

ASYMPTOTIC NORMALITY SINGLE FUNCTIONAL INDEX QUANTILE REGRESSION UNDER RANDOMLY CENSORED DATA

ABDESSAMAD DIB¹, MOHAMED MEHDI HAMRI², ABBES RABHI³

Manuscript received: 16.08.2022; Accepted paper: 02.11.2022;

Published online: 30.12.2022.

Abstract. *The main objective of this paper is to estimate non-parametrically the quantiles of a conditional distribution based on the single-index model in the censorship model when the sample is considered as an independent and identically distributed (i.i.d.) random variables. First of all, a kernel type estimator for the conditional cumulative distribution function (cond-cdf) is introduced. Afterwards, we give an estimation of the quantiles by inverting this estimated cond-cdf, the asymptotic properties are stated when the observations are linked with a single-index structure. Finally, a simulation study is carried out to evaluate the performance of this estimate.*

Keywords: *Asymptotic normality; conditional quantile; functional single-index process; functional random variable; nonparametric estimation; small ball probability.*

1. INTRODUCTION

The estimation of a conditional model, because of the variety of its application possibilities, is an important question in statistics. This subject can (and must) be approached from several angles depending on the complexity of the problem posed: the possible presence of censorship in the observed sample (a common phenomenon in medical applications for example), the possible presence of dependence between the observed variables (e.g. a common phenomenon in seismological, and econometric applications), and the presence of explanatory variables. Many techniques have been studied in the literature to deal with these different situations, but they all only deal with real or multi-dimensional explanatory random variables.

The technical progress made in the collection and storage of data make it possible to have more and more often functional statistical data: curves, images, tables, etc. These data are modeled as being the realizations of a random variable taking its values in an abstract space of infinite dimension, and the scientific community has naturally been interested in recent years in the development of statistical tools capable of processing this type of sample.

Thus, the estimation of conditional models in the presence of a functional explanatory variable from a simple index regression model is a topical question to which this article proposes to provide a first element of answer. After a brief bibliographic overview presented in Section 1, the conditional model for a functional explanatory variable is presented in Section 2, to which this work proposes an extension of the simple index model, when considering an explanatory random variable with values in an infinite dimensional space.

¹ University Djillali LIABES, Laboratory of Mathematics, 22000 Sidi Bel Abbès, Algeria.

E-mail: dibabdassamad@gmail.com.

² University Djillali LIABES, LABRI, ESI, 22000 Sidi Bel Abbès, Algeria. E-mail: hamri.m.m@gmail.com.

³ University Djillali LIABES, 22000 Sidi Bel Abbès, Algeria. E-mail: rabhi_abbès@yahoo.fr.

Such a model was designated generically by a simple functional index model. Naturally, these methods have some drawbacks, and to overcome these, an alternative approach is naturally provided by semi-parametric modeling which supposes the introduction of a parameter on the regressors – these models are known in the literature as simple index models, which have two major advantages. First it is possible to generalize existing models, and then to remedy the problems of the scourge of dimension. These are single revealing direction models (or simple functional index models), and have the advantage of specifying the model to a minimum. The authors used a non-parametric link function having previously determined linear combinations of explanatory variables which contain the maximum information, thus alleviating the scourge of dimension. The idea of these models, in the case of conditional density estimation or regression, consists in reducing to covariates of a dimension smaller than the dimension of the space of variables, thus making it possible to overcome the problem of the scourge of dimension. These models make it possible to obtain a compromise between a parametric model, generally too restrictive, and a non-parametric model where the speed of convergence of the estimators deteriorates quickly in the presence of a large number of explanatory variables, for example, in the partially linear model one decomposes the quantity which one seeks to estimate, in a linear part and a functional part. This last quantity does not pose an estimation problem since it is expressed as a function of the explanatory variables of the defined dimension, thus avoiding the problems linked to the scourge of dimension.

This work proposes an extension of the simple index model when considering an explanatory random variable with values in an infinite dimensional space. Such a model is generically referred to as a model with a simple functional index. The main contribution of this paper lies in a double generalization of the simple index model. On the one hand, the authors place themselves in a framework of functional random variables, and on the other, introduce hypotheses on the law of the explanatory random variable that are less restrictive than those usually used in the vector framework.

First point convergence results were established. The non-parametric method only considers regularity assumptions. Naturally, these methods have some drawbacks, therefore an alternative approach was provided by semi-parametric modeling which supposes the introduction of a parameter on the regressors, these models are known in the literature as simple index models, with two major advantages which, firstly makes it possible to generalize the already existing models, and then to remedy the problems of the scourge of the dimension.

Non-parametric methods based on convolution kernel ideas, which are known to perform well in model estimation problems (conditional or not), are thus widely used in non-parametric estimation of conditional models. A wide range of literature in this area is provided by the bibliographic reviews of [1-5]. The immediate consequence of progress in data collection processes is to offer statisticians the opportunity to increasingly have observations of functional variables. In [6] and [7] it is proposed a wide range of statistical methods, parametric or non-parametric, recently developed to deal with various estimation problems involving functional random variables (i.e., with values in a space of infinite dimension). To date, such statistical developments for directionally revealing functional variables occurred rarely in this context, despite the obvious potential for their application. In practice, in medical applications in particular, one may be in the presence of censored variables. This problem is usually modeled by considering positive variable C called (censorship), and the observed random variables. Such censoring models have been extensively studied in the literature on real and multidimensional random variables, and in non-parametric frameworks, particularly in kernel techniques (see [8-11]), for not a necessarily exhaustive sample of the literature in this field.

Other authors have been interested in the estimation of conditional models from censored or truncated observations (see, e.g. [12-18]). Many statistical applications had to

involve a variable of duration denoted T , designating the time elapsed until the occurrence of the event of interest. These types of variables are observed in various fields such as in reliability (first failure for a machine, lifespan of a material, etc.), in medicine (death or remission for a patient, etc.) in economics and insurance (duration of unemployment, time between two successive breakdowns of a device, etc.). A specificity of these models is the existence of incomplete observations, for which the variable of interest is not completely observed for all the data in the sample. This work studied models where the duration is likely to be right-censored by then calling on techniques adapted to this type of context to take into account the censored observations without losing too much information on it. This study was only interested in the case of right-censored random data. This corresponds to the model frequently used in practice. For example, during a therapeutic trial this can be caused by a loss of sight (the patient leaves the study in progress), the stopping or the change of a treatment, in which case the patients are excluded from the study, or the study ends when some individuals have not experienced the event.

The well-known functional regression model with scalar response postulates a relation between real random variable and functional random variable X .

A large class of flexible and useful tools for modeling regression operator r is presented by the simple functional index model. This consists in putting a semi-parametric dimension reduction approach on the model by introducing functional parameter θ . The main idea was to find the direction of θ on which the projection of covariate X captures the most information about answer Y . The considered model was a single revealing direction model (or simple functional index model). This approach arouses various interests. Firstly, to avoid the problems due to dimensionality that can be encountered in the purely non-parametric approach [19]. The non-parametric estimation of the regression would no longer be affected by the scourge of dimension since it is a dependent function of θ which is of dimension 1. Finally, the estimation of functional parameter θ provides an easily interpretable tool. The simple index approach is well-known in the standard multivariate context for its interest in its predictive abilities, and for its interpretability attested by various works that appeared over the past two decades [20]. Extensions to the functional framework of such functional semi-parametric methodology have been the subject of extensive study in the literature. The first work linking the single index model and the non-parametric regression model for functional variables is made [21] in the case of independent observations, and they established almost complete convergence. Their results were extended to dependent cases by [22]. Ait-Saidi et al. [23] studied the case where the simple functional index is unknown; they proposed an estimator of this parameter based on the cross-validation technique. These results were extended to the multiple functional index models by [24]. Ferraty and Park [25] proposed a new estimator of this parameter based on the idea of functional derivative estimation; the problem of the single index model to the functional data where the observations are censored does not seem to have been considered much in the literature, which makes this paper one of the more recent research work on the subject.

Moreover, the analysis of functional data being a branch of statistics that has been the subject of several recent studies and developments, this paper makes it possible to adapt the functional conditional models to censored data based on a single functional index structure.

The rest of the paper is arranged as follows, in Section 2, we present our model and some basic assumptions. In Section 3, we state the main results as well as their proofs. As then application, we study the asymptotic normality of the conditional quantile for functional data in the single functional index model in Section 4. After that, in Section 5, we carry out a simulation study in order to illustrate some properties of the resulted estimator.

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 G , where $Y_i = \min\{T_i, C_i\}$ and $\delta_i = 1_{T_i \leq C_i}$.

To ensure the identifiability of the model, we suppose that $(T_i)_i$ are independent of $(C_i)_i$. Let $\{(Y_i, \delta_i, X_i)_i\}$ be a sequence of strictly stationary random vectors where $(X_i)_{i \geq 1}$ is valued in infinite dimensional semi-metric vector space, and Y_k 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 patient's modality. Furthermore this condition permits to get an unbiased Kernel estimator.

2. MODEL AND SOME BASIC ASSUMPTIONS

Consider a random pair (X, T) where T is real-valued random variable \mathbb{R} and X be a functional random variable (frv) who takes its values in a separable real Hilbert space \mathcal{H} with the norm $\|\cdot\|$ generated by an inner product $\langle \cdot, \cdot \rangle$ and consider that, given the $(X_i, T_i)_{i=1, \dots, n}$ is the statistical sample of pairs which are identically distributed like (X, T) . Hence for the, X is called a 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:

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

By saying that, one is implicitly assuming the existence of a regular version of the conditional distribution T given $\langle \theta, X \rangle = \langle \theta, x \rangle$.

In this infinite dimensional purpose, the term functional nonparametric was used, where the word functional refers to the infinite dimensionality of the data and where non-parametric refers to the infinite dimensionality of the model. Such *functional non-parametric* statistics can also be called doubly infinite dimensional (see [26]). The authors also used the term operational statistics since the target object to be estimated (the *cond-cdf* $F(\theta, \cdot, x)$) can be viewed as a non-linear operator.

2.1. THE ESTIMATORS

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

$$F_n(\theta, t, x) = \frac{\sum_{i=1}^n K(h_K^{-1}(\langle x - X_i, \theta \rangle)) H(h_H^{-1}(t - T_i))}{\sum_{i=1}^n K(h_K^{-1}(\langle x - X_i, \theta \rangle))}, \quad (2.1)$$

with the convention $0/0 = 0$, where K is a kernel function, H a cumulative distribution function and $h_K := h_{n,K}$ (resp. $h_H := h_{n,H}$) is a sequence of bandwidths that decrease to zero as n goes to infinity.

Let, for any $x \in \mathcal{H}$, $i = 1, \dots, n$ and $t \in \mathbb{R}$ $K_i(\theta, x) := K(h_K^{-1}|\langle x - X_i, \theta \rangle|)$, $H_i(t) = H(h_H^{-1}(t - T_i))$. We denote by $B_\theta(x, h) = \{\chi \in \mathcal{H} : 0 < |\langle x - \chi, \theta \rangle| < h\}$ be a ball of center x and radius h , and let $d_\theta(x, X_i) = |\langle x - X_i, \theta \rangle|$ denote a random variable such that its

cumulative distribution function is given by $\phi_{\theta,x}(u) = \mathbb{P}(d_{\theta}(x, X_i) \leq u) = \mathbb{P}(X_i \in B_{\theta}(x, u))$.

In practice, in particularly medical applications, one can be deal with censored variables. This problem is usually modeled by considering positive C variable-censorship, and the observed random variables are not couples (T_i, X_i) , but rather (Y_i, δ_i, X_i) where $Y_i = \min\{T_i, C_i\}$ and $\delta_i = 1_{T_i \leq C_i}$. 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 .

The objective of this section is to adapt these ideas under functional random variable X , and build a kernel type estimator of the conditional distribution $F(\theta, \cdot, x)$ adapted for censored samples. Thus one can reformulate the expression (2.1) as follows:

$$\tilde{F}(\theta, t, x) = \frac{\sum_{i=1}^n \frac{\delta_i}{\bar{G}(Y_i)} K(h_K^{-1}(\langle x - X_i, \theta \rangle)) H(h_H^{-1}(t - T_i))}{\sum_{i=1}^n K(h_K^{-1}(\langle x - X_i, \theta \rangle))} \quad (2.2)$$

In practice $\bar{G}(\cdot) = 1 - G(\cdot)$ is unknown, hence it is impossible to use the estimator (2.2). Next, the authors replaced $\bar{G}(\cdot)$ by its Kaplan and Meier [27] estimate $\bar{G}_n(\cdot)$ given by

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

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

$$\hat{F}(\theta, t, x) = \frac{\sum_{i=1}^n \frac{\delta_i}{\bar{G}_n(Y_i)} K(h_K^{-1}(\langle x - X_i, \theta \rangle)) H(h_H^{-1}(t - T_i))}{\sum_{i=1}^n K(h_K^{-1}(\langle x - X_i, \theta \rangle))}. \quad (2.3)$$

2.2. ASSUMPTIONS ON THE FUNCTIONAL VARIABLES

Let N_x be a fixed neighborhood of x in \mathcal{H} and $S_{\mathbb{R}}$ is a fixed compact of \mathbb{R}^+ . Assume that, $(C_i)_{i \geq 1}$ and $(T_i)_{i \geq 1}$ are independent.

For any dfL, let $\tau_L := \sup\{t, \text{such that } L(t) < 1\}$ be its support's right endpoint. Let $S_{\mathbb{R}}$ be a compact set such that $\zeta_{\theta}(\gamma, x) \in S_{\mathbb{R}} \cup (-\infty, \tau]$, where $\tau < \min(\tau_G, \tau_F)$. Assume that $(C_i)_{i \geq 1}$ are independent and let's consider the following hypotheses:

(H1) $\forall h > 0, \mathbb{P}(X \in B_{\theta}(x, h)) = \phi_{\theta,x}(h) > 0$.

2.3. THE NONPARAMETRIC MODEL

As usually in non-parametric estimation, it is supposed that the *cond-cdf* $F(\theta, \cdot, x)$ verifies some smoothness constraints. Let b_1 and b_2 be two positive numbers; such that:

(H2) $\forall (x_1, x_2) \in N_x \times N_x, \forall (t_1, t_2) \in S_{\mathbb{R}} \times S_{\mathbb{R}}$

- (i) $|F(\theta, t_1, x_1) - F(\theta, t_2, x_2)| \leq C_{\theta, x}(\|x_1 - x_2\|^{\alpha_1} + |t_1 - t_2|^{\alpha_2})$
- (ii) $\int t f(\theta, t, x) dt < \infty$ for all $\theta, x \in \mathcal{H}$.

To this end, we need some assumptions concerning the kernel estimator $\hat{F}(\theta, \cdot, x)$.

(H3) The kernel H is a positive bounded function such that $\forall (t_1, t_2) \in \mathbb{R}^2, |H(t_1) - H(t_2)| \leq C|t_1 - t_2|, \int H^2(t) dt < \infty$ and $\int |t|^{b_2} H(t) dt < \infty$.

(H4) The kernel K is a positive bounded function supported on $[0, 1]$ and is differentiable on $[0, 1]$ with derivative such that: $\exists C_1, C_2, -\infty < C_1 < K'(t) < C_2 < 0$, for $0 < t < 1$.

(H5) The df of the censored random variable, G has bounded first derivative G' .

(H6) There exists a function $\beta_{\theta, x}(\cdot)$ such that $\lim_{n \rightarrow \infty} \frac{\phi_{\theta, x}(sh_K)}{\phi_{\theta, x}(h_K)} = \beta_{\theta, x}(s)$, for $\forall s \in [0, 1]$.

(H7) The bandwidth h_K and h_H , small ball probability $\phi_{\theta, x}(h_K)$ satisfying

- (i) $nh_H^2 \phi_{\theta, x}^2(h_K) \rightarrow 0$ and $\lim_{n \rightarrow \infty} \frac{nh_H^3 \phi_{\theta, x}(h_K) \log n}{\log^2 n} \rightarrow \infty, as n \rightarrow \infty$.
- (ii) $nh_H^2 \phi_{\theta, x}^3(h_K) \rightarrow 0, as n \rightarrow \infty$.

3. MAIN RESULTS

In this section the asymptotic normality of the estimator $\hat{F}(\theta, \cdot, x)$ in the single functional index model is established.

Theorem 3.1 Under Assumptions we have (H1)-(H6)-(ii) for all $x \in \mathcal{H}$

$$\frac{n\phi_{\theta, x}(h_K)}{\sigma^2(\theta, t, x)} (\hat{F}(\theta, t, x) - F(\theta, t, x)) \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1),$$

where $\sigma^2(\theta, t, x) = \frac{a_2(\theta, x)}{(a_1(\theta, x))^2} F(\theta, t, x) \left(\frac{1}{\bar{G}(t)} - F(\theta, t, x) \right)$ with $a_l(\theta, x) = K^l(1) - \int_0^1 (K^l)'(u) \beta_{\theta, x}(u) du, l = 1, 2, \xrightarrow{\mathcal{D}}$ means the convergence in distribution.

Proof: In order to establish the asymptotic normality of $\hat{F}(\theta, t, x)$, we need further notations and definitions. First we consider the following decomposition

$$\begin{aligned}
\hat{F}(\theta, t, x) - F(\theta, t, x) &= \frac{\hat{F}_N(\theta, t, x)}{\hat{F}_D(\theta, x)} - \frac{a_1(\theta, x)F(\theta, t, x)}{a_1(\theta, x)} \\
&= \frac{1}{\hat{F}_D(\theta, x)} \left\{ \left(\hat{F}_N(\theta, t, x) - \mathbb{E}\hat{F}_N(\theta, t, x) \right) - \left(a_1(\theta, x)F(\theta, t, x) - \mathbb{E}\hat{F}_N(\theta, t, x) \right) \right\} \\
&\quad + \frac{F(\theta, t, x)}{\hat{F}_D(\theta, x)} \left\{ \left(a_1(\theta, x) - \mathbb{E}\hat{F}_D(\theta, x) \right) - \left(\hat{F}_D(\theta, x) - \mathbb{E}\hat{F}_D(\theta, x) \right) \right\} \\
\hat{F}(\theta, t, x) - F(\theta, t, x) &= \frac{1}{\hat{F}_D(\theta, x)} A_n(\theta, t, x) + B_n(\theta, t, x) \tag{3.1}
\end{aligned}$$

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

$$\begin{aligned}
A_n(\theta, t, x) &= \frac{1}{n\mathbb{E}(K_1(\theta, x))} \sum_{i=1}^n \left\{ \left(\frac{\delta_i}{\bar{G}_n(Y_i)} H_i(t) - F(\theta, t, x) \right) K_i(\theta, x) \right. \\
&\quad \left. - \mathbb{E} \left[\left(\frac{\delta_i}{\bar{G}_n(Y_i)} H_i(t) - F(\theta, t, x) \right) K_i(\theta, x) \right] \right\} \\
&= \frac{1}{n\mathbb{E}(K_1(\theta, x))} \sum_{i=1}^n N_i(\theta, t, x).
\end{aligned}$$

It follows that,

$$\begin{aligned}
n\phi_{\theta, x}(h_K) \text{Var}(A_n(\theta, t, x)) &= \frac{\phi_{\theta, x}(h_K)}{\mathbb{E}^2(K_1(\theta, x))} \text{Var}(N_1(\theta, t, x)) \\
&= V_n(\theta, t, x)
\end{aligned}$$

and $B_n(\theta, t, x) = a_1(\theta, x)F(\theta, t, x) - \mathbb{E}\hat{F}_N(\theta, t, x) + F(\theta, t, x) \left(a_1(\theta, x) - \mathbb{E}\hat{F}_D(\theta, x) \right)$.

Then, the proof of Theorem 3.1 can be deduced from the following Lemmas.

Lemma 3.1. Under assumptions of Theorem 3.1, we have

$$\sqrt{n\phi_{\theta, x}(h_K)} A_n(\theta, t, x) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma^2(\theta, t, x)).$$

Proof:

$$\begin{aligned}
V_n(\theta, t, x) &= \frac{\phi_{\theta, x}(h_K)}{\mathbb{E}^2(K_1(\theta, x))} \mathbb{E} \left[K_1^2(\theta, x) \left(\frac{\delta_1}{\bar{G}(Y_1)} H_1(t) - F(\theta, t, x) \right)^2 \right] \\
&= \frac{\phi_{\theta, x}(h_K)}{\mathbb{E}^2(K_1(\theta, x))} \mathbb{E} \left[K_1^2(\theta, x) \mathbb{E} \left(\left(\frac{\delta_1}{\bar{G}(Y_1)} H_1(t) - F(\theta, t, x) \right)^2 \middle| \langle \theta, X_1 \rangle \right) \right] \tag{3.2}
\end{aligned}$$

Then, the proof of Theorem 3.1 can be deduced from the following Lemmas.

Lemma 3.1. Under assumptions of Theorem 3.1, we have

$$\sqrt{n\phi_{\theta, x}(h_K)} A_n(\theta, t, x) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma^2(\theta, t, x)).$$

Proof:

$$\begin{aligned} V_n(\theta, t, x) &= \frac{\phi_{\theta, x}(h_K)}{\mathbb{E}^2(K_1(\theta, x))} \mathbb{E} \left[K_1^2(\theta, x) \left(\frac{\delta_1}{\bar{G}(Y_1)} H_1(t) - F(\theta, t, x) \right)^2 \right] \\ &= \frac{\phi_{\theta, x}(h_K)}{\mathbb{E}^2(K_1(\theta, x))} \mathbb{E} \left[K_1^2(\theta, x) \mathbb{E} \left(\left(\frac{\delta_1}{\bar{G}(Y_1)} H_1(t) - F(\theta, t, x) \right)^2 \middle| \langle \theta, X_1 \rangle \right) \right] \end{aligned}$$

Using the definition of conditional variance, we have

$$\mathbb{E} \left(\left(\frac{\delta_1}{\bar{G}(Y_1)} H_1(t) - F(\theta, t, x) \right)^2 \middle| \langle \theta, X_1 \rangle \right) = J_{1n} + J_{2n},$$

where $J_{1n} = \text{Var} \left(\frac{\delta_1}{\bar{G}(Y_1)} H_1(t) \middle| \langle \theta, X_1 \rangle \right)$, $J_{2n} = \left[\mathbb{E} \left(\frac{\delta_1}{\bar{G}(Y_1)} H_1(t) \middle| \langle \theta, X_1 \rangle \right) - F(\theta, t, x) \right]^2$.

- Concerning J_{1n} ,

$$J_{1n} = \mathbb{E} \left(\frac{\delta_1}{\bar{G}^2(Y_1)} H^2 \left(\frac{t - Y_1}{h_H} \right) \middle| \langle \theta, X_1 \rangle \right) - \left[\mathbb{E} \left(\frac{\delta_1}{\bar{G}(Y_1)} H_1 \left(\frac{t - Y_1}{h_H} \right) \middle| \langle \theta, X_1 \rangle \right) \right]^2 = \mathcal{J}_1 + \mathcal{J}_2.$$

As for \mathcal{J}_1 , by the property of double conditional expectation, we get that,

$$\begin{aligned} \mathcal{J}_1 &= \mathbb{E} \left\{ \mathbb{E} \left(\frac{\delta_1}{\bar{G}^2(Y_1)} H^2 \left(\frac{t - Y_1}{h_H} \right) \middle| \langle \theta, X_1 \rangle, T_1 \right) \right\} = \mathbb{E} \left\{ \frac{\delta_1}{\bar{G}^2(T_1)} H^2 \left(\frac{t - T_1}{h_H} \right) \mathbb{E}[\mathbf{1}_{T_1 \leq c_1} | T_1] \middle| \langle \theta, X_1 \rangle \right\} \\ &= \mathbb{E} \left(\frac{\delta_1}{\bar{G}(T_1)} H^2 \left(\frac{t - T_1}{h_H} \right) \middle| \langle \theta, X_1 \rangle \right) = \int \frac{1}{\bar{G}(v)} H^2 \left(\frac{t - v}{h_H} \right) dF(\theta, v, X_1) \\ &= \int \frac{1}{\bar{G}(t - u h_H)} H^2(u) dF(\theta, t - u h_H, X_1) \end{aligned} \quad (3.3)$$

By the first order Taylor's expansion of the function $\bar{G}^{-1}(\cdot)$ around zero, one gets

$$\begin{aligned} \mathcal{J}_1 &= \int \frac{1}{\bar{G}(t)} H^2(u) dF(\theta, t - u h_H, X_1) \\ &+ \frac{h_H^2}{\bar{G}^2(t)} \int u H(u) \bar{G}^{(1)}(t^*) f(\theta, t - u h_H, X_1) du + o(1), \end{aligned}$$

where t^* is between t and $t - u h_H$.

Under hypothesis (H7) and using hypothesis (H3)-(ii), we get

$$\mathcal{J}'_1 = \frac{h_H^2}{\bar{G}^2(t)} \int u H^2(u) \bar{G}^{(1)}(t^*) f(\theta, t - u h_H, X_1) du = \mathcal{O}(h_H^2).$$

Indeed

$$\mathcal{J}'_1 \leq h_H^2 \left(\sup_{u \in \mathbb{R}} |G'(u)| / \bar{G}^2(t) \right) \int u f(\theta, t - u h_H, X_1) du.$$

On the other hand, by integrating by part and under assumption (H3)-(i), we have

$$\begin{aligned} \int \frac{H^2(u)}{\bar{G}(t)} dF(\theta, t - uh_H, X_1) &= \frac{1}{\bar{G}(t)} \int 2H(u)H'(u) F(\theta, t - uh_H, X_1) du \\ &\quad - \frac{1}{\bar{G}(t)} \int 2H(u)H'(u) F(\theta, t, x) du \\ &\quad + \frac{1}{\bar{G}(t)} \int 2H(u)H'(u) F(\theta, t, x) du \end{aligned}$$

Clearly we have

$$\int 2H(u)H'(u) dF(\theta, t, x) = [H^2(u)F(\theta, t, x)]_{-\infty}^{+\infty} = F(\theta, t, x), \quad (3.4)$$

thus

$$\int \frac{1}{\bar{G}(t)} H^2(u) dF(\theta, t - uh_H, X_1) = \frac{F(\theta, t, x)}{\bar{G}(t)} + \mathcal{O}(h_K^{\alpha_1} + h_H^{\alpha_2}). \quad (3.5)$$

As for J_{2n} , by (H2), (H4)-(H5), and using Lemma 3.2 in [28] we obtain that $J_{2n} \rightarrow 0$ as $n \rightarrow \infty$.

- Concerning \mathcal{J}_2

$$\begin{aligned} \mathcal{J}'_2 &= \mathbb{E} \left(\frac{\delta_1}{\bar{G}(Y_1)} H_1(t) | \langle \theta, X_1 \rangle \right) = \mathbb{E} \left\{ \mathbb{E} \left(\frac{\delta_1}{\bar{G}(Y_1)} H_1(t) | \langle \theta, X_1 \rangle, T_1 \right) \right\} \\ &= \mathbb{E} \left(\frac{1}{\bar{G}(T_1)} H \left(\frac{t - T_1}{h_H} \right) \mathbb{E}[\mathbf{1}_{T_1 \leq c_1} | T_1] | T_1 \right) = \mathbb{E} \left(H \left(\frac{t - T_1}{h_H} \right) | \langle \theta, X_1 \rangle \right) \\ &= \int H \left(\frac{t - v}{h_H} \right) f(\theta, t, X_1) dv \end{aligned}$$

Moreover, we have by integration by parts and changing variables

$$\mathcal{J}'_2 = F(\theta, t, x) \int H'(u) du + \int H'(u) (F(\theta, t - uh_H, x) - F(\theta, t, x)) du,$$

the last equality is due to the fact that $H0$ is a probability density.

Thus we have:

$$\mathcal{J}'_2 = F(\theta, t, x) + \mathcal{O}(h_K^{\alpha_1} + h_H^{\alpha_2}). \quad (3.6)$$

Finally by hypothesis (H5) we get $\mathcal{J}_2 \rightarrow F^2(\theta, t, x)$. Meanwhile, by (H1), (H4), (H6) and (H8), it follows that:

$$\frac{\phi_{\theta, x}(h_K) \mathbb{E} K_1^2(\theta, x)}{\mathbb{E}^2(K_1(\theta, x))} \xrightarrow{n \rightarrow \infty} \frac{a_2(\theta, x)}{(a_1(\theta, x))^2},$$

which leads to combining equations (3.2)-(3.6)

$$V_n(\theta, t, x) \xrightarrow{n \rightarrow \infty} \frac{a_2(\theta, x)}{(a_1(\theta, x))^2} F(\theta, t, x) \left(\frac{1}{\bar{G}(t)} - F(\theta, t, x) \right). \quad (3.7)$$

Lemma 3.2. If the assumptions (H1)-(H6) are satisfied, we have

$$\sqrt{n\phi_{\theta,x}(h_K)} B_n(\theta, t, x) \rightarrow 0, \text{ in Probability.}$$

Proof: We have

$$\begin{aligned} \sqrt{n\phi_{\theta,x}(h_K)} B_n(\theta, t, x) &= \frac{\sqrt{n\phi_{\theta,x}(h_K)}}{\hat{F}_D(\theta, x)} \{ \mathbb{E} \hat{F}_N(\theta, t, x) - a_1(\theta, x) F(\theta, t, x) \\ &\quad + F(\theta, t, x) (a_1(\theta, x) - \mathbb{E} \hat{F}_D(\theta, x)) \} \end{aligned}$$

Firstly, observed that the results below as $n \rightarrow \infty$

$$\frac{1}{\phi_{\theta,x}(h_K)} \mathbb{E} \left[K^l \left(\frac{\langle \theta, x - X_i \rangle}{h_K} \right) \right] \rightarrow a_l(\theta, x), \text{ for } l = 1, 2, \quad (3.8)$$

$$\mathbb{E} \hat{F}_D(\theta, x) \rightarrow a_1(\theta, x), \quad (3.9)$$

and

$$\mathbb{E} \hat{F}_N(\theta, t, x) \rightarrow a_1(\theta, x) F(\theta, t, x), \quad (3.10)$$

can be proved in the same way as in [29] corresponding to their Lemmas 5.1 and 5.2, and then their proofs are omitted.

Secondly, on the one hand, making use of (3.8), (3.9) and (3.10), we have as $n \rightarrow \infty$

$$\left\{ \mathbb{E} \hat{F}_N(\theta, t, x) - a_1(\theta, x) F(\theta, t, x) + F(\theta, t, x) (a_1(\theta, x) - \mathbb{E} \hat{F}_D(\theta, x)) \right\} \rightarrow 0.$$

On other hand,

$$\frac{\sqrt{n\phi_{\theta,x}(h_K)}}{\hat{F}_D(\theta, x)} = \frac{\sqrt{n\phi_{\theta,x}(h_K)} \tilde{F}'(\theta, t, x)}{\hat{F}_D(\theta, x) \tilde{F}'(\theta, t, x)} = \frac{\sqrt{n\phi_{\theta,x}(h_K)} \tilde{F}'(\theta, t, x)}{\tilde{F}'_N(\theta, t, x)}.$$

Then using Proposition 3.2 in [28], it suffices to show that $\frac{\sqrt{n\phi_{\theta,x}(h_K)}}{\tilde{F}'_N(\theta, t, x)}$ tends to zero as n goes to infinity.

Indeed

$$\tilde{F}'_N(\theta, t, x) = \frac{1}{nh_H \mathbb{E}(K_1(\theta, x))} \sum_{i=1}^n \frac{\delta_i}{\bar{G}(Y_i)} K(h_K^{-1}(\langle x - X_i, \theta \rangle)) H'(h_H^{-1}(t - T_i)).$$

Because K and H' are continuous with support on $[0, 1]$ then by (H5)-(ii) and (H6) $\exists m = \min_{[0,1]} K(t)H'(t)$ it follows that

$$\tilde{F}'_N(\theta, t, x) \geq \frac{m}{h_H \phi_{\theta,x}(h_K)}$$

wich give

$$\frac{n\phi_{\theta,x}(h_K)}{\tilde{F}'_N(\theta, t, x)} \leq \frac{\sqrt{nh_H^2 \phi_{\theta,x}^3(h_K)}}{m}.$$

Finally, using (H7)-(ii), completes the proof of Lemma 3.2.

4. APPLICATION: THE CONDITIONAL QUANTILE IN FUNCTIONAL SINGLE-INDEX MODEL

The main objective of this section is to establish the asymptotic normality of the conditional quantile estimator of T given $\langle \theta, X \rangle = \langle \theta, x \rangle$ denoted by $\zeta_\theta(\gamma, x)$. Saying that, we are implicitly assuming the existence of a regular version for the conditional distribution of T given $\langle \theta, X \rangle = \langle \theta, x \rangle$. Now, let $\zeta_\theta(\gamma, x)$ be the γ th-conditional quantile of the distribution of T given $\langle \theta, X \rangle = \langle \theta, x \rangle$ denoted by $\zeta_\theta(\gamma, x)$. Formally, $\zeta_\theta(\gamma, x)$ is defined 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 $F(\theta, t, x)$ is strictly increasing and continuous in a neighborhood of $\zeta_\theta(\gamma, x)$. This is insuring that the conditional quantile $\zeta_\theta(\gamma, x)$ is uniquely defined by:

$$\zeta_\theta(\gamma, x) = F^{-1}(\theta, \gamma, x) \text{ equivalently } \hat{F}(\theta, \hat{\zeta}_\theta(\gamma, x), x) = \gamma. \quad (4.1)$$

As a by-product of (4.1) and (2.1), it is easy to derive an estimator $\zeta_{\theta,n}(\gamma, x)$ of $\zeta_\theta(\gamma, x)$:

$$\zeta_{\theta,n}(\gamma, x) = F_n^{-1}(\theta, \gamma, x). \quad (4.2)$$

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

$$\hat{\zeta}_\theta(\gamma, x) = \hat{F}^{-1}(\theta, \gamma, x) = \inf\{t \in \mathbb{R} : \hat{F}(\theta, t, x) \geq \gamma\}; \quad (4.3)$$

which satisfies

$$\hat{F}(\theta, \hat{\zeta}_\theta(\gamma, x), x) = \gamma. \quad (4.4)$$

Theorem 4.1. If the assumptions (H1)-(H7) are satisfied and if γ is the unique order of the quantile such that $\gamma = F(\theta, \zeta_\theta(\gamma, x), x) = \hat{F}(\theta, \hat{\zeta}_\theta(\gamma, x), x)$, $x \in \mathcal{H}$

$$\left(\frac{n\phi_{\theta,x}(h_K)}{\Sigma^2(\theta, \zeta_\theta(\gamma, x), x)} \right)^{1/2} \left(\hat{\zeta}_\theta(\gamma, x) - \zeta_\theta(\gamma, x) \right) \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1),$$

$$\text{where } \Sigma^2(\theta, \zeta_\theta(\gamma, x), x) = \frac{\sigma^2(\theta, \zeta_\theta(\gamma, x), x)}{f^2(\theta, \zeta_\theta(\gamma, x), x)} = \frac{a_2(\theta, x)}{(a_1(\theta, x))^2} \frac{\gamma[\bar{G}^{-1}(\zeta_\theta(\gamma, x)) - \gamma]}{f^2(\theta, \zeta_\theta(\gamma, x), x)}.$$

Proof: For Theorem 4.1, making use of (3.1), we have

$$\begin{aligned} \sqrt{n\phi_{\theta,x}(h_K)} \left(\hat{\zeta}_{\theta}(\gamma, x) - \zeta_{\theta}(\gamma, x) \right) &= \sqrt{n\phi_{\theta,x}(h_K)} \frac{F_n(\theta, \zeta_{\theta}(\gamma, x), x)}{F'_n(\theta, \hat{\zeta}_{\theta}^*(\gamma, x), x)} \\ &\quad - \sqrt{n\phi_{\theta,x}(h_K)} \frac{F(\theta, \zeta_{\theta}(\gamma, x), x)}{F'_n(\theta, \hat{\zeta}_{\theta}^*(\gamma, x), x)} \\ &= \frac{\sqrt{n\phi_{\theta,x}(h_K)} A_n(\theta, t, x)}{F'_n(\theta, \hat{\zeta}_{\theta}^*(\gamma, x), x)} - \frac{\sqrt{n\phi_{\theta,x}(h_K)} B_n(\theta, t, x)}{F'_n(\theta, \hat{\zeta}_{\theta}^*(\gamma, x), x)} \end{aligned} \quad (4.7)$$

Then using Theorem 3.1 and Lemma 3.2 we obtain the result.

4.1. APPLICATION AND CONFIDENCE BANDS

The asymptotic variances $\sigma^2(\theta, t, x)$ and $\Sigma^2(\theta, \zeta_{\theta}(\gamma, x), x)$ in theorems Theorem 3.1 and Theorem 4.1 depend on some unknown quantities including a_1 , a_2 , $\phi(u)$, $\zeta_{\theta}(\gamma, x)$, $\bar{G}(\cdot)$ and $f(\theta, \zeta_{\theta}(\gamma, x), x)$. Therefore, $\bar{G}(\cdot)$, $\zeta_{\theta}(\gamma, x)$, and $f(\theta, \zeta_{\theta}(\gamma, x), x)$ should be replaced, respectively, by the Kaplan-Meier's estimator $\bar{G}_n(\cdot)$, the kernel-type estimator of the joint distribution $\hat{f}(\theta, \zeta_{\theta}(\gamma, x), x)$ and $\hat{\zeta}_{\theta}(\gamma, x)$ the conditional quantile estimator given by equation (4.3). Moreover, using the decomposition given by assumption (H1), one can estimate $\phi_{\theta,x}(z)$ by $F_{x,n}(z) = 1/n \sum_{i=1}^n \mathbf{1}_{\{X_i \in B_{\theta}(x,z)\}}$. Because the unknown functions $a_j := a_j(\theta, x)$ and $f(\theta, t, x)$ intervening in the expression of the variance. So we need to estimate the quantities $a_1(\theta, x)$, $a_2(\theta, x)$ and $F(\theta, t, x)$, respectively.

By the assumptions (H1)-(H4) we know that $a_j(\theta, x)$ can be estimated by $\hat{a}_j(\theta, x)$ which is defined as:

$$\hat{a}_j(\theta, x) = \frac{1}{n\hat{\phi}_{\theta,x}(h)} \sum_{i=1}^n K_i^j(\theta, x), \quad \text{where } \hat{\phi}_{\theta,x}(h) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{\{|\langle x - X_i, \theta \rangle| < h\}},$$

with $\mathbf{1}_{\{\cdot\}}$ being the indicator function.

By applying the kernel estimator of $f(\theta, t, x)$ given above, the quantity $\sigma^2(\theta, t, x)$ can be estimated by:

$$\hat{\sigma}^2(\theta, t, x) = \frac{\hat{a}_2(\theta, x)}{(\hat{a}_1(\theta, x))^2} \hat{f}(\theta, t, x) \int H^2(u) du$$

so we can derive the following corollary:

Corollary 4.1. Under the assumptions of Theorem 3.1, K' and $(K^2)'$ are integrable functions, then we get as $n \rightarrow \infty$.

1.

$$\frac{\hat{a}_1}{\sqrt{\hat{a}_2}} \sqrt{\frac{n\hat{\phi}_{\theta,x}(h_K)}{\hat{F}(\theta, t, x) [\bar{G}_n^{-1}(t) - \hat{F}(\theta, t, x)]}} \left(\hat{F}(\theta, t, x) - F(\theta, t, x) \right) \xrightarrow{\mathcal{D}} \mathcal{N}(0,1).$$

2.

$$\frac{\hat{a}_1 \hat{f}(\theta, \hat{\zeta}_\theta(\gamma, x), x)}{\sqrt{\hat{a}_2}} \sqrt{\frac{n\hat{\phi}_{\theta,x}(h_K)}{\gamma [\bar{G}_n^{-1}(\hat{\zeta}_\theta(\gamma, x)) - \gamma]}} \left(\hat{\zeta}_\theta(\gamma, x) - \zeta_\theta(\gamma, x) \right) \xrightarrow{\mathcal{D}} \mathcal{N}(0,1).$$

Proof: Observe that

1.

$$\begin{aligned} & \frac{\hat{a}_1}{\sqrt{\hat{a}_2}} \sqrt{\frac{n\hat{\phi}_{\theta,x}(h_K)}{\hat{F}(\theta, t, x) [\bar{G}_n^{-1}(t) - \hat{F}(\theta, t, x)]}} \left(\hat{F}(\theta, t, x) - F(\theta, t, x) \right) \\ &= \frac{\hat{a}_1 \sqrt{\hat{a}_2}}{a_1 \sqrt{\hat{a}_2}} \sqrt{\frac{n\hat{\phi}_{\theta,x}(h_K) [\bar{G}^{-1}(t) - F(\theta, t, x)] F(\theta, t, x)}{\hat{F}(\theta, t, x) [\bar{G}_n^{-1}(t) - \hat{F}(\theta, t, x)] n\phi_{\theta,x}(h_K)}} \\ & \times \frac{a_1}{\sqrt{\hat{a}_2}} \sqrt{\frac{n\phi_{\theta,x}(h_K)}{F(\theta, t, x) [\bar{G}^{-1}(t) - F(\theta, t, x)]}} \left(\hat{F}(\theta, t, x) - F(\theta, t, x) \right) \end{aligned}$$

Via Theorem 3.1, we have

$$\frac{a_1}{\sqrt{\hat{a}_2}} \sqrt{\frac{n\phi_{\theta,x}(h_K)}{F(\theta, t, x) [\bar{G}^{-1}(t) - F(\theta, t, x)]}} \left(\hat{F}(\theta, t, x) - F(\theta, t, x) \right) \xrightarrow{\mathcal{D}} \mathcal{N}(0,1).$$

Next, by [30], we can prove that

$$\hat{a}_1 \xrightarrow{\mathbb{P}} a_1, \hat{a}_2 \xrightarrow{\mathbb{P}} a_2 \text{ and } \frac{\hat{\phi}_{\theta,x}(h_K)}{\sqrt{\phi_{\theta,x}(h_K)}} \xrightarrow{\mathbb{P}} 1 \text{ as } n \rightarrow \infty.$$

If in addition, we consider Lemma 3.2 and (4.7), the consistency of $\bar{G}_n^{-1}(\cdot)$ to $\bar{G}^{-1}(\cdot)$ according to [31], we obtain

$$\frac{\hat{a}_1 \sqrt{\hat{a}_2}}{a_1 \sqrt{\hat{a}_2}} \sqrt{\frac{n\hat{\phi}_{\theta,x}(h_K) [\bar{G}^{-1}(t) - F(\theta, t, x)] F(\theta, t, x)}{\hat{F}(\theta, t, x) [\bar{G}_n^{-1}(t) - \hat{F}(\theta, t, x)] n\phi_{\theta,x}(h_K)}} \rightarrow 1 \text{ a. s.}$$

2.

$$\begin{aligned}
& \frac{\hat{a}_1 \hat{f}(\theta, \hat{\zeta}_\theta(\gamma, x), x)}{\sqrt{\hat{a}_2}} \sqrt{\frac{n \hat{\phi}_{\theta, x}(h_K)}{\gamma [\bar{G}_n^{-1}(\hat{\zeta}_\theta(\gamma, x)) - \gamma]}} (\hat{\zeta}_\theta(\gamma, x) - \zeta_\theta(\gamma, x)) \\
&= \frac{\hat{a}_1 \sqrt{a_2}}{a_1 \sqrt{\hat{a}_2}} \sqrt{\frac{n \hat{\phi}_{\theta, x}(h_K) [\bar{G}^{-1}(\zeta_\theta(\gamma, x)) - \gamma]}{[\bar{G}_n^{-1}(\hat{\zeta}_\theta(\gamma, x)) - \gamma] n \phi_{\theta, x}(h_K)}} \frac{\hat{f}(\theta, \hat{\zeta}_\theta(\gamma, x), x)}{f(\theta, \zeta_\theta(\gamma, x), x)} \\
&\times \frac{a_1}{\sqrt{a_2}} \sqrt{\frac{n \phi_{\theta, x}(h_K)}{\gamma [\bar{G}^{-1}(\zeta_\theta(\gamma, x)) - \gamma]}} f(\theta, \zeta_\theta(\gamma, x), x) (\hat{\zeta}_\theta(\gamma, x) - \zeta_\theta(\gamma, x))
\end{aligned}$$

Making use of Theorem 4.1, we obtain

$$\frac{a_1}{\sqrt{a_2}} \sqrt{\frac{n \phi_{\theta, x}(h_K)}{\gamma [\bar{G}^{-1}(\zeta_\theta(\gamma, x)) - \gamma]}} f(\theta, \zeta_\theta(\gamma, x), x) (\hat{\zeta}_\theta(\gamma, x) - \zeta_\theta(\gamma, x)) \xrightarrow{D} \mathcal{N}(0, 1).$$

Further, by considering Lemma 3.2, (4.7), and the consistency of $\bar{G}_n^{-1}(\cdot)$ to $\bar{G}^{-1}(\cdot)$ (see [31]), we obtain as $n \rightarrow \infty$.

$$\frac{\hat{a}_1 \sqrt{a_2}}{a_1 \sqrt{\hat{a}_2}} \sqrt{\frac{n \hat{\phi}_{\theta, x}(h_K) [\bar{G}^{-1}(\zeta_\theta(\gamma, x)) - \gamma]}{[\bar{G}_n^{-1}(\hat{\zeta}_\theta(\gamma, x)) - \gamma] n \phi_{\theta, x}(h_K)}} \frac{\hat{f}(\theta, \hat{\zeta}_\theta(\gamma, x), x)}{f(\theta, \zeta_\theta(\gamma, x), x)} \xrightarrow{\mathbb{P}} 1.$$

Hence, the proof is completed.

Remark 4.1. Thus, following the Corollary 4.1, the asymptotic $(1 - \xi)$ confidence interval of $F(\theta, t, x)$ and $\zeta_\theta(\gamma, x)$ respectively, which are expressed as follows:

$$\hat{F}(\theta, t, x) \pm \eta_{\gamma/2} \frac{\hat{a}_1}{\sqrt{\hat{a}_2}} \sqrt{\frac{\hat{F}(\theta, t, x) [\bar{G}_n^{-1}(t) - \hat{F}(\theta, t, x)]}{n \hat{\phi}_{\theta, x}(h_K)}},$$

and

$$\hat{\zeta}_\theta(\gamma, x) \pm \eta_{\gamma/2} \frac{\hat{a}_1 \hat{f}(\theta, \hat{\zeta}_\theta(\gamma, x), x)}{\sqrt{\hat{a}_2}} \sqrt{\frac{\gamma [\bar{G}_n^{-1}(\hat{\zeta}_\theta(\gamma, x)) - \gamma]}{n \hat{\phi}_{\theta, x}(h_K)}},$$

where $\eta_{\gamma/2}$ is the upper $\gamma/2$ quantile of the normal distribution $\mathcal{N}(0, 1)$.

5. SIMULATION STUDY

In this section we consider simulated data studies to assess the finite sample performance of the proposed estimator and compare it to its competitor. To study the behavior

of our conditional quantiles estimator, we consider in this part a comparison of our CFSIM (2.4) model (functional single index model with censored data) with that of CNPFDA (5.1) (censored nonparametric functional data analysis), for more details, we can refer be made to the works ([32] or [33]) 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.

$$\hat{F}_n(t, x) = \frac{\sum_{k=1}^n \frac{\delta_k}{\bar{G}_n(Y_k)} K(h_K^{-1} d(x, X_k)) H(h_H^{-1}(t - T_k))}{\sum_{k=1}^n K(h_K^{-1} d(x, X_k))}. \quad (5.1)$$

Furthermore, some tuning parameters have to be specified. The kernel $K(\cdot)$ is chosen to be the quadratic function defined as $K = \frac{3}{2}(1 - u^2)\mathbf{1}_{[0;1]}$ and the cumulative $dfH(u) = \int_{-\infty}^u \frac{3}{4}(1 - z^2)\mathbf{1}_{[-1;1]}(z)dz$.

The semi-metric $d(\cdot, \cdot)$ will be specified according to the choice of the functional space \mathcal{H} discussed in the scenarios below. It is well-known that one of the crucial H parameters in semi-parametric models is the smoothing parameters which are involved in defining the shape of the link function between the response and the covariate.

Now for simplifying the implementation of our methodology, we take the bandwidths $h_K \sim h_H = h$, where h will be chosen by the cross-validation method on the k -nearest neighbors (see [6], p. 102).

5.1. SIMULATION 1: CASE OF SMOOTH CURVES

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

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

where ϵ_i a sequence of i.i.d. random variables normally distributed with a variance equal to 0.1.

The functional covariate X is assumed to be a diffusion process defined on $[0, 1]$ and generated by the following equation:

$$X(t) = a \cos(b + \pi Wt) + c \sin(d + \pi Wt) + (1 - A) \sin(\pi Wt), \quad d \in [0, 1],$$

where Wb and d are independent of normal distributions respectively $\rightsquigarrow \mathcal{N}(0, 1)$, $\rightsquigarrow \mathcal{N}(0, 0.03)$ and $\rightsquigarrow \mathcal{N}(0, 0.05)$. The variables a and c are Bernoulli's laws Bernoulli $\mathcal{B}(0.5)$. Fig. 1 depicts a sample of 200 curves representing a realization of the functional random variable X .

Take into account of the smoothness of the curves $X_i(t)$ (see Fig. 1), we choose the distance $deriv_1$ (the semi-metric based on the first derivatives of the curves) 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. Then, we consider a nonlinear regression function defined as

$$R(X) = 4 \log \left\{ \frac{1}{\left(\int_0^1 (X'(t))^2 dt + \left[\int_0^1 X'(t) dt \right]^2 \right)} \right\}.$$

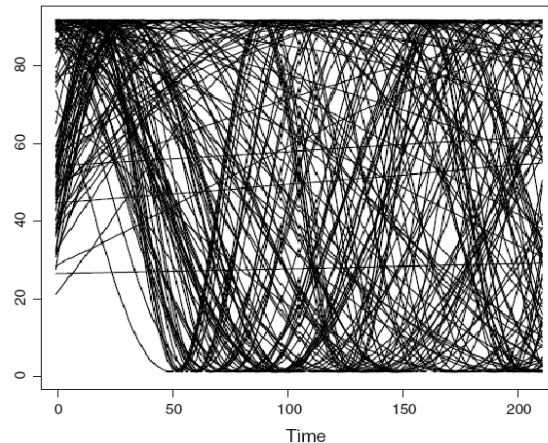


Figure 1. A sample of 200 curves $X_{i=1,\dots,200}(t_j)$, $t_j=1,\dots,200 \in [0, 1]$.

Given $X = x$, $Y \rightsquigarrow \mathcal{N}(R(x), 0.2)$, 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 $(X_i, Y_i)_{i=1,\dots,n}$ and the single index θ which is unknown and has to be estimated.

In practice this parameter can be selected by cross-validation approach (see [23]). In this passage it may be that one can select the real-valued function $\theta(t)$ among the eigenfunctions of the covariance operator $\mathbb{E}[(X' - \mathbb{E}X')\langle X', \cdot \rangle_{\mathcal{H}}]$ where $X(t)$ is a diffusion processes defined on a real interval $[a, b]$ and $X'(t)$ its first derivative (see [34]). 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: $\frac{1}{\mathcal{L}} \sum_{i \in \mathcal{L}} (X'_i - \mathbb{E}X'_i)^t (X'_i - \mathbb{E}X'_i)$, 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 of the empirical covariance operator, which will replace θ during the simulation step.

In the following graphs, the covariance operator for $\mathcal{L} = \{1, \dots, 200\}$ gives the discretization of the eigenfunctions $\theta_i(t)$ (presented as a continuous curve) (Fig. 2).

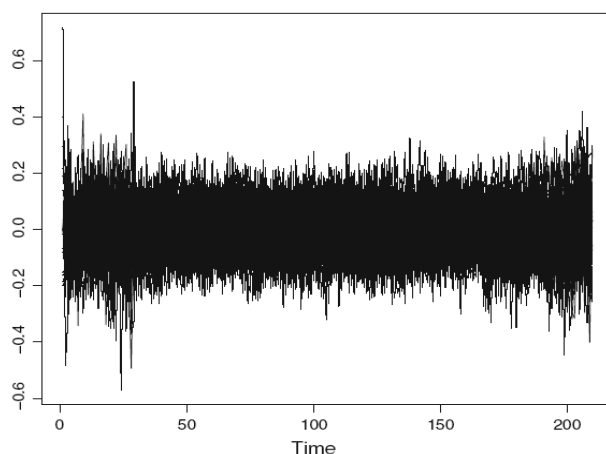


Figure 2. The curves $\theta_{i=1,\dots,200}(t_j)$, $t_{j=1,\dots,200} \in [0, 1]$.

In this simulation part, we divide our sample of size 200 into two parts. The first one from 1 to 125 will be used to make the simulation and the second from 126 to 200 will serve us for the prediction.

We follow the following steps:

Step 1: Compute the inner product: $\langle \theta^*, X_1 \rangle, \dots, \langle \theta^*, X_{200} \rangle$, generate independently the variables $\epsilon_1, \dots, \epsilon_{200}$, then simulate the response variables $Y_k = r(\langle \theta^*, X_k \rangle) + \epsilon_k$, where $r(\langle \theta^*, X_k \rangle) = \exp(10(\langle \theta^*, X_k \rangle - 0.05))$ and generate independently the variables $\epsilon_1, \dots, \epsilon_{200}$.

Step 2: For each k in the test sample $\mathcal{J} = 126, \dots, 200$, we compute: $\hat{Y}_{nk} = \hat{\zeta}(\gamma, X_k)$, where

$$\zeta(\gamma, \chi) := \inf\{y \in \mathbb{R} : F^\chi(y) \geq \gamma\},$$

and

$$\hat{F}^\chi(y) = \frac{\sum_{k=1}^n K(h^{-1}d(\chi, X_k)) H(h^{-1}(y - Y_k))}{\sum_{k=1}^n K(h^{-1}d(\chi, X_k))}, \quad \forall y \in \mathbb{R}. \quad (5.2)$$

Step 3: Finally, we present the results by plotting the predicted values versus the true values and compute the mean squared error (MSE):

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

then, using the learning sample to compute the estimator of

$$\hat{Y}_{nk} = \hat{\zeta}(\gamma, x) \text{ for } j = \{126, \dots, 200\}.$$

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) (2.4) and (5.2) which are defined as:

$$FSIM.MSE = \frac{1}{75} \sum_{k=126}^{200} (Y_k - \hat{Y}_k)^2, \quad NPFDA.MSE = \frac{1}{75} \sum_{k=126}^{200} (Y_k - \hat{Y}_{nk})^2,$$

respectively.

We see that the sum of mean square error (MSE) of our method (Censored-Single-Index-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 methods (CFSIM) against the conditional quantile by nonparametric functional data analysis (NPFDA) (Figs. 3- 4). Our estimator is so acceptable.

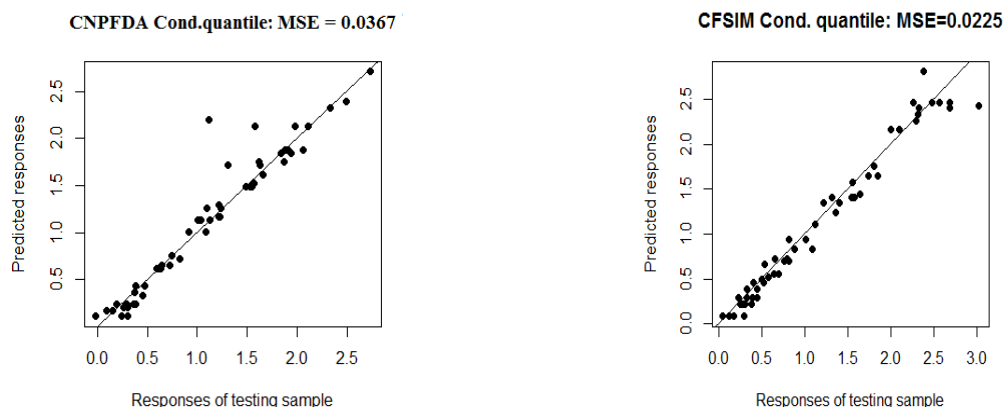


Figure 3. Comparison between NPFDA and CFSIM with CR~3%.

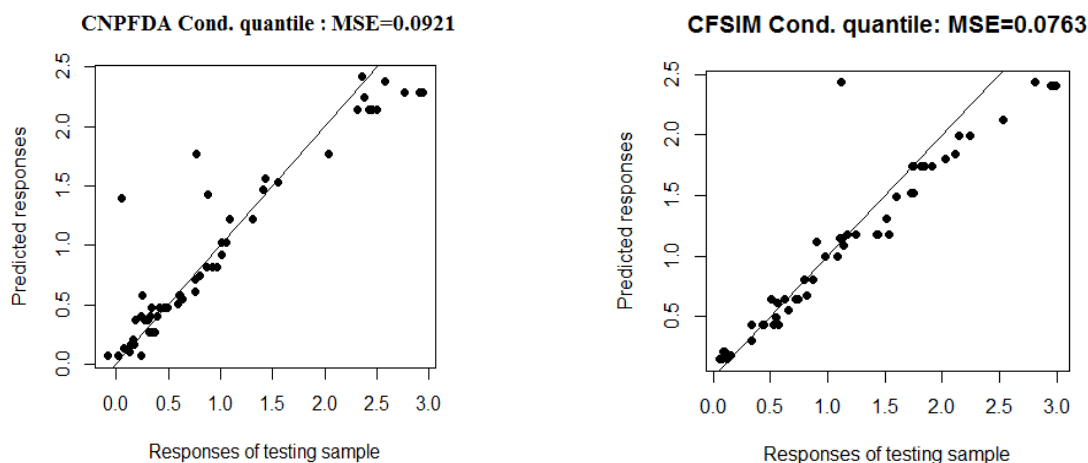


Figure 4. Comparison between NPFDA and CFSIM with CR~18%.

4. CONCLUSIONS

This paper focused on the non-parametric estimation of a conditional quantile for independent data under random censorship. Both the almost complete convergence (with rates), and the resulting estimator were shown to be asymptotically normally distributed under some regularity conditions. Naturally, the plug-in rules were used to obtain an estimator of the asymptotic variance term. The authors point out that here it is possible to prove that the variance estimator is almost completely consistent, using analogous ideas.

The proofs are based on a combination of the existing techniques. The author's prime aim was to improve the performance of this model for the conditional quantile with the censored response variable. The simulations experiments in this paper show that this methodology can be easily implemented and works very well for both simulated and real data. It is well known that the kernel choice does not affect substantially the quality of the

estimator. By contrast, the bandwidth choice is very crucial in non-parametric estimation. In addition, in order to explore the effectiveness of this method in real situations, the authors applied the CNPFDA estimator to data constituting hourly electricity demand for the Rocky Mountain region of the United States, as well as spectrometric data.

This paper examines conditional distribution based on the single-index model in the censorship model when the sample is considered as an independent and identically distributed (i.i.d.) random variables. The asymptotic properties such as point-wise almost complete consistency, and the uniform almost complete convergence of the kernel estimator with rates, are presented under some mild conditions. In this case, the asymptotic properties of the estimation of the conditional hazard function and the asymptotic normality of the conditional quantile in the single functional index model are being investigated in other works by these authors.

REFERENCES

- [1] Gefeller, O., Michels, P. A review on smoothing methods for the estimation of the hazard rate based on kernel functions, In Dodge, Y., Whittaker, J. (Eds), *Computational Statistics*, Springer -Verlag, New York, 1992, pp. 459-464.
- [2] Hassani, S., Sarda, P., Vieu, P., *Annual Review of Statistics and Its Application*, **35**, 27, 1986.
- [3] Izenman, A., *Journal of the American Statistical Association*, **86**, 205, 1991.
- [4] Pascu, M., Vaduva, I., *Revue Roumaine de Mathématique Pures et Appliquées*, **48**, 173, 2003.
- [5] Singpurwallam, N., Wong, M.Y., *Communications in Statistics - Theory and Methods*, **12**, 559, 1983.
- [6] Ferraty, F., Vieu, P., *Nonparametric Functional Data Analysis: Theory and Practice*, Springer, New York, 2006, pp. 99-112.
- [7] Ramsay, J., Silverman, B., *Functional Data Analysis*, 2nd Ed., Springer, New York, 2005.
- [8] Lecoutre, J.P., Ould-Saïd, E., *Journal of Nonparametric Statistics*, **5**, 83, 1995.
- [9] Padgett, W.J., Nonparametric Estimation of Density and Hazard Rate Functions when Samples are Censored. In Krishnaiah, P.R., Rao, C.R. (Eds.), *Handbook of Statistics*, Vol. **7**, Elsevier, Amsterdam, 1988, pp. 313-331.
- [10] Tanner, M., Wong, W.H., *Annals of Statistics*, **11**, 989, 1983.
- [11] Van-Keilegom, I., Veraverbeke, N., *Annals of the Institute of Statistical Mathematics*, **53**, 730, 2001.
- [12] Khardani, S., Lemdani, M., Ould-Saïd, E., *Journal of the Korean Statistical Society*, **39**, 455, 2010.
- [13] Khardani, S., Lemdani, M., Ould-Saïd, E., *Journal of Statistical Planning and Inference*, **42**, 475, 2011.
- [14] Khardani, S., Lemdani, M., Ould-Saïd, E., *Metrika*, **75**, 229, 2012.
- [15] Lemdani, M., Ould-Saïd, E., Poulin, N., *Journal of Multivariate Analysis*, **100**, 546, 2009.
- [16] Liang, H., De Uña-Alvarez, J., *Metrika*, **72**, 1, 2010.
- [17] Ould Saïd, E., Djabrane, Y., *Communications in Statistics - Theory and Methods*, **40**, 2605, 2011.
- [18] Ould Saïd, E., Tatachak, A.A., *International Journal of Statistics and Economics*, **6**, 76, 2011.

- [19] Ferraty, F., Vieu, P., *Computational Statistics*, **17**(4), 545, 2002.
- [20] Xia, Y., Härdle, W., *Journal of Multivariate Analysis*, **97**, 1162, 2006.
- [21] Ferraty, F., Peuch, A., Vieu, P., *Comptes rendus de l'Académie des Sciences*, **336**, 1025, 2003.
- [22] Aït-Saïdi, A., Ferraty, F., Kassa, R., *Revue Roumaine de Mathématique Pures et Appliquées*, **50**, 321, 2005.
- [23] Aït-Saïdi, A., Ferraty, F., Kassa, R., Vieu, P., *Statistics*, **24**, 475, 2008.
- [24] Bouraine M., Aït-Saïdi, A., Ferraty, F., Vieu, P., *Revue Roumaine de Mathématique Pures et Appliquées*, **55**(5), 355, 2010.
- [25] Ferraty, F., Park, J., Vieu, P., Estimation of a Functional Single Index Model. In Ferraty, F. (Ed.), *Recent Advances in Functional Data Analysis and Related Topics*, Springer, New York, 2011, pp. 111-116.
- [26] Ferraty, F., Vieu, P., Functional Nonparametric Statistics: A Double Infinite Dimensional Framework. In Akritas, M., Politis, D. (eds.), *Recent Advances and Trends in Nonparametric Statistics*, Elsevier, Amsterdam, 2003, pp. 61-76.
- [27] Kaplan, E., Meier, P., *Journal of the American Statistical Association*, **53**, 457, 1958.
- [28] Kadiri, N., Rabhi, A., Bouchentouf, A.A., *Journal Dependence Modeling*, **6**(1), 197, 2018.
- [29] Ezzahrioui, M., Ould Saïd, E., *Communications in Statistics - Theory and Methods*, **37**(16-17), 2735, 2008.
- [30] Laïb, N., Louani, D., *Journal of Multivariate Analysis*, **101**, 2266, 2010.
- [31] Deheuvels, P., Einmahl, J.H.J., *The Annals of Probability*, **28**, 1301, 2000.
- [32] Chaouch, M., Bouchentouf, A.A., Traore, A., Rabhi, A., *Journal of Nonparametric Statistics*, **32**(2), 725, 2020.
- [33] Kadiri, N., Rabhi, A., Khardani, S., Akkal, F., *Acta Universitatis Sapientiae, Mathematica*, **13**(1), 45, 2021.
- [34] Attaoui, S., Ling, N., *Metrika: International Journal for Theoretical and Applied Statistics*, **79**(5), 485, 2016.