Remerciments

Je tiens à exprimer toute ma reconnaissance à mon Directeur de thèse Monsieur Abdelkader Gheriballah. Je le remercie de m'avoir encadrée, orientée, et conseillée.

Je tiens à remercier sincèrement Monsieur le professeur Ali Laksaci, qui, s'est toujours montré à l'écoute et très disponible tout au long de la réalisation de ce mémoire, ainsi pour l'inspiration, l'aide et le temps qu'il a bien voulu me consacrer.

Mes remerciements vont aussi à Messieurs Fethallah Tebboune , Mohammed Kadi Attouch, Ali Laksaci et El Cheikh Mezouar qui me font l'honneur d'être parmi les membres du jury.

Je remercie particulierement mes parents, qui ont su croire en moi et qui m'ont apporté toute leur aide quand j'en ai eu besoin.

Je voudrais aussi remercier mon mari pour son soutien et ses encouragements.

Enfin, Mes remerciements les plus sincères à toutes les personnes qui ont contribué de près ou de loin à l'élaboration de ce mémoire.

Table des matières

1 Introduction **4**

1.1 Présentation 4

1.2 Contexte bibliographique 5

1.2.1 Données et variables fonctionnelles 5

1.2.2 Variables ergodiques 7

1.3 Brève présentations des résultats existants 8

1.3.1 Résultats existants Cas i.i.d 8

1.3.2 résultats existants : Cas α mélangeant 9

1.3.3 résultats existants : Cas spatial 11

1.4 Présentation de modèle 13

1.5 Contribution de la thèse 14

2 Nonparametric M-regression for functional ergodic data **22**

2.1 Introduction 23

2.2 The robust model and its estimate 25

2.3 Notations, hypotheses and comments 26

2.4 Results 28

2.5 Discussions 31

2.5.1 Some particulars cases 31

2.5.2 Some prospects 32
2.6 Appendix 34

3 Asymptotic normality **39**

3.1 Introduction 40
3.2 The ψ- regression model and its estimate 42
3.3 Notations, hypotheses and comments 43
3.4 Results 45
3.5 Some special cases 47
3.6 Some applications 48
3.6.1 Conditional Confidence curve 48
3.6.2 Smoothing parameter selection 49
3.6.3 The functional times series prediction 50
3.7 Application 51
3.8 Appendix 54

4 Conclusion **65**

Chapitre 1

Introduction

1.1 Présentation

Cette thèse est consacrée à l'étude de la régression robuste dans le cas où la variable explicative est fonctionnelle. Plus précisément, Nous nous proposons d'étudier le problème de la modélisation robuste non paramétrique lorsque les données statistiques sont à la fois fonctionnelles et ergodiques. Il s'agit d'un sujet très important motivé par l'importance de la régression robuste comme outil de prévision permet d'éliminer les observations aberrantes.

Dans un premier temps, nous considérons une suite d'observations dépendantes, et nous construisons un estimateur, par la méthode du noyau pour ce modèle de régression. Nous étudions ensuite sa convergence presque complète, sous des conditions standard.

Dans un second temps, nous établissons la normalité asymptotique de ce même estimateur et on va démontrer que cette propriété asymptotique est très utile dans des nombreuses analyses statistiques telle la détermination des intervalles de confiance, le choix du paramètre de lissage et la convergence des moments. Ce résultat est aussi obtenu sous des conditions standard.

1.2 Contexte bibliographique

1.2.1 Données et variables fonctionnelles

L'étude des données fonctionnelles est largement développée ces derniers temps. En effet, Les données fonctionnelles provenant de diverses branches des sciences (économétrie, biologie, environnement, ...). Ce développement est aussi justifié par le perfectionnement des outils de mesure et le progrès des outils informatiques, permettant la modernisation des méthodes d'observations. La statistique fonctionnelles est une nouvelle branche de la statistique à comme objectif la modernisation des modèles qui traite ce genre de données. D'où l'importance de cette thématique. Depuis la publication des monographies de Ramsay et Silvermann (2005) pour les modèles paramétriques et Ferraty et Vieu (2006) pour les modèles non paramétriques, un corpus travaux de recherche considérables ont été dédiés à cette thématique. Historiquement, les premiers résultats considérant des observations sous forme de trajectoires ont été obtenus par Obukhov (1960), Holmstrom (1963) en climatologie, Deville (1974) en démographie, Molenaar et Boomsma (1987) puis Kirkpatrick et Heckman (1989) en génétique.

En statistique, les modèle de régression (paramétrique ou non-paramétrique) ont une place très importante dans ces dernières années. Par ailleurs, dans le contexte paramétriques, la contribution de Ramsay et Silverman (1997, 2002) présente un ouvrage joue un rôle pivotal qui a remporté un grand succès sur la modélisation statistique pour données fonctionnelles, tandis que Bosq (2000) a traité des modèles fonctionnels linéaires tel le processus d'autorégression puis s'intéresse à la prédiction à partir des observations dépendantes (Bosq (2005)). Cardot et al. (1999) ont construit un estimateur pour le modèle de régression linéaire Hilbertien en utilisant l'analyse des composantes principales aux variables fonctionnelles. Cet estimateur est construit à l'aide des propriétés spectrales de la version empirique de l'opérateur de variance-covriance de la variable explicative fonctionnelle. Ils ont montré la convergence presque complète ainsi que la convergence en norme L^2 de cet estimateur. En utilisant une version régularisée, Cardot et al. (2004), ont introduit un estimateur pour les quantiles conditionnels vu comme forme linéaire continue définie sur un espace de Hilbert. Récemment Crambes et al. (2010) ont appliqué ce modèle à la prédiction dans le modèle linéaire fonctionnel lorsque la variable d'intérêt est fonctionnelle. Concernant les modèles du test en statistique fonctionnelle, nous citons à Cardot et al. (2003) et à Cuevas et al. (2004), Horváth et Reeder (2011) . D'autres auteurs se sont intéressés au cas où la variable réponse est qualitative. Nous renvoyons à Hastie et al. (1995), Hall et Heckman (2002) pour une étude bibliographique.

La régression non-paramétrique s'est développée considérablement ces dernières décennies, les premiers résultats ont été réalisés par Ferraty et Vieu (2000), et ils ont établi la convergence presque complète d'un estimateur à noyau de ce modèle non paramétrique dans le cas i.i.d. Le cas dépendent a été traité par Ferraty et al. (2002) . En considérant la propriété de concentration sur des petites boules pour la variable explicative., Dabo-Niang et Rhomari (2003) ont obtenu la convergence en norme L^p de l'estimateur de la régression. Masry (2005) a étudié la normalité asymptotique de l'estimateur de la régression dans le cas où les observations sont α-mélengeantes. En 2006 Ferraty et al. ont montré la convergence presque complète de l'estimateur à noyau de quelque modèle conditionnel tel que la fonction de répartition conditionnelle, la densité conditionnelle, les quantiles conditionnels ainsi que pour le mode conditionnel. Dans le même contexte, Dabo-Niang et al. (2004) ont établi la convergence en norme L_p de l'estimateur à noyau du mode conditionnel à variable explicative à valeur dans un espace vectoriel semi-normé, de dimension infinie. Récemment, J. Demongeot et al. (2013) ont introduit une nouvelle méthode d'estimation locale linéaire de la densité conditionnelle des données fonctionnelles dans un espace semi- métrique, ils ont obtenu, sous des conditions standards, les convergences ponctuelle et uniforme presque complète ainsi que les vitesses de convergence. L'estimation des quantiles conditionnels comme inverse de la fonction de répartition conditionnelle a également été largement étudiée dans des différents types de corrélation par Ferraty, Rabhi et al. (2005), Ferraty et al. (2006), Ferraty et Vieu (2006a) et Ezzahrioui (2007). Laksaci en collaboration avec Maref (2009) ont prouvé la version uniforme de la convergence presque complète de ce dernier estimateur. Ces auteurs ont abordé cet estimateur dans le cas où les observations sont spatialement dépendantes. En ce qui concerne la convergence en moyenne quadratique On trouve aussi La convergence en moyenne quadratique, les premiers résultats ont été établis par Ferraty et al. (2007). Dans cet article, les auteurs ont explicité le terme asymptotique exacte de l'erreur quadratique. On trouvera des applications de ce résultat sur le choix du paramètre de lissage dans l'article de Rachdi et Vieu (2007) pour le cas global et Benhenni et al. (2007) le choix local. Dans un article récent, une méthode alternative a été développée par Burba et al. (2008). Cette méthode est une généralisation de l'approche des k plus proches voisins dans le cadre de l'estimation à noyau de l'opérateur de régression, et ils ont obtenu la convergence presque complète de l'estimateur construit.

Dans le même contexte, Attouch et Benchikh (2012), sous l'hypothèse de concentration sur les petites boules de la mesure de probabilité, de la variable explicative fonctionnelle. Ils ont établit la normalité asymptotique de l'estimateur non paramétrique robuste de la fonction de régression avec la méthode des k plus proches voisins lorsque les variables sont fonctionnelle. D'autres méthodes d'estimation ont été proposées, la propriété de robuste.

La méthode robuste utilisée dans ce travail appartient à la classe de M-estimations introduites par Huber (1964) . La littérature sur cette méthode d'estimation est très importante dans le cas où les observations sont de dimension fini (voir par exemple Robinson (1984) , Collomb et Härdle (1986) et Boente et Fraiman (1989, 1990)). Pour le cas fonctionnel, Cadre (2001) a étudié l'estimation de L_1-médiane d'une variable aléatoire dans un espace de Banach . Ils ont utilisé cette approche robuste pour étudier le modèle de régression linéaire sur les quantiles avec les valeurs des variables explicatives prise dans un espace de Hilbert. Dans le cas non paramétrique Azzedine et al. (2008) ont considéré le problème de l'estimation de la régression non paramétrique robuste pour une variable explicative fonctionnelle. Attouch et al. (2009) ont étudié la convergence presque complète d'une famille d'estimateurs robustes basée sur la méthode du noyau, lorsque des observations sont indépendantes. La vitesse de convergence en norme L^p fait l'objet d'un travail de Crambes et al. (2008) en considérant les deux types d'observations indépendantes et α-mélangeantes. En considérant des indépendantes identiquement distribuées Attouch et Benchikh (2012), ont établit la normalité asymptotique de l'estimateur non paramétrique robuste de la fonction de régression avec la méthode des k plus proches voisins lorsque les variables sont fonctionnelle. Nous renvoyons, Attouch et al. (2011) pour le cas des données spatiales fonctionnelles.

1.2.2 Variables ergodiques

La théorie ergodique est une hypothèse originaire de la mécanique. Plus précisément, le mot Ergodique est un mot grec signifie (travail, énergie). Elle a été proposée par Boltzmann en 1885, afin de modéliser la théorie cinétique des gaz. Depuis cette contribution, les chercheurs en mécanique ont montré que cette hypothèse est importante pour modéliser plusieurs phénomènes. Ce qui a motivé les mathématiciens Neumann et Birkhoff en 1931 pour chercher la formulation mathématique de l'ergodicité. Et depuis, la théorie ergodique occupe une place dans des différents branches de la mathématique tel l' analyse fonctionnelle et théorie des groupes ; calcul des probabilités et plus précisément processus markoviens ; théorie de l'information, etc.... Nous nous référons au livre de Krengel (1985) pour une panoplie des résultats de la théorie ergodique. En ce qui concerne le cadre générale de la thèse qui la statistique non paramétrique, La littérature sur le cas ergodique est très restreinte. Nous citons pour ce cas, l'article de Laïb et Ould-Saïd (2000). Ces derniers ont étudie l'estimateur de Collomb et Härdle (1986) pour le modèle d'auto-régression d'un processus stationnaire ergodique. Ils ont obtenu la convergence uniforme de cet estimateur même lorsque la fonction objective est non bornée. Ould-said en 1997 a abordé l'estimateur à noyau du mode conditionnelle. Récemment, Laïb et Louani (2010) ont étudier les propriétés asymptotiques de l'estimation de la régression du noyau quand des données fonctionnels sont ergodiques stationnaires, et ils ont établi la convergence uniforme presque complète ainsi que la normalité asymptotique de l'estimateur à noyau de la fonction de régression.

1.3 Brève présentations des résultats existants

Afin de montrer l'importance de notre contribution comparativement aux résultats existants. Nous rappelons dans cette section les différents cas traités dans la statistique fonctionnelle a savoir le cas i.i.d, la cas fortement mélangeant et le cas spatial.

1.3.1 Résultats existants Cas i.i.d

Sous les conditions suivantes

(H1) $\mathbb{P}(X \in B(x,h)) = \phi_x(h) > 0 \ \forall h > 0$ et $\lim_{h \to 0} \phi_x(h) = 0$.

(H2) Il existe $C_1 > 0$ et $b > 0$ tel que $\forall x_1, x_2 \in \mathcal{N}_x$, $\forall t \in \mathbb{R}$
$|\Psi(t,x_1) - \Psi(t,x_2)| \leq C_1 d^b(x_1,x_2)$.

(H3) La fonction ψ_x est strictement monotone, bornée, continûment dérivable, et sa dérivée est telle que $|\psi'_x(t)| > C_2 > 0, \forall t \in \mathbb{R}$.

(H4) K est une fonction continue à support [0, 1] telle que $0 < C_3 < K(t) < C_4 < \infty$.

(H5) $\lim_{n\to\infty} h_k = 0$ et $\lim_{n\to\infty} \log n/n\phi_x(h_k) = 0$.

On obtient le théorème

Théorème 1 *Supposons que (H1)-(H5) sont satisfaites ; puis $\widehat{\theta}_x$ existe et est unique p.co. pour tout n suffisamment grand, et nous avons*

$$\widehat{\theta_x} - \theta_x = O(h_k^b) + O\left(\sqrt{\frac{logn}{n\phi_x(h_k)}}\right) \ p.co.$$

Ce résultat a été obtenu par Azzedine et al. (2008). Tandis que la normalité asymptotique a été étudié par Attouch et al. (2009). Elle est donné dans le théorème suivant

Théorème 2 *Sous des hypothèses de concentration de la mesure de probabilité de la variable fonctionnelle et des conditions techniques standards le noyau et la fenêtre, on a pour tout $x \in A$,*

$$\left(\frac{n\phi(h)}{\sigma^2(x,\theta_x)}\right)^{1/2}\left(\widehat{\theta_x} - \theta_x - B_n(x)\right) \to^L N(0,1) \quad lorsque\ n \to \infty$$

où
$B_n(x) = \dfrac{h}{\phi(h)\alpha_1\Gamma_1(x,\theta_x)}\displaystyle\int_0^1 K(t)\varphi_x(th)\phi'(th)dt + 0(1)$ *(avec $\varphi_x(s) = \mathbb{E}[\psi_x(Y,\theta_x) \mid d(X,x) = s]$),*

$$\sigma^2(x,\theta_x) = \frac{\alpha_2\lambda_2(x,\theta_x)}{\alpha_1^2 g(x)(\Gamma_1(x,\theta_x))^2} \text{ (avec } \alpha_j = -\int_0^1 (K^j)'(s)\beta(s)ds, \text{ pour, } j=1,2),$$

La démonstration et les détails des conditions imposées pour aboutir à ces résultats sont donnés dans l'article de Attouch et al.(2009).

1.3.2 résultats existants : Cas α mélangeant

Nous commençons par rappeler la définition de la propriété forte de mélange. Pour cela, nous introduisons les notations suivantes. Soit $\mathcal{F}_i^k(Z)$ désigner le σ- algèbre engendrée par $\{Z_j, i \leq j \leq k\}$.

Définition

Soit $\{Z_i, i = 1, 2, ...\}$ une suite de va. Étant donné un entier positif n, posons

$$\alpha(n) = \sup\{\mid \mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B) \mid : \text{A} \in \mathcal{F}_1^k(Z) \text{ et B} \in \mathcal{F}_{k+n}^\infty(Z), k \in \mathbb{N}^*\}.$$

La suite $\{Z_i, i = 1, 2, ...\}$ est dit α-mélange (fort mélange) si le coefficient de mélange $\alpha(n) \to 0$ quand $n \to \infty$.

Il existe de nombreux processus répondant à e propriété de mélange forte. Nous citons, ici, les processus ARMA habituels qui sont géométriquement fortement mélange, c'est à dire, il existe $\rho \in (0,1)$ et $a > 0$ de telle sorte que, pour tout $n \geq 1$, $\alpha(n) \leq a\rho^n$ (voir, par exemple, Jones (1978)).
Dans le reste de cette section, nous supposons que $(X_n, Y_n)_{n\geq 1}$ est fortement mélange. Sous cette corrélation Attouch et al (2013) ont obtenu le résultat suivant

Théorème 3 *Sous des hypothèses peu restrictives sur le coefficient du mélange et la mesure de concentration de la variable fonctionnelle, on a*

$$\widehat{\theta}_x - \theta_x = O(h^{b_1}) + O\left(\sqrt{\frac{\log n}{n\phi_x(h)}}\right) \quad p.co. \quad lorsque\ n \to \infty$$

où $\phi_x(h_K)$ est la concentration de la mesure de probabilité de la variable fonctionnelle X dans la boule de centre x et de rayon h_K.

Afin de démonter la normalité asymptotique de ce modèle dans le cas de mélange fort, Attouch et al. (2010) ont consideré les conditions suivante

(H1) $\mathbb{P}(X \in B(x,h)) =: \phi_x(h) > 0$,

(H2) ψ_x est monotone, bornée, fonction différentiable continue, w.r.t le second composant, et son dérivé est $\dfrac{\partial \psi_x(y,t)}{\partial t}$ est bornée et continue à θ_x uniformément en y.

(H3) La fonction $\lambda_\gamma(.,.)$ satisfait à la condition de Hölder par rapport à la première variable, qui est la suivante : il existe une constante positive b_γ telle que :
$\forall (u_1,u_2) \in N_x \times N_x, \forall t \in \mathbb{R}, \mid \lambda_\gamma(u_1,t) - \lambda_\gamma(u_2,t) \mid \leq C_1 d^{b_\gamma}(u_1,u_2)$.

(H4) La fonction $\Gamma_\gamma(.,.)$ satisfait à la condition de Hölder par rapport à la première variable, qui est la suivante : il existe une constante d_γ strictement positif tel que :
$\forall (u_1,u_2) \in N_x \times N_x, \forall t \in \mathbb{R}, \mid \Gamma_\gamma(u_1,t) - \Gamma_\gamma(u_2,t) \mid \leq C_2 d^{d_\gamma}(u_1,u_2)$.

(H5) La bande passante h satisfait :
$$h \downarrow 0\ ,\ \ n\phi_x(h) \to \infty \text{ quand } n \to \infty,$$
et il existe une fonction $\beta_x(.)$ de telle sorte que
$$\forall t \in [0,1] \ \ \lim_{h\to 0} \frac{\phi_x(th)}{\phi_x(h)} = \beta_x(t).$$

(H6) Le noyau K est une fonction différentiable positif supporté sur [0,1]. K′ son dérivé existe et satisfait $-\infty < C_3 < K'(t) < C_4 < 0$ pour $0 \leq t \leq 1$ où C_3 et C_4 sont des constantes.

(H7) $(X_i, Y_i)_{i\in\mathbb{N}}$ est une séquence de α-mélange dont les coefficients satisfaire

$$\exists a > 0,\ \ \exists C > 0:\ \forall n \in \mathbb{N}\ \ \alpha(n) \leq Cn^{-a}.$$

(H8) $0 < \sup_{i\neq j} \mathbb{P}((X_i, X_j) \in B(x,h) \times B(x,h)) = O\left(\frac{(\phi_x(h))^{(a+1)/a}}{n^{1/a}}\right)$.

(H9) Il existe $\eta > 0$ tel que, $Cn^{1-a+\eta} \leq \phi_x(h) \leq C'n^{\frac{1}{1-a}+\eta}$, avec a > 2.

Théorème 4 *Supposons que (H1) - (H9) vrai, puis $\widehat{\theta}_x$ existe et est unique avec une grande probabilité et pour tout $x \in A$, nous avons*

$$\left(\frac{n\phi_x(h)}{\sigma^2(x,\theta_x)}\right)^{1/2} (\widehat{\theta}_x - \theta_x - B_n(x)) \to^D N(0,1) \ \ \textit{lorsque } n \to \infty$$

ou

$B_n(x) = \dfrac{h}{\phi_x(h)\alpha_1\Gamma_1(x,\theta_x)} \displaystyle\int_0^1 K(t)\varphi_x(th)\phi_x'(th)dt + 0(1)$ *(avec* $\varphi_x(s) = \mathbb{E}[\psi_x(Y,\theta_x) \mid d(X,x) = s]$,

$$\sigma^2(x,\theta_x) = \frac{\alpha_2\lambda_2(x,\theta_x)}{\alpha_1^2(\Gamma_1(x,\theta_x))^2} \ \textit{(avec } \alpha_j = -\int_0^1 (K^j)'(s)\beta(s)ds, \textit{ for, j=1,2),}$$

$$A = \{x \in \mathcal{F}, \lambda_2(x,\theta_x)\Gamma_1(x,\theta_x) \neq 0\}$$

1.3.3 résultats existants : Cas spatial

Dans ce cas, on considére $Z_i = (X_i, Y_i)$, $i \in \mathbb{N}^N$ être un processus spatial strictement stationnaire , définie sur un espace de probabilité $(\Omega, A, \mathbb{P})$. Nous supposons que ce processus est observée sur un domaine rectangulaire $I_n = \{\mathbf{i} = (i_1, ...i_n) \in \mathbb{N}^N, 1 \leq i_k \leq n_k, k = 1, ...N\}$, $\mathbf{n} = (n_1, ..., n_N) \in \mathbb{N}^N$. Un point $\mathbf{i}$ sera considéré comme un site. Nous écrirons $\mathbf{n} \to \infty$ si $\min\{n_k\} \to \infty$ et $\mid n_j/n_k \mid < C$ pour une constante C telle que $0 < C < \infty$ pour tout j, k tel que $1 \leq j\,;\ k \leq N$. Pour $\mathbf{n} = (n_1, ..., n_N) \in \mathbb{N}^N$, nous avons mis en $\hat{\mathbf{n}} = n_1 \times ... \times n_N$. Par ailleurs, on suppose que le champ aléatoire $(Z_{\mathbf{i}}, \mathbf{i} \in \mathbb{N}^N)$ satisfait à la condition de mélange suivant :

$$\left\{\begin{array}{l} Il \text{ existe une fonction } \varphi(t) \downarrow 0 \text{ quand } t \to \infty, \text{ tels que} \\ \forall E, E' \text{ sous} - \text{ensembles de } \mathbb{N}^N \text{ avec fini cardinaux} \\ \alpha(\beta(E), \beta(E')) = \displaystyle\sup_{B\in\beta(E), C\in\beta(E')} |\mathbb{P}(B \cap C) - \mathbb{P}(B)\mathbb{P}(C)| \\ \qquad \leq s(card(E), card(E'))\varphi(dist(E, E'))\ , \end{array}\right.$$

où B(E) (resp. B(E')) désigne la σ-champ Borel généré par $(Z_{\mathbf{i}}, \mathbf{i} \in E)(resp.(Z_{\mathbf{i}}, \mathbf{i} \in E'))$. Card(E)(resp. Card(E')) la cardinalité de E (resp. E'), dist(E,E') la distance euclidienne entre E et E' et s : $\mathbb{N}^2 \to \mathbb{R}^+$ est fonction positive asymétrique non décroissante dans chaque variable de telle sorte que soit
$s(n,m) \leq C \min(n,m), \ \forall n, m \in \mathbb{N}$
ou
$s(n,m) \leq C(n+m+1)^{\tilde{\beta}}$
pour certains $\tilde{\beta} \geq 1$ et certains $C > 0$.

Notons que ces conditions ont été utilisées par Tran (1990), Carbon et al.(1996) et elle sont vérifiées par de nombreux modèles spatiaux (voir Guyon (1987)).

Par la suite, on note par N_x un voisinage de x, on pose et on considère les hypothèses suivantes :

(H1) $\mathbb{P}(X \in B(x,h)) = \phi_x(h) > 0$.

(H2) Il existe $C_1 > 0$, $\delta_0 > 0$ et $b_1 > 0$, tels que,

$$\forall x_1, x_2 \in N_x \quad \forall t \in [\theta(x)-\delta_0, \theta(x)+\delta_0] \ \mid \Psi(x_1,t) - \Psi(x_2,t) \mid \leq C_1 d^{b_1}(x_1,x_2).$$

(H2′) La fonction $\Gamma(\cdot,\cdot) := \mathbb{E}\left[\psi'(Y-\cdot)|X=\cdot\right]$ vérifie :

$$\exists C_2 > 0, \ b_2 > 0, \ \text{ tels que}, \forall x_1, x_2 \in N_x \quad \forall t \in [\theta(x)-\delta_0, \theta(x)+\delta_0] : \ |\Gamma(x_1,t) - \Gamma(x_2,t)| \leq$$

(H3) La fonction $\psi(\cdot)$ est bornée, monotone, differentiable de dérivée bornée et uniformément continue.

(H4) La suite des observations $(X_i, Y_i)_{i \in \mathbb{N}}$ est α-mélangeante dont le coefficient de mélange satisfait :

$$\exists\, a > 0, \ \exists\, C_2 > 0 : \ \forall n \in \mathbb{N} \ \alpha_n \leq C_2 n^{-a}.$$

(H5) $0 < \sup_{i \neq j} \mathbb{P}\left((X_i, X_j) \in B(x,h) \times B(x,h)\right) = O\left(\frac{(\phi_x(h))^{(a+1)/a}}{n^{1/a}}\right)$.

(H6) K est un noyau continu et à support compact $[0,1]$ tel que $0 < C_3 < K(t) < C_4 < \infty$.

(H7) $\lim_{n\to\infty} h = 0$ et il existe $\eta > 0$, tel que, $C_5 n^{\frac{3-a}{a+1}+\eta} \leq \phi_x(h) \leq C_6 n^{\frac{1}{1-a}}$, avec $a > \frac{5+\sqrt{17}}{2}$.

Remarque 1 : Nos hypothèses sont relativement standard, les conditions (H1)-(H2), (H4)-(H7)ont été utilisées pour la régression classique (voir Ferraty et

Vieu (2006)), tandis que la condition (H3) sur la fonction objective ψ est la même qu'en dimension finie (voir Collomb et Hardle (1986)). Notons que l'hypothèse ψ bornée est non seulement pour simplifier la démonstration du résultat, mais c'est une condition structurelle pour conserver la robustesse du modèle en limitant l'influence des observations aberrantes.

On a le résultat suivant :

Theorem 1 *Sous les hypothèses (H1)-(H7) et si $\Gamma(x, \theta_x) \neq 0$, alors l'estimateur $\widehat{\theta}(x)$ existe et unique, pour n assez grand, et il vérifie,*

$$\widehat{\theta}(x) - \theta(x) = O\left(h^{b_1}\right) + O\left(\sqrt{\frac{\log n}{n\,\phi_x(h)}}\right) \qquad p.co. \tag{1.1}$$

1.4 Présentation de modèle

Soit $(X_i, Y_i)_{i=1,\ldots,n}$ une suite de processus ergodique stationnaire à valeurs dans $\mathcal{F} \times \mathbb{R}$, ou $\mathcal{F}$ est un semi-métrique. On note d(.,.) la semi-métrique sur $\mathcal{F}$. Pour tout x dans $\mathcal{F}$, on considère une fonction réelle, mesurable, notée ψ_x satisfaisant certaines conditions de régularité énoncés ci-dessous et nous modélisons la co-variation de X_i et Y_i via la régression non paramétrique robuste, désigné par θ_x , implicitement défini comme nul par rapport à t de l'équation suivante :

$$\Psi(x, t) = \mathbb{E}[\psi_x(Y_i, t)|X_i = x] = 0. \tag{1.2}$$

Nous supposons que, pour tout x $\in \mathcal{F}$, θ_x existe et est unique, est un zéro par rapport à t de (1.2) (voir, par exemple, Boente et Fraiman (1989) ou Koul et Stute (1998) pour l'existence et l'unicité de θ_x).
Pour tout (x,t) $\in \mathcal{F} \times \mathbb{R}$, nous proposons un estimateur non paramétrique de $\Psi(x, t)$ définie par

$$\widehat{\Psi}(x, t) := \frac{\sum_{i=1}^{n} K(h^{-1}d(x, X_i))\psi_x(Y_i, t)}{\sum_{i=1}^{n} K(h^{-1}d(x, X_i))}, \ \forall \text{ t} \in \mathbb{R}$$

o*ù* K est un noyau et h = h_n (est une suite de nombres réels positifs qui tend vers zéro quand n tend vers l'infini). Un estimateur naturel du paramètre θ_x, noté $\widehat{\theta_x}$, est un zéro par rapport à t de

$$\widehat{\Psi}(x,t) = 0.$$

De toute évidence, lorsque $\psi_x(Y,t)$ = Y - t, alors $\hat{\theta_n}$ est l'estimateur donné dans Ferraty et Vieu (2006) pour la régression non paramétrique fonctionnel. Alors que pour $\psi_x(Y,t) = 1_{y>t} - (1-\alpha)$, On obtient l'estimation du quantile conditionnel étudié par Laksaci et al. (2009).

Nous rappelons que La régression robuste utilisé ici appartient à la classe de M-estimation introduites par Huber (1964) et qui est un thème ayant eu un grand intéret en statistique non-paramétrique.

1.5 Contribution de la thèse

La régression non paramétrique robuste de la statistique fonctionnelle a été récemment introduit dans statistique fonctionnelle par Azzeddine et al. (2008), ces auteurs ont montré la convergence presque complète d'un estimateur de noyau dans le cas ou les observations sont i.i.d. La convergence en norme L^p pour ce modéle a été obtenue par Crambes et al. (2008) en considérant les deux cas d'observations cas indépendantes et α mélangeantes. Attouch et al. (2009) ont établi la normalité assymptotiques lorsque les observations sont i.i.d. La généralisation de ce résultat le cas de mélange a été élaborer par Attouch et al. (2010). Dans ce travaille on s'interèsse a la convergence presque complète de ce modèle dans le cas ou les observations sont ergodique.

Cette thèse est organisé comme suit, Le premier chapitre est consacré à une brève présentation de la thèse. Ainsi, nous avons présenté, en détails les résultats obtenus en statistique non-paramétrique quand les données sont fonctionnelles. Dans ce contexte, nous traitons la convergence et la normalité asymptotique dans le cas i.i.d. Ensuite, de la convergence ponctuelle et uniforme presque complète de l'estimateur non-paramétrique de la fonction de régression robuste au cas alpha-mélangeant, et aussi la normalité asymptotique, aprés on s'intéresse a la cohérence presque complète et la normalité asymptotique de l'estimateur non paramétrique robuste pour la régression spatiale.

Le second chapitre, on a étudié la régression robuste dans le cas ou la variable explicative est fonctionnelle et les observations sont ergodique. L'objectif principal de cet travail est de prouver la convergence presque complète (avec taux) pour l'estimateur proposé . Ce résultat est obtenu sous une hypothèse processus stationnaire ergodique , sans l'aide de conditions de mélange traditionnels. Ce travail a fait l'objet d'un article soumis.

Enfin dans le troisiéme chapitre traitons le problème de la normalité asymptotique d'une famille d'estimateurs robustes basée sur la méthode du noyau quand les données fonctionnels sont ergodiques stationnaires.

Bibliographie

M. Attouch and T. Benchikh. (2012). Asymptotic distribution of robust k- nearest neighbour estimation for functional nonparametric models.

M. Attouch, A. Gheriballah, A. Laksaci. (2011). Robust Nonparametric Estimation for Functional Spatial Regression. Springer. pp.27-31.

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

N. Azzeddine, A. Laksaci, E. Ould-Saïd. (2008). On the robust nonparametric regression estimation for functional regressor. Statist. & Probab. Lett., in press.

K. Benhenni, F. Ferraty, M. Rachdi, P. Vieu.. (2007). Local smoothing regression with functional data. Computational Statist., 22, 53-71.

G. Boente, and R. Fraiman (1989). Nonparametric regression estimation. J. Multivariate Anal. 29, 180-198.

T. Bollerslev, (1986). General autoregressive conditional heteroskedasticity. J. Econom., 31, 307- 327.

D. Bosq, (1996). Limit theorems for Banach-valued autoregressive processes. Applications to real continuous time processes. Bull. Belg. Math. Soc. 3, 537-555.

D. Bosq. (2000). Linear processes in function spaces. Theory and applications. Lecture Notes in Statistics, 149. Springer-Verlag.

F. Burba, F. Ferraty, P. Vieu. (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.

B. Cadre. (2001). Convergent estimators for the L_1-median of a Banach valued random variable. Statistics 35 (4), 509-521.

M. Carbon, M. Hallin, L.T. Tran,(1996). Kernel density estimation for random fields : the L_1 theory. Nonparametric Statist.6, 157-170.

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

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

H. Cardot, F. Ferraty, P. Sarda. (2000). Etude asymptotique d'un estimateur spline hybride pour le modèle linéaire fonctionnel. C. R. Math. Acad. Sci., Paris, 330, 501-504.

H. Cardot, F. Ferraty, A. Mas, P. Sarda, Testing hypotheses in the functional linear model. Scand. J. Statist. (2003), 30, No.1, 241-255.

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

C. Crambes, L. Delsol, A. Laksaci. (2008). Robust nonparametric estimation for functional data. J. Nonpparametric Statist., 20, No.7, 573-598.

A. Cuevas, M. Febrero, R. Fraiman, An anova test for functional data. Comput. Statist. Data Anal. (2004), 47, no.1, 111-122.

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

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

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

R.F. Engle, (1982). Autoregressive conditional heteroskedasticity with estimates of the variance of U.K. inflation. Econometrica, 50, 987-1007.

M. Ezzahrioui. (2007) Prévision dans les modèles conditionnels en dimension infinie. (French) PhD Thesis.

F. Ferraty, A. Laksaci, and P. Vieu. (2006). Estimating some characteristics of the conditional distribution in nonparametric functional models. Stat. Inference Stoch. Process 9 (1) 47-76.

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

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

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

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

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

F. Ferraty, A. Rabhi, et P. Vieu. (2005). Conditional quantiles for dependent functional data with application to the climatic El Niño phenomenon. Sankhyā 67 (2) 378-398.

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

A. Gheriballah, A. Laksaci, and R. Rouane. (2010). Robuste non-paeametric estimation for spatial regression. J. Stat. Plann. and Inf, 140, 1656-1670.

X. Guyon, (1987). Estimation d'un champ par pseudo-vraisemblance conditionnelle : Etude asymptotique et application au cas Markovien. In :Proceedings of the Sixth Franco-Belgian Meeting of Statisticia.

P. Hall, N.E. Heckman. (2002). Estimating and depicting the structure of a distribution of random functions. Biometrika. 89, No.1, 145-158.

T. Hastie, A. Buja, R. Tibshirani. (1995). Penalized discriminant analysis. Ann. Statist. 23, No.1, 73-102.

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

L. Horváth, and R. Reeder. (2011). A test of significance in functional quadratic regression. arXiv :1105.0014v1 [math.ST], forthcoming in Bernoulli Journal

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

D.A. Jones, (1978). Nonlinear autoregressive processes. Proc. Roy. Soc. London A, 360, 71-95.

M. Kirkpatrick, N. Heckman (1989). A quantitative genetic model for growth, shape, reaction norms, and other infinite-dimensional characters. J. Math. Biol., 27, No.4, 429-450.

H.L. Koul, W. Stute, Regression model fitting with long memory errors, J. Statist. Plann. Inference 71 (1998), 35-56.

U. Krengel, Ergodic Theorems, Walter de Gruyter & Co., Berlin, 1985.

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

N. Laïb, D. Louani, Strong consistency of the regression function estimator for functional stationary ergodic data, J. Statist. Plann. Inference (2010), in press.

N. Laïb, and D. Louani. (2011). Strong consistency of the regression function estimator for functional stationary ergodic data. J. Statist. Plan. and Inference, 141, 359-372.

A. Laksaci, M. Lemdani, and E. Ould-Saïd. (2009). L^1-norm kernel estimator of conditional quantile for functional regressors : consistency and asymptotic normality. Statist. Probab. Lett. 79, 1065-1073.

A. Laksaci et F. Maref. (2009). Estimation non paramétrique de quantiles conditionnels pour des variables fonctionnelles spatialement dépendantes. C. R. Acad. Sci. Paris, Sér. I,347,1075,1080.

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

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

T. Ozaki, (1979). Nonlinear time series models for nonlinear random vibrations. Technical report. Univ. of Manchester.

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

R. Robinson, 1984. Robust Nonparametric Autoregression. In : Lecture Notes in Statistics, vol. 26. Springer-Verlag, New York, pp. 247-255.

J.O. Ramsay, B.W Silverman. (1997). Functional Data Analysis. Springer, NewYork.

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

J.O. Ramsay, B.W. Silverman. (2005). Functional Data Analysis. Springer, New-York, 2nd Edition.

Chapitre 2

Nonparametric M-regression for functional ergodic data

Ce chapitre fait l'objet d'une publication au journal Statistics & Probability Letters.

Nonparametric M-regression for functional ergodic data

Abdelkader Ghriballah, Ali Laksaci [1] and Soumeya Sekkal

e-mail : (gheribaek, alilak or s_sekkal)@yahoo.fr

Abstract
Let $(X_i, Y_i)_{i=1,\dots,n}$ be a sequence of stationary ergodic processes valued in $\mathcal{F} \times \mathbb{R}$, where $\mathcal{F}$ is a semi-metric space. We consider the problem of estimating the regression function of Y_i given X_i by the robust M-estimation method. The principal aim of the paper is to prove the almost complete convergence (with rate) for the proposed estimator. This result is obtained under a stationary ergodic process assumption, without using traditional mixing conditions.

2.1 Introduction

It is well known that the ergodicity is a fundamental hypothesis in statistical physics, thermodynamic or in signal processing. In all these areas, this condition is to be studied in continuous path. Thus, it is really necessary to develop statistical tools allowing to treat the continuous ergodic processes in its own dimension by exploring its functional character. This is the general framework of the present work. More precisely, the main purpose of this paper is the robust nonparametric modilization of the functional ergodic data.

Noting that, functional data analysis (FDA) has become a major topic of research in the last decade, as evidenced the several special issues dedicated to this topic by various statistical journals (see, for instance Davidian et al. (2004), Gonzalez-Manteiga and Vieu (2007), Valderrama (2007) or Ferraty (2010)). The nonparametric treatment of such data has also, been widely developed in the last years (see Ferraty and Vieu (2006), Ferraty and Romain

1. Corresponding author. Email : alilak@yahoo.fr

(2011) for recent advances and references). The first result on the robust estimate of the functional nonparametric models was given by Azzedine *et al.* (2008). They established the almost complete convergence of robust estimators of the regression function when the regressors is functional and the observations are i.i.d. The robust analysis in functional time series, has been investigated by many authors (see, for examples, Crambes *et al.* (2008), Chen and Zhang (2009) and the references therein).

Recently, Laib and Louani (2011) consider the problem of functional estimation for nonparametric regression operators under less restrictive dependence structure that is the ergodicity condition. They studied the almost complete convergence of an estimator of the classical regression function constructed by the kernel method. Considering the same functional time series context, we establish, in this paper, the same asymptotic propriety for an alternative estimator of the functional nonparametric regression. This estimator is constructed by combining the ideas of robustness with those of smoothed regression which allows us to obtain reliable estimations when outlier observations are present. Thus, the originality of our result to the weak dependence structure imposed on the observation process by simply assuming its ergodicity and, also, to the robust estimation method used which is insensitive to the presence of outliers or heteroskedatic variables. Finally, let us mention that the ergodicity assumption has great consideration in practice. Indeed, it is one of a principal postulate of statistical physics in order to control the thermodynamic properties of gases, atoms, electrons or plasmas. This hypothesis is also used in signal processing, to study the evolution of a random signal. Moreover, as discussed by Laib and Louani (2011) the random variables generated with the functional autoregressive models is a particular example of the functional ergodic random variables. Such models are widely considered in functional data analysis to carry out some concrete problem (see, Bosq (1996) for some example and references).

The setup of this paper is as follows : We present our model in Section 2. Section 3 is dedicated to fixing notations and hypotheses. Our main result is given in Section 4. In Section 5 we show the generality of our model and the flexibility of our conditions by studying some particular case. The proofs of the auxiliary results are relegated to the Appendix.

2.2 The robust model and its estimate

Consider $Z_i = (X_i, Y_i)_{i=1,...n}$ be a $\mathcal{F} \times \mathbb{R}$-valued measurable strictly stationary process, defined on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$, where $\mathcal{F}$ is a semi-metric space, d denoting the semi-metric. For $x \in \mathcal{F}$, we consider a real measurable function ψ_x and we model the co-variation between X_i and Y_i via the nonparametric robust regression, denoted by θ_x, implicitly defined as a zero with respect to (w.r.t.) t of the equation

$$\Psi(x,t) := \mathbb{E}\left[\psi_x(Y_i,t) \mid X_i = x\right] = 0, \tag{2.1}$$

where ψ_x is a real-valued Borel function satisfying some regularity conditions to be stated below. In the following, we assume that equation (3.1) allows θ_x as unique solution (see for instance Boente and Fraiman (1989) for sufficient conditions for existence and uniqueness of θ_x). We point out that this robustification method belongs to the class of M-estimates introduced by Huber (1964) and it covers and includes many important nonparametric models, for example, $\psi_x(y,t) = (y-t)$ yields the classical regression, $\psi_x(y,t) = 1_{y\geq t} - 1_{y<t}$ leads to the conditional median function $m(x) = med(Y|X = x)$ and the α^{th} conditional quantile is obtained by setting $\psi_x(y,t) = 1_{y>t} - (1-\alpha)$, $\alpha \in (0,1)$. In addition, our robustification method allows us to consider the functional nonparametric regression model with a scale of the error assumed to be known by taking $\psi_x(.,.) = \psi(. - ./\sigma(x))$, where $\sigma(.)$ is a measure of spread for the conditional distribution of Y given $X = x$. We return to Stone (2005) for other examples of the function ψ.

For all $(x,t) \in \mathcal{F} \times \mathbb{R}$, we propose a nonparametric estimator of $\Psi(x,t)$ given by

$$\widehat{\Psi}(x,t) := \frac{\sum_{i=1}^{n} K(h^{-1}d(x,X_i))\psi_x(Y_i,t)}{\sum_{i=1}^{n} K(h^{-1}d(x,X_i))},$$

where K is a kernel and $h = h_{\mathbf{n}}$ is a sequence of positive real numbers. A natural estimator $\widehat{\theta_x}$ of θ_x is a zero w.r.t. t of the equation

$$\widehat{\Psi}(x,t) = 0.$$

Obviously, when $\psi_x(Y,t) = Y - t$, then $\widehat{\theta}_n$ is the estimator given in Ferraty and Vieu (2006) for the functional nonparametric regression. While for $\psi_x(y,t) = 1_{y>t} - (1-\alpha)$, we obtain the α^{th} conditional quantile estimate studied by Laksaci *et al.* (2009).

In this work, we will assume that the the underlying process Z_i is functional stationary ergodic (see Laib and Louani (2011) for the definition and some examples). Of course, this work includes the finite dimensional case ($\mathcal{F} = \mathbb{R}^p$) but its importance is due to the fact that it covers also the infinite dimensional case. Because, these questions in infinite dimension are particularly interesting, not only for the fundamental problems they formulate, but also for many applications they may allow, see Bosq (2000), Ramsay and Silverman (2005) and Ferraty and Vieu (2006).

2.3 Notations, hypotheses and comments

All along the paper, when no confusion is possible, we will denote by C and C' some strictly positive generic constants, x is a fixed point in $\mathcal{F}$ and $\mathcal{N}_x$ denote a fixed neighborhood of x. For $r > 0$, let $B(x,r) := \{x' \in \mathcal{F}/\ d(x',x) < r\}$. Moreover, for $i = 1,\dots,n$, we put $\mathfrak{F}_k$ is the σ-field generated by $((X_1,Y_1),\dots(X_k,Y_k)))$ and we pose $\mathfrak{G}_k$ is the σ-field generated by $((X_1,Y_1),\dots(X_k,Y_k),X_{k+1})$.

In order to establish our asymptotic results we need the following hypotheses :

(H1) The processes $(X_i,Y_i)_{i\in\mathbb{N}}$ satisfies :

$$\begin{cases} \text{(i) The function } \phi(x,r) := \mathbb{P}(X \in B(x,r)) > 0, \quad \forall\, r > 0. \\ \text{(ii) For all } \ i = 1,\dots,n \ \text{ there exist a deterministic function } \ \phi_i(x,\cdot) \text{ such that} \\ \qquad\qquad 0 < \mathbb{P}\left(X_i \in B(x,r)|\mathfrak{F}_{i-1}\right) \le \phi_i(x,r), \ \ \forall\, r > 0. \\ \text{(iii) For all } \ r > 0, \ \frac{1}{n\phi(x,r)} \sum_{i=1}^n \mathbb{P}\left(X_i \in B(x,r)|\mathfrak{F}_{i-1}\right) \to 1 \qquad a.co. \end{cases}$$

(H2) The function Ψ is such that :

$$\begin{cases} \text{(i) The function } \Psi(x,\cdot) \text{ is of class } \mathcal{C}^1 \text{ on } \ [\theta_x - \delta, \theta_x + \delta],\ \delta > 0. \\ \text{(ii) For each fixed } t \in [\theta_x - \delta, \theta_x + \delta], \text{ the function } \Psi(\cdot,t) \text{ is continuous at the point} \\ \text{(ii)} \quad \forall (t_1,t_2) \in [\theta_x - \delta, \theta_x + \delta] \times [\theta_x - \delta, \theta_x + \delta], \ \forall (x_1,x_2) \in \mathcal{N}_x \times \mathcal{N}_x \\ \qquad\qquad |\Psi(x_1,t_1) - \Psi(x_2,t_2)| \le C d^{b_1}(x_1,x_2) + |t_1 - t_2|^{b_2}, \quad b_1 > 0, b_2 > 0. \end{cases}$$

(H3) For each fixed $t \in [\theta_x - \delta, \theta_x + \delta]$, $\forall j \ge 1$, $\mathbb{E}\left[\psi_x^j(Y,t)|\mathfrak{G}_{i-1}\right] = \mathbb{E}\left[\psi_x^j(Y,t)|X_i\right] < Cj! < \infty$, a.s.,

(H4) The function ψ_x is monotone w.r.t. the second component.

(H5) K is a function with support $(0,1)$ such that $0 < C1_{(0,1)} < K(t) < C'1_{(0,1)} < \infty$.

(H6) $\lim_{n\to\infty} h = 0$ and $\displaystyle\lim_{n\to\infty} \frac{\varphi(x,h)\log n}{n^2\phi^2(x,h)} = 0$ where $\displaystyle\varphi(x,h) = \sum_{i=1}^n \phi_i(x,h)$.

Comments on the hypotheses

Our assumptions are fairly mild in this context of nonparametric statistic in functional time series, in order to involve a larger class of precesses and/or of models. Moreover, it highlights the three structural axes of this subject, namely the ergodicity in functional statistic, the “dimensionality" and the robustness of the model. In the rest of this paragraph, we discuss the impact of each of these axes in our study.

- *On the ergodicity of functional data :* The latter is exploited together with condition (H1) which is less restrictive to the conditions imposed by Laib and Louani (2011) because, here, it is not necessary to write (approximatively) the concentration function $\mathbb{P}(X_i \in B(x,r))$ and the conditional concentration function $\mathbb{P}(X_i \in B(x,r)|\mathfrak{F}_{i-1})$ as product of two independent nonnegative functions of the center and of the radius. This weakness on the conditions is very important in functional case as well as in finite dimensional case. Indeed, in functional case, this asymptotic decomposition of these small ball probability functions requires the boundness of its associated Onsager-Machlup function (see Bogachev (1999)) but, here it is not necessary to employ this function. In addition, such writing form in multivariate case needs the differentiability of the marginal (resp. the conditional) distribution function and also the strict positivity of its densities, then, here, we can also proceed without the existence of these densities and even if these densities are vanishing at a center. Nevertheless, if this writing form of the concentration functions is verified, both conditions are equivalent and checked for several functional ergodic processes (see Laib and Louani (2011) for some examples). Furthermore, in the multivariate case, when the marginal density (resp. the conditional density) of X given $\mathfrak{F}_{i-1}$ exist and are continuous strictly positives, the hypothesis (H1)(iii) is a direct consequence of BeckŠs theorem (sse Gyorfi *et al.*, 1989, p. 49).
- *On the nonparametric model :* As all nonparametric analysis, the functional space of the model is characterized by some regularity conditions allow to evaluate the bias term. In this work, we control the nonparametric aspect of our study by considering two kinds of conditions the first one (H2(ii)) is the continuity of the model, while the second one is based on Lipschitz-type condition (H2(iii)). The first one is necessary to get the convergence, while the second consideration is used to precise the convergence rate of the estimate. We recall that, usually, in vectorial nonparametric statistic, we suppose that the underling model is k-times continuously differentiable, but, in our functional setting, the topological structure is insufficient to define the derivative operators. So, we can say that the Lipschitz condition is an alternative condition which permits to evaluate the bias term without using the differentiability.

– *On the robust method :* It is well known that the fundamentals constraints of the robustness properties of the M-estimators are the convexity and the boundness of the score function. The first one is important for the existence and the unicity of the estimate while the second one is essential to reduce the influence of large errors. In this work, the convexity propriety is controlled by means of the monotony condition (H4). However, we have opted for a presentation without the boundness condition in order to cover the classical regression studied in this ergodic functional context by Laib and Louani (2011). In conclusion, we can say that the smoothness condition imposed here on the score function ψ_x is mild and is a quite lower than that considered in Crambes *et al.* (2008), Azzedine *et al.* (2008) where ψ_x is assumed to be continuously differentiable which gives more flexibility in the practical choice of this function. It is worth noting that, except the quantiles regression model which required some special treatment, all the classical M-estimators are included such as the Winsor regression (ψ_x), the trimming regression (ψ),. . . .

2.4 Results

Our main first result is given in the following theorem which deals with pointwise almost complete (a.co.) [2] convergence.

Theorem 1 *Assume that (H1), (H2)((i)-(ii)) and (H3)-(H6) are satisfied then $\widehat{\theta_x}$ exists for all sufficiently large n. Furthermore, we have*

$$\widehat{\theta_x} - \theta_x \to 0 \qquad a.co.$$

In order to give a more accurate asymptotic result, we replace (H2) (ii) by H2(iii) and we obtain the following result

Theorem 2 *Assume that (H1), (H2) ((i)-(iii)) and (H3)-(H6) are satisfied, then $\widehat{\theta_x}$ exists a.s. for all sufficiently large n. Furthermore, if $\Psi'(x, \theta_x) \neq 0$*

2. Let $(z_n)_{n\in\mathbb{N}}$ be a sequence of real r.v.'s ; we say that z_n converges almost completely (a.co.) to zero if, and only if, $\forall\epsilon > 0$, $\sum_{n=1}^{\infty} P(|z_n| > \epsilon) < \infty$. Moreover, we say that the rate of almost complete convergence of z_n to zero is of order u_n (with $u_n \to 0$) and we write $z_n = O_{a.co.}(u_n)$ if, and only if, $\exists\epsilon > 0$, $\sum_{n=1}^{\infty} P(|z_n| > \epsilon u_n) < \infty$. This kind of convergence implies both almost sure convergence and convergence in probability.

we have

$$\widehat{\theta}_x - \theta_x = O\left(h^{b_1}\right) + O\left(\sqrt{\frac{\varphi(x,h)\log n}{n^2\phi^2(x,h)}}\right) \qquad a.co.$$

Proof of the main result : For the proofs of Theorems (1) and (2) we use the fact that ψ_x is monotone w.r.t. the second component. We give the proof for the case of an increasing $\psi_x(Y,\cdot)$, decreasing case being obtained by the same manner. For this consideration, we write, $\forall\epsilon > 0$

$$\begin{aligned}\sum_n \mathbb{P}\left[|\widehat{\theta}_x - \theta_x| \geq \epsilon\right] &\leq \sum_n \mathbb{P}\left[\left(|\widehat{\theta}_x - \theta_x| 1_{\left\{|\widehat{\theta}_x - \theta_x| \leq \delta\right\}}\right) \geq \epsilon\right] \\ &\quad + \sum_n \mathbb{P}\left[\left(|\widehat{\theta}_x - \theta_x| \mathbb{P}_{\left\{|\widehat{\theta}_x - \theta_x| > \delta\right\}}\right) \geq \epsilon\right].\end{aligned}$$

Because $\psi_x(Y,\cdot)$ is increasing, it follows that,

$$\begin{aligned}\mathbb{P}\left(\left(|\widehat{\theta}_x - \theta_x| 1_{\left\{|\widehat{\theta}_x - \theta_x| > \delta_0\right\}}\right) \geq \epsilon\right) &\leq \mathbb{P}\left(|\widehat{\theta}_x - \theta_x| > \delta_0\right) \\ &\leq \mathbb{P}\left(|\widehat{\Psi}(x,\theta_x+\delta) - \Psi(x,\theta_x+\delta)| \geq \Psi(x,\theta_x+\delta)\right) \\ &\qquad + \mathbb{P}\left(|\widehat{\Psi}(x,\theta_x-\delta) - \Psi(x,\theta_x-\delta)| \geq -\Psi(x,\theta_x-\delta)\right).\end{aligned}$$

Moreover, under ((H2) (i)), we have

$$\left(\widehat{\theta}_x - \theta_x\right) 1_{\left\{|\widehat{\theta}_x - \theta_x| \leq \delta\right\}} = \frac{\Psi(x,\widehat{\theta}_x) - \widehat{\Psi}(x,\widehat{\theta}_x)}{\Psi'(x,\xi_n)}$$

where ξ_n is between $\widehat{\theta}_x$ and θ_x.

Therefore, all what is left to do, is to study the convergence rate of

$$\sup_{t\in[\theta_x-\delta,\,\theta_x+\delta]} |\Psi(x,t) - \widehat{\Psi}(x,t)|$$

and to show that

$$\exists\tau > 0,\ \sum_{n=1}^{\infty} \mathbb{P}\left(\Psi'(x,\xi_n) < \tau\right) \ < \ \infty. \tag{2.2}$$

To do that, we write

$$\widehat{\Psi}(x,t) = B_n(x,t) + \frac{R_n(x,t)}{\widehat{\Psi}_D(x)} + \frac{Q_n(x,t)}{\widehat{\Psi}_D(x)}$$

where

$$
\begin{aligned}
Q_n(x,t) &:= (\widehat{\Psi}_N(x,t) - \bar{\Psi}_N(x,t)) - \Psi(x,t)(\widehat{\Psi}_D(x) - \bar{\Psi}_D(x)) \\
B_n(x,t) &:= \frac{\bar{\Psi}_N(x,t)}{\bar{\Psi}_D(x)} - \Psi(x,t), \quad \text{and} \quad R_n(x,t) := -B_n(t)(\widehat{\Psi}_N(x,t) - \bar{\Psi}_N(x,t))
\end{aligned}
$$

with

$$
\begin{aligned}
\widehat{\Psi}_N(x,t) &:= \frac{1}{n\mathbb{E}[K(h^{-1}d(x,X_1))]} \sum_{i=1}^{n} K(h^{-1}d(x,X_i))\psi_x(Y_i,t), \\
\bar{\Psi}_N(x,t) &:= \frac{1}{n\mathbb{E}[K(h^{-1}d(x,X_1))]} \sum_{i=1}^{n} \mathbb{E}\left[K(h^{-1}d(x,X_i))\psi_x(Y_i,t)|\mathfrak{F}_{i-1}\right], \\
\widehat{\Psi}_D(x) &:= \frac{1}{n\mathbb{E}[K(h^{-1}d(x,X_1))]} \sum_{i=1}^{n} K(h^{-1}d(x,X_i)), \\
\bar{\Psi}_D(x) &:= \frac{1}{n\mathbb{E}[K(h^{-1}d(x,X_1))]} \sum_{i=1}^{n} \mathbb{E}\left[K(h^{-1}d(x,X_i))|\mathfrak{F}_{i-1}\right].
\end{aligned}
$$

Thus, both Theorems are a consequence of the following intermediates results, where their proofs are given at the end.

Lemma 1 *Under Hypotheses (H1),(H5) and (H6), we have,*

$$
\widehat{\Psi}_D(x) - \bar{\Psi}_D(x) = O\left(\sqrt{\frac{\varphi(x,h)\log n}{n^2\phi^2(x,h)}}\right) \qquad a.co.
$$

Corollary 1 *Under Hypotheses of Lemma 2, we have,*

$$
\exists\, C > 0 \qquad \sum_{n=1}^{\infty} \mathbb{P}\left(\widehat{\Psi}_D(x) < C\right) < \infty.
$$

Lemma 2 *Under Hypotheses (H1),(H2)((i)-(ii)), (H5) and (H6), we have,*

$$
\sup_{t\in[\theta_x-\delta,\theta_x+\delta]} |B_n(x,t)| = o(1).
$$

If we replace ((H2) (ii)) by ((H2) (iii)), we have

$$
\sup_{t\in[\theta_x-\delta,\theta_x+\delta]} |B_n(x,t)| = O(h^{b_1}).
$$

Lemma 3 *Under Hypotheses (H1) and (H3)-(H6), we have,*

$$\sup_{t\in[\theta_x-\delta,\theta_x+\delta]} |\widehat{\Psi}_N(x,t) - \bar{\Psi}_N(x,t)| = O\left(\sqrt{\frac{\varphi(x,h)\log n}{n^2\phi^2(x,h)}}\right) \qquad a.co.$$

Lemma 4 *Under Assumptions (H1), (H2) ((i)-(ii)) and (H3)-(H6), $\widehat{\theta}_x$ exists a.s. for all sufficiently large n and there exists $\zeta_1 > 0$ such that*

$$\sum_{n\geq 1} \mathbb{P}\left\{\Psi'(x,\xi_n) < \zeta_1\right\} < \infty.$$

2.5 Discussions

2.5.1 Some particulars cases

In this section we emphasize the generality of our study over several existing results in both functional or nonfunctional statistic. To do that, we consider three special cases such as the classical regression case, the independent case and the multivariate case which are respectively studied, Laib and Louani(2011), Azzedine *et al.* (2008),

- *The classical regression case :* As noticed earlier the classical regression defined by conditional expectation is a particular case of our study with $\psi_x(Y,t) = (Y-t)$. Clearly, condition (H4) is verified with this score function. So, for this particular case we obtain the following convergence rate.

  Corollary 2 *Under the hypotheses (H1) (H2)((i)-(iii)),(H3), (H5), (H6) and if $\frac{\varphi(x,h)}{n\phi(x,h)} = O_{a.s.}(1)$, we have*

  $$\widehat{\theta}_x - \theta_x = O\left(h^{b_1}\right) + O\left(\sqrt{\frac{\log n}{n\phi(x,h)}}\right) \qquad a.co.$$

  Remark 1

  It is clear that, this convergence rate is exactly what is obtained by Laib and Louani (2011) for the standard regression model
- *The independent case :* In this situation, condition (H1(ii)) and (H1(iii)) are automatically verified and for all $i = 1, \ldots n$ take $\phi_i(x,r) = \phi(x,r)$. Therefore, condition (H1) is restricted to $\phi(x,r) > 0$, for all $r > 0$. Thus, our Theorem leads to the next Corollary,

Corollary 3 *Under assumptions (H1), (H2)((i)-iii)) and (H3)-(H6) we have :*

$$O\left(h^{b_1}\right) + O\left(\sqrt{\frac{\log n}{n\phi(x,h)}}\right) \qquad a.co.$$

Remark 2
We point out that in this case where the (X_i, Y_i) are independent, we obtain the same convergence rate given by Azzedine *et al* . (2008).

– *The real case* As mentioned in the previous section, in the real case when $\mathcal{F} = \mathbb{R}$, and if the probability density of the random variable X (resp. the conditional density of X given $\mathfrak{F}_{i-1}$) denoted by f (resp. by $f_i^{\mathfrak{F}_{i-1}}$), is of $\mathcal{C}^1$ class, then $\phi(x,h) = f(x)h + o(h)$ and $\mathbb{P}(X_i \in [x-h, x+h]|\mathfrak{F}_{i-1}) = f_i^{\mathfrak{F}_{i-1}}(x)h + o(h))$. Moreover the ergodic Theorem insure that

$$\|\frac{1}{n}\sum_{i=1}^{n} f_i^{\mathfrak{F}_{i-1}} - f\| \to 0,$$

where $\|.\|$ is a norm in sparable Banach space $\mathcal{C}^1$. Therefore condition (H1) is verified and Theorem 2 can be reformulated in the following way.

Corollary 4 *Under assumptions H2 (i), H2 (iii) and (H3)-(H6) we have :*

$$O\left(h^{b_1}\right) + O\left(\sqrt{\frac{\log n}{nh}}\right) \qquad a.co.$$

Remark 3
A similar thing can be concluded if $\mathcal{F} = \mathbb{R}^p$ we show that the convergence rate is given by

$$O\left(h^{b_1}\right) + O\left(\sqrt{\frac{\log n}{nh^p}}\right) \qquad a.co.$$

Moreover, we recall that, our result can be stated even if X has no density with respect to Lebesgue measure. Indeed, it suffices to express $\phi(x,h)$ by means of the cumulative distribution of X as follows

$$\phi(x,h) = F(x+h) - F(x-h).$$

2.5.2 Some prospects

This Section is devoted to discuss some interesting open questions as natural prospects of the present work.

- *The asymptotic normality of the estimator :* Motivated by its interest in various statistical problems such as the confidence's interval or the test's problem, the asymptotic normality of our estimator is one of the important prospects of our work. Recall that, this asymptotic propriety has particular interest in the robust statistic, because it permits to determine the asymptotic variance which is one of the main quantifiers of robustness. Of course, the asymptotic normality can be obtained by a standard arguments (Lindeberg condition), but it seems more interesting, if we can adopt the recent development of Delsol (2009) in α-mixing to functional ergodic data. The latter has established the asymptotic normality of the classical regression under very mild assumptions.
- *The choice of the smoothing parameter :* It is clear that, the selection of the bandwidth parameter h play a crucial role on the efficiency of the estimate in practice. Despite its importance this problem has not yet been fully explored in nonparametric functional data analysis. The few existing results concerns only the kernel estimation of the classical functional nonparametric regression in the i.i.d case (see Benheni et al, 2007). Their ideas can be combined together with the cross-validation rule of Heng-Yan Leung (2005) in order to develop a data-driven method for choosing automatically the optimal smoothing parameter which keeps the robust features of our model in this context of functional ergodic data. The theoretical validity of this procedure is very interesting open question in the future.
- *The local M-estimator :* it is well known that a local polynomial smoothing has various advantages over the kernel method, in particular, this method has superior bias properties to the previous one (see Fan and Gijbels (1996) for an extensive discussion on the comparison between the both methods). Recently, this estimation method has been considered in nonparametric functional statistic by Barrientos-Marin et al. (2010). They proposed a local linear estimate of the standard nonparametric regression with functional regressor. The robustification of this estimate is also an important prospect of this work. Let us point out that the ideas of Cai and Ould-said (2003) in the vectorial case could be useful in this subject

2.6 Appendix

The proof of Lemma 2 and Lemma 2 are, respectively, very close to those of Lemmas 1 and Lemma 2 in Laib and Louani (2011). Moreover, the proof of the first part of Lemma 5 is similar to the Lemma 5 in Azzedine *et al.* (2008) and its second part is a direct consequence of the regularity assumption (H2) (i) on $\Psi(x,\cdot)$). Thus, for sake of shortness, we omit the proof of these three Lemmas

Proof of Corollary 1

It is clear that, under (H5), there exists $0 < C < C' < \infty$

$$0 < C\frac{1}{n\phi(x,h)}\sum_{i=1}^{n}\mathbb{P}(X_i \in B(x,r)|\mathfrak{F}_{i-1}) < \bar{\Psi}_D(x) < \left|\widehat{\Psi}_D(x) - \bar{\Psi}_D(x)\right| + \widehat{\Psi}_D(x).$$

Hence,

$$\begin{aligned}\mathbb{P}\left(\widehat{\Psi}_D(x) \le \tfrac{C}{2}\right) \quad &\le \mathbb{P}\left(\frac{C}{n\phi(x,h)}\sum_{i=1}^{n}\mathbb{P}(X_i \in B(x,r)|\mathfrak{F}_{i-1}) < \frac{C}{2} + \left|\widehat{\Psi}_D(x) - \bar{\Psi}_D(x)\right|\right)\\ &\le \mathbb{P}\left(\left|\frac{C}{n\phi(x,h)}\sum_{i=1}^{n}\mathbb{P}(X_i \in B(x,r)|\mathfrak{F}_{i-1}) - \left|\widehat{\Psi}_D(x) - \bar{\Psi}_D(x)\right| - C\right| > \frac{C}{2}\right).\end{aligned}$$

It is obvious that the previous Lemma and (H1)(iii) allows to get

$$\sum_n \mathbb{P}\left(\left|\frac{C}{n\phi(x,h)}\sum_{i=1}^{n}\mathbb{P}(X_i \in B(x,r)|\mathfrak{F}_{i-1}) - \left|\widehat{\Psi}_D(x) - \bar{\Psi}_D(x)\right| - C\right| > \frac{C}{2}\right).$$

which gives the result.

Proof of Lemma 4

Using the compactness of $[\theta_x - \delta,\, \theta_x + \delta]$, we can write $[\theta_x - \delta,\, \theta_x + \delta] \subset \bigcup_{j=1}^{d_n}(t_j - l_n, t_j + l_n)$ with $l_n = n^{-1/2b_2}$ and $d_n = O\left(n^{1/2b_2}\right)$. We consider the intervals extremities grid

$$\mathcal{G}_n = \{t_j - l_n, t_j + l_n, 1 \le j \le d_n\}. \tag{2.3}$$

Similarly to Lemma 13 of Ferraty et al. (2010), we use the monotony of ψ_x to show that

$$\sup_{t\in[\theta_x-\delta,\,\theta_x+\delta]}\left|\widehat{\Psi}_N(x,t) - \bar{\Psi}_N(x,t)\right| \le \max_{1\le j\le d_n}\max_{z\in\{t_j-l_n,t_j+l_n\}}\left|\widehat{\Psi}_N(x,z) - \bar{\Psi}_N(x,z)\right| + 2^{b_2}C_2 l_n^{b_2}\bar{\Psi}_D(x).$$

From Lemma 2, we deduce that

$$l_n^{b_2}\bar{\Psi}_D(x) = O_{a.co.}\left(\sqrt{\frac{1}{n}}\right) = O_{a.co.}\left(\sqrt{\frac{\varphi(x,h)\log n}{n^2\phi^2(x,h)}}\right).$$

Thus, it suffices to show that

$$\max_{1\leq j\leq d_n}\max_{z\in\{t_j-l_n,t_j+l_n\}}\left|\widehat{\Psi}_N(x,z)-\bar{\Psi}_N(x,z)\right| = O_{a.co.}\left(\sqrt{\frac{\varphi(x,h)\log n}{n^2\phi^2(x,h)}}\right).$$

The proof of (2.6) is based on the application of the exponential inequality of Lemma 1 in Louani and Laib (2011, P.365) on the triangular array of martingale differences (according the σ-fields $(\mathfrak{F}_{i-1})_i$)

$$\Lambda_i(x,z) = K_i(x)\psi_x(Y_i,z) - \mathbb{E}\left[K_i(x)\psi_x(Y_i,z)|\mathfrak{F}_{i-1}\right], \qquad z\in\mathcal{G}_n$$

where $K_i(x) = K(h^{-1}d(x,X_i))$. For this, we must evaluate the quantity $\mathbb{E}[\Lambda_i^p(x,z)|\mathfrak{F}_{i-1}]$. The Latter can be evaluated by the same arguments as those invoked for proving Lemma 5 in Laib and Louani (2011) which allow to write, under (H3)

$$\mathbb{E}[\Lambda_i^p(x,z)|\mathfrak{F}_{i-1}] \leq C\phi_i(x,h).$$

Thus, we are now in position to apply the mentioned exponential inequality and we get : for all $\eta>0$ and $d_n = l_n^{-1}$, we have

$$l_n^{-1}\max_{z\in\mathcal{G}_n}\mathbb{P}\left(\left|\widehat{\Psi}_N(x,z)-\bar{\Psi}_N(x,z)\right| > \eta\sqrt{\frac{\varphi(x,h)\log n}{n^2\,\phi^2(h)}}\right) \leq C'n^{-C\eta^2+1/2b_2}.$$

Consequently, an appropriate choice of η completes the proof of this lemma.

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

Bibliographie

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

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

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

Boente, G. and Fraiman, R. (1989). Nonparametric regression estimation. *J. Multivariate Anal.* **29**, 180-198.

Bogachev, V.I., 1999. Gaussian measures. Math surveys and monographs. 62, Amer. Math. Soc.

Bosq, D. (1996) Limit theorems for Banach-valued autoregressive processes. Applications to real continuous time processes. *Bull. Belg. Math. Soc.* **3**, 537–555.

Bosq, D. (2000). Linear processs in function spaces. Théorie and Application. *Lectures Notes in Statistics. Springer Verlag, New York.* **149**.

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

Crambes, C. Delsol, L. and Laksaci, A. (2008). Robust nonparametric estimation for functional data. *J. of Nonparametric Statistics*

20, 573-598.

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

Davidian, M. Lin, X. Wang, J.L. (2004). Introduction. *Statist. Sinica* **14**, 613–614.

Delsol, L (2009). Advances on asymptotic normality in non-parametric functional time series analysis. *Statistics*, **43**, 13–33.

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

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

Ferraty, F. (2010). High-dimensional data : a fascinating statistical challenge. *J. Multivariate Anal.* **101**, 305–306

Ferraty, F., Laksaci, A., Tadj, A., and Vieu, P. (2010). Rate of uniform consistency for nonparametric estimates with functional variables. *J. Statist. Plan. and Inference,* **140**, 335–352.

Ferraty, F. and Romain, Y. (2011). *The Oxford handbook of functional data analysis.* Oxford University Press.

Györfi, L., Härdle, W., Sarda, P. et Vieu, P. (1989). *Nonparametric curve estimation for time series.* Lecture Notes in Statistics, **60**, Springer-Verlag.

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

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

Laksaci, A. Lemdani, M. Ould-Saïd, E. (2009) L^1-norm kernel estimator of conditional quantile for functional regressors : consistency and asymptotic normality. *Statist. Probab. Lett.* **79**, 1065-1073 ;

Laïb, N. and Louani, D. (2011). Strong consistency of the regression function estimator for functional stationary ergodic data. *J. Statist. Plan. and Inference,* **141**, 359-372.

Manteiga, W. G. Vieu, P. (2007) Statistics for functional data. *Comput. Statist. Data Anal.* **51**, 4788–4792.

Ramsay, J. O. and Silverman, B. W. (2005). *Functional Data Analysis*, 2*nd* Edition, Springer, New-York.

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

Valderrama, M. J. (2007). An overview to modelling functional data.*Comput. Statist.* **22**, 331–334.

Chapitre 3

Asymptotic normality

Ce travail traite, sous l'hypothèse d'un processus ergodique stationnaire, la normalité asymptotique de même estimateur. Ce travail fait l'objet d'un article soumis.

Asymptotic normality of kernel estimator of ψ-regression function for functional ergodic data

Abdelkader Gheriballah, Ali Laksaci[1] and Soumeya Sekkal
Laboratoire de statistique et processus stochastiques
Univ. Djillali Liabès
BP 89, 22000 Sidi Bel Abbès, Algeria
e-mail : (gheribaek, alilak or s_sekkal)@yahoo.fr

Abstract
In this paper we consider the problem of the estimation of the ψ-regression function when the covariates take values in an infinite dimensional space. Our main aim is to establish, under a stationary ergodic process assumption, the asymptotic normality of this estimate.

3.1 Introduction

The statistical analysis of functional data has received much attention in the last few years. The are several special issues dedicated to this topic by various statistical journals (see, for instance Davidian et al. (2004), Gonzalez-Manteiga and Vieu (2007), Valderrama (2007) or Ferraty (2010)). This particular interest of this topic is due to the diversity of application's fields in which the data to be treated are curves. Indeed, it is well documented that this kind of data can be found, for instance, in chemometrics, environmetrics, sciences medical, speech recognition, economics, ... The main purpose of this contribution is to study the asymptotic properties of the nonparametric robust regression, under less restrictive dependence conditions that is the ergodicity assumption.

The nonparametric robust analysis of functional data is a very recent field of investigations. The first results in this subject was given by Azzedine *et al.* (2008). Since this work, a considerable corpus of research has been dedicated to the robust nonparametric modeling in functional data analysis. This great consideration is motivated by the fact that the robust nonparametric regression provides an alternative approach to the classical methods which is insensitive to the presence of outliers or heteroskedastic variables. Considering a sample of i.i.d, Azzedine *et al.* (2008) have established the almost complete convergence of the robust estimators of the regression function when the regressors is functional. In the same context, Attouch *et al.* (2009) studied the asymptotic normality of this model. The robust analysis in functional time series, has been investigated by many authors. We refer, for example, to Crambes *et al.* (2008) for the convergence in L^q norm, Attouch *et al.* (2010) for the asymptotic normality and Chen and Zhang (2009) for the weak and strong consistency of the nonparametric functional conditional location estimate in the mixing case.

While in all these literatures, the functional time series data is modeled by using the mixing conditions, our main aim in this paper is to consider the problem of the robust nonparametric analysis of functional ergodic time series data. This dependence structure covers several case does not satisfy the usual mixing structures. Moreover, the ergodic framework avoid the widely used strong mixing condition and its variants to measure the dependency and the very involved probabilistic calculations that it implies. In addition, from practical point of view the ergodicity condition is one of a principal postulate of statistical physics. It models the thermodynamic properties of gases, atoms, electrons or plasmas. This hypothesis is also used in signal processing, for studying the evolution of a random signal. Despite this importance in applications, the litterateur in this functional dependency is still limited. This problem has been initiated by Laib and Louani (2010, 2011). They consider the problem of functional estimation for nonparametric regression operators under the ergodicity condition. Recently Gheiriballah et al (2013) gave the almost complete convergence (with rate) of a family of robust nonparametric estimators for regression function. More recently Benziadi et al. (2014) stated the almost complete convergence of a recursive kernel estimates of the conditionals quantiles.

Our main goal in this paper is to establish the asymptotic normality of the estimator proposed by Gheiriballah et al (2013). This results is obtained under, some general condition. We recall that the asymptotic normality is a fundamental preliminary result to build confidence interval of our model or to precis the asymptotic dominant terms of the moments of order q by explicating the asymptotic bias and the asymptotic variance which are a basic ingredients of this error. All these derivatives of our asymptotic results has been discussed in this paper. It should be noted that our hypotheses and results unify the both cases of finite or infinite dimension of the regressors. Thus, we can say that our result it also even in vectorial statistic.

The setup of this paper is as follows : We present our model in Section 2. Section 3 is dedicated to fixing notations and hypotheses. Our main result is given in Section 4. In Section 5 we show the generality of our model and the flexibility of our conditions by studying some particular case. In Section 6, we discuss the importance of the asymptotic normality property by giving some direct applications. The proofs of the auxiliary results are relegated to the Appendix.

3.2 The ψ- regression model and its estimate

Let $Z_i = (X_i, Y_i)_{i=1,...n}$ be a $\mathcal{F} \times \mathbb{R}$-valued measurable strictly stationary process, defined on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$, where $\mathcal{F}$ is a semi-metric space, d denoting the semi-metric. For any x in $\mathcal{F}$, we consider ψ_x a real-valued Borel function satisfying some regularity conditions to be stated below. The nonparametric model studied in this paper, denoted by θ_x, is implicitly defined as a zero with respect to (w.r.t.) t of the following equation

$$\Psi(x,t) := \mathbb{E}\left[\psi_x(Y_i,t) \mid X_i = x\right] = 0. \tag{3.1}$$

We suppose that, for all $x \in \mathcal{F}$, θ_x exists and is the unique zero w.r.t. t of (3.1) (see, for instance Boente and Fraiman (1989) or Koul and Stute (1998) for the existence and uniqueness of θ_x). We point out that this robustification method belongs to the class of M-estimates introduced by Huber (1964) and it covers and includes many important nonparametric models, for example, $\psi_x(y,t) = (y-t)$ yields the classical regression, $\psi_x(y,t) = 1\!\!1_{y \geq t} - 1\!\!1_{y<t}$ leads to the conditional median function $m(x) = med(Y|X = x)$ and the α^{th} conditional quantile is obtained by setting $\psi_x(y,t) = 1\!\!1_{y>t} - (1-\alpha)$, $\alpha \in (0,1)$. In addition, our robustification method allows us to consider the functional nonparametric regression model with a scale of the error assumed to be known by taking $\psi_x(.,.) = \psi(. - ./\sigma(x))$, where $\sigma(.)$ is a measure of spread for the conditional distribution of Y given $X = x$. We return to Stone (2005) for other examples of the function ψ.

For all $(x,t) \in \mathcal{F} \times \mathbb{R}$, we propose a nonparametric estimator of $\Psi(x,t)$ given by

$$\widehat{\Psi}(x,t) := \frac{\sum_{i=1}^{n} K(h^{-1}d(x,X_i))\psi_x(Y_i,t)}{\sum_{i=1}^{n} K(h^{-1}d(x,X_i))},$$

where K is a kernel and $h = h_{\mathbf{n}}$ is a sequence of positive real numbers. A natural estimator $\widehat{\theta_x}$ of θ_x is a zero w.r.t. t of the equation

$$\widehat{\Psi}(x,t) = 0.$$

Obviously, when $\psi_x(Y,t) = Y - t$, then $\widehat{\theta}_n$ is the estimator given in Ferraty and Vieu (2006) for the functional nonparametric regression. While for $\psi_x(y,t) = \mathbb{1}_{y>t} - (1-\alpha)$, we obtain the α^{th} conditional quantile estimate studied by Laksaci et al. (2009).

In this work, we will assume that the underlying process Z_i is functional stationary ergodic (see Laib and Louani (2011) for the definition and some examples). Of course, this work includes the finite dimensional case ($\mathcal{F} = \mathbb{R}^p$) but its importance is due to the fact that it covers also the infinite dimensional case. Because, these questions in infinite dimension are particularly interesting, not only for the fundamental problems they formulate, but also for many applications they may allow, see Bosq (2000), Ramsay and Silverman (2005) and Ferraty and Vieu (2006).

3.3 Notations, hypotheses and comments

All along the paper, when no confusion is possible, we will denote by C and C' some strictly positive generic constants, x is a fixed point in $\mathcal{F}$ and $\mathcal{N}_x$ denote a fixed neighborhood of x. For $r > 0$, let $B(x,r) := \{x' \in \mathcal{F}/\ d(x',x) < r\}$. Moreover, for $i = 1,\dots,n$, we put $\mathfrak{F}_k$ is the σ-field generated by $((X_1,Y_1),\dots(X_k,Y_k)))$ and we pose $\mathfrak{G}_k$ is the σ-field generated by $((X_1,Y_1),\dots(X_k,Y_k),X_{k+1})$.

In order to establish our asymptotic results we need the following hypotheses :

(H1) The processes $(X_i,Y_i)_{i\in\mathbb{N}}$ satisfies :

$$\begin{cases} \text{(i) The functions } \phi(x,r) := \mathbb{P}(X \in B(x,r)) > 0, \\ \qquad\qquad\qquad\qquad \text{and} \\ \phi_i(x,r) = \mathbb{P}\left(X_i \in B(x,r) | \mathfrak{F}_{i-1}\right) > 0 \ \forall\, r > 0. \\ \text{(ii) For all } \ r > 0,\ \dfrac{1}{n\phi(x,r)} \displaystyle\sum_{i=1}^{n} \phi_i(x,r) \to^p 1 \qquad \text{and} \qquad n\phi(x,h) \to \infty \quad as \quad h \to 0. \end{cases}$$

(H2) The functions Ψ such that :

$$\begin{cases} \text{(i) The function } \Psi(x,\cdot) \text{ is of class } \mathcal{C}^1 \text{ at } N_x \text{ a fixed neighborhood of } \theta_x. \\ \text{(ii) For each fixed } \ t \text{ in } N_x \text{ the functions } \Psi(\cdot,t) \text{ and } \lambda_2(\cdot,t) = \mathbb{E}[\psi_x^2(Y,t)|X=\cdot] \\ \quad \text{are continuous at the point } x. \\ \text{(iii) The derivative of the real function } \ \Phi(s,z) = \mathbb{E}\left[\Psi(X_1,z) - \Psi(x,z)|d(x,X_1) = s\right] \\ \qquad\qquad \text{exists at } s = 0 \ \text{ and is continuous w.r.t. the second component at } N_x \end{cases}$$

(H3) For each fixed t in the neighborhood of θ_x $\forall j \geq 2$,

$$\mathbb{E}\left[\psi_x^j(Y,t)|\mathfrak{G}_{i-1}\right] = \mathbb{E}\left[\psi_x^j(Y,t)|X_i\right] < C < \infty, a.s.,$$

(H4) The function ψ_x is continuous and monotone w.r.t. the second component.

(H5) The kernel K is a positive function supported on $(0, 1[$. Its derivative K' exists on $(0, 1)$ and satisfies $K'(t) < 0$ for $0 < t < 1$.

(H6) There exists a function $\tau_x(\cdot)$ such that

$$\forall t \in [0,1] \quad \lim_{h\to 0} \frac{\phi(x,th)}{\phi(x,h)} = \tau_x(t),$$

$$K^2(1) - \int_0^1 (K^2(u))'\tau_x(u)du > 0 \quad \text{and} \qquad K(1) - \int_0^1 K'(u)\tau_x(u)du \neq 0.$$

Comments on the hypotheses

Our assumptions are quite mild. Indeed, the ergodicity of functional data : The latter is exploited together with condition (H1ii) which is less restrictive to the conditions imposed by Laib and Louani (2011) (see Gheriballah et al.(2013) for more discussion). In this work, the functional space of our model is characterized by the regularity condition (H2iii). This condition replace the Lipschitz condition usually assumed in nonparametric functional data analysis. This change is useful in order to explicit asymptotically the bias term. However, the Lipschitz condition gives inexact/inaccurate asymptotic bias term which not interesting for the asymptotic normality. The robustness property is controlled by (H4) where only the convexity (which is fundamentals constraints of the robustness properties of the M-estimators) of the score function is needed. In order to cover the classical regression studied in this ergodic functional context by Laib and Louani (2011) we establish our asymptotic normality without the boundedness condition for the score function. Condition (H3), (H5) and (H6) are very similar to those used by Ferraty et al. (2007). Moreover, the function $\tau_x(\cdot)$ defined in (H5) plays a fundamental role in the asymptotic normality result. It permits to give the variance term explicitly. Note that this function can be specified in several situations where the function $\phi(x, h)$ is known and (H5) is fulfilled. We quote the following cases (which can be found in Ferraty *et al.* (2007)) :
i) $\phi(x, h) = C_x h^\gamma + o(h^\gamma)$ for some $\gamma > 0$ with $\tau_x(s) = s^\gamma$,
ii) $\phi(x, h) = C_x h^{\gamma_1} \exp\{-Ch^{-\gamma_2}\} + o\left(h^{\gamma_1} \exp\{-Ch^{-\gamma_2}\}\right)$ for some $\gamma_1 > 0$ and $\gamma_2 > 0$ with $\tau_x(\cdot)$ being Dirac's function in 1,

3.4 Results

Our main result is given in the following theorem

Theorem 2 *Assume that (H1)-(H6) hold, then $\widehat{\theta_x}$ exists and is unique with great probability and for any $x \in \mathcal{A}$, we have*

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

where

$$B_n(x) = h\Phi'(0,\theta_x)\frac{\beta_0}{\beta_1} + o(h) \;\; and\, \sigma^2(x,\theta_x) = \frac{\beta_2\lambda_2(x,\theta_x)}{\beta_1^2(\Gamma_1(x,\theta_x))^2}$$

with

$$\beta_0 = -\int_0^1 (sK(s))'\beta_x(s)ds, \quad \beta_j = -\int_0^1 (K^j)'(s)\beta_x(s)ds, \; for, \; j = 1,\, 2),$$

$$\Gamma_1(x,\theta_x) = \frac{\partial}{\partial t}\Psi(x,\theta_x) \quad and \quad \mathcal{A} = \{x \in \mathcal{F}, \; \lambda_2(x,\theta_x)\Gamma_1(x,\theta_x) \neq 0\}$$

and $\xrightarrow{\mathcal{D}}$ means the convergence in distribution.

In order to remove the bias term $B_n(x)$, we need an additional condition on the bandwidth parameter h.

Corollary 5 *Under the hypotheses of Theorem 2 and if the bandwidth parameter h satisfies $nh^2\phi(x,h) \to 0$ as $n \to \infty$, we have*

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

Proof of Theorem 2 and Corollary 5
We give the proof for the case of a increasing ψ_x, decreasing case being obtained by considering $-\psi_x$. In this case, we define, for all $u \in \mathbb{R}$, $z = \theta_x - B_n(x) + u\,[n\phi(x,h)]^{-1/2}\,\sigma(x,\theta_x)$. Let us remark that,

$$\begin{aligned}\mathbb{P}\left\{\left(\frac{n\phi(x,h)}{\sigma^2(x,\theta_x)}\right)^{1/2}\left(\widehat{\theta_x} - \theta_x + B_n(x)\right) < u\right\} &= \mathbb{P}\left\{\widehat{\theta_x} < \theta_x - B_n(x) + u\,[n\phi(x,h)]^{-1/2}\,\sigma(x,\theta_x)\right\} \\ &= \mathbb{P}\left\{0 < \widehat{\Psi}(x,z)\right\}.\end{aligned}$$

It is clear thar we can write

$$\widehat{\Psi}(x,t) = B_n(x,t) + \frac{R_n(x,t)}{\widehat{\Psi}_D(x)} + \frac{Q_n(x,t)}{\widehat{\Psi}_D(x)}$$

where

$$\begin{aligned} Q_n(x,t) &:= (\widehat{\Psi}_N(x,t) - \bar{\Psi}_N(x,t)) - \Psi(x,t)(\widehat{\Psi}_D(x) - \bar{\Psi}_D(x)) \\ B_n(x,t) &:= \frac{\bar{\Psi}_N(x,t)}{\bar{\Psi}_D(x)}, \quad \text{and} \quad R_n(x,t) := -\left(\frac{\bar{\Psi}_N(x,t)}{\bar{\Psi}_D(x)} - \Psi(x,t)\right)(\widehat{\Psi}_N(x,t) - \bar{\Psi}_N(x,t)) \end{aligned}$$

with

$$\begin{aligned} \widehat{\Psi}_N(x,t) &:= \frac{1}{n\mathbb{E}[K(h^{-1}d(x,X_1))]} \sum_{i=1}^n K(h^{-1}d(x,X_i))\psi_x(Y_i,t), \\ \bar{\Psi}_N(x,t) &:= \frac{1}{n\mathbb{E}[K(h^{-1}d(x,X_1))]} \sum_{i=1}^n \mathbb{E}\left[K(h^{-1}d(x,X_i))\psi_x(Y_i,t)|\mathfrak{F}_{i-1}\right], \\ \widehat{\Psi}_D(x) &:= \frac{1}{n\mathbb{E}[K(h^{-1}d(x,X_1))]} \sum_{i=1}^n K(h^{-1}d(x,X_i)), \\ \bar{\Psi}_D(x) &:= \frac{1}{n\mathbb{E}[K(h^{-1}d(x,X_1))]} \sum_{i=1}^n \mathbb{E}\left[K(h^{-1}d(x,X_i))|\mathfrak{F}_{i-1}\right]. \end{aligned}$$

It follows that

$$\mathbb{P}\left\{\left(\frac{n\phi(x,h)}{\sigma^2(x,\theta_x)}\right)^{1/2}\left(\widehat{\theta_x} - \theta_x + B_n(x)\right) < u\right\} = \mathbb{P}\left\{-\widehat{\Psi}_D(x)B_n(x,z) - R_n(x,z) < Q_n(x,z\right.$$

Therefore, our main result is a consequence of the following intermediates results.

Lemma **1** *Under the hypotheses of Theorem 2, we have for any* $x \in \mathcal{A}$

$$\left(\frac{n\phi(x,h)\beta_1^2}{\beta_2\lambda_2(x,\theta_x)}\right)^{1/2} Q_n(x,z) \xrightarrow{\mathcal{D}} \mathcal{N}(0,1) \quad \text{as } n \to \infty.$$

Lemma **2** *(see, Laib and Louani, 2010) Under Hypotheses (H1) and (H4)-(H6), we have,*

$$\widehat{\Psi}_D(x) - 1 = o_P(1).$$

LEMMA **3** *Under hypotheses (H1), (H2), and (H4)-(H6) we have*

$$\left(\frac{n\phi(x,h)\beta_1^2}{\beta_2\lambda_2(x,\theta_x)}\right)^{1/2} B_n(x,z) = u + o(1), \text{ as } \quad n \to +\infty.$$

LEMMA **4** *Under hypotheses (H1), (H2), and (H4)-(H6) we have,*

$$\left(\frac{n\phi(x,h)\beta_1^2}{\beta_2\lambda_2(x,\theta_x)}\right)^{1/2} R_n(x,z) = o_P(1) \qquad a.co.$$

LEMMA **5** *Under Hypotheses (H1) and (H4)-(H6), $\widehat{\theta_x}$ exists a.s. for all sufficiently large n*

3.5 Some special cases

In this section we discus the generality of our study by comparing it to some popular case. More precisely, we consider three special cases such as the classical regression case, the independent case and the multivariate case which are respectively studied, Laib and Louani(2010), Attouch *et al.* (2009).

- *The classical regression case :* As noticed earlier the classical regression defined by conditional expectation is a particular case of our study with $\psi_x(Y,t) = (Y - t)$. So, for this particular case we obtain the following convergence rate.

  Corollaire 1 *Under the hypotheses (H1)-(H6), we have*

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

  where

  $$B_n(x) = h\Phi'(0,\theta_x)\frac{\beta_0}{\beta_1} + o(h) \text{ and } \sigma^2(x,\theta_x) = \frac{\beta_2(\mathbb{E}[Y^2|X=x] - \mathbb{E}^2[Y|X=x])}{\beta_1^2}$$

  Remark 1

  Clearly, this convergence rate is exactly what is obtained by Laib and Louani (2011) for the standard regression model.

- *The independent case :* In this situation, condition (H1(ii)) is automatically verified and for all $i = 1, \ldots n$ take $\phi_i(x,r) = \phi(x,r)$. Therefore, condition (H1) is restricted to $\phi(x,r) > 0$, for all $r > 0$. Thus, our Theorem leads to the next Corollary,

Corollaire 2 *Under assumptions (H1)-(H6) we have :*

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

Remark 2
We point out that in this case where the (X_i, Y_i) are independent, we obtain the same convergence rate given by Attouch *et al* . (2009).

– *The real case* As mentioned in the introduction on, in the real case when $\mathcal{F} = \mathrm{I\!R}$, and if the probability density of the random variable X (resp. the conditional density of X given $\mathfrak{F}_{i-1}$) denoted by f (resp. by $f_i^{\mathfrak{F}_{i-1}}$), is strictly positive and of $\mathcal{C}^1$ class, then $\phi(x,h) = f(x)h + o(h)$ and $\shortmid(X_i \in [x-h, x+h]|\mathfrak{f}_{i-1}) = f_i^{\mathfrak{f}_{i-1}}(x)h + o(h))$. Moreover the ergodic Theorem insure that

$$\|\frac{1}{n}\sum_{i=1}^{n} f_i^{f_{i-1}} - f\| \to 0,$$

where $\|.\|$ is a norm in sparable Banach space $\mathcal{C}^1$. Therefore condition (H1) is verified and Theorem 2 can be reformulated in the following way.

Corollaire 3 *Under assumptions (H2)-(H6) and if the probability density of the random variable X (resp. the conditional density of X given $\mathfrak{F}_{i-1}$) denoted by f (resp. by $f_i^{\mathfrak{f}_{i-1}}$), is strictly positive and of $\mathcal{C}^1$ class, we have :*

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

Remark 3
A similar thing can be concluded if $\mathcal{F} = \mathrm{I\!R}^p$. It is worth to noting that this last consistency is also new in vectorial statistic

3.6 Some applications

3.6.1 Conditional Confidence curve

The most important application of the asymptotic normality result is the building of confidence intervals for the true value of θ_x given curve $X = x$. However, the latter requires an estimation of the bias $B_n(x)$ term and the standard deviation $\sigma(x, \theta_x)$. For sake of shortness, we neglect the bias term and we estimate $\sigma(x,\theta_x)$ by plug-in method as follows. Indeed, if ψ_x is of class C^1, w.r.t the second component, the quantities $\lambda_2(x,\theta_x)$ and $\Gamma_1(x,\theta_x)$ can be estimated by

$$\widehat{\lambda_2}(x,\widehat{\theta_x}) = \frac{\sum_{i=1}^n K(h^{-1}d(x,X_i))\psi_x^2(Y_i,\widehat{\theta_x})}{\sum_{i=1}^n K(h^{-1}d(x,X_i))},$$

and

$$\widehat{\Gamma_1}(x,\widehat{\theta_x}) = \frac{\sum_{i=1}^n K(h^{-1}d(x,X_i))\frac{\partial}{\partial t}\psi_x(Y_i,\widehat{\theta_x})}{\sum_{i=1}^n K(h^{-1}d(x,X_i))}.$$

Furthermore, the quantities β_1 and β_2 can be estimated empirically by

$$\widehat{\beta_1} = \frac{1}{n\phi(x,h)}\sum_{i=1}^n K(h^{-1}d(x,X_i)) \quad \text{and} \quad \widehat{\beta_2} = \frac{1}{n\phi(x,h)}\sum_{i=1}^n K^2(h^{-1}d(x,X_i))$$

It follows that $\widehat{\sigma}(x,\widehat{\theta_x}) := \left(\frac{\widehat{\beta_2}\widehat{\lambda_2}(x,\widehat{\theta_x})}{(\widehat{\beta_1})^2\widehat{\Gamma_1}^2(x,\widehat{\theta_x})}\right)^{1/2}$. Thus, we get the following approximate $(1-\zeta)$ confidence interval for θ_x

$$\widehat{\theta_x} \pm t_{1-\zeta/2} \times \left(\frac{\widehat{\sigma}_n^2(x,\widehat{\theta_x})}{n\phi(x,h)}\right)^{1/2}$$

where $t_{1-\zeta/2}$ denotes the $1-\zeta/2$ quantile of the standard normal distribution. It should be to note that the function $\phi(x,\cdot)$ does not appear in the calculation of the confidence interval by simplification.

3.6.2 Smoothing parameter selection

As all nonparametric estimation with the smoothing technique, the choice of the smoothing parameters, plays a primordial role. The most celebrated criterium for the smoothing parameter choice is the mean squared error where the optimal one have to balance the bias and variance terms. However, in our robust approach, this L_2 error is not adequate if we suspect that outliers are present. In this context, the L_1 error is more adapted to this case. For both criteria, our asymptotic result is a basic ingredient to determine the leading term in these errors. Indeed, by using the same arguments as those used by Crambes et al. (2008) we show that

$$\mathbb{E}\left[\left|\widehat{\theta}_n - \theta_x\right|^q\right] = \mathbb{E}\left[\left|B_n(x) + \sqrt{\frac{\sigma(x,\theta_x)}{n\phi(x,h)}}W\right|^q\right] + o\left(\frac{1}{\sqrt{n\phi(x,h)}^q}\right)$$

where W is a standard gaussian variable. Therefore, an ideal theoretical rule permitting to obtain an optimal bandwidth, for both (robust or no robust case) is the minimization of the leading term :

$$\mathbb{E}\left[\left|B_n(x)+\sqrt{\frac{\sigma(x,\theta_x)}{n\phi(x,h)}}W\right|^q\right].$$

Of course that the practical utilization of this method requires some additional computational efforts. More precisely, it requires the estimation of the unknown quantities $B_n(x)$ and $\sigma(x,\theta_x)$. The latter is estimated by $\widehat{\sigma}(x,\widehat{\theta_x})$ while $B_n(x)$ can be estimated by the same fashion where the real function $\Phi(t,s)$ is treated as a real regression function with response variable $\Psi(X,t)-\Psi(x,t)$. In conclusion, we can say that, the practical utilization of the present approach is possible, but it needs the determination of a pilot estimator of the conditional density function.

3.6.3 The functional times series prediction

It is well known that one of the most application of the nonparametric functional data analysis in dependent data is the prediction of a future real characteristic of an continuous processus by taking into account the whole past continuously. Indeed, let $(Z_t)_{t\in[0,b[}$ be a continuous time real valued random process. From Z_t we may construct N functional random variables $(X_i)_{i=1,\dots,N}$ defined by :

$$\forall t\in[0,b[,\qquad X_i(t)=Z_{((i-1)b+t)/N}.$$

The prediction aims is to evaluate a real characteristic denoted Y given X_N. The definition (3.1) shows that the random variable $\widehat{\theta_{X_N}}$, defined by (3.2), is the best approximation of this characteristic with respect to the loss function $\rho_x(,t)=\int_0^t\psi_x(,s)ds$, where (3.2) is given by using the $N-1$ pairs of r.v $(X_i,G(X_{i+1}))_{i=1,\dots,N-1}$ with G is the function which describes this characteristic.

3.7 Application

The aim of this short section is to apply our method to some chemiometrical real data. More precisely, our main aim is to compare the sensitivity to outliers of the classical regression defined as the conditional expectation m (x) = $\mathbb{E}[Y|X = x]$ estimated by

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

and the robust regression θ_x associated to $\psi(t) = \dfrac{1}{\sqrt{1 + t^2/2}}$ in the presence of outliers. This score function known in the literature by $L_1 - L_2$'s function. This choice is motivated by the fact that the estimation with this function take both the advantage of the estimators to reduce the influence of large errors and that of estimators to be convex which are a fundamentals constraints of the robustness properties. For this purpose, we consider the autoregressive process, the latter represents a special case of ergodic data not verified the conditions of melanges, and we can not conclude the asymptotic properties of its data types from the work of Crambes et al. that require strong melanges condition. In this part we will test the efficiency of our procedure on this type of data. for this, one takes a sample of an autoregressive process of size 110 functional random variables discretized on 60 points.
The graphics division of 110 variables are represented in the following Figure :

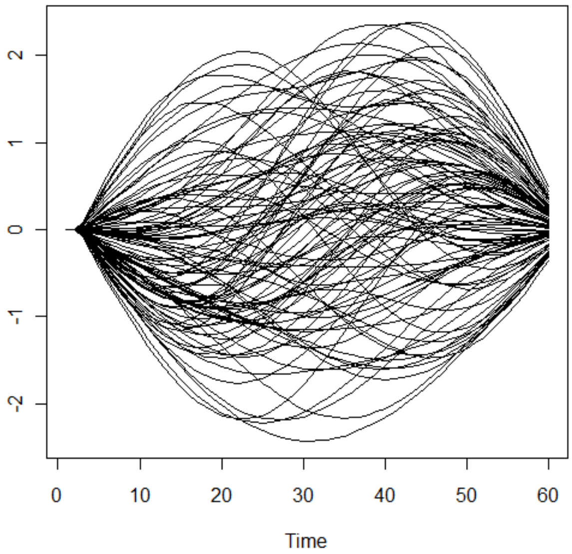

FIGURE 3.1 – 110 curves, $\{X_i(t)$, t $\in$ [1, 60], i = 1, ..., 110.$\}$

In order to introduce the outliers in this sample we multiply by 100 the response variable of a number of observations. In practice, we consider 110-observations splitted into two samples (learning sample (100 observations) and test sample (10 observations).

To explored the performance of our approach, we compare it to the classical regression in the both cases (presence or absence of outliers) and in the both cases we proceed by the following algorithm :

- *step 1.* We split our data into two subsets
  $(X_i, Y_j)_{j=1,...100}$ training sample,
  $(X_i, Y_i)_{i=101,...110}$ test sample.
- *step 2.* We compute the kernel estimator $\widehat{\theta_{X_j}}$, for all j by using the training sample.
- *step 3.* The error used to evaluate this comparison is the mean of absolute error (MAE) expressed by :

$$\frac{1}{10}\sum_{i=101}^{110}(Y_i - \widehat{T}(X_i))^2,$$

where $\widehat{T}$ designate the estimator used : Classical or robust regression.

The first illustration is given in the Fig 3.2 where we show that in the absence of outliers, the two methods are basically equivalent and both give the good behavior of our functional forecasting procedure.

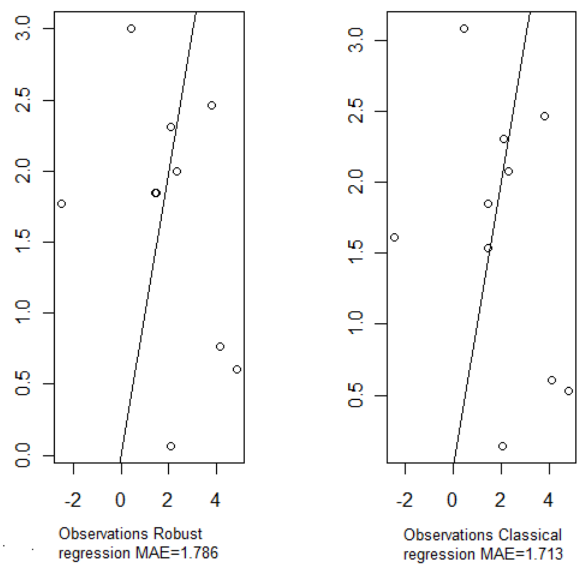

FIGURE 3.2 – Comparaison between the both methods in the absence of outliers.

The second illustration is given in the following table where we observe that in the presence of outliers, the robust regression gives better results than the classical method, in sense that, even if the MAE value of the both methods increases substantially relatively to the number of the perturbed points, but it remaining very low for the robust one.

The number of the perturbed observations	MAE.Robust Method	MAE. Classical Method
5 observations are perturbed	1.908	2.421
16 observations are perturbed	2.860	5.323
32 observations are perturbed	4.305	9.641

TABLE 3.1 – Comparaison between the both methods in the presence of outliers..

3.8 Appendix

Proof of Lemma 1 For all $i = 1, \dots, n$ we put $K_i(x) = K(h^{-1}d(x, X_i))$ and

$$\eta_{ni} = \left(\frac{\phi(x,h)\beta_1^2}{\beta_2\lambda_2(x,\theta_x)}\right)^{1/2} (\psi_x(Y_i, z) - \Psi(x, z))\frac{K_i(x)}{\mathbb{E}K_1(x)} \tag{3.2}$$

and define $\zeta_{ni} := \eta_{ni} - \mathbb{E}[\eta_{ni} \mid F_{i-1}]$. Under this consideration we write

$$\left(\frac{n\phi(x,h)\beta_1^2}{\beta_2\lambda_2(x,\theta_x)}\right)^{1/2} Q_n(x,z) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\zeta_{ni}$$

As ζ_{ni} is a triangular array of martingale differences according the σ-fields $(\mathfrak{F}_{i-1})_i$, we are in position to apply the central limit theorem based on unconditional Lindeberg condition (see, Gaenssler et al. (1978)). More precisely, we must to check the following condition

$$\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[\zeta_{ni}^2 \mid F_{i-1}] \to 1 \quad \text{in probability} \tag{3.3}$$

and

$$\text{for every} \quad \epsilon > 0 \qquad \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[\zeta_{ni}^2 \mathbb{1}_{\zeta_{ni}^2 > \epsilon n}] \to 0. \tag{3.4}$$

Firstly, we prove (3.3). To do that,we write

$$\mathbb{E}[\zeta_{ni}^2 \mid F_{i-1}] = \mathbb{E}[\eta_{ni}^2 \mid F_{i-1}] - \mathbb{E}^2[\eta_{ni} \mid F_{i-1}].$$

Therefore, it suffices to show the

$$\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}^2[\eta_{ni} \mid F_{i-1}] \to^{\mathbb{P}} 0 \tag{3.5}$$

and

$$\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[\eta_{ni}^2 \mid F_{i-1}] \to^{\mathbb{P}} 1. \tag{3.6}$$

For th first convergence we have

$$\begin{aligned} \mid \mathbb{E}[\eta_{ni} \mid F_{i-1}] \mid &= \frac{1}{\mathbb{E}K_1(x)}\left(\frac{\phi(x,h)\beta_1^2}{\beta_2\lambda_2(x,\theta_x)}\right)^{1/2} \mid \mathbb{E}[(\Psi(X_i,t) - \Psi(x,t))K_i(x) \mid F_{i-1}] \mid \\ &\leq \frac{1}{\mathbb{E}K_1(x)}\left(\left(\frac{\phi(x,h)\beta_1^2}{\beta_2\lambda_2(x,\theta_x)}\right)\right)^{1/2} \sup_{u\in B(x,h)} \mid \Psi(u,t) - \Psi(x,t) \mid \mathbb{E}[K_i(x) \mid F_{i-1}]. \end{aligned}$$

Obviuosly, under (H1) and (H5) we have

$$C\phi_i(x,h) \leq \mathbb{E}\left[K_i(x)|\mathfrak{F}_{i-1}\right] \leq C'\phi_i(x,h).$$

and

$$C\phi(x,h) \leq \mathbb{E}\left[\Delta_i(x)\right] \leq C'\phi(x,h).$$

On other hand condition (H2ii) implies that

$$\sup_{u\in B(x,h)} \mid \Psi(u,t) - \Psi(x,t) \mid = o(1).$$

Combining these lasts three results, we obtain

$$\begin{aligned}(\mid \mathbb{E}[\eta_{ni} \mid F_{i-1}])^2 &\leq \sup\nolimits_{u\in B(x,h)} \mid \Psi(u,t) - \Psi(x,t) \left(\tfrac{\beta_1^2}{\beta_2\lambda_2(x,\theta_x)}\right) \mid \tfrac{1}{\phi(x,h)}\phi_i^2(x,h)\\ &\leq \sup\nolimits_{u\in B(x,h)} \mid \Psi(u,t) - \Psi(x,t) \mid \left(\tfrac{\beta_1^2}{\beta_2\lambda_2(x,\theta_x)}\right) \tfrac{1}{\phi(x,h)}\phi_i(x,h).\end{aligned}$$

Finally, under the fact that (see, (H1ii))

$$\frac{1}{n\phi(x,h)}\sum_{i=1}^{n}\phi_i(x,h) \to^{\mathbb{P}} 1$$

we obtain that

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n}(\mathbb{E}[\eta_{ni} \mid F_{i-1}])^2 &= \sup_{u\in B(x,h)} \mid \Psi(u,t) - \Psi(x,t) \mid \left(\frac{\beta_1^2}{\beta_2\lambda_2(x,\theta_x)}\right)\left(\frac{1}{n\phi(x,h)}\sum_{i=1}^{n}\phi_i(x,h)\right)\\ &= o_p(1).\end{aligned}$$

Now, we treat the convergence (3.6), Indeed, we write

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[\eta_{ni}^2 \mid F_{i-1}] &= \frac{1}{n(\mathbb{E}K_1(x))^2}\left(\frac{\phi(x,h)\beta_1^2}{\beta_2\lambda_2(x,\theta_x)}\right)\sum_{i=1}^{n}\mathbb{E}[(\psi_x(Y_i,z) - \Psi(x,z))^2K_i^2(x) \mid F_{i-1}]\\ &= \frac{1}{n(\mathbb{E}K_1(x))^2}\left(\frac{\phi(x,h)\beta_1^2}{\beta_2\lambda_2(x,\theta_x)}\right)\left(\sum_{i=1}^{n}\mathbb{E}[\psi_x^2(Y_i,z)\Delta_i^2(x) \mid F_{i-1}]\right.\\ &\qquad -2\Psi(x,z)\sum_{i=1}^{n}\mathbb{E}[\psi_x(Y_i,z)\Delta_i^2(x) \mid F_{i-1}]\\ &\qquad \left.+\Psi^2(x,z)\sum\nolimits_{i=1}^{n}\mathbb{E}[\Delta_i^2(x) \mid F_{i-1}]\right).\end{aligned}$$

Denote

$$J_1 = \sum_{i=1}^{n} \mathbb{E}[(\psi_x^2(Y_i, z)\Delta_i^2(x) \mid F_{i-1}], \quad J_2 = \sum_{i=1}^{n} \mathbb{E}[(\psi_x(Y_i, z)\Delta_i^2(x) \mid F_{i-1}]$$

and

$$J_3 = \sum_{i=1}^{n} \mathbb{E}[\Delta_i^2(x) \mid F_{i-1}].$$

It is easily seen that

$$\begin{aligned} J_1 &= \lambda_2(x,z) \sum_{i=1}^{n} \mathbb{E}[(K_i^2(x)|\mathfrak{F}_{i-1}] + \sum_{i=1}^{n} \left[\mathbb{E}\left[\psi^2(Y_i,z)K_i^2(x)|\mathfrak{F}_{i-1}\right] - \lambda_2(x,z)\mathbb{E}\left[K_i^2(x)|\mathfrak{F}_{i-1}\right]\right] \\ &= \lambda_2(x,z) \sum_{i=1}^{n} \mathbb{E}[(K_i^2(x)|\mathfrak{F}_{i-1}] + \sum_{i=1}^{n} \left[\mathbb{E}\left[K_i^2(x)\mathbb{E}[\psi^2(Y_i,z)|\mathfrak{G}_{i-1}]|\mathfrak{F}_{i-1}\right] - \lambda_2(x,z)\mathbb{E}\left[K_i^2(x\right.\right. \\ &= \lambda_2(x,z) \sum_{i=1}^{n} \mathbb{E}[(K_i^2(x)|\mathfrak{F}_{i-1}] + \sum_{i=1}^{n} \left[\mathbb{E}\left[K_i^2(x)\mathbb{E}[\psi^2(Y_i,z)|X_i]|\mathfrak{F}_{i-1}\right] - \lambda_2(x,z)\mathbb{E}\left[K_i^2(x)|\mathfrak{F}\right.\right. \end{aligned}$$

Using the same arguments as those used in (3.5), to evaluate the second term. Then, we have,

$$\frac{1}{n\mathbb{E}\left[K_1(x)\right]} \sum_{i=1}^{n} \left[\mathbb{E}\left[K_i^2(x)\mathbb{E}[\psi^2(Y_i,z)|X_i]|\mathfrak{F}_{i-1}\right] - \lambda_2(x,z)\mathbb{E}\left[K_i^2(x)|\mathfrak{F}_{i-1}\right]\right]$$

$$\leq \sup_{u\in B(x,h)} \mid \lambda_2(x,z) - \lambda_2(x,z) \mid \left(\frac{1}{n\phi(x,h)} \sum_{i=1}^{n} \mathbb{P}(X_i \in B(x,h)|\mathfrak{F}_{i-1})\right).$$

Furthermore, we use the continuity of $\lambda_2(x,.)$, to write

$$\lambda_2(x,z) = \lambda_2(x,\theta_x) + o(1)$$

Thus,

$$\frac{1}{n\mathbb{E}\left[K_1(x)\right]} J_1 = \lambda_2(x,\theta_x) \frac{1}{n\mathbb{E}\left[K_1(x)\right]} \sum_{i=1}^{n} \mathbb{E}\left[K_i^2(x)|\mathfrak{F}_{i-1}\right] + o(1)$$

and by the same manner

$$\frac{1}{n\mathbb{E}\left[K_1(x)\right]} J_2 = \Psi_x(x,\theta_x) \frac{1}{n\mathbb{E}\left[K_1(x)\right]} \sum_{i=1}^{n} \mathbb{E}\left[K_i^2(x)|\mathfrak{F}_{i-1}\right] + o(1) = o(1)$$

Finally we have

$$\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[\eta_{ni}^2 \mid F_{i-1}] = \frac{1}{n(\mathbb{E}K_1(x))^2}\left(\frac{\phi(x,h)\beta_1^2}{\beta_2}\right)\sum_{i=1}^{n}\mathbb{E}\left[(K_i^2(x)|\mathfrak{F}_{i-1}\right] + o(1)$$

Next, we use the same ideas used in Ferraty et al. (2009) to get

$$\mathbb{E}\left[K_i^2(x)|\mathfrak{F}_{i-1}\right] = K^2(1)\phi_i(x,h) - \int_0^1 (K^2(u))'\phi_i(x,uh)du$$

and

$$\mathbb{E}\left[K_1(x)\right] = K(1)\phi(x,h) - \int_0^1 (K(u))'\phi(x,uh)du$$

it follows that

$$\begin{aligned}\frac{1}{n\phi(x,h)}\sum_{i=1}^{n}\mathbb{E}\left[(K_i^2(x)|\mathfrak{F}_{i-1}\right] &= \frac{K^2(1)}{n\phi(x,h)}\sum_{i=1}^{n}\phi_i(x,h)\\ &\quad - \int_0^1 (K^2(u))'\frac{\phi(x,uh)}{n\phi(x,h)\phi(x,uh)}\sum_{i=1}^{n}\phi_i(x,uh)du\\ &\quad = K^2(1) - \int_0^1 (K^2(u))'\tau_x(u)du + o_p(1) = \beta_2 + o_p(1).\end{aligned}$$

and

$$\frac{1}{n\phi(x,h)}\mathbb{E}\left[K_1(x)\right] = \beta_1 + o(1).$$

We deduce that

$$\lim_{n\mapsto\infty}\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[\eta_{ni}^2 \mid F_{i-1}] = 1$$

which completes the proof of (3.3).
Concerning (3.4), we write

$$\zeta_{ni}^2 \mathbb{1}_{\zeta_{ni}^2 > \epsilon n} \leq \frac{|\zeta_{ni}|^{2+\delta}}{\sqrt{(\epsilon n)^\delta}} \quad \text{for every} \quad \delta > 0$$

Observe that

$$\begin{aligned}\mathbb{E}[\zeta_{ni}^{2+\delta}] &= \mathbb{E}\left[\left|\eta_{ni}(x) - \mathbb{E}[\eta_{ni} \mid F_{i-1}]\right|^{2+\delta}\right]\\ &\leq 2^{1+\delta}\mathbb{E}\left[|\eta_{ni}(x)|^{2+\delta}\right] + 2^{1+\delta}|\mathbb{E}\left[\mathbb{E}[\eta_{ni} \mid F_{i-1}]\right]^{2+\delta}\right]|\end{aligned}$$

By the Jensen inequality we obtain

$$\mathbb{E}[\zeta_{ni}^{2+\delta}] \leq C\mathbb{E}\left[|\eta_{ni}(x)|^{2+\delta}\right].$$

So, it remains to evaluate $\mathbb{E}\left[|\eta_{ni}(x)|^{2+\delta}\right]$. For this, once again we use the C_r-inequality

$$\mathbb{E}\left[|\eta_{ni}(x)|^{2+K}\right] \leq C\left(\frac{\phi(x,h)\beta_1^2}{\beta_2\lambda_2(x,\theta_x)\mathbb{E}^2\left[K_1\right]}\right)^{1+\delta/2} \mathbb{E}\left[K_i^{2+\delta}(x)\psi^{2+\delta}(Y_i,t)\right]+\Psi^{2+\delta}(x,z)\mathbb{E}\left[K_i^{2+\delta}\right]$$

We conditione by X_i, and we use the fact that

$$\mathbb{E}\left[\psi^{2+\delta}(Y_i,t) \mid X_i\right] < \infty.$$

It follows that

$$\begin{aligned}\mathbb{E}\left[|\eta_{ni}(x)|^{2+\delta}\right] &\leq C\left(\frac{1}{\phi(x,h)}\right)^{1+\delta/2} \mathbb{E}([K_i(x)]^{2+\delta}) \\ &\leq C\left(\frac{1}{\phi(x,h)}\right)^{\delta/2} \qquad (3.7)\end{aligned}$$

Consequently

$$\frac{1}{n}\sum_{i=1}^{n} \mathbb{E}[\zeta_{ni}^2 \mathbb{1}_{\zeta_{ni}^2>\epsilon n}] \leq C\left(\frac{1}{n\phi(x,h)}\right)^{\delta/2} \to 0$$

which completes the proof of the Lemma . ■

Proof of Lemma 3

By a simple manipulation we have

$$\begin{aligned}\frac{\bar{\Psi}_N(x,z)}{\bar{\Psi}_D(x)} &= \frac{1}{\sum_{i=1}^n \mathbb{E}\left[K_i \mid F_{i-1}\right]}\sum_{i=1}^{n} \mathbb{E}\left[K_i\left[\mathbb{E}[\psi_x(Y,z)|X_1] - \mathbb{E}[\psi_x(Y,z)|X=x]\right] \mid F_{i-1}\right]\right] \\ &\quad +\mathbb{E}[\psi_x(Y,z)|X=x] - \mathbb{E}[\psi_x(Y,\theta(x))|X=x] =: I_1 + I_2. \qquad (3.8)\end{aligned}$$

For $I_1(x)$ we use the same ideas asi in Ferraty *et al.* (2007), we obtain under (H2iii)

$$\begin{aligned}A_i &= \mathbb{E}\left[K_i\left[\mathbb{E}[\psi_x(Y,z)|X_i] - \mathbb{E}[\psi_x(Y,z)|X=x]\right] \mid F_{i-1}\right] \\ &= \mathbb{E}\left[K_i\left[\mathbb{E}[\Psi(X_i,z) - \Psi(x,z)|d(x,X_i)\right] \mid F_{i-1}\right]\right]\right] \\ &= \mathbb{E}\left[K_i\Phi(d(x,X_i),z) \mid F_{i-1}\right]\right] \\ &= \int \Phi(th,z)K(t)dP^{F_{i-1}}(th) \\ &= h\Phi'(0,z)\int tK(t)dP^{F_{i-1}}(th). \qquad (3.9)\end{aligned}$$

Using the continuity of $\Phi'(0,\cdot)$ and the fact that

$$\int tK(t)dP^{F_{i-1}}(th) = K(1)\phi_i(x,h) - \int_0^1 (sK(s))'\phi_i(x,sh)ds$$

to obtain that

$$\frac{1}{n}\sum_{i=1}^{n} A_i = h\Phi'(0,\theta_x)\left(K(1) - \int_0^1 (sK(s))'\tau_x(s)ds\right) + o_p(h).$$

Similarly, we have

$$\frac{1}{n}\sum_{i=1}^{n} \mathbb{E}\left[K_i \mid F_{i-1}\right]] = \left(K(1) - \int_0^1 K'(s)\tau_x(s)ds\right) + o_p(1).$$

At last,

$$I_1 = B_n(x) + o(h).$$

Concerning I_2 we use a Taylor expansion to get, under (H2)

$$I_2 = -B_n(x) + u\left[n\phi(x,h)\right]^{-1/2}\sigma(x,\theta_x)\frac{\partial}{\partial t}\Psi(x,\theta_x) + o\left(\left[n\phi(x,h)\right]^{-1/2}\right).$$

The result is then a consequence of the decomposition (3.8).

■

Proof of Lemma 4 Clearly, it suffices to show that

$$\frac{\bar{\Psi}_N(x,t) - \Psi(x,t)\bar{\Psi}_D(x)}{\bar{\Psi}_D(x)} = o_p(1)$$

and

$$|\widehat{\Psi}_N(x,t) - \bar{\Psi}_N(x,t)| = o_p(1).$$

On the one hand

$$\begin{aligned}\frac{\bar{\Psi}_N(x,t) - \Psi(x,t)\bar{\Psi}_D(x)}{\bar{\Psi}_D(x)} &= \frac{1}{n\mathbb{E}\left[K_1(x)\right]\bar{\Psi}_D(x)}\sum_{i=1}^{n}\left[\mathbb{E}\left[K_i(x)\mathbb{E}[\psi(Y_i,t)|\mathfrak{G}_{i-1}]|\mathfrak{F}_{i-1}\right] - \Psi(x,t)\mathbb{E}\left[K_i(x)\right.\right.\\ &= \frac{1}{n\mathbb{E}\left[K_1(x)\right]\bar{\Psi}_D(x)}\sum_{i=1}^{n}\left[\mathbb{E}\left[K_i(x)\mathbb{E}[\psi(Y_i,t)|X_i]|\mathfrak{F}_{i-1}\right] - \Psi(x,t)\mathbb{E}\left[K_i(x)|\mathfrak{F}\right.\right.\\ &\leq \frac{1}{n\mathbb{E}\left[K_1(x)\right]\bar{\Psi}_D(x)}\sum_{i=1}^{n}\left[\mathbb{E}\left[K_i(x)|\Psi(X_i,t) - \Psi(x,t)|\mathfrak{F}_{i-1}\right]\right]\end{aligned}$$

Finally by (H1ii), we deduce that

$$\left| \frac{\bar{\Psi}_N(x,t) - \Psi(x,t)\bar{\Psi}_D(x)}{\bar{\Psi}_D(x)} \right| \leq \sup_{x' \in B(x,h)} |\Psi(x',t) - \Psi(x,t)| \to 0.$$

On the other hand, the convergence

$$\widehat{\Psi}_N(x,z) - \bar{\Psi}_N(x,z) = o_p(1)$$

will be established by showing the following two results

$$\mathbb{E}\left[\widehat{\Psi}_N(x,z) - \bar{\Psi}_N(x,z)\right] \to 0$$

and

$$Var\left[\widehat{\Psi}_N(x,z) - \bar{\Psi}_N(x,z)\right] \to 0.$$

The first one is a consequence of the definitions of $\widehat{\Psi}_N(x,z)$ and $\bar{\Psi}_N(x,z)$. Next, for the second one, we have

$$\widehat{\Psi}_N(x,z) - \bar{\Psi}_N(x,z) = \sum_{i=1}^{n} \Delta_i(x,z)$$

where

$\Delta_i(x,z) = \dfrac{1}{n\mathbb{E}[K_1]} K_i\psi(Y_i,z) - \mathbb{E}\left[K_i\psi(Y_i,z)|\mathfrak{F}_{i-1}\right]$. By Burkholder 's inequality, we have

$$\mathbb{E}\left[\sum_{i=1}^{n} \Delta_i(x,z)\right]^2 \leq \sum_{i=1}^{n} \mathbb{E}\left[\Delta_i(x,z)\right]^2.$$

Furthermore, by Jensen's inequality we show that

$$\mathbb{E}\left[\Delta_i(x,z)\right]^2 \leq \frac{1}{n^2\mathbb{E}^2[K_1]}\mathbb{E}\left[K_i^2\psi^2(Y_i,z)\right] \leq \frac{1}{n^2\mathbb{E}^2[K_1]}\mathbb{E}\left[K_i^2\right] \leq \frac{1}{n\phi^2(x,h)}\phi_i(x,h)$$

Hence, (H1) gives

$$Var\left[\widehat{\Psi}_N(x,z) - \bar{\Psi}_N(x,z)\right] \to 0$$

■

Proof of the Lemma 5 It is clear that, the monotony of $\psi_x(Y,\cdot)$, for all $\epsilon > 0$

$$\Psi(x,\theta_x - \epsilon) \leq \Psi(x,\theta_x) \leq \Psi(x,\theta_x + \epsilon).$$

By using a similar argument as those used in the previous Lemmas we show that

$$\widehat{\Psi}(x,t) \longrightarrow \Psi(x,t) \text{ in probability}$$

for all real fixed $t \in N_x$. So, for sufficiently large n and for all ϵ small enough

$$\widehat{\Psi}(x,\theta_x - \epsilon) \leq 0 \leq \widehat{\Psi}(x,\theta_x + \epsilon) \quad \text{holds with probability tending to 1.}$$

Since ψ_x is continuous function, then as $\widehat{\Psi}(x,t)$ is a continuous function of t, there exists a $\widehat{\theta_x} \in [\theta_x - \epsilon, \theta_x + \epsilon]$ such that $\widehat{\Psi}(x,\widehat{\theta_x}) = 0$. Finally, the uniqueness of $\widehat{\theta_x}$ is a direct consequence of the strict monotonicity of ψ_x, w.r.t. the second component, and the fact that

$$\mathbb{P}\left(\sum_{i=1}^{n} K_i = 0\right) = \mathbb{P}\left(\widehat{\Psi}_D(x) = 0\right) \to 0 \quad \text{as} \quad n \to \infty.$$

which imply $(\sum_{i=1}^{n} K_i \neq 0)$ with probability tending to 1. Moreover, as $\widehat{\theta_x} \in [\theta_x - \epsilon, \theta_x + \epsilon]$ in probability, then

$$\widehat{\theta_x} \to \theta_x \quad \text{in probability as } n \longrightarrow \infty$$

■

Bibliographie

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

Attouch, M. Laksaci, A. and Ould Saïd E. (2010). Asymptotic normality of a robust estimator of the regression function for functional time series data. *J. of the Korean Statist. Soc.*, **39**, 489-500.

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

Boente, G. and Fraiman, R. (1989). Nonparametric regression estimation. *J. Multivariate Anal.* **29**, 180-198.

Bosq, D. (2000). Linear processs in function spaces. Théorie and Application. *Lectures Notes in Statistics. Springer Verlag, New York.* **149**.

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

Davidian, M. Lin, X. and Wang, J.L. (2004). Introduction. *Statist. Sinica* **14**, 613–614.

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

Ferraty, F. (2010). High-dimensional data : a fascinating statistical challenge. *J. Multivariate Anal.* **101**, 305–306.

Ferraty, F. and Romain, Y. (2011). *The Oxford handbook of functional data analysis.* Oxford University Press.

Gaenssler, P. and Haeusler, E., 1986. *On martingale central limit theory,* in Eberlein, E., Taqqu, M. S. (Eds), Dependence in Probability and Statistics, Birkhauser, Boston, pp. 303-334.

Gonzalez Manteiga, W. and Vieu, P. (2007). Statistics for functional data. *Comput. Statist. Data Anal.* **51**, 4788–4792.

Györfi, L., Härdle, W., Sarda, P. and Vieu, P. (1989). *Nonparametric curve estimation for time series.* Lecture Notes in Statistics, **60**, Springer-Verlag.

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

Laksaci, A. Lemdani, M. and Ould-Saïd, E. (2009). L^1-norm kernel estimator of conditional quantile for functional regressors : consistency and asymptotic normality. *Statist. Probab. Lett.* **79**, 1065-1073.

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

Laïb, N. and Louani, D. (2010). Nonparametric kernel reghression estimate for functional stationary ergodic data : Asymptotic properties. *J. Multivariate Anal.*, **101**, 2266-2281.

Laïb, N. and Louani, D. (2011). Strong consistency of the regression function estimator for functional stationary ergodic data. *J. Statist. Plan. and Inference,* **141**, 359-372.

Ramsay, J. O. and Silverman, B. W. (2005). *Functional Data Analysis*, *2nd* Edition, Springer, New-York.

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

Valderrama, M. J. (2007). An overview to modelling functional data.*Comput. Statist.* **22**, 331–334.

Chapitre 4

Conclusion

Le travail de recherche développé dans cette thèse est la modélisation de la régression robuste non paramétrique quand les données sont ergodiques.

Dans un premier temps nous avons présenté, en détails les résultats obtenus en statistique non-paramétrique quand les données sont fonctionnelles.

Dans la deuxième partie, nous avons considéré une approche pour l'estimation de la régression robuste quand les données sont ergodiques. Le paramètre $\theta(x)$, appelé ψ_x-régression dans Laïb et Ould-Saïd (2000), et est une généralisation de la fonction de régression classique. Comme résultats asymptotiques nous avons établi la convergence presque complète (avec taux), ces résultats asymptotiques sont obtenus sous des conditions d'ergodicité en statistique non-paramétrique fonctionnelle. Sous ces même conditions, on a étudier la normalité asymptotique du même estimateur (cf. le chapitre 3) ; cette propriété asymptotique permettra de déterminer les intervalles de confiance, le choix du paramètre de lissage et la convergence des moments.