Statistica Sinica Preprint No: SS-2018-0069

Title	Recursive nonparametric regression estimation for independent functional data
Manuscript ID	SS-2018-0069
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202018.0069
Complete List of Authors	Yousri Slaoui
Corresponding Author	Yousri Slaoui
E-mail	yousri.slaoui@math.univ-poitiers.fr

Notice: Accepted version subject to English editing.

Recursive nonparametric regression estimation for independent functional data

Yousri Slaoui

Université de Poitiers

Abstract: In this paper we propose an automatic selection of the bandwidth of the recursive nonparametric estimation of the regression function defined by the stochastic approximation algorithm, when the explanatory data are curves and the response is real. We compared our recursive estimators with the non-recursive one proposed by Ferraty and Vieu (2002), the two methods are based on wild bootstraping idea, where resampling is done from a suitably estimated residual distribution. Moreover, we established a central limit theorem for our proposed recursive estimators, using the wild bootstrap selected bandwidth and some special stepsizes, the proposed recursive estimators will be very competitive to the nonrecursive one in terms of estimation error but much better in terms of computational costs. The proposed estimators are used both on simulated and real functional datasets.

Key words and phrases: Stochastic approximation algorithm, Asymptotic Normality, Functional Data, Regression estimation, Wild Functional Bootrstrap, Smoothing, curve fitting.

1. Introduction

The progress of the computer tool, both in memory capacity and in calculation, makes it possible to record more and more voluminous data. Thus, a very large number of variables can be observed for the study of the same phenomenon. This is especially the case when one has 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 $\{\theta_i\}_{i=1,\ldots,I}$ of Θ in order to consider that the behavior of $\{X(\theta)\}_{\theta\in\Theta}$ is close to the one of $\{X(\theta_i)\}_{i=1,\ldots,I}$. Such a family $\{X(\theta)\}_{\theta\in\Theta}$ of random variables are called a functional random variable (see Ferraty and Vieu (2002, 2004)).

There has been an increasing interest in functional data analysis. For an introduction to this field, the reader can check the popular monograph Ramsy and Silverman (2002) that gives a detailed expostion of both the theoretical and practical aspects of functional data analysis. The existing literature contains numerous studies on functional linear models (see, among many others, Cardot et al. (1999), Cai and Hall (2006), Hall and Horowitz (1997)). In the framework of non-parametric estimation with functional predictors and scalar responses (see, among others Ferraty and Vieu (2002, 2004, 2006), Preda (2007) and Biau, Gerou and Guyader (2010)).

In this paper, we are interested in the problem of the recursive estimation of the 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 $\mathbb{E}|Y_i| < \infty$ and define the regression functional as

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

The model (1.1) can be written as follows

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

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

The purpose of the current paper is to adapt the recursive estimator proposed in Slaoui (2016) to the case when the covariate is functional and the response is real. Then, our proposal estimators to estimate recursively the operator r is the following :

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

with

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

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

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

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

Several assumptions will be made on the kernel K, on the bandwidth (h_n) and on the stepsize (γ_n). The recursive property (1.4) is particularly useful in large sample size since $\widehat{r}_n$ can be easily updated with each additional observation.

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

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

where F is the comulative distribution function of the random variable $\|\mathcal{X}-\chi\|$. One can check easily that, this estimators are a special case of our proposed recusive estimators (1.4), with the choice of the stepsize $(\gamma_n)=\left(h_nF\left(h_n\right)^{-l}\left[\sum_{k=1}^n h_kF\left(h_k\right)^{-l}\right]^{-1}\right)$.

Further, we show in the special case when $\mathcal{X}$ is geometric process (or fractal) with $F\left(h_n\right)\sim h_n^{\kappa}$, with $\kappa>0$, that the optimal bandwidth which minimize the $\mathbb{E}\left[\widehat{r}_n\left(\chi,h\right)-r\left(\chi\right)\right]^2$ depends on the choice of the stepsize (γ_n); we show in particular that under some conditions of regularity of the functional regression r and using the stepsizes $(\gamma_n)=(\gamma_0n^{-1})$, where $\gamma_0>0$, the bandwidth (h_n) must equal

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

The first purpose of this paper is to propose an automatic selection of such bandwidth through a plug-in method and then through a wild bootstrap method, and the second purpose is to compare the proposed recursive estimators r_n to the nonrecursive functional regression estimator introduced by Ferraty and Vieu (2002) in the case of independent data. They constructed the functional estimate of the operator r by following the standard kernel methods (Nadaraya (1964) and Watson (1964)), and defined as

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

This estimator was considered by Ferraty and Vieu (2004, 2006), while Masry (2005) considered the asymptotic normality of (1.5) in the dependent case. Benhenni (2010) consider the case of fixed-design with correlated errors. Lian (2012) consider the case when both predictors and responses are 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 dataset (subsection 3.2). We conclude the article in Section 4. Appendix A gives the proof 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

$$\lim_{n\to+\infty} n\left[1 - \frac{v_{n-1}}{v_n}\right] = \gamma. \tag{2.6}$$

Condition (2.6) was introduced by Galambos and Seneta (1973) to define regularly varying sequences (see also Bojanic and Seneta (1995)) and by

Mokkadem and Pelletier (2007) in the context of stochastic approximation algorithms. Noting that the acronym $\mathcal{GS}$ stand 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 asymptotic properties of the proposed estimators (1.4). The assumptions to which we shall refer are the following:

(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 $s \in [0,1]$, $\tau_h(s) := \frac{F(hs)}{F(h)} \to \tau_0(s) < \infty$ as $h \to 0$.

(A4) *i)* $(\gamma_n) \in \mathcal{GS}(-\alpha)$ with $\alpha \in]1/2, 1]$.

ii) $(h_n) \in \mathcal{GS}(-a)$ with $a \in]0,1[$.

iii) $(F(h_n)) \in \mathcal{GS}(-\mathcal{F}_a)$ with $\mathcal{F}_a \in]0, \alpha[$.

iv) $\lim_{n\to\infty}(n\gamma_n) \in]\min\{\mathcal{F}_a, (\alpha + \mathcal{F}_a)/2 - a\}, \infty]$.

v) $(g_n) \in \mathcal{GS}(-g)$ with $g \in]0, a[$.

vi) $(F(g_n)) \in \mathcal{GS}(-\mathcal{F}_g)$ with $\mathcal{F}_g \in]0, \mathcal{F}_a[$.

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 notations:

$$\begin{aligned}
\xi &= \lim_{n\to\infty} (n\gamma_n)^{-1}, && (2.7)\\
M_0 &= K(1) - \int_0^1 (tK(t))' \tau_0(t)\,dt,\\
M_1 &= K(1) - \int_0^1 K'(t)\,\tau_0(t)\,dt,\\
M_2 &= K^2(1) - \int_0^1 \left(K^2(t)\right)' \tau_0(t)\,dt.
\end{aligned}$$

Our first result is the following proposition, which gives the bias and the variance of 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*

$$\mathbb{E}\left[\widehat{r}_n(\chi,h)\right] - r(\chi) = h_n\phi'(0)\,\frac{1-(\mathcal{F}_a - a)\,\xi}{1-\mathcal{F}_a\xi}\,\frac{M_0}{M_1}\left[1+o(1)\right]. \quad (2.8)$$

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*

$$Var[\widehat{r}_n(\chi, h)] = \sigma_\varepsilon^2(\chi) \frac{(1-(\mathcal{F}_a - a)\xi)^2}{(2-(\mathcal{F}_a+\alpha-2a)\xi)} \frac{M_2}{M_1^2} \frac{\gamma_n}{F(h_n)} [1+o(1)]. \quad (2.9)$$

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

$$Var[\widehat{r}_n(\chi, h)] = o\left(h_n^2\right). \quad (2.10)$$

3. *If* $\lim_{n\to\infty}(n\gamma_n) > \max\{\mathcal{F}_a, (\mathcal{F}_a+\alpha)/2 - a\}$*, then (2.8) and (2.9) hold simultaneously.*

The bias and the variance of the estimator 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 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*

$$\sqrt{\gamma_n^{-1} F(h_n)} (\widehat{r}_n(\chi, h) - r(\chi)) \xrightarrow{\mathcal{D}} \mathcal{N}\left(c^{1/2}\phi'(0) \frac{1-(\mathcal{F}_a-a)\xi}{1-\mathcal{F}_a\xi} \frac{M_0}{M_1}, \sigma_\varepsilon^2(\chi) \frac{(1-(\mathcal{F}_a-a)\xi)^2}{(2-(\mathcal{F}_a+\alpha-2a)\xi)} \frac{M_2}{M_1^2}\right).$$

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

$$\frac{1}{h_n^2}\left(\widehat{r}_n(\chi,h)-r(\chi)\right)\overset{\mathbb{P}}{\to}\phi'(0)\frac{1-(\mathcal{F}_a-a)\xi}{1-\mathcal{F}_a\xi}\frac{M_0}{M_1},$$

where $\overset{\mathcal{D}}{\to}$ *denotes the convergence in distribution,* $\mathcal{N}$ *the Gaussian-distribution and* $\overset{\mathbb{P}}{\to}$ *the 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), the proposed estimator fulfils the following central limit theorem

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

We let ϕ denote the distribution function of the $\mathcal{N}(0,1)$, 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

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

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

$$\widehat{M}_i = \Pi_n\sum_{k=1}^n\Pi_k^{-1}\gamma_k\widehat{F}(h_k)^{-1}K^i\left(\frac{\|\chi-\mathcal{X}_k\|}{h_k}\right),\quad i\in\{1,2\}\tag{2.11}$$

$$\widehat{\sigma}_\varepsilon^2(\chi) = \frac{\Pi_n\sum_{k=1}^n\Pi_k^{-1}\gamma_kK\left(\frac{\|\chi-\mathcal{X}_k\|}{h_k}\right)Y_k^2}{\Pi_n\sum_{k=1}^n\Pi_k^{-1}\gamma_kK\left(\frac{\|\chi-\mathcal{X}_k\|}{h_k}\right)}-\left(\widehat{r}_n(\chi,h)\right)^2.\tag{2.12}$$

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 following 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*

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

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

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

2.1 Stepsize selection

In the framework of the nonparametric kernel estimators, to determine the optimal choice of stepsize, Mokkadem et al. (2009a) consider $(\gamma_n) \in$

$\mathcal{GS}(-1)$ to ensure the optimal convergence rate and consider two points of view: pointwise estimation and estimation by confidence intervals. From the pointwise estimation point of view, the criteria they consider to find the optimal stepsize is minimizing the mean squared error (MSE) or the integrated mean squared error ($MISE$). In our context, we display a set of stepsizes (γ_n) minimizing the MSE or the MISE of the estimator $\widehat{r}_n(\chi, h)$ defined by (1.4); we show in particular that the sequence $(\gamma_n) = (n^{-1})$ belongs to this set. Let us underline 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 pointwise estimation and when rapid updating is not such important, it is preferable to use the nonrecursive estimator rather than any recursive estimator $\widehat{r}_n(\chi, h)$ defined by (1.4).

Moreover, from the confidence interval point of view, the criteria they consider to find the optimal stepsize is minimizing the variance.In our context, we display a set of stepsizes (γ_n) minimizing the variance of $\widehat{r}_n(\chi, h)$; it follows from (2.9) that, the sequence $(\gamma_n) = ([1-a]\, n^{-1})$ 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 nonrecursive estimator $\widetilde{r}_n(\chi, h)$ defined by (1.5). Consequently, even in the case when the on-line aspect is not quite important, it is preferable to use recursive estimators to construct confidence intervals.

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

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

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

2.2 Bandwidth selection

In the framework of the nonparametric kernel estimators, the bandwidth selection methods studied in the literature can be divided into three broad classes: the cross-validation techniques, the plug-in methods and the bootstrap idea. A detailed comparison of the three practical bandwidth selection can be found in Delaigle and Gijbels (2004). They concluded that chosen appropriately plug-in and bootstrap selectors both outperform the cross-validation bandwidth, and that none of the two can be claimed to be best in all cases. In this paper, we consider first 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 geometric process (or fractal) with $F(h_n) \sim h_n^{\kappa}$, with $\kappa > 0$.

Recursive estimators The following corollary indicates that the bandwidth which minimizes the MSE of 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

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

The following corollary shows that, for a special choice of the stepsize $(\gamma_n)=(\gamma_0 n^{-1})$, which fulfilleds that $\lim_{n\to\infty} n\gamma_n=\gamma_0$ and that $(\gamma_n)\in\mathcal{GS}(-1)$, the optimal value for h_n depends on γ_0 and then the corresponding $AMSE$ depend 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* $\mathcal{GS}(-1)$*, the bandwidth* (h_n) *must equal*

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

Then, we have

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

where $\texttt{Coeff}(\kappa,\gamma_0)=3\times(\kappa/4)^{2/(\kappa+2)}\left(\sigma_\varepsilon^2(\chi)\right)^{2/(\kappa+2)}\gamma_0^{-2/(\kappa+2)}\frac{\left(\gamma_0+\frac{\kappa+1}{\kappa+2}\right)^2}{\left(\gamma_0+\frac{\kappa}{\kappa+2}\right)^{2\kappa/(\kappa+2)}}$.

Then, under assumptions $(A1)-(A4)$. The plug-in bandwidth (h_n) must equal

$$\left(\left\{\frac{\kappa}{2}\frac{\left(\gamma_0+\frac{\kappa}{\kappa+a}\right)^2}{2\gamma_0}\right\}^{1/(\kappa+2)}\widehat{\sigma}_{\varepsilon}^2\left(\chi\right)\frac{\widehat{M_2}}{\widehat{I}_0^2}n^{-1/(\kappa+2)}\right). \tag{2.13}$$

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

$$\begin{aligned} AMSE\left[\widehat{r}_n\left(\chi,h\right)\right] &= \texttt{Coeff}\left(\kappa,\gamma_0\right)\left(\widehat{\sigma}_{\varepsilon}^2\left(\chi\right)\right)^{2/(\kappa+2)} \\ &\quad \widehat{M_1}^{-2}\widehat{M_2}^{2/(\kappa+2)}\widehat{I}_0^{2\kappa/(\kappa+2)}n^{-2/(\kappa+2)}, \end{aligned} \tag{2.14}$$

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

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

where, $\widehat{M_i}$, for $i\in\{1,2\}$ and $\widehat{\sigma}_{\varepsilon}^2\left(\chi\right)$ are given in (2.11) and (2.12).

Non-recursive estimator Let us first recall that under the assumptions $(A1)-(A3)$ and $(A4)\,ii)$ the bias and variance of $\widetilde{r}_n\left(\chi,h\right)$ 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,

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

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

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

Then, we have

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

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

$$\left(\left\{\frac{\kappa}{2}\widetilde{M_2}\frac{\widetilde{\sigma}_\varepsilon^2(\chi)}{\widetilde{I}_0^2}\right\}^{1/(\kappa+2)} n^{-1/(\kappa+2)}\right).$$

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

$$AMSE\left[\widetilde{r}_n(\chi,h)\right] = \texttt{Coeff}(\kappa)\left(\sigma_\varepsilon^2(\chi)\right)^{2/(\kappa+2)}\widetilde{I}_0^{2\kappa/(\kappa+2)}\widetilde{M_1}^{-2}\widetilde{M_2}^{2/(\kappa+2)} n^{-2/(\kappa+2)},$$

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

$$\begin{aligned}
\widetilde{I}_0 &= \frac{1}{n\widehat{F}(h_n)}\sum_{k=1}^{n}(Y_k - r(\chi))\,K\left(\frac{\|\chi-\mathcal{X}_k\|}{h_n}\right)\\
\widetilde{M}_i &= \frac{1}{n\widehat{F}(h_n)}\sum_{k=1}^{n}K^i\left(\frac{\|\chi-\mathcal{X}_k\|}{h_n}\right), \quad i\in\{1,2\}\\
\widetilde{\sigma}_{\varepsilon}^2(\chi) &= \frac{\sum_{k=1}^{n}K\left(\frac{\|\chi-\mathcal{X}_k\|}{h_n}\right)Y_k^2}{\sum_{k=1}^{n}K\left(\frac{\|\chi-\mathcal{X}_k\|}{h_n}\right)} - \left(\widetilde{r}_n(\chi,h)\right)^2.
\end{aligned}$$

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

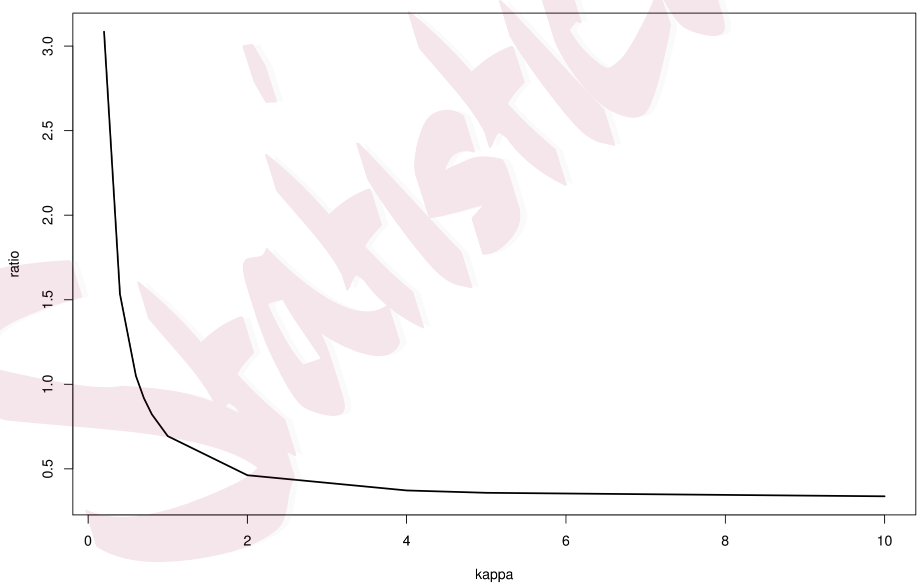

Figure 1: the ratio of $AMSE\left[\widetilde{r}_n(x)\right]$ and $AMSE\left[\widehat{r}_n(x)\right]$ with the optimal γ_0 in function of κ.

	Non-recursive	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]

Table 1: Some numerical results concerning the coefficient of the $AMSE$ of the non-recursive and the proposed estimator with the optimal γ_0 obtained for each chosen $\kappa \in \{0.2, 0.4, 0.6, 0.7, 0.8, 1, 2, 4\}$.

2.2.2 Wild bootstrap method

The main idea of the wild bootstrap proposed in Härdle and Marron (1991) and adapted to functional version in Ferraty et al. (2007) is rather to use the naive bootstrap approach of resampling from the pairs $\{Y_i, \mathcal{X}_i\}_{i=1}^n$, is to resample from the estimated residuals $\widehat{\varepsilon}_i = Y_i - \widehat{r}_n(\mathcal{X}_i)$, and then use the obtained data to construct an estimator whose distribution will approximate the distribution of the original estimator and where each bootstrap residual ε_i^* is drawn from two-point distribution, such that $\mathbb{E}\left(\varepsilon_i^*\right) = 0$, $\mathbb{E}\left(\varepsilon_i^{*2}\right) = \widehat{\varepsilon}_i^2$ and $\mathbb{E}\left(\varepsilon_i^{*3}\right) = \widehat{\varepsilon}_i^3$. Such distribution is equal to

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

Our adapted procedure for bandwidth selection to estimate the operator r recursively in the case of functional setting :

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

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

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

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,

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

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 the number of replications. The wild bootstrap method in the case of the non-recursive regression estimator when the explanatory data are curves and the response is real is given in Ferraty et al. (2007).

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

$$\begin{aligned}\widehat{b}_n(\chi,h,g) &= \mathbb{E}^*\left[\widehat{r}_n^*(\chi,h)\right] - \widehat{r}_n(\chi,g)\\ &= \Pi_n\sum_{k=1}^{n}\Pi_k^{-1}\gamma_k h_k^{-1}K\left(\frac{\|\chi-\mathcal{X}_k\|}{h_k}\right)\frac{\widehat{r}_n(\chi,g)}{\widehat{f}_n(\chi)} - \widehat{r}_n(\chi,g)\\ &= \frac{\phi_n(\chi,h,g)}{\widehat{f}_n(\chi)} - \widehat{r}_n(\chi,g), \end{aligned} \tag{2.15}$$

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 we prove the following Theorem.

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

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

where,

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

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

Computational cost The advantage of recursive estimators on their nonrecursive version is that their update, from a sample of size n to one of size $n+1$, require less computations. This property can be generalized, one can check that it follows from (1.2) that for all $n_1 \in [0, n-1]$,

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

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 $n_1 \in [0, n-1]$,

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

here, we suppose that we receive a first sample of size $n_1 = \lfloor n/2 \rfloor$ (the lower integer part of $n/2$) an then, we suppose that we receive an additional sample of size $n - n_1$. It is clear, that we can use a plug-in idea or a wild bootstrap 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 $n - n_1$, and then the proposed estimator can be viewed as a linear combination of two estimators, which improve considerably the computational cost.

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:

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

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

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

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

We provided the box-plot (see, Figure 3) of the quantities $\left(\widehat{Y}^{[j]} - Y^{[j]}\right)^2$, where $\widehat{Y}^{[j]}$ represent the predicted value at the $j^{\mathtt{th}}$ iteration of the simulation ($j = 1, \ldots, 500$). From these resuls, we observed that the non-recursive estimator proposed by Ferraty and Vieu (2002) is better than our proposed recursive estimators in terms of estimation error but the main interest of using our recursive estimators resides on the fact that it can give much better computational time. Performing the two methods, the running time using the recursive regression estimators (1.4) was roughly 28s on the author's workstation, while the running time using the non-recursive regression estimator (1.5) was roughly 50s on the author's workstation.

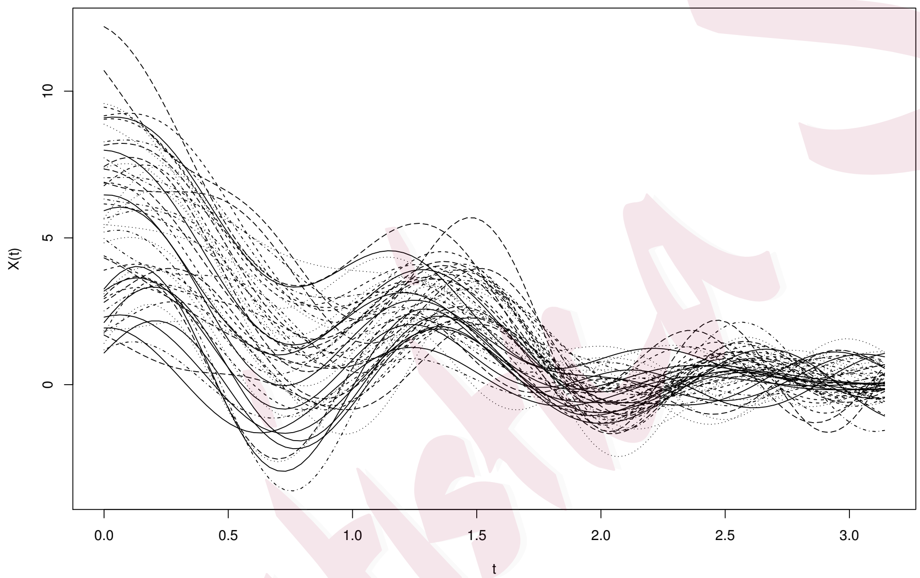

Figure 2: A sample of 50 simulated curves.

3.2 A real data chemometric application

This data is available on line at

`http://www.lsp.ups-tlse.fr/staph/npfda/npfda-spectrometric.dat`.

This time series of spectra have been measured from wavelengths $\lambda = 850$

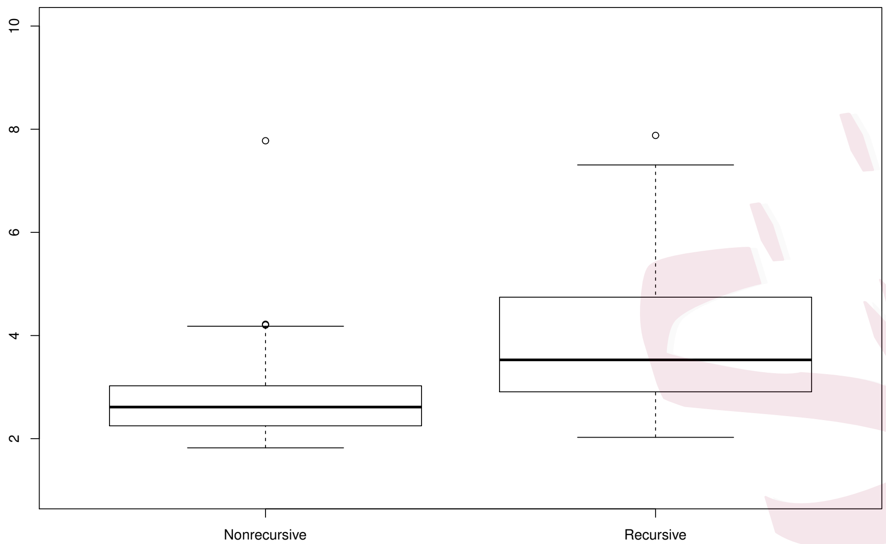

Figure 3: The Mean Square Prediction Error (MSPE) of the non-recursive estimator (1.5) and the proposed recursive estimator (1.4) over 500 bootstrap replications of $n = 200$ curves.

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. In addition, the response are presented by the percentage of fatness. These curves are graphed in Figure 4. Moreover, as measures of proximity, we focus on the family of semi-metrics

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

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.

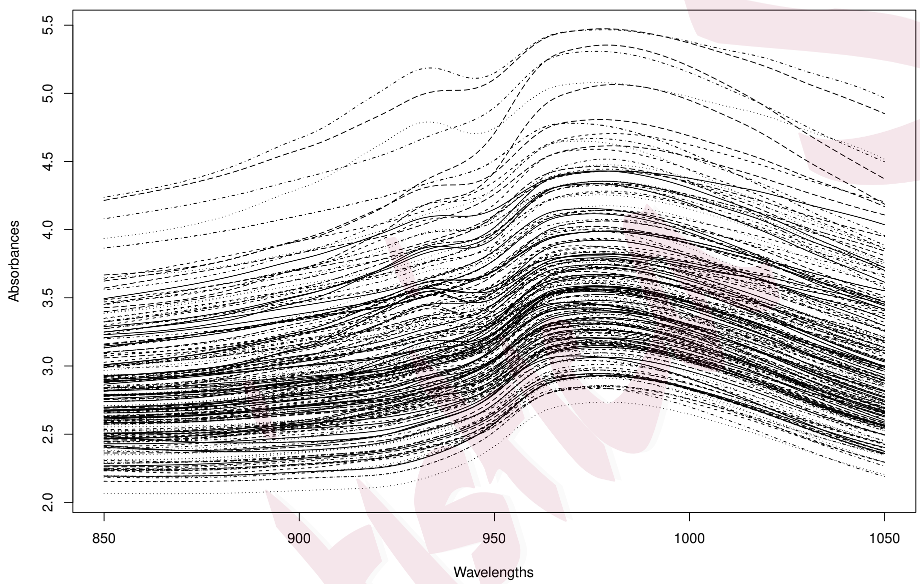

Figure 4: Spectrometric curves data.

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

Our sample of 215 pairs $(\mathcal{X}_i, Y_i)$ will be decomposed into a learning

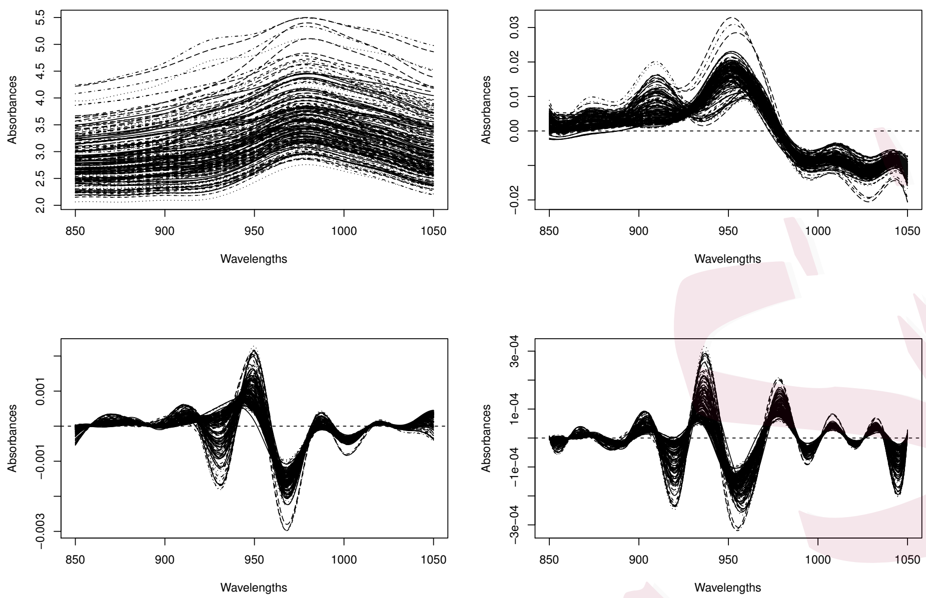

Figure 5: Shape of the dervatives of the Spectrometric Curves, $m = 0$ (in the top left panel), the first derivative (in the top right panel), the second derivative (in the down left panel) and the third derivative (in the down right panel).

sample ($\mathcal{L}$) of size 160 on which the various statistical methods will be constructed and a second sample ($\mathcal{T}$) of size 55 on which the predictive performances of these methods will be tested. We will 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 with either the recursive estimator ($MSPE$ (`Recursive`) $= 3.831634$) or the estimator non-recursive ($MSPE$ (`Non-recursive`) $= 1.211091$). The non-recursive estimator gives a smaller $MSPE$ compared to the recursive one. Performing the two methods, the running time using the recursive regression estimators (1.4) was roughly 63s on the author's workstation, while the running time using the non-recursive regression estimator (1.5) was roughly 176s on the author's workstation. Moreover, we plot in Figure 6 the predicted values obtained using the two methods as function of the true one for the 55 spectra in our testing sample.

4. Conclusion

This paper propose an automatic selection of the bandwidth of the recursive kernel density estimators for length-biased data. The proposed estimators asymptotically follows normal distribution. The estimators are compared to the nonrecursive Ferraty and Vieu's regression estimator for functional data. We showed that, using some selected bandwidth and some particularly stepsizes, the proposed recursive estimators will be very competitive to the non-recursive one. The simulation study confirms the nice features of our proposed recursive estimators and statisfactory improvement in the `CPU` time in comparison to the non-recursive estimator.

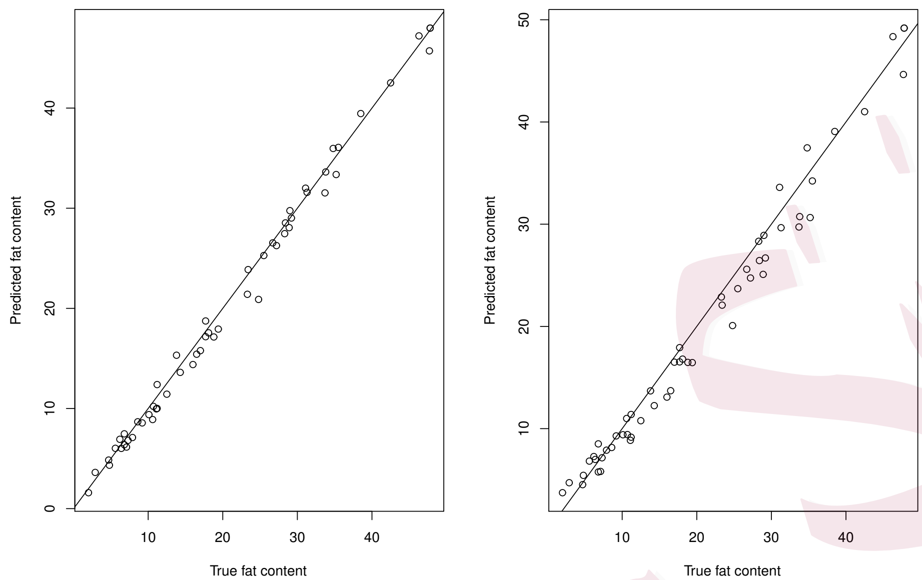

Figure 6: Spectrometric data: Predicted values on the testing sample using the non-recursive estimator (in the left panel) and using the recursive estimator (in the right panel).

In conclusion, the proposed method allowed us to obtain quite competitive estimaor to the non-recursive one proposed by Ferraty and Vieu (2002). Moreover, we plan to make an extensions of this work by proposing other bandwidth selection methods.

A. Proofs

Throughout this section we use the following notation:

$$\begin{aligned}\Pi_n &= \prod_{j=1}^{n}\left(1-\gamma_j\right),\\ a_n\left(\chi\right) &= \Pi_n \sum_{k=1}^{n} \Pi_k^{-1}\gamma_k h_k^{-1} K\left(\frac{\|\chi-\mathcal{X}_k\|}{h_k}\right) Y_k,\\ f_n\left(\chi\right) &= \Pi_n \sum_{k=1}^{n} \Pi_k^{-1}\gamma_k h_k^{-1} K\left(\frac{\|\chi-\mathcal{X}_k\|}{h_k}\right),\end{aligned} \tag{A.16}$$

Let us first state the following technical lemma.

Lemma 1. *Let* $(v_n) \in \mathcal{GS}\left(v^*\right)$, $(\gamma_n) \in \mathcal{GS}\left(-\alpha\right)$, *and* $m > 0$ *such that* $m - v^*\xi > 0$ *where* ξ *is defined in* (2.7). *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.

Acknowledgements

The authors would like to thank the Editor, the associate Editor and the two referees for their very helpful comments, which led to a considerable improvement of the original version of the paper and a more sharply focused presentation.

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 Series in Statistics, 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.

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

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

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

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.