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 ABBÈS

THESE DE DOCTORAT

Présentée par: **Omar Fetitah**

Domaine : Mathématique informatique
Filière : Mathématique
Intitulé de la formation : Probabilité et Statistiques Appliquées

Intitulée

Contribution à la régression non paramétrique fonctionnelle robuste pour des données manquantes et ses applications

Soutenue le 07/07/2021.
Devant le jury composé de :

Président :

Mr BENCHIKH Tawfik	Professeur	à l'université de S.B.A.

Examinateurs :

Mr KHARDANI Salah	Professeur	à l'université de Tunisie.
Mme RAHMANI Saâdia	Professeur	à l'université de Saida.

Directeur de thèse :

Mr ATTOUCH Mohammed Kadi	Professeur	à l'université de S.B.A.

Co-Directeur de thèse :

Mr RIGHI Ali	M.C.B .	à l'université de S.B.A.

Année universitaire : *2020/2021*

Dedication

This thesis is dedicated to my parents *Miloud* and *Kheira*.
I hope that this achievement will complete the dream that you had for me
all those many years ago.

A big thank to my brothers *Mohammed*, *Walid* and my sister *Zahra*.

"Statistics is the grammar of science"

—*Karl Pearson*

Acknowledgments

First and foremost, praises and thanks to ALLAH, the Almighty, for His showers of blessings throughout my research work to complete the research successfully.

I would like to express my deep and sincere gratitude to my research supervisor, M^{r}. Attouch Mohammed Kadi for giving me the opportunity to do research and providing invaluable guidance throughout this research. His dynamism, vision, sincerity and motivation have deeply inspired me. He has taught me the methodology to carry out the research and to present the research works as clearly as possible. It was a great privilege and honor to work and study under the supervision of M^{r}. Righi Ali. I am extremely grateful for what he has offered me. I am extending my heartfelt thanks to him for his acceptance and patience during the research work and thesis preparation.

I want to thank very warmly M^{r}. Bencheikh Tawfik for chairing my committee members. I heartily thank him for his help, advice and support throughout the course of this thesis.

I am extending my thanks to M^{r}. Khardani Salah for the warm welcome and hospitality he showed to me during my first official visit to Tunisia, for his support during my research work. I also thank him for giving me the honor to participate in my jury.

I am very honored that M^{me}. Rahmani Saâdia has agreed to participate in my jury, thank you for your time and effort in reviewing this work.

Finally, my thanks go to all the people who have supported me to complete the research work directly or indirectly.

Abstract

In this thesis, we study the asymptotic properties of functional parameters in nonparametric statistics for incomplete data. More precisely, we are interested in the robust and relative regression for which we build estimators and we study the asymptotic behavior in the censored and missing model.

We first studied, the asymptotic properties of a nonparametric estimator of the relative error regression given a functional explanatory variable, when the scalar response is right censored, in the both i.i.d. case and $\alpha-$mixing case. We establish the strong almost complete convergence rate and asymptotic normality of these estimators. A simulation study and real data application are performed to illustrate how this fact allows getting higher predictive performances than those obtained with standard estimates.

Finally, it seems possible to us to study the robust model, in the case of a scalar missing at random (MAR) response, for both cases, without and with unknown scale parameter. We establish, the almost complete convergence rate of our estimators in the two proposed models.
The numerical study is based on the statistical software R.

Key words: Relative error regression; Censored data; Functional data analysis; almost complete convergence; asymptotic normality; robust regression; $\alpha-$mixing data; missing at random data; scale parameter.

Résumé

Dans cette thèse, nous étudions les propriétés asymptotiques des paramètres fonctionnels en statistique non paramétrique pour des données incomplètes. Plus précisement, nous nous intéressons à la regression robuste et relative pour lesquelles nous construisons des estimateurs et nous étudions le comportement asymptotique dans le modèle censuré et manquantes.

Nous avons d'abord étudié les propriétés asymptotiques d'un estimateur non paramétrique de la régression d'erreur relative étant donné une variable explicative fonctionnelle, lorsque la réponse scalaire est censurée à droite, dans les deux cas i.i.d. et $\alpha-$mélange. Nous établissons la convergence uniforme presque complète et la normalité asymptotique de ces estimateurs. Une étude de simulation et une application de données réelles sont réalisées pour illustrer comment ce fait permet d'obtenir des performances prédictives supérieures à celles obtenues avec des estimations classiques.

Enfin, il nous semble possible d'étudier le modèle robuste, dans le cas d'une réponse scalaire manquante (MAR), dans les deux cas, sans et avec paramètre d'échelle. Nous établissons, la convergence presque complète de nos estimateurs, dans les deux modèles proposés.
Du coté numérique, notre étude est fondée sur le logiciel statistique R.

Mots- clés: Régression d'erreur relative; Données censurées; Analyse des données fonctionnelles; convergence presque complète; normalité asymptotique; régression robuste; données $\alpha-$ mélange; données manquant; paramètre d'échelle.

List of articles

1. O. FETITAH, I-M. ALMANJAHIE, M-K. ATTOUCH and A. RIGHI (2020). Strong convergence of the functional nonparametric relative error regression estimator under right censoring. Mathematica Slovaca 70(6), 1469-1490, DOI: https://doi.org/10.1515/ms-2017-0443.

2. O. FETITAH, M-K. ATTOUCH, S. KHARDANI and A. RIGHI (2020). Nonparametric relative error regression for dependent Functional data under random censorship. Submitted for publication.

3. M-K. ATTOUCH, O. FETITAH, S. KHARDANI and A. RIGHI (2020). Robust nonparametric equivariant regression for functional data with responses missing at random. Submitted for publication.

List of communications

1. O. FETITAH. Nonparametric Relative Error Regression Under Random Censorship Data. RAMA11'2019. Sidi Bel Abbès, 21-24 Nouvembre 2019.

2. O. FETITAH and M-K. ATTOUCH. Relative error regression with responses missing at random. JDFSE'2019. Sidi Bel Abbès, 14-15 Décembre 2019.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

GENERAL INTRODUCTION

1.1 Nonparametric Functional Data Analysis (NPFDA): Motivation and Examples

1.1.1 Motivation

Functional data analysis is a branch of statistics that has been the object of many studies and developments during the last few years. This type of data appears in many practical situations, as soon as one is interested in a continuous-time phenomenon for instance. For this reason, the possible application fields promising for the use of functional data are extensive: climatology, economics, linguistics, medicine. In the digital age, the development of computer systems and the computing power of machines have led to a constant increase in the quantity of data to be processed. This increase reveals, each time, the limitations of the analytical techniques used, which poses a real challenge to statisticians for the construction and development of new statistical methods, adapted to this profusion of data. Aforementioned is one of the reasons why, during the last twenty years, functional data analysis has become one of the most used tools for studying data in their dimensions, providing among other things several elements of answer to the problem.

Functional data analysis is used for twofold: In practice, it is often possible to collect data for which the observation times are different for each individual, which makes it difficult to approach by conventional methods of multivariate statistics. On the other hand, when the temporal variability is extremely frequent in the real data (in particular when one has high-resolution observations of any phenomenon), the functional statistic then makes it possible to avoid the simplification of some observations by replacing them by example by an average.

In the case of infinite dimensional spaces, the problem that always arises is that there is no existing for Lebesgue or any analogous measures. Also, it should be kept in mind that the Haar measure, which could be an alternative choice, does not generally exist in infinite dimensional space even for Hilbert spaces. Then, it becomes interesting to address the estimation problems in infinite dimensional spaces.

There are many results for non-parametric models. For instance, Ferraty and Vieu (2004) established the strong consistency of kernel estimators of the regression function when the explanatory variable is functional and the response is scalar. Their study is an extension of a previous work Ferraty and Vieu (2002) with non-standard regression problems such as time series prediction or curves discrimination. They highlighted the issue of the curse of dimensionality for functional data and gave methods to overcome the problem. Dabo-Niang (2004) studied density estimation in a Banach space with an application to the estimation density of a diffusion process with respect to Wiener's measure. The kernel type estimation of some characteristics of the conditional cumulative distribution function as well as the successive derivatives of the conditional density was introduced by Ferraty and Vieu (2006). Some asymptotic properties were established with a particular application to the conditional mode and quantile. The almost complete convergence with rates for the kernel type estimates is established and illustrated by an application to El Nino data Ferraty et al. (2006). It should be noted that there exist previous results for the conditional mode and quantile problems in abstract metric spaces. Finally, for more details on the subject, we refer the reader to the monograph of Ferraty and Vieu (2006).

1.1.2 Examples

Now we will introduce some statistical problematics and associated functional data. In fact, there are many nonparametrical statistical problems which occur in the functional setting. Sometimes, they appear purely in terms of statistical modelling or, on the contrary, they can be drawn directly from some specific functional datasets. This section describes various functional data. As we will see, these data have been choosen to cover different applied statistics fields, different shapes of curves (smooth, unsmooth), various grids of discretization (fine, sparse).

1. Near-infrared spectrometry provides benchmark examples coming from chemometrics. This is a non-destructive technology able to measure numerous chemi-

cal compounds in a wide variety of products for example in petroleum industry. The diesel fuel data set investigated here has been used in a number of previous studies (Boger (2003); Esteban-Dıéz et al. (2004); Feng et al. (2015)) for testing new variable selection and calibration algorithms. For instance, let us consider a sample of $n = 784$ diesel fuels samples. Each sample is illuminated by a light beam at 401 equally spaced wavelengths $(\omega_1, \ldots, \omega_{401})$ in the near-infrared range $750 - 1550$ nm. For each wavelength ω and each diesel sample i, the absorption $X_i(\omega)$ of radiation is measured. The ith discretized spectrometric curve is given by $X_i(\omega_1), \ldots, X_i(\omega_{401})$; Figure 1.1 displays the spectrometric curves.

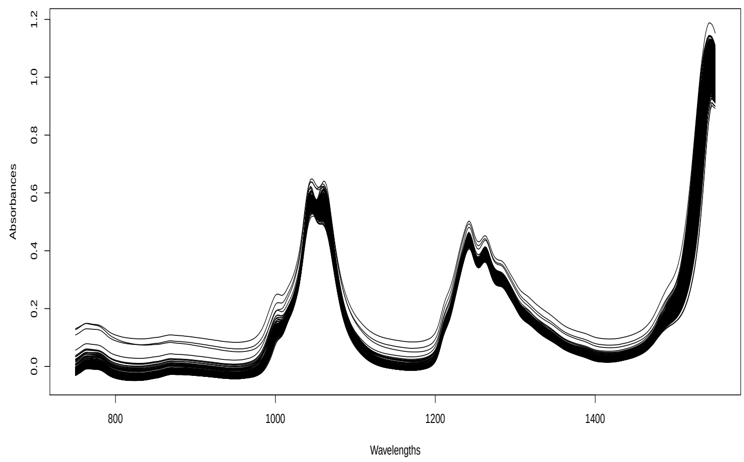

Figure 1.1: The spectrometric curves of the diesel fuels data.

2. In medicine, Diffusion Tensor Imaging (DTI) is a magnetic resonance imaging methodology which is used to measure the diffusion of water in the brain. Water diffuses isotropically (i.e. the same in all directions) in the brain except in white mater where it diffuses anisotropically (i.e. differently in different directions). This allows researchers to utilize DTI to generate images of white matter in the brain. Understanding the structure of the brain is important for a wide range of neurological conditions and diseases including Multiple Sclerosis. These data have been previously analyzed using several methods Goldsmith et al. (2011), Randolph et al. (2012). The Figure 1.2 shows the tract summaries sorted. This data set is available in the R package refund (Crainiceanu et al. (2012)), available on CRAN. A total of 376 patients are considered, with each tract measured at 93 equally spaced locations

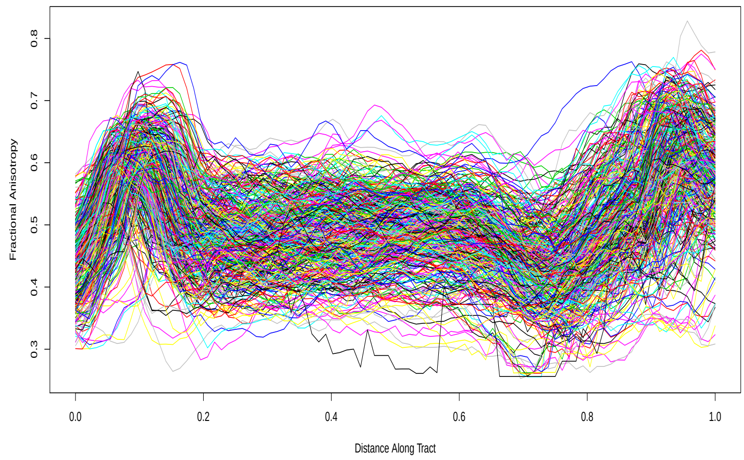

Figure 1.2: Fractional anisotropy (FA) curves along the right corticospinal tract.

3. One of the main specific problems in economic is to predict future or maximum consumption of electricity, and usual statistical models (either parametric or non-parametric) achieve that by taking into consideration a finite number of past data. However, one could think that it is more reasonable to take into account as explanatory variable the continuous time series over some period. We have acquired a large dataset, consisting of number of 8784 records, containing the hourly electricity consumption for the year 2016 (measured in MWh), retrieved from the smart metering device of a commercial center type of consumer (a large hypermarket) (see Pîrjan et al. (2017) for more description on this data set). For our example, we decided to choose the whole past day as explanatory period. That means that the set of explanatory variables to be included in our statistical method is composed with 366 curves data which are the 366 daily continuous time series. These functional data are presented in figure 1.3.

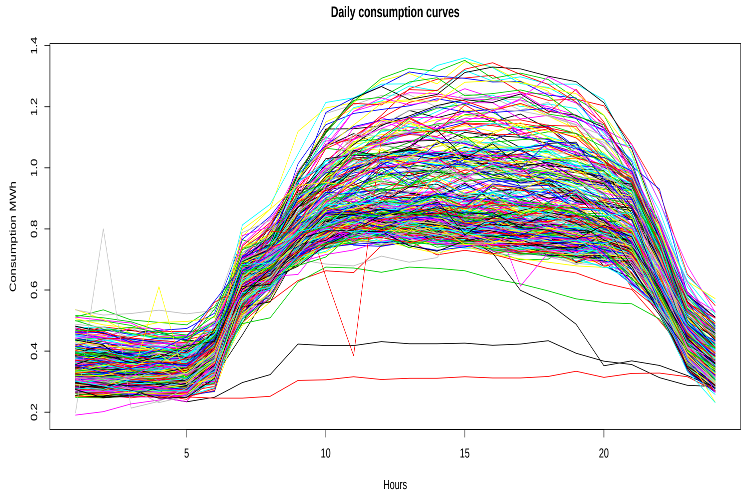

Figure 1.3: Electricity Consumption: Daily Curves.

4. Air pollution in developed countries has become a major problem in the daily lives of people living in these countries and has arisen and manifested itself by the high levels of smoke produced by industries or traffic, forcing authorities to search mechanisms to better control air quality in real time. For this purpose, our final example concerning the analyze the relationship between the palling gases such as the Ozone (O_3), Nitric Oxides (NO), Nitrogen Dioxide (NO_2) and Sulphur Dioxide (SO_2). The data of this contribution are acquired from real-time measurement by Marylebone road monitoring site. Marylebone Road is an important thoroughfare in central London, within the City of Westminster. The data used here are provided by the website `https://www.airqualityengland.co.uk/site/data?site_id=MY1`. It consist the hourly measurements during the period from January 1^{st} to the 31^{st} December for the year 2017. The daily emission of the gases observed are plotted in figure 1.4

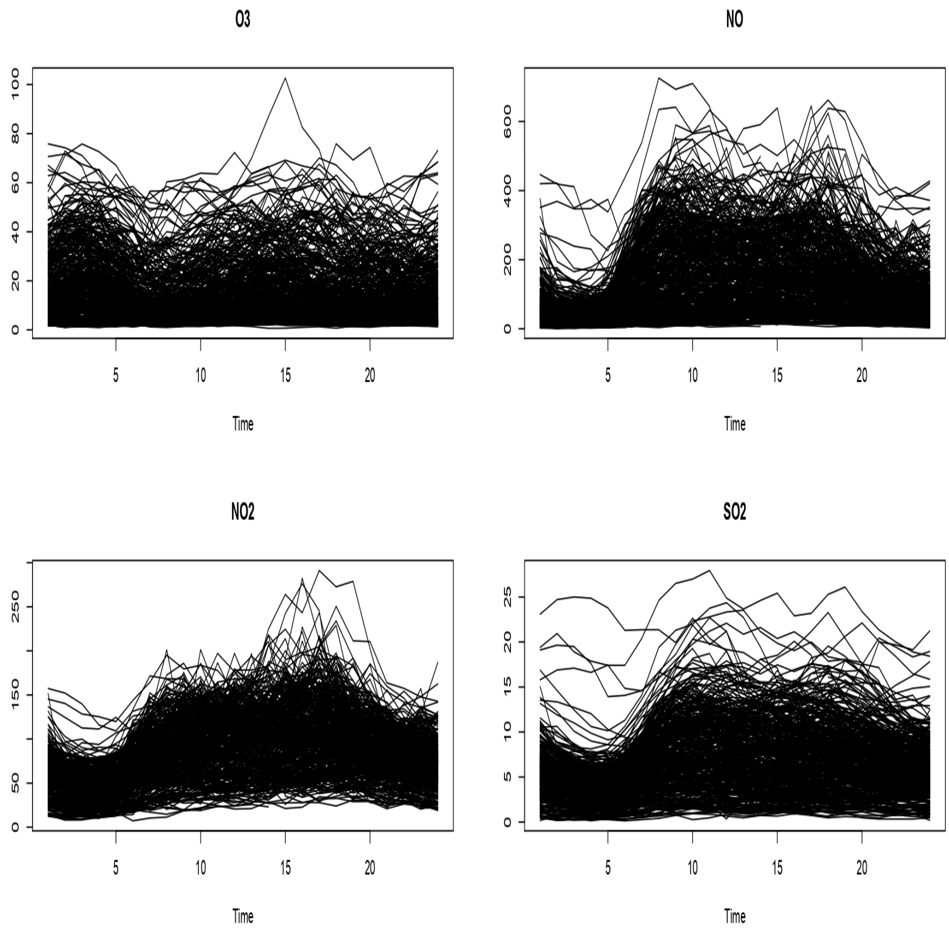

Figure 1.4: The curves of the daily emission of the gases in $\mu g/m^3$.

1.2 Robustness in Nonparametric Statistic

The robustness of a standard statistical procedure (estimation, test) is a very important question in statistics. It makes it possible to control the stability of this procedure relative to the deviation of the model and / or of the observations. Note that this problem was the subject of a long debate at the end of the XIX century, several scientists already had a relatively clear idea of this notion of robustness. In fact, the first mathematical work on robust estimation seems to have gone back in 1818 with the work of Laplace (1818). in his second supplement to the analytical theory of probabilities. More precisely, the term "robust" was introduced in 1954 by Box et al. (1954). But this notion was not recognized as a field of research until the mid-sixties. It is especially with the work of Huber (1992), Hampel (1971) that a coherent theory of robust statistics has been developed based on criteria of the minmax type and essentially uses arguments of convexity. From another point of view, other authors (Huber et al. (1973) and Huber and Ronchetti (1981)). Andrews (1974), Krasker and Welsch (1982), have developed automatic methods of robust adjustment, which is effective as the method of least squares when there are no aberrant points, but more effective in the presence of atypical observations or when the distribution of the error in the model follows a distribution with heavy tails.

It is well-known that the classical regression methods are sensitive to the outliers. The treatment of outliers is an essential step in highlighting the features of any data set. In this situation, outlying observations can be even more dangerous since the shape of the estimated curve is highly sensitive to outlying observations. Therefore, in order to overcome this problem, we consider a robust approach. More precisely, we are interested in the class of M-estimates, which was introduced by Huber (1965). The first work of robust estimates for nonparametric regression was given by Cleveland (1979), who give local polynomial fit versions by introducing weights to deal with large residuals. Härdle and Gasser (1984) and Tsybakov (1982), also studied pointwise asymptotic properties of a robust version of the Nadaraya−Watson method. These results were extended to M−type scale equivariant kernel estimates by Boente and Fraiman (1989) and Härdle et al. (1988) who also considered robust equivariant nonparametric estimates using nearest neighbor weights.

In particular, the presence of outliers can lead to unreasonable results since all variables have the same weight. Now, to overcome this limitation an alternative robust tool can be used in this kind of situation as relative error method. The literature on the relative error regression in Nonparametric Functional Data Analysis (NFDA) is still limited. The first consistent results were obtained in by Campbell and Donner (1989), where relative regression was used as a classification tool. Jones et al. (2008) studied the nonparametric prediction via relative error regression. They investigated the asymptotic properties of an estimator minimizing the sum of the squared relative errors by considering both (kernel method and local linear approach). Recently, Mechab and Laksaci (2016) studied this regression model when the observations are weakly dependent. For spatial data, Attouch et al. (2017) proved the almost complete consistency and the asymptotic normality of this estimator. Altendji et al. (2018) investigated the relative error in functional regression under random left-truncation model.

1.3 Incomplete Data

1.3.1 Censored data

Many researchers consider survival data analysis to be merely the application of two conventional statistical methods to a special type of problem: parametric if the distribution of survival times is known to be normal and nonparametric if the distribution is unknown. This assumption would be true if the survival times of all the subjects were exact and known; however, some survival times are not.

For example, some patients may still be alive or disease free at the end of the study period. The exact survival times of these subjects are unknown. These are called censored observations or censored times and can also occur when people are lost to follow up after a period of study. When these are not censored observations, the set of survival times is complete. There are three types of right censoring.

Type I censorship :

Instead of observing the variables $Y_1, Y_2, \ldots, Y_n$ which interest us, we observe Y_i when it is less than a fixed duration C, otherwise we only know that Y_i is greater than C. We therefore observe a variable T_i such that $T_i = \min(Y_i, C)$.

Type II censorship:

We observe the lifetimes of n patients until r of them have died and we stop at this point. If we order the $Y_1, Y_2, \ldots, Y_n$, we get the order statistics $Y_{(1)}, Y_{(2)}, \ldots, Y_{(n)}$. The censorship date is then $Y_{(r)}$ and we observe $T_{(1)} = Y_{(1)}, T_{(2)} = Y_{(2)}, \ldots, T_{(r)} = Y_{(r)}, T_{(r+1)} = Y_{(r)}, \ldots, T_{(n)} = Y_{(r)}$.

Type III censorship:

Definition: Given an $n-$sample $Y_1, Y_2, \ldots, Y_n$ of a positive random variable Y, we say that there is random censorship of this sample if there exists an $n-$dimensional random variable $C_1, C_2, \ldots, C_n$ such that, instead of observing $Y_1, Y_2, \ldots, Y_n$, we observe

$$(T_i, \delta_i) \quad \text{with} \quad T_i = \min(Y_i, C_i) \quad \text{and} \quad \delta_i = \mathbb{1}_{Y_i \le C_i}. \tag{1.1}$$

Here T_i is the duration actually observed. We know, moreover, what is the nature of this duration: if $\delta_i = 1$, it is a survival, if $\delta_i = 0$, it is a censorship.

1.3.2 Missing At Random (MAR) data

In many practical works including for instance sampling survey, pharmaceutical tracing or reliability, data are often incompletely observed and part of the responses are missing at random (MAR). The literature in multivariate setting for MAR samples is rather developed (see for among other Cheng (1994), Little and Rubin (2019) and Efromovich (2011)). When the explanatory variable is infinite dimensional, as far as we know the only contribution dealing with MAR sample is by Ferraty et al. (2013) and concerns the simple (parametric) problem of estimating the response mean.

Let (X, Y, δ) be a random variables (rv) in $\mathcal{F} \times \mathbb{R} \times \{0, 1\}$, where $(\mathcal{F}, d)$ is a semi-metric space (i.e. X is a functional random variable (f.r.v) and d a semi-metric). Let x be a fixed element of $\mathcal{F}$. One has an incomplete sample of size n from (X, Y, δ) which is classically denoted by $\{(X_i, Y_i, \delta_i), 1 \le i \le n\}$, where $\delta_i = 1$ if Y_i is observed, and $\delta_i = 0$ otherwise. The Bernoulli random variable δ is supposed to be such that

$$\mathbb{P}(\delta = 1 \mid X = x, Y = y) = \mathbb{P}(\delta = 1 \mid X = x) = p(x),$$

where $p(x)$ is a functional operator. This last condition models the fact that the censoring process δ is, conditionally on X, independent of the response Y.

1.4 Brief presentation of results

In this section, we give a brief presentation of the different results obtained for each chapter of this thesis.

1.4.1 Presentation of the estimators

We consider model 1.1, assuming that Y and C admit continues c.d.f. H and G, respectively. Consider a functional variable X of $\mathcal{F}$ representing a covariate variable of the regression function r.

Now, we estimate the regression function $r(.)$ under our relative loss function as

$$\widetilde{r}_n(x) = \frac{\sum_{i=1}^{n} \frac{\delta_i T_i^{-1}}{\bar{G}_n(T_i)} K\left(\frac{d(x, X_i)}{h}\right)}{\sum_{i=1}^{n} \frac{\delta_i T_i^{-2}}{\bar{G}_n(T_i)} K\left(\frac{d(x, X_i)}{h}\right)}, \tag{1.2}$$

with $\bar{G}_n(.) = 1 - G_n(.)$ where $G_n(.)$ is the Kaplan-Meier estimator of $G(.)$, K is the kernel and $h := h_n$ is a sequence of positive reals tending towards zero when $n \to \infty$.

Now we aim to generalize the results of Boente and Vahnovan (2015) which are obtained in the complete case to missing case (see 1.3.2).
We consider a real, measurable function, denoted ψ. The functional parameter studied in this note, noted ϑ_x, is solution of the following equation:

$$\Gamma(x,t,\sigma) := \mathbb{E}\left(\psi\left(\frac{Y-t}{\sigma}\right) \mid X = x\right) = 0, \tag{1.3}$$

where σ is a robust measure of the conditional scale. This measure can be taken as the conditional median of the absolute deviation from the conditional median, that is,

$$\sigma := s(x) = \text{MED}(|Y - m(x)|\,|X = x) = \text{MAD}_C\left(F_Y^x(\cdot)\right).$$

Denote by $\widehat{s}(x)$ a robust estimator of the conditional scale, for instance, $\widehat{s}(x) = \text{MAD}_\text{C}(\widehat{F}(\cdot \mid X = x))$, the scale measure. On the other hand, the robust nonparametric estimator of ϑ_x is given by the solution $\widehat{\vartheta}_x$ of $\widehat{\Gamma}(x,.,\widehat{s}(x)) = 0$, where $\widehat{\Gamma}(x,t,\widehat{s}(x))$ as an estimate of $\Gamma(x,t,s(x))$ by

$$\widehat{\Gamma}(x,t,\widehat{s}(x)) = \frac{\sum_{i=1}^n \delta_i K\left(h^{-1} d\left(x, X_i\right)\right) \psi\left(\frac{Y_i - t}{\widehat{s}(x)}\right)}{\sum_{i=1}^n \delta_i K\left(h^{-1} d\left(x, X_i\right)\right)}.$$

We also note by $\widehat{\theta}_x$ the estimator of θ_x which is the solution of 1.3 in the case with $\sigma = 1$.

1.4.2 Results: Relative i.i.d. Case

In this part, we assume that our observations are i.i.d. If the regression function $r(.)$ Satisfies certain regularity conditions then, under general technical hypotheses, we establish the almost complete uniform convergence and the asymptotic normality of our proposed estimator.

Theorem 1.4.1. *Under certain assumptions mentioned in 2.3. We have*

$$\sup_{x \in \mathcal{F}} |\widetilde{r}_n(x) - r(x)| = O_{a.co.}\left(h^{k_1}\right) + O_{a.co.}\left(h^{k_2}\right) + O_{a.co.}\left(\sqrt{\frac{\psi_{S_\mathcal{F}}\left(\frac{\log n}{n}\right)}{n\varphi_x(h)}}\right). \tag{1.4}$$

Theorem 1.4.2. *Under certain assumptions mentioned in 2.3. We have*

$$\left(\frac{n\varphi_x(h)}{\sigma^2(x)}\right)^{1/2}(\widetilde{r}_n(x)-r(x)-B_n(x)-o(h))\xrightarrow{\mathcal{D}}\mathcal{N}(0,1)\quad as\quad n\to\infty, \tag{1.5}$$

where $\xrightarrow{\mathcal{D}}$ *means the convergence in distribution. Also,*

$$B_n(x)=\frac{(\Psi_1'(0)-r(x)\Psi_2'(0))\,\beta_0}{\beta_1 g_2(x)} \tag{1.6}$$

and

$$\sigma^2(x)=\frac{(q_2(x)-2r(x)q_3(x)+r^2(x)q_4(x))\,\beta_2}{\beta_1^2} \tag{1.7}$$

with

$$\beta_0=K(1)-\int_0^1(sK(s))'\chi_x(s)\mathrm{d}s\quad and\quad \beta_j=K^j(1)-\int_0^1\left(K^j\right)'(s)\chi_x(s)\mathrm{d}s$$

for $j=1,2$.

1.4.3 Results: Relative $\alpha-$mixing Case

In order to generalize the results obtained in chapter 2 to the dependent observations, we reinforce the previous hypotheses, by adding hypotheses on the concentration of joint law (X_i,X_j) and on the mixing coefficient. We establish the asymptotic properties of estimator.

Theorem 1.4.3. *Under restrictive assumptions on the mixing coefficient (see 3.3), the kernel and the regression function* r*, we have*

$$|\widetilde{r}_n(x)-r(x)|=O\left(h^{k_1}\right)+O\left(h^{k_2}\right)+O_{a.s.}\left(\sqrt{\frac{\log n}{n\varphi_x(h)}}\right). \tag{1.8}$$

Theorem 1.4.4. *Under regularity assumptions mentioned in 2.3, we have*

$$\left(\frac{n\varphi_x(h)}{\sigma^2(x)}\right)^{1/2}\left(\widetilde{r}_n(x)-r(x)-hB_n(x)-o(h)\right)\xrightarrow{\mathcal{D}}\mathcal{N}(0,1)\quad as\quad n\to\infty. \tag{1.9}$$

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

$$B_n(x)=\frac{\left(\Psi_1'(0)-r(x)\Psi_2'(0)\right)\beta_0}{\beta_1 g_2(x)} \tag{1.10}$$

and

$$\sigma^2(x) = \frac{\left(q_2(x) - 2r(x)q_3(x) + r^2(x)q_4(x)\right)\beta_2}{\beta_1^2}. \tag{1.11}$$

with

$$\beta_0 = K(1) - \int_0^1 (sK(s))'\chi_x(s)\mathrm{d}s \quad \textit{and} \quad \beta_j = K^j(1) - \int_0^1 \left(K^j\right)'(s)\chi_x(s)\mathrm{d}s \neq 0$$

for $j = 1, 2$.

1.4.4 Results: Robust i.i.d. Case

In this part the observations are considered independent. If the model satisfies certain regularity conditions, we have the following asymptotic properties for the estimators $\widehat{\theta}_x$ and $\widehat{\vartheta}_x$.

Theorem 1.4.5. *Under concentration hypothesis of the probability measure of the functional variable and standard technical conditions for the kernel and the bandwidth (see 4.2.2), we have*

$$\left|\widehat{\theta}_x - \theta_x\right| = O_{a.co.}\left(h^b + \sqrt{\frac{\log n}{n\varphi_x(h)}}\right). \tag{1.12}$$

Theorem 1.4.6. *Under standard assumptions and technical conditions on the scale parameter* σ *(see 4.2.3), we have*

$$\left|\widehat{\vartheta}_x - \vartheta_x\right| = O_{a.co.}\left(h^b + \sqrt{\frac{\log n}{n\varphi_x(h)}}\right). \tag{1.13}$$

BIBLIOGRAPHY

B. Altendji, J. Demongeot, A. Laksaci, and M. Rachdi. Functional data analysis: estimation of the relative error in functional regression under random left-truncation model . *Journal of Nonparametric Statistics*, 30:472–490, 2018.

D. F. Andrews. A robust method for multiple linear regression. *Technometrics*, 16:523–531, 1974.

M. Attouch, A. Laksaci, and N. Messabihi. Nonparametric relative error regression for spatial random variables. *Stat Papers DOI*, 58, 2017.

G. Boente and R. Fraiman. Robust nonparametric regression estimation. *Journal of Multivariate Analysis*, 29:180–198, 1989.

G. Boente and A. Vahnovan. Strong convergence of robust equivariant nonparametric functional regression estimators. *Statistics and Probability Letters*, 11:1–11, 2015.

Z. Boger. Selection of quasi-optimal inputs in chemometrics modeling by artificial neural network analysis. *Analytica Chimica Acta*, 490:31–40, 2003.

G. E. Box et al. Some theorems on quadratic forms applied in the study of analysis of variance problems, i. effect of inequality of variance in the one-way classification. *The annals of mathematical statistics*, 25(2):290–302, 1954.

M. Campbell and A. Donner. Classification efficiency of multinomial logistic regression relative to ordinal logistic regression . *J. Amer. Statist. Asso*, 84:587–591, 1989.

P. E. Cheng. Nonparametric estimation of mean functionals with data missing at random. *Journal of the American Statistical Association*, 89:81–87, 1994.

W. S. Cleveland. Robust locally weighted regression and smoothing scatterplots. *Journal of the American statistical association*, 74:829–836, 1979.

C. Crainiceanu, P. Reiss, J. Goldsmith, L. Huang, L. Huo, F. Scheipl, S. Greven, J. Harezlak, M. Kundu, and Y. Zhao. refund: Regression with functional data. *R package version 0.1-6*, 2012.

S. Dabo-Niang. Kernel density estimator in an infinite–dimensional space with a rate of convergence in the case of diffusion process. *Applied Mathematics Letters*, 17:381–386, 2004.

S. Efromovich. Nonparametric regression with predictors missing at random. *Journal of the American Statistical Association*, 106:306–319, 2011.

I. Esteban-Dıéz, J. González-Sáiz, and C. Pizarro. Owavec: a combination of wavelet analysis and an orthogonalization algorithm as a pre-processing step in multivariate calibration. *Analytica chimica acta*, 515:31–41, 2004.

F. Feng, Q. Wu, and L. Zeng. Rapid analysis of diesel fuel properties by near infrared reflectance spectra. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, 149:271–278, 2015.

F. Ferraty and P. Vieu. The functional nonparametric model and application to spectrometric data. *Computational Statistics*, 17:545–564, 2002.

F. Ferraty and P. Vieu. Nonparametric models for functional data, with application in regression, time series prediction and curve discrimination. *Journal of Nonparametric Statistics*, 16:111–125, 2004.

F. Ferraty and P. Vieu. *Nonparametric Functional Data Analysis: Theory and Practice*. Springer Series in Statistics, New York, springer edition, 2006.

F. Ferraty, A. Laksaci, and P. Vieu. Estimating some characteristics of the conditional distribution in nonparametric functional models. *Statistical Inference for Stochastic Processes*, 9:47–76, 2006.

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

J. Goldsmith, J. Bobb, C. Crainiceanu, B. Caffo, and D. Reich. Penalized functional regression. *Journal of computational and graphical statistics*, 20:830–851, 2011.

F. R. Hampel. A general qualitative definition of robustness. *The Annals of Mathematical Statistics*, pages 1887–1896, 1971.

W. Härdle and T. Gasser. Robust nonparametric function fitting. *Journal of the Royal Statistical Society: Series B (Methodological)*, 46:42–51, 1984.

W. Härdle, A. B. Tsybakov, et al. Robust nonparametric regression with simultaneous scale curve estimation. *The annals of statistics*, 16:120–135, 1988.

P. Huber and E. Ronchetti. Robust statistics john wiley & sons. *New York*, 1(1), 1981.

P. J. Huber. A robust version of the probability ratio test. *The Annals of Mathematical Statistics*, 36:1753–1758, 1965.

P. J. Huber. Robust estimation of a location parameter. In *Breakthroughs in statistics*, pages 492–518. Springer, 1992.

P. J. Huber et al. Robust regression: asymptotics, conjectures and monte carlo. *The annals of statistics*, 1:799–821, 1973.

M. Jones, P. Heungsun, S. Key-Il, and S. Vines. Relative error prediction via kernel regression smoothers. . *J.Statist. Plann. Inference*, 138:2887–2898, 2008.

W. S. Krasker and R. E. Welsch. Efficient bounded-influence regression estimation. *Journal of the American statistical Association*, 77:595–604, 1982.

P. Laplace. Deuxiéme supplément to laplace (1812). *Oeuvr compl.*, 7:531–580, 1818.

R. J. A. Little and D. B. Rubin. *Statistical analysis with missing data*, volume 793. Wiley, 2019.

W. Mechab and A. Laksaci. Nonparametric relative regression for associated random variables. *Metron*, 74:75–97, 2016.

A. Pîrjan, S. V. Oprea, G. Cărutasu, D. M. Petrosanu, A. Bâra, and C. Coculescu. Devising hourly forecasting solutions regarding electricity consumption in the case of commercial center type consumers. *Energies*, 10:17–27, 2017.

T. Randolph, J. Harezlak, and Z. Feng. Structured penalties for functional linear models—partially empirical eigenvectors for regression. *Electronic journal of statistics*, 6: 323, 2012.

A. B. Tsybakov. Robust estimates of a function. *Problemy Peredachi Informatsii*, 18: 39–52, 1982.

CHAPTER 2

STRONG CONVERGENCE OF THE FUNCTIONAL NONPARAMETRIC RELATIVE ERROR REGRESSION ESTIMATOR UNDER RIGHT CENSORING

This chapter is the subject of a publication in Journal of Mathematica Slovaca.

Strong convergence of the functional nonparametric relative error regression estimator under right censoring

[1]**Omar FETITAH**, [2]**Ibrahim M. Almanjahie**, [2]**Mohammed Kadi ATTOUCH** and [1]**Ali Righi**.

[1] Laboratory of Statistics and Stochastic Processes University of Djillali Liabes BP 89, Sidi Bel Abbes 22000, Algeria.

[2]Department of Mathematics Faculty of Sciences King Khalid University PO. Box 9004, Abha, Saudi Arabia.

Abstract : In this paper, we investigate the asymptotic properties of a nonparametric estimator of the relative error regression given a functional explanatory variable, in the case of a scalar censored response, we use the mean squared relative error as a loss function to construct a nonparametric estimator of the regression operator of these functional censored data. We establish the strong almost complete convergence rate and asymptotic normality of these estimators. A simulation study is performed to illustrate and compare the higher predictive performances of our proposed method to those obtained with standard estimators.

Keywords : Relative error regression, Censored data, Nonparametric kernel estimation, Functional data analysis, Almost complete convergence, Asymptotic normality, Small ball probability.

Mathematics Subject Classification: 62G05, 62G08, 62G20, 62G35, 62N01.

2.1 Introduction

Functional data analysis is a branch of statistics that has been the object of many studies and developments during the last few years. This type of data appears in many practical situations, as soon as one is interested in a continuous-time phenomenon, for instance. This increase reveals, each time, the limitations of the analytical techniques used, which poses a real challenge to statisticians for the construction and development of new statistical methods, adapted to this profusion of data. Aforementioned is one of the reasons why, during the last twenty years, functional data analysis has become one of the most used tools for studying data in their dimensions, providing among other things several elements of answer to the problem (See, for instance, Ramsay (2004) and Ferraty and Vieu (2006)).

Modeling functional variables have received increasing interest in the last few years from mathematical or application points of view. There are many results for nonparametric models for more details on the subject, and we refer the reader to the monograph of Ferraty and Vieu (2006).

The study of a scalar response variable Y given a new value for the explanatory variable X is an important subject in nonparametric statistics. This regression relation is modeled by:

$$Y = r(X) + \epsilon, \tag{2.1}$$

where $r(.)$ is the regression function and ϵ a sequence of error independent to X.

Usually, $r(.) = \mathbb{E}[Y|X = .]$ is estimated by minimizing the mean squared loss function. However, this loss function is based on some restrictive conditions that is the variance of the residual is the same for all the observations, which is inadequate when the data contains some outliers.

When the predicted values are large or when the data contain many outliers, the following criterium

$$\mathbb{E}\left[\left(\frac{Y - r(X)}{Y}\right)^2 |X\right], \text{ for } , Y > 0 \tag{2.2}$$

is a more meaningful measure of the prediction performance than the least square error. Notice that this kind of model, so-called relative error regression, has been widely studied in parametric regression analysis. When the first two conditional inverse moments of Y given X are finite, the solution is given by the minimization of the sum of absolute relative errors for a linear model of the following ratio:

$$r(x) = \frac{\mathbb{E}[Y^{-1}|X = x]}{\mathbb{E}[Y^{-2}|X = x]}. \tag{2.3}$$

The least absolute relative error estimation for multiplicative regression models was proposed by Chen et al. (2010), who proved consistency and asymptotic normality of their estimator and also provided an inference approach via random weighting. Ruiz-Velasco (1991) discussed the asymptotic efficiency of relative logistic regression in a parametric context, particularly when explanatory variables are normally distributed. Moreover, Jones et al. (2008) has built a consistent estimator for this model using the kernel method.

They established asymptotic properties, especially its quadratic convergence, in the case where the observations are independent and identically distributed.

The literature on the relative error regression (RER) in nonparametric functional data analysis is still not very developed. The first consistent results were obtained by Campbell and Donner (1989), where relative regression was used as a classification tool. For the kernel method combined with the local linear method, Jones et al. (2008) gives the asymptotic properties of the nonparametric prediction via relative error regression. Recently, Attouch et al. (2017) proposed a kernel regression estimator version in the spatial framework context and derived asymptotically and numerically the effectiveness of this kind of estimator, whereas Demongeot et al. (2016) proposed a functional version of the relative kernel regression estimator while Thiam (2018) proposed a nonparametric method estimation for deconvolution regression model using relative error prediction.

On the other hand, the literature on the nonparametric analysis of incomplete functional data is quite restricted. There are very few results on this topic (see, for instance, Altendji et al. (2018) for estimation of the relative error in functional regression under the random left-truncation model). In the right censorship model, the kernel estimator of the classical regression was introduced by Carbonez, Carbonez et al. (1995), and improved by Köhler et al. (2002). This model was used later by Ould-Saïd and Guessoum (2008), Horrigue and Ould-Saïd (2011) and Horrigue and Ould-Saïd (2014) for the conditional quantile estimation when regressors are functional. In addition, Helal and Ould-Saïd (2016) used the same model with truncated data.

In this work, we focus on the prediction problem in models with incomplete data lifetimes (randomly censored on the right). Beyond the historical origins of the statistical analysis of lifetime, demographic, and actuarial life, the three main areas of current analysis of survival data are reliability, biostatistics, and economics.

Generally, we can say that censored regression models have received a great deal of attention in both theoretical and applied statistics literature. In this work, we give an alternative approach to traditional estimation models by considering the minimization of the least relative error for regressions models when the data are randomly right-censored, then we establish the asymptotic properties of the kernel estimator of the functional RER. More precisely, we define the relative error estimator for both complete and incomplete data (randomly censored data), and establish the strong almost complete convergence (a.co.), with rate, and obtain the asymptotic normality of this estimator.

The rest of the paper is organized as follows. Section 2.2 deal with nonparametric relative regression and regression under random censorship. Section 2.3 summarizes the assumptions and the main results, while Section 2.4 is devoted to our simulations results. The proofs of the auxiliary results given in Appendix 2.5.

2.2 Model

2.2.1 Nonparametric relative regression

Let $\{Z_i = (X_i, Y_i)_{1 \leq i \leq n}\}$ be n independent pairs, identically distributed as $Z = (X, Y)$ and valued in $\mathcal{F} \times \mathbb{R}$, where $(\mathcal{F}, d)$ is a semi-metric space (i.e. X is a functional random variable (f.r.v) and d a semi-metric). Let x and y be a fixed element of $\mathcal{F}$ and $\mathbb{R}$ respectively.

By simple algebra, we explicitly determine a kernel-based estimator of the equation (2.2) by

$$\hat{r}(x) = \frac{\sum_{i=1}^{n} Y_i K\left(\frac{d(x,X_i)}{h}\right)}{\sum_{i=1}^{n} K\left(\frac{d(x,X_i)}{h}\right)}, \tag{2.4}$$

where K is an asymmetrical kernel and h (depending on n) is a strictly positive real number.

Note that the formula in (2.4) is a functional extension of the familiar Nadaraya-Watson estimate. The main change comes from the semi-metric d, which measures the proximity between functional objects. However, the use of the previous loss function as a measure of prediction performance may be not suitable in some situations. In particular, the presence of outliers can lead to unreasonable results since all variables have the same weight.

It is clear that the criterion in (2.3) is a more meaningful measure of prediction performance than the least squares error, in particular, when the range of predicted values is large. Moreover, the solution of (2.3) can be explicitly expressed by the ratio of first two conditional inverse moments of Y given X, i.e.

$$r(.) = \frac{\mathbb{E}[Y^{-1}|X = .]}{\mathbb{E}[Y^{-2}|X = .]} =: \frac{g_1(.)}{g_2(.)}. \tag{2.5}$$

Now, we estimate the regression function $r(.)$ under our relative loss function as

$$\hat{r}(x) = \frac{\sum_{i=1}^{n} Y_i^{-1} K\left(\frac{d(x,X_i)}{h}\right)}{\sum_{i=1}^{n} Y_i^{-2} K\left(\frac{d(x,X_i)}{h}\right)} =: \frac{\hat{g}_1(x)}{\hat{g}_2(x)}, \tag{2.6}$$

where

$$\hat{g}_l(x) = \frac{1}{n\mathbb{E}\left[K\left(\frac{d(x,X_1)}{h}\right)\right]} \sum_{i=1}^{n} Y_i^{-l} K\left(\frac{d(x,X_i)}{h}\right) \qquad \text{for} \quad l = 1, 2.$$

2.2.2 Relative error regression under random censorship

In the censoring case, instead of observing the lifetimes Y (which has a continuous distribution function (df) H) we observe the censored lifetimes of items. That is, assuming that $(C_i)_{1\leq i\leq n}$ is a sequence of i.i.d. censoring random variable (r.v.) with common unknown continuous df G. Then, in the right censorship model, we only observe the n pairs (T_i, δ_i) with

$$T_i = Y_i \wedge C_i \quad and \quad \delta_i = \mathbb{1}_{\{Y_i \leq C_i\}}, 1 \leq i \leq n, \tag{2.7}$$

where $\mathbb{1}_A$ denotes the indicator function of the set A.

Now, we assume that $(C_i)_{1\leq i\leq n}$ and $(X_i, Y_i)_{1\leq i\leq n}$ are independent. In censorship model, only the $(X_i, T_i, \delta_i)_{1\leq i\leq n}$ are observed. For any df L, we will write $\tau_L = sup\{t : \bar{L}(t) > 0\}$, where $\bar{L}(.) = 1 - L(.)$ On the other hand, $L_n(.)$ will denote a functional estimator of $L(.)$. Denote by $\widetilde{r}(x)$ the estimator of $r(x)$ in presence of censored data. Then,

$$\widetilde{r}(x) = \frac{\sum_{i=1}^{n} \frac{\delta_i T_i^{-1}}{\bar{G}(T_i)} K\left(\frac{d(x, X_i)}{h}\right)}{\sum_{i=1}^{n} \frac{\delta_i T_i^{-2}}{\bar{G}(T_i)} K\left(\frac{d(x, X_i)}{h}\right)} =: \frac{\widetilde{g}_1(x)}{\widetilde{g}_2(x)}, \tag{2.8}$$

where

$$\widetilde{g}_l(x) = \frac{\sum_{i=1}^{n} \frac{\delta_i T_i^{-l}}{\bar{G}(T_i)} K\left(\frac{d(x, X_i)}{h}\right)}{n\mathbb{E}\left[K\left(\frac{d(x,X_1)}{h}\right)\right]}, \qquad \text{for} \quad l = 1, 2$$

In practice, G is unknown. So, we use the Kaplan-Meier estimator in Kaplan et al. (1958) of $\bar{G}$ given by

$$\bar{G}_n(t) = \begin{cases} \prod_{i=1}^{n}\left(1 - \frac{1-\delta_{(i)}}{n-i+1}\right)^{\mathbb{1}_{\{T_{(i)}\leq t\}}} & \text{if } t \leq T_{(n)} \\ 0 & \text{otherwise,} \end{cases} \tag{2.9}$$

where $T_{(1)} \leq T_{(2)} \leq ... \leq T_{(n)}$ are the order statistics of $(T_i)_{1\leq i\leq n}$ and $\delta_{(i)}$ is the concomitant of $T_{(i)}$. Therefore, the estimator of $r(x)$ is given by

$$\widetilde{r}_n(x) = \frac{\sum_{i=1}^{n} \frac{\delta_i T_i^{-1}}{\bar{G}_n(T_i)} K\left(\frac{d(x, X_i)}{h}\right)}{\sum_{i=1}^{n} \frac{\delta_i T_i^{-2}}{\bar{G}_n(T_i)} K\left(\frac{d(x, X_i)}{h}\right)} =: \frac{\widetilde{g}_{1,n}(x)}{\widetilde{g}_{2,n}(x)}, \tag{2.10}$$

where

$$\widetilde{g}_{l,n}(x) = \frac{1}{n\mathbb{E}\left[K\left(\frac{d(x,X_1)}{h}\right)\right]} \sum_{i=1}^{n} \frac{\delta_i T_i^{-l}}{\bar{G}_n(T_i)} K\left(\frac{d(x, X_i)}{h}\right) \qquad \text{for} \quad l = 1, 2.$$

2.3 Assumptions and main results

The main purpose of this section is to study the uniform almost-complete convergence[1](a.co.) of $\widetilde{r}_n(x)$ toward $r(x)$.

From now on, for all x in $\mathcal{F}$, for all positive real h, and denote by $\mathcal{N}_x$ the neighborhood of the point x, when no confusion is possible, we will denote by c and c' generic constants and define $K_i(x)$ by

$$K_i(x) = K\left(\frac{d(x, X_i)}{h}\right) \quad \text{for} \;\; i = 1, ..., n,$$

where K is a kernel function and $h := h_{n,K}$ is a sequence of positive numbers decreasing toward 0. We will also use the notation

$$\varphi_x(h) = \mathbb{P}(X \in B(x, h)), \tag{2.11}$$

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

We recall the definition of the Kolmogorov's entropy which is an important tool to obtain uniform convergence results. Given a subset $S_{\mathcal{F}} \subset S$ and $\varepsilon > 0$, denote $N_\varepsilon(S)$ or N the minimal number of open balls of radius ε needed to cover S. Then, the quantity $\psi_{S_{\mathcal{F}}} = \log(N)$ is called Kolmogorov's ε-entropy of the set S. In what follows, we will need the following assumptions:

(H1) $\mathbb{P}(X \in B(x, h)) =: \varphi_x(h) > 0$ for all $h > 0$ and $\lim_{h \to 0} \varphi_x(h) = 0$.

(H2) For all $(x_1, x_2) \in \mathcal{N}_x^2$, we have

$$|g_l(x_1) - g_l(x_2)| \leq c d^{k_l}(x_1, x_2) \quad for \;\; k_l > 0.$$

(H3) The kernel K is a bounded Lipshitzian and differentiable function on its support $(0; 1)$ and satisfying:

$$0 < c \leq K(.) \leq c' < +\infty,$$

and its first derivative function K' is such that: $-\infty < c < K'(.) < c' < 0$.

(H4) The bandwidth h satisfies:

(i) $\sqrt{\frac{\log \log n}{n}} = o(\varphi_x(h))$;

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

(ii) $\frac{n\varphi_x(h)}{\log n} \to \infty \ \ as \ \ n \to \infty.$

(H5) The response variable Y is such that: $|Y| > c > 0$ for all $x \in \mathcal{F}$ and

$$\inf_{x\in\mathcal{F}} g_2(x) \geq \gamma > 0.$$

(H6) The functions φ_x and $\psi_{S_\mathcal{F}}$ are such that:

(H6a) there exists $\eta_0 > 0$ such that for all $\eta < \eta_0$, $\varphi'_x(\eta) < c$, where φ'_x denotes the first derivative function of φ_x.

(H6b) for a large enough integer n, we have:

$$\frac{(\log n)^2}{n\varphi_x(h)} < \psi_{S_\mathcal{F}}\left(\frac{\log n}{n}\right) < \frac{n\varphi_x(h)}{\log n},$$

(H6c) the Kolmogorov's ε-entropy of $S_\mathcal{F}$ satisfies:

$$\sum_{n=1}^{\infty} \exp\left[(1-\beta)\psi_{S_\mathcal{F}}\left(\frac{\log n}{n}\right)\right] < \infty \text{ for some } \beta > 1.$$

(N1) There exists a function $\chi_x(.)$ such that:

$$\text{for all } s \in [0,1], \quad \lim_{r\to 0} \frac{\varphi_x(sr)}{\varphi_x(r)} = \chi_x(s)$$

(N2) For $\gamma \in \{1,2\}$, the functions $\Psi_\gamma(.) = \mathbb{E}\left[g_\gamma(X) - g_\gamma(x)|d(x,X) = .\right]$ are derivable at 0.

(N3) The small ball probability satisfies: $n\varphi_x(h) \longrightarrow \infty$.

(N4) For $m \in \{1,2,3,4\}$, the functions $q_m(.) = \mathbb{E}\left[G^{-1}Y^{-m}|X = .\right]$ are continuous in a neighborhood of x.

Comments on the hypotheses: All these conditions are very standard and usually assumed in this context of nonparametric functional estimation. Specifically, Assumption **(H1)** is classic in the asymptotic theory of nonparametric functional statistic, which was linked to the functional structure of the functional covariate. Such a function can be explicitly obtained for several continuous processes (see Ferraty and Vieu (2006)).

The nonparametric aspects of our model are ensured by means of Assumptions **(H2)**, **(N1)** and **(N4)**. These hypotheses are regularity conditions which characterize the functional space of our model and are needed to evaluate the bias term in the asymptotic properties. Assumptions **(H3)** and **(H4)** concern the kernel $K(.)$ and the smoothing parameter h and **(H5)** is a technical condition for getting the proof of our results, while

Hypothesis **(H6)** deals with topological considerations by controlling the entropy of $S_{\mathcal{F}}$. For a radius not too large, one requires that $\psi_{S_{\mathcal{F}}}\left(\frac{\log n}{n}\right)$ is not too small and not too large. Moreover, **(H6)b** implies that $\frac{\psi_{S_{\mathcal{F}}}\left(\frac{\log n}{n}\right)}{n\varphi_x(h)}$ tends to 0 when n tends to $+\infty$. In a different way, Assumption **(H6c)** acts on Kolmogorov's ϵ-entropy of $S_{\mathcal{F}}$. Similarly to the concentration property, this additional argument also controls the contribution of the topological structure of $\mathcal{F}$ in the uniform convergence rate. The Assumptions **(N3)** and **(N4)** are very similar to those used by Ferraty et al. (2007).

Now we are in a position to give our main result.

Theorem 2.3.1. *Under Assumptions **(H1)-(H6)**, we have*

$$\sup_{x\in\mathcal{F}}|\widetilde{r}_n(x)-r(x)| = O_{a.co.}\left(h^{k_1}\right)+O_{a.co.}\left(h^{k_2}\right)+O_{a.co.}\left(\sqrt{\frac{\psi_{S_{\mathcal{F}}}\left(\frac{\log n}{n}\right)}{n\varphi_x(h)}}\right). \tag{2.12}$$

Theorem 2.3.2. *Under Assumptions **(H1),(H3)**and **(N1)-(N4)**, for any $x\in\mathcal{F}$, we have:*

$$\left(\frac{n\varphi_x(h)}{\sigma^2(x)}\right)^{1/2}\left(\widetilde{r}_n(x)-r(x)-B_n(x)-o(h)\right)\xrightarrow{\mathcal{D}}\mathcal{N}(0,1) \quad as \quad n\to\infty, \tag{2.13}$$

where $\xrightarrow{\mathcal{D}}$ means the convergence in distribution. Also,

$$B_n(x)=\frac{\left(\Psi_1'(0)-r(x)\Psi_2'(0)\right)\beta_0}{\beta_1 g_2(x)} \tag{2.14}$$

and

$$\sigma^2(x)=\frac{\left(q_2(x)-2r(x)q_3(x)+r^2(x)q_4(x)\right)\beta_2}{\beta_1^2} \tag{2.15}$$

with

$$\beta_0=K(1)-\int_0^1(sK(s))'\chi_x(s)\mathrm{d}s \quad and \quad \beta_j=K^j(1)-\int_0^1\left(K^j\right)'(s)\chi_x(s)\mathrm{d}s$$

for $j=1,2$.

2.3.1 Proofs of Theorem 2.3.1

From (2.12), we can see that:

$$
\begin{aligned}
\sup_{x\in\mathcal{F}}|\widetilde{r}_n(x)-r(x)| \leq & \sup_{x\in\mathcal{F}}\left\{\left|\frac{\widetilde{g}_{1,n}(x)}{\widetilde{g}_{2,n}(x)}-\frac{\widetilde{g}_1(x)}{\widetilde{g}_{2,n}(x)}\right|+\left|\frac{\widetilde{g}_1(x)}{\widetilde{g}_{2,n}(x)}-\frac{\mathbb{E}(\widetilde{g}_1(x))}{\widetilde{g}_{2,n}(x)}\right|\right.\\
& + \left.\left|\frac{\mathbb{E}(\widetilde{g}_1(x))}{\widetilde{g}_{2,n}(x)}-\frac{g_1(x)}{\widetilde{g}_{2,n}(x)}\right|+\left|\frac{g_1(x)}{\widetilde{g}_{2,n}(x)}-\frac{g_1(x)}{g_2(x)}\right|\right\}\\
\leq & \frac{1}{\inf_{x\in\mathcal{F}}|\widetilde{g}_{2,n}(x)|}\left\{\sup_{x\in\mathcal{F}}|\widetilde{g}_{1,n}(x)-\widetilde{g}_1(x)|+\sup_{x\in\mathcal{F}}|\widetilde{g}_1(x)-\mathbb{E}(\widetilde{g}_1(x))|\right.\\
& + \left.\sup_{x\in\mathcal{F}}|\mathbb{E}(\widetilde{g}_1(x))-g_1(x)|\right\}+\frac{\sup_{x\in\mathcal{F}}|g_1(x)|\gamma^{-1}}{\inf_{x\in\mathcal{F}}|\widetilde{g}_{2,n}(x)|}\left\{\sup_{x\in\mathcal{F}}|\widetilde{g}_{2,n}(x)-\widetilde{g}_2(x)|\right.\\
& + \left.\sup_{x\in\mathcal{F}}|\widetilde{g}_2(x)-\mathbb{E}(\widetilde{g}_2(x))|+\sup_{x\in\mathcal{F}}|\mathbb{E}(\widetilde{g}_2(x))-g_2(x)|\right\}.
\end{aligned}
$$

Therefore, Theorem 2.3.1's result is a consequence of the following intermediate results, where their proofs are postponed to the appendix.

Lemma 2.3.1. *Under assumptions **(H2)-(H5)**, we have*

$$\sup_{x\in\mathcal{F}}|\widetilde{g}_{l,n}(x)-\widetilde{g}_l(x)| = O_{a.s.}\left(\sqrt{\frac{\log\log n}{n}}\right),\ \textit{with } l\in\{1,2\}. \tag{2.16}$$

Where $O_{a.s.}$ means the rate of the almost sure convergence.

Lemma 2.3.2. *Under assumptions **(H1)-(H3)** and **(H5)**, we have*

$$\sup_{x\in\mathcal{F}}|\mathbb{E}(\widetilde{g}_l(x))-g_l(x)| = O\left(h^{k_l}\right), \tag{2.17}$$

withe $l\in\{1,2\}$.

Lemma 2.3.3. *Under assumptions **(H1)-(H3)** and **(H6)**, we have*

$$\sup_{x\in\mathcal{F}}|\widetilde{g}_l(x)-\mathbb{E}(\widetilde{g}_l(x))| = O_{a.co.}\left(\sqrt{\frac{\psi_{S_\mathcal{F}}\left(\frac{\log n}{n}\right)}{n\varphi_x(h)}}\right), \tag{2.18}$$

withe $l\in\{1,2\}$.

Corollary 2.3.1. *Under the assumptions of lemma 2.3.2 and 2.3.3, we obtain:*

$$\textit{there exists } \delta>0 \textit{ such that } \sum_{n=1}^{\infty}\mathbb{P}\left(\inf_{x\in\mathcal{F}}|\widetilde{g}_{2,n}(x)|<\delta\right)<\infty. \tag{2.19}$$

2.3.2 Proofs of Theorem 2.3.2

To prove Theorem 2.3.2, we use the following decomposition

$$\widetilde{r}_n(x) - r(x) = \frac{1}{\widetilde{g}_{2,n}(x)} \left[D_n + A_n\left(\widetilde{g}_{2,n}(x) - \mathbb{E}\left[\widetilde{g}_2(x)\right]\right)\right] + A_n,$$

where

$$A_n = \frac{1}{\mathbb{E}\left[\widetilde{g}_2(x)\right] g_2(x)} \left[\mathbb{E}\left[\widetilde{g}_1(x)\right] g_2(x) - \mathbb{E}\left[\widetilde{g}_2(x)\right] g_1(x)\right]$$

and

$$D_n = \frac{1}{g_2(x)} \left[\left[\widetilde{g}_{1,n}(x) - \mathbb{E}\left[\widetilde{g}_1(x)\right]\right] g_2(x) + \left[\mathbb{E}\left[\widetilde{g}_2(x)\right] - \widetilde{g}_{2,n}(x)\right] g_1(x)\right].$$

Consequently, the proof of Theorem 2.3.2 can be deduced from the convergence rate of Lemma 2.3.1 and the following intermediate results (cf. Lemmas 2.3.4, 2.3.5 and 2.3.6).

Lemma 2.3.4. *Under assumptions of Theorem 2.3.2, we have*

$$\left(\frac{n\varphi_x(h)}{g_2^2(x)\sigma^2(x)}\right)^{1/2} \left(\left[\widetilde{g}_1(x) - \mathbb{E}\left[\widetilde{g}_1(x)\right]\right] g_2(x) - \left[\widetilde{g}_2(x) - \mathbb{E}\left[\widetilde{g}_2(x)\right]\right] g_1(x)\right) \xrightarrow{\mathcal{D}} \mathcal{N}(0,1).$$

Lemma 2.3.5. *Under assumptions of Theorem 2.3.2, we obtain*

$$A_n = hB_n + o(h).$$

Lemma 2.3.6. *Under assumptions of Theorem 2.3.2, we obtain*

$$\widetilde{g}_2(x) \to g_2(x), \textit{in probability}$$

and

$$\left(\frac{n\varphi_x(h)}{g_2^2(x)\sigma^2(x)}\right)^{1/2} A_n\left(\widetilde{g}_2(x) - \mathbb{E}[\widetilde{g}_2(x)]\right) \to 0, \textit{in probability}.$$

2.4 Simulation study

In order to see the behavior of our proposed estimator, we consider the curves generated in the following way:

$$X_i(t) = a_i \sin(4(b_i - t)) + b_i + \eta_{i,t} \quad i = 1:200 \quad t \in [0,1[,$$

where $a_i \sim \mathcal{N}(5,2)$, $b_i \sim \mathcal{N}(0, \ 0.1)$ and $\eta_{i,t} \sim \mathcal{N}(0, \ 0.2)$. All the curves are discretized on the same grid generated from $m = 150$ equispaced points $t \in [0,1[$. The observations

Y_i's for $i = 1, ..., n$ are generated from the model

$$Y_i = r(X_i) + \epsilon_i \quad where \ \epsilon_i \sim \mathcal{N}(0, \ 0.01),$$

where

$$r(x) = \int_0^1 \frac{dt}{1 + |x(t)|}.$$

In practice, the semi-metric choice is based on the regularity of the curves which are under study. In our case, regarding the shape of the curves X_i, it is clear that the PCA-type semi-metric (cf. Ferraty and Vieu (2006)) is well adapted to this data set. It should also be noticed that the best results concerning prediction are obtained for $q = 4$ (the number of components in the PCA-type semi-metric). The optimal bandwidth h is chosen by the cross-validation method for the k nearest neighbors (kNN) in a local way.
We select the quadratic kernel for both classic and relative estimators defined by

$$K(u) = \frac{3}{2}(1 - u^2)\mathbb{1}_{(0,1)}.$$

Next, we consider a sample of size $n = 200$ and we split the data generated from the model above into two subsets: a training sample $(X_i, T_i, \delta_i), i = 1, ..., 150$ and a test sample $(X_i, T_i, \delta_i), i = 151, ..., 200$. Then, we calculate the estimator $\hat{\theta}(X_i)$ for any $i \in \{151, ..., 200\}$.

We also, simulate n i.i.d. rv's C_i, $i = 1, ..., n$ with law $\mathcal{E}(\lambda)$ (the exponential law with the λ parameter that controls the censorship rate).

2.4.1 Prediction regression

The performance of both estimators was compared under the mean squared prediction error (MSE) criterion:

$$MSE = \frac{1}{50} \sum_{j=151}^{200} (\theta(X_j) - \hat{\theta}(X_j))^2,$$

where $\hat{\theta}(X_j)$ means the estimator of both regression models and $\theta(X_j)$ the response variable.

1) Data without outliers : The obtained results are shown in Figure 2.2. With the censorship rate $CR = 1.33\%$, it is clear that there is no meaningful difference between the two estimation methods: the Classical Kernel Estimator (CKE) and the Relative Error

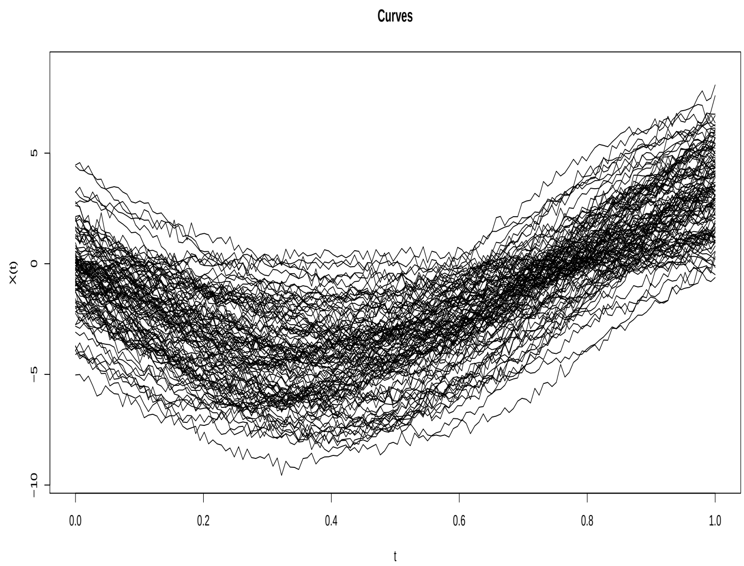

Figure 2.1: The curves $X_{i=1,\ldots,100}(t), \ \ t \in [0,1[$.

Estimator (REE) (MSE_{CKE} = 0.00038, MSE_{REE} = 0.00048).

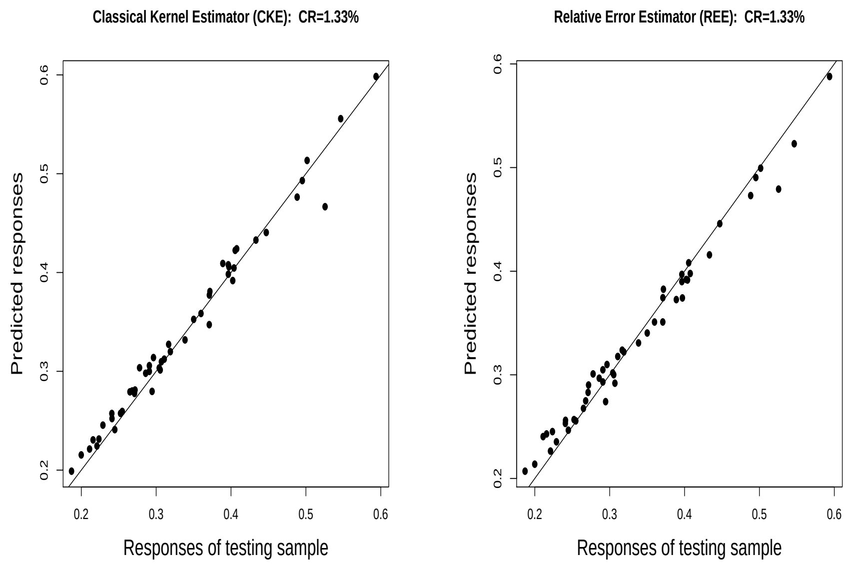

Figure 2.2: Comparison between the Classical Kernel Estimator (CKE) and the Relative Error Estimator (REE) without outliers.

2) Data with outliers : Here, we concentrate on the comparison of both models' performances in the presence of outliers. For this aim, we introduce artificial outliers by multiplying some values of Y in the training sample by 10. The estimators of both models are obtained by the same previous selection methods of the smoothing parameter, i.e., the same metric d and also the same kernel K. Finally, the obtained results are shown in

Table 2.1 and displayed in Figure 2.3. Note that, in Figure 2.2 the two estimators are

Table 2.1: MSE for the Classical Kernel Estimator (CKE) and the Relative Error Estimator (REE) according to numbers of introduced artificial outliers.

Number of artificial outliers	0	10	30	50
Classical Kernel Estimator MSE_{CKE}	0.00076	0.02520	3.98068	434.82333
Relative Error Estimator MSE_{REE}	0.00060	0.00064	0.00072	0.00054

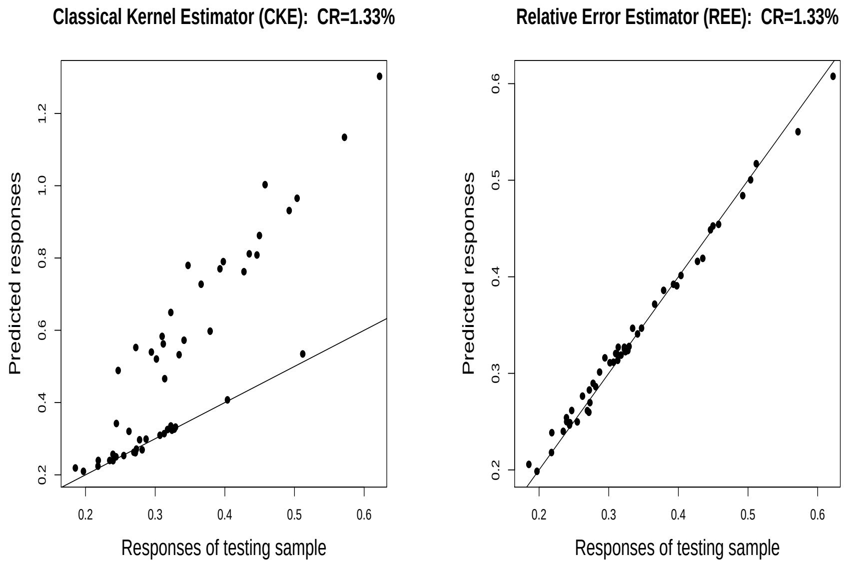

Figure 2.3: Comparison between the Classical Kernel Estimator (CKE) and the Relative Error Estimator (REE) in the presence of outliers.

equivalent but in Figure 2.3, in which we considered the presence of outliers, the relative error regression is robust than the classical kernel regression; i.e., the classical kernel method is susceptible to the presence of outliers. Now, we will study the behavior of our estimator with different censored rates (CR). The results are shown in Table 2.2. We see that the quality of fit is affected and becomes worse as the CR increases, but the relative error estimator is more efficient than the classical one in the presence of censoring data.

Another point of view, and in order to verify the superiority of our methodology, we provide a comparative study with the case of missing data. Note that, in the case of missing data the estimator is given as follow:

$$\widehat{r}(x) = \frac{\sum_{i=1}^{n} \delta_i T_i^{-1} K\left(\frac{d(x, X_i)}{h}\right)}{\sum_{i=1}^{n} \delta_i T_i^{-2} K\left(\frac{d(x, X_i)}{h}\right)}. \tag{2.20}$$

Table 2.2: MSE for the Classical Kernel Estimator (CKE) and the Relative Error Estimator (REE) according to to the censoring rates with different sample size.

Sample size	CR	Classical Kernel Estimator MSE_{CKE}	Relative Error Estimator MSE_{REE}
100	10%	0.00239	0.00203
	20%	0.00613	0.00366
	60%	0.01483	0.00679
200	10%	0.00175	0.00182
	20%	0.00545	0.00292
	60%	0.01155	0.00576
600	10%	0.00100	0.00051
	20%	0.00860	0.00284
	60%	0.01179	0.00408

We obtain the following result resumed in this table:

Table 2.3: MSE for the REE estimator under missing case and the REE estimator under censored case.

Incomplete data	9%	24%	50%	64%
Relative Error Estimator $MSE_{missing}$	0.00628	0.07332	0.11894	0.12236
Relative Error Estimator $MSE_{censoring}$	0.00315	0.00457	0.00644	0.00702

We note that the model with censorship gives better results than the missing model when the censorship rate increases.

2.4.2 Confidence Interval

Based on Theorem 2.3.2, we aim in this subsection to build confidence interval for the true value of $r(x)$ given curve $X = x$. Plug-in estimates for the asymptotic standard deviation $(n\varphi_x(h)/\sigma^2(x))^{1/2}$ and the bias term $hB_n(x) + o(h)$. Then, estimate $q_m(x)$ by

$$\hat{q}_m(x) = \frac{\sum_{i=1}^n K_i \delta_i \bar{G}_n^{-1}(T_i) T_i^{-m}}{\sum_{i=1}^n K_i}.$$

Whereas β_1 and β_2 are empirically estimated by

$$\hat{\beta}_1 = \frac{1}{n\varphi_x(h)} \sum_{i=1}^n K_i \delta_i G_n^{-1}(T_i) \text{ and } \hat{\beta}_2 = \frac{1}{n\varphi_x(h)} \sum_{i=1}^n K_i^2 \delta_i G_n^{-1}(T_i) \tag{2.21}$$

Finally, the practical estimator of the normalized deviation is

$$\left(\frac{\left(\sum_{i=1}^{n} K_i\delta_i G_n^{-1}(T_i)\right)^2 \widetilde{q}_2^2(x)}{\left(\sum_{i=1}^{n} K_i^2\delta_i G_n^{-1}(T_i)\right)\left(\widetilde{q}_2(x) - 2\widetilde{r}(x)\widetilde{q}_3(x) + \widetilde{r}^2(x)\widetilde{q}_4(x)\right)}\right)^{1/2}.$$

We point out that the function $\varphi_x(.)$ in (2.21) does not intervene in the calculation of the confidence interval.Finally, the approximated $1-\zeta/2$ confidence interval for $r(x)$, for any $x \in \mathcal{F}$, is

$$[b_{inf}, b_{sup}],$$

where

$$b_{inf} = \widetilde{r}_n(x) - t_{1-\zeta/2}\left(\frac{\left(\sum_{i=1}^{n} K_i\delta_i G_n^{-1}(T_i)\right)^2 \widetilde{q}_2^2(x)}{\left(\sum_{i=1}^{n} K_i^2\delta_i G_n^{-1}(T_i)\right)\left(\widetilde{q}_2(x) - 2\widetilde{r}(x)\widetilde{q}_3(x) + \widetilde{r}^2(x)\widetilde{q}_4(x)\right)}\right)^{1/2}$$

and

$$b_{sup} = \widetilde{r}_n(x) + t_{1-\zeta/2}\left(\frac{\left(\sum_{i=1}^{n} K_i\delta_i G_n^{-1}(T_i)\right)^2 \widetilde{q}_2^2(x)}{\left(\sum_{i=1}^{n} K_i^2\delta_i G_n^{-1}(T_i)\right)\left(\widetilde{q}_2(x) - 2\widetilde{r}(x)\widetilde{q}_3(x) + \widetilde{r}^2(x)\widetilde{q}_4(x)\right)}\right)^{1/2},$$

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

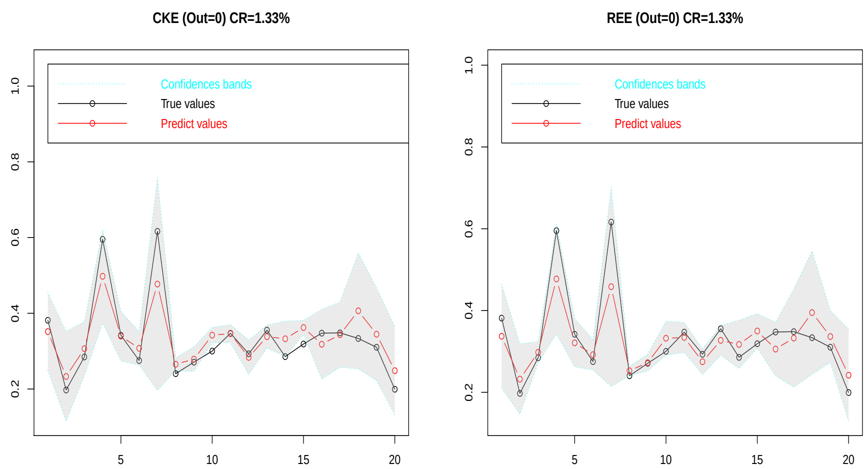

Figure 2.4: Extremities of the predicted values versus the true values and the confidence bands (simulation data without outliers).

We see clearly in Figure 2.5 that our predicted values of REE fit very well the real values and the latter are all in the confidence interval in the presence of outliers.

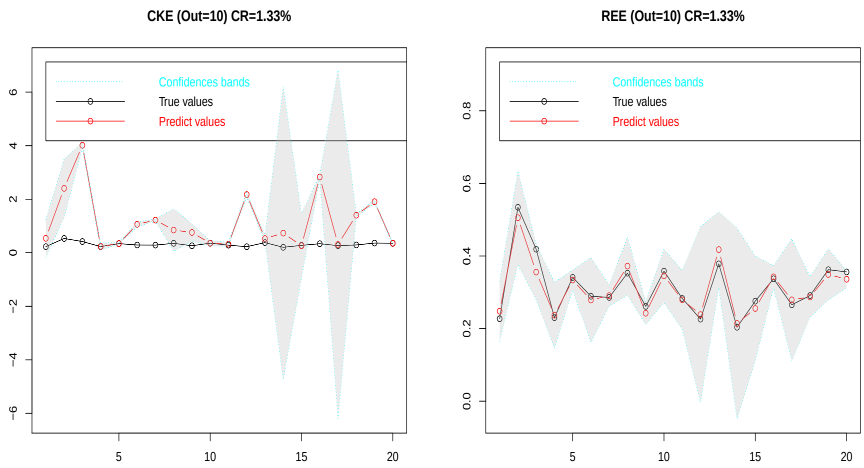

Figure 2.5: Extremities of the predicted values versus the true values and the confidence bands (simulation data with 10 outliers).

2.5 Appendix

Proof of Lemma 2.3.1:

For all $l = 1, 2$, we have:

$$\begin{aligned} |\widetilde{g}_{l,n}(x) - \widetilde{g}_l(x)| &\leq \frac{1}{\mathbb{E}(K_1(x))} \sum_{i=1}^{n} \left| \frac{\delta_i T_i^{-l}}{\bar{G}_n(T_i)} K_i(x) - \frac{\delta_i T_i^{-l}}{\bar{G}(T_i)} K_i(x) \right| \\ &\leq \frac{\sup_{t \in \mathbb{R}} |\bar{G}_n(t) - \bar{G}(t)|}{\bar{G}_n(\tau_H)\bar{G}(\tau_H)} \frac{\sum_{i=1}^{n} |Y_i^{-l} K_i(x)|}{n\mathbb{E}(K_1(x))}. \end{aligned}$$

Since $\bar{G}(\tau_H) > 0$, in conjunction with the SLLN and the LIL on the censoring law (see Formula (4.28) in Deheuvels and Einmahl (2000)), we have:

$$\sup_{x \in \mathcal{F}} |\widetilde{g}_{l,n}(x) - \widetilde{g}_l(x)| \leq \frac{\sum_{i=1}^{n} |Y_i^{-l} K_i(x)|}{n\mathbb{E}(K_1(x))} \frac{1}{\bar{G}^2(\tau_H)} \sqrt{\frac{\log \log n}{n}}.$$

(2.3) allows to achieve the proof. □

Proof of Lemma 2.3.2:

Since $(X_1, Y_1), \ldots, (X_n, Y_n)$ are independent and identically distributed pairs, and for all

$l = 1, 2$, we have:

$$\begin{aligned}\forall x \in \mathcal{F}, \quad |\mathbb{E}(\widetilde{g}_l(x)) - g_l(x)| &= \left|\mathbb{E}\left(\frac{K_1(x)}{\mathbb{E}(K_1(x))}\mathbb{E}\left[\frac{\mathbb{E}(\mathbb{1}_{Y_1 \le C_i}|Y_1)Y_1^{-l}}{\bar{G}(Y_1)}|X_1\right]\right) - g_l(x)\right| \\ &= \frac{1}{\mathbb{E}(K_1(x))}\left|\mathbb{E}\left\{\left[\mathbb{E}(Y_1^{-l}|X_1) - g_l(x)\right]\mathbb{1}_{B(x,h)}(X_1)K_1(x)\right\}\right|.\end{aligned}$$

Then, by the Hölder assumption (**H2**) we get that

$$\forall x \in \mathcal{F}, \quad |g_l(X_1) - g_l(x)| \le ch^{k_l}.$$

Thus,

$$\sup_{x \in \mathcal{F}} |\mathbb{E}(\widetilde{g}_l(x)) - g_l(x)| \le ch^{k_l}.$$

□

Proof of Lemma 2.3.3:
The proof of this Lemma is based on the exponential inequality given in Corollary A.8.ii in Ferraty and Vieu (2006) with

$$Z_{i,l} = \frac{1}{\mathbb{E}(K_1(x))}\left[\frac{\delta_i T_i^{-l}}{\bar{G}(T_i)}K_i(x) - \mathbb{E}\left(\frac{\delta_i T_i^{-l}}{\bar{G}(T_i)}K_i(x)\right)\right]. \tag{2.22}$$

To do that, we have to show that:

$$\exists c > 0, \forall m \ge 2, \quad \mathbb{E}(|Z_{1,l}^m|) = c\varphi_x^{-m+1}(h). \tag{2.23}$$

First, we prove for $m \ge 2$ that:

$$\frac{1}{\mathbb{E}^m(K_1(x))}\mathbb{E}\left[\left|\frac{\delta_1 T_1^{-lm}}{\bar{G}^m(T_1)}K_1^m(x)\right|\right] = O(\varphi_x^{-m+1}(h)). \tag{2.24}$$

Then, we write:

$$\begin{aligned}\mathbb{E}\left[\left|\frac{\delta_1 T_1^{-lm}}{\bar{G}^m(T_1)}K_1^m(x)\right|\right] &= \mathbb{E}\left[\left|\frac{\mathbb{1}_{\{Y_1 \le C_1\}}Y_1^{-lm}}{\bar{G}^m(Y_1)}K_1^m(x)\right|\right] \\ &= \mathbb{E}\left[\mathbb{E}\left(\frac{\mathbb{E}(\mathbb{1}_{\{Y_1 \le C_1\}}|Y_1)|Y_1^{-lm}|}{\bar{G}^m(Y_1)}|X_1\right)K_1^m(x)\right] \\ &= \mathbb{E}\left[\mathbb{E}\left(\frac{|Y_1^{-lm}|}{\bar{G}^{m-1}(Y_1)}|X_1\right)K_1^m(x)\right] \\ &\le \frac{c}{\bar{G}(\tau_H)}\mathbb{E}(K_1^m(x)) \\ &\le c\varphi_x(h)\end{aligned}$$

which implies that

$$\frac{1}{\mathbb{E}^m(K_1(x))}\mathbb{E}\left[\left|\frac{\delta_1 T_1^{-lm}}{\bar{G}^m(T_1)}K_1^m(x)\right|\right] = O(\varphi_x^{-m+1}(h))$$

and

$$\frac{1}{\mathbb{E}(K_1(x))}\mathbb{E}\left[\left|\frac{\delta_1 T_1^{-l}}{\bar{G}(T_1)}K_1(x)\right|\right] \leq c.$$

Next, by the Newton's binomial expansion we obtain:

$$\begin{aligned}
\mathbb{E}(|Z_{1,l}^m|) &\leq c\sum_{k=0}^{m}\frac{\mathbb{E}\left[\left|\frac{\delta_1 T_1^{-lk}}{\bar{G}^k(T_1)}K_1^k(x)\right|\right]}{\mathbb{E}^k(K_1(x))}\left[\frac{\mathbb{E}\left[\left|\frac{\delta_1 T_1^{-l}}{\bar{G}(T_1)}K_1(x)\right|\right]}{\mathbb{E}(K_1(x))}\right]^{m-k} \\
&\leq c\max_{k=0,\dots,m}\varphi_x^{-k+1}(h) \\
&\leq c\varphi_x^{-m+1}(h).
\end{aligned}$$

It follows that:

$$\mathbb{E}(|Z_{1,l}^m|) = O(\varphi_x^{-m+1}(h)). \tag{2.25}$$

Thus, we apply the exponential inequality given in Corollary A.8.ii in Ferraty and Vieu (2006) with $a^2 = \varphi_x^{-1}(h)$ and obtain

$$\frac{1}{n}\sum_{i=1}^{n}Z_{1,l} = O_{a.co.}\left(\frac{\log n}{n\varphi_x(h)}\right).$$

Let $x_1, \dots, x_N$ be a finite set of points in $\mathcal{F}$ and $S_{\mathcal{F}} \subset \mathcal{F}$ such that

$$S_{\mathcal{F}} = \bigcup_{k=1}^{N} B(x_k, \varepsilon) \quad with \ \ \varepsilon = \frac{\log n}{n}.$$

For all $x \in S_{\mathcal{F}}$, we set $k(x) = \arg\min_{k\in\{1,2,\dots,N\}} d(x, x_k)$.

Then, we have the following decomposition:

$$\begin{aligned}
\sup_{x\in S_{\mathcal{F}}}|\widetilde{g}_l(x) - \mathbb{E}(\widetilde{g}_l(x))| \leq \quad & \underbrace{\sup_{x\in S_{\mathcal{F}}}\left|\widetilde{g}_l(x) - \widetilde{g}_l(x_{k(x)})\right|}_{F_1} + \underbrace{\max_{k\in\{1,2,\dots,N\}}\left|\widetilde{g}_l(x_{k(x)}) - \mathbb{E}(\widetilde{g}_l(x_{k(x)}))\right|}_{F_2} \\
& + \underbrace{\sup_{x\in S_{\mathcal{F}}}\left|\mathbb{E}(\widetilde{g}_l(x_{k(x)})) - \mathbb{E}(\widetilde{g}_l(x))\right|}_{F_3}.
\end{aligned}$$

We would have demonstrated the expected result if we show that

$$F_i \to 0 \ \ for \ \ i = 1, 2, 3.$$

For the term F_1, a direct consequence of the assumption **(H3)** is that:

$$c\varphi_x(h) \leq \mathbb{E}\left(K\left(\frac{d(x,X)}{h}\right)\right) \leq c'\varphi_x(h).$$

Therefore,

$$\begin{aligned} F_1 &\leq \sup_{x\in S_{\mathcal{F}}} \frac{1}{n}\sum_{i=1}^{n}\left|\frac{\delta_i T_i^{-l}}{\bar{G}(T_i)\mathbb{E}(K_1(x))}K_i(x) - \frac{\delta_i T_i^{-l}}{\bar{G}(T_i)\mathbb{E}(K_1(x_{k(x)}))}K_i(x_{k(x)})\right| \\ &\leq \frac{c}{\varphi_x(h)}\sup_{x\in S_{\mathcal{F}}}\frac{1}{n}\sum_{i=1}^{n}\frac{\delta_i|T_i^{-l}|}{\bar{G}(T_i)}\left|K_i(x) - K_i(x_{k(x)})\right|\mathbb{1}_{B(x,h)\cup B(x_{k(x)},h)}(X_i), \end{aligned}$$

In this situation K is lipshitzian on $(0,1)$. One has to decompose F_1 into three terms as follows:

$$F_1 \leq c\sup_{x\in S_{\mathcal{F}}}(F_{11} + F_{12} + F_{13}),$$

with

$$\begin{aligned} F_{11} &= \frac{1}{n\varphi_x(h)}\sum_{i=1}^{n}\frac{\delta_i|T_i^{-l}|}{\bar{G}(T_i)}\left|K_i(x) - K_i(x_{k(x)})\right|\mathbb{1}_{B(x,h)\cap B(x_{k(x)},h)}(X_i), \\ F_{12} &= \frac{1}{n\varphi_x(h)}\sum_{i=1}^{n}\frac{\delta_i|T_i^{-l}|}{\bar{G}(T_i)}K_i(x)\mathbb{1}_{B(x,h)\cap \overline{B(x_{k(x)},h)}}(X_i), \\ F_{13} &= \frac{1}{n\varphi_x(h)}\sum_{i=1}^{n}\frac{\delta_i|T_i^{-l}|}{\bar{G}(T_i)}\left|K_i(x) - K_i(x_{k(x)})\right|\mathbb{1}_{\overline{B(x,h)}\cap B(x_{k(x)},h)}(X_i). \end{aligned}$$

Next, we write:

$$F_{11} \leq \frac{c}{n}\sum_{i=1}^{n}U_{i,l} \;\; with \quad U_{i,l} = \frac{\varepsilon}{h\varphi_x(h)}\frac{\delta_i|T_i^{-l}|}{\bar{G}(T_i)}\mathbb{1}_{B(x,h)\cap B(x_{k(x)},h)}(X_i),$$

$$F_{12} \leq \frac{c}{n}\sum_{i=1}^{n}V_{i,l} \;\; with \quad V_{i,l} = \frac{1}{\varphi_x(h)}\frac{\delta_i|T_i^{-l}|}{\bar{G}(T_i)}\mathbb{1}_{B(x,h)\cap \overline{B(x_{k(x)},h)}}(X_i),$$

and

$$F_{13} \leq \frac{c}{n}\sum_{i=1}^{n}W_{i,l} \;\; with \quad W_{i,l} = \frac{1}{\varphi_x(h)}\frac{\delta_i|T_i^{-l}|}{\bar{G}(T_i)}\mathbb{1}_{\overline{B(x,h)}\cap B(x_{k(x)},h)}(X_i).$$

Following the same steps as in $Z_{i,l}$ to obtain the result of F_{11}.

$$\mathbb{E}(|U_{1,l}^m|) \leq \frac{c\varepsilon^m}{h^m\varphi_x(h)^{m-1}}.$$

Also, we apply the inequality of Corollary A.8.ii in Ferraty and Vieu (2006),with $a =$

$\sqrt{\frac{\varepsilon}{h\varphi_x(h)}}$ to get

$$F_{11} = O_{a.co.}\left(\sqrt{\frac{\varepsilon \log n}{nh\varphi_x(h)}}\right).$$

Following the same ideas for studying F_{12} and F_{13} we obtain:

$$F_{12} = O\left(\frac{\varepsilon}{\varphi_x(h)}\right) + O_{a.co.}\left(\sqrt{\frac{\varepsilon \log n}{n\varphi_x(h)^2}}\right)$$

and

$$F_{13} = O\left(\frac{\varepsilon}{\varphi_x(h)}\right) + O_{a.co.}\left(\sqrt{\frac{\varepsilon \log n}{n\varphi_x(h)^2}}\right).$$

To end the proof of F_1, it suffices to put together all the intermediate results and to use again **(H6b)** for determining

$$F_1 = O_{a.co.}\left(\sqrt{\frac{\psi_{S_\mathcal{F}}(\varepsilon)}{n\varphi_x(h)}}\right) \to 0.$$

For F_3, it is clear that $F_3 \leq \mathbb{E}\left(\sup_{x\in S_\mathcal{F}} \left|\widetilde{g}_l(x) - \widetilde{g}_l(x_{k(x)})\right|\right)$ and by following a similar proof to the one used for F_1, it comes

$$F_3 = O_{a.co.}\left(\sqrt{\frac{\psi_{S_\mathcal{F}}(\varepsilon)}{n\varphi_x(h)}}\right) \to 0.$$

Now, we deal with F_2, for all $\eta > 0$, we have

$$\begin{aligned} \mathbb{P}(F_2 > \eta) &= \mathbb{P}\left(\max_{k\in\{1,2,\ldots,N\}} \left|\widetilde{g}_l(x_{k(x)}) - \mathbb{E}(\widetilde{g}_l(x_{k(x)}))\right| > \eta\right) \\ &\leq N \max_{k\in\{1,2,\ldots,N\}} \mathbb{P}\left(\left|\widetilde{g}_l(x_{k(x)}) - \mathbb{E}(\widetilde{g}_l(x_{k(x)}))\right| > \eta\right). \end{aligned}$$

Put $\eta = \eta_0\sqrt{\frac{\psi_{S_\mathcal{F}}(\varepsilon)}{n\varphi_x(h)}}$, and apply the exponential inequality given by Corollary **A.8.*ii*** in Ferraty and Vieu (2006) for

$$\Delta_{l,i} = \frac{1}{\mathbb{E}(K_1(x_k)))}\left[\frac{\delta_i T_i^{-l}}{\bar{G}(T_i)}K_i(x_k) - \mathbb{E}\left(\frac{\delta_i T_i^{-l}}{\bar{G}(T_i)}K_i(x_k)\right)\right].$$

Since $\mathbb{E}[|\Delta_{l,i}|^m] = O(\varphi_x(h)^{-m+1})$, then, we can take $a^2 = \frac{1}{\varphi_x(h)}$. Hence, for all $\eta_0 > 0$:

$$\begin{aligned}\mathbb{P}\left(\left|\widetilde{g}_l(x_{k(x)}) - \mathbb{E}(\widetilde{g}_l(x_{k(x)}))\right| > \eta_0\sqrt{\frac{\psi_{S_\mathcal{F}}(\varepsilon)}{n\varphi_x(h)}}\right) &= \mathbb{P}\left(\frac{1}{n}\left|\sum_{i=1}^n \Delta_{l,i}\right| > \eta_0\sqrt{\frac{\psi_{S_\mathcal{F}}(\varepsilon)}{n\varphi_x(h)}}\right) \\ &\leq 2\exp\left(-c\eta_0^2\psi_{S_\mathcal{F}}(\varepsilon)\right).\end{aligned}$$

By using the fact $\psi_{S_\mathcal{F}}(\varepsilon) = \log N$ and by choosing η_0 such that $c\varepsilon_0^2 = \beta$, we obtain:

$$N \max_{k\in\{1,2,\ldots,N\}} \mathbb{P}\left(\left|\widetilde{g}_l(x_{k(x)}) - \mathbb{E}(\widetilde{g}_l(x_{k(x)}))\right| > \eta_0\sqrt{\frac{\psi_{S_\mathcal{F}}}{n\varphi_x(h)}}\right) \leq cN^{1-\beta}. \tag{2.26}$$

Since $\sum_{N=1}^\infty N^{1-\beta} < \infty$, we obtain that

$$F_2 = O_{a.co.}\left(\sqrt{\frac{\psi_{S_\mathcal{F}}(\varepsilon)}{n\varphi_x(h)}}\right) \to 0.$$

□

Proof of Corollary 2.3.1:

It is easy to remark that:

$\inf_{x\in\mathcal{F}} |\widetilde{g}_{2,n}(x)| \leq \dfrac{g_2(x)}{2}$ implies that there exists $x \in \mathcal{F}$ such that $g_2(x) - \widetilde{g}_{2,n}(x) \geq \frac{g_2(x)}{2}$ which implies that $\sup_{x\in\mathcal{F}} |g_2(x) - \widetilde{g}_{2,n}(x)| \geq \dfrac{g_2(x)}{2}$.

We deduce, from Lemma 2.3.2 and 2.3.3, that

$$\mathbb{P}\left(\inf_{x\in\mathcal{F}} |\widetilde{g}_{2,n}(x)| \leq \frac{g_2(x)}{2}\right) \leq \mathbb{P}\left(\sup_{x\in\mathcal{F}} |g_2(x) - \widetilde{g}_{2,n}(x)| > \frac{g_2(x)}{2}\right)$$

Consequently

$$\sum_{n=1}^\infty \mathbb{P}\left(\inf_{x\in\mathcal{F}} |\widetilde{g}_{2,n}(x)| < \frac{g_2(x)}{2}\right) < \infty.$$

□

Proof of Lemma 2.3.4:

We write

$$\sqrt{n\varphi_x(h)}\left(g_2^2(x)\sigma(x)\right)^{-1}\left[\left(\widetilde{g}_2(x) - \mathbb{E}\left[\widetilde{g}_2(x)\right]\right)g_1(x) - \left(\widetilde{g}_1(x) - \mathbb{E}\left[\widetilde{g}_1(x)\right]\right)g_2(x)\right] = \frac{S_n}{g_2^2(x)\sigma(x)}$$

with $S_n = \sum_{i=1}^{n}(L_i(x) - \mathbb{E}(L_i(x))$, where

$$L_i(x) := \frac{\sqrt{n\varphi_x(h)}}{n\mathbb{E}[K_1]} \frac{\delta_i}{\bar{G}(T_i)} K_i \left(g_1(x)T_i^{-2} - g_2(x)T_i^{-1}\right) \tag{2.27}$$

Thus, to achieve this lemma's proof it suffices to show the asymptotic normality of S_n. This last is reached by applying the Lyapounov central limit Theorem Feller (1966) on $L_i(x)$, i.e., it suffices to show, for some $\kappa > 0$, that:

$$\frac{\sum_{i=1}^{n}\mathbb{E}\left[|L_i(x) - \mathbb{E}[L_i(x)]|^{2+\kappa}\right]}{\left(\text{var}\left[\sum_{i=1}^{n} L_i(x)\right]\right)^{(2+\kappa)/2}} \to 0. \tag{2.28}$$

Clearly

$$\begin{aligned}\text{var}\left[\sum_{i=1}^{n} L_i(x)\right] &= = n\varphi_x(h)\left(\text{var}\left[\widetilde{g}_1(x)\right] g_2^2(x) + \text{var}\left[\widetilde{g}_2(x)\right] g_1^2(x)\right. \\ &\quad \left. -2g_1(x)g_2(x)\,\text{Cov}\left[\widetilde{g}_1(x), \widetilde{g}_2(x)\right]\right).\end{aligned}$$

We have for $l \in \{1, 2\}$:

$$\text{var}(\widetilde{g}_l(x)) = \frac{1}{(n\mathbb{E}[K_1])^2}\sum_{i=1}^{n}\text{var}\left[\frac{\delta_i}{\bar{G}(T_i)}T_i^{-l}K_i\right] = \frac{1}{n(E[K_1])^2}\text{var}\left[\frac{\delta_1}{\bar{G}(T_1)}T_1^{-l}K_1\right]$$

By conditioning on the random variable X, by the same ideas in the proof of lemma 2.3.2 and by using assumptions **(N1)** and **(N3)**, we get:

$$\begin{aligned}\mathbb{E}\left(\left(\frac{\delta_1}{\bar{G}(Y_1)}\right)^2 Y_1^{-2l}K_1^2\right) &= \mathbb{E}\left(\mathbb{E}(\bar{G}^{-1}(Y_1)Y_1^{-2l}|X = x)K_1^2\right) \\ &= \varphi_x(h)\mathbb{E}\left[\bar{G}^{-1}(Y_1)Y_1^{-2l}|X = x\right]\left(K^2(1) - \int_0^1 \left(K^2(s)\right)' \chi_x(u)du\right) \\ &\quad +o(\varphi_x(h))\end{aligned}$$

and

$$\mathbb{E}\left(\frac{\delta_1}{\bar{G}(Y_1)}Y_1^{-l}K_1\right) = O(\varphi_x(h)).$$

Thus,

$$\operatorname{var}\left[\frac{\delta_1}{\bar{G}(T_1)}T_1^{-l}K_1\right] = \varphi_x(h)E\left[\bar{G}^{-1}(Y)Y^{-2l}|X=x\right]\left(K^2(1)-\int_0^1\left(K^2(s)\right)'\chi_x(u)du\right) + o(\varphi_x(h)) + O\left(\varphi_x^2(h)\right).$$

Then, we obtain

$$\operatorname{var}\left[\widetilde{g}_l(x)\right] = \frac{E\left[\bar{G}^{-1}(Y)Y^{-2l}|X=x\right]\left(K^2(1)-\int_0^1\left(K^2(s)\right)'\chi_x(u)du\right)}{n\varphi_x(h)\left(K(1)-\int_0^1 K'(s)\chi_x(s)ds\right)^2} + o\left(\frac{1}{n\varphi_x(h)}\right).$$

Concerning the covariance term, we follow the same steps as for the variance to get

$$\operatorname{Cov}\left[\widetilde{g}_1(x),\widetilde{g}_2(x)\right] = \frac{E\left[\bar{G}^{-1}(Y)Y^{-3}|X=x\right]\left(K^2(1)-\int_0^1\left(K^2(s)\right)'\chi_x(u)du\right)}{n\varphi_x(h)\left(K(1)-\int_0^1 K'(s)\chi_x(s)ds\right)^2} + o\left(\frac{1}{n\varphi_x(h)}\right).$$

Hence,

$$\operatorname{var}\left(\sum_{i=1}^n L_i(x)\right) = g_2^2(x)\sigma^2(x) + o(1).$$

Therefore, to complete the proof of this Lemma, it is enough to show that the numerator of (2.28) converges to 0. To do this, we use the C_r-inequality (cf. Loeve (1963), page 155) to show that:

$$\sum_{i=1}^n \mathbb{E}\left[|L_i(x)-\mathbb{E}\left[L_i(x)\right]|^{2+\kappa}\right] \le c\sum_{i=1}^n \mathbb{E}\left[|L_i(x)|^{2+\kappa}\right] + c'\sum_{i=1}^n |\mathbb{E}\left[L_i(x)\right]|^{2+\kappa} \tag{2.29}$$

Recall that, for all $j>0$, $\mathbb{E}[K_1^j] = O(\varphi_x(h))$, and assumption **(H5)**, we get:

$$\sum_{i=1}^n \mathbb{E}\left[|L_i(x)|^{2+\kappa}\right] \le c\left(n\varphi_x(h)\right)^{-\kappa/2}\left(\mathbb{E}\left[K_1^{2+\kappa}\right]/\varphi_x(h)\right) \to 0.$$

Hence, the second term of (2.29) becomes

$$\sum_{i=1}^n |\mathbb{E}\left[L_i(x)\right]|^{2+\kappa} \le cn^{-\kappa/2}\left(\varphi_x(h)\right)^{1+\kappa/2} \to 0,$$

which completes the proof. □

Proof of Lemma 2.3.5:

By standard arguments we show that:

$$\mathbb{E}[\widetilde{r}_n(x)] = \frac{\mathbb{E}\left[\widetilde{g}_1(x)\right]}{\mathbb{E}\left[\widetilde{g}_2(x)\right]} + O\left(\frac{1}{n\varphi_x(h)}\right).$$

So, it suffices to evaluate $\mathbb{E}[\widetilde{g}_1(x)]$ and $\mathbb{E}[\widetilde{g}_2(x)]$. Indeed, by the same ideas as in the proof of Lemma 2.3.2, we obtain

$$\mathbb{E}\left[\widetilde{g}_l(x)\right] = \frac{1}{\mathbb{E}\left[K_1\right]}\mathbb{E}\left[K_1(x)\mathbb{E}\left[Y_1^{-l}|X_1\right]\right] \text{ for } l \in \{1,2\}.$$

Now, for $l = 1, 2$, we can write

$$\begin{aligned}\mathbb{E}\left[Y_1^{-l}|X_1\right] &= g_l(x)\mathbb{E}\left[K_1\right] + \mathbb{E}\left[K_1\mathbb{E}\left[g_l\left(X_1\right) - g_l(x)|d\left(X_1,x\right)\right]\right] \\ &= g_l(x)\mathbb{E}\left[K_1\right] + \mathbb{E}\left[K_1\left(\Psi_l\left(d\left(X_1,x\right)\right)\right)\right].\end{aligned}$$

By the same arguments as those used by Ferraty et al. (2007) for the regression operator, we show that

$$\mathbb{E}\left[\widetilde{g}_l(x)\right] = g_l(x) + h\Psi_l'(0)\left[\frac{K(1) - \int_0^1 (uK(u))'\chi_x(u)du}{K(1) - \int_0^1 K'(u)\chi_x(u)du}\right] + o(h).$$

Then, we deduce that

$$A_n = \frac{\mathbb{E}\left[\widetilde{g}_1(x)\right]}{\mathbb{E}\left[\widetilde{g}_2(x)\right]} - r(x) = hB_n + o(h).$$

□

Proof of Lemma 2.3.6:

For the first limit in the Lemmas 5 and 6's results, we have

$$\mathbb{E}\left[\widetilde{g}_2(x) - g_2(x)\right] \to 0,$$

and

$$\text{var}\left[\widetilde{g}_2(x)\right] \to 0.$$

Hence,

$$\widetilde{g}_2(x) - g_2(x) \to 0, \text{ in probability}.$$

Next, for the last needed convergence, we obtain by the same manner:

$$\mathbb{E}\left[\left(\frac{n\varphi_x(h)}{g_1(x)^2\sigma^2(x)}\right)^{1/2} A_n\left(\widetilde{g}_2(x) - \mathbb{E}\left[\widetilde{g}_2(x)\right]\right)\right] = 0$$

and

$$\operatorname{var}\left[\left(\frac{n\varphi_x(h)}{g_1(x)^2\sigma^2(x)}\right)^{1/2} A_n\left(\widetilde{g}_2(x) - \mathbb{E}\left[\widetilde{g}_2(x)\right]\right)\right] = O(A_n^2) = O(h^2) \to 0.$$

It follows that

$$\left(\frac{n\varphi_x(h)}{g_1(x)^2\sigma^2(x)}\right)^{1/2} A_n\left(\widetilde{g}_2(x) - \mathbb{E}\left[\widetilde{g}_2(x)\right]\right) \to 0, \text{ in probability}.$$

□

BIBLIOGRAPHY

B. Altendji, J. Demongeot, A. Laksaci, and M. Rachdi. Functional data analysis: estimation of the relative error in functional regression under random left-truncation model . *Journal of Nonparametric Statistics*, 30:472–490, 2018.

M. Attouch, A. Laksaci, and N. Messabihi. Nonparametric relative error regression for spatial random variables. *Stat Papers DOI*, 58, 2017.

M. Campbell and A. Donner. Classification efficiency of multinomial logistic regression relative to ordinal logistic regression . *J. Amer. Statist. Asso*, 84:587–591, 1989.

A. Carbonez, L. Gyorfi, and C. Edward. Partitioning-estimates of a regression function under random censoring . *Statist. Decisions*, 13:21–38, 1995.

K. Chen, S. Guo, Y. Lin, , and Z. . Ying. Least absolute relative error estimation. . *Journal of the American Statistical Association*, 105:1104–1112, 2010.

P. Deheuvels and J. H. J. Einmahl. Functional limit laws for the increments of Kaplan Meier product-limit processes and applications . *The Annals of Probability*, 28:1301–1335, 2000.

J. Demongeot, A. Hamie, A. Laksaci, and M. Rachdi. Relative-Error Prediction in Nonparametric Functional Statistics: Theory and Practice . *Journal of Multivariate analysis*, 147:261–268, 2016.

W. Feller. An introduction to probability theory and ito applications ii. *John Willy, New York*, 1966.

F. Ferraty and P. Vieu. *Nonparametric Functional Data Analysis: Theory and Practice*. Springer Series in Statistics, New York, springer edition, 2006.

F. Ferraty, A. Mas, and P. Vieu. Nonparametric Regression on Functional Data: Inference and Practical Aspects. *Australian and New Zealand Journal of Statistics*, 49:267–286, 2007.

N. Helal and E. Ould-Saïd. Kernel Conditional Quantile Estimator under Left Truncation for Functional Regressors . *Opuscula Mathematica*, 36:25–48, 2016.

W. Horrigue and E. Ould-Saïd. Strong Uniform Consistency of a Nonparametric Estimator of a Conditional Quantile for Censored Dependent Data and Functional Regressors . *Random Operators and Stochastic Equations*, 19:131–156, 2011.

W. Horrigue and E. Ould-Saïd. Nonparametric Regression Quantile Estimation for Dependant Functional Data under Random Censorship: Asymptotic Normality . *Communications in Statistics - Theory and Methods*, 44:4307–4332, 2014.

M. Jones, P. Heungsun, S. Key-Il, and S. Vines. Relative error prediction via kernel regression smoothers. . *J.Statist. Plann. Inference*, 138:2887–2898, 2008.

E. Kaplan, L, and P. , Meier. Nonparametric estimation from incomplete observations . *Journal of the American statistical association*, 53:457–481, 1958.

M. Köhler, K. Máthé, and M. Pintér. Prediction from randomly right censored data . *J. Multivariate Anal.*, 80:73–100, 2002.

M. Loeve. *Probability theory*, volume 1963. Springer, 1963.

E. Ould-Saïd and Z. Guessoum. On nonparametric estimation of the regression function under random censorship model . *Statistics and Decisions*, 26:159–177, 2008.

J. O. Ramsay. Functional data analysis. *Encyclopedia of Statistical Sciences*, 4, 2004.

S. Ruiz-Velasco. Asymptotic efficiency of logistic regression relative to linear discriminant analysis . *Biometrika*, 78:235–243, 1991.

B. Thiam. Relative error prediction in nonparametric deconvolution regression model . *Statistica Neerlandica*, 10:1–15, 2018.

CHAPTER 3

NONPARAMETRIC RELATIVE ERROR REGRESSION FOR DEPENDENT FUNCTIONAL DATA UNDER RANDOM CENSORSHIP

This chapter is the subject of a publication submitted.

Nonparametric relative error regression for dependent Functional data under random censorship

[1]**Mohammed Kadi ATTOUCH**, [2]**Omar FETITAH**, [3]**Salah KHARDANI** and [2]**Ali RIGHI**.

[1]Department of Mathematics Faculty of Sciences King Khalid University PO. Box 9004, Abha, Saudi Arabia.

[2] Laboratory of Statistics and Stochastic Processes University of Djillali Liabes BP 89, Sidi Bel Abbes 22000, Algeria.

[3] Laboratoire des Réseaux Intelligents et Nanotechnologie. Ecole Nationale des Sciences et Technologies Avancées á Borj-Cédria, Tunisia.

Abstract : In this paper, we investigate the asymptotic properties of a nonparametric estimator of the relative error regression given a dependent functional explanatory variable, in the case of a scalar censored response. We use the mean squared relative error as a loss function to construct a nonparametric estimator of the regression operator of these functional censored data. We establish the almost complete convergence (with rates) and the asymptotic normality of the proposed estimator. A simulation study is performed to lend further support to our theoretical results and to compare the quality of predictive performances of the relative error regression estimator than those obtained with standard kernel regression estimates.

Keywords : Kernel method, Nonparametric estimation, Functional data analysis, Almost complete convergence, Censored data, Small ball probability, α-mixing dependency, Functional regression, Mean square relative error.

Mathematics Subject Classification: 62G05, 62G08, 62G20, 62G35, 62N01.

3.1 Introduction

Functional data analysis is a branch of statistics that has received an increasing interest in the last few years from mathematical or applications points of view. This type of data appears in many practical situations such as continuous phenomena (climatology, economics, linguistics, medicine, and so on.). Since the work by Ramsay and Dalzell (1991), many developments have been investigated, in order to build theory and methods around functional data. The monographs of Ramsay and Silverman (2005) provide an overview of both the theoretical and practical aspects of functional data analysis and Ferraty and Vieu (2006) for the nonparametric approaches. From a nonparametric point of view, several models have been developed. For instance, Ferraty and Vieu (2004) established the strong consistency of the regression function when the explanatory variable is functional and the response is scalar, and their study extended to non-standard regression problems such as time series prediction and curves' discrimination (see Ferraty et al. (2002); Ferraty and Vieu (2003)), we can also cite (Attouch et al. (2009)) for the robust estimation. The asymptotic normality of the same estimator is established by Masry (2005) under an α-mixing assumption. Dabo-Niang (2004) studied density estimation in a Banach space with an application to the estimation density of a diffusion process with respect to Wiener's measure. The kernel type estimation of some characteristics of the conditional cumulative distribution function as well as the successive derivatives of the conditional density were introduced by Ferraty and Vieu (2006). The almost complete convergence (a.co.) with rates for the kernel type estimates is established and illustrated by an application to El Nin$\tilde{o}$ data.

In this paper, we consider the problem of estimating the regression function based on the minimization of the mean squared relative error (MSRE). We consider a regression model in which the response variable is subject to random right censoring dependent data when the covariates takes values in an infinite dimensional semi-metric vector space $(\mathcal{F}, d(.,.))$.

This problem can be formulated by considering that $(Y_i, X_i) \quad i = 1 \dots n$ is a stationary $\alpha-$ mixing couples, where Y_i is real-valued and X_i takes values in some functional space $\mathcal{F}$. Assume that $\mathbb{E}|Y_i| < \infty$ and define the regression functional as

$$r(x) = \mathbb{E}[Y_i | X_i = x], \quad x \in \mathcal{F}, \quad \forall i \in \mathbb{N}. \tag{3.1}$$

The model (3.1), can be written as follows

$$Y_i = r(X_i) + \varepsilon_i, \quad i = 1 \cdots n,$$

where ε_i is a random variable, such that $\mathbb{E}[\varepsilon_i | X_i] = 0$ and $\mathbb{E}[\varepsilon_i^2 | X_i] = \sigma_i^2(X_i) < +\infty$.

Unlike to the multivariate case, there exists various versions of the functional regression estimate. But, all these versions are based on two common procedures. The first one is the functional operator which is supposed smooth enough to be locally well approximated by a polynomial. The second one is the use of the following least square error

$$r(x) = \arg\min_{r^*} \left(\mathbb{E}\left[(Y - r^*(x))^2 | X = x \right] \right), \tag{3.2}$$

as a loss function to determine the estimates of $r(\cdot)$.

In complete data, a typical kernel regression estimator based on (3.2) (see Ferraty et al. (2007)), is given by

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

where K is a kernel, $(h = h_n)$ is sequence of bandwidths. For an overview results for both theoretical and application points of view considering independent or dependent case, we refer the reader to the studies of Chahad et al. (2017), Attouch et al. (2017). Amiri et al. (2014) studied the regression function of a real random variable with functional explanatory variable by using a recursive nonparametric kernel approach.

In the presence of right random censoring, the problem has been studied by (Buckley and James (1979)) using parametric methods. For nonparametric approaches, we refer to Amiri and Khardani (2018), Stute (1993). Some asymptotic properties were established with a particular application to the conditional mode and quantile by Chaouch and Khardani (2015) and Khardani and Thiam (2016). Horrigue and Ould-Saïd (2014) studied a regression quantile estimation for dependent functional data.

However, the use of previous loss function (3.2) as a measure of prediction performance may be not suitable in some situation. In particular, the presence of outliers can lead to unreasonable results since all variables have the same weight. Now, to overcome this limitation we propose to estimate the function r by an alternative loss function. In the relative regression analysis $r(x)$ is obtained by minimizing the mean squared relative error (MSRE) ie: $r(x)$ is the solution of the optimisation problem:

$$r(x) = \arg\min_{r^*} \left(\mathbb{E}\left[\left(\frac{Y - r^*(X)}{Y} \right)^2 | X = x \right] \right).$$

As mentioned in Jones et al. (2008) where outlier data are present and the response variable of the model is positive, the MSRE is to be minimized. Moreover, the solution of this problem can be expressed by the ratio of first two conditional inverse moments of Y given X. As discussed by Park and Stefanski (1998), for $Y > 0$

$$r(\cdot) = \frac{\mathbb{E}\left[Y^{-1} | X = \cdot\right]}{\mathbb{E}\left[Y^{-2} | X = \cdot\right]} := \frac{g_1(\cdot)}{g_2(\cdot)}, \tag{3.3}$$

where $g_l(\cdot) = \mathbb{E}\left[Y^{-l} | X = \cdot\right]$, for $l = 1, 2$, is the best MSRE predictor of Y given $X = x$.

The literature on the relative error regression in Nonparametric Functional Data Analysis (NFDA) is still limited. The first consistent results were obtained in by Campbell and Donner (1989), where relative regression was used as a classification tool. Jones et al. (2008) studied the nonparametric prediction via relative error regression. They investigated the asymptotic properties of an estimator minimizing the sum of the squared relative errors by considering both (kernel method and local linear approach). Recently, Mechab and Laksaci (2016) studied this regression model when the observations are weakly dependent. For spatial data, Attouch et al. (2017) proved the almost complete consistency and the asymptotic normality of this estimator. Fetitah et al. (2020) investigated the relative error in functional regression under random censorship in the independent case.

Note that, data are truncated when the data set does not include observations in the analysis that are beyond a boundary value. Having a value beyond the boundary eliminates that individual from being in the analysis. In contrast, data are censored when we have partial information about the value of a variable—we know it is beyond some boundary, but not how far above or below it. Thus, the work that we discuss in this paper is completely different from that studied in Altendji et al. (2018).

In this paper we define and study a new estimator of the relative-error regression function when the interest random variable is subject to random right-censoring and the explanatory variable is functional. Notice that the main feature of our approach is to develop a prediction model alternative to the classical regression which is not sensitive to the presence of the outliers.

The paper is organized as follows: in Section 3.2 we define our parameter of interest and its corresponding estimators. In Section 3.3 we give some assumptions and state an almost sure (a.c.) consistency and asymptotic normality for the proposed estimator. A simulation study is performed in Section 3.4, whereas the technical details and the proofs are deferred to Section 3.5.

3.2 Model

3.2.1 Construction of the estimator

To construct our estimator, let us recall that in the case of complete data, a well-known estimator of the regression function is based on the Nadaraya-Watson weights. Let $\{Z_i = (X_i, Y_i)_{1 \le i \le n}\}$ be n pairs, identically distributed as $Z = (X, Y)$ and valued in $\mathcal{F} \times \mathbb{R}$, where $(\mathcal{F}, d)$ is a semi-metric space (i.e. X is a functional random variable (f.r.v) and d a semi-metric). Let x be a fixed element of $\mathcal{F}$.

For the complete data, it is well known that the kernel estimator of (3.3), and is given by

$$\hat{r}(x) = \frac{\sum_{i=1}^{n} Y_i^{-1} K\left(\frac{d(x,X_i)}{h}\right)}{\sum_{i=1}^{n} Y_i^{-2} K\left(\frac{d(x,X_i)}{h}\right)} := \frac{\hat{g}_1(x)}{\hat{g}_2(x)}, \tag{3.4}$$

where $\hat{g}_l(x) = \dfrac{1}{n\mathbb{E}(K_1)} \sum_{i=1}^{n} Y_i^{-l} K\left(\dfrac{d(x,X_i)}{h}\right)$, for $l = 1,2$, with K is an asymmetrical kernel and $h = h_{n,K}$ (depending on n) is a strictly positive real numbers and $K_1 = K\left(\frac{d(x,X_1)}{h}\right)$. It is a functional extension of the familiar Nadaraya-Watson estimate.

The main change comes from the semi-metric d which measures the proximity between functional objects. In the censoring case, instead of observing the lifetimes Y we observe the censored lifetimes of items under study. That is, assuming that $(C_i)_{1\le i\le n}$ is a sequence of i.i.d. censoring random variable (r.v.) with common unknown continuous distribution function (df) G.

Then in the right censorship model, we only observe the n pairs (T_i, δ_i) with

$$T_i = Y_i \wedge C_i \quad and \quad \delta_i = \mathbb{1}_{\{Y_i \le C_i\}}, 1 \le i \le n, \tag{3.5}$$

where $\mathbb{1}_A$ denotes the indicator function of the set A.

In censorship model only the $(X_i, T_i, \delta_i)_{1\le i\le n}$ are observed, we define $\widetilde{r}(x)$ as an estimate of $r(x)$ by

$$\widetilde{r}(x) = \frac{\sum_{i=1}^{n} \frac{\delta_i T_i^{-1}}{\bar{G}(T_i)} K\left(\frac{d(x,X_i)}{h}\right)}{\sum_{i=1}^{n} \frac{\delta_i T_i^{-2}}{\bar{G}(T_i)} K\left(\frac{d(x,X_i)}{h}\right)} =: \frac{\widetilde{g}_1(x)}{\widetilde{g}_2(x)}, \tag{3.6}$$

where $\widetilde{g}_l(x) = \dfrac{1}{n\mathbb{E}(K_1)} \sum_{i=1}^{n} \dfrac{\delta_i T_i^{-l}}{\bar{G}(T_i)} K\left(\dfrac{d(x,X_i)}{h}\right)$ for $l = 1,2$.

In practice G is unknown, we use the Kaplan-Meier estimator Deheuvels and Einmahl (2000) of $\bar{G}$ given by:

$$\bar{G}_n(t) = \begin{cases} \prod_{i=1}^{n}\left(1 - \frac{1-\delta_{(i)}}{n-i+1}\right)^{\mathbb{1}_{\{T_{(i)}\le t\}}} & \text{if } t \le T_{(n)} \\ 0 & \text{otherwise,} \end{cases} \tag{3.7}$$

where $T_{(1)} \le T_{(2)} \le \dots \le T_{(n)}$ are the order statistics of $(T_i)_{1\le i\le n}$ and $\delta_{(i)}$ is the concomitant of $T_{(i)}$.

Finally, the estimator of $r(x)$ can be written as:

$$\widetilde{r}_n(x) = \frac{\sum_{i=1}^{n} \frac{\delta_i T_i^{-1}}{\bar{G}_n(T_i)} K\left(\frac{d(x,X_i)}{h}\right)}{\sum_{i=1}^{n} \frac{\delta_i T_i^{-2}}{\bar{G}_n(T_i)} K\left(\frac{d(x,X_i)}{h}\right)} := \frac{\widetilde{g}_{1,n}(x)}{\widetilde{g}_{2,n}(x)}, \tag{3.8}$$

where $\widetilde{g}_{l,n}(x) = \frac{1}{n\mathbb{E}(K_1)} \sum_{i=1}^{n} \frac{\delta_i T_i^{-l}}{\bar{G}_n(T_i)} K\left(\frac{d(x,X_i)}{h}\right)$ for $l = 1, 2$.

Remark 3.2.1. *In* (3.6) *and* (3.8) *the sums are taken for the subscripts i for which $\bar{G}_n(T_i) \neq 0$ and $\bar{G}(T_i) \neq 0$. The same convention is followed in the forthcoming formulae. Note that under the assumptions on the model, the sets $\{i, \bar{G}(Y_i) = 0\}$ and $\{i, \bar{G}_n(Y_i) = 0\}$ are $\mathbb{P}$-negligible.*

3.3 Assumptions and main results

In what follows, we define the endpoints of F and G by $\tau_F = \sup\{t : \bar{F}(t) > 0\}$, and $\tau_G = \sup\{t : \bar{G}(t) > 0\}$ where $\bar{F}(\cdot) = 1 - F(\cdot)$ and $\bar{G}(\cdot) = 1 - G(\cdot)$. We assume that $\tau_F < \infty$ and $\bar{G}(\tau_F) > 0$, (this implies $\tau_F < \tau_G$).

Throughout this paper, x is a fixed element of the functional space $\mathcal{F}$. To formulate our assumptions, some notations are required. and we denote by $\mathcal{N}_x$ a neighborhood of the point x. Hereafter, when no confusion is possible, we will denote by c and c' some strictly positive generic constants.

Let $B(x,h)$ be the closed ball centered at x with radius h, and consider the cumulative distribution function (CFD) of $d(x,X)$ defined by

$$\varphi_x(h) = \mathbb{P}(X \in B(x,h)) = \mathbb{P}(d(x,X) \leq h),$$

h being positive and satisfies $\varphi_x(0) = 0$ and $\varphi_x(h) \to 0$ when $h \to 0$. Let us consider the following definition.

Definition 3.3.1. *Let $(Z_n)_{n\in\mathbb{N}}$ be a sequence of rv's. Given a positive integer n, set*

$$\alpha(n) = \sup_k \left\{ |\mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)|, A \in \mathcal{F}_1^k(Z) \text{ and } B \in \mathcal{F}_{k+n}^{\infty}(Z) \right\},$$

where $\mathcal{F}_i^k(Z)$ denotes the σ–field generated by $\{Z_j, i \leq j \leq k\}$.

The sequence is said to be α–mixing if the mixing coefficient $\alpha(n) \to 0$ when $n \to \infty$.

3.3.1 Asymptotic consistency

Our main first result is the pointwise almost sure convergence. In order to state this result, we will need some assumptions which are gathered together in order to make our results reading easier.

In what follows, we will assume that the following assumptions hold:

(H1) $P(X \in B(x,h)) =: \varphi_x(h) > 0$ for all $h > 0$.

(H2) For all $(x_1, x_2) \in \mathcal{N}_x^2$, we have

$$|g_l(x_1) - g_l(x_2)| \leq C d^{k_l}(x_1, x_2) \text{ for an integer } k_l > 0.$$

(H3) The kernel K is a bounded and Lipschitzian function on its support $(0,1)$ and satisfying:

$$0 < c \leq K(.) \leq c' < +\infty.$$

(H4) The bandwidth h satisfies:$h \to 0$, $\dfrac{\log n}{n\varphi_x(h)} \to 0$ as $n \to \infty$.

(H5) The inverse moments of the response variable verify:

$$\text{for all } m \geq 2, \quad \mathbb{E}[Y^{-m}|X = x] < c < \infty.$$

(H6) **(i)** $(X_n, Y_n)_{n\geq 1}$ is a sequence of stationary $\alpha-$mixing rv's with coefficient $\alpha(n) = O(n^{-a})$, for some $a \in \mathbb{R}_+^*$.

(ii) $(C_n)_{n\geq 1}$ and $(X_n, Y_n)_{n\geq 1}$ are independent.

(H7)

$$\forall i \neq j, \mathbb{E}[Y_i^{-1} Y_j^{-2} | (X_i, X_j)] \leq c < \infty,$$

and

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

(H8) There exists $\eta > 0$, such that, $c n^{\frac{3-a}{a+1}+\eta} \leq \varphi_x(h) \leq c' n^{\frac{1}{1-a}}$, with $a > 2$.

We are in state to give our main result.

Theorem 3.3.1. *Under Assumptions **(H1)-(H8)**, we have*

$$|\widetilde{r}_n(x) - r(x)| = O\left(h^{k_1}\right) + O\left(h^{k_2}\right) + O_{a.s.}\left(\sqrt{\frac{\log n}{n\varphi_x(h)}} \right). \tag{3.9}$$

Proofs of Theorem 3.3.1: From (3.8), we have:

$$\begin{aligned}
|\widetilde{r}_n(x) - r(x)| &\leq \frac{1}{|\widetilde{g}_{2,n}(x)|}\Big\{ |\widetilde{g}_{1,n}(x) - \widetilde{g}_1(x)| + |\widetilde{g}_1(x) - \mathbb{E}(\widetilde{g}_1(x))| \\
&+ |\mathbb{E}(\widetilde{g}_1(x)) - g_1(x)| \Big\} + \frac{|r(x)|}{|\widetilde{g}_{2,n}(x)|}\Big\{ |\widetilde{g}_{2,n}(x) - \widetilde{g}_2(x)| \\
&+ |\widetilde{g}_2(x) - \mathbb{E}(\widetilde{g}_2(x))| + |\mathbb{E}(\widetilde{g}_2(x)) - g_2(x)| \Big\}.
\end{aligned}$$

Therefore, Theorem 3.3.1's result is a consequence of the following intermediate results, where their proofs are postponed to the appendix.

Lemma 3.3.1. *Under hypotheses **(H2)-(H5)**, we have*

$$|\widetilde{g}_{l,n}(x) - \widetilde{g}_l(x)| = O_{a.s.}\left(\sqrt{\frac{\log\log n}{n}}\right), \tag{3.10}$$

for $l \in \{1,2\}$.

Lemma 3.3.2. *Under hypotheses **(H1)-(H3)** and **(H5)**, we have*

$$|\mathbb{E}(\widetilde{g}_l(x)) - g_l(x)| = O\left(h^{k_l}\right), \tag{3.11}$$

for $l \in \{1,2\}$.

Lemma 3.3.3. *Under hypotheses **(H1)-(H4)** and **(H6)-(H8)**, we have*

$$|\widetilde{g}_l(x) - \mathbb{E}(\widetilde{g}_l(x))| = O_{a.co.}\left(\sqrt{\frac{\log n}{n\varphi_x(h)}}\right), \tag{3.12}$$

for $l \in \{1,2\}$.

Corollary 3.3.1. *Under the hypotheses of lemma 3.3.2 and 3.3.3, we obtain:*

$$\text{there exists } \delta > 0; \text{ such that } \sum_{n=1}^{\infty} \mathbb{P}\left(|\widetilde{g}_{2,n}(x)| < \delta \right) < \infty. \tag{3.13}$$

3.3.2 Asymptotic normality

This section is devoted to the study of the asymptotic normality of $\widetilde{r}_n(x)$. To do that, we replace assumptions **(H1)**, **(H3)** and **(H4)** respectively by the following assumptions:

(N1) The concentration property **(H1)** holds. Moreover, There exists a function $\chi_x(.)$, such that:

$$\text{for all } \ s \in [0,1], \quad \lim_{r\to 0} \frac{\varphi_x(sr)}{\varphi_x(r)} = \chi_x(s).$$

(N2) For $\gamma \in \{1,2\}$, the functions $\Psi_\gamma(.) = \mathbb{E}\left[g_\gamma(X) - g_\gamma(x) | d(x,X) = .\right]$ are derivable at 0.

(N3) The kernel function K satisfies **(H3)** and is a differentiable function on $]0,1[$ where its first derivative function K' is such that; $-\infty < c < K'(.) < c' < 0$.

(N4) The small ball probability satisfies, $n\varphi_x(h) \longrightarrow \infty$.

(N5) For $m \in \{1,2,3,4\}$, the functions $q_m(\cdot) = \mathbb{E}\left[\bar{G}(Y)^{-1} Y^{-m} | X = \cdot\right]$ are continuous in a neighborhood of x.

Remarks on the assumptions

Assumption **(H1)** is the same as that used by Ferraty and Vieu (2006) which is linked to the functional structure of the functional covariates. Assumptions **(H2)**, **(H3)** and **(H7)** deal with the functional aspect of the covariates and the associated small ball probability techniques used in this paper. Assumptions **(H6)** and **(H8)** specify the model and the rate of mixing coefficient. Condition **(N5)** stands as regularity condition that is useful to establish the asymptotic properties of the estimators. Assumptions **(H3)**, **(H4)**, **(N3)** and **(N4)** concern the kernel $K(\cdot)$ and the smoothing parameter h and are technical conditions.

The fractal or geometric process is a family of infinite dimensional processes for which the small balls have the property

$$\varphi_x(t) = \mathbb{P}\left(\|x - X\| < t\right) \sim c_x t^\gamma,$$

where c_x and γ are positive constants. In this case, setting $h_n = An^{-a}$ with $0 < a < 1$ and $0 < A$ implies $\varphi_x(h) = c_x A n^{-\gamma a}$. Thus, **(H1)**, **(H4)** and **(H8)** hold when $\gamma < 1/a$.

Theorem 3.3.2. *Under Assumptions **(H6)-(H8)**and **(N1)-(N4)**, we have*

$$\left(\frac{n\varphi_x(h)}{\sigma^2(x)}\right)^{1/2} \left(\widetilde{r}_n(x) - r(x) - hB_n(x) - o(h)\right) \xrightarrow{\mathcal{D}} \mathcal{N}(0,1) \quad as \quad n \to \infty. \tag{3.14}$$

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

$$B_n(x) = \frac{\left(\Psi_1'(0) - r(x)\Psi_2'(0)\right)\beta_0}{\beta_1 g_2(x)}, \tag{3.15}$$

$$\sigma^2(x) = \frac{\left(q_2(x) - 2r(x)q_3(x) + r^2(x)q_4(x)\right)\beta_2}{\beta_1^2}, \tag{3.16}$$

with

$$\beta_0 = K(1) - \int_0^1 (sK(s))'\chi_x(s)\mathrm{d}s \quad \textit{and} \quad \beta_j = K^j(1) - \int_0^1 \left(K^j\right)'(s)\chi_x(s)\mathrm{d}s, \neq 0$$

for $j = 1, 2$.

Remark 3.3.1. *(***Comeback to complete data***). In absence of censoring, ($\bar{G}(\cdot) = 1$), the asymptotic variance becomes*

$$\sigma^2(x) = \frac{\left(a_2(x) - 2r(x)a_3(x) + r^2(x)a_4(x)\right)\beta_2}{\beta_1^2},$$

where

$$a_j(\cdot) = \mathbb{E}\left[Y^{-j}|X = \cdot\right],$$

corresponding to the result obtained in Demongeot et al. (2016).

Proofs of Theorem 3.3.2.

From (3.8), we adopt the following decomposition:

$$\widetilde{r}_n(x) - r(x) = \widetilde{r}_n(x) - \widetilde{r}(x) + \widetilde{r}(x) - r(x) =: I_{1n}(x) + I_{2n}(x),$$

where

$$I_{1n}(x) =: \widetilde{r}_n(x) - \widetilde{r}(x) \quad \text{and} \quad I_{2n}(x) =: \widetilde{r}(x) - r(x).$$

The proof can be deduced by showing that $I_{1n}(x)$ is negligible, whereas $I_{2n}(x)$ is asymptotically normal. From Lemma 3.3.1 and Corollary 3.3.1, we deduce that

$$I_{1n}(x) \to 0, \text{in probability}. \tag{3.17}$$

Now, we can write that

$$I_{2n}(x) = \frac{1}{\widetilde{g}_2(x)}\left[D_n + A_n\left(\mathbb{E}\left[\widetilde{g}_2(x)\right] - \widetilde{g}_2(x)\right)\right] + A_n, \tag{3.18}$$

where

$$\begin{aligned} A_n &= \frac{1}{\mathbb{E}\left[\widetilde{g}_2(x)\right] g_2(x)}\left\{\mathbb{E}\left[\widetilde{g}_1(x)\right] g_2(x) - \mathbb{E}\left[\widetilde{g}_2(x)\right] g_1(x)\right\} \\ D_n &= \frac{1}{g_2(x)}\left[V_{1n}(x)\, g_2(x) - V_{2n}(x)\, g_1(x)\right], \end{aligned}$$

whit

$$V_{ln}(x) = \widetilde{g}_l(x) - \mathbb{E}\left[\widetilde{g}_l(x)\right], \quad \text{for} \quad l = 1, 2.$$

Then, it follows from (3.18), that

$$\begin{aligned} \widetilde{r}(x) - r(x) - A_n &= \frac{1}{\widetilde{g}_2(x)}\left[D_n + A_n\left(\mathbb{E}\left[\widetilde{g}_2(x)\right] - \widetilde{g}_2(x)\right)\right] \\ &=: \frac{D_n - A_n V_{2n}(x)}{\widetilde{g}_2(x)}, \end{aligned}$$

where

$$V_{ln}(x) = \widetilde{g}_{l,n}(x) - \mathbb{E}\left[\widetilde{g}_{l,n}(x)\right], \quad \text{for} \quad l = 1, 2.$$

Consequently, the proof of Theorem 3.3.2 can be deduced from the convergence in (3.17), and the following intermediate results (cf. Lemmas 3.3.4, 3.3.5 and 3.3.6), which the proves are postponed into the Appendix.

Lemma 3.3.4. *Under hypotheses of Theorem 3.3.2, we have*

$$\left(\frac{n\varphi_x(h)}{g_2^2(x)\sigma^2(x)}\right)^{1/2}\left(\left[\widetilde{g}_1(x) - \mathbb{E}\left[\widetilde{g}_1(x)\right]\right] g_2(x) - \left[\widetilde{g}_2(x) - \mathbb{E}\left[\widetilde{g}_2(x)\right]\right] g_1(x)\right) \xrightarrow{\mathcal{D}} \mathcal{N}(0,1).$$

Lemma 3.3.5. *Under assumptions of Theorem 3.3.2, we obtain*

$$A_n = hB_n + o(h).$$

Lemma 3.3.6. *Under assumptions of Theorem 3.3.2, we get*

$$\widetilde{g}_2(x) \to g_2(x), \textit{in probability},$$

and

$$\left(\frac{n\varphi_x(h)}{g_2^2(x)\sigma^2(x)}\right)^{1/2} A_n\left(\mathbb{E}[\widetilde{g}_2(x)] - \widetilde{g}_2(x)\right) \to 0, \textit{in probability}.$$

3.4 Simulation study

In this section, we treat a simulation example to show the behaviour of our estimator $\widetilde{r}_n(x)$, and compare the obtained result in presence of outliers to the classical regression defined as the conditional expectation $m(x) = \mathbb{E}[Y|X = x]$, estimated by

$$\widehat{m}(x) = \frac{\sum_{i=1}^{n} \frac{\delta_i T_i}{\bar{G}_n(T_i)} K\left(h_n^{-1} d\left(x, X_i\right)\right)}{\sum_{i=1}^{n} K\left(h_n^{-1} d\left(x, X_i\right)\right)},$$

and the relative error estimator $\widetilde{r}_n(x)$ previously defined in the equation 3.8.

To do this, we consider the classical nonparametric functional regression model

$$Y = r(X) + \varepsilon,$$

where the operator r is defined by

$$r(X) = \frac{10}{1 + \int_0^1 X^2(t)\mathrm{d}t}. \tag{3.19}$$

We consider two diffusion processes on the interval $[0,1]$, $Z_1(t) = 2 - \cos(\pi t W)$ and $Z_2(t) = \cos(\pi t W)$, $(W \to \mathcal{N}(0,1))$ and we take $X(t) = AZ_1(t) + (1 - A)Z_2(t)$, where A is a Bernoulli distributed random variable.

We carried out the simulation with 200-sample of the curve X, and ε is an α-mixing process generated by the following model

$$\varepsilon_i = \frac{1}{\sqrt{2}}\left(\varepsilon_{i-1} + \eta_i\right), \quad i = 1, \ldots, 200,$$

η_i being centered Gaussian rv's with variance 0.5, and is independent of η_i. We also, simulate n i.i.d. rv's $C_i, i = 1, ..., n$ with law $\mathcal{E}(\lambda)$ (that is exponentially distributed with density $\lambda e^{-\lambda x}\mathbb{1}_{x\geq 0}$).

Simulated data from our model are plotted in Figure3.1. To compute our estimator based on the observed data $(X_i, T_i, \delta_i), i = 1, ..., n$, where $T_i = Y_i \wedge C_i$ and $\delta_i = \mathbb{1}_{\{Y_i \leq C_i\}}$.

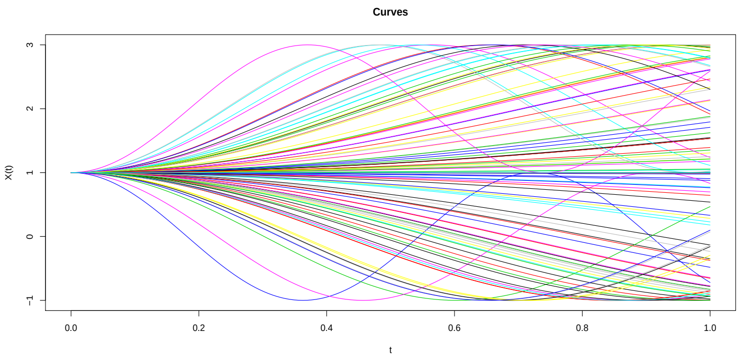

Figure 3.1: The curves $X_{i=1,...,100}(t), \quad t \in [0,1[$.

We choose the quadratic kernel defined by

$$K(x) = \frac{3}{2}\left(1 - x^2\right) \mathbb{1}_{(0,1)}.$$

In practice, the semi-metric choice is based on the regularity of the curves $X(.)$ which are under study. In our case, we take the semi-metric based on the second derivatives of the curves x. More precisely, we take

$$d(x, x') = \left(\int_0^1 (x^{(i)}(t) - x'^{(i)}(t))^2 dt \right)^{1/2},$$

where $x^{(i)}$ denotes the i^{-th} derivative of the curve x.

For the bandwidth, we choose the automatic selection with a cross validation procedure introduced by (Ferraty and Vieu (2006), Chapter 13).

We split the data generated from the model above into two subsets: a training sample $(X_i, T_i, \delta_i), i = 1, ..., 150$ and a test sample $(X_j, T_j, \delta_j), j = 151, ..., 200$. Then, we calculate the estimator $\hat{\theta}(X_j)$ for any $j \in \{151, ..., 200\}$.

The performance of both estimators is described by the mean squared prediction error:

$$MSE = \frac{1}{50} \sum_{j=151}^{200} (Y_j - \tilde{r}(X_j))^2,$$

where $\tilde{r}(X_j)$ means the estimator of both regression models. We note that the result of our simulation study is evaluated over 100 independent replications.

The obtained results are shown in Figure 3.2 with the censorship rate $CR = 20.67\%$ it is clear that there is no meaningful difference between the two estimation methods: the Classical Kernel Estimator and the Relative Error Estimator ($MSE_{CKE} = 0.2209$, $MSE_{REE} = 0.1579$)

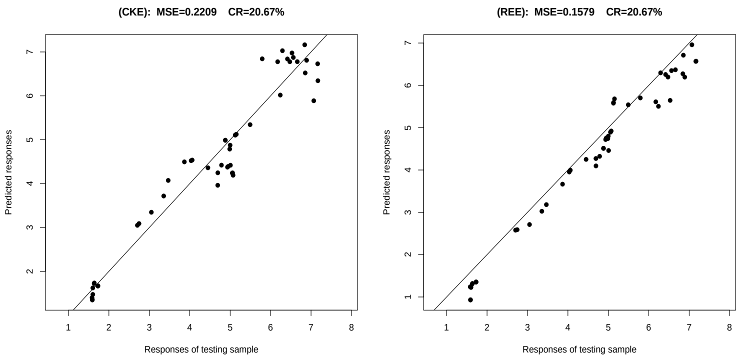

Figure 3.2: Comparison between the Classical Kernel Estimator (CKE) and the Relative Error Estimator (REE).

The second illustration is given in the following table where we observe that in the presence of outliers (0, 10, 20) with different values of censorship rate ($CR = 3\%, 30\%, 60\%$), the Relative Error regression gives better results than the classical method, in sense that, even if the MSE value of the both methods increases substantially relatively to the number of the perturbed points and censorship rate, but it remains very low for the Relative Error one.

Table 3.1: MSE for the Classical Kernel Estimator (CKE) and the Relative Error Estimator (REE) according to numbers of introduced artificial outliers and different censorship rate.

number of artificial outliers	→ censorship rate CR ↓	0	10	20
	3%	0.0921	2856.646	6499.6945
Classical Kernel Estimator	30%	0.8766	14126.2706	19358.5386
MSE_{CKE}	60%	2.8038	32182.8188	56681.7038
	3%	0.0551	0.0579	0.0665
Relative Error Estimator	30%	0.0949	0.1048	0.1258
MSE_{REE}	60%	0.1455	0.1903	0.2712

Confidence Interval:

Our main application of Theorem 3.3.2 is to build confidence interval for the true value of $r(x)$ given curve $X = x$. A plug-in estimate for the asymptotic standard deviation $(n\varphi_x(h)/\sigma^2(x))^{1/2}$ and the bias term $hB_n(x) + o(h)$. Precisely, we estimate $q_m(x)$ by

$$\widetilde{q}_m(x) = \frac{\sum_{i=1}^{n} K_i \delta_i \bar{G}_n^{-2}(T_i) T_i^{-m}}{\sum_{i=1}^{n} K_i}.$$

Whereas we estimate empirically β_1 and β_2 by

$$\hat{\beta}_1 = \frac{1}{n\varphi_x(h)} \sum_{i=1}^{n} K_i, \text{ and } \hat{\beta}_2 = \frac{1}{n\varphi_x(h)} \sum_{i=1}^{n} K_i^2.$$

Finally, the practical estimator of the normalised deviation is

$$\widetilde{\sigma}_n(x) = \left(\frac{\left(\sum_{i=1}^{n} K_i^2\right)\left(\widetilde{q}_2(x) - 2\widetilde{r}(x)\widetilde{q}_3(x) + \widetilde{r}^2(x)\widetilde{q}_4(x)\right)}{\left(\sum_{i=1}^{n} K_i\right)^2 \widetilde{q}_2^2(x)} \right)^{1/2}.$$

We point out that the function $\varphi_x(.)$ do not intervene in the calculation of the confidence interval by simplification. Finally, the approximate $1 - \zeta/2$ confidence interval for $r(x)$, for any $x \in \mathcal{F}$, is

$$\left[\widetilde{r}_n(x) - t_{1-\zeta/2}\widetilde{\sigma}_n(x) \quad , \widetilde{r}_n(x) + t_{1-\zeta/2}\widetilde{\sigma}_n(x)\right],$$

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

In order to compare our confidence interval with that of the classical regression, we will generalize the Corollary 1 in Ferraty et al. (2007), for the censored case, we get

$$\sqrt{n\varphi_x(h)} \frac{\beta_1}{\sigma_\varepsilon(x)\sqrt{\beta_2}} \left(\widehat{m}(x) - m(x)\right) \to \mathcal{N}(0, 1),$$

where $\sigma_\varepsilon^2(x) = \mathbb{E}\left[(Y - m(x))^2 | X = x\right]$ and β_1, β_2 are define previously.

With simple calculus, we can estimate $\sigma_\varepsilon^2(x)$ by:

$$\widehat{\sigma}_\varepsilon^2(x) = \widehat{\rho}_2(x) - 2\widehat{m}(x)\widehat{\rho}_1(x) + \widehat{m}^2(x),$$

where

$$\widehat{\rho}_m(x) = \frac{\sum_{i=1}^{n} K_i \delta_i \bar{G}_n^{-1}(T_i) T_i^m}{\sum_{i=1}^{n} K_i} \text{ for all } m \in \{1, 2\}.$$

Finally, the approximate $1 - \zeta/2$ confidence interval for $m(x)$ (the classical regression), for any $x \in \mathcal{F}$, is

$$\left[\widehat{m}(x) - t_{1-\zeta/2} \frac{\sqrt{\widehat{\beta}_2}\widehat{\sigma}_\varepsilon(x)}{\widehat{\beta}_1} \quad , \widehat{m}(x) + t_{1-\zeta/2} \frac{\sqrt{\widehat{\beta}_2}\widehat{\sigma}_\varepsilon(x)}{\widehat{\beta}_1}\right].$$

In order to construct confidence bands (for both CKE and REE), we proceed by the following algorithm:

Step 1 We split our data into randomly chosen subsets:$(X_i, Y_i)_{i \in I}$: training sample and $(X_j, Y_j)_{j \in J}$: test sample.

Step 2 We calculate the estimator $\widetilde{r}_n(X_i)$ for all $i \in I$ by using the training sample.

Step 3 For each X_j in the test sample, we set: $i^* := \arg\min_{i \in I} d(X_j, X_i)$.

Step 4 For all, $j \in J$, we define the confidence bands by

$$[\widetilde{r}_n(X_{i^*}) - t_{0.975}\widetilde{\sigma}_n(X_{i^*}) \quad , \widetilde{r}_n(X_{i^*}) + t_{0.975}\widetilde{\sigma}_n(X_{i^*})],$$

where $t_{0.975}$ is the 2.5% quantile of a standard normal distribution.

Step 5 We present our results by plotting the extremities of the predicted values versus the true values and the confidence bands.

Figures (3.3) and (3.4) shows clearly a good behaviour of our estimator compared to the classical regression, with censorship rate $(CR = 30\%)$, without and in the presence of outliers.

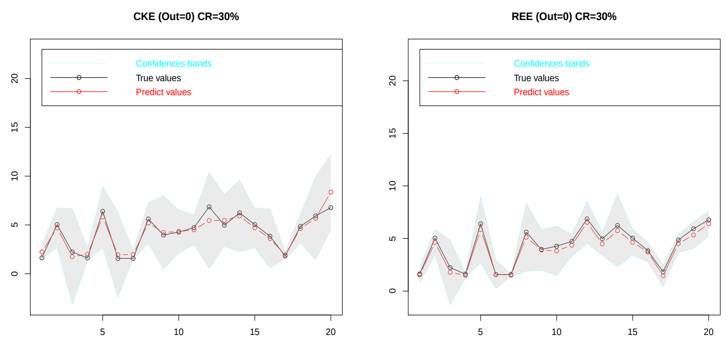

Figure 3.3: Extremities of the predicted values versus the true values and the confidence bands (simulation data without outliers). The solid black curve connects the true values. The dashed Blue curves connect the lower and upper predicted values. The solid Red curve connects the crossed points which give the predicted values.

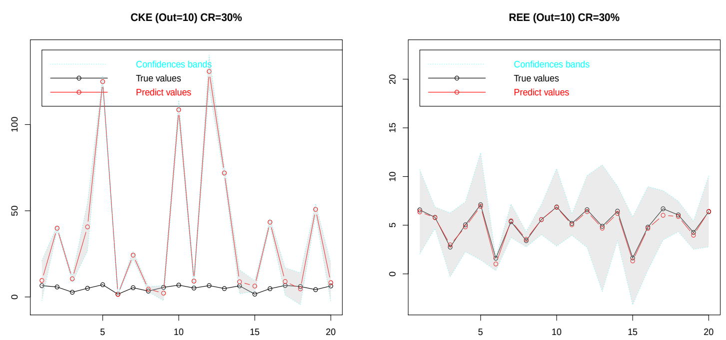

Figure 3.4: Extremities of the predicted values versus the true values and the confidence bands (simulation data in the presence of 10 outliers). The solid black curve connects the true values. The dashed Blue curves connect the lower and upper predicted values. The solid Red curve connects the crossed points which give the predicted values.

3.5 Appendix

We recall the exponential type inequality used above in the proof of the Theorem 3.3.1.

Lemma 3.5.1. *[Fuk-Nagaev](see Rio (1999), p. 87, 6.19b). Let $\{U_i, i \in \mathbb{N}\}$ be a sequence of centered real rv's, with strong mixing coefficient $\alpha(n) = O(n^{-\nu}), \nu > 1$, such that $\forall n \in \mathbb{N}, 1 \le i \le n \quad |U_i| < +\infty$. Then for each $r > 1$:*

$$\mathbb{P}\left\{\left|\sum_{i=1}^{n} U_i\right| > \epsilon\right\} \le c\left(1 + \frac{\epsilon^2}{16 r S_n^2}\right)^{-r/2} + ncr^{-1}\left(\frac{2r}{\epsilon}\right)^{\nu+1},$$

where $S_n^2 = \sum_{1\le i,j\le n} |cov(U_i, U_j)|$.

Let denote by, $K_i(x) = K\left(\frac{d(x,X_i)}{h}\right)$.

Proof of Lemma 3.3.1:

For all $l = 1, 2.$ we have:

$$\begin{aligned} |\widetilde{g}_{l,n}(x) - \widetilde{g}_l(x)| &\le \frac{1}{\mathbb{E}(K_1(x))} \sum_{i=1}^{n} \left| \frac{\delta_i T_i^{-l}}{\bar{G}_n(T_i)} K_i(x) - \frac{\delta_i T_i^{-l}}{\bar{G}(T_i)} K_i(x) \right| \\ &\le \frac{\sup_{t\in\mathbb{R}} |\bar{G}_n(t) - \bar{G}(t)|}{\bar{G}_n(\tau_F)\bar{G}(\tau_F)} \frac{\sum_{i=1}^{n} |Y_i^{-l} K_i(x)|}{n\mathbb{E}(K_1(x))}. \end{aligned}$$

Since $\bar{G}(\tau_F) > 0$, in conjunction withe SLLN[1] and the LIL[2] on the censoring law (see formula (4.28) in Deheuvels and Einmahl (2000)), we have

$$|\widetilde{g}_{l,n}(x) - \widetilde{g}_l(x)| \le \frac{\mathbb{E}|Y_1^{-l} K_1(x)|}{\mathbb{E}(K_1(x))} \frac{1}{\bar{G}^2(\tau_H)} \sqrt{\frac{\log\log n}{n}}$$

(**H5**) concludes the proof. □

Proof of Lemma 3.3.2:

For all $l = 1, 2$, we have

$$\begin{aligned} |\mathbb{E}(\widetilde{g}_l(x)) - g_l(x)| &= \left| \mathbb{E}\left(\frac{K_1(x)}{\mathbb{E}(K_1(x))} \mathbb{E}\left[\frac{\mathbb{E}(\mathbb{1}_{Y_1 \le C_1} |Y_1) Y_1^{-l}}{\bar{G}(Y_1)} |X_1 \right] \right) - g_l(x) \right| \\ &= \frac{1}{\mathbb{E}(K_1(x))} \left| \mathbb{E}\left\{ \left[\mathbb{E}(Y_1^{-l}|X_1) - g_l(x) \right] \mathbb{1}_{B(x,h)}(X_1) K_1(x) \right\} \right|. \end{aligned}$$

[1] Strong law of large numbers
[2] Law of the iterated logarithm

Then, by the Hölder hypothesis (**H2**), we get

$$|g_l(X_1) - g_l(x)| \le ch^{k_l}.$$

Thus,

$$|\mathbb{E}(\widetilde{g}_l(x)) - g_l(x)| \le ch^{k_l}.$$

□

Proof of Lemma 3.3.3:

For $l = 1, 2$, we note

$$\Delta_i(x) = \frac{\delta_i T_i^{-l}}{\bar{G}(T_i)} K\left(\frac{d(x, X_i)}{h}\right) - \mathbb{E}\left[\frac{\delta_i T_i^{-l}}{\bar{G}(T_i)} K\left(\frac{d(x, X_i)}{h}\right)\right].$$

The use of the Fuk-Nagaev's inequality (see lemma (3.5.1)) which is based on

$$\begin{aligned} S_n^2 &= \sum_{i=1}^{n}\sum_{j=1}^{n} |\text{Cov}\left(\Delta_i(x), \Delta_j(x)\right)| \\ &= \sum_{i \neq j} |\text{Cov}\left(\Delta_i(x), \Delta_j(x)\right)| + n\,\text{var}\left(\Delta_1(x)\right). \end{aligned}$$

By using (**H5**), we get

$$\begin{aligned} \text{var}\left(\Delta_1(x)\right) &\le \mathbb{E}\left[\frac{\delta_1 Y_1^{-2l}}{\bar{G}^2(Y_1)} K_1^2(x)\right] + \mathbb{E}^2\left[\frac{\delta_1 Y_1^{-l}}{\bar{G}(Y_1)} K_1(x)\right] \\ &\le \frac{c}{\bar{G}(\tau_H)} \mathbb{E}\left[K_1^2(x)\right] + c\varphi_x^2(h) \\ &\le c\left(\varphi_x(h) + \varphi_x^2(h)\right). \end{aligned}$$

On the other hand, for $i \neq j$, we have

$$\begin{aligned} |\text{Cov}\left(\Delta_i(x), \Delta_j(x)\right)| &= |\mathbb{E}\left(\Delta_i(x)\Delta_j(x)\right)| \\ &\le c|\mathbb{E}\left(K_i(x)K_j(x)\right) + \mathbb{E}\left(K_i(x)\right)\mathbb{E}\left(K_j(x)\right)|. \end{aligned}$$

Now, following Masry (1986), we define the sets

$$E_1 = \{(i,j) \text{ such that } 1 \le |i-j| \le \nu_n\} \text{ and } E_2 = \{(i,j) \text{ such that } \nu_n + 1 \le |i-j| \le n\},$$

where $\nu_n \to \infty$ as $n \to \infty$, we can write

$$\sum_{i \neq j} |\text{Cov}\left(\Delta_i(x), \Delta_j(x)\right)| = J_{1,n} + J_{2,n},$$

where $J_{1,n}$ and $J_{2,n}$ be the sums of the covariances over E_1 and E_2 respectively.

Therefore, under (**H7**), we get

$$\begin{aligned} J_{1,n} &= \sum_{E_1} |\text{Cov}(\Delta_i(x), \Delta_j(x))| \leq c \sum_{E_1} \left|\mathbb{E}(K_i(x)K_j(x)) + \mathbb{E}(K_1(x))^2\right| \\ &\leq c \sum_{E_1} \left|\mathbb{P}((X_i, X_j) \in B(x,h) \times B(x,h)) + \varphi_x(h)^2\right| \\ &\leq cn\nu_n\varphi_x(h)\left[\left(\frac{\varphi_x(h)}{n}\right)^{\frac{1}{a}} + \varphi_x(h)\right]. \end{aligned}$$

For the second term, we use the modified Davydov covariance inequality for mixing processes (see Rio (1999), Formula 1.12a, p.10), we have

$$\forall i \neq j, |\text{Cov}(\Delta_i(x), \Delta_j(x))| \leq c\alpha(|i-j|).$$

Then, we get by (**H6**)

$$J_{2,n} \leq \sum_{E_2} |\text{Cov}(K_i(x), K_j(x))| \leq n^2 \nu_n^{-a}.$$

So, for $\nu_n = \left(\frac{\varphi_x(h)}{n}\right)^{-1/a}$, we will have

$$\sum_{i \neq j} |\text{Cov}(\Delta_i(x), \Delta_j(x))| = O(n\varphi_x(h)). \tag{3.20}$$

Finally, combining previous result

$$S_n^2 = O(n\varphi_x(h)). \tag{3.21}$$

Using Fuk-Nagaev's inequality (see Rio (1999), Formula 6.19b, p.87), we get for all $l = 1, 2$, $\varepsilon > 0$ and $r > 1$

$$\begin{aligned} \mathbb{P}\left[\left|\mathbb{E}[\widetilde{g}_l(x)] - \widetilde{g}_l(x)\right| > \varepsilon\right] &= \mathbb{P}\left[\left|\frac{1}{n\mathbb{E}(K_1(x))} \sum_{i=1}^{n} \Delta_i(x)\right| > \varepsilon\right] \\ &= \mathbb{P}\left[\left|\sum_{i=1}^{n} \Delta_i(x)\right| > \varepsilon n\mathbb{E}(K_1(x))\right] \\ &\leq c\left\{\left(1 + \frac{\varepsilon^2 n^2 \mathbb{E}(K_1(x))^2}{rS_n^2}\right)^{-r/2} + nr^{-1}\left(\frac{r}{\varepsilon n\mathbb{E}(K_1(x))}\right)^{a+1}\right\} \\ &\leq c(A_1 + A_2), \end{aligned}$$

where

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

Therefore, by (3.21) and putting

$$\varepsilon = \varepsilon_0 \sqrt{\frac{\log n}{n\varphi_x(h)}} \quad \text{and} \quad r = (\log n)^2.$$

It follow that

$$A_2 \le c n^{1-(a+1)/2} \varphi_x(h)^{-(a+1)/2} (\log n)^{(3a-1)/2}.$$

Next, using the left side of (**H8**), we obtain

$$A_2 \le c n^{-1-\eta(a+1)/2} (\log n)^{(3a-1)/2}.$$

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

$$A_2 \le c n^{-1-\nu}. \tag{3.22}$$

Because of $r = (\log n)^2$, we show that

$$A_1 \le \left(1 + \frac{\varepsilon_0^2}{\log n}\right)^{-\frac{(\log n)^2}{2}} = \exp\left(-\frac{(\log n)^2}{2} \log\left(1 + \frac{\varepsilon_0^2}{\log n}\right)\right).$$

Using the fact that, $\log(1+x) = x - x^2/2 + o(x^2)$ when $x \to 0$, we get

$$A_1 \le e^{-\frac{\varepsilon_0^2 \log n}{2}} = n^{-\frac{\varepsilon_0^2}{2}}.$$

The last result allows us to get directly that, there exist some ε_0 and some ν', such that

$$A_1 \le c n^{-1-\nu'}. \tag{3.23}$$

Finally, by the results (3.23) and (3.22), we can deduce that

$$\sum_{n\ge 1} \mathbb{P}\left[\left|\mathbb{E}[\widetilde{g}_l(x)] - \widetilde{g}_l(x)\right| > \varepsilon_0 \sqrt{\frac{\log n}{n\varphi_x(h)}}\right] < \infty.$$

□

Proof of Corollary 3.3.1:

It is easy to remark that:

$$|\widetilde{g}_{2,n}(x)| \leq \frac{g_2(x)}{2},$$

which implies that

$$|g_2(x) - \widetilde{g}_{2,n}(x)| \geq \frac{g_2(x)}{2}.$$

We deduce, from Lemmas 3.3.2 and 3.3.3, that

$$\mathbb{P}\left(|\widetilde{g}_{2,n}(x)| \leq \frac{g_2(x)}{2}\right) \leq \mathbb{P}\left(|g_2(x) - \widetilde{g}_{2,n}(x)| > \frac{g_2(x)}{2}\right)$$

Consequently:

$$\sum_{n=1}^{\infty} \mathbb{P}\left(|\widetilde{g}_{2,n}(x)| < \frac{g_2(x)}{2}\right) < \infty.$$

□

Proof of Lemma 3.3.4:

It is easy to see that

$$\sqrt{n\varphi_x(h)}\left[\left(\widetilde{g}_2(x) - \mathbb{E}\left[\widetilde{g}_2(x)\right]\right)g_1(x) - \left(\widetilde{g}_1(x) - \mathbb{E}\left[\widetilde{g}_1(x)\right]\right)g_2(x)\right] = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} L_i(x),$$

where

$$L_i(x) := \frac{\sqrt{\varphi_x(h)}}{\mathbb{E}\left[K_1\right]}\left\{\frac{\delta_i}{\bar{G}\left(T_i\right)}K_i\left(g_1(x)T_i^{-2} - g_2(x)T_i^{-1}\right) - \mathbb{E}\left[\frac{\delta_i}{\bar{G}\left(T_i\right)}K_i\left(g_1(x)T_i^{-2} - g_2(x)T_i^{-1}\right)\right]\right\}. \tag{3.24}$$

The proof of this Lemma is based on the central limit theorem of Doukhan et al. (1994). We have then to consider the asymptotic behavior of the variance term, and the following assumption

$$\int_0^1 \alpha^{-1}(u)\left(Q_{L_1}(u)\right)^2 du < +\infty, \tag{3.25}$$

where Q_{L_1} is the "upper tail" quantile function defined by

$$Q_{L_1}(u) = \inf\left\{t \geq 0 : \mathbb{P}\left(L_1 > t\right) \leq u\right\},$$

and $\alpha^{-1}(u) = \sum_{n\in\mathbb{N}} \mathbb{1}_{u<\alpha_n}$.

Clearly,

$$\begin{aligned}\operatorname{Var}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}L_i(x)\right) &= n\varphi_x(h)\operatorname{Var}\left(\frac{g_1(x)}{n\mathbb{E}[K_1]}\sum_{i=1}^{n}\frac{\delta_i}{\bar{G}(T_i)}K_iT_i^{-2}-\frac{g_2(x)}{n\mathbb{E}[K_1]}\sum_{i=1}^{n}\frac{\delta_i}{\bar{G}(T_i)}K_iT_i^{-1}\right)\\ &= n\varphi_x(h)\Bigg(\operatorname{Var}[\widetilde{g}_1(x)]\,g_2^2(x)+\operatorname{Var}[\widetilde{g}_2(x)]\,g_1^2(x)\\ &\qquad -2g_1(x)g_2(x)\operatorname{Cov}[\widetilde{g}_1(x),\widetilde{g}_2(x)]\Bigg).\end{aligned}$$

By definition of, $\widetilde{g}_l(x)$ for $l=1;2$, we have

$$\begin{aligned}n\varphi_x(h)\operatorname{Var}[\widetilde{g}_l(x)] &= \frac{\varphi_x(h)}{(\mathbb{E}[K_1])^2}\operatorname{Var}\left[\frac{\delta_1}{\bar{G}(T_1)}K_1T_1^{-l}\right]\\ &\quad+\frac{\varphi_x(h)}{n(\mathbb{E}[K_1])^2}\sum_{i=1}^{n}\sum_{\substack{j=1\\|i-j|>0}}^{n}\operatorname{Cov}\left[\frac{\delta_i}{\bar{G}(T_i)}K_iT_i^{-l},\frac{\delta_j}{\bar{G}(T_j)}K_jT_j^{-l}\right]\\ &= \frac{\varphi_x(h)}{(\mathbb{E}[K_1])^2}J_{1,1}+\frac{\varphi_x(h)}{n(\mathbb{E}[K_1])^2}J_{2,n},\end{aligned}$$

where

$$\begin{aligned}J_{1,1} &= \operatorname{Var}\left[\frac{\delta_1}{\bar{G}(T_1)}K_1T_1^{-l}\right],\\ J_{2,n} &= \sum_{i=1}^{n}\sum_{\substack{j=1\\|i-j|>0}}^{n}\operatorname{Cov}\left[\frac{\delta_i}{\bar{G}(T_i)}K_iT_i^{-l},\frac{\delta_j}{\bar{G}(T_j)}K_jT_j^{-l}\right].\end{aligned}$$

By conditioning on the random variable X, by the same ideas in the proof of Lemma 3.3.2, and by using assumptions **(H5)**,**(N1)** and **(N5)**, we get

$$\begin{aligned}\mathbb{E}\left(\left(\frac{\delta_1}{\bar{G}(Y_1)}\right)^2Y_1^{-2l}K_1^2\right) &= \varphi_x(h)\mathbb{E}\left[\bar{G}^{-1}(Y_1)Y_1^{-2l}|X=x\right]\left(K^2(1)-\int_0^1\left(K^2(s)\right)'\chi_x(u)du\right)\\ &\quad+o(\varphi_x(h)),\end{aligned}$$

and

$$\mathbb{E}\left(\frac{\delta_1}{\bar{G}(Y_1)}Y_1^{-l}K_1\right)=O(\varphi_x(h)).$$

Thus,

$$\begin{aligned}\operatorname{Var}\left[\frac{\delta_1}{\bar{G}(T_1)}T_1^{-l}K_1\right] &= \varphi_x(h)\mathbb{E}\left[\bar{G}^{-1}(Y)Y^{-2l}|X=x\right]\left(K^2(1)-\int_0^1\left(K^2(s)\right)'\chi_x(u)du\right)\\ &\quad+O\left(\varphi_x^2(h)\right).\end{aligned}$$

We obtain

$$\frac{\varphi_x(h)}{(\mathbb{E}[K_1])^2} J_{1,1} \to \frac{q_{2l}(x)\beta_2}{\beta_1^2}. \tag{3.26}$$

Let us turn to $J_{2,n}$, for this we use the same technic used in Masry (1986). We define the same sets, E_1 and E_2, used in the proof of Lemma 3.3.3. Let $J_{2,n}^1$ and $J_{2,n}^2$, be the sums of covariances over E_1 and E_2 respectively. On the one hand, we have

$$J_{2,n}^1 = \sum_{E_1} \left| \mathrm{Cov}\left[\frac{\delta_i}{\bar{G}(T_i)} K_i T_i^{-l}, \frac{\delta_j}{\bar{G}(T_j)} K_j T_j^{-l} \right] \right| \le c \sum_{E_1} \left| \mathbb{E}[K_i K_j] - \mathbb{E}[K_i]\mathbb{E}[K_j] \right|.$$

Because of the assumptions of Lemma 3.3.3, we can write

$$J_{2,n}^1 \le c n \nu_n \varphi_x(h) \left(\left(\frac{\varphi_x(h)}{n} \right)^{\frac{1}{a}} + \varphi_x(h) \right).$$

On the other hand, for the summation over E_2, we use Davydov-Rio's inequality (Rio (1999), p.87) for mixing processes. This leads, for all $i \neq j$, to

$$|\mathrm{Cov}(K_i, K_j)| \le c\alpha(|i-j|).$$

Therefore,

$$\sum_{E_2} |\mathrm{Cov}(K_i, K_j)| \le n^2 \nu_n^{-a}.$$

The choice $\nu_n = \dfrac{1}{\varphi_x(h)\log(n)}$, motivated by the upper bound in **(H8)**, permits to get

$$\sum_{i \neq j}^{n} \mathrm{Cov}(K_i, K_j) = o(n\varphi_x(h)),$$

then

$$\frac{\varphi_x(h)}{n(\mathbb{E}[K_1])^2} J_{2,n} = o(1) \text{ as } n \longrightarrow \infty. \tag{3.27}$$

Thanks to (3.26) and (3.27), we have

$$n\varphi_x(h)\,\mathrm{Var}(\widetilde{g}_l(x)) \underset{n\longrightarrow\infty}{\longrightarrow} \frac{\beta_2 q_{2l}(x)}{\beta_1^2}. \tag{3.28}$$

Concerning the covariance term, we follow the same steps as for the variance (3.28), then we get

$$n\varphi_x(h)\,\mathrm{Cov}(\widetilde{g}_1(x), \widetilde{g}_2(x)) \underset{n\longrightarrow\infty}{\longrightarrow} \frac{\beta_2 q_3(x)}{\beta_1^2}. \tag{3.29}$$

Let us now prove the claimed result. Clearly, the function Q_{L_1} is nonincreasing, then

$$\sum_{n=1}^{\infty} \int_0^{\alpha_n} [Q_{L_1}(u)]^2 \, du \leq \sum_{n=1}^{\infty} \alpha_n Q_{L_1}^2(0).$$

By Hypotheses **(H1)**, **(H3)** and **(H5)**, we can write

$$c \frac{1}{\sqrt{\varphi_x(h)}} \leq |L_1| \leq c' \frac{1}{\sqrt{\varphi_x(h)}},$$

then

$$Q_{L_1}(0) \leq c' \frac{1}{\sqrt{\varphi_x(h)}}.$$

Therefore, we have

$$\sum_{i=1}^{\infty} \int_0^{\alpha_n} [Q_{L_1}(u)]^2 \, du \leq \sum_{n=1}^{\infty} \alpha_n \left(\varphi_x(h)\right)^{-1}.$$

It follows from **(H7)** and **(H8)**, that

$$\sum_{i=1}^{\infty} \int_0^{\alpha_n} [Q_{L_1}(u)]^2 \, du < \infty. \tag{3.30}$$

From (3.28), (3.29) and by noting $\sigma^2(x) = \frac{\left(q_2(x) - 2r(x)q_3(x) + r^2(x)q_4(x)\right)\beta_2}{\beta_1^2}$, we conclude that:

$$\operatorname{Var}\left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} L_i(x)\right) \overset{n\to\infty}{\to} \sigma^2(x). \tag{3.31}$$

Now, the Lemma can be easily deduced from (3.30), (3.31) and the central limit theorem of Doukhan et al. (1994):

$$\frac{1}{\sqrt{n g_2^2 \sigma^2(x)}} \sum_{i=1}^{n} L_i(x) =$$

$$\left(\frac{n\varphi_x(h)}{g_2^2(x)\sigma^2(x)}\right)^{1/2} \left(\left[\widetilde{g}_1(x) - \mathbb{E}\left[\widetilde{g}_1(x)\right]\right] g_2(x) - \left[\widetilde{g}_2(x) - \mathbb{E}\left[\widetilde{g}_2(x)\right]\right] g_1(x)\right) \overset{\mathcal{D}}{\to} \mathcal{N}(0,1).$$

□

Proof of Lemma 3.3.5:

As in Ferraty et al. (2007), we show that

$$\mathbb{E}[\widetilde{r}_n(x)] = \frac{\mathbb{E}\left[\widetilde{g}_1(x)\right]}{\mathbb{E}\left[\widetilde{g}_2(x)\right]} + O\left(\frac{1}{n\varphi_x(h)}\right).$$

So, it suffices to evaluate $\mathbb{E}[\widetilde{g}_l(x)]$ for $l \in \{1,2\}$, we obtain

$$\begin{aligned}\mathbb{E}\left[\widetilde{g}_1(x)\right] &= \frac{1}{\mathbb{E}\left[K_1\right]}\mathbb{E}\left(K_1(x)\mathbb{E}\left[Y_1^{-l}|X_1\right]\right)\\ &= \frac{1}{\mathbb{E}\left[K_1\right]}\left(g_l(x)\mathbb{E}\left[K_1\right]+\mathbb{E}\left[K_1\mathbb{E}\left(g_l\left(X_1\right)-g_l(x)|d\left(X_1,x\right)\right)\right]\right)\\ &= g_l(x)+\frac{\int_0^1 K(t)\Psi_l(ht)\mathrm{d}\mathbb{P}^{d(x,X)/h}(t)}{\int_0^1 K(t)\mathrm{d}\mathbb{P}^{d(x,X)/h}(t)}.\end{aligned}$$

By using the first-order Taylor's expansion for $\Psi_l(.)$ around 0, where $\Psi_l(0)=0$, we have

$$\mathbb{E}\left[\widetilde{g}_l(x)\right]=g_l(x)+h\Psi_l'(0)\left[\frac{\int_0^1 tK(t)\mathrm{d}\mathbb{P}^{d(x,X)/h}(t)}{\int_0^1 K(t)\mathrm{d}\mathbb{P}^{d(x,X)/h}(t)}\right]+o(h).$$

According to Lemma 2 in Ferraty et al. (2007), we get, under **(N1)**

$$\frac{\int_0^1 tK(t)\mathrm{d}\mathbb{P}^{d(x,X)/h}(t)}{\int_0^1 K(t)\mathrm{d}\mathbb{P}^{d(x,X)/h}(t)}\longrightarrow\frac{\beta_0}{\beta_1}\text{ and }\int_0^1 K(t)\mathrm{d}\mathbb{P}^{d(x,X)/h}(t)\longrightarrow\beta_1.$$

Consequently

$$\mathbb{E}\left[\widetilde{g}_l(x)\right]=g_l(x)+h\Psi_l'(0)\frac{\beta_0}{\beta_1}+o(h),$$

then we can deduce that,

$$A_n=\frac{\mathbb{E}\left[\widetilde{g}_1(x)\right]}{\mathbb{E}\left[\widetilde{g}_2(x)\right]}-r(x)=hB_n+o(h).$$

□

Proof of Lemma 3.3.6:

Thanks to Lemmas 3.3.4 and 3.3.5's results, we have:

$$\mathbb{E}\left[\widetilde{g}_2(x)-g_2(x)\right]\to 0,$$

and

$$\mathrm{Var}\left[\widetilde{g}_2(x)\right]\to 0.$$

Then,

$$\widetilde{g}_2(x)-g_2(x)\to 0,\ \text{in probability}.$$

Next, for the last needed convergence, we obtain by the same manner:

$$\mathbb{E}\left[\left(\frac{n\varphi_x(h)}{g_1(x)^2\sigma^2(x)}\right)^{1/2} A_n\left(\widetilde{g}_2(x)-\mathbb{E}\left[\widetilde{g}_2(x)\right]\right)\right]=0,$$

and

$$\operatorname{Var}\left[\left(\frac{n\varphi_x(h)}{g_1(x)^2\sigma^2(x)}\right)^{1/2} A_n\left(\widetilde{g}_2(x)-\mathbb{E}\left[\widetilde{g}_2(x)\right]\right)\right]=O(A_n^2)=O(h^2)\to 0.$$

Consequently,

$$\left(\frac{n\varphi_x(h)}{g_1(x)^2\sigma^2(x)}\right)^{1/2} A_n\left(\widetilde{g}_2(x)-\mathbb{E}\left[\widetilde{g}_2(x)\right]\right)\to 0,\ \text{in probability}.$$

which completes the proof. □

BIBLIOGRAPHY

B. Altendji, J. Demongeot, A. Laksaci, and M. Rachdi. Functional data analysis: estimation of the relative error in functional regression under random left-truncation model . *Journal of Nonparametric Statistics*, 30:472–490, 2018.

A. Amiri and S. Khardani. Nonparametric regression estimation using recursive kernel for dependent censored data. *Int J Math Stat*, 19:52–82, 2018.

A. Amiri, C. Crambes, and B. Thiam. Recursive estimation of nonparametric regression with functional covariate. *Computational Statistics & Data Analysis*, 69:154–172, 2014.

M. Attouch, A. Laksaci, and E. Ould-Saïd. Asymptotic Distribution of Robust Estimator for Functional Nonparametric Models. *Communications in Statistics Theory and Methods*, 38:1317–1335, 2009.

M. Attouch, A. Laksaci, and N. Messabihi. Nonparametric relative error regression for spatial random variables. *Stat Papers DOI*, 58, 2017.

J. Buckley and I. James. Linear regression with censored data. *Biometrika*, 66:429–436, 1979.

M. Campbell and A. Donner. Classification efficiency of multinomial logistic regression relative to ordinal logistic regression . *J. Amer. Statist. Asso*, 84:587–591, 1989.

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

M. Chaouch and S. Khardani. Randomly censored quantile regression estimation using functional stationary ergodic data. *Journal of Nonparametric Statistics*, 27:65–87, 2015.

S. Dabo-Niang. Kernel density estimator in an infinite–dimensional space with a rate of convergence in the case of diffusion process. *Applied Mathematics Letters*, 17:381–386, 2004.

P. Deheuvels and J. H. J. Einmahl. Functional limit laws for the increments of Kaplan Meier product-limit processes and applications . *The Annals of Probability*, 28:1301–1335, 2000.

J. Demongeot, A. Hamie, A. Laksaci, and M. Rachdi. Relative-Error Prediction in Nonparametric Functional Statistics: Theory and Practice . *Journal of Multivariate analysis*, 147:261–268, 2016.

P. Doukhan, P. Massart, and E. Rio. The functional central limit theorem for strongly mixing processes. In *Annales de l'IHP Probabilités et statistiques*, volume 30, pages 63–82, 1994.

F. Ferraty and P. Vieu. Curves discrimination: a nonparametric functional approach. *Computational Statistics and Data Analysis*, 44:161–173, 2003.

F. Ferraty and P. Vieu. Nonparametric methods for functional data, with applications in regression, time–series prediction and curves discrimination. *J. Nonparametr. Stat*, 16: 111–125, 2004.

F. Ferraty and P. Vieu. *Nonparametric Functional Data Analysis: Theory and Practice*. Springer Series in Statistics, New York, springer edition, 2006.

F. Ferraty, A. Goia, and P. Vieu. Functional nonparametric model for time series: a fractal approach for dimension reduction. *Test*, 11:317–344, 2002.

F. Ferraty, A. Mas, and P. Vieu. Nonparametric Regression on Functional Data: Inference and Practical Aspects. *Australian and New Zealand Journal of Statistics*, 49:267–286, 2007.

O. Fetitah, I. M. Almanjahie, M. K. Attouch, and A. Righi. Strong convergence of the functional nonparametric relative error regression estimator under right censoring. *Mathematica Slovaca*, 70(6):1469–1490, 2020.

W. Horrigue and E. Ould-Saïd. Nonparametric Regression Quantile Estimation for Dependant Functional Data under Random Censorship: Asymptotic Normality . *Communications in Statistics - Theory and Methods*, 44:4307–4332, 2014.

M. Jones, P. Heungsun, S. Key-Il, and S. Vines. Relative error prediction via kernel regression smoothers. . *J.Statist. Plann. Inference*, 138:2887–2898, 2008.

S. Khardani and B. Thiam. Strong consistency result of a non parametric conditional mode estimator under random censorship for functional regressors . *Communications in Statistics-Theory and Methods*, 45:1863–1875, 2016.

E. Masry. Recursive probability density estimation for weakly dependent stationary processes. *IEEE Transactions on Information Theory*, 32:254–267, 1986.

E. Masry. Nonparametric regression estimation for dependent functional data: asymptotic normality. *Stochastic Processes and their Applications*, 115:155–177, 2005.

W. Mechab and A. Laksaci. Nonparametric relative regression for associated random variables. *Metron*, 74:75–97, 2016.

H. Park and L. Stefanski. Relative-error prediction. *Statistics & probability letters*, 40: 227–236, 1998.

J. O. Ramsay and C. J. Dalzell. Some tools for functional data analysis. *Journal of the Royal Statistical Society: Series B (Methodological)*, 53:539–561, 1991.

J. O. Ramsay and B. W. Silverman. Principal components analysis for functional data. In *Functional Data Analysis*, pages 147–172. Springer, 2005.

E. Rio. *Théorie asymptotique des processus aléatoires faiblement dépendants*, volume 31. Springer Science & Business Media, 1999.

W. Stute. Consistent estimation under random censorship when covariables are present. *Journal of Multivariate Analysis*, 45:89–103, 1993.

CHAPTER 4

ROBUST NONPARAMETRIC EQUIVARIANT REGRESSION FOR FUNCTIONAL DATA WITH RESPONSES MISSING AT RANDOM

This chapter is the subject of an article submitted.

Robust nonparametric equivariant regression for functional data with responses missing at random

[1]**Omar FETITAH**, [2]**Mohammed Kadi ATTOUCH**, [3]**Salah KHARDANI** and [1]**Ali RIGHI**.

[1] Laboratory of Statistics and Stochastic Processes University of Djillali Liabes BP 89, Sidi Bel Abbes 22000, Algeria.

[2]Department of Mathematics Faculty of Sciences King Khalid University PO. Box 9004, Abha, Saudi Arabia.

[3] Laboratoire des Réseaux Intelligents et Nanotechnologie. Ecole Nationale des Sciences et Technologies Avancées á Borj-Cédria, Tunisia.

Abstract : The paper deal with the robust nonparametric regression in a functional space when the response variables are missing at random (MAR), for both cases, without and with unknown scale parameter. We establish, the almost complete convergence rate of our estimators the two proposed models. Some simulations study is given to illustrate the higher predictive performances of our proposed method.

Keywords : Robust regression, Functional data analysis, Almost complete convergence, Missing data, Scale parameter.

Mathematics Subject Classification: Primary 62G08, 62G35, 62G20. Secondary 49N30.

4.1 Introduction

The regression function plays an important role in the nonparametric prediction. Indeed, it provides a very informative summary of the relationship between the variable of interest Y and the covariate X. There is an extensive literature on the regression function estimation when the data are incomplete given a functional random covariate (i.e. valued in an $\mathcal{F}$ space on a real interval).

In this paper, we consider the problem of the co-variability analysis between a functional variable X and a scalar response variable Y which is not completely observed. We use a non parametric approach to model this relationship by constructing an estimator of the robust regression function when missing data occur in the response variable.

The purpose of this paper is to generalize the estimators proposed by Attouch et al. (2009), Boente and Fraiman (1989) and Boente and Vahnovan (2015) to the case of missing data and for both cases, with and without unknown scale parameters. Moreover, one can check that the estimators proposed by Ferraty et al. (2013) are a special case of our proposed estimators.

In the literature, functional data analysis (FDA) has received considerable interest in the statistical area because of its wide applications in many practical fields such as climatology, economics and medicine. For more details related to this topic, one could refer to the monographs of Ramsay and Silverman (2005), Bosq (2012), Ferraty and Vieu (2006), Yao et al. (2005) and Cai et al. (2006).

As Known, the regression function is used to study the relationship between two variables denoted by,

$$r(x) = \mathbb{E}[Y|X = x], \quad x \in \mathcal{F}. \tag{4.1}$$

The link function in (4.1) is of the form

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

where ϵ is a random variable such that $\mathbb{E}[\epsilon|X] = 0$ and $\mathbb{E}[\epsilon^2|X] = \sigma^2(X) < +\infty$.

However, in many practical situations such as in pharmaceutical tracing, econometric life-test study or reliability, data are often incompletely observed, and part of the responses are missing at random (MAR). In the regression method with missing data, a standard approach is to attribute the incomplete observations, then proceed to estimate the conditional or unconditional mean of the response variable with the completed sample.

It is well-known that the above regression methods are outlier-sensitive. The treatment of outliers is an essential step in highlighting the features of any data set. In this situation, outlying observations can be even more dangerous since the shape of the estimated curve is highly sensitive to outlying observations.

Therefore, in order to overcome this problem, we consider a robust approach. More precisely, we are interested in the class of M-estimates which was introduced by Huber (1965). In the statistical literature, several papers have been devoted to the study the properties of the nonparametric M-estimator defined as a solution of (4.3) when the variable of interest Y is completely observed. One can refer, among others, to Laïb and Ould-Saïd (2000) for stationary ergodic processes, Collomb and Härdle (1986), Boente and Fraiman (1989), Boente and Rodriguez (2006) for mixing processes, Huber (1992) and

Härdle (1984) for the independent and identically distributed (i.i.d.) case. Härdle and Gasser (1984) and Tsybakov (1982) also studied pointwise asymptotic properties of a robust version of the Nadaraya−Watson method. These results were extended to M−type scale equivariant kernel estimates by Boente and Fraiman (1989) and Härdle et al. (1988) who also considered robust equivariant nonparametric estimates using nearest neighbor weights. While Gheriballah et al. (2013) established the almost complete convergence with rate in the setting of functional and stationary ergodic data.

Inspired by all the papers above, our work in this paper aims to contribute to the research on functional nonparametric regression model by giving (provide) an alternative estimation of regression based on missing data. More precisely, we construct a robust regression estimator of a missing scalar response and a functional covariate. Noting that in nonparametric modeling, robust regression is an essential regression analysis tool because it is less sensitive to outliers in the data compared with the classical regression. On the other hand, the statistical analysis of infinite-dimensional data has been the subject of several works in the recent statistical literature.

Moreover, the regression analysis of incomplete data (missing data) has gained a particular interest in the statistics literature. Such kinds of data occur in many fields of applications such as in astronomy, economics, epidemiology, biometry, and medical studies. Consider n independent pairs of random variables (X_i, Y_i) for $i = 1, ..., n$ that we assume drawn from the pair (X, Y). The latter is valued in $\mathcal{F} \times \mathbb{R}$, where $\mathcal{F}$ is a semi-metric space and d denotes a semi-metric. Our main goal is to study the co-variation between X_i and Y_i by the nonparametric robust regression function.

For $x \in \mathcal{F}$ the nonparametric robust regression, denoted by θ_x, is defined as the unique minimizer of

$$\theta_x = \arg\min_{t\in\mathbb{R}} \mathbb{E}\left[\rho\left(Y - t\right)|X = x\right] \tag{4.2}$$

where $\rho(.)$ is a real-valued Borel function satisfying some regularity conditions, for more information we refer the reader to Maronna and Martin (2006). This kind of models belongs to the class of M-estimates introduced by Huber (1992). It includes many usual nonparametric models. For example, when $\rho(y) = y^2$ we obtain the classical regression, $\rho(y) = |y||\alpha - \mathbb{1}_{y<0}|$ leads to the α^{th} conditional quantile. In this case, Chaouch and Khardani (2015) considered the conditional quantile estimation based on functional stationary ergodic data.

The paper is organized as follows. In Section 4.2, we describe our model in two cases and precisely construct the robust estimator of $r(\cdot)$ based on the functional stationary data with MAR and we give the main results. In Sections 4.3, we illustrate our methodology by a simulation study to compare the classical nonparametric functional model with the complete data and the model with MAR. Finally, the proofs of the main results are postponed to Section 4.4.

4.2 Model

4.2.1 Robust estimation under complete data

Let (X, Y) be a pair of random variables (rv) in $\mathcal{F} \times \mathbb{R}$, where $(\mathcal{F}, d)$ is a semi-metric space (i.e. X is a functional random variable (f.r.v) and d a semi-metric). Let x (resp. t) be a fixed element of $\mathcal{F}$ (resp. $\mathbb{R}$).

In this framework, let ψ be a real-valued Borel function satisfying some regularity conditions to be stated below. We denote by θ_x, which is implicitly defined as a zero with respect to (w.r.t.) t of the equation

$$\Psi(x,t) := \mathbb{E}\left(\psi(Y-t)|X=x\right). \tag{4.3}$$

The robust nonparametric estimator of θ_x is given by the solution $\widetilde{\theta}_x$ of $\widetilde{\Psi}(x,t) = 0$, where

$$\widetilde{\Psi}(x,t) = \frac{\sum_{i=1}^{n} K(h^{-1}d(x,X_i))\psi\left(Y_i - t\right)}{\sum_{i=1}^{n} K(h^{-1}d(x,X_i))} \tag{4.4}$$

where K is real-valued kernel function and $h = h_n > 0$ is a smoothing parameter satisfying $h_n \to 0$ as $n \to \infty$.

4.2.2 Robust estimation under missing data

Now, we will focus on the case of missing response, one has an incomplete sample of size n from (X, Y, δ) which is classically denoted by $\{(X_i, Y_i, \delta_i), 1 \le i \le n\}$, where $\delta_i = 1$ if Y_i is observed, and $\delta_i = 0$ otherwise. The Bernoulli random variable δ is supposed to be such that

$$\mathbb{P}(\delta = 1|X = x, Y = y) = \mathbb{P}(\delta = 1|X = x) = p(x), \tag{4.5}$$

where $p(x)$ is a functional operator. This last condition models the fact that the censoring process δ is, conditionally on X, independent of the response Y.

In the missing model only, the $(X_i, Y_i, \delta_i)_{1 \le i \le n}$ are observed, the robust nonparametric estimator of θ_x is given by the solution $\widehat{\theta}_x$ of $\widehat{\Psi}(x,t) = 0$, where $\widehat{\Psi}(x,t)$ as an estimate of

$\Psi(x,t)$ defined by

$$\widehat{\Psi}(x,t)=\frac{\sum_{i=1}^{n}\delta_i K(h^{-1}d(x,X_i))\psi\left(Y_i-t\right)}{\sum_{i=1}^{n}\delta_i K(h^{-1}d(x,X_i))}:=\frac{\widehat{\Psi}_N(x,t)}{\widehat{\Psi}_D(x)},\tag{4.6}$$

with

$$\widehat{\Psi}_N(x,t)=\frac{\sum_{i=1}^{n}\delta_i K(h^{-1}d(x,X_i))\psi\left(Y_i-t\right)}{n\mathbb{E}(K(h^{-1}d(x,X_1)))}$$

and

$$\widehat{\Psi}_D(x)=\frac{\sum_{i=1}^{n}\delta_i K(h^{-1}d(x,X_i))}{n\mathbb{E}(K(h^{-1}d(x,X_1)))}.$$

Assumptions and main results

The main purpose of this section is to study the almost-complete convergence[1] (a.co.) of $\widehat{\theta}_x$ toward θ_x.

From now on, for all x in $\mathcal{F}$ and for all positive real h, when no confusion is possible, we will denote by C and C' some strictly positive generic constants and by:

$$K_i(x)=K\left(\frac{d(x,X_i)}{h}\right)\quad for\ \ i=1,...,n,$$

where K is a kernel function and $h:=h_{n,K}$ is a sequence of positive numbers decreasing toward 0. We will also use the notation:

$$\varphi_x(h)=\mathbb{P}(X\in B(x,h)),\tag{4.7}$$

where $B(x,h)=\{x'\in\mathcal{F},d(x',x)\le h\}$.

In what follows, we will need the following assumptions:

(H1) $\mathbb{P}(X\in B(x,h))=:\varphi_x(h)>0$ for all $h>0$ and $\lim_{h\to 0}\varphi_x(h)=0$ and

$$0<C\varphi_x(h)\le\mathbb{P}(X\in B(x,h))\le C'\varphi_x(h).$$

(H2) The bandwidth h satisfies: $\dfrac{n\varphi_x(h)}{\log n}\to\infty\ \ as\ \ n\to\infty.$

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

(H3) The kernel K is a bounded and continuous function on its support $(0;1)$ and satisfying:

$$0 < C \leq K(.) \leq C' < +\infty.$$

(H4) The function Ψ is such that:

i) Ψ is of class $\mathcal{C}^1$ on $[\theta_x - \kappa, \theta_x + \kappa]$ $\forall \kappa > 0$.

ii) $\Psi(.,.)$ satisfies Hölder's condition with first variable, that is : there exists strictly positive constants b, b' and κ, such that:

$$\forall (t_1, t_2) \in [\theta_x - \kappa, \theta_x + \kappa]^2, \forall\, (x_1, x_2) \in \mathcal{N}_x \times \mathcal{N}_x,$$
$$|\Psi\,(x_1, t) - \Psi\,(x_2, t)| \leq C\Big(d^b\,(x_1, x_2) + |t_1 - t_2|^{b'}\Big).$$

iii) For each fixed $t \in [\theta_x - \kappa, \theta_x + \kappa]$, the function $\Psi(., t)$ is continuous at the point x.

iv) $p(x)$ is positive continuous in a neighbourhood of x.

(H5) The function ψ is strictly monotone, and

$$\mathbb{E}\left[|\psi(Y - t)|^m | X\right] < C < \infty, \quad m \geq 1.$$

Theorem 4.2.1. *Assume that the Assumptions **(H1)-(H5)**are satisfied, then $\widehat{\theta}_x$ exists and is unique a.co. for all sufficiently large n. Furthermore, if $\Psi'\,(x, \theta_x) \neq 0$*

$$\left|\widehat{\theta}_x - \theta_x\right| = O_{a.co.}\left(h^b + \sqrt{\frac{\log n}{n\varphi_x(h)}}\right). \tag{4.8}$$

4.2.3 Robust equivariant estimation under missing data

Our robustification method allows us to consider the functional nonparametric regression model with a scale of the error assumed to be unknown by taking $\psi(x, t) = \psi(\frac{x-t}{\sigma})$, where σ is a measure of spread for the conditional distribution of Y given $X = x$.

In this case, we denote by ϑ_x as a zero with respect to (w.r.t.) t of the equation

$$\Gamma(x, t, \sigma) := \mathbb{E}\left(\psi\left(\frac{Y - t}{\sigma}\right)|X = x\right) = 0, \tag{4.9}$$

where σ is a robust measure of the conditional scale. This measure can be taken as the conditional median of the absolute deviation from the conditional median, that is,

$$\sigma := s(x) = \mathrm{MED}(|Y - m(x)||X = x) = \mathrm{MAD}_C\,(F_Y^x(\cdot)) \tag{4.10}$$

where $F_Y^x(.) = F(.|X = x) = \mathbb{P}(Y \le .|X = x)$ the conditional distribution of Y given $X = x$ and $m(x) = \text{MED}(Y|X = x) = \inf\left\{y \in \mathbb{R} : F(y|X = x) \ge \frac{1}{2}\right\}$ is the median of the conditional distribution.

In the missing model only, the $(X_i, Y_i, \delta_i)_{1\le i\le n}$ are observed, noting that the conditional distribution

$$F(y|X = x) = \mathbb{P}(Y \le y|X = x) = \mathbb{E}(\mathbb{1}_{(-\infty;y]}(Y)|X = x),$$

where $\mathbb{1}_A$ denotes the indicator function of the set A, and using the Nadaraya-Watson estimator an estimator $\widehat{F}(y|X = x)$ of $F(y|X = x)$ can be defined as

$$\widehat{F}(y|X = x) = \frac{\sum_{i=1}^{n} \delta_i K(h^{-1}d(x, X_i))\mathbb{1}_{(-\infty;y]}(Y_i)}{\sum_{i=1}^{n} \delta_i K(h^{-1}d(x, X_i))} := \frac{\widehat{R}_N(x, y)}{\widehat{\Psi}_D(x)}, \tag{4.11}$$

where

$$\widehat{R}_N(x, y) = \frac{\sum_{i=1}^{n} \delta_i K(h^{-1}d(x, X_i))\mathbb{1}_{(-\infty;y]}(Y_i)}{n\mathbb{E}(K(h^{-1}d(x, X_1)))}$$

and $\widehat{\Psi}_D(x)$ is define previously.

Denote by $\widehat{s}(x)$ a robust estimator of the conditional scale, for instance, $\widehat{s}(x) = \text{MAD}_\text{C}(\widehat{F}(\cdot|X = x))$, the scale measure. On the other hand, the robust nonparametric estimator of ϑ_x is given by the solution $\widehat{\vartheta}_x$ of $\widehat{\Gamma}(x, ., \widehat{s}(x)) = 0$, where $\widehat{\Gamma}(x, t, \widehat{s}(x))$ as an estimate of $\Gamma(x, t, s(x))$ by

$$\widehat{\Gamma}(x, t, \widehat{s}(x)) = \frac{\sum_{i=1}^{n} \delta_i K(h^{-1}d(x, X_i))\psi\left(\frac{Y_i - t}{\widehat{s}(x)}\right)}{\sum_{i=1}^{n} \delta_i K(h^{-1}d(x, X_i))} := \frac{\widehat{\Gamma}_N(x, t, \widehat{s}(x))}{\widehat{\Psi}_D(x)}, \tag{4.12}$$

with

$$\widehat{\Gamma}_N(x, t, \widehat{s}(x)) = \frac{\sum_{i=1}^{n} \delta_i K(h^{-1}d(x, X_i))\psi\left(\frac{Y_i - t}{\widehat{s}(x)}\right)}{n\mathbb{E}(K(h^{-1}d(x, X_1)))}.$$

Assumptions and main results

The main goal of this section is to study the almost-complete convergence (a.co.) of $\widehat{\vartheta}_x$ toward ϑ_x.

In what follows, we will need the following assumptions:

(E1) $\mathbb{P}(X \in B(x,h)) =: \varphi_x(h) > 0$ for all $h > 0$ and $\lim_{h\to 0} \varphi_x(h) = 0$ and

$$0 < C\phi(h) \le \mathbb{P}(X \in B(x,h)) \le C'\phi(h).$$

(E2) The bandwidth h satisfies: $\dfrac{n\varphi_x(h)}{\log n} \to \infty \ \ as \ \ n \to \infty.$

(E3) The kernel K is a bounded and continuous function on its support $(0;1)$ and satisfying:

$$0 < C \le K(.) \le C' < +\infty.$$

(E4) Let $\mathcal{N}_x$ be a neighborhood of the point x:

i) $F(.|X = x)$ is continuous function symmetric around ϑ_x.

ii) $F(y|X = x)$ has a unique median $m(x)$.

iii) for any fixed $y \in [\vartheta_x - \kappa, \vartheta_x + \kappa]$, $F(y|X = .) : \mathcal{F} \to [0;1]$ is a continuous function.

iv) Let be $\mathcal{N}_x$ a neighborhood of the point $x \in \mathcal{F}$, there exists a $b_1 > 0$ and $b_2 > 0$ such that:

$$\forall\,(y_1,y_2) \in [\vartheta_x - \kappa, \vartheta_x + \kappa] \times [\vartheta_x - \kappa, \vartheta_x + \kappa], \forall\,(x_1,x_2) \in \mathcal{N}_x \times \mathcal{N}_x,$$
$$|F(y_1|X = x_1) - F(y_2|X = x_2)| \le C\left(d^{b_1}(x_1,x_2) + |y_1 - y_2|^{b_2}\right).$$

(E5) The function Γ is such that:

i) The function Γ is of class $\mathcal{C}^1$ on $[\vartheta_x - \kappa, \vartheta_x + \kappa]$.

ii) $\Gamma(.,.,\sigma)$ satisfies Hölder's condition with the first variable, that is: there exist strictly positive constants b and κ such that :

$$\forall t \in [\vartheta_x - \kappa, \vartheta_x + \kappa], \forall \sigma > 0, \forall\,(x_1,x_2) \in \mathcal{N}_x \times \mathcal{N}_x,$$
$$|\Gamma(x_1,t,\sigma) - \Gamma(x_2,t,\sigma)| \le Cd^b(x_1,x_2).$$

iii) For each fixed $t \in [\vartheta_x - \kappa, \vartheta_x + \kappa]$ and $\sigma > 0$, the function $\Gamma(.,t,\sigma)$ is continuous at the point x.

iv) $p(x)$ is positive continuous in a neighbourhood of x.

(E6) $\psi : \mathbb{R} \to \mathbb{R}$ is an odd function, strictly monotone, bounded and continuous differentiable, with bounded derivative ψ' such that $\zeta(u) = u\psi'(u)$ is bounded.

Property 4.2.1. *Under Assumptions **(E1)-(E4)** and **(E5 iv)**, we obtain*

$$\sup_{y\in[\vartheta_x-\kappa,\vartheta_x+\kappa]} \left|\widehat{F}(y|X=x) - F(y|X=x)\right| = O_{a.co.}\left(h^{b_1} + \sqrt{\frac{\log n}{n\varphi_x(h)}}\right). \tag{4.13}$$

Lemma 4.2.1. *Under Assumptions of proposition 4.2.1 and if*

$$\sup_{y\in[\vartheta_x-\kappa,\vartheta_x+\kappa]} \left|\widehat{F}(y|X=x) - F(y|X=x)\right| \to 0$$

there exist positive constants $A \le B$, such that, $\widehat{s}(x) = \mathrm{MAD_C}(\widehat{F}(\cdot|X=x))$ verifies

$$A \le \widehat{s}(x) \le B \, for \, n \ge n_0. \tag{4.14}$$

Theorem 4.2.2. *Under Assumptions **(E1)-(E3)** and **(E5)-(E6)**, then $\widehat{\vartheta}_x$ exists and is unique $a.co.$ for all sufficiently large n. Furthermore, if $\Gamma'(x, \vartheta_x, \widehat{s}(x)) \neq 0$, we have*

$$\left|\widehat{\vartheta}_x - \vartheta_x\right| = O_{a.co.}\left(h^{b} + \sqrt{\frac{\log n}{n\varphi_x(h)}}\right). \tag{4.15}$$

4.3 Numerical study

In this section, we discuss a simulation example to check the behavior of our estimator. The functional variable X is taken as a function with support $[0,1]$ whose based on the following observation:

$$X_i(t) = A_i t^2 + \cos(\pi B_i t) \qquad i = 1,\dots,200; t \in [0,1];$$

where A_i are i.i.d. $\sim U(0,1)$ and B_i are i.i.d. $\sim \mathcal{N}(0,1)$, and are independent from A_i and B_i. For simplicity, figure4.1 presents a sample of $n = 200$ of the covariable curves $X(t)$. We define the response variable Y by $Y = r(X) + \epsilon$, where r is the regression operator with

$$r(x) = \left(\int_0^1 x'(t)\mathrm{d}t\right)^2. \tag{4.16}$$

and $\epsilon \sim \mathcal{N}(0, 0.075)$.

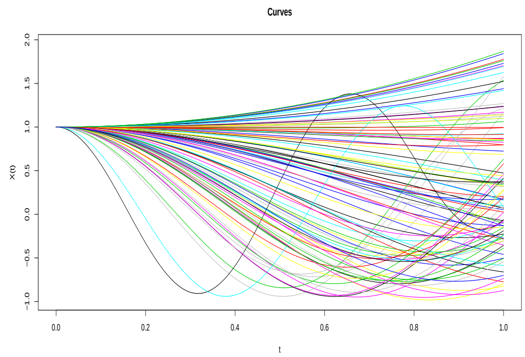

Figure 4.1: The curves $X_{i=1,\dots,100}(t), \ \ t \in [0,1]$.

Our main goal is to compare the sensitivity of the classical classical methods to outliers estimated by

$$\widehat{r}_x = \frac{\sum_{i=1}^{n} \delta_i Y_i K\left(h_n^{-1} d\left(x, X_i\right)\right)}{\sum_{i=1}^{n} \delta_i K\left(h_n^{-1} d\left(x, X_i\right)\right)} \tag{4.17}$$

and the robust regression estimator $\widehat{\vartheta}_x$ where

$$\widehat{\vartheta}_x = \arg\min_t \frac{\sum_{i=1}^{n} \delta_i K(h^{-1} d(x, X_i)) \psi\left(\frac{Y_i - t}{\widehat{s}(x)}\right)}{\sum_{i=1}^{n} \delta_i K(h^{-1} d(x, X_i))},$$

with $\widehat{s}(x) = \mathrm{MAD_C}(\widehat{F}(\cdot|X = x))$. We take $\psi(t) = \frac{t}{\sqrt{1+t^2/2}}$. For more details related to the choice of the score function, we refer the reader to Attouch et al. (2009). We choose the semi-metric on $\mathcal{F}$:

$$d\left(x_i, x_j\right) = \sqrt{\int_0^1 \left(x_i'(t) - x_j'(t)\right)^2 \mathrm{d}t}, \quad \text{for } \forall x_i, x_j \in \mathcal{F}.$$

We choose the quadratic kernel defined as:

$$K(x) = \frac{3}{2}\left(1 - x^2\right) \mathbb{1}_{(0,1)}.$$

Then, we split the sample of size 200 into a learning sub-sample $(X_i, Y_i), i = 1, \dots, 150$ and a testing sub-sample $(X_j, Y_j), j = 151, \dots, 200$. For the missing mechanism, we adopted it as in Ferraty et al. (2013):

$$p(x) = \mathbb{P}(\delta = 1|X = x) = \mathrm{expit}\left(2\alpha \int_0^1 x^2(t)\mathrm{d}t\right),$$

where $\mathrm{expit}(u) = e^u/(1 + e^u)$ for $\forall u \in \mathbb{R}$. The parameter α controls the degree of dependency between the functional curve X and the variable δ. To keep control the quantity $p(x)$, we compute $\overline{\delta} = 1 - \frac{1}{150}\sum_{i=1}^{150}\delta_i$. The optimal bandwidth h is selected by the cross-validation method for the k nearest neighbours ($k-$NN) in a local way (see Ferraty and Vieu (2006) for more details). Then we calculate the $\widehat{r}_{X_j}$ and $\widehat{\vartheta}_{X_j}$ for $j = 151, \ldots, 200$.

To highlight the performance of our results, we plot the true values versus, the predicted values for the MSE for both cases complete data and response missing at random MAR, respectively.

1. Complete case, the mean square error (CMSE) is

$$\mathrm{CMSE}_{class} = \frac{1}{50}\sum_{j=151}^{200}\left(\widetilde{r}_{X_j} - r(X_j)\right)^2 \text{ and } \mathrm{CMSE}_{robust} = \frac{1}{50}\sum_{j=151}^{200}\left(\widetilde{\vartheta}_{X_j} - r(X_j)\right)^2$$

2. Incomplete case response MAR, the mean square error (MMSE) is

$$\mathrm{MMSE}_{class} = \frac{1}{50}\sum_{j=151}^{200}\left(\widehat{r}_{X_j} - r(X_j)\right)^2 \text{ and } \mathrm{MMSE}_{robust} = \frac{1}{50}\sum_{j=151}^{200}\left(\widehat{\vartheta}_{X_j} - r(X_j)\right)^2$$

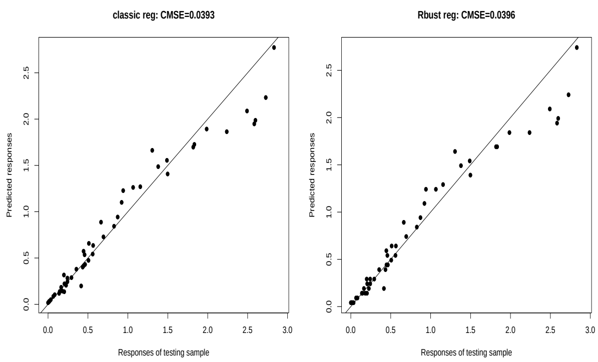

Figure 4.2: The complete data case: $CMSE$.

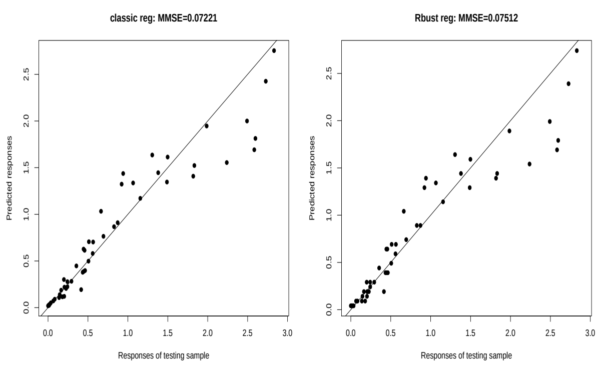

Figure 4.3: The missing at random case: $MMSE$.

Table 4.1: MMSE comparison between both methods for the combinations of parameters (α).

α	$\overline{\delta}$	MMSE$_{class}$	MMSE$_{robust}$
0	0.50	0.0816188	0.0827697
0.5	0.33	0.07811728	0.0798221
1	0.23	0.07026415	0.07212116
1.5	0.15	0.06799575	0.06894988
2	0.10	0.06621263	0.06748479

Table 4.2: MMSE for the Classical Kernel Estimator and the Robust Estimator according to numbers of introduced artificial outliers.

number of artificial outliers	0	10	20	40
Classical Estimator $MMSE_{class}$	0.04605103	34.41909	104.3417	1112.265
Robust Estimator $MMSE_{Robust}$	0.04678676	0.08507463	0.1022152	0.422661

As illustrated in Figures 4.2 and 4.3, one could notice, that there is no difference between the two methods: the Classical Kernel method and the Robust method. On the other hand, we can see that our estimator $\widehat{\vartheta}_x$ on MAR works almost as well as if we had the complete dataset and using $\widetilde{\vartheta}_x$.

In all the following, we have picked $n = 200$ and have treated different values for the parameter α. We carried 100 independent replications of the model, for each value of α, and we have computed the mean squared error (MMSE) of both estimators. The results are shown in Table 4.1.

In Table 4.1, we can see the good performance of both estimators $\widehat{r}_x$ and $\widehat{\vartheta}_x$ when $\overline{\delta}$ is small, but it goes bad when $\overline{\delta}$ is bigger.

Now, we will move to compare the performance of both estimators(classic and robust) in the presence of outliers. To do this, we introduced artificial outliers by multiplying some values of responses by 100 with a fixed degree of dependence (α). The Robust estimator has a better performance than the classical one, even if the MMSE of both estimators increases substantially relative to the number of outliers, but it remains very low for the Robust method, as shown in table 4.2.

4.4 Appendix

4.4.1 Proofs of Theorem 4.2.1:

For the proofs of the theorem 4.2.1, we use the fact that ρ is a strictly convex function and continuously differentiable w.r.t. the second component, then ψ is strictly monotone and continuous w.r.t. the second component. We give the proofs for the case of an increasing $\psi(Y-.)$, decreasing case being obtained by considering $-\psi(Y-.)$. Therefore, we can write, under this consideration, for all $\kappa>0$

$$\Psi\left(x,\theta_x-\kappa\right)\leq\Psi\left(x,\theta_x\right)=0\leq\Psi\left(x,\theta_x+\kappa\right)$$

and

$$\widehat{\Psi}\left(x,\widehat{\theta}_x-\kappa\right)\leq\widehat{\Psi}\left(x,\widehat{\theta}_x\right)=0\leq\widehat{\Psi}\left(x,\widehat{\theta}_x+\kappa\right).$$

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

$$\begin{aligned}\mathbb{P}\left(\left|\widehat{\theta}_x-\theta_x\right|\geq\kappa\right) &\leq \mathbb{P}\left(\left|\widehat{\Psi}\left(x,\theta_x+\kappa\right)-\Psi\left(x,\theta_x+\kappa\right)\right|\geq\Psi\left(x,\theta_x+\kappa\right)\right)\\ &\quad+\mathbb{P}\left(\left|\widehat{\Psi}\left(x,\theta_x-\kappa\right)-\Psi\left(x,\theta_x-\kappa\right)\right|\geq-\Psi\left(x,\theta_x-\kappa\right)\right).\end{aligned}$$

So, it suffices to show that

$$\widehat{\Psi}(x,t)-\Psi(x,t)\to0\quad\text{a.co. for}\quad t:=\theta_x\pm\kappa. \tag{4.18}$$

Moreover, under ((**H4**) ***(i)***), we get that

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

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

$$\exists\tau>0,\quad\sum_{n=1}^{\infty}\mathbb{P}\left(\Psi'\left(x,\xi_n\right)<\tau\right)<\infty, \tag{4.19}$$

we would have

$$\left|\widehat{\theta}_x-\theta_x\right|=O_{a.co.}\left(\sup_{t\in[\theta_x-\kappa,\theta_x+\kappa]}\left|\Psi(x,t)-\widehat{\Psi}(x,t)\right|\right).$$

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

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

To do that, we consider the following decomposition

$$\begin{aligned}\widehat{\Psi}(x,t) - \Psi(x,t) &= \frac{1}{\widehat{\Psi}_D(x)}\left\{\left(\widehat{\Psi}_N(x,t) - \mathbb{E}\left[\widehat{\Psi}_N(x,t)\right]\right)\right. \\ &\quad \left. - \left(\mathbb{E}\left[\widehat{\Psi}_D(x)\right]\Psi(x,t) - \mathbb{E}\left[\widehat{\Psi}_N(x,t)\right]\right)\right\} \\ &\quad + \frac{\Psi(x,t)}{\widehat{\Psi}_D(x)}\left(\mathbb{E}\left[\widehat{\Psi}_D(x)\right] - \widehat{\Psi}_D(x)\right). \qquad (4.20)\end{aligned}$$

Therefore, Theorem 4.2.1's result is a consequence of the following intermediate results, where their proofs are postponed to the appendix.

Lemma 4.4.1. *Under hypotheses **(H1)-(H4 iv)**, we obtain*

$$|\widehat{\Psi}_D(x) - \mathbb{E}\left[\widehat{\Psi}_D(x)\right]| = O_{a.co.}\left(\sqrt{\frac{\log n}{n\varphi_x(h)}}\right). \qquad (4.21)$$

and

$$\lim_{n\to\infty}\widehat{\Psi}_D(x) = \lim_{n\to\infty}\mathbb{E}\left[\widehat{\Psi}_D(x)\right] = p(x), \quad a.co. \qquad (4.22)$$

Corollary 4.4.1. *Under hypotheses of lemma 4.4.1, we obtain:*

$$\text{there exists } \eta > 0, \text{ such that } \sum_{n=1}^{\infty}\mathbb{P}\left(\left|\widehat{\Psi}_D(x)\right| < \eta\right) < \infty. \qquad (4.23)$$

Lemma 4.4.2. *Under hypotheses **(H3)** and **(H4)**, we have*

$$\sup_{t\in[\theta_x-\kappa,\theta_x+\kappa]}\left|\mathbb{E}\left[\widehat{\Psi}_D(x)\right]\Psi(x,t) - \mathbb{E}\left[\widehat{\Psi}_N(x,t)\right]\right| = O\left(h^b\right). \qquad (4.24)$$

Lemma 4.4.3. *Under hypotheses **(H1)-(H5)**, we have*

$$\sup_{t\in[\theta_x-\kappa,\theta_x+\kappa]}\left|\widehat{\Psi}_N(x,t) - \mathbb{E}\left[\widehat{\Psi}_N(x,t)\right]\right| = O_{a.co.}\left(\sqrt{\frac{\log n}{n\varphi_x(h)}}\right). \qquad (4.25)$$

Lemma 4.4.4. *Under the hypotheses of Theorem 4.2.1, $\widehat{\theta}_x$ exists and is unique a.co. for all sufficiently large n and there exists $\tau > 0$ such that*

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

4.4.2 Proofs of Proposition 4.2.1:

Similarly to (4.20), we have

$$\begin{aligned}\widehat{F}(y|X=x)-F(y|X=x) &= \frac{1}{\widehat{\Psi}_D(x)}\left\{\left(\widehat{R}_N(x,y)-\mathbb{E}\left[\widehat{R}_N(x,y)\right]\right)\right.\\ &\left.-\left(\mathbb{E}\left[\widehat{\Psi}_D(x)\right]F(y|X=x)-\mathbb{E}\left[\widehat{R}_N(x,y)\right]\right)\right\}\\ &+\frac{F(y|X=x)}{\widehat{\Psi}_D(x)}\left(\mathbb{E}\left[\widehat{\Psi}_D(x)\right]-\widehat{\Psi}_D(x)\right).\end{aligned}$$

Then, Proposition 4.2.1 can be deduced from the following intermediate results, together with Lemma 4.4.1 and Corollary 4.4.1.

Lemma 4.4.5. *Under hypotheses **(E3)** and **(E4)**, we have*

$$\sup_{y\in[\vartheta_x-\kappa,\vartheta_x+\kappa]}\left|\mathbb{E}\left[\widehat{\Psi}_D(x)\right]F(y|X=x)-\mathbb{E}\left[\widehat{R}_N(x,y)\right]\right|=O\left(h^{b_1}\right). \tag{4.26}$$

Lemma 4.4.6. *Under hypotheses **(E1)-(E4)**, we have*

$$\sup_{y\in[\vartheta_x-\kappa,\vartheta_x+\kappa]}\left|\widehat{R}_N(x,y)-\mathbb{E}\left[\widehat{R}_N(x,y)\right]\right|=O_{a.co.}\left(\sqrt{\frac{\log n}{n\varphi_x(h)}}\right). \tag{4.27}$$

4.4.3 Proofs of Theorem 4.2.2:

Assumption (**E5 (*i*)**) leads to

$$\widehat{\vartheta}_x-\vartheta_x=\frac{\Gamma\left(x,\widehat{\vartheta}_x,\widehat{s}(x)\right)-\widehat{\Gamma}\left(x,\widehat{\vartheta}_x,\widehat{s}(x)\right)}{\Gamma'\left(x,\xi_n,\widehat{s}(x)\right)}$$

where ξ_n is between $\widehat{\vartheta}_x$ and ϑ_x. Assumption (**E5 (*i*)**) and Lemma 4.2.1 imply that

$$\exists\tau>0,\quad\sum_{n=1}^{\infty}\mathbb{P}\left(\Gamma'\left(x,\xi_n,\widehat{s}(x)\right)<\tau\right)<\infty, \tag{4.28}$$

we would have

$$\left|\widehat{\vartheta}_x-\vartheta_x\right|=O_{a.co.}\left(\sup_{y\in[\vartheta_x-\kappa,\vartheta_x+\kappa]}\sup_{A\le\sigma\le B}\left|\Gamma(x,y,\sigma)-\widehat{\Gamma}(x,y,\sigma)\right|\right).$$

This result is based on the same kind of decomposition as (4.20). Indeed, we can write:

$$\begin{aligned}\widehat{\Gamma}(x,y,\sigma)-\Gamma(x,y,\sigma) &= \frac{1}{\widehat{\Psi}_D(x)}\Big\{\Big(\widehat{\Gamma}_N(x,y,\sigma)-\mathbb{E}\big[\widehat{\Gamma}_N(x,y,\sigma)\big]\Big)\\ &\quad -\Big(\mathbb{E}\big[\widehat{\Psi}_D(x)\big]\Gamma(x,y,\sigma)-\mathbb{E}\big[\widehat{\Gamma}_N(x,y,\sigma)\big]\Big)\Big\}\\ &\quad +\frac{\Gamma(x,y,\sigma)}{\widehat{\Psi}_D(x)}\Big(\mathbb{E}\big[\widehat{\Psi}_D(x)\big]-\widehat{\Psi}_D(x)\Big). \end{aligned} \tag{4.29}$$

Finally, the proof of Theorem 4.2.2 is achieved via the following lemmas, together with Lemma 4.4.1 and Corollary 4.4.1.

Lemma 4.4.7. *Under hypotheses **(E3)** and **(E5)**, we have*

$$\sup_{y\in[\vartheta_x-\kappa,\vartheta_x+\kappa]}\ \sup_{A\le\sigma\le B}\left|\mathbb{E}\big[\widehat{\Psi}_D(x)\big]\Gamma(x,y,\sigma)-\mathbb{E}\big[\widehat{\Gamma}_N(x,y,\sigma)\big]\right| = O\left(h^{b_1}\right). \tag{4.30}$$

Lemma 4.4.8. *Under hypotheses **(E1)-(E3)** and **(E5)-(E6)**, we have*

$$\sup_{y\in[\vartheta_x-\kappa,\vartheta_x+\kappa]}\ \sup_{A\le\sigma\le B}\left|\widehat{\Gamma}_N(x,y,\sigma)-\mathbb{E}\big[\widehat{\Gamma}_N(x,y,\sigma)\big]\right| = O_{a.co.}\left(\sqrt{\frac{\log n}{n\varphi_x(h)}}\right). \tag{4.31}$$

Lemma 4.4.9. *Under the hypotheses of Theorem 4.2.2, $\widehat{\vartheta}_x$ exists and is unique $a.co.$ for all sufficiently large n.*

4.4.4 Proofs of Lemmas:

Proof of Lemma 4.4.1:

First, we have:

$$\widehat{\Psi}_D(x)-\mathbb{E}\big[\widehat{\Psi}_D(x)\big] = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{\delta_i K_i(x)}{\mathbb{E}(K_1(x))}-\frac{\mathbb{E}(\delta_i K_i(x))}{\mathbb{E}(K_1(x))}\right) = \frac{1}{n}\sum_{i=1}^{n}\left(\tilde{\Delta}_i-\mathbb{E}(\tilde{\Delta}_i)\right),$$

where $\tilde{\Delta}_i = \dfrac{\delta_i K_i(x)}{\mathbb{E}(K_1(x))}$. Because of **(H1)** and **(H3)**, we can write

$$C\varphi_x(h) < \mathbb{E}(K_1(x)) < C'\varphi_x(h)$$

So, we can get directly that

$$\left|\tilde{\Delta}_i\right| < C/\varphi_x(h) \quad \text{and} \quad \mathbb{E}\left|\tilde{\Delta}_i\right|^2 < C'/\varphi_x(h).$$

Thus, the use of the classical Bernstein's inequality allows us to write for all $\eta > 0$:

$$\mathbb{P}\left(\left|\widehat{\Psi}_D(x) - \mathbb{E}\left[\widehat{\Psi}_D(x)\right]\right| > \eta\sqrt{\frac{\log n}{n\varphi_x(h)}}\right) \leq C' n^{-C\eta^2}.$$

For the proof of (4.22), we only need to establish

$$\mathbb{E}\left[\widehat{\Psi}_D(x)\right] \to p(x), \quad a.co. \text{ as } n \to \infty. \tag{4.32}$$

By the properties of conditional expectation and the mechanism of MAR and **(H4 *iv*)**, it follows that

$$\begin{aligned}
\mathbb{E}\left[\widehat{\Psi}_D(x)\right] &= \frac{1}{nE(K_1(x))}\sum_{i=1}^{n}\mathbb{E}(\delta_i K_i(x)) = \frac{1}{nE(K_1(x))}\sum_{i=1}^{n}\mathbb{E}(\mathbb{E}[\delta_i|X_i]K_i(x)) \\
&= \frac{1}{nE(K_1(x))}[p(x)+o(1)]\sum_{i=1}^{n}\mathbb{E}(K_i(x)) \to p(x), \quad a.co. \text{ as } n\to\infty.
\end{aligned}$$

Therefore, (4.22) follows from (4.21) and (4.32). □

Proof of Corollary 4.4.1:

It is easy to remark that:
$\left|\widehat{\Psi}_D(x)\right| \leq \frac{p(x)}{2}$ implies that $p(x) - \widehat{\Psi}_D(x) \geq \frac{p(x)}{2}$ which implies that $\left|\widehat{\Psi}_D(x) - p(x)\right| \geq \frac{p(x)}{2}$.

We deduce, from Lemma 4.4.1, that

$$\mathbb{P}\left(\left|\widehat{\Psi}_D(x)\right| \leq \frac{p(x)}{2}\right) \leq \mathbb{P}\left(\left|\widehat{\Psi}_D(x) - p(x)\right| > \frac{p(x)}{2}\right).$$

Consequently

$$\sum_{n=1}^{\infty}\mathbb{P}\left(\left|\widehat{\Psi}_D(x)\right| < \frac{p(x)}{2}\right) < \infty.$$

□

Proof of Lemma 4.4.2:

Since $(X_1, Y_1, \delta_1), ..., (X_n, Y_n, \delta_n)$ are independents identically distributed, we have:

$$\begin{aligned}
\forall t \in [\theta_x - \kappa, \theta_x + \kappa] \quad & \left|\mathbb{E}\left[\widehat{\Psi}_D(x)\right]\Psi(x,t) - \mathbb{E}\left[\widehat{\Psi}_N(x,t)\right]\right| \\
= \quad & \frac{\left|\mathbb{E}\left[\delta_1 K_1(x)\psi(Y_1 - t) - \delta_1 K_1(x)\Psi(x,t)\right]\right|}{\mathbb{E}(K_1(x))} \\
= \quad & \frac{\left|\mathbb{E}\left(\mathbb{1}_{B(x,h)}(X_1)K_1(x)p(X_1)\left[\mathbb{E}(\psi(Y_1 - t)|X = X_1) - \Psi(x,t)\right]\right)\right|}{\mathbb{E}(K_1(x))} \\
= \quad & \frac{\left|\mathbb{E}\left(\mathbb{1}_{B(x,h)}(X_1)K_1(x)p(X_1)\left[\Psi(X_1,t) - \Psi(x,t)\right]\right)\right|}{\mathbb{E}(K_1(x))}.
\end{aligned}$$

Then, by the Hölder hypothesis (**H4 (*ii*)**), (**H3**) and the continuity of $p(x)$, we get that:

$$\begin{aligned}
& \forall t \in [\theta_x - \kappa, \theta_x + \kappa] \\
\left|\mathbb{E}\left[\widehat{\Psi}_D(x)\right]\Psi(x,t) - \mathbb{E}\left[\widehat{\Psi}_N(x,t)\right]\right| \quad & \leq \quad Ch^b[p(x) + o(1)]\frac{\left|\mathbb{E}\left(K_1(x)\right)\right|}{\mathbb{E}(K_1(x))} \\
& = \quad O(h^b).
\end{aligned}$$

□

Proof of Lemma 4.4.3:

Using the compactness of $[\theta_x - \kappa, \theta_x + \kappa]$, we can write that $[\theta_x - \kappa, \theta_x + \kappa] \subset \bigcup_{k=1}^{s_n} S_k$ where $S_k = (y_k - l_n, y_k + l_n)$.

We consider the intervals extremities gride

$$\mathcal{H}_n = \left\{y_j - l_n, y_j + l_n, 1 \leq j \leq s_n\right\}.$$

Then the monotony of $\mathbb{E}\left[\widehat{\Psi}_N(x,t)\right]$ and $\widehat{\Psi}_N(x,t)$ gives, for $1 \leq j \leq s_n$

$$\begin{aligned}
\mathbb{E}\left[\widehat{\Psi}_N\left(x, y_j - l_n\right)\right] \quad & \leq \quad \sup_{t \in (y_j - l_n, y_j + l_n)} \mathbb{E}\left[\widehat{\Psi}_N(x,t)\right] \leq \mathbb{E}\left[\widehat{\Psi}_N\left(x, y_j + l_n\right)\right] \\
\widehat{\Psi}_N\left(x, y_j - l_n\right) \quad & \leq \quad \sup_{t \in (y_j - l_n, y_j + l_n)} \widehat{\Psi}_N(x,t) \leq \widehat{\Psi}_N\left(x, y_j + l_n\right).
\end{aligned} \tag{4.33}$$

Now, from (**H4 (*ii*)**) we have, for any $t_1, t_2 \in [\theta_x - \kappa, \theta_x + \kappa]$

$$\left|\mathbb{E}\left[\widehat{\Psi}_N\left(x, t_1\right)\right] - \mathbb{E}\left[\widehat{\Psi}_N\left(x, t_2\right)\right]\right| \leq C\left|t_1 - t_2\right|^{b'}. \tag{4.34}$$

So, we deduce from (4.33) and (4.34) that

$$\begin{aligned}
\sup_{t \in [\theta_x - \kappa, \theta_x + \kappa]} \quad & \left|\mathbb{E}\left[\widehat{\Psi}_N\left(x,t\right)\right] - \widehat{\Psi}_N\left(x,t\right)\right| \\
\leq \quad & \max_{1 \leq j \leq s_n} \max_{z \in \{y_j - l_n, y_j + l_n\}} \left|\widehat{\Psi}_N(x,z) - \mathbb{E}\left[\widehat{\Psi}_N(x,z)\right]\right| + 2Cl_n^{b'}.
\end{aligned} \tag{4.35}$$

Take now $l_n = n^{-\alpha/b'}$ for some $\alpha > b'/2$, and note that because of $\lim_{n\to\infty} l_n = 0$ and (**H2**), we have

$$\sup_{t\in[\theta_x-\kappa,\theta_x+\kappa]} \left|\mathbb{E}\left[\widehat{\Psi}_N(x,t)\right] - \widehat{\Psi}_N(x,t)\right| \le \max_{z\in\mathcal{H}_n}\left|\widehat{\Psi}_N(x,z) - \mathbb{E}\left[\widehat{\Psi}_N(x,z)\right]\right| + O\left(\sqrt{\frac{\log n}{n\varphi_x(h)}}\right). \tag{4.36}$$

The proof of this part is based on the exponential inequality given in Corollary A.8.ii in Ferraty and Vieu (2006) with $Z_i = \frac{1}{\mathbb{E}(K_1(x))}\left[\delta_i K_i(x)\psi(Y_i - z) - \mathbb{E}\left(\delta_i K_i(x)\psi(Y_i - z)\right)\right]$.

To do that, we have to show that:

$$\exists C > 0, \forall m \ge 2, \quad \mathbb{E}(|Z_1^m|) = C\varphi_x^{-m+1}(h). \tag{4.37}$$

First, we prove for $m \ge 2$ that:

$$\frac{1}{\mathbb{E}^m(K_1(x))}\mathbb{E}\left[|\delta_1 K_1(x)\psi(Y_1 - z)|^m\right] = O(\varphi_x^{-m+1}(h)). \tag{4.38}$$

Then, using (**H3**) and (**H5**) we write:

$$\begin{aligned} \mathbb{E}\left[|\delta_1 K_1^m(x)\psi(Y_1 - z)^m|\right] &\le \mathbb{E}\left[\mathbb{E}\left(|\psi(Y_1 - z)|^m |X_1\right) K_1^m(x)\right] \\ &\le C\mathbb{E}(K_1^m(x)) \\ &\le C\varphi_x(h). \end{aligned}$$

Which implies that

$$\frac{1}{\mathbb{E}^m(K_1(x))}\mathbb{E}\left[|\delta_1\psi(Y_1 - z)^m K_1^m(x)|\right] = O(\varphi_x^{-m+1}(h))$$

and

$$\frac{1}{\mathbb{E}(K_1(x))}\mathbb{E}\left[|\delta_1\psi(Y_1 - z) K_1(x)|\right] \le C.$$

Next, by the Newton's binomial expansion we obtain:

$$\begin{aligned} \mathbb{E}(|Z_1^m|) &\le C\sum_{k=0}^{m} \frac{\mathbb{E}\left[\left|\delta_1\psi(Y_1 - z)^k K_1^k(x)\right|\right]}{\mathbb{E}^k(K_1(x))}\left[\frac{\mathbb{E}\left[|\delta_1\psi(Y_1 - z) K_1(x)|\right]}{\mathbb{E}(K_1(x))}\right]^{m-k} \\ &\le C\max_{k=0,\ldots,m}\varphi_x^{-k+1}(h) \\ &\le C\varphi_x^{-m+1}(h). \end{aligned}$$

It follows that:

$$\mathbb{E}(|Z_1^m|) = O(\varphi_x^{-m+1}(h)). \tag{4.39}$$

Now apply the exponential inequality given by Corollary **A.8.***ii* in Ferraty and Vieu (2006) for Z_i. Since $\mathbb{E}[|Z_i|^m] = O(\varphi_x(h)^{-m+1})$, then, we can take $a^2 = \frac{1}{\varphi_x(h)}$. Hence, for all $\eta_0 > 0$:

$$\begin{aligned}\mathbb{P}\left(\left|\widehat{\Psi}_N(x,z) - \mathbb{E}\left(\widehat{\Psi}_N(x,z)\right)\right| > \eta\sqrt{\frac{\log n}{n\varphi_x(h)}}\right) &= \mathbb{P}\left(\frac{1}{n}\left|\sum_{i=1}^n Z_i\right| > \eta\sqrt{\frac{\log n}{n\varphi_x(h)}}\right)\\ &\leq 2\exp\left(-C\eta^2\log n\right).\end{aligned}$$

Then, we have for any $\eta > 0$

$$\begin{aligned}&\mathbb{P}\left(\max_{z\in\mathcal{H}_n}\left|\widehat{\Psi}_N(x,z) - \mathbb{E}\left(\widehat{\Psi}_N(x,z)\right)\right| > \eta\sqrt{\frac{\log n}{n\varphi_x(h)}}\right)\\ &\qquad\leq s_n\max_{z\in\mathcal{H}_n}\mathbb{P}\left(\left|\widehat{\Psi}_N(x,z) - \mathbb{E}\left(\widehat{\Psi}_N(x,z)\right)\right| > \eta\sqrt{\frac{\log n}{n\varphi_x(h)}}\right)\\ &\qquad\leq C s_n n^{-C\eta^2}\\ &\qquad\leq \frac{C}{l_n} n^{-C\eta^2}.\end{aligned}$$

Thus, by choosing η such that $C\eta^2 = 1/2 + 2\alpha/b'$, we obtain

$$s_n\max_{z\in\mathcal{H}_n}\mathbb{P}\left(\left|\widehat{\Psi}_N(x,z) - \mathbb{E}\left(\widehat{\Psi}_N(x,z)\right)\right| > \eta\sqrt{\frac{\log n}{n\varphi_x(h)}}\right) \leq n^{-1/2-2\alpha/b'}.$$

Now, we can conclude

$$\sup_{t\in[\theta_x-\kappa,\theta_x+\kappa]}\left|\widehat{\Psi}_N(x,t) - \mathbb{E}\left(\widehat{\Psi}_N(x,t)\right)\right| = O_{a.co.}\left(\sqrt{\frac{\log n}{n\varphi_x(h)}}\right). \tag{4.40}$$

□

Proof of Lemma 4.4.4:

We give the proof for the case of an increasing $\psi(Y - .)$, decreasing case being obtained by considering $-\psi(Y - .)$. Therefore we can Therefore write, under this consideration, for all $\kappa > 0$

$$\Psi(x,\theta_x-\kappa) \leq \Psi(x,\theta_x) = 0 \leq \Psi(x,\theta_x+\kappa)$$

The results of lemmas 4.4.1,4.4.2,4.4.3 and 4.4.1 show that

$$\widehat{\Psi}(x,t) \to \Psi(x,t), \qquad \text{in a.co. as } n\to\infty.$$

for all real fixed $t \in [\theta_x - \kappa, \theta_x + \kappa]$. So, for sufficiently large n.

$$\widehat{\Psi}\left(x, \widehat{\theta}_x - \kappa\right) \leq 0 \leq \widehat{\Psi}\left(x, \widehat{\theta}_x + \kappa\right), \qquad \text{in a.co. .}$$

Since ψ is a continuous function, then as $\widehat{\Psi}(x,t)$ is a continuous function of t, there exists a $\widehat{\theta}_x \in [\theta_x - \kappa, \theta_x + \kappa]$ such that $\widehat{\Psi}(x, \widehat{\theta}_x) = 0$. Finally, the uniqueness (in $a.co.$) of $\widehat{\theta}_x$ is a direct consequence of the strict monotonicity of ψ, while the second part is a direct consequence of the regularity assumption (**H4 (*i*)**). □

Proof of Lemma 4.4.5:

The same idea in the proof of lemma 4.4.2 □

Proof of Lemma 4.4.6:

As in *Lemma 3* of Attouch et al. (2013) using the compactness of $[\vartheta_x - \kappa, \vartheta_x + \kappa]$, we can write

$$[\vartheta_x - \kappa, \vartheta_x + \kappa] \subset \bigcup_{j=1}^{d_n} (y_j - l_n, y_j + l_n).$$

We consider the intervals extremities gride

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

Then the monotony of $\mathbb{E}\left[\widehat{R}_N(x,y)\right]$ and $\widehat{R}_N(x,y)$ gives, for $1 \leq j \leq d_n$

$$\begin{aligned} \mathbb{E}\left[\widehat{R}_N(x, y_j - l_n)\right] &\leq \sup_{y \in (y_j - l_n, y_j + l_n)} \mathbb{E}\left[\widehat{R}_N(x,y)\right] \leq \mathbb{E}\left[\widehat{R}_N(x, y_j + l_n)\right] \\ \widehat{R}_N(x, y_j - l_n) &\leq \sup_{y \in (y_j - l_n, y_j + l_n)} \widehat{R}_N(x,y) \leq \widehat{R}_N(x, y_j + l_n). \end{aligned} \tag{4.41}$$

Now, from (**E4 (*iv*)**) we have, for any $y_1, y_2 \in [\vartheta_x - \kappa, \vartheta_x + \kappa]$

$$\left|\mathbb{E}\left[\widehat{R}_N(x, y_1)\right] - \mathbb{E}\left[\widehat{R}_N(x, y_2)\right]\right| \leq C\left|y_1 - y_2\right|^{b_2}. \tag{4.42}$$

So, we deduce from (4.41) and (4.42) that

$$\begin{aligned} \sup_{y \in [\vartheta_x - \kappa, \vartheta_x + \kappa]} & \left|\mathbb{E}\left[\widehat{R}_N(x,y)\right] - \widehat{R}_N(x,y)\right| \\ &\leq \max_{1 \leq j \leq d_n} \max_{z \in \{y_j - l_n, y_j + l_n\}} \left|\widehat{R}_N(x,z) - \mathbb{E}\left[\widehat{R}_N(x,z)\right]\right| + 2C l_n^{b_2} \qquad (4.43) \\ &\leq \max_{z \in \mathcal{G}_n} \left|\widehat{R}_N(x,z) - \mathbb{E}\left[\widehat{R}_N(x,z)\right]\right| + O\left(\sqrt{\frac{\log n}{n\varphi_x(h)}}\right). \qquad (4.44) \end{aligned}$$

Let $\Lambda_i = (\delta_i K_i(x) \mathbb{1}_{(-\infty;z]}(Y_i) - \mathbb{E}[\delta_i K_i(x) \mathbb{1}_{(-\infty;z]}(Y_i)]) / \mathbb{E}(K_1(x))$. By using similar arguments for the proof of Lemma 4.4.1, we deduce that $\mathbb{E}|\Lambda_i| \leq C/\varphi_x(h)$ and $\mathbb{E}(\Lambda_i^2) \leq$

$C'/\varphi_x(h)$. By applying again the Bernstein's exponential inequality, we get

$$\max_{z\in\mathcal{G}_n}\left|\mathbb{E}\left[\widehat{R}_N(x,z)\right]-\widehat{R}_N(x,z)\right| = O_{a.co.}\left(\sqrt{\frac{\log n}{n\varphi_x(h)}}\right). \tag{4.45}$$

which concluded the proof. □

Proof of Lemma 4.2.1:

The proof of this lemma is analogous to *Lemma A.4* of Boente and Vahnovan (2015). By Proposition 4.2.1 and Assumption (**E2**) we can check that:

$$\sup_{y\in[\vartheta_x-\kappa,\vartheta_x+\kappa]}\left|\widehat{F}(y|X=x)-F(y|X=x)\right| \to 0. \tag{4.46}$$

Otherwise by (**E4**) for a fixed $x \in \mathcal{F}$ there exist a, b such that $F(b|X = x) > \frac{7}{8}$ and $F(a|X = x) < \frac{1}{8}$. Let $m_n(x)$ be the median of $\widehat{F}(y|X = x)$. Then, (4.46) implies that there exists $n_0 \in \mathbb{N}$ such that $\widehat{F}(a|X = x) < \frac{1}{4}$ and $\widehat{F}(b|X = x) > \frac{3}{4}$ for all $n \geq n_0$. Hence, we have that $a < m_n(x) < b$. It is easy to see that for a good choice of a, b and n also, implies that $\widehat{s}(x) < b - a$ for all $n \geq n_0$.

For the lower bound, using that $F(y|X = x)$ is a continuous distribution function in $x \in \mathcal{F}$, there exist $a(x)$ and $b(x)$ such that

$$F(a(x)|X=x) = \frac{1}{3}, \quad F(b(x)|X=x) = \frac{7}{10}.$$

Let $a(x) = a$ and $b(x) = b$. Assumption (**E4 (*i*)**) entails that for any $\varepsilon < \frac{1}{30}$ there exists $\eta > 0$, such that

$$\frac{1}{3}-\varepsilon < F(a-\eta|X=x) < \frac{1}{3}+\varepsilon, \quad \frac{7}{10}-\varepsilon < F(b+\eta|X=x) < \frac{7}{10}+\varepsilon.$$

Finally, (4.46) implies that for all $n \geq n_0$ we have that $\widehat{F}(a|X = x) < \frac{1}{2}$, $\widehat{F}(b|X = x) > \frac{1}{2}$, $\widehat{F}(a-\eta|X = x) > \frac{1}{4}$ and $\widehat{F}(b+\eta|X = x) < \frac{3}{4}$. Hence, $a < m_n(x) < b$, and $\widehat{s}(x) > \eta$ for all $n \geq n_0$. □

Proof of Lemma 4.4.7:

The proof is analogous to the proof of lemma 4.4.2. □

Proof of Lemma 4.4.8:

$\forall y \in [\vartheta_x-\kappa,\vartheta_x+\kappa], \forall A \leq \sigma \leq B$ consider a finite covering of $[\vartheta_x-\kappa,\vartheta_x+\kappa]$, we have $[\vartheta_x-\kappa,\vartheta_x+\kappa] \subset \bigcup_{k=1}^{d_n}(y_k-l_n, y_k+l_n)$. Taking $k_y = \arg\min_{t\in\{y_1,\ldots,y_{d_n}\}}|y-t|$. On the other hand, by the same way covering $[A,B] \subset \bigcup_{k=1}^{s_n}(\sigma_k-\nu_n,\sigma_k+\nu_n)$ and taking $k_\sigma = \arg\min_{s\in\{\sigma_1,\ldots,\sigma_{s_n}\}}|\sigma-s|$.

With same manner in proof of lemma 4.4.3, by taking $l_n = \nu_n = n^{-\alpha}$ for some $\alpha > 1/2$, we have

$$l_n = \nu_n = o\left(\sqrt{\frac{\log n}{n\varphi_x(h)}}\right). \tag{4.47}$$

Note that,

$$\begin{aligned}
\sup_{y\in[\vartheta_x-\kappa,\vartheta_x+\kappa]}\sup_{\sigma\in[A,B]} & \left|\widehat{\Gamma}_N(x,y,\sigma)-\mathbb{E}\left(\widehat{\Gamma}_N(x,y,\sigma)\right)\right| \\
\leq & \underbrace{\sup_{y\in[\vartheta_x-\kappa,\vartheta_x+\kappa]}\sup_{\sigma\in[A,B]}\left|\widehat{\Gamma}_N(x,y,\sigma)-\widehat{\Gamma}_N(x,k_y,\sigma)\right|}_{F_1} \\
& + \underbrace{\sup_{y\in[\vartheta_x-\kappa,\vartheta_x+\kappa]}\sup_{\sigma\in[A,B]}\left|\widehat{\Gamma}_N(x,k_y,\sigma)-\widehat{\Gamma}_N(x,k_y,k_\sigma)\right|}_{F_2} \\
& + \underbrace{\sup_{y\in[\vartheta_x-\kappa,\vartheta_x+\kappa]}\sup_{\sigma\in[A,B]}\left|\widehat{\Gamma}_N(x,k_y,k_\sigma)-\mathbb{E}\left(\widehat{\Gamma}_N(x,k_y,k_\sigma)\right)\right|}_{F_3} \\
& + \underbrace{\sup_{y\in[\vartheta_x-\kappa,\vartheta_x+\kappa]}\sup_{\sigma\in[A,B]}\left|\mathbb{E}\left(\widehat{\Gamma}_N(x,k_y,k_\sigma)\right)-\mathbb{E}\left(\widehat{\Gamma}_N(x,k_y,\sigma)\right)\right|}_{F_4} \\
& + \underbrace{\sup_{y\in[\vartheta_x-\kappa,\vartheta_x+\kappa]}\sup_{\sigma\in[A,B]}\left|\mathbb{E}\left(\widehat{\Gamma}_N(x,k_y,\sigma)\right)-\mathbb{E}\left(\widehat{\Gamma}_N(x,y,\sigma)\right)\right|.}_{F_5}
\end{aligned}$$

• Concerning F_1 and F_5, follow the same steps of T_1 and T_3 in Lemma 4.4.3 by conditions (**E3**) and (**E6**), we obtain

$$\begin{aligned}
\sup_{y\in[\vartheta_x-\kappa,\vartheta_x+\kappa]}\sup_{\sigma\in[A,B]} & \left|\widehat{\Gamma}_N(x,y,\sigma)-\widehat{\Gamma}_N(x,k_y,\sigma)\right| \\
\leq & \sup_{y\in[\vartheta_x-\kappa,\vartheta_x+\kappa]}\sup_{\sigma\in[A,B]}\frac{1}{n\mathbb{E}(K_1(x))}\sum_{i=1}^n\left|\psi\left(\frac{Y_i-y}{\sigma}\right)-\psi\left(\frac{Y_i-k_y}{\sigma}\right)\right|\delta_iK_i(x) \\
\leq & \sup_{y\in[\vartheta_x-\kappa,\vartheta_x+\kappa]}\frac{\|\psi'\|_\infty}{A}|y-k_y|\left(\frac{1}{n\mathbb{E}(K_1(x))}\sum_{i=1}^n\delta_iK_i(x)\right) \\
\leq & Cl_n.
\end{aligned}$$

Now, for n large enough, we can write

$$\mathbb{P}\left(\sup_{y\in[\vartheta_x-\kappa,\vartheta_x+\kappa]}\sup_{\sigma\in[A,B]}\left|\widehat{\Gamma}_N(x,y,\sigma)-\widehat{\Gamma}_N(x,k_y,\sigma)\right| > \frac{\eta}{5}\sqrt{\frac{\log n}{n\varphi_x(h)}}\right) = 0 \tag{4.48}$$

and

$$\mathbb{P}\left(\sup_{y \in [\vartheta_x - \kappa, \vartheta_x + \kappa]} \sup_{\sigma \in [A,B]} \left| \mathbb{E}\left(\widehat{\Gamma}_N(x, k_y, \sigma)\right) - \mathbb{E}\left(\widehat{\Gamma}_N(x, y, \sigma)\right) \right| > \frac{\eta}{5} \sqrt{\frac{\log n}{n\varphi_x(h)}} \right) = 0. \quad (4.49)$$

• Concerning F_2 and F_4, by assumptions (**E3**) and (**E6**)(using that $\zeta(u) = u\psi'(u)$ is bounded), it follows

$$\begin{aligned}
\sup_{y \in [\vartheta_x - \kappa, \vartheta_x + \kappa]} \sup_{\sigma \in [A,B]} & \left| \widehat{\Gamma}_N(x, k_y, \sigma) - \widehat{\Gamma}_N(x, k_y, k_\sigma) \right| \\
&\le \sup_{\sigma \in [A,B]} \frac{1}{n\mathbb{E}(K_1(x))} \sum_{i=1}^n \left| \psi\left(\frac{Y_i - k_y}{\sigma}\right) - \psi\left(\frac{Y_i - k_y}{k_\sigma}\right) \right| \delta_i K_i(x) \\
&\le \frac{\|\zeta\|_\infty}{A} \nu_n \left(\frac{1}{n\mathbb{E}(K_1(x))} \sum_{i=1}^n \delta_i K_i(x) \right) \\
&\le C\nu_n.
\end{aligned}$$

Thus, for n large enough, we can write

$$\mathbb{P}\left(\sup_{y \in [\vartheta_x - \kappa, \vartheta_x + \kappa]} \sup_{\sigma \in [A,B]} \left| \widehat{\Gamma}_N(x, k_y, \sigma) - \widehat{\Gamma}_N(x, k_y, k_\sigma) \right| > \frac{\eta}{5} \sqrt{\frac{\log n}{n\varphi_x(h)}} \right) = 0 \quad (4.50)$$

and

$$\mathbb{P}\left(\sup_{y \in [\vartheta_x - \kappa, \vartheta_x + \kappa]} \sup_{\sigma \in [A,B]} \left| \mathbb{E}\left(\widehat{\Gamma}_N(x, k_y, \sigma)\right) - \mathbb{E}\left(\widehat{\Gamma}_N(x, k_y, k_\sigma)\right) \right| > \frac{\eta}{5} \sqrt{\frac{\log n}{n\varphi_x(h)}} \right) = 0. \quad (4.51)$$

• Concerning F_3: Let $\Omega_i = \frac{1}{\mathbb{E}(K_1(x))} \left[\delta_i K_i(x) \psi\left(\frac{Y_i - k_y}{k_\sigma}\right) - \mathbb{E}\left(\delta_i K_i(x) \psi\left(\frac{Y_i - k_y}{k_\sigma}\right)\right)\right]$.
Note that K and ψ are bounded and we have $\widehat{\Gamma}_N(x, k_y, k_\sigma) - \mathbb{E}\left(\widehat{\Gamma}_N(x, k_y, k_\sigma)\right) = \frac{1}{n}\sum_{i=1}^n \Omega_i$, we deduce that $\mathbb{E}|\Omega_i| \le C/\varphi_x(h)$ and $\mathbb{E}(\Omega_i^2) \le C'/\varphi_x(h)$.

We apply now again the Bernstein's exponential inequality to get

$$\begin{aligned}
&\mathbb{P}\left(F_3 > \frac{\eta}{5} \sqrt{\frac{\log n}{n\varphi_x(h)}} \right) \\
&\quad = \mathbb{P}\left(\max_{k_y \in \{y_1, \ldots, y_{d_n}\}} \max_{k_\sigma \in \{\sigma_1, \ldots, \sigma_{s_n}\}} \left| \widehat{\Gamma}_N(x, k_y, k_\sigma) - \mathbb{E}\left(\widehat{\Gamma}_N(x, k_y, k_\sigma)\right) \right| > \frac{\eta}{5} \sqrt{\frac{\log n}{n\varphi_x(h)}} \right) \\
&\quad \le d_n s_n \max_{k_y \in \{y_1, \ldots, y_{d_n}\}} \max_{k_\sigma \in \{\sigma_1, \ldots, \sigma_{s_n}\}} \mathbb{P}\left(\frac{1}{n} \left| \sum_{i=1}^n \Omega_i \right| > \frac{\eta}{5} \sqrt{\frac{\log n}{n\varphi_x(h)}} \right) \\
&\quad \le d_n s_n 2 \exp\left(-C\eta^2 \log n\right).
\end{aligned}$$

Because $l_n = \nu_n = n^{-\alpha}$ for $\alpha > 1/2$, and by choosing η such that $C\eta^2 = 1/2 + 3\alpha$, we have

$$\mathbb{P}\left(\max_{k_y \in \{y_1,\ldots,y_{d_n}\}} \max_{k_\sigma \in \{\sigma_1,\ldots,\sigma_{s_n}\}} \left|\widehat{\Gamma}_N\left(x, k_y, k_\sigma\right) - \mathbb{E}\left(\widehat{\Gamma}_N\left(x, k_y, k_\sigma\right)\right)\right| > \frac{\eta}{5}\sqrt{\frac{\log n}{n\varphi_x(h)}}\right) \leq Cn^{-\alpha-1/2}. \tag{4.52}$$

Now, from (4.48), (4.49), (4.50),(4.51) and (4.52), we conclude

$$\sup_{y \in [\vartheta_x - \kappa, \vartheta_x + \kappa]} \sup_{A \leq \sigma \leq B} \left|\widehat{\Gamma}_N(x, y, \sigma) - \mathbb{E}\left[\widehat{\Gamma}_N(x, y, \sigma)\right]\right| = O_{a.co.}\left(\sqrt{\frac{\log n}{n\varphi_x(h)}}\right). \tag{4.53}$$

□

BIBLIOGRAPHY

M. Attouch, A. Laksaci, and E. Ould-Saïd. Asymptotic Distribution of Robust Estimator for Functional Nonparametric Models. *Communications in Statistics Theory and Methods*, 38:1317–1335, 2009.

M. Attouch, A. Laksaci, and E. Ould-Saïd. Strong uniform convergence rate of robust estimator of the regression function for functional and dependent processes. *Journal of the Japan Statistical Society*, 42:125–143, 2013.

G. Boente and R. Fraiman. Robust nonparametric regression estimation. *Journal of Multivariate Analysis*, 29:180–198, 1989.

G. Boente and D. Rodriguez. Robust estimators of high order derivatives of regression functions. *Statistics & probability letters*, 76:1335–1344, 2006.

G. Boente and A. Vahnovan. Strong convergence of robust equivariant nonparametric functional regression estimators. *Statistics and Probability Letters*, 11:1–11, 2015.

D. Bosq. *Linear processes in function spaces: theory and applications*, volume 149. Springer Science & Business Media, 2012.

T. Cai, P. Hall, et al. Prediction in functional linear regression. *The Annals of Statistics*, 34:2159–2179, 2006.

M. Chaouch and S. Khardani. Randomly censored quantile regression estimation using functional stationary ergodic data. *Journal of Nonparametric Statistics*, 27:65–87, 2015.

G. Collomb and W. Härdle. Strong uniform convergence rates in robust nonparametric time series analysis and prediction: Kernel regression estimation from dependent observations. *Stochastic processes and their applications*, 23:77–89, 1986.

F. Ferraty and P. Vieu. *Nonparametric Functional Data Analysis: Theory and Practice*. Springer Series in Statistics, New York, springer edition, 2006.

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

A. Gheriballah, A. Laksaci, and S. Sekkal. Nonparametric m-regression for functional ergodic data. *Statistics & Probability Letters*, 83:902–908, 2013.

W. Härdle. Robust regression function estimation. *Journal of Multivariate Analysis*, 14: 169–180, 1984.

W. Härdle and T. Gasser. Robust nonparametric function fitting. *Journal of the Royal Statistical Society: Series B (Methodological)*, 46:42–51, 1984.

W. Härdle, A. B. Tsybakov, et al. Robust nonparametric regression with simultaneous scale curve estimation. *The annals of statistics*, 16:120–135, 1988.

P. J. Huber. A robust version of the probability ratio test. *The Annals of Mathematical Statistics*, 36:1753–1758, 1965.

P. J. Huber. Robust estimation of a location parameter. In *Breakthroughs in statistics*, pages 492–518. Springer, 1992.

N. Laïb and E. Ould-Saïd. A robust nonparametric estimation of the autoregression function under an ergodic hypothesis. *Canadian Journal of Statistics*, 28:817–828, 2000.

R. A. Maronna and D. Martin. Yohai. robust statistics. *Wiley Series in Probability and Statistics. John Wiley and Sons*, 2:3, 2006.

J. O. Ramsay and B. W. Silverman. Principal components analysis for functional data. In *Functional Data Analysis*, pages 147–172. Springer, 2005.

A. B. Tsybakov. Robust estimates of a function. *Problemy Peredachi Informatsii*, 18: 39–52, 1982.

F. Yao, H. G. Müller, and J. L. Wang. Functional data analysis for sparse longitudinal data. *Journal of the American Statistical Association*, 100:577–590, 2005.

CHAPTER 5

A REAL DATA APPLICATION

The objective of this chapter is to apply the theoretical results obtained in the previous chapters to real data. More precisely, we study two datasets. First, our contribution in section 5.1 leads to the prediction of the peak electricity demand given its daily temperature curve by using the nonparametric relative error regression under random censorship studied in the chapter 3. The second application discussed in section 5.2 has an object of determine some diesel fuels parameters by analysing its spectral data using the robust nonparametric equivariant regression under MAR (see chapter 4). A comparative study of these models is provided in order to emphasize their possible advantages.

5.1 Peak electricity demand (censored case)

5.1.1 Materials and methods

First, we have acquired a large dataset, consisting of number of 8784 records, containing the hourly energy consumption for the year 2016 (measured in MWh), retrieved from the smart metering device of a commercial center type of consumer (a large hypermarket). We have also acquired a dataset containing the historical hourly meteorological data regarding the temperature (measured in Celsius degrees), recorded by the meteorological sensors of a specialized institute for the year 2016, consisting in a number of 8784 records (see Pîrjan et al. (2017) and Mebsout et al. (2020) for more description on this data set).

Now, we are interested in the estimation of interval prediction of peak consumption of energy. For a fixed day i let us denote by $(E_i(t_j))_{j=1,\ldots,24}$ the hourly measurements of some consumption of energy. The peak demand observed for the day i is defined as

$$P_i = \max_{j=1,\ldots,24} E_i(t_j).$$

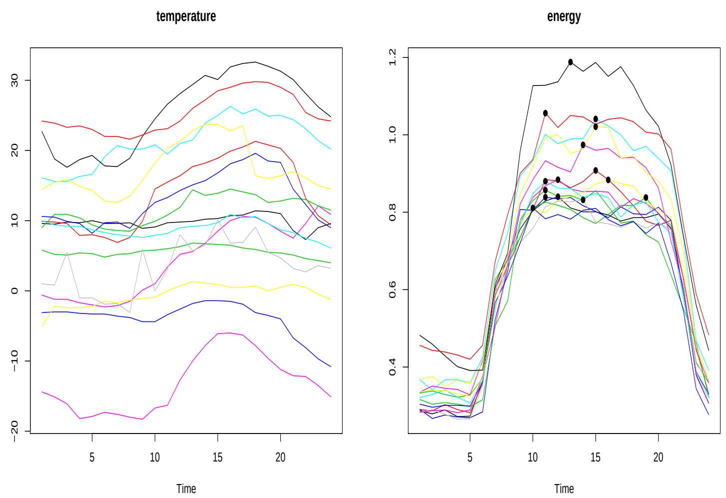

Figure 5.1: Sample of 15 daily temperature curves and the associated energy consumption curves.

It is well known that peak demand is very correlated with temperature measurements. Figure 5.1 provides a sample of 15 curves of hourly temperature measures and the associated electricity consumption curves. The observed peak, for each day. We split our sample of 366 days into a learning sample containing the first 300 days and a testing sample with the last 66 days. From the learning sample, we selected 30% of days within which we generated the censorship randomly. Figure 5.2 provides a sample of four censored daily load curves. For those days, we observe the electricity consumption until a certain time $t_c \in [1, 24]$ which corresponds to the time of censorship which is plotted in a dashed line in Figure 5.2. For a censored day, we define the censored random variable

$$C_i = \max_{j=1,\ldots,t_c} E_i\left(t_j\right).$$

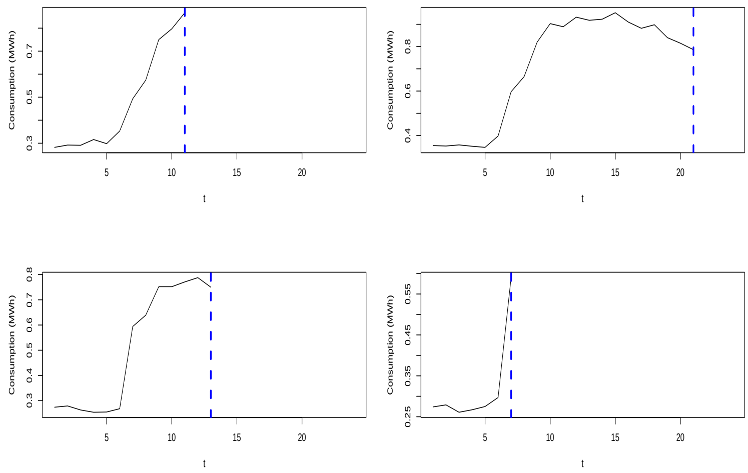

Figure 5.2: Sample of four censored daily load curves, the dashed line corresponds to the time of censorship t_c.

Therefore, our sample is formed as follows $(X_i, Y_i, \delta_i)_{i=1,\dots,300}$, where $\delta_i = 1$ if $Y_i = P_i$ and $\delta_i = 0$ if $Y_i = C_i$. In order to introduce the outliers in this sample we multiplies by 10 the response variable of a number of observations.

The selection of the bandwidth parameter is an important and basic problem in all kernel smoothing techniques. Another important point for ensuring a good behavior of the method, is to use a semi-metric that is well adapted to the kind of data we have to deal with. Ours is based on the m eigenfunctions of the empirical covariance operator associated to the m greatest eigenvalues (see Ferraty and Vieu (2006), Chapter 13). The estimators are obtained by choosing the optimal bandwidths by L^1 cross-validation method and the kernel K is the quadratic function defined by :

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

The error used is the mean of squared error (MSE) expressed by

$$MSE_{CKE} = \frac{1}{66}\sum_{i=301}^{366}\left(Y_i - \widehat{m}(X_i)\right)^2 \text{ and } MSE_{REE} = \frac{1}{66}\sum_{i=301}^{366}\left(Y_i - \widetilde{r}(X_i)\right)^2.$$

5.1.2 Results and discussions

The results are given in the Figure 5.3 where two curves corresponding to the observed values (black curve) the predicted values (dashed curve green for the classical regression and red for the relative one) are drawn. Clearly, this Figure 5.3 shows the good behavior of our procedure. We observe that the relative approach gives better results than the classical regression approach ($MSE_{CKE} = 0.0883$ and $MSE_{REE} = 0.0034$).

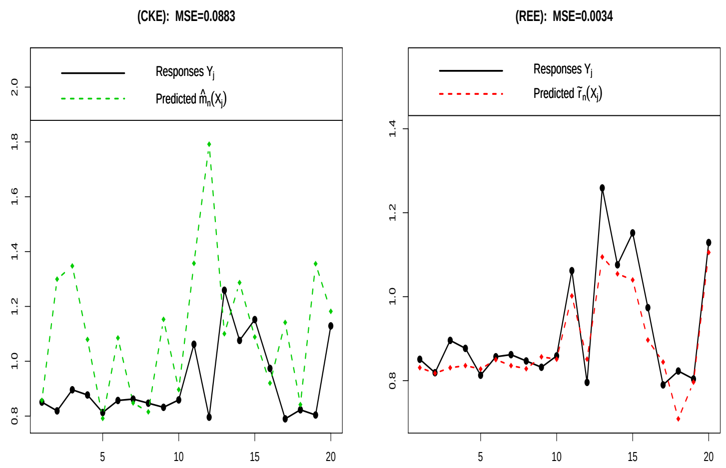

Figure 5.3: Prediction by classical and relative regression.

Now, we give in Table 5.1 the 90% predictive intervals of the concentrations for the peak load of the 20 last values in the sample test. This conclusion shows the good performance of our asymptotic normality.

Table 5.1: The 90% predictive intervals of the peak demand for the last 20 days.

The true value	Predicted value	Predictive intervals $IC_{90\%}$
0.851	0.8310	[0.6078 , 1.0542]
0.819	0.8177	[0.7376 , 0.8978]
0.896	0.8307	[0.7697 , 0.8918]
0.877	0.8358	[0.4879 , 1.1838]
0.813	0.8277	[0.4660 , 1.1894]
0.857	0.8501	[0.5713 , 1.1289]
0.862	0.8358	[0.7802 , 0.8914]
0.847	0.8284	[0.3206 , 1.3363]
0.832	0.8568	[0.7976 , 0.9160]
0.859	0.8511	[0.7328 , 0.9694]
1.062	1.0017	[0.8279 , 1.1756]
0.796	0.8514	[0.7592 , 0.9435]
1.259	1.0946	[0.9344 , 1.2548]
1.076	1.0545	[0.8648 , 1.2441]
1.152	1.0399	[0.9289 , 1.1508]
0.974	0.8968	[0.7833 , 1.0103]
0.790	0.8444	[0.7913 , 0.8974]
0.823	0.7091	[0.0456 , 1.3727]
0.804	0.7965	[0.6710 , 0.9219]
1.129	1.1054	[0.8670 , 1.3437]

5.2 NIR Spectrometric of Diesel Fuels (MAR case)

5.2.1 Materials and methods

The near infrared spectra of diesel fuel samples, together with six properties that were measured at the Southwest Research Institute, are obtained from the web site of Eigenvector Research Corporation (http://software.eigenvector.com/Data/index.html). It contains six different data sets for the parameters: cetane number, boiling point, freezing point, total aromatic content, viscosity, and density. The above physical and chemical properties of the samples were determined independently using standard reference methods before the near infrared spectra were recorded. The diesel fuel data set investigated here has also been used in a number of previous studies (Boger (2003); Esteban-Dıéz et al. (2004); Feng et al. (2015)) for testing new variable selection and calibration algorithms. For instance, let us consider a sample of $n = 480$ diesel fuels samples. Each sample is illuminated by a light beam at 401 equally spaced wavelengths $(\omega_1, \ldots, \omega_{401})$ in the near-infrared range $750 - 1550$ nm. For each wavelength ω and each diesel sample i, the absorption $X_i(\omega)$ of radiation is measured. The ith discretized spectrometric curve is given by $X_i(\omega_1), \ldots, X_i(\omega_{401})$; Figure 5.4 displays the 480 spectrometric curves.

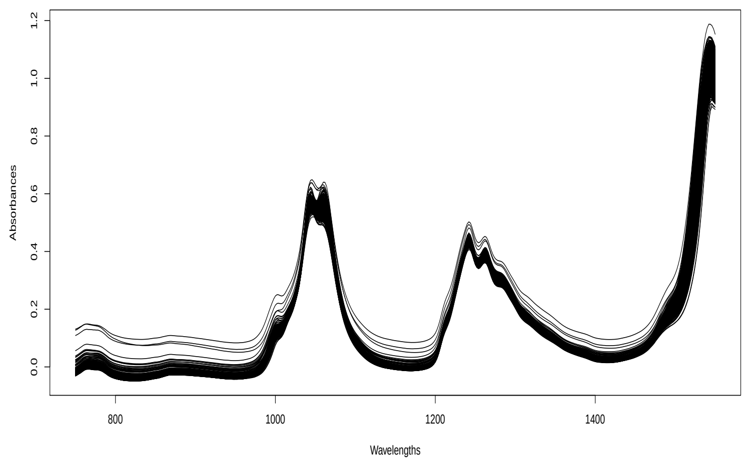

Figure 5.4: The 480 NIR spectroscopy curves of the diesel fuels data.

To fix the ideas, let's present our prediction problem. Indeed, assume that we aim to predict the content of certain diesel fuels parameters (in our case we are interested to predict the total aromatic content), denoted by Y_i, using the spectrometric curves associated X_i. Some of the properties Y_i are not measured on some of the samples, so diesel fuels parameters has some missing values (NaNs) in it (20% missing data). Therefore, our sample is formed as follows $(X_i, Y_i, \delta_i)_{i=1,\dots,480}$, where $\delta_i = 1$ if Y_i is observed and $\delta_i = 0$ else. We assume that the output variable Y and the input variable X are linked by the following regression formula,

$$Y = m(X) + \varepsilon, \tag{5.1}$$

We would like to estimate $m(x)$ by the MAR robust equivariant estimator $\widehat{\vartheta}_x$ and compare it with the models $\widehat{r}(x)$ for the classical missing estimator and the MAR robust estimator $\widehat{\theta}_x$.

The performance of all above mentioned models is closely linked with the use of different parameters involved in the estimation. In fact, the kernel is supposed to be the quadratic kernel defined by $K(u) = 1.5(1-u^2)\mathbb{1}_{[0,1]}$. Because the curves are very smooth, the L_2 distance between the second derivative of the curves is considered as a semi-metric $d(.,.)$ defined as:

$$d(x_i, x_j) = \sqrt{\int_0^1 \left(x_i''(\omega) - x_j''(\omega)\right)^2 d\omega}.$$

Finally, we considered the optimal bandwidth $h := h_{n,K}$ chosen by the cross-validation procedure:

$$h_{opt} = \arg\min_h CV(h) \qquad \text{where} \quad CV(h) = \sum_{i=1}^{n} \left(Y_i - \widetilde{Y}_{(-i)}(X_i)\right)^2,$$

with $\widetilde{Y}_{(-i)}(X_i)$ the values of the estimator $\widehat{m}(.)$, $\widehat{\theta}_.$ or $\widehat{\vartheta}_.$ calculate at X_i.

To evaluate the efficiency of the proposed models in this prediction issue, we randomly split the $n-$sample into two parts : one is a training sample $(X_i, Y_i)_{i=1,\dots,400}$ which is used to model, and the other is a testing sample $(X_j, Y_j)_{j=401,\dots,480}$ which is used to verify the prediction effect.
The testing sample provides the mean absolute error MAE of prediction:

$$MAE = \frac{1}{80}\sum_{j=401}^{480} \left|Y_j - \widetilde{Y}(X_j)\right|$$

where $\widetilde{Y}(X_j)$ the prediction values of the estimators $\widehat{m}(.)$, $\widehat{\theta}_.$ and $\widehat{\vartheta}_.$ calculate at X_j.

5.2.2 Results and discussions

The obtained prediction results are shown in the following Figure 5.5.

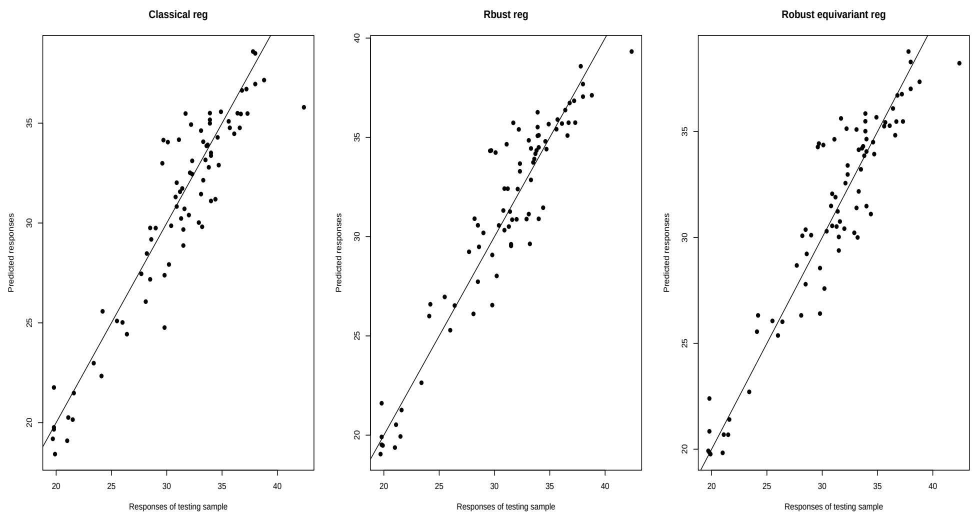

Figure 5.5: Prediction of the testing sample (Classical, Robust and Robust equivariant respectively).

Figure 5.5 gives an idea on the accuracy of the predictions corresponding to one run. They presents the 80 of the predictions: the observed values (horizontal axis), the predicted values (vertical axis). It is depicted in Figure 5.5 that there is no significant difference between the three models when there is no outliers in the learning sample.

To further explore the performances of our models, we carry out $M = 100$ independent replications which allows to compute 100 values for MAE and to display their distribution by means of a beanplot. Figure 5.6 shows the bean-plots of the MAE of the prediction values.

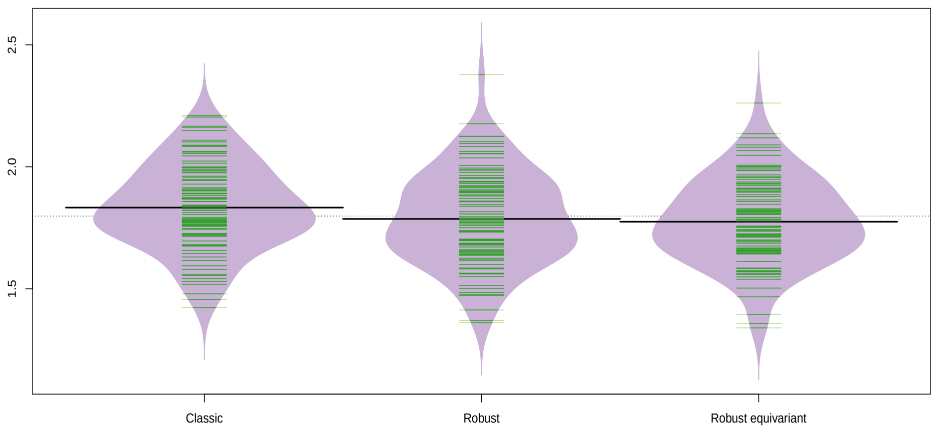

Figure 5.6: The bean-plots of the MAE of the prediction values by the three methods without outliers (Classical, Robust and Robust equivariant respectively).

Now, we concentrate on the comparison of models performances in the presence of outliers. For this aim, we introduce artificial outliers by multiplying 5 values of response Y in the learning sample by 100. The estimators of our models are obtained by the same previous selection methods of the smoothing parameter, i.e., the same metric d and also the same kernel K. Finally, the obtained results of the bean-plots of the MAE are shown in Figure 5.7. Note that, in Figure 5.6 the three estimators are equivalent but in Figure 5.7, in which we considered the presence of outliers, the Robust equivariant regression gives better results than the Classical and the robust methods: There is a small difference between the Robust equivariant regression and the Robust one, while the MAE is significantly large for the Classical model, in sense that; i.e., the classical method is susceptible to the presence of outliers.

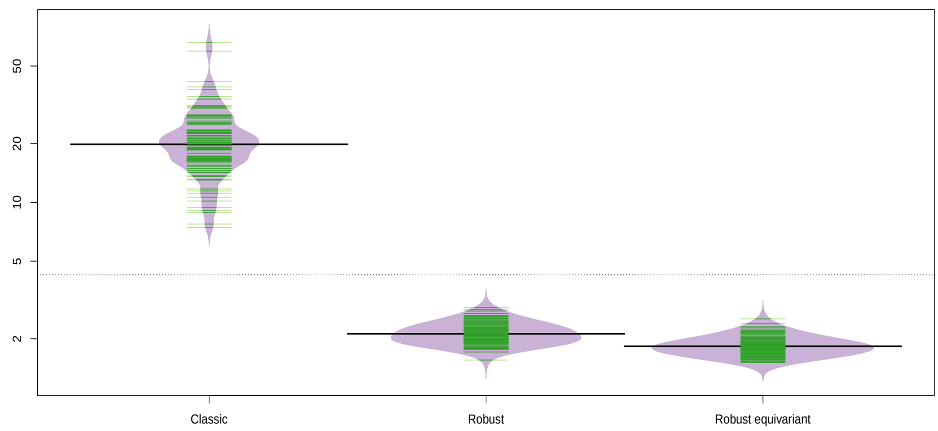

Figure 5.7: The bean-plots of the MAE of the prediction values by the three methods in the presence of outliers (Classical, Robust and Robust equivariant respectively).

BIBLIOGRAPHY

Z. Boger. Selection of quasi-optimal inputs in chemometrics modeling by artificial neural network analysis. *Analytica Chimica Acta*, 490:31–40, 2003.

I. Esteban-Dıéz, J. González-Sáiz, and C. Pizarro. Owavec: a combination of wavelet analysis and an orthogonalization algorithm as a pre-processing step in multivariate calibration. *Analytica chimica acta*, 515:31–41, 2004.

F. Feng, Q. Wu, and L. Zeng. Rapid analysis of diesel fuel properties by near infrared reflectance spectra. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, 149:271–278, 2015.

F. Ferraty and P. Vieu. *Nonparametric Functional Data Analysis: Theory and Practice*. Springer Series in Statistics, New York, springer edition, 2006.

M. Mebsout, K. M. Attouch, and O. Fetitah. Nonparametric m-regression with scale parameter for functional dependent data. *Applications & Applied Mathematics*, 15(2), 2020.

A. Pîrjan, S. V. Oprea, G. Cărutasu, D. M. Petrosanu, A. Bâra, and C. Coculescu. Devising hourly forecasting solutions regarding electricity consumption in the case of commercial center type consumers. *Energies*, 10:17–27, 2017.

GENERAL CONCLUSION AND PROSPECTS

Conclusion

The literature on missing data is still relevant, especially with regard to the estimation of the functional parameters present in this model. Recall that one of the main motivations for the craze of nonparametric functional statistics is the solution it offers for the problem of the scourge of dimension, and the power of computers which have made it possible to process data in very large dimensions.

In this thesis, we are interested in the robust estimation of the regression operator in the presence of missing data. It is clear that the superiority of this approach over the classical method is the main motivation for this subject. In order to highlight this superiority in NPFDA, we first studied, the asymptotic properties of a nonparametric estimator of the relative error regression given a functional explanatory variable, when the scalar response is right censored, in the i.i.d. case. We establish the strong almost complete convergence rate and asymptotic normality of these estimators.

As a first idea of extension, is to establish similar results when one frees oneself from the assumption of independence. It is well known that in practice several processes have a certain dependence. The second part of this thesis is devoted to the problem of estimating the relative regression operator when the observation are $\alpha-$mixing. We establish the almost complete convergence rate of these estimators. A simulation study and real data application are performed to illustrate how this fact allows getting higher predictive performances than those obtained with standard estimates.

Finally, it seems possible to us interested in studying the robust model given a functional explanatory variable, in the case of a scalar missing at random (MAR) response, for both cases, without and with unknown scale parameter. We establish, the almost complete convergence rate of our estimators in the two proposed models.

Prospects

To conclude the work of this thesis, many questions remain unanswered. We believe we will invest in the future on a few issues in order to improve and extend our results.

- We think it is possible to adapt our results to another type of dependency such as the quasi-associated and the ergodic case.
- Other issues are possible, such that extensions our estimators to the local linear ideas.
- Another possible prospect is to obtain the asymptotic normality of the robust equivariant regression for functional data with responses missing at random.
- Robust estimation with single functional index model can be approached in the missing case.
- We will be able to elaborate the asymptotic properties of our estimators based on the k nearest neighbor ($k-$NN) method or other methods on the bandwidth selection, because it allows the improvement of the quality of the estimator.
- We can generalize the results obtained using other models such as the aditive model or the semi-functional partial linear model.
- An important issue about the comparison of the constructed estimators when there are surrogate outputs.

BIBLIOGRAPHY

B. Altendji, J. Demongeot, A. Laksaci, and M. Rachdi. Functional data analysis: estimation of the relative error in functional regression under random left-truncation model . *Journal of Nonparametric Statistics*, 30:472–490, 2018.

A. Amiri and S. Khardani. Nonparametric regression estimation using recursive kernel for dependent censored data. *Int J Math Stat*, 19:52–82, 2018.

A. Amiri, C. Crambes, and B. Thiam. Recursive estimation of nonparametric regression with functional covariate. *Computational Statistics & Data Analysis*, 69:154–172, 2014.

D. F. Andrews. A robust method for multiple linear regression. *Technometrics*, 16:523–531, 1974.

M. Attouch, A. Laksaci, and E. Ould-Saïd. Asymptotic Distribution of Robust Estimator for Functional Nonparametric Models. *Communications in Statistics Theory and Methods*, 38:1317–1335, 2009.

M. Attouch, A. Laksaci, and E. Ould-Saïd. Strong uniform convergence rate of robust estimator of the regression function for functional and dependent processes. *Journal of the Japan Statistical Society*, 42:125–143, 2013.

M. Attouch, A. Laksaci, and N. Messabihi. Nonparametric relative error regression for spatial random variables. *Stat Papers*, 58, 2017.

G. Boente and R. Fraiman. Robust nonparametric regression estimation. *Journal of Multivariate Analysis*, 29:180–198, 1989.

G. Boente and D. Rodriguez. Robust estimators of high order derivatives of regression functions. *Statistics & probability letters*, 76:1335–1344, 2006.

G. Boente and A. Vahnovan. Strong convergence of robust equivariant nonparametric functional regression estimators. *Statistics and Probability Letters*, 11:1–11, 2015.

Z. Boger. Selection of quasi-optimal inputs in chemometrics modeling by artificial neural network analysis. *Analytica Chimica Acta*, 490:31–40, 2003.

D. Bosq. *Linear processes in function spaces: theory and applications*, volume 149. Springer Science & Business Media, 2012.

G. E. Box et al. Some theorems on quadratic forms applied in the study of analysis of variance problems, i. effect of inequality of variance in the one-way classification. *The annals of mathematical statistics*, 25(2):290–302, 1954.

J. Buckley and I. James. Linear regression with censored data. *Biometrika*, 66:429–436, 1979.

T. Cai, P. Hall, et al. Prediction in functional linear regression. *The Annals of Statistics*, 34:2159–2179, 2006.

M. Campbell and A. Donner. Classification efficiency of multinomial logistic regression relative to ordinal logistic regression . *J. Amer. Statist. Asso*, 84:587–591, 1989.

A. Carbonez, L. Gyorfi, and C. Edward. Partitioning-estimates of a regression function under random censoring . *Statist. Decisions*, 13:21–38, 1995.

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

M. Chaouch and S. Khardani. Randomly censored quantile regression estimation using functional stationary ergodic data. *Journal of Nonparametric Statistics*, 27:65–87, 2015a.

M. Chaouch and S. Khardani. Randomly censored quantile regression estimation using functional stationary ergodic data. *Journal of Nonparametric Statistics*, 27:65–87, 2015b.

K. Chen, S. Guo, Y. Lin, , and Z. . Ying. Least absolute relative error estimation. . *Journal of the American Statistical Association*, 105:1104–1112, 2010.

P. E. Cheng. Nonparametric estimation of mean functionals with data missing at random. *Journal of the American Statistical Association*, 89:81–87, 1994.

W. S. Cleveland. Robust locally weighted regression and smoothing scatterplots. *Journal of the American statistical association*, 74:829–836, 1979.

G. Collomb and W. Härdle. Strong uniform convergence rates in robust nonparametric time series analysis and prediction: Kernel regression estimation from dependent observations. *Stochastic processes and their applications*, 23:77–89, 1986.

C. Crainiceanu, P. Reiss, J. Goldsmith, L. Huang, L. Huo, F. Scheipl, S. Greven, J. Harezlak, M. Kundu, and Y. Zhao. refund: Regression with functional data. *R package version 0.1-6*, 2012.

S. Dabo-Niang. Kernel density estimator in an infinite–dimensional space with a rate of convergence in the case of diffusion process. *Applied Mathematics Letters*, 17:381–386, 2004.

P. Deheuvels and J. H. J. Einmahl. Functional limit laws for the increments of Kaplan Meier product-limit processes and applications . *The Annals of Probability*, 28:1301–1335, 2000.

J. Demongeot, A. Hamie, A. Laksaci, and M. Rachdi. Relative-Error Prediction in Nonparametric Functional Statistics: Theory and Practice . *Journal of Multivariate analysis*, 147:261–268, 2016.

P. Doukhan, P. Massart, and E. Rio. The functional central limit theorem for strongly mixing processes. In *Annales de l'IHP Probabilités et statistiques*, volume 30, pages 63–82, 1994.

S. Efromovich. Nonparametric regression with predictors missing at random. *Journal of the American Statistical Association*, 106:306–319, 2011.

I. Esteban-Dıéz, J. González-Sáiz, and C. Pizarro. Owavec: a combination of wavelet analysis and an orthogonalization algorithm as a pre-processing step in multivariate calibration. *Analytica chimica acta*, 515:31–41, 2004.

W. Feller. An introduction to probability theory and ito applications ii. *John Willy, New York*, 1966.

F. Feng, Q. Wu, and L. Zeng. Rapid analysis of diesel fuel properties by near infrared reflectance spectra. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, 149:271–278, 2015.

F. Ferraty and P. Vieu. The functional nonparametric model and application to spectrometric data. *Computational Statistics*, 17:545–564, 2002.

F. Ferraty and P. Vieu. Curves discrimination: a nonparametric functional approach. *Computational Statistics and Data Analysis*, 44:161–173, 2003.

F. Ferraty and P. Vieu. Nonparametric methods for functional data, with applications in regression, time–series prediction and curves discrimination. *J. Nonparametr. Stat*, 16: 111–125, 2004a.

F. Ferraty and P. Vieu. Nonparametric models for functional data, with application in regression, time series prediction and curve discrimination. *Journal of Nonparametric Statistics*, 16:111–125, 2004b.

F. Ferraty and P. Vieu. *Nonparametric Functional Data Analysis: Theory and Practice*. Springer Series in Statistics, New York, springer edition, 2006.

F. Ferraty, A. Goia, and P. Vieu. Functional nonparametric model for time series: a fractal approach for dimension reduction. *Test*, 11:317–344, 2002.

F. Ferraty, A. Laksaci, and P. Vieu. Estimating some characteristics of the conditional distribution in nonparametric functional models. *Statistical Inference for Stochastic Processes*, 9:47–76, 2006.

F. Ferraty, A. Mas, and P. Vieu. Nonparametric Regression on Functional Data: Inference and Practical Aspects. *Australian and New Zealand Journal of Statistics*, 49:267–286, 2007.

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

O. Fetitah, I. M. Almanjahie, M. K. Attouch, and A. Righi. Strong convergence of the functional nonparametric relative error regression estimator under right censoring. *Mathematica Slovaca*, 70(6):1469–1490, 2020.

A. Gheriballah, A. Laksaci, and S. Sekkal. Nonparametric m-regression for functional ergodic data. *Statistics & Probability Letters*, 83:902–908, 2013.

J. Goldsmith, J. Bobb, C. Crainiceanu, B. Caffo, and D. Reich. Penalized functional regression. *Journal of computational and graphical statistics*, 20:830–851, 2011.

F. R. Hampel. A general qualitative definition of robustness. *The Annals of Mathematical Statistics*, pages 1887–1896, 1971.

W. Härdle. Robust regression function estimation. *Journal of Multivariate Analysis*, 14: 169–180, 1984.

W. Härdle and T. Gasser. Robust nonparametric function fitting. *Journal of the Royal Statistical Society: Series B (Methodological)*, 46:42–51, 1984.

W. Härdle, A. B. Tsybakov, et al. Robust nonparametric regression with simultaneous scale curve estimation. *The annals of statistics*, 16:120–135, 1988.

N. Helal and E. Ould-Saïd. Kernel Conditional Quantile Estimator under Left Truncation for Functional Regressors . *Opuscula Mathematica*, 36:25–48, 2016.

W. Horrigue and E. Ould-Saïd. Strong Uniform Consistency of a Nonparametric Estimator of a Conditional Quantile for Censored Dependent Data and Functional Regressors . *Random Operators and Stochastic Equations*, 19:131–156, 2011.

W. Horrigue and E. Ould-Saïd. Nonparametric Regression Quantile Estimation for Dependant Functional Data under Random Censorship: Asymptotic Normality . *Communications in Statistics - Theory and Methods*, 44:4307–4332, 2014.

P. Huber and E. Ronchetti. Robust statistics john wiley & sons. *New York*, 1(1), 1981.

P. J. Huber. A robust version of the probability ratio test. *The Annals of Mathematical Statistics*, 36:1753–1758, 1965.

P. J. Huber. Robust estimation of a location parameter. In *Breakthroughs in statistics*, pages 492–518. Springer, 1992.

P. J. Huber et al. Robust regression: asymptotics, conjectures and monte carlo. *The annals of statistics*, 1:799–821, 1973.

M. Jones, P. Heungsun, S. Key-Il, and S. Vines. Relative error prediction via kernel regression smoothers. . *J.Statist. Plann. Inference*, 138:2887–2898, 2008.

E. Kaplan, L, and P. , Meier. Nonparametric estimation from incomplete observations . *Journal of the American statistical association*, 53:457–481, 1958.

S. Khardani and B. Thiam. Strong consistency result of a non parametric conditional mode estimator under random censorship for functional regressors . *Communications in Statistics-Theory and Methods*, 45:1863–1875, 2016.

M. Köhler, K. Máthé, and M. Pintér. Prediction from randomly right censored data . *J. Multivariate Anal.*, 80:73–100, 2002.

W. S. Krasker and R. E. Welsch. Efficient bounded-influence regression estimation. *Journal of the American statistical Association*, 77:595–604, 1982.

N. Laïb and E. Ould-Saïd. A robust nonparametric estimation of the autoregression function under an ergodic hypothesis. *Canadian Journal of Statistics*, 28:817–828, 2000.

P. Laplace. Deuxiéme supplément to laplace (1812). *Oeuvr compl.*, 7:531–580, 1818.

R. J. A. Little and D. B. Rubin. *Statistical analysis with missing data*, volume 793. Wiley, 2019.

M. Loeve. *Probability theory*, volume 1963. Springer, 1963.

R. A. Maronna and D. Martin. Yohai. robust statistics. *Wiley Series in Probability and Statistics. John Wiley and Sons*, 2:3, 2006.

E. Masry. Recursive probability density estimation for weakly dependent stationary processes. *IEEE Transactions on Information Theory*, 32:254–267, 1986.

E. Masry. Nonparametric regression estimation for dependent functional data: asymptotic normality. *Stochastic Processes and their Applications*, 115:155–177, 2005.

M. Mebsout, K. M. Attouch, and O. Fetitah. Nonparametric m-regression with scale parameter for functional dependent data. *Applications & Applied Mathematics*, 15(2), 2020.

W. Mechab and A. Laksaci. Nonparametric relative regression for associated random variables. *Metron*, 74:75–97, 2016.

E. Ould-Saïd and Z. Guessoum. On nonparametric estimation of the regression function under random censorship model . *Statistics and Decisions*, 26:159–177, 2008.

H. Park and L. Stefanski. Relative-error prediction. *Statistics & probability letters*, 40: 227–236, 1998.

A. Pîrjan, S. V. Oprea, G. Cărutasu, D. M. Petrosanu, A. Bâra, and C. Coculescu. Devising hourly forecasting solutions regarding electricity consumption in the case of commercial center type consumers. *Energies*, 10:17–27, 2017.

J. O. Ramsay. Functional data analysis. *Encyclopedia of Statistical Sciences*, 4, 2004.

J. O. Ramsay and C. J. Dalzell. Some tools for functional data analysis. *Journal of the Royal Statistical Society: Series B (Methodological)*, 53:539–561, 1991.

J. O. Ramsay and B. W. Silverman. Principal components analysis for functional data. In *Functional Data Analysis*, pages 147–172. Springer, 2005.

T. Randolph, J. Harezlak, and Z. Feng. Structured penalties for functional linear models—partially empirical eigenvectors for regression. *Electronic journal of statistics*, 6: 323, 2012.

E. Rio. *Théorie asymptotique des processus aléatoires faiblement dépendants*, volume 31. Springer Science & Business Media, 1999.

S. Ruiz-Velasco. Asymptotic efficiency of logistic regression relative to linear discriminant analysis . *Biometrika*, 78:235–243, 1991.

W. Stute. Consistent estimation under random censorship when covariables are present. *Journal of Multivariate Analysis*, 45:89–103, 1993.

B. Thiam. Relative error prediction in nonparametric deconvolution regression model . *Statistica Neerlandica*, 10:1–15, 2018.

A. B. Tsybakov. Robust estimates of a function. *Problemy Peredachi Informatsii*, 18: 39–52, 1982.

F. Yao, H. G. Müller, and J. L. Wang. Functional data analysis for sparse longitudinal data. *Journal of the American Statistical Association*, 100:577–590, 2005.

ملخص

في هذه الرسالة، قمنا بدراسة بعض خصائص مقاربة في الإحصائيات اللامعلمية للبيانات الدالية غير الكاملة. بتعبير أدق، نحن مهتمون بالانحدار المتين والنسبي الذي نبني من أجله المقدرات ونقوم بدراسة نمط التقارب في النموذج البيانات الخاضعة للرقابة أو المفقودة. ندرس أولاً خصائص المقاربة لمقدر غير معلمي لانحدار الخطأ النسبي في ضوء متغير دالي، عندما يكون متغير الاستجابة غير دالي وخاضع للرقابة على اليمين، في حالة الانفصال العشوائي أو الارتباط من نوع α. و بعد ذلك، يبدو من الممكن لنا دراسة النموذج المتين، في حالة الاستجابة العددية المفقودة (MAR)، في كلتا الحالتين، بدون ومع معامل مقياس غير معروف.

Résumé

Dans cette thèse, nous étudions les propriétés asymptotiques des paramètres fonctionnels en statistique non paramétrique pour des données incomplètes. Plus précisément, nous nous intéressons à la régression robuste et relative pour lesquelles nous construisons des estimateurs et nous étudions le comportement asymptotique dans le modèle censuré et manquantes. Nous avons d'abord étudié les propriétés asymptotiques d'un estimateur non paramétrique de la régression d'erreur relative étant donné une variable explicative fonctionnelle, lorsque la réponse scalaire est censurée à droite, dans les cas i.i.d. et $\alpha-$mélange. Ensuite, il nous semble possible d'étudier le modèle robuste, dans le cas d'une réponse scalaire manquante (MAR), dans les deux cas, sans et avec paramètre d'échelle inconnu.

Abstract

In this thesis, we study the asymptotic properties of functional parameters in nonparametric statistics for incomplete data. More precisely, we are interested in the robust and relative regression for which we build estimators, and we study the asymptotic behavior in the censored and missing model. We first studied, the asymptotic properties of a nonparametric estimator of the relative error regression given a functional explanatory variable, when the scalar response is right censored, in the i.i.d. case and $\alpha-$mixing case. Then, it seems possible to us to study the robust model, in the case of a scalar missing at random (MAR) response, for both cases, without and with unknown scale parameter.