Recursive regression estimation based on the two-time-scale stochastic approximation method and Bernstein polynomials

Yousri Slaoui, Salima Helali

Research Article

Yousri Slaoui* and Salima Helali

Recursive regression estimation based on the two-time-scale stochastic approximation method and Bernstein polynomials

https://doi.org/10.1515/mcma-2022-2104
Received March 10, 2021; revised January 25, 2022; accepted January 30, 2022

Abstract: In this paper, we propose a recursive estimators of the regression function based on the two-time-scale stochastic approximation algorithms and the Bernstein polynomials. We study the asymptotic properties of this estimators. We compare the proposed estimators with the classic regression estimator using the Bernstein polynomial defined by Tenbusch. Results showed that, our proposed recursive estimators can overcome the problem of the edges associated with kernel regression estimation with a compact support. The proposed recursive two-time-scale estimators are compared to the non-recursive estimator introduced by Tenbusch and the performance of the two estimators are illustrated via simulations as well as two real datasets.

Keywords: Two-time-scale stochastic approximation algorithms, Bernstein polynomials, regression estimation

MSC 2010: 62G08, 62L20, 65D10

1 Introduction

Nonparametric regression estimation methods have attracted a great deal of attention as researchers have realized that parametric regression is not suitable for adequately fitting curves to many datasets that arise in practice. There are many reasons for choosing the nonparametric regression methods, no assumption should be made on the form of the regression function, the complexity of the model will be determined completely by the data, it is applicable for various design situations and it is easy to interpret.

There have been several monograph on the nonparametric regression estimation, see, e.g., [5, 8, 10, 25], where it is shown that nonparametric regression techniques have much to offer in applications, such as observational astronomy, forecasting future opportunities and risks in business, causal relationships between parameters in biological systems and various other situations.

The most famous nonparametric estimator of the regression function $r : x \to \mathbb{E}(Y \mid X = x)$, was proposed by Nadaraya [26] and Watson [39]. In the current work, we are concerned by the recursive estimation of a regression function. Recursive estimation have become an increasingly important area of research. In many situations, data arrives regularly so that it is impossible to store them in a traditional database. In such a context, building a recursive estimator which does not require to store all the data in memory is of great interest.

***Corresponding author: Yousri Slaoui,** Laboratoire de Mathématiques et Applications, UMR 7348 du CNRS Site du Futuroscope, Université de Poitiers, Téléport 2, Bâtiment H3 11 Bd Marie et Pierre Curie, 86073 Poitiers cedex 9, France, e-mail: yousri.slaoui@math.univ-poitiers.fr. https://orcid.org/0000-0001-5295-3311
Salima Helali, Laboratoire Angevin de Recherche en Mathématiques, Faculté des Sciences Bâtiment I, Université d'Angers, 2 Boulevard Lavoisier 49045 Angers cedex 01, France, e-mail: salima.helali@univ-angers.fr

The recursive regression estimation was introduced first by Kiefer and Wolfowitz [16], extended by Révész [28] using the Robbins–Monro's procedure, generalized by Mokkadem, Pelletier and Slaoui [24] and studied by Slaoui [30–32]. The semi-recursive approach was introduced in [33], while the recursive regression estimation for independent functional data was established in [34, 35].

However, kernel methods and their recursive improvements have estimation problems in the edges when the regression function has a bounded support. To overcome this problem, their have been many methods such as the approach of regression estimation using Bernstein polynomials proposed by Tenbusch [37].

Bernstein polynomial was used in several directions to estimate a density of probability and distribution functions. See for instance, the original work of [38], extended in [1, 7, 11–14, 19–21, 27, 36, 37].

Let $(X, Y), (X_1, Y_1), \ldots, (X_n, Y_n)$ be independent, identically distributed pairs of random variables with joints density function $g(x, y)$ and let f such that $f(x) = \int_{\mathbb{R}} g(x, t)\, dt$, denote the density of X which is supported on $[0, 1]$. We denote by

$$a(x) = f(x)r(x), \quad Z_{k,m} = \mathbb{1}_{\{\frac{k}{m} < X_i \frac{k+1}{m}\}} b_k(m-1, x),$$

where $b_k(m, x) = C_k^m x^k (1-x)^{m-k}$ is the Bernstein polynomial of order m and let $\mathbb{E}(|Y|) < \infty$. The non-recursive estimator defined in [37] such that $f(x) \neq 0$, for $x \in [0, 1]$,

$$\widehat{r}_n(x) = \begin{cases} \dfrac{\sum_{i=1}^n Y_i \sum_{k=1}^m Z_{k,m}}{\sum_{i=1}^n \sum_{k=1}^m Z_{k,m}} & \text{if } \displaystyle\sum_{k=1}^m Z_{k,m} \neq 0, \\ 0 & \text{if } \displaystyle\sum_{k=1}^m Z_{k,m} = 0, \end{cases} \tag{1.1}$$

The aim of this paper is to introduce two-time-scale stochastic approximation algorithms in order to define a class of recursive estimators of a regression function based on Bernstein polynomials. The two-time-scale stochastic approximation algorithms have been defined in [2, 3, 17], their convergence rate studied in [18, 22, 23, 29].

In order to construct a recursive regression estimator defined by the two-time-scale stochastic algorithm, at a point x such as $f(x) \neq 0$, we define an algorithm of search of the common zero of the functions:

$$l_1 : (y, z) \mapsto f(x) - y \quad \text{and} \quad l_2 : (y, z) \mapsto \frac{r(x)f(x)}{y} - z.$$

We proceed in the following way, for $x \in [0, 1]$:

(i) $f_0(x) > 0$ and $r_0(x) \in \mathbb{R}$,

(ii) for $n \geq 1$,

$$\begin{cases} f_n(x) = f_{n-1}(x) + \gamma_n W_n^{(1)}(x), \\ r_n(x) = r_{n-1}(x) + \beta_n W_n^{(2)}(x), \end{cases}$$

where $W_n^{(1)}$ and $W_n^{(2)}$ are the observations of the functions l_1 and l_2 at the point $(f_{n-1}(x), r_{n-1}(x))$, and where the stepsizes (γ_n) and (β_n) are two sequences of positive real numbers that go to zero such that $\gamma_n \leq 1$ and $\lim_{n\to\infty} \beta_n \gamma_n^{-1} = 0$. Then, we estimate $f(x)$ and $a(x)$ respectively by

$$m \sum_{k=0}^{m-1} \mathbb{1}_{\{\frac{k}{m} < X_n \leq \frac{k+1}{m}\}} b_k(m-1, x) \quad \text{and} \quad mY_n \sum_{k=0}^{m-1} \mathbb{1}_{\{\frac{k}{m} < X_n \leq \frac{k+1}{m}\}} b_k(m-1, x).$$

Then the two-time-scale stochastic approximation algorithm can be rewritten as

$$\begin{cases} f_n(x) = (1-\gamma_n) f_{n-1}(x) + \gamma_n m \displaystyle\sum_{k=0}^{m-1} \mathbb{1}_{\{\frac{k}{m} < X_n \leq \frac{k+1}{m}\}} b_k(m-1, x), \\ r_n(x) = (1-\beta_n) r_{n-1}(x) + \dfrac{\beta_n m Y_n}{f_{n-1}(x)} \displaystyle\sum_{k=0}^{m-1} \mathbb{1}_{\{\frac{k}{m} < X_n \leq \frac{k+1}{m}\}} b_k(m-1, x). \end{cases} \tag{1.2}$$

The aim of this paper is to study the properties of the regression estimators defined in (1.2), as a competitor for the non-recursive estimator (1.1).

The remainder of this paper is organized as follows. In the next section, we present the assumptions and notations that we need to give the proprieties of our proposed two-time-scale estimators (1.2). In Section 3, we state our main results. Section 4 is devoted to give some numerical comparison between our proposed recursive estimators (1.2) and the non-recursive estimator 3 through some simulation studies and then through two real datasets. Some concluding remarks and possible future developments are mentioned in Section 6. While, all the mathematical developments are deferred to Section 7.

2 Assumptions and notations

We define the following class of regularly varying sequences.

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

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

This condition was introduced by [6] to define regularly varying sequences. To study asymptotic the behaviors of the estimator r_n defined by equations (1.2) inside the interval $[0, 1]$, we consider the following assumptions for $a \in (0, \frac{1}{3})$:

(A1) (i) $(m_n) \in \mathcal{GS}(a)$ with $a \in (0, 1)$.
(ii) $(\beta_n) \in \mathcal{GS}(-\beta)$ with $\beta \in (3a, 1]$.
(iii) $(\gamma_n) \in \mathcal{GS}(-\alpha)$ with $\alpha \in (\min\{3a, a + 1 - \beta\}, \beta]$ such that $\lim_{n\to\infty} \beta_n^{-1}\gamma_n(\ln \sum_{k=1}^n \gamma_k)^{-1} = \infty$.
(iv) $\lim_{n\to\infty}(n\beta_n) > \min\{a, \frac{1}{2}(\beta - \frac{a}{2})\}\mathbb{1}_{\{x\in(0,1)\}} + \min\{a, \frac{1}{2}(\beta - a)\}\mathbb{1}_{\{x\in\{0,1\}}}$.

(A2) (i) $v \to g(v, w)$ is continuous on $\mathbb{R}$.
(ii) For $t > 0$, $v \to \int_{\mathbb{R}} |w|^t g(v, w)\, dw$ is a bounded function.

Discussion of the assumptions:

(i) Assumptions (A1) and (A2) are regularity conditions which permit us to evaluate the bias term and the variance term of the proposed estimator.

(ii) The intuition behind the use of such bandwidth (h_n) belonging to $\mathcal{GS}(-a)$ is that the ratio $\frac{h_{n-1}}{h_n}$ is equal to $1 + \frac{a}{n} + o(\frac{1}{n})$, similarly, we use the stepsize (β_n) belonging to $\mathcal{GS}(-\beta)$; then the ratio $\frac{\beta_{n-1}}{\beta_n}$ is equal to $1 + \frac{\beta}{n} + o(\frac{1}{n})$. The application of Lemma 1 ensures that the bias and the variance will depend only on h_n and β_n and not on $h_1, \ldots, h_n$ and $\beta_1, \ldots, \beta_n$; then the MISE will depend only on h_n and β_n, which will be helpful to deduce an optimal bandwidth and an optimal stepsize.

(iii) In order to help the readers to follow the main results obtained in this paper, we underline that the application of Lemma 1 under assumption (A2) ensures

$$\Pi_n \sum_{k=1}^n \Pi_k^{-1}\beta_k = 1 + o(1),$$

$$\Pi_n \sum_{k=1}^n \Pi_k^{-1}\beta_k h_k^2 = O(h_n^2),$$

$$\Pi_n^2 \sum_{k=1}^n \Pi_k^{-2}\beta_k^2 h_k^{-1} = O(\beta_n h_n^{-1}).$$

(iv) Assumption (A2)(iii) on the limit of $(n\beta_n)$ as n goes to infinity is usual in the framework of stochastic approximation algorithms. It implies in particular that the limit of $([n\beta_n]^{-1})$ is finite.

Throughout this paper, we will use the following notations:

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

$$\psi(x) = (4\pi x(1-x))^{-\frac{1}{2}},$$

$$\mathcal{W}_n(x) = m_n Y_n \sum_{k=0}^{m_n-1} \mathbb{1}_{\{\frac{k}{m_n} < X_n \leq \frac{k+1}{m_n}\}} b_k(m_n - 1, x),$$

$$\mathcal{Z}_n(x) = m_n \sum_{k=0}^{m_n-1} \mathbb{1}_{\{\frac{k}{m_n} < X_n \leq \frac{k+1}{m_n}\}} b_k(m_n - 1, x),$$

$$\Delta_1(x) = \frac{1}{2}[(1-2x)f'(x) + x(1-x)f^2(x)],$$

$$\Delta_2(x) = \frac{1}{2}[(1-2x)(r'(x)f(x) - r(x)f'(x)) + x(1-x)(r^{(2)}(x)f(x) + r(x)f^{(2)}(x) + 2f'(x)r'(x))],$$

$$\mathrm{Bias}(x) = \frac{(\Delta_2(x) - r(x)\Delta_1(x))}{f(x)},$$

$$\mathcal{V}(x) = \frac{\mathrm{Var}[Y \mid X = x]\psi(x)}{f(x)(2 - (\beta - \frac{a}{2})\xi)} \mathbb{1}_{\{x \in (0,1)\}} + \frac{\mathrm{Var}[Y \mid X = x]}{f(x)(2 - (\beta - a)\xi)} \mathbb{1}_{\{x \in \{0,1\}\}}.$$

3 Main results

Throughout this paper we consider the two following sequences:

$$(v_n) = (m_n^{\frac{1}{2}} \mathbb{1}_{\{x \in (0,1)\}} + m_n \mathbb{1}_{\{x \in \{0,1\}\}}) \quad \text{and} \quad (\mathcal{V}_n) = (\beta_n^{-\frac{1}{2}} v_n^{-1}).$$

For simplicity, we let

$$\mathcal{A}^1_{a,x} = \left\{a, x : a \in \left(0, \frac{\beta}{3}\right],\ x \in \{0,1\}\right\}, \quad \mathcal{A}^2_{a,x} = \left\{a, x : a \in \left(0, \frac{2}{5}\beta\right],\ x \in (0,1)\right\},$$

$$\mathcal{A}^3_{a,x} = \left\{a, x : a \in \left(\frac{\beta}{3}, 1\right),\ x \in \{0,1\}\right\}, \quad \mathcal{A}^4_{a,x} = \left\{a, x : a \in \left(\frac{2}{5}\beta, 1\right),\ x \in (0,1)\right\}.$$

Our first result is the following proposition, which gives the bias and the variance of the proposed recursive estimator r_n.

Proposition 1 (Bias and variance of r_n). *Let Assumptions* (A1)–(A2) *hold such as* $f(x) > 0$. *Then*

$$\mathbb{E}[r_n(x)] - r(x) = m_n^{-1}\,\mathrm{Bias}(x)\frac{1}{1-a\xi}\mathbb{1}_{\mathcal{A}^1_{a,x}\cup\mathcal{A}^2_{a,x}}[1+o(1)] + \mathbb{1}_{\mathcal{A}^4_{a,x}} o\Big(\sqrt{\beta_n m_n^{\frac{1}{2}}}\Big) + \mathbb{1}_{\mathcal{A}^3_{a,x}} o\Big(\sqrt{\beta_n m_n}\Big)$$

and

$$\begin{aligned}\mathrm{Var}[r_n(x)] = {} & \beta_n m_n^{\frac{1}{2}} \frac{\mathrm{Var}[Y \mid X = x]\psi(x)}{2 - (\beta - \frac{a}{2})\xi} \mathbb{1}_{\mathcal{A}^4_{a,x}}[1 + o_x(1)] \\ & + \beta_n m_n \frac{\mathrm{Var}[Y \mid X = x]}{2 - (\beta - a)\xi} \mathbb{1}_{\mathcal{A}^3_{a,x}}[1 + o(1)] + \mathbb{1}_{\mathcal{A}^1_{a,x}\cup\mathcal{A}^2_{a,x}} o(m_n^{-2}).\end{aligned} \tag{3.1}$$

The following proposition gives the mean squared error (MSE) of the proposed two-time-scale recursive estimators r_n.

Proposition 2 (MSE of r_n). *Let Assumptions* (A1)–(A2) *hold. Then*

$$\begin{aligned}\mathrm{MSE}[r_n(x)] = {} & m_n^{-2}\frac{\mathrm{Bias}^2(x)}{(1-a\xi)^2}\mathbb{1}_{\mathcal{A}^1_{a,x}\cup\mathcal{A}^2_{a,x}}[1+o(1)] + \beta_n m_n^{\frac{1}{2}}\frac{\mathrm{Var}[Y \mid X = x]\psi(x)}{2 - (\beta - \frac{a}{2})\xi}\mathbb{1}_{\mathcal{A}^4_{a,x}}[1+o_x(1)] \\ & + \beta_n m_n \frac{\mathrm{Var}[Y \mid X = x]}{2 - (\beta - a)\xi}\mathbb{1}_{\mathcal{A}^3_{a,x}}[1+o(1)] + \mathbb{1}_{\mathcal{A}^1_{a,x}\cup\mathcal{A}^2_{a,x}} o(m_n^{-2}) \\ & + \mathbb{1}_{\mathcal{A}^4_{a,x}} o(\beta_n m_n^{\frac{1}{2}}) + \mathbb{1}_{\mathcal{A}^3_{a,x}} o(\beta_n m_n).\end{aligned}$$

To minimize the MSE of $r_n(x)$ for $x \in (0,1)$ such that $\psi(x) > 0$, the stepsize (β_n) must be chosen in $\mathcal{GS}(-1)$ and (m_n) must be in $\mathcal{GS}(\frac{2}{5})$. To minimize the MSE of $r_n(x)$ for $x \in \{0,1\}$ such that $\psi(x) > 0$, the stepsize (β_n) must be chosen in $\mathcal{GS}(-1)$ and (m_n) must be in $\mathcal{GS}(\frac{1}{3})$.

Corollary 1. *Let Assumptions* (A1)–(A2) *hold. To minimize the* MSE *of the proposed two-time-scale recursive estimators* $r_n(x)$, *the stepsize* (γ_n) *must be chosen equal to* (n^{-1}). *The optimal order* (m_n) *should be equal to*

$$\left(2^{\frac{6}{5}}\left(\frac{5}{3}\right)^{\frac{2}{5}}\left(\frac{\mathrm{Bias}^{\frac{4}{5}}(x)}{\psi(x)^{\frac{2}{5}}\mathrm{Var}[Y \mid X = x]^{\frac{2}{5}}}\right)n^{\frac{2}{5}}\mathbb{1}_{\{x\in(0,1)\}} + 6^{\frac{1}{3}}\left(\frac{\mathrm{Bias}^{\frac{2}{3}}(x)}{\mathrm{Var}[Y \mid X = x]^{\frac{1}{3}}}\right)n^{\frac{1}{3}}\mathbb{1}_{\{x\in\{0,1\}\}}\right),$$

and the corresponding MSE *must be equal to*

$$\mathrm{MSE}[\widehat{r}_n(x)] = \frac{5}{4}2^{-\frac{2}{5}}\Big(\frac{5}{3}\Big)^{\frac{6}{5}}\psi(x)^{\frac{4}{5}}\mathrm{Var}[Y \mid X = x]^{\frac{4}{5}}\,\mathrm{Bias}^{\frac{2}{5}}(x)n^{-\frac{4}{5}}[1 + o(1)]\mathbb{1}_{\{x\in(0,1)\}} + \frac{3}{3}2^{-\frac{1}{3}}\Big(\frac{3}{2}\Big)^{\frac{4}{3}}\mathrm{Var}[Y \mid X = x]^{\frac{2}{3}}\,\mathrm{Bias}^{\frac{2}{3}}(x)n^{-\frac{2}{3}}[1 + o(1)]\mathbb{1}_{\{x\in\{0,1\}\}}.$$

Finally, the following proposition shows the asymptotic normality of the recursive estimators r_n.

Theorem 1 (Asymptotic normality of r_n). *The following statements hold:*

(i) *In the case when $x \in (0, 1)$, if there exists $c \geq 0$ such that $\beta_n^{-1}m_n^{-\frac{5}{2}} \to c$ (resp. the case when $x \in \{0, 1\}$, if $\beta_n^{-1}m_n^{-3} \to c$), then*

$$\mathcal{V}_n(r_n(x) - r(x)) \xrightarrow{\mathcal{D}} \mathcal{N}\Big(\frac{\sqrt{c}}{(1 - a\xi)}\,\mathrm{Bias}(x), \mathcal{V}(x)\Big).$$

(ii) *In the case when $x \in (0, 1)$, if $\beta_n^{-1}m_n^{-\frac{5}{2}} \to \infty$ (resp. the case when $x \in \{0, 1\}$, if $\beta_n^{-1}m_n^{-3} \to \infty$), then*

$$m_n(r_n(x) - r(x)) \xrightarrow{\mathbb{P}} \frac{1}{1 - a\xi}\,\mathrm{Bias}(x).$$

4 Numerical studies

The aim of this subsection is to compare our proposed two-time-scale estimators (1.2) with the non-recursive estimator (1.1), through a simulation. We consider the regression model

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

where $X \sim \mathcal{N}(0, 1)$ and $\varepsilon \sim \mathcal{N}(0, \sigma^2)$, where $\sigma \in \{0.05, 0.1, 1\}$. When using our proposed two-time-scale algorithm (1.2), the stepsizes (γ_n, β_n) are chosen to be equal to $(n^{-0.98}, 0.31n^{-0.99})$. Moreover, in order to select the smoothing parameter (m_n), we consider a Monte Carlo procedure for each point $x \in [0, 1]$. We determine the parameter m for $1 \leq m \leq 400$ by minimizing

$$\frac{1}{N}\sum_{i=1}^{N}[r_{n,m}^i(x) - r(x)]^2,$$

where $r_{n,m}^i$ is the estimator of $r_{n,m}$ computed from the ith sample of size n. We choose $N = 500$ trials for the Monte Carlo simulation. In our simulation study, we consider three sample sizes; $n = 50$, $n = 100$, $n = 200$ and the following regression functions:

(a) $r(x) = \cos(x)$,
(b) $r(x) = (1 + \exp(x))^{-1}$,
(c) $r(x) = 1 + 0.6x$.

For each model and sample size n, we approximate the average integrated squared error ($\overline{\mathrm{ISE}}$) and the integrated absolute error ($\overline{\mathrm{IAE}}$) of the estimator using $N = 500$ trials of sample size n:

$$\overline{\mathrm{ISE}} = \frac{1}{N}\sum_{k=1}^{N}\mathrm{ISE}[\overline{r_k}], \quad \overline{\mathrm{IAE}} = \frac{1}{N}\sum_{k=1}^{N}\mathrm{IAE}[\overline{r_k}],$$

where r_k is the estimator computed from the kth sample, and

$$\mathrm{ISE}(r_k) = \int_0^1 (r_k(x) - r(x))^2\,dx, \quad \mathrm{IAE}(r_k) = \int_{\mathbb{R}} |r_k(x) - r(x)|\,dx.$$

Table 1 shows that in terms of the average ISE and IAE of the two considered estimators, the proposed two-time-scale recursive estimators (1.2) have a smaller $\overline{\mathrm{ISE}}$ and smaller $\overline{\mathrm{IAE}}$ compared to the non-recursive estimator (1.1) by using any one of the three regression functions and by considering any one of the three different noise variance.

			ISE		IAE	
Model		***n***	**Recursive**	**Non-recursive**	**Recursive**	**Non-recursive**
$\sigma = 0.05$	(a)	50	**0.108152**	0.263466	**0.278830**	0.512572
		100	**0.146173**	0.275628	**0.342211**	0.523425
		200	**0.172898**	0.265569	**0.385824**	0.514904
	(b)	50	**0.149485**	0.261717	**0.374100**	0.511469
		100	**0.180104**	0.250942	**0.416693**	0.500793
		200	**0.219903**	0.256299	**0.461882**	0.506157
	(c)	50	**0.192147**	0.246308	**0.375262**	0.495552
		100	**0.166492**	0.241962	**0.345006**	0.491480
		200	0.315283	**0.242560**	**0.488220**	0.492026
$\sigma = 0.1$	(a)	50	**0.237895**	0.279316	**0.459012**	0.527103
		100	**0.142256**	0.273119	**0.339965**	0.521741
		200	**0.205788**	0.261468	**0.429494**	0.510851
	(b)	50	**0.240268**	0.262336	**0.478994**	0.512000
		100	**0.186827**	0.263747	**0.422515**	0.513192
		200	**0.168894**	0.260439	**0.403971**	0.510155
	(c)	50	0.241451	**0.236557**	**0.432593**	0.485448
		100	**0.169090**	0.2400389	**0.354381**	0.489659
		200	0.669683	**0.237169**	0.714953	**0.486410**
$\sigma = 1$	(a)	50	**0.153426**	0.245387	**0.256094**	0.495342
		100	**0.089536**	0.231082	**0.242418**	0.480270
		200	**0.081775**	0.216917	**0.234569**	0.465104
	(b)	50	**0.076953**	0.238752	**0.257034**	0.487440
		100	**0.069710**	0.214250	**0.231290**	0.460268
		200	**0.086527**	0.240928	**0.263234**	0.488604
	(c)	50	**0.044846**	0.259154	**0.113061**	0.508533
		100	**0.030788**	0.236227	**0.074095**	0.483667
		200	**0.026170**	0.227478	**0.078804**	0.474521

Table 1: Average ISE and IAE for $N = 500$ trials of non-recursive estimator and the proposed estimator with the choice $(\gamma_n, \beta_n) = (n^{-0.98}, 0.31n^{-99})$. The bold values indicate the smallest values regarding the average ISE and IAE.

5 Real dataset

An interesting subject of any dataset is to estimate the unknown regression in order to predict the response variable Y when we know the explanatory variable X. In this section, we apply our proposed two-time-scale estimators r_n defined in (1.2) and the non-recursive estimator $\widehat{r}_n$ given in (1.1) on the following two datasets.

(1) Firstly, we consider the `CO2` dataset[1] which contained 60 observations on two variables: `Year` and `CO2` in August. Scientists recorded `CO2` levels, in parts per million (ppm), in the atmosphere for each `Year` from the start of 1958 through 2018. Finally, we used the Monte-Carlo method to obtain $m = 60$ for our recursive estimator r_n and $m = 50$ for the non-recursive estimator $\widehat{r}_n$. We observe that our estimator is more close to the observed data, than the non-recursive estimator, especially near the boundaries.

(2) Secondly, we consider the `wage1` dataset which appear in `R` package `np` (see [40]). Cross-section wage data consisting of a random sample taken from the U.S. Current Population Survey for the year 1976. There are 526 observations of the average hourly earnings and the years of education. We used the Monte-Carlo method to obtain $m = 68$ for our recursive estimator r_n and $m = 170$ for the non-recursive estimator $\widehat{r}_n$. We observe that our estimator is more close to the observed data, than the non-recursive estimator, especially near the boundaries.

Figures 1 and 2 show that our proposed two-time-scale recursive estimators (1.2) can give better results compared to the non-recursive estimator (1.1).

1 See https://www.co2.earth/monthly-co2.

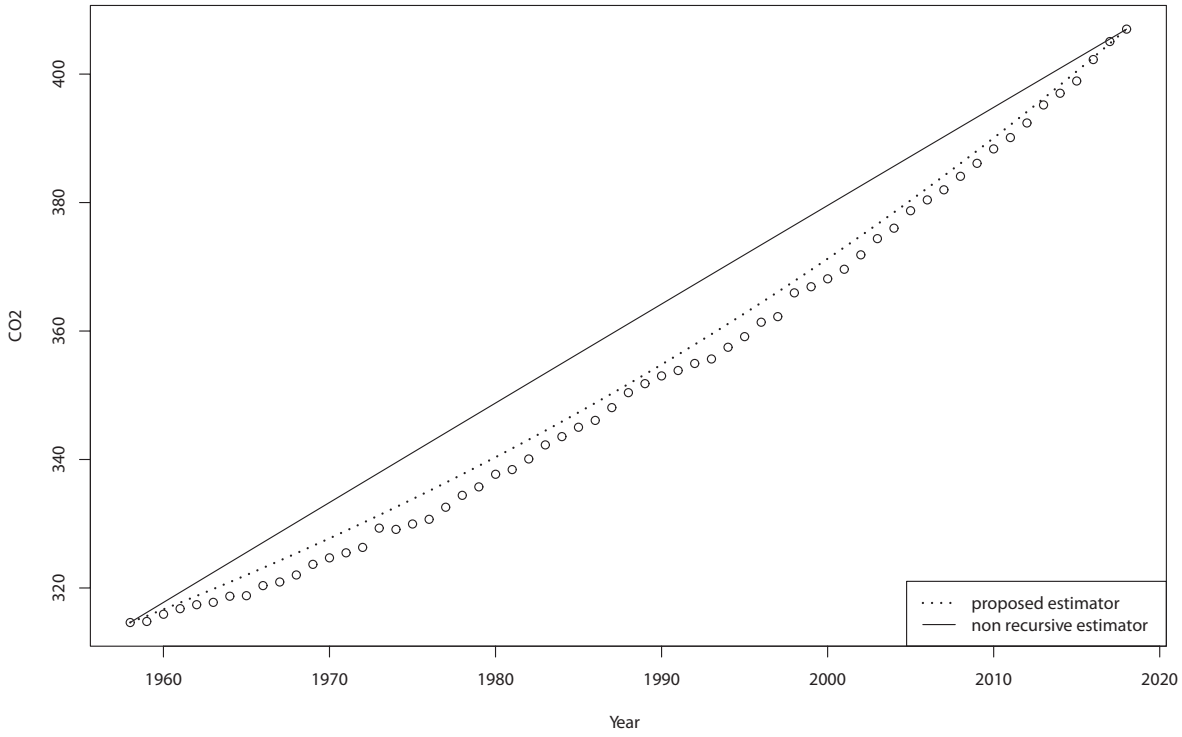

Figure 1: Qualitative comparison between the two considered estimators. Here we consider the `CO2` dataset using our proposed estimators defined in (1.2) with the stepsize $(\gamma_n, \beta_n) = (n^{-0.98}, 0.31n^{-0.99})$ and the non-recursive estimator defined in (1.1).

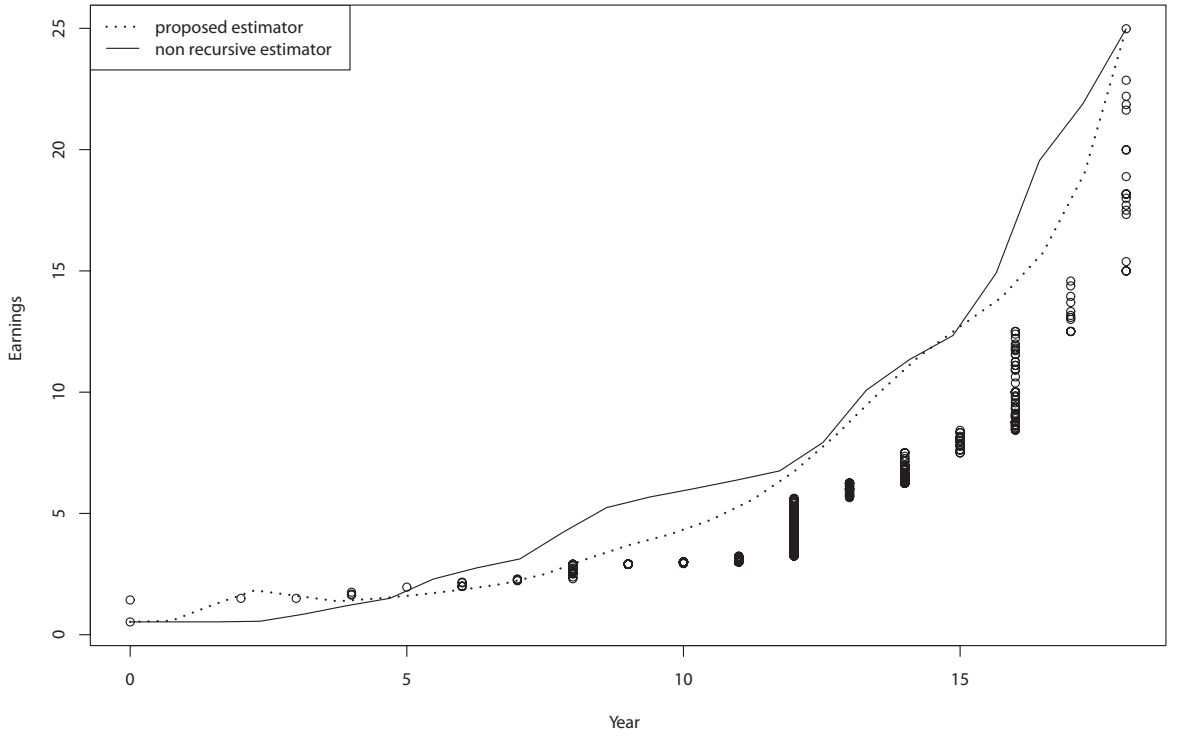

Figure 2: Qualitative comparison between the two considered estimators. Here we consider the `wage1` dataset of the package `np` using our proposed estimators defined in (1.2) with the stepsize $(\gamma_n, \beta_n) = (n^{-0.98}, 0.31n^{-0.99})$ and the non-recursive estimator defined in (1.1).

6 Concluding remarks

In the present paper we investigated a recursive nonparametric regression estimator to overcome the edge estimation problem based on Bernstein polynomials and stochastic algorithm with two-time-scale. The proposed estimator asymptotically follows normal distribution. Moreover, our proposed estimators attained the asymptotic convergence rate $O(n^{-\frac{4}{5}})$ within the interval $(0, 1)$ and $O(n^{-\frac{2}{3}})$ near the edges $\{0, 1\}$. A future research direction would be to extend our findings to the setting of serially dependent observations, α-mixing framework like in [15]. Another direction is to investigate the data-driven bandwidth selection procedures (see [32]), which requires nontrivial mathematics, this would go well beyond the scope of the present paper. We plan also to extend our approach to the case of locally linear two-time-scale recursive regression (see [9] in the case of locally linear regression).

7 Auxiliary results and mathematical developments

This section is devoted to the detailed proofs of our results. Throughout this paper we consider the following notations:

$$\xi' = \lim_{n\to\infty}(n\gamma_n)^{-1}, \quad s_n(x) = \sum_{k=1}^{n}\gamma_k,$$

$$\Pi_n = \prod_{j=1}^{n}(1-\beta_j), \qquad \mathcal{Q}_n = \prod_{j=1}^{n}(1-\gamma_j), \quad C = \frac{1}{\sqrt{2}} + 4\Big(1-\sqrt{\frac{2}{3}}\Big).$$

Let us first state the following technical lemmas.

Lemma 1 (Mokkadem, Pelletier and Slaoui [24]). *Let $(\nu_n) \in \mathcal{GS}(\nu)$, let $(\gamma_n) \in \mathcal{GS}(-\gamma)$, and let $l > 0$ such that $l - \nu\xi > 0$. We have*

$$\lim_{n\to\infty}\nu_n\Pi_n^l\sum_{k=1}^{n}\Pi_k^{-l}\frac{\gamma_k}{\nu_k} = \frac{1}{l-\nu\xi}.$$

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

$$\lim_{n\to\infty}\nu_n\Pi_n^l\Big[\sum_{k=1}^{n}\Pi_k^{-l}\frac{\gamma_k}{\nu_k}\alpha_k + \delta\Big] = 0.$$

Let us now use the following decomposition:

$$r_n(x) - r(x) = \frac{1}{f(x)}[\mathcal{T}_n(x) + \mathcal{R}_{n,1}(x) + \mathcal{R}_{n,2}(x) + \mathcal{R}_{n,3}(x) + \mathcal{R}_{n,4}(x)] + \mathcal{R}_{n,5}(x),$$

where

$$\mathcal{T}_n(x) = \Pi_n\sum_{k=1}^{n}\Pi_k^{-1}\beta_k[\mathcal{W}_k(x) - r(x)\mathcal{Z}_k(x)], \tag{7.1}$$

$$\mathcal{R}_{n,1}(x) = r(x)\Pi_n\sum_{k=1}^{n}\frac{\Pi_k^{-1}\beta_k}{\gamma_k}[f_k(x) - f_{k-1}(x)],$$

$$\mathcal{R}_{n,2}(x) = \Pi_n\sum_{k=1}^{n}\Pi_k^{-1}\beta_k[\mathcal{W}_k(x) - \mathbb{E}(\mathcal{W}_k(x))]\Big[\frac{f(x) - f_{k-1}(x)}{f_{k-1}(x)}\Big],$$

$$\mathcal{R}_{n,3}(x) = \Pi_n\sum_{k=1}^{n}\Pi_k^{-1}\beta_k[\mathbb{E}(\mathcal{W}_k(x)) - a(x)]\Big[\frac{f(x) - f_{k-1}(x)}{f_{k-1}(x)}\Big],$$

$$\mathcal{R}_{n,4}(x) = r(x)\Pi_n\sum_{k=1}^{n}\Pi_k^{-1}\beta_k\Big[\frac{(f(x) - f_{k-1}(x))^2}{f_{k-1}(x)}\Big],$$

$$\mathcal{R}_{n,5}(x) = \Pi_n(r_0(x) - r(x)).$$

Lemma 2. *Under assumptions* (A1)–(A2)*, we have*

$$f_n(x) - f(x) = O\Big(\max\{\sqrt{\gamma_n \ln(s_n)v_n}, v_n^{-2}\}\Big).$$

Lemma 3. *Under assumptions* (A1)–(A2)*, we have*

$$\mathcal{R}_{n,1}(x) = \mathcal{R}_{n,2}(x) = \mathcal{R}_{n,3}(x) = \mathcal{R}_{n,4}(x) = \mathcal{R}_{n,5}(x) = o\Big(\max\Big\{\sqrt{\beta_n v_n}, v_n^{-2}\Big\}\Big).$$

Lemma 4. *Let* $\mathcal{E}_m(x) = \sum_{k=0}^{m} b_k^2(m, x)$*. We have:*
(i) $0 \le \mathcal{E}_m(x) \le 1$*, for* $x \in [0, 1]$.
(ii) $\mathcal{E}_m(x) = m^{-\frac{1}{2}}[\psi(x) + o_x((1)]$*, for* $x \in (0, 1)$.
(iii) $\mathcal{E}_m(0) = \mathcal{E}_m(1) = 1$.
(iv) *Let* g *be any continuous function on* $[0, 1]$*. Then*

$$m^{\frac{1}{2}} \int_0^1 g(x)\mathcal{E}_m(x)\,dx = \int_0^1 g(x)\psi(x)\,dx + o(1).$$

Lemma 5. *For* $x \in [0, 1]$*, we have*

$$\mathbb{E}(\mathcal{Z}_k(x)) = f(x) + \Delta_1(x)m_k^{-1} + o(m_k^{-1}), \tag{7.2}$$

$$\mathbb{E}(\mathcal{W}_k(x)) = r(x)f(x) + \Delta_2(x)m_k^{-1} + o(m_k^{-1}), \tag{7.3}$$

$$\begin{aligned}\mathrm{Var}(\mathcal{W}_k(x)) &= \Big(m_k^{\frac{1}{2}}\mathbb{E}(Y^2 \mid X = x)f(x)\psi(x) + o_x\Big(m_k^{\frac{1}{2}}\Big)\Big)\mathbb{1}_{\{x\in(0,1)\}} \\ &\quad + \big(\mathbb{E}[Y^2 \mid X = x]f(x)m_k + o_x(m_k)\big)\mathbb{1}_{\{x\in\{0,1\}\}},\end{aligned} \tag{7.4}$$

$$\mathrm{Var}(\mathcal{Z}_k(x)) = \Big(f(x)\psi(x)m_k^{\frac{1}{2}} + o_x\Big(m_k^{\frac{1}{2}}\Big)\Big)\mathbb{1}_{\{x\in(0,1)\}} + (f(x)m_k + o(m_k))\mathbb{1}_{\{x\in\{0,1\}\}}, \tag{7.5}$$

$$\mathrm{Cov}(\mathcal{W}_k(x), \mathcal{Z}_k(x)) = \Big(r(x)f(x)\psi(x)m_k^{\frac{1}{2}} + o_x\Big(m_k^{\frac{1}{2}}\Big)\Big)\mathbb{1}_{\{x\in(0,1)\}} + (r(x)f(x)m_k + o_x(m_k))\mathbb{1}_{\{x\in\{0,1\}\}}. \tag{7.6}$$

7.1 Proof of Lemma 2

In order to prove this lemma, we referred to [36, Proposition 3.1] and we followed the same steps of proof of [29, Theorem 2, p. 44].

7.2 Proof of Lemma 3

Let us first state the following notation:

$$(\zeta_n)_n = \begin{cases} \sqrt{\beta_n v_n} & \text{if } \lim_{n\to\infty} \beta_n m_n^2 v_n = \infty, \\ v_n^{-2} & \text{otherwise.} \end{cases}$$

First, we have $\mathcal{R}_{n,5}(x) = o(\zeta_n)$.

7.2.1 Proof of Lemma 3 for $\mathcal{R}_{n,1}$

We first note that $\mathcal{R}_{n,1}$ can be written as,

$$\begin{aligned}\mathcal{R}_{n,1}(x) &= \Pi_n \sum_{k=1}^{n} \frac{\Pi_k^{-1}\beta_k}{\gamma_k}[f_k(x) - f_{k-1}(x)] \\ &= \Pi_n \sum_{k=1}^{n} \frac{\Pi_k^{-1}\beta_k}{\gamma_k}[f_k(x) - f(x)] - \Pi_n \sum_{k=1}^{n} \frac{\Pi_k^{-1}\beta_k}{\gamma_k}[f_{k-1}(x) - f(x)]\end{aligned}$$

$$
\begin{aligned}
&= \Pi_n \sum_{k=1}^{n-1} \frac{\Pi_k^{-1}\beta_k}{\gamma_k}[f_k(x) - f(x)] + \frac{\beta_n}{\gamma_n}[f_n(x) - f(x)] - \Pi_n \frac{\Pi_1^{-1}\beta_1}{\gamma_1}[f_0(x) - f(x)] \\
&\quad - \Pi_n \sum_{k=1}^{n-1} \frac{\Pi_{k+1}^{-1}\beta_{k+1}}{\gamma_{k+1}}[f_k(x) - f(x)] \\
&= \Pi_n \sum_{k=1}^{n-1} \frac{\Pi_k^{-1}\beta_k}{\gamma_k}\Big(1 - \frac{\Pi_k}{\Pi_{k+1}}\frac{\beta_{k+1}}{\beta_k}\frac{\gamma_k}{\gamma_{k+1}}\Big)(f_k(x) - f(x)) \\
&\quad + \frac{\beta_n}{\gamma_n}(f_n(x) - f(x)) - \Pi_n \frac{\beta_1}{(1-\beta_1)\gamma_1}(f_0(x) - f(x)).
\end{aligned} \tag{7.7}
$$

Moreover, in view of (A1), we infer that

$$
1 - \frac{\Pi_k}{\Pi_{k+1}}\frac{\beta_{k+1}}{\beta_k}\frac{\gamma_k}{\gamma_{k+1}} = 1 - \frac{1}{(1-\beta_{k+1})}\frac{\beta_k^{-1}}{\beta_{k+1}^{-1}}\frac{\gamma_k}{\gamma_{k+1}} = O(\beta_k). \tag{7.8}
$$

Then the application of Lemma 2 together with (7.7) and (7.8) ensures that

$$
\begin{aligned}
\mathcal{R}_{n,1}(x) &= O\Big(\Pi_n \sum_{k=1}^{n} \Pi_k^{-1}\frac{\beta_k^2}{\gamma_k}(f_k(x) - f(x)) + \frac{\beta_n}{\gamma_n}(f_n(x) - f(x)) + \Pi_n\Big) \\
&= O\Big(\Pi_n \sum_{k=1}^{n} \Pi_k^{-1}\frac{\beta_k^2}{\gamma_k}(\gamma_k v_k \ln s_k)^{\frac{1}{2}} + \Pi_n \sum_{k=1}^{n} \Pi_k^{-1}\frac{\beta_k^2}{\gamma_k}v_k^{-2} + \frac{\beta_n}{\gamma_n}(\gamma_n v_n \ln s_n)^{\frac{1}{2}} + \frac{\beta_n}{\gamma_n}v_n^{-2}\Big) + O(\Pi_n) \\
&= O\Big(\Pi_n \sum_{k=1}^{n} \Pi_k^{-1}\beta_k o(\sqrt{\beta_k v_k}) + \Pi_n \sum_{k=1}^{n} \Pi_k^{-1}\beta_k o(v_k^{-2}) o(\sqrt{\beta_n v_n}) + o(v_n^{-2}) + \Pi_n\Big) \\
&= O\Big(\Pi_n \sum_{k=1}^{n} \Pi_k^{-1}\beta_k o(\zeta_n)\Big) + o(\zeta_n) + O(\Pi_n) \\
&= o(\zeta_n).
\end{aligned}
$$

7.2.2 Proof of Lemma 3 for $\mathcal{R}_{n,2}$

Throughout this proof we use the following notation:

$$
\mathcal{L}_k(x) = \mathcal{W}_k(x) - \mathbb{E}(\mathcal{W}_k(x)), \quad G_k(x) = \frac{f(x) - f_k(x)}{f_k(x)}, \quad M_n(x) = \sum_{k=1}^{n} \Pi_k^{-1}\beta_k \mathcal{L}_k(x) G_{k-1}(x).
$$

First, we set $\mathcal{F}_k = \sigma((X_1, Y_1), \ldots, (X_k, Y_k))$, and we note that the increasing process of the martingale $(M_n(x)$ satisfies

$$
\begin{aligned}
\langle M\rangle_n(x) &= \sum_{k=1}^{n} \mathbb{E}[\Pi_k^{-2}\beta_k^2 \mathcal{L}_k^2(x) G_{k-1}^2(x) \mid \mathcal{F}_{k-1}] \\
&= \sum_{k=1}^{n} \Pi_k^{-2}\beta_k^2 G_{k-1}^2(x)\mathbb{E}[\mathcal{L}_k^2(x) \mid \mathcal{F}_{k-1}] \\
&= \sum_{k=1}^{n} \Pi_k^{-2}\beta_k^2 G_{k-1}^2(x)\mathbb{E}(\mathcal{L}_k^2(x)) \\
&= \sum_{k=1}^{n} \Pi_k^{-2}\beta_k^2 G_{k-1}^2(x) Var(\mathcal{W}_k(x)).
\end{aligned}
$$

In view of (7.4), the application of Lemma 2 ensures that

$$
\begin{aligned}
\langle M\rangle_n(x) &= O\Big(\sum_{k=1}^{n} \Pi_k^{-2}\beta_k^2\Big(\frac{f(x) - f_{k-1}(x)}{f_{k-1}(x)}\Big)^2 v_k\Big) = O\Big(\sum_{k=1}^{n} \Pi_k^{-2}\beta_k^2(\gamma_k \ln(s_k)v_k + v_k^{-4})v_k\Big) \\
&= O\Big(\sum_{k=1}^{n} \Pi_k^{-2}\beta_k^2 v_k^2 \gamma_k \ln(s_k) + \sum_{k=1}^{n} \Pi_k^{-2}\beta_k^2 v_k^{-3}\Big).
\end{aligned}
$$

Let us first consider the case when $\lim_{n\to\infty} n\beta_n = \infty$. The application of Lemma 1 gives

$$\langle M\rangle_n(x) = O(\Pi_n^{-2}\beta_n\gamma_n \ln(s_n)v_n^2 + \Pi_n^{-2}\beta_n v_n^{-3}).$$

Moreover, we note that for all $\varepsilon > 0$ we have

$$\ln(\Pi_n^{-2}) = \sum_{k=1}^{n} \ln(1-\beta_k)^{-2} = \sum_{k=1}^{n}(2\beta_k + o(\beta_k)) = O\Big(\sum_{k=1}^{n}\beta_k k^{\varepsilon}\Big).$$

Since we have $\beta_n n^{\varepsilon} \in \mathcal{GS}(-(\beta-\varepsilon))$ with $(\beta-\varepsilon) < 1$, the application of Lemma 1 ensures that

$$\lim_{n\to\infty} \frac{n\beta_n n^{\varepsilon}}{\sum_{k=1}^{n}\beta_k k^{\varepsilon}} = 1-(\beta-\varepsilon).$$

It comes that

$$\ln(\Pi_n^{-2}) = O(n^{1+\varepsilon}\beta_n).$$

Moreover, the sequences $(\beta_n\gamma_n\ln(s_n)v_n^2)$ and $(\beta_n v_n^{-3})$ tend to zero. It follows that

$$\ln\langle M\rangle_n(x) = O(\ln(\Pi_n^{-2})) = O(n^{1+\varepsilon}\beta_n).$$

Then the application of [4, Theorem 1.3.15], ensures that for any $\delta > 0$,

$$\begin{aligned}|M_n(x)| &= o\Big(\langle M\rangle_n^{\frac{1}{2}}(x)(\ln\langle M\rangle_n(x))^{\frac{1+\delta}{2}}\Big) + O(1)\\ &= o\Big(\Pi_n^{-1}\Big(\beta_n^{\frac{1}{2}}\gamma_n^{\frac{1}{2}}\ln(s_n)^{\frac{1}{2}}v_n + \beta_n^{\frac{1}{2}}v_n^{-\frac{3}{2}}\Big)(n^{1+\varepsilon}\beta_n)^{\frac{1+\delta}{2}}\Big) + O(1).\end{aligned}$$

Now, we set $\varepsilon > 0$ and $\delta > 0$ such that $((\gamma_n v_n^2)^{\frac{1}{2}}(n^{1+\varepsilon}\beta_n)^{\frac{1+\delta}{2}}) \in \mathcal{GS}(\mu)$ with $\mu < 0$. For more precision, the existence of ε and δ should be ensured by the condition $\alpha > (a\mathbb{1}_{\{x\in\{0,1\}\}} + \frac{a}{2}\mathbb{1}_{\{x\in(0,1)\}}) + 1 - \beta$. Moreover, we obtain

$$\begin{aligned}\Pi_n|M_n(x)| &= o\Big((\beta_n v_n)^{\frac{1}{2}}(\gamma_n v_n)^{\frac{1}{2}}\ln(s_n)^{\frac{1}{2}}(n^{1+\varepsilon}\beta_n)^{\frac{1+\delta}{2}} + (\beta_n v_n)^{\frac{1}{2}}v_n^{-2}(n^{1+\varepsilon}\beta_n)^{\frac{1+\delta}{2}}\Big) + O(\Pi_n)\\ &= o\Big((\beta_n v_n)^{\frac{1}{2}} + v_n^{-2}\Big) + o(\zeta_n)\\ &= o(\zeta_n).\end{aligned}$$

Now, in the case when the sequence $(n\beta_n)$ is bounded when n goes to infinity, the application of Lemma 1 for all $(\mathcal{C}_n) \in \mathcal{GS}(0)$ together with (7.9) ensures that

$$\begin{aligned}\langle M\rangle_n(x) &= O\Big(\sum_{k=1}^{n}\Pi_k^{-2}\beta_k[(\beta_k v_k)(\gamma_k v_k\ln(s_k)) + (\beta_k v_k)v_k^{-4}]\Big)\\ &= O\Big(\sum_{k=1}^{n}\Pi_k^{-2}\beta_k[o(\beta_k v_k\mathcal{C}_k) + o(v_k^{-4}\mathcal{C}_k)]\Big)\\ &= o\Big(\sum_{k=1}^{n}\Pi_k^{-2}\beta_k o(\zeta_k^2\mathcal{C}_k)\Big)\\ &= o(\Pi_n^{-2}\zeta_n^2\mathcal{C}_n).\end{aligned}$$

Moreover, since the sequence $(n\beta_n)$ is bounded when n goes to infinity, in this case, we have $\Pi_n^{-1} \in \mathcal{GS}(\xi^{-1})$ and $\ln(\Pi_n^{-2}\zeta_n^2\mathcal{L}_n) = O(\ln n)$. Then, the application of [4, Theorem 1.3.15], ensures that for all $\delta > 0$,

$$\begin{aligned}|M_n(x)| &= o\Big(\langle M\rangle_n^{\frac{1}{2}}(x)(\ln\langle M\rangle_n(x))^{\frac{1+\delta}{2}}\Big) + O(1)\\ &= o\Big(\Pi_n^{-1}\zeta_n\mathcal{C}_n^{\frac{1}{2}}\ln(n)^{\frac{1+\delta}{2}}\Big) + O(1).\end{aligned}$$

Thus, for $\delta = 1$ and $\mathcal{L}_n = (\ln n)^{-2}$, we get $\Pi_n|M_n(x)| = o(\zeta_n)$, which ensures that

$$\mathcal{R}_{n,2}(x) = o(\zeta_n).$$

7.2.3 Proof of Lemma 3 for $\mathcal{R}_{n,3}$

The application of Lemma 1 together with Lemma 2, ensures that

$$\begin{aligned}\mathcal{R}_{n,3}(x) &= \Pi_n \sum_{k=0}^{n-1} \Pi_{k+1}^{-1} \beta_{k+1} [\mathbb{E}(\mathcal{W}_{k+1}(x)) - a(x)] \Big[\frac{f(x) - f_k(x)}{f_k(x)} \Big] \\ &= \Pi_n \sum_{k=0}^{n-1} \Pi_{k+1}^{-1} \beta_{k+1} [\Delta_2(x) m_{k+1}^{-1} + o(m_{k+1}^{-1})] \Big[\frac{f(x) - f_k(x)}{f_k(x)} \Big] \\ &= O\Big(\Pi_n \sum_{k=0}^{n-1} \Pi_{k+1}^{-1} \beta_{k+1} m_{k+1}^{-1} |f(x) - f_k(x)| \Big) \\ &= o(\zeta_n).\end{aligned}$$

7.2.4 Proof of Lemma 3 for $\mathcal{R}_{n,4}$

The application of Lemma 2 ensures that

$$\begin{aligned}\mathcal{R}_{n,4}(x) &= \Pi_n \sum_{k=1}^{n} \Pi_k^{-1} \beta_k O((f(x) - f_{k-1}(x))^2) \\ &= O\Big(\Pi_n \sum_{k=1}^{n} \Pi_k^{-1} \beta_k ((\gamma_k v_k \ln s_k) + v_k^{-4}) \Big) \\ &= o(\zeta_n).\end{aligned}$$

7.3 Proof of Proposition 1

The application of Lemma 1 together with (7.1) and (7.2) ensures that in the case when $a, x \in \mathcal{A}_{a,x}^1 \cup \mathcal{A}_{a,x}^2$,

$$\mathbb{E}(\mathcal{T}_n(x)) = \frac{1}{1 - a\xi} (\Delta_2(x) - r(x)\Delta_1(x) + o(1)) m_n^{-1},$$

in the case when $a, x \in \mathcal{A}_{a,x}^3$, we have

$$\mathbb{E}(\mathcal{T}_n(x)) = o\Big(\sqrt{\beta_n m_n} \Big),$$

and in the case when $a, x \in \mathcal{A}_{a,x}^4$, we have

$$\mathbb{E}(\mathcal{T}_n(x)) = o\Big(\sqrt{\beta_n m_n^{\frac{1}{2}}} \Big).$$

Let assumptions (A1)–(A2) hold such as $f(x) > 0$ for $x \in \{0, 1\}$. The asymptotic behavior of $r_n - r$ is given by the one of $\mathcal{T}_n$. More precisely, we establish the following lemma.

Lemma 6. *The following statements hold:*

(i) *In the case when $x \in (0, 1)$, if there exists $c \geq 0$ such that $\beta_n^{-1} m_n^{-\frac{5}{2}} \to c$ (resp. the case when $x \in \{0, 1\}$, if $\beta_n^{-1} m_n^{-3} \to c$), then*

$$\mathcal{V}_n \mathcal{T}_n(x) \xrightarrow{\mathcal{D}} \mathcal{N}\Big(\frac{\sqrt{c}}{(1 - a\xi)} \operatorname{Bias}(x) f(x), \mathcal{V}(x) f^2(x) \Big).$$

(ii) *In the case when $x \in (0, 1)$, if $\beta_n^{-1} m_n^{-\frac{5}{2}} \to \infty$ (resp. the case when $x \in \{0, 1\}$, if $\beta_n^{-1} m_n^{-3} \to \infty$), then*

$$m_n \, \mathcal{T}_n(x) \xrightarrow{\mathbb{P}} \frac{1}{(1 - a\xi)} \operatorname{Bias}(x) f(x).$$

Proof. Let us first note that

$$\mathcal{T}_n(x) - \mathbb{E}(\mathcal{T}_n(x)) = \Pi_n \sum_{k=1}^{n} \{v_k(x) - \mathbb{E}(v_k(x))\},$$

where

$$v_k(x) = \Pi_k^{-1}\beta_k(\mathcal{W}_k(x) - r(x)\mathcal{Z}_k(x)).$$

Let us now assume that, when $a, x \in \mathcal{A}_{a,x}^3 \cup \mathcal{A}_{a,x}^4$, we have

$$\mathcal{V}_n(\mathcal{T}_n(x) - \mathbb{E}[\mathcal{T}_n(x)]) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \mathcal{V}(x)f^2(x)). \tag{7.9}$$

In the case when $a, x \in \mathcal{A}_{a,x}^1 \cup \mathcal{A}_{a,x}^2$, equation (3.1) gives $m_n(\mathcal{T}_n(x) - \mathbb{E}(\mathcal{T}_n(x))) \xrightarrow{\mathbb{P}} 0$.

Let us now prove (7.9). For this purpose, we set

$$\mathcal{B}_k(x) = \Pi_k^{-1}\beta_k\{v_k(x) - \mathbb{E}(v_k(x))\}.$$

Moreover, since

$$\mathrm{Var}(\mathcal{T}_k(x)) = O\Big(m_k^{\frac{1}{2}}[1 + o_x(1)]\mathbb{1}_{\mathcal{A}_{a,x}^4} + m_k[1 + o(1)]\mathbb{1}_{\mathcal{A}_{a,x}^3}\Big),$$

the application of Lemma 1 ensures that

$$\begin{aligned}\mathcal{Y}_n^2 &= \sum_{k=1}^n \mathrm{Var}(\mathcal{B}_k(x)) = \sum_{k=1}^n \Pi_k^{-2}\beta_k^2\mathrm{Var}(v_k(x)),\\ &= \frac{\beta_n m_n^{\frac{1}{2}}}{\Pi_n^2}\mathcal{V}(x) + o_x\Big(\frac{\beta_n m_n^{\frac{1}{2}}}{\Pi_n^2}\Big)\mathbb{1}_{\mathcal{A}_{a,x}^4} + o\Big(\frac{\gamma_n m_n}{\Pi_n^2}\Big)\mathbb{1}_{\mathcal{A}_{a,x}^3}.\end{aligned}$$

Further, the application of Lemma 1 ensure that for all $p > 0$,

$$\begin{aligned}\mathbb{E}[|\mathcal{B}_k(x)|^{2+p}] &= m_k^{2+p}\mathbb{E}\Bigg(|Y_k - r(x)|^{2+p}\Big\{\sum_{j=0}^{m_k-1}\mathbb{1}_{\{\frac{j}{m_k} < T_k \le \frac{j+1}{m_k}\}}b_j(m_k - 1, x)\Big\}^{2+p}\Bigg)\\ &= O\Big(m_k^{\frac{3}{4}(2+p)}\Big)\mathbb{1}_{\mathcal{A}_{a,x}^4} + O\Big(m_k^{2+p}\Big)\mathbb{1}_{\mathcal{A}_{a,x}^3}.\end{aligned}$$

Using the fact that

$$\lim_{n\to\infty}(n\gamma_n) > \frac{2\alpha - a}{4}\mathbb{1}_{\mathcal{A}_{a,x}^4} + \frac{\alpha - a}{2}\mathbb{1}_{\mathcal{A}_{a,x}^3},$$

there exists $p > 0$ such that

$$\lim_{n\to\infty}(n\gamma_n) > \Big(\frac{1+p}{2+p}\alpha - \frac{3}{4}\Big)\mathbb{1}_{\mathcal{A}_{a,x}^4} + \Big(\frac{1+p}{2+p}\alpha - \frac{3}{2}\Big)\mathbb{1}_{\mathcal{A}_{a,x}^3}.$$

Lemma 1 gives

$$\begin{aligned}\sum_{k=1}^n \mathbb{E}[|\mathcal{B}_k((x)|^{2+p}] &= O\Big(\sum_{k=1}^n \Pi_k^{-2-p}\beta_k^{2+p}\mathbb{E}[|\mathcal{T}_k(x)|^{2+p}]\Big)\\ &= O\Big(\gamma_n^{1+p}\Pi_n^{-2-p}m_n^{\frac{3}{4}(2+p)}\Big)\mathbb{1}_{\mathcal{A}_{a,x}^4} + O\Big(\gamma_n^{1+p}\Pi_n^{-2-p}m_n^{2+p}\Big)\mathbb{1}_{\mathcal{A}_{a,x}^3},\end{aligned}$$

from which we deduce that

$$\frac{1}{\mathcal{Y}_n^{2+p}}\sum_{k=1}^n \mathbb{E}[|\mathcal{B}_k((x)|^{2+p}] = O(m_n[\gamma_n m_n]^{\frac{p}{2}}) = o(1).$$

The convergence in (7.9) then follows from the application of Lyapunov's Theorem. □

References

[1] G. J. Babu, A. J. Canty and Y. P. Chaubey, Application of Bernstein polynomials for smooth estimation of a distribution and density function, *J. Statist. Plann. Inference* **105** (2002), no. 2, 377–392.

[2] S. Bhatnagar, M. C. Fu, S. I. Marcus and S. Bathnagar, Two timescale algorithms for simulation optimization of hidden Markov models, *IEEE Trans.* **3** (2001), 245–258.

[3] V. S. Borkar, Stochastic approximation with two time scales, *Systems Control Lett.* **29** (1997), no. 5, 291–294.
[4] M. Duflo, *Random Iterative Models*, Appl. Math. (New York) 34, Springer, Berlin, 1997.
[5] R. L. Eubank, *Spline Smoothing and Nonparametric Regression*, Stat. Textb. Monogr. 90, Marcel Dekker, New York, 1988.
[6] J. Galambos and E. Seneta, Regularly varying sequences, *Proc. Amer. Math. Soc.* **41** (1973), 110–116.
[7] S. Ghosal, Convergence rates for density estimation with Bernstein polynomials, *Ann. Statist.* **29** (2001), no. 5, 1264–1280.
[8] W. Härdle, *Applied Nonparametric Regression*, Econom. Soc. Monogr. 19, Cambridge University, Cambridge, 1990.
[9] W. Härdle, M. Müller, S. Sperlich and A. Werwatz, *Nonparametric and Semiparametric Models*, Springer Ser. Statist., Springer, New York, 2004.
[10] T. J. Hastie and R. J. Tibshirani, *Generalized Additive Models*, Monogr. Statist. Appl. Probab. 43, Chapman and Hall, London, 1990.
[11] G. Igarashi and Y. Kakizawa, On improving convergence rate of Bernstein polynomial density estimator, *J. Nonparametr. Stat.* **26** (2014), no. 1, 61–84.
[12] A. Jmaei, Y. Slaoui and W. Dellagi, Recursive distribution estimator defined by stochastic approximation method using Bernstein polynomials, *J. Nonparametr. Stat.* **29** (2017), no. 4, 792–805.
[13] Y. Kakizawa, Bernstein polynomial probability density estimation, *J. Nonparametr. Stat.* **16** (2004), no. 5, 709–729.
[14] Y. Kakizawa, A note on generalized Bernstein polynomial density estimators, *Stat. Methodol.* **8** (2011), no. 2, 136–153.
[15] S. Khardani and Y. Slaoui, Recursive kernel density estimation and optimal bandwidth selection under α: Mixing data, *J. Stat. Theory Pract.* **13** (2019), no. 2, Paper No. 36.
[16] J. Kiefer and J. Wolfowitz, Stochastic estimation of the maximum of a regression function, *Ann. Math. Statist.* **23** (1952), 462–466.
[17] V. R. Konda and V. S. Borkar, Actor-critic-type learning algorithms for Markov decision processes, *SIAM J. Control Optim.* **38** (1999), no. 1, 94–123.
[18] V. R. Konda and J. N. Tsitsiklis, Convergence rate of linear two-time-scale stochastic approximation, *Ann. Appl. Probab.* **14** (2004), no. 2, 796–819.
[19] A. Leblanc, A bias-reduced approach to density estimation using Bernstein polynomials, *J. Nonparametr. Stat.* **22** (2010), no. 3–4, 459–475.
[20] A. Leblanc, On estimating distribution functions using Bernstein polynomials, *Ann. Inst. Statist. Math.* **64** (2012), no. 5, 919–943.
[21] A. Leblanc, On the boundary properties of Bernstein polynomial estimators of density and distribution functions, *J. Statist. Plann. Inference* **142** (2012), no. 10, 2762–2778.
[22] A. Mokkadem and M. Pelletier, Convergence rate and averaging of nonlinear two-time-scale stochastic approximation algorithms, *Ann. Appl. Probab.* **16** (2006), no. 3, 1671–1702.
[23] A. Mokkadem and M. Pelletier, A generalization of the averaging procedure: The use of two-time-scale algorithms, *SIAM J. Control Optim.* **49** (2011), no. 4, 1523–1543.
[24] A. Mokkadem, M. Pelletier and Y. Slaoui, Revisiting Révész's stochastic approximation method for the estimation of a regression function, *ALEA Lat. Am. J. Probab. Math. Stat.* **6** (2009), 63–114.
[25] H.-G. Müller, *Nonparametric Regression Analysis of Longitudinal Data*, Lect. Notes Stat. 46, Springer, Berlin, 1988.
[26] E. A. Nadaraya, On estimating regression, *Theory Probab. Appl.* **10** (1964), 186–190.
[27] B. L. S. Prakasa Rao, Estimation of distribution and density functions by generalized Bernstein polynomials, *Indian J. Pure Appl. Math.* **36** (2005), no. 2, 63–88.
[28] P. Révész, Robbins–Monro procedure in a Hilbert space and its application in the theory of learning processes. I, *Studia Sci. Math. Hungar.* **8** (1973), 391–398.
[29] Y. Slaoui, *Application des méhodes d'approximations stochastiques à l'estimation de la densité et de la régression*, Doctoral dissertation, Université de Versailles-Saint Quentin en Yvelines, 2006.
[30] Y. Slaoui, Large and moderate deviation principles for recursive kernel density estimators defined by stochastic approximation method, *Serdica Math. J.* **39** (2013), no. 1, 53–82.
[31] Y. Slaoui, Moderate deviation principles for recursive regression estimators defined by stochastic approximation method, *Int. J. Math. Stat.* **16** (2015), no. 2, 51–60.
[32] Y. Slaoui, Plug-in bandwidth selector for recursive kernel regression estimators defined by stochastic approximation method, *Stat. Neerl.* **69** (2015), no. 4, 483–509.
[33] Y. Slaoui, Optimal bandwidth selection for semi-recursive kernel regression estimators, *Stat. Interface* **9** (2016), no. 3, 375–388.
[34] Y. Slaoui, Wild bootstrap bandwidth selection of recursive nonparametric relative regression for independent functional data, *J. Multivariate Anal.* **173** (2019), 494–511.
[35] Y. Slaoui, Recursive nonparametric regression estimation for independent functional data, *Statist. Sinica* **30** (2020), no. 1, 417–437.
[36] Y. Slaoui and A. Jmaei, Recursive density estimators based on Robbins–Monro's scheme and using Bernstein polynomials, *Stat. Interface* **12** (2019), no. 3, 439–455.
[37] A. Tenbusch, Two-dimensional Bernstein polynomial density estimators, *Metrika* **41** (1994), no. 3–4, 233–253.

[38] R. A. Vitale, Bernstein polynomial approach to density function estimation, in: *Statistical Inference and Related Topics. Vol. 2*, Indiana University, Bloomington (1975), 87–99.
[39] G. S. Watson, Smooth regression analysis, *Sankhyā Ser. A* **26** (1964), 359–372.
[40] J. M. Wooldridge, *Basic Regression Analysis with Time Series Data. Introductory Econometrics: A Modern Approach*, South-Western College, Cincinnati, 2000.