ECONOMETRICS. EKONOMETRIA
Advances in Applied Data Analysis
Year 2023, Vol. 27, No. 1
ISSN 2449-9994

The Estimating of the Conditional Density with Application to the Mode Function in Scalar-On-Function Regression Structure: Local Linear Approach with Missing at Random*

Wahiba Bouabsa
University Djillali LIABES of Sidi Bel Abbes, Sidi Bel Abbes, Algeria
e-mail: wahiba_bouab@yahoo.fr
ORCID: 0000-0002-8106-5038

Quote as: Bouabsa, W. (2023). The Estimating of the Conditional Density with Application to the Mode Function in Scalar-On-Function Regression Structure: Local Linear Approach with Missing at Random. *Econometrics. Ekonometria. Advances in Applied Data Analysis*, *27*(1).

DOI: 10.15611/eada.2023.1.02

JEL Classification: C13, C14, C15

Abstract: The aim of this research was to study a nonparametric estimator of the density and mode function of a scalar response variable given a functional variable, when the observations are i.i.d. This proposed estimator is given by combining Missing At Random (MAR) with the local linear approach. Finally, a comparison study based on simulated data is also provided to illustrate the finite sample performances and the usefulness of the local linear approach with MAR to the presence of even a small proportion of outliers in the data.

Keywords: functional data, local linear estimation, conditional mode function, functional non-parametric statistics.

1. Introduction

It is very well recognised that robust regression in statistics is an attractive research method. It is used to overcome some of the weaknesses of classical regression, namely when outliers contain heteroscedastic data.

The study of the connection between a random variable W and set of covariates Z is a common problem in statistics. In the literature, these variables are generally known as functional variables. Remember that the robustification method is an old statistical issue, investigated first by Hurber (1964) who studied an estimation of alocation parameter (see also (Collomb & Härdale, 1979; Laib & Ouled-Said 2000)) for some

* Received 12 July 2022, accepted 26 March 2023.

results containing the multivariate time series case under a mixing or an ergodic condition).

Note that many writers have tackled the functional local linear estimating (LLE) problem since the appearance of the monograph by Ferraty and Vieu (2006) in NFDA, which relates to some responses to question number five given there, local linear estimation technique has several advantages in the case of finite-dimensional data over the kernel method, such as bias reduction and the adaptation of edge effects. Ballo and Grané (2009) provided the first response to this question, demonstrating the L^2-consistency of an LLE of the Hilbertian regression operator, whereas Barrientos et al. (2010) presented a different solution that may be used to a more general functional regressor. In turn, Demongeot et al. (2013) developed the first results on the LLE of the conditional density using functional variables. In this, the (a.co.) almost--complete consistency of an LLE of the conditional density was proved (see, p. 230, definition A.3 in Ferraty and Vieu (2006)). In Rachdi, Laksaci, Demongeot et al. (2014), the leading term of the mean quadratic error of the LLE of the conditional density was then explicitly mentioned. Additionally, Zhou and Lin (2016) established the asymptotic normality of the LLE of the regression operator, and recently Almanjahie et al. (2022) treated the k Nearest Neighbour kNN LLE of the conditional density in a scalar-On-Function regression structure, for additional research on the LLE in NFDA (Nonparametric Functional Data Analysis), see, Chahad et al. (2017); Attouch et al. (2017).

Note that the LLE's significant advantages over the traditional kernel technique function as the primary inspiration for all of the cited works on functional LLE. Specifically, the kernel method's bias error is reduced using the LLE approach (see, for instance (Fan & Gijbels, 1996; Rachdi et al., 2014) for more discussions on advantages of this method).

The primary innovation in this study is the estimation of the conditional density mode using the LLE approach weighted by MAR procedure for the functional and i.i.d data, while many studies in this field use the kernel estimation approach to estimate nonparametric functional models.

There are many references on the estimation of the conditional density, and there has been an increasing interest in the study of functional variables, the prime results about the conditional mode estimation obtained by Ferraty and Vieu (2006), about the almost complete convergence of the kernel estimator, and in the both cases i.i.d. and strong mixing, Ezzahrioui and Ould-Said (2006) established the asymptotic normality of the kernel estimator of the conditional mode. Lately the nonparametric estimation of the conditional mode using kNN method for independent and dependent functional data was examined by Attouch and Bouabsa (2013) and Attouch, Bouabsa and Chiker el mozoaur (2018). One can also mention for example Amiri and Dabo-Niang (2016) who investigated the recursive version of the kernel estimator of the density function. for spatial data, they stated the asymptotic qualities of the built estimator. The author referred to Giraldo et al. (2018), for a thorough discussion of spatio-functional data analysis and its viability as an attractive method in spatial big data modeling. Taking the ergodicity structure into account, Chaouch and Laib (2017) demonstrated the uniform consistency of the nonparametric estimator of functional modal regression. Maillot and Chesneau (2018) established the asymptotic properties of the conditional density for continuous time processes valued in functional space for current results. They also used this to calculate the convergence of the conditional mode and conditional expectation. Kirkby et al. (2021) investigated an alternate estimate of the conditional density, and employed the Galerkin method was to create an effective cross-validation technique for bandwidth parameter choice.

This study's innovation lies in its combination of the LLE technique with MAR data to provide a new estimator of functional conditional density. The rest of the paper is organized as follows. Section 2, presents for functional data the local linear estimation of the conditional mode. The principal hypotheses and notations are given in Section 3, All the results and proof are determined in Section 4. Section 5 presents a simulation study to prove the effectiveness of this study. The conclusions and prospects for future research are provided in the last section.

2. The model

$(A_i, B_i), i = 1, .., n$ is n copies of random vector identically distributed as (A, B), A is valued in infinite dimensional semimetric vector space (F, d), and B is valued in $\mathbb{R}$, S in handy applications is a normed space that can be of infinite dimension, for example Banach or Hilbert space with norm $\|.\|$ that $d(a, a') = \| a - a' \|$.

Recall that G^a (resp$G^{a(j)}$) isconditional density (resp its j^{th} order derivative) of B. The article presents the almost complete convergence results and rates for nonparametric estimates of $G^{a(j)}$. Thus when $G^a = G^{a(0)}$, the convergence of conditional density estimate is deduced instantly from the general results concerning $G^{a(j)}$.

Moreover, density G^a is continuous with respect to Lebesgue's measure over $\mathbb{R}$. In the local linear modeling the estimation of conditional mode function $\widehat{G}^a(\vartheta(a))$ is based on the approximation of conditional density function $G^a(b)$ by a linear function. To expand the local linear ideas to the infinite dimensional framework the author used the technique of Bàillo and Grané (2009), (see Barrientos-Marin et al. (2010) for some examples and references). Function $G^a(.)$ as marked by Fan and Gijbels (1996) can be viewed as a nonparametric regression model with response variable $h_Q^{-j} Q^{(j)}$ $\left(h_Q^{-1}(. - B_i)\right)$ where Q is some cumulative distribution function and h_Q is a sequence of positive real numbers. This estimation is motivated by the fact that:

$$E\left[h_Q^{-j} Q^{(j)}\left(h_Q^{-1}(b - B_i)\right) \middle| A_i = a\right] \to G^{a(j)}(b) \; as \; h_Q \to 0 \; (j = 0; 1).$$

Here, let us adopted fast functional local modelling, i.e. conditional density function $\widehat{G}^a$ is estimated by $\hat{x}$ where the pair $(\hat{x}; \hat{y})$ is obtained by the optimisation rule:

$$(\hat{x},\hat{y}) = \arg\min_{(x,y)\in\mathbb{R}^2} \sum_{i=1}^{n} h_Q^{-1} Q\left(h_Q^{-1}(b - B_i)\right) - x - y\ell(A_i,a)^2 W(h_W^{-1}\wp(a,A_i)), \quad (2.1)$$

where, $\ell(.\,;.)$ and $\wp(.\,;.)$ are locating functions defined from $\mathcal{F}^2$ into ℝ, such that:

$$\forall\xi \in \mathcal{F}; \ell(\xi;\xi) = 0 \; and \; d(.\,;.) = |\wp(.\,;.)|$$

and where function W: kernel function, Q: distribution function (df) and $h = h_W := h_{W,n}$ and $h_Q = h_{Q,n}$ are suites of positive real numbers, as n goes to infinity, and goes to zero. Clearly, estimator $\hat{x}$, given by (3.1), can be explicitly written as follows

$$\hat{G}^a(b) = \frac{\sum_{1\le i,j\le n} \nu_{ij}(x) G^{(1)}(h_G^{-1}(b-B_i))}{h_G \sum_{1\le i,j\le n} \nu_{ij}(a)} \forall b \in \text{IR}, \quad (2.2)$$

where, $Q^{(1)}$ is the derivative of Q, with

$$\nu_{ij}(a) = \ell_i(\ell_i - \ell_j) W(h_W^{-1}\wp(a,A_i))$$

and $\ell_i = \ell(A_i,a)$ and convention $0/0 = 0$;

Suppose that $G^a(.)$ has an only mode, noted by $\vartheta(a)$ assumed uniquely defined in compact set $\mathcal{S}$ which is given by

$$G^a(\vartheta(a)) = \sup_{b\in\mathcal{S}} G^a(b). \quad (2.3)$$

$\vartheta(a)$ is a kernel estimator of the conditional mode which given as random variable $\hat{\vartheta}(a)$ that maximises kernel estimator $\hat{G}^a(.)$ of $G^a(.)$

$$\hat{G}^a(\hat{\vartheta}(a)) = \sup_{b\in\mathcal{S}} \hat{G}^a(b). \quad (2.4)$$

3. Principal hypotheses and notations

This section contains all the assumptions that are necessary in deriving the almost--complete convergence (a.co.) of the functional locally modelled estimator of $\hat{G}^a(\hat{\vartheta}(a))$. Then, (resp. b) denotes a fixed point in ($\mathcal{F}(resp.$ ℝ), $\mathcal{N}_a$ (resp. $\mathcal{N}_b$) denote a fixed neighbourhood of fixed point x (resp. of y) and $\chi_a(z_1;z_2) = P(z_2 < \sigma(A;a) < z_1)$. Thus, one assumes that this nonparametric model satisfies the following conditions:

- (A1) For any $r > 0$, $\chi_a(z) := \chi_a(-z;z) > 0$.
- (A2) Conditional density function G^a) is such that there exist some positive constants y_1 and y_2, $\forall(b_1;b_2) \in \mathcal{N}_b \times \mathcal{N}_b$ and $\forall(a_1;a_2) \in \mathcal{N}_a \times \mathcal{N}_a$

$$|G^{a_1(j)}(b_1) - G^{a_2(j)}(b_2)| \le C(\sigma(a_1,a_2)^{y_1} + |b_1 - b_2|^{y_2}),$$

where, C is a positive constant depending on a.

- (A3) Function $\ell(.,.)$ and $\wp(.,.)$ and are such that:

$$\forall r \in F, |\wp(a,r)| = d(a,r) and C_1|\wp(a,r)| \leq |\ell(a,r)| \leq C_2|\wp(a,r)|,$$

where, $C_1 > 0; C_2 > 0$.

- (A4) Kernel W is a positive, differentiable function which is supported within $(-1;1)$.
- (A5) Kernel Q is a differentiable function and $Q^{(1)}$ is a positive, bounded, Lipschitzian continuous function such that:

$$\int |t|^{y_2} Q^{(1)}(t)dt < \infty and \int Q^{(2)}(t)dt < \infty.$$

Bandwidth h_W satisfies that there exists positive integer n_0, such that, $\forall n > n_0$,

$$-\frac{1}{\chi_a(h_W)} \int_{-1}^{1} \chi(rh_W; h_W) \frac{d}{dr}(r^2 W(r))dr > C_3 > 0,$$

and

$$h_W \int_{B(a;h_W)} \ell(u;a)dP(u) = o\left(\int_{B(a;h_W)} \ell^2(u;x)dP(u)\right).$$

where, $B(a;r) = \{r \in \mathcal{F}/|\wp(r;a)| \leq r\}$ and $d\mathbb{P}$(a) is the cumulative distribution of A; also that $\lim_{n\to\infty} h_Q = 0$ width $\lim_{n\to\infty} n^\gamma h_Q = \infty$ forsome $\gamma > 0$ and $\frac{\log(n)}{nh_Q^{2j+1}\chi_x(h_W)} \to 0$.

- (A6) $\exists \eta > 0, G^a \nearrow$ on $(\vartheta - \eta, \vartheta)$ and $G^a \searrow$ on $(\vartheta, \vartheta + \eta)$.
- (A7) G^a is j-times continuously differentiable with respect to a on $(\vartheta - \eta, \vartheta + \eta)$,
- (A8) $\begin{cases} G^{a(l)}(\vartheta) = 0, if\ 1 \leq l < j, \\ \text{and} \left|G^{a(j)}(\vartheta)\right| > 0. \end{cases}$
- (A9) Operator $P(.)$ is continuous on N_a and such that $P(A) > 0$.

4. Remarks on the hypotheses

Condition (A1) is straight forward tailoring (in Ferraty et al., (2008)) of hypothesis (A1), when one substitutes semi-metric d(., .) by $\wp(.,.)$. Hypotheses (A2) typify the structural functional space of the model. As in Barrientos-Marin et al. (2010) and Demongeot et al., (2010), this study used the same conditions (A3) and the first part of (A5). In order to simplify the proofs, the author supposed (A4) and imposed the technical conditions (A5) as in Fan and Gijbels (1996) for non-parametric estimation. The convergence of the estimator can be obtained under minimal assumption (A6). (A7) and (A8) are classical hypotheses in the functional estimation in finite or infinite dimension spaces,while assumption (A9) is a supposition for missing at random, hence it is a technical condition for the concision of the proof of the main results.

5. Result and proof

The following denotes by C and C' some strictly positive generic constants. and defines the quantities, for any $a \in \mathcal{F}$ and for all $i = 1, \cdots, n$. $W_i = W(h_W^{-1} \wp(a, A_i))$ and

$$Q_i = Q(h_Q^{-1} \wp(a, A_i)).$$

Theorem 5.1. Under conditions (A1)-(A6), one has

$$\boldsymbol{\hat{\vartheta}(a) - \vartheta(a) = O\left(h_W^{y_1} + h_Q^{y_2}\right) + O\left(\frac{\log(n)}{n h_Q \chi_a(h_W)}\right)^{\frac{1}{2}}, a.co.}$$

Proof

1. Conditional density $G^a(.)$ is continuous, see (A2) and (A6) thus

$$\forall \varepsilon > 0, \exists \sigma(\varepsilon) > 0, \forall b \in (\vartheta(a) - \eta, \vartheta(a) + \eta), |G^a(b) - G^a(\vartheta(a))| \leq \sigma(\varepsilon) \Rightarrow |b - \vartheta(a)| \leq \epsilon.$$

By construction $\hat{\vartheta}(a) \in (\vartheta(a) - \eta, \vartheta(a) + \eta)$ then

$$\forall \epsilon > 0, \exists \sigma(\epsilon) > 0, |G^a(\hat{\vartheta}) - G^a(\vartheta(a))| \leq \sigma(\varepsilon) \Rightarrow |b - \vartheta(a)| \leq \varepsilon.$$

Therefore,one arrives finally at

$$\exists \sigma(\epsilon) > 0, \mathbb{P}(|\hat{\vartheta}(a) - \vartheta(a)| > \epsilon) \leq \mathbb{P}(|G^a(\hat{\vartheta}(a)) - G^a(\vartheta(a))| > \wp(\epsilon)).$$

In the other case, it comes directly by the definition of $\vartheta(a)$ and $\hat{\vartheta}(a)$ that:

$$\begin{aligned} \left|G^a(\hat{\vartheta}(a)) - G^a(\vartheta(a))\right| &= \left|G^a(\hat{\vartheta}(a)) - \hat{G}^a(\hat{\vartheta}(a)) + \hat{G}^a(\hat{\vartheta}(a)) - G^a(\vartheta(a))\right| \\ &\leq \left|G^a(\hat{\vartheta}(a)) - \hat{G}^a(\hat{\vartheta}(a))\right| + \left|\hat{G}^a(\hat{\vartheta}(a)) - G^a(\vartheta(a))\right| \\ &\leq 2\sup_{b \in \mathcal{S}} \left|\hat{G}^a(b) - G^a(b)\right|. \end{aligned} \tag{5.1}$$

The uniform complete convergence of the conditional density estimate beyond the compact set $|\vartheta - \eta, \vartheta + \eta|$ (see 1 below) can be used leading directly from both precedent inequalities to

$$\forall \epsilon > 0, \sum_{i=1}^{n} \mathbf{P}(|\hat{\vartheta}(a) - \vartheta(a)| > \epsilon) < \infty. \tag{5.2}$$

Finally, the following consistency result is checked as long as the following Lemmas could be verified.

Proposition 5.1. Under conditions of Theorem 5.1 one has that

$$\lim_{n \to \infty} \sup_{y \in \mathcal{S}} |\hat{G}^a(b) - G^a(b)| = 0 \text{a.co.}$$

$$\sup_{b \in \mathcal{S}} \left|\hat{G}^a(b) - G^a(b)\right| = O\left(h_W^{y1}\right) + O\left(h_Q^{y2}\right) + O\left(\sqrt{\frac{\log n}{n h_Q \chi_a(h_W)}}\right) \text{ a.} co.$$

Proof

It was already shown in (5.1) that

$$|G^a(\hat{\vartheta}) - G^a(\vartheta)| = 2\sup_{b\in\mathcal{S}}|\hat{G}^a(b) - G^a(b)|. \tag{5.3}$$

Now write the following Taylor expansion of function f^a:

$$G^a(\hat{\vartheta}) = G^a(\vartheta) + \frac{1}{j!}G^{a(j)}(\vartheta^*)(\hat{\vartheta} - \vartheta)^j,$$

for some ϑ^* between ϑ and $\hat{\vartheta}$, use (5.3), as along as it is possible to check that

$$\forall\tau > 0, \sum_{i=1}^{n} \mathbb{P}\left(G^{a(j)}(\vartheta^*) < \tau\right) < \infty. \tag{5.4}$$

Hence

$$(\hat{\vartheta} - \vartheta)^j = O\left(\sup_{b\in\mathcal{S}}|\hat{G}^a(b) - G^a(b)|\right), a.co., \tag{5.5}$$

so it suffices to check (5.4), and this is done directly by using the second part of (A8) together with (1).

The proof is a direct consequence of the following decomposition:

$$\begin{aligned}\widehat{G^a}(b) - G^a(b) &= \frac{1}{\hat{g}_D^a}\{(\hat{G}_N^a(b) - \mathbb{E}[\hat{G}_N^a(b)]) - (G^a(b) - \mathbb{E}[\hat{G}_N^a(b)])\}\\ &+ \frac{G^a(b)}{\hat{g}_D^a}(1 - \hat{g}_D^a),\end{aligned}$$

where

$$\hat{G}_D^a = \frac{1}{n(n-1)\mathbb{E}[\nu_{12}(a)]}\sum_{i\neq j} \nu_{ij}(a),$$

and

$$\hat{G}_N^a(b) = \frac{1}{n(n-1)h_Q\mathbb{E}[\nu_{12}(a)]}\sum_{i\neq j} \nu_{ij}(x)Q^{(1)}\left(h_Q^{-1}(b - B_j)\right),$$

and of the Lemmas below.

Lemma 5.1. Under assumptions (A1), (A3), and (A4) one has

$$1 - \hat{g}_D^a = O\left(\sqrt{\frac{\log(n)}{nh_Q\chi_a(h_W)}}\right), a.co.$$

$$\exists\wp > 0, \sum_n \mathbb{P}[|\hat{g}_D^a| < \wp] < +\infty.$$

Proof

One can write

$$|1 - \hat{g}^a(b)| \leq (1 - \hat{g}^a(b))/2 \Rightarrow |\hat{G}^a(b) - G^a(b)| \geq G^a(b)/2,$$

to arrive finally at

$$\mathbb{P}|1-\hat{g}^a(b)| \leq (1-\hat{G}^a(b))/2 \leq \mathbb{P}|\hat{g}^a(b)-g^a(b)| \geq g^a(b)/2.$$

It is enough to take, $\eta = (1-\hat{G}^a(b))/2$, to show the result.

Lemma 5.2. Under all suppositions (A1), (A2), (A4) and (A5,) one has

$$\sup_{b\in\mathcal{S}}\left|G^a(b)-\mathbb{E}[\hat{G}_N^a(b)]\right| = O\left(h_W^{y1}\right)+O\left(h_Q^{y2}\right), a.co.$$

Proof

From assumption (A4), and when pairs (A_i, B_i) are identically distributed, one obtains

$$\mathbb{E}[\hat{G}_N^a] = \mathbb{E}[\nu_{12}(a)[\mathbb{E}[h_Q^{-1}Q_1^1|A]]].$$

By using classical change of variables $t=\frac{b-r}{h_Q}$, one obtains

$$h_Q^{-1}\mathbf{E}[Q^{(1)}|A] = \int_{\mathbb{R}} Q^{(1)}G^A(b-h_G t)dt,$$

so

$$|\mathbb{E}[Q^{(1)}|A]-G^a(b)| \leq \int_{\mathbb{R}} Q^{(1)}|G^A(b-h_Q t)-G^a(b)|dt,$$

and by supposition (A2)

$$\mathrm{I}_{B(a,h_Q)}(A)\left|\mathbb{E}\left[Q^{(1)}\middle|A\right]-G^a(b)\right| \leq \int_{\mathbb{R}} Q^{(1)}(t)\left(h_Q^{y1}+|t|^{y2}h_Q^{y2}\right)dt.$$

Thus we obtain the claimed result of this Lemma from the direct consequence ofassumption (A5), and $Q^{(1)}$ is a probability density.

Lemma 5.3. Under the assumptions of Theorem 5.1 one has

$$\sup_{b\in\mathcal{S}}\left|\hat{G}_N^{a(j)}(b)-\mathbb{E}[\hat{G}_N^a(b)\right| = O\left(\sqrt{\frac{\log(n)}{nh_Q\chi_a(h_W)}}\right), a.co.$$

Proof

The demonstration of this lemma is given by a straightforward adaptation of the proof of Lemma 5.2 in (Barrientos-Marin et al., 2010) by writing

$$G_N^a(b) = E_1(E_2E_3-E_4E_5),$$

where

$$G_N^a(b) = \underbrace{\frac{n^2h_W\chi_a(h_W)}{n(n-1)\mathbb{E}[V_{12}]}}_{E_1}[\underbrace{\left(\frac{1}{n}\sum_{j=1}^n\frac{W_j(a)Q_j(y)}{h_Q\chi_x(h_W)}\right)}_{E_2}\underbrace{\left(\frac{1}{n}\sum_{i=1}^n\frac{W_i(a)\ell_i^2(a)}{h_W^2\chi_a(h_W)}\right)}_{E_3}$$
$$-\underbrace{\left(\frac{1}{n}\sum_{j=1}^n\frac{W_j(a)\ell_j(a)Q_j(b)}{h_Qh_W\chi_a(h_W)}\right)}_{E_4}\underbrace{\left(\frac{1}{n}\sum_{i=1}^n\frac{W_i(a)\ell_i(a)}{h_W\chi_a(h_W)}\right)}_{E_5}].$$

With the similar method utilised above, one can show that

$$E_i - \mathbb{E}[E_i] = O_{a.co.}\left(\sqrt{\frac{\log(n)}{nh_Q\chi_a(h_W)}}\right), for\ i = 2,4. \quad (5.6)$$

$$\mathbb{E}[E_l] = O(1)\ for\ l = 2,4. \quad (5.7)$$

$$Cov(E_2, E_3) = o\left(\sqrt{\frac{\log(n)}{nh_Q\chi_a(h_W)}}\right), \quad (5.8)$$

and

$$Cov(E_4, E_5) = o\left(\sqrt{\frac{\log(n)}{nh_Q\chi_a(h_W)}}\right). \quad (5.9)$$

First, note that (5.7) was already set by Barrientos et al., (2010). Thus, it suffices to show (5.6), (5.8) and (5.9) to finish the proof of the Lemma. To show the result (5.6), set

$$E_{l,\varsigma} - \mathbb{E}[E_{l,\kappa}] = \frac{1}{nh_Q\chi_a(h_W)}\sum_{i=1}^{n} U_i^{l,\varsigma} for l = 0,1,2, and \varsigma = 0,1$$

where

$$U_i^{l,\varsigma} = \frac{1}{nh_W^l h_Q^\varsigma \chi_a(h_W)}(W_i Q_i^\varsigma \ell_i^l - \mathbb{E}[W_i Q_i^\varsigma \ell_i^l]).$$

By (A3), one has $\frac{1}{h_W^l h_Q^\varsigma}(W_i \ell_i^l) < C$ and since $Q < 1$, therefore

$$\left|U_i^{l,\varsigma}\right| \le \frac{C}{\chi_a(h_W)}\ and\ \mathbb{E}[U_i^{l,\varsigma^2}] \le \frac{C\prime}{\chi_a(h_W)}.$$

Thus, using the classical Bernstein's inequality (Uspensky, 1937, p. 205) allows to give for all $\eta \in (0; \frac{C'}{C})$,

$$\mathbb{P}\left\{E_{l,\varsigma} - \mathbb{E}[U_{l,\varsigma}] > \eta\sqrt{\frac{logn}{nh_Q\chi_a(h_W)}}\right\} \le C'n^{-C\eta^2},$$

and the relevant choice of η permits to subtract that

$$\mathbb{P}\left\{E_{l,\varsigma} - \mathbb{E}[E_{l,\varsigma}] > \eta\sqrt{\frac{\log n}{nh_Q^2\chi_a(h_W)}}\right\} \le C'n^{-1-\gamma},$$

for l = 0,1,2 and ς = 0,1. Now, let us continue to show the results of (5.8) and (5.9). For both equations using the case that pairs $(A_i; B_i)$, $i = 1, \cdots, n$ are identically distributed, one obtains

$$\begin{cases} Cov(E_2, E_3) = \frac{1}{nh_Q h_W^2 \chi_a^2(h_W)} [\mathbb{E}[W_1^2 Q_1 \ell_1^2] - \mathbb{E}[W_1 Q_1]\mathbb{E}[W_1 \ell_1^2]], \\ Cov(E_4, E_5) = \frac{1}{nh_Q h_W^2 \chi_a^2(h_W)} \left[\mathbb{E}[W_1^2 Q_1 \ell_1^2] - \mathbb{E}[W_1 G_1 \ell_1]\mathbb{E}[W_1 \ell_1]\right]. \end{cases}$$

In fact, one has to evaluate for both results

$$\mathbb{E}\left[W_i Q_i^{\varsigma} \ell_i^l\right] for\ l = 0; 1; 2;\ and\ \varsigma = 0; 1.$$

Once more, as $Q < 1$, then for all l = 0; 1; 2; and ς = 0; 1; one has that

$$\mathbb{E}[W_i Q_i^{\varsigma} \ell_i^l] = O(\mathbb{E}[W_i \ell_i^l]),$$

and by using (Lemma 5.3 in Barrientos et al. (2010)), one obtains

$$\mathbb{E}[W_i Q_i^{\varsigma} \ell_i^l] = O(h_W^l h_Q^{\varsigma} \chi_a(h_W)),$$

which implies that

$$Cov(E_2, E_3) = O\left(\frac{1}{nh_Q \chi_a(h_W)}\right) = o\left(\frac{logn}{nh_Q \chi_a(h_W)}\right),$$

$$and\ Cov(E_4, E_5) = O\left(\frac{1}{nh_Q \chi_a(h_W)}\right) = o\left(\frac{\log n}{nh_Q \chi_a(h_K)}\right).$$

6. Application: simulated data

The aim of this simulation was to prove the effectiveness of this study by comparing the classical regression and the conditional mode with MAR in the presence of or without outliers.

Functional variable A is taken as a function with support [0,1] based on the following observation

$$A_i(t) = X_i t^2 + \cos(\pi Y_i t), i = 1, \dots, 200;\ t \in [0,1],$$

where X_i are i.i.d. $\sim U(0,1)$ and Y_i are i.i.d. $\sim \mathcal{N}(0,1)$, and are independent from X_i and Y_i. For simplicity, Figure 1 presents a sample of $n = 200$ of the covariable curves $A(t)$. The author defined response variable B by $B = \lambda(A) + \epsilon$, where λ is the regression operator with

$$\lambda(a) = \left(\int_0^1 a'(t)\mathrm{d}t\right)^2,$$

and $\epsilon \sim \mathcal{N}(0{,}0.075)$.

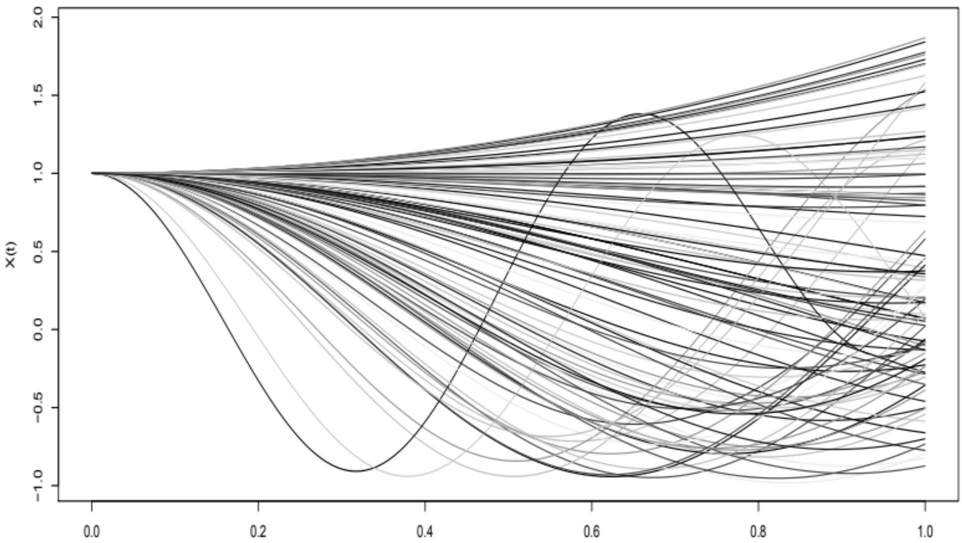

Fig. 1. The results of the analysis with the assumption that A_(i = 1, …, 100) (t), t ∈ [0,1]

Source: own calculations.

The main goal was to compare the sensitivity of the regression methods with MAR and the mode with MAR to outliers, the two estimators are defined as following $\forall b \in$ IR, hence

$$\hat{\lambda}^a(b) = \frac{\sum_{1\le i,j\le n} \nu_{ij}(a)B_j}{\sum_{1\le i,j\le n} \nu_{ij}(a)}, \tag{6.1}$$

and the mode estimator, and $\hat{G}^a(b)$ defined in (6.1).

Choosing the semi-metric on $\mathcal{F}$

$$d(a_i, a_j) = \sqrt{\int_0^1 \left(a_i'(t) - a_j'(t)\right)^2 dt}, for a_i, a_j \in \mathcal{F},$$

and the quadratic kernel defined as:

$$K(x) = \frac{3}{2}(1 - a^2)1 = \mathrm{I} = \mathrm{I}_{(0,1)}.$$

One splits the sample of size 200 into a learning sub-sample (A_i, B_i), $i = 1, \dots, 150$ and a testing sub-sample $(A_j, B_j), j = 151, \dots, 200$. For the missing mechanism, the author adopted the following expression

$$P(x) = \mathbb{P}(\wp = 1|A = a) = exp\left(2\alpha \int_0^1 a^2(t)dt\right),$$

where, exp $(\omega) = e^{\omega}/(1 + e^{\omega})$ for $\forall\omega \in \mathbb{R}$. To control the quantity $P(a)$, compute $\overline{\wp} = 1 - \frac{1}{150}\sum_{i=1}^{150} \wp_i$. For optimal bandwidth h_W, choose the automatic selection with a cross validation procedure introduced by (Ferraty & Vieu, 2006). Then calculate $\hat{\lambda}_{A_j}$ and $\hat{G}_{A_j}$ for $j = 151, \dots, 200$.

To highlight the performance of the results, the author plots the true values versus the predicted values for the MSE for both cases with complete data without outlier (CMSE), and MAR with outlier (MMSE), respectively.

Predicted values for the MSE for both cases complete the data and response missing at random MAR, respectively.

1. Complete case, the mean square error (CMSE) is

$$CMSE_{reg} = \frac{1}{50}\sum_{j=151}^{200}\left(\hat{\lambda}_{A_j} - T(A_j)\right)^2. \ And\ CMSE_{mode} = \frac{1}{50}\sum_{j=151}^{200}\left(\hat{G}_{A_j} - T(A_j)\right)^2.$$

2. Incomplete case response MAR, the mean square error (MMSE) is

$$MMSE_{reg} = \frac{1}{50}\sum_{j=151}^{200}\left(\hat{\lambda}_{A_j} - T(A_j)\right)^2. \ And\ MMSE_{mode} = \frac{1}{50}\sum_{j=151}^{200}\left(\hat{G}_{A_j} - T(A_j)\right)^2,$$

$T(A_j)$ means the response variable.

1) Complete case: the obtained results are shown in Figure 2, it is clear that there is no meaningful difference between the two estimation methods.

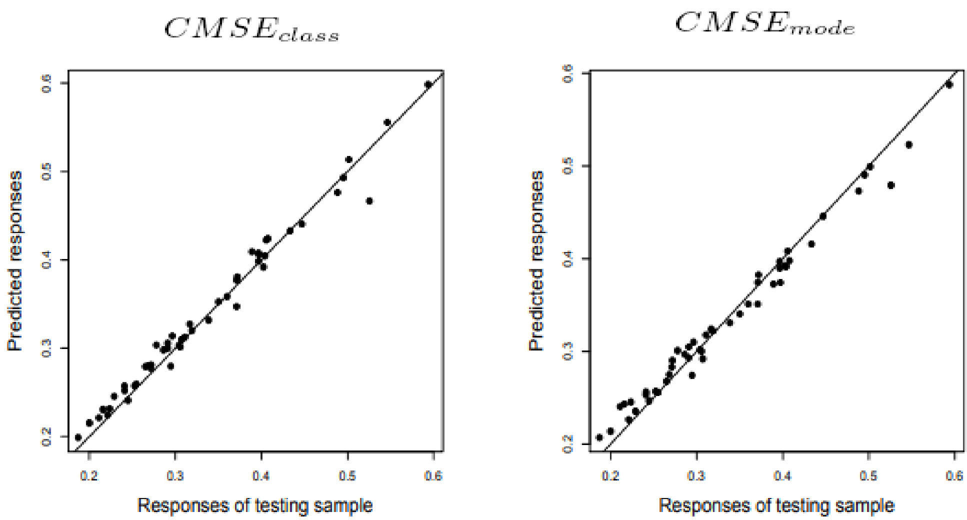

Fig. 2. The complete data case: CMSE

Source: own calculation.

2) Incomplete case response MAR: the obtained results in Figure 3 show that the mode estimation is better than the classical kernel regression; i.e. the classical kernel method with MAR.

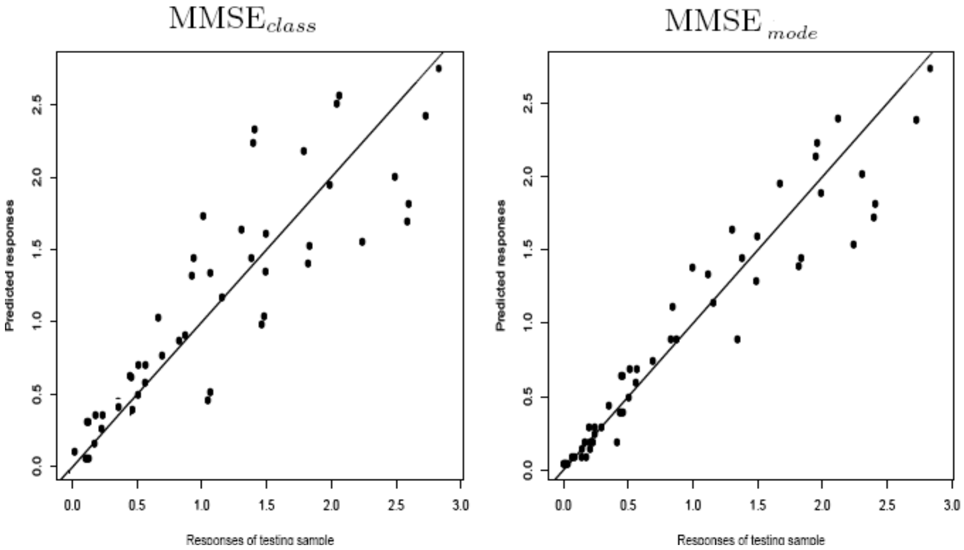

Fig. 3. The missing at random case: MMSE

Source: own calculation.

Table 1. MMSE comparison between both methods for the combinations of parameters (α)

α	$\bar{\wp}$	$MMSE_{class}$	$MMSE_{mode}$
0	0.50	0.0705089	0.07207680
0.5	0.33	0.06701626	0.0687210
1	0.23	0.05916404	0.06102103
1.5	0.15	0.05689561	0.05784979
2	0.10	0.05511259	0.05638462

Source: own calculation.

Now, compare the performance of both estimators(classic and mode) in the presence of outliers. To do this, the author introduced artificial outliers by multiplying some values of responses B by 100 with a fixed degree of dependence (α). The mode estimator has a better performance than the classical one, even if the MMSE of both estimators increases substantially relative to the number of outliers, but it remains very low for the mode method, as shown in Table 2.

Table 2. MMSE for the Classical Kernel Estimator and the mode Estimator according to the numbers of the introduced artificial outliers

Number of artificial outliers	0	10	20	40
Classical Estimator $MMSE_{reg}$	0.04605103	34.41909	104.3417	1112.265
Mode Estimator $MMSE_{mode}$	0.04678676	0.08507463	0.1022152	0.422661

Source: own calculation.

The mode estimator with MAR has a better performance than the classical MAR, even if the MMSE of both estimators increases substantially relative to the number of outliers, but it remains very low for the mode MAR method, as shown in Table 1.

7. Conclusion

In this paper, the author studied the problem of the nonparametric estimation of the conditional density with application to the mode function using the local linear approach. The feature of this study is the possibility to cover the incomplete data situation characterised by the missing phenomena. Empirical analysis showed the excellent performance of the proposed methodology, which varied with respect to the missing level. In addition to these features, the presented study opened up some significant avenues for the future research. In particular, it will be interesting to investigate other types of incomplete functional data, such as censored or truncated data. Another possible direction is to study the asymptotic property of kNN local linear estimator in the functional times series case (complete or incomplete cases). In addition, the asymptotic distribution of the proposed estimators is an interesting open question. Such asymptotic property is essential as preliminary statistical analyses, including the confidence interval or hypotheses testing. Moreover, extending this kind of estimation to other nonparametric models, such as the conditional hazard function or the conditional distribution function, is also a natural prospect of this study.

References

Almanjahie, I., Kaid, Z., Laksaci, A., & Rachdi, M. (2022). Estimating the conditional density in scalar--on-function regression structure: k-N-N local linear approach.*Mathematics*, *10*(6), 1-16.

Attouch, M.., Bouabsa, W., & Chiker el Mozoaur, Z. (2018). The k-nearest neighbors estimation of the conditional mode for functional data under dependency. *International Journal of Statistics & Economics*, *19*(1), 48-60.

Attouch, M., & Bouabsa, W. (2013). The k-nearest neighbors estimation of the conditional mode for functional data. *Rev. Roumaine Math. Pures Appl.*, *58*(4), 393-415.

Attouch, M., Laksaci, A., & Messabihi, N. (2015). Nonparametric relative error regression for spatial random variables. *Statistical Papers*, *58*(4), 987-1008.

Amiri, A., Dabo-Niang, S., & Yahaya, M. (2016). Nonparametric recursive density estimation for spatial data. *Comptes Rendus Math.*, *354* , 205-210.

Baìllo, A. & Grané, A. (2009). Local linear regression for functional predictor and scalar response. *Journal of Multivariate Analysis*, *100*, 102-111.

Barrientos-Marin, J., Ferraty, F., & Vieu, P. (2010). Locally modelled regression and functional data. *Journal of Nonparametric Statistics*, *22*(5), 617-632.

Boj, E., Delicado, P., & Fortiana, J. (2010). Distance-based local linear regression for functional predictors, *Computational Statistics and Data Analysis*, 54, 429-437.

Chaouch, M., Laïb, N., & Louani, D. (2014). Rate of uniform consistency for a class of mode regression on functional stationary ergodic data. *Stat. Meth. Appl.*, 26, 19-47.

Chahad, A., Aït-Hennani, L., & Laksaci, L. (2017). Functional local linear estimate for functional relative-error regression. *Journal of Statistical Theory and Practice*, *13*(11), 771-789.

Collomb, G., & Härdale, H. (1979). Strong uniform convergence rates in robust nonparametric time series analysis and prediction: Kernel regression estimation from dependent observations. *Stochastic Processes and Their Application*, *23*, 77-89.

Demongeot, J., Laksaci, A., Madani, F., & Rachdi, M. (2010). Local linear estimation of the conditional density for functional data. *C. R., Math., Acad. Sci. Paris*, *348*, 931-934.

Demongeot, J., Laksaci, A., Madani, F., & Rachdi, M. (2013). Functional data: local linear estimation of the conditional density and its application. *Statistics*, *47*, 26-44.

Ezzahrioui, M., & Ould-Said, E. (2006). *On the asymptotic properties of a nonparametric estimator of the conditional mode for functional dependent data*. Preprint, LMPA No 277, Univ. du Littoral Côte d'Opale.

Fan, J. (1992). Design-adaptative nonparametric regression. *Journal of the American Statistical Association*, *87*, 998-1004.

Fan, J., & Gijbels, I. (1996). *Local polynomial modelling and its applications.* Chapman & Hall.

Ferraty, F., Rabhi, A., & Vieu, P. (2008). Estimation non-paramétrique de la fonction de hasard avec variable explicative fonctionnelle. *Revue de Mathématiques Pures et Appliquées*, *53*, 1-18.

Ferraty, F., & Vieu, P. (2006). *Nonparametric functional data analysis. Theory and practice.* Springer--Verlag.

Ferraty, F. (2010). High-dimensional data: a fascinating statistical challenge. *J. Multivariariate Anal.*, *101*, 305-306.

Giraldo, R. Dabo-Niang S., & Martìnez, S. (2018). Statistical modeling of spatial big data: an approach from a functional data analysis perspective. *Statist. Probab. Lett.*, 136, 126-129.

Hurber, P. (1964). Robust estimation of a location parameter. *Ann. Math. Statist.*, *35*, 73-101.

Kirkby, J. L., Leitao, A., & Nguyen, D. (2021). Nonparametric density estimation and bandwidth selection with B-spline bases: a novel Galerkin method. *Comput. Statist. Data Anal.*, *159*.

Laib, N., & Ouled-Said, E. (2000). A robust nonparametric estimation of the autoregression function under an ergodic hypothesis. *Canad. J. Statist.*, *28*, 817-828.

Maillot, B., & Chesneau, C. (2021). On the conditional density estimation for continuous time processes with values functional in spaces. *Statist. Probab. Lett.*, *178*.

Rachdi, M., Laksaci, A., Demongeot, J., Abdali, A., & Madani, F. (2014). Theoretical and practical aspects on the quadratic error in the local linear estimation of the conditional density for functional data. *Computational Statistics & Data Analysis*, *73*, 53-68.

Uspensky, J. V. (1937). *Introduction to mathematical probability*. Mc Graw-Hill.

Zhou, Z., & Lin, Z. (2016). Asymptotic normality of locally modeled regression estimator for functional data. *Journal of Nonparametric Statistics*, *30*(28), 116-131.

Szacowanie gęstości warunkowej z wykorzystaniem modelu w strukturze regresji skalarnej na funkcji: lokalne podejście liniowe z losowym brakiem

Streszczenie: Celem analizy było zbadanie nieparametrycznego estymatora funkcji gęstości i trybu skalarnej zmiennej odpowiedzi na zmienną funkcyjną, gdy obserwacje są i.i.d. Ten proponowany estymator jest tworzony przez połączenie metody Missing At Random (MAR) z lokalnym podejściem liniowym. Na koniec zapewniono również badanie porównawcze oparte na symulowanych danych, aby zilustrować wydajność skończonej próbki i przydatność lokalnego podejścia liniowego z MAR do obecności nawet niewielkiej części wartości odstających w danych.

Słowa kluczowe: dane funkcjonalne, lokalna estymacja liniowa, funkcja trybu warunkowego, funkcjonalna statystyka nieparametryczna.