Applications and Applied Mathematics: An International Journal (AAM)

Volume 15 | Issue 2 | Article 8

12-2020

Nonparametric M-regression with Scale Parameter For Functional Dependent Data

Mebsout Mokhtaria
Djillali Liabes University

Attouch M. Kadi
Djillali Liabes University of Sidi Bel Abbes, King Khalid University

Fetitah Omar
Djillali Liabes University of Sidi Bel Abbes

Available at
http://pvamu.edu/aam
Appl. Appl. Math.
ISSN: 1932-9466

Applications and Applied Mathematics:
An International Journal
(AAM)

Vol. 15, Issue 2 (December 2020), pp. 846 – 874

Nonparametric M-regression with Scale Parameter For Functional Dependent Data

[1]Mebsout Mokhtaria, [2,*]Attouch Mohammed Kadi and [3]Fetitah Omar

[1,2,3] Laboratoire de Statistique Processus Stochastiques
Djillali Liabes University
Sidi Bel Abbes, 22000
Algeria
[1]mokhtariamebsout@yahoo.com; [3]fetitah-omar@hotmail.com

[2] Mathematic Department, King Khalid University
Abha, 61421
Saudi Arabia
[2]attou_kadi@yahoo.fr

*Corresponding Author

Received: June 12, 2020; Accepted: October 25, 2020

Abstract

In this paper, we study the equivariant nonparametric robust regression estimation relationship between a functional dependent random covariable and a scalar response. We consider a new robust regression estimator when the scale parameter is unknown. The consistency result of the proposed estimator is studied, namely the uniform almost complete convergence (with rate). Thus, suitable topological considerations are needed, implying changes in the convergence rates, which are quantified by entropy considerations. The benefits of considering robust estimators are illustrated on two real data sets where the robust fit reveals the presence of influential outliers.

Keywords: Scale parameter; Robust function; Functional dependent data; Nonparametric estimation

MSC 2010 No.: 62G35, 62G08, 62G05

1. Introduction

Studying the relationship between a random variable Y and a set of covariates X in comparison with usual regression methods is a very relevant topic and at the same time it considered a common problem in nonparametric statistics, and there are several ways to explain this relationship. In many applications, the covariates X can be seen as functions recorded over a period of time instead of finite-dimensional vectors, (see Ferraty and Vieu (2006)) for an extensive discussion on nonparametric statistics for functional data.

In this general framework, statistical models adapted to infinite-dimensional data have been recently studied. We refer to Ramsay and Silverman (2002), Ramsay and Silverman (2005) and Ferraty and Vieu (2006) for a description of different procedures for functional data. Linear nonparametric regression estimators in the functional setting, that is, estimators based on a weighted average of the response variables, have been considered by several authors such as Benhenni et al. (2007) and Ferraty et al. (2006), who also considered estimators of the conditional quantiles. The literature on robust proposals for nonparametric regression estimation is sparse. Motivated by its flexibility when data are affected by outliers, the robust regression was widely studied in nonparametric functional statistics. Indeed, it was firstly introduced by Azzedine et al. (2008) who proved the almost-complete convergence of this model in the independent and identically distributed (i.i.d.) case. Since this work, several results on the nonparametric robust functional regression were realized (see, for instance, Attouch et al. (2010), Attouch et al. (2012), Attouch et al. (2019), Gheriballah et al. (2013), Boente and Vahnovanb (2015) and references therein for some key references on this topic). Notice that all these results are obtained when the scale parameter is supposed to be known.

In this sense, we extend some of the previous works in two directions. On the one hand, we generalize the proposal given in the Euclidean case by Boente and Fraiman (1989a) to provide robust equivariant estimators for the regression function in the functional case, that is, in the case where the covariates are in an infinite-dimensional space. On the other side, we extend the proposal given in Azzedine et al. (2008) to allow for an unknown scale, and heteroscedastic models are provided in Boente and Vahnovanb (2015). The main goal of this paper is to study the uniform convergence of this nonparametric estimator with an unknown scale parameter when the explanatory variable X is valued in infinite dimension space and the observations $(X_n, Y_n)_{n\geq 1}$ are strongly mixing.

The paper is organized as follows. In Section 2, we state our notation and introduce the robust equivariant estimators. Section 3 contains the main results of this paper, namely, the uniform convergence consistency and uniform convergence rates over compact sets of the equivariant local M-estimators. In Section 4, we examine the performances of our proposed estimator with two real data sets applications.

2. Basic definitions and notation

Consider $Z_i = (X_i, Y_i)_{i=1\cdots n}$ be n copies of random vector, identically distributed as (X, Y) and is valued in $\mathcal{F} \times \mathbb{R}$, where $\mathcal{F}$ is a semi-metric space, d denoting the semi-metric. For any $x \in \mathcal{F}$, we consider a real-valued Borel function ψ, and stated the model of the covariation between X_i and Y_i. Our nonparametric regression function, denoted by θ_x is implicitly defined as a zero with respect to (w.r.t.) t of the following equation.

Let us define (X, Y) be a random element in $\mathcal{F} \times \mathbb{R}$ and let

$$\Psi(x, t, \sigma) = \mathbb{E}\left(\psi\left(\frac{Y-t}{\sigma}\right)/X = x\right), \tag{1}$$

where $\psi : \mathbb{R} \times \mathbb{R}$ is an odd, bounded and continuous function satisfying some regularity conditions to be stated below. In the following, we assume that Equation (1) allows θ as a unique solution(see, for instance, (Boente and Fraiman (1989a)) for sufficient conditions for existence and uniqueness of θ). In addition, our robustification method allows us to consider the functional nonparametric regression model with a scale of the error assumed to be unknown, where $\sigma(.)$ is a measure of spread for the conditional distribution of Y given $X = x$. We return to Stone (2005) for other examples of the function ψ.

The conditional scale measure can be taken as the conditional median of the absolute deviation from the conditional median, that is,

$$s(x) = MED(|Y - m(x)|/X = x) = MAD_c(F_Y^x(.)), \tag{2}$$

where $m(x) = MED(Y/X = x)$ is the median of the conditional distribution.

Note that $s(x)$, which corresponds to a robust measure of the conditional scale, usually equals $\sigma(X)$ up to a multiplicative constant, when $\epsilon = \sigma(X)u$ with u independent of X. For instance, the median of the absolute deviation is usually calibrated so that $MAD(\phi) = 1$, where ϕ states for the distribution function of a standard normal random variable. In this case, when the errors u have a Gaussian distribution, we x have that $s(x) = MAD_c(F_Y^x(.)) = \sigma(x)$. To obtain estimators of $\theta(x)$ we plug into (1) an estimator of $F_Y^x(y)$, which will be taken as $\widehat{F}(y/X = x)$. Denote by $\widehat{s}(x)$ a robust estimator of the conditional scale, for instance, $\widehat{s}(x) = MAD_c\left(\widehat{F}(./X = x)\right)$, the scale measure defined in (2) evaluated in $\widehat{F}(y/X = x)$. With this notation, the robust nonparametric estimator of $\theta(x)$ is given by the solution $\widehat{\theta}(x)$ of $\widehat{\Psi}(x, t, \widehat{s}(x)) = 0$, where

$$\widehat{\Psi}(x, t, \widehat{\sigma}) = \int \psi\left(\frac{y-t}{\widehat{\sigma}}\right) d\widehat{F}(y/X = x) = \sum_{i=1}^{n} w_i(x)\psi\left(\frac{Y_i - t}{\widehat{\sigma}}\right), \tag{3}$$

where $w_i(x) = \dfrac{K\left(\frac{d(X_i,x)}{h_k}\right)}{\sum_{i=1}^{n} K\left(\dfrac{d(X_i, x)}{h_k}\right)}$.

3. Main results

Uniform convergence results and uniform convergence rates over compacts for the local M-estimators are derived under some general assumptions that are described below. From now on, $\xrightarrow{a.co}$ and a.co stand for almost complete convergence while $\xrightarrow{a.s}$ stands for almost sure convergence.

Throughout this paper, when no confusion will be possible, we will denote by C and C' some strictly positive generic constants. x is a fixed point in $\mathcal{F}$ and $\mathcal{N}_x$ denotes a fixed neighborhood of x. We consider $S_{\mathbb{R}} \subset \mathbb{R}$ and $S_{\mathcal{F}} \subset \mathcal{F}$ compact subsets of non empty interior. For $r > 0$, let $B(x,r) =: \{x' \in \mathcal{F}/d(x',x) < r\}$.

In order to define the strong mixing property, introduce the following notations. Denote by $\mathcal{F}_1^k$ the σ-algebra generated by $(X_1, Y_1), \ldots, (X_k, Y_k)$ and $\mathcal{F}_{k+n}^{\infty}$ that generated by $(X_{k+n}, Y_{k+n}), \ldots$.

Let's define, for any $n \geq 1$,

$$\alpha(n) = \sup_{A \in \mathcal{F}_1^k} \sup_{B \in \mathcal{F}_{k+n}^{\infty}} \{|\mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)|\}. \tag{4}$$

The process $(X_n, Y_n)_{n \geq 1}$ is said to be strongly mixing if

$$\lim_{n \to \infty} \alpha(n) = 0. \tag{5}$$

There exists many processes fulfilling the strong mixing property. We quote, here, the usual ARMA processes which are geometricaly strongly mixing, i.e., there exists $\rho \in (0,1)$ and $a > 0$ such that, for any $n \geq 1$, $\alpha(n) \leq a\rho^n$ (see, e.g., Rio (2000) for more detail).

Before giving the main asymptotic result, we need some assumptions.

(**H1**) Let's denote by $\phi_x(h) = \mathbb{P}(X \in B(x,h)) = \mathbb{P}[X \in \{x' \in \varepsilon; d(x,x') < h\}]$, and we suppose that $\phi_x(h)$ is continuous, strictly increasing in a neighborhood of 0 and $\phi_x(0) = 0$.

(**H2**) The function Ψ is such that

- **i)** The function $\Psi(x,t,\sigma)$ is of class $\mathcal{C}^2$ on $[\theta_x - \tau, \theta_x + \tau],\ \tau > 0$.
- **ii)** $\forall (a_1, a_2) \in [\theta_x - \tau, \theta_x + \tau] \times [\theta_x - \tau, \theta_x + \tau]$, $\forall (x_1, x_2) \in \mathcal{N}_x \times \mathcal{N}_x$,
$$|\Psi(x_1, a_1, \sigma) - \Psi(x_2, a_2, \sigma)| \leq C d^{\gamma_1}(x_1, x_2) + |a_1 - a_2|^{\gamma_1}, \gamma_1,\ \gamma_2 > 0.$$
- **iii)** For each fixed $t \in [\theta_x - \tau, \theta_x + \tau]$, the function $\Psi(x,t,\sigma)$ is continuous at the point x.

(**H3**) **i)** ψ is a monotone function w.r.t. the second component.

- **ii)** $\psi : \mathbb{R} \to \mathbb{R}$ is an odd, strictly convex and increasing function, continuously differentiable with bounded derivative ψ, such that $\xi(u) = u\psi'(u)$.
- **iii)** For each fixed $t \in [\theta_x - \tau, \theta_x + \tau]$, $\mathbb{E}\left[\psi\left(\frac{Y-t}{\sigma}\right)/X\right] < C < \infty, a.s.$

(**H4**) The kernel K is a bounded nonnegative function with support [0,1), such that $|\int_0^1 K(u)du| < \infty$, and satisfies Lipschitz condition of order $\gamma_1 > 0$,

$$\exists L_K < \infty, |K(u) - K(v)| \leq L_K |u - v|^{\gamma_1},$$

and

i) If $K(1) = 0$, K is a differentiable with derivative K' and

$$-\infty < \inf_{u \in [0,1]} K'(u) \leq \sup_{u \in [0,1]} K'(u) = \|K'\|_\infty < 0.$$

ii) If $K(1) > 0$, there exist two real constants $C_1, C_2, 0 < C_1 < C_2 < \infty$ such that

$$0 < C_1 \mathbb{I}_{[0,1]} < K < C_2 \mathbb{I}_{[0,1]} < \infty.$$

(**H5**) Let $S_{\mathcal{F}}$ be a compact of $\mathcal{F}$ such that:

i) $F(y/X = x)$ is symmetric around θ_x.
ii) $F(y/X = x)$ has a unique median $m(x)$.
iii) For each y fixed $F(y/X = x)$ is a uniformly continuous function of x in a neighborhood of $S_{\mathcal{F}}$.
iv) The following equicontinuity condition holds,

$$\forall \epsilon > 0, \exists \delta > 0; |u - v| < \delta \Rightarrow \sup_{x \in S_{\mathcal{F}}} |F(u/X = x) - F(v/X = x)| < \epsilon.$$

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

i) $\exists\, C > 0$ and $\exists\, \eta_0 > 0$ such that for all $\eta < \eta_0,\ \phi_x'(\eta) < C$. If $K(1) = 0$, the function ϕ_x satisfy the additional condition:

$$\exists\, C > 0, \text{ and } \exists\, \eta_0 > 0, \text{such that} \ \forall\, 0 < \eta < \eta_0 \quad \int_0^\eta \phi_x(u)du > C\eta\phi_x(\eta).$$

ii) For n large enough,

$$\frac{(\log(n))^2}{n\phi_x(h)} < \Gamma_{S_{\mathcal{F}}}\left(\frac{\log(n)}{n}\right) < \frac{n\phi_x(h)}{\log(n)}.$$

(**H7**) The sequence $h = h_n$ is such that $h_n \to 0$,

$$n\phi_x(h_n) \to \infty \text{ and } \frac{n\phi_x(h_n)}{\log(n)} \text{ as } n \to \infty.$$

(**H8**) **i)** $\forall (y_1, y_2) \in S_{\mathbb{R}} \times S_{\mathbb{R}}, \forall (x_1, x_2) \in S_{\mathcal{F}} \times S_{\mathcal{F}}$,

$$|F(y_1/X = x_1) - F(y_2/X = x_2)| \leq C\left(d(x_1, x_2)^{\gamma_1} + C|y_1 - y_2|^{\gamma_2}\right).$$

ii) The function $F(y/X = x)$ is uniformly Lipschitz in a neighborhood $S_{\mathcal{F}}$ of $\mathcal{F}$, and there exists a constant $C > 0$ and $\gamma_1 > 0$, such that, for $x_1, x_2 \in S_{\mathcal{F}}$,

$$\sup_{y \in \mathbb{R}} |F(y/X = x_1) - F(y/X = x_2)| \leq C d^{\gamma_1}(x_1, x_2).$$

(**H9**) The Kolmogov's ϵ-entropy of $S_{\mathcal{F}}$, satisfy

$$\sum_{i=1}^{\infty} n \exp\left\{(1-\beta)\Gamma_{S_{\mathcal{F}}}\left(\frac{\log(n)}{n}\right)\right\} < \infty \qquad \text{for some } \beta > 1. \tag{6}$$

(**H10**) The sequence $(X_i, Y_i)_{i\in\mathbb{N}}$ satisfies:
i) $\exists a > 0, \exists c > 0 : \forall n \in \mathbb{N}, \alpha(n) \leq cn^{-a}$.
ii)

$$\begin{cases} \forall i \neq j, \forall t \in [\theta_x - \tau, \theta_x + \tau], \quad \mathbb{E}\left[\psi\left(\frac{Y_i - t}{\sigma}\right)\psi\left(\frac{Y_j - t}{\sigma}\right) \mid X_i, X_j\right] \leq C < \infty, \\ \mathbb{P}\left((X_i, X_j) \in B(x,r) \times B(x,r)\right) = \varphi_x(r) > 0. \end{cases}$$

iii) There exists $\eta > 0$, such that

$$Cn^{2(2-a)/(a+1)+\eta} \leq \max\left(\phi_x^2(r), \varphi_x(r)\right) \leq n^{2a/(2-a)p},$$

where $a > \max\left(5/2, \frac{4p+1+\sqrt{24p+1}}{p-1}\right)$.

The quantity $\Gamma_S(\varepsilon) = \log(N_\varepsilon(S))$, where $N_\varepsilon(S)$ denote the minimal number of open balls in $\mathcal{F}$ of radius ε which is necessary to cover $S_{\mathcal{F}}$. This concept was introduced by Kolmogorov in the mid-1950s (see Kolmogorov and Tikhomirov (1959)) and it represents a measure of the complexity of a set, in sense that, high entropy means that much information is needed to describe an element with an accuracy ε.

3.1. Uniform strong convergence results

In this section, we will obtain uniform convergence results over compact sets of the regression estimator $\widehat{\theta}$ defined as a solution of $\widehat{\Psi}(x, t, \widehat{s}(x)) = 0$. Theorem 3.1 generalizes the result obtained in Lemma (6.5) in Ferraty and Vieu (2006). Note that it is the functional version of Theorem (3.1) in Boente and Fraiman (1991).

Theorem 3.1.

Let $S_{\mathcal{F}} \subset \mathcal{F}$ be a compact set. Assume that $(H1), (H4), (H5)iii)iv), (H6)$ and $(H7)$ holds. Then, we have that

$$\sup_{x\in S_{\mathcal{F}}} |\widehat{F}(y/X=x) - F(y/X=x)| \overset{a.s}{\to} 0 \ \text{ as } \ n \to 0. \tag{7}$$

Proof:

The proofs are analogous to the one given in Boente and Vahnovanb (2015). Moreover, to prove Theorem 3.1 and 3.2, we need to fixing some notation.

Given fixed $y \in \mathbb{R}$, we denote by $W_i^j = \mathbb{I}_{(-\infty,y]}(Y_i)$, and for $j = 0, 1$,

$$\tilde{v}_j(x) = \frac{1}{n}\sum_{i=1}^{n} W_i^j \frac{K_i(x)}{\mathbb{E}K_1(x)}, \forall n \geq 1, \tag{8}$$

$$\upsilon_j(x) = \frac{1}{n}\sum_{i=1}^{n} W_i^j \frac{K_i(x)}{\phi_x(h)}, \tag{9}$$

with

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

hence,

$$\widehat{F}(y/X = x) = \frac{\tilde{\upsilon}_1(x)}{\tilde{\upsilon}_0(x)}.$$

As in Collomb (1982), we have the following bounds:

$$\sup_{x\in S_{\mathcal{F}}} |\widehat{F}(y|X = x) - F(y|X = x)| \leq \frac{1}{\inf_{x\in S_{\mathcal{F}}} \tilde{\upsilon}_0(t)} \sup_{x\in S_{\mathcal{F}}} |\tilde{\upsilon}_1(x) - \mathbb{E}(\tilde{\upsilon}_1(x))|$$

$$+\frac{1}{\inf\limits_{x\in S_{\mathcal{F}}} \tilde{\upsilon}_0(t)} \sup_{x\in S_{\mathcal{F}}} |\tilde{\upsilon}_0(x) - \mathbb{E}(\tilde{\upsilon}_0(x))|$$

$$+\frac{1}{\inf\limits_{x\in S_{\mathcal{F}}} \tilde{\upsilon}_0(x)} \sup_{x\in S_{\mathcal{F}}} |\mathbb{E}(\tilde{\upsilon}_1(x)) - F(y/X = x)\mathbb{E}(\tilde{\upsilon}_0(x))|.$$

To prove (7), we need to show that, $\forall y \in \mathbb{R}$,

$$\sup_{x\in S_{\mathcal{F}}} |\tilde{\upsilon}_j(x) - \mathbb{E}(\tilde{\upsilon}_j(x))| \stackrel{a.co}{\to} 0, \tag{10}$$

$$\sup_{x\in S_{\mathcal{F}}} |\mathbb{E}(\tilde{\upsilon}_1(x)) - F(y/X = x)\mathbb{E}(\tilde{\upsilon}_0(x))| \to 0, \tag{11}$$

and for some $a > 0$, we have

$$\sum_{n\geq 1} \mathbb{P}\left(\inf_{x\in S_{\mathcal{F}}} \tilde{\upsilon}_0(x) < a\right) < \infty. \tag{12}$$

Note that Lemmas (4.3) and (4.4) in Ferraty and Vieu (2006), the assumptions $(H1), (H4)$ and $(H6)$ imply that there exists a constants such that

$$\forall x \in S_{\mathcal{F}}, \quad \exists\, 0 < C < C' < \infty,\ \ C\phi_x(h) < \mathbb{E}\left[K_1(x)\right] < C'\phi_x(h). \tag{13}$$

Then, if $\tilde{C} = 1/C$, we have

$$|\tilde{\upsilon}_j(x) - \mathbb{E}(\tilde{\upsilon}_j(x))| \leq \tilde{C}|\upsilon_j(x) - \mathbb{E}(\upsilon_j(x))|.$$

Therefore, to obtain (10), it will be enough to show that

$$\sup_{x\in S_{\mathcal{F}}} |\upsilon_j(x) - \mathbb{E}(\upsilon_j(x))| \stackrel{a.co}{\to} 0. \tag{14}$$

Note that $\mathbb{E}[\upsilon_0(x)] = 1$, and using the fact that

$$\inf_{x\in S_{\mathcal{F}}} \tilde{\upsilon}_0(x) \geq \inf_{x\in S_{\mathcal{F}}} \mathbb{E}[\upsilon_0(x)] - \sup_{x\in S_{\mathcal{F}}} |\tilde{\upsilon}_0(x) - \mathbb{E}[\tilde{\upsilon}_0(x)]|,$$

we can deduce (14) and (12) by using the following Lemmas.

Lemma 3.1.

Let, $S_{\mathcal{F}} \subset \mathcal{F}$, be a compact set of non empty interior. Assume $(H1)$, $(H4)$ and $(H6)$. Then, for $j = 0, 1$, we have

a) For any $\epsilon > 0$,

$$\sup_{y\in\mathbb{R}} \sup_{x\in S_{\mathcal{F}}} \mathbb{P}\{|\upsilon_j(x) - \mathbb{E}[\upsilon_j(x)]| > \epsilon\} \leq C \exp\left(-\frac{\epsilon^2 n\phi_x(h)}{2C'\|K\|_\infty^2 \left(1 + \frac{2\epsilon}{C'\|K\|_\infty}\right)}\right) + Cnr^{-1}\left(\frac{r}{\epsilon}\right).$$

b) For any $\epsilon > 0$,

$$\mathbb{P}\{\sup_{x\in S_{\mathcal{F}}} \mathbb{P}|\upsilon_j(x) - \mathbb{E}[\upsilon_j(x)]| > \epsilon\} \leq N_\rho(S_{\mathcal{F}}) \exp\left(-\frac{\epsilon^2 n\phi_x(h)}{a_1(1+\epsilon a_2)}\right) + Cnr^{-1}\left(\frac{r}{\epsilon}\right).$$

c) There exists $c > 2$ such that, for any $\epsilon_0 > c$ and $n \geq n_0$, we have

$$\sup_{y\in\mathbb{R}} \mathbb{P}\left\{\theta_n^{-1} \sup_{x\in S_{\mathcal{F}}} \mathbb{P}|\upsilon_j(x) - \mathbb{E}[\upsilon_j(x)]| > \epsilon_0\right\} \leq \exp\left(-\frac{\epsilon_0^2}{8(1+\epsilon_0)}\Gamma_{S_{\mathcal{F}}}\left(\frac{\log(n)}{n}\right)\right) + n(\log n)^{-2}\left(\frac{4C(\log n)^2}{n\epsilon}\right)^{a+1}.$$

Proof:

Proof of Part a).

Let's denote by $Z_i = W_i^j \Pi_i - \mathbb{E}(W_i^j \Pi_i)$, where $|W_i^j| \leq 1$ and $\Pi_i = K_i(x)/\phi_x(h)$, the kernel K is bounded, and let $C_1 = \|K\|_\infty$, such that

$$D_n(x) = \upsilon_j(x) - \mathbb{E}[\upsilon_j(x)]. \tag{15}$$

Thus,

$$D_n(x) = \frac{1}{n\phi_x(h)} \sum_{i=1}^n \left[W_i^j K_i(x) - \mathbb{E}\left[W_i^j K_i(x)\right]\right], \quad \text{for } j = 0, 1.$$

We have

$$\frac{1}{n\phi_x(h)} \left[W_i^j K_i(x) - \mathbb{E}\left[W_i^j K_i(x)\right]\right] \leq \frac{1}{n\phi_x(h)} \left[K_i(x) - \mathbb{E}\left[K_i(x)\right]\right],$$

and denote by

$$\Delta_i(x) = \frac{1}{\phi_x(h)} \left[K_i(x) - \mathbb{E}\left[K_i(x)\right]\right].$$

We need to evaluate the variance term $S_n^2(x)$,

$$S_n^2(x) = \sum_{i=1}^n \sum_{j=1}^n cov|\Delta_i(x), \Delta_j(x)| =: S_n^{2*}(x) + nVar[\Delta_1(x)],$$

where

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

Next, we evaluate the asymptotic behavior of $S_n^{2*}(x)$.

Following Masry (2005), we define the sets

$$L_1 = \{(i,j) \quad \text{such that} \quad 1 \leq |i-j| \leq m_n\},$$

and

$$L_2 = \{(i,j) \quad \text{such that} \quad m_n + 1 \leq |i-j| \leq n-1\},$$

where, $m_n \longrightarrow \infty$, as $n \to \infty$.

Let, E_1 and E_2 , be the sums of the covariances over L_1 and L_2 , respectively. Then

$$S_n^{2*}(x) = \underbrace{\sum_{L_1} |cov(\Delta_i(x), \Delta_j(x))|}_{J_1} + \underbrace{\sum_{L_2} |cov(\Delta_i(x), \Delta_j(x))|}_{J_2}.$$

We stared by J_1,

$$J_1 = \sum_{|i-j| \leq m_n} |cov(\Delta_i(x), \Delta_j(x))|,$$

where

$$cov(\Delta_i(x), \Delta_j(x)) = \frac{1}{\phi_x^2(h)} cov(K_i(x) - \mathbb{E}[K_i(x)], K_j(x) - \mathbb{E}[K_j(x)]).$$

So

$$\begin{aligned} J_1 &\leq \frac{1}{\phi_x^2(h)} \sum_{L_2} |\mathbb{E}[K_i(x)K_j(x)] - \mathbb{E}^2[K_1(x)]| \\ &\leq \tilde{C} n m_n \phi_x(h)^{\frac{1-a}{a}} n^{\frac{-1}{a}}. \end{aligned}$$

For J_2, we have

$$J_2 = \sum_{|i-j| \geq m_n} |cov(\Delta_i(x), \Delta_j(x))|.$$

We use Davydov-Rio's inequality (Rio (2000), p. 87) for mixing processes, to leads, for all $i \neq j$,

$$cov|\Delta_i(x), \Delta_j(x)| \leq 4\alpha(|i-j|).$$

Finally,

$$J_2 \leq C n^2 m_n^{-a}.$$

So

$$\sum_{i \neq j} |cov(\Delta_i(x), \Delta_j(x))| = C n m_n \phi_x(h)^{\frac{1-a}{a}} n^{\frac{-1}{a}} + \tilde{C} n^2 m_n^{-a}.$$

Choosing $m_n = \left(\frac{\phi_x(h)}{n}\right)^{-\frac{1}{a}}$, then,

$$\begin{aligned} J_1 + J_2 &= C n \phi_x(h)^{-\frac{1}{a}} n^{\frac{1}{a}} \phi_x(h)^{\frac{a+1}{a}} n^{\frac{-1}{a}} + \tilde{C} n^2 \left(\left(\frac{\phi_x(x)}{n} \right)^{-\frac{1}{a}} \right)^{-a} \\ &= C n \phi_x^{-1}(h) + \frac{\tilde{C} n}{\phi_x(h)}. \end{aligned}$$

Putting $C = 2C_2/n$ and $\tilde{C} = 4C_1\epsilon/n\phi_x^2(h)$, we conclude that

$$S_n^2(x) \leq \frac{2C_2}{\phi_x(h)}\left(1 + \frac{4\epsilon C_1/\phi_x(h)}{2C_2/\phi_x(h)}\right).$$

Then, applying Fuk-Nagaev's inequality, (see Ferraty and Vieu (2006), p. 237, Proposition A.11), we can get

$$\mathbb{P}\{|\upsilon_j(x) - \mathbb{E}[\upsilon_j(x)]| > \epsilon\} = \mathbb{P}(|\frac{1}{n}\sum_{i=1}^{n} Z_i| > \epsilon)$$
$$= \leq C\left\{exp\left(-\frac{\epsilon^2 n\phi_x(h)}{2C_2\left(1 + \frac{2\epsilon}{C'C_1}\right)}\right) + nr^{-1}\left(\frac{r}{\epsilon}\right)\right\}.$$

Proof of Part (b).

Let's denote by $\rho = \rho_n > 0$ a numeric sequence such that, $\rho_n \to 0$. Consider a covering of $S_{\mathcal{F}}$ by balls of radius ρ , i.e., $S_{\mathcal{F}} \subset \bigcup_{k=1}^{l} B(x_k, \rho)$, where $l = N_\rho(S_{\mathcal{F}})$.

Let, $D_n(x) = \upsilon_j(x) - \mathbb{E}\upsilon_j(x)$, where $\upsilon_j(x) = (1/n)\sum_{i=1}^{n} W_i^j K_i(x)/\phi_x(h)$, and for all $x \in x \in B(x_k, \rho)$, consider

$$k(x) = \arg \min_{k \in 1,2 .... N_\rho(S_{\mathcal{F}})} d(x - x_k).$$

Then,

$$D_n(x) = \tilde{D}_n(x) + D_n(x_k).$$

Therefore,

$$\sup_{x \in S_{\mathcal{F}}} |\upsilon_j(x) - \mathbb{E}\left[\upsilon_j(x)\right]| \leq \max_{1 \leq k \leq l} |D_n(x_k)| + \max_{1 \leq k \leq l} \sup_{x \in S_{\mathcal{F}} \cap B(x_k,\rho)} |\tilde{D}_n(x)|,$$

it entails that

$$\mathbb{P}\left(\sup_{x \in S_{\mathcal{F}}} |\upsilon_j(x) - \mathbb{E}\left[\upsilon_j(x)\right]| > \epsilon\right) \leq \lambda_n + \delta_n,$$

$$\text{where} \quad \delta_n = \mathbb{P}\left(\max_{1 \leq k \leq l} |D_n(x_k)| > \frac{\epsilon}{2}\right) \quad \text{and} \quad \lambda_n = \mathbb{P}\left(\max_{1 \leq k \leq l} \sup_{x \in S_{\mathcal{F}} \cap B(x_k,\rho)} |\tilde{D}_n(x)| > \frac{\epsilon}{2}\right).$$

Using the obtained result in part a), leads to

$$\lambda_n \leq \sum_{k=1}^{l} \mathbb{P}\left(|D_n(x_k)| > \frac{\epsilon}{2}\right) \leq l \sup_{x \in S_{\mathcal{F}}} \mathbb{P}\left(|D_n(x)| > \frac{\epsilon}{2}\right)$$

$$\leq l\left\{\exp\left(-\frac{\epsilon^2 n\phi_x(h)}{8C_2\left(1 + C_3\frac{\epsilon}{2}\right)}\right) + nr^{-1}\left(\frac{r}{\epsilon}\right)\right\},$$

where $C_3 = 2/C'\|K\|_\infty$ and $C_2 = C'\|K\|_\infty^2$.

Then, the proof of the part $b)$ can be obtained by using the following inequality

$$\delta_n \leq \sum_{k=1}^{l} \mathbb{P}\left(\sup_{x\in S_{\mathcal{F}}\cap B(x_k,\rho)} |\tilde{D}_n(x)| > \frac{\epsilon}{2}\right) \leq l \max_{1\leq k\leq l} \mathbb{P}\left(\sup_{x\in S_{\mathcal{F}}\cap B(x_k,\rho)} |\tilde{D}_n(x)| > \frac{\epsilon}{2}\right).$$

We consider two cases $K(1) = 0$ and $K(1) > 0$.

We begin by $K(1) = 0$, where K is Lipschitz of order one in $[0, 1]$. We have that

$$|\tilde{D}_n(x)| = \left|\frac{1}{n\phi_x(h)}\sum_{i=1}^{n}\left(W_i^j K_i(x) - W_i^j K_i(x_k)\right) - \mathbb{E}\left[\frac{1}{n\phi_x(h)}\sum_{i=1}^{n}\left(W_i^j K_i(x) - W_i^j K_i(x_k)\right)\right]\right|$$
$$\leq \tfrac{1}{n\phi_x(h)} \textstyle\sum_{i=1}^{n} \left\{|K_i(x) - K_i(x_k)| + \mathbb{E}|K_i(x) - K_i(x_k)|\right\},$$

when

$$\frac{1}{\phi_x(h)}\sum_{i=1}^{n}|K_i(x_k) - K_i(x_k)| = \frac{1}{\phi_x(h)}\sum_{i=1}^{n}|K_i(x_k) - K_i(x_k)|\,\mathbb{I}_{B(x,h)\cup B(x_k,h)}(X_i).$$

Thus, we conclude,

$$\sup_{x\in S_{\mathcal{F}}\cap B(x_k,\rho)} |\tilde{D}_n(x)| \leq \frac{C\rho}{h\phi_x(h)} \sup_{x\in S_{\mathcal{F}}}\left\{\frac{1}{n}\sum_{i=1}^{n}\mathbb{I}_{B(x_k,h)\cup B(x_k,h)}(X_i) + \mathbb{E}(\mathbb{I}_{B(x_k,h)\cup B(x_k,h)}(X_1))\right\}$$
$$\leq \frac{C\rho}{h\phi_x(h)} \sup_{x\in S_{\mathcal{F}}}\left\{\frac{1}{n}\sum_{i=1}^{n}\mathbb{I}_{B(x_k,h+\rho)}(X_i) + \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}(\mathbb{I}_{B(x_k,h+\rho)}(X_i))\right\}$$
$$\leq \frac{C\rho}{h\phi_x(h)} \sup_{x\in S_{\mathcal{F}}}\left\{\frac{1}{n}\sum_{i=1}^{n}\left(\mathbb{I}_{B(x_k,h+\rho)}(X_i) + \mathbb{E}\mathbb{I}_{B(x_k,h+\rho)}(X_1)\right)\right\}. \tag{16}$$

We denote by $Z_i = \frac{\rho}{h\phi_x(h)}\mathbb{I}_{B(x_k,h+\rho)}(X_i)$, and suppose that

$$\phi_x(h+\rho) \leq \phi_x(h) + C\rho \leq 2\phi_x(h). \tag{17}$$

Therefore,

$$\frac{\rho}{h\phi_x(h)}\left\{\frac{1}{n}\sum_{i=1}^{n}\mathbb{I}_{B(x_k,h+\rho)}(X_i) + \mathbb{E}(\mathbb{I}_{B(x_k,h+\rho)}(X_i))\right\} \leq \frac{1}{n}\sum_{i=1}^{n}|Z_i - \mathbb{E}[Z_i]|.$$

We can deduce that

$$\mathbb{P}\left(\sup_{x\in S_{\mathcal{F}}\cap B(x_k,\rho)} |\tilde{D}_n(x)| > \frac{\epsilon}{2}\right) \leq \mathbb{P}\left(\frac{1}{n}\sum_{i=1}^{n}|Z_i - \mathbb{E}[Z_i]| > \frac{\epsilon}{4C}\right).$$

Moreover, we apply Fuck-Nagaev's exponential inequality, (Proposition A.11 in Ferraty and Vieu (2006)), to get

$$\mathbb{P}\left(\sup_{x\in S_{\mathcal{F}}\cap B(x_k,\rho)} |\tilde{D}_n(x)| > \frac{\epsilon}{2}\right) \leq C(A_1(x) + A_2(x)), \tag{18}$$

where

$$A_1(x) = \left(1 + \frac{\epsilon^2}{rS_n^2}\right)^{\frac{-r}{2}} \quad \text{and} \quad A_2(x) = nr^{-1}\left(\frac{r}{\epsilon}\right)^{a+1}. \tag{19}$$

Firstly, we must calculate the term

$$S_n^2(x) = \sum_{i=1}^{n}\sum_{j=1}^{n} |cov(\Lambda_i(x), \Lambda_j(x))| = S_n^{2*}(x) + nVar(\Lambda_1(x)),$$

where

$$S_n^{2*} = \sum_{i \neq j} |cov(\Lambda_i(x), \Lambda_j(x))|,$$

and $\Lambda_i = \mathbb{I}_{B(x_k,h+\rho)}(X_i) + \mathbb{E}\mathbb{I}_{B(x_k,h+\rho)}(X_1)$.

Next, we evaluate the asymptotic behavior of $S_n^{2*}(x)$.

Following the decomposition used in Masry (2005), we define the sets

$$F_1 = \{(i,j) \quad \text{such that} \quad 1 \leq |i-j| \leq \nu_n\},$$

and

$$F_2 = \{(i,j) \quad \text{such that} \quad \nu_n + 1 \leq |i-j| \leq n-1\},$$

where $\nu_n \longrightarrow \infty$, as $n \to \infty$. Let, Γ_1 and Γ_2, be the sums of covariances over F_1 and F_2, respectively.

Hence,

$$S_n^{2*}(x) = \underbrace{\sum_{F_1} |cov(\Lambda_i(x), \Lambda_j(x))|}_{\Gamma_1} + \underbrace{\sum_{F_2} |cov(\Lambda_i(x), \Lambda_j(x))|}_{\Gamma_2}.$$

We stared by evaluate the term Γ_1. Note that

$$\begin{aligned} cov(\Lambda_i(x), \Lambda_j(x)) &= cov(\mathbb{I}_{B(x_k,h+\rho)}(X_i) + \mathbb{E}\mathbb{I}_{B(x_k,h+\rho)}(X_j), \mathbb{I}_{B(x_k,h+\rho)}(X_j) + \mathbb{E}\mathbb{I}_{B(x_k,h+\rho)}(X_j)) \\ &= \mathbb{E}[\mathbb{I}_{B(x_k,h+\rho)}(X_i)\mathbb{I}_{B(x_k,h+\rho)}(X_j)] + \mathbb{E}[\mathbb{I}_{B(x_k,h+\rho)}(X_i)]\mathbb{E}[\mathbb{I}_{B(x_k,h+\rho)}(X_j)]. \end{aligned}$$

Thus,

$$\begin{aligned} |cov(\Lambda_i(x), \Lambda_j(x))| &\leq C_1\mathbb{E}\left(\mathbb{I}_{B(x_k,h+\rho)\times B(x_k,h+\rho)}(X_i, X_j)\right) + C_1\mathbb{E}[\mathbb{I}_{B(x_k,h+\rho)}(X_i)]\mathbb{E}[\mathbb{I}_{B(x_k,h+\rho)}(X_j)] \\ &\leq C_1\mathbb{P}\left((X_i, X_j) \in B(x_k, h+\rho) \times B(x_k, h+\rho)\right) \\ &\quad + C_1\mathbb{P}\left(X_i \in B(x_k, h+\rho)\right)\mathbb{P}\left(X_j \in B(x_k, h+\rho)\right). \end{aligned}$$

Under the assumptions $(H2)$, $(H4)$ and $(H10-i)$, we have

$$\begin{aligned} \Gamma_1 &\leq Cn\nu_n\phi_x(h+\rho)\left(\left(\frac{\phi_x(h+\rho)}{n}\right)^{\frac{1}{a}} + \phi_x(h+\rho)\right) \\ &\leq Cn\nu_n\phi_x(h+\rho)^{\frac{a+1}{a}}n^{-\frac{1}{a}}. \end{aligned}$$

For F_2, we can write

$$\Gamma_2 = \sum_{|i-j|\geq \nu_n} |cov(\Lambda_i(x), \Lambda_j(x))|,$$

by Davydov-Rio's inequality (Rio (2000), p. 87) for mixing processes, we have, for all $i \neq j$,

$$cov|\Lambda_i(x), \Lambda_j(x))| \leq 4\alpha(|i-j|).$$

Then,

$$\Gamma_2 = C \sum_{|i-j|\geq \nu_n} \alpha(|i-j|),$$

and by α-mixing condition $(H10, i)$, we have

$$\sum_{\Gamma_2} cov|\Lambda_i(x), \Lambda_j(x))| \leq C_1 n^2 \alpha(\nu_n) \leq C_1 n^2 \nu_n^{-a}.$$

We put, $\nu_n = \left(\frac{\phi_x(h+\rho)}{n}\right)^{\frac{-1}{a}}$, we obtain

$$\Gamma_1 + \Gamma_2 \leq \bar{C} n \phi_x(h+\rho)^{\frac{a+1}{a}} n^{\frac{1}{a}} n^{\frac{-1}{a}} + \acute{C} n^2 \left(\left(\frac{\phi_x(h+\rho)}{n}\right)^{\frac{-1}{a}}\right)^{-a},$$

where $\bar{C} = C\dfrac{\rho^2}{h^2 n \phi_x(h)^2}$ and $\acute{C} = \dfrac{\rho}{nh}\epsilon$.

When we replace ν_n, and we use (17), we can deduce that

$$\Gamma_1 + \Gamma_2 \leq \frac{\rho^2 \phi_x(h+\rho)}{h^2 \phi_x(h)^2} + \frac{C\epsilon\rho\phi_x(h+\rho)}{h\phi_x(h)^2} \leq C\left(2\frac{\rho^2}{h^2\phi_x(h)} + 2\epsilon\frac{\rho}{h\phi_x(h)}\right).$$

Finally,

$$S_n^2(x) = O\left(2C\left(\frac{\rho^2}{h^2\phi_x(h)} + \epsilon\frac{\rho}{h\phi_x(h)}\right)\right).$$

We use (19) to conclude that

$$A_1(x) \leq \left(1 + \frac{\epsilon^2}{r2Cn\left(2\frac{\rho^2}{h^2\phi_x(h)} + 2\epsilon\frac{\rho}{h\phi_x(h)}\right)}\right)^{\frac{-r}{2}} \leq \exp\left(-\frac{\epsilon^2}{2C\left(2\frac{\rho^2}{h^2\phi_x(h)} + 2\epsilon\frac{\rho}{h\phi_x(h)}\right)}\right).$$

We put $r = (\log n)^2$. Thus,

$$A_2(x) = n(\log n)^{-2}\left(\frac{(\log n)^2}{\epsilon}\right)^{a+1}.$$

Moreover, we use $A_1(x)$ and $A_2(x)$, to get

$$\mathbb{P}\left(\sup_{x\in S_{\mathcal{F}}\cap B(x_k,\rho)} |\tilde{D}_n(x)| > \frac{\epsilon}{2}\right) \leq \mathbb{P}\left(\frac{1}{n}\sum_{i=1}^{n} |Z_i - \mathbb{E}[Z_i]| > \frac{\epsilon}{4C}\right)$$

$$\leq exp\left(-\epsilon^2 n\phi_x(h)\frac{1}{32C^3\left(2\frac{\rho^2}{h^2} + \frac{\epsilon}{2C}\frac{\rho}{h}\right)}\right) + n(\log n)^{-2}\left(\frac{4C(\log n)^2}{n\epsilon}\right)^{a+1}. \quad (20)$$

Then, using the fact that, $\frac{\rho}{h} \to 0$, then, if $n \geq n_0$, we have $\frac{\rho}{h} < \left(\frac{1}{4}\right)\left(\frac{1}{2 \parallel K \parallel_\infty}\right)$, and $\frac{2\rho^2}{h^2} < \frac{1}{4}$.

Thus,

$$32C^3\left(2\frac{\rho^2}{h^2} + \frac{\epsilon n}{2C}\frac{\rho}{h}\right) + n(\log n)^{-2}\left(\frac{4C(\log n)^2}{n\epsilon}\right)^{a+1} \leq$$

$$8\left(1 + \epsilon\frac{1}{\parallel K \parallel_\infty}\right) + n(\log n)^{-2}\left(\frac{4C(\log n)^2}{n\epsilon}\right)^{a+1}.$$

Then, the case $K(1) = 0$ is achieved.

The second case is $K(1) > 0$. Let's define

$$\tilde{\tilde{D}}_n(x) = \frac{1}{n}\sum_{i=1}^{n}\left(W_j^i\frac{K_i(x)}{\phi_x(h)} - W_j^i\frac{K_i(x_k)}{\phi_x(h)}\right).$$

Then, $|\tilde{D}_n(x)| \leq |\tilde{\tilde{D}}_n(x)| + \mathbb{E}|\tilde{\tilde{D}}_n(x)|$. As in Lemma 8 of Ferraty et al. (2010), if $x \in S_{\mathcal{F}} \cap B(x_k, \rho)$, we have

$$|\tilde{\tilde{D}}_n(x)| \leq \frac{1}{n\phi_x(h)}\left\{\sum_{i=1}^{n}|K_i(x) - K_i(x_k)|\mathbb{I}_{B(x_k,\rho)\cap B(x,\rho)}(X_i) + \sum_{i=1}^{n}|K_i(x)|\mathbb{I}_{B(x_k,\rho)^c\cap B(x,\rho)}(X_i)\right.$$

$$\left. + \sum_{i=1}^{n}|K_i(x)|\mathbb{I}_{B(x_k,\rho)\cap B(x,\rho)^c}(X_i)\right\}$$

$$\leq C\frac{\rho}{h\phi_x(h)}\frac{1}{n}\sum_{i=1}^{n}\mathbb{I}_{B(x_k,\rho)\cap B(x,\rho)}(X_i) + \parallel K \parallel_\infty \frac{1}{\phi_x(h)}\frac{1}{n}\sum_{i=1}^{n}\mathbb{I}_{B(x_k,\rho)^c\cap B(x,\rho)}(X_i)$$

$$+ \parallel K \parallel_\infty \frac{1}{\phi_x(h)}\frac{1}{n}\sum_{i=1}^{n}\mathbb{I}_{B(x_k,\rho)\cap B(x,\rho)^c}(X_i)$$

$$\leq C\frac{\rho}{h\phi_x(h)}\frac{1}{n}\sum_{i=1}^{n}\mathbb{I}_{B(x_k,\rho)}(X_i) + \parallel K \parallel_\infty \frac{1}{\phi_x(h)}\frac{1}{n}\sum_{i=1}^{n}\mathbb{I}_{B(x_k,\rho)^c\cap B(x_k,h+\rho)}(X_i)$$

$$+ \parallel K \parallel_\infty \frac{1}{\phi_x(h)}\frac{1}{n}\sum_{i=1}^{n}\mathbb{I}_{B(x_k,\rho)\cap B(x_{k,h-\rho})^c}(X_i) \leq \frac{1}{n}\sum_{i=1}^{n}Z_i + Z_{n,1} + Z_{n,2},$$

where Z_i is defined in the first case, and $Z_{n,j} = \frac{1}{n}\sum_{i=1}^{n}W_{i,j}$ where

$$W_{i,1} = \parallel K \parallel_\infty \frac{2}{\phi_x(h)}\mathbb{I}_{B(x_k,\rho)^c\cap B(x_k,h+\rho)}(X_i) \quad \text{and} \quad W_{i,2} = \|K\|_\infty\frac{2}{\phi_x(h)}\mathbb{I}_{B(x_k,\rho)\cap B(x_k,h-\rho)^c}(X_i).$$

Therefore, $W_{i,j} \leq 2/|K\|_\infty/\phi_x(h)$, and let's consider

$$\Lambda_{i,1} = \parallel K \parallel_\infty \frac{2}{\phi_x(h)}\left[\mathbb{I}_{B(x_k,\rho)^c\cap B(x_k,h+\rho)}(X_i) - \mathbb{E}[\mathbb{I}_{B(x_k,\rho)^c\cap B(x_k,h+\rho)}(X_i)]\right],$$

and

$$\Lambda_{i,2} =\| K \|_\infty \frac{2}{\phi_x(h)} \left[\mathbb{I}_{B(x_k,\rho)\cap B(x_k,h-\rho)^c}(X_i) - \mathbb{E}[\mathbb{I}_{B(x_k,\rho)\cap B(x_k,h-\rho)^c}(X_i)]\right].$$

Thus, for $k = 1, 2$, we have

$$S^2_{n_{i,k}}(x) = \sum_{i=1}^{n}\sum_{j=1}^{n} |cov(\Lambda_{i,k}(x), \Lambda_{j,k}(x))| = S^{2*}_{n_{i,k}}(x) + nVar(\Lambda_1(x)) \le S^2_n(x),$$

with

$$S^{2*}_{n_{i,k}} = \sum_{i\neq j} |cov(\Lambda_{i,k}(x), \Lambda_{j,k}(x))|.$$

Then, we can get

$$\sup_{x\in S_{\mathcal{F}}\cap B(x_k,\rho)} \tilde{\tilde{D}}_n(x) \le \frac{C}{n}\sum_{i=1}^{n} Z_i + Z_{n,1} + Z_{n,2}.$$

Notice that,

$$\mathbb{P}\left(\sup_{x\in S_{\mathcal{F}}\cap B(x_k,\rho)} |\tilde{D}_n(x)| > \frac{\epsilon}{2}\right) \le \mathbb{P}\left(\sup_{x\in S_{\mathcal{F}}\cap B(x_k,\rho)} |\tilde{\tilde{D}}_n(x)| > \frac{\epsilon}{4}\right)$$

$$\le \mathbb{P}\left(\frac{C}{n}\sum_{i=1}^{n} Z_i + Z_{n,1} + Z_{n,2} > \frac{\epsilon}{4}\right).$$

Using the result obtained in the case $K(1) = 0$, we deduce that

$$\mathbb{P}\left(\frac{1}{n}\sum_{i=1}^{n} Z_i > \frac{\epsilon}{8C}\right) \le exp\left(-\epsilon^2 n\phi_x(h)\frac{1}{128\left(1+\epsilon\frac{1}{16\|K\|_\infty}\right)}\right) + n(\log n)^{-2}\left(\frac{8C(\log n)^2}{n\epsilon}\right)^{a+1}.$$

Then, for concluding the proof we use a similar inequality for $\mathbb{P}(Z_{n,j} > \frac{\epsilon}{16})$, for $j = 1, 2$.

We can find the bound of $\mathbb{P}\left(\frac{1}{n}\sum_{i=1}^{n} |W_{i,j} - \mathbb{E}[W_{i,j}]| > \frac{\epsilon}{32}\right)$, for $j = 1, 2$, by using the Proposition A.11 in Ferraty and Vieu (2006). If $C_1 = 4 \| K \|^2_\infty$ and $C_2 = 4 \| K \|_\infty$, its follows that

$$\mathbb{P}\left(\frac{1}{n}\sum_{i=1}^{n} |W_{i,j} - \mathbb{E}[W_{i,j}]| > \frac{\epsilon}{32}\right) \le 2exp\left(-\epsilon^2 n\phi_x(h)\frac{1}{2048\left(C_1\frac{\rho}{\phi_x(h)} + \epsilon C_2\right)}\right)$$

$$+n(\log n)^{-2}\left(\frac{4C(\log n)^2}{n\epsilon}\right)^{a+1}.$$

Proof of Part c).
The proof is similar to the one given in part b), and then omitted. ∎

Lemma 3.2.

Let, $\tilde{v}_j(x)$ be defined in (8) for $j = 0, 1$. Under $(H1)$ and $(H4)$, if $h_n \to 0$, we have that a)

$$\sup_{y\in\mathbb{R}}\sup_{x\in S_{\mathcal{F}}}\left|\mathbb{E}\left[\tilde{\upsilon}_1(x)\right]-F(y/X=x)\mathbb{E}\left[\tilde{\upsilon}_0(x)\right]\right|\to 0.$$

b) If the assumption $H8$ holds also, we deduce that

$$\sup_{y\in\mathbb{R}}\sup_{x\in S_{\mathcal{F}}}\left|\mathbb{E}\left[\tilde{\upsilon}_1(x)\right]-F(y/X=x)\mathbb{E}\left[\tilde{\upsilon}_0(x)\right]\right|=O(h^{\eta_1}).$$

Proof:

Proof of Part b).

Let, $Z(x)$ be a random variable where, $Z(x)=d(x,X)/h$ and $P_{Z(x)}$ his probability function. By the assumption $(H8-ii)$, and the boundness of K, we get

$$\delta_1(x,y)=C\int_{S_{\mathcal{F}}^{\epsilon}}d^{\gamma_1}(x,u)K\left(\frac{d(x,u)}{h}\right)dP_X(u)\leq Dd^{\gamma_1}\int v^{\gamma_1}K(v)dP_{Z(x)}(v)$$

$$\leq Dd^{\gamma_1}\int_0^1 K(v)dP_{Z(x)}(v)\leq D_1 d^{\gamma_1}\phi_x(h).$$

Then, for all $y\in\mathbb{R}$ and $x\in S_{\mathcal{F}}$

$$\frac{C_1}{\phi_x(h)}\delta_1(x,y)\leq D_1 d^{\gamma_1}\phi_x(h), \tag{21}$$

which conclude the proof.

The part $a)$ is direct consequence of the result of part $b)$. ■

Let, $F_n(y/X=x)$ be a sequence of conditional distribution function verifying

$$\sup_{y\in\mathbb{R}}\sup_{x\in S_{\mathcal{F}}}\left|F_n(y/X=x)-F(y/X=x)\right|\to 0. \tag{22}$$

Then, if F verifies the assumption $(H5)$, there exist a positive constants $T_1\leq T_2$, such that $s_n(x)=MAD_C(F_n(./X=x))$ satisfies

$$T_1\leq s_n(x)\leq T_2, \forall x\in S_{\mathcal{F}},\ \text{and} n>n_0.$$

The proof of this result is analogous to Lemma 3.1 in Boente and Fraiman (1991). Using Lemma 3.1 part b), when $\rho_n=\frac{\log n}{n}$, we obtain, for any $\epsilon>0$,

$$\mathbb{P}\left(\sup_{x\in S_{\mathcal{F}}}\left|\upsilon_j(x)-\mathbb{E}\left[\upsilon_j(x)\right]\right|>\epsilon\right)\leq\exp\left(\Gamma_{S_{\mathcal{F}}}(\rho_n)-\frac{\epsilon^2 n\phi_x(h)}{a_1(1+\epsilon a_2)}\right)+Cnr^{-1}\left(\frac{r}{\epsilon}\right)$$

$$\leq\exp\left\{-n\phi_x(h)\left(\frac{\epsilon^2}{a_1(1+\epsilon a_2)}-\frac{\Gamma_{S_{\mathcal{F}}}(\rho_n)}{n\phi_x(h)}\right)\right\}+Cnr^{-1}\left(\frac{r}{\epsilon}\right).$$

On the other hand, $H6\ ii)$ implies that, for $n>n_0$,

$$\frac{\Gamma_{S_{\mathcal{F}}}(\rho_n)}{n\phi_x(h)}<\frac{1}{2}\frac{\epsilon^2}{a_1(1+\epsilon a_2)}.$$

If we take $c = 2a_1$, we obtain

$$\mathbb{P}\left(\sup_{x\in S_{\mathcal{F}}} |\upsilon_j(x) - \mathbb{E}\left[\upsilon_j(x)\right]|\right) \le \exp\left\{-n\phi_x(h)\frac{\epsilon^2}{c(1+\epsilon a_2)}\right\} + Cnr^{-1}\left(\frac{r}{\epsilon}\right).$$

Since, $|\tilde{\upsilon}_j(x) - \mathbb{E}\left[\tilde{\upsilon}_j(x)\right]| \le \tilde{C}|\upsilon_j(x) - \mathbb{E}\left[\upsilon_j(x)\right]|$, thus,

$$\mathbb{P}\left(\sup_{x\in S_{\mathcal{F}}} |\tilde{\upsilon}_j(x) - \mathbb{E}\left[\tilde{\upsilon}_j(x)\right]| > \tilde{C}\epsilon\right) \le \exp\left\{-n\phi_x(h)\frac{\epsilon^2}{c(1+\epsilon a_2)}\right\} + Cnr^{-1}\left(\frac{r}{\epsilon}\right).$$

Then, let $\mathcal{A} = \{\inf_{x\in S_{\mathcal{F}}} \tilde{\upsilon}_0(x) \le \frac{1}{2}\}$ and remark that $\mathbb{E}\left[\tilde{\upsilon}_0(x)\right] = 1$, we can deduce that

$$\mathbb{P}(\mathcal{A}) \le \mathbb{P}\left(\sup_{x\in S_{\mathcal{F}}} |\tilde{\upsilon}_j(x) - \mathbb{E}\left[\tilde{\upsilon}_j(x)\right]| > \frac{1}{2}\right) \le \exp(-An\phi_x(h)),$$

where $A^{-1} = 2\tilde{C}c(2\tilde{C} + a_2)$.

Therefore, using that $\dfrac{n\phi_x(h)}{\log(n)} \to \infty$, then $\sup_{x\in S_{\mathcal{F}}} |\tilde{\upsilon}_j(x) - \mathbb{E}\left[\tilde{\upsilon}_j(x)\right]| \overset{a.co}{\to} 0$, for $j = 0, 1$ and

$$\mathbb{P}\left(\inf_{x\in S_{\mathcal{F}}} \tilde{\upsilon}_0(x) \le \frac{1}{2}\right) < \exp(-An\phi_x(h)). \tag{23}$$

This results give the proof of (14), and combined with 3.1.a) and (10), we can deduce that, $\forall y \in \mathbb{R}$,

$$\sup_{x\in S_{\mathcal{F}}} |\hat{F}(y|X = x) - F(y|X = x)| \overset{a.s}{\to} 0. \tag{24}$$

For each, $q \in \mathbb{Q}$, define $\mathcal{N} = \{w \in \Omega : \sup_{x\in S_{\mathcal{F}}} |\widehat{F}(q/X = x) - F(q/X = x)| \nrightarrow 0\}$ and $\mathcal{N} = \cup_{q\in\mathbb{Q}}\mathcal{N}(q)$.

Then, (24) implies that $\mathbb{P}(\mathcal{N}) = 0$. Let $w \in \Omega,\ w \in \mathcal{N}$, then,

$$\sup_{x\in S_{\mathcal{F}}} |\widehat{F}(q/X = x) - F(q/X = x)| \to 0, \text{ for all } q \in \mathbb{Q}.$$

Given $\epsilon > 0$, by the assumption $H5\ iii)$, there exist $a, b \in \mathbb{Q}$, such that, $F(b/X = x) > 1 - \epsilon$ and $F(a/X = x) < \epsilon$, $\forall x \in S_{\mathcal{F}}$.

Moreover, the equicontinuity condition of F, given in the hypothesis $(H5 - iv)$, entails that there exist $a = y_1 < y_2 < \cdots < y_l = b; y_i \in \mathbb{Q}$, such that, $|F(y/X = x) - F(y_i/X = x)| < \epsilon$, for $x \in S_{\mathcal{F}}$ and for any y.

Let, $n_0 \in \mathbb{N}$, such that for, $n > n_0$,

$$\max_{1\le i\le l} \sup_{x\in S_{\mathcal{F}}} |\widehat{F}(y_i/X = x) - F(y/X = x) < \epsilon.$$

Then, it is easy to see that

$$\max_{y\in\mathbb{R}} \sup_{x\in S_{\mathcal{F}}} |\widehat{F}(y/X = x) - F(y/X = x) < 2\epsilon,\ \forall n \ge n_0,$$

which is the claimed result. ∎

3.2. Uniform strong convergence rates

Theorems 3.2 give almost complete convergence rates for the estimators of the empirical conditional distribution and for the local M-estimators of the regression function.

Theorem 3.2.

Let, $S_{\mathcal{F}} \subset \mathcal{F}$ be a compact set. Assume that $(H1)$, $(H4)$, $(H5\ iii)iv)$, $(H6)$, $(H7)$, $(H8)$, and $(H9)$ holds. If γ_1 defined in $(H8)$, is such that $\gamma_1 < \frac{1}{2}$, furthermore, assume that, $\left(\frac{n}{\log(n)}\right)^{1-\gamma_1} \phi_x(h) \leq C$ holds. Let $\widehat{\theta}(x)$ be a robust estimator such that with probability 1, there exists real constants, $0 < T_1 < T_2$, such that, $T_1 < \widehat{\theta}(x) < T_2$, for all $x \in S_{\mathcal{F}}$ and $n > n_0$.

Then,

$$\sup_{x\in S_{\mathcal{F}}} \sup_{y\in\mathbb{R}} |\widehat{\theta}(x) - \theta(x)| = O_{a.co}\left(h^{\eta} + \sqrt{\frac{\Gamma_{S_{\mathcal{F}}}(\log(n)/n)}{n\phi_x(h)}}\right).$$

Proof:

Suppose that $T_1 \leq \widehat{\theta} \leq T_2$. Thus, for any $\varepsilon \in \mathbb{R}$,

$$\sup_{x\in S_{\mathcal{F}}} |\Psi(x, \theta(x)+\varepsilon, \widehat{s}(x)) - \widehat{\Psi}(x, \theta(x)+\varepsilon, \widehat{s}(x))| \leq \frac{1}{T_1}\|\psi\|_V \sup_{x\in S_{\mathcal{F}}} \sup_{y\in\mathbb{R}} |F(y/X = x) - \widehat{F}(y/X = x)|.$$

So, if $\tilde{\theta}_n = h^{\gamma_1} + \sqrt{\frac{\Gamma_{S_{\mathcal{F}}}(\log(n)/n)}{n\phi_x(h)}}$, and for each $\tau > 0$, we have

$$\sup_{x\in S_{\mathcal{F}}} \sup_{-\tau<\varepsilon<\tau} |\Psi(x, \theta(x) + \varepsilon, \widehat{s}(x)) - \widehat{\Psi}(x, \theta(x) + \varepsilon, \widehat{s}(x))| = O_{a.s}(\tilde{\theta}_n).$$

Then, using the fact that $\sup_{x\in S_{\mathcal{F}}} |\widehat{\theta}(x) - \theta(x)| \overset{a.s}{\to} 0$, we can easily get

$$\sup_{x\in S_{\mathcal{F}}} |\Psi(x, \widehat{\theta}(x), \widehat{s}(x)) - \widehat{\Psi}(x, \widehat{\theta}(x), \widehat{s}(x))| = O_{a.s}(\tilde{\theta}_n). \tag{25}$$

Note that, $\Psi(x, \theta(x), \widehat{s}(x)) - \Psi(x, \widehat{\theta}(x), \widehat{s}(x)) + \Psi(x, \widehat{\theta}(x), \widehat{s}(x)) - \widehat{\Psi}(x, \widehat{\theta}(x), \widehat{s}(x)) = 0$ and using results of (25), we get

$$\sup_{x\in S_{\mathcal{F}}} |\widehat{\Psi}(x, \theta(x), \widehat{s}(x)) - \widehat{\Psi}(x, \widehat{\theta}(x), \widehat{s}(x))| = O_{a.s}(\tilde{\theta}_n). \tag{26}$$

Denote by $\Psi'(x, t, \sigma) = \partial\Psi(x, u, \sigma)/\partial u|_{u=t}$. The Mean Value Theorem leads to

$$\widehat{\Psi}(x, \theta(x), \widehat{s}(x)) - \widehat{\Psi}(x, \widehat{\theta}(x), \widehat{s}(x)) = (\widehat{\theta}(x) - \theta(x))[\Psi(x, \xi_n, \widehat{s}(x))]), \tag{27}$$

where $\xi_n(x) \in [\theta(x), \widehat{\theta}(x)]$.

Thanks to assumption $(H3 - ii)$, we get

$$\inf_{-\tau<\varepsilon<\tau} \inf_{x\in S_{\mathcal{F}}} -\frac{\partial}{\partial u}\Psi(x, \theta(x), \widehat{s}(x))|_{u=a} > c_0 > 0. \tag{28}$$

Let $\mathcal{N}$ be the set where (26), (28) and $\sup_{x\in S_{\mathcal{F}}} |\widehat{\theta}(x) - \theta(x)| \longrightarrow 0$ hold, then, $\mathbb{P}(\mathcal{N}) = 0$. Since, for $w \in \mathcal{N}$ and $\tau > 0$, $|\xi_n - \theta(x)| \leq \tau$, $\sup_{x\in S_{\mathcal{F}}} |\Psi(x, \theta(x), \widehat{s}(x)) - \Psi(x, \widehat{\theta}(x), \widehat{s}(x))| = (\tilde{\theta}_n)$.

Therefore, we get that

$$\inf_{x\in S_{\mathcal{F}}} -\Psi'(x,\xi_n,\widehat{s}(x)) > c_0 > 0.$$

The combination of this result together with (26),(27) and (28) achieves the proof of Theorem 3.2. ∎

Theorem 3.3.

Let $S_{\mathcal{F}} \cup \mathcal{F}$ be a compact set. Assume that $(H1),(H4),(H5),(H6),(H7)$ and $(H9)$ holds. Moreover, assume that σ and θ are Lipschitz function of order γ_1 and γ_2 respectively. Then, if $\gamma = \min(\gamma_1,\gamma_2)$, such that $\gamma > 1$, and assume that

$$h\left(\frac{n}{\log(n)}\right)^{1-\gamma} \leq C_\gamma \text{ for all } n \geq 1 \tag{29}$$

or

$$\phi_x(h)\left(\frac{n}{\log(n)}\right)^{1-\gamma} \leq C_\gamma \text{ for all } n \geq 1 \;\; for\ all\ n \geq 1 \tag{30}$$

is fulfilled.

Let, $\widehat{s}(x)$ be a robust scale estimator with probability 1, and $\theta(x)$ be the unique solution of $\Psi(x,a,s(x)) = 0$. Suppose that exists real constants, $0 < T_1 < T_2$, such that $T_1 < \widehat{s}(x) < T_2$ for all $x \in S_{\mathcal{F}}$, and $n > n_0$.

Then, if $\widehat{\theta}(x)$ is a solution of (3), such that $\sup_{x\in S_{\mathcal{F}}} |\widehat{\theta}(x) - \theta(x)| \overset{a.so}{\to} 0$, we have

$$sup_{x\in S_{\mathcal{F}}} |\widehat{\theta}(x) - \theta(x)| = O_{a.co}\left(h^\gamma + \sqrt{\frac{\Gamma_{S_{\mathcal{F}}}(\log(n)/n)}{n\phi_x(h)}}\right).$$

Proof (Theorem 3.3):

Our goal is to show,

$$A_n = \sup_{x\in S_{\mathcal{F}}} \sup_{|\varepsilon|<\tau} \sup_{T_1<\sigma<T_2} |\widehat{\Psi}(x,\theta(x)+\varepsilon,\sigma) - \Psi(x,\theta(x)+\varepsilon,\sigma)| = O_{a.co}(\tilde{\theta}_n).$$

Let denote

$$W_{i,\varepsilon,\sigma}(x) = \psi\left(\frac{Y_i - \theta(x) - \varepsilon}{\sigma}\right), \qquad \tilde{v}_1(x,\varepsilon,\sigma) = \frac{1}{n}\sum_{i=1}^n W_{i,\varepsilon,\sigma}(x)\frac{K_i(x)}{\mathbb{E}K_1(x)},$$

and

$$v_1(x,\varepsilon,\sigma) = \frac{1}{n}\sum_{i=1}^n W_{i,\varepsilon,\sigma}(x)\frac{K_i(x)}{\phi_x(h)}.$$

Since $\widehat{\Psi}(x,\theta(x)+\varepsilon,\sigma) = \tilde{v}_1(x,\varepsilon,\sigma)/\tilde{v}_0(x)$, where $\tilde{v}_0(x)$ is defined in (10), so

$$|\widehat{\Psi}(x,\theta(x)+\varepsilon,\sigma) - \Psi(x,\theta(x)+\varepsilon,\sigma)| \leq \frac{1}{\inf_{x\in S_{\mathcal{F}}} \tilde{v}_0(x)}\left[\sup_{x\in S_{\mathcal{F}}} |\tilde{v}_0(x) - \mathbb{E}\tilde{v}_0(x)|\right.$$

$$+ \sup_{x\in S_{\mathcal{F}}}\sup_{|\varepsilon|<\tau}\sup_{T_1<\sigma<T_2}|\tilde{\upsilon}_1(x,\varepsilon,\sigma)-\mathbb{E}\tilde{\upsilon}_1(x,\varepsilon,\sigma)|$$

$$\left. + \sup_{x\in S_{\mathcal{F}}}\sup_{|\varepsilon|<\tau}\sup_{T_1<\sigma<T_2}|\mathbb{E}\tilde{\upsilon}_1(x,\varepsilon,\sigma)-\Psi(x,\theta(x)+\varepsilon,\sigma)\mathbb{E}\tilde{\upsilon}_0(x)|\right]. \quad (31)$$

By the assumption $(H2-ii)$, and for some constant $C>0$, if $d(X_1,x)<h_n<1$, then

$$\sup_{x\in S_{\mathcal{F}}}\sup_{|\varepsilon|<\tau}\sup_{T_1<\sigma<T_2}|\Psi(x_1,\theta(x)+\varepsilon,\sigma)-\Psi(x,\theta(x)+\varepsilon,\sigma)|\leq C[d(X_1,x)^{\gamma_1}+d(X_1,x)^{\gamma_2}]$$

$$\leq Cd(X_1,x)^{\gamma},$$

where $\gamma=\min\{\gamma_1,\gamma_2\}$.

Since $K_1(x)\leq \mathbb{I}_{B(x,h)}(X_1)$, we have, for n large enough,

$$|\mathbb{E}\tilde{\upsilon}_1(x,\varepsilon,\sigma)-\Psi(x,\theta(x)+\varepsilon,\sigma)\mathbb{E}\tilde{\upsilon}_0(x)|\leq \mathbb{E}|\Psi(X_1,\theta(x)+\varepsilon,\sigma)-\Psi(x,\theta(x)+\varepsilon,\sigma)|\frac{K_1(x)}{\mathbb{E}K_1(x)}$$

$$\leq Ch^{\gamma}.$$

Therefore, if $\tilde{\theta}_n=\theta_n+h^{\gamma}$, then,

$$B_{2n}=\sup_{x\in S_{\mathcal{F}}}\sup_{|\varepsilon|<\tau}\sup_{T_1<\sigma<T_2}|\mathbb{E}\tilde{\upsilon}_1(x,\varepsilon,\sigma)-\Psi(x,\theta(x)+\varepsilon,\sigma)\mathbb{E}\tilde{\upsilon}_0(x)|\leq C\tilde{\theta}_n.$$

Let $\mathcal{A}_n=\inf_{x\in S_{\mathcal{F}}}\tilde{\upsilon}_0(x)<\frac{1}{2}$ and $\epsilon_0>C$. Using (31), we conclude that

$$\mathbb{P}(A_n>4\epsilon_0)\leq \mathbb{P}(\mathcal{A}_n)+\mathbb{P}(\sup_{x\in S_{\mathcal{F}}}|\tilde{\upsilon}_0(x)-\mathbb{E}\tilde{\upsilon}_0(x)|>\epsilon_0\theta_n)+\mathbb{P}(\tilde{B}_{1n}>\epsilon_0\theta_n),$$

where

$$\tilde{B}_{1n}=\sup_{x\in S_{\mathcal{F}}}\sup_{|\varepsilon|<\tau}\sup_{T_1<\sigma<T_2}|\tilde{\upsilon}_1(x,\varepsilon,\sigma)-\mathbb{E}\tilde{\upsilon}_1(x,\varepsilon,\sigma)|.$$

By Lemma 3.1 part b), with $\rho_n=\log(n)/n$, and using (6), we get

$$\sum_{n\geq 1}\mathbb{P}(\mathcal{A}_n<\infty).$$

When Lemma 3.1 part c), permit to obtain

$$\mathbb{P}(\sup_{x\in S_{\mathcal{F}}}|\tilde{\upsilon}_0(x)-\mathbb{E}\tilde{\upsilon}_0(x)|>\epsilon_0\theta_n))<\infty,\ \text{for } \epsilon_0>c.$$

So, it suffices to show that

$$\sum_{n\geq 1}\mathbb{P}(\tilde{B}_{1n}>\epsilon_0\theta_n)<\infty.$$

Note that $\mathbb{P}(\tilde{B}_{1n}>\epsilon_0\theta_n)\leq \mathbb{P}(B_{1n}>\epsilon_1\theta_n)$, where $\epsilon_1=\epsilon_0 C$, and

$$B_{1n}=\sup_{x\in S_{\mathcal{F}}}\sup_{|\varepsilon|<\tau}\sup_{T_1<\sigma<T_2}|\upsilon_1(x,\varepsilon,\sigma)-\mathbb{E}\upsilon_1(x,\varepsilon,\sigma)|.$$

Therefore, we need to prove that, for some c_1 and for some $\epsilon_1>c_1$, $\sum_{n\geq 1}\mathbb{P}(B_{1n}>\epsilon_1\theta_n)<\infty$.

Let $\rho_n = \log(n)/n$ and $x_1, \cdots, x_l$, such that, $S_{\mathcal{F}} \cup \subset_{j=1}^{l} B(x_j, \rho_n)$, where $l = N_{\rho_n}(S_{\mathcal{F}})$.

Then, if

$$S_n(x, \varepsilon, \sigma) = \upsilon_1(x, \varepsilon, \sigma) - \mathbb{E}\upsilon_1(x, \varepsilon, \sigma) \text{ and } \tilde{S}_{n,k}(x, \varepsilon, \sigma) = S_n(x, \varepsilon, \sigma) - S_n(x_k, \varepsilon, \sigma),$$

we have that

$$S_n(x, \varepsilon, \sigma) = S_n(x_k, \varepsilon, \sigma) + \tilde{S}_{n,k}(x, \varepsilon, \sigma),$$

and

$$\begin{aligned}\mathbb{P}(B_{1n} > \epsilon_1\theta_n) = & \quad \mathbb{P}(\sup_{x\in S_{\mathcal{F}}} \sup_{|\varepsilon|<\tau} \sup_{T_1<\sigma<T_2} S_n(x, \varepsilon, \sigma) > \epsilon_1\theta_n) \\ \leq & \quad \mathbb{P}(\sup_{|\varepsilon|<\tau} \sup_{T_1<\sigma<T_2} \max_{1\leq k\leq l} |S_n = (x_k, \varepsilon, \sigma)| > \tfrac{\epsilon_1}{2}\theta_n) \\ & + \mathbb{P}(\sup_{T_1<\sigma<T_2} \sup_{|\varepsilon|<\tau} \max_{1\leq k\leq l}) \sup_{x\in B(x_k,\rho_n)\cap S_{\mathcal{F}}} |\tilde{S}_{n,k}(x, \varepsilon, \sigma)| > \tfrac{\epsilon_1}{2}\theta_n).\end{aligned}$$

Note that,

$$|\tilde{S}_{n,k}(x, \varepsilon, \sigma)| \leq |\upsilon_1(x, \varepsilon, \sigma) - \upsilon_1(x_k, \varepsilon, \sigma)| + |\mathbb{E}\upsilon_1(x, \varepsilon, \sigma) - \mathbb{E}\upsilon_1(x_k, \varepsilon, \sigma)|.$$

Then, by Lipschitz condition of ψ, we have

$$|W_{i,\varepsilon,\sigma}(x_1) - W_{i,\varepsilon,\sigma}(x_2)| \leq c_\theta \|\psi'\|_\infty d(x_1, x_2)^{\gamma_1}/A,$$

for all ε and $T_1 < \sigma < T_2$, where c_θ stands for the Lipschitz constant of θ.

Therefore, if $d(x_k, x) \leq \rho_n < 1$, we have

$$|\upsilon_1(x, \varepsilon, \sigma) - \upsilon_1(x_k, \varepsilon, \sigma)| \leq \|\psi\|_\infty \frac{1}{n\phi_x(h)} \sum_{i=1}^{n} |K_i(x) - K_i(x_k)| + \frac{c_\theta}{T_1} \|\psi'\|_\infty \frac{\rho_n^n}{n\phi_x(h)} \sum_{i=1}^{n} |K_i(x_k)|.$$

Denote by, $\tilde{T}_1 = c_\theta \|K\|_\infty \|\psi'\|_\infty / T_1$, then

$$\max_{1\leq k\leq l} \sup_{x\in B(x_k,\rho_n)\cap S_{\mathcal{F}}} \sup_{T_1<\sigma<T_2} \sup_{|\varepsilon|<\tau} |\upsilon_1(x, \varepsilon, \sigma) - \upsilon_1(x_k, \varepsilon, \sigma)| \leq \tilde{T}\upsilon_1 \max_{1\leq k\leq l} \frac{\rho_n^n}{n\phi_x(h)} \sum_{i=1}^{n} \mathbb{I}_{B(x_k,\rho)}(X_i)$$

$$+ \|\psi\|_\infty \max_{1\leq k\leq l} \sup_{x\in B(x_k,\rho_n)\cap S_{\mathcal{F}}} \frac{1}{n\phi_x(h)} \sum_{i=1}^{n} |K_i(x) - K_i(x_k)|. \quad (32)$$

We consider two situation which that $\gamma \geq 1$ or $\gamma \leq 1$.

i) If $\gamma \geq 1$, we have that $\rho_n^n < \rho_n$, so the bound (32), leads to

$$\max_{1\leq k\leq l} \sup_{x\in B(x_k,\rho_n)\cap S_{\mathcal{F}}} \sup_{T_1<\sigma<T_2} \sup_{|\varepsilon|<\tau} |\upsilon_1(x, \varepsilon, \sigma) - \upsilon_1(x_k, \varepsilon, \sigma)| \leq \tilde{T}_1 \max_{1\leq k\leq l} \frac{\rho_n}{n\phi_x(h)} \sum_{i=1}^{n} \mathbb{I}_{B(x_k,\rho)}(X_i)$$

$$+ \|\psi\|_\infty \max_{1\leq k\leq l} \sup_{x\in B(x_k,\rho_n)\cap S_{\mathcal{F}}} \frac{1}{n\phi_x(h)} \sum_{i=1}^{n} |K_i(x) - K_i(x_k)|,$$

since we have that

$$\sum_{i\geq 1} \mathbb{P}(\sup_{T_1<\sigma<T_2} \sup_{|\varepsilon|<\tau} \max_{1\leq k\leq l} \sup_{x\in B(x_k,\rho_n)\cap S_{\mathcal{F}}} |\tilde{S}_{n,k}(x, \varepsilon, \sigma)| > \frac{\epsilon_1}{2}\theta_n) < \infty. \quad (33)$$

ii) If $\gamma < 1$ and (29) holds, taking $C_\gamma \rho_n / h < 1$, we conclude that

$$\max_{1\le k\le l} \sup_{x\in B(x_k,\rho_n)\cap S_{\mathcal{F}}} \sup_{T_1<\sigma<T_2} \sup_{|\varepsilon|<\tau} |\upsilon_1(x,\varepsilon,\sigma) - \upsilon_1(x_k,\varepsilon,\sigma)| \le \tilde{T}_1 C_\gamma \max_{1\le k\le l} \frac{\rho_n}{nh\phi_x(h)} \sum_{i=1}^{n} \mathbb{I}_{B(x_k,\rho)}(X_i)$$

$$+\|\psi\|_\infty \max_{1\le k\le l} \sup_{x\in B(x_k,\rho_n)\cap S_{\mathcal{F}}} \frac{1}{n\phi_x(h)} \sum_{i=1}^{n} |K_i(x) - K_i(x_k)|,$$

then (33) can be deduced from the proof of the Lemma 3.1 part c).

iii) If $\gamma < 1$ and (30) holds, the sequence $\theta_n^{-1}\rho_n^n$ is bounded. Then, defining

$$Z_i = \frac{\rho_n^n}{n\phi_x(h)} \mathbb{I}_{B(x_k,\rho)}(X_i).$$

There exist $\tilde{c}$ such that for $\epsilon_1 > \tilde{c}$, we can write

$$\sum_{i\ge 1} \mathbb{P}\left(\max_{1\le k\le l} \frac{\rho_n^n}{n\phi_x(h)} \sum_{i=1}^{n} \mathbb{I}_{B(x_k,\rho)}(X_i) > \epsilon_1 \frac{\theta_n}{4} \right) < \infty,$$

which concludes the proof of (33).

Now, we need to find a bound for χ_n, where

$$\chi_n = \mathbb{P}\left(\sup_{T_1<\sigma<T_2} \sup_{|\varepsilon|<\tau} \max_{1\le k\le l} |S_n(x_k,\varepsilon,\sigma)| > \epsilon_1 \frac{\theta_n}{2} \right).$$

Let, $|\varepsilon| \le \tau$, a finite covering by intervals $I_j^\varepsilon = [\varepsilon_j, \varepsilon_{j+1}]$, when $|\varepsilon_{j+1} - \varepsilon_j| < \alpha_n$ with $\alpha_n = 1/\sqrt{n}$.

Then, we have at most $M_{1,\alpha_n} = 2\tau\alpha_n^{-1}$ intervals. On the other hand, considered a covering of $T_1 < \sigma < T_2$ by intervals $I_j^\sigma = [\sigma_j, \sigma_{j+1}]$, such that, $|\sigma_{j+1} - \sigma_j| < \alpha_n$, where we have $M_{2,\alpha_n} = 2(T_1 - T_2)\alpha_n^{-1}$ intervals. Thus,

$$\chi_n \le \chi_{n,1} + \chi_{n,2},$$

with

$$\chi_{n,1} = \mathbb{P}\left(\max_{1\le s\le M_{2,\alpha_n}} \max_{1\le j\le M_{1,\alpha_n}} \max_{1\le k\le l} |S_n(x_k,\varepsilon_j,\sigma_s)| > \epsilon_1 \frac{\theta_n}{2} \right), \tag{34}$$

and

$$\chi_{n,2} = \mathbb{P}\left(\max_{1\le j\le M_{1,\alpha_n}} \max_{1\le k\le l} \sup_{T_1<\sigma<T_2} \sup_{|\varepsilon|<\tau} |S_n(x_k,\varepsilon_j,\sigma_s) - S_n(x_k,\varepsilon,\sigma)| > \epsilon_1 \frac{\theta_n}{2} \right). \tag{35}$$

Since $|W_{i,\varepsilon,\sigma}| \le \|\psi\|_\infty$, similar arguments to those used in Lemma 3.1, we obtain

$$\sup_{T_1<\sigma<T_2} \sup_{|\varepsilon|<\tau} \max_{1\le k\le l} \mathbb{P}\left(|S_n(x_k,\varepsilon,\sigma)| > \theta_n \frac{\epsilon_1}{4} \right) \le \exp\left\{ (1 - C\epsilon_1^2)\Gamma_{S_{\mathcal{F}}} \left(\frac{\log(n)}{n} \right) \right\}.$$

Moreover, if $C_2 = 2\tau(T_1 - T_2)$, we have that

$$\mathbb{P}\left(\max_{1\le j\le M_{1,\alpha_n}} \max_{1\le s\le M_{2,\alpha_n}} \max_{1\le k\le l} |S_n(x_k,\varepsilon_j,\sigma_s)| > \theta_n \frac{\epsilon_1}{4} \right)$$

$$\leq 2M_{1,\alpha_n} M_{2,\alpha_n} M_{\rho_n}(S_{\mathcal{F}}) \exp\left\{-C\epsilon_1^2 \Gamma_{S_{\mathcal{F}}}\left(\frac{\log(n)}{n}\right)\right\}$$

$$\leq 2C_2\alpha_n^{-2} \exp\left\{(1 - C\epsilon_1^2)\Gamma_{S_{\mathcal{F}}}\left(\frac{\log(n)}{n}\right)\right\}.$$

Taking ϵ_1 such that $(1 - C\epsilon_1^2) > 1 - \beta$, and $\alpha_n = 1/\sqrt{n}$, then

$$\chi_{n,1} \leq 2C_2 n \exp\left\{(1 - \beta)\Gamma_{S_{\mathcal{F}}}\left(\frac{\log(n)}{n}\right)\right\}.$$

The result $\sum_{n\geq 1} \chi_{n,1} < \infty$ follows as a direct consequence of (6), for n a large enough.

Concerning

$$\sum_{n\geq 1} \chi_{n,2} < \infty, \tag{36}$$

note that, if $a \in I_j^{(a)}$, for all $T_1 < \sigma < T_2$, we have

$$|\upsilon_1(x_k, \varepsilon_j, \sigma_s) - \upsilon_1(x_k, \varepsilon, \sigma)| \leq \frac{\|\psi'\|_\infty}{T_1} \frac{\alpha_n}{n\phi_x(h)} \sum_{i=1}^{n} K_i(x_k).$$

Then, if $\sigma \in I_s^{(\sigma)}$, using that $\zeta(t) = t\psi'(t)$ is bounded, we obtain

$$|\upsilon_1(x_k, \varepsilon_j, \sigma_s) - \upsilon_1(x_k, \varepsilon, \sigma)| \leq \frac{\|\zeta\|_\infty}{T_1} \frac{\alpha_n}{n\phi_x(h)} \sum_{i=1}^{n} K_i(x_k).$$

Consequently,

$$|\upsilon_1(x_k, \varepsilon_j, \sigma_s) - \upsilon_1(x_k, \varepsilon, \sigma)| \leq C_3 \frac{\alpha_n}{n\phi_x(h)} \sum_{i=1}^{n} K_i(x_k),$$

where $C_3 = (\|\psi'\|_\infty + \|\zeta\|_\infty)/T_1$.

Noting that $(1/n\phi_x(h)) \sum_{i=1}^{n} K_i(x_k) = \upsilon_0(x_k)$, we get

$$|S_n(x_k, \varepsilon, \sigma) - S_n(x_k, \varepsilon, \sigma)| \leq |\upsilon_1(x_k, \varepsilon_j, \sigma_s) - \upsilon_1(x_k, \varepsilon_j, \sigma_s)| + \mathbb{E}|\upsilon_1(x_k, \varepsilon_j, \sigma_s) - \upsilon_1(x_k, \varepsilon_j, \sigma_s)|$$

$$\leq C_3 \frac{\alpha_n}{n\phi_x(h)} \sum_{i=1}^{n} K_i(x_k) + C_3 \frac{\alpha_n}{\phi_x(h)} \mathbb{E}K_1(x_k)$$

$$\leq C_3\alpha_n |\upsilon_0(x_k) - \mathbb{E}\upsilon_0(x_k)| + 2C_3 \frac{\alpha_n}{\phi_x(h)} \mathbb{E}K_1(x_k).$$

Using the fact that, $\mathbb{E}K_1(x_k) \leq C'\phi_x(h), \alpha_n = 1/\sqrt{n}$ and $\theta_n^{-1}\alpha_n \to 0$, then $\exists n_0$, such that, for $n > n_0$, we have $\theta_n^{-1}\alpha_n \leq \min(1/C_3, \epsilon_1/(16C'C_3))$.

Therefore,

$$\theta_n^{-1}|S_n(x_k, \varepsilon, \sigma) - S_n(x_k, \varepsilon, \sigma)| \leq C_3\theta_n^{-1}\alpha_n |\upsilon_0(x_k) - \mathbb{E}\upsilon_0(x_k)| + 2C'C_3\theta_n^{-1}\alpha_n$$

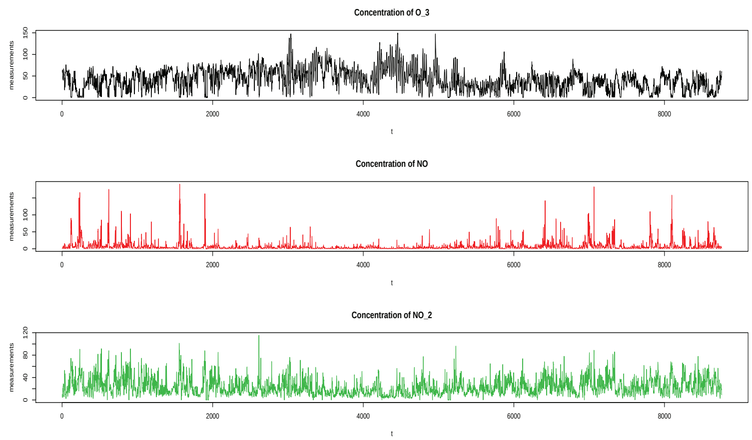

Figure 1. The hourly measurements of O_3,NO and NO_2 concentration for the year 2018

$$\leq |\upsilon_0(x_k) - \mathbb{E}\upsilon_0(x_k)| + \frac{\epsilon_1}{8}.$$

Then, the right hand side of the last inequality does not depend of σ and a, so we can write

$$\chi_{n,2} \leq \max_{1\leq k\leq l} \mathbb{P}\left(|\upsilon_0(x_k) - \mathbb{E}\upsilon_0(x_k)| > \frac{\epsilon_1}{8} \right).$$

So, for all $\epsilon_0 > 0$,

$$\sum_{n\geq 1} \mathbb{P}\left(\sup_{x\in S_{\mathcal{F}}} |\upsilon_0(x_k) - \mathbb{E}\upsilon_0(x_k)| > \epsilon_0 \right) < \infty,$$

this completes the proof of (36) and therefore, the proof of Theorem 3.3. ■

4. Real data applications

The main objective of this part is to evaluate the good behavior of the proposed robust estimator for two real data applications and to show the efficiency of the robust estimator with unknown scale parameter compared to the one with fixed scale parameter.

4.1. Maximum Ozone Concentration

This section, we are interested in forecasting maximum values of ozone concentration. The data consist of hourly measurements of ozone (O_3) concentration together with additional chemical measurements such as NO and NO_2 concentrations $(\mu\mathrm{g/m}^3)$ (see Figure 1) in Leicester University monitoring site during the period from January 1^{st} to the 31^{st} December for the year 2018, (365 days). Data are available on the website `https://uk-air.defra.gov.uk`.

The original time series are:

$$O_{3,t}, NO_t \text{ and } NO_{2,t}, \quad t = 1, \ldots, 8760.$$

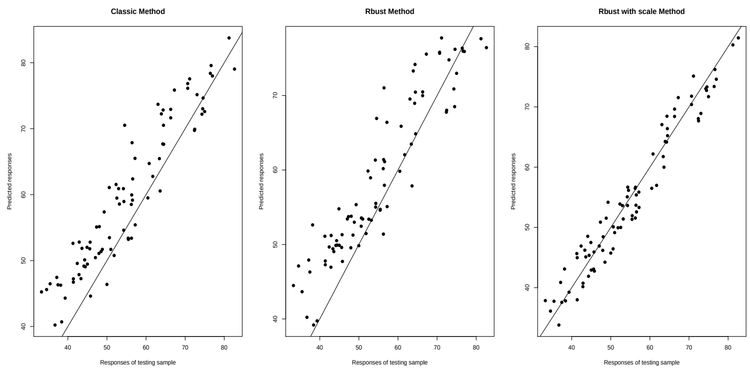

Figure 2. Prediction of the maximum ozone of the last 64 days by classical and robust regression

To fix the ideas, let's present the mathematical formulation of our prediction problem. Indeed, assume that we aim to predict the maximum air pollutant (Ozone O_3) concentration at day i, denoted by Y, using the curve of the daily emission of the gases observed the day before $i-1$. Formally, we assume that the output variable Y and the input variables are $Z = \left(X_{O_{3,i-1}}, X_{NO_{i-1}}, X_{NO_{2,i-1}}\right)$ is linked by the following regression formula,

$$Y_i = r(Z_i) + \epsilon_i, \qquad i = 1, \ldots, 364,$$

is cut into 364 daily curves,

$$Z_i = \left\{(O_{3,24(i-1)+t}, NO_{24(i-1)+t}, NO_{2,24(i-1)+t}), t \in [\,0, 24\,[\,\right\}, \quad i = 1, \ldots, 364.$$

Second, we need to select a suitable semi-metric $d(.,.)$, kernel $K(.)$, smoothing parameter h_{opt} for our estimator. For that purpose, we choose the asymmetrical quadratic kernel defined as $K(u) = \frac{3}{4}\left(\frac{12}{11} - u^2\right)\mathbb{I}_{[0,1]}(u)$. According to the general guidelines provided in Ferraty and Vieu (2006), we suggest to use standard PCA semi-metrics as follows:

$$d^{PCA}(z_i, z_j) = d^{PCA}\left(x_{O_{3,i}}, x_{O_{3,j}}\right) + d^{PCA}\left(x_{NO_i}, x_{NO_j}\right) + d^{PCA}\left(x_{NO_{2,i}}, x_{NO_{2,j}}\right).$$

We choose $h_{opt} = \arg\min_h CV(h)$ for the estimator, where $CV(h) = \sum_{i=1}^{n}\left(Y_i - \widehat{\theta}_{(-i)}(Z_i)\right)^2$, and $\widehat{\theta}_{(-i)}(z)$ is the solution at t of:

$$\sum_{j=1, j\neq i}^{n} w_j(x)\psi\left(\frac{Y_j - t}{\widehat{\sigma}}\right) = 0.$$

Finally, we split our sample of 364 days into a learning sample containing the first 300 days and a testing sample with the last 64 days. Figure 2 shows the results obtained for the ozone prediction derived from the testing sample. The left panel represents the prediction under the classical method (Ferraty and Vieu (2006)). The central group represents the prediction by the robust method without scale parameter introduced by Azzedine et al. (2008), while the right panel shows the forecast under our model (robust with scale parameter).

The error used is the mean of squared error (MSE), expressed by

$$MSE = \frac{1}{64}\sum_{i=301}^{364}\left(Y_i - \widehat{\theta}(Z_i)\right)^2.$$

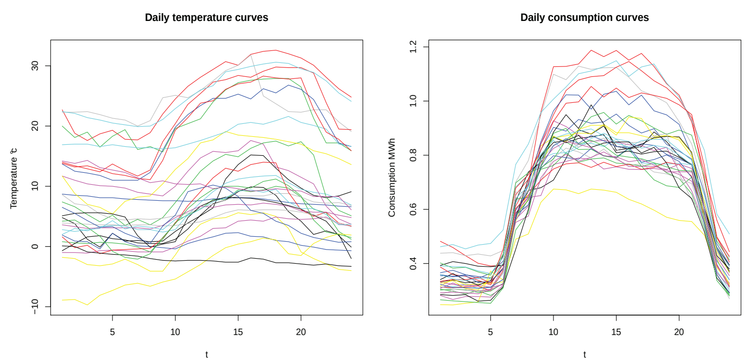

Figure 3. Sample of 30 daily temperature curves and the associated energy consumption curves

This is illustrated by the $MSE = 3.71488$ for the classic method, $MSE = 3.5238$ for the robust with fixed scale parameter, and $MSE = 3.025231$ for our proposed estimator.

Another important point for ensuring a good behavior of our method is to introduce the outliers in this learning sample. We multiply by 100 the response variable of some observations Y. We observe that the robust method gives better results than the classical method in the presence of outliers (MSE for the classic is 67.97584, for the robust is 4.671861 and for the robust with scale parameter 4.15175).

As we can see, the robust with scale parameter method always gives good behavior in the tow cases, with and without outlier variables.

4.2. Peak electricity demand

The evolution of peak electricity demand can be considered as an essential system design metric for grid operators and planners. The peak demand forecasting of aggregated electricity demand has been widely studied in the statistical literature, and several approaches have been proposed to solve this issue (see, for instance, Chikobvu and Sigauke (2012) and Goia et al. (2010) for short-term peak forecasting and Hyndman et al. (2010) for long-term density peak forecasting).

In this subsection, we are interested in the estimation of peak demand at the customer level. For a fixed day d let us denote by $(E_d\,(t_j))_{j=1,\ldots,24}$ the hourly measurements 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 $(T_d\,(t_j))_{j=1,\ldots,24}$ (measured in Celsius degrees). The peak demand observed for the day d is defined as

$$P_d = \max_{j=1,\ldots,24} E_d\,(t_j)\,.$$

Figure 3 provides a sample of 30 curves of hourly temperature measures and the associated electricity consumption curves.

Therefore, our sample is formed as follows $(T_d, P_d)_{d=1,\ldots,366}$, where T_d is the predicted temperature

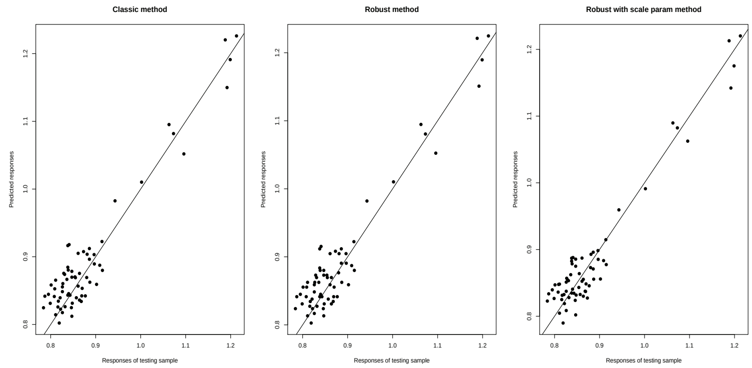

Figure 4. Prediction of the last 66 days by three models

curve for the day d and P_d the peak demand observed for the day d. 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.

The choice of the kernel, the semi-metric and the bandwidth are similar to these used in the maximum ozone application.

Figure 4 shows the results obtained for the peak energy prediction derived from the testing sample for the three models (classical, robust and robust with scale parameter method). The associated MSE are 0.00102, 0.00098 and 0.00072, respectively, for each methods.

5. Conclusion

We provide in this paper the uniform almost complete consistency with rates of the robust regression function in case of unknown scale parameter. These results were obtained under sufficient standard conditions that allows one to explore different structural axes of the subject, such as the functional naturalness of the model and the data as well as the robustness of the regression function and the dependency of the observation. The real data applications (Maximum Ozone Concentration, Peak electricity demand) have also highlighted several attractive features of the robust regression with unknown scale parameter estimator. In terms of mean squared error (MSE) the proposed estimator performs competitively in comparison to existing estimators with know scale parameter.

Based on the experience of this paper on robust regression with unknown scale parameter, we guess that most of the techniques using nonparametric functional kernel smothers could be efficiently extended. For instance, challenging open questions in this sense could concern as well extensions to other forms of nonparametric predictors (like functional local linear ones, functional kNN ones, and many other ones). Extensions to other kinds of prediction models in which a preliminary kernel stage plays a crucial role (this would include many semiparametric regression models like functional single index models, and partial linear models, and many other ones).

Acknowledgment:

The authors would like to thank the Editor-in-Chief of AAM, and the anonymous reviewers for their valuable comments and suggestions which improved substantially the quality of this paper.

REFERENCES

Attouch, M.K., Kaid, Z., and Louhab, H. (2019). Asymptotic normality of a robust kernel estimator of the regression function for functional ergodic data: Case of an unknown scale parameter, C. R. Math. Acad. Sci. Paris, Vol. 357, No. 5, pp. 478-481.

Attouch, M., Laksaci, A. and Ould-Said, E. (2010). Asymptotic normality of a robust estimator of the regression function for functional time series data, J. of the Korean Statist. Soc., Vol. 39, No. 4, pp. 489-500.

Attouch, M., Laksaci, A. and Ould Said, E. (2012). Strong uniform convergence rate of a robust estimator of the regression function for functional and dependent processes, Journal of Japan Statistical Society, Vol. 42, No. 3, pp. 125-143.

Azzedine, N., Laksaci, A. and Ould-Said, E. (2008). On robust nonparametric regression estimation for a functional regressor, Statistics & Probability Letters, Vol. 78, No. 18, pp. 3216-3221.

Benhenni, K., Ferraty, F., Rachdi, M. and Vieu, P. (2007). Local smoothing regression with functional data, Comput. Statist., Vol. 22, No. 3, pp. 353-369.

Boente, G. and Fraiman, R. (1989a). Nonparametric regression estimation, J. Multivariate Anal. Vol. 29, No. 2, pp. 180-198.

Boente, G. and Fraiman, R. (1989b). Robust nonparametric regression estimation for dependent observations, Annals of Statistics, Vol. 17, No. 3, pp. 1242-1256.

Boente, G. and Fraiman, R. (1991). Strong uniform convergence rates for some robust equivariant nonparametric regression estimates for mixing processes, International Statistical Institute, Vol. 59, No. 3, pp. 355-372.

Boente, G. and Vahnovanb, A. (2015). Strong convergence of robust equivariant nonparametric functional regression estimators, Statistic & Probability Letters, Vol. 100, pp. 1-11.

Chikobvu, D. and Sigauke, C. (2012). Regression-SARIMA modelling of daily peak electricity demand in South Africa, Journal of Energy in Southern Africa, Vol. 23, No. 3, pp. 23-30.

Collomb, G. (1982). Prédiction non paramétrique: Etude de l'erreur quadratique du préedictogramme, C. R.Acad. Sci. Paris Sér. I Math, Vol. 294, No. 4, pp. 59-62.

Ferraty, F., Laksaci, A., Tadj, A. and Vieu, Ph. (2010). Rate of uniform consistency for nonparametric estimates with functional variables, Journal of Statistical Planning and Inference, Vol. 140, No. 2, pp. 335-352.

Ferraty, F., Laksaci, A. and Vieu, Ph. (2006). Estimating some characteristics of the conditional distribution in nonparametric functional models, Statistical Inference for Stochastic Processes, Vol. 9, No. 3, pp. 47-76.

Ferraty, F. and Vieu, Ph., (2006). *Nonparametric Functional Data Analysis*, Springer-Verlag Publisher, United States of America.

Gheriballah, A., Laksaci, A. and Sekkal, S. (2013). Nonparametric M-regression for functional ergodic data, Journal Statistics et Probability Letters, Vol. 83, No. 3, pp. 902-908.

Goia, A., May, C. and Fusai, G. (2010). Functional clustering and linear regression for peak load forecasting, International Journal of Forecasting, Vol. 26, No. 4, pp. 700-711.

Hyndman, R.J. and Fan, S. (2010). Density forecasting for long-term peak electricity demand, IEEE Transaction on Power Systems, Vol. 25, No. 2, pp. 1142-1153.

Kolmogorov, A.N. (1956). Asymptotic characteristics of some completely bounded metric spaces, Doklady Akademii Nauk SSSR, Vol. 1, No. 8, pp. 585-589.

Kolmogorov, A.N. and Tikhomirov, V.M. (1959). ϵ-entropy and ϵ-capacity, Uspekhi Mat. Nauk, Vol. 14, pp. 3-86. (English version in Transl. Amer. Math. Soc. Transl. Ser, Vol. 2, No. 1, pp. 277-364 (1961).

Masry, E. (2005). Nonparametric regression estimation for dependent functional data: Asymptotic normality, Stoch. Proc. and their Appl., Vol. 115, No. 1, pp. 155-177.

Rachdi, M., Laksaci, A., Demongeot, J., Abdali, A. and Madani F. (2014). Theoretical and practical aspects of the quadratic error in the local linear estimation of the conditional density for functional data, Comput. Statist Data Anal., Vol. 73, No. 2, pp. 53-68.

Ramsay, J. and Silverman, B. (2002). *Applied Functional Data Analysis: Methods and Case Studies*, Springer.

Ramsay, J. and Silverman, B. (2005). *Functional Data Analysis*, Springer Series in Statistics.

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

Stone, C.J. (2005). Nonparametric M-regression with free knot splines, J. Stat. Plan. Inf., Vol. 130, No. 1-2, pp. 183-206.