

Asymptotic normality of kernel estimator of ψ -regression function for functional ergodic data

Fatima Benziadi, Abdelkader Gheriballah and Ali Laksaci

Agence Thmatique de Recherche en Sciences et Technologie, Laboratoire de statistique et processus stochastiques, Univ. Djillali Liabès, BP 89, 22000 Sidi Bel Abbès, Algeria

Received: 16 July 2015, Revised: 30 September 2015, Accepted: 15 October 2015

Published online: 21 February 2016.

Abstract: In this paper we consider the problem of the estimation of the ψ -regression function when the covariates take values in an infinite dimensional space. Our main aim is to establish, under a stationary ergodic process assumption, the asymptotic normality of this estimate.

Keywords: Kernel estimate, ergodic data, functional data, robust regression.

1 Introduction

The statistical analysis of functional data has received much attention in the last few years. There are several special issues dedicated to this topic by various statistical journals (see, for instance Davidian et al. (2004), Gonzalez-Manteiga and Vieu (2007), Valderrama (2007) or Ferraty (2010)). This particular interest of this topic is due to the diversity of application's fields in which the data to be treated are curves. Indeed, it is well documented that this kind of data can be found, for instance, in chemometrics, environmetrics, sciences medical, speech recognition, economics. The main purpose of this contribution is to study the asymptotic properties of the nonparametric robust regression, under less restrictive dependence conditions that is the ergodicity assumption.

The nonparametric robust analysis of functional data is a very recent field of investigations. The first results in this subject was given by Azzedine *et al.* (2008). Since this work, a considerable corpus of research has been dedicated to the robust nonparametric modeling in functional data analysis. This great consideration is motivated by the fact that the robust nonparametric regression provides an alternative approach to the classical methods which is insensitive to the presence of outliers or heteroskedastic variables. Considering a sample of i.i.d, Azzedine *et al.* (2008) have established the almost complete convergence of the robust estimators of the regression function when the regressors is functional. In the same context, Attouch *et al.* (2009) studied the asymptotic normality of this model. The robust analysis in functional time series, has been investigated by many authors. We refer, for example, to Crambes *et al.* (2008) for the convergence in L^q norm, Attouch *et al.* (2010) for the asymptotic normality and Chen and Zhang (2009) for the weak and strong consistency of the nonparametric functional conditional location estimate in the mixing case.

While in all these literatures, the functional time series data is modeled by using the mixing conditions, our main aim in this paper is to consider the problem of the robust nonparametric analysis of functional ergodic time series data. This dependence structure covers several case does not satisfy the usual mixing structures. Moreover, the ergodic framework

avoid the widely used strong mixing condition and its variants to measure the dependency and the very involved probabilistic calculations that it implies. In addition, from practical point of view the ergodicity condition is one of a principal postulate of statistical physics. It models the thermodynamic properties of gases, atoms, electrons or plasmas. This hypothesis is also used in signal processing, for studying the evolution of a random signal. Despite this importance in applications, the litterateur in this functional dependency is still limited. This problem has been initiated by Laib and Louani (2010, 2011). They consider the problem of functional estimation for nonparametric regression operators under the ergodicity condition. Recently Gheiriballah et al (2013) gave the almost complete convergence (with rate) of a family of robust nonparametric estimators for regression function. More recently Benziadi et al. (2014) stated the almost complete convergence of a recursive kernel estimates of the conditionals quantiles.

Our main goal in this paper is to establish the asymptotic normality of the estimator proposed by Gheiriballah et al (2013). This results is obtained under, some general condition. We recall that the asymptotic normality is a fundamental preliminary result to build confidence interval of our model or to precis the asymptotic dominant terms of the moments of order q by explicating the asymptotic bias and the asymptotic variance which are a basic ingredients of this error. All these derivatives of our asymptotic results has been discussed in this paper. It should be noted that our hypotheses and results unify the both cases of finite or infinite dimension of the regressors. Thus, we can say that our result it also even in vectorial statistic.

The setup of this paper is as follows: We present our model in Section 2. Section 3 is dedicated to fixing notations and hypotheses. Our main result is given in Section 4. In Section 5 we show the generality of our model and the flexibility of our conditions by studying some particular case. In Section 6, we discuss the importance of the asymptotic normality property by giving some direct applications. The proofs of the auxiliary results are relegated to the Appendix.

2 The ψ - regression model and its estimate

Let $Z_i = (X_i, Y_i)_{i=1, \dots, n}$ be a $\mathcal{F} \times \mathbb{R}$ -valued measurable strictly stationary process, defined on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$, where \mathcal{F} is a semi-metric space, d denoting the semi-metric. For any x in \mathcal{F} , we consider ψ_x a real-valued Borel function satisfying some regularity conditions to be stated below. The nonparametric model studied in this paper, denoted by θ_x , is implicitly defined as a zero with respect to (w.r.t.) t of the following equation

$$\Psi(x, t) := \mathbb{E}[\psi_x(Y_i, t) | X_i = x] = 0. \quad (1)$$

We suppose that, for all $x \in \mathcal{F}$, θ_x exists and is the unique zero w.r.t. t of (1) (see, for instance Boente and Fraiman (1989) or Koul and Stute (1998) for the existence and uniqueness of θ_x). We point out that this robustification method belongs to the class of M-estimates introduced by Huber (1964) and it covers and includes many important nonparametric models, for example, $\psi_x(y, t) = (y - t)$ yields the classical regression, $\psi_x(y, t) = \mathbb{I}_{y \geq t} - \mathbb{I}_{y < t}$ leads to the conditional median function $m(x) = \text{med}(Y|X = x)$ and the α^{th} conditional quantile is obtained by setting $\psi_x(y, t) = \mathbb{I}_{y > t} - (1 - \alpha)$, $\alpha \in (0, 1)$. In addition, our robustification method allows us to consider the functional nonparametric regression model with a scale of the error assumed to be known by taking $\psi_x(\cdot, \cdot) = \psi(\cdot - \cdot / \sigma(x))$, where $\sigma(\cdot)$ 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 .

For all $(x, t) \in \mathcal{F} \times \mathbb{R}$, we propose a nonparametric estimator of $\Psi(x, t)$ given by

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

where K is a kernel and $h = h_n$ is a sequence of positive real numbers. A natural estimator $\widehat{\theta}_x$ of θ_x is a zero w.r.t. t of the equation

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

Obviously, when $\psi_x(Y, t) = Y - t$, then $\widehat{\theta}_n$ is the estimator given in Ferraty and Vieu (2006) for the functional nonparametric regression. While for $\psi_x(y, t) = \mathbb{I}_{y>t} - (1 - \alpha)$, we obtain the α^{th} conditional quantile estimate studied by Laksaci et al. (2009).

In this work, we will assume that the underlying process Z_i is functional stationary ergodic (see Laib and Louani (2011) for the definition and some examples). Of course, this work includes the finite dimensional case ($\mathcal{F} = \mathbb{R}^p$) but its importance is due to the fact that it covers also the infinite dimensional case. Because, these questions in infinite dimension are particularly interesting, not only for the fundamental problems they formulate, but also for many applications they may allow, see Bosq (2000), Ramsay and Silverman (2005) and Ferraty and Vieu (2006).

3 Notations, hypotheses and comments

All along the paper, when no confusion is possible, we will denote by C and C' some strictly positive generic constants, x is a fixed point in \mathcal{F} and N_x denote a fixed neighborhood of x . For $r > 0$, let $B(x, r) := \{x' \in \mathcal{F} / d(x', x) < r\}$. Moreover, for $i = 1, \dots, n$, we put \mathfrak{F}_k is the σ -field generated by $((X_1, Y_1), \dots, (X_k, Y_k))$ and we pose \mathfrak{G}_k is the σ -field generated by $((X_1, Y_1), \dots, (X_k, Y_k), X_{k+1})$. In order to establish our asymptotic results we need the following hypotheses:

(H1) The processes $(X_i, Y_i)_{i \in \mathbb{N}}$ satisfies:

- (i) The functions $\phi(x, r) := \mathbb{P}(X \in B(x, r)) > 0$, and $\phi_i(x, r) = \mathbb{P}(X_i \in B(x, r) | \mathfrak{F}_{i-1}) > 0 \forall r > 0$.
- (ii) For all $r > 0$, $\frac{1}{n\phi(x, r)} \sum_{i=1}^n \phi_i(x, r) \rightarrow^p 1$ and $n\phi(x, h) \rightarrow \infty$ as $h \rightarrow 0$.

(H2) The functions Ψ such that:

- (i) The function $\Psi(x, \cdot)$ is of class \mathcal{C}^1 at N_x a fixed neighborhood of θ_x .
- (ii) For each fixed t in N_x the functions $\Psi(\cdot, t)$ and $\lambda_2(\cdot, t) = \mathbb{E}[\psi_x^2(Y, t) | X = \cdot]$ are continuous at the point x .
- (iii) The derivative of the real function

$$\Phi(s, z) = \mathbb{E}[\Psi(X_1, z) - \Psi(x, z) | d(x, X_1) = s]$$

exists at $s = 0$ and is continuous w.r.t. the second component at N_x .

(H3) For each fixed t in the neighborhood of $\theta_x \forall j \geq 2$,

$$\mathbb{E}[\psi_x^j(Y, t) | \mathfrak{G}_{i-1}] = \mathbb{E}[\psi_x^j(Y, t) | X_i] < C < \infty, a.s.,$$

(H4) The function ψ_x is continuous and monotone w.r.t. the second component.

(H5) The kernel K is a positive function supported on $(0, 1[$. Its derivative K' exists on $(0, 1)$ and satisfies $K'(t) < 0$ for $0 < t < 1$.

(H6) There exists a function $\tau_x(\cdot)$ such that

$$\forall t \in [0, 1] \quad \lim_{h \rightarrow 0} \frac{\phi(x, th)}{\phi(x, h)} = \tau_x(t),$$

$$K^2(1) - \int_0^1 (K^2(u))' \tau_x(u) du > 0 \quad \text{and} \quad K(1) - \int_0^1 K'(u) \tau_x(u) du \neq 0.$$

Comments on the hypotheses.

Our assumptions are quite mild. Indeed, the ergodicity of functional data: The latter is exploited together with condition (H1 (ii)) which is less restrictive to the conditions imposed by Laib and Louani (2011) (see Gheriballah et al. (2013) for more discussion). In this work, the functional space of our model is characterized by the regularity condition (H2iii). This condition replaces the Lipschitz condition usually assumed in nonparametric functional data analysis. This change is useful in order to explicitly asymptotically the bias term. However, the Lipschitz condition gives inexact/inaccurate asymptotic bias term which is not interesting for the asymptotic normality. The robustness property is controlled by (H4) where only the convexity (which is a fundamental constraint of the robustness properties of the M-estimators) of the score function is needed. In order to cover the classical regression studied in this ergodic functional context by Laib and Louani (2011) we establish our asymptotic normality without the boundedness condition for the score function. Condition (H3), (H5) and (H6) are very similar to those used by Ferraty et al. (2007). Moreover, the function $\tau_x(\cdot)$ defined in (H5) plays a fundamental role in the asymptotic normality result. It permits to give the variance term explicitly. Note that this function can be specified in several situations where the function $\phi(x, h)$ is known and (H5) is fulfilled. We quote the following cases (which can be found in Ferraty *et al.* (2007)).

- (i) $\phi(x, h) = C_x h^\gamma + o(h^\gamma)$ for some $\gamma > 0$ with $\tau_x(s) = s^\gamma$,
- (ii) $\phi(x, h) = C_x h^{\gamma_1} \exp\{-Ch^{-\gamma_2}\} + o(h^{\gamma_1} \exp\{-Ch^{-\gamma_2}\})$ for some $\gamma_1 > 0$ and $\gamma_2 > 0$ with $\tau_x(\cdot)$ being Dirac's function in 1.

4 Results

Our main result is given in the following theorem

Theorem 1. Assume that (H1)-(H6) hold, then $\hat{\theta}_x$ exists and is unique with great probability and for any $x \in \mathcal{A}$, we have

$$\left(\frac{n\phi(x, h)}{\sigma^2(x, \theta_x)} \right)^{1/2} \left(\hat{\theta}_x - \theta_x - B_n(x) \right) \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1) \quad \text{as } n \rightarrow \infty$$

where

$$B_n(x) = h\Phi'(0, \theta_x) \frac{\beta_0}{\beta_1} + o(h) \quad \text{and} \quad \sigma^2(x, \theta_x) = \frac{\beta_2 \lambda_2(x, \theta_x)}{\beta_1^2 (\Gamma_1(x, \theta_x))^2}$$

with

$$\beta_0 = - \int_0^1 (sK(s))' \beta_x(s) ds, \quad \beta_j = - \int_0^1 (K^j)'(s) \beta_x(s) ds, \quad \text{for } j = 1, 2,$$

$$\Gamma_1(x, \theta_x) = \frac{\partial}{\partial t} \Psi(x, \theta_x) \quad \text{and} \quad \mathcal{A} = \{x \in \mathcal{F}, \lambda_2(x, \theta_x) \Gamma_1(x, \theta_x) \neq 0\}$$

and $\xrightarrow{\mathcal{D}}$ means the convergence in distribution.

In order to remove the bias term $B_n(x)$, we need an additional condition on the bandwidth parameter h .

Corollary 1. Under the hypotheses of Theorem 1 and if the bandwidth parameter h satisfies $nh^2\phi(x, h) \rightarrow 0$ as $n \rightarrow \infty$, we have

$$\left(\frac{n\phi(x, h)}{\sigma^2(x, \theta_x)}\right)^{1/2} (\hat{\theta}_x - \theta_x) \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1) \quad \text{as } n \rightarrow \infty.$$

Proof. (Proof of Theorem 1 and Corollary 1). We give the proof for the case of an increasing ψ_x , decreasing case being obtained by considering $-\psi_x$. In this case, we define, for all $u \in \mathbb{R}$, $z = \theta_x - B_n(x) + u[n\phi(x, h)]^{-1/2} \sigma(x, \theta_x)$. Let us remark that,

$$\mathbb{P} \left\{ \left(\frac{n\phi(x, h)}{\sigma^2(x, \theta_x)}\right)^{1/2} (\hat{\theta}_x - \theta_x + B_n(x)) < u \right\} = \mathbb{P} \left\{ \hat{\theta}_x < \theta_x - B_n(x) + u[n\phi(x, h)]^{-1/2} \sigma(x, \theta_x) \right\} = \mathbb{P} \left\{ 0 < \hat{\Psi}(x, z) \right\}.$$

It is clear that we can write

$$\hat{\Psi}(x, t) = B_n(x, t) + \frac{R_n(x, t)}{\hat{\Psi}_D(x)} + \frac{Q_n(x, t)}{\hat{\Psi}_D(x)}$$

where

$$Q_n(x, t) := (\hat{\Psi}_N(x, t) - \bar{\Psi}_N(x, t)) - \Psi(x, t)(\hat{\Psi}_D(x) - \bar{\Psi}_D(x))$$

$$B_n(x, t) := \frac{\bar{\Psi}_N(x, t)}{\bar{\Psi}_D(x)}, \quad \text{and} \quad R_n(x, t) := - \left(\frac{\bar{\Psi}_N(x, t)}{\bar{\Psi}_D(x)} - \Psi(x, t) \right) (\hat{\Psi}_N(x, t) - \bar{\Psi}_N(x, t))$$

with

$$\hat{\Psi}_N(x, t) := \frac{1}{n\mathbb{E}[K(h^{-1}d(x, X_1))]} \sum_{i=1}^n K(h^{-1}d(x, X_i)) \psi_x(Y_i, t),$$

$$\bar{\Psi}_N(x, t) := \frac{1}{n\mathbb{E}[K(h^{-1}d(x, X_1))]} \sum_{i=1}^n \mathbb{E} [K(h^{-1}d(x, X_i)) \psi_x(Y_i, t) | \mathfrak{F}_{i-1}],$$

$$\hat{\Psi}_D(x) := \frac{1}{n\mathbb{E}[K(h^{-1}d(x, X_1))]} \sum_{i=1}^n K(h^{-1}d(x, X_i)),$$

$$\bar{\Psi}_D(x) := \frac{1}{n\mathbb{E}[K(h^{-1}d(x, X_1))]} \sum_{i=1}^n \mathbb{E} [K(h^{-1}d(x, X_i)) | \mathfrak{F}_{i-1}].$$

It follows that

$$\mathbb{P} \left\{ \left(\frac{n\phi(x, h)}{\sigma^2(x, \theta_x)}\right)^{1/2} (\hat{\theta}_x - \theta_x + B_n(x)) < u \right\} = \mathbb{P} \left\{ -\hat{\Psi}_D(x)B_n(x, z) - R_n(x, z) < Q_n(x, z) \right\}.$$

Therefore, our main result is a consequence of the following intermediates results.

Lemma 1. Under the hypotheses of Theorem 1, we have for any $x \in \mathcal{A}$

$$\left(\frac{n\phi(x, h)\beta_1^2}{\beta_2\lambda_2(x, \theta_x)}\right)^{1/2} Q_n(x, z) \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1) \quad \text{as } n \rightarrow \infty.$$

Lemma 2. (see, Laib and Louani, 2010) Under Hypotheses (H1) and (H4)-(H6), we have,

$$\hat{\Psi}_D(x) - 1 = o_P(1).$$

Lemma 3. Under hypotheses (H1), (H2), and (H4)-(H6) we have

$$\left(\frac{n\phi(x, h)\beta_1^2}{\beta_2\lambda_2(x, \theta_x)}\right)^{1/2} B_n(x, z) = u + o(1), \quad \text{as } n \rightarrow +\infty.$$

Lemma 4. Under hypotheses (H1), (H2), and (H4)-(H6) we have,

$$\left(\frac{n\phi(x, h)\beta_1^2}{\beta_2\lambda_2(x, \theta_x)} \right)^{1/2} R_n(x, z) = o_P(1) \quad a.co.$$

Lemma 5. Under Hypotheses (H1) and (H4)-(H6), $\hat{\theta}_x$ exists a.s. for all sufficiently large n

5 Some special cases

In this section we discuss the generality of our study by comparing it to some popular case. More precisely, we consider three special cases such as the classical regression case, the independent case and the multivariate case which are respectively studied, Laib and Louani(2010), Attouch *et al.* (2009).

- (i) *The classical regression case:* As noticed earlier the classical regression defined by conditional expectation is a particular case of our study with $\psi_x(Y, t) = (Y - t)$.

So, for this particular case we obtain the following convergence rate.

Corollary 2. Under the hypotheses (H1)-(H6), we have

$$\left(\frac{n\phi(x, h)}{\sigma^2(x, \theta_x)} \right)^{1/2} \left(\hat{\theta}_x - \theta_x - B_n(x) \right) \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1) \quad \text{as } n \rightarrow \infty$$

where

$$B_n(x) = h\Phi'(0, \theta_x) \frac{\beta_0}{\beta_1} + o(h) \quad \text{and} \quad \sigma^2(x, \theta_x) = \frac{\beta_2(\mathbb{E}[Y^2|X=x] - \mathbb{E}^2[Y|X=x])}{\beta_1^2}$$

Remark. Clearly, this convergence rate is exactly what is obtained by Laib and Louani (2011) for the standard regression model.

- (i) *The independent case:* In this situation, condition (H1(ii)) is automatically verified and for all $i = 1, \dots, n$ take $\phi_i(x, r) = \phi(x, r)$.

Therefore, condition (H1) is restricted to $\phi(x, r) > 0$, for all $r > 0$. Thus, our Theorem leads to the next Corollary.

Corollary 3. Under assumptions (H1)-(H6) we have:

$$\left(\frac{n\phi(x, h)}{\sigma^2(x, \theta_x)} \right)^{1/2} \left(\hat{\theta}_x - \theta_x - B_n(x) \right) \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1) \quad \text{as } n \rightarrow \infty$$

Remark. We point out that in this case where the (X_i, Y_i) are independent, we obtain the same convergence rate given by Attouch *et al.* (2009).

- (i) *The real case* As mentioned in the introduction on, in the real case when $\mathcal{F} = \mathbb{R}$, and if the probability density of the random variable X (resp. the conditional density of X given \mathfrak{F}_{i-1}) denoted by f (resp. by $f_i^{\mathfrak{F}_{i-1}}$), is strictly positive and of \mathcal{C}^1 class, then $\phi(x, h) = f(x)h + o(h)$ and $BBP(X_i \in [x-h, x+h] | \mathfrak{F}_{i-1}) = f_i^{\mathfrak{F}_{i-1}}(x)h + o(h)$. Moreover the ergodic Theorem insure that

$$\left\| \frac{1}{n} \sum_{i=1}^n f_i^{\mathfrak{F}_{i-1}} - f \right\| \rightarrow 0,$$

where $\|\cdot\|$ is a norm in separable Banach space \mathcal{C}^1 .

Therefore condition (H1) is verified and Theorem ?? can be reformulated in the following way.

Corollary 4. *Under assumptions (H2)-(H6) and if the probability density of the random variable X (resp. the conditional density of X given \mathfrak{F}_{i-1}) denoted by f (resp. by $f_i^{\mathfrak{F}_{i-1}}$), is strictly positive and of \mathcal{C}^1 class, we have:*

$$\left(\frac{n\phi(x, h)}{\sigma^2(x, \theta_x)} \right)^{1/2} \left(\hat{\theta}_x - \theta_x - B_n(x) \right) \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1) \quad \text{as } n \rightarrow \infty$$

Remark. A similar thing can be concluded if $\mathcal{F} = \mathbb{R}^p$. It is worth to noting that this last consistency is also new in vectorial statistic

6 Some applications

6.1 Conditional Confidence curve

The most important application of the asymptotic normality result is the building of confidence intervals for the true value of θ_x given curve $X = x$. However, the latter requires an estimation of the bias $B_n(x)$ term and the standard deviation $\sigma(x, \theta_x)$. For sake of shortness, we neglect the bias term and we estimate $\sigma(x, \theta_x)$ by plug-in method as follows. Indeed, if ψ_x is of class C^1 , w.r.t the second component, the quantities $\lambda_2(x, \theta_x)$ and $\Gamma_1(x, \theta_x)$ can be estimated by

$$\hat{\lambda}_2(x, \hat{\theta}_x) = \frac{\sum_{i=1}^n K(h^{-1}d(x, X_i)) \psi_x^2(Y_i, \hat{\theta}_x)}{\sum_{i=1}^n K(h^{-1}d(x, X_i))},$$

and

$$\hat{\Gamma}_1(x, \hat{\theta}_x) = \frac{\sum_{i=1}^n K(h^{-1}d(x, X_i)) \frac{\partial}{\partial t} \psi_x(Y_i, \hat{\theta}_x)}{\sum_{i=1}^n K(h^{-1}d(x, X_i))}.$$

Furthermore, the quantities β_1 and β_2 can be estimated empirically by

$$\hat{\beta}_1 = \frac{1}{n\phi(x, h)} \sum_{i=1}^n K(h^{-1}d(x, X_i)) \quad \text{and} \quad \hat{\beta}_2 = \frac{1}{n\phi(x, h)} \sum_{i=1}^n K^2(h^{-1}d(x, X_i))$$

It follows that $\hat{\sigma}(x, \hat{\theta}_x) := \left(\frac{\hat{\beta}_2 \hat{\lambda}_2(x, \hat{\theta}_x)}{(\hat{\beta}_1)^2 \hat{\Gamma}_1^2(x, \hat{\theta}_x)} \right)^{1/2}$. Thus, we get the following approximate $(1 - \zeta)$ confidence interval for θ_x

$$\hat{\theta}_x \pm t_{1-\zeta/2} \times \left(\frac{\hat{\sigma}_n^2(x, \hat{\theta}_x)}{n\phi(x, h)} \right)^{1/2}$$

where $t_{1-\zeta/2}$ denotes the $1 - \zeta/2$ quantile of the standard normal distribution. It should be to note that the function $\phi(x, \cdot)$ does not appear in the calculation of the confidence interval by simplification.

6.2 Smoothing parameter selection

As all nonparametric estimation with the smoothing technique, the choice of the smoothing parameters, plays a primordial role. The most celebrated criterium for the smoothing parameter choice is the mean squared error where the optimal one have to balance the bias and variance terms. However, in our robust approach, this L_2 error is not adequate if we suspect

that outliers are present. In this context, the L_1 error is more adapted to this case. For both criteria, our asymptotic result is a basic ingredient to determine the leading term in these errors. Indeed, by using the same arguments as those used by Crambes et al. (2008) we show that

$$\mathbb{E} \left[\left| \widehat{\theta}_n - \theta_x \right|^q \right] = \mathbb{E} \left[\left| B_n(x) + \sqrt{\frac{\sigma(x, \theta_x)}{n\phi(x, h)}} W \right|^q \right] + o \left(\frac{1}{\sqrt{n\phi(x, h)^q}} \right)$$

where W is a standard gaussian variable. Therefore, an ideal theoretical rule permitting to obtain an optimal bandwidth, for both (robust or no robust case) is the minimization of the leading term:

$$\mathbb{E} \left[\left| B_n(x) + \sqrt{\frac{\sigma(x, \theta_x)}{n\phi(x, h)}} W \right|^q \right].$$

Of course that the practical utilization of this method requires some additional computational efforts. More precisely, it requires the estimation of the unknown quantities $B_n(x)$ and $\sigma(x, \theta_x)$. The latter is estimated by $\widehat{\sigma}(x, \widehat{\theta}_x)$ while $B_n(x)$ can be estimated by the same fashion where the real function $\Phi(t, s)$ is treated as a real regression function with response variable $\Psi(X, t) - \Psi(x, t)$. In conclusion, we can say that, the practical utilization of the present approach is possible, but it needs the determination of a pilot estimator of the conditional density function.

6.3 The functional times series prediction

It is well known that one of the most application of the nonparametric functional data analysis in dependent data is the prediction of a future real characteristic of an continuous processus by taking into account the whole past continuously. Indeed, let $(Z_t)_{t \in [0, b[}$ be a continuous time real valued random process. From Z_t we may construct N functional random variables $(X_i)_{i=1, \dots, N}$ defined by:

$$\forall t \in [0, b[, \quad X_i(t) = Z_{((i-1)b+t)/N}.$$

The prediction aims is to evaluate a real characteristic denoted Y given X_N . The definition (1) shows that the random variable $\widehat{\theta}_{X_N}$, defined by (2), is the best approximation of this characteristic with respect to the loss function $\rho_x(\cdot, t) = \int_0^t \psi_x(\cdot, s) ds$, where (2) is given by using the $N - 1$ pairs of r.v $(X_i, G(X_{i+1}))_{i=1, \dots, N-1}$ with G is the function which describes this characteristic.

Acknowledgements

The authors greatly thank an Associate Editor and an anonymous referee for a careful reading of the paper. The authors also thank the Agence Thmatique de Recherche en Sciences et Technologie for their financial support

References

- [1] Attouch, M. Laksaci, A. and Ould Saïd E (2009). Asymptotic distribution of robust estimator for functional nonparametric models. *Comm. Statist. Theory and Methods* **38**, 1317-1335.
- [2] Attouch, M. Laksaci, A. and Ould Saïd E. (2010). Asymptotic normality of a robust estimator of the regression function for functional time series data. *J. of the Korean Statist. Soc.*, **39**, 489-500.

[3] Azzedine, N. Laksaci, A. and Ould Saïd E. (2008). On the robust nonparametric regression estimation for functional regressor. *Statist. Probab. Lett.* **78**, 3216-3221. Boente, G. and Fraiman, R. (1989). Nonparametric regression estimation. *J. Multivariate Anal.* **29**, 180-198.

[4] Bosq, D. (2000). Linear processs in function spaces. Théorie and Application. *Lectures Notes in Statistics. Springer Verlag, New York.* **149**.

[5] Crambes, C. Delsol, L. and Laksaci, A. (2008). Robust nonparametric estimation for functional data. *J. of Nonparametric Statistics* **20**, 573-598.

[6] Davidian, M. Lin, X. and Wang, J.L. (2004). Introduction. *Statist. Sinica* **14**, 613–614.

[7] Ferraty, F. and Vieu, P. (2006). *Nonparametric functional data analysis Theory and Practice.* Springer, New York.

[8] Ferraty, F. (2010). High-dimensional data: a fascinating statistical challenge. *J. Multivariate Anal.* **101**, 305–306.

[9] Ferraty, F. and Romain, Y. (2011). *The Oxford handbook of functional data analysis.* Oxford University Press.

[10] Gaenssler, P. and Haeusler, E., 1986. *On martingale central limit theory*, in Eberlein, E., Taqqu, M. S. (Eds), *Dependence in Probability and Statistics*, Birkhauser, Boston, pp. 303-334.

[11] Gonzalez Manteiga, W. and Vieu, P. (2007). Statistics for functional data. *Comput. Statist. Data Anal.* **51**, 4788–4792.

[12] Györfi, L., Härdle, W., Sarda, P. and Vieu, P. (1989). *Nonparametric curve estimation for time series.* Lecture Notes in Statistics, **60**, Springer-Verlag. Huber, P. J. (1964). Robust estimation of a location parameter. *Ann. Math. Statist.* **35**, 73-101.

[13] Laksaci, A. Lemdani, M. and Ould-Saïd, E. (2009). L^1 -norm kernel estimator of conditional quantile for functional regressors: consistency and asymptotic normality. *Statist. Probab. Lett.* **79**, 1065-1073.

[14] Laïb, N. and Ould Saïd, E. (2000). A robust nonparametric estimation of the autoregression function under an ergodic hypothesis. *Canad. J. Statist.* **28**, 817-828.

[15] b, N. and Louani, D. (2010). Nonparametric kernel regression estimate for functional stationary ergodic data: Asymptotic properties. *J. Multivariate Anal.*, **101**, 2266-2281.

[16] Laïb, N. and Louani, D. (2011). Strong consistency of the regression function estimator for functional stationary ergodic data. *J. Statist. Plan. and Inference*, **141**, 359-372.

[17] Ramsay, J. O. and Silverman, B. W. (2005). *Functional Data Analysis, 2nd Edition*, Springer, New-York.

[18] Stone C J, (2005). Nonparametric M-regression with free knot splines. *J. Statist. Plan. and Inference*, **130**, 183-206.

[19] Valderrama, M. J. (2007). An overview to modelling functional data. *Comput. Statist.* **22**, 331–334.

7 Appendix

Proof. (Proof of Lemma 1). For all $i = 1, \dots, n$ we put $K_i(x) = K(h^{-1}d(x, X_i))$ and

$$\eta_{ni} = \left(\frac{\phi(x, h)\beta_1^2}{\beta_2\lambda_2(x, \theta_x)} \right)^{1/2} (\psi_x(Y_i, z) - \Psi(x, z)) \frac{K_i(x)}{\mathbb{E}K_1(x)} \tag{2}$$

and define $\zeta_{ni} := \eta_{ni} - \mathbb{E}[\eta_{ni} | \mathcal{F}_{i-1}]$. Under this consideration we write

$$\left(\frac{n\phi(x, h)\beta_1^2}{\beta_2\lambda_2(x, \theta_x)} \right)^{1/2} Q_n(x, z) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \zeta_{ni}$$

As ζ_{ni} is a triangular array of martingale differences according the σ -fields $(\mathcal{F}_{i-1})_i$, we are in position to apply the central limit theorem based on unconditional Lindeberg condition (see, Gaenssler et al. (1978)). More precisely, we must to check the following condition

$$\frac{1}{n} \sum_{i=1}^n \mathbb{E}[\zeta_{ni}^2 | \mathcal{F}_{i-1}] \rightarrow 1 \quad \text{in probability} \tag{3}$$

and

$$\text{for every } \varepsilon > 0 \quad \frac{1}{n} \sum_{i=1}^n \mathbb{E}[\zeta_{ni}^2 \mathbb{I}_{\zeta_{ni}^2 > \varepsilon n}] \rightarrow 0. \tag{4}$$

Firstly, we prove (3). To do that, we write

$$\mathbb{E}[\zeta_{ni}^2 | \mathfrak{F}_{i-1}] = \mathbb{E}[\eta_{ni}^2 | \mathfrak{F}_{i-1}] - \mathbb{E}^2[\eta_{ni} | \mathfrak{F}_{i-1}].$$

Therefore, it suffices to show the

$$\frac{1}{n} \sum_{i=1}^n \mathbb{E}^2[\eta_{ni} | \mathfrak{F}_{i-1}] \xrightarrow{\mathbb{P}} 0 \tag{5}$$

and

$$\frac{1}{n} \sum_{i=1}^n \mathbb{E}[\eta_{ni}^2 | \mathfrak{F}_{i-1}] \xrightarrow{\mathbb{P}} 1. \tag{6}$$

For the first convergence we have

$$\begin{aligned} |\mathbb{E}[\eta_{ni} | \mathfrak{F}_{i-1}]| &= \frac{1}{\mathbb{E}K_1(x)} \left(\frac{\phi(x, h)\beta_1^2}{\beta_2\lambda_2(x, \theta_x)} \right)^{1/2} |\mathbb{E}[(\Psi(X_i, t) - \Psi(x, t))K_i(x) | F_{i-1}]| \\ &\leq \frac{1}{\mathbb{E}K_1(x)} \left(\left(\frac{\phi(x, h)\beta_1^2}{\beta_2\lambda_2(x, \theta_x)} \right) \right)^{1/2} \sup_{u \in B(x, h)} |\Psi(u, t) - \Psi(x, t)| \mathbb{E}[K_i(x) | F_{i-1}]. \end{aligned}$$

Obviously, under (H1) and (H5) we have

$$C\phi_i(x, h) \leq \mathbb{E}[K_i(x) | \mathfrak{F}_{i-1}] \leq C'\phi_i(x, h).$$

and

$$C\phi(x, h) \leq \mathbb{E}[\Delta_i(x)] \leq C'\phi(x, h).$$

On other hand condition (H2ii) implies that

$$\sup_{u \in B(x, h)} |\Psi(u, t) - \Psi(x, t)| = o(1).$$

Combining these last three results, we obtain

$$\begin{aligned} (|\mathbb{E}[\eta_{ni} | F_{i-1}]|)^2 &\leq \sup_{u \in B(x, h)} |\Psi(u, t) - \Psi(x, t)| \left(\frac{\beta_1^2}{\beta_2\lambda_2(x, \theta_x)} \right) \frac{1}{\phi(x, h)} \phi_i^2(x, h) \\ &\leq \sup_{u \in B(x, h)} |\Psi(u, t) - \Psi(x, t)| \left(\frac{\beta_1^2}{\beta_2\lambda_2(x, \theta_x)} \right) \frac{1}{\phi(x, h)} \phi_i(x, h). \end{aligned}$$

Finally, under the fact that (see, (H1ii))

$$\frac{1}{n\phi(x, h)} \sum_{i=1}^n \phi_i(x, h) \xrightarrow{\mathbb{P}} 1$$

we obtain that

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^n (\mathbb{E}[\eta_{ni} | F_{i-1}])^2 &= \sup_{u \in B(x, h)} |\Psi(u, t) - \Psi(x, t)| \left(\frac{\beta_1^2}{\beta_2\lambda_2(x, \theta_x)} \right) \left(\frac{1}{n\phi(x, h)} \sum_{i=1}^n \phi_i(x, h) \right) \\ &= o_p(1). \end{aligned}$$

Now, we treat the convergence (6), Indeed, we write

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^n \mathbb{E}[\eta_{ni}^2 | F_{i-1}] &= \frac{1}{n(\mathbb{E}K_1(x))^2} \left(\frac{\phi(x, h)\beta_1^2}{\beta_2\lambda_2(x, \theta_x)} \right) \sum_{i=1}^n \mathbb{E}[(\psi_x(Y_i, z) - \Psi(x, z))^2 K_i^2(x) | F_{i-1}] \\ &= \frac{1}{n(\mathbb{E}K_1(x))^2} \left(\frac{\phi(x, h)\beta_1^2}{\beta_2\lambda_2(x, \theta_x)} \right) \left(\sum_{i=1}^n \mathbb{E}[\psi_x^2(Y_i, z)\Delta_i^2(x) | F_{i-1}] \right. \\ &\quad \left. - 2\Psi(x, z) \sum_{i=1}^n \mathbb{E}[\psi_x(Y_i, z)\Delta_i^2(x) | F_{i-1}] \right. \\ &\quad \left. + \Psi^2(x, z) \sum_{i=1}^n \mathbb{E}[\Delta_i^2(x) | F_{i-1}] \right). \end{aligned}$$

Denote

$$J_1 = \sum_{i=1}^n \mathbb{E}[(\psi_x^2(Y_i, z)\Delta_i^2(x) | F_{i-1})], \quad J_2 = \sum_{i=1}^n \mathbb{E}[(\psi_x(Y_i, z)\Delta_i^2(x) | F_{i-1})]$$

and

$$J_3 = \sum_{i=1}^n \mathbb{E}[\Delta_i^2(x) | F_{i-1}].$$

It is easily seen that

$$\begin{aligned} J_1 &= \lambda_2(x, z) \sum_{i=1}^n \mathbb{E}[(K_i^2(x)|\mathfrak{F}_{i-1})] + \sum_{i=1}^n [\mathbb{E}[\psi^2(Y_i, z)K_i^2(x)|\mathfrak{F}_{i-1}] - \lambda_2(x, z)\mathbb{E}[K_i^2(x)|\mathfrak{F}_{i-1}]] \\ &= \lambda_2(x, z) \sum_{i=1}^n \mathbb{E}[(K_i^2(x)|\mathfrak{F}_{i-1})] + \sum_{i=1}^n [\mathbb{E}[K_i^2(x)\mathbb{E}[\psi^2(Y_i, z)|\mathfrak{G}_{i-1}]|\mathfrak{F}_{i-1}] - \lambda_2(x, z)\mathbb{E}[K_i^2(x)|\mathfrak{F}_{i-1}]] \\ &= \lambda_2(x, z) \sum_{i=1}^n \mathbb{E}[(K_i^2(x)|\mathfrak{F}_{i-1})] + \sum_{i=1}^n [\mathbb{E}[K_i^2(x)\mathbb{E}[\psi^2(Y_i, z)|X_i]|\mathfrak{F}_{i-1}] - \lambda_2(x, z)\mathbb{E}[K_i^2(x)|\mathfrak{F}_{i-1}]]. \end{aligned}$$

Using the same arguments as those used in (5), to evaluate the second term. Then, we have,

$$\begin{aligned} &\frac{1}{n\mathbb{E}[K_1(x)]} \sum_{i=1}^n [\mathbb{E}[K_i^2(x)\mathbb{E}[\psi^2(Y_i, z)|X_i]|\mathfrak{F}_{i-1}] - \lambda_2(x, z)\mathbb{E}[K_i^2(x)|\mathfrak{F}_{i-1}]] \\ &\leq \sup_{u \in B(x, h)} |\lambda_2(x, z) - \lambda_2(x, u)| \left(\frac{1}{n\phi(x, h)} \sum_{i=1}^n \mathbb{P}(X_i \in B(x, h)|\mathfrak{F}_{i-1}) \right). \end{aligned}$$

Furthermore, we use the continuity of $\lambda_2(x, \cdot)$, to write

$$\lambda_2(x, z) = \lambda_2(x, \theta_x) + o(1)$$

Thus,

$$\frac{1}{n\mathbb{E}[K_1(x)]} J_1 = \lambda_2(x, \theta_x) \frac{1}{n\mathbb{E}[K_1(x)]} \sum_{i=1}^n \mathbb{E}[K_i^2(x)|\mathfrak{F}_{i-1}] + o(1)$$

and by the same manner

$$\frac{1}{n\mathbb{E}[K_1(x)]} J_2 = \Psi_x(x, \theta_x) \frac{1}{n\mathbb{E}[K_1(x)]} \sum_{i=1}^n \mathbb{E}[K_i^2(x)|\mathfrak{F}_{i-1}] + o(1) = o(1)$$

Finally we have

$$\frac{1}{n} \sum_{i=1}^n \mathbb{E}[\eta_{ni}^2 | F_{i-1}] = \frac{1}{n(\mathbb{E}K_1(x))^2} \left(\frac{\phi(x, h)\beta_1^2}{\beta_2} \right) \sum_{i=1}^n \mathbb{E}[(K_i^2(x)|\mathfrak{F}_{i-1})] + o(1)$$

Next, we use the same ideas used in Ferraty et al. (2009) to get

$$\mathbb{E}[K_i^2(x)|\mathfrak{F}_{i-1}] = K^2(1)\phi_i(x, h) - \int_0^1 (K^2(u))' \phi_i(x, uh) du$$

and

$$\mathbb{E}[K_1(x)] = K(1)\phi(x, h) - \int_0^1 (K(u))' \phi(x, uh) du$$

it follows that

$$\begin{aligned} \frac{1}{n\phi(x, h)} \sum_{i=1}^n \mathbb{E}[(K_i^2(x)|\mathfrak{F}_{i-1})] &= \frac{K^2(1)}{n\phi(x, h)} \sum_{i=1}^n \phi_i(x, h) \\ &\quad - \int_0^1 (K^2(u))' \frac{\phi(x, uh)}{n\phi(x, h)\phi(x, uh)} \sum_{i=1}^n \phi_i(x, uh) du \\ &= K^2(1) - \int_0^1 (K^2(u))' \tau_x(u) du + o_p(1) = \beta_2 + o_p(1). \end{aligned}$$

and

$$\frac{1}{n\phi(x, h)} \mathbb{E}[K_1(x)] = \beta_1 + o(1).$$

We deduce that

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \mathbb{E}[\eta_{ni}^2 | F_{i-1}] = 1$$

which completes the proof of (3).

Concerning (4), we write

$$\zeta_{ni}^2 \mathbb{I}_{\zeta_{ni}^2 > \varepsilon n} \leq \frac{|\zeta_{ni}|^{2+\delta}}{\sqrt{(\varepsilon n)^\delta}} \quad \text{for every } \delta > 0$$

Observe that

$$\begin{aligned} \mathbb{E}[\zeta_{ni}^{2+\delta}] &= \mathbb{E} \left[\left| \eta_{ni}(x) - \mathbb{E}[\eta_{ni} | F_{i-1}] \right|^{2+\delta} \right] \\ &\leq 2^{1+\delta} \mathbb{E} \left[|\eta_{ni}(x)|^{2+\delta} \right] + 2^{1+\delta} \mathbb{E} \left[\left| \mathbb{E}[\eta_{ni} | F_{i-1}] \right|^{2+\delta} \right] \end{aligned}$$

By the Jensen inequality we obtain

$$\mathbb{E}[\zeta_{ni}^{2+\delta}] \leq C \mathbb{E} \left[|\eta_{ni}(x)|^{2+\delta} \right].$$

So, it remains to evaluate $\mathbb{E} \left[|\eta_{ni}(x)|^{2+\delta} \right]$. For this, once again we use the C_r -inequality

$$\mathbb{E} \left[|\eta_{ni}(x)|^{2+K} \right] \leq C \left(\frac{\phi(x, h)\beta_1^2}{\beta_2 \lambda_2(x, \theta_x) \mathbb{E}^2[K_1]} \right)^{1+\delta/2} \mathbb{E} \left[K_i^{2+\delta}(x) \psi^{2+\delta}(Y_i, t) \right] + \Psi^{2+\delta}(x, z) \mathbb{E} \left[K_i^{2+\delta} \right]$$

We condition by X_i , and we use the fact that

$$\mathbb{E} \left[\psi^{2+\delta}(Y_i, t) | X_i \right] < \infty.$$

It follows that

$$\begin{aligned} \mathbb{E} \left[|\eta_{mi}(x)|^{2+\delta} \right] &\leq C \left(\frac{1}{\phi(x, h)} \right)^{1+\delta/2} \mathbb{E}([K_i(x)]^{2+\delta}) \\ &\leq C \left(\frac{1}{\phi(x, h)} \right)^{\delta/2} \end{aligned} \tag{7}$$

Consequently

$$\frac{1}{n} \sum_{i=1}^n \mathbb{E}[\zeta_{mi}^2 \mathbb{I}_{\zeta_{mi}^2 > \varepsilon n}] \leq C \left(\frac{1}{n\phi(x, h)} \right)^{\delta/2} \rightarrow 0$$

which completes the proof of the Lemma.

Proof. (Proof of Lemma 3). By a simple manipulation we have

$$\begin{aligned} \frac{\bar{\Psi}_N(x, z)}{\bar{\Psi}_D(x)} &= \frac{1}{\sum_{i=1}^n \mathbb{E}[K_i | F_{i-1}]} \sum_{i=1}^n \mathbb{E}[K_i [\mathbb{E}[\psi_x(Y, z) | X_i] - \mathbb{E}[\psi_x(Y, z) | X = x]] | F_{i-1}] \\ &\quad + \mathbb{E}[\psi_x(Y, z) | X = x] - \mathbb{E}[\psi_x(Y, \theta(x)) | X = x] =: I_1 + I_2. \end{aligned} \tag{8}$$

For $I_1(x)$ we use the same ideas as in Ferraty *et al.* (2007), we obtain under (H2iii)

$$\begin{aligned} A_i &= \mathbb{E}[K_i [\mathbb{E}[\psi_x(Y, z) | X_i] - \mathbb{E}[\psi_x(Y, z) | X = x]] | F_{i-1}] \\ &= \mathbb{E}[K_i [\mathbb{E}[\Psi(X_i, z) - \Psi(x, z) | d(x, X_i)] | F_{i-1}]] \\ &= \mathbb{E}[K_i \Phi(d(x, X_i), z) | F_{i-1}] \\ &= \int \Phi(th, z) K(t) dP^{F_{i-1}}(th) \\ &= h\Phi'(0, z) \int tK(t) dP^{F_{i-1}}(th). \end{aligned} \tag{9}$$

Using the continuity of $\Phi'(0, \cdot)$ and the fact that

$$\int tK(t) dP^{F_{i-1}}(th) = K(1)\phi_i(x, h) - \int_0^1 (sK(s))' \phi_i(x, sh) ds$$

to obtain that

$$\frac{1}{n} \sum_{i=1}^n A_i = h\Phi'(0, \theta_x) \left(K(1) - \int_0^1 (sK(s))' \tau_x(s) ds \right) + o_p(h).$$

Similarly, we have

$$\frac{1}{n} \sum_{i=1}^n \mathbb{E}[K_i | F_{i-1}] = \left(K(1) - \int_0^1 K'(s) \tau_x(s) ds \right) + o_p(1).$$

At last,

$$I_1 = B_n(x) + o(h).$$

Concerning I_2 we use a Taylor expansion to get, under (H2)

$$I_2 = -B_n(x) + u [n\phi(x, h)]^{-1/2} \sigma(x, \theta_x) \frac{\partial}{\partial t} \Psi(x, \theta_x) + o\left([n\phi(x, h)]^{-1/2}\right).$$

The result is then a consequence of the decomposition (8).

Proof. (Proof of Lemma 4) Clearly, it suffices to show that

$$\frac{\bar{\Psi}_N(x, t) - \Psi(x, t)\bar{\Psi}_D(x)}{\bar{\Psi}_D(x)} = o_p(1)$$

and

$$|\hat{\Psi}_N(x, t) - \bar{\Psi}_N(x, t)| = o_p(1).$$

On the one hand

$$\begin{aligned} \frac{\bar{\Psi}_N(x, t) - \Psi(x, t)\bar{\Psi}_D(x)}{\bar{\Psi}_D(x)} &= \frac{1}{n\mathbb{E}[K_1(x)]\bar{\Psi}_D(x)} \sum_{i=1}^n [\mathbb{E}[K_i(x)\mathbb{E}[\psi(Y_i, t)|\mathfrak{G}_{i-1}]|\mathfrak{F}_{i-1}] - \Psi(x, t)\mathbb{E}[K_i(x)|\mathfrak{F}_{i-1}]] \\ &= \frac{1}{n\mathbb{E}[K_1(x)]\bar{\Psi}_D(x)} \sum_{i=1}^n [\mathbb{E}[K_i(x)\mathbb{E}[\psi(Y_i, t)|X_i]|\mathfrak{F}_{i-1}] - \Psi(x, t)\mathbb{E}[K_i(x)|\mathfrak{F}_{i-1}]] \\ &\leq \frac{1}{n\mathbb{E}[K_1(x)]\bar{\Psi}_D(x)} \sum_{i=1}^n [\mathbb{E}[K_i(x)|\Psi(X_i, t) - \Psi(x, t)|\mathfrak{F}_{i-1}]] \end{aligned}$$

Finally by (H1ii), we deduce that

$$\left| \frac{\bar{\Psi}_N(x, t) - \Psi(x, t)\bar{\Psi}_D(x)}{\bar{\Psi}_D(x)} \right| \leq \sup_{x' \in B(x, h)} |\Psi(x', t) - \Psi(x, t)| \rightarrow 0.$$

On the other hand, the convergence

$$\hat{\Psi}_N(x, z) - \bar{\Psi}_N(x, z) = o_p(1)$$

will be established by showing the following two results

$$\mathbb{E} \left[\hat{\Psi}_N(x, z) - \bar{\Psi}_N(x, z) \right] \rightarrow 0$$

and

$$\text{Var} \left[\hat{\Psi}_N(x, z) - \bar{\Psi}_N(x, z) \right] \rightarrow 0.$$

The first one is a consequence of the definitions of $\hat{\Psi}_N(x, z)$ and $\bar{\Psi}_N(x, z)$. Next, for the second one, we have

$$\hat{\Psi}_N(x, z) - \bar{\Psi}_N(x, z) = \sum_{i=1}^n \Delta_i(x, z)$$

where

$$\Delta_i(x, z) = \frac{1}{n\mathbb{E}[K_1]} K_i \psi(Y_i, z) - \mathbb{E}[K_i \psi(Y_i, z)|\mathfrak{F}_{i-1}]. \text{ By Burkholder's inequality, we have}$$

$$\mathbb{E} \left[\sum_{i=1}^n \Delta_i(x, z) \right]^2 \leq \sum_{i=1}^n \mathbb{E} [\Delta_i(x, z)]^2.$$

Furthermore, by Jensen's inequality we show that

$$\mathbb{E} [\Delta_i(x, z)]^2 \leq \frac{1}{n^2 \mathbb{E}^2[K_1]} \mathbb{E} [K_i^2 \psi^2(Y_i, z)] \leq \frac{1}{n^2 \mathbb{E}^2[K_1]} \mathbb{E} [K_i^2] \leq \frac{1}{n\phi^2(x, h)} \phi_i(x, h)$$

Hence, (H1) gives

$$\text{Var} \left[\widehat{\Psi}_N(x, z) - \bar{\Psi}_N(x, z) \right] \rightarrow 0.$$

Proof. (Proof of the Lemma 5) It is clear that, the monotony of $\psi_x(Y, \cdot)$, for all $\varepsilon > 0$

$$\Psi(x, \theta_x - \varepsilon) \leq \Psi(x, \theta_x) \leq \Psi(x, \theta_x + \varepsilon).$$

By using a similar argument as those used in the previous Lemmas we show that

$$\widehat{\Psi}(x, t) \rightarrow \Psi(x, t) \text{ in probability}$$

for all real fixed $t \in N_x$. So, for sufficiently large n and for all ε small enough

$$\widehat{\Psi}(x, \theta_x - \varepsilon) \leq 0 \leq \widehat{\Psi}(x, \theta_x + \varepsilon) \text{ holds with probability tending to 1.}$$

Since ψ_x is continuous function, then as $\widehat{\Psi}(x, t)$ is a continuous function of t , there exists a $\widehat{\theta}_x \in [\theta_x - \varepsilon, \theta_x + \varepsilon]$ such that $\widehat{\Psi}(x, \widehat{\theta}_x) = 0$. Finally, the uniqueness of $\widehat{\theta}_x$ is a direct consequence of the strict monotonicity of ψ_x , w.r.t. the second component, and the fact that

$$\mathbb{P} \left(\sum_{i=1}^n K_i = 0 \right) = \mathbb{P} \left(\widehat{\Psi}_D(x) = 0 \right) \rightarrow 0 \text{ as } n \rightarrow \infty.$$

which imply $(\sum_{i=1}^n K_i \neq 0)$ with probability tending to 1. Moreover, as $\widehat{\theta}_x \in [\theta_x - \varepsilon, \theta_x + \varepsilon]$ in probability, then

$$\widehat{\theta}_x \rightarrow \theta_x \text{ in probability as } n \rightarrow \infty.$$