



REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE
MINISTÈRE DE L'ENSEIGNEMENT SUPÉRIEUR ET DE LA
RECHERCHE SCIENTIFIQUE

UNIVERSITÉ DJILLALI LYABES
FACULTÉ DES SCIENCES

THÈSE DE DOCTORAT 3ÈME CYCLE
EN : MATHÉMATIQUES
SPÉCIALITÉ : PROBABILITÉS ET STATISTIQUES

Par

Saliha DERRAR

Titre

*Estimation Robuste De La
Fonction De Régression Pour Un
Modèle Tronqué Aléatoirement À
Gauche À Co-variables
Fonctionnelles*

Soutenu publiquement le .. / 11 / 2016, devant le jury composé de :

M	E. OULD SAÏD	MCF-HDR	Univ. Littoral, France	Encadreur
M	A. LAKSACI	Professeur	Univ. Sidi Bel-Abbès	Co-Encadreur
M	A. GHERIBALLAH	Professeur	Univ. Sidi Bel-Abbès	Président
M	A. TATACHAK	Professeur	U.S.T.H.B	Examineur
M	M. ATTOUCH	Maître de Conférence	Univ. Sidi Bel-Abbès	Examineur

*À mes parents, à mon époux
et à ma chère petite Tasnim.*

Remerciements

Mes premiers remerciements vont naturellement à mes directeurs de thèse les Professeurs **Elies Oul Saïd** et **Ali Laksaci** pour leur encadrement et leurs encouragements durant toute la période de la réalisation de ce travail malgré toutes leurs occupations. Merci pour leur disponibilité, leur grande patience et leur convivialité tout au long de ces années.

Je remercie vivement Monsieur le Professeur **Abdelkader Gheriballah** de me faire l'honneur de présider ce jury, malgré ses nombreuses responsabilités.

Je remercie chaleureusement Monsieur **Abdelkader Tatachak** de l'université d'Alger (U.S.T.H.B) pour l'intérêt qu'il a porté à mon travail en acceptant de faire partie du jury.

Je tiens à exprimer ma sincère reconnaissance à Monsieur **Mohammed Kadi Attouch** pour l'honneur qu'il m'a fait en acceptant d'être examinateur de mon travail.

A titre plus personnel, un grand merci aux membres de ma famille, plus particulièrement mes parents qui ont toujours cru en moi, m'ont constamment encouragée et soutenue tout au long de mes années d'études. Je ne saurai passer sous silence l'apport inestimable des autres membres de ma famille (frères, beaux parents, etc.) qui m'ont soutenue, de près ou de loin durant mes études doctorales.

Les mots me manquent pour remercier, à sa juste valeur mon époux, pour ses encouragements et son soutien perpétuel et réconfortant, je ne pourrais finir mes remerciements sans une pensée très personnelle à ma chère fille pour laquelle je dédie ce travail. Merci ma famille, je n'aurais rien fait de tout cela sans votre amour.

Je ne citerai pas, mais ils se reconnaîtront, tous mes amies et collègues, qui m'ont soutenu. Je leur en suis reconnaissante.

Résumé

La régression robuste est une approche statistique sur la régression qui admet un comportement insensible aux observations aberrantes (atypiques, données tronquées). Nous nous intéressons plus particulièrement dans cette thèse à l'étude de l'estimation robuste de la fonction de régression dans laquelle la variable réponse est tronquée à gauche par une autre variable aléatoire tandis que la variable explicative est fonctionnelle, c'est à dire à valeurs dans un espace de dimension infinie. Notre étude porte sur des données indépendantes identiquement distribuées (i.i.d.) ainsi que sur des données fortement mélangeantes. Nous rappelons certains résultats établis sur le comportement asymptotique de quelques estimateurs non paramétriques robustes de la fonction de régression, dans le cas où les données sont complètement observées, dans le cas de données incomplètes et dans le cas où les co-variables appartiennent à un espace de dimension finie ou infinie.

Dans un premier temps, nous supposons que les variables sont i.i.d., et nous présentons des travaux sur la M-estimation de la fonction de régression dans le cas de données tronquées aléatoirement à gauche et dans la présence de co-variables fonctionnelles. Nous illustrons les résultats obtenus (consistance et normalité asymptotique) par des simulations pour montrer les performances des estimateurs proposés.

Dans un second temps, nous supposons que les variables sont fortement mélangeantes et nous établissons la convergence presque complète avec taux de convergence de notre estimateur .

Autant que l'on sache le problème de l'estimation non paramétrique robuste de la régression pour des données fonctionnelles dans la présence de données incomplètes n'a jamais été abordé, d'où l'originalité et la nouveauté de l'étude menée dans cette thèse.

Mots-clés : *Normalité asymptotique , données fonctionnelles, estimateur à noyau, estimateur de Lynden-Bell , estimation robuste, probabilités de petites boules, convergence presque complète, données tronquées.*

Abstract

The robust regression is a statistical approach of regression which admits an insensitive behavior to outliers (atypical , truncated data). We are interested especially in this thesis to study the robust estimate of the regression function in which the response variable is left truncated by another random variable and the explanatory variable is functional i.e. with values in an infinite dimensional space.

Our study focuses on independent identically distributed data (i.i.d.) and on strongly mixing data. We recall some results established on asymptotic propriety of some robust nonparametric estimators of the regression function in the case of completely observed data, and for incomplete data also in the case where the covariates are in a finite-dimensional space or infinite-dimensional space.

Firstly, we assume that the variables are i.i.d. , and we present our results on the M- estimation of the regression function in the case of left-truncated random data and in the presence of functional covariates. We illustrate the results (consistency and asymptotic normality) by simulations to show the performance of the proposed estimator.

Secondly , we suppose that the variables are strongly mixing and we establish the almost complete convergence with rates of convergence of our estimator.

As far as we know the problem of nonparametric robust estimation of regression for functional data in the presence of incomplete data was never been approached , hence the originality and novelty of the study conducted in this thesis.

key-words : *Asymptotic normality, functional data, Kernel estimator, Lynden-Bell estimator, robust estimation, small balls probability, almost complete convergence , truncated data.*

Liste des travaux

Publications

1. Derrar, S., A. Laksaci and Ould Saïd, E. (2015). On the nonparametric estimation of the functional ψ -regression for a random left-truncation model. *Journal of statistical Theory and Practice*, **9**, 823-849.
2. Derrar, S., A. Laksaci and Ould Saïd, E. M -estimation of the regression function under random left truncation and functional time series model. *Eletronic Journal of statistics*. Soumis

Communications internationales

1. Derrar, S., A. Laksaci and Ould Saïd, E. On the nonparametric estimation of the functional ψ -regression for a random left-truncation model. Colloque Internationale de statistiques et ses applications (ICSA2015) du 06 au 07 Mai 2015, Saida, Algérie.
2. Derrar, S., A. Laksaci and Ould Saïd, E. On the nonparametric estimation of the functional ψ -regression for a random left-truncation model. Journée de statistique, le 10 avril 2016, Biskra, Algérie.
3. Derrar, S., A. Laksaci and Ould Saïd, E. On the nonparametric estimation of the functional ψ -regression for a random left-truncation model. Conférence internationale de probabilité et Statistique (MICPS-2016) du 25 au 28 avril, Merrakech, Maroc.

Table des matières

1	Introduction	8
1.1	Statistique non paramétrique fonctionnelle pour des données complètes	8
1.2	Données incomplètes	9
1.2.1	Durée de vie	9
1.2.2	Censure	10
1.2.3	troncature	10
1.2.4	Troncature à gauche	10
1.2.5	Troncature à droite	11
1.3	la statistique non paramétrique robuste pour des données incomplètes : Motivations	11
1.4	Estimation de la fonction de régression	13
1.5	Description de la thèse	14
1.6	Présentation des résultats obtenus	15
1.6.1	Notations	15
2	Estimation nonparamétrique robuste	22
2.1	Les M—estimateurs	22
2.2	Exemples de M—estimateurs	23
2.2.1	Estimateur biweight	23
2.2.2	Estimateur de Huber	23
2.2.3	Estimateur de Andrew	23
2.2.4	Le choix de la fonction objective	23
2.3	La régression robuste : Cas de dimension finie	24
2.3.1	Pour des données complètes	24
2.3.2	Pour des données incomplètes	24
2.4	La régression robuste : Cas de dimension infinie	26
2.4.1	Pour des données complètes	26
2.4.2	Pour des données incomplètes	28

3	Convergence presque complète et normalité asymptotique : Cas iid	29
3.1	Introduction	30
3.2	Presentation of model and estimators	32
3.3	Main results	34
3.3.1	Consistency	35
3.3.2	Asymptotic normality	36
3.4	Special cases	37
3.4.1	Comeback to complete data	37
3.4.2	Comeback to classical regression case	37
3.4.3	Comeback to real case	38
3.5	Simulation study	38
3.5.1	Consistency	40
3.5.2	Asymptotic normality	47
3.6	Appendix	49
4	Convergence presque complète : Cas α-mélange	65
4.1	Introduction	66
4.2	Construction of the estimate and background for truncation model . .	69
4.3	Hypotheses and results	70
4.4	Appendix	73
5	Conclusion et perspectives	89
5.1	Conclusion	89
5.2	Perspectives	90

Chapitre 1

Introduction

1.1 Statistique non paramétrique fonctionnelle pour des données complètes

La statistique pour des données fonctionnelles complètes ou analyse des données fonctionnelles étudie des observations qui sont des courbes aléatoires ou des surfaces. L'utilisation de ce type de données est devenue très fréquente avec l'automatisation et l'informatisation des procédures de mesure et leurs capacités de stockage. En conséquence, la littérature consacrée à l'étude de données fonctionnelle s'est considérablement développée, les courbes de croissance, les enregistrements sonores, les images satellites, les séries chronologiques, les courbes spectrométriques ne sont que quelques exemples illustrant le grand nombre et la diversité des données de nature fonctionnelle. Historiquement, les premières contributions sur la modélisation par la régression ont été consacrées à l'étude des modèles paramétriques (voir les monographies de Ramsay et Silverman (1997, 2002, 2005) pour le cas i.i.d ou Bosq (2000) pour le cas dépendant). Cardot *et al.* (1999) ont construit un estimateur pour le modèle de régression linéaire Hilbertien en utilisant l'analyse en composantes principales fonctionnelles. Cet estimateur est construit à l'aide des propriétés spectrales de la version empirique de l'opérateur de variance—covariance de la variable explicative fonctionnelle. En ce qui concerne le cas non paramétrique et dans le cas des espaces de Banach on peut citer Lecoutre (1990). par la suite, il y a les travaux de Ferraty et Vieu (2000), étendus en (2002) au cas non standard de la régression telle la prévision en séries chronologiques. En utilisant le développement récent de la théorie des probabilités des petites boules, Ferraty et Vieu (2003) ont obtenu des résultats très fins et qui offrent une solution originale pour le problème de fléau de la dimension. Niang et Rhomari (2004) ont utilisé cet estimateur de la régression pour établir la convergence en norme L_p , une application à la discrimination des courbes a été proposée par ces auteurs. Masry (2005), a étudié la normalité asymptotique de l'esti-

mateur de la régression dans le cas dépendant. L'ouvrage de Ferraty et Vieu (2006) constitue une contribution de référence en la matière et fournit de plus amples détails concernant le comportement asymptotique des estimateurs à noyau pour plusieurs modèles fonctionnels à savoir la régression, le mode (avec ou sans conditionnement), la médiane (avec ou sans conditionnement) ainsi que les quantiles conditionnels. Notons, que cet ouvrage contient aussi de nombreux exemples de problèmes concrets pour lesquels les données sont de nature fonctionnelle. Nous renvoyons à Ferraty *et al.* (2007) pour des résultats précurseurs sur la vitesse de convergence asymptotique exacte en norme L_2 . En utilisant ce résultat, Rachdi et Vieu (2007) construisent un critère de choix automatique du paramètre de lissage pour l'estimateur à noyau de la fonction de régression à co-variables fonctionnelles. Récemment, Delsol (2007) généralise au cas dépendant les résultats de Ferraty *et al.* (2007) en donnant les termes asymptotiques dominants des moments centrés et des erreurs L_q de l'estimateur à noyau de la fonction de régression. Benhenni *et al.* (2007) ont établi la consistance de l'estimateur de la fonction de régression pour des co-variables fonctionnelles sous conditions de longue mémoire. Tandis que Burba *et al.* (2008) se sont intéressés à l'estimation de la régression via la méthode des k plus proches voisins (kNN), ils ont obtenu la convergence presque complète de l'estimateur construit. Un des premiers articles sur les modèles non paramétriques conditionnels avec variables explicatives fonctionnelles fut celui de Ferraty *et al.* (2006). Nous renvoyons à Ferraty *et al.* (2005) pour l'utilisation de ces modèles conditionnels fonctionnels à la prévision des séries temporelles fonctionnelles. Ezzahrioui et Ould-Saïd (2006, 2008a, 2008b, 2008c) et Niang et Laksaci (2007, 2008) ont une contribution déterminante dans ce domaine. Le lecteur pourra également se rapporter au travail de Manteiga et Vieu (2007). Et la contribution de Ferraty *et al.* (2014) pour une bibliographie récente et exhaustive en la matière. Récemment, des estimateurs robustes ont été considérés pour des modèles avec données fonctionnelles. nous développerons ce point dans le paragraphe 3.

1.2 Données incomplètes

1.2.1 Durée de vie

Un modèle de survie est un modèle basé sur des durées de vie. Une durée de vie est une variable aléatoire (va) souvent positive, précisément c'est le temps nécessaire de passer d'un état A à un état B. Le terme de durée de vie est employé de manière générale pour désigner le temps qui s'écoule jusqu'à la survenue d'un événement particulier qui n'est pas forcément la mort : par exemple, le délai de rémission d'une maladie (temps écoulé entre le début du traitement et la rechute), le délai de guérison (temps qui sépare le diagnostic de la guérison) ou encore le délai de panne (durée de fonctionnement d'une machine, en fiabilité) etc... Lorsque les durées de vie sont observées dans leur totalité,

on parle de données complètes, dans le cas contraire, les données sont incomplètes et nécessitent un traitement statistique particulier, si la durée de vie d'un individu n'est observable que dans une période donnée, on parle de censure. Il s'agit de troncature, si l'individu doit survivre longtemps pour être observé. La littérature est beaucoup plus riche en ce qui concerne la censure que la troncature qui est plus récente. Dans cette thèse, nous nous intéressons particulièrement, à la troncature gauche qui est le cadre dans le quel nous avons apporté une certaine contribution.

1.2.2 Censure

Définition 1.2.1 *Soit Y la variable d'intérêt, si au lieu d'observer Y on observe une constante C et si on sait que $Y > C$, on dira alors qu'il y a censure à droite.*

Il arrive souvent que la constante de censure soit elle même une variable aléatoire, dans ce cas, on parle de censure aléatoire à droite et on a :

Définition 1.2.2 *Etant donnée une variable aléatoire Y , on dit qu'il y a censure aléatoire à droite s'il existe une variable aléatoire C , telle qu'au lieu d'observer Y on observe :*

$$\{(Z, \delta) \text{ où } Z = \min(Y, C), \delta = \mathbb{1}_{Y \leq C}\}$$

1.2.3 troncature

Le modèle de troncature est apparu tout d'abord en astronomie, mais il est observé dans plusieurs domaines comme la médecine, l'épidémiologie, la biométrie et l'économie. La recherche d'objets cachés qui devront être assez grand pour être détectés, comme les réserves de pétrole est un champ d'application pour les données tronquées. De plus, les enquêtes de suivi médical où la troncature gauche peut apparaître si le temps d'origine de la durée de vie précède le temps d'origine de l'étude.

Définition 1.2.3 *Soit Y la variable d'intérêt et T une autre variable aléatoire, si Y et T sont observables seulement si $Y > T$, et sinon rien n'est observable, on dira que la variable Y est aléatoirement tronquée à gauche par la variable de troncature T . Dans le cas contraire, on parle de troncature aléatoire à droite.*

1.2.4 Troncature à gauche

En cas de troncature gauche, nous ne sommes capable d'observer que les durées de vie Y pour les quelles $Y \geq T$; ici T est la variable de troncature. Un exemple de troncature à gauche concerne la durée de vie des retraités. On enregistre l'âge au décès et l'âge d'entrée au centre. Un individu donné, doit vivre suffisamment longtemps pour être parmi la communauté du centre de retraite, et s'il décède avant, il y a troncature à

gauche. La durée de vie d'une population est étudiée à partir d'une cohorte tirée au sort dans cette population, seule la survie des sujets vivants à l'inclusion pourra être étudiée (il y a troncature à gauche car seuls les sujets ayant survécu jusqu'à la date d'inclusion dans la cohorte sont observables).

Dans ce cas, nous disposons d'un échantillon de taille n , dont la variable d'intérêt Y est observable, cet échantillon est extrait d'un échantillon de plus grande taille N inconnue.

La proportion de la population avec laquelle nous pouvons disposer d'une observation joue un rôle important en estimation sous troncature. Cette probabilité notée $\tau = P(Y \geq T)$ pourrait être estimée par la quantité $\frac{n}{N}$ mais malheureusement cet estimateur ne peut pas être calculé, car N est inconnue et la taille n de l'échantillon observé est elle-même une variable aléatoire de loi binomiale $B(N, \tau)$. En utilisant des estimateurs produit-limites de Lynden-Bell (1971), He et Yang (1998) donnent un estimateur calculable de τ . Ainsi que des résultats de convergence asymptotique dont la normalité. Des résultats asymptotiques ont été obtenus pour certains estimateurs et prédicteurs pour des données incomplètes (sous troncature ou censure). Rappelons les travaux de Ould Saïd et Lemdani (2006) pour la fonction de régression sous troncature, Ould Saïd et Sadki (2008) concernant les quantiles conditionnels dans un modèle de censure à droite, Ould Saïd et Tatachak (2009) pour le mode conditionnel sous troncature à gauche et finalement Lemdani *et al.* (2009) ont étudié la fonction des quantiles conditionnels pour des données tronquées mais dans le cas de variables indépendantes et identiquement distribués (i.i.d.).

1.2.5 Troncature à droite

seuls les individus avec une durée de vie inférieure à un certain seuil d'événement sont inclus dans l'étude. A titre d'exemple, si vous demandez à un groupe de lycéens fumeurs à quel âge ils ont commencé à fumer, vous avez nécessairement des données tronquées, que les personnes qui commencent à fumer après avoir quitté l'école ne sont pas inclus dans l'étude. Un autre exemple de troncature à droite survient dans l'estimation de la distribution des étoiles à partir de la terre. Les étoiles très lointaines ne sont pas visibles, et par conséquent, il y a troncature à droite.

1.3 la statistique non paramétrique robuste pour des données incomplètes : Motivations

La robustesse traduit le plus souvent la résistance de l'estimation aux données aberrantes, on définit les données aberrantes mathématiquement par le plus petit nombre de données extrêmes qui modifie la valeur de l'estimation ramené à la taille de l'échantillon. Historiquement, le terme robuste a été pour la première fois introduit en sta-

tistique par G.E.P. Box en 1953. L'estimation robuste de la fonction de régression a connu un réel essor en statistique non paramétrique, il s'agit d'un domaine qui a été popularisé par Huber (1964) dont il a obtenu la consistance et la normalité asymptotique d'une classe d'estimateurs pour cette fonction. Robinson (1984), Härdle (1984) et Härdle et Tsybakov (1989) ont établi sous des conditions de mélange la normalité asymptotique d'une famille d'estimateurs à pondération issue de la méthode du noyau pour la fonction de régression. La convergence uniforme de l'estimateur robuste de la fonction de régression a été obtenu par Collomb et Härdle (1986) en considérant des observations α -mélangeantes. Laïb et Ould-Saïd (2000) ont adapté l'estimateur de Collomb et Härdle (1986) pour le modèle d'auto-régression d'un processus stationnaire ergodique. Cai et Ould-Saïd (2003) ont utilisé une version robuste de l'estimation par la méthode des polynômes locaux pour la fonction de la régression. Durant ces dernières années, l'utilisation des données fonctionnelles est devenue très fréquente car cela offre un énorme potentiel en terme d'applications, notamment, en imagerie, en agro-alimentaire, en reconnaissance de formes, en géophysique, en économétrie, en environnement. Grâce au progrès technologique, les techniques infini-dimensionnelles sont maintenant plus puissantes, et la quantité des informations collectées devient volumineuse, ce qui peut engendrer un risque de mauvaise spécification de la relation entre les observations. La régression robuste présente une approche alternative aux méthodes de régression classiques qui sont sensibles à ce type d'erreurs qui proviennent d'une mauvaise lecture ou lorsque qu'on ne peut observer qu'une partie de données et non pas la totalité dans un modèle de survie pour des données incomplètes ou d'un mauvais enregistrement ou de toute autre cause due à l'environnement expérimental. La littérature sur la statistique fonctionnelle robuste pour des modèles de survie est très restreinte. En effet, les premiers résultats conséquents dans le domaine des données complètes ont été fournis par Cadre (2001). Ce dernier a étudié l'estimation de la médiane de la distribution d'une variable aléatoire à valeurs dans un espace de Banach. Azzedine *et al.* (2008) ont étudié la convergence presque complète d'une famille d'estimateurs robustes basée sur la méthode du noyau, en considérant des observations 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, Attouch *et al.* (2009, 2010, 2012) ont établie la convergence presque complète ponctuelle et uniforme d'une famille d'estimateurs robustes basée sur la méthode du noyau, ainsi que la normalité asymptotique.

Une des caractéristiques des données de survie est l'existence d'observations incomplètes. En effet, les données sont souvent recueillies partiellement, notamment, à cause des processus de censure et de troncature. Les données censurées ou tronquées proviennent du fait qu'on n'a pas accès à toute l'information : au lieu d'observer des réalisations indépendantes et identiquement distribuées (i.i.d.) de durées X , on observe la réalisation de la variable X soumise à diverses perturbations, indépendantes ou non du phénomène étudié, ce type de problème est réglé par la robustesse ce qui

motive notre travail. La littérature en ce domaine est très restreinte dans le cas infini on peut citer Hellal et Ould said(2016) qui ont étudié l'estimateur à noyau du quantile conditionnel pour des données tronquées à gauche pour des régresseurs fonctionnels, dans le cas de régresseurs en dimension finie Wang et Liang (2012), ils ont construit un M-estimateur de la fonction de régression pour un modèle tronqué à gauche quand les données sont dépendantes et ont établi la convergence faible et forte de l'estimateur ainsi la normalité asymptotique.

1.4 Estimation de la fonction de régression

Le problème de l'estimation en utilisant la régression a une longue histoire comme l'atteste la littérature. Déjà en 1632, Galileo Galilei utilisa une procédure qui peut être interprétée comme étant un ajustement linéaire de données contaminées. Un tel ajustement d'un nuage de points par une droite est le principe du problème classique de la régression linéaire. Une solution à ce problème est fournie par le célèbre principe des moindres carrés, mis en oeuvre indépendamment par A. M. Legendre et C. F. Gauss, et publié en (1805) et (1809), respectivement. Le premier estimateur de la régression non paramétrique de type "local averaging" a été proposé par J. W. Tukey (1947). Parmi les principaux estimateurs on rappelle l'estimateur de Nadaraya (1964) et Watson (1964)

$$\hat{r}_n(x) := \frac{\sum_{i=1}^n Y_i K_d \left(\frac{x - X_i}{h_n} \right)}{\sum_{i=1}^n K_d \left(\frac{x - X_i}{h_n} \right)}$$

L'étude de modèles de régression adaptés à des données fonctionnelles est un domaine important de la statistique fonctionnelle, que certaines situations ont été plus étudiées que d'autres et qu'il reste encore de nombreuses questions ouvertes. Au cours des années, des modèles fonctionnels plus généraux ont été proposés. De nombreux travaux se sont intéressés au modèle général dans lequel on ne fait pas d'hypothèse sur la forme de l'opérateur de régression mais seulement sur sa régularité. Les premiers résultats ont été obtenus à partir de l'étude d'un estimateur généralisé introduit par Ferraty et Vieu (2000). Pour tout x de $\mathcal{F}$ il s'écrit de la manière suivante

$$\hat{r}_n(x) := \frac{\sum_{i=1}^n Y_i K \left(\frac{d(x, X_i)}{h_n} \right)}{\sum_{i=1}^n K \left(\frac{d(x, X_i)}{h_n} \right)}.$$

Différents résultats ont déjà été obtenus pour cet estimateur en ce qui concerne la convergence presque sûre dans le cas d'un échantillon indépendant (voir Ferraty et Vieu, 2000, 2002) ou α -mélangeant (voir Ferraty *et al.*, 2002a, 2002b), la normalité asymptotique dans le cas indépendant (voir Ferraty *et al.*, 2007) ou dans le cas α -

mélangeant (voir Masry, 2005).

Cette estimation de r vu comme la moyenne conditionnelle de Y sachant X peut être inadaptée dans certaines situations. Par exemple, la présence de données aberrantes peut amener à des résultats non pertinents. La régression robuste a été introduite pour résoudre ce genre de problèmes. Cadre (2001), ainsi que Cardot *et al.* (2005) sont les principales références. Récemment, Azzedine *et al.* (2008) ont étudié la convergence presque complète d'estimateurs robustes à noyau. Dans le même cadre, Attouch *et al.* (2007) ont étudié la normalité asymptotique de ces estimateurs.

Le paramètre fonctionnel étudié dans ce travail, noté θ_x , est la solution (supposée unique) de l'équation en t définie par

$$\Psi(x, t) := \mathbb{E}[\psi(Y, t) \mid X = x] = 0.$$

Un estimateur à noyau de $\Psi(x, t)$ est donné par

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

1.5 Description de la thèse

La modélisation statistique de données fonctionnelles considère des données sous forme de courbes ou plus généralement des fonctions. Ce type de données se rencontre de plus en plus avec les récentes innovations réalisées sur les appareils de mesure et les méthodes d'acquisition ainsi que l'utilisation intensive de moyens informatiques qui permettent souvent de récolter ce type de données. Mais par fois les données sont volumineuses ce qui peut causer des erreurs lors de la détermination de la relation entre les observations. Le modèle de régression usuel est très sensible à ce type d'erreurs, d'où l'intérêt d'étudier la régression robuste en statistique fonctionnelle. Dans ce cadre, on se propose dans cette thèse d'étudier l'estimation robuste de la fonction de régression, dans le cas où les observations sont de nature fonctionnelle et la variable réponse est tronquée à gauche par une autre variable aléatoire. Cette thèse est divisée en quatre chapitres. Le premier chapitre est une introduction générale, introduisant des concepts généraux utiles pour faciliter la lecture des chapitres suivants. Dans le second chapitre, Nous rappelons certains résultats établis sur le comportement asymptotique de quelques estimateurs non paramétriques robustes de la fonction de régression, dans le cas où les données sont complètement observées et dans le cas de données incomplètes que ce soit dans le cadre de dimension finie ou infinie.

Dans le troisième chapitre, nous présentons des travaux sur la régression robuste dans le cas de données i.i.d., tronquées aléatoirement à gauche. Nous illustrons les résultats obtenus (consistance et normalité asymptotique) par des simulations pour montrer les performances des estimateurs proposés.

le quatrième chapitre est une extension des résultats du troisième chapitre au cas de données fortement mélangeantes. Seule la consistance est présentée, la normalité asymptotique fait l'objet d'un travail en cours.

Dans le cinquième chapitre, une conclusion suivie de quelques perspectives permettant d'étendre les résultats de cette thèse.

1.6 Présentation des résultats obtenus

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

1.6.1 Notations

Soit (X, Y) un couple de variables aléatoires à valeurs dans $\mathcal{F} \times \mathbb{R}$ où $\mathcal{F}$ est un espace semi-métrique.

On note d la semi-métrique sur $\mathcal{F}$. D'après Ferraty et Vieu (2006) où les données sont complètes, il est bien connu que l'estimateur à noyau classique $\widehat{m}(\cdot)$, de $m(\cdot)$ est un estimateur de moindres carrés, tel que

$$\widehat{m}(x) = \arg \min_{t \in \mathbb{R}} \sum_{i=1}^N W_{N,i}(x) (Y_i - t)^2,$$

où

$$W_{N,i}(x) = \frac{K(h_N^{-1}d(x, X_i))}{\sum_{i=1}^N K(h_N^{-1}d(x, X_i))}, \quad (1.1)$$

sont les poids de Nadaraya-Watson (NW).

K est un noyau à valeur réelles et $h := h_N$ est une suite de nombres réels positifs qui tend vers zero quand n tend vers l'infini.

Il est bien connu que la régression classique est non robuste. La ψ -régression méthode est largement développée pour atténuer les effets des valeurs abérrantes.

On considère une fonction réelle, mesurable, notée ψ_x , pour tout $x \in \mathcal{F}$. La fonction de ψ -régression est le paramètre fonctionnel noté θ_x , est solution de l'équation suivante :

$$t \mapsto \Psi(x, t) := \mathbb{E}[\psi(Y, t) \mid X = x]$$

On a alors , $\Psi(x, \theta_x) = 0$.

On suppose que pour $x \in \mathcal{F}$, θ_x existe et unique (Boente et Fraiman, 1989 et Boente *et al...*, 2009).

Soit $(T_i)_{i=1, \dots, N}$ un ensemble de variables aléatoires i.i.d. de même loi que T , où N est inconnu mais déterministe. En troncature à gauche, la variable aléatoire d'intérêt Y et la variable aléatoire de troncature T , sont observable seulement si $Y \geq T$ aucune n'est observé si $Y < T$. Dans ce cas, nous disposons d'un échantillon de taille ($n \leq N$) observé, dont la variable d'intérêt Y est observable, $(Y_i, T_i), i = 1, 2, \dots, n$; avec n aléatoire mais connu. On note que si les données originales $(Y_i, T_i)_i, i = 1, 2, \dots, N$ sont i.i.d., alors les données observées $(Y_i, T_i)_i, i = 1, 2, \dots, n$ sont aussi i.i.d. (voir Lemdani et Ould Saïd, 2007).

Comme conséquence de troncature, la taille de l'échantillon observé, n , est une variable aléatoire de loi binomiale $Bin(N, \tau)$, avec $\tau := \mathbb{P}(Y \geq T)$. C'est clair que si $\tau = 0$, aucune des données n'est observée alors, on suppose que $\tau \neq 0$. On construit notre estimateur à partir des variables observées $(X_i, Y_i, T_i)_{i=1, \dots, n}$.

On suppose que la variable d'intérêt Y et la variable de troncature T sont indépendantes et admettent des fonctions de répartition inconnues F et G , respectivement. D'après Stute (1993) les fonctions de répartition de Y et T , sous la condition de troncature à gauche, sont exprimées, respectivement, par

$$F^*(y) := \tau^{-1} \int_{-\infty}^y G(v) dF(v) \quad \text{et} \quad G^*(t) := \tau^{-1} \int_{-\infty}^{\infty} G(t \wedge v) dF(v)$$

où $t \wedge v = \min(t, v)$ et elles sont estimées par leurs estimateurs empiriques.

Soit

$$R(y) = G^*(y) - F^*(y) = \tau^{-1} G(y) [1 - F(y)].$$

estimé par son estimateur empirique.

On rappelle que les fonctions de répartition F et G sont estimées par l'estimateur de Lynden-Bell (1971) comme suit

$$F_n(y) = 1 - \prod_{s \leq y} \left[1 - \frac{F_n^*(s)}{R_n(s)} \right] \quad \text{and} \quad G_n(y) = \prod_{s > y} \left[1 - \frac{G_n^*(s)}{R_n(s)} \right].$$

He et Wang (1998) ont montrés que τ peut être estimé par

$$\tau_n = \frac{G_n(y) \bar{F}_n(y)}{R_n(y)}.$$

et indépendant de y .

les propriétés asymptotiques de F_n et G_n ont été étudié par Woodroffe (1985), tandis que He and Yang (1998) ont obtenu la convergence presque sûre de τ_n . La consistance

du modèle de troncature serait établie en considérant la probabilité conditionnelle $\mathbf{P}(\cdot)$ (de l'ensemble n) et non la mesure de probabilité $\mathbb{P}(\cdot)$ (de l'ensemble N). On note que $\mathbf{E}(\cdot)$ et $\mathbb{E}(\cdot)$ (respectivement) les opérateurs d'espérance de $\mathbf{P}(\cdot)$ et $\mathbb{P}(\cdot)$. Finalement les propriétés asymptotiques de F_n, G_n et τ_n sont obtenues sous les conditions d'identifiabilité tels que $a_G \leq a_F$, $b_G \leq b_F$ où a_W et b_W les points de bord du support de la fonction de répartition W définis

$$a_W = \inf\{x : W(x) > 0\} \quad \text{et} \quad b_W = \sup\{x; W(x) < 1\}.$$

Soit a, b deux réels positifs tel que $[a, b] \subset [a_F, b_F]$. De plus on suppose que notre paramètre est tel que $\theta_x \in (a, b)$.

Soit le triplet de variables aléatoires observées $(X_1, Y_1, T_1), \dots, (X_n, Y_n, T_n)$, en combinant les idées de Ferraty et Vieu (2006) et Ould Saïd et Lemdani (2006), l'estimateur à noyau de la fonction ψ -régression θ_x noté $\hat{\theta}_x$ est solution par rapport à t de l'équation

$$\hat{\Psi}(x, t) = 0, \tag{1.2}$$

où

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

et $h := h_n$.

1.6.1.1 Résultats : Cas i.i.d.

Dans cette partie les observations seront considérées indépendantes. Si la fonction objective ψ_x vérifie certaine condition de régularité, et si, la propriété de concentration de la mesure de probabilité de la variable fonctionnelle est telle que $\forall r > 0, \mathbf{P}(X \in B(x, r)) =: \phi_x(r) > 0$ avec $B(x, r) := \{x' \in \mathcal{F} / d(x', x) < r\}$. Alors, sous des hypothèses techniques assez générales on établit la convergence presque complète avec taux de convergence de notre estimateur $\hat{\theta}_x$ de θ_x .

Théoreme 1.6.1 *Sous des hypothèses de concentration de la mesure de probabilité de la variable fonctionnelle et des conditions techniques standards sur le noyau et le fenêtrage, de plus θ_x existe pour n suffisamment grand. On a*

$$\hat{\theta}_x - \theta_x \rightarrow 0 \quad p.c.$$

Afin de donner des résultats asymptotiques plus précis, on a obtenu le résultat suivant

Théoreme 1.6.2 *Sous des hypothèses de concentration de la mesure de probabilité de la variable fonctionnelle et des conditions techniques, $\hat{\theta}_x$ existe p.c., De plus, si $\Psi'(x, \theta_x) \neq 0$ on a*

$$\hat{\theta}_x - \theta_x = O(h^{b_1}) + O\left(\sqrt{\frac{\log n}{n\phi_x(h)}}\right) \quad p.c.$$

La démonstration et les détails des conditions imposées pour aboutir à ces résultats sont donnés dans le chapitre 3. Le deuxième résultat est sur la normalité asymptotique de notre estimateur.

Théoreme 1.6.3 *Sous des hypothèses de régularité de la fonction objective ainsi que des conditions standards sur le noyau et la fenêtre, on a pour tout $x \in \mathcal{A}$,*

$$\left(\frac{n\phi_x(h)}{\sigma^2(x, \theta_x)}\right)^{1/2} \left(\hat{\theta}_x - \theta_x + B_n(x)\right) \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1) \quad pc \quad n \rightarrow \infty.$$

où

$$\sigma^2(x, \theta_x) = \tau \frac{\alpha_2 \Psi^{2,G}(x, \theta_x)}{\left(\alpha_1 \frac{\partial}{\partial t} \Psi(x, \theta_x)\right)^2} \quad (\text{avec } \alpha_j = K(1) - \int_0^1 (K^j)'(s) \beta_x(s) ds, \text{ for, } j = 1, 2),$$

$$B_n(x) = h \Phi'(0, \theta_x) \alpha_1^{-1} \left((K(1) - \int_0^1 (sK(s))'(s) \beta_x(s) ds) \right) + o(h).$$

$$\mathcal{A} = \{x \in \mathcal{F}, \Psi^{2,G}(x, \theta_x) \frac{\partial}{\partial t} \Psi(x, \theta_x) \neq 0\}$$

et $\xrightarrow{\mathcal{D}}$ symbolise la convergence en loi.

La démonstration et les conditions imposées afin d'obtenir ce théorème sont données au chapitre 3.

1.6.1.2 Résultats : Cas α mélangeant

Afin de généraliser les résultats obtenus dans le chapitre précédent à des observations fortement mélangeantes, nous renforçons les hypothèses précédentes, en ajoutant des hypothèses supplémentaires sur la concentration de loi conjointe $(X_i; X_j)$ et sur le coefficient de mélange, on montre le résultat suivant :

Théoreme 1.6.4 *Sous des hypothèses de régularité de la fonction robuste ainsi que des conditions standards sur le noyau et la fenêtre et sur le coefficient de mélange, on a pour tout n suffisamment grand $\hat{\theta}_x$ existe, et on a*

$$\hat{\theta}_x - \theta_x \rightarrow 0 \quad p.c.$$

Afin de donner des résultats asymptotiques plus précis, on remplace quelques hypothèse et on obtient le résultat suivant

Théoreme 1.6.5 *pour tout n suffisamment grand $\hat{\theta}_x$ existe p.s. Si $\Gamma''(x, \cdot)(x, \theta_x) \neq 0$ on a*

$$\hat{\theta}_x - \theta_x = O(h^b) + O\left(\sqrt{\frac{\log n}{n\phi_x(h)}}\right) \quad p.c.$$

La démonstration de ce théorème ainsi que les hypothèses utilisées sont détaillées au chapitre 4.

Références

1. Attouch, M., Laksaci, A. and Ould Saïd, E. (2009). Asymptotic distribution of robust estimator in nonparametric functional models, *Comm. in Statist. Theory and Methods*, **38**, 1317-1335.
2. Attouch, M., Laksaci, A. and Ould Saïd, E. (2010). Asymptotic normality of a robust estimator of the regression function for functional time series, *J. Korean Statist. Soc.*, **39**, 489-500.
3. Attouch, M., Laksaci, A. and Ould Saïd, E. (2012). Robust regression for functional time series data. *J. Japan Statist. Soc.*, **42**, 125-143.
4. Azzedine, N., Laksaci, A. and Ould Saïd, E. (2008). On the robust nonparametric regression estimation for functional regressor, *Statist. Probab. Lett.*, **78**, 3216-3221.
5. Benhenni, K., Ferraty, F., Rachdi, M. et P. Vieu (2007). Local smoothing regression with functional data. *Comp. Statistics*, **22**, 353-369.
6. Bosq, D. (2000). Linear Processes in Function Spaces : Theory and Applications Lecture Notes in Statistics *Springer-Verlag*, **149**, New York.
7. Burba, F., Ferraty, F. et Vieu, P. (2008). Propriétés asymptotiques de l'estimateur kNN en régression fonctionnelle non-paramétrique. *C. R. Math. Acad. Sci. Ser.*, **I 346**, 339-342.
8. Bertrand, Retali Convergence uniforme de certains estimateurs de la densité par la méthode du noyau. C.R.A.S. Série A, **279**, 527-529.
9. Collomb, G. and Härdle, W. (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.
10. Crambes C, Delsol L. and Laksaci A. (2008). Robust nonparametric estimation for functional data. *J. Nonparametric Statist.*, **20**, 573-598.
11. Ferraty, F. et Vieu, P. (2002). The functional nonparametric model and application to spectrometric data. *Comput. Statist.*, **17 (4)**, 545-564.
12. Ferraty, F. and Vieu, P. (2003). Curves discrimination : a nonparametric functional approach *Computational Statistics and Data Analysis*, **44 (1-2)**, 161-173.
13. Ferraty, F. and Vieu, P. (2006). Nonparametric Functional Data Analysis. *Theory and Practice. Springer-Verlag*, New York.
14. Holmstrom, I. (1961). On method for parametric representation of the state of the atmosphere. *Tellus*, **15**, 127-149.
15. Huber, P.J. (1964). Robust estimation of a location parameter. *Annals of the Mathematical Statistics*, **35**, 73-101.

16. González-Manteiga W. and Vieu P. (2007). Editorial Statistics for Functional Data. *Comp. Stat. Data Analysis*, **51** (10), 4788-4792.
17. Lecoutre, J.P. (1990) Uniform consistency of a class of regression function estimators for Banach-space valued random variable. *Statist. Probab. Lett.* **10**, 145-149.
18. Masry, E. (2005). Nonparametric regression estimation for dependent functional data : asymptotic normality Stochastic Process. *Appl*, **115** (1), 155-177.
19. Ould Saïd, E. and Lemdani, M. (2006). Asymptotic properties of a nonparametric regression function estimator with randomly truncated data. *Ann. Inst. Statist. Math.*, **58**, 357-378.
20. Ramsay, J. and Silverman, B. (1997). Functional Data Analysis *Springer-Verlag*, New York.
21. Ramsay, J. and Silverman, B. (2002). Applied functional data analysis : Methods and case studies *Springer-Verlag*, New York.
22. Ramsay, J. and Silverman, B. (2005). Functional Data Analysis (Second Edition) *Springer-Verlag*, New York.
23. Wang, J., Liang, H.Y. and Fan, G. (2012). Asymptotic properties for an M-estimator of the regression function with truncation and dependent data. *Journal of the Korean Statistical Society*, **41**, 35-367.
24. Woodroffe, M. (1985). Estimating a distribution function with truncated data. *The Annals of Statistics*, **13**, 163-177.

Chapitre 2

Estimation nonparamétrique robuste

La robustesse permet d'avoir une certaine insensibilité aux quelques erreurs localisées qu'il peut y avoir dans les données. Une méthode robuste doit produire un résultat qui ne changera que légèrement si une petite partie des données est remplacée par de nouvelles valeurs qui peuvent être très différentes des originales. Les méthodes robustes doivent accorder une importance assez grande à la partie principale des données et laisser une part très faible aux possibles "lointains". Par exemple, la médiane est une statistique robuste tandis que la moyenne ne l'est pas. On trouve dans la littérature trois types d'estimateurs robustes (M, L et R)-estimateurs qui correspondent respectivement aux estimateurs de type maximum de vraisemblance, combinaisons linéaires des statistiques d'ordre et estimateur dérive de test de rang. Les M-estimateurs sont les plus utilisés.

2.1 Les M-estimateurs

Les M-estimateurs ont été introduits par Huber (1964), où il a proposé de généraliser l'estimateur du maximum de vraisemblance en le considérant comme un problème de minimisation d'une certaine fonction objectives. Plutôt que de prendre le carré de l'écart entre chaque observation x_i et l'estimateur t , il a utilisé une fonction $\rho(x, t)$ et former la fonction objective en faisant la somme sur tout l'échantillon $\sum_{i=1}^n \rho(x, t)$. A noter que le choix de $\rho(x, t) = -\log f(x, \theta)$ donne l'estimateur de maximum de vraisemblance ordinaire. Cette fonction $\rho(x, t)$ dépend souvent de $x - t$. Alors le M-estimateur pour la fonction $\rho(x, t)$ est la valeur de t qui minimise la fonction objective, ainsi pour trouver notre estimateur, il suffira de trouver la valeur de t qui vérifie :

$$\psi(x, t) = 0$$

tel que la fonction ψ est la dérivée par rapport à t de la fonction ρ

2.2 Exemples de M-estimateurs

Dans ce paragraphe nous présentons des exemples de classes de M-estimateurs les plus fréquemment utilisées :

2.2.1 Estimateur biweight

L'estimateur biweight souvent surnommé biweight de Tukey, est défini par

$$\psi(u) = \begin{cases} u(1 - u^2)^2 & |u| \leq 1 \\ 0 & |u| > 1. \end{cases}$$

2.2.2 Estimateur de Huber

L'estimateur de Huber est défini par

$$\psi(u) = \begin{cases} u & |u| \leq k \\ k \operatorname{signe}(u) & |u| > k. \end{cases}$$

2.2.3 Estimateur de Andrew

L'estimateur de Andrew, souvent appelé vague de Andrew, est défini par

$$\psi(u) = \begin{cases} \frac{1}{\pi} \sin \pi u & |u| \leq 1 \\ k \operatorname{signe}(u) & |u| > 1. \end{cases}$$

2.2.4 Le choix de la fonction objective

Le choix de la fonction objective (appelée souvent fonction d'influence dans la littérature) est très important pour la régression robuste. En effet, cette fonction contrôle la contribution de chaque observation dans l'erreur de l'estimation. Un des critères judicieux sur la fonction ψ est qu'elle soit bornée, l'existence d'un estimateur robuste unique impose la monotonie de la fonction objective. Chaffai (1989) a fait une étude comparative entre différentes fonctions objectives, et il a constaté qu'en plus de la fonction objective, la robustesse dépend aussi du paramètre d'échelle retenu. Il conclut dans cette étude que la fonction objective de Fair en pratique est peu sensible à ce paramètre. Cependant, de manière générale, le modèle de Huber est souvent considéré comme l'un des meilleurs estimateurs car il associe une décroissance forte à une zone de confiance contrôlée par le paramètre k , tandis qu'à l'extérieur de cette zone il se comporte comme une médiane, Attouch et al. (2009) ont fait une étude comparative entre plusieurs fonctions robustes dans la présence des données fonctionnelles en calculant l'erreur absolue en moyenne et ils ont remarqué que les fonctions d'Andrew et Tukey sont les meilleures.

2.3 La régression robuste : Cas de dimension finie

2.3.1 Pour des données complètes

Dans la statistique non paramétrique la méthode la plus communément utilisée pour étudier la relation entre deux variables X et Y est la régression, tels que Y une variable de réponse et X un vecteur aléatoire de prédicteurs (co-variables) prenant ses valeurs dans $\mathbb{R}^d$. La fonction de régression à un point $x \in \mathbb{R}^d$ est l'espérance conditionnelle de Y sachant $X = x$

$$m(x) := \mathbb{E}(Y|X = x).$$

Il existe plusieurs méthodes pour estimer une fonction de régression comme l'estimateur à noyau, les méthodes de spline de régression, etc. Cependant, ces méthodes ne sont pas robustes, il sont sensibles aux valeurs aberrantes. Alors que les valeurs aberrantes ou sont observées très souvent dans de nombreux domaines d'application comme la finance, pour cela un traitement approprié des valeurs aberrantes est une étape importante pour mettre en évidence les caractéristiques d'ensemble de données. Afin de surmonter l'inconvénient de robustesse, La régression robuste est une approche statistique sur la régression qui admet un comportement stable ou résistant lorsque le modèle est affecté par des données aberrantes (atypique, données tronquées). Au cours des dernières décennies, il y a une vaste littérature consacrée à l'étude de M -estimateurs de la fonction de régression non paramétrique qui ont été introduit par Huber (1964). Par la suite Collomb et Härdle (1986) ont étudié la convergence uniforme de l'estimateur robuste de la fonction de régression en considérant des observations ϕ -mélangeantes dans le cas réel (en dimension finie).

2.3.2 Pour des données incomplètes

La statistique non paramétrique robuste sur les données incomplète est un domaine en plein essor, que certaines situations ont été étudiées et qu'il reste encore de nombreux problèmes à résoudre, dans le cas fini-dimensionnel beaucoup de résultats sont donnés on citera quelques uns.

Wang et Liang (2012) ont construit un M -estimateur de la fonction de régression pour un modèle tronqué à gauche avec des données dépendantes et ont établi la convergence faible et forte. Ould Saïd et Lemdani (2006) ont construit un estimateur de type Nadaraya-Watson (NW) pour un modèle tronqué à gauche comme suit

$$m_n(x) = \frac{\sum_{i=1}^n G_n^{-1}(Y_i) K\left(\frac{x - X_i}{h_n}\right) Y_i}{\sum_{i=1}^n G_n^{-1}(Y_i) K\left(\frac{x - X_i}{h_n}\right)}$$

où K un noyau, h_n un paramètre de lissage et

$$G_n(y) = \prod_{T_i > y} \left[\frac{nR_n(T_i) - 1}{nR_n(T_i)} \right]$$

et

$$R_n(y) = \frac{1}{n} \sum_{i=1}^n I(T_i \leq y \leq Y_i).$$

En se basant sur cet estimateur, Wang et Liang (2012) ont construit un M-estimateur $\hat{m}_n(x)$ qui est défini comme solution par rapport à τ de l'équation

$$\sum_{i=1}^n G_n^{-1}(Y_i) K\left(\frac{x - X_i}{h_n}\right) \psi(Y_i - \tau) = 0.$$

Ils ont obtenu des résultats asymptotiques sous des conditions de mélange comme suit :

Théoreme 2.3.1 *Soit $x \in \mathbb{R}^d$ et $\alpha(n) = O(n^{-\gamma})$ pour $\gamma > 3$. On suppose que $m(x)$ est une solution de l'équation $\mathbb{E}[\psi(Y_1 - s) | X_1 = x] = 0$. Sous des hypothèses sur le noyau, la fenêtre et la fonction ψ*

$$\hat{m}_n(x) \xrightarrow{P} m(x).$$

Théoreme 2.3.2 *Sous les hypothèses du théorème précédent, s'il existe une suite de nombre réels A_n tel que A_n et $\alpha(l)$ satisfaisant*

$$A_n \rightarrow \infty, \sum_{n=1}^{\infty} A_n^{-r} < \infty, \frac{nh_n^d}{A_n^2 \log n} \rightarrow \infty, \sum_{n=1}^{\infty} \frac{A_n \alpha([A_n])}{h_n^d} < \infty \text{ alors}$$

$$\hat{m}_n(x) \xrightarrow{p.s.} m(x).$$

Lemdani et Ould Saïd (2015) se sont intéressés à un autre type de données incomplètes qui est la censure. Ils ont considéré un estimateur robuste de la régression avec une variable d'intérêt censurée à droite et des covariables X_i iid dans $\mathbb{R}^d$ avec une densité $\gamma > 0$. Ils ont adopté le modèle qui est solution $m(x)$ de l'équation suivante

$$\Psi(x, \theta) := \mathbb{E}[\psi_x(T_1 - \theta) | X_1 = x] \gamma(x) = 0$$

qui est estimé par $\hat{\Psi}$ donné par

$$\hat{\Psi}(x, \theta) := \frac{1}{nh^d} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right) \hat{\psi}^*(Y_i, \theta),$$

avec

$$\widehat{\psi^*}(Y_i, \theta) := \frac{\delta\psi_x(T_1 - \theta)}{\bar{G}_n(T_1)}.$$

où $\bar{G}(\cdot)$ est l'estimateur de Kaplan-Meier (1958) de la survie de la variable aléatoire de la censure.

Alors l'estimateur de $m(x)$ noté $\widehat{m}(x)$ est la solution par rapport à θ de l'équation

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

Sous des conditions classiques et en utilisant les VC-classes, les auteurs ont établi la convergence uniforme avec vitesse sur un compact de $\mathbb{R}^d$ et la normalité asymptotique de l'estimateur robuste.

Théoreme 2.3.3 *Sous des hypothèses classiques.*

$$(i) \sup_{x \in \mathcal{S}} |\widehat{m}(x) - m(x)| \xrightarrow{p.s.} 0.$$

$$(ii) \sup_{x \in \mathcal{S}} |\widehat{m}(x) - m(x)| = O(h^2) + O\left(\sqrt{\frac{\log n}{n^{1-2\zeta}h^d}}\right) \quad p.s.$$

Théoreme 2.3.4 *Soit $x \in \mathcal{S}$ sous des hypothèses techniques. On a*

$$\sqrt{nh^d}(\widehat{m}(x) - m(x)) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma^2(x))$$

$$\sigma^2(x) = \frac{\Gamma(x, x, m(x)) \int K^2(v) dv}{\left[\frac{\partial \Psi}{\partial \theta}(x, m(x))\right]^2}.$$

Avec $\Gamma(x, u, \theta) := \mathbb{E} \left[\frac{\psi_x^2(T_1 - \theta)}{\bar{G}(T_1)/X_1 = u} \right] \gamma(u)$

2.4 La régression robuste : Cas de dimension infinie

2.4.1 Pour des données complètes

Dans cette partie on suppose que la variable à expliquer Y est à valeurs réelles et la variable explicative X est à valeurs dans un espace de fonctions $\mathcal{F}$ muni d'une semi-métrique d . Ce type de variables, connu sous le nom de variables fonctionnelles dans la littérature, permet de considérer des variables aléatoires en tant que fonctions (du temps par exemple), ce qui est adapté dans le cas où les observations sont par nature fonctionnelles : on renvoie notamment à Ramsay et Silverman (2002, 2005) pour une vue d'ensemble sur les données fonctionnelles. Dans ce contexte, le

modèle le plus général est le modèle de régression lorsque la variable explicative est une courbe. Concernant la régression robuste pour ce type de données, la littérature est beaucoup plus restreinte : Cadre (2001), ainsi que Cardot et al. (2005) sont les principales références. Récemment, Azzedine et al. (2008) ont étudié la convergence presque complète d'un estimateur robuste à noyau qui s'écrit comme suit :

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

Azzedine et al. (2008) ont étudié la convergence presque complète d'estimateurs robustes à noyau sous des hypothèses technique et standard sur le noyau et la fenêtre, ils ont supposé que la fonction objective ψ est monotone, différentiable et bornée et si la propriété de concentration de la mesure de probabilité de la variable fonctionnelle. Sous les hypothèses suivantes

- (H1) $\forall r > 0, \mathbf{P}(X \in B(x, r)) =: \phi_x(r) > 0, \phi_x(r) \rightarrow 0$ as $r \rightarrow 0$.
- (H2) $\exists C > 0$ et $b > 0$ tel que $\forall (x_1, x_2) \in \mathcal{N}_x, \forall t \in \mathbb{R}$

$$| \Psi(t, x_1) - \Psi(t, x_2) | \leq C d^b(x_1, x_2).$$

- (H3) La fonction ψ est strictement monotone, bornée, différentiable et sa dérivée est tel que $|\psi'(t)| > C_1$.

Théoreme 2.4.1

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

Attouch *et al.* (2009) ont obtenu des résultats pour des données indépendantes, sous des hypothèses techniques assez générales sur le noyau et la fenêtre et sur la meme propriété de concentration de la mesure de probabilité de la variable fonctionnelle et les mêmes conditions sur la fonction objective et sous les hypothèses suivantes, ces auteurs ont établi la normalité asymptotique sur l'ensemble des x appartenant à l'ensemble

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

avec $\lambda_2(x)(u, t) = \mathbb{E}[\psi_x^2(Y, t)/X = u]$, $\Gamma_1(u, t) = \mathbb{E}[\psi'_x(Y, t)/X = u]$.

- (H1) La fonction $\lambda_\gamma(\cdot, \cdot)$ satisfait la condition de Lipschitz. Ils existent les constantes strictement positives b_γ tel que :

$$\forall (u_1, u_2) \in \mathcal{N}_x \times \mathcal{N}_x, \forall t \in \mathbb{R}, | \lambda_\gamma(u_1, t) - \lambda_\gamma(u_2, t) | \leq C_1 d(u_1, u_2)^{b_\gamma}$$

- (H2) La fonction $\Gamma_\gamma(\cdot, \cdot)$ satisfait la condition de Lipschitz. Ils existent les constantes strictement positives b_γ tel que :

$$\forall (u_1, u_2) \in \mathcal{N}_x \times \mathcal{N}_x, \forall t \in \mathbb{R}, |\Gamma_\gamma(u_1, t) - \Gamma_\gamma(u_2, t)| \leq C_1 d(u_1, u_2)^{d_\gamma}$$

Théoreme 2.4.2

$$\left(\frac{n\phi(h)}{\sigma^2(x, \theta_x)} \right)^{1/2} \left(\hat{\theta}_n - \theta_x - B_n(x) \right) \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1),$$

où $B_n(x) = \frac{h}{\phi(h)\alpha_1\Gamma_1(x, \theta_x)} \int_0^1 K(t)\varphi(th)\phi'(th)dt + o(1)$
avec $(\varphi(s) = \mathbb{E}[\psi(Y, \theta)/d(X, x) = s])$,

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

Sous des observations fortement mélangantes, en ajoutant des hypothèses supplémentaires sur la concentration de loi conjointe (X_i, X_j) et sur le coefficient de mélange.

Théoreme 2.4.3 *Sous des hypothèses peu restrictives sur le coefficient du mélange et la mesure de concentration de la variable fonctionnelle, on a*

$$\hat{\theta}_x - \theta_x = O(h^{b_1}) + O\left(\sqrt{\frac{\log n}{n\phi_x(h_k)}}\right) \text{ p.co.}$$

Théoreme 2.4.4 *Sous des hypothèses de régularité de la fonction robuste ainsi que des conditions standards sur le noyau et la fenêtre, on a pour tout $x \in \mathcal{A}$.*

$$\left(\frac{n\phi(h)}{\sigma^2(x, \theta)} \right)^{1/2} \left(\hat{\theta}_x - \theta_x - B_n(x) \right) \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1),$$

2.4.2 Pour des données incomplètes

La littérature est très restreinte en ce qui concerne la régression pour des données fonctionnelle tronquée à gauche, le but de notre étude est d'étudier les propriétés asymptotiques d'un estimateur robuste de la régression pour des co-variables fonctionnelles dans la présence de données tronquées à gauche, qui sont cités dans le troisième et quatrième chapitres.

Chapitre 3

Convergence presque complète et normalité asymptotique : Cas iid

Ce chapitre traite la convergence presque complète et la normalité asymptotique d'un estimateur robuste du régresseur fonctionnel tronquée à gauche, ce travail fait l'objet d'une publication au Journal of Statistical Theory and Practice.

On the nonparametric estimation of the functional ψ -regression under random left-truncation model

Saliha Derrar,[†] Ali Laksaci,[†]

[†] Laboratoire de Statistique et Processus Stochastiques,

Université de Sidi Bel Abbès,

BP 89 Sidi Bel Abbès 22000. Algeria.

derrarsaliha@yahoo.fr, alilak@yahoo.fr

Elias Ould Saïd,^{††,*}

^{††} Univ. Lille Nord de France,

* ULCO, LMPA, Centre de la Mi-Voix,

CS : 80699, Calais, France.

ouldsaid@lmpa.univ-littoral.fr¹

Abstract

In this paper we study a family of the nonparametric estimation of the ψ -regression model when the response variable is subject to left-truncation by an other random variable. Under standard assumptions, we get the almost complete convergence rate as well as asymptotic normality of a robust nonparametric estimators for the truncated and functional regressors. Clearly this permit to get confidence interval that is usable in practice since it does not depend on any unknown quantity. Some special cases have been studied to show how our work extend many results. Our results have been conformed by simulation study for consistency as well as for probability coverage of a confidence interval.

AMS subject classifications : 60G42, 62F12, 62G20, 62G05.

Key words : Asymptotic normality, Functional data, Kernel estimator, Lynden-Bell estimator, Robust estimation, Small balls probabilities, Strong consistency, Truncated data.

3.1 Introduction

The regression analysis of incomplete data (censored and/or truncated data) has gained a particular interest in the statistics literatures. Such kind of data occur in many fields of applications such as in astronomy, economics, epidemiology, biometry and medical studies (see, Woodroffe (1985), Andersen *et al.* (1993) or He and Yang

1. *corresponding author*

(1994)). For example, in a retirement center, subjects are observed only if they live long enough to enter the retirement house. The lifetime, say Y , is then left truncated by the retirement house entry age, say T . People who enter the retirement house earlier may get better medical attention and therefore live longer. On the other hand, people with poor health and shorter expected lifetime may retire earlier.

In this paper, we are interested on the ψ -regression when the interest random variable is subject to left-truncation and in the presence of functional covariable by using a robust approach.

Recall that, for complete data and in finite dimensional spaces, the pioneer work is of Huber (1964), where he studied an estimation of some location parameter. Collomb and Hardle (1986), proved the uniform almost complete convergence with rates of the robust nonparametric kernel estimator, when data satisfies a some of kind of dependency. Later, Laïb and Ould Saïd (2000), improved this result when the data satisfy an ergodicity condition.

The nonparametric modeling of functional data becomes a major topic of research in the last decade (see Ferraty and Romain 2011, for recent advances and references). In this context the robust estimation of the regression function is an interesting problem in statistical inference. It is used as an alternative approach to classical methods, in particular when the data are affected by the presence of outliers. This model was firstly introduced, as far as we know, in nonparametric functional statistics by Azzedine *et al.* (2008). They obtained the almost complete convergence of a family of robust estimators of the regression function in the independent and identically distributed (i.i.d.) case. Since this work, various advances on the nonparametric robust functional regression have been provided. Without pretend to the exhaustivity, we quote for instance, Crambes *et al.* (2008), Chen and Zhang (2009), Attouch *et al.* (2009, 2010, 2012), Gheriballah *et al.* (2013) and the references therein. All these results are obtained in the complete data. There are very few results in the nonparametric functional statistic when the data are incomplete. At the best of our knowledge, only the papers of Horrigue and Ould Saïd (2011, 2014) have paid attention to study the conditional quantiles estimation for censorship model and functional regressors. However, for the finite dimensional spaces, the nonparametric estimation of the regression function with randomly truncated data has been widely considered by many authors. For example, and to cite only a few, Ould Saïd and Lemdani (2006) studied the asymptotic properties of the kernel estimator of the regression function in i.i.d. case, however Wang *et al.* (2012a) used M-estimation method and Wang *et al.* (2012b) for local M-estimation method under dependent condition.

The main goal of this paper is to generalize the results of Wang *et al.* (2012a) from the real case to the functional case and those of Azzedine *et al.* (2008) and Attouch *et al.* (2009) to incomplete data. More precisely, we establish the almost-complete convergence, with rates as well as the asymptotic normality of the estimator constructed by

combining the ideas of robustness with those of smoothed regression. This induces a confidence interval that which is useful for the practitioners. Finally, noting that the main feature of our approach is to develop an alternative prediction model to the classical regression that is not sensitive to outliers or heteroscedastic data. Our results have been emphasized by simulation study for consistency as well as for probability coverage of a confidence interval.

The rest of the paper is organized as follows : Section 2 is devoted to the presentation of the model and give the estimator of the parameter. The assumptions and main results are given in Section 3. Some special cases have been made in Section 4. In Section 5, we deal with simulations data where where we study the convergence part and the coverage probability of a confidence interval, whereas the proofs of the main results are relegated to the Appendix.

3.2 Presentation of model and estimators

Consider N pairs of independent random variables (X_i, Y_i) for $i = 1, \dots, N$ that we assume drawn from the pair (X, Y) which is valued in $\mathcal{F} \times \mathbb{R}$, where $\mathcal{F}$ is a semi-metric space and d denoting the semi-metric.

Since Ferraty and Vieu (2006) and when there is no truncation, it is well-known that the classical kernel smoothed estimator, say $\hat{m}(\cdot)$, of $m(\cdot)$ may be viewed as a local least squares estimator, that is

$$\hat{m}(x) = \arg \min_{t \in \mathbb{R}} \sum_{i=1}^N W_{N,i}(x) (Y_i - t)^2,$$

where

$$W_{N,i}(x) = \frac{K(h_N^{-1}d(x, X_i))}{\sum_{i=1}^N K(h_N^{-1}d(x, X_i))}, \quad (3.1)$$

are the Nadaraya-Watson (NW) weights. Here K is a real-valued kernel function and $h := h_N$ be a sequence of positive real numbers which decreases to zero as n tends to infinity. It is well-known that the classical regression suffer from the not robustness. As mentioned before, the ψ -regression method have largely developed to ease the effects of the outliers values.

Recall that the ψ -regression function is the nonparametric parameter, denoted by θ_x , implicitly defined, for all $x \in \mathcal{F}$, as a value which vanishes with respect to (w.r.t.) $t \in \mathbb{R}$ the function

$$t \mapsto \Psi(x, t) := \mathbb{E}[\psi(Y, t) \mid X = x]$$

where ψ is a real-valued Borel function satisfying some regularity conditions to be stated below. Clearly we have, $\Psi(x, \theta_x) = 0$.

In what follows, we suppose that, for all $x \in \mathcal{F}$, θ_x exists and is unique (see, for instance, Boente and Fraiman, 1989 and Boente *et al.*, 2009).

Let $(T_i)_{i=1, \dots, N}$ be a sample of i.i.d. random variables which are distributed as T , where N is unknown but deterministic. In the random left-truncation model, the random variable (rv) of interest Y is interfered by the rv T , such that both quantities Y and T are observable only if $Y \geq T$ whereas neither is observed if $Y < T$. Without possible confusion, we still denote $(Y_i, T_i), i = 1, 2, \dots, n; (n \leq N)$ the actually observed r.v., with n being random but known. We point out that if the original data $(Y_i, T_i)_i, i = 1, 2, \dots, N$ is i.i.d., the observed data $(Y_i, T_i)_i, i = 1, 2, \dots, n$ are still i.i.d. (see Lemdani and Ould Saïd, 2007).

As a consequence of truncation, the size of the actually observed sample, n , is a $\text{Bin}(N, \tau)$ r.v., with parameter $\tau := \mathbb{P}(Y \geq T)$. It is clear that if $\tau = 0$, no data can be observed and therefore, we suppose through this paper that $\tau \neq 0$. Thus, taking into account the truncation's effect the practical estimator is constructed by the observed variables $(X_i, Y_i, T_i)_{i=1, \dots, n}$. To do that, we suppose that the variable of interest Y and the truncating variable T are independent and have, respectively, an unknown distribution function F and G . Following Stute (1993) the distribution functions of Y and T , under the left-truncation condition, are expressed, respectively, by

$$F^*(y) := \tau^{-1} \int_{-\infty}^y G(v) dF(v) \quad \text{and} \quad G^*(t) := \tau^{-1} \int_{-\infty}^{\infty} G(t \wedge v) dF(v)$$

where $t \wedge v = \min(t, v)$ and they are estimated by their empirical ones.

Let

$$R(y) = G^*(y) - F^*(y) = \tau^{-1} G(y) [1 - F(y)].$$

and also it is estimated by its empirical one.

Recall that the distribution function F and G are estimated by the Lynden-Bell (1971) as follows

$$F_n(y) = 1 - \prod_{s \leq y} \left[1 - \frac{F_n^*(s)}{R_n(s)} \right] \quad \text{and} \quad G_n(y) = \prod_{s > y} \left[1 - \frac{G_n^*(s)}{R_n(s)} \right].$$

He and Wang (1998) showed that τ can be estimated by

$$\tau_n = \frac{G_n(y) \bar{F}_n(y)}{R_n(y)}.$$

and is independent of y .

The asymptotic properties of F_n and G_n were studied by Woodroffe (1985), while the almost sure consistency of τ_n was obtained by He and Yang (1998). Noting that as N is

unknown and n is known (although random), the stochastic consistency in truncation model would be stated with respect to the conditional probability $\mathbf{P}(\cdot)$ (related to the n -sample) and not the probability measure $\mathbb{P}(\cdot)$ (related to the N -sample). We denote $\mathbf{E}(\cdot)$ and $\mathbb{E}(\cdot)$ the respective expectation operators of $\mathbf{P}(\cdot)$ and $\mathbb{P}(\cdot)$. Finally let's point out that the asymptotic proprieties of F_n, G_n and τ_n are obtained under an identifiability conditions that are $a_G \leq a_F$, $b_G \leq b_F$ where a_W and b_W the endpoints of the support of a distribution function W defined by

$$a_W = \inf\{x : W(x) > 0\} \quad \text{and} \quad b_W = \sup\{x; W(x) < 1\}.$$

Let a, b be two real numbers such that $[a, b] \subset [a_F, b_F]$. We point out that we need this assumption because we need the uniform consistency of the distribution law $G(\cdot)$ of the truncated rv T , which stated over compact set (see Remark 6 in Woodrooffe (1985)). Furthermore we suppose that our parameter is such that $\theta_x \in (a, b)$.

Given an observed triplet random variables $(X_1, Y_1, T_1), \dots, (X_n, Y_n, T_n)$, and combining the ideas in Ferraty and Vieu (2006) and Ould Saïd and Lemdani (2006), the kernel estimate of the ψ -regression function θ_x denoted $\hat{\theta}_x$ is a zero w.r.t. t of the equation

$$\hat{\Psi}(x, t) = 0, \tag{3.2}$$

where

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

and $h := h_n$.

3.3 Main results

From now on, x stand for a fixed point in $\mathcal{F}$, and all along the paper, when no confusion is possible, we denote by C and/or C' any generic positive constant. For $r > 0$, let $B(x, r) := \{x' \in \mathcal{F} / d(x', x) < r\}$. In order to derive the almost complete convergence (*a. co.*) of the kernel estimate $\hat{\theta}_x$ of θ_x , some conditions are necessary. Recall that a sequence Z_n is said to converge *a. co.* to Z if and only if, for any $\epsilon > 0$, $\sum_n \mathbf{P}(|Z_n - Z| > \epsilon) < \infty$.

We start by establishing the almost complete consistency with rates of our estimator.

3.3.1 Consistency

Our result will be stated under some assumptions which are gathered together for easy reference

(H1) $\forall r > 0, \mathbf{P}(X \in B(x, r)) =: \phi_x(r) > 0$. Moreover, $\phi_x(r) \rightarrow 0$ as $r \rightarrow 0$.

(H2) The function Ψ is such that :

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

(H3) The function ψ_x is continuous monotone w.r.t. the second component and for each fixed $t \in [\theta_x - \delta, \theta_x + \delta]$, $\forall j \geq 1$, $\mathbf{E}[\psi_x^j(Y, t)] < C < \infty$.

(H4) K is a function with support $(0, 1)$ such that

$$0 < C \mathbb{1}_{(0,1)} < K(t) < C' \mathbb{1}_{(0,1)} < \infty.$$

(H5) $\lim_{n \rightarrow \infty} h = 0$ and $\lim_{n \rightarrow \infty} \frac{\log n}{n \phi_x(h)} = 0$.

Remarks on the assumptions

Our conditions are very standard in this context. Indeed, the conditions (H1) is the same as those used by Ferraty and Vieu (2006). Conditions (H2) and (H3) on the score function is a quite lower than that considered in Crambes *et al.* (2008) or Attouch *et al.* (2010) where ψ is assumed to be continuously differentiable. Our robustification condition is verified for the usual functions (Andrews, Huber, Hampel, Tuckey, ...) which gives more flexibility in the practical choice. Furthermore, they are needed to evaluate the bias term in the asymptotic properties. Assumption (H4) concerns the kernel $K(\cdot)$ which is technical and imposed for sake of simplicity whereas (H5) is classical for consistency results.

Théoreme 3.3.1 *Assume that (H1), (H2)((i), (ii)) and (H3)-(H5) are satisfied, then $\hat{\theta}_x$ exists for all sufficiently large n . Furthermore, we have*

$$\hat{\theta}_x - \theta_x \rightarrow 0 \quad a.c.$$

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

Théoreme 3.3.2 *Assume that (H1), (H2) ((i), (iii)) and (H3)-(H5) are satisfied, then $\hat{\theta}_x$ exists a.s. for all sufficiently large n . Furthermore, if $\Psi'(x, \theta_x) \neq 0$ we have*

$$\hat{\theta}_x - \theta_x = O(h^{b_1}) + O\left(\sqrt{\frac{\log n}{n \phi_x(h)}}\right) \quad a.c.$$

3.3.2 Asymptotic normality

Now, we study the asymptotic normality of $\widehat{\theta}_x$. We replace (H1), (H2), (H4) and (H5) by the following hypotheses, respectively.

(H1') The concentration property (H1) holds. Moreover, there exists a function $\beta_x(\cdot)$ such that

$$\forall s \in [0, 1], \quad \lim_{r \rightarrow 0} \phi_x(sr)/\phi_x(r) = \beta_x(s).$$

(H2') Conditions (H2(i)) and (H2(ii)) hold and the derivative function

$\Phi(s, t) = \mathbf{E}[\Psi(X_1, t) - \Psi(x, t) | d(x, X_1) = s]$ exists at $s = 0$ and is continuous w.r.t. the second component at a neighborhood of θ_x . The function $\Psi^{2,G}(\cdot, \cdot) = \mathbf{E}\left[\frac{\psi^2(Y, \cdot)}{G(y)} | X = \cdot\right]$ is continuous at the point (x, θ_x) .

(H4') The kernel K satisfies (H4) and is a differentiable function on $]0, 1[$ with derivative K' such that $-\infty < C < K'(\cdot) < C' < 0$.

(H5') $n\phi_x(h) \rightarrow \infty$ and $nh^2\phi_x(h) \rightarrow C$ as $n \rightarrow \infty$.

Théoreme 3.3.3 *Assume that (H1')-(H5') hold. Then we have for any $x \in \mathcal{A}$,*

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

where

$$\sigma^2(x, \theta_x) = \tau \frac{\alpha_2 \Psi^{2,G}(x, \theta_x)}{\left(\alpha_1 \frac{\partial}{\partial t} \Psi(x, \theta_x)\right)^2} \quad (\text{with } \alpha_j = K(1) - \int_0^1 (K^j)'(s) \beta_x(s) ds, \text{ for } j = 1, 2),$$

$$B_n(x) = h\Phi'(0, \theta_x)\alpha_1^{-1} \left((K(1) - \int_0^1 (sK(s))'(s) \beta_x(s) ds) \right) + o(h).$$

$$\mathcal{A} = \{x \in \mathcal{F}, \Psi^{2,G}(x, \theta_x) \frac{\partial}{\partial t} \Psi(x, \theta_x) \neq 0\}$$

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

Corollary 1 *Under the assumptions of the Theorem 3.3.3 and if $nh^2\phi_x(h) \rightarrow 0$, then*

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

where $\sigma^2(x, \theta_x)$ is as before.

As it is well-known, the mean application for the asymptotic normality it to get confidence interval, which is given here after.

Corollary 2 *Based on $\overline{G}_n(\cdot), \hat{\theta}_x, \hat{\alpha}_j (j = 1, 2), \hat{\Psi}^{2, \hat{G}}(\cdot, \cdot)$ and the derivative $\frac{\partial \hat{\Psi}}{\partial t}(x, \hat{\theta}_x)$ estimators of $G(\cdot), \theta_x, \alpha_j (j = 1, 2), \Psi^{2, G}(\cdot, \cdot)$ and $\frac{\partial \Psi(x, \theta_x)}{\partial t}$ respectively, we easily get a (consistent) plug-in estimator $\hat{\sigma}^2(x, \hat{\theta}_x)$ for $\sigma^2(x, \theta_x)$ which, under the assumptions of Theorem 3.3.3, gives a confidence interval of asymptotic level $1 - \zeta$ for θ_x given by*

$$\left[\hat{\theta}_x - \frac{u_{1-\zeta/2} \hat{\sigma}(x, \hat{\theta}_x)}{\sqrt{n \phi_x(h)}}, \quad \hat{\theta}_x + \frac{u_{1-\zeta/2} \hat{\sigma}(x, \hat{\theta}_x)}{\sqrt{n \phi_x(h)}} \right]$$

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

3.4 Special cases

3.4.1 Comeback to complete data

In absence of truncation ($T = -\infty$) we have $\tau = 1$ and the asymptotic variance becomes

$$\sigma_{com.dat.}^2(x, \theta_x) = \frac{\alpha_2 \mathbf{E}[\psi^2(Y, \theta_x) | X = x]}{(\alpha_1 \frac{\partial}{\partial t} \Psi(x, \theta_x))^2}$$

So, for this case, we obtain the following Corollary.

Corollary 3 *Under the assumptions of the Theorem 3.3.3 and if $nh^2 \phi_x(h) \rightarrow 0$, then*

$$\left(\frac{n \phi_x(h)}{\sigma_{com.dat.}^2(x, \theta_x)} \right)^{1/2} (\hat{\theta}_x - \theta_x) \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1) \quad \text{as } n \rightarrow \infty.$$

This corollary is the same result obtained by Attouch *et al.* (2009).

3.4.2 Comeback to classical regression case

Clearly the classical regression defined by conditional expectation is a particular case of our study with $\psi_x(Y, t) = (Y - t)$. For this case

$$\sigma_{Cla.Reg.}^2(x) = \tau \frac{\alpha_2 \mathbf{E} \left[\frac{Y^2}{G(y)} | X = x \right] - \mathbf{E}^2 \left[\frac{Y}{G(y)} | X = x \right]}{\alpha_1^2}.$$

For this case, we obtain the following Corollary.

Corollary 4 *Under the assumptions of the Theorem 3.3.3 and if $nh^2\phi_x(h) \rightarrow 0$, then*

$$\left(\frac{n\phi_x(h)}{\sigma_{Cla.Reg.}^2(x, \theta_x)} \right)^{1/2} (\hat{\theta}_x - \theta_x) \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1) \quad a.c. \quad n \rightarrow \infty.$$

It should be noted that the result of this corollary is also new in statistics functional. Nevertheless, in non functional case, this asymptotic normality is comparable to the results of Lemdani and Ould-Said (2006).

3.4.3 Comeback to real case

In the real case when $\mathcal{F} = \mathbb{R}$, and if the probability density of the random variable X , denoted by f , is of $\mathcal{C}^1$ class, then $\phi_x(h) = f(x)h + o(h)$ and $\beta_x(s) = s$

$$\sigma_{Real}^2(x) = \tau \frac{\Psi^{2,G}(x, \theta_x) \int_0^1 K^2(s) ds}{\left(\frac{\partial}{\partial t} \Psi(x, \theta_x) \int_0^1 K(s) ds \right)^2}.$$

In the real case, we obtain the following Corollary.

Corollary 5 *Under the assumptions of the Theorem 3.3.3 and if $nh^3 \rightarrow 0$, then*

$$\left(\frac{nhf(x)}{\sigma_{Real}^2(x, \theta_x)} \right)^{1/2} (\hat{\theta}_x - \theta_x) \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1) \quad a.c. \quad n \rightarrow \infty.$$

Observe that in this case we obtain the same convergence rate given by Wang *et al.* (2012a). In the same context we obtain as almost complete convergence rate

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

Clearly, we attain the usual optimal rate in nonparametric statistic of Stone(1981) that is $\left(\frac{n}{\log n} \right)^{-b_1/2b_1+1}$ which corresponds the optimal $h = C \left(\frac{n}{\log n} \right)^{-1/2b_1+1}$

3.5 Simulation study

The aim of this part is to examine the performance of our estimator $\hat{\theta}_x$ for the consistence as well as the asymptotic normality by comparing it to the classical regression. More precisely, we compare the resistance to the presence of outliers for both

models in the two functional prediction problem's (the pointwise predication controlled by the mean absolute error (MAE) and the prediction bands controlled through the coverage probabilities of the proposed confidence intervals. Moreover, in order to check the effect of the truncation in this practical study we vary the percentage of truncation and we consider several different sample-size. Specifically, we proceed with the following algorithm :

We fix the random size n .

– *Step 1 :*

We fix the random size n and we generate a random variable $T_1, X_1(t)$, $t \in [0, 1]$ in the following manner :

$T_1 \hookrightarrow \mathcal{N}(\mu, 1)$, and we adapt μ in such way to get a different rate of truncation, $X_1(t)$ will be generate as $X_1(t) = AS_1(t) + (1 - A)S_2(t)$ with $A \hookrightarrow \mathcal{B}(0.5)$, $S_1(t) = 2 - \cos(\pi tW)$, $S_2(t) = \cos(\pi tW)$ and $W \hookrightarrow \mathcal{U}(0, 1)$. Furthermore we simulate $\varepsilon_1 \hookrightarrow \mathcal{N}(0, 0.5)$.

– *Step 2 :*

We calculate $Y_1 = R(X_1(t)) + \varepsilon_1$ where $R(X_1(t)) = \int_0^1 (X_1(t))^2 dt$.

– *Step 3 : TEST*

We begin by :

$N=0$,

$j=0$,

While $j \leq n$:

We put $N = N + 1$ We test : if $Y_1 < T_1$ we reject the triplet $(X_1(t), Y_1, T_1)$

Otherwise we keep the triplet $(X_1(t), Y_1, T_1)$ and we put $j = j + 1$

at the end of this count we get the deterministic N ,

which permits to get the rate of truncation $\tau = \frac{n}{N}$.

More precisely, the rate of the observed triplet.

We get n random vectors $(X_j(t), Y_j, T_j)$ $j = 1, \dots n$.

– *Step 4 :*

We calculate the Lynden-Bell estimator by the observed couple (Y_j, T_j) for $i = 1, \dots, n$

– *Step 5 :*

We calculate our estimator as the argument value (say $\hat{\theta}_x$) which is the upper negative value of $\Psi(x, t)$ with respect to t before $\Psi(x, t)$ becoming positive. Noting that, we have used for the robust regression the score function defined by $\psi_1(y, t) = \frac{y-t}{\sqrt{(1+((y-t)^2)/2)}}$ while for the classical case we have $\psi_2(y, t) = y - t$. It

should be noted that we simulate by a quadratic kernel defined by :

$$K(x) = \frac{3}{2}(1 - x^2)\mathbb{I}_{[0,1]},$$

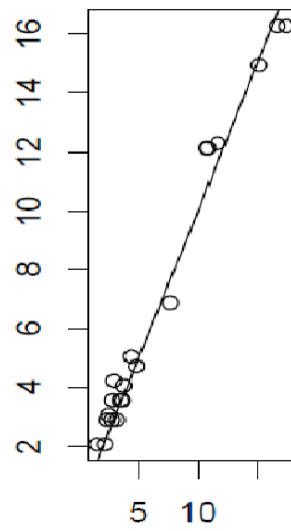
and we choose the optimal bandwidth h by the cross-validation method. Clearly, the regularity of the curves insures that the semi-metric defined by the L2-distance between the first derivatives of the curves is more adapted than functional PCA-metric (see Ferraty and Vieu (2006) for more discussion in this subject).

Remark 1 *Here we point out that unlike to complete and censored data, the explanatory random function $X(t)$ is not kept in each step but only if the interest random variable Y is bigger than the truncated random variable T .*

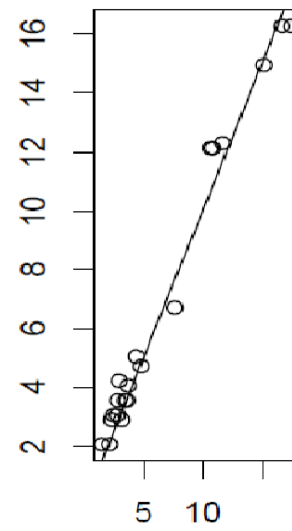
3.5.1 Consistency

Our main purpose in this part is to illustrate the consistency propriety of our estimate by examining its accuracy as predictor in the following situations :

For the first case. We fix the percentage of truncation evaluated by $\mu = 2$ and we vary the sample-size $n = 50, 100, 500$. For each sample-size case, we split our data into two subsets (learning sample and test sample). From the learning sample we compute the predicted values $\hat{\theta}_x$ (for the two score functions ψ_1 and ψ_2), corresponding to the curves of the test sample. In the following figures we plot the predicted values estimated by $\hat{\theta}_x$ versus the true values. The continuous line represents what would be the perfect prediction. Typically, the efficiency of the prediction method is quantified by the closeness of the round point to this continuous line.

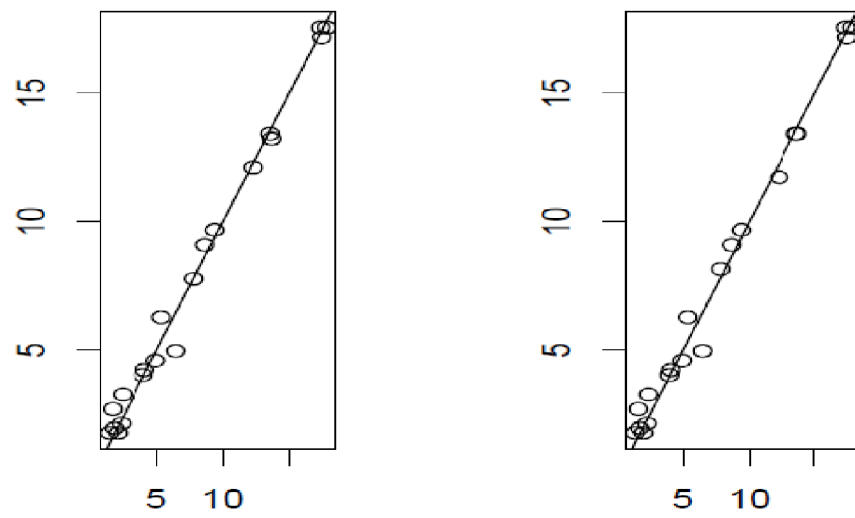


The robust regression, MAE=0.57



The classical regression, MAE=0.59

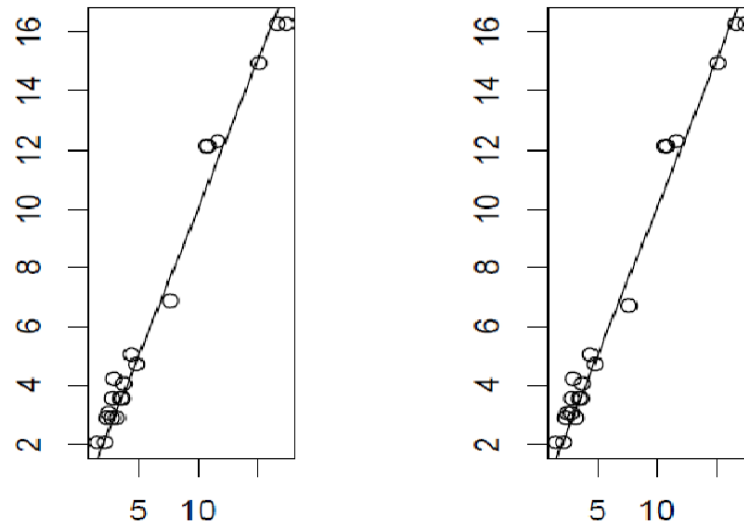
Figure 1. Case : $n = 50$



The robust regression, MAE=0.40

The classical regression, MAE=0.43

Figure 2. Case : $n = 100$



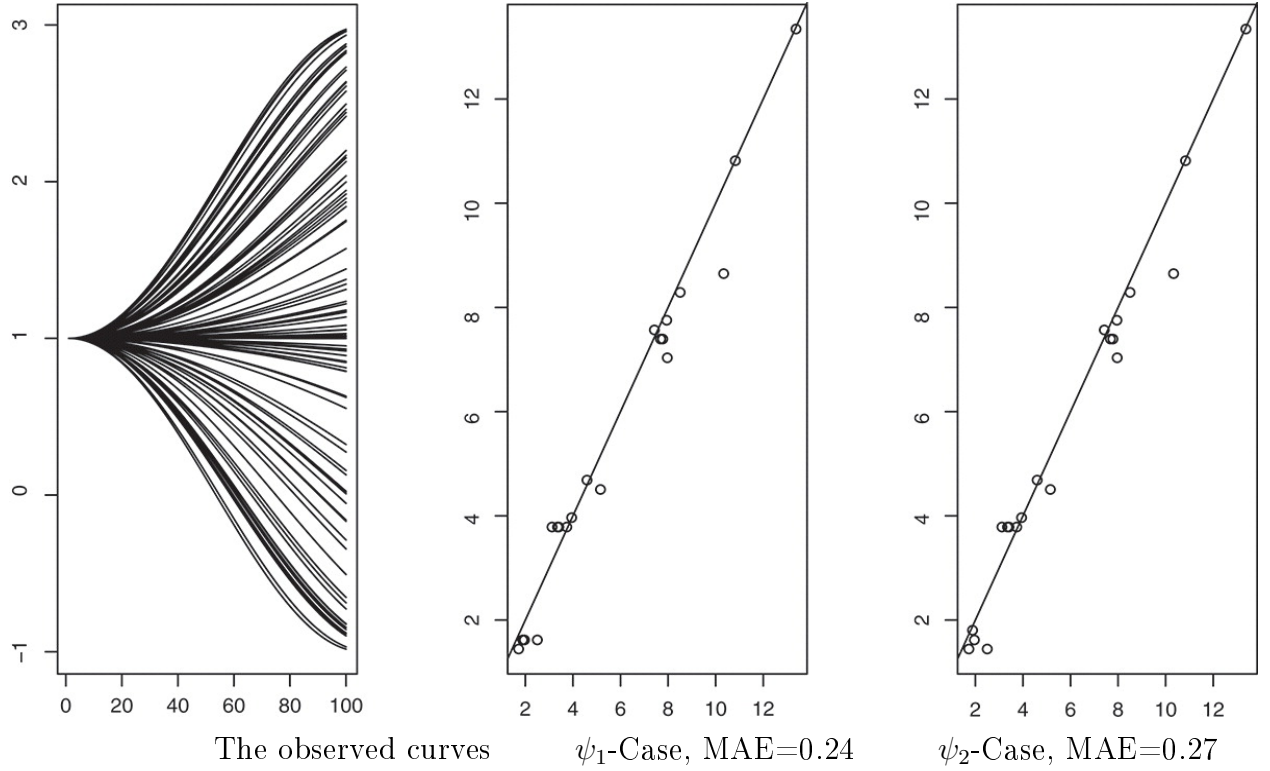
The robust regression, MAE=0.32

The classical regression, MAE=0.31

Figure 3. Case : $n = 500$

It can be seen that, the two methods are basically equivalent and both show the good behavior of our functional procedure. This fact is confirmed by the MAE error given for each cases. Moreover, although this error increase substantially with respect the sample-size n the predict results admit sufficiently good performances even for a small sample size $n = 50$. It should be noted that, our functional prediction procedure automatic selection method computes faster the results also for large sample sizes $n = 100$ or $n = 500$.

For the seconde case. We fix the sample-size $n = 100$ and we vary percentage of truncation for three values of $\mu = 0, 4$ and 8 giving 100%, 50% and 25% of observed data. Similarly to the first case, we split our data in two sub-samples (learning and test sample) we plot the predicted values versus the true values



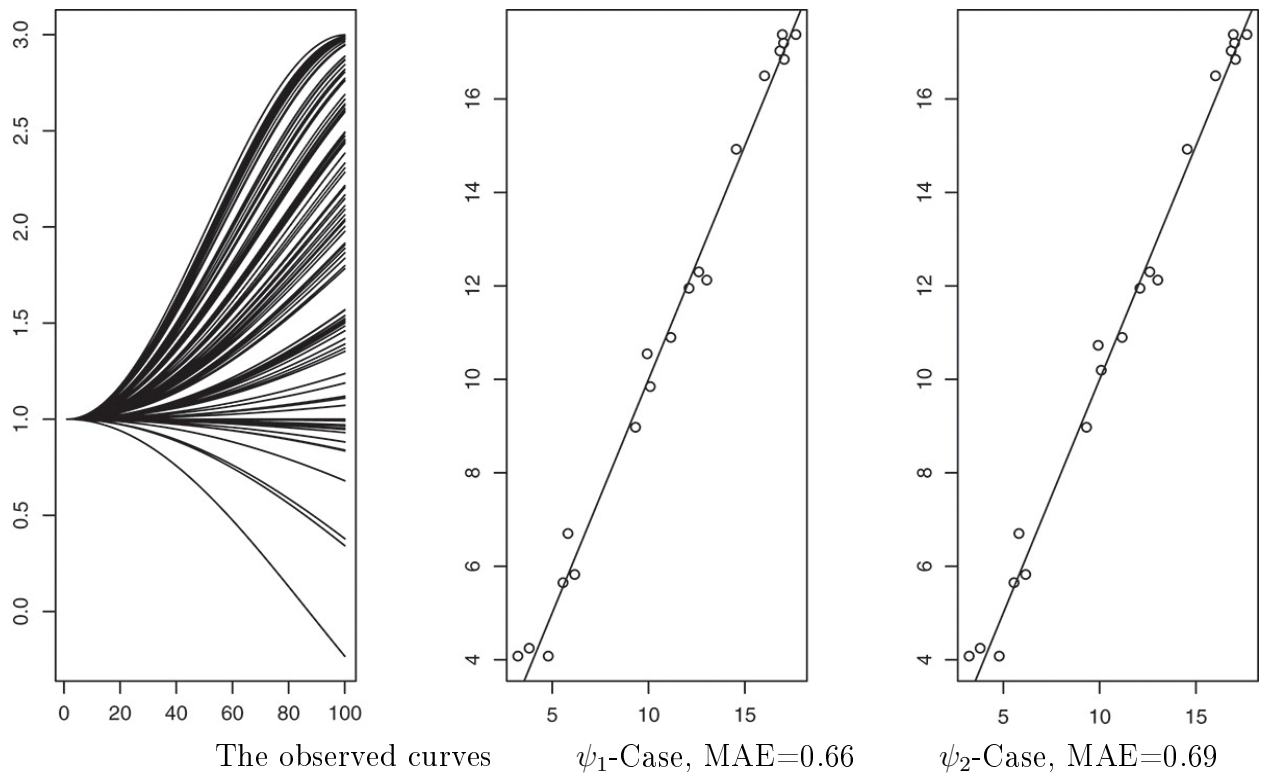


Figure 5. Case : 50% observed

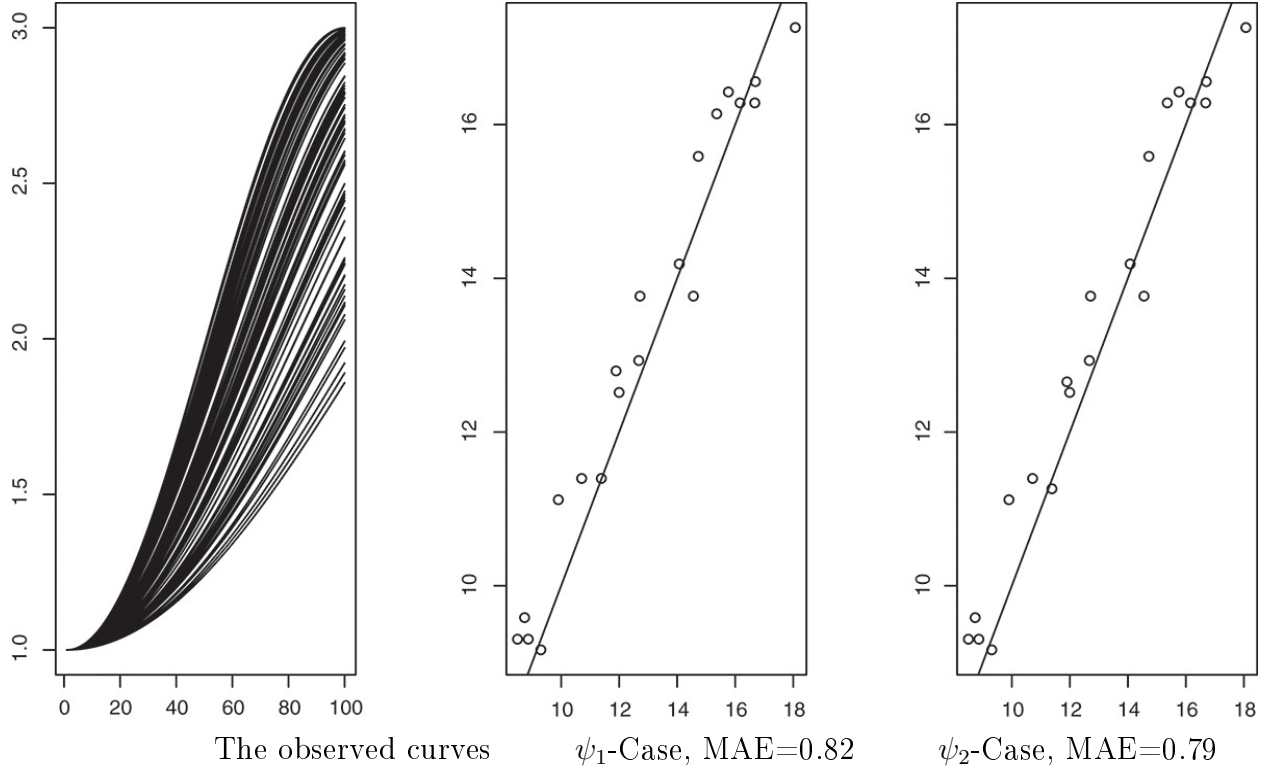


Figure 6. Case : 23% observed

The performance of both predictors seems to be less affected by the truncation percentage. The quality of prediction is dramatically destroyed in the strong truncated case compared to complete date case. However, all in all the comparison of both scatterplots indicates that the two predictors are equivalent in the three case.

For the last case. We keep the same data of the previous case. However, in order to compare the sensitivity to outliers between the robust method and the classical one we introduced artificial outliers. For this purpose, we multiple by a MF (multiplier factor) same values of the response variable in the learning sample. The obtained results are summarized in the following Table. It gives MAE error for both model for various values of $MF = 5, 10, 50, 100$

μ	MF	M.A.E.(Robust)	M.A.E.(Claassic)
$\mu = 0$ P.T=0%	1	0.27	0.24
	5	0.97	3.10
	10	1.74	4.05
	50	5.46	20.92
	100	8.74	26.72
$\mu = 4$ P.T=49%	1	0.69	0.66
	5	2.47	5.65
	10	3.42	7.72
	50	6.81	33.96
	100	10.75	46.59
$\mu = 8$ P.T=77%	1	0.79	0.82
	5	3.18	6.19
	10	7.74	13.69
	50	8.61	45.53
	100	11.81	128.01

Table 1 *MAE*-Results.

We observe that in the presence of outliers, the robust regression is more stable than the classical method, in sense that, even if the *MAE* value of the both methods increases substantially relatively to the values of MF but its variability in the robust method is not very important compared to the classical one.

3.5.2 Asymptotic normality

Our second illustration concerns the the asymptotic normality propriety of our estimate. More precisely, we examine the coverage probability (C.P.) of the proposed confidence intervals as direct consequence of the asymptotic normality. To do that, we estimate empirically α_1 and α_2 by

$$\widehat{\alpha}_1 = \frac{1}{n\phi_x(h)} \sum_{i=1}^n K(h^{-1}d(x, X_i)) \quad \text{and} \quad \widehat{\alpha}_2 = \frac{1}{n\phi_x(h)} \sum_{i=1}^n K^2(h^{-1}d(x, X_i)).$$

This last estimation is justified by the fact that,

$$\frac{1}{\phi_x(h)} \mathbb{E}[K_1^j] \rightarrow \beta_j, \quad j = 1, 2.$$

The functional operator $\Psi^{2,G}(x, \theta_x)$ and $\frac{\partial}{\partial t}\Psi(x, \theta_x)$ are respectively estimated by

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

and

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

Thus, the functions $\phi_x(\cdot)$ does not intervene in the calculation of the normalized deviation. We use the same data and the same arguments as those used in the calculation of $\hat{\theta}_x$ to compute the of the normalized deviation. Finally, we test the efficiency of our asymptotic normality by comparing the C.P. of the two models. The latter is evaluated by the percentage of the true values belong the estimated confidence intervals.

μ	$M.F$	C.P.(Robust)	C.P.(Claassic)
$\mu = 0$ P.T=0%	1	0.75	0.75
	5	0.65	0.35
	10	0.55	0.25
	50	0.55	0.1
	100	0.45	0.05
$\mu = 4$ P.T=49%	1	0.65	0.65
	5	0.5	0.15
	10	0.45	0.05
	50	0.4	0.05
	100	0.4	0
$\mu = 8$ P.T=77%	1	0.6	0.65
	5	0.5	0.1
	10	0.4	0.05
	50	0.4	0
	100	0.35	0

Table 1 *C.P.* of a confidence interval of asymptotic level 95%.

This Table presents the C.P values in both case (absence and presence of outliers). It is clear that there is no outliers the two regression models give better results. But,

in the presence of outliers the C.P. of the robust case is widely important compared to the classical one.

3.6 Appendix

Proof Theorems 4.3.1 and 4.3.2.

Similarly to Azzedine et al. (2008), the proof of both Theorems 4.3.1 and 4.3.2 is based in the fact that ψ_x is monotone w.r.t. the second component. Noting that, the proofs of both cases increasing or decreasing case can be obtained by the same manner. Thus, we give only the proof for the case of a increasing $\psi_x(Y, \cdot)$. For this case we write, $\forall \epsilon > 0$

$$\begin{aligned} \sum_n \mathbf{P} \left(\left| \hat{\theta}_x - \theta_x \right| \geq \epsilon \right) &\leq \sum_n \mathbf{P} \left(\left(\left| \hat{\theta}_x - \theta_x \right| \mathbb{I}_{\{|\hat{\theta}_x - \theta_x| \leq \delta\}} \right) \geq \epsilon \right) \\ &+ \sum_n \mathbf{P} \left(\left(\left| \hat{\theta}_x - \theta_x \right| \mathbb{I}_{|\hat{\theta}_x - \theta_x| > \delta} \right) \geq \epsilon \right). \end{aligned}$$

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

$$\begin{aligned} \sum_n \mathbf{P} \left(\left(\left| \hat{\theta}_x - \theta_x \right| \mathbb{I}_{|\hat{\theta}_x - \theta_x| > \delta} \right) \geq \epsilon \right) &\leq \sum_n \mathbf{P} \left(\left| \hat{\theta}_x - \theta_x \right| > \delta \right) \\ &\leq \sum_n \mathbf{P} \left(\left| \hat{\Psi}(x, \theta_x + \delta) - \Psi(x, \theta_x + \delta) \right| \geq \Psi(x, \theta_x + \delta) \right) \\ &+ \sum_n \mathbf{P} \left(\left| \hat{\Psi}(x, \theta_x - \delta) - \Psi(x, \theta_x - \delta) \right| \geq -\Psi(x, \theta_x - \delta) \right). \end{aligned}$$

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

$$\left(\hat{\theta}_x - \theta_x \right) \mathbb{I}_{\{|\hat{\theta}_x - \theta_x| \leq \delta\}} = \frac{\Psi(x, \hat{\theta}_x) - \hat{\Psi}(x, \hat{\theta}_x)}{\frac{\partial}{\partial t} \Psi(x, \xi_n)}$$

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

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

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

and to show that

$$\exists C > 0, \quad \frac{\partial}{\partial t} \Psi(x, \xi_n) > C \quad \text{almost surely.} \quad (3.3)$$

Unlike to Azzedine et al. (2008) we must introduce the following pseudo-estimators

$$\tilde{\Psi}_D(x) = \frac{\tau}{n\mathbf{E}[K_1]} \sum_{i=1}^n \frac{1}{G(Y_i)} K\left(\frac{d(x, X_i)}{h}\right)$$

and

$$\tilde{\Psi}_N(x, t) = \frac{\tau}{n\mathbf{E}[K_1]} \sum_{i=1}^n \frac{1}{G(Y_i)} K\left(\frac{d(x, X_i)}{h}\right) \psi(Y_i, t).$$

Moreover, we put

$$\hat{\Psi}_N(x, t) = \frac{\tau_n}{n\mathbf{E}[K_1]} \sum_{i=1}^n \frac{1}{G_n(Y_i)} K\left(\frac{d(x, X_i)}{h}\right) \psi(Y_i, t)$$

and

$$\hat{\Psi}_D(x) = \frac{\tau_n}{n\mathbf{E}[K_1]} \sum_{i=1}^n \frac{1}{G_n(Y_i)} K\left(\frac{d(x, X_i)}{h}\right).$$

Now, we consider the following decomposition

$$\begin{aligned} \Psi(x, t) - \hat{\Psi}(x, t) &= \frac{\hat{\Psi}_N(x, t)}{\hat{\Psi}_D(x)} - \Psi(x, t) \\ &= \frac{\hat{\Psi}_N(x, t)}{\hat{\Psi}_D(x)} - \frac{\Psi(x, t)}{\hat{\Psi}_D(x)} + \frac{\Psi(x, t)}{\hat{\Psi}_D(x)} - \Psi(x, t) \\ &= \frac{1}{\hat{\Psi}_D(x)} \left(\hat{\Psi}_N(x, t) - \Psi(x, t) \right) + \frac{\Psi(x, t)}{\hat{\Psi}_D(x)} \left(-\hat{\Psi}_D(x) + 1 \right) \\ &= \frac{1}{\hat{\Psi}_D(x)} \left(\hat{\Psi}_N(x, t) - \tilde{\Psi}_N(x, t) \right) + \frac{1}{\hat{\Psi}_D(x)} \left(\tilde{\Psi}_N(x, t) - \mathbf{E} \left[\tilde{\Psi}_N(x, t) \right] \right) \\ &\quad + \frac{1}{\hat{\Psi}_D(x)} \left(\mathbf{E} \left[\tilde{\Psi}_N(x, t) \right] - \Psi(x, t) \right) \\ &+ \frac{\Psi(x, t)}{\hat{\Psi}_D(x)} \left\{ \left(\tilde{\Psi}_D(x) - \hat{\Psi}_D(x) \right) + \left(\mathbf{E} \left[\tilde{\Psi}_D(x) \right] - \tilde{\Psi}_D(x) \right) + \left(-\mathbf{E} \left[\tilde{\Psi}_D(x) \right] + 1 \right) \right\} \end{aligned}$$

Thus, both Theorems are a consequence of the following intermediates results.

Lemma 3.6.1 *Under Hypotheses (H1) and (H3)-(H5), we have,*

$$\sup_{t \in [\theta_x - \delta, \theta_x + \delta]} \left| \tilde{\Psi}_N(x, t) - \mathbf{E} \left[\tilde{\Psi}_N(x, t) \right] \right| = O \left(\sqrt{\frac{\log n}{n\phi_x(h)}} \right) \quad a.co.$$

We first state the following lemmas which can be found in Ferraty and Vieu (2006). They are needed for the strong convergence of our estimates. Their proofs will then be omitted.

Lemma 3.6.2 *Let Z_i a sequence of i.i.d centred random variables such that $\forall m \geq 2, \exists C_m > 0, \mathbb{E}(|Z_1^m|) < C_m a^{2(m-1)}$ then*

$$\forall \varepsilon > 0, \mathbf{P} \left(\left| \sum_{i=1}^m Z_i \right| > \varepsilon n \right) \leq 2 \exp \left(-\frac{\varepsilon^2 n}{2a^2(1+\varepsilon)} \right)$$

Lemma 3.6.3 *Let Z_i a sequence of i.i.d centred random variables such that $|Z_i| \leq d$ and $\mathbf{E}(|Z_1^2|) < C\delta^2$ then*

$$\forall \varepsilon \in]0, \delta^2/d[\quad \mathbf{P} \left(\left| \sum_{i=1}^m Z_i \right| > \varepsilon n \right) \leq 2 \exp \left(-\frac{\varepsilon^2 n}{4\delta^2} \right)$$

Proof of Lemma 3.6.1

For $l_n = n^{-1/2b_2}$ we cover the compact by $[\theta_x - \delta, \theta_x + \delta] \subset \bigcup_{j=1}^{d_n} (t_j - l_n, t_j + l_n)$ with $d_n = O(n^{1/2b_2})$. We use the monotony of ψ_x to show that

$$\sup_{t \in [\theta_x - \delta, \theta_x + \delta]} \left| \tilde{\Psi}_N(x, t) - \mathbf{E} \left[\tilde{\Psi}_N(x, t) \right] \right| \leq \max_{1 \leq j \leq d_n} \max_{t \in \{t_j - l_n, t_j + l_n\}} \left| \tilde{\Psi}_N(x, t) - \mathbf{E} \left[\tilde{\Psi}_N(x, t) \right] \right| + 2^{b_2} C_2 l_n^{b_2}.$$

Since $l_n = n^{-1/2b_2}$ then

$$l_n^{b_2} = O \left(\sqrt{\frac{\log n}{n\phi_x(h)}} \right).$$

Therefore, all it remains to show that

$$\max_{1 \leq j \leq d_n} \max_{t \in \{t_j - l_n, t_j + l_n\}} \left| \tilde{\Psi}_N(x, t) - \mathbf{E} \left[\tilde{\Psi}_N(x, t) \right] \right| = O_{a.co.} \left(\sqrt{\frac{\log n}{n\phi_x(h)}} \right). \quad (3.4)$$

To do that, we denote by $\mathcal{G}_n$ the subset of the intervals extremities grid

$$\mathcal{G}_n = \{t_j - l_n, t_j + l_n, 1 \leq j \leq d_n\}. \quad (3.5)$$

and we use the fact that

$$\mathbf{P} \left(\max_{t \in \mathcal{G}_n} \left| \tilde{\Psi}_N(x, t) - \mathbf{E} \left[\tilde{\Psi}_N(x, z) \right] \right| > \varepsilon \right) \leq \sum_{t \in \mathcal{G}_n} \mathbf{P} \left(\left| \tilde{\Psi}_N(x, t) - \mathbf{E} \left[\tilde{\Psi}_N(x, t) \right] \right| > \varepsilon \right).$$

Next, we apply the exponential inequality given in Lemma 3.6.3 to evaluate the quantities $\mathbf{P} \left(\left| \tilde{\Psi}_N(x, t) - \mathbf{E} \left[\tilde{\Psi}_N(x, t) \right] \right| > \varepsilon \right)$. Indeed, for a fixed $t \in \mathcal{G}_n$, we write

$$\tilde{\Psi}_N(x, t) - \mathbf{E} \left[\tilde{\Psi}_N(x, t) \right] = \frac{1}{n} \sum_{i=1}^n Z_i$$

where

$$Z_i = \frac{1}{\mathbf{E}[K_1]} \frac{\tau}{G(Y_i)} K \left(\frac{d(x, X_i)}{h} \right) \psi(Y_i, t) - \mathbf{E} \left[\frac{\tau}{G(Y_i)} \frac{1}{\mathbf{E}[K_1]} K \left(\frac{d(x, X_i)}{h} \right) \psi(Y_i, t) \right]$$

The application of the mentioned inequality is based on evaluation of $\mathbf{E}(\|Z_1^m\|)$ for all $m \geq 2$. The latter is evaluated by the same arguments used by Ferraty and Vieu (2006). Indeed, by Newton binomial expansion

$$\begin{aligned} \mathbf{E}[\|Z_1^m\|] &= \frac{1}{\mathbf{E}[K_1]^m} \sum_{k=1}^m C_m^k \mathbf{E} \left[\frac{\tau^k}{G^k(Y_1)} K^k \left(\frac{d(x, X_1)}{h} \right) \psi^k(Y_1, t) \right] \\ &\quad \times \left(\mathbf{E} \left[\frac{\tau}{G(Y_1)} K \left(\frac{d(x, X_1)}{h} \right) \psi(Y_1, t) \right] \right)^{m-k} \end{aligned}$$

Observe that, under (H3) and (H4), we have, for all $k > 0$

$$\mathbf{E} \left[\frac{\tau}{G(Y_1)} K \left(\frac{d(x, X_1)}{h} \right) \psi(Y_1, t) \right]^k \leq C \mathbf{E} \left[\frac{\tau}{G(Y_1)} K \left(\frac{d(x, X_1)}{h} \right) \right]^k \leq \frac{C \tau^k}{G^k(a_F)} \mathbf{E}[K_1].$$

and

$$\left(\mathbf{E} \left[\frac{\tau}{G(Y_1)} K \left(\frac{d(x, X_1)}{h} \right) \psi(Y_1, z) \right] \right)^{m-k} \leq C \left(\mathbf{E} \left[K \left(\frac{d(x, X_i)}{h} \right) \right] \right)^{m-k} \leq C (\mathbf{E}[K_1]^{m-k}).$$

It follows that

$$\mathbf{E}[\|Z_1^m\|] \leq C \mathbf{E}[K_1]^{-m+1}.$$

Then, we can apply Lemma 3.6.2 with $a = \sqrt{\mathbf{E}[K_1]^{-1}}$. Thus,

$$\mathbf{P} \left(\left| \tilde{\Psi}_N(x, t) - \mathbf{E} \left[\tilde{\Psi}_N(x, t) \right] \right| > \varepsilon \right) = \mathbf{P} \left(\left| \sum_{i=1}^m Z_i \right| > \varepsilon n \right) \leq 2 \exp \left(-\frac{\varepsilon^2 n}{2 \mathbf{E}[K_1] (1 + \varepsilon)} \right).$$

Consequently, for $\varepsilon = \varepsilon_0 \sqrt{\frac{\log n}{n \mathbf{E}[K_1]}}$ we get

$$\mathbf{P} \left(\left| \tilde{\Psi}_N(x, t) - \mathbf{E} \left[\tilde{\Psi}_N(x, t) \right] \right| > \varepsilon \right) \leq 2 \exp \left(-\frac{\varepsilon_0^2 n \frac{\log n}{n \mathbf{E}[K_1]}}{\frac{2}{\mathbf{E}[K_1]} \left(1 + \varepsilon_0 \sqrt{\frac{\log n}{n \mathbf{E}[K_1]}} \right)} \right)$$

$$\leq 2 \exp \left(- \frac{\varepsilon_0^2 \log n}{2 \left(1 + \varepsilon_0 \sqrt{\frac{\log n}{n \mathbf{E}[K_1]}} \right)} \right).$$

Furthermore, using the fact that $\mathbf{E}[K_1] = O(\phi_x(h))$ we obtain

$$\mathbf{P} \left(\left| \tilde{\Psi}_N(x, t) - \mathbf{E} \left[\tilde{\Psi}_N(x, t) \right] \right| > \varepsilon \right) \leq 2 \exp \left(-C \varepsilon_0^2 \log n \phi_x(h) \right) \leq 2n^{-C \varepsilon_0^2}.$$

It follows that

$$\sum_{t \in \mathcal{G}_n} \mathbf{P} \left(\left| \tilde{\Psi}_N(x, t) - \mathbf{E} \left[\tilde{\Psi}_N(x, t) \right] \right| > \varepsilon \right) \leq 4d_n n^{-C \varepsilon_0^2} = C n^{-C \varepsilon_0^2 + 1/2b_2}.$$

Finally, an appropriate choice of ε_0 and Borel-Cantelli Lemma's give

$$\sup_{t \in [\theta_x - \delta, \theta_x + \delta]} \left| \tilde{\Psi}_N(x, t) - \mathbf{E} \left[\tilde{\Psi}_N(x, t) \right] \right| = O \left(\sqrt{\frac{\log n}{n \phi_x(h)}} \right) \quad a.co.$$

which completes the proof. ■

Lemma 3.6.4 *Under Hypotheses (H1), (H2)((i)-(ii)), (H4) and (H5), we have,*

$$\sup_{t \in [\theta_x - \delta, \theta_x + \delta]} \left| \mathbf{E} \left[\tilde{\Psi}_N(x, t) \right] - \tau \Psi(x, t) \right| = o(1).$$

Furthermore if we add (H2)(iii), we have

$$\sup_{t \in [\theta_x - \delta, \theta_x + \delta]} \left| \mathbf{E} \left[\tilde{\Psi}_N(x, t) \right] - \tau \Psi(x, t) \right| = O(h^{b_1}).$$

Proof of Lamma 3.6.4.

By using truncation effect, we have for all $t \in [\theta_x - \delta, \theta_x + \delta]$

$$\begin{aligned} \mathbf{E}[\tilde{\Psi}_N(x, t)] &= \frac{\tau}{\mathbf{E}[K_1]} \mathbf{E} \left[\frac{1}{G(Y_1)} K \left(\frac{d(x, X_1)}{h} \right) \psi(Y_1, t) \right] \\ &= \frac{\tau}{\mathbf{E}[K_1]} \int \int \frac{K \left(\frac{d(x, u)}{h} \right) \psi(y, t)}{G(y)} \mathbf{f}^*(y|u) dy d\mathbf{P}_x^X(u) \\ &= \frac{\tau}{\mathbf{E}[K_1]} \int \int \frac{G(y) K \left(\frac{d(x, u)}{h} \right) \psi(y, t)}{\tau G(y)} \mathbf{f}(y|u) dy d\mathbf{P}_x^X(u) \\ &= \frac{1}{\mathbf{E}[K_1]} \int \int K \left(\frac{d(x, u)}{h} \right) \psi(y, t) \mathbf{f}(y|u) dy d\mathbf{P}_x^X(u) \\ &= \frac{1}{\mathbf{E}[K_1]} \mathbf{E} \left[K \left(\frac{d(x, X_1)}{h} \right) \psi(Y_1, t) \right] \end{aligned}$$

$$\begin{aligned}
&= \frac{1}{\mathbf{E}[K_1]} \mathbf{E} \left[\mathbf{E} \left(K \left(\frac{d(x, X_1)}{h} \right) \mathbb{I}_{B(x,h)} \psi(Y_1, t) | X_1 \right) \right] \\
&= \frac{1}{\mathbf{E}[K_1]} \mathbf{E} \left[K \left(\frac{d(x, X_1)}{h} \right) \mathbb{I}_{B(x,h)} \mathbf{E}(\psi(Y_1, t) | X_1) \right] \\
&= \frac{1}{\mathbf{E}[K_1]} \mathbf{E} \left[K \left(\frac{d(x, X_1)}{h} \right) \mathbb{I}_{B(x,h)} \Psi(X_1, t) \right].
\end{aligned}$$

Therefore, for all $t \in [\theta_x - \delta, \theta_x + \delta]$, with (H2) (ii)

$$| \mathbf{E}[\tilde{\Psi}_N(x, t)] - \Psi(x, t) | = o(1).$$

On the other hand, if we added (H2) (iii), we have

$$\begin{aligned}
| \mathbf{E}[\tilde{\Psi}_N(x, t)] - \Psi(x, t) | &\leq \frac{\tau}{\mathbf{E}[K_1]} \mathbf{E} \left[K_1 \left(\frac{d(x, X_i)}{h} \right) \mathbb{I}_{B(x,h)} | \Psi(X_1, t) - \Psi(x, t) | \right] \\
&\leq \mathbf{E} \left[K_1 \left(\frac{d(x, X_i)}{h} \right) \mathbb{I}_{B(x,h)} C_1 d^b(X_1, x) \right] \\
&\leq C_1 h^{b_1}.
\end{aligned}$$

The proof of this Lemma is finished. ■

Lemma 3.6.5 *Under Hypotheses (H1) and (H3)-(H5), we have,*

$$\sup_{t \in [\theta_x - \delta, \theta_x + \delta]} \left| \hat{\Psi}_N(x, t) - \tilde{\Psi}_N(x, t) \right| = O \left(\sqrt{\frac{1}{n}} \right) \quad a.s.$$

Proof of Lemma 3.6.5.

We write for all $t \in [\theta_x - \delta, \theta_x + \delta]$

$$\begin{aligned}
\left| \hat{\Psi}_N(x, t) - \tilde{\Psi}_N(x, t) \right| &= \left| \frac{\tau_n}{n \mathbf{E}[K_1]} \sum_{i=1}^n \frac{1}{G_n(Y_i)} K \left(\frac{d(x, X_i)}{h} \right) \psi(Y_i, t) \right. \\
&\quad \left. - \frac{\tau}{n \mathbf{E}[K_1]} \sum_{i=1}^n \frac{1}{G(Y_i)} K \left(\frac{d(x, X_i)}{h} \right) \psi(Y_i, t) \right| \\
&= \left| \frac{\tau_n - \tau}{n \mathbf{E}[K_1]} \sum_{i=1}^n \frac{1}{G_n(Y_i)} K \left(\frac{d(x, X_i)}{h} \right) \psi(Y_i, t) \right. \\
&\quad \left. + \tau \left(\frac{G_n(Y) - G(Y)}{G_n(Y)G(Y)} \right) \frac{1}{n \mathbf{E}[K_1]} \sum_{i=1}^n K \left(\frac{d(x, X_i)}{h} \right) \psi(Y_i, t) \right|
\end{aligned}$$

$$\leq \left[\frac{|\tau_n - \tau|}{G_n(a_F)} + \tau \frac{\sup_{y \in [a, b]} |G_n(y) - G(y)|}{G_n(a_F)G(a_F)} \right] \\ \times \frac{1}{n\mathbf{E}[K_1]} \sum_{i=1}^n K\left(\frac{d(x, X_i)}{h}\right) \psi(Y_i, t).$$

Concerning the first part of the last term we recall that, it is shown in Theorem 3.2 of He and Yang(1998) $|\tau_n - \tau| = O_{a.s.}(n^{-\frac{1}{2}})$, while remark 6 of Woodroffe(1985)) gives $|G_n(a_F) - G(a_F)| = O_{a.s.}(n^{-\frac{1}{2}})$. Furthermore, for the second part we follow the same line of the proof of Lemma 3.6.1 to show that

$$\sup_{t \in [\theta_x - \delta, \theta_x + \delta]} \frac{1}{n\mathbf{E}[K_1]} \sum_{i=1}^n \left[K\left(\frac{d(x, X_i)}{h}\right) \psi(Y_i, t) - \mathbf{E} \left[K\left(\frac{d(x, X_i)}{h}\right) \psi(Y_i, t) \right] \right] = O\left(\sqrt{\frac{\log n}{n\phi_x(h)}}\right)$$

and we combine the latter with Lemma 3.1 in Azzedine *et al.* (2008) to get

$$\sup_{t \in [\theta_x - \delta, \theta_x + \delta]} \frac{1}{n\mathbf{E}[K_1]} \sum_{i=1}^n K\left(\frac{d(x, X_i)}{h}\right) \psi(Y_i, t) = O_{a.co.}(1).$$

So, we have

$$\left| \widehat{\Psi}_N(x, t) - \widetilde{\Psi}_N(x, t) \right| = O_{a.s.}(n^{-\frac{1}{2}})$$

which is the claimed result ■

Lemma 3.6.6 *Under Hypotheses (H1), (H2), (H4) and (H5), we have,*

$$\widetilde{\Psi}_D(x) - \mathbf{E} [\widetilde{\Psi}_D(x)] = O\left(\sqrt{\frac{\log n}{n\phi_x(h)}}\right) \quad a.co.$$

Proof of Lemma 3.6.6.

The proof of this lemma is very close to Lemma 3.6.1. It's based on the exponential inequality given in Lemma 3.6.2 and the following consideration

$$\widetilde{\Psi}_D(x) - \mathbf{E} [\widetilde{\Psi}_D(x)] = \frac{1}{n} \sum_{i=1}^n L_i$$

where

$$L_i = \frac{1}{\mathbf{E}[K_1]} \frac{\tau}{G(Y_i)} K\left(\frac{d(x, X_i)}{h}\right) - \frac{1}{\mathbf{E}[K_1]} \mathbf{E} \left[\frac{\tau}{G(Y_i)} K\left(\frac{d(x, X_i)}{h}\right) \right].$$

Similarly to Lemma 3.6.1, we obtain by (H4)

$$|L_i| < \frac{C}{\mathbf{E}[K_1]} \quad \text{and} \quad \mathbf{E} [L_i^2] < \frac{C'}{\mathbf{E}[K_1]}.$$

So, we apply the exponential inequality of Lemma 3.6.2 to get for all $\eta_0 > 0$

$$\mathbf{P} \left(\left| \tilde{\Psi}_D(x) - \mathbf{E} [\tilde{\Psi}_D(x)] \right| > \eta_0 \sqrt{\frac{\log n}{n \mathbf{E}(K_1)}} \right) \leq 2 \exp\{-'n^{-C\eta_0^2 \log n}\}.$$

Because of (H1) and (H4) we have $\mathbf{E}(K_1) = O(\phi_x(h))$. Thus, there exist an $\eta_0 > 0$ such that

$$\sum_n \mathbf{P} \left(\left| \tilde{\Psi}_D(x) - \mathbf{E} [\tilde{\Psi}_D(x)] \right| > \eta_0 \sqrt{\frac{\log n}{n \phi_x(h)}} \right) < \infty.$$

The last result complete the proof of lemma. ■

Lemma 3.6.7 *Under Hypotheses (H1),(H4) and (H5), we have,*

$$\mathbf{E} [\tilde{\Psi}_D(x)] = 1.$$

Proof of Lemma 7.

By the same fashion of Lemma 3.6.4, we write

$$\begin{aligned} \mathbf{E}[\tilde{\Psi}_D(x)] &= \frac{\tau}{\mathbf{E}[K_1]} \mathbf{E} \left[\mathbf{E} \left[K \left(\frac{d(x, X_1)}{h} \right) \frac{\mathbb{1}_{Y_1 \geq T_1}}{\tau G(Y_1)} | X_1, Y_1 \right] \right] \\ &= \frac{1}{\mathbf{E}[K_1]} \mathbf{E} \left[K \left(\frac{d(x, X_1)}{h} \right) \right] = 1. \end{aligned}$$
■

Lemma 3.6.8 *Under Hypotheses (H1),(H4) and (H5), we have,*

$$\hat{\Psi}_D(x) - \tilde{\Psi}_D(x) = O \left(\sqrt{\frac{1}{n}} \right) \quad a.s.$$

Proof of Lemma 3.6.8.

The proof of this lemma follows the same lines of Lemma 3.6.1. Indeed, it suffices to write,

$$\begin{aligned} \left| \hat{\Psi}_D(x) - \tilde{\Psi}_D(x) \right| &= \left| \frac{\tau_n}{n \mathbf{E}[K_1]} \sum_{i=1}^n \frac{1}{G_n(Y_i)} K \left(\frac{d(x, X_i)}{h} \right) \right. \\ &\quad \left. - \frac{\tau}{n \mathbf{E}[K_1]} \sum_{i=1}^n \frac{1}{G(Y_i)} K \left(\frac{d(x, X_i)}{h} \right) \right| \\ &\leq \left[\frac{|\tau_n - \tau|}{G_n(a_F)} + \tau \frac{\sup_{y \in [a, b]} |G_n(y) - G(y)|}{G_n(a_F) G(a_F)} \right] \end{aligned}$$

$$\times \frac{1}{n\mathbf{E}[K_1]} \sum_{i=1}^n K\left(\frac{d(x, X_i)}{h}\right).$$

Once again we use the result of Theorem 3.2 of He and Yang(1998) and remark 6 of Woodroffe(1985)) where we have, respectively, $|\tau_n - \tau| = O_{a.s.}(n^{-\frac{1}{2}})$ and $|G_n(a_F) - G(a_F)| = O_{a.s.}(n^{-\frac{1}{2}})$. While the quantity $\frac{1}{n\mathbf{E}[K_1]} \sum_{i=1}^n K\left(\frac{d(x, X_i)}{h}\right)$ is bounded (it can be proved by adding and substracting its expectation). Finally, we have

$$\left| \widehat{\Psi}_D(x) - \widetilde{\Psi}_D(x) \right| = O_{a.s.} \left(\sqrt{\frac{1}{n}} \right)$$

which is the claimed result. ■

Corollary 6 *Under Assumptions (H1), (H2) ((i)-(ii)), (H3)-(H5) and if $\Psi'(x, \theta_x) \neq 0$, then $\widehat{\theta}_x$ exists a.s. for all sufficiently large n such that*

$$\exists C > 0 \quad \frac{\partial}{\partial t} \Psi(x, \xi_n) > C.$$

Proof of Corollary 6.

It is clear that, if $\psi_x(Y, \cdot)$ is increasing, we have for all $\epsilon > 0$

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

The results of Lemmas 3.6.1-3.6.6 show that

$$\widehat{\Psi}(x, t) \longrightarrow \Psi(x, t) \quad \text{a. s.}$$

for all real fixed $t \in [\theta_x - \delta, \theta_x + \delta]$. So, for sufficiently large n and for all $\epsilon \leq \delta$

$$\widehat{\Psi}(x, \theta_x - \epsilon) \leq 0 \leq \widehat{\Psi}(x, \theta_x + \epsilon) \quad \text{a. s.}$$

Since $\widehat{\Psi}(x, t)$ is a continuous function of t , there exists a $\widehat{\theta}_x \in [\theta_x - \epsilon, \theta_x + \epsilon]$ such that $\widehat{\Psi}(x, \widehat{\theta}_x) = 0$. Concerning, the uniqueness of $\widehat{\theta}_x$, we point out that the latter is a direct consequence of the strict monotonicity of ψ_x , w.r.t. the second component and the positivity of K .

Finally, The second part of this corollary is a direct consequence of the regularity assumption (H2) (i) on $\Psi(x, \cdot)$ and the convergence

$$\widehat{\theta}_x \rightarrow \theta_x \quad n \longrightarrow \infty \quad \text{a. s.}$$

■

Proof of Theorem 3.3.3

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, if $\widehat{\Psi}_D(x)$ is not equal to zero, the definition of the estimator by (4.4) is equivalent to

$$\widehat{\Psi}_N(x, \widehat{\theta}_x) = 0.$$

Hence, if $\widehat{\Psi}_D(x) \neq 0$ we can write

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

Therefore, Theorem 3.3.3 is a consequence of Lemma 3.6.1 and the following intermediates results.

Lemma 3.6.9 *Under Hypotheses (H1)-(H3) and (H5), we have*

$$\mathbf{P} \left\{ \left(\widehat{\Psi}_D(x) = 0 \right) \right\} \longrightarrow 0 \quad \text{as } n \longrightarrow \infty.$$

Proof of Lemma 3.6.9.

Since the X_i 's are independent we get, under (H1), (H3) and (H4')

$$\begin{aligned} \mathbf{P} \left\{ \widehat{\Psi}_D(x) = 0 \right\} &= \mathbf{P} \left(\bigcap_{i=1}^n \{ \omega \in \Omega \text{ such that } X_i(\omega) \notin B(x, h) \} \right) \\ &= (1 - \phi_x(h))^n \leq \exp(-n\phi_x(h)) \rightarrow 0. \end{aligned} \quad \blacksquare$$

Lemma 3.6.10 *Under the Hypotheses of Theorem 3.3.3, we have for any $x \in \mathcal{A}$*

$$\left(\frac{n\phi_x(h)}{\left(\frac{\partial}{\partial t} \Psi(x, \theta_x) \right)^2 \sigma^2(x, \theta_x)} \right)^{1/2} \left(\widetilde{\Psi}_N(x, z) - \mathbf{E} [\widetilde{\Psi}_N(x, z)] \right) \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1) \quad \text{as } n \rightarrow \infty.$$

Proof of Lemma 3.6.10

Firstly we show the following limit

$$n\phi_x(h) \text{Var} [\widetilde{\Psi}_N(x, z)] \rightarrow \left(\frac{\partial}{\partial t} \Psi(x, \theta(x)) \right)^2 \sigma^2(x, \theta(x)). \quad (3.6)$$

Definition of $\tilde{\Psi}_N(x, z)$ allows to write that

$$\begin{aligned} \text{Var} \left[\tilde{\Psi}_N(x, t) \right] &= \frac{\tau^2}{(n \mathbf{E}[K_1])^2} \sum_{i=1}^n \text{Var} \left[\frac{K_i \psi(Y_i, t)}{G(Y_i)} \right] \\ &= \frac{\tau^2}{n \mathbf{E}^2[K_1]} \text{Var} \left[\frac{K_1 \psi(Y_1, t)}{G(Y_1)} \right] \\ &= \frac{\tau^2}{n \mathbf{E}^2[K_1]} \mathbf{E} \left[\frac{K_1^2 \psi^2(Y_1, t)}{G^2(Y_1)} \right] - \frac{\tau^2}{n \mathbf{E}^2[K_1]} \left(\mathbf{E} \left[\frac{K_1 \psi(Y_1, t)}{G(Y_1)} \right] \right)^2. \end{aligned} \quad (3.7)$$

Now, using the truncation effect to the first term of (3.7), we have

$$\begin{aligned} \mathbf{E} \left[\frac{K_1^2 \psi^2(Y_1, t)}{G^2(Y_1)} \right] &= \int \int \frac{K^2 \left(\frac{d(x, u)}{h} \right) \psi^2(y, t)}{G^2(y)} \mathbf{f}^*(y|u) dy d\mathbf{P}_x^X(u) \\ &= \int \int \frac{G(y) K^2 \left(\frac{d(x, u)}{h} \right) \psi^2(y, t)}{\tau G^2(y)} \mathbf{f}(y|u) dy d\mathbf{P}_x^X(u) \\ &= \frac{1}{\tau} \mathbf{E} \left[\frac{K_1^2 \psi^2(Y_1, t)}{G(Y_1)} \right]. \\ &= \frac{\tau \mathbf{E}[K_1^2]}{n \mathbf{E}^2[K_1]} \mathbf{E} \left[\frac{K_1^2 \psi^2(Y_1, t) \mathbb{I}_{Y_1 \geq T_1}}{G^2(Y_1) \mathbf{E}[K_1^2]} \right] - \frac{1}{n} \left(\mathbf{E} \left[\frac{K_1 \psi(Y, t) \mathbb{I}_{Y_1 \geq T_1}}{G(Y_1) \mathbf{E}[K_1]} \right] \right)^2. \end{aligned} \quad (3.8)$$

So,

$$\begin{aligned} n \phi_x(h) \text{Var} \left[\hat{\Psi}_N(x, z) \right] &= \frac{\phi_x(h) \tau \mathbf{E}[K_1^2]}{\mathbf{E}^2[K_1]} \left(\mathbf{E} \left[\frac{K_1^2 \mathbb{I}_{Y_1 \geq T_1} (\psi(Y_1, z))^2}{G^2(Y_1) \mathbf{E}[K_1^2]} \right] \right) \\ &\quad - \phi_x(h) \left(\mathbf{E} \left[\frac{K_1 \mathbb{I}_{Y_1 \geq T_1} \psi(Y_1, z)}{G(Y_1) \mathbf{E}[K_1]} \right] \right)^2. \end{aligned} \quad (3.9)$$

Noting that the limit of the second term of the right hand side of (3.9) has been treated in Lemma 3.6.4 where we have shown

$$\forall t \in \mathbb{R}, \quad \mathbf{E} \left[\frac{K_1 \mathbb{I}_{Y_1 \geq T_1} \psi(Y, t)}{G(Y_1) \mathbf{E}[K_1]} \right] \longrightarrow \Psi(x, t). \quad (3.10)$$

Then,

$$\phi_x(h) \left(\mathbf{E} \left[\frac{K_1 \mathbb{I}_{Y_1 \geq T_1} \psi(Y, t)}{G(Y_1) \mathbf{E}[K_1]} \right] \right) \longrightarrow 0.$$

Now, we compute the limit of the first term of the right hand side of (3.9). Clearly, under (H2)

$$\mathbf{E} \left[\frac{K_1^2 \mathbb{I}_{Y_1 \geq T_1} \psi_x^2(Y, t)}{G^2(Y_1) \mathbf{E}[K_1^2]} \right] = \frac{1}{\mathbf{E}[K_1^2]} \mathbf{E} \left[K^2 \left(\frac{d(x, X_1)}{h} \right) \frac{\mathbb{I}_{Y_1 \geq T_1}}{G^2(Y_1)} \psi^2(Y_1, t) \right]$$

$$\begin{aligned}
&= \frac{1}{\mathbf{E}[K_1^2]} \mathbf{E} \left[\mathbf{E} \left[K^2 \left(\frac{d(x, X_1)}{h} \right) \frac{\mathbb{I}_{Y_1 \geq T_1}}{G^2(Y_1)} \psi^2(Y_1, t) | X_1, Y_1 \right] \right] \\
&= \frac{1}{\mathbf{E}[K_1^2]} \mathbf{E} \left[K^2 \left(\frac{d(x, X_1)}{h} \right) \frac{1}{G(Y_1)} \psi^2(Y_1, t) \right] \\
&= \frac{1}{\mathbf{E}[K_1^2]} \mathbf{E} \left[K \left(\frac{d(x, X_1)}{h} \right) \mathbb{I}_{B(x, h)} \Psi^{2, G}(X_1, t) \right].
\end{aligned}$$

The continuity of $\Psi^{2, G}(\cdot, \cdot)$ ensure that

$$\mathbf{E} \left[\frac{K_1^2 \mathbb{I}_{Y_1 \geq T_1} \psi_x^2(Y, t)}{G^2(Y_1) \mathbf{E}[K_1^2]} \right] = \Psi^{2, G}(x, \theta_x) + o(1).$$

We return to Ferraty *et al.* (2007) for the asymptotic expression of

$$\frac{1}{\phi_x(h)} E[K_1^j] = \alpha_j + o(1); \quad j = 1, 2. \quad (3.11)$$

Therefore

$$n\phi_x(h) \text{Var} \left[\tilde{\Psi}_N(x, t) \right] \rightarrow \left(\frac{\partial}{\partial t} \Psi(x, \theta(x)) \right)^2 \sigma^2(x, \theta(x)).$$

Let's now prove the claimed result. The latter is based on the version of the central limit Theorem given in Loève (1963, p. 275). To do that, we write

$$\tilde{\Psi}_N(x, t) - \mathbf{E} \left[\tilde{\Psi}_N(x, t) \right] = \frac{1}{n} \sum_{i=1}^n Z_i$$

where

$$Z_i = \frac{1}{\mathbf{E}[K_1]} \frac{\tau}{G(Y_i)} K \left(\frac{d(x, X_i)}{h} \right) \psi(Y_i, t) - \mathbf{E} \frac{\tau}{G(Y_i)} \frac{1}{\mathbf{E}[K_1]} K \left(\frac{d(x, X_i)}{h} \right) \psi(Y_i, t).$$

It is enough to show that for some $\delta > 0$

$$\frac{\sum_{i=1}^n \mathbf{E} [|Z_i - \mathbf{E}[Z_i]|^{2+\delta}]}{\left(\text{Var} \left(\sum_{i=1}^n Z_i \right) \right)^{(2+\delta)/2}} \rightarrow 0.$$

Clearly, the above calculations show that $n\phi_x(h) \text{Var} \left(\frac{1}{n} \sum_{i=1}^n Z_i \right)$ converges to $\left[\frac{\partial}{\partial t} \Psi(x, \theta(x)) \right]^2 \sigma^2(x, \theta(x))$ as n tends to infinity. Therefore, to conclude the proof of this Lemma, it is enough

to show that the numerator of the above expression converges to 0. For this, we use the C_r -inequality (see Loève (1963), p. 155) we show that,

$$\begin{aligned}
& (n\phi_x(h))^{(1+\delta/2)} \sum_{i=1}^n \mathbf{E} \left[\left| Z_i - \mathbf{E}[Z_i] \right|^{2+\delta} \right] \\
& \leq 2^{1+\delta} (n\phi_x(h))^{(2+\delta)/2} \sum_{i=1}^n \mathbf{E} [|Z_i|^{2+\delta}] \\
& + 2^{1+\delta} (n\phi_x(h))^{(2+\delta)/2} \sum_{i=1}^n |\mathbf{E}[Z_i]|^{2+\delta}. \tag{3.12}
\end{aligned}$$

Observe that, according to (3.11), we have, for all $j > 0$, $\mathbf{E}[K_1^j] = O(\phi(h))$, then, because G^{-1} is bounded and $\mathbf{E}[\psi^{2+\delta}(Y, t)|X] < \infty$, we have

$$\begin{aligned}
2^{1+\delta} (n\phi_x(h))^{(2+\delta)/2} \sum_{i=1}^n \mathbf{E} [|Z_i|^{2+\delta}] &= 2^{1+\delta} n^{-\delta/2} (\phi(h))^{-1-\delta/2} \left(\frac{\phi(h)}{\mathbf{E}[K_1]} \right)^{2+\delta} \mathbf{E}[K_1^{2+\delta} \mathbf{E}[\psi^{2+\delta}(Y, t)|X]] \\
&\leq C(n\phi_x(h))^{-\delta/2} \left(\frac{\phi(h)}{\mathbf{E}[K_1]} \right)^{2+\delta} \left(\mathbf{E}[K_1^{2+\delta}] / \phi_x(h) \right) \rightarrow 0.
\end{aligned}$$

The last limit is a consequence of Assumption $(H5')$.

For the second term of (3.12) we use once again the fact that $\mathbf{E}[\psi(Y, t)|X] < \infty$ to get

$$\begin{aligned}
2^{1+\delta} (n\phi(h))^{(2+\delta)/2} \sum_{i=1}^n |\mathbf{E}[Z_i]|^{2+\delta} &\leq 2^{1+\delta} n^{-\delta/2} (\phi(h))^{(2+\delta)/2} \mathbf{E}^{-2-\delta}[K_1] \left| \mathbf{E}[K_1 \psi(Y_1, t)] \right|^{2+\delta} \\
&\leq C n^{-\delta/2} (\phi(h))^{(2+\delta)/2} \rightarrow 0,
\end{aligned}$$

which completes the proof. ■

Lemma 3.6.11 *Under hypotheses $(H1')$, $(H3)$ and $(H6)$ we have*

$$\left(\frac{n\phi_x(h)}{\left(\frac{\partial}{\partial t} \Psi(x, \theta(x)) \right)^2 \sigma^2(x, \theta_x)} \right)^{1/2} \mathbf{E} [\tilde{\Psi}_N(x, z)] = u + o(1), \text{ as } n \rightarrow +\infty.$$

Proof of Lemma 3.6.11

Observe that in Lemma 3.6.4 we have shown that

$$\mathbf{E} [\tilde{\Psi}_N(x, t)] = \frac{1}{\mathbf{E}[K_1(x)]} \mathbf{E} [K_1(x) \psi_x(Y, t)]$$

and by a simple manipulations we have

$$\begin{aligned}
\mathbf{E} \left[\tilde{\Psi}_N(x, t) \right] &= \frac{1}{\mathbf{E}[K_1(x)]} \mathbf{E} [K_1(x) \psi_x(Y, t)] \\
&= \frac{\mathbf{E} \{ K_1(x) [\mathbf{E}[\psi_x(Y, t)|X_1] - \mathbf{E}[\psi_x(Y, t)|X = x]] \}}{\mathbf{E}[K_1(x)]} \\
&\quad + \mathbf{E}[\psi_x(Y, t)|X = x] - \mathbf{E}[\psi_x(Y, \theta(x))|X = x] \\
&=: I_1 + I_2.
\end{aligned} \tag{3.13}$$

For $I_1(x)$ we use the same arguments as those used in Ferraty *et al.* (2007) and under (H2) to write that

$$\begin{aligned}
\mathbf{E} [K_1(x) [\mathbf{E}[\psi_x(Y, t)|X_1] - \mathbf{E}[\psi_x(Y, t)|X = x]]] &= \mathbf{E} [K_1(x) [\mathbf{E}[\Psi(X_1, t) - \Psi(x, z)|d(x, X_1)]]] \\
&= \mathbf{E} [K_1(x) [\mathbf{E}[\Phi(d(x, X_1), t)]]] \\
&= \int \Phi(th, t) K(t) d\mathbf{P}^{d(X, x)/h}(t) \\
&= h\Phi'(0, t) \int tK(t) d\mathbf{P}^{d(X, x)/h}(t).
\end{aligned} \tag{3.14}$$

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

$$\int tK(t) d\mathbf{P}^{d(X, x)/h}(t) = K(1) - \int_0^1 (sK(s))' \beta_x(s) ds + o(1)$$

to obtain that

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

Concerning I_2 we use a Taylor expansion to get, under (H1') and (H4')

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

The result is then a consequence of (3.13). ■

References

- Andersen, P.K., Borgan, O., Gill, R.D. and Keiding, N. (1993). *Statistical Models Based on Counting Processes*. Springer, New York.
- Attouch, M., Laksaci, A. and Ould Saïd, E. (2009). Asymptotic distribution of robust estimator in nonparametric functional models, *Comm. in 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, *J. Korean Statist. Soc.*, **39**, 489-500.

- Attouch, M., Laksaci, A. and Ould Saïd, E. (2012). Robust regression for functional time series data. *J. Japan Statist. Soc.* , **42** , 125–143.
- 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.
- Boente, G., Gonzalez-Manteiga, W. & Pérez-Gonzalez, A. (2009). Robust nonparametric estimation with missing data. *J. Statist. Plann. Inference.* **139**, 571-592.
- Collomb, G. and Härdle, W. (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.
- Chen, J. and Zhang, L. (2009). Asymptotic properties of nonparametric M-estimation for mixing functional data, *J. Statist. Plan. and Inference* **139**, 533-546.
- Crambes C, Delsol L. and Laksaci A. (2008). Robust nonparametric estimation for functional data, *J. Nonparametric Statist.* **20**, 573-598.
- Ferraty, F. and Vieu, P. *Nonparametric Functional Data Analysis. Theory and Practice*, Springer-Verlag, New York, 2006.
- Ferraty, F. and Romain, Y. (2011). *The Oxford Handbook of Functional Data Analysis*. Oxford University Press.
- Gheriballah, A. Laksaci, A. and Sekkal, S. (2013) Nonparametric M -regression for functional ergodic data. *Statist. Probab. Lett.* , **83**, 902–908.
- He, S., Yang, G. (1994). Estimating a lifetime distribution under different sampling plan. In S.S. Gupta J.O. Berger (Eds.) *Statistical decision theory and related topics* (Vol. 5. pp. 73-85) Berlin Heidelberg, NewYork : Springer.
- He, S. and Yang, G. (1998). Estimation of the truncation probability in the random truncation model. *The Annals of Statistics*, **26**, 1011–1027.
- Horrigue, W. and Ould Saïd, E. (2011). Strong uniform consistency of a nonparametric estimator of a conditional quantile for censored dependent data and functional regressors. *Random Oper. Stoch. Equ.*, **19** , 131–156.
- Horrigue, W. and Ould Saïd, E. (2014). Nonparametric regression quantile estimation for dependant functional data under random censorship : Asymptotic normality. To appear in Comm. in Statistics. Theory and Methods. DOI number : 10.1080/03610926.2013.784993.
- Huber, P.J. (1964). Robust estimation of a location parameter. *Annals of the Mathematical Statistics*, **35**, 73-101.
- Laïb, N. and Ould Saïd, E. (2000). A robust nonparametric estimation of the autoregression function under an ergodic hypothesis. *Canadian Journal of Statistics*, **28**, 817-828.
- Lemdani, M. and Ould Saïd, E. (2007). Asymptotic behavior of hazard rate kernel estimator under truncated and censored data, *Comm. in Statist. Theory and Methods*, **37**, 155–173.

- Loève, M. *Probability theory I*. Springer-Verlag. New-York, 1977.
- Lynden-Bell, D. (1971) A method of allowing for known observational selection in small samples applied to 3CR quasars. *Monthly Notices of the Royal Astronomical Society*, **155**, 95–118.
- Ould Saïd, E. and Lemdani, M. (2006). Asymptotic properties of a nonparametric regression function estimator with randomly truncated data. *Ann. Inst. Statist. Math.*, **58**, 357–378.
- Stone, C. (1981). Optimal rates of convergence for nonparametric regression. *Ann. Statist.*, **9**, 1348–1360.
- Stute, W. (1993). Almost sure representations of the product-limit estimator for truncated data. *Ann. Statist.*, **21**, 146–156.
- Wang, J., Liang, H.Y. and Fan, G. (2012a). Asymptotic properties for an M-estimator of the regression function with truncation and dependent data. *Journal of the Korean Statistical Society*, **41**, 35–367.
- Wang, J., Liang, H.Y. and Fan, G. (2012b). Local M-estimation of nonparametric regression with left-truncated and dependent data. *Scientia Sinica Mathematica*, **42**, 995–1015.
- Woodroffe, M. (1985). Estimating a distribution function with truncated data. *The Annals of Statistics*, **13**, 163–177.

Chapitre 4

Convergence presque complète : Cas α -mélange

Ce chapitre traite la convergence presque complète d'un estimateur robuste du régresseur fonctionnel tronquée à gauche pour des données α -mélangeantes, ces résultats font l'objet d'un article soumis dans le journal *Electronic Journal of statistics*.

M -estimation of the regression function under random left truncation and functional time series model

Saliha Derrar,[†] Ali Laksaci,[†]

[†] Laboratoire de Statistique et Processus Stochastiques,
Université de Sidi Bel Abbès,
BP 89 Sidi Bel Abbès 22000. Algeria.

derrarsaliha@yahoo.fr, alilak@yahoo.fr

Elias Ould Saïd,^{††,*}

^{††} Univ. Lille Nord de France,

* ULCO, LMPA, Centre de la Mi-Voix,

CS : 80699, Calais, France.

ouldsaid@lmpa.univ-littoral.fr¹

Abstract In this paper we study the M -estimation of the functional nonparametric regression when the response variable is subject to left-truncation by an other random variable. Under standard assumptions, we get the almost complete convergence rate of this robust estimate when the sample is an α -mixing sequence. This approach can be applied in time series analysis to the prediction problem. We point out that we treat the general case where the derivative of the lost function is unbounded.

AMS subject classifications : 60G42, 62F12, 62G20, 62G05.

Key words : Asymptotic normality, Functional data, Kernel estimator, Lynden-Bell estimator, Robust estimation, Small balls probabilities, Strong consistency, Truncated data.

4.1 Introduction

Let $(\mathcal{Y}_i)_{i=1,\dots,N}$ be a sample of real random variables (rv) with common unknown distribution function (df) F . Here the integer N is unknown but deterministic. Let $(\mathcal{X}_i)_{i=1,\dots,N}$ a corresponding sample which is valued in $\mathcal{F}$, where $\mathcal{F}$ is a semi-metric space and endowed with the semi-metric d . The $\mathcal{Y}_i$'s are regarded as the lifetimes of the items under study and are supposed to be subject to left-truncation which may occur if the time origin of the lifetime precedes the time origin of the study.

1. *corresponding author*

Only subjects that fail after the start of the study are being followed, the others being truncated. We denote by $(\mathcal{T}_i)_{i=1,\dots,N}$ the sample of truncation rv with d.f. G . The $\mathcal{T}_i$'s are assumed to be independent of the $\mathcal{Y}_i$'s. Then $(\mathcal{Y}_i, \mathcal{T}_i)$ are observed only when $(\mathcal{Y}_i \geq \mathcal{T}_i)$. Clearly the initial sample is not completely observed and only n observations (among N) are obtained. We point out that n is random but known. Such data, arises in many fields such as astronomy, economics and medical studies (see e.g., Woodroffe (1985)). In what follows, we denote again (if there is no-confusion for the index) $\{(Y_i, T_i, X_i), i = 1, \dots, n\}, (n \in \mathbb{N})$, the observed sample from the original N -sample. As a consequence of truncation, the size of the actually observed sample, n is a $\text{Bin}(N, \mu)$ random variable, with $\tau := \mathbb{P}(\mathcal{Y} \geq \mathcal{T})$. It is clear that if $\tau = 0$, no data can be observed and therefore, we suppose throughout this paper that $\tau > 0$. By the strong law of large numbers (SLLN) we have, as $N \rightarrow \infty$

$$\hat{\tau}_n := \frac{n}{N} \longrightarrow \tau, \quad \mathbb{P} - a.s. \quad (4.1)$$

Since the N is unknown and the n is known (although random), our results would not be stated with respect to the probability measure $\mathbb{P}$ (related to the N -sample) but will involve the conditional probability $\mathbf{P}(\cdot) = \mathbb{P}(\cdot | \mathcal{Y} > \mathcal{T})$ with respect to the actually observed n -sample. Also $\mathbf{E}$ and $\mathbb{E}$ will denote the expectation operators under $\mathbf{P}$ and $\mathbb{P}$, respectively.

Considering this situation we model the co-variation between X and Y by the M -regression function. Specifically, for a given function $\rho(\cdot, \cdot)$ on $\mathbb{R}$ with the unique minimizer at the origin with respect to (w.r.t.) the second component, define the M -regression θ_x such that it minimizes w.r.t. t

$$\Gamma(x, t) = \mathbb{E}[\rho(Y, t) | X = x] \quad (4.2)$$

that is

$$\theta_x = \arg \min_{t \in \mathbb{R}} \Gamma(x, t). \quad (4.3)$$

Noting that this definition covers and includes many important nonparametric models, for example, $\rho(y, t) = (y - t)^2$ yields the classical regression, $\rho(y, t) = |y - t|$ leads to the conditional median function $m(x) = \text{med}(Y | X = x)$, the α^{th} conditional quantile is obtained by setting $\rho(y, t) = |y - t| + (2\alpha - 1)(y - t)$. The M -regression function is an old model in nonparametric robust statistic. Precisely, this model has introduced by Huber (1964). In general the nonparametric robust regression model has received extensive attention in the literature when the data takes values in finite dimensional space. The key references on the nonparametric robust regression estimation are Robinson (1984), Collomb and Härdle (1986), Boente and Fraiman (1989 and 1990), Fan *et al.* (1994) Laïb and Ould Saïd (2000), Boente *et al.* (2009). The nonparametric robust regression estimation was firstly introduced, as far as we

know, in nonparametric functional statistics by Azzedine *et al.* (2008). They obtained the almost complete convergence of a family of robust estimators of the regression function in the independent and identically distributed (i.i.d.) case. Crambes *et al.* (2008) stated the convergence in L_q norm in both cases (i.i.d and strong mixing). The asymptotic normality of this last model has been established by Attouch *et al.* (2010), under the concentration properties on small balls of the probability measure of the underlying functional variable. We return to Chen and Zhang (2009), Attouch *et al.* (2009, 2010, 2012), Gheriballah *et al.* (2013) for some asymptotic results in functional time series data of the the nonparametric robust regression. While all these results are obtained in the complete data. In this work we study the nonparametric robust regression estimation the data are incomplete. For the censored case, Lemdani and Ould Saïd (2015) introduced a robust estimate of the regression function and established its uniform strong consistency with rate over a compact set as well as its asymptotic normality in iid case. Chaouch *et al.* (2014) extended the last results to ergodic case. At the best of our knowledge, only the papers of Horrigue and Ould Saïd (2011, 2015), Helal and Ould Saïd (2016) and Derrar *et al.* (2015) have paid attention to the nonparametric estimation with censored and truncated data and functional regressors respectively. The first authors have studied the conditional quantiles estimation for censorship model and functional regressors in dependent case. While in the second paper the authors study the conditional quantile in the truncated data and functional regressors for the iid case. In the later paper, the authors introduced, for the same model as in Helal and Ould Saïd (2016) a new family of the ψ -regression function in the i.i.d case. We recall that in the finite dimensional case, the nonparametric estimation of the regression function with randomly truncated data has been widely considered by many authors. For example, and to cite only a few, the pioneer work of Ould Saïd and Lemdani (2006) where they studied the asymptotic properties of the kernel estimator of the regression function in i.i.d. case, however Wang *et al.* (2012a) extended the last results to and Wang *et al.* (2012b) studied M-estimation method by using the local polynomial technique under dependent condition.

The main object of this paper is to generalize to dependent case the results obtained by Derrar *et al.* (2015) in the i.i.d case. More precisely, we establish the almost-complete convergence, with rates of an estimator constructed by combining the ideas of robustness with those of smoothed regression. We point out that the main feature of our approach is to develop an alternative prediction model to the classical regression that is not sensitive to outliers or heteroscedastic data.

The paper is organized as follows : Section 2 is devoted to the presentation of our estimate and some background for the truncated data with the estimation of their distribution functions. Section 3 is dedicated to fixing notations and hypotheses. We state our main result in Section 4, whereas the proofs of the main results are relegated to the Appendix.

4.2 Construction of the estimate and background for truncation model

Consider $(\mathcal{T}_i)_{i=1,\dots,N}$ be a sample of i.i.d. random variables which are distributed as $\mathcal{T}$, where N is unknown but deterministic. Without possible confusion, we still denote $(Y_i, T_i), i = 1, 2, \dots, n; (n \leq N)$ the actually observed r.v., with n being random but known. Thus, taking into account the truncation's effect the practical estimator is constructed by the observed variables $(X_i, Y_i, T_i)_{i=1,\dots,n}$. To do that, we suppose that the variable of interest Y and the truncating variable T are independent and have, respectively, an unknown distribution function F and G . Following Stute (1993) the distribution functions of Y and T , under the left-truncation condition, are expressed, respectively, by

$$F^*(y) := \tau^{-1} \int_{-\infty}^y G(v) dF(v) \quad \text{and} \quad G^*(t) := \tau^{-1} \int_{-\infty}^{\infty} G(t \wedge v) dF(v)$$

where $t \wedge v = \min(t, v)$ and they are estimated by their empirical ones.

Let

$$R(y) = G^*(y) - F^*(y) = \tau^{-1} G(y) [1 - F(y)].$$

and also it is estimated by its empirical one.

Recall that the distribution function F and G are estimated by the Lynden-Bell (1971) as follows

$$F_n(y) = 1 - \prod_{s \leq y} \left[1 - \frac{F_n^*(s)}{R_n(s)} \right] \quad \text{and} \quad G_n(y) = \prod_{s > y} \left[1 - \frac{G_n^*(s)}{R_n(s)} \right].$$

He and Wang (1998) showed that τ can be estimated by

$$\tau_n = \frac{G_n(y) \bar{F}_n(y)}{R_n(y)}.$$

and is independent of y .

The asymptotic properties of F_n and G_n were studied by Woodroffe (1985), while the almost sure consistency of τ_n was obtained by He and Yang (1998). Finally let's point out that the asymptotic proprieties of F_n, G_n and τ_n are obtained under an identifiability conditions that are $a_G \leq a_F, b_G \leq b_F$ where a_W and b_W the endpoints of the support of a distribution function W defined by

$$a_W = \inf\{x : W(x) > 0\} \quad \text{and} \quad b_W = \sup\{x; W(x) < 1\}.$$

Let a, b be two real numbers such that $[a, b] \subset [a_F, b_F]$. We point out that we need this assumption because we need the uniform consistency of the distribution law $G(\cdot)$

of the truncated rv T , which stated over compact set (see Remark 6 in Woodroffe (1985)). Furthermore we suppose that our parameter is such that $\theta_x \in (a, b)$.

Given an observed triplet random variables $(X_1, Y_1, T_1), \dots, (X_n, Y_n, T_n)$, and combining the ideas in Ferraty and Vieu (2006) and Ould Saïd and Lemdani (2006), the kernel estimate of the M -regression function θ_x denoted $\hat{\theta}_x$ is

$$\hat{\theta}_x = \arg \min_{t \in \mathbb{R}} \hat{\Gamma}(x, t), \quad (4.4)$$

where

$$\hat{\Gamma}(x, t) = \frac{\sum_{i=1}^n G_n^{-1}(Y_i) K(h^{-1}d(x, X_i)) \rho(Y_i, t)}{\sum_{i=1}^n G_n^{-1}(Y_i) K(h^{-1}d(x, X_i))}$$

and $h := h_n$.

Noting that, in nonparametric modeling, robust regression is a very important regression analysis tool because it is less sensitive to outliers in the data compared with the classical regression. On the other hand the statistical analysis of infinite dimensional data has been the subject of several works in the recent statistical literature. We cite for instance, Ramsay and Silverman (2005), Ferraty and Vieu (2006), Horváth and Kokoszka (2012) or Bongiorno et al. (2014) for general books). Moreover, the regression analysis of incomplete data (censored and/or truncated data) has gained a particular interest in the statistics literatures. Such kind of data occur in many fields of applications such as in astronomy, economics, epidemiology, biometry and medical studies (see, Woodroffe (1985), Andersen *et al.* (1993) or He and Yang (1994)). For example, in a retirement center, subjects are observed only if they live long enough to enter the retirement house. The lifetime, say Y , is then left truncated by the retirement house entry age, say T . People who enter the retirement house earlier may get better medical attention and therefore live longer. On the other hand, people with poor health and shorter expected lifetime may retire earlier. The main aim of this contribution is to bring together these three areas, namely, the robust analysis, the functional data and the regression analysis of incomplete data. The main asymptotic result of this contribution is the establishment of the almost complete consistency of (3) with rates, under the the strong mixing conditions.

4.3 Hypotheses and results

We begin by recalling the definition of the strong mixing property. For this we introduce the following notations. Let $\mathcal{F}_i^k(Z)$ denote the σ -algebra generated by $\{Z_j, i \leq j \leq k\}$.

Definition 1. Let $\{Z_i, i = 1, 2, \dots\}$ be a strictly stationary sequence of random variables. Given a positive integer n , set

$$\alpha(n) = \sup \left\{ \left| \mathbf{P}(A \cap B) - \mathbf{P}(A)\mathbf{P}(B) \right| : A \in \mathcal{F}_1^k(Z) \text{ and } B \in \mathcal{F}_{k+n}^\infty(Z), \right. \\ \left. k \text{ is strictly positive integer} \right\}.$$

The sequence is said to be α -mixing (strong mixing) if the mixing coefficient $\alpha(n) \rightarrow 0$ as $n \rightarrow \infty$.

There exist many processes fulfilling the strong mixing property. We quote, here, the usual ARMA processes (with innovations satisfying some existing moment conditions) are geometrically strongly mixing, *i.e.*, there exist $v \in (0, 1)$ and $C > 0$ such that, for any $n \geq 1$, $\alpha(n) \leq C v^n$ (see, *e.g.*, Jones (1978)). The threshold models, the EXPAR models (see, Ozaki (1979)), the simple ARCH models (see Engle (1982)), their GARCH extension (see Bollerslev (1986)) and the bilinear Markovian models are geometrically strongly mixing under some general ergodicity conditions. For more details we refer the reader to the monographs of Bradley (2007) or Dedecker *et al.* (2007).

In the remainder of the paper, we suppose that $(X_n, Y_n)_{n \geq 1}$ is strongly mixing whose coefficient $\alpha(n)$, x is a fixed point in $\mathcal{F}$ and when no confusion is possible, we will denote by C and C' some strictly positive generic constants.

We need the following hypotheses which gathered here together to easy references.

(H1) $\mathbf{P}(X \in B(x, r)) =: \phi_x(r) > 0$, $\forall r > 0$, where $B(x, r) = \{x' \in \mathcal{F} : d(x', x) < r\}$.

(H2) The function Γ is such that :

$$\left\{ \begin{array}{l} \text{(i) The function } \Gamma(x, \cdot) \text{ is of class } \mathcal{C}^2 \text{ on } [\theta_x - \delta, \theta_x + \delta], \delta > 0. \\ \text{(ii) } \forall t \in [\theta_x - \delta, \theta_x + \delta], \forall (x_1, x_2) \in \mathcal{N}_x \times \mathcal{N}_x, |\Gamma(x_1, t) - \Gamma(x_2, t)| \leq C d^b(x_1, x_2), \\ \quad \text{where } \mathcal{N}_x \text{ is a fixed neighborhood of } x. \\ \text{(iii) For each fixed } t \in [\theta_x - \delta, \theta_x + \delta], \text{ the function } \Gamma(\cdot, t) \text{ is continuous at the point } x. \end{array} \right.$$

(H3) ρ is a strictly convex function, continuously differentiable w.r.t. the second component, and its derivative, $\psi(y, t) = \frac{\partial \rho(y, t)}{\partial t}$, such that

$$\mathbf{E} [|\psi(Y, t)|^2 | X] < C < \infty \quad \text{and} \quad \mathbf{E} [|\psi(Y, t)|^p] < C < \infty, \quad p > 2.$$

(H4) The observed sequence $(X_i, Y_i)_{i \in \mathbb{N}}$ satisfies : $\exists a > 0, \exists c > 0 : \forall n \in \mathbb{N}, \alpha(n) \leq cn^{-a}$ and

$$\left\{ \begin{array}{l} \forall i \neq j, \forall t \in [\theta_x - \delta, \theta_x + \delta], \quad \mathbf{E} [\psi(Y_i, t)\psi(Y_j, t) | X_i, X_j] \leq C < \infty, \\ \mathbf{P}((X_i, X_j) \in B(x, r) \times B(x, r)) = O \left(\left(\frac{\phi_x(h)}{n^{2/p}} \right)^{a/(a-1)} \right). \end{array} \right.$$

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

(H6) There exists $\eta > 0$, such that

$$Cn^{\frac{(4-a)}{(a+1)}+\eta} \leq \phi_x(h) \leq n^{\frac{2a}{(2-a)p}},$$

$$\text{where and } a > \max\left(4, p, \frac{3p+1+\sqrt{p^2+22p+1}}{p-2}\right).$$

Remarks on the assumptions

Our assumptions are standard in nonparametric functional statistics. Hypothesis (H1) is checked for several continuous time processes (see for instance Bogachev (1999) for a Gaussian measure and Li and Shao (2001) for a general Gaussian process). This conditions is the same as those used by Ferraty and Vieu (2006). More precisely, even if $\phi_x(r)$ depends strongly on the center ball x , but this function can be asymptotically explicated for several well-known continuous time processes. Indeed, in many examples of continuous time processes, the function $\phi_x(r)$ can be written approximatively as the product of two independent functions $g(x)$ and $\phi(r)$. For instance, it is shown in It is well known that, the latter can be quantified for several continuous time processes such as a Gaussian processes, Diffusion processes where the function $\phi(r)$ has the following general form Corollary 4.7.8 in Bogachev (1999, page 186), that the expression of the Onsager-Machlup function of the couple (x, z) , for the Gaussian measures on a semi-normed space $(\mathcal{F}, \|\cdot\|)$, is given by :

$$F(x, z) = \log \left(\lim_{r \rightarrow 0} \frac{\mathbf{P}(X \in B(x, r))}{\mathbf{P}(X \in B(z, r))} \right) = \frac{1}{2} \|\pi(z)\|_H^2 - \frac{1}{2} \|\pi(x)\|_H^2$$

where $\|\cdot\|_H$ is the Hilbert norm on the Cameron-Martin space of $\mathcal{F}$ associated to a Gaussian measure, denoted by H , and $\pi(\cdot)$ is the orthogonal projection onto the orthogonal complement of the set $\{a \in H, \text{ such that } \|a\| = 0\}$. So, in this case, the small ball probability function $\phi_x(r)$ can be written as $\phi_x(r) = g(x)\phi(r) + o(\phi(r))$ with

$$g(x) = \exp\left(-\frac{1}{2}\|\pi(x)\|_H^2\right) \quad \text{and} \quad \phi(r) = \mathbf{P}(X \in B(0, r)).$$

Then

$$\phi(r) = r^\gamma \exp\left(-\frac{C}{r^p}\right) + o\left(r^\gamma \exp\left(-\frac{C}{r^p}\right)\right) \quad \text{for some } \gamma > 0 \text{ and } p > 0.$$

Conditions (H2) and (H3) on the score function is a quite lower than that considered in Crambes *et al.* (2008) or Attouch *et al.* (2010) where ψ is assumed to be continuously differentiable. The robustification condition (H3) is verified for the usual functions (Andrews, Huber, Hampel, Tuckey, ...) which gives more flexibility in the practical choice of ρ . Condition (H4) controls the local dependency of the functional time series data. Assumptions (H5) and (H6) are standard technical conditions in nonparametric estimation. They are imposed for the sake of simplicity and brevity of the proofs.

Théoreme 4.3.1 *Assume that (H1), (H2)((i), (iii)) and (H3)-(H6) are satisfied, then $\hat{\theta}_x$ exists for all sufficiently large n . Furthermore, we have*

$$\hat{\theta}_x - \theta_x \rightarrow 0 \quad a.s.$$

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

Théoreme 4.3.2 *Assume that (H1), (H2) ((i), (ii)) and (H3)-(H6) are satisfied, then $\hat{\theta}_x$ exists a.s. for all sufficiently large n . Furthermore, if $\Gamma''(x, \cdot)(x, \theta_x) \neq 0$ we have*

$$\hat{\theta}_x - \theta_x = O(h^b) + O\left(\sqrt{\frac{\log n}{n\phi_x(h)}}\right) \quad a.s.$$

4.4 Appendix

For the proofs of the Theorem 4.3.1 and 4.3.2 we use the fact that ρ is a strictly convex function and continuously differentiable w.r.t. the second component, then ψ is strictly monotone and continuous w.r.t. the second component. We give the proof for the case of a increasing $\psi(Y, \cdot)$, decreasing case being obtained by considering $-\psi(Y, \cdot)$. Therefore, we can write, under this consideration, for all $\epsilon > 0$

$$\Gamma'(x, \theta_x - \epsilon) \leq \Gamma'(x, \theta_x) = 0 \leq \Gamma'(x, \theta_x + \epsilon) \quad \text{and} \quad \hat{\Gamma}'(x, \hat{\theta}_x - \epsilon) \leq \hat{\Gamma}'(x, \hat{\theta}_x) = 0 \leq \hat{\Gamma}'(x, \hat{\theta}_x + \epsilon).$$

Hence, for all $\epsilon > 0$, we have

$$\begin{aligned} \mathbf{P}\left(|\hat{\theta}_x - \theta_x| \geq \epsilon\right) &\leq \mathbf{P}\left(|\hat{\Gamma}'(x, \theta_x + \epsilon) - \Gamma'(x, \theta_x + \epsilon)| \geq \Gamma'(x, \theta_x + \epsilon)\right) \\ &\quad + \mathbf{P}\left(|\hat{\Gamma}'(x, \theta_x - \epsilon) - \Gamma'(x, \theta_x - \epsilon)| \geq -\Gamma'(x, \theta_x - \epsilon)\right). \end{aligned}$$

So, it suffices to show that

$$\hat{\Gamma}'(x, t) - \Gamma'(x, t) \rightarrow 0 \quad a.co. \text{ for } t := \theta_x \pm \epsilon. \quad (4.5)$$

Moreover, under ((H2) (i)), we get that

$$\hat{\theta}_x - \theta_x = \frac{\Gamma'(x, \hat{\theta}_x) - \hat{\Gamma}'(x, \hat{\theta}_x)}{\Gamma''(x, \xi_n)}$$

where ξ_n is between $\widehat{\theta}_x$ and θ_x . As long as we could be able to check that

$$\exists \tau > 0, \sum_{n=1}^{\infty} \mathbf{P}(\Gamma''(x, \xi_n) < \tau) < \infty, \quad (4.6)$$

we would have

$$\widehat{\theta}_x - \theta_x = O_{a.co.} \left(\sup_{t \in [\theta_x - \delta, \theta_x + \delta]} |\Gamma'(x, t) - \widehat{\Gamma}'(x, t)| \right).$$

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

$$\sup_{t \in [\theta_x - \delta, \theta_x + \delta]} |\Gamma'(x, t) - \widehat{\Gamma}'(x, t)|.$$

To do that, we write

$$\widehat{\Gamma}'(x, t) = \frac{\widehat{\Psi}_N(x, t)}{\widehat{\Psi}_D(x)}$$

with

$$\begin{aligned} \widehat{\Psi}_N(x, t) &= \frac{\tau_n}{n\mathbf{E}[K_1]} \sum_{i=1}^n \frac{1}{G_n(Y_i)} K\left(\frac{d(x, X_i)}{h}\right) \psi(Y_i, t), \\ \widehat{\Psi}_D(x) &= \frac{\tau_n}{n\mathbf{E}[K_1]} \sum_{i=1}^n \frac{1}{G_n(Y_i)} K\left(\frac{d(x, X_i)}{h}\right) \end{aligned}$$

and we consider the following decomposition

$$\begin{aligned} \widehat{\Gamma}'(x, t) - \Gamma'(x, t) &= \frac{1}{\widehat{\Psi}_D(x)} \left[\left(\widehat{\Psi}_N(x, t) - \mathbf{E} \left[\widehat{\Psi}_N(x, t) \right] \right) - \left(\Gamma'(x, t) - \mathbf{E} \left[\widehat{\Psi}_N(x, t) \right] \right) \right] \\ &\quad + \frac{\Gamma'(x, t)}{\widehat{\Psi}_D(x)} \left[\mathbf{E} \left[\widehat{\Psi}_D(x) \right] - \widehat{\Psi}_D(x) \right]. \end{aligned} \quad (4.7)$$

Unlike to Attouch *et al.* (2012) we must introduce the following pseudo-estimators

$$\widetilde{\Psi}_D(x) = \frac{\tau}{n\mathbf{E}[K_1]} \sum_{i=1}^n \frac{1}{G(Y_i)} K\left(\frac{d(x, X_i)}{h}\right)$$

and

$$\widetilde{\Psi}_N(x, t) = \frac{\tau}{n\mathbf{E}[K_1]} \sum_{i=1}^n \frac{1}{G(Y_i)} K\left(\frac{d(x, X_i)}{h}\right) \psi(Y_i, t).$$

Moreover, we put

$$\widehat{\Psi}_N(x, t) = \frac{\tau_n}{n\mathbf{E}[K_1]} \sum_{i=1}^n \frac{1}{G_n(Y_i)} K\left(\frac{d(x, X_i)}{h}\right) \psi(Y_i, t)$$

and

$$\widehat{\Psi}_D(x) = \frac{\tau_n}{n\mathbf{E}[K_1]} \sum_{i=1}^n \frac{1}{G_n(Y_i)} K\left(\frac{d(x, X_i)}{h}\right).$$

Now, we consider the following decomposition

$$\begin{aligned} \widehat{\Gamma}'(x, t) - \Gamma'(x, t) &= \frac{\widehat{\Psi}_N(x, t)}{\widehat{\Psi}_D(x)} - \Psi(x, t) \\ &= \frac{\widehat{\Psi}_N(x, t)}{\widehat{\Psi}_D(x)} - \frac{\Psi(x, t)}{\widehat{\Psi}_D(x)} + \frac{\Psi(x, t)}{\widehat{\Psi}_D(x)} - \Psi(x, t) \\ &= \frac{1}{\widehat{\Psi}_D(x)} \left(\widehat{\Psi}_N(x, t) - \Psi(x, t) \right) + \frac{\Psi(x, t)}{\widehat{\Psi}_D(x)} \left(-\widehat{\Psi}_D(x) + 1 \right) \\ &= \frac{1}{\widehat{\Psi}_D(x)} \left(\widehat{\Psi}_N(x, t) - \widetilde{\Psi}_N(x, t) \right) + \frac{1}{\widehat{\Psi}_D(x)} \left(\widetilde{\Psi}_N(x, t) - \mathbf{E} \left[\widetilde{\Psi}_N(x, t) \right] \right) \\ &\quad + \frac{1}{\widehat{\Psi}_D(x)} \left(\mathbf{E} \left[\widetilde{\Psi}_N(x, t) \right] - \Psi(x, t) \right) \\ &\quad + \frac{\Psi(x, t)}{\widehat{\Psi}_D(x)} \left\{ \left(\widetilde{\Psi}_D(x) - \widehat{\Psi}_D(x) \right) + \left(\mathbf{E} \left[\widetilde{\Psi}_D(x) \right] - \widetilde{\Psi}_D(x) \right) + \left(-\mathbf{E} \left[\widetilde{\Psi}_D(x) \right] + 1 \right) \right\}. \end{aligned}$$

Thus, both Theorems are a consequence of the following intermediates results.

Lemma 1.

Under Hypotheses (H1), (H4)-(H6), we have,

$$\widetilde{\Psi}_D(x) - \mathbf{E} \left[\widetilde{\Psi}_D(x) \right] = O \left(\sqrt{\frac{\log n}{n\phi_x(h)}} \right) \quad a.co.$$

Proof. For all $x \in \mathcal{F}$, for all $i = 1, \dots, n$, we put $K_i(x) = K(h^{-1}d(x, X_i))$ and

$$\Delta_i(x) = K_i(x) \frac{\tau}{G(Y_i)} - \mathbf{E} \left[K_i(x) \frac{\tau}{G(Y_i)} \right].$$

It follows that

$$\widetilde{\Psi}_D(x) - \mathbf{E}[\widetilde{\Psi}_D(x)] = \frac{1}{n\mathbf{E}[K_1(x)]} \sum_{i=1}^n \Delta_i(x).$$

Now, we apply the Fuck-Nagaev exponential inequality (Rio, 2000 p. 87) to get for all $\ell > 0$ and $\varepsilon > 0$, we have

$$\begin{aligned} \mathbf{P} \left\{ \left| \tilde{\Psi}_D(x) - \mathbf{E}[\tilde{\Psi}_D](x) \right| > \varepsilon \right\} &= \mathbf{P} \left\{ \left| \sum_{i=1}^n \Delta_i(x) \right| > \varepsilon n \mathbf{E}[K_1(x)] \right\} \\ &\leq C(A_1 + A_2) \end{aligned} \quad (4.8)$$

where

$$A_1 = \left(1 + \frac{\varepsilon^2 n^2 (\mathbf{E}[K_1(x)])^2}{S_n^2 \ell} \right)^{-\ell/2} \quad \text{and} \quad A_2 = n \ell^{-1} \left(\frac{\ell}{\varepsilon n \mathbf{E}[K_1(x)]} \right)^{a+1}$$

and

$$S_n^2 = \sum_{i=1}^n \sum_{j=1}^n \text{Cov}(\Delta_i(x), \Delta_j(x)) = S_n^{2*} + n \text{Var}[\Delta_1(x)]$$

with

$$S_n^{2*} = \sum_{i=1}^n \sum_{i \neq j}^n \text{Cov}(\Delta_i(x), \Delta_j(x)).$$

Next, we evaluate the asymptotic behavior of S_n^{2*} . Using the technique of Masry (1986), we define the sets

$$S_1 = \{(i, j) \text{ such that } 1 \leq i - j \leq m_n\}$$

and

$$S_2 = \{(i, j) \text{ such that } m_n + 1 \leq i - j \leq n - 1\}$$

where $m_n \rightarrow \infty$, as $n \rightarrow \infty$. Thus,

$$J_{1,n} := \sum_{S_1} |\text{Cov}(\Delta_i(x), \Delta_j(x))| \leq \sum_{S_1} |\mathbf{E}[K_i(x)K_j(x)] - \mathbf{E}[K_i(x)]\mathbf{E}[K_j(x)]|.$$

Because of (H1), (H5) and (H6) we can write

$$J_{1,n} \leq C n m_n \left(\frac{\phi_x(h)}{n^{2/p}} \right)^{a/(a-1)}.$$

For the summation over S_2 , we use Davydov-Rio's inequality (Rio, 2000, p. 87) for mixing processes. This leads, for all $i \neq j$, to

$$|\text{Cov}(K_i(x), K_j(x))| \leq C \alpha(|i - j|).$$

Therefore, under (H3) we have

$$J_{2,n} := \left| \sum_{(i,j) \in S_2} \text{Cov} [\Delta_i(x, z), \Delta_j(x, z)] \right| \leq C \frac{nm_n^{-a+1}}{a-1}. \quad (4.9)$$

Putting $m_n = \left(\left(\frac{\phi_x(h)}{n^{2/p}} \right)^{2/(1-a)} \right)$ we get

$$\sum_{i \neq j}^n \text{Cov}(K_i(x), K_j(x)) = O(n^{(p-2)/p} \phi_x(h)).$$

For the variance term, we deduce from (H1) that

$$\text{Var}(\Delta_1(x)) \leq C(\phi_x(h) + (\phi_x(h))^2) \leq C\phi_x(h).$$

Finally, as $a > 2$

$$S_n^2 = O(n\phi_x(h)). \quad (4.10)$$

Now, we apply (4.8) to $\varepsilon = \lambda \frac{\sqrt{S_n^2 \log n}}{n \mathbf{E}[K_1(x)]}$ and $\ell = C(\log n)^2$. We show that at

$$A_2 \leq Cn^{1-(a+1)/2} \phi_x(h)^{-(a+1)/2} (\log n)^{(3a-1)/2}.$$

The left side of (H6) allows to write

$$A_2 \leq Cn^{-1-\eta(a+1)/2} (\log n)^{(3a-1)/2} \leq Cn^{-1-\nu}$$

for some real $\nu > 0$.

Concerning the term A_1 , from (4.10), we show that

$$A_1 \leq C \left(1 + \frac{\lambda^2 \log n}{\ell} \right)^{-\ell/2} = C \exp \left(-\ell/2 \log \left(1 + \frac{\lambda^2 \log n}{\ell} \right) \right)$$

using the fact that $\ell = C(\log n)^2$, to get

$$A_1 \leq C \exp \left(-\lambda^2 \frac{\log n}{2} \right) = Cn^{-\lambda^2/2}.$$

Thus, for λ large enough :

$$\exists \nu' > 0, \quad A_1 \leq Cn^{-\lambda^2/2} \leq Cn^{-1-\nu'}.$$

which complete the proof of Lemma 4.4. ■ **Lemma 2.** [?]
Under Hypotheses (H1),(H4) and (H5), we have,

$$\mathbf{E} \left[\tilde{\Psi}_D(x) \right] = 1.$$

Lemma 3.

Under Hypotheses (H1),(H4) and (H5), we have,

$$\widehat{\Psi}_D(x) - \tilde{\Psi}_D(x) = O_{a.s.} \left(\sqrt{\frac{1}{n}} \right).$$

Proof. Firstly, we write

$$\begin{aligned} \left| \widehat{\Psi}_D(x) - \tilde{\Psi}_D(x) \right| &= \left| \frac{\tau_n}{n\mathbf{E}[K_1]} \sum_{i=1}^n \frac{1}{G_n(Y_i)} K \left(\frac{d(x, X_i)}{h} \right) \right. \\ &\quad \left. - \frac{\tau}{n\mathbf{E}[K_1]} \sum_{i=1}^n \frac{1}{G(Y_i)} K \left(\frac{d(x, X_i)}{h} \right) \right| \\ &\leq \left[\frac{|\tau_n - \tau|}{G_n(a_F)} + \tau \frac{\sup_{y \in [a, b]} |G_n(y) - G(y)|}{G_n(a_F)G(a_F)} \right] \\ &\quad \times \frac{1}{n\mathbf{E}[K_1]} \sum_{i=1}^n K \left(\frac{d(x, X_i)}{h} \right). \end{aligned}$$

Secondly, the quantity $|\tau_n - \tau|$ has been studied by Ould Saïd and Tatachak (2009) where we show that $|\tau_n - \tau| = O_{a.s.}(n^{-\frac{1}{2}})$. Moreover, the remark 6 of Woodroffe(1985)) allows to write $|G_n(a_F) - G(a_F)| = O_{a.s.}(n^{-\frac{1}{2}})$. Furthermore, we point out it is shown in Attouch *et al.* (2012) that

$$\frac{1}{n\mathbf{E}[K_1]} \sum_{i=1}^n K \left(\frac{d(x, X_i)}{h} \right) = O(1).$$

Consequently,

$$\left| \widehat{\Psi}_D(x) - \tilde{\Psi}_D(x) \right| = O_{a.s.} \left(\sqrt{\frac{1}{n}} \right)$$

which is the claimed result. ■ **Lemma 4.**

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

$$\sup_{t \in [\theta_x - \delta, \theta_x + \delta]} \left| \tilde{\Psi}_N(x, t) - \mathbf{E} \left[\tilde{\Psi}_N(x, t) \right] \right| = O \left(\sqrt{\frac{\log n}{n\phi_x(h)}} \right) \quad a.co.$$

Proof. The compactness of $[\theta_x - \delta, \theta_x + \delta]$, allows to write

$$[\theta_x - \delta, \theta_x + \delta] \subset \bigcup_{j=1}^{d_n} (y_j - l_n, y_j + l_n) \quad (4.11)$$

with $l_n = n^{-1/2}$ and $d_n = O(n^{1/2})$. We put

$$\mathcal{G}_n = \{y_j - l_n, y_j + l_n, 1 \leq j \leq d_n\}. \quad (4.12)$$

We combine (H4) together with the monotony of $\mathbf{E}[\tilde{\Psi}_N(x, \cdot)]$ and $\tilde{\Psi}_N(x, \cdot)$ to write

$$\begin{aligned} \sup_{t \in [\theta_x - \delta, \theta_x + \delta]} \left| \tilde{\Psi}_N(x, t) - \mathbf{E}[\tilde{\Psi}_N(x, t)] \right| &\leq \max_{1 \leq j \leq d_n} \max_{z \in \{y_j - l_n, y_j + l_n\}} \left| \tilde{\Psi}_N(x, z) - \mathbf{E}[\tilde{\Psi}_N(x, z)] \right| \\ &\quad + 2Cl_n. \end{aligned} \quad (4.13)$$

On the one hand condition (H6) implies that

$$l_n = o\left(\sqrt{\frac{\log n}{n\phi_x(h)}}\right). \quad (4.14)$$

On the other hand we have, for any $\varepsilon > 0$

$$\mathbf{P}\left(\max_{z \in \mathcal{G}_n} \left| \tilde{\Psi}_N(x, z) - \mathbf{E}[\tilde{\Psi}_N(x, z)] \right| > \varepsilon\right) \leq \sum_{z \in \mathcal{G}_n} \mathbf{P}\left(\left| \tilde{\Psi}_N(x, z) - \mathbf{E}[\tilde{\Psi}_N(x, z)] \right| > \varepsilon\right). \quad (4.15)$$

Now, we use a truncation method to prove our result for general case where ψ is not necessary bounded. Indeed, we consider the following random variable

$$\tilde{\Psi}_N^*(x, t) = \frac{1}{n\mathbf{E}[K(h^{-1}d(x, X_1))]} \sum_{i=1}^n K(h^{-1}d(x, X_i)) \frac{\tau}{G(Y_i)} \psi^*(Y_i, t)$$

where $\psi^*(\cdot, t) = \psi(\cdot, t)\mathbf{1}_{(\psi(\cdot, t) < \gamma_n)}$ with $\gamma_n = n^{a/p}$. So, the claimed result is a consequence of the three intermediates results.

$$\max_{z \in \mathcal{G}_n} \left| \mathbf{E}[\tilde{\Psi}_N^*(x, z)] - \mathbf{E}[\tilde{\Psi}_N(x, z)] \right| = o\left(\sqrt{\frac{\log n}{n\phi_x(h)}}\right), \quad (4.16)$$

$$\sum_n d_n \max_{z \in \mathcal{G}_n} \mathbf{P}\left(\left| \tilde{\Psi}_N^*(x, z) - \tilde{\Psi}_N(x, z) \right| > \epsilon_0 \left(\sqrt{\frac{\log n}{n\phi_x(h)}}\right)\right) < \infty \quad \text{for some } \epsilon_0 > 0$$

(4.17)

and

$$\sum_n d_n \max_{z \in \mathcal{G}_n} \mathbf{P} \left(\left| \tilde{\Psi}_N^*(x, z) - \mathbf{E}[\tilde{\Psi}_N^*(x, z)] \right| > \epsilon_0 \left(\sqrt{\frac{\log n}{n\phi_x(h)}} \right) \right) < \infty. \quad (4.18)$$

Proof of (4.16)

It clear that, for all $z \in \mathcal{G}_n$ we can write

$$\left| \mathbf{E}[\tilde{\Psi}_N^*(x, z)] - \mathbf{E}[\tilde{\Psi}_N(x, z)] \right| \leq C \frac{1}{\phi_x(h)} \mathbf{E} \left[|\psi(Y, z)| \mathbf{1}_{\{\psi(Y, z) \geq \gamma_n\}} K(h^{-1}d(x, X)) \right].$$

Applying Holder inequality, for $\kappa = \frac{p}{2}$ with ζ such that $\frac{1}{\kappa} + \frac{1}{\zeta} = 1$, to write that, for all $z \in \mathcal{G}_n$

$$\begin{aligned} \mathbf{E} \left[|\psi(Y, z)| \mathbf{1}_{\{\psi(Y, z) \geq \gamma_n\}} K(h^{-1}d(x, X_1)) \right] &\leq \mathbf{E}^{1/\kappa} \left[|\psi^\kappa(Y, z)| \mathbf{1}_{\{\psi(Y, z) \geq \gamma_n\}} \right] \mathbf{E}^{1/\zeta} \left[K^\zeta(h^{-1}d(x, X_1)) \right] \\ &\leq \gamma_n^{-1} \mathbf{E}^{1/\kappa} \left[|\psi^{2\kappa}(Y, z)| \right] \mathbf{E}^{1/\zeta} \left[K^\zeta(h^{-1}d(x, X_1)) \right] \\ &\leq C \gamma_n^{-1} \phi_x^{1/\zeta}(h). \end{aligned}$$

Thus,

$$\max_{z \in \mathcal{G}_n} \left| \tilde{\Psi}_N^*(x, z) - \mathbf{E}[\tilde{\Psi}_N^*(x, z)] \right| \leq n^{-a/p} \phi_x^{(1-\zeta)/\zeta}.$$

Therefore, (4.16) is consequence of the fact that $a > p > 2$.

Proof of (4.17)

To do so, we use the Markov's inequality $\forall z \in \mathcal{G}_n, \forall \epsilon > 0$

$$\begin{aligned} \mathbf{P} \left(\left| \tilde{\Psi}_N^*(x, z) - \tilde{\Psi}_N(x, z) \right| > \epsilon \right) &\leq \sum_{i=1}^n \mathbf{P} \left(\psi(Y_i, z) > n^{a/p} \right) \\ &\leq n^{1-a} \mathbf{E}[\psi^p(Y, z)]. \end{aligned}$$

In particular, for $\epsilon = \epsilon_0 \left(\sqrt{\frac{\log n}{n\phi_x(h)}} \right)$ and thanks to $a > 4$, we have

$$d_n \max_{z \in \mathcal{G}_n} \mathbf{P} \left(\left| \tilde{\Psi}_N(x, z) - \tilde{\Psi}_N^*(x, z) \right| > \epsilon_0 \left(\sqrt{\frac{\log n}{n\phi_x(h)}} \right) \right) \leq n^{3/2-a} < C n^{-1-\nu}.$$

Proof of (4.18)

The latter is based on the similar arguments as those used to proving Lemma 4.4. Indeed, we define, for all $z \in \mathcal{G}_n$,

$$\Lambda_i(z) = K_i(x) \psi^*(Y_i, z) \frac{\tau}{G(Y_i)} - \mathbf{E} \left[K_1(x) \frac{\tau}{G(Y_i)} \psi^*(Y_i, z) \right].$$

Therefore, for all $\epsilon > 0$

$$\mathbf{P} \left\{ \left| \tilde{\Psi}_N^*(x, z) - \mathbf{E} \left[\tilde{\Psi}_N^*(x, z) \right] \right| > \epsilon \right\} = \mathbf{P} \left\{ \left| \sum_{i=1}^n \Lambda_i(z) \right| > \epsilon n \mathbf{E}[K_1(x)] \right\}.$$

As in Lemma 4.4, the evaluation of the last quantity is based on the asymptotic behavior of the following quantity

$$S_n'^2 = \sum_{i=1}^n \sum_{j=1}^n \text{Cov}(\Lambda_i(z), \Lambda_j(z)) = \sum_{i=1}^n \sum_{i \neq j}^n \text{Cov}(\Lambda_i(z), \Lambda_j(z)) + n \text{Var}[\Lambda_1(z)].$$

For this term we use the same technique developed by Masry (1986) by splitting the sum into two sets defined by

$$S_1' = \{(i, j) \text{ such that } 1 \leq i - j \leq u_n\}$$

and

$$S_2' = \{(i, j) \text{ such that } u_n + 1 \leq i - j \leq n - 1\}.$$

We note by $J_{1,n}'$ and $J_{2,n}'$ be the sum of covariance over S_1' and S_2' respectively. On S_1' , we have, under (H4)

$$J_{1,n}' \leq C \sum_{S_1'} |\mathbf{E}[K_i(x)K_j(x)]| + |\mathbf{E}[K_i(x)] \mathbf{E}[K_j(x)]|.$$

Because of (H1), (H5) and (H6) we have

$$J_{1,n}' \leq C n u_n \left(\frac{\phi_x(h)}{n^{2/p}} \right)^{a/(a-1)}.$$

Concerning $J_{2,n}'$, we use again Davydov-Rio's inequality in the L^∞ cases and we obtain

$$|\text{Cov}(\Lambda_i(z), \Lambda_j(z))| \leq C \gamma_n^2 \alpha(|i - j|).$$

Therefore,

$$J_{2,n}' = \sum_{S_2'} |\text{Cov}(\Lambda_i(z), \Lambda_j(z))| \leq \frac{n \gamma_n^2 u_n^{-a+1}}{a-1}.$$

Taking $u_n = \left(\frac{\gamma_n^2}{\left(\frac{\phi_x(h)}{n^{2/p}} \right)^{a/(a-1)}} \right)^{1/a}$, we prove that

$$\sum_{i=1}^n \sum_{i \neq j}^n \text{Cov}(\Lambda_i(z), \Lambda_j(z)) = O(n \phi_x(h)).$$

On the other hand, under (H3), we have

$$\text{Var}(\Lambda_1(z)) \leq \mathbf{E}[K_i(x)\psi^*(Y_i, z)]^2 \leq \mathbf{E}[K_i(x)\psi(Y_i, z)]^2 = O(\phi_x(h)).$$

Under (H6) we have

$$S_n'^2 = O(n\phi_x(h)). \quad (4.19)$$

Now, we apply Fuk-Nagaev's inequality for $\Lambda_i(z)$. We have, for all $\ell > 0$ and for all $\varepsilon > 0$,

$$\begin{aligned} \mathbf{P} \left\{ \left| \mathbf{E} [\widehat{\Psi}_N^*(x, z)] - \widehat{\Psi}_N^*(x, z) \right| > \varepsilon \right\} &\leq \mathbf{P} \left\{ \left| \sum_{i=1}^n \Lambda_i(z) \right| > \varepsilon n \mathbf{E}[K_1(x)] \right\} \\ &\leq C(A_1'(x) + A_2'(x)) \end{aligned}$$

where

$$A_1' = \left(1 + \frac{\varepsilon^2 n^2 (\mathbf{E}[K_1(x)])^2}{S_n'^2 \ell} \right)^{-\ell/2} \quad \text{and} \quad A_2' = n \ell^{-1} \left(\frac{\ell}{\varepsilon n \mathbf{E}[K_1(x)]} \right)^{a+1}.$$

In particular, for $\varepsilon = \epsilon_0 \frac{\sqrt{n \log n \phi_x(h)}}{n \mathbf{E}[K_1(x)]}$ and $\ell = C(\log n)^2$, we obtain by (H6)

$$d_n A_2' \leq C n^{3/2-(a+1)/2} \phi_x(h)^{-(a+1)/2} (\log n)^{(3a-1)/2} \leq C n^{-1-\nu_1'} \quad (4.20)$$

for some $\nu_1' > 0$ and

$$d_n A_1' \leq C \left(1 + \frac{\epsilon_0^2 \log n}{\ell} \right)^{-\ell/2} \leq C n^{-1-\nu_2'} \quad \text{for some } \nu_2' > 0. \quad (4.21)$$

From (4.20) and (4.21) we obtain

$$\sum_n d_n \max_{z \in \mathcal{G}_n} \mathbf{P} \left(\left| \widetilde{\Psi}_N^*(x, z) - \mathbf{E}[\widetilde{\Psi}_N^*(x, z)] \right| > \epsilon_0 \left(\sqrt{\frac{\log n}{n \phi_x(h)}} \right) \right) < \infty.$$

■ **Lemma 5** [?]

Under Hypotheses (H1), (H2)((i), (iii)) and (H5), we have,

$$\sup_{t \in [\theta_x - \delta, \theta_x + \delta]} \left| \mathbf{E} [\widetilde{\Psi}_N(x, t)] - \tau \Psi(x, t) \right| = o(1).$$

Furthermore if we add (H2)(ii), we have

$$\sup_{t \in [\theta_x - \delta, \theta_x + \delta]} \left| \mathbf{E} [\widetilde{\Psi}_N(x, t)] - \tau \Psi(x, t) \right| = O(h^{b_1}).$$

Lemma 6

Under Hypotheses (H1) and (H3)-(H5), we have,

$$\sup_{t \in [\theta_x - \delta, \theta_x + \delta]} \left| \widehat{\Psi}_N(x, t) - \widetilde{\Psi}_N(x, t) \right| = O\left(n^{-\frac{1}{2}}\right) \quad a.s.$$

Proof. By a simple algebra we write for all $t \in [\theta_x - \delta, \theta_x + \delta]$

$$\begin{aligned} \left| \widehat{\Psi}_N(x, t) - \widetilde{\Psi}_N(x, t) \right| &= \left| \frac{\tau_n - \tau}{n\mathbf{E}[K_1]} \sum_{i=1}^n \frac{1}{G_n(Y_i)} K\left(\frac{d(x, X_i)}{h}\right) \psi(Y_i, t) \right. \\ &\quad \left. + \tau \left(\frac{G_n(Y) - G(Y)}{G_n(Y)G(Y)} \right) \frac{1}{n\mathbf{E}[K_1]} \sum_{i=1}^n K\left(\frac{d(x, X_i)}{h}\right) \psi(Y_i, t) \right| \\ &\leq \left[\frac{|\tau_n - \tau|}{G_n(a_F)} + \tau \frac{\sup_{y \in [a, b]} |G_n(y) - G(y)|}{G_n(a_F)G(a_F)} \right] \\ &\quad \times \frac{1}{n\mathbf{E}[K_1]} \sum_{i=1}^n K\left(\frac{d(x, X_i)}{h}\right) \psi(Y_i, t). \end{aligned}$$

Similarly to Lemma 4.4 we use the result of Ould Saïd and Tatachak (2009) for which we have $|\tau_n - \tau| = O_{a.s.}(n^{-\frac{1}{2}})$ and the remark 6 of Woodroffe(1985)) where we obtain $|G_n(a_F) - G(a_F)| = O_{a.s.}(n^{-\frac{1}{2}})$. Furthermore, for the second part, we recall that this term has been studied by Attouch *et al.* (2012) where we show that

$$\sup_{t \in [\theta_x - \delta, \theta_x + \delta]} \frac{1}{n\mathbf{E}[K_1]} \sum_{i=1}^n \left[K\left(\frac{d(x, X_i)}{h}\right) \psi(Y_i, t) - \mathbf{E} \left[K\left(\frac{d(x, X_i)}{h}\right) \psi(Y_i, t) \right] \right] = o(1)$$

It follows that

$$\sup_{t \in [\theta_x - \delta, \theta_x + \delta]} \frac{1}{n\mathbf{E}[K_1]} \sum_{i=1}^n K\left(\frac{d(x, X_i)}{h}\right) \psi(Y_i, t) = O_{a.s.}(1).$$

Thus, we have

$$\left| \widehat{\Psi}_N(x, t) - \widetilde{\Psi}_N(x, t) \right| = O_{a.s.}\left(n^{-\frac{1}{2}}\right)$$

which completes the proof of Lemma 4.4. ■ **Corollary 1**

Under Assumptions (H1), (H2) ((i)-(ii)), (H3)-(H5) and if $\Psi'(x, \theta_x) \neq 0$, then $\widehat{\theta}_x$ exists a.s. for all sufficiently large n such that

$$\exists C > 0 \quad \frac{\partial}{\partial t} \Psi(x, \xi_n) > C.$$

Proof. It is clear that, if $\psi_x(Y, \cdot)$ is increasing, we have for all $\epsilon > 0$

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

The results of Lemmas 1-6 show that

$$\widehat{\Psi}(x, t) \longrightarrow \Psi(x, t) \text{ a. s.}$$

for all real fixed $t \in [\theta_x - \delta, \theta_x + \delta]$. So, for sufficiently large n and for all $\epsilon \leq \delta$

$$\widehat{\Psi}(x, \theta_x - \epsilon) \leq 0 \leq \widehat{\Psi}(x, \theta_x + \epsilon) \text{ a. s..}$$

Since $\widehat{\Psi}(x, t)$ is a continuous function of t , there exists a $\widehat{\theta}_x \in [\theta_x - \epsilon, \theta_x + \epsilon]$ such that $\widehat{\Psi}(x, \widehat{\theta}_x) = 0$. Concerning, the uniqueness of $\widehat{\theta}_x$, we point out that the latter is a direct consequence of the strict monotonicity of ψ_x , w.r.t. the second component and the positivity of K .

Finally, The second part of this corollary is a direct consequence of the regularity assumption (H2) (i) on $\Psi(x, \cdot)$ and the convergence

$$\widehat{\theta}_x \rightarrow \theta_x \text{ a.s. as } n \rightarrow \infty.$$

■ References

- ANDERSEN, P.K., BORGAN, O., GILL, R.D. and KEIDING, N. (1993). *Statistical Models Based on Counting Processes*. Springer, New York.
- ATTOUCH, M., LAKSACI, A. and OULD SAÏD, E. (2009). Asymptotic distribution of robust estimator in nonparametric functional models, *Comm. in 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, *J. Korean Statist. Soc.*, 39, 489-500.
- ATTOUCH, M., LAKSACI, A. and OULD SAÏD, E. (2012). Robust regression for functional time series data. *J. Japan Statist. Soc.*, 42, 125-143.
- 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.
- BOENTE, G. and FRAIMAN, R. (1990). Asymptotic distribution of robust estimators for nonparametric models from mixing processes. *Ann. Statist.*, 18, 891-906.

BOENTE, G., GONZALEZ-MANTEIGA, W. and PÉREZ-GONZALEZ, A. (2009). Robust nonparametric estimation with missing data. *J. Statist. Plann. Inference.* 139, 571-592.

BOGACHEV, V.I. *Gaussian Measures*. Math surveys and monographs Amer. Math. Soc. 62, 1999.

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

BONGIORNO, E.G., SALINELLI, E., GOIA, A. and VIEU, P. (Eds.) *Contributions in infinite-dimensional statistics and related topics*. Società editrice Esculapio. Bologna. 2014.

BRADLEY, R. C. (2007). *Introduction to strong mixing conditions*. Vol I-III, Kendrick Press, Utah.

CHAOUCH. M., LAÏB, N. and OULD SAÏD E. (2014). Nonparametric M-estimation for right censored regression model with stationary ergodic data. Under revision.

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

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

CRAMBES, C., DELSOL, L. and LAKSACI, A. (2008). Robust nonparametric estimation for functional data, *J. Nonparametric Statist.*, 20, 573-598.

DEDECKER, J. DOUKHAN, P. LANG, G. LEON, JR. LOUHICHI, S. and PRIEUR, C. (2007). Weak dependence : With examples and applications. *Lecture Notes in Statistics*, 190 Springer-Verlag, New York.

DERRAR, S., A. LAKSACI and OULD SAÏD, E. (2015). On the nonparametric estimation of the functional ψ -regression for a random left-truncation model. *Journal of statistical Theory and Practice*, 9, 823-849.

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

FAN, J. HU, T. C. and TRUONG, Y. K. (1994). Robust nonparametric function estimation. *Scand. J. Statist.* 21, 433-446.

FERRATY, F. and VIEU, P. *Nonparametric Functional Data Analysis. Theory and Practice*, Springer-Verlag, New York, 2006.

FERRATY, F. and ROMAIN, Y. (2011). *The Oxford Handbook of Functional Data Analysis*. Oxford University Press.

GHERIBALLAH, A. LAKSACI, A. and SEKKAL, S. (2013). Nonparametric M -regression for functional ergodic data. *Statist. Probab. Lett.*, 83, 902-908.

HE, S. and YANG, G. (1994). Estimating a lifetime distribution under different sampling plan. In S.S. Gupta J.O. Berger (Eds.) *Statistical decision theory and related topics* (Vol. 5. pp. 73-85) Berlin Heidelberg, NewYork : Springer.

HE, S. and YANG, G. (1998). Estimation of the truncation probability in the random truncation model. *The Annals of Statistics*, 26, 1011-1027.

HORRIGUE, W. and OULD SAÏD, E. (2011). Strong uniform consistency of a nonparametric estimator of a conditional quantile for censored dependent data and functional regressors. *Random Oper. Stoch. Equa.*, 19, 131-156.

HORRIGUE, W. and OULD SAÏD, E. (2015). Nonparametric regression quantile estimation for dependent functional data under random censorship : Asymptotic normality. *Comm. in Statistics. Theory and Methods*, 44, 4307-4332.

HORVÁTH, L. and KOKOSZKA, P. (2012). *Inference for Functional Data with Applications*. Springer-Verlag, New York. 200.

HUBER, P.J. (1964). Robust estimation of a location parameter. *Annals of the Mathematical Statistics*, 35, 73-101.

LAÏB, N. and OULD SAÏD, E. (2000). A robust nonparametric estimation of the autoregression function under an ergodic hypothesis. *Canadian Journal of Statistics*, 28, 817-828.

LEMDANI, M. and OULD SAÏD, E. (2015). Nonparametric robust regression estimation for censored data *Statistical Papers*, DOI : 10.1007/s00362-015-0709-8.

LI, W. V. and SHAO, Q. M. (2001). *Gaussian processes : inequalities, small ball*

probabilities and applications. In Stochastic processes : Theory and Methods, Ed. C.R. Rao and D. Shanbhag. Hanbook of Statistics, 19 : North-Holland, Amsterdam.

LOÈVE, M. *Probability theory I*. Springer-Verlag. New-York, 1977.

LYNDEN-BELL, D. (1971) A method of allowing for known observational selection in small samples applied to 3CR quasars. *Monthly Notices of the Royal Astronomical Society*, 155, 95–118.

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

OULD SAÏD, E. and LEMDANI, M. (2006). Asymptotic properties of a nonparametric regression function estimator with randomly truncated data. *Ann. Inst. Statist. Math.*, 58, 357–378.

OULD SAÏD, E. and TATACHAK, A. (2009) Strong consistency rate for the kernel mode estimator under strong mixing hypothesis and left truncation. *Comm. Statist. Theory Methods*, 38. 1154–1169

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

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

RIO, E. Théorie asymptotique des processus aléatoires faiblement dépendants. *Mathématiques & Applications 31*, Springer-Verlag, Berlin.

STONE, C. (1981). Optimal rates of convergence for nonparametric regression. *Ann. Statist.*, 9, 1348-1360.

STUTE, W. (1993). Almost sure representations of the product-limit estimator for truncated data. *Ann. Statist.*, 21, 146–156.

WANG, J., LIANG, H.Y. and FAN, G. (2012a). Asymptotic properties for an M-estimator of the regression function with truncation and dependent data. *Journal of the Korean Statistical Society*, 41, 35–367.

WANG, J., LIANG, H.Y. and FAN, G. (2012b). Local M-estimation of nonparametric regression with left-truncated and dependent data. *Scientia Sinica Mathematica*,

42, 995–1015.

WOODROOFE, M. (1985). Estimating a distribution function with truncated data. *The Annals of Statistics*, **13**, 163–177.

Chapitre 5

Conclusion et perspectives

5.1 Conclusion

Dans cette thèse, nous avons traité la ψ -estimation qui est un type de M-estimation de la fonction de régression, lorsque le régresseur est fonctionnel et la variable réponse est tronquée aléatoirement à gauche par une autre variable aléatoire en utilisant la méthode de noyau. Après avoir rappelé quelques résultats existant dans ce contexte, on a établi la convergence forte et la normalité asymptotique de l'estimateur robuste pour des données indépendantes puis pour des données α -mélangeantes.

Du point de vue théorique, notre modèle est une généralisation de la régression classique dans le sens que cette dernière est obtenue en prenant $\psi_x(y, t) = (y - t)$. Cependant, d'un point de vue pratique, notre modèle fournit une estimation alternative de la régression classique qui a plus d'avantage en sens de robustesse. La supériorité de notre modèle par rapport à la régression classique dans la présence des valeurs atypiques est confirmée par notre étude de simulation.

En outre, la méthode d'estimation robuste proposée peut être utilisée pour robustifier d'autres méthodes d'estimation de la régression comme : la méthode k-NN, le modèle linéaire local, le modèle d'indice fonctionnel, le modèle linéaire partiel,...

Ainsi, on peut dire que notre contribution est très importante dans la théorie comme dans la pratique. Cet impact est évalué par la généralité et la robustesse du modèle étudié, l'exactitude et l'utilité des résultats asymptotiques obtenus et aussi par les nombreux problèmes ouverts en statistique nonparamétrique.

5.2 Perspectives

Le sujet que nous avons étudié dans cette thèse, nous permet d'envisager à entamer plusieurs pistes de recherche :

- Les résultats établis au deuxième chapitre pour la normalité asymptotique peuvent être étendus au cas où le modèle présente une certaine forme de dépendance(α -mélange).
- Au meilleur de nos connaissances, il n'existe pas d'estimateurs récurrents pour des données tronquées, que ce soit pour la densité ou la fonction de régression.
- il nous semble possible d'étendre les résultats obtenus pour d'autres type de variable : censurées, associées, quasi-associées ou spatiales, ou le cas où la variable réponse est fonctionnelle qui peut être un sujet de recherche très prometteur.

Bibliographie

1. Attouch, M., Laksaci, A. and Ould Saïd, E. (2009). Asymptotic distribution of robust estimator in nonparametric functional models, *Comm. in Statist. Theory and Methods*, **38**, 1317-1335.
2. Attouch, M., Laksaci, A. and Ould Saïd, E. (2010). Asymptotic normality of a robust estimator of the regression function for functional time series, *J. Korean Statist. Soc.*, **39**, 489-500.
3. Attouch, M., Laksaci, A. and Ould Saïd, E. (2012). Robust regression for functional time series data. *J. Japan Statist. Soc.*, **42**, 125-143.
4. Azzedine, N., Laksaci, A. and Ould Saïd, E. (2008). On the robust nonparametric regression estimation for functional regressor, *Statist. Probab. Lett.*, **78**, 3216-3221.
5. Benhenni, K., Ferraty, F., Rachdi, M. et P. Vieu (2007). Local smoothing regression with functional data. *Comp. Statistics*, **22**, 353-369.
6. Boente, G. and Fraiman, R. (1989). Nonparametric regression estimation. *J. Multivariate Anal.*, **29**, 180-198.
7. Bosq, D. (2000). Linear Processes in Function Spaces : Theory and Applications Lecture Notes in Statistics *Springer-Verlag*, **149**, New York.
8. Burba, F., Ferraty, F. et Vieu, P. (2008). Propriétés asymptotiques de l'estimateur kNN en régression fonctionnelle non-paramétrique. *C. R. Math. Acad. Sci. Ser.*, **I 346**, 339-342.
9. Bertrand, Retali Convergence uniforme de certains estimateurs de la densité par la méthode du noyau. *C.R.A.S. Série A*, **279**, 527-529.
10. Cadre, B. (2001). Convergent estimators for the L_1 -median of a Banach valued random variable. *Statistics*, **35 (4)**, 509-521.
11. Chaffai, M. E. (1989). Les estimateurs robustes sont-ils vraiment robustes en pratique?. *Revue de Statistique Appliquée*, **37**, No.3, 57-73.
12. Collomb, G. and Härdle, W. (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.
13. Crambes C, Delsol L. and Laksaci A. (2008). Robust nonparametric estimation for functional data. *J. Nonparametric Statist.*, **20**, 573-598.
14. Derrar, S., A. Laksaci and Ould Saïd, E. (2015). On the nonparametric estimation of the functional ψ -regression for a random left-truncation model. *Journal of statistical Theory and Practice*, **9**, 823-849.

15. Delsol, L. (2007a). CLT and L_q errors in nonparametric functional regression *C. R. Math. Acad. Sci*, **345** (7), 411-414.
16. Ferraty, F. et Vieu, P. (2002). The functional nonparametric model and application to spectrometric data. *Comput. Statist*, **17** (4), 545-564.
17. Ferraty, F. and Vieu, P. (2003). Curves discrimination : a nonparametric functional approach *Computational Statistics and Data Analysis*, **44** (1-2), 161-173.
18. Ferraty, F. and Vieu, P. (2006). Nonparametric Functional Data Analysis. *Theory and Practice*. Springer-Verlag, New York.
19. He, S. and Yang, G. (1998). Estimation of the truncation probability in the random truncation model. *The Annals of Statistics*, **26**, 1011-1027.
20. Huber, P.J. (1964). Robust estimation of a location parameter. *Annals of the Mathematical Statistics*, **35**, 73-101.
21. González-Manteiga W. and Vieu P. (2007). Editorial Statistics for Functional Data. *Comp. Stat. Data Analysis*, **51** (10), 4788-4792.
22. Kaplan, E. L. and Meier, P. (1958) Nonparametric estimation from incomplete observations. *J. Amer. Statist. Assoc.*, **53**, 457-481.
23. Lemdani, M. and Ould Saïd, E. (2015). Nonparametric robust regression estimation for censored data *Statistical Papers*, DOI : 10.1007/s00362-015-0709-8.
24. Lecoutre, J.P. (1990) Uniform consistency of a class of regression function estimators for Banach-space valued random variable. *Statist. Probab. Lett.* **10**, 145-149.
25. Masry, E. (2005). Nonparametric regression estimation for dependent functional data : asymptotic normality Stochastic Process. *Appl*, **115** (1), 155-177.
26. Nadaraya, E.A (1964). On estimating regression. *theor. Prob. Appl*, **10**, 186-196.
27. Ould Saïd, E. and Lemdani, M. (2006). Asymptotic properties of a nonparametric regression function estimator with randomly truncated data. *Ann. Inst. Statist. Math.*, **58**, 357-378.
28. Ould Saïd, E. and Tatachak, A. (2009) Strong consistency rate for the kernel mode estimator under strong mixing hypothesis and left truncation. *Comm. Statist. Theory Methods*, **38**. 1154-1169.
29. Ramsay, J. and Silverman, B. (1997). Functional Data Analysis *Springer-Verlag*, New York.
30. Ramsay, J. and Silverman, B. (2002). Applied functional data analysis : Methods and case studies *Spinger-Verlag*, New York.
31. Ramsay, J. and Silverman, B. (2005). Functional Data Analysis (Second Edition) *Spinger-Verlag*, New York.

-
32. Stute, W. (1993). Almost sure representations of the product-limit estimator for truncated data. *Ann. Statist.*, **21**, 146-156.
 33. Wang, J., Liang, H. Y. and Fan, G. (2012a). Asymptotic properties for an M-estimator of the regression function with truncation and dependent data. *Journal of the Korean Statistical Society*, **41**, 35-367.
 34. Wang, J., Liang, H. Y. and Fan, G. (2012b). Local M-estimation of nonparametric regression with left-truncated and dependent data. *Scientia Sinica Mathematica*, **42**, 995-1015.
 35. Watson, G.S. (1964). Smooth regression analysis. *Sankhya Ser, A* **26**, 359-372.
 36. Woodroffe, M. (1985). Estimating a distribution function with truncated data. *The Annals of Statistics*, **13**, 163-177.