Research Article

Open Access

Nadia Kadiri*, Abbes Rabhi, and Amina Angelika Bouchentouf

Strong uniform consistency rates of conditional quantile estimation in the single functional index model under random censorship

<https://doi.org/10.1515/demo-2018-0013>

Received March 31, 2018; accepted September 17, 2018

Abstract: The main objective of this paper is to non-parametrically estimate the quantiles of a conditional distribution in the censorship model when the sample is considered as an α -mixing sequence. First of all, a kernel type estimator for the conditional cumulative distribution function (*cond-cdf*) is introduced. Afterwards, we estimate the quantiles by inverting this estimated *cond-cdf* and state the asymptotic properties when the observations are linked with a single-index structure. The pointwise almost complete convergence and the uniform almost complete convergence (with rate) of the kernel estimate of this model are established. This approach can be applied in time series analysis.

Keywords: Conditional quantile, conditional cumulative distribution, derivatives of conditional cumulative distribution, functional data, kernel estimator, nonparametric estimation; probabilities of small balls, strong mixing processes

MSC: 62G05, 62G99, 62M10

1 Introduction

Estimating quantiles of any distribution allows one to build confidence ranges with applications in numerous fields, such as chemistry, geophysics, medicine and meteorology. Furthermore, statistics for functional random variables are important in their own right. Numerous multivariate statistical techniques, mainly parametric in the functional model terminology, have been extended to functional data and we recommend the interested reader to Ramsay and Silverman ([39] and [40]) or Bosq [10]. Recently nonparametric methods considering functional variables have been studied with very interesting practical motivations dealing with environmetrics, (see Damon and Guillas [15], Fernández *et al.* [17], Aneiros *et al.* [3]), chemometrics (see Ferraty and Vieu [23]), meteorological sciences (see Besse *et al.* [8], Hall and Heckman [34]), speech recognition (see Ferraty and Vieu [24]), radar range profile (see Hall *et al.* [33], Dabo-Niang *et al.* [14]) and medical data (see Gasser *et al.* [30]).

On the other hand, the single index model is a powerful tool to incorporate multivariate regressors into a regression model, thus ensuring some flexibility, while avoiding the "curse of dimensionality".

***Corresponding Author: Nadia Kadiri:** Statistics Laboratory Stochastic Processes University Djillali LIABES of Sidi Bel Abbes, PO. BOX. 89, Sidi Bel Abbes 22000, Algeria, E-mail: nad.kad06@yahoo.com

Abbes Rabhi: Laboratory of Mathematics, University Djillali LIABES of Sidi Bel Abbes, PO. BOX. 89, Sidi Bel Abbes 22000, Algeria, E-mail: rabhi_abbes@yahoo.fr

Amina Angelika Bouchentouf: Laboratory of Mathematics, University Djillali LIABES of Sidi Bel Abbes, PO. BOX. 89, Sidi Bel Abbes 22000, Algeria, E-mail: amina_bouchentouf@yahoo.com

For functional regressors, the single functional index model is a very simple approach for reducing dimensionality by assuming that a functional explanatory variable acts on a scalar response only through its production on one functional direction. This model was first introduced by Ferraty *et al.* [21] for regression problems, and since then has received a lot of attention.

For example, Aït Saidi *et al.* [1] studied the single functional index model for functional time series data; Aït Saidi *et al.* [2] proposed to estimate the unknown functional index via the cross-validation technique. Ferraty *et al.* [20] proposed a new estimator of this parameter based on the idea of functional derivative estimation.

Chen *et al.* [13] introduced a new technique for estimating the link function nonparametrically, and they proposed multi-index modeling using an adaptive linear projections approach.

Attaoui *et al.* [4] dealt with the single index via its conditional density and established the pointwise and the uniform almost complete convergence of the kernel density estimator.

Their results were extended to the α -mixing case by Ling *et al.* [37].

Recently Attaoui [5] established the a.co. of the kernel conditional density and the mode estimators for α -mixing data. For the same data, Attaoui [6] investigated the strong uniform convergence rate and the asymptotic normality of the conditional density estimator in the single index model, and gave an approach to estimate this functional parameter.

In a recent survey of semiparametric FDA, we refer to Goia and Vieu [31, 32] where the authors gave a methodology to approximate the unknown regression operator semiparametrically through a single index approach.

Our purpose in this work is to investigate nonparametrically the estimation of the conditional distribution of a scalar response variable Y , given a functional Hilbertian explanatory variable X , when the observations are linked with a single index relationship.

In the censoring case, instead of observing the lifetime T , we observe the censored lifetime of items under study. That is, assuming that $(T_i)_{i \geq 1}$ is a stationary sequence of lifetimes which satisfy some kind of dependency and $(C_i)_{i \geq 1}$ is a sequence of i.i.d censoring rv with common unknown continuous *df* G , where $Y_i = \min\{T_i, C_i\}$ and $\delta_i = \mathbf{1}_{T_i \leq C_i}$.

To ensure the identifiability of the model, we suppose that $(T_i)_i$ are independent of $(C_i)_i$. Let $\{(Y_i, \delta_i, X_i)\}$ be a sequence of strictly stationary random vectors where $X_{i \geq 1}$ is valued in infinite dimensional semi-metric vector space, and Y_i is real valued. To follow the convention in biomedical studies and as indicated before, we assume that $(C_i)_{i \geq 1}$ and $\{(X_i, T_i)_{i \geq 1}\}$ are independent; this condition is plausible whenever the censoring is independent of the patients modality. Furthermore this condition permits to get an unbiased Kernel estimator. Along this paper we assume that $\tau_G < \infty$ where $\tau_G := \sup\{t : G(t) < 1\}$ and let τ be a positive real number such that $\tau < \tau_G$.

2 The model and the estimates

2.1 The functional nonparametric framework

Consider a random pair (X, T) where T is valued in $\mathbb{R}$ and X is valued in some infinite dimensional Hilbertian space $\mathcal{H}$ with scalar product $\langle \cdot, \cdot \rangle$. Let $(X_i, T_i)_{i=1, \dots, n}$ be the statistical sample of pairs which are identically distributed like (X, T) , but not necessarily independent. Henceforward, X is called functional random variable *f.r.v.*. Let x be fixed in $\mathcal{H}$ and let $F(\theta, t, x)$ be the conditional cumulative distribution function (*cond-cdf*) of T given $\langle \theta, X \rangle = \langle \theta, x \rangle$, specifically:

$$\forall t \in \mathbb{R}, F(\theta, t, x) = \mathbb{P}(T \leq t \mid \langle X, \theta \rangle = \langle x, \theta \rangle).$$

Saying that, we are implicitly assuming the existence of a regular version for the conditional distribution of T given $\langle \theta, X \rangle$. Now, let $\zeta_\theta(\gamma, x)$ be the γ -order quantile of the distribution of T given $\langle \theta, X \rangle = \langle \theta, x \rangle$.

From the *cond-cdf* $F(\theta, \cdot, x)$, the general definition of the γ -order quantile is given as:

$$\zeta_{\theta}(\gamma, x) = \inf\{t \in \mathbb{R} : F(\theta, t, x) \geq \gamma\}, \quad \forall \gamma \in (0, 1).$$

In order to simplify our framework and to focus on the main interest of our paper (the functional feature of $\langle \theta, X \rangle$), we assume that $F(\theta, \cdot, x)$ is strictly increasing and continuous in a neighborhood of $\zeta_{\theta}(\gamma, x)$. This ensures that the conditional quantile $\zeta_{\theta}(\gamma, x)$ is uniquely defined by:

$$\zeta_{\theta}(\gamma, x) = F^{-1}(\theta, \gamma, x). \quad (2.1)$$

Next, in what follows, we assume only smoothness restrictions for the *cond-cdf* $F(\theta, \cdot, x)$ through non-parametric modelling (Section 2.4). We suppose also that $(X_i, T_i)_{i \in \mathbb{N}}$ is an α -mixing sequence, which is one among the most general mixing structures. The α -mixing condition together with the functional approach allow to deal with continuous time processes (see Section 4 for instance).

In our infinite dimensional purpose, we use the terminology *functional nonparametric*, where the word *functional* referees to the infinite dimensionality of the data and where the word *nonparametric* referees to the infinite dimensionality of the model. Such *functional nonparametric* statistics are also called *doubly infinite dimensional* (see Ferraty and Vieu [25], for more details). We also use the terminology *operational statistics* since the target object to be estimated (the *cond-cdf* $F(\theta, \cdot, x)$) can be viewed as a nonlinear operator.

2.2 The estimators

The kernel estimator $F_n(\theta, \cdot, x)$ of $F(\theta, \cdot, x)$ is presented as follows:

$$F_n(\theta, t, x) = \frac{\sum_{i=1}^n K\left(h_K^{-1}(\langle x - X_i, \theta \rangle)\right) H\left(h_H^{-1}(t - T_i)\right)}{\sum_{i=1}^n K\left(h_K^{-1}(\langle x - X_i, \theta \rangle)\right)}, \quad (2.2)$$

where K is a kernel function, H a cumulative distribution function and $h_K = h_{K,n}$ (resp. $h_H = h_{H,n}$) a sequence of positive real numbers. Note that using similar ideas, Roussas [42] introduced some related estimates but in the special case when X is real, while Samanta [43] produced the previously mentioned asymptotic study.

As a by-product of (2.1) and (2.2), it is easy to derive an estimator $\zeta_{\theta,n}(\gamma, x)$ of $\zeta_{\theta}(\gamma, x)$:

$$\zeta_{\theta,n}(\gamma, x) = F_n^{-1}(\theta, \gamma, x). \quad (2.3)$$

Such an estimator is unique as soon as H is an increasing continuous function. Such an approach has been largely used in the case where the variable X is of finite dimension (see e.g. Whang and Zhao, [47], Cai [12], Zhou and Liang [48] or Gannoun *et al.* [29]).

Estimation of the quantile function when the data are censored is an important problem in medicine. This problem is usually modeled by considering a positive variable C called censorship, and the observed random variables. In the censored framework, the observed random variables are the triplets (T_i, δ_i, X_i) where $Y_i = \min(T_i, C_i)$ and $\delta_i = I_{T_i \leq C_i}$, where both of Y_i and C_i are expected to exhibit some kind of dependence which ensures the identifiability of the model. In the following we will use the notations F_1^X and f_1^X to describe the conditional distribution function and the conditional density C knowing the covariate X .

In biomedical case studies, it is assumed that C_i and (T_i, X_i) are independent, this condition is plausible whenever the censoring is independent of the patient's modality.

Our kernel type estimator of the conditional distribution $F(\theta, \cdot, X)$ adapted for censored samples is based on "a pseudo-estimator" of $F(\theta, \cdot, X)$ given by equations (2.4), (2.5) and (2.6).

Such censorship models have been amply studied in the literature for real or multi-dimensional random variables, and in nonparametric frameworks the kernel techniques are particularly used (see Tanner and Wong [45], Padgett [38], Lecoutre and Ould-Saïd [36] and Van Keilegom and Veraverbeke [46], for a necessarily non-exhaustive sample of literature in this area).

The objective of this paragraph is to adapt these ideas under a functional random variable X , and build a kernel type estimator of the conditional distribution $F(\theta, \cdot, X)$ adapted for censored samples. Thus we can reformulate the expression (2.2) as follows:

$$\tilde{F}(\theta, t, x) = \frac{\sum_{i=1}^n \frac{\delta_i}{\tilde{G}(Y_i)} K\left(h_K^{-1}(\langle x - X_i, \theta \rangle)\right) H\left(h_H^{-1}(t - Y_i)\right)}{\sum_{i=1}^n K\left(h_K^{-1}(\langle x - X_i, \theta \rangle)\right)}, \quad (2.4)$$

In practice $\tilde{G}(\cdot) = 1 - G(\cdot)$ is unknown, hence it is impossible to use the estimator (2.4). Then, we replace $\tilde{G}(\cdot)$ by its Kaplan and Meier (1958) estimate $\tilde{G}_n(\cdot)$, given by

$$\tilde{G}_n(t) = 1 - G_n(t) = \begin{cases} \prod_{i=1}^n \left(1 - \frac{1 - \delta_{(i)}}{n - i + 1}\right)^{\mathbf{1}_{\{Y_{(i)} \leq t\}}}, & \text{if } t < Y_{(n)}; \\ 0, & \text{if } t \geq Y_{(n)}. \end{cases} \quad (2.5)$$

where $Y_{(1)} < Y_{(2)} < \dots < Y_{(n)}$ are the order statistics of Y_i and $\delta_{(i)}$ is the non-censoring indicator corresponding to $Y_{(i)}$.

Therefore a feasible estimator of the conditional distribution function $F(\theta, \cdot, x)$ is given by

$$\hat{F}(\theta, t, x) = \frac{\sum_{i=1}^n \frac{\delta_i}{\tilde{G}_n(Y_i)} K\left(h_K^{-1}(\langle x - X_i, \theta \rangle)\right) H\left(h_H^{-1}(t - Y_i)\right)}{\sum_{i=1}^n K\left(h_K^{-1}(\langle x - X_i, \theta \rangle)\right)}. \quad (2.6)$$

Then a natural estimator of $\zeta_\theta(\gamma, x)$ is given by

$$\hat{\zeta}_\theta(\gamma, x) = \hat{F}^{-1}(\theta, \gamma, x) \quad (2.7)$$

$$= \inf\{t \in \mathbb{R} : \hat{F}(\theta, t, x) \geq \gamma\}, \quad (2.8)$$

which satisfies

$$\hat{F}(\theta, \hat{\zeta}_\theta(\gamma, x), x) = \gamma \quad (2.9)$$

2.3 Assumptions on the functional variable

Let N_x be a fixed neighborhood of x and let $B_\theta(x, h)$ be the ball of center x and radius h , namely $B_\theta(x, h) = \{f \in \mathcal{H} : 0 < \langle x - f, \theta \rangle < h\}$ and $S_{\mathbb{R}}$ is a fixed compact of $\mathbb{R}^+$. Assume that, $(C_i)_{i \geq 1}$ are independent and let us consider the following hypotheses:

(H0) $\forall h > 0, \mathbb{P}(X \in B_\theta(x, h)) = \phi_{\theta, x}(h) > 0$,

(H1) $(X_i, Y_i)_{i \in \mathbb{N}}$ is an α -mixing sequence whose the coefficients of mixture verify:

$$\exists a > 0, \exists c > 0 : \forall n \in \mathbb{N}, \alpha(n) \leq cn^{-a}.$$

(H2) $0 < \sup_{i \neq j} \mathbb{P}((X_i, X_j) \in B_\theta(x, h) \times B_\theta(x, h)) = \mathcal{O}\left(\frac{(\phi_{\theta, x}(h_K))^{(a+1)/a}}{n^{1/a}}\right).$

• (H0) can be interpreted as a concentration hypothesis acting on the distribution of the f.r.v. X , while (H2) concerns the behavior of the joint distribution of the pairs (X_i, X_j) . Indeed, this hypothesis is equivalent to assume that, for n large enough for some positive constant C

$$\sup_{i \neq j} \frac{\mathbb{P}((X_i, X_j) \in B_\theta(x, h) \times B_\theta(x, h))}{\mathbb{P}(X \in B_\theta(x, h))} \leq C \left(\frac{\phi_{\theta, x}(h_K)}{n}\right)^{1/a}.$$

This is one way to control the local asymptotic ratio between the joint distribution and its margin. Note that the upper bound increases with a . In other words, the stronger the dependence, the more restrictive (H2) is. The hypothesis (H1) specifies the asymptotic behavior of the α -mixing coefficients.

2.4 The nonparametric model

As is standard in nonparametric estimation, we suppose that the *cond-cdf* $F(\theta, \cdot, x)$ verifies some smoothness constraints. Let b_1 and b_2 be two positive numbers; such that:

$$(H3) \quad \forall (x_1, x_2) \in N_x \times N_x, \forall (t_1, t_2) \in \mathbb{S}_{\mathbb{R}}^2,$$

$$(i) \quad |F(\theta, t_1, x_1) - F(\theta, t_2, x_2)| \leq C_{\theta, x} \left(\|x_1 - x_2\|^{b_1} + |t_1 - t_2|^{b_2} \right),$$

$$(ii) \quad \int_{\mathbb{R}} t f(\theta, t, x) dt < \infty \text{ for all } \theta, x \in \mathcal{H}.$$

$$(H4) \quad F(\theta, \cdot, x) \text{ is } l\text{-times continuously differentiable in some neighborhood of } \zeta_{\theta}(\gamma, x),$$

$$(H5) \quad \forall (x_1, x_2) \in N_x \times N_x, \forall (t_1, t_2) \in \mathbb{S}_{\mathbb{R}}^2,$$

$$|F^{(l)}(\theta, t_1, x_1) - F^{(l)}(\theta, t_2, x_2)| \leq C_{\theta, x} \left(\|x_1 - x_2\|^{b_1} + |t_1 - t_2|^{b_2} \right),$$

where, for any positive integer l , $F^{(l)}(\theta, z, x)$ denotes its l th derivative (i.e. $\left. \frac{\partial^l F(\theta, t, x)}{\partial t^l} \right|_{t=z}$).

Let us note that (H3) is used for the proof of the almost complete convergence of $\hat{\zeta}_{\theta}(\gamma, x)$ whereas (H4) and (H5) are needed to establish the rate of convergence.

3 Asymptotic study

This part of paper is devoted to the theoretical analysis. We begin by giving the almost complete convergence (*a.co.*) of the estimated conditional quantile $\hat{\zeta}_{\theta}(\gamma, x)$. After that, we will focus on the rate of convergence. Concerning notation, C and C' will denote generic constants. Moreover, from now on, h_H (resp. h_K) is a sequence which tends to 0 with n .

3.1 Pointwise almost complete convergence

Let us begin with the statement of an almost complete convergence property ¹. To this end, we need some assumptions concerning the kernel estimator $\hat{F}(\theta, \cdot, x)$:

$$(H6) \quad (i) \quad \forall (t_1, t_2) \in \mathbb{R}^2, |H(t_1) - H(t_2)| \leq C|t_1 - t_2| \text{ and } \int |t|^{b_2} H^{(1)}(t) dt < \infty,$$

$$\text{where, for all } l \in \mathbb{N}^*, H^{(l)}(t) = \left. \frac{d^l H(z)}{dz^l} \right|_{z=t} \text{ and } \lim_{n \rightarrow \infty} n^{\zeta} h_H = \infty, \text{ for some } \zeta > 0$$

$$(ii) \quad \text{The support of } H^{(1)} \text{ is compact and } \forall l \geq j, H^{(l)} \text{ exists and is bounded.}$$

$$(H7) \quad \text{The restriction of } H \text{ to the set } \{u \in \mathbb{R}, H(u) \in (0, 1)\} \text{ is a strictly increasing function.}$$

$$(H8) \quad K \text{ is a positive bounded function with support } [0, 1] \text{ such that } \forall u \in (0, 1) \quad 0 < K(u),$$

$$(H9) \quad (i) \quad \frac{\log n}{n \phi_{\theta, x}(h_K)} \xrightarrow{n \rightarrow \infty} 0.$$

$$(ii) \quad \exists C > 0, \phi_{\theta, x}(h_K) \geq C/n^{2/a+1} \text{ and } \left(\frac{\phi_{\theta, x}(h_K)}{n} \right)^{1/a} + \phi_{\theta, x}(h_K) = o\left(\frac{1}{n^{a/a+1}}\right)$$

$$(H10) \quad (X_i, Y_i) \text{ for } i = 1, \dots, n \text{ are strongly mixing with arithmetic coefficient of order } a > 1, \text{ and } \exists \beta > 2 \text{ such that}$$

$$(i) \quad s_{n,l}^{-(a+1)} = o\left(n^{-\beta}\right) \text{ for } l = 1, 2, 3;$$

¹ Recall that a sequence $(S_n)_{n \in \mathbb{N}}$ of random variables is said to converge almost completely to some variable S , if for any $\epsilon > 0$, we have $\sum_n \mathbb{P}(|S_n - S| > \epsilon) < \infty$. This mode of convergence implies both almost sure and in probability convergence (see for instance Bosq and Lecoutre, [11]).

(ii) $s_{n,k}^{-(a+1)} = o\left(n^{-\beta}\right)$ for $k = 4, 5, 6, 7, 8$;

Remark 3.1. • (H7) ensures the existence of $\hat{\zeta}_{\theta}(\gamma, x)$, while (H6) ensures its uniqueness.

• (H0)-(H5) and (H8) are standard assumptions for the distribution conditional estimation in single functional index model, which have been adopted by Bouchentouf *et al.* [9] for i.i.d case.

• (H9) is a technical condition for our results.

• (H10) is similar to analysis in Ferraty and Vieu [27], and it shows the influence of covariance structure on the convergence rate. $s_{n,l}$ and $s_{n,k}$ will be defined below.

First observe that (2.6) can be rewritten as:

$$\hat{F}(\theta, t, x) = \frac{\hat{F}_N(\theta, t, x)}{\hat{F}_D(\theta, x)} \quad (3.1)$$

Proposition 3.1. Under conditions (H0)-(H4), (H7)-(H9), and assume that (H6)-(i) and (H10)-(i) are satisfied, then we have

$$\sup_{t \in \mathbb{S}_{\mathbb{R}}} \left| \hat{F}(\theta, t, x) - F(\theta, t, x) \right| = \mathcal{O}\left(h_K^{b_1} + h_H^{b_2}\right) + \mathcal{O}_{a.co.}\left(\frac{\sqrt{s_n^2 \log n}}{n}\right) \quad (3.2)$$

where $s_n^2 = \max\{s_{n,1}^2, s_{n,2}^2\}$

Proof. Consider now, for $i = 1, \dots, n$, in what follows, we denote:

$$K_i(\theta, x) = K(h_K^{-1}(\langle x - X_i, \theta \rangle)), \quad H_i(t) = H\left(h_H^{-1}(t - Y_i)\right), \quad \bar{G}_i = \bar{G}(Y_i),$$

$$\hat{F}_N(\theta, t, x) = \frac{1}{n \mathbb{E}(K_1(\theta, x))} \sum_{i=1}^n \frac{\delta_i}{\bar{G}_n(Y_i)} K_i(\theta, x) H_i(t)$$

$$\tilde{F}_N(\theta, t, x) = \frac{1}{n \mathbb{E}(K_1(\theta, x))} \sum_{i=1}^n \frac{\delta_i}{\bar{G}(Y_i)} K_i(\theta, x) H_i(t)$$

$$\hat{F}_D(\theta, x) = \frac{1}{n \mathbb{E}(K_1(\theta, x))} \sum_{i=1}^n K_i(\theta, x)$$

$$\Delta_i(x, \theta) = \frac{K(h_K^{-1}(\langle x - X_i, \theta \rangle))}{\mathbb{E}K_1(\theta, x)}$$

and

$$s_{n,1}^2 = \sum_{i=1}^n \sum_{j=1}^n |\text{Cov}(\Delta_i(x, \theta), \Delta_j(x, \theta))|,$$

$$s_{n,2}^2 = \sum_{i=1}^n \sum_{j=1}^n \left| \text{Cov}\left(\frac{\delta_i}{\bar{G}(Y_i)} H_i(t) \Delta_i(x, \theta), \frac{\delta_j}{\bar{G}(Y_j)} H_j(t) \Delta_j(x, \theta)\right) \right|.$$

The proof is based on the following decomposition, valid for any $t \in \mathbb{S}_{\mathbb{R}}$:

$$\begin{aligned} \sup_{t \in \mathbb{S}_{\mathbb{R}}} \left| \hat{F}(\theta, t, x) - F(\theta, t, x) \right| &\leq \frac{1}{\hat{F}_D(\theta, x)} \sup_{t \in \mathbb{S}_{\mathbb{R}}} \left\{ \left| \hat{F}_N(\theta, t, x) - \tilde{F}_N(\theta, t, x) \right| \right\} \\ &+ \frac{1}{\hat{F}_D(\theta, x)} \sup_{t \in \mathbb{S}_{\mathbb{R}}} \left\{ \left| \tilde{F}_N(\theta, t, x) - \mathbb{E} \tilde{F}_N(\theta, t, x) \right| \right\} \\ &+ \frac{1}{\hat{F}_D(\theta, x)} \sup_{t \in \mathbb{S}_{\mathbb{R}}} \left\{ \left| \mathbb{E} \tilde{F}_N(\theta, t, x) - F(\theta, t, x) \right| \right\} \\ &+ \frac{F(\theta, t, x)}{\hat{F}_D(\theta, x)} \sup_{t \in \mathbb{S}_{\mathbb{R}}} \left| 1 - \hat{F}_D(\theta, x) \right|, \end{aligned} \quad (3.3)$$

Finally, the proof of this proposition is a direct consequence of the following intermediate results.

Lemma 3.1. Assume that either (H6)-(i) is satisfied together with under conditions (H7)-(H8) and if

$$\left(\frac{\log \log n}{n}\right)^{1/2} = o(\phi_{\theta,x}(h_K)),$$

we have

$$\sup_{t \in \mathcal{S}_{\mathbb{R}}} \left\{ \left| \widehat{F}_N^{(l)}(\theta, t, x) - \widetilde{F}_N^{(l)}(\theta, t, x) \right| \right\} = \mathcal{O}_{a.s.} \left(\frac{\log \log n}{n} \right), \text{ for } l \geq 0$$

The following lemma shows the asymptotic bias term of $\widetilde{F}_N(\theta, t, x)$ and $\widehat{F}_D(\theta, x)$ as n tends to infinity.

Lemma 3.2. Under hypotheses (H0), (H4) and (H6)-(i), we have as $n \rightarrow \infty$

$$\sup_{t \in \mathcal{S}_{\mathbb{R}}} \left| \mathbb{E} \left[\widetilde{F}_N(\theta, t, x) \right] - F(\theta, t, x) \right| = \mathcal{O} \left(h_K^{b_1} + h_H^{b_2} \right). \quad (3.4)$$

The following result deals with the variance term of the right-hand side of (3.3) which is expressed by: $\sup_{t \in \mathcal{S}_{\mathbb{R}}} \left\{ \left| \widetilde{F}_N(\theta, t, x) - \mathbb{E} \widetilde{F}_N(\theta, t, x) \right| \right\}$. For $\widehat{F}_D(\theta, x) - \mathbb{E} \left[\widehat{F}_D(\theta, x) \right]$ the same arguments will be used with a slight difference.

Lemma 3.3. Under the assumptions of Proposition 3.1 and if

$$\left(\frac{\log \log n}{n}\right)^{1/2} = o(\phi_{\theta,x}(h_K)),$$

we have

$$(i) \quad \widehat{F}_D(\theta, x) - \mathbb{E} \widehat{F}_D(\theta, x) = \mathcal{O}_{a.co.} \left(\frac{\sqrt{s_{n,1}^2 \log n}}{n} \right),$$

$$(ii) \quad \sup_{t \in \mathcal{S}_{\mathbb{R}}} \left\{ \left| \widetilde{F}_N(\theta, t, x) - \mathbb{E} \widetilde{F}_N(\theta, t, x) \right| \right\} = \mathcal{O}_{a.co.} \left(\frac{\sqrt{s_{n,2}^2 \log n}}{n} \right)$$

We conclude the proof of Proposition 3.1 by making use of (3.3), in conjunction with lemmas 3.1, 3.2 and 3.3. $\square$

The proof of these results will be collected in Section 5.

Theorem 3.1. Put $s_n = \max\{s_{n,1}; s_{n,2}\}$, and under the hypotheses of Proposition 3.1, thus we have:

$$\widehat{\zeta}_{\theta}(\gamma, x) - \zeta_{\theta}(\gamma, x) \xrightarrow[n \rightarrow \infty]{} 0, \text{ a.co.} \quad (3.5)$$

Proof of Theorem 3.1. The proof is based on the pointwise convergence of $\widehat{F}(\theta, \cdot, x)$.

First of all, note that because of (H6)-(i) and (H7), $\widehat{F}(\theta, \cdot, x)$ is a continuous and strictly increasing function. So we have:

$$\forall \epsilon > 0, \exists \delta(\epsilon) > 0, \forall y, |\widehat{F}(\theta, t, x) - \widehat{F}(\theta, \zeta_{\theta}(\gamma, x), x)| \leq \delta(\epsilon) \Rightarrow |t - \zeta_{\theta}(\gamma, x)| \leq \epsilon.$$

This leads us to write

$$\begin{aligned} \forall \epsilon > 0, \exists \delta(\epsilon) > 0, \mathbb{P} \left(|\widehat{\zeta}_{\theta}(\gamma, x) - \zeta_{\theta}(\gamma, x)| > \epsilon \right) &\leq \mathbb{P} \left(|\widehat{F}(\theta, \widehat{\zeta}_{\theta}(\gamma, x), x) - \widehat{F}(\theta, \zeta_{\theta}(\gamma, x), x)| \geq \delta(\epsilon) \right) \\ &= \mathbb{P} \left(|F(\theta, \zeta_{\theta}(\gamma, x), x) - \widehat{F}(\theta, \zeta_{\theta}(\gamma, x), x)| \geq \delta(\epsilon) \right), \end{aligned}$$

since (2.1) and (2.9) and is implying that $\widehat{F}(\theta, \widehat{\zeta}_{\theta}(\gamma, x), x) = \gamma = F(\theta, \zeta_{\theta}(\gamma, x), x)$.

Moreover, we have

$$|F(\theta, \widehat{\zeta}_{\theta}(\gamma, x), x) - F(\theta, \zeta_{\theta}(\gamma, x), x)| = |F(\theta, \widehat{\zeta}_{\theta}(\gamma, x), x) - \widehat{F}(\theta, \widehat{\zeta}_{\theta}(\gamma, x), x)|$$

$$\leq \sup_{t \in \mathcal{S}_{\mathbb{R}}} |\widehat{F}(\theta, t, x) - F(\theta, t, x)|. \quad (3.6)$$

The consistency of $\widehat{\zeta}_{\theta}(\gamma, x)$ follows then immediately from Proposition 3.1 and the following inequality

$$\sum_{n \geq 1} \mathbb{P}(|\zeta_{\theta, n}(\gamma, x) - \zeta_{\theta}(\gamma, x)| \geq \epsilon) \leq \sum_{n \geq 1} \mathbb{P}\left(\sup_{t \in \mathcal{S}_{\mathbb{R}}} |\widehat{F}(\theta, t, x) - F(\theta, t, x)| \geq \delta(\epsilon)\right).$$

□

3.2 Pointwise almost complete rate of convergence

In this section we study the rate of convergence of our conditional quantile estimator $\widehat{\zeta}_{\theta}(\gamma, x)$. Because this kind of result is stronger than the previous one, we have to introduce some additional assumptions. As is standard in conditional quantiles estimation, the rate of convergence can be linked with the flatness of *cond-cdf* $F(\theta, \cdot, x)$ around the conditional quantile $\zeta_{\theta}(\gamma, x)$. This is one reason why we introduced hypotheses (H4) and (H5). But a complementary way to take this local shape constraint into account is to suppose that:

$$(H11) \quad \exists l > 0, \forall m, 1 \leq m < l, \widehat{F}^{(m)}(\theta, \zeta_{\theta}(\gamma, x), x) = 0 \text{ and } 0 < |F^{(l)}(\theta, \zeta_{\theta}(\gamma, x), x)| < \infty.$$

Because we focus on the local behavior of $F(\theta, \cdot, x)$ around $\zeta_{\theta}(\gamma, x)$ via its derivatives, that leads us to consider the successive derivatives of $\widehat{F}(\theta, \cdot, x)$ and subsequently some assumptions on the successive derivatives of the cumulative kernel H :

$$(H12) \quad \forall i \neq j, \text{ the conditional density of } (Y_i, Y_j) \text{ given } (\langle X_i, \theta \rangle, \langle X_j, \theta \rangle) \text{ is continuous at } (\zeta_{\theta}(\gamma, x), \zeta_{\theta}(\gamma, x)).$$

Proposition 3.2. Suppose that the hypotheses (H0)-(H2), (H4)-(H9) are satisfied and if

$$\lim_{n \rightarrow \infty} \frac{\log n}{n h_H^{2l-1} \phi_{\theta, x}(h_K)} = 0.$$

Then we have

$$|\widehat{F}^{(l)}(\theta, t, x) - F^{(l)}(\theta, t, x)| = \mathcal{O}\left(h_K^{b_1} + h_H^{b_2}\right) + \mathcal{O}_{a.co.}\left(\frac{\sqrt{s_n'^2 \log n}}{n}\right).$$

Theorem 3.2. Put $s_n'^2 = \max\{s_{n,1}^2; s_{n,3}^2\}$, and assume that either (H10)-(i) is satisfied together with hypotheses (H0)-(H9) and (H11)-(H12), we have

$$\widehat{\zeta}_{\theta}(\gamma, x) - \zeta_{\theta}(\gamma, x) = \mathcal{O}\left(\left(h_K^{b_1} + h_H^{b_2}\right)^{\frac{1}{l}}\right) + \mathcal{O}_{a.co.}\left(\left(\frac{s_n'^2 \log n}{n^2}\right)^{\frac{1}{2l}}\right), \quad (3.7)$$

where

$$s_{n,3}^2 = \sum_{i=1}^n \sum_{j=1}^n \left| \text{Cov}\left(h_H^{-l} \frac{\delta_i}{\widehat{G}(Y_i)} H_i^{(l)}(t) \Delta_i(x, \theta), h_H^{-l} \frac{\delta_j}{\widehat{G}(Y_j)} H_j^{(l)}(t) \Delta_j(x, \theta)\right) \right|$$

Proof of Theorem 3.2. The proof is based on the Taylor expansion of $\widehat{F}(\theta, \cdot, x)$ at $\zeta_{\theta}(\gamma, x)$ and on the use of (H11):

$$\begin{aligned} \widehat{F}(\theta, \zeta_{\theta}(\gamma, x), x) - \widehat{F}(\theta, \widehat{\zeta}_{\theta}(\gamma, x), x) &= \sum_{m=1}^{l-1} \frac{(\zeta_{\theta}(\gamma, x) - \widehat{\zeta}_{\theta}(\gamma, x))^{m-1}}{m!} \widehat{F}^{(m)}(\theta, \zeta_{\theta}(\gamma, x), x) \\ &\quad + \frac{(\zeta_{\theta}(\gamma, x) - \widehat{\zeta}_{\theta}(\gamma, x))^l}{l!} \widehat{F}^{(l)}(\theta, \zeta_{\theta}^*, x), \\ &= \sum_{m=1}^{l-1} \frac{(\zeta_{\theta}(\gamma, x) - \widehat{\zeta}_{\theta}(\gamma, x))^{m-1}}{m!} (\widehat{F}^{(m)}(\theta, \zeta_{\theta}(\gamma, x), x) - F^{(m)}(\theta, \zeta_{\theta}(\gamma, x), x)) \end{aligned}$$

$$+ \frac{(\zeta_\theta(\gamma, x) - \widehat{\zeta}_\theta(\gamma, x))^l}{l!} \widehat{F}^{(l)}(\theta, \zeta_\theta^*, x),$$

where, for all $t \in \mathbb{R}$,

$$\begin{aligned} \widehat{F}^{(l)}(\theta, t, x) &= \frac{h_H^{-l} \sum_{i=1}^n \frac{\delta_i}{\bar{G}_n(Y_i)} K(h_K^{-1}(\langle x - X_i, \theta \rangle)) H^{(l)}(h_H^{-1}(t - Y_i))}{\sum_{i=1}^n K(h_K^{-1}(\langle x - X_i, \theta \rangle))}, \\ \widetilde{F}^{(l)}(\theta, t, x) &= \frac{h_H^{-l} \sum_{i=1}^n \frac{\delta_i}{\bar{G}(Y_i)} K(h_K^{-1}(\langle x - X_i, \theta \rangle)) H^{(l)}(h_H^{-1}(t - Y_i))}{\sum_{i=1}^n K(h_K^{-1}(\langle x - X_i, \theta \rangle))}, \end{aligned}$$

and where $\min(\zeta_\theta(\gamma, x), \widehat{\zeta}_\theta(\gamma, x)) < \zeta_\theta^* < \max(\zeta_\theta(\gamma, x), \widehat{\zeta}_\theta(\gamma, x))$. Suppose now that we have the following result.

Because of Theorem 3.1, Proposition 3.2 and (H11), we have:

$$\widehat{F}^{(l)}(\theta, \zeta_\theta^*, x) \xrightarrow[n \rightarrow \infty]{} F^{(l)}(\theta, \zeta_\theta(\gamma), x) \neq 0, \text{ a.co.},$$

then we derive

$$\begin{aligned} \left| \zeta_\theta(\gamma, x) - \widehat{\zeta}_\theta(\gamma, x) \right|^l &= \mathcal{O} \left(\widehat{F}(\theta, \zeta_\theta(\gamma, x), x) - F(\theta, \zeta_\theta(\gamma, x), x) \right) \\ &+ \mathcal{O} \left(\sum_{m=1}^{l-1} (\zeta_\theta(\gamma, x) - \widehat{\zeta}_\theta(\gamma, x))^m (\widehat{F}^{(m)}(\theta, \zeta_\theta(\gamma, x), x) - F^{(m)}(\theta, \zeta_\theta(\gamma, x), x)) \right), \text{ a.co.} \end{aligned}$$

Now, comparing the convergence rates given in Proposition 3.1 and 3.2, we get

$$\left| \zeta_\theta(\gamma, x) - \widehat{\zeta}_\theta(\gamma, x) \right|^l = \mathcal{O} \left(\widehat{F}(\theta, t, x) - F(\theta, t, x) \right), \text{ a.co.}$$

Thus, first part of Lemma 3.3 with lemmas 3.4 and 3.5 allow us to get the stated result.

Lemma 3.4. Under the hypotheses (H0) and (H5)-(H8) we have

$$\sup_{t \in \mathcal{S}_{\mathbb{R}}} |F^{(l)}(\theta, t, x) - \mathbb{E}(\widetilde{F}_N^{(l)}(\theta, t, x))| = \mathcal{O} \left(h_K^{b_1} + h_H^{b_2} \right),$$

Lemma 3.5. Under the assumptions of Proposition 3.2, we have

$$\sup_{t \in \mathcal{S}_{\mathbb{R}}} |\widetilde{F}_N^{(l)}(\theta, t, x) - \mathbb{E}(\widetilde{F}_N^{(l)}(\theta, t, x))| = \mathcal{O}_{a.co.} \left(\sqrt{\frac{s_{n,3}^2 \log n}{n^2}} \right),$$

□

4 Uniform almost complete convergence and rate of convergence

In this section we derive the uniform version of Theorems 3.1 and 3.2. The study of the uniform consistency is a crucial tool for studying the asymptotic properties of all estimates of the functional index if it is unknown. In the multivariate case, the uniform consistency is a standard extension of the pointwise one, nevertheless, in the studied case, it requires some additional tools and topological conditions (see Ferraty *et al.* [19]). Consequently, coupled with the conditions introduced previously, we need the following results. Firstly, consider

$$\mathcal{S}_{\mathcal{H}} \subset \bigcup_{k=1}^{d_n^{\mathcal{S}_{\mathcal{H}}}} B_\theta(x_k, r_n) \text{ and } \Theta_{\mathcal{H}} \subset \bigcup_{q=1}^{d_n^{\Theta_{\mathcal{H}}}} B_\theta(\theta_q, r_n) \quad (4.1)$$

with x_k (resp. θ_q) $\in \mathcal{H}$ and $r_n, d_n^{\mathcal{S}_{\mathcal{H}}}, d_n^{\Theta_{\mathcal{H}}}$ are sequences of positive real numbers which tend to infinity as n goes to infinity and suppose that $d_n^{\mathcal{S}_{\mathcal{H}}}, d_n^{\Theta_{\mathcal{H}}}$ are the minimal numbers of open balls with radius r_n in $\mathcal{H}$, which are required to cover $\mathcal{S}_{\mathcal{H}}$ and $\Theta_{\mathcal{H}}$.

4.1 Conditional quantile distribution estimation

In this subsection we will investigate the uniform almost complete convergence of our estimator (2.3), to this end, we need to state the following assumptions

(A1) There exists a differentiable function $\phi(\cdot)$ such that $\forall x \in \mathcal{S}_{\mathcal{H}}$ and $\forall \theta \in \Theta_{\mathcal{H}}$,

$$0 < C\phi(h) \leq \phi_{\theta,x}(h) \leq C'\phi(h) < \infty \text{ and } \exists \eta_0 > 0, \forall \eta < \eta_0, \phi'(\eta) < C,$$

(A2) The kernel K satisfies (H3) and the Lipschitz condition holds

$$|K(x) - K(y)| \leq C\|x - y\|,$$

(A3) $\forall (t_1, t_2) \in \mathcal{S}_{\mathbb{R}} \times \mathcal{S}_{\mathbb{R}}, \forall (x_1, x_2) \in \mathcal{S}_{\mathcal{H}} \times \mathcal{S}_{\mathcal{H}}$ and $\forall \theta \in \Theta_{\mathcal{H}}$,

$$|F(\theta, t_1, x_1) - F(\theta, t_2, x_2)| \leq C \left(\|x_1 - x_2\|^{b_1} + |t_1 - t_2|^{b_2} \right),$$

(A4) For some $\nu \in (0, 1)$, $\lim_{n \rightarrow \infty} n^\nu h_H = \infty$, and for $r_n = \mathcal{O}\left(\frac{\log n}{n}\right)$, the sequences $d_n^{\mathcal{S}_{\mathcal{H}}}$ and $d_n^{\Theta_{\mathcal{H}}}$ satisfy:

$$\begin{cases} \text{(i)} \frac{(\log n)^2}{n\phi(h_K)} < \log d_n^{\mathcal{S}_{\mathcal{H}}} + \log d_n^{\Theta_{\mathcal{H}}} < \frac{n\phi(h_K)}{\log n}, \\ \text{(ii)} \sum_{n=1}^\infty n^{1/2b_2} (d_n^{\mathcal{S}_{\mathcal{H}}} d_n^{\Theta_{\mathcal{H}}})^{1-\beta} < \infty \text{ for some } \beta > 1, \\ \text{(iii)} n\phi(h_K) = \mathcal{O}((\log n)^2). \end{cases}$$

(A5) $\forall (t_1, t_2) \in \mathcal{S}_{\mathbb{R}} \times \mathcal{S}_{\mathbb{R}}, \forall (x_1, x_2) \in \mathcal{S}_{\mathcal{H}} \times \mathcal{S}_{\mathcal{H}}$ and $\forall \theta \in \Theta_{\mathcal{H}}$,

$$|F^{(l)}(\theta, t_1, x_1) - F^{(l)}(\theta, t_2, x_2)| \leq C \left(\|x_1 - x_2\|^{b_1} + |t_1 - t_2|^{b_2} \right),$$

(A6) For some $\nu \in (0, 1)$, $\lim_{n \rightarrow \infty} n^\nu h_H = \infty$, and for $r_n = \mathcal{O}\left(\frac{\log n}{n}\right)$, the sequences $d_n^{\mathcal{S}_{\mathcal{H}}}$ and $d_n^{\Theta_{\mathcal{H}}}$ satisfy:

$$\begin{cases} \text{(i)} \frac{(\log n)^2}{nh_H^{2l-1}\phi(h_K)} < \log d_n^{\mathcal{S}_{\mathcal{H}}} + \log d_n^{\Theta_{\mathcal{H}}} < \frac{nh_H^{2l-1}\phi(h_K)}{\log n}, \\ \text{(ii)} nh_H^{2l-1}\phi(h_K) = \mathcal{O}((\log n)^2) \end{cases}$$

And let

$$\begin{aligned} s_{n,4}^2 &= \sum_{i=1}^n \sum_{j=1}^n |Cov(\Lambda_i, \Lambda_j)|, \quad s_{n,5}^2 = \sum_{i=1}^n \sum_{j=1}^n |Cov(\Omega_i, \Omega_j)| \\ s_{n,6}^2 &= \sum_{i=1}^n \sum_{j=1}^n |Cov(\Delta_i(x_{k(x)}, \theta_{q(\theta)}), \Delta_j(x_{k(x)}, \theta_{q(\theta)}))|, \quad s_{n,7}^2 = \sum_{i=1}^n \sum_{j=1}^n |Cov(\Gamma_i, \Gamma_j)| \\ s_{n,8}^2 &= \sum_{i=1}^n \sum_{j=1}^n |Cov(\Gamma_i^{(l)}, \Gamma_j^{(l)})| \end{aligned}$$

where

$$\begin{aligned} \Lambda_i(x, \theta) &= \frac{1}{h_K \phi(h_K)} \mathbf{1}_{B_\theta(x, h) \cup B_\theta(x_{k(x)}, h)}(X_i), \\ \Omega_i(x, \theta) &= \frac{1}{h_K \phi(h_K)} \mathbf{1}_{B_\theta(x_{k(x)}, h) \cup B_{\theta_{q(\theta)}}(x_{k(x)}, h)}(X_i), \end{aligned}$$

$$\Delta_i(x_{k(x)}, \theta_{q(\theta)}) = \frac{K(h_K^{-1} < x_{k(x)} - X_i, \theta_{q(\theta)} >)}{\mathbb{E}K(h_K^{-1} < x_{k(x)} - X_i, \theta_{q(\theta)} >)},$$

$$\begin{aligned} \Gamma_i &= \frac{K(h_K^{-1} < x_{k(x)} - X_i, \theta_{q(\theta)} >)}{\mathbb{E}K(h_K^{-1} < x_{k(x)} - X_i, \theta_{q(\theta)} >)} \frac{\delta_i}{\bar{G}(Y_i)} H(h_H^{-1}(v_{k_t} - Y_i)) \\ &- \mathbb{E} \left(\frac{K(h_K^{-1} < x_{k(x)} - X_i, \theta_{q(\theta)} >)}{\mathbb{E}K(h_K^{-1} < x_{k(x)} - X_i, \theta_{q(\theta)} >)} \frac{\delta_i}{\bar{G}(Y_i)} H(h_H^{-1}(v_{k_t} - Y_i)) \right) \end{aligned}$$

and

$$\begin{aligned} \Gamma_i^{(l)} &= \frac{1}{h_H^l} \frac{K(h_K^{-1} < x_{k(x)} - X_i, \theta_{q(\theta)} >)}{\mathbb{E}K(h_K^{-1} < x_{k(x)} - X_i, \theta_{q(\theta)} >)} H^{(l)}(h_H^{-1}(v_{k_t} - Y_i)) \\ &- \frac{1}{h_H^l} \mathbb{E} \left(\frac{K(h_K^{-1} < x_{k(x)} - X_i, \theta_{q(\theta)} >)}{\mathbb{E}K(h_K^{-1} < x_{k(x)} - X_i, \theta_{q(\theta)} >)} H^{(l)}(h_H^{-1}(v_{k_t} - Y_i)) \right) \end{aligned}$$

Theorem 4.1. Set $s_n^* = \max\{s_{n,4}; s_{n,5}; s_{n,6}; s_{n,7}\}$, and assume that either (H10)-(ii) is satisfied together with under hypotheses (H0)-(H3) and (H6)-(H9), (A1)-(A4), we have

$$\sup_{x \in \mathcal{S}_{\mathcal{H}}} |\hat{\zeta}_\theta(\gamma, x) - \zeta_\theta(\gamma, x)| \xrightarrow[n \rightarrow \infty]{} 0, \text{ a.co.} \quad (4.2)$$

Proof of Theorem 4.1. The proof of the theorem can be completed by using the following results.

Lemma 4.1. Under the conditions (H0)-(H3) and (H6)-(H9), we have

$$\sup_{\theta \in \Theta_{\mathcal{H}}} \sup_{x \in \mathcal{S}_{\mathcal{H}}} \sup_{t \in \mathcal{S}_{\mathcal{R}}} |F(\theta, t, x) - \mathbb{E}\hat{F}_N(\theta, t, x)| = \mathcal{O}(h_K^{b_1}) + \mathcal{O}(h_H^{b_2}). \quad (4.3)$$

Lemma 4.2. Under the assumptions of Theorem 4.1, we have:

$$\begin{aligned} 1. \sup_{\theta \in \Theta_{\mathcal{F}}} \sup_{x \in \mathcal{S}_{\mathcal{F}}} |\hat{F}_D(\theta, x) - \mathbb{E}\hat{F}_D(\theta, x)| &= \mathcal{O}_{a.co.} \left(\frac{\sqrt{\max\{s_{n,4}^2; s_{n,5}^2; s_{n,6}^2\} \log d_n^{\mathcal{S}_{\mathcal{F}}} d_n^{\Theta_{\mathcal{F}}}}}{n} \right), \\ 2. \end{aligned}$$

$$\begin{aligned} \sup_{\theta \in \Theta_{\mathcal{H}}} \sup_{x \in \mathcal{S}_{\mathcal{H}}} \sup_{t \in \mathcal{S}_{\mathcal{R}}} \hat{F}_N(\theta, t, x) - \mathbb{E}\hat{F}_N(\theta, t, x) &= \mathcal{O}_{a.co.} \left(\frac{\sqrt{\max\{s_{n,4}^2; s_{n,5}^2; s_{n,7}^2\} \log d_n^{\mathcal{S}_{\mathcal{F}}} d_n^{\Theta_{\mathcal{F}}}}}{n} \right) \\ &+ \mathcal{O}_{a.co.} \left(\sqrt{\frac{\log d_n^{\mathcal{S}_{\mathcal{F}}} d_n^{\Theta_{\mathcal{F}}}}{n\phi(h_K)}} \right). \end{aligned}$$

□

Theorem 4.2. Under hypotheses (H0)-(H3), (H6)-(H10), (A1)-(A2) and (A5)-(A6), we have

$$\begin{aligned} \sup_{\theta \in \Theta_{\mathcal{F}}} \sup_{x \in \mathcal{S}_{\mathcal{F}}} |\hat{\zeta}_\theta(\gamma, x) - \zeta_\theta(\gamma, x)| &= \mathcal{O} \left(\left(h_K^{b_1} + h_H^{b_2} \right)^{\frac{1}{l}} \right) + \mathcal{O}_{a.co.} \left(\left(\frac{\log d_n^{\mathcal{S}_{\mathcal{F}}} d_n^{\Theta_{\mathcal{F}}}}{nh_H^{2l-1}\phi(h_K)} \right)^{\frac{1}{2l}} \right) \\ &+ \mathcal{O}_{a.co.} \left(\left(\frac{s_n''^2 \log d_n^{\mathcal{S}_{\mathcal{F}}} d_n^{\Theta_{\mathcal{F}}}}{n^2} \right)^{\frac{1}{2l}} \right) \end{aligned}$$

where $s_n'' = \max\{s_{n,4}; s_{n,5}; s_{n,6}; s_{n,8}\}$

Proof. Obviously, the proof of these results, can be deduced from the decomposition (7.2) and the following intermediate results which are only uniform version of Proposition 3.2.

Lemma 4.3. Under the conditions (H0)-(H3) and (H6)-(H9), we have

$$\sup_{\theta \in \Theta_{\mathcal{J}_C}} \sup_{x \in \mathcal{S}_{\mathcal{J}_C}} \sup_{t \in \mathcal{S}_{\mathbb{R}}} \left| F^{(l)}(\theta, t, x) - \mathbb{E} \widehat{F}_N^{(l)}(\theta, t, x) \right| = \mathcal{O} \left(h_K^{b_1} + h_H^{b_2} \right). \quad (4.4)$$

Lemma 4.4. Under the assumptions of Theorem 4.1, we have:

$$\sup_{\theta \in \Theta_{\mathcal{J}_C}} \sup_{x \in \mathcal{S}_{\mathcal{J}_C}} \left| \widehat{F}_D(\theta, x) - \mathbb{E} \widehat{F}_D(\theta, x) \right| = \mathcal{O}_{a.co.} \left(\frac{\sqrt{\max\{s_{n,3}^2; s_{n,4}^2; s_{n,5}^2\} \log d_n^{\mathcal{S}_{\mathcal{J}_C}} d_n^{\Theta_{\mathcal{J}_C}}}}{n} \right).$$

Lemma 4.5. Under the assumptions of Theorem 4.1, we have:

$$\begin{aligned} \sup_{\theta \in \Theta_{\mathcal{J}_C}} \sup_{x \in \mathcal{S}_{\mathcal{J}_C}} \sup_{t \in \mathcal{S}_{\mathbb{R}}} \widehat{F}_N^{(l)}(\theta, t, x) - \mathbb{E} \widehat{F}_N^{(l)}(\theta, t, x) &= \mathcal{O}_{a.co.} \left(\frac{\sqrt{s_n'^{*2} \log d_n^{\mathcal{S}_{\mathcal{J}_C}} d_n^{\Theta_{\mathcal{J}_C}}}}{n} \right) \\ &+ \mathcal{O}_{a.co.} \left(\sqrt{\frac{\log d_n^{\mathcal{S}_{\mathcal{J}_C}} d_n^{\Theta_{\mathcal{J}_C}}}{n h_H^{2l-1} \phi(h_K)}} \right). \end{aligned}$$

□

Remark 4.1. 1. These results extends Theorem 3 or Theorem 4 given in Bouchentouf *et al.* [9] to the mixing case. The effect of covariance structure for dependence case on the convergence rate is reflected in the last term. Specially, if the functional single-index is fixed, it is easy to prove the following corollary that are similar the one given in Bouchentouf *et al.* [9].

2. Theorems 4.1 and 4.2 are obtained when the functional parameter θ is known. However, if the functional parameter θ is unknown, it would be very important for practical use of the model and would be more interesting.

Corollary 4.1. Under the conditions of Theorem 4.1 for $l = 0$, we have

$$\sup_{x \in \mathcal{S}_{\mathcal{J}_C}} \sup_{t \in \mathcal{S}_{\mathbb{R}}} |\widehat{F}(\theta, t, x) - F(\theta, t, x)| = \mathcal{O} \left(h_K^{b_1} + h_H^{b_2} \right) + \mathcal{O}_{a.co.} \left(\sqrt{\frac{\log d_n^{\mathcal{S}_{\mathcal{J}_C}}}{n \phi(h_K)}} \right) + \mathcal{O}_{a.co.} \left(\frac{\sqrt{s_n'^{*2} \log d_n^{\mathcal{S}_{\mathcal{J}_C}}}}{n} \right)$$

where $s_n'^{*} = \max\{s_{n,4}; s_{n,6}; s_{n,7}\}$

5 Simulation study

To study the behavior of our conditional quantiles estimator, in this section we consider two examples of simulation. First, we compare our model CESIM (functional single index model with censored data in mixing case) with that of NPFDA (non-parametric functional data analysis) and in the latter, knowing the distribution of the regression model (the distribution is known and usual), we look to the behavior of our estimator of the conditional distribution function with respect to this distribution.

Example 5.1. Let us consider the following regression model where the covariate is a curve and the response is a scalar:

$$T_i = R(X_i) + \epsilon_i, \quad i = 1, \dots, n \quad (5.1)$$

where ϵ_i is the error supposed to be generated by an autoregressive model defined by

$$\epsilon_i = \frac{1}{\sqrt{2}} (\epsilon_{i-1} + \eta_i), \quad i = 1, \dots, n$$

with $(\eta_i)_i$ a sequence of i.i.d. random variables normally distributed with a variance equal to 0.1.

The functional covariate X is supposed to be a diffusion process defined on $[0, 2\pi/3]$ and generated by the following equation:

$$X(t) = 2 - \cos\left(W\left(t - \frac{2\pi}{3}\right)\right), \quad t \in]0, 2\pi/3[$$

where W is an α -mixing process generated by $W_i = \frac{2}{9}W_{i-1} + \epsilon_i$, $i = 1, \dots, 215$ and ϵ_i are i.i.d $\mathcal{N}(0, 1)$ (standard normal distribution) independently from W_i (the standard normal W_0 is independently generated). We carried out the simulation with a 100-sample of the curve X which is represented by the following graph above. Here is a nonlinear regression function is considered such that

$$R(X) = \frac{1}{4} \exp\left\{2 - 1/\left(\int_0^1 X'(t)dt\right)^2\right\}.$$

On the other hand, n i.i.d. random variables $(C_i)_i$ are simulated through the exponential distribution $\mathcal{E}(1.5)$.

Given $X = x$, $T \rightsquigarrow \mathcal{N}(R(x), 0.2)$, and thus, the conditional median, the conditional mode and the conditional mean functions will coincide and will be equal to $R(x)$, for any fixed x .

The computation of our estimator is based on the observed data $(X_i, Y_i, \delta_i)_{i=1, \dots, n}$ where $Y_i = \min(T_i, C_i)$; $\delta_i = \mathbb{I}_{\{T_i \leq C_i\}}$ and the single index θ which is unknown and has to be estimated.

In practice this parameter can be selected by cross-validation (see Aït Saidi *et al.* (2008)).

In this passage it may be that one can select the real-valued function $\theta(t)$ among the eigenfunctions of the covariance operator $\mathbb{E}[(X' - \mathbb{E}X') < X', \cdot]_{\mathcal{H}}$ where $X(t)$ is a diffusion processes defined on a real interval $[a, b]$ and $X'(t)$ its first derivative (see Attaoui & Ling (2016)). So for a chosen training sample $\mathcal{L}$, by applying the principal component analysis (PCA) method, the computation of the eigenvectors of the covariance operator estimated by its empirical covariance operator: $\frac{1}{|\mathcal{L}|} \sum_{i \in \mathcal{L}} (X'_i - \mathbb{E}X')^t (X'_i - \mathbb{E}X')$, will be the one best

approximation of our functional parameter θ . Now, let us denote θ^* the first eigenfunction corresponding to the first higher eigenvalue, which will replace θ during the simulation step.

In practice, some tuning parameters have to be fixed: the kernel $K(\cdot)$ is chosen to be the quadratic function defined as $K(u) = \frac{3}{2}(1 - u^2)\mathbb{I}_{[0,1]}$ and the cumulative df $H(u) = \int_{-\infty}^u \frac{3}{4}(1 - z^2)\mathbb{I}_{[-1,1]}(z) dz$. Take into account of the smoothness of the curves $X_i(t)$ (Figure 1), we choose the distance in $\mathcal{H}$ as:

$$d(\chi_1, \chi_2) = \left(\int_0^1 (\chi'_1(t) - \chi'_2(t))^2 dt\right)^{1/2}$$

as semi-metric.

In the following graphs, the covariance operator for $\mathcal{L} = \{1, \dots, 215\}$ gives the discretization of the eigenfunctions $\theta_i(t)$ (presented by continuous curves). First, we show two eigenfunctions θ_1, θ_2 (Figure 2), next we show twenty eigenfunctions $\theta_1, \dots, \theta_{20}$ (Figure 3). Finally, in Figure 4, we give all the 215 eigenfunctions $\theta_i(t)$.

Now to simplify the implementation of our methodology, we take the bandwidths $h_H \sim h_K = h$, where h will be chosen by the cross-validation method on the k -nearest neighbors (see Ferraty and Vieu, 2006, p. 102) and we denote by θ^* the first eigenfunction corresponding to the first higher eigenvalue of the empirical covariance operator:

$$\frac{1}{|\mathcal{L}|} \sum_{i \in \mathcal{L}} (X'_i - \mathbb{E}X')^t (X'_i - \mathbb{E}X').$$

In this simulation, we divide our sample of size 215 into two parts. The first one from 1 to 160 will be used to make the simulation and the second from 161 to 215 will serve us for the prediction.

We follow the following steps:

Figure 1: A sample of curves $\{X_i(t), t \in [0, 1]\}_{i=1, \dots, 215}$

Step 1. Simulate the response variables Y_i .

Step 2. For each j in the test sample $\mathcal{J} = 161, \dots, 215$, we compute: $\hat{Y}_j = \hat{t}_{\alpha, \theta^*}(X_j)$ and $\hat{Y}_j = \hat{t}_{\alpha, \theta^*}(X_j)$, where

$$t_{\alpha}(\chi) = \inf \{y \in \mathbb{R}, F^{\chi}(y) \geq \alpha\},$$

and

$$\widehat{F^{\chi}}(y) = \frac{\sum_{i=1}^n K(h^{-1}d(\chi, X_i))(h^{-1}(y - Y_i))}{\sum_{i=1}^n K(h^{-1}d(\chi, X_i))} \quad \forall y \in \mathbb{R}.$$

Finally, we present the results by plotting the predicted values against the true values and compute the sum of squared residuals (SSR):

$$SSR = \frac{1}{|\mathcal{J}|} \sum_{j \in \mathcal{J}} (Y_j - \hat{Y}_j)^2.$$

We see that the sum of squared residuals (SSR) of our method (Censored-Single-Index-Mixing-Method) is less than the one of the Non-Parametric-Functional-Data-Analysis (NPFDA). This is confirmed by the following graphs, when we compare the conditional quantile by censored single index mixing methods (CFSIM) against the conditional quantile by nonparametric functional data analysis (NPFDA) (Figure 5 and Figure 6). Our estimator is consequently acceptable. To make a decision, we choose another example (5.2) in which the distribution of the model is known and standard.

Example 5.2. Consider a linear model with defined functional index by :

$$Y_i = \frac{|\langle \theta, X_i \rangle|}{25} + \epsilon_i, \quad \forall i = 1, \dots, 215.$$

where $(\epsilon_i)_i$ are i.i.d and $\epsilon_i \rightsquigarrow \mathcal{N}(0, 1)$. We keep the values of θ^* and $(X_i)_{i=1, \dots, 215}$ of the previous example (θ is replaced by θ^*).

According to this model, it is clear that, when $X = x$, $Y \rightsquigarrow \mathcal{N}\left(\frac{|\langle \theta, x \rangle|}{25}, 1\right)$.

After the calculation of the errors, we find for our method an error $SSR = 0.0225$, the NPFDA method gives an error $SSR = 0.0367$, while the real error (knowing that $Y \rightsquigarrow \mathcal{N}\left(\frac{|\langle \theta, x \rangle|}{25}, 1\right)$) is equal to $SSR = 2.0938 \cdot 10^{-31}$. This confirms once again that our estimator is much better than that of the NPFDA case. So, in the context of censored data, our estimator is much preferable.

Figure 2: The curves $\theta_{i=1,2}(t_j)$, $t_j \in [0, 1]$

Figure 3: The curves $\theta_{i=1,\dots,20}(t_j)$, $t_j \in [0, 1]$

Example 5.3. In this section we carry out a numerical simulation to evaluate the performance of the proposed method for finite samples data. For this aim, we compare our model CFSIM (2.6) (censored functional single index model) with CNPFDA (5.2) (censored nonparametric functional data analysis) in censored data.

$$\hat{F}_n(t, x) = \frac{\sum_{i=1}^n \frac{\delta_i}{\bar{G}_n(Y_i)} K\left(h_K^{-1} d(x, X_i)\right) H\left(h_H^{-1}(t - Y_i)\right)}{\sum_{i=1}^n K\left(h_K^{-1} d(x, X_i)\right)}. \quad (5.2)$$

Consider the curves generated in the following way:

$$X_i(t) = 1 - \sin\left(W_i\left(t - \frac{\pi}{3}\right)\right), \quad i = 1, \dots, 200; \quad t \in \left[0, \frac{\pi}{3}\right] \quad (5.3)$$

where $W_i = \frac{1}{3} W_{i-1} + \epsilon$ and ϵ and standard normal distribution ($\epsilon \sim \mathcal{N}(0, 1)$) and ($W_0 \sim \mathcal{N}(0, 1)$). We carry out the simulation with a 200 sample of the curves $X(t)$ (see Figure 7).

Figure 4: The curves $\theta_{i=1,\dots,215}(t_j)$, $t_j \in [0, 1]$

The scalar response is defined as $T_i = r(X_i) + \varepsilon_i$, where $r(X_i) = \left(\int_0^{\pi/3} X'_i(t) dt \right)^2$ and $\varepsilon \sim \mathcal{N}(0, 0.075)$. We also simulate n i.i.d rv C_i , $i = 1, \dots, n$ with the exponential distribution $\mathcal{E}(1, 5)$.

Noting that the computation of these estimators are based on the observed data $(X_i, Y_i, \delta_i)_{i=1,\dots,n}$, where $Y_i = \min(T_i, C_i)$ and $\delta_i = I_{\{T_i \leq C_i\}}$. On the other hand, taking into account the smoothness of the curves $X_i(t)$, we choose the semi-metric in $\mathcal{H}$:

$$d(x_i, x_j) = \sqrt{\int_0^{\pi/3} (x'_i(t) - x'_j(t))^2 dt}, \quad x_i, x_j \in \mathcal{H}$$

Then, we choose the quadratic kernel defined by :

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

and the distribution function $H(\cdot)$

$$H(t) = \int_{-\infty}^t \frac{3}{4}(1 - t^2)I_{(-1,1)}(t) dt$$

The bandwidths $h_H \sim h_K =: h$ are automatically selected by cross-validation method based on the K-nearest neighbours ([27]).

In the second part of the simulation studies, we simulate the single functional index model as follows.

First, we choose the functional parameter θ . For a training sample $\mathcal{L} = \{1, \dots, 150\}$, the best approximation of θ is to estimate the eigenfunctions of the covariance operator $\mathbb{E}[(X' - \mathbb{E}(X')) < X', \cdot >_{\mathcal{H}}]$ by its empirical covariance $\frac{1}{L} \sum_{i \in \mathcal{L}} (X'_i - \mathbb{E}(X'))^t (X'_i - \mathbb{E}(X'))$ [6]. Figure 8 show the discretization of all the eigenfunctions $\theta_i(t)$.

Taking θ^* the first eigenfunction corresponding to the first higher eigenvalue, and compute the inner product $< \theta^*, X_1 >, \dots, < \theta^*, X_{200} >$. Then simulate the response variables $Y_i = r(< \theta^*, X_i >) + \varepsilon$, where $r(< \theta^*, X_i >) = \exp(10(< \theta^*, X_i > - 0.05))$ and generate independently the variables $\varepsilon_1, \dots, \varepsilon_{200}$.

In our simulation, sample sizes are $n=200$, which are subdivided into two parts, one is a learning sample of 150 observations and the other 50 observations are a test sample.

Then using the learning sample to compute the estimator of $\hat{Y}_i = \hat{\zeta}_{\theta^*}(\gamma, x)$ and $\hat{Y}_{ni} = \hat{\zeta}(\gamma, x)$ for $j=\{151, \dots, 200\}$.

Figure 5: CR $\sim 3.5\%$, MSE=0.0132 on the left, CR $\sim 18\%$, MSE=0.0611 on the center, and CR $\sim 48\%$, MSE=0.3214 on the right

Figure 6: CR $\sim 3.5\%$, MSE=0.0074 on the left, CR $\sim 18\%$, MSE=0.0191 on the center, and CR $\sim 48\%$, MSE=0.1455 on the right

Figure 9 shows the results by plotting the true values versus the predicted values for the MSE in complete data for both estimators.

Finally, to show how the different censored rates (CRs) affect the prediction results, we present some CRs and their corresponding MSE, which are defined as :

$$CFSIM.MSE = \frac{1}{50} \sum_{i=151}^{200} (Y_i - \hat{Y}_i)^2, \quad CNPFDA.MSE = \frac{1}{50} \sum_{i=151}^{200} (Y_i - \hat{Y}_{ni})^2$$

respectively.

For each explicit value of CRs and n , we carried out 100 independent replications of the experiment, and then computed the average of the mean squared error. These quantities are presented in Table (1).

Figure 7: The curves $X_{i=1,\dots,200}(t)$, $t \in [0, 1]$.

Figure 8: The curves $\theta_{i=1,\dots,150}(t)$, $t \in [0, 1]$.

Example 5.4. This example aims to illustrate our study of forecasting via the conditional quantile. More precisely, we will compare our model CFSIM (2.6) (censored functional single index model) with CNPFDA (5.4) (censored nonparametric functional data analysis) in censored data.

$$\hat{F}_n(t, x) = \frac{\sum_{i=1}^n \frac{\delta_i}{\bar{G}_n(Y_i)} K\left(h_K^{-1} d(x, X_i)\right) H\left(h_H^{-1}(t - Y_i)\right)}{\sum_{i=1}^n K\left(h_K^{-1} d(x, X_i)\right)}. \quad (5.4)$$

We consider a diffusion process on the interval $[0, 1]$:

$$X_i(t) = \cos(\pi b_i t) + a_i t^2, \quad i = 1, \dots, 200; \quad t \in [0, 1] \quad (5.5)$$

Figure 9: Prediction via the conditional quantile for complete data

Table 1: MSE comparison for CFSIM and CNPFDA

n	CR%	CNPFDA	CFSIM
200	4%	0.0784	0.0471
	20%	0.2694	0.1915
	45%	0.7030	0.4844
300	4%	0.0653	0.0237
	20%	0.2235	0.1219
	45%	0.6172	0.3014

where a_i are uniformly distributed on $[0, 1]$ ($a \sim \mathcal{U}(0, 1)$) and $b_i = \frac{1}{3}b_{i-1} + \xi_i$, $\xi_i \sim \mathcal{N}(0, 1)$ are independent from a_i and b_i , which is generated by standard normal distribution ($b_0 \sim \mathcal{N}(0, 1)$). We carry out the simulation with 200 samples of the curves $X(t)$; see Figure 10.

The scalar response is defined as $T_i = r(X_i) + \varepsilon_i$, where $r(X_i) = \left(\int_0^1 X'_i(t) dt \right)^2$ and $\varepsilon \sim \mathcal{N}(0, 0.075)$. We also simulate n i.i.d rv C_i , $i = 1, \dots, n$ with the exponential distribution $\mathcal{E}(1, 5)$.

Noting that the computation of these estimators are based on the observed data $(X_i, Y_i, \delta_i)_{i=1, \dots, n}$, where $Y_i = \min(T_i, C_i)$ and $\delta_i = \mathbf{1}_{\{T_i \leq C_i\}}$. On the other hand, taking into account of the smoothness of the curves $X_i(t)$, we choose the semi-metric in $\mathcal{H}$:

$$d(x_i, x_j) = \sqrt{\int_0^1 (x'_i(t) - x'_j(t))^2 dt}, \quad x_i, x_j \in \mathcal{H}$$

Then, we choose the quadratic kernels defined by :

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

and the distribution function $H(\cdot)$

$$H(t) = \int_{-\infty}^t \frac{3}{4}(1 - t^2)I_{(-1,1)}(t)dt$$

Figure 10: The curves $X_{i=1,\dots,200}(t)$, $t \in [0, 1]$.

The bandwidths $h_H \sim h_K =: h$ are automatically selected by cross-validation method on the K-nearest neighbours ([27]).

In the second part of the simulation study, we simulate the single functional index model as follows.

First, we choose the functional parameter θ . So for a training sample $\mathcal{L} = \{1, \dots, 150\}$, the best approximation of θ is to estimate the eigenfunctions of the covariance operator $\mathbb{E}[(X' - \mathbb{E}(X')) \langle X', \cdot \rangle_{\mathcal{H}}]$ by its empirical covariance $\frac{1}{\mathcal{L}} \sum_{i \in \mathcal{L}} (X'_i - \mathbb{E}(X'))^t (X'_i - \mathbb{E}(X'))$ [6]. Figure 11 shows the discretization of all the eigenfunctions $\theta_i(t)$ respectively.

Figure 11: The curves $\theta_{i=1,\dots,150}(t)$, $t \in [0, 1]$.

Taking θ^* as the first eigenfunction corresponding to the first higher eigenvalue, and computing the inner product $\langle \theta^*, X_1 \rangle, \dots, \langle \theta^*, X_{200} \rangle$, we then simulate the response variables Y_i .

In our simulation, sample sizes are $n=200$, subdivided into two parts, one a learning sample of 150 observations and the other 50 observations are a test sample.

Then using the learning sample to compute the estimator of $\hat{Y}_i = \hat{\zeta}_{\theta^*}(\gamma, x)$ and $\hat{Y}_{ni} = \hat{\zeta}(\gamma, x)$ for $j=\{151, \dots, 200\}$.

Finally we show the results by plotting the true values versus the predicted values for the MSE under censored data for both estimators with different censored rate (CR) 2.6 and 5.2 which are defined as :

$$FSIM.MSE = \frac{1}{50} \sum_{i=151}^{200} (Y_i - \hat{Y}_i)^2$$

$$NPFDA.MSE = \frac{1}{50} \sum_{i=151}^{200} (Y_i - \hat{Y}_{ni})^2$$

respectively.

6 Remarks on the single index

It is well known that one of the main advantages of the single index model is its ability to deal with the problem of high dimensional data. A straightforward example is the optimal convergence rate of type $\mathcal{O}(n^{-2k/(2k+p)})$ for the estimation of a k -times differentiable regression function, this rate goes to zero dramatically slowly if k is small compared to the dimension p of the explanatory variable ($X \in \mathbb{R}^p$). In this regard, Gaïffas and Lecué [28] showed that the optimal rate of convergence of regression function in the single index model, is of order $\mathcal{O}(n^{-2k/(2k+p)})$ (rather than $\mathcal{O}(n^{-2k/(2k+p)})$), thereby answering a conjecture of Stone [44].

The same idea was adopted in the abstract metric spaces by the choice of the semi-metric increasing the probability measure concentration of the explanatory variable in small balls (see Ferraty and Vieu [27]), Section 13.2). Among this family of semi-metrics, one can consider the semi-metric induced by the functional single index estimate.

In the classical nonparametric multivariate regression context the first problem appearing is the so-called curse of dimensionality caused by the sparsity of data in high dimensional spaces.

One possibility to circumvent this problem is to impose additional assumptions on the regression function, in order to reduce the effect of the dimensionality of the regressor variables.

6.1 About the Functional Single Index Estimate

Besides using the pseudo-maximum likelihood method proposed by Attaoui *et al.* (2011) to estimate θ , the cross-validation method presented by Aït-Saïdi *et al.* [2]) for the estimation of θ in the single functional index regression model would be the first natural candidate method for choosing θ in this case.

The second potential approach to estimate θ may be the wild bootstrap method, which was shown by Ferraty *et al.* [19] to estimate the functional nonparametric regression operator. Thus, the theoretical justification and practice have been established.

Among the interesting aspects of our study is to show how the conditional density estimate can be used to derive an estimate of the functional single index if the latter is unknown. The leave-out-one-curve cross-validation procedure was adapted by Aït Saïdi *et al.* [2] to estimate the single index. Recently, Hall and Müller [35] proposed a method for estimating functional derivatives and Ferraty *et al.* [20] adopted this technique to estimate the parameter θ . Alternatively, this parameter can be estimated via the pseudo-maximum likelihood method which is based on the preliminary estimation of the conditional density of Y given X by

$$\hat{\theta} = \arg \max_{\theta \in \Theta_{\mathcal{F}}} \hat{L}(\theta)$$

Figure 12: CR $\sim 3.5\%$, MSE=0.0132 on the left, CR $\sim 18\%$, MSE=0.0611 on the center, and CR $\sim 48\%$, MSE=0.3214 on the right

Figure 13: CR $\sim 3.5\%$, MSE=0.0074 on the left, CR $\sim 18\%$, MSE=0.0191 on the center, and CR $\sim 48\%$, MSE=0.1455 on the right

where

$$\hat{L}(\theta) = \sum_{k=1}^n \log \hat{f}(\theta, Y_k, X_k).$$

Note that, this method has been studied by Delecroix *et al.* [16] in the real case where they showed that this technique has minimal variance among all estimators. The asymptotic optimality of this procedure in functional statistic, is an important prospect of the present work.

As an application, this approach can be used for answering the semi-metric choice question. Indeed, it is well known that, in nonparametric functional statistic, the projection-type semi-metric is very important for

increasing the concentration property. The functional index model is a particular case of this family of semi-metric, because it is based on the projection on one functional direction. So, the estimation procedures of this direction permit us to compute adaptive semi-metrics in the general context of nonparametric functional data analysis. Finally, the theoretical justification and practice should be established.

6.2 Concluding Remarks

Single-index models and varying coefficient models are powerful tools for dimension reduction and semi-parametric modeling because they can effectively avoid the "curse of dimensionality". Both models largely relax some restrictive assumptions on linear and nonparametric models. Specially, the single-index model is usually treated as the first step of the famous projection pursuit regression. The single-index model has also been extensively used in the projection pursuit regression.

For example, firstly the motivation of this model also comes from an analysis of environmental data, consisting of daily measurement of pollutants and other environment factors. Of interest is to examine the association between the levels of pollutants, two environment factors (temperature and relative humidity) and the total number of daily hospital admissions for respiratory problems.

Secondly, single-index models cannot reflect the additivity of covariates, while the performance of varying coefficient models can be poor if the varying coefficient contains moderate or high dimensional covariates.

In standard multivariate regression, the single-index model is a good compromise between nonparametric and parametric regression models when one wishes to regress a response variable on several real-valued explanatory ones. It assumes the existence of a latent one-dimensional explanatory variable which allows us to explain the response through a nonparametric regression model.

In addition, the latent explanatory variable is supposed to be a linear combination of the explanatory variables. The vector of the linear combination is called a "single index". Such models are useful tools for interpreting some situations. This is particularly the case in econometrics where single-index models and various extensions have been intensively studied.

Single-index models, or projection pursuit regression, have proven to be an efficient way of coping with the high dimensional problem in nonparametric regressions. The idea is restricting the general multivariate regression function to a special form.

For the functional single-index models, the literature is closely limited, and only a few theoretical results have been obtained until now.

For the past two decades, the single-index model, a special case of projection pursuit regression, has proven to be an efficient way of coping with the high dimensional problem in nonparametric regression. Here we deal with single index modeling when the explanatory variable is functional. More precisely, we consider the problem of estimating the conditional density of a real variable Y given a functional variable X when the explanation of Y given X is done through its projection on one functional direction.

The single-index approach is widely applied in econometrics as a reasonable compromise between non-parametric and parametric models. Such kind of modeling is intensively studied in the multivariate case.

In practice, this study has great importance, because it permits us to construct a prediction method based on the conditional mode estimator. Moreover, in the case where the functional single index is unknown, our estimate can be used to estimate this parameter via the pseudo-maximum likelihood estimation method. Noting that the estimation of the functional single-index has great interest on the semi-metric choice in non-parametric functional data analysis but it has been not considered in this paper.

7 Proofs of technical lemmas

In order to highlight the main contribution of our paper (i.e. α -mixing and functional variables) some details are voluntarily omitted.

Proof of Lemma 3.1. Let

$$\begin{aligned} |\widehat{F}_N^{(l)}(\theta, t, x) - \widetilde{F}_N^{(l)}(\theta, t, x)| &\leq \frac{h_H^{-l}}{n\mathbb{E}K_1(\theta, x)} \sum_{i=1}^n \left| \frac{\delta_i}{\bar{G}_n(Y_i)} K_i(\theta, x) H_i^{(l)}(t) - \frac{\delta_i}{\bar{G}(Y_i)} K_i(\theta, x) H_i^{(l)}(t) \right| \\ &\leq \frac{h_H^{-l}}{n\mathbb{E}K_1(\theta, x)} \sum_{i=1}^n |\delta_i K_i(\theta, x) H_i(t)| \left| \frac{1}{\bar{G}_n(Y_i)} - \frac{1}{\bar{G}(Y_i)} \right| \\ &\leq \frac{h_H^{-l}}{\phi_{\theta, x}(h_K)} \frac{C}{\bar{G}_n(\tau_F) \bar{G}(\tau_F)} \sup_{t \in \mathcal{S}_{\mathbb{R}}} |\bar{G}_n(t) - \bar{G}(t)| \frac{1}{n} \sum_{i=1}^n \delta_i \end{aligned}$$

In conjunction with the SLLN and the LIL on the censoring law (see Deheuvels and Einmahl (2000)), hypotheses (H2), (H7) and $\left(\frac{\log \log n}{n}\right)^{1/2} = o(\phi_{\theta, x}(h_K))$ complete the proof. $\square$

Proof of Lemma 3.2. The asymptotic behavior of the bias term is standard, in the sense that it is not affected by the dependence structure of the data. We have

$$\begin{aligned} \mathbb{E}\widetilde{F}_N(\theta, t, x) - F(\theta, t, x) &= \frac{1}{\mathbb{E}K_1(x, \theta)} \mathbb{E}\left(\frac{\delta_i}{\bar{G}(Y_i)} K_i(x, \theta) H_i(t)\right) - F(\theta, t, x) \\ &= \frac{1}{\mathbb{E}K_1(x, \theta)} \mathbb{E}\left(\frac{\delta_i}{\bar{G}(Y_i)} K_i(x, \theta) \left[\mathbb{E}(H_i(t) | X_1, \theta >)\right]\right) - F(\theta, t, x). \quad (7.1) \end{aligned}$$

integrating by parts and using the fact that H is a *cdf* and the use a double conditioning with respect to T_1 , we can easily get

$$\begin{aligned} I &= \mathbb{E}\left(\frac{\delta_i}{\bar{G}(Y_i)} H_i(t) | X_1, \theta >\right) \\ &= \mathbb{E}\left(\mathbb{E}\left[\frac{\mathbf{1}_{T_1 \leq C_1}}{\bar{G}(T_1)} H\left(\frac{t - T_1}{h_H}\right) | X_1, \theta >, T_1\right]\right) \\ &= \mathbb{E}\left(\frac{1}{\bar{G}(T_1)} H\left(\frac{t - T_1}{h_H}\right) \mathbb{E}[\mathbf{1}_{T_1 \leq C_1} | T_1] | X_1, \theta >\right) \\ &= \mathbb{E}\left[H\left(\frac{t - T_1}{h_H}\right) | X_1, \theta >\right] \\ &= \int_{\mathbb{R}} H\left(\frac{t - u}{h_H}\right) f(\theta, u, X_1) du, \\ &= \int_{\mathbb{R}} H\left(\frac{t - u}{h_H}\right) dF(\theta, u, X_1), \\ &= \int_{\mathbb{R}} H^{(1)}\left(\frac{t - u}{h_H}\right) F(\theta, u, X_1) du, \\ &= \int_{\mathbb{R}} H^{(1)}(v) F(\theta, t - vh_H, X_1) dv, \\ &= \int_{\mathbb{R}} H^{(1)}(v) (F(\theta, t - vh_H, X_1) - F(\theta, t, x)) dv + F(\theta, t, x) \int_{\mathbb{R}} H^{(1)}(v) dv, \end{aligned}$$

we can write, because of (H3) and (H6)-(i):

$$\begin{aligned} I &\leq C_{x, \theta} \int_{\mathbb{R}} H^{(1)}(v) \left(h_K^{b_1} + |v|^{b_2} h_H^{b_2}\right) dv + F(\theta, t, x) \\ &= \mathcal{O}\left(h_K^{b_1} + h_H^{b_2}\right) + F(\theta, t, x). \end{aligned}$$

Combining this last result with (7.1) allows us to achieve the proof. $\square$

Proof of Lemma 3.3. Using the compactness of $\mathcal{S}_{\mathbb{R}}$, we can write that

$$\mathcal{S}_{\mathbb{R}} \subset \bigcup_{k=1}^{\tau_n} (z_k - l_n, z_k + l_n) \text{ with } l_n \text{ and } \tau_n \text{ can be chosen such that } l_n = C\tau_n^{-1} \sim Cn^{-\zeta-1/2}. \text{ Taking } k_t = \arg \min_{\{z_1, \dots, z_{\tau_n}\}} |t - z_k|.$$

Thus, we have the following decomposition:

$$\begin{aligned} \frac{1}{\widehat{F}_D(\theta, x)} \sup_{t \in \mathcal{S}_{\mathbb{R}}} \left| \widetilde{F}_N(\theta, t, x) - \mathbb{E}\widetilde{F}_N(\theta, t, x) \right| &\leq \frac{1}{\widehat{F}_D(\theta, x)} \sup_{t \in \mathcal{S}_{\mathbb{R}}} \left| \widetilde{F}_N(\theta, t, x) - \widetilde{F}_N(\theta, t_k, x) \right| \\ &\quad + \frac{1}{\widehat{F}_D(\theta, x)} \sup_{t \in \mathcal{S}_{\mathbb{R}}} \left| \widetilde{F}_N(\theta, t_k, x) - \mathbb{E}\widetilde{F}_N(\theta, t_k, x) \right| \\ &\quad + \frac{1}{\widehat{F}_D(\theta, x)} \sup_{t \in \mathcal{S}_{\mathbb{R}}} \left| \mathbb{E}\widetilde{F}_N(\theta, t_k, x) - \mathbb{E}\widetilde{F}_N(\theta, t, x) \right| \\ &\leq B_1 + B_2 + B_3 \end{aligned}$$

On the one hand, as the first and the third terms can be treated in the same manner, we deal only with first term. Making use of (H6)-(i) we get

$$\begin{aligned} \sup_{t \in \mathcal{S}_{\mathbb{R}}} \left| \widetilde{F}_N(\theta, t, x) - \widetilde{F}_N(\theta, t_k, x) \right| &\leq \frac{1}{n\mathbb{E}K_1(\theta, x)} \sup_{t \in \mathcal{S}_{\mathbb{R}}} \sum_{i=1}^n \left| \frac{\delta_i}{\widehat{G}(Y_i)} H_i(t) - \frac{\delta_i}{\widehat{G}(Y_i)} H_i(t_k) \right| |K_i(\theta, x)| \\ &\leq \frac{1}{n\mathbb{E}K_1(\theta, x)} \sup_{t \in \mathcal{S}_{\mathbb{R}}} \sum_{i=1}^n \left| \frac{\delta_i}{\widehat{G}(Y_i)} H_i(t) - \frac{\delta_i}{\widehat{G}_n(Y_i)} H_i(t_k) \right| |K_i(\theta, x)| \\ &\leq \frac{C}{n\mathbb{E}K_1(\theta, x)} \sup_{t \in \mathcal{S}_{\mathbb{R}}} \frac{|t - t_k|}{h_H} \times \left(\sum_{i=1}^n K_i(\theta, x) \left(\frac{1}{\widehat{G}(Y_i)} - \frac{1}{\widehat{G}_n(Y_i)} \right) \right) \\ &\leq \frac{Cl_n}{h_H \widehat{G}_n(a_F) \widehat{G}(a_F)} \sup_{t \in \mathcal{S}_{\mathbb{R}}} |G_n(t) - G(t)| \widehat{F}_D(\theta, x) \end{aligned}$$

Using $l_n = n^{-\zeta-1/2}$ we obtain

$$A_1 \leq \frac{Cn^{-\zeta-1/2}}{h_H \widehat{G}_n(a_F) \widehat{G}(a_F)} \left(\frac{\log n \log n}{n} \right)^{1/2}$$

and note that, because of (H6)-(i), we have

$$\frac{l_n}{h_H} = o \left(\sqrt{\frac{\log n}{n\phi_{\theta, x}(h_K)}} \right).$$

Thus, for n large enough, we have

$$B_1 = \mathcal{O}_{a.co} \left(\sqrt{\frac{\log n}{n\phi_{\theta, x}(h_K)}} \right).$$

Following similar arguments, we can write

$$B_3 \leq B_1.$$

Concerning B_2 , let us consider $\varepsilon = \epsilon_0 \sqrt{\frac{S_{n,2}^2 \log n}{n^2}}$. Since for all $\epsilon_0 > 0$, we have that

$$\begin{aligned} \mathbb{P} \left(\sup_{t \in \mathcal{S}_{\mathbb{R}}} \left| \mathbb{E}\widehat{F}_N(\theta, t_k, x) - \mathbb{E}\widehat{F}_N(\theta, t, x) \right| > \varepsilon \right) &\leq \mathbb{P} \left(\max_{k \in \{1, \dots, \tau_n\}} \left| \mathbb{E}\widehat{F}_N(\theta, t_k, x) - \mathbb{E}\widehat{F}_N(\theta, t, x) \right| > \varepsilon \right) \\ &\leq \tau_n \mathbb{P} \left(\left| \mathbb{E}\widehat{F}_N(\theta, t_k, x) - \mathbb{E}\widehat{F}_N(\theta, t, x) \right| > \varepsilon \right). \end{aligned}$$

The application of Fuk-Nagaev's inequality (see Proposition A.11-ii of Ferraty and Vieu [27]) with $r = (\log n)^2$ and $q = a + 1$, we get that

$$\mathbb{P} \left(\left| \mathbb{E}\widehat{F}_N(\theta, t_k, x) - \mathbb{E}\widehat{F}_N(\theta, t, x) \right| > \varepsilon \right) \leq \left(1 + \frac{\epsilon_0^2}{(\log n)^2} \right)^{-(\log n)^2/2} + n(\log n)^{-2} \left(\frac{\sqrt{\log n}}{\epsilon_0 S_{n,2}} \right)^{a+1}$$

$$\leq C_{\theta,x}F_1 + C'_{\theta,x}F_2.$$

Finally, the use that $s_{n,2}^2 = \mathcal{O}(n\phi_{\theta,x}(h_K))$, allows to get directly that there exist some $\eta > 0$ such that

$$F_1 + F_2 \leq Cn^{-1-\eta}.$$

Finally, we arrive at

$$\sup_{t \in \mathcal{S}_{\mathbb{R}}} |\tilde{F}_N(\theta, t, x) - \mathbb{E}\tilde{F}_N(\theta, t, x)| = \mathcal{O}_{a.co.} \left(\frac{\sqrt{s_{n,2}^2 \log n}}{n} \right)$$

□

Proof of Proposition 3.2. It is omitted as it very similar to that of Proposition 3.1, we use again the same kind of decomposition as (3.3):

$$\begin{aligned} \sup_{t \in \mathcal{S}_{\mathbb{R}}} |\hat{F}^{(l)}(\theta, t, x) - F^{(l)}(\theta, t, x)| &= \frac{1}{\hat{F}_D(\theta, x)} \sup_{t \in \mathcal{S}_{\mathbb{R}}} \left\{ |\hat{F}_N^{(l)}(\theta, t, x) - \tilde{F}_N^{(l)}(\theta, t, x)| \right\} \\ &+ \frac{1}{\hat{F}_D(\theta, x)} \sup_{t \in \mathcal{S}_{\mathbb{R}}} \left\{ |\tilde{F}_N^{(l)}(\theta, t, x) - \mathbb{E}\tilde{F}_N^{(l)}(\theta, t, x)| \right\} \\ &+ \frac{1}{\hat{F}_D(\theta, x)} \sup_{t \in \mathcal{S}_{\mathbb{R}}} \left\{ |\mathbb{E}\tilde{F}_N^{(l)}(\theta, t, x) - F^{(l)}(\theta, t, x)| \right\} \\ &+ \frac{F^{(l)}(\theta, t, x)}{\hat{F}_D(\theta, x)} \sup_{t \in \mathcal{S}_{\mathbb{R}}} |1 - \hat{F}_D(\theta, x)|. \end{aligned} \quad (7.2)$$

This proof is very similar to the one of Proposition 3.1. Indeed the proof of the Proposition 3.2 by making use the decomposition (7.2), in conjunction of first part of Lemma 3.3 with lemmas Lemma 3.1, Lemma 3.4 and Lemma 3.5.

□

Proof of Lemma 3.4. Using the same arguments in the proof of Lemma 3.2, replacing $F(\theta, t, x)$ (resp. $\tilde{F}(\theta, t, x)$) with $F^{(l)}(\theta, t, x)$ (resp. $\tilde{F}_N^{(l)}(\theta, t, x)$) and considering hypotheses (H5) and (H6) we get:

$$F^{(l)}(\theta, t, x) - \mathbb{E}\tilde{F}_N^{(l)}(\theta, t, x) = \mathcal{O} \left(h_K^{b_1} + h_K^{b_2} \right). \quad (7.3)$$

□

Proof of Lemma 3.5. To get the asymptotic behaviour of the quantity $\tilde{F}_N^{(l)}(\theta, t, x) - \mathbb{E}\tilde{F}_N^{(l)}(\theta, t, x)$, we comeback to the proof of Lemma 3.3, and we replace $F(\theta, t, x)$ (resp. $\tilde{F}(\theta, t, x)$) with $F^{(l)}(\theta, t, x)$ (resp. $\tilde{F}_N^{(l)}(\theta, t, x)$).

Note that (H6)-(ii) and (H12) permit to show that

$$\mathbb{E} \left(\frac{\delta_i}{\bar{G}(Y_i)} H^{(l)} \left(h_H^{-1}(t - Y_i) \right) \frac{\delta_j}{\bar{G}(Y_j)} H^{(l)} \left(h_H^{-1}(t - Y_j) \right) | (X_i, X_j) \right) = \mathcal{O}(h_H^2),$$

while (H1) and (H5) imply that

$$\mathbb{E} \left(\frac{\delta_i}{\bar{G}(Y_i)} H^{(l)} \left(h_H^{-1}(t_{\theta}(\gamma) - Y_i) \right) | X_i \right) = \mathcal{O}(h_H).$$

Consequently, we have by using successively (H0), (H2), (H8) and (H10)-(i)

$$\text{Cov}(\Xi_i(\theta, t, x), \Xi_j(\theta, t, x)) = \mathcal{O} \left(h_H^2 \left(\frac{\phi_{\theta,x}(h_K)}{n} \right)^{1/a} \right),$$

where

$$\Xi_i = h_H^{-l} \frac{\delta_i}{\bar{G}(Y_i)} H_i^{(l)}(t) \Delta_i(x, \theta) - \mathbb{E} h_H^{-l} \frac{\delta_i}{\bar{G}(Y_i)} H_i^{(l)}(t) \Delta_i(x, \theta)$$

has zero mean and satisfies $|\Xi_i(\theta, t, x)| \leq Ch_H^{-l} \phi_{\theta, x}^{-1}(h_K)$. Now, we show that because $H^{(l)}$ is bounded. Indeed, it can be found that

$$\tilde{F}_N^{(l)}(\theta, t, x) - \mathbb{E}\tilde{F}_N^{(l)}(\theta, t, x) = \frac{1}{n} \sum_{i=1}^n \Xi_i(\theta, t, x),$$

then it allows us to use directly similar arguments of second part by Lemma 3.3, we obtain

$$\tilde{F}_N^{(l)}(\theta, t, x) - \mathbb{E}\tilde{F}_N^{(l)}(\theta, t, x) = \mathcal{O}_{a.co.} \left(\frac{\sqrt{s_{n,2} \log n}}{n} \right),$$

which leads directly to the result of Lemma 3.5 □

Proof of Lemma 4.3. The proof can be completed following the same steps as Lemma 3.4. □

Proof of Lemma 4.4. Similar to the proof of Lemma 3.4 in Ling *et al.* (2014), it can be completed easily. Here we omit its proof. □

Proof of Lemma 4.5. For all $x \in \mathcal{S}_{\mathcal{H}}$ and $\forall \theta \in \Theta_{\mathcal{H}}$, we set

$$k(x) = \arg \min_{k \in \{1 \dots d_n^{\mathcal{S}_{\mathcal{H}}}\}} \|x - x_k\| \text{ and } q(\theta) = \arg \min_{m \in \{1 \dots d_n^{\Theta_{\mathcal{H}}}\}} \|\theta - \theta_q\| \text{ and by the compact property of } \mathcal{S}_{\mathbb{R}} \subset \mathbb{R},$$

we have $\mathcal{S}_{\mathbb{R}} \subset \bigcup_{k=1}^{\tau_n} (v_k - l_n, v_k + l_n)$ with l_n and τ_n can be selected such as $l_n = \mathcal{O}(\tau_n^{-1}) = \mathcal{O}(n^{-(3\zeta+1)/2})$. Taking

$$k_t = \arg \min_{\{v_1, \dots, v_{\tau_n}\}} |t - v_k|.$$

Let us consider the following decomposition

$$\begin{aligned} \sup_{\theta \in \Theta_{\mathcal{H}}} \sup_{x \in \mathcal{S}_{\mathcal{H}}} \sup_{t \in \mathcal{S}_{\mathbb{R}}} \left| \tilde{F}_N^{(l)}(\theta, t, x) - \mathbb{E} \left(\tilde{F}_N^{(l)}(\theta, t, x) \right) \right| &\leq \sup_{\theta \in \Theta_{\mathcal{H}}} \sup_{x \in \mathcal{S}_{\mathcal{H}}} \sup_{t \in \mathcal{S}_{\mathbb{R}}} \left\{ \left| \tilde{F}_N^{(l)}(\theta, t, x) - \tilde{F}_N^{(l)}(\theta, t, x_{k(x)}) \right| \right. \\ &\quad + \left| \tilde{F}_N^{(l)}(\theta, t, x) - \tilde{F}_N^{(l)}(\theta, t, x_{k(x)}) \right| \\ &\quad + \left| \tilde{F}_N^{(l)}(\theta_{q(\theta)}, t, x_{k(x)}) - \tilde{F}_N^{(l)}(\theta_{q(\theta)}, v_{k_t}, x_{k(x)}) \right| \\ &\quad + \left| \tilde{F}_N^{(l)}(\theta_{q(\theta)}, v_{k_t}, x_{k(x)}) - \mathbb{E} \left(\tilde{F}_N^{(l)}(\theta_{q(\theta)}, v_{k_t}, x_{k(x)}) \right) \right| \\ &\quad + \left| \mathbb{E} \left(\tilde{F}_N^{(l)}(\theta_{q(\theta)}, v_{k_t}, x_{k(x)}) \right) - \mathbb{E} \left(\tilde{F}_N^{(l)}(\theta_{q(\theta)}, t, x_{k(x)}) \right) \right| \\ &\quad + \left| \mathbb{E} \left(\tilde{F}_N^{(l)}(\theta_{q(\theta)}, t, x_{k(x)}) \right) - \mathbb{E} \left(\tilde{F}_N^{(l)}(\theta, t, x_{k(x)}) \right) \right| \\ &\quad + \left. \left| \mathbb{E} \left(\tilde{F}_N^{(l)}(\theta, t, x_{k(x)}) \right) - \mathbb{E} \left(\tilde{F}_N^{(l)}(\theta, t, x) \right) \right| \right\} \\ &\leq \Psi_1 + \Psi_2 + \Psi_3 + \Psi_4 + \Psi_5 + \Psi_6 + \Psi_7. \end{aligned}$$

• Concerning Ψ_3 and Ψ_5 ; by conditions (H6)-(ii) and (A6), boundness of K , we obtain

$$\begin{aligned} \left| \tilde{F}_N^{(l)}(\theta_{q(\theta)}, t, x_{k(x)}) - \tilde{F}_N^{(l)}(\theta_{q(\theta)}, v_{k_t}, x_{k(x)}) \right| &\leq \frac{1}{nh_H^l \mathbb{E}K_1(\theta, x)} \sup_{t \in \mathcal{S}_{\mathbb{R}}} \sum_{i=1}^n \left| \frac{\delta_i}{\bar{G}(Y_i)} K_i(\theta_{q(\theta)}, x_{k(x)}) \right| \\ &\quad \left| H^{(l)} \left(\frac{t - Y_i}{h_H} \right) H^{(l)} \left(\frac{v_{k_t} - Y_i}{h_H} \right) \right| \\ &\leq \sup_{t \in \mathcal{S}_{\mathbb{R}}} C \frac{|t - v_{k_t}|}{h_H^{l+1}} \left(\frac{1}{n \mathbb{E}(K_1(\theta_{q(\theta)}, x_{k(x)}))} \right. \\ &\quad \left. \sum_{i=1}^n \left| K_i(\theta_{q(\theta)}, x_{k(x)}) \frac{1}{\bar{G}(Y_i)} \right| \right) \\ &\leq \frac{Cl_n}{\phi(h_K) h_H^{l+1}} = \mathcal{O} \left(\frac{l_n}{h_H^{l+1} \phi(h_K)} \right). \end{aligned}$$

Now, the fact that $\lim_{n \rightarrow \infty} n^\nu h_H = \infty$, and choosing $l_n = n^{-(3\nu+1)/2}$ and using the second part of (A6), imply that

$$\frac{l_n}{h_H^{l+1} \phi(h_K)} = o \left(\sqrt{\frac{\log n}{nh_H^{2l-1} \phi(h_K)}} \right)$$

as $n \rightarrow \infty$, therefore, it follows

$$\Psi_5 \leq \Psi_3 = \mathcal{O}_{a.co.} \left(\sqrt{\frac{\log d_n^{\mathcal{S}_{\mathcal{H}}} d_n^{\Theta_{\mathcal{H}}}}{nh_H^{2l-1} \phi(h_K)}} \right).$$

Concerning Ψ_4 , let us consider $\varepsilon = \varepsilon_0 \sqrt{\frac{s_{n,8}^2 \log d_n^{\mathcal{S}_{\mathcal{F}}} d_n^{\Theta_{\mathcal{F}}}}{n^2}}$. Since

$$\begin{aligned} \mathbb{P} \left(\Psi_4 > \varepsilon_0 \sqrt{\frac{s_{n,8}^2 \log d_n^{\mathcal{S}_{\mathcal{F}}} d_n^{\Theta_{\mathcal{F}}}}{n^2}} \right) &= \mathbb{P} \left(\max_{q \in \{1 \dots d_n^{\mathcal{S}_{\mathcal{H}}}\}} \max_{k \in \{1 \dots d_n^{\mathcal{S}_{\mathcal{H}}}\}} \max_{k_t \in \{1, 2, \dots, \tau_n\}} \left| \Gamma_i^{(l)} - \mathbb{E} \Gamma_i^{(l)} \right| > \varepsilon \right) \\ &\leq \tau_n d_n^{\mathcal{S}_{\mathcal{H}}} d_n^{\Theta_{\mathcal{H}}} \mathbb{P} \left(\left| \Gamma_i^{(l)} - \mathbb{E} \Gamma_i^{(l)} \right| > \varepsilon \right) \end{aligned} \quad (7.4)$$

the application of Fuk-Nagaev's inequality (Proposition A.11-(ii), see Ferraty and Vieu [27]) with $r = (\log n)^2 > 1$ and $q = a + 1$, we get

$$\begin{aligned} \mathbb{P} \left(\left| \sum_{i=1}^n (\Gamma_i^{(l)} - \mathbb{E} \Gamma_i^{(l)}) \right| > \varepsilon \right) &\leq C \left(1 + \frac{\varepsilon_0^2 (\log d_n^{\mathcal{S}_{\mathcal{F}}} d_n^{\Theta_{\mathcal{F}}})}{(\log n)^2} \right)^{-(\log n)^2/2} \\ &\quad + \frac{n(\log n)^{2a} \varepsilon_0^{-(a+1)}}{(\log d_n^{\mathcal{S}_{\mathcal{F}}} d_n^{\Theta_{\mathcal{F}}})^{(a+1)/2} s_{n,8}^{a+1}} \\ &\leq C(J_1 + J_2). \end{aligned} \quad (7.5)$$

By assumptions (A6), it has $\frac{\log d_n^{\mathcal{S}_{\mathcal{H}}} d_n^{\Theta_{\mathcal{H}}}}{(\log n)^2} \rightarrow \infty$ as $n \rightarrow \infty$, which leads to

$$J_1 \leq \left(d_n^{\mathcal{S}_{\mathcal{H}}} d_n^{\Theta_{\mathcal{H}}} \right)^{-\lambda} \quad (7.6)$$

for some $\lambda > 1$ and $\varepsilon_0 > 0$ such that $\varepsilon_0^2 = 2\lambda$.

On the other hand,

$$J_2 \leq Cn (\log n)^{2a} \left(d_n^{\mathcal{S}_{\mathcal{H}}} d_n^{\Theta_{\mathcal{H}}} \right)^{-\frac{a+1}{2}} n^{-\beta} \leq C' n^{1+\eta-\beta} \quad (7.7)$$

where $\eta > 0$ such that $\beta > \beta - \eta > 2$. Meanwhile, by the selection of λ and β , we can find that

$$\left(d_n^{\mathcal{S}_{\mathcal{H}}} d_n^{\Theta_{\mathcal{H}}} \right)^\lambda = \mathcal{O} \left(n^{\beta-\eta-1} \right) \quad (7.8)$$

Combining (7.4)-(7.8) with assumption (A6)-(ii)

$$\Psi_4 = \mathcal{O}_{a.co.} \left(\sqrt{\frac{s_{n,8}^2 \log d_n^{\mathcal{S}_{\mathcal{F}}}}{n^2}} \right). \quad (7.9)$$

• Concerning Ψ_1 and Ψ_2

$$\begin{aligned} \sup_{\theta \in \Theta_{\mathcal{H}}} \sup_{x \in \mathcal{S}_{\mathcal{H}}} \sup_{t \in \mathcal{S}_{\mathbb{R}}} |\tilde{F}_N^{(l)}(\theta, t, x) - \tilde{F}_N^{(l)}(\theta, t, x_{k(x)})| &\leq \frac{1}{nh_H^l \mathbb{E} K_1(\theta, x)} \sup_{\theta \in \Theta_{\mathcal{H}}} \sup_{x \in \mathcal{S}_{\mathcal{H}}} \sup_{t \in \mathcal{S}_{\mathbb{R}}} \sum_{i=1}^n \left| \frac{\delta_i}{\bar{G}(Y_i)} \right| \left| H_i^{(l)(t)} \right| \\ &\quad \left| (K_i(\theta, x) - K_i(\theta, x_k)) \right| \\ &\leq \frac{1}{nh_H^l \phi(h_K)} \sup_{x \in \mathcal{S}_{\mathcal{H}}} \sup_{\theta \in \Theta_{\mathcal{H}}} \sum_{i=1}^n |\Delta_i(x, \theta) - \Delta_i(x_{k(x)}, \theta)| \end{aligned}$$

$$\begin{aligned}
&\leq \frac{1}{h_H^l \phi(h_K)} \sup_{x \in \mathcal{S}_{\mathcal{J}_C}} \sup_{\theta \in \Theta_{\mathcal{J}_C}} \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{B_\theta(x, h) \cup B_{\theta}(x_{k(x)}, h)}(X_i) \\
&\leq \frac{C}{h_H^l} \sup_{x \in \mathcal{S}_{\mathcal{J}_C}} \sup_{\theta \in \Theta_{\mathcal{J}_C}} \frac{1}{n} \sum_{i=1}^n \Lambda_i(x, \theta)
\end{aligned}$$

Therefore, similar to the arguments for (7.9), we can get that

$$\Psi_1 = \mathcal{O}_{a.co.} \left(\sqrt{\frac{s_{n,4}^2 \log d_n^{\mathcal{S}_{\mathcal{F}}}}{n^2}} \right).$$

$$\begin{aligned}
\sup_{\theta \in \Theta_{\mathcal{J}_C}} \sup_{x \in \mathcal{S}_{\mathcal{J}_C}} \sup_{t \in \mathcal{S}_{\mathbb{R}}} |\tilde{F}_N^{(l)}(\theta, t, x) - \tilde{F}_N^{(l)}(\theta_{q(\theta)}, t, x_{k(x)})| &\leq \frac{h_H^{-l}}{n \mathbb{E} K_1(\theta, x)} \sup_{\theta \in \Theta_{\mathcal{J}_C}} \sup_{x \in \mathcal{S}_{\mathcal{J}_C}} \sup_{t \in \mathcal{S}_{\mathbb{R}}} \sum_{i=1}^n \left| \frac{\delta_i}{\bar{G}(Y_i)} \right| |H_i^{(l)}(t)| \\
&\quad |(K_i(\theta, x_k) - K_i(\theta_{q(\theta)}, x_k))| \\
&\leq \frac{Ch_H^{-l}}{n \phi(h_K)} \sup_{x \in \mathcal{S}_{\mathcal{J}_C}} \sup_{\theta \in \Theta_{\mathcal{J}_C}} \sum_{i=1}^n |\Delta_i(\theta, x_k) - \Delta_i(\theta_{q(\theta)}, x_{k(x)})| \\
&\leq \frac{Ch_H^{-l}}{\phi(h_K)} \sup_{x \in \mathcal{S}_{\mathcal{J}_C}} \sup_{\theta \in \Theta_{\mathcal{J}_C}} \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{B_\theta(x_k, h) \cup B_{\theta_{q(\theta)}}(x_{k(x)}, h)}(X_i) \\
&\leq \frac{C}{h_H^l} \sup_{x \in \mathcal{S}_{\mathcal{J}_C}} \sup_{\theta \in \Theta_{\mathcal{J}_C}} \frac{1}{n} \sum_{i=1}^n \Omega_i(x, \theta)
\end{aligned}$$

Similar to the deduce of (7.9), it yields

$$\Psi_2 = \mathcal{O}_{a.co.} \left(\sqrt{\frac{s_{n,5}^2 \log d_n^{\mathcal{S}_{\mathcal{F}}}}{n^2}} \right).$$

On the other hand, since $\Psi_7 \leq \Psi_1$ and $\Psi_6 \leq \Psi_2$, it also leads to

$$\Psi_6 = \mathcal{O}_{a.co.} \left(\sqrt{\frac{s_{n,4}^2 \log d_n^{\mathcal{S}_{\mathcal{F}}}}{n^2}} \right).$$

and

$$\Psi_7 = \mathcal{O}_{a.co.} \left(\sqrt{\frac{s_{n,5}^2 \log d_n^{\mathcal{S}_{\mathcal{F}}}}{n^2}} \right).$$

Then the proof of Lemma 4.5 can be completed. $\square$

Acknowledgement: Authors are thankful to the anonymous referees for its careful checking and invaluable comments that improved this paper.

References

- [1] Ait-Saïdi, A., F. Ferraty, and R. Kassa (2005). Single functional index model for a time series. *Rev. Roumaine Math. Pures Appl.* 50(4), 321–330.
- [2] Ait-Saïdi, A., F. Ferraty, R. Kassa, and P. Vieu (2008). Cross-validated estimations in the single-functional index model. *Statistics* 42(6), 475–494.
- [3] Aneiros-Pérez, G., H. Cardot, G. Estévez-Pérez, and P. Vieu (2004). Maximum ozone concentration forecasting by functional non-parametric approaches. *Environmetrics* 15(7), 675–685.

- [4] Attaoui, S., A. Laksaci, and E. Ould-Saïd (2011). A note on the conditional density estimate in the single functional index model. *Statist. Probab. Lett.* 81(1), 45–53.
- [5] Attaoui, S. (2014a). On the nonparametric conditional density and mode estimates in the single functional index model with strongly mixing data. *Sankhyā A* 76(2), 356–378.
- [6] Attaoui, S. (2014b). Strong uniform consistency rates and asymptotic normality of conditional density estimator in the single functional index modeling for time series data. *ASIA Adv. Stat. Anal.* 98(3), 257–286.
- [7] Benhenni, K., S. Hedli-Grice, M. Rachdi, and P. Vieu (2008). Consistency of the regression estimator with functional data under long memory conditions. *Statist. Probab. Lett.* 78(8), 1043–1049.
- [8] Besse, P. C., H. Cardot, and D. B. Stephenson (2000). Autoregressive forecasting of some functional climatic variations. *Scand. J. Statist.* 27(4), 673–687.
- [9] Bouchentouf, A. A., T. Djebbouri, A. Rabhi, and K. Sabri (2014). Strong uniform consistency rates of some characteristics of the conditional distribution estimator in the functional single-index model. *Appl. Math. (Warsaw)* 41(4), 301–322.
- [10] Bosq, D. (2000). *Linear Processes in Function Spaces*. Springer, New York.
- [11] Bosq, D. and J.-P. Lecoutre (1987). *Théorie de l'Estimation Fonctionnelle*. Economica, Paris.
- [12] Cai, Z. (2002). Regression quantiles for time series. *Econometric Theory* 18(1), 169–192.
- [13] Chen, D., P. Hall, and H.-G. Müller (2011). Single and multiple index functional regression models with nonparametric link. *Ann. Statist.* 39(3), 1720–1747.
- [14] Dabo-Niang, S., F. Ferraty, and P. Vieu (2004). Nonparametric unsupervised classification of satellite wave altimeter forms. In J. Antoch (Ed.), *COMPSTAT 2004 - Proceedings in Computational Statistics*, pp. 879–886. Physica, Heidelberg.
- [15] Damon, J. and S. Guillas (2002). The inclusion of exogenous variables in functional autoregressive ozone forecasting. *Environmetrics* 13(7), 759–774.
- [16] Delecroix, M., W. Härdle, and M. Hristache (2003). Efficient estimation in conditional single-index regression. *J. Multivariate Anal.* 86(2), 213–226.
- [17] Fernández de Castro, B., S. Guillas, and W. González Manteiga (2005). Functional samples and bootstrap for predicting sulfur dioxide levels. *Technometrics* 47(2), 212–222.
- [18] Ferraty, F., A. Goia, and P. Vieu (2002). Functional nonparametric model for times series: a fractal approach for dimensional reduction. *Test* 11(2), 317–344.
- [19] Ferraty, F., A. Laksaci, A. Tadj, and P. Vieu (2010). Rate of uniform consistency for nonparametric estimates with functional variables. *J. Statist. Plann. Inference* 140(2), 335–352.
- [20] Ferraty, F., J. Park, and P. Vieu (2011). Estimation of a functional single index model. In F. Ferraty (Ed.), *Recent Advances in Functional Data Analysis and Related Topics*, pp. 111–116. Physica, Heidelberg.
- [21] Ferraty, F., A. Peuch, and P. Vieu (2003). Modèle à indice fonctionnel simple. *C. R. Math. Acad. Sci. Paris* 336(12), 1025–1028.
- [22] Ferraty, F., A. Rabhi, and P. Vieu (2005). Conditional quantiles for dependent functional data with application to the climatic El Niño phenomenon. *Sankhyā* 67(2), 378–398.
- [23] Ferraty, F. and P. Vieu (2002). The functional nonparametric model and application to spectrometric data. *Comput. Statist.* 17(4), 545–564.
- [24] Ferraty, F. and P. Vieu (2003a). Curves discrimination: a nonparametric functional approach. *Comput. Statist. Data Anal.* 44(1-2), 161–173.
- [25] Ferraty, F. and P. Vieu (2003b). Functional nonparametric statistics: a double infinite dimensional framework. In M. G. Akritas and D. N. Politis (Eds.), *Recent Advances and Trends in Nonparametric Statistics*, pp. 61–76. Elsevier, Amsterdam.
- [26] Ferraty, F. and P. Vieu (2004). Nonparametric models for functional data, with application in regression, time-series prediction and curve discrimination. *J. Nonparametr. Stat.* 16(1-2), 111–125.
- [27] Ferraty, F. and P. Vieu (2006). *Nonparametric Functional Data Analysis*. Springer, New York.
- [28] Gaïffas, S. and G. Lecué (2007). Optimal rates and adaptation in the single-index model using aggregation. *Electron. J. Stat.* 1, 538–573.
- [29] Gannoun, A., J. Saracco, and K. Yu (2003). Nonparametric prediction by conditional median and quantiles. *J. Statist. Plann. Inference* 117(2), 207–223.
- [30] Gasser, T., P. Hall, and B. Presnell (1998). Nonparametric estimation of the mode of a distribution of random curves. *J. R. Stat. Soc. Ser. B Stat. Methodol.* 60(4), 681–691.
- [31] Goia, A. and P. Vieu (2014). Some advances on semi-parametric functional data modelling. In E. G. Bongiorno, A. Goia, E. Salinelli, and P. Vieu (Eds.), *Contributions in Infinite-Dimensional Statistics and Related Topics*, pp. 135–140. Esculapio, Bologna.
- [32] Goia, A. and P. Vieu (2015). A partitioned single functional index model. *Comput. Statist.* 30(3), 673–692.
- [33] Hall, P., D. S. Poskitt, and B. Presnell (2001). A functional data-analytic approach to signal discrimination. *Technometrics* 43(1), 1–9.
- [34] Hall, P. and N. Heckman (2002). Estimating and depicting the structure of a distribution of random functions. *Biometrika* 89(1), 145–158.
- [35] Hall, P., H.-G. Müller, and F. Yao (2009). Estimation of functional derivatives. *Ann. Statist.* 37(6A), 3307–3329.
- [36] Lecoutre, J.-P. and E. Ould-Saïd (1995). Hazard rate estimation for strong-mixing and censored processes. *J. Nonparametr. Stat.* 5(1), 83–89.

- [37] Ling, N., Z. Li, and W. Yang (2014). Conditional density estimation in the single functional index model for α -mixing functional data. *Comm. Statist. Theory Methods* 43(3), 441–454.
- [38] Padgett, W. J. (1988). Nonparametric estimation of density and hazard rate functions when samples are censored. In P. R. Krishnaiah and C. R. Rao (Eds.), *Handbook of Statistics* 7, pp. 313–331. Elsevier, Amsterdam.
- [39] Ramsay, J. O. and B. W. Silverman (1997). *Functional Data Analysis*. Springer, New York.
- [40] Ramsay, J. O. and B. W. Silverman (2002). *Applied Functional Data Analysis*. Springer, New York.
- [41] Rio, E. (2000). *Théorie Asymptotique des Processus Aléatoires Faiblement Dépendants*. Springer, Heidelberg.
- [42] Roussas, G. G. (1969). Nonparametric estimation of the transition distribution function of a Markov process. *Ann. Math. Statist.* 40(4), 1386–1400.
- [43] Samanta, M. (1989). Nonparametric estimation of conditional quantiles. *Statist. Probab. Lett.* 7(5), 407–412.
- [44] Stone, C. J. (1982). Optimal global rates of convergence for nonparametric regression. *Ann. Statist.* 10(4), 1040–1053.
- [45] Tanner, M. A. and W.H. Wong (1983). The estimation of the hazard function from randomly censored data by the kernel method. *Ann. Statist.* 11(3), 989–993.
- [46] Van Keilegom, I. and N. Veraverbeke (2001). Hazard rate estimation in nonparametric regression with censored data. *Ann. Inst. Statist. Math.* 53(4), 730–745.
- [47] Wang, H. J. and Y. S. Zhao (1999). A kernel estimator for conditional t -quantiles for mixing samples and its strong uniform convergence (in chinese). *Math. Appl. (Wuhan)* 12(3), 123–127.
- [48] Zhou, Y. and H. Liang (2003). Asymptotic properties for L_1 norm kernel estimator of conditional median under dependence. *J. Nonparametr. Stat.* 15(2), 205–219.