Optimal bandwidth selection for semi-recursive kernel regression estimators

Yousri Slaoui
Université de Poitiers

Abstract: 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.

Key words and phrases: Nonparametric regression; Stochastic approximation algorithm; Smoothing, curve fitting.
Mathematics Subject Classification (2010) Primary 62G08. 62L20. 65D10.

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 (2004). 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, plug-in bandwidth selection method for recursive kernel density estimators defined by stochastic approximation method have been done by Slaoui (2014a) and for recursive kernel distribution estimators have been done by Slaoui (2014b). 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 \(a : x \mapsto \mathbb{E}(Y \mid X = x)f(x)\) 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 $(\beta_n)$ is a sequence of positive real numbers that goes to zero. To define $W_n(x)$, we follow the approach of Révész (1973), Tsybakov (1990) and of Mokkadem et al. (2009a,b) and introduces a kernel $K$ (that is, a function satisfying $\int R K(x) dx = 1$), and a bandwidth $(h_n)$ (that is, a sequence of positive real numbers that goes to zero), and sets $W_n(x) = h_n^{-1} Y_n K (h_n^{-1} (x - X_n)) - a_{n-1}(x)$. Then, the estimator $a_n$ to recursively estimate the function $a$ at the point $x$ can be written as

$$a_n(x) = (1 - \beta_n) a_{n-1}(x) + \beta_n h_n^{-1} Y_n K \left( \frac{x - X_n}{h_n} \right).$$  \hspace{1cm} (1)

This estimator was proposed by Slăou (2015) 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. (2009a) to estimate recursively the density function $f$ at the point $x$

$$f_n(x) = (1 - \gamma_n) f_{n-1}(x) + \gamma_n h_n^{-1} K \left( h_n^{-1} [x - X_n] \right),$$  \hspace{1cm} (2)

where the stepsize $(\gamma_n)$ 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 $x$

$$r_n(x) = \begin{cases} \frac{a_n(x)}{f_n(x)} & \text{if } f_n(x) \neq 0, \\ 0 & \text{otherwise}. \end{cases}$$  \hspace{1cm} (3)

Moreover, we show that the optimal bandwidth which minimize the $E[|r_n(x) - r(x)|^2] dx$ of $r_n$ depends on the choice of the stepsize $(\gamma_n)$ and $(\beta_n)$; 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

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

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 (1964) and Watson (1964), and defined as

$$\tilde{r}_n(x) = \begin{cases} \frac{\tilde{a}_n(x)}{\tilde{f}_n(x)} & \text{if } \tilde{f}_n(x) \neq 0, \\ 0 & \text{otherwise}, \end{cases}$$  \hspace{1cm} (4)
with
\[ \tilde{a}_n(x) = \frac{1}{nh_n} \sum_{i=1}^{n} Y_i K \left( \frac{x - X_i}{h_n} \right) \quad \text{and} \quad \tilde{f}_n(x) = \frac{1}{nh_n} \sum_{i=1}^{n} K \left( \frac{x - X_i}{h_n} \right). \]

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 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

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

Condition (5) 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 \( b \in \mathbb{R} \),

\[ n^{\gamma (\log n)^b}, \quad n^{\gamma (\log \log n)^b}, \quad \text{and so on.} \]

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, \( \int_{\mathbb{R}} zK(z) \, dz = 0 \) and \( \int_{\mathbb{R}} z^2K(z) < \infty \).

(A2) i) \((\beta_n) \in \mathcal{GS}(-\beta)\) with \( \beta \in [1/2, 1] \).

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

\( \bullet \) 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 \( s \).

\( \bullet \) 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.

\( \bullet \) iii) For \( q \in \{0, 1\}, \) \( \int_{\mathbb{R}} |t|^q |\partial_s^q g(x, t)| \, dt < \infty \), and \( s \mapsto \int_{\mathbb{R}} t^q \partial_s^q g(s, t) \, dt \) 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 \((n\beta_n)^{-1}\) is finite. For simplicity, we introduce the following notations:

\[ \xi = \lim_{n \to \infty} (n\beta_n)^{-1}, \tag{6} \]
\[ R(K) = \int_{\mathbb{R}} K^2(z) \, dz, \quad \mu_j(K) = \int_{\mathbb{R}} z^j K(z) \, dz, \]

\[ \Theta(K) = R(K)^{4/5} \mu_2(K)^{2/5}, \quad I_1 = \int_{\mathbb{R}} (a^{(2)}(x))^2 f(x) \, dx, \]

\[ I_2 = \int_{\mathbb{R}} a^{(2)}(x) f^{(2)}(x) r(x) f(x) \, dx, \quad I_3 = \int_{\mathbb{R}} (f^{(2)}(x))^2 r^2(x) f(x) \, dx, \]

\[ I_4 = \int_{\mathbb{R}} \mathbb{E}[Y^2|X = x] f^2(x) \, dx, \quad I_5 = \int_{\mathbb{R}} r^2(x) f^2(x) \, dx, \]

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 \(M\), in order to provide a comparison with the nonrecursive estimator \(\hat{r}\). Moreover, it was shown in Mokkadem et al. (2009a) and considered in Slaoui (2013) 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 \((\gamma_n)\) must be chosen in \(GS(-1)\) 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 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

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

2. If \(a \in [\beta/5, 1[\), then

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

3. If \(a \in [0, \beta/5[\), then

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

4. 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 \(M\) then heavily depend on the choice of the stepsizes \((\gamma_n)\) and \((\beta_n)\). Let us first state the following theorem, which gives the weak convergence rate of the estimator \(r_n\) defined in \(M\) in the case of \((\gamma_n) = (n^{-1})\).

**Theorem 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 \( \beta^{-1} h_n^5 \to c \), then
\[
\sqrt{\beta^{-1} h_n} (r_n(x) - r(x)) \overset{D}{\to} \mathcal{N} \left( \sqrt{c} B_{a,\xi}(x) , V_{a,\xi,\beta}(x) \right),
\]
where
\[
B_{a,\xi}^{(1)}(x) = \frac{1}{2f(x)} \left( \frac{a^{(2)}(x)}{(1 - 2a\xi)} - \frac{r(x)f^{(2)}(x)}{(1 - 2a)} \right) \mu_2(K),
\]
\[
V_{a,\xi,\beta}^{(1)}(x) = \left\{ \frac{\mathbb{E}[Y^2|X = x]}{2 - (\beta - a)\xi} f(x) - \left( \frac{2\xi}{1 + a\xi} - \frac{\xi}{1 + a} \right) \frac{r^2(x)}{f(x)} \right\} R(K).
\]

2. If \( nh_n^5 \to \infty \), then
\[
\frac{1}{h_n^2} (r_n(x) - r(x)) \overset{P}{\to} B_{a,\xi}(x),
\]
where \( \overset{D}{\to} \) denotes the convergence in distribution, \( \mathcal{N} \) the Gaussian-distribution and \( \overset{P}{\to} \) 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 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
\[
\sqrt{nh_n} (r_n(x) - r(x)) \overset{D}{\to} \mathcal{N} \left( \sqrt{c} B_{a,1}^{(1)}(x) , V_{a,1,1}^{(1)}(x) \right).
\]

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

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

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

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

**Proposition 2** (\( MWISE \) of \( r_n \)). Let Assumptions (A1) – (A3) hold, and suppose that \( (\gamma_n) = (n^{-1}) \).

1. If \( a \in [0, \beta/5] \), then
\[
MWISE[r_n] = \frac{1}{4} \left( \frac{I_1}{(1 - 2a\xi)^2} + \frac{I_3}{(1 - 2a)^2} - 2 \frac{I_2}{(1 - 2a)(1 - 2a\xi)} \right) h_n^4 \mu_2(K) + o(h_n^4).
\]
2. If \( a = \beta / 5 \), then

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

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

**Corollary 2.** Let Assumptions (A1) – (A3) hold, and suppose that \( (\gamma_n) = (n^{-1}) \). To minimize the \( MWISE \) of \( r_n \), the stepsize \( (\beta_n) \) must be chosen in \( GS (-1) \), the bandwidth \( (h_n) \) must equal

\[
\left( \frac{I_4}{(2 - (\beta - a) \xi)} - \left( \frac{2\xi}{1 + a\xi} - \frac{\xi}{1 + a} \right) I_5 \right)^{1/5} \left\{ \frac{R(K)}{\mu_2^2(K)} \right\}^{1/5} \beta_n^{1/5}.
\]

Then, we have

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

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

**Corollary 3.** Let Assumptions (A1) – (A3) hold, and suppose that \( (\gamma_n) = (n^{-1}) \). To minimize the \( MWISE \) of \( r_n \), the stepsize \( (\beta_n) \) must be chosen in \( GS (-1) \), \( \lim_{n \to \infty} n^{\beta_n} = \beta_0 \), the bandwidth \( (h_n) \) must equal

\[
\left( \frac{\beta_0 - 2/5}{2} \right)^{1/5} \left( I_4 - \frac{(7\beta_0 - 1)\beta_0 - 2/5}{3\beta_0^2(\beta_0 + 1/5)} I_5 \right)^{1/5} \left\{ \frac{R(K)}{\mu_2^2(K)} \right\}^{1/5} n^{-1/5},
\]

and we then have

\[
MWISE [r_n] = \frac{5}{4} \frac{\beta_0^2}{2^{4/5}(\beta_0 - 2/5)^{6/5}} \left( I_4 - \frac{(7\beta_0 - 1)\beta_0 - 2/5}{3\beta_0^2(\beta_0 + 1/5)} I_5 \right)^{4/5} \times \left( I_1 + \frac{25}{3\beta_0^2(\beta_0 - 2/5)} I_3 - \frac{10}{3} \frac{(\beta_0 - 2/5)}{\beta_0} I_2 \right)^{1/5} \Theta(K) n^{-4/5} + o(n^{-4/5}).
\]
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
\[
\left( \frac{3}{10} \right)^{1/5} \left( \frac{I_4 - I_5}{I_1 + I_3 - 2I_2} \right)^{1/5} \left\{ \frac{R(K)}{\mu_2^2(K)} \right\}^{1/5} n^{-1/5},
\]
and we then have
\[
MWISE[r_n] = \frac{5}{4^{2/5}} \left( \frac{5}{3} \right)^{6/5} (I_4 - I_5)^{4/5} \times (I_1 + I_3 - 2I_2)^{1/5} \Theta(K) n^{-4/5} + o(n^{-4/5}).
\]

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 (1995) and Slaoui (2014a,b)], which is called the plug-in estimate, and we use the following kernel estimators of \( I_1, I_2, I_3, I_4 \) and \( I_5 \):
\[
\hat{I}_1 = \frac{Q_2^2}{n} \sum_{i,j,k=1}^n Q_{-1}Q_{-1}^{-1} \beta_j b_j^{-3} b_k^{-3} K_{b(2)} \left( \frac{X_i - X_j}{b_j} \right) K_{b(2)} \left( \frac{X_i - X_k}{b_k} \right) Y_j Y_k, \quad (15)
\]
\[
\hat{I}_2 = \frac{\Pi_n Q_n}{n} \sum_{i,j,k=1}^n \Pi_{-1}Q_{-1}^{-1} \gamma_j b_j^{-3} b_k^{-3} K_{b(2)} \left( \frac{X_i - X_j}{b_j} \right) K_{b(2)} \left( \frac{X_i - X_k}{b_k} \right) Y_i Y_j, \quad (16)
\]
\[
\hat{I}_3 = \frac{\Pi_n^2}{n} \sum_{i,j,k,l=1}^n \Pi_{-1}Q_{-1}^{-1} \gamma_j \gamma_k b_j^{-3} b_k^{-3} K_{b(2)} \left( \frac{X_i - X_j}{b_j} \right) K_{b(2)} \left( \frac{X_i - X_k}{b_k} \right) Y_i Y_l, \quad (17)
\]
\[
\hat{I}_4 = \frac{\Pi_n}{n} \sum_{i,j,k,l=1}^n \Pi_{-1}Q_{-1}^{-1} \gamma_j b_j^{-1} K_{b(2)} \left( \frac{X_i - X_k}{b_k} \right) Y_i^2, \quad (18)
\]
\[
\hat{I}_5 = \frac{Q_n}{n} \sum_{i,j,k,l=1}^n Q_{-1}^{-1} \beta_j b_j^{-1} K_{b(2)} \left( \frac{X_i - X_k}{b_k} \right) Y_i Y_l, \quad (19)
\]
where \( K_{b} \) is a kernel and \( b_n \) is the associated bandwidth.

In practice, we take
\[
b_n = n^{-\beta} \min \left\{ \hat{s}, \frac{Q_3 - Q_1}{1.349} \right\}, \quad \beta \in ]0, 1[
\]
(see [Silverman (1986)]) where \( \hat{s} \) the sample standard deviation, and \( Q_1, Q_3 \) denoting the first and third quartiles, respectively.

We followed the same steps as in [Slaoui (2014a)] and we showed that in order to minimize the \( MISE \) of \( \hat{I}_1 \) respectively of \( \hat{I}_2, \hat{I}_3, \hat{I}_4 \) and \( \hat{I}_5 \), the pilot bandwidth \( (h_n) \) must belong to \( GS (-3/14) \), respectively to \( GS (-3/14), GS (-3/14), GS (-2/5) \) and \( GS (-2/5) \).

Finally, the plug-in estimator of the bandwidth \( (h_n) \) using the semi-recursive estimators defined in (13) with the stepsizes \( (\gamma_n, \beta_n) = (n^{-1}, n^{-1}) \),
\[
\left( \frac{3}{10} \right)^{1/5} \left( \frac{\hat{I}_4 - \hat{I}_5}{\hat{I}_1 + \hat{I}_3 - 2\hat{I}_2} \right)^{1/5} \left\{ \frac{R(K)}{\mu_2^2(K)} \right\}^{1/5} n^{-1/5},
\]
(21)
\[
\hat{M\text{W}ISE}[r_n] = 5 \cdot \frac{1}{4} \cdot 24/5 \cdot \left(\frac{5}{3}\right)^{6/5} \left(\hat{I}_4 - \hat{I}_5\right)^{4/5} \times \left(\hat{I}_1 + \hat{I}_3 - 2\hat{I}_2\right)^{1/5} \Theta(K) n^{-4/5} \\
+ o\left(n^{-4/5}\right).
\]

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

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

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

\[
\left(\frac{1}{5}\right)^{1/5} \left(\frac{\hat{I}_4 - \frac{23}{24} \hat{I}_5}{\hat{I}_1 + \frac{25}{36} \hat{I}_3 - \frac{5}{3} \hat{I}_2}\right)^{1/5} \left\{\frac{R(K)}{\mu^2(K)}\right\}^{1/5} n^{-1/5},
\]

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

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

Let us now provide the case when the stepsize \((\gamma_n)\) is chosen to minimize the variance of \(f_n\). It was shown in Mokkadem et al. (2009a) and considered in Slaoui (2013) that to minimize the variance of \(f_n\), the stepsize \((\gamma_n)\) must be chosen in \(GS(-1)\) and must satisfy \(\lim_{n \to \infty} n\gamma_n = 1 - a\). We consider here the case \((\gamma_n) = ((1 - a) n^{-1})\). Our first result is the following proposition, which gives the bias and the variance of \(r_n\) in the special case of \((\gamma_n) = ((1 - a) n^{-1})\).

**Proposition 3** (Bias and variance of \(r_n\)). Let Assumptions (A1) – (A3) hold, and suppose that \((\gamma_n) = ((1 - a) n^{-1})\).

1. If \(a \in ]0, \beta/5]\), then

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

If \(a \in ]\beta/5, 1]\), then

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

2. If \(a \in [\beta/5, 1]\), then

\[
\text{Var}[r_n(x)] = \frac{\beta_n}{h_n} \left\{\mathbb{E}[Y^2|X = x] \left(\frac{1}{(2 - (\beta - a)\xi)} f(x) - (1 - a) \xi \frac{r^2(x)}{f(x)}\right) R(K)\right\} + o\left(\beta_n h_n^{-1}\right).
\]
If \( a \in [0, \beta/5] \), then
\[
\text{Var} [r_n (x)] = o \left( h_n^4 \right).
\] (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 \( (\gamma_n) \) and \( (\beta_n) \).

Let us first state the following theorem, which gives the weak convergence rate of the estimator \( r_n \) defined in [3].

**Theorem 2** (Weak pointwise convergence rate). Let Assumptions (A1) – (A3) hold, and suppose that \( (\gamma_n) = ((1 - a) n^{-1}) \).

1. If there exists \( c \geq 0 \) such that \( \beta_n^{-1} h_n^5 \to c \), then
\[
\sqrt{\beta_n^{-1} h_n} (r_n (x) - r (x)) \xrightarrow{D} \mathcal{N} \left( \sqrt{\mathcal{E} B_{a, \xi}^{(2)} (x), V_{a, \xi, \beta}^{(2)} (x)} \right),
\]
where
\[
B_{a, \xi}^{(2)} (x) = \frac{1}{2f (x)} \left( \frac{a^{(2)} (x)}{(1 - 2a \xi)} - \frac{(1 - a)}{(1 - 3a)} r (x) f^{(2)} (x) \right) \mu_2 (K),
\]
\[
V_{a, \xi, \beta}^{(2)} (x) = \left\{ \frac{\mathbb{E} [Y^2 | X = x]}{(2 - (\beta - a) \xi f (x)) - (1 - a) \xi f^2 (x)} - (1 - a) \xi f (x) \right\} R (K).
\]

2. If \( nh_n^5 \to \infty \), then
\[
\frac{1}{h_n^2} (r_n (x) - r (x)) \xrightarrow{p} B_{a, \xi}^{(2)} (x).
\]

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

**Corollary 4** (Weak pointwise convergence rate). Let Assumptions (A1) – (A3) hold.

1. If we suppose that the stepsizes \((\gamma_n, \beta_n) = ((1 - a) n^{-1}, n^{-1})\), and if there exists \( c \geq 0 \) such that \( nh_n^5 \to c \), then
\[
\sqrt{nh_n} (r_n (x) - r (x)) \xrightarrow{D} \mathcal{N} \left( \sqrt{\mathcal{E} B_{a, \xi}^{(2)} (x), V_{a, \xi, \beta}^{(2)} (x)} \right).
\]

2. If we suppose that the stepsizes \((\gamma_n, \beta_n) = ((1 - a) n^{-1}, (1 - a) n^{-1})\), and if there exists \( c \geq 0 \) such that \( nh_n^5 \to c \), then
\[
\sqrt{nh_n} (r_n (x) - r (x)) \xrightarrow{D} \mathcal{N} \left( \sqrt{\mathcal{E} B_{a, (1 - a)^{-1}}^{(2)} (x), V_{a, (1 - a)^{-1}, \beta}^{(2)} (x)} \right).
\]

The following proposition gives the \textit{MWISE} of \( r_n \) in the case when \( (\gamma_n) \) is chosen to minimize the variance of \( f_n \).

**Proposition 4** (\textit{MWISE} of \( r_n \)). Let Assumptions (A1) – (A3) hold, and suppose that \((\gamma_n) = ((1 - a) n^{-1})\).
1. If \( a \in ]0, \beta/5[ \), then

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

2. If \( a = \beta/5 \), then

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

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

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 \( (\beta_n) \) and then the corresponding \( MWISE \) depend also on the stepsize \( (h_n) \).

**Corollary 5.** Let Assumptions (A1) - (A3) hold, and suppose that \( (\gamma_n) = ((1-a)n^{-1}) \).

To minimize the \( MWISE \) of \( r_n \), the stepsize \( (\beta_n) \) must be chosen in \( GS(-1) \), the bandwidth \( (h_n) \) must equal

\[
\left( \frac{I_4}{(2-\beta-a)\xi} - (1-a)\xi I_5 \right) R(K) \left\{ \frac{1}{\mu_2^2(K)} \right\}^{1/5} \beta_n^{1/5}.
\]

Then, we have

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

\[
+ o \left( \beta_n^{4/5} \right).
\]

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

**Corollary 6.** Let Assumptions (A1) - (A3) hold, and suppose that \( (\gamma_n) = ((1-a)n^{-1}) \).

To minimize the \( MWISE \) of \( r_n \), the stepsize \( (\beta_n) \) must be chosen in \( GS(-1) \), \( \lim_{n \to \infty} n\beta_n = \beta_0 \), the bandwidth \( (h_n) \) must equal

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

(28)
and we then have

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

(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

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

and we then have

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

(30)

In order to estimate the optimal bandwidth \( [15] \), 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 \( \hat{I}_1 \) respectively of \( \hat{I}_2, \hat{I}_3, \hat{I}_4 \) and \( \hat{I}_5 \), the pilot bandwidth \( (h_n) \) must belong to \( GS(-3/4) \), respectively to \( GS(-3/4), GS(-2/5) \) and \( GS(-2/5) \).

Finally, the plug-in estimator of the bandwidth \( (h_n) \) using the semi-recursive estimators defined in \( [3] \) with the stepizes \( (\gamma_n, \beta_n) = ((1 - a) n^{-1}, n^{-1}) \).

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

(31)

\[
MWISE[r_n] = \frac{5}{4} \frac{1}{2^{4/5}} \frac{5}{3} \left( \hat{I}_4 - \frac{24}{25} \hat{I}_5 \right)^{4/5} \times \left( \hat{I}_1 + \frac{36}{25} \hat{I}_3 - \frac{12}{5} \hat{I}_2 \right)^{1/5} \Theta(K) n^{-4/5} + o(n^{-4/5}).
\]

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

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

(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

\[
\left( \frac{1}{5} \right)^{1/5} \left( \frac{\hat{I}_4 - \hat{I}_5}{\hat{I}_1 + \hat{I}_3 - 2\hat{I}_2} \right)^{1/5} \left\{ \frac{R(K)}{\mu_2^2(K)} \right\}^{1/5} n^{-1/5},
\]

(33)
and it follows from (29), that the plug-in MWISE of the proposed estimator (3) using the stepsizes \((\gamma_n, \beta_n) = ((1 - a) n^{-1}, (1 - a) n^{-1})\), is given by

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

Now, let us recall that the bias and variance of Nadaraya-Watson’s estimator \(\tilde{r}_n\) are given by

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

and

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

It follows that,

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

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

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

and we have

\[
MWISE [\tilde{r}_n] = \frac{5}{4} (I_4 - I_5)^{4/5} (I_1 + I_3 - 2I_2)^{1/5} \Theta (K) n^{-4/5} + o (n^{-4/5}).
\]

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 \(I_5\):

\[
\tilde{I}_1 = \frac{1}{n3^3b_n^6} \sum_{i,j,k=1}^{n} K_b^{(2)} \left( X_i - X_j \right) b_n \left( X_i - X_k \right) b_n Y_j Y_k,
\]

\[
\tilde{I}_2 = \frac{1}{n3^3b_n^6} \sum_{i,j,k=1}^{n} K_b^{(2)} \left( X_i - X_j \right) b_n \left( X_i - X_k \right) b_n Y_i Y_j,
\]

\[
\tilde{I}_3 = \frac{1}{n4^4b_n^6} \sum_{i,j,k,l=1}^{n} K_b^{(2)} \left( X_i - X_j \right) b_n \left( X_i - X_k \right) b_n Y_i Y_l,
\]

\[
\tilde{I}_4 = \frac{1}{n2b_n} \sum_{i,j=1}^{n} K_b \left( X_i - X_j \right) b_n Y_i^2,
\]

\[
\tilde{I}_5 = \frac{1}{n2b_n} \sum_{i,j=1}^{n} K_b \left( X_i - X_j \right) b_n Y_j Y_k,
\]

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 \(\tilde{I}_1\) respectively of \(\tilde{I}_2, \tilde{I}_3, \tilde{I}_4\) and
\[
\tilde{I}_5, \text{ the pilot bandwidth (}b_n\text{) must belong to } G_3(-3/14) \text{, respectively to } G_3(-3/14), G_5(-3/14), G_5(-2/5) \text{ and } G_5(-2/5). \\
\text{Then the plug-in estimator of the bandwidth (}h_n\text{) using the nonrecursive estimator (1), is given by }
\left( \frac{\tilde{I}_4 - \tilde{I}_5}{\tilde{I}_1 + \tilde{I}_3 - 2\tilde{I}_2} \right)^{1/5} \left\{ \frac{R(K)}{\mu_2(K)} \right\}^{1/5} n^{-1/5}, \tag{36}
\]
and the plug-in of the MWISE of the nonrecursive estimator (1), is given by
\[
\text{MWISE} [\tilde{r}_n] = \frac{5}{4} \left( \tilde{I}_4 - \tilde{I}_5 \right)^{4/5} \left( \tilde{I}_1 + \tilde{I}_3 - 2\tilde{I}_2 \right)^{1/5} \Theta(K) n^{-4/5} + o(n^{-4/5}).
\]
Finally, it follows from (14), (22), (30), (32) and (35), that:

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

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

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

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 (1).

When applying \(r_n\) one need to choose three quantities:

- The function \(K\), we choose the Normal kernel.
- The stepsizes \((\gamma_n, \beta_n)\) equal respectively to \((n^{-1}, n^{-1})\), \((n^{-1}, (1 - a) n^{-1})\), \(((1 - a) n^{-1}, n^{-1})\) or \(((1 - a) n^{-1}, (1 - a) n^{-1})\). 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), (24) for (Recursive 2), (33) for (Recursive 3) and (31) for (Recursive 4).

When applying \(\tilde{r}_n\) 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 \( \varepsilon \) is \( \mathcal{N}(0, \sigma) \)-distributed, with \( \sigma \) 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 \( (MSE = n^{-1} \sum_i (r_i - r^*_i)^2) \).

**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 \( \sigma \) 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 propose an automatic selection of the bandwidth of the semi-recursive kernel estimators of a regression function defined by the stochastic approximation algorithm (4). The proposed estimators asymptotically follows 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
| n = 100 | Nadaraya | Recursive 1 | Recursive 2 | Recursive 3 | Recursive 4 |
|--------|----------|-------------|-------------|-------------|-------------|
| 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**|
| CPU    | 835      | 514         | 509         | **464**     | 470         |
| n = 500| MSE      | 0.000284    | 0.000279    | 0.000294    | **0.000217**|
| CPU    | 3679     | 2185        | 1973        | 1966        | **1865**    |

| n = 100 | MSE      | 0.004486    | 0.004447    | 0.004286    | **0.003729**|
| CPU    | 231      | 143         | 135         | 137         | **129**     |
| n = 200| MSE      | 0.002331    | 0.002337    | 0.002142    | **0.001929**|
| CPU    | 885      | 568         | 549         | 485         | **457**     |
| n = 500| MSE      | 0.001372    | 0.001411    | 0.001265    | **0.001174**|
| CPU    | 3498     | 2049        | 1943        | 2242        | **2045**    |

| n = 100 | 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**    |

Table 1: Quantitative comparison between the nonrecursive estimator (4) and four recursive estimators; recursive 1 correspond to the estimator (3) with the choice \((\gamma_n, \beta_n) = (n^{-1}, n^{-1})\), 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 \(r(x) = \cos(x)\), \(X \sim \mathcal{N}(0, 1)\) 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.

The simulation study confirms the nice features of our proposed semi-recursive estimators and satisfactory 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 (1964) and Watson (1964). Moreover, we plan to make an extension of our method in future and to consider the case of the averaged Révész’s regression estimators (see Mokkadem et al. (2009b) and Slaoui (2015a,b)).
|        | Nadaraya  | Recursive 1 | Recursive 2 | Recursive 3 | Recursive 4 |
|--------|-----------|-------------|-------------|-------------|-------------|
| $n = 100$ | 1.31e-04 | 1.15e-04 | **6.22e-05** | 1.71e-04 | 1.03e-04 |
| CPU   | 249       | 184        | **135**     | 146         | 146         |
| $n = 200$ | 4.38e-05 | 3.87e-05 | **1.50e-05** | 8.03e-05 | 3.63e-05 |
| CPU   | 909       | 524        | **475**     | 601         | 458         |
| $n = 500$ | 5.70e-06 | 5.02e-06 | **3.20e-06** | 2.32e-05 | 4.29e-06 |
| CPU   | 3708      | 1803       | **1672**    | 1855        | 1483        |

|        | $\sigma = 0.1$ | $\sigma = 0.5$ | $\sigma = 2$ |
|--------|----------------|----------------|---------------|
| $n = 100$ | 0.000351 | 0.000325 | **0.000252** | 0.000350 | 0.000296 |
| CPU   | 256       | 144        | **132**      | 134         | 125         |
| $n = 200$ | 0.000189 | 0.000171 | 0.000154 | 0.000163 | **0.000151** |
| CPU   | 873       | 524        | **483**      | 576         | 451         |
| $n = 500$ | 2.30e-05 | 2.25e-05 | 2.42e-05 | 3.351e-05 | **2.06e-05** |
| CPU   | 4389      | 2113       | **1987**     | 1999        | 1973        |

|        | $\sigma = 0.5$ | $\sigma = 2$ |
|--------|----------------|---------------|
| $n = 100$ | 0.003447 | 0.003294 | 0.003155 | **0.003132** | 0.003137 |
| CPU   | 294       | 155        | **173**     | 143         | 148         |
| $n = 200$ | 0.000160 | 0.000152 | 0.000162 | **0.000111** | 0.000189 |
| CPU   | 917       | 503        | **581**     | 515         | 477         |
| $n = 500$ | 6.56e-05 | 7.03e-05 | **5.01e-05** | 6.70e-05 | 5.39e-05 |
| CPU   | 3643      | 2105       | **1951**    | 1947        | 1877        |

Table 2: Quantitative comparison between the nonrecursive estimator (1) 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.

and the case of time series as in [Hart and Vieu (1990)] in recursive way (see [Huang et al. (2014)]).
Figure 1: The daily carbon dioxide measurements data with automatically bandwidth selection using the nonrecursive Nadaraya’s estimator (4) and two semi-recursive estimators (Recursive 1 and Recursive 4).

A Proofs

Throughout this section we use the following notation:

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

\[ W_n(x) = h_n^{-1} K \left( \frac{x - X_n}{h_n} \right). \tag{37} \]

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

Let us first state the following technical lemma.

**Lemma 1.** Let \((v_n) \in \mathcal{G}\mathcal{S}(v^*), (\beta_n) \in \mathcal{G}\mathcal{S}(-\beta), \) and \(m > 0 \) such that \(m - v^*\xi > 0 \) where \(\xi\) is defined in (4). We have

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

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

\[ \lim_{n \to +\infty} v_n Q_n^m \left[ \sum_{k=1}^{n} Q_k^{-m} \frac{\beta_k}{v_k} \alpha_k + \delta \right] = 0. \tag{40} \]
Lemma 1 is widely applied throughout the proofs. Let us underline that it is its application, which requires Assumption (A2)(iii) on the limit of \((n\gamma_n)\) as \(n\) goes to infinity.

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

### A.1 Proof of Proposition 1

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

\[
    r_n(x) - r(x) = B_n(x) \frac{f(x)}{f_n(x)} ,
\]

with

\[
    B_n(x) = \frac{1}{f(x)} (a_n(x) - a(x)) - \frac{r(x)}{f(x)} (f_n(x) - f(x)) .
\]

It follows from Lemma 1, 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. (2009a).

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

1. If \(a \in \left[0, \frac{1}{5}\right]\), then

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

If \(a \in \left[1/5, 1\right]\), then

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

2. If \(a \in \left[1/5, 1\right]\), then

\[
    \text{Var} [f_n(x)] = \frac{1}{1 + a n h_n} f(x) R(K) + o \left( \frac{1}{n h_n} \right) .
\]

If \(a \in \left[0, 1/5\right]\), then

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

Following similar steps as the proof of the Proposition 1 of Mokkadem et al. (2009a), we show that

**Lemma 3 (Bias and variance of \(a_n\)).** Let Assumptions (A1) – (A3) hold.

1. If \(a \in \left[0, \frac{1}{5}\right]\), then

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

If \(a \in \left[\frac{1}{5}, 1\right]\), then

\[
    \mathbb{E} [a_n(x)] - a(x) = o \left( \sqrt{\beta_n h_n^{-1}} \right) .
\]
2. If \( a \in [\beta/5, 1[ \), then
\[
Var [a_n (x)] = E[Y^2 | X = x] f (x) \beta_n \frac{R (K)}{h_n} + o \left( \frac{\beta_n}{h_n} \right).
\] (49)

If \( a \in ]0, \beta/5[ \), then
\[
Var [a_n (x)] = o (h_n^4).
\] (50)

Then, (7) follows from (43), (47) and (41) and (8) follows from (44), (48) and (41).

Now, it follows from (42) that
\[
Var [B_n (x)] = 1 f^2 (x) \left\{ Var [a_n (x)] + r^2 (x) Var [f_n (x)] - 2r (x) Cov (a_n (x), f_n (x)) \right\}.
\] (51)

In view of (A3), and with the choice of the stepsize \((\gamma_n) = (n^{-1})\) and using Lemma 1, classical computations gives
\[
Cov (a_n (x), f_n (x)) = \frac{\xi}{1 + a \xi} h_n r (x) f (x) R (K) + o \left( \frac{\beta_n}{h_n} \right).
\] (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

Following similar steps as the proof of the Proposition 2 of Mokkadem et al. (2009a), we proof the Proposition 2.

A.3 Proof of Theorem 1

Let us at first assume that, if \( a \geq \beta/5 \), then
\[
\sqrt{\beta_n^{-1} h_n (r_n (x) - E [r_n (x)])} \overset{D}{\to} \mathcal{N} \left( 0, V^{(1)}_{a, \xi, \beta} \right).
\] (53)

In the case when \( a > \beta/5 \), Part 1 of Theorem 1 follows from the combination of (8) and (53). In the case when \( a = \beta/5 \), Parts 1 and 2 of Theorem 1 follow from the combination of (7) and (53). In the case \( a < \beta/5 \), (10) implies that
\[
h_n^{-2} (r_n (x) - E (r_n (x))) \overset{p}{\to} 0,
\]
and the application of (7) gives Part 2 of Theorem 1.

We now prove (53). In view of (42), we have
\[
B_n (x) - E [B_n (x)] = \frac{1}{f (x)} Q_n \sum_{k=1}^{n} (T_k (x) - E [T_k (x)]),
\] (54)
with
\[
T_k (x) = Q^{-1}_k \left( \beta_k Z_k (x) - r (x) \zeta_n \xi_k^{-1} \gamma_k W_k (x) \right).
\] (55)
In the case when \((\gamma_n) = (n^{-1})\), we have \(\zeta_n = (nQ_n)^{-1}\) et \(\zeta_k^{-1} = Q_k\), then

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

Set

\[
Y_k(x) = T_k(x) - \mathbb{E}(T_k(x)).
\]

Moreover, we have

\[
v_n^2 = \sum_{k=1}^{n} \text{Var}(Y_k(x))
\]

\[
= \sum_{k=1}^{n} Q_k^{-2} \beta_k^2 \text{Var}(Z_k(x)) + r^2(x)(nQ_n)^{-2} \sum_{k=1}^{n} \text{Var}(W_k(x))
\]

\[
-2r(x)(nQ_n)^{-1} \sum_{k=1}^{n} Q_k^{-1} \beta_k \text{Cov}(Z_k(x), W_k(x)).
\]

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

\[
\text{Var}(Z_k(x)) = \frac{1}{h_k} \mathbb{E}[Y^2 | X = x] f(x)R(K) + o(1),
\]

\[
\text{Var}(W_k(x)) = \frac{1}{h_k} [f(x)R(K) + o(1)],
\]

\[
\text{Cov}(Z_k(x), W_k(x)) = \frac{1}{h_k} [r(x)f(x)R(K) + o(1)].
\]

The application of Lemma 1 ensures that

\[
v_n^2 = \sum_{k=1}^{n} Q_k^{-2} \beta_k^2 \frac{1}{h_k} \left[\mathbb{E}[Y^2 | X = x] f(x)R(K) + o(1)\right]
\]

\[
+ \frac{r(x)}{n^2 Q_n^2} \sum_{k=1}^{n} \frac{1}{h_k} [f(x)R(K) + o(1)]
\]

\[
-2 \frac{r(x)}{n Q_n} \sum_{k=1}^{n} Q_k^{-1} \beta_k \frac{1}{h_k} [r(x)f(x)R(K) + o(1)]
\]

\[
= \frac{f^2(x) \beta_n}{Q_n^2 h_n} V_{a,\xi,\beta}^{(1)} + o(1).
\]

On the other hand, we have, for all \(p > 0\),

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

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 1 we get

\[
\sum_{k=1}^{n} \mathbb{E}[|Y_k(x)|^{2+p}] = O\left(\sum_{k=1}^{n} Q_k^{-2-p} \beta_k^{2+p} \mathbb{E}[|T_k(x)|^{2+p}]\right)
\]

\[
= O \left(\sum_{k=1}^{n} Q_k^{-2-p} \beta_k^{2+p} \frac{1}{h_k^{1+p}}\right)
\]

\[
= O \left(\frac{\beta_1^{1+p}}{Q_n^{2+p} h_n^{1+p}}\right),
\]
and we thus obtain
\[
\frac{1}{v_n^{2+p}} \sum_{k=1}^{n} E \left[ |Y_k(x)|^{2+p} \right] = O \left( \left[ \beta_n h_n^{-1} \right]^{p/2} \right) = o(1) .
\]

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

### A.4 Proof of Proposition 3

The following Lemma follows from the Proposition 1 of Mokkadem et al. (2009a).

**Lemma 4 (Bias and variance of \(f_n)\).** Let Assumptions \((A_1) – (A_3)\) and suppose that the stepsize \((\gamma_n) = ([1 - a]n^{-1}).\)

1. If \(a \in ]0, 1/5]\), then
\[
E \left[ f_n (x) \right] - f (x) = \frac{1 - a}{2 (1 - 3a)} f^{(2)} (x) h_n^2 \mu_2 (K) + o \left( h_n^2 \right) . \tag{57}
\]

If \(a \in ]1/5, 1]\), then
\[
E \left[ f_n (x) \right] - f (x) = o \left( \sqrt{n^{-1}h_n^{-1}} \right) . \tag{58}
\]

2. If \(a \in ]1/5, 1]\), then
\[
Var \left[ f_n (x) \right] = \frac{1 - a}{nh_n} R (K) + o \left( \frac{1}{nh_n} \right) . \tag{59}
\]

If \(a \in ]0, 1/5]\), then
\[
Var \left[ f_n (x) \right] = o \left( h_n^4 \right) . \tag{60}
\]

Then, \((23)\) follows from \((57), (47)\) and \((41)\) and \((25)\) follows from \((58), (48)\) and \((41)\). Moreover, in view of \((A_3)\), and using the choice of the stepsize \((\gamma_n) = ([1 - a]n^{-1})\) and using Lemma [4] classical computations gives
\[
Cov \left( a_n (x), f_n (x) \right) = (1 - a) \xi\beta \frac{n h}{n Q} r (x) f (x) R (K) + o \left( \frac{\beta}{h_n} \right) . \tag{61}
\]

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 4

Following similar steps as the proof of the Proposition 2 of Mokkadem et al. (2009a), we proof the Proposition 4.

### A.6 Proof of Theorem 2

Following similar steps as the proof of the Theorem 1 and using the fact that in the case when \((\gamma_n) = ([1 - a]n^{-1})\), we have \(Q_k^{-1} \zeta_k^{-1} \gamma_k = (1 - a) h_k / (nh_n Q_n)\), and then it follows from \((55)\), that
\[
T_k (x) = Q_k^{-1} \beta_k Z_k (x) - r (x) \frac{(1 - a) h_k}{nh_n Q_n} W_k (x) ,
\]
we prove Theorem 2.
References

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

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

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

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

Hart, J. D. and Vieu, P. (1990). Data-Driven Bandwidth Choice for Density Estimation Based on Dependent Data. *Ann. Statist.*, **18**, 873–890.

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

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

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

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

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

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.

Révész, P. (1977). How to apply the method of stochastic approximation in the nonparametric estimation of a regression function. *Math. Operationsforsch. Statist., Ser. Statistics*, **8**, 119–126.

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

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

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

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

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

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

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

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

Université de Poitiers, Laboratoire de Mathématiques et Application, Futuroscope Chasseneuil, France
E.mail: Yousri.Slaoui@math.univ-poitiers.fr