N° d'ordre :

REPUBLIQUE ALGERIENNE DEMOCRATIQUE & POPULAIRE

MINISTERE DE L'ENSEIGNEMENT SUPERIEUR & DE LA RECHERCHE SCIENTIFIQUE

UNIVERSITE DJILLALI LIABES
FACULTE DES SCIENCES EXACTES
SIDI BEL ABBES

THESE DE DOCTORAT

Présentée par :

CHEMIKH Souheyla

Domaine : Mathématiques Informatique

Filière : Mathématiques

Intitulé de la formation : Probabilité et Statistiques appliquées

Intitulée

Régression Trimmed en statistique fonctionnelle

Soutenue le : 22/06/2020

Devant le jury composé de :

Président :

Mr BENAISSA Samir Professeur à L'Université S.B.A

Examinateurs :

Mr RABHI Abbes Professeur à L'Université de S.B.A.

Mr BOUKHARI Fakhr eddine Maitre de Conférences A à L'Université de Tlemecn

Mr MECHAB Boubaker Maitre de Conférences A à L'Université S.B.A.

Directeur de thèse :

Mme LIMAM-BELARBI Faiza Professeur à L'Université S.B.A

Année universitaire : 2019/2020

Aucune investigation humaine ne peut s'intituler véritable science, si elle ne passe par la démonstration mathématique.

Léonard de vinci

Je dédie cette thèse :

A ma chère mère qui a oeuvré pour ma réussite, par son amour, son soutien, tous les sacrifices consentis et ses précieux conseils.

A mon mari ZETILI Noaman, qui a su m'apporter courage et confiance en moi, même lorsque l'avenir paraissait incertain.

A mes frères Ilyes et Hanane.

A toute ma famille et mes meilleurs amis.

REMERCIEMENTS

Je tiens tout d'abord à remercier Dieu le tout puissant et miséricordieux, qui nous a donné la force et la patience d'accomplir ce modeste travail.

je tiens à remercier vivement ma directrice de thèse Professeur **Faiza BELARBI** pour son soutien constant durant toute la période du travail. Vous avez réussi à me transmettre le plaisir de la recherche et à mener à bien ce travail. Encore une fois merci infiniment pour votre disponibilité et vos conseils et pour tout l'intérêt que vous avez accordé à ce travail.

La réalisation de ce travail n'aurait pas été possible sans le soutien du Professeur **Ali LAKSACI** durant toute la période du travail. Il est difficile en quelques mots de vous exprimer toute ma reconnaissance et ma profonde gratitude.

Mes remerciements vont aussi au Professeur **Samir BENAISSA** pour l'honneur qui me fait en présidant le jury.

Je tiens à remercier le Professeur **Abbes RABHI**, Docteur **Fakhr eddine BOUKHARI** et Docteur **Boubaker MECHAB** d'avoir accepté d'examiner cette thèse.

Enfin, j'adresse mes remerciements à l'ensembles des enseignants de la faculté des sciences exactes.

TABLE DES MATIÈRES

RÉSUMÉ

Dans cette thèse, nous nous intéressons au problème de prévision via la régression locale linéaire robuste qui a un comportement insensible et efficace en présence des observations aberrantes. Nous nous intéressons plus particulièrement à l'estimation locale linéaire non paramétrique de la fonction de régression d'une variable réponse réelle sachant une variable explicative fonctionnelle en utilisant l'approche Trimmed. Nos études portent sur des données indépendantes identiquement distribuées (i.i.d.) et ainsi que sur des données dépendantes.
Dans un premier temps, nous considérons que les observations sont indépendantes identiquement distribuées. Nous établissons dans ce cas, la vitesse de convergence presque-complète et la normalité asymptotique d'une famille d'estimateurs robustes basée sur la méthode locale linéaire. Ensuite, nous construisons l'estimateur local linéaire Trimmed de la fonction de régression et nous établissons la vitesse de convergence presque-complète et la normalité asymptotique de l'estimateur construit.
Dans un second temps, nous supposons que les observations sont de type α- mélangeantes, sous certaines conditions, nous établissons les propriétés asymptotiques, telles que la vitesse de convergence presque-complète et la normalité asymptotique de cette famille d'estimateurs locaux linéaires robusts. Afin d'illustrer nos résultats asymptotiques et de montrer la performance de prédiction de nos estimateurs dans le cadre fonctionnel, nous donnerons des exemples sur données simulées.
Autant que l'on sache, le problème de l'estimation non paramétrique de la régression Trimmed pour des données fonctionnelles n'a jamais été abordé. d'où l'originalité et la nouveauté de l'étude menée dans cette thèse.
Mots clés : Données fonctionnelles ; Méthode locale linéaire ; Estimation robuste ; α-mélange ; Prévision non paramétrique ; Convergence presque-complète ; Normalité asymptotique.

ENGLISH SUMMARY

In this thesis, we are interested in the problem of prediction via the robust linear local regression that has an insensitive and effective behavior in the presence of the outliers observations. We are particularly interested in the nonparametric local linear estimation of the regression function of a scalar response variable given a functional explanatory variable by using the Trimmed approach. Our studies focus on identically independent data distributed (i.i.d.) as well as on dependent data.
In a first step, we consider that the observations are independent identically distributed. We establish in this case, the almost complete convergence rate and the asymptotic normality of a family of robust estimators based on the local linear method. Then, we construct the Trimmed local linear estimator of the regression function and we establish the almost complete convergence rate and the asymptotic normality of the constructed estimator.
In a second step, we suppose that the observations are of α- mixing type, under certain conditions, we establish the asymptotic properties, such as, the almost complete convergence rate and the asymptotic normality of this family of robust local linear estimators. In order to illustrate our asymptotic results and to show the prediction performance of our estimators in the functional framework, we will give examples on simulated data.
As far as we know, the problem of nonparametric estimation of Trimmed regression for functional data was never been approached, hence the originality and novelty of the study conducted in this thesis.
Keywords : Functional data ; Local linear method ; Robust estimation ; α- mixing ; Nonparametric prediction ; Almost complete convergence ; Asymptotic normality.

LISTE DES PUBLICATIONS ET COMMUNICATIONS

Publications

1. Belarbi,F., Chemikh,S.,Laksaci, A.(2017).Asymptotic properties of the robust local linear regression with functional regressor. *Tech. report,* https://www.researchgate.net/profile/Ali_Laksaci.
2. Belarbi,F.,Chemikh,S., Laksaci, A.(2018).Local linear estimate of the nonparametric robust regression in function data. *Statistic and Probability letters,* **134**, 128-133.
3. Chemikh,S. and Belarbi,F.(2019a).Local linear estimation of the nonparametric trimmed regression in functional data. *Journal of Statistics & Economics,* **20**,29-41.
4. Chemikh,S., Belarbi,F., Laksaci, A.(2019b).On the local linear modelization of the nonparametric robust regression for functional time series data. En révision.

Communications

1. Chemikh,S. Asymptotic study of the robust estimator of the regression function. ECMISciTech'2017, Octobre 09-12, 2017 à Constantine, Algérie.
2. Chemikh,S. Asymptotic properties of the robust regression for functional data. (HDDA-VIII), Avril 09-13, 2018 à Marrakech, Maroc.
3. Chemikh,S. Nonparametric estimate of robust regression for functional data. CMA, Mai 12-13, 2018 à Boumerdès, Algérie.
4. Chemikh,S. Belarbi,F. On the local linear of the nonparametric robust regression for dependent functional data . RAMA 11, novembre 21-24, 2019 à Sidi Bel Abbès, Algérie.

CHAPITRE 1

INTRODUCTION GÉNÉRALE

1.1 Statistique non paramétrique des données fonctionnelles

1.1.1 Présentation

La statistique tente toujours à répondre aux questions qui se posent sur les phénomènes observés et comment les anticiper. Ce rôle privilégié l'a rendu un champ de recherche important que les statisticiens cherchent à aborder dans différents domaines. Notons que l'objectif principal de la statistique est de fournir une analyse d'un phénomène passé et de prédire un phénomène à venir de nature similaire. Un phénomène courant que la statistique se propose à traiter, consiste à essayer d'expliquer le lien entre une variable d'intérêt Y et une variable explicative X d'un modèle statistique, pour prédire l'une d'elles. D'un point de vue historique, le premier résultat revient à Galileo Galilei (1632), dans un contexte géométrique. Plusieurs modèles sont rencontrés en statistique expriment le lien entre les variables, un des modèles les plus fréquemment rencontrés est la régression, il s'agit d'une méthode d'estimation qui est basé sur l'espérance mathématique conditionnelle qui s'écrit

$$Y = \mathbb{E}(Y|X) + \epsilon, \qquad \text{où } \epsilon \text{ est une variable aléatoire d'erreur.}$$

La régression non paramétrique est un outil de prévision qui exprime la relation entre une variable dépendante Y en fonction d'une ou de plusieurs variables explicatives, en se basant sur l'espérance mathématique conditionnelle, sans spécifier une forme stricte pour cette relation. De façon plus générale, un modèle statistique pour une variable réelle est dite non paramétrique lorsque la classe $\mathcal{C}$ n'est pas indexée par un nombre fini de paramètres réels. D'un point de vue historique, les premiers ouvrages sur l'approche non paramétrique reviennent au 19éme siècles, précisément, au milieu des années 50, cette dernière a connu un développement considérable avec le travail

de Rosenblatt (1956) dans lequel il a proposé une estimation non paramétrique de la fonction de densité par la méthode de noyau, nous citons également le travail de Tukey (1961) et de Parzen (1962) pour le même modèle. Ensuite, des estimateurs de type noyau du modèle de régression ont été introduits, indépendamment, par Nadaraya et Watson (1964). La généralisation à l'estimation non paramétrique locale polynomiale de la fonction de régression a été vue par Cleveland et Develin (1988).
Le problème de prévision est l'une des applications importantes de l'estimation non paramétrique. C'est ce qui explique son grand succès qui s'est développé considérablement avec l'étude des modèles non paramétrique pour des variables dites fonctionnelles, c'est-à-dire, des variables aléatoires, qui sont observées dans plusieurs points d'un intervalle, à valeur dans un espace de dimension infini (appelé espace fonctionnel). Par exemple, la variable fonctionnelle $X = \{X(t); t \in T\}$ avec $T \subset \mathbb{R}$ représente une courbe observée sur l'intervalle T de $\mathbb{R}$. Il est important de noter que la notion de variable fonctionnelle couvre une zone plus grande que l'analyse des courbes. En particulier, une variable fonctionnelle peut être une surface aléatoire, une image, où un vecteur de courbes et dans ces cas $T \subset \mathbb{R}^2$, ou tout autre objet mathématique de dimension infinie plus compliqué. Il s'agit, donc, d'un nouveau champ de la statistique, qui est consacré à l'étude de données fonctionnelles. Ce dernier a suscité un fort engouement au début des années quatre-vingt, sous l'impulsion, notamment, des travaux de Grenander (1981), Dauxois et al. (1982) et Ramsay (1982). Il a été popularisé par Ramsay et Silverman (1997). puis par d'autres ouvrages, Nous citons Bosq (2000), Ramsay et Silverman (2002 ,2005), et Ferraty et Vieu (2006).

1.1.2 Modélisation non paramétrique fonctionnelle

La modélisation statistique des variables fonctionnelles connaît un grand essor chez de nombreux auteurs et dans différents domaines, la preuve de ce succès est les nombreux ouvrages qui sont publiés sur ce sujet et les potentialités d'applications sur les problèmes pratiques. Il est à noter que la modélisation statistique fonctionnelle est liée à l'étude des observations qui ne sont pas réelles ou vectorielles mais des courbes, des surfaces ou des images. Ceci revient aux progrès technologique, notamment en ce qui concerne les outils informatiques et leurs capacités de stockage qui aident à enregistrer des données de plus en plus volumineuses. Dans ces dernières décennies, une énorme innovation en termes d'applications est apparue afin de traiter ce type de données dans différentes disciplines de science appliquée telles que, l'environnement, la médecine, la biométrique, l'économétrie, l'imagerie, l'agro-alimentaire et dans d'autres domaines, et c'est l'une des raisons qui justifie la grand réussite de cette branche statistique qui est devenue le centre d'intérêt de plusieurs études, tant sur le plan pratique que sur le plan théorique.
Signalons également que grâce au développement qu'a vu l'outil informatique dans la façon de récolter les données, d'autres alternatives ont été élaborées afin de surmonter

cette difficulté en étudiant ce type de données dans sa propre dimension, c'est-à-dire, on peut conserver le caractère fonctionnel. Historiquement, depuis les années soixante, plusieurs statisticiens se sont intéressés à traiter les observations sous forme de trajectoires, nous pouvons citer leurs travaux les plus populaires tels que, Obukhov(1960), Holmstrom (1963) en climatologie, Deville (1974) en démographie, Molenaar et Boomsma (1987) puis Kirkpatrick et Heckman (1989) en génétique. Mais cette thématique n'a connu une vraie importance qu'après la publication des monographies de Ramsay et Silvermann (2005) pour les modèles paramétriques et Ferraty et Vieu (2006) pour les modèles non paramétriques. D'où un énorme travail de recherche considérable a été dédié à cette thématique.
En dimension finie, il est bien connu que les vitesses de convergences obtenues pour certains modèles non paramétriques sont optimales mais elles n'en restent pas moins relativement mauvaises dès que la dimension grandit. Cette notion est fortement liée à la rareté des données dans un espace à grande dimension, nous nous référons à Friedman et Stuetzle (1981) et à Huber (1985) pour plus de discussions sur ce problème. En effet, cette rareté des données impose dans la pratique un choix de la largeur de fenêtre $h = h_n$ d'autant plus grand que la dimension est grande, pour stabiliser la variance de l'estimateur, mais cela conduit à un estimateur fortement biaisé. Il est important de noter que ce problème lié uniquement à la nature des modèles purement non paramétriques et n'est absolument pas est lié à la technique d'estimation par noyau, nous nous référons aux ouvrages récents de Gu (2000) et Nychka (2000) pour une description du comportement des estimateurs de type Spline face à ce problème de dimension, et à celui de Wu (2000) pour les estimateurs de type polynomes locaux. Nous renvoyons aussi au travail de de Härdle et Müller (2000).
Rappelons que dans les espaces à dimension infinie, les variables aléatoires admettent toujours la fonction de densité contrairement au cas de la dimension infinie où il n'y a pas de mesure de Lebesgue, dans ce cas, les variables aléatoires fonctionnelles ne possèdent pas toujours la fonction de densité comme dans le cas réel ou encore dans le cas vectoriel, donc une alternative à ce problème est la mesure de probabilités des petites boules qui est une écriture probabiliste qui permet d'agir directement sur la distribution de la variable aléatoire plutôt que sur la densité comme dans le cas de dimension finie. Notons que la propriété de concentration (la mesure de probabilités des petites boules) permet de proposer une alternative au problème de fléau de la dimension. De plus, elle apparait dans les expressions des vitesses de convergence et elle est définie par $F_x(h_n) = \mathbb{P}(d(x, X) \leq h_n)$ où X est une variable aléatoire à valeur dans un espace semi-métrique $(\mathcal{F}, d)$ de dimension infini et x est un point fixé de $\mathcal{F}$. Dans la littérature, on trouve beaucoup d'articles qui donnent des résultats probabilistes qui traitent la manière dont ces probabilités des petites boules tendent vers 0 quand d est une norme (voir par exemple, Li et Shao (2001), Lifshits et al. (2006) et Gao et Li (2007)). On pourra également citer le travail de Dereich (2003) qui a étudié

le comportement des probabilités des petites boules dont les centres sont aléatoires. Ces petites boules de probabilités ont une forme exponentielle (en fonction de h_n) dans le cas de processus non lisses (mouvement Brownian, processus Ornstein- Ullenbeck...) et par conséquent la vitesse de convergence est une puissance de $\log(n)$ (voir Ferraty et Vieu(2006a)). les différents types de semi-métriques existants permettent de trouver une topologie qui mesure de façon pertinente les proximités entre les courbes, dans des situations diverses,(voir Ferraty et Vieu(2006a)). A titre d'exemple, le choix d'une projection semi-métrique (basé sur des composants principaux fonctionnels, la décomposition de Fourier, splines...) constitue une alternative intéressante au problème de dimension en affichant les données dans un espace dimensionnel réduit. Quand la variable explicative est Hilbertienne Ferraty et Vieu(2006a)(Lemme 13.6,p.213) ont justifié qu'il existe une spécifique semi-métrique qui permet de se ramener aux petites boule fractal de probabilités (i.e.$\exists C, \delta > 0, F_x(h_n) \sim C_x h_n^\delta$ quand $h_n \to 0$). Notons qu'il existe d'autres situations où nous pouvons trouver des données très lisses (comme les courbes voir Ferraty et Vieu (2006a)). Dans ce cas, il est de préférence d'utiliser la semi métrique basée sur des dérivées.

1.1.3 Prévision via la régression fonctionnelle : Note bibliographique

En statistique fonctionnelle, il existe plusieurs modèles qui sont consacrés à l'étude du lien entre deux ou plusieurs variables aléatoires, l'un de ces modèles est la régression fonctionnelle (paramétrique ou non paramétrique) qui est un outil important dans la prévision en constituant un champ de recherche privilégié et important dans ces dernier temps. En effet, concernant le contexte paramétrique, et dans le cas de dimension infinie, les travaux de Ramsay et Silverman présentent une large collection de méthodes statistiques pour des variable fonctionnelles dans le cas i.i.d. De même, le manuscrit de Bosq (2000) contribue au traitement des modèles fonctionnels linéaires dans le cas dépendant (processus d'autorégression). Cardot et al. (1999) ont utilisé l'analyse en composantes principales fonctionnelles pour construire un estimateur pour le modèle de la régression linéaire Hilbertienne. Notons que cet estimateur est construit à l'aide des propriétés spectrales de la version empirique de l'opérateur de variance-covariance de la variable explicative fonctionnelle. Leurs résultats asymptotiques portent sur la convergence presque-complète ainsi que la convergence en norme L_2 pour une version régularisée de l'estimateur construit. Cardot et al. (2004) ont proposé un estimateur pour les quantiles conditionnels, vu comme forme linéaire continue définie sur un espace de Hilbert, les auteurs ont obtenu la vitesse de convergence en norme L_2 de l'estimateur proposé sous des hypothèses sur les valeurs propres de l'opérateur de covariance de la variable explicative et sur la densité de la loi conditionnelle.

En ce qui concerne le contexte non paramétrique, les premiers résultats ont été fournis par Ferraty et Vieu (2000). Ces résultats ont été étendus au cas non standard de la régression telle la prévision dans le contexte des séries chronologiques en utilisant le

développement récent de la théorie des probabilités des petites boules par Ferraty et al. (2002). En (2003), Ferraty et Vieu ont abordé une nouvelle approche pour le modèle de régression lorsque la variable explicative à valeurs dans un espace semi-métrique, ils ont obtenu des résultats asymptotiques très fins en utilisant de nouvelles méthodes sur l'estimateur de Nadaraya- Watson dans le cas dépendant. Cette nouvelle approche offre une solution originale pour le problème de fléau de la dimension en utilisant la propriété de concentration sur des petites boules pour la variable explicative. Niang et Rhomari (2003) ont utilisé l'estimateur proposé par Ferraty et Vieu (2003) pour étudier la convergence en norme L_p de l'estimateur proposé par Ferraty et Vieu (2003), les auteurs ont proposé une application à la discrimination des courbes. Dans la même ligne, Masry(2005) a donné la normalité asymptotique de l'estimateur de la fonction de régression. La monographie de Ferraty et Vieu (2006) est considérée comme une référence en statistique non-paramétrique pour des données fonctionnelles en étudiant plusieurs modèles fonctionnelles tels que la régression, le mode (avec ou sans conditionnement), la médiane (avec ou sans conditionnement) ainsi que les quantiles conditionnels, dans les deux cas (i.i.d et mélange fort). De plus, cet ouvrage offre aux lecteurs plusieurs exemples des problèmes concrets pour lesquels les données sont de nature fonctionnelle. Nous renvoyons au travail de Rachdi et Vieu (2007) dans lequel, ils ont étudié l'estimation de la régression lorsque la variable explicative est fonctionnelle, le but de leur travail a été de construire un critère basé sur les données pour choisir ce paramètre de lissage. Delsol (2007) a généralisé les résultats obtenus par Ferraty et al. (2007) au cas dépendant en donnant les termes asymptotiquement dominants des moments centrés et des erreurs L_q de l'estimateur à noyau de la fonction de régression. Burba et al. (2008) ont obtenu la convergence presque-complète de l'estimateur de la fonction de régression via la méthode des k plus proches voisins (k-NN). D'autres auteurs se sont intéressés à l'estimation de la régression fonctionnelle en utilisant la méthode locale linéaire. Nous citons pour cela, le travail de Barrientos et al (2010) où on trouve la convergence presque-complète de l'estimateur local linéaire de la régression fonctionnelle. Chouaf et Laksaci (2012) ont établi la convergence presque-complète de la version spatiale de l'estimateur de Barrientos et al (2010). Récemment, Chouaf (2015) a étudié l'erreur quadratique de l'estimateur de la régression fonctionnelle. Cette procédure d'estimation a été adaptée pour d'autres paramètres conditionnels, à savoir le mode conditionnel (cf. Demongeot et al (2013), Bouanani et al.(2018)), les quantiles conditionnels (cf Messaci et al. (2015)), ainsi que la fonction de hasard conditionnelle (cf.Massiim et Mechab (2016)).
Rappelons que l'estimation à noyau permet de donner des estimateurs réguliers et faciles à utiliser, mais cette technique reste très sensible aux observations aberrantes. Dans le but de résoudre ce problème, Azzedine et al. (2008) ont introduit la régression non paramétrique robuste en montrant la convergence presque-complète des M-estimateurs, La normalité asymptotique de cet estimateur a été abordé par Attouch et

al. (2009).Notons aussi que plusieurs résultats ont été donnés pour l'estimation robuste de la régression quand la co-variable est fonctionnelle Nous citons, à titre d'exemple, Attouch et al (2010, 2012) et Gheriballah et al. (2012) pour le cas de données spatiales et Gheriballah et al. (2013) quand les données sont ergodiques.
D'autres études, sur l'estimation de la régression non paramétrique, ont été menées dans le cas où les deux variables (réponse et explicative) sont fonctionnelles. Nous renvoyons à Ferraty et al. (2011) pour la convergence presque-complète de l'estimateur à noyau. Dabo-Niang et Rhomari (2012) ont obtenu la convergence en norme L_p de l'estimateur proposé par Ferraty et al. (2011), puis Ferraty et al (2012) ont donné la vitesse de convergence presque-complète de l'estimateur de la fonction de régression en considérant des observations β-mélangeantes. Nous citons Ferraty et al. (2013) qui ont étudié la convergence en loi. Heng (2011) a établi la convergence forte d'un estimateur construit par la méthode des k plus proches.

1.2 Statistique non paramétrique robuste

1.2.1 Etat de l'art

Les procédures statistiques robustes (estimation et test) sont des procédures qui ont la capacité de maintenir ses performances malgré les modifications dans les données ou dans les paramètres du modèle choisi. Plus précisément, la robustesse d'une estimation est sa capacité d'être relativement insensible aux larges déviations dues à certaines observations aberrantes. Tandis que, un test statistique est dit robuste s'il reste valable alors que les hypothèses d'application ne sont pas toutes réunies.
Cette notion de robustesse est une question cruciale en statistique. Notons qu'une longue discussion sur ce sujet a été vue à la fin du XIX siècle, plusieurs scientifiques s'y sont intéressés. En fait, le premier résultat sur l'estimation robuste revient à Laplace P. S. (1818) dans son deuxième ouvrage de la théorie analytique des probabilités. De façon plus précise, le terme " robuste " a été introduit au milieu des années cinquante par G. E. P. Box en (1954) mais au milieu des années soixante, cette notion a été reconnue comme un champ de recherche surtout avec les travaux de Huber P.J. (1964), Hampel F.R. (1971). Ils ont développé la statistique robuste en utilisant des arguments de convexité et des critères de type min-max. Nous nous référons à Huber (1973 et 1981), Andrews (1974), Krasker et Welsh (1982) qui ont aussi développé les méthodes automatiques d'ajustement robustes qui sont équivalentes à la méthode des moindres carrés en absence des valeurs atypiques, par contre en présence de ces valeurs ou lorsque la distribution de l'erreur dans le modèle suit une distribution à queues lourdes, ces méthodes surclassent la méthode des moindres carrés.
L'analyse robuste, en particulier l'analyse de la régression robuste est un outil statistique important qui est habituellement appliqué dans la plupart des sciences, et qui a eu un grand intérêt en statistique non paramétrique. Comme nous avons mentionné

précédemment, les premiers résultats ont été obtenus par Huber (1964) qui a étudié la consistance et la normalité asymptotique d'une classe d'estimateurs pour la fonction de régression. Sous des conditions de mélange, Robinson(1984), Härdle (1984) et Härdle et Tsybakov (1989) ont obtenu la normalité asymptotique d'une famille d'estimateurs qui ont une pondération issue de la méthode à noyau de la fonction de régression. Dans la ligne de ces travaux, Boente et Fraiman (1989, 1990) ont étudié les deux paramètres de position et d'échèle en utilisant l'estimateur de Robinson (1984), sous des conditions générales et dans le cas indépendant et dépendant, ils ont obtenu la consistance des estimateurs. Ensuite, Collomb et Härdle (1986) ont établi la convergence uniforme de l'estimateur robuste de la fonction de régression dans le cas où les observations sont ϕ-mélangeantes. En 1994, Fan et al. ont adopté une méthode alternative à l'estimation robuste de la fonction de régression en obtenant la normalité asymptotique de l'estimateur robuste proposé. Laïb et Ould-Saïd (2000) ont introduit une famille d'estimateur non paramétrique robuste pour le modèle d'auto-régression en établissant la convergence uniforme de ces estimateurs sous des hypothèses d'ergodicité. Une version robuste de l'estimation par la méthode des polynômes locaux a été adoptée par Cai et Ould-Saïd (2003), leurs résultats portent sur la normalité asymptotique et la convergence presque-sûre lorsque les observations sont α-mélangeantes, une famille de M-estimateurs robustes pour le paramètre de dispersion a été introduite par Ghement et al. (2008), ils ont donné la consistance et la normalité asymptotiques dans le cas des variables

Par ailleurs, plusieurs auteurs se sont intéressés aux mêmes problématiques quand les variables sont fonctionnelles (cf. Azzedine et al(2008) pour des données indépendantes et Gheriballah et al. (2010) pour le cas où les observations sont spatialement dépendantes).

1.2.2 Analyse non paramétrique robuste des données fonctionnelles

Rappelons tout d'abord que la modélisation statistique des données fonctionnelles est liée à l'étude d'ensembles de données dans lesquels les observations ne sont plus considérées comme un simple vecteur mais elles sont des courbes, des tableaux, ... etc. Ou plus généralement elles sont présentées sous forme de fonctions aléatoires (fonction du temps par exemple). Ce type de données devient de plus en plus courant avec l'évènement de l'informatique, en particulier, les outils informatiques et leurs capacités de stockage de ces données. Un risque de mauvaise spécification de la relation entre les observations peut se produire à cause du volume des données stockées. La régression usuelle reste incapable devant ce genre d'erreurs qui proviennent d'un mauvais enregistrement ou d'une mauvaise lecture ou de toute autre cause dû à l'environnement expérimental. Pour résoudre ce genre de problème, la régression robuste a été adoptée. Autant que nous sachions, l'estimation de régression robuste non paramétrique a été introduite pour la première fois dans le cadre fonctionnel non paramétrique, par Azzedine et al. (2008), où ils ont obtenu la convergence presque-complète d'une famille d'estimateurs robustes de la fonction de régression dans le cas indépendant et identiquement distribué (i.i.d.). Crambes et al. (2008) ont étudié la convergence de la norme L_q dans les deux cas (i.i.d et mélange fort). La normalité asymptotique de ce dernier modèle a été établie par Attouch et al. (2009) sous la propriété de concentration de la mesure de probabilité de la variable fonctionnelle sur des petites boules. La normalité asymptotique de la version dépendante du même modèle a été abordée par les mêmes auteurs en (2010). Le cadre spatial a été étudié dans l'article de Attouch et al. (2010), dans lequel ils ont donné la normalité asymptotique d'une famille d'estimateurs non paramétriques robustes pour la fonction de régression basée sur la méthode à noyau. Après ceci, Attouch et al. (2012), quant à eux, ont généralisé les résultats dus à Azzedine et al (2008) au cas où les observations sont fortement mélangeantes, De plus, ils ont appliqué leurs résultats sur des données réelles. Dans la ligne de ces travaux, Gheriballah et al. (2013) ont déterminé la vitesse la convergence presque-complète du même estimateur sous une hypothèse de processus ergodique stationnaire, sans utiliser les conditions de mélange traditionnelles. Nous renvoyons à Chen et Zhang (2009) pour le même modèle. Tous les résultats que nous avons cités ont été obtenus dans le cas de données complètes. Concernant le cas de données fonctionnelles incomplètes, nous citons le travail de Derrar et al. (2015), leurs résultats asymptotiques portent sur l'étude de la convergence presque-complète et la normalité asymptotique d'une famille d'estimateurs non paramétriques pour le modèle de ψ-régression. L'étude de la convergence presque-complète du même modèle a été récemment généralisé au cas dépendent par Derrar et al. (2018). Comme c'est indiqué plus haut, le modèle de régression usuel n'est pas robuste dans certaines situations, précisément lorsqu'il y a des valeurs aberrantes. Contrairement au cas fonctionnel où nous ne pouvons pas détecter des valeurs aberrantes (vu la nature de données fonctionnelles) mais plutôt des courbes aberrantes qui sont générées par un processus stochastique dont la distribution est différente d'une autre collection de courbes. Pour une méthode de détection de ce genre de courbes, nous nous référons à Febrero et al. (2007a, 2007b), ils ont utilisé la mesure de profondeur qui permet de mesurer la centralité d'une courbe donnée dans un groupe de trajectoires fournissant ainsi un ordre centre-extérieur de l'ensemble de courbes. Selon cette définition, on dit qu'une courbe est atypique si sa profondeur est significativement la plus faible.

1.3 Art de la modélisation locale linéaire en statistique non paramétrique

La modélisation locale linéaire est une approche statistique alternative à l'estimation à noyau, et qui possède de nombreux avantages par rapport à cette dernière. En particulier, le plus grand avantage de la méthode locale linéaire sur la méthode à noyau (locale constante) est la réduction du biais de l'estimateur et la résistance aux effets de bords. De plus, la méthode à noyau peut être traitée comme un cas particulier de la méthode locale linéaire. Les premiers résultats sur ce sujet furent établis à la fin des années soixante-dix par Stone (1977) qui a obtenu la vitesse de convergence minimax d'une famille d'estimateurs construits à partir d'une fonction du poids pour une suite des variables aléatoires indépendantes et identiquement distribuées (i.i.d). Il a également abordé l'étude de plusieurs modèles non paramétriques tels que l'estimation de quantile, des corrélations, des écarts types, des covariances et l'espérance conditionnelle. L'ajustement par la méthode locale linéaire a été largement développé après Stone (1977) et le travail de Fan (1992) a été considéré comme révélateur dans cette direction. Ce dernier a prouvé que la modélisation locale linéaire est asymptotiquement plus efficace que la méthode de base (locale constante) en établissant la vitesse de convergence de l'erreur quadratique moyenne conditionnelle et celle de l'erreur quadratique moyenne conditionnelle intégrée de l'estimateur local linéaire de la régression. En (1993) le même auteur a étudié le problème du risque minimax. Ensuite, Fan et Gijbels (1992) ont traité le problème du choix du paramètre de lissage de l'estimateur local linéaire de la fonction de régression en donnant la forme explicite de l'erreur quadratique moyenne (MSE) et l'erreur quadratique moyenne intégrée(MISE). En outre, ils ont montré que la méthode locale linéaire résiste aux effets de bord et qu'elle produit un biais moindre que celui de la méthode à noyau. Par suite, Fan et al.(1994) ont donné la normalité asymptotique de l'estimateur robuste de la fonction de régression en donnant l'avantage des estimateurs de noyau qui sont basés sur des ajustements linéaires locaux. Une généralisation des résultats de Fan et Gijbels (1992) au cas où la variable explicative est multidimensionnelle a été donnée par Ruppert et Wand (1994). Dans cet article, les auteurs ont explicité le terme dominat de biais et la variance conditionnelle de l'estimateur local linéaire de la régression. D'un autre coté, Fan et Gijbels (1994) ont traité le cas où la variable réponse est observée avec censure pour l'estimation locale linéaire de la fonction de régression et ils ont donné une application sur des données médicales. En (1995) Fan et al ont généralisé cette approche en une approche polynomiale en introduisant une pondération dans la méthode classique et ils ont également prouvé que le degré du polynôme dans l'approximation locale a une influence sur la partie biais. Plus précisément, plus le degré du polynôme augmente plus le gain de la partie biais augmente. Nous nous référons aussi au travail de Fan et Gijbels (1996) pour une présentation assez précise de cette approche et ses applications. durant la même année Masry a établi la normalité asymptotique et la convergence forte lorsque les observations sont fortement mélangeantes de l'estimateur de la fonction de régression multivariée et de ses dérivées. Après ceci, Masry et Fan (1997) ont pris le cas des observations fortement mélangeantes et ρ-mélangeantes et ont étudié la normalité asymptotique des estimateurs des moments conditionnels de la fonction de répartition conditionnelle ainsi que de la densité conditionnelle via la méthode de polynômes locaux. Ces résultats ont été généralisés au cadre spatial par Hallin et al (2004). Nous renvoyons à Masry (2001), qui a repris le travail de Masry (1996) en obtenant la convergence forte du même estimateur, mais il a considéré une corrélation alternative au mélange fort. Nous citons aussi Masry et Mielniczuk (1999) pour l'estimation locale linéaire des dérivées de la fonction de régression et Cai et Masry (2000) pour l'estimation locale linéaire des différentes composantes du modèle additif, ils ont donné la normalité asymptotique de ces estimateurs sous des conditions de mélange fort

Notons que plusieurs auteurs se sont intéressés à l'approche locale linéaire quand les variables sont fonctionnelles ou spatiales. A titre d'exemple, Douge (2010) pour le cas de variables quasi-associées et (Dabo-Niang et al. (2011), Laib et Louani (2011)) pour des données dépendantes. D'autres auteurs ont été intéressés par des versions alternatives à la méthode du noyau classique comme la modélisation locale linéaire par la méthode des k plus proche voisins k-NN. Nous citons le travail de Burba et al. (2009), qui ont établi la vitesse de convergence presque-complète de l'estimateur de la fonction de régression. Nous citons aussi l'article de Attouch et al (2017) pour l'estimation locale linéaire de la fonction de régression par la méthode de k-NN.

1.3.1 Estimation locale linéaire fonctionnelle

Comme nous l'avons mentionné plus haut, la technique de l'estimation locale linéaire a des avantages certains par rapport à la méthode du noyau (locale constante). D'autre part cette approche repose sur une approximation linéaire du modèle au voisinage de la variable observée. De plus, il est bien connu que l'estimation non paramétrique classique (voir Ferraty et Vieu(2006)) est basée sur une pondération des observations les plus proches de la variable observée.

La plupart des résultats que nous avons cités précédemment ont été traités dans le cas où la variable explicative réelle ou vectorielle. La généralisation au cas des variables explicatives fonctionnelles a été abordée très récemment, le premier article intéressant sur ce sujet revient à Baìllo et Grané (2009) qui ont proposé la première version de l'estimation locale linéaire fonctionnelle, ils ont donné la convergence en norme L_2 d'un estimateur local linéaire de l'opérateur de régression lorsque la variable explicative appartient à un espace de Hilbert. Ensuite, Barrientos et al (2010) ont donné une autre version fonctionnelle facile à utiliser même quand la covariable prend ses valeurs dans un espace semi-métrique, Les auteurs dans cet article ont montré la convergence presque-complète de l'estimateur local linéaire de la fonction de régression en appliquant le résultat obtenu sur des données réelles. Chouaf et Laksaci (2012) ont généralisé les résultats de Barrientos et al (2010) au cas d'observations spatio-fonctionnelles en donnant la convergence presque-complète et la vitesse de convergence de la version spatiale de l'estimateur de Barrientos et al. (2010). On peut citer aussi le travail de Berlinet et al. (2011) pour une troisième version. Ces auteurs ont construit un esti-

mateur local linéaire de la fonction de régression ainsi que sa dérivée. Leurs résultats asymptotiques portent sur la convergence en moyenne quadratique de l'estimateur construit. Nous renvoyons à Methni et Rachdi (2010) pour les degrés supérieurs. Nous nous référons à Demongeot et al. (2010) pour l'estimation locale linéaire de la densité conditionnelle. Sous certaines conditions, ils ont établi la convergence presque-complète ponctuelle et uniforme. De plus, ils ont utilisé les résultats obtenus pour déterminer quelques propriétés asymptotiques de l'estimateur local linéaire du mode conditionnel. En 2011, les mêmes auteurs ont abordé le cas de données fonctionnelles dépendantes (cas α-mélangeantes (strong mixing)), les résultats obtenus ont été appliqués au problème de prévision d'une série temporelle par l'estimation du mode conditionnel. La convergence en moyenne quadratique de l'estimateur de la densité conditionnelle dans le cas où les observations sont i.i.d a été obtenue par Rachdi et al (2014). Notons que le cadre spatial a été abordé relativement au thème de la NPFDA. En effet, Laksaci et al. (2013) ont proposé une version spatiale de l'estimation locale linéaire de la fonction de répartition et ses dérivées telles que la densité conditionnelle, le mode conditionnel et la fonction de hasard conditionnelle. Leurs résultats asymptotiques portent sur la convergence presque-complète.
Afin de terminer notre propre littérature sur la méthode locale linéaire fonctionnelle, nous mentionnons quelques travaux récents. En effet, Messaci et al. (2015) ont proposé l'estimation locale linéaire des quantiles conditionnels. En suite. Demongeot et al (2016) ont construit et étudié l'estimation locale linéaire de la fonction de régression pour des variables hilbertiennes. Nous nous référons à Massim et Mechab (2016) pour l'estimation locale linéaire de la fonction de hasard conditionnelle Dans cet article, les auteurs ont établi la convergence presque-complète de l'estimateur construit. Demongeot et al (2017) ont proposé un estimateur local linéaire de la fonction de régression où les deux variables (réponse et explicative) sont fonctionnelles en donnant la convergence presque-complète de l'estimateur proposé. Parallèlement, Chahad et al. (2017) ont montré la convergence presque-complète uniforme et ponctuelle de l'estimateur de la fonction de régression relative lorsque les observations sont (i.i.d). Plus récemment, Bouanani et al (2018) se sont intéressés à l'estimation de quelques paramètres conditionnels fonctionnels via la méthode locale linéaire. Ces derniers ont établi la normalité asymptotique des estimateurs non paramétriques (i) de la fonction de réparation conditionnelle et (ii) les dérivées successives de la densité conditionnelle. De plus, ils ont déduit la normalité asymptotique du mode conditionnel et ils ont appliqué leurs résultats sur des données simulées puis sur des données réelles. Nous nous référons aussi à Almanjahie et al.(2019) pour l'estimation locale linéaire de de la distribution conditionnelle en introduisant la méthode de k-NN.

1.4 Régression non paramétrique Trimmed

Il est bien connu que les estimations des moindres carrés ordinaires d'un modèle de régression sont très sensibles même à une petite quantité de contamination des données. Cela a motivé l'intérêt pour l'étude d'estimateurs robustes. L'estimation des moindres carrés tronqués (LTS pour Least Trimmed Squares) est une des premières méthodes de régression de type robuste apparue dans le travail de Rousseeuw (1984) qui a aussi introduit une approche différente dans laquelle la somme est remplacée par la médiane des résidus au carré, contrairement à la méthode des moindres carrés qui consiste à minimiser la somme des résidus au carré. Nous nous référons à la monographie de Rousseeuw et Leroy (1987) pour plus de détails. D'un point de vue historique, la méthode de LTS a été adoptée par plusieurs auteurs dans le contexte paramétrique. En effet, Rousseeuw et van Driessen (1999) proposent un algorithme rapide pour le calcul de LTS qui a finalement été publié sous le nom de Rousseeuw et van Driessen (2006). Par la suite, Víšek (1999) a montré la normalité asymptotique de l'estimateur des moindres carrés tronqués de la fonction de régression. Après ceci, Agulló(2001) s'est intéressé à l'estimation de LTS de la régression multiple, il a proposé deux algorithmes pour calculer l'estimateur LTS, Le premier algorithme est probabiliste et est basé sur une procédure d'échange. Le deuxième algorithme est exact et basé sur une technique " branch-and-bound "qui garantit une optimalité globale sans évaluation exhaustive. Giloni et Padberg (2002) (T3) ont étudié la régression de LTS et LMS (moindre médiane des carrés) en montrant que la régression LTS est un problème d'optimisation non linéaire qui peut être traité comme un problème de minimisation concave sur un polytope. Ensuite, Jung (2005) a proposé un estimateur robuste du modèle de régression multi-varié basé sur l'estimateur des moindres carrés tronqués (LTS) en effectuant des simulations pour illustrer l'efficacité de l'estimateur proposé. Sous des conditions générales, Víšek (2006) a établi la consistance de l'estimateur LTS dans le cas uni-varié. Par la suite, Agulló et al. (2008) ont introduit l'estimateur LTS pour la régression multi-variée en donnant trois formulations équivalentes de l'estimateur introduit et en obtenant son point de rupture. De plus, ils ont proposé un algorithme rapide pour son calcul, une étude de simulation a été faite pour montrer la robustesse de l'estimateur. Récemment, Morteza et Roozbeh(2019) ont introduit une famille d'estimateurs de crêtes robustes pour les paramètres de régression et la partie non linéaire.
La méthode LTS a été introduite afin d'améliorer la faible efficacité dont souffre la moindre médiane des carrés (LMS) (Least Median of Squares). Cette méthode se base sur la minimisation de la somme des k plus petites valeurs résiduelles au carré. Plus précisément, elle tente à remplacer la somme des résidus au carré sur n points par la somme des résidus au carré sur un sous-ensemble, k, de ces points, en éliminant $(n-k)$ points (pour plus de discussion sur le choix de k, nous citons la monographie de Rousseeuw et Leroy (1987)). Notons que, la méthode LTS a un taux de convergence de $n^{-1/2}$ avec la même efficacité asymptotique à la distribution normale que le M-

estimateur qui est défini par une fonction objective discontinue $\psi(y) = y\mathbb{1}_{|y| \leq q}$, où q est le quantile d'ordre$(1-\alpha/2)$. Ce M-estimateur est connu, dans la littérature, sous le nom de la moyenne sautée de type Huber (Huber-type skipped mean) (cf. Hampel et al. (1986), Rousseeuw (1984), Rousseeuw et Leroy (1987)).
Rappelons que tous les résultats que nous avons cités précédemment ont été obtenus dans le cadre paramétrique (uni-varié ou multi-varié). Dans notre thèse, nous avons abordé pour la première fois, dans le cadre non paramétrique fonctionnel, la régression Trimmed en utilisant la fonction objective ψ définie précédemment, autrement dit, nous avons traité l'étude de la moyenne sautée de type Huber (Huber-type skipped mean) dans le contexte non paramétrique fonctionnel.

1.5 Description de la thèse

Dans notre travail de thèse, nous avons positionné notre contribution dans la vaste littérature de l'analyse non paramétrique des données fonctionnelles (NPFDA). Le travail présenté dans cette thèse porte sur la méthode d'estimation locale linéaire de deux type de régression, à savoir la régression robuste et la régression Trimmed, lorsque la variable repense est réelle et la covariable est éventuellement dans un espace de dimension infinie.
Comme dans la littérature, il n'y a pas de résultats asymptotiques existent dans le cadre de l'estimation par la méthode locale linéaire de la régression robuste et de la régression Trimmed, nous nous sommes intéressés à enrichir ce cadre d'estimation par quelques travaux originaux de recherche en étudiant des propriétés asymptotiques qui sont énoncées en termes de convergence presque-complète qui est connue pour impliquer à la fois la convergence presque-sûre et la convergence en probabilité et la normalité asymptotique.
Ce manuscrit est divisé en six chapitres. Le premier chapitre est un chapitre introductif où nous présentons les différents thèmes abordés dans notre axe de recherche. Nous préférons commencer par un bref historique sur l'estimation non paramétrique fonctionnelle et une brève présentation de la statistique non paramétrique robuste, ainsi nous donnons une courte introduction sur la méthode locale linéaire et un dernier paragraphe sur la régression non paramétrique Trimmed pour finaliser notre chapitre.
Le second chapitre traite l'estimation locale linéaire de la régression robuste. Nous construisons un estimateur en combinant les idées locales linéaires et la technique de M-estimation et afin d'étendre la portée du présent travail à plusieurs cas habituels, nous donnons une présentation générale sans la condition de bornitude de la fonction objective. Les principaux résultats sont l'établissement de la convergence presque-complète et la normalité asymptotique de l'estimateur construit dans le cas où les observations sont i.i.d. Notons que ces résultats sont obtenus sous des conditions générales. Ce travail a fait l'objet d'une publication parue au journal international *Statistics*

and Probability Letters.
Le troisième chapitre est consacré à l'estimation locale linéaire de la régression Trimmed dans le cadre non paramétrique, nous étudions, dans ce chapitre, la convergence presque-complète et la normalité asymptotique de l'estimateur local linéaire Trimmed dans le cas i.i.d., De plus, nous présentons une application à travers une estimation de l'intervalle de confiance. Ce travail a fait l'objet d'une publication acceptée dans le journal International *Journal of Statistics & Economics.*
Dans le quatrième chapitre, on s'est intéressé à étendre les résultats des deux chapitres précédents au cas de variables aléatoires α-mélangeantes en prenant une fonction objective continue et bornée. Nous trouverons aussi dans ce chapitre un exemple sur des données simulées dont l'objectif est l'étude comparative entre quatre types de régression. Ce travail a été soumis pour une éventuelle publication.
Afin de compléter notre travail, nous consacrons le cinquième chapitre à la mise en application des résultats obtenus au troisième chapitre, pour des données simulées, en faisant une étude comparative entre la régression classique et la régression Trimmed. Dans ce chapitre, nous donnerons aussi des commentaires sur le rôle des paramètres qui interfèrent dans la qualité de l'estimation.
Le sixième chapitre est une suite du deuxième chapitre dans lequel nous trouverons la démonstration de résultats asymptotiques qui ont été donnés dans le chapitre deux. Ce chapitre a fait l'objet d'un rapport technique trouvé sur https://www.researchgate.net/profile/Ali_Laksaci.
Enfin, notre manuscrit, se termine par une conclusion et quelques perspectives de recherche.

1.6 Brève présentation des résultats

Nous donnons dans cette partie une brève présentation des résultats obtenus dans cette thèse.

1.6.1 Notations

Considérons le couple de variables aléatoire (X_i, Y_i) pour $i = 1, \ldots, n$ de même loi que (X, Y) à valeur dans $\mathcal{F} \times \mathbb{R}$, où $\mathcal{F}$ est un espace semi-métrique. On note par d la semi-métrique sur $\mathcal{F}$, Pour toute $x \in \mathcal{F}$, le paramètre fonctionnel étudié dans cette thèse, noté θ_x , est définie comme la solution du problème d'optimisation suivant

$$\theta_x = \arg\min_{t \in \mathbb{R}} \mathbb{E}\left[\rho(Y - t) | X = x\right],$$

où ρ (connue sous le nom de la fonction de perte) est une fonction réelle qui a une dérivée première ψ (la fonction objective).
Ce type de modèles couvre et inclut de nombreux modèles non paramétriques habituels,

par exemple, pour $\rho(y) = y^2$ on obtient la régression classique, $\rho(y) = |y||\alpha - \mathbb{1}_{y<0}|$ conduit au α^{th} quantile conditionnel. Pour d'autres exemples, nous renvoyons à Stone (2005).
Rappelons que l'idée de base de l'ajustement local linéaire consiste à approximer localement le modèle non paramétrique par une fonction linéaire. En statistique fonctionnelle, il y a plusieurs manières d'étendre cette approche. Dans notre travail de thèse, nous adoptons la version rapide proposée par Barrientos-Marin et al. (2010) pour laquelle la fonction θ_x est approximée par

$$\forall z \quad \text{dans le voisinage de } x \qquad \theta_z = a + b\beta(z, x),$$

où a et b sont estimés par $\widehat{a}$ et $\widehat{b}$, ils sont la solution de

$$\min_{(a,b)\in\mathbb{R}^2} \sum_{i=1}^{n} \rho(Y_i - a - b\beta(X_i, x))K\left(h^{-1}\delta(x, X_i)\right)$$

où $\beta(.,.)$ et $\delta(.,.)$ sont des opérateurs connus de $\mathcal{F} \times \mathcal{F}$ dans $\mathbb{R}$ tels que

$$\forall \xi \in \mathcal{F}, \beta(\xi, \xi) = 0 \quad \text{et} \qquad d(.,.) = |\delta(.,.)|.$$

La fonction K est un noyau, $h = h_n$ est une suite de nombres réels positifs qui converge vers 0 quand $n \to \infty$. Il est clair que sous cette considération, nous pouvons écrire

$$\theta_x = a \quad \text{et} \qquad \widehat{\theta}_x = \widehat{a}.$$

1.6.2 Résultats : cas i.i.d.

Nous supposons pour commencer que les obseravations sont indépendente et nous établissons la convergence presque-complète (p.co) de nos estimateurs en donnant l'expression explicite des termes de vitesse de convergence et la normalité asymptotique.

I. Régression robuste

Si les deux fonctions ρ et ψ vérifient certaines conditions telles que ρ est une fonction strictement convexe, continuement différentiable et possède une dérivée lipschitzienne ψ telle que

$$\mathbb{E}\left[|\psi(Y - t)|^p | X = x\right] < C < \infty \quad \text{presque-sûrement,} \quad \text{pour tous } p \geq 2,$$

alors, sous certaines hypothèses (cf. le chapitre suivant),nous obtenons les deux thèorème suivants

Thèoreme 1.1. *Sous certaines conditions, et si* $\frac{\partial \Gamma}{\partial t}(x, \theta_x) > 0$, *alors nous obtenons*

$$|\widehat{\theta}_x - \theta_x| = O(h^{\min(k_1,k_2)}) + O\left(\left(\frac{\log n}{n\phi_x(h)}\right)^{1/2}\right), \quad p.co.$$

Thèoreme 1.2. *Sous certaines conditions, et si* $\frac{\partial \Gamma}{\partial t}(x,\theta_x) > 0$, *alors pour tout* $x \in \mathcal{A}$, *nous avons*

$$\left(\frac{n\phi_x(h)}{\sigma^2(x)}\right)^{1/2} \left(\widehat{\theta}_x - \theta_x - o(h)\right) \xrightarrow{\mathcal{D}} \mathcal{N}(0,1) \quad quand \quad n \to \infty.$$

où $\xrightarrow{\mathcal{D}}$ *symbolise la convergence en distribution,*

$$\sigma^2(x) = \frac{\mathbb{E}\left[\psi^2(Y-\theta_x)|X=x\right]}{\left(\frac{\partial \Gamma}{\partial t}(x,\theta_x)\right)^2} \frac{a_3^2 D_1 - 2a_2a_3D_2 + a_2^2 D_3}{(a_1a_3 - a_2^2)^2} \quad et \quad \mathcal{A} = \{x \in \mathcal{F}, \ \sigma^2(x) \neq 0\}$$

avec $\qquad \Gamma(x,\theta_x) := \mathbb{E}\left[\psi(Y-\theta_x)|X=x\right],$

$$a_j = K(1) - \int_{-1}^{1} (s^{j-1}K(s))'\chi_x(s)ds \ \ et \ \ D_j = K(1) - \int_{-1}^{1} (s^{j-1}K^2(s))'\chi_x(s)ds, \ \ pour \ j = 1,2,3.$$

La démonstration et le détail des conditions imposées font l'objet d'un rapport technique sur https://www.researchgate.net/profile/Ali_Laksaci et ils sont présentés également dans le chapitre six de cette thèse.

II. Régression Trimmed

Dans cette partie nous prenons une fonction objective particulière $\psi(y) = y\mathbb{1}_{|y| \leq q}$ où q est le quantile d'ordre$(1-\alpha/2)$. Sous des conditions imposées (cf. le chapitre 3), nous obtenons les deux théorème suivants.

Thèoreme 1.3. *Sous certaines conditions, et si* $\gamma(x,\theta_x) > 0$ *où* $\gamma(x,\theta_x) := \dfrac{d}{dt}\Gamma_1(x,\theta_x)$ *avec* $\Gamma_\lambda(x,\theta_x) := \mathbb{E}[\psi^\lambda(Y-\theta_x)|X=x]$ *pour* $\lambda \in \{1,2\}$, *alors nous obtenons*

$$|\widehat{\theta}_x - \theta_x| = O(h^{\min(k_1,k_2)}) + O\left(\left(\frac{\log n}{n\phi_x(h)}\right)^{1/2}\right), \quad p.co.$$

Thèoreme 1.4. *Sous certaines conditions, et si* $\gamma(x,\theta_x) > 0$ *alors pour tout* $x \in \mathcal{A}$, *nous avons*

$$\left(\frac{n\phi_x(h)}{\sigma^2(x)}\right)^{1/2} \left(\widehat{\theta}_x - \theta_x - o(h)\right) \xrightarrow{\mathcal{D}} \mathcal{N}(0,1) \quad quand \quad n \to \infty.$$

$$\sigma^2(x) = \frac{\Gamma_2(x,\theta_x)}{(\gamma(x,\theta_x))^2} \frac{a_3^2 D_1 - 2a_2a_3D_2 + a_2^2 D_3}{(a_1a_3 - a_2^2)^2} \quad et \quad \mathcal{A} = \{x \in \mathcal{F}, \ \sigma^2(x) \neq 0\}$$

La démonstration et les détails des conditions afin d'obtenir ces deux résultats sont donnés au chapitre 3.

1.6.3 Résultats : cas α mélangeant

Afin de généraliser les résultats obtenus dans les chapitres deux et trois à des observations dépendantes, nous renforçons les hypothèses précédentes en ajoutant des hypothèses sur la concentration de loi conjointe (X_i, X_j) et sur le coefficient de mélange. Nous montrons la convergence presque-complète (avec le taux) en donnant l'expression explicite des termes de vitesse de convergence et la normalité asymptotique. Rappelons que nous avons pris une fonction objective ψ continue et bornée. Alors, sous cetraines conditions (cf. le chapitre 4), nous obtenons les deux théorème suivants.

Thèoreme 1.5. *Sous certaines conditions, et si* $\gamma(x,\theta_x) > 0$ *alors nous obtenons*

$$|\widehat{\theta_x} - \theta_x| = O(h^{\min(k_1,k_2)}) + O\left(\left(\frac{\log n}{n\phi_x(h)}\right)^{1/2}\right), \quad p.co.$$

Thèoreme 1.6. *Sous certaines conditions, et si* $\gamma(x,\theta_x) > 0$ *alors pour tout* $x \in \mathcal{A}$, *nous avons*

$$\left(\frac{n\phi_x(h)}{\sigma^2(x)}\right)^{1/2}\left(\widehat{\theta_x} - \theta_x - o(h)\right) \xrightarrow{\mathcal{D}} \mathcal{N}(0,1) \quad quand \quad n \to \infty,$$

$$\sigma^2(x) = \frac{\Gamma_2(x,\theta_x)}{(\gamma(x,\theta_x))^2}\frac{a_3^2D_1 - 2a_2a_3D_2 + a_2^2D_3}{(a_1a_3 - a_2^2)^2} \quad et \;\; \mathcal{A} = \{x \in \mathcal{F},\; \sigma^2(x) \neq 0\}.$$

Références

Agulló,J.(2001). New algorithms for computing the least trimmed squares regression estimator. Computational *Statistics & Data Analysis,* **36,**425-439.

Agulló,J., Croux,C., Aelst,S.V. (2008). The multivariate least-trimmed squares estimator. *Journal of Multivariate Analysis,***99,** 311 - 338.

Attouch, M., Gheriballah, A., Laksaci, A.(2012). Convergence presque complète d'un estimateur robuste de la régression non paramétrique fonctionnelle : cas spatial. *Ann. I.S.U.P.,* **56,** 3-16.

Attouch, M., A. Laksaci, Ould Saïd,E. (2009) . Asymptotic distribution of robust estimator in nonparametric functional models. *Comm. Stat. Theory Methods,***38,** 1317-1335.

Attouch, M., Laksaci, A., Ould Saïd, E.(2010). Asymptotic normality of robust estimator of the regression function for functional time series data. *Journal Korean statistical society,* **39,** 489-500.

Attouch, M., Laksaci, A.,Rafaa,F.(2017).Estimation locale linéaire de la régression non paramétrique fonctionnelle par la méthode des k plus proches voisins. *Comptes Rendus Mathematique,* **355,** 824-829

Attouch, M., Laksaci, A., Ould Saïd, E. (2012). Robust regression for functional time series data. *Journal of Japan Statist.Soc,* **42**, 125-143.

Azzedine, N., Laksaci, A., Ould-Said, E.(2008). On robust nonparametric regression estimation for a functional regressor. Statist. *Probab. Lett,* **78 ,** 3216-3221.

Baìllo, A. and Grané, A.(2009). Local linear regression for functional predictor and scalar response. *J. Multivariate Anal.,***100,** 102-111.

Barrientos-Marin, J., Ferraty, F. and Vieu, P. (2010). Locally modelled regression and functional data. *J. of Nonparametric Statistics,* **22**, 617-632.

Berlinet, A., Elamine, A.,Mas, A. (2011). Local linear regression for functional data. *Ann. Inst. Statist. Math,* **63,** 1047-1075.

Bouanani,O., Laksaci,A.,Rachdi,A., et Rahmani,S.(2018).Asymptotic normality of some conditional nonparametric functional parameters in high-dimensional statistics.*Behaviormetrika,***45,**,1-35.

Boente, G., Fraiman, R. (1989). Robust nonparametric regression estimation for dependent observations. *Ann. Statist.,* **17,** 1242-1256.

Boente, G., Fraiman, R. (1990). Asymptotic distribution of robust estimators for nonparametric models from mixing processes. *Ann. Statist.,* **18,** 891-906.

Bosq, D. (2000). Linear processes in function spaces. Theory and Application. *Lectures Notes in Statistics. Springer Verlag,* New-York, Vol. 149.

Box,G.E. P. (1954). Some Theorems on Quadratic Forms Applied in the Study of Analysis of Variance Problems, I. E_ect of Inequality of Variance in the One-Way Classi_cation. *The Annals of Mathematical Statistics,* **25,**, 290-302.

Burba,F.,Ferraty,F., Vieu,F. (2008). Convergence de l'estimateur à noyau des k plus proches voisins en régression fonctionnelle non-paramétrique. *C. R. Acad. Sci., Paris,* **346,** 339-342.

Burba, F., Ferraty, F., Vieu, P.(2009). k-nearest neighbour method in functional non-parametric regression. *J. Nonparametr. Stat,* **21,** 453-469.

Chahad,A., Ait-Hennani,L, Laksaci,A.(2017).Functional local linear estimate for functional relative-error regression.*Journal of statistical theory and practice,***11** , 771-789 .

Cai, Z., Masry, E.(2000). Nonparametric estimation of additive nonlinear ARX time series : local linear fitting and projections. *Econometric Theory,* **16,** 465-501.

Cai,Z., Ould Saïd,E. (2003). Local M-estimator for nonparametric time series. Statist. *And Probab. Lett,* **65,** 433-449.

Cardot,H. Crambes,C., Sarda,P. (2004). Spline estimation of conditional quantiles for functional covariates. *C. R. Math. Acad. Sci.,* **339,** 141-144.

Cardot,H., Ferraty,F., Sarda,P. (1999). Functional linear model. *Statist.Probab. Lett.,* **45,** 11-22.

Chouaf, A., Laksaci, A.(2012). On the functional local linear estimate for spatial regression. *Stat. Risk Model,* **29,** 189-214.

Chen, J., Zhang, L. (2009). Asymptotic properties of nonparametric M-estimation for mixing functional data. *J Stat Plann Inference,***139,** 533-546.

Collomb, G., Härdle, W. (1986). Strong uniform convergence rates in robust non-parametric time series analysis and prediction : Kernel regression estimation from dependent observations. *Stoch. Proc. and their Appl,* **23,** 77-89.

Crambes, C., L. Delsol, Laksaci, A.(2008) . Robust nonparametric estimation for functional data. *J. Nonparametric Stat,***20,** 573-598.

Dabo-Nian,S., Rhomari,N. (2003). Estimation non paramétrique de la régression avec variable explicative dans un espace métrique. *C. R. Acad. Sci,* **336,** 75-80.

Dabo-Niang, S., Rachdi, M., Yao, A.(2011). Kernel regression estimation for spatial functional random variables. *Far East J. Theor. Stat.,***37,** 77-113.

Dauxois, J., Pousse, , A. and Romain, Y. (1982). Asymptotic theory for the principal component analysis of a vector random function : *some applications to statistical inference J. Multivariate Anal.* **12,** 136-154.

Delsol. (2007). Régression non paramétrique fonctionnelle : expression asymptotique des moments. *Ann. I. S. U. P., L1,* 43-68.

Demongeot, J., Laksaci, A., Madani, F. et Rachdi, M.(2010). Estimation locale linéaire de la densité conditionnelle pour des données fonctionnelles. *C. R. Acad. Sci,* **348,** 931-934.

Demongeot, J., Laksaci, A., Madani, F. et Rachdi, M. (2011). A fast functional locally modeled conditional density and mode for functional time-series. *In : Recent Advances in Functional Data Analysis and Related Topics. In : Contributions to Statistics, Physica-Verlag/Springer, pp,* **85-90.** http ://dx.doi.org/10.1007/978-3-7908-2736-1-13.

Demongeot, J., Laksaci, A., Madani, F. et Rachdi, M. (2013). Functional data : local linear estimation of the conditional density and its application. *Statistics,***47,** 26-44.

Demongeot, J., Laksaci, A., Naceri, A. et Rachdi, M. (2016). Estimation locale linéaire de la fonction de régression pour des variables hilbertiennes. *C.R. Math. Acad. Sci,* **354,** 847-850.

Demongeot, J., Laksaci, A., Naceri, A. et Rachdi, M. (2017). Local linear regression modelization when all variables are curves. *Statistics and Probability Letters,***121,** 37-

Dereich, S. (2003). High resolution coding of stochastic processes and small ball probabilities. *PhD Thesis.*

Derrar,S., Laksaci,A., Ould Saïd, E. (2015). On the Nonparametric Estimation of the Functional ψ-Regression for a Random Left-Truncation Model. *Journal of Statistical Theory and Practice,* **9,** 823-849 .

Derrar,S., Laksaci,A., Ould Saïd, E. (2018). M-estimation of the regression function under random left truncation and functional time series model. *Statistical Papers,* 1-22.

Douge, L.(2000). Théorèmes limites pour des variables quasi-associées hilbertiennes. *Ann. I.S.U.P.,* **54,** 1-2, 51-60.

Deville.J.C. (1974). Méthodes statistiques et numériques de l'analyse harmonique. *Ann. Insee,* **15,** 3-101.

El Methni, M.,Rachdi, M. (2010) Local weighted average estimation of the regression operator for functional data, *Commun. in Statist.- Theory and Meth.,* **40,** 3141-3153.

Fan, J. (1992). Design-adaptive nonparametric regression. *J. Amer. Statist. Assoc,* **87,** 998-1004.

Fan, J.(1993). Local linear regression smoothers and their minimax efficiencies. *Ann. Statist,* 21 (1993), no. 1, 196-216.

Fan, J., Gijbels, I. (1992).Variable bandwidth and local linear regression smoothers. *Ann. Statist,* **20,** 2008-2036.

Fan, J., Gijbels, I.(1994). Censored regression : local linear approximations and their applications. *J. Amer. Statist. Assoc,* **89,** 560- 570.

Fan, J., Gijbels, I.(1996). Local Polynomial Modelling and its Applications. *London, Chapman and Hall.*

Fan, J., Heckman, N.E., Wand, M.P.(1995). Local polynomial kernel regressixon for generalized linear models and quasi-likelihood functions. *J. Amer. Statist. Assoc,* **90,** 141-150.

Fan, J., Hu, T.C., Truong, Y.K.(1994) . Robust non-parametric function estimation. *Scand. J. Statist,* **21,** 433-446.

Febrero,M., Galeno,P., Gonzàlez-Manteiga,W. (2007a). A functional analysis of NOx levels :, location and scale estimation and outlier detection. *Computational Statistics,* **22,** 411-427.

Febrero,M., Galeno,P., Gonzàlez-Manteiga,W.(2007b) Outlier detection in functional data by depth measures with application to identify abnormal NOx levels. *Environmetrics,* **19,** 331-347.

Friedman,J.,Stuetzle,W.(1981). Projection pursuit regression, J. Amer. *Statist. Assoc,* **76,** 817-823.

Ferraty, F., Laksaci, A., Tadj, A., et Vieu, P. (2011). Kernel regression with functional response. *Electron. J. Statist,* **5,** 159-171.

Ferraty, F., Laksaci, A., Tadj, A., et Vieu, P. (2012). Estimation de la fonction de régression pour variable explicative et réponse fonctionnelles dépendantes. *C.R. Math. Acad. Sci,* **350,** 717-720.

Ferraty, F., Sued, M. and Vieu, P. (2013). Mean estimation with data missing at random for functional covariables. *Statistics,* **47** 688-706.

Ferraty, F., Van Keilegom, I.,Vieu, P.(2012). Regression when both response and predictor are functions. *J. Multivariate Anal,* **109,** 10-28.

Ferraty, F., Vieu, P.(2000). Dimension fractale et estimation de la régression dans des espaces vectoriels semi-normés. *C. R. Acad. Sci,***330,** 139-142.

Ferraty, F., Vieu, P. (2002). The functional nonparametric model and application to spectrometric data. Comput. *Statist. and Data Anal,* **17,** 545-564.

Ferraty, F., Vieu, P. (2003). Curves discrimination : a nonparametric functional approach. Comput. Statist. *and Data Anal,* **44,** 161-173.

Ferraty, F., Vieu, P. (2006). Nonparametric functional data analysis. *Theory and Practice. Springer- Verlag.*

Ferraty, F., Vieu, P. (2006a). Nonparametric modelling for functional data. *Springer Verlag, New York.*

Ferraty,F., Mas.,A., Vieu,p. (2007). Advances in nonparametric regression for functional variables. *Aust. New Zeal. J. of Satist,* **49,** 1-20.

Gao, F., et Li, W.V., (2007). Small ball probabilities for the Slepian Gaussian fields. *Trans. Amer. Math. Soc,* **359,** 1339-1350.

Ghement,I., Ruiz,M., Zamar,R., (2008). Robust Estimation of Error Scale in Nonparametric Regression Models. *J. of Statist. Plan. and Inf,* **138,** 3200-3216.

Gheriballah, A., Laksaci, A., Rouane, R. (2010). Robust nonparametric estimation for spatial regression. *Journal of Statistical Planning and Inference,* **140,** 1656-1670.

Gheriballah, A., Laksaci, A., Sekkal, S.(2013). Nonparametric M-regression for functional ergodic data. Statist. *Probab. Lett,* **83,** 902-908.

GILONI,A., PADBERG,M. (2002). Least Trimmed Squares Regression, Least Median Squares Regression, and Mathematical Programming. *Mathematical and Computer Modelling ,* **35,** 1043-1060.

Grenander,U. (1981). Abstract inference (Wiley series in probability and mathematical statistics)

Gu,C.(2000) Multivariate spline regression. Smoothing and Regression : Approaches computation and application. *Ed. M.G. Schimek. Wiley Series in Probability and Statistics,* 329-356.

Hallin, M., Lu, Z. and Tran, L.T. (2004). Local linear spatial regression. *Ann. of Statist,* **32,** 2469-2500.

Hampel.F.R. (1971). A general qualitative definition of robustness. *Ann. Math. Stat,* **42,** 1887-1896.

Hampel,F.R., Ronchetti,E.M,. Rousseeuw, P. J., Stahel.,W.A. (1986). Robust Statistics : The Approach Based on Influence Functions. *John Wiley & Sons, New York.*

Härdle, W. (1984). Robust and nonlinear time series analysis. *Lecture Notes in Statistics, Springer- Verlag, New York,26*

Härdle, W., Müller, M. (2000). Multivariate and semiparametric kernel regression. *Smoothing and Regression : Approaches, computation, and application, Ed. M.G. Schimek,* 357-392, Wiley Series in Probability and Statistics.

Härdle, W., Tsybakov, A. B. (1988). Robust nonparametric regression with simultaneous scale curve estimation. *Ann. Statist,* **16,** 120-135.

Holmstrome,I. (1963). On a method for parametric representation of the state of the atmosphere. *Tellus,* **15,** 127-149

Huber,P.J.(1964). Robust estimation of a location parameter. *Ann. Math. Statist,***35,** 73-101.

Huber,P.J.(1985). Projection Pursuit. *Ann. Statist,* **13,** 435-475.

Huber,P.J. (1973). Robust regression. Asymptotics, conjonctures and monte Carlo. *Ann. Stat,* **43,** 799-821.

Huber,P.J.(1981). Robust statistics. *John Wiley.*

Jung,K.M.(2005). Multivariate least-trimmed squares regression estimator. *Comput. Statist. Data Anal.* **48** 307-316.

Kirkpatrick, M., Heckman,N.(1989). A quantitative genetic model for growth, shape, reaction norms, and other infinite-dimensional characters. *J. Math. Biol,* **27,** 429-450.

Krasker,W.S., Welsch,R.E (1982). Efficient bounded influence regression estimation. *J. A. S. A,* **77,** 595-604.

Laïb, N., Louani, D. (2011). Rates of strong consistencies of the regression function estimator for functional stationary ergodic data. *J. Statistical Planning and Inference,* **141,** 359-372.

Laïb, N., Ould-Saïd, E. (2000). A robust nonparametric estimation of the autoregression function under ergodic hypothesis. *Canad. J. Statist,* **28,** 817-828.

Laksaci, A., Rachdi, M. and Rahmani, S. (2013). Spatial modelization : Local linear estimation of the conditional distribution for functional data. *Spatial Statistics,* **6** 1-23.

Laplace,P.S. (1818). Deuxième supplément de Laplace.

Li, W.V., Shao, Q.M. (2001).Gaussian processes : inequalities, small ball probabilities and applications. *In :C.R. Rao and D. Shanbhag (eds.) Stochastic processes, Theory and Methods. Handbook of Statitics,* **19,** North-Holland, Amsterdam.

Lifshits, M.A., Linde, W., Shi, Z. (2006). Small deviations of Riemann- Liouville processes in L_q-spaces with respect to fractal measures. *Proc. London Math. Soc,* **92** 224-250.

Masry, E.(1996). Multivariate local polynomial regression for time series : uniform strong consistency and rates. *J. Time Ser. Anal., 17,* **6,** 571-599.

Masry, E.(2001). Local linear regression estimation under long-range dependence : strong consistency and rates. *IEEE Trans. Inform. Theory.,* **47,** 2863-2875.

Masry,E. (2005). Nonparametric regression estimation for dependent functional data : asymptotic normality. *Stoch. Proc. and their Appl.,* **115,** 155-177.

Masry, E.,Fan, J. (1997). Local polynomial estimation of regression functions for mixing processes. *Scandinavian Journal of Statistics,* **24,** 165-179.

Masry, E., Mielniczuk, J.(1999). Local linear regression estimation for timeseries with long-range dependence. *Stochastic Process. Appl,* **82,**173-193.

Massim, I.,Mechab,B. (2016).Local linear estimation of the conditional hazard function. *Journal of Statistics & Economics.***17,** 1-11.

Messaci, F., Nemouchi, N., Ouassou, I. et Rachdi, M.(2015). Local polynomial modelling of the conditional quantile for functional data. *Stat. Methods Appl,* **24,** 597-622.

Molenaar,p.,Boomsma,D.(1987). The genetic analysis of repeated measures : the karhunen-loeve expansion. *Behavior Genetics,* **17,** 229-242.

Morteza, A., Roozbeh, M. (20016).Least trimmed squares ridge estimation in partially linear regression models. *Journal of Statistical Computation and Simulation,* **86,** 2766-2780.

Nychka,D.W.(2000). Spatial process estimates as smoothers. Smoothing and Regression : Approaches, Computation, and application. *Ed. M.G. Schimek, Wiley Series in Probability and Statistics.* 393-424.

Obukhov,O. (1960). The statistically orthogonal expansion of empirical functions.*American Geophysical Union,* 288-291.

Rachdi, M., Laksaci, A., Demongeot, J., Abdali, A. et Madani, F.(2014). Theoretical and practical aspects of the quadratic error in the local linear estimation of the conditional density for functional data. *Comput. Statist. Data Anal,* **73,** 53-68.

Rachdi,M, Vieu.P. (2007). Nonparametric regression functional data : Automatic smoothing parameter selection. *J. Stat. Plan. Inf,* **137,** 2784-2801.

Ramsay, J.,Silverman, B. (1997). Functional Data Analysis, *Springer- Verlag, New York.*

Ramsay, J., Silverman, B.(2002). Applied functional data analysis : Methods and case studies. *Springer, New-York.*

Ramsay, J., Silverman, B. (2005). Functional Data Analysis (Second Edition). *Spinger- Verlag, New York.*

Robinson,R. (1984). Robust nonparametric autoregression. *Lecture Notes in Statistics,* **26,** 247-255.

Rousseeuw, P. J. (1984). Least median of squares regression. *Journal of the American Statistical Association,* **79,** 871-880.

Rousseeuw, P. J., and Leroy, A. M. (1987). Robust Regression and Outlier Detection. *New York : John Wiley & Sons.*

Rousseeuw, P. J., van Driessen, K. (1999). Computing lts regression for large data sets. *Discussion paper, Institute of Mathematical Statistics Bulletin.*

Rousseeuw, P. J.,van Driessen, K. (2006). Computing lts regression for large data sets. *Data mining and Knowledge Discovery,***12,** 29-45.

Ruppert, D. Wand, M.P. (1994). Multivariate locally weighted least squares regression. *Ann. Statist,* **22,** 1346-1370.

Stone, C.J. (1977).Consistent nonparametric regression. With discussion and a reply by the author. *Ann. Statist,* **5** 595-645.

Víšek, J. A. (1999). The Least Trimmed Squares - random carriers. *Bulletin of the Czech Econometric Society,* **6,** 1-30.

Víšek., J. A. (2006). The Least Trimmed Squares. Part II : p n-consistency. *Kybernetika,* **42,** 181-202.

Wu,C.O.(2000). Local polynomial regression with selection biased data. *Statistica Sinica,* **10,** 789-817.

CHAPITRE 2

LOCAL LINEAR ESTIMATE OF THE NONPARAMETRIC ROBUST REGRESSION IN FUNCTIONAL DATA

Ce chapitre a fait l'objet d'une publication internationale *Statistic and probability letters.*

Local linear estimate of the nonparametric robust regression in functional data

Faiza Belarbi [a], **Souheyla Chemikh**[a], **Ali Laksaci** [b]

[a]Laboratoire de statistique et processus stochastiques, Univ. Djillali Liabès, Algérie.

faiza_belarbi@yahoo.fr and chemikh_souheyla@yahoo.com

[b]Department of Mathematics, College of Science, King Khalid University, P.O. Box : 9004, Postal Code : 61413, Abha, Saudi Arabia.

alilak@yahoo.fr

Abstract : *In this paper, we study the robust estimation of the functional local linear regression model. The main results of this work are the establishment of the almost complete convergence as well as the asymptotic normality for the constructed estimator.*

Keywords **:** Functional data ; analysis Local linear method ; Robust estimation ; Asymptotic normality ; Almost complete convergence.

2.1 Introduction

The nonparametric functional data analysis is an emergent field of research in nonparametric statistics. In this context, the modelization of the relationship between a scalar response variable and functional covariate has attracted a considerable interest. The most used approaches are based on local constant fits. In this work, we propose to model this relationship by using the local linear M-regression.
It is well known that the local linear method has several advantages over the local constant fitting. In particular, it allows to reduce the bias term in a various situations. We refer to Fan and Gijbels (1996) for more discussions on the importance of this approach. Noting that the robust local linear estimation has investigated by many authors in the multivariate case. See Fan et al. (1994), for the i.i.d. case, Cai and Ould-Said (2003) for the α-mixing case. We return to Boente et al. (2009) for some asymptotic results in the robust local constant method for both procedures (kernel and KNN methods). In this work, we will focus on the case where the covariate is of infinite dimension. It should be noted that, these questions of statistical analysis of infinite dimensional data are arising more and more in applied statistics. For a

bibliographical survey or an overview on recent developments in this topic, we cite, for example, Geenens (2011), Horvàth and Kokozsca (2012), Zhang (2013), Hsing and Eubank (2015) and Goia and Vieu (2016).

The local linear method has been considered for functional data. The first interesting results, on this topic, were obtained by Baìllo and Grané (2009). They established the L2 consistency of a local linear estimator of the regression operator when the explanatory variable is Hilbertian. Barrientos-Marin et al. (2010) consider an alternative fast version of the functional local linear estimator which can be used for more general functional regressor. This last contribution gives the almost complete convergence (with rate) of the proposed estimate. Berlinet et al (2011) propose another local linear estimate based on the inverse of the local covariance operator of the functional explanatory variable. We return to Zhou and Lin (2016) for the asymptotic normality of the functional local linear regression estimate. For a more recent advances on the functional kernel estimate or on its alternative functional smoothers that is the KNN-method we cite Kara-Zaitri et al (2017a, 2017b). All these studies focus mainly on the classical regression. However, it is well known that this model is very sensitive to outliers and do not perform well when the errors are heavy-tailed. Such kind of data is observed very often in econometrics and finance as well as in many other applied fields. To attenuate the lack of robustness of this model, we propose in this work to robustify the functional local linear regression model.

In recent years, there is a growing body of studies that consider the nonparametric robust regression for functional data. We cite, for instance, Chen and Zhang (2009) for previous results and Boente and Vahnovan (2015) for recent advances and references. All these works use an estimation procedure based on the local constant fits. As far as we are aware, there has not been any attempt so far to study the robust-type of local linear regression in functional data. This is the main goal of this paper. Precisely, we combine the ideas Barrientos-Marin et al. (2010) on the functional local linear method with those of the robustness developed by Fan et al. (1994) to construct an estimator resistant to the presence of outliers or heteroscedasticity of data and inherits many nice statistical properties of the local linear approach. The asymptotic behavior of this approach is studied under some standard conditions of nonparametric functional data analysis. Precisely, under the concentration properties on small balls of the probability measure of this functional variable, we establish the almost complete consistency (with rate) and the asymptotic normality of this estimator.

This paper is organized as follows : We present our model in Section 2.2. The needed conditions and the main results are given in Section 2.3. In Section 2.4 we highlight the main features of our approach and we compare it to the existing approaches. We present also in this section some prospects of the present contribution. A sketch of the proof of the main result is done in the Appendix. See the details of the proof in Belarbi et al. (2017).

2.2 The robust local linear estimator

Consider n independent pairs of random variables (X_i, Y_i) for $i = 1, ..., n$ that we assume drawn from the pair (X, Y). The latter is valued in $\mathcal{F} \times \mathbb{R}$ where $\mathcal{F}$ is a semi-metric space and d denotes a semi- metric. The object of this paper is to study the co-variation between X_i and Y_i by the nonparametric robust regression function. For $x \in \mathcal{F}$ the nonparametric robust regression, denoted by θ_x, is defined as the unique minimizer of

$$\theta_x = \arg\min_{t \in \mathbb{R}} \mathbb{E}\left[\rho(Y - t)|X = x\right] \tag{2.1}$$

where $\rho(.)$ is a real-valued Borel function satisfying some regularity conditions to be stated below. This kind of models belongs to the class of M-estimates introduced by Huber (1964). It covers and includes many usual nonparametric models, for example, for $\rho(y) = y^2$ we obtain the classical regression, $\rho(y) = |y||\alpha - \mathbb{1}_{y<0}|$ leads to the α^{th} conditional quantile. The α^{th} conditional expectile is obtained by setting $\rho(y) = y^2|\alpha - \mathbb{1}_{y<0}|$. For more others examples we refer the reader to Stone (2005).
Recall that the basic idea of local linear fitting consists in approximating locally the nonparametric model by a linear function. In functional statistics, there are several ways for extending this approach (see, Baìllo and Grané (2009) or Barrientos-Marin et al. (2010) for some examples). In this paper, we adopt the fast version proposed by Barrientos-Marin et al. (2010) for which the function θ_x is approximated by

$$\forall z \quad \text{in neighborhood of } x \qquad \theta_z = a + b\beta(z, x)$$

where a and b are estimated by $\widehat{a}$ and $\widehat{b}$ are solution of

$$\min_{(a,b)\in\mathbb{R}^2} \sum_{i=1}^{n} \rho(Y_i - a - b\beta(X_i, x))K\left(h^{-1}\delta(x, X_i)\right) \tag{2.2}$$

where $\beta(., .)$ is a known function from $\mathcal{F} \times \mathcal{F}$ into $\mathbb{R}$ such that, $\forall\xi \in \mathcal{F}$, $\beta(\xi, \xi) = 0$, with K is kernel and $h = h_n$ is a sequence of positive real numbers and $\delta(., .)$ is a function of $\mathcal{F} \times \mathcal{F}$ such that $d(., .) = |\delta(., .)|$. It is clear that, under this consideration, we can write

$$\theta_x = a \quad \text{and} \quad \widehat{\theta_x} = \widehat{a}.$$

We point out that unlike to the classical regression case studied by Barrientos-Marin et al.(2010) the robust local linear estimator $\widehat{\theta_x}$ cannot explicitly expressed. Thus, the establishment of the asymptotic proprieties of our estimate is very difficult, it requires some additional tools.

2.3 Main results

In what follows, when no confusion is possible, we will denote by C and C' some strictly positive generic constants. Moreover, x denotes a fixed point in $\mathcal{F}$, N_x denotes a fixed neighborhood of x. For $i = 1, \ldots, n$, we denote by $K_i = K(h^{-1}\delta(x, X_i))$, and $\beta_i = \beta(X_i, x)$. Furthermore, we put $\phi_x(r_1, r_2) = \mathbb{P}(r_2 \leq \delta(x, X) \leq r_1)$ and we assume the following hypotheses :

(H1) For any $r > 0$, $\phi_x(r) := \phi_x(-r, r) > 0$ and there exists a function $\chi_x(\cdot)$ such that :

$$\forall t \in (-1, 1), \qquad \lim_{h\to 0} \frac{\phi_x(th, h)}{\phi_x(h)} = \chi_x(t).$$

(H2) ρ is a strictly convex function, continuously differentiable and has a Lipschitzian derivative ψ such that

$$\mathbb{E}\left[|\psi(Y - t)|^p | X = x\right] < C < \infty \quad \text{almost surely,} \quad \text{for all } p \geq 2.$$

(H3) The function $\Gamma(x, \cdot) := \mathbb{E}\left[\psi(Y - \cdot)|X = x\right]$ is of class C^1 on $[\theta_x - \delta, \theta_x + \delta]$; $\delta > 0$. such that

$$\begin{cases} (i)\, \forall (t_1, t_2) \in [\theta_x - \delta, \theta_x + \delta]^2,\ \forall (x_1, x_2) \in N_x \times N_x, \\ \qquad\qquad |\Gamma(x_1, t_1) - \Gamma(x_2, t_2)| \leq C d^{k_1}(x_1, x_2) + C'|t_1 - t_2|^{k_2}. \\ (ii) \text{ The variable } \delta(x, X) \text{ is } \sigma(\beta(X, x)) - \text{ measurable and the two partial derivatives} \\ \text{of the function, } \Upsilon_x(s, t) := \mathbb{E}[\Gamma(X, t)|\beta(X, x) = s] \text{ at } (0, \theta_x) \text{ exist.} \end{cases}$$

(H4) The function $\beta(.,.)$ is such that :

$$\forall z \in \mathcal{F},\ C|\delta(x, z)| \leq |\beta(z, x)| \leq C'|\delta(x, z)| \text{ and } \sup_{u \in B(x,r)} |\beta(u, x) - \delta(x, u)| = o(r).$$

where $B(x, r) = \{x' \in \mathcal{F}/d(x, x') \leq r\}$.

(H5) The kernel K is a positive, differentiable function which is supported within $(-1, 1)$ such that

$$\begin{pmatrix} K(1) - \int_{-1}^{1} K'(t)\chi_x(t)dt & K(1) - \int_{-1}^{1} (tK(t))'\chi_x(t)dt \\ K(1) - \int_{-1}^{1} (tK(t))'\chi_x(t)dt & K(1) - \int_{-1}^{1} (t^2K(t))'\chi_x(t)dt \end{pmatrix} \text{ is a positive de definite matrix.}$$

(H6) The bandwidth h satisfies : $\displaystyle\lim_{n\to\infty} \frac{\log n}{n\phi_x(h)} = 0.$

Conditions (H1)-(H6) are very mild in the nonparametric functional data analysis. Indeed, Condition (H1) is usually assumed in this context and is verified for a large class of the usual processes. Such a condition is extensively discussed and commented in the literature of nonparametric functional statistics (see, Ferraty and Vieu (2006)). Conditions (H2)-(H3) are quite lower than those considered by Attouch et al. (2009).

Finally, Assumptions (H4)-(H6) are similar to those used by Barrientos- Marin et al.(2010). Our main results are summarized in the following Theorems

Theorem 2.1. *Under hypotheses (H1), (H2), (H3(i)), (H4)-(H6) and if* $\frac{\partial\Gamma}{\partial t}(x,\theta_x) > 0$ *then*

$$|\widehat{\theta_x} - \theta_x| = O(h^{\min(k_1,k_2)}) + O\left(\left(\frac{\log n}{n\phi_x(h)}\right)^{1/2}\right), \quad a.co.$$

Theorem 2.2. *Under hypotheses (H1)-(H6) and if* $\frac{\partial\Gamma}{\partial t}(x,\theta_x) > 0$ *then, for any* $x \in \mathcal{A}$, *we have*

$$\left(\frac{n\phi_x(h)}{\sigma^2(x)}\right)^{1/2}\left(\widehat{\theta_x} - \theta_x - o(h)\right) \xrightarrow{\mathcal{D}} \mathcal{N}(0,1) \quad as \quad n \to \infty.$$

where $\xrightarrow{\mathcal{D}}$ *denotes the convergence in distribution,*

$$\sigma^2(x) = \frac{\mathbb{E}\left[\psi^2(Y-\theta_x)|X=x\right]}{\left(\frac{\partial\Gamma}{\partial t}(x,\theta_x)\right)^2}\frac{a_3^2D_1 - 2a_2a_3D_2 + a_2^2D_3}{(a_1a_3 - a_2^2)^2} \quad and \quad \mathcal{A} = \{x \in \mathcal{F},\ \sigma^2(x) \neq 0\}$$

with

$$a_j = K(1) - \int_{-1}^{1}(s^{j-1}K(s))'\chi_x(s)ds \ \ and \ \ D_j = K(1) - \int_{-1}^{1}(s^{j-1}K^2(s))'\chi_x(s)ds, \ \ for\ j = 1,2,3.$$

2.4 Conclusions and prospects

On the bias of the estimate : As discussed in the introduction section the behavior of the bias term is one of the most principal motivations for the local linear approach (LL). Although, the asymptotic behavior of this term is closely linked to the regularity assumption of the nonparametric model, the bias term can easily improved in the local linear approach for various situations. This important feature is due to the specific form of the weighting functions in LL approach (see, Fan (1993, P. 198) for the non functional regression). The same conclusion can be stated for this functional robust approach. In this case, under a standard regularity assumption, we have improvement the bias term of the Nadaraya- Watson (NW) estimator obtained by Attouch et al. (2009). Specifically, here, we have a bias of $o(h)$ which is significantly better than $O(h)$ the bias of the NW estimator. This gain is due to the nature of the weighting functions of $\widehat{\Phi}(x)$ (see, Corollary 2.1, below) which is the leading part of $(\widehat{\theta_x} - \theta_x)$. These functions such that $\mathbb{E}\left[\beta_i(A_3K_i - A_2h^{-1}\beta_iK_i)\right] = 0$. Under this consideration, we prove, by a simple calculation, that $\mathbb{E}\left[\widehat{\theta_x} - \theta_x\right] = o(h)$. Moreover, if we suppose that the function $\Upsilon_x(s,t)$ is of class C^2 at neighborhood of $(0,\theta_x)$, we obtain a bias term of order $O(h^2)$. In conclusion, we can say that even if the robust LL and the robust NW approaches has, generally, the same asymptotic properties, the specific forme of the LL method allows to improve the bias term in a certain situations.
On the variance term : It is well documented that, the LL estimate keeps the same

variance term of the NW estimator. Of course this statement is asymptotically true of this functional robust model. Nevertheless, we can also keep exactly the same constants of the variance of the NW estimator by considering some additional condition. Indeed, recall that, in the multivariate case we consider a kernel K of order 2 to obtain the same variance expression of the NW estimator. In this functional context, this condition can be replaced by

$$K(1) - \int_{-1}^{1} (tK(t))'\chi_x(t)dt = 0.$$

This last is very restrictive in the functional data analysis, because, it depends on some structural parameter that is $\chi_x(t)$ the asymptotic variation of the small probability function. Alternatively, we can use an other condition that is

$$h \int_{B(x,h)} \beta(u,x)dP(u) = o\left(\int_{B(x,h)} \beta^2(u,x)dP(u)\right).$$

Such additional assumption has been considered by Barrientos-Marin et al.(2010). They proved that, under this condition, we have $\mathbb{E}[\beta K] = o(h\phi_x(h))$ which yields a variance term

$$\sigma_2(x) = \frac{D_1}{a_1^2} \frac{\mathbb{E}\left[\psi^2(Y-\theta_x)|X=x\right]}{\left(\frac{\partial\Gamma}{\partial t}(x,\theta_x)\right)^2}.$$

This expression is the same as of the kernel method. Of course, this additional condition is not very restrictive, but, it concerns the function β which is a fundamental element of the functional structure of the space $\mathcal{F}$. So, unlike to the multivariate case, when under some technical condition, the two estimators LL and NW have the same variance term, in this functional context, this statement requires some additional assumptions closely related to the functional structure of the data. Thus, we have opted in this contribution to express the variance term in a general form without supplementary conditions.

On the impact of our robust approach : In addition to the nice properties of the bias and the variance terms, the present approach has also two important features : The first one is the robustness and the second one is the generality of the model. Usually the robustness property of the M-estimators is controlled by the two fundamental constraints that are the convexity of the loss function ρ and the boundness of the score function ψ. The robustness property is very useful in practice, it allows to provide a nonparametric model not affected by the presence of the outliers or the heteroscedasticity phenomena. On the other hand, in this paper we have opted for a general presentation without the boundness condition. This consideration extends the scope of the present work to include a several usual cases. In order to highlight the generality of the consider model we cite some particular cases : First, if we replace $b = 0$ in (2.2) we obtain the definition of the classical functional robust regression (see, Attouch et al. (2009)).

Secondly, if we take $\rho(y) = y^2$ the model (2) becomes

$$\min_{(a,b)\in \mathbb{R}^2} \sum_{i=1}^{n} (Y_i - a - b\beta(X_i, x))^2 K\left(h^{-1}\delta(x, X_i)\right).$$

This estimate is the same as proposed by Barrientos-Marin et al.(2010) for the classical regression. Moreover, the real case treated by Fan et al. (1994) can be viewed as a particular case of this work. Indeed, it suffices to put, $\mathcal{F} = \mathbb{R}$ and $\beta(X, x) = \delta(x, X) = X - x$. Finally, we can say that the robust model studied in the present work is particularly interesting for its statistical proprieties (bias property, robustness) and its generality which allows to bring together a several usual cases.

Some prospects : Noting that even with the mild smoothness condition imposed on ρ, there exist some models which can not be include here. Such models require some additional tools. Among these models, we cite, for instance, the conditional quantiles and the Winsor regression. Thus the treatment of these models in an important prospect of this work. Finally, we mention that there exist more other prospects which can be studied in the future, such as the functional time series, the functional spatial case or the incomplete data case, ...

2.5 Appendix

Firstly, we state the following lemmas which are needed to establish our asymptotic results

Lemma 2.1. *Let $(V_n)_n$ a sequence of vectorial functions such that*

i) For all $\lambda \geq 1$ and a vector δ, we have, ${}^t(\delta)V_n(\lambda\delta) \leq {}^t(\delta)V_n(\delta)$.

ii) For a positive definite matrix D and vectorial sequences A_n such that $\mathbb{P}(\|A_n\| \geq A) \to 0$, for some $A > 0$ we have

$$\sup_{\|\delta\|\leq M} \|V_n(\delta) + \lambda_0 D\delta - A_n\| = o_p(1) \quad \textit{for} \quad \frac{A}{\lambda_0\lambda_1(D)} < M < \infty.$$

with $\lambda_1(D)$ is the minimum eigenvalue of D and $\lambda_0 > 0$. Then, for any vectorial sequence δ_n such that $V_n(\delta_n) = o_p(1)$ we have

$$\|\delta_n\| \leq M, \quad \textit{in probability.}$$

Furthermore, if $\|A_n\| = o_{a.co.}(1)$ and $\sup_{\|\delta\|\leq M} \|V_n(\delta) + \lambda_0 D\delta - A_n\| = o_{a.co.}(1)$. Then, for any vectorial sequence δ_n such that $V_n(\delta_n) = o_{a.co.}(1)$ we have

$$\|\delta_n\| \leq M, \quad \textit{almost completely.}$$

Now, for the proof of both of Theorems, we define, for $\delta = \begin{pmatrix} c \\ d \end{pmatrix}$ and $\delta_0 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$

the following functions

$$\varphi(\delta) = \psi(Y_i - ((c+a) + (h^{-1}d + b)\beta_i)),\ \varphi'(\delta) = \psi\left(Y_i - \left(\left(\frac{1}{\sqrt{n\phi_x(h)}}c + a\right) + \left(\frac{1}{h\sqrt{n\phi_x(h)}}d + b\right)\beta_i\right)\right)$$

and the following vectorial sequences

$$V_n(\delta) = \frac{1}{n\phi_x(h)}\sum_{i=1}^{n}\varphi(\delta)\begin{pmatrix}1\\ h^{-1}\beta_i\end{pmatrix}K_i,\ A_n = V_n(\delta_0),\ \delta_n = \begin{pmatrix}\widehat{a} - a\\ h(\widehat{b} - b)\end{pmatrix}$$

and

$$V_n'(\delta) = \frac{1}{\sqrt{n\phi_x(h)}}\sum_{i=1}^{n}\varphi'(\delta)\begin{pmatrix}1\\ h^{-1}\beta_i\end{pmatrix}K_i,\ A_n' = V_n'(\delta_0),\ \delta_n' = \sqrt{n\phi_x(h)}\delta_n.$$

Therefor, the proof of the first Theorem 2.1 is based on the application of the second part of Lemma 2.1 to (V_n, A_n, δ_n). While the Theorem 2.2 is obtained by applying the first part of Lemma 2.1 to (V_n', A_n', δ_n'). Clearly, for both cases the Condition (i) of Lemma 2.1 holds, because ψ is monotone increasing function.
Thus, Theorem 2.1 is consequence of the following Lemmas :

Lemma 2.2. *Under Hypotheses (H1), (H2), (H3(i)) and (H4)-(H6), we have*

$$\|A_n\| = O(h^{\min(k_1,k_2)}) + O\left(\left(\frac{\log n}{n\phi_x(h)}\right)^{1/2}\right), \quad a.co.$$

Lemma 2.3. *Under Hypotheses (H1), (H2), (H3(i)) and (H4)-(H6), we have*

$$\sup_{\|\delta\|\le M}\|V_n(\delta) + \lambda_0 D\delta - A_n\| = O(h^{\min(k_1,k_2)}) + O\left(\left(\frac{\log n}{n\phi_x(h)}\right)^{1/2}\right), \quad a.co.$$

with

$$D = \begin{pmatrix} K(1) - \int_{-1}^{1}K'(t)\chi_x(t)dt & K(1) - \int_{-1}^{1}(tK(t))'\chi_x(t)dt \\ K(1) - \int_{-1}^{1}(tK(t))'\chi_x(t)dt & K(1) - \int_{-1}^{1}(t^2K(t))'\chi_x(t)dt \end{pmatrix} \quad \text{and}\ \ \lambda_0 = \frac{\partial\Gamma}{\partial t}(x,\theta_x).$$

Concerning the asymptotic normality result we use the following Lemmas :

Lemma 2.4. *Under Hypotheses (H1), (H2), (H3(i)) and (H4)-(H6), we have*

$$\|A_n'\| = O_p(1)\ \text{and}\ \sup_{\|\delta'\|\le M}\|V_n(\delta) + \frac{\partial\Gamma}{\partial t}(x,\theta_x)D\delta' - A_n'\| = o_p(1).$$

Corollary 2.1. *Under Hypotheses of Lemma 2.4 , we have*

$$\widehat{\theta}_x - \theta_x = \frac{1}{\frac{\partial\Gamma}{\partial t}(x,\theta_x)(a_1a_3 - a_2^2)}\widehat{\Phi}(x) + o_p(1)$$

where $\widehat{\Phi}(x) = \dfrac{1}{n\phi_x(h)}\sum_{i=1}^{n}\varphi'(\delta_0)(A_3K_i - A_2h^{-1}\beta_iK_i)$ *with* $A_j = \phi_x^{-1}(h)h^{1-j}\mathbb{E}[\beta^{j-1}K]$; $j = 2, 3$.

Lemma 2.5. *Under the hypotheses of Theorem 2.2, we have,*

$$\mathbb{E}\widehat{\Phi}(x) = o(h) \quad \textit{and} \quad \sqrt{n\phi_x(h)}\sigma_1^{-1/2}\left(\widehat{\Phi}(x) - \mathbb{E}[\widehat{\Phi}(x)]\right) \xrightarrow{\mathcal{D}} \mathcal{N}(0,1) \quad \textit{as} \quad n \to \infty$$

where $\sigma_1 = \mathbb{E}\left[\psi^2(Y - \theta_x)|X = x\right]\left(a_3^2D_1 - 2a_2a_3D_2 + a_2^2D_3\right)$.

Acknowledgements. The authors would like to thank the Associate-Editor and the three anonymous reviewers for their valuable comments and suggestions which improved substantially the quality of this paper.

Références

Attouch, M., Laksaci, A., Ould Said, E. (2009). Asymptotic distribution of robust estimator for functional nonparametric mo dels. *Comm. Statist. Theory and Methods,* **38**, 1317-1335.

Barrientos-Marin, J., Ferraty, F.,Vieu, P. (2010). Locally modeled regression and functional data. *J. of Nonparametric Statistics,* **22**, 617-632

Baìllo, A. and Grané, A. (2009). Local linear regression for functional predictor and scalar response. *J. Multivariate Anal.,* **100**, 102-111.

Belarbi, F., Chemikh, S.,Laksaci, A. (2017). Asymptotic properties of the robust local linear regression with functional regressor.*Tech. report,*https://www.researchgate.net/profile/Ali_Laksaci.

Berlinet, A., Elamine, A., and Mas, A. (2011). Local linear regression for functional data. *Ann. Inst. Statist. Math.,* **63**, 1047-1075.

Boente, G., Manteiga, W.G., Péerez-Gonzàlez, A. (2009). Robust nonparametric estimation with missing data. *J. Statist. Plan. and Inference,* **139**, 571-592.

Boente, G.,Vahnovan, A. (2015) . Strong convergence of robust equivariant nonparametric functional regression estimators. *Statist. Probab. Lett.* **100**, 1-11.

Cai, Z. Ould Said., E. (2003). Local M-estimator for nonparametric time series. *Statist. Probab. Lett.* **65**, 433-449.

Chen, J., Zhang, L. (2009). Asymptotic properties of nonparametric M-estimation for mixing functional data, *J. Statist. Plan. and Inference,* **139**, 533-546.

Geenens, G. (2011) Curse of dimensionality and related issues in nonparametric functional regression. *Stat. Surv.*, **5**, 30-43.

Goia, A., Vieu, P. (2016). An introduction to recent advances in high/infinite dimensional statistics. *J. Multivariate Anal.,* **146**, 1-6.

Fan, J., Hu, T. C. and Truong, Y. K. (1994). Robust non-parametric function estimation. *Scand. J. Statist.* **21**, 433-446.

Fan, J. and Gijbels, I. (1996). Local polynomial modeling and its applications. *London, Chapman & Hall.*

Fan, J. (1993). Local Linear Regression Smoothers and Their Minimax Efficiencies. *The Annals of Statistics,* **21**, 196-216.

Ferraty, F. and Vieu, P. (2006). Nonparametric functional data analysis. Theory and Practice. *Springer Series in Statistics. New York.*

Horvàth, L. and Kokoszka, P. (2012). Inference for functional data with applications. *Springer Series in Statistics. Springer-Verlag, New York.*

Hsing, T. and Eubank, R. (2015). Theoretical foundations of functional data analysis, with an introduction to linear operators. *Wiley Series in Probability and Statistics. John Wiley & Sons, Chichester.*

Huber, P. J. (1964). Robust estimation of a location parameter. *Ann. Math. Statist.* **35**, 73- 101.

Kara-Zaitri, L., Laksaci, A., Rachdi, M. and Vieu, P, (2017a). Uniform in bandwidth consistency for various kernel estimators involving functional data. *J. of Nonparametric Statistics,*, **29** , 85-107.

Kara-Zaitri, L., Laksaci, A., Rachdi, M. and Vieu, P, (2017b). Data-driven kNN estimation in nonparametric functional data analysis. *J. Multivariate Anal.,* **153**, 176-188.

Stone C J, (2005). Nonparametric M-regression with free knot splines. *J. Statist. Plan. and Inference,* **130**, 183-206.

Zhang, J. (2014). Analysis of variance for functional data. *Monographs on Statistics and Applied Probability, 127. CRC Press, Boca Raton, FL.*

Zhou, Z. and Lin, Z. (2016). Asymptotic normality of locally modeled regression estimator for functional data. *J. of Nonparametric Statistics,* **28**, 116-131.

CHAPITRE 3

LOCAL LINEAR ESTIMATION OF THE NONPARAMETRIC TRIMMED REGRESSION IN FUNCTIONAL DATA

Ce chapitre a fait l'objet d'une publication internationale dans *Journal of Statistics & Economics.*

Local linear estimation of the nonparametric trimmed regression in functional data

Souheyla Chemikh and Faiza Belarbi

Laboratory of Statistics and Stochastic Processes
Department of Probability and Statistics
Djillali Liabes University
Sidi Bel Abbes 22000, Algeria
chemikh_souheyla@yahoo.com
faiza_belarbi@yahoo.fr

Abstract

In this work, we study the nonparametric estimation of the local linear regression model when the regressors are a functional random variables. We construct an estimator based on the local linear method using a trimmed approach. We establish the almost complete convergence rate as well as the asymptotic normality of this estimator.

Keywords : Functional data analysis, Local linear method, Robust estimation, Asymptotic normality, Almost complete convergence.

2000 Mathematics Subject Classification : 62F12, 62G20, 62G05.

3.1 Introduction

Nonparametric estimation is an important field of research in statistic, this field is based on the study of nonparametric models for random variables which take values in an infinite-dimensional space. This kind of variables well-known as functional variables in literature, we refer to Ferraty and Vieu (2006) for more discussions. In this work we focus on the nonparametric estimation of the trimmed regression by using the local linear method.

The trimmed estimator of the regression function outclasses the classical estimator which is introduced by Nadarya-watson(NW). In the particular, it is insensitive to the presence of outliers or heteroscedasticity because it is defined by minimizing a robust measure of the scatter of the residuals.(See Rousseeuw and Leroy (1987)).

It is well known that, the local linear technique has various advantage aver the classical kernel method. (see Fan and Gijbels(1996)) for more discussions on the importance of this approach).In the case of functional data analysis, the local linear method dates back to Bàillo and Grané(2009), they constructed a local linear estimator of the regression function and established the L_2 convergence rate of the constructed estimator when the explanatory variable is Hilbertian. Barrientos et al.(2010) investigated the asymptotic behavior (the rate of the almost complete convergence) of the locally modelled regression estimator of the regression function. we return to Berlinet et al (2011) for another local linear estimate of the regression operator. The estimator of the conditional density function based on the local modeling approach was proposed by Demongeot et al 2013, they studied the pointwise and uniform almost complete convergence (with the rate) of this estimator. Recently, Chouaf (2015) established the asymptotic mean square error of the nonparametric regression function. Massim and Mechab (2016) showed the almost complete convergence of the local linear estimation of the conditional hazard function when the explanatory variable takes values in a functional space. Kaid and Laksaci (2017) studied the almost complete convergence and the asymptotic normality of the local linear estimator of the conditional quantiles. Belarbi et al.(2018) proposed a robust local linear estimator of the regression function by showing the almost complete convergence (with the rate)and the asymptotic normality of the proposed estimator.

In this paper, we construct a new nonparametric robust estimator of the regression function based on the idea of LTS. Precisely, we use a score function discontinuous which corresponds to Huber-Type skipped mean.(See Rousseeuw and Leroy (1987), (p.135,181,190) and Hampel el al.(1986)). Specifically, we use also the functional local linear procedure proposed by Barrientos et al.(2010). Our estimator keeps the robustness of the trimmed approach and the advantages of the local linear method. Under a standards assumptions we study the almost complete convergence (with rate) and the asymptotic normality of this estimator for independent and identically distributed observations.

This work is organized as follows, Section 3.2 is dedicated to presenting our model. We fix the notations, hypotheses and state our main results in Section 3.3. An application to estimate the confidence interval is given in section 3.4. We give the proofs of the results in Section 3.5. Finally, in the last Section, we give our conclusion.

3.2 The trimmed local linear estimator

Let (X_1, Y_2),$\cdots$,(X_n, Y_n) be n copies of random vectors, independent and identically distributed as (X, Y) which is a random pair valued in $\mathcal{F} \times \mathbb{R}$, where $\mathcal{F}$ is a semi-metric space, of eventually infinite dimension and d denotes the semi-metric. For any x in $\mathcal{F}$, The nonparametric trimmed regression, denoted by θ_x is solution with respect to (w.r.t) t of following problem

$$\mathbb{E}\left[\psi(Y-t)|X=x\right]=0 \quad t\in\mathbb{R} \tag{3.1}$$

where $\psi(y) = y\mathbb{1}_{|y| \le q}$. $\mathbb{1}$ is indicator function, and $q = F^{-1}(1-\alpha/2)$ for $\alpha \in \left]0, \frac{1}{2}\right[$ where F is a cumulative distribution function which has a symmetric density.
The basic idea of trimmed approach is to replace the set of n residuals by the subset which contains just the residuals which are between -q and q. Precisely, it keeps the square function and eliminates a percentage of residuals.(See Rousseeuw and Leroy (1987)for more discussions).
Our main purpose of this work is to study the functional local linear estimate of the trimmed regression function by adopting the fast functional local modeling proposed by Barrientoset al.(2010) for which the function θ_x is approximated by

$$\forall z \qquad \text{in neighborhood of } x \qquad \theta_z = a + b\beta(x,z) \tag{3.2}$$

where a and b are estimated by $\widehat{a}$ and $\widehat{b}$ are solution of

$$\min_{(a,b)\in\mathbb{R}^2} \sum_{i=1}^{n} \rho(Y_i - a - b\beta(X_i,x))K(h^{-1}\delta(x,X_i)) \tag{3.3}$$

where ρ is the primitive of ψ and K is a kernel function and $h := h_n$ (to simplify the notation) is a sequence of positive real numbers with goes to zero as n goes to infinity and $\beta(\cdot,\cdot)$ is a know function from $\mathcal{F}^2$ into $\mathbb{R}$ such that, $\forall \xi \in \mathcal{F}, \beta(\xi,\xi) = 0$ and $\delta(\cdot,\cdot)$ is function of $\mathcal{F} \times \mathcal{F}$ such that $d(\cdot,\cdot) = |\delta(\cdot,\cdot)|$. It is clear that, under this consideration, we can write

$$\theta_x = a \;\text{ and }\; \widehat{\theta_x} = \widehat{a}.$$

(H1) For any $r > 0$, $\phi_x(r) := \phi_x(-r,r) > 0$ and there exists a function $\chi_x(\cdot)$ such that :

$$\forall t \in (-1,1), \qquad \lim_{h\to 0} \frac{\phi_x(th,h)}{\phi_x(h)} = \chi_x(t).$$

(H2) The function $\Gamma_\lambda(x,\cdot) := \mathbb{E}[\psi^\lambda(Y-\cdot)|X=x]$ is of class $\mathcal{C}^1$ on $[\theta_x - \delta, \theta_x + \delta]$: $\delta > 0$, and $\lambda \in \{1,2\}$. we put $\gamma(x,\cdot) := \dfrac{d}{dt}\Gamma_1(x,\cdot)$, such that :

$$\begin{cases} \forall (t_1,t_2) \in [\theta_x - \delta, \theta_x + \delta]^2, \;\forall (x_1,x_2) \in \mathcal{N}_x \times \mathcal{N}_x, \quad \text{and for } k_1,k_2 > 0 \\ \\ (i) \qquad |\Gamma_\lambda(x_1,t_1) - \Gamma_\lambda(x_2,t_2)| \le Cd^{k_1}(x_1,x_2) + |t_1-t_2|^{k_2}. \\ \\ (ii) \qquad |\gamma(x_1,t_1) - \gamma(x_2,t_2)| \le C'd^{k_1}(x_1,x_2) + |t_1-t_2|^{k_2}. \end{cases}$$

(H2') The variable $\delta(x,X)$ is $\sigma(\beta(x,X))-$ measurable and the two partial derivatives of the function, $\Upsilon_x(s,t) = \mathbb{E}[\Gamma_1(X,t)|\beta(x,X)=s]$, at $(0,\theta_x)$ exist.

(H3) The function $\beta(.,.)$ is such that :

$$\forall z \in \mathcal{F},\, C|\delta(x,z)| \le |\beta(z,x)| \le C'|\delta(x,z)| \text{ and } \sup_{u\in B(x,r)} |\beta(u,x) - \delta(x,u)| = o(r).$$

(H4) The kernel K is a positive, differentiable function which is supported within $(-1,1)$ such that

$$\begin{pmatrix} K(1)-\int_{-1}^{1}K'(t)\chi_x(t)dt & K(1)-\int_{-1}^{1}(tK(t))'\chi_x(t)dt \\ K(1)-\int_{-1}^{1}(tK(t))'\chi_x(t)dt & K(1)-\int_{-1}^{1}(t^2K(t))'\chi_x(t)dt \end{pmatrix} \text{ is a positive de definite matrix.}$$

(H5) The bandwidth h satisfies : $\lim_{n\to\infty}\dfrac{\log n}{n\phi_x(h)}=0.$

Our main results are summarized in the following Theorems

Theorem 3.1. *Under hypotheses (H1),(H2),(H3)-(H5) and if $\gamma(x,\theta_x)>0$ then*

$$|\widehat{\theta}_x-\theta_x| = O(h^{\min(k_1,k_2)}) + O\left(\left(\frac{\log n}{n\phi_x(h)}\right)^{1/2}\right), \quad a.co.$$

Theorem 3.2. *Under hypotheses (H1),(H2),(H2'),(H3)-(H5) and if $\gamma(x,\theta_x)>0$ then, for any $x\in\mathcal{A}$, we have*

$$\left(\frac{n\phi_x(h)}{\sigma^2(x)}\right)^{1/2}\left(\widehat{\theta}_x-\theta_x-o(h)\right)\xrightarrow{\mathcal{D}}\mathcal{N}(0,1) \quad as \quad n\to\infty.$$

where $\xrightarrow{\mathcal{D}}$ denotes the convergence in distribution,

$$\sigma^2(x)=\frac{\Gamma_2(x,\theta_x)}{(\gamma(x,\theta_x))^2}\frac{a_3^2D_1-2a_2a_3D_2+a_2^2D_3}{(a_1a_3-a_2^2)^2} \text{ and } \mathcal{A}=\{x\in\mathcal{F},\ \sigma^2(x)\neq 0\},$$

with

$$a_j=K(1)-\int_{-1}^{1}(s^{j-1}K(s))'\chi_x(s)ds \text{ and } D_j=K(1)-\int_{-1}^{1}(s^{j-1}K^2(s))'\chi_x(s)ds, \text{ for } j=1,2,3.$$

We can remove the bias term $o(h)$ but we need an additional assumption on the bandwidth parameter h

Corollary 3.1. *Under Hypotheses (H1),(H2) and (H3)-(H5)and if the bandwidth parameter h satisfies $nh^{2\min(k_1,k_2)}\phi_x(h)\to 0$ as$\to\infty$ and if $\gamma(x,\theta_x)>0$. We have*

$$\left(\frac{n\phi_x(h)}{\sigma^2(x)}\right)^{1/2}\left(\widehat{\theta}_x-\theta_x\right)\xrightarrow{\mathcal{D}}\mathcal{N}(0,1) \quad as \quad n\to\infty.$$

3.3 Application

3.3.1 Conditional Confidence Interval

It is well clear that the most important application of the asymptotic normality result is to build confidence interval for the true value of θ_x given that curve $X=x$. On the one hand, the building of the confidence intervals requires an estimation of standard $\sigma(x)$

deviation and the bias term $o(h)$ because these terms contain an unknown functions $\Gamma_2(x,\theta_x)$,$\gamma(x,\theta_x)$,a_j and D_j. In our case, we neglect the bias term and we focus on the estimation of $\sigma(x)$.
On the other hand, the building of these intervals need also the estimation of $\phi_x(h)$ which can be empirically estimated by :

$$\widehat{\phi_x}(h) = \frac{\sharp\{i : |\delta(x,X_i)| \leq h\}}{n}$$

where $\sharp\{A\}$ denotes the cardinal number of the set A(see Demongeot et al.(2014)). The estimation of $\sigma(x)$ can be obtained using the estimators $\widehat{\Gamma}_2(x,\widehat{\theta}_x)$,$\widehat{\gamma}(x,\widehat{\theta}_x)$,$\widehat{a_j}$ and $\widehat{D_j}$ of $\Gamma_2(x,\theta_x)$,$\gamma(x,\theta_x)$,a_j and D_j. respectively. We have

$$\widehat{\Gamma}_2(x,\widehat{\theta}_x) = \frac{\sum_{i=1}^n K(h^{-1}\delta(x,X_i))\psi^2(Y_i-\widehat{\theta}_x)}{\sum_{i=1}^n K(h^{-1}\delta(x,X_i))}$$

and

$$\widehat{\gamma}(x,\widehat{\theta}_x) = \frac{\sum_{i=1}^n K(h^{-1}\delta(x,X_i))\dfrac{d}{dt}\psi(Y_i-\widehat{\theta}_x)}{\sum_{i=1}^n K(h^{-1}\delta(x,X_i))}$$

Furthermore, we estimate the quantities a_j and D_j empirically by

$$\widehat{a_j} = \frac{1}{n\widehat{\phi_x}(h)}\sum_{i=1}^{n}\left(\frac{\delta(x,X_i)}{h}\right)^{j-1} K\left(\frac{\delta(x,X_i)}{h}\right).$$

and

$$\widehat{D_j} = \frac{1}{n\widehat{\phi_x}(h)}\sum_{i=1}^{n}\left(\frac{\delta(x,X_i)}{h}\right)^{j-1} K\left(\frac{\delta(x,X_i)}{h}\right)^2.$$

The last estimations is justified by the fact that, under (H1),(H3) and (H5). we have (see Demongeot et al.(2014))

$$\frac{\mathbb{E}[\beta_i^{j-1}K_i]}{h^{j-1}\phi_x(h)} \to a_j \qquad\qquad \frac{\mathbb{E}[\beta_i^{j-1}K_i^2]}{h^{j-1}\phi_x(h)} \to D_j.$$

It follows that $\widehat{\sigma(x)} = \left(\dfrac{\widehat{\gamma}(x,\widehat{\theta}_x)}{(\widehat{\gamma}(x,\widehat{\theta}_x))^2}\dfrac{\widehat{a_3}^2\widehat{D_1} - 2\widehat{a_2}\widehat{a_3}\widehat{D_2} + \widehat{a_2}^2\widehat{D_3}}{(\widehat{a_1}\widehat{a_3} - \widehat{a_2}^2)^2}\right)^{1/2}$.Finally, we can obtain confidence intervals of θ_x at asymptotic level $(1-\vartheta)$

$$\left[\widehat{\theta}_x - t_{1-\vartheta/2}\times\left(\frac{\widehat{\sigma(x)}^2}{n\widehat{\phi_x}(h)}\right)^{1/2}, \widehat{\theta}_x + t_{1-\vartheta/2}\times\left(\frac{\widehat{\sigma(x)}^2}{n\widehat{\phi_x}(h)}\right)^{1/2}\right],$$

where$t_{1-\vartheta/2}$ denotes $(1-\vartheta/2)$ quantile of the standard normal distribution.

Remark 3.1. *It should be noted that, we can neglect the estimation of $\phi_x(h)$ because* *it does not appear in the calculations of the confidence interval by simplification.*

3.4 Appendix

Firstly, we state the following lemma which is needed to established the both asymptotic results.(see Belarbi et al.(2017))

Lemma 3.1. *Let $(V_n)_n$ a sequence of vectorial functions such that*

i) For all $\lambda \geq 1$ and a vector δ, we have, ${}^t(\delta)V_n(\lambda\delta) \leq {}^t(\delta)V_n(\delta)$.

ii) For a positive definite matrix D and vectorial sequences A_n such that $\mathbb{P}(\|A_n\| \geq A) \to 0$, for some $A > 0$ we have

$$\sup_{\|\delta\|\leq M} \|V_n(\delta) + \lambda_0 D\delta - A_n\| = o_p(1) \quad \textit{for} \quad \frac{A}{\lambda_0\lambda_1(D)} < M < \infty.$$

with $\lambda_1(D)$ is the minimum eigenvalue of D and $\lambda_0 > 0$. Then, for any vectorial sequence δ_n such that $V_n(\delta_n) = o_p(1)$ we have

$$\|\delta_n\| \leq M, \quad \textit{in probability.}$$

Furthermore, if $\|A_n\| = o_{a.co.}(1)$ and $\sup_{\|\delta\|\leq M} \|V_n(\delta) + \lambda_0 D\delta - A_n\| = o_{a.co.}(1)$. Then, for any vectorial sequence δ_n such that $V_n(\delta_n) = o_{a.co.}(1)$ we have

$$\|\delta_n\| \leq M, \quad \textit{almost completely.}$$

Now, for the proof of the theorems we define, for all $\delta = \begin{pmatrix} c \\ d \end{pmatrix}$, ,$\delta_0 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$ and$\lambda = \sqrt{n\phi_x(h)}$ the following functions

$$\begin{aligned}\varphi_i(\delta) &= \psi_i(Y_i - ((c+a) + (h^{-1}d + b)\beta_i)) \\ &= (Y_i - ((a+c) + (h_n^{-1}d + b)\beta_i))\mathbb{1}_{|Y_i - ((a+b) + (h_n^{-1}d+b)\beta_i)| \leq q}\end{aligned}$$

and

$$\begin{aligned}\varphi_i'(\delta) &= \psi(Y_i - ((\lambda^{-1}c + a) - (h^{-1}\lambda^{-1}d + b)\beta_i)) \\ &= (Y_i - ((\lambda^{-1}c + a) - (h^{-1}\lambda^{-1}d + b)\beta_i))\mathbb{1}_{|Y_i - ((\lambda^{-1}c+a) - (h^{-1}\lambda^{-1}d+b)\beta_i)| \leq q}\end{aligned}$$

and the following vectorial sequences

$$V_n(\delta) = \frac{1}{n\phi_x(h)} \sum_{i=1}^{n} \varphi_i(\delta) \begin{pmatrix} 1 \\ h^{-1}\beta_i \end{pmatrix} K_i, \;\; A_n = V_n(\delta_0), \;\; \delta_n = \begin{pmatrix} \widehat{a} - a \\ h(\widehat{b} - b) \end{pmatrix}$$

and

$$V_n'(\delta)=\frac{1}{\sqrt{n\phi_x(h)}}\sum_{i=1}^{n}\varphi_i'(\delta)\begin{pmatrix}1\\ h^{-1}\beta_i\end{pmatrix}K_i,\ A_n'=V_n'(\delta_0),\ \delta_n'=\sqrt{n\phi_x(h)}\delta_n.$$

Now, the proof of the result of Theorem 3.1 is based on the application of the second part of Lemma 3.1 to (V_n, A_n, δ_n) . While the seconde result is obtained by applying the first part of Lemma 3.1 to (V_n', A_n', δ_n').

Lemma 3.2. *Under Hypotheses (H1),(H2) and (H3)-(H5), we have*

$$\|A_n\|=O(h^{\min(k_1,k_2)})+O\left(\left(\frac{\log n}{n\phi_x(h)}\right)^{1/2}\right),\quad a.co.$$

Lemma 3.3. *Under Hypotheses (H1),(H2) and (H3)-(H5), we have*

$$\sup_{\|\delta\|\le M}\|V_n(\delta)+\lambda_0 D\delta-A_n\|=O(h^{\min(k_1,k_2)})+O\left(\left(\frac{\log n}{n\phi_x(h)}\right)^{1/2}\right),\quad a.co.$$

with

$$D=\begin{pmatrix}K(1)-\int_{-1}^{1}K'(t)\chi_x(t)dt & K(1)-\int_{-1}^{1}(tK(t))'\chi_x(t)dt\\ K(1)-\int_{-1}^{1}(tK(t))'\chi_x(t)dt & K(1)-\int_{-1}^{1}(t^2K(t))'\chi_x(t)dt\end{pmatrix}\quad and\ \ \lambda_0=\gamma(x,\theta_x).$$

Concerning the asymptotic normality result given in Theorem 3.2 and in Corollary 3.1 we use the following Lemmas :

Lemma 3.4. *Under Hypotheses (H1),(H2), and (H3)-(H5), we have*

$$\|A_n'\|=O_p(1)\ \ and\ \ \sup_{\|\delta'\|\le M}\|V_n(\delta)+\gamma(x,\theta_x)D\delta'-A_n'\|=o_p(1).$$

Proposition 3.1. *(see Belarbiet al.(2017)) Under Hypotheses of Lemma 3.4 we have :*

$$\widehat{\theta_x}-\theta_x=\frac{1}{\gamma(x,\theta_x)(a_1a_3-a_2^2)}\widehat{\Psi}(x)+o(1)$$

where $\widehat{\Psi}(x)=\frac{1}{n\phi_x(h)}\sum_{i=1}^{n}\varphi_i(\delta_0)(A_3K_i-A_2h_n^{-1}\beta_iK_i)$ *with* $A_j=\phi_x(h)^{-1}h^{1-j}\mathbb{E}[\beta^{j-1}K]=a_j+o(1).$

for $j=2,3$.
Then,we can write

$$\sqrt{n\phi_x(h)}(\widehat{\theta_x}-\theta_x)=\frac{1}{\gamma(x,\theta_x)(a_1a_3-a_2^2)}\left(\sqrt{n\phi_x(h)}(\widehat{\Psi}(x)-\mathbb{E}[\widehat{\Psi}(x)])\right)$$
$$+\frac{\sqrt{n\phi_x(h)}}{\gamma(x,\theta_x)(a_1a_3-a_2^2)}\mathbb{E}[\widehat{\Psi}(x)].\qquad(3.4)$$

Then, to state asymptotic normality, we show that the first term of the right-hand side of (3.4) suitably normalized is asymptotically normally distributed. and the second term is equal to $\sqrt{n\phi_x(h)}o(h)$. On the other hand, we obtain the asymptotic normality version given in Corollary 3.1 if we show that the second term of the right-hand side of (3.4) is asymptotically negligible under an additional assumption of the bandwidth parameter h.

Lemma 3.5. *Under the hypotheses of Theorem 3.2, we have,*

$$\mathbb{E}\widehat{\Psi}(x) = o(h) \quad and \quad \sqrt{n\phi_x(h)}\sigma_1^{-1/2}\left(\widehat{\Psi}(x) - \mathbb{E}[\widehat{\Psi}(x)]\right) \xrightarrow{\mathcal{D}} \mathcal{N}(0,1) \quad as \quad n \to \infty$$

where $\sigma_1 = \mathbb{E}\left[\psi^2(Y-\theta_x)|X=x\right]\left(a_3^2 D_1 - 2a_2a_3D_2 + a_2^2D_3\right)$.

Lemma 3.6. *Under Under Hypotheses (H1),(H2) and (H3)-(H5),if the bandwidth parameter h satisfies $nh^{2\min(k_1,k_2)}\phi_x(h) \to 0$ as$\to \infty$, we have*

$$\frac{\sqrt{n\phi_x(h)}}{\gamma(x,\theta_x)(a_1a_3^2 - a_2^2)^2}\mathbb{E}[\widehat{\Psi}(x)] \longrightarrow 0, \qquad as \quad n \longrightarrow +\infty$$

Proof of Lemmas

Proof of Lemma 3.2

Firstly, we write
$$A_n - \mathbb{E}[A_n] = \begin{pmatrix} A_n^1 := \dfrac{1}{n\phi_x(h)}\displaystyle\sum_{i=1}^n Z_i^1 \\ A_n^2 := \dfrac{1}{nh\phi_x(h)}\displaystyle\sum_{i=1}^n Z_i^2 \end{pmatrix}.$$

Where

$$Zi^1 = \varphi_i(\delta_0)K_i - \mathbb{E}[\varphi_i(\delta_0)K_i]. \quad \text{and} \quad Z_i^2 = \varphi_i(\delta_0)\beta_iK_i - \mathbb{E}[\varphi_i(\delta_0)\beta_iK_i].$$

By using the bounded of φ_i and K we have

$$|Z_i^1| \le C \qquad and \qquad |Z_i^2| \le C'h$$

Moreover

$$\mathbb{E}[Z_i^1]^2 \le C\phi_x(h) \qquad \text{and} \qquad \mathbb{E}[Z_i^2]^2 \le C'h^2\phi_x(h)$$

By applying the exponential inequality for unbounded variables (see Ferraty and Vieu (2006)). We obtain

$$A_n^1 - \mathbb{E}[A_n^1] = O_{a.co}\left(\sqrt{\frac{\log(n)}{n\phi_x(h)}}\right) \qquad \text{and} \qquad A_n^2 - \mathbb{E}[A_n^2] = O_{a.co}\left(\sqrt{\frac{\log(n)}{n\phi_x(h)}}\right).$$

On the other hand, by (H2), we obtain

$$\begin{aligned}\mathbb{E}[A_n^1] &\leq \frac{1}{\phi_x(h)}|\Gamma_1(x,\theta_x) - \Gamma_1(X, a+b\beta_1)|]\\ &= O(h^{k_1}) + o(h^{k_2})\end{aligned}$$

Similarly

$$\mathbb{E}[A_n^2] = O(h^{k_1}) + o(h^{k_2})$$

Which achieves the proof of Lemma

Proof of Lemma 3.3

We have to prove that

$$\sup_{||\delta||\leq M} ||\mathbb{E}[V_n(\delta) - A_n] + \gamma(x,\theta_x)D\delta|| = O(h^{\min(k_1,k_2)}) \tag{3.5}$$

$$\sup_{||\delta||\leq M} ||V_n(\delta) - A_n - \mathbb{E}[V_n(\delta) - A_n]|| = O_{a.co.}\left(\sqrt{\frac{\log(n)}{n\phi_x(h)}}\right). \tag{3.6}$$

For the first result, we write

$$V_n(\delta) - A_n = \begin{pmatrix} W_n^1(\delta) := \dfrac{1}{n\phi(h)}\displaystyle\sum_{i=1}^n(\varphi_i(\delta) - \varphi_i(\delta_0))K_i. \\ W_n^2(\delta) := \dfrac{1}{nh\phi(h)}\displaystyle\sum_{i=1}^n(\varphi_i(\delta) - \varphi_i(\delta_0))\beta_i K_i. \end{pmatrix}$$

Observe that, by (H2) we get

$$\begin{aligned}\mathbb{E}[W_n^1(\delta)] &= \frac{1}{\phi_x(h)}\mathbb{E}[(\Gamma_1(x,(c+a)+(h^{-1}d+b)\beta_1) - \Gamma_1(x,a+b\beta_1))K_1] + O(h^{k_1})\\ &= \frac{1}{\phi_x(h}\mathbb{E}\left[\gamma(x,a+b\beta_1)\left(1,h^{-1}\beta_1\right)\delta K_1\right] + O(h^{k_1}) + o(||\delta||)\\ &= \gamma(x,\theta_x)\frac{1}{\phi_x(h)}\left(\mathbb{E}[K_1], h^{-1}\mathbb{E}[\beta_i K_1]\right)\delta + O(h^{\min(k_1,k_2)}) + o(||\delta||).\end{aligned}$$

and by the same ideas we calculate $\mathbb{E}[W_n^2(\delta)]$. Therefore,

$$\mathbb{E}[V_n(\delta) - A_n] = \gamma(x,\theta_x)\frac{1}{\phi_x(h)}\begin{pmatrix}\mathbb{E}[K_i] & \mathbb{E}[K_i h^{-1}\beta_i]\\ \mathbb{E}[K_i h^{-1}\beta_i] & \mathbb{E}[h^{-2}\beta_i^2 K_i^2]\end{pmatrix}\delta + O(h^{\min(k_1,k_2)}) + o(||\delta||).$$

Using the same ideas of Demongeot et al.(2013), under the seconde part of (H3) that

$$h_n^{-a}\mathbb{E}[\beta^a K_i^b] = \phi_x(h)\left(K^b(1) - \int_{-1}^1 (u^a K(u))'\chi_x(u)du\right) + o(\phi_x(h)).$$

It follow that

$$\sup_{||\delta|| \leq M} ||\mathbb{E}[V_n(\delta) - A_n] + \gamma(x, \theta_x) D\delta + o(||\delta||)|| = O(h^{\min(k_1, k_2)}).$$

Which implies the result (3.5).Concerning (3.6), we use the compactness of the ball $B(0, M)$ in $\mathbb{R}^2$ and we write

$$B(0, M) \subset \bigcup_{j=1}^{d_n} B(\delta_j, l_n) \quad \delta_j = \begin{pmatrix} c_j \\ d_j \end{pmatrix}, \quad \text{and} \quad l_n = d_n = n^{-1/2}.$$

Let $j(\delta) = \operatorname{argmin}_j |\delta - \delta_j|$, then can write

$$\sup_{||\delta|| \leq M} ||V_n(\delta) - A_n - \mathbb{E}[V_n(\delta) - A_n]|| \leq \sup_{||\delta|| \leq M} ||V_n(\delta) - V_n(\delta_j)||$$

$$+ \sup_{||\delta|| \leq M} ||V_n(\delta_j) - A_n - \mathbb{E}[V_n(\delta_j) - A_n]|| + \sup_{||\delta|| \leq M} ||\mathbb{E}[V_n(\delta) - V_n(\delta_j)]||.$$

Now, we trait the first term we use the fact that K is bounded

$$\sup_{||\delta|| \leq M} ||V_n(\delta) - V_n(\delta_j)|| \leq \frac{1}{n\phi_x(h)} \sum_{i=1}^{n} \widetilde{\Delta}_i$$

It is clear that the function $\varphi_i(.)$ is locally Lipschitzian on $[-q, q]$, then we can write

$$\widetilde{\Delta}_i = \sup_{||\delta|| \leq M} |\varphi(\delta) - \varphi(\delta_j)| \left\| \begin{pmatrix} 1 \\ h_n^{-1}\beta_i \end{pmatrix} \right\| K_i \leq C l_n$$

Then

$$\sup_{||\delta|| \leq M} ||V_n(\delta) - V_n(\delta_j)|| \leq \frac{C l_n}{\phi_x(h)} = o\left(\sqrt{\frac{\log(n)}{n\phi_x(h)}}\right)$$

Finally

$$\sup_{||\delta|| \leq M} ||V_n(\delta) - V_n(\delta_j)|| = O_{a.co.}\left(\sqrt{\frac{\log(n)}{n\phi_x(h)}}\right).$$

Concerning the last term, analogously to first term and we use the fact that $\mathbb{E}[K_i] = O(\phi_x(h))$(see Ferraty and Vieu (2006)) we obtain

$$\sup_{||\delta|| \leq M} ||\mathbb{E}[V_n(\delta) - V_n(\delta_j)]|| = o_{a.co}\left(\sqrt{\frac{\log(n)}{n\phi_x(h)}}\right).$$

Now,we treat the quantity

$$\sup_{||\delta|| \leq M} ||V_n(\delta_j) - A_n - \mathbb{E}[V_n(\delta_j) - A_n]||$$

we set

$$V_n(\delta_j) - A_n - \mathbb{E}[V_n(\delta_j) - A_n] = \begin{pmatrix} W_n^1(\delta_j) := \dfrac{1}{n\phi_x(h)} \sum_{i=1}^{n} \Lambda_i^1 \\ W_n^2(\delta_j) := \dfrac{1}{nh\phi_x(h)} \sum_{i=1}^{n} \Lambda_i^1 \end{pmatrix}$$

where $\qquad \Lambda_i^1 = (\varphi_i(\delta_j) - \varphi_i(\delta_0))K_i - \mathbb{E}[(\varphi_i(\delta_j) - \varphi_i(\delta_0))K_i].$

and $\qquad \Lambda_i^2 = (\varphi_i(\delta_j) - \varphi_i(\delta_0))\beta_i K_i - \mathbb{E}[(\varphi_i(\delta) - \varphi_i(\delta_0))\beta_i K_i].$

thanks to the boundedness of K and $\varphi_i(.)$ we have

$$|\Lambda_i^1| \leq C \qquad |\Lambda_i^1| \leq C'$$

It is clear that

$$\mathbb{E}[\Lambda_i^1]^2 \leq C\phi_x(h) \qquad \mathbb{E}[\Lambda_i^2]^2 \leq C' h\phi_x(h)$$

Now, we apply the exponential inequality for unbounded variables we have

$$\sup_{||\delta|| \leq M} \|V_n(\delta_j) - A_n - \mathbb{E}[V_n(\delta_j) - A_n]\| = O_{a.co}\left(\sqrt{\frac{\log n}{n\phi_x(h)}}\right).$$

which complete the proof of (3.6)

Proof of Lemma 3.4

Concerning the first part, it suffices to prove that

$$n\phi_x(h) Var[e_1 A_n^1 + e_2 A_n^2] \longrightarrow Cst \quad for \quad \begin{pmatrix} e_1 \\ e_2 \end{pmatrix} \in \mathbb{R}^2$$

where A_n^1 and A_n^2 are defined in the proof of Lemma 3.2.

and $\qquad Var[e_1 A_n^1 + e_2 A_n^2] = Var[e_1 A_n^1] + Var[e_2 A_n^2] + 2e_1 e_2 Cov(A_n^1, A_n^2).$

By using (H2) we have

$$\frac{1}{\phi_x(h)} Var(Z_1^1) \to \Gamma_2(x, \theta_x)\left(K^2(1) - \int_{-1}^{1} (K^2(t))' \chi_x(t) dt\right).$$

$$\frac{1}{h^2\phi_x(h)} Var(Z_1^2) \to \Gamma_2(x, \theta_x)\left(K^2(1) - \int_{-1}^{1} (t^2 K^2(t))' \chi_x(t) dt\right).$$

$$\frac{1}{h\phi_x(h)} Cov(Z_1^1 Z_1^2,) \to \Gamma_2(x, \theta_x)\left(K^2(1) - \int_{-1}^{1} (tK^2(t))' \chi_x(t) dt\right).$$

The proof of the second part is based on the same assertions of Lemma 3.3, by following the same line as (3.5) we have

$$\mathbb{E}[V_n'(\delta)-A_n'] = \gamma(x,\theta_x)\frac{1}{\phi_x(h)}\begin{pmatrix} \mathbb{E}[K_i] & \mathbb{E}[K_ih^{-1}\beta_i] \\ \mathbb{E}[K_ih^{-1}\beta_i] & \mathbb{E}[h^{-2}\beta_i^2K_i^2] \end{pmatrix}\delta+O(h^{\min(k_1,k_2)})+o(||\delta||).$$

Now, for the dispersion the second part we keep the notation of lemma 3.3 with anther choice of l_nd_n. Specifically we put $l_n = d_n^{-1} = O\left(\frac{1}{\log n}\right)$ and we write

$$\sup_{||\delta||\le M}||V_n'(\delta) - A_n' - \mathbb{E}[V_n'(\delta) - A_n']|| \le \sup_{||\delta||\le M}||V_n'(\delta) - V_n'(\delta_j)||$$

$$+\sup_{||\delta||\le M}||V_n'(\delta_j) - A_n' - \mathbb{E}[V_n'(\delta_j) - A_n']|| + \sup_{||\delta||\le M}||\mathbb{E}[V_n'(\delta) - V_n'(\delta_j)]||.$$

By simple calculation and under (H1),(H3)-(H4), and we use the fact that the function ψ is Lipschitzian on $[-q,q]$ we have

$$\sup_{||\delta||\le M}||V_n'(\delta) - V_n'(\delta_j)|| = o_p(1) \quad \text{and} \quad \sup_{||\delta||\le M}||\mathbb{E}[V_n'(\delta) - V_n'(\delta_j)]|| \le Cl_n = o(1). \tag{3.7}$$

Furthermore, for the quantity

$$V_n'(\delta_j) - A_n' - \mathbb{E}[V_n'(\delta_j) - A_n'] = \begin{pmatrix} \sqrt{n\phi_x(h)}G_n^1(\delta_j) \\ \sqrt{n\phi_x(h)}G_n^2(\delta_j) \end{pmatrix},$$

where

$$\begin{cases} G_n^1(\delta_j) := \dfrac{1}{n\phi_x(h)}\displaystyle\sum_{i=1}^n \Delta_i^1 \\ G_n^2(\delta_j) := \dfrac{1}{nh\phi_x(h)}\displaystyle\sum_{i=1}^n \Delta_i^2 \end{cases}$$

and

$$\Delta_i^1 = (\varphi_i'(\delta_j)-\varphi_i(\delta_0))K_i-\mathbb{E}[(\varphi_i'(\delta_j)-\varphi_i(\delta_0))K_i] \quad \Delta_i^2 = (\varphi_i'(\delta_j)-\varphi_i(\delta_0))\beta_iK_i-\mathbb{E}[(\varphi_i'(\delta_j)-\varphi_i(\delta_0))\beta_iK_i].$$

$$\begin{aligned} Var[\sqrt{n\phi_x(h)}(G_n^1(\delta_j) - \mathbb{E}[G_n^1(\delta_j)])] &= O\left(\frac{1}{n\phi_x(h)}\right) \\ Var[\sqrt{n\phi_x(h)}(G_n^2(\delta_j) - \mathbb{E}[G_n^2(\delta_j)])] &= O\left(\frac{1}{n\phi_x(h)}\right) \end{aligned} \tag{3.8}$$

Thus, the proof of this Lemma is consequence of (H6), (3.7) and (3.8).

Proof of Lemma 3.5

For the first result :(see Belarbi et al.(2017)), and under (H2') we show that

$$\mathbb{E}[\widehat{\Psi}(x)] = \mathbb{E}[\Upsilon_x(\beta,\theta+a+\beta)(A3K - A_2h^{-1}K)] = o(h).$$

The proof of the seconde result is based on the version of central limit Theorem given in (see Loève (1963), p.175) in the first step we consider the asymptotic of the variance term.

$$\begin{aligned} n\phi_x(h)Var[\widehat{\Psi}(x)] &= \frac{1}{n\phi_x(h)}\sum_{i=1}^{n} Var[A_3K_1 - A_2h^{-1}\beta_1K_1)\varphi(\delta_0)] \\ &= \frac{1}{\phi_x(h)}Var[\varphi(\delta_0)(A_3K_1 - A_2h^{-1}\beta_1K_1)] \\ &= \frac{1}{\phi_x(h)}\mathbb{E}[(A_3K_1 - A_2h^{-1}\beta_1K_1)^2]\mathbb{E}\left[\frac{(\varphi(\delta_0))^2(A_3K_1 - A_2h^{-1}\beta_1K_1)^2}{\mathbb{E}[(A_3K_1 - A_2h^{-1}\beta_1K_1)^2]}\right] \\ &- \frac{1}{\phi_x(h)}\mathbb{E}^2\left[\varphi(\delta_0)(A_3K_1 - A_2h^{-1}\beta_1K_1)\right]. \end{aligned}$$

By using (H2i) and (H2') we obtain

$$\frac{1}{\phi_x(h)}\mathbb{E}[(A_3K_1 - A_2h^{-1}\beta_1K_1)^2]\mathbb{E}\left[\frac{(\varphi(\delta_0))^2(A_3K_1 - A_2h^{-1}\beta_1K_1)^2}{\mathbb{E}[(A_3K_1 - A_2h^{-1}\beta_1K_1)^2]}\right] \to \Gamma_2(x,\theta_x)(a_3^2D_1 - 2a_2a_3D_2 + a_2^2D_3)$$

and
$$\frac{1}{\phi_x(h)}\mathbb{E}^2\left[\varphi(\delta_0)(A_3K_1 - A_2h^{-1}\beta_1K_1)\right] = o(h).$$

It follows that

$$n\phi_x(h)Var[\widehat{\Psi}(x)] = \Gamma_2(x,\theta_x)\left(a_3^2D_1 - 2a_2a_3D_2 + a_2^2D_3\right). \quad as \quad n \to \infty. \tag{3.9}$$

Secondly, by using the same ideas of Attouch et al.(2009) we state the required conditions of Liapounov's Theorem that is

$$\frac{\sum_{i=1}^{n}\mathbb{E}[|L_i(x) - \mathbb{E}[L_i(x)]|^{2+\delta}]}{\left(Var\left(\sum_{i=1}^{n}L_i(x)\right)\right)^{(2+\delta)/2}} \to 0 \quad as \quad n \to \infty. \quad for\ some \quad \delta > 0 \tag{3.10}$$

where

$$L_i(x) = \frac{1}{n\phi_x(h)}(a_3K_i - a_2h^{-1}\beta_iK_i\varphi(\delta_0))$$

which concludes the proof.

Proof of Lemma 3.6

It is clear that
$$\frac{1}{|\gamma(x,\theta_x)(a_1a_3 - a_2^2)|} \le C.$$

Moreover

$$\mathbb{E}[\widehat{\Psi}(x)] \leq \frac{C}{\mathbb{E}[A_3K_1 - A_2h^{-1}\beta_1K_1]}|\mathbb{E}[\varphi(\delta_0)(A_3K_1 - A_2h^{-1}K_1)]|$$

By simple calculations using (H3) we get :

$$\begin{aligned}\left|\mathbb{E}\left[\frac{\varphi(\delta_0)(A_3K_1 - A_2h^{-1}K_1)}{\mathbb{E}[A_3K_1 - A_2h^{-1}\beta_1K_1]}\right] - \Gamma_1(x,\theta_x)\right| &\leq \frac{C}{\mathbb{E}[A_3K_1 - A_2h^{-1}K_1]} \\ &\times \mathbb{E}[A_3K_1 - A_2h^{-1}\beta_1K_1|\Gamma_1(X_1, a + b\beta_1) - \Gamma_1(x,\theta_x)|] \\ &\leq Ch^{\min(k_1,k_2)}.\end{aligned}$$

Finely, we have

$$\frac{\sqrt{n\phi_x(h)}}{\gamma(x,\theta_x)(a_1a_3 - a_2^2)^2}\mathbb{E}[\widehat{\Psi}(x)] \leq C\sqrt{nh_n^{2\min(k_1,k_2)}\phi_x(h)} \longrightarrow 0 \qquad as \quad n \to \infty.$$

3.5 Conclusion

Our objective in this talk was to study the regression model in the case where the explanatory variable X is functional by using the trimmed approach and the local linear method. The results obtained (the almost complete convergence rate and the asymptotic normality) confirm that the trimmed local linear estimator of the regression function is statistically good.

Références

Attouch, M.,Laksaci, A., Ould Saïd, E.(2009).Asymptotic distribution of robust estimator for functional models.*Journal Korean statistical society,* **38,** 1317-1335.

Baìllo, A., Grané, A.(2009).Local linear regression for functional predictor and scalar response.*Multivariate Analysis,* **100,** 102-111.

Barrientos-Marin,J. and Ferray, F. and Vieu,P.(2010).Locally Modelled Regression And Functional Data.*Nonparametric Statistics,* **22,** 617-632.

Belarbi,F., Chemikh,S.,Laksaci, A.(2017).Asymptotic properties of the robust local linear regression with functional regressor.*Tech. report,* https://www.researchgate.net/profile/Ali_Laksaci.

Belarbi,F.,Chemikh,S., Laksaci, A.(2018).Local linear estimate of nonparametric robust regression in function data.*Statistic and Probability letters,* **134,** 128–133.

Berlinet, A., Elamine, A., and Mas, A. (2011). Local linear regression for functional data. *Ann. Inst. Statist. Math,* **63**, 1047-1075.

Chouaf, A. (2015).Modelization local linear regression for functional random variables. *International Journal of Statistics & Economics,* **16**, 54-68.

Demongeot, J., Laksaci, A.,Madani, F. and Rachdi, M.(2013).Modelization local linear regression for functional random variables. *Statistics,* **47**, 26-44.

Demongeot, J., Laksaci, A., Rachdi, M., Abdali, A. and Rahmani, S.(2014).On the Local Linear Modelization of the Conditional Distribution for Functional Data. *Sankhya A,* **67**, 328-355.

Fan, J. and Gijbels, I. (1996). Local polynomial modeling and its applications. *London, Chapman & Hall.*

Ferraty, F. and Vieu, P. (2006). Nonparametric functional data analysis. Theory and Practice. *Springer Series in Statistics. New York.*

Hampel,F.R., Ronchetti,E.M,. Rousseeuw, P. J., Stahel.,W.A. (1986). Robust Statistics The Approach Based on Influence Functions. *John Wiley & Sons, New York.*

Kaid,Z.,Laksaci,A.(2017).Functional quantile regression :local linear modelisation. *Functional Statistics and Related Fields,* 155-156.

Loéve, M.(1963).Probability Theory.*3rd ed. Princeton, D. van Nostrand, N.J.-Toronto-New York-London*

Massim, I.,Mechab,B. (2016).Local linear estimation of the conditional hazard function. *Journal of Statistics & Economics.***17,** 1-11.

Rousseeuw, P. J., and Leroy, A. M. (1987). Robust Regression and Outlier Detection. *New York : John Wiley & Sons.*

CHAPITRE 4

ON THE LOCAL LINEAR MODELIZATION OF THE NONPARAMETRIC ROBUST REGRESSION FOR FUNCTIONAL TIME SERIES DATA

Ce travail traite la convergence presque-complète et la normalité asymptotique de la régression locale linéaire robuste lorsque les données sont fonctionnelles dépendantes. Ce travail a été soumis pour une eventuelle publication.

On the local linear modelization of the nonparametric robust regression for functional time series data

Souheyla Chemikh[1], **Faiza Belarbi**[2] **and Ali Laksaci**[3]

[1,2]Laboratory of Statistics and Stochastic Processes,Djillali Liabès University Bp,89, Sidi Bel Abbès 22000, Algeria.
chemikh_souheyla@yahoo.com
faiza_belarbi@yahoo.fr
[3]Department of Mathematics, College of Science, King Khalid University, P.O. Box : 9004,Postal Code : 61413, Abha, Saudi Arabia.
alilak@yahoo.fr

Abstract

In this work, we construct a robust nonparametric local linear estimator for a regression function when the regressors are functional random variables by combining both M-estimation technique and an local linear ideas, we establish the almost complete convergence rate as well as the asymptotic normality of this estimator under an α-mixing assumption and on some topological characteristics of the data. Moreover, a simulation study is given in order to evaluate, on a finite sample, the performance of our estimator.

Keywords : Functional data analysis ; Local linear method ; Robust estimation ; Asymptotic normality ; Almost complete convergence ; Strong mixing .

2000 Mathematics Subject Classification : 62F12, 62G20, 62G05.

4.1 Introduction

This paper deals with the nonparametric local linear estimation of the regression function when the observations are strongly mixing and the explanatory variable is functional (i.e. any observation lying into an infinite-dimensional space) by using a robust approach. This approach has an important role in statistics, because, it has more advantage over the classical method. In particular, it is not overly affected by the presence of outliers. Historically, the first consequent result in this area dates back to Huber (1964) and it has been widely studied for real data (see for instance Robinson(1984), Collomb and Härdle(1986). and Boente et al.(2009)). Concerning the robust local linear estimation, it has been studied by several authors in multivariate cases, we refer to Cai and Ould-Saïd (2003), for the α-mixing case. In this work, we introduce the local linear estimator of the robust regression.

Recall first that, the local linear method has various advantage over the classical kernel method. Precisely, it has superior bias properties to the previous one (see Fan and Gijbels (1996) for more discussions on this approach).
The local linear method has been considered for functional data,The literature on this subject dates back to Baìllo and Grané (2009), they proposed a local linear estimator of the regression function and they established the L_2-convergence rate of proposed estimator when the explanatory variable takes values in a Hilbert space. Barrientos-Marin et al (2010) studied the almost complete convergence (give the rate) of the locally modelled estimator of the regression function in the functional case. Another local linear estimate of the regression function was proposed by Berlinet et al.(2011). Demongeot et al. (2013) introduced the local linear estimator of the conditional density function. They treated the pointwise and uniform almost complete consistencies with convergence rates of this estimator. Recently, Chouaf (2015) obtained the asymptotic mean square error of the nonparametric regression function. Massim et al.(2016) considered the local linear estimation of the conditional hazard function when the explanatory variable is functional, showing the almost complete convergence of the construct estimate. Kaid and Laksaci (2017) proposed the local linear estimator of the conditional quantiles when the explanatory variable is functional, they showed the almost complete convergence and the asymptotic normality of the proposed estimator. More recently, Bouanani et al.(2018) were interested in the estimation of some functional conditional parameters by using the linear local method. These latter studied the asymptotic normality of the nonparametric local linear estimators such as la function de distribution conditional and the successive derivatives of the conditional density and they applied their results on simulated data then on real data. We return to Attach et al.(2017) for the local linear estimate of the regression function by the KNN method and Chahad et al. (2017) for the local linear estimation of the relative regression function. we cite also Xiong et al.(2017) for the local linear estimate of the conditional density under an α- mixing condition.
In an earlier paper (see Belarbi et al.(2018)), we studied the almost complete convergence (with rate)and the asymptotic normality of a robust local linear estimator of regression function for independent and identically distributed observations. Our interest in this work is to generalize these results to dependent case by using a score function bounded and we carry out a simulation study in order to show the advantages of this estimate .
This work is organized as follows, section 4.2 is dedicated to presenting our model. We fix the notations, hypotheses and state our main results in section 4.3. A simulation is given in the section 4.4. We give the proofs of the results in the appendix. Finally, we give our conclusion, in the last section.

4.2 The model and the estimator

Let $(X_1, Y_1),\cdots,(X_n, Y_n)$ be n pairs of random variables, identically distributed as (X, Y) which is a random pair valued in $\mathcal{F} \times \mathbb{R}$, where $\mathcal{F}$ is a semi-metric space, of eventually infinite dimension and d denotes the semi-metric. For any x in $\mathcal{F}$, The nonparametric robust regression, denoted by θ_x, it defined as the unique minimizer of the following optimization problem

$$\theta_x = \arg\min_{t\in\mathbb{R}} \mathbb{E}\left[\rho(Y-t)|X=x\right] \tag{4.1}$$

where $\rho(.)$ is a real-valued Borel function satisfying some regularity condition to be state below. This model belongs to the class of M-estimator introduced by Huber(1964), it includes many important nonparametric models, for example, The classical regression is obtained by setting $\rho(Y-t) = (Y-t)^2$, $\rho(Y-t) = |Y-t| + (2\alpha - 1)(Y-t)$ permits to obtain the α^{th} conditional quantile. For more others examples of this function we refer Stone (2005).
Our main purpose of this paper is to study the functional local linear estimate of the regression function. Noting that in functional statistics, there are several ways for extending this approach(see Baìllo Garné and (2009) and Barrientos-Mairin et al. (2010)). In this work, we adopt the fast version proposed by Barrientos-Mairin et al. (2010) for which the function θ_x is approximated by

$$\forall z \qquad \text{in neighborhood of } x \qquad \theta_z = a + b\beta(x,z) \tag{4.2}$$

where a and b are estimated by $\widehat{a}$ and $\widehat{b}$ are solution of

$$\min_{(a,b)\in\mathbb{R}^2} \sum_{i=1}^{n} \rho(Y_i - a - b\beta(X_i,x))K(h^{-1}\delta(x,X_i)) \tag{4.3}$$

where K is a kernel function and $h := h_n$ (to simplify the notation) is a sequence of positive real numbers with goes to zero as n goes to infinity and $\beta(\cdot,\cdot)$ is a know function from $\mathcal{F}^2$ into $\mathbb{R}$ such that, $\forall\xi \in \mathcal{F}, \beta(\xi,\xi) = 0$ and $\delta(\cdot,\cdot)$ is function of $\mathcal{F}\times\mathcal{F}$ such that $d(\cdot,\cdot) = |\delta(\cdot,\cdot)|$. It is clear that, under this consideration, we can write $\theta_x = a$ and $\widehat{\theta}_x = \widehat{a}$.

4.3 Hypotheses and results

We begin by recalling the definition of the strong mixing property.

Definition 4.1. *Let $\{\Delta_i, i = 1,2,\cdots\}$ be a strictly stationary sequence of random variables for any pair $(i,j) \in \mathbb{N}^*$ we note by $\Im_i^j$ the σ-algebra generated by $\{\Delta_k, i < k < j\}$ .The strong mixing coefficient is defined to be the following quantity, for any*

$n \geq 1$

$$\alpha(n) = \sup\left\{|\mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)| : k \in \mathbb{N}^*, A \in \Im_1^k(\Delta) \ \textit{and} \ B \in \Im_{k+n}^{\infty}(\Delta)\right\}.$$

The sequence is said to be α-mixing (or strongly mixing), if $\lim_{n\to\infty} \alpha(n) = 0$.

Throughout the paper, we suppose that$(X_n, Y_n)_{n\geq 1}$is strongly mixing. When no confusion is possible, we denote by C and C' some strictly positive generic constants. Moreover, x denotes a fixed point in $\mathcal{F}$, $\mathcal{N}_x$ denotes a fixed neighborhood of x. Furthermore, we put $\phi_x(r_1, r_2) = \mathbb{P}(r_1 \leq \delta(x, X) \leq r_2)$ where $B(x, r) = \{x' \in \mathcal{F}/|\delta(x', x)| \leq r\}$. For $i = 1, \cdots, n$, we denote by $\beta_i = \beta(X_i, x)$, and $K_i = K(h^{-1}\delta(x, X_i))$, and we introduce the following assumptions :

(H1) For any $r > 0$ $\phi_x(r) := \phi_x(-r, r) > 0$ and there exists a function $\chi_x(.)$ such that :

$$\forall t \in (-1, 1) \lim_{h\to 0} \frac{\phi_x(th, h)}{\phi_x(h)} = \chi_x(t).$$

(H2) ρ is a strictly convex function, continuously differentiable and has a Lipschitzian derivative ψ such that

$$|\psi(Y - t)| < C < \infty, \qquad \forall t \in \mathbb{R}.$$

(H3) The function $\Gamma_\lambda(x, \cdot) := \mathbb{E}[\psi^\lambda(Y - \cdot)|X = x]$ is of class $\mathcal{C}^1$ on $[\theta_x - \delta, \theta_x + \delta]$: $\delta > 0$, and $\lambda \in \{1, 2\}$. we put $\gamma(x, \cdot) := \dfrac{d}{dt}\Gamma_1(x, \cdot)$, such that :

$$\begin{cases} \forall (t_1, t_2) \in [\theta_x - \delta, \theta_x + \delta]^2, \ \forall (x_1, x_2) \in \mathcal{N}_x \times \mathcal{N}_x, \quad \text{and for } k_1, k_2 > 0 \\ (i) \qquad |\Gamma_\lambda(x_1, t_1) - \Gamma_\lambda(x_2, t_2)| \leq C d^{k_1}(x_1, x_2) + |t_1 - t_2|^{k_2}. \\ (ii) \qquad |\gamma(x_1, t_1) - \gamma(x_2, t_2)| \leq C' d^{k_1}(x_1, x_2) + |t_1 - t_2|^{k_2}. \end{cases}$$

(H3') The variable $\delta(x, X)$ is $\sigma(\beta(x, X))-$ measurable and the two partial derivatives of the function, $\Upsilon_x(s, t) = \mathbb{E}[\Gamma_1(X, t)|\beta(x, X) = s]$, at $(0, \theta_x)$ exist.

(H4) The function $\beta(\cdot, \cdot)$ is such that :

$$\forall z \in \mathcal{F}, \ C|\delta(x, z)| \leq |\beta(x, z)| \leq C'|\delta(x, z)| \quad \text{and} \quad \sup_{u\in B(x,r)} |\beta(u, x) - \delta(x, u)| = o(r).$$

(H5) $(X_1, Y_1)_{i\geq 1}$ is an α-mixing sequence whose coefficient satisfy

$$\alpha(n) = O(n^{-a^*}) \qquad \text{for some } a^* > 0.$$

(H6) $0 < \sup_{i\neq j} \mathbb{P}((X_i, X_j) \in (B(x, h) \times B(x, h))) = O\left(\dfrac{(\phi_x(h))^{(a^*+1)/a^*}}{n^{1/a^*}}\right).$

(H7) The kernel K is a positive, differentiable function within (-1,1) such that

$$\begin{pmatrix} K(1)-\int_{-1}^{1}K'(t)\chi_x(t)dt & K(1)-\int_{-1}^{1}(tK(t))'\chi_x(t)dt \\ K(1)-\int_{-1}^{1}(tK(t))'\chi_x(t)dt & K(1)-\int_{-1}^{1}(t^2K(t))'\chi_x(t)dt \end{pmatrix} \text{ is a positive definite matrix.}$$

(H8) There exists $\eta > 0$, such that, $Cn^{\frac{3-a^*}{a^*+1}+\eta} \leq \phi_x(h) \leq C'n^{\frac{1}{1-a^*}}$, with, $a^* > \dfrac{5+\sqrt{17}}{2}$.

Condition (H1)characterizes the classical concentration property . It is used in most of recent works in nonparametric statistics for functional data (see Ferraty and Vieu (2006)). Assumptions (H3) and (H3') are a regularity conditions which characterize the functional space of our model and are needed to evaluate the asymptotic bias.
Recall that,the robustness property of M-estimators is controlled by the convexity of loss function ρ and the boundedness of score function ψ. On the other hand, the boundedness condition of ψ can be dropped while using the truncation methods as in Laïb and Ould-Saïd (2000) which permits to obtain a general presentations of our model.
In order to establish the same results as for the i.i.d. case (see Belarbi et al.(2017)), we reinforce the mixing condition by (H5)and (H6). Note that we can establish the convergence results without these hypotheses. However, in that case,the convergence rate expression would be perturbed by the presence of the observations covariance.
Assumptions (H4) and (H7) used here are just the same as conditions used in Barrientos-Marin et al. (2010) and Demongeot et al. (2013). Assumption (H8) is technique condition imposed for the brevity of proofs.
Our main result is given in the following theorem.

Theorem 4.1. *Under hypotheses $(H1)-(H3)$, $(H4)-(H8)$ and if $\gamma(x,\theta_x) > 0$ then*

$$|\widehat{\theta_x}-\theta_x| = O(h^{\min(k_1,k_2)}) + O\left(\sqrt{\frac{\log(n)}{n\phi_x(h)}}\right) \qquad a.co.$$

Theorem 4.2. *Under hypotheses $(H1)-(H8)$, and if $\gamma(x,\theta_x),> 0$, for any $x \in \mathcal{A}$, we have :*

$$\left(\frac{n\phi_x(h)}{\sigma^2(x)}\right)^{1/2}(\widehat{\theta_x}-\theta_x-o(h)) \xrightarrow{\mathcal{D}} \mathcal{N}(0,1) \qquad as \quad n\to\infty$$

where $\xrightarrow{\mathcal{D}}$ denote convergence in distribution,

$$\sigma^2(x) = \frac{\Gamma_2(x,\theta_x)}{(\gamma(x,\theta_x))^2}\frac{a_3^2D_1-2a_2a_3D_2+a_2^2D_3}{(a_1a_3-a_2^2)^2} \quad and\ \mathcal{A}=\{x\in\mathcal{F},\ \sigma^2(x)\neq 0\}$$

with

$$a_j(x) = K(1) - \int_{-1}^{1} (s^{j-1}K(s))'\chi_x(s)ds \quad and \quad D_j(x) = K(1) - \int_{-1}^{1} (s^{j-1}K(s)^2)'\chi_x(s)ds \quad for\ j = 1, 2, 3.$$

4.4 Simulation study

Recall first that, our model has several advantages, it is considered as a generalization of several models of the functional regression. More precisely, we can easily obtain the definition the classical functional robust regression (C.F.R.R.) proposed by Attouch et al.(2010)) if we take $b = 0$ in 4.3,its estimator is given by

$$\min_{a \in \mathbb{R}} \sum_{i=1}^{n} \rho(Y_i - a) K(h^{-1}d(x, X_i)).$$

We obtain the local linear functional regression (L.L.F.R.) constructed by Barrientos-Marin et al.(2010) if we we take $\rho(y) = y^2$ in the model 4.3 , its estimator is defined by

$$\min_{(a,b) \in \mathbb{R}^2} \sum_{i=1}^{n} (Y_i - a - b\beta(X_i, x))^2 K(h^{-1}\delta(x, X_i)).$$

And if we replace b by 0 and we take $\rho(y) = y^2$ in the same model, we obtain the classical functional regression (C.F.R) which was introduced in NDFA by Ferraty and Vieu (2006), its estimator is defined by

$$\min_{a \in \mathbb{R}} \sum_{i=1}^{n} (Y_i - a)^2 K(h^{-1}d(x, X_i)).$$

The objective of this section is to see the behavior of our proposed estimator (the local linear functional robust regression (L.L.F.R.R)) for a finite sample .In particular, our first purpose is to show how can implement easily our estimator in the practice and the second purpose is to compare the sensitivity to outliers of the four estimators of ((C.F.R), (C.F.R.R.),(L.L.F.R.),(L.L.F.R.R)).For this we consider the explanatory curves generated in the following way :

$$X_i(t) = \cos(4(W_i - t)\pi) + a_i t^2 \quad for\ i = 1, 2, \ldots, n \ and\ \forall t \in [0, 1],$$

withe W_i is an $\alpha-$mixing process generated by $W_i = \dfrac{2}{9}W_{i-1} + \eta_i$ where η_i are i.i.d and follow the standard normal distribution $\mathcal{N}(0, 1)$ independently from W_i (W_0 is generated independently from a standard gaussian distribution $\mathcal{N}(0, 1)$) and the n random variables a_i's are distributed as $\mathcal{N}(4, 3)$. All the curves X_i's are discretized on the same grid which is composed of 100 equidistant values in [0,1]. and are represented in Figure 4.1

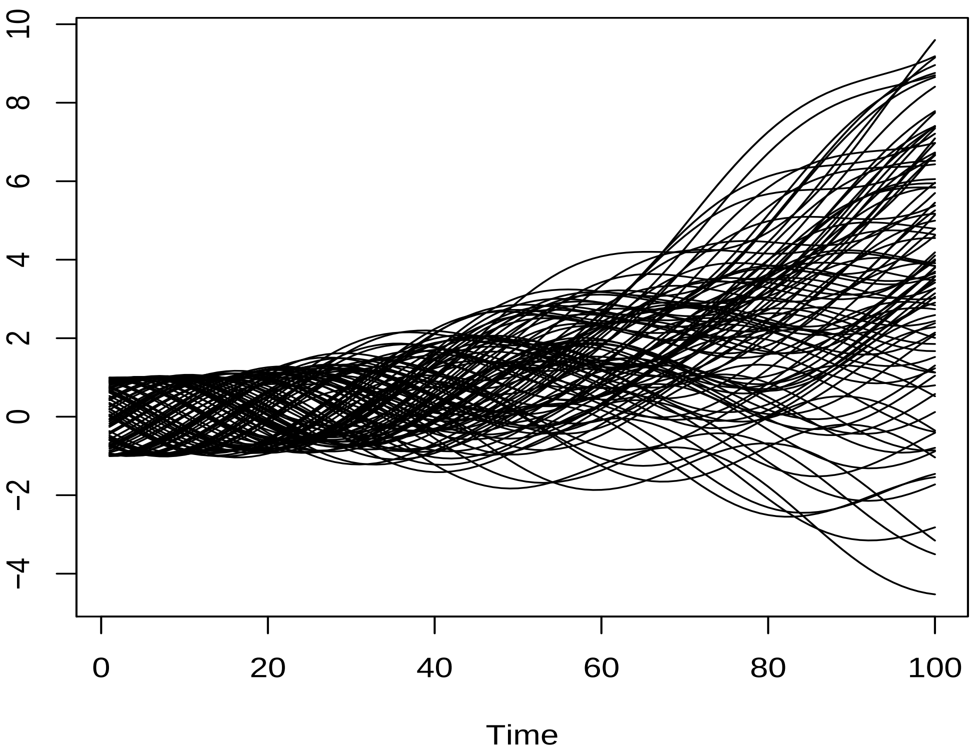

FIGURE 4.1 – The curves $X_{i=1,\ldots,100}(t), \quad t_{j=1,\ldots,100} \in [0,1]$

Then, we define the scalars responses variables Y_i's by :

$$Y_i = R(X_i) + \epsilon_i$$

where the errors ϵ_i's are i.i.d normally distributed as $\mathcal{N}(0, 0.2)$ and

$$R(X_i) = \exp\left\{ -\int_0^1 \frac{dt}{1 + X_i(t)^2} \right\}$$

In order to study this comparison, firstly, we take $\rho(y) = 2\sqrt{1 + y^2/2}$, for the both robust regressions models ((C.F.R),(L.L.F.R.R)). This loss function known in the literature by $L1 - L2$'s function. Noting that, this choice is motivated by the fact that the function $\rho(.)$ respects the fundamentals constraints of the properties of the M-estimations, namely, the strict convexity and the the boundedness of the score function $\psi(.)$. The first one is important for the unicity of the estimator, while the importance of the second propriety is the reduction the influence of large errors. Furthermore, the assumption (H2) is verified in this case.
Secondly, we select the optimal bandwidth h, for the four regressions models,by the

cross-validation method on the k nearest neighbors in a local way and we use the quadratic kernel which is defined by

$$K(x) = \frac{3}{4}(1-x^2)1\!\!1_{[-1,1]}$$

On the other hand, the choice of the locating functions has an important role in the local linear estimate. It's clear that, we can take $\beta(.,.) = \delta(.,.)$ in theory but the operators β and δ do not play the same role. More precisely, β refers to the local performance of the regression and δ concerns the local weighting (see section 2 in the paper of Barrientas-Marin et al.(2010)). Considering $\beta(.,.) \neq \delta(.,.)$ can lead us on the one hand to a more adaptative method and the other hand to a better fit to the data from a practical point of view. For some concrete examples and more discussion on this subject, we refer to (Barrientas-Marin et al.(2010)(in the section 3),(Rachdi et al.(2014)(in Section 5.1) and Xiongy et al. (2017) (in Section 5)). Furthermore, the choices of two locating functions δ and β depend on the shape of data. Here the curves X_i's are sufficiently smooth, then we take the following types (see Barrientas-Marin et al.(2010) for more discussion of this types)

$$\delta(x_1,x_2) = d(x_1,x_2) = \sqrt{\int_0^1 (x_1^{(i)}(t) - x_2^{(i)}(t))^2 dt} \quad and \quad \beta(x_1,x_2) = \int_0^1 \theta(t)(x_1(t) - x_2(t))dt$$

where θ is the eigenfunction of the empirical covariance operator

$$\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X_n})^t(X_i - \overline{X_n})$$

corresponding to the q-greatest eigenvalues,where $\overline{X_n} = \frac{1}{n}\sum_{i=1}^{n} X_i$ and $x^{(i)}(\cdot)$ denotes the i^{th} of derivative of the curve $x(\cdot)$. Noticing that, the locating functions depend on the order of derivatives i and the eigenvalues q but their choices are not very difficult in practice. In our simulative study, we present the results in the case where $q = 1$ and $i = 2$.

As we said previously, our objective is to test the sensitivity to outliers the four regressions models. For this aim, , we compare them over two samples (one is of the size n=100 and the other is of size n=200) and we split each sample in two samples : learning sample $((X_i, Y_i)_{j\in J})$ and test sample $((X_i, Y_i)_{i\in I})$. Then, we calculate $\widehat{m(X_i)}$ for any $i \in I$, where $\widehat{m(X_i)}$ designate the estimator used : (C.F.R), (C.F.R.R.),(L.L.F.R.) or (L.L.F.R.R). Finally, we use the mean square error (MSE) for evaluate this comparison.

In the first illustration which is given in Table 4.1, we compare between the four regressions models in the case of the absence of outliers where it appears clearly that the models ((C.F.R), (C.F.R.R.),(L.L.F.R.),(L.L.F.R.R)) have a good behavior but the MSE for ((L.L.F.R.),(L.L.F.R.R)) is smaller than for (C.F.R.) and (C.F.R.R.).

Table 4.1 – *Comparison between the four regressions models in the absence of outliers*

Size of sample	C.F.R.	C.F.R.R	L.L.F.R.	L.L.F.R.R
n=100	0.960	0.954	0.860	0.758
n=200	0.664	0.668	0.591	0.442

The second illustration is given in Table 4.2 where we introduced the outliers by multiplying some values of the response variable Y, in the learning sample by 50. In this case, we observe that (C.F.R.R.) and (L.L.F.R.R) models give better results than (C.F.R) and (L.L.F.R.). More precisely, the MSE of the (C.F.R)and (L.L.F.R.) increases substantially relatively to the number of the outliers, but it remains is very low for (C.F.R.R) and (L.L.F.R.). we observe also that the error of (L.L.F.R.) and (L.L.F.R.R) is smaller than that of (C.F.R) and (C.F.R.R.) respectively. Another remarkable point, the error of (L.L.F.R.R) remain very low with respect to the other regressions models studied, in the all cases.

Table 4.2 – *Comparison between the four regressions models in the presence of outliers*

Size of sample	number of artificial outliers	C.F.R.	C.F.R.R	L.L.F.R.	L.L.F.R.R
n=100	One outlier	1.981	1.051	1.516	0.971
	10 outliers	80.189	8.280	75.830	7.811
	25 outliers	359.33	32.468	321.970	28.409
n=200	One outlier	1.280	0.981	1.115	0.795
	10 outliers	68.871	6.921	59.124	5.142
	25 outliers	311.652	26.422	270.354	21.789

4.5 Appendix

Firstly, we state the following lemma which are needed to establish our asymptotic results (see Belarbi et al.(2017)).

Lemma 4.1. *Let $(V_n)_n$ a sequence of vectorial functions such that*

i) For all $\lambda \geq 1$ and a vector δ, we have, ${}^t(\delta)V_n(\lambda\delta) \leq {}^t(\delta)V_n(\delta)$.

ii) For a positive definite matrix D and vectorial sequences A_n such that $\mathbb{P}(\|A_n\| \geq A) \to 0$, for some $A > 0$ we have

$$\sup_{\|\delta\|\leq M} \|V_n(\delta) + \lambda_0 D\delta - A_n\| = o_p(1) \quad \textit{for} \quad \frac{A}{\lambda_0\lambda_1(D)} < M < \infty.$$

with $\lambda_1(D)$ is the minimum eigenvalue of D and $\lambda_0 > 0$. Then, for any vectorial sequence δ_n such that $V_n(\delta_n) = o_p(1)$ we have

$$\|\delta_n\| \leq M, \quad \textit{in probability.}$$

Furthermore, if $\|A_n\| = o_{a.co.}(1)$ and $\sup_{\|\delta\|\leq M} \|V_n(\delta) + \lambda_0 D\delta - A_n\| = o_{a.co.}(1)$. Then, for any vectorial sequence δ_n such that $V_n(\delta_n) = o_{a.co.}(1)$ we have

$$\|\delta_n\| \leq M, \quad \textit{almost completely.}$$

Proof of Theorems

We define, for all $\delta = \begin{pmatrix} c \\ d \end{pmatrix}$, $\varphi_i(\delta) = \psi(Y_i - ((a+c) + (h^{-1}d + b)\beta_i))$ and we consider the following vectorial sequence

$$V_n(\delta) = \frac{1}{n\phi_x(h)} \sum_{i=1}^{n} \varphi_i(\delta) \begin{pmatrix} 1 \\ h^{-1}\beta_i \end{pmatrix} K_i \quad \text{and} \quad A_n = V_n(\delta_0) \quad \text{with} \quad \delta_0 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

Now, the proof of the result of Theorem 4.1 is based on the application of the second part of Lemma 4.1 to (V_n, A_n, δ_n) with $\delta_n = \begin{pmatrix} \widehat{a} - a \\ h^{-1}(\widehat{b} - b) \end{pmatrix}$.While the second result is obtained by applying the first part of Lemma 4.1 to (V_n', A_n', δ_n') where

$$\varphi_i'(\delta) = \psi\left(Y_i - \left(a + \frac{1}{\sqrt{n\phi_x(h)}}c\right) + \left(\frac{1}{h\sqrt{n\phi_x(h)}}d + b\right)\beta_i\right)$$

$$V_n'(\delta) = \frac{1}{\sqrt{n\phi_x(h)}} \sum_{i=1}^{n} \varphi_i'(\delta) \begin{pmatrix} 1 \\ h^{-1}\beta_i \end{pmatrix} K_i \quad \text{and} \quad A_n' = V_n'(\delta_0) \quad \text{and} \quad \delta_n' = \sqrt{n\phi_x(h)}\delta_n'.$$

Clearly, the condition i) of Lemma 4.1 is holds, because ψ is monotone increasing function, decreasing case being obtained by considering $-\psi$.
So, the result of Theorem 4.1 is consequence of the following lemmas :

Lemma 4.2. *Under Hypotheses (H1),(H2), (H3i) and (H4)-(H8), we have*

$$\|A_n\| = O(h^{\min(k_1,k_2)}) + O_{a.co}\left(\sqrt{\frac{\log(n)}{n\phi_x(h)}}\right).$$

Lemma 4.3. *Under Hypotheses (H1)-(H3) and (H4)-(H8) we have*

$$\sup_{||\delta||\leq M} ||V_n(\delta) + \lambda_0 D\delta - A_n|| = o_{a.co.}(1).$$

$$\textit{with} \qquad D = \begin{pmatrix} K(1) - \int_{-1}^{1} K'(t)\chi_x(t)dt & K(1) - \int_{-1}^{1}(tK(t))'\chi_x(t)dt \\ K(1) - \int_{-1}^{1}(tK(t))'\chi_x(t)dt & K(1) - \int_{-1}^{1}(t^2K(t))'\chi_x(t)dt \end{pmatrix}$$

$$and\ \lambda_0 = \gamma(x, \theta_x).$$

Concerning the asymptotic normality result we use the following Lemmas :

Lemma 4.4. *Under Hypotheses (H1),(H3),(H4) and (H7)-(H8), we have*

$$\|A'_n\| = O_p(1).$$

Lemma 4.5. *Under Hypotheses (H1)-H3),(H4)- (H8) we have*

$$\sup_{||\delta'|| \leq M} ||V'_n(\delta) + \gamma(x_1, \theta_x) D\delta' - A'_n|| = o_p(1).$$

Proposition 4.1. *(see Belarbi et al. (2017).) Under Hypotheses of Lemma 4.4 we have :*

$$\widehat{\theta}_x - \theta_x = \frac{1}{\gamma(x, \theta_x)(a_1 a_3 - a_2^2)} \widehat{\Psi}(x) + o(1)$$

where $\widehat{\Psi}(x) = \dfrac{1}{n\phi_x(h)} \displaystyle\sum_{i=1}^{n} \varphi_i(\delta_0)(A_3 K_i - A_2 h_n^{-1} \beta_i K_i)$ *with* $A_j = \phi_x(h)^{-1} h^{1-j} \mathbb{E}[\beta^{j-1} K] = a_j + o(1).$

for $j = 2, 3..$
Then,we can write

$$\sqrt{n\phi_x(h)}(\widehat{\theta}_x - \theta_x) = \frac{1}{\gamma(x, \theta_x)(a_1 a_3 - a_2^2)} \left(\sqrt{n\phi_x(h)}(\widehat{\Psi}(x) - \mathbb{E}[\widehat{\Psi}(x)])\right) + \frac{\sqrt{n\phi_x(h)}}{\gamma(x, \theta_x)(a_1 a_3 - a_2^2)} \mathbb{E}[\widehat{\Psi}(x)]. \tag{4.4}$$

Then, to state asymptotic normality, we show that the first term of the right-hand side of (4.4) suitably normalized is asymptotically normally distributed. On the other hand, the second term is equal $\sqrt{n\phi_x(h)} o(h)$.

Lemma 4.6. *Under the hypotheses of Theorem 4.2, we have :*

$$\mathbb{E}[\widehat{\Psi}(x)] = o(h)$$
$$\sqrt{n\phi_x(h)} \sigma_1^{-1/2} (\widehat{\Psi}(x) - \mathbb{E}[\widehat{\Psi}(x)]) \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1) \qquad as \quad n \to \infty.$$
$$\sigma_1 = \Gamma_2(x, \theta_x)(a_3^2 D_1 - 2a_2 a_3 D_2 + a_2^2 D_3)$$

whenever

$$\exists \eta > 0,\ \ such\ that, \quad C n^{\frac{1}{1-a^*}} \leq \phi_x(h) \leq C' n^{\frac{1}{3-2a^*} - \eta}, \quad with\ a^* > \frac{5 + \sqrt{17}}{2}. \tag{4.5}$$

Proof of Lemmas

Proof of Lemma 4.2

Firstly, we set $$A_n - \mathbb{E}[A_n] = \begin{pmatrix} A_n^1 := \dfrac{1}{n\phi_x(h)} \displaystyle\sum_{i=1}^n Z_i^1 \\ A_n^2 := \dfrac{1}{n\phi_x(h)} \displaystyle\sum_{i=1}^n Z_i^2 \end{pmatrix}$$

where $$A_n^{l+1} := \frac{1}{n\phi_x(h)} \sum_{i=1}^n Z_i^{l+1} \qquad \text{for } l = 0, 1$$

and $$Z_i^{l+1} = \varphi_i(\delta_0)h^{-l}\beta_i^l K_i - \mathbb{E}[\varphi_i(\delta_0)h^{-l}\beta_i^l K_i].$$

Moreover, under (H1),(H2) and (H4)

$$\begin{aligned} |\varphi_i(\delta_0)h^{-l}\beta_i^l K_i| &\leq Ch^{-l}|\delta(X_i, x)|^l K_i \\ &\leq CK_i \leq C. \end{aligned}$$

So, we apply Fuk-Nagaev exponential inequality(Rio,2000,p.87) to get for all $r > 0$ and $\varepsilon_l > 0$.

$$\mathbb{P}\left[|A_n^{l+1}| > \varepsilon_l\right] \leq \mathbb{P}\left[\left|\sum_{i=1}^n Z_i^{l+1}\right| > \varepsilon_l n\phi_x(h)\right] \leq C(L1(x) + L2(x)) \tag{4.6}$$

where

$$L1(x) = \left(1 + \frac{\varepsilon_l^2 n^2 \phi_x(h)^2}{r S_{n,l+1}^2}\right)^{-r/2} \qquad L2(x) = nr^{-1}\left(\frac{r}{\varepsilon_l n\phi_x(h)}\right)^{(a^*+1)}$$

and

$$S_{n,l+1}^2 = \sum_{i=1}^n \sum_{j=1}^n Cov(Z_i^{l+1}, Z_j^{l+1}) = S_{n,l+1}^{2*} + nVar[Z_1^{l+1}] \tag{4.7}$$

with

$$S_{n,l+1}^{2*} = \sum_{i \neq j} Cov(Z_i^{l+1}, Z_j^{l+1}) = I_{l+1,n} + I'_{l+1,n}. \tag{4.8}$$

Next, we evaluate the asymptotic behavior of (4.8). Following Masry (1986), we define the sets

$$E_1 = \{(i,j) \quad such\ that \quad 1 \leq |i-j| \leq v_n\}$$

$$E_2 = \{(i,j) \quad such\ that \quad v_n + 1 \leq |i-j| \leq n-1\}$$

where $$v_n \to \infty, \qquad as\ n \quad \to \infty.$$

Now,we treat the quantities $I_{l+1,n}$

$$I_{l+1,n} \leq \sum_{E_1} |Cov(Z_i^{l+1}, Z_j^{l+1})| \leq \sum_{E_1} |\mathbb{E}[\varphi_i(\delta_0)\varphi_j(\delta_0)h^{-2l}\beta_i^l\beta_j^l K_i K_j] - \mathbb{E}[\varphi_i(\delta_0)h^{-l}\beta_i^l K_j] \times \mathbb{E}[\varphi_j(\delta_0)h^{-l}\beta_j^l K_j]|.$$

Under (H1),(H2),(H4), (H6)and (H7) permits to get

$$I_{l+1,n} \leq Cnv_n\phi_x(h)\left(\left(\frac{\phi_x(h)}{n}\right)^{(1/a^*)} + \phi_x(h)\right).$$

For $(I'_{l+1,n})$ we use Davydov-Rio inequality (Rio,2000,p.87) for mixing processes. This leards, for all $i \neq j$, we have

$$|Cov(Z_i^{l+1}, Z_j^{l+1})| \leq C\alpha(|i-j|).$$

Under (H5) we obtain

$$I'_{l+1,n} \leq \sum_{E_2} |Cov(Z_i^{l+1}, Z_j^{l+1})| \leq Cn^2 v_n^{-a^*}.$$

Choosing $v_n = \left(\frac{\phi_x(h)}{n}\right)^{-1/a^*}$ permits to get, under (H8)

$$S_{n,l+1}^{2*} = I_{l+1,n} + I'_{l+1,n} = O(n\phi_x(h)). \tag{4.9}$$

Concerning the variance term, we deduce from (H1) and (H2)

$$Var(Z_1^{l+1}) \leq C(\phi_x(h) + (\phi_x(h))^2). \tag{4.10}$$

Finally form (4.7), (4.8), (4.9) and (4.10) we obtain

$$S_{n,l+1}^2 = O(n\phi_x(h)). \tag{4.11}$$

Now, we apply (4.6) with

$$\varepsilon_l = \varepsilon'_l \frac{\sqrt{S_{n,l+1}^2 \log(n)}}{n\phi_x(h)} \qquad \text{and we chooser such that } r = C(\log n)^2. \tag{4.12}$$

It follows that, under (H8)

$$L2(x) \leq Cn^{-1-\eta(a^*+1)/2}(\log n)^{(3a^*-1)/2}.$$

So, it exists some real $\nu_l > 0$ such that

$$L2(x) \leq Cn^{-1-\nu_l}. \tag{4.13}$$

By means of (4.11) and (4.12) we get

$$L1(x) \leq C\exp\left(-C\varepsilon_l'^2\frac{\log n}{2}\right) = Cn^{-C\varepsilon_l'^2/2}.$$

Thus, for ε_l' large enough :

$$\exists \nu_l' > 0 \quad \text{such that} \quad L1(x) \leq Cn^{-C\varepsilon_l'^2/2} \leq Cn^{-1-\nu_l'}. \tag{4.14}$$

The results is easily deduced from (4.6), (4.13) and (4.14).Therefore

$$A_n - \mathbb{E}[A_n] = O_{a.co.}\left(\sqrt{\frac{\log n}{n\phi_x(h)}}\right). \tag{4.15}$$

On the other hand

$$\begin{aligned}
\mathbb{E}[A_n^{l+1}] &= \frac{1}{\phi_x(h)}\mathbb{E}[\varphi_1(\delta_0)h^{-l}\beta_1^l K_1] \\
&\leq \frac{1}{\phi_x(h)}\mathbb{E}[h^{-l}\beta_1^l K_1 1\!\!1_{B(x,h)}(X_1)|\Gamma_1(x,\theta_x) - \Gamma_1(X_1, a + b\beta_1)|] \\
&= O(h^{k_1}) + o(h^{k_2})
\end{aligned}$$

which achieves the proof of Lemma 4.2 .

Proof of lemma 4.3

We prove that

$$\sup_{||\delta||\leq M} ||\mathbb{E}[V_n(\delta) - A_n] + \gamma(x,\theta_x)D\delta|| = O(h^{\min(k_1,k_2)}) \tag{4.16}$$

$$\sup_{||\delta||\leq M} ||V_n(\delta) - A_n - \mathbb{E}[V_n(\delta) - A_n]|| = O_{a.co.}\left(\sqrt{\frac{\log(n)}{n\phi_x(h)}}\right). \tag{4.17}$$

For the first result, we write

$$V_n(\delta) - A_n = \begin{pmatrix} W_n^1(\delta) \\ W_n^2(\delta) \end{pmatrix}$$

where

$$W_n^{l+1}(\delta) := \frac{1}{nh^l\phi(h)}\sum_{i=1}^n (\varphi_i(\delta) - \varphi_i(\delta_0))\beta_i^l K_i.$$

Observe that,by (H3) we have

$$\begin{aligned}
\mathbb{E}[W_n^{l+1}(\delta)] &= \frac{1}{h^l\phi_x(h)}\mathbb{E}[\beta_1^l K_1 1\!\!1_{B(x,h)}(X_1)(\Gamma(X_1,(a+c)+(h_n^{-1}d+b)\beta_1)-\Gamma(X_1,a+b\beta_1))]\\
&= \frac{1}{h^l\phi_x(h}\mathbb{E}\left[\beta_1^l K_1 1\!\!1_{B(x,h)}(X_1)\gamma(X_1,a+b\beta_1)\left(1,h^{-1}\beta_1\right)\delta\right]+o(||\delta||)\\
&= \gamma(x,\theta_x)\frac{1}{h^l\phi_x(h)}\left(\mathbb{E}[\beta_1^l K_1],h^{-1}\mathbb{E}[\beta_1^{l+1}K_1]\right)\delta+O(h^{\min(k_1,k_2)})+o(||\delta||).
\end{aligned}$$

Therefore

$$\mathbb{E}[V_n(\delta)-A_n]=\gamma(x,\theta_x)\frac{1}{\phi_x(h)}\begin{pmatrix}\mathbb{E}[K_i] & \mathbb{E}[K_ih^{-1}\beta_i]\\ \mathbb{E}[K_ih^{-1}\beta_i] & \mathbb{E}[h^{-2}\beta_i^2K_i^2]\end{pmatrix}\delta+O(h^{\min(k_1,k_2)})+o(||\delta||).$$

Using the same ideas of Demongeot et al.(2013), under the seconde part of (H4) that

$$h_n^{-a}\mathbb{E}[\beta^a K_i^b]=\phi_x(h)\left(K^b(1)-\int_{-1}^{1}(u^aK(u))'\chi_x(u)du)\right)+o(\phi_x(h)).$$

It follow that

$$\sup_{||\delta||\le M}||\mathbb{E}[V_n(\delta)-A_n]+\gamma(x,\theta_x)D\delta+o(||\delta||)||=O(h^{\min(k_1,k_2)}).$$

Which complete the proof of (4.16). Concerning (4.17), we use the compactness of the ball $B(0,M)$ in $\mathbb{R}$ and we write

$$B(0,M)\subset\bigcup_{j=1}^{d_n}B(\delta_j,l_n)\quad \delta_j=\begin{pmatrix}c_j\\ d_j\end{pmatrix},\quad \text{and}\quad l_n=d_n=n^{-1/2}.$$

We take $j(\delta)=\mathrm{argmin}_j|\delta-\delta_j|$ we use fact that

$$\sup_{||\delta||\le M}||V_n(\delta)-A_n-\mathbb{E}[V_n(\delta)-A_n]||\le\underbrace{\sup_{||\delta||\le M}||V_n(\delta)-V_n(\delta_j)||}_{T_1}$$

$$+\underbrace{\sup_{||\delta||\le M}||V_n(\delta_j)-A_n-\mathbb{E}[V_n(\delta_j)-A_n]||}_{T_2}+\underbrace{\sup_{||\delta||\le M}||\mathbb{E}[V_n(\delta)-V_n(\delta_j)]||}_{T_3}.$$

Concerning T_1 and T_3 we use the fact that ψ is Lipshitzian function and K is bounded and under (H4) we get

$$\sup_{||\delta||\le M}||V_n(\delta)-V_n(\delta_j)||\le\frac{Cl_n}{\phi_x(h)}=o\left(\sqrt{\frac{\log(n)}{n\phi_x(h)}}\right).$$

It follows that

$$\sup_{||\delta||\leq M} ||V_n(\delta) - V_n(\delta_j)|| = O_{a.co.}\left(\sqrt{\frac{\log(n)}{n\phi_x(h)}} \right).$$

Concerning T_3 By using the same steps as in T_1 we have :

$$\sup_{||\delta||\leq M} ||\mathbb{E}[V_n(\delta) - V_n(\delta_j)]|| = o\left(\sqrt{\frac{\log(n)}{n\phi_x(h)}} \right).$$

Now, we treat the quantity T_2, we set

$$V_n(\delta_j) - A_n - \mathbb{E}[V_n(\delta_j) - A_n] = W_n(\delta_j) - \mathbb{E}[W_n(\delta_j)] = \begin{pmatrix} W_n^1(\delta_j) \\ W_n^2(\delta_j) \end{pmatrix}$$

where $$W_n^{l+1}(\delta_j) = \frac{1}{n\phi_x(h)} \sum_{i=1}^{n} \Lambda_i^{l+1}$$

and $$\Lambda_i^{l+1} = (\varphi_i(\delta_j) - \varphi_i(\delta_0))h^{-l}\beta_i^l K_i - \mathbb{E}[(\varphi_i(\delta_j) - \varphi_i(\delta_0))h^{-l}\beta_i^l K_i].$$

For all real $\mu > 0$ we have

$$\begin{aligned}
\mathbb{P}\left(T_2 > \mu\sqrt{\frac{\log n}{n\phi_x(h)}}\right) &= \mathbb{P}\left(\max_{j\in\{1,\cdots,d_n\}} \|W_n(\delta_j) - \mathbb{E}[W_n(\delta_j)]\| > \mu\sqrt{\frac{\log n}{n\phi_x(h)}}\right) \\
&\leq \sum_{j=1}^{d_n} \mathbb{P}\left(\|W_n(\delta_j) - \mathbb{E}[W_n(\delta_j)]\| > \mu\sqrt{\frac{\log n}{n\phi_x(h)}}\right) \\
&\leq d_n \max_{j\in\{1,\cdots,d_n\}} \mathbb{P}\left(\|W_n(\delta_j) - \mathbb{E}[W_n(\delta_j)]\| > \mu\sqrt{\frac{\log n}{n\phi_x(h)}}\right).
\end{aligned}$$

Then,for all real $\mu_l > 0$

$$\mathbb{P}\left(\mid W_n^{l+1}(\delta_j) \mid > \mu_l\sqrt{\frac{\log n}{n\phi_x(h)}}\right) \tag{4.18}$$

By using steps as in Lemma 4.2, where (H1), (H4),(H6),(H7), and (H8) are holds

$$\mathbb{P}\left(\mid W_n^{l+1}(\delta_j) \mid > \mu_l\sqrt{\frac{\log n}{n\phi_x(h)}}\right) \leq C\left(n^{-C\mu_l^2/2} + n^{(1-a^*)/2}(\log n)^{(3a^*-1)/2}\phi_x(h)^{-(a^*+1)/2}\right). \tag{4.19}$$

Therefore,

$$d_n\mathbb{P}\left(\mid W_n^{l+1}(\delta_j)\mid > \mu_l\sqrt{\frac{\log n}{n\phi_x(h)}}\right) \leq C\left(n^{1/2-C\mu_l^2/2} + n^{1/2+(1-a^*)/2}(\log n)^{(3a^*-1)/2}\phi_x(h)^{-(a^*+1)/2}\right). \tag{4.20}$$

Now, using the left hand side of (H8) and choosing μ_l a large enough , we obtain

$$\exists \upsilon_l > 0 \qquad d_n\mathbb{P}\left(\mid W_n^{l+1}(\delta_j)\mid > \mu_l\sqrt{\frac{\log n}{n\phi_x(h)}}\right) \leq Cn^{-1-\upsilon_l}. \tag{4.21}$$

Finally, by (4.21) we have :

$$\sup_{||\delta||\leq M}||V_n(\delta_j) - A_n - \mathbb{E}[V_n(\delta_j) - A_n]|| = O_{a.co.}\left(\sqrt{\frac{\log n}{n\phi_x(h)}}\right)$$

which complete the proof of (4.17).

Proof of lemma 4.4

We use the notations the proof of Lemma 4.2. It suffices to prove that

$$n\phi_x(h)Var[c_1A_n^1 + c_2A_n^2] = O(1) \quad \text{and} \quad \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} \in \mathbb{R}$$

where

$$Var[c_1A_n^1 + c_2A_n^2] = Var[c_1A_n^1] + Var[c_2A_n^2] + 2c_1c_2Cov(A_n^1, A_n^2).$$

So

$$\begin{aligned} Var[\sqrt{n\phi_x(h)}A_n^{l+1}] &= \frac{1}{n\phi_x(h)}\sum_{i=1}^{n}\sum_{j=1}^{n}Cov(Z_i^{l+1}, Z_j^{l+1}) \\ &= \frac{1}{\phi_x(h)}Var(Z_1^{l+1}) + \frac{1}{n\phi_x(h)}\sum_{i\neq j}Cov(Z_i^{l+1}, Z_j^{l+1}). \\ &= \frac{1}{\phi_x(h)}Var(Z_1^{l+1}) + \frac{1}{n\phi_x(h)}S_{n,l+1}^{2*}. \end{aligned}$$

It is shown in Belarbi et al.(2017) that

$$\frac{1}{\phi_x(h)}Var(Z_1^1) \to \Gamma_2(x,\theta_x)\left(K^2(1) - \int_{-1}^{1}(K^2(t))'\chi_x(t)dt\right).$$

$$\frac{1}{\phi_x(h)}Var(Z_1^2) \to \Gamma_2(x,\theta_x)\left(K^2(1) - \int_{-1}^{1}(t^2K^2(t))'\chi_x(t)dt\right).$$

$$\frac{1}{\phi_x(h)}Cov(Z_1^1 Z_1^2,) \to \Gamma_2(x,\theta_x)\left(K^2(1)-\int_{-1}^{1}(tK^2(t))'\chi_x(t)dt\right).$$

and by (4.9)
$$\frac{1}{n\phi_x(h)}S_{n,l+1}^{2*} = O(1).$$

Proof of lemma 4.5

The proof is based on the same assertions of Lemma 4.3. Firstly, by following the same line as (4.16) we have

$$\mathbb{E}[V_n'(\delta)-A_n'] = \gamma(x,\theta_x)\frac{1}{\phi_x(h)}\begin{pmatrix} \mathbb{E}[K_i] & \mathbb{E}[K_i h^{-1}\beta_i] \\ \mathbb{E}[K_i h^{-1}\beta_i] & \mathbb{E}[h^{-2}\beta_i^2 K_i^2] \end{pmatrix}\delta + O(h^{\min(k_1,k_2)}) + o(||\delta||).$$

Now, for the dispersion part we keep the notation of Lemma 4.3 and we write

$$\sup_{||\delta||\le M}||V_n'(\delta)-A_n'-\mathbb{E}[V_n'(\delta)-A_n']|| \le \sup_{||\delta||\le M}||V_n'(\delta)-V_n'(\delta_j)||$$

$$+\sup_{||\delta||\le M}||V_n'(\delta_j)-A_n'-\mathbb{E}[V_n'(\delta_j)-A_n']|| + \sup_{||\delta||\le M}||\mathbb{E}[V_n'(\delta)-V_n'(\delta_j)]||.$$

By simple calculation and under (H1),(H4),(H7) and (H8) and we use the fact that the function ψ is Lipschitzian we have

$$\sup_{||\delta||\le M}||V_n'(\delta)-V_n'(\delta_j)|| = o_p(1) \quad \text{and} \quad \sup_{||\delta||\le M}||\mathbb{E}[V_n'(\delta)-V_n'(\delta_j)]|| \le Cl_n = o(1). \tag{4.22}$$

Furthermore, for the quantity

$$V_n'(\delta_j)-A_n'-\mathbb{E}[V_n'(\delta_j)-A_n'] = \begin{pmatrix} \sqrt{n\phi_x(h)}G_n^1(\delta_j) \\ \sqrt{n\phi_x(h)}G_n^2(\delta_j) \end{pmatrix}$$

$$G_n(\delta_j)^{l+1} = \frac{1}{n\phi_x(h)}\sum_{i=1}^{n}\Delta_i^{l+1}$$

and
$$\Delta_i^{l+1} = (\varphi_i'(\delta_j)-\varphi_i(\delta_0))h^{-l}\beta_i^l K_i - \mathbb{E}[(\varphi_i'(\delta_j)-\varphi_i(\delta_0))h^{-l}\beta_i^l K_i].$$

$$\begin{aligned} Var[\sqrt{n\phi_x(h)}G_n^{l+1}(\delta_j)] &= \frac{1}{n\phi_x(h)}\sum_{i=1}^{n}\sum_{j=1}^{n}Cov(\Delta_i^{l+1},\Delta_j^{l+1}) \\ &= \frac{1}{\phi_x(h)}Var(\Delta_1^{l+1}) + \frac{1}{n\phi_x(h)}\sum_{i\neq j}Cov(\Delta_i^{l+1},\Delta_j^{l+1}) \\ &= \frac{1}{\phi_x(h)}Var(\Delta_1^{l+1}) + \frac{1}{n\phi_x(h)}L_n^{l+1}. \end{aligned}$$

By simple calculation we have

$$\frac{1}{\phi_x(h)} Var(\Delta_1^{l+1}) = O\left(\frac{1}{n\phi_x(h)}\right). \tag{4.23}$$

Next, we treat quantities L_n^{l+1},by using the fact that the function ψ is Lipschitzian we have we have

$$\begin{aligned} L_n^{l+1} \leq \sum_{i\neq j} |Cov(\Delta_i^{l+1}, \Delta_j^{l+1})| &\leq \sum_{i\neq j} |Cov((\varphi_i'(\delta_j) - \varphi_i(\delta_0))K_i, (\varphi_j'(\delta_j) - \varphi_j(\delta_0))K_j)| \\ &\leq C \frac{1}{n\phi_x(h)} \sum_{i\neq j} \mathbb{E}[K_iK_j] + \mathbb{E}[K_i]\mathbb{E}[K_j] \end{aligned}$$

Then,by using the same steps of calculus of $I_{n,l+1}$ and $I'_{n,l+1}$ in Lemma 4.2 we have

$$\sum_{i\neq j} \mathbb{E}[K_iK_j] + \mathbb{E}[K_i]\mathbb{E}[K_j] = O(n\phi_x(h)) \qquad \text{So} \qquad \frac{1}{n\phi_x(h)} L_n^{l+1} = O\left(\frac{1}{n\phi_x(h)}\right). \tag{4.24}$$

Thus, the proof of this Lemma is consequence of (H8), (4.22), (4.23) and (4.24).

Proof of lemma 4.6

For the first result :(see Belarbi et al. (2017)), and under (H3') we show that

$$\mathbb{E}[\widehat{\Psi}(x)] = \mathbb{E}[\Upsilon_x(\beta, \theta + a + \beta)(A3K - A_2h^{-1}K) = o(h).$$

The proof of the seconde result runs along the lines of lemma 1 in Attouch et et al (2009).

$$\sqrt{n\phi_x(h)}(\widehat{\Psi}(x) - \mathbb{E}[\widehat{\Psi}(x)]) = \sum_{i=1}^{n} \widetilde{\Delta}_i(x)$$

$$\text{where} \quad \widetilde{\Delta}_i(x) = \frac{1}{\sqrt{n\phi_x(h)}} \left[\varphi_i(\delta_0)(A_3K_i - A_2h^{-1}\beta_iK_i) - \mathbb{E}\left[\varphi_i(\delta_0)(A_3K_i - A_2h^{-1}\beta_iK_i)\right]\right].$$

The proof is based of the central limit Theorem of Liebscher (2001)(Corollary 2,2,p196) which rests on the asymptotic behavior of the following quantity .

$$\lim_{n\to\infty} \sum_{i=1}^{n} \mathbb{E}[\widetilde{\Delta}_i(x)^2] \tag{4.25}$$

and the additional assumptions

$$\begin{cases} \text{There exists a sequence } w_n = o(\sqrt{n}), w_n \to \infty \text{ and} w_n \leq \left(\max_{i=1,\cdots,n} \zeta_i\right)^{-1} \text{ where } \zeta_i = ess\sup_{\omega\in\Omega} |\Lambda_i(x)| \\ \text{and} \dfrac{n}{w_n}\alpha(\varepsilon w_n) \longrightarrow 0 \text{ fo all } \varepsilon > 0. \end{cases} \tag{4.26}$$

$$\begin{cases} \text{There exists a sequence } (v_n) \text{ of positive integers tending to } \infty \text{ such that } nv_n\rho_n = o(1) \text{ where} \\ \rho_n := \max_{1\leq i\neq j\leq n} (\mathbb{E}[|\widetilde{\Delta}_i(x)\widetilde{\Delta}_j(x)|]) \text{ and } \left(\sum_{j=v_n+1}^{\infty} \alpha(j)\right) \sum_{i=1}^{n} \zeta_i = o(1). \end{cases} \tag{4.27}$$

Firstly, we evaluate the limit of (4.25). To do that, let us remark that

$$\begin{aligned} \sum_{i=1}^{n} \mathbb{E}[\widetilde{\Delta}_i(x)^2] &= \frac{1}{n\phi_x(h)} \sum_{i=1}^{n} Var[A_3K_1 - A_2h^{-1}\beta_1K_1)\varphi(\delta_0)] \\ &= \frac{1}{\phi_x(h)} Var[\varphi(\delta_0)(A_3K_1 - A_2h^{-1}\beta_1K_1)] \\ &= \frac{1}{\phi_x(h)} \mathbb{E}[(A_3K_1 - A_2h^{-1}\beta_1K_1)^2]\mathbb{E}\left[\frac{(\varphi(\delta_0))^2(A_3K_1 - A_2h^{-1}\beta_1K_1)^2}{\mathbb{E}[(A_3K_1 - A_2h^{-1}\beta_1K_1)^2]}\right] \\ &- \frac{1}{\phi_x(h)} \mathbb{E}^2\left[\varphi(\delta_0)(A_3K_1 - A_2h^{-1}\beta_1K_1)\right]. \end{aligned}$$

It is shown in Belarbi et al.(2017)that

$$\frac{1}{\phi_x(h)} \mathbb{E}[(A_3K_1 - A_2h^{-1}\beta_1K_1)^2]\mathbb{E}\left[\frac{(\varphi(\delta_0))^2(A_3K_1 - A_2h^{-1}\beta_1K_1)^2}{\mathbb{E}[(A_3K_1 - A_2h^{-1}\beta_1K_1)^2]}\right] \to \Gamma_2(x,\theta_x) \times \left(a_3^2D_1 - 2a_2a_3D_2 + a_2^2D_3\right)$$

and
$$\frac{1}{\phi_x(h)} \mathbb{E}^2\left[\varphi(\delta_0)(A_3K_1 - A_2h^{-1}\beta_1K_1)\right] = o(h).$$

It follows that

$$\lim_{n\to\infty} \sum_{i=1}^{n} \mathbb{E}[\widetilde{\Delta}_i(x)^2] = \Gamma_2(x,\theta_x)\left(a_3^2D_1 - 2a_2a_3D_2 + a_2^2D_3\right). \tag{4.28}$$

Now, we treat (4.26), using the boundness $\varphi_.(.)$ and K and under (H4) we have :

$$\begin{aligned} |\widetilde{\Delta}_i(x)| &\leq \frac{1}{\sqrt{n\phi_x(h)}} \left|\left[\varphi_i(\delta_0)(A_3K_i - A_2h^{-1}\beta_iK_i) - \mathbb{E}\left[\varphi_i(\delta_0)(A_3K_i - A_2h^{-1}\beta_iK_i)\right]\right]\right| \\ &\leq \frac{C}{\sqrt{n\phi_x(h)}}. \end{aligned}$$

Then

$$\zeta_i = ess\sup_{\omega\in\Omega}|\widetilde{\Delta}_i(x)| = O\left(\frac{1}{\sqrt{n\phi_x(h)}}\right). \text{ Therefore, we can take} w_n = \sqrt{\frac{n\phi_x(h)}{\log n}}$$

.

Furthermore, this choice give, for all ε, and by (4.5) we get :

$$\begin{aligned}
\frac{n}{w_n}\alpha(\varepsilon w_n) &\leq C\left(n^{1-(a^*+1)/2}(\phi_x(h)^{-(a^*+1)/2}(\log n)^{(a^*+1)/2}\right)\\
&\leq C\left(n^{1-(a^*+1)/2+(a^*+1)/2(a^*-1)}(\log n)^{(a^*+1)/2}\right)\\
&\leq C\left(n^{(3a^*-a^{*2})/2(a^*-1)}(\log n)^{(a^*+1)/2}\right)\longrightarrow 0 \quad \text{since} \quad a^*>3.
\end{aligned}$$

Concerning (4.27),On the one hand as $\varphi_.(.)$ is bounded and under (H4), then

$$\begin{aligned}
\forall\ i\neq j \quad \mathbb{E}[|\widetilde{\Delta}_i(x)\widetilde{\Delta}_j(x))|] &\leq \frac{1}{n\phi_x(h)}(\mathbb{E}[|\varphi_i(\delta_0)(A_3K_i - A_2h_n^{-1}\beta_iK_i)\varphi_j(\delta_0)(A_3K_j - A_2h^{-1}\beta_jK_j)|]\\
&+ \mathbb{E}[|\varphi_i(\delta_0)(A_3K_i - A_2h^{-1}\beta_iK_i)|]\mathbb{E}[|\varphi_j(\delta_0)(A_3K_j - A_2h^{-1}\beta_jK_j)|])\\
&\leq C\frac{1}{n\phi_x(h)}(\mathbb{E}[K_iK_j] + \mathbb{E}[K_i]\mathbb{E}[K_j]).
\end{aligned}$$

Using the argument of Ferraty and Vieu (2006) $\mathbb{E}[K_i] = O(\phi_x(h))$ and under (H1)and (H6) we get :

$$\mathbb{E}[|\widetilde{\Delta}_i(x)\widetilde{\Delta}_j(x)|] \leq C\frac{1}{n}\left(\left(\frac{\phi_x(h)}{n}\right)^{1/a^*} + \phi_x(h)\right).$$

Because of (4.5) we have $\rho_n = \max_{1\leq i\neq j\leq n}(\mathbb{E}[|\widetilde{\Delta}_i(x)\widetilde{\Delta}_j(x)|]) = O\left(\frac{\phi_x(h)}{n}\right)$.

On the other hand , by (H5) we have :

$$\sum_{j\geq v_n+1}^{\infty}\alpha(j) \leq \sum_{j\geq v_n}^{\infty}\alpha(j) \leq \int_{u\geq v_n} u^{-a^*}du = \frac{v_n^{1-a^*}}{a^*-1}$$

thus

$$\left(\sum_{j=v_n+1}^{\infty}\alpha(j)\right)\sum_{i=1}^{n}\zeta_i = O\left(\frac{v_n^{1-a^*}}{a^*-1}\sqrt{\frac{n}{\phi_x(h)}}\right).$$

Choosing $v_n = \left[\left(\frac{\phi_x(h)}{n\log n}\right)^{1/(2(1-a^*)}\right]$ where [.] denote the function integer part. and by using (4.5) , $v_n\to\infty$.. Now, we replace v_n by its expression, we obtain :

$$\left(\sum_{j=v_n+1}^{\infty}\alpha(j)\right)\sum_{i=1}^{n}\zeta_i = O\left(\frac{1}{\sqrt{\log n}}\right) = o(1).$$

Using (4.5) we have :

$$\begin{aligned} v_n\rho_n &\leq Cn^{-1-1/(2(1-a^*))}(\phi_x(h))^{1+1/(2(1-a^*))}(\log n)^{-1/(2(1-a^*))} \\ &\leq Cn^{(-3+2a^*)/(2(1-a^*))}(\phi_x(h))^{(3-2a^*)/(2(1-a^*))}(\log n)^{-1/(2(1-a^*))} \\ &\leq n^{-1-\eta(3-2a^*)/(2(1-a^*))}(\log n)^{-1/(2(1-a^*))} = o(n^{-1}). \end{aligned}$$

The result can be easily from (4.25), (4.26) and (4.27) and corollary of Liebscher (2001).

4.6 Conclusion

In this work, we have positioned our contribution in the extensive literature of non-parametric analysis of functional data by giving the asymptotic results in the case α-mixing. The results obtained, in the simulation part, confirm that the robust local linear estimator has an insensitive and effective behavior in the presence of the outliers observations with a reduced bias term.

Références

Attouch, M., Laksaci, A., Ould Saïd, E. (2010). Asymptotic normality of robust estimator of the regression function for functional time series data. *Journal Korean statistical society* , **39,** 489-500.

Attouch, M., Laksaci, A., Ould Saïd, E. (2012). Robust regression for functional time series data. *Journal of Japan Statist.Soc,* **42,** 125-143.

Attouch, M., Laksaci, A., Rafaa,M.(2017).Estimation locale linéaire de la régression non paramétrique fonctionnelle par la méthode des k plus proches voisins. *Journal of Comptes Rendus Mathematique,***355,** 824-829.

Barrientos-Marin,J. and Ferray, F. and Vieu,P.(2010).Locally Modelled Regression And Functional Data.*Nonparametric Statistics,* **22,** 617-632.

Bouanani, O., Laksaci, A., Rachdi, M., Rahmani, S. (2018). Asymptotic normality of some conditional nonparametric functional parameters in high-dimensional statistics.*Journal of Behaviormetrika,* **45,** 1-35.

Belarbi,F., Chemikh,S.,Laksaci, A.(2017).Asymptotic properties of the robust local linear regression with functional regressor.*Tech. report,* https://www.researchgate.net/profile/Ali_Laksaci.

Belarbi,F.,Chemikh,S., Laksaci, A.(2018).Local linear estimate of nonparametric robust regression in function data.*Statistic and Probability letters,* **134,** 128–133.

Berlinet, A., Elamine, A., and Mas, A. (2011). Local linear regression for functional data. *Ann. Inst. Statist. Math,* **63**, 1047-1075.

Boente, G., Manteiga, W.G., Péerez-Gonzàlez, A. (2009). Robust nonparametric estimation with missing data. *J. Statist. Plan. and Inference,* **139**, 571-592.

Cai, Z. Ould Said., E. (2003). Local M-estimator for nonparametric time series. *Statist. Probab. Lett.* **65**, 433-449.

Chahad, A., Ait-Hennani, L., Laksaci, A.(2017). Functional local linear estimate for functional relative-error regression. *Journal of Statistical Theory and Practice,* **11,**771-789.

Chouaf, A. (2015).Modelization local linear regression for functional random variables. *International Journal of Statistics & Economics,* **16**, 54-68.

Collomb, G., Härdle, W. (1986). Strong uniform convergence rates in robust nonparametric times eries analysis and prediction : Kernel regression estimation from dependent observations. *Stoch. Proc. Appl,* **23,** 77-89.

Demongeot, J., Laksaci, A.,Madani, F. and Rachdi, M.(2013).Modelization local linear regression for functional random variables. *Statistics,* **47**, 26-44.

Fan, J., Hu, T. C. and Truong, Y. K. (1994). Robust non-parametric function estimation. *Scand. J. Statist.* **21**, 433-446.

Fan, J. and Gijbels, I. (1996). Local polynomial modeling and its applications. *London, Chapman & Hall.*

Ferraty, F. and Vieu, P. (2006). Nonparametric functional data analysis. Theory and Practice. *Springer Series in Statistics. New York.*

Huber, P. J. (1964). Robust estimation of a location parameter. *Ann. Math. Statist.* **35**, 73- 101.

Kaid,Z.,Laksaci,A.(2017).Functional quantile regression :local linear modelisation. *Functional Statistics and Related Fields,* 155-156.

Liebscher, E.(2001). Central limit theorems for α-mixing triangular arrays with applications to nonparametric statistics.*Mathematical Methods of Statistics,* **10,** 194 -214.

Laïb,N., Ould Saïd,E.,(2000). A robust nonparametric estimation of the autoregression function under an ergodic hypothesis. *Canad. J. Statist.,***28,** 817-228.

Massim, I.,Mechab,B. (2016).Local linear estimation of the conditional hazard function. *Journal of Statistics & Economics.***17,** 1-11.

Masry, E. (1986). Recursive probability density estimation for weakly dependent stationary processes. *IEEE. Trans. Inform. Theory,***32,** 254-267.

Rachdi, M., Laksaci, A., Demongeot, J., Abdali, A. et Madani, F.(2014). Theoretical and practical aspects of the quadratic error in the local linear estimation of the conditional density for functional data. *Comput. Statist. Data Anal,* **73,** 53-68.

Rio, E. (2000). Théorie asymptotique des processus aléatoires faiblements dépendants.(In french). *Mathématiques & Applications,* **31,** Springer Verlag, New-York.

Robinson, R. (1984).Robust Nonparametric Autoregression. *Lecture Notes in Statistics,* **26,** 247-255. Springer-Verlag, New-York.

Stone C J, (2005). Nonparametric M-regression with free knot splines. *J. Statist. Plan. and Inference,* **130**, 183-206.

Xiong, X., Zhou, Z., Ailian, C. (2017). Asymptotic normality of the local linear estimation of the conditional density for functional time series data. *Journal of Communications in Statistics - Theory and Methods,***47,** 3418-3440.

Zhou, Z. and Lin, Z. (2016). Asymptotic normality of locally modelled regression estimator for functional data. *J. of Nonparametric Statistics,* **28** , 116-131.

CHAPITRE 5

ESTIMATION LOCALE LINÉAIRE DE LA RÉGRESSION TRIMMED POUR DES DONNÉES FONCTIONNELLES : APPLICATION SUR DES DONNÉES SIMULÉES

La régression robuste en particulier la régression Trimmed est l'un des outils fondamentaux qui explique le lien entre deux variables aléatoires. Dans ce chapitre, nous allons déterminer cette relation en utilisant l'estimation non paramétrique locale linéaire. Plus précisément, nous allons présenter ce chapitre en deux sections, la première section est consacrée à montrer la performance de notre estimateur en le comparant avec d'autres types de régressions à l'aide de données simulées. Tandis que, dans la deuxième section, nous essayerons de répondre, autant que nous sachions, à quelques questions pratiques.

5.1 Illustration de la régression Trimmed

L'objectif principal de cette section est d'examiner le comportement de l'approche Trimmed en comparant entre la technique locale linéaire et la méthode à noyau classique (locale constante), sur des échantillons finis. Précisément, nous comparons, dans un premier temps, entre la prédiction via la régression locale constante Trimmed (R.L.C.T.) c'est-à-dire dans le cas où $b = 0$, et la prédiction par la régression classique (R.C.) qui a été introduite dans NDFA par Ferraty et Vieu (2006). Deuxièmement,

nous comparons entre la prédiction via la régression locale linéaire Trimmed (R.L.L.T.) et l'estimation par la régression locale linéaire classique (R.L.L.C.) qui a été construite par Barrientos-Marin et al. (2010). Finalement, afin de montrer l'intérêt de notre modèle de régression (R.L.L.T.), nous comparons la sensibilité aux valeurs aberrantes des quatre types de régression à savoir, (R.L.C.T.), (R.C.), (R.L.L.T.) et (R.L.L.C.).
Nous générons nos observations fonctionnelles $X_{i,\{i=1,\ldots,n\}}$ à l'aide du processus suivant :

$$X_i(t) = \pi \cos(\pi(tW_i - \eta_i)) + \sin(\eta_i^2 t^2), \quad \forall t \in [0,1] \text{ et } n = 150.$$

où les variables W_i (respectivement η_i) sont distribuées selon la loi normal $\mathcal{N}(0,1)$ (respectivement uniforme $\mathcal{U}(1,1)$). Toutes les courbes X_i sont discrétisées sur la même grille qui est composée de 100 valeurs équidistantes dans $[0,1]$ et elles sont représentées dans la Figure 5.1.

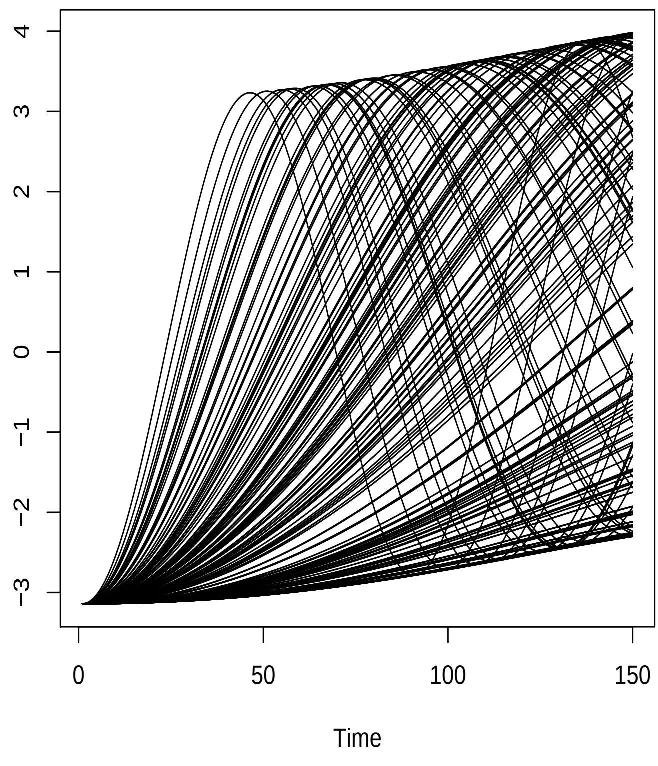

FIGURE 5.1 – Les courbes $X_{i=1,\ldots,150}(t), \quad t_{j=1,\ldots,150} \in [0,1]$

Ensuite, nous définissons les variables de réponses scalaires Y_i par

$$Y_i = R(X_i) + \epsilon_i$$

où les erreurs ϵ_i sont i.i.d et elles suivent la loi normale $\mathcal{N}(0, 0.2)$ et

$$R(X_i) = \int_0^1 (1 + X_i(t))dt$$

Rappelons maintenant que notre estimateur $\widehat{\theta_x}$ (voir les chapitres précédents) qui est défini par la solution du problème de minimisation suivant :

$$\min_{(a,b)\in\mathbb{R}^2} \sum_{i=1}^{n} \rho(Y_i - a - b\beta(X_i, x))K(h^{-1}\delta(x, X_i)), \tag{5.1}$$

où ρ est la fonction primitive de ψ, généralise plusieurs estimateurs de la fonction de régression. Nous pouvons facilement obtenir la régression classique (C.R.) qui a été introduite par Ferraty et Vieu(2006) si nous prenons $\psi(y) = y$ et nous remplaçons b par 0, tandis que nous pouvons atteindre sa version locale linéaire (R.L.L.C.) qui a été introduite par Barrientos-Marin et al. (2010) si nous prenons $\psi(y) = y$. Nous obtenons la définition de la régression locale constante Trimmed (R.L.C.T.) (respectivement la régression locale linéaire Trimmed (R.L.L.T.)) si nous travaillions par une fonction $\psi(y) = y\mathbb{1}_{|y| \leq q}$ où $\mathbb{1}$ est une fonction indicatrice et $b = 0$ (respectivement $\psi(y) = y\mathbb{1}_{|y| \leq q}$, et $b \neq 0$)(Voir Chapitre 3).

Notre but initial de cette illustration est de montrer l'utilité de l'approche Trimmed dans un contexte de prévision. A cette fin, nous choisissons $\alpha = 0.01$ et F est la fonction de répartition de ϵ_i. Ensuite, nous divisons notre échantillon en deux paquets d'observations : un échantillon d'apprentissage $(X_i, Y_i)_{i=1,\ldots,120}$ et un échantillon de test $(X_i, Y_i)_{i=121,\ldots,150}$. Pour ce dernier sous-échantillon (échantillon de test), nous supposons que les valeurs réponses sont inconnues et nous les estimons par $\widehat{\theta_{Xi}}$ qui désigne l'estimateur utilisé de : R.C.,R.L.L.C.,R.L.C.T. et R.L.L.T. Dans le but de vérifier l'efficacité de notre estimateur nous comparons les quantités $(Y_i)_{i=121,\ldots,150}$ aux quantités $(\widehat{\theta_{Xi}})_{i=121,\ldots,150}$. et nous sélectionnons le paramètre de lissage optimale h ,pour les quatre types de régression, par la méthode de L^1 validation croisée sur le nombre de voisin le plus proche de manière locale et nous utilisons le noyau quadratique défini par

$$K(x) = \frac{3}{4}(1 - x^2)\mathbb{1}_{[-1,1]}$$

pour les quatre types de régressions étudiés.

Maintenant, nous pésentons les résultats en tracant le graphe des valeurs prédites par (R.C.) et le graphe des valeurs éstimées via (R.L.C.T.) et nous les comparons au graphe des valeurs réelles $(Y_i)_{i=121,\ldots,150}$. Les résultats sont illustrés dans la Figure.5.2 où nous montrons qu'en l'absence de valeurs aberrantes, les deux types de régression sont fondamentalement équivalents et donnent le bon comportement de notre procédure de prévision fonctionnelle.

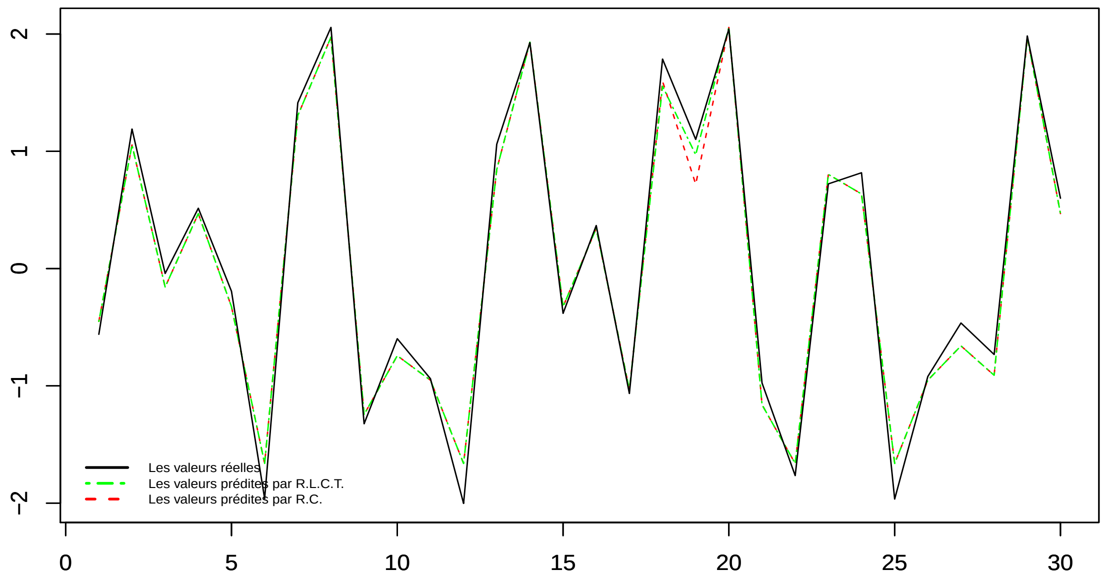

FIGURE 5.2 – Comparaison entre les méthodes d'estimation (R.C. et R.L.C.T.)

Dans la deuxième illustration, nous essayons de comparer entre les deux types de régression, locale linéaire R.L.L.C. et R.L.L.T. qui dépendent du choix de fonctions de localisation δ et β. Rappelons que le choix de ces fonctions dépend aussi de la forme des données. Ici les courbes X_i sont suffisamment lisses, nous prenons les types suivants (voir Barrientas-Marin et al. (2010) pour plus de discussion sur ce type)

$$\delta(x_1, x_2) = d(x_1, x_2) = \sqrt{\int_0^1 (x_1^{(i)}(t) - x_2^{(i)}(t))^2 dt}$$

et

$$\beta(x_1, x_2) =< \theta, x_1^{(j)} - x_2^{(j)} >= \int_0^1 \theta(t)(x_1^{(j)}(t) - x_2^{(j)}(t))dt$$

où θ est le vecteur propre fonctionnel (la fonction propre) de l'opérateur de covariance empirique

$$\frac{1}{n}\sum_{i=1}^{n}(X_i^{(j)} - \overline{X_n}^{(j)})^t(X_i^{(j)} - \overline{X_n}^{(j)})$$

correspondant aux q-plus grandes valeurs propres, où $\overline{X_n}^{(j)} = \frac{1}{n}\sum_{i=1}^{n} X_i^{(j)}$,et $x^{(i)}$ (respectivement $x^{(j)}$) désigne la $i^{\text{ème}}$ (respectivement la $j^{\text{ème}}$) dérivée de la courbe x. Dans cette simulation, nous avons travaillé par plusieurs valeurs de i, q et j, cela nous a permis de remarquer que nos résultats montrent une performance légèrement meilleure pour $q = 1$, $i = 1$ et $j = 2$.

La deuxième illustration est donnée dans la Figure 5.3 où nous comparons entre R.L.L.C. et R.L.L.T. en l'absence des valeurs aberrantes en utilisant l'erreur quadratique moyenne (mean square error (MSE))comme un outil de comparaison défini par

$$\frac{1}{30}\sum_{i=121}^{150}(Y_i - \widehat{\theta_{Xi}})^2$$

Il apparaît clairement que les deux types de régression ont un bon comportement mais l'MSE de R.L.L.T. est plus petite que R.L.L.C.

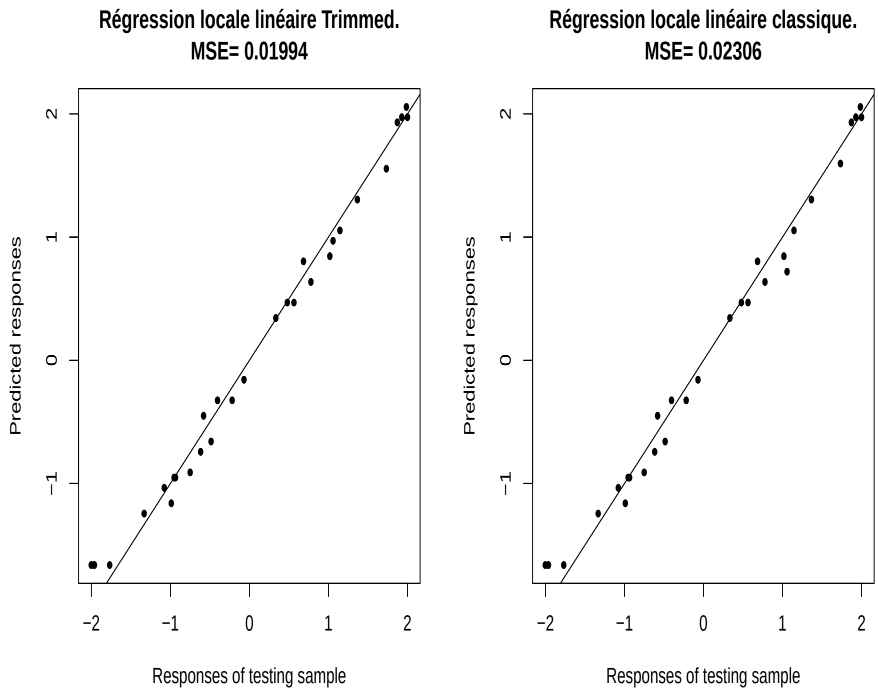

FIGURE 5.3 – Comparaison entre R.L.L.C. et R.L.L.T. en l'absence de valeurs aberrantes.

Rappelons que notre dernier objectif est de tester la sensibilité aux valeurs aberrantes des quatre types de régression. Pour cela, nous les comparons sur deux échantillons (l'un est de taille $n = 100$ et l'autre est de taille $n = 200$). Ensuite, nous divisons chaque échantillon en deux échantillons : échantillon d'apprentissage $(X_j, Y_j)_{j\in J}$ et échantillon de test $(X_i, Y_i)_{i\in I}$.
Afin d'introduire les valeurs aberrantes dans ces échantillons, nous multiplions par 100, $M\%$ des variables réponses de l'échantillon d'apprentissage $(Y_j)_{j\in J}$ et nous utilisons l'MSE pour évaluer cette comparaisons. les résultats sont illustrés dans la Table 5.1

ci-dessous.

Taille de l'échantillon	M %	R.C.	R.L.C.T.	R.L.L.C.	R.L.L.T.
n=100	0%	0.0307	0.02482	0.0270	0.0201
	1 %	0.1054	0.0504	0.0903	0.0381
	10 %	60.1298	7.8201	58.5147	5.9012
	25 %	491.7982	34.1369	452.1192	30.9012
n=200	0%	0.0196	0.0189	0.0192	0.0169
	1 %	0.0707	0.0314	0.0513	0.0271
	10 %	54.8051	6.6990	50.1261	4.7140
	25 %	363.3972	24.9225	300.3581	20.9271

Table 5.1 – Comparaison entre les quatre types de régression dans les deux cas (présence et l'absence de valeurs aberrantes)

On voit clairement que l'efficacité de ces fonctions varie en fonction de M. Cependant, les deux modèles R.L.C.T. et R.L.L.T. sont plus stables que R.C. et R.L.L.C. Au sens que les MSE de C.R. et L.L.R. augmentent considérablement par rapport aux valeurs de M, alors que la variabilité de cette erreur dans les modèles R.L.C.T. et R.L.L.T est très faible. Nous remarquons également que l'erreur pour R.L.L.C. et R.L.L.T. est plus petite que pour R.C. et R.L.C.T. dans la plupart des cas.

5.2 Quelque questions pratiques

5.2.1 Choix de la semi-métrique

Le choix de la semi métrique consiste une alternative intéressante à la problématique de dimension. De plus, ce choix a une influence positive sur la vitesse de convergence en augmentant la concentration de la mesure de probabilité sur les petites boules. Une question qui se pose souvent est comment choisir la semi métrique la mieux adaptée au données, surtout qu'il y a une grande diversité des semi-métriques que nous pouvons construire. Ferraty et al.(2002) ont proposé un exemple explicite de choix de semi-métrique. Cependant, à présent, il n'y a aucune méthode automatique qui permet d'offrir une semi-métrique optimale. Ferraty et Vieu (2006) ont présenté trois familles de semi-métriques mais bien sûr, beaucoup d'autres peuvent être construites. Nous présentons ici seulement deux familles de semi-métriques que nous avons utilisées dans notre simulation. La première a été proposée par Ferraty et Vieu (2004), elle est définie par

$$\forall (x_1, x_2) \in \mathcal{F} \times \mathcal{F} \quad d_q^{deriv}(x_1, x_2) = \sqrt{\int (x_1^{(q)}(t) - x_2^{(q)}(t))^2 dt}, \quad q \in \mathbb{N}$$

où $x^{(q)}$ est la dérivée d'ordre q. Notons que cette famille est bien adaptée aux courbes tout à fait lisses.
La deuxième est la semi-métrique de modèle d'indice fonctionnel, nous pouvons la définir, en prenant les outils de Ferraty et al(2002), par

$$\forall \theta, x_1, x2 \in \mathcal{H} \quad d_\theta(x_1, x_2) = |< \theta, x_1 - x_2 >|$$

où $\mathcal{H}$ est l'espace de Hilbert muni du produit scalaire $< \cdot, \cdot >$ dans lequel la variable fonctionnelle X est définie.
En pratique, ce paramètre θ peut être sélectionné par une technique de validation croisée (cf. Aït Saidi et al. (2008)). Dans nos simulations, nous avons sélectionné θ parmi les vecteurs propres de l'opérateur de covariance (voir Attaoui et al.(2011), Attaoui et Nengxiang (2015) pour plus de détails).

Il est fondamental de noter ici qu'il y a plusieurs familles de semi-métriques qui sont utilisées mais chacune d'entre elles est adaptée à certain genre de données, par exemple, la semi-métrique de l'analyse en composante principale (ACP) qui a été introduite par Besse et al.(1997) est bien adaptée aux données irrégulières, tandis que, la semi-métrique basée sur la dérivé sera employée quand nous avons des données lisses (ou régulières). Notons également que ce choix n'est pas lié seulement à la nature de données mais il faut également prendre en compte la nature de problème étudié.

5.2.2 Choix du paramètre de lissage

Il est bien connu qu'en estimation non paramétrique fonctionnelle, le choix du paramètre de lissage h joue un rôle crucial puisque ce dernier intervient dans la construction de l'estimateur ainsi que dans le terme de biais et dans le terme de dispersion. Précisément, une grande valeur de h nous conduit à une détérioration du biais alors qu'une trop petite valeur de h entraîne un estimateur à forte variabilité. En cette occurrence, nous avons besoin d'un paramètre de lissage qui optimise à la fois les deux termes asymptotiques afin d'assurer de bonnes performances sur le plans pratique.
À notre connaissance, la question sur le choix du paramètre de lissage optimale a été posée dans le cadre d'estimer une fonction de régression non paramétrique fonctionnelle par Rachdi et Vieu (2005a) et Rachdi et Vieu (2005b) dont le but est de construire un critère de choix automatique de ce paramètre, ce critère a été formulé sous la forme d'une validation croisée fonctionnelle, ainsi, ils ont montré, sous certaines conditions, que cette procédure est optimale. Nous citons aussi Rachdi et Vieu (2007), Benhenni et al.(2007) pour le choix optimal du paramètre de lissage dans la régression classique

et Madani et al. (2008) pour le cas de la densité conditionnelle. Notons que nombreuse autre techniques existent pour des variables à dimension finie (voir Härdle et Vieu (1992), Härdle et Simar(1993) et Youndjé (1993)).
Dans la littérature, on trouve plusieurs techniques de sélection du paramètres de lissage mais la méthode L^2 validation croisée est la plus utilisée pour la régression classique. Mais cette méthode n'est pas un sélecteur du paramètre de lissage satisfaisant car elle est trop influencée par des erreurs de prédiction extrêmes causées par l'existence de ces valeurs aberrantes. Le problème de choix de paramètre de lissage dans la régression robuste multivariée a fait l'objet de plusieurs travaux. Citons par-exemple Wang et Scott (1994) pour la version L^1 validation croisée qui a été considérée comme une alternative à la méthode de L^2 validation croisée. La généralisation de cette méthode a été réalisée par Leung (2005). Il s'agit d'une méthode de validation croisée plus robuste qui réduit les erreurs de prédiction extrêmes en remplaçant la fonction "valeur absolue" par une fonction "$\rho(\cdot)$" (cf. Attouch (2009), (Section 5.1.2.) et Atouch et al.(2009)). Dans nos simulations, nous avons choisi le paramètre de lissage par une adaptation de la méthode L^1 validation croisée sur le nombre de voisin le plus proche de manière locale (nous renvoyons à Ferraty et Vieu(2006), Section.., ... et ... pour plus de détail).

5.2.3 Choix des fonctions de localisation

Rappelons tout d'abord que l'idée de base de l'approche locale linéaire consiste à approximer localement le modèle non paramétrique par une fonction linéaire. Dans la statistique fonctionnelle, il existe plusieurs façons d'étendre cette approche. Nous citions par exemple Baìllo et Grané (2009) et Barrientos Marin et al. (2010). Rappelons également que dans nos travaux, nous avons adopté la version rapide proposée par Barrientos-Marin et al. (2010) pour laquelle la fonction θ_x est approximée par $a + b\beta(\cdot, x)$ (voir les chapitres précédents). Il est bien clair que le comportement de l'estimation locale linéaire dépend de choix d'opérateurs $\delta(\cdot,\cdot)$ et $\beta(\cdot,\cdot)$ qui jouent un rôle crucial. Concrètement, les résultats théoriques offrent, une grande souplesse dans le choix de ces deux opérateurs, sous certaines conditions (cf. les chapitres 2-4).
Barrientos-Marin et al. (2010) ont proposé des formes particulières de ces opérateurs que nous rappelons dans ce qui suit :
Si les données fonctionnelles sont des courbes "lisses", nous pouvons essayer d'utiliser la famille de fonctions de localisation qui est issue de la semi-métrique de l'indice fonctionnel en remplaçant le produit scalaire $< x_1, x_2 >$ par $< x_1^{(q)}, x_2^{(q)} >$ (cf. la sous-section 5.2.1). Plus précisément, nous considérons

$$loc_a^{(q)}(x_1, x_2) = \int \theta(t)(x_1^{(q)}(t) - x_2^{(q)}(t))dt =< \theta, x_1^{(q)} - x_2^{(q)} >_{\mathcal{F}},$$

où θ est une fonction donnée qui peut être adaptée aux données (voir la sous-section

5.2.1) et $x^{(q)}$ désigne la $q^{\text{ème}}$ dérivée de x). **Il est clair, qu'avec ce choix, on peut contrôler l'approximation des courbes avec signe (ou courbes signées).**
Les métriques, ou plus généralement les semi-métriques qui sont basées sur des dérivés, pourraient être également de bons candidats pour localiser une courbe par rapport à une autre. Par exemple, nous pouvons définir

$$loc_b^{(q)}(x_1, x_2) = \sqrt{\int (x_1^{(q)}(t) - x_2^{(q)}(t))^2 dt}$$

qui est une semi-métrique (voir la sous-section 5.2.1). Cette famille de fonctions de localisation est particulièrement bien adaptée à l'opérateur δ qui mesure la proximité entre deux éléments de $\mathcal{F}$. Il est important de noter qu'il y a de nombreuses fonctions de localisation qui peuvent être définies. En effet, nous pouvons construire des semi-métriques basées sur l'analyse fonctionnelle en composante principale (ACP) si nos données sont irrégulières. Nous nous référons au travail de Bouanani et al.(2018), dans lequel ils ont utilisé la semi-métrique (ACP) comme un choix de l'opérateur $\delta(\cdot,\cdot)$.
Les deux fonctions de localisation $\delta(\cdot,\cdot)$ et $\beta(\cdot,\cdot)$ désignent la localisation d'un élément de $\mathcal{F}$ par rapport à un autre, ce qui signifie que les deux fonctions de localisation peuvent prendre des valeurs négatives. Autrement dit, $\delta(\cdot,\cdot)$ et $\beta(\cdot,\cdot)$ permettent de calculer une proximité "signée" entre deux courbes. Remarquons que la théorie permet de prendre $\beta(\cdot,\cdot) = \delta(\cdot,\cdot)$ mais ces deux opérateurs ne jouent pas le même rôle. Précisément, $\beta(\cdot,\cdot)$ concerne la performance locale de la régression tandis que l'opérateur $\delta(\cdot,\cdot)$ se réfère à la pondération locale (voir la section 2 de l'article de Barrientas-Marin et al.(2010)). Dans nos simulations nous avons considéré $\beta(\cdot,\cdot) \neq \delta(\cdot,\cdot)$, ce choix peut nous conduire, d'une part, à un meilleur ajustement des données, et d'autre part, d'un point de vue pratique, à une méthode plus souple et adaptative . Pour des exemples concrets et des discussions plus approfondies sur ce sujet, nous renvoyons à (Barrientas-Marin et al.(2010)(Section 3),(Rachdi et al. (2014) (Section 5.1) et Xiongy et al. (2017) (Section 5), la section 4.4 et la section 5.1 de cette thèse).

Références

Aït Saidi, A., Ferraty ,F., Kassa, R. et Vieu, P.(2008). Cross-validated estimation in the single functional index model. *Statistics,* **42,** 475 -494.

Attaoui, S., Laksaci, A., Ould-Saïd, E. (2011). A note on the conditional density estimate in the single functional index model. *Stat Probab Lett,* **81**, 45-53.

Attaoui, S., Ling,N.(2015).Asymptotic results of a nonparametric conditional cumulative distribution estimator in the single functional index modeling for time series data with applications. *Metrika,* **79**, 485-511.

Attouch, M.(2009).Estimation robuste de la fonction de régression pour des variables fonctionnelles.*PhD Thesis.*

Attouch, M.,Laksaci, A., Ould Saïd, E.(2009).Asymptotic distribution of robust estimator for functional models.*Journal Korean statistical society,* **38,** 1317-1335.

Barrientos-Marin, J., Ferraty, F.,Vieu, P. (2010). Locally modeled regression and functional data. *J. of Nonparametric Statistics,* **22**, 617-632

Baìllo, A. and Grané, A. (2009). Local linear regression for functional predictor and scalar response. *J. Multivariate Anal.,* **100**, 102-111.

Benhenni, K., Ferraty, F., Rachdi, M., Vieu, P. (2007). Local smoothing regression with functional data. *Comput. Statist,* **22**, 353-369.

Besse, P., Cardot, H., Ferraty, F. (1997). Simultaneous nonparametric regressions of unbalanced longitudinal data. *Computational Statistics and Data Analysis,* **24,** 255-270.

Bouanani, O., Laksaci, A., Rachdi, M., Rahmani, S. (2018). Asymptotic normality of some conditional nonparametric functional parameters in high-dimensional statistics.*Journal of Behaviormetrika,* **45,** 1-35.

Ferraty, F., Vieu, P. (2002). The functional nonparametric model and application to spectrometric data. *Comput. Statist,* **17** , 545-564.

Ferraty, F., Vieu, P. (2004). Nonparametric models for functional data, with applications in regression, time series prediction and curves discrimination. *J. Nonparametr. Stat,* **16,** 11-127.

Ferraty, F., Vieu, P. (2006). Nonparametric functional data analysis. Theory and Practice. *Springer Series in Statistics. New York.*

Härdle,W., Simar,L. (1993). Computer Intensive methods in Statistics. *Statistics and Computing, Physica Verlag, Berlin,* 134-149.

Härdle, W., Vieu, P. (1992). Kernel regression smoothing of time series. *J. Time Ser. Anal,* **13** , 209-232.

Heng-Yan Leung D. (2005). Cross-validation in nonparametric regression with outliers. *The Ann. of Stat,* **33,** 2291-2310.

Madani,F.(2012).Aspects théoriques et pratiques dans l'estimation non paramétrique de la densité conditionnelle pour des données fonctionnelles. *PhD Thesis.*

Rachdi, M., Laksaci, A., Demongeot, J., Abdali, A. et Madani, F.(2014). Theoretical and practical aspects of the quadratic error in the local linear estimation of the conditional density for functional data. *Comput. Statist. Data Anal,* **73,** 53-68.

Rachdi, M., Vieu, P.(2005a). Sélection automatique du paramètre de lissage pour l'estimation non paramétrique de la régression pour des données fonctionnelles (in french) (2005a). *Compte Rendus Acad. Sci. Paris, Ser. I,* **341**, 365-368.

Rachdi, M., Vieu, P.(2005b). Nonparametric regression for functional data : automatic smoothing parameter selection. *Preprint* .

Rachdi, M., Vieu, P.(2007). Nonparametric regression for functional data : automatic smoothing parameter selection. *J. Statist. Plann. and Inference,* **137**, 2784-2801.

Xiong, X., Zhou, Z., Ailian, C. (2017). Asymptotic normality of the local linear estimation of the conditional density for functional time series data. *Journal of Communications in Statistics - Theory and Methods,***47,** 3418-3440.

Youndjé,E.(1993). Estimation non param´etrique de la densit´e conditionnelle par la m´ethode de noyau. *PhD Thesis, university of Rouen.*

CHAPITRE 6

ASYMPTOTIC PROPERTIES OF THE ROBUST LOCAL LINEAR REGRESSION WITH FUNCTIONAL REGRESSOR

Ce chapitre fait l'objet d'un rapport technique sur https://www.researchgate.net/profile/Ali_Laksaci

Asymptotic properties of the robust local linear regression with functional regressor

Faiza Belarbi , Souheyla Chemikh, Ali Laksaci

Laboratoire de statistique et processus stochastiques
Univ. Djillali Liabès
BP.22000 Sidi Bel Abbès, Algeria

e-mai : faiza_belarbi@yahoo.fr, chemikh_souheyla@yahoo.com and alilak@yahoo.fr

June 23, 2017

Abstract : *In this paper, we introduce a robust version of the local linear regression of a scalar response variable Y given a functional random variable X. We construct an estimator by combining both local linear ideas and an M-estimation techniques. The main results of this work are the establishment of the almost complete convergence as well as the asymptotic normality for the constructed estimator. These asymptotic results are stated with rate and are proved under a general condition.*

6.1 Introduction

The functional statistic is an important field in statistics, it concerns the modelisation of random variables takes values in infinite dimensional space. In practice, there is an increasing number of situations coming from different fields of applied sciences in which the data are of functional nature (such as soil science, geology, oceanography, econometrics, epidemiology, environmental science, forestry,...). In this paper, we are interested to the nonparametric estimation of the robust regression by using the local linear method.

It is well known that the robust regression has more advantages over the classical regression. In particular, it insensitive to the presence of outliers or the heteroscedasticity phenomena. This kind of regression model has been widely studied in nonparametric statistics. The literature on this subject dates back to Huber (1964) and has received, since, continuous developments (see for instance Robinson(1984) Collomb and Härdle (1968), Fan et al. (1994)) for previous results and Boente et al. (2009) for recent advances and references). Concerning the robust local linear approach, the literature is

limited. We cite, only Fan et al (1994), for the i.i.d. case, Cai and ouled-said (2003) for the α-mixing case. Recall that the local linear method is a more general alternative smoothing approach to the kernel method. See, Fan et Gijbels (1996) for more discussions on the importance of this approach.
The local linear method has been considered for functional data. In this area, the first results were obtained by Bàillo and Grané (2009). We return to Barrientos-Marin et al. (2010), Berlinet et al. (2011) for some alternatives versions of the functional local linear model and to Zhou and Lin(2016) for asymptotic results on this model. In parallel, the robust local constant method has been widely studied in functional data analysis. We cite, for instance, Attouch et al.(2009) for previous results and Derrar et al.(2016) for recent advances and references. The main purpose of this paper is to combine the ideas Barrientos-Marin et al (2010) on the functional local linear method with those of the robustness developed by Fan et al.(1994) to construct a smooth estimator resistant to the presence of outliers or heteroscedasticity of data and inherits many nice statistical properties of the local linear approach. Under a standards assumptions we establish the almost complete consistency (with rate) and the asymptotic normality of this estimate. This paper is organized as follows : We present our model in Section 6.2 . The needed conditions and the main results are given in Section 6.3. The proofs of the auxiliary results are relegated to the Appendix.

6.2 The robust local linear estimator

Consider n independent pairs of random variables (X_i, Y_i) for $i = 1, ..., n$ that we assume drawn from the pair (X, Y). The latter is valued in $\mathcal{F} \times \mathbb{R}$, where $\mathcal{F}$ is a semi-metric space and d denotes a semi-metric. The object of this paper is to study the co-variation between X_i and Y_i by the nonparametric robust regression function. For $x \in \mathcal{F}$ the nonparametric robust regression, denoted by θ_x, is defined as the unique minimizer of

$$\theta_x = \arg\min_{t \in \mathbb{R}} \mathbb{E}\left[\rho(Y - t)|X = x\right] \tag{6.1}$$

where $\rho(.)$ is a real-valued Borel function satisfying some regularity conditions to be stated below.
This kind of models belongs to the class of M-estimates introduced by Huber (1964). It covers and includes many usual nonparametric models, for example, for $\rho(y) = y^2$ we obtain the classical regression, $\rho(y) = |y||\alpha - \mathbb{1}_{y<0}|$ leads to the α^{th} conditional quantile. The α^{th} conditional expectile is obtained by setting $\rho(y) = y^2|\alpha - \mathbb{1}_{y<0}|$. For more others examples we refer the reader to Stone (2005).
Noting that the robust local linear estimation has investigated by many authors in the multivariate case. We refer to Fan et al.(1994), for the i.i.d. case, Cai and Ouled-Said (2003) for the α-mixing case and Hallinet al.(2009) for the spatial case. In this work we will focus on the case where the covariate X is of infinite dimension. It should be

noted that, these questions of statistical analysis of infinite dimensional data are arising more and more in applied statistics. For an overview on recent developments in this topic, we cite, for example, Cuevas(2014), Zhang(2013) and Hsing and Eubank (2015). Recall that the basic idea of local linear fitting consists in approximating locally the nonparametric model by a linear function. In functional statistics, there are several ways for extending this approach (see, Baillo and Grané (2009) or Barrientos-Marinet al.(2010) for some examples). In this paper, we adopt the fast version proposed by Barrientos-Marinet al.(2010)for which the function θ_x is approximated by

$$\forall z \qquad \text{in neighborhood of } x \qquad \theta_z = a + b\beta(x, z) \tag{6.2}$$

where a and b are estimated by $\widehat{a}$ and $\widehat{b}$ are solution of

$$\min_{(a,b)\in\mathbb{R}^2} \sum_{i=1}^{n} \rho(Y_i - a - b\beta(X_i, x))K(h^{-1}\delta(x, X_i)) \tag{6.3}$$

where $\beta(\cdot,\cdot)$ is a know function from $\mathcal{F}\times\mathcal{F}$ into $\mathbb{R}$ such that, $\forall\xi \in \mathcal{F}, \beta(\xi,\xi) = 0$, with K is kernel and $h = h_n$ is a sequence of positive real numbers and $\delta(\cdot,\cdot)$ is a function of $\mathcal{F}\times\mathcal{F}$ such that $d(\cdot,\cdot) = |\delta(\cdot,\cdot)|$. It is clear that, under this consideration, we can write

$$\theta_x = a \quad \text{and} \quad \widehat{\theta_x} = \widehat{a}.$$

We point out that unlike to the classical regression case studied by Barrientos-Marin el al. (2010) the robust local linear estimator $\widehat{\theta_x}$ cannot explicitly expressed. Thus, the establishment of the asymptotic proprieties of our estimate is very difficult, it requires some additional tools.

6.3 Main results

In what follows, when no confusion is possible, we will denote by C and C' some strictly positive generic constants. Moreover, x denotes a fixed point in $\mathcal{F}, N_x$ denotes a fixed neighborhood of x. For $i = 1, ..., n$, we denote by $K_i = K(h^{-1}\delta(x, X_i))$, and $\beta_i = \beta(X_i, x)$. Furthermore, we put $\phi_x(r_1, r_2) = \mathbb{P}(r_2 \leq \delta(x, X) \leq r_1)$ and we assume the following hypotheses :

(H1) For any $r > 0$, $\phi_x(r) := \phi_x(-r, r) > 0$ and there exists a function $\chi_x(.)$ such that :

$$\forall t \in (-1, 1), \lim_{h\to 0} \frac{\phi_x(th, h)}{\phi_x(h)} = \chi_x(t).$$

(H2)ρ is a strictly convex function, continuously differentiable and has a Lipschitzian derivative ψ such that

$$\mathbb{E}[|\psi(Y - t)^p|X = x] < C < \infty \qquad \text{almost surely,} \qquad p > 2.$$

(H3) The function $\Gamma(x,\cdot) := \mathbb{E}[|\psi(Y-\cdot)||X=x]$ is of class $\mathcal{C}^1$ on $[\theta_x-\delta,\theta_x+\delta]$: $\delta>0$. Such that :

$$\begin{cases} (i)\ \forall(t_1,t_2)\in[\theta_x-\delta,\theta_x+\delta]^2, \qquad \forall(x_1,x_2)\in\mathcal{N}_x\times\mathcal{N}_x, \\ \qquad |\Gamma(x_1,t)-\Gamma(x_2,t)|\le Cd^{k_1}(x_1,x_2)+|t_1-t_2|^{k_2}. \\ (ii)\ \text{The variable}\ \delta(x,X)\ \text{is}\ \sigma(\beta(X,x))\text{-measurable and the two partial derivatives} \\ \qquad \text{of the function, } \Upsilon_x(s,t)=\mathbb{E}[\Gamma(X,t)|\beta(X,x)=s])\ \text{at}(0,\theta_x)\ \text{exist.} \end{cases}$$

(H4) The function $\beta(\cdot,\cdot)$ is such that :

$$\forall z\in\mathcal{F},\ C|\delta(x,z)|\le|\beta(x,z)|\le C'|\delta(x,z)| \quad \text{and} \quad \sup_{u\in B(x,r)}|\beta(u,x)-\delta(x,u)|=o(r).$$

(H5) The kernel K is a positive, differentiable function within $(-1,1)$ such that

$$\begin{pmatrix} K(1)-\int_{-1}^1 K'(t)\chi_x(t)dt & K(1)-\int_{-1}^1 (tK(t))'\chi_x(t)dt \\ K(1)-\int_{-1}^1 (tK(t))'\chi_x(t)dt & K(1)-\int_{-1}^1 (t^2K(t))'\chi_x(t)dt \end{pmatrix} \text{ is a positive definite matrix}$$

(H6) The bandwidth h satisfies : $\lim_{n\to\infty}\frac{\log n}{n\phi_x(h)}=0$

Our main results are summarized in the following Theorems

Theorem 6.1. *Under hypotheses (H1), (H2), (H3(i)), (H4)-(H6) and if* $\frac{\partial\Gamma}{\partial t}(x,\theta_x)>0$ *then*

$$|\widehat{\theta_x}-\theta_x|=O(h^{\min(k_1,k_2)})+O\left(\left(\frac{\log(n)}{n\phi_x(h)}\right)^{1/2}\right) \qquad a.co.$$

Theorem 6.2. *Under hypotheses (H1)-(H6) and if* $\frac{\partial\Gamma}{\partial t}(x,\theta_x)>0$, *for any* $x\in\mathcal{A}$ *then*

$$\left(\frac{n\phi_x(h)}{\sigma^2(x)}\right)^{1/2}(\widehat{\theta_x}-\theta_x-o(h))\xrightarrow{\mathcal{D}}\mathcal{N}(0,1) \qquad as \quad n\to\infty$$

where $\xrightarrow{\mathcal{D}}$ *denote convergence in distribution,*

$$\sigma^2(x)=\frac{\mathbb{E}[\psi^2(Y-\theta_x)|X=x]}{(\frac{\partial\Gamma}{\partial t}(x,\theta_x))^2}\frac{a_3^2D_1-2a_2a_3D_2+a_2^2D_3}{(a_1a_3-a_2^2)^2} \quad et\ \mathcal{A}=\{x\in\mathcal{F},\ \sigma^2(x)\neq0\}$$

with

$$a_j(x)=K(1)-\int_{-1}^1(s^{j-1}K(s))'\chi_x(s)ds \quad and \quad D_j(x)=K(1)-\int_{-1}^1(s^{j-1}K(s)^2)'\chi_x(s)ds \ \ for\ j=1,2,3. \tag{6.4}$$

6.4 Appendix

Firstly, we state the following lemmas which are needed to establish our asymptotic results The proof of both theorems is based on the following lemma

Lemma 6.1. *Let V_n a sequence of vectorial function such that*

i) For all $\lambda \geq 1$ and the vector δ

$$^{t}(\delta)V_n(\lambda\delta) \leq \ ^{t}(\delta)V_n(\delta)$$

ii) For a positive definite matrix D and vectorial sequences A_n such that $\mathbb{P}(A_n \geq 0) \to 0$, for some $A > 0$ we have

$$\sup_{||\delta|| \leq M} ||V_n(\delta) + \lambda_0 D\delta - A_n|| = o_p(1) \quad for \quad \frac{A}{\lambda_0\lambda_1(D)} < M < \infty.$$

with $\lambda_1(D)$ is the minimum eigenvalue of D and $\lambda_0 > 0$.

Then for any vectorial sequence δ_n such that $V_n(\delta_n) = o_p(1)$ we have

$$||\delta_n|| \leq M \qquad In \ \ probability. \tag{6.5}$$

Furthermore, if $||A_n|| = o_{a.co.}(1)$ and $\sup_{||\delta|| \leq M} ||V_n(\delta) + \lambda_0 D\delta - A_n|| = o_{a.co.}(1)$.

Then, for any vectorial sequence δ_n such that $V_n(\delta_n) = o_{a.co.}(1)$ we have

$$||\delta_n|| \leq M \qquad almost \ \ completely. \tag{6.6}$$

Proof of Lemma 6.1

The The proof of both results is similar. It is based on the same arguments as in Koenker and Zhao (1996). For the sake of shortness, we prove the second case which is more general. Indeed, for $\eta > 0$, we have

$$\begin{aligned} \mathbb{P}(\| \delta_n \| \geq M) &= \mathbb{P}(\| \delta_n \| \geq M, \| V_n(\delta_n) \| < \eta) + \mathbb{P}(\| V_n(\delta_n) \| \geq \eta) \\ &\leq \mathbb{P}\left(\inf_{\|\delta\| \geq M} \| V_n(\delta) \| < \eta\right) + \mathbb{P}(\| V_n(\delta_n) \| \geq \eta) \end{aligned}$$

Since, for any $\eta > 0$,

$$\sum_n \mathbb{P}(\| V_n(\delta_n) \| \geq \eta) < \infty$$

then, all it remains to show that

$$\sum_n \mathbb{P}\left(\inf_{\|\delta\| \geq M} \| V_n(\delta) \| < \eta\right) < \infty.$$

Now, it is clear that for all vectorδ such that $\| \delta \| \geq M$ can be written as $\delta = \lambda\delta_1$ for

$\lambda \geq 1$ and $\parallel \delta \parallel = M$. So, by condition i) we have

$$\parallel V_n(\delta) \parallel = \left\| -\frac{{}^t\delta_1 V_n(\delta)}{M} \right\| \geq -\frac{{}^t\delta_1 V_n(\delta)}{M} \geq -\frac{{}^t\delta_1 V_n(\delta_1)}{M}$$

which implies that

$$\mathbb{P}\left(\inf_{\|\delta\| \geq M} \|V_n(\delta) < \eta \right) \leq \mathbb{P}\left(\inf_{\|\delta_1\|=M} \left[-\frac{{}^t\delta_1 V_n(\delta_1)}{M} \right] < \eta \right).$$

Thus, it suffices to evaluate this last quantity. To do that, we write

$$\begin{aligned}
\mathbb{P}\left(\inf_{\|\delta_1\|=M} \left[-\frac{{}^t\delta_1 V_n(\delta_1)}{M} \right] < \eta \right) &\leq \mathbb{P}(\inf_{\|\delta_1\|=M} \left[-{}^t\delta_1 V_n(\delta_1) \right] < \eta M, \\
&\quad \inf_{\|\delta_1\|=M} \left[{}^t\delta_1(-\lambda_0 D \delta_1 + A_n) \right] \geq 2\eta M) \\
&\quad + \mathbb{P}\left(\inf_{\|\delta_1\|=M} \left[{}^t\delta_1(-\lambda_0 D \delta_1 + A_n) \right] \geq 2\eta M \right) \\
&\leq \mathbb{P}\left(\sup_{\|\delta_1\|=M} \|V_n(\delta_1) + \lambda_0 D \delta_1 - A_n\| \geq \eta \right) \\
&\qquad + \mathbb{P}(\|A_n\| \geq \lambda_0 \lambda_1(D) M - 2\eta).
\end{aligned}$$

Finally, under the condition (ii), we can chose η for witch

$$\sum_n \mathbb{P}\left(\sup_{\|\delta\|=M} \|V_n(\delta_1 + \lambda_0 - A_n\| \geq \eta \right) < \infty$$

and

$$\sum_n \mathbb{P}\left(\parallel A_n \parallel \geq \lambda_0 \lambda_1(D) M - 2\eta \right) < \infty.$$

It follows that

$$\sum_n \mathbb{P}\left(\parallel \delta_n \parallel \geq M \right) < \infty.$$

Proof of Theorems :

The proof of Theorem is based on the fact that $\psi$is monotone function, We give the proof for the case of an increasing, the decreasing case begin obtained by replacing ψ by $-\psi$. Under this consideration, we define, for $\delta = \begin{pmatrix} c \\ d \end{pmatrix}$, and $\delta_0 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$, for following functions

$\varphi(\delta) = \psi(Y_i - ((c+a) + (h_n^{-1} d + b)\beta_i))$ and $\varphi'(\delta) = \psi\left(Y_i - \left(\left(\frac{1}{\sqrt{n\phi_x(h)}} c + a \right) + \left(\frac{1}{h_n \sqrt{n\phi_x(h)}} d + b \right) \right.\right.$

and the following vectorial sequences

$$V_n(\delta) = \frac{1}{n\phi_x(h)} \sum_{i=1}^{n} \varphi(\delta) \begin{pmatrix} 1 \\ h^{-1}\beta_i \end{pmatrix} K_i, \quad A_n = V_n(\delta_0), \quad \delta_n = \begin{pmatrix} \widehat{a} - a \\ h(\widehat{b} - b) \end{pmatrix}$$

and

$$V_n'(\delta) = \frac{1}{\sqrt{n\phi_x(h)}} \sum_{i=1}^{n} \varphi'(\delta) \begin{pmatrix} 1 \\ h^{-1}\beta_i \end{pmatrix} K_i, \quad A_n' = V_n'(\delta_0) \quad \text{and } \delta_n' = \sqrt{n\phi_x(h)}\delta_n.$$

Now, the proof of the first result of Theorem 6.1 is based on application of the seconde part of Lemma 6.1 to (V_n, A_n, δ_n). While the seconde result is obtained by applying the first part of Lemma 6.1 to (V_n', A_n', δ_n'). Clearly, for both cases the condition(i) of Lemma 6.1 is holds, because ψ is monotone increasing function.
Thus, the first result of Theorem 6.1 is consequence of following lemmas :

Lemma 6.2. *Under Hypotheses (H1),(H2),(H3(i)) and (H4)-(H6), we have*

$$\|A_n\| = O(h^{\min(k_1,k_2)}) + O_{a.co}\left(\left(\frac{\log(n)}{n\phi_x(h)}\right)^{1/2}\right).$$

Proof of lemma 6.2

We write $$A_n - \mathbb{E}[A_n] = \begin{pmatrix} A_n^1 := \dfrac{1}{n\phi_x(h)} \sum_{i=1}^{n} \Delta_i^1 \\ A_n^2 := \dfrac{1}{nh\phi_x(h)} \sum_{i=1}^{n} \Delta_i^2 \end{pmatrix}.$$ Where

$$\Delta_i^1 = \varphi(\delta_0)K_i - \mathbb{E}[\varphi(\delta_0)K_i]. \quad \text{and} \quad \Delta_i^2 = \varphi_i(\delta_0)\beta_i^1 K_i - \mathbb{E}[\varphi(\delta_0)\beta_i^1 K_i].$$

From (H2) and (H3) we get

$$\mathbb{E}[|\Delta_i^j|^m] \leq \frac{C}{\phi_x(h)^{m-1}} \qquad \text{for} \qquad j = 1.2.$$

Then we can apply the exponential inequality for unbounded variables, we obtain from Corollary A.8 in Ferraty and Vieu with $a^2 = O\left(\frac{1}{\phi_x(h)}\right)$

$$A_n^1 - \mathbb{E}[A_n^1] = O_{a.co0}\left(\sqrt{\frac{\log(n)}{n\phi_x(h)}}\right) \qquad \text{and} \qquad A_n^2 - \mathbb{E}[A_n^2] = O_{a.co0}\left(\sqrt{\frac{\log(n)}{n\phi_x(h)}}\right).$$

On the other hand, by (H3), we obtain

$$\begin{aligned}\mathbb{E}[A_n^1] &= \frac{1}{\phi_x(h)}|\Gamma(x,\theta_x) - \Gamma(X, a + b\beta_1)|] \\ &= O(h^{k_1}) + o(h^{k_2})\end{aligned}$$

Similarly

$$\mathbb{E}[A_n^2] = O(h^{k_1}) + o(h^{k_2})$$

Which achieves the proof of Lemma

Lemma 6.3. *Under Hypotheses (H1),(H2),(H3(i))and (H4)-(H6), we have*

$$\sup_{||\delta|| \le M} ||V_n(\delta) + \lambda_0 D\delta - A_n|| = o_{a.co.}(1).$$

with $D = \begin{pmatrix} K(1) - \int_{-1}^{1} K'(t)\chi_x(t)dt & K(1) - \int_{-1}^{1}(tK(t))'\chi_x(t)dt \\ K(1) - \int_{-1}^{1}(tK(t))'\chi_x(t)dt & K(1) - \int_{-1}^{1}(t^2K(t))'\chi_x(t)dt \end{pmatrix}$

and $\lambda_0 = \frac{\partial \Gamma}{\partial t}(x,\theta_x)$.

Proof of Lemma 6.3

The proof is based on the following assertions

$$\sup_{||\delta|| \le M} ||\mathbb{E}[V_n(\delta) - A_n] + \frac{\partial \Gamma}{\partial t}(x,\theta_x)D\delta|| = O(h^{\min(k_1,k_2)}) \quad (6.7)$$

$$\sup_{||\delta|| \le M} ||V_n(\delta) - A_n - \mathbb{E}[V_n(\delta) - A_n]|| = O_{a.co.}\left(\sqrt{\frac{\log(n)}{n\phi_x(h)}}\right). \quad (6.8)$$

For the first result, we write

$$V_n(\delta) - A_n = \begin{pmatrix} W_n^1(\delta) := \frac{1}{n\phi(h)} \sum_{i=1}^{n} (\varphi(\delta) - \varphi(\delta_0))K_i. \\ W_n^2(\delta) := \frac{1}{nh\phi(h)} \sum_{i=1}^{n} (\varphi(\delta) - \varphi(\delta_0))\beta_i K_i. \end{pmatrix}$$

Observe that, by (H3) we get

$$\begin{aligned}\mathbb{E}[W_n^1(\delta)] &= \frac{1}{\phi_x(h)}\mathbb{E}[(\Gamma(x,(c+a) + (h^{-1}d + b)\beta_1) - \Gamma(x, a + b\beta_1))K_1] + O(h^{k_1}) \\ &= \frac{1}{\phi_x(h}\mathbb{E}\left[\frac{\partial \Gamma}{\partial t}(x, a + b\beta_1)\left(1, h^{-1}\beta_1\right)\delta K_1\right] + O(h^{k_1}) + o(||\delta||) \\ &= \frac{\partial \Gamma}{\partial t}(x,\theta_x)\frac{1}{\phi_x(h)}\left(\mathbb{E}[K_1], h^{-1}\mathbb{E}[\beta_i K_1]\right)\delta + O(h^{\min(k_1,k_2)}) + o(||\delta||).\end{aligned}$$

and by the same ideas we calculate $\mathbb{E}[W_n^2(\delta)]$. Therefore,

$$\mathbb{E}[V_n(\delta)-A_n] = \frac{\partial\Gamma}{\partial t}(x,\theta_x)\frac{1}{\phi_x(h)}\begin{pmatrix} \mathbb{E}[K_i] & \mathbb{E}[K_i h^{-1}\beta_i] \\ \mathbb{E}[K_i h^{-1}\beta_i] & \mathbb{E}[h^{-2}\beta_i^2 K_i^2] \end{pmatrix}\delta + O(h^{\min(k_1,k_2)}) + o(||\delta||).$$

It is show in Demongeot et al. (2013), under the seconde part of (H4) that

$$h_n^{-a}\mathbb{E}[\beta^a K_i^b] = \phi_x(h)\left(K^b(1) - \int_{-1}^{1}(u^a K(u))'\chi_x(u)du \right) + o(\phi_x(h)).$$

Consequently

$$\sup_{||\delta||\leq M} ||\mathbb{E}[V_n(\delta)-A_n] + \frac{\partial\Gamma}{\partial t}(x,\theta_x)D\delta + o(||\delta||)|| = O(h^{\min(k_1,k_2)}).$$

Which implies the result (6.7).
Concerning (6.8), we use the compactness of the ball $B(0,M)$ in $\mathbb{R}^2$ and we write

$$B(0,M) \subset \bigcup_{j=1}^{d_n} B(\delta_j, l_n) \quad \delta_j = \begin{pmatrix} c_j \\ d_j \end{pmatrix}, \quad \text{and} \quad l_n = d_n^{-1} = 1/\sqrt{n}.$$

Let $j(\delta) = \text{argmin}_j |\delta - \delta_j|$, then can write

$$\sup_{||\delta||\leq M} ||V_n(\delta) - A_n - \mathbb{E}[V_n(\delta)-A_n]|| \leq \sup_{||\delta||\leq M} ||V_n(\delta) - V_n(\delta_j)||$$

$$+ \sup_{||\delta||\leq M} ||V_n(\delta_j) - A_n - \mathbb{E}[V_n(\delta_j)-A_n]|| + \sup_{||\delta||\leq M} ||\mathbb{E}[V_n(\delta) - V_n(\delta_j)]||.$$

For the first and last terms we use the fact that ψ is a Lipschitzian function and K is bounded to prove that

$$\sup_{||\delta||\leq M} ||V_n(\delta) - V_n(\delta_j)|| = O_{a.co}\left(\left(\frac{\log(n)}{n\phi_x(h)}\right)^{1/2}\right)$$

and

$$\sup_{||\delta||\leq M} ||\mathbb{E}[V_n(\delta) - V_n(\delta_j)]|| = o_{a.co}\left(\left(\frac{\log(n)}{n\phi_x(h)}\right)^{1/2}\right).$$

Now, for the quantity

$$\sup_{||\delta||\leq M} ||V_n(\delta_j) - A_n - \mathbb{E}[V_n(\delta_j)-A_n]||$$

We put $V_n(\delta_j)-A_n-\mathbb{E}[V_n(\delta_j)-A_n] = W_n(\delta_j)-\mathbb{E}[W_n(\delta_j)] = \begin{pmatrix} W_n^1(\delta_j) := \dfrac{1}{n\phi_x(h)}\sum_{i=1}^n \Lambda_i^1 \\ W_n^2(\delta_j) := \dfrac{1}{nh\phi_x(h)}\sum_{i=1}^n \Lambda_i^2 \end{pmatrix}$

where $\Lambda_i^1 = (\varphi(\delta_j) - \varphi(\delta_0))K_i - \mathbb{E}[(\varphi(\delta_j) - \varphi(\delta_0))K_i]$.
and $\Lambda_i^2 = (\varphi(\delta_j) - \varphi(\delta_0))\beta_i K_i - \mathbb{E}[(\varphi(\delta) - \varphi(\delta_0))\beta_i K_i]$.
Once again we apply the exponential inequality for unbounded variables on Λ_i^j for which

$$\mathbb{E}[|\Lambda_i^j|^m] \leq \frac{C}{\phi_x(h)^{m-1}} \qquad \text{for} \qquad j = 1.2.$$

It follows that, there exists $\eta > 0$ such that

$$\sum_n \mathbb{P}\left(\sup_{\|\delta\| \leq M} \|V_n(\delta_j) - A_n - \mathbb{E}[V_n(\delta_j) - A_n]\| > \eta\sqrt{\frac{\log n}{n\phi_x(h)}}\right)$$
$$\leq \sum_n d_n \mathbb{P}\left(V_n(\delta_j) - A_n - \mathbb{E}[V_n(\delta_j) - A_n]\| > \eta\sqrt{\frac{\log n}{n\phi_x(h)}}\right) \leq \infty$$

which complete the proof of (6.8)
Concerning the asymptotic normality result we use the following Lemmas :

Lemma 6.4. *Under Hypotheses (H1), (H2), (H3(i)) and (H4)-(H6), we have*

$$\|A_n^{'}\| = O_p(1) \quad \text{and} \quad \sup_{\|\delta^{'}\| \leq M} \|V_n^{'}(\delta) + \frac{\partial\Gamma}{\partial t}(x, \theta_x)D\delta^{'} - A_n^{'}\| = o_p(1).$$

Proof of the first part

It suffices to prove that

$$n\phi_x(h)Var[c_1 A_n^1 + c_2 A_n^2] = O(1), \quad for \quad \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} \in mathrmIR^2$$

where A_n^1 and A_n^2 are defined in the proof of Lemma 6.2. The latter is a simple consequence of the fact that

$$\frac{1}{\phi_x(h)}Var[\Delta_1^1] \to \mathbb{E}\left[\psi^2(Y - \theta_x)|X = x\right]\left(K^2(1) - \int_{-1}^1 (K^2(t))^{'}\chi_x(t)dt\right),$$

$$\frac{1}{h^2\phi_x(h)}Var[\Delta_1^2] \to \mathbb{E}\left[\psi^2(Y - \theta_x)|X = x\right]\left(K^2(1) - \int_{-1}^1 (t^2K^2(t))^{'}\chi_x(t)dt\right)$$

and

$$\frac{1}{h\phi_x(h)}cov(\Delta_1^1, \Delta_1^2) \to \mathbb{E}\left[\psi^2(Y - \theta_x)|X = x\right]\left(K^2(1) - \int_{-1}^1 (tK^2(t))^{'}\chi_x(t)dt\right).$$

Proof of the second part :

The proof is based on the same assertions of Lemma 6.3. Firstly, by following the same lines as for (6.7) we have

$$\mathbb{E}[V_n^{'}(\delta) - A_n^{'}] = \frac{\partial \Gamma}{\partial t}(x, \theta_x)\frac{1}{\phi_x(h)}\begin{pmatrix} \mathbb{E}[K_i] & \mathbb{E}[K_i h^{-1}\beta_i] \\ \mathbb{E}[K_i h^{-1}\beta_i] & \mathbb{E}[h^{-2}\beta_i^2 K_i] \end{pmatrix}\delta + o(1) + o(\|\delta\|).$$

Now, for the dispersion part we keep the notation of Lemma 6.3 with another choice of $l_n d_n$. Specifically, we put $l_n = d_n^{-1} = O(1/\log n)$. By evaluating the variance and the expectation of the quantity $\sup_{\|\delta\|\leq M}\|V_n(\delta) - V_n(\delta_j)\|$ we get

$$\sup_{\|\delta\|\leq M}\|V_n^{'}(\delta) - V_n^{'}(\delta_j)\| = o_p(1) \quad and \quad \sup_{\|\delta\|\leq M}\|\mathbb{E}[V_n^{'}(\delta) - V_n^{'}(\delta_j)] \leq Cl_n = o(1). \tag{6.9}$$

Furthermore, for the quantity

$$V_n^{'}(\delta_j) - A_n^{'} - \mathbb{E}[V_n^{'}(\delta_j) - A_n^{'}] = \begin{pmatrix} \sqrt{n\phi_x(h)}W_n^1(\delta_j) \\ \sqrt{n\phi_x(h)}W_n^2(\delta_j) \end{pmatrix}$$

we prove by a simple calculation

$$\begin{gathered} Var|\sqrt{n\phi_x(h)}(W_n^1(\delta_j) - \mathbb{E}[W_n^1(\delta_j)]) = O\left(\frac{1}{\sqrt{n\phi_x(h)}}\right) \\ and \\ Var|\sqrt{n\phi_x(h)}(W_n^2(\delta_j) - \mathbb{E}[W_n^2(\delta_j)]) = O\left(\frac{1}{\sqrt{n\phi_x(h)}}\right) \end{gathered} \tag{6.10}$$

Thus, the proof of this Lemma is consequence of (H6), (6.9) and (6.10)

Corollary 6.1. *Under Hypotheses of Lemma 6.4, we have*

$$\widehat{\theta}_x - \theta_x = \frac{1}{\frac{\partial \Gamma}{\partial t}(x, \theta_x)(a_1 a_3 - a_2^2)}\widehat{\Phi}(x) + o_p(1)$$

where $\widehat{\Phi}(x) = \frac{1}{n\phi_x(h)}\sum_{i=1}^{n}\varphi^{'}(\delta_0)(A_3 K_i - A_2 h^{-1}\beta_i K_i)$ *with* $A_j = \phi_x^{-1}(h)h^{1-j}\mathbb{E}[\beta^{j-1}K]$.

Corollary 6.1

As $\|\delta_n^{'}\| \leq M$ and $V_n^{'}(\delta_n^{'}) = 0$ then, form Lemma 6.4

$$\delta_n^{'} = \frac{1}{\frac{\partial \Gamma}{\partial t}(x, \theta_x)}D^{-1}A_n^{'} + o_p(1).$$

Finally, it suffices to replace $A_n^{'}$ and D^{-1} by their expressions and to use the fact that $A_j = a_j + o(1)$ to achieve the proof of this Corollary.

Lemma 6.5. *Under the hypotheses of Theorem 6.2, we have,*

$$\mathbb{E}\left[\widehat{\Phi}(x)\right] = o(h)$$

and

$$\sqrt{n\phi_x(h)}\sigma_1^{-1/2}\left(\widehat{\Phi}(x) - \mathbb{E}\left[\widehat{\Phi}(x)\right]\right) \xrightarrow{\mathcal{D}} \mathcal{N}(0,1) \quad as \quad n \to \infty$$

where

$$\sigma_1 = \mathbb{E}\left[\psi^2(Y-\theta_x)|X=x\right](a_3^2 D_1 - 2a_2 a_3 D_2 + a_2^2 D_3)$$

Proof of Lemma 6.5

For the first result : We use the same arguments as in Ferraty et al .(2007) we show that

$$\begin{aligned}
\mathbb{E}\left[\widehat{\Phi}(x)\right] &= \mathbb{E}[\Upsilon_x(\beta, \theta_x + b\beta)(A_3K - A_2h^{-1}\beta K)] \\
&= \Upsilon_x(0,\theta_x)\mathbb{E}[(A_3K - A_2h^{-1}\beta K)] \\
&\quad + \frac{\partial}{\partial s}\Upsilon_x(0,\theta_x)\mathbb{E}[\beta(A_3K - A_2h^{-1}\beta K)] \\
&\quad + b\frac{\partial}{\partial t}\Upsilon_x(0,\theta_x)\mathbb{E}[\beta(A_3K - A_2h^{-1}\beta K)] \\
&\quad + \mathbb{E}[o(\beta)(A_3K - A_2h^{-1}\beta K)].
\end{aligned}$$

Since $\mathbb{E}[\beta(A_3K - A_2h^{-1}\beta K)] = 0$ and $\Upsilon_x(0,\theta_x) = 0$ we obtain

$$\mathbb{E}[\widehat{\Phi}(x)] = \mathbb{E}[o(\beta)(A_3K - A_2h^{-1}\beta K)] = o(h).$$

For the second result : We use Liapounov's theorem (see Loeve, 1947(P.275)). Firstly, from Lemma 6.4, we have

$$n\phi_x(h)Var\left[\widehat{\Phi}(x)\right] \to \mathbb{E}[\psi^2(Y-\theta_x)|X=x]a_3^2D_1 - 2a_2a_3D_2 + a_2^2D_3 \quad as \quad n \to \infty. \tag{6.11}$$

Secondly, by using the same ideas of Atouch et al. (2009) we state the required conditions of Liapounov's Theorem that is

$$\frac{\sum_{i=1}^{n}\mathbb{E}[|L_i - \mathbb{E}[L_i]|^{2+\delta}]}{\left(Var\left(\sum_{i=1}^{n}L_i\right)\right)^{(2+\delta)/2}} \to 0 \quad as \quad n \to \infty. \quad for \;\; some \quad \delta > 0 \tag{6.12}$$

where

$$L_i = \frac{1}{n\phi_x(h)}(a_3 K_i - a_2 h^{-1} \beta_i K_i \varphi(\delta_0))$$

which concludes the proof.

CONCLUSION ET PERSPECTIVES

Conclusion générale

Dans cette thèse, nous nous sommes focalisés sur l'estimation locale linéaire de la régression robuste lorsque la variable reponse est réelle et la variable explicative est de type fonctionnel. Il est bien connu que la technique de M-estimation a la capacité d'être relativement insensible aux valeurs aberrantes, contrairement à la régression usuelle qui est incapable devant ce genre de valeurs. Il est également bien connu que l'approche locale linéaire est une méthode d'estimation alternative à la méthode à noyau. Dans ce manuscrit nous avons combiné les idées locales linéaires et les techniques de M-estimation afin de mettre en évidence la supériorité de ces deux approches en NPFDA qui est la motivation principale de ce sujet. Nous avons étudié dans un premier temps, la convergence presque-complète et la normalité asymptotique de l'estimateur local linéaire de la régression robuste dans le cas i.i.d. Ensuite, nous avons construit l'estimateur local linéaire de la régression Trimmed qui est considéré comme un cas particulier de l'estimateur précédent. Comme résultats asymptotiques, nous avons établi la vitesse de convergence presque- complète et la normalité asymptotique de l'estimateur construit dans un cadre fonctionnel indépendant.
Le dernier problème traité dans cette thèse est le cas où les observations sont α- mélangeantes. Dans ce cas, nous avons étudié la vitesse de convergence presque-complète et la normalité asymptotique de l'estimateur local linéaire de la régression robuste. Cet estimateur généralise plusieurs estimateurs de la fonction de régression, dans le cadre non paramétrique, nous citons par exemple, l'estimateur proposé par Attouch et al. (2012). Nous avons aussi donné dans cette partie, un exemple d'application sur des donnée simulées dont l'objectif est l'étude comparative entre quelques types de régression. Il est bien clair que nos résultats théoriques confirment que nos estimateurs possèdent de bonnes propriétés asymptotiques dans tous les cas de notre étude. Plus précisément, la convergence presque complète couvre tous les différents types de

convergence stochastique et la normalité asymptotique nous permet de construire des intervalles de confiance ou de faire des tests statistiques. Notons que ces résultats théoriques ont été obtenus sous des hypothèses qui sont divisées en trois catégories, des hypothèses structurales, des hypothèses sur la variable explicative et des hypothèses techniques.
La dernière partie de ce manuscrit a été consacré à un exemple d'application sur des données simulées, A partir de cette étude applicative nous avons pu constater que la régression locale linéaire Trimmed propose des outils de prévisions alternatives à la régression locale constante Tirmmed, la régression locale linéaire et à la régression classique avec une précision plus efficace, notamment en présence des valeurs aberrantes.

Perspectives

L'importance du travail que nous avons développé dans cette thèse donne la possibilité à nombreuse perspectives. Nous allons proposer dans ce qui suit quelques questions ouvertes à traiter dans le futur.

1. **Sur le choix du paramètre q de la régression Trimmed :** Le choix de ce paramètre est très important en pratique, car il permet d'améliorer la qualité de prévision. Dans nos simulations, nous avons travaillé par plusieurs valeurs de α et nous avons remarqué que les résultats ont été meilleur pour le choix que nous avons pris. Signalons que, à l'heure actuelle, il n'existe aucune méthode automatique pour sélectionner le paramètre q optimale. Dans un futur proche, nous essayerons de résoudre ce problème.
2. **Sur le choix du paramètre de lissage :** le choix de ce paramètre a une grande importance dans la prévision. Précisément, il permet d'améliorer la qualité de l'estimateur en améliorant sa vitesse de convergence. Il pourrait être possible d'adopter la généralisation de Heng-Yan Leung (2005) pour traiter ce problème.
3. **Sur le choix de la semi-métrique :** Vu l'importance de la semi-métrique en pratique et son rôle dans l'optimalité et la performance de l'estimateur, nous proposerons de généraliser les résultats en utilisant d'autres familles de semi-métriques.
4. **Généralisation au cadre spatiale et ergodique :** la généralisation de nos résultats au cas où les données sont spatiales ou ergodiques est une suite logique à suivre. Nous essayerons d'étendre les résultats de Gheriballah et al. (2010) et Gheriballah et al. (2013) sur la méthode à noyau pour nos estimateurs locaux linéaires.
5. **Cas de données incomplètes :** le travail sur des données incomplètes (données censurées et tronquées) lorsque la variable explicative est fonctionnelle n'est pas encore traité. Il est donc intéressant d'aborder ce cas dans les futures recherches.

D'autres perspectives pourraient être abordées à long terme telles que, les cas où les deux variables (réponse et explicative) sont fonctionnelles et la généralisation de nos résultats au cas où l'ordre de polynômes locaux est supérieur strictement à un.

Références

Attouch, M., Laksaci, A., Ould Said, E. (2009). Asymptotic distribution of robust estimator for functional nonparametric mo dels. *Comm. Statist. Theory and Methods,* **38**, 1317-1335.

Baìllo, A. and Grané, A. (2009). Local linear regression for functional predictor and scalar response. *J. Multivariate Anal.,* **100**, 102-111.

Barrientos-Marin, J., Ferraty, F.,Vieu, P. (2010). Locally modeled regression and functional data. *J. of Nonparametric Statistics,* **22**, 617-632

Berlinet, A., Elamine, A., and Mas, A. (2011). Local linear regression for functional data. *Ann. Inst. Statist. Math,* **63**, 1047-1075.

Cai, Z. Ould Said., E. (2003). Local M-estimator for nonparametric time series. *Statist. Probab. Lett.* **65**, 433-449.

Cuevas, A. (2014).A partial overview of the theory of satistics with functional data. *J.statist.Plann* **147**, 1 -23.

Cuevas, A. (2014).A partial overview of the theory of satistics with functional data. *J.statist.Plann* **147**, 1 -23.

Derrar,S., Laksaci,A., Ould Saïd, E. (2015). On the Nonparametric Estimation of the Functional ψ-Regression for a Random Left-Truncation Model. *Journal of Statistical Theory and Practice,* **9,** 823-849 .

Fan, J., Hu, T. C. and Truong, Y. K. (1994). Robust non-parametric function estimation. *Scand. J. Statist.* **21**, 433-446.

Fan, J. and Gijbels, I. (1996). Local polynomial modeling and its applications. *London, Chapman & Hall.*

Fan, J. (1993). Local Linear Regression Smoothers and Their Minimax Efficiencies. *The Annals of Statistics,* **21**, 196-216.

Ferraty, F. and Vieu, P. (2006). Nonparametric functional data analysis. Theory and Practice. *Springer Series in Statistics. New York.*

Gheriballah, A., Laksaci, A., Rouane, R. (2010). Robust nonparametric estimation for spatial regression. *Journal of Statistical Planning and Inference,* **140,** 1656-1670.

Gheriballah, A., Laksaci, A., Sekkal, S.(2013). Nonparametric M-regression for functional ergodic data. Statist. *Probab. Lett,* **83,** 902-908.

Hallin, M., Lu, Z. and Yu,K. (2004). Local linear spatial quantile regression. *Bernoulli,* **15,** 73-101.

Hsing, T. and Eubank, R. (2015). Theoretical foundations of functional data analysis, with an introduction to linear operators. *Wiley Series in Probability and Statistics. John Wiley & Sons, Chichester.*

Huber, P. J. (1964). Robust estimation of a location parameter. *Ann. Math. Statist.* **35**, 73- 101.

Liebscher, E.(2001). Central limit theorems for α-mixing triangular arrays with applications to nonparametric statistics.*Mathematical Methods of Statistics,* **10,** 194-214.

Rachdi, M., Laksaci, A., Demongeot, J., Abdali, A. et Madani, F.(2014). Theoretical and practical aspects of the quadratic error in the local linear estimation of the conditional density for functional data. *Comput. Statist. Data Anal,* **73,** 53-68.

Stone C J, (2005). Nonparametric M-regression with free knot splines. *J. Statist. Plan. and Inference,* **130**, 183-206.

Zhang, J. (2014). Analysis of variance for functional data. *Monographs on Statistics and Applied Probability, 127. CRC Press, Boca Raton, FL.*

Zhou, Z. and Lin, Z. (2016). Asymptotic normality of locally modeled regression estimator for functional data. *J. of Nonparametric Statistics,* **28**, 116-131.