Przegląd Statystyczny. Statistical Review, 2023, vol. 70, 2, 20–45 https://doi.org/10.59139/ps.2023.02.2
Received: 22.07.2023, revised: 2.10.2023, accepted: 13.10.2023

Some asymptotic results of the estimators for conditional mode for functional data in the single index model missing data at random

Souad Mekkaoui,[a] Nadia Kadiri,[b] Abbes Rabhi[c]

Abstract. In this work, we consider the problem of non-parametric estimation of a regression function, namely the conditional density and the conditional mode in a single functional index model (SFIM) with randomly missing data. The main result of this work is the establishment of the asymptotic properties of the estimator, such as almost complete convergence rates. Moreover, the asymptotic normality of the constructs is obtained under certain mild conditions. We finally discuss how to apply our result to construct confidence intervals.

Keywords: functional data analysis, functional single-index process, kernel estimator, missing at random, non-parametric estimation, small ball probability

JEL: C10, C13, C14, C19, C24

1. Introduction

The Single Index Model (SIM) is a popular framework used for reducing dimensionality and modelling complex relationships between covariates and responses in a simplified way. When dealing with functional data, where each observation is a curve or a function, the SIM is extended to handle functional predictors and responses. When dealing with missing data in the SIM framework, the missingness is assumed to be at random (MAR). This means that the probability of missing values is related to the observed data but not to the missing values themselves. The key idea is that, given the observed data, the missingness mechanism is unrelated to the values that are missing. It is important to note that the choice of approach depends on the specifics of your data, the extent of the missingness, and the assumptions you are willing to make. A careful consideration of the nature of your data and consulting domain experts when handling missing data in the SIM or any other modeling framework is always recommended.

[a] Djillali LIABES University of Sidi Bel Abbes, Laboratory of Mathematics, Algeria, ORCID: https://orcid.org/0009-0005-8611-3333.

[b] Djillali LIABES University of Sidi Bel Abbes, Higher School of Economics of Oran 22000, Sidi Bel Abbes, Algeria, e-mail: nad.kad06@yahoo.com, ORCID: https://orcid.org/0000-0002-3405-7414.

[c] Djillali LIABES University of Sidi Bel Abbes, Laboratory of Mathematics, Algeria, ORCID: https://orcid.org/0000-0001-6740-0226.

The asymptotic properties of semi-parametric estimators of the conditional mode for functional data in the Single Index Model (SIM) with data missing at random (MAR) are an active area of research, and specific results may depend on the particular assumptions and estimation methods employed. However, this work provides a general overview of some relevant concepts and approaches in this context. In the SIM framework, functional data refers to observations that are functions rather than scalar values. The goal is to estimate the conditional mode of a functional-response variable given a set of functional predictors and a single-index variable.

To establish the asymptotic properties of the semi-parametric estimators of the conditional mode for functional data in the SIM with data missing at random, various theoretical conditions need to be satisfied. These conditions often involve assumptions about the functional data, the missing data mechanism, and the model specification. Some common conditions include consistency and efficiency. Specific results in this area may depend on the assumptions and estimation techniques employed in each study. Therefore, it is important to refer to the literature and research articles that focus on the specific estimation method and assumptions one is interested in to obtain more detailed and precise asymptotic properties of the estimators.

One of the most common problems in non-parametric statistics is forecasting. In some situations, regression is the best forecasting tool. Sometimes, however, e.g. in the case where the conditional density is asymmetrical or multimodal, this tool is inadequate. Therefore, the conditional quantile predicts the impact of the variable of interest Y on the explanatory variable X more efficiently. There is scarce literature investigating the statistical properties of a functional non-parametric regression model for missing data when the explanatory variable is infinite dimensional or it is of a functional nature. Recently, Ferraty et al. (2013) proposed to estimate the mean of a scalar response based on an independent and identically distributed (i.i.d) functional sample in which explanatory variables were observed for every subject, and a part of the responses were missing at random (MAR) for some of them. It generalised the results in Cheng (1994) to the case where the explanatory variables are of a functional nature.

To the best of our knowledge, the estimation of a non-parametric conditional distribution in the functional single index structure combining missing data and stationary processes of a functional nature has not yet been studied in statistical literature. Therefore, in this work we investigate a conditional quantile estimation when the data are MAR. Our aim is to develop a functional methodology for dealing with MAR samples in non-parametric problems (namely in the conditional quantile estimation). Then, the asymptotic properties of the estimator are obtained under some mild conditions. Our study considers a model in which the response variable is missing.

2. Literature review

Therefore, within this framework, the independence of the variables was assumed. As far as we know, the estimation of a conditional quantile combining censored data, an independent theory and functional data with single-index structure has not been studied in statistical literature yet. Our paper extends the work of Ling et al. (2015, 2016) and Mekki et al. (2021) to the functional single-index model case.

For the above-mentioned theoretical and application reasons, the statistical community has displayed a great interest in estimating conditional quantiles, especially the conditional median function, as an interesting, alternative predictor to the conditional mean (thanks to its robustness to the presence of outliers) (see Chaudhuri et al., 1997). The estimation of the conditional mode of a scalar response given a functional covariate has attracted the attention of many researchers. Ferraty et al. (2005) introduced a non-parametric estimator of the conditional quantile, defined as the inverse of the conditional distribution function when data are dependent. Ezzahrioui and Ould-Saïd (2008) established the asymptotic normality of the kernel conditional mode estimator. In the censored case, Ould-Saïd and Cai (2005) established a uniform strong consistency of the kernel estimator for the conditional mode function. In this context, we recommend referring to Lemdani et al. (2009) for the estimation of conditional quantiles. Other authors have been interested in the estimation of conditional models when the observations were censored or truncated, eg. Hamri et al. (2022), Liang and de Uña-Alvarez (2010), Ould-Saïd and Tatachak (2011), Ould-Saïd and Yahia (2011), Rabhi et al. (2021), etc.

For instance, Aït-Saidi et al. (2008) were interested in using SFIM to estimate the regression operator, and suggested using a cross-validation procedure allowing the estimation of the unknown link function as well as the unknown functional index. Attaoui (2014) and Attaoui and Ling (2016) studied, respectively, the estimation of the conditional density and the conditional cumulative distribution function based on a SFIM with the assumption that the data satisfy a strong mixing condition. Kadiri et al. (2018) studied the asymptotic properties of the kernel-type estimator of the conditional quantiles when the response was right-censored and the data was sampled from a strong mixing process.

The remaining part of the paper is arranged in the following way: in Section 3, we present the non-parametric estimator of the functional conditional model when the data are MAR. In Section 4, we make assumptions for the theoretical study. The point-wise almost-complete convergence and the uniform almost-complete convergence of the kernel estimator for our models (with rates) are established in Section 5.

3. Model and estimator

3.1. The functional non-parametric framework

Consider a random pair (X, Y), where Y is valued in $\mathbb{R}$ and X is valued in some infinite dimensional Hilbertian space $\mathcal{H}$ with a scalar product $<\cdot, \cdot>$. Let the $(X_i, Y_i)_{i=1,\ldots,n}$ be the statistical sample of pairs which are identically distributed like (X, Y), but not necessarily independent. Henceforward, X will be 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 T given $<\theta, X> = < \theta, x >$, specifically:

$$\forall y \in \mathbb{R},\ F(\theta, y, x) = \mathbb{P}(Y \leq y | <\theta, X> = <\theta, x>).$$

By the above, we are implicitly assuming the existence of a regular version of the conditional distribution Y, given $<\theta, X> = <\theta, x>$.

In our infinite dimensional purpose, we use the 'functional non-parametric' term, where the word 'functional' refers to the infinite dimensionality of the data, and the word 'non-parametric' denotes the infinite dimensionality of the model. Such 'functional non-parametric' statistics is also called 'doubly infinite dimensional' (see Ferraty & Vieu, 2003, for more details). We also use the 'operational statistics' term, since the target object to be estimated (the *cond-df* $f(\theta, ., x)$) can be viewed as a non-linear operator.

3.2. The estimators

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

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

where K and H are kernel functions, and h_n(resp. g_n) is a sequence of positive real numbers. Note that using similar ideas, Roussas (1969) introduced some related estimates, but in the special case where X was real, while Samanta (1989) produced an earlier asymptotic study on the subject.

Meanwhile, in an incomplete case with data missing at random for the response variable, we 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 and $\delta_i = 0$ otherwise. We define 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 conditional only on X.

Therefore, the estimator of $f(\theta, y, x)$ in the single-index model with response MAR is presented as

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

where $K_i(\theta,x) := \mathrm{K}(h_n^{-1}|< x - X_i, \theta >|)$,$H_i(y) = H\left(g_n^{-1}(y - Y_i)\right)$,

$\hat{f}_D(\theta,x) = \frac{\sum_{i=1}^{n}\delta_i K_i(\theta,x)}{n\mathbb{E}(K_1(\theta,x))}$, and $\hat{f}_N(\theta,y,x) = \frac{\sum_{i=1}^{n}\delta_i K_i(\theta,x)H_i(y)}{ng_n\mathbb{E}(K_1(\theta,x))}$.

3.3. Assumptions on the functional variable

Let N_x be a fixed neighborhood of x in $\mathcal{H}$ and let $B_\theta(x,h)$ be a ball of center x and radius h, namely $B_\theta(x,h) = \{\chi\in\mathcal{H}: 0 <| <x-\chi,\theta >| < h\}$, $d_\theta(x,X_i) = | < x - X_i,\theta>|$ denote a random variable such that its cumulative distribution function is given by $\phi_{\theta,x}(u) = \mathbb{P}(d_\theta(x,X_i) \leq u) = \mathbb{P}\left(X_i \in B_\theta(x,u)\right)$.

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

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

3.4. The non-parametric model

As is usually the case in non-parametric estimation, we suppose that the *cond-dff*$(\theta,.,x)$ verifies some smoothness constraints. Let α_1 and α_2 be two positive numbers, such that

(H2) $\forall(x_1,x_2) \in N_x \times N_x, \forall(y_1,y_2) \in S_\mathbb{R} \times S_\mathbb{R}$

(i) $|f(\theta,y_1,x_1) - f(\theta,y_2,x_2)| \leq C_{\theta,x}(\|x_1 - x_2\|^{\alpha_1} + |y_1 - y_2|^{\alpha_2})$

(ii) $\int yf(\theta,y,x)dy < \infty$ for all $\theta, x \in \mathcal{H}$.

4. Asymptotic study

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

4.1. Point-wise almost-complete convergence

Besides the assumptions introduced in Section 3.4, we will need additional conditions. The assumptions we will need later, concerning the parameters of our estimator, i.e. K,H, h_n and g_n, 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, we will introduce the following conditions to ensure the performance of the estimator $\hat{f}(\theta, ., x)$:

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

(i) $\forall (y_1, y_2) \in \mathbb{R}^2, |H(y_1) - H(y_2)| \leq C|y_1 - y_2|$, $\int |y|^{\alpha_2} H(y)dy < \infty$ and $\int yH(y)dy = 0$.

(ii) $H^{(1)}$ and $H^{(2)}$ are bounded with $\int \left(H^{(1)}(t)\right)^2 dt < \infty$.

(H4) K is a positive bounded kernel function with the support of [0,1]: $\forall u \in (0,1)$, $0 < \mathrm{K}(u)$, and the derivative K' exists on [0,1] with $K'(t) < 0$ for all $t \in [0, 1]$ and $\int_0^1 \left(K^j\right)'(t)dt < \infty$ for $j = 1,2$.

(H5) $p(x, \theta)$ is continuous in the neighbourhood of x: $0 < p(x, \theta) < 1$.

THEOREM 1: Suppose that hypotheses (H1)–(H5) are satisfied if $\exists \beta > 0, n^{\beta} g_n \underset{n\to\infty}{\longrightarrow} \infty$, and if

$$\frac{\log n}{n g_n^2 \phi_{\theta,x}(h_n)} \underset{n\to\infty}{\longrightarrow} 0,$$

[1] We say that a sequence $(S_n)_{n\in\mathbb{N}}$ converges almost completely to S if and only if, for any $\epsilon > 0$, we have $\sum_n \mathbb{P}(|S_n - S| > \epsilon) < \infty$.

then we have

$$\sup_{y\in S_{\mathbb{R}}}\left|\hat{f}(\theta,y,x)-f(\theta,y,x)\right| = \mathcal{O}\left(h_n^{\alpha_1}+g_n^{\alpha_2}\right)+\mathcal{O}_{a.co.}\left(\sqrt{\frac{\log n}{ng_n^2\phi_{\theta,x}(h_n)}}\right).$$

PROOF: The proof is based on the following decomposition, valid for any $y \in S_{\mathbb{R}}$:

$$\begin{aligned}\sup_{t\in S_{\mathbb{R}}}\left|\hat{f}(\theta,y,x)-f(\theta,y,x)\right| \le \frac{1}{\hat{f}_D(\theta,x)}\sup_{y\in S_{\mathbb{R}}}\left|\hat{f}_N(\theta,y,x)-\mathbb{E}\hat{f}_N(\theta,y,x)\right| + \\ +\frac{1}{\hat{f}_D(\theta,x)}\sup_{t\in S_{\mathbb{R}}}\left|\mathbb{E}\hat{f}_N(\theta,y,x)-f(\theta,y,x)\right| + \frac{f(\theta,y,x)}{\hat{f}_D(\theta,x)}\sup_{y\in S_{\mathbb{R}}}\left|\hat{f}_D(\theta,x)-\mathbb{E}\hat{f}_D(\theta,x)\right|.\end{aligned} \tag{1}$$

Finally, the proof of this theorem is a direct consequence of the following intermediate results:

LEMMA 1: Suppose that hypotheses (H1)–(H3) and (H5) are satisfied, then we have

$$\sup_{y\in S_{\mathbb{R}}}\left|\mathbb{E}\hat{f}_N(\theta,y,x)-f(\theta,y,x)\right| = \mathcal{O}\left(h_n^{\alpha_1}+g_n^{\alpha_2}\right).$$

PROOF: We have

$$\begin{aligned}I = \mathbb{E}\hat{f}_N(\theta,y,x)-f(\theta,y,x) = \mathbb{E}\left(\frac{1}{ng_n\mathbb{E}(K_1(\theta,x))}\sum_{i=1}^n \delta_i K_i(\theta,x)H_i(y)\right) + \\ -f(\theta,y,x) = \frac{1}{ng_n\mathbb{E}(K_1(\theta,x))}\sum_{i=1}^n \mathbb{E}([\mathbb{E}(\delta_i K_i(\theta,x)H_i(y)|<\theta,X_i>)]) - f(\theta,y,x) = \\ = \frac{1}{g_n\mathbb{E}(K_1(\theta,x))}\mathbb{E}\left(p(x,\theta)K_1(\theta,x)\mathbb{E}(H_1(y))\right) - f(\theta,y,x).\end{aligned}$$

Moreover, by changing variables and using the fact that H is a *df* and uses a double conditioning with respect to Y_1, we can easily obtain

$$\begin{aligned}\mathbb{E}\left(H\left(g_n^{-1}(y-Y_1)\right)|<\theta,X_1>\right) = \int_{\mathbb{R}} H\left(\frac{y-u}{g_n}\right) f(\theta,u,X_1)du = \\ = \int_{\mathbb{R}} H(v)f(\theta,y-vg_n,X_1)dv = \\ = g_n\int_{\mathbb{R}} H(v)\left(f(\theta,y-vg_n,X_1)-f(\theta,u,x)\right)dv + g_n f(\theta,u,x)\int_{\mathbb{R}} H(v)dv.\end{aligned}$$

We can write, because of (H2) and (H3):

$$\begin{aligned}I = \frac{1}{\mathbb{E}K_1}\mathbb{E}\left(p(x,\theta)K_1(\theta,x)\int_{\mathbb{R}} H(v)\left(f(\theta,y-vg_n,X_1)-f(\theta,y,x)\right)dv\right) \le \\ \le C_{\theta,x}\left(p(x,\theta)+o(1)\right)\int_{\mathbb{R}} H(v)\left(h_n^{\alpha_1}+|v|^{\alpha_2}g_n^{\alpha_2}\right)dv \le \mathcal{O}\left(h_n^{\alpha_1}+g_n^{\alpha_2}\right).\end{aligned}$$

Finally, the proof is achieved.

LEMMA 2: Under hypotheses of Theorem 1, we have, as $n \to \infty$,

$$\sup_{y\in S_{\mathbb{R}}}\left|\hat{f}_N(\theta,y,x)-\mathbb{E}\hat{f}_N(\theta,y,x)\right| = \mathcal{O}_{a.co.}\left(\sqrt{\frac{\log n}{ng_n^2\phi_{\theta,x}(h_n)}}\right).$$

PROOF: Using the compactness of $S_{\mathbb{R}}$, we can write that $S_{\mathbb{R}} \subset \bigcup_{j=1}^{\tau_n}(z_j - l_n, z_j + l_n)$, with l_n and τ_n which can be chosen such that $l_n = C\tau_n^{-1} \sim Cn^{-\varsigma-1/2}$. Taking $m_y = \arg\min_{j\in\{z_1,\cdots,z_{\tau_n}\}}|y-m_j|$. Thus, we have the following decomposition:

$$\begin{aligned}\sup_{y\in S_{\mathbb{R}}}\left|\hat{f}_N(\theta,y,x)-\mathbb{E}\hat{f}_N(\theta,y,x)\right| &\le \sup_{y\in S_{\mathbb{R}}}\left|\hat{f}_N(\theta,t,x)-\hat{f}_N(\theta,m_y,x)\right| \\ &+ \sup_{y\in S_{\mathbb{R}}}\left|\hat{f}_N(\theta,m_y,x)-\mathbb{E}\hat{f}_N(\theta,m_y,x)\right| \\ &+ \sup_{y\in S_{\mathbb{R}}}\left|\mathbb{E}\hat{f}_N(\theta,m_y,x)-\mathbb{E}\hat{f}_N(\theta,y,x)\right| \\ &\le B_1+B_2+B_3.\end{aligned}$$

As the first and the third term can be treated in the same manner, we deal only with the first term. By (H3)-(i), which in particular implies that H is a Hölder continuous with order one, we can write:

$$B_1 \le \frac{1}{ng_n\mathbb{E}(K_1(\theta,x))}\sup_{y\in S_{\mathbb{R}}}\sum_{i=1}^{n}\delta_i\left|H_i(y)-H_i(m_y)\right|K_i(\theta,x) \le$$
$$\le \frac{C}{ng_n\mathbb{E}(K_1(\theta,x))}\sup_{y\in S_{\mathbb{R}}}\frac{|y-m_y|}{g_n}\times\sum_{i=1}^{n}\delta_iK_i(\theta,x) \le \frac{Cl_n}{ng_n^2\mathbb{E}(K_1(\theta,x))}\times\sum_{i=1}^{n}\delta_iK_i(\theta,x).$$

Using $\mathbb{E}\hat{f}_{\mathrm{D}}(\theta,x) = p(x,\theta)$, (H3)-(i) and $\lim_{n\to\infty} n^{\beta}g_n = \infty$, it follows that

$$B_1 \xrightarrow[n\to\infty]{} \infty\,.$$

Thus, for n large enough, we have

$$B_1 = \mathcal{O}_{a.co.}\left(\sqrt{\frac{\log n}{ng_n^2\phi_{\theta,x}(h_n)}}\right).$$

Following similar arguments, we can have

$$B_3 \le B_1.$$

Concerning B_2, let us consider $\varepsilon = \varepsilon_0 \sqrt{\frac{\log n}{n g_n^2 \phi_{\theta,x}(h_n)}}$, for all $\varepsilon_0 > 0$, we have

$$\mathbb{P}\left(\sup_{y\in S_{\mathbb{R}}}\left|\hat{f}_N(\theta,m_y,x) - \mathbb{E}\hat{f}_N(\theta,m_y,x)\right| > \varepsilon\right) \leq$$
$$\leq \mathbb{P}\left(\max_{j\in\{1,\cdots,\tau_n\}}\left|\hat{f}_N(\theta,m_y,x) - \mathbb{E}\hat{f}_N(\theta,m_y,x)\right| > \varepsilon\right) \leq$$
$$\leq \tau_n\, \mathbb{P}\left(\left|\hat{f}_N(\theta,m_y,x) - \mathbb{E}\hat{f}_N(\theta,m_y,x)\right| > \varepsilon\right).$$

Applying Berstain's exponential inequality to

$$\Pi_i = \frac{1}{g_n \mathbb{E}\big(K_1(\theta,x)\big)}\left[\delta_i K_i(\theta,x) H_i(m_y) - \mathbb{E}\left(\delta_i K_i(\theta,x) H_i(m_y)\right)\right].$$

Firstly, it follows from the fact that kernels K and H are bounded that we get

$$\mathbb{P}\left(\left|\hat{f}_N(\theta,m_y,x) - \mathbb{E}\hat{f}_N(\theta,m_y,x)\right| > \varepsilon\right) \leq \mathbb{P}\left(\frac{1}{n}\left|\sum_{i=1}^{n}\Pi_i\right| > \varepsilon\right) \leq 2n^{-C\varepsilon_0^2}.$$

Finally, by choosing ε_0 large enough, the proof can be concluded by the use of the Borel-Cantelli lemma, and the result can be easily deduced.

LEMMA 3: Under hypotheses (H1) and (H4)–(H5), we have, as $n \to \infty$,

(i) $\sup_{y\in S_{\mathbb{R}}}\left|\hat{f}_{\mathrm{D}}(\theta,x) - \mathbb{E}\hat{f}_{\mathrm{D}}(\theta,x)\right| = \mathcal{O}_{a.co.}\left(\sqrt{\frac{\log n}{n\phi_{\theta,x}(h_n)}}\right)$.

(ii) $\sum_{n\geq 1}\mathbb{P}\left(\hat{f}_{\mathrm{D}}(\theta,x) < 1/2\right) < \infty$.

PROOF: For the demonstration of the first part of this lemma, we use the same arguments as the previous lemma, the only change is in $\Delta_i(\theta,x)$, where

$$\hat{f}_{\mathrm{D}}(\theta,x) - \mathbb{E}\hat{f}_{\mathrm{D}}(\theta,x) = \frac{1}{n\mathbb{E}\left(K_1(\theta,x)\right)}\sum_{i=1}^{n}\Delta_i(\theta,x),$$

with $\Delta_i(\theta,x) = \delta_i K_i(\theta,x) - \mathbb{E}\delta_i K_i(\theta,x)$.

All the calculus previously made with the variables $\Pi_i(\theta,x)$ remains valid for the variables $\Delta_i(\theta,x)$, and we obtain

$$\mathbb{P}\left(\left|\hat{f}_{\mathrm{D}}(\theta,x) - \mathbb{E}\hat{f}_{\mathrm{D}}(\theta,x)\right| > \varepsilon\sqrt{\frac{\log n}{n\phi_{\theta,x}(h_n)}}\right) \leq 2n^{-C'\varepsilon^2} < \infty.$$

For the proof of the second part of this lemma, we only need to establish $\mathbb{E}\hat{f}_{\mathrm{D}}(\theta,x) \underset{n\to\infty}{\longrightarrow} p(x,\theta)\ a.co.$

By the properties of the conditional expectation and the mechanism of MAR and (H5), it follows that

$$\mathbb{E}\hat{f}_{\mathrm{D}}(\theta,x) = \frac{1}{n\mathbb{E}\big(K_1(\theta,x)\big)}\sum_{i=1}^{n}\mathbb{E}\big(\delta_i K_i(\theta,x)\big)$$
$$= \frac{1}{n\mathbb{E}\big(K_1(\theta,x)\big)}\sum_{i=1}^{n}\mathbb{E}[\mathbb{E}(\delta_i| <\ \theta, X_i >)K_i(\theta,x)]$$
$$= \frac{\big(p(x,\theta)+o(1)\big)}{n\mathbb{E}\big(K_1(\theta,x)\big)}\sum_{i=1}^{n}\mathbb{E}\big(K_i(\theta,x)\big) \underset{n\to\infty}{\longrightarrow} p(x,\theta)\ a.co$$

Therefore (ii) of Lemma 3 follows from (i), and because $\hat{f}_{\mathrm{D}}(\theta,x) \underset{n\to\infty}{\longrightarrow} p(x,\theta)\ a.co.$

Concerning the last part, we have

$$\big\{\hat{f}_{\mathrm{D}}(\theta,x) < p(x,\theta)/2\big\} \subseteq \big\{\big|\hat{f}_{\mathrm{D}}(\theta,x) - p(x,\theta)\big| > p(x,\theta)/2\big\} \Rightarrow$$
$$\Rightarrow \mathbb{P}\big\{\hat{f}_{\mathrm{D}}(\theta,x) < p(x,\theta)/2\big\} \le \mathbb{P}\big\{\big|\hat{f}_{\mathrm{D}}(\theta,x) - p(x,\theta)\big| > p(x,\theta)/2\big\} \le$$
$$\le \mathbb{P}\big\{\big|\hat{f}_{\mathrm{D}}(\theta,x) - \mathbb{E}\hat{f}_{\mathrm{D}}(\theta,x)\big| > 1/2\big\},$$

and because $\lim_{n\to\infty}\hat{f}_{\mathrm{D}}(\theta,x) = p(x,\theta)$, we show that

$$\sum_{n\ge 1}\mathbb{P}\big(\hat{f}_{\mathrm{D}}(\theta,x) < p(x,\theta)/2\big) < \infty.$$

We conclude the proof of the Theorem 1 by making use of Inequality (1), in conjunction with Lemma 1, Lemma 2 and Lemma 3.

4.2. Conditional mode estimation

In this section, we will study the rate of convergence of our conditional mode estimator $\widehat{M}_\theta(x)$. Obviously, obtaining these results will require more sophisticated technical developments than those presented so far. To ensure a good readability of this paragraph, we introduce conditions related to the flatness of the *cond-dff* $(\theta,.,x)$ around the conditional quantile $M_\theta(x)$.

Then a natural estimator of the conditional mode $M_\theta(x)$ is defined as:

$$\widehat{M}_\theta(x) = \arg \sup_{y \in \mathcal{S}_{\mathbb{R}}} \hat{f}(\theta, y, x),$$

where $M_\theta(x) = \arg \sup_{y \in \mathcal{S}_{\mathbb{R}}} f(\theta, y, x)$, $\mathcal{S}_{\mathbb{R}}$ is a fixed compact subset of $\mathbb{R}$.

But a complementary way to take this local shape constraint into account is to suppose that:

(H6) The conditional density $f(\theta,.,x)$ satisfies

(i) $\exists \epsilon_0$, such that $f(\theta,.,x)$ is strictly increasing on $\left(M_\theta(x) - \epsilon_0, M_\theta(x)\right)$ and strictly decreasing on $(M_\theta(x), M_\theta(x) + \epsilon_0)$, with respect to x.

(ii) $f(\theta, y, x)$ is twice continuously differentiable around the point $M_\theta(x)$ with $f^{(1)}(\theta, M\theta(x), x) = 0$ and $f^{(2)}(\theta,.,x)$ is uniformly continuous on $S_{\mathbb{R}}$, such that $f^{(2)}(\theta, M\theta(x), x) \neq 0$, where $f^{(j)}(\theta,.,x)$ (j = 1,2) is the j-th order derivative of the conditional density $f(\theta, y, x)$.

(H7) $\forall \varepsilon > 0, \exists \eta > 0, \forall \varphi |M_\theta(x) - \varphi(x)| \geq \varepsilon \Rightarrow |f(\theta, \varphi(x), x) - f(\theta, M_\theta(x), x)| \geq \eta$.

The difficulty of the problem is naturally linked to the flatness of function $f(\theta, y, x)$ around mode M_θ. This flatness can be controlled by the number of vanishing derivatives at point M_θ, and this parameter will also have a significant influence on the asymptotic rates of our estimates. More precisely, we introduce the following additional smoothness condition.

(H8) There exists some integer $j > 1$, such that $\forall\, x$ and the function $f(\theta,.,x)$ is j-times continuously differentiable w.r.t y on $S_{\mathbb{R}}$ with

$$\begin{cases} f^{(j)}(\theta, M_\theta(x), x) = 0, if; 1 \leq j < l \\ f^{(j)}(\theta,.,x) \text{ is uniformly continuous on } S_{\mathbb{R}} \\ \text{such that } f^{(j)}(\theta, M_\theta(x), x) \neq 0. \end{cases}$$

PROPOSTION 1: Suppose that the hypotheses (H1), (H3)–(H8) are satisfied if $\exists \beta > 0, n^{\beta} g_n \xrightarrow[n\to\infty]{} \infty$, and if

$$\lim_{n\to\infty} \frac{\log n}{n g_n^{2l} \phi_{\theta,x}(h_n)} = 0,$$

then we have

$$\left|\widehat{M}_\theta(\gamma, x) - M_\theta(\gamma, x)\right| = O\left(\left(h_n^{\alpha_1} + g_n^{\alpha_2}\right)^{\frac{1}{j}}\right) + O_{a.co.}\left(\left(\frac{\log n}{n g_n^2 \phi_{\theta,x}(h_n)}\right)^{\frac{1}{2j}}\right).$$

PROOF: The proof is based on the Taylor expansion of $f(\theta,.,x)$. In the neighborhood of $M_\theta(\gamma,x)$, we get

$$\hat{f}\left(\theta,\widehat{M}_\theta(x),x\right) = f(\theta,M_\theta(x),x) + \frac{f^{(j)}(\theta,M_\theta^*(x),x)}{j!}\left(\widehat{M}_\theta(x) - M_\theta(x)\right)^j,$$

where $M_\theta^*(x)$ is between $M_\theta(x)$ and $\widehat{M}_\theta(x)$, combining the last equality with the fact that

$$\left|\hat{f}\left(\theta,\widehat{M}_\theta(x),x\right) - f(\theta,M_\theta(x),x)\right| \leq 2 \sup_{y\in S_{\mathbb{R}}}\left|\hat{f}(\theta,y,x) - f(\theta,y,x)\right|,$$

which makes it possible to write:

$$\left|M_\theta(x) - \widehat{M}_\theta(x)\right|^j \leq \frac{j!}{f^{(j)}\left(\theta,M_\theta^*(x),x\right)} \sup_{y\in S_{\mathbb{R}}}\left|\hat{f}(\theta,y,x) - f(\theta,y,x)\right|.$$

Using the second part of (H8), we obtain

$$\exists\delta > 0, \sum_{n\geq 1} \mathbb{P}\left(f^{(j)}(\theta,M_\theta^*(x),x) \geq \delta\right) < \infty.$$

So, we have

$$\left|\widehat{M}_\theta(\gamma,x) - M_\theta(\gamma,x)\right|^j = O_{a.co.}\left(\sup_{y\in S_{\mathbb{R}}}\left|\hat{f}(\theta,y,x) - f(\theta,y,x)\right|\right).$$

Finally, Proposition 1 can be deduced from Theorem 1.

COROLLARY 1: Under hypotheses of Theorem 1, we have

$$\widehat{M}_\theta(x) - M_\theta(x) \underset{n\to\infty}{\longrightarrow} 0, a.co.$$

PROOF: The proof is based on the point-wise convergence of $\widehat{f}(\theta,.,x)$, and the Lipschitz property introduced in (H3)-(i) and hypothesis (H7), $f(\theta,t,x)$ is a continuous. We therefore have:

$\forall\, \epsilon > 0, \exists\eta(\epsilon) > 0$, such that

$$\left|f(\theta,y,x) - \hat{f}(\theta,M_\theta(x),x)\right| \leq \eta(\epsilon) \Rightarrow |y - M_\theta(x)| \leq \epsilon.$$

Therefore, for $y = \widehat{M}_\theta(x)$,

$$\mathbb{P}\left(\left|\widehat{M}_\theta(x) - M_\theta(x)\right| > \epsilon\right) \leq \mathbb{P}\left(\left| f\left(\theta, \widehat{M}_\theta(x), x\right) - f\left(\theta, M_\theta(x), x\right)\right| \geq \eta(\epsilon)\right). \quad (2)$$

Then, according to the theorem, $\widehat{M}_\theta - M_\theta$ go almost completely to 0, as n goes to infinity.

5. Asymptotic normality

The asymptotic normality of the semi-parametric estimators of the conditional mode 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 might vary depending on the assumptions and estimation methods used, it allows us to construct confidence intervals and hypothesis tests for the estimated mode. In this Section, the asymptotic normality of the estimator $\hat{f}(\theta,.,x)$ in the single functional index model is established.

(N1) There exists a function $\beta_{\theta,x}(\cdot)$, such that $\lim_{n\to\infty} \frac{\phi_{\theta,x}(sh_n)}{\phi_{\theta,x}(h_n)} = \beta_{\theta,x}(s)$, for $\forall s \in [0,1]$.

(N2) The bandwidth $h_n$and g_n, small ball probability $\phi_{\theta,x}(h_n)$ satisfying

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

(ii) $ng_n^2\phi_{\theta,x}^3(h_n) \longrightarrow 0, asn \to \infty$.

(N3) The conditional density $f(\theta,y,x)$ satisfies: $\exists \alpha > 0, \forall (y_1, y_2) \in S_\mathbb{R} \times S_\mathbb{R}$,

$$\left|f^{(j)}(\theta, y_1, x_1) - f^{(j)}(\theta, y_2, x_2)\right| \leq C(|y_1 - y_2|^\alpha), j = 1{,}2.$$

THEOREM 2: Under the assumptions of Theorem 1 and (N1)–(N3) for all x $\in \mathcal{H}$, and if

$$\sqrt{ng_n\phi_{\theta,x}(h_n)}\left(h_n^{\alpha_1} + g_n^{\alpha_2}\right) \underset{n\to\infty}{\longrightarrow} 0,$$

then we have

$$\sqrt{\frac{ng_n\phi_{\theta,x}(h_n)}{\sigma^2(\theta,y,x)}}\left(\hat{f}(\theta, y, x) - f(\theta, y, x)\right) \xrightarrow{\mathcal{D}} \mathcal{N}(0{,}1),$$

where $\sigma^2(\theta, y, x) = \frac{M_2(\theta,x)}{(M_1(\theta,x))^2} \frac{f(\theta,y,x)}{p(\theta,x)} \int H^2(u)du$ with $M_l(\theta, x) = K^l(1) - \int_0^1 (K^l)'(u)\, \beta_{\theta,x}(u)du$, $l = 1, 2$.

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

$$\begin{aligned}\hat{f}(\theta,y,x)-f(\theta,y,x) &= \frac{\hat{f}_N(\theta,y,x)}{\hat{f}_D(\theta,x)}-\frac{M_1(\theta,x)f(\theta,y,x)}{M_1(\theta,x)}= \\ &= \frac{1}{\hat{f}_D(\theta,x)}\Big(\hat{f}_N(\theta,y,x)-\mathbb{E}\hat{f}_N(\theta,y,x)\Big)+ \\ &-\frac{1}{\hat{f}_D(\theta,x)}\Big(M_1(\theta,x)f(\theta,y,x)-\mathbb{E}\hat{f}_N(\theta,y,x)\Big)+ \\ &+\frac{f(\theta,y,x)}{\hat{f}_D(\theta,x)}\Big(M_1(\theta,x)-\mathbb{E}\hat{F}_D(\theta,x)\Big)-\frac{f(\theta,y,x)}{\hat{f}_D(\theta,x)}\Big(\hat{F}_D(\theta,x)-\mathbb{E}\hat{F}_D(\theta,x)\Big)= \\ &= \frac{1}{\hat{f}_D(\theta,x)}A_n(\theta,y,x)+B_n(\theta,y,x),\end{aligned}$$

where:

$$\begin{aligned}A_n(\theta,y,x) &= \frac{1}{ng_n\mathbb{E}(K_1(\theta,x))}\sum_{i=1}^{n}\{\big(H_i(t)-g_nf(\theta,y,x)\big)\delta_iK_i(\theta,x) \\ &-\mathbb{E}\big[\big(H_i(t)-g_nf(\theta,y,x)\big)\delta_iK_i(\theta,x)\big]\}=\frac{1}{ng_n\mathbb{E}(K_1(\theta,x))}\sum_{i=1}^{n}N_i\,(\theta,t,x)\end{aligned}$$

and

$$B_n(\theta,y,x)=M_1(\theta,x)f(\theta,y,x)-\ \mathbb{E}\hat{f}_N(\theta,y,x)+f(\theta,y,x)\Big(M_1(\theta,x)-\mathbb{E}\hat{f}_D(\theta,x)\Big).$$

It follows that,

$$\begin{aligned}ng_n\phi_{\theta,x}(h_n)Var\big(A_n(\theta,y,x)\big) &= \frac{\phi_{\theta,x}(h_n)}{g_n\mathbb{E}^2\big(K_1(\theta,x)\big)}Var\big(N_1(\theta,y,x)\big)= \\ &= V_n(\theta,y,x).\end{aligned}$$

Then, the proof of Theorem 2 can be deduced from the following Lemmas.

LEMMA 4: Under assumptions of Theorem 2, we have

$$\sqrt{ng_n\phi_{\theta,x}(h_n)}A_n(\theta,y,x)\xrightarrow{\mathcal{D}}\mathcal{N}\big(0,\sigma^2(\theta,y,x)\big).$$

PROOF:

$$V_n(\theta,y,x)=\frac{\phi_{\theta,x}(h_n)}{g_n\mathbb{E}^2\big(K_1(\theta,x)\big)}\mathbb{E}\left[\delta_1K_1^2(\theta,x)\big(H_1(y)-g_nf(\theta,y,x)\big)^2\right]$$

$$V_n(\theta,y,x) = \\ = \frac{\phi_{\theta,x}(h_n)}{g_n\mathbb{E}^2\big(K_1(\theta,x)\big)}\mathbb{E}\left[K_1^2(\theta,x)\mathbb{E}\left(\delta_1\left(H_1(y)-g_nf(\theta,y,x)\right)^2\Big|\langle\theta,X_1\rangle\right)\right] \quad (3)$$

Using the definition of conditional variance, we have

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

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

Concerning J_{1n},

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

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

$$\mathcal{J}_1 = \mathbb{E}\left(\delta_1H^2\left(\tfrac{y-Y_1}{g_n}\right)\Big|\langle\theta,X_1\rangle\right) = p(x,\theta)\int H^2\left(\tfrac{y-v}{g_n}\right)f(\theta,v,X_1)dv \\ = p(x,\theta)\int H^2(u)\,dF(\theta,y-ug_n,X_1).$$

On the other hand, under assumptions (H2)–(H3), we have

$$\mathcal{J}_1 = \int H^2(u)\,dF(\theta,y-ug_n,X_1) = h_n\int H^2(u)\,f(\theta,y-ug_n,X_1)du \leq \\ \leq g_n\int H^2(u)\big(f(\theta,y-ug_n,X_1)-f(\theta,y,x)\big)du + \\ +g_n\int H^2(u)\,f(\theta,y,x)du \leq \\ \leq g_n\left(C_{\theta,x}\int H^2(u)\big(h_n^{\alpha_1}+|v|^{\alpha_2}g_n^{\alpha_2}\big)du + f(\theta,y,x)\int H^2(u)\,du\right) = \\ = \mathcal{O}\big(h_n^{\alpha_1}+g_n^{\alpha_2}\big) + g_nf(\theta,y,x)\int H^2(u)\,du. \quad (4)$$

As for J_2, $\mathcal{J}_2' = \mathbb{E}(\delta_1H_1(y)|\langle\theta,X_1\rangle) = p(x,\theta)\int H\left(\frac{y-v}{g_n}\right)f(\theta,y,X_1)dv.$

Moreover, by changing variables, we obtain:

$$\mathcal{J}'_2 = h_n \int H(u)\big(f(\theta, y - ug_n, x) - f(\theta, y, x)\big)du + g_n f(\theta, y, x) \int H(u)du.$$

The last equality is due to the fact that H is a probability density, thus we have

$$\mathcal{J}'_2 = O\big(h_n^{\alpha_1} + g_n^{\alpha_2}\big) + g_n f(\theta, y, x).$$

Finally, we get $J_2 \xrightarrow[n\to\infty]{} \infty$. As for J_{2n}, by (H1)–(H3), we obtain $J_{2n} \xrightarrow[n\to\infty]{} \infty$. Meanwile, from (H1)–(H3), it follows that

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

which leads to combining equations (3) and (4):

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

LEMMA 5: If the assumptions (H1)–(H7) are satisfied, we have

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

PROOF: We have

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

Firstly, it can be observed that, as $n \to \infty$,

$$\frac{1}{\phi_{\theta,x}(h_n)} \mathbb{E}\left[K^l\left(\frac{\langle \theta, x - X_i\rangle}{h_n}\right)\right] \to M_l(\theta, x), \text{for } l = 1,2, \tag{5}$$

$$\mathbb{E}\hat{f}_D(\theta, x) \to M_1(\theta, x) p(x, \theta) \textit{ and } \mathbb{E}\hat{f}_N(\theta, y, x) \to M_1(\theta, x) f(\theta, y, x), \tag{6}$$

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, using (5) and (6), we have, on the one hand, as $n \to \infty$,

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

On the other hand,

$$\frac{\sqrt{ng_n\phi_{\theta,x}(h_n)}}{\hat{f}_D(\theta,x)} = \frac{\sqrt{ng_n\phi_{\theta,x}(h_n)}\hat{f}(\theta,y,x)}{\hat{f}_D(\theta,x)\hat{f}(\theta,y,x)} = \frac{\sqrt{ng_n\phi_{\theta,x}(h_n)}\hat{f}(\theta,y,x)}{\hat{f}_N(\theta,y,x)}.$$

Because K and H are continuous with the support on [0,1], then from (H3) and (H4) $\exists m = \min_{[0,1]} K(t)H(t)$, it follows that

$$\hat{f}_N(\theta, y, x) \geq \frac{m}{g_n\phi_{\theta,x}(h_n)},$$

which yields

$$\frac{\sqrt{ng_n\phi_{\theta,x}(h_n)}}{\hat{f}_N(\theta,y,x)} \leq \frac{\sqrt{ng_n^3\phi_{\theta,x}^3(h_n)}}{m}.$$

Finally, (N2)–(i) completes the proof of Lemma 5.

5.1. Application: The conditional mode in functional single-index model

The main objective of this part of our work is to establish the asymptotic normality of the conditional mode estimator of Y, given $< \theta, X > = < \theta, x >$ denoted by $M_\theta(x)$.

COROLLARY 2: Under the assumptions of Theorem 2, and if (H6) holds true, and in addition if

$$ng_n^3\phi_{\theta,x}(h_n) \underset{n\to\infty}{\longrightarrow} 0,$$

then we have, as $n \to \infty$,

$$\sqrt{ng_n^3\phi_{\theta,x}(h_n)}\left(\widehat{M}_\theta(x) - M_\theta(x)\right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \varrho^2(\theta, M_\theta(x), x)\right),$$

where

$$\varrho^2(\theta, M_\theta(x), x) = \frac{M_2(\theta,x) f(\theta,M_\theta(x),x)}{p(\theta,x)\left(M_1(\theta,x) f^{(2)}(\theta,M_\theta(x),x)\right)^2}.$$

PROOF: By the first-order Taylor expansion for $\hat{f}^{(1)}(\theta, y, x)$ at point $M_\theta(x)$, and the fact that $\widehat{f}^{(1)}\big(\theta, \widehat{M}_\theta(x), x\big) = 0$, it follows that

$$\sqrt{n g_n^3 \phi_{\theta,x}(h_n)}\left(\widehat{M}_\theta(x) - M_\theta(x)\right) = -\sqrt{n g_n^3 \phi_{\theta,x}(h_n)}\frac{\hat{f}^{(1)}(\theta,M_\theta(x),x)}{\hat{f}^{(2)}(\theta,M_\theta^*(x),x)},$$

where $M_\theta^*(x)$ is between $\widehat{M}_\theta(x)$ and $M_\theta(x)$. Similarly to the proof of Theorem 2, it follows that

$$-\sqrt{n g_n^3 \phi_{\theta,x}(h_n)}\hat{f}^{(1)}(\theta, M_\theta(x), x) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \varrho_0^2(\theta, M_\theta(x), x)\right), \tag{7}$$

where

$$\varrho_0^2(\theta, M_\theta(x), x) = \frac{M_2(\theta,x)}{(M_1(\theta,x))^2}\frac{f(\theta,M_\theta(x),x)}{p(\theta,x)}\int\left(H'(u)\right)^2 du.$$

Thus, as above, similarly to Ferraty and Vieu (2006), we can obtain $\hat{f}^{(2)}(\theta, y, x) \xrightarrow{\mathbb{P}} f^{(2)}(\theta, y, x)$, as $n \to \infty$, which implies that $\widehat{M}_\theta(x) \to M_\theta(x)$. Therefore, we get

$$\hat{f}^{(2)}(\theta, M_\theta^*(x), x) \underset{n \to \infty}{\longrightarrow} f^{(2)}(\theta, M_\theta(x), x) \neq 0. \tag{8}$$

By (H3), (H6) and (N3), similarly to the proof of lemmas, Lemma 4 and Lemma 5, respectively, (7) follows directly. Then, the proof of Corollary 2 is completed.

5.2. Confidence bands

The asymptotic variances $\sigma^2(\theta, t, x)$ and $\varrho^2(\theta, M_\theta(x), x)$ in Theorem 2 and Corollary 2 depend on some unknown quantities including M_1, M_2, $\phi(u)$, $M_\theta(x)$, $p(\theta, x)$ and $f(\theta, M_\theta(x), x)$. Therefore, $p(\theta, x)$,$M_\theta(x)$, and $f(\theta, M_\theta(x), x)$ can be estimated by $P_n(\theta, x)$, $\widehat{M}_\theta(x)$ and $\hat{f}(\theta, M_\theta(x), x)$ and $\widehat{M}_\theta(x)$, respectively. Moreover, using the decomposition given by the assumption (H1), one can estimate $\phi_{\theta,x}(.)$ by $\hat{\phi}_{\theta,x}(.)$. Because the unknown functions $M_j := M_j(\theta, x)$ and $f(\theta, y, x)$ are intervening in the expression of the variance, we need to estimate the mode $M_1(\theta, x)$, $M_2(\theta, x)$ and $f(\theta, y, x)$, respectively.

From the assumptions (H1)–(H4), we know that $M_j(\theta, x)$ can be estimated by $\widehat{M}_j(\theta, x)$, which is defined as:

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

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:

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

Finally, in order to show the asymptotic $(1-\xi)$ confidence interval of $M_\theta(x)$, we need to consider the estimator of $\varrho^2(\theta,M_\theta(x),x)$, as follows:

$$\widehat{\varrho}^2(\theta,M_\theta(x),x)=\frac{\widehat{M}_2(\theta,x)}{(\widehat{M}_1(\theta,x))^2}\frac{\widehat{f}(\theta,\widehat{M}_\theta(x),x)}{P_n(\theta,x)\left(\widehat{f}^{(2)}(\theta,\widehat{M}_\theta(x),x)\right)^2}\int\left(H'(u)\right)^2du,$$

so we can derive the corollary below.

COROLLARY 3: Under the assumptions of Theorem 2, K' and $(\mathrm{K}^2)'$ are integrable functions, then we get, as $\mathrm{n}\to\infty$,

(a) $\sqrt{\frac{ng_n\widehat{\phi}_{\theta,x}(h_n)}{\widehat{\sigma}^2(\theta,y,x)}}\left(\widehat{f}(\theta,y,x)-f(\theta,y,x)\right)\xrightarrow{\mathcal{D}}\mathcal{N}(0,1).$

(b) $\sqrt{\frac{ng_n^3\widehat{\phi}_{\theta,x}(h_n)}{\widehat{\varrho}^2(\theta,M_\theta(x),x)}}\left(\widehat{M}_\theta(x)-M_\theta(x)\right)\xrightarrow{\mathcal{D}}\mathcal{N}(0,1).$

PROOF: Observe that

$$\text{(a) } \Sigma(\theta,y,x)=\frac{\widehat{M}_1}{M_1}\sqrt{\frac{M_2}{\widehat{M}_2}}\sqrt{\frac{ng_n\widehat{\phi}_{\theta,x}(h_n)P_n(\theta,x)f(\theta,y,x)}{p(\theta,x)\widehat{f}(\theta,y,x)ng_n\phi_{\theta,x}(h_n)}}\times$$
$$\times\frac{M_1}{\sqrt{M_2}}\sqrt{\frac{ng_n\phi_{\theta,x}(h_n)}{\sigma^2(\theta,y,x)}}\left(\widehat{f}(\theta,y,x)-f(\theta,y,x)\right),$$

where $\Sigma(\theta,y,x)=\sqrt{\frac{ng_n\widehat{\phi}_{\theta,x}(h_n)}{\widehat{\sigma}^2(\theta,y,x)}}\left(\widehat{f}(\theta,y,x)-f(\theta,y,x)\right)$, by Theorem 2, we have, as $n\to\infty$,

$$\sqrt{\frac{ng_n\phi_{\theta,x}(h_n)}{\sigma^2(\theta,y,x)}}\left(\hat{f}(\theta,y,x)-f(\theta,y,x)\right)\xrightarrow{\mathcal{D}}\mathcal{N}(0,1).$$

In order to prove (a), we need to show that

$$\frac{\widehat{M}_1}{M_1}\sqrt{\frac{M_2}{\widehat{M}_2}}\sqrt{\frac{ng_n\widehat{\phi}_{\theta,x}(h_n)P_n(\theta,x)f(\theta,y,x)}{p(\theta,x)\hat{f}(\theta,y,x)ng_n\phi_{\theta,x}(h_n)}}\left(\hat{f}(\theta,y,x)-f(\theta,y,x)\right)\xrightarrow{\mathbb{P}}\mathcal{N}(0,1),$$

using the result given by Laib and Louani (2010), we have

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

On the other hand, from Proposition 2 in Laib and Louani (2010), it follows that

$$P_n(\theta,x)\underset{n\to\infty}{\longrightarrow}\mathbb{E}(\delta|\langle X,\theta\rangle=\langle x,\theta\rangle)=\mathbb{P}(\delta=1|\langle X,\theta\rangle=\langle x,\theta\rangle)=p(x,\theta).$$

In addition, from Theorem 1, we have $\hat{f}(\theta,y,x)\longrightarrow f(\theta,y,x)$, as $n\to\infty$. This yields the proof for the first part of Corollary 3.

(b) $$\frac{\widehat{M}_1\hat{f}^{(2)}(\theta,\widehat{M}_\theta(x),x)}{\sqrt{\widehat{M}_2}}\sqrt{\frac{ng_n^3\widehat{\phi}_{\theta,x}(h_n)P_n(\theta,x)}{\hat{f}(\theta,\widehat{M}_\theta(x),x)}}\left(\widehat{M}_\theta(x)-M_\theta(x)\right)=$$
$$=\frac{\widehat{M}_1\sqrt{M_2}}{M_1\sqrt{\widehat{M}_2}}\sqrt{\frac{ng_n^3\widehat{\phi}_{\theta,x}(h_n)P_n(\theta,x)f(\theta,M_\theta(x),x)}{ng_n^3\phi_{\theta,x}(h_n)p(\theta,x)}}\frac{\hat{f}^{(2)}(\theta,\widehat{M}_\theta(x),x)}{f^{(2)}(\theta,M_\theta(x),x)}\times$$
$$\times\frac{M_1}{\sqrt{M_2}}\sqrt{\frac{ng_n^3\phi_{\theta,x}(h_n)p(\theta,x)}{f(\theta,M_\theta(x),x)}}f^{(2)}(\theta,M_\theta(x),x)\left(\widehat{M}_\theta(x)-M_\theta(x)\right).$$

Applying Corollary 2, we obtain

$$\frac{M_1}{\sqrt{M_2}}\sqrt{\frac{ng_n^3\phi_{\theta,x}(h_n)p(\theta,x)}{f(\theta,M_\theta(x),x)}}f^{(2)}(\theta,M_\theta(x),x)\left(\widehat{M}_\theta(x)-M_\theta(x)\right)\longrightarrow\mathcal{N}(0,1).$$

Further, by considering Lemma 5, (2) and (8), we obtain, as $n\to\infty$,

$$\frac{\widehat{M}_1\sqrt{M_2}}{M_1\sqrt{\widehat{M}_2}}\sqrt{\frac{ng_n^3\widehat{\phi}_{\theta,x}(h_n)P_n(\theta,x)f(\theta,M_\theta(x),x)}{ng_n^3\phi_{\theta,x}(h_n)p(\theta,x)}}\frac{\hat{f}^{(2)}(\theta,\widehat{M}_\theta(x),x)}{f^{(2)}(\theta,M_\theta(x),x)}\xrightarrow{\mathbb{P}}1.$$

Hence, the proof is completed.

REMARK 1: Thus, following Corollary 3, the asymptotic $(1-\xi)$ confidence interval of the conditional density $f(\theta,y,x)$, and the conditional mode $M_\theta(x)$, respectively, are expressed as follows:

$$\hat{f}(\theta,y,x) \pm \eta_{\gamma/2}\sqrt{\frac{\widehat{\sigma}^2(\theta,y,x)}{ng_n\widehat{\Phi}_{\theta,x}(h_n)}} \text{and} \widehat{M}_\theta(x) \pm \eta_{\gamma/2}\sqrt{\frac{\widehat{\varrho}^2(\theta,M_\theta(x),x)}{ng_n^3\widehat{\Phi}_{\theta,x}(h_n)}},$$

where $\widehat{\sigma}^2(\theta,y,x)$ and $\hat{\varrho}^2(\theta,M_\theta(x),x)$ are defined in Section 5.2, and $\eta_{\gamma/2}$ is the upper $\gamma/2$ quantile of the normal distribution $\mathcal{N}(0,1)$.

6. Simulation study on finite samples

6.1. A numerical study

In this Section, we will consider simulated data studied to assess the finite sample performance of the proposed estimator and compare it to the competing estimator. For studying the behavior of our estimators, and in order to illustrate our results, we evaluate the performance of our estimation approach using a single-index dimensional reduce model in order to prove the effectiveness of our model. More precisely, we will compare the finite sample behavior of estimator $\widetilde{f}$ with the complete functional data and the estimator $\hat{f}$ under functional data with MAR.

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

$$H(u)=\int_{-\infty}^{u}\frac{3}{4}(1-z^2)\mathbf{1}_{[-1;1]}(z)dz.$$

The semi-metric $d(.,.)$ will be specified according to the choice of the functional space $\mathcal{H}$ discussed in the scenarios below. It is well known that some of the crucial H parameters in semi-parametric models are 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 [12], p. 102).

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

$$Y_i = R(\langle\theta, X_i\rangle)+\epsilon_i; \quad i=1,\dots,n=300,$$

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

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

with $(\eta_i)_i$ is 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:

$$X(t) = \Omega(2 - \cos(\pi\Upsilon t)) + (1 - \Omega)\cos(\pi\Upsilon t), \quad t \in [0,1],$$

where Υ is generated from a standard normal distribution, and Ω is Bernoulli's law $\mathfrak{B}(0.5)$.

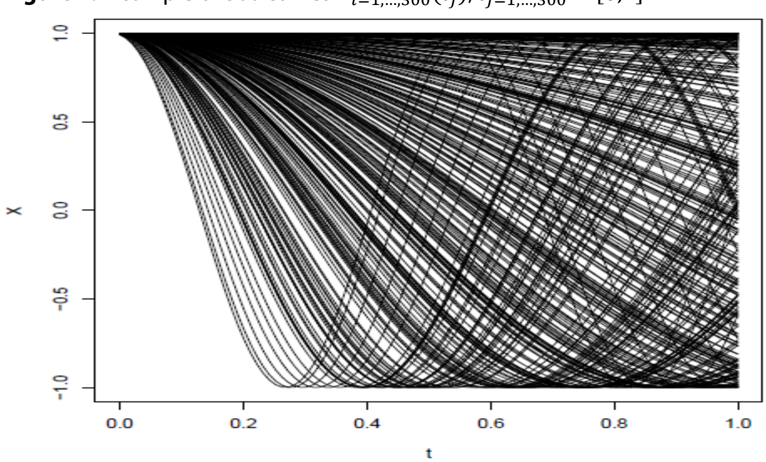
Let us take into account the smoothness of curves $X_i(t)$ (see Figure 1).We choose the distance $deriv_1$ (the semi-metric based on the first derivatives of the curves) in $\mathcal{H}$ as:

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

as semi-metric. Then, we consider a non-linear regression function defined as:

$$R(X) = 4exp\left(\frac{1}{2 + \int_0^{\pi/2} X^2(t)dt}\right).$$

Figure 1. A sample of 300 curves $X_{i=1,\ldots,300}(t_j)$, $t_{j=1,\ldots,300} \in [0,1]$



Source: authors' work.

The missing mechanism is similar to that in Ferraty et al. (2013):

$$p(x) = \mathbb{P}(\delta = 1|X = x) = exp\,it\left(2\alpha \int_0^1 x^2(t)dt\right),$$

where $exp\,it(v) = e^v/(1 + e^v)$ for $v \in \mathbb{R}$, and the degree of dependency between the functional covariate X and the missing variable δ is controlled by parameter α (where we compare three missing rates: strong, medium and weak cases, with α=0.05, $\alpha = 0.5$ and $\alpha = 3$ respectively). The missing proportions are quantified by the following benchmark:

$$\bar{\delta} = 1 - \frac{1}{n}\sum_{i=1}^{n} \delta_i.$$

In practice, this parameter can be selected by a cross-validation approach (see [2]). It might be that one selected the real-valued function $\theta(t)$ from among the eigenfunctions of the covariance operator $\mathbb{E}[(X' - \mathbb{E}X')\langle X', . \rangle_{\mathcal{H}}]$, where $X(t)$ is a diffusion process defined on a real interval $[a, b]$, and $X'(t)$ is its first derivative (see [3]). 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 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 our simulation, the sample size is $n = 300$. We divide it into two parts: one is a learning sample of 250 observations, and the other 50 observations are a test sample.

The first one from 1 to 250 will be used to make the simulation, and the second, from 251 to 300, will serve for the prediction. We then perform the following steps:

- **Step 1:** Compute the inner product: $\langle\theta^*, X_1\rangle, \ldots, \langle\theta^*, X_{300}\rangle$, generate independently variables $\epsilon_1, \ldots, \epsilon_{300}$, then simulate response variables $Y_i = r(\langle\theta^*, X_i\rangle) + \epsilon_i$, where $r(\langle\theta^*, X_i\rangle) = exp\big(10(\langle\theta^*, X_i\rangle - 0.05)\big)$, and generate independently variables $\epsilon_1, \ldots, \epsilon_{300}$.
- **Step 2:** For each j in, the test sample $\mathcal{J} = 251, \ldots, 300$, we compute: $\tilde{Y}_j = \widetilde{M}_{\theta^*}(X_j)$ and $\hat{Y}_j = \widehat{M}_{\theta^*}(X_j)$, where:

$$\widetilde{M}_\theta(\chi) = \arg \sup_{y\in\mathcal{S}_{\mathbb{R}}} \tilde{f}\,(\theta, y, \chi).$$

- **Step 3:** Finally, we show the results by juxtaposing the true values and the predicted values for the MSE, both for the option of having complete data and the option of having a MAR response.

$$CMSE = \frac{1}{50}\sum_{j=251}^{300}\left(Y_j - \tilde{Y}_j\right)^2 \quad and \quad MMSE = \frac{1}{50}\sum_{j=251}^{300}\left(Y_j - \hat{Y}_j\right)^2.$$

Figure 2. Complete data case

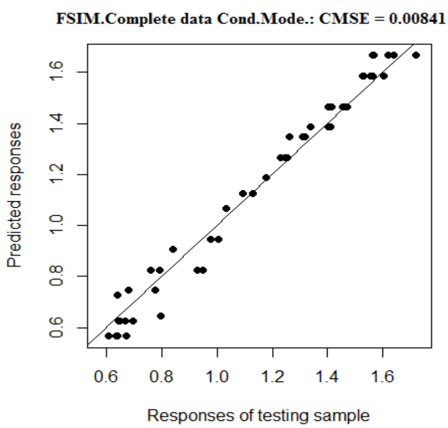
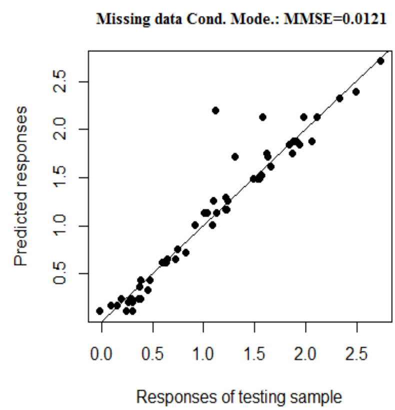


Figure 3. MAR with $\alpha = 2$



Source: authors' work.

Figure 4. MAR with $\alpha = 1.5$

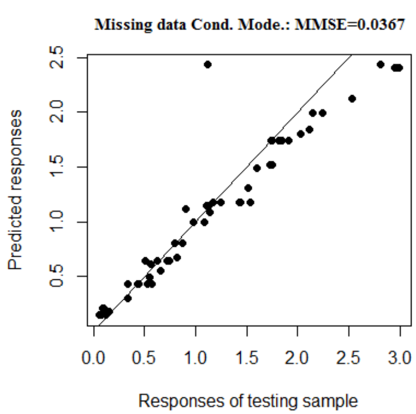
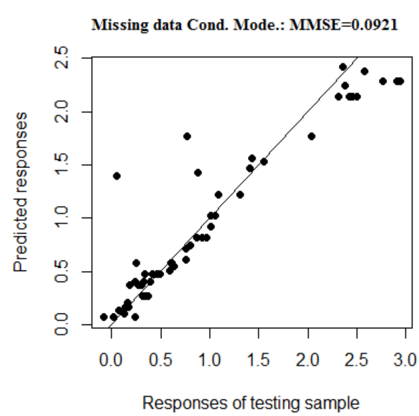


Figure 5. MAR with $\alpha = 0.5$



Source: authors' work.

In the MAR responses, the proportion $\bar{\delta}$ of missing response data is the key parameter which is shown by α. In Figures 2–5, we can see that when $\bar{\delta}$ is small (or

α is big), the $\widehat{M}_\theta(x)$ of the MAR response works almost as efficiently as if we had a complete data set and used $\widetilde{M}_\theta(x)$. In this case, when one has missing response data, the estimator $\widetilde{M}_\theta(x)$ is not useful, but the $\widehat{M}_\theta(x)$ is a benchmark analysis, and the fact that $\widehat{M}_\theta(x)$ is almost as effective as $\widetilde{M}_\theta(x)$ is what one is really expecting.

Acknowledgement

The authors are grateful to anonymous referees and the Editor-in-Chief for their valuable comments and suggestions towards the improvement of this paper.

References

Aït-Saidi, A., Ferraty, F., Kassa, R., & Vieu, P. (2008). Cross-validated estimation in the single-functional index model. *Statistics. A Journal of Theoretical and Applied Statistics*, *42*(6), 475–494. https://doi.org/10.1080/02331880801980377.

Attaoui, S.(2014). On the non-parametric conditional density and mode estimates in the single functional index model with strongly mixing data. *Sankhyã. The Indian Journal of Statistics*, *76*(2), 356–378. https://doi.org/10.1007/s13171-014-0051-6.

Attaoui, S., & Ling, N. (2016). Asymptotic results of a non-parametric conditional cumulative distribution estimator in the single functional index modeling for time series data with applications. *Metrika*. *International Journal for Theoretical and Applied Statistics*, *79*(5), 485–511. https://doi.org/10.1007/s00184-015-0564-6.

Chaudhuri, P., Doksum, K., & Samarov, A. (1997). On average derivative quantile regression. *The Annals of Statistics*, *25*(2), 715–744. https://doi.org/10.1214/aos/1031833670.

Cheng, P. E. (1994). Non-parametric Estimation of Mean Functional with Data Missing at Random. *Journal of the American Statistical Association*, *89*(425), 81–87. https://doi.org/10.2307/2291203.

Ezzahrioui, M., & Ould-Saïd, M. (2008). Asymptotic normality of a non-parametric estimator of the conditional mode function for functional data. *Journal of Non-parametric Statistics*, *20*(1), 3–18. https://doi.org/10.1080/10485250701541454.

Ferraty, F., Rabhi, A., & Vieu, P. (2005). Conditional Quantiles for Dependent Functional Data with Application to the Climatic *El Niño* Phenomenon. *Sankhyã*. *The Indian Journal of Statistics*, *67*(2), 378–398.

Ferraty, F., Sued, F., & Vieu, P. (2013). Mean estimation with data missing at random for functional covariables. *Statistics. A Journal of Theoretical and Applied Statistics*, *47*(4), 688–706.

Ferraty, F., & Vieu, P. (2003). Functional Non-parametric Statistics. A Double Infinite Dimensional Framework. In M. G. Akritas & D. N. Politis (Eds.), *Recent Advances and Trends in Non-parametric Statistics* (pp. 61–78). Elsevier.

Ferraty, F., & Vieu, P. (2006). *Non-parametric Functional Data Analysis. Theory and Practice*. Springer. https://doi.org/10.1007/0-387-36620-2.

Hamri, M. M., Mekki, S. D., Rabhi, A., & Kadiri, N. (2022). Single functional index quantile regression for independent functional data under right-censoring. *Econometrics. Ekonometria*, *26*(1), 31–62. https://doi.org/10.15611/eada.2022.1.03.

Kadiri, N., Rabhi, A., & Bouchentouf, A. A. (2018). Strong uniform consistency rates of conditional quantile estimation in the single functional index model under random censorship. *Dependence Modeling*, *6*(1), 197–227. https://doi.org/10.1515/demo-2018-0013.

Laib, N., & Louani, D. (2010). Non-parametric Kernel Regression Estimation for Functional Stationary Ergodic Data: Asymptotic Properties. *Journal of Multivariate Analysis*, *101*(10), 2266–2281. https://doi.org/10.1016/j.jmva.2010.05.010.

Lemdani, M., Ould-Saïd, E., & Poulin, N. (2009). Asymptotic properties of a conditional quantile estimator with randomly truncated data. *Journalof Multivariate Analysis*, *100*(3), 546–559. https://doi.org/10.1016/j.jmva.2008.06.004.

Liang, H.-Y., & de Uña-Álvarez, J. (2010). Asymptotic normality for estimator of conditional mode under left-truncated and dependent observations. *Metrika. International Journal for Theoretical and Applied Statistics*, *72*(1), 1–19. https://doi.org/10.1007/s00184-009-0237-4.

Ling, N., Liang, L., & Vieu, P. (2015). Non-parametric regression estimation for functional stationary ergodic data with missing at random. *Journal of Statistical Planning and Inference*, *162*, 75–87. https://doi.org/10.1016/j.jspi.2015.02.001.

Ling, N., Liu, Y., & Vieu, P. (2016). Conditional mode estimation for functional stationary ergodic data with responses missing at random. *Statistics. A Journal of Theoretical and Applied Statistics*, *50*(5), 991–1013. https://doi.org/10.1080/02331888.2015.1122012.

Mekki, S. D., Kadiri, N., & 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. https://doi.org/10.6092/issn.1973-2201/10472.

Ould-Saïd, E., & Cai, Z. (2005). Strong uniform consistency of non-parametric estimation of the censored conditional mode function. *Journal of Non-parametric Statistics*, *17*(7), 797–806. https://doi.org/10.1080/10485250500038561.

Ould-Saïd, E., & Tatachak, A. (2011). A non-parametric conditional mode estimate under RLT model and strong mixing condition. *International Journal of Statistics&Economics*, *6*(S11), 76–92.

Ould-Saïd, E., & Yahia, D. (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. https://doi.org/10.1080/03610926.2010.489171.

Rabhi, A., Kadiri, N., & 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*, *16*(2), 844–866. https://digitalcommons.pvamu.edu/aam/vol16/iss2/4.

Roussas, G. G. (1969). Non-parametric estimation of the transition distribution function of a Markov Process. *The Annals of Mathematical Statistics*, *40*(4), 1386–1400. https://doi.org/10.1214/aoms/1177697510.

Samanta, M. (1989). Non-parametric estimation of conditional quantiles. *Statistics & Probability Letters*, *7*(5), 407–412. https://doi.org/10.1016/0167-7152(89)90095-3.