Smoothed Influence Function : Another View at Robust Nonparametric Regression *

Julien Tamine
Université de la Méditerranée
GREQAM Vieille Charité
2 rue de la Charité
13002 Marseille France
e mail : julientamine@yahoo.fr

July 2001

Abstract

In this work, we introduce a smoothed influence function that constitute a theoretical tool for studying the outliers robustness properties of a large class of nonparametric estimators. With this tool, we first show the nonrobustness of the Nadaraya-Watson estimator of regression. Then we show that the M, the L and the R-estimators of the regression achieve robustness (when estimated by kernel). Our results are illustrated performing Monte-Carlo simulation.

Keywords *robustness, influence function, M-estimator, L-estimator, R-estimator, nonparametric, regression, Von-mises statistical functional, generalized Delta theorem.*

JEL Classification C13, C14, C15

*Acknowledgements: The author received financial support from Greqam, Université de la Méditerranée, Marseille and from the Deutsche Forschungsgemeinschaft via Sonderforschungsbereich 373 "Quantifikation und Simulation Ökonomischer Prozesse" at Humboldt-Universität zu Berlin.

1 Introduction

The robustness properties of parametric estimators have received much attention since the seminal work of Huber (see [12]) and Hampel (see [8]). The problem of robustness of nonparametric estimators may seem, at first view, less important since nonparametric estimators are often considered as a way of robustifying parametric estimators. This thought is justified when the robustness under study concerns departures from the parametric null model since, with nonparametric techniques, less stringent assumptions are made on the underlying data generating process. Nevertheless, when robustness to outliers is concerned, the behavior of nonparametric estimators is also worth studying.

The concept of outlier has often been used rather informally (for a general discussion see Lucas [13]). Nevertheless, some attempts of definition have been given in the past : for instance, Barnett and Lewis (see[3]) define an outlier as an observation (or subset of observations) which appears to be inconsistent with the remainder data set; as underlined by Lucas (see [13]), this is not what we could call a precise definition. A quantitative definition has been given by Davis and Gather (see[7]): if F is the null distribution, for instance, the standard univariate distribution with mean μ and variance σ^2, an observation x is said to be an α ($\in]0;1[$) outlier if $|x-\mu| > \sigma z_{1-\alpha/2}$ where $z_{1-\alpha/2}$ is the $1-\alpha/2$ quantile of the null distribution F. This definition is interesting since it states that outliers need not be generated out of the null data generating process. In this sense, outliers may also affect nonparametric estimators.

In the parametric field, the functional approach of Von-Mises (see [18]) has been extensively used for assessing the robustness properties of estimators through the concept of influence function (see hampel [8]). In comparison, the literature on the links between the robustness properties of nonparametric estimators and functional analysis is much more restricted and only a few references, limited to the special case of nonparametric regression are available (see for instance Boente and Fraiman [4]). This is, in our sense, due to the lack of functional framework for their analysis. Indeed, for a long time, Von-Mises statistical functionals have found no applications in the field of nonparametric estimators so that, in particular, the concept of influence function has completely been occulted.

Recently, Aït Sahalia (see [1]) has shown how a large category of nonparametric kernel estimators can be considered in a functional framework. His functional approach has been used in various fields such as for instance game theory (see Protopopescu [15]) or noisy differential equations (see Vanhems [17]) but it hasn't been used yet as a tool for studying the robustness properties of nonparametric estimators. Our purpose is twofold. First, we aim at filling this gap by defining an analogous of Hampel's influence function for nonparametric estimators : the smoothed influence function. Next, we want to illustrate the usefulness of this smoothed influence function; for this, we will study the M, the L and the R estimators (see Huber [12]) of the nonparametric regression in a functional framework and we will calculate their smoothed influence function. This study will constitute a theoretical basis which will confirm the conclusions of the empirical work that has already been led on the subject (see Härdle [10] for a survey).

Our work is organized as follows : in section 2, we set the functional tools introduced by Aït-Sahalia and we show how the smoothed influence function can be defined in this framework. In section 3, we show how this framework can be useful in the particular case of an estimator implicitly defined by an equation. In section 4, we study the Nadaraya-Watson estimator of the regression and we show its nonrobustness in the sense of the smoothed influence function. In section 5, we study the M, L and R estimators of the regression and we give conditions ensuring their robustness in the sense of the smoothed influence function. In section 6, we discuss the differences between the estimators we analyze in this work an those that have already been proposed in the literature. Finally, we give a Monte-Carlo simulation illustrating the performances of the M, L and R estimators of the regression in the presence of outliers in the sample. Proofs are given in appendix.

2 Functional analysis of nonparametric estimators

2.1 Aït-Sahalia Generalized Delta Theorem

First of all, we're going to define the concept of strong Frechet differentiability in a general setting that we will restrict later.

Definition 1 *let $(E_1, \|\circ\|_1)$ and $(E_2, \|\circ\|_2)$ be Banach spaces and let T be an application from E_1 to E_2. The functional T will be strongly Frechet differentiable at the point $F \in E_1$ with respect to the norm $\|\circ\|_1$ if there exists a linear continuous operator $D_F T : E_1 \to E_2$ such that*

$$T(F+H) = T(F) + D_F T.H + O\left(\|H\|_1^2\right)$$

for all H in a neighborhood of zero.

- One of the appealing features of the strong Frechet differentiability is the validity of the chain rule.
- If E_1 is the product of two (or more spaces), one can define in an analogous way the concept of partial strong Frechet derivative. The existence of the strong Frechet derivative implies the existence of the partial strong Frechet derivatives too.

In this work we will be denote by C^s, the space of s times continuously differentiable functions on R^d, whose derivatives of all order are compactly supported on

R^d. We will restrict ourselves to the particular case where E_1 is a closed subspace of C^s (that we will precise for each study) and $E_2 = R$. More precisely, we will be interested in estimating the quantity $T\left(F^Z\right)$ where $F^Z \in E_1$ is the cumulative distribution of a $d-$dimensional random vector Z and $T : E_1 \to R$ is a real valued functional. Our goal in this section will be to give sufficient conditions for $T\left(F^Z\right)$ to be consistently estimated by $T\left(\hat{F}_n^Z\right)$ where $\hat{F}_n^Z(\circ) = \dfrac{1}{n}\sum_{i=1}^{n} K_I\left(\dfrac{\circ - Z_i}{h}\right)$ is the kernel[1] estimator of F.

We now need to define a norm on the space C^s which will induce a norm on $E_1 \subset C^s$:

Definition 2 *for any function $H \in C^s$, the $m^{th}(m \leq s)$ order uniform Sobolev norm is classicaly defined by*

$$\|H\|_{L(\infty,m)} = sup_{0\leq c\leq m} sup_{|\Delta|=c} sup_{u\in R^d} \left|\partial^{(\Delta)}H(u)\right|$$

where $\Delta = (\Delta_1, ..., \Delta_d) \in N^d$ with $|\Delta| = \Delta_1 + ... + \Delta_d$ and $\partial^{(\Delta)}H(u) = \dfrac{\partial^{|\Delta|}H(u)}{\partial u_1.....\partial u_d}$.

We are now going to precise further which functionals will be considered in this work :

Definition 3 *Let $u \in R^d$ denote an integration variable and $u^{d(l)}$ a $d(l)$ dimensional subvector of u. Let $z^{d(l)}$ denote a fixed point in $R^{d(l)}$,and let $\Delta^l \in N^{d(l)}$ denote an indice set. We suppose that $\alpha_l(F)(u)$ has continuous partial derivatives up to order $\left|\Delta^l\right|$, all belonging to L^1, for all $1 \leq d(l) \leq d$. The functional T will be said to admit the representation (R) if it's strongly Frechet differentiable at the point F^Z for the norm $\|\circ\|_{L(\infty,m)}$ and if $D_{F^Z}T.H$ can be written under the form*

$$D_F\Phi.H = \int \left[A_F(u) + B_F(u)\right] dH(u)$$

where $A_{F^Z}(u) = \sum_{l=1}^{L} \alpha_l(F)(u)\partial^{(\Delta^l)}\delta_{z^{(l)}}(u^{(l)})$ and $B_{F^Z}(u)$ is a cadlag[2] function

- (R) implicitly supposes the function H to be an element of $C^{\left|\Delta^l\right|}$ since

$$\int \alpha_l(F)(u)\partial^{(\Delta^l)}\delta_{z^{(l)}}(u^{(l)})dH(u) = \int \partial^{(\Delta^l)}\left(\alpha_l(F).h\right)(z^{(l)}, u^{(-l)})du^{(-l)}.$$

[1] K_I is an integrated kernel ie a cumulative distribution function. We will precise later conditions on the kernel in order to ensure the consistency of the estimators considered.

[2] right continuous with left hand limits.

- The aim of this characterization is to allow for the identification of terms of different order when plugging $H = F_n - F$. Intuitively speaking, $d(l)$ indicates the dimension of the random vector which density is to be estimated (i.e. the so called curse of dimensionality) and $\left|\Delta^l\right|$ indicates the order of this density's derivative.

Before enunciating Aït-Sahalia's Generalized Delta Theorem (see [1]), we have to state a few assumptions :

(A1) : We suppose $\{Z_1, ..., Z_n\}$ to be a sample of strictly stationary and β mixing realizations of the random vector Z.

(A2)(s) : We suppose K to be an even kernel[3] of order r on R^d which partial derivatives exist up to order $s + d$ and are square summable on R^d.

(A3)(e, m) : we suppose the bandwidth h to satisfy the following conditions for values of m and e which will be specified later : for $n \to \infty$, we have $h \to 0$ in such a way that $n^{\frac{1}{2}} h^e \to 0$ et $n^{\frac{1}{2}} h^{2m} \to \infty$.[4]

Theorem 1 ***Aït-Sahalia's Generalized Delta Theorem :***

We suppose the functional T to be strongly Frechet differentiable at the point F^Z for the norm $\|\circ\|_{L(\infty,m)}$ in such a way that $D_{F^Z}T.H$ admits the representation (Rep). Then, under the assumptions (A1) and (A2)(s) :

(i) if $A_{F^Z}(u) = 0$, under the assumption (A3)(r, m) we have

$$\sqrt{n}\left(T\left(\hat{F}_n^Z\right) - T\left(F^z\right)\right) \Longrightarrow N(0, V_T(F^Z))$$

where

$$V_T(F^Z) = var(B_{F^Z}(Z_i)) + 2\sum_{k=1}^{\infty} cov\left(B_{F^Z}(Z_i), B_F(Z_{i+k})\right)$$

(ii) if $A_{F^Z}(u) \neq 0$, let $\varkappa^* = max\left\{d(l) + 2\left|\Delta^l\right|,\ l = 1....L\right\}$ and let L^* denote the subset of $\{1....L\}$ defined by $L^* = \left\{l \in \{1....L\} \backslash d(l) + 2\left|\Delta^l\right| = \varkappa^*\right\}$. Finally let $d^* = max\left\{d(l) + 2\left|\Delta^l\right|,\ l \in L^*\right\}$ then, under the assumption $A3(r + \dfrac{d^*}{2}, m)$ we have :

$$\sqrt{n} h^{(\varkappa^*)\backslash 2}\left(T\left(\hat{F}_n^Z\right) - T\left(F^Z\right)\right) \Longrightarrow N(0, V_T(F^Z))$$

and

$$V_T(F^Z) = \sum_{l \in L^*} \int \left[\partial^{\Delta^l} K^{(l)}(u^{(l)})\right]^2 du^{(l)} \int \alpha_l(F^Z)(z^{(l)}, u^{(-l)}) f^Z(z^{(l)}, u^{(-l)}) du^{(-l)}$$

[3] In order to simplify the notations, K will always be the product of d univariate kernels and the variables will be supposed to be scaled so that the bandwith chosen is the same in each direction.

[4] These conditions can only be satisfied if $e > 2m$

where f denotes the density of the random vector Z and $K^{(l)}(u^{(l)}) = \int K(u^{(l)}, u^{(-l)})du^{(-l)}$.

- the assumption $A3(r + \frac{d^*}{2}, m)$ ensures, among others things, that the asymptotic bias is equal to zero. It implicitly determines the order of the kernel to be used since it implies $r > m - \frac{d^*}{2}$.

2.2 Smoothed influence function

In this section, we suppose satisfied the assumptions of the generalized delta theorem so that we are ensured of the convergence of $T(\hat{F}_n^Z)$ to $T(F^Z)$ and of the strong Frechet differentiability of the functional T.

We are now going to define the smoothed influence function by analogy with Hampel's influence function (see [8]). Using the expansion

$$T(\hat{F}_n^Z) = T(F^Z) + D_{F^Z}T.(\hat{F}_n^Z - F^Z) + O_p\left(\left\|\hat{F}_n^Z - F^Z\right\|^2_{L(\infty,m)}\right)$$

and the expression of

$$\hat{F}_n(\circ) - F^Z = \frac{1}{n}\sum_{i=1}^{n}\left[K_I\left(\frac{\circ - Z_i}{h}\right) - F^Z\right]$$

with the linearity of the Frechet differential, we obtain

$$T(\hat{F}_n^Z) = T(F^Z) + \frac{1}{n}\sum_{i=1}^{n} D_{F^Z}T.\left[K_I\left(\frac{\circ - Z_i}{h}\right) - F^Z\right] + O_p\left(\left\|\hat{F}_n^Z - F^Z\right\|^2_{L(\infty,m)}\right)$$

This expression tells us that the asymptotic error in estimating $T(F^Z)$ by $T(\hat{F}_n^Z)$ can be split into the sum of the contribution of each observation Z_i : $D_{F^Z}T.\left[K_I\left(\frac{\circ - Z_i}{h}\right) - F^Z\right]$. The first idea would then be to define the asymptotic influence of the observation Z_i as $D_{F^Z}T.\left(K_I\left(\frac{\circ - Z_i}{h}\right) - F^Z\right)$. This approach wouldn't be satisfactory since $D_{F^Z}T.\left(K_I\left(\frac{\circ - Z_i}{h}\right) - F^Z\right)$ contains terms of different order, some of which are asymptotically negligible and don't deserve to enter the expression of the influence function. Once again, the representation (R) will allow us to separate these terms and to define the smoothed influence function using only the asymptotic leading terms.

Definition 4 *(i) if $A_{F^Z}(u) = 0$ then we define the smoothed influence function by*

$$SIF(z_i) = D_{F^Z}\Phi.\left(K_I\left(\frac{\circ - Z_i}{h}\right) - F^Z\right)$$

(ii) if $A_{F^Z}(u) \neq 0$ *then we define the smoothed influence function by*

$$SIF(z_i) = \sum_{l \in L^*} \int \alpha_l(F^Z)(u)\partial^{(\Delta^l)}\delta_{z^{(l)}}(u^{(l)})d\left(K_I\left(\frac{u - Z_i}{h}\right) - F^Z\right)$$

An estimator will then be robust in the sense of the smoothed influence function if $SIF(z_i)$ is bounded for $\|z_i\| \to \infty$ (and h fixed).

3 Identification problems

Let's now see how the previous functional theory can be applied in the particular case of a parameter $\theta^0 \in R$ implicitly defined by an equation of the form $\Omega(F^Z, \theta^0) = 0$. The question that naturally arises is to know under which conditions on the functional Ω the equation $\Omega(\hat{F}_n^Z, \theta^0) = 0$, with $\hat{F}_n^Z$ defined as in section 2.1, admits a solution $\hat{\theta}_n^0$. This framework will be useful for the analysis of robustified non-parametric regression.

Theorem 2

Let $\Omega : E_1 \times R \to R$ *be a functional supposed to be strongly Frechet differentiable at the point* (F^Z, θ^0) *with respect to the norm max* $\left\{\|H\|_{L(\infty,q)}, |\theta|\right\}$, $q \leq s$. *We further suppose that* $\frac{\partial\Omega}{\partial\theta}(F^Z, \theta^0) > 0$ *and that* $\frac{\partial\Omega}{\partial\theta}$ *is continuous at the point* (F^Z, θ^0). *If* $\Omega(F^z, \theta^0) = 0$ and if $\hat{F}_n^Z$ *is a consistent estimator of* F^Z *for the norm* $\|H\|_{L(\infty,q)}$, *then :*

(i) There exists a functional T *defined on an open subset* $O \subset E_1$, *continuous at the point* (F^Z) *for the norm* $\|H\|_{L(\infty,q)}$ *and such as the equation* $\Omega(\hat{F}_n^Z, \theta) = 0$ *asymptotically admits (with probability one) a locally unique solution satisfying* $\hat{\theta}_n^0 = T(\hat{F}_n^Z)$.

(ii) T *is strongly Frechet differentiable at the point* F^Z *for the norm* $\|H\|_{L(\infty,q)}$ *with differential given by*

$$D_{F^Z}T.H = -\frac{1}{\frac{\partial\Omega}{\partial\theta}(F^Z, \theta^0)}\frac{\partial\Omega}{\partial F}(F^Z, \theta^0).H$$

where $\frac{\partial\Omega}{\partial F}(F^Z, \theta^0)$ *and* $\frac{\partial\Omega}{\partial F}(F^Z, \theta^0)$ *are the strong partial Frechet differential of* Ω *with respect to* F *and* θ.

4 Non robustness of the Nadaraya Watson regression estimator

We are now going to apply the previous concepts to the particular case of regression estimation. The $d-$dimensional random vector Z is now supposed to be partitioned under the form $Z = (X, Y)$ where Y is the one dimensional stochastic response to the $d-1$ dimensional set of stochastic regressors X.

We first need to state a further assumption that will allow us to show the Frechet differentiability of the functionals we will consider.

(**A4**) : We suppose the density of the vector Z to admit a strictly positive lower bound on its support.

Let's consider the problem of estimating $m(x) = E(Y/X = x)$ using the well-known Nadaraya-Watson estimator (see Nadaraya [14] or Watson [19]). Our aim is to calculate its smoothed influence function and to show its non-robustness in this sense.

We can see $m(x) = \displaystyle\int y\frac{f(x,y)}{f(x)}dy$[5] as a functional of F, we will denote it by $T(F)$. The Nadaraya-Watson estimator of the regression is then defined by :

$$T(\hat{F}_n) = \frac{1}{n}\sum_{i=1}^{n}\frac{K\left(\dfrac{x-X_i}{h}\right)}{\displaystyle\sum_{j=1}^{n}K\left(\dfrac{x-X_j}{h}\right)}Y_i$$

This estimator being a weighted mean of the response Y_i, its nonrobustness has been studied by analogy with results on the estimation of a non conditional expectation. The smoothed influence function we're going to calculate will provide us with a direct criterion to quantify this nonrobustness and to take into account the effect of smoothing. Furthermore, it will allow us to rediscover many properties of the Nadaraya-Watson estimator.

Theorem 3 *We suppose F^Z to be an element of $E_1 = C^d$ and we suppose satisfied the assumptions* $(A1)$, $(A2)(d)$, $(A3)(r + \dfrac{d-1}{2}, d)$ *and* $(A4)$ *then :*

(i) the functional T is strongly Frechet differentiable at the point $F^Z$for the norm $\|\circ\|_{L(\infty,d)}$ and its differential is given by

$$D_{F^Z}T.H = \frac{1}{f(x)}\left[\int yh(x,y)dy - h(x)m(x)\right]$$

[5]With a slight abuse of notations, we will denote by $f(x,y)$ the joint density of the random vector (X,Y) and by $f(x)$ the joint density of the random vector X.

(ii) the functional $T(F^Z)$ is consistently estimated by $T(\hat{F}_n^Z)$ and

$$\sqrt{n}h^{\frac{d-1}{2}}\left(T\left(\hat{F}_n^Z\right)-T\left(F^Z\right)\right)\Longrightarrow N(0,V_T(F^Z))$$

where

$$V_T(F^Z)=\frac{\int K^2(x)dx}{f(x)}E\left[(y-m(x))^2\,|X=x\right]\cdot$$

Corollary 4 *Under the assumptions of theorem 3, the smoothed influence function of the Nadaraya-Watson regression estimator at the point $X=x$ is given by :*

$$SIF(x_i,y_i)=\frac{K(\frac{x-x_i}{h})}{h^{d-1}f(x)}\left[y_i-m(x)\right]$$

- $SIF(x_i,y_i)$ is unbounded for $|y_i|\to\infty$. An outlier Y_i can have an unlimited influence on the estimator $\hat{m}(x)$. We recover with theoretical foundations what intuition suggested to us.
- $SIF(x_i,y_i)$ is bounded for $|x_i|\to\infty$. This means that an outlier X_i has a limited influence on the estimator $\hat{m}(x)$. This property is intuitively coherent since the weight given to X_i decreases with its distance to x.
- the more X_i is distant from x, the weaker the influence of Y_i is.
- the more the density $f(x)$ is important, the weaker the influence of Y_i is. Intuitively, this means that the observation Y_i is flooded in the observations round the point x.

5 Robustification of the Nadaraya-Watson estimator

In this section, we will denote by $F_x(\circ)$ the cumulative distribution function of the random variable Y conditional on the event $\{X=x\}$. We will suppose F_x to be symmetric round $m(x)$ so that the robust conditional location estimators we will propose will be equal to the conditional expectation of Y.

5.1 M-robustified nonparametric regression

We are now going to study the M-estimator of regression (see among others Härdle [9] or Boente and Fraiman [4]). We are interested in the robustness properties an the estimation of the functional T_M implicitly defined by the relation :

$$\int \Psi\left(y - T_M(F^Z)\right) \frac{f(x,y)}{f(x)} dy = 0$$

We will make the following assumption on the score function :

$(M1)$: Ψ is twice continuously differentiable on R that satisfies $\int \frac{\partial \Psi}{\partial t}\left(y - m(x)\right) dF_x(y) > 0$.

$(M2)$: Ψ is an odd function..

$(M3)$: Ψ is bounded on R.

Remark : these assumptions are stronger than those that are usually required for the study of the M estimator of the regression. It has to be noted that our aim is not to lead a study under minimal conditions but to develop a functional framework suitable for the robustness properties of the M-estimator to be analyzed.

Assumption $(M2)$, together with the symmetry or F_x insures that $m(x)$ is a solution to $\int \Psi\left(y - T_M(F)\right) \frac{f(x,y)}{f(x)} dy = 0$. Together with assumption $(M1)$, it ensures that $m(x)$ is locally the unique solution to that equation, and it will allow us to establish the Frechet differentiability of the functional T_M. Assumption $(M3)$ will ensures the robustness of the M-estimator with respect to the smoothed influence function. For more details on the role of each of these assumptions, see the proof of theorem 5 in appendix.

Theorem 5 *We suppose F to be an element of $E_1 = C^d$ and we suppose satisfied the assumptions* $(A1)$, $(A2)(d)$, $(A3)(r + \frac{d-1}{2}, d)$ *and* $(A4)$ *then :*

(i) the functional T_M is Frechet differentiable at the point F^Z for the norm $\|\circ\|_{L(\infty,d)}$ and its differential is given by :

$$D_{F^Z} T_M . H = \frac{\int \Psi\left(y - m(x)\right) h(x,y) dy}{\int \frac{\partial \Psi}{\partial t}\left(y - m(x)\right) f(x,y) dy}$$

(ii) $T_M(\hat{F}_n^Z)$ is a consistent estimator of $T_M(F^Z)$ and we have

$$\sqrt{n} h^{\frac{d}{2}} \left(T_M\left(\hat{F}_n^Z\right) - T_M\left(F^Z\right)\right) \Longrightarrow N(0, V_{T_M}(F^Z))$$

where

$$V_{T_M}(F^Z) = \int \left(\frac{\Psi\left(u_y - m(x)\right)}{\displaystyle\int \frac{\partial \Psi}{\partial t}\left(y - m(x)\right) f(x,y)dy} \right)^2 f(x,y)dy. \int K^2(x)dx$$

Corollary 6 *The smoothed influence function of the M -robustified estimator of the regression is given by*

$$SIF(x_i, y_i) = \frac{\dfrac{1}{h}K\left(\dfrac{x - x_i}{h}\right)\displaystyle\int \Psi\left(hu + y_i - m(x)\right) K\left(u\right) du}{\displaystyle\int \Psi'\left(y - m(x)\right) f(x,y)dy}$$

- $SIF(x_i, y_i)$ is bounded for $|y_i| \to \infty$ (this stems from assumption $(M3)$) so that outliers in the response Y can only have a limited influence on $T_M(\hat{F}_n^Z)$.
- $SIF(x_i, y_i)$ is bounded for $|x_i| \to \infty$ so that $T_M(\hat{F}_n^Z)$ is also robust (in the sense of the smoothed influence function) to the presence of outliers in the regressors.

5.2 L-robustified nonparametric regression

We are now going to consider the L estimator of the regression as proposed by Boente and Fraiman (see [5]) so that we are interested in the estimation of the conditional L-estimate

$$T_L(F^Z) = \int_0^1 J(p) F_x^{-1}(p) dp$$

where J is a score function defined on $[0;1]$.

We will make the following assumptions on the score :

$(L1)$: J_L has compact support $[a,b] \subset]0;1[$.

$(L2)$: J_L is twice continuously differentiable on its support.

The assumption $(L2)$ will ensure the robustness of the L-estimator and the assumption $(L3)$ will allow us to establish the strong Frechet differentiability of the functional T_L.

Theorem 7 *We suppose F^Z to be an element of $E1$, the closed subspace of the functions $F \in C^d$ such as $\dfrac{\partial^d F}{\partial z_1 .... \partial z_d}$ admits a strictly lower bound on its supports. We suppose further satisfied the assumptions $(A1)$, $(A2)(d)$, $(A3)(r + \dfrac{d-1}{2}, d)$ and $(A4)$ then :*

(i) the functional T_R is strongly Frechet differentiable at the point F^Z for the norm $\|\circ\|_{L(\infty,d)}$ and its differential is given by :

$$D_{F^Z} T_L.H = \int_0^1 \frac{J_L(p)}{f(x, F_x^{-1}(p))} \left[ph(x) - \frac{\partial^{d-1} h}{\partial x^1 ....\partial x^{d-1}}(x, F_x^{-1}(p)) \right] dp$$

(ii) $T_L(\hat{F}_n)$ is a consistent estimator of $T_L(F)$ and :

$$\sqrt{n} h^{\frac{d-1}{2}} \left(T_L\left(\hat{F}_n^Z\right) - T_L\left(F^Z\right) \right) \Longrightarrow N(0, V_{T_L}(F^Z))$$

where

$$V_{T_L}(F^Z) = \int \left[\int_0^1 J_L(p) \frac{p - I_{]-\infty,q_0]}(y)}{f(x, F_x^{-1}(p))} dp \right]^2 f(x,y)dy \cdot \int K^2(x)dx$$

Corollary 8 *The smoothed influence function of the conditional L-estimate $T_L(\hat{F}_n)$ is given by :*

$$\begin{aligned} SIF(x_i, y_i) &= \left[\frac{1}{h^d} K\left(\frac{x - x_i}{h}\right) - f(x) \right] \int_a^b \frac{J_L(p)}{f(x, F_x^{-1}(p))} pdp \\ &\quad - \frac{1}{h^{d-1}} K\left(\frac{x - x_i}{h}\right) \int_a^b \frac{J_L(p)}{f(x, F_x^{-1}(p))} K_I\left(\frac{F_x^{-1}(p) - y_i}{h}\right) dp \\ &\quad + \int_b^a \frac{J_L(p)}{f(x, F_x^{-1}(p))} \frac{\partial^{d-1} F}{\partial x^1 ....\partial x^{k-1}}(x, F_x^{-1}(p)) dp \end{aligned}$$

- $SIF(xi, y_i)$ is bounded for $|y_i| \to \infty$ (under assumption $(L2)$). An outlier in the observations Y_i can only have a limited influence on the estimation of the conditional L-estimate.
- $SIF(xi, y_i)$ is bounded for $|x_i| \to \infty$ which means that outliers among the regressors have a limited influence.

5.3 R-robustified nonparametric regression

We are now going to consider the conditional R-estimator of the regression as proposed by Cheng and Cheng (see [6]) so that we are now interested in the estimation and the robustness properties of the functional $T_R(F)$ implicitly defined by

$$\int_{-\infty}^{+\infty} J_R\left[2^{-1}\left(F_x(y) + 1 - F(2T_R(F) - y)\right)\right] dF_x(y) = 0$$

where J_R is a score function defined on $[0;1]$.

We will make the following assumptions on the score :

$(R1)$: J_R is twice continuously differentiable on $[0;1]$ $\int\limits_{-\infty}^{+\infty} J'_R\left(F_x(y)\right)\left(F'_x(y)\right)^2 dy >$ 0.

$(R2)$: J_R si an increasing function on $[0;1]$ with $J_R(t) = -J_R(1-t)$.

The assumption $(R1)$ will allow us to establish the strong Frechet differentiability of the functional T_R. The assumption $(R2)$ ensures the identification of $T_R(F^Z)$. Together with the symmetry assumptions on F_x it also ensures that $T_R(F^Z) = m(x)$.

Theorem 9 *we suppose F^Z to be an element of $E_1 = C^{d+1}$ and we suppose satisfied the assumptions $(A1)$, $(A2)(d+1)$, $(A3)(r + \dfrac{d-1}{2}, d+1)$ and $(A4)$ then :*

(i) the functional T_R is Frechet differentiable for the norm $\|\cdot\|_{L(\infty,d+1)}$ and its differential is given by :

$$D_{F^Z} T_R . H = \frac{\int\limits_{-\infty}^{+\infty} J_R\left[F_x(y)\right] h(x,y) dy}{\int\limits_{-\infty}^{+\infty} J'_R\left[F_x(y)\right]\left(F'_x(y)\right)^2 dy}$$

(ii) $T_R(\hat{F}_n)$ is a consistent estimator of $T_R(F)$ and :

$$\sqrt{n} h^{\frac{d-1}{2}} \left(T_R\left(\hat{F}^Z_n\right) - T_R\left(F^Z\right)\right) \Longrightarrow N(0, V_{T_R}(F))$$

where

$$V_{T_R}(F^Z) = \frac{\int\limits_{-\infty}^{+\infty} J^2_R(t) dt}{\left[\int\limits_{-\infty}^{+\infty} J'_R\left[F_x(y)\right]\left(F'_x(y)\right)^2 dy\right]^2} \cdot \int K^2(x) dx$$

Corollary 10 *Under the assumptions of theorem 5, the smoothed influence function of the conditional R-estimate $T_R(\hat{F}^Z_n)$ is given by :*

$$SIF(x_i,y_i)=\frac{K\left(\frac{x-x_i}{h}\right)\int\limits_{-\infty}^{+\infty}J\left(F_x(hu+y_i\right)K(u)du}{\int\limits_{-\infty}^{+\infty}J'\left[F_x(y)\right]\left(F_x'(y)\right)^2dy}$$

- $SIF(x_i,y_i)$ is bounded both for $x_i\to\infty$ and for $y_i\to\infty$ so that outliers among the regressors or among the regressand can only have a limited influence on the estimation of $m(x)$.

Remark 1 *the boundedness of $SIF(x_i,y_i)$ for $y_i\to\infty$ stems from the fact that the score is supposed to be twice continuously differentiable on $[0;1]$, this assumptions excludes the normal score.*

6 Computational concepts

6.1 Smoothing's effects

In this work, we have considered the M, L and R estimators of the regression as functionals of F^Z, the joint cumulative distribution function of $Z=(X,Y)$. The estimation procedure consists of plugging the smoothed cumulative distribution function $\hat{F}_n^Z$. Another approach, studied by various authors (see for instance Boente and Fraiman [4]), consists of studying these estimators as functionals of F_x, the cumulative distribution of Y, conditional on the event $\{X=x\}$. The estimation procedure they propose is based on the plug-in of an empirical estimate of F_x :

$$F_x^n(y)=\sum_{i=1}^n\frac{K\left(\frac{x-X_i}{h}\right)}{\sum\limits_{j=1}^n K\left(\frac{x-X_j}{h}\right)}I_{[Y_i\leq y]}$$

Our estimation procedure can be linked to this estimation procedure since it can also be viewed as the plug-in of a smoothed estimate of F_x :

$$\hat{F}_x^n(y)=\sum_{i=1}^n\frac{K\left(\frac{x-X_i}{h}\right)}{\sum\limits_{j=1}^n K\left(\frac{x-X_j}{h}\right)}\int\limits_{-\infty}^{\frac{y-Y_i}{h}}K(u)du$$

There are many differences between these estimators : ours introduce an additional small sample bias which is due to the effect of smoothing on Y. From an asymptotic point of view, this bias is made negligible by the conditions imposed to the bandwidth so that the asymptotic distributions of the M, L and R estimators are identical, no matter on the estimate of F_x which is plugged.

More important a difference between these estimators relies in their regularity properties : the empirical conditional distribution function has a step structure function whereas the smoothed conditional distribution function is at least a continuous function. This difference is of major importance on a theoretical point of view since it has allowed us to solve the identification problem using the implicit function theorem (see section 3) and to define the M, L and R estimators uniquely. This difference has also practical consequences since it allows for the use of classical numerical algorithms for inversion (estimation of the conditional quantiles for the L-estimator) and search of zeros (for the R estimator).

6.2 Monte-Carlo simulation

In order to illustrate the behavior of the robustified estimators, we performed a Monte-Carlo simulation. The data generating process we have chosen is the following :

$$Y = m(X) + \varepsilon$$

where m is the quadratic function $m(x) = x^2$, X is uniformly distributed on $[-1; 1]$ and ε is a mixture of normal laws with probability density function

$$f(\varepsilon) = \frac{1-\nu}{\sqrt{2\pi}} exp\left(-\frac{\varepsilon^2}{2}\right) + \frac{\nu}{\sqrt{2\pi k}} exp\left(-\frac{\varepsilon^2}{2k}\right)$$

ν represents the degree of contamination of the standard normal law. This degree of contamination has successively been set equal to 5%, 10% and 15%. k represents the variance of the contaminating law and has been set equal to 9.

In the case $\nu = 10\%$, this model of contamination implies that, in a sample of 100 observations, 10 are expected to come from the contaminating law. But on these 10 observations, only those who will be outside the interval $[m(x) - 2.5; m(x) + 2.5]$ will be noticeable as outliers (in the interval $[-2.5; 2.5]$, the cdf of the contaminating is almost confounded with those of the standard normal law). For the contaminating law $N(0, 9)$, the probability of finding observations outside of the interval $[m(x) - 2.5; m(x) + 2.5]$ is equal to 0.4 so that we can only expect 4 observations of the contaminating law to lie outside the interval $[m(x) - 2.5; m(x) + 2.5]$ (for more details see Huber [11]). The identification of these observations would necessitate a preliminary estimation of $m(x)$ and this estimation would surely be imprecise due to the outliers. Furthermore, on the 4 observations lying outside the interval $[m(x) - 2.5; m(x) + 2.5]$, some of them may come from the standard normal law and would be indistinguishable from those coming from the contaminating law.

With this data generating process, the law of Y conditional on the regressors X is symmetric so that with the scores we have chosen (see below), the conditional expectation is equal to both $T_M(F^Z)$, $T_L(F^Z)$ and $T_R(F^Z)$.

For the M-robustified estimator, we chose the score function

$$\Psi(e) = \begin{cases} e \text{ if } |e| \leq c_1 \\ \tilde{\Psi}(e) \text{ if } c_1 < |e| \leq c_2 \\ 0 \text{ if } |e| > c_2 \end{cases}$$

where $\tilde{\Psi}$ is a fifth order polynomial chosen in such a way that Ψ is twice continuously differentiable and c_1, c_2 are trimming constants. Following Lucas (see [13]), we chose $c_1 = \sqrt{\chi^{-1}(0.99)}$ and $c_2 = \sqrt{\chi^{-1}(0.999)}$. If $Y_i - m(x)$ follows a standardized normal law, such a choice ensures that the observations for which $|Y_i - m(x)| > 3.5$ are discarded and the observations for which $3.5 > |Y_i - m(x)| > 2.8$ are downweighted if $Y_i - m(x)$ (for more details see Lucas [13]). $T_M(\hat{F}_n)$ is obtained as

$$T_M(\hat{F}_n^z) = argsol_\theta \left\{ \sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right) \int \Psi\left(Y_i + hu - \theta\right) K(u)du = 0 \right\}$$

where K is the Gaussian kernel.

For the L-robustified estimator, we chose the score function $J_L(p) = \dfrac{1}{1-2\alpha} I_{[\alpha;1-\alpha]}(p)$ which correspond to the α conditional trimmed expectation. The trimming parameter was set equal to 5%. We took

$$T_L(\hat{F}_n^Z) = \frac{1}{1-2\alpha} \int_{\alpha}^{1-\alpha} (\hat{F}_n^x)^{-1}(p)dp$$

with $\hat{F}_n^x$ defined in section 6.1, the Kernel K being the Gaussian kernel. The integral was approximated by a Riemann sum calculated on a grid of 50 points regularly spaced on the interval $[\alpha; 1-\alpha]$.

For the R-robustified estimator, we chose the score $J_R(e) = e - \frac{1}{2}$ which corresponds to the Hodges-Lehmann estimator. $T_R(\hat{F}_n)$ was obtained as

$$T_R(\hat{F}_n^Z) = argsol_\theta \left\{ \int_{-\infty}^{+\infty} J_R\left[2^{-1}\left(\hat{F}_n^x(y) + 1 - \hat{F}_n^x(2\theta - y)\right)\right] d\hat{F}_n^x(y) = 0 \right\}$$

where $\hat{F}_n^x$ is calculated as in section 6.1 with the kernel $K(u) = \frac{1}{2\sqrt{2\pi}}(3-u^2)\exp(-\frac{u^2}{2})$ which is a fourth order kernel as required by assumption $A(r, \frac{3}{2})$ (see theorem 5 with $d = 2$).

In order to compare the performances of our estimators, we estimated the regression function on a grid $\{t_1, ...t_{20}\}$ equally spaced on $[-0.7; 0.7]$ [6]. Our criterion of

[6] We restricted the grid to $[-0.7; 0.7]$ in order to avoid the boundary effects.

comparison is $C = \frac{1}{20}\sum_{i=1}^{20}(m(t_i) - \hat{m}(t_i))^2$ where $\hat{m}$ will be either the Nadaraya watson estimator of the regression or one of the robustified estimators of the regression. Our results are consigned in table 1 below :

	degree	of	contamination
estimator	5%	10%	15%
Nadaraya-watson	1.20	1.24	1.36
M-robustified	0.54	0.50	0.38
L-robustified	0.57	0.59	0.64
R-robustified	0.40	0.36	0.47

Table 1 shows us the improvement achieved by robustified estimators since in each contamination case, the criterion C for the Nadaraya-Watson estimator is twice as large as the criterion C of the robustified estimator. Of course, we would obtain different results if we would choose different tuning constant for the M-estimator or a different trimming parameter for the L-estimator but, the choices we have made are reasonable[7] choices when one doesn't know the amount of outliers in the sample. The good performances of the R-robustified estimator have to be noticed, since, contrarily to the M and L estimators, no parameter has to be chosen for its use; in this sense, one can think it's less subjective an estimator.

7 Conclusion

In this work, we have developed the concept of smoothed influence function that allows us to quantify the robustness properties of a large class of nonparametric estimators. With this concept, we have shown that M, L and R regression achieved robustness. Such robustness properties had already been stated by empirical means but so far, no tool was available for a theoretical analysis.

Of course, the field of application of the smoothed influence is much wider than those of regression and such robustness studies could be lead for many others nonparametric and semi-parametric estimators that enter in the framework of Von-Mises statistical functionals. This will be done in further work.

A proofs

The proofs of Aït-Sahalia's generalized delta theorem 1 and of theorem 3 can be found in Aït-Sahalia (see [1]).

[7] by reasonable, we mean choices we would recommend when one suspects the presence of outliers in the sample.

A.1 proof of theorem 2

The point (i) of this theorem is exactly the implicit function theorem since strong Frechet differentiability implies Frechet differentiability : there exists an open subset $O \subset E_1$ and a continuous functional T such as, for all $F \in O$, the equation $\Omega(F, \theta) = 0$ is equivalent to $\theta = T(F)$.

$\hat{F}_n^Z$ being supposed to be a consistent estimator of F^Z for the norm $\|\circ\|_{L(\infty,q)}$ it will asymptotically be in O so that $\Omega(\hat{F}_n, \theta) = 0$ will admit the (locally) unique solution $\hat{\theta}_n^o = T(\hat{F}_n^Z)$.

The point (ii) is a slight modification of the proof of the implicit function theorem using the concept of strong Frechet differentiability. This proof can be found for instance in Zeidler (see [20])

A.2 Proof of corollary 4

Under the assumptions of theorem 1, the functional T is Frechet differentiable with

$$D_{F^Z}T.H = \frac{1}{f(x)}\left[\int yh(x,y)dy - h(x)m(x)\right]$$

The smoothed influence function of $T(\hat{F}_n^Z)$ is defined by

$$SIF(x_i, y_i) = D_{F^Z}T.\left(K_I\left(\frac{\circ - x_i}{h}\right)K_I\left(\frac{\circ - y_i}{h}\right) - F^Z\right)$$

with $D_{F^Z}T.F = 0$, we obtain

$$SIF(x_i, y_i) = \frac{1}{f(x)}\int y\frac{1}{h^d}K\left(\frac{x - x_i}{h}\right)K\left(\frac{y - y_i}{h}\right)dy - \frac{1}{h^{d-1}f(x)}K\left(\frac{x - x_i}{h}\right)m(x)$$

which can also be written

$$SIF(x_i, y_i) = \frac{K\left(\frac{x - x_i}{h}\right)}{h^{d-1}f(x)}\left[\frac{1}{h}\int yK\left(\frac{y - y_i}{h}\right)dy - m(x)\right]$$

if we make the change of variable $u = \frac{y - y_i}{h}$ in the integral, we get

$$SIF(x_i, y_i) = \frac{K\left(\frac{x - x_i}{h}\right)}{h^{d-1}f(x)}\left[\int uK(u)\,du + y_i\int K(u)\,du - m(x)\right]$$

with $\int uK(u)\,du = 0$ and $\int K(u)\,du = 1$, this expression is reduced to

$$SIF(x_i, y_i) = \frac{K(\frac{x - x_i}{h})}{h^{d-1}f(x)}\left[y_i - m(x)\right]$$

Preliminary : we are first going to introduce some notations and establish some results that we will need in the following :

for $t \in [0;1]$, let's define

$$\varphi_{x,y}(t) = \frac{f(x,y) + th(x,y)}{f(x) + th(x)}$$

We have

$$\varphi'_{x,y}(t) = \frac{h(x,y)f(x) - f(x,y)h(x)}{(f(x) + th(x))^2}$$

and

$$\varphi''_{x,y}(t) = -2h(x)\frac{h(x,y)f(x) - f(x,y)h(x)}{(f(x) + th(x))^3}$$

Following Aït-Sahalia, Bickel and Stoker (see [2]), we are going to establish a few majorations that will be useful to us. From assumption $(A4)$, there exists a constant $c > 0$ such as $|f(x)| > c$. For H in a neighborhood of zero for the norm $\|\circ\|_{L(\infty,d)}$ we can suppose $|h(x)| < \frac{c}{2}$. Applying a triangular inequality to $|f(x) + th(x)|$, we obtain $\left|\frac{1}{f(x) + th(x)}\right| < \frac{2}{c}$ for all $t \in [0;1]$. Using this majoration, we obtain immediately

$$\varphi'_{x,y}(t) = O\left(\|H\|_{L(\infty,d)}\right)$$

and

$$\varphi''_{x,y}(t) = O\left(\|H\|^2_{L(\infty,d)}\right)$$

A.3 Proof of theorem 5

With the notations of theorem 1, let's define $\Omega_M(F,\theta) = \int \Psi\left(y-\theta\right)\frac{f(x,y)}{f(x)}dy$. Note that since Ψ is continuous (assumption $(M1)$), $\int \Psi\left(y-\theta\right)\frac{f(x,y)}{f(x)}dy$ is defined for all $F \in C^d$ and $\theta \in R$ (i.e. we take $E_1 = C^d$).

Let's also define the function $\omega_M \begin{cases} t \to \Omega_M(F + tH, \theta + tk) \\ [0;1] \to R \end{cases}$. f and h being compactly supported and Ψ being twice continuously differentiable (assumption $(M1)$), differentiation with respect to t under the integral operator yields :

$$\omega'_M(t) = -k\int \Psi'\left(y - (\theta + tk)\right)\varphi_{x,y}(t)dy + \int \Psi\left(y - (\theta + tk)\right)\varphi'_{x,y}(t)dy$$

and

$$\begin{aligned}\omega_M''(t) &= k^2\int \Psi''\left(y-(\theta+tk)\right)\varphi_{x,y}(t)dy - 2k\int \Psi'\left(y-(\theta+tk)\right)\varphi_{x,y}'(t)dy \\ &\quad + \int \Psi\left(y-(\theta+tk)\right)\varphi_{x,y}''(t)dy\end{aligned}$$

If we make a Taylor expansion of ω_M between 0 and 1, we find

$$\omega_M(1) = \omega_M(0) + \omega_M'(0) + \frac{1}{2}\omega_M''(c) \quad (1)$$

where $c \in \left]0;1\right[$.

Using the majorations established in the preliminary and the assumption $(M1)$, we readily find that

$$\omega_M''(c) = O\left(sup\left(\|H\|_{L(\infty,d)}, |k|\right)^2\right)$$

Expansion (1) can be written $\Omega_M(F+H,\theta+k) = \Omega_M(F,\theta) - k\int \Psi'(y-\theta)\frac{f(x,y)}{f(x)}dy$ $+\int \Psi(y-\theta)\frac{h(x,y)f(x)-h(x)f(x,y)}{f^2(x)}dy + O\left(sup\left(\|H\|_{L(\infty,d)}, |k|\right)^2\right)$ where the functional
$(H,k) \to \left[-k\int \Psi'(y-\theta)\frac{f(x,y)}{f(x)}dy + \int \Psi(y-\theta)\frac{h(x,y)f(x)-h(x)f(x,y)}{f^2(x)}dy\right]$ is linear and continuous. This establishes the Frechet differentiability of the functional Ω_M.

The symmetry of F_x, together with assumption $(M2)$ imply that the equation $\int \Psi(y-\theta)\frac{f(x,y)}{f(x)}dy = 0$ admits for solution $\theta = m(x)$. Furthermore $\frac{\partial \Omega_M}{\partial \theta}(F, m(x)) = -\int \Psi'(y-m(x))\frac{f(x,y)}{f(x)}dy$ is strictly positive[8] from assumption $(M1)$ so that, together with the Frechet differentiability of the functional Ω_M previously established, and the convergence of $\hat{F}_n^Z$ to F^Z for the norm $\|\circ\|_{L(\infty,d)}$ (under the assumptions of theorem 3, see Aït Sahalia [1]), we can apply the point (i) of theorem 1 : there exists a functional T_M such as $m(x) = T_M(F)$ and such as the equation $\Omega_M(\hat{F}_n,\theta) = 0$ asymptotically admits (locally) a unique solution $T_M(\hat{F}_n)$.

Applying the point (ii) of theorem 1, we obtain the Frechet differentiability of the functional T_M for the norm $\|\circ\|_{L(\infty,d)}$ with

$$D_{F^Z}T_M.H = \frac{\int \Psi\left(y-m(x)\right)h(x,y)dy}{\int \Psi'\left(y-m(x)\right)f(x,y)dy}$$

[8]Note that differentiation under the integral operator is, one's more justified by the regularity properties of F and Ψ.

$D_F T_M$ can also be written under the form

$$D_{F^z} T_M.H = \int \frac{\Psi\left(u_y - m(u_x)\right)}{\displaystyle\int \frac{\partial \Psi}{\partial t}\left(y - m(u_x)\right) f(u_x, y)dy} \delta_x\left(u_x\right) h(u_x, u_y) du_x du_y$$

so that $\varkappa^* = d - 1$ and $\sqrt{n} h^{\frac{d-1}{2}} \left[T_M(\hat{F}_n^Z) - T_M(F^Z)\right] \Longrightarrow N(0, V_{T_M}(F^Z))$ with

$$V_{T_M}(F^Z) = \int \left(\frac{\Psi\left(u_y - m(x)\right)}{\displaystyle\int \Psi'\left(y - m(x)\right) f(x, y)dy} \right)^2 f(x, u_y) du_y . \int K^2(x)dx$$

A.4 Proof of corollary 6

The smoothed influence function of the estimator $T_M(\hat{F}_n^Z)$ is given by

$$SIF(x_i, y_i) = D_{F^Z} T_M . \left(K_I \left(\frac{\circ - x_i}{h} \right) K_I \left(\frac{\circ - y_i}{h} \right) - F \right)$$

with

$$D_{F^Z} T_M.H = \frac{\displaystyle\int \Psi\left(y - m(x)\right) h(x, y)dy}{\displaystyle\int \Psi'\left(y - m(x)\right) f(x, y)dy}$$

we obtain

$$SIF(x_i, y_i) = \frac{\frac{1}{h} K \left(\frac{x - x_i}{h} \right) \displaystyle\int \Psi\left(y - m(x)\right) \frac{1}{h} K \left(\frac{y - y_i}{h} \right) dy - \int \Psi\left(y - m(x)\right) f(x, y)dy}{\displaystyle\int \Psi'\left(y - m(x)\right) f(x, y)dy}$$

i.e. with $\displaystyle\int \Psi\left(y - m(x)\right) f(x, y)dy = f(x) \int \Psi\left(y - m(x)\right) \frac{f(x, y)}{f(x)} dy = 0$ and the change of variable $u = \dfrac{y - y_i}{h}$, we finally get

$$SIF(x_i, y_i) = \frac{\frac{1}{h} K \left(\frac{x - x_i}{h} \right) \displaystyle\int \Psi\left(hu + y_i - m(x)\right) K\left(u\right) du}{\displaystyle\int \Psi'\left(y - m(x)\right) f(x, y)dy}$$

A.5 Proof of theorem 7

With the notations of theorem 1, let's define

$\Omega_L(F,\theta) = \theta - \int_a^b J_L(p)F_x^{-1}(p)dp = \theta - \int_{F_x^{-1}(a)}^{F_x^{-1}(b)} yJ_L\left(F_x(y)\right)dF_x(y)$. Let's take for E_1 the subspace of the functions $F \in C^d$ such as $\frac{\partial^d F}{\partial z_1 .... \partial z_d}$ admits a strictly lower bound on its support. E_1 is a closed subspace of C^d and $\Omega_L(F,\theta)$ is defined for all $F \in E_1$ and $\theta \in R$.

Let's also define the function $\omega_L \begin{cases} [0;1] \rightarrow R \\ t \rightarrow \Omega_L(F+tH, \theta+tk) \end{cases}$. f and h being compactly supported and J_L being twice continuously differentiable (assumption $(L3)$), differentiation with respect to t under the integral operator yields :

$$\begin{aligned}
\omega'_L(t) &= k - \int_{F_x^{-1}(a)}^{F_x^{-1}(b)} y\left(\int_{-\infty}^{y} \varphi'_{x,v}(t)dv\right) J'_L\left(\int_{-\infty}^{y} \varphi_{x,v}(t)dv\right)\varphi_{x,y}(t)dy \\
&\quad - \int_{F_x^{-1}(a)}^{F_x^{-1}(b)} yJ_L\left(\int_{-\infty}^{y} \varphi_{x,v}(t)dv\right)\varphi'_{x,y}(t)dy
\end{aligned}$$

and

$$\begin{aligned}
\omega''_L(t) &= -\int_{F_x^{-1}(a)}^{F_x^{-1}(b)} y\left(\int_{-\infty}^{y} \varphi''_{x,v}(t)dv\right) J'_L\left(\int_{-\infty}^{y} \varphi_{x,v}(t)dv\right)\varphi_{x,y}(t)dy \\
&\quad - \int_{F_x^{-1}(a)}^{F_x^{-1}(b)} y\left(\int_{-\infty}^{y} \varphi'_{x,v}(t)dv\right)^2 J''_L\left(\int_{-\infty}^{y} \varphi_{x,v}(t)dv\right)\varphi_{x,y}(t)dy \\
&\quad -2\int_{F_x^{-1}(a)}^{F_x^{-1}(b)} y\left(\int_{-\infty}^{y} \varphi'_{x,v}(t)dv\right) J'_L\left(\int_{-\infty}^{y} \varphi_{x,v}(t)dv\right)\varphi'_{x,y}(t)dy \\
&\quad - \int_{F_x^{-1}(a)}^{F_x^{-1}(b)} yJ_L\left(\int_{-\infty}^{y} \varphi_{x,v}(t)dv\right)\varphi''_{x,y}(t)dy
\end{aligned}$$

Using the majorations established in the preliminary and the assumption $(L3)$ on the score J_L we readily find $\omega''_L(t) = O\left(sup\left(\|H\|_{L(\infty,d)}, |k|\right)^2\right)$. A Taylor expansion of ω_L between o and 1 (see proof of theorem 3) establishes the Frechet differentiability of the functional Ω_L.

The equation $\Omega_L(F,\theta)=0$ obviously admits the solution $\theta = \int\limits_{F_x^{-1}(a)}^{F_x^{-1}(b)} yJ_L\left(F_x(y)\right)dF_x(y)$.

Furthermore, we have $\dfrac{\partial\Omega_L}{\partial\theta}(F,\theta) = 1 > 0$ so that, using the Frechet differentiability of the functional Ω_L previously established, and the convergence of $\hat{F}_n^Z$ to F^Z for the norm $\|\circ\|_{L(\infty,d)}$ (under the assumptions of theorem 3, see Aït Sahalia [1]), we can apply the point (i) of theorem 2 : there exists a functional $T_L(F) = \int\limits_{F_x^{-1}(a)}^{F_x^{-1}(b)} yJ_L\left(F_x(y)\right)dF_x(y)$ such as the equation $\Omega_L(\hat{F}_n,\theta)=0$ asymptotically admits (locally) a unique solution $T_R(\hat{F}_n)$.

Note : It would not be necessary here to apply the theorem 2 since the functional T_L is defined in an explicit way by $T_L(F^Z) = \int\limits_{F_x^{-1}(a)}^{F_x^{-1}(b)} yJ_L\left(F_x(y)\right)dF_x(y)$. Nevertheless, its use allows us to solve rapidly the problem of the existence of $T(\hat{F}_n^Z)$.

Applying the point (ii) of theorem 2, with the change of variable $p=F_x(y)$, we obtain the strong Frechet differentiability of the functional T_L for the norm $\|\circ\|_{L(\infty,d)}$ with

$$D_{F^Z}T_L.H = -\int\limits_a^b J_L(p)\frac{f(x)}{f(x,F_x^{-1}(p))}\left(\int\limits_{-\infty}^{F_x^{-1}(p)}\frac{h(x,v)f(x)-f(x,v)h(x)}{f^2(x)}dv\right)dp$$

which can also be written

$$D_{F^Z}T_L.H = \int\limits_a^b \frac{J_L(p)}{f(x,F_x^{-1}(p))}\left[ph(x) - \frac{\partial^{d-1}h}{\partial x^1....\partial x^{d-1}}(x,F_x^{-1}(p))\right]dp$$

Using generalized functions, $D_{F^Z}T_L$ can also be written under the form

$$D_{F^Z}T_L.H = \int\left[\int\limits_a^b \frac{J_L(p)}{f\left(u_x,F_{u_x}^{-1}(p)\right)}\left[p - I_{\left[-\infty;F_{u_x}^{-1}(p)\right]}\right]dp\right]h(u_x,u_y)\delta_x(u_x)du_x du_y$$

so that $\varkappa^* = d-1$ and $\sqrt{n}h^{\frac{d-1}{2}}\left[T_L(\hat{F}_n^Z) - T_L(F^Z)\right] \Longrightarrow N(0,V_{T_L}(F^Z))$ with

$$V_{T_L}(F^Z) = \int\left[\int\limits_a^b J_L(p)\frac{p - I_{]-\infty,q_0]}(y)}{f(x,q_0)}dp\right]^2 f(x,y)dy\cdot\int K^2(x)dx$$

A.6 proof of corollary 8

The smoothed influence function for the estimator $T(\hat{F}_n^Z)$ is given by

$$SIF(x_i^k, x_i^{-k}, y_i) = D_{F^Z}T_L.\left(K_I\left(\frac{\circ - x_i}{h}\right)K_I\left(\frac{\circ - y_i}{h}\right) - F\right)$$

with

$$D_{F^Z}T_L.H = \int_a^b \frac{J_L(p)}{f(x, F_x^{-1}(p))}\left[ph(x) - \frac{\partial^{d-1}H}{\partial x^1....\partial x^{d-1}}(x, F_x^{-1}(p))\right]dp$$

we obtain $SIF(x_i, y_i) = \int_a^b \frac{J_L(p)}{f(x, F_x^{-1}(p))}p\left[\frac{1}{h^d}K\left(\frac{x - x_i}{h}\right) - f(x)\right]dp$

$$-\int_a^b \frac{J_L(p)}{f(x, F_x^{-1}(p))}\left[\frac{1}{h^{d-1}}K\left(\frac{x - x_i}{h}\right)K_I\left(\frac{F_x^{-1}(p) - y_i}{h}\right) - \frac{\partial^{d-1}F}{\partial x^1....\partial x^{d-1}}(x, F_x^{-1}(p))\right]dp$$

i.e.

$$\begin{aligned}SIF(x_i, y_i) &= \left[\frac{1}{h^d}K\left(\frac{x - x_i}{h}\right) - f(x)\right]\int_a^b \frac{J_L(p)}{f(x, F_x^{-1}(p))}pdp \\ &\quad -\frac{1}{h^{d-1}}K\left(\frac{x - x_i}{h}\right)\int_a^b \frac{J_L(p)}{f(x, F_x^{-1}(p))}K_I\left(\frac{F_x^{-1}(p) - y_i}{h}\right)dp \\ &\quad +\int_a^b \frac{J_L(p)}{f(x, F_x^{-1}(p))}\frac{\partial^{d-1}F}{\partial x^1....\partial x^{d-1}}(x, F_x^{-1}(p))dp\end{aligned}$$

A.7 proof of theorem 9

With the notations of theorem 2, let's define $\Omega_R(F, \theta) = \int_{-\infty}^{+\infty} J_R\left[2^{-1}\left(F_x(y) + 1 - F_x(2\theta - y)\right)\right]dF_xy.$

Since J_R is continuous on $[0;1]$ from assumption $(R1)$, Ω_R is defined for all $F \in C^{d+1}$ (so that we will take $E_1 = C^{d+1}$) and $\theta \in R$.

Let's also define the function $\omega_R\left\{\begin{array}{c}[0;1] \to R \\ t \to \Omega_R(F + tH, \theta + tk)\end{array}\right.$. f, $\frac{\partial f}{\partial y}$ and h being compactly supported and J_R being twice continuously differentiable (assumption

$(R1)$), differentiation with respect to t under the integral operator yields :

$$
\begin{aligned}
\omega_R'(t) \;=\; & 2^{-1}\int_{-\infty}^{+\infty}\left[\int_{-\infty}^{y}\varphi_{x,v}'(t)dv-\int_{-\infty}^{2(\theta+tk)-y}\varphi_{x,v}'(t)dv-2k\varphi_{x,2(\theta+tk)-y}(t)\right]\\
& .J_R'\left[2^{-1}\left(\int_{-\infty}^{y}\varphi_{x,v}(t)dv+1-\int_{-\infty}^{2(\theta+tk)-y}\varphi_{x,v}(t)dv\right)\right]\varphi_{x,y}(t)dy\\
& +\int_{-\infty}^{+\infty}J_R\left[2^{-1}\left(\int_{-\infty}^{y}\varphi_{x,v}(t)dv+1-\int_{-\infty}^{2(\theta+tk)-y}\varphi_{x,v}(t)dv\right)\right]\varphi_{x,y}'(t)dy
\end{aligned}
$$

and

$$
\begin{aligned}
\omega_R''(t) \;=\; & 2^{-1}\int_{-\infty}^{+\infty}\left[\int_{-\infty}^{y}\varphi_{x,v}''(t)dv-\int_{-\infty}^{2(\theta+tk)-y}\varphi_{x,v}''(t)dv-2k\left[\left(\varphi_{x,2(\theta+tk)-y}(t)\right)'+\varphi_{x,2(\theta+tk)-y}'(t)\right]\right]\\
& J_R'\left[2^{-1}\left(\int_{-\infty}^{y}\varphi_{x,v}(t)dv+1-\int_{-\infty}^{2(\theta+tk)-y}\varphi_{x,v}(t)dv\right)\right]\varphi_{x,y}(t)dy\\
& +4^{-1}\int_{-\infty}^{+\infty}\left[\int_{-\infty}^{y}\varphi_{x,v}'(t)dv-\int_{-\infty}^{2(\theta+tk)-y}\varphi_{x,v}'(t)dv-2k\varphi_{x,2(\theta+tk)-y}(t)\right]^2\\
& J_R''\left[2^{-1}\left(\int_{-\infty}^{y}\varphi_{x,v}'(t)dv+1-\int_{-\infty}^{2(\theta+tk)-y}\varphi_{x,v}'(t)dv\right)\right]\varphi_{x,y}(t)dy\\
& +\int_{-\infty}^{+\infty}\left[\int_{-\infty}^{y}\varphi_{x,v}'(t)dv-\int_{-\infty}^{2(\theta+tk)-y}\varphi_{x,u}'(t)du-2k\varphi_{x,2(\theta+tk)-y}(t)\right]\\
& J_R'\left[2^{-1}\left(\int_{-\infty}^{y}\varphi_{x,v}'(t)dv+1-\int_{-\infty}^{2(\theta+tk)-y}\varphi_{x,v}'(t)dv\right)\right]\varphi_{x,y}'(t)dy\\
& +\int_{-\infty}^{+\infty}J_R\left[2^{-1}\left(\int_{-\infty}^{y}\varphi_{x,v}(t)dv+1-\int_{-\infty}^{2(\theta+tk)-y}\varphi_{x,v}(t)dv\right)\right]\varphi_{x,y}''(t)dy
\end{aligned}
$$

where $\left(\varphi_{x,2(\theta+tk)-y}(t)\right)'=\dfrac{2k\left(\dfrac{\partial f}{\partial y}\left(x,2\left(\theta+tk\right)-y\right)+t\dfrac{\partial h}{\partial y}\left(x,2\left(\theta+tk\right)-y\right)\right)}{\left(f(x)+th(x)\right)^2}$

$+\dfrac{h(x,2(\theta+tk))f(x)-h(x)f(x,2(\theta+tk))}{(f(x)+th(x))^2}$ which can be shown (using the same kind of majorations as in preliminary and the fact that $F\in c^{d+1}$) to be $O\left(sup\left\{\|H\|_{L(\infty,d+1)},|k|\right\}\right)$ so that $\omega''_R(t)=O\left(sup\left(\|H\|_{L(\infty,d+1)},|k|\right)^2\right)$. A Taylor expansion of ω_R between o and 1 (see proof of theorem 5) shows us that Ω_R is strongly Frechet differentiable.

The symmetry of F_x round $m(x)$ implies $2^{-1}\left[F_x(y)+1-F_x(2m(x)-y)\right]=F_x(y)$. This relation, together with the property of the score $J_R(t)=-J_R(1-t)$ (assumption $(R2)$), shows us that the equation $\int_{-\infty}^{+\infty} J_R\left[2^{-1}\left(F_x(y)+1-F(2\theta-y)\right)\right]dF_x(y)=0$ admits for solution $\theta=m(x)$.

Differentiating[9] with respect to θ the expression $\Omega_R(F,\theta)=\int_{-\infty}^{+\infty} J_R\left[2^{-1}\left(F_x(y)+1-F_x(2\theta-y)\right)\right]dF_xy$, we obtain $\dfrac{\partial\Omega_M}{\partial\theta}(F,m(x))=$

$$\int_{-\infty}^{+\infty} F'_x(2m(x)-y)J\left[2^{-1}\left(F_x(y)+1-F_x(2m(x)-y)\right)\right]F'_x(y)dy$$

$=\int_{-\infty}^{+\infty} J\left[F_x(y)\right]\left(F'_x(y)\right)^2dy$ which is strictly positive from assumption $(R1)$. This strict positivity, together with the Frechet differentiability of the functional Ω_R previously established and the convergence of $\hat{F}_n^Z$ to F^Z for the norm $\|\circ\|_{L(\infty,d+1)}$ (under the assumptions of theorem 5, see Aït Sahalia [1]), allows us to apply the point (i) of theorem 2 : there exists a functional T_R such as $m(x)=T_R(F)$ and such as the equation $\Omega_M(\hat{F}_n,\theta)=0$ asymptotically admits (locally) a unique solution $T_R(\hat{F}_n)$.

Applying the point (ii) of theorem 2, we obtain the Frechet diffentiability of the functional T_R with Frechet differential given by

$$D_{F^Z}T_R.H=\frac{2^{-1}\displaystyle\int_{-\infty}^{+\infty}\left(\int_{-\infty}^{y}\frac{h(x,v)f(x)-f(x,v)h(x)}{f^2(x)}dv-\int_{-\infty}^{2m(x)-y}\frac{h(x,v)f(x)-f(x,v)h(x)}{f^2(x)}dv\right) J'_R\left(2^{-1}\left(\int_{-\infty}^{y}\frac{f(x,v)}{f(x)}dv+1-\int_{-\infty}^{2m(x)-y}\frac{f(x,v)}{f(x)}dv\right)\right)\frac{f(x,y)}{f(x)}dy}{\displaystyle\int_{-\infty}^{+\infty}J\left[F_x(y)\right]\left(F'_x(y)\right)^2dy}$$

[9] differentiation under the integral operator is justified by the regularity properties of F and J_R

$$\frac{\int\limits_{-\infty}^{+\infty} J_R\left(2^{-1}\left(\int\limits_{-\infty}^{y}\frac{f(x,v)}{f(x)}dv+1-\int\limits_{-\infty}^{2m(x)-y}\frac{f(x,v)}{f(x)}dv\right)\right)\frac{h(x,y)f(x)-f(x,y)h(x)}{f^2(x)}dy}{\int\limits_{-\infty}^{+\infty} J\left[F_x(y)\right]\left(F'_x(y)\right)^2 dy}$$

Fortunately, this expression can be conderably simplified. Indeed, if we denote $H_x(y)=\int\limits_{-\infty}^{y}\frac{h(x,v)}{h(x)}dv$, we have

$$\int\limits_{-\infty}^{+\infty}\left[H_x(y)-H_x(2m(x)-y)\right]J'_R\left[2^{-1}\left(F_x(y)+1-F_x(2m(x)-y)\right)\right]\frac{f(x,y)}{f(x)}dy=0$$

$$\int\limits_{-\infty}^{+\infty}\left[F_x(y)-F_x(2m(x)-y)\right]J'_R\left[2^{-1}\left(F_x(y)+1-F_x(2m(x)-y)\right)\right]\frac{f(x,y)}{f(x)}dy=0$$

$\int\limits_{-\infty}^{+\infty} J_R\left[2^{-1}\left(F_x(y)+1-F_x(2m(x)-y)\right)\right]\frac{f(x,y)}{f(x)}dy=0$ (these relations come from the symmetry of F_x round $m(x)$ and from the relation $J_R(t)=-J_R(1-t)$)

Finally, we obtain :

$$D_{F^Z}T_R.H=\frac{\int\limits_{-\infty}^{+\infty} J_R\left[F_x(y)\right]h(x,y)dy}{\int\limits_{-\infty}^{+\infty} J'_R\left[F_x(y)\right]\left(F'_x(y)\right)^2 dy}$$

Using generalized functions, $D_{F^Z}T_R$ can also be written under the form

$$D_{F^Z}T_R.H=\int\limits_{-\infty}^{+\infty}\frac{J_R\left[F_{u_x}(u_y)\right]}{\int\limits_{-\infty}^{+\infty} J'_R\left[F_{u_x}(y)\right]\left(F'_{u_x}(y)\right)^2 dy}\delta_x\left(u_x\right)h(u_x,u_y)du_xdu_y$$

so that $\varkappa^*=d-1$ and $\sqrt{n}h^{\frac{d-1}{2}}\left[T_R(\hat{F}_n^z)-T_R(F^Z)\right]\Longrightarrow N(0,V_{T_R}(F^Z))$ with

$$V_{T_R}(F^Z)=\int\left(\frac{J_R\left[F_{u_x}(u_y)\right]}{\int\limits_{-\infty}^{+\infty} J'_R\left[F_x(y)\right]\left(F'_x(y)\right)^2 dy}\right)^2 f(x,u_y)du_y.\int K^2(x)dx$$

A.8 proof of corollary 10

We have

$$SIF(x_i, y_i) = D_F z T_R. \left(K_I \left(\frac{\circ - x_i}{h} \right) K_I \left(\frac{\circ - y_i}{h} \right) - F \right)$$

i.e.

$$SIF(x_i, y_i) = \frac{\int_{-\infty}^{+\infty} J_R \left[F_x(y)\right] \frac{1}{h} K \left(\frac{x - x_i}{h} \right) \frac{1}{h} K \left(\frac{y - y_i}{h} \right) dy - \int_{-\infty}^{+\infty} J_R \left[F_x(y)\right] f(x,y) dy}{\int_{-\infty}^{+\infty} J_R' \left[F_x(y)\right] \left(F_x'(y)\right)^2 dy}$$

with $\int_{-\infty}^{+\infty} J_R \left[F_x(y)\right] f(x,y) dy = f(x) \int_{-\infty}^{+\infty} J_R \left[F_x(y)\right] dF_x(y) = 0$, we obtain :

$$SIF(x_i, y_i) = \frac{\frac{1}{h} K \left(\frac{x - x_i}{h} \right) \int_{-\infty}^{+\infty} J_R \left[F_x(y)\right] \frac{1}{h} K \left(\frac{y - y_i}{h} \right) dy}{\int_{-\infty}^{+\infty} J_R' \left[F_x(y)\right] \left(F_x'(y)\right)^2 dy}$$

and with the change of variable $u = \frac{y - y_i}{h}$, we finally get

$$SIF(x_i, y_i) = \frac{\frac{1}{h} K \left(\frac{x - x_i}{h} \right) \int_{-\infty}^{+\infty} J_R \left[F_x(hu + y_i)\right] K(u) \, du}{\int_{-\infty}^{+\infty} J_R' \left[F_x(y)\right] \left(F_x'(y)\right)^2 dy}$$

References

[1] Aït-Sahalia, Y (1995) : "The Delta Method for Nonlinear Kernel Functionals", University of Chicago, USA.

[2] Aït-Sahalia, Y & Bickel, P.J. & Stocker, T.M (2000) : "Goodness of fit tests for kernel regression with an application to option implied volatilities" forthcoming in journal of econometrics.

[3] Barnett, V. & T. Lewis (1979) : *"Outliers in statistical data"*, Wiley, New York.

[4] Boente, G. & R. Fraiman (1991) : "A functional approach to robust non-parametric regression" in Directions in Robusts Statistics and Diagnostics, W. Stahel & S. Weisberg (eds), IMA series 33, p 33-46, Springer Verlag, New-York.

[5] Boente, G. & R. Fraiman (1994) : "Local L-estimators for nonparametric regression under dependence", Journal of nonparametric statistics, vol 4, p 91-101.

[6] Cheng, K.F. & P.E. Cheng (1990) : "Asymptotic normality for robust R-estimators of regression function", Journal of statistical planning and inference, vol 24, p 137-149.

[7] Davis, R.A. & U. Gather (1993) : "The identification of multiple outliers", Journal of the american statistical association, Vol 88, p 782-801.

[8] Hampel, F.R. (1994) : "The influence curve and its role in robust estimation", Journal of the american statistical association, Vol 62, p 1179-118.

[9] Härdle, W. (1984) : "Robust regression function estimation",Journal of multivariate analysis, vol 14, p 169-180.

[10] Härdle, W (1990) : "*Applied nonparametric regression*", Cambridge University press.

[11] Huber, C. (1981) : "Voisinages d'un modèle", in "théorie de la robustesse et estimation d'un paramètre", Société Mathématique de France, col Astérisque Vol 43-44, p11-21.

[12] Huber, P.J. (1981) : "*Robust statistics*", Wiley, New-York.

[13] Lucas, A. (1996) : "Outlier robust unit root analysis", Ph D Thesis, Rotterdam, Tinbergen Institute.

[14] Nadaraya, E.A. (1964) : "On estimating regression", Theory. Prob. App. Vol 10, p 186-190.

[15] Protopopescu, C. (1995) : "Estimation fonctionnelle des modèles issus de la théorie des jeux", Ph D Thesis, Gremaq, Toulouse, France.

[16] Schwartz, L (1966) : "*Théorie des distributions*", Hermann, Paris.

[17] Vanhems, A. (2001) : "Etude non-paramétrique d'équations différentielles bruitées : application à la théorie du consommateur représentatif", Ph D Thesis, Gremaq, Toulouse, France.

[18] Von Mises, R. (1947) : "On the asymptotic distribution of differentiable statistical functions", Annals of Mathematical Statistics, Vol 18, p 309-348.

[19] Watson, G.S. (1964) : "Smooth regression analysis", Sankhya , series A, vol 26, p359-372.

[20] Zeidler, E. (1986) : *"Nonlinear functional analysis and its applications"*, Springer Verlag, New York.