



HAL
open science

On the robust nonparametric regression estimation for a functional regressor

Nadjia Azzedine, Ali Laksaci, Elias Ould-Saïd

► To cite this version:

Nadjia Azzedine, Ali Laksaci, Elias Ould-Saïd. On the robust nonparametric regression estimation for a functional regressor. *Statistics and Probability Letters*, 2008, 78 (18), pp.3216-3221. 10.1016/j.spl.2008.06.018 . hal-00510975

HAL Id: hal-00510975

<https://hal.science/hal-00510975>

Submitted on 23 Aug 2010

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Accepted Manuscript

On the robust nonparametric regression estimation for a functional regressor

Nadjia Azzedine, Ali Laksaci, Elias Ould-Saïd

PII: S0167-7152(08)00301-5
DOI: [10.1016/j.spl.2008.06.018](https://doi.org/10.1016/j.spl.2008.06.018)
Reference: STAPRO 5114

To appear in: *Statistics and Probability Letters*

Received date: 8 November 2006
Revised date: 24 January 2008
Accepted date: 2 June 2008

Please cite this article as: Azzedine, N., Laksaci, A., Ould-Saïd, E., On the robust nonparametric regression estimation for a functional regressor. *Statistics and Probability Letters* (2008), doi:10.1016/j.spl.2008.06.018

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.



On the robust nonparametric regression estimation for a functional regressor

NADJIA AZZEDINE, ALI LAKSACI

Département de Mathématiques

Univ. Djillali Liabès

BP 89, 22000 Sidi Bel Abbès, Algeria

e-mail: (azzedine or alilak)@yahoo.fr

ELIAS OULD-SAÏD *

L.M.P.A. J. Liouville

Univ. du Littoral Côte d'Opale

BP 699, 62228 Calais, France

e-mail: ouldsaid@lmpa.univ-littoral.fr

11th January 2008

Abstract

We study a family of robust nonparametric estimators for a regression function based on a kernel method when the regressors are functional random variables. We establish the almost complete convergence rate of these estimators under the probability measure's concentration property on small balls of the functional variable. Simulations are given to show our estimator's behavior and the prediction quality for functional data.

1 Introduction

Regression function estimation is the most important tool for nonparametric prediction problem. The goal of this paper is to study this functional parameter when the explanatory variable is a curve by using a robust approach. The robust method used in this paper is belongs to the class of M-estimates

*corresponding author

introduced by Huber (1964). The literature on this estimation method is quite important when the data are real (see for instance Robinson (1984), Collomb and Härdle (1986), Boente and Fraiman (1989), (1990), for previous results and Laïb and Ould-Saïd (2000), Boente and Rodriguez (2006) for recent advances and references). For the functional case, this literature is relatively limited, indeed, Cadre (2001) studied the estimation of the L_1 -median of a Banach space-valued random variable. Cardot *et al.* (2004) used this robust approach to study the linear regression model on quantiles with explanatory variable taking values in a Hilbert space. They established the L_2 -convergence rate. We refer the reader to Ferraty and Vieu (2006) for the prediction problem in functional nonparametric statistics via the regression function, the conditional mode and the conditional quantiles estimation by the kernel method. The asymptotic normality of these parameters have been obtained by Masry (2005) and Ezzahrioui and Ould-Saïd (2007, 2008) respectively.

Our interest in this paper is to generalize, to infinite dimension, the robust nonparametric estimation of regression function developed by Collomb and Härdle (1986) in the real case. We establish, under suitable conditions, the almost complete (a.co.)¹ convergence rate of the M-estimator with the regression function kernel weights when the observations are independent and identically distributed. This rate is closely related to the concentration property on small balls of the functional variable's probability measure. Thus, by using recent results in the probability theory of small balls, we can clarify our results for some continuous-time stochastic processes.

The paper is organized as follows. We present our model in Section 2. In Section 3 we give some notations, hypotheses and presentat the main results. Section 4 is devoted to some comments. Simulation results appear in the last section. The proofs of the auxiliary results are relagated to the Appendix.

¹We say that the sequence $(U_n)_n$ converges a.co. to zero if and only if $\forall \varepsilon > 0$, $\sum_{n \geq 1} \mathbb{P} \{|U_n| > \varepsilon\} < \infty$. Futhermore, we say that $U_n = o(x_n)$ a.co. if there exists $\varepsilon_0 > 0$ such that $\sum_{n \geq 1} \mathbb{P} \{|U_n| > \varepsilon_0 x_n\} < \infty$.

2 The model

Let (X, Y) be a pair of random variables (rv) in $\mathcal{F} \times \mathbb{R}$, where the space \mathcal{F} is dotted with a semi-metric $d(\cdot, \cdot)$ (this covers the case of normed spaces of possibly infinite dimension). In this framework, X can be a functional random variable. For any x in \mathcal{F} , let ψ_x be a real-valued Borel function satisfying some regularity conditions to be stated below. The nonparametric parameter studied in this paper, denoted by θ_x , is implicitly defined as a zero with respect to (w.r.t.) t of the equation

$$(1) \quad \Psi(t, x) := \mathbb{E}(\psi_x(Y - t) | X = x) = 0.$$

We suppose that, for all $x \in \mathcal{F}$, θ_x exists and is unique (see, for instance, Boente and Fraiman (1989)). The model θ_x , called ψ_x -regression in Laïb and Ould-Saïd (2000) is a generalization of the classical regression function. Indeed, if $\psi_x(t) = t$ we get $\theta_x = \mathbb{E}[Y | X = x]$.

Let $(X_1, Y_1), \dots, (X_n, Y_n)$ be n independent pairs, identically distributed as (X, Y) . We then estimate $\Psi(t, x)$ by

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

where K is a kernel function and $h_K = h_{K,n}$ is a sequence of positive real numbers which decreases to zero as n goes to infinity. A natural estimator of θ_x denoted by $\widehat{\theta}_x$, is a zero w.r.t. t of

$$(2) \quad \widehat{\Psi}(t, x) = 0.$$

Our main goal is to obtain the rate of the almost complete convergence for $\widehat{\theta}_x$. Some simulation has been given to show how to implement our methodology for the functional data and the behavior of our estimator.

3 Main result

In the following, x is a fixed point in \mathcal{F} , \mathcal{N}_x denotes a fixed neighborhood of x , and we introduce the following assumptions:

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

(H2) There exist $C_1 > 0$ and $b > 0$ such that $\forall x_1, x_2 \in N_x, \forall t \in \mathbb{R}$
 $|\Psi(t, x_1) - \Psi(t, x_2)| \leq C_1 d^b(x_1, x_2)$.

(H3) The function ψ_x is strictly monotone, bounded, continuously differentiable, and its derivative is such that, $|\psi'_x(t)| > C_2 > 0, \forall t \in \mathbb{R}$.

(H4) K is a continuous function with support $[0, 1]$ such that $0 < C_3 < K(t) < C_4 < \infty$.

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

Our main result is given in the following theorem.

Theorem 1 *Assume that (H1)-(H5) are satisfied, then $\hat{\theta}_x$ exists and is unique a.s. for all sufficiently large n , and we have*

$$(3) \quad \hat{\theta}_x - \theta_x = O(h_K^b) + O\left(\sqrt{\frac{\log n}{n \phi_x(h_K)}}\right) \quad a.co.$$

Proof. In what follows, we will denote by C some strictly positive generic constant and we put $K_i = K\left(\frac{d(x, X_i)}{h_K}\right)$. Under to (H3), we have

$$\widehat{\Psi}(\hat{\theta}_x, x) = \widehat{\Psi}(\theta_x, x) + (\hat{\theta}_x - \theta_x) \widehat{\Psi}'(\xi_{x,n}, x)$$

for some $\xi_{x,n}$ between $\hat{\theta}_x$ and θ_x . The condition on the derivative of ψ_x in (H3), leads us to write

$$\begin{aligned} \exists C > 0, \forall \epsilon_0 > 0, \quad & \mathbb{P}\left(|\hat{\theta}_x - \theta_x| \geq \epsilon_0 \left(h^b + \sqrt{\frac{\log n}{n \phi_x(h)}}\right)\right) \\ & \leq \mathbb{P}\left(|\widehat{\Psi}(\theta_x, x) - \Psi(\theta_x, x)| \geq C^{-1} \epsilon_0 \left(h^b + \sqrt{\frac{\log n}{n \phi_x(h)}}\right)\right). \end{aligned}$$

Then, (3) is proved as soon as the following result could be checked

$$(4) \quad \widehat{\Psi}(\theta_x, x) - \Psi(\theta_x, x) = O\left(h_K^b + \sqrt{\frac{\log n}{n \phi_x(h_K)}}\right) \quad a.co.$$

The proof of (4) is based on the decomposition

$$(5) \quad \forall t \in \mathbb{R}, \widehat{\Psi}(t, x) - \Psi(t, x) = \frac{1}{\widehat{\Psi}_D(x)} \left[\left(\widehat{\Psi}_N(t, x) - \mathbb{E} \left[\widehat{\Psi}_N(t, x) \right] \right) - \left(\Psi(t, x) - \mathbb{E} \left[\widehat{\Psi}_N(t, x) \right] \right) \right] \\ - \frac{\Psi(t, x)}{\widehat{\Psi}_D(x)} \left[\widehat{\Psi}_D(x) - \mathbb{E} \left[\widehat{\Psi}_D(x) \right] \right]$$

where

$$\widehat{\Psi}_D(x) = \frac{1}{n\mathbb{E}[K_1]} \sum_{i=1}^n K_i, \quad \widehat{\Psi}_N(t, x) = \frac{1}{n\mathbb{E}[K_1]} \sum_{i=1}^n K_i \psi_x(Y_i - t)$$

and with the fact that $\widehat{\Psi}(t, x) = \frac{\Psi_N(t, x)}{\widehat{\Psi}_D(x)}$ and $\mathbb{E} \left[\widehat{\Psi}_D(x) \right] = 1$. Finally, the proof of Theorem 1 is achieved with the following lemmas

Lemma 3.1 *Under Hypotheses (H1), (H4) and (H5), we have,*

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

This lemma gives straightforwardly the following corollary

Corollary 3.2 *Under the hypotheses of Lemma 3.1, we have,*

$$\sum_{n \geq 1} \mathbb{P} \left(\left| \widehat{\Psi}_D(x) \right| \leq 1/2 \right) \leq \sum_{n \geq 1} \mathbb{P} \left(\left| \widehat{\Psi}_D(x) - 1 \right| > 1/2 \right) < \infty.$$

Lemma 3.3 *Under Hypotheses (H1), (H2), (H4) and (H5), we have, for all $t \in \mathbb{R}$*

$$\Psi(t, x) - \mathbb{E} \left[\widehat{\Psi}_N(t, x) \right] = O(h_K^b).$$

Lemma 3.4 *Under Hypotheses (H1) and (H3)-(H5), we have, for all $t \in \mathbb{R}$*

$$\widehat{\Psi}_N(t, x) - \mathbb{E} \left[\widehat{\Psi}_N(t, x) \right] = O \left(\sqrt{\frac{\log n}{n\phi_x(h_K)}} \right) \quad a.co.$$

Lemma 3.5 *Under the hypotheses of Theorem 1, $\widehat{\theta}_x$ exists and is unique a.s. for all sufficiently large n .*

4 Comments

1. *Remarks on the functional variable:*

The concentration hypothesis (H1) is less restrictive than the strict positivity of the explanatory variable's density X which is usually assumed in most of previous works in the finite-dimensional case (see Collomb and Härdle (1986) and Laïb and Ould-Saïd (2000)). Moreover, it is checked for a great class of continuous time processes (see for instance Bogachev (1999) for a Gaussian measure and Li and Shao (2001) for a general Gaussian process).

2. *Remarks on the nonparametric model:*

The functional character of our model is well exploited in this work. Indeed, Hypothesis (H2) is a regularity condition which characterizes the functional space.

3. *Remarks on the robustness properties:*

In this paper, we consider a family of ψ -functions indexed by x , in order to cover most of M-estimates classes (see Collomb and Härdle (1986) for some examples of ψ_x). It is also worth noting that we keep the same conditions on the function ψ_x (assumption (H3)) given by Collomb and Härdle (1986) in the multivariate case. Furthermore, the boundedness assumption on ψ_x is made only for the simplicity of the proof. It can be dropped while using truncation methods as to those used in Laïb and Ould-Saïd (2000).

4. *Remarks on the convergence rate:*

The expression of convergence rate (3) is identical with those of Ferraty and Vieu (2006) and Collomb and Härdle (1986) for the regression model in the functional and the multivariate cases respectively. Thus, by considering the same arguments to those of Ferraty *et al.* (2005), we obtain the almost complete convergence rate $O\left((\log n)^{-b/2}\right)$ for the estimator $\hat{\theta}_x$ for continuous-time stochastic process having a probability measure which is absolutely continuous with respect to the Wiener measure, under suitable bandwidth choice ($h_K \sim \eta(\log n)^{-1/2}$) and for the L^∞ metric. The almost complete convergence rate of the estimator $\hat{\theta}_x$ for the fractional Brownian

motion of parameter δ , ($0 < \delta < 2$), $h_K \sim \eta(\log n)^{-\delta/2}$ and $d(\cdot, \cdot)$ is L^∞ -norm, is of order $O\left((\log n)^{-\delta b/2}\right)$. A convergence rate of order $O\left(n^{\frac{-b}{(2b+p)}}\right)$ is obtained in the finite-dimensional case ($X \in \mathbb{R}^p$) under positivity of the density of X and for $h_K \sim n^{-1/(2b+p)}$.

5 Simulation study

In this section, we examine one example for which the forecast via the estimated conditional median which is a robust method (the medianogram corresponds to $\psi_x(t) = \mathbb{1}_{(t \geq 0)} - 1/2$) is better than that obtained by the kernel method defined as

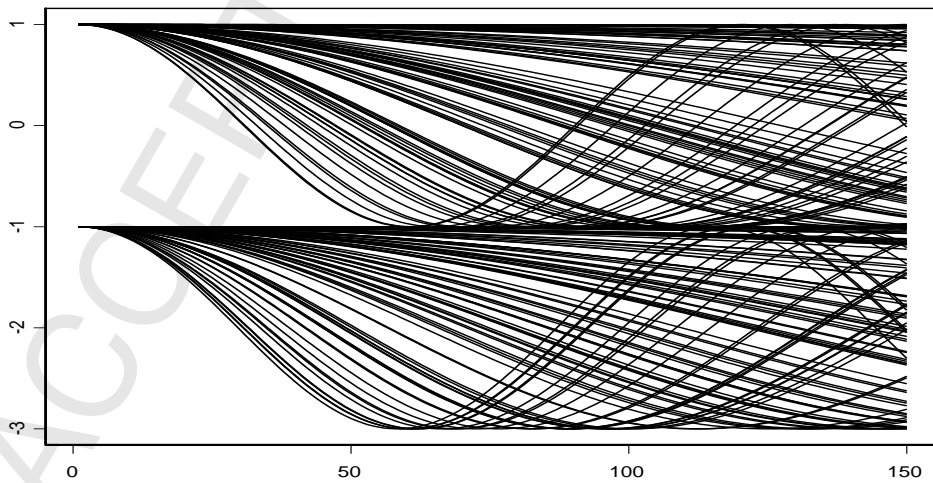
$$\widehat{m}(x) = \inf\{y \text{ such that } \widehat{F}^x(y) \geq 0.5\}$$

where

$$\widehat{F}^x(y) = \frac{\sum_{i=1}^n K(h_K^{-1}d(x, X_i))H(h_H^{-1}(y - Y_i))}{\sum_{i=1}^n K(h_K^{-1}d(x, X_i))}$$

is a kernel estimator of the conditional cumulative distribution and H is a distribution function.

In this example, we consider two diffusion processes on the interval $[0, 1]$, $Z_1(t) = 2 - \cos(\pi t W)$ and $Z_2(t) = \cos(\pi t W)$, ($W \rightsquigarrow N(0, 1)$) and we take $X(t) = AZ_1(t) + (1 - A)Z_2(t)$, where A is a Bernoulli random variable distributed. We carried out the simulation with a 200-sample of the curve X which is represented by the following graphes:



The curves $X_{i=1, \dots, 200}(t)$, $t_{j=1, \dots, 150} \in [0, 1]$

The scalar response variable is defined by $Y = Ar_1(X) + (1 - A)r_2(X)$ where r_1 (resp. r_2) is the nonlinear regression model $r_1(X) = 0.25 \times \left(\int_0^1 X'(t)dt \right)^2 + \epsilon$, with ϵ is $U_{[0,0.5]}$ (resp. $r_2(X)$ is the null function). The selection of the bandwidth h_K (resp. h_H) is an important and basic problem in all kernel smoothing techniques. In this simulation, for both methods, the optimal bandwidths were chosen by the cross-validation method on the k -nearest neighbours. We choose the quadratic kernel:

$$K(x) = \frac{3}{2}(1 - x^2)\mathbb{I}_{[0,1]} \quad K(1) > 0 \quad \text{and} \quad H(t) = \int_{-\infty}^t K(u)du.$$

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. Here, we used the semi-metric defined by the L_2 -distance between the q^{th} derivatives of the curves (for further discussion see Ferraty *et al.* (2005)). This choice is motivated by the regularity of the curves X . In order to compare these two methods we proceed by the following algorithm:

Step 1. Split the data randomly into two subsets:

- $(X_j, Y_j)_{j \in J}$: training sample,
- $(X_i, Y_i)_{i \in I}$: test sample.

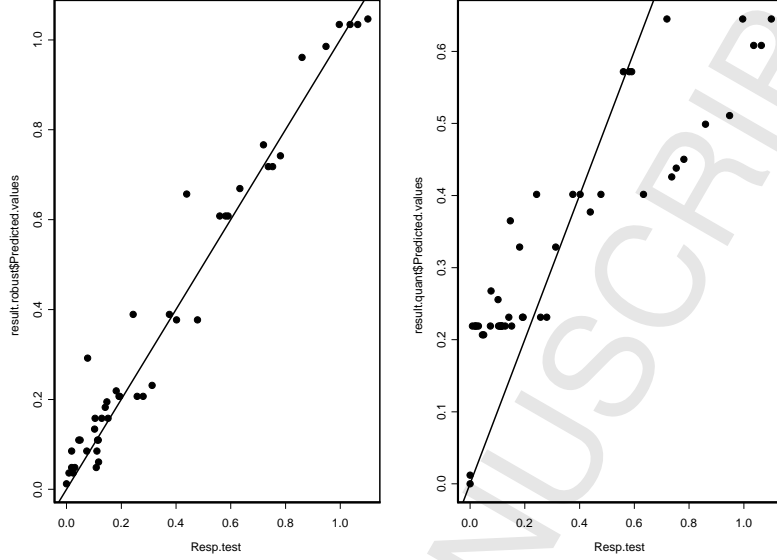
Step 2. Compute the robust estimator $\widehat{\theta}(X_j)$ and the kernel estimator $\widehat{m}(X_j)$, for all $j \in J$ by using the training sample.

Step 3. For each X_i in the test sample, set: $i_* = \arg \min_{j \in J} d(X_i, X_j)$. *Step 4.* For all $i \in I$, take

$$\widehat{Y}_i^{robust} = \widehat{\theta}(X_{i_*}) \quad \text{and} \quad \widehat{Y}_i^{KN} = \widehat{m}(X_{i_*}).$$

Step 5. Present the results by plotting the predicted values versus the true values and compute the sum of squared residuals (SSR).

It is clear that, the number of null response variable has inverse effect in the choice of the smoothing parameter h_H in the estimation of the conditional cumulative distribution function, in particular in $y = 0$. Consequently, the robust method gives better results than the kernel method. This is confirmed by the scatterplots and the test error $SSR_{robust} = 0.002$ and $SSR_{KN} = 0.020$.



Robust method SSR=0.002

KN-method SSR=0.020

6 Appendix

Proof of Lemma 3.1 The proof of this Lemma runs along the lines of Lemma 3.1 in Ferraty *et al.* (2005).

Let $\tilde{\Delta}_i := \frac{K_i}{\mathbb{E}[K_1]}$. From (H1) and (H4) we deduce

$$\left| \tilde{\Delta}_i \right| < C/\phi_x(h_K) \quad \text{and} \quad \mathbb{E} \left[\left| \tilde{\Delta}_i \right|^2 \right] < C'/\phi_x(h_K).$$

So, we apply the Bernstein exponential inequality to get for all $\eta > 0$

$$\mathbb{P} \left(\left| \hat{\Psi}_D(x) - \mathbb{E} \left[\hat{\Psi}_D(x) \right] \right| > \eta \sqrt{\frac{\log n}{n \phi_x(h_K)}} \right) \leq C' n^{-C\eta^2}.$$

Proof of Lemma 3.3

The equidistribution of the couples (X_i, Y_i) and (H4) imply

$$(6) \quad \Psi(t, x) - \mathbb{E} \left[\hat{\Psi}_N(t, x) \right] = \frac{1}{\mathbb{E}[K_1]} \mathbb{E} \left[(K_1 \mathbb{1}_{B(x, h_K)}(X_1)) (\Psi(t, x) - \mathbb{E} [\psi_x(Y_1 - t) | X = X_1]) \right],$$

where $\mathbb{1}$ is indicator function. Conditioning w.r.t. X_1 , the Hölder hypothesis and under (H2), we prove that (H2) allows us to write that

$$K_1 \mathbb{1}_{B(x, h_K)}(X_1) |\Psi(t, X_1) - \Psi(t, x)| \leq C_1 h_K^b,$$

then

$$|\Psi(t, x) - \mathbb{E} [\widehat{\Psi}_N(t, x)]| \leq C_1 h_K^b.$$

Proof of Lemma 3.4

The proof of this result is similar to the proof of the Lemma 3.1. We put

$$\Lambda_i = \frac{\{K_i \psi_x(Y_i - t) - \mathbb{E} [K_1 \psi_x(Y_1 - t)]\}}{\mathbb{E} [K_1]}.$$

Because ψ_x is bounded, we get $|\Lambda_i| \leq C/\phi_x(h_K)$ and $\mathbb{E} [\Lambda_i^2] \leq C'/\phi_x(h_K)$, for all $i \leq n$. As in Lemma 3.1, Bernstein's inequality is used to finish the proof.

Proof of Lemma 3.5

We prove this Lemma by similar arguments to those used in Theorem 1 of Collomb and Härdle (1986). Indeed, for all $\epsilon > 0$, the strict monotony of ψ_x implies

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

The Lemmas 3.1, 3.4, 3.5 and Corollary (3.2) show that

$$\widehat{\Psi}(\theta_x, x) - \Psi(\theta_x, x) = O \left(h_K^b + \sqrt{\frac{\log n}{n \phi_x(h_K)}} \right) \quad a.co.$$

for all fixed real t . So, for sufficiently large n

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

Since ψ_x and K are continuous functions then $\widehat{\Psi}(t, x)$ is a continuous function of t , then there exists a $t_0 = \widehat{\theta}_x \in [\theta_x - \epsilon, \theta_x + \epsilon]$ such that $\widehat{\Psi}(\widehat{\theta}_x, x) = 0$. Finally, the unicity of $\widehat{\theta}_x$ is a direct consequence of the strict monotony of ψ_x and the positivity of K .

References

- [1] Boente, G., Fraiman, R., 1989. Nonparametric regression estimation. *J. Multivariate Anal.* 29, No.2, 180-198.
- [2] Boente, G., Fraiman, R., 1990. Asymptotic distribution of robust estimators for nonparametric models from mixing processes. *Ann. Statist.* 18, No.2, 891-906.
- [3] Bogachev, V.I., 1999. Gaussian measures. *Math surveys and monographs.* 62, Amer. Math. Soc.
- [4] Boente, G., Rodriguez, D., 2006. Robust estimators of high order derivatives of regression function. *Statist. Probab. Lett.* 76, 1335-1344.
- [5] Cadre, B., 2001. Convergent estimators for the L_1 -median of a Banach valued random variable. *Statistics* 35, No.4, 509-521.
- [6] Cardot, H., Crambes, C., Sarda, P., 2004. Spline estimation of conditional quantities for functional covariates. *C. R. Math.* 339, No.2, 141-144.
- [7] Collomb, G., Härdle, W., 1986. Strong uniform convergence rates in robust nonparametric time series analysis and prediction: Kernel regression estimation from dependent observations. *Stoch. Proc. and their Appl.* 23, 77-89.
- [8] Ezzahrioui, M., Ould-Saïd, E., 2007. Asymptotic normality of the kernel estimators of the conditional quantile in the normed space. *Far East J. Theoretical Statist.* (In press).
- [9] Ezzahrioui, M., Ould-Saïd, E., 2008. Asymptotic normality of a nonparametric estimator of the conditional mode function for functional data. *Nonparametric Statist. J.* (In press).
- [10] Ferraty, F., Vieu, P., 2006. *Nonparametric functional data analysis. Theory and Practice.* Springer-Verlag. New-York.
- [11] Ferraty, F., Laksaci, A., Vieu, P., 2005. Estimating some characteristics of the conditional distribution in nonparametric functional models. *Statist. Inf. for Stoch. Proc.* 9, 47-76

- [12] Huber, P.J., 1964. Robust estimation of a location parameter. *Ann. Math. Statist.* 35, 73-101.
- [13] Laïb, N., Ould-Saïd, E., 2000. A robust nonparametric estimation of the autoregression function under an ergodic hypothesis. *Canad. J. Statist.* 28, 817-828.
- [14] Li, W.V., Shao, Q.M., 2001. Gaussian processes: inequalities, small ball probabilities and applications. In *Stochastic processes: Theory and Methods*, Ed. C.R. Rao and D. Shanbhag. *Handbook of Statistics*, 19, North-Holland, Amsterdam.
- [15] Masry, E., 2005. Nonparametric regression estimation for dependent functional data: Asymptotic normality. *Stoch. Proc. and their Appl.* 115, 155-177.
- [16] Robinson, R., 1984. *Robust Nonparametric Autoregression*. *Lecture Notes in Statistics*, Springer-Verlag, New York 26, 247-255.