Statistica Sinica **30** (2020), 417-437
doi:https://doi.org/10.5705/ss.202018.0069

RECURSIVE NONPARAMETRIC REGRESSION ESTIMATION FOR INDEPENDENT FUNCTIONAL DATA

Yousri Slaoui

Université de Poitiers

Abstract: We propose an automatic selection of the bandwidth of the recursive nonparametric estimation of the regression function defined by the stochastic approximation algorithm. Here the explanatory data are curves and the response is real. We compare our recursive estimators with the nonrecursive estimator proposed by Ferraty and Vieu (2002). The two methods are based on the wild bootstrapping approach, where resampling is done from a suitably estimated residual distribution. Moreover, we establish a central limit theorem for our proposed recursive estimators. We use the wild bootstrap to select the bandwidth and some special stepsizes. As such, the proposed recursive estimators are competitive in terms of the estimation error, but much better in terms of computational costs. The proposed estimators are used in simulated and real functional data sets.

Key words and phrases: Asymptotic normality, curve fitting, functional data, regression estimation, smoothing, stochastic approximation algorithm, wild functional bootstrap.

1. Introduction

The progress in computing technology, in terms of both memory capacity and computing speed, has made it possible to record vast amounts of data. Thus, many variables can be observed when studying the same phenomenon. This is especially true when we have a family of variables $\{X(\theta)\}_{\theta\in\Theta}$ indexed by a parameter θ, varying in a space Θ (e.g., $\mathbb{R}$, $\mathbb{R}^p$, or $\Theta \subset H$, where H is a Hilbertian space). Obviously, it is technically impossible to measure $X(\theta)$ for each $\theta \in \Theta$. Nevertheless, it is possible to consider a smooth discretization $\{\theta_i\}_{i=1,\dots,I}$ of Θ in order to consider that the behavior of $\{X(\theta)\}_{\theta\in\Theta}$ is close to that of $\{X(\theta_i)\}_{i=1,\dots,I}$. Such a family $\{X(\theta)\}_{\theta\in\Theta}$ of random variables is called a functional random variable (see Ferraty and Vieu (2002, 2004)).

There has been growing increasing interest in functional data analysis. For an introduction to this field, refer to Ramsy and Silverman (2002), who provide a detailed exposition of both the theoretical and practical aspects of functional data analyses. The existing literature contains numerous studies on functional linear models (e.g., see among many others, Cardot, Ferraty and Sarda (1999), Cai and Hall (2006), Hall and Horowitz (2007)). In the framework of nonparametric estimations with functional predictors and scalar responses (e.g., see among others Ferraty and Vieu (2002, 2004, 2006), Preda (2007), and Biau, Gerou and Guyader (2010)).

Here, we are interested in the problem of recursively estimating a regression when the explanatory data are curves and the response is real. This problem can be formulated by considering that $\{Y_i, \mathcal{X}_i\}_{i=1}^n$ is 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 $\|.\|$. Assume that $\mathbb{E}\left|Y_i\right| < \infty$ and define the regression functional as

$$r\left(u\right) := \mathbb{E}\left[Y_i|\mathcal{X}_i = u\right]; \quad u \in \mathcal{E}, \quad \forall i \in \mathbb{N}. \tag{1.1}$$

Model (1.1) can be written as follows:

$$Y_i = r\left(\mathcal{X}_i\right) + \varepsilon_i, \quad i =, \ldots, n,$$

where ε_i is a random variable, such that $\mathbb{E}\left[\varepsilon_i|\mathcal{X}_i\right] = 0$ and $\mathbb{E}\left[\varepsilon_i^2|\mathcal{X}_i\right] = \sigma_\varepsilon^2\left(\mathcal{X}_i\right) < \infty$.

The purpose of this study is to adapt the recursive estimator proposed in Slaoui (2016) to the case when the covariate is functional and the response is real. Thus, our proposed recursive estimator for the operator r is the following :

$$\widehat{r}_n\left(\chi, h\right) = \frac{\widehat{a}_n\left(\chi, h\right)}{\widehat{f}_n\left(\chi, h\right)},$$

with

$$\widehat{a}_n\left(\chi, h\right) = \left(1 - \gamma_n\right)\widehat{a}_{n-1}\left(\chi, h\right) + \gamma_n h_n^{-1} K\left(\frac{\|\chi - \mathcal{X}_n\|}{h_n}\right) Y_n, \tag{1.2}$$

$$\widehat{f}_n\left(\chi, h\right) = \left(1 - \gamma_n\right)\widehat{f}_{n-1}\left(\chi, h\right) + \gamma_n h_n^{-1} K\left(\frac{\|\chi - \mathcal{X}_n\|}{h_n}\right), \tag{1.3}$$

where (γ_n) and (h_n) are sequences of positive real numbers that go to zero, and K is a kernel. Throughout this paper, we suppose that $\widehat{a}_0\left(\chi, h\right) = \widehat{f}_0\left(\chi, h\right) = 0$, and let $\Pi_n = \prod_{i=1}^n \left(1 - \gamma_i\right)$. Then, we can estimate the operator r by:

$$\widehat{r}_n\left(\chi, h\right) = \frac{\Pi_n \sum_{k=1}^n \Pi_k^{-1}\gamma_k h_k^{-1} K\left(\|\chi - \mathcal{X}_k\|/h_k\right) Y_k}{\Pi_n \sum_{k=1}^n \Pi_k^{-1}\gamma_k h_k^{-1} K\left(\|\chi - \mathcal{X}_k\|/h_k\right)}. \tag{1.4}$$

Several assumptions are made on the kernel K, bandwidth (h_n), and stepsize (γ_n). The recursive property (1.4) is particularly useful for large samples, because $\widehat{r}_n$ can be updated easily using each additional observation.

The first results for the recursive kernel estimator of the operator r when the response variable is real and the covariable is functional are obtained by Amiri, Crambes and Thiam (2014). They proposed the following estimators

$$r_n^{[l]}(\chi, h) = \frac{\sum_{i=1}^{n} Y_i/(F(h_i)^l)K(\|\chi - \mathcal{X}_i\|/h_i)}{\sum_{i=1}^{n} 1/(F(h_i)^l)K(\|\chi - \mathcal{X}_i\|/h_i)}, \quad l \in [0,1],$$

where F is the cumulative distribution function of the random variable $\|\mathcal{X} - \chi\|$. We can check easily that these estimators are a special case of our proposed recursive estimators (1.4), with a stepsize $(\gamma_n) = (h_n F(h_n)^{-l}[\sum_{k=1}^{n} h_k F(h_k)^{-l}]^{-1})$.

Furthermore, we show the special case when $\mathcal{X}$ is a geometric process (or fractal), with $F(h_n) \sim h_n^{\kappa}$, and $\kappa > 0$. Here, the optimal bandwidth that minimizes $\mathbb{E}[\widehat{r}_n(\chi, h) - r(\chi)]^2$ depends on the choice of the stepsize (γ_n). In particular, we show that under some regularity conditions of the functional regression r, and using the stepsizes $(\gamma_n) = (\gamma_0 n^{-1})$, where $\gamma_0 > 0$, the bandwidth (h_n) must equal

$$\left(\left\{\frac{\kappa}{2}\frac{\sigma_{\varepsilon}^2(\chi)}{(\phi'(0))^2}\frac{(\gamma_0 + \kappa/(\kappa+a))^2}{2\gamma_0}\frac{M_2}{M_0^2}\right\}^{1/(\kappa+2)} n^{-1/(\kappa+2)}\right).$$

The first purpose of this study is to propose an automatic selection for the bandwidth using the plug-in method and then through the wild bootstrap method. Second, we compare the proposed recursive estimators $\widehat{r}_n$ to the nonrecursive functional regression estimator introduced by Ferraty and Vieu (2002) for independent data. They constructed the functional estimate of the operator r using the standard kernel methods (Nadaraya (1964) and Watson (1964)), defined as

$$\widetilde{r}_n(\chi, h) = \frac{\sum_{i=1}^{n} Y_i K(\|\chi - \mathcal{X}_i\|/h_n)}{\sum_{i=1}^{n} K(\|\chi - \mathcal{X}_i\|/h_n)}. \tag{1.5}$$

This estimator was considered by Ferraty and Vieu (2004, 2006), whereas Masry (2005) considered the asymptotic normality of (1.5) in the dependent case. Benhenni, Hedli-Griche and Rachdi (2010) considered the case of a fixed-design with correlated errors. Lian (2012) examined the case when the predictors and responses are both functions. The remainder of the paper is organized as follows. In Section 2, we state our main results. Section 3 is devoted to our application results, first by simulation (subsection 3.1) and second using a real data set (subsection 3.2). We conclude the article in Section 4. Appendix A gives the proofs of our theoretical results.

2. Assumptions and Main Results

Let F be the cumulative distribution function of the random variable $\|\mathcal{X} - \chi\|$:

$$F(t) = \mathbb{P}\left(\|\mathcal{X} - \chi\| \leq t\right).$$

We first define the following class of regularly varying sequences.

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

$$\lim_{n\to+\infty} n\left[1 - \frac{v_{n-1}}{v_n}\right] = \gamma. \tag{2.1}$$

Condition (2.1) was introduced by Galambos and Seneta (1973) to define regularly varying sequences (see also Bojanic and Seneta (1973)) and by Mokkadem and Pelletier (2007) 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 estimators (1.4). Here, we refer to the following assumptions:

(A1) The function $\phi(u) := \mathbb{E}\left[\{r(\mathcal{X}) - r(\chi)\} \mid \|\mathcal{X} - \chi\| = u\right]$ is assumed to be derivable at $t = 0$.

(A2) $K : \mathbb{R} \to \mathbb{R}$ is a continuous, bounded function with support on the compact $[0,1]$, such that $\inf_{t\in[0,1]} K(t) > 0$.

(A3) For any $s \in [0,1]$, $\tau_h(s) := (F(hs))/(F(h)) \to \tau_0(s) < \infty$ as $h \to 0$.

(A4) *i)* $(\gamma_n) \in \mathcal{GS}(-\alpha)$ with $\alpha \in \left]1/2, 1\right]$.
ii) $(h_n) \in \mathcal{GS}(-a)$ with $a \in \left]0, 1\right[$.
iii) $(F(h_n)) \in \mathcal{GS}(-\mathcal{F}_a)$ with $\mathcal{F}_a \in \left]0, \alpha\right[$.
iv) $\lim_{n\to\infty} (n\gamma_n) \in \left]\min\{\mathcal{F}_a, (\alpha + \mathcal{F}_a)/2 - a\}, \infty\right]$.
v) $(g_n) \in \mathcal{GS}(-g)$ with $g \in \left]0, a\right[$.
vi) $(F(g_n)) \in \mathcal{GS}(-\mathcal{F}_g)$ with $\mathcal{F}_g \in \left]0, \mathcal{F}_a\right[$.

Assumption $(A4)\,(iv)$ on the limit of $(n\gamma_n)$ as n goes to infinity 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. For simplicity, we introduce the following notation:

$$\xi = \lim_{n\to\infty} (n\gamma_n)^{-1}, \tag{2.2}$$

$$M_0 = K(1) - \int_0^1 (tK(t))' \, \tau_0(t)\, dt,$$

$$M_1 = K(1) - \int_0^1 K'(t)\, \tau_0(t)\, dt,$$

$$M_2 = K^2(1) - \int_0^1 \left(K^2(t)\right)' \tau_0(t)\, dt.$$

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

Proposition 1 (Bias and variance of $\widehat{r}_n$)**.** *Let Assumptions* (A1)-(A4) *hold.*

1. *If* $a \in]0, (\alpha - \mathcal{F}_a)/2]$, *then*
$$\mathbb{E}\left[\widehat{r}_n(\chi, h)\right] - r(\chi) = h_n \phi'(0) \frac{1 - (\mathcal{F}_a - a)\xi}{1 - \mathcal{F}_a \xi} \frac{M_0}{M_1} \left[1 + o(1)\right]. \tag{2.3}$$
If $a \in](\alpha - \mathcal{F}_a)/2, 1[$, *then*
$$\mathbb{E}\left[\widehat{r}_n(\chi, h)\right] - r(\chi) = o\left(\sqrt{\gamma_n F(h_n)^{-1}}\right).$$

2. *If* $a \in [(\alpha - \mathcal{F}_a)/2, 1[$, *then*
$$\begin{aligned} &Var\left[\widehat{r}_n(\chi, h)\right] \\ &= \sigma_\varepsilon^2(\chi) \frac{(1 - (\mathcal{F}_a - a)\xi)^2}{(2 - (\mathcal{F}_a + \alpha - 2a)\xi)} \frac{M_2}{M_1^2} \frac{\gamma_n}{F(h_n)} \left[1 + o(1)\right]. \end{aligned} \tag{2.4}$$
If $a \in]0, (\alpha - \mathcal{F}_a)/2[$, *then*
$$Var\left[\widehat{r}_n(\chi, h)\right] = o\left(h_n^2\right). \tag{2.5}$$

3. *If* $\lim_{n\to\infty}(n\gamma_n) > \max\{\mathcal{F}_a, (\mathcal{F}_a + \alpha)/2 - a\}$, *then* (2.3) *and* (2.4) *hold simultaneously.*

The bias and the variance of the estimator $\widehat{r}_n$ defined by the stochastic approximation algorithm (1.4) then heavily depend on the choice of the stepsize (γ_n).

Let us first state the following theorem, which gives the weak convergence rate of the estimator $\widehat{r}_n$ defined in (1.4).

Theorem 1 (Weak pointwise convergence rate)**.** *Let Assumptions* (A1)-(A4) *hold.*

1. *If there exists* $c \geq 0$ *such that* $\gamma_n^{-1} h_n^2 F(h_n) \to c$, *then*
$$\begin{aligned} &\sqrt{\gamma_n^{-1} F(h_n)} \left(\widehat{r}_n(\chi, h) - r(\chi)\right) \\ &\stackrel{\mathcal{D}}{\rightarrow} \mathcal{N}\left(c^{1/2}\phi'(0) \frac{1 - (\mathcal{F}_a - a)\xi}{1 - \mathcal{F}_a \xi} \frac{M_0}{M_1}, \sigma_\varepsilon^2(\chi) \frac{(1 - (\mathcal{F}_a - a)\xi)^2}{(2 - (\mathcal{F}_a + \alpha - 2a)\xi)} \frac{M_2}{M_1^2}\right). \end{aligned}$$

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

$$\frac{1}{h_n}\left(\widehat{r}_n(\chi,h)-r(\chi)\right)\stackrel{\mathbb{P}}{\to}\phi'(0)\,\frac{1-(\mathcal{F}_a-a)\,\xi}{1-\mathcal{F}_a\xi}\,\frac{M_0}{M_1},$$

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

Let us now consider the case when the bandwidth (h_n) is chosen such that $\lim_{n\to\infty}\gamma_n^{-1}h_n^2F(h_n)=0$ (which corresponds to undersmoothing). Thus, the proposed estimator fulfills the following central limit theorem:

$$\sqrt{\gamma_n^{-1}F(h_n)}\left(\widehat{r}_n(\chi,h)-r(\chi)\right)\stackrel{\mathcal{D}}{\to}\mathcal{N}\left(0,\sigma_\varepsilon^2(\chi)\,\frac{\left(1-(\mathcal{F}_a-a)\,\xi\right)^2}{\left(2-(\mathcal{F}_a+\alpha-2a)\,\xi\right)}\,\frac{M_2}{M_1^2}\right).$$

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

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

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

$$\widehat{M}_i=\Pi_n\sum_{k=1}^n\Pi_k^{-1}\gamma_k\widehat{F}(h_k)^{-1}K^i\left(\frac{\|\chi-\mathcal{X}_k\|}{h_k}\right),\quad i\in\{1,2\}\tag{2.6}$$

$$\widehat{\sigma}_\varepsilon^2(\chi)=\frac{\Pi_n\sum_{k=1}^n\Pi_k^{-1}\gamma_kK\left(\|\chi-\mathcal{X}_k\|/h_k\right)Y_k^2}{\Pi_n\sum_{k=1}^n\Pi_k^{-1}\gamma_kK\left(\|\chi-\mathcal{X}_k\|/h_k\right)}-\left(\widehat{r}_n(\chi,h)\right)^2.\tag{2.7}$$

In order to measure the quality of our recursive estimator (1.4), we use the following quantity:

$$MSE\left[\widehat{r}_n(\chi,h)\right]=\left(\mathbb{E}\left(\widehat{r}_n(\chi,h)\right)-r(\chi)\right)^2+Var\left(\widehat{r}_n(\chi,h)\right).$$

The next proposition gives the MSE of the recursive estimators defined in (1.4).

Proposition 2 (MSE of $\widehat{r}_n(\chi,h)$)**.** *Let Assumptions* (A1)-(A4) *hold.*

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

$$MSE\left[\widehat{r}_n(\chi,h)\right]=h_n^2\left(\phi'(0)\right)^2\left(\frac{1-(\mathcal{F}_a-a)\,\xi}{1-\mathcal{F}_a\xi}\right)^2\frac{M_0^2}{M_1^2}+o\left(h_n^2\right).$$

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

$$MSE\left[\widehat{r}_n(\chi,h)\right]=h_n^2\left(\phi'(0)\right)^2\left(\frac{1-(\mathcal{F}_a-a)\,\xi}{1-\mathcal{F}_a\xi}\right)^2\frac{M_0^2}{M_1^2}$$

$$+\sigma_{\varepsilon}^{2}(\chi)\frac{(1-(\mathcal{F}_a-a)\,\xi)^2}{(2-(\mathcal{F}_a+\alpha-2a)\,\xi)}\frac{M_2}{M_1^2}\frac{\gamma_n}{F(h_n)}$$
$$+o\left(h_n^2+\frac{\gamma_n}{F(h_n)}\right).$$

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

$$MSE\left[\widehat{r}_n(\chi,h)\right]=\sigma_{\varepsilon}^{2}(\chi)\frac{(1-(\mathcal{F}_a-a)\,\xi)^2}{(2-(\mathcal{F}_a+\alpha-2a)\,\xi)}\frac{M_2}{M_1^2}\frac{\gamma_n}{F(h_n)}+o\left(\frac{\gamma_n}{F(h_n)}\right).$$

2.1. Stepsize selection

In the framework of nonparametric kernel estimators, to determine the optimal choice of stepsize, Mokkadem, Pelletier and Slaoui (2009a) consider $(\gamma_n)\in\mathcal{GS}(-1)$ in order to ensure an optimal convergence rate. Furthermore, they consider two points of view: a pointwise estimation and estimation by confidence intervals. For the pointwise estimation, the criteria they use to find the optimal stepsize to minimize the mean squared error (MSE) or the integrated mean squared error ($MISE$). In our context, we display a set of stepsizes (γ_n) that minimize the MSE or the MISE of the estimator $\widehat{r}_n(\chi,h)$ defined by (1.4); in particular, we show that the sequence $(\gamma_n)=\left(n^{-1}\right)$ belongs to this set. Note that these minimum MSE and $MISE$ are larger than those obtained for the nonrecursive estimator $\widetilde{r}(\chi,h)$, defined by (1.5). Thus, for a pointwise estimation and when rapid updating is not as important, it is preferable to use the nonrecursive estimator rather than any recursive estimator $\widehat{r}_n(\chi,h)$, defined by (1.4).

Moreover, for the confidence interval point of view, they find the optimal stepsize by minimizing the variance. In our context, we display a set of stepsizes (γ_n) that minimize the variance of $\widehat{r}_n(\chi,h)$; it follows from (2.4) that the sequence $(\gamma_n)=\left([1-a]\,n^{-1}\right)$ belongs to this set. Let us underline that the variance of the estimator $\widehat{r}_n(\chi,h)$ defined with this stepsize is smaller than that of the nonrecursive estimator $\widetilde{r}_n(\chi,h)$, defined by (1.5). Consequently, even when the online aspect is not as important, it is preferable to use recursive estimators when constructing confidence intervals.

Remark 1. Under assumptions (A1)-(A4), and $(\gamma_n)=\left([1-a]\,n^{-1}\right)$, the variance of $\widehat{r}_n(\chi,h)$ is equal to:

$$Var\left[\widehat{r}_n(\chi,h)\right]=(1-\mathcal{F}_a)\,\sigma_{\varepsilon}^{2}(\chi)\frac{M_2}{M_1^2}\frac{1}{nF(h_n)}\left[1+o(1)\right],$$

whereas the variance of $\widetilde{r}_n(\chi,h)$ is equal to:

$$Var\left[\widetilde{r}_n(\chi,h)\right]=\sigma_\varepsilon^2(\chi)\frac{M_2}{M_1^2}\frac{1}{nF(h_n)}\left[1+o(1)\right].$$

Furthermore, Mokkadem, Pelletier and Slaoui (2009b) consider (γ_n), such that $n\gamma_n\to\infty$, and then use the averaging principle of the stochastic approximation algorithm to ensure an optimal convergence rate. Throughout this paper, we consider $(\gamma_n)\in\mathcal{GS}(-1)$ and the pointwise estimation point of view.

2.2. Bandwidth selection

In the framework of nonparametric kernel estimators, the bandwidth selection methods studied in the literature can be divided into three broad classes: cross-validation techniques, plug-in methods, and the bootstrap idea. A detailed comparison of the three techniques can be found in Delaigle and Gijbels (2004). They concluded that, chosen appropriately, the plug-in and bootstrap selectors both outperform the cross-validation bandwidth, and neither of the two is best in all cases. Therefor, we consider a plug-in method in the special case when $\mathcal{X}$ is geometric process (or fractal) with $F(h_n)\sim h_n^\kappa$, with $\kappa>0$. Then, for a more general context, we used a wild bootstrap to approximate the distribution of the error of the recursive kernel regression estimators (1.4).

2.2.1. Plug-in method

In this subsection, we consider the special case when $\mathcal{X}$ is a geometric process (or fractal) with $F(h_n)\sim h_n^\kappa$, with $\kappa>0$.

Recursive estimators. The following corollary indicates that the bandwidth that minimizes the MSE of $\widehat{r}_n$ depends on the stepsize (γ_n) and then the corresponding MSE depends also on the stepsize (γ_n).

Corollary 1. *Let Assumptions* (A1)-(A4) *hold. To minimize the MSE of $\widehat{r}_n$, the stepsize (γ_n) must be chosen in $\mathcal{GS}(-1)$, the bandwidth (h_n) must equal*

$$\left(\left\{\frac{\kappa}{2}\frac{\sigma_\varepsilon^2(\chi)}{(\phi'(0))^2}\frac{M_2}{M_0^2}\frac{(1-\mathcal{F}_a\xi)^2}{(2-(\mathcal{F}_a+\alpha-2a)\xi)}\right\}^{1/(\kappa+2)}\gamma_n^{1/(\kappa+2)}\right).$$

Then, we have

$$\begin{aligned}&AMSE\left[\widehat{r}_n(\chi,h)\right]\\&=3\times(\kappa/2)^{2/(\kappa+2)}\frac{(1-(\mathcal{F}_a-a)\xi)^2}{(2-(\mathcal{F}_a+\alpha-2a)\xi)^{2/(\kappa+2)}(1-\mathcal{F}_a\xi)^{2\kappa/(\kappa+2)}}\\&\quad\left(\sigma_\varepsilon^2(\chi)\right)^{2/(\kappa+2)}\left(\phi'(0)M_0\right)^{2\kappa/(\kappa+2)}M_2^{2/(\kappa+2)}M_1^{-2}\gamma_n^{2/(\kappa+2)}.\end{aligned}$$

The following corollary shows that, for a special choice of the stepsize $(\gamma_n) = \left(\gamma_0 n^{-1}\right)$, where $\lim_{n\to\infty} n\gamma_n = \gamma_0$ and $(\gamma_n) \in \mathcal{GS}(-1)$, the optimal value of h_n depends on γ_0. Then, the corresponding $AMSE$ depends on γ_0.

Corollary 2. *Let Assumptions* (A1)-(A4) *hold. To minimize the AMSE of* $\widehat{r}_n(\chi, h)$*, the stepsize* (γ_n) *must be chosen in* $\mathcal{GS}(-1)$*, and the bandwidth* (h_n) *must equal*

$$\left(\left\{\frac{\kappa}{2}\frac{\sigma_\varepsilon^2(\chi)}{(\phi'(0))^2}\frac{M_2}{M_0^2}\frac{(\gamma_0+\kappa/(\kappa+a))^2}{2\gamma_0}\right\}^{1/(\kappa+2)} n^{-1/(\kappa+2)}\right).$$

Then, we have

$$\begin{aligned}
&AMSE\left[\widehat{r}_n(\chi,h)\right]\\
&= 3\times(\kappa/4)^{2/(\kappa+2)}\left(\sigma_\varepsilon^2(\chi)\right)^{2/(\kappa+2)}\gamma_0^{-2/(\kappa+2)}\frac{(\gamma_0+(\kappa+1)/(\kappa+2))^2}{(\gamma_0+\kappa/(\kappa+2))^{2\kappa/(\kappa+2)}}\\
&\quad\left(\phi'(0)M_0\right)^{2\kappa/(\kappa+2)}M_2^{2/(\kappa+2)}M_1^{-2}n^{-2/(\kappa+2)}\\
&= \mathtt{Coeff}(\kappa,\gamma_0)\left(\sigma_\varepsilon^2(\chi)\right)^{2/(\kappa+2)}\left(\phi'(0)M_0\right)^{2\kappa/(\kappa+2)}M_2^{2/(\kappa+2)}M_1^{-2}n^{-2/(\kappa+2)},
\end{aligned}$$

where $\mathtt{Coeff}(\kappa,\gamma_0) = 3\times(\kappa/4)^{2/(\kappa+2)}\gamma_0^{-2/(\kappa+2)}((\gamma_0+(\kappa+1)/(\kappa+2))^2)/((\gamma_0+\kappa/(\kappa+2))^{2\kappa/(\kappa+2)})$.

Under assumptions (A1)-(A4), the plug-in bandwidth (h_n) must equal

$$\left(\left\{\frac{\kappa}{2}\frac{(\gamma_0+\kappa/(\kappa+a))^2}{2\gamma_0}\right\}^{1/(\kappa+2)}\widehat{\sigma}_\varepsilon^2(\chi)\frac{\widehat{M}_2}{\widehat{I}_0^2}n^{-1/(\kappa+2)}\right). \tag{2.8}$$

Then, the corresponding plug-in $AMSE$ is equal to

$$\begin{aligned}
AMSE\left[\widehat{r}_n(\chi,h)\right] = \mathtt{Coeff}(\kappa,\gamma_0)&\left(\widehat{\sigma}_\varepsilon^2(\chi)\right)^{2/(\kappa+2)}\\
&\widehat{M}_1^{-2}\widehat{M}_2^{2/(\kappa+2)}\widehat{I}_0^{2\kappa/(\kappa+2)}n^{-2/(\kappa+2)},
\end{aligned}\tag{2.9}$$

where $\widehat{I}_0$, $\widehat{M}_1$, $\widehat{M}_2$, and $\widehat{\sigma}_\varepsilon^2(\chi)$ are asymptotically unbiased estimators of $\phi'(0)M_0$, M_1, M_2, and $\sigma_\varepsilon^2(\chi)$, respectively.

$$\widehat{I}_0 = \Pi_n\sum_{k=1}^n \Pi_k^{-1}\gamma_k\widehat{F}(h_k)^{-1}(Y_k-\widehat{r}_n(\chi,h))K\left(\frac{\|\chi-\mathcal{X}_k\|}{h_k}\right),$$

where, $\widehat{M}_i$, for $i\in\{1,2\}$, and $\widehat{\sigma}_\varepsilon^2(\chi)$ are given in (2.6) and (2.7).

Nonrecursive estimator. First, recall that under assumptions (A1)-(A3) and (A4)ii), the bias and variance of $\widetilde{r}_n(\chi,h)$ are given by

$$\mathbb{E}\left[\widetilde{r}_n\left(\chi,h\right)\right]-r\left(\chi\right)=h_n\phi'\left(0\right)\frac{M_0}{M_1}\left[1+o\left(1\right)\right]$$

and

$$Var\left[\widetilde{r}_n\left(\chi,h\right)\right]=\sigma_\varepsilon^2\left(\chi\right)\frac{M_2}{M_1^2}\frac{1}{nF\left(h_n\right)}\left[1+o\left(1\right)\right].$$

It follows that

$$MSE\left[\widetilde{r}_n\left(\chi,h\right)\right]=h_n^2\left(\phi'\left(0\right)\right)^2\frac{M_0^2}{M_1^2}+\sigma_\varepsilon^2\left(\chi\right)\frac{M_2}{M_1^2}\frac{1}{nF\left(h_n\right)}+o\left(h_n^2+\frac{1}{nF\left(h_n\right)}\right).$$

Then, to minimize the MSE of $\widetilde{r}_n\left(\chi,h\right)$, the bandwidth $\left(h_n\right)$ must equal

$$\left(\left\{\frac{\kappa}{2}\frac{\sigma_\varepsilon^2\left(\chi\right)}{\left(\phi'\left(0\right)\right)^2}\frac{M_2}{M_0^2}\right\}^{1/(\kappa+2)}n^{-1/(\kappa+2)}\right).$$

Then, we have

$$\begin{aligned}AMSE\left[\widetilde{r}_n\left(\chi,h\right)\right]=\texttt{Coeff}\left(\kappa\right)&\left(\sigma_\varepsilon^2\left(\chi\right)\right)^{2/(\kappa+2)}\left(\phi'\left(0\right)\right)^{2\kappa/(\kappa+2)}\\&M_2^{2/(\kappa+2)}M_0^{2\kappa/(\kappa+2)}M_1^{-2}n^{-2/(\kappa+2)},\end{aligned}$$

where $\texttt{Coeff}\left(\kappa\right)=\left((\kappa+2)/\kappa\right)\left(\kappa/2\right)^{2/(\kappa+2)}$. The plug-in bandwidth $\left(h_n\right)$ must equal

$$\left(\left\{\frac{\kappa}{2}\widetilde{M}_2\frac{\widetilde{\sigma}_\varepsilon^2\left(\chi\right)}{\widetilde{I}_0^2}\right\}^{1/(\kappa+2)}n^{-1/(\kappa+2)}\right).$$

Then, the corresponding plug-in $AMSE$ is equal to

$$AMSE\left[\widetilde{r}_n\left(\chi,h\right)\right]=\texttt{Coeff}\left(\kappa\right)\left(\sigma_\varepsilon^2\left(\chi\right)\right)^{2/(\kappa+2)}\widetilde{I}_0^{2\kappa/(\kappa+2)}\widetilde{M}_1^{-2}\widetilde{M}_2^{2/(\kappa+2)}n^{-2/(\kappa+2)},$$

where $\widetilde{I}_0$, $\widetilde{M}_1$, $\widetilde{M}_2$, and $\widetilde{\sigma}_\varepsilon^2\left(\chi\right)$ are asymptotically unbiased estimators of $\phi'\left(0\right)M_0$, M_1, M_2, and $\sigma_\varepsilon^2\left(\chi\right)$, respectively.

$$\widetilde{I}_0=\frac{1}{n\widehat{F}\left(h_n\right)}\sum_{k=1}^n\left(Y_k-\widetilde{r}_n\left(\chi,h\right)\right)K\left(\frac{\|\chi-\mathcal{X}_k\|}{h_n}\right),$$

$$\widetilde{M}_i=\frac{1}{n\widehat{F}\left(h_n\right)}\sum_{k=1}^nK^i\left(\frac{\|\chi-\mathcal{X}_k\|}{h_n}\right),\quad i\in\{1,2\},$$

$$\widetilde{\sigma}_\varepsilon^2\left(\chi\right)=\frac{\sum_{k=1}^nK\left(\|\chi-\mathcal{X}_k\|/h_n\right)Y_k^2}{\sum_{k=1}^nK\left(\|\chi-\mathcal{X}_k\|/h_n\right)}-\left(\widetilde{r}_n\left(\chi,h\right)\right)^2.$$

We observe from Table 1 and Figure 1 that, for κ bigger than 0.7, the $AMSE$ of the nonrecursive estimators is smaller than that of the recursive estimator.

2.2.2. Wild bootstrap method

The main idea of the wild bootstrap proposed in Härdle and Marron (1991),

Table 1. Numerical results of the coefficient of the $AMSE$ of the nonrecursive and the proposed estimator with the optimal γ_0 obtained for each chosen $\kappa \in \{0.2, 0.4, 0.6, 0.7, 0.8, 1, 2, 4\}$.

	nonrecursive	Recursive	
κ	`Coeff`	`Coeff`	$[\gamma_0]$
$\kappa = 0.2$	1.356131	**0.439570**	[0.624817]
$\kappa = 0.4$	1.569193	**1.023768**	[0.718623]
$\kappa = 0.6$	1.716357	**1.635163**	[0.794197]
$\kappa = 0.7$	**1.772305**	1.930109	[0.827012]
$\kappa = 0.8$	**1.818969**	2.211670	[0.857143]
$\kappa = 1$	**1.889882**	2.725681	[0.910684]
$\kappa = 2$	**2.000000**	4.326444	[1.093070]
$\kappa = 4$	**1.889882**	5.076874	[1.270579]

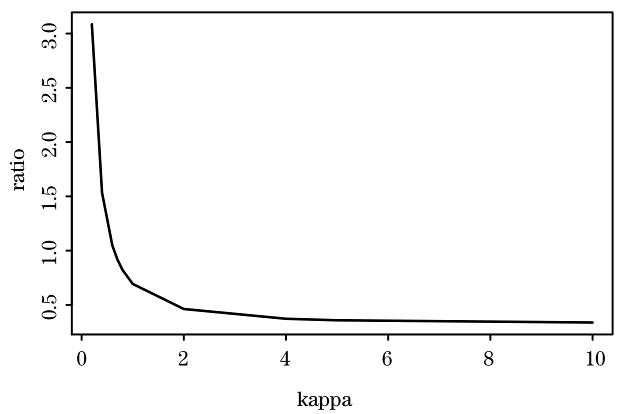

Figure 1. the ratio of $AMSE\left[\widetilde{r}_n(\chi, h)\right]$ and $AMSE\left[\widehat{r}_n(\chi, h)\right]$ with the optimal γ_0 in the function of κ.

and adapted to a functional version in Ferraty, Mas and Vieu (2007) is that, rather that using the naive bootstrap approach of resampling from the pairs $\{Y_i, \mathcal{X}_i\}_{i=1}^n$, we resample from the estimated residuals $\widehat{\varepsilon}_i = Y_i - \widehat{r}_n(\mathcal{X}_i)$. 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 $\mathbb{E}(\varepsilon_i^*) = 0$, $\mathbb{E}\left(\varepsilon_i^{*2}\right) = \widehat{\varepsilon}_i^2$, and $\mathbb{E}\left(\varepsilon_i^{*3}\right) = \widehat{\varepsilon}_i^3$. Such a distribution is equal to

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

Our adapted procedure for the bandwidth selection when estimating the operator r recursively in the case of functional setting is as follows:

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

Step 2: Resampling, new observations $Y_i^* = \widehat{r}_n^* (\chi_i, g) + \varepsilon_i^*$, where

$$\widehat{r}_n^* (\chi, g) = \frac{\Pi_n \sum_{k=1}^n \Pi_k^{-1} \gamma_k g_k^{-1} K \left(\|\chi - \mathcal{X}_k\| / g_k \right) Y_k^*}{\Pi_n \sum_{k=1}^n \Pi_k^{-1} \gamma_k g_k^{-1} K \left(\|\chi - \mathcal{X}_k\| / g_k \right)},$$

and g should be larger than h (an explanation of why it is essential to oversmooth g is given later).

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

$$\widehat{r}_n^* (\chi, h) = \frac{\Pi_n \sum_{k=1}^n \Pi_k^{-1} \gamma_k h_k^{-1} K \left(\|\chi - \mathcal{X}_k\| / h_k \right) Y_k^*}{\Pi_n \sum_{k=1}^n \Pi_k^{-1} \gamma_k h_k^{-1} K \left(\|\chi - \mathcal{X}_k\| / h_k \right)}.$$

The bootstrapped bandwidth h^* is then defined by:

$$h^* = h^* (\chi) = \arg\min_{h \in H} \left(\frac{1}{N_B} \sum_{b=1}^{N_B} \left(\widehat{r}_n^* (\chi, h) - \widehat{r}_n (\chi, g) \right)^2 \right),$$

where H is a fixed set of bandwidths and N_B is the number of replications. The wild bootstrap method for the nonrecursive regression estimator when the explanatory data are curves and the response is real is given in Ferraty, Mas and Vieu (2007).

The bootstrap bias of the estimator constructed from the resampled data is

$$\begin{aligned} \widehat{b}_n (\chi, h, g) &= \mathbb{E}^* \left[\widehat{r}_n^* (\chi, h) \right] - \widehat{r}_n (\chi, g) \\ &= \Pi_n \sum_{k=1}^n \Pi_k^{-1} \gamma_k h_k^{-1} K \left(\frac{\|\chi - \mathcal{X}_k\|}{h_k} \right) \frac{\widehat{r}_n (\chi, g)}{\widehat{f}_n (\chi)} - \widehat{r}_n (\chi, g) \\ &= \frac{\phi_n (\chi, h, g)}{\widehat{f}_n (\chi)} - \widehat{r}_n (\chi, g), \end{aligned} \tag{2.10}$$

where

$$\begin{aligned} \phi_n (\chi, h, g) &= \Pi_n \sum_{k=1}^n \Pi_k^{-1} \gamma_k h_k^{-1} K \left(\frac{\|\chi - \mathcal{X}_k\|}{h_k} \right) \widehat{r}_n (\chi, g) \\ &= \frac{\psi_{n,1} (\chi, h, g)}{\widehat{f}_n (\chi, g)} + \frac{\psi_{n,2} (\chi, h, g)}{\widehat{f}_n (\chi, g)}, \end{aligned}$$

with

$$\psi_{n,1} (\chi, h, g) = \Pi_n^2 \sum_{k=1}^n \Pi_k^{-2} \gamma_k^2 h_k^{-1} g_k^{-1} K \left(\frac{\|\chi - \mathcal{X}_k\|}{h_k} \right) K \left(\frac{\|\chi - \mathcal{X}_k\|}{g_k} \right) Y_k,$$

$$\psi_{n,2} (\chi, h, g) = \Pi_n^2 \sum_{\substack{k,k'=1 \\ k \neq k'}}^n \Pi_k^{-1} \Pi_{k'}^{-1} \gamma_k \gamma_{k'} h_k^{-1} g_{k'}^{-1} K \left(\frac{\|\chi - \mathcal{X}_k\|}{h_k} \right) K \left(\frac{\|\chi - \mathcal{X}_{k'}\|}{g_{k'}} \right) Y_{k'},$$

$$\widehat{f}_n(\chi, g) = \Pi_n \sum_{k=1}^{n} \Pi_k^{-1} \gamma_k g_k^{-1} K\left(\frac{\|\chi - \mathcal{X}_k\|}{g_k}\right).$$

For an explanation of why the bandwidth g_n should be larger than h_n, we let

$$\widehat{b}_n(\chi, h) = \widehat{r}_n(\chi, h) - r(\chi),$$

and then prove the following theorem.

Theorem 2. *Let Assumptions* (A1)-(A4) *hold. Then,*

$$\begin{aligned}&\mathbb{E}\left[\left(\widehat{b}_n(\chi, h, g) - \widehat{b}_n(\chi, h)\right)^2\right]\\&\simeq C_1 \frac{\gamma_n}{F(g_n)} + C_2 \frac{F(h_n)}{F(g_n)} + C_3 \frac{\gamma_n}{F(h_n)} + C_4 h_n^2 + C_5 g_n^2,\end{aligned}$$

where,

$$\begin{aligned}C_1 &= \frac{(2 - (\mathcal{F}_a + \alpha - 2a)\xi)(2 - (\mathcal{F}_g + \alpha - 2a)\xi)}{(4 - (\mathcal{F}_a + 3\alpha - 2a - 2g)\xi)}\left(r^2(\chi) + \sigma_\varepsilon^2(\chi)\right)\frac{K^2(0)}{M_2}\\&\quad + \sigma_\varepsilon^2(\chi)\frac{(1 - (\mathcal{F}_g - g)\xi)^2}{(1 - (\mathcal{F}_g + \alpha - 2g)\xi)}\frac{M_2}{M_1^2},\\C_2 &= \frac{(2 - (\mathcal{F}_a + \alpha - 2a)\xi)(2 - (\mathcal{F}_g + \alpha - 2a)\xi)}{(2 - (\mathcal{F}_a + \alpha - a - g)\xi)^2} r^2(\chi) K^2(0)\frac{M_1^2}{M_2^2},\\C_3 &= \sigma_\varepsilon^2(\chi)\frac{(1 - (\mathcal{F}_a - a)\xi)^2}{(1 - (\mathcal{F}_a + \alpha - 2a)\xi)}\frac{M_2}{M_1^2},\\C_4 &= \left(\phi'(0)\right)^2 \frac{(1 - (\mathcal{F}_a - a)\xi)^2}{(1 - \mathcal{F}_a\xi)^2}\frac{M_0^2}{M_1^2},\\C_5 &= \left(\phi'(0)\right)^2 \frac{(1 - (\mathcal{F}_g - g)\xi)^2}{(1 - \mathcal{F}_g\xi)^2}\frac{M_0^2}{M_1^2}.\end{aligned}$$

Theorem 2 shows that the distribution of $\widehat{r}_n(\chi, h) - r(\chi)$ is approximated by the distribution $\widehat{r}_n^*(\chi, h) - \widehat{r}_n(\chi, g)$. Moreover, we need that $\widehat{r}_n(\chi, h)$ tends to $r(\chi)$. This requires choosing g_n tending to zero at a rate slower than the optimal bandwidth h_n for estimating $\widehat{r}_n(\chi)$.

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; we can check whether it follows from (1.2) that, for all $n_1 \in [0, n-1]$,

$$\widehat{a}_n(\chi, h) = \prod_{j=n_1+1}^{n} (1 - \gamma_j)\, \widehat{a}_{n_1}(\chi, h)$$

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

where $\alpha_1=\prod_{j=n_1+1}^{n}(1-\gamma_j)$ and $\beta_k=\prod_{j=k+1}^{n}(1-\gamma_j)$.

Similarly, it follows from (1.3) that for all $n_1\in[0,n-1]$,

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

Here, we suppose that we receive a first sample of size $n_1=\lfloor n/2\rfloor$ (the lower integer part of $n/2$). Then, we receive an additional sample of size $n-n_1$. It is clear that we can use a plug-in or a wild bootstrap to construct an optimal bandwidth based on the first sample of size n_1 and 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 improve the computational cost significantly.

Remark 2. It is possible to suppose that we receive more than two samples separately.

3. Applications

The aim of our applications is to compare the performance of the recursive estimators defined in (1.4) to the nonrecursive estimator defined in (1.5) using a resampling bootstrap method.

3.1. Simulation studies

We construct random curves in the following way:

$$
\begin{aligned}
\mathcal{X}(t)=&\,a\cos(4t)+b\cos(5t)+c\cos(6t)+d\sin(5t)+e\sin(6t)\\
&+f\sin(7t)+g(t-\pi)^2,\quad t\in[0,2\pi],
\end{aligned}
$$

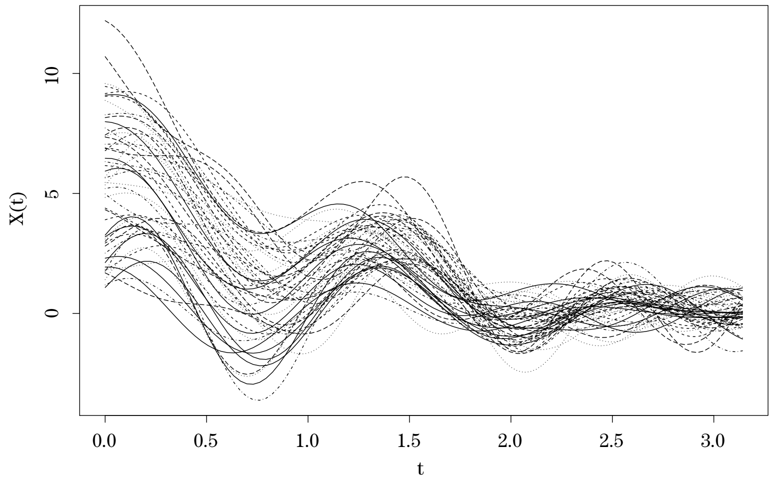

Figure 2. A sample of 50 simulated curves.

where a, b, d, e, and g are real random variables drawn from a uniform distribution on $(0,1)$, and c and f are real random variables drawn from a normal distribution $\mathcal{N}(0, 0.5)$. Each curve is discretized into $p = 100$ equidistant points on $[0, \pi]$.

The response variable is simulated from the following regression model:

$$Y = r(\mathcal{X}) + \varepsilon, \quad \text{with} \quad \varepsilon \sim \mathcal{N}(0, 1),$$

and where

$$r(\mathcal{X}) = \int_0^\pi \left|\mathcal{X}'(t)\right| \sin\left(\frac{\pi}{2}t\right) dt.$$

Some of these curves are presented in Figure 2. For our application, we simulated two samples: a learning sample of size $n_l = 200$, on which all estimates are computed, and a testing sample of size $n_t = 100$, which is used to examine the behavior of our method. The number of bootstrap replications is $N_B = 500$, for each application. In this functional context, the proposed estimator depends on the following parameters; First, the semi-norm $\|.\|$ of the functional space $\mathcal{E}$ is taken to be the L_2- norm between the first-order derivatives of the curves. Second, because the choice of the kernel function K is not crucial, we use the quadratic kernel $K(u) = \left(1 - u^2\right) \mathbb{1}_{[0,1]}(u)$ for all $u \in \mathbb{R}$. The bandwidth h is assumed to belong to some grid in terms of nearest neighbors, $h \in \{h_1, \ldots, h_{50}\}$, where h_k is the radius of the ball of center χ, containing exactly k among the curves data $\mathcal{X}_1, \ldots, \mathcal{X}_{200}$.

We provide a box-plot (see, Figure 3) of the quantities $(\widehat{Y}^{[j]} - Y^{[j]})^2$, where

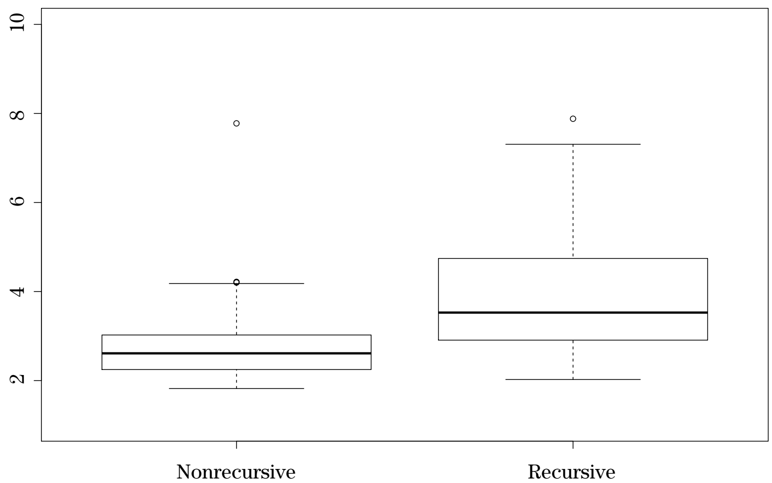

Figure 3. The mean square prediction error (MSPE) of the nonrecursive estimator (1.5) and the proposed recursive estimator (1.4) over 500 bootstrap replications of $n = 200$ curves.

$\widehat{Y}^{[j]}$ represents the predicted value at the j^{th} iteration of the simulation ($j = 1, \ldots, 500$). From these results, we observe that the nonrecursive estimator proposed by Ferraty and Vieu (2002) is better than our proposed recursive estimators in terms of the estimation error, but the main interest of using our recursive estimators is because it can give much better computational time. Performing the two methods, the running time using the recursive regression estimators (1.4) was roughly 28 s on the author's workstation, whereas that using the nonrecursive regression estimator (1.5) was roughly 50 s on the same workstation.

3.2. A real-data chemometric application

These data are available online at `http://www.lsp.ups-tlse.fr/staph/npfda/npfda-spectrometric.dat`. This time series of spectra has been measured from wavelengths $\lambda = 850$ to $\lambda = 1{,}050$ nm for 215 finely chopped pieces of meat. From this time 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. In addition, the responses include the percentage of fatness. These curves are graphed in Figure 4. Moreover, as measures of proximity, we focus on the family of semi-metrics

$$\sqrt{\int \left(\chi_i^{(m)}(t) - \chi_j^{(m)}(t)\right)^2 dt}, \quad m \in \{0, 1, 2, 3\},$$

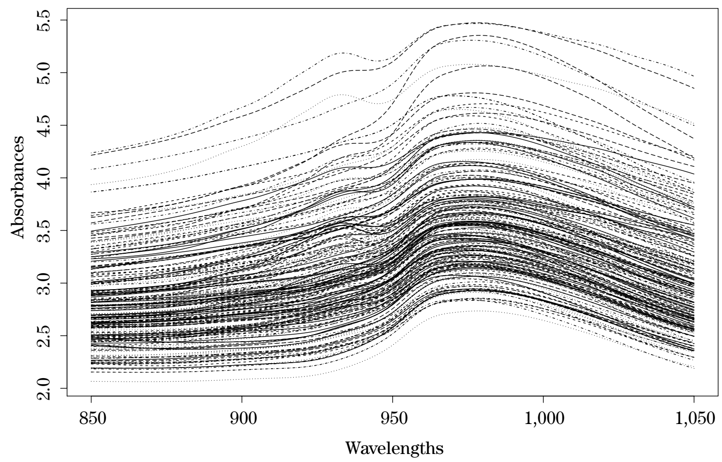

Figure 4. Spectrometric curves data.

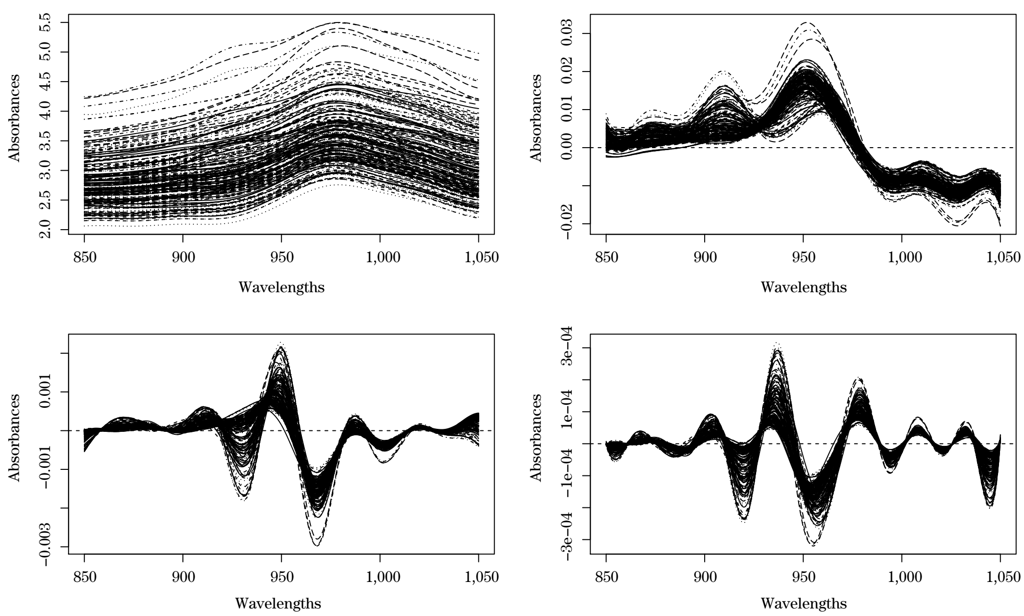

Figure 5. Shape of the dervatives of the spectrometric curves, $m = 0$ (in the top left panel), the first derivative (in the top right panel), the second derivative (in the down left panel), and the third derivative (in the down right panel).

where $\chi^{(m)}$ denotes the mth derivatives of χ and $\chi^{(0)} = \chi$. We plot the successive derivatives in Figure 5 (using B-spline approximation, see Febrero-Bande and Oviedo de la Fuente (2012)). We can observe that the second derivative act as a filter and then can select more pertinent information.

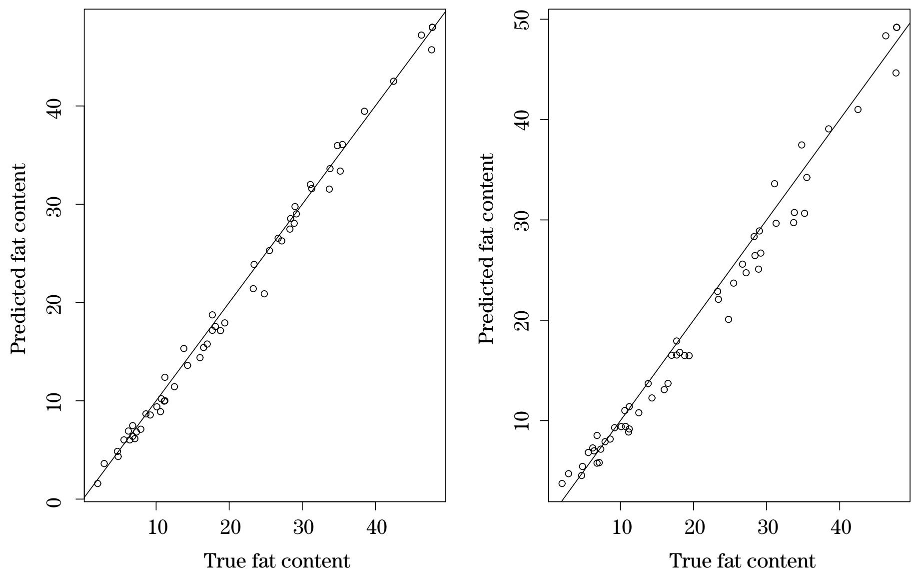

Figure 6. Spectrometric data: Predicted values on the testing sample using the nonrecursive estimator (in the left panel) and using the recursive estimator (in the right panel).

Our main interest in this section is to compare the performance of the two methods by determining the relation between the spectrum and the fatness. We estimate functional regression model using the nonrecursive estimator (1.5) and the recursive estimator (1.4).

Our sample of 215 pairs $(\mathcal{X}_i, Y_i)$ will be decomposed into a learning sample $(\mathcal{L})$ of size 160 on which the various statistical methods are constructed and a second sample $(\mathcal{T})$ of size 55 on which the predictive performances of these methods is tested. We measure the performance of the two estimators using the MSPE:

$$MSPE = \frac{1}{55} \sum_{i \in \mathcal{T}} \left(\widehat{Y}_i - Y_i \right)^2,$$

where $\widehat{Y}_i$ is the prediction for Y_i obtained for each new curve $\mathcal{X}_i$, $i \in \mathcal{T}$ using one of the two estimators. The $MSPE$ was computed using the recursive estimator ($MSPE\,(\texttt{Recursive}) = 3.831634$) or the nonrecursive estimator ($MSPE\,(\texttt{nonrecursive}) = 1.211091$). The nonrecursive estimator gives a smaller $MSPE$ than that of the recursive estimator. Performing the two methods, the running time using the recursive regression estimators (1.4) was roughly 63 s on the author's workstation, whereas that of the nonrecursive estimator (1.5) was roughly 176 s on the author's workstation. Moreover, in Figure 6, we plot the predicted values obtained using the two methods as a function of the true value for the 55 spectra in our testing sample.

4. Conclusion

We propose an automatic selection of the bandwidth of a recursive non-parametric regression estimation for independent functional data. The proposed estimators asymptotically follow a normal distribution. The estimators are compared with the nonrecursive Ferraty and Vieu regression estimator for functional data. We showed that, using some selected bandwidth and stepsizes, the proposed recursive estimators will be very competitive. Our simulation confirms the nice features of our proposed recursive estimators and satisfactory improvement in `CPU` time compared with the nonrecursive estimator.

In conclusion, the proposed method allows us to obtain a competitive estimate to that of the nonrecursive Ferraty and Vieu (2002). Moreover, we plan to extend this work by proposing other bandwidth selection methods.

Supplementary Material

The Supplementary Material contains the proofs for the main results stated in the paper.

Acknowledgments

The author would like to thank the editor, associate editor and two referees for their helpful comments, which helped me to focus on improving the original version of the paper.

Appendix

A. Proofs

Throughout this section we use the following notation:

$$\Pi_n = \prod_{j=1}^{n} (1-\gamma_j),$$

$$a_n(\chi) = \Pi_n \sum_{k=1}^{n} \Pi_k^{-1} \gamma_k h_k^{-1} K\left(\frac{\|\chi - \mathcal{X}_k\|}{h_k}\right) Y_k,$$

$$f_n(\chi) = \Pi_n \sum_{k=1}^{n} \Pi_k^{-1} \gamma_k h_k^{-1} K\left(\frac{\|\chi - \mathcal{X}_k\|}{h_k}\right), \tag{A.1}$$

Let us first state the following technical lemma.

Lemma 1. *Let* $(v_n) \in \mathcal{GS}(v^*)$, $(\gamma_n) \in \mathcal{GS}(-\alpha)$, *and* $m > 0$ *such that* $m - v^*\xi > 0$ *where* ξ *is defined in* (2.2). *We have*

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

Moreover, for all positive sequence (b_n) *such that* $\lim_{n\to+\infty} b_n = 0$, *and all* $\delta \in \mathbb{R}$,

$$\lim_{n\to+\infty} v_n \Pi_n^m \left[\sum_{k=1}^{n} \Pi_k^{-m} \frac{\gamma_k}{v_k} b_k + \delta\right] = 0.$$

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.

The proof Proposition 1, Theorem 1 and Theorem 2 are given in Supplementary material.

References

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

Benhenni, K., Hedli-Griche, S. and Rachdi, M. (2010). Estimation of the regression operator from functional fixed-design with correlated errors. *J. Multivariate Anal.* **101**, 476–490.

Biau, G., Cerou, F. and Guyader, A. (2010). Rates of convergence of the functional k-nearest neighbor estimate. *IEEE Trans. Inform. Theory* **56**, 2034–2040.

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

Cardot, H., Ferraty, F. and Sarda, P. (1999). Functional linear model. *Statist. Probab. Lett.* **45**, 11–22.

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

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

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

Ferraty, F. and Vieu, P. (2002). The functional nonparametric model and application to spectrometric data. *Comput. Statist.* **17**, 545–564.

Ferraty, F. and Vieu, P. (2004). Nonparametric models for functional data, with application in regression, time-series prediction and curve discrimination. *J. Nonparametr. Stat.* **16**, 111–125.

Ferraty, F. and Vieu, P. (2006). Nonparametric Functional Data Analysis: Theory and Practice. *Springer Ser. Statist.*, Springer, NewYork.

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

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

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

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

Lian, H. (2012). Convergence of nonparametric functional regression estimates with functional responses. *Electron. J. Stat.* **6**, 1373–1391.

Masry, E. (2005). Nonparametric regression estimation for dependent functional data: Asymptotic normality. *Stochastic Process. Appl.* **115**, 155–177.

Mokkadem, A. and Pelletier, M. (2007). A companion for the Kiefer-Wolfowitz-Blum stochastic approximation algorithm. *Ann. Statist.* **35**, 1749–1772.

Mokkadem, A., Pelletier, M. and Slaoui, Y. (2009a). The stochastic approximation method for the estimation of a multivariate probability density. *J. Statist. Plann. Inference* **139**, 2459–2478.

Mokkadem, A., Pelletier, M. and Slaoui, Y. (2009b). Revisiting Révész's stochastic approximation method for the estimation of a regression function, *ALEA Lat. Am. J. Probab. Math. Stat.* **6**, 63–114.

Nadaraya, E. A. (1964). On estimating regression. *Theory Probab. Appl.* **10**, 186–190.

Preda, C. (2007). Regression models for functional data by reproducing kernel Hilbert spaces methods. *J. Statist. Plann. Inference* **137**, 829–840.

Ramsay, J. O. and Silverman, B. W. (2002). *Applied Functional Data Analysis: Methods and Case Studies*. Springer, New York.

Slaoui, Y. (2016). Optimal bandwidth selection for semi-recursive kernel regression estimators. *Stat. Interface.* **9**, 375–388.

Watson, G. S. (1964). Smooth regression analysis. *Sankhya A* **26**, 359–372.

Laboratoire de Mathématiques et Applications, University of Poitiers, UMR 7348 du CNRS, Téléport 2- BP 30179, 11 Boulevard Marie et Pierre Curie, 86962 Futuroscope Chasseneuil, France.

E-mail: yousri.slaoui@math.univ-poitiers.fr

(Received October 2017; accepted April 2018)