(\mathbf{X}_i-\mathbf{W}_i)^\tau\right]\\&+\rho\boldsymbol{a}^\tau\mathbf{L}^\tau\mathbf{L}\boldsymbol{a}\Bigg\},\end{aligned} \tag{III.26}$$

where $\mathbf{L}$ is a $p\times p$ matrix and ρ is a regularization parameter allowing to deal with the ill-conditioning of the design matrix $\mathbf{W}^\tau\mathbf{W}$ (see Golub, Hansen and O'Leary, 1999). Indeed, the TLS solution in $\boldsymbol{a}\in\mathbb{R}^p$ to the minimization problem (III.26) is given by

$$\widehat{\boldsymbol{\alpha}}_{TLS,pen}=\left(\mathbf{W}^\tau\mathbf{W}+\rho\mathbf{L}^\tau\mathbf{L}-\sigma_k^2\mathbf{I}_p\right)^{-1}\mathbf{W}^\tau\mathbf{Y}, \tag{III.27}$$

where σ_k is the smallest non-zero singular value of the matrix $(\mathbf{W},\mathbf{Y})$ and $\mathbf{I}_p$ is the $p\times p$ identity matrix.

In our functional situation, we consider model (III.15) and using the same matricial notations as in the previous section, we write

$$\mathbf{W}=\mathbf{X}+\boldsymbol{\delta},$$

where $\mathbf{W}$ and $\boldsymbol{\delta}$ are the $n \times p$ matrices with respective general terms $W_i(t_j)$ and δ_{ij}. So, the minimization problem we consider now is the following one : we are looking for an estimation $\widehat{\boldsymbol{\alpha}}^*_{FTLS}$ of $\boldsymbol{\alpha}$, solution of the minimization problem

$$\text{(III.28)} \min_{\boldsymbol{a}\in\mathbb{R}^p, \mathbf{X}_i\in\mathbb{R}^p} \left\{ \frac{1}{n} \sum_{i=1}^{n} \left[\left(Y_i - \frac{1}{p} \mathbf{X}_i^\tau \boldsymbol{a} \right)^2 + \frac{1}{p} \|\mathbf{X}_i - \mathbf{W}_i\|^2 \right] + \frac{\rho}{p} \boldsymbol{a}^\tau \mathbf{A}_m \boldsymbol{a} \right\},$$

where the matrix $\mathbf{A}_m$ is the one introduced previously. Now, with these notations, we have the following result.

Proposition III.6. — *The solution in $\boldsymbol{a} \in \mathbb{R}^p$ of the minimization problem (III.28) is given by*

$$\text{(III.29)} \qquad \widehat{\boldsymbol{\alpha}}^*_{FTLS} = \frac{1}{np} \left(\frac{1}{np^2} \mathbf{W}^\tau \mathbf{W} + \frac{\rho}{p} \mathbf{A}_m - \sigma_k^2 \mathbf{I}_p \right)^{-1} \mathbf{W}^\tau \mathbf{Y},$$

where σ_k^2 is the smallest non-zero eigenvalue of the matrix

$$\frac{1}{n} \left(\frac{\mathbf{W}}{p}, \mathbf{Y} \right)^\tau \left(\frac{\mathbf{W}}{p}, \mathbf{Y} \right) + \frac{\rho}{p} \begin{pmatrix} \mathbf{A}_m & \mathbf{0} \\ \mathbf{0} & 0 \end{pmatrix}.$$

In equation (III.29), computational problems can appear due to the value of σ_k^2 which may be close to zero. Indeed, the eigenvalues of $\frac{1}{n}\left(\frac{\mathbf{W}}{p}, \mathbf{Y}\right)^\tau \left(\frac{\mathbf{W}}{p}, \mathbf{Y}\right)$ are known to decrease rapidly to zero, and this can of course cause numerical problems with the computation of σ_k^2. Nevertheless, we can circumvent this problem using the following result.

Proposition III.7. — *Suppose that for some constant $D_5 > 0$ independant of n and p,*

$$(H.3) \qquad \mathbb{E}\left(\delta_{ir}^4\right) \leq D_5.$$

Then, if moreover $(H.2)$ holds, we have

(III.30)
$$\frac{1}{np^2}\mathbf{W}^\tau\mathbf{W} = \frac{1}{np^2}\mathbf{X}^\tau\mathbf{X} + \frac{\sigma_\delta^2}{p^2}\mathbf{I}_p + \mathbf{R},$$

where $\mathbf{R}$ *is a matrix such that* $\|\mathbf{R}\| = O_{\mathbb{P}}\left(\frac{1}{n^{1/2}p}\right)$, $\|.\|$ *being the usual norm of a matrix.*

The last problem is that σ_δ^2 is not always known. There are several ways to estimate it. We choose to use the estimator presented in Gasser, Sroka and Jennen-Steinmetz (1986) and given by (as we are in the case of equispaced measurement points)

(III.31)
$$\widehat{\sigma}_\delta^2 = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{6(p-2)}\sum_{j=2}^{p-1}\left[W_i(t_{j-1}) - W_i(t_j) + W_i(t_{j+1}) - W_i(t_j)\right]^2.$$

This leads us to change the former estimator of $\boldsymbol{\alpha}$ given by (III.29) and to take instead

(III.32)
$$\widehat{\boldsymbol{\alpha}}_{FTLS} = \frac{1}{np}\left(\frac{1}{np^2}\mathbf{W}^\tau\mathbf{W} + \frac{\rho}{p}\mathbf{A}_m - \frac{\widehat{\sigma}_\delta^2}{p^2}\mathbf{I}_p\right)^{-1}\mathbf{W}^\tau\mathbf{Y},$$

and again a corresponding estimator of α is provided by $\widehat{\alpha}_{FTLS}$.

The asymptotic behavior $\widehat{\boldsymbol{\alpha}}_{FTLS}$ is given in the following theorem.

Theorem III.2. — *Under assumptions* $(H.1)$ - $(H.3)$, *if we assume moreover that* Y_i *and* δ_{ij} *are independent for all* $i = 1, \ldots, n$, $j = 1, \ldots, p$ *and that there exists a constant* $0 < D_6 < +\infty$ *such that, for all* n *and* p *sufficiently large* $p^{1/2}\left\|\frac{1}{np^2}\mathbf{X}^\tau\mathbf{X}\boldsymbol{\alpha}\right\| \geq D_6$, *then we have*

(III.33)
$$\|\widehat{\boldsymbol{\alpha}}_{FTLS} - \widehat{\boldsymbol{\alpha}}_{FLS,X}\|_{\Gamma_{X,n,p}} = O_{\mathbb{P}}\left(\frac{1}{n^{1/2}p^{1/2}\rho^{1/2}} + \frac{1}{n^{1/2}}\right).$$

III.3.4. Some comments

(i) In the expression (III.32) of the estimator of $\boldsymbol{\alpha}$, the term $-\frac{\widehat{\sigma}_\delta^2}{p^2}\mathbf{I}_p$ acts as a deregularization term. It allows us to deal with the bias introduced by the fact that we only know the matrix $\mathbf{W}$ instead of the "true" one $\mathbf{X}$.

(ii) In theorem III.2, the hypothesis $p^{1/2}\left\|\frac{1}{np^2}\mathbf{X}^\tau\mathbf{X}\boldsymbol{\alpha}\right\| \geq D_6$ means (in the case of the random design) that α does not belong to the kernel of the covariance operator Γ_X.

(iii) An immediate corollary of theorems III.1 and III.2 is

$$\|\widehat{\boldsymbol{\alpha}}_{FTLS}-\boldsymbol{\alpha}\|^2_{\Gamma_{X,n,p}} = O_{\mathbb{P}}\left(\frac{1}{n\rho}+\rho+\frac{1}{np\rho}+\frac{1}{n}\right).$$

If we compare these terms, we can see that, for p large enough, it remains

$$\|\widehat{\boldsymbol{\alpha}}_{FTLS}-\boldsymbol{\alpha}\|^2_{\Gamma_{X,n,p}} = O_{\mathbb{P}}\left(\frac{1}{n\rho}+\rho\right),$$

and then, for $\rho=\rho_n\sim n^{-1/2}$,

$$\|\widehat{\boldsymbol{\alpha}}_{FTLS}-\boldsymbol{\alpha}\|^2_{\Gamma_{X,n,p}} = O_{\mathbb{P}}\left(n^{-1/2}\right).$$

This means that we obtain the same upper bound for the convergence speed of the $FTLS$ estimator as the FLS estimator using the true curves $X_1,\ldots,X_n$. This result is in accordance with the intuition. The estimation will be improved for a high number p of discretization points.

(iv) Let us see what happens for the FLS estimator using the noisy curves $W_1,\ldots,W_n$. The estimator of $\boldsymbol{\alpha}$ is then given by

$$\widehat{\boldsymbol{\alpha}}_{FLS,W} = \frac{1}{np}\left(\frac{1}{np^2}\mathbf{W}^\tau\mathbf{W} + \frac{\rho}{p}\mathbf{A}_m\right)^{-1}\mathbf{W}^\tau\mathbf{Y}. \tag{III.34}$$

A calculus analogous to the one used in the proof of theorem III.2 leads us to

$$\|\widehat{\boldsymbol{\alpha}}_{FLS,W} - \widehat{\boldsymbol{\alpha}}_{FLS,X}\|_{\Gamma_{X,n,p}} = O_{\mathbb{P}}\left(\frac{1}{n^{1/2}p^{1/2}\rho^{1/2}} + \frac{1}{n^{1/2}}\right),$$

that is to say we have the same upper bound of convergence speed for $\widehat{\boldsymbol{\alpha}}_{FLS,W}$ and $\widehat{\boldsymbol{\alpha}}_{FTLS}$. Nevertheless, the results obtained in the simulations allow us to think that we improve the estimation (see last remark) using the $FTLS$ estimator instead of the FLS estimator with the noisy curves $W_1, \ldots, W_n$.

(v) Using some heuristic arguments to expand the mean quadratic error of estimation of $\boldsymbol{\alpha}$ (similarly to what is done in Cardot, 2000), we can see that it is generally better to consider the $FTLS$ estimator compared to the FLS one with the variable W. More precisely, using the same notations as before, let us denote

$$\widehat{\boldsymbol{\alpha}}(\lambda) = \frac{1}{np}\left(\frac{1}{np^2}\mathbf{W}^\tau\mathbf{W} + \frac{\rho}{p}\mathbf{A}_m - \lambda\mathbf{I}_p\right)^{-1}\mathbf{W}^\tau\mathbf{Y},$$

where λ is a positive real number such that the matrix $\frac{1}{np^2}\mathbf{W}^\tau\mathbf{W} + \frac{\rho}{p}\mathbf{A}_m - \lambda\mathbf{I}_p$ is positive definite. Then we have the following result, which proof is given in the last section.

Proposition III.8. — *Let* $MISE(\lambda) = \mathbb{E}\left[(\widehat{\boldsymbol{\alpha}}(\lambda) - \boldsymbol{\alpha})^\tau(\widehat{\boldsymbol{\alpha}}(\lambda) - \boldsymbol{\alpha})\right]$. *If we assume that* $(\mathbf{W}^\tau\mathbf{W})^{-1}$ *exists and if* $\rho\,\|\mathbf{A}_m\|$ *is negligible compared to* $\left\|\frac{1}{np}\mathbf{W}^\tau\mathbf{W}\right\|$, *then we have*

$$\frac{\partial}{\partial\lambda}MISE(\lambda)_{|\lambda=0} < 0.$$

In other words, this result means that it is advantageous to put a deregularization term $-\lambda \mathbf{I}_p$ (with a small positive λ) in order to improve the quality of the estimation relatively to the $MISE$ criterion.

III.3.5. A simulation study

Presentation of the simulation. — The aim of this simulation is to evaluate the performances of our estimator $\widehat{\boldsymbol{\alpha}}_{FTLS}$, and to compare it with $\widehat{\boldsymbol{\alpha}}_{FLS,W}$. We also compare $\widehat{\boldsymbol{\alpha}}_{FTLS}$ to $\widehat{\boldsymbol{\alpha}}_{FLS,\widetilde{W}}$, which is given by the same formula (III.34) where the curve W is now replaced by a smoothed version $\widetilde{W}$. We can think that this smoothing step has a correcting effect on the noisy curve W, and then this smoothed curve $\widetilde{W}$ can be expected to be closer than W to the unknown "true" curve X. This gives us the intuition that the estimator $\widehat{\boldsymbol{\alpha}}_{FLS,\widetilde{W}}$ should be better than $\widehat{\boldsymbol{\alpha}}_{FLS,W}$. To obtain a smoothed version $\widetilde{W}$ of W, we choose to use the Nadaraya-Watson kernel estimator (see for example Härdle,1991 or Sarda and Vieu, 2000). In the simulations, the kernel is the standard normal kernel. For the bandwidth we have tried at first a value chosen by cross validation for each curve (see Härdle,1991). We have also tried several other bandwidth values applying to this cross-validated bandwidth a decreasing or increasing factor. In order to synthesize results, we only give the simulation results when X is non-random (when X is random, the simulation we have done lead to the same kind of conclusions). We have simulated $N = 100$ samples, each being composed of $n = 200$ observations $(W_i, Y_i)_{i=1,\ldots,n}$ from the model given by (III.12) and (III.13), where the fixed design curves $X_1, \ldots, X_n$ are defined on $I = [0, 1]$ by

$$X_i(t) = \begin{cases} 10\sin(2\pi i t) \text{ if } i \text{ is even,} \\ 10\cos(2\pi i t) \text{ if } i \text{ is odd,} \end{cases}$$

similarly to what is used for the simulation in Cuevas, Febrero and Fraiman. Each sample is randomly split into a learning sample of length $n_l = 100$ (this sample is used to build the estimator) and a test sample of length $n_t = 100$ (this sample is used to see the quality of the estimator by the way of computation of error terms). We made simulations for different numbers of discretization points, $p = 50$, $p = 100$ and $p = 200$. Two functions α were considered, either $\alpha(t) = 10\sin(2\pi t)$ or $\alpha(t) = 10\sin^3(2\pi t^3)$. Finally, the error terms were chosen as follows : $\epsilon \sim \mathcal{N}(0, \sigma_\epsilon^2)$ with $\sigma_\epsilon = 0.2$ and $\delta_{ij} \sim \mathcal{N}(0, \sigma_\delta^2)$ for all $i = 1, \ldots, n$, $j = 1, \ldots, p$ with either $\sigma_\delta = 0.1$, $\sigma_\delta = 0.2$ or $\sigma_\delta = 0.5$. Concerning the parameters of the spline functions, the order of differentiation in the penalization is fixed to the value $m = 2$. The most important parameter to choose is the smoothing parameter value ρ (see Marx and Eilers, 1999). We present in the next subsection a criterion allowing to select reasonable values and we check the effectiveness of this criterion in the simulations.

Generalized Cross Validation criteria. — In the setting of the estimation of a function f by smoothing splines described in section 2, the most popular method for the selection of ρ is *generalized cross-validation* (see Wahba, 1990). The first idea is to use *cross-validtion* that is to choose the ρ that yields the best prediction (in a mean squares sense) when prediction of a value is done with the remaining observations. After this, a computational simplification of the cross-validation criterion has been proposed in the literature that leads to the generalized cross-validation : see Wahba (1990). In our Functional Least Squares estimation, we can easily adapt this generalized cross-validation (GCV) in the following way. The GCV criterion is defined by

$$GCV_{FLS,W}(\rho) = \frac{\frac{1}{n}\sum_{i=1}^{n}(Y_i - \widehat{Y}_i)^2}{\left(1 - \frac{1}{n}\mathrm{Tr}(\mathbf{H}_{FLS,W}(\rho))\right)^2}, \tag{III.35}$$

where $\mathbf{H}_{FLS,W}(\rho)$ is the "hat matrix" given by

$$\mathbf{H}_{FLS,W}(\rho) = \frac{1}{np}\mathbf{W}\left(\frac{1}{np^2}\mathbf{W}^\tau\mathbf{W} + \frac{\rho}{p}\mathbf{A}_m\right)^{-1}\mathbf{W}^\tau,$$

and $\widehat{\mathbf{Y}} = \mathbf{H}_{FLS,W}(\rho)\mathbf{Y}$. Then, we select the optimal parameter ρ_{GCV} as the one that minimizes the GCV criterion (III.35). The criterion (III.35) is a direct adaptation of the one introduced in Wahba (1990) except that the "hat matrix" has been changed for our setting.

Concerning the Functional Total Least Squares estimation, although *Cross Validation* has already been studied in Sima and Van Huffel (2004), what we want to propose here is a generalization of the GCV criterion (III.35), in the following way. The prediction of Y_i for $i = 1, \ldots, n$ is slightly different in the context of TLS. The estimation of the unknown $\mathbf{X}_i$, noted $\widehat{\mathbf{X}}_i$, is given by

$$\widehat{\mathbf{X}}_i = \mathbf{W}_i + \frac{Y_i - \frac{1}{p}\widehat{\boldsymbol{\alpha}}^\tau \mathbf{W}_i}{1 + \frac{1}{p}\left\|\widehat{\boldsymbol{\alpha}}\right\|^2}\,\widehat{\boldsymbol{\alpha}}, \tag{III.36}$$

obtained as in Fuller (1987) by differentiating equation (III.28) with respect to $\mathbf{X}_i$. Then, we take $\widehat{Y}_i = \langle \widehat{\boldsymbol{\alpha}}, \widehat{\mathbf{X}}_i \rangle_p$ as the prediction of Y_i. Then, the proposed GCV criterion is given by

$$GCV_{FTLS}(\rho) = \frac{\frac{1}{n}\sum_{i=1}^{n}(Y_i - \langle \widehat{\boldsymbol{\alpha}}_{FTLS}, \widehat{\mathbf{X}}_i \rangle_p)^2}{\left(1 - \frac{1}{n}\mathrm{Tr}(\mathbf{H}_{FTLS}(\rho))\right)^2}, \tag{III.37}$$

where $\mathbf{H}_{FTLS}(\rho)$ is the "hat matrix" given by

$$\mathbf{H}_{FTLS}(\rho) = \frac{1}{np}\mathbf{W}\left(\frac{1}{np^2}\mathbf{W}^\tau\mathbf{W} + \frac{\rho}{p}\mathbf{A}_m - \frac{\widehat{\sigma}_\delta^2}{p^2}\mathbf{I}_p\right)^{-1}\mathbf{W}^\tau$$

Then, the optimal parameters ρ_{GCV} as obtained by minimizing the GCV criterion (III.37). In our simulations, these GCV criteria have been computed for ρ over a grid taking its values among $10^{-2}, 10^{-3}, \ldots, 10^{-8}$.

Results of the simulation. — We use two error criteria to see the quality of the prediction. The first one is the relative mean square error of the estimator of $\boldsymbol{\alpha}$, given by

$$E_1 = \frac{\sum_{j=1}^{p}\left[\widehat{\boldsymbol{\alpha}}(t_j) - \boldsymbol{\alpha}(t_j)\right]^2}{\sum_{j=1}^{p}\boldsymbol{\alpha}(t_j)^2}, \tag{III.38}$$

and the second one is the mean square error of the prediction of $\mathbf{Y}$, given by

$$E_2 = \frac{1}{n} \sum_{i=1}^{n} \left(\widehat{Y}_i - Y_i \right)^2 . \tag{III.39}$$

These errors, evaluated on $N = 100$ simulated test samples, are given in tables 1 and 3 for the different values of p and the different functions α. We have computed the FLS estimator of α using the unknown true curves X (in order to have a reference), the observed curves W and the smooth version $\widetilde{W}$ of the observed curves W. Adopting the strategy described in 5.1, we have selected bandwidth values for $\widetilde{W}$ leading to best results in term of estimation of α. Indeed, for the first example, it appears to us that undersmoothing the curves W had some benefit on the estimation of α whereas the cross-validated bandwidth was selected for the second example. Although it is beyond the scope of this paper, we note that this problem of the choice of the bandwidth in the "smoothing" procedure has to be investigated more deeply since it is a crucial problem as it is pointed out for instance in Chiou, Müller and Wang (2003).

We can see that the $FTLS$ estimator always improves the prediction compared to FLS, W, and the improvement is really interesting when p is small with a relatively important noise level σ_δ. We can also see that the estimators $FTLS$ and $FLS, \widetilde{W}$ are quite close. $FLS, \widetilde{W}$ seems to be better when the noise level σ_δ is small whereas $FTLS$ seems to be better when this noise level becomes high. Nevertheless, it is important to note that the $FTLS$ estimator is faster to compute compared to the $FLS, \widetilde{W}$ one. Moreover, the choice of the parameter h is not evident (choosing it by cross validation is not always the best thing to do whereas it implies additional long computation times, particularly when n is large).

Moreover, it has to be noticed that the prediction is also improved when the number of discretization points increases. We can also see that the error increases between table 1 and table 3, mainly because of the shape of the second function α, which is less smooth than the first one.

Table 2 gives the estimated values of σ_δ using the estimator defined by (III.31) and given in Gasser, Sroka and Jennen-Steinmetz (1986). We can see that we get good estimations of σ_δ, and an increasing accuracy with the number of discretization points. It also seems that the quality of the estimation is not much related to the value of σ_δ. Finally, we have plotted on figure 1 an example of the estimation of α in the case where $p = 100$ and $\sigma_\delta = 0.5$, in the case of the function $\alpha(t) = 10\sin^3(2\pi t^3)$. In order not to have too many curves on a same graphic, we choose to plot only the estimators $FTLS$, FLS, X and FLS, W. This graphic tends to confirm the values given in tables 1 and 3.

		E_1			E_2		
		$\sigma_\delta = 0.1$	$\sigma_\delta = 0.2$	$\sigma_\delta = 0.5$	$\sigma_\delta = 0.1$	$\sigma_\delta = 0.2$	$\sigma_\delta = 0.5$
FLS, X	$p = 50$	0.00015	0.00014	0.00013	0.0031	0.0032	0.0032
	$p = 100$	0.00009	0.00010	0.00009	0.0027	0.0026	0.0027
	$p = 200$	0.00005	0.00006	0.00004	0.0024	0.0026	0.0025
$FTLS$	$p = 50$	0.00018	0.00061	0.00232	0.0044	0.0067	0.0180
	$p = 100$	0.00013	0.00065	0.00219	0.0040	0.0063	0.0139
	$p = 200$	0.00009	0.00057	0.00204	0.0035	0.0056	0.0091
$FLS, \widetilde{W}$	$p = 50$	0.00017	0.00080	0.00245	0.0040	0.0065	0.0209
	$p = 100$	0.00011	0.00063	0.00226	0.0036	0.0062	0.0154
	$p = 200$	0.00006	0.00056	0.00210	0.0029	0.0056	0.0112
FLS, W	$p = 50$	0.00020	0.00098	0.00366	0.0050	0.0081	0.0305
	$p = 100$	0.00015	0.00079	0.00344	0.0045	0.0072	0.0245
	$p = 200$	0.00011	0.00063	0.00329	0.0039	0.0067	0.0124

TABLE 1. Error E_1 on α given by $\alpha(t) = 10\sin(2\pi t)$ and error E_2 of prediction.

	$\sigma_\delta = 0.1$	$\sigma_\delta = 0.2$	$\sigma_\delta = 0.5$
$p = 50$	0.1141	0.2075	0.5034
$p = 100$	0.1011	0.2005	0.5005
$p = 200$	0.0999	0.1999	0.4999

TABLE 2. Estimated values of σ_δ according to the different values of σ_δ and the different values of p.

III.3.6. Proof of the results

Proof of theorem III.1. — First consider relation (III.23), and note that

$$\mathbb{E}\left(\widehat{\boldsymbol{\alpha}}_{FLS,X}\right) = \frac{1}{np^2}\left(\frac{1}{np^2}\mathbf{X}^\tau\mathbf{X} + \frac{\rho}{p}\mathbf{A}_m\right)^{-1}\mathbf{X}^\tau\mathbf{X}\boldsymbol{\alpha}.$$

		E_1			E_2		
		$\sigma_\delta = 0.1$	$\sigma_\delta = 0.2$	$\sigma_\delta = 0.5$	$\sigma_\delta = 0.1$	$\sigma_\delta = 0.2$	$\sigma_\delta = 0.5$
	$p = 50$	0.0508	0.0509	0.0510	0.0427	0.0426	0.0426
FLS, X	$p = 100$	0.0504	0.0504	0.0503	0.0422	0.0423	0.0424
	$p = 200$	0.0503	0.0502	0.0502	0.0414	0.0414	0.0416
	$p = 50$	0.0513	0.0526	0.0630	0.0439	0.0491	0.0830
$FTLS$	$p = 100$	0.0509	0.0522	0.0618	0.0434	0.0476	0.0762
	$p = 200$	0.0506	0.0517	0.0607	0.0429	0.0460	0.0735
	$p = 50$	0.0510	0.0525	0.0645	0.0435	0.0490	0.0851
FLS, $\widetilde{W}$	$p = 100$	0.0507	0.0520	0.0627	0.0429	0.0475	0.0790
	$p = 200$	0.0504	0.0516	0.0614	0.0422	0.0458	0.0763
	$p = 50$	0.0516	0.0530	0.0850	0.0447	0.0504	0.0960
FLS, W	$p = 100$	0.0512	0.0527	0.0822	0.0442	0.0496	0.0889
	$p = 200$	0.0508	0.0521	0.0799	0.0438	0.0488	0.0834

TABLE 3. Error E_1 on α given by $\alpha(t) = 10\sin^3\left(2\pi t^3\right)$ and error E_2 of prediction.

It follows that $\mathbb{E}\left(\widehat{\boldsymbol{\alpha}}_{FLS,X}\right)$ is solution of the minimization problem

$$\min_{\boldsymbol{a}\in\mathbb{R}^p}\left\{\frac{1}{n}\left\|\frac{1}{p}\mathbf{X}\boldsymbol{\alpha} - \frac{1}{p}\mathbf{X}\boldsymbol{a}\right\|^2 + \frac{\rho}{p}\boldsymbol{a}^\tau\mathbf{A}_m\boldsymbol{a}\right\}.$$

This implies, comparing this criterion above for $\boldsymbol{a} = \mathbb{E}\left(\widehat{\boldsymbol{\alpha}}_{FLS,X}\right)$ and $\boldsymbol{a} = \boldsymbol{\alpha}$,

$$\frac{1}{n}\left\|\frac{1}{p}\mathbf{X}\boldsymbol{\alpha} - \frac{1}{p}\mathbf{X}\mathbb{E}\left(\widehat{\boldsymbol{\alpha}}_{FLS,X}\right)\right\|^2 + \frac{\rho}{p}\mathbb{E}\left(\widehat{\boldsymbol{\alpha}}_{FLS,X}\right)^\tau\mathbf{A}_m\mathbb{E}\left(\widehat{\boldsymbol{\alpha}}_{FLS,X}\right) \leq \frac{\rho}{p}\boldsymbol{\alpha}^\tau\mathbf{A}_m\boldsymbol{\alpha}.$$

But definition of $\mathbf{A}_m$ and as well (III.17) lead to

$$\frac{1}{p}\boldsymbol{\alpha}^\tau\mathbf{A}_m\boldsymbol{\alpha} = \frac{1}{p}\boldsymbol{\alpha}^\tau\mathbf{P}_m\boldsymbol{\alpha} + \int_I s_{\boldsymbol{\alpha}}^{(m)}(t)^2dt \leq \frac{1}{p}\boldsymbol{\alpha}^\tau\mathbf{P}_m\boldsymbol{\alpha} + \int_I \alpha^{(m)}(t)^2dt,$$

and (III.23) is an immediate consequence, noticing that

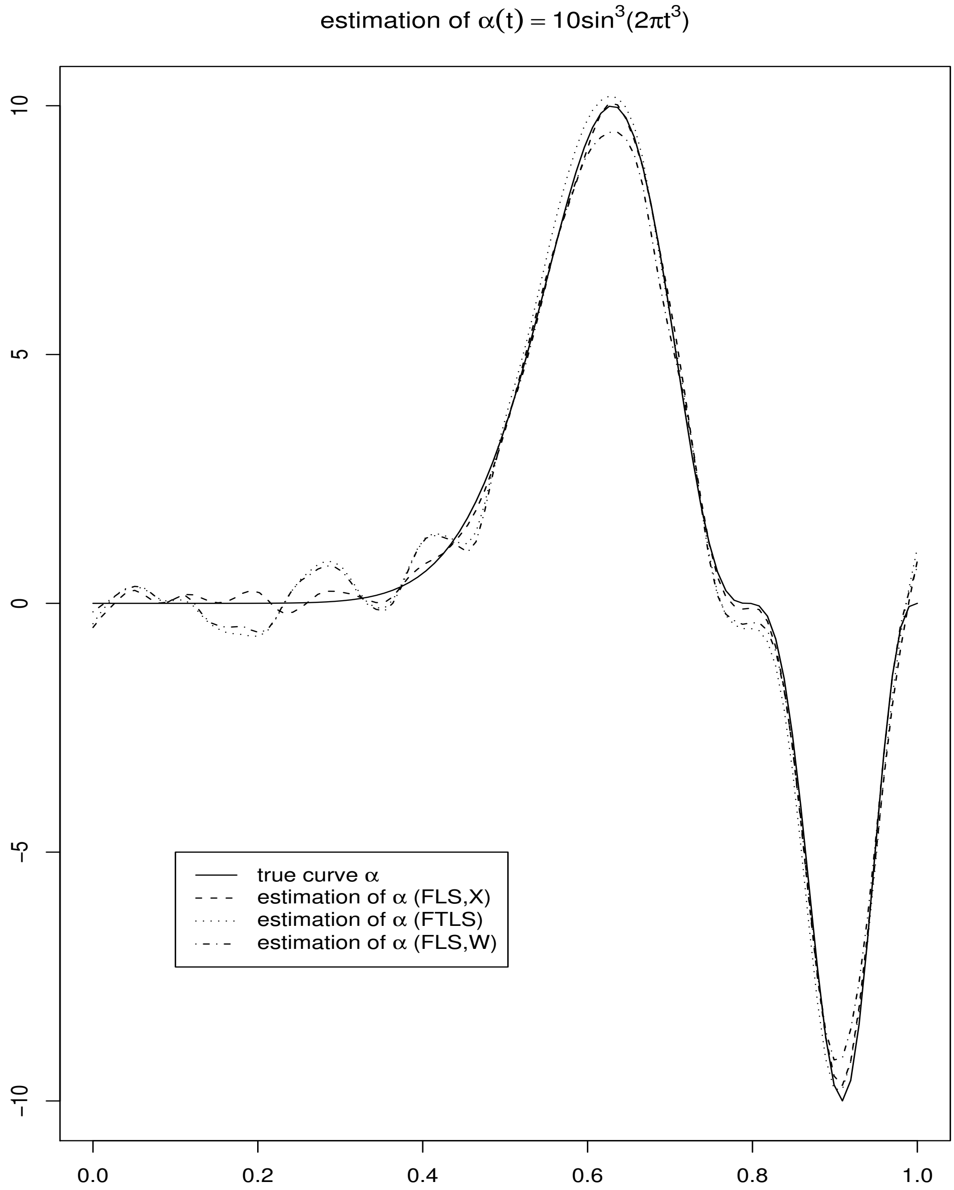


FIGURE 1. Estimation of α (solid line) with functional least squares using X (dashed line), functional least squares using W (dashed and dotted line) and functional total least squares (dotted line) in cases $\alpha(t) = 10\sin(2\pi t)$ and $\alpha(t) = 10\sin^3(2\pi t^3)$.

$$\|\mathbb{E}(\widehat{\boldsymbol{\alpha}}_{FLS,X}) - \boldsymbol{\alpha}\|^2_{\Gamma_{X,n,p}} = \frac{1}{np^2}\|\mathbf{X}\boldsymbol{\alpha} - \mathbf{X}\mathbb{E}(\widehat{\boldsymbol{\alpha}}_{FLS,X})\|^2.$$

Relation (III.24) follows from

$$\begin{aligned}
&\frac{1}{p}\mathbb{E}\left(\left[\widehat{\boldsymbol{\alpha}}^{\tau}_{FLS,X} - \mathbb{E}\left(\widehat{\boldsymbol{\alpha}}^{\tau}_{FLS,X}\right)\right]\left[\widehat{\boldsymbol{\alpha}}_{FLS,X} - \mathbb{E}(\widehat{\boldsymbol{\alpha}}_{FLS,X})\right]\right)\\
= \; & \frac{1}{p}\mathbb{E}\left(\frac{1}{n^2p^2}\boldsymbol{\epsilon}^{\tau}\mathbf{X}\left(\frac{1}{np^2}\mathbf{X}^{\tau}\mathbf{X} + \frac{\rho}{p}\mathbf{A}_m\right)^{-2}\mathbf{X}^{\tau}\boldsymbol{\epsilon}\right)\\
= \; & \frac{\sigma^2_{\epsilon}}{n}\mathrm{Tr}\left[\left(\frac{1}{np}\mathbf{X}^{\tau}\mathbf{X} + \rho\mathbf{A}_m\right)^{-2}\frac{1}{np}\mathbf{X}^{\tau}\mathbf{X}\right]\\
\leq \; & \frac{\sigma^2_{\epsilon}}{n}\mathrm{Tr}\left[\left(\frac{1}{np}\mathbf{X}^{\tau}\mathbf{X} + \rho\mathbf{A}_m\right)^{-1}\right] \leq \frac{\sigma^2_{\epsilon}}{n}\mathrm{Tr}\left[(\rho\mathbf{A}_m)^{-1}\right] \leq \frac{\sigma^2_{\epsilon}}{n\rho}D_1.
\end{aligned}$$

This completes the proof of the theorem III.1.

Proof of proposition III.6. — We have

$$\left(\left(\frac{\mathbf{W}}{p}, \mathbf{Y}\right) - \left(\frac{\boldsymbol{\delta}}{p}, \boldsymbol{\epsilon}\right)\right)\begin{pmatrix}\boldsymbol{\alpha}\\ -1\end{pmatrix} = 0, \qquad \text{(III.40)}$$

which allows us now to write the minimisation problem (III.28) as follows

$$\min_{\left(\left(\frac{\mathbf{W}}{p},\mathbf{Y}\right)-\left(\frac{\boldsymbol{\delta}}{p},\boldsymbol{\epsilon}\right)\right)\binom{\boldsymbol{a}}{-1}=0}\left\{\frac{1}{n}\left\|\left(\frac{\boldsymbol{\delta}}{\sqrt{p}}, \boldsymbol{\epsilon}\right)\right\|^2_F + \frac{\rho}{p}\boldsymbol{a}^{\tau}\mathbf{A}_m\boldsymbol{a}\right\},$$

where the notation $\|.\|_F$ stands for the usual Frobenius norm, more precisely $\|\mathbf{A}\|^2_F = \mathrm{Tr}\ (\mathbf{A}^{\tau}\mathbf{A})$ for every matrix $\mathbf{A}$. Then, we are led to consider the minimization problem

$$\text{(III.41)} \qquad \min_{\mathbf{C}\boldsymbol{x}=\mathbf{E}\boldsymbol{x}} \left\{ \frac{1}{n} \left\| \left(\frac{\boldsymbol{\delta}}{\sqrt{p}}, \boldsymbol{\epsilon} \right) \right\|_F^2 + \frac{\rho}{p} \boldsymbol{x}^\tau \mathbf{B}_m \boldsymbol{x} \right\},$$

with $\mathbf{C} = \left(\frac{\mathbf{W}}{p}, \mathbf{Y}\right)$, $\mathbf{E} = \left(\frac{\boldsymbol{\delta}}{p}, \boldsymbol{\epsilon}\right)$, $\boldsymbol{x} = \binom{\boldsymbol{a}}{-1}$ and $\mathbf{B}_m = \begin{pmatrix} \mathbf{A}_m & \mathbf{0} \\ \mathbf{0} & 0 \end{pmatrix}$. If we denote $\boldsymbol{\gamma}$ the $(p+1) \times (p+1)$ matrix defined by

$$\boldsymbol{\gamma} = \begin{pmatrix} \mathbf{diag}(1/\sqrt{p}, \ldots, 1/\sqrt{p}) & \mathbf{0} \\ \mathbf{0} & 0 \end{pmatrix},$$

we have

$$\begin{aligned} \frac{1}{n} \boldsymbol{x}^\tau \boldsymbol{\gamma}^\tau \left(\frac{\boldsymbol{\delta}}{\sqrt{p}}, \boldsymbol{\epsilon} \right)^\tau \left(\frac{\boldsymbol{\delta}}{\sqrt{p}}, \boldsymbol{\epsilon} \right) \boldsymbol{\gamma} \boldsymbol{x} &= \frac{1}{n} \boldsymbol{x}^\tau \mathbf{E}^\tau \mathbf{E} \boldsymbol{x} = \frac{1}{n} \boldsymbol{x}^\tau \mathbf{C}^\tau \mathbf{C} \boldsymbol{x} \\ &= \frac{1}{n} \boldsymbol{x}^\tau \boldsymbol{\gamma}^\tau \left(\frac{\mathbf{W}}{\sqrt{p}}, \mathbf{Y} \right)^\tau \left(\frac{\mathbf{W}}{\sqrt{p}}, \mathbf{Y} \right) \boldsymbol{\gamma} \boldsymbol{x}, \end{aligned}$$

and then we can see that the quantity

$$\begin{aligned} & \frac{1}{n} \boldsymbol{x}^\tau \boldsymbol{\gamma}^\tau \left(\frac{\mathbf{W}}{\sqrt{p}}, \mathbf{Y} \right)^\tau \left(\frac{\mathbf{W}}{\sqrt{p}}, \mathbf{Y} \right) \boldsymbol{\gamma} \boldsymbol{x} + \frac{\rho}{p} \boldsymbol{x}^\tau \mathbf{B}_m \boldsymbol{x} \\ = \; & \frac{1}{n} \boldsymbol{x}^\tau \boldsymbol{\gamma}^\tau \left(\frac{\mathbf{W}}{\sqrt{p}}, \mathbf{Y} \right)^\tau \left(\frac{\mathbf{W}}{\sqrt{p}}, \mathbf{Y} \right) \boldsymbol{\gamma} \boldsymbol{x} + \boldsymbol{x}^\tau \boldsymbol{\gamma}^\tau \left(\rho \mathbf{B}_m \right) \boldsymbol{\gamma} \boldsymbol{x} \end{aligned}$$

is minimized for $\boldsymbol{x}$ eigenvector of the matrix

$$\begin{aligned} & \frac{1}{n} \boldsymbol{\gamma}^\tau \left(\frac{\mathbf{W}}{\sqrt{p}}, \mathbf{Y} \right)^\tau \left(\frac{\mathbf{W}}{\sqrt{p}}, \mathbf{Y} \right) \boldsymbol{\gamma} + \boldsymbol{\gamma}^\tau \left(\rho \mathbf{B}_m \right) \boldsymbol{\gamma} \\ = \; & \frac{1}{n} \left(\frac{\mathbf{W}}{p}, \mathbf{Y} \right)^\tau \left(\frac{\mathbf{W}}{p}, \mathbf{Y} \right) + \frac{\rho}{p} \mathbf{B}_m, \end{aligned}$$

corresponding to the smallest non-zero eigenvalue, which is denoted σ_k^2. Using the definition of this eigenvalue, we deduce that

$$\left(\frac{1}{n}\left(\frac{\mathbf{W}}{p},\mathbf{Y}\right)^{\tau}\left(\frac{\mathbf{W}}{p},\mathbf{Y}\right)+\frac{\rho}{p}\mathbf{B}_m\right)\widehat{\boldsymbol{x}}=\sigma_k^2\widehat{\boldsymbol{x}}.$$

This gives, keeping the p first rows,

$$\widehat{\boldsymbol{\alpha}}=\frac{1}{np}\left(\frac{1}{np^2}\mathbf{W}^{\tau}\mathbf{W}+\frac{\rho}{p}\mathbf{A}_m-\sigma_k^2\mathbf{I}_p\right)^{-1}\mathbf{W}^{\tau}\mathbf{Y},$$

and the proof of the proposition III.6 is now complete.

Proof of proposition III.7. — We can write

$$\frac{1}{np^2}\mathbf{W}^{\tau}\mathbf{W}=\frac{1}{np^2}\mathbf{X}^{\tau}\mathbf{X}+\left(\frac{1}{np^2}\sum_{i=1}^{n}M_{irs}\right)_{r,s=1,\ldots,p}$$

where $M_{irs}=X_i(t_r)\delta_{is}+\delta_{ir}X_i(t_s)+\delta_{ir}\delta_{is}$. If we define $\mathbf{R}$ as the matrix such that

$$\left(\frac{1}{np^2}\sum_{i=1}^{n}M_{irs}\right)_{r,s=1,\ldots,p}=\frac{\sigma_\delta^2}{p^2}\mathbf{I}_p+\mathbf{R},$$

for every unitary vector $\boldsymbol{u}=(u_1,\ldots,u_p)^{\tau}\in\mathbb{R}^p$, we get $\mathbb{E}\left(\mathbf{R}\boldsymbol{u}\right)=\mathbf{0}$. Moreover, with hypotheses $(H.2)$ and $(H.3)$, we have, $\mathbb{E}$

$$\begin{aligned}
\mathbb{E}\left(\|\mathbf{R}\boldsymbol{u}\|^2\right) &= \boldsymbol{u}^\tau \mathbb{E}\left(\mathbf{R}^\tau \mathbf{R}\right)\boldsymbol{u} \\
&= \frac{1}{n^2 p^4}\sum_{r,s=1}^{p}\left(\sum_{i=1}^{n}\sum_{j=1}^{p}\mathbb{E}\left(M_{irj}M_{ijs}\right)u_r u_s\right) \\
&= \frac{1}{n^2 p^4}\sum_{r=1}^{p}\left(\sum_{i=1}^{n}\sum_{j=1}^{p}\mathbb{E}\left(M_{irj}M_{ijr}\right)u_r^2\right) \\
&= O\left(\frac{1}{np^2}\right),
\end{aligned}$$

hence we deduce that

$$\|\mathbf{R}\| = O_{\mathbb{P}}\left(\frac{1}{n^{1/2}p}\right).$$

Proof of theorem III.2. — We can write

$$\widehat{\boldsymbol{\alpha}}_{FTLS} - \widehat{\boldsymbol{\alpha}} = \left(\frac{1}{np^2}\mathbf{X}^\tau\mathbf{X} + \frac{\rho}{p}\mathbf{A}_m\right)^{-1}\mathbf{V} + \mathbf{S}\left(\frac{1}{np}\mathbf{W}^\tau\mathbf{Y}\right),$$

with

$$\mathbf{V} := \frac{1}{np}\mathbf{W}^\tau\mathbf{Y} - \frac{1}{np}\mathbf{X}^\tau\mathbf{Y} = \frac{1}{np}\boldsymbol{\delta}^\tau\mathbf{Y},$$

$$\mathbf{S} := \left(\frac{1}{np^2}\mathbf{X}^\tau\mathbf{X} + \frac{\rho}{p}\mathbf{A}_m + \mathbf{T}\right)^{-1} - \left(\frac{1}{np^2}\mathbf{X}^\tau\mathbf{X} + \frac{\rho}{p}\mathbf{A}_m\right)^{-1},$$

and

$$\mathbf{T} := \mathbf{R} - \frac{\widehat{\sigma}_\delta^2 - \sigma_\delta^2}{p^2}\mathbf{I}_p.$$

First noticing that

$$\mathbb{E}\left[\left(\frac{1}{np^2}\mathbf{X}^\tau\mathbf{X}+\frac{\rho}{p}\mathbf{A}_m\right)^{-1}\mathbf{V}\right]=0,$$

we also have, with hypotheses $(H.2)$ and $(H.3)$,

$$\begin{aligned}
&\mathbb{E}\left[\left\|\left(\frac{1}{np^2}\mathbf{X}^\tau\mathbf{X}+\frac{\rho}{p}\mathbf{A}_m\right)^{-1}\mathbf{V}\right\|^2_{\Gamma_{X,n,p}}\right]\\
=\;&\mathbb{E}\left[\frac{1}{n^2p}\mathbf{Y}^\tau\boldsymbol{\delta}\left(\frac{1}{np}\mathbf{X}^\tau\mathbf{X}+\rho\mathbf{A}_m\right)^{-1}\frac{1}{np}\mathbf{X}^\tau\mathbf{X}\left(\frac{1}{np}\mathbf{X}^\tau\mathbf{X}+\rho\mathbf{A}_m\right)^{-1}\boldsymbol{\delta}^\tau\mathbf{Y}\right]\\
\leq\;&\frac{D_8}{np}\mathrm{Tr}\left[\left(\frac{1}{np}\mathbf{X}^\tau\mathbf{X}+\rho\mathbf{A}_m\right)^{-1}\right],
\end{aligned}$$

where $D_8<+\infty$ does not depend on n and p. This allows us to get

$$\left\|\left(\frac{1}{np^2}\mathbf{X}^\tau\mathbf{X}+\frac{\rho}{p}\mathbf{A}_m\right)^{-1}\mathbf{V}\right\|_{\Gamma_{X,n,p}}=O_{\mathbb{P}}\left(\frac{1}{(np\rho)^{(1/2)}}\right). \tag{III.42}$$

The convergence result given in Gasser, Sroka and Jennen-Steinmetz (1986) implies that for the estimator $\widehat{\sigma}^2_\delta$ of σ^2_δ defined by (III.31) we have

$$\widehat{\sigma}^2_\delta=\sigma^2_\delta+O_{\mathbb{P}}\left(\frac{1}{n^{1/2}p}\right). \tag{III.43}$$

Then, using this and the result (III.30) of the proposition III.7, we can write

$$\left(\frac{1}{np^2}\mathbf{W}^\tau\mathbf{W}+\frac{\rho}{p}\mathbf{A}_m-\frac{\widehat{\sigma}^2_\delta}{p^2}\mathbf{I}_p\right)^{-1}=\left(\frac{1}{np^2}\mathbf{X}^\tau\mathbf{X}+\frac{\rho}{p}\mathbf{A}_m+\mathbf{R}-\frac{\widehat{\sigma}^2_\delta-\sigma^2_\delta}{p^2}\mathbf{I}_p\right)^{-1}.$$

Using the result (III.43) and the fact that the norm of $\mathbf{I}_p$ is 1, we deduce

$$\left\| \frac{\widehat{\sigma}_\delta^2 - \sigma_\delta^2}{p^2} \mathbf{I}_p \right\| = O_{\mathbb{P}}\left(\frac{1}{n^{1/2} p^3} \right).$$

If we recall the order of $\|\mathbf{R}\|$ given in proposition III.7, we finally obtain

$$\|\mathbf{T}\| = O_{\mathbb{P}}\left(\frac{1}{n^{1/2} p} \right).$$

Now, if we use the first inequality in Demmel (1992), we can write

$$\begin{aligned}
& \frac{1}{p^{1/2}} \left\| \mathbf{S}\left(\frac{1}{np} \mathbf{W}^\tau \mathbf{Y} \right) \right\| \\
= \;& \frac{1}{p^{1/2}} \left\| \left[\left(\frac{1}{np^2} \mathbf{X}^\tau \mathbf{X} + \frac{\rho}{p} \mathbf{A}_m + \mathbf{T} \right)^{-1} - \left(\frac{1}{np^2} \mathbf{X}^\tau \mathbf{X} + \frac{\rho}{p} \mathbf{A}_m \right)^{-1} \right] \right. \\
& \qquad\qquad\qquad\qquad \left. \times \left(\frac{1}{n^2 p^2} \mathbf{W}^\tau \mathbf{Y} \mathbf{Y}^\tau \mathbf{W} \right)^{1/2} \right\|_F \\
\leq \;& \frac{D_9}{p^{1/2}} \left\| \left(\frac{1}{np^2} \mathbf{X}^\tau \mathbf{X} + \frac{\rho}{p} \mathbf{A}_m \right)^{-1} \left(\frac{1}{np} \mathbf{W}^\tau \mathbf{Y} \right) \right\|^2 \|\mathbf{T}\| \left\| \frac{1}{np} \mathbf{W}^\tau \mathbf{Y} \right\|^{-1},
\end{aligned}$$

where $D_9 < +\infty$ does not depend on n and p. We notice that

$$\left\| \left(\frac{1}{np^2} \mathbf{X}^\tau \mathbf{X} + \frac{\rho}{p} \mathbf{A}_m \right)^{-1} \left(\frac{1}{np} \mathbf{W}^\tau \mathbf{Y} \right) \right\| = O_{\mathbb{P}}(p^{1/2}),$$

and $\|\mathbf{T}\| = O_{\mathbb{P}}\left(1/(n^{1/2}p)\right)$. Moreover, with the hypothesis on $p^{1/2} \left\| \frac{1}{np^2} \mathbf{X}^\tau \mathbf{X} \boldsymbol{\alpha} \right\|$, we also have

$$\left\| \frac{1}{np} \mathbf{W}^\tau \mathbf{Y} \right\|^{-1} = O_{\mathbb{P}}\left(p^{1/2}\right),$$

so we get

$$\left\| \mathbf{S}\left(\frac{1}{np}\mathbf{W}^\tau \mathbf{Y}\right)\right\|_{\Gamma_{X,n,p}} = O_{\mathbb{P}}\left(\frac{1}{n^{1/2}}\right). \tag{III.44}$$

Finally, we combine relations (III.42) and (III.44) to get the result of theorem III.2.

Proof of proposition III.8. — Let us expand the $MISE(\lambda)$,

$$MISE(\lambda) = \mathbb{E}\left(\widehat{\boldsymbol{\alpha}}(\lambda)^\tau \widehat{\boldsymbol{\alpha}}(\lambda)\right) - 2\boldsymbol{\alpha}^\tau \mathbb{E}\left(\widehat{\boldsymbol{\alpha}}(\lambda)\right) + \boldsymbol{\alpha}^\tau \boldsymbol{\alpha},$$

to deduce, using the matricial expression of $\widehat{\boldsymbol{\alpha}}(\lambda)$

$$\begin{aligned}\frac{\partial}{\partial\lambda} MISE(\lambda)_{|\lambda=0} = 2\mathbb{E}\Bigg[& \frac{1}{n^2p^2}\mathbf{Y}^\tau \mathbf{W}\left(\frac{1}{np^2}\mathbf{W}^\tau\mathbf{W} + \frac{\rho}{p}\mathbf{A}_m\right)^{-3}\mathbf{W}^\tau\mathbf{Y} \\ & -\frac{1}{np}\boldsymbol{\alpha}^\tau\left(\frac{1}{np^2}\mathbf{W}^\tau\mathbf{W} + \frac{\rho}{p}\mathbf{A}_m\right)^{-2}\mathbf{W}^\tau\mathbf{Y}\Bigg].\end{aligned} \tag{III.45}$$

Now, using the fact that $\mathbf{Y} = \frac{1}{p}\mathbf{W}\boldsymbol{\alpha} - \frac{1}{p}\boldsymbol{\delta}\boldsymbol{\alpha} + \boldsymbol{\epsilon}$

$$\begin{aligned}
& \frac{1}{np}\mathbf{Y}^\tau\mathbf{W}\left(\frac{1}{np^2}\mathbf{W}^\tau\mathbf{W}+\frac{\rho}{p}\mathbf{A}_m\right)^{-1}-\boldsymbol{\alpha}^\tau \\
= \;& \frac{1}{np}\mathbf{Y}^\tau\mathbf{W}\left(\frac{1}{np^2}\mathbf{W}^\tau\mathbf{W}+\frac{\rho}{p}\mathbf{A}_m\right)^{-1}-\frac{1}{np^2}\boldsymbol{\alpha}^\tau\mathbf{W}^\tau\mathbf{W}\left(\frac{1}{np^2}\mathbf{W}^\tau\mathbf{W}\right)^{-1} \\
= \;& \frac{1}{np}\left[\frac{1}{p}\boldsymbol{\alpha}^\tau\mathbf{W}^\tau\mathbf{W}\left(\frac{1}{np^2}\mathbf{W}^\tau\mathbf{W}+\frac{\rho}{p}\mathbf{A}_m\right)^{-1}\right. \\
& \left.-\frac{1}{p}\boldsymbol{\alpha}^\tau\boldsymbol{\delta}^\tau\mathbf{W}\left(\frac{1}{np^2}\mathbf{W}^\tau\mathbf{W}+\frac{\rho}{p}\mathbf{A}_m\right)^{-1}+\boldsymbol{\epsilon}^\tau\mathbf{W}\left(\frac{1}{np^2}\mathbf{W}^\tau\mathbf{W}+\frac{\rho}{p}\mathbf{A}_m\right)^{-1}\right] \\
& -\frac{1}{np}\left[\frac{1}{p}\boldsymbol{\alpha}^\tau\mathbf{W}^\tau\mathbf{W}\left(\frac{1}{np^2}\mathbf{W}^\tau\mathbf{W}\right)^{-1}\right].
\end{aligned}$$

Considering the quantity $\left(\frac{1}{np^2}\mathbf{W}^\tau\mathbf{W}+\frac{\rho}{p}\mathbf{A}_m\right)^{-1}-\left(\frac{1}{np^2}\mathbf{W}^\tau\mathbf{W}\right)^{-1}$, if we make an approximation at first order, we get

$$\begin{aligned}
& \left(\frac{1}{np^2}\mathbf{W}^\tau\mathbf{W}+\frac{\rho}{p}\mathbf{A}_m\right)^{-1}-\left(\frac{1}{np^2}\mathbf{W}^\tau\mathbf{W}\right)^{-1} \\
\approx \;& -\left(\frac{1}{np^2}\mathbf{W}^\tau\mathbf{W}\right)^{-1}\left(\frac{\rho}{p}\mathbf{A}_m\right)\left(\frac{1}{np^2}\mathbf{W}^\tau\mathbf{W}\right)^{-1},
\end{aligned}$$

what gives us, coming back to relation (III.45)

$$\begin{aligned}
\text{(III.46)} \quad & \frac{\partial}{\partial \lambda} MISE(\lambda)_{|\lambda=0} \\
\approx \quad & 2\mathbb{E}\left[-\frac{1}{n^2p^3} \boldsymbol{\alpha}^\tau \mathbf{W}^\tau \mathbf{W} \left(\left(\frac{1}{np^2} \mathbf{W}^\tau \mathbf{W} \right)^{-1} \frac{\rho}{p} \mathbf{A}_m \left(\frac{1}{np^2} \mathbf{W}^\tau \mathbf{W} \right)^{-1} \right) \right. \\
& \left. \qquad \times \left(\frac{1}{np^2} \mathbf{W}^\tau \mathbf{W} + \frac{\rho}{p} \mathbf{A}_m \right)^{-2} \mathbf{W}^\tau \mathbf{Y} \right] \\
& + 2\mathbb{E}\left[-\frac{1}{n^2p^3} \boldsymbol{\alpha}^\tau \boldsymbol{\delta}^\tau \mathbf{W} \left(\frac{1}{np^2} \mathbf{W}^\tau \mathbf{W} + \frac{\rho}{p} \mathbf{A}_m \right)^{-3} \mathbf{W}^\tau \mathbf{Y} \right] \\
& + 2\mathbb{E}\left[\frac{1}{n^2p^2} \boldsymbol{\epsilon}^\tau \mathbf{W} \left(\frac{1}{np^2} \mathbf{W}^\tau \mathbf{W} + \frac{\rho}{p} \mathbf{A}_m \right)^{-3} \mathbf{W}^\tau \mathbf{Y} \right].
\end{aligned}$$

Using the fact that δ and ϵ are both independent from W and Y, the last two terms in relation (III.46) are zero, and we obtain finally

$$\begin{aligned}
& \frac{\partial}{\partial \lambda} MISE(\lambda)_{|\lambda=0} \\
\approx \quad & 2\mathbb{E}\left[-\frac{1}{n^2p^4} \boldsymbol{\alpha}^\tau \mathbf{W}^\tau \mathbf{W} \left(\left(\frac{1}{np^2} \mathbf{W}^\tau \mathbf{W} \right)^{-1} \frac{\rho}{p} \mathbf{A}_m \left(\frac{1}{np^2} \mathbf{W}^\tau \mathbf{W} \right)^{-1} \right) \right. \\
& \left. \qquad \times \left(\frac{1}{np^2} \mathbf{W}^\tau \mathbf{W} + \frac{\rho}{p} \mathbf{A}_m \right)^{-2} \mathbf{W}^\tau \mathbf{W} \boldsymbol{\alpha} \right].
\end{aligned}$$

This last quantity is negative, what achieves the proof of proposition III.8.

III.4. RÉGRESSION SUR COMPOSANTES PRINCIPALES

Dans ce cadre d'une variable explicative bruitée, ce qui précédait visait à généraliser la méthode des moindres carrés orthogonaux au cas d'une variable explicative fonctionnelle. Dans cette partie, on va présenter une autre approche, qui concerne un travail en cours. Cette approche est finalement au départ plus directe que les moindres carrés orthogonaux (qui traite globalement les courbes bruitées). L'idée de départ est d'effectuer un lissage (par exemple un lissage à noyau) de chaque courbe bruitée, puis de produire une estimation du paramètre fonctionnel à l'aide par exemple d'une régression sur composantes principales. Rappelons que dans cette partie, le modèle considéré est donné par (2) et (8) et que le but est de donner une méthode d'estimation de α à l'aide des observations (W_1, Y_1), $\ldots$, (W_n, Y_n).

III.4.1. Procédure d'estimation

L'idée de cette procédure de lissage est donc la suivante. Lorsqu'on doit prendre en compte des courbes bruitées $W_1, \ldots, W_n$, on les lisse dans le but de construire une estimation des "vraies" courbes $X_1, \ldots, X_n$. On peut alors utiliser ces nouvelles courbes lissées $\widetilde{W}_1, \ldots, \widetilde{W}_n$ pour produire un estimateur de α au moyen d'une régression sur composantes principales fonctionnelle (voir les travaux de Cardot, Ferraty et Sarda, 1999, 2003). Cette régression sur composantes principales fonctionnelle sera néanmoins adaptée, reprenant une idée utilisée par Kneip et Utikal (2001) ainsi que Benko, Härdle et Kneip (2005). La procédure d'estimation comprend ainsi deux étapes qui vont être détaillées dans ce qui suit.

Étape 1 : lissage des courbes bruitées. — Cette première étape consiste à lisser les observations bruitées $W_1, \ldots, W_n$ à l'aide d'un estimateur à noyau de type Nadaraya-Watson, introduit à l'origine simultanément par Nadaraya (1964) et Watson (1964). Pour plus de détails sur cet estimateur à noyau tant d'un point de vue théorique qu'appliqué, on renvoie à Härdle (1991) ou Sarda et Vieu (2000). Plus précisément, on définit, pour $i = 1, \ldots, n$ et pour $t \in [0, 1]$,

$$\widetilde{W}_i(t) = \frac{\displaystyle\sum_{j=1}^{p} W_i(t_j) K\left(\frac{t - t_j}{h_i}\right)}{\displaystyle\sum_{j=1}^{p} K\left(\frac{t - t_j}{h_i}\right)}. \tag{III.47}$$

Dans cette expression (III.47), la fonction K, paire et d'intégrale égale à 1, est appelée noyau et le nombre réel $h_i > 0$ est appelé largeur de fenêtre. C'est ce paramètre qui permet de contrôler le lissage de la courbe estimée $\widetilde{W}_i$, alors que le choix du noyau est moins fondamental, si ce n'est que l'estimateur construit hérite des propriétés de régularité du noyau choisi (continuité, dérivabilité, ...). Le choix de la largeur de fenêtre étant très important, de nombreux travaux ont été réalisés pour permettre de déterminer ce paramètre en pratique, comme par exemple la validation croisée (voir Härdle ,1991).

Étape 2 : régression sur composantes principales fonctionnelle. — Cette étape consiste à construire une estimation de α au moyen d'une régression sur composantes principales fonctionnelle (voir Cardot, Ferraty et Sarda, 1999, 2003) en utilisant comme variable explicative la version lissée $\widetilde{W}_i$ de la courbe bruitée W_i. Cette méthode est basée sur la diagonalisation de l'opérateur de covariance empirique $\Gamma_{X,n}$ associé à X. On note $(\lambda_r)_{r\geq 1}$ la suite des valeurs propres de $\Gamma_{X,n}$ (rangées par ordre décroissant) et $(g_r)_{r\geq 1}$ la suite de fonctions propres associées, telles que pour tout $r \geq 1$,

$$\Gamma_{X,n} g_r = \lambda_r g_r.$$

En introduisant les coefficients

$$\eta_{ir} = \langle X_i, g_r \rangle, \tag{III.48}$$

pour tout $i = 1, \dots, n$ et pour tout $r \geq 1$, on a alors

$$\sum_{i=1}^{n} \eta_{ir} = 0,$$

pour tout $r \geq 1$ et

$$\sum_{i=1}^{n} \eta_{ir}\eta_{is} = \lambda_r \mathbb{1}_{[r=s]},$$

pour $r, s \geq 1$, avec $\mathbb{1}_{[r=s]} = 1$ si $r = s$ et 0 sinon. On utilise alors l'dée suivante provenant de Kneip et Utikal (2001), idée également reprise dans Benko, Härdle et Kneip (2005). Pour déterminer les fonctions propres g_r, $r \geq 1$, il n'est pas nécessaire de passer par l'opérateur $\Gamma_{X,n}$. En effet, on peut considérer à la place la matrice $\mathbf{M}$ de taille $n \times n$ définie par

$$M_{i_1 i_2} = \frac{1}{n}\langle X_{i_1}, X_{i_2} \rangle, \tag{III.49}$$

pour $i_1, i_2 = 1, \dots, n$. L'avantage d'utiliser cette matrice est que l'on estime les produits scalaires entre les courbes, donc des nombres réels. On obtiendra ainsi des vitesses de convergences avec un biais en h_i^2 et une variance en $1/n$. L'étude de ces vitesses sera détaillée dans la section suivante. En utilisant des résultats d'algèbre sur la diagonalisation de matrices (voir par exemple Good, 69), on montre que les valeurs propres non nulles de $\Gamma_{X,n}$ et les valeurs propres de $\mathbf{M}$ sont les mêmes. De plus, si on note $\boldsymbol{p}_r = (p_{1r}, \dots, p_{nr})^\tau$ le vecteur propre de $\mathbf{M}$ associé à la valeur propre λ_r, on a la relation

$$\eta_{ir} = \sqrt{\lambda_r} p_{ir}, \tag{III.50}$$

pour tout $i = 1, \ldots, n$ et pour tout $r \geq 1$ tel que $\lambda_r > 0$. On obtient alors les fonctions propres g_r, pour $r \geq 1$, par

$$g_r = \frac{1}{\sqrt{\lambda_r}} \sum_{i=1}^{n} p_{ir} X_i = \frac{\sum_{i=1}^{n} \eta_{ir} X_i}{\sum_{i=1}^{n} \eta_{ir}^2}. \tag{III.51}$$

Ainsi, en utilisant ce qui a été fait dans la première étape, on construit une estimation de la matrice $\mathbf{M}$ définie par (III.49), en utilisant les estimations $\widetilde{W}_i$ de X_i, pour $i = 1, \ldots, n$. L'estimateur le plus naturel $\widehat{\mathbf{M}}$ de $\mathbf{M}$ semble être la matrice de taille $n \times n$ et de terme général $\widehat{M}_{i_1 i_2} = \frac{1}{n}\langle \widetilde{W}_{i_1}, \widetilde{W}_{i_2} \rangle$, pour $i_1, i_2 = 1, \ldots, n$. Cependant, comme cela a été souligné, on cherche à estimer ici des produits scalaires entre des courbes et non les courbes elles-mêmes. Si on regarde l'estimateur $\langle \widetilde{W}_{i_1}, \widetilde{W}_{i_2} \rangle$ lorsque $i_1 = i_2 = i$, on remarque que ce terme s'écrit

$$\begin{aligned}
\langle \widetilde{W}_i, \widetilde{W}_i \rangle &= \sum_{j_1=1}^{p} \sum_{j_2=1}^{p} W_i(t_{j_1}) W_i(t_{j_2}) \int_I \frac{K\left(\frac{t-t_{j_1}}{h_i}\right) K\left(\frac{t-t_{j_2}}{h_i}\right)}{\left[\sum_{j_3=1}^{p} K\left(\frac{t-t_{j_3}}{h_i}\right)\right]^2} \, dt \\
&= \sum_{j_1=1}^{p} \sum_{\substack{j_2=1 \\ j_2 \neq j_1}}^{p} W_i(t_{j_1}) W_i(t_{j_2}) \int_I \frac{K\left(\frac{t-t_{j_1}}{h_i}\right) K\left(\frac{t-t_{j_2}}{h_i}\right)}{\left[\sum_{j_3=1}^{p} K\left(\frac{t-t_{j_3}}{h_i}\right)\right]^2} \, dt \\
&\quad + \sum_{j_1=1}^{p} W_i(t_{j_1})^2 \int_I \frac{K\left(\frac{t-t_{j_1}}{h_i}\right)^2}{\left[\sum_{j_3=1}^{p} K\left(\frac{t-t_{j_3}}{h_i}\right)\right]^2} \, dt.
\end{aligned}$$

Ainsi, le terme

$$\sum_{j_1=1}^{p} W_i(t_{j_1})^2 \int_I \frac{K\left(\frac{t-t_{j_1}}{h_i}\right)^2}{\left[\sum_{j_3=1}^{p} K\left(\frac{t-t_{j_3}}{h_i}\right)\right]^2} \, dt$$

produit un biais dans l'estimation de M_{ii}. L'idée est donc de le supprimer dans l'estimateur. Cette idée a été utilisée par Kneip et Utikal (2001) dans le cadre de l'estimation de densités puis par Benko, Härdle et Kneip (2005) concernant l'analyse en composantes principales fonctionnelle. On se rend compte cependant que cette idée avait déjà été proposée auparavant par Hall et Marron (1987) et Jones et Sheater (1991) dans le cadre de l'intégration de densités. Concernant l'estimation de l'intégrale du carré de fonctions de régression (qui nous intéressera plus particulièrement ici), on peut citer les travaux de Benhenni et Cambanis (1992), Ruppert, Sheater et Wand (1993), ainsi que Huang et Fan (1999). Ceci permet de gagner au niveau du biais de l'estimation de la matrice $\mathbf{M}$. Dans la suite, on considère donc l'estimateur, pour $i_1, i_2 = 1, \ldots, n$,

$$\widehat{M}_{i_1 i_2} = \begin{cases} \dfrac{1}{n} \displaystyle\sum_{j_1=1}^{p} \sum_{j_2=1}^{p} W_{i_1}(t_{j_1}) W_{i_2}(t_{j_2}) \int_I \frac{K\left(\frac{t-t_{j_1}}{h_{i_1}}\right) K\left(\frac{t-t_{j_2}}{h_{i_2}}\right)}{\left[\sum_{j_3=1}^{p} K\left(\frac{t-t_{j_3}}{h_{i_1}}\right)\right] \left[\sum_{j_4=1}^{p} K\left(\frac{t-t_{j_4}}{h_{i_2}}\right)\right]} \, dt, \\ \hfill \text{si } i_1 \neq i_2, \\ \dfrac{1}{n} \displaystyle\sum_{j_1=1}^{p} \sum_{\substack{j_2=1 \\ j_2 \neq j_1}}^{p} W_{i_1}(t_{j_1}) W_{i_2}(t_{j_2}) \int_I \frac{K\left(\frac{t-t_{j_1}}{h_{i_1}}\right) K\left(\frac{t-t_{j_2}}{h_{i_2}}\right)}{\left[\sum_{j_3=1}^{p} K\left(\frac{t-t_{j_3}}{h_{i_1}}\right)\right]^2} \, dt, \quad \text{si } i_1 = i_2. \end{cases}$$

De plus, en ayant à l'esprit que l'on estime des produits scalaires plutôt que des courbes, il semble plus approprié de choisir une même largeur de fenêtre $h_{i_1 i_2}$ pour l'estimation du produit scalaire entre la paire de courbes $\{X_{i_1}, X_{i_2}\}$, pour $i_1, i_2 = 1, \ldots, n$. Une légère modification de l'estimateur ci-dessus nous donne alors

(III.52)
$$\widehat{M}_{i_1 i_2} = \begin{cases} \dfrac{1}{n} \displaystyle\sum_{j_1=1}^{p} \sum_{j_2=1}^{p} W_{i_1}(t_{j_1}) W_{i_2}(t_{j_2}) \int_I \frac{K\left(\frac{t-t_{j_1}}{h_{i_1 i_2}}\right) K\left(\frac{t-t_{j_2}}{h_{i_1 i_2}}\right)}{\left[\sum_{j_3=1}^{p} K\left(\frac{t-t_{j_3}}{h_{i_1 i_2}}\right)\right]^2} \, dt, \quad \text{si } i_1 \neq i_2, \\ \dfrac{1}{n} \displaystyle\sum_{j_1=1}^{p} \sum_{\substack{j_2=1 \\ j_2 \neq j_1}}^{p} W_{i_1}(t_{j_1}) W_{i_2}(t_{j_2}) \int_I \frac{K\left(\frac{t-t_{j_1}}{h_{i_1 i_2}}\right) K\left(\frac{t-t_{j_2}}{h_{i_1 i_2}}\right)}{\left[\sum_{j_3=1}^{p} K\left(\frac{t-t_{j_3}}{h_{i_1 i_2}}\right)\right]^2} \, dt, \quad \text{si } i_1 = i_2. \end{cases}$$

Avec cette estimation $\widehat{\mathbf{M}}$ de $\mathbf{M}$, on calcule les valeurs propres $\widehat{\lambda}_r$ et les vecteurs propres $\widehat{\boldsymbol{p}}_r$ correspondants, pour $r = 1, \ldots, n$. On en déduit les estimations $\widehat{\eta}_{ir}$ et $\widehat{g}_r$ de η_{ir} et g_r en utilisant les relations (III.50) et (III.51). Finalement, on construit l'estimateur de α donné par l'approximation d'ordre $L \geq 1$, c'est-à-dire en utilisant les L premières composantes principales (voir Cardot, Ferraty et Sarda, 1999, 2003). Notre estimateur est ainsi donné par

$$\widehat{\alpha}_L = \frac{1}{n} \sum_{r=1}^{L} \sum_{i=1}^{n} \frac{Y_i}{\widehat{\lambda}_r} \langle \widetilde{W}_i, \widehat{g}_r \rangle \widehat{g}_r. \tag{III.53}$$

III.4.2. Intégrale du carré de la régression

Cette partie est relativement indépendante de ce qui précède. Son but est d'établir des résultats concernant le biais et la variance d'un estimateur à noyau de l'intégrale du carré d'une fonction de régression. Ce résultat sera ensuite utilisé pour en déduire le biais et la variance concernant l'estimation de la matrice $\mathbf{M}$ par (III.52). Le cas qui nous intéresse est celui d'un modèle à effets fixes ($t_1, \ldots, t_p$ sont fixés et forment une subdivision régulière de $[0, 1]$). On considère donc ici le modèle

$$Y_i = r(x_i) + \epsilon_i,$$

pour tout $i = 1, \ldots, n$, avec $\mathbb{E}(\epsilon_i) = 0$, $\mathbb{E}(\epsilon_i^2) = \sigma^2$, et $r \in L^2([0, 1])$. On suppose que $x_1, \ldots, x_n$ forment une subdivision régulière de $[0, 1]$. L'estimateur à noyau de r (de Nadaraya-Watson) s'écrit

$$\widehat{r}(x) = \frac{\sum_{i=1}^n Y_i K\left(\frac{x_i - x}{h}\right)}{\sum_{i=1}^n K\left(\frac{x_i - x}{h}\right)}.$$

Le but final est d'estimer l'intégrale de r^2 sur $[0, 1]$ (voir cas 3 ci-après). On va cependant profiter de cette étude pour donner d'autres résultats pouvant être utiles (cas 1 et cas 2). Les preuves de ces résultats sont relativement techniques et seront données dans l'annexe (voir partie V.2.).

Cas1. — On souhaite dans ce premier cas estimer la quantité

$$\theta = \int_0^1 r(x)\alpha(x)dx,$$

où α est une fonction donnée de $L^2([0,1])$. On considère pour cela l'estimateur

$$\widehat{\theta} = \sum_{i=1}^n Y_i \int_0^1 \frac{K\left(\frac{x_i-x}{h}\right)\alpha(x)}{\sum_{i=1}^n K\left(\frac{x_i-x}{h}\right)}dx.$$

On suppose que K est un noyau d'ordre 2 à support sur $[-1,1]$ et on notera $\mu_2(K) = \int_{-1}^1 s^2K(s)ds$ et $R(K) = \int_{-1}^1 K(s)^2ds$. On supposera de plus que n et h sont tels que $nh \longrightarrow +\infty$. Le biais et la variance de θ sont donnés dans les propositions III.9 et III.10.

Proposition III.9. — *On a*

$$\mathbb{E}(\widehat{\theta}) - \theta = \frac{\mu_2(K)}{2}\left(\int_0^1 r''(x)\alpha(x)dx\right)h^2 + o(h^2). \tag{III.54}$$

Proposition III.10. — *On a*

$$\mathbb{V}\left(\widehat{\theta}\right) = 2\sigma^2\left(\int_0^1 \alpha(x)^2dx\right)\left(\int_0^2 \Psi(z)dz\right)\frac{1}{n} + o\left(\frac{1}{n}\right), \tag{III.55}$$

avec

$$\Psi(z) = \int_{-1+z}^1 K(s)K(s-z)dz.$$

Cas2. — On suppose maintenant qu'on dispose d'autres observations $Z_1, \ldots, Z_n$ telles que, pour tout $i = 1, \ldots, n$,

$$Z_i = s(x_i) + \delta_i,$$

avec $\mathbb{E}(\delta_i) = 0$, $\mathbb{E}(\delta_i^2) = \tau^2$, Y_i indépendant de Z_i, ϵ_i indépendant de δ_i, et $s \in L^2([0,1])$. On souhaite cette fois estimer la quantité

$$\theta = \int_0^1 r(x)s(x)dx.$$

On considère pour cela l'estimateur

$$\widehat{\theta} = \sum_{i=1}^n \sum_{j=1}^n Y_i Z_j \int_0^1 \frac{K\left(\frac{x_i - x}{h}\right) K\left(\frac{x_j - x}{h}\right)}{\sum_{i=1}^n \sum_{j=1}^n K\left(\frac{x_i - x}{h}\right) K\left(\frac{x_j - x}{h}\right)} dx.$$

Les hypothèses et notations sur K sont les mêmes qu'au cas précédent. Le biais et la variance de θ sont donnés dans les propositions III.11 et III.12.

Proposition III.11. — *On a*

$$\text{(III.56)} \qquad \mathbb{E}(\widehat{\theta}) - \theta = \frac{\mu_2(K)}{2} \left(\int_0^1 \left[r(x)s''(x) + r''(x)s(x) \right] dx \right) h^2 + o(h^2).$$

Proposition III.12. — *On a*

$$\text{(III.57)} \qquad \mathbb{V}\left(\widehat{\theta}\right) = 2\left[\tau^2 \left(\int_0^1 r(x)^2 dx \right) + \sigma^2 \left(\int_0^1 s(x)^2 dx \right) \right] \times \left(\int_0^2 \Psi(z)dz \right) \frac{1}{n} + o\left(\frac{1}{n}\right).$$

Cas3. — Avec les mêmes notations qu'au cas 1, on considère l'estimation de

$$\theta = \int_0^1 r(x)^2 dx.$$

On introduit pour cela l'estimateur

$$\begin{aligned}
\widehat{\theta} &= \int_0^1 \frac{\sum_{i=1}^n \sum_{\substack{j=1\\ j\neq i}}^n Y_i Y_j K\left(\frac{x_i-x}{h}\right) K\left(\frac{x_j-x}{h}\right)}{\left[\sum_{i=1}^n K\left(\frac{x_i-x}{h}\right)\right]^2} dx \\
&= \sum_{i=1}^n \sum_{\substack{j=1\\ j\neq i}}^n Y_i Y_j \int_0^1 \frac{K\left(\frac{x_i-x}{h}\right) K\left(\frac{x_j-x}{h}\right)}{\left[\sum_{i=1}^n K\left(\frac{x_i-x}{h}\right)\right]^2} dx.
\end{aligned}$$

Le biais et la varaince de θ sont donnés dans les propositions III.13 et III.14.

Proposition III.13. — *On a*

$$\begin{aligned}
\text{(III.58)} \qquad \mathbb{E}(\widehat{\theta}) - \theta &= \mu_2(K)\left(\int_0^1 r(x) r''(x) dx\right) h^2 \\
&\quad - R(K)\left(\int_0^1 r(x)^2 dx\right)\frac{1}{nh} + o\left(h^2 + \frac{1}{nh}\right).
\end{aligned}$$

Proposition III.14. — *On a*

$$\text{(III.59)} \qquad \mathbb{V}(\widehat{\theta}) = 8\sigma^2 \left(\int_0^1 r(x)^2 dx\right)\left(\int_0^2 \Psi(z) dz\right)\frac{1}{n} + o\left(\frac{1}{n}\right).$$

III.4.3. Résultats asymptotiques

Les résultats que l'on va établir ici sont directement inspirés des travaux de Kneip et Utikal (2001) et Benko, Härdle et Kneip (2005). Les preuves des résultats qui suivent sont elles aussi directement inspirées de ces travaux, elles seront données dans l'annexe de la thèse (voir partie V.3.).

Proposition III.15. — *Pour tous $i_1, i_2 = 1, \ldots, n$, si on prend $h_{i_1 i_2}$ de la forme $p^{-\zeta}$ avec $\zeta \in [1/4, 1/2[$, on a*

$$\widehat{M}_{i_1 i_2} - M_{i_1 i_2} = O_{\mathbb{P}}\left(\frac{1}{np^{1/2}}\right).$$

Proposition III.16. — *Si on note $\|.\|$ la norme matricielle euclidienne usuelle, on a*

$$\left\|\widehat{\mathbf{M}} - \mathbf{M}\right\|^2 = O_{\mathbb{P}}\left(\frac{1}{p}\right).$$

Les deux propositions suivantes donnent le comportement des valeurs propres et des vecteurs propres de $\widehat{\mathbf{M}}$ par rapport à ceux de $\mathbf{M}$.

Proposition III.17. — *On suppose que les valeurs propres de $\mathbf{M}$ sont telles que, pour tout $r = 1, \ldots, L$, il existe des constantes $0 < C_{1r} < +\infty$ et $0 < C_{2r} \leq C_{3r} < +\infty$ vérifiant*

$$\min_{s=1,\ldots,n, s\neq r} |\lambda_r - \lambda_s| \geq C_{1r},$$

et

$$C_{2r} \leq \lambda_r \leq C_{3r}.$$

Alors, on a, pour tout $r = 1, \ldots, L$,

$$\widehat{\lambda}_r - \lambda_r = O_{\mathbb{P}}\left(\frac{1}{n^{1/2}p^{1/2}} + \frac{1}{p}\right).$$

Proposition III.18. — *Sous les mêmes hypothèses qu'à la proposition III.17, on a*

$$\|\widehat{\boldsymbol{p}}_r - \boldsymbol{p}_r\| = O_{\mathbb{P}}\left(\frac{1}{p^{1/2}}\right).$$

III.4.4. Perspectives

Ce début de travail donne des résultats assez encourageants. D'un point de vue théorique, il faut maintenant prouver un résultat de convergence concernant l'estimateur $\widehat{\alpha}_L$ de α défini par (III.53), ce qui devrait être raisonnablement envisageable au vu des résultats de convregence précédents concernant les valeurs propres et les vecteurs propres de l'estimateur de la matrice $\mathbf{M}$. D'un point de vue un peu plus appliqué, il reste à faire tout un travail sur le choix du nombre L de composantes principales. Une méthode par validation croisée peut à priori être envisagée, alors que Kneip et Utikal (2001) proposent une procédure de test dans leur contexte d'estimation de densités. Enfin, il paraît assez intéressant de comparer cette procédure d'estimation avec celle des moindres carrés orthogonaux.

PARTIE IV

APPLICATION À LA PRÉVISION DE PICS DE POLLUTION

IV.1. PRÉVISION PAR LES QUANTILES CONDITIONNELS

Dans ce premier chapitre, on s'intéresse à l'estimation spline de quantiles conditionnels qui a été présenté dans la partie I de la thèse. Comme cela a été souligné, le problème de minimisation (I.7) ne possède pas de solution explicite. On adopte ici une stratégie déjà utilisée par Lejeune et Sarda (1988). Elle est basée sur un algorithme itératif (appelé algorithme des moindres carrés itérés pondérés) qui consiste, à chaque étape de l'algorithme, à résoudre un problème de moindres carrés pondérés (voir Ruppert et Caroll, 1988). Cette étude pratique a fait l'objet d'un chapitre de e-book dont le but était de faire de la prévision de pics de pollution. Ce chapitre de e-book, ainsi que la présentation de données de pollution étudiées sont donnés dans les chapitres IV.3 et IV.4 de la thèse.

IV.1.1. Algorithme d'estimation

Considérons tout d'abord la quantité

$$\omega_i(\alpha) = 2\alpha \mathbb{1}_{\{Y_i - \langle \mathbf{B}_{k,q}^\tau \theta, X_i\rangle \geq 0\}} + 2(1-\alpha)\mathbb{1}_{\{Y_i - \langle \mathbf{B}_{k,q}^\tau \theta, X_i\rangle < 0\}},$$

où $\mathbb{1}$ désigne la fonction indicatrice d'un ensemble. En utilisant cette notation, on peut écrire le problème de minimisation (I.7) sous la forme

$$\min_{\boldsymbol{\theta}\in\mathbb{R}^{k+q}} \left\{ \frac{1}{n}\sum_{i=1}^{n} \omega_i(\alpha)\left|Y_i - \langle \mathbf{B}_{k,q}^{\tau}\boldsymbol{\theta}, X_i\rangle\right| + \rho \left\|(\mathbf{B}_{k,q}^{\tau}\boldsymbol{\theta})^{(m)}\right\|_{L^2}^2 \right\}. \tag{IV.1}$$

Le principe de l'algorithme des moindres carrés itérés pondérés consiste alors à remplacer la valeur absolue par un terme quadratique pondéré. On obtient ainsi, à chaque étape de l'algorithme, une expression explicite de la solution du problème de minimisation. On décrit l'algorithme ci-dessous.

- Initialisation : on détermine $\widehat{\boldsymbol{\theta}}^{(1)}$ solution du problème de minimisation

  $$\min_{\boldsymbol{\theta}\in\mathbb{R}^{k+q}} \left\{ \frac{1}{n}\sum_{i=1}^{n} \left(Y_i - \langle \mathbf{B}_{k,q}^{\tau}\boldsymbol{\theta}, X_i\rangle\right)^2 + \rho \left\|(\mathbf{B}_{k,q}^{\tau}\boldsymbol{\theta})^{(m)}\right\|_{L^2}^2 \right\},$$

  dont l'expression explicite est donnée par

  $$\widehat{\boldsymbol{\theta}}^{(1)} = \frac{1}{n}\left(\frac{1}{n}\mathbf{D}_X^{\tau}\mathbf{D}_X + \rho\mathbf{G}_k\right)^{-1}\mathbf{D}_X^{\tau}\mathbf{Y},$$

  où $\mathbf{D}_X$ est la matrice $n\times(k+q)$ de terme général $\langle B_j, X_i\rangle$ pour $i=1,\ldots,n$ et $j=1,\ldots,(k+q)$, $\mathbf{G}_k$ est la matrice $(k+q)\times(k+q)$ de terme général $\langle B_j^{(m)}, B_l^{(m)}\rangle$ pour $j,l=1,\ldots,(k+q)$ et $\mathbf{Y}=(Y_1,\ldots,Y_n)^{\tau}$.
- Étape $r+1$: connaissant $\widehat{\boldsymbol{\theta}}^{(r)}$, on détermine $\widehat{\boldsymbol{\theta}}^{(r+1)}$ solution du problème de minimisation

  $$\min_{\boldsymbol{\theta}\in\mathbb{R}^{k+q}} \left\{ \frac{1}{n}\sum_{i=1}^{n} \frac{\omega_i^{(r)}(\alpha)\left(Y_i - \langle \mathbf{B}_{k,q}^{\tau}\boldsymbol{\theta}, X_i\rangle\right)^2}{\left[\left(Y_i - \langle \mathbf{B}_{k,q}^{\tau}\boldsymbol{\theta}, X_i\rangle\right)^2 + \eta^2\right]^{1/2}} + \rho \left\|(\mathbf{B}_{k,q}^{\tau}\boldsymbol{\theta})^{(m)}\right\|_{L^2}^2 \right\},$$

  où η^2 est une constante strictement positive que l'on se fixe pour éviter un dénominateur nul, et $\omega_i^{(r)}(\alpha)$ est $\omega_i(\alpha)$ à l'étape r de l'algorithme, à savoir

  $$\omega_i^{(r)}(\alpha) = 2\alpha\mathbb{1}_{\{Y_i - \langle \mathbf{B}_{k,q}^{\tau}\widehat{\boldsymbol{\theta}}^{(r)}, X_i\rangle \geq 0\}} + 2(1-\alpha)\mathbb{1}_{\{Y_i - \langle \mathbf{B}_{k,q}^{\tau}\widehat{\boldsymbol{\theta}}^{(r)}, X_i\rangle < 0\}}.$$

  En définissant la matrice $\boldsymbol{\Omega}^{(r)}$ de taille $n\times n$, diagonale dont les éléments diagonaux sont donnés, pour tout $i=1,\ldots,n$, par

$$\frac{\omega_i^{(r)}(\alpha)}{[(Y_i - \langle \mathbf{B}_{k,q}^\tau \boldsymbol{\theta}, X_i\rangle)^2 + \eta^2]^{1/2}},$$

on obtient la solution du problème de minimisation de l'étape $r+1$ donnée par

$$\widehat{\boldsymbol{\theta}}^{(r+1)} = \frac{1}{n}\left(\frac{1}{n}\mathbf{D}_X^\tau \boldsymbol{\Omega}^{(r)}\mathbf{D}_X + \rho \mathbf{G}_k\right)^{-1} \mathbf{D}_X^\tau \boldsymbol{\Omega}^{(r)}\mathbf{Y}.$$

- Critère d'arrêt : on décide d'arrêter l'algorithme lorsque

$$\left|C^{(r+1)} - C^{(r)}\right| < err,$$

où la quantité err est fixée et où $C^{(r)}$ est défini par

$$C^{(r)} = \frac{1}{n}\sum_{i=1}^n l_\alpha(Y_i - \langle \mathbf{B}_{k,q}^\tau \widehat{\boldsymbol{\theta}}^{(r)}, X_i\rangle) + \rho \left\|(\mathbf{B}_{k,q}^\tau \widehat{\boldsymbol{\theta}}^{(r)})^{(m)}\right\|_{L^2}^2 .$$

IV.1.2. Choix des paramètres

La procédure d'estimation qui vient d'être présentée dépend de beaucoup de paramètres : le nombre de nœuds k, le degré des fonctions splines q, l'ordre de dérivation dans la pénalisation m, et le paramètre de lissage ρ. Outre le nombre de nœuds k, le choix du paramètre ρ est crucial pour donner une bonne estimation de $\boldsymbol{\theta}$ (voir à ce sujet Marx et Eilers, 1999). On fixe donc $m = 2$, $q = 3$ (splines cubiques) et $k = 8$. Le paramètre ρ est quant à lui fixé par validation croisée généralisée (voir Wahba, 1990) à chaque étape de l'algorithme des moindres carrés itérés pondérés. Avant de donner ce critère, on revient à la validation croisée ordinaire. À l'étape r de l'algorithme des moindres carrés itérés pondérés, on peut définir le critère de validation croisée par

$$CV^{(r)}(\rho) = \frac{1}{n}\sum_{i=1}^n \left(Y_i - \langle \mathbf{B}_{k,q}^\tau \widehat{\boldsymbol{\theta}}^{(r),[-i]}, X_i\rangle\right)^2,$$

où $\widehat{\boldsymbol{\theta}}^{(r),[-i]}$ désigne l'estimation de $\boldsymbol{\theta}$ à l'étape r de l'algorithme, en utilisant toutes les données sauf la $i^{\text{ème}}$, (X_i, Y_i). On choisit alors pour valeur de ρ celle qui minimise ce critère $CV(\rho)$. On peut alors montrer (voir Wahba, 1990) que ce critère s'écrit sous la forme

$$CV^{(r)}(\rho) = \frac{1}{n}\sum_{i=1}^{n} \frac{\left(Y_i - \langle \mathbf{B}_{k,q}^{\tau}\widehat{\boldsymbol{\theta}}^{(r)}, X_i\rangle\right)^2}{\left(1 - h_{ii}^{(r)}\right)^2},$$

où $h_{ii}^{(r)}$ est le $i^{\text{ème}}$ élément diagonal de la "hat matrix" $\mathbf{H}^{(r)}$ (à l'étape r de l'algorithme) définie par

$$\mathbf{H}^{(r)} = \mathbf{D}_X \left(\mathbf{D}_X^{\tau}\boldsymbol{\Omega}^{(r)}\mathbf{D}_X + \rho\mathbf{G}_k\right)^{-1} \mathbf{D}_X^{\tau}\boldsymbol{\Omega}^{(r)}.$$

Le critère de validation croisée généralisée (à l'étape r de l'algorithme) est alors obtenu en remplaçant dans le critère de validation croisée $h_{ii}^{(r)}$ par $\frac{1}{n}\sum_{i=1}^{n} h_{ii}^{(r)} = \frac{1}{n}\mathrm{Tr}\left(\mathbf{H}^{(r)}\right)$. Ce critère s'écrit donc, en notant $\widehat{\mathbf{Y}}^{(r)} = \mathbf{H}^{(r)}\mathbf{Y}$,

$$GCV^{(r)}(\rho) = \frac{\frac{1}{n}\left(\mathbf{Y} - \widehat{\mathbf{Y}}^{(r)}\right)^{\tau} \boldsymbol{\Omega}^{(r)} \left(\mathbf{Y} - \widehat{\mathbf{Y}}^{(r)}\right)}{\left(1 - \frac{1}{n}\mathrm{Tr}\left(\mathbf{H}^{(r)}\right)\right)^2},$$

et on choisit alors pour valeur de ρ celle qui minimise ce critère $GCV(\rho)$. Notons que le calcul numérique de ce critère est plus rapide que celui du critère de validation croisée ordinaire, et donne en pratique de bons résultats. En utilisant les propriétés de la trace, on peut gagner aussi en calcul sur les tailles de matrices en écrivant

$$\mathrm{Tr}\left(\mathbf{H}^{(r)}\right) = \mathrm{Tr}\left(\mathbf{D}_X^{\tau}\boldsymbol{\Omega}^{(r)}\mathbf{D}_X \left(\mathbf{D}_X^{\tau}\boldsymbol{\Omega}^{(r)}\mathbf{D}_X + \rho\mathbf{G}_k\right)^{-1}\right).$$

IV.1.3. Modèle avec plusieurs variables explicatives

On souhaite étendre ici le modèle (2) au cas où on dispose de plusieurs variables explicatives. Notons dans la suite ces v variables (centrées) $X_i^1, \ldots, X_i^v$

appartenant à $L^2([0,1])$ pour tout $i = 1, \ldots, n$. On se restreint à un modèle additif s'écrivant, pour tout $i = 1, \ldots, n$

$$Y_i = \int_0^1 \Psi_\alpha^1(t) X_i^1(t) dt + \ldots + \int_0^1 \Psi_\alpha^v(t) X_i^v(t) dt + \epsilon_i,$$

avec

$$\mathbb{P}\left(\epsilon_i \leq 0 / X_i^1 = x^1, \ldots, X_i^v = x^v\right) = \alpha.$$

Pour déterminer des estimateurs (par splines de régression) de $\Psi_\alpha^1, \ldots, \Psi_\alpha^v$, on utilise l'algorithme backfitting de Hastie et Tibshirani (1990), que l'on décrit ici.

- Initialisation : on pose $\widehat{\Psi}_\alpha^{1,(1)} = \ldots = \widehat{\Psi}_\alpha^{v,(1)} = 0$.
- Étape $r+1$: on considère, pour tout $l = 1, \ldots, v$, la variable

$$Y_i^{l,r+1} = Y_i - \sum_{s=1}^{l-1} \int_0^1 \widehat{\Psi}_\alpha^{s,(r+1)}(t) X_i^s(t) dt - \sum_{s=r+1}^{v} \int_0^1 \widehat{\Psi}_\alpha^{s,(r)}(t) X_i^s(t) dt,$$

  et on considère le modèle à une variable explicative

$$Y_i^{l,r+1} = \int_0^1 \Psi_\alpha(t) X_i^l(t) dt + \epsilon_i.$$

  En utilisant l'algorithme des moindres carrés itérés pondérés présenté précédemment, on obtient ainsi $\widehat{\Psi}_\alpha^{l,(r+1)}$ pour tout $l = 1, \ldots, v$.

IV.2. PRÉVISION PAR LA MOYENNE CONDITIONNELLE

Le chapitre précédent concernait l'estimation de quantiles conditionnels pour variable explicative fonctionnelle en pratique. Dans ce chapitre, on présente brièvement le travail qui a été fait dans le cas de l'estimation de la moyenne conditionnelle pour variable explicative fonctionnelle.

IV.2.1. Estimation par splines de régression

Concernant les splines de régression, Cardot, Ferraty et Sarda (1999, 2003) ont là aussi évalué les performances de leur estimateur au moyen de simulations. Une extension de cet estimateur au cas de plusieurs variables explicatives est possible (de façon analogue a ce qui a été fait pour les quantiles conditionnels à la partie précédente avec un modèle additif et un algorithme "backfitting"). Ceci est proposé par Cardot, Crambes et Sarda (2006), dans le chapitre de e-book évoqué précédemment (voir les chapitres IV.3 et IV.4 pour les détails de cette étude de prévision de pics de pollution).

IV.2.2. Estimation par splines de lissage

Concernant les estimateurs présentés dans les parties II et III de la thèse (estimateur de la moyenne conditionnelle basé sur les splines de lissage lorsque la variable explicative est bruitée ou non), l'expression explicite de l'estimateur (basée sur une écriture matricielle simple) permet une programmation immédiate. On peut alors évaluer ses performances au moyen de simulations.

Celles-ci peuvent être trouvées dans l'article de Cardot, Crambes, Kneip et Sarda (2006) (donné au chapitre III.3. de la thèse), ce qui permet de se faire une idée des performances des estimateurs construits, notamment par la méthode des moindres carrés orthogonaux.

IV.3. DONNÉES DE POLLUTION

Dans cette partie, on donne une rapide description des données de pollution fournies par l'ORAMIP, ainsi qu'une rapide étude de celles-ci. Les différentes variables sont mesurées toutes les heures, la période concernant nos données allant du 15 Mai au 15 Septembre, pour les années 1997, 1998, 1999 et 2000. Les variables mesurées sont des polluants ou des variables météorologiques,

- le monoxyde d'azote NO (en $\mu g/m^3$) : NO,
- le dioxyde d'azote NO_2 (en $\mu g/m^3$) : N2,
- les poussières PM_{10} (en $\mu g/m^3$) : PS,
- l'ozone O_3 (en $\mu g/m^3$) : O3,
- la vitesse du vent (en m/s) : VV,
- la direction du vent (en degrés) : DV,
- la température (en degrés Celsius) : TE,
- l'humidité relative (en pourcentage) : HR.

Ces variables sont mesurées dans différentes stations de l'agglomération de Toulouse,

- la station Jacquier : JAC,
- la station Léo Lagrange : LEO,
- la station des Mazades : MAZ,
- la station Berthelot : BRT,
- la station Colomiers : COL,
- la station Balma : BAL.

De nombreuses données manquantes apparaissent, principalement en raison de pannes ou en l'absence d'appareils de mesures, certaines variables, n'étant jamais observées. On supprime les variables qui ont trop de données manquantes, et on garde pour les autres leur moyenne sur toutes les stations (car les variables sont très similaires d'une station à l'autre). On dispose ainsi de 5 variables NO, N2, O3, DV et VV. On peut avoir une première idée du comportement des variables (par exemple la variable O3 qui nous intéressera plus par la suite) en regardant les courbes journalières. La figure 2 donne une partie des courbes journalières de cette variable O3, ainsi que la courbe moyenne en trait plus clair. Notons que les courbes journalières vont de 18 h à 17 h le lendemain. On constate ainsi une diminution de l'ozone pendant la nuit, alors que le pic d'ozone intervient dans le milieu de l'après-midi.

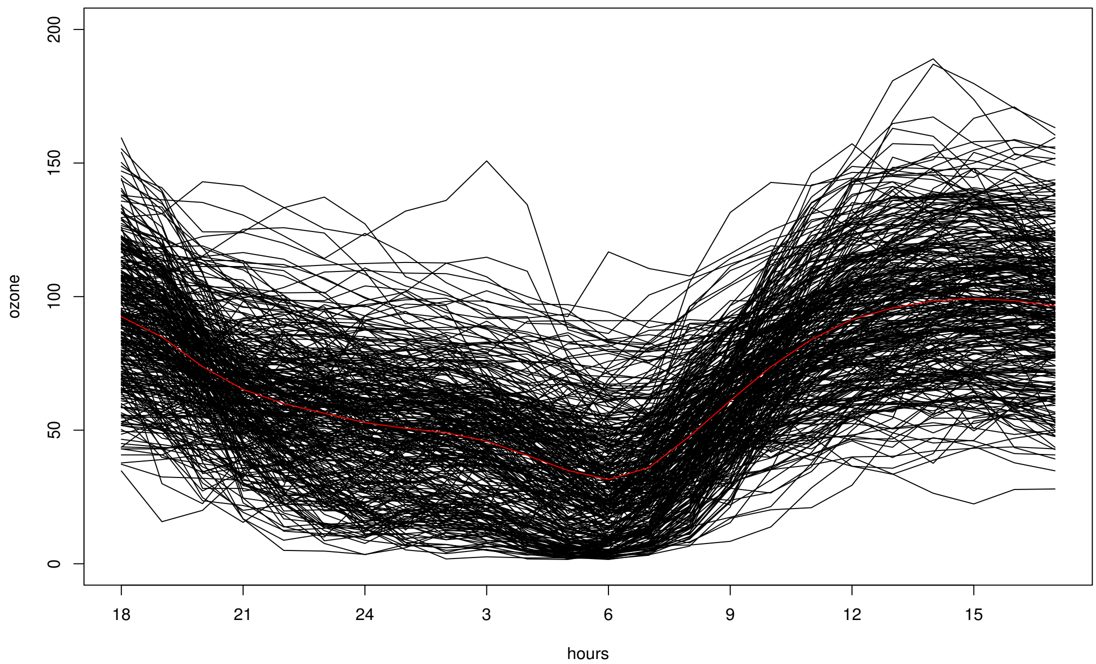


FIGURE 2. Courbes journalières d'ozone.

Une étude descriptive de ces données a été faite (analyses en composantes principales). Une première étude est faite par Cardot, Crambes et Sarda (2004b). Une étude plus détaillée peut être trouvée dans Cardot, Crambes et Sarda (2006). Ce travail regroupe toute une étude de ces données à l'aide du logiciel Xplore. Les méthodes d'estimation vues précédemment (moyenne conditionnelle, quantiles conditionnels) ont été utilisées et les résultats obtenus sont donnés dans le chapitre de ce e-book (voir chapitre suivant).

L'étude de ces données de pollution va être prolongée par un travail en cours, en collaboration avec Hervé Cardot, Alois Kneip et Pascal Sarda, dans le cadre de l'étude de l'estimateur de la moyenne conditionnelle par splines de lissage (estimateur présenté dans la partie II de la thèse). En supposant la normalité des erreurs dans le modèle (2), on considère l'estimation d'une nouvelle donnée Y_{n+1} connaissant une nouvelle variable explicative fonctionnelle X_{n+1}, le but étant de donner un intervalle de prédiction pour Y_{n+1}.

IV.4. OZONE POLLUTION FORECASTING

IV.4.1. Introduction

Prediction of Ozone pollution is currently an important field of research, mainly in a goal of prevention. Many statistical methods have already been used to study data dealing with pollution. For example, Ghattas (1999) used a regression tree approach, while a functional approach has been proposed by Damon and Guillas (2002) and by Aneiros-Perez, Cardot, Estevez-Perez and Vieu (2004). Pollution data often consist now in hourly measurements of pollutants and meteorological data. These variables are then comparable to curves known in some discretization points, usually called *functional data* in the literature (see ramsay and Silverman, 1997). Many examples of such data have already been studied in various fields (see Frank and Friedman, 1993, Ramsay and Silverman, 2002, Ferraty and Vieu, 2002). It seems then natural to propose some models that take into account the fact that the variables are functions of time.

The data we study here were provided by the ORAMIP[(1)], which is an air observatory located in the city of Toulouse (France). We are interested in a pollutant like Ozone. We consider the prediction of the maximum of pollution for a day (maximum of Ozone) knowing the Ozone temporal evolution the day before. To do this, we consider two models. The first one is the functional linear model introduced by ramsay and Dalzell (1991). It is based on the prediction of the conditional mean. The second one is a generalization of the linear model for quantile regression introduced by Koenker and Bassett (1978) when the covariates are curves. It consists in forecasting the conditional median. More

[(1)] "Observatoire Régional de l'Air en Midi-Pyrénées" generally, we introduce this model for the α-conditional quantile, with $\alpha \in]0,1[$. This allows us to give prediction intervals. For both models, a spline estimator of the functional coefficient is introduced, in a way similar to Cardot, Ferraty and Sarda (2003).

This work is divided into four parts. First, we give a brief statistical description and analysis of the data, in particular by the use of principal components analysis (PCA), to study the general behaviour of the variables. Secondly, we present the functional linear model and we propose a spline estimator of the functional coefficient. Similarly, we propose in the third part a spline estimator of the functional coefficient for the α-conditional quantile. In both models, we describe the algorithms that have been implemented to obtain the spline estimator. We also extend these algorithms to the case where there are several functional predictors by the use of a *backfitting* algorithm. Finally, these approaches are illustrated using the real pollution data provided by the ORAMIP.

IV.4.2. A brief analysis of the data

Description of the data. — The data provided by ORAMIP consist in hourly measurements during the period going from the 15$^{\text{th}}$ May to the 15$^{\text{th}}$ September for the years 1997, 1998, 1999 and 2000, of the following variables :

- Nitrogen Monoxide (noted NO),
- Nitrogen Dioxide (noted N2),
- Ozone (noted O3),
- Wind Direction (noted WD),
- Wind Speed (noted WS).

These variables were observed in six different stations in Toulouse. There are some missing data, mainly because of breakdowns. There were also other variables (such as the temperature) for which the missing data were too numerous and we could not use them, so, in the following, we just consider the five variables mentioned above. We first noticed that these variables take values which are very similar from one station to another. Thus, for each variable, we consider the mean of the measurements in the different stations. This approach is one way to deal with missing values.

A descriptive analysis of the variables can show simple links between them. For example, we can see that the mean daily curves of the first three variables NO, N2 and O3 (cf. figure 3) have a similar evolution for NO and N2 (at least in the first part of the day). On the contrary, the curves for NO and O3 have opposite variations. These observations are also confirmed by the correlation matrix of the variables NO, N2 and O3.

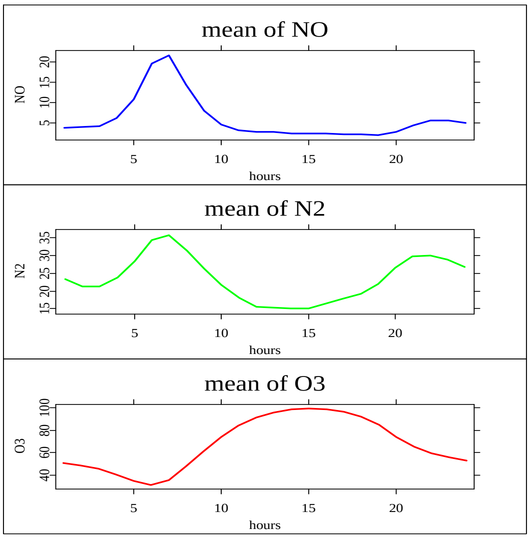


FIGURE 3. Daily mean curves for the variables NO (blue curve), N2 (green curve and O3 (red curve).

Principal Component Analysis. — A first PCA has been done on the matrix whose columns are the different daily mean variables. As these variables have different units, we also consider the reduced matrix. The first two components allow to explain more than 80% of the variance. To visualize the results of this PCA, we have represented the mean hours (figure 4) and the variables (figure 5) in the plane formed by the two first principal axes. We notice on figure 4 that the first axis separates the morning and the afternoon evolution while the second axis separates the day and the night. Concerning figure 5, the first axis separates Nitrogen Monoxide and Nitrogen Dioxide of Ozone. We can also remark that, if we put the graphic 4 on the graphic 5, we find that the maximum of Ozone is in the afternoon and that the quantity of Ozone is low in the morning. It is the contrary for Nitrogen Monoxide and Nitrogen Dioxide.

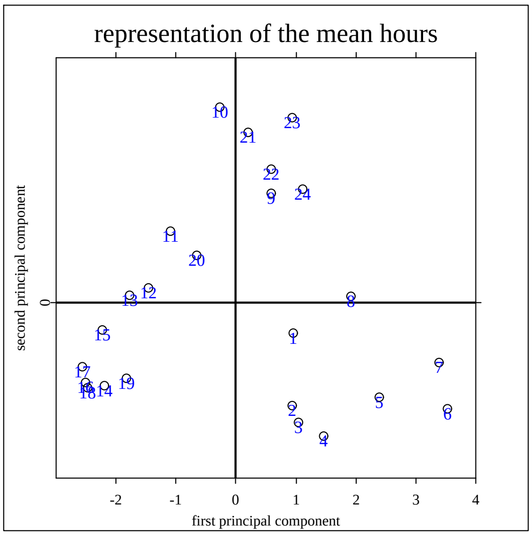


FIGURE 4. Representation of the mean hours 1, . . . , 24 in the plane generated by the two first principal components.

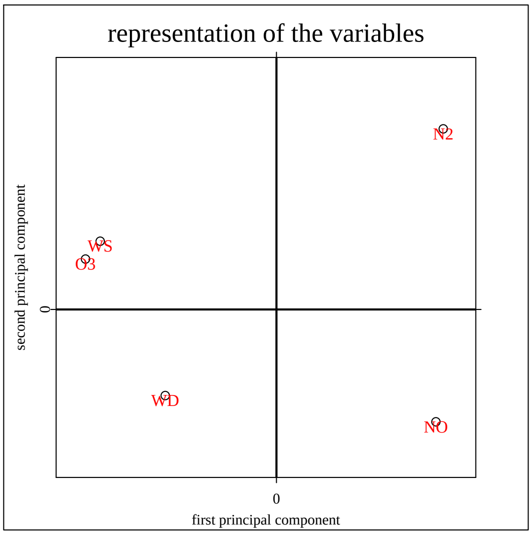


FIGURE 5. Representation of the variables NO, N2, O3, WD and WS in the plane generated by the two first principal components.

Functional Principal Component Analysis. — We also performed a functional PCA (see Ramsay and Silverman, 1997) of the different variables.

We come back here to the functional background where we consider each variable as a curve discretized in some points. We can look at the variations of each variable around its mean by representing the functions μ, $\mu + C\xi$ and $\mu - C\xi$, where μ is the mean curve of the variable, C is a constant and ξ is a principal component. For example, for Ozone, we make this representation for the first principal component (that represents nearly 80% of the information) on figure 6. The constant C has been fixed arbitrarily in this example equal to 10, to obtain a figure easily interpretable. We can see that the first principal component highlights variations around the mean at 3:00 pm. It is the time of the maximum of Ozone in the middle of the afternoon.

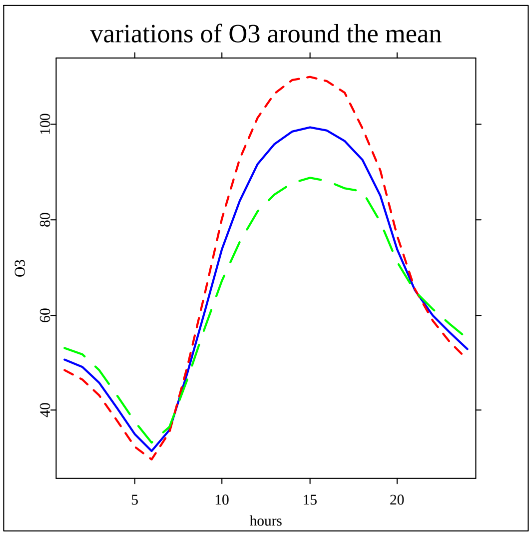


FIGURE 6. Variations of O3 around the mean. The blue solid curve represents the mean curve μ of Ozone, the red dotted curve represents $\mu+10\xi$ where ξ is the first principal component, and the green dashed curve represents $\mu - 10\xi$.

IV.4.3. Functional linear model

We describe now the functional linear model presented for example by Ramsay and Silverman (1997). Let us consider a sample $(X_i, Y_i)_{i=1,\dots,n}$ of pairs of random variables, independent and identically distributed, with the same distribution as (X, Y), with X belonging to the functional space $L^2(D)$ of the integrable square functions defined on a bounded interval D of $\mathbb{R}$, and Y belonging to $\mathbb{R}$. We center each function X_i by introducing $\widetilde{X_i} = X_i - \mathbb{E}(X_i)$. The functional linear model is then defined by

$$Y_i = \mu + \int_D \alpha(t)\widetilde{X_i}(t)\, dt + \epsilon_i, \tag{IV.2}$$

with $\mathbb{E}(\epsilon_i|X_i) = 0$. We have $\mathbb{E}(Y_i) = \mu$ and $\mathbb{E}(Y_i|X_i) = \mu + \int_D \alpha(t)\widetilde{X_i}(t)\, dt$.

In practice, each function X_i is known in $p = 24$ equispaced discretization points $t_1, \ldots, t_p \in D$ (with $t_1 \leq \ldots \leq t_p$). So, the integral above is approximated by

$$\int_D \alpha(t)\widetilde{X_i}(t)\, dt \simeq \frac{\lambda(D)}{p} \sum_{j=1}^{p-1} \alpha(t_j)\widetilde{X_i}(t_j),$$

where $\lambda(D)$ stands for the length of the interval D. More generally, when the discretization points are not equispaced, the integral can be easily approximated by

$$\int_D \alpha(t)\widetilde{X_i}(t)\, dt \simeq \sum_{j=1}^{p-1} (t_{j+1} - t_j)\alpha(t_j)\widetilde{X_i}(t_j).$$

Spline estimation of α. — We choose to estimate the functional coefficient of regression $\alpha : D \longrightarrow \mathbb{R}$ by a spline function (see de Boor, 1978, for details). Let us consider $k \in \mathbb{N}^\star$ and $q \in \mathbb{N}$. We split D into k intervals of the same length. A spline function is a piecewise polynomial function of degree $q \in \mathbb{N}^\star$ on each sub-interval, $(q-1)$ times differentiable on D. The extremities of the sub-intervals are called knots. It is known that the space of such splines functions is a vectorial space of dimension $k+q$. We consider the basis $\mathbf{B}_{k,q}$ of this space called B-splines basis and that we write $\mathbf{B}_{k,q} = {}^t(B_1, \cdots, B_{k+q})$.

We estimate α by a linear combination of the functions B_l, $l = 1, \ldots, k+q$, that leads us to find $\widehat{\mu} \in \mathbb{R}$ and a vector $\widehat{\boldsymbol{\theta}} \in \mathbb{R}^{k+q}$ such that

$$\widehat{\alpha} = \sum_{l=1}^{k+q} \widehat{\theta}_l B_l = {}^t\mathbf{B}_{k,q}\widehat{\boldsymbol{\theta}},$$

with $\widehat{\mu}$ and $\widehat{\boldsymbol{\theta}}$ solutions of the following minimization problem

$$\min_{\mu\in\mathbb{R},\boldsymbol{\theta}\in\mathbb{R}^{k+q}} \left\{ \frac{1}{n}\sum_{i=1}^{n}(Y_i - \mu - \langle {}^t\mathbf{B}_{k,q}\boldsymbol{\theta}, \widetilde{X_i}\rangle)^2 + \rho \left\| ({}^t\mathbf{B}_{k,q}\boldsymbol{\theta})^{(m)} \right\|_{L^2}^2 \right\}, \tag{IV.3}$$

where $({}^t\mathbf{B}_{k,q}\boldsymbol{\theta})^{(m)}$ is the m^{th} derivative of ${}^t\mathbf{B}_{k,q}\boldsymbol{\theta}$ and ρ is a penalization parameter that allows to control the smoothness of the estimator (see Cardot, Ferraty and Sarda, 2003). The notation $\langle .,. \rangle$ refers to the usual inner product of $L^2(D)$ and $\|.\|_{L^2}$ is the norm induced by this inner product.

If we set $\boldsymbol{\beta} = \begin{pmatrix} \mu \\ \boldsymbol{\theta} \end{pmatrix} \in \mathbb{R}^{k+q+1}$, then, the solution of the minimization problem (IV.3) above is given by

$$\widehat{\boldsymbol{\beta}} = \frac{1}{n}(\frac{1}{n}\,{}^t\mathbf{D}_X\mathbf{D}_X + \rho\mathbf{K}_k)^{-1}\,{}^t\mathbf{D}_X\mathbf{Y},$$

with

$$\mathbf{D}_X = \begin{pmatrix} 1 & \langle B_1, X_1\rangle & \dots & \langle B_{k+q}, X_1\rangle \\ \vdots & \vdots & & \vdots \\ 1 & \langle B_1, X_n\rangle & \dots & \langle B_{k+q}, X_n\rangle \end{pmatrix} \quad \text{and} \quad \mathbf{K}_k = \begin{pmatrix} 0 & 0 \\ 0 & \mathbf{G}_k \end{pmatrix},$$

where $\mathbf{G}_k$ is the $(k+q)\times(k+q)$ matrix with elements $\langle B_j^{(m)}, B_l^{(m)}\rangle$. It also satisfies

$${}^t\boldsymbol{\theta}\mathbf{G}_k\boldsymbol{\theta} = \left\| ({}^t\mathbf{B}_{k,q}\boldsymbol{\theta})^{(m)} \right\|_{L^2}^2 .$$

The computation of the matrices $\mathbf{D}_X$ and $\mathbf{G}_k$ is performed with the Xplore functions `bspline` and `bsplineini`.

Let us notice that a convergence result for this spline estimator is given by Cardot, Ferraty and Sarda (2003).

Selection of the parameters. — The estimator defined by (IV.3) depends on a large number of parameters : the number of knots k, the degree q of splines, the order m of derivation in the penalization term, and the smoothing parameter ρ. It seems (see Marx and Eilers, 1999, Besse, Cardot and Ferraty, 1997) that only the penalization parameter ρ is really important provided that the number of knots is large enough.

The parameter ρ is chosen by the generalized cross validation criterion (see Wahba, 1990) which is described below.

Consider the "hat matrix" $\mathbf{H}(\rho) = \frac{1}{n}\mathbf{D}_X(\frac{1}{n}\,{}^t\mathbf{D}_X\mathbf{D}_X + \rho\mathbf{K}_k)^{-1}\,{}^t\mathbf{D}_X$. It satisfies $\widehat{\mathbf{Y}} = \mathbf{H}(\rho)\mathbf{Y}$. The generalized cross validation criterion is then given by

$$GCV(\rho) = \frac{\frac{1}{n}\sum_{l=1}^{n}(Y_l - \widehat{Y}_l)^2}{\Big(1 - \frac{1}{n}Tr(\mathbf{H}(\rho))\Big)^2}\,. \tag{IV.4}$$

We select the optimal parameter ρ_{GCV} as the one that minimizes the GCV criterion (IV.4). Let us notice that we do not have to compute the matrix $\mathbf{H}(\rho)$ (whose size is $n \times n$) since we have $Tr(\mathbf{H}(\rho)) = Tr(\frac{1}{n}\,{}^t\mathbf{D}_X\mathbf{D}_X(\frac{1}{n}\,{}^t\mathbf{D}_X\mathbf{D}_X + \rho\mathbf{K}_k)^{-1})$.

The Xplore function `sflmgcv` uses this GCV criterion and gives the estimations of μ, θ and α.

Multiple functional linear model. — We now want to generalize the model (IV.2) to the case where there are several (centered) functional covariates $\widetilde{X}^1, \ldots, \widetilde{X}^v$. We consider the following additive model

$$\text{(IV.5)} \qquad Y_i = \mu + \int_D \alpha_1(t)\widetilde{X_i^1}(t)\,dt + \ldots + \int_D \alpha_v(t)\widetilde{X_i^v}(t)\,dt + \epsilon_i.$$

To get the estimates of $\mu, \alpha_1, \ldots, \alpha_v$, we used the *backfitting* algorithm (see Hastie and Tibshirani, 1990), which principle is described below. It allows us to avoid inverting large scale matrices and leads to a faster estimation procedure. The Xplore function giving the estimates of $\mu, \alpha_1, \ldots, \alpha_v$ using the backfitting algorithm for v covariates is `sflmgcvmult`.

- **Step 1**
  We initialize $\widehat{\alpha_1}^{(1)}, \ldots, \widehat{\alpha_{v-1}}^{(1)}$ to 0 and $\widehat{\mu}$ to $\frac{1}{n}\sum_{i=1}^n Y_i$. Then, we determine $\widehat{\mu}^{(1)}$ and $\widehat{\alpha_v}^{(1)}$ by using the spline estimation procedure for the functional linear model with one covariate.
- **Step 2**
  For $r = 1, \ldots, v$, we consider

$$Y_i^{r,2} = Y_i - \sum_{l=1}^{r-1} \int_D \widehat{\alpha_l}^{(2)}(t)\widetilde{X_i^l}(t)\,dt - \sum_{l=r+1}^{v} \int_D \widehat{\alpha_l}^{(1)}(t)\widetilde{X_i^l}(t)\,dt,$$

  and we make a simple functional regression

$$Y_i^{r,2} = \mu + \int_D \alpha_r^{(2)}(t)\widetilde{X_i^r}(t)\,dt + \epsilon_i.$$

  Then, we obtain $\widehat{\mu}^{(2)}$ and $\widehat{\alpha_r}^{(2)}$, for $r = 1, \ldots, v$. The optimal penalization parameter is determined for each estimator with generalized cross validation.
- **Step** $j+1$
  While $\max_{r=1,\ldots,v}(\|\widehat{\alpha_r}^{(j)} - \widehat{\alpha_r}^{(j-1)}\|) > \xi$ (where ξ is an error constant to be fixed), we consider

$$Y_i^{r,j+1} = Y_i - \sum_{l=1}^{r-1} \int_D \widehat{\alpha_l}^{(j+1)}(t)\widetilde{X_i^l}(t)\,dt - \sum_{l=r+1}^{v} \int_D \widehat{\alpha_l}^{(j)}(t)\widetilde{X_i^l}(t)\,dt,$$

  and we make a simple functional regression

$$Y_i^{r,j+1} = \mu + \int_D \alpha_r^{(j+1)}(t)\widetilde{X_i^r}(t)\,dt + \epsilon_i,$$

by using the estimator defined for the functional linear model with one covariate. We then deduce $\widehat{\mu}^{(j+1)}$ and $\widehat{\alpha_r}^{(j+1)}$, for $r = 1, \ldots, v$. The optimal penalization parameter is determined for each estimator with generalized cross validation.

IV.4.4. Conditional quantiles estimation

Our goal is now to find the Ozone threshold value such that the conditional probability to exceed this value is equal to a certain given risk $\alpha \in]0,1[$. More precisely, if Y is a real random value, we define its α-quantile by the real number q_α such that

$$\mathbb{P}(Y \leq q_\alpha) = \alpha.$$

Koenker and Bassett (1978) use the following property to define quantile estimators (which can be naturally generalized to conditional quantiles) :

$$q_\alpha = \arg\min_{a \in \mathbb{R}} \mathbb{E}(l_\alpha(Y - a)),$$

with

$$l_\alpha(u) = \mid u \mid + (2\alpha - 1)u.$$

Let us now come back to our functional case. We still consider the sample $(X_i, Y_i)_{i=1,\ldots,n}$ of pairs of random variables, independent and identically distributed, with the same distribution as (X, Y), with X belonging to the functional space $L^2(D)$, and Y belonging to $\mathbb{R}$. Without loss of generality, we suppose that X is a centered variable, that is to say $\mathbb{E}(X) = 0$. Let α be a real number in $]0,1[$ and x a function in $L^2(D)$. We suppose that the *conditional α-quantile* of Y given $[X = x]$ is the unique scalar $g_\alpha(x)$ such that

$$\mathbb{P}(Y \leq g_\alpha(x) | X = x) = \alpha, \tag{IV.6}$$

where $\mathbb{P}(.|X = x)$ is the conditional probability given $[X = x]$.

Let us remark that $g_\alpha(x)$ can be defined in an equivalent way as the solution of the minimization problem

$$\min_{a\in\mathbb{R}} \mathbb{E}(l_\alpha(Y - a)|X = x). \tag{IV.7}$$

We assume now that there exists a unique function $\Psi_\alpha \in L^2(D)$ such that g_α can be written in the following way

$$g_\alpha(X) = c + \langle \Psi_\alpha, X\rangle = c + \int_D \Psi_\alpha(t)X(t)\, dt. \tag{IV.8}$$

This condition can be seen as a direct generalization of the model introduced by Koenker and Bassett (1978), the difference being that here, the covariates are functions.

Spline estimator of Ψ_α. — Our goal is now to give a nonparametric estimator of the function Ψ_α. In the case where the covariate X is real, many nonparametric estimators have already been proposed (see for example Bhattacharya and Gangopadhyay, 1990, Fan, Hu and Truong, 1994, Lejeune and Sarda, 1988 or He and Shi, 1994).

As for the spline estimator of the conditional mean, we consider the vectorial space of spline functions with $k - 1$ interior knots and of degree q, and its B-splines basis $\mathbf{B}_{k,q} = {}^t(B_1, \ldots, B_{k+q})$. We estimate Ψ_α by a linear combination of the B_l functions for l going from 1 to $k + q$. This leads us to find a vector $\widehat{\boldsymbol{\theta}} = {}^t(\widehat{\theta}_1, \ldots, \widehat{\theta}_{k+q})$ in $\mathbb{R}^{k+q}$ such that

$$\widehat{\Psi}_\alpha = \sum_{l=1}^{k+q} \widehat{\theta}_l B_l = {}^t\mathbf{B}_{k,q}\widehat{\boldsymbol{\theta}}. \tag{IV.9}$$

The vector $\widehat{\boldsymbol{\theta}}$ will be solution of the following minimization problem, which is the penalized empirical version of (IV.7),

$$\text{(IV.10)} \qquad \min_{c\in\mathbb{R},\theta\in\mathbb{R}^{k+q}} \left\{ \frac{1}{n}\sum_{i=1}^{n} l_\alpha(Y_i - c - \langle {}^t\mathbf{B}_{k,q}\boldsymbol{\theta}, X_i\rangle) + \rho \left\| ({}^t\mathbf{B}_{k,q}\boldsymbol{\theta})^{(m)} \right\|_{L^2}^2 \right\},$$

where $({}^t\mathbf{B}_{k,q}\boldsymbol{\theta})^{(m)}$ is the m-th derivative of the spline function ${}^t\mathbf{B}_{k,q}\boldsymbol{\theta}$ and ρ is a penalization parameter which role is to control the smoothness of the estimator, as for the minimization problem (IV.3). This criterion is similar to (IV.3), the quadratic function being here replaced by the loss function l_α. In this case, we have to deal with an optimization problem that does not have an explicit solution, contrary to the estimation of the conditional mean. That is why we adopted the strategy proposed by Lejeune and Sarda (1988). It is based on an algorithm that consists in performing iterative weighted least squares (see Ruppert and Caroll, 1988). Let us consider the function δ_i defined by

$$\delta_i(\alpha) = 2\alpha \mathbb{1}_{\{Y_i - c - \langle {}^t\mathbf{B}_{k,q}\theta, X_i\rangle \geq 0\}} + 2(1-\alpha)\mathbb{1}_{\{Y_i - c - \langle {}^t\mathbf{B}_{k,q}\theta, X_i\rangle < 0\}}.$$

The minimization problem (IV.10) is then equivalent to

$$\text{(IV.11)} \quad \min_{c\in\mathbb{R},\boldsymbol{\theta}\in\mathbb{R}^{k+q}} \left\{ \frac{1}{n}\sum_{i=1}^{n} \delta_i(\alpha) \mid Y_i - c - \langle {}^t\mathbf{B}_{k,q}\boldsymbol{\theta}, X_i\rangle \mid + \rho \left\| ({}^t\mathbf{B}_{k,q}\boldsymbol{\theta})^{(m)} \right\|_{L^2}^2 \right\}.$$

Then, we can approximate this criterion by replacing the absolute value by a weighted quadratic term, hence we can obtain a sequence of explicit solutions. The principle of this Iterative Reweighted Least Squares algorithm is described below.

– **Initialization**
  We determine $\boldsymbol{\beta}^1 = {}^t(c^1, \boldsymbol{\theta}^1)$ solution of the minimization problem

$$\min_{c\in\mathbb{R},\boldsymbol{\theta}\in\mathbb{R}^{k+q}} \left\{ \frac{1}{n}\sum_{i=1}^{n} (Y_i - c - \langle {}^t\mathbf{B}_{k,q}\boldsymbol{\theta}, X_i\rangle)^2 + \rho \left\| ({}^t\mathbf{B}_{k,q}\boldsymbol{\theta})^{(m)} \right\|_{L^2}^2 \right\},$$

  which solution $\boldsymbol{\beta}^1$ is given by $\boldsymbol{\beta}^1 = \frac{1}{n}(\frac{1}{n}\ {}^t\mathbf{D}_X\mathbf{D}_X + \rho\mathbf{K}_k)^{-1}\ {}^t\mathbf{D}_X\mathbf{Y}$, with $\mathbf{D}_X$ and $\mathbf{K}_k$ defined previously.
– **Step j+1**

Knowing $\boldsymbol{\beta}^j = {}^t(c^j, \boldsymbol{\theta}^j)$, we determine $\boldsymbol{\beta}^{j+1} = {}^t(c^{j+1}, \boldsymbol{\theta}^{j+1})$ solution of the minimization problem

$$\min_{c\in\mathbb{R},\boldsymbol{\theta}\in\mathbb{R}^{k+q}} \left\{ \frac{1}{n}\sum_{i=1}^{n} \frac{\delta_i^j(\alpha)(Y_i - c - \langle {}^t\mathbf{B}_{k,q}\boldsymbol{\theta}, X_i\rangle)^2}{[(Y_i - c - \langle {}^t\mathbf{B}_{k,q}\boldsymbol{\theta}, X_i\rangle)^2 + \eta^2]^{1/2}} + \rho \left\| ({}^t\mathbf{B}_{k,q}\boldsymbol{\theta})^{(m)} \right\|_{L^2}^2 \right\},$$

where $\delta_i^j(\alpha)$ is $\delta_i(\alpha)$ on step j of the algorithm, and η is a strictly positive constant that allows us to avoid a denominator equal to zero. Let us define the $n \times n$ diagonal matrix $\mathbf{W}_j$ with diagonal elements given by

$$[\mathbf{W}_j]_{ll} = \frac{\delta_1^j(\alpha)}{n[(Y_l - c - \langle {}^t\mathbf{B}_{k,q}\boldsymbol{\theta}, X_l\rangle)^2 + \eta^2]^{1/2}}.$$

Then, $\beta^{j+1} = ({}^t\mathbf{D}_X\mathbf{W}_j\mathbf{D}_X + \rho\mathbf{K}_k)^{-1}\, {}^t\mathbf{D}_X\mathbf{W}_j\mathbf{Y}$.

Remark : Since our algorithm relies on weighted least squares, we can derive a generalized cross validation criterion to choose the penalization parameter value ρ at each step of the algorithm. Indeed, the "hat matrix" defined by $\mathbf{H}(\rho) = \mathbf{D}_X({}^t\mathbf{D}_X\mathbf{W}\mathbf{D}_X + \rho\mathbf{K}_k)^{-1}\, {}^t\mathbf{D}_X\mathbf{W}$ satisfies $\widehat{\mathbf{Y}} = \mathbf{H}(\rho)\mathbf{Y}$, where $\mathbf{W}$ is the weight matrix obtained at the previous step of the algorithm. The generalized cross validation criterion is then given by

$$GCV(\rho) = \frac{\frac{1}{n}\, {}^t(\mathbf{Y} - \widehat{\mathbf{Y}})\mathbf{W}(\mathbf{Y} - \widehat{\mathbf{Y}})}{\left(1 - \frac{1}{n}Tr(\mathbf{H}(\rho))\right)^2}, \tag{IV.12}$$

where $Tr(\mathbf{H}(\rho)) = Tr(\, {}^t\mathbf{D}_X\mathbf{W}({}^t\mathbf{D}_X\mathbf{W}\mathbf{D}_X + \rho\mathbf{K}_k))$.

We select the optimal parameter ρ_{GCV} as the one that minimizes the GCV criterion (IV.12). The Xplore function `squantgcv` uses this GCV criterion and gives the estimations of c, θ and Ψ_α.

A convergence result of the estimator $\widehat{\Psi}_\alpha$ is also available in Cardot, Crambes and Sarda (2005).

Multiple conditional quantiles. — Assuming we have now v functional covariates $X^1, \ldots, X^v$, this estimation procedure can be easily extended. We consider the following model

$$P(Y_i \leq g_\alpha^1(X_i^1) + \ldots + g_\alpha^v(X_i^v) / X_i^1 = x_i^1, \ldots, X_i^v = x_i^v) = \alpha. \tag{IV.13}$$

Similarly as before, we assume that $g_\alpha^1(X_i^1) + \ldots + g_\alpha^v(X_i^v) = c + \langle \Psi_\alpha^1, X_i^1 \rangle + \ldots + \langle \Psi_\alpha^v, X_i^v \rangle$ with $\Psi_\alpha^1, \ldots, \Psi_\alpha^v$ in $L^2(D)$. The estimation of each function Ψ_α^r is obtained using the iterative backfitting algorithm combined with the Iterative Reweighted Least Squares algorithm. The Xplore function giving the estimates of $c, \Psi_\alpha^1, \ldots, \Psi_\alpha^v$ is `squantgcvmult`.

IV.4.5. Application to Ozone prediction

We want to predict the variable maximum of Ozone one day i, noted Y_i, using the functional covariates observed the day before until 5 :00 pm. We consider covariates with length of 24 hours. We can assume that beyond 24 hours, the effects of the covariate are negligible knowing the last 24 hours, so each curve X_i begins at 6 :00 pm the day $i-2$.

We ramdomly splitted the initial sample $(X_i, Y_i)_{i=1,\ldots,n}$ into two sub-samples :

- a learning sample $(X_{a_i}, Y_{a_i})_{i=1,\ldots,n_l}$ whose size is $n_l = 332$, used to compute the estimators $\widehat{\mu}$ and $\widehat{\alpha}$ for the functional linear model and the estimators $\widehat{c}$ and $\widehat{\Psi}_\alpha$ for the model with quantiles,
- a test sample $(X_{t_i}, Y_{t_i})_{i=1,\ldots,n_t}$ whose size is $n_t = 142$, used to evaluate the quality of the models and to make a comparison between them.

We also have chosen to take $k = 8$ for the number of knots, $q = 3$ for the degree of spline functions and $m = 2$ for the order of the derivative in the penalization.

To predict the value of Y_i, we use the conditional mean and the conditional median (*i.e.* $\alpha = 0.5$). To judge the quality of the models, we give a prediction of the maximum of Ozone for each element of the test sample,

$$\widehat{Y_{t_i}} = \widehat{\mu} + \int_D \widehat{\alpha}(t) X_{t_i}(t)\, dt,$$

for the prediction of the conditional mean, and

$$\widehat{Y_{t_i}} = \widehat{c} + \int_D \widehat{\Psi}_\alpha(t) X_{t_i}(t)\, dt$$

for the prediction of the conditional median.

Then, we consider three criteria given by

$$C_1 = \frac{\frac{1}{n_t}\sum_{i=1}^{n_t}(Y_{t_i} - \widehat{Y_{t_i}})^2}{\frac{1}{n_t}\sum_{i=1}^{n_t}(Y_{t_i} - \overline{Y}_l)^2},$$

$$C_2 = \frac{1}{n_t}\sum_{i=1}^{n_t} \mid Y_{t_i} - \widehat{Y_{t_i}} \mid,$$

$$C_3 = \frac{\frac{1}{n_t}\sum_{i=1}^{n_t} l_\alpha(Y_{t_i} - \widehat{Y_{t_i}})}{\frac{1}{n_t}\sum_{i=1}^{n_t} l_\alpha(Y_{t_i} - q_\alpha(Y_l))},$$

where $\overline{Y}_l$ is the empirical mean of the learning sample $(Y_{a_i})_{i=1,\ldots,n_l}$ and $q_\alpha(Y_l)$ is the empirical α-quantile of the learning sample $(Y_{a_i})_{i=1,\ldots,n_l}$. This last criterion C_3 is similar to the one proposed by Koenker and Machado (1999). We remark that, the more these criteria take low values (close to zero), the better is the prediction. These three criteria are all computed on the test sample.

Prediction of the conditional mean. — The values of the criteria C_1 and C_2 are given in the table 4. It appears that the best model with one covariate to predict the maximum of Ozone is the one that use the curve of Ozone the day before. We have also built multiple functional linear models, in order to improve the prediction. The errors for these models are also given in table 4. It appears that the best model is the one that use the four covariates Ozone, Nitrogen Monoxide, Wind Direction and Wind Speed. So, adding other covariates allows to improve the prediction, even if the gain is low.

Models	Variables	C_1	C_2
models with 1 covariate	NO	0.828	16.998
	N2	0.761	16.153
	O3	**0.416**	**12.621**
	WD	0.910	18.414
	WS	0.796	16.756
models with 2 covariates	O3, NO	0.409	12.338
	O3, N2	0.410	12.373
	O3, WD	0.405	12.318
	O3, WS	0.400	12.267
models with 3 covariates	O3, NO, N2	0.408	12.305
	O3, NO, WD	0.394	11.956
	O3, NO, WS	0.397	12.121
	O3, N2, WD	0.397	12.003
	O3, N2, WS	0.404	12.156
	O3, WD, WS	0.397	12.101
models with 4 covariates	**O3, NO, WD, WS**	**0.391**	**11.870**
	O3, NO, N2, WD	0.395	11.875
	O3, NO, N2, WS	0.398	12.069
	O3, N2, WD, WS	0.394	11.962
model with 5 covariates	O3, NO, N2, WD, WS	0.392	11.877

TABLE 4. Prediction error criteria C_1 and C_2 for the different functional linear models.

Prediction of the conditional median. — Table 5 gathers the prediction errors of the different models. As for the functional linear model, the best prediction using one covariate is the one obtained by using the Ozone curve the day before. Moreover, the prediction is slightly improved by adding other covariates. The best prediction for the criterion C_3 is obtained for the model using the covariates Ozone, Nitrogen Monoxide, Nitrogen Dioxide and Wind Speed. For this model with these four covariates, we have represented on figure 7 the GCV criterion versus $-\log(\rho)$ for the different values of ρ from 10^{-5} to 10^{-10}. The minimum value of the GCV criterion is reached for $\rho = 10^{-8}$. Figure 8 represents the predicted maximum of Ozone (with this model of 4 covariates) versus the measured maximum of Ozone for the test sample. We see on this graphic that the points are quite close to the straight line of equation $y = x$.

Another interest of the conditional quantiles is that we can build some prediction intervals for the maximum of Ozone, which can be quite useful in the

Models	Variables	C_1	C_2	C_3
models with 1 covariate	NO	0.826	16.996	0.911
	N2	0.805	16.800	0.876
	O3	**0.425**	**12.332**	**0.661**
	WD	0.798	18.836	0.902
	WS	0.885	18.222	0.976
models with 2 covariates	O3, NO	0.412	12.007	0.643
	O3, N2	0.405	11.936	0.640
	O3, WD	0.406	12.109	0.649
	O3, WS	0.406	11.823	0.633
models with 3 covariates	O3, NO, N2	0.404	11.935	0.639
	O3, NO, WD	0.404	12.024	0.644
	O3, NO, WS	0.407	11.832	0.638
	O3, N2, WD	0.402	11.994	0.642
	O3, N2, WS	0.403	12.108	0.641
	O3, WD, WS	0.403	12.123	0.640
models with 4 covariates	O3, NO, WD, WS	0.399	11.954	0.641
	O3, NO, N2, WD	0.397	11.921	0.639
	O3, NO, N2, WS	**0.397**	**11.712**	**0.634**
	O3, N2, WD, WS	0.398	11.952	0.640
model with 5 covariates	O3, NO, N2, WD, WS	0.397	11.864	0.638

TABLE 5. Prediction error criteria C_1, C_2 and C_3 for the different functional quantile regression models.

context of prevention of Ozone pollution. Coming back to the initial sample (that is to say when the days are chronologically ordered), we have plotted on figure 9 the measures of the maximum of Ozone during the first 40 days of our sample, that is to say from the 17$^{\text{th}}$ May of 1997 to the 25$^{\text{th}}$ June of 1997 (blue solid curve). The red dotted curve above represents the values of the 90% quantile and the green dashed curve below represents the values of the 10% quantile predicted for these measures. The prediction model used is again the quantile regression model with the 4 covariates O3, NO, N2 and WS.

Analysis of the results. — Both models, the functional linear model and the model with conditional quantiles for functional covariates, give satisfying results concerning the maximum of Ozone prediction. Concerning figure 8, it seems that few values are not well predicted. This highlights a common problem for statistical models, which get into trouble when predicting extreme values

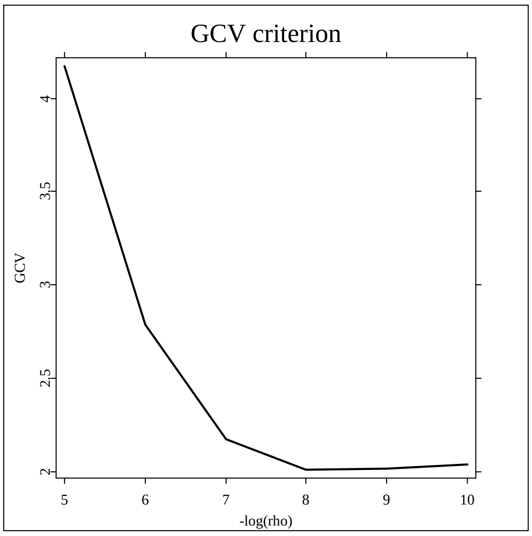


FIGURE 7. Generalized Cross Validation criterion for different values of ρ in the quantile regression model using the covariates O3, NO, N2, WS.

(outliers). The interval of prediction given by the 90% and 10% conditional quantiles can be an interesting answer to that problem, as seen on figure 9.

In spite of the lack of some important variables in the model, such as temperature for example, we can produce good estimators of maximum of pollution knowing the data the day before. The most efficient variable to estimate the maximum of Ozone is the Ozone curve the day before ; however, we noticed that prediction accuracy can be improved by adding other variables in the model. We can suppose that it will be possible to improve again these results when other covariates will be available from ORAMIP, such as temperature curves.

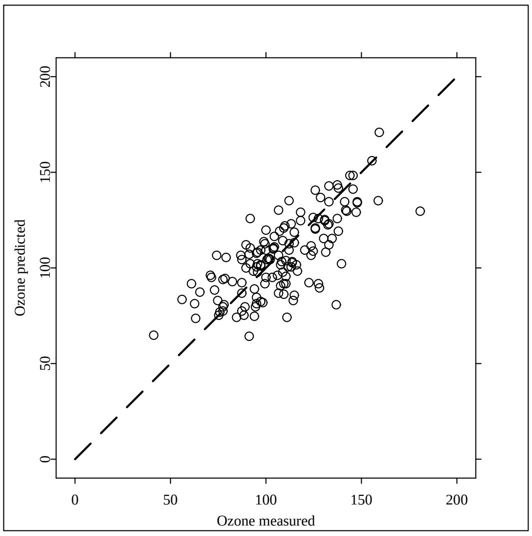


FIGURE 8. Predicted Ozone versus measured Ozone for the test sample, using the prediction quantile regression model with the covariates O3, NO, N2, WS.

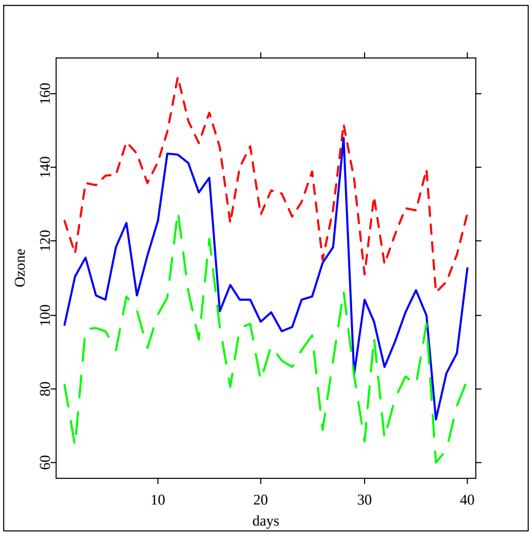


FIGURE 9. Prediction interval of the measures of maximum of Ozone for the period going from the 17^{th} May of 1997 to the 25^{th} June of 1997 (blue solid curve). The red dotted curve and the green dashed curve represent respectively the values of the 90% and 10% quantiles predicted for these measures.

PARTIE V

ANNEXE

V.1. VARIABLE EXPLICATIVE BRUITÉE - PREUVES

Preuve de la proposition III.3. — La solution du problème de minimisation (III.7) est donnée par

$$\widehat{\boldsymbol{\theta}}_{FTLS} = \frac{1}{n}\left(\frac{1}{n}\mathbf{D}_W^\tau \mathbf{D}_W + \rho \mathbf{G}_k - \sigma_{min}^2 \mathbf{B}_k\right)^{-1} \mathbf{D}_W^\tau \mathbf{Y}.$$

Preuve: En utilisant la matrice $\boldsymbol{\beta}$, on écrit $\boldsymbol{\alpha} = \boldsymbol{\beta}\boldsymbol{\theta}$ avec $\boldsymbol{\theta} \in \mathbb{R}^{k+q}$. On tire alors de l'écriture du modèle

$$\left(\left(\frac{\mathbf{W}\boldsymbol{\beta}}{p}, \mathbf{Y}\right) - \left(\frac{\boldsymbol{\delta}\boldsymbol{\beta}}{p}, \boldsymbol{\epsilon}\right)\right)\begin{pmatrix} \boldsymbol{\theta} \\ -1 \end{pmatrix} = 0, \tag{V.1}$$

ce qui nous permet d'écrire le problème de minimisation (III.7) sous la forme

$$\min_{(V.1)} \left\{ \frac{1}{n} \left\| \left(\frac{\boldsymbol{\delta}}{\sqrt{p}}, \boldsymbol{\epsilon} \right) \right\|_F^2 + \rho \boldsymbol{\theta}^\tau \mathbf{G}_k \boldsymbol{\theta} \right\},$$

où la notation $\|.\|_F$ désigne toujours la norme de Frobenius matricielle. On considère alors le problème de minimisation

$$\min_{\mathbf{A}\boldsymbol{x}=\mathbf{E}\boldsymbol{x}} \left\{ \frac{1}{n} \left\| \left(\frac{\boldsymbol{\delta}}{\sqrt{p}}, \boldsymbol{\epsilon} \right) \right\|_F^2 + \rho \boldsymbol{x}^\tau \mathbf{K}_k \boldsymbol{x} \right\}, \tag{V.2}$$

avec $\mathbf{A} = \left(\frac{\mathbf{W}\boldsymbol{\beta}}{p}, \mathbf{Y} \right)$, $\mathbf{E} = \left(\frac{\boldsymbol{\delta}\boldsymbol{\beta}}{p}, \boldsymbol{\epsilon} \right)$ et $\boldsymbol{x} = \binom{\boldsymbol{\theta}}{-1}$. En notant que

$$\begin{aligned}
\frac{1}{n} \boldsymbol{x}^\tau \boldsymbol{\gamma}^\tau \left(\frac{\boldsymbol{\delta}}{\sqrt{p}}, \boldsymbol{\epsilon} \right)^\tau \left(\frac{\boldsymbol{\delta}}{\sqrt{p}}, \boldsymbol{\epsilon} \right) \boldsymbol{\gamma} \boldsymbol{x} &= \frac{1}{n} \boldsymbol{x}^\tau \mathbf{E}^\tau \mathbf{E} \boldsymbol{x} \\
&= \frac{1}{n} \boldsymbol{x}^\tau \mathbf{A}^\tau \mathbf{A} \boldsymbol{x} \\
&= \frac{1}{n} \boldsymbol{x}^\tau \boldsymbol{\gamma}^\tau \left(\frac{\mathbf{W}}{\sqrt{p}}, \mathbf{Y} \right)^\tau \left(\frac{\mathbf{W}}{\sqrt{p}}, \mathbf{Y} \right) \boldsymbol{\gamma} \boldsymbol{x},
\end{aligned}$$

on voit que la quantité

$$\begin{aligned}
& \frac{1}{n} \boldsymbol{x}^\tau \boldsymbol{\gamma}^\tau \left(\frac{\mathbf{W}}{\sqrt{p}}, \mathbf{Y} \right)^\tau \left(\frac{\mathbf{W}}{\sqrt{p}}, \mathbf{Y} \right) \boldsymbol{\gamma} \boldsymbol{x} + \rho \boldsymbol{x}^\tau \mathbf{K}_k \boldsymbol{x} \\
= \; & \frac{1}{n} \boldsymbol{x}^\tau \boldsymbol{\gamma}^\tau \left(\frac{\mathbf{W}}{\sqrt{p}}, \mathbf{Y} \right)^\tau \left(\frac{\mathbf{W}}{\sqrt{p}}, \mathbf{Y} \right) \boldsymbol{\gamma} \boldsymbol{x} + \boldsymbol{x}^\tau \boldsymbol{\gamma}^\tau \boldsymbol{\gamma} \left(\boldsymbol{\gamma}^\tau \boldsymbol{\gamma} \right)^{-1} \left(\rho \mathbf{K}_k \right) \left(\boldsymbol{\gamma}^\tau \boldsymbol{\gamma} \right)^{-1} \boldsymbol{\gamma}^\tau \boldsymbol{\gamma} \boldsymbol{x}
\end{aligned}$$

est minimisée pour $\boldsymbol{\gamma}\boldsymbol{x}$ vecteur propre de la matrice

$$\frac{1}{n} \left(\frac{\mathbf{W}}{\sqrt{p}}, \mathbf{Y} \right)^\tau \left(\frac{\mathbf{W}}{\sqrt{p}}, \mathbf{Y} \right) + \boldsymbol{\gamma} \left(\boldsymbol{\gamma}^\tau \boldsymbol{\gamma} \right)^{-1} \left(\rho \mathbf{K}_k \right) \left(\boldsymbol{\gamma}^\tau \boldsymbol{\gamma} \right)^{-1} \boldsymbol{\gamma}^\tau,$$

correspondant à la plus petite valeur propre non nulle, notée σ^2_{min}. En utilisant la définition de cette valeur propre, on déduit que

$$\left(\frac{1}{n} \left(\frac{\mathbf{W}}{\sqrt{p}}, \mathbf{Y} \right)^\tau \left(\frac{\mathbf{W}}{\sqrt{p}}, \mathbf{Y} \right) + \boldsymbol{\gamma} \left(\boldsymbol{\gamma}^\tau \boldsymbol{\gamma} \right)^{-1} \left(\rho \mathbf{K}_k \right) \left(\boldsymbol{\gamma}^\tau \boldsymbol{\gamma} \right)^{-1} \boldsymbol{\gamma}^\tau \right) \boldsymbol{\gamma} \widehat{\boldsymbol{x}} = \sigma^2_{min} \boldsymbol{\gamma} \widehat{\boldsymbol{x}},$$

ce qui donne, en prémultipliant par $\boldsymbol{\gamma}^\tau$,

$$\frac{1}{n}\left(\frac{\mathbf{W}\boldsymbol{\beta}}{p},\mathbf{Y}\right)^{\tau}\left(\frac{\mathbf{W}\boldsymbol{\beta}}{p},\mathbf{Y}\right)\begin{pmatrix}\widehat{\boldsymbol{\theta}}\\ -1\end{pmatrix}=\sigma^2_{min}\begin{pmatrix}\frac{\boldsymbol{\beta}^{\tau}\boldsymbol{\beta}}{p}\widehat{\boldsymbol{\theta}}\\ -1\end{pmatrix}.$$

Finalement, en gardant les $k+q$ premières lignes, on obtient

$$\widehat{\boldsymbol{\theta}}=\frac{1}{n}\left(\frac{1}{n}\mathbf{D}_W^{\tau}\mathbf{D}_W+\rho\mathbf{G}_k-\sigma^2_{min}\mathbf{B}_k\right)^{-1}\mathbf{D}_W^{\tau}\mathbf{Y},$$

ce qui achève la preuve de la proposition III.3. □

Preuve de la proposition III.4. — On a

$$\frac{1}{n}\mathbf{D}_W^{\tau}\mathbf{D}_W=\frac{1}{n}\mathbf{D}_X^{\tau}\mathbf{D}_X+\frac{\sigma_\delta^2}{p}\mathbf{B}_k+\mathbf{R}_1,$$

avec

$$\|\mathbf{R}_1\|=O_{\mathbb{P}}\left(\frac{1}{n^{1/2}p^{1/2}k^{1/2}}\right).$$

Preuve: En utilisant le fait que $W_i(t_j)=X_i(t_j)+\delta_{ij}$ pour $i=1,\dots,n$ et $j=1,\dots,p$, on peut écrire

$$\frac{1}{n}\mathbf{D}_W^{\tau}\mathbf{D}_W=\frac{1}{n}\mathbf{D}_X^{\tau}\mathbf{D}_X+\left(\frac{1}{n}\sum_{i=1}^{n}M_{irs}\right)_{r,s=1,\dots,k+q},$$

avec $M_{irs}=\langle B_r,X_i\rangle\langle B_s,\delta_i\rangle+\langle B_r,\delta_i\rangle\langle B_s,X_i\rangle+\langle B_r,\delta_i\rangle\langle B_s,\delta_i\rangle$. Étudions maintenant cette variable aléatoire M_{irs}. Tout d'abord, en utilisant l'indépendance entre X_i et δ_i, on peut écrire

$$\text{(V.3)} \qquad \begin{aligned} \mathbb{E}\left(M_{irs}\right) &= \mathbb{E}\left(\langle B_r,\delta_i\rangle\langle B_s,\delta_i\rangle\right)\\ &= \frac{1}{p^2}\sum_{j_1=1}^{p}\sum_{j_2=1}^{p} B_r(t_{j_1})B_s(t_{j_2})\mathbb{E}\left(\delta_i(t_{j_1})\delta_i(t_{j_2})\right)\\ &= \frac{\sigma_\delta^2}{p^2}\sum_{j=1}^{p} B_r(t_j)B_s(t_j)\\ &= \frac{\sigma_\delta^2}{p}\langle B_r,B_s\rangle. \end{aligned}$$

D'autre part, on a

$$\text{(V.4)} \qquad \begin{aligned} \mathbb{E}\left(M_{irs}^2\right) &= \mathbb{E}\left(\langle B_r,X_i\rangle^2\right)\mathbb{E}\left(\langle B_s,\delta_i\rangle^2\right)\\ &\quad+\mathbb{E}\left(\langle B_r,\delta_i\rangle^2\right)\mathbb{E}\left(\langle B_s,X_i\rangle^2\right)\\ &\quad+\mathbb{E}\left(\langle B_r,\delta_i\rangle^2\langle B_s,\delta_i\rangle^2\right)\\ &\quad+2\mathbb{E}\left(\langle B_r,X_i\rangle\langle B_s,X_i\rangle\right)\mathbb{E}\left(\langle B_r,\delta_i\rangle\langle B_s,\delta_i\rangle\right). \end{aligned}$$

En utilisant des résultats techniques sur les B-splines (voir Cardot, 2000), on note que

$$\text{(V.5)} \qquad \left|\mathbb{E}\left(\langle B_r,\delta_i\rangle\langle B_s,\delta_i\rangle\right)\right| = \left|\frac{\sigma_{\delta_i}^2}{p^2}\sum_{j=1}^{p} B_r(t_j)B_s(t_j)\right| = O\left(\frac{1}{pk}\right),$$

et, avec l'hypothèse $(B.0)$,

$$\text{(V.6)} \qquad \left|\mathbb{E}\left(\langle B_r,X_i\rangle\langle B_s,X_i\rangle\right)\right| = O\left(\frac{1}{k^2}\right).$$

Avec le même type de calculs, on a aussi

$$\begin{aligned}
&\mathbb{E}\left(\langle B_r,\delta_i\rangle^2\langle B_s,\delta_i\rangle^2\right)\\
=\;&\frac{1}{p^4}\sum_{j_1=1}^{p}\sum_{j_2=1}^{p}\sum_{j_3=1}^{p}\sum_{j_4=1}^{p}B_r(t_{j_1})B_r(t_{j_2})B_s(t_{j_3})B_s(t_{j_4})\mathbb{E}\left(\delta_{ij_1}\delta_{ij_2}\delta_{ij_3}\delta_{ij_4}\right)\\
=\;&\frac{1}{p^4}\sum_{j=1}^{p}B_r(t_j)^2B_s(t_j)^2\mathbb{E}\left(\delta_{ij}^4\right)\\
&+\frac{1}{p^4}\sum_{j_1=1}^{p}\sum_{j_2=1}^{p}B_r(t_{j_1})B_r(t_{j_2})B_s(t_{j_1})B_s(t_{j_2})\mathbb{E}\left(\delta_{ij_1}^2\right)\mathbb{E}\left(\delta_{ij_2}^2\right),
\end{aligned}$$

d'où

$$\mathbb{E}\left(\langle B_r,\delta_i\rangle^2\langle B_s,\delta_i\rangle^2\right)=O\left(\frac{1}{p^4}\left[\sum_{j=1}^{p}B_r(t_j)B_s(t_j)\right]^2\right),$$

ce qui donne finalement

$$\mathbb{E}(\langle B_r,\delta_i\rangle^2\langle B_s,\delta_i\rangle^2)=O\left(\frac{1}{p^2k^2}\right). \tag{V.7}$$

Maintenant, avec (V.5), (V.6) et (V.7), la relation (V.4) devient

$$\mathbb{E}\left(M_{irs}^2\right)=O\left(\frac{1}{pk^3}\right)+O\left(\frac{1}{p^2k^2}\right),$$

soit, en prenant $p>k$,

$$\mathbb{E}\left(M_{irs}^2\right)=O\left(\frac{1}{pk^3}\right). \tag{V.8}$$

On peut maintenant conclure la preuve de la proposition III.4. En utilisant (V.3) et (V.8), on a donc

$$\frac{1}{n}\sum_{i=1}^{n} M_{irs} = \frac{\sigma_\delta^2}{p}\langle B_r, B_s\rangle + O_{\mathbb{P}}\left(\frac{1}{n^{1/2}p^{1/2}k^{3/2}}\right).$$

C'est donc là qu'apparaissent les matrices $\mathbf{B}_k$ and $\mathbf{R}_1$: il existe une matrice $\mathbf{R}_1$ telle que

$$\left(\frac{1}{n}\sum_{i=1}^{n} M_{irs}\right)_{r,s=1,\dots,k+q} = \frac{\sigma_\delta^2}{p}\mathbf{B}_k + \mathbf{R}_1,$$

avec, pour $r, s = 1, \dots, k+q$, $R_{1_{rs}} = O_{\mathbb{P}}\left(\frac{1}{n^{1/2}p^{1/2}k^{3/2}}\right)$, soit, avec le théorème 1.19 de Chatelin (1983),

$$\|\mathbf{R}_1\| = O_{\mathbb{P}}\left(\frac{1}{n^{1/2}p^{1/2}k^{1/2}}\right),$$

ce qui termine la preuve de la proposition III.4. $\square$

V.2. INTÉGRALE DU CARRÉ DE LA RÉGRESSION - PREUVES

Preuve de la proposition III.9. — On a

$$\mathbb{E}(\widehat{\theta}) - \theta = \frac{\mu_2(K)}{2}\left(\int_0^1 r''(x)\alpha(x)dx\right)h^2 + o(h^2).$$

Preuve:

$$\begin{aligned}
\text{(V.9)} \qquad & \mathbb{E}(\widehat{\theta}) - \theta \\
= \ & \sum_{i=1}^n \mathbb{E}(Y_i)\int_0^1 \frac{K\left(\frac{x_i-x}{h}\right)\alpha(x)}{\sum_{i=1}^n K\left(\frac{x_i-x}{h}\right)}dx - \int_0^1 r(x)\alpha(x)dx \\
= \ & \int_0^1 \frac{\sum_{i=1}^n r(x_i)K\left(\frac{x_i-x}{h}\right) - r(x)\sum_{i=1}^n K\left(\frac{x_i-x}{h}\right)}{\sum_{i=1}^n K\left(\frac{x_i-x}{h}\right)}\alpha(x)dx.
\end{aligned}$$

Calculons d'abord la somme

$$\sum_{i=1}^n K\left(\frac{x_i-x}{h}\right).$$

On utilise l'approximation d'une somme par une intégrale, ce qui donne, avec le changement de variable $s = \frac{u-x}{h}$,

$$\text{(V.10)} \qquad \begin{aligned} \sum_{i=1}^{n} K\left(\frac{x_i - x}{h}\right) &= n\int_{x-h}^{x+h} K\left(\frac{u-x}{h}\right) du + o(nh) \\ &= nh\int_{-1}^{1} K(s)ds + o(nh) \\ &= nh + o(nh). \end{aligned}$$

Calculons maintenant

$$\sum_{i=1}^{n} r(x_i)K\left(\frac{x_i - x}{h}\right) - r(x)\sum_{i=1}^{n} K\left(\frac{x_i - x}{h}\right).$$

On a, par un développement limité,

$$\begin{aligned} &\sum_{i=1}^{n} r(x_i)K\left(\frac{x_i - x}{h}\right) - r(x)\sum_{i=1}^{n} K\left(\frac{x_i - x}{h}\right) \\ = &\left[r'(x)\sum_{i=1}^{n}(x_i - x)K\left(\frac{x_i - x}{h}\right) + \frac{r''(x)}{2}\sum_{i=1}^{n}(x_i - x)^2 K\left(\frac{x_i - x}{h}\right)\right. \\ &\left. + o\left(\sum_{i=1}^{n}(x_i - x)^2 K\left(\frac{x_i - x}{h}\right)\right)\right], \end{aligned}$$

ce qui donne, en utilisant le fait que K est un noyau d'ordre 2 et une approximation par une intégrale de $\sum_{i=1}^{n}(x_i - x)^2 K\left(\frac{x_i - x}{h}\right)$,

$$\text{(V.11)} \qquad \begin{aligned} &\sum_{i=1}^{n} r(x_i)K\left(\frac{x_i - x}{h}\right) - r(x)\sum_{i=1}^{n} K\left(\frac{x_i - x}{h}\right) \\ = &\frac{\mu_2(K)}{2} r''(x) nh^3 + o\left(nh^3\right). \end{aligned}$$

En revenant à la relation (V.9) et en utilisant les résultats (V.10) et (V.11), on obtient

$$\mathbb{E}(\widehat{\theta}) - \theta = \int_0^1 \frac{\frac{\mu_2(K)}{2} r''(x) nh^3 + o(nh^3)}{nh + o\,(nh)} dx,$$

ce qui donne le biais. □

Preuve de la proposition III.10. — On a

$$\mathbb{V}\left(\widehat{\theta}\right) = 2\sigma^2 \left(\int_0^1 \alpha(x)^2 dx\right) \left(\int_0^2 \Psi(z) dz\right) \frac{1}{n} + o\left(\frac{1}{n}\right).$$

Preuve: D'après la définition de $\widehat{\theta}$, on a

$$\begin{aligned} \widehat{\theta}^2 \quad = \quad & \sum_{i=1}^n \sum_{j=1}^n Y_i Y_j \left(\int_0^1 \frac{K\left(\frac{x_i - x}{h}\right) \alpha(x)}{\sum_{i=1}^n K\left(\frac{x_i - x}{h}\right)} dx\right) \\ & \times \left(\int_0^1 \frac{K\left(\frac{x_j - y}{h}\right) \alpha(y)}{\sum_{i=1}^n K\left(\frac{x_i - y}{h}\right)} dy\right), \end{aligned}$$

ce qui donne, vu que $\mathbb{E}(Y_i^2) = r(x_i)^2 + \sigma^2$,

$$\begin{aligned} \text{(V.12)} \quad \mathbb{V}\left(\widehat{\theta}\right) &= \mathbb{E}\left(\widehat{\theta}^2\right) - \mathbb{E}\left(\widehat{\theta}\right)^2 \\ &= \sum_{i=1}^n \left[\mathbb{E}(Y_i^2) - r(x_i)^2\right] \left(\int_0^1 \frac{K\left(\frac{x_i - x}{h}\right) \alpha(x)}{\sum_{i=1}^n K\left(\frac{x_i - x}{h}\right)} dx\right) \\ &\qquad \times \left(\int_0^1 \frac{K\left(\frac{x_i - y}{h}\right) \alpha(y)}{\sum_{i=1}^n K\left(\frac{x_i - y}{h}\right)} dy\right) \\ &= \sigma^2 \sum_{i=1}^n \left(\int_0^1 \frac{K\left(\frac{x_i - x}{h}\right) \alpha(x)}{\sum_{i=1}^n K\left(\frac{x_i - x}{h}\right)} dx\right) \left(\int_0^1 \frac{K\left(\frac{x_i - y}{h}\right) \alpha(y)}{\sum_{i=1}^n K\left(\frac{x_i - y}{h}\right)} dy\right) \\ &= \sigma^2 \int_0^1 \int_0^1 \frac{1}{S(x) S(y)} \left[\sum_{i=1}^n K\left(\frac{x_i - x}{h}\right) K\left(\frac{x_i - y}{h}\right)\right] \alpha(x) \alpha(y) dx dy, \end{aligned}$$

en notant $S(x) = \sum_{i=1}^n K\left(\frac{x_i-x}{h}\right)$. Si on calcule la somme à l'aide d'une approximation par une intégrale, on a

$$\begin{aligned}&\sum_{i=1}^n K\left(\frac{x_i-x}{h}\right)K\left(\frac{x_i-y}{h}\right)\\ =\ & n\int_{[x-h,x+h]\cup[y-h,y+h]} K\left(\frac{u-x}{h}\right)K\left(\frac{u-y}{h}\right)du + o(nh).\end{aligned}$$

Considérons le cas $x \le y$ (le cas $x \ge y$ est analogue). En posant $s = \frac{u-x}{h}$, ceci donne

$$\begin{aligned}&\sum_{i=1}^n K\left(\frac{x_i-x}{h}\right)K\left(\frac{x_i-y}{h}\right)\\ =\ & nh\int_{[-1,1]\cup[-1+\frac{y-x}{h},1+\frac{y-x}{h}]} K(s)K\left(s-\frac{y-x}{h}\right)ds + o(nh).\end{aligned}$$

Remarquons que les quantités $K\left(\frac{u-x}{h}\right)$ et $K\left(\frac{u-y}{h}\right)du$ ont des supports non disjoints uniquement pour y compris entre $x-2h$ et $x+2h$, soit $\frac{x-y}{h}$ compris entre -2 et 2. On introduit alors la fonction Ψ définie par

$$\Psi(a) = \int_{-1+a}^{1} K(s)K(s-a)ds,$$

pour tout $a \in [0,2]$, et on déduit alors

$$\sum_{i=1}^n K\left(\frac{x_i-x}{h}\right)K\left(\frac{x_i-y}{h}\right) = nh\Psi\left(\frac{y-x}{h}\right) + o(nh).$$

En revenant alors à (V.12) et en posant $z = \frac{y-x}{h}$, il vient (en utilisant aussi (V.10))

$$\mathbb{V}\left(\widehat{\theta}\right) = 2\sigma^2 \int_0^1 \frac{nh^2}{[nh + o(nh)]^2} \left(\int_0^2 \Psi(z)\alpha(x+hz)dz \right) \alpha(x)dx,$$

ce qui donne la variance. □

Preuve de la proposition III.11. — On a

$$\mathbb{E}(\widehat{\theta}) - \theta = \frac{\mu_2(K)}{2} \left(\int_0^1 \left[r(x)s''(x) + r''(x)s(x) \right] dx \right) h^2 + o(h^2).$$

Preuve: On a, avec les notations introduites précédemment

$$\begin{aligned}
& \mathbb{E}(\widehat{\theta}) - \theta \\
= \; & \sum_{i=1}^n \sum_{j=1}^n \mathbb{E}(Y_i)\mathbb{E}(Z_j) \int_0^1 \frac{K\left(\frac{x_i-x}{h}\right) K\left(\frac{x_j-x}{h}\right)}{\sum_{i=1}^n \sum_{j=1}^n K\left(\frac{x_i-x}{h}\right) K\left(\frac{x_j-x}{h}\right)} dx - \int_0^1 r(x)s(x)dx \\
= \; & \int_0^1 \frac{\left[\sum_{i=1}^n r(x_i)K\left(\frac{x_i-x}{h}\right)\right] \left[\sum_{j=1}^n s(x_j)K\left(\frac{x_j-x}{h}\right) - s(x)\sum_{j=1}^n K\left(\frac{x_j-x}{h}\right)\right]}{S(x)^2} dx \\
& + \int_0^1 \frac{\left[\sum_{i=1}^n r(x_i)K\left(\frac{x_i-x}{h}\right) - r(x)\sum_{i=1}^n K\left(\frac{x_i-x}{h}\right)\right] \left[s(x)\sum_{j=1}^n K\left(\frac{x_j-x}{h}\right)\right]}{S(x)^2} dx.
\end{aligned}$$

Avec ce qui a été fait précédemment pour établir les relations (V.10) et (V.11), on a aussi

$$\sum_{i=1}^n r(x_i)K\left(\frac{x_i-x}{h}\right) = r(x)nh + o\left(nh\right),$$

$$\sum_{j=1}^n s(x_j)K\left(\frac{x_j-x}{h}\right) - s(x)\sum_{j=1}^n K\left(\frac{x_j-x}{h}\right) = \frac{\mu_2(K)}{2}s''(x)nh^3 + o\left(nh^3\right),$$

$$\sum_{i=1}^{n} r(x_i)K\left(\frac{x_i-x}{h}\right) - r(x)\sum_{i=1}^{n} K\left(\frac{x_i-x}{h}\right) = \frac{\mu_2(K)}{2}r''(x)nh^3 + o\left(nh^3\right),$$

$$s(x)\sum_{j=1}^{n} K\left(\frac{x_j-x}{h}\right) = s(x)nh + o\left(nh\right),$$

ce qui donne, en revenant à la relation de départ

$$\begin{aligned}&\mathbb{E}(\widehat{\theta}) - \theta\\ = &\int_0^1 \frac{\frac{\mu_2(K)}{2}r(x)s''(x)n^2h^4 + \frac{\mu_2(K)}{2}r''(x)s(x)n^2h^4 + o\left(n^2h^4\right)}{\left[nh + o(nh)\right]^2}dx,\end{aligned}$$

ce qui achève la preuve du biais pour ce cas 2. $\square$

Preuve de la proposition III.12. — On a

$$\mathbb{V}\left(\widehat{\theta}\right) = 2\left[\tau^2\left(\int_0^1 r(x)^2dx\right) + \sigma^2\left(\int_0^1 s(x)^2dx\right)\right]\left(\int_0^2 \Psi(z)dz\right)\frac{1}{n} + o\left(\frac{1}{n}\right).$$

Preuve: Par un calcul analogue à celui du calcul de la variance dans le cas 1, et on a

$$\begin{aligned}\widehat{\theta}^2 = &\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\sum_{l=1}^{n} Y_iZ_jY_kZ_l\left(\int_0^1 \frac{K\left(\frac{x_i-x}{h}\right)K\left(\frac{x_j-x}{h}\right)}{\sum_{i=1}^{n}\sum_{j=1}^{n}K\left(\frac{x_i-x}{h}\right)K\left(\frac{x_j-x}{h}\right)}dx\right)\\ &\times\left(\int_0^1 \frac{K\left(\frac{x_k-y}{h}\right)K\left(\frac{x_l-y}{h}\right)}{\sum_{k=1}^{n}\sum_{l=1}^{n}K\left(\frac{x_k-y}{h}\right)K\left(\frac{x_l-y}{h}\right)}dy\right).\end{aligned}$$

Les différents cas de figure de la quadruple somme sont (a) $i = k$, $j = l$, (b) $i = k$, $j \neq l$, (c) $i \neq k$, $j = l$ et (d) $i \neq k$, $j \neq l$. En utilisant $\mathbb{E}(Y_i^2) = r(x_i)^2 + \sigma^2$ et $\mathbb{E}(Z_j^2) = s(x_j)^2 + \tau^2$, il vient

$$
\begin{aligned}
&\mathbb{V}\left(\widehat{\theta}\right) = \mathbb{E}\left(\widehat{\theta}^2\right) - \mathbb{E}\left(\widehat{\theta}\right)^2 \\
= \;& \sum_{i=1}^{n}\sum_{j=1}^{n}\left[\sigma^2 s(x_j)^2 + \tau^2 r(x_j)^2 + \sigma^2\tau^2\right]\left(\int_0^1 \frac{K\left(\frac{x_i-x}{h}\right)K\left(\frac{x_j-x}{h}\right)}{S(x)^2}dx\right) \\
&\times\left(\int_0^1 \frac{K\left(\frac{x_i-y}{h}\right)K\left(\frac{x_j-y}{h}\right)}{S(y)^2}dy\right) \\
&+\tau^2\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{\substack{k=1\\k\neq i}}^{n} r(x_i)r(x_k)\left(\int_0^1 \frac{K\left(\frac{x_i-x}{h}\right)K\left(\frac{x_j-x}{h}\right)}{S(x)^2}dx\right) \\
&\times\left(\int_0^1 \frac{K\left(\frac{x_k-y}{h}\right)K\left(\frac{x_j-y}{h}\right)}{S(y)^2}dy\right) \\
&+\sigma^2\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{\substack{l=1\\l\neq j}}^{n} s(x_j)s(x_l)\left(\int_0^1 \frac{K\left(\frac{x_i-x}{h}\right)K\left(\frac{x_j-x}{h}\right)}{S(x)^2}dx\right) \\
&\times\left(\int_0^1 \frac{K\left(\frac{x_i-y}{h}\right)K\left(\frac{x_l-y}{h}\right)}{S(y)^2}dy\right) \\
= \;& \sigma^2\int_0^1\!\!\int_0^1 \frac{\left[\sum_{i=1}^n K\left(\frac{x_i-x}{h}\right)K\left(\frac{x_i-y}{h}\right)\right]\left[\sum_{j=1}^n s(x_j)^2 K\left(\frac{x_j-x}{h}\right)K\left(\frac{x_j-y}{h}\right)\right]}{S(x)^2S(y)^2}dxdy \\
&+\tau^2\int_0^1\!\!\int_0^1 \frac{\left[\sum_{i=1}^n r(x_i)^2 K\left(\frac{x_i-x}{h}\right)K\left(\frac{x_i-y}{h}\right)\right]\left[\sum_{j=1}^n K\left(\frac{x_j-x}{h}\right)K\left(\frac{x_j-y}{h}\right)\right]}{S(x)^2S(y)^2}dxdy \\
&+\sigma^2\tau^2\int_0^1\!\!\int_0^1 \frac{\left[\sum_{i=1}^n K\left(\frac{x_i-x}{h}\right)K\left(\frac{x_i-y}{h}\right)\right]\left[\sum_{j=1}^n K\left(\frac{x_j-x}{h}\right)K\left(\frac{x_j-y}{h}\right)\right]}{S(x)^2S(y)^2}dxdy \\
&+\tau^2\int_0^1\!\!\int_0^1 \frac{\left[\sum_{i=1}^n\sum_{\substack{k=1\\k\neq i}}^n r(x_i)r(x_k)K\left(\frac{x_i-x}{h}\right)K\left(\frac{x_k-y}{h}\right)\right]\left[\sum_{j=1}^n K\left(\frac{x_j-x}{h}\right)K\left(\frac{x_j-y}{h}\right)\right]}{S(x)^2S(y)^2}dxdy \\
&+\sigma^2\int_0^1\!\!\int_0^1 \frac{\left[\sum_{i=1}^n K\left(\frac{x_i-x}{h}\right)K\left(\frac{x_i-y}{h}\right)\right]\left[\sum_{j=1}^n\sum_{\substack{l=1\\l\neq j}}^n s(x_j)s(x_l)K\left(\frac{x_j-x}{h}\right)K\left(\frac{x_l-y}{h}\right)\right]}{S(x)^2S(y)^2}dxdy
\end{aligned}
$$

Or, on a vu au calcul de la variance dans le cas 1 que

$$\sum_{i=1}^{n} K\left(\frac{x_i - x}{h}\right) K\left(\frac{x_i - y}{h}\right) = nh\Psi\left(\frac{y-x}{h}\right) + o(nh).$$

Par des calculs analogues, on montrerait aussi que

$$\sum_{j=1}^{n} s(x_j)^2 K\left(\frac{x_j - x}{h}\right) K\left(\frac{x_j - y}{h}\right) = nhs(x)\Psi\left(\frac{y-x}{h}\right) + o(nh),$$

et on remarque également que

$$\sum_{i=1}^{n} \sum_{\substack{k=1 \\ k \neq i}}^{n} r(x_i) r(x_k) K\left(\frac{x_i - x}{h}\right) K\left(\frac{x_k - y}{h}\right) = n^2 h^2 r(x) r(y) + o\left(n^2 h^2\right).$$

Ainsi, en revenant à notre calcul de variance, on obtient finalement (après changement de variable $z = \frac{y-x}{h}$)

$$\begin{aligned} & \mathbb{V}\left(\widehat{\theta}\right) \\ = \; & \sigma^2 \times O\left(\frac{1}{n^2 h}\right) + \tau^2 \times O\left(\frac{1}{n^2 h}\right) + \sigma^2 \tau^2 \times O\left(\frac{1}{n^2 h}\right) \\ & +2\left[\tau^2 \left(\int_0^1 r(x)^2 dx\right) + \sigma^2 \left(\int_0^1 s(x)^2 dx\right)\right] \left(\int_0^2 \Psi(z) dz\right) \frac{1}{n} + o\left(\frac{1}{n}\right), \end{aligned}$$

ce qui donne la variance annoncée, en prenant n et h tels que $nh \longrightarrow +\infty$. $\square$

Preuve de la proposition III.13. — On a

$$\begin{aligned}\mathbb{E}(\widehat{\theta})-\theta &= \mu_2(K)\left(\int_0^1 r(x)r''(x)dx\right)h^2\\ &\quad -R(K)\left(\int_0^1 r(x)^2dx\right)\frac{1}{nh}+o\left(h^2+\frac{1}{nh}\right).\end{aligned}$$

Preuve:

$$\begin{aligned}&\mathbb{E}(\widehat{\theta})-\theta\\ =\ & \sum_{i=1}^n\sum_{\substack{j=1\\ j\neq i}}^n r(x_i)r(x_j)\int_0^1\frac{K\left(\frac{x_i-x}{h}\right)K\left(\frac{x_j-x}{h}\right)}{\left[\sum_{i=1}^n K\left(\frac{x_i-x}{h}\right)\right]^2}dx-\int_0^1 r(x)^2dx\\ =\ & \int_0^1\left[\sum_{i=1}^n\sum_{\substack{j=1\\ j\neq i}}^n r(x_i)r(x_j)\frac{K\left(\frac{x_i-x}{h}\right)K\left(\frac{x_j-x}{h}\right)}{\left[\sum_{i=1}^n K\left(\frac{x_i-x}{h}\right)\right]^2}-r(x)^2\right]dx\\ =\ & \int_0^1\frac{\left[\sum_{i=1}^n r(x_i)K\left(\frac{x_i-x}{h}\right)\right]^2-\sum_{i=1}^n r(x_i)^2K\left(\frac{x_i-x}{h}\right)^2-\left[\sum_{i=1}^n r(x)K\left(\frac{x_i-x}{h}\right)\right]^2}{\left[\sum_{i=1}^n K\left(\frac{x_i-x}{h}\right)\right]^2}dx.\end{aligned}$$

Calculons d'abord la somme

$$\sum_{i=1}^n r(x_i)^2K\left(\frac{x_i-x}{h}\right)^2.$$

En procédant de façon analogue aux calculs précédents (approximation d'une somme par une intégrale, même changement de variable, on écrit de plus $r(x+hs)=r(x)+o(1)$), on obtient

$$\text{(V.13)}\qquad \sum_{i=1}^n r(x_i)^2K\left(\frac{x_i-x}{h}\right)^2=R(K)r(x)^2nh+o(nh).$$

Calculons à présent

$$\left[\sum_{i=1}^{n} r(x_i)K\left(\frac{x_i-x}{h}\right)\right]^2 - \left[\sum_{i=1}^{n} r(x)K\left(\frac{x_i-x}{h}\right)\right]^2.$$

On a, comme précédemment,

$$\begin{aligned}
& \left[\sum_{i=1}^{n} r(x_i)K\left(\frac{x_i-x}{h}\right)\right]^2 - \left[\sum_{i=1}^{n} r(x)K\left(\frac{x_i-x}{h}\right)\right]^2 \\
= & \left[\sum_{i=1}^{n} r(x_i)K\left(\frac{x_i-x}{h}\right) - \sum_{i=1}^{n} r(x)K\left(\frac{x_i-x}{h}\right)\right] \\
& \times \left[\sum_{i=1}^{n} r(x_i)K\left(\frac{x_i-x}{h}\right) + \sum_{i=1}^{n} r(x)K\left(\frac{x_i-x}{h}\right)\right] \\
= & \left[r'(x)\sum_{i=1}^{n}(x_i-x)K\left(\frac{x_i-x}{h}\right) + \frac{r''(x)}{2}\sum_{i=1}^{n}(x_i-x)^2K\left(\frac{x_i-x}{h}\right)\right. \\
& \left. \qquad + o\left(\sum_{i=1}^{n}(x_i-x)^2K\left(\frac{x_i-x}{h}\right)\right)\right] \\
& \times \left[2r(x)\sum_{i=1}^{n} K\left(\frac{x_i-x}{h}\right) + o\left(\sum_{i=1}^{n} K\left(\frac{x_i-x}{h}\right)\right)\right].
\end{aligned}$$

On utilise alors (V.10), le fait que K est un noyau d'ordre 2 et une approximation par une intégrale de $\sum_{i=1}^{n}(x_i-x)^2K\left(\frac{x_i-x}{h}\right)$, et on obtient

$$\begin{aligned}
& \left[\sum_{i=1}^{n} r(x_i)K\left(\frac{x_i-x}{h}\right)\right]^2 - \left[\sum_{i=1}^{n} r(x)K\left(\frac{x_i-x}{h}\right)\right]^2 \\
= & \left[\frac{\mu_2(K)}{2}r''(x)nh^3 + o(nh^3)\right] \times [2r(x)nh + o(nh)] \\
= & \mu_2(K)r(x)r''(x)n^2h^4 + o(n^2h^4).
\end{aligned} \tag{V.14}$$

En revenant au calcul du biais et en utilisant les résultats (V.10), (V.12) et (V.14), on obtient

$$\mathbb{E}(\widehat{\theta}) - \theta = \int_0^1 \frac{\mu_2(K)r(x)r''(x)n^2h^4 + o(n^2h^4) - R(K)r(x)^2nh + o(nh)}{n^2h^2 + o(n^2h^2)}dx,$$

ce qui donne le biais annoncé. $\square$

Preuve de la proposition III.14. — On a

$$\mathbb{V}(\widehat{\theta}) = 8\sigma^2 \left(\int_0^1 r(x)^2 dx\right)\left(\int_0^2 \Psi(z)dz\right)\frac{1}{n} + o\left(\frac{1}{n}\right).$$

Preuve: D'après la définition de $\widehat{\theta}$, on a

$$\widehat{\theta}^2 = \sum_{i=1}^n \sum_{\substack{j=1 \\ j\neq i}}^n \sum_{k=1}^n \sum_{\substack{l=1 \\ l\neq k}}^n Y_iY_jY_kY_l \left(\int_0^1 \frac{K\left(\frac{x_i-x}{h}\right)K\left(\frac{x_j-x}{h}\right)}{\left[\sum_{i=1}^n K\left(\frac{x_i-x}{h}\right)\right]^2}dx\right)$$
$$\times \left(\int_0^1 \frac{K\left(\frac{x_k-y}{h}\right)K\left(\frac{x_l-y}{h}\right)}{\left[\sum_{k=1}^n K\left(\frac{x_k-x}{h}\right)\right]^2}dy\right).$$

On commence par chercher les différents cas de figure pour la quadruple somme ci-dessus. On répertorie ces cas ci-dessous.

- $k = i \quad \begin{cases} l = j \rightarrow Y_i^2Y_j^2 \rightarrow n(n-1) \\ l \neq j \rightarrow Y_i^2Y_jY_l \rightarrow n(n-1)(n-2) \end{cases}$
- $k \neq i \quad \begin{cases} k = j & \begin{cases} l = i \rightarrow Y_i^2Y_j^2 \rightarrow n(n-1) \\ l \neq i \rightarrow Y_iY_j^2Y_l \rightarrow n(n-1)(n-2) \end{cases} \\ & \\ k \neq j & \begin{cases} l = i \rightarrow Y_i^2Y_jY_k \rightarrow n(n-1)(n-2) \\ l = j \rightarrow Y_iY_j^2Y_k \rightarrow n(n-1)(n-2) \\ l \neq i,j \rightarrow Y_iY_jY_kY_l \rightarrow n(n-1)(n-2)(n-3) \end{cases} \end{cases}$

Comme $Y_i^2 = r(x_i)^2 + 2r(x_i)\epsilon_i + \epsilon_i^2$, on a $\mathbb{E}\left(Y_i^2\right) = r(x_i)^2 + \sigma^2$, et on obtient donc dans les cas de figures suivants

$$\mathbb{E}\left(Y_i^2 Y_j^2\right) = r(x_i)^2 r(x_j)^2 + \sigma^2\left[r(x_i)^2 + r(x_j)^2\right] + \sigma^4,$$

pour $j \neq i$,

$$\mathbb{E}\left(Y_i^2 Y_j Y_k\right) = r(x_i)^2 r(x_j) r(x_k) + \sigma^2 r(x_j) r(x_k),$$

pour $j \neq i$ et $k \neq i, j$,

$$\mathbb{E}\left(Y_i Y_j Y_k Y_l\right) = r(x_i) r(x_j) r(x_k) r(x_l),$$

pour $j \neq i$, $k \neq i, j$ et $l \neq i, j, k$. On déduit donc de ceci

$$\begin{aligned}
\text{(V.15)} \qquad \mathbb{V}(\widehat{\theta}) &= \mathbb{E}(\widehat{\theta}^2) - \mathbb{E}(\widehat{\theta})^2 \\
&= \sum_{i=1}^n \sum_{\substack{j=1 \\ j\neq i}}^n \sum_{k=1}^n \sum_{\substack{l=1 \\ l\neq k}}^n \left[\mathbb{E}(Y_i Y_j Y_k Y_l) - r(x_i) r(x_j) r(x_k) r(x_l)\right] \\
&\quad \times \left(\int_0^1 \frac{K\left(\frac{x_i - x}{h}\right) K\left(\frac{x_j - x}{h}\right)}{\left[\sum_{i=1}^n K\left(\frac{x_i - x}{h}\right)\right]^2} dx\right) \times \left(\int_0^1 \frac{K\left(\frac{x_k - y}{h}\right) K\left(\frac{x_l - y}{h}\right)}{\left[\sum_{k=1}^n K\left(\frac{x_k - y}{h}\right)\right]^2} dy\right) \\
&= 2 \sum_{i=1}^n \sum_{\substack{j=1 \\ j\neq i}}^n \left(\sigma^2\left[r(x_i)^2 + r(x_j)^2\right] + \sigma^4\right) \\
&\quad \times \left(\int_0^1 \frac{K\left(\frac{x_i - x}{h}\right) K\left(\frac{x_j - x}{h}\right)}{\left[\sum_{i=1}^n K\left(\frac{x_i - x}{h}\right)\right]^2} dx\right) \times \left(\int_0^1 \frac{K\left(\frac{x_i - y}{h}\right) K\left(\frac{x_j - y}{h}\right)}{\left[\sum_{i=1}^n K\left(\frac{x_i - y}{h}\right)\right]^2} dy\right) \\
&\quad + 4 \sum_{i=1}^n \sum_{\substack{j=1 \\ j\neq i}}^n \sum_{\substack{k=1 \\ k\neq i,j}}^n \sigma^2 r(x_j) r(x_k) \\
&\quad \times \left(\int_0^1 \frac{K\left(\frac{x_i - x}{h}\right) K\left(\frac{x_j - x}{h}\right)}{\left[\sum_{i=1}^n K\left(\frac{x_i - x}{h}\right)\right]^2} dx\right) \times \left(\int_0^1 \frac{K\left(\frac{x_i - y}{h}\right) K\left(\frac{x_k - y}{h}\right)}{\left[\sum_{k=1}^n K\left(\frac{x_k - y}{h}\right)\right]^2} dy\right).
\end{aligned}$$

Le premier terme de la double somme de (V.15) s'écrit

$$
\text{(V.16)} \quad \sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n} r(x_i)^2 \left(\int_0^1 \frac{K\left(\frac{x_i-x}{h}\right)K\left(\frac{x_j-x}{h}\right)}{\left[\sum_{i=1}^{n} K\left(\frac{x_i-x}{h}\right)\right]^2}dx\right) \times \left(\int_0^1 \frac{K\left(\frac{x_i-y}{h}\right)K\left(\frac{x_j-y}{h}\right)}{\left[\sum_{i=1}^{n} K\left(\frac{x_i-y}{h}\right)\right]^2}dy\right)
$$

$$
\begin{aligned}
= \;& \int_0^1\!\!\int_0^1 \frac{1}{S(x)^2S(y)^2}\left[\sum_{i=1}^{n}\sum_{j=1}^{n} r(x_i)^2 K\left(\frac{x_i-x}{h}\right)K\left(\frac{x_j-x}{h}\right)\right.\\
&\qquad \left.\times K\left(\frac{x_i-y}{h}\right)K\left(\frac{x_j-y}{h}\right)\right] dxdy\\
&- \int_0^1\!\!\int_0^1 \frac{1}{S(x)^2S(y)^2}\left[\sum_{i=1}^{n} r(x_i)^2 K\left(\frac{x_i-x}{h}\right)^2 K\left(\frac{x_i-y}{h}\right)^2\right] dxdy.
\end{aligned}
$$

Or, en utilisant à nouveau l'approximation d'une somme par une intégrale (double cette fois), on a

$$
\begin{aligned}
&\sum_{i=1}^{n}\sum_{j=1}^{n} r(x_i)^2 K\left(\frac{x_i-x}{h}\right)K\left(\frac{x_j-x}{h}\right)K\left(\frac{x_i-y}{h}\right)K\left(\frac{x_j-y}{h}\right)\\
= \;& n^2 \iint_{([x-h,x+h]\cup[y-h,y+h])^2} r(u)^2 K\left(\frac{u-x}{h}\right)K\left(\frac{v-x}{h}\right)\\
&\qquad \times K\left(\frac{u-y}{h}\right)K\left(\frac{v-y}{h}\right) dudv + o(n^2h^2)\\
= \;& n^2 \int_{[x-h,x+h]\cup[y-h,y+h]} r(u)^2 K\left(\frac{u-x}{h}\right)K\left(\frac{u-y}{h}\right) du\\
&\qquad \times \int_{[x-h,x+h]\cup[y-h,y+h]} K\left(\frac{v-x}{h}\right)K\left(\frac{v-y}{h}\right) dv + o(n^2h^2).
\end{aligned}
$$

Considérons le cas $x \leq y$ (le cas $x \geq y$ est analogue). En posant $s = \frac{u-x}{h}$ et $t = \frac{v-x}{h}$, ceci donne

$$
\begin{aligned}
&\sum_{i=1}^{n}\sum_{j=1}^{n} r(x_i)^2 K\left(\frac{x_i-x}{h}\right) K\left(\frac{x_j-x}{h}\right) K\left(\frac{x_i-y}{h}\right) K\left(\frac{x_j-y}{h}\right)\\
= \;& n^2h^2 \int_{[-1,1]\cup[-1+\frac{y-x}{h},1+\frac{y-x}{h}]} r(x+hs)^2 K(s) K\left(s-\frac{y-x}{h}\right) ds\\
&\qquad \times \int_{[-1,1]\cup[-1+\frac{y-x}{h},1+\frac{y-x}{h}]} K(t) K\left(t-\frac{y-x}{h}\right) dt + o(n^2h^2)\\
= \;& n^2h^2 \int_{-1+\frac{y-x}{h}}^{1} r(x+hs)^2 K(s) K\left(s-\frac{y-x}{h}\right) ds\\
&\qquad\qquad \times \int_{-1+\frac{y-x}{h}}^{1} K(t) K\left(t-\frac{y-x}{h}\right) dt + o(n^2h^2).
\end{aligned}
$$

Remarquons que, par exemple dans $\int r(u)^2 K\left(\frac{u-x}{h}\right) K\left(\frac{u-y}{h}\right) du$, les quantités $K\left(\frac{u-x}{h}\right)$ et $K\left(\frac{u-y}{h}\right) du$ ont des supports non disjoints uniquement pour y compris entre $x-2h$ et $x+2h$, soit $\frac{x-y}{h}$ compris entre -2 et 2. On déduit alors, en utilisant la fonction Ψ vue dans le cas 1

$$
\begin{aligned}
&\sum_{i=1}^{n}\sum_{j=1}^{n} r(x_i)^2 K\left(\frac{x_i-x}{h}\right) K\left(\frac{x_j-x}{h}\right) K\left(\frac{x_i-y}{h}\right) K\left(\frac{x_j-y}{h}\right)\\
= \;& n^2h^2 \left[r(x)+o(1)\right]^2 \Psi\left(\frac{y-x}{h}\right)^2 + o(n^2h^2).
\end{aligned}
$$

En utilisant ce résultat, on revient au calcul de (V.16) qui donne alors

$$\begin{aligned}
&\int_0^1\int_0^1 \frac{1}{S(x)^2S(y)^2}\left[\sum_{i=1}^n\sum_{j=1}^n r(x_i)^2 K\left(\frac{x_i-x}{h}\right)K\left(\frac{x_j-x}{h}\right)\right.\\
&\qquad\qquad\left.\times K\left(\frac{x_i-y}{h}\right)K\left(\frac{x_j-y}{h}\right)\right]dxdy\\
=\;& 2\int_0^1\int_{x-2h}^{x+2h}\frac{n^2h^2\left[r(x)+o(1)\right]^2}{\left[nh+o(nh)\right]^4}\Psi\left(\frac{y-x}{h}\right)^2 dxdy\\
&\qquad\qquad +o\left(\frac{n^2h^3}{\left[nh+o(nh)\right]^4}\right),
\end{aligned}$$

d'où, en posant $z=\frac{y-x}{h}$ et en supposant que $nh\to+\infty$

$$\begin{aligned}
&\int_0^1\int_0^1 \frac{1}{S(x)^2S(y)^2}\left[\sum_{i=1}^n\sum_{j=1}^n r(x_i)^2 K\left(\frac{x_i-x}{h}\right)K\left(\frac{x_j-x}{h}\right)\right.\\
&\qquad\qquad\left.\times K\left(\frac{x_i-y}{h}\right)K\left(\frac{x_j-y}{h}\right)\right]dxdy\\
=\;& 2\left(\int_0^1 r(x)^2dx\right)\left(\int_0^2\Psi(z)^2dz\right)\frac{1}{n^2h}+o\left(\frac{1}{n^2h}\right)=o\left(\frac{1}{n}\right).
\end{aligned}$$

Par un calcul très analogue au précédent, on a aussi

$$\int_0^1\int_0^1 \frac{1}{S(x)^2S(y)^2}\left[\sum_{i=1}^n r(x_i)^2K\left(\frac{x_i-x}{h}\right)^2K\left(\frac{x_i-y}{h}\right)^2\right]dxdy=o\left(\frac{1}{n}\right).$$

Ainsi, en revenant à la relation (V.16), on a finalement

$$
\text{(V.17)} \qquad \sum_{i=1}^{n}\sum_{\substack{j=1\\j\neq i}}^{n} r(x_i)^2 \left(\int_0^1 \frac{K\left(\frac{x_i-x}{h}\right)K\left(\frac{x_j-x}{h}\right)}{\left[\sum_{i=1}^n K\left(\frac{x_i-x}{h}\right)\right]^2}dx\right) \times \left(\int_0^1 \frac{K\left(\frac{x_i-y}{h}\right)K\left(\frac{x_j-y}{h}\right)}{\left[\sum_{i=1}^n K\left(\frac{x_i-y}{h}\right)\right]^2}dy\right) = o\left(\frac{1}{n}\right).
$$

Passons maintemant au calcul de la somme triple dans (V.15). On a

$$
\begin{aligned}
\text{(V.18)} \sum_{i=1}^{n}\sum_{\substack{j=1\\j\neq i}}^{n}\sum_{\substack{k=1\\k\neq i,j}}^{n} r(x_j)r(x_k) &\left(\int_0^1 \frac{K\left(\frac{x_i-x}{h}\right)K\left(\frac{x_j-x}{h}\right)}{\left[\sum_{i=1}^n K\left(\frac{x_i-x}{h}\right)\right]^2}dx\right) \\
&\times \left(\int_0^1 \frac{K\left(\frac{x_k-y}{h}\right)K\left(\frac{x_l-y}{h}\right)}{\left[\sum_{k=1}^n K\left(\frac{x_k-x}{h}\right)\right]^2}dy\right) \\
= \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}[\,] - \underbrace{\sum_{i=1}^{n}\sum_{\substack{j=1\\j\neq i}}^{n}[\,]}_{\text{cas } k=i} - \underbrace{\sum_{i=1}^{n}\sum_{\substack{j=1\\j\neq i}}^{n}[\,]}_{\text{cas } k=j} &- \underbrace{\sum_{i=1}^{n}\sum_{\substack{k=1\\k\neq i}}^{n}[\,]}_{\text{cas } j=i} - \underbrace{\sum_{i=1}^{n}[\,]}_{\text{cas } k=j=i}.
\end{aligned}
$$

Dans cette expression, par des calculs analogues aux précédents, les quatre dernières sommes donnent des $o\left(\frac{1}{n}\right)$. Il reste à calculer la première somme, et là aussi, un calcul similaire (en approximant la triple somme par une intégrale triple) conduit à

$$
\begin{aligned}
&\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n} r(x_j)r(x_k)K\left(\frac{x_i-x}{h}\right)K\left(\frac{x_j-x}{h}\right)K\left(\frac{x_i-y}{h}\right)K\left(\frac{x_k-y}{h}\right) \\
= \;& n^3\left(\int_{[x-h,x+h]\cup[y-h,y+h]} K\left(\frac{u-x}{h}\right)K\left(\frac{u-y}{h}\right)du\right) \\
&\times \left(\int_{x-h}^{x+h} r(v)K\left(\frac{v-x}{h}\right)dv\right)\left(\int_{y-h}^{y+h} r(w)K\left(\frac{w-y}{h}\right)dw\right) + o(n^3h^3).
\end{aligned}
$$

Ceci donne, pour $x \leq y$ et en posant $s = \frac{u-x}{h}$, $t = \frac{v-x}{h}$,

$$
\begin{aligned}
&\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n} r(x_j)r(x_k)K\left(\frac{x_i-x}{h}\right)K\left(\frac{x_j-x}{h}\right)K\left(\frac{x_i-y}{h}\right)K\left(\frac{x_k-y}{h}\right)\\
= \;& n^3h^3\Psi\left(\frac{y-x}{h}\right)\left(\int_{-1}^{1} r(x+ht)K(t)dt\right)\left(\int_{-1}^{1} r(y+hz)K(z)dz\right)+o(n^3h^3)\\
= \;& n^3h^3\Psi\left(\frac{y-x}{h}\right)\left[r(x)+o(1)\right]\left[r(y)+o(1)\right]+o(n^3h^3).
\end{aligned}
$$

En posant $z=\frac{y-x}{h}$, on obtient finalement en revenant à (V.18)

$$
\begin{aligned}
\text{(V.19)}\quad & \sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n}\sum_{\substack{k=1\\ k\neq i,j}}^{n} r(x_j)r(x_k)\left(\int_0^1 \frac{K\left(\frac{x_i-x}{h}\right)K\left(\frac{x_j-x}{h}\right)}{\left[\sum_{i=1}^{n}K\left(\frac{x_i-x}{h}\right)\right]^2}dx\right)\\
&\qquad\times\left(\int_0^1 \frac{K\left(\frac{x_k-y}{h}\right)K\left(\frac{x_l-y}{h}\right)}{\left[\sum_{k=1}^{n}K\left(\frac{x_k-x}{h}\right)\right]^2}dy\right)\\
= \;& 2\int_0^1 \frac{n^3h^4\left[r(x)+o(1)\right]^2}{\left[nh+o(nh)\right]^4}\left(\int_0^2\Psi(z)dz\right)dx+o\left(\frac{n^3h^4}{\left[nh+o(nh)\right]^4}\right)\\
= \;& 2\left(\int_0^1 r(x)^2dx\right)\left(\int_0^2\Psi(z)dz\right)\frac{1}{n}+o\left(\frac{1}{n}\right).
\end{aligned}
$$

En revenant finalement à l'expression (V.15) de la variance et en utilisant les relations (V.17) et (V.19), on obtient le résultat. □

V.3. RÉGRESSION SUR COMPOSANTES PRINCIPALES - PREUVES

Preuve de la proposition III.15. — Pour tous $i_1, i_2 = 1, \ldots, n$, si on prend $h_{i_1 i_2}$ de la forme $p^{-\zeta}$ avec $\zeta \in [1/4, 1/2[$, on a

$$\widehat{M}_{i_1 i_2} - M_{i_1 i_2} = O_{\mathbb{P}}\left(\frac{1}{np^{1/2}}\right).$$

Preuve: On va utiliser les résultats prouvés dans la section III.4.2. Dans le cas où $i_1 \neq i_2$, on a, d'après les propositions III.11 et III.12,

$$\begin{aligned} & \mathbb{E}\left(\widehat{M}_{i_1 i_2} - M_{i_1 i_2}\right) \\ = \; & \frac{1}{n}\left\{\frac{\mu_2(K)}{2}\left(\int_0^1 \left[X_{i1}(t)X''_{i2}(t) + X''_{i1}(t)X_{i2}(t)\right] dt\right) h^2_{i_1 i_2} + o\left(h^2_{i_1 i_2}\right)\right\}, \end{aligned}$$

et

$$\begin{aligned} & \mathbb{V}\left(\widehat{M}_{i_1 i_2} - M_{i_1 i_2}\right) \\ = \; & \frac{1}{n^2}\left\{2\left[\sigma^2_\delta \int_0^1 \left[X_{i1}(t)^2 + X_{i2}(t)^2\right] dt\right]\left(\int_0^2 \Psi(z)dz\right)\frac{1}{p} + o\left(\frac{1}{p}\right)\right\}. \end{aligned}$$

De même, dans le cas où $i_1 = i_2 = i$, on a, d'après les propositions III.13 et III.14,

$$\mathbb{E}\left(\widehat{M}_{ii} - M_{ii}\right) = \frac{1}{n}\left\{\mu_2(K)\left(\int_0^1 X_i(t)X_i''(t)dt\right)h_{ii}^2 + o\left(h_{ii}^2\right)\right\},$$

et

$$\mathbb{V}\left(\widehat{M}_{ii} - M_{ii}\right) = \frac{1}{n^2}\left\{4\sigma_\delta^2\left(\int_0^1 X_i(t)^2dt\right)\left(\int_0^2 \Psi(z)dz\right)\frac{1}{p} + o\left(\frac{1}{p}\right)\right\}.$$

On en déduit, pour tous $i_1, i_2 = 1, \ldots, n$,

$$\mathbb{E}\left(\widehat{M}_{i_1i_2} - M_{i_1i_2}\right) = O\left(\frac{h_{i_1i_2}^2}{n}\right),$$

et

$$\mathbb{V}\left(\widehat{M}_{i_1i_2}\right) = O\left(\frac{1}{n^2p}\right).$$

Ainsi, en prenant $h_{i_1i_2}$ de la forme $p^{-\zeta}$ avec $\zeta \in [1/4, 1/2[$, on s'assure d'un biais négligeable, et le résultat de la proposition III.15 est immédiat. $\square$

Preuve de la proposition III.16. — On a

$$\left\|\widehat{\mathbf{M}} - \mathbf{M}\right\|^2 = O_{\mathbb{P}}\left(\frac{1}{p}\right).$$

Preuve: On a

$$\left\|\widehat{\mathbf{M}} - \mathbf{M}\right\|^2 \leq \mathrm{Tr}\left[\left(\widehat{\mathbf{M}} - \mathbf{M}\right)^\tau\left(\widehat{\mathbf{M}} - \mathbf{M}\right)\right] = \sum_{i_1=1}^n\sum_{i_2=1}^n\left(\widehat{M}_{i_1i_2} - M_{i_1i_2}\right)^2,$$

et le résultat est alors immédiat en utilisant la proposition III.15. $\square$

Preuve de la proposition III.17. — Sous les hypothèses de la proposition III.17, on a, pour tout $r = 1, \ldots, L$,

$$\widehat{\lambda}_r - \lambda_r = O_{\mathbb{P}}\left(\frac{1}{n^{1/2}p^{1/2}} + \frac{1}{p}\right).$$

Preuve: On commence par donner un résultat provenant de Kneip et Utikal (2001). On a

$$\widehat{\lambda}_r - \lambda_r = \mathrm{Tr}\left[\mathbf{P}_{\mathcal{E}_r}\left(\widehat{\mathbf{M}} - \mathbf{M}\right)\right] + R_1,$$

où $\mathbf{P}_{\mathcal{E}_r}$ désigne la matrice de projection sur le sous-espace propre associé à la $r^{\text{ème}}$ valeur propre λ_r, et R_1 vérifie

$$|R_1| \le \frac{6\left\|\widehat{\mathbf{M}} - \mathbf{M}\right\|^2}{\min_{s=1,\ldots,n, s\neq r}|\lambda_r - \lambda_s|}.$$

En utilisant la proposition III.16 et l'hypothèse de cette proposition concernant les valeurs propres de $\mathbf{M}$, on en déduit que

$$R_1 = O_{\mathbb{P}}\left(\frac{1}{p}\right),$$

ce qui donne

$$\widehat{\lambda}_r - \lambda_r = \boldsymbol{p}_r^{\tau}\left(\widehat{\mathbf{M}} - \mathbf{M}\right)\boldsymbol{p}_r + O_{\mathbb{P}}\left(\frac{1}{p}\right).$$

D'autre part, on a

$$
\begin{aligned}
&\mathbb{E}\left[\left(\boldsymbol{p}_r^\tau\left(\widehat{\mathbf{M}}-\mathbf{M}\right)\boldsymbol{p}_r\right)^2\right]\\
= \;& \sum_{i_1=1}^n\sum_{i_2=1}^n\sum_{i_3=1}^n\sum_{i_4=1}^n p_{i_1r}p_{i_2r}p_{i_3r}p_{i_4r}\mathbb{E}\left[\left(\widehat{M}_{i_1i_2}-M_{i_1i_2}\right)\left(\widehat{M}_{i_3i_4}-M_{i_3i_4}\right)\right]\\
= \;& \sum_{i_1=1}^n\sum_{i_2=1}^n\sum_{i_3=1}^n p_{i_1r}^2p_{i_2r}p_{i_3r}\mathbb{E}\left[\left(\widehat{M}_{i_1i_2}-M_{i_1i_2}\right)\left(\widehat{M}_{i_1i_3}-M_{i_1i_3}\right)\right]\\
\leq \;& \sum_{i_1=1}^n p_{i_1r}^2\sum_{i_2=1}^n |p_{i_2r}|\sum_{i_3=1}^n |p_{i_3r}|\times O\left(\frac{1}{n^2p}\right),
\end{aligned}
$$

ce qui donne finalement, vu que $\sum_{i_1=1}^n p_{i_1r}^2=1$ et que $\sum_{i_2=1}^n |p_{i_2r}|=O\left(n^{1/2}\right)$,

$$\boldsymbol{p}_r^\tau\left(\widehat{\mathbf{M}}-\mathbf{M}\right)\boldsymbol{p}_r=O_{\mathbb{P}}\left(\frac{1}{n^{1/2}p^{1/2}}\right),$$

et achève la preuve de la proposition III.17. $\square$

Preuve de la proposition III.18. — Sous les hypothèses de la proposition III.17, on a

$$\|\widehat{\boldsymbol{p}}_r-\boldsymbol{p}_r\|=O_{\mathbb{P}}\left(\frac{1}{p^{1/2}}\right).$$

Preuve: On commence là aussi par énoncer un résultat provenant de Kneip et Utikal (2001). On a

$$\widehat{\boldsymbol{p}}_r-\boldsymbol{p}_r=-\mathbf{S}_r\left(\widehat{\mathbf{M}}-\mathbf{M}\right)\boldsymbol{p}_r+\mathbf{R}_2,$$

où $\mathbf{S}_r$ est la matrice définie par

$$\mathbf{S}_r=\sum_{s\neq r}\frac{1}{\lambda_s-\lambda_r}\mathbf{P}_{\mathcal{E}_s},$$

et $\mathbf{R}_2$ vérifie

$$\|\mathbf{R}_2\| \leq \frac{6\left\|\widehat{\mathbf{M}} - \mathbf{M}\right\|^2}{\min_{\lambda \neq \lambda_r} |\lambda - \lambda_r|^2}.$$

En utilisant la proposition III.16 et l'hypothèse de cette proposition concernant les valeurs propres de $\mathbf{M}$, on en déduit que

$$\|\mathbf{R}_2\| = O\left(\frac{1}{p}\right).$$

En posant $\boldsymbol{q}_r = \left(\widehat{\mathbf{M}} - \mathbf{M}\right)\boldsymbol{p}_r$, on obtient donc

$$\|\widehat{\boldsymbol{p}}_r - \boldsymbol{p}_r + \mathbf{S}_r\boldsymbol{q}_r\| = O_{\mathbb{P}}\left(\frac{1}{p}\right).$$

Calculons maintenant $\|\mathbf{S}_r\|$. On a

$$\begin{aligned}
\|\mathbf{S}_r\| &= \sup_{\|\boldsymbol{v}\|=1}\left[\boldsymbol{v}^\tau\left(\sum_{s_1\neq r}\frac{1}{\lambda_{s_1}-\lambda_r}\boldsymbol{p}_{s_1}\boldsymbol{p}_{s_1}^\tau\right)\left(\sum_{s_2\neq r}\frac{1}{\lambda_{s_2}-\lambda_r}\boldsymbol{p}_{s_2}\boldsymbol{p}_{s_2}^\tau\right)\boldsymbol{v}\right]^{1/2}\\
&\leq \frac{1}{\ell_r}\sup_{\|\boldsymbol{v}\|=1}\left[\boldsymbol{v}^\tau\left(\sum_{s_1\neq r}\boldsymbol{p}_{s_1}\boldsymbol{p}_{s_1}^\tau\right)\left(\sum_{s_2\neq r}\boldsymbol{p}_{s_2}\boldsymbol{p}_{s_2}^\tau\right)\boldsymbol{v}\right]^{1/2},
\end{aligned}$$

où $\ell_r = \min\{|\lambda_{r-1}-\lambda_r|, |\lambda_{r+1}-\lambda_r|\}$. Comme le suprémum ci-dessus n'est autre que $\left\|\sum_{s\neq r}\boldsymbol{p}_s\boldsymbol{p}_s^\tau\right\|$, on en déduit qu'il est inférieur à $\|\sum\boldsymbol{p}_s\boldsymbol{p}_s^\tau\| = 1$, d'où

$$\|\mathbf{S}_r\| \leq \frac{1}{\ell_r}.$$

D'après les hypothèses faites sur les λ_r, on en déduit alors que

$$\|\mathbf{S}_r\| = O\left(1\right).$$

Pour tout $\boldsymbol{v} \in \mathbb{R}^n$, on a alors

$$\begin{aligned}
\mathbb{E}\left[\left(\boldsymbol{v}^\tau \mathbf{S}_r \boldsymbol{q}_r\right)^2\right] &= \boldsymbol{v}^\tau \mathbf{S}_r \mathbb{E}\left(\boldsymbol{q}_r \boldsymbol{q}_r^\tau\right) \mathbf{S}_r \boldsymbol{v} \\
&\leq \|\boldsymbol{v}\|^2 \|\mathbf{S}_r\|^2 \mathbb{E}\left[\boldsymbol{p}_r^\tau \left(\widehat{\mathbf{M}} - \mathbf{M}\right)^\tau \left(\widehat{\mathbf{M}} - \mathbf{M}\right) \boldsymbol{p}_r\right].
\end{aligned}$$

D'après ce qui précède, on obtient ainsi

$$\mathbb{E}\left[\left(\boldsymbol{v}^\tau \mathbf{S}_r \boldsymbol{q}_r\right)^2\right] = O\left(\frac{\|\boldsymbol{v}\|^2}{np}\right).$$

Finalement, en prenant comme vecteur $\boldsymbol{v}$ le vecteur dont toutes les composantes sont nulles sauf la $i^{\text{ème}}$ qui vaut 1, il vient

$$\mathbb{E}\left[\left(\mathbf{S}_r \boldsymbol{q}_r\right)_i^2\right] = O\left(\frac{1}{np}\right),$$

d'où

$$\left|\widehat{p}_{ir} - p_{ir}\right| = O_{\mathbb{P}}\left(\frac{1}{n^{1/2} p^{1/2}}\right),$$

ce qui prouve le résultat de la proposition III.18. $\square$

BIBLIOGRAPHIE

[1] Aneiros-Perez, G., Cardot, H., Estevez-Perez, G. and Vieu, P. (2004). Maximum ozone concentration forecasting by functional nonparametric approaches. *Environmetrics*, **15**, 675-685.

[2] Averous, J. and Meste, M. (1997). Median balls : an extension of the interquartile intervals to multivariate distributions. *Journal of Multivariate Analysis*, **63**, 222-241.

[3] Bassett, G. and Koenker, R. (1978). Asymptotic theory of least absolute error regression. *Journal of the American Statistical Association*, **73**, 618-622.

[4] Benhenni, K. and Cambanis, S. (1992). Sampling designs for estimating integrals of stochastic processes. *Annals of Statistics*, **20**, 161-194.

[5] Benko, M., Härdle, W. and Kneip, A. (2005). Common functional principal components. *SFB 649 Economic Risk Discussion Paper*, 2006-010.

[6] Berlinet, A., Biau, G. and Rouvière, L. (2005). Functional classification with wavelets. *Preprint*.

[7] Berlinet, A., Cadre, B. and Gannoun, A. (2001). On the conditional L_1-median and its estimation. *Journal of Nonparametric Statistics*, **13**, 631-645.

[8] Besse, P., Cardot, H. and Ferraty, F. (1997). Simultaneous nonparametric regression of unbalanced longitudinal data. *Computational Statistics and Data Analysis*, **24**, 255-270.

[9] Besse, P., Cardot, H. and Stephenson, D. (2000). Autoregressive forecasting of some functional climatic variations. *Scandinavian Journal of Statistics*, **27**, 673-687.

[10] Besse, P. and Ramsay, J.O. (1986). Principal components analysis of sampled functions. *Psychometrika*, **51**, 285-311.

[11] Bhattacharya, P.K. and Gangopadhyay, A.K. (1990). Kernel and nearest-neighbor estimation of a conditional quantile. *Annals of Statistics*, **18**, 1400-1415.

[12] Bosq, D. (2000). *Linear processes in function spaces.* Lecture Notes in Statistics, **149**, Springer.

[13] Breiman, L., Friedman, J.H., Olshen, R.A. and Stone, C.J. (1984). *Classification and regression trees.* Wadsworth Statistics and Probability Series, Wadsworth Advanced Books and Software, Belmont.

[14] Cadre, B. (2001). Convergent estimators for the L_1-median of a Banach valued random variable. *Statistics*, **35**, 509-521.

[15] Cardot, H. (2000). Nonparametric estimation of smoothed principal components analysis of sampled noisy functions. *Nonparametric Statistics*, **12**, 503-538.

[16] Cardot, H. (2006). Conditional functional principal components analysis. *Scandinavian Journal of Statistics*, to appear.

[17] Cardot, H., Crambes, C., Kneip, A. and Sarda, P. (2006). Smoothing splines estimators in functional linear regression with errors-in-variables. *Computational Statistics and Data Analysis*, special issue on functional data analysis, to appear.

[18] Cardot, H., Crambes, C. and Sarda, P. (2004a). Estimation spline de quantiles conditionnels pour variables explicatives fonctionnelles. *Comptes Rendus de l'Académie des Sciences*, **339**, 141-144.

[19] Cardot, H., Crambes, C. and Sarda, P. (2004b). Conditional quantiles with functional covariates : an application to ozone pollution forecasting. In *Compstat 2004 Proceedings*, J. Antoch editor, Physica-Verlag, 769-776.

[20] Cardot, H., Crambes, C. and Sarda, P. (2005). Quantile regression when the covariates are functions. *Journal of Nonparametric Statistics*, **17**, 841-856.

[21] Cardot, H., Crambes, C. and Sarda, P. (2006). Conditional quantiles with functional covariates : an application to ozone pollution forecasting. In *Applied Biostatistics : Case Studies and Interdisciplinary Methods*, Xplore e-book, to appear.

[22] Cardot, H., Ferraty, F. and Sarda, P. (1999). Functional linear model. *Statistic and Probability Letters*, **45**, 11-22.

[23] Cardot, H., Ferraty, F. and Sarda, P. (2003). Spline estimators for the functional linear model. *Statistica Sinica*, **13**, 571-591.

[24] Cardot, H. and Sarda, P. (2005). Estimation in generalized linear models for functional data via penalized likelihood. *Journal of Multivariate Analysis*, **92**, 24-41.

[25] Chatelin, F. (1983). *Spectral approximation of linear operators.* Academic Press, New-York.

[26] Chiou, J-M., Müller, H.-G., Wang, J-L and Carey, J.R. (2003). A functional multiplicative effects model for longitudinal data, with application to reproductive histories of female medflies. *Statistica Sinica*, **13**, 1119-1133.

[27] Chiou, J-M., Müller, H.-G. and Wang, J-L. (2003). Functional quasi-likelihood regression models with smooth random effects. *Journal of the Royal Statistical Society, Series B*, **65**, 405-423.

[28] Chiou, J-M., Müller, H.-G. and Wang, J-L. (2004). Functional response models. *Statistica Sinica*, **14**, 675-693.

[29] Cohen, A. (2003). *Numerical analysis of wavelets methods.* Elsevier, Amsterdam.

[30] Crambes, C. (2005). Total least squares for functional data. Invited paper in *ASMDA 2005 Proceedings*, 619-626.

[31] Cuevas, A., Febrero, M. and Fraiman, R. (2002). Linear functional regression : the case of a fixed design and functional response. *Canadian Journal of Statistics*, **30**, 285-300.

[32] Damon, J. and Guillas, S. (2002). The inclusion of exogenous variables in functional autoregressive ozone forecasting. *Environmetrics*, **13**, 759-774.

[33] Daubechies, I. (1992). *Ten lectures on wavelets.* SIAM, Philadelphia.

[34] Dauxois, J. et Pousse, A. (1976). *Les analyses factorielles en calcul des probabilités et en statistique : essai d'étude synthétique*. Thèse de doctorat, Université Paul Sabatier, Toulouse.

[35] Dauxois, J., Pousse, A. and Romain, Y. (1982). Asymptotic theory for the principal component analysis of a random vector function : some applications to statistical inference. *Journal of Multivariate Analysis*, **12**, 136-154.

[36] de Boor, C. (1978). *A practical guide to splines*. Applied Mathematical Sciences, Springer, New York

[37] Demmel, J. (1992). The componentwise distance to the nearest singular matrix. *SIAM, Journal of Matrix Analysis and Applications*, **13**, 10-19.

[38] Deville, J.-C. (1974). Méthodes statistiques et numériques de l'analyse harmonique. *Annales de l'I.N.S.E.E.*, **15**, 3-101.

[39] Devroye, L., Györfi, L. and Lugosi, G. (1996). *A probabilistic theory of pattern recognition*. Applications of Mathematics, Springer, New York.

[40] Dierckx, P. (1993). *Curve and surface fitting with splines*. Monographs on Numerical Analysis, Oxford Science Publications, New York.

[41] Ducharme, G., Gannoun, A., Guertin, M.-C. and Jéquier, J.-C. (1995). Reference values obtained by kernel-based estimation of quantile regression. *Biometrics*, **51**, 1105-1116.

[42] Dunford, N. and Schwarz, J. (1963). *Linear operators*. Interscience, New York.

[43] Eubank, R.L. (1988). *Spline smoothing and nonparametric regression*. Marcel Dekker.

[44] Fan, J., Hu, T.-C. and Truong, Y (1994). Robust nonparametric function estimation. *Scandinavian Journal of Statistics*, **21**, 433-446.

[45] Ferraty, F., Rahbi, A. and Vieu, P. (2005). Conditional quantiles for functional dependent data with application to the climatic El Nino phenomenon. *Sankhya*, **67**, 378-399.

[46] Ferraty, F. and Vieu, P. (2002). The functional nonparametric model and application to spectrometric data. *Computational Statistics*, **17**, 545-564.

[47] Ferraty, F. and Vieu, P. (2003). Curves discrimination : a nonparametric functional approach. Special issue in honour of Stan Azen : a birthday celebration. *Computational Statistics and Data Analysis*, **44**, 161-173.

[48] Ferraty, F. and Vieu, P. (2006). *Nonparametric functional data analysis : theory and practice*. Springer, New York.

[49] Frank, I.E. and Friedman, J.H. (1993). A statistical view of some chemometrics regression tools. *Technometrics*, **35**, 109-135.

[50] Fuller, W.A. (1987). *Measurement error models*. Wiley, New York.

[51] Gasser, T., Sroka, L. and Jennen-Steinmetz, C. (1986). Residual variance and residual pattern in nonlinear regression. *Biometrika*, **3**, 625-633.

[52] Ghattas, B. (1999). Prévisions des pics d'ozone par arbres de régression, simples et agrégés par bootstrap. *Revue de Statistique Appliquée*, **XLVII**, 61-80.

[53] Gleser, L.J. (1981). Estimation in a multivariate "errors-in-variables" regression model : large sample results. *Annals of Statistics*, **9**, 24-44.

[54] Goberg, I.C. et Krein, M.G. (1971). *Introduction à la théorie des opérateurs linéaires non auto-adjoints dans un espace hilbertien*. Dunod, Paris.

[55] Golub, G.H., Hansen, P.C and O'Leary, D.P. (1999). Tikhonov regularization and total least squares. *SIAM, Journal of Matrix Analysis and Applications*, **21**, 185-194.

[56] Golub, G.H. and Van Loan, C.F. (1980). An analysis of the total least squares problem. *SIAM, Journal of Numerical Analysis*, **17**, 883-893.

[57] Golub, G.H. and Van Loan, C.F. (1996). *Matrix computations*. Johns Hopkins University Press, Baltimore.

[58] Good, I.J. (1969). Some applications of the singular value decomposition of a matrix. *Technometrics*, **11**, 823-831.

[59] Goutis, C. (1998). Second derivative functional regression with applications to near infrared spectroscopy. *Journal of the Royal Statistical Society, Series B*, **60**, 103-114.

[60] Green, P.J. and Silverman, B.W. (1994). *Nonparametric regression and generalized linear models : a roughness penalty approach*. Monographs on Statistics and Applied Probability, Chapman and Hall, London.

[61] Hall, P. and Marron, J.S. (1987). Estimation of integrated squared density derivatives. *Statistics and probability Letters*, **6**, 109-115.

[62] Härdle, W. (1991). *Smoothing techniques with implementation in S.* Springer, New-York.

[63] Hastie, T. and Tibshirani, R. (1990). *Generalized additive models*. Monographs on Statistics and Applied Probability, Chapman and Hall, London.

[64] Hastie, T., Buja, A. and Tibshirani, R. (1995). Penalized discriminant analysis. *Annals of Statistics*, **23**, 73-102.

[65] Hastie, T. and Mallows, C. (1993). Discussion of "A statistical view of some chemometrics regression tools." by Frank, I.E. and Friedman, J.H. *Technometrics*, **35**, 140-143.

[66] He, X. and Shi, P. (1994). Convergence rate of B-spline estimators of nonparametric conditional quantile functions. *Nonparametric Statistics*, **3**, 299-308.

[67] Helland, I.S. (1990). Partial least squares regression and statistical models. *Scandinavian Journal of Statistics*, **17**, 97-114.

[68] Hoerl, A.E. and Kennard, R.W. (1980). Ridge regression : advances, algorithms and applications. *American Journal of Mathematical Management Sciences*, **1**, 5-83.

[69] Huang, L.-S. and Fan, J. (1999). Nonparametric estimation of quadratic regression functionals. *Bernoulli*, **5**, 927-949.

[70] Jones, M.C. and Sheater, S.J. (1991). Using non-stochastic terms to advantage in kernel-based estimation of integrated squared density derivatives. *Statistics and probability Letters*, **11**, 511-514.

[71] Kneip, A., Li, X., Mac Gibbon, K.B. and Ramsay, J.O. (2000). Curve registration by local regression. *Canadian Journal of Statistics*, **28**, 19-29.

[72] Kneip, A. and Utikal, K.J. (2001). Inference for density families using functional principal component analysis. *Journal of the American Statistical Association*, **96**, 519-542.

[73] Koenker, R. (2005). *Quantile regression.* Econometric Society Monographs, Cambridge.

[74] Koenker, R. and Bassett, G. (1978). Regression quantiles. *Econometrica*, **46**, 33-50.

[75] Koenker, R. and Machado, J. (1999). Goodness of fit and related inference processes for quantile regression. *Journal of the American Statistical Association*, **94**, 1296-1310.

[76] Koenker, R., Ng, P. and Portnoy, S. (1994). Quantile smoothing splines. *Biometrika*, **81**, 673-680.

[77] Lejeune, M. and Sarda, P. (1988). Quantile regression : a nonparametric approach. *Computational Statistics and Data Analysis* **6**, 229-239.

[78] Marx, B.D. and Eilers, P.H. (1999). Generalized linear regression on sampled signals and curves : a P-spline approach. *Technometrics*, **41**, 1-13.

[79] McCullagh, P. and Nelder, J.A. (1989). *Generalized linear models* (Second Edition). Monographs on Statistics and Applied Probability, Chapman and Hall, London.

[80] Mosteller, F. and Tukey, J. (1977). *Data analysis and regression : a second course in statistics.* Addison-Wesley, Reading.

[81] Müller, H.-G. (2005). Functional modeling and classification of longitudinal data. *Scandinavian Journal of Statistics*, **32**, 223-240.

[82] Müller, H.-G. and Stadtmüller, U. (2005). Generalized functional linear models. *Annals of Statistics*, **33**, 774-805.

[83] Nadaraya, E.A. (1964). On estimating regression. *Theory of Probability and its Applications*, **10**, 186-190.

[84] Poiraud-Casanova, S. et Thomas-Agnan, C. (1998). Quantiles conditionnels. *Journal de la Société Française de Statistique*, **139**, 31-44.

[85] Ramsay, J.O. (1982). When the data are functions. *Psychometrika*, **47**, 379-396.

[86] Ramsay, J.O. (2000). Functional components of variation in handwriting. *Journal of the American Statistical Association*, **95**, 9-15.

[87] Ramsay, J.O. and Dalzell, C.J. (1991). Some tools for functional data analysis. *Journal of the Royal Statistical Society, Series B*, **53**, 539-572.

[88] Ramsay, J.O. and Li, X. (1998). Curve registration. *Journal of the Royal Statistical Society, Series B*, **60**, 351-363.

[89] Ramsay, J.O. and Silverman, B.W. (1997). *Functional data analysis* (First Edition). Springer, New York.

[90] Ramsay, J.O. and Silverman, B.W. (2002). *Applied functional data analysis*. Springer, New York.

[91] Ramsay, J.O. and Silverman, B.W. (2005). *Functional data analysis* (Second Edition). Springer, New York.

[92] Rao, C.R. (1958). Some statistical methods for comparison of growth curves. *Biometrics*, **14**, 1-17.

[93] Rio, E. (2000). *Théorie asymptotique des processus aléatoires faiblement dépendants*. Springer-Verlag, Berlin.

[94] Ruppert, D. and Caroll, J. (1988). *Transformation and weighting in regression*. Chapman and Hall, New York.

[95] Ruppert, D., Sheater, S.J. and Wand, M.P. (1993). An effective bandwidth selector for local least squares regression. *Working paper*, 93-017.

[96] Sarda, P. and Vieu, P. (2000). Kernel regression. In *Smoothing and Regression : Approches, Computation and Application*, M.G. Schimek editor, Wiley Series in Probability and Statistics, 43-70.

[97] Schumaker, L.L. (1981). *Spline functions : basic theory*. Wiley, New York.

[98] Sima, D.M. and Van Huffel, S. (2004). Appropriate cross validation for regularized error-in-variables linear models. In *Compstat 2004 Proceedings*, J. Antoch editor, Physica-Verlag, 1815-1822.

[99] Stone, C.J. (1982). Optimal rates of convergence for nonparametric models. *Annals of Statistics*, **10**, 1040-1053.

[100] Stone, C.J. (1985). Additive regression and other nonparametric models. *Annals of Statistics*, **13**, 689-705.

[101] Tsybakov, A.B. (1986). Robust reconstruction of functions by the local-approximation method. *Problems of Information Transmission*, **22**, 133-146.

[102] Tucker, L.R. (1958). Determination of parameters of a functional relation by factor analysis. *Psychometrika*, **23**, 19-23.

[103] Uspensky, J.V. (1937). *Introduction to mathematical probability.* McGraw-Hill Book Company, New York.

[104] Utreras, F. (1983). Natural spline functions, their associated eigenvalue problem. *Numerische Mathematik*, **42**, 107-117.

[105] Van Huffel, S. and Vandewalle, J. (1991). *The total least squares problem : computational aspects and analysis.* SIAM, Philadelphia.

[106] Wahba, G. (1990). *Spline models for observational data.* SIAM, Philadelphia.

[107] Watson, G.S. (1964). Smooth regression analysis. *Sankhya, Series A*, **26**, 359-372.

[108] Wedderburn, R.W.M. (1974). Quasi-likelihood functions, generalized linear models, and the Gauss-Newton method. *Biometrika*, **61**, 439-447.

[109] Weinberger, H.F. (1974). *Variational methods for eigenvalue approximation.* SIAM, Philadelphia.

[110] Yao, F., Müller, H.-G. and Wang, J.-L. (2005a). Functional data analysis for sparse longitudinal data. *Journal of the American Statistical Association*, **100**, 577-590.

[111] Yao, F., Müller, H.-G. and Wang, J.-L. (2005b). Functional linear regression analysis for longitudinal data. *Annals of Statistics*, **33**, 2873-2903.