Hacettepe Journal of
Mathematics & Statistics

Hacet. J. Math. Stat.
Volume 53 (2) (2024), 537 – 555
DOI : 10.15672/hujms.1294382

Research Article

Local linear-kNN smoothing for semi-functional partial linear regression

Kedir Nassima Houda[1], Benchikh Tawfik*[1,2], Naceri Amina[1,3]
Fetitah Omar[1,3]

[1] *Statistical and Stochastic Process Laboratory, Djillali Liabes University, Sidi Bel Abbes, Algeria*
[2] *Fauclty of Medecine, Djillali Liabes University, Sidi Bel Abbes, Algeria*
[3] *School of Computer Science of Sidi Bel Abbes (ESI-SBA), Sidi Bel Abbes, Algeria*

Abstract

The aim of this paper is to study a semi-functional partial linear regression model. The estimators are constructed by k-nearest neighbors local linear method. Some asymptotic results are established for an i.i.d sample under certain conditions, including asymptotic normality of the parametric component and the almost certain convergence (with rate) of the non-parametric component. Lastly, using cross-validation, the performances of our estimation method are presented on simulated data and on real data by comparing them with other known approaches for semi-functional partial linear regression models.

Mathematics Subject Classification (2020). 60G25, 62G05, 62G08,62G20,62F12

Keywords. Functional data analysis, partial linear regression, kNN estimation, local linear estimation, asymptotic normality

1. Introduction

The study of the influence of a functional random variable on a scalar variable, called functional regression, has become a major subject in functional data analysis (FDA) both by the methods used and by the diversification of application areas such as this is evidenced by various works on this subject [16,23,24,36,38] as well as through various recent bibliographical discussions such as [3]. In particular, among the themes studied on this subject, we find semi-parametric functional regression models which have the properties of retaining the flexibility of parametric regression models and overcoming the sensitivity to dimensional effects of non-parametric approaches (for a complete discussion with a state of the art on semi-parametric modeling by functional regression, we refer you to the bibliographical surveys of [25,34]).

One of the most important semi-parametric models is the partially linear regression model introduced by [7], which is called semi-functional partial linear regression (SPFLR) model. This model is expressed as:

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

*Corresponding Author.
Email addresses: houda.kedir@univ-sba.dz (K.N. Houda), benchikh.tawfik@gmail.com (B. Tawfik), am.naceri@esi-sba.dz (N. Amina), o.fetitah@esi-sba.dz (F. Omar)
Received: 09.05.2023; Accepted: 19.12.2023

where Y is the scalar response variable, $\mathbf{X} = (X_1, X_2, \ldots, X_p)$ is a p-vector of explanatory variables, ξ is functional explanatory variable, β is unknown p-dimensional parameter vector, $m(.)$ is an unknown smooth functional operator and ϵ is the centered random error with finite unknown variance. We generally assume an additional condition of independence of the variable ϵ with respect to the random vector $(\mathbf{X}, \xi)$. The authors use the classical kernel method (Nadaraya-Watson type weights method) to prove the convergence rate of m and the asymptotic normality of β. This model was extended to dependent data by [6]. Furthermore, Aneiros-Pérez and Vieu [4] proposed a bootstrap procedure to approximate the distribution of these estimators. The case where the linear component is also functional was proposed by [29]. More recently, Aneiros-Pérez et al. [5] studied different bootstrapping procedures for this model under dependency structures while Ling et al. [31] considers SPFLR models with random missing responses. A procedure for testing linearity in partially linear functional models is proposed in [39] while the extension of this model for spatial data was proposed by [13].Other approaches have been proposed to estimate SPFLR model parameters, we cite for example, the local linear approach (LLE) used by [22], the robust procedures considered by [15], the k nearest neighbors (kNN) procedure used by [30] and Bayesian approaches proposed by [37]. For the most recent contributions in this area, we can consult the bibliographic reviews in [33, 34].

In this paper, we consider using the kNN estimation under the local linear approach (kNN)-LLE for i.i.d data, which was recently developed by [10]. The combination of the two approaches makes it possible to build an estimator with good statistical properties which inherits their advantages. This is confirmed by the construction of an attractive and robust estimator which converges quickly with lower bias and is very easy to implement in practice (see [1] for more details).

Indeed, in nonparmetric statistics, when the variables are of a functional nature, it is necessary to take into account the local structures of the data. The kNN approach is generally the most widely used. This method makes it possible to select a bandwidth parameter adapted to the structure of the data and the estimator which deduces from it can be updated with any new observation. Attracting by its characteristics, this approach has been increasingly considered in recent years the FDA, in particular to the construction of the kNN kernel estimator of the regression operator and to obtain its asymptotic properties. The kNN approach for functional data was studied by [18] and reconsidered in other approaches by [8, 26–28]. Some recent advances for this method can be found in [32, 35].

On the other hand, compared to the classical Nadaraya-Watson method, local linear fitting is known to have various advantages. More precisely, this approach improves the quality of the bias error (see [21] for more motivations on this method). Thus, the functional local linear modeling has become an attractive topic in nonparametric functional regression in recent years. The results developed from this approach lead to several versions of estimators (see. [11, 12, 14] and [19] for spatial data) The asymptotic distribution of the estimator was established by [40] while the case where all the variables are curves is considered by [20]. We also cite the work of [9] where they study the asymptotic properties uniformly on the smoothing parameter.

For all that we have just mentioned, the objective of this article is to consider an approach based on a combination of the estimations by the kNN-LLE method with semi-parametric partially linear models. Our contribution can be considered as the first study in this direction and our main results are to prove the asymptotic distribution of the parametric component and the almost certain convergence (with rate) of the non-parametric component.

Our paper is organized as follows. In Section 2, the estimators of the model based on a sample of i.i.d random vectors are defined, whereas Section 3 is devoted to the assumptions and the asymptotic results. In section 4, some results on simulated data and real data are carried out to highlight how the kNN-LLE approach allows a nice improvement compared to the usual global approach. Finally, the Proof section is devoted to proofs of certain technical lemmas and proofs of the results of this article.

2. The model and its estimation

Let the observed data $\{(Y_i, X_{i1}, X_{i2}, \ldots, X_{ip}, \xi_i)^T, 1 \leq i \leq n\}$, which are independant and identically distributed as $(Y, X_1, X_2, \ldots, X_p, \xi)^T$, be generated by means of a semi functional partial linear regression model

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

where Y_i, $\mathbf{X}_i = (X_{i1}, X_{i2}, \ldots, X_{ip})$, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^T$, m and ϵ_i are defined as before such that the random errors satisfying

$$\mathbb{E}(\epsilon_i|X_{i1}, \ldots, X_{ip}, \xi_i) = 0 \text{ and } \mathbb{E}(\epsilon_i^2|X_{i1}, \ldots, X_{ip}, \xi_i) < \infty \quad .$$

We assume that ξ is valued in semi-metric space $\mathcal{F}$ with associated semi-metric denoted by $d(\cdot, \cdot)$ and we denote by $B(\xi, h) = \{\xi' \in \mathcal{F}\}$ the the topological closed ball in $\mathcal{F}$ such that $d(\xi, \xi') \leq h\}$.

Now, by conditioning Y_i on ξ, we have

$$\mathbb{E}(Y_i|\xi_i = \xi) = \mathbb{E}(\mathbf{X}_i|\xi_i = \xi)^T \beta + m(\xi)\,. \tag{2.2}$$

Then, by the equations (2.1) and (2.2) , we can write

$$Y_i - \mathbb{E}(Y_i|\xi_i = \xi) = (\mathbf{X}_i - \mathbb{E}(\mathbf{X}_i|\xi_i = \xi))^T \boldsymbol{\beta} + \epsilon_i\,. \tag{2.3}$$

In the following we denote $g_1(\xi) = \mathbb{E}(\mathbf{X}_i|\xi_i = \xi)$, $g_2(\xi) = \mathbb{E}(Y_i|\xi_i = \xi)$, $\overline{Y}_i = Y_i - g_2(\xi_i)$ and $\overline{\mathbf{X}}_i = \mathbf{X}_i - g_1(\xi_i)$.

Note that, by equation (2.2), we have

$$m(\xi) = g_2(\xi) - g_1^T(\xi)\boldsymbol{\beta}, \tag{2.4}$$

and, by equations (2.3), we have

$$\overline{Y}_i = \overline{\mathbf{X}}_i^T \boldsymbol{\beta} + \epsilon_i\,. \tag{2.5}$$

Thus, under the condition that $g_1(\xi)$ and $g_2(\xi)$ are known, the least squares estimator (LSE) of $\boldsymbol{\beta}$ is

$$\overline{\boldsymbol{\beta}}_n = \arg\min_{\boldsymbol{\beta}} \sum_{i=1}^{n} (\overline{Y}_i - \overline{\mathbf{X}}_i^T \boldsymbol{\beta})^2,$$

which is obtained by the following formula

$$\hat{\boldsymbol{\beta}}_n = \left(\sum_{i=1}^{n} \overline{\mathbf{X}}_i^T \overline{\mathbf{X}}_i\right)^{-1} \sum_{i=1}^{n} \overline{\mathbf{X}}_i^T \overline{Y}_i. \tag{2.6}$$

However, $g_1(\xi)$ and $g_2(\xi)$ are generally unknown and they must be estimated for applied (2.6). Assuming that g_1 and g_2 are smooth functions of ξ_i, these can be estimated using nonparametric estimators noted by $\hat{g}_1$ and $\hat{g}_2$.

As cited in the introduction, the functional nonparametric estimator methods used to estimate g_1 and g_2 are: Nadaraya-Watson (N-W) kernel approach, kNN method, local linear approach and robust estimators.

The main purpose of this article is to using local linear approach weighted by the k-nearest neighbors smoothing procedure (kNN-LLE) to estimate g_1 and g_2. Recall that the estimation of $g_l(\xi)$ for $l = 1, 2$ by the kNN-LLE method was introduced by [10]. They assumed that the $g_l(\xi)$ is locally approximated, for all $\xi_0 \in \mathcal{N}_\xi$ ($\mathcal{N}_\xi$ a neighborhood of ξ), by

$$g_l(\xi_0) = a + b\varrho(\xi_0, \xi) + o(d(\xi, \xi_0)),$$

where $\varrho(\cdot,\cdot)$ is a known function from $\mathcal{F}^2$ into $\mathbb{R}$ such that $\varrho(\xi,\xi)=0$. Under this, if we denote by $h_k = h_{k,n}$ the random sequence of positive real numbers such that

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

Thus, the LLE-kNN estimator of $g_1(\xi)$ and $g_2(\xi)$ is expressed by

$$\widehat{g}_1(\xi) = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} W_{ij}(\xi,h_k)X_j}{\sum_{i=1}^{n}\sum_{j=1}^{n} W_{ij}(\xi,h_k)} = \sum_{j=1}^{n} \mathbf{W_j}(\xi)\mathbf{X_j}\,, \tag{2.7}$$

and

$$\widehat{g}_2(\xi) = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} W_{ij}(\xi,h_k)Y_j}{\sum_{i=1}^{n}\sum_{j=1}^{n} W_{ij}(\xi,h_k)} = \sum_{j=1}^{n} \mathbf{W_j}(\xi)Y_j, \tag{2.8}$$

where $W_{ij}(\xi,h_k) = \varrho_i(\varrho_i - \varrho_j)K_iK_j$, $\mathbf{W_j}(\xi) = \sum_{i=1}^{n} W_{ij}(\xi,h_k)/\sum_{i=1}^{n}\sum_{j=1}^{n} W_{ij}(\xi,h_k)$, with $\varrho_i = \varrho(\xi_i,\xi)$ and $K_i = K(h_k^{-1}d(\xi,\xi_i))$ (K stands for the kernel function).

As the parameter k is unknown, this estimator cannot be used directly. It is therefore necessary to estimate the parameter k and for this we will use the cross-validation method. Precisely, we choose k according to the following cross-validation rule

$$k_{opt} = \arg\min_k CV(k) = \arg\min_k \sum_{i=1}^{n}\left(Y_i - \widetilde{Y}_{(-i)}^{kNN}\left((X_i,\xi_i)\right)\right)^2, \tag{2.9}$$

where $\widetilde{Y}_{(-i)}^{kNN}(X_i)$ is the values of the leave one-out kNN-LLE estimator at $(X_i,\xi_i))$ (see [1] for more details).

Hence, an estimator of $\boldsymbol{\beta}$ after estimating g_1 and g_2 is given by

$$\widehat{\boldsymbol{\beta}}_n = (\widetilde{\mathbf{X}}^T\widetilde{\mathbf{X}})^{-1}\widetilde{\mathbf{X}}^T\widetilde{Y} = \left(\sum_{i=1}^{n}\widetilde{\mathbf{X}}_i^T\widetilde{\mathbf{X}}_i\right)^{-1}\sum_{i=1}^{n}\widetilde{\mathbf{X}}_i^T\widetilde{Y}_i, \tag{2.10}$$

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

Finally, a nonparametric estimator for m can be obtained by (2.4) and (2.10)

$$\widehat{m}_n(\xi) = \widehat{g}_2(\xi) - \widehat{g}_1(\xi)^T\widehat{\boldsymbol{\beta}}_n = \sum_{j=1}^{n}\mathbf{W_j}(\xi)Y_j - \left(\sum_{j=1}^{n}\mathbf{W_j}(\xi)\mathbf{X_j}\right)\widehat{\boldsymbol{\beta}}_n. \tag{2.11}$$

3. Asymptotic properties

The aim of this section is to establish some asymptotic properties of the estimators $\widehat{\boldsymbol{\beta}}_n$ and $\widehat{m}_n(\xi)$ defined in (2.10) and (2.11), respectively. To do that, we'll start by introducing some following notations:

$$g_{1s}(\xi) = \mathbb{E}(X_{is}|\xi_i = \xi), \quad (i=1,\ldots,n, s=1,\ldots,p),$$

$$\eta_{is} = X_{is} - g_{1s}(\xi_i), \quad \boldsymbol{\eta}_i = (\eta_{i1},\ldots,\eta_{ip})^T\,,$$

and we need the following assumptions.

3.1. Technical assumptions

(A1): For any $h > 0$, $\mathbb{P}(B(\xi, h)) := \phi_\xi(h) > 0$ is an invertible function and continuous on a neighborhood of 0 such that $\phi_\xi(0) = 0$.

(A2): There exists a positive number α such that, $\forall (u, v) \in \mathcal{F}^2$ and $\forall f \in \{m, g_{11}, g_{12}, \ldots, g_{1p}\}$,
$$|f(u) - f(v)| \leq Cd^\alpha(u, v).$$

(A3): The kernel K is a continuous, derivable and compact supported on $]0, 1[$ such that there exist two constants C and C' satisfying
$$-\infty < C' < K'(.) < C < 0\,.$$

(A4): $\forall z \in \mathcal{F}$, the function ϱ is such that
$$Cd(\xi, z) \leq \mid \varrho(\xi, z) \mid \leq C'd(\xi, z).$$

(A5): For all sequence $h := h_n \to 0$, we have
$$h\left(\int_{B(\xi,h/2)} \varrho(u, \xi)dP^\xi(u)\right) / \left(\int_{B(\xi,h/2)} \varrho^2(u, \xi)dP^\xi(u)\right) \to 0,$$
where dP^ξ denotes the probability distribution of the regressor ξ.

(A6): The number of neighbours k is such that
$$\frac{k}{n} \to 0 \text{ and } \frac{\log n}{k} \to 0 \text{ as } n \to \infty.$$

(A7): There exists two positive constants C and C' such that $\forall m \geq 3$, $\mathbb{E}(|X_{1s}|^m|\xi_1 = \xi) \leq C < \infty$ for $s = 1, \ldots, p$, and $\mathbb{E}(|Y|^m|\xi_1 = \xi) < a_m(\xi) < C' < \infty$, with $a_m(\cdot)$ is a continuous function on N_ξ.

(A8): $\boldsymbol{\Sigma} = \mathbb{E}(\boldsymbol{\eta}_1\boldsymbol{\eta}_1^T)$ is a positive definite matrix.

Comments on the assumptions: We note that the hypotheses considered are quite usual in local linear smoothing by the kNN approach in this context of non-parametric functional models and/or partial linear functional models. In particular, the conditions (A1)-(A3), (A7) and (A8) are used for estimation in the SFPLR model (see [7]), while the conditions (A4)-(A6) are used even for the kNN-LLE functional estimator which is a special case of local linear modeling (see [1], [10]).

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)-(A8), if in addition* $\frac{\sqrt{n}\log n}{k} \to 0$ *as* $n \to \infty$, $\frac{\sqrt{n}\log^2 n}{k} \to 0$ *as* $n \to \infty$, $\sqrt{n}\phi^{-1}(\frac{k}{n})^\alpha \to 0$ *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*

(i): $n^{1/2}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \xrightarrow{\mathcal{D}} N(0, \sigma^2\boldsymbol{\Sigma}^{-1})$,

(ii): $\limsup_{n\to\infty} \sqrt{\frac{n}{2\log\log n}}|\widehat{\beta}_{ns} - \beta_s| = (\sigma^2\sigma_{ss})^{\frac{1}{2}}$ *a.s.* , *where* $\sigma_{ss} = (\boldsymbol{\Sigma}^{-1})_{ss}$ *for* $s = 1, \ldots, p$.

Theorem 3.2. *Under the assumptions of Theorem 3.1, we have that*
$$|\widehat{m}(\xi) - m(\xi)| = O\left(\phi^{-1}\left(\frac{k}{n}\right)^\alpha\right) + O\left(\sqrt{\frac{\log n}{k}}\right) \text{ a.s..}$$

4. Application results

4.1. Simulation study

The main objective of this section is to examine the behavior of the local linear kNN approach on finite samples. More precisely, in order to highlight the superiority of this approach compared to other approaches, we compare the mean square error of prediction behavior of the two models: functional nonparametric regression and Semi-functional partial linear regression for the four estimators: Kernel estimator, linear local estimator, kNN estimator and kNN-LLE estimator. The proposed estimators regression are:

- Functional Non-Parametric Kernel regression (FNP.KR) proposed by [23];
- Functional Non-Parametric kNN regression (FNP.kNN) introduced by [18];
- Functional Non-Parametric Local-Linear regression (FNP.LLE) proposed by [12];
- Functional Non-Parametric kNN-Local Linear regression (FNP.kNN-LLE) developed by [1];
- Semi-functional partial linear kernel regression (SFPLR.KR) introduced by [7];
- Semi-functional partial linear kNN regression (SFPLR.kNN) proposed by [30];
- Semi-functional partial linear Local Linear regression (SFPLR.LLE), our estimator given by the equation (2.11) by replacing h_k with h_n (through the utilization of a cross-validation procedure);
- Semi-semi-functional partial linear Local-Linear kNN regression (SFPLR.kNN-LLE), our estimator given by the equation (2.11).

In this simulation, we used the routine code available at https://www.math.univ-toulouse.fr/~ferraty/SOFTWARES/NPFDA/npfda-content.html and the 'fda' R-package, as discussed in [23], with some enhancements within the SPLR framework.
Formally, we generate our observations (Y_i, X_i, ξ_i) from the following SFPLR model:

$$Y_i = m(\xi_i) + X_i^T\beta + \epsilon_i,\ i = 1, \dots, n = 300\,,$$

where $X_i \sim Exp(0.5)$, $\beta = 2$, $\epsilon_i \sim \mathcal{N}(0, 1)$, and take the nonparametric operators $m(.)$ as follows:

$$m(\xi) = \frac{10}{1 + \int_0^1 \xi(t)dt}.$$

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

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

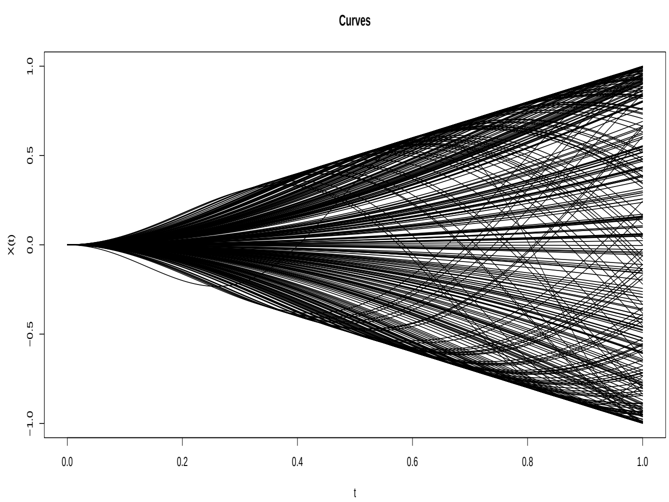

Figure 1. A sample of $n = 300$ functional explanatory variables ξ.

Recall that the FNP model is defined as follows:

$$Y_i = m(\xi_i) + \epsilon_i, \ i = 1, \ldots, n = 300.$$

In order to compute our estimators we use the class of semi-metrics based on derivatives which is well adapted to this type of data (smooth data):

$$d(\xi_1, \xi_2) = \sqrt{\int_0^1 (\xi_1^{(s)}(t) - \xi_2^{(s)}(t))^2 dt}, \tag{4.1}$$

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

$$K(u) = \frac{3}{4}(1 - u^2)\mathbb{I}_{[0,1/2]}(u). \tag{4.2}$$

While the Local-Linear estimator (for the two models FNP and SFPLR) is constructed by the same procedure proposed by [12] for which the locating function ϱ is defined by

$$\varrho(\xi_1, \xi_2) = \int_0^1 \theta(t)(\xi_1^{(s)}(t) - \xi_2^{(s)}(t))dt \tag{4.3}$$

where θ is the eigenfunction of the empirical covariance operator:

$$\frac{1}{|\Lambda|}\sum_{i\in\Lambda}(\xi_i^{(s)}(t) - \overline{\xi^{(s)}(t)})^t(\xi_i^{(s)}(t) - \overline{\xi^{(s)}(t)}), \ \ with \ \ \overline{\xi^{(s)}(t)} = \frac{1}{|\Lambda|}\sum_{i\in\Lambda}\xi_i^{(s)}(t),$$

associated with the q-greatest eigenvalue.

On other hand, the models' 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 optimal bandwidth (h_{opt}) is obtained using the kernel cross-validation (CV) method and we use the selection rule proposed in [23].

$$h_{opt} = \arg\min_h CV(h) \text{ where } CV(h) = \sum_{i=1}^n \left(Y_i - \widetilde{Y}_{l,(-i)}\right)^2, \, l = 1, 2,$$

with $\widetilde{Y}_{1,(-i)}$ (resp$\widetilde{Y}_{2,(-i)}$) is the leave-one-out values of the functional nonparametric regression estimators calculate at (ξ_i) (resp. the leave-one-out values of the semi-functional nonparametric regression estimators calculate at $(\mathbf{X}_i, \xi_i)$).

In the second stage, the k-NN 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:

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

where

$$k_{opt} = \arg\min_h CV(k) \text{ with } CV(k) = \sum_{i=1}^n \left(Y_i - \widetilde{Y}_{l,(-i)}^{kNN}\right)^2 \ l = 1, 2.$$

The $\widetilde{Y}_{1,(-i)}^{kNN}$ (resp $\widetilde{Y}_{2,(-i)}^{kNN}$) is the leave-one-out values of the functional nonparametric regression estimators calculate at (ξ_i) (resp. the leave-one-out values of the semi-functional nonparametric regression estimators calculate at $(\mathbf{X}_i, \xi_i)$).

In this simulation study, we take the following parameters: $s = 2$, $q = 8$, $\Lambda = \{1, \ldots, 300\}$, and $k \in \{10, 10 + \lfloor \frac{n}{100} \rfloor, 10 + 2\lfloor \frac{n}{100} \rfloor, \ldots, \lfloor \frac{n}{2} \rfloor\}$ ($\lfloor . \rfloor$ is the ceiling function).

We randomly split the n-sample into two parts: one is a training sample $S_{train} = \{(Y_i, X_i, \xi_i)_{i\in \text{ Train}}\}$ (for example, we take 75% of the n-sample) which is used for modeling procedure, and the other is a testing sample $S_{test} = \{(Y_i, X_i, \xi_i)_{i\in \text{ Test}}\}$ which is used to verify the prediction effect. To assess the effectiveness of the proposed model for this prediction problem, we calculated the mean square error of prediction ($MSEP$) on the test sample:

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

where n_{Test} is the length of the testing sample and $\widetilde{Y}_1$ (resp $\widetilde{Y}_{2,}$) is the prediction values of the functional nonparametric regression estimators calculate at (ξ_i) (resp. the prediction values of the semi-functional nonparametric regression estimators calculate at (X_i, ξ_i)).

The experiment was replicated $M = 100$ times, which allow us to compute M values for $MSEP$ and display their distribution through a boxplot. More precisely, every $MSEP$ is calculated by M independent replications of sample size n,

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

where $MSEP_u$ is the Mean square errors of prediction computed from the u^{th} sample.

The $MSEP$ errors are showing in Figure 2 and summarize in Table 1. This figure shows the boxplots of the $MSEP$ of the prediction values in the testing sample for the eight proposed regression models.

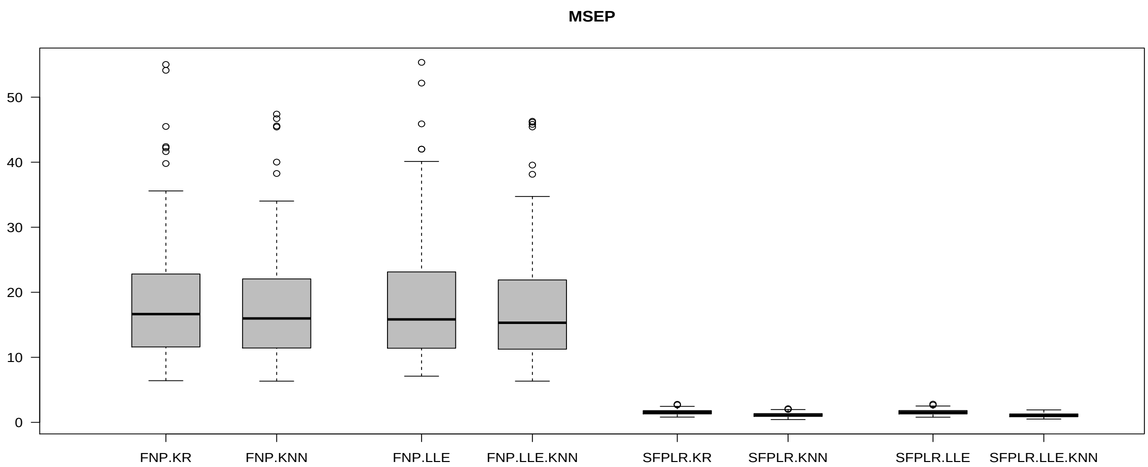

Figure 2. Mean square errors of prediction $MSEP$ from the M replicates of the experiment.

Table 1. $\overline{MSEP}$ (Mean squared error) for the six models.

	Model	Kernel	kNN	LLE	kNN-LLE
FNP	$Y = m(\xi) + \epsilon$	17.930123	16.769152	17.192430	16.715073
SFPLR	$Y = m(\xi) + \sum_{k=1}^{3} X_k \beta^k + \epsilon$	1.530868	1.144981	1.510142	1.143406

The prediction results are given in Figure 3 where we plot the predicted values against the true values for both models and using the four estimation methods. This Figure gives an idea on the accuracy of the predictions corresponding to one run.

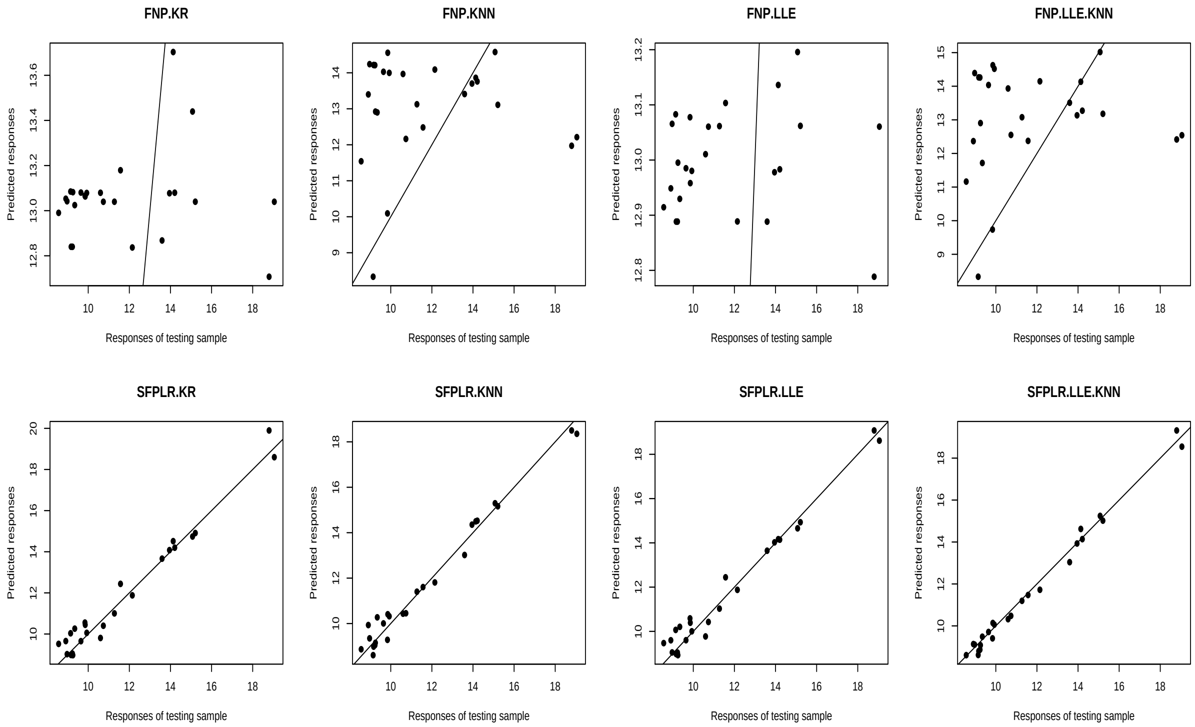

Figure 3. Prediction in the testing for the eight methods.

It clearly appears that the SFPLR model shows the best prediction effects compared to the FNR model. This conclusion can be confirmed by the MSEP error. In fact, the comparison of Mean Squared Error of Prediction ($MSEP$) for the eight methods, as illustrated in Figure 2, demonstrates that the forecasting accuracy of the SFPLR model. Additionally, the results suggest that the kNN-LLE approach outperforms other methods when applied to SFPLR models.

Now, we aim to determine how the degree of of derivation, denoted as s, influences the distance measure between curves, represented as $d(\cdot,\cdot)$, in semi-functional partial linear models using the bias of the β measurement. Table 2 provides insights into the potential improvement in predictive power through the Square Errors SE for β:

$$SE = |\widehat{\beta}_n - \beta|^2.$$

Table 2. The SE of $\widehat{\beta}_n$ for the four semi-functional partial linear models varies according to the degree of derivation, denoted as s.

Degree of derivation s	SFPLR.KR	SFPLR.KNN	SFPLR.LLE	SFPLR.LLE.KNN
0	0.04603	0.03000	0.03065	0.02729
1	0.04642	0.03605	0.03426	0.03604
2	0.04490	0.02886	0.03166	0.02922
3	0.07023	0.02591	0.03074	0.02647
4	0.23835	0.02912	0.03021	0.02941
5	0.16881	0.03394	0.02949	0.03396
6	0.32991	0.02966	0.03120	0.02958

4.2. Application to real data

As in the simulation part, the goal of this section is to compare, on a real dataset, our kNN-LLE estimators to other types of estimators proposed for SFLPR models and to the estimators used for nonparametric functional regression (FNR), namely kNN estimator, local linear estimator (LLE) and kernel estimator.

We show, on the one hand, the ease of practical implementation of the estimators obtained and on the other hand their usefulness compared to other estimation methods for the two models (FNR and SFLPR).

Data Description

The dataset contains NIR spectroscopy measurements resulting from an experiment to determine the composition of 72 samples of cookie dough pieces (formed but unbaked cookies) studied by [17] which had been analysed by [2]. The main objective is to predict the percentage of fat content Y for cookie dough datasets from the corresponding content, namely, sucrose (X_1), dry flour (X_2) and water (X_3) as well as from the spectra of near-infrared absorbance ξ, using SFPLR. For this purpose, near-infrared spectroscopy curves representing $\xi_i(\omega)$ absorption steps from 1100 to 2498 nm were measured at intervals of 2 nm steps. The absorption $\xi_i(\omega)$ of light is measured for each wavelength ω and each sample i, $i = 1, \ldots, 72$. $\xi_i(\omega_1), \ldots, \xi_i(\omega_{700})$ represent the i th discretized spectrometric curve.

The sample of curves $\{\xi_i\}_1^{72}$ can be observed in Figure 4.

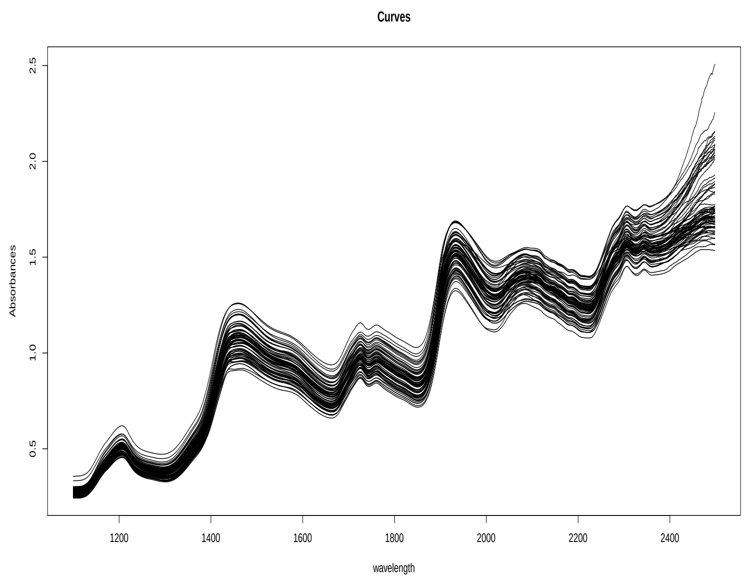

Figure 4. A sample of 72 curves ξ of NIR spectroscopy.

Statistical Models The 72 i.i.d. observations were randomly divided into two subsets. The first subset, 75% of the cookie dough sample, was used for the modeling procedure (training sample). The second subset was used as a test sample to evaluate the prediction quality.

Assuming a Semi-Functional Partial Linear Regression (SFPLR) model, we establish a relationship between the variables as follows:

$$Y = m(\xi) + \sum_{k=1}^{3} X_k \beta^k + \epsilon,$$

where $\beta^1, \beta^2, \beta^3$ and $m(\cdot)$ are unknown modelling the relationship between X_1, X_2, X_3, ξ and Y. The statistical difficulty is to come up with a suitable estimator. We will concentrate on regression models such that $\mathbb{E}(\epsilon \mid X_1, X_2, X_3, \xi) = 0$.

To calculate our estimators, we will keep the same tools as in the simulation. We carry out $M = 100$ independent replications, which allow us to compute M values for $MSEP$ and display their distribution through a boxplot.

The $MSEP$ results are reported in Table 3.

Table 3. $\overline{MSEP}$ (Mean squared error) for FNP.KR, FNP.*k*NN, FNP.LLE, FNP.*k*NN-LLE, SFPLR.KR, SFPLR.*k*NN, SFPLR.LLE and SFPLR.*k*NN-LLE models.

	Model	Kernel	*k*NN	LLE	*k*NN-LLE
FNP	$Y = m(\xi) + \epsilon$	3.2072918	2.0579926	2.7751550	1.9710179
SFPLR	$Y = m(\xi) + \sum_{k=1}^{3} X_k \beta^k + \epsilon$	0.5793651	0.5166147	0.1919556	0.1038148

Additionally, one hundred random partitions were generated to reduce the impact of S_{train} and S_{test} partitions. One hundred $MSEP$ values were obtained for each procedure from the FNP and SFPLR models. Figure 5 represents the boxplots corresponding to the M values of $MSEP$ for the eight models.

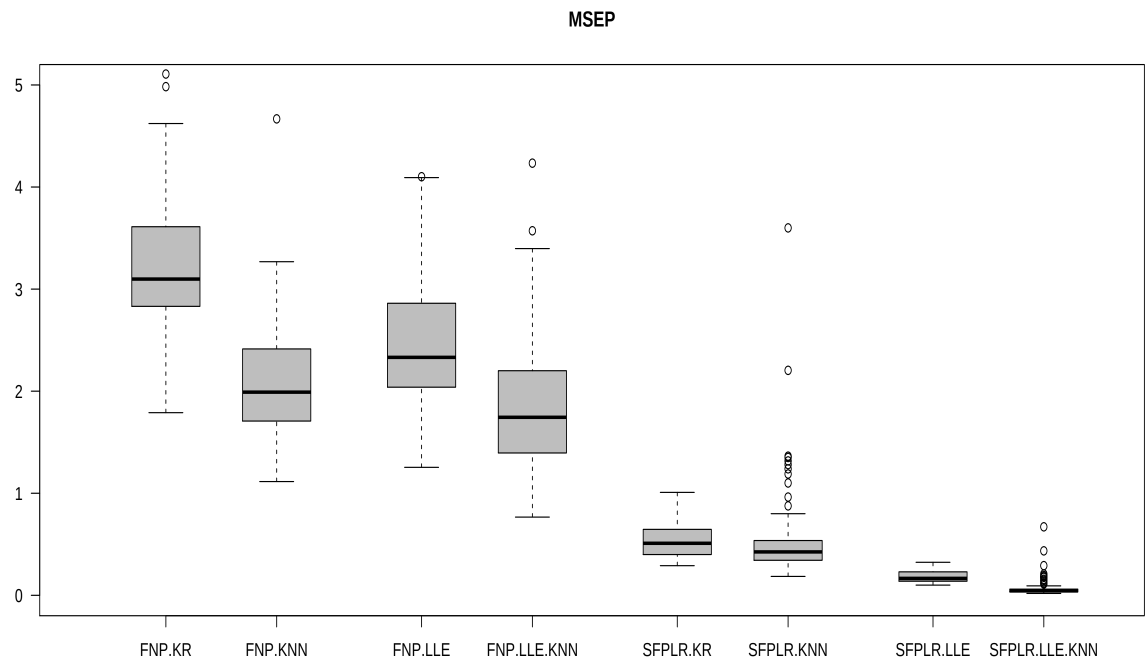

Figure 5. The $MSEP$ box plots of the cookie prediction values dough data for both Models with the 4 estimation methods.

In Figure 6, we plot the predicted values versus the true values.

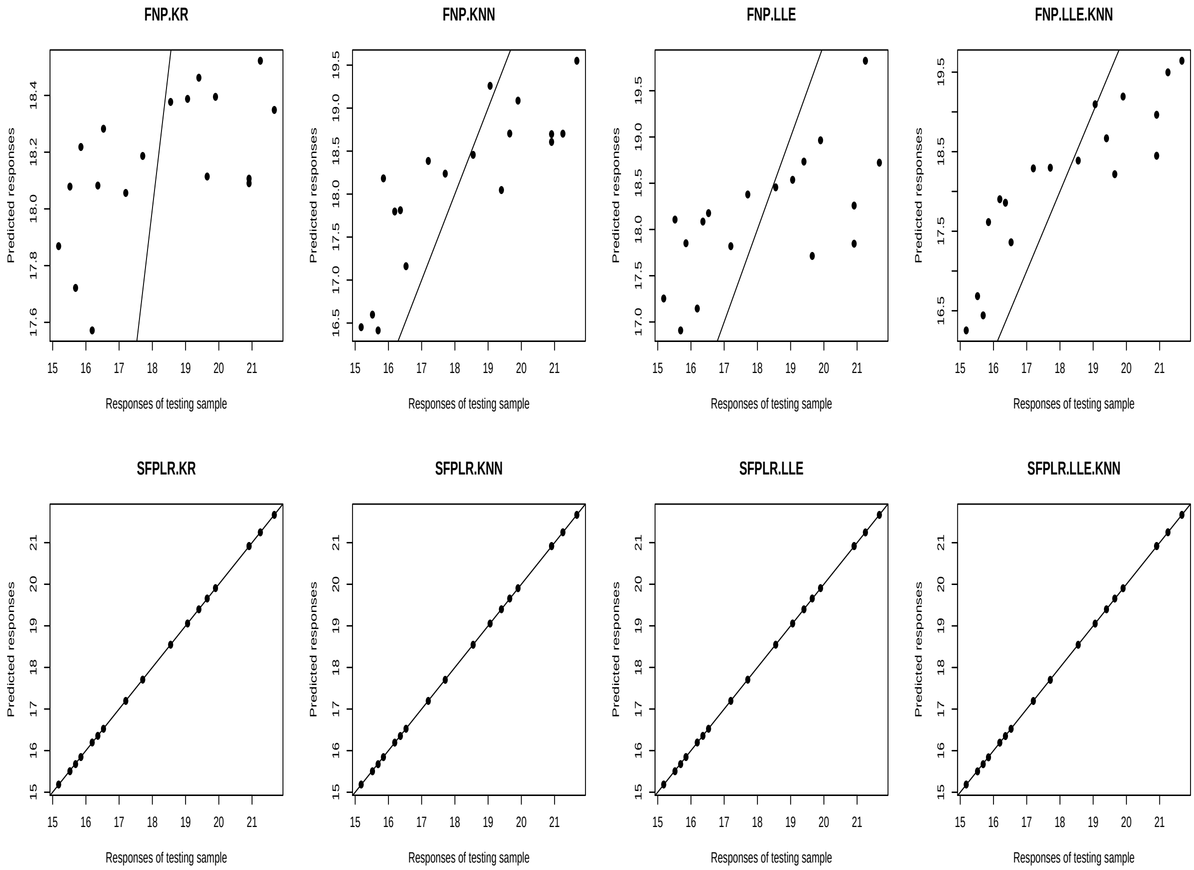

Figure 6. Prediction of the testing cookie dough samples

5. Concluding remarks

The considered dataset contains NIR spectroscopy measurements from an experiment to determine the composition of 72 samples of cookie dough pieces, specifically formed but unbaked cookies. The primary objective is to predict the fat content percentage of cookie dough based on key ingredients, including sucrose, dry flour, and water, as well as the near-infrared absorbance spectra. The model of choice for this predictive task is SFPLR. Table 3 provides a comprehensive view of Mean Squared Error ($MSEP$) values for various predictive models. These models encompass different configurations, including FNP.KR, FNP.kNN, FNP.LLE, FNP.kNN-LLE, SFPLR.KR, SFPLR.kNN, SFPLR.LLE, and SFPLR.kNN-LLE. The results highlight that SFPLR models consistently outperform FNP models in terms of prediction accuracy, and in particular, the kNN-LLE approach shows promise in the SFPLR framework. The MSEP obtained (Figure 5) confirm the previous conclusions by affirming the emergence of the kNN-LLEN approach as an outstanding player in the SFPLR category. Figure 6 takes the evaluation a step further by depicting the alignment between predicted values and true values. This scatter plot visually confirms the accuracy of predictions and provides an intuitive assessment of model performance. Once again, it reinforces the previous conclusions.

Acknowledgment. The authors are indebted to the Editor-in-Chief and the referees for their very valuable comments and suggestions which led to a considerable improvement of the manuscript. They also extend their appreciation to Thematic Research Agency in

Science and Technology (ATRST) for funding this work through research groups program under the project number PRFU, C00L03UN220120220002.

References

[1] I.M. Almanjahie, W. Mesfer, A. Laksaci and M. Rachdi, *Computational aspects of the kNN local linear smoothing for some conditional models in high dimensional statistics*, Comm. Statist. Simulation Comput. **52** (7), 2985-3005, 2023.

[2] I.M. Almanjahie, O. Fetitah, M.K. Attouch and T. Benchikh, *Functional nonparametric predictions in food industry using near-infrared spectroscopy measurement*, Comput. Mater. Contin. **74** (3), 6307-6319, 2023.

[3] G. Aneiros-Pérez, I. Horová, M. Hušková and P. Vieu, *Editorial for the special issue on functional data analysis and related fields*, J. Multivariate Anal. **189**, 2022.

[4] G. Aneiros-Pérez and P. Vieu, *Automatic estimation procedure in partial linear model with functional data*, Stat. Pap. **52** (4), 751-771, 2011.

[5] G. Aneiros Pérez, P. Raña, P. Vieu and J. Vilar, *Bootstrap in semi-functional partial linear regression under dependence*, Test **27** (3), 659-679, 2018.

[6] G. Aneiros-Pérez and P. Vieu, *Nonparametric time series prediction: A semi-functional partial linear modeling*, J. Multivariate Anal. **99**, 834-857, 2008.

[7] G. Aneiros-Pérez and P. Vieu, *Semi-functional partial linear regression*, Stat. Probab. Lett. **76** (11), 1102-1110, 2006.

[8] M.K Attouch and T. Benchikh, *Asymptotic distribution of robust k-nearest neighbour estimator for functional nonparametric models*, Mat. Vesnik **644**, 275-285, 2012.

[9] M.K Attouch, A. Laksaci and F. Rafaa, *On the local linear estimate for functional regression: uniform in bandwidth consistency*, Comm. Statist. Theory Methods **48**, 1836-1853, 2019.

[10] M.K. Attouch, A. Laksaci and F. Rafaa, *Estimation locale linéaire de la régression non paramétrique fonctionnelle par la méthode des k plus proches voisins*, Comptes Rendus. Mathématique **355** (7), 824-829, 2017.

[11] A. Baíllo and A. Grané, *Local linear regression for functional predictor and scalar response*, J. Multivariate Anal. **100** (1), 102-111, 2009.

[12] J. Barrientos-Marin, F. Ferraty and P. Vieu, *Locally modelled regression and functional data*, J Nonparametr Stat. **22** (5-6), 617-632, 2010.

[13] M. Benallou, M.K. Attouch, T. Benchikh and O. Fetitah, *Asymptotic results of semi-functional partial linear regression estimate under functional spatial dependency*, Comm. Statist. Theory Methods **51**, 1-21, 2021.

[14] A. Berlinet, A. Elamine and A. Mas, *Local linear regression for functional data*, Ann. Inst. Statist. Math. **63** (5), 1047-1075, 2011.

[15] G. Boente and A. Vahnovan, *Robust estimators in semi-functional partial linear regression models*, J. Multivariate Anal. **154** (C), 59-84, 2017.

[16] D. Bosq and D. Blanke, *Inference and Prediction in Large Dimension*, Wiley Series in Probability and Statistics, Chichester, 2007.

[17] P. Brown, T. Fearn and M. Vannucci, *Bayesian wavelet regression on curves with application to a spectroscopic calibration problem*, J. Amer. Statist. Assoc. **96**, 398-408, 2001.

[18] F. Burba, F. Ferraty and P. Vieu, *k-Nearest Neighbour method in functional nonparametric regression.* J. Nonparametr. Stat. **21** (4), 453-469, 2009.

[19] A. Chouaf and A. Laksaci, *On the functional local linear estimate for spatial regression*, Stat. Risk Model **29**, 189-214, 2013.

[20] J. Demongeot, A. Naceri, A. Laksaci and M. Rachdi, *Local linear regression modelization when all variables are curves*, Statist. Probab. Lett. **121**, 37-44, 2017.

[21] J. Fan and I. Gijbels. *Local Polynomial Modelling and Its Applications*, London: Chapman and Hall, 1996.

[22] S. Feng and L. Xue, *Partially functional linear varying coefficient model*, Statistics **50** (4), 717-732, 2016.

[23] F. Ferraty and P. Vieu, *Nonparametric Functional Data Analysis. Theory and Practice*, Springer Series in Statistics, New York, 2006.

[24] T. Hsing and R.L. Eubank, *Theoretical Foundations of Functional Data Analysis, with an Introduction to Linear Operators*, John Wiley and Sons, 2015.

[25] S. Greven and F. Scheipl, *A general framework for functional regression modelling*, Stat. Model. **17** (1-2), 1-35, 2017.

[26] L. Kara-Zaitri, A. Laksaci, M. Rachdi and P. Vieu, *Uniform in bandwidth consistency for various kernel estimators involving functional data*, J. Nonparametr. Stat. **29** (1), 85-107, 2017.

[27] N. Kudraszow and P. Vieu P, *Uniform consistency of kNN regressors for functional variables*, Statist. Probab. Lett. **83** (8), 1863-1870, 2013.

[28] H. Lian, *Convergence of functional k-nearest neighbor regression estimate with functional responses*, Electron. J. Stat. **5**, 31-40, 2011.

[29] H. Lian, *Functional partial linear model*, J. Nonparametr. Stat. **23** (1), 115-128, 2011.

[30] N. Ling, G. Aneiros-Pérez and P. Vieu, *knn estimation in functional partial linear modeling*, Statist. Papers **61** (1), 423-444, 2020.

[31] N. Ling, R. Kan, P. Vieu and S. Meng, *Semi-functional partially linear regression model with responses missing at random*, Metrika **82** (1), 39-70, 2019.

[32] N. Ling, S. Meng and P. Vieu, *Uniform consistency rate of kNN regression estimation for functional time series data*, J. Nonparametr. Stat. **31** (2),451-468, 2019.

[33] N. Ling and P. Vieu, *On semiparametric regression in functional data analysis*, Wiley Interdiscip. Rev.: Comput. Stat. **12** (6), 20-30, 2020.

[34] N. Ling and P. Vieu, *Nonparametric modelling for functional data: selected survey and tracks for future*, Statistics **52** (4), 934-949, 2018.

[35] S. Novo, G. Aneiros and P. Vieu. *A kNN procedure in semiparametric functional data analysis*, Statist. Probab. Lett. **17**, 2021.

[36] J. Ramsay and B. Silverman, *Functional Data Analysis (Second Edition)*, Spinger-Verlag, New York, 2005.

[37] H. Shang, *Bayesian bandwidth estimation for a semi-functional partial linear regression model with unknown error density*, Comput. Stat. **29** (3-4), 829-848, 2014.

[38] J. Zhang, *Analysis of Variance for Functional Data*, Monographs on Statistics and Applied Probability, CRC Press, **127**, 2014.

[39] F. Zhao and B. Zhang, *Testing linearity in functional partially linear models*, Acta Math. Appl. Sin., Doi: 10.1007/s10255-023-1040-0, 2022.

[40] Z. Zhou and Z. Lin, *Asymptotic normality of locally modelled regression estimator for functional data*, J. Nonparametr. Stat. **28** (1), 116-131, 2016.

Appendix A. Proofs section

The proof of asymptotic results follow the same ideas as in [7]. They are based on the following intermediate results.

Lemma A.1. *(Theorem 2 in [10]) Under assumptions (A1)-(A7), we have*

$$\left| g_{1s}(\xi) - \sum_{j=1}^{n} \mathbf{W_j}(\xi) X_{js} \right| = O_{a.co.} \left(\phi^{-1}\left(\frac{k}{n}\right)^{\alpha} + \sqrt{\frac{\log n}{k}} \right). \tag{A.1}$$

$$\left| g_{2}(\xi) - \sum_{j=1}^{n} \mathbf{W_j}(\xi) Y_{j} \right| = O_{a.co.} \left(\phi^{-1}\left(\frac{k}{n}\right)^{\alpha} + \sqrt{\frac{\log n}{k}} \right). \tag{A.2}$$

Lemma A.2. *(Lemma 3 in [7]) Let $Z_1, \dots, Z_n$ be independent centered r.v.'s such that for some $r \geq 2$, $\max_{1\leq l\leq n} \mathbb{E}|Z_l|^r \leq C < \infty$. Assume that $\{a_{jl}, \quad j,l = 1,\dots,n\}$ is a sequence of positive numbers such that $\max_{1\leq j,l\leq n}|a_{jl}| = O(a_n)$ and $\max_{1\leq j\leq n}\sum_{l=1}^{n}|a_{jl}| = O(b_n)$. If, in addition,*

$$\exp\left(-\frac{b_n^{1/2}(\log n)^2}{b_n^{1/2} + a_n^{1/2} n^{1/r}\log n} \right) = O(n^{-a}), (a > 2),$$

and

$$a_n^{1/2} n^{1/r+b} = O(b_n^{1/2}\log n), (b > 0),$$

then

$$\max_{1\leq j\leq n} \left| \sum_{l=1}^{n} a_{jl} V_l \right| = O(a_n^{1/2} b_n^{1/2} \log n) \quad a.s. \tag{A.3}$$

We then state the following lemmas, similar to Lemmas 6 and 7 in [7], which are necessary to establish our asymptotic results.

Lemma A.3. *Under assumptions (A3) and (A4), if in addition $\xi_1, \dots, \xi_n$ are independent distributed as ξ, we have*

$$\max_{1\leq j,l\leq n} |\mathbf{W_j}(\xi_l)| = O\left(\frac{1}{k-1}\right). \tag{A.4}$$

Proof. Let $J_j = \{j \in \{1,\dots,n\}; d(\xi_j,\xi) \leq h_k\}$, according to (A3), for all $\xi \in \mathcal{F}$, we have

$$C_1 h_k \leq C_1 d(\xi_i,\xi) \leq \varrho(\xi_i,\xi) \leq C_2 d(\xi_i,\xi) \leq C_2 h_k.$$

Then

$$C' h_k^2 \leq \varrho(\xi_i,\xi)(\varrho(\xi_i,\xi) - \varrho(\xi_j,\xi)) \leq C'' h_k^2.$$

Moreover, from assumption (A4) and the fact that $|J_i| = k$, it follows that

$$\sum_{j\in J_i} W_{ij}(\xi,h_k) \geq kCh_k^2\phi_\xi^2(h) \text{ and } \sum_{j\in \overline{J_i}} \mathbf{W_j}(\xi) = 0. \tag{A.5}$$

Similarly, we verify that

$$\sum\sum_{i\neq j} W_{ij}(\xi,h_k) \geq k(k-1)Ch_n^2\phi_\xi^2(h). \tag{A.6}$$

Then we have

$$\frac{\sum_{i=1}^{n} W_{ij}(\xi,h_k)}{\sum\sum_{j\neq i} W_{ij}(\xi,h_k)} \geq \frac{1}{k-1},$$

from which we deduce that

$$\max_{1\leq j,l\leq n} |\mathbf{W_j}(\xi_l)| = O\left(\frac{1}{k-1}\right).$$

□

Lemma A.4. *Under the conditions (A1)-(A8), we have*

$$\frac{1}{n}\widetilde{\mathbf{X}}^T\widetilde{\mathbf{X}} \to \mathbf{\Sigma} \quad a.s. \tag{A.7}$$

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

$$\begin{aligned} n^{-1}\textstyle\sum_{i=1}^n \widetilde{X}_{lr}^T\widetilde{X}_{ls} &= n^{-1}[\textstyle\sum_{l=1}^n \eta_{lr}\eta_{ls} + \sum_{l=1}^n \eta_{lr}\left(g_{1s}(\xi_l) - \sum_{j=1}^n \mathbf{W}_j(\xi_l)X_{js}\right) \\ &\quad + \textstyle\sum_{l=1}^n \eta_{ls}\left(g_{1r}(\xi_l) - \sum_{j=1}^n \mathbf{W}_j(\xi_l)X_{jr}\right) \\ &\quad + \textstyle\sum_{l=1}^n \left(g_{1r}(\xi_l) - \sum_{j=1}^n \mathbf{W}_j(\xi_l)X_{jr}\right) \\ &\quad \times \left(g_{1s}(\xi_l) - \textstyle\sum_{j=1}^n \mathbf{W}_j(\xi_l)X_{js}\right)]\,. \end{aligned} \tag{A.8}$$

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

$$n^{-1}\sum_{l=1}^n \eta_{lr}\eta_{ls} \to \Sigma_{rs} \quad a.s. \tag{A.9}$$

Furthermore, by applying directly the Lemma A.1 and using again the strong law of large numbers for i.i.d. variables, we can see that

$$n^{-1}\sum_{l=1}^n \eta_{lr}\left(g_{1s}(\xi_l) - \sum_{j=1}^n \mathbf{W}_j(\xi_l)X_{js}\right) \to 0 \quad a.s. \tag{A.10}$$

$$n^{-1}\sum_{l=1}^n \eta_{ls}\left(g_{1r}(\xi_l) - \sum_{j=1}^n \mathbf{W}_j(\xi_l)X_{jr}\right) \to 0 \quad a.s. \tag{A.11}$$

and

$$n^{-1}\sum_{l=1}^n \left(g_{1r}(\xi_l) - \sum_{j=1}^n \mathbf{W}_j(\xi_l)X_{jr}\right) \times \left(g_{1s}(\xi_l) - \sum_{j=1}^n \mathbf{W}_j(\xi_l)X_{js}\right) \to 0 \quad a.s. \tag{A.12}$$

Finally, we conclude the proof by using results of A.8-A.12. □

Proof. (**Theorem 3.1**)

Part (i):

Similar to [30], we can write

$$\begin{aligned} \sqrt{n}\left(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}\right) &= \left(n^{-1}\widetilde{\mathbf{X}}^T\widetilde{\mathbf{X}}\right)^{-1}\frac{1}{\sqrt{n}}\left\{\sum_{l=1}^n \widetilde{\mathbf{X}}_l\overline{m}_n(\xi_l) - \sum_{l=1}^n \widetilde{\mathbf{X}}_l\left(\sum_{j=1}^n \mathbf{W}_j(\xi_l)\epsilon_j\right) + \sum_{l=1}^n \widetilde{\mathbf{X}}_l\epsilon_l\right\} \\ &= \left(n^{-1}\widetilde{\mathbf{X}}^T\widetilde{\mathbf{X}}\right)^{-1}\frac{1}{\sqrt{n}}(S_{n1} - S_{n2} + S_{n3}). \end{aligned} \tag{A.13}$$

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

$$\begin{aligned} \widetilde{X}_{ls} &= \eta_{ls} - \sum_{j=1}^n \mathbf{W}_j(\xi_l)\eta_{js} + g_{1s}(\xi_l) - \sum_{j=1}^n \mathbf{W}_j(\xi_l)g_{1s}(\xi_j), \\ &= \eta_{ls} - \sum_{j=1}^n \mathbf{W}_j(\xi_l)\eta_{js} + \widetilde{m}_{ls} \quad (l = 1,\dots,n, s = 1,\dots,p)\,. \end{aligned} \tag{A.14}$$

Then each element of the vectors S_{n1}, S_{n2} and S_{n3} can be decomposed into three summands, noted $S_{nq,1}$, $S_{nq,2}$ and $S_{nq,3}$ for $q = 1,2,3$, whose asymptotic behavior can be obtained from the Lemma A.1, Lemma A.2 and the Lemma A.3. More specifically, if in the Lemma A.2, we take $a_{jl} = \mathbf{W}_j(\xi_l)$, $a_n = \frac{1}{k-1}$ and $b_n = 1$ with $V_l = \epsilon_l(V_l = \eta_{ls})$, it follows that

$$\mid \overline{m}(\xi_l) \mid = O_{a.s.}\left(\phi^{-1}\left(\frac{k}{n}\right)^{\alpha} + \sqrt{\frac{\log n}{k}} + \frac{\log n}{\sqrt{k-1}}\right), \tag{A.15}$$

and

$$\mid \widetilde{m}_{ls} \mid = O_{a.a.}\left(\phi^{-1}\left(\frac{k}{n}\right)^{\alpha} + \sqrt{\frac{\log n}{k}} + \frac{\log n}{\sqrt{k-1}}\right). \tag{A.16}$$

Let us treat term S_{n1}. It follows from (A.15), (A.16) and Abel's inequality that

$$\begin{aligned} S_{n1,3} &= \sum_{l=1}^{n} \widetilde{m}_{ls}\overline{m}(\xi_l) \leq n \mid \overline{m}(\xi_l) \mid\mid \widetilde{m}_{ls} \mid \\ &= O\left(n\left(\phi^{-1}\left(\frac{k}{n}\right)^{2\alpha} + \frac{\log n}{k} + \frac{\log n^2}{k-1}\right)\right) = o(\sqrt{n}) \quad a.s. \end{aligned} \tag{A.17}$$

Again, using the same techniques, if we consider $a_{jl} = \overline{m}(\xi_l)$, $a_n = \phi^{-1}\left(\frac{k}{n}\right)^{\alpha} + \sqrt{\frac{\log n}{k}} + \frac{\log n}{\sqrt{k-1}}$, $b_n = na_n$ and $V_l = \eta_{ls}$ in lemma A.2, it follow that

$$\begin{aligned} S_{n1,1} &= \sum_{l=1}^{n} \eta_{ls}\overline{m}(\xi_l) = O\left[\left(\phi^{-1}\left(\frac{k}{n}\right)^{\alpha} + \sqrt{\frac{\log n}{k}} + \frac{\log n}{\sqrt{k-1}}\right)\sqrt{n}\log n\right] \\ &= o(\sqrt{n}) \quad a.s., \end{aligned} \tag{A.18}$$

and

$$\begin{aligned} S_{n1,2} &= \sum_{l=1}^{n}\left(\sum_{j=1}^{n} \mathbf{W}_j(\xi_l)\eta_{js}\right)\overline{m}(\xi_l) \\ &\leq n\left|\overline{m}(\xi_l)\right| \max_{1\leq j,l\leq n}\left|\sum_{j=1}^{n} \mathbf{W}_j(\xi_l)\eta_{js}\right| \\ &= O\left[\left(\phi^{-1}\left(\frac{k}{n}\right)^{\alpha} + \sqrt{\frac{\log n}{k}} + \frac{\log n}{\sqrt{k-1}}\right)\frac{n\log n}{k-1}\right] \\ &= o(\sqrt{n}) \quad a.s. \end{aligned} \tag{A.19}$$

For the term S_{n2}, similar to the arguments used to obtain A.17-A.19, given A.16 and Abel's inequality, it follows that

$$S_{n2,1} = \sum_{l=1}^{n}\left(\sum_{j=1}^{n} \mathbf{W}_j(\xi_l)\epsilon_j\right)\eta_{ts} = O\left(\frac{\sqrt{n}\log^2 n}{\sqrt{k-1}}\right) = o(\sqrt{n}) \quad a.s., \tag{A.20}$$

$$\begin{aligned} S_{n2,2} &= \sum_{l=1}^{n}\left(\sum_{j=1}^{n}\mathbf{W}_j(\xi_l)\eta_{js}\right)\left(\sum_{j=1}^{n}\mathbf{W}_j(\xi_l)\epsilon_j\right) \\ &\leq n \max_{1\leq j,l\leq n}\left|\sum_{j=1}^{n}\mathbf{W}_j(\xi_l)\eta_{js}\right| \max_{1\leq j,l\leq n}\left|\sum_{j=1}^{n}\mathbf{W}_j(\xi_l)\epsilon_j\right| \\ &= O\left(\frac{n\log^2 n}{k-1}\right) \\ &= o(\sqrt{n}) \quad a.s., \end{aligned} \tag{A.21}$$

and

$$\begin{aligned} S_{n2,3} &= \sum_{l=1}^{n}\widetilde{m}_{ls}\left(\sum_{j=1}^{n}\mathbf{W}_j(\xi_l)\epsilon_j\right) \\ &= O\left(\left(\phi^{-1}\left(\frac{k}{n}\right)^{\alpha} + \sqrt{\frac{\log n}{k}} + \frac{\log n}{\sqrt{k-1}}\right)\frac{n\log n}{\sqrt{k-1}}\right) \\ &= o(\sqrt{n}) \quad a.s. \end{aligned} \tag{A.22}$$

Finally, for S_{n3}, we have

$$S_{n3,1} = \sum_{l=1}^{n}\eta_l\epsilon_l\,, \tag{A.23}$$

$$S_{n3,2} = \sum_{l=1}^{n}\left(\sum_{j=1}^{n}\mathbf{W}_j(\xi_l)\eta_{js}\right)\epsilon_l = O\left(\frac{\sqrt{n}\log n}{\sqrt{k-1}}\right) = o(\sqrt{n}) \quad a.s., \tag{A.24}$$

and

$$\begin{aligned} S_{n3,3} &= \sum_{l=1}^{n}\widetilde{m}_{ls}\epsilon_l \\ &= O\left(\left(\phi^{-1}\left(\frac{k}{n}\right)^{\alpha} + \sqrt{\frac{\log n}{k}}\right)\frac{\sqrt{n}\log n}{\sqrt{k-1}}\right) = o(\sqrt{n}) \quad a.s. \end{aligned} \tag{A.25}$$

Then, by (A.13) and (A.17)-(A.25), it follows that

$$\sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) = \left(n^{-1}\widetilde{\mathbf{X}}^T\widetilde{\mathbf{X}}\right)^{-1}\frac{1}{\sqrt{n}}\left(\sum_{l=1}^{n}\eta_l\epsilon_l + o(\sqrt{n})\right). \tag{A.26}$$

Therefore, the proof of part (i) of the theorem 3.1 follows directly from central limit theorem and the Lemma A.4.

Part (ii): using the result A.26 and lemma A.4, the proof of these results is obtained by following the same steps as those of [7]. □

Proof. (**Theorem 3.2**)

From equation 2.11, we can write

$$\begin{aligned} |\widehat{m}(\xi_l) - m(\xi_l)| &\leq |\textstyle\sum_{j=1}^{n}\mathbf{W_j}(\xi_l)\mathbf{X_j}^T(\widehat{\beta}_n - \beta)| + |\textstyle\sum_{j=1}^{n}\mathbf{W_j}(\xi_l)(m(\xi_l) + \varepsilon_l)| \\ &\leq |S_1| + |S_2|\,. \end{aligned} \tag{A.27}$$

On the one hand, we have

$$\begin{array}{lll} |S_1| & \leq & \max_{1\leq j,l\leq n} |\sum_{j=1}^{n} \mathbf{W_j}(\xi_l)\mathbf{X_j}|||\widehat{\beta}_n - \beta|| \\ & \leq & \max_{1\leq j,t\leq n} |\sum_{j=1}^{n} \mathbf{W_j}(\xi_l)\left(\mathbf{X_j} - \mathbb{E}(\mathbf{X}_l/\xi_l)\right)|||\widehat{\beta}_n - \beta|| + ||\mathbb{E}(\mathbf{X}_l/\xi_l)||||\widehat{\beta}_n - \beta||\,. \end{array}$$

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

$$|S_1| \to 0 \; a.s. \tag{A.28}$$

In other hand, from Lemma A.1, we have

$$|S_2| = O_{a.s.}\left(\phi^{-1}\left(\frac{k}{n}\right)^{\alpha} + \sqrt{\frac{\log n}{k}}\right)\,. \tag{A.29}$$

So by using Equations A.27, A.28 and A.29, we have

$$|\widehat{m}(\xi) - m(\xi)| = O_{a.s.}\left(\phi^{-1}\left(\frac{k}{n}\right)^{\alpha} + \sqrt{\frac{\log n}{k}}\right)\,.$$

$\square$