Statistics

A Journal of Theoretical and Applied Statistics

ISSN: 0233-1888 (Print) 1029-4910 (Online) Journal homepage: http://www.tandfonline.com/loi/gsta20

On the asymptotic behaviour of the recursive Nadaraya–Watson estimator associated with the recursive sliced inverse regression method

Bernard Bercu, Thi Mong Ngoc Nguyen & Jerome Saracco

Statistics, 2015
Vol. 49, No. 3, 660–679, http://dx.doi.org/10.1080/02331888.2014.884097

Taylor & Francis
Taylor & Francis Group

On the asymptotic behaviour of the recursive Nadaraya–Watson estimator associated with the recursive sliced inverse regression method

Bernard Bercu[a], Thi Mong Ngoc Nguyen[b] and Jerome Saracco[c*]

[a]*Université de Bordeaux, Institut de Mathématiques de Bordeaux, UMR CNRS 5251, 351 cours de la libération, 33405 Talence cedex, France;* [b]*Université de Strasbourg, Institut de Recherche Mathématique Avancée, UMR CNRS 7501, 7 rue René Descartes 67084 Strasbourg cedex, France;* [c]*Institut Polytechnique de Bordeaux, Institut de Mathématiques de Bordeaux, UMR CNRS 5251, 351 cours de la libération, 33405 Talence cedex, France*

(*Received 25 September 2012; accepted 10 January 2014*)

We investigate the asymptotic behaviour of the recursive Nadaraya–Watson estimator for the estimation of the regression function in a semiparametric regression model. On the one hand, we make use of the recursive version of the sliced inverse regression method for the estimation of the unknown parameter of the model. On the other hand, we implement a recursive Nadaraya–Watson procedure for the estimation of the regression function which takes into account the previous estimation of the parameter of the semiparametric regression model. We establish the almost sure convergence as well as the asymptotic normality for our Nadaraya–Watson estimate. We also illustrate our semiparametric estimation procedure on simulated data.

Keywords: semi-parametric regression; recursive estimation; Nadaraya–Watson estimator; sliced inversion regression

2000 Mathematics Subject Classification: Primary: 62H12; Secondary: 62G05, 60F05, 62L12

1. Introduction

The goal of this paper is to investigate the asymptotic behaviour of the recursive Nadaraya–Watson estimator of the regression function f in the semiparametric regression model given, for all $k \geq 1$, by

$$Y_k = f(\theta' X_k) + \varepsilon_k, \tag{1}$$

where (X_k) is a sequence of independent and identically distributed random vectors of $\mathbb{R}^p$ and the driven noise (ε_k) is a real martingale difference sequence independent of (X_k). We assume in all the sequel that the unknown p-dimensional parameter $\theta \neq 0$. On the one hand, we make use of the recursive version of the sliced inverse regression (SIR) method, originally proposed by Li [1] and Duan and Li,[2] in order to estimate θ. On the other hand, we estimate the unknown regression function f via a recursive Nadaraya–Watson estimator which takes into account the previous estimation of the parameter θ. Our purpose is precisely to investigate the asymptotic

*Corresponding author. Email: jerome.saracco@math.u-bordeaux1.fr

behaviour of the recursive Nadaraya–Watson estimator of f. One can observe that single index models have been extensively studied in the literature, see for instance Horowitz.[3]

We wish to point out that we propose two recursive estimation procedures for the estimation of θ and f. As matter of fact, in many practical situations, where the data come online with relatively high speed, it is more convenient to implement recursive estimation procedure. It is the case for high-frequency financial data where one should wait the best time to trade in the stock market. It is also the case for real-time electricity data where it is important to forecast electricity peak consumption times.

One can find a wide range of literature on nonparametric estimation of a regression function. We refer the reader to [4–7] for some excellent books on density and regression function estimation. In the classical situation without any parameter θ, the almost sure convergence of the Nadaraya–Watson estimator [8,9] was proved by Noda [10] and its asymptotic normality was established by Schuster.[11] Moreover, Choi et al. [12] proposed three data-sharpening versions of the Nadaraya–Watson estimator in order to reduce the asymptotic bias in the central limit theorem as well as the asymptotic mean-squared error. In our situation, we propose to make use of a recursive Nadaraya–Watson estimator [13] of f which takes into account the previous estimation of the parameter θ. It is given, for all $x \in \mathbb{R}^p$ and $n > p$, by

$$\hat{f}_n(x) = \frac{\sum_{k=p+1}^{n} W_k(x) Y_k}{\sum_{k=p+1}^{n} W_k(x)} \tag{2}$$

with

$$W_k(x) = \frac{1}{h_k} K\left(\frac{x - \hat{\theta}_{k-1}' X_k}{h_k}\right),$$

where the kernel K is a chosen probability density function and the bandwidth (h_n) is a sequence of positive real numbers decreasing to zero, such that nh_n tends to infinity. For the sake of simplicity, we propose to make use of $h_n = 1/n^{\alpha}$ with $\alpha \in]0, 1[$. The main difficulty arising here is that we have to deal with the recursive SIR estimator $\hat{\theta}_n$ of θ inside the kernel K. Note that the SIR estimator $\hat{\theta}_n$ in its recursive version is well defined for $n > p$ since it is necessary to invert the sample covariance matrix of the X_k's, see next section for details.

The paper is organized as follows. In Section 2, we recall some results on the recursive SIR estimator $\hat{\theta}_n$. Our main results on the asymptotic behaviour of $\hat{f}_n$ are given in Section 3. Under standard regularity assumptions on the kernel K, we establish the almost sure pointwise convergence of $\hat{f}_n$ together with its asymptotic normality. Section 4 contains some numerical experiments on simulated data, illustrating the good performances of our semiparametric estimation procedure. A conclusion is provided in Section 5. All the technical proofs are postponed to Appendices 2 and 3.

2. On the recursive SIR method

From the seminal work of Li [1] and Duan and Li [2] devoted to the SIR theory, we know that the eigenvector associated with the maximum eigenvalue of the matrix $\Sigma^{-1}\Gamma$ is collinear with θ where $\Sigma = \mathbb{V}(X_k)$ is positive definite, $\Gamma = \mathbb{V}(\mathbb{E}[X_k \mid T(Y_k)])$ and T is a slicing of the range of Y_k into H non overlapping slices $s_1, \ldots, s_H$. One can observe that since the link function f is unknown in the semiparametric regression model (1), the parameter θ is not entirely identifiable. Only its direction can be identified without assuming additional constraints. Li [1] called effective dimension reduction (EDR), any direction collinear with θ. One can note that the notion of EDR space was also clarified by Cook and his collaborators in their numerous papers introducing the notions of central subspace and central mean subspace, see for details Cook [14] or Cook and Li.[15] Moreover, the SIR theory mainly relies on the so-called linearity condition (LC) which imposes that for all $b \in \mathbb{R}^p$, $\mathbb{E}[b'X_k \mid \theta' X_k]$ is linear in $\theta' X_k$. It means that one can find $\alpha, \beta \in \mathbb{R}$ such that

$$\mathbb{E}[b'X_k \mid \theta' X_k] = \alpha + \beta\theta' X_k. \tag{LC}$$

This condition is required to hold only for the true parameter θ. Since θ is unknown, it is not possible in practice to verify it a priori. Hence, we can assume that (LC) holds for all possible values of θ, which is equivalent to elliptical symmetry of the distribution of the identically distributed sequence (X_k). Finally, Hall and Li [16] mentioned that (LC) is not a severe restriction because (LC) holds to a good approximation in many problems as the dimension p of the regression vector X_k increases. Chen and Li [17] or Cook and Ni [18] also provide interesting discussions on the linearity condition.

The SIR estimates based on the first inverse moment have been studied extensively in the literature, see for instance Prendergast [19] or Szretter and Yohai [20] among others. In order to avoid the choice of a slicing in SIR, pooled slicing, kernel or spline versions of SIR have been investigated, see for example Zhu and Yu,[21] Wu [22] or Azais et al. [23] Note that these methods are hard to implement compared to the basic SIR approach and are often computationally slow. Sparse SIR has been proposed, see for example Li and Nachtsheim.[24] Regularized versions for SIR have also been proposed for high-dimensional covariates, see for instance Scrucca [25] or Li and Yin.[26] Hybrid methods of inverse regression-based algorithms have been studied, see for instance Zhu et al. [27]

However all these methods are not recursive. In the following, we will describe a recursive way to obtain an EDR direction estimated with SIR approach.

In order to obtain a recursive version of an EDR direction estimated with SIR approach, we need an analytic expression of the maximum eigenvector of $\Sigma^{-1}\Gamma$. It is easily tractable when the range of Y_k is divided into two non overlapping slices s_1 and s_2. Hereafter we shall assume that $H = 2$. In this special case, it is not hard to see that $\Gamma = p_1 z_1 z_1' + p_2 z_2 z_2'$ where $p_h = P(Y_k \in s_h)$ and $z_h = \mathbb{E}[X_k \mid Y_k \in s_h] - \mathbb{E}[X_k]$ with $p_h \neq 0$ for $h = 1, 2$. Moreover, it is straightforward to show that the eigenvector associated with the maximum eigenvalue of $\Sigma^{-1}\Gamma$ can be written as

$$\tilde{\theta} = \Sigma^{-1}(z_1 - z_2),$$

see Appendix 1 for details. This vector $\tilde{\theta}$ is therefore an EDR direction. For the sake of simplicity, we identify in all the sequel the EDR direction $\tilde{\theta}$ with θ. Our purpose is now to propose an estimator of the EDR direction θ. First of all, let us recall the non recursive SIR estimator $\tilde{\theta}_n$ of θ given by Nguyen and Saracco.[28] The estimator $\tilde{\theta}_n$ can be easily obtained from the sample $(X_1, Y_1), \ldots, (X_n, Y_n)$ by substituting the theoretical moments by their sample counterparts. More precisely, $\tilde{\theta}_n$ is given by

$$\tilde{\theta}_n = \Sigma_n^{-1}(z_{1,n} - z_{2,n}), \tag{3}$$

where

$$\Sigma_n = \frac{1}{n}\sum_{k=1}^{n}(X_k - \bar{X}_n)(X_k - \bar{X}_n)', \quad \bar{X}_n = \frac{1}{n}\sum_{k=1}^{n} X_k \tag{4}$$

and, for $h = 1, 2$, $z_{h,n} = m_{h,n} - \bar{X}_n$ where

$$m_{h,n} = \frac{1}{n_{h,n}}\sum_{k=1}^{n} X_k \mathrm{I}_{\{Y_k \in s_h\}}, \quad n_{h,n} = \sum_{k=1}^{n} \mathrm{I}_{\{Y_k \in s_h\}}. \tag{5}$$

Next, we focus our attention on the recursive SIR estimator $\hat{\theta}_n$ of θ proposed by Nguyen and Saracco.[28] We split the sample into two parts: the subsample of the first $(n-1)$ observations

$(X_1, Y_1), \ldots, (X_{n-1}, Y_{n-1})$, and the new observation (X_n, Y_n). On the one hand, the inverse of the matrix Σ_n given by Equation (6) may be recursively calculated via the Riccati equation,[13]

$$\Sigma_n^{-1} = \frac{n}{n-1}\Sigma_{n-1}^{-1} - \frac{n}{(n-1)(n+\rho_n)}\Sigma_{n-1}^{-1}\Phi_n\Phi_n'\Sigma_{n-1}^{-1}, \tag{6}$$

where $\rho_n = \Phi_n'\Sigma_{n-1}^{-1}\Phi_n$ and $\Phi_n = X_n - \bar{X}_{n-1}$. On the other hand, we can also obtain the recursive form of $z_{h,n}$. As a matter of fact, we have for $h = 1, 2$,

$$z_{h,n} = \begin{cases} z_{h^*,n-1} - \dfrac{1}{n}\Phi_n + \dfrac{1}{n_{h^*,n-1}+1}\Phi_{h^*,n} & \text{if } h = h^*, \\ z_{h,n-1} - \dfrac{1}{n}\Phi_n & \text{otherwise,} \end{cases} \tag{7}$$

where h^* denotes the slice containing the observation Y_n and $\Phi_{h^*,n} = X_n - m_{h^*,n-1}$. We deduce from Equations (6) and (7) that the recursive SIR estimator $\hat{\theta}_n$ is given by

$$\begin{aligned}\hat{\theta}_n = \left(\frac{n}{n-1}\right)\hat{\theta}_{n-1} - \frac{n}{(n-1)(n+\rho_n)}\Sigma_{n-1}^{-1}\Phi_n\Phi_n'\hat{\theta}_{n-1} \\ - \frac{(-1)^{h^*}, n}{(n_{h^*,n-1}+1)(n-1)}\left(\Sigma_{n-1}^{-1} - \frac{1}{n+\rho_n}\Sigma_{n-1}^{-1}\Phi_n\Phi_n'\Sigma_{n-1}^{-1}\right)\Phi_{h^*,n}.\end{aligned} \tag{8}$$

Note that the recursive SIR procedure can only be used when $n > p$ in order to have an initial value for the inverse of Σ_{n-1}, the corresponding initial value of $\hat{\theta}_{n-1}$ being given by $\tilde{\theta}_{n-1}$.

The SIR estimators $\tilde{\theta}_n$ and $\hat{\theta}_n$ share the same asymptotic properties, previously established in [28], under the following classical hypothesis.

(H_1): The random vectors (X_k) are square integrable, independent and identically distributed and $(X_1, Y_1), \ldots, (X_n, Y_n)$ are independently drawn from Equation (1).

LEMMA 2.1 *Assume that (LC) and (H_1) hold. Then, $\hat{\theta}_n$ converges a.s. to θ,*

$$\|\hat{\theta}_n - \theta\|^2 = \mathcal{O}\left(\frac{\log(\log n)}{n}\right) \quad \text{a.s.} \tag{9}$$

In addition, we also have the asymptotic normality

$$\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \Delta), \tag{10}$$

where the limiting covariance matrix Δ may be explicitly calculated.

Since this recursive SIR estimator is based on $H = 2$ slices, only one EDR direction can be estimated. When we consider a number H of slices greater or equal to 3, it seems not possible to obtain an analytical expression of the EDR direction, that is of the major eigenvector of $\Sigma^{-1}\Gamma$ where

$$\Gamma = \sum_{h=1}^{H} p_h z_h z_h'.$$

Bercu et al. [29] provided a recursive estimator of $\Sigma^{-1}\Gamma$, but not of the corresponding major eigenvector. They also proposed a new method called SIRoneslice that can be used when the regression model is a single index model. The SIRoneslice estimator of the EDR direction is based on the use of only one ‘optimal’ slice chosen among the H slices. They provided its recursive version and established some asymptotic results for the SIRoneslice approach. Good numerical performances have been obtained on simulations and the main advantage of using recursive versions of the SIR and SIRoneslice methods is clearly the gain in term of computing times.

In this paper, we will show that the recursive SIR procedure based on $H = 2$ slices is clearly faster than the standard SIR approach based on $H = 2$ slices. The reader can also find numerical results on these comparisons in [28]. One explanation of this gain is certainly due to the recursive calculation of the $p \times p$ matrix Σ_n^{-1}. In Section 4, we provide the computing times for our proposed estimator of $f(\theta' x)$, based on the recursive SIR estimator of θ and the associated recursive Nadaraya–Watson estimator of f.

3. Main results

Our purpose is to investigate the asymptotic properties of the recursive Nadaraya–Watson estimator $\hat{f}_n$ of the link function f given by Equation (2). First of all, we assume that the kernel K is a positive symmetric function, bounded with compact support, twice differentiable with bounded derivatives, satisfying

$$\int_{\mathbb{R}} K(x)\,\mathrm{d}x = 1 \quad \text{and} \quad \int_{\mathbb{R}} K^2(x)\,\mathrm{d}x = \nu^2.$$

Moreover, it is necessary to add the following standard hypothesis.
(H_2): The probability density function g associated with (X_n) is continuous, positive on all $\mathbb{R}^p$, twice differentiable with bounded derivatives.

(H_3): The link function f is Lipschitz.
Our first result deals with the almost sure convergence of the estimator $\hat{f}_n$.

THEOREM 3.1 *Assume that* (*LC*) *and* (H_1) *to* (H_3) *hold. In addition, suppose that the sequence* (X_n) *has a finite moment of order* $a > 2$. *Then, for any* $x \in \mathbb{R}$, *we have*

$$\lim_{n\to\infty} \hat{f}_n(x) = f(x) \quad \text{a.s.} \tag{11}$$

More precisely, if the bandwidth (h_n) *is given by* $h_n = 1/n^{\alpha}$ *with* $0 < \alpha < \frac{1}{3}$,

$$\hat{f}_n(x) - f(x) = \mathcal{O}(n^{-\alpha}) + \mathcal{O}\left(n^{1/a}\sqrt{\frac{\log(\log n)}{n}}\right) \quad \text{a.s.} \tag{12}$$

while, if $\frac{1}{3} \leq \alpha < 1$,

$$\hat{f}_n(x) - f(x) = \mathcal{O}(\sqrt{n^{\alpha-1}}\log n) + \mathcal{O}\left(n^{1/a}\sqrt{\frac{\log(\log n)}{n}}\right) \quad \text{a.s.} \tag{13}$$

Proof The proof is given in Appendix 2. ∎

Remark 3.1 In the particular case where (X_n) is a sequence of independent random vectors of $\mathbb{R}^p$ sharing the same $\mathcal{N}(m, \Sigma)$ distribution where the covariance matrix Σ is positive definite, we can replace $n^{1/a}$ by $\log n$ into Equations (12) and (13). Consequently, for any $x \in \mathbb{R}$, we obtain that if $0 < \alpha < \frac{1}{3}$,

$$\hat{f}_n(x) - f(x) = \mathcal{O}(n^{-\alpha}) \quad \text{a.s.}$$

while, if $\frac{1}{3} \le \alpha < 1$,

$$\hat{f}_n(x) - f(x) = \mathcal{O}(\sqrt{n^{\alpha-1}} \log n) \quad \text{a.s.}$$

The asymptotic normality of the estimator $\hat{f}_n$ is as follows.

THEOREM 3.2 *Assume that (LC) and (H_1) to (H_3) hold. In addition, suppose that the sequence (X_n) has a finite moment of order $a = 6$ and that the sequence (ε_n) has a finite conditional moment of order $b > 2$. Then, as soon as the bandwidth (h_n) satisfies $h_n = 1/n^\alpha$ with $\frac{1}{3} < \alpha < 1$, we have for any $x \in \mathbb{R}$, the pointwise asymptotic normality*

$$\sqrt{nh_n}(\hat{f}_n(x) - f(x)) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \frac{\sigma^2 \nu^2}{(1+\alpha)h(\theta, x)}\right), \tag{14}$$

where $h(\theta, x)$ stands for the probability density function associated with ($\theta' X_n$).

Proof The proof is given in Appendix 3. ■

4. Numerical simulations

The goal of this section is to illustrate via some numerical experiments the theoretical results of Section 3. We will provide the numerical behaviour of our recursive estimators combining the recursive Nadaraya–Watson estimator of the link function f together with the recursive SIR estimator of the parameter θ. First of all, we describe in Section 4.1 the simulated model used in the numerical study and we present the estimation procedure, in particular the choice of the bandwidth parameter α by a cross-validation criterion.

We compare in Section 4.2 the computing times of the four methods and we will observe that the recursive procedures are the fastest ones as it was expected.

Then, we illustrate in Sections 4.3 and 4.4 the almost sure convergence and the asymptotic normality of our recursive Nadaraya–Watson estimator of f.

4.1. *Simulated model and estimation procedures*

We consider the semiparametric regression model given, for all $k \ge 1$, by

$$Y_k = f(\theta' X_k) + \varepsilon_k, \tag{M}$$

where the link function f is defined, for all $x \in \mathbb{R}$, by

$$f(x) = x \exp\left(\frac{3x}{4}\right).$$

The parameter θ belongs to $\mathbb{R}^p$ with $p = 10$ and it is given by

$$\theta = \frac{1}{\sqrt{10}}(1, 2, -2, -1, 0, \ldots, 0).$$

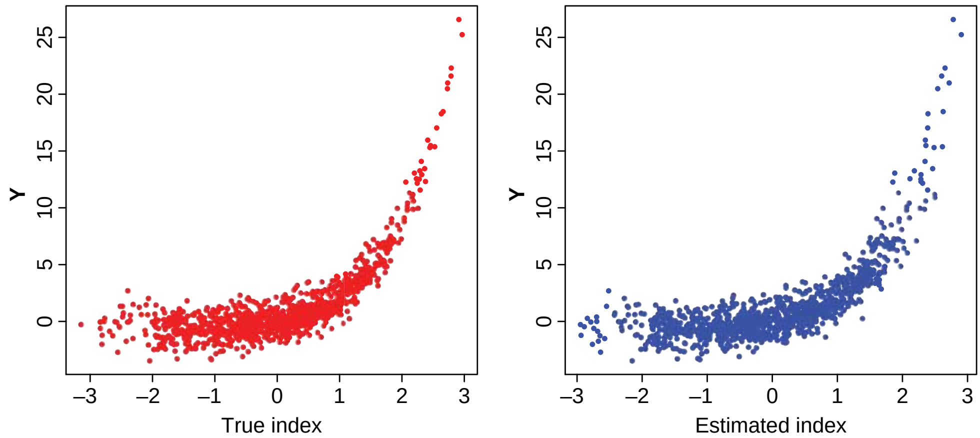

Figure 1. Scatterplots of $\{(\theta' X_1, Y_1), \dots, (\theta' X_n, Y_n)\}$ (on the left) and $\{(\hat{\theta}_n' X_1, Y_1), \dots, (\hat{\theta}_n' X_n, Y_n)\}$ (on the right), when $p = 10$, $\Sigma = \mathrm{I}_p$ and $= 1000$.

Moreover, (X_k) is a sequence of independent random vectors of $\mathbb{R}^p$ sharing the same $\mathcal{N}(0, \Sigma)$ distribution, while (ε_k) is a sequence of independent random variables with standard $\mathcal{N}(0, 1)$ distribution, independent of (X_k). We consider two kinds of covariance matrix $\Sigma : \Sigma = \mathrm{I}_p$ and $\Sigma = VV' + 0.1\mathrm{I}_p$ where the components v_{kl} of the $p \times p$ matrix V are generated from $\mathcal{N}(0, 0.3)$ distribution. In Figure 1, we present two scatterplots for a sample of size $n = 1000$ generated from model (M) with $\Sigma = \mathrm{I}_p$. On the left side, one can observe the data in the 'true' reduction subspace, that is the scatterplot of $(\theta' X_1, Y_1), \dots, (\theta' X_n, Y_n)$ based on the 'true' EDR direction θ. On the right side, we plot the data obtained from the estimated EDR direction $\hat{\theta}_n$ calculated via our recursive SIR procedure, that is the scatterplot of $(\hat{\theta}_n' X_1, Y_1), \dots, (\hat{\theta}_n' X_n, Y_n)$. One can clearly notice that the EDR direction has been well estimated.

For the recursive Nadaraya–Watson estimator $\hat{f}_n$ of f, we have chosen the well-known Epanechnikov kernel

$$K(x) = \tfrac{3}{4}(1 - x^2)\mathrm{I}_{\{|x| \leq 1\}}$$

and the bandwidth $h_n = 1/n^\alpha$ with $0 < \alpha < 1$. We now need to evaluate an optimal value for the smoothing parameter α. The problem of deciding how much to smooth is of great importance in nonparametric regression. We propose to make use of the optimal data-driven bandwidth α which minimizes the cross-validation criterion

$$\mathrm{CV}(\alpha) = \sum_{k=p+1}^{n} (Y_k - \hat{Y}_{k,\alpha})^2 \quad \text{where } \hat{Y}_{k,\alpha} = \hat{f}_{k-1}(\hat{\theta}_{k-1}' X_k).$$

To illustrate the numerical behaviour of our proposed cross-validation criterion, we consider simulated samples of sizes $n = 200$, 500, 1000 and 2000 generated from model (M) with $p = 10$. In Figure 2, we present the corresponding $\mathrm{CV}(\alpha)$ criteria. We can observe that the $\mathrm{CV}(\alpha)$ functions are all convex and the corresponding optimal data-driven bandwidth α lies into the interval $[0.33, 0.38]$. Consequently, in all Section 4, we have chosen the optimal value $\alpha = 0.35$.

4.2. *Comparison of computing times between non recursive and recursive procedures*

In this part of the simulation study, we only focus on the computing times of two estimators: the proposed recursive estimator and the corresponding non recursive one, based on usual SIR in the first step and classical Nadaraya–Watson estimator in the second step.

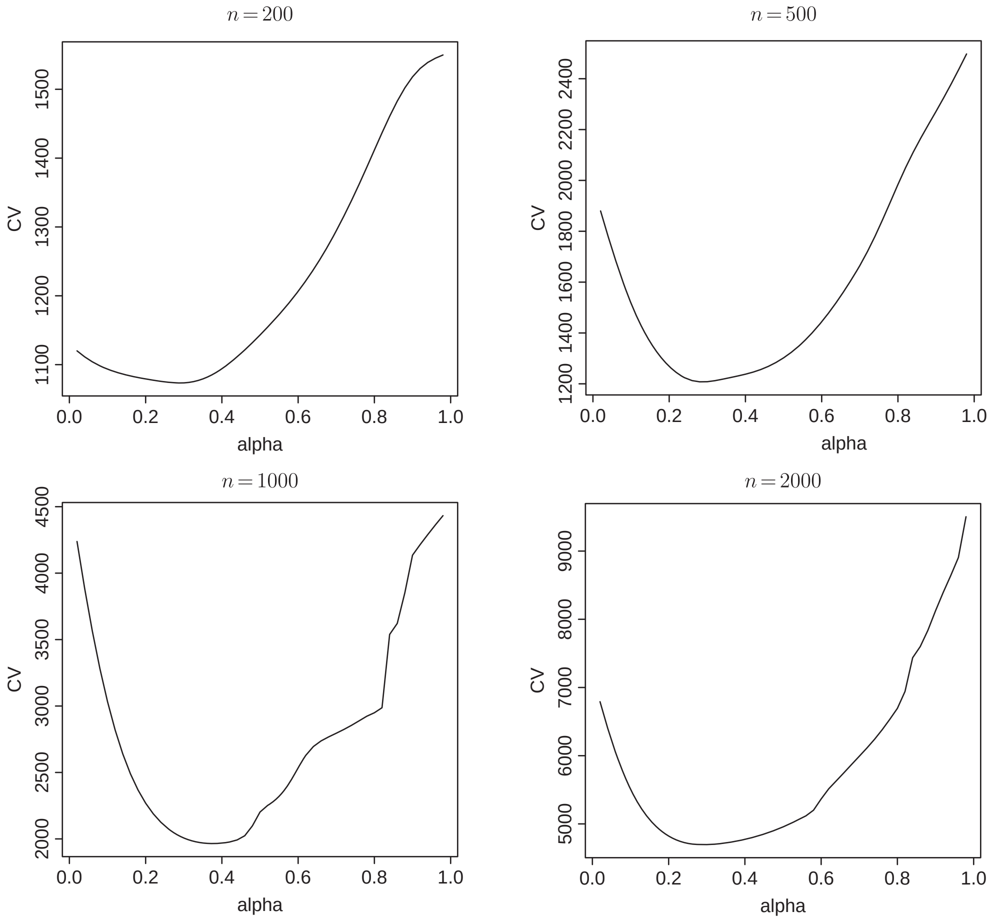

Figure 2. Examples of the cross-validation function $\mathrm{CV}(\alpha)$ for the simulated sample of fixed sizes $n = 200, 500, 1000$ and 2000 generated from model (M) with $p = 10$.

For each method and for a given simulated sample of size n, we measure in seconds the computing (CPU) time needed to calculate the corresponding estimators $\hat{f}_n(\hat{\theta}_n' x)$ of $f(\theta' x)$ for n going from $N_0 = p + 1$ to N, where $\hat{\theta}_n$ and $\hat{f}_n$ are the estimators of f and θ only based on the first n observations of the sample. More precisely, the computing time is the global time needed to calculate the $N - N_0 + 1$ estimators: $\hat{f}_n(\hat{\theta}_{N_0}' x), \hat{f}_n(\hat{\theta}_{N_0+1}' x), \ldots, \hat{f}_n(\hat{\theta}_{N-1}' x)$ and $\hat{f}_n(\hat{\theta}_N' x)$.

For various values of p and N, we generate $\mathcal{B} = 100$ replicated samples from model (M). Then, for each method and each simulated sample, we estimate $f(\theta' x)$ for 10 values of x, randomly generated from the uniform distribution on $[-1.5; 1.5]^p$, using the two above-mentioned estimators.

In Table 1, we give the means of computing times evaluated on the $\mathcal{B} = 100$ replicated samples for different values of $p = 5, 10, 25, 50$ and $N = 500, 1000$ and 2000.

From the reading of Table 1, one can give the following comments. For both methods, not surprisingly, the larger are the dimension p or the size N, the larger is the mean of computing times. However the recursive approach clearly provides much smaller mean of computing times in comparison with the non recursive approach. One explanation of this gain in term of computing time is certainly due to the recursive calculation of the $p \times p$ matrix $\hat{\Sigma}_n^{-1}$ in the SIR step. Note that when the dimension p increases, the computing time in mean of recursive method only slightly increases contrary to the non recursive one: for instance, the computing time is multiplied by around 2 from $p = 5$ to $p = 50$ for the recursive approaches, whereas it is multiplied by more than 5 for non recursive method. According to size N, the computing time increases faster for the non recursive approach than for the recursive one (for which the computing time is multiplied by two when the sample size doubles). To conclude, we clearly exhibit the great

Table 1. Mean of computing (CPU) times in seconds, over $\mathcal{B} = 100$ replicated samples from the simulated model (M) for different values of p and N, for calculating estimates $\hat{f}_n(\hat{\theta}_n' x)$ of $f(\theta' x)$ (for n going from $p+1$ to N) with our proposed recursive method and the corresponding non recursive one.

		$p = 5$	$p = 10$	$p = 25$	$p = 50$
$N = 500$	Recursive estimator	0.25	0.27	0.31	0.48
	Non recursive estimator	1.52	1.88	3.31	7.44
$N = 1000$	Recursive estimator	0.51	0.52	0.63	1.02
	Non recursive estimator	3.33	4.32	8.05	18.63
$N = 2000$	Recursive estimator	1.02	1.05	1.28	2.08
	Non recursive estimator	7.88	10.77	21.95	51.38

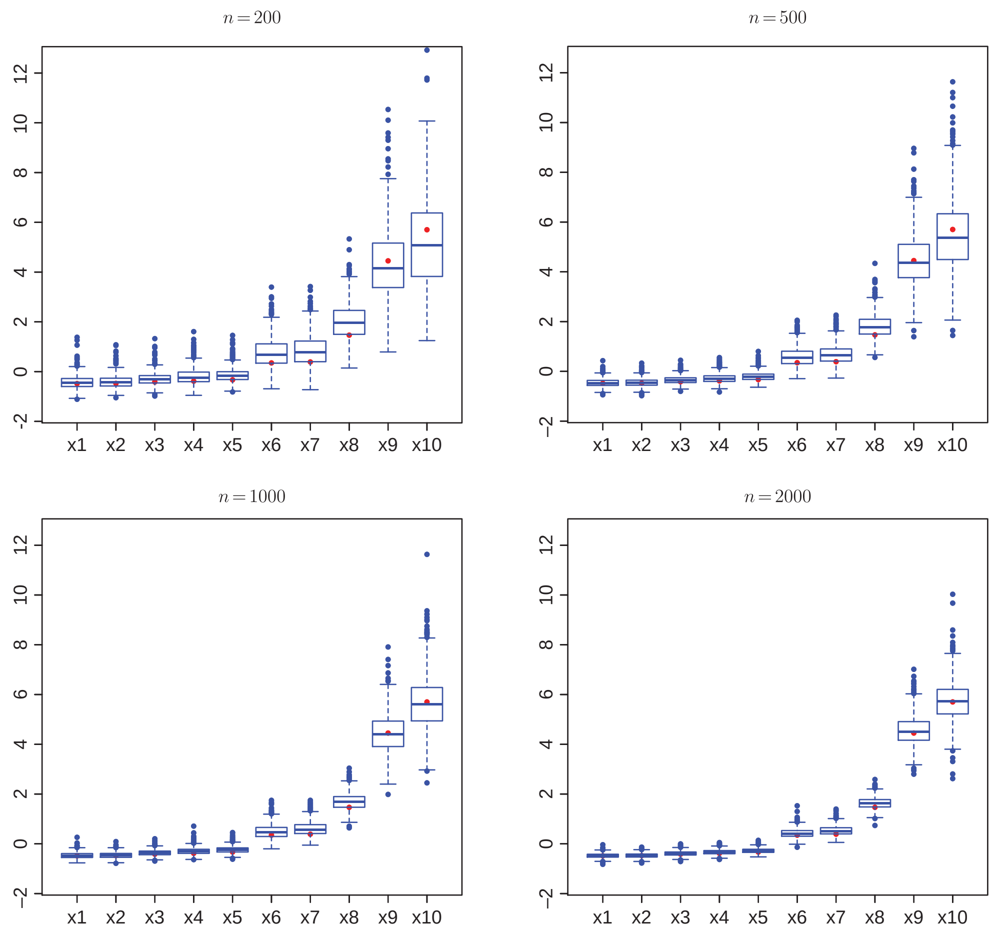

Figure 3. Almost sure convergence of $\hat{f}_n(\hat{\theta}_n' x)$ to $f(\theta' x)$ for 10 different values of x, when $p = 10$ and $\Sigma = \mathrm{I}_p$.

advantage of using the proposed recursive approach in terms of computing times: for instance for high-dimensional data ($p = 50$ and $N = 2000$), the recursive approach is around 25 times faster than the non recursive approach.

4.3. *Almost sure convergence*

The good numerical performances of the recursive SIR estimator $\hat{\theta}_n$ were already illustrated in [28,29]. In order to keep this section brief, we only focus our attention on the almost sure convergence of $\hat{f}_n$. We generate $N = 1000$ samples of different sizes $n = 200, 500, 1000, 2000$ from model (M) with $p = 10$. For each sample and each kind of matrix Σ, we calculate the estimation $\hat{f}_n(\hat{\theta}_n' x)$ of $f(\theta' x)$ for 10 different values $x_1, \ldots, x_{10}$ of $x \in \mathbb{R}^p$, randomly generated from the uniform distribution on $[-1.5; 1.5]^{10}$. To obtain pleasant-looking graphics, we have sorted the x_j's in ascending order according to their corresponding 'true' value $f(\theta' x_j)$, see the red dots in Figures 3 and 4.

The boxplots of the $\hat{f}_n(\hat{\theta}_n' x_j)$'s are respectively given in Figure 3 (resp. Figure 4) when $\Sigma = \mathrm{I}_p$ (resp. when $\Sigma = VV' + 0.1\mathrm{I}_p$). Note that the red circle point in each boxplot which represents the true value $f(\theta' x_j)$ allows us to easily judge the quality of the estimator. One can observe that the dispersion of the $\hat{f}_n(\hat{\theta}_n' x)$'s are small and the mean is very close to the true value $f(\theta' x)$. One can also notice that the larger is the sample size n, the greater is the quality measure. As it was expected, the quality of the estimation decreases for large values of $f(\theta' x)$ since the number of observations around x decreases, see the scatterplots of Figure 1 to be convinced. Finally,

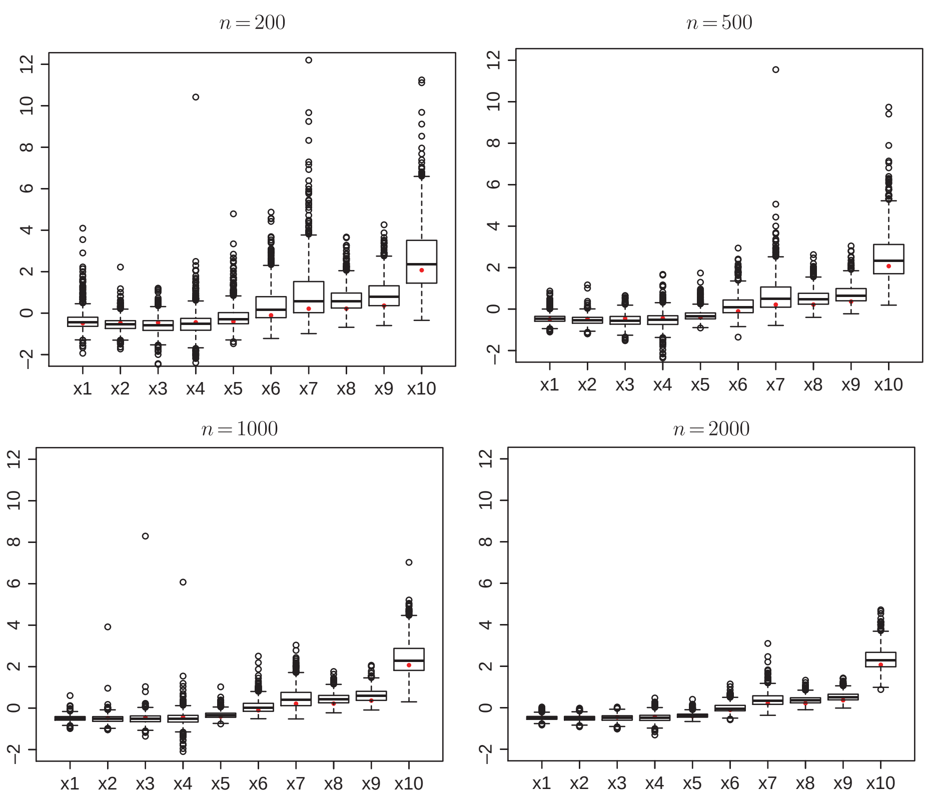

Figure 4. Almost sure convergence of $\hat{f}_n(\hat{\theta}_n' x)$ to $f(\theta' x)$ for 10 different values of x, when $p = 10$ and $\Sigma = VV' + 0.1\mathrm{I}_p$.

one can also observe that the form of covariance matrix Σ does not impact the quality of the estimates.

4.4. *Asymptotic normality*

In order to illustrate the asymptotic normality of our recursive Nadaraya–Watson estimator, we generate $N = 1000$ realizations of $\hat{f}_n(\hat{\theta}_n' x)$ based on samples from model (M) for various values of $n = 500, 1000, 2000$ and $p = 5, 10, 20$.

In Figure 5, we plot histograms of the standardized values of the $\hat{f}_n(\hat{\theta}_n' x)$'s for one value x_1 of $x \in \mathbb{R}^p$ randomly generated from the uniform distribution on $[-1.5; 1.5]^p$ with $p = 10$, various sample sizes $n = 500$, 2000 and 5000, and two covariance matrices $\Sigma = I_p$ and $VV' + 0.1\mathrm{I}_p$. On each histogram, we add the curve of the standard normal density. One can clearly see that the

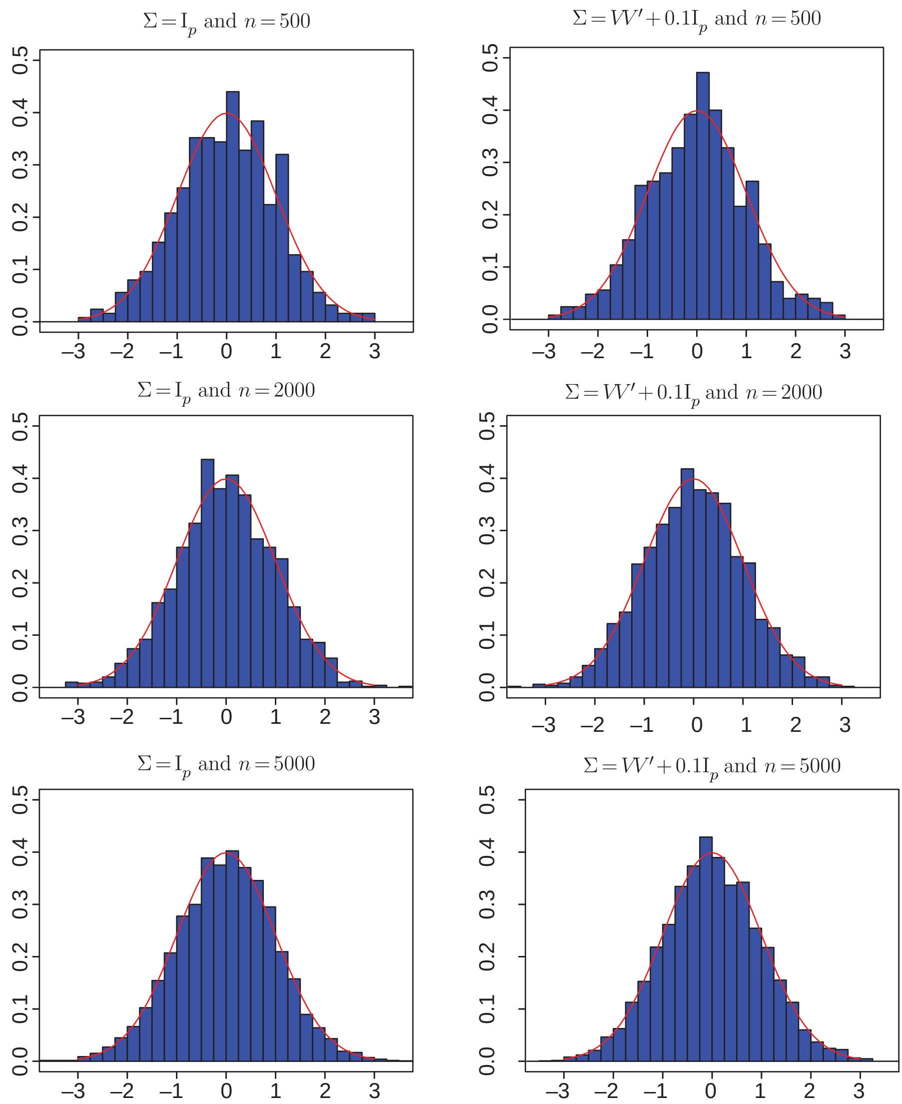

Figure 5. Asymptotic normality of $\hat{f}_n(\hat{\theta}_n' x)$ to $f(\theta' x)$ with $x = x_1$ and $p = 10$, for various values of $n = 500$, 2000 and 5000 and $\Sigma = \mathrm{I}_p$ and $VV' + 0.1\mathrm{I}_p$.

Table 2. Mean over 10 different values of x of skewness and standardized kurtosis coefficients of the distributions of the $\hat{f}_n(\hat{\theta}_n' x)$'s, standard deviations are given in parentheses.

		$n = 500$	$n = 1000$	$n = 2000$
$p = 5$	Skewness	0.148 (0.102)	0.123 (0.087)	0.063 (0.056)
	Kurtosis	0.139 (0.112)	−0.077 (0.086)	0.078 (0.064)
$p = 10$	Skewness	0.159 (0.125)	0.124 (0.099)	0.070 (0.066)
	Kurtosis	0.159 (0.179)	0.084 (0.93)	−0.053 (0.044)
$p = 20$	Skewness	0.191 (0.167)	0.141 (0.104)	0.082 (0.081)
	Kurtosis	−0.148 (0.121)	0.095 (0.087)	0.073 (0.066)

normal density coincides pretty well with all the histograms as the sample size n increases. This visually shows the asymptotic normality of our recursive Nadaraya–Watson estimator $\hat{f}_n$ of f.

Furthermore, to enforce understanding asymptotic normality, skewness and kurtosis of distribution of simulated estimator have been calculated and tabulated in Table 2 for the different values of n and p and data generated from model (M) with Σ of the form $VV' + 0.1\mathrm{I}_p$. Note that we used a standardized version of kurtosis such that the sample values of skewness and kurtosis should be close to zero when the simulated sample is generated from a normal distribution. These coefficients are theoretically equal to zero when they are defined as measures of the asymmetry and the 'peakedness' of the probability distribution of a real-valued random variable.

For each values of n and p, we used 10 different values of x randomly generated from the uniform distribution on $[-1.5; 1.5]^p$. For each value of x, we calculate the corresponding sample skewness and standardized kurtosis. Finally, for each pair (n, p), we calculate the mean and the standard deviation of these 10 sample skewness (resp. kurtosis) coefficients. One can observe in Table 2 that the mean values of skewness and standardized kurtosis are close to zero, the proximity to zero seems to increases with the sample size n and to slightly decrease with the dimension p of the covariable. From these numerical results, we highlight the asymptotic normality of our recursive Nadaraya–Watson estimate. Note that we obtain very similar results when data were generated from model (M) with $\Sigma = \mathrm{I}_p$

5. Conclusion

In this paper, the asymptotic behaviour of the recursive Nadaraya–Watson estimator for the estimation of the regression function in a single index regression model has been investigated from theoretical and numerical points of view. First, we use a recursive version of SIR for the estimation of the unknown euclidean parameter of the underlying semiparametric regression model. Then, we consider a recursive Nadaraya–Watson procedure for the estimation of the regression function which takes into account the previous estimation of the parameter of the underlying model. The almost sure convergence as well as the asymptotic normality for our Nadaraya–Watson estimate have been established under standard regularity assumptions on the kernel. The good numerical behaviour of the proposed estimation procedure has been illustrated on simulated data. A major advantage of this recursive estimation procedure is that the recursive approach clearly provides very smaller computing times in comparison with the non recursive approach.

Acknowledgements

The authors thank the Editor, the Associate Editor and the two anonymous referees for their helpful remarks and their valuable comments and suggestions which greatly improved the paper.

References

[1] Li K-C. Sliced inverse regression for dimension reduction (with discussions). J Am Stat Assoc. 1991;86:316–342.
[2] Duan N, Li K-C. Slicing regression: a link-free regression method. Ann Stat. 1991;19:505–530.
[3] Horowitz JL. Semiparametric methods in econometrics. Lecture Notes in Statistics, Vol. 131. New York: Springer-Verlag; 1998.
[4] Devroye L, Lugosi G. Combinatorial methods in density estimation. Springer Series in Statistics. New York: Springer-Verlag; 2001.
[5] Nadaraya EA. Nonparametric estimation of probability densities and regression curves. Mathematics and its applications. Dordrecht: Kluwer Academic Publishers Group; 1989.
[6] Silverman BW. Density estimation for statistics and data analysis. Monographs on Statistics and Applied Probability. London: Chapman and Hall; 1986.
[7] Tsybakov AB. Introduction à l'estimation non-paramétrique (in French). Mathématiques et Applications. Berlin: Springer-Verlag; 2004.
[8] Nadaraya EA. On a regression estimate. Teor. Verojatnost. i Primenen. 1964;9:157–159.
[9] Watson GS. Smooth regression analysis. Sankhya Ser A. 1964;26:359–372.
[10] Noda K. Estimation of a regression function by the Parzen kernel-type density estimators. Ann Inst Stat Math. 1976;28:221–234.
[11] Schuster EF. Joint asymptotic distribution of the estimated regression function at a finite number of distinct points. Ann Math Stat. 1972;43:84–88.
[12] Choi E, Hall P, Rousson V. Data sharpening methods for bias reduction in nonparametric regression. Ann Stat. 2000;28:1339–1355.
[13] Duflo M. Random iterative models. Berlin: Springer-Verlag; 1997.
[14] Cook RD. Principal Hessian directions revisited (with discussion). J Am Stat Assoc. 1998;93:84–100.
[15] Cook RD, Li B. Dimension reduction for conditional mean in regression. Ann Stat. 2002;30:455–474.
[16] Hall P, Li K-C. On almost linearity of low-dimensional projections from high-dimensional data. Ann Stat. 1993;21:867–889.
[17] Chen C-H, Li K-C. Can SIR be as popular as multiple linear regression? Statist Sin. 1998;8:289–316.
[18] Cook RD, Ni L. Sufficient dimension reduction via inverse regression: a minimum discrepancy approach. J Am Stat Assoc. 2005;100:410–418.
[19] Prendergast LA. Influence functions for sliced inverse regression. Scand J Stat. 2005;32:385–404.
[20] Szretter ME, Yohai VJ. The sliced inverse regression algorithm as a maximum likelihood procedure. J Stat Plan Inference. 2009;139:3570–3578.
[21] Zhu L-P, Yu Z. On spline approximation of sliced inverse regression. Sci China Ser A: Math. 2007;50:1289–1302.
[22] Wu H-M. Kernel sliced inverse regression with applications to classification. J Comput Graph Stat. 2008;17:590–610.
[23] Azaïs R, Gégout-Petit A, Saracco J. Optimal quantization applied to sliced inverse regression. J Stat Plan Inference. 2012;142:481–492.
[24] Li L, Nachtsheim CJ. Sparse sliced inverse regression. Technometrics. 2006;48:503–510.
[25] Scrucca L. Class prediction and gene selection for DNA microarrays using regularized sliced inverse regression. Comput Stat Data Anal. 2007;52:438–451.
[26] Li L, Yin X. Sliced inverse regression with regularizations. Biometrics. 2008;64:124–131.
[27] Zhu LX, Ohtaki M, Li Y. On hybrid methods of inverse regression-based algorithms. Comput Stat. 2007;51:2621–2635.
[28] Nguyen TMN, Saracco J. Recursive estimation for sliced inverse regression (in French). Journal de la Société Française de Statistique. 2010;151:19–46.
[29] Bercu B, Nguyen TMN, Saracco J. A new approach on recursive and non-recursive SIR methods. J Korean Stat Soc. 2012;41:17–36.

Appendix 1. Analytical expression of the EDR direction $\tilde{\theta}$

Let us recall that $\tilde{\theta}$ is the eigenvector of $\Sigma^{-1}\Gamma$ associated with the non-null eigenvalue λ when we consider $H = 2$ slices. We assume in the following that $0 < p_1 < 1$, so $p_2 = 1 - p_1 > 0$. We clearly have

$$p_1 z_1 + p_2 z_2 = p_1 m_1 + p_2 m_2 - (p_1 + p_2)\mu = p_1 m_1 + p_2 m_2 - \mu = \mu - \mu = 0. \tag{A1}$$

Moreover, we already saw that

$$\Gamma = p_1 z_1 z_1' + p_2 z_2 z_2'. \tag{A2}$$

Clearly, the rank of the matrix Γ is one. We will show that $\tilde{\theta} = \Sigma^{-1}(z_1 - z_2)$ is the major eigenvector of $\Sigma^{-1}\Gamma$. From Equation (A2), we get

$$\begin{aligned}\Sigma^{-1}\Gamma\tilde{\theta} &= \Sigma^{-1}\Gamma\Sigma^{-1}(z_1 - z_2)\\ &= q_1p_1\Sigma^{-1}z_1 - q_{12}p_1\Sigma^{-1}z_1 + q_{12}p_2\Sigma^{-1}z_2 - q_2p_2\Sigma^{-1}z_2,\end{aligned}$$

where $q_1 = z_1'\Sigma^{-1}z_1$, $q_2 = z_2'\Sigma^{-1}z_2$ and $q_{12} = z_1'\Sigma^{-1}z_2 = z_2'\Sigma^{-1}z_1$. Moreover, from Equation (A1), we also have $p_1q_1 = -p_2q_{12}$ et $p_2q_2 = -p_1q_{12}$. Then, we easily deduce that

$$\begin{aligned}\Sigma^{-1}\Gamma\tilde{\theta} &= p_2q_{12}\Sigma^{-1}(z_2 - z_1) + p_1q_{12}\Sigma^{-1}(z_2 - z_1)\\ &= -q_{12}\Sigma^{-1}(z_1 - z_2).\end{aligned}$$

Using once again Equation (A1), we find that

$$-q_{12} = \frac{p_1}{p_2}z_1'\Sigma^{-1}z_1.$$

Finally, by denoting

$$\lambda = \frac{p_1}{p_2}z_1'\Sigma^{-1}z_1,$$

we obtain that $\Sigma^{-1}\Gamma\tilde{\theta} = \lambda\tilde{\theta}$. The vector $\tilde{\theta}$ is thus the eigenvector associated with the largest eigenvalue λ of the matrix $\Sigma^{-1}\Gamma$.

Appendix 2: Proof of Theorem 3.1

In order to prove the almost sure pointwise convergence of Theorem 3.1, we shall denote for all $x \in \mathbb{R}$

$$P_n(x) = \sum_{k=p+1}^{n} W_k(x)\varepsilon_k, \quad N_n(x) = \sum_{k=p+1}^{n} W_k(x),$$

and

$$Q_n(x) = \sum_{k=p+1}^{n} W_k(x)(f(\Phi_k) - f(x))$$

where $\Phi_n = \theta' X_n$. We clearly obtain from Equation (1) the main decomposition

$$\hat{f}_n(x) - f(x) = \frac{P_n(x) + Q_n(x)}{N_n(x)}. \tag{A3}$$

We shall establish the asymptotic behaviour of each sequence $(P_n(x))$, $(Q_n(x))$ and $(N_n(x))$. Let $(\mathcal{F}_n)$ be the filtration given by $\mathcal{F}_n = \sigma(X_1, \ldots, X_n, Y_1, \ldots, Y_n)$. First of all, we can split $N_n(x)$ into two terms,

$$N_n(x) = M_n^{(N)}(x) + R_n^{(N)}(x), \tag{A4}$$

where

$$M_n^{(N)}(x) = \sum_{k=p+1}^{n} (W_k(x) - \mathbb{E}[W_k(x)|\mathcal{F}_{k-1}]) \quad \text{and} \quad R_n^{(N)}(x) = \sum_{k=p+1}^{n} \mathbb{E}[W_k(x)|\mathcal{F}_{k-1}].$$

On the one hand, we have

$$\mathbb{E}[W_n(x) \mid \mathcal{F}_{n-1}] = \frac{1}{h_n}\int_{\mathbb{R}^p} K\left(\frac{x - \hat{\theta}_{n-1}' x_n}{h_n}\right) g(x_n)\, dx_n.$$

We can assume without loss of generality that, for n large enough, at least one component of $\hat{\theta}_n$ is different from zero a.s. As a matter of fact, we already saw from Lemma 2.1 that $\hat{\theta}_n$ converges a.s. to θ which is different from the null vector.

For the sake of simplicity, suppose that the first component $\hat{\theta}_{n-1,1} \neq 0$ a.s. We can make the change of variables

$$z = \frac{x - \hat{\theta}'_{n-1} x_n}{h_n}$$

and $z_2 = x_{n,2}, \ldots, z_p = x_{n,p}$. The Jacobian of this linear transformation is given by

$$J = -\frac{h_n}{\hat{\theta}_{n-1,1}}.$$

Consequently, we obtain that

$$\mathbb{E}[W_n(x) \mid \mathcal{F}_{n-1}] = \int_{\mathbb{R}} K(z) h(\hat{\theta}_{n-1}, x - z h_n)\, \mathrm{d}z, \tag{A5}$$

where

$$h(\hat{\theta}_{n-1}, x) = \frac{1}{|\hat{\theta}_{n-1,1}|} \int_{\mathbb{R}^{p-1}} g\left(\frac{1}{\hat{\theta}_{n-1,1}}\left(x - \sum_{k=2}^{p} \hat{\theta}_{n-1,k} z_k\right), z_2, \ldots, z_p\right) \mathrm{d}z_2 \ldots \mathrm{d}z_p.$$

One can observe that $h(\theta, x)$ is exactly the probability density function associated with the identically distributed sequence $(\theta' X_n)$. Therefore, as the probability density function g is continuous, twice differentiable with bounded derivatives, we deduce from Equation (A5) together with Taylor's formula that

$$\begin{aligned}
\mathbb{E}[W_n(x) \mid \mathcal{F}_{n-1}] &= \int_{\mathbb{R}} K(z) \Bigg(h(\hat{\theta}_{n-1}, x) - z h_n h'(\hat{\theta}_{n-1}, x) \\
&\qquad + \frac{z^2 h_n^2}{2} h''(\hat{\theta}_{n-1}, x - z h_n \xi) \Bigg) \mathrm{d}z, \\
&= h(\hat{\theta}_{n-1}, x) + \frac{h_n^2}{2} \int_{\mathbb{R}} z^2 K(z) h''(\hat{\theta}_{n-1}, x - z h_n \xi)\, \mathrm{d}z,
\end{aligned}$$

where $0 < \xi < 1$. Consequently, for n large enough,

$$|\mathbb{E}[W_n(x)|\mathcal{F}_{n-1}] - h(\hat{\theta}_{n-1}, x)| \leq M_h \tau^2 h_n^2 \quad \text{a.s.} \tag{A6}$$

where

$$M_h = \sup_{x \in \mathbb{R}} |h''(\hat{\theta}_{n-1}, x)| \quad \text{and} \quad \tau^2 = \frac{1}{2} \int_{\mathbb{R}} x^2 K(x)\, \mathrm{d}x.$$

Hence, we find from Equation (A6) that

$$\sum_{k=p+1}^{n} |\mathbb{E}[W_k(x) \mid \mathcal{F}_{k-1}] - h(\hat{\theta}_{k-1}, x)| = \mathcal{O}\left(\sum_{k=p+1}^{n} h_k^2\right) \quad \text{a.s.}$$

It follows from the continuity of h together with the fact that $\hat{\theta}_n$ converges to θ a.s. and h_n goes to zero that

$$\lim_{n \to \infty} \frac{1}{n} \sum_{k=p+1}^{n} \mathbb{E}[W_k(x)|\mathcal{F}_{k-1}] = h(\theta, x) \quad \text{a.s.} \tag{A7}$$

which of course immediately implies that for all $x \in \mathbb{R}$

$$\lim_{n \to \infty} \frac{R_n^{(N)}(x)}{n} = h(\theta, x) \quad \text{a.s.} \tag{A8}$$

On the other hand, $(M_n^{(N)}(x))$ is a square integrable martingale difference sequence with predictable quadratic variation given by

$$\begin{aligned}
\langle M^{(N)}(x) \rangle_n &= \sum_{k=p+1}^{n} \mathbb{E}[(M_k^{(N)}(x) - M_{k-1}^{(N)}(x))^2 | \mathcal{F}_{k-1}], \\
&= \sum_{k=p+1}^{n} (\mathbb{E}[W_k^2(x)|\mathcal{F}_{k-1}] - \mathbb{E}^2[W_k(x)|\mathcal{F}_{k-1}]).
\end{aligned}$$

Via the same change of variables as in Equation (A5), we obtain that

$$\mathbb{E}[W_n^2(x)|\mathcal{F}_{n-1}] = \frac{1}{h_n}\int_{\mathbb{R}} K^2(z)h(\hat{\theta}_{n-1}, x - zh_n)\,\mathrm{d}z,$$

$$= \frac{1}{h_n}\int_{\mathbb{R}} K^2(z)\left(h(\hat{\theta}_{n-1}, x) - zh_n h'(\hat{\theta}_{n-1}, x) + \frac{z^2h_n^2}{2}h''(\hat{\theta}_{n-1}, x - zh_n\xi)\right)\mathrm{d}z,$$

where $0 < \xi < 1$. Consequently, for n large enough,

$$\left|\mathbb{E}[W_n^2(x)|\mathcal{F}_{n-1}] - \frac{\nu^2}{h_n}h(\hat{\theta}_{n-1}, x)\right| \le M_h\mu^2 h_n \quad \text{a.s.} \tag{A9}$$

where

$$\nu^2 = \int_{\mathbb{R}} K^2(x)\,\mathrm{d}x \quad \text{and} \quad \mu^2 = \frac{1}{2}\int_{\mathbb{R}} x^2K^2(x)\,\mathrm{d}x.$$

Hence, Equation (A9) ensures that

$$\sum_{k=p+1}^{n}\left|\mathbb{E}[W_k^2(x) \mid \mathcal{F}_{k-1}] - \frac{\nu^2}{h_k}h(\hat{\theta}_{k-1}, x)\right| = \mathcal{O}\left(\sum_{k=p+1}^{n} h_k\right) \quad \text{a.s.}$$

However, it is not hard to see that

$$\lim_{n\to\infty}\frac{1}{n^{1+\alpha}}\sum_{k=p+1}^{n}\frac{1}{h_k} = \frac{1}{1+\alpha}.$$

Therefore, it follows from Equation (A9) together with the almost sure convergence of $h(\hat{\theta}_n, x)$ to $h(\theta, x)$ and Toeplitz's lemma that

$$\lim_{n\to\infty}\frac{1}{n^{1+\alpha}}\sum_{k=p+1}^{n}\mathbb{E}[W_k^2(x) \mid \mathcal{F}_{k-1}] = \frac{\nu^2}{1+\alpha}h(\theta, x) \quad \text{a.s.} \tag{A10}$$

Furthermore, we also have from Equation (A6) that

$$\lim_{n\to\infty}\frac{1}{n}\sum_{k=p+1}^{n}\mathbb{E}^2[W_k(x)|\mathcal{F}_{k-1}] = h^2(\theta, x) \quad \text{a.s.} \tag{A11}$$

Consequently, we deduce from (A10) and (A11) that for all $x \in \mathbb{R}$,

$$\lim_{n\to\infty}\frac{\langle M^{(N)}(x)\rangle_n}{n^{1+\alpha}} = \frac{\nu^2}{1+\alpha}h(\theta, x) \quad \text{a.s.} \tag{A12}$$

We are now in position to make use of the strong law of large numbers for martingales given e.g. by Theorem 1.3.15 of Duflo.[13] As the probability density function g is positive on its support, we have for all $x \in \mathbb{R}$, $h(\theta, x) > 0$, which implies that $\langle M^{(N)}(x)\rangle_n$ goes to infinity a.s. Hence, for any $\gamma > 0$, $(M_n^{(N)}(x))^2 = o(n^{1+\alpha}(\log n)^{1+\gamma})$ a.s. which leads to

$$M_n^{(N)}(x) = o(n) \quad \text{a.s.} \tag{A13}$$

Then, we obtain from Equations (A4), (A8) and (A13) that for all $x \in \mathbb{R}$

$$\lim_{n\to\infty}\frac{N_n(x)}{n} = h(\theta, x) \quad \text{a.s.} \tag{A14}$$

We shall now investigate the asymptotic behaviour of the sequence $(P_n(x))$. Since (X_n) and (ε_n) are independent, $(P_n(x))$ is a square integrable martingale difference sequence with predictable quadratic variation given by

$$\langle P(x)\rangle_n = \sum_{k=p+1}^{n}\mathbb{E}[(P_k(x) - P_{k-1}(x))^2|\mathcal{F}_{k-1}] = \sigma^2\sum_{k=p+1}^{n}\mathbb{E}[W_k^2(x)|\mathcal{F}_{k-1}].$$

Then, it follows from convergence (A10) that

$$\lim_{n\to\infty}\frac{\langle P(x)\rangle_n}{n^{1+\alpha}} = \frac{\sigma^2\nu^2}{1+\alpha}h(\theta, x) \quad \text{a.s.} \tag{A15}$$

Consequently, we obtain from the strong law of large numbers for martingales that for any $\gamma > 0$ and that for all $x \in \mathbb{R}$,

$$P_n(x) = o(\sqrt{n^{1+\alpha}(\log n)^{1+\gamma}}) = o(n) \quad \text{a.s.} \tag{A16}$$

It remains to study the asymptotic behaviour of the sequence $(Q_n(x))$. We can split $Q_n(x)$ into two terms,

$$Q_n(x) = \Sigma_n(x) + \Delta_n(x), \tag{A17}$$

where $\hat{\Phi}_n = \hat{\theta}'_{n-1}X_n$,

$$\Sigma_n(x) = \sum_{k=p+1}^{n} W_k(x)(f(\Phi_k) - f(\hat{\Phi}_k)) \qquad \text{and} \qquad \Delta_n(x) = \sum_{k=p+1}^{n} W_k(x)(f(\hat{\Phi}_k) - f(x)).$$

The right-hand side of Equation (A17) is easy to handle. As a matter of fact, the kernel K is compactly supported which means that one can find a positive constant A such that K vanishes outside the interval $[-A, A]$. Thus, for all $n \geq 1$ and all $x \in \mathbb{R}$,

$$W_n(x) = \frac{1}{h_n}K\left(\frac{x - \hat{\theta}'_{n-1}X_n}{h_n}\right)\mathrm{I}_{\{|\hat{\theta}'_{n-1}X_n - x| \leq Ah_n\}}.$$

In addition, the function f is Lipschitz, so it exists a positive constant C_f such that for all $n \geq 1$

$$|f(\hat{\Phi}_n) - f(x)| \leq C_f|\hat{\Phi}_n - x| \leq C_f|\hat{\theta}'_{n-1}X_n - x|. \tag{A18}$$

Consequently, we obtain from Equation (A18) that for all $x \in \mathbb{R}$

$$\begin{aligned} |\Delta_n(x)| &\leq C_f \sum_{k=p+1}^{n} W_k(x)|\hat{\theta}'_{k-1}X_k - x|, \\ &\leq AC_f \sum_{k=p+1}^{n} h_k W_k(x). \end{aligned} \tag{A19}$$

Moreover, via the same lines as in the proof of Equation (A7), we find that

$$\lim_{n\to\infty} \frac{1}{n^{1-\alpha}} \sum_{k=p+1}^{n} h_k \mathbb{E}[W_k(x)|\mathcal{F}_{k-1}] = \frac{1}{1-\alpha}h(\theta, x) \quad \text{a.s.} \tag{A20}$$

Furthermore, denote

$$M_n^{(\Delta)}(x) = \sum_{k=p+1}^{n} h_k(W_k(x) - \mathbb{E}[W_k(x)|\mathcal{F}_{k-1}]).$$

One can observe that $(M_n^{(\Delta)}(x))$ is a square integrable martingale difference sequence with bounded increments and predictable quadratic variation given by

$$\begin{aligned} \langle M^{(\Delta)}(x)\rangle_n &= \sum_{k=p+1}^{n} \mathbb{E}[(M_k^{(\Delta)}(x) - M_{k-1}^{(\Delta)}(x))^2|\mathcal{F}_{k-1}], \\ &= \sum_{k=p+1}^{n} h_k^2(\mathbb{E}[W_k^2(x)|\mathcal{F}_{k-1}] - \mathbb{E}^2[W_k(x)|\mathcal{F}_{k-1}]). \end{aligned}$$

Hence, it follows from Equations (A6) and (A9) together with the almost sure convergence of $h(\hat{\theta}_n, x)$ to $h(\theta, x)$ and Toeplitz's lemma that

$$\lim_{n\to\infty} \frac{\langle M^{(\Delta)}(x)\rangle_n}{n^{1-\alpha}} = \frac{\nu^2}{1-\alpha}h(\theta, x) \quad \text{a.s.} \tag{A21}$$

Consequently, we obtain from the strong law of large numbers for martingales that

$$(M_n^{(\Delta)}(x))^2 = \mathcal{O}(n^{1-\alpha}\log n) \quad \text{a.s.} \tag{A22}$$

Then, we infer from the conjunction of Equations (A19), (A20) and (A22) that for all $x \in \mathbb{R}$

$$|\Delta_n(x)| = \mathcal{O}(n^{1-\alpha}) \quad \text{a.s.} \tag{A23}$$

The left-hand side of Equation (A17) is much more difficult to handle. We can use once again the assumption that the function f is Lipschitz to deduce that it exists a positive constant C_f such that for all $n \geq 1$

$$|f(\hat{\Phi}_n) - f(\Phi_n)| \leq C_f|\pi_n|, \tag{A24}$$

where $\pi_n = (\hat{\theta}_{n-1} - \theta)'X_n$. Hence, it immediately follows from Equation (A24) that for all $x \in \mathbb{R}$

$$|\Sigma_n(x)| \leq C_f \sum_{k=p+1}^{n} W_k(x)|\pi_k|. \tag{A25}$$

Denote

$$\mathcal{A}_n = \{|\hat{\theta}'_{n-1}X_n - x| \leq Ah_n\} \quad \text{and} \quad \mathcal{B}_n = \{|\theta' X_n - x| \leq Ah_n + b_n\}$$

where (b_n) is a sequence of positive real numbers which will be explicitly given later. On the one hand, we immediately have from the triangle inequality that on the set $\mathcal{A}_n \cap \mathcal{B}_n$,

$$|\pi_n| \leq 2Ah_n + b_n.$$

On the other hand, we also have on the set $\mathcal{A}_n \cap \bar{\mathcal{B}}_n$,

$$Ah_n + b_n < |\theta' X_n - x| \leq |\pi_n| + |\hat{\theta}'_{n-1}X_n - x| \leq |\pi_n| + Ah_n,$$

which implies that $|\pi_n| > b_n$. Consequently, we obtain from Equation (A25) that

$$|\Sigma_n(x)| \leq 2AC_f \sum_{k=p+1}^{n} h_k W_k(x) + C_f \sum_{k=p+1}^{n} b_k W_k(x) + C_f \sum_{k=p+1}^{n} W_k(x)|\pi_k| \mathrm{I}_{\{|\pi_k|>b_k\}}. \tag{A26}$$

We already saw from Equation (A23) that

$$\sum_{k=p+1}^{n} h_k W_k(x) = \mathcal{O}(n^{1-\alpha}) \quad \text{a.s.} \tag{A27}$$

Moreover, it is assumed that the sequence (X_n) has a finite moment of order $a > 2$ which ensures that

$$\sup_{1 \leq k \leq n} \|X_k\| = o(n^{1/a}) \quad \text{a.s.}$$

Consequently, we find from Lemma 2.1 that

$$|\pi_n| = o(b_n) \quad \text{a.s.} \tag{A28}$$

where we can choose

$$b_n = n^{1/a}\sqrt{\frac{\log(\log n)}{n}}.$$

Therefore, we clearly have

$$\sum_{k=p+1}^{n} W_k(x)|\pi_k| \mathrm{I}_{\{|\pi_k|>b_k\}} < +\infty \quad \text{a.s.} \tag{A29}$$

Furthermore, it is not hard to see that

$$\sum_{k=p+1}^{n} b_k = \mathcal{O}(n^{1/a}\sqrt{n \log(\log n)}).$$

Hence, via the same lines as in the proof of Equation (A23), we obtain that

$$\sum_{k=p+1}^{n} b_k W_k(x) = \mathcal{O}(n^{1/a}\sqrt{n \log(\log n)}) \quad \text{a.s.} \tag{A30}$$

Then, we deduce from the conjunction of Equations (A26), (A27), (A29), and (A30) that

$$|\Sigma_n(x)| = \mathcal{O}(n^{1-\alpha}) + \mathcal{O}(n^{1/a}\sqrt{n\log(\log n)}) \quad \text{a.s.} \tag{A31}$$

Consequently, we infer from Equations (A23) and (A31) that for all $x \in \mathbb{R}$

$$Q_n(x) = \mathcal{O}(n^{1-\alpha}) + \mathcal{O}(n^{1/a}\sqrt{n\log(\log n)}) \quad \text{a.s.} \quad \text{a.s.} \tag{A32}$$

Finally, we can conclude from Equation (A3) together with Equations (A14), (A16) and (A32) that

$$\lim_{n\to\infty} \hat{f}_n(x) = f(x) \quad \text{a.s.}$$

with the almost sure rates of convergence given by Equations (12) and (13), which completes the proof of Theorem 3.1.

Appendix 3. Proof of Theorem 3.2

We already saw that $(P_n(x))$ is a square integrable martingale difference sequence with predictable quadratic variation satisfying

$$\lim_{n\to\infty} \frac{\langle P(x)\rangle_n}{n^{1+\alpha}} = \frac{\sigma^2\nu^2}{1+\alpha} h(\theta, x) \quad \text{a.s.}$$

In order to establish the asymptotic normality of Theorem 3.2, it is necessary to prove that the sequence $(P_n(x))$ satisfies the Lindeberg condition, that is for all $\varepsilon > 0$,

$$\mathcal{P}_n(x) = \frac{1}{n^{1+\alpha}} \sum_{k=p+1}^{n} \mathbb{E}[|\Delta P_k(x)|^2 \mathrm{I}_{\{|\Delta P_k(x)| \geq \varepsilon\sqrt{n^{1+\alpha}}\}} | \mathcal{F}_{k-1}] \xrightarrow{\mathcal{P}} 0, \tag{A33}$$

where $\Delta P_n(x) = P_n(x) - P_{n-1}(x)$. We have assumed that the sequence (ε_n) has a finite conditional moment of order $b > 2$ which means that

$$\sup_{n\geq 0} \mathbb{E}[|\varepsilon_n|^b | \mathcal{F}_{n-1}] < +\infty \quad \text{a.s.}$$

Consequently, for all $\varepsilon > 0$, we have

$$\begin{aligned}
\mathcal{P}_n(x) &\leq \frac{1}{\varepsilon^{b-2} n^c} \sum_{k=p+1}^{n} \mathbb{E}[|\Delta P_k(x)|^b | \mathcal{F}_{k-1}], \\
&\leq \frac{1}{\varepsilon^{b-2} n^c} \sum_{k=p+1}^{n} \mathbb{E}[W_k^b(x) | \mathcal{F}_{k-1}] \mathbb{E}[|\varepsilon_k|^b | \mathcal{F}_{k-1}], \\
&\leq \frac{1}{\varepsilon^{b-2} n^c} \sup_{1\leq k\leq n} \mathbb{E}[|\varepsilon_k|^b | \mathcal{F}_{k-1}] \sum_{k=p+1}^{n} \mathbb{E}[W_k^b(x) | \mathcal{F}_{k-1}]
\end{aligned} \tag{A34}$$

where $c = b(1+\alpha)/2$. In addition, via the same lines as in the proof of Equation (A10), we obtain that

$$\lim_{n\to\infty} \frac{1}{n^{1+\alpha(b-1)}} \sum_{k=p+1}^{n} \mathbb{E}[W_k^b(x) \mid \mathcal{F}_{k-1}] = \frac{\xi^b}{1+\alpha(b-1)} h(\theta, x) \quad \text{a.s.} \tag{A35}$$

where

$$\xi^b = \int_{\mathbb{R}} K^b(x)\, \mathrm{d}x.$$

Therefore, we deduce from Equation (A33) together with Equations (A34) and (A35) that, for all $\varepsilon > 0$,

$$\mathcal{P}_n(x) = \mathcal{O}(n^d) \quad \text{a.s.}$$

where $d = (2-b)(1-\alpha)/2$. We recall that $b > 2$ which means that $d < 0$. It ensures that the Lindeberg condition is satisfied. Hence, it follows from the central limit theorem for martingales given e.g. by Corollary 2.1.10 of Duflo [13] that for all $x \in \mathbb{R}$,

$$\frac{P_n(x)}{\sqrt{n^{1+\alpha}}} \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \frac{\sigma^2 \nu^2}{1+\alpha} h(\theta, x)\right). \tag{A36}$$

Furthermore, as soon as $a \geq 6$ and $\frac{1}{3} < \alpha < 1$, we clearly obtain from Equation (A32) that

$$\lim_{n \to \infty} \frac{Q_n(x)}{\sqrt{n^{1+\alpha}}} = 0 \quad \text{a.s.} \tag{A37}$$

Finally, we find from Equation (A3) together with Equations (A14), (A36), (A37) and Slutsky's lemma that, for all $x \in \mathbb{R}$,

$$\sqrt{nh_n}(\hat{f}_n(x) - f(x)) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \frac{\sigma^2 \nu^2}{(1+\alpha) h(\theta, x)}\right),$$

which achieves the proof of Theorem 3.2.