AIMS Mathematics, 10(7): 15929–15954.
DOI: 10.3934/math.2025714
Received: 24 March 2025
Revised: 17 June 2025
Accepted: 08 July 2025
Published: 15 July 2025

https://www.aimspress.com/journal/Math

Research article

The k nearest neighbors local linear estimator of semi functional partial linear model with missing response at random

Amina Naceri[1,3], Tawfik Benchikh[1,2,*], Ibrahim M. Almanjahie[4], Omar Fetitah[1,3], Mohammed Kadi Attouch[1] and Fatimah Alshahrani[5]

[1] Laboratory of Statistics and Stochastic Processes, University of Djillali Liabes BP 89, Sidi Bel Abbes 22000, Algeria; Email: benchikh.tawfik@gmail.com, attou_kadi@yahoo.fr
[2] Medical Faculty, Djillali Liabes University BP 89, Sidi Bel Abbes, 22000, Algeria
[3] Ecole Supérieure en Informatique, Sidi Bel Abbes, 22000, Algeria; Email: am.naceri@esi-sba.dz, o.fetitah@esi-sba.dz
[4] Department of Mathematics, College of Science, King Khalid University, Abha 62223, Saudi Arabia; Email: imalmanjahi@kku.edu.sa
[5] Department of Mathematical Sciences, College of Science, Princess Nourah bint Abdulrahman University, P.O. Box 84428, Riyadh 11671, Saudi Arabia; Email: fmalshahrani@pnu.edu.sa

* **Correspondence:** Email: benchikh.tawfik@gmail.com.

Abstract: This paper aims to investigate a semi-functional partial linear regression model in the presence of missing data in the response variable under the missing at random mechanism. We construct estimators using the kNN-local linear method and establish the asymptotic distribution of the parametric component. Additionally, the uniform almost complete consistency rates for the nonparametric component with respect to the number of neighbors under appropriate conditions is derived. Through simulations and real data analysis, we assess the effectiveness of the proposed approach and demonstrate its superiority by comparing it with existing methods for semi-functional partial linear regression models.

Keywords: functional data analysis; partial linear regression; missing at random data; kNN estimation; local linear estimation
Mathematics Subject Classification: 60G25, 62G05, 62G08, 62G20

1. Introduction

In recent years, functional data analysis (FDA) has emerged as a powerful framework, allowing each observation to be treated as a function. FDA provides tools and techniques to model, analyze, and

interpret complex data structures inherent in various application fields. By leveraging the continuous nature of data, FDA enables insights and predictions that traditional statistical methods cannot achieve (see [1–3] and bibliographic discussions such as [4–6]).

In this context, combining the flexibility of nonparametric methods with the functional nature of data provides a robust framework for modeling complex relationships. One prominent approach is nonparametric functional regression (FNR), which establishes a relationship between a scalar response variable and a functional explanatory variable (see [7–9] for more details). However, in many practical applications, multiple covariates are involved. In particular, it is common to have both functional and scalar explanatory variables linked to the response.

To address such scenarios, semi-functional partial linear regression (SFPLR) models offer an appealing approach. These models are widely applied in various fields where understanding dynamic relationships between variables over time or space is crucial. The SFPLR model, introduced by [10], is given by:

$$Y = \mathbf{X}^T\beta + m(\xi) + \epsilon\,,$$

where Y is the scalar response variable, $\mathbf{X} = (X_1, X_2, \ldots, X_p)$ is a p-dimensional vector of scalar explanatory variables, ξ is functional explanatory variable, β is an unknown p-dimensional parameter vector, $m(.)$ is an unknown smooth functional operator, and ϵ is the centered random error with finite unknown variance. Extensive research has been conducted in this area: Aneiros-Pérez et al. [10] introduced kernel-based estimation, Boente et al. [11] proposed robust estimation methods, Feng et al. [12] developed local linear estimators, and Ling et al. [13] studied k-nearest neighbors (kNN) approaches. Extensions of this model have been explored for dependent data [14] and spatial data [15]. The SFPLR model is particularly attractive due to its ability to balance interpretability (through the linear component) and flexibility (via the nonparametric functional component). For recent advances, see [16, 17].

An alternative to SFPLR is the single-index semi-functional partial linear model (SFPLSIM), which integrates functional single-index concepts for handling the functional variable ξ while maintaining the partially linear structure for multivariate covariates. This model filters the functional variable ξ to extract the component that explains Y. This model was studied by [18], where a novel automatic and locally adaptive procedure to estimate SFPLSIM components using kNN techniques was proposed, achieving uniform convergence rates for all model parameters.

More recently, Kedir et al. [19] introduced an estimation method based on kNN-local linear estimation (kNN-LLE) for independent and identically distributed (i.i.d.) data, building on earlier work by [20]. This approach is innovative in nonparametric functional data analysis (NFDA), as it significantly reduces bias compared to traditional kernel methods. By combining kNN and local linear estimation, it yields an estimator with enhanced statistical properties, leading to faster convergence, lower bias, and practical ease of implementation. Motivated by the advantages of this approach, several studies have been conducted, including regression operator estimation [20], conditional cumulative distribution estimation under dependence [21], the strong consistency of kNN-LLE estimation for functional conditional density and mode [22], conditional expectation estimation [23], and conditional density estimation for spatial functional data [24].

In real-world applications, missing data in the response variable frequently occurs, often following a "missing at random" (MAR) scenario. While missing data problems are extensively studied in multivariate analysis, research on this issue remains limited within non-functional data analysis

(NFDA). Early contributions to addressing missing data in NFDA include [25] for regression operator estimation, [26] for ergodic data, [27] for conditional mode estimation, and [28] for spatial functional regression. Other significant contributions include [29] on regression operator estimation and [30] on conditional distribution estimation for scalar response variables with missing data using kNN-local linear methods. For semi-functional partial linear regression (SFPLR) models with a MAR response, the first results were established by [31] using kernel methods for i.i.d. data. Subsequently, [32] further explored the SFPLR model under the setting of responses missing at random for spatially dependent data.

Building on these contributions, our work aims to advance research on SFPLR models for i.i.d. data in the presence of missing values in the response variable. Specifically, we construct estimators for both the parametric and nonparametric components using the kNN-LL approach and establish their asymptotic properties. However, since the bandwidth parameter is a random variable, analyzing the asymptotic properties requires additional theoretical tools and techniques.

The paper is organized as follows: Section 2 introduces our model. Section 3 presents the necessary notations and assumptions, along with the main theoretical results. Section 4 discusses simulation results and an application to real data. Finally, the proofs of our results are provided in the last section.

2. The model and its estimation

Let $(Y_i, \mathbf{X}_i, \xi_i)$, for $1 \leq i \leq n\}$, be a sequence of independent random variables drawn from the triplet $(Y, \mathbf{X}, \xi) \in \mathbb{R} \times \mathbb{R}^{\mathbf{p}} \times \mathcal{F}$, where $\mathcal{F}$ is a semi-metric space equipped with a semi-metric $d(.,.)$. The topological closed ball in $\mathcal{F}$ centered at ξ with radius h is denoted by $B(\xi, h) = \{\xi^{'} \in \mathcal{F}/d(\xi, \xi^{'}) \leq h\}$ and $\mathcal{N}_{\xi}$ denotes a neighborhood of ξ.

We consider the semi-functional partial linear regression (SFPLR) model, where the scalar response Y is linearly related to the p-dimensional random vector $\mathbf{X}$ and nonparametrically related to an independent functional covariate ξ. This model is defined as:

$$Y_i = \sum_{s=1}^{p} X_{is}\beta_s + m(\xi_i) + \epsilon_i = \mathbf{X}_i^T\boldsymbol{\beta} + m(\xi_i) + \epsilon_i \quad (i = 1, \ldots, n), \tag{2.1}$$

where $\mathbf{X}_i = (X_i^{(1)}, X_i^{(2)}, \ldots, X_i^{(p)})^T$, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^T$, $m(.)$ and ε_i are defined as before such that the random error satisfies

$$\mathbb{E}(\epsilon_i|\mathbf{X}_i, \xi_i) = 0 \text{ and } \mathbb{E}(\epsilon_i^2|\mathbf{X}_i, \xi_i) < \infty \ .$$

We assume that for any $\xi_0 \in \mathcal{N}_{\xi}$, the functions $m(\xi)$ can be locally approximated by

$$m(\xi_0) = a + b\varrho(\xi_0, \xi) + o(d(\xi_0, \xi)) \tag{2.2}$$

where $\varrho(\cdot, \cdot)$ is a bilinear continuous function from $\mathcal{F}^2$ into $\mathbb{R}$ such that $\varrho(\xi, \xi) = 0$.

Let $h_k = h_{k,n}$ be a sequence of bandwidths decreasing to zero as $\mathbf{n}$ tends to infinity, defined by:

$$h_k = \min\{h \in \mathbb{R}^+ \text{ such that } \sum_{i=1}^{n} \mathbb{1}_{B(\xi,h)}(\xi_i) = k\} \ .$$

Then, the kNN-LLE estimators of $\boldsymbol{\beta}_{\mathbf{n}}$ and $m_{\mathbf{n}}$ (as proposed in [19]) are given by:

$$\widehat{\boldsymbol{\beta}}_n = (\widetilde{\mathbf{X}}^T\widetilde{\mathbf{X}})^{-1}\widetilde{\mathbf{X}}^T\widetilde{Y} = \left(\sum_{i=1}^{n}\widetilde{\mathbf{X}}_i^T\widetilde{\mathbf{X}}_i\right)^{-1}\sum_{i=1}^{n}\widetilde{\mathbf{X}}_i^T\widetilde{Y}_i, \tag{2.3}$$

and

$$\widehat{m}_n(\xi) = \sum_{j=1}^{n}\mathbf{W_j}(\xi)Y_j - \left(\sum_{j=1}^{n}\mathbf{W_j}(\xi)\mathbf{X_j}\right)\widehat{\boldsymbol{\beta}}_n \tag{2.4}$$

where $\widetilde{\mathbf{X}} = (\widetilde{\mathbf{X}}_1, \ldots, \widetilde{\mathbf{X}}_n)$, $\widetilde{Y} = (\widetilde{Y}_1, \ldots, \widetilde{Y}_n)$ with $\widetilde{\mathbf{X}}_i = \mathbf{X}_i - \sum_{j=1}^{n}\mathbf{W}_j(\xi)\mathbf{X}_j$ and $\widetilde{Y}_i = Y_i - \sum_{j=1}^{n}\mathbf{W}_j(\xi)Y_j$. The weights $\mathbf{W}_j(\xi)$ are defined as

$$\mathbf{W}_j(\xi) = \sum_{i=1}^{n}W_{ij}(\xi, h_k)/\sum_{i=1}^{n}\sum_{j=1}^{n}W_{ij}(\xi, h_k)$$

where $W_{ij}(\xi, h_k) = \varrho_i(\varrho_i - \varrho_j)K_iK_j$, with $\varrho_i = \varrho(\xi_i, \xi)$ and $K_i = K(h_k^{-1}d(\xi, \xi_i))$ (K is a real-valued kernel function).

The main objective of this work is to adapt the above estimation procedure for SFPLR model in the presence of incomplete data. Specifically, we consider the case where the response variable Y is MAR, while the covariates X and ξ are fully observed.

To formalize this setting, we assume that the study is conducted on an incomplete sample of size n: $\{(Y_i, X_i, \xi_i, \delta_i), i = 1, \ldots, n\}$, where $\delta_i = 1$ if Y_i is observed and $\delta_i = 0$ if Y_i is missing. The missing data mechanism is assumed to satisfy the MAR condition:

$$\mathbb{P}(\delta_i = 1|Y_i, X_i, \xi_i) = \mathbb{P}(\delta_i = 1|X_i, \xi_i) = p(X, \xi).$$

The function p is generally unknown.

Given this, our model becomes:

$$\delta_iY_i = \delta_i\mathbf{X}_i^T\boldsymbol{\beta} + \delta_im(\xi_i) + \delta_i\epsilon_i \quad i = 1, \ldots, n. \tag{2.5}$$

Conditioning on $\xi_i = \xi$ gives

$$\mathbb{E}(\delta_iY_i|\xi_i = \xi) = \mathbb{E}(\delta_i\mathbf{X}_i|\xi_i = \xi)^T\boldsymbol{\beta} + \mathbb{E}(\delta_i|\xi_i = \xi)m(\xi)\,. \tag{2.6}$$

This yields:

$$m(\xi) = \frac{\mathbb{E}(\delta_iY_i/\xi_i=\xi)}{\mathbb{E}(\delta_i/\xi_i=\xi)} - \left(\frac{\mathbb{E}(\delta_i\mathbf{X}_i/\xi_i=\xi)}{\mathbb{E}(\delta_i/\xi_i=\xi)}\right)^T\boldsymbol{\beta} = m_2(\xi) - m_1^T(\xi)\boldsymbol{\beta}\,. \tag{2.7}$$

Substituting into (2.5), we get:

$$\delta_i\left(Y_i - m_2(\xi)\right) = \delta_i\left(\mathbf{X}_i - m_1(\xi)\right)^T\boldsymbol{\beta} + \delta_i\epsilon_i\,. \tag{2.8}$$

Thus, if the functions $m_1(\xi)$ and $m_2(\xi)$ are known, the least squares estimator (LSE) of $\boldsymbol{\beta}$ is given by

$$\overline{\boldsymbol{\beta}}_n = \arg\min_{\boldsymbol{\beta}}\sum_{i=1}^{n}\delta_i\left(Y_i - m_2(\xi_i) - (\mathbf{X}_i - m_1(\xi_i))^T\boldsymbol{\beta}\right)^2,$$

with explicit solution:

$$\hat{\boldsymbol{\beta}}_{\mathbf{n}} = \left[\sum_{i=1}^{n} \delta_i \left(\mathbf{X}_i - m_1(\xi_i)\right)^{\otimes 2}\right]^{-1} \left(\sum_{i=1}^{n} \delta_i \left(\mathbf{X}_i - m_1(\xi_i)\right)^T \left(Y_i - m_2(\xi_i)\right)\right), \tag{2.9}$$

where $A^{\otimes 2}$ denotes $A^T A$.

In practice, the functions $m_1(\xi)$ and $m_2(\xi)$ are unknown and must be estimated to apply Eq (2.9). Under the local approximation:

$$m_l(\xi_0) = a + b\varrho(\xi_0, \xi) + o(d(\xi, \xi_0)), l = 1, 2. \tag{2.10}$$

We define the the LLE-kNN estimator of $m_1(\xi)$ and $m_2(\xi)$ (see [29]) as:

$$\widehat{m}_1(\xi) = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} W_{ij}(\xi, h_k) X_j}{\sum_{i=1}^{n}\sum_{j=1}^{n} W_{ij}(\xi, h_k)} = \sum_{j=1}^{n} \mathbf{W}_j(\xi)\mathbf{X_j}\,, \tag{2.11}$$

and

$$\widehat{m}_2(\xi) = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} W_{ij}(\xi, h_k) Y_j}{\sum_{i=1}^{n}\sum_{j=1}^{n} W_{ij}(\xi, h_k)} = \sum_{j=1}^{n} \mathbf{W}_j(\xi) Y_j\,, \tag{2.12}$$

where $W_{ij}(\xi, h_k) = \delta_i\delta_j\varrho_i(\varrho_i - \varrho_j)K_iK_j$. Here, the weight function is defined as

$$\mathbf{W}_j(\xi) = \frac{\sum_{i=1}^{n} W_{ij}(\xi, h_k)}{\sum_{i=1}^{n}\sum_{j=1}^{n} W_{ij}(\xi, h_k)}\,.$$

Similar to classical kNN methods, the choice of the parameter k is critical and not known a priori. In this work, the optimal number of neighbors k is selected via cross-validation:

$$k_{opt} = \arg \min_{k\in]k_{1,n},k_{2,n}[} CV(k) = \arg \min_{k\in]k_{1,n},k_{2,n}[} \sum_{i=1}^{n} \delta_i \left(Y_i - \widetilde{Y}_{(-i)}^{kNN}\left((X_i, \xi_i)\right)\right)^2, \tag{2.13}$$

where $\widetilde{Y}_{(-i)}^{kNN}$ represents the leave one-out kNN-LLE estimator, and $k_{1,n}$ and $k_{2,n}$ are two sequences of strictly positive integers. The existence of such sequences is ensured by the results of [33] (see also [23] for more details).

Using the estimates $\widehat{m}_1(\xi)$ and $\widehat{m}_2(\xi)$, we define the adjusted estimator of $\boldsymbol{\beta}$ as

$$\widehat{\boldsymbol{\beta}}_n = (\delta\widetilde{\mathbf{X}}^T\widetilde{\mathbf{X}})^{-1}\delta\widetilde{\mathbf{X}}^T\widetilde{Y} = \left(\sum_{l=1}^{n} \delta_l \widetilde{\mathbf{X}}_l^T\widetilde{\mathbf{X}}_l\right)^{-1} \sum_{l=1}^{n} \delta_l \widetilde{\mathbf{X}}_l^T\widetilde{Y}_l, \tag{2.14}$$

where $\widetilde{\mathbf{X}} = (\widetilde{\mathbf{X}}_1, \ldots, \widetilde{\mathbf{X}}_n)$ and $\widetilde{Y} = (\widetilde{Y}_1, \ldots, \widetilde{Y}_n)$ with $\widetilde{\mathbf{X}}_i = \mathbf{X}_i - \sum_{j=1}^{n} \mathbf{W_j}(\xi)\mathbf{X_j}$, $\widetilde{Y}_i = Y_i - \sum_{j=1}^{n} \mathbf{W_j}(\xi)Y_j$.

Finally, using Eqs (2.7) and (2.14), we obtain a nonparametric estimator of m,

$$\widehat{m}_n(\xi) = \widehat{m}_2(\xi) - \widehat{m}_1(\xi)^T\widehat{\boldsymbol{\beta}}_n = \sum_{j=1}^{n} \mathbf{W_j}(\xi)Y_j - \left(\sum_{j=1}^{n} \mathbf{W_j}(\xi)\mathbf{X_j}\right)^T \widehat{\boldsymbol{\beta}}_n. \tag{2.15}$$

3. Notations and assumptions

Before presenting the main results, we introduce some notations that will be used throughout the analysis. For all $l = 1, \ldots, n$ and $s = 1, \ldots, p$, we define

$$m^s_{1,l}(\xi) = \mathbb{E}(\delta X_{ls}|\xi_l = \xi), \quad \eta_{ls} = X_{ls} - m^s_{1,l}(\xi_l)\,.$$

We also set

$$\boldsymbol{\eta}_l = (\eta_{l1}, \ldots, \eta_{lp})^T\,, \quad \boldsymbol{\theta}_l = \boldsymbol{\eta}_l \varepsilon_l \qquad \boldsymbol{\Theta} = (\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_n)\,.$$

Let $\mathcal{S}$ be a subset of $\mathcal{F}$ such that: $\mathcal{S} \subset \bigcup_{l=1}^{d_n} B(Z_l; r_n)$, where $d_n > 1$ and r_n are sequences of positive real numbers, and $Z_l \in \mathcal{F}$, for $l = 1, \ldots, d_{\mathbf{n}}$.

Our objective is to establish the almost complete (a.co.) convergence of $\widehat{m}_n(\xi)$ and to derive the asymptotic properties of the estimator $\widehat{\boldsymbol{\beta}}_n$. To achieve this, we fix a point ξ in $\mathcal{F}$ and a neighborhood $\mathcal{N}_\xi$ of ξ and assume the following conditions:

3.1. Technical assumptions

(A1) For any $h > 0$, the function $\phi_\xi(h) := \mathbb{P}(B(\xi, h)) > 0$ is continuous in the neighborhood of 0 with $\phi_\xi(0) = 0$.

A2) There exists a bounded, positive function τ and a differentiable, invertible, nonnegative function ϕ such that

$$\sup_{\mathcal{S}} \left| \frac{\phi_\xi(h)}{\phi(h)} - \tau(\xi) \right| = O(h^\nu)\,, as\, h \to 0\,, \text{for some } \nu > 0.$$

The function $\phi(.)$, is such that

i) for all $\xi \in \mathcal{S}$: $0 < C_1\phi(h) \leq \phi_\xi(h) < C_2\phi(h)$, for $C_1 > 0, C_2 > 0$,

ii) $\exists h_0 > 0$, such that for all $h_0 > h$, $\phi'(h) < C$, where ϕ' is the derivative of ϕ.

iii) There exists a function $\iota(t)$ such that, for all $t \in]0, 1[$:

$$\lim_{h_{\mathbf{n}} \to 0} \frac{\phi(t\, h_{\mathbf{n}})}{\phi(h_{\mathbf{n}})} = \iota(t) = t^\upsilon\,,\ \upsilon > 0.$$

(A3) The functions m_1 and m_2 satisfy condition (2.10) with continuous operator ϱ. Moreover, there exists two positive constants C and C' such that $\forall r \geq 3$, $\mathbb{E}(|X_{1s}|^r|\xi_1 = \xi) < a_r(\xi) < C < \infty$ for $s = 1, \ldots, p$, and $\mathbb{E}(|Y|^r|\xi_1 = \xi) < b_r(\xi) < C' < \infty$, where $a_r(\cdot)$ and $b_r(\xi)$ are a continuous functions on $\mathcal{N}_\xi$.

(A4) $\exists \alpha > 0$, $\exists C > 0$ such that

$$m, m^1_{1,i}, m^2_{1,i}, \ldots, m^p_{1,i} \in \{f : \mathcal{F} \to \mathbb{R}, |f(\xi_1) - f(\xi_2)| \leq C\, d^\alpha(\xi_1, \xi_2), \forall (\xi_1, \xi_2) \in \mathcal{N}^2_\xi\}.$$

(A5) The kernel function K is supported within $[0, 1/2]$ and it has a continuous first derivative on $[0, 1/2]$ such that there exist two constants C and C' satisfying

$$-\infty < C' < K'(t) < C < 0\,,$$

$$K(1/2)\iota(1/2) - \int_0^{1/2} K'(t)\iota(t)dt > 0,$$

and

$$(1/4)K(1/2)\iota(1/2) - \int_0^{1/2} (u^2K'(t)\iota(t)dt) > 0.$$

(A6) Let the functions class $\mathcal{K}^a = \{\cdot \mapsto \gamma^{-a}K(\gamma^{-1}d(\xi,\cdot))\varrho^a(\cdot,\xi), \gamma > 0\}$, for $a = 0, 1, 2$. We assume that these functions classes are a pointwise measurable, and satisfies the following condition:

$$\sup_{\mathbb{Q}} \int_0^1 \sqrt{1 + \log \mathcal{N}(\epsilon\|F\|_{\mathbb{Q},2}, \mathcal{K}^a, d_{\mathbb{Q},2})}d\epsilon < \infty,$$

where F is the envelope function of the set $\mathcal{K}^a$ and where the supremum is taken over all probability measures $\mathbb{Q}$ on the space $\mathcal{F}$ with $Q(F^2) < \infty$. $\|.\|_{\mathbb{Q},2}$ is the norm $L^2(\mathbb{Q})$ and $d_{\mathbb{Q},2}$ is the metric associated with the norm $\|.\|_{\mathbb{Q},2}$. Finally, $\mathcal{N}(\epsilon, \mathcal{K}^a, d_{\mathbb{Q},2})$ denotes the minimal number of open balls with radius ϵ which are needed to cover the functions class $\mathcal{K}^a$ in the topological space given by $d_{\mathbb{Q},2}$.

(A7) There exist positive constants C_1, C_2 and C'' such that the bi-function ϱ satisfies the following conditions:

$$\forall \xi^{'} \in \mathcal{F}\,, \quad C_1\, d(\xi, z) \leq |\, \varrho(\xi, z)\, | \leq C_2\, d(\xi, z)\,,$$

$$\forall \xi_1, \xi_2 \in \mathcal{S}\,, \qquad |\varrho(\xi_1, \xi) - \varrho(\xi_2, \xi)| \leq C''d(\xi_1, \xi_2)\,.$$

(A8) For any sequence $h := h_n \in]\phi^{-1}(k_{2,n}/n), \phi^{-1}(k_{1,n}/n)[$, we have

$$h\left(\int_{B(\xi,h/2)} \varrho(u,\xi)dP^{\xi}(u)\right) = o\left(\int_{B(\xi,h/2)} \varrho^2(u,\xi)dP^{\xi}(u)\right),$$

where dP^{ξ} is the cumulative distribution of r.v. ξ.

(A9) The sequences $(k_{1,n})$ and $(k_{2,n})$ are such that

$$\phi^{-1}(k_{2,n}/n) \to 0 \text{ and } \frac{\log n}{\min\left(\phi^{-1}(k_{1,n}/n), k_{1,n}\right)} \to 0 \text{ as } n \to \infty.$$

(A10) The subset $\mathcal{S}$ is such that, for $r_n = O(\frac{\log n}{n})$, the sequence d_n satisfies:

$$\frac{(\log n)^2}{k} < \log d_n < \frac{k}{\log n}$$

and there exits $\varsigma > 1$ such that

$$\sum_n d_n^{1-\varsigma} < \infty.$$

(A11) Let $\mathbf{\Sigma} = \mathbb{E}\left[p(X_1, \xi_1)\left(\boldsymbol{\eta}_1(\boldsymbol{\eta}_1)^T\right)\right]$ and $\mathbf{B} = \mathbb{E}\left[p(X_1, \xi_1)\left(\mathbf{\Theta}_1(\mathbf{\Theta}_1)^T\right)\right]$.

- We assume that $\mathbf{\Sigma}$ is an invertible matrix.
- The matrix $\mathbf{B}$ is assumed to be positive definite.

Comments on the assumptions:

Most of these conditions are standard in the literature on local linear smoothing using the kNN approach (see [23]). Although they are not the weakest possible, they are sufficient to derive our theoretical results. More specifically, and in line with the discussions in [34], assumptions (A1) and (A2) are standard conditions related to small-ball probability. In particular, assumption (A2) states that the small-ball probability can be approximately decomposed as the product of two independent functions, ϕ and τ and the assumption (A2) (i-ii) ensures the usual concentration property of the functional variable, a well-documented key feature for capturing the functional nature of the data. Meanwhile, the assumption (A2)(iii) provides information on the variability of the small-ball probability, which plays a crucial role in controlling the bias of nonparametric estimators. This condition holds for various continuous-time processes, including Gaussian processes, diffusion processes, and general Gaussian processes (see [7, 35] for examples). Furthermore, the assumptions (A5), (A7) and (A8) are identical to those introduced by [36] in the context of functional local linear regression, and the assumptions (A6) and (A9) follow the framework established by [37] to ensure uniform integrated bias (UIB) consistency for any kNN-based estimator. Assumptions (A3), (A4), and (A11) are standard in SFPLR models (see, for example, [10]). The entropy condition in (A10) is met in several common cases (see [38]). These conditions help establish uniform convergence rates over the functional variable. Naturally, these conditions impose constraints on the small-ball probability function ϕ, as reflected in assumption (A2). However, this restriction can be relaxed by imposing alternative conditions on the set $S_{\mathcal{F}}$ (see [31] for more details).

3.2. Main results

We are now in position to give our asymptotic results. The first one gives the asymptotic distribution of the estimator for the parametric component of the model, whereas the second one precises the rate of almost complete convergence for the nonparametric component.

Theorem 3.1. *Under assumptions (A1)–(A11), if in addition* $\frac{\sqrt{n}\log d_n}{k_{1,n}} \to 0$ *as* $n \to \infty$, $\frac{\sqrt{n}\log^2 n}{k_{1,n}} \to 0$ *as* $n \to \infty$, $\sqrt{n}\phi^{-1}(\frac{k_{2,n}}{n})^{\alpha} \to 0$ *as* $n \to \infty$ *and* $k \geq n^{(2/r)+b-1}/(\log n)^2$ *for some constant* $b > 0$ *satisfying* $(\frac{2}{r}) + b > 1/2$ *(where* $r \geq 3$*) and n large enough, then we have*

$$\sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \xrightarrow{\mathcal{D}} N(0, \boldsymbol{\Sigma}^{-1}\mathbf{B}\boldsymbol{\Sigma}^{-1}).$$

Theorem 3.2. *Under the assumptions of Theorem 3.1, we have*

$$\sup_{k_{1,n}\leq k\leq k_{2,n}} \sup_{\xi\in\mathcal{S}} |\widehat{m}(\xi) - m(\xi)| = O\left(\phi^{-1}\left(\frac{k_{2,n}}{n}\right)^{\alpha}\right) + O_{a.co.}\left(\sqrt{\frac{\log d_n}{k_{1,n}}}\right).$$

Note that the proofs of the asymptotic results are given in the appendix.

4. Simulation and application studies

4.1. Simulation study

The main objective of this section is to examine the performance of the kNN-LLE approach on finite samples. More specifically, to highlight the superiority of this method compared to others, we compare the prediction accuracy of two models: Functional nonparametric regression and semi-functional partial linear regression. To evaluate the performance, we examine the mean square error (MSE) of prediction for three estimators: kNN estimator and kNN-LLE estimators. The proposed regression estimators are as follows:

- Functional nonparametric kNN regression (FNP.kNN) introduced by [39];
- Functional nonparametric kNN-local linear regression (FNP.kNN-LLE) developed by [29];
- Semi-functional partial linear kNN regression (SFPLR.kNN) proposed by [13];
- Semi-functional partial linear local-linear kNN regression (SFPLR.kNN-LLE), our estimator given by the Eq (2.15).

Recall that the functional nonparametric (FNP) model is defined as:

$$Y_i = m(\xi_i) + \epsilon_i,\ i = 1, \ldots, n,$$

and the kNN-kernel estimators for the SFPL and FNP models are given, respectively, by the following expressions:

- For the semi-functional partial linear (SFPL) model:

$$\widehat{m}_n(\xi) = \widehat{m}_2(\xi) - \widehat{m}_1(\xi)^T\widehat{\boldsymbol{\beta}}_n = \sum_{i=1}^{n} \delta_i w_i(\xi) Y_i - \left(\sum_{i=1}^{n} \delta_i w_i(\xi)\mathbf{X_i}\right)^T \widehat{\boldsymbol{\beta}}_n, \tag{4.1}$$

- and for the functional nonparametric (FNP) model:

$$\widehat{m}_n(\xi) = \sum_{i=1}^{n} \delta_i w_i(\xi) Y_i. \tag{4.2}$$

In both cases, the weights $w_i(\xi)$ are defined by:

$$w_i(\xi) = \frac{K\left(d(\xi, \xi_i)/h_k\right)}{\sum_{j=1}^{n} \delta_j K\left(d(\xi, \xi_j)/h_k\right)},$$

where:$K(\cdot)$ is a kernel function, $d(\cdot, \cdot)$ is a semi-metric, h_k is the bandwidth associated with the number of neighbors k, δ_i is the missingness indicator for observation i, and $\widehat{\boldsymbol{\beta}}_n$ is the estimated parametric component in the SFPL model.

Additionally, this section includes naive estimators based on a missing at random (MAR) mechanism. These estimators are based on complete case analysis, where all observations with missing

responses are discarded, without accounting for the potential information contained in the covariates $\mathbf{X}$ and ξ. It should be emphasized that the naive estimator is inconsistent unless the missingness indicator δ is independent of the covariates X and ξ, which is a strong and often unrealistic assumption (see [25] for details). In fact, while such estimators are simple to implement, excluding incomplete cases can lead to substantial bias in the estimation of model parameters, particularly when the missingness is informative, the proportion of missing data is high, or the missingness is related to the response variable. Moreover, restricting the analysis to complete cases reduces the effective sample size, resulting in decreased precision and a loss of efficiency. Ignoring missing data can also lead to model misspecification and potentially misleading conclusions about the relationships between variables. For these reasons, the naive estimator is included in our study solely as a benchmark, and not as a competing method.

Formally, we generate our observations (Y_i, X_i, ξ_i) from the SFPLR model, i.e.,

$$Y_i = m(\xi_i) + \sum_{j=1}^{2} X_{i,j}\beta_j + \epsilon_i,\ i = 1, \ldots, n = 250\,,$$

where $X_{i,j} \sim Exp(0.5)$, $\beta = (2, 1)^T$, $\epsilon_i \sim \mathcal{N}(0, 1)$, and we take the nonparametric operators $m(.)$ as

$$m(\xi) = \int_0^{\pi} \xi^2(t)dt.$$

The functional explanatory variables $\xi_i(t)$ are defined as: $\xi_i(t) = 2\sin(W_i t)^3 + 3\sin(2W_i + t)^2 + W_i t$, for $t \in [0, \pi]$ and $W_i \sim \mathcal{N}(0, 1)$.

The sample of curves $\{\xi_i\}_1^n$ can be observed in Figure 1.

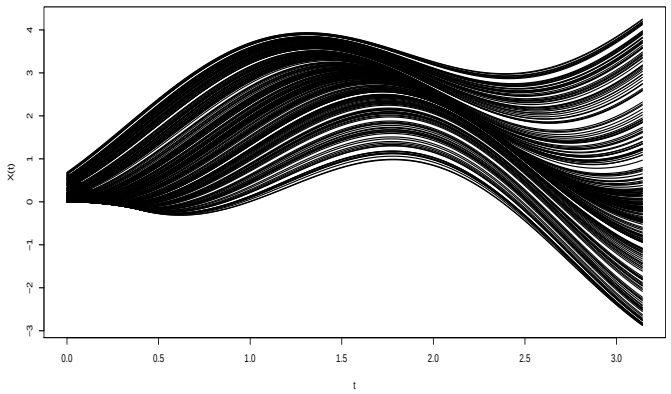


Figure 1. A sample of $n = 250$ functional explanatory variables ξ.

Moreover, similar to that described in [31], we adopted the following missing data mechanism,

$$p(x, z) = \mathbb{P}(\delta = 1|X = x, \xi = z) = \text{expit}\left(2\alpha\left(\sum_{j=1}^{2}|x_j| + \int_0^{\pi} z^2(t)\mathrm{d}t\right)\right),$$

where expit$(u) = e^u/(1+e^u)$ for all $u \in \mathbb{R}$ and we will take for α the following values: $\alpha = 0.25, 0.5, 1$. Recalling that the degree of dependence between the variable (X, ξ) and the variable δ is controlled by the parameter α and to check the value of $p(x, z)$, we compute $\overline{\delta} = 1 - \frac{1}{n}\sum_{i=1}^{n} \delta_i$. In order to compute our estimators, we use the class of semi-metrics $d(.,.)$ based on derivatives which are well adapted to this type of data (smooth data),

$$d(\xi_1, \xi_2) = \sqrt{\int_0^{\pi} (\xi_1^{(s)}(t) - \xi_2^{(s)}(t))^2 dt}\,, \tag{4.3}$$

where $\xi^{(s)}(t)$ denotes the s^{th} derivative of the curve $\xi(t)$, and we selected the asymmetric quadratic kernel K defined by

$$K(u) = \frac{3}{4}(1-u^2)\mathbb{1}_{[0,1/2]}(u). \tag{4.4}$$

The local-linear estimator (for the two models FNP and SFPLR) is constructed by the same procedure proposed by [36] for which the locating function ϱ is defined by

$$\varrho(\xi_1, \xi_2) = \int_0^{\pi} \theta(t)(\xi_1^{(s)}(t) - \xi_2^{(s)}(t))dt\,, \tag{4.5}$$

where θ is the eigenfunction of the empirical covariance operator, i.e.,

$$\frac{1}{n}\sum_{i=1}^{n}(\xi_i^{(s)}(t) - \overline{\xi^{(s)}(t)})^t(\xi_i^{(s)}(t) - \overline{\xi^{(s)}(t)}), \quad with \ \overline{\xi^{(s)}(t)} = \frac{1}{n}_{i=1}\sum_{i=1}^{n}\xi_i^{(s)}(t),$$

associated with the q-greatest eigenvalue.

On other hand, the model's performance depends on the parameters used in the estimation process. In fact, bandwidth parameters play a critical role in nonparametric estimation, affecting all asymptotic properties, and in particular the rate of convergence. In our study, the k-nearest neighbors technique is utilized to derive $h_{k_{opt}}$, which represents the bandwidth associated with the optimal number of neighbors, as determined by following cross-validation,

$$h_{k_{opt}} = \min\left\{h \in \mathbb{R}^+ \text{ such that } \sum_{i=1}^{n}\mathbb{1}_{B(\xi,h)}(\xi_i) = k_{opt}\right\},$$

where

$$k_{opt} = \arg\min_{k\in]k_{1,n},k_{2,n}[} CV(k) = \arg\min_{k\in]k_{1,n},k_{2,n}[}\sum_{i=1}^{n}\delta_i\left(Y_i - \widetilde{Y}_{l,(-i)}^{kNN}((X_i,\xi_i))\right)^2,\ l=1,2,$$

where $k_{1,n}$ and $k_{2,n}$ are two sequences of strictly positive integers, and $\widetilde{Y}_{1,(-i)}^{kNN} = \widehat{m}_n^{(-i)}(\xi_j)$ (resp $\widetilde{Y}_{2,(-i)}^{kNN} = (\widehat{m}_n^{(-i)}(\xi_j) + \mathbf{X}_j^T\widehat{\beta}_n)$) are the leave-one-out values of the functional nonparametric regression estimators calculate without observation (ξ_i) (resp. the leave-one-out values of the semi-functional nonparametric regression estimators calculate without observation $(\mathbf{X}_i, \xi_i)$ after estimating $\beta_{\mathbf{n}}$). (see [40, 41] for more details).

Remark 4.1. *For simplicity, the optimal parameter k_{opt} is selected using cross-validation, following the procedure described in [23]. Note that k_{opt} is a data-driven quantity, as it depends on the entire sample. This dependency makes the theoretical analysis of the resulting kNN-LLE estimator particularly challenging. To date, no theoretical or empirical results are available for the case where k is selected via cross-validation. Nevertheless, one can investigate the performance of the estimator by comparing the predictive accuracy obtained for several fixed values of k with that achieved using k_{opt} (see [13]). Simulation studies suggest that estimators based on k_{opt} are highly competitive in practice. Alternative methods for selecting k can also be considered, such as Bayesian approaches (see [42]) or minimax criteria (see [43]). For insights into uniform selection methods, we refer the reader to [37,44].*

In this simulation study, we take the parameters: $s = 2$ and $q = 8$. The sample of size n is randomly split into two parts: A training set $S_{train} = \{(Y_i, \delta_i, X_i, \xi_i)_{i\in \text{Train}}\}$, consisting of $n - 50$ observations, used for model estimation, and a testing set $S_{test} = \{(Y_i, \delta_i, X_i, \xi_i)_{i\in \text{Test}}\}$ consisting of 50 observations, used to evaluate prediction performance.

To assess the effectiveness of the proposed model for this prediction problem, we calculated the mean square error of prediction ($MSEP$) on the test set:

$$MSEP = \frac{1}{n_{\text{Test}}} \sum_{i\in \text{Test}} \left(Y_i - \widetilde{Y}_l\right)^2 \quad l = 1, 2,$$

where $n_{\text{Test}} = 50$ (the length of the testing sample) and $\widetilde{Y}_1$ (resp $\widetilde{Y}_{2,}$) denotes the predicted value from the functional nonparametric regression estimators calculate at (ξ_i) (resp. from the semi-functional nonparametric regression estimator calculate at (X_i, ξ_i)).

For each fixed value of α, we performed $M = 100$ independent replications of the experiment. Each replication yields an estimate of the mean squared error of prediction (MSEP), resulting in a total of M independent MSEP values. We summarize the distribution of these estimates using boxplots and compute the average MSEP as follows:

$$\overline{MSEP} = \frac{1}{M} \sum_{u=1}^{n} MSEP_u,$$

where $MSEP_u$ is the prediction error computed from the u^{th} replication.

Table 1 reports the values of $\overline{\text{MSEP}}$ for the six models at $k = 50,\ 70,\ 90$, and at the optimal value k_{opt}, as well as the $\overline{\text{MSEP}}$ of the naive estimator computed using k_{opt}.

Table 1. $\overline{MSEP}$ (mean squared error) for the six models for $k = 50, 70, 90$ and k_{opt}.

Model	α	Missing	kNN				kNN-LLE				Naive
			k_{opt}	$k = 50$	$k = 70$	$k = 90$	k_{opt}	$k = 50$	$k = 70$	$k = 90$	k_{opt}
	0,25	30%	23.963	25.786	25.632	25.547	22.731	25.719	25.560	25.492	24.993
FNP	0,5	16%	22.886	23.973	23.828	23.782	22.641	23.058	22.867	22.789	24.096
	1	05%	22.625	23.105	22.900	22.897	22.514	22.859	22.764	22.673	23.915
	Complete	00%	22.481	23.194	23.155	22.658	22.447	22.782	22.667	22.497	23.005
	0,25	30%	0.419	1.152	0.802	0.523	0.384	0.743	0.694	0.672	0.645
SFPLR	0,5	16%	0.404	0.936	0.794	0.452	0.167	0.389	0.283	0.192	0.638
	1	05%	0.391	0.565	0.463	0.482	0.131	0.241	0.233	0.151	0.596
	Complete	00%	0.385	0.479	0.441	0.406	0.105	0.182	0.167	0.109	0.481

Table 1 confirms that selecting the optimal k significantly improves prediction accuracy. The kNN-LLE method consistently outperforms the standard kNN and naive approaches, especially under the SFPLR model. Moreover, its performance remains robust even in the presence of missing data.

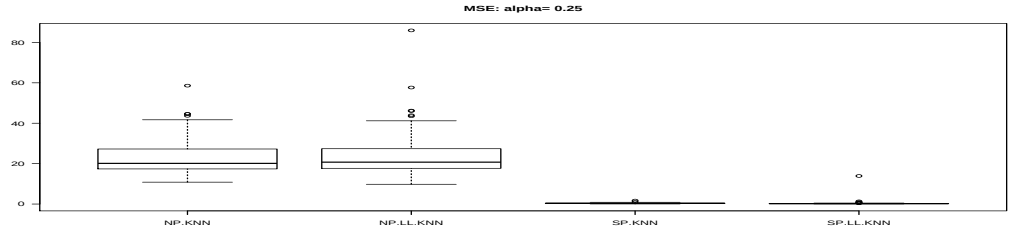
The main conclusions drawn from the Table 1 are as follows:

- The estimators are highly sensitive to the choice of their tuning parameters;
- The cross-validation selectors perform competitively;
- The predictive performance of the kNN-LL estimator surpasses that of the kNN-kernel estimator;
- The SFPL model demonstrates superior goodness-of-fit compared to the FNP model.

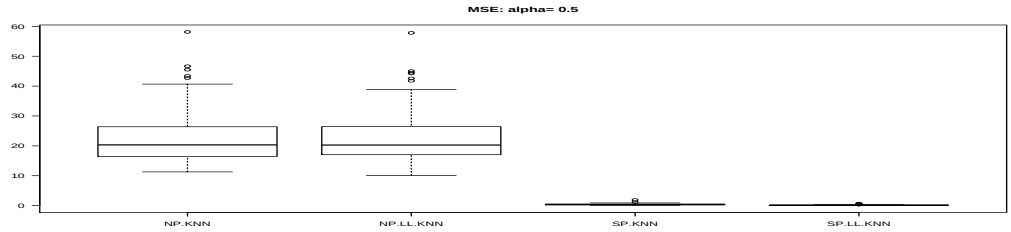
To summarize the simulation study, Figure 2 presents the boxplots of the MSEP values for the four regression models evaluated on the test sets using the optimal number of neighbors, k_{opt}. Table 2 reports the corresponding average prediction errors, $\overline{\text{MSEP}}$, also computed with k_{opt}. The figure illustrates the distribution and variability of prediction errors across replications, highlighting the relative performance of each method. As shown in Table 2, the naive estimator performs significantly worse than the other methods for both the FNP and SFPL models. Moreover, the predictive performance of the kNN-LL estimator consistently outperforms that of the kNN-kernel estimator.

Table 2. $\overline{MSEP}$ (mean squared error) for the six models, computed using k_{opt}.

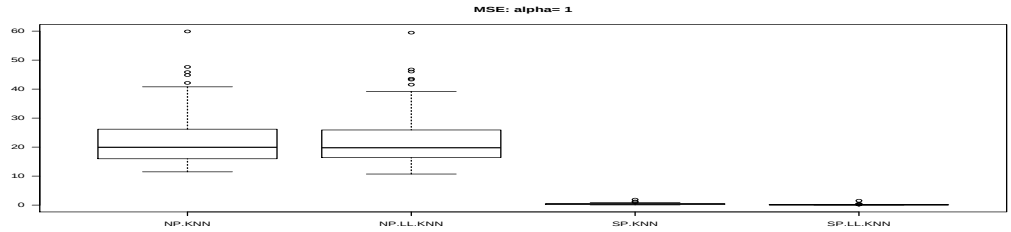
Model	α	Missing Rate	kNN	kNN-LLE	Naive
	0,25	30%	23.9632	22.7319	24.9931
FNP	0,5	16%	22.8861	22.6411	24.0969
$(Y = m(\xi) + \epsilon)$	1	05%	22.6256	22.5140	23.9153
	Complete	00%	22.4817	22.4472	23.0058
	0,25	30%	0.4195	0.3845	0.6453
SFPLR	0,5	16%	0.4044	0.1675	0.6385
$(Y = m(\xi) + \sum_{j=1}^{2} X_j\beta_j + \epsilon)$	1	05%	0.3918	0.1319	0.5961
	Complete	00%	0.3855	0.1053	0.4819



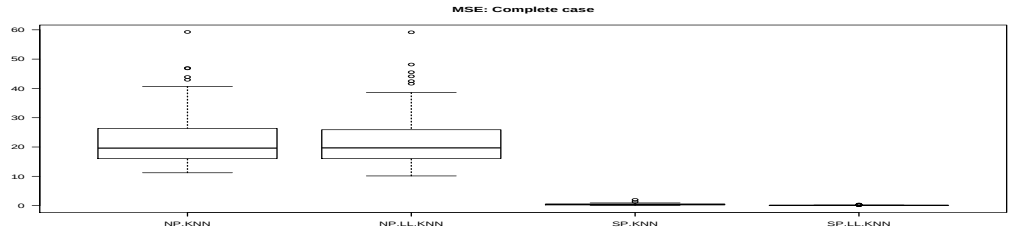
(a) $\alpha = 0.25$.



(b) $\alpha = 0.5$.



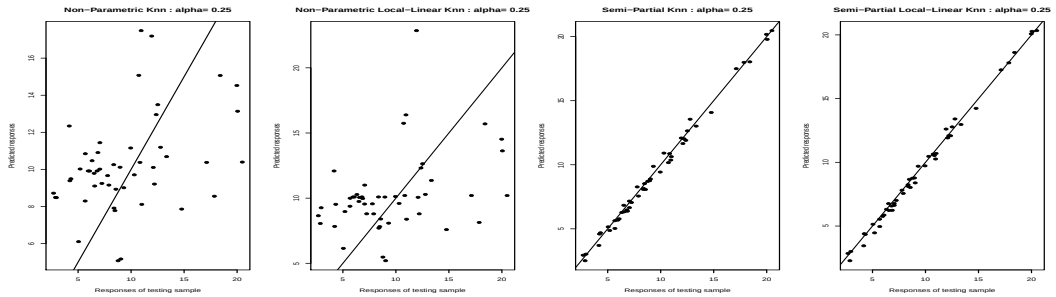
(c) $\alpha = 1.0$.



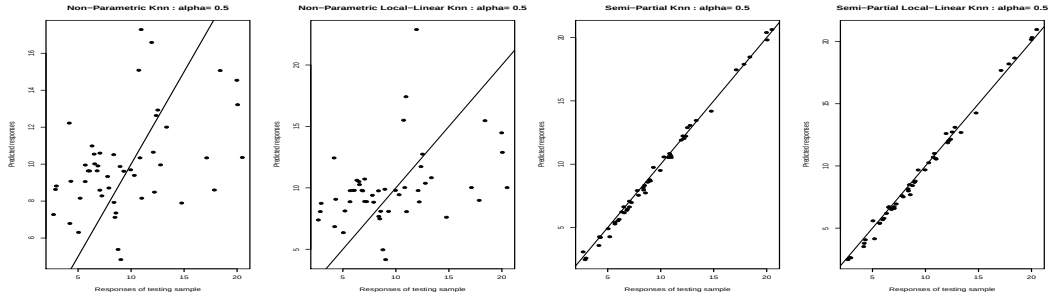
(d) Complete case.

Figure 2. Boxplots of the prediction MSE of componentwise prediction values by the two methods without and with missing data.

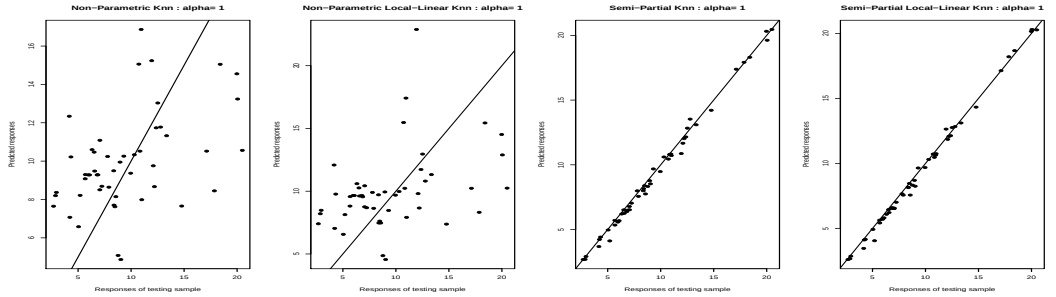
Figure 3 presents the prediction results, plotting the predicted values against the true values for both models, using the four estimation methods. This figure offers insight into the accuracy of the predictions from a single run.



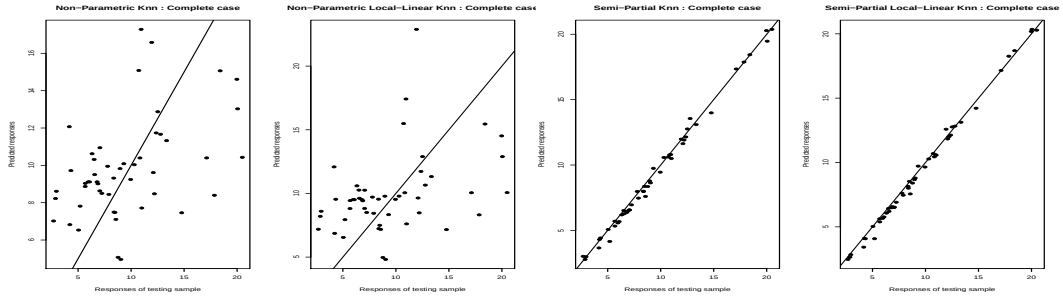
(a) $\alpha = 0.25$.



(b) $\alpha = 0.5$.



(c) $\alpha = 1.0$.



(d) Complete case.

Figure 3. Prediction of the testing simple by the four methods without and with missing data.

It's clear the SFPLR model offers superior predictive performance compared to the FNR model. This is strongly supported by the Mean Squared Error of Prediction (MSEP). As Figure 2 illustrates, the MSEP comparison across the four estimation methods confirms the SFPLR model's better forecasting accuracy. This finding is also reinforced by Figure 3 and Table 2. Furthermore, when applied to SFPLR models, the kNN-LLE approach outperforms other methods.

Table 3 presents the convergence behavior of the estimated coefficients $\widehat{\beta}_n$ for the two models (kNN and kNN-LLE) under different scenarios, considering varying sample sizes n and different levels of missing data (α).

Table 3. Convergence of $\widehat{\beta}_n$.

n	α	Missing Rate	kNN $(\widehat{\beta}_{1,n}, \widehat{\beta}_{2,n})$	kNN-LLE $(\widehat{\beta}_{1,n}, \widehat{\beta}_{2,n})$
	0,25	30%	(2.0906 , 1.0861)	(2.0272 , 1.0288)
150	0,5	16%	(1.9697 , 0.9792)	(2.0226 , 1.0239)
	1	05%	(1.9801 , 0.9899)	(2.0210 , 1.0185)
	Complete	00%	(2.0135 , 0.9916)	(2.0200 , 1.0151)
	0,25	30%	(2.0510 , 1.0569)	(2.0210 , 1.0216)
250	0,5	16%	(2.0187 , 1.0180)	(2.0201 , 1.0163)
	1	05%	(1.9871 , 1.0100)	(2.0197 , 1.0162)
	Complete	00%	(2.0066 , 0.9970)	(2.0111 , 1.0147)
	0,25	30%	(1.9674 , 1.0383)	(2.0206 , 1.0185)
550	0,5	16%	(1.9862 , 0.9880)	(2.0119 , 1.0157)
	1	05%	(2.0044 , 0.9942)	(2.0074 , 1.0147)
	Complete	00%	(2.0024 , 0.9979)	(2.0069 , 1.0119)

From Table 3, it is evident that the kNN-LLE estimator yields superior performance compared to the standard kNN estimator, particularly in imputation tasks involving missing data. While an increase in sample size correlates with improved estimation accuracy, the pervasive effect of missing data remains discernible. These observations underscore the imperative of deploying robust estimation methodologies, such as kNN-LLE, in analyses affected by substantial data incompleteness.

4.2. Application to real data

We aim to compare our SFPLR-kNN-LLE estimators with other established estimators for SFPLR models, as well as with those commonly used in nonparametric functional regression (FNR). The FNR estimators include the kNN estimator, the local linear estimator (LLE), and the kernel estimator. This comparison aligns with the objectives outlined in the simulation section.

Sugar Quality Assessment Dataset: Our analysis uses a real dataset related to sugar quality assessment through fluorescence data. Sugar quality is typically evaluated based on two key parameters: ash content and color. Ash content, expressed as a percentage, indicates the amount of inorganic impurities in refined sugar, determined by its conductivity. Color is measured by the absorption at 420 nm in a membrane-filtered sugar solution adjusted to pH=7. Its values are derived from absorbance units, with 45 being the maximum allowable value for standard sugar.

The primary goal of this study is to investigate the relationship between the color indexation and

both ash content and Near-Infrared Spectroscopy (NIR) curves. The NIR curves represent emission spectra, measured between 275*nm* and 560*nm* at 0.5*nm* intervals (571 wavelengths), for four excitation wavelengths (290*nm*, 305*nm*, 325*nm* and 340*nm*). This dataset is publicly available at: `https://ucphchemometrics.com/sugar-process-data/`.

The dataset contains missing observations, which were replaced by NaN, resulting in an overall missing rate of 15.67%. Our analysis compares two distinct scenarios to account for this:

- Without missing observations: Data after imputation.
- With missing observations: The original dataset.

Figure 4 displays the observed curves for all 265 samples.

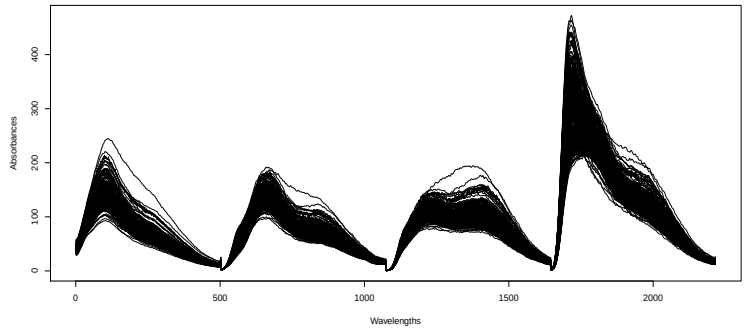


Figure 4. A sample of 265 curves ξ of NIR spectroscopy.

The primary goal of this computational study is to evaluate the efficiency of the SFPLR-kNN-LLE estimator in approximating the average color indexation.

To achieve this, we randomly divide the 265 independent and identically distributed (i.i.d.) observations into two subsets:

- Training set: 215 observations used for model estimation.
- Test set: The remaining observations used to evaluate the prediction quality.

We assume a semi-functional partial linear regression (SFPLR) model to establish the relationship between the variables:

$$Y = m(\xi) + X\beta + \epsilon,$$

where β and $m(\cdot)$ are unknown components modeling the relationship between X (ash content), ξ (near-infrared spectroscopy curves), and Y (color indexation).

To compute our estimators, we adopt the same methodological framework as in the simulation study. Specifically, we conduct $M = 100$ independent replications, generating M values of the mean squared error of prediction (MSEP). The distribution of these MSEP values is visualized using boxplots for both scenarios: With and without missing observations.

The MSEP results are visualized in Figures 5 and 6.

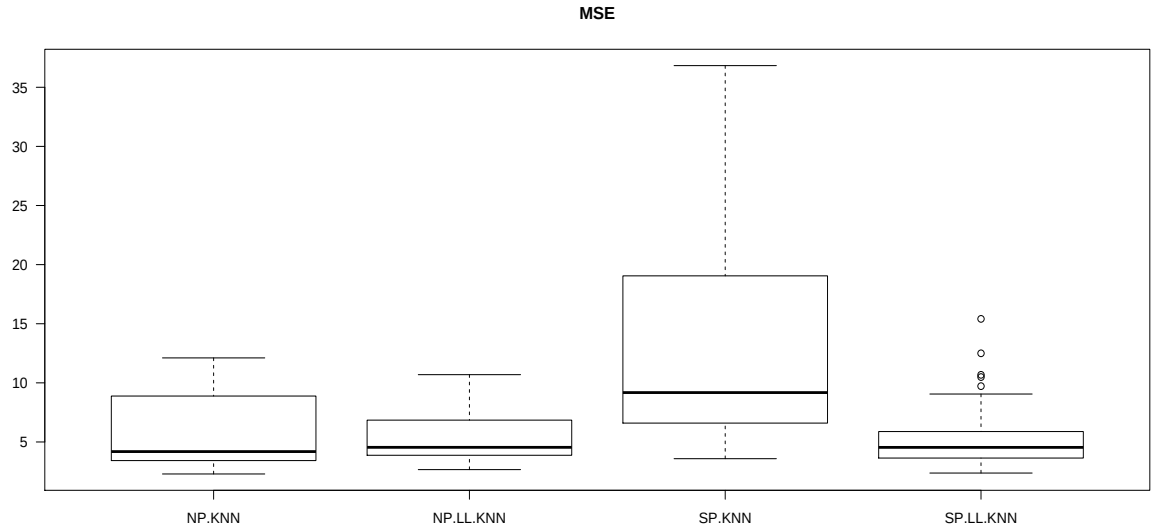


Figure 5. The $MSEP$ box plots with missing observations of the prediction values for both models with the 4 estimation methods.

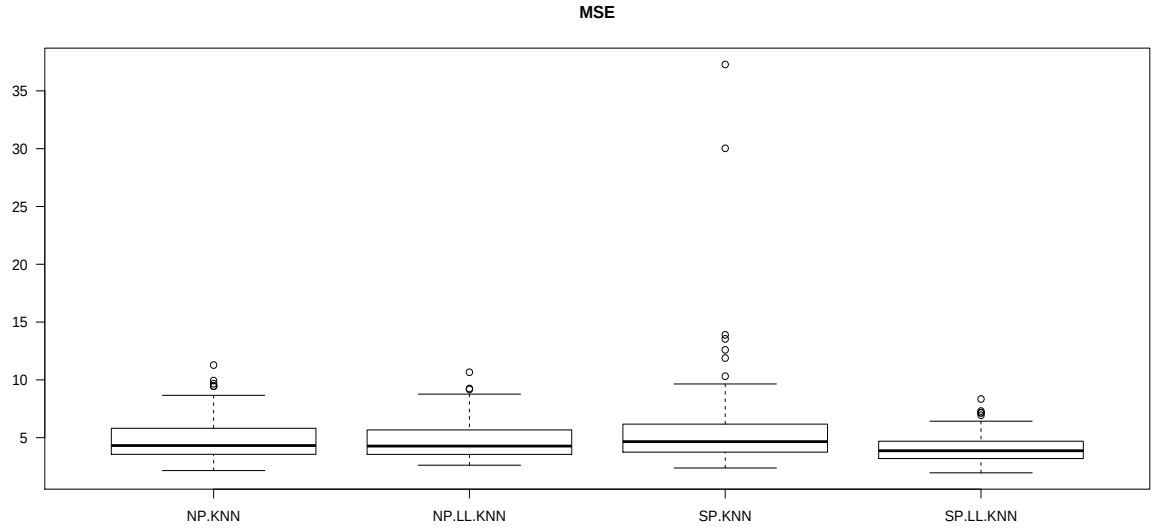


Figure 6. The $MSEP$ box plots without missing observations of the prediction values for both models with the 4 estimation methods.

The dataset for this study comprises near-infrared (NIR) spectroscopy measurements from an experiment designed to determine the composition of 265 sugar quality samples. Our main goal is to predict color indexation using ash content and the NIR curves, which are derived from light absorbance measurements at various wavelengths. To achieve this, we use the semi-functional partial linear regression (SFPLR) model as our predictive framework.

Figures 5 and 6 offer a detailed analysis of the mean squared error of prediction (MSEP) for several predictive models, including: FNP.kNN, FNP.kNN-LLE, SFPLR.kNN and SFPLR.kNN-LLE.

The results clearly show that SFPLR-based models deliver superior prediction accuracy compared to FNP models. Within the SFPLR framework, the kNN-LLE approach stands out as particularly effective. The accompanying scatter plot visually confirms the accuracy of these predictions, providing an intuitive assessment of model performance and reinforcing our findings.

5. Conclusions

This paper introduces the semi-functional partial linear regression (SFPLR) model for independent and identically distributed (i.i.d.) data. This model combines the strengths of partial linear regression with the flexibility of functional data analysis, explicitly accounting for missing at random (MAR) data in the response variable.

Our primary contribution is the development of novel estimators through the integration of local linear estimation (LLE) with the k-Nearest Neighbor (kNN) smoothing method. This hybrid approach effectively addresses the common problem of bandwidth selection in nonparametric estimation, yielding estimators with reduced bias. Beyond its solid theoretical properties, this estimator proves to be highly practical: It is fast, robust, and more accurate than competing alternatives. We establish the asymptotic distribution of the parametric component and the quasi-complete uniform consistency rates of the nonparametric component, relative to the number of neighbors, under appropriate conditions.

The advantages of the LLE-kNN approach are twofold:

- Improved bias component: While the proposed estimator's convergence rate is consistent with that of existing methods in the SFPLR framework, it significantly improves the bias component. Utilizing the local linear method not only reduces computational costs but also enhances implementation efficiency, leading to substantial gains in predictive performance.

- Efficient bandwidth selection: The integration of kNN smoothing offers an elegant and efficient solution to the complex, long-standing challenge of bandwidth selection in nonparametric statistics. One challenge remains: determining an appropriate rule for selecting the optimal smoothing parameter and identifying the relevant subset for optimization. Nevertheless, by reformulating the problem as selecting an integer $k \in \{1, \ldots, n\}$, the kNN approach simplifies this task while maintaining high performance.

Finally, we evaluate the finite-sample performance of the proposed estimators through simulations and real-world data analysis.The results clearly demonstrate that the LLE-kNN estimator outperforms its competitor. This superiority is evidenced by the lowest mean squared error (MSE) obtained on both simulated and real-world datasets.

Author contributions

A. Naceri and T. Benchikh: Formal analysis; F. Alshahrani and O. Fetitah: Validation; I. M. Almanjahie and M. Kadi Attouch: Writing – review & editing. The authors contributed approximately equally to this work. All authors have read and agreed to the final version of the manuscript.

Use of Generative-AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

Acknowledgments

The authors thank and extend their appreciation to the funders of this work. This research was funded by Princess Nourah bint Abdulrahman University Researchers Supporting Project number (PNURSP2025R358), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia; and the Deanship of Scientific Research and Graduate Studies at King Khalid University through the Research Groups Program under grant number RGP2/456/46.

Data availability

The data used in this study are available through the link `https://ucphchemometrics.com/sugar-process-data/`.

Conflict of interest

The authors declare no conflict of interest.

References

1. J. Ramsay, B. Silverman, *Functional data analysis*, 2 Eds., Spinger-Verlag, New York, 2005.
2. P. Kokoszka, M. Reimherr, *Introduction to functional data analysis*, 1st Edition, eBook, Chapman and Hall, 2017. https://doi.org/10.1201/9781315117416
3. T. Hsing, R. L. Eubank, *Theoretical foundations of functional data analysis, with an introduction to linear operators*, John Wiley and Sons, 2015.
4. G. Aneiros-Pérez, R. Cao, R. Fraiman, C. Genest, P. Vieu, Recent advances in functional data analysis and high-dimensional statistics, *J. Multivariate Anal.*, **170** (2019), 3–9. https://doi.org/10.1016/j.jmva.2018.11.007
5. G. Aneiros-Pérez, I. Horová, M. Hušková, P. Vieu, Editorial for the special issue on functional data analysis and related fields, *J. Multivariate Anal.*, **189** (2022). https://doi.org/10.1016/j.jmva.2021.104861
6. M. Rachdi, Functional data analysis: Theory and applications to different scenarios, *Mathematics*, an Open Access Journal by MDPI 2023. Avaiable from: `https://www.mdpi.com/journal/mathematics/specialissues/45POZ9BG9S`.

7. F. Ferraty, F. Vieu, *Nonparametric functional data analysis. Theory and Practice*, Springer Series in Statistics, New York, 2006.
8. A. Cuevas, A partial overview of the theory of statistics with functional data, *J. Statist. Plann. Inference*, **147** (2014), 1–23. https://doi.org/10.1016/j.jspi.2013.04.002
9. S. Greven, F. Scheipl, A general framework for functional regression modelling, *Stat. Model.*, **17** (2017), 1–35. https://doi.org/10.1177/1471082X16681317
10. G. Aneiros-Pérez, P. Vieu, Semi-functional partial linear regression, *Stat. Probab. Lett.*, **76** (2006), 1102–1110. https://doi.org/10.1016/j.spl.2005.12.007
11. G. Boente, A. Vahnovan, Robust estimators in semi-functional partial linear regression models, *J. Multivariate Anal.*, **154** (2017), 59–84. https://doi.org/10.1016/j.jmva.2016.10.005
12. S. Feng, L. Xue, Partially functional linear varying coefficient model, *Statistics*, **50** (2016), 717–732. https://doi.org/10.1080/02331888.2016.1138954
13. N. Ling, G. Aneiros-Pérez, P. Vieu, kNN estimation in functional partial linear modeling, *Statist. Papers*, **61** (2020), 423–444. https://doi.org/10.1007/s00362-017-0946-0
14. G. Aneiros-Pérez, P. Vieu, Nonparametric time series prediction: A semi-functional partial linear modeling, *J. Multivariate Anal.*, **99** (2008), 834–857. https://doi.org/10.1016/j.jmva.2007.04.010
15. M. Benallou, M. K. Attouch, T. Benchikh, O. Fetitah, Asymptotic results of semi-functional partial linear regression estimate under functional spatial dependency, *Commun. Stat.-Theor. Method.*, **51** (2021), 1–21. https://doi.org/10.1080/03610926.2020.1871021
16. N. Ling, P. Vieu, Nonparametric modelling for functional data: Selected survey and tracks for future, *Statistics*, **52** (2018), 934–949. https://doi.org/10.1080/02331888.2018.1487120
17. N. Ling, P. Vieu, On semiparametric regression in functional data analysis, *WIRES Comput. Stat.*, **12** (2020), 20–30. https://doi.org/10.1002/wics.1538
18. S. Novo, G. Aneiros-Pérez, P. Vieu, A kNN procedure in semiparametric functional data analysis, *Statist. Probab. Lett.*, **171** (2021). https://doi.org/10.1016/j.spl.2020.109028
19. N. H. Kedir, T. Benchikh, A. Naceri, O. Fetitah, Local linear-kNN smoothing for semi-functional partial linear regression, *Hacet. J. Math. Stat.*, **53** (2024), 537–555. https://doi.org/10.15672/hujms.1294382
20. M. K. Attouch, A. Laksaci, F. Rafaa, Estimation locale linéaire de la régression non paramétrique fonctionnelle par la méthode des k plus proches voisins, *CR Math.*, **355** (2017), 824–829. https://doi.org/10.1016/j.crma.2017.05.007
21. I. M. Almanjahie, Z. Chikr-Elmezouar, A. Laksaci, M. Rachdi, kNN local linear estimation of the conditional cumulative distribution function: Dependent functional data case, *C. R. Acad. Sci. Paris, Ser. I*, **356** (2018), 1036–1039. https://doi.org/10.1016/j.crma.2018.09.001
22. Z. Chikr-Elmezouar, I. M. Almanjahie, A. Laksaci, M. Rachdi, FDA: Strong consistency of the kNN local linear estimation of the functional conditional density and mode, *J. Nonparametr. Stat.*, **31** (2019), 175–195. https://doi.org/10.1080/10485252.2018.1538450
23. I. M. Almanjahie, W. Mesfer, A. Laksaci, M. Rachdi, Computational aspects of the kNN local linear smoothing for some conditional models in high dimensional statistics, *Commun. Stat. Simul. Comput.*, **52** (2023), 2985–3005. https://doi.org/10.1080/03610918.2021.1923745

24. F. Alshahrani, W. Bouabsa, I. M. Almanjahie, M. K. Attouch, kNN local linear estimation of the conditional density and mode for functional spatial high dimensional data, *AIMS Math.*, **8** (2023), 15844–15875. https://doi.org/10.3934/math.2023809

25. F. Ferraty, F. Sued, P. Vieu, Mean estimation with data missing at random for functional covariables, *Statistics*, **47** (2013), 688–706. https://doi.org/10.1080/02331888.2011.650172

26. N. Ling, L. Liang, P. Vieu, Nonparametric regression estimation for functional stationary ergodic data with missing at random, *J. Stat. Plan. Inference*, **162** (2015), 75–87. https://doi.org/10.1016/j.jspi.2015.02.001

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

28. F. Alshahrani, I. M. Almanjahi, T. Benchikh, O. Fetitah, M. K. Attouch, Asymptotic normality of nonparametric kernel regression estimation for missing at random functional spatial data, *J. Math.*, **2** (2023), 1–20. https://doi.org/10.1155/2023/8874880.4/math.2023809

29. M. Rachdi, A. Laksaci, K. Kaid, A. Benchiha, F. Al-Awadh, k-Nearest neighbors local linear regression for functional and missing data at random, *Stat. Neerl.*, **75** (2021), 42–65. https://doi.org/10.1111/stan.12224

30. I. M. Almanjahie, W. M. Alahmari, A. Laksaci, The k nearest neighbors local linear estimator of functional conditional density when there are missing data, *Hacet. J. Math. Stat.*, **51** (2022), 914–931. https://doi.org/10.15672/hujms.796694

31. N. Ling, R. Kan, P. Vieu, S. Meng, Semi-functional partially linear regression model with responses missing at random, *Metrika*, **82** (2019), 39–70. https://doi.org/10.1007/s00184-018-0688-6

32. T. Benchikh, I. M. Almanjahie, O. Fetitah, M. K. Attouch, Estimation for spatial semi-functional partial linear regression model with missing response at random, *Demonstr. Math.*, **58** (2025), https://doi.org/10.1515/dema-2025-0108

33. A. Naceri, A. Laksaci, M. Rachdi, *Exact quadratic error of the local linear regression operator estimator for functional covariates*, In : Functional statistics and applications, Springer Cham Heidelberg, New York, 2015, 79–90. https://doi.org/10.1007/978-3-319-22476-3_5

34. N. Kudraszow, P. Vieu, Uniform consistency of kNN regressors for functional variables, *Statist. Probab. Lett.*, **83** (2013), 1863–1870. https://doi.org/10.1016/j.spl.2013.04.017

35. F. Ferraty, A. Mas, P. Vieu, Nonparametric regression on functional data: Inference and practical aspects, *Aust. N. Z. J. Stat.*, **49** (2007), 267–286. https://doi.org/10.1111/j.1467-842X.2007.00480.x

36. J. Barrientos-Marin, F. Ferraty, P. Vieu, Locally modelled regression and functional data, *J. Nonparametr Stat.*, **22** (2010), 617–632. https://doi.org/10.1080/10485250903089930

37. L. Kara-Zaitri, A. Laksaci, M. Rachdi, P. Vieu, Data-driven kNN estimation in nonparametric functional data analysis, *J. Multivariate Anal.*, **153** (2017), 176–188. https://doi.org/10.1016/j.jmva.2016.09.016

38. F. Ferraty, A. Laksaci, A. Tadj, P. Vieu, Rates of uniform consistency for nonparametric estimates with functional variables, *J. Stat. Plan. Infer.*, **140** (2010), 335–352. https://doi.org/10.1016/j.jspi.2009.07.019

39. F. Burba, F. Ferraty, P. Vieu, k-Nearest Neighbour method in functional nonparametric regression, *J. Nonparametr. Stat.*, **21** (2009), 453–469. https://doi.org/10.1080/10485250802668909

40. M. Rachdi, P. Vieu, Nonparametric regression for functional data: Automatic smoothing parameter selection, *J. Statist. Plann. Inference*, **137** (2007), 2784–2801. https://doi.org/10.1016/j.jspi.2006.10.001

41. G. Aneiros-Pérez, P. Vieu, Automatic estimation procedure in partial linear model with functional data, *Stat. Papers*, **52** (2011), 751–771. https://doi.org/10.1007/s00362-009-0280-2

42. H. L. Shang, Bayesian bandwidth estimation for a semi-functional partial linear regression model with unknown error density, *Comput Stat.*, **29** (2014), 829–848. https://doi.org/10.1007/s00180-013-0463-0

43. G. Chagny, A. Roche, Adaptive estimation in the functional nonparametric regression model, *J. Multivariate Anal.*, **146** (2016), 105–118. https://doi.org/10.1016/j.jmva.2015.07.001

44. L. Kara-Zaitri, A. Laksaci, M. Rachdi, P. Vieu, Uniform in bandwidth consistency for various kernel estimators involving functional data, *J. Nonparametr. Stat.*, **29** (2017), 85–107. https://doi.org/10.1080/10485252.2016.1254780

Appendix

The proofs of the asymptotic results follow the same ideas as in [31], so they are given briefly. They are based on the following intermediate results.

Lemma 5.1. *Under assumptions (A1)–(A10), we have*

$$\sup_{k_{1,n}\leq k\leq k_{2,n}}\sup_{\xi\in\mathcal{S}}\left|m^s_{1,l}(\xi)-\sum_{j=1}^{n}\mathbf{W_j}(\xi)X_{js}\right|=O\left(\phi^{-1}\left(\frac{k_{2,n}}{n}\right)^{\alpha}\right)+O_{a.co.}\left(\sqrt{\frac{\log d_n}{k_{1,n}}}\right). \tag{5.1}$$

$$\sup_{k_{1,n}\leq k\leq k_{2,n}}\sup_{\xi\in\mathcal{S}}\left|m_2(\xi)-\sum_{j=1}^{n}\mathbf{W_j}(\xi)Y_{j}\right|=O\left(\phi^{-1}\left(\frac{k_{2,n}}{n}\right)^{\alpha}\right)+O_{a.co.}\left(\sqrt{\frac{\log d_n}{k_{1,n}}}\right). \tag{5.2}$$

Remark 5.1. *The Lemma 5.1 extends Theorem 2 of [29] in the case where the functional space is a semi-metric space and Theorem 2.1 of [23] with MAR data. The proof of this result is a combination of the same demonstration techniques used in these references. It is not given here.*

Lemma 5.2. *Under the conditions (A1)–(A10), we have*

$$\frac{1}{n}\sum_{l=1}^{n}\delta_l\widetilde{\mathbf{X}}_l^T\widetilde{\mathbf{X}}_l\to\mathbf{\Sigma}\quad a.s. \tag{5.3}$$

Proof. Denote $\widetilde{\mathbf{X}}_{ls}=X_{ls}-\sum_{j=1}^{n}\mathbf{W}_j(\xi_l)X_{js}=\eta_{ls}-g_{1s}(\xi_l)-\sum_{j=1}^{n}\mathbf{W}_j(\xi_l)X_{js}$ $(l=1,\ldots,n,s=1,\ldots,p)$. Then the (r,s)th element of $\frac{1}{n}\widetilde{\mathbf{X}}^T\widetilde{\mathbf{X}}$ can be written as

$$
\begin{aligned}
n^{-1}\sum_{l=1}^{n}\delta_l\widetilde{X}_{lr}^{T}\widetilde{X}_{ls} &= n^{-1}[\sum_{l=1}^{n}\delta_l\eta_{lr}\eta_{ls}+\sum_{l=1}^{n}\delta_l\eta_{lr}\left(g_{1s}(\xi_l)-\sum_{j=1}^{n}\mathbf{W}_j(\xi_l)X_{js}\right)\\
&\quad+\sum_{l=1}^{n}\delta_l\eta_{ls}\left(g_{1r}(\xi_l)-\sum_{j=1}^{n}\mathbf{W}_j(\xi_l)X_{jr}\right)\\
&\quad+\sum_{l=1}^{n}\delta_l\left(g_{1r}(\xi_l)-\sum_{j=1}^{n}\mathbf{W}_j(\xi_l)X_{jr}\right)\\
&\quad\times\left(g_{1s}(\xi_l)-\sum_{j=1}^{n}\mathbf{W}_j(\xi_l)X_{js}\right)].
\end{aligned}
\tag{5.4}
$$

Thus, using the strong law of large numbers for i.i.d. variables, we get, as $n\to\infty$,

$$
n^{-1}\sum_{l=1}^{n}\delta_l\eta_{lr}\eta_{ls}\to\Sigma_{rs}\quad a.s. \tag{5.5}
$$

Furthermore, by applying directly the Lemma 5.1 and using again the strong law of large numbers for i.i.d. variables, we can see that

$$
n^{-1}\sum_{l=1}^{n}\delta_l\eta_{lr}\left(g_{1s}(\xi_l)-\sum_{j=1}^{n}\mathbf{W}_j(\xi_l)X_{js}\right)\to 0\quad a.s. \tag{5.6}
$$

$$
n^{-1}\sum_{l=1}^{n}\delta_l\eta_{ls}\left(g_{1r}(\xi_l)-\sum_{j=1}^{n}\mathbf{W}_j(\xi_l)X_{jr}\right)\to 0\quad a.s. \tag{5.7}
$$

and

$$
n^{-1}\sum_{l=1}^{n}\delta_l\left(g_{1r}(\xi_l)-\sum_{j=1}^{n}\mathbf{W}_j(\xi_l)X_{jr}\right)\times\left(g_{1s}(\xi_l)-\sum_{j=1}^{n}\mathbf{W}_j(\xi_l)X_{js}\right)\to 0\quad a.s. \tag{5.8}
$$

Finally, we conclude the proof by using results of (5.4)–(5.8). □

Proof of Theorem 3.1.

Similar to [13, 31], we can write

$$
\begin{aligned}
\sqrt{n}\left(\widehat{\boldsymbol{\beta}}_n-\boldsymbol{\beta}\right) &= \left(n^{-1}\widetilde{\mathbf{X}}^{T}\widetilde{\mathbf{X}}\right)^{-1}\frac{1}{\sqrt{n}}\left\{\sum_{l=1}^{n}\widetilde{\mathbf{X}}_l\overline{m}_n(\xi_l)-\sum_{l=1}^{n}\widetilde{\mathbf{X}}_l\left(\sum_{j=1}^{n}\mathbf{W}_j(\xi_l)\epsilon_j\right)+\sum_{l=1}^{n}\widetilde{\mathbf{X}}_l\epsilon_l\right\}\\
&= \left(n^{-1}\widetilde{\mathbf{X}}^{T}\widetilde{\mathbf{X}}\right)^{-1}\frac{1}{\sqrt{n}}(L_{n1}-L_{n2}+L_{n3}),
\end{aligned}
\tag{5.9}
$$

where $\overline{m}(\xi_l)=m(\xi_l)-\sum_{j=1}^{n}\mathbf{W}_j(\xi_l)m(\xi_j)$.

As, the sth element of $\widetilde{\mathbf{X}}_l$ is written as

$$
\begin{aligned}
\widetilde{X}_{ls} &= \eta_{ls}-\sum_{j=1}^{n}\mathbf{W}_j(\xi_l)\eta_{js}+g_{1s}(\xi_l)-\sum_{j=1}^{n}\mathbf{W}_j(\xi_l)g_{1s}(\xi_j),\\
&= \eta_{ls}-\sum_{j=1}^{n}\mathbf{W}_j(\xi_l)\eta_{js}+\widetilde{m}_{ls}\quad(l=1,\ldots,n,s=1,\ldots,p).
\end{aligned}
\tag{5.10}
$$

Then each element of the vectors L_{n1}, L_{n2} and L_{n3} can be decomposed into three summands, noted $L_{nq,1}$, $L_{nq,2}$ and $L_{nq,3}$ for $q=1,2,3$, whose asymptotic behavior can be obtained from the Lemmas 5.1 and 5.2, Lemma 3 in [10] and Lemma A.3 in [19].

More specifically, by Lemma 5.1 we have

$$| \overline{m}(\xi_l) |= O\left(\phi^{-1}\left(\frac{k_{2,n}}{n}\right)^{\alpha} + \frac{\log n}{\sqrt{k_{1,n}-1}}\right) + O_{a.co.}\left(\sqrt{\frac{\log d_n}{k_{1,n}}}\right), \tag{5.11}$$

and

$$| \widetilde{m}_{ls} |= O\left(\phi^{-1}\left(\frac{k_{2,n}}{n}\right)^{\alpha} + \frac{\log n}{\sqrt{k_{1,n}-1}}\right) + O_{a.co.}\left(\sqrt{\frac{\log d_n}{k_{1,n}}}\right). \tag{5.12}$$

It follows from (5.11), (5.12), Lemma 3 in [10], Lemma A.3 in [19] and Abel's inequality that

$$L_{n1,1} = O\left(n\left(\left(\phi^{-1}\left(\frac{k_{2,n}}{n}\right)\right)^{2\alpha} + \frac{\log n^2}{k_{1,n}-1}\right)\right) + O_{a.co.}\left(n\frac{\log d_n}{k_{1,n}}\right) = o(\sqrt{n}) \quad a.s. \tag{5.13}$$

$$L_{n1,2} = O\left[\left(\phi^{-1}\left(\frac{k_{2,n}}{n}\right)^{\alpha} + + \sqrt{\frac{\log d_n}{k_{1,n}}} + \frac{\log n}{\sqrt{k_{1,n}-1}}\right)\sqrt{n}\log n\right] = o(\sqrt{n}) \quad a.s.,$$

$$L_{n1,3} = O\left[\left(\phi^{-1}\left(\frac{k_{2,n}}{n}\right)^{\alpha} + \sqrt{\frac{\log d_n}{k_{1,n}}} + \frac{\log n}{\sqrt{k_{1,n}-1}}\right)n\frac{\log n}{\sqrt{k_{1,n}-1}}\right] = o(\sqrt{n}) \quad a.s.$$

In the same way, we have

$$L_{n2,1} = O\left(\left(\phi^{-1}\left(\frac{k_{2,n}}{n}\right)^{\alpha} + \sqrt{\frac{\log d_n}{k_{1,n}}} + \frac{\log n}{\sqrt{k_{1,n}-1}}\right)\frac{n\log n}{\sqrt{k_{1,n}-1}}\right) = o(\sqrt{n}) \quad a.s.,$$

$$L_{n2,2} = = O_{a.co.}\left(\frac{\sqrt{n}\log^2 n}{\sqrt{k_{1,n}-1}}\right) = o(\sqrt{n}) \quad a.s. \tag{5.14}$$

$$L_{n2,3} = O\left(\frac{n\log^2 n}{k_{1,n}-1}\right) = o(\sqrt{n}) \quad a.s.$$

Finally, for L_{n3}, we have

$$L_{n3,1} = O\left[\left(\phi^{-1}\left(\frac{k_{2,n}}{n}\right)^{\alpha} + \sqrt{\frac{\log d_n}{k_{1,n}}} + \frac{\log n}{\sqrt{k_{1,n}-1}}\right)\frac{\sqrt{n}\log n}{\sqrt{k_{1,n}-1}}\right] = o(\sqrt{n}) \quad a.s.$$

$$L_{n3,2} = O\left(\frac{\sqrt{n}\log n}{\sqrt{k_{1,n}-1}}\right) = o(\sqrt{n}) \quad a.s.,$$

$$L_{n3,3} = \sum_{l=1}^{n}\eta_l\epsilon_l\,.$$

Then, by (5.9) and (5.13)–(5.15), it follows that

$$\sqrt{n}(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}) = \left(n^{-1}\widetilde{\mathbf{X}}^T\widetilde{\mathbf{X}}\right)^{-1}\frac{1}{\sqrt{n}}\left(\sum_{l=1}^{n}\eta_l\epsilon_l+o(\sqrt{n})\right). \tag{5.15}$$

Therefore, applying Slutsky' lemma and central limit theorem with the Lemma 5.2, the proof of Theorem 3.1 is concluded.

Proof of Theorem 3.2.

From the fact that

$$\begin{aligned}\widehat{m}(\xi) &= \textstyle\sum_{j=1}^{n}\mathbf{W}_j(\xi)Y_j-\sum_{j\in\mathcal{I}_{\mathbf{n}}}\mathbf{W}_j(\xi)X_j^T\widetilde{\boldsymbol{\beta}}_n\\ &= \textstyle\sum_{j=1}^{n}\mathbf{W}_j(\xi)\left(m(\xi_j)+\varepsilon_j\right)-\sum_{j=1}^{n}\mathbf{W}_j(\xi)\mathbf{X}_j^T\left(\widetilde{\boldsymbol{\beta}}_n-\boldsymbol{\beta}_n\right),\end{aligned}$$

we can deduce that

$$\begin{aligned}\sup_{k\in]k_{1,n},k_{2,n}[}\sup_{\xi\in\mathcal{S}}|\widehat{m}(\xi)-m(\xi)| &\leq \sup_{k\in]k_{1,n},k_{2,n}[}\sup_{\xi\in\mathcal{S}}|\textstyle\sum_{j=1}^{n}\mathbf{W}_j(\xi)\left(m(\xi_j)+\varepsilon_j\right)-m(\xi)|\\ &\quad+\sup_{k\in]k_{1,n},k_{2,n}[}\sup_{\xi\in\mathcal{S}}|\textstyle\sum_{j=1}^{n}\mathbf{W}_j(\xi)\mathbf{X}_j^T|\|\widehat{\beta}_n-\beta)\|\\ &\leq \sup_{k\in]k_{1,n},k_{2,n}[}\sup_{\xi\in\mathcal{S}}|S_1|+\sup_{k\in]k_{1,n},k_{2,n}[}\sup_{\xi\in\mathcal{S}}|S_2|.\end{aligned} \tag{5.16}$$

On the one hand, from Lemma 5.1, we have

$$\sup_{k\in]k_{1,n},k_{2,n}[}\sup_{\xi\in\mathcal{S}}|S_1|=O\left(\phi^{-1}\left(\frac{k_{2,n}}{n}\right)\right)+O_{a.co}\left(\sqrt{\frac{\log d_n}{k_{1,n}}}\right). \tag{5.17}$$

In other hand,

$$\begin{aligned}\sup_{k\in]k_{1,n},k_{2,n}[}\sup_{\xi\in\mathcal{S}}|S_2| &\leq \sup_{k\in]k_{1,n},k_{2,n}[}\sup_{\xi\in\mathcal{S}}|\textstyle\sum_{j=1}^{n}\mathbf{W}_j(\xi_l)\mathbf{X}_j|\|\widehat{\beta}_n-\beta\|\\ &\leq \sup_{k\in]k_{1,n},k_{2,n}[}\sup_{\xi\in\mathcal{S}}|\textstyle\sum_{j=1}^{n}\mathbf{W}_j(\xi_l)\left(\mathbf{X}_j\right)-\mathbb{E}\left(\mathbf{X}_l/\xi_l\right)|\,|\widehat{\beta}_{\mathbf{n}}-\beta|+\\ &\quad\sup_{\xi\in\mathcal{S}}|\mathbb{E}(\mathbf{X}_l/\xi_l)|\,\|\widehat{\beta}_{\mathbf{n}}-\beta\|.\end{aligned}$$

Under Theorem 3.1(ii), we have $\|\widehat{\beta}_n-\beta\|\to 0$ and according to the fact that $\sup_{\xi\in\mathcal{S}}|\mathbb{E}(\mathbf{X}_l/\xi_l)|<\infty$, 5.1 implies that

$$\sup_{k\in]k_{1,n},k_{2,n}[}\sup_{\xi\in\mathcal{S}}|S_2|=O\left(\phi^{-1}\left(\frac{k_{2,n}}{\widehat{\mathbf{n}}}\right)\right)+O_{a.co}\left(\sqrt{\frac{\log d_{\mathbf{n}}}{k_{1,n}}}\right). \tag{5.18}$$

So by using Eqs (5.16)–(5.18), we have

$$\sup_{k\in]k_{1,n},k_{2,n}[}\sup_{\xi\in\mathcal{S}}|\widehat{m}(\xi)-m(\xi)|=O\left(\phi^{-1}\left(\frac{k_{2,n}}{\widehat{\mathbf{n}}}\right)\right)+O_{a.co}\left(\sqrt{\frac{\log d_{\mathbf{n}}}{k_{1,n}}}\right).$$