Statistics and Its Interface Volume 9 (2016) 375–388

Optimal bandwidth selection for semi-recursive kernel regression estimators

Yousri Slaoui

In this paper we propose an automatic selection of the bandwidth of the semi-recursive kernel estimators of a regression function defined by the stochastic approximation algorithm. We showed that, using the selected bandwidth and some special stepsizes, the proposed semi-recursive estimators will be very competitive to the nonrecursive one in terms of estimation error but much better in terms of computational costs. We corroborated these theoretical results through simulation study and a real dataset.

AMS 2000 subject classifications: Primary 62G08, 62L20; secondary 65D10.
Keywords and phrases: Nonparametric regression, Stochastic approximation algorithm, Smoothing, Curve fitting.

1. INTRODUCTION

In recent years, there has been a lot of interest in big data. In such a large sample data context, building a semi-recursive estimator which does not require to store all the data in memory and can be updated easily in order to deal with online data is of great interest.

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 ideas and the bootstrap. A detailed comparison of the three practical bandwidth selection can be found in Delaigle and Gijbels [3]. They concluded that chosen appropriately, plug-in and bootstrap selectors both outperform the cross-validation bandwidth, and that neither of the two can be claimed to be better in all cases. Recently, a plug-in bandwidth selection method for recursive kernel density estimators defined by stochastic approximation method have been done by Slaoui [15] and for recursive kernel distribution estimators have been done by Slaoui [16]. In this paper, we developed a specific plug-in bandwidth selection method of the semi-recursive kernel estimators of a regression function defined by stochastic approximation method.

Let $(X_1, Y_1), \ldots, (X_n, Y_n)$ be independent, identically distributed pairs of random variables with joint density function $g(x, y)$, and let f denote the probability density of X. In order to construct a stochastic algorithm for the estimation of the regression function $a : x \mapsto \mathbb{E}(Y|X = x)f(x)$ at a point x, we define an algorithm of search of the zero of the function $h : y \to a(x) - y$. Following Robbins-Monro's procedure, this algorithm is defined by setting $a_0(x) \in \mathbb{R}$, and, for all $n \geq 1$,

$$a_n(x) = a_{n-1}(x) + \beta_n W_n,$$

where $W_n(x)$ is an "observation" of the function h at the point $a_{n-1}(x)$, and the stepsize (β_n) is a sequence of positive real numbers that goes to zero. To define $W_n(x)$, we follow the approach of Révész [11, 12], Tsybakov [20] and of Mokkadem et al. [8, 9] and introduces a kernel K (that is, a function satisfying $\int_{\mathbb{R}} K(x)dx = 1$), and a bandwidth (h_n) (that is, a sequence of positive real numbers that goes to zero), and sets $W_n(x) = h_n^{-1} Y_n K\left(h_n^{-1}(x - X_n)\right) - a_{n-1}(x)$. Then, the estimator a_n to recursively estimate the function a at the point x can be written as

$$\begin{aligned} a_n(x) &= (1 - \beta_n) a_{n-1}(x) \\ &\quad + \beta_n h_n^{-1} Y_n K\left(h_n^{-1}[x - X_n]\right). \end{aligned} \tag{1}$$

This estimator was proposed by Slaoui [19] to estimate recursively the regression function with a fixed design setting. The recursive property (1) is particularly useful in large sample size since a_n can be easily updated with each additional observation.

Let us underline that, we consider $a_0(x) = 0$ and we let $Q_n = \prod_{j=1}^n (1 - \beta_j)$, then it follows from (1) that, one can estimate a recursively at the point x by

$$a_n(x) = Q_n \sum_{k=1}^n Q_k^{-1} \beta_k h_k^{-1} Y_k K\left(\frac{x - X_k}{h_k}\right).$$

Moreover, we use the estimator introduced in Mokkadem et al. [8] to estimate recursively the density f at the point x

$$\begin{aligned} f_n(x) &= (1 - \gamma_n) f_{n-1}(x) \\ &\quad + \gamma_n h_n^{-1} K\left(h_n^{-1}[x - X_n]\right), \end{aligned} \tag{2}$$

where the stepsize (γ_n) is a sequence of positive real numbers that goes to zero. Let us underline that we consider $f_0(x) = 0$, and we let $\Pi_n = \prod_{j=1}^n (1 - \gamma_j)$, then it follows from (2) that, one can estimate f recursively at the point x by

$$f_n(x) = \Pi_n \sum_{k=1}^{n} \Pi_k^{-1} \gamma_k h_k^{-1} K\left(\frac{x - X_k}{h_k}\right).$$

Then, we consider the semi-recursive estimator for the regression function r at the point x

$$(3) \qquad r_n(x) = \begin{cases} \frac{a_n(x)}{f_n(x)} & \text{if} \quad f_n(x) \neq 0, \\ 0 & \text{otherwise.} \end{cases}$$

Moreover, we show that the optimal bandwidth which minimize the $\mathbb{E}\int_{\mathbb{R}} [r_n(x) - r(x)]^2 dx$ of r_n depends on the choice of the stepsizes (γ_n) and (β_n); we show in particular that under some conditions of regularity of r and using the stepsizes $(\gamma_n, \beta_n) = (n^{-1}, n^{-1})$, the bandwidth (h_n) must equal

$$\left(\left(\frac{3}{10}\right)^{1/5} \left\{\frac{\int_{\mathbb{R}} Var\left[Y^2|X=x\right] f^{-1}(x)\, dx}{\int_{\mathbb{R}} \left(a^{(2)}(x) - r(x) f^{(2)}(x)\right)^2 f^{-2}(x)\, dx}\right\}^{1/5} \times \left\{\frac{\int_{\mathbb{R}} K^2(z)\, dz}{\left(\int_{\mathbb{R}} z^2 K(z)\, dz\right)^2}\right\}^{1/5} n^{-1/5}\right).$$

The first aim of this paper is to propose an automatic selection of such bandwidth through a plug-in method, and the second aim is to give the conditions under which the semi-recursive estimator r_n will be approximately similar to the nonrecursive kernel regression estimators introduced by Nadaraya [10] and Watson [21], and defined as

$$(4) \qquad \widetilde{r}_n(x) = \begin{cases} \frac{\widetilde{a}_n(x)}{\widetilde{f}_n(x)} & \text{if} \quad \widetilde{f}_n(x) \neq 0 \\ 0 & \text{otherwise,} \end{cases}$$

with

$$\widetilde{a}_n(x) = \frac{1}{nh_n} \sum_{i=1}^{n} Y_i K\left(\frac{x - X_i}{h_n}\right)$$

and

$$\widetilde{f}_n(x) = \frac{1}{nh_n} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h_n}\right).$$

The applications results given in Section 3 corroborate these theoretical results. 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

We define the following class of regularly varying sequences.

Definition 2.1. *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*

$$(5) \qquad \lim_{n\to+\infty} n\left[1 - \frac{v_{n-1}}{v_n}\right] = \gamma.$$

Condition (5) was introduced by Galambos and Seneta [4] to define regularly varying sequences (see also Bojanic and Seneta [2]) and by Mokkadem and Pelletier [7] 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 (3). The assumptions to which we shall refer are the following:

(A1) $K : \mathbb{R} \to \mathbb{R}$ is a continuous, bounded function satisfying $\int_{\mathbb{R}} K(z)\, dz = 1$, and, $\int_{\mathbb{R}} zK(z) = 0$ and $\int_{\mathbb{R}} z^2 K(z) < \infty$.

(A2) *i)* $(\beta_n) \in \mathcal{GS}(-\beta)$ with $\beta \in]1/2, 1]$.
ii) $(h_n) \in \mathcal{GS}(-a)$ with $a \in]0, 1[$.
iii) $\lim_{n\to\infty} (n\beta_n) \in]\min\{2a, (\beta - a)/2\}, \infty]$.

(A3) *i)* $g(s,t)$ is twice continuously differentiable with respect to s.
ii) For $q \in \{0, 1, 2\}$, $s \mapsto \int_{\mathbb{R}} t^q g(s,t)\, dt$ is a bounded function continuous at $s = x$.
For $q \in [2, 3]$, $s \mapsto \int_{\mathbb{R}} |t|^q g(s,t)\, dt$ is a bounded function.
iii) For $q \in \{0, 1\}$, $\int_{\mathbb{R}} |t|^q \left|\frac{\partial g}{\partial x}(x,t)\right| dt < \infty$, and $s \mapsto \int_{\mathbb{R}} t^q \frac{\partial^2 g}{\partial s^2}(s,t)\, dt$ is a bounded function continuous at $s = x$.

Assumption $(A2)(iii)$ on the limit of $(n\beta_n)$ as n goes to infinity is standard in the framework of stochastic approximation algorithms. It implies in particular that the limit of $\left([n\beta_n]^{-1}\right)$ is finite. For simplicity, we introduce the following notations:

$$\begin{aligned} (6) \qquad \xi &= \lim_{n\to\infty} (n\beta_n)^{-1}, \\ R(K) &= \int_{\mathbb{R}} K^2(z)\, dz, \\ \mu_j(K) &= \int_{\mathbb{R}} z^j K(z)\, dz, \\ \Theta(K) &= R(K)^{4/5} \mu_2(K)^{2/5}, \\ I_1 &= \int_{\mathbb{R}} \left(a^{(2)}(x)\right)^2 f(x)\, dx, \\ I_2 &= \int_{\mathbb{R}} a^{(2)}(x) f^{(2)}(x) r(x) f(x)\, dx, \\ I_3 &= \int_{\mathbb{R}} \left(f^{(2)}(x)\right)^2 r^2(x) f(x)\, dx, \\ I_4 &= \int_{\mathbb{R}} \mathbb{E}\left[Y^2|X=x\right] f^2(x)\, dx, \end{aligned}$$

$$I_5 = \int_{\mathbb{R}} r^2(x) f^2(x)\, dx,$$

where $L^{(2)}(x)$ is the second derivative of the function L at a point x. In this section, we explicit the choice of (h_n) through a plug-in method, which minimize the Mean Weighted Integrated Squared Error $MWISE$ of the semi-recursive estimators (3), in order to provide a comparison with the nonrecursive estimator (4). Moreover, it was shown in Mokkadem et al. [8] and considered in Slaoui [14] that to minimize the Mean Integrated Squared Error $MISE$ of f_n ($MISE[f_n] = \mathbb{E}\int_{\mathbb{R}}[f_n(x) - f(x)]^2 dx$), the stepsize (γ_n) must be chosen in $\mathcal{GS}(-1)$ and must satisfy $\lim_{n\to\infty} n\gamma_n = 1$. We consider here the case $(\gamma_n) = (n^{-1})$. Our first result is the following proposition, which gives the bias and the variance of r_n in the special case of $(\gamma_n) = (n^{-1})$.

Proposition 2.1 (Bias and variance of r_n). *Let Assumptions $(A1)-(A3)$ hold, and suppose that the stepsize $(\gamma_n) = (n^{-1})$*

1. *If $a \in]0, \beta/5]$, then*

$$\mathbb{E}[r_n(x)] - r(x) = \frac{1}{2f(x)}\left(\frac{a^{(2)}(x)}{(1-2a\xi)} - \frac{r(x) f^{(2)}(x)}{(1-2a)}\right) h_n^2 \mu_2(K) + o\left(h_n^2\right). \tag{7}$$

If $a \in]\beta/5, 1[$, then

$$\mathbb{E}[r_n(x)] - r(x) = o\left(\sqrt{\beta_n h_n^{-1}}\right). \tag{8}$$

2. *If $a \in [\beta/5, 1[$, then*

$$Var[r_n(x)] = \frac{\beta_n}{h_n}\left\{\frac{\mathbb{E}\left[Y^2|X=x\right]}{(2-(\beta-a)\xi) f(x)} - \left(\frac{2\xi}{1+a\xi} - \frac{\xi}{1+a}\right)\frac{r^2(x)}{f(x)}\right\} R(K) + o\left(\frac{\beta_n}{h_n}\right). \tag{9}$$

If $a \in]0, \beta/5[$, then

$$Var[r_n(x)] = o\left(h_n^4\right). \tag{10}$$

3. *If $\lim_{n\to\infty}(n\beta_n) > \max\{2a, (a-\beta)/2\}$, then (7) and (9) hold simultaneously.*

The bias and the variance of the estimator r_n defined by the stochastic approximation algorithm (3) then heavily depend on the choice of the stepsizes (γ_n) and (β_n).
Let us first state the following theorem, which gives the weak convergence rate of the estimator r_n defined in (3) in the case of $(\gamma_n) = (n^{-1})$.

Theorem 2.1 (Weak pointwise convergence rate). *Let Assumptions $(A1)-(A3)$ hold, and suppose that $(\gamma_n) = (n^{-1})$.*

1. *If there exists $c \geq 0$ such that $\beta_n^{-1} h_n^5 \to c$, then*

$$\sqrt{\beta_n^{-1} h_n}\left(r_n(x) - r(x)\right) \stackrel{\mathcal{D}}{\to} \mathcal{N}\left(\sqrt{c} B_{a,\xi}^{(1)}(x), V_{a,\xi,\beta}^{(1)}(x)\right),$$

where

$$B_{a,\xi}^{(1)}(x) = \frac{1}{2f(x)}\left(\frac{a^{(2)}(x)}{(1-2a\xi)} - \frac{r(x) f^{(2)}(x)}{(1-2a)}\right)\mu_2(K),$$

$$V_{a,\xi,\beta}^{(1)}(x) = \left\{\frac{\mathbb{E}\left[Y^2|X=x\right]}{(2-(\beta-a)\xi) f(x)} - \left(\frac{2\xi}{1+a\xi} - \frac{\xi}{1+a}\right)\frac{r^2(x)}{f(x)}\right\} R(K).$$

2. *If $nh_n^5 \to \infty$, then*

$$\frac{1}{h_n^2}\left(r_n(x) - r(x)\right) \stackrel{\mathbb{P}}{\to} B_{a,\xi}^{(1)}(x),$$

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

The following corollary gives the weak convergence rate of r_n in the two special cases; $(\gamma_n, \beta_n) = (n^{-1}, n^{-1})$ and $(\gamma_n, \beta_n) = (n^{-1}, (1-a) n^{-1})$ respectively.

Corollary 2.1 (Weak pointwise convergence rate). *Let Assumptions $(A1)-(A3)$ hold.*

1. *If we suppose that the stepsizes $(\gamma_n, \beta_n) = (n^{-1}, n^{-1})$ and if there exists $c \geq 0$ such that $nh_n^5 \to c$, then*

$$\sqrt{nh_n}\left(r_n(x) - r(x)\right) \stackrel{\mathcal{D}}{\to} \mathcal{N}\left(\sqrt{c} B_{a,1}^{(1)}(x), V_{a,1,1}^{(1)}(x)\right).$$

2. *If we suppose that the stepsizes $(\gamma_n, \beta_n) = (n^{-1}, (1-a) n^{-1})$, and if there exists $c \geq 0$ such that $nh_n^5 \to c$, then*

$$\sqrt{nh_n}\left(r_n(x) - r(x)\right) \stackrel{\mathcal{D}}{\to} \mathcal{N}\left(\sqrt{c} B_{a,(1-a)^{-1}}^{(1)}(x), V_{a,(1-a)^{-1},1}^{(1)}(x)\right).$$

In order to measure the quality of our semi-recursive estimator (3) in the case when the stepsize (γ_n) is chosen to minimize the $MISE$ of f_n, we use the following quantity,

$$\begin{aligned} MWISE[r_n] &= \mathbb{E}\int_{\mathbb{R}}[r_n(x) - r(x)]^2 f^3(x)\, dx \\ &= \int_{\mathbb{R}}\left(\mathbb{E}(r_n(x)) - r(x)\right)^2 f^3(x)\, dx \\ &\quad + \int_{\mathbb{R}} Var(r_n(x)) f^3(x)\, dx. \end{aligned}$$

The following proposition gives the $MWISE$ of the semi-recursive estimators defined in (3) in the case when (γ_n) is chosen to minimize the $MISE$ of f_n.

Proposition 2.2 (*MWISE of* r_n). *Let Assumptions* $(A1)-(A3)$ *hold, and suppose that* $(\gamma_n)=\left(n^{-1}\right)$.

1. *If* $a\in]0,\beta/5[$, *then*

$$MWISE[r_n] = \frac{1}{4}\left(\frac{I_1}{(1-2a\xi)^2}+\frac{I_3}{(1-2a)^2} -2\frac{I_2}{(1-2a)(1-2a\xi)}\right)h_n^4\mu_2^2(K) +o\left(h_n^4\right).$$

2. *If* $a=\beta/5$, *then*

$$MWISE[r_n] = \frac{\beta_n}{h_n}\left(\frac{I_4}{(2-(\beta-a)\xi)} -\left(\frac{2\xi}{1+a\xi}-\frac{\xi}{1+a}\right)I_5\right)R(K) +\frac{1}{4}\left(\frac{I_1}{(1-2a\xi)^2}+\frac{I_3}{(1-2a)^2} -2\frac{I_2}{(1-2a)(1-2a\xi)}\right)h_n^4\mu_2^2(K) +o\left(h_n^4\right).$$

3. *If* $a\in]\beta/5,1[$, *then*

$$MWISE[r_n] = \frac{\beta_n}{h_n}\left(\frac{I_4}{(2-(\beta-a)\xi)} -\left(\frac{2\xi}{1+a\xi}-\frac{\xi}{1+a}\right)I_5\right)R(K) +o\left(\frac{\beta_n}{h_n}\right).$$

The following corollary indicates that the bandwidth which minimizes the $MWISE$ of r_n depends on the stepsize (β_n) and then the corresponding $MWISE$ depends also on the stepsize (β_n).

Corollary 2.2. *Let Assumptions* $(A1)-(A3)$ *hold, and suppose that* $(\gamma_n)=\left(n^{-1}\right)$. *To minimize the* $MWISE$ *of* r_n, *the stepsize* (β_n) *must be chosen in* $\mathcal{GS}(-1)$, *the bandwidth* (h_n) *must equal*

$$\left(\left\{\frac{\frac{I_4}{(2-(\beta-a)\xi)}-\left(\frac{2\xi}{1+a\xi}-\frac{\xi}{1+a}\right)I_5}{\frac{I_1}{(1-2a\xi)^2}+\frac{I_3}{(1-2a)^2}-2\frac{I_2}{(1-2a)(1-2a\xi)}}\right\}^{1/5} \times\left\{\frac{R(K)}{\mu_2^2(K)}\right\}^{1/5}\beta_n^{1/5}\right).$$

Then, we have

$$MWISE[r_n] = \frac{5}{4}\left(\frac{I_4}{(2-(\beta-a)\xi)} -\left(\frac{2\xi}{1+a\xi}-\frac{\xi}{1+a}\right)I_5\right)^{4/5} \times\left(\frac{I_1}{(1-2a\xi)^2}+\frac{I_3}{(1-2a)^2} -2\frac{I_2}{(1-2a)(1-2a\xi)}\right)^{1/5}\Theta(K)\beta_n^{4/5} +o\left(\beta_n^{4/5}\right).$$

The following corollary shows that, for a special choice of the stepsize $(\beta_n)=\left(\beta_0n^{-1}\right)$, which fulfilleds that $\lim_{n\to\infty}n\beta_n=\beta_0$ and that $(\beta_n)\in\mathcal{GS}(-1)$, the optimal value for h_n depends on β_0 and then the corresponding $MWISE$ depend on β_0.

Corollary 2.3. *Let Assumptions* $(A1)-(A3)$ *hold, and suppose that* $(\gamma_n)=\left(n^{-1}\right)$. *To minimize the* $MWISE$ *of* r_n, *the stepsize* (β_n) *must be chosen in* $\mathcal{GS}(-1)$, $\lim_{n\to\infty}n\beta_n=\beta_0$, *the bandwidth* (h_n) *must equal*

$$\left(\left(\frac{\beta_0-2/5}{2}\right)^{1/5}\left(\frac{I_4-\frac{(7\beta_0-1)(\beta_0-2/5)}{3\beta_0^2(\beta_0+1/5)}I_5}{I_1+\frac{25}{9}\left(\frac{\beta_0-2/5}{\beta_0}\right)^2I_3-\frac{10}{3}\left(\frac{\beta_0-2/5}{\beta_0}\right)I_2}\right)^{1/5}\left\{\frac{R(K)}{\mu_2^2(K)}\right\}^{1/5}n^{-1/5}\right), \tag{11}$$

and we then have

$$MWISE[r_n] = \frac{5}{4}\frac{1}{2^{4/5}}\frac{\beta_0^2}{(\beta_0-2/5)^{6/5}}\left(I_4 -\frac{(7\beta_0-1)(\beta_0-2/5)}{3\beta_0^2(\beta_0+1/5)}I_5\right)^{4/5} \times\left(I_1+\frac{25}{9}\left(\frac{\beta_0-2/5}{\beta_0}\right)^2I_3 -\frac{10}{3}\left(\frac{\beta_0-2/5}{\beta_0}\right)I_2\right)^{1/5}\Theta(K)n^{-4/5} +o\left(n^{-4/5}\right). \tag{12}$$

Moreover, the minimum of $\beta_0^2(\beta_0-2/5)^{-6/5}$ *is reached at* $\beta_0=1$, *then the bandwidth* (h_n) *must equal*

$$\left(\left(\frac{3}{10}\right)^{1/5}\left(\frac{I_4-I_5}{I_1+I_3-2I_2}\right)^{1/5}\left\{\frac{R(K)}{\mu_2^2(K)}\right\}^{1/5}n^{-1/5}\right), \tag{13}$$

and we then have

$$MWISE[r_n] = \frac{5}{4}\frac{1}{2^{4/5}}\left(\frac{5}{3}\right)^{6/5}(I_4-I_5)^{4/5}$$

$$\times\left(I_1+I_3-2I_2\right)^{1/5}\Theta\left(K\right)n^{-4/5}$$
$$+o\left(n^{-4/5}\right). \tag{14}$$

In order to estimate the optimal bandwidth (13), we must estimate I_1, I_2, I_3, I_4 and I_5. We followed the approach of Altman and Leger [1] and Slaoui [15, 16], which is called the plug-in estimate, and we use the following kernel estimators of I_1, I_2, I_3, I_4 and I_5:

$$\widehat{I}_1 = \frac{Q_n^2}{n}\sum_{\substack{i,j,k=1\\ j\neq k}}^{n} Q_j^{-1}Q_k^{-1}\beta_j\beta_k b_j^{-3}b_k^{-3}$$
$$\times K_b^{(2)}\left(\frac{X_i-X_j}{b_j}\right)K_b^{(2)}\left(\frac{X_i-X_k}{b_k}\right)Y_jY_k, \tag{15}$$

$$\widehat{I}_2 = \frac{\Pi_nQ_n}{n}\sum_{\substack{i,j,k=1\\ j\neq k}}^{n} \Pi_k^{-1}Q_j^{-1}\gamma_k\beta_j b_k^{-3}b_j^{-3}$$
$$\times K_b^{(2)}\left(\frac{X_i-X_k}{b_k}\right)K_b^{(2)}\left(\frac{X_i-X_j}{b_j}\right)Y_iY_j, \tag{16}$$

$$\widehat{I}_3 = \frac{\Pi_n^2}{n}\sum_{\substack{i,j,k,l=1\\ j\neq k\neq l}}^{n} \Pi_j^{-1}\Pi_k^{-1}\gamma_j\gamma_k b_j^{-3}b_k^{-3}$$
$$K_b^{(2)}\left(\frac{X_i-X_j}{b_j}\right)K_b^{(2)}\left(\frac{X_i-X_k}{b_k}\right)Y_iY_l, \tag{17}$$

$$\widehat{I}_4 = \frac{\Pi_n}{n}\sum_{\substack{i,k=1\\ i\neq k}}^{n} \Pi_k^{-1}\gamma_k b_k^{-1}$$
$$K_b\left(\frac{X_i-X_k}{b_k}\right)Y_i^2, \tag{18}$$

$$\widehat{I}_5 = \frac{Q_n}{n}\sum_{\substack{i,k=1\\ i\neq k}}^{n} Q_k^{-1}\beta_k b_k^{-1}$$
$$K_b\left(\frac{X_i-X_k}{b_k}\right)Y_iY_k, \tag{19}$$

where K_b is a kernel and b_n is the associated bandwidth. In practice, we take

$$b_n = n^{-\beta}\min\left\{\widehat{s},\frac{Q_3-Q_1}{1.349}\right\},\quad \beta\in\,]0,1[\tag{20}$$

(see Silverman [13]) where $\widehat{s}$ the sample standard deviation, and Q_1, Q_3 denoting the first and third quartiles, respectively.
We followed the same steps as in Slaoui [15] and we showed that in order to minimize the $MISE$ of $\widehat{I}_1$ respectively of $\widehat{I}_2$, $\widehat{I}_3$, $\widehat{I}_4$ and $\widehat{I}_5$, the pilot bandwidth (b_n) must belong to $\mathcal{GS}(-3/14)$, respectively to $\mathcal{GS}(-3/14)$, $\mathcal{GS}(-3/14)$, $\mathcal{GS}(-2/5)$ and $\mathcal{GS}(-2/5)$.

Finally, the plug-in estimator of the bandwidth (h_n) using the semi-recursive estimators defined in (3) with the stepsizes $(\gamma_n,\beta_n)=\left(n^{-1},n^{-1}\right)$.

$$\left(\left(\frac{3}{10}\right)^{1/5}\left(\frac{\widehat{I}_4-\widehat{I}_5}{\widehat{I}_1+\widehat{I}_3-2\widehat{I}_2}\right)^{1/5}\right.$$
$$\left.\left\{\frac{R(K)}{\mu_2^2(K)}\right\}^{1/5}n^{-1/5}\right), \tag{21}$$

$$\widehat{MWISE}[r_n] = \frac{5}{4}\frac{1}{2^{4/5}}\left(\frac{5}{3}\right)^{6/5}\left(\widehat{I}_4-\widehat{I}_5\right)^{4/5}$$
$$\times\left(\widehat{I}_1+\widehat{I}_3-2\widehat{I}_2\right)^{1/5}\Theta(K)n^{-4/5}$$
$$+o\left(n^{-4/5}\right).$$

Let us now consider the stepsize $(\beta_n)=\left((1-a)n^{-1}\right)$, the case which minimizes the variance of $a_n(x)$ combined with the stepsize $(\gamma_n)=\left(n^{-1}\right)$, the case which minimizes the $MISE$ of f_n, it follows from (12), that

$$MWISE[r_n] = 5^{1/5}\left(I_4-\frac{23}{24}I_5\right)^{4/5}$$
$$\times\left(I_1+\frac{25}{36}I_3-\frac{5}{3}I_2\right)^{1/5}\Theta(K)n^{-4/5}$$
$$+o\left(n^{-4/5}\right), \tag{22}$$

and from (11), that the plug-in estimator of the bandwidth (h_n) using the semi-recursive estimators defined in (3) is given by

$$\left(\left(\frac{1}{5}\right)^{1/5}\left(\frac{\widehat{I}_4-\frac{23}{24}\widehat{I}_5}{\widehat{I}_1+\frac{25}{36}\widehat{I}_3-\frac{5}{3}\widehat{I}_2}\right)^{1/5}\right.$$
$$\left.\left\{\frac{R(K)}{\mu_2^2(K)}\right\}^{1/5}n^{-1/5}\right), \tag{23}$$

and it follows from (12), that the plug-in $MWISE$ of the proposed estimator (3) using the stepsizes $(\gamma_n,\beta_n)=\left(n^{-1},(1-a)n^{-1}\right)$ is given by

$$\widehat{MWISE}[r_n] = 5^{1/5}\left(\widehat{I}_4-\frac{23}{24}\widehat{I}_5\right)^{4/5}$$
$$\times\left(\widehat{I}_1+\frac{25}{36}\widehat{I}_3-\frac{5}{3}\widehat{I}_2\right)^{1/5}\Theta(K)n^{-4/5}$$
$$+o\left(n^{-4/5}\right).$$

Let us now provide the case when the stepsize (γ_n) is chosen to minimize the variance of f_n. It was shown in Mokkadem et al. [8] and considered in Slaoui [14] that to minimize the variance of f_n, the stepsize (γ_n) must be chosen in $\mathcal{GS}(-1)$

and must satisfy $\lim_{n\to\infty} n\gamma_n = 1-a$. We consider here the case $(\gamma_n) = \left((1-a)\,n^{-1}\right)$. Our first result is the following proposition, which gives the bias and the variance of r_n in the special case of $(\gamma_n) = \left((1-a)\,n^{-1}\right)$.

Proposition 2.3 (Bias and variance of r_n)**.** *Let Assumptions (A1) − (A3) hold, and suppose that* $(\gamma_n) = \left((1-a)\,n^{-1}\right)$.

1. *If* $a \in]0, \beta/5]$, *then*

$$\mathbb{E}\left[r_n(x)\right] - r(x) = \frac{1}{2f(x)}\left(\frac{a^{(2)}(x)}{(1-2a\xi)} - \frac{1-a}{(1-3a)}r(x)\,f^{(2)}(x)\right)h_n^2\mu_2(K) + o\left(h_n^2\right). \tag{24}$$

If $a \in]\beta/5, 1[$, *then*

$$\mathbb{E}\left[r_n(x)\right] - r(x) = o\left(\sqrt{\beta_n h_n^{-1}}\right). \tag{25}$$

2. *If* $a \in [\beta/5, 1[$, *then*

$$Var\left[r_n(x)\right] = \frac{\beta_n}{h_n}\left\{\frac{\mathbb{E}\left[Y^2|X=x\right]}{(2-(\beta-a)\,\xi)\,f(x)} - (1-a)\,\xi\frac{r^2(x)}{f(x)}\right\}R(K) + o\left(\frac{\beta_n}{h_n}\right). \tag{26}$$

If $a \in]0, \beta/5[$, *then*

$$Var\left[r_n(x)\right] = o\left(h_n^4\right). \tag{27}$$

3. *If* $\lim_{n\to\infty}(n\beta_n) > \max\{2a, (a-\beta)/2\}$, *then (24) and (26) hold simultaneously.*

The bias and the variance of the estimator r_n defined by the stochastic approximation algorithm (3) then heavily depend on the choice of the stepsizes (γ_n) and (β_n).
Let us first state the following theorem, which gives the weak convergence rate of the estimator r_n defined in (3).

Theorem 2.2 (Weak pointwise convergence rate)**.** *Let Assumptions (A1) − (A3) hold, and suppose that* $(\gamma_n) = \left((1-a)\,n^{-1}\right)$.

1. *If there exists* $c \geq 0$ *such that* $\beta_n^{-1}h_n^5 \to c$, *then*

$$\sqrt{\beta_n^{-1}h_n}\left(r_n(x) - r(x)\right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(\sqrt{c}B_{a,\xi}^{(2)}(x), V_{a,\xi,\beta}^{(2)}(x)\right),$$

where

$$B_{a,\xi}^{(2)}(x) = \frac{1}{2f(x)}\left(\frac{a^{(2)}(x)}{(1-2a\xi)} - \frac{(1-a)}{(1-3a)}r(x)\,f^{(2)}(x)\right)\mu_2(K),$$

$$V_{a,\xi,\beta}^{(2)}(x) = \left\{\frac{\mathbb{E}\left[Y^2|X=x\right]}{(2-(\beta-a)\,\xi)\,f(x)} - (1-a)\,\xi\frac{r^2(x)}{f(x)}\right\}R(K).$$

2. *If* $nh_n^5 \to \infty$, *then*

$$\frac{1}{h_n^2}\left(r_n(x) - r(x)\right) \xrightarrow{\mathbb{P}} B_{a,\xi}^{(2)}(x).$$

The following corollary gives the weak convergence rate of r_n in the two special cases; $(\gamma_n, \beta_n) = \left((1-a)\,n^{-1}, n^{-1}\right)$ and $(\gamma_n, \beta_n) = \left((1-a)\,n^{-1}, (1-a)\,n^{-1}\right)$ respectively.

Corollary 2.4 (Weak pointwise convergence rate)**.** *Let Assumptions (A1) − (A3) hold.*

1. *If we suppose that the stepsizes* $(\gamma_n, \beta_n) = \left((1-a)\,n^{-1}, n^{-1}\right)$, *and if there exists* $c \geq 0$ *such that* $nh_n^5 \to c$, *then*

$$\sqrt{nh_n}\left(r_n(x) - r(x)\right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(\sqrt{c}B_{a,1}^{(2)}(x), V_{a,1,1}^{(2)}(x)\right).$$

2. *If we suppose that the stepsizes* $(\gamma_n, \beta_n) = \left((1-a)\,n^{-1}, (1-a)\,n^{-1}\right)$, *and if there exists* $c \geq 0$ *such that* $nh_n^5 \to c$, *then*

$$\sqrt{nh_n}\left(r_n(x) - r(x)\right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(\sqrt{c}B_{a,(1-a)^{-1}}^{(2)}(x), V_{a,(1-a)^{-1},1}^{(2)}(x)\right).$$

The following proposition gives the $MWISE$ of r_n in the case when (γ_n) is chosen to minimize the variance of f_n.

Proposition 2.4 ($MWISE$ of r_n)**.** *Let Assumptions (A1) − (A3) hold, and suppose that* $(\gamma_n) = \left((1-a)\,n^{-1}\right)$.

1. *If* $a \in]0, \beta/5[$, *then*

$$MWISE\left[r_n\right] = \frac{1}{4}\left(\frac{I_1}{(1-2a\xi)^2} + \frac{(1-a)^2}{(1-3a)^2}I_3 - 2\frac{(1-a)}{(1-3a)(1-2a\xi)}I_2\right)h_n^4\mu_2^2(K) + o\left(h_n^4\right).$$

2. *If* $a = \beta/5$, *then*

$$MWISE\left[r_n\right] = \frac{\beta_n}{h_n}\left(\frac{I_4}{(2-(\beta-a)\,\xi)} - (1-a)\,\xi I_5\right)R(K) + \frac{1}{4}\left(\frac{I_1}{(1-2a\xi)^2} + \frac{(1-a)^2}{(1-3a)^2}I_3\right.$$

$$- 2\frac{(1-a)}{(1-3a)(1-2a\xi)}I_2\Bigg) h_n^4\mu_2^2(K) + o\left(h_n^4\right).$$

3. *If $a\in]\beta/5,1[$, then*

$$MWISE[r_n] = \frac{\beta_n}{h_n}\left(\frac{I_4}{(2-(\beta-a)\xi)} - (1-a)\xi I_5\right)R(K) + o\left(\frac{\beta_n}{h_n}\right).$$

The following corollary ensures that the bandwidth which minimize the $MWISE$ depend on the stepsize (β_n) and then the corresponding $MWISE$ depend also on the stepsize (β_n).

Corollary 2.5. *Let Assumptions $(A1)-(A3)$ hold, and suppose that $(\gamma_n) = \left((1-a)n^{-1}\right)$. To minimize the MWISE of r_n, the stepsize (β_n) must be chosen in $\mathcal{GS}(-1)$, the bandwidth (h_n) must equal*

$$\left(\left\{\frac{\frac{I_4}{(2-(\beta-a)\xi)} - (1-a)\xi I_5}{\frac{I_1}{(1-2a\xi)^2} + \frac{(1-a)^2}{(1-3a)^2}I_3 - 2\frac{(1-a)}{(1-3a)(1-2a\xi)}I_2}\right\}^{1/5}\left\{\frac{R(K)}{\mu_2^2(K)}\right\}^{1/5}\beta_n^{1/5}\right).$$

Then, we have

$$\begin{aligned} MWISE[r_n] &= \frac{5}{4}\left(\frac{I_4}{(2-(\beta-a)\xi)} - (1-a)\xi I_5\right)^{4/5} \\ &\times\left(\frac{I_1}{(1-2a\xi)^2} + \frac{(1-a)^2}{(1-3a)^2}I_3 - 2\frac{(1-a)}{(1-3a)(1-2a\xi)}I_2\right)^{1/5}\Theta(K)\beta_n^{4/5} \\ &+ o\left(\beta_n^{4/5}\right). \end{aligned}$$

The following corollary shows that, for a special choice of the stepsize $(\beta_n) = \left(\beta_0 n^{-1}\right)$, which fulfilled that $\lim_{n\to\infty} n\beta_n = \beta_0$ and that $(\beta_n)\in\mathcal{GS}(-1)$, the optimal value for h_n depend on β_0 and then the corresponding $MWISE$ depend on β_0.

Corollary 2.6. *Let Assumptions $(A1)-(A3)$ hold, and suppose that $(\gamma_n) = \left((1-a)n^{-1}\right)$. To minimize the MWISE of r_n, the stepsize (β_n) must be chosen in $\mathcal{GS}(-1)$, $\lim_{n\to\infty} n\beta_n = \beta_0$, the bandwidth (h_n) must equal*

$$\left(\left(\frac{\beta_0-2/5}{2}\right)^{1/5}\left(\frac{I_4 - \frac{8}{5}\frac{(\beta_0-2/5)}{\beta_0^2}I_5}{I_1 + 4\left(\frac{\beta_0-2/5}{\beta_0}\right)^2 I_3 - 4\left(\frac{\beta_0-2/5}{\beta_0}\right)I_2}\right)^{1/5}\left\{\frac{R(K)}{\mu_2^2(K)}\right\}^{1/5}n^{-1/5}\right), \quad (28)$$

and we then have

$$\begin{aligned} MWISE[r_n] &= \frac{5}{4}\frac{1}{2^{4/5}}\frac{\beta_0^2}{(\beta_0-2/5)^{6/5}}\left(I_4 - \frac{8}{5}\frac{\beta_0-2/5}{\beta_0^2}I_5\right)^{4/5} \\ &\times\left(I_1 + 4\left(\frac{\beta_0-2/5}{\beta_0}\right)^2 I_3 - 4\left(\frac{\beta_0-2/5}{\beta_0}\right)I_2\right)^{1/5}\Theta(K)n^{-4/5} \\ &+ o\left(n^{-4/5}\right). \end{aligned} \quad (29)$$

Moreover, the minimum of $\beta_0^2(\beta_0-2/5)^{-6/5}$ is reached at $\beta_0 = 1$, then the bandwidth (h_n) must equal

$$\left(\left(\frac{3}{10}\right)^{1/5}\left(\frac{I_4 - \frac{24}{25}I_5}{I_1 + \frac{36}{25}I_3 - \frac{12}{5}I_2}\right)^{1/5}\left\{\frac{R(K)}{\mu_2^2(K)}\right\}^{1/5}n^{-1/5}\right),$$

and we then have

$$\begin{aligned} MWISE[r_n] &= \frac{5}{4}\frac{1}{2^{4/5}}\left(\frac{5}{3}\right)^{6/5}\left(I_4 - \frac{24}{25}I_5\right)^{4/5} \\ &\times\left(I_1 + \frac{36}{25}I_3 - \frac{12}{5}I_2\right)^{1/5}\Theta(K)n^{-4/5} \\ &+ o\left(n^{-4/5}\right). \end{aligned} \quad (30)$$

In order to estimate the optimal bandwidth (13), we must estimate I_1, I_2, I_3, I_4 and I_5. We use the kernel estimators defined in (15), (16), (17), (18) and (19). We showed that in order to minimize the $MISE$ of $\widehat{I}_1$ respectively of $\widehat{I}_2$, $\widehat{I}_3$, $\widehat{I}_4$ and $\widehat{I}_5$, the pilot bandwidth (b_n) must belong to $\mathcal{GS}(-3/14)$, respectively to $\mathcal{GS}(-3/14)$, $\mathcal{GS}(-3/14)$, $\mathcal{GS}(-2/5)$ and $\mathcal{GS}(-2/5)$.
Finally, the plug-in estimator of the bandwidth (h_n) using the semi-recursive estimators defined in (3) with the stepsizes $(\gamma_n,\beta_n) = \left((1-a)n^{-1}, n^{-1}\right)$.

$$\left(\left(\frac{3}{10}\right)^{1/5}\left(\frac{\widehat{I}_4 - \frac{24}{25}\widehat{I}_5}{\widehat{I}_1 + \frac{36}{25}\widehat{I}_3 - \frac{12}{5}\widehat{I}_2}\right)^{1/5}\left\{\frac{R(K)}{\mu_2^2(K)}\right\}^{1/5}n^{-1/5}\right), \quad (31)$$

$$\begin{aligned} \widehat{MWISE}[r_n] &= \frac{5}{4}\frac{1}{2^{4/5}}\left(\frac{5}{3}\right)^{6/5}\left(\widehat{I}_4 - \frac{24}{25}\widehat{I}_5\right)^{4/5} \\ &\times\left(\widehat{I}_1 + \frac{36}{25}\widehat{I}_3 - \frac{12}{5}\widehat{I}_2\right)^{1/5}\Theta(K)n^{-4/5} \end{aligned}$$

$$+o\left(n^{-4/5}\right).$$

Let us now consider the stepsize $(\beta_n) = \left((1-a)\,n^{-1}\right)$, the case which minimizes the variance of $a_n(x)$ combined with the stepsize $(\gamma_n) = \left((1-a)\,n^{-1}\right)$, the case which minimizes the variance of f_n, it follows from (29), that

$$MWISE\left[r_n\right] = 5^{1/5}\left(I_4-I_5\right)^{4/5}\times\left(I_1+I_3-2I_2\right)^{1/5} \Theta(K)\,n^{-4/5}+o\left(n^{-4/5}\right), \tag{32}$$

and from (28), that the plug-in estimator of the bandwidth (h_n) using the semi-recursive estimators defined in (3) is given by

$$\left(\left(\frac{1}{5}\right)^{1/5}\left(\frac{\widehat{I}_4-\widehat{I}_5}{\widehat{I}_1+\widehat{I}_3-2\widehat{I}_2}\right)^{1/5} \left\{\frac{R(K)}{\mu_2^2(K)}\right\}^{1/5} n^{-1/5}\right), \tag{33}$$

and it follows from (29), that the plug-in $MWISE$ of the proposed estimator (3) using the stepsizes $(\gamma_n,\beta_n) = \left((1-a)\,n^{-1},(1-a)\,n^{-1}\right)$, is given by

$$\widehat{MWISE}\left[r_n\right] = 5^{1/5}\left(\widehat{I}_4-\widehat{I}_5\right)^{4/5}\times\left(\widehat{I}_1+\widehat{I}_3-2\widehat{I}_2\right)^{1/5} \Theta(K)\,n^{-4/5}+o\left(n^{-4/5}\right).$$

Now, let us recall that the bias and variance of Nadaraya-Watson's estimator $\widetilde{r}_n$ are given by

$$\mathbb{E}\left[\widetilde{r}_n(x)\right]-r(x) = \frac{1}{2}\left(a^{(2)}(x)-r(x)\,f^{(2)}(x)\right) \times f^{-1}(x)\,h_n^2\mu_2(K)+o\left(h_n^2\right),$$

and

$$Var\left[\widetilde{r}_n(x)\right] = \frac{1}{nh_n}Var\left[Y|X=x\right]f^{-1}(x)\,R(K) +o\left(\frac{1}{nh_n}\right).$$

It follows that,

$$MWISE\left[\widetilde{r}_n\right] = \frac{1}{nh_n}\left(I_4-I_5\right)R(K) +\frac{1}{4}\left(I_1+I_3-2I_2\right)h_n^4\mu_2^2(K) +o\left(h_n^4+\frac{1}{nh_n}\right).$$

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

$$\left(\left(\frac{I_4-I_5}{I_1+I_3-2I_2}\right)^{1/5}\left\{\frac{R(K)}{\mu_2^2(K)}\right\}^{1/5} n^{-1/5}\right), \tag{34}$$

and we have

$$MWISE\left[\widetilde{r}_n\right] = \frac{5}{4}\left(I_4-I_5\right)^{4/5}\left(I_1+I_3-2I_2\right)^{1/5} \times\Theta(K)\,n^{-4/5}+o\left(n^{-4/5}\right). \tag{35}$$

To estimate the optimal bandwidth (34), we must estimate I_1, I_2, I_3, I_4 and I_5. We use the following kernel estimator of I_1, I_2, I_3, I_4 and I_5:

$$\widetilde{I}_1 = \frac{1}{n^3b_n^6}\sum_{\substack{i,j,k=1\\ j\neq k}}^{n} K_b^{(2)}\left(\frac{X_i-X_j}{b_n}\right)K_b^{(2)}\left(\frac{X_i-X_k}{b_n}\right)Y_jY_k,$$

$$\widetilde{I}_2 = \frac{1}{n^3b_n^6}\sum_{\substack{i,j,k=1\\ j\neq k}}^{n} K_b^{(2)}\left(\frac{X_i-X_j}{b_n}\right)K_b^{(2)}\left(\frac{X_i-X_k}{b_n}\right)Y_iY_j,$$

$$\widetilde{I}_3 = \frac{1}{n^4b_n^6}\sum_{\substack{i,j,k,l=1\\ j\neq k\neq l}}^{n} K_b^{(2)}\left(\frac{X_i-X_j}{b_n}\right)K_b^{(2)}\left(\frac{X_i-X_k}{b_n}\right)Y_iY_l,$$

$$\widetilde{I}_4 = \frac{1}{n^2b_n}\sum_{\substack{i,j=1\\ i\neq j}}^{n} K_b\left(\frac{X_i-X_j}{b_n}\right)Y_i^2,$$

$$\widetilde{I}_5 = \frac{1}{n^2b_n}\sum_{\substack{i,k=1\\ i\neq k}}^{n} K_b\left(\frac{X_i-X_k}{b_n}\right)Y_iY_k,$$

where K_b is a kernel and b_n is the associated bandwidth given in (20).
We showed that in order to minimize the $MISE$ of $\widetilde{I}_1$ respectively of $\widetilde{I}_2$, $\widetilde{I}_3$, $\widetilde{I}_4$ and $\widetilde{I}_5$, the pilot bandwidth (b_n) must belong to $\mathcal{GS}(-3/14)$, respectively to $\mathcal{GS}(-3/14)$, $\mathcal{GS}(-3/14)$, $\mathcal{GS}(-2/5)$ and $\mathcal{GS}(-2/5)$.
Then the plug-in estimator of the bandwidth (h_n) using the nonrecursive estimator (4), is given by

$$\left(\left(\frac{\widetilde{I}_4-\widetilde{I}_5}{\widetilde{I}_1+\widetilde{I}_3-2\widetilde{I}_2}\right)^{1/5}\left\{\frac{R(K)}{\mu_2^2(K)}\right\}^{1/5} n^{-1/5}\right), \tag{36}$$

and the plug-in of the $MWISE$ of the nonrecursive estimator (4), is given by

$$\widetilde{MWISE}\left[\widetilde{r}_n\right] = \frac{5}{4}\left(\widetilde{I}_4-\widetilde{I}_5\right)^{4/5}\left(\widetilde{I}_1+\widetilde{I}_3-2\widetilde{I}_2\right)^{1/5}\Theta(K)\,n^{-4/5} +o\left(n^{-4/5}\right).$$

Finally, it follows from (14), (22), (30), (32) and (35), that:

The $MWISE$ of the proposed estimator (3) with the choice of the stepsizes $(\gamma_n,\beta_n) = \left(n^{-1},n^{-1}\right)$ is 1.06 larger than the nonrecursive estimator (4).

The $MWISE$ of the proposed estimator (3) with the choice of the stepsizes $(\gamma_n, \beta_n) = \left((1-a)\, n^{-1}, (1-a)\, n^{-1}\right)$ is 1.1 larger than the nonrecursive estimator (4).

We can't compare the $MWISE$ of the proposed estimator (3) with the choice of the stepsizes $(\gamma_n, \beta_n) = \left(n^{-1}, (1-a)\, n^{-1}\right)$ (respectively, the $MWISE$ of the proposed estimator (3) with the choice of the stepsizes $(\gamma_n, \beta_n) = \left((1-a)\, n^{-1}, n^{-1}\right)$) neither to the $MWISE$ of the others proposed estimators nor to the $MWISE$ of the nonrecursive estimator (4).

3. APPLICATIONS

The aim of our applications is to compare the performance of the semi-recursive estimators defined in (3) with that of the nonrecursive Nadaraya-Watson's estimator defined in (4).

When applying r_n one need to choose three quantities:

- The function K, we choose the Normal kernel.
- The stepsizes (γ_n, β_n) equal respectively to $\left(n^{-1}, n^{-1}\right)$, $\left(n^{-1}, (1-a)\, n^{-1}\right)$, $\left((1-a)\, n^{-1}, n^{-1}\right)$ or $\left((1-a)\, n^{-1}, (1-a)\, n^{-1}\right)$. These four choices are referred to as `Recursive 1`, `2`, `3` and `4` respectively.
- The bandwidth (h_n) is chosen to be equal respectively to (21) for (`Recursive 1`), (23) for (`Recursive 2`), (33) for (`Recursive 3`) and (31) for (`Recursive 4`).

When applying $\widetilde{r}_n$ one need to choose two quantities:

- The function K, as in the semi-recursive framework, we use the Normal kernel.
- The bandwidth (h_n) is chosen to be equal to (36).

3.1 Simulations

Throughout this subsection, we consider the regression model

$$Y = r(X) + \varepsilon,$$

where X is $\mathcal{N}(0,1)$-distributed and ε is $\mathcal{N}(0,\sigma)$-distributed, with σ is chosen in the interval $[0.1, 2]$.

In order to investigate the comparison between the proposed estimators, we consider two regression functions: cosine function $r(x) = \cos(x)$ (see Table 1) and the following function $r(x) = (1+\exp(x))^{-1}$ (see Table 2). For each fixed $\sigma \in [0.1, 2]$, the number of simulations is 500. We denote by r_i^* the true regression function, and by r_i the considered regression estimators, and then we compute the Mean Squared Error ($MSE = n^{-1} \sum_i (r_i - r_i^*)^2$).

3.1.0.1. 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. Performing all the proposed methods, we report the total `CPU` time values for each considered regression function and for each fixed σ and for each sample size in Tables 1 and 2. For the two tables we give the `CPU` time in seconds.

3.2 Real dataset

The CO2 dataset was available in the R package `Stat2Data` and contained 237 observations on the following two variables; Day and CO2, for more details see the station information system (GAWSIS). Scientists at a research station in Brotjacklriegel, Germany recorded CO2 levels, in parts per million, in the atmosphere for each day from the start of April through November in 2001.

Figure 1 and Tables 1 and 2 indicate that

- The `Recursive 1` is very close to the nonrecursive estimator (4).
- The two estimators `Recursive 2` and `Recursive 3` can be better than the others estimators in many situations.
- The `CPU` time are always faster using the proposed semi-recursive estimators and the reduction of `CPU` time goes from a minimum of 22.3% to a maximum of 60% compared to the nonrecursive estimator.

4. CONCLUSION

This paper proposes an automatic selection of the bandwidth of the semi-recursive kernel estimators of a regression function defined by the stochastic approximation algorithm (3). The proposed estimators asymptotically follow normal distribution. The estimators are compared to the nonrecursive Nadaraya-Watson's regression estimator. We showed that, using some selected bandwidth and some particularly stepsizes, the proposed semi-recursive estimators will be very competitive to the nonrecursive one. The simulation study confirms the nice features of our proposed semi-recursive estimators and statisfactory improvement in the `CPU` time in comparison to the nonrecursive estimator.

In conclusion, the proposed method allowed us to obtain quite similar results as the nonrecursive estimator proposed by Nadaraya [10] and Watson [21]. Moreover, we plan to make extensions of our method in the future and to consider the case of the averaged Révész's regression estimators (see Mokkadem et al. [9] and Slaoui [17, 18]) and the case of time series as in Hart and Vieu [5] in a recursive way (see Huang et al. [6]).

APPENDIX A. PROOFS

Throughout this section we use the following notation:

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

$$W_n(x) = h_n^{-1} K\left(\frac{x-X_n}{h_n}\right). \qquad (37)$$

$$Z_n(x) = h_n^{-1} Y_n K\left(\frac{x-X_n}{h_n}\right). \qquad (38)$$

Let us first state the following technical lemma.

Table 1. Quantitative comparison between the nonrecursive estimator (4) and four recursive estimators; recursive 1 correspond to the estimator (3) with the choice $(\gamma_n,\beta_n)=\left(n^{-1},n^{-1}\right)$, recursive 2 correspond to the estimator (3) with the choice $(\gamma_n,\beta_n)=\left(n^{-1},(1-a)\,n^{-1}\right)$, recursive 3 correspond to the estimator (3) with the choice $(\gamma_n,\beta_n)=\left((1-a)\,n^{-1},n^{-1}\right)$ and recursive 4 correspond to the estimator (3) with the choice $(\gamma_n,\beta_n)=\left((1-a)\,n^{-1},(1-a)\,n^{-1}\right)$. Here we consider the regression function $r(x)=\cos(x)$, $X\sim\mathcal{N}(0,1)$ and $\varepsilon\sim\mathcal{N}(0,\sigma)$ with $\sigma=0.1$ in the first block, $\sigma=0.5$ in the second block and $\sigma=1$ in the last block, we consider three sample sizes $n=100$, $n=200$ and $n=500$, the number of simulations is 500, and we compute the Mean squared error (MSE) and the CPU time in seconds

	Nadaraya	Recursive 1	Recursive 2	Recursive 3	Recursive 4
$n=100$			$\sigma=0.1$		
MSE	0.000812	0.000748	0.000764	**0.000567**	0.000667
CPU	238	184	170	**154**	164
$n=200$					
MSE	0.000507	0.000483	0.000508	**0.000366**	0.000443
CPU	835	514	509	**464**	470
$n=500$					
MSE	0.000284	0.000279	0.000294	**0.000217**	0.000260
CPU	3679	2185	1973	1966	**1865**
$n=100$			$\sigma=0.5$		
MSE	0.004486	0.004447	0.004286	**0.003729**	0.004184
CPU	231	143	135	137	**129**
$n=200$					
MSE	0.002331	0.002337	0.002142	**0.001929**	0.002141
CPU	885	568	549	485	**457**
$n=500$					
MSE	0.001372	0.001411	0.001265	**0.001174**	0.001291
CPU	3498	2049	**1943**	2242	2045
$n=100$			$\sigma=1$		
MSE	**0.013960**	0.021204	0.020982	0.021476	0.021832
CPU	246	166	**136**	146	137
$n=200$					
MSE	**0.006016**	0.010935	0.008714	0.012524	0.011657
CPU	831	580	519	541	**505**
$n=500$					
MSE	0.001916	**0.001816**	0.002268	0.003018	0.001972
CPU	3801	2193	2043	2024	**1875**

Lemma A.1. *Let $(v_n)\in\mathcal{GS}(v^*)$, $(\beta_n)\in\mathcal{GS}(-\beta)$, and $m>0$ such that $m-v^*\xi>0$ where ξ is defined in (6). We have*

$$\lim_{n\to+\infty} v_nQ_n^m\sum_{k=1}^{n}Q_k^{-m}\frac{\beta_k}{v_k}=\frac{1}{m-v^*\xi}. \tag{39}$$

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

$$\lim_{n\to+\infty} v_nQ_n^m\left[\sum_{k=1}^{n}Q_k^{-m}\frac{\beta_k}{v_k}\alpha_k+\delta\right]=0. \tag{40}$$

Lemma A.1 is widely applied throughout the proofs. Let us underline that it is its application, which requires Assumption $(A2)(iii)$ on the limit of $(n\gamma_n)$ as n goes to infinity.

Our proofs are organized as follows. Propositions 2.1 and 2.2 in Sections A.1 and A.2 respectively, Theorem 2.1 in Section A.3. Propositions 2.3 and 2.4 in Sections A.4 and A.5 respectively, Theorem 2.2 in Section A.6.

A.1 Proof of Proposition 2.1

Let us first note that, for x such that $f_n(x)\neq 0$, we have

$$r_n(x)-r(x) = B_n(x)\frac{f(x)}{f_n(x)}, \tag{41}$$

with

$$\begin{aligned} B_n(x) &= \frac{1}{f(x)}\left(a_n(x)-a(x)\right)\\ &\quad -\frac{r(x)}{f(x)}\left(f_n(x)-f(x)\right). \end{aligned} \tag{42}$$

It follows from (41), that the asymptotic behaviour of $r_n(x)-r(x)$ can be deduced from the one of $B_n(x)$. Moreover, the following Lemma follows from the Proposition 1 of Mokkadem et al. [8].

Table 2. Quantitative comparison between the nonrecursive estimator (4) and four recursive estimators; recursive 1 correspond to the estimator (3) with the choice $(\gamma_n, \beta_n) = (n^{-1}, n^{-1})$, recursive 2 correspond to the estimator (3) with the choice $(\gamma_n, \beta_n) = (n^{-1}, (1-a)n^{-1})$, recursive 3 correspond to the estimator (3) with the choice $(\gamma_n, \beta_n) = ((1-a)n^{-1}, n^{-1})$ and recursive 4 correspond to the estimator (3) with the choice $(\gamma_n, \beta_n) = ((1-a)n^{-1}, (1-a)n^{-1})$. Here we consider the regression function $r(x) = (1+\exp(x))^{-1}$, $X \sim \mathcal{N}(0,1)$ and $\varepsilon \sim \mathcal{N}(0,\sigma)$ with $\sigma = 0.1$ in the first block, $\sigma = 0.5$ in the second block and $\sigma = 2$ in the last block, we consider three sample sizes $n = 100$, $n = 200$ and $n = 500$, the number of simulations is 500, and we compute the Mean squared error (MSE) and the CPU time in seconds

	Nadaraya	Recursive 1	Recursive 2	Recursive 3	Recursive 4
$n = 100$			$\sigma = 0.1$		
MSE	$1.31e^{-04}$	$1.15e^{-04}$	$\mathbf{6.22e^{-05}}$	$1.71e^{-04}$	$1.03e^{-04}$
CPU	249	184	**135**	146	146
$n = 200$					
MSE	$4.38e^{-05}$	$3.87e^{-05}$	$\mathbf{1.50e^{-05}}$	$8.03e^{-05}$	$3.63e^{-05}$
CPU	909	524	475	601	**458**
$n = 500$					
MSE	$5.70e^{-06}$	$5.02e^{-06}$	$\mathbf{3.20e^{-06}}$	$2.32e^{-05}$	$4.29e^{-06}$
CPU	3708	1803	1672	1855	**1483**
$n = 100$			$\sigma = 0.5$		
MSE	0.000351	0.000325	**0.000252**	0.000350	0.000296
CPU	256	144	132	134	**125**
$n = 200$					
MSE	0.000189	0.000171	0.000154	0.000163	**0.000151**
CPU	873	524	483	576	**451**
$n = 500$					
MSE	$2.30e^{-05}$	$2.25e^{-05}$	$2.42e^{-05}$	$3.351e^{-05}$	$\mathbf{2.06e^{-05}}$
CPU	4389	2113	1987	1999	**1973**
$n = 100$			$\sigma = 2$		
MSE	0.003447	0.003294	0.003155	**0.003132**	0.003137
CPU	294	155	173	**143**	148
$n = 200$					
MSE	0.000160	0.000152	0.000162	**0.000111**	0.000189
CPU	917	503	581	515	**477**
$n = 500$					
MSE	$6.56e^{-05}$	$7.03e^{-05}$	$\mathbf{5.01e^{-05}}$	$6.70e^{-05}$	$5.39e^{-05}$
CPU	3643	2105	1951	1947	**1877**

Lemma A.2 (Bias and variance of f_n)**.** *Let Assumptions $(A1)-(A3)$ and suppose that the stepsize $(\gamma_n) = (n^{-1})$.*

1. *If $a \in]0, 1/5]$, then*

$$\mathbb{E}[f_n(x)] - f(x) = \frac{1}{2(1-2a)} f^{(2)}(x) h_n^2 \mu_2(K) + o(h_n^2). \quad (43)$$

If $a \in]1/5, 1[$, then

$$\mathbb{E}[f_n(x)] - f(x) = o\left(\sqrt{n^{-1}h_n^{-1}}\right). \quad (44)$$

2. *If $a \in [1/5, 1[$, then*

$$Var[f_n(x)] = \frac{1}{1+a}\frac{1}{nh_n} f(x) R(K) + o\left(\frac{1}{nh_n}\right). \quad (45)$$

If $a \in]0, 1/5[$, then

$$Var[f_n(x)] = o(h_n^4). \quad (46)$$

Following similar steps as the proof of the Proposition 1 of Mokkadem et al. [8], we show that

Lemma A.3 (Bias and variance of a_n)**.** *Let Assumptions $(A1)-(A3)$ hold.*

1. *If $a \in]0, \beta/5]$, then*

$$\mathbb{E}[a_n(x)] - a(x) = \frac{1}{2(1-2a\xi)} a^{(2)}(x) h_n^2 \mu_2(K) + o(h_n^2). \quad (47)$$

If $a \in]\beta/5, 1[$, then

$$\mathbb{E}[a_n(x)] - a(x) = o\left(\sqrt{\beta_n h_n^{-1}}\right). \quad (48)$$

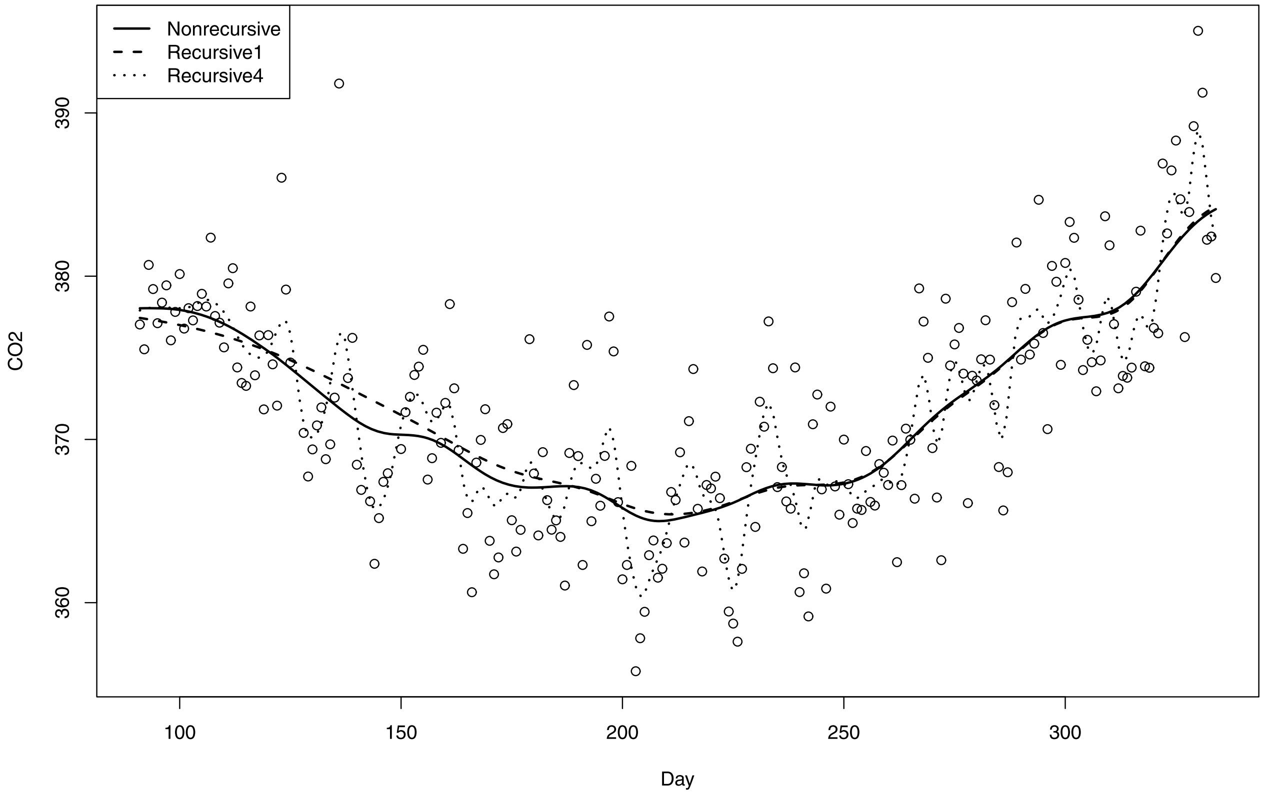

Figure 1. The daily carbon dioxide measurements data with automatic bandwidth selection using the nonrecursive Nadaraya's estimator (4) and two semi-recursive estimators (3) (`Recursive 1` and `Recursive 4`).

2. *If $a \in [\beta/5, 1[$, then*

$$Var\left[a_n(x)\right] = \frac{\mathbb{E}\left[Y^2|X=x\right]f(x)}{(2-(\beta-a)\xi)}\frac{\beta_n}{h_n}R(K) + o\left(\frac{\beta_n}{h_n}\right). \tag{49}$$

If $a \in]0, \beta/5[$, then

$$Var\left[a_n(x)\right] = o\left(h_n^4\right). \tag{50}$$

Then, (7) follows from (43), (47) and (41) and (8) follows from (44), (48) and (41).
Now, it follows from (42) that

$$\begin{aligned} Var\left[B_n(x)\right] &= \frac{1}{f^2(x)}\left\{Var\left[a_n(x)\right] + r^2(x)Var\left[f_n(x)\right]\right. \\ &\quad \left. -2r(x)Cov\left(a_n(x), f_n(x)\right)\right\}. \end{aligned} \tag{51}$$

In view of $(A3)$, and with the choice of the stepsize $(\gamma_n) = (n^{-1})$ and using Lemma A.1, classical computations gives

$$Cov\left(a_n(x), f_n(x)\right) = \frac{\xi}{1+a\xi}\frac{\beta_n}{h_n}r(x)f(x)R(K) + o\left(\frac{\beta_n}{h_n}\right). \tag{52}$$

Then, the combination of (41), (51), (45), (49) and (52), gives (9), and the combination of (41), (51), (46), (50) and (52), gives (10).

A.2 Proof of Proposition 2.2

Following similar steps as the proof of the Proposition 2 of Mokkadem et al. [8], we proof the Proposition 2.2.

A.3 Proof of Theorem 2.1

Let us at first assume that, if $a \geq \beta/5$, then

$$\sqrt{\beta_n^{-1}h_n}\left(r_n(x) - \mathbb{E}\left[r_n(x)\right]\right) \stackrel{\mathcal{D}}{\rightarrow} \mathcal{N}\left(0, V_{a,\xi,\beta}^{(1)}\right). \tag{53}$$

In the case when $a > \beta/5$, Part 1 of Theorem 2.1 follows from the combination of (8) and (53). In the case when $a = \beta/5$, Parts 1 and 2 of Theorem 2.1 follow from the combination of (7) and (53). In the case $a < \beta/5$, (10) implies that

$$h_n^{-2}\left(r_n(x) - \mathbb{E}\left(r_n(x)\right)\right) \stackrel{\mathbb{P}}{\rightarrow} 0,$$

and the application of (7) gives Part 2 of Theorem 2.1.

We now prove (53). In view of (42), we have

$$\begin{aligned} B_n(x) - \mathbb{E}\left[B_n(x)\right] &= \frac{1}{f(x)}Q_n\sum_{k=1}^{n}\left(T_k(x) - \mathbb{E}\left[T_k(x)\right]\right), \end{aligned} \tag{54}$$

with

$$T_k(x) = Q_k^{-1}\left(\beta_k Z_k(x) - r(x)\zeta_n \zeta_k^{-1}\gamma_k W_k(x)\right). \tag{55}$$

In the case when $(\gamma_n) = (n^{-1})$, we have $\zeta_n = (nQ_n)^{-1}$ et $\zeta_k^{-1}\gamma_k = Q_k$, then

$$T_k(x) = Q_k^{-1}\beta_k Z_k(x) - r(x)(nQ_n)^{-1} W_k(x).$$

Set

$$Y_k(x) = T_k(x) - \mathbb{E}(T_k(x)). \tag{56}$$

Moreover, we have

$$\begin{aligned} v_n^2 &= \sum_{k=1}^n Var(Y_k(x)) \\ &= \sum_{k=1}^n Q_k^{-2}\beta_k^2 Var(Z_k(x)) \\ &\quad + r^2(x)(nQ_n)^{-2}\sum_{k=1}^n Var(W_k(x)) \\ &\quad -2r(x)(nQ_n)^{-1}\sum_{k=1}^n Q_k^{-1}\beta_k Cov(Z_k(x), W_k(x)). \end{aligned}$$

Moreover, in view of $(A3)$, classical computations give

$$\begin{aligned} Var(Z_k(x)) &= \frac{1}{h_k}\left[\mathbb{E}\left[Y^2|X=x\right]\right. \\ &\quad \left.\times f(x)R(K) + o(1)\right], \\ Var(W_k(x)) &= \frac{1}{h_k}\left[f(x)R(K) + o(1)\right], \\ Cov(Z_k(x), W_k(x)) &= \frac{1}{h_k}\left[r(x)f(x)R(K) + o(1)\right]. \end{aligned}$$

The application of Lemma A.1 ensures that

$$\begin{aligned} v_n^2 &= \sum_{k=1}^n \frac{Q_k^{-2}\beta_k^2}{h_k}\left[\mathbb{E}\left[Y^2|X=x\right]f(x)R(K) + o(1)\right] \\ &\quad + \frac{r(x)}{n^2Q_n^2}\sum_{k=1}^n \frac{1}{h_k}\left[f(x)R(K) + o(1)\right] \\ &\quad -2\frac{r(x)}{nQ_n}\sum_{k=1}^n \frac{Q_k^{-1}\beta_k}{h_k}\left[r(x)f(x)R(K) + o(1)\right] \\ &= \frac{f^2(x)}{Q_n^2}\frac{\beta_n}{h_n}\left[V_{a,\xi,\beta}^{(1)} + o(1)\right]. \end{aligned}$$

On the other hand, we have, for all $p > 0$,

$$\mathbb{E}\left[|T_k(x)|^{2+p}\right] = O\left(\frac{1}{h_k^{1+p}}\right),$$

and, since $\lim_{n\to\infty}(n\beta_n) > (\beta - a)/2$, there exists $p > 0$ such that $\lim_{n\to\infty}(n\beta_n) > \frac{1+p}{2+p}(\beta - a)$. Applying Lemma A.1, we get

$$\begin{aligned} \sum_{k=1}^n \mathbb{E}\left[|Y_k(x)|^{2+p}\right] &= O\left(\sum_{k=1}^n Q_k^{-2-p}\beta_k^{2+p}\mathbb{E}\left[|T_k(x)|^{2+p}\right]\right) \\ &= O\left(\sum_{k=1}^n \frac{Q_k^{-2-p}\beta_k^{2+p}}{h_k^{1+p}}\right) \\ &= O\left(\frac{\beta_n^{1+p}}{Q_n^{2+p}h_n^{1+p}}\right), \end{aligned}$$

and we thus obtain

$$\frac{1}{v_n^{2+p}}\sum_{k=1}^n \mathbb{E}\left[|Y_k(x)|^{2+p}\right] = O\left(\left[\beta_n h_n^{-1}\right]^{p/2}\right) = o(1).$$

The convergence in (53) then follows from the application of Lyapounov's Theorem.

A.4 Proof of Proposition 2.3

The following Lemma follows from the Proposition 1 of Mokkadem et al. [8].

Lemma A.4 (Bias and variance of f_n). *Let Assumptions $(A1) - (A3)$ and suppose that the stepsize $(\gamma_n) = ([1-a]n^{-1})$.*

1. *If $a \in]0, 1/5]$, then*

$$\mathbb{E}[f_n(x)] - f(x) = \frac{1-a}{2(1-3a)}f^{(2)}(x)h_n^2\mu_2(K) + o\left(h_n^2\right). \tag{57}$$

If $a \in]1/5, 1[$, then

$$\mathbb{E}[f_n(x)] - f(x) = o\left(\sqrt{n^{-1}h_n^{-1}}\right). \tag{58}$$

2. *If $a \in [1/5, 1[$, then*

$$Var[f_n(x)] = \frac{1-a}{nh_n}R(K) + o\left(\frac{1}{nh_n}\right). \tag{59}$$

If $a \in]0, 1/5[$, then

$$Var[f_n(x)] = o\left(h_n^4\right). \tag{60}$$

Then, (24) follows from (57), (47) and (41) and (25) follows from (58), (48) and (41).
Moreover, in view of $(A3)$, and using the choice of the stepsize $(\gamma_n) = ([1-a]n^{-1})$ and using Lemma A.1, classical computations gives

$$Cov(a_n(x), f_n(x)) = (1-a)\xi\frac{\beta_n}{h_n}r(x)f(x)R(K) + o\left(\frac{\beta_n}{h_n}\right). \tag{61}$$

Then, the combination of (41), (51), (45), (49) and (61), gives (26), and the combination of (41), (51), (60), (50) and (61), gives (27).

A.5 Proof of Proposition 2.4

Following similar steps as the proof of the Proposition 2 of Mokkadem et al. [8], we proof the Propostion 2.4.

A.6 Proof of Theorem 2.2

Following similar steps as the proof of the Theorem 2.1 and using the fact that in the case when $(\gamma_n) = \left([1-a]\, n^{-1}\right)$, we have $Q_k^{-1}\zeta_n\zeta_k^{-1}\gamma_k = (1-a)\, h_k/(nh_nQ_n)$, and then it follows from (55), that

$$T_k(x) \quad = \quad Q_k^{-1}\beta_k Z_k(x) - r(x)\,\frac{(1-a)\,h_k}{nh_nQ_n} W_k(x)\,,$$

we prove Theorem 2.2.

ACKNOWLEDGEMENTS

We are grateful to two referees and an Editor for their helpful comments, which have led to this substantially improved version of the paper.

Received 1 April 2015

REFERENCES

[1] Altman, N. and Leger, C. (1995). Bandwidth selection for kernel distribution function estimation. *J. Statist. Plann. Inference* **46** 195–214. MR1354087

[2] Bojanic, R. and Seneta, E. (1973). A unified theory of regularly varying sequences. *Math. Z.* **134** 91–106. MR0333082

[3] Delaigle, A. and Gijbels, I. (2004). Practical bandwidth selection in deconvolution kernel density estimation. *Comput. Statist. Data Anal.* **45** 249–267. MR2045631

[4] Galambos, J. and Seneta, E. (1973). Regularly varying sequences. *Proc. Amer. Math. Soc.* **41** 110–116. MR0323963

[5] Hart, J. D. and Vieu, P. (1990). Data-driven bandwidth choice for density estimation based on dependent data. *Ann. Statist.* **18** 873–890. MR1056341

[6] Huang, Y. Chen, X. and Wu, W. B. (2014). Recursive nonparametric estimation for times series. *IEEE Trans. Inform. Theory.* **60** 1301–1312. MR3164976

[7] Mokkadem, A. and Pelletier, M. (2007). A companion for the Kiefer-Wolfowitz-Blum stochastic approximation algorithm. *Ann. Statist.* **35** 1749–1772. MR2351104

[8] 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. MR2508006

[9] 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. MR2496230

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

[11] Révész, P. (1973). Robbins-Monro procedure in a Hilbert space and its application in the theory of learning processes I. *Studia Sci. Math. Hung.* **8** 391–398. MR0373198

[12] Révész, P. (1977). How to apply the method of stochastic approximation in the non-parametric estimation of a regression function. *Math. Operationsforsch. Statist., Ser. Statistics.* **8** 119–126. MR0501557

[13] Silverman, B. W. (1986). Density estimation for statistics and data analysis. *Chapman and Hall*, London. MR0848134

[14] Slaoui, Y. (2013). Large and moderate deviation principles for recursive kernel density estimators defined by stochastic approximation method. *Serdica Math. J.* **39** 53–82. MR3087768

[15] Slaoui, Y. (2014a). Bandwidth selection for recursive kernel density estimators defined by stochastic approximation method. *J. Probab. Stat.* **2014**, ID 739640, doi:10.1155/2014/739640. MR3219439

[16] Slaoui, Y. (2014b). The stochastic approximation method for the estimation of a distribution function. *Math. Methods Statist.* **23** 306–325. MR3295061

[17] Slaoui, Y. (2015a). Moderate deviation principles for recursive regression estimators defined by stochastic approximation method. *Int. J. Math. Stat.* **16** 51–60. MR3378190

[18] Slaoui, Y. (2015b). Large and moderate deviation principles for recursive regression estimators defined by stochastic approximation method. *Serdica Math. J.* **41** 307–328. MR3363608

[19] Slaoui, Y. (2015c). Plug-In Bandwidth selector for recursive kernel regression estimators defined by stochastic approximation method. *Stat. Neerl.* **69**, 483–509.

[20] Tsybakov, A. B. (1990). Recurrent estimation of the mode of a multidimensional distribution. *Probl. Inf. Transm.* **8** 119–126. MR1051586

[21] Watson, G. S. (1964). Smooth regression analysis. *Sankhya A.* **26** 359–372. MR0185765

Yousri Slaoui
Laboratoire de Mathématiques et Application
86962 Futuroscope
Chasseneuil
France
E-mail address: yousri.slaoui@math.univ-poitiers.fr