Stat Papers (2020) 61:423–444
https://doi.org/10.1007/s00362-017-0946-0

REGULAR ARTICLE

# *k*NN estimation in functional partial linear modeling

**Nengxiang Ling**[1] · **Germán Aneiros**[2] · **Philippe Vieu**[3]

Received: 17 January 2017 / Revised: 31 May 2017 / Published online: 29 August 2017


**Abstract** A statistical procedure combining the local adaptivity and the easiness of implementation of $k$-nearest-neighbours ($k$NN) estimates together with the semi-parametric flexibility of partial linear modeling is developed for regression problems involving functional variable. Various asymptotic results are stated, both for the linear parameters and for the nonparametric operator involved in the model. A simulation study compares the finite sample behaviour of the $k$NN method with alternative estimation procedures. Finally, comparison with alternative functional regression models is carried out by means of a real curves data application which exhibits the interest both of the $k$NN method and of the semi-parametric modeling.

**Keywords** $k$NN estimate · Functional data analysis · Partial linear regression · Semi-parametrics

✉ Nengxiang Ling
hfut.lnx@163.com

Germán Aneiros
ganeiros@udc.es

Philippe Vieu
philippe.vieu@math.univ-toulouse.fr

1 School of Mathematics, Hefei University of Technology, Hefei, China

2 Departamento de Matemáticas, Universidade da Coruña, A Coruña, Spain

3 Institut de Mathématiques, Université Paul Sabatier, Toulouse, France

## 1 Introduction

This paper focuses on the semi-functional partially linear regression (SFPLR) model

$$Y = \sum_{j=1}^{p} X_j \beta_j + m(\boldsymbol{\zeta}) + \varepsilon, \tag{1.1}$$

where $X_j$ $(j = 1, 2, \ldots, p)$ are real explanatory variables, $\boldsymbol{\zeta}$ is another explanatory variable but of functional feature, $\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots, \beta_p)^T$ is a vector of unknown real parameters, $m(\cdot)$ is an unknown smooth functional operator and $\varepsilon$ is the random error with $E(\varepsilon) = 0$ and unknown variance $0 < Var(\varepsilon) < \infty$. We assume that $\boldsymbol{\zeta}$ is valued in $S_{\mathcal{F}} \subset \mathcal{F}$, where $\mathcal{F}$ is some abstract infinite-dimensional space with associated semi-metric denoted by $d(\cdot, \cdot)$. Model (1.1) was first introduced in Engle et al. (1986) in the setting where the dimension of $\mathcal{F}$ is finite and, since then, many attention has been dedicated to it. See, for instance, the monograph by Härdle et al. (2000) and the references therein for the finite-dimensional case.

In recent years, however, modern sciences allow us to collect continuous data (such as curves, surface and images) and statisticians are in front of new challenges. This field, which is called functional data analysis (FDA), has been popularized in the last two decades (see Ramsay and Silverman 2005 for pioneer developments; Cuevas 2014; Goia and Vieu 2016; Aneiros et al. 2017 for discussions on the recent literature in the field). FDA is receiving an increasing interest in recent literature and many approaches have been revisited for this infinite dimensional framework, and one has now for instance at hand wide scope of contributions in functional nonparametric modeling (see e.g. Ferraty and Vieu 2006), in functional variance analysis (see e.g. Zhang 2013), in dependent functional data analysis (see e.g. Bosq and Blanke 2007; Horváth and Kokoszka 2012). The recent monograph by Hsing and Eubank (2015) provides extensive mathematical materials for theoretical study of functional data. In functional regression, semiparametric models have been recently introduced (see Goia and Vieu 2014, for a short survey) in order to balance the too few flexibility of linear regression and the too high sensitivity to dimensional effects of nonparametric approaches. The partial linear model (1.1) takes a large part in this functional semiparametric literature. For example, Aneiros-Pérez and Vieu (2006) presented their first work on model (1.1). More precisely, the estimators of unknown parametric and unknown nonparametric regression operator are constructed based a functional version of Nadaraya-Watson (N-W)-type weights method, and some asymptotic results with convergence rates are given under a sample of i.i.d. Aneiros-Pérez and Vieu (2011) and Shang (2014) proposed and studied cross-validation and Bayesian bandwidth selectors, respectively. Lian (2011) extended the model (1.1) to the case where the linear variable is also of functional nature.

On the other hand, an important point when dealing with infinite-dimensional variable is the need for taking into account local structures of the data. In nonparametric setting one way to do that is to construct location adaptive methods. One of the most

popular way to do is to consider $k$ nearest neighbour ($k$NN) ideas. Compared with the the standard N–W kernel method, the main appealing feature of the $k$NN method consists in its local adaptive bandwidth parameter (a bandwidth depending on sample) which is controlled by a discrete parameter $k \in \mathbb{N}$ (rather than by a continuous one $h$). The main difficulty comes from the randomness of the new bandwidth parameter which avoids for decomposing estimates into sums of independent variables and which therefore requires an additional high level of technicality to investigate its theoretical properties. In multivariate nonparametric estimation this question has been addressed in various different ways. The precursor work goes back to Cover (1968) and a set of important advances includes Collomb (1979) or Devroye et al. (1994) in regression estimation, Devroye and Wagner (1977) or Berlinet and Servien (2011) in density estimation, Collomb et al. (1985) or Dette and Gefeller (1995) in hazard estimation, Devroye and Wagner (1982) in discrimination, Paindaveine and Van Bever (2015) in unsupervised classification, Li et al. (2016) for model specification, Sancetta (2010) for time series analysis, ... The monograph by Györfi et al. (2002) proposes a wide presentation on $k$NN ideas in nonparametric finite dimensional statistics.

Starting with Laloë (2008) and Burba et al. (2009), there are now a few theoretical advances in the functional data analysis setting; the main challenge being to look whether and how $k$NN ideas could be adapted to deal infinite dimensional variables. These few advances concern mainly application of the $k$NN method to the functional nonparametric regression model, including the construction of the $k$NN kernel estimator of regression operator and the obtention of its asymptotic properties such as, for instance, uniform almost-complete convergence with rates (see Biau et al. 2010; Kudraszow and Vieu 2013; Attouch and Benchikh 2012, for recent advances on $k$NN functional regression); see also Cerou and Guyader (2006) for similar approach in the related curves classification problem. The most recent results in $k$NN nonparametric functional data analysis can be found in Kara-Zaitri et al. (2017a) and references therein. However, as far as we know, all the existing literature concerns only nonparametric functional approaches and there is currently no advances in this direction in the semiparametric framework, as it could be the case in multivariate statistics (see e.g. Hong 1992; Robinson 1995 or Chu et al. 2013). The aim of this paper is to make a first step in this direction by developing $k$NN procedures for the model (1.1). More precisely, we construct estimators of both unknown parameters and unknown regression operator in (1.1) with the aid of $k$NN and least squares approach, and some asymptotic properties such as convergence with rates of such estimators are obtained.

The organization of the paper is as follows. In Sect. 2, we construct precisely the $k$NN estimators based on a sample of i.i.d. random vectors, while in Sect. 3 we present the main theoretical results of the paper: asymptotic normality and law of iterated logarithm for the estimator of the parameter vector $\boldsymbol{\beta}$, and uniform almost-complete convergence with rates for the estimator of the functional operator $m(\cdot)$. A simulation study designed to highlight how the local feature of the $k$NN approach allows for nice improvement upon usual global techniques is described in Sect. 4. The usefulness in practice of the proposed methodology is reported in Sect. 5 by means of an application to real data. Some concluding comments are given in Sect. 6. The proofs of the main results are postponed to Sect. 7.

## 2 Model and estimators

Let $\{(Y_i, X_{i1}, X_{i2}, \ldots, X_{ip}, \boldsymbol{\zeta}_i)^T,\ 1 \le i \le n\}$ be a sequence of i.i.d. random vectors from $(Y, X_1, X_2, \ldots, X_p, \boldsymbol{\zeta})^T$ in the SFPLR model (1.1). Thus, we have that

$$Y_i = X_{i1}\beta_1 + X_{i2}\beta_2 + \ldots + X_{ip}\beta_p + m(\boldsymbol{\zeta}_i) + \varepsilon_i,\ (i = 1, 2, \ldots, n),$$

where

$$E\,(\varepsilon_i) = 0 \text{ and } 0 < Var(\varepsilon_i) < \infty.$$

Throughout the paper, we suppose that $\{(X_{i1}, X_{i2}, \ldots, X_{ip}, \boldsymbol{\zeta}_i)^T, 1 \le i \le n\}$ is independent of $\{\varepsilon_i, 1 \le i \le n\}$.

Before building the estimators to be studied in this paper, let us introduce some notations. We denote

$$\omega_{n,k}(\zeta, \boldsymbol{\zeta}_i) = \frac{K\left(\frac{d(\zeta,\boldsymbol{\zeta}_i)}{H_{n,k}(\zeta)}\right)}{\sum_{j=1}^{n} K\left(\frac{d(\zeta,\boldsymbol{\zeta}_j)}{H_{n,k}(\zeta)}\right)}\ (i = 1, \ldots, n), \tag{2.1}$$

where $K(\cdot)$ is an asymmetrical kernel function and $H_{n,k}(\zeta) > 0$ is a random bandwidth parameter defined as

$$H_{n,k}(\zeta) = \min\left\{h > 0; \sum_{i=1}^{n} I_{B(\zeta,h)}(\boldsymbol{\zeta}_i) = k\right\},$$

with $B\,(\zeta, h) = \left\{z' \in \mathcal{F}; d\left(z', \zeta\right) \le h\right\}$. In addition, we denote

$$\mathbf{X}_i = (X_{i1}, \ldots, X_{ip})^T,\ \widetilde{\mathbf{X}}_i = \mathbf{X}_i - \sum_{j=1}^{n} \omega_{n,k}(\boldsymbol{\zeta}_i, \boldsymbol{\zeta}_j)\mathbf{X}_j,$$

$$\widetilde{Y}_i = Y_i - \sum_{j=1}^{n} \omega_{n,k}(\boldsymbol{\zeta}_i, \boldsymbol{\zeta}_j)Y_j\ (i = 1, \ldots, n),$$

$$\widetilde{\mathbf{X}} = (\widetilde{\mathbf{X}}_1, \ldots, \widetilde{\mathbf{X}}_n)^T \text{ and } \widetilde{\mathbf{Y}} = (\widetilde{Y}_1, \cdots, \widetilde{Y}_n)^T.$$

If $\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots, \beta_p)^T$ is known, the $k$NN kernel estimator of $m(\cdot)$ can be written as

$$\begin{aligned}\widetilde{m}_n(\zeta) \equiv \widetilde{m}_n(\zeta, \boldsymbol{\beta}) &= \sum_{i=1}^{n} \omega_{n,k}(\zeta, \boldsymbol{\zeta}_i)(Y_i - \mathbf{X}_i^T\boldsymbol{\beta}) \\ &= \widehat{g}_{1n}(\zeta) - \widehat{g}_{2n}(\zeta)^T\boldsymbol{\beta},\end{aligned} \tag{2.2}$$

where $\widehat{g}_{1n}(\zeta) = \sum_{i=1}^{n} \omega_{n,k}(\zeta, \boldsymbol{\zeta}_i)Y_i$ and $\widehat{g}_{2n}(\zeta) = \sum_{i=1}^{n} \omega_{n,k}(\zeta, \boldsymbol{\zeta}_i)\mathbf{X}_i$ are the $k$NN kernel estimators of $g_1(\zeta) = E(Y_1|\boldsymbol{\zeta}_1 = \zeta)$ and $g_2(\zeta) = E(\mathbf{X}_1|\boldsymbol{\zeta}_1 = \zeta)$, respectively.

In practice, $\boldsymbol{\beta}$ is unknown, but we can estimate it by means of the least squares method, which is defined by

$$\widehat{\boldsymbol{\beta}}_n = \arg_{\boldsymbol{\beta}} \sum_{j=1}^{n} \left(Y_j - \mathbf{X}_j^T \boldsymbol{\beta} - \widetilde{m}_n(\boldsymbol{\zeta}_j, \boldsymbol{\beta})\right)^2 = \arg_{\boldsymbol{\beta}} (\widetilde{\mathbf{Y}} - \widetilde{\mathbf{X}}\boldsymbol{\beta})^T (\widetilde{\mathbf{Y}} - \widetilde{\mathbf{X}}\boldsymbol{\beta}) = (\widetilde{\mathbf{X}}^T \widetilde{\mathbf{X}})^{-1} \widetilde{\mathbf{X}}^T \widetilde{\mathbf{Y}}. \tag{2.3}$$

(Note that the second equality in (2.3) is a consequence of (2.2)) Finally, the estimator of operator $m(\cdot)$ is obtained by

$$\widehat{m}_n(\zeta) = \widehat{m}_n(\zeta, \widehat{\boldsymbol{\beta}}_n) = \sum_{i=1}^{n} \omega_{n,k}(\zeta, \boldsymbol{\zeta}_i)(Y_i - \mathbf{X}_i^T \widehat{\boldsymbol{\beta}}_n) = \widehat{g}_{1n}(\zeta) - \widehat{g}_{2n}(\zeta)^T \widehat{\boldsymbol{\beta}}_n. \tag{2.4}$$

*Remark 1* It is clear that although the number of neighbors $k$ is fixed, the bandwidth $H_{n,k}(\zeta)$ changes as $\zeta$ does. This fact gives the local adaptive property of the $k$NN-based estimators. Another advantage of the $k$NN-based estimators against the kernel-based ones appears when they are put in practice. We refer to the computational cost to select the smoothing parameter $k$ or $h$, respectively. In SFPLR models, it is usual to choose such parameter from the cross-validation procedure. So, one must to minimize a loss function over a discrete or a continuous set, respectively. Numeric algorithms are faster when the parameter to find is discrete. Finally, it is worth being noted what is the price to pay by these nice features of the $k$NN-based estimators: remembering that $H_{n,k}(\zeta)$ is a random variable (which depends on $(\boldsymbol{\zeta}_1, \ldots, \boldsymbol{\zeta}_n)$), one should expect that additional technical difficulties will appear along the proofs of asymptotic properties. To fix the idea on this point, note that the random elements involved in the nonparametric components of the method (i.e. (2.1), (2.2) and (2.4)) can not be decomposed as sums of independent variables (as it is the case for instance with kernel-based estimators), and hence its treatment will need more sophisticated probabilistic developments than standard limit theorems for sums of iid variables (see Sect. 7).

## 3 Asymptotic properties

This section is devoted to establish some asymptotic properties of the estimators $\widehat{\boldsymbol{\beta}}_n$ and $\widehat{m}_n(\zeta)$. For that, we will use the following general technical assumptions, where, for any set $S \subset \mathcal{F}$ and $\epsilon > 0$, $\psi_S(\epsilon) = \log(N_\epsilon(S))$ denotes the Kolmogorov's $\epsilon$-entropy of the set S ($N_\varepsilon(S)$ is the minimal number of open balls in $\mathcal{F}$ of radius $\epsilon$ which is necessary to cover $S$), while $g_{2j}\left(\boldsymbol{\zeta}_i\right) = E\left(X_{ij} \mid \boldsymbol{\zeta}_i\right)$ and $\boldsymbol{\eta}_i = \left(\eta_{i1}, \ldots, \eta_{ip}\right)^T$ with $\eta_{ij} = X_{ij} - g_{2j}\left(\boldsymbol{\zeta}_i\right)$ $(i = 1, \ldots, n,\ j = 1, \ldots, p)$. Other specific assumptions will be stated along the theorem below.

### 3.1 Technical assumptions

(A1) $\forall \epsilon > 0$, $\varphi_\zeta(\epsilon) := P(\boldsymbol{\zeta} \in B(\zeta, \epsilon)) > 0$, with $\varphi_\zeta(\cdot)$ continuous on a neighbourhood of 0 and $\varphi_\zeta(0) = 0$.

(A2) There exist a nonnegative function $\phi(\cdot)$ regularly varying at 0 with nonnegative index, a positive function $g(\cdot)$ and a positive number $\alpha$ such that:

(A2.i) $\phi(0) = 0$ and $\lim_{\epsilon \to 0} \phi(\epsilon) = 0$.
(A2.ii) $\exists\, C > 0$ and $\exists\, \eta_0 > 0$ such that, $\forall 0 < \eta < \eta_0$, $\phi'(\eta) < C$.
(A2.iii) $\sup_{\zeta \in S_{\mathcal{F}}} |\frac{\varphi_\zeta(\epsilon)}{\phi(\epsilon)} - g(\zeta)| = O(\epsilon^\alpha)$ as $\epsilon \to 0$.
(A2.iv) $\exists C < \infty$ such that $\forall\, (u, v) \in S_{\mathcal{F}} \times S_{\mathcal{F}}$, $\forall\, f \in \left\{m, g_{21}, \ldots, g_{2p}\right\}$, $|f(u) - f(v)| \leq C d(u, v)^\alpha$.
(A3) The kernel function $K(\cdot)$:
(A3.i) $K(\cdot)$ is a nonnegative, bounded and non increasing function with support $[0, 1]$ and Lipschitz on $[0, 1)$.
(A3.ii) If $K(1) = 0$, it must also be such that $-\infty < C < K'(t) < C' < 0$.
(A4) Kolmogorov's $\epsilon$-entropy of $S_{\mathcal{F}}$ satisfies

$$\sum_{n=1}^{n} \exp\left\{(1-\omega)\psi_{S_{\mathcal{F}}}\left(\frac{\log n}{n}\right)\right\} < \infty \text{ for some } \omega > 1.$$

(A5) $k = k_n$ is a sequence of positive real numbers such that:
(A5.i) $\frac{k}{n} \to 0$ and $\frac{\log n}{k} \to 0$ as $n \to \infty$.
(A5.ii) For $n$ large enough, $\frac{(\log n)^2}{k} < \psi_{S_{\mathcal{F}}}(\frac{\log n}{n}) < \frac{k}{\log n}$.
(A6) Moment conditions:
(A6.i) $\forall r \geq 3,\ 1 \leq j \leq p$, $E(|Y_1|^r | \boldsymbol{\zeta}_1 = \zeta) \leq C_1 < \infty$ and $E(|X_{1j}|^r | \boldsymbol{\zeta}_1 = \zeta) \leq C_2 < \infty$.
(A6.ii) $\boldsymbol{\Sigma} = E\left(\boldsymbol{\eta}_1 \boldsymbol{\eta}_1^T\right)$ is a positive definite matrix.

*Remark 2* As the reader will see in Sect. 7, Theorem 2 in Kudraszow and Vieu (2013) plays a main role in order to prove our asymptotic results. So, we need to impose the same assumptions as in such theorem; we refer to assumptions (A1)–(A5) (for comments on these hypotheses, see Kudraszow and Vieu 2013). The additional assumption (A6) is usual in the SFPLR context (see, for instance, Aneiros-Pérez and Vieu 2006). It imposes moment conditions on the scalar variables in the model and certainly it is not restrictive.

### 3.2 Main results

**Theorem 1** *Under assumptions (A1)–(A6), if in addition* $\frac{\sqrt{n}\log^2 n}{k} \to 0$, $\sqrt{n}\phi^{-1}(\frac{k}{n})^\alpha \to 0$, $\frac{\sqrt{n}\psi_{S_{\mathcal{F}}}(\frac{\log n}{n})}{k} \to 0$ *as* $n \to \infty$ *and* $k \geq n^{(2/r)+b}/\log^2 n$ *for* $n$ *large enough and some constant* $b > 0$ *with* $\frac{2}{r} + b > \frac{1}{2}$ *(where* $r \geq 3$*), then we have*

(i) $\sqrt{n}(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}) \to N(0, \sigma^2 \boldsymbol{\Sigma}^{-1})$,
(ii) $\limsup_{n\to\infty} \left(\frac{n}{2\log\log n}\right)^{\frac{1}{2}} |\widehat{\beta}_{nj} - \beta_j| = (\sigma^2 \sigma_{jj})^{\frac{1}{2}}$ *a.s., where* $\sigma_{jj} = (\boldsymbol{\Sigma}^{-1})_{jj}$ *for* $j = 1, 2, \ldots, p$

*and*

(iii) $\sup_{\zeta \in S_{\mathcal{F}}} |\widehat{m}_n(\zeta) - m(\zeta)| = O\left(\phi^{-1}(\frac{k}{n})^\alpha + \sqrt{\frac{\psi_{S_{\mathcal{F}}}(\frac{\log n}{n})}{k}}\right)$ *a.s.*

*Remark 3* Firstly, it is worth noting that Theorem 1 extends in two ways the results in Aneiros-Pérez and Vieu (2006). On the one hand, a more general assumption on the

topology of $S_{\mathcal{F}}$ is considered (based on the notion of Kolmogorov's entropy instead of imposing a more restrictive condition on a set of balls covering $S_{\mathcal{F}}$). On the other hand, $k$NN-based estimators are studied (instead of kernel-based ones). Secondly, Theorem 1(i,ii) shows that the $k$NN estimator of $\boldsymbol{\beta}$ presents a similar asymptotic behavior as that of the kernel estimator (see Aneiros-Pérez and Vieu 2006). Finally, Theorem 1(iii) indicates that, from an asymptotic point of view, the fact that $\boldsymbol{\beta}$ is unknown does not affect to the rate of uniform convergence of the estimator of $m(\cdot)$ (see Kudraszow and Vieu 2013).

*Remark 4* It is worth to be noted that, in order to use the result established in Theorem 1(i) for making statistical inference on $\boldsymbol{\beta}$, consistent estimators for $\boldsymbol{\Sigma}$ and $\sigma^2$ are needed. For that, one can consider the consistent kernel-based estimators proposed in Aneiros and Vieu (2006) (see Lemma 7 in that paper) and Aneiros et al. (2015) (see Theorem 1 in that paper), respectively. Another option is to use $k$NN-based estimators. In that case, Lemma 3 in Sect. 7.1 gives the consistence for a $k$NN-based estimator for $\boldsymbol{\Sigma}$. In addition, one can construct a $k$NN-based estimator for $\sigma^2$ by considering, in the kernel-based estimator proposed in Aneiros et al. (2015), $k$NN-based residuals instead of kernel-based ones. The consistence of such estimator can be obtained in a similar way as in Theorem 1 in Aneiros et al. (2015).

## 4 Simulation study

This section is devoted to illustrate the finite sample behaviour of the proposed $k$NN-based methodology against the corresponding kernel-based one studied in Aneiros-Pérez and Vieu (2006). Specifically, we will compare the accurate of the estimators of $\boldsymbol{\beta}$ and $m(\cdot)$ in the SFPLR model (1.1), as well as the predictive power of the SFPLR model, when estimated from the $k$NN- and kernel-based estimators. The simulation study is designed to highlight how the local feature of the $k$NN approach allows for nice improvement when applied on datasets of curves that change from one area of the space to another one (this fact being usual in practice when dimension increases, as is the case of functional data). As a byproduct, the behaviour of the cross-validation selector for choosing the number of neighbors, $k$, is also shown.

First, note that, to construct the referred kernel-based estimators, all one has to do is to change in (2.3) and (2.4) the weights $\omega_{n,k}(\zeta, \boldsymbol{\zeta}_i)$ (see (2.1)) by the ones

$$\omega_{n,h}(\zeta, \boldsymbol{\zeta}_i) = \frac{K\left(\frac{d(\zeta,\boldsymbol{\zeta}_i)}{h}\right)}{\sum_{j=1}^{n} K\left(\frac{d(\zeta,\boldsymbol{\zeta}_j)}{h}\right)} \ (i = 1, \ldots, n),$$

where $h > 0$ is the smoothing parameter.

### 4.1 The simulated model

We simulate $N = 300$ vectors, $\{(Y_i, X_i, \boldsymbol{\zeta}_i)\}_{i=1}^{N}$, verifying the SFPLR model

$$Y = X\beta + m(\boldsymbol{\zeta}) + \varepsilon. \tag{4.1}$$

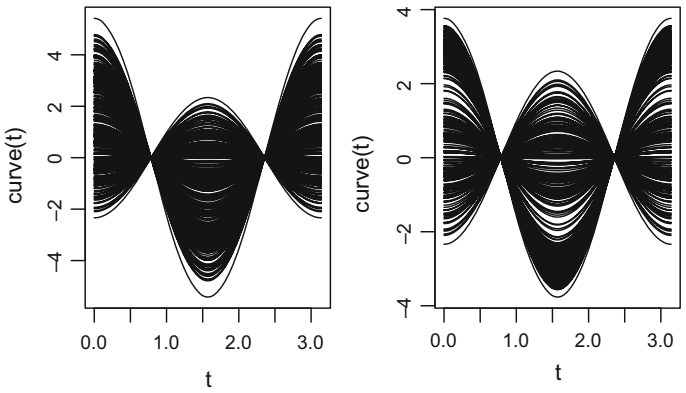


**Fig. 1** Simulated curves for the most homogeneous case ($\sigma_a^2 = 1$; *left panel*) and the most heterogeneous one ($\sigma_a^2 = 0.1$; *right panel*)

The curves $\{\boldsymbol{\zeta}_i\}_{i=1}^N$ were

$$\boldsymbol{\zeta}_i(t) = a_i \cos(2t), \ t \in [0, 2\pi]$$

where the random variables $\{a_i\}_{i=1}^{N/2}$ and $\{a_j\}_{j=N/2+1}^N$ were i.i.d. according to $N(0, 1)$ and $N(3, \sigma_a)$ distributions, respectively. Note that, although the dataset $\{\boldsymbol{\zeta}_i\}_{i=1}^N$ is heterogeneous, such heterogeneity decreases as $\sigma_a^2$ goes to 1, while it increases as $\sigma_a^2$ goes to 0 (for a nice study on the effect of $\sigma_a^2$ on the concentration of the curves, see Burba et al. 2009). In this simulation study, values $\sigma_a^2 = 1$ and $\sigma_a^2 = 0.1$ were considered; in addition, the curves were observed on a grid generated from 100 equispaced points in $[0, 2\pi]$ (for graphics displaying the curves, see Fig. 1). The nonparametric operator $m(\cdot)$ was

$$m(\boldsymbol{\zeta}_i) = a_i^2 \ (i = 1, \ldots, N).$$

(Note that we are considering the same both curves and nonparametric operator as in Burba et al. 2009)

The scalar covariates $\{X_i\}_{i=1}^N$ were i.i.d. according to a $N(0, 1)$ distribution. The value of the parameter $\beta$ was 3. Finally, the independent random errors $\{\varepsilon_i\}_{i=1}^N$ were generated from a $N(0, \sigma_\varepsilon)$ distribution, where $\sigma_\varepsilon = c\sigma_r$ with $\sigma_r$ denoting the standard deviation of the regression functional $r(X_i, \boldsymbol{\zeta}_i) = X_i\beta + m(\boldsymbol{\zeta}_i)$ (note that $c$ can be seen as the signal-to-noise ratio). Values $c = 0.025$, $c = 0.05$ and $c = 0.1$ were considered.

The experiment was replicated $M = 500$ times.

### 4.2 Tuning parameters selection

$k$NN-based (kernel-based) estimators depend on two tuning parameters, which must be selected from the observed data. Those tuning parameters consist of the quantity of

neighbours, $k$, (the bandwidth, $h$,) and the semi-metric, $d(\cdot,\cdot)$, to be considered in the estimators. On the one hand, because the simulated curves are smooth (see Fig. 1), we considered (as recommended in Ferraty and Vieu 2006) a class of semi-metrics based on derivatives, $\{d_q^{[0,2\pi]}(\cdot,\cdot),\ q=0,1\}$, where

$$d_q^{[a,b]}(\boldsymbol{\zeta}_{j_1},\boldsymbol{\zeta}_{j_2})=\left(\int_a^b(\boldsymbol{\zeta}_{j_1}^{(q)}(t)-\boldsymbol{\zeta}_{j_2}^{(q)}(t))^2dt\right)^{1/2}. \tag{4.2}$$

On the other hand, the size of the grid where the parameter $h$ was searched was twofold the size of the grid corresponding to the parameter $k$ (remember that $k$ and $h$ move on a discrete and continuous space, respectively; so, to achieve accuracy in the selection of $h$, more values than in the case of $k$ must be considered). The Epanechnikov kernel was considered in both the $k$NN and the kernel procedures.

Each of the $M=500$ simulated datasets, $\{(Y_i,X_i,\boldsymbol{\zeta}_i)\}_{i=1}^{N=300}$, was randomly split into two samples: a training sample, $\{(Y_i,X_i,\boldsymbol{\zeta}_i),\ i\in\mathcal{S}_{train}\}$, and a test sample, $\{(Y_j,X_j,\boldsymbol{\zeta}_j),\ j\in\mathcal{S}_{test}\}$, in such a way that $\#(\mathcal{S}_{train})=n=250$, $\mathcal{S}_{train}\cup\mathcal{S}_{test}=\{1,\ldots,300\}$ and $\mathcal{S}_{train}\cap\mathcal{S}_{test}=\emptyset$. The training sample was used to choose the tuning parameters ($(k,q)$ and $(h,q)$ for the $k$NN- and kernel-based estimators, respectively); then, the test sample allowed to measure the accurate of the predictions.

Specifically, the selected tuning parameters $(\widehat{k},\widehat{q})$ of the $k$NN-based estimators were obtained by minimizing a global cross-validation criterion on the training sample; that is:

$$(\widehat{k},\widehat{q})=\text{argmin}_{k,q}CV(k,q),\quad\text{where } CV(k,q)=\sum_{i\in\mathcal{S}_{train}}(Y_i-\widehat{r}_{k,q}^{(-i)}(X_i,\boldsymbol{\zeta}_i))^2 \tag{4.3}$$

with $\widehat{r}_{k,q}^{(-i)}(\cdot,\cdot)$ denoting the $k$NN-based leave-one-out estimator, with tuning parameters $k$ and $q$, of the regression function obtained from the training sample once the observation $(Y_i,X_i,\boldsymbol{\zeta}_i)$ is eliminated. Note that the selected tuning parameters $(\widehat{h},\widehat{q})$ of the kernel-based estimators can be defined in a similar way (by changing $k$ by $h$ in the expressions above). In both cases, the value $\widehat{q}=0$ was selected.

At this point, it is worth being noted that, in the setting of SFPLR models, Aneiros-Pérez and Vieu (2011) obtained the asymptotic optimality of selectors for the smoothing parameter $h$ based on global and local cross-validation ideas. Nevertheless, to the best of our knowledge, no results (neither theoretical nor empirical ones) exist for the case of the choice of the number of neighbors, $k$, by means of the cross-validation procedure. So, before to attack the aim of this Sect. 4 ($k$NN-based methodology vs kernel-based one), we will show the behaviour of such data-driven $k$ selector when applied on the simulated data introduced in the previous Sect. 4.1.

Figure 2 displays boxplots constructed from the Mean Square Error of Prediction (MSEP) obtained from each of the $M=500$ replicates of the experiment and the combinations of signal-to-noise ratios ($c=0.025,0.05,0.1$) and measures of homogeneity ($\sigma_a^2=1,0.1$). Note that

$$\text{MSEP}=\frac{1}{\#(\mathcal{S}_{test})}\sum_{j\in\mathcal{S}_{test}}(Y_j-\hat{Y}_j)^2, \tag{4.4}$$

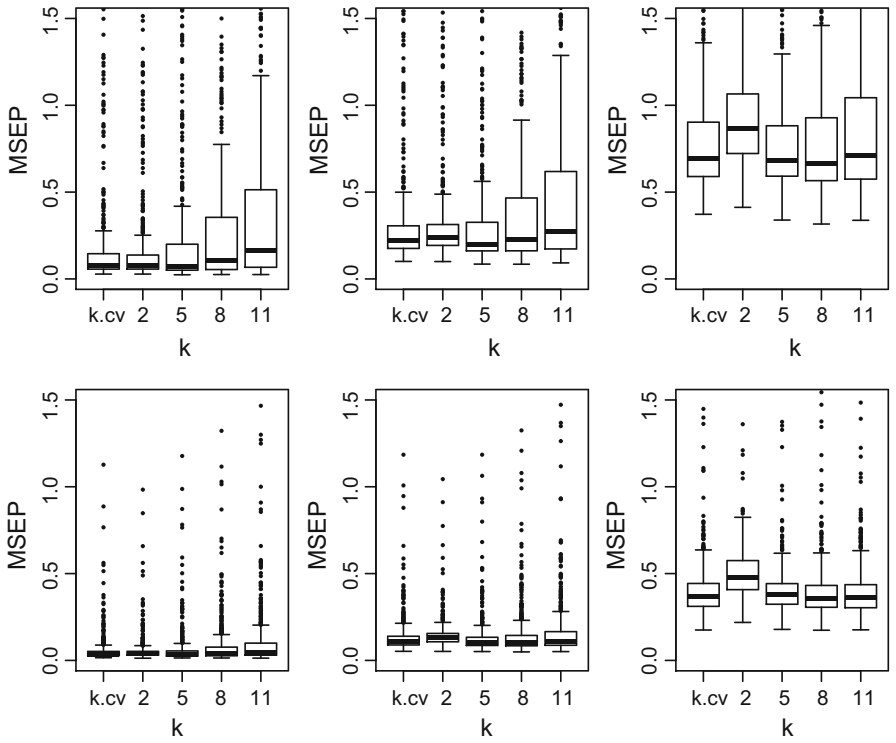


**Fig. 2** (Predictive power of the $k$NN-based estimators for various values of $k$) Mean square errors of prediction from the $M = 500$ replicates of the experiment. *Top panels* $\sigma_a^2 = 1$ (the most homogeneous case); *bottom panels* $\sigma_a^2 = 0.1$ (the most heterogeneous case). From *left* to *right*, values $c = 0.025, 0.05$ and $0.1$ are considered

where $\hat{Y}_j$ denotes a prediction of $Y_j$. Specifically, for each $j \in \mathcal{S}_{test}$, several predictions $\hat{Y}_j = \hat{Y}_j(k)$ were obtained from the SFPLR model (4.1) and the proposed $k$NN estimators (2.3) and (2.4). Such predictions differ because different values for $k$ were considered. On the one hand, $k = k_{cv}$ was selected from the cross-validation procedure (so, $k_{cv}$ depends on the replicate). On the other hand, fixed values $k = 2, 5, 8, 11$ were also used. Figure 2 shows that the $k$NN-based estimators with number of neighbors selected from the cross-validation procedure are competitive estimators. In fact, in all the six considered scenarios, the predictive power of such proposal is similar to that of the best one.

### 4.3 Results

As noted in Sect. 4.1, for each combination of signal-to-noise ratios ($c = 0.025, 0.05, 0.1$) and measures of homogeneity ($\sigma_a^2 = 1, 0.1$) in the dataset of curves, $M = 500$ samples $\{(Y_i, X_i, \boldsymbol{\zeta}_i)\}_{i=1}^N$ were drawn. Then, following the indications given in the previous Sect. 4.2, $M$ training samples (and their corresponding $M$ test samples)

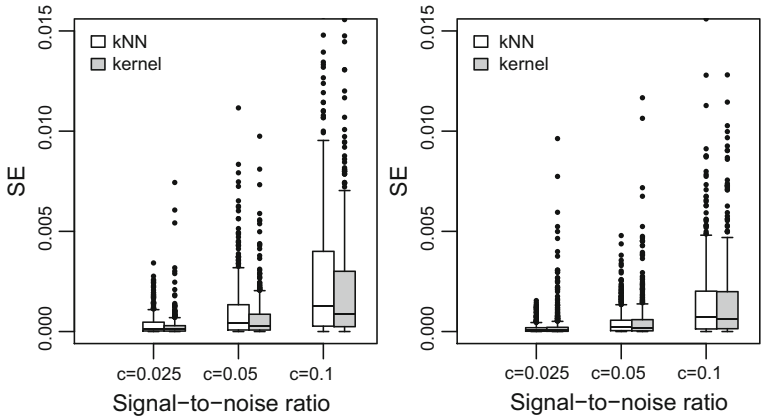


**Fig. 3** ($\beta$ estimation) Square errors from the $M = 500$ replicates of the experiment. *Left panel* $\sigma_a^2 = 1$ (the most homogeneous case). *Right panel* $\sigma_a^2 = 0.1$ (the most heterogeneous case)

were obtained. From each training sample and for both $k$NN- and kernel-based procedures, the tuning parameters were selected by means of the global cross-validation procedure, both $\beta$ and $m(\boldsymbol{\zeta}_j)$ ($j \in \mathcal{S}_{test}$) were estimated, and $Y_j$ ($j \in \mathcal{S}_{test}$) was predicted. On the one hand, the accurate of estimates, $\widehat{\beta}$ and $\widehat{m}(\cdot)$, of $\beta$ and $m(\cdot)$ was measured through the Square and Mean Square Errors (SE and MSE):

$$\text{SE} = |\widehat{\beta} - \beta|^2 \text{ and MSE} = \frac{1}{\#(\mathcal{S}_{test})} \sum_{j \in \mathcal{S}_{test}} (\widehat{m}(\boldsymbol{\zeta}_j) - m(\boldsymbol{\zeta}_j))^2,$$

respectively. On the other hand, the accurate of predictions, $\hat{Y}_j$, of $Y_j$ was measured by means of the MSEP (see 4.4).

Figures 3 and 4 display boxplots of the corresponding $M$ values of SE and MSE, respectively. As expected, accuracy of the estimators of $\beta$ and $m(\cdot)$ gets worse as the signal-to-noise ratio increases. In the particular case of $\beta$ estimation, there are no large differences in the behaviour of the $k$NN- and kernel-based estimators (see Fig. 3). Note that this conclusion agrees with the known fact that the smoothing parameter (i.e., $k$ or $h$) has little effect on the $\beta$ estimator. Focusing now in the estimation of the nonparametric component $m(\cdot)$ (see Fig. 4), the better behaviour of the $k$NN-based estimator against the kernel-based one is apparent. Note that, as in pure nonparametric models (see Burba et al. 2009), the difference in their behaviours is dramatic in scenarios of high heterogeneity in the dataset of curves.

Figure 5 shows boxplots of the MSEPs. The main conclusion is the superiority, from the predictive power point of view, of the SFPLR model when estimated from the proposed $k$NN methodology against the kernel one. Again, this superiority is dramatic in scenarios of high heterogeneity in the dataset of curves.

## 5 Real data analysis

The aim of this section is to compare, on a real dataset, the predictive power of the SFPLR model (1.1) when estimated from the $k$NN-based estimators (2.3) and (2.4)

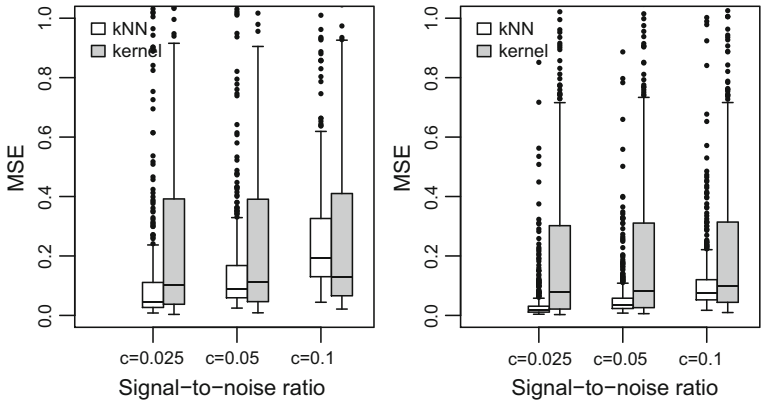


**Fig. 4** ($m(\cdot)$ estimation) Mean square errors from the $M = 500$ replicates of the experiment. *Left panel* $\sigma_a^2 = 1$ (the most homogeneous case). *Right panel* $\sigma_a^2 = 0.1$ (the most heterogeneous case)

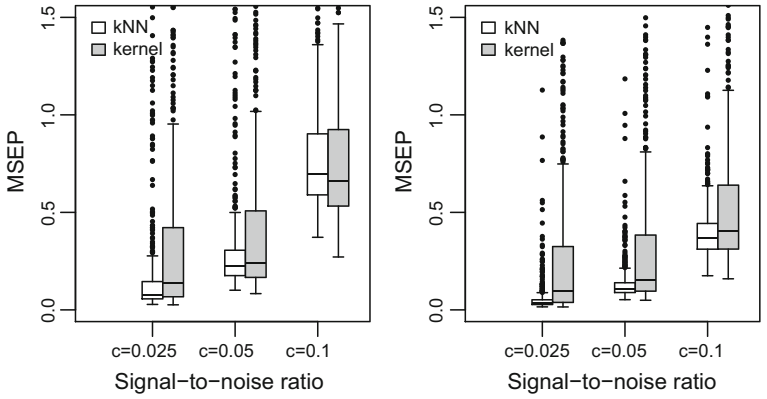


**Fig. 5** (Predictive power) Mean square errors of prediction from the $M = 500$ replicates of the experiment. *Left panel* $\sigma_a^2 = 1$ (the most homogeneous case). *Right panel* $\sigma_a^2 = 0.1$ (the most heterogeneous case)

against the case where the kernel-based estimators proposed in Aneiros-Pérez and Vieu (2006) were considered. For that, we will predict the percentage of fat content ($Y$) in a piece of meat from the corresponding percentages of both protein ($X_1$) and moisture ($X_2$) contents, as well as from the near-infrared absorbance spectra ($\boldsymbol{\zeta}$) in the range 850–1050 nm. In addition to the SFPLR model, some other models will be considered.

Specifically, the dataset consists in $N = 215$ observations, $\{(Y_i, X_{i1}, X_{i2}, \boldsymbol{\zeta}_i)\}_{i=1}^{215}$, from $(Y, X_1, X_2, \boldsymbol{\zeta})$, the spectrometric curves, $\boldsymbol{\zeta}_i$, being observed on 100 equally spaced wavelengths, $t_1 = 850 < t_2 < \cdots < t_{99} < t_{100} = 1050$. Note that this dataset, known as Tecator's dataset, is a benchmark in the setting of functional modeling (for more details on the Tecator's dataset, see, for instance, Ferraty and Vieu 2006; Aneiros-Pérez and Vieu 2006; Shang 2014). They are available at http://lib.stat.cmu.edu/datasets/tecator. Figure 6 shows the spectrometric curves and their first derivatives.

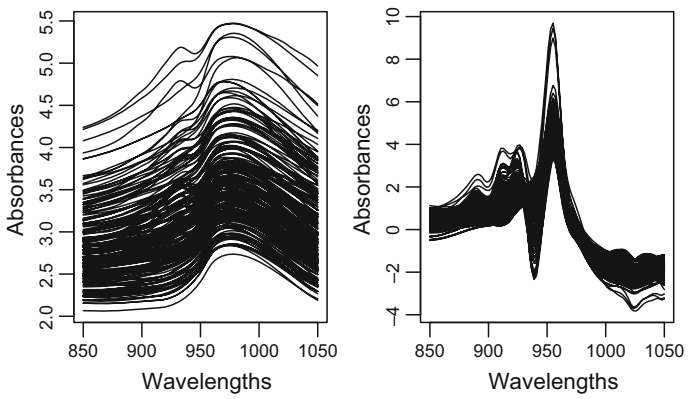


**Fig. 6** Absorbance curves (*left panel*) and their first derivatives (*right panel*)

To compare the predictive power of each procedure ($k$NN- and kernel-based procedures), the Tecator's dataset was split into two samples: a training sample, $\{(Y_i, X_{i1}, X_{i2}, \boldsymbol{\zeta}_i),\ i \in \mathcal{S}_{train}\}$, and a test sample, $\{(Y_j, X_{j1}, X_{j2}, \boldsymbol{\zeta}_j),\ j \in \mathcal{S}_{test}\}$, in such a way that $\#(\mathcal{S}_{train}) = n = 160$, $\mathcal{S}_{train} \cup \mathcal{S}_{test} = \{1, \ldots, 215\}$ and $\mathcal{S}_{train} \cap \mathcal{S}_{test} = \emptyset$. As in the previous Sect. 4: (i) the training sample was used to choose the tuning parameters, (ii) the test sample was allowed to measure the accurate of the predictions, (iii) noting that the spectrometric curves are smooth (see left panel in Fig. 6), a class of semi-metrics based on derivatives, $\{d_q^{[850,1050]}(\cdot,\cdot),\ q = 0, 1, 2, 3\}$ (see 4.2), was considered, (iv) the tuning parameters were selected by minimizing a global cross-validation criterion on the training sample (for the case of the $k$NN-based estimators, see (4.3), (v) the size of the grid used to research $h$ was twofold the size of the grid corresponding to $k$, (vi) the Epanechnikov kernel was considered in both the $k$NN and the kernel procedures, and (vii) the accurate of the predictions was measured by means of the MSEP (see (4.4)).

In first attempt, we considered the partition given by $\mathcal{S}_{train} = \{1, \ldots, 160\}$ and $\mathcal{S}_{test} = \{161, \ldots, 215\}$, this being the more used partition in the literature on functional data. Table 1 reports the values of both the MSEP and the selected semi-metric ($\widehat{q}$) obtained from the SFPLR model (1.1) when estimated from $k$NN- and kernel-based estimators. In addition, to show the adequacy of the SFPLR model to deal with the spectrometric dataset, both the functional nonparametric regression (FNR) and the semi-functional additive regression (SFAR) models were implemented. As in the case of the SFPLR model, their tuning parameters were selected by means of the cross-validation procedure on the training sample. The corresponding values of both the MSEPs and $\widehat{q}$s are also reported in Table 1.

Several conclusions can be drawn from Table 1. Before that, keep in mind that with a functional linear model (that is, with a model which does not involve any nonparametric component) the MSEP is 7.17: on the one hand this is clearly indicating that the effect of the spectrometric curve on the content of fat is nonlinear, while on the other hand this value 7.17 may serve as benchmark for analyzing the results in Table 1. Table 1 indicates that for each model better predictions are obtained from $k$NN-based

**Table 1** Some functional models and values of both the MSEP and the selected semi-metric ($\widehat{q}$) when estimated from $k$NN and kernel estimators

| | Model | $k$NN | Kernel | $\widehat{q}$ |
|---|---|---|---|---|
| FNR | $Y = m(\boldsymbol{\zeta}) + \varepsilon$ | 1.79 | 4.04 | 2 |
| SFAR | $Y = m_1(X_1) + m_2(X_2) + m_3(\boldsymbol{\zeta}) + \varepsilon$ | 1.11 | 2.47 | 2 |
| SFPLR | $Y = X_1\beta_1 + X_2\beta_2 + m(\boldsymbol{\zeta}) + \varepsilon$ | 0.57 | 0.82 | 1 |

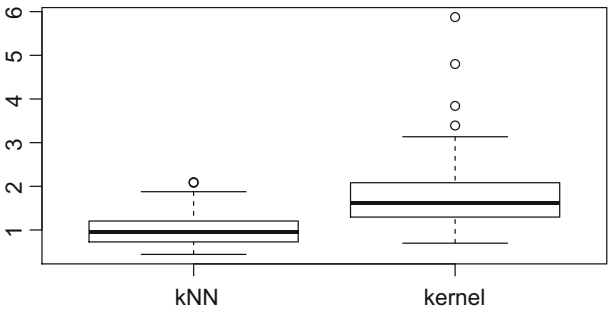


**Fig. 7** Boxplots of the MSEP's obtained when one predicts from the SFPLR model, using the $k$NN and the kernel estimates, the fat contents in each of 100 test samples from the corresponding training samples

estimators than from kernel-based ones. It is also apparent that the SFPLR model shows the better predictive power since the MSEP (with kNN procedure) is around 15 times lower than the benchmark 7.17 obtained with functional linear model.

So, from now on, we will focus on the SFPLR model. In a second attempt, to attenuate the effect of the considered partition, $\{\mathcal{S}_{train}, \mathcal{S}_{test}\}$, 100 partitions were generated at random. Then, for each procedure ($k$NN or kernel), 100 values for the MSEP were obtained from the SFPLR model. Figure 7 displays boxplots from such values. Figure 7 clearly shows a better predictive power for the $k$NN methodology (again).

Finally, it is worth being noted that the computational time needed for implementing the kernel estimate (equivalently, for selecting the bandwidth $h$) was twofold the time corresponding to the $k$NN estimate. Obviously, this is a consequence of the fact that the size of the grid used to research $h$ was twofold the size of the grid corresponding to $k$. Actually, this larger size (and, therefore, the higher computational cost) is the price to pay by the kernel estimator in order to attain accuracy in the bandwidth selection. But, even with such a doble grid, predictive behaviour of the kernel-based estimate is much worse than that of the $k$NN-based one (see again Fig. 7).

## 6 Concluding remarks

Our paper has stated some asymptotic theory for a $k$NN procedure in the functional partial linear regression framework. The finite sample size study highlighted the double interest of the $k$NN approach: on the one hand its location adaptive feature allows for

strong reducing of prediction errors, while on the other hand its simple structure involving a single discrete parameter reduces strongly the computational costs. As far as we know, this is the first paper in location adaptive semiparametric framework for functional data. Extensions to other kinds of functional semiparametric models are challenging open problems for which the technique used in this paper could be helpful.

## 7 Proofs

As discussed in Remark 1, the main challenge when studying theoretical properties of $k$NN estimates is to have to deal with random quantities that cannot be expressed by means of sums of i.i.d. variables. That means that one needs to develop more sophisticated probabilistic tools than the standard limit theorems for i.i.d. sequences which are commonly used for studying other estimates (kernel for instance). In the usual multivariate literature one can distinguish four general routes for that purpose: one based on order statistics properties (see for instance Györfi et al. 2002), a second one based on martingale differences properties (see Chen et al. 2016), a third one based on uniform in bandwidth (UIB) results (see Ouadah 2013) and a fourth one aiming in stating general analytic results about almost sure consistency of random sums being not necessarily i.i.d. (as initiated by Collomb 1979). In nonparametric infinite dimensional setting the last two multivariate routes have been adapted (see for instance Kara-Zaitri et al. 2017a, b, for functional UIB ideas, and Burba et al. 2009 and Kudraszow and Vieu 2013 for functional versions of Collomb's ideas), while a third way (Laloe, 2008) consists in truncating the functional data and to apply then standard multivariate ideas.

In the functional semiparametric framework, the easiest to do is to follow Collomb's route and the proofs of this paper are therefore constructed in this way. First, we state some preliminary lemmas to be used in the proof of Theorem 1. Then, we will prove the theorem.

### 7.1 Some technical lemmas

**Lemma 1** (Theorem 2 in Kudraszow and Vieu 2013) *If assumptions* $(A1)$–$(A6)(i)$ *are satisfied then we have that*

$$\sup_{\zeta\in S_{\mathcal{F}}}\left|g_1(\zeta)-\sum_{i=1}^{n}\omega_{n,k}(\zeta,\boldsymbol{\zeta}_i)Y_i\right|\to 0\ a.s.$$

*and*

$$\sup_{\zeta\in S_{\mathcal{F}}}\left|g_{2j}(\zeta)-\sum_{i=1}^{n}\omega_{n,k}(\zeta,\boldsymbol{\zeta}_i)X_{ij}\right|\to 0\ a.s.$$

**Lemma 2** *Under Assumption (A3) we have that*

$$\max_{1\le i,j\le n} |\omega_{n,k}(\boldsymbol{\zeta}_i, \boldsymbol{\zeta}_j)| = O\left(\frac{1}{k}\right).$$

*Proof* Let us denote $J_i = \{j \in \{1, \ldots, n\}; d(\boldsymbol{\zeta}_i, \boldsymbol{\zeta}_j) < H_{n,k}(\boldsymbol{\zeta}_i)\}$. We can write

$$\sum_{j=1}^{n} K\left(\frac{d(\boldsymbol{\zeta}_i, \boldsymbol{\zeta}_j)}{H_{n,k}(\boldsymbol{\zeta}_i)}\right) = \sum_{j\in J_i} K\left(\frac{d(\boldsymbol{\zeta}_i, \boldsymbol{\zeta}_j)}{H_{n,k}(\boldsymbol{\zeta}_i)}\right) + \sum_{j\in \overline{J_i}} K\left(\frac{d(\boldsymbol{\zeta}_i, \boldsymbol{\zeta}_j)}{H_{n,k}(\boldsymbol{\zeta}_i)}\right).$$

From the conditions on the kernel $K(\cdot)$, and taking into account that $\#(J_i) = k$, it follows that

$$\sum_{j\in J_i} K\left(\frac{d(\boldsymbol{\zeta}_i, \boldsymbol{\zeta}_j)}{H_{n,k}(\boldsymbol{\zeta}_i)}\right) \ge C_1 k \text{ and } \sum_{j\in \overline{J_i}} K\left(\frac{d(\boldsymbol{\zeta}_i, \boldsymbol{\zeta}_j)}{H_{n,k}(\boldsymbol{\zeta}_i)}\right) = 0.$$

Finally, the definition of $\omega_{n,k}(\boldsymbol{\zeta}_i, \boldsymbol{\zeta}_j)$ together with the bounded of $K(\cdot)$ give the result of the lemma. □

**Lemma 3** *Under the conditions of Lemma 1 we have that*

$$\frac{1}{n}\widetilde{\mathbf{X}}^T\widetilde{\mathbf{X}} \to \boldsymbol{\Sigma} \ a.s.$$

*Proof* Firstly, notice that $\widetilde{X}_{ij} = X_{ij} - \sum_{s=1}^{n} \omega_{n,k}(\boldsymbol{\zeta}_i, \boldsymbol{\zeta}_s)X_{sj} = \eta_{ij} + g_{2j}(\boldsymbol{\zeta}_i) - \sum_{s=1}^{n} \omega_{n,k}(\boldsymbol{\zeta}_i, \boldsymbol{\zeta}_s)X_{sj}$ $(i = 1, \ldots, n,\ j = 1, \ldots, p)$; remember that $g_{2j}(\boldsymbol{\zeta}_i) = E\left(X_{ij} \mid \boldsymbol{\zeta}_i\right)$. Then the $(u, v)$th element of $\widetilde{\mathbf{X}}^T\widetilde{\mathbf{X}}$ can be written as

$$\begin{aligned}
\sum_{t=1}^{n} \widetilde{X}_{tu}\widetilde{X}_{tv} &= \sum_{t=1}^{n} \eta_{tu}\eta_{tv} + \sum_{t=1}^{n} \eta_{tu}\left(g_{2v}(\boldsymbol{\zeta}_t) - \sum_{s=1}^{n} \omega_{n,k}(\boldsymbol{\zeta}_t, \boldsymbol{\zeta}_s)X_{sv}\right) \\
&\quad + \sum_{t=1}^{n} \eta_{tv}\left(g_{2u}(\boldsymbol{\zeta}_t) - \sum_{s=1}^{n} \omega_{n,k}(\boldsymbol{\zeta}_t, \boldsymbol{\zeta}_s)X_{su}\right) \\
&\quad + \sum_{t=1}^{n} \left(g_{2u}(\boldsymbol{\zeta}_t) - \sum_{s=1}^{n} \omega_{n,k}(\boldsymbol{\zeta}_t, \boldsymbol{\zeta}_s)X_{su}\right) \\
&\quad \times \left(g_{2v}(\boldsymbol{\zeta}_t) - \sum_{s=1}^{n} \omega_{n,k}(\boldsymbol{\zeta}_t, \boldsymbol{\zeta}_s)X_{sv}\right) \\
&=: I_{uv1} + I_{uv2} + I_{uv3} + I_{uv4}. \qquad (7.1)
\end{aligned}$$

By the strong law of large numbers for i.i.d. variables, we have that, as $n \to \infty$,

$$\frac{1}{n} I_{uv1} = \frac{1}{n}\sum_{t=1}^{n} \eta_{tu}\eta_{tv} \to \sigma_{uv} \text{ a.s.}, \qquad (7.2)$$

where $\sigma_{uv} = Cov(\eta_{1u}, \eta_{1v})$ is the $(u, v)$th element of $\mathbf{\Sigma}$ $(1 \leq u, v \leq p)$. Using the strong law of large numbers again, it follows that

$$\frac{1}{n}\sum_{t=1}^{n}|\eta_{tv}| \to E|\eta_{1v}| < \infty \text{ a.s.} \tag{7.3}$$

Thus, from the second assertion in Lemma 1 and (7.3) we obtain

$$\frac{1}{n}|I_{uv2}| \leq \sup_{\zeta\in S_{\mathcal{F}}}\left|g_{2v}(\zeta) - \sum_{s=1}^{n}\omega_{n,k}(\zeta, \boldsymbol{\zeta}_s)X_{sv}\right|\sum_{t=1}^{n}|\eta_{tv}|/n \to 0 \text{ a.s.} \tag{7.4}$$

Similarly to the proof of (7.4), it follows that

$$\frac{1}{n}|I_{uvr}| \to 0 \text{ a.s. for } r = 3, 4. \tag{7.5}$$

The proof concludes from (7.1), (7.2), (7.4) and (7.5). □

**Lemma 4** (Lemma 3 in Aneiros-Pérez and Vieu 2006) *Let $V_1, \ldots, V_n$ be independent r.v. with zero means and such that for some $r \geq 2$, $\max_{1\leq j\leq n} E\left|V_j\right|^r \leq C < \infty$. Assume that $\left\{a_{ij},\ i, j = 1, \ldots, n\right\}$ is a sequence of positive numbers such that $\max_{1\leq i,j\leq n}\left|a_{ij}\right| = O\left(a_n\right)$ and $\max_{1\leq i\leq n}\sum_{j=1}^{n}\left|a_{ij}\right| = O\left(b_n\right)$. If, in addition,*

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

*and*

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

*then*

$$\max_{1\leq i\leq n}\left|\sum_{j=1}^{n}a_{ij}V_j\right| = O\left(a_n^{1/2}b_n^{1/2}\log n\right) \text{ a.s.}$$

*As a matter of fact, the conclusion of Lemma 4 remains unchanged when $\left\{a_{ij},\ i, j = 1, \ldots, n\right\}$ is a random sequence satisfying the conditions above almost surely.*

### 7.2 Proof of Theorem 1

*Proof of (i)* Similar to Aneiros-Pérez and Vieu (2006), we can write

$$\sqrt{n}(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}) = \left(\frac{\widetilde{\mathbf{X}}^T\widetilde{\mathbf{X}}}{n}\right)^{-1} n^{-\frac{1}{2}}\left\{\sum_{i=1}^{n}\widetilde{\mathbf{X}}_i\overline{m}_{n,k}(\boldsymbol{\zeta}_i)\right.$$

$$
\begin{aligned}
&\quad - \sum_{i=1}^{n} \widetilde{\mathbf{X}}_i \left( \sum_{j=1}^{n} \omega_{n,k}(\boldsymbol{\zeta}_i, \boldsymbol{\zeta}_j)\varepsilon_j \right) + \sum_{i=1}^{n} \widetilde{\mathbf{X}}_i \varepsilon_i \Bigg\} \\
&= \left( \frac{\widetilde{\mathbf{X}}^T \widetilde{\mathbf{X}}}{n} \right)^{-1} n^{-\frac{1}{2}} (S_{n1} - S_{n2} + S_{n3}), \qquad (7.6)
\end{aligned}
$$

where $\overline{m}_{n,k}(\boldsymbol{\zeta}_i) = m(\boldsymbol{\zeta}_i) - \sum_{j=1}^{n} \omega_{n,k}(\boldsymbol{\zeta}_i, \boldsymbol{\zeta}_j) m(\boldsymbol{\zeta}_j)$. Next, noticing the decomposition

$$
\begin{aligned}
\widetilde{X}_{ij} &= \eta_{ij} - \sum_{s=1}^{n} \omega_{n,k}(\boldsymbol{\zeta}_i, \boldsymbol{\zeta}_s)\eta_{sj} + g_{2j}\left( \boldsymbol{\zeta}_i - \sum_{s=1}^{n} \omega_{nk}(\boldsymbol{\zeta}_i, \boldsymbol{\zeta}_s) g_{2j}(\boldsymbol{\zeta}_s) \right) \\
&=: \eta_{ij} - \sum_{s=1}^{n} \omega_{n,k}(\boldsymbol{\zeta}_i, \boldsymbol{\zeta}_s)\eta_{sj} + \widetilde{m}_{ij} \quad (i = 1, \ldots, n,\ j = 1, \ldots, p),
\end{aligned}
$$

we will investigate the asymptotic properties of each $j$th component of $S_{nk}$ ($j = 1, \ldots, p$, $k = 1, 2, 3$) by decomposing it into three summands: $S_{nk,j1}$, $S_{nk,j2}$ and $S_{nk,j3}$.

First, by Theorem 2 in Kudraszow and Vieu (2013), taking $a_{ij} = \omega_{n,k}(\boldsymbol{\zeta}_i, \boldsymbol{\zeta}_j)$, $a_n = \frac{1}{k}$ and $b_n = 1$ with $V_j = \varepsilon_j$ ($V_j = \eta_{ij}$) in Lemma 4, and considering the result in Lemma 2, it follows that

$$
\max_{1 \le i \le n} |\overline{m}_{n,k}(\boldsymbol{\zeta}_i)| = O\left( \phi^{-1}\left( \frac{k}{n} \right)^{\alpha} + \sqrt{\frac{\psi_{S_{\mathcal{F}}}\left( \frac{\log n}{k} \right)}{k}} \right) + O\left( \frac{\log n}{\sqrt{k}} \right) \text{ a.s.} \qquad (7.7)
$$

and

$$
\max_{1 \le i \le n} |\widetilde{m}_{ij}| = O\left( \phi^{-1}\left( \frac{k}{n} \right)^{\alpha} + \sqrt{\frac{\psi_{S_{\mathcal{F}}}\left( \frac{\log n}{k} \right)}{k}} \right) + O\left( \frac{\log n}{\sqrt{k}} \right) \text{ a.s. } (j = 1, \ldots, p). \qquad (7.8)
$$

Now, from (7.7), (7.8) and the Abel's inequality, we obtain that

$$
\begin{aligned}
S_{n1,j1} &= \sum_{i=1}^{n} \widetilde{m}_{ij} \overline{m}_{n,k}(\boldsymbol{\zeta}_i) \le n \max_{1 \le i \le n} |\overline{m}_{n,k}(\boldsymbol{\zeta}_i)| \max_{1 \le i \le n} |\widetilde{m}_{i,j}| \\
&= O\left\{ n \left( \frac{\log^2 n}{k} + \frac{\psi_{S_{\mathcal{F}}}\left( \frac{\log n}{k} \right)}{k} + \left[ \phi^{-1}\left( \frac{k}{n} \right)^{\alpha} \right]^2 \right) \right\} = o(\sqrt{n}) \text{ a.s.} \qquad (7.9)
\end{aligned}
$$

Next, by Lemma 4 again, taking $a_{ij} = \overline{m}_{n,k}(\boldsymbol{\zeta}_i)$, $a_n = \phi^{-1}(\frac{k}{n})^{\alpha} + \sqrt{\frac{\psi_{S_{\mathcal{F}}}(\frac{\log n}{k})}{k}} + \frac{\log n}{\sqrt{k}}$, $b_n = n a_n$ and $V_j = \eta_{ij}$, it follows that

$$S_{n1,j2} = \sum_{i=1}^{n} \eta_{ij} \overline{m}_{n,k}(\boldsymbol{\zeta}_i) = O\left[\left(\phi^{-1}\left(\frac{k}{n}\right)^{\alpha} + \sqrt{\frac{\psi_{S_{\mathcal{F}}}\left(\frac{\log n}{k}\right)}{k}} + \frac{\log n}{\sqrt{k}}\right)\sqrt{n}\log n\right]$$
$$= o(\sqrt{n}) \text{ a.s.} \tag{7.10}$$

Furthermore, by Abel's inequality and Lemma 4 again, we have that

$$\begin{aligned} S_{n1,j3} &= \sum_{i=1}^{n}\left(\sum_{s=1}^{n} \omega_{n,k}(\boldsymbol{\zeta}_i, \boldsymbol{\zeta}_s)\eta_{sj}\right)\overline{m}_{n,k}(\boldsymbol{\zeta}_i) \\ &\le n \max_i |\overline{m}_{n,k}(\boldsymbol{\zeta}_i)| \max_i \left|\sum_{s=1}^{n} \omega_{n,k}(\boldsymbol{\zeta}_i, \boldsymbol{\zeta}_s)\eta_{sj}\right| \\ &= O\left[\left(\phi^{-1}\left(\frac{k}{n}\right)^{\alpha} + \sqrt{\frac{\psi_{S_{\mathcal{F}}}\left(\frac{\log n}{k}\right)}{k}} + \frac{\log n}{\sqrt{k}}\right)\frac{n\log n}{\sqrt{k}}\right] = o(\sqrt{n}) \text{ a.s.} \end{aligned} \tag{7.11}$$

Secondly, observing that from Lemma 2 it holds that $\max_i |\sum_{j=1}^{n} \omega_{n,k}(\boldsymbol{\zeta}_i, \boldsymbol{\zeta}_j)\varepsilon_j| = O(\frac{\log n}{\sqrt{k}})$ a.s., we have that similar to the arguments used to obtain (7.9)–(7.11), considering (7.8) and the Abel's inequality, it follows that

$$\begin{aligned} S_{n2,j1} &= \sum_{i=1}^{n} \widetilde{m}_{ij}\left(\sum_{j=1}^{n} \omega_{n,k}(\boldsymbol{\zeta}_i, \boldsymbol{\zeta}_j)\varepsilon_j\right) \\ &= O\left[\left(\phi^{-1}\left(\frac{k}{n}\right)^{\alpha} + \sqrt{\frac{\psi_{S_{\mathcal{F}}}\left(\frac{\log n}{k}\right)}{k}} + \frac{\log n}{\sqrt{k}}\right)\frac{n\log n}{\sqrt{k}}\right] = o(\sqrt{n}) \text{ a.s.}, \end{aligned} \tag{7.12}$$

$$S_{n2,j2} = \sum_{i=1}^{n}\left(\sum_{j=1}^{n} \omega_{n,k}(\boldsymbol{\zeta}_i, \boldsymbol{\zeta}_j)\varepsilon_j\right)\eta_{ij} = O\left(\frac{\sqrt{n}\log^2 n}{\sqrt{k}}\right) = o(\sqrt{n}) \text{ a.s.} \tag{7.13}$$

and

$$\begin{aligned} S_{n2,j3} &= \sum_{i=1}^{n}\left(\sum_{s=1}^{n} \omega_{n,k}(\boldsymbol{\zeta}_i, \boldsymbol{\zeta}_s)\eta_{sj}\right)\left(\sum_{l=1}^{n} \omega_{n,k}(\boldsymbol{\zeta}_i, \boldsymbol{\zeta}_l)\varepsilon_l\right) \\ &\le n \max_i \left|\sum_{s=1}^{n} \omega_{n,k}(\boldsymbol{\zeta}_i, \boldsymbol{\zeta}_s)\eta_{sj}\right| \max_i \left|\sum_{l=1}^{n} \omega_{n,k}(\boldsymbol{\zeta}_i, \boldsymbol{\zeta}_l)\varepsilon_l\right| \\ &= O\left(\frac{n\log^2 n}{k}\right) = o(\sqrt{n}) \text{ a.s.} \end{aligned} \tag{7.14}$$

Finally, for $S_{n3}$, we have that

$$S_{n3,j1} = \sum_{i=1}^{n} \widetilde{m}_{ij}\varepsilon_i = O\left[\left(\phi^{-1}\left(\frac{k}{n}\right)^{\alpha} + \sqrt{\frac{\psi_{S_{\mathcal{F}}}\left(\frac{\log n}{k}\right)}{k}} + \frac{\log n}{\sqrt{k}}\right)\frac{\sqrt{n}\log n}{\sqrt{k}}\right] = o(\sqrt{n}) \text{ a.s.} \tag{7.15}$$

and

$$S_{n3,j2} = \sum_{i=1}^{n}\left(\sum_{s=1}^{n}\omega_{n,k}(\boldsymbol{\zeta}_i, \boldsymbol{\zeta}_s)\eta_{sj}\right)\varepsilon_i = O\left(\frac{\sqrt{n}\log n}{\sqrt{k}}\right) = o(\sqrt{n}) \text{ a.s.} \tag{7.16}$$

Then, from (7.6) and (7.9)–(7.16), it follows that

$$\begin{aligned}\sqrt{n}(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}) &= \left(\frac{\widetilde{\mathbf{X}}^T\widetilde{\mathbf{X}}}{n}\right)^{-1} n^{-\frac{1}{2}}(S_{n1} - S_{n2} + S_{n3}) \\ &= \left(\frac{\widetilde{\mathbf{X}}^T\widetilde{\mathbf{X}}}{n}\right)^{-1} n^{-\frac{1}{2}}\left(\sum_{i=1}^{n}\boldsymbol{\eta}_i\varepsilon_i + o(\sqrt{n})\right).\end{aligned} \tag{7.17}$$

Furthermore, a central limit theorem for i.i.d. random variables shows that

$$n^{-\frac{1}{2}}\sum_{i=1}^{n}\boldsymbol{\eta}_i\varepsilon_i \to N(0, \sigma^2\boldsymbol{\Sigma}). \tag{7.18}$$

The proof concludes from Lemma 3, (7.17) and (7.18). □

*Proof of (ii)* Taking into account both Lemma 3 and the decomposition (7.17), this proof can be obtained by following the same steps as those in Aneiros-Perez and Vieu (2006); so, we omit it here. □

*Proof of (iii)* Firstly, we obtain the following decomposition:

$$\begin{aligned}\widehat{m}_n(\zeta) &= \sum_{i=1}^{n}\omega_{n,k}(\zeta, \boldsymbol{\zeta}_i)(m(\boldsymbol{\zeta}_i) + \varepsilon_i) - \sum_{i=1}^{n}\omega_{n,k}(\zeta, \boldsymbol{\zeta}_i)\mathbf{X}_i^T(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}) \\ &=: \widehat{m}_n^*(\zeta) - \sum_{i=1}^{n}\omega_{n,k}(\zeta, \boldsymbol{\zeta}_i)\mathbf{X}_i^T(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}).\end{aligned}$$

Therefore, we can write

$$\sup_{\zeta\in S_{\mathcal{F}}}|\widehat{m}_n(\zeta) - m(\zeta)| \le \sup_{\zeta\in S_{\mathcal{F}}}|\widehat{m}_n^*(\zeta) - m(\zeta)| + \sup_{\zeta\in S_{\mathcal{F}}}\left\|\sum_{i=1}^{n}\omega_{n,k}(\zeta, \boldsymbol{\zeta}_i)\mathbf{X}_i^T\right\|\left\|\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}\right\|,$$

where $\|\cdot\|$ denotes the Euclidean norm. Now, Theorem 2 in Kudraszow and Vieu (2013) together with the first assertion in Theorem 1 above are enough to conclude this proof. □

**Acknowledgements** This work has received financial support from the National Social Science Funds of China (NSSF 14ATJ005), the NNSF of China (NNSF 11501005), the Spanish Ministerio de Economía y Competitividad (Grant MTM2014-52876-R), the Xunta de Galicia (Centro Singular de Investigación de Galicia accreditation ED431G/01 2016-2019 and Grupos de Referencia Competitiva ED431C2016-015) and the European Union (European Regional Development Fund—ERDF). The authors would like to thank the Associate Editor and the two anonymous referees for their constructive and helpful comments, which have greatly improved the paper.

## References

Aneiros G, Bongiorno EG, Cao R, Vieu P (2017) Functional statistics and related fields. Contributions to statistics. Springer, Cham

Aneiros G, Ling N, Vieu P (2015) Error variance estimation in semi-functional partially linear regression models. J Nonparametr Stat 27(3):316–330

Aneiros-Pérez G, Vieu P (2006) Semi-functional partial linear regression. Stat Probab Lett 76(11):1102–1110

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

Attouch M, Benchikh T (2012) Asymptotic distribution of robust k-nearest neighbour estimator for functional nonparametric models. Mat Vesnik 644:275–285

Berlinet A, Servien R (2011) Necessary and sufficient condition for the existence of a limit distribution of the nearest-neighbour density estimator. J Nonparametr Stat 23(3):633–643

Biau G, Cérou F, Guyader A (2010) Rates of convergence of the functional k-nearest neighbor estimate. IEEE Trans Inf Theory 56(4):2034–2040

Bosq D, Blanke D (2007) Inference and prediction in large dimension. Wiley series in probability and statistics. Wiley, Chichester

Burba F, Ferraty F, Vieu P (2009) $k$-Nearest Neighbour method in functional nonparametric regression. J Nonparametr Stat 21(4):453–469

Cerou F, Guyader A (2006) Nearest neighbor classification in infinite dimension. ESAIM Probab Stat 10:340–355

Chen Z, Wang H, Wang X (2016) The consistency for the estimator of nonparametric regression model based on martingale difference errors. Stat Pap 57(2):451–469

Chu B, Huynh K, Jacho-Chávez D (2013) Functionals of order statistics and their multivariate concomitants with application to semiparametric estimation by nearest neighbours. Sankhya B 75(2):238–292

Collomb G (1979) Estimation de la régression par la méthode des k points les plus proches: propriétés de convergence ponctuelle (French). C R Acad Sci Paris 289(3):245–247

Collomb G, Hassani S, Sarda P, Vieu P (1985) Convergence uniforme d'estimateurs de la fonction de hazard pour des observations dépendantes: méthodes du noyau et des k points les plus proches. (French). C R Acad Sci Paris Sér I Math 301(12):653–656

Cover T-M (1968) Estimation by the nearest neighbor rule. IEEE Trans Inf Theory IT–14:50–55

Cuevas A (2014) A partial overview of the theory of statistics with functional data. J Stat Plann Inference 147:1–23

Dette H, Gefeller O (1995) The impact of different definitions of nearest neighbour distances for censored data on nearest neighbour kernel estimators of the hazard rate. J Nonparametr Stat 4(3):271–282

Devroye L, Györfi L, Krzyzak A, Lugosi G (1994) On the strong universal consistency of nearest neighbor regression function estimates. Ann Stat 22:1371–1385

Devroye L, Wagner T (1977) The strong uniform consistency of nearest neighbor density estimates. Ann Stat 5(3):536–540

Devroye L, Wagner T (1982) Nearest neighbor methods in discrimination. In: Classification, pattern recognition and reduction of dimensionality. Handbook of statistics, vol 2. North-Holland, Amsterdam. pp 193–197

Engle RF, Granger CWJ, Rice J, Weiss A (1986) Semiparametric estimates of the relation between weather and electricity sales. J Am Stat Assoc 81:310–320
Ferraty F, Vieu P (2006) Nonparametric functional data analysis. Theory and practice. Springer, New York
Goia A, Vieu P (2014) Some advances in semiparametric functional data modelling. In: Contributions in infinite-dimensional statistics and related topics. Esculapio, Bologna, pp 135–141
Goia A, Vieu P (2016) An introduction to recent advances in high/infinite dimensional statistics. J Multivar Anal 146:1–6
Györfi L, Kohler M, Krzyzak A, Walk H (2002) A distribution-free theory of nonparametric regression. Springer series in statisics. Springer, New York
Härdle W, Liang H, Gao J (2000) Partially linear models. Physica-Verlag, Heidelberg
Hong S (1992) Estimation theory of a class of semiparametric regression models. Sci China Ser A 35(6):657–674
Horváth L, Kokoszka P (2012) Inference for functional data with applications. Springer series in statistics. Springer, New York
Hsing T, Eubank R (2015) Theoretical foundations to functional data analysis with an introduction to linear operators. Wiley series in probability and statistics. Wiley, Chichester
Kara-Zaitri L, Laksaci A, Rachdi M, Vieu P (2017a) Data-driven kNN estimation in nonparametric functional data-analysis. J Multivar Anal 153:176–188
Kara-Zaitri L, Laksaci A, Rachdi M, Vieu P (2017b) Uniform in bandwidth consistency for various kernel estimators involving functional data. J Nonparametr Stat 29(1):85–107
Kudraszow N, Vieu P (2013) Uniform consistency of *k*NN regressors for functional variables. Stat Probab Lett 83(8):1863–1870
Laloë T (2008) A k-nearest approach for functional regression. Stat Prob Lett 10:1189–1193
Li H, Li Q, Liu R (2016) Consistent model specification tests based on k-nearest-neighbor estimation methods. J Econom 194(1):187–202
Lian H (2011) Functional partial linear model. J Nonparametr Stat 23(1):115–128
Ouadah S (2013) Uniform-in-bandwidth nearest-neighbor density estimation. Stat Probab Lett 83(8):1835–1843
Paindaveine D, Van Bever G (2015) Nonparametrically consistent depth-based classifiers. Bernoulli 21(1):69–82
Ramsay J, Silverman B (2005) Functional data analysis. Springer series in statistics. Springer, New York
Robinson P (1995) Nearest-neighbour estimation of semiparametric regression models. J Nonparametr Stat 5(1):33–41
Sancetta A (2010) Nearest neighbor conditional estimation for Harris recurrent Markov chains. J Multivar Anal 100(10):2224–2236
Shang H (2014) Bayesian bandwidth estimation for a semi-functional partial linear regression model with unknown error density. Comput Stat 29(3–4):829–848
Zhang J (2013) Analysis of variance for functional data. Monographs on statistics and applied probability. Chapman & Hall/CRC, Boca Raton