Wild bootstrap bandwidth selection of recursive nonparametric relative regression for independent functional data

Yousri Slaoui

Version of Record: https://www.sciencedirect.com/science/article/pii/S0047259X19300363
Manuscript_4cf604e6ff82c021eeb9c772208e9354

Wild Bootstrap Bandwidth Selection of Recursive Nonparametric Relative Regression for Independent Functional Data

Yousri Slaoui*

Université de Poitiers, France

Abstract

We propose and investigate a new kernel regression estimator based on the minimization of the mean squared relative error. We study the properties of the proposed recursive estimator and compare it with the recursive estimator based on the minimization of the mean squared error proposed by Slaoui (2018). It turns out that, with an adequate choice of the parameters, the proposed estimator performs better than the recursive estimator based on the minimization of the mean squared error. We illustrate these theoretical results through a real chemometric dataset.

Key words: Asymptotic normality, Bootstrap, Functional data analysis, Functional nonparametric statistics, Mean square relative error, Nonparametric estimation, Stochastic approximation algorithm.
2010 MSC: Primary: 62G08, 62L20, 62G09. Secondary: 65D10.

1. Introduction

Functional data have become more and more popular in modern statistics because of the progress in computing technology, in terms of both memory capacity and computing speed which have made it possible to record vast amounts of data. It concerns many statistical methods dealing with random variables valued in some infinite-dimensional space, called functional variables. Thus, a very large number of variables can be observed for the study of the same phenomenon. This type of data appears in several fields including climatology, economics, psychology, linguistics, medicine, and so on.

There has been an increasing interest in Functional Data Analysis (FDA) in the past decades, as it is highlighted by the popular monograph of Ramsay and Silverman [43], who provide a detailed exposition of both theoretical and practical aspects of functional data analysis. Statistical inference for FDA has been widely investigated (see, e.g., [28, 29]). The existing literature contains numerous studies on functional linear models (see, e.g., [7, 8, 26]). The nonparametric treatment has been popularized by the book of Ferraty and Vieu [23] and now takes a large place in the FDA literature; see the discussions in the recent surveys by Cuevas [15] who offers a short tutorial as well as a partial survey of the FDA theory. We point also to the work [36] which gives a survey of nonparametric FDA and presents a wide scope of open questions. Aneiro et al. [2] present various contributions into two categories: papers promoting new methodology for data varying over a continuum, and papers concerned with very high- but finite-dimensional problems.

In the last decade, data streams have become an increasingly important area of research. Common data streams include Internet packet data, Twitter activity, Facebook news stream, and credit card transactions. In those situations, data arrives regularly so that it is impossible to store them in a traditional database. In such a context, building a recursive estimator which does not require to store all the data in memory and can be updated easily in order to deal with online data is of great interest.

This work concerns nonparametric recursive estimation of the regression operator when the explanatory data are curves and the response is real-valued, based on the minimization of the Mean Squared Relative Error (MSRE). This

*Corresponding author. Email address: yousri.slaoui@math.univ-poitiers.fr

problem can be formulated by considering $\{Y_i, \mathcal{X}_i\}_{i=1}^n$ a sample of independent and identically distributed couples, where Y_i is real-valued and $\mathcal{X}_i$ takes values in some functional space $\mathcal{E}$ equipped with a semi-norm $\|\cdot\|$. A common nonparametric model of this relationship is for $i \in \{1, \ldots, n\}$ given by

$$Y_i = r(\mathcal{X}_i) + \varepsilon_i,$$

where ε_i is a random variable. In ordinary predictions, we estimate the operator r by minimizing the expected squared loss function:

$$\mathrm{E}\left[(Y - r(\mathcal{X}))^2 | \mathcal{X}\right],$$

and we obtain as predictor the quantity $\mathrm{E}[Y|\mathcal{X}]$, which gives the Mean Squared Error (MSE). However, when $Y > 0$ or in the presence of outliers, the use of the classical loss function can lead to unreasonable results, since all variables have the same weight. Therefore, it is of interest to consider the response level $(Y - r(\mathcal{X}))/Y$ rather than $(Y - r(\mathcal{X}))$ and then it is necessary to consider the MSRE:

$$\mathrm{E}\left[\left(\frac{Y - r(\mathcal{X})}{Y}\right)^2 | \mathcal{X}\right].$$

Minimizing this loss function leads to predicting the quantity (see Park and Stefanski [40])

$$\frac{\mathrm{E}\left[Y^{-1}|\mathcal{X}\right]}{\mathrm{E}\left[Y^{-2}|\mathcal{X}\right]}.$$

By assuming that $\mathrm{E}\left|Y_i^{-l}|\mathcal{X}\right| < \infty$, for $l \in \{1, 2\}$, we define the regression functional as

$$r(u) := \frac{\mathrm{E}\left[Y_i^{-1}|\mathcal{X}_i = u\right]}{\mathrm{E}\left[Y_i^{-2}|\mathcal{X}_i = u\right]}; \quad u \in \mathcal{E}, \quad \forall i \in \mathbb{N}.$$

Moreover, we set $g_l(u) = \mathrm{E}\left[Y_i^{-l}|\mathcal{X}_i = u\right]$, for all $l \in \mathbb{N}$, $u \in \mathcal{E}$, and $i \in \mathbb{N}$. Then, we have the following relation, $g_1(u) = r(u)\, g_2(u)$.

The purpose of this study is to extend the work proposed first in the case of a real explanatory variable in Slaoui [46] and then in the case of a functional explanatory variable in Slaoui [47]. The two previous estimators are based on the minimization of the MSE. In the current work we propose to use the MSRE rather than the MSE criterion. Thus, our proposed estimator is:

$$\hat{r}_n(\chi, h) = \frac{\hat{m}_{n,1}(\chi, h)}{\hat{m}_{n,2}(\chi, h)},$$

with

$$\hat{m}_{n,j}(\chi, h) = (1 - \gamma_n)\, \hat{m}_{n-1,j}(\chi, h) + \gamma_n h_n^{-1} K\left(\frac{\|\chi - \mathcal{X}_n\|}{h_n}\right) Y_n^{-j}; \quad j \in \{1, 2\}, \tag{1}$$

where (γ_n) and (h_n) are sequences of positive real numbers which are converging to 0, K is a kernel and $h \in \{h_1, \ldots, h_n\}$. The recursive property (1) is particularly useful for large samples, because $\hat{r}_n$ can be updated easily using additional observation. Throughout this paper, we suppose that $\hat{m}_{0,j}(\chi, h) = 0$, for $j \in \{1, 2\}$; and we let $\prod_n = \prod_{i=1}^n (1 - \gamma_i)$. Then, we can estimate the operator r by

$$\hat{r}_n(\chi, h) = \frac{\prod_n \sum_{k=1}^n \prod_k^{-1} \gamma_k h_k^{-1} Y_k^{-1} K\left(\frac{\|\chi - \mathcal{X}_k\|}{h_k}\right)}{\prod_n \sum_{k=1}^n \prod_k^{-1} \gamma_k h_k^{-1} Y_k^{-2} K\left(\frac{\|\chi - \mathcal{X}_k\|}{h_k}\right)}. \tag{2}$$

Despite that the MSRE is widely used in practice as a measure of performance, the theoretical properties of this alternative regression were not frequently used until the work of Narula and Wellington [39]. Since this work, we

can find some criteria based on minimizing the sum of Absolute Relative Errors (ARE) and others on the sum of squared relative errors (SRE). For example, see [33] for some models in software engineering, [9] for some examples in medicine or [10] for some financial applications. Note that most of the recent methods are concerned with the estimation of unknown parameters (see, [49] for recent advances and references).

In the framework of nonparametric estimations, Jones et al. [31] studied the asymptotic properties of an estimator minimizing the sum of the squared relative errors by considering the kernel estimation methods and a local linear approach. Moreover, in the functional framework, only the paper by Demongeot et al. [17] has paid attention to the study of nonparametric prediction via relative error regression. As far as we know, there is no work on recursive estimators in the functional nonparametric literature based on the minimization of the MSRE. The works of Amiri et al. [1] and Slaoui [47], both proposed recursive estimation of the operator r based on the minimization of the MSE.

Furthermore, the functional bootstrap literature is not widely developed. Poltis and Romano [41] developed some weak convergence results for approximating sums of weakly dependent stationary Hilbert space valued random variables under the asymptotic validity of a stationary bootstrap method. Cuevas et al. [14] presented a Monte Carlo study analyzing the performance of the bootstrap confidence bands of several functional estimators. Raña et al. [42] considered a naive and a wild bootstrap procedure to construct pointwise confidence intervals for a nonparametric regression function when the predictor is of functional nature and when the data are dependent. More recently, Aneiro et al. [3] proposed a naive and a wild bootstrap procedure to approximate the distribution of kernel-based estimators under α-mixing conditions, whereas Shang [45] considered bootstrap methods for estimating the long-run covariance of stationary functional time series.

The first purpose of this paper, is to study, under some general conditions, the asymptotic properties of an alternative functional recursive kernel estimator of the regression operator r. Secondly, we propose an automatic bandwidth selection through a wild bootstrap method. Thirdly, we compare the performance of the proposed recursive estimator $\hat{r}_n$ based on the minimization of the MSRE to the recursive functional regression estimator introduced by Slaoui [47], the latter is based on the minimization of the MSE and defined as

$$\check{r}_n(\chi, h) = \frac{\prod_n \sum_{k=1}^{n} \prod_k^{-1} \gamma_k h_k^{-1} Y_k K\left(\frac{\|\chi - \mathcal{X}_k\|}{h_k}\right)}{\prod_n \sum_{k=1}^{n} \prod_k^{-1} \gamma_k h_k^{-1} K\left(\frac{\|\chi - \mathcal{X}_k\|}{h_k}\right)}. \tag{3}$$

The layout of the present paper is as follows. Section 2 is devoted to the main results of the present work. Section 3 is dedicated to application results by using a real dataset. We conclude the article in Section 4. To avoid interrupting the flow of this paper, all mathematical developments are relegated to Appendix A.

2. Assumptions and main results

Firstly, we define the following class of regularly varying sequences.

Definition 1. *Let $\gamma \in \mathbb{R}$ and $(v_n)_{n\geq 1}$ be a non-random positive sequence. We say that $(v_n) \in \mathcal{GS}(\gamma)$ if*

$$\lim_{n\to+\infty} n\left[1 - \frac{v_{n-1}}{v_n}\right] = \gamma. \tag{4}$$

Condition (4) was introduced by Galambos and Seneta [24] to define regularly varying sequences (see also Bojanic and Seneta [6]) and by Mokkadem and Pelletier [37] in the context of stochastic approximation algorithms. Note that $\mathcal{GS}$ stands for *Galambos and Seneta*. Typical sequences in $\mathcal{GS}(\gamma)$ are, for $b \in \mathbb{R}$, $n^{\gamma}(\log n)^{b}$, $n^{\gamma}(\log\log n)^{b}$, and so on. In this section, we investigate the asymptotic properties of the proposed estimator (2). Before presenting the assumptions, let us give some notations. Let F be the cumulative distribution function of the random variable $\|\mathcal{X} - \chi\|$:

$$F_{\chi}(t) = \Pr(\|\mathcal{X} - \chi\| \leq t)$$

and let $B(\chi, t)$ be the ball in $\mathcal{E}$ with center χ and radius t:

$$B(\chi, t) = \{\chi_1 \in \mathcal{E}; \|\chi_1 - \chi\| \leq t\}.$$

Moreover, we let for any $s \in [0,1]$

$$\tau_{\chi,h}(s) = \frac{F_\chi(hs)}{F_\chi(h)}.$$

For any $\beta \in \{1,2\}$

$$\psi_{\chi,\beta}(u) = \mathrm{E}\left[\left\{g_\beta(\mathcal{X}) - g_\beta(\chi)\right\} | \, \|\mathcal{X} - \chi\| = u\right].$$

For simplicity, we introduce the following notation:

$$\xi = \lim_{n\to+\infty} (n\gamma_n)^{-1}, \tag{5}$$

$$\mathcal{B}(\chi) = \frac{\left(\psi'_{\chi,1}(0) - r(\chi)\,\psi'_{\chi,2}(0)\right)}{g_2(\chi)}, \quad \mathcal{V}(\chi) = \frac{\left(g_2(\chi) - 4r(\chi)\,g_3(\chi) + 3r^2(\chi)\,g_4(\chi)\right)}{g_2^2(\chi)}.$$

We denote by C or/and C' some real generic constants supposed to be strictly positive. The assumptions which we shall refer to are the following:

(A1) For $\beta \in \{1,2,3,4\}$ and for all $u \in \mathcal{E}$ the functions $g_\beta(u)$ are bounded and continuous in a neighborhood of u.

(A2) For $\beta \in \{1,2\}$ and for all $u \in \mathcal{E}$ the function $\psi_{\chi,\beta}(u)$ is assumed to admit a derivative at $t = 0$ and $\psi'_{\chi,\beta}(0)$ is uniformly Lipschitz continuous of order $\alpha \in (0,1]$ in χ.

(A3) For all $\chi_1 \in \mathcal{E}$, $F_{\chi_1}(0) = 0$ and $F_{\chi_1}(t)/F_\chi(t)$ is Lipschitz continuous of order $\alpha \in (0,1]$ in χ_1, uniformly in t in a neighborhood of 0.

(A4) $K : \mathbb{R} \to \mathbb{R}$ is a continuous, bounded function with support on the compact $[0,1]$, such that $\min_{t\in[0,1]} K(t) > 0$, and is a differentiable function on $(0,1)$ where its first derivative function K' is such that: $-\infty < C < K'(t) < C' < 0$.
Now, it is possible to introduce the following notation used first by Ferraty et al. [21]:

$$M_{\chi,0} = K(1) - \int_0^1 (tK(t))'\,\tau_{\chi,0}(t)\,dt, \quad M_{\chi,1} = K(1) - \int_0^1 K'(t)\,\tau_{\chi,0}(t)\,dt,$$

$$M_{\chi,2} = K^2(1) - \int_0^1 \left(K^2(t)\right)'\,\tau_{\chi,0}(t)\,dt.$$

(A5) For any $s \in [0,1]$, there exists a function $\tau_{\chi,0}(.)$ such that, $\tau_{\chi,0}(s) = \lim_{h\to 0}\tau_{\chi,h}(s)$, $\tau_{\chi,0}(s) < \infty$, $\sup_{s\in[0,1]}\left|\tau_{\chi,h}(s) - \tau_{\chi,0}(s)\right| = o(1)$, $M_{\chi,j} > 0$ for $j \in \{0,1,2\}$, and $M_{\chi,j}$ is Lipschitz continuous of order $\alpha \in (0,1]$ for $j \in \{0,1,2\}$.

(A6) (i) $(\gamma_n) \in \mathcal{GS}(-\gamma)$ with $\gamma \in (1/2, 1]$.
(ii) $(h_n) \in \mathcal{GS}(-a)$ with $a \in (0,1)$.
(iii) $\left(F_\chi(h_n)\right) \in \mathcal{GS}(-\mathcal{F}_a)$ with $\mathcal{F}_a \in (0,\gamma)$.
(iv) $\lim_{n\to\infty}(n\gamma_n) \in (\min\{\mathcal{F}_a, (\gamma+\mathcal{F}_a)/2 - a\}, \infty]$.
(v) $(b_n) \in \mathcal{GS}(-b)$ with $b \in ((\gamma - \mathcal{F}_a)/2, a)$.
(vi) $\left(F_\chi(b_n)\right) \in \mathcal{GS}(-\mathcal{F}_b)$ with $\mathcal{F}_b \in (0, \mathcal{F}_a)$.

(A7) For each n, there exist $r_n \geq 1$, $l_n \in \mathcal{GS}(-l)$ with $l > b + (\gamma - \mathcal{F}_a)/2$ and $t_{1,n}, \ldots, t_{r_n,n}$ such that $B(\chi,h) \subset \bigcup_{k=1}^{r_n} B(t_{k,n}, l_n)$, with $r_n = O\left(n^{b_n/h_n}\right)$.

Assumption $(A1)$ ensures that the inverse moments of the response variable $\mathrm{E}\left[Y_i^{-\beta}|\mathcal{X}_i = u\right]$, for all, $\beta \in \{1,2,3,4\}$, and all $u \in \mathcal{E}$ are bounded and continuous in a neighborhood of u. Assumption $(A6)\,(iii)$ indicates that the small ball probability satisfies $\gamma_n^{-1} F_\chi(h_n) \to \infty$ as n goes to ∞ for any $\chi \in \mathcal{E}$. Assumption $(A6)\,(iv)$ is standard in the framework of stochastic approximation algorithms, it implies in particular that the limit of $\left([n\gamma_n]^{-1}\right)$ is finite. Assumption $(A6)\,(v)$ ensures that $b_n^{1+\alpha}\left(\gamma_n^{-1} F_\chi(h_n)\right)^{1/2} \to 0$ as n goes to ∞ for $\alpha \in (0,1]$. Assumption $(A6)\,(vi)$ indicates that $\left[F_\chi(h_n)/F_\chi(b_n)\right]\log n \to 0$ as n goes to ∞. Assumption $(A7)$ ensures that $l_n b_n^{-1}\left(\gamma_n^{-1} F_\chi(h_n)\right)^{1/2} \to 0$ as n goes to ∞. Moreover, we give two examples of functional spaces for which Assumption $(A7)$ is fulfilled.

Example 1 (Functional spaces satisfying the assumption $(A7)$).

1. *Let $\mathcal{E}$ is a separable Hilbert space, with inner product $<,>$ and with orthonormal basis $\left\{e_j : j \in \{1,\ldots,\infty\}\right\}$. For any $(\chi_1,\chi_2) \in \mathcal{E}\times\mathcal{E}$, and for $k>0$, we let d_k be the semi-metric defined by:*

$$d_k(\chi_1,\chi_2) = \sqrt{\sum_{j=1}^{k} <\chi_1-\chi_2, e_j>^2}.$$

   The space $(\mathcal{E}, d_k)$ fulfills Assumption $(A7)$.

2. *Let $\mathcal{E}$ is the space of all continuous functions $\chi : [a,b] \to \mathbb{R}$ with $\|\chi\|_\gamma \le C$, where $-\infty < a < b < \infty$, $0 < \gamma < \infty$, let γ_1 be the largest integer strictly smaller than γ and $\|.\|_2$ be the Euclidean norm:*

$$\|\chi\|_\gamma = \max_{k\le\gamma_1}\sup_t\left|\chi^{(k)}(t)\right| + \sup_{t_1,t_2}\frac{\left|\chi^{(\gamma_1)}(t_1)-\chi^{(\gamma_1)}(t_2)\right|}{\|t_1-t_2\|_2^{\gamma-\gamma_1}}.$$

   The space $(\mathcal{E}, d_{L^p})$ fulfills the Assumption $(A7)$, where d_{L^p} is the L^p distance in $\mathcal{E}$ and $p \in [1,\infty]$.

Remark 1.
The intuition behind the use of such bandwidth h_n belonging to $\mathcal{GS}(-a)$ is that the ratio h_{n-1}/h_n is equal to $1 + a/n + o(1/n)$, the application of Lemma 1 (given in the Appendix) under the assumption $(A6)$, ensures that the bias and the variance depend only on h_n and not on $h_1,\ldots,h_n$.

Our first result is the following proposition, which gives the bias and the variance of $\hat{r}_n$.

Proposition 1 (Bias and variance of $\hat{r}_n$). *Let Assumptions $(A1)-(A6)$ be satisfied.*

1. *If $a \in (0, (\gamma-\mathcal{F}_a)/2]$, then*

$$\mathrm{E}\left[\hat{r}_n(\chi,h)\right] - r(\chi) = \mathcal{B}(\chi)\frac{1-(\mathcal{F}_a-a)\xi}{1-\mathcal{F}_a\xi}\frac{M_{\chi,0}}{M_{\chi,1}}h_n\left[1+o(1)\right]. \tag{6}$$

   If $a \in ((\gamma-\mathcal{F}_a)/2, 1)$, then

$$\mathrm{E}\left[\hat{r}_n(\chi,h)\right] - r(\chi) = o\left(\sqrt{\gamma_n F_\chi(h_n)^{-1}}\right).$$

2. *If $a \in (0, (\gamma-\mathcal{F}_a)/2)$, then*

$$\mathrm{var}\left[\hat{r}_n(\chi,h)\right] = o\left(h_n^2\right). \tag{7}$$

   If $a \in [(\gamma-\mathcal{F}_a)/2, 1)$, then

$$\mathrm{var}\left[\hat{r}_n(\chi,h)\right] = \mathcal{V}(\chi)\frac{(1-(\mathcal{F}_a-a)\xi)^2}{(2-(\mathcal{F}_a+\gamma-2a)\xi)}\frac{M_{\chi,2}}{M_{\chi,1}^2}\frac{\gamma_n}{F_\chi(h_n)}\left[1+o(1)\right]. \tag{8}$$

3. *If* $\lim_{n\to\infty}(n\gamma_n) > \max\{\mathcal{F}_a, (\mathcal{F}_a+\gamma)/2-a\}$*, then (6) and (8) hold simultaneously.*

The bias and the variance of the estimator $\hat{r}_n$ defined by the stochastic approximation algorithm (2) then heavily depends on the choice of the stepsize γ_n. Let us now recall that the bias and variance of the nonrecursive estimator $\widetilde{r}_n(\chi,h)$, proposed in Demongeot et al. [17] are given by

$$\begin{aligned}\mathrm{E}\left[\widetilde{r}_n(\chi,h)\right]-r(\chi) &= \mathcal{B}(\chi)\frac{M_{\chi,0}}{M_{\chi,1}}h_n\left[1+o(1)\right],\\ \mathrm{var}\left[\widetilde{r}_n(\chi,h)\right] &= \mathcal{V}(\chi)\frac{M_{\chi,2}}{M_{\chi,1}^2}\frac{\gamma_n}{F_\chi(h_n)}\left[1+o(1)\right].\end{aligned}$$

To illustrate the benefit of choosing non-standard weights γ_n, we give some possible choices of the sequence γ_n and compare the bias and variance of $\hat{r}_n$ with those of $\widetilde{r}_n$.

Example 2 (Choices of γ_n minimizing the bias of $\hat{r}_n$).
In view of (6)*, the asymptotic bias of* $\hat{r}_n$ *is minimum when* $\xi=0$*, that is, when* (γ_n) *is chosen so that* $\lim_{n\to\infty}(n\gamma_n)=\infty$*, and then we have*

$$\mathrm{E}\left[\hat{r}_n(\chi,h)\right]-r(\chi)=\mathrm{E}\left[\widetilde{r}_n(\chi,h)\right]-r(\chi)=\mathcal{B}(\chi)\frac{M_{\chi,0}}{M_{\chi,1}}h_n\left[1+o(1)\right].$$

This choice is not interesting since we have

$$\lim_{n\to\infty}\frac{\mathrm{var}\left(\widetilde{r}_n(\chi,h)\right)}{\mathrm{var}\left(\hat{r}_n(\chi,h)\right)}=0.$$

It is advised in this situation to use the averaging principle for the stochastic approximation algorithm to ensure an optimal convergence rate (see, [38]).

Example 3 (Choices of γ_n minimizing the variance of $\hat{r}_n$).
It follows from (8)*, that, to minimize the asymptotic variance of* $\hat{r}_n$*,* γ *should be equal to* 1*,* (γ_n) *must satisfy* $\lim_{n\to\infty}(n\gamma_n)=1-a$*, and we then have*

$$\mathrm{var}\left[\hat{r}_n(\chi,h)\right] = (1-\mathcal{F}_a)\mathcal{V}(\chi)\frac{M_{\chi,2}}{M_{\chi,1}^2}\frac{\gamma_n}{F_\chi(h_n)}\left[1+o(1)\right].$$

Using an adequate choice of γ_n*, the variance of the proposed recursive estimator* $\hat{r}_n$ *can be smaller than that of the nonrecursive estimator. To conclude this example, let us mention that the two most simple stepsizes that minimize the variance of* $\hat{r}_n$ *are* $\gamma_n=[1-a]\,n^{-1}$ *and* $\gamma_n=h_n\left(\sum_{k=1}^n h_k\right)^{-1}$.

Let us now state the following theorem, which gives the convergence rate of the estimator $\hat{r}_n$ defined in (2).

Theorem 1 (Weak pointwise convergence rate). *Let Assumptions* $(A1)-(A6)$ *be satisfied.*

1. *If there exists* $c\geq 0$ *such that* $\gamma_n^{-1}h_n^2F_\chi(h_n)\to c$*, then*

$$\begin{aligned}&\sqrt{\gamma_n^{-1}F_\chi(h_n)}\left(\hat{r}_n(\chi,h)-r(\chi)\right)\\ &\quad\xrightarrow{\mathcal{D}}\mathcal{N}\left(c^{1/2}\mathcal{B}(\chi)\frac{1-(\mathcal{F}_a-a)\xi}{1-\mathcal{F}_a\xi}\frac{M_{\chi,0}}{M_{\chi,1}},\mathcal{V}(\chi)\frac{(1-(\mathcal{F}_a-a)\xi)^2}{(2-(\mathcal{F}_a+\gamma-2a)\xi)}\frac{M_{\chi,2}}{M_{\chi,1}^2}\right).\end{aligned}$$

2. *If* $\gamma_n^{-1}h_n^2F_\chi(h_n)\to\infty$*, then*

$$\frac{1}{h_n}\left(\hat{r}_n(\chi,h)-r(\chi)\right)\xrightarrow{\mathrm{Pr}}\mathcal{B}(\chi)\frac{1-(\mathcal{F}_a-a)\xi}{1-\mathcal{F}_a\xi}\frac{M_{\chi,0}}{M_{\chi,1}},$$

where $\overset{\mathcal{D}}{\to}$ *denotes convergence in distribution,* $\mathcal{N}$ *is the Gaussian distribution, and* $\overset{\Pr}{\to}$ *denotes convergence in probability.*

Let us now consider the case where the bandwidth h_n is chosen so that $\lim_{n\to\infty}\gamma_n^{-1}h_n^2F_\chi(h_n)=0$ (which corresponds to under-smoothing). Thus, the proposed estimator fulfils the following central limit theorem:

$$\sqrt{\gamma_n^{-1}F_\chi(h_n)}\left(\hat{r}_n(\chi,h)-r(\chi)\right)\overset{\mathcal{D}}{\to}\quad\mathcal{N}\left(0,\mathcal{V}(\chi)\,\frac{(1-(\mathcal{F}_a-a)\xi)^2}{(2-(\mathcal{F}_a+\gamma-2a)\xi)}\,\frac{M_{\chi,2}}{M_{\chi,1}^2}\right).$$

Let ϕ denote the distribution function $\mathcal{N}(0,1)$, and $t_{\alpha/2}$ be so that $\phi(t_{\alpha/2})=1-t_{\alpha/2}$ (where $\alpha\in(0,1)$). Then the approximate asymptotic confidence band of $r(\chi)$, with level $1-\alpha$, is given by

$$\left[\hat{r}_n(\chi,h)-\phi(t_{\alpha/2})\,\sqrt{\gamma_n^{-1}\hat{F}_n(h_n)}\sqrt{\frac{(2-(\mathcal{F}_a+\gamma-2a)\xi)}{(1-(\mathcal{F}_a-a)\xi)^2}\,\frac{\hat{M}_1^2}{\hat{M}_2\hat{\mathcal{V}}(\chi)}},\right.$$
$$\left.\hat{r}_n(\chi,h)+\phi(t_{\alpha/2})\,\sqrt{\gamma_n^{-1}\hat{F}_n(h_n)}\sqrt{\frac{(2-(\mathcal{F}_a+\gamma-2a)\xi)}{(1-(\mathcal{F}_a-a)\xi)^2}\,\frac{\hat{M}_1^2}{\hat{M}_2\hat{\mathcal{V}}(\chi)}}\right],$$

where $\hat{F}_n$ is the empirical distribution function, and

$$\begin{aligned}
\hat{M}_i &= \sum_{k=1}^{n}\frac{\prod_n\prod_k^{-1}\gamma_k K^i\left(\frac{\|\chi-\mathcal{X}_k\|}{h_k}\right)}{\hat{F}(h_k)},\quad i\in\{1,2\},\\
\hat{\mathcal{V}}(\chi) &= \hat{g}_{n,2}(\chi,h)-4\hat{r}_n(\chi,h)\,\hat{g}_{n,3}(\chi,h)+3\hat{r}_n^2(\chi)\,\hat{g}_{n,4}(\chi,h),\\
\hat{g}_{n,\beta}(\chi,h) &= \frac{\prod_n\sum_{k=1}^{n}\prod_k^{-1}\gamma_k h_k^{-1}Y_k^{-\beta}K\left(\frac{\|\chi-\mathcal{X}_k\|}{h_k}\right)}{\prod_n\sum_{k=1}^{n}\prod_k^{-1}\gamma_k h_k^{-1}\mathrm{E}\left[K\left(\frac{\|\chi-\mathcal{X}_k\|}{h_k}\right)\right]}.
\end{aligned}\tag{9}$$

Remark 2. *By considering the choice of the stepsize* γ_n *that minimize the variance of* $\hat{r}_n$*, the approximate asymptotic confidence band of* $r(\chi)$ *with level* $1-\alpha$ *of the proposed recursive estimator is given by*

$$\left[\hat{r}_n(\chi,h)-\sqrt{1-\mathcal{F}_a}\phi(t_{\alpha/2})\,\sqrt{n\hat{F}_n(h_n)}\frac{\hat{M}_1^2}{\hat{M}_2\hat{\mathcal{V}}(\chi)},\hat{r}_n(\chi,h)+\sqrt{1-\mathcal{F}_a}\phi(t_{\alpha/2})\,\sqrt{n\hat{F}_n(h_n)}\frac{\hat{M}_1^2}{\hat{M}_2\hat{\mathcal{V}}(\chi)}\right],$$

however the approximate asymptotic confidence band of $r(\chi)$ *with level* $1-\alpha$ *of the nonrecursive estimator* $\widetilde{r}_n$ *is given by*

$$\left[\widetilde{r}_n(\chi,h)-\phi(t_{\alpha/2})\,\sqrt{n\hat{F}_n(h_n)}\frac{\widetilde{M}_1^2}{\widetilde{M}_2\widetilde{\mathcal{V}}(\chi)},\widetilde{r}_n(\chi,h)+\phi(t_{\alpha/2})\,\sqrt{n\hat{F}_n(h_n)}\frac{\widetilde{M}_1^2}{\widetilde{M}_2\widetilde{\mathcal{V}}(\chi)}\right],$$

where

$$\begin{aligned}
\widetilde{M}_i &= \frac{1}{n\hat{F}(h_n)}\sum_{k=1}^{n}K^i\left(\frac{\|\chi-\mathcal{X}_k\|}{h_n}\right),\quad i\in\{1,2\},\\
\widetilde{\mathcal{V}}(\chi) &= \widetilde{g}_{n,2}(\chi,h)-4\widetilde{r}_n(\chi,h)\,\widetilde{g}_{n,3}(\chi,h)+3\widetilde{r}_n^2(\chi)\,\widetilde{g}_{n,4}(\chi,h),\\
\widetilde{g}_{n,\beta}(\chi,h) &= \frac{\sum_{k=1}^{n}Y_k^{-\beta}K\left(\frac{\|\chi-\mathcal{X}_k\|}{h_n}\right)}{\sum_{k=1}^{n}\mathrm{E}\left[K\left(\frac{\|\chi-\mathcal{X}_k\|}{h_n}\right)\right]}.
\end{aligned}$$

Now, in order to measure the quality of our recursive estimator (2), we use the following quantity:

$$MSE\left[\hat{r}_n(\chi,h)\right] = \left(\mathrm{E}\left(\hat{r}_n(\chi,h)\right)-r(\chi)\right)^2+\mathrm{var}\left(\hat{r}_n(\chi,h)\right).$$

The next proposition gives the MSE of the proposed recursive estimator based on the minimization of the MSRE and defined in (2).

Proposition 2 (MSE of $\hat{r}_n(\chi, h)$). *Let Assumptions* $(A1) - (A6)$ *be satisfied,*

1. *If* $a \in (0, (\gamma - \mathcal{F}_a)/2)$*, then*

$$MSE\left[\hat{r}_n(\chi, h)\right] = \mathcal{B}^2(\chi)\left(\frac{1-(\mathcal{F}_a - a)\xi}{1-\mathcal{F}_a\xi}\right)^2 \frac{M_{\chi,0}^2}{M_{\chi,1}^2} h_n^2 + o\left(h_n^2\right).$$

2. *If* $a = (\gamma - \mathcal{F}_a)/2$*, then*

$$\begin{aligned} MSE\left[\hat{r}_n(\chi, h)\right] &= \mathcal{B}^2(\chi)\left(\frac{1-(\mathcal{F}_a - a)\xi}{1-\mathcal{F}_a\xi}\right)^2 \frac{M_{\chi,0}^2}{M_{\chi,1}^2} h_n^2 + \mathcal{V}(\chi) \frac{(1-(\mathcal{F}_a - a)\xi)^2}{(2-(\mathcal{F}_a + \gamma - 2a)\xi)} \frac{M_{\chi,2}}{M_{\chi,1}^2} \frac{\gamma_n}{F_\chi(h_n)} \\ &\quad + o\left(h_n^2 + \frac{\gamma_n}{F_\chi(h_n)}\right). \end{aligned}$$

3. *If* $a \in ((\gamma - \mathcal{F}_a)/2, 1)$*, then*

$$MSE\left[\hat{r}_n(\chi, h)\right] = \mathcal{V}(\chi) \frac{(1-(\mathcal{F}_a - a)\xi)^2}{(2-(\mathcal{F}_a + \gamma - 2a)\xi)} \frac{M_{\chi,2}}{M_{\chi,1}^2} \frac{\gamma_n}{F_\chi(h_n)} + o\left(\frac{\gamma_n}{F_\chi(h_n)}\right).$$

2.1. Bandwidth selection

In the framework of nonparametric kernel estimators, the data-driven bandwidth selection methods studied in the literature can be divided into three broad classes: cross-validation techniques, plug-in methods, and the bootstrap approach. A detailed comparison of the three techniques can be found in [16]. Here, we adapt the wild bootstrap procedure proposed in [27] to approximate the distribution of the error of our proposed recursive kernel regression estimator based on the minimization of the MSRE.

2.1.1. Wild bootstrap method

The wild bootstrap adapted first to functional data in [21], then under dependence conditions in [42], and more recently to the recursive estimation of the functional variable in [47]. The main idea of the wild bootstrap is that, rather than using the naive bootstrap approach of resampling from the pairs $\{Y_i, \mathcal{X}_i\}_{i=1}^n$, we resample from the estimated residuals $\hat{\varepsilon}_i = Y_i - \hat{r}_n(\mathcal{X}_i, h)$. Then, we use the obtained data to construct an estimator with a distribution that approximates the distribution of the original estimator, and where each bootstrap residual ε_i^* is drawn from a two-point distribution, such that $\mathrm{E}\left(\varepsilon_i^*\right) = 0$, $\mathrm{E}\left(\varepsilon_i^{*2}\right) = \hat{\varepsilon}_i^2$, and $\mathrm{E}\left(\varepsilon_i^{*3}\right) = \hat{\varepsilon}_i^3$. Such a distribution equals

$$G_i^* = \left(\frac{5+\sqrt{5}}{10}\right)\delta_{\hat{\varepsilon}_i\frac{(1-\sqrt{5})}{2}} + \left(\frac{5-\sqrt{5}}{10}\right)\delta_{\hat{\varepsilon}_i\frac{(1+\sqrt{5})}{2}}.$$

Our adapted procedure for the bandwidth selection when estimating recursively the operator r based on the minimization of the MSRE in the case of functional setting is performed in three steps:

Step 1: We construct the bootstrapped residuals ε_i^* drawn from the distribution G_i^*.

Step 2: Resampling, new observations $Y_i^{*-l} = \hat{g}_{n,l}(\chi_i, b) + Y_i^{-l}\varepsilon_i^*$, for $l \in \{1, 2\}$, where

$$\hat{g}_{n,l}(\chi, b) = \frac{\prod_n \sum_{k=1}^n \prod_k^{-1} \gamma_k b_k^{-1} K\left(\frac{\|\chi - \mathcal{X}_k\|}{b_k}\right) Y_k^{-l}}{\prod_n \sum_{k=1}^n \prod_k^{-1} \gamma_k b_k^{-1} \mathrm{E}\left[K\left(\frac{\|\chi - \mathcal{X}_k\|}{b_k}\right)\right]},$$

$b \in \{b_1, \ldots, b_n\}$ and b_n should be larger than h_n.

Step 3: Given the bootstrapped data $\left\{\mathcal{X}_i, Y_i^*\right\}_{i=1}^n$, we compute the kernel regression estimator,

$$\hat{r}_n^*(\chi, h) = \frac{\prod_n \sum_{k=1}^n \prod_k^{-1} \gamma_k h_k^{-1} K\left(\frac{\|\chi - \mathcal{X}_k\|}{h_k}\right) Y_k^{*-1}}{\prod_n \sum_{k=1}^n \prod_k^{-1} \gamma_k h_k^{-1} K\left(\frac{\|\chi - \mathcal{X}_k\|}{h_k}\right) Y_k^{*-2}}.$$

We repeat N_B times this wild bootstrapping scheme, and we use the empirical distribution $\hat{r}_n^*(\chi, h) - \hat{r}_n(\chi, b)$ for selecting an optimal bandwidth.

Given N_B replications of the proposed wild bootstrap procedure and given a fixed set H of bandwidths, the bootstrapped bandwidth h^* is defined by:

$$h^* = h^*(\chi) = \arg\min_{h \in H}\left(\frac{1}{N_B}\sum_{J=1}^{N_B}\left(\hat{r}_n^*(\chi, h) - \hat{r}_n(\chi, b)\right)^2\right). \tag{10}$$

We expect that the bootstrapped bandwidths will be close to the optimal ones. The good behavior of the bootstrapping approach (10) as an automatic bandwidth selection procedure was shown in Ferraty et al. [21] via some simulations in the case of regression estimation based on the minimization of the MSE. They indicate that both the theoretical quadratic loss and its data-driven bootstrapping approach have the same convex shape. Moreover, they compare the theoretical minimal quadratic loss with the error obtained by using the bootstrapped bandwidth h^*. As mentioned in Ferraty et al. [21], the theoretical support for this bootstrap bandwidth procedure is still an open question.

The wild bootstrap method in the case of the nonrecursive regression estimator based on the minimization of MSE is given in Ferraty et al. [21]. The case of the recursive regression estimator based on the minimization of the MSE is given in Slaoui [47].

Theorem 2. *Assume that Assumptions* $(A1) - (A7)$ *hold. Then, we have*

$$\sup_{y \in \mathbb{R}}\left|\Pr^{\mathcal{S}}\left(\sqrt{\gamma_n^{-1} F_\chi(h_n)}\left\{\hat{r}_n^*(\chi, h) - \hat{r}_n(\chi, b)\right\} \le y\right) - \Pr\left(\sqrt{\gamma_n^{-1} F_\chi(h_n)}\left\{\hat{r}_n(\chi, h) - r(\chi)\right\} \le y\right)\right| \overset{a.s.}{\to} 0,$$

where $\Pr^{\mathcal{S}}$ *denotes probability, conditionally on the the sample* $\mathcal{S}$ *(i.e.,* $(\mathcal{X}_i, Y_i)$*,* $i \in \{1, \ldots, n\}$*).*

Computational cost. The advantage of recursive estimators over their nonrecursive counterparts is that their update, from a sample of size n to one of size $n+1$, requires fewer computations. This property can be generalized, if we suppose that we receive separately two sets of data, the first one of cardinal n_1 that should be smaller or equal to $n-1$ and the second set of cardinal $n - n_1$. We can check that it follows from (1) that for $j \in \{1, 2\}$

$$\begin{aligned}
\hat{g}_{n,j}(\chi, h) &= \prod_{j=n_1+1}^{n}\left(1-\gamma_j\right)\hat{g}_{n_1,j}(\chi, h) + \sum_{k=n_1}^{n-1}\left[\prod_{j=k+1}^{n}\left(1-\gamma_j\right)\right]\frac{\gamma_k}{h_k}K\left(\frac{\|\chi - \mathcal{X}_k\|}{h_k}\right)Y_k^{-j} + \frac{\gamma_n}{h_n}K\left(\frac{\|\chi - \mathcal{X}_n\|}{h_n}\right)Y_n^{-j} \\
&= \alpha_1 \hat{g}_{n_1,j}(\chi, h) + \sum_{k=n_1}^{n-1}\beta_k \frac{\gamma_k}{h_k}K\left(\frac{\|\chi - \mathcal{X}_k\|}{h_k}\right)Y_k^{-j} + \frac{\gamma_n}{h_n}K\left(\frac{\|\chi - \mathcal{X}_n\|}{h_n}\right)Y_n^{-j},
\end{aligned}$$

where $\alpha_1 = \prod_{j=n_1+1}^{n}\left(1-\gamma_j\right)$ and $\beta_k = \prod_{j=k+1}^{n}\left(1-\gamma_j\right)$. It is clear, that we can use a bootstrap procedure to construct an optimal bandwidth based on the first sample of size n_1 and separately an optimal bandwidth based on the second sample of size $n - n_1$. Then the proposed estimator can be viewed as a linear combination of two estimators, which improves the computational cost significantly.

Remark 3. *The proposed recursive estimator is much better than the nonrecursive estimator in terms of computational times.*

3. Applications

The aim of this section is to compare the performance of the proposed recursive estimator $\hat{r}_n$ based on the minimization of the MSRE defined in (2) to the recursive functional regression estimator introduced by Slaoui [47] based on the minimization of the MSE and defined in (3) using a resampling bootstrap method.

3.1. A chemometric application

This data is available online at [20]. This time series of spectra has been measured from wavelengths $\lambda = 850$ to $\lambda = 1050$nm for 215 fined chopped pieces of meat. From this times series, we extracted the 215 spectra of light absorbance curves $\mathcal{X}_1, \ldots, \mathcal{X}_{215}$ as functions of the wavelength, discretized into $p = 100$ points. The response variable is the percentage of fat. These curves are displayed in Fig. 1.

In this functional context, the proposed estimator depends on the following parameters: Firstly, as measure of proximity we consider the semi-metric $d(\chi_1, \chi_2) = \sqrt{\int \left(\chi_1^{(1)}(t) - \chi_2^{(1)}(t)\right)^2 dt}$, where $\chi^{(1)}$ denotes the first derivatives of χ (since $\chi^{(1)}$ is unknown we used B-spline approximation (see, [18, 19])). Secondly, since the choice of the kernel function K was not crucial, we used the quadratic kernel $K(u) = \left(1 - u^2\right) \mathbb{1}_{[0,1]}(u)$ for all $u \in \mathbb{R}$. Moreover, in order to improve the speed of our proposed algorithm, the bandwidth h is assumed to belong to some grid in terms of k nearest neighbors, $h \in \{h_1, \ldots, h_{50}\} = H$, where h_k is the radius of the ball of center χ and containing exactly k data curves from $\mathcal{X}_1, \ldots, \mathcal{X}_{215}$ and where the number k is selected through a cross-validation procedure (here $k = 8$).

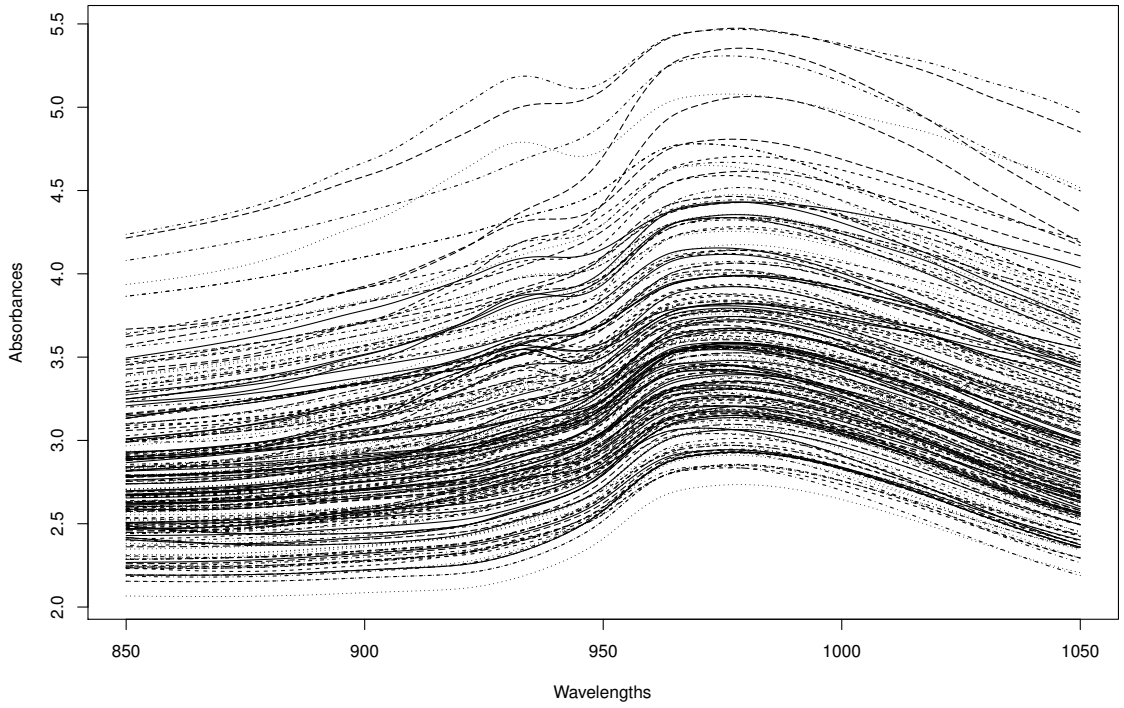

Fig. 1. These spectrometric curves data are available in [20].

Our main interest in this section is to compare the performance of two discussed estimators by determining the relation between the spectrum and the fatness by estimating a functional regression model using the recursive estimator (3) and using the recursive estimator (2). Notice that the routine `ODM` in the `R` Package `OutlierDM` [13] detected 20-outliers in the response variable Y.

For that reason, we decompose our sample of 215 pairs $(\mathcal{X}_i, Y_i)$ in a learning and a testing sample. The learning sample ($\mathcal{L}$) has size 195, on which the various statistical methods are constructed and a testing sample ($\mathcal{T}$) of size 20, which is used to examine the behavior of our method. We then measure the performance of the two estimators by the Mean Squared Prediction Error (MSPE):

$$\text{MSPE} = \frac{1}{20} \sum_{i \in \mathcal{T}} \left(\hat{Y}_i - Y_i\right)^2,$$

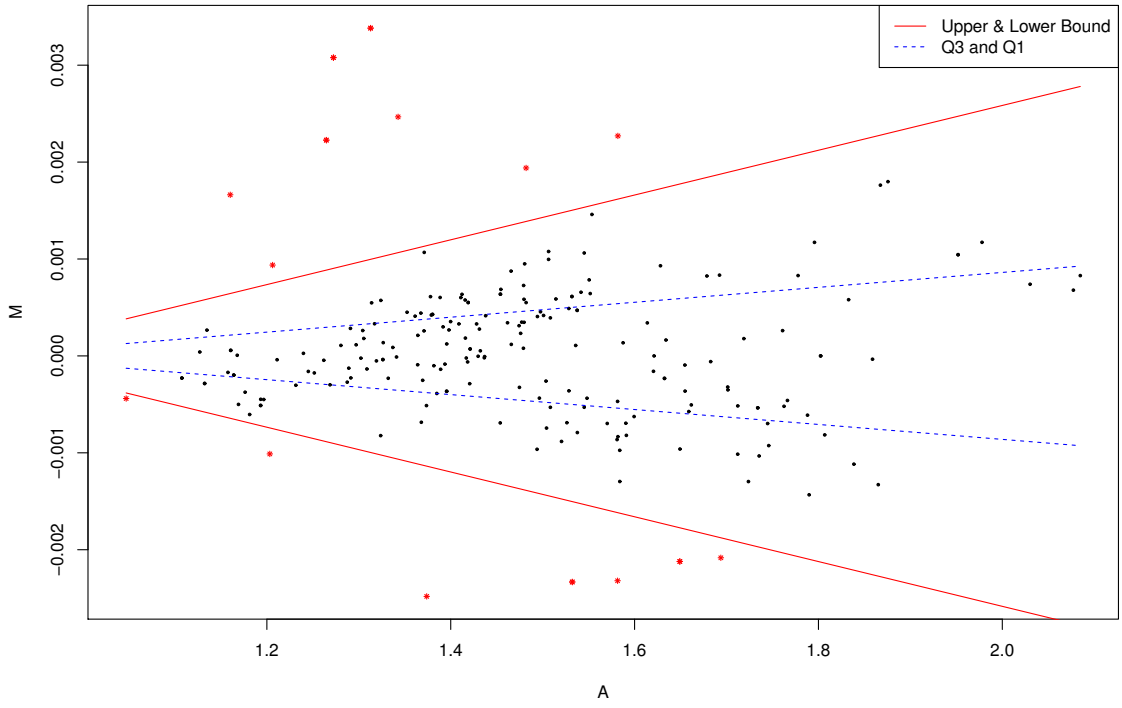

Fig. 2. Outlier detection for spectrometric data based on the MA plot. The MA plot visualizes the differences between measurements taken in two samples, by transforming the data onto M (log ratio) and A (mean average) scales.

and the Mean Squared Prediction Relative Error (MSPRE):

$$\text{MSPRE} = \frac{1}{20}\sum_{i\in\mathcal{T}}\left(\hat{Y}_i - Y_i\right)^2 Y_i^{-2},$$

where $\hat{Y}_i$ is the prediction for Y_i obtained for each new curve $\mathcal{X}_i$, $i \in \mathcal{T}$, using one of the two estimators.

Table 1
MSPE and MSPRE results for the spectrometric data. Two approaches were implemented, (1) "$\check{r}_n$" is the recursive estimator based on the minimization of the MSE, and (2) "$\hat{r}_n$" is our recursive estimator based on the minimization of the MSRE.

Estimator	MSPE	MSPRE
$\check{r}_n$	13.131	1.3302
$\hat{r}_n$	1.769	0.0915

We can observe from Table 1, that the proposed recursive estimator $\hat{r}_n$ based on the minimization of the MSRE gives a smaller MSPRE compared to the recursive estimator $\check{r}_n$ based on the minimization of the MSE. Imposing the quantile regression fence lines on a MA plot (see Fig. 2), then we can classify data points into outliers and non-outliers. In Fig. 2, the dashed line corresponds to the first and the third quantile and the continuous line corresponds to the lower and the upper bound. The lower and upper fences are respectively equal to $Q_1(A) - 1.5IQR(A)$ and $Q_1(A) + 1.5IQR(A)$, where $IQR(A) = Q_3(A) - Q_1(A)$. The q-quantile is given by

$$\arg\min_{\theta_i} q\sum_{i,M_i\geq\theta_i}|M_i-\theta_i| + (1-q)\sum_{i,M_i\leq\theta_i}|M_i-\theta_i|,$$

where $q \in]0,1[$ and $\theta_i = \beta_0 + \beta_1 A_i$, used first by Koenker and Bassett [34] and more recently by Cho et al. [12].

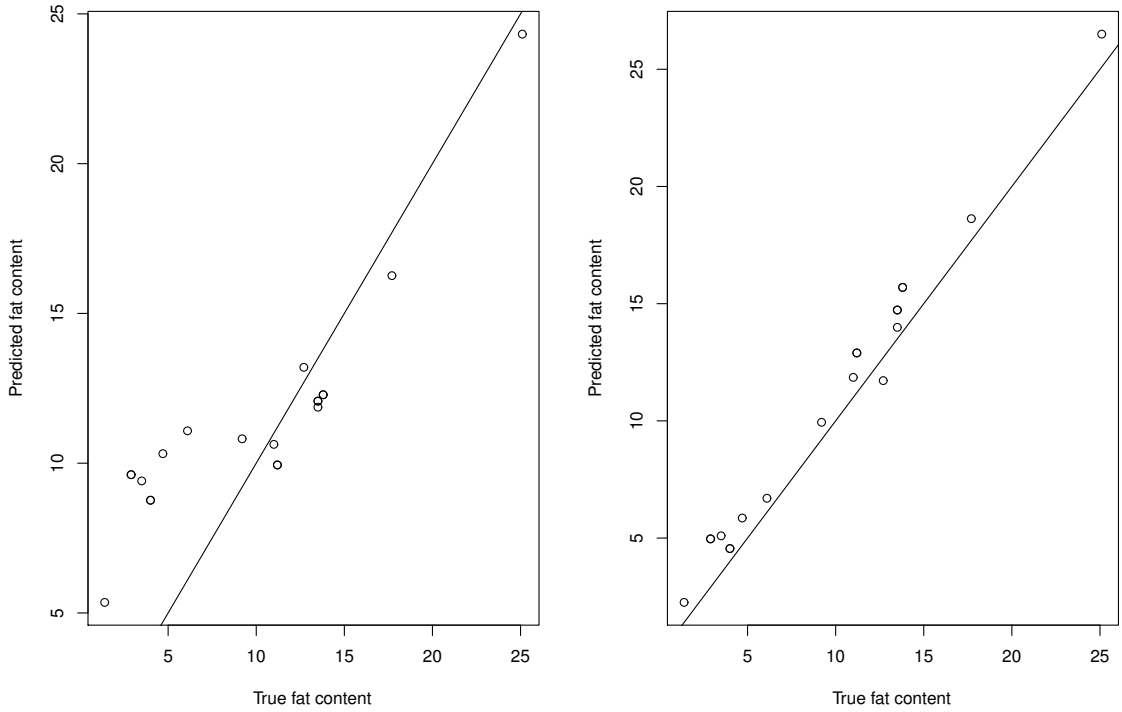

Fig. 3. Predicted values. Left plot: estimator (3) based on the MSE; Right plot: estimator (2) based on the MSRE.

Moreover, we plot in Fig. 3 the predicted values obtained using the two considered estimators as a function of the true one for the 20 spectra in our testing sample.

4. Conclusion

In this article, we propose an automatic selection of the bandwidth of recursive nonparametric relative regression for independent functional data. The proposed estimator asymptotically follows a normal distribution. The proposed estimator is compared to the recursive estimator based on the minimization of the MSE for functional data proposed by Slaoui [47] as an extension of the method introduced in Slaoui [46]. We showed that using some selected parameters, the proposed recursive estimator outperformed the recursive estimator introduced in [47]. The application study illustrates our finding.

We plan to extend this work by considering Bernstein polynomials rather than kernels and to propose an adaptation of the estimators developed in Jmaei et al. [30] and Slaoui and Jmaei [48] in the case of functional data. We plan also to compare these estimators to the kernel nearest-neighbor approach developed in Kara et al. [32], the semi-parametric functional projection pursuit regression [11], the single index model [25], the partial linear models [4, 35] and the sparse modeling approach [5].

A. Proofs

Before proving our main results, we state the following technical lemma.

Lemma 1. *Let $v_n \in \mathcal{GS}(v^*)$, $\gamma_n \in \mathcal{GS}(-\gamma)$, and $m > 0$ such that $m - v^*\xi > 0$ where ξ is defined in* (5)*. We have*

$$\lim_{n\to+\infty} v_n \sum_{k=1}^{n} \frac{\prod_n^m \gamma_k}{\prod_k^m v_k} = \frac{1}{m - v^*\xi}.$$

Moreover, for all positive sequences b_n such that $\lim_{n\to+\infty} b_n = 0$*,*

$$\lim_{n\to+\infty} v_n \sum_{k=1}^{n} \frac{\prod_n^m \gamma_k}{\prod_k^m v_k} b_k = 0.$$

Lemma 1 is widely applied throughout the proofs. Let us underline that it is its application, which requires Assumption (A2) (iii) on the limit of $(n\gamma_n)$ as n goes to infinity.
Our proofs are organized as follows. Proposition 1 and Theorem 1 are proved respectively, in Sections A.1 and A.2.

A.1. Proof of Proposition 1

First, we consider the following decomposition:

$$\begin{aligned}\mathrm{E}\left[\hat{r}_n(\chi,h)\right] - r(\chi) &= \frac{\mathrm{E}\left[\hat{g}_{n,1}(\chi,h)\right]}{\mathrm{E}\left[\hat{g}_{n,2}(\chi,h)\right]} - r(\chi) - \frac{\mathrm{E}\left\{\hat{g}_{n,1}(\chi,h)\left[\hat{g}_{n,2}(\chi,h) - \mathrm{E}\left(\hat{g}_{n,2}(\chi,h)\right)\right]\right\}}{\left\{\mathrm{E}\left[\hat{g}_{n,2}(\chi,h)\right]\right\}^2} \\ &\quad + \frac{\mathrm{E}\left\{\hat{r}_n(\chi,h)\left[\hat{g}_{n,2}(\chi,h) - \mathrm{E}\left(\hat{g}_{n,2}(\chi,h)\right)\right]^2\right\}}{\left\{\mathrm{E}\left[\hat{g}_{n,2}(\chi,h)\right]\right\}^2}. \end{aligned} \tag{11}$$

Now, we focus on computing the expectation of $\hat{g}_{n,\beta}(\chi,h)$. It follows from (9), that

$$\mathrm{E}\left[\hat{g}_{n,\beta}(\chi,h)\right] = \frac{\prod_n \sum_{k=1}^n \prod_k^{-1} \gamma_k h_k^{-1} \mathrm{E}\left[Y_k^{-\beta} K\left(\frac{\|\chi-\mathcal{X}_k\|}{h_k}\right)\right]}{\prod_n \sum_{k=1}^n \prod_k^{-1} \gamma_k h_k^{-1} \mathrm{E}\left[K\left(\frac{\|\chi-\mathcal{X}_k\|}{h_k}\right)\right]}.$$

Moreover, for $\alpha \in \mathbb{R}$, we have

$$\begin{aligned}\mathrm{E}\left[K^\alpha\left(\frac{\|\chi-\mathcal{X}_k\|}{h_k}\right)\right] &= \int_0^1 K^\alpha(u)\, d\mathrm{Pr}^{\left(\frac{\|\chi-\mathcal{X}_k\|}{h_k}\right)}(u) \\ &= F_\chi(h_k)\left[K^\alpha(1) - \int_0^1 \left(K^\alpha(u)\right)' \tau_{h_k}(u)\, du\right].\end{aligned}$$

Then, since $\lim_{n\to\infty}(n\gamma_n) > \mathcal{F}_a - a$, we apply Lemma 1 to infer that

$$\sum_{k=1}^{n} \frac{\prod_n \gamma_k}{\prod_k h_k} \mathrm{E}\left[K\left(\frac{\|\chi-\mathcal{X}_k\|}{h_k}\right)\right] = \frac{1}{1-(\mathcal{F}_a - a)\xi} M_{\chi,1} h_n^{-1} F_\chi(h_n)\left[1 + o(1)\right]. \tag{12}$$

Further, we have

$$\begin{aligned}\mathrm{E}\left[Y_k^{-\beta} K^\alpha\left(\frac{\|\chi-\mathcal{X}_k\|}{h_k}\right)\right] &= \mathrm{E}\left[K^\alpha\left(\frac{\|\chi-\mathcal{X}_k\|}{h_k}\right)\mathrm{E}\left[Y_k^{-\beta}|\mathcal{X}_k\right]|\mathcal{X}_k\right] \\ &= g_\beta(\chi)\,\mathrm{E}\left[K^\alpha\left(\frac{\|\chi-\mathcal{X}_k\|}{h_k}\right)\right] + \mathrm{E}\left[K^\alpha\left(\frac{\|\chi-\mathcal{X}_k\|}{h_k}\right)\mathrm{E}\left[\left\{g_\beta(\mathcal{X}_k) - g_\beta(\chi)\right\}|\,\|\mathcal{X}_k-\chi\|\right]\right] \\ &= g_\beta(\chi)\,\mathrm{E}\left[K^\alpha\left(\frac{\|\chi-\mathcal{X}_k\|}{h_k}\right)\right] + \mathrm{E}\left[K^\alpha\left(\frac{\|\chi-\mathcal{X}_k\|}{h_k}\right)\psi_{\chi,\beta}(\|\mathcal{X}_k-\chi\|)\right].\end{aligned}$$

A Taylor expansion of $\psi_{\chi,\beta}$ around 0 ensures that

$$\begin{aligned}\mathrm{E}\left[K^\alpha\left(\frac{\|\chi-\mathcal{X}_k\|}{h_k}\right)\psi_{\chi,\beta}(\|\mathcal{X}_k-\chi\|)\right] &= \int_0^1 \psi_{\chi,\beta}(h_k u)\, K^\alpha(u)\, d\mathrm{Pr}^{\left(\frac{\|\chi-\mathcal{X}_k\|}{h_k}\right)}(u) \\ &= h_k \psi'_{\chi,\beta}(0) \int_0^1 u K^\alpha(u)\, d\mathrm{Pr}^{\left(\frac{\|\chi-\mathcal{X}_k\|}{h_k}\right)}(u) + o(h_k).\end{aligned}$$

Moreover, it follows from the proof of Lemma 2 in Ferraty et al. [21], the Assumption $(A2)$ and Fubini's theorem

$$\int_0^1 uK^\alpha(u)\, d\Pr^{\left(\frac{\|\chi-\mathcal{X}_k\|}{h_k}\right)}(u) = F_\chi(h_k)\left[K^\alpha(1) - \int_0^1 (uK^\alpha(u))'\,\tau_{h_k}(u)\,du\right].$$

Then, we have, for all α and β in $\mathbb{R}$

$$\begin{aligned}\mathrm{E}\left[Y_k^{-\beta}K^\alpha\left(\frac{\|\chi-\mathcal{X}_k\|}{h_k}\right)\right] &= g_\beta(\chi)\,F_\chi(h_k)\left[K^\alpha(1) - \int_0^1 (K^\alpha(u))'\,\tau_{h_k}(u)\,du\right]\\ &\quad + h_k\psi'_{\chi,\beta}(0)\,F_\chi(h_k)\left[K^\alpha(1) - \int_0^1 (uK^\alpha(u))'\,\tau_{h_k}(u)\,du\right]. \qquad (13)\end{aligned}$$

Whence, since we have $\lim_{n\to\infty}(n\gamma_n) > \mathcal{F}_a$, the application of Lemma 1 gives

$$\mathrm{E}\left[\hat{g}_{n,\beta}(\chi,h)\right] = g_\beta(\chi) + \psi'_{\chi,\beta}(0)\,\frac{1-(\mathcal{F}_a-a)\,\xi}{1-\mathcal{F}_a\xi}\,\frac{M_{\chi,0}}{M_{\chi,1}}h_n\left[1+o(1)\right]. \qquad (14)$$

Now, we focus on computing the variance of $\hat{g}_{n,\beta}(\chi,h)$. It follows from (9), that

$$\mathrm{var}\left[\hat{g}_{n,\beta}(\chi,h)\right] = \frac{\prod_n^2\sum_{k=1}^n\prod_k^{-2}\gamma_k^2h_k^{-2}\mathrm{var}\left[Y_k^{-\beta}K\left(\frac{\|\chi-\mathcal{X}_k\|}{h_k}\right)\right]}{\left(\prod_n\sum_{k=1}^n\prod_k^{-1}\gamma_kh_k^{-1}\mathrm{E}\left[K\left(\frac{\|\chi-\mathcal{X}_k\|}{h_k}\right)\right]\right)^2}.$$

Moreover, it follows from (13)

$$\mathrm{var}\left[Y_k^{-\beta}K\left(\frac{\|\chi-\mathcal{X}_k\|}{h_k}\right)\right] = g_{2\beta}(\chi)\,F_\chi(h_k)\left[K^2(1) - \int_0^1\left(K^2(u)\right)'\tau_{h_k}(u)\,du\right]\left[1+o(1)\right].$$

Then, since $\lim_{n\to\infty}(n\gamma_n) > (\mathcal{F}_a+\gamma)/2-a$, we make use of Lemma 1 to infer that

$$\sum_{k=1}^n\frac{\prod_n^2\gamma_k^2}{\prod_k h_k^2}\mathrm{var}\left[Y_k^{-\beta}K\left(\frac{\|\chi-\mathcal{X}_k\|}{h_k}\right)\right] = \frac{1}{2-(\mathcal{F}_a+\gamma-2a)\,\xi}g_{2\beta}(\chi)\,M_{\chi,2}\frac{\gamma_n}{h_n^2}F_\chi(h_n)\left[1+o(1)\right]. \qquad (15)$$

The combination of (12) and (15) ensures that

$$\mathrm{var}\left[\hat{g}_{n,\beta}(\chi)\right] = \frac{(1-(\mathcal{F}_a-a)\,\xi)^2}{2-(\mathcal{F}_a+\alpha-2a)\,\xi}g_{2\beta}(\chi)\,\frac{M_{\chi,2}}{M_{\chi,1}^2}\,\frac{\gamma_n}{F_\chi(h_n)}\left[1+o(1)\right]. \qquad (16)$$

In order to compute the covariance between $\hat{g}_{n,1}(\chi,h)$ and $\hat{g}_{n,2}(\chi,h)$, we use (9), then we readily have

$$\begin{aligned}\mathrm{E}\left[\hat{g}_{n,1}(\chi,h)\,\hat{g}_{n,2}(\chi,h)\right] &= \frac{\prod_n^2\sum_{k=1}^n\prod_k^{-2}\gamma_k^2h_k^{-2}\mathrm{E}\left[Y_k^{-3}K^2\left(\frac{\|\chi-\mathcal{X}_k\|}{h_k}\right)\right]}{\left(\prod_n\sum_{k=1}^n\prod_k^{-1}\gamma_kh_k^{-1}\mathrm{E}\left[K\left(\frac{\|\chi-\mathcal{X}_k\|}{h_k}\right)\right]\right)^2}\\ &\quad + \frac{\prod_n^2\sum_{\substack{k,k'=1\\k\neq k'}}^n\prod_k^{-1}\Pi_{k'}^{-1}\gamma_k\gamma_{k'}h_k^{-1}\mathrm{E}\left[Y_k^{-1}K\left(\frac{\|\chi-\mathcal{X}_k\|}{h_k}\right)\right]h_{k'}^{-1}\mathrm{E}\left[Y_{k'}^{-2}K\left(\frac{\|\chi-\mathcal{X}_{k'}\|}{h_{k'}}\right)\right]}{\left(\prod_n\sum_{k=1}^n\prod_k^{-1}\gamma_kh_k^{-1}\mathrm{E}\left[K\left(\frac{\|\chi-\mathcal{X}_k\|}{h_k}\right)\right]\right)^2}.\end{aligned}$$

Since $\lim_{n\to\infty}(n\gamma_n) > (\mathcal{F}_a+\gamma)/2-a$, the application of Lemma 1 together with (12) and (13), ensures that

$$\mathrm{cov}\left[\hat{g}_{n,1}(\chi,h),\hat{g}_{n,2}(\chi,h)\right] = \frac{(1-(\mathcal{F}_a-a)\,\xi)^2}{2-(\mathcal{F}_a+\alpha-2a)\,\xi}g_3(\chi)\,\frac{M_{\chi,2}}{M_{\chi,1}^2}\,\frac{\gamma_n}{F_\chi(h_n)}\left[1+o(1)\right]. \qquad (17)$$

Let us now compute the expectation of $\hat{r}_n(\chi,h)$. First, it follows from (16), that

$$\mathrm{E}\left\{\hat{g}_{n,1}(\chi,h)\left[\hat{g}_{n,2}(\chi,h) - \mathrm{E}\left(\hat{g}_{n,2}(\chi,h)\right)\right]\right\} = O\left(\frac{\gamma_n}{F_\chi(h_n)}\right), \qquad (18)$$

$$\mathrm{E}\left\{\hat{r}_n(\chi,h)\left[\hat{g}_{n,2}(\chi,h) - \mathrm{E}\left(\hat{g}_{n,2}(\chi,h)\right)\right]^2\right\} = O\left(\frac{\gamma_n}{F_\chi(h_n)}\right). \qquad (19)$$

Then (6) follows from (11), (14), (18) and (19). Now, in order to compute the variance of $\hat{r}_n(\chi, h)$, we use the following decomposition

$$\begin{aligned}\operatorname{var}\left[\hat{r}_n(\chi,h)\right] &\simeq \frac{\operatorname{var}\left[\hat{g}_{n,1}(\chi,h)\right]}{\left\{\mathrm{E}\left[\hat{g}_{n,2}(\chi,h)\right]\right\}^2} - 4\frac{\mathrm{E}\left[\hat{g}_{n,1}(\chi,h)\right]\operatorname{cov}\left(\hat{g}_{n,1}(\chi,h)),\hat{g}_{n,2}(\chi,h)\right)}{\left\{\mathrm{E}\left[\hat{g}_{n,2}(\chi,h)\right]\right\}^3} \\ &\quad +3\operatorname{var}\left[\hat{g}_{n,2}(\chi,h)\right]\frac{\left\{\mathrm{E}\left[\hat{g}_{n,1}(\chi,h)\right]\right\}^2}{\left\{\mathrm{E}\left[\hat{g}_{n,2}(\chi,h)\right]\right\}^4}. \end{aligned} \tag{20}$$

The combination of (14), (16), (17) and (20), implies that

$$\operatorname{var}\left[\hat{r}_n(\chi,h)\right] = \mathcal{V}(\chi)\frac{M_{\chi,2}}{M_{\chi,1}^2}\frac{\left(1-(\mathcal{F}_a-a)\xi\right)^2}{\left(2-(\mathcal{F}_a+\gamma-2a)\xi\right)}\frac{\gamma_n}{F_\chi(h_n)}\left[1+o(1)\right].$$

A.2. Proof of Theorem 1

Let us at first assume that, if $a \geq (\gamma+\mathcal{F}_a)/2$, then

$$\sqrt{\gamma_n^{-1}F_\chi(h_n)}\left(\hat{r}_n(\chi,h)-\mathrm{E}\left[\hat{r}_n(\chi,h)\right]\right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0,\mathcal{V}(\chi)\frac{\left(1-(\mathcal{F}_a-a)\xi\right)^2}{\left(2-(\mathcal{F}_a+\gamma-2a)\xi\right)}\frac{M_{\chi,2}}{M_{\chi,1}^2}\right). \tag{21}$$

In the case when $\gamma_n^{-1}h_n^2F_\chi(h_n)\to c$, Part 1 of Theorem 1 follows from the combination of (6) and (21). In the case $\gamma_n^{-1}h_n^2F_\chi(h_n)\to\infty$, (7) implies that

$$h_n^{-2}\left(\hat{r}_n(\chi,h)-\mathrm{E}\left(\hat{r}_n(\chi,h)\right)\right) \xrightarrow{\Pr} 0,$$

and the application of (6) gives Part 2 of Theorem 1.

We now prove (21). Using the fact that, for x such that $\hat{g}_{n,2}(\chi,h)\neq 0$, we have the following decomposition:

$$\hat{r}_n(\chi,h)-r(\chi) = \mathcal{D}_n(\chi,h)\frac{g_2(\chi)}{\hat{g}_{n,2}(\chi,h)}, \tag{22}$$

with

$$\mathcal{D}_n(\chi,h) = \frac{1}{g_2(\chi)}\left\{\hat{g}_{n,1}(\chi,h)-r(\chi)\,\hat{g}_{n,2}(\chi,h)\right\}.$$

It follows from (22), that the asymptotic behavior of $\hat{r}_n(\chi,h)-r(\chi)$ can be deduced from the one of $\mathcal{D}_n(\chi,h)$. Let us set

$$\begin{aligned} Z_k(\chi) &= \frac{\gamma_k}{\prod_k h_k}\left\{Y_k^{-1}-r(\chi)\,Y_k^{-2}\right\}K\left(\frac{\|\chi-\mathcal{X}_k\|}{h_k}\right), \\ \hat{f}_n(\chi) &= \sum_{k=1}^n\frac{\prod_n\gamma_k}{\prod_k h_k}K\left(\frac{\|\chi-\mathcal{X}_k\|}{h_k}\right) \end{aligned}$$

and

$$T_k(\chi) = Z_k(\chi)-\mathrm{E}\left[Z_k(\chi)\right]. \tag{23}$$

Hence, we directly have

$$\mathcal{D}_n(\chi,h)-\mathrm{E}\left[\mathcal{D}_n(\chi,h)\right] = \frac{\prod_n}{g_2(\chi)\,\mathrm{E}\left[\hat{f}_n(\chi)\right]}\sum_{k=1}^n T_k(\chi).$$

Since $\lim_{n\to\infty}(n\gamma_n) > (\mathcal{F}_a+\gamma)/2-a$, we apply Lemma 1 as well as (12), (13) and (23), to deduce that

$$\begin{aligned} v_n^2 &= \sum_{k=1}^{n}\operatorname{var}\left(T_k(\chi)\right)=\sum_{k=1}^{n}\frac{\gamma_k^2}{\prod_k^2 h_k^2}\operatorname{var}\left[\left\{Y_k^{-1}-r(\chi)\,Y_k^{-2}\right\}K\left(\frac{\|\chi-\mathcal{X}_k\|}{h_k}\right)\right] \\ &= O\left(\sum_{k=1}^{n}\frac{\gamma_k^2}{\prod_k^2 h_k^2}F_\chi(h_k)\right)=O\left(\frac{\gamma_n F_\chi(h_n)}{\prod_n^2 h_n^2}\right). \end{aligned}$$

Moreover, using the fact that, for all $p>0$,

$$\mathrm{E}\left[|Z_k(\chi)|^{2+p}\right] = O\left(\frac{F_\chi(h_k)}{h_k^{1+p}}\right),$$

and, since $\lim_{n\to\infty}(n\gamma_n) > (\mathcal{F}_a+\gamma)/2-a$, there exists a $p>0$ such that $\lim_{n\to\infty}(n\gamma_n) > \frac{1+p}{2+p}\left((\mathcal{F}_a+\gamma)/2-a\right)$. Applying Lemma 1, we get

$$\begin{aligned} \sum_{k=1}^{n}\mathrm{E}\left[|T_k(\chi)|^{2+p}\right] &= O\left(\sum_{k=1}^{n}\frac{\gamma_k^{2+p}}{\prod_k^{2+p}}\mathrm{E}\left[|Z_k(\chi)|^{2+p}\right]\right)=O\left(\sum_{k=1}^{n}\frac{\prod_k^{-2-p}\gamma_k^{2+p}F_\chi(h_k)}{h_k^{1+p}}\right) \\ &= O\left(\frac{\gamma_n^{1+p}F_\chi(h_n)}{\prod_n^{2+p}h_n^{1+p}}\right), \end{aligned}$$

and we thus obtain

$$\frac{1}{v_n^{2+p}}\sum_{k=1}^{n}\mathrm{E}\left[|T_k(\chi)|^{2+p}\right] = O\left(\left[\gamma_n h_n^{-1}F_\chi(h_n)\right]^{p/2}\right)=o(1).$$

The convergence in (21) then follows from the application of Lyapounov's theorem.

A.3. Proof of Theorem 2

The proof is based on the following decomposition:

$$\begin{aligned} \Pr{}^{\mathcal{S}}\left(\sqrt{\gamma_n^{-1}F_\chi(h_n)}\left\{\hat{r}_n^*(\chi,h)-\hat{r}_n(\chi,b)\right\}\le y\right)- \Pr\left(\sqrt{\gamma_n^{-1}F_\chi(h_n)}\left\{\hat{r}_n(\chi,h)-r(\chi)\right\}\le y\right) \\ =\mathcal{T}_1(y)+\mathcal{T}_2(y)+\mathcal{T}_3(y), \end{aligned}$$

where

$$\begin{aligned} \mathcal{T}_1(y) &= \Pr{}^{\mathcal{S}}\left(\sqrt{\gamma_n^{-1}F_\chi(h_n)}\left\{\hat{r}_n^*(\chi,h)-\hat{r}_n(\chi,b)\right\}\le y\right)-\Phi\left(\frac{y-\sqrt{\gamma_n^{-1}F_\chi(h_n)}\left\{\mathrm{E}^{\mathcal{S}}\left[\hat{r}_n^*(\chi,h)\right]-\hat{r}_n(\chi,b)\right\}}{\sqrt{\gamma_n^{-1}F_\chi(h_n)\,Var^{\mathcal{S}}\left[\hat{r}_n^*(\chi,h)\right]}}\right), \\ \mathcal{T}_2(y) &= \Phi\left(\frac{y-\sqrt{\gamma_n^{-1}F_\chi(h_n)}\left\{\mathrm{E}\left[\hat{r}_n(\chi,h)\right]-r(\chi)\right\}}{\sqrt{\gamma_n^{-1}F_\chi(h_n)\operatorname{var}\left[\hat{r}_n(\chi,h)\right]}}\right)-\Pr\left(\sqrt{\gamma_n^{-1}F_\chi(h_n)}\left\{\hat{r}_n(\chi,h)-r(\chi)\right\}\le y\right), \\ \mathcal{T}_3(y) &= \Phi\left(\frac{y-\sqrt{\gamma_n^{-1}F_\chi(h_n)}\left\{\mathrm{E}^{\mathcal{S}}\left[\hat{r}_n^*(\chi,h)\right]-\hat{r}_n(\chi,b)\right\}}{\sqrt{\gamma_n^{-1}F_\chi(h_n)\,Var^{\mathcal{S}}\left[\hat{r}_n^*(\chi,h)\right]}}\right)-\Phi\left(\frac{y-\sqrt{\gamma_n^{-1}F_\chi(h_n)}\left\{\mathrm{E}\left[\hat{r}_n(\chi,h)\right]-r(\chi)\right\}}{\sqrt{\gamma_n^{-1}F_\chi(h_n)\operatorname{var}\left[\hat{r}_n(\chi,h)\right]}}\right), \end{aligned}$$

where $\mathrm{E}^{\mathcal{S}}$ and $Var^{\mathcal{S}}$ denote expectation and variance, conditionally on the sample $\mathcal{S}$, and Φ denotes the standard normal distribution function. The first part of Theorem 1 ensures that

$$\mathcal{T}_2(y)\to 0 \quad \text{a.s.} \quad \forall y\in\mathbb{R}. \tag{24}$$

Moreover,

$$\mathcal{T}_1(y)\to 0 \quad \text{a.s.} \quad \forall y\in\mathbb{R}, \tag{25}$$

follows from the next lemma.

Lemma 2. *Assume that Assumptions* $(A1) - (A6)$ *hold. Then, we have*

$$\frac{\hat{r}_n^*(\chi,h) - \mathrm{E}^{\mathcal{S}}\left[\hat{r}_n^*(\chi,h)\right]}{\sqrt{Var^{\mathcal{S}}\left[\hat{r}_n^*(\chi,h)\right]}} \xrightarrow{d} \mathcal{N}(0,1).$$

Now, from (24), (25) and Polya's theorem (see, e.g., Serfling [44], p. 18) together with the continuity of the function Φ, we arrive at

$$\sup_{y\in\mathbb{R}} |\mathcal{T}_1(y)| + \sup_{y\in\mathbb{R}} |\mathcal{T}_2(y)| \to 0 \quad \text{a.s.} \quad \forall y \in \mathbb{R}.$$

Finally, it remains to study the term $\mathcal{T}_3(y)$. Using the fact that, for any $a > 0$ and $c \in \mathbb{R}$,

$$\sup_{y\in\mathbb{R}} |\Phi(c + ay) - \Phi(y)| \leq |c| + \max\left\{a, a^{-1}\right\} - 1,$$

and considering

$$a = \sqrt{\frac{\mathrm{var}\left[\hat{r}_n(\chi,h)\right]}{Var^{\mathcal{S}}\left[\hat{r}_n^*(\chi,h)\right]}}$$

and

$$c = \frac{\sqrt{\gamma_n^{-1}F_\chi(h_n)}\left\{\mathrm{E}\left[\hat{r}_n(\chi,h)\right] - r(\chi) - \mathrm{E}^{\mathcal{S}}\left[\hat{r}_n^*(\chi,h)\right] + \hat{r}_n(\chi,b)\right\}}{\sqrt{\gamma_n^{-1}F_\chi(h_n)\, Var^{\mathcal{S}}\left[\hat{r}_n^*(\chi,h)\right]}},$$

we get

$$\begin{aligned}\sup_{y\in\mathbb{R}} |\mathcal{T}_3(y)| \quad \leq \quad & \left| \frac{\sqrt{\gamma_n^{-1}F_\chi(h_n)}\left\{\mathrm{E}\left[\hat{r}_n(\chi,h)\right] - r(\chi) - \mathrm{E}^{\mathcal{S}}\left[\hat{r}_n^*(\chi,h)\right] + \hat{r}_n(\chi,b)\right\}}{\sqrt{\gamma_n^{-1}F_\chi(h_n)\, Var^{\mathcal{S}}\left[\hat{r}_n^*(\chi,h)\right]}} \right| \\ & + \max\left\{ \sqrt{\frac{\mathrm{var}\left[\hat{r}_n(\chi,h)\right]}{Var^{\mathcal{S}}\left[\hat{r}_n^*(\chi,h)\right]}}, \sqrt{\frac{Var^{\mathcal{S}}\left[\hat{r}_n^*(\chi,h)\right]}{\mathrm{var}\left[\hat{r}_n(\chi,h)\right]}} \right\} - 1. \end{aligned} \tag{26}$$

The combination of (8), (26) and the following two lemmas ensure the convergence of $\sup_{y\in\mathbb{R}} |\mathcal{T}_3(y)|$.

Lemma 3. *Assume that Assumptions* $(A1) - (A6)$ *hold. Then*

$$\sqrt{\frac{\mathrm{var}\left[\hat{r}_n(\chi,h)\right]}{Var^{\mathcal{S}}\left[\hat{r}_n^*(\chi,h)\right]}} \to 1 \quad a.s.$$

Lemma 4. *Assume that Assumptions* $(A1) - (A7)$ *hold. Then*

$$\sqrt{\gamma_n^{-1}F_\chi(h_n)}\left\{\mathrm{E}\left[\hat{r}_n(\chi,h)\right] - r(\chi) - \mathrm{E}^{\mathcal{S}}\left[\hat{r}_n^*(\chi,h)\right] + \hat{r}_n(\chi,b)\right\} \to 0 \quad a.s.$$

A.3.1. Proof of Lemma 2

Using the fact that, for x such that $\hat{g}_{n,2}^*(\chi) \neq 0$, we have the following decomposition:

$$\hat{r}_n^*(\chi,h) - r(\chi) \quad = \quad \mathcal{D}_n^*(\chi,h)\frac{g_2(\chi)}{\hat{g}_{n,2}^*(\chi,h)}, \tag{27}$$

with

$$\mathcal{D}_n^*(\chi,h) \quad = \quad \frac{1}{g_2(\chi)}\left\{\hat{g}_{n,1}^*(\chi,h) - r(\chi)\,\hat{g}_{n,2}^*(\chi,h)\right\}.$$

It follows from (27), that the asymptotic behavior of $\hat{r}_n^*(\chi,h) - r(\chi)$ can be deduced from the one of $\mathcal{D}_n^*(\chi,h)$. Let us set

$$Z_k^*(\chi) = \frac{\gamma_k}{\prod_k h_k}\left\{Y_k^{*-1} - r(\chi)\, Y_k^{*-2}\right\} K\left(\frac{\|\chi - \mathcal{X}_k\|}{h_k}\right)$$

and

$$T_k^*(\chi) = Z_k^*(\chi) - \mathrm{E}^{\mathcal{S}}\left[Z_k^*(\chi)\right]. \tag{28}$$

Hence, we readily infer that

$$\mathcal{D}_n^*(\chi,h) - \mathrm{E}^{\mathcal{S}}\left[\mathcal{D}_n^*(\chi,h)\right] = \frac{\prod_n}{g_2(\chi)\,\mathrm{E}\left[\hat{f}_n(\chi)\right]}\sum_{k=1}^n T_k^*(\chi).$$

Moreover, since we have for $l \in \{1,2,3\}$

$$Var^{\mathcal{S}}\left[Y_k^{*-l}\right] = \mathrm{E}\left[Y_k^{-2l}\right]\varepsilon_k^2, \quad \mathrm{E}^{\mathcal{S}}\left[Y_k^{*-l}\right] = \hat{g}_{n,l}(\chi_k, b),$$

and given that $\lim_{n\to\infty}(n\gamma_n) > (\mathcal{F}_a + \gamma)/2 - a$, the application of Lemma 1 together with (12), (13) and (28), ensures that

$$\begin{aligned} v_n^{*2} &= \sum_{k=1}^n Var^{\mathcal{S}}\left(T_k^*(\chi)\right) = \sum_{k=1}^n \frac{\gamma_k^2}{\prod_k^2 h_k^2} Var^{\mathcal{S}}\left[\left\{Y_k^{*-1} - r(\chi)\, Y_k^{*-2}\right\} K\left(\frac{\|\chi - \mathcal{X}_k\|}{h_k}\right)\right] \\ &= O\left(\sum_{k=1}^n \frac{\gamma_k^2}{\prod_k^2 h_k^2} F_\chi(h_k)\right) = O\left(\frac{\gamma_n F_\chi(h_n)}{\prod_n^2 h_n^2}\right). \end{aligned}$$

Now, we have, for all $p > 0$,

$$\mathrm{E}\left[\left|Z_k^*(\chi)\right|^{2+p}\right] = O\left(\frac{F_\chi(h_k)}{h_k^{1+p}}\right),$$

and, since $\lim_{n\to\infty}(n\gamma_n) > (\mathcal{F}_a + \gamma)/2 - a$, there exists $p > 0$ such that $\lim_{n\to\infty}(n\gamma_n) > \frac{1+p}{2+p}\left((\mathcal{F}_a + \gamma)/2 - a\right)$. Applying again Lemma 1, we get

$$\begin{aligned} \sum_{k=1}^n \mathrm{E}^{\mathcal{S}}\left[\left|T_k^*(\chi)\right|^{2+p}\right] &= O\left(\sum_{k=1}^n \frac{\gamma_k^{2+p}}{\prod_k^{2+p}} \mathrm{E}^{\mathcal{S}}\left[\left|Z_k^*(\chi)\right|^{2+p}\right]\right) = O\left(\sum_{k=1}^n \frac{\prod_k^{-2-p}\gamma_k^{2+p} F_\chi(h_k)}{h_k^{1+p}}\right) \\ &= O\left(\frac{\gamma_n^{1+p} F_\chi(h_n)}{\prod_n^{2+p} h_n^{1+p}}\right), \end{aligned}$$

and thus we obtain

$$\frac{1}{\left|v_n^*\right|^{2+p}}\sum_{k=1}^n \mathrm{E}^{\mathcal{S}}\left[\left|T_k^*(\chi)\right|^{2+p}\right] = O\left(\left[\gamma_n h_n^{-1} F_\chi(h_n)\right]^{p/2}\right) = o(1).$$

The convergence in (21) then follows from the application of Lyapounov's theorem.

A.3.2. Proof of Lemma 3

First, we use the following decomposition:

$$\begin{aligned} Var^{\mathcal{S}}\left[\hat{r}_n^*(\chi,h)\right] \simeq\ & \frac{Var^{\mathcal{S}}\left[\hat{g}_{n,1}^*(\chi,h)\right]}{\left\{\mathrm{E}^{\mathcal{S}}\left[\hat{g}_{n,2}^*(\chi,h)\right]\right\}^2} - 4\frac{\mathrm{E}^{\mathcal{S}}\left[\hat{g}_{n,1}^*(\chi,h)\right] Cov^{\mathcal{S}}\left(\hat{g}_{n,1}^*(\chi,h)), \hat{g}_{n,2}^*(\chi,h)\right)}{\left\{\mathrm{E}^{\mathcal{S}}\left[\hat{g}_{n,2}^*(\chi,h)\right]\right\}^3} \\ & + 3Var^{\mathcal{S}}\left[\hat{g}_{n,2}^*(\chi,h)\right]\frac{\left\{\mathrm{E}^{\mathcal{S}}\left[\hat{g}_{n,1}^*(\chi,h)\right]\right\}^2}{\left\{\mathrm{E}^{\mathcal{S}}\left[\hat{g}_{n,2}^*(\chi,h)\right]\right\}^4}. \end{aligned} \tag{29}$$

Moreover, we have for $l \in \{1,2\}$

$$Var^{\mathcal{S}}\left[\hat{g}^*_{n,l}(\chi,h)\right] = \frac{\prod_n^2 \sum_{k=1}^n \prod_k^{-2} \gamma_k^2 h_k^{-2} K^2\left(\frac{\|\chi-\mathcal{X}_k\|}{h_k}\right) Var^{\mathcal{S}}\left[Y_k^{*-l}\right]}{\left(\prod_n \sum_{k=1}^n \prod_k^{-1} \gamma_k h_k^{-1} \mathrm{E}\left[K\left(\frac{\|\chi-\mathcal{X}_k\|}{h_k}\right)\right]\right)^2}, \tag{30}$$

$$\mathrm{E}^{\mathcal{S}}\left[\hat{g}^*_{n,l}(\chi,h)\right] = \frac{\prod_n \sum_{k=1}^n \prod_k^{-1} \gamma_k h_k^{-1} K\left(\frac{\|\chi-\mathcal{X}_k\|}{h_k}\right) \mathrm{E}^{\mathcal{S}}\left[Y_k^{*-l}\right]}{\prod_n \sum_{k=1}^n \prod_k^{-1} \gamma_k h_k^{-1} \mathrm{E}\left[K\left(\frac{\|\chi-\mathcal{X}_k\|}{h_k}\right)\right]} \tag{31}$$

and

$$\begin{aligned}\mathrm{E}^{\mathcal{S}}\left[\hat{g}^*_{n,1}(\chi,h)\,\hat{g}^*_{n,2}(\chi,h)\right] &= \frac{\prod_n^2 \sum_{k=1}^n \prod_k^{-2} \gamma_k^2 h_k^{-2} K^2\left(\frac{\|\chi-\mathcal{X}_k\|}{h_k}\right) \mathrm{E}^{\mathcal{S}}\left[Y_k^{*-3}\right]}{\left(\prod_n \sum_{k=1}^n \prod_k^{-1} \gamma_k h_k^{-1} \mathrm{E}\left[K\left(\frac{\|\chi-\mathcal{X}_k\|}{h_k}\right)\right]\right)^2} \\ &+ \frac{\prod_n^2 \sum_{\substack{k,k'=1 \\ k\neq k'}}^n \prod_k^{-1} \Pi_{k'}^{-1} \gamma_k \gamma_{k'} h_k^{-1} h_{k'}^{-1} K\left(\frac{\|\chi-\mathcal{X}_k\|}{h_k}\right) K\left(\frac{\|\chi-\mathcal{X}_{k'}\|}{h_{k'}}\right) \mathrm{E}^{\mathcal{S}}\left[Y_k^{*-1}\right] \mathrm{E}^{\mathcal{S}}\left[Y_{k'}^{*-2}\right]}{\left(\prod_n \sum_{k=1}^n \prod_k^{-1} \gamma_k h_k^{-1} \mathrm{E}\left[K\left(\frac{\|\chi-\mathcal{X}_k\|}{h_k}\right)\right]\right)^2}. \end{aligned} \tag{32}$$

Since we have for $l \in \{1,2,3\}$

$$Var^{\mathcal{S}}\left[Y_k^{*-l}\right] = \mathrm{E}\left[Y_k^{-2l}\right]\varepsilon_k^2, \quad \mathrm{E}^{\mathcal{S}}\left[Y_k^{*-l}\right] = \hat{g}_{n,l}(\chi_k,b), \tag{33}$$

the combination of (29), (30), (31), (32), (33) and some computational analysis ensures that

$$Var^{\mathcal{S}}\left[\hat{r}^*_n(\chi,h)\right] = \mathrm{var}\left[\hat{r}_n(\chi,h)\right](1+o(1)).$$

A.3.3. Proof of Lemma 4

This proof follows the same steps as those used in Ferraty et al. [22]. First, we use the following decomposition:

$$\mathrm{E}^{\mathcal{S}}\left[\hat{r}^*_n(\chi,h)\right] - \hat{r}_n(\chi,b) = U_1 + U_2 + U_3,$$

where

$$\begin{aligned} U_1 &= \mathcal{J}_1^{-1} \sum_{k=1}^n \frac{\prod_n \gamma_k}{\prod_k h_k} K\left(\frac{\|\chi-\mathcal{X}_k\|}{h_k}\right) \{\hat{g}_{n,1}(\chi_k,b) - \hat{r}_n(\chi,b)\,\hat{g}_{n,2}(\chi_k,b) \\ &\quad - \mathrm{E}\left[\hat{g}_{n,1}(\chi_k,b)\right] + \mathrm{E}\left[\hat{r}_n(\chi,b)\,\hat{g}_{n,2}(\chi_k,b)\right]\}, \\ U_2 &= \mathcal{J}_1^{-1} \sum_{k=1}^n \frac{\prod_n \gamma_k}{\prod_k h_k} K\left(\frac{\|\chi-\mathcal{X}_k\|}{h_k}\right) \{\mathrm{E}\left[\hat{g}_{n,1}(\chi_k,b)\right] - \mathrm{E}\left[\hat{r}_n(\chi,b)\,\hat{g}_{n,2}(\chi_k,b)\right] \\ &\quad - g_1(\chi_k) + r(\chi)\,g_2(\chi_k)\}, \\ U_3 &= \mathcal{J}_1^{-1} \sum_{k=1}^n \frac{\prod_n \gamma_k}{\prod_k h_k} K\left(\frac{\|\chi-\mathcal{X}_k\|}{h_k}\right) \{g_1(\chi) - r(\chi)\,g_2(\chi_k)\}, \\ \mathcal{J}_1 &= \sum_{k=1}^n \frac{\prod_n \gamma_k}{\prod_k h_k} K\left(\frac{\|\chi-\mathcal{X}_k\|}{h_k}\right) \hat{g}_{n,2}(\chi_k,b). \end{aligned}$$

Moreover, we can check that $U_3 = \mathrm{E}\left[\hat{r}_n(\chi,h)\right] - r(\chi) + o\left(\sqrt{\gamma_n^{-1} F_\chi(h_n)}\right)$ a.s., whereas U_1 and U_2 are $o\left(\sqrt{\gamma_n^{-1} F_\chi(h_n)}\right)$ a.s. by Lemmas 5 and 6.

Lemma 5. *Assume that Assumptions* $(A1) - (A6)$ *hold. Then*

$$\sup_{\|\chi-\chi_1\|\leq h} \left|\mathrm{E}\left[\hat{g}_{n,1}(\chi,b)\right] - \mathrm{E}\left[\hat{r}_n(\chi,b)\,\hat{g}_{n,2}(\chi_1,b)\right] - g_1(\chi_1) + r(\chi)\,g_2(\chi_1)\right| = o\left(\sqrt{\gamma_n^{-1} F_\chi(h_n)}\right) \quad a.s.$$

Lemma 6. *Assume that Assumptions* $(A1) - (A6)$ *hold. Then*

$$\sup_{\|\chi-\chi_1\|\leq h} \left|\mathrm{E}\left[\hat{g}_{n,1}\left(\chi_1, b\right)\right)\right] - \mathrm{E}\left[\hat{r}_n\left(\chi, b\right)\hat{g}_{n,2}\left(\chi_1, b\right)\right] - \hat{g}_{n,1}\left(\chi_1, b\right) + \hat{r}_n\left(\chi, b\right)\hat{g}_{n,2}\left(\chi_1, b\right)\right| = o\left(\sqrt{\gamma_n^{-1} F_\chi\left(h_n\right)}\right) \quad a.s.$$

The proof of Lemmas 5 and 6 are obtained by following the same lines and decompositions used in the proof of Lemmas A.5 and A.6 given in Ferraty et al. [22].

Acknowledgments

The author would like to thank the Editor-in-Chief of JMVA, Prof. D. von Rosen and three anonymous reviewers for their helpful comments, which helped me to focus on improving the original version of the paper.

References

[1] A. Amiri, C. Crambes, B. Thiam, Recursive estimation of nonparametric regression with functional covariate. *Comput. Statist. Data Anal.* 69 (2014) 154–172.

[2] G. Aneiros, R. Cao, R. Fraiman, C. Genest, P. Vieu, Recent advances in functional data analysis and high-dimensional statistics, *J. Multivariate Anal.* 170 (2019) 3–9.

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

[4] G. Aneiros, P. Vieu, Semi-functional partial linear regression, *Statist. Probab. Lett.* 6 (2006) 1102–1110.

[5] G. Aneiros and P. Vieu, Partial linear modelling with multi-functional covariates, *Comput. Statist* 30 (2015) 647–671.

[6] R. Bojanic, E. Seneta, A unified theory of regularly varying sequences, *Math. Z.* 134 (1973) 91–106.

[7] H. Cardot, P. Sarda, Functional linear regression. *The Oxford handbook of functional data analysis*, (2011) 21–46, Oxford Univ. Press, Oxford.

[8] T.T. Cai, P. Hall. Prediction in functional linear regression, *Ann. Statist.* 34 (2006) 2159–2179.

[9] C. Chatfield, The joys of consulting, *Significance* 4 (2007) 33–36.

[10] K. Chen, S. Guo, Y. Lin, Z. Ying, Least absolute relative error estimation, *J. Amer. Statist. Assoc.* 105 (2010) 1104–1112.

[11] D. Chen, P. Hall, H.G. Müller, Single and multiple index functional regression models with nonparametric link, *Ann. Statist.* 38 (2011) 3458–3486.

[12] H. Cho, Y-J. Kim, H-J. Jung, S-W. Lee, W. Lee, OutlierD: an R package for outlier detection using quantile regression on mass spectrometry data. *Bioinform.* 24 (2008) 882–884.

[13] S-H. Eo, H. Cho, OutlierDM: Outlier Detection for Multi-replicated High-throughput Data. `https://CRAN.R-project.org/package=OutlierDM`, (2014) R package version 1.1.1.

[14] A. Cuevas, M. Febrero, R. Fraiman, On the use of the bootstrap for estimating functions with functional data. *Comput. Statist. Data Anal.* 51 (2006) 1063–1074.

[15] A. Cuevas, A partial overview of the theory of statistics with functional data. *J. Statist. Plann. Inference*, 147 (2014) 1–23.

[16] A. Delaigle, I. Gijbels, Practical bandwidth selection in deconvolution kernel density estimation. *Comput. Statist. Data Anal.* 45 (2004) 249–267.

[17] J. Demongeot, A. Hamie, A. Laksaci, M. Rachdi, Relative-error prediction in nonparametric functional statistics: Theory and practice, *J. Multivariate Anal.* 146 (2016) 261–268.

[18] M. Febrero-Bande, M. Oviedo de la Fuente, Statistical computing in Functional Data Analysis: The R Package fda.usc. *J. Stat. Software*, 51 (2012) 1–28.

[19] M. Febrero-Bande, M. Oviedo de la Fuente, P. Galeano, A. Nieto, E. Garcia-Portugue, fda.usc: Functional Data Analysis and Utilities for Statistical Computing. `https://CRAN.R-project.org/package=fda.usc`, (2019) R package version 1.5.0.

[20] F. Ferraty, `http://www.lsp.ups-tlse.fr/staph/npfda/npfda-spectrometric.dat` (2006).

[21] F. Ferraty, A. Mas, P. Vieu, Nonparametric regression on functional data: Inference and practical aspects, *Aust. N. Z. J. Stat.* 49 (2007) 267–286.

[22] F. Ferraty, I. Van Keilegom, P. Vieu, On the validity of the bootstrap in non-parametric functional regression, *Scand. J. Statist.* 37 (2010) 286–306.

[23] F. Ferraty, P. Vieu, Non parametric Functional Data Analysis, Theory and Practice, *Springer-Verlag*, New York, 2006.

[24] J. Galambos, E. Seneta, Regularly varying sequences, *Proc. Amer. Math. Soc.* 41 (1973) 110–116.

[25] A. Goia, P. Vieu, A partitioned single functional index model, *Comput. Statist* 30 (2015) 676–692.

[26] P. Hall, J.L. Horowitz, Methodology and convergence rates for functional linear regression, *Ann. Statist.* 35 (2007) 70–91.

[27] W. Härdle, J.S. Marron, Bootstrap simultaneous error bars for nonparametric regression, *Ann. Statist.* 16 (1991) 1696–1708.

[28] L. Horvàth, P. Kokoszka, Inference for Functional Data With Applications. *Springer*, New York, 2012.

[29] T. Hsing, R. Eubank, Theoretical Foundations of Functional Data Analysis, With an Introduction to Linear Operators. *Wiley*, Chichester, 2015.

[30] A. Jmaei, Y. Slaoui, W. Dellagi, Recursive distribution estimators defined by stochastic approximation method using Bernstein polynomials. *J. Nonparametr. Stat.* 29 (2017) 792–805.

[31] M.C. Jones, H. Park, K-I. Shin, S.K. Vines, S-O. Jeong, Relative error prediction via kernel regression smoothers. *J. Statist. Plann. Inference* 138 (2008) 2887–2898.
[32] L.Z. Kara, A. Laksaci, M. Rachdi, P. Vieu, Data-driven *k*NN estimation in nonparametric functional data analysis. *J. Multivariate Anal.* 153 (2017) 176–188.
[33] T.M. Khoshgoftaar, B.B. Bhattacharyya, G.D. Richardson, Predicting software errors, during development, using nonlinear regression models: a comparative study, *IEEE Trans. Reliab.* 41 (1992) 390–395.
[34] R. Koenker, B. Bassett, Regression quantiles, *Econometrics* 46 (1978) 33–50.
[35] H. Lian, Functional partial linear model. *J. Nonparametr. Stat.* 23 (2011) 115–128.
[36] N. Ling, P. Vieu, Nonparametric modelling for functional data: selected survey and tracks for future. Statistics 52 (2018) 934–949.
[37] A. Mokkadem, M. Pelletier, A companion for the Kiefer-Wolfowitz-Blum stochastic approximation algorithm, *Ann. Statist.* 35 (2007) 1749–1772.
[38] A. Mokkadem, M. Pelletier, Y. Slaoui, Revisiting Révész's stochastic approximation method for the estimation of a regression function, *ALEA Lat. Am. J. Probab. Math. Stat.* 6 (2009) 63–114.
[39] S.C. Narula, J.F. Wellington, Prediction, linear regression and the minimum sum of relative errors, *Technometrics* 19 (1977) 185–190.
[40] H. Park, L.A. Stefanski, Relative-error prediction. *Statist. Probab. Lett.* 40 (1998) 227–236.
[41] D.N. Politis, J.P. Romano, Limit theorems for weakly dependent Hilbert space valued random variables with application to the stationarity bootstrap, *Statist. Sinica* 4 (1994) 461–476.
[42] P, Raña, G. Aneiros, J. Vilar, P. Vieu, Bootstrap confidence intervals in functional nonparametric regression under dependence, *Electron. J. Stat.* 10 (2016) 1973–1999.
[43] J.O. Ramsay, B. W. Silverman, Applied Functional Data Analysis: Methods and Case Studies, *Springer*, New York, 2002.
[44] R.J. Serfling, Approximation Theorems of Mathematical Statistics, *Wiley*, New York, 1980.
[45] H.L. Shang, Bootstrap methods for stationary functional time series, *Statist. Comput.* 28 (2018) 1–10.
[46] Y. Slaoui, Optimal bandwidth selection for semi-recursive kernel regression estimators. *Stat. Interface* 9 (2016) 375–388.
[47] Y. Slaoui, Recursive nonparametric regression estimation for independent functional data, *Statist. Sinica* (2018) `http://dx.doi.org/10.5705/ss.202018.0069`.
[48] Y. Slaoui, A. Jmaei, Recursive density estimators based on Robbins-Monro's scheme and using Bernstein polynomials, *Stat. Interface* (2019) To appear.
[49] Y. Yang, F. Ye, General relative error criterion and M-estimation, *Front. Math. China* 8 (2013) 695–715.