

N° d'ordre :

REPUBLIQUE ALGERIENNE DEMOCRATIQUE & POPULAIRE
MINISTERE DE L'ENSEIGNEMENT SUPERIEUR & DE LA RECHERCHE
SCIENTIFIQUE



UNIVERSITE DJILLALI LIABES
FACULTE DES SCIENCES EXACTES
SIDI BEL ABBES

THESE DE DOCTORAT de 3^{ème} Cycle

Présentée par

Mademoiselle RAFAA Fatima

Domaine : Mathématiques–Informatique.

Filière : Mathématiques.

***Intitulé de la formation : Statistique Mathématiques Appliquées
à l'économie et la finance.***

Intitulée

***Estimation non paramétrique locale linéaire par k
plus proches voisins***

Soutenue le 04/07/2018.

Devant le jury composé de :

Président : GHERIBALLAH Aek

Examineur : BENCHIKH Tawfik

Directeur de thèse : ATTOUCH Med Kadi.

Co-directeur de thèse: LAKSACI Ali

MCA Univ Sidi Bel Abbès

Pr. Univ. Sidi Bel Abbès

Pr. Univ. Sidi Bel Abbès

Pr. Univ Abha, Arabie Saoudite

2017/2018

1	Introduction	3
1.1	Introduction Générale	4
1.1.1	La statistique fonctionnelle	4
1.1.2	L'analyse locale linéaire	6
1.1.3	La méthode kNN	7
1.2	Présentation des modèles étudiés	9
1.2.1	L'estimateur locale linéaire	9
1.2.2	L'estimateur Knn	10
1.3	Résultats existants	10
1.3.1	Résultats existants sur la régression locale linéaire en statis- tique fonctionnelle	11
1.3.2	Résultats existants sur la régression kNN locale en statistique fonctionnelle	11
2	L'estimation locale linéaire de la régression non paramétrique fonc- tionnelle par la méthode des k plus proches voisins	13
2.1	Introduction	14
2.2	Le Modèle et son estimateur	15
2.3	Convergence presque complète ponctuelle	16
2.4	Convergence presque complète pour un k fixé	17
3	On the local linear estimate for functional regression : Uniform in bandwidth consistency	25
3.1	Introduction	26
3.2	The local linear estimate for functional regression	28
3.3	Main result	30
3.4	Discussions and Applications	31
3.4.1	Discussions	31
3.4.2	Simulation results	33
3.4.3	Application on a real data	36
3.5	Appendix	38
4	On the k-Nerest Neighbor local linear estimation for functional regres- sion : Uniform and bandwidth consistency	49
4.1	introduction	50
4.2	The representation of the estimator :	51
4.3	useful results and assumptions	52
4.4	Uniform convergence of kNN local linear regression estimator	53

.1	Appendix	54
A	Conclusion	61

Contents

4.1 introduction	50
4.2 The representation of the estimator :	51
4.3 useful results and assumptions	52
4.4 Uniform convergence of kNN local linear regression estimator	53
.1 Appendix	54

1.1 Introduction Générale

1.1.1 La statistique fonctionnelle

Le développement technologique constaté ces dernières années a imposé la modernisation des méthodes statistiques à l'effet d'analyser et de contrôler les différents paramètres observés tels que, changement climatique, pollution, l'évolution des indices boursiers. Ainsi est développée une nouvelle branche de la statistique appelée statistique fonctionnelle qui étudie des observations qui ne sont pas des variables réelles ou vectorielles mais des données qui sont dans espace de dimension infinie.

Ces données fonctionnelles sont employées dans plusieurs domaines. Nous citons à titre d'exemple les auteurs qui les ont utilisées : Frank et Friedman (1993), Ferraty et Vieu (2002), Abraham, Cornillon, Matzner et Molinari (2003) les ont utilisées dans la chimie ;

Hall, Poskitt, et Presnell (2001), Dabo-Niang, Ferraty, et Vieu (2004) dans les radars, Ramsay, Altman, et Bock, 1994 ; Gasser, hall, et Presnell (1998) dans la biométrie, Abramovitch, Antoniadis, Sapatinas, et Vidakovic (2004) ; Abramovitch et Angelini (2006) ; Antoniadis et Sapatinas (2007) dans la physiologie.

D'autres exemples existent dans les livres : Ramsay et Silverman (2002, 2005) ; Ferraty et Vieu (2006) ; Dabo-Niang et Ferraty (2008).

Les premiers apports sur ce sujet ont été consacrés à l'étude des modèles paramétriques ; Ramsay et Silverman en (1997, 2002, 2005) ont étudié le cas indépendant , Bosq (2000) le cas dépendant.

Le traitement non paramétrique des données fonctionnelles est beaucoup plus inédit que l'analyse paramétrique. Effectivement, les premiers résultats obtenus par Gasser et al. (1998) et qui se sont intéressés à l'estimation non paramétrique du mode de la distribution d'une variable fonctionnelle vérifiant une la condition fractale.

Sur la même condition fractale, Ferraty et Vieu (2000) ont étudié la convergence presque complète d'un estimateur à noyau de la fonction de régression, lorsque les observations sont indépendantes et identiquement distribuées ; Ferraty et Vieu (2004) ont généralisé ces derniers résultats au cas mélangeant.

En 2002, Dabo-Niang a obtenu la convergence presque sûre et la normalité asymptotique d'un estimateur de type histogramme de la densité d'une variable aléatoire dans un espace de dimension infinie.

En utilisant le développement récent de la théorie des petites boules de probabilités, 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 et à la classification des courbes a été proposée par ces auteurs. En (2005) Masry a étudié la normalité asymptotique de l'estimateur 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.

La convergence en moyenne quadratique a été étudiée par Ferraty et al. (2007), ils ont explicité exactement le terme asymptotique de l'erreur quadratique. Ce résultat a été utilisé par Rachdi et Vieu (2007) pour déterminer un critère de sélection automatique du paramètre de lissage basé sur la validation croisée. La version locale de ce critère a été étudiée par Benhenni et al. (2007). Nous trouverons dans cet article une étude comparative entre l'approche locale et globale. Comme travaux bibliographiques récents en régression, nous renvoyons le lecteur à l'ouvrage de Ferraty et Vieu (2011) ainsi qu'à Delsol (2011). Des résultats sur l'uniforme intégrabilité ont été établis par Delsol (2007,2009) et Delsol et al. (2011). D'autres travaux se sont intéressés à l'estimation de la fonction de régression en utilisant différentes approches : la méthode des k plus proches voisins par Burba et al. (2008), les techniques robustes par Azzidine et al. (2008), Attouch et al. (2009) et Crambes et al. (2008).

Delsol (2008) a étudié la régression sur les variables fonctionnelles en donnant des résultats liés aux propriétés asymptotiques de l'estimateur à noyau généralisé au cas d'une variable explicative fonctionnelle, Laïb et Louani (2010-2011) ont étudié l'estimation de la fonction de régression fonctionnelle, Attouch et Benchikh (2012) ont établis la normalité asymptotique de l'estimation robuste de la fonction de régression.

Notons que le modèle le plus fréquemment rencontré en statistique non paramétrique est celui de régression qui décrit la relation entre deux ou plusieurs variables aléatoires.

La présentation dans notre thèse s'intéresse à l'estimateur non paramétrique localement linéaire, lorsque la variable explicative est fonctionnelle et la variable réponse est réelle.

1.1.2 L'analyse locale linéaire

La théorie de l'estimation non paramétrique de la fonction de régression s'est développée autour des méthodes de constructions des estimateurs, de leurs propriétés statistiques, de leurs optimalités, ainsi que leur estimation adaptative.

En 1964 et pour la première fois, Nadaraya et Watson ont estimé la fonction de régression en utilisant la méthode du noyau qui a connu par la suite un grand développement par plusieurs auteurs. Il est à noter qu'il existe d'autres méthodes de régression, telle la méthode par polynômes locaux dont son utilisation est préférée à celle du noyau.

Malgré le fait que l'estimateur par la méthode du noyau soit très simple et facile à comprendre, elle présente certaines anomalies. En 1996, Simonoff a révélé des problèmes de biais aux extrémités des données, de manque de variation locale du lissage et également d'aplatissement des pics et des vallées. Pour ces causes, l'utilisation de la méthode par polynômes locaux est préférable à celle de la méthode du noyau, en effet la variance des deux méthodes est la même, mais le biais aux bornes des données de la méthode par polynômes locaux est d'ordre $O(h^{p+1})$, alors qu'il est d'ordre $O(h)$ pour la méthode du noyau. La régression par polynômes locaux est très similaire à l'estimation du noyau, ce qui diffère c'est que l'estimation est obtenue par régression polynomiale de y sur x . Il est important de souligner le fait que la méthode du noyau est en réalité qu'un cas particulier de la méthode par polynômes locaux.

Lorsque le p prend la valeur de 1, nous sommes devant une régression localement linéaire dont son estimateur a été trouvé par Simonoff(1996) pour une variable explicative vectorielle, Le lecteur pourra se référer à Lejeune (1985), Muller (1988), Tsybakov (1986,1990), Fan (1992, 1993), Hastie et Loader (1993). Fan et Gijbels (1992, 1995) ont étudié et comparé cette méthode d'estimation sur plusieurs modèles non paramétrique et ils ont aussi abordé la question du choix de différents paramètres intervenants dans la définition de l'estimateur tel le paramètre de lissage, le noyau, ainsi que le degré du polynôme.

Wand (1994) et Fan et al. (1997) ont traité le cas des variables indépendantes et les propriétés asymptotiques de l'estimateur localement linéaire; Masry et Fan (1997), Masry (1996), Hardle et Tsybakov (1997), Hardle et al. (1998), Opsomer (1997) et Vilar-Fernández (1998, 2000) se sont intéressés aux cas dépendant en appliquant des conditions de mélange de l'estimateur polynôme local et ils ont étudié la convergence en moyenne quadratique en précisant le terme biais et la variance asymptotique; Kang et al. (2011) ont abordé le cas où la variable réponse est censurée. Chu et Marron (1991), Fan (1992) et Hastie et Loader (1993) ont comparé entre la méthode du noyau et l'ajustement par la méthode locale linéaire, La contribution de Hafen(2010) sur l'ajustement local linéaire constitué

un avancée importante dans le domaine aussi bien en pratique qu'en théorique. Beaucoup plus récemment Vidaurre et al. (2012) combinent l'approche lasso avec l'ajustement localement linéaire, dont l'objectif est de trouver une alternative estimation de la lasso-régression de Tibshirani(1996).

En 2006, Welsh en collaboration avec Yee ont abordé le cas où la variable réponse est multi dimensionnelle. Ils ont évalué le biais et la variance asymptotique de l'estimateur local linéaire de la fonction de régression et sa dérivée. Irizarry(2001), généralise la régression linéaire.

L'étude consacrée à la méthode d'estimation par polynômes locaux dans le cadre fonctionnel a été très limitée. Nous citons les quelques études réalisées ci-après : L'article de Baillo and Grané (2009) sont les premiers dans ce domaine. Ils ont établi la convergence en norme L2 d'une méthode locale linéaire lorsque la variable explicative est hilbertienne, en (2007), Barrientos avait approfondi l'estimation fonctionnelle par polynômes locaux en généralisant le résultat obtenu par Fan et Gibels(1996), et en étudiant aussi la convergence presque complète de l'estimateur de la fonction de régression estimée par la méthode localement linéaire. En 2006, Dony et al ont étudié l'estimateur et démontré la convergence uniforme presque sûre dans le cas non fonctionnel.

La performance de la méthode locale linéaire en comparant l'estimateur de Nadaraya-Watson et la méthode de Monte-Carlo dans le cas fonctionnel a été établie par Baillo et Grané. Un procédé alternatif pour l'estimation locale linéaire fonctionnelle a été réalisé par Boj et al. En 2012, Chouaf et Laksaci ont généralisé les résultats de Barrientos et al. aux données spatialement dépendantes.

Devant les résultats de convergence trouvés par Barrientos(2007) en tenant compte du paramètre de lissage déterministe, la question qui se pose :

Est-ce que les résultats restent valables lorsque le paramètre de lissage est aléatoire ?

Dans le même ordre d'idées, nous allons étudier la convergence uniforme sur le paramètre de lissage de l'estimateur de Barrientos.

1.1.3 La méthode kNN

Notons que l'estimateur à noyau et celui localement linéaire prennent le paramètre de lissage h comme un nombre réel positive et si h est pris dans un ensemble discret, dans ce cas on parle d'estimateur kNN de la fonction de régression.

La méthode des plus proches voisins notée parfois k-PPV ou k-NN pour (Nearest Neighbor) consiste à classer et déterminer pour chaque nouvel individu la liste des plus proches voisins parmi les individus déjà classés. Cette méthode raisonne

avec le dicton « dis- moi qui sont tes amis, je te dirais qui tu es ». Exemple : Si l'on veut prédire la probabilité de survenue d'un cancer chez un nouveau patient on procède en deux étapes : On recherche selon les caractéristiques de ce patient avec ceux qui lui ressemblent ; si parmi ses voisins, il y a eu plus de cancer, alors le patient a une forte probabilité de le contacter.

La méthode kNN nécessite de choisir une distance, la plus classique est la distance euclidienne, et le nombre de voisins à prendre en compte.

L'avantage de cette méthode permet de construire un voisinage adapté à la structure locale des données. Cependant, quand les données sont hétérogènes, la méthode k-NN résout ce cas de données par le choix du nombre des voisins et lorsque les données sont fonctionnelles, la méthode k-NN devient efficace grâce au paramètre adaptatif k , en outre dans le cas fonctionnelle, la méthode k-NN respecte la structure locale des données. De même son utilisation pratique est facile car elle contrôle un seul paramètre k (Voir Ferraty et Vieu(2006)).

Néanmoins sa démonstration pose certaines difficultés car la sélection des plus proches voisins donne une variable aléatoire et dans ce cas nous ne pouvons pas utiliser les mêmes outils que ceux de la méthode des noyaux standards.

Néanmoins sa démonstration pose certaines difficultés car la sélection des plus proches voisins donne une variable aléatoire et dans ce cas nous ne pouvons pas utiliser les mêmes outils que ceux de la méthode des noyaux standard.

Cover(1968), Stone (1977), Collomb (1981) or Devroye et al. (1994) se sont penchés sur la méthode kNN dans le cas données vectorielle tandis que le cas des données fonctionnelles quelques études ont été faites d'autre auteurs avaient aussi l'intérêt d'étudier cet estimateur, trouvons par exemple Collomb (1980) qui a montré les différents types de convergence (probabilité et presque complète), Mack (1981) a étudié la convergence en L^2 et la distribution asymptotique, la convergence uniforme est donnée par Devroye (1978) et (1981). Pour le cas des données fonctionnelles, Ferraty et Vieu (2004) ont commencé par une introduction sur l'estimation des k-NN, Burba et al. (2008) ont obtenu la convergence presque complète de l'estimateur de la fonction de régression avec des données indépendantes et identiquement distribuées, Burba et al. (2009) for the pointwise almost complete consistency, Kudraszow and Vieu (2013) for the uniform almost complete convergence and to Lian (2011) for the functional regressor case. Attouch et Benchikh (2012) ont établi la normalité asymptotique de la fonction de régression, Belabed et Attouch The k-nearest neighbors estimation of the conditional hazard function for functional data.

Laksaci et Kara(2017) ont discuté l'importance de la convergence uniforme sur le nombre des voisins.

Nous allons combiner les idées de la régression locale linéaire à ceux de la méthode du kNN afin de construire un estimateur localement adaptatif avec un biais réduit. Comme c'est indiquée dans Burba et al. (2009).

Deux propriétés asymptotiques de cet estimateur seront établies. La première est la convergence presque complète ponctuelle, la deuxième est la convergence presque complète uniforme sur le nombre des voisins k .

Ensuite, sous certaines hypothèses nous établissons la convergence presque complète uniforme sur le nombre des voisins k et sur l'ensemble $S_{\mathcal{F}}$.

1.2 Présentation des modèles étudiés

1.2.1 L'estimateur locale linéaire

la régression non paramétrique est un outil statistique permettant de décrire la relation entre une variable dépendante Y et une variable explicative X .

Nous proposons d'étudier le modèle de régression non paramétrique de la forme :

$$Y = R(X) + \epsilon, \quad \mathbb{E}(\epsilon|X) = 0$$

Lorsque la variable explicative X est évaluée dans certains espaces de dimension infinis $\mathcal{H}$ et Y est la réponse scalaire

on considère n paires $(X_i, Y_i)_{i=1, \dots, n}$ identically and independently distributed as (X, Y) , nous essayons de résoudre le problème de minimisation suivant :

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

1

Nous obtenons l'estimateur de Barriontos(2010)

$$\widehat{R}(x) = \frac{\sum_{i,j=1}^n W_{ij}(x) Y_j}{\sum_{i,j=1}^n W_{ij}(x)}$$

where

$$W_{ij}(x) = \beta(X_i, x) \left(\beta(X_i, x) - \beta(X_j, x) \right) K(h^{-1}d(x, X_i)) K(h^{-1}d(x, X_j))$$

où K est une fonction à noyau univarié, d mesure la proximité d'un élément $\mathcal{H}$ par rapport à un autre et $\beta(\cdot, \cdot)$ est un opérateur inconnu $\mathcal{H}^2$ in $\mathbb{R}$ $h = h_{k,n}$ is a sequence of positive real numbers

9

1.2.2 L'estimateur Knn

lestimateur a noyau prend le paremetre de lissage h comme un nombre réel positive et si h est pris dans un ensemble discret finis, dans ce cas on parle lestimateur kNN localement linéaire. nous essayons de résoudre le problème de minimisation suivant :

$$\min_{(a,b) \in \mathbb{R}^2} \sum_{i=1}^n (Y_i - a - b\beta(X_i, \chi))^2 K(H_{k,x}^{-1}d(\chi, X_i)) \quad 2$$

où K est une fonction à noyau univarié, d mesure la proximité d'un élément $\mathcal{H}$ by par rapport à un autre et $\beta(\cdot, \cdot)$ est un opérateur inconnu $\mathcal{H}^2$ in $\mathbb{R}$, et nous avons conduit à la solution de 2 :

$$\begin{pmatrix} \hat{m}(\chi) \\ \hat{m}_1(\chi) \end{pmatrix} = (G'_\beta K Q_\beta)^{-1} G'_\beta K Y$$

where

$$G'_\beta = \begin{bmatrix} 1 & \dots & 1 \\ \beta(X_1, \chi) & \dots & \beta(X_n, \chi) \end{bmatrix}.$$

et $K = \text{diag}(K(H_{k,x}^{-1}d(X_1, \chi)), \dots, K(H_{k,x}^{-1}d(X_n, \chi)))$ et $Y = {}^t[Y_1, \dots, Y_n]$; ou

$H_{k,x} = \min \left\{ h \in \mathbb{R}^+ \mid \sum_{i=1}^n 1_{B(x,h)}(X_i) = k \right\}$ Il peut être réécrit comme suit :

$$\begin{pmatrix} \hat{m}(\chi) \\ \hat{m}_1(\chi) \end{pmatrix} = \frac{1}{\sum_{i=1}^n \sum_{j=1}^n \beta_i(\beta_i - \beta_j) K_i K_j} \begin{pmatrix} \sum_{i=1}^n \sum_{j=1}^n \beta_i(\beta_i - \beta_j) K_i K_j Y_j \\ \sum_{i=1}^n \sum_{j=1}^n (\beta_j - \beta_i) K_i K_j Y_j \end{pmatrix}.$$

$\hat{m}(\chi)$ Est l'estimateur de la fonction de régression $m(\chi)$.

1.3 Résultats existants

Hypothèses :

(H1) $\forall r > 0, \quad \mathbb{P}(r_1 < \delta(X, x) < r_2) := \phi_x(r_1, r_2)$ with $\phi_x(r) = \phi_x(-r, r) > 0$ and

$$\text{For all } s \in (0, 1), \quad \lim_{r \rightarrow 0} \frac{\phi_x(-r/2, sr)}{\phi_x(-r, r)} = \tau_x(s) < \infty.$$

(H2) The sequence (a_n) verifies :

$$\frac{\log n}{n \min(a_n, \phi_x(a_n))} \rightarrow 0.$$

(H3) La fonction $\phi_x(r) = \phi_x(-r, r)$, $r > 0$ est continue au voisinage de 0 et vérifie $\phi_x(r) > 0, \forall r > 0$.

(H4) Les moments conditionnels de Y vérifient $\forall m \geq 2, E[Y^m|X = \cdot] := \sigma_m(\cdot)$

(H5) La condition (H1) est vérifiée pour un $\phi_x(\cdot)$ telle que pour tout $s \in (0, 1)$,

$$\lim_{r \rightarrow 0} \frac{\phi_x(sr)}{\phi_x(r)} = \tau_x(s).$$

(H6) The functions φ and ψ_{S_F} verifies :

$$\text{For } n \text{ large enough, } \frac{(\log n)^2}{n\varphi(h_K)} < \psi_{S_F}\left(\frac{\log n}{n}\right) < \frac{n\varphi(h_K)}{\log n}$$

1.3.1 Résultats existants sur la régression locale linéaire en statistique fonctionnelle

Sous les hypothèses (H1)-(H2), on a

$$\sup_{a_n \leq h \leq b_n} |\hat{r}(\chi) - r(\chi)| = O(b_n^v) + O_{a.c} \left(\sqrt{\frac{\log n}{n\varphi_x(a_n)}} \right).$$

1.3.2 Résultats existants sur la régression kNN locale en statistique fonctionnelle

Sous les hypothèses (H3)-(H4), on a

$$\widehat{m}(x) - m(x) = O \left(\left(\phi_x^{-1} \left(\frac{k}{n} \right) \right)^\beta + \left(\sqrt{\frac{\log n}{k}} \right) \right) \quad p.co. \quad \text{3}$$

Sous les hypothèses (H3)-(H4), on a

$$\sup_{k_{1,n} \leq k \leq k_{2,n}} |\widehat{m}(x) - m(x)| = O \left(\phi_x^{-1} \left(\frac{k_{2,n}}{n} \right)^\beta \right) + O_{a.co.} \left(\sqrt{\frac{\log n}{k_{1,n}}} \right). \quad \text{4}$$

Sous les hypothèses (H3)-(H6), on a

$$\sup_{k_{1,n} \leq k \leq k_{2,n}} \sup_{x \in S_{\mathcal{F}}} |\hat{m}(x) - m(x)| = O\left(\varphi_x^{-1}\left(\frac{k_{2,n}}{n}\right)^v\right) + O_{a.c}\left(\sqrt{\frac{\psi_{S_{\mathcal{F}}}\left(\frac{\log n}{n}\right)}{k_{1,n}}}\right).$$

Chapitre 2: L'estimation locale linéaire de la régression non paramétrique fonctionnelle par la méthode des k plus proches voisins

Contents

L'estimation locale linéaire de la régression non paramétrique fonctionnelle par la méthode des k plus proches voisins

Mohammed ATTOUCH¹, Ali LAKSACI² et Fatima RAFAA³

²auteur Correspondant : Laboratoire de Statistique et Processus Stochastiques,
Université Djillali Liabes, Sidi Bel Abbes. alilak@yahoo.fr

^{1,3}Laboratoire de Statistique et Processus Stochastiques, Université Djillali
Liabes, Sidi Bel Abbes(attou_kadi@yahoo.fr ou rafaa_fatima@yahoo.com)

Résumé

Nous étudions l'estimation locale linéaire de l'opérateur de régression lorsque la variable explicative prend ses valeurs dans un espace semi-métrique. Nous construisons un estimateur par la méthode des k plus proches voisins. Deux propriétés asymptotiques de cet estimateur seront établies. La première est la convergence presque complète ponctuelle, la deuxième est la convergence presque complète uniforme sur le nombre des voisins.

Abstract

Local linear estimate of the nonparametric function regression by the k NN method We consider the problem of the local linear estimate of the regression operator when the regressor is functional. We construct an estimator by the k NN method and we study its asymptotic proprieties. Precisely, we establish the almost complete consistency of this estimator with rate in both pointwise and uniform in number of neighbors.

2.1 Introduction

Le cadre général de cette note est la modélisation locale linéaire de la fonction de régression. Plus précisément, nous construisons, par la méthode des k plus proches voisins (k NN), un estimateur pour ce modèle et nous établissons ces propriétés asymptotiques dans le cas où la variable explicative est fonctionnelle.

Notons que la modélisation local linéaire est une approche alternative à la estimation de Nadaraya-Watson (NW) et qui a plus d'avantages sur cette dernière (voir [11]). En particulier, le plus grand avantage de la méthode des polynômes

locaux sur la méthode du noyau est la réduction considérable du biais de l'estimateur. Motivée par cette supériorité sur l'estimation de NW, l'estimation locale linéaire de la régression fonctionnelle a été l'objet des plusieurs travaux en statistique fonctionnelle. Nous citons, par exemple, [3], [4], [7]. Dans la plupart de ces travaux, les auteurs utilisent une pondération issue de la méthode du noyau. Dans cette note, nous construisons notre estimateur à l'aide de méthode des k plus proches voisins. Rappelons que l'importance de l'estimation par la méthode des k plus proches voisins provient essentiellement du fait que le paramètre de lissage est bien adapté aux données. La souplesse et l'efficacité, de la méthode des k plus proches voisins ont motivé plusieurs auteurs d'introduire cette approche en statistique fonctionnelle. Parmi les travaux précurseurs dans cette thématique nous citons [6], [13], [14], [15]. Le but principal de cette note est l'accumulation des avantages de deux approches (local linéaire et la méthode du k NN) . Plus précisément, nous allons combiner les idées de la régression locale linéaire à ceux de la méthode du k NN pour construire un estimateur localement adaptatif avec un biais réduit. Comme c'est indiqué dans [6], l'estimateur de la méthode du k NN est plus pratique que la méthode à noyau, car, le nombre optimal des voisins est à sélectionner sur un ensemble fini. Comme résultat asymptotique, nous établissons la convergence presque complète de notre estimateur. Nous traitons, dans un premier temps, le cas où le nombre des voisins est fixé. Ensuite, nous généralisons ce résultat au cas uniforme sur le nombre des voisins. Nous mentionnons, également, que ce résultat est extrêmement important en pratique où le nombre optimale des voisins est variable en fonction des données. Nous renvoyons à [13] pour plus de discussion sur l'importance de la convergence uniforme sur le nombre des voisins. Citons également les récentes monographies [12], [19] sur l'analyse statistique des données fonctionnelle pour un état de l'art sur le sujet. Nous présentons notre modèle et son estimateur dans le paragraphe suivant. Dans le troisième paragraphe, nous étudions la convergence presque complète de notre estimateur lorsque le nombre des voisins est fixé. La version uniforme de cette convergence est donnée au dernier paragraphe.

2.2 Le Modèle et son estimateur

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 $\mathcal{F}$.

Généralement, l'estimation de la fonction de régression r par la méthode des polynômes locaux est fondée sur une condition de régularité permettant d'approximer localement r par un polynôme. En analyse statistique des données fonctionnelles, plusieurs approches par polynômes locaux ont été proposées, afin de construire des estimateurs de r . Nous procédons par la même approximation

de [3] :

$$r(x_0) = a + b\beta(x_0, x) + o(\beta(x_0, x)), \forall x_0 \text{ dans un voisinage de } x \text{ de } \mathcal{F} \quad 1$$

où $\beta(\cdot, \cdot)$ est une fonction de $\mathcal{F} \times \mathcal{F}$ à valeurs dans $\mathbb{R}$ vérifie $\beta(\xi, \xi) = 0$, pour tout $\xi \in \mathcal{F}$. Les deux paramètres a et b sont à estimer en cherchant la solution du problème de minimisation :

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

où $\delta(\cdot, \cdot)$ est une fonction de $\mathcal{F} \times \mathcal{F}$ à valeurs dans $\mathbb{R}$ vérifie $|\delta(\xi, \xi')| = d((\xi, \xi'))$, K est un noyau et $H_{k,x} = \min \{h \in \mathbb{R}^+ \text{ tel que } \sum_{i=1}^n \mathbb{1}_{B(x,h)}(X_i) = k\}$. La solution de ce problème est explicitement donné par

$$\begin{pmatrix} \hat{a}(x) \\ \hat{b}(x) \end{pmatrix} = \frac{1}{\sum_{i=1}^n \sum_{j=1}^n \beta_i(\beta_i - \beta_j) K_i K_j} \begin{pmatrix} \sum_{i=1}^n \sum_{j=1}^n \beta_i(\beta_i - \beta_j) K_i K_j Y_j \\ \sum_{i=1}^n \sum_{j=1}^n (\beta_j - \beta_i) K_i K_j Y_j \end{pmatrix}$$

où $\beta_i = \beta(X_i, x)$ et $K_i = K(H_{k,x}^{-1} \delta(x, X_i))$.

En utilisant (1) et le fait que $\beta(x, x) = 0$, on estime $r(x)$ par

$$\hat{r}(x) = \hat{a}(x) = \frac{1}{\sum_{i=1}^n \sum_{j=1}^n \beta_i(\beta_i - \beta_j) K_i K_j} \sum_{i=1}^n \sum_{j=1}^n \beta_i(\beta_i - \beta_j) K_i K_j Y_j.$$

Nous rappelons que la littérature sur l'estimation par la méthode de kNN est très abondante dans le cas de la statistique vectorielle (voir [9] pour une liste de références).

2.3 Convergence presque complète ponctuelle

Par la suite, on fixe un point x de $\mathcal{F}$, on note par N_x un voisinage de x et par C, C' des constantes strictement positives. On pose $\phi_x(r_1, r_2) = \mathbb{P}(r_2 \leq \delta(X, x) \leq r_1)$ et on considère les hypothèses suivantes :

(H1) La fonction $\phi_x(r) = \phi_x(-r, r)$, $r > 0$ est continue au voisinage de 0 et vérifie $\phi_x(r) > 0$, $\forall r > 0$.

(H2) $\forall x_1, x_2 \in N_x \mid r(x_1) - r(x_2) \mid \leq C_1 d^\beta(x_1, x_2)$.

(H3) La fonction $\beta(\cdot, \cdot)$ vérifie $\forall x' \in \mathcal{F}$, $C |\delta(x, x')| \leq |\beta(x, x')| \leq C' |\delta(x, x')|$.

(H4) K est un noyau continu, dérivable et à support compact $[-1, 1]$.

(H5) Le nombre de voisins k est tel que $k/n \rightarrow 0$ et $\lim_{n \rightarrow \infty} \log n/k = 0$

(H6) Pour toute suite $h_n \rightarrow 0$ on a :

Il existe un n_0 , tel que $\forall n > n_0$, $-\frac{1}{\phi_x(h_n)} \int_{-1}^1 \phi_x(zh_n, h_n) \frac{d}{dz} (z^2 K(z)) dz > C > 0$,

$$\text{et } h_n \int_{B(x, h_n)} \beta(u, x) dP(u) = o \left(\int_{B(x, h_n)} \beta^2(u, x) dP(u) \right)$$

où $B(x, r) = \{x' \in \mathcal{F} / |\delta(x', x)| \leq r\}$

(H7) Les moments conditionnels de Y vérifient $\forall m \geq 2$, $E[Y^m | X = \cdot] := \sigma_m(\cdot)$ est une fonction continue et bornée

Remarque 1. : Nos hypothèses sont standards, les conditions (H1)-(H2), (H4), (H5) et (H7) ont été utilisées pour l'estimation k NN de la régression classique (voir [6]), tandis que la condition (H3) et (H6) sont similaires à ceux utilisées par [3] pour la méthode locale linéaire.

On obtient le résultat suivant :

Theorem 2.1. *Sous les hypothèses (H1)-(H7), on a*

$$\widehat{r}(x) - r(x) = O \left(\left(\phi_x^{-1} \left(\frac{k}{n} \right) \right)^\beta + \left(\sqrt{\frac{\log n}{k}} \right) \right) \quad p.co. \quad \text{3}$$

Schéma de la démonstration : Il suffit de combiner les techniques de [6] a ceux utilisées par [3]. La démonstration détaillée peut être obtenue sur simple demande.

2.4 Convergence presque complète pour un k fixé

Dans cette section, nous établissons la version uniforme de la vitesse de convergence précédente. Nous rappelons que ce type de la convergence uniforme est très important en pratique. En effet, dans la plupart des cas l'estimateur optimal est associé à un nombre des voisins, k , variable et qui est fonction aux observations $(X_i, Y_i)_i$. Ainsi, la convergence presque complète pour un k fixé est insuffisante pour implémenter cet estimateur en pratique. Notons que ces questions de la convergence uniforme sur le paramètre de lissage ont été largement étudiés dans l'estimation par la méthode à noyau (voir [10], [8]). Cependant, à l'exception de l'article [13] qui traite la régression classique ce problème n'a pas été abordé dans la méthode de k NN. Afin d'établir cette propriété asymptotique, on suppose que $k \in (k_{1,n}, k_{2,n})$ et on introduit les conditions suivantes :

(U1) La condition (H1) est vérifiée pour un $\phi_x(\cdot)$ telle que pour tout $s \in (0, 1)$,

$$\lim_{r \rightarrow 0} \frac{\phi_x(sr)}{\phi_x(r)} = \tau_x(s).$$

(U2) La condition (H2) est vérifiée.

(U3) La condition (H3) est vérifiée

(U4) K est un noyau continu, dérivable de dérivée continue et à support compact $(-1/2, 1/2)$. De plus, la classe de fonctions

$\mathcal{K}_k = \{\cdot \mapsto \gamma^{-k} K(\gamma^{-1} \delta(x, \cdot)) \beta^k(x, \cdot), \gamma > 0\}$ est mesurable ponctuellement

vérifiant :

$$\sup_Q \int_0^1 \sqrt{1 + \log \mathcal{N}(\epsilon \|F\|_{\mathcal{Q},2}, \mathcal{K}_k, d_Q)} d\epsilon < \infty$$

où le supremum est pris sur toutes les mesures de probabilité Q de l'espace $\mathcal{F}$ avec $\|F\|_{\mathcal{Q},2} < \infty$ et où F est la fonction enveloppe de la classe $\mathcal{K}_k$. Ici, d_Q est la métrique $L_2(Q)$ et $\mathcal{N}(\epsilon, \mathcal{K}_k, d_Q)$ est le nombre minimal de boules ouvertes (par rapport à la métrique $L_2(Q)$) de rayon ϵ qui sont nécessaires pour couvrir la classe de fonction $\mathcal{K}_k$.

(U5) Les deux suites $k_{1,n}, k_{2,n}$ sont telles que

$$\phi_x^{-1} \left(\frac{k_{2,n}}{n} \right) \rightarrow 0 \quad \text{and} \quad \frac{\log n}{\min(n \phi_x^{-1} \left(\frac{k_{1,n}}{n} \right), k_{1,n})} \rightarrow 0.$$

(U6) La deuxième partie de (H6) est vérifiée et la première est remplacée par

$$K(1/2) \tau_x(1/2) - \int_{-1/2}^{1/2} K'(s) \tau_x(s) ds > 0$$

$$\text{et} \quad (1/4) K(1/2) \tau_x(1/2) - \int_{-1/2}^{1/2} (s^2 K'(s)) \tau_x(s) ds > 0.$$

(U7) La condition (H7) est vérifiée

Remarque 2. : De même que le premier résultat, les conditions utilisées sont classiques. En effet, toutes les conditions supplémentaires sont similaires à ceux utilisés par [13].

On obtient le résultat suivant :

Theorem 2.2. *Sous les hypothèses (U1)-(U7), on a*

$$\sup_{k_{1,n} \leq k \leq k_{2,n}} |\widehat{r}(x) - r(x)| = O \left(\phi_x^{-1} \left(\frac{k_{2,n}}{n} \right)^\beta \right) + O_{a.co.} \left(\sqrt{\frac{\log n}{k_{1,n}}} \right).$$

Schéma de la démonstration : La preuve est basée sur des idées similaires à ceux utilisées par [13]. La démonstration détaillée peut être obtenue sur simple demande.

- [1] Barrientos-Marin, J., Ferraty, F. and Vieu, P. (2010). Locally Modelled Regression and Functional Data. *J. of Nonparametric Statistics*, **22**, Pages 617–632.
- [2] Baillo, A. and Grané, A. (2009). Local linear regression for functional predictor and scalar response. *J. of Multivariate Analysis*, **100**, Pages 102–111.
- [3] Burba, F., Ferraty, F. and Vieu, P. (2009) k -nearest neighbor method in functional nonparametric regression. *J. Nonparametric Statist.*, **21**, 453-469.
- [4] Demongeot, J., Naceri, A. , Laksaci, A. and Rachdi, M. (2017). Local linear regression modelization when all variables are curves. *Statist. Probab. Lett.* **121**, 37–44.
- [5] Deheuvels, P. and Ouadah, S. (2013). Uniform-in-bandwidth functional limit laws. *J. Theoret. Probab.*, **26**(3), 697-721.
- [6] Devroye, L. Györfi, L. Krzyżak, A. and Lugosi, G. (1994). On the strong universal consistency of nearest neighbor regression function estimates. *Ann. Statist.*, **22**, 1371-1385.
- [7] Einmahl, U. and Mason, D. (2005). Uniform in bandwidth consistency of kernel-type function estimators. *Ann. Statist.*, **33**, 1380-1403.
- [8] Fan, J. and Gijbels, I. (1996). *Local Polynomial Modelling and its Applications*. London, Chapman & Hall.
- [9] Hsing, T. and Eubank, R. (2015). Theoretical foundations of functional data analysis, with an introduction to linear operators. Wiley Series in Probability and Statistics. John Wiley & Sons, Chichester.
- [10] Karaa, L-Z. Laksaci, A, Rachdi, M. and Vieu, P. (2017) Data-driven k NN estimation in nonparametric functional data analysis. *J. Multivariate Anal.* **153**, 176–188.
- [11] Kudraszow, N. and Vieu, P. (2013). Uniform consistency of k NN regressors for functional variables. *Statist. Probab. Lett.*, **83**, 1863-1870.
- [12] Lian, H. (2011). Convergence of functional k -nearest neighbor regression estimate with functional responses. *Electron. J. Statist.*, **5**, 31-40.
- [13] Zhang, J. (2014). *Analysis of variance for functional data*. Monographs on Statistics and Applied Probability, 127. CRC Press, Boca Raton, FL.

Démonstration du théorème 2.1 Il suffit d'appliquer le lemme 3.2 du Burba et al [6] sur les termes S_i , $i = 1, \dots, 4$ de Barrientos et al [3].

Démonstration du théorème 2.2 On pose

$$z_n = O\left(\phi_x^{-1}\left(\frac{k_{2,n}}{n}\right)^\beta + \sqrt{\frac{\log n}{k_{1,n}}}\right)$$

pour $a \in]0, 1[$ on a :

$$\begin{aligned} \sup_{k_{1,n} \leq k \leq k_{2,n}} |\widehat{r}(x) - r(x)| &= \sup_{k_{1,n} \leq k \leq k_{2,n}} |\widehat{r}(x) - r(x)| \mathbb{I}_{\left\{\phi_x^{-1}\left(a \frac{k_{1,n}}{n}\right) \leq H_{k,x} \leq \phi_x^{-1}\left(\frac{k_{2,n}}{an}\right)\right\}} \\ &+ \sup_{k_{1,n} \leq k \leq k_{2,n}} |\widehat{r}(x) - r(x)| \mathbb{I}_{\left\{H_{k,x} \notin \left(\phi_x^{-1}\left(\frac{ak_{1,n}}{n}\right), \phi_x^{-1}\left(\frac{k_{2,n}}{an}\right)\right)\right\}}. \end{aligned}$$

Ainsi, pour tout $\epsilon > 0$, on a :

$$\begin{aligned} &\mathbb{P}\left\{\sup_{k_{1,n} \leq k \leq k_{2,n}} |\widehat{r}(x) - r(x)| \geq \epsilon z_n\right\} \\ &\leq \mathbb{P}\left\{\sup_{k_{1,n} \leq k \leq k_{2,n}} |\widehat{r}(x) - r(x)| \mathbb{I}_{\left\{\phi_x^{-1}\left(\frac{ak_{1,n}}{n}\right) \leq H_{k,x} \leq \phi_x^{-1}\left(\frac{k_{2,n}}{na}\right)\right\}} \geq \frac{\epsilon z_n}{2}\right\} \\ &+ \mathbb{P}\left\{H_{k,x} \notin \left(\phi_x^{-1}\left(\frac{ak_{1,n}}{n}\right), \phi_x^{-1}\left(\frac{k_{2,n}}{na}\right)\right)\right\}. \end{aligned}$$

Donc pour montrer que $\sup_{k_{1,n} \leq k \leq k_{2,n}} |\widehat{r}(x) - r(x)| = O_{a.co.}(z_n)$, il suffit de montrer

$$\sum_n \sum_{k=k_{1,n}}^{k_{2,n}} \mathbb{P}\left(H_{k,x} \leq \phi_x^{-1}\left(\frac{ak_{1,n}}{n}\right)\right) < \infty \quad 5$$

$$\sum_n \sum_{k=k_{1,n}}^{k_{2,n}} \mathbb{P}\left(H_{k,x} \geq \phi_x^{-1}\left(\frac{k_{2,n}}{na}\right)\right) < \infty \quad 6$$

$$\sup_{\phi_x^{-1}\left(\frac{k_{1,n}}{n}\right) \leq h \leq \phi_x^{-1}\left(\frac{k_{2,n}}{n}\right)} |\widehat{r}(x) - r(x)| = O_{a.co.}(z_n) \quad 7$$

où

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

La preuve de de équations (5 et 6) sont donnée dans Kara et al [13], tandis que la démonstration de 7 est dans Attouch et al [1]

- [1] Attouch M., Laksaci A. Raffa F. (2017) On the local linear estimate for functional regression : Uniform in bandwidth consistency. Prprint
- [2] Barrientos-Marin, J., Ferraty, F. and Vieu, P. (2010). Locally Modelled Regression and Functional Data. *J. of Nonparametric Statistics*, **22**, Pages 617–632.
- [3] Burba, F., Ferraty, F. and Vieu, P. (2008) Convergence de l'estimateur à noyau des k plus proches voisins en régression fonctionnelle non-paramétrique. *C. R. Math. Acad. Sci. Paris* 346 (2008), no. 5-6, 339–342
- [4] Karaa, L-Z. Laksaci, A, Rachdi, M. and Vieu, P. (2017) Data-driven k NN estimation in nonparametric functional data analysis. *J. Multivariate Anal.* **153**, 176–188.

Chapitre **3: On the local linear estimate for functional regression : Uniform in bandwidth consistency**

Contents

On the local linear estimate for functional regression : Uniform in bandwidth consistency

Mohammed Attouch, Ali Laksaci and Fatima Rafea

Laboratoire de Statistique et Processus Stochastiques,

Université de Sidi Bel Abbès,

BP 89 Sidi Bel Abbès 22000. Algeria.

attou_kadi@yahoo.fr,

alilak@yahoo.fr,

rafaa_fatima@yahoo.com

Abstract

We consider the problem of local linear estimation of the regression function when the regressor is functional. The main result of this paper is to prove the strong convergence (with rates), uniformly in bandwidth parameters (UIB), of the considered estimator. The main interest of this result is the possibility to derive the asymptotic properties of our estimate even if the bandwidth parameter is a random variable.

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

Key words : Functional data analysis, UIB consistency, Functional local linear regression, Bandwidth parameter, Almost complete consistency.

3.1 Introduction

The regression operator is one of the most used models to explore the relationship between two random variables. In this paper, we deal with the problem of the estimation of the functional regression operator by using the local linear approach. Noting that the functional statistical analysis is an emerging field of mathematical statistics. Such area becomes increasingly popular as evidenced in several special issues dedicated to this topic by several statistical journals (see, for instance, Davidian et al. [7], Gonzalez-Manteiga and Vieu [14] and Goia and Vieu [13]). Concerning the nonparametric estimation of the functional regression, the literature is also very abundant (see, for instance Kudraszowa and Vieu [16] and its references therein). Most of this literature considers the kernel estimation method. However, it is known that the latter produces high bias compared to the local linear method (see Fan and Gijbels [10] for a thorough discussion on the comparison between the two methods). The local linear estimation method has been considered for functional statistics in the last few years. Precisely, the first

result in this topic was obtained by Baïllo and Grané [4]. They obtained the L^2 -convergence rate of the local linear estimate of the regression function when the explanatory variable takes values in a Hilbert space. Barrientos-Marin *et al.* [3] propose an alternative simplified version which can be used, even if the regressors are not Hilbertian. As asymptotic results, the authors establish the almost complete convergence (with rate) of the constructed estimate. Berline *et al.* [4] constructed another local linear estimate based on the inverse of the local covariance operator of the functional explanatory variable. Boj *et al.* [5] treat the local linear Distance- Based Regression for functional regressor. Recently, Zhou and Lin [21] gave the asymptotic normality of the functional local linear regression estimate.

The main purpose of this work is to study the asymptotic properties of the local linear estimate uniformly on the smoothing parameter. Precisely, we consider the functional local linear version proposed by Barrientos-Marin *et al.* [3] and we state its convergence rate of the almost complete consistency uniformly on the bandwidth parameter. We point out that the importance of our results is that they allow to conclude the asymptotic properties of the local linear estimate even if the smoothing parameter is a random variable. Recall that this situation is very usual in practice. Indeed, the bandwidth parameter of the optimal estimate is usually variable in function of the data. Moreover, this contribution can be viewed as a generalization of both works Barrientos-Marin *et al.* [3] and Kara *et al.* [15]. The first work can be obtained by fixing the bandwidth parameter, while the standard kernel estimate (the local constant estimator) is a particular case of the local linear approach (see Barrientos-Marin *et al.* [3]). Finally, let's mention that the present work is the first one that treats the uniform in bandwidths (UIB) asymptotic properties in local polynomial estimation even in the nonfunctional case. All the existing results treat only the standard kernel estimate (see, Deheuvels and Ouadah [8]) for a list of references in the multivariate case and Kara *et al.* [15] for the functional case).

The paper is organized as follows : after establishing notations and listing the required assumptions in Section 2, we state our main results in Section 3. In Section 4, we conduct a simulation study to compare the finite sample performance of our estimator against the kernel one. Finally, Section 5 is devoted to the technical proofs of the results of this paper.

3.2 The local linear estimate for functional regression

Let us introduce n pairs of random variables (X_i, Y_i) for $i = 1, \dots, n$ that we assume drawn from the pair (X, Y) which is valued in $\mathcal{F} \times \mathcal{R}$, where $\mathcal{F}$ is a semi-metric space equipped with a semi-metric d . Furthermore, we assume that the link between X and Y is modeled by the following relation

$$Y = R(X) + \epsilon,$$

where R is an operator from $\mathcal{F}$ to $\mathcal{R}$ and ϵ is a random error variable such that $E[\epsilon|X] = 0$.

The local polynomial smoothing is based on the assumption that is the functional operator is smooth enough to be locally approximated by a polynomial. In functional statistics, there are several ways for extending this approach. Here we adopt the fast version proposed by Barrientos-Marin *et al.* [3] for which the regression function $R(x)$ is estimated by

$$\widehat{R}(x) = \frac{\sum_{i,j=1}^n W_{ij}(x) Y_j}{\sum_{i,j=1}^n W_{ij}(x)}$$

where

$$W_{ij}(x) = \beta(X_i, x) \left(\beta(X_i, x) - \beta(X_j, x) \right) K(h^{-1} \delta(x, X_i)) K(h^{-1} \delta(x, X_j))$$

with $\beta(.,.)$ is a known function from $\mathcal{F}^2$ into $\mathbb{R}$ such that, $\forall \xi \in \mathcal{F}$, $\beta(\xi, \xi) = 0$ and $\delta(.,.)$ is a function of $\mathcal{F} \times \mathcal{F}$ such that $d(.,.) = |\delta(.,.)|$. Furthermore, $h = h_{K,n}$ is a sequence of positive real numbers belonging to an interval (a_n, b_n) with $\lim_{n \rightarrow \infty} a_n = \lim_{n \rightarrow \infty} b_n = 0$.

Our main goal in this contribution is to establish the almost complete consistency of $\widehat{R}(x)$ uniformly in bandwidth h . To do that, we fix a point x in $\mathcal{F}$, N_x neighborhood of x and we consider the following assumptions

(H1) $\forall r > 0$, $\mathbb{P}(r_1 < \delta(X, x) < r_2) := \phi_x(r_1, r_2)$ with $\phi_x(r) = \phi_x(-r, r) > 0$ and

$$\text{For all } s \in (0, 1), \quad \lim_{r \rightarrow 0} \frac{\phi_x(-r/2, sr)}{\phi_x(-r, r)} = \tau_x(s) < \infty.$$

(H2) There exists $C_0 > 0$, $C_1 > 0$, $p \geq 2$ and $b_1 > 0$ such that $E[|Y|^p | X = x] < C_0$ and

$$\forall (x_1, x_2) \in N_x \times N_x; \quad |R(x_1) - R(x_2)| \leq C_1 d^{b_1}(x_1, x_2).$$

(H3) For $k = 0, 1, 2$ the class of functions :

$\mathcal{K}_k = \{\cdot \mapsto \gamma^{-k} K(\gamma^{-1} \delta(x, \cdot)) \beta^k(x, \cdot), \gamma > 0\}$ is a pointwise measurable class ¹

such that :

$$\sup_Q \int_0^1 \sqrt{1 + \log \mathcal{N}(\epsilon \|F\|_{Q,2}, \mathcal{K}_k, d_Q)} d\epsilon < \infty$$

where the supremum is taken over all probability measures Q on the space $\mathcal{F}$ with $Q(F^2) < \infty$ and where F is the envelope function ² of the set $\mathcal{K}_k$. Here, d_Q is the $L_2(Q)$ -metric and $\mathcal{N}(\epsilon, \mathcal{K}_k, d_Q)$ is the minimal number of open balls (with respect to the $L_2(Q)$ -metric) with radius ϵ which are needed to cover the function class $\mathcal{K}_k$. We will denote by $\|\cdot\|_{Q,2}$ the $L_2(Q)$ -norm.

(H4) The kernel K is supported within $(-1/2, 1/2)$ and has a continuous first derivative on $(-1/2, 1/2)$ which is such that :

$$0 < C_2 \mathbb{I}_{(-1/2, 1/2)}(\cdot) \leq K(\cdot) \leq C_3 \mathbb{I}_{(-1/2, 1/2)}(\cdot) \quad (\mathbb{I}_A \text{ is the indicator function of the set } A)$$

$$K(1/2) \mathbb{I}_x(1/2) - \int_{-1/2}^{1/2} K'(s) \mathbb{I}_x(s) ds > 0$$

$$\text{and} \quad (1/4)K(1/2) \mathbb{I}_x(1/2) - \int_{-1/2}^{1/2} (s^2 K'(s)) \mathbb{I}_x(s) ds > 0.$$

(H5) The function $\beta(\cdot, \cdot)$ is such that :

$$\forall x' \in \mathcal{F}, \quad C_4 |\delta(x, x')| \leq |\beta(x, x')| \leq C_5 |\delta(x, x')| \quad C_4, C_5 > 0.$$

(H6) For all $h \in (a_n, b_n)$ we have

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

where $B(x, r) = \{x' \in \mathcal{F} / d(x', x) \leq r\}$ and $dP(x)$ is the cumulative distribution.

(H7) The sequence (a_n) verifies :

$$\frac{\log n}{n \min(a_n, \phi_x(a_n))} \rightarrow 0.$$

Remark

We note that the considered assumptions are classic in this context of nonpa-

1. A class of functions C is said to be a pointwise measurable class if, there exists a countable subclass C_0 such that for any function $g \in C$ there exists a sequence of functions $(g_m)_{m \in \mathbb{N}}$ in C_0 such that : $|g_m(z) - g(z)| = o(1)$.

2. An envelope function G for a class of functions C is any measurable function such that : $\sup_{g \in C} |g(z)| \leq G(z)$, for all z .

rametric functional data analysis. Conditions (H1), (H5) and (H6) are the same as introduced by Barrientos-Marin *et al.* [3] in functional local linear regression. Moreover, Assumptions (H2)-(H4) (H7) are the same as those used by Kara *et al.* [15] for obtaining the UIB consistency of the kernel estimate of the functional regression.

3.3 Main result

The following theorem gives the UIB consistency of $\widehat{R}(x)$.

Theorem 3.3. *Under assumptions (H1)-(H7), we have :*

$$\sup_{a_n \leq h_K \leq b_n} |\widehat{R}(x) - R(x)| = O(b_n^\beta) + O_{a.co.} \left(\sqrt{\frac{\log n}{n\phi_x(a_n)}} \right).$$

The proof of Theorem 3.3 is based on the following decomposition

$$\mathbb{R}(x) - R(x) = \widehat{B}(x) + \frac{\widehat{D}(x)}{\widehat{r}_{1,h}(x)} + \frac{\widehat{Q}(x)}{\widehat{r}_{1,h}(x)}$$

where

$$\begin{aligned} \widehat{Q}(x) &:= (\widehat{r}_{2,h}(x) - \mathbb{E}[\widehat{r}_{2,h}(x)]) - R(x)(\widehat{r}_{1,h}(x) - \mathbb{E}[\widehat{r}_{1,h}(x)]) \\ \widehat{B}(x) &:= \frac{\mathbb{E}[\widehat{r}_{2,h}(x)]}{\mathbb{E}[\widehat{r}_{1,h}(x)]} - R(x) \quad \text{and} \quad \widehat{D}(x) := -\widehat{B}(x)(\widehat{r}_{1,h}(x) - \mathbb{E}[\widehat{r}_{1,h}(x)]) \end{aligned}$$

with

$$\widehat{r}_{2,h}(x) = \frac{1}{n(n-1)h^2\phi_x^2(h)} \sum_{i \neq j} W_{ij}(x) Y_j \quad \text{and} \quad \widehat{r}_{1,h}(x) = \frac{1}{n(n-1)h^2\phi_x^2(h)} \sum_{i \neq j} W_{ij}(x).$$

Then, Theorem 3.3 is a direct consequence of the following Lemmas.

Lemma 3.1. *Under the hypotheses (H1), (H3)-(H7), we have that :*

$$\sup_{a_n \leq h \leq b_n} |\widehat{r}_{1,h}(x) - \mathbb{E}[\widehat{r}_{1,h}(x)]| = O_{a.co.} \left(\sqrt{\frac{\log n}{n\phi_x(a_n)}} \right).$$

Corollary 3.1. *Under the hypotheses of Lemma 0.9, there exists a real number $C_6 > 0$ such that :*

$$\sum_{n=1}^{\infty} \mathbb{P} \left(\inf_{a_n \leq h \leq b_n} \widehat{r}_{1,h}(x) < C_6 \right) < \infty.$$

Lemma 3.2. *Under the hypotheses (H1)-(H6) we have that :*

$$\sup_{a_n \leq h \leq b_n} |\widehat{B}(x)| = O(b_n^\beta).$$

Lemma 3.3. *Under the assumptions of Theorem 3.3, we have that :*

$$\sup_{a_n \leq h \leq b_n} |\widehat{g}(x) - E[\widehat{g}(x)]| = O_{a.co.} \left(\sqrt{\frac{\log n}{n\phi_x(a_n)}} \right).$$

3.4 Discussions and Applications

3.4.1 Discussions

— *On the functional local linear approach*

As mentioned in the introduction the local polynomial smoothing has various advantages over the kernel method. In particular the local linear approach has a small bias compared to the standard kernel method. Such an advantage is also produced in the UIB consistency. Indeed, it is well-known that the main idea in the local linear smoothing is based on the fact that the regression operator admits a linear approximation in the neighborhood of the conditioning point. Notice that, in the non-functional case, this linear approximation is due to a Taylor expansion of $R(\cdot)$ but, in the functional setup, this can be expressed as follows,

$$\text{for any } z \text{ in } N_x, \quad R(z) = R(x) + b\beta(x, z) + O(\beta^2(x, z)). \quad \mathbf{1}$$

In other words, if we replace the assumption (H2) by the decomposition (1) and we use the fact that $\sum_{i,j=1}^n \beta(X_j, x) W_{ij} = 0$, then we reformulate the bias term of the estimator $\mathbb{R}(x)$ as follows :

$$\begin{aligned} E[\mathbb{R}(x)] - R(x) &= E \left[\frac{\sum_{i,j=1}^n W_{ij}(x) Y_j - R(X_j)}{\sum_{i,j=1}^n W_{ij}(x)} \right] \\ &\quad + E \left[\frac{\sum_{i,j=1}^n W_{ij}(x) (R(X_j) - R(x) + b\beta(X_j, x))}{\sum_{i,j=1}^n W_{ij}(x)} \right]. \end{aligned}$$

So,

$$E[\mathbb{R}(x)] - R(x) = O \left(E \left[\frac{\sum_{i,j=1}^n W_{ij}(x) \beta^2(X_j, x)}{\sum_{i,j=1}^n W_{ij}(x)} \right] \right).$$

By the same arguments as those used for proving Theorem 3.3 we show that

$$\sup_{a_n \leq h \leq b_n} E[\mathbb{R}(x)] - R(x) = O(b_n^2).$$

This bias term is significantly better than in the standard kernel case, studied by Ferraty et al. [12], which under a similar assumption, have obtained

a bias term of order Ch which is $\leq Cb_n$.

– *On the impact of the UIB consistency*

As all smoothing estimation method the choice of the smoothing parameters has a crucial role in the performance of the local linear estimate. Moreover, in the local linear fitting, the selection of the smoothing parameters admits an additional role. Indeed, as pointed-out by Fan and Gijbels [10], the bandwidth parameter contributes to determine the complexity of the model by defining the local neighborhood for which the linear approximation of the model is valid. In practice the cross-validation method is the most popular selector technique to determine the optimal bandwidth parameter. Typically, the optimal estimator $\widehat{R}_{opt}$ is obtained by choosing h that minimizes the following least square prediction criterion :

$$\widetilde{h}_{CV} = \arg \min_{h \in (a_n, b_n)} CV(h),$$

where

$$CV(h) = \sum_{k=1}^n (Y_k - \mathbb{R}^{-k}(X_k))^2 \quad 2$$

with

$$\widehat{R}^{-k}(x) = \frac{\sum_{i \neq k}^n \sum_{j \neq k}^n W_{ij} Y_j}{\sum_{i \neq k}^n \sum_{j \neq k}^n W_{ij}}.$$

However, this usual rule is used in practice without mathematical justification, in the sense that, there is no result (to the best of knowledge) concerning the convergence of $\widehat{R}_{opt}$ to R . All the existing results it concerns $\widehat{R}$ for fixed h , whereas, here $\widetilde{h}_{CV}$ is function of the observations $(X_1, Y_1), \dots, (X_n, Y_n)$. Furthermore, the main feature of the UIB consistency studied in the present contribution is that we can conclude the convergence $\widehat{R}_{opt}$ to R . Specifically, we have the following result.

Corollary 3.2. *If $\widetilde{h}_{CV} \in (a_n, b_n)$ and if the conditions of Theorem 3.3 are hold, we have*

$$|\widehat{R}_{opt}(x) - R(x)| = O(b_n^\beta) + O_{a.co.} \left(\sqrt{\frac{\log n}{n\phi(a_n)}} \right).$$

In conclusion, we can say that our result is interesting for practical purposes, because, it gives a mathematical justification of the popular cross-validation selection rule. Finally, let's note that, despite the practical importance of this question of the smoothing parameter choice, few studies have been published on this topic in functional sadistic. We cite for instance Rachdi and Vieu [17] or Benhenni et al [3] for the functional cross-validation or Chaghy [6] and Shang [19] for some alternative approaches.

3.4.2 Simulation results

The main purpose of this Section is to examine the behaviour of the local functional linear approach on finite samples. Precisely, we show, in the first step, how we can implement the cross-validation data-driven of the previous section in practice. Secondly, in order to emphasize the superiority of the local linear approach over the classical kernel method, we compare the bias behavior of both methods. This numerical study is carried out using two nonparametric functional models (homoscedastic and heteroscedastic model) with two kinds of functional covariates (smooth and discontinuous curves). Formally, we generate our observations $(X_i, Y_i)_{i=1, \dots, n}$ from the following functional nonparametric model :

$$Y_i = r(X_i) + \sigma(X_i)\varepsilon_i, \quad i = 1, \dots, n = 200,$$

where the ε_i 's are generated independently according to a $\mathcal{N}(0, 0.5)$ distribution. The operator $\sigma \equiv 1$ for the homoscedastic case, while, for the heteroscedastic case, we take

$$\sigma(X) = \int |X(t)| \log(|X(t)|) dt.$$

The sampled functional explanatory variables $X_i, i = 1, \dots, 200$, are generated according the following expression

$$X_i(t) = a_i \sin(4(b_i - t)) + b_i + A\eta_i, \quad \forall t \in [0, \pi] \quad \text{and } i = 1, 2, \dots, 200$$

where a_i (resp. b_i and η_i) is $(-3, 0.5)$ (resp is $(0, 3)$ and $(0, 0.5)$). The smooth case (resp. the discontinuous case) is obtained by taking $A = 0$ (resp. $A = 1$). All the curves X_i 's are discretized on the same grid generated from 100 equispaced measurements in $[0, \pi]$. The curves for both cases are displayed on the following figures :

Concerning the regression operator r , we use the following functional :

$$r(x) = \int_0^\pi \frac{dt}{(x(t))^2}.$$

In order to highlight the importance of the data-driven (2) as well as the superiority of our estimate to the kernel method we compare the curve of

$$MSE_1(h) := \frac{1}{n} \sum_{i=1}^n (Y_i - \widehat{Y}_i^{(1,h)})^2 \quad \text{where} \quad \widehat{Y}_i^{(1,h)} = \mathbb{R}(X_i).$$

to both : (1)- $MSE_1(\widetilde{h}_{CV})$ = MSE of the local linear estimate corresponds to the opti-

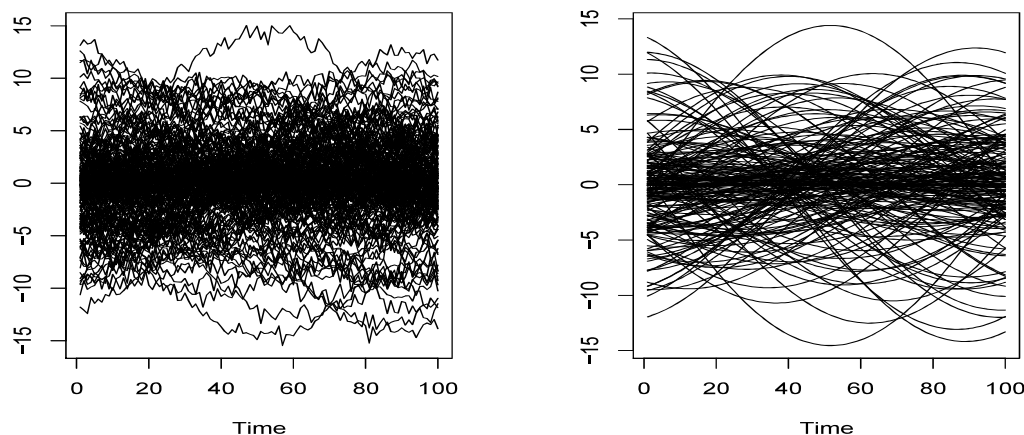


FIGURE 3.1: The functional explanatory variables

mal smoothing parameter and

(2)- MSE of kernel estimate defined by

$$MSE_2(h) := \frac{1}{n} \sum_{i=1}^n (Y_i - \widehat{Y}_i^{(2,h)})^2$$

where

$$\widehat{Y}_i^{(2,h)} = \widetilde{R}(X_i) = \frac{\sum_{j \neq i} K(h^{-1}d(x, X_j))Y_j}{\sum_{j \neq i} K(h^{-1}d(x, X_j))}.$$

These errors are evaluated over a subset

$\mathcal{H} = \{\text{sequence of the 20 quantiles of the distance vector between the functional variable}\}.$

For practical purposes, the estimate $\widehat{Y}_i^{(2,h)}$ is obtained by the R-routine named *funopare.kernel.cv* (available at the website : “www.lsp.ups-tlse.fr/staph/npfda”). While the estimate $\widehat{Y}_i^{(1,h)}$ is constructed by the same procedure proposed by Barrientos-Marin *et al.* [3] for which the locating functions δ and β are defined by

$$\delta(X, X') = d(X, X'), \quad \beta(X, X') = \langle \vartheta, X - X' \rangle$$

where the functional ϑ is selected over the set of eigenvalues of the empirical covariance operator :

$$\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^t (X_i - \bar{X}_n) \text{ where } \bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i.$$

Of course, both functions β and δ can be expressed in function of the semi-metric d . Specifically, the practical computation of d is closely linked to the smoothing

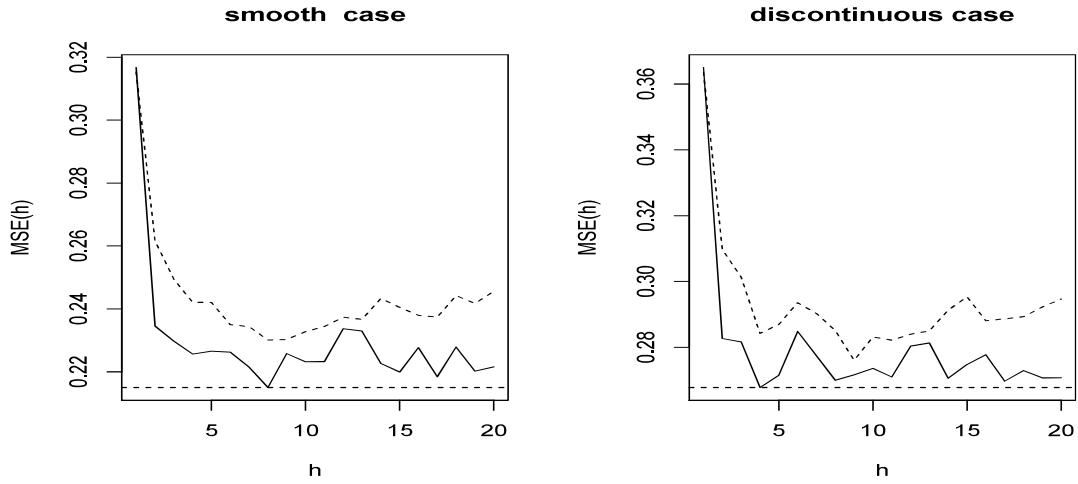


FIGURE 3.2: MSE-results of the homoscedastic case, $MSE_2(h)$: dash curve and $MSE_1(h)$: continuous curve

	Homoscedastic case	Heteroscedastic case
Smooth case	0.21	1.26
Discontinuous case	0.25	2.14

TABLE 3.1: $MSE_1(\tilde{h}_{CV})$

degree of the functional variables $(X_i)_i$. In particular, for the smooth case we use the semi-metric defined by the L_2 -distance between the 2^{th} derivatives of the curves. While of the discontinuous case, we consider a semi-metric based on the $q = 3$ first eigenfunctions of the empirical covariance operator associated with the q greatest eigenvalues (cf. Ferraty and Vieu (2006) for more discussions).

Finally, we point out that, we have simulated by the quadratic kernel K for both estimates. The efficiency of our approach is illustrated in the following figures where we draw, for the various situations, the $MSE_1()$ and $MSE_2()$ functions. In each plot the horizontal line shows the $MSE(\tilde{h}_{CV})$.

It clear that the data-driven (2) works perfectly in practice. In the sense that the optimal estimate associated to $\tilde{h}_{CV}$ reduces substantially the MSE error. However, the efficiency of this procedure is strongly related to the nature of the data. In the following table we summarize the MSE errors of the optimal estimate for the four cases.

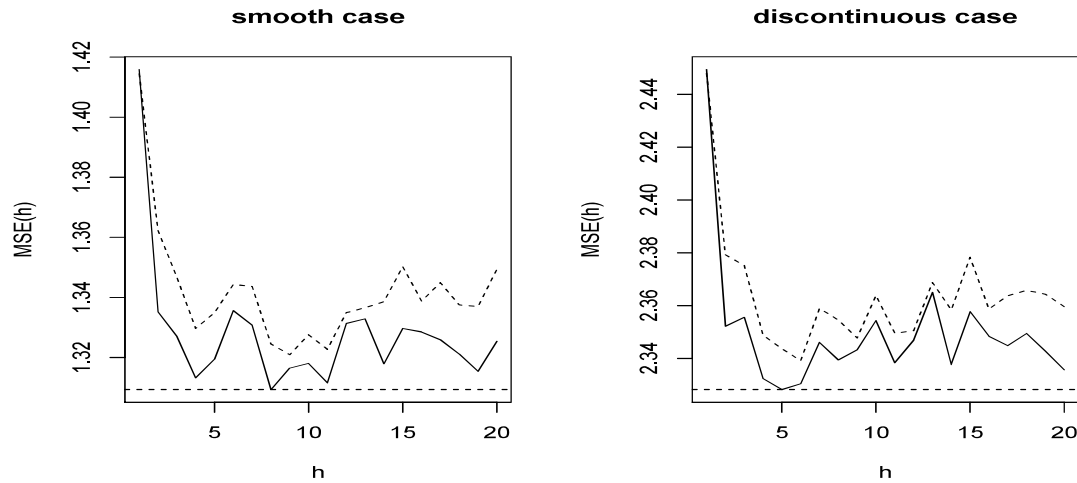


FIGURE 3.3: MSE-results of the heteroscedastic case, $MSE_2(h)$: dash curve and $MSE_1(h)$: continuous curve

On the other hand, it appears clearly that the local linear method (L.L.) gives better results than the standard kernel method (local constant method (L.C.)). Furthermore, this superiority increases in function of h , which allows to conclude that this gain in efficiency is essentially due to the bias part of the MSE-error. To confirm this statement, we run 100 replicates of the simulation process and we compare the two estimators to the true operator r at $x = X_0$ for the simple case (smooth-homoscedastic case) by plotting the box-plot of the estimated $\mathbb{R}(X_0)$ and $\tilde{\mathbb{R}}(X_0)$ jointly with the $r(X_0)$ (dashed horizontal line).

We observe that, there is an important gain in the bias term in the local linear method compared to the kernel method.

3.4.3 Application on a real data

The main goal of this short application example is the analyze of the chemical composition of the soil by using the functional local linear regression. Precisely, we use this model to predict the ergosterol content in soil given the spectrometric curve. It is well known that the ergosterol is an important indicator of soil fungal biomass. In this contribution, we consider the dataset studied by Rinnan [18] which contains a sample soil of size 108 observations obtained at subarctic fell heath in Abisko, northern Sweden. The following figure plots the spectrometric curve X defined by the absorbance versus wavelength 400-2500nm

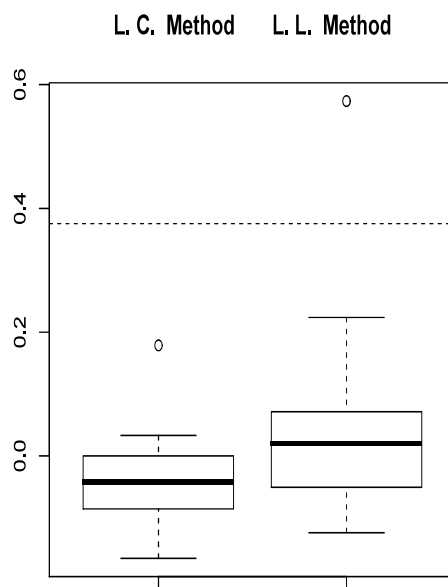
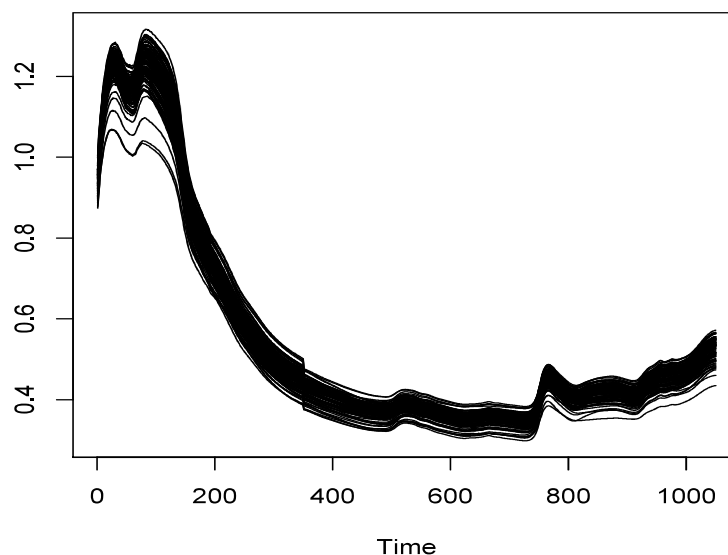


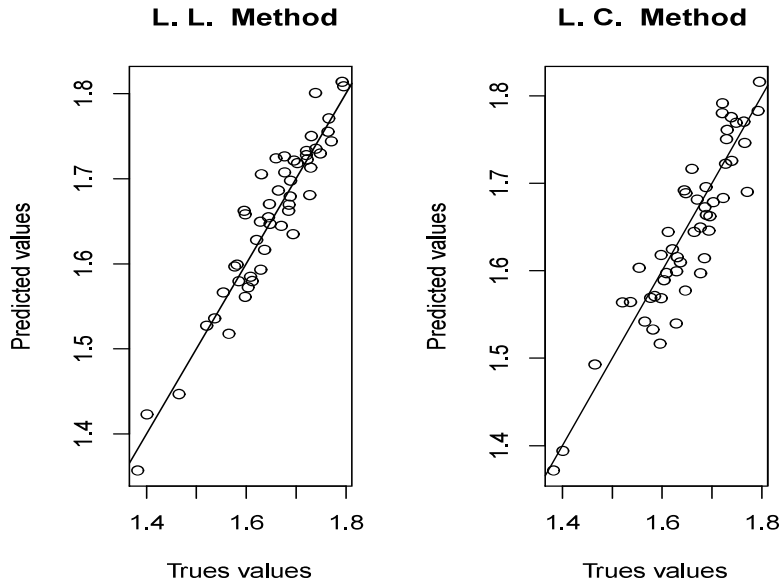
FIGURE 3.4: The estimated regression operator VS the true operator



The spectrometric curves,

This dataset contains 72 samples from the 0 to 5 cm soil depth and 36 samples from the 5 to 10 cm depth. This measurement in different depths induces a several outliers in the data. So, to circumvent the effect of these outliers we take as response variable Y the logarithm of the ergosterol content in soil. For this prediction

problem, we compare the functional local linear model $\mathbb{R}(\cdot)$ to the local constant model $\widetilde{R}(\cdot)$. These functional predictors are computed by the same manner as in the previous section. Now, for our comparison study, we randomly split our data into two subsets. The learning sample, of size 90, will be our statistical sample from which the estimators are calculated. Then, with this sub-sample, we compute the predicted values corresponding to the 18 remaining curves (considered as the test sample). In the following figure we plot the predicted values versus the true values



Comparison results

Once again, it appears clear that, the local linear model outperforms better than the classical regression. This fact is confirmed by the scatterplots and the MSE-error 0.67 for the local linear regression versus 1.29 for the kernel method.

3.5 Appendix

In what follows, when no confusion is possible, we will denote by C and C' some strictly positive generic constants. Moreover, we put, for any $x \in \mathcal{F}$, and for all $i = 1, \dots, n$:

$$K_i = K(h^{-1} \delta(x, X_i)), \beta_i = \beta(X_i, x) \text{ and } \delta_i = \delta(x, X_i).$$

First of all we state the following lemmas which are needed to establish our asymptotic results

Lemma 3.4. (cf. Theorem 2.14.1 in Van der Vaart and Wellner [20], page 239).

Let $Z_1, Z_2, \dots, Z_n$ be independent and identically distributed taking values in a mea-

measurable space $(\mathcal{E}, \Upsilon)$ and consider C a pointwise measurable class of functions $g : \mathcal{E} \rightarrow \mathbb{R}$ with envelope function F , then :

$$\|a_n(g)\|_C \leq CJ(1, C)\|F\|_{p \vee 2}$$

where

$$a_n(g) = \frac{1}{\sqrt{n}} \sum_{i=1}^n (g(Z_i) - \mathbf{E}g(Z_i)) \text{ and } J(1, C) = \sup_{\mathcal{G}} \int_0^1 \sqrt{1 + \log N(\epsilon \|F\|_{\mathcal{G}, 2}, C, d_{\mathcal{G}})} d\epsilon$$

with $\|\cdot\|_p = \mathbf{E}^{1/p}[\cdot]^p$, $\|a_n(g)\|_C = \sup_{g \in C} |a_n(g)|$ and $s \vee t$ denoting the maximum of s and t .

Lemma 3.5. (see Dony and Einmahl [9]).

Let $Z_1, Z_2, \dots, Z_n$ be independent and identically distributed taking values in a measurable space $(\mathcal{E}, \Upsilon)$ and consider C a pointwise measurable class of functions $g : \mathcal{E} \rightarrow \mathbb{R}$ satisfying :

$$\mathbf{E}\|a_n(g)\|_C \leq C\|F\|_2$$

with F is an envelope function of C . Then, for any $A \in \Upsilon$, we have :

$$\mathbf{E}\|a_n(g, \mathbb{1}_A)\|_C \leq 2C\|F, \mathbb{1}_A\|_2.$$

Lemma 3.6. (see Dony and Einmahl [9]) Let $Z_1, Z_2, \dots, Z_n$ be independent and identically distributed taking values in a measurable space $(\mathcal{E}, \Upsilon)$ and consider C a pointwise measurable class of functions $g : \mathcal{E} \rightarrow \mathbb{R}$ with envelope function F . Assume that for some $H > 0$:

$$\mathbf{E}[F^p(Z)] \leq \frac{p!}{2} \sigma^2 H^{p-2} \text{ where } \sigma^2 \geq \mathbf{E}[F^2(Z)].$$

Then, for $\beta_n = \mathbf{E}[\|\sqrt{n}a_n(g)\|_C]$, we have for any $t > 0$:

$$\mathbb{P}\left\{\max_{1 \leq k \leq n} \|a_n(g)\|_C \geq \beta_n + t\right\} \leq \exp\left(-\frac{t^2}{2n\sigma^2 + 2tH}\right).$$

Proof of Lemma 0.9. We use the same arguments as in Barrientos-Marin *et al.* [3] based on the following decomposition

$$\widehat{r}_{1,h}(x) = \underbrace{\frac{n^2 h^2 \phi_x^2(h)}{n(n-1)h^2 \phi_x^2(h)}}_{T_1} \left| \underbrace{\left(\frac{1}{n} \sum_{j=1}^n \frac{K_j}{\phi_x(h)}\right)}_{T_2} \underbrace{\left(\frac{1}{n} \sum_i \frac{K_i \beta_i^2}{h^2 \phi_x(h)}\right)}_{T_4} \right|$$

$$- \underbrace{\left(\frac{1}{n} \sum_{j=1}^n \frac{K_j \beta_j}{h \phi_x(h)} \right)^2}_{T_3} \Bigg]$$

The, the claimed result is a consequence of the following assertions

$$\sum_n \mathbb{P} \left\{ \sup_{a_n \leq h_K \leq b_0} \sqrt{\frac{n \phi_x(a_n)}{\log n}} |T_l - \mathbb{E}[T_l]| \geq \eta_0 \right\} < \infty, \quad (3)$$

for some positive number b_0 and for $l = 2, 3, 4$.

$$\mathbb{E}[T_l] = O(1) \text{ for } l = 2, 3, 4. \quad (4)$$

$$\text{Cov}(T_2, T_4) = o \left(\left(\frac{\log n}{n \phi_x(a_n)} \right)^{1/2} \right) \quad (5)$$

$$\text{and } \text{Var}[T_3] = o \left(\left(\frac{\log n}{n \phi_x(a_n)} \right)^{1/2} \right). \quad (6)$$

Note first that (4) was proved by Barrientos-Marin *et al.* [3]. On the other side, under (H1),(H4) and (H5) we have

$$\mathbb{E}[h^{-k} K_i^l \beta_i^k] \leq \phi_x(h/2).$$

Now, to prove (5) and (6), it suffices to write that

$$\text{Cov}(T_2, T_4) = \frac{1}{n^2 \phi_x^2(h)} \sum_i \text{Cov} \left(K_i, \frac{K_i \beta_i^2}{h^2} \right)$$

and

$$\text{Var}[T_3] = \frac{1}{n^2 \phi_x^2(h)} \sum_i \text{Var} \left[K_i \frac{\beta_i}{h} \right].$$

It follows that

$$\text{Cov}(T_2, T_4) = O \left(\frac{1}{n \phi_x(h)} \right) = o \left(\sqrt{\frac{\log n}{n \phi_x(a_n)}} \right)$$

and

$$\text{Var}[T_3] = O \left(\frac{1}{n \phi_x(h)} \right) = o \left(\sqrt{\frac{\log n}{n \phi_x(a_n)}} \right).$$

Concerning (3). We Define

$$\Delta_i^k = \frac{1}{h^k} \left(K_i \beta_i^k - \mathbb{E} \left[K_i \beta_i^k \right] \right) \text{ for } k = 0, 1, 2$$

and we put

$$h_j = 2^j a_n \text{ and } L(n) = \max\{j : h_j \leq 2b_0\}.$$

Therefrom, we can write

$$\sup_{a_n \leq h \leq b_0} \sqrt{\frac{n\phi_x(a_n)}{\log n}} |T_l - \mathbb{E}[T_l]| \leq \max_{1 \leq j \leq L(n)} \sup_{h_{j-1} \leq h \leq h_j} \sqrt{\frac{n\phi_x(h)}{\log n}} |T_l - \mathbb{E}[T_l]|.$$

Now, we have, for $l = 2, 3, 4$

$$T_l - \mathbb{E}[T_l] := \frac{1}{\sqrt{n}\phi_x(h)} a_n^l(K),$$

where $a_n^l(K) = \frac{1}{\sqrt{n}} \sum_{i=1}^n (\Delta_i^{4-l} - \mathbb{E}[\Delta_i^{4-l}])$ corresponds to the empirical process based on the variables $X_1, X_2, \dots, X_n$. Next, we consider the following class of functions :

$$\mathcal{G}_j^l = \left\{ z \mapsto \gamma^{4-l} K(\gamma^{-1} \delta(x, z)) \beta^{4-l}(x, z) \text{ where } h_{j-1} \leq \gamma \leq h_j \right\}.$$

Therefore

$$\begin{aligned} & \mathbb{P} \left\{ \sup_{a_n \leq h \leq b_0} \sqrt{\frac{n\phi_x(a_n)}{\log n}} |T_l - \mathbb{E}[T_l]| \geq \eta_0 \right\} \\ & \leq L(n) \max_{j=1, \dots, L(n)} \mathbb{P} \left\{ \max_{1 \leq k \leq n} \|\sqrt{k} a_k^l(K)\|_{\mathcal{G}_j^l} \geq \eta_0 \sqrt{n\phi_x(h_j/2) \log n} \right\}. \end{aligned}$$

Furthermore, using the Bernstein's inequality on the a pointwise measurable class of functions to evaluate

$$\mathbb{P} \left\{ \max_{1 \leq k \leq n} \|\sqrt{k} a_k^l(K)\|_{\mathcal{G}_j^l} \geq \eta_0 \sqrt{n\phi_x(h_j/2) \log n} \right\}.$$

For this aim, we must study the asymptotic behavior of :

$$\beta_n = \mathbb{E} \left[\|\sqrt{n} a_n^l(K)\|_{\mathcal{G}_j^l} \right] \quad \text{and} \quad \sigma^2 = \mathbb{E} \left[G_j^2(X) \right]$$

where G_j is the envelope function of the class $\mathcal{G}_j^l$. It clear that under (H4) and (H5) the envelope function G_j satisfies that :

$$G_j(z) \leq C \mathbb{I}_B(x, h_j/2)(z).$$

Hence, for all $p \geq 1$, we have, from (H1) that :

$$\mathbb{E} \left[G_j^p(X) \right] \leq C^p \phi_x(h_j/2)$$

and particularly :

$$\sigma^2 = O(\phi_x(h_j/2)).$$

Concerning the term β_n , we combine (H3) together with the Lemma 3.4's result to get that :

$$E[\|a_n^l(g)\|_{\mathcal{G}_j^l}] \leq CJ(1, \mathcal{K}_{4-l})\|F\|_2$$

where F is the envelope function of $\mathcal{K}_{4-l}$. Thus, Lemma 3.5's conditions are verified for the class $\mathcal{G}_j^l$ and the envelope function F . It follows that :

$$E\|a_n^l(g, \mathbb{1}_{B(x, h_j/2)})\|_{\mathcal{G}_j^l} \leq 2CJ(1, \mathcal{K}_{4-l})\|F, \mathbb{1}_{B(x, h_j/2)}\|_2.$$

Finally, we obtain that :

$$\mathbb{E} \left[\|\sqrt{n}a_n^l(K)\|_{\mathcal{G}_j^l} \right] \leq C\sqrt{n\phi_x(h_j/2)}.$$

We are now in position to apply the Bernstein's inequality for :

$$\beta_n = O\left(\sqrt{n\phi_x(h_j/2)}\right), \sigma^2 = O\left(\phi_x(h_j/2)\right), D = C \text{ and } t = \eta_0/2 \sqrt{n\phi_x(h_j/2) \log n}.$$

Therefrom, we obtain that :

$$\begin{aligned} \mathbb{P} \left\{ \max_{1 \leq k \leq n} \|\sqrt{k}a_k^l(K)\|_{\mathcal{G}_j^l} > \eta_0 \sqrt{n\phi_x(h_j/2) \log n} \right\} &\leq \mathbb{P} \left\{ \max_{1 \leq k \leq n} \|\sqrt{k}a_k^l(K)\|_{\mathcal{G}_j^l} > \beta_n + t \right\} \\ &\leq \exp \left\{ -\eta_0^2 \frac{\log n}{8 + C\sqrt{\frac{\log n}{n\phi_x(h_j/2)}}} \right\} \\ &\leq n^{-C'\eta_0^2}. \end{aligned}$$

The last inequality is a consequence of (H7). Moreover, since $L(n) \leq 2 \log n$, we get that :

$$L(n) \max_{j=1, \dots, L(n)} \mathbb{P} \left\{ \max_{1 \leq k \leq n} \|\sqrt{k}a_k^l(K)\|_{\mathcal{G}_j^l} > \eta_0 \sqrt{n\phi_x(h_j/2) \log n} \right\} \leq C(\log n)n^{-C'\eta_0^2}.$$

Choosing now η_0 such that $C'\eta_0^2 > 1$ permits to get :

$$\sup_{a_n \leq h \leq b_n} |T_l - \mathbb{E}[T_l]| = O_{a.co.} \left(\sqrt{\frac{\log n}{n\phi(a_n)}} \right).$$

Which permits to achieve the proof of this Lemma

Proof of Corollary 3.1. By an analytical arguments we obtain, under (H1) and

(H4)

$$\frac{E[\delta_1^2 K_1]}{h^2 \phi_x(h)} \longrightarrow (1/4)K(1/2)\tau_x(1/2) - \int_{-1/2}^{1/2} (s^2 K(s))' \tau_x(s) ds > 0$$

and

$$\frac{E[K_1]}{\phi_x(h)} \longrightarrow K(1/2)\tau_x(1/2) - \int_{-1/2}^{1/2} (K(s))' \tau_x(s) ds > 0.$$

Moreover, (H4) and (H5) imply that

$$E[\beta_1 K_1] = o(h\phi_x(h/2)) \text{ and } E[\beta_1 K_1] \geq CE[\delta_1^2 K_1].$$

So, there exists a constant $C' > 0$, such that :

$$E[\widehat{r}_{1,h}(x)] \geq C' \text{ for all } h \in (a_n, b_n).$$

Therefor

$$\inf_{h \in (a_n, b_n)} \widehat{r}_{1,h}(x) \leq \frac{C'}{2} \quad \text{implies that there exists } h \in (a_n, b_n) \text{ such that } |E[\widehat{r}_{1,h}(x)] - \widehat{r}_{1,h}(x)| \geq \frac{C'}{2}$$

which allows to write $\sup_{h \in (a_n, b_n)} |E[\widehat{r}_{1,h}(x)] - \widehat{r}_{1,h}(x)| \geq \frac{C'}{2}.$

Thus the result of Lemma 0.9 allows to write that for a $C = \frac{C'}{2}$:

$$\sum_n \mathbb{P}\left(\inf_{h \in (a_n, b_n)} \widehat{r}_{1,h}(x) \leq C\right) \leq \sum_n \mathbb{P}\left(\sup_{h \in (a_n, b_n)} |E[\widehat{r}_{1,h}(x)] - \widehat{r}_{1,h}(x)| \geq C\right) < \infty.$$

Proof of Lemma 0.10. We write

$$\begin{aligned} \widehat{B}(x) &= \frac{E[\widehat{r}_{2,h}(x)] - R(x)E[\widehat{r}_{1,h}(x)]}{E[\widehat{r}_{1,h}(x)]} = \frac{1}{h^2 \phi_x^2(h)E[\widehat{r}_{1,h}(x)]} E[W_{12}(x) [E(Y_2|X_2) - R(x)]] \\ &= \frac{1}{h^2 \phi_x^2(h)E[\widehat{r}_{1,h}(x)]} E[W_{12}(x) [R(X_2) - R(x)]] . \end{aligned}$$

From conditions (H2) and (H4) we have :

$$\mathbb{I}_{B(x,h/2)} |m(X) - m(x)| \leq C(h/2)^b.$$

Consequently

$$\sup_{a_n \leq h \leq b_n} |\widehat{B}(x)| \leq Cb_n^b.$$

Proof of Lemma 0.11. The proof of this Lemma is very similar to Lemma 0.9. Then the same notations of Lemma 0.9 are used. Specifically we use the following

decomposition

$$\widehat{g}(x) = \underbrace{\frac{n^2 h^2 \phi_x^2(h)}{n(n-1)h^2 \phi_x^2(h)}}_{T_1} \left[\underbrace{\left(\frac{1}{n} \sum_{j=1}^n \frac{K_j Y_j}{\phi_x(h)} \right)}_{S_1} \underbrace{\left(\frac{1}{n} \sum_{i=1}^N \frac{K_i \beta_i^2}{h^2 \phi_x(h)} \right)}_{T_4} - \underbrace{\left(\frac{1}{n} \sum_{j=1}^n \frac{K_j \beta_j Y_j}{h \phi_x(h)} \right)}_{S_2} \underbrace{\left(\frac{1}{n} \sum_{i=1}^n \frac{K_i \beta_i}{h \phi_x(h)} \right)}_{T_3} \right].$$

So, the claimed result will be obtained as soon as the three following ones have been checked

$$\sum_n \mathbb{P} \left\{ \sup_{a_n \leq h \leq b_0} \sqrt{\frac{n \phi_x(a_n)}{\log n}} |S_m - \mathbb{E}[S_m]| > \eta \right\} < \infty, \text{ for } m = 1, 2, \text{ for a certain } b_0.$$

7

$$\text{Cov}(S_1, T_4) = o \left(\sqrt{\frac{\log n}{n \phi_x(a_n)}} \right)$$

8

$$\text{and } \text{Cov}(S_2, T_3) = o \left(\sqrt{\frac{\log n}{n \phi_x(a_n)}} \right).$$

9

Firstly, the required result (7) can be treated exactly as (3). Indeed, we put

$$\Lambda_i^m = h^{1-m} K \beta_i^{m-1} Y_i - h^{1-m} \mathbb{E} \left[K \beta_i^{m-1} Y_i \right] \text{ for } k = 1, 2$$

and we consider the empirical process :

$$a_n^{'m}(K) = \frac{1}{\sqrt{n}} \sum_{i=1}^n (\Lambda_i^m - \mathbb{E}[\Lambda_i^m]).$$

We define the following class of functions :

$$\mathcal{G}_j^{'m} = \left\{ (z, y) \mapsto \gamma^{1-m} y K \left(\gamma^{-1} \delta(x, z) \beta^{m-1}(x, z) \right) \text{ for } h_{j-1} \leq \gamma \leq h_j \right\}.$$

It is clear that this class of functions admits an envelope function $F_j(\cdot, \cdot)$ such that :

$$F_j(z, y) \leq C y \mathbb{I}_B(x, h_j/2)(z)$$

From (H1), (H2) and (H4) we obtain that :

$$E[F_j^p(X, Y)] \leq C^p \phi_x(h_j/2) \quad \text{and} \quad E[F^2(X, Y)] = O(\phi_x(h_j/2)).$$

Next, similar ideas as those used in Lemma 0.9's proof allow to get :

$$\beta'_n = E\left[\|\sqrt{n}a'_n(K)\|_{\mathcal{G}'_j}\right] = O\left(\sqrt{n\phi_x(h_j/2)}\right).$$

We apply again the Bernstein's inequality to the empirical process $a'_n(K)$ to obtain :

$$\begin{aligned} \mathbb{P}\left\{\max_{1 \leq k \leq n} \|\sqrt{k}a'_k(K)\|_{\mathcal{G}'_j} > \eta'_0 \sqrt{n\phi_x(h_j/2) \log n}\right\} &\leq \mathbb{P}\left\{\max_{1 \leq k \leq n} \|\sqrt{k}a'_k(K)\|_{\mathcal{G}'_j} > \beta'_n + t\right\} \\ &\leq n^{-C'\eta_0'^2}. \end{aligned}$$

Therefore :

$$\begin{aligned} &\mathbb{P}\left\{\sup_{a_n \leq h \leq b_0} \sqrt{\frac{n\phi_x(a_n)}{\log n}} |S_m - \mathbb{E}[S_m]| > \eta\right\} \\ &\leq L(n) \mathbb{P}\left\{\max_{1 \leq k \leq n} \|\sqrt{k}a'_k(K)\|_{\mathcal{G}'_j} > \eta'_0 \sqrt{n\phi_x(h_j/2) \log n}\right\} \leq \log(n) n^{-C'\eta_0'^2}. \end{aligned}$$

Consequently, an appropriate choice of η'_0 permits to deduce that :

$$\sup_{a_n \leq h \leq b_n} |S_m - \mathbb{E}[S_m]| = O_{a.co.}\left(\sqrt{\frac{\log n}{n\phi(a_n)}}\right).$$

Now, let us show the results (8) and (9). The proof of these lasts follows the same lines as the proof of (5) and (6), it suffices to write that

$$\text{Cov}(S_1, T_4) = \frac{1}{n^2 \phi_x^2(h)} \sum_i \text{Cov}\left(K_i Y_i, \frac{K_i \beta_i^2}{h^2}\right)$$

and

$$\text{Cov}(S_2, T_3) = \frac{1}{n^2 \phi_x^2(h)} \sum_i \text{Cov}\left(\frac{K_i \beta_i Y_i}{h}, \frac{K_i \beta_i}{h}\right).$$

Observe that, under (H4) et (H8), for $k = 0, 1$ and $l = 1, 2$, we have

$$\begin{aligned} \left|\text{Cov}\left(\frac{K_i \beta_i^k Y_i}{h^k}, \frac{K_i \beta_i^l}{h^l}\right)\right| &= \left|\mathbb{E}\left[\frac{K_i^2 \beta_i^{k+l} Y_i}{h^{k+l}}\right] - \mathbb{E}\left[\frac{K_i \beta_i^k Y_i}{h^k}\right] \mathbb{E}\left[\frac{K_i \beta_i^l}{h^l}\right]\right| \\ &\leq C \left|\mathbb{E}[K_i^2 Y_i]\right| + |\mathbb{E}[K_i Y_i]| |\mathbb{E}[K_i]| \\ &\leq C \left[\phi_x(h/2) + \phi_x^2(h/2)\right] \leq C \phi_x(h/2) \end{aligned}$$

it follows that

$$\sum_i \text{Cov}\left(\frac{K_i \beta_i^k Y_i}{h^k}, \frac{K_i \beta_i^l}{h^l}\right) \leq n \phi_x(h/2).$$

Finally, the case $(k, l) = (0, 2)$ implies that

$$\text{Cov}(S_1, T_4) \leq C\left(\frac{1}{n \phi_x(h)}\right) \leq C\left(\frac{1}{n \phi_x(h)}\right) = O_{a.co.}\left(\sqrt{\frac{\log n}{n \phi(a_n)}}\right)$$

While, if we consider the case of $(k, l) = (1, 1)$ we show that

$$\text{Cov}(S_2, T_3) C\left(\frac{1}{n \phi_x(h)}\right) \leq C\left(\frac{1}{n \phi_x(h)}\right) = O_{a.co.}\left(\sqrt{\frac{\log n}{n \phi(a_n)}}\right)$$

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

- [1] Baillo, A. and Grané, A. (2009). Local linear regression for functional predictor and scalar response, *J. Multiv. Anal.*, **100**, 102-111.
- [2] Barrientos-Marin, J., Ferraty, F. and Vieu, P. (2010). Locally Modelled Regression and Functional Data. *J. Nonparametr. Statist.*, **22**, 617-632.
- [3] Benhenni, K., Ferraty, F., Rachdi, M., Vieu, P. (2007). Local smoothing regression with functional data. *Comput. Statistics*, **22**, 353-370.
- [4] Berlinet, A., Elamine, A., and Mas, A. (2011). Local linear regression for functional data. *Ann. Inst. Statist. Math.*, **63**, 1047-1075.
- [5] Boj, E., Delicado, P. and Fortiana, J. (2010). Distance-based local linear regression for functional predictors. *Comput. Statist. Data Anal.*, **54**, Pages 429-437.
- [6] Chagny, G. and Roche, A. (2016) Adaptive estimation in the functional non-parametric regression model. *J. Multiv. Anal.*, **146**, 105-118.
- [7] Davidian, M., Lin, X. and Wang, J.L. (2004). Introduction (Emerging Issues in Longitudinal and Functional Data Analysis). *Statist. Sinica*, **14**, 613-614.
- [8] Deheuvels, P. and Ouadah, S. (2013). Uniform-in-bandwidth functional limit laws. *J. Theoret. Probab.*, **26**, 697-721.
- [9] Dony, J. and Einmahl, U. Uniform in bandwidth consistency of kernel regression estimators at a fixed point. *High dimensional probability V : the Luminy volume*, 308325, Inst. Math. Stat. Collect., 5, Inst. Math. Statist., Beachwood, OH, 2009.
- [10] Fan, J. and Gijbels, I. (1996). *Local Polynomial Modelling and its Applications*. London, Chapman & Hall.
- [11] Ferraty, F. and Vieu, P. (2006). *Nonparametric functional data analysis. Theory and Practice*. Springer Series in Statistics. New York.
- [12] Ferraty, F., Mas, A. and Vieu, P. (2007). Nonparametric regression on functional data : inference and practical aspects. *Aust. N. Z. J. Stat.*, **49**, Pages 267-286.
- [13] Goia, A. and Vieu, P. (2016). An introduction to recent advances in high/infinite dimensional statistics. *J. Multiv. Anal.*, **146**, 1-6.
- [14] Gonzalez-Manteiga, W. and Vieu, P. (2007). Statistics for functional data. *Comput. Statist. Data Anal.*, **51**, 4788-4792.
- [15] Karaa, L-Z. Laksaci, A, Rachdi, M. and Vieu, P. (2015) Uniform in bandwidth consistency results for kernel estimates in various statistical models with functional. *J. Nonparametr. Statist.* (accepted).

- [16] Kudraszow, N. and Vieu, P. (2013). Uniform consistency of k NN regressors for functional variables. *Statist. Probab. Lett.*, **83**, 1863-1870.
- [17] Rachdi, M. and Vieu, P. (2007). Nonparametric regression for functional data : automatic smoothing parameter selection. *J. Stat. Plann. Inference.*, **137**, 2784-2801.
- [18] Rinnana R. and Rinnan S. (2007). Application of near infrared reflectance (NIR) and fluorescence spectroscopy to analysis of microbiological and chemical properties of arctic soil. *Soil Biology & Biochemistry* , **39** , 1664-1673
- [19] Shang, H. L. (2016). A Bayesian approach for determining the optimal semi-metric and bandwidth in scalar-on-function quantile regression with unknown error density and dependent functional data. *J. Multiv. Anal.*, **146**, 95-104
- [20] Van der Vaart, A.W. and Wellner, J.A. (1996). *Weak Convergence and Empirical Processes : With Applications to Statistics*. Springer-Verlag, New York, 1996.
- [21] Zhou, Z. and Lin, Z. (2016). Asymptotic normality of locally modelled regression estimator for functional data. *J. Nonparametr. Statist.*, **28**, 116-131

Chapitre **4: On the k -Nearest Neighbor local linear estimation for functional regression : Uniform and bandwidth consistency**

Contents

On the k -Nerest Neighbor local linear estimation for functional regression : Uniform and bandwidth consistency

Mohammed Attouch, Ali Laksaci and Fatima Rafaa

Laboratoire de Statistique et Processus Stochastiques,

Université de Sidi Bel Abbès,

BP 89 Sidi Bel Abbès 22000. Algeria.

attou_kadi@yahoo.fr,

alilak@yahoo.fr,

rafaa_fatima@yahoo.com

Abstract : In this article, we are interested in estimating the regression function by the method of kNN local linear. Our goal is to establish the almost complete consistency uniformly in number of neighbourhood and on the set $S_{\mathcal{F}}$ in the proposed estimator.

Key word : Functional data analysis, Uniformity in bandwidth, Functional regression, kNN local linear estimator

4.1 introduction

Modeling by the linear local method has taken on great importance by several authors as Fan et al. (1996) [11], Masry et Fan (1997) [17]. Loader's (2004) [16] book is an important pedagogical document for the estimation of the linear local method. In 2006, Welsh in collaboration with Yee [18] addressed the case where the response variable is multi-dimensional. All previous results were obtained for a vector explanatory variable. the case of a functional explanatory variable has been very recently discussed.

At this stage, the literature is very limited, let's mention some author research : Baillo and Grané (2009) [4](the first in this field), Barrientos-Marin et al. (2010)[3] and Boj et al. (2010) [5].

If the data are heterogeneous and h taken in a discrete finite set we speak in this case of the kNN method. This method has interested several authors : Cover (1968), Stone (1977), Collomb (1981) or Devroye and al. (1994) [9] have been looking at the kNN method in the vector data case while the case of the functional data some studies have been done by Burba et al. (2009) for the pointwise almost complete consistency, Kudraszow and Vieu (2013) [14] for the uniform almost complete convergence and to Lian (2011)[15] for the functional regressor case. If we combine the local linear method and the kNN method we find the linear

local kNN estimator introduced in the article in the article Attouch, Laksaci and Rafea(2017) [1] or they established almost complete punctual and uniform convergence on the number of neighbors k .

In this paper, the regression function is estimated by the locally linear kNN estimator. As a result, the almost complete uniform convergence on the number of neighbors and on the set $S_{\mathcal{F}}$ is studied.

4.2 The representation of the estimator :

Nonparametric regression is a statistical tool allowing to describe the relationship between a dependent variable Y and an explanatory variable X .

Consider n pairs $(X_i, Y_i)_{i=1, \dots, n}$ identically and independently distributed as (X, Y) , which is valued in $\mathcal{F} \times \mathcal{R}$, where $\mathcal{F}$ is a semi-metric space equipped with a semi-metric d .

We proposes to study the nonparametric regression model of the form :

$$Y = m(X) + \epsilon, \quad \mathbb{E}(\epsilon|X) = 0$$

1

Considering initially the locally linear functional estimator studied in detail by Barrientos (2007).

for which the regression function $m(x)$ is estimated by

$$\hat{m}(x) = \frac{\sum_{i,j=1}^n W_{ij}(x) Y_j}{\sum_{i,j=1}^n W_{ij}(x)}$$

where

$$W_{ij}(x) = \beta(X_i, x) \left(\beta(X_i, x) - \beta(X_j, x) \right) K(H_{k,x}^{-1} \delta(x, X_i)) K(H_{k,x}^{-1} \delta(x, X_j))$$

with $\beta(.,.)$ is a known function from $\mathcal{F}^2$ into $\mathbb{R}$ such that, $\forall \xi \in \mathcal{F}$, $\beta(\xi, \xi) = 0$ and $\delta(.,.)$ is a function of $\mathcal{F} \times \mathcal{F}$ such that $d(.,.) = |\delta(.,.)|$. Furthermore,

$$H_{k,x} = \min \left\{ h \in R^+ \quad \sum_{i=1}^n 1_{B(x,h)}(X_i) = k \right\}$$

4.3 useful results and assumptions

We consider the following assumptions

(H1) The function $\phi_x(r) = \phi_x(-r, r)$, $r > 0$ is continuous at the neighborhood of 0 and checks

$$\phi_x(r) > 0, \forall r > 0.$$

(H2) $\forall x_1, x_2 \in S_{\mathcal{F}} \mid r(x_1) - r(x_2) \leq C_1 d^\beta(x_1, x_2)$.

(H3) the function $\beta(., .)$ checked $\forall x' \in \mathcal{F}$, $C|\delta(x, x')| \leq |\beta(x, x')| \leq C'|\delta(x, x')|$.

(H4) K is a continuous kernel, differentiable on its support $(-1/2, 1/2)$. Furthermore, the class of functions

$$\mathcal{K}_k = \{\cdot \mapsto \gamma^{-k} K(\gamma^{-1} \delta(x, \cdot)) \beta^k(x, \cdot), \gamma > 0\} \text{ is measurable punctually}$$

checking :

$$\sup_Q \int_0^1 \sqrt{1 + \log N(\epsilon \|F\|_{Q,2}, \mathcal{K}_k, d_Q)} d\epsilon < \infty$$

where the supremum is taken on all probability measures Q from space $\mathcal{F}$ with $\|F\|_{Q,2} < \infty$ and where F is the envelope function of the class $\mathcal{K}_k$. Here, d_Q is the metric $L_2(Q)$ and $N(\epsilon, \mathcal{K}_k, d_Q)$ is the minimum number of open balls (relative to the metric $L_2(Q)$) radius ϵ that are needed to cover the function class $\mathcal{K}_k$.

(H5) the two suites $k_{1,n}, k_{2,n}$ are such that

$$\phi_x^{-1}\left(\frac{k_{2,n}}{n}\right) \rightarrow 0 \quad \text{and} \quad \frac{\log n}{\min(n\phi_x^{-1}\left(\frac{k_{1,n}}{n}\right), k_{1,n})} \rightarrow 0.$$

(H6)

$$K(1/2)\tau_x(1/2) - \int_{-1/2}^{1/2} K'(s)\tau_x(s)ds > 0$$

$$\text{et} \quad (1/4)K(1/2)\tau_x(1/2) - \int_{-1/2}^{1/2} (s^2 K'(s))\tau_x(s)ds > 0.$$

(H7) The conditional moments of Y check $\forall m \geq 2$, $E[Y^m|X = \cdot] := \sigma_m(\cdot)$ is a continuous and bounded function

(H8) The functions φ and $\psi_{S_{\mathcal{F}}}$ verifies :

$$\text{For } n \text{ large enough, } \frac{(\log n)^2}{n\varphi(h_K)} < \psi_{S_{\mathcal{F}}}\left(\frac{\log n}{n}\right) < \frac{n\varphi(h_K)}{\log n}$$

Lemma 4.7. inequality of Bernstein's

Let $Z_1, Z_2, \dots, Z_n$ random variables i.i.d and either $\mathcal{G}$ class of measurable punctual functions with an envelope function G Suppose that for some $H > 0$

$$\forall m \geq 2, \quad \mathbb{E}[G^m(Z)] \leq \left(\frac{m!}{2}\right) \sigma^2 H^{m-2},$$

Where $\sigma^2 \geq \mathbb{E}[G^2(Z)]$. and for $\beta_n = \mathbb{E}[\|\sqrt{n}a_n(g)\|_{\mathcal{G}}]$, we have for all $t > 0$

$$\mathbb{P}\left\{\max_{1 \leq k \leq n} \|\sqrt{k}a_n(g)\|_{\mathcal{G}} \geq \beta_n + t\right\} \leq \exp\left(-\frac{t^2}{2n\sigma^2 + 2tH}\right)$$

Lemma 4.8. (theorem(3.2) of Einmahl and masson (2008))

Let $Z_1, Z_2, \dots, Z_n$ random variables i.i.d and $\mathcal{G}$ class of measurable punctual functions with an envelope function G satysfying for some $0 < \beta < \infty$, $\mathbb{E}G^2(Z) \leq \beta^2$ then we have

$$\mathbb{E}\left\|\sum_{i=1}^n (g(Z_i) - \mathbb{E}g(Z))\right\|_{\mathcal{G}} \leq C\sqrt{n\beta^2}, \quad n \geq 1.$$

4.4 Uniform convergence of kNN local linear regression estimator

Theorem 4.4. Under the assumption **H1** and **H8** we have :

$$\sup_{k_{1,n} \leq k \leq k_{2,n}} \sup_{x \in S_{\mathcal{F}}} |\hat{m}(x) - m(x)| = O\left(\varphi_x^{-1}\left(\frac{k_{2,n}}{n}\right)^v\right) + O_{a.c}\left(\sqrt{\frac{\psi_{S_{\mathcal{F}}}\left(\frac{\log n}{n}\right)}{k_{1,n}}}\right).$$

Proof of the Theorem 4.4 .: The proof of theorem 4.4 is based on the following decomposition : for some $a \in]0, 1[$

$$\begin{aligned} \sup_{k_{1,n} \leq k \leq k_{2,n}} \sup_{x \in S_{\mathcal{F}}} |\hat{m}(x) - m(x)| &= \sup_{k_{1,n} \leq k \leq k_{2,n}} \sup_{x \in S_{\mathcal{F}}} |\hat{m}(x) - m(x)| 1_{\varphi_x^{-1}\left(a \frac{k_{1,n}}{n}\right) \leq H_{k,x} \leq \varphi_x^{-1}\left(\frac{k_{2,n}}{an}\right)} \\ &+ \sup_{k_{1,n} \leq k \leq k_{2,n}} \sup_{x \in S_{\mathcal{F}}} |\hat{m}(x) - m(x)| 1_{H_{k,x} \notin \left(\varphi_x^{-1}\left(a \frac{k_{1,n}}{n}\right), \varphi_x^{-1}\left(\frac{k_{2,n}}{an}\right)\right)} \end{aligned}$$

thus for all $\varepsilon > 0$

$$\begin{aligned} &\mathbb{P}\left\{\sup_{k_{1,n} \leq k \leq k_{2,n}} \sup_{x \in S_{\mathcal{F}}} |\hat{m}(x) - m(x)| \geq \varepsilon Z_n\right\} \\ &\leq \mathbb{P}\left\{\sup_{k_{1,n} \leq k \leq k_{2,n}} \sup_{x \in S_{\mathcal{F}}} |\hat{m}(x) - m(x)| 1_{\varphi_x^{-1}\left(a \frac{k_{1,n}}{n}\right) \leq H_{k,x} \leq \varphi_x^{-1}\left(\frac{k_{2,n}}{an}\right)} \geq \frac{\varepsilon Z_n}{2}\right\} \\ &+ \mathbb{P}\left\{H_{k,x} \notin \left(\varphi_x^{-1}\left(a \frac{k_{1,n}}{n}\right), \varphi_x^{-1}\left(\frac{k_{2,n}}{an}\right)\right)\right\} \end{aligned}$$

finally it suffices to prove that

$$\sum_n \sum_{k=k_{1,n}}^{k_{2,n}} \mathbb{P}\left(H_{k,x} \leq \varphi_x^{-1}\left(a \frac{k_{1,n}}{n}\right)\right) < \infty, \quad \sum_n \sum_{k=k_{1,n}}^{k_{2,n}} \mathbb{P}\left(H_{k,x} \geq \varphi_x^{-1}\left(\frac{k_{2,n}}{an}\right)\right) < \infty \quad 2$$

and

$$\sup_{\varphi_x^{-1}\left(\frac{k_{1,n}}{an}\right) \leq h \leq \varphi_x^{-1}\left(\frac{k_{2,n}}{an}\right)} \sup_{x \in S_{\mathcal{F}}} |\hat{m}(x) - m(x)| = O_{a.s.}(z_n) \quad 3$$

with

$$z_n = O\left(\varphi_x^{-1}\left(\frac{k_{2,n}}{n}\right)^v\right) + O_{a.c.}\left(\sqrt{\frac{\psi_{S_{\mathcal{F}}}\left(\frac{\log n}{n}\right)}{k_{1,n}}}\right).$$

■

.1 Appendix

The quantity 2 is already proven(c.f article Laksaci and Kara(2017)[13]).
The proof of 3 is based on the following decomposition

$$\hat{m}(x) - m(x) = \widehat{B}(x) + \frac{\widehat{D}(x)}{\widehat{r}_{1,h}(x)} + \frac{\widehat{Q}(x)}{\widehat{r}_{1,h}(x)}$$

where

$$\begin{aligned} \widehat{Q}(x) &:= (\widehat{r}_{2,h}(x) - \mathbb{E}[\widehat{r}_{2,h}(x)]) - m(x)(\widehat{r}_{1,h}(x) - \mathbb{E}[\widehat{r}_{1,h}(x)]) \\ \widehat{B}(x) &:= \frac{\mathbb{E}[\widehat{r}_{2,h}(x)]}{\mathbb{E}[\widehat{r}_{1,h}(x)]} - m(x) \quad \text{and} \quad \widehat{D}(x) := -\widehat{B}(x)(\widehat{r}_{1,h}(x) - \mathbb{E}[\widehat{r}_{1,h}(x)]) \end{aligned}$$

with

$$\widehat{r}_{2,h}(x) = \frac{1}{n(n-1)h^2\varphi_x^2(h)} \sum_{i \neq j} W_{ij}(x) Y_j \quad \text{and} \quad \widehat{r}_{1,h}(x) = \frac{1}{n(n-1)h^2\varphi_x^2(h)} \sum_{i \neq j} W_{ij}(x).$$

Then, Theorem 4.4 is a direct consequence of the following Lemmas.

Lemma 0.9. *Under the hypotheses (H1), (H4)-(H5), we have that :*

$$\sup_{k_{1,n} \leq k \leq k_{2,n}} \sup_{x \in S_{\mathcal{F}}} |\widehat{r}_{1,h}(x) - \mathbb{E}[\widehat{r}_{1,h}(x)]| = O_{a.c.}\left(\sqrt{\frac{\psi_{S_{\mathcal{F}}}\left(\frac{\log n}{n}\right)}{k_{1,n}}}\right).$$

Lemma 0.10. *Under the hypotheses (H1)-(H2) we have that :*

$$\sup_{k_{1,n} \leq k \leq k_{2,n}} \sup_{x \in S_{\mathcal{F}}} |\widehat{B}(x)| = O\left(\varphi_x^{-1}\left(\frac{k_{2,n}}{n}\right)^v\right).$$

Lemma 0.11. *Under the assumptions of Theorem 4.4, we have that :*

$$\sup_{k_{1,n} \leq k \leq k_{2,n}} \sup_{x \in S_{\mathcal{F}}} |\widehat{g}(x) - E[\widehat{g}(x)]| = O_{a.c} \left(\sqrt{\frac{\psi_{S_{\mathcal{F}}} \left(\frac{\log n}{n} \right)}{k_{1,n}}} \right).$$

Proof of Lemma 0.11. We have

$$\widehat{g}(x) = \underbrace{\frac{n^2 h^2 \varphi_X(h/2)^2}{n(n-1) \mathbb{E}(w_{12})}}_Q \left[\underbrace{\left(\frac{1}{n} \sum_{j=1}^n \frac{K_j Y_j}{\varphi_X(h/2)} \right)}_{S_1} \underbrace{\left(\frac{1}{n} \sum_{i=1}^n \frac{K_i \beta_i^2}{h^2 \varphi_X(h/2)} \right)}_{S_2} - \underbrace{\left(\frac{1}{n} \sum_{j=1}^n \frac{K_j \beta_j Y_j}{h \varphi_X(h/2)} \right)}_{S_3} \underbrace{\left(\frac{1}{n} \sum_{i=1}^n \frac{K_i \beta_i}{h \varphi_X(h/2)} \right)}_{S_4} \right] \quad 4$$

which allows to write

$$\widehat{g}(x) - E[\widehat{g}(x)] = Q[S_1 S_2 - \mathbb{E}(S_1 S_2) - (S_3 S_4 - \mathbb{E}(S_3 S_4))]$$

in a first, we has :

$$\begin{aligned} S_1 S_2 - \mathbb{E}(S_1 S_2) &= (S_1 - \mathbb{E}(S_1))(S_2 - \mathbb{E}(S_2)) + S_1 \mathbb{E}(S_2) + S_2 \mathbb{E}(S_1) \\ &\quad - \mathbb{E}(S_1) \mathbb{E}(S_2) - \mathbb{E}(S_1) \mathbb{E}(S_2) + \mathbb{E}(S_1) \mathbb{E}(S_2) - \mathbb{E}(S_1 S_2) \\ &= (S_1 - \mathbb{E}(S_1))(S_2 - \mathbb{E}(S_2)) + (S_2 - \mathbb{E}(S_2)) \mathbb{E}(S_1) \\ &\quad + (S_1 - \mathbb{E}(S_1)) \mathbb{E}(S_2) + \mathbb{E}(S_1) \mathbb{E}(S_2) - \mathbb{E}(S_1 S_2). \end{aligned} \quad 5$$

It remains to study each term of the expression 5 based on the idea of Tadj Amel (2011)

$$\begin{aligned} \sup_{\varphi_X^{-1} \left(\frac{k_{1,n}}{an} \right) \leq h \leq \varphi_X^{-1} \left(\frac{k_{2,n}}{an} \right)} \sup_{x \in S_{\mathcal{F}}} |S_1(x) - \mathbb{E}(S_1(x))| &\leq \underbrace{\sup_{\varphi_X^{-1} \left(\frac{k_{1,n}}{an} \right) \leq h \leq \varphi_X^{-1} \left(\frac{k_{2,n}}{an} \right)} \sup_{x \in S_{\mathcal{F}}} |S_1(x) - S_1(x_{k(x)})|}_{G_1} \\ &+ \underbrace{\sup_{\varphi_X^{-1} \left(\frac{k_{1,n}}{an} \right) \leq h \leq \varphi_X^{-1} \left(\frac{k_{2,n}}{an} \right)} \sup_{x \in S_{\mathcal{F}}} |S_1(x_{k(x)}) - \mathbb{E}(S_1(x_{k(x)}))|}_{G_2} \\ &+ \underbrace{\sup_{\varphi_X^{-1} \left(\frac{k_{1,n}}{an} \right) \leq h \leq \varphi_X^{-1} \left(\frac{k_{2,n}}{an} \right)} \sup_{x \in S_{\mathcal{F}}} |\mathbb{E}(S_1(x_{k(x)})) - \mathbb{E}(S_1(x))|}_{G_3} \end{aligned}$$

$$\begin{aligned}
G_1 &= \sup_{\varphi_x^{-1}\left(\frac{k_{1,n}}{an}\right) \leq h \leq \varphi_x^{-1}\left(\frac{k_{2,n}}{an}\right)} \sup_{x \in S_F} \left| \frac{1}{n\varphi_x(h/2)} \sum_{i=1}^n K_i(x) Y_i - K_i(x_{k(x)}) Y_i \right| \\
&\leq \sup_{\varphi_x^{-1}\left(\frac{k_{1,n}}{an}\right) \leq h \leq \varphi_x^{-1}\left(\frac{k_{2,n}}{an}\right)} \sup_{x \in S_F} \frac{1}{n\varphi_x(h/2)} \sum_{i=1}^n |Y_i| |K_i(x) - K_i(x_{k(x)})| 1_{B(x, h/2) \cup B(x_{k(x)}, h/2)} \\
&\leq \sup_{\varphi_x^{-1}\left(\frac{k_{1,n}}{an}\right) \leq h \leq \varphi_x^{-1}\left(\frac{k_{2,n}}{an}\right)} \frac{1}{n\varphi_x(h/2)} \sum_{i=1}^n |Y_i| |K_i(x) - K_i(x_{k(x)})| \sup_{x \in S_F} 1_{B(x, h/2) \cup B(x_{k(x)}, h/2)}
\end{aligned}$$

(1) We pose $n_k = 2^k$, $k \geq 0$ and we define $h_{k,j} = 2^j a_{n_k}$ and $L(k) = \max\{j : h_{k,j} \leq 2b_0\}$. therefrom, we have

$$\sup_{\varphi_x^{-1}\left(\frac{k_{1,n}}{an}\right) \leq h \leq \varphi_x^{-1}\left(\frac{k_{2,n}}{an}\right)} \sqrt{\frac{n(k_{1,n})}{\log n}} |Z_i - \mathbb{E}(Z_i)| \leq \max_{1 \leq j \leq L(k)} \sup_{h_{k,j-1} \leq h \leq h_{k,j}} \sqrt{\frac{n\varphi_x(h)}{\log n}} |Z_i - \mathbb{E}(Z_i)|.$$

We write

$$Z_i - \mathbb{E}(Z_i) = \frac{1}{\sqrt{n}\varphi_x(h/2)} a_n(K)$$

such as, $a_n(K) = \frac{1}{\sqrt{n}} \sum_{i=1}^n Z_i$ empirical process, and we define the class of function

$$\mathcal{G}_{k,j} = \left\{ z \rightarrow 1_{B(z, h/2) \cup B(z_{k(z)}, h/2)}, \quad h_{k,j-1} \leq h \leq h_{k,j} \right\}$$

as a result,

$$\begin{aligned}
\mathbb{P} \left\{ \sup_{\varphi_x^{-1}\left(\frac{k_{1,n}}{an}\right) \leq h \leq \varphi_x^{-1}\left(\frac{k_{2,n}}{an}\right)} k_{1,n} \log n |Z_i - \mathbb{E}(Z_i)| \geq \rho \right\} \\
\leq \sum_{j=1}^{L(k)} \mathbb{P} \left\{ \frac{1}{\sqrt{n\varphi_x(h_{k,j}/2) \log n}} \left\| \sqrt{n} a_n(K) \right\|_{\mathcal{G}_{k,j}} \geq \rho \right\} \\
\leq L(k) \mathbb{P} \left\{ \max_{n_{k-1} \leq n \leq n_k} \left\| \sqrt{n} a_n(K) \right\|_{\mathcal{G}_{k,j}} \geq \rho \sqrt{n_k \varphi_x(h_{k,j}/2) \log n_k} \right\}
\end{aligned}$$

6

for evaluating 6 we use Lemma 4.7 Therefore, from inequality we are oblige to seek the quantities

$$\sigma^2 \geq \mathbb{E} \left[G_{k,j}^2(Z) \right] \quad \text{and} \quad \beta_n = \mathbb{E} \left[\left\| \sqrt{n} a_n(g) \right\|_{\mathcal{G}_{k,j}} \right].$$

From the assumption H4

$$G_{k,j}(z) \leq C 1_{B(x, h_{k,j}/2)}(z)$$

from which $\forall m \geq 1$,

$$\mathbb{E} \left[G_{k,j}^m(X) \right] \leq C^m \varphi_x(h_{k,j}/2).$$

in particular

$$\sigma^2 = O(\varphi_x(h_{k,j}/2)).$$

to increase β_n , we use Lemma 4.8 which gives the following result :

$$\mathbb{E} \left[\left\| \sqrt{n} a_n(g) \right\|_{\mathcal{G}_{k,j}} \right] \leq C \sqrt{n \varphi_x(h_{k,j}/2)}.$$

Now we can apply inequality of Bernstein's with $\beta_n = O(\sqrt{n_k \varphi_x(h_{k,j}/2)})$,

$$\sigma^2 = O(\varphi_x(h_{k,j}/2)), H = C \text{ and } t = \rho/2 \sqrt{n_k \varphi_x(h_{k,j}/2) \log n_k}$$

$$\begin{aligned} \mathbb{P} \left\{ \max_{n_{k-1} \leq n \leq n_k} \left\| \sqrt{n} a_n(K) \right\|_{\mathcal{G}_{k,j}} \geq \rho \sqrt{n_k \varphi_x(h_{k,j}/2) \log n_k} \right\} &\leq \mathbb{P} \left\{ \max_{n_{k-1} \leq n \leq n_k} \left\| \sqrt{n} a_n(K) \right\|_{\mathcal{G}_{k,j}} > \beta_n + t \right\} \\ &\leq \exp \left\{ -\rho^2 \frac{\log n_k}{8 + C \sqrt{\frac{\log n_k}{n_k \varphi_x(h_{k,j}/2)}}} \right\} \\ &\leq (\log n_k)^{-C' \rho^2}. \end{aligned}$$

and because $L(k) \leq 2 \log n_k$ we will have that

$$\begin{aligned} L(k) \mathbb{P} \left\{ \max_{n_{k-1} \leq n \leq n_k} \left\| \sqrt{n} a_n(K) \right\|_{\mathcal{G}_{k,j}} > \rho \sqrt{n_k \varphi_x(h_{k,j}/2) \log n_k} \right\} \\ \leq 2(\log n_k)^{1-C' \rho^2} \end{aligned}$$

and hence,

$$\mathbb{P} \left\{ \sup_{\varphi_x^{-1}\left(\frac{k_{1,n}}{an}\right) \leq h \leq \varphi_x^{-1}\left(\frac{k_{2,n}}{an}\right)} k_{1,n} \log n |S_1 - \mathbb{E}(S_1)| \geq \rho \leq 2(\log n_k)^{1-C' \rho^2} \right\}$$

and at the end it suffices to use the lemma of Borel Cantelli to conclude that

$$\sup_{\varphi_x^{-1}\left(\frac{k_{1,n}}{an}\right) \leq h \leq \varphi_x^{-1}\left(\frac{k_{2,n}}{an}\right)} |S_1 - \mathbb{E}(S_1)| = O_{a.c} \left(\sqrt{\frac{\log n}{k_{1,n}}} \right).$$

and it suffices to use (H5) to conclude that

$$\sup_{\varphi_x^{-1}\left(\frac{k_{1,n}}{an}\right) \leq h \leq \varphi_x^{-1}\left(\frac{k_{2,n}}{an}\right)} |S_1 - \mathbb{E}(S_1)| = O_{a.c} \left(\sqrt{\frac{\psi_{S_F}\left(\frac{\log n}{n}\right)}{k_{1,n}}} \right).$$

(2) We use the same procedures, we will obtain :

$$\sup_{\varphi_x^{-1}\left(\frac{k_{1,n}}{an}\right) \leq h \leq \varphi_x^{-1}\left(\frac{k_{2,n}}{an}\right)} |S_2 - \mathbb{E}(S_2)| = O_{a.c} \left(\sqrt{\frac{\psi_{S_F}\left(\frac{\log n}{n}\right)}{k_{1,n}}} \right).$$

Finally, we get :

$$\sup_{\varphi_x^{-1}\left(\frac{k_{1,n}}{an}\right) \leq h \leq \varphi_x^{-1}\left(\frac{k_{2,n}}{an}\right)} |S_1 S_2 - \mathbb{E}(S_1 S_2)| = O_{a.c} \left(\sqrt{\frac{\psi_{S_F}\left(\frac{\log n}{n}\right)}{k_{1,n}}} \right). \quad 7$$

In the same way and with similar arguments one can conclude that

$$\sup_{\varphi_x^{-1}\left(\frac{k_{1,n}}{an}\right) \leq h \leq \varphi_x^{-1}\left(\frac{k_{2,n}}{an}\right)} |S_3 S_4 - \mathbb{E}(S_3 S_4)| = O_{a.c} \left(\sqrt{\frac{\psi_{S_F}\left(\frac{\log n}{n}\right)}{k_{1,n}}} \right). \quad 8$$

To prove lemma 0.9 and lemma 0.10, refer to the article Attoch, Laksaci and Rafaa [2] ■

- [1] Attouch, M., Laksaci, A. and Rafaa, F. (2017). L'estimation locale linéaire de la régression non paramétrique fonctionnelle par la méthode des k plus proches voisins. *CRASS1-5935*.
- [2] Attouch, M., Laksaci, A. and Rafaa, F. (2017). On the local linear estimate for functional regression :Uniform in bandwidth consistency. *Communications in Statistics-Theory and Methods-Decision on Manuscript ID LSTA-2017-0001*.
- [3] Barrientos-Marin, J., Ferraty, F. and Vieu, P. (2010). Locally Modelled Regression and Functional Data. *J. of Nonparametric Statistics*, **22**, Pages 617–632.
- [4] Baillo, A. and Grané, A. (2009). Local linear regression for functional predictor and scalar response. *J. of Multivariate Analysis*, **100**, Pages 102–111.
- [5] Boj, E. Delicado, P. and Fortiana, J. (2010). Distance-based local linear regression for functional predictors. *Computational Statistics and Data Analysis*, **54**, Pages 429–437
- [6] Burba, F., Ferraty, F. and Vieu, P. (2009) *k*-nearest neighbor method in functional nonparametric regression. *J. Nonparametric Statist.*, **21**, 453-469.
- [7] Demongeot, J., Naceri, A. , Laksaci, A. and Rachdi, M. (2017). Local linear regression modelization when all variables are curves. *Statist. Probab. Lett.* **121**, 37–44.
- [8] Deheuvels, P. and Ouadah, S. (2013). Uniform-in-bandwidth functional limit laws. *J. Theoret. Probab.*, **26**(3), 697-721.
- [9] Devroye, L. Györfi, L. Krzyżak, A. and Lugosi, G. (1994). On the strong universal consistency of nearest neighbor regression function estimates. *Ann. Statist.*, **22**, 1371-1385.
- [10] Einmahl, U. and Mason, D. (2005). Uniform in bandwidth consistency of kernel-type function estimators. *Ann. Statist.*, **33**, 1380-1403.
- [11] Fan, J. and Gijbels, I. (1996). *Local Polynomial Modelling and its Applications*. London, Chapman & Hall.
- [12] Hsing, T. and Eubank, R. (2015). Theoretical foundations of functional data analysis, with an introduction to linear operators. Wiley Series in Probability and Statistics. John Wiley & Sons, Chichester.
- [13] Karaa, L-Z. Laksaci, A, Rachdi, M. and Vieu, P. (2017) Data-driven *k*NN estimation in nonparametric functional data analysis. *J. Multivariate Anal.* **153**, 176–188.
- [14] Kudraszow, N. and Vieu, P. (2013). Uniform consistency of *k*NN regressors for functional variables. *Statist. Probab. Lett.*, **83**, 1863-1870.

- [15] Lian, H. (2011). Convergence of functional k -nearest neighbor regression estimate with functional responses. *Electron. J. Statist.*, **5**, 31-40.
- [16] Loader, C. (2004). *Smoothing : local regression techniques. Handbook of computational statistics*, 539-563, Springer, Berlin,.
- [17] Masry, E. and Fan, J. (1997). Local polynomial estimation of regression functions for mixing processes. *Scand. J. Statist.*, **24**, 165-179.
- [18] Welsh, A. H. ; Yee, T. W. (2006). Local regression for vector responses. *Scand. J. Statist. Plann. Inference*, **136**, 3007-3031.
- [19] Zhang, J. (2014). *Analysis of variance for functional data*. Monographs on Statistics and Applied Probability, 127. CRC Press, Boca Raton, FL.

Dans la présente thèse, nous avons abordé l'étude globale de l'estimation de la fonction de régression lorsque la variable explicative fonctionnelle est la réponse réelle. Deux manières d'estimation ont été considérées, la première concerne l'estimation par la méthode locale linéaire tandis que la seconde concerne la méthode d'estimation par la méthode Knn locale.

Pour la première méthode nous avons choisi la méthode locale linéaire car elle est préférée à celle du noyau pour plusieurs raisons, et comme propriété asymptotique, nous avons traité la convergence uniforme presque complète sur le paramètre de lissage en précisant la vitesse de convergence.

Pour la deuxième méthode, nous avons estimé la fonction de régression par la méthode kNN locale laquelle a un biais réduit par rapport à la méthode kNN largement connu. Aussi et sous certaines conditions, nous avons établi la convergence uniforme et ponctuelle presque complète sur le nombre de voisins, de même nous avons établi la convergence uniforme sur l'ensemble $S_{\mathcal{F}}$.

Compte tenu de leurs importances, ces deux méthodes ont fait l'objet d'un nombre important de perspectives et de recherche. Dans la section suivante, nous avons listé quelques exemples de ces perspectives.

