ECONOMETRICS
EKONOMETRIA
Advances in Applied Data Analysis

Econometrics. Ekonometria.
Advances in Applied Data Analysis
Year 2024, Vol. 28, No. 1
ISSN 2449-9994

Asymptotic Normality of Single Functional Index Quantile Regression for Functional Data with Missing Data at Random

Anis Allal

University Djillali LIABES of Sidi Bel Abbes, Algeria

e-mail: anis.allal@univ-sba.dz

ORCID: 0000-0002-1553-8132

Nadia Kadiri

University Djillali LIABES of Sidi Bel Abbes, Algeria

e-mail: nad.kad06@yahoo.com

ORCID: 0000-0002-3405-7414

Abbes Rabhi

University Djillali LIABES of Sidi Bel Abbes, Algeria

e-mail: rabhi_abbes@yahoo.fr

ORCID: 0000-0001-6740-0226

Quote as: Allal, A., Kadiri, N., and Rabhi, A. (2024). Asymptotic Normality of Single Functional Index Quantile Regression for Functional Data with Missing Data at Random. *Econometrics. Ekonometria. Advances in Applied Data Analysis*, *28*(1), 26-38.

DOI: 10.15611/eada.2024.1.03

JEL Classification: C02, C13, C14, C21

Abstract: This work addresses the problem of the nonparametric estimation of the regression function, namely the conditional distribution and the conditional quantile in the single functional index model (SFIM) under the independent and identically distributed condition with randomly missing data. The main result of this study was the establishment of the asymptotic properties of the estimator, such as the almost complete convergence rates. Moreover, the asymptotic normality of the constructs was obtained under certain mild conditions. Lastly, the authors discussed how to apply the result to construct confidence intervals.

Keywords: asymptotic normality, functional data analysis, functional single-index process, missing at random, nonparametric estimation, small ball probability

1. Introduction

Nonparametric methods for functional data analysis are statistical techniques used to analyse data observed as curves or functions, rather than traditional scalar values. These methods do not rely on explicit assumptions about the functional form or distribution of the data and are wellsuited for handling complex and high-dimensional functional data. Below are some commonly used nonparametric methods for functional data analysis:

1. Functional Data Visualisation: visualisation techniques help in understanding the patterns and structures within functional data. Plotting functional data as curves or surfaces provides a visual representation of the variability and trends present in the data.
2. Functional Principal Component Analysis (FPCA): FPCA is an extension of traditional PCA to functional data. It aims to identify the dominant modes of variation in the data by decomposing the functional data into orthogonal components, known as functional principal components (FPCs). FPCA provides dimension reduction and can be used for tasks such as denoising, clustering, and visualisation of functional data.
3. Kernel Smoothing: kernel smoothing methods, such as kernel density estimation and kernel regression, can be adapted to functional data. These methods involve smoothing the functional data using kernel functions to estimate the underlying density or regression function. Kernel smoothing is useful for estimating functional summaries and making predictions for new functional observations.
4. Nonparametric Regression: nonparametric regression techniques, like local linear regression or spline-based regression, can be applied to functional data. These methods estimate the relationship between the response variable and the functional predictors without assuming a specific parametric form. Nonparametric regression is useful for modelling complex relations and making predictions for functional outcomes.

In the case of missing data at random (MAR), and when the predictor takes values in a semi-metric space, one can use nonparametric methods to estimate the quantile of a conditional distribution. One approach is to use the concept of inverse probability weighting. Missing data often appear in different areas, including surveys, clinical trials and longitudinal studies. Responses may be missing, and methods for processing missing data often depend on the mechanism that generates the missing values (see Efromovich, 2011).

The difference between missing and non-missing values lies in whether or not a value is present or recorded for a particular variable in a dataset. Non-missing values refer to observations or data points where a value is available for a given variable. These values are typically recorded, observed, or measured directly. Non-missing values contain information that can be used for analysis and inference, however missing values refer to observations or data points where a value is absent or unknown for a variable. Missing values can occur due to various reasons, such as non-response in surveys, data entry errors, equipment malfunction, or simply the absence of data for certain variables in a particular case or record.

Missing values can introduce challenges when analysing data, as they can lead to biased or incomplete results if not properly handled. Depending on the mechanism generating the missing values, such as Missing Completely at Random (MCAR), Missing at Random (MAR), or Missing Not at Random (MNAR), different methods and techniques can be employed to handle missing data. Itis important to appropriately identify and address missing values to avoid any potential bias or loss of information in data analysis and modeling. Techniques such as data imputation, deletion of cases or variables with missing values, or specialised statistical models can be used to handle missing data effectively.

Nonparametric estimation by kernel methods is a popular approach for estimating the conditional models when dealing with missing data that are missing at random (MAR). MAR refers to a situation where the probability of missingness depends on the observed data but not on the unobserved data itself. In this context, kernel methods can be used to estimate the conditional distribution of the observed data given the available information. Fairly recently, (Ferraty, Sued, and Vieu, 2013) were first to propose estimating the mean of a scalar response based on an i.i.d. functional sample in which explanatory variables are observed for every subject, while part of the responses are missing at random (MAR) for some of them. It generalised the results in (Cheng 1994) to cases where the explanatory variables are of functional nature. Other authors were interested in the estimation of conditional models when the observations are censored or truncated (see for instance: Hamri et al., 2022; Liang and de Uña-Alvarez, 2010; Ould-Saïd and Djabrane, 2011; Ould-Saïd and Tatachak, 2011; Rabhi et al., 2021, etc.).

Kernel density estimation is a nonparametric method that estimates the underlying probability density function (pdf) based on the observed data. It involves smoothing the data using a kernel function, which is a symmetric probability density function centred at each data point. The individual kernel functions are then summed up or averaged to obtain the overall density estimate. When dealing with missing data that are MAR, the missing values can be considered as additional variables in the estimation process. The kernel density estimation can be modified to handle the missing data by incorporating appropriate weighting schemes.

One common approach is the multiple imputation kernel distribution estimation, which involves imputing multiple plausible values for the missing data points using techniques such as regression imputation or predictive mean matching. Next, the kernel distribution estimation is performed on each imputed dataset separately, and the resulting distribution estimates are combined to obtain a final estimate that incorporates the uncertainty due to the missing data. Another approach is the weighted kernel distribution estimation. In this method, the observed data points are given weights that depend on the likelihood of missingness. The weights are used to adjust the contribution of each data point to the kernel distribution estimate, taking into account the missingness mechanism. Note that the specific implementation and choice of kernel function, bandwidth, and imputation method may depend on the characteristics of the data and the research question at hand. Additionally, it is important to consider the assumptions made about the missing data mechanism and assess their validity in the given context. Generally, nonparametric estimation by kernel methods can be adapted to handle missing data that are MAR and provide estimates of the conditional models based on the available information.

This study extends, to the functional single index model case, the results of (Ling, Liang, and Vieu, 2016; Ling, Liu, and Vieu, 2016; Mekki et al., 2022). The rest of the paper is arranged as follows. Section 2 presents the non-parametric estimator of the functional conditional model when the data are MAR. In Section 3, the authors pose useful assumptions for their theoretical study, and then the point-wise almost complete convergence, and the asymptotic normality of the kernel estimator for the models are established in Section 4.

2. Model and estimator

2.1. The functional nonparametric framework

Consider random pair (X, Y) where Y is valued in $\mathbb{R}$ and X is valued in some infinite dimensional Hilbertian space $\mathcal{H}$ with scalar product $<\cdot, \cdot>$. Consider that, given $(X_i, Y_i)_{i=1,\dots,n}$ is the statistical sample of pairs which are identically distributed like (X, Y), but not necessarily independent. Henceforth, X is called functional random variable *f.r.v*.

Let x be fixed in $\mathcal{H}$ and let $F(\theta, y, x)$ be the conditional cumulative distribution function *(cond-cdf)* of Y given $<\vartheta,X>=<\vartheta,x>$ specifically: $\forall y \in \mathbb{R}$, $F(\vartheta,y,x) = \mathbb{P}(Y \leq y| <\vartheta,X>=<\vartheta,x>)$. In saying that, one is implicitly assuming the existence of a regular version of conditional distribution Y given $<\vartheta,X>=<\vartheta,x>$.

For the infinite dimensional purpose, the authors used the term *functional nonparametric*, where *functional* refers to the infinite dimensionality of the data, and *nonparametric* refers to the infinite dimensionality of the model. Such *functional nonparametric* statistics is also called *doubly infinite dimensional* (see Ferraty, and Vieu, 2003 for more details). The term *operational statistics* was used since the target object to be estimated (*cond-cdf* $F(\theta,.,x)$) can be viewed as a nonlinear operator.

2.2. The estimators

In the case of complete data, the kernel estimator $\tilde{F}_n(\theta,.,x)$ of $F(\theta,.,x)$ is presented as follows:

$$\tilde{F}(\theta,t,x)=\frac{\sum_{i=1}^{n}K\left(h_n^{-1}(|<x-X_i,\theta>|)\right)H\left(g_n^{-1}(y-Y_i)\right)}{\sum_{i=1}^{n}K\left(h_n^{-1}(<x-X_i,\theta>)\right)}, \qquad (2.1)$$

where K is a kernel function, H a cumulative distribution function and h_n(resp. g_n) a sequence of positive real numbers. Note that using similar ideas (Roussas, 1969) introduced some related estimates but only in a special case when *X* is real, whilst (Samanta, 1989) produced an earlier asymptotic study.

Such an estimator is unique as soon as H is an increasing continuous function. This approach has been largely where variable *X* is of a finite dimension (see e.g. Cai, 2002; Gannoun et al., 2003; Whang and Zhao, 1999; Zhou and Liang, 2003), the general definition of the γ-order quantile is given as:

$$\vartheta_\theta(\gamma,x)=\inf\{y\in\mathbb{R},F(\theta,y,x)\geq\gamma\}.$$

In order to simplify the framework and to focus on the main interest of this paper (the functional feature of <*ϑ,X*>), it was assumed that $F(\theta,.,x)$ is strictly increasing and continuous in the neighbourhood of $\vartheta_\theta(\gamma,x)$. This ensures that conditional quantile $\vartheta_\theta(\gamma,x)$ is uniquely defined by:

$$\vartheta_\theta(\gamma,x)=F^{-1}(\theta,\gamma,x),\ \forall\gamma\in(0,1). \qquad (2.2)$$

As a by-product of (2.1) and (2.2), it is easy to derive estimator $\tilde{\vartheta}_\theta(\gamma,x)$ of $\vartheta_\theta(\gamma,x)$:

$$\tilde{\vartheta}_\theta(\gamma,x)=\tilde{F}^{-1}(\theta,\gamma,x).$$

Meanwhile, in an incomplete case with missing at random for the response variable, observe $(X_i,Y_i,\delta_i)_{1\leq i\leq n}$ where X_i is observed completely, and $\delta_i=1$ if Y_i is observed and $\delta_i=0$ otherwise. Thus, one defines the Bernoulli random variable δ by

$$\mathbb{P}(\delta=1|\langle X,\theta\rangle=\langle x,\theta\rangle,Y=y)=\mathbb{P}(\delta=1|\langle X,\theta\rangle=\langle x,\theta\rangle)=p(x,\theta),$$

where $p(x,\theta)$ is a functional operator which is conditionally only on X. Therefore, the estimator of $F(\theta,y,x)$ in the single index model with response MAR is given by:

$$\hat{F}(\theta,t,x)=\frac{\sum_{i=1}^{n}\delta_iK\left(h_n^{-1}(|<x-X_i,\theta>|)\right)H\left(g_n^{-1}(y-Y_i)\right)}{\sum_{i=1}^{n}\delta_iK\left(h_n^{-1}(<x-X_i,\theta>)\right)}=\frac{\hat{F}_N(\theta,y,x)}{\hat{F}_D(\theta,x)},$$

where $K_i(\theta,x):=\mathrm{K}(h_n^{-1}|<x-X_i,\theta>|)$, $H_i(y)=H\left(g_n^{-1}(y-Y_i)\right)$, $\hat{F}_N(\theta,y,x)=\frac{\sum_{i=1}^{n}\delta_iK_i(\theta,x)H_i(y)}{n\mathbb{E}(K_1(\theta,x))}$ and $\hat{F}_D(\theta,x)=\frac{\sum_{i=1}^{n}\delta_iK_i(\theta,x)}{n\mathbb{E}(K_1(\theta,x))}$.

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

$$\hat{\vartheta}_\theta(\gamma,x)=\hat{F}^{-1}(\theta,\gamma,x)\ =\inf\{t\in\mathbb{R},\hat{F}(\theta,y,x)\geq\gamma\},$$

which satisfies

$$\hat{F}\left(\theta,\hat{\vartheta}_\theta(\gamma,x),x\right)=\gamma.$$

2.3. Assumptions on the functional variable

Let N_x be a fixed neighbourhood of *x* in $\mathcal{H}$ and let $B_\vartheta(x, h)$ be the ball of centre *x* and radius *h*, namely $B_\vartheta(x, h) = \{\chi \in \mathcal{H}: 0 <| < x - \chi, \vartheta>| < h\}$, $d_\vartheta(x, X_i) = | < x - X_i, \vartheta>|$ denote a random variable such that its cumulative distribution function is given by $\phi_{\theta,x}(u) = \mathbb{P}(d_\theta(x, X_i) \leq u) = \mathbb{P}(X_i \in B_\theta(x, u))$, where $S_\mathbb{R}$ is a fixed compact of $\mathbb{R}^+$.

Now, consider the following basic assumptions that are necessary in deriving the main result of this paper.

(H1) $\mathbb{P}(X \in B_\theta(x, h_n)) =: \phi_{\theta,x}(h_n) > 0; \phi_{\theta,x}(h_n) \longrightarrow 0$ as $h_n \rightarrow 0$.

2.4. The nonparametric model

As usual in nonparametric estimation, suppose that *cond-cdf* $F(\theta,.,x)$ verifies some smoothness constraints. Let α_1 and α_2 be two positive numbers; such that:

(H2) $F(\theta,.,x)$ is differentiable continuous and it has a first derivative uniformly bounded, such that:

$$\exists \alpha_1, \alpha_2 > 0 \forall (x_1, x_2) \in N_x \times N_x, \forall (y_1, y_2) \in S_\mathbb{R} \times S_\mathbb{R}$$

(i) $|F^{(j)}(\theta, y_1, x_1) - F^{(j)}(\theta, y_2, x_2)| \leq C_{\theta,x}(\|x_1 - x_2\|^{\alpha_1} + |y_1 - y_2|^{\alpha_2}), for j = 0,1.$
(ii) $\int y dF(\theta, y, x) < \infty$ for all $\theta, x \in \mathcal{H}$.

3. Asymptotic study

The objective of this paragraph was to adapt these ideas to the framework of a functional explanatory variable, and to construct a kernel-type estimator of conditional distribution function $F(\theta, y, x)$ adapted to the MAR response samples. The objective was to establish almost complete convergence[1] of the kernel estimator $\hat{F}(\theta, y, x)$ when considering a model in which the response variable is missing. The results presented are accompanied by the data on the rate of convergence. In what follows *C* and *C'* denote generic strictly positive real constants, h_n (resp. g_n) is a sequence which tends to 0 with n.

3.1. Pointwise almost complete convergence

In addition to the assumptions introduced in Section 2.4, additional conditions were necessary. These assumptions, needed later, concerning the parameters of the estimator, i.e. K, H, h_n and g_n, which are not very restrictive. Indeed, on the one hand, they are rather inherent in the estimation problem of $F(\theta, y, x)$, and on the other, they correspond to the assumptions usually made in the context of non-functional variables. More precisely, the authors introduced the following conditions which guarantee the good behaviour of the estimator $\widehat{F}(\theta,.,x)$:

(H3) Kernel *H* is a positive bounded function such that:

(i) $\forall (y_1, y_2) \in \mathbb{R}^2, |H(y_1) - H(y_2)| \leq C|y_1 - y_2|$, $\int |y|^{\alpha_2} H^{(1)}(y) dy < \infty$ and $\lim_{n\to\infty} n^\varsigma g_n = \infty$, for some $\varsigma > 0$.
(ii) The restriction of H to set $\{u \in \mathbb{R}, H(u) \in (0,1)\}$ is a strictly increasing function.
(iii) The support of $H^{(1)}$ is compact and $H^{(1)}$ exists and is bounded.

[1] Recall that sequence $(S_n)_{n\in\mathbb{N}}$ of random variables is said to converge almost completely to some variable $S, \epsilon > 0$, one has $\sum_n \mathbb{P}(|S_n - S| > \epsilon) < \infty$. This mode of convergence implies both almost sure and in probability convergence (see e.g. Bosq and Lecoutre, 1987).

(H4) K is a positive bounded function with support [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) $p(x,\theta)$ is continuous in the neighbourhood of x, such that $0 < p(x,\theta) < 1$.

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

(H7) Bandwidth h_K and h_H, small-ball probability $\varphi_{\vartheta,x}(h_n)$ satisfying

(i) $ng_n^2\phi_{\theta,x}^2(h_n) \longrightarrow 0$ and $\frac{ng_n^3\phi_{\theta,x}(h_n)\log n}{\log^2 n} \longrightarrow \infty, as n \to \infty$.
(ii) $ng_n^2\phi_{\theta,x}^3(h_n) \longrightarrow 0, as n \to \infty$.

Remark 3.1

(H6) ensures the existence of $\hat{\vartheta}_\theta(\gamma,x)$, while (H5) ensures its uniqueness.

(H1)-(H4) and (H7) are standard assumptions for the distribution conditional estimation in a single functional index model, adopted by (Bouchentouf et al., 2014) for the i.i.d case.

Theorem 3.1. Suppose that hypotheses (H1)-(H5) are satisfied, if $\exists\beta > 0, n^\beta g_n \underset{n\to\infty}{\longrightarrow} \infty$, and if

$$\frac{\log n}{ng_n\phi_{\theta,x}(h_n)} \underset{n\to\infty}{\longrightarrow} 0,$$

then for $j = 0,1$ one has

1. $\sup_{t\in S_{\mathbb{R}}}\left|\hat{F}^{(j)}(\theta,y,x) - F^{(j)}(\theta,y,x)\right| = O\left(h_n^{\alpha_1} + g_n^{\alpha_2}\right) + O_{a.co.}\left(\sqrt{\frac{\log n}{ng_n^{j+1}\phi_{\theta,x}(h_n)}}\right)$.
2. $\hat{\vartheta}_\theta(\gamma,x) - \vartheta_\theta(\gamma,x) \underset{n\to\infty}{\longrightarrow} 0,\ a.co.$
3. $\hat{\vartheta}_\theta(\gamma,x) - \vartheta_\theta(\gamma,x) = O\left(h_n^{\alpha_1} + g_n^{\alpha_2}\right) + O_{a.co.}\left(\frac{\log n}{ng_n\phi_{\theta,x}(h_n)}\right)$.

Proof. Similarly to the proof of Proposition 3.1 and Theorem 3.2 in (Kadiri et al., 2023)), it can be completed easily. Here the authors omitted its proof.

4. Asymptotic normality

The asymptotic normality of the semi-parametric estimators of the conditional quantile for functional data in the Single Index Model (SIM) with missing data at random (MAR) is an important property that establishes the limiting distribution of the estimators as the sample size increases. Although specific results may vary depending on the assumptions and estimation methods used, it allows is to construct confidence intervals and hypothesis tests for the estimated quantile; in this section the asymptotic normality of estimator $\hat{F}(\theta,.,x)$ in the single functional index model is established.

Theorem 4.1 Under assumptions one has (H1) to (H7)-(ii) for all $x \in \mathcal{H}$

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

where $\sigma^2(\theta,t,x) = \frac{M_2(\theta,x)}{(M_1(\theta,x))^2}F(\theta,t,x)(p(x,\theta) - F(\theta,t,x))$ with $M_l(\theta,x) = K^l(1) - \int_0^1 \left(K^l\right)'(u)$ $\beta_{\theta,x}(u)du,\ l = 1,2,$ $\overset{\mathcal{D}}{\to}$means the convergence in distribution.

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

$$\widehat{F}(\theta,t,x)-F(\theta,t,x)=\frac{\widehat{F}_N(\theta,t,x)}{\widehat{F}_{\mathrm{D}}(\theta,x)}-\frac{M_1(\theta,x)F(\theta,t,x)}{M_1(\theta,x)}=\frac{1}{\widehat{F}_{\mathrm{D}}(\theta,x)}\left(\widehat{F}_N(\theta,t,x)-\mathbb{E}\widehat{F}_N(\theta,t,x)\right)$$
$$-\frac{1}{\widehat{F}_{\mathrm{D}}(\theta,x)}\left(M_1(\theta,x)F(\theta,t,x)-\mathbb{E}\widehat{F}_N(\theta,t,x)\right)+\frac{F(\theta,t,x)}{\widehat{F}_{\mathrm{D}}(\theta,x)}\left(M_1(\theta,x)-\mathbb{E}\widehat{F}_D(\theta,x)\right) \tag{4.1}$$
$$-\frac{F(\theta,t,x)}{\widehat{F}_{\mathrm{D}}(\theta,x)}\left(\widehat{F}_D(\theta,x)-\mathbb{E}\widehat{F}_D(\theta,x)\right)=\frac{1}{\widehat{F}_{\mathrm{D}}(\theta,x)}A_n(\theta,t,x)+B_n(\theta,t,x),$$

where

$$A_n(\theta,t,x)=\frac{1}{n\mathbb{E}\left(K_1(\theta,x)\right)}\sum_{i=1}^{n}\{\left(H_i(t)-F(\theta,t,x)\right)\delta_iK_i(\theta,x)$$
$$-\mathbb{E}\left[\left(H_i(t)-F(\theta,t,x)\right)\delta_iK_i(\theta,x)\right]\}=\frac{1}{n\mathbb{E}\left(K_1(\theta,x)\right)}\sum_{i=1}^{n}N_i\,(\theta,t,x)$$

and

$$B_n(\theta,t,x)=M_1(\theta,x)F(\theta,t,x)-\mathbb{E}\widehat{F}_N(\theta,t,x)+F(\theta,t,x)\ \left(M_1(\theta,x)-\mathbb{E}\widehat{F}_D(\theta,x)\right).$$

It follows that

$$n\phi_{\theta,x}(h_K)Var\left(A_n(\theta,t,x)\right)=\frac{\phi_{\theta,x}(h_K)}{\mathbb{E}^2\left(K_1(\theta,x)\right)}Var\left(N_1(\theta,t,x)\right)=V_n(\theta,t,x).$$

Then, the proof of Theorem 4.1 can be deduced from the following lemmas.

Lemma 4.1. Under the assumptions of Theorem 4.1

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

Proof.

$$V_n(\theta,t,x)=\frac{\phi_{\theta,x}(h_K)}{\mathbb{E}^2\left(K_1(\theta,x)\right)}\mathbb{E}\left[\delta_1K_1^2(\theta,x)\left(H_1(t)-F(\theta,t,x)\right)^2\right]$$
$$=\frac{\phi_{\theta,x}(h_K)}{\mathbb{E}^2\left(K_1(\theta,x)\right)}\mathbb{E}\left[K_1^2(\theta,x)\mathbb{E}\left(\delta_1\left(H_1(t)-F(\theta,t,x)\right)^2\middle|\langle\theta,X_1\rangle\right)\right]. \tag{4.2}$$

Using the definition of conditional variance

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

where $J_{1n}=Var(\delta_1H_1(t)|\langle\theta,X_1\rangle)$, $J_{2n}=[\mathbb{E}(H_1(t)|\langle\theta,X_1\rangle)-F(\theta,t,x)]^2$.

- Concerning J_{1n}

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

As for $\mathcal{J}_1$, by the property of double conditional expectation, one obtains

$$\mathcal{J}_1=\mathbb{E}\left(\delta_1H^2\left(\frac{t-Y_1}{h_H}\right)\middle|\langle\theta,X_1\rangle\right)=p(x,\theta)\int H^2\left(\frac{t-v}{h_H}\right)dF(\theta,v,X_1)$$
$$=p(x,\theta)\int H^2(u)\,dF(\theta,t-uh_H,X_1). \tag{4.3}$$

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

$$\int H^2(u)\, dF(\theta, t-uh_H, X_1) = \int 2H(u)H'(u)\, F(\theta, t-uh_H, X_1)du$$
$$= \int 2H(u)H'(u)\big(F(\theta, t-uh_H, X_1) - F(\theta, t, x)\big)du$$
$$+ \int 2H(u)H'(u)\, F(\theta, t, x)du.$$

Clearly,

$$\int 2\, H(u)H'(u)dF(\theta, t, x) = [H^2(u)F(\theta, t, x)]_{-\infty}^{+\infty} = F(\theta, t, x), \tag{4.4}$$

thus

$$\int H^2(u)dF(\theta, t-uh_H, X_1) = F(\theta, t, x) + \mathcal{O}\big(h_K^{\alpha_1} + h_H^{\alpha_2}\big). \tag{4.5}$$

As for J_{2n}, by (H2)-(H3), (H5), and using Lemma 3.1, one obtains that $J_{2n} \to 0$ as $n \to \infty$.

- Concerning $\mathcal{J}_2$

$$\mathcal{J}_2^{'} = \mathbb{E}(\delta_1 H_1(t)|\langle\theta, X_1\rangle)$$
$$= p(x,\theta)\int H\left(\frac{t-v}{h_H}\right) f(\theta, t, X_1)dv^{\cdot}$$

Moreover, by integration by parts and changing variables

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

the last equality is due to the fact that H' is a probability density. Thus

$$\mathcal{J}^{'}_2 = F(\theta, t, x) + \mathcal{O}\big(h_K^{\alpha_1} + h_H^{\alpha_2}\big). \tag{4.6}$$

Finally, by hypothesis (H3) one obtains $\mathcal{J}_2 \to F^2(\theta, t, x)$. Meanwhile, by (H1)-(H2), (H4) and (H6), it follows that:

$$\frac{\phi_{\theta,x}(h_K)\mathbb{E}K_1^2(\theta, x)}{\mathbb{E}^2\big(K_1(\theta, x)\big)} \xrightarrow[n\to\infty]{} \frac{M_2(\theta, x)}{(M_1(\theta, x))^2},$$

which leads to combining equations (4.2)-(4.6).

$$V_n(\theta, t, x) \xrightarrow[n\to\infty]{} \frac{M_2(\theta, x)}{(M_1(\theta, x))^2} F(\theta, t, x)(p(x,\theta) - F(\theta, t, x)).$$

Lemma 4.2. If the assumptions (H1) to (H7) are satisfied, then

$$\sqrt{n\phi_{\theta,x}(h_n)}B_n(\theta, t, x) \to 0, \text{ in probability.}$$

Proof.

One has

$$\sqrt{n\phi_{\theta,x}(h_n)}B_n(\theta, t, x) = \frac{\sqrt{n\phi_{\theta,x}(h_n)}}{\widehat{F}_D(\theta, x)}\{\mathbb{E}\widehat{F}_N(\theta, t, x) - M_1(\theta, x)F(\theta, t, x)$$
$$+F(\theta, t, x)\left(M_1(\theta, x) - \mathbb{E}\widehat{F}_D(\theta, x)\right).$$

Firstly, observe that the results below as $n \to \infty$

$$\frac{1}{\phi_{\theta,x}(h_n)}\mathbb{E}\left[K^l\left(\frac{\langle\theta, x - X_i\rangle}{h_n}\right)\right] \to M_l(\theta, x), \text{for } l = 1,2, \tag{4.7}$$

$$\mathbb{E}\hat{F}_D(\theta, x) \to M_1(\theta, x) \text{ and } \mathbb{E}\hat{F}_N(\theta, t, x) \to M_1(\theta, x)F(\theta, t, x), \tag{4.8}$$

can be proved in the same way as in (Ezzahrioui and Ould-Saïd, 2008) corresponding to their Lemmas 5.1 and 5.2, and then their proofs are omitted.

Secondly, on the one hand, making use of (4.7) and (4.8), as$n \to \infty$

$$\left\{\mathbb{E}\hat{F}_N(\theta, t, x) - M_1(\theta, x)F(\theta, t, x) + F(\theta, t, x)\left(M_1(\theta, x) - \mathbb{E}\hat{F}_D(\theta, x)\right)\right\} \to 0.$$

On the other hand,

$$\frac{\sqrt{n\phi_{\theta,x}(h_n)}}{\hat{F}_{\mathrm{D}}(\theta, x)} = \frac{\sqrt{n\phi_{\theta,x}(h_n)}\hat{F}'(\theta, t, x)}{\hat{F}_{\mathrm{D}}(\theta, x)\hat{F}'(\theta, t, x)} = \frac{\sqrt{n\phi_{\theta,x}(h_n)}\hat{F}'(\theta, t, x)}{\hat{F}'_N(\theta, t, x)}.$$

Then, using Proposition 3.1, it suffices to show that $\frac{\sqrt{n\phi_{\theta,x}(h_n)}}{\hat{F}'_N(\theta,t,x)}$ tends to zero as $n$goes to infinity. Indeed

$$\hat{F}'_N(\theta, t, x) = \frac{1}{ng_n\mathbb{E}\left(K_1(\theta, x)\right)}\sum_{i=1}^{n}\delta_i K\left(h_n^{-1}(< x - X_i, \theta >)\right)H'\left(g_n^{-1}(y - Y_i)\right).$$

Since K *and* H' *are* continuous with support on [0,1], then by (H3) and (H4)

$\exists m = \min_{[0,1]} K(t)H'(t)$ it follows that

$$\hat{F}'_N(\theta, t, x) \geq \frac{m}{g_n\phi_{\theta,x}(h_n)},$$

which gives

$$\frac{\sqrt{n\phi_{\theta,x}(h_n)}}{\hat{F}'_N(\theta, t, x)} \leq \frac{\sqrt{ng_n^2\phi_{\theta,x}^3(h_n)}}{m}.$$

Finally, using (H7)-(ii), completes the proof of Lemma 4.2.

4.1. Application: The conditional quantile in functional single-index model

The main objective of this section was to establish the asymptotic normality of the conditional quantile estimator of *T* given *<ϑ,X>=<ϑ,x>* denoted by $\vartheta_\theta(\gamma, x)$.

Corollary 4.1. If assumptions (H1) to (H7) are satisfied and if γ is the unique order of the quantile such that $\gamma = F(\theta, \vartheta_\theta(\gamma, x), x) = \hat{F}\left(\theta, \hat{\vartheta}_\theta(\gamma, x), x\right)$:

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

where $\Sigma^2(\theta, \vartheta_\theta(\gamma, x), x) = \frac{\sigma^2(\theta,\vartheta_\theta(\gamma,x),x)}{f^2(\theta,\vartheta_\theta(\gamma,x),x)} = \frac{M_2(\theta,x)}{(M_1(\theta,x))^2}\frac{\gamma[p(\theta,x)-\gamma]}{f^2(\theta,\vartheta_\theta(\gamma,x),x)}$.

Proof.

For Corollary 4.1, making use of (4.1), one has

$$\begin{aligned}\sqrt{n\phi_{\theta,x}(h_K)}\left(\hat{\vartheta}_\theta(\gamma,x)-\vartheta_\theta(\gamma,x)\right) &= \sqrt{n\phi_{\theta,x}(h_n)}\frac{\hat{F}(\theta,\vartheta_\theta(\gamma,x),x)}{\hat{F}'\left(\theta,\hat{\vartheta}^*_\theta(\gamma,x),x\right)} \\ &-\sqrt{n\phi_{\theta,x}(h_n)}\frac{F(\theta,\vartheta_\theta(\gamma,x),x)}{\hat{F}'\left(\theta,\hat{\vartheta}^*_\theta(\gamma,x),x\right)} \\ &= \frac{\sqrt{n\phi_{\theta,x}(h_n)}A_n(\theta,t,x)}{\hat{F}'\left(\theta,\hat{\vartheta}^*_\theta(\gamma,x),x\right)}-\frac{\sqrt{n\phi_{\theta,x}(h_n)}B_n(\theta,t,x)}{\hat{F}'\left(\theta,\hat{\vartheta}^*_\theta(\gamma,x),x\right)}\end{aligned} \quad . \tag{4.9}$$

Then, using Theorem 4.1 and Lemma 4.1 provides the result.

4.2. Confidence bands

The asymptotic variances $\sigma^2(\theta,t,x)$ and $\Sigma^2(\theta,\vartheta_\theta(\gamma,x),x)$ in Theorem 4.1 and Corollary 4.1 depend on some unknown quantities including M_1, M_2, $\phi(u)$, $\vartheta_\theta(\gamma,x)$, $p(\theta,x)$ and $f(\theta,\vartheta_\theta(\gamma,x),x)$. Therefore, $p(\theta,x)$,$\vartheta_\theta(\gamma,x)$, and $f(\theta,\vartheta_\theta(\gamma,x),x)$ should can be estimated respectively by $P_n(\theta,x)$, $\hat{\vartheta}_\theta(\gamma,x)$ and $\hat{f}(\theta,\vartheta_\theta(\gamma,x),x)$ and $\hat{\vartheta}_\theta(\gamma,x)$. Moreover, using the decomposition given by assumption (H1), one can estimate $\phi_{\theta,x}(.)$ by $\hat{\phi}_{\theta,x}(.)$ As the unknown functions $M_j := M_j(\theta,x)$ and $f(\theta,y,x)$intervening in the expression of the variance, hence it was necessary to estimate quantities $M_1(\theta,x)$, $M_2(\theta,x)$ and $F(\theta,y,x)$, respectively.

By assumptions (H1) to (H4) it is known that $M_j(\theta,x)$ can be estimated by $\widehat{M}_j(\theta,x)$ which is defined as:

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

with $\mathbf{1}_{\{.\}}$ being the indicator function. Finally, the estimator of $p(\theta,x)$ is denoted by

$$P_n(\theta,x)=\frac{\sum_{i=1}^n \delta_i K\left(h_n^{-1}(< x-X_i,\theta >)\right)}{\sum_{i=1}^n K\left(h_n^{-1}(< x-X_i,\theta >)\right)}.$$

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

$$\hat{\sigma}^2(\theta,y,x)=\frac{\widehat{M}_2(\theta,x)}{\left(\widehat{M}_1(\theta,x)\right)^2}\hat{f}(\theta,y,x)\int H^2(u)du$$

so one can derive the following corollary:

Corollary 4.2. Under the assumptions of Theorem 4.1, $K'$and $(K^2)'$are integrable functions, then as $n\to\infty$.

1. $\frac{\widehat{M}_1}{\sqrt{\widehat{M}_2}}\sqrt{\frac{n\hat{\phi}_{\theta,x}(h_n)}{\hat{F}(\theta,y,x)[P_n(\theta,x)-\hat{F}(\theta,y,x)]}}\left(\hat{F}(\theta,y,x)-F(\theta,y,x)\right)\xrightarrow{\mathcal{D}}\mathcal{N}(0,1)$.
2. $\frac{\widehat{M}_1\hat{f}(\theta,\hat{\vartheta}_\theta(\gamma,x),x)}{\sqrt{\widehat{M}_2}}\sqrt{\frac{n\hat{\phi}_{\theta,x}(h_n)}{\gamma[P_n(\theta,x)-\gamma]}}\left(\hat{\vartheta}_\theta(\gamma,x)-\vartheta_\theta(\gamma,x)\right)\xrightarrow{\mathcal{D}}\mathcal{N}(0,1)$.

Proof.

Observe that

1.

$$\frac{\widehat{M}_1}{\sqrt{\widehat{M}_2}}\sqrt{\frac{n\hat{\phi}_{\theta,x}(h_n)}{\hat{F}(\theta,y,x)\left[P_n(\theta,x)-\hat{F}(\theta,y,x)\right]}}\left(\hat{F}(\theta,y,x)-F(\theta,y,x)\right)$$
$$=\frac{\widehat{M}_1\sqrt{M_2}}{M_1\sqrt{\widehat{M}_2}}\sqrt{\frac{n\hat{\phi}_{\theta,x}(h_n)[p(\theta,x)-F(\theta,y,x)]F(\theta,y,x)}{\hat{F}(\theta,y,x)\left[P_n(\theta,x)-\hat{F}(\theta,y,x)\right]n\phi_{\theta,x}(h_n)}}$$
$$\times\frac{M_1}{\sqrt{M_2}}\sqrt{\frac{n\phi_{\theta,x}(h_n)}{F(\theta,y,x)[p(\theta,x)-F(\theta,y,x)]}}\left(\hat{F}(\theta,y,x)-F(\theta,y,x)\right).$$

Via Theorem 4.1,

$$\frac{M_1}{\sqrt{M_2}}\sqrt{\frac{n\phi_{\theta,x}(h_n)}{F(\theta,y,x)[p(\theta,x)-F(\theta,y,x)]}}\left(\hat{F}(\theta,y,x)-F(\theta,y,x)\right)\xrightarrow{\mathcal{D}}\mathcal{N}(0,1).$$

Next, by (Laib and Louani, 2010), one can prove that

$$\widehat{M}_1\xrightarrow{\mathbb{P}}M_1,\widehat{M}_2\xrightarrow{\mathbb{P}}M_2 \text{ and } \frac{\hat{\phi}_{\theta,x}(h_n)}{\sqrt{\phi_{\theta,x}(h_n)}}\xrightarrow{\mathbb{P}}1 \text{ as } n\to\infty.$$

If in addition, one considers Lemma 4.2 and (4.9), the consistency of $P_n(\theta,x)$ to $p(\theta,x)$ according to (Deheuvels and Einmahl, 2000), then

$$\frac{\widehat{M}_1\sqrt{M_2}}{M_1\sqrt{\widehat{M}_2}}\sqrt{\frac{n\hat{\phi}_{\theta,x}(h_n)[p(\theta,x)-F(\theta,y,x)]F(\theta,y,x)}{\hat{F}(\theta,y,x)\left[P_n(\theta,x)-\hat{F}(\theta,y,x)\right]n\phi_{\theta,x}(h_n)}}\to 1\ a.s.$$

2.

$$\frac{\widehat{M}_1\hat{f}\left(\theta,\hat{\vartheta}_\theta(\gamma,x),x\right)}{\sqrt{\widehat{M}_2}}\sqrt{\frac{n\hat{\phi}_{\theta,x}(h_n)}{\gamma[P_n(\theta,x)-\gamma]}}\left(\hat{\vartheta}_\theta(\gamma,x)-\vartheta_\theta(\gamma,x)\right)$$
$$=\frac{\widehat{M}_1\sqrt{M_2}}{M_1\sqrt{\widehat{M}_2}}\sqrt{\frac{n\hat{\phi}_{\theta,x}(h_n)[p(\theta,x)-\gamma]}{[P_n(\theta,x)-\gamma]n\phi_{\theta,x}(h_n)}}\frac{\hat{f}\left(\theta,\hat{\vartheta}_\theta(\gamma,x),x\right)}{f(\theta,\vartheta_\theta(\gamma,x),x)}.$$
$$\times\frac{M_1}{\sqrt{M_2}}\sqrt{\frac{n\phi_{\theta,x}(h_n)}{\gamma[p(\theta,x)-\gamma]}}f(\theta,\vartheta_\theta(\gamma,x),x)\left(\hat{\vartheta}_\theta(\gamma,x)-\vartheta_\theta(\gamma,x)\right)$$

Making use of Corollary 4.1, one obtains

$$\frac{M_1}{\sqrt{M_2}}\sqrt{\frac{n\phi_{\theta,x}(h_n)}{\gamma[p(\theta,x)-\gamma]}}f(\theta,\vartheta_\theta(\gamma,x),x)\left(\hat{\vartheta}_\theta(\gamma,x)-\vartheta_\theta(\gamma,x)\right)\xrightarrow{\mathcal{D}}\mathcal{N}(0,1).$$

Further, by considering Lemma 4.2, (4.9), and the consistency of $P_n(\theta,x)$ to $p(\theta,x)$, as $n\to\infty$.

$$\frac{\widehat{M}_1\sqrt{M_2}}{M_1\sqrt{\widehat{M}_2}}\sqrt{\frac{n\hat{\phi}_{\theta,x}(h_K)[p(\theta,x)-\gamma]}{[P_n(\theta,x)-\gamma]n\phi_{\theta,x}(h_K)}}\frac{\hat{f}\left(\theta,\hat{\vartheta}_\theta(\gamma,x),x\right)}{f(\theta,\vartheta_\theta(\gamma,x),x)}\xrightarrow{\mathbb{P}}1.$$

Hence, the proof is completed.

Remark 4.1.

Thus, following Corollary 4.2, the asymptotic $(1-\xi)$confidence interval of $F(\theta,y,x)$ and $\vartheta_\theta(\gamma,x)$ respectively, which are expressed as follows:

$$\hat{F}(\theta,y,x) \pm \eta_{\gamma/2}\frac{\widehat{M}_1}{\sqrt{\widehat{M}_2}}\sqrt{\frac{\hat{F}(\theta,y,x)\left[P_n(\theta,x)-\hat{F}(\theta,y,x)\right]}{n\hat{\phi}_{\theta,x}(h_n)}},$$

and

$$\hat{\vartheta}_\theta(\gamma,x) \pm \eta_{\gamma/2}\frac{\widehat{M}_1\hat{f}\left(\theta,\hat{\vartheta}_\theta(\gamma,x),x\right)}{\sqrt{\widehat{M}_2}}\sqrt{\frac{\gamma[P_n(\theta,x)-\gamma]}{n\hat{\phi}_{\theta,x}(h_n)}},$$

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

5. Conclusions

In this study, the authors were mainly interested in the nonparametric estimation of the conditional distribution function and/or conditional quantile estimator for a variable explanatory functionally conditioned to an actual response variable via a functional single index model, when the variable of interest is subject to the presence of randomly missing data, involving both some (semi-parametric) single model structure and also some censoring process on the variables. The paper shows that the estimator provided good predictions under this model.

One of the main contributions of this work is the choice of semi-metrics. Indeed, it is well known that, in non-parametric functional statistics, the semi-metric of the projection type is very important for increasing the concentration property. 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 it was possible to prove that the variance estimator is almost completely consistent, using analogous ideas. The proofs were based on certain standard assumptions in Functional Data Analysis (FDA). The functional index model is a special case of this family of semi- -metrics because it is based on the projection on a functional direction which is important for the implementation of this method in practice.

References

Bosq, D., and Lecoutre, J. P. (1987). *Théorie de L'estimation Fonctionnelle*. Economica.

Bouchentouf, A. A., Djebbouri, T., Rabhi, A., and Sabri, K. (2014). Strong Uniform Consistency Rates of Some Characteristics of the Conditional Distribution Estimator in the Functional Single Index Model. *Appl. Math*., *41*(4), 301-322.

Cai, Z. (2002). *Regression Quantiles for Time Series*. *Econometric Theory*, (18), 169-192.

Cheng, P. E. (1994). Nonparametric Estimation of Mean Functional with Data Missing at Random. *Journal of the American Statistical Association*, *89*(425), 81-87.

Deheuvels, P., and Einmahl, J. H. J. (2000). Functional Limit Laws for the Increments of Kaplan-Meier Product-Limit Processes and Applications. *The Annals of Probability*, (28), 1301-1335.

Efromovich, S. (2011). Nonparametric Regression with Responses Missing at Random. *Journal of Statistical Planning and Inference*, *141*, 3744-3752.

Ezzahrioui, M., and Ould-Saïd, E. (2008). Asymptotic Results of a Nonparametric Conditional Quantile Estimator for Functional Time Series. *Comm. Statist. Theory and Methods*, *37*(16-17), 2735-2759.

Ferraty, F., Sued, F., and Vieu, P. (2013). Mean Estimation with Data Missing at Random for Functional Covariables. *Statistics*, *47*(4), 688-706.

Ferraty, F., and Vieu, P. (2003). *Functional Nonparametric Statistics: A Double Infinite Dimensional Framework. Recent Advances and Trends in Nonparametric Statistics*, M. Akritas and D. Politis (Ed.). Elsevier.

Gannoun, A., Saracco, J., and Yu, K. (2003). Nonparametric Prediction by Conditional Median and Quantiles. *J. Stat. Plann. and Inf.,* (117), 207-223.

Hamri, M. M., Mekki, S. D., Rabhi, A. and Kadiri, N. (2022). Single Functional Index Quantile Regression for Independent Functional Data Under Right-Censoring. *Econometrics*, *26*(1), 31-62. https://doi.org/10.15611/eada.2022.1.03

Kadiri, N., Mekki, S. D., and Rabhi, A. (2023). Single Functional Index Quantile Regression for Functional Data with Missing Data at Random. *Econometrics. Ekonometria. Advances in Applied Data Analysis*, *27*(3), 1-19. DOI 10.15611/eada.2023.3.01

Liang, H., and de Uña-Alvarez, J. (2010). Asymptotic Normality for Estimator of Conditional Mode under Left-Truncated and Dependent Observations. *Metrika*, *72*(1), 1-19.

Ling, N., Liang, L., and Vieu, P. (2015). Nonparametric Regression Estimation for Functional Stationary Ergodic Data with Missing at Random. *Journal of Statistical Planning and Inference*, (162), 75-87.

Ling, N., Liu, Y., and Vieu, P. (2016). Conditional Mode Estimation for Functional Stationary Ergodic Data with Responses Missing at Random. *Statistics*, *50*(5), 991-1013.

Mekki, S. D., Kadiri, N., and Rabhi, A. (2021). Asymptotic Properties of the Semi-Parametric Estimators of the Conditional Density for Functional Data in the Single Index Model with Missing Data at Random. *Statistica*, *81*(4), 399-422.

Ould-Saïd, E., and Djabrane, Y. (2011). Asymptotic Normality of a Kernel Conditional Quantile Estimator under Strong Mixing Hypothesis and Left-Truncation. *Communications in Statistics. Theory and Methods*, *40*(14), 2605-2627.

Ould-Saïd, E., and Tatachak, A. (2011). A Nonparametric Conditional Mode Estimate under RLT Model and Strong Mixing Condition. *International Journal of Statistics and Economics*, (6), 76-92.

Rabhi, A., Kadiri, N., and Akkal, F. (2021). On the Central Limit Theorem for Conditional Density Estimator in the Single Functional Index Model.*Applications and Applied Mathematics: An International Journal (AAM), 16*(2), 844-866.

Wang, H., and Zhao, Y. (1999). A Kernel Estimator for Conditional *T*-Quantiles for Mixing Samples and Its Strong Uniform Convergence (in Chinese). *Math. Appl. (Wuhan)*, (12), 123-127.

Zhou, Y., and Liang, H. (2003). Asymptotic Properties for L1Norm Kernel Estimator of Conditional Median under Dependence. *J. Nonparametr. Stat*., (15), 205-219.

Asymptotyczna normalność regresji kwantylowej pojedynczego wskaźnika funkcyjnego dla danych funkcjonalnych z losowymi brakującymi danymi

Streszczenie: W artykule autorzy prowadzą rozważania dotyczące problemu nieparametrycznej estymacji funkcji regresji, a mianowicie rozkładu warunkowego i kwantyla warunkowego w modelu pojedynczego indeksu funkcjonalnego (SFIM) przy założeniu niezależnych i z identycznym rozkładem danych z losowymi brakami danych. Głównym rezultatem przeprowadzonych badań było ustalenie asymptotycznych właściwości estymatora, takich jak prawie całkowite współczynniki zbieżności. Co więcej, asymptotyczną normalność konstruktów uzyskano dla pewnych łagodnych warunków. Na koniec omówiono, jak zastosować uzyskany wynik do skonstruowania przedziałów ufności.

Słowa kluczowe: asymptotyczna normalność, funkcjonalna analiza danych, funkcjonalny proces pojedynczego indeksu, estymator jądra, losowe braki, estymacja nieparametryczna, prawdopodobieństwo małej kuli