

Vol. 14 (2009), Paper no. 87, pages 2492–2526.

Journal URL

<http://www.math.washington.edu/~ejpecp/>

## Asymptotic analysis for bifurcating autoregressive processes via a martingale approach

Bernard Bercu\*

Benoîte de Saporta<sup>†</sup>

Anne Gégout-Petit<sup>‡</sup>

### Abstract

We study the asymptotic behavior of the least squares estimators of the unknown parameters of general  $p$ th-order bifurcating autoregressive processes. Under very weak assumptions on the driven noise of the process, namely conditional pair-wise independence and suitable moment conditions, we establish the almost sure convergence of our estimators together with the quadratic strong law and the central limit theorem. All our analysis relies on non-standard asymptotic results for martingales.

**Key words:** bifurcating autoregressive process ; tree-indexed times series; martingales ; least squares estimation ; almost sure convergence ; quadratic strong law; central limit theorem.

**AMS 2000 Subject Classification:** Primary 60F15; Secondary: 60F05, 60G42.

Submitted to EJP on September 9, 2008, final version accepted November 2, 2009.

---

\*Université de Bordeaux, IMB, CNRS, UMR 5251, and INRIA Bordeaux, team CQFD, France, [bernard.bercu@math.u-bordeaux1.fr](mailto:bernard.bercu@math.u-bordeaux1.fr)

<sup>†</sup>Université de Bordeaux, GREThA, CNRS, UMR 5113, IMB, CNRS, UMR 5251, and INRIA Bordeaux, team CQFD, France, [saporta@math.u-bordeaux1.fr](mailto:saporta@math.u-bordeaux1.fr)

<sup>‡</sup>Université de Bordeaux, IMB, CNRS, UMR 5251, and INRIA Bordeaux, team CQFD, France, [anne.petit@u-bordeaux2.fr](mailto:anne.petit@u-bordeaux2.fr)

# 1 Introduction

Bifurcating autoregressive (BAR) processes are an adaptation of autoregressive (AR) processes to binary tree structured data. They were first introduced by Cowan and Staudte [2] for cell lineage data, where each individual in one generation gives birth to two offspring in the next generation. Cell lineage data typically consist of observations of some quantitative characteristic of the cells over several generations of descendants from an initial cell. BAR processes take into account both inherited and environmental effects to explain the evolution of the quantitative characteristic under study.

More precisely, the original BAR process is defined as follows. The initial cell is labelled 1, and the two offspring of cell  $n$  are labelled  $2n$  and  $2n + 1$ . Denote by  $X_n$  the quantitative characteristic of individual  $n$ . Then, the first-order BAR process is given, for all  $n \geq 1$ , by

$$\begin{cases} X_{2n} &= a + bX_n + \varepsilon_{2n}, \\ X_{2n+1} &= a + bX_n + \varepsilon_{2n+1}. \end{cases}$$

The noise sequence  $(\varepsilon_{2n}, \varepsilon_{2n+1})$  represents environmental effects while  $a, b$  are unknown real parameters with  $|b| < 1$ . The driven noise  $(\varepsilon_{2n}, \varepsilon_{2n+1})$  was originally supposed to be independent and identically distributed with normal distribution. However, two sister cells being in the same environment early in their lives,  $\varepsilon_{2n}$  and  $\varepsilon_{2n+1}$  are allowed to be correlated, inducing a correlation between sister cells distinct from the correlation inherited from their mother.

Several extensions of the model have been proposed. On the one hand, we refer the reader to Huggins and Basawa [10] and Basawa and Zhou [1; 15] for statistical inference on symmetric bifurcating processes. On the other hand, higher order processes, when not only the effects of the mother but also those of the grand-mother and higher order ancestors are taken into account, have been investigated by Huggins and Basawa [10]. Recently, an asymmetric model has been introduced by Guyon [5; 6] where only the effects of the mother are considered, but sister cells are allowed to have different conditional distributions. We can also mention a recent work of Delmas and Marsalle [3] dealing with a model of asymmetric bifurcating Markov chains on a Galton Watson tree instead of regular binary tree.

The purpose of this paper is to carry out a sharp analysis of the asymptotic properties of the least squares (LS) estimators of the unknown parameters of general asymmetric  $p$ th-order BAR processes. There are several results on statistical inference and asymptotic properties of estimators for BAR models in the literature. For maximum likelihood inference on small independent trees, see Huggins and Basawa [10]. For maximum likelihood inference on a single large tree, see Huggins [9] for the original BAR model, Huggins and Basawa [11] for higher order Gaussian BAR models, and Zhou and Basawa [15] for exponential first-order BAR processes. We also refer the reader to Zhou and Basawa [14] for the LS parameter estimation, and to Hwang, Basawa and Yeo [12] for the local asymptotic normality for BAR processes and related asymptotic inference. In all those papers, the process is supposed to be stationary. Consequently,  $X_n$  has a time-series representation involving an holomorphic function. In Guyon [5], the LS estimator is also investigated, but the process is not stationary, and the author makes intensive use of the tree structure and Markov chain theory. Our goal is to improve and extend the previous results of Guyon [5] via a martingale approach. As previously done by Basawa and Zhou [1; 14; 15] we shall make use of the strong law of large numbers [4] as well as the central limit theorem [7; 8] for martingales. It will allow

us to go further in the analysis of general  $p$ th-order BAR processes. We shall establish the almost sure convergence of the LS estimators together with the quadratic strong law and the central limit theorem.

The paper is organised as follows. Section 2 is devoted to the presentation of the asymmetric  $p$ th-order BAR process under study, while Section 3 deals with the LS estimators of the unknown parameters. In Section 4, we explain our strategy based on martingale theory. Our main results about the asymptotic properties of the LS estimators are given in Section 5. More precisely, we shall establish the almost sure convergence, the quadratic strong law (QSL) and the central limit theorem (CLT) for the LS estimators. The proof of our main results are detailed in Sections 6 to 10, the more technical ones being gathered in the appendices.

## 2 Bifurcating autoregressive processes

In all the sequel, let  $p$  be a non-zero integer. We consider the asymmetric BAR( $p$ ) process given, for all  $n \geq 2^{p-1}$ , by

$$\begin{cases} X_{2n} &= a_0 + \sum_{k=1}^p a_k X_{[\frac{n}{2^{k-1}}]} + \varepsilon_{2n}, \\ X_{2n+1} &= b_0 + \sum_{k=1}^p b_k X_{[\frac{n}{2^{k-1}}]} + \varepsilon_{2n+1}, \end{cases} \quad (2.1)$$

where  $[x]$  stands for the largest integer less than or equal to  $x$ . The initial states  $\{X_k, 1 \leq k \leq 2^{p-1} - 1\}$  are the ancestors while  $(\varepsilon_{2n}, \varepsilon_{2n+1})$  is the driven noise of the process. The parameters  $(a_0, a_1, \dots, a_p)$  and  $(b_0, b_1, \dots, b_p)$  are unknown real numbers. The BAR( $p$ ) process can be rewritten in the abbreviated vector form given, for all  $n \geq 2^{p-1}$ , by

$$\begin{cases} \mathbb{X}_{2n} &= A\mathbb{X}_n + \eta_{2n}, \\ \mathbb{X}_{2n+1} &= B\mathbb{X}_n + \eta_{2n+1}, \end{cases} \quad (2.2)$$

where the regression vector  $\mathbb{X}_n = (X_n, X_{[\frac{n}{2}]}, \dots, X_{[\frac{n}{2^{p-1}}]})^t$ ,  $\eta_{2n} = (a_0 + \varepsilon_{2n})e_1$ ,  $\eta_{2n+1} = (b_0 + \varepsilon_{2n+1})e_1$  with  $e_1 = (1, 0, \dots, 0)^t \in \mathbb{R}^p$ . Moreover,  $A$  and  $B$  are the  $p \times p$  companion matrices

$$A = \begin{pmatrix} a_1 & a_2 & \cdots & a_p \\ 1 & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \vdots \\ 0 & 0 & 1 & 0 \end{pmatrix}, \quad B = \begin{pmatrix} b_1 & b_2 & \cdots & b_p \\ 1 & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \vdots \\ 0 & 0 & 1 & 0 \end{pmatrix}.$$

This process is a direct generalization of the symmetric BAR( $p$ ) process studied by Huggins, Basawa and Zhou [10; 14]. One can also observe that, in the particular case  $p = 1$ , it is the asymmetric BAR process studied by Guyon [5; 6]. In all the sequel, we shall assume that  $\mathbb{E}[X_k^8] < \infty$  for all  $1 \leq k \leq 2^{p-1} - 1$  and that matrices  $A$  and  $B$  satisfy the contracting property

$$\beta = \max\{\|A\|, \|B\|\} < 1,$$

where  $\|A\| = \sup\{\|Au\|, u \in \mathbb{R}^p \text{ with } \|u\| = 1\}$ .

As explained in the introduction, one can see this BAR( $p$ ) process as a  $p$ th-order autoregressive process on a binary tree, where each vertex represents an individual or cell, vertex 1 being the original ancestor, see Figure 1 for an illustration. For all  $n \geq 1$ , denote the  $n$ th generation by

$$\mathbb{G}_n = \{2^n, 2^n + 1, \dots, 2^{n+1} - 1\}.$$

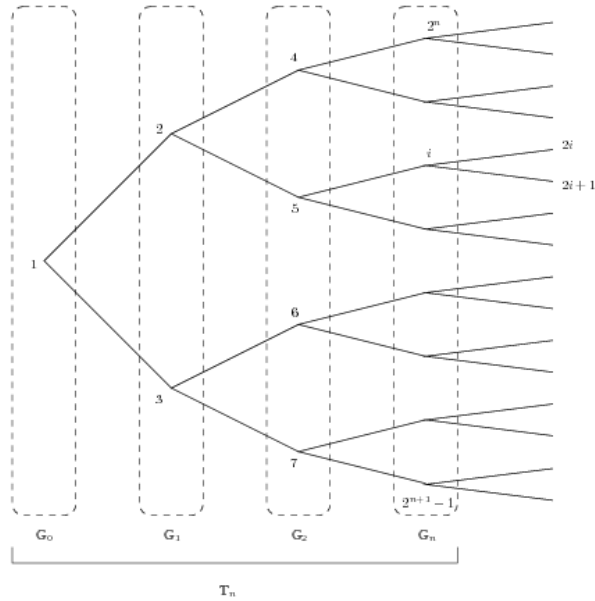


Figure 1: The tree associated with the bifurcating auto-regressive process.

In particular,  $\mathbb{G}_0 = \{1\}$  is the initial generation and  $\mathbb{G}_1 = \{2, 3\}$  is the first generation of offspring from the first ancestor. Let  $\mathbb{G}_{r_n}$  be the generation of individual  $n$ , which means that  $r_n = \log_2(n)$ . Recall that the two offspring of individual  $n$  are labelled  $2n$  and  $2n + 1$ , or conversely, the mother of individual  $n$  is  $\lfloor n/2 \rfloor$ . More generally, the ancestors of individual  $n$  are  $\lfloor n/2 \rfloor, \lfloor n/2^2 \rfloor, \dots, \lfloor n/2^{r_n} \rfloor$ . Furthermore, denote by

$$\mathbb{T}_n = \bigcup_{k=0}^n \mathbb{G}_k$$

the sub-tree of all individuals from the original individual up to the  $n$ th generation. It is clear that the cardinality  $|\mathbb{G}_n|$  of  $\mathbb{G}_n$  is  $2^n$  while that of  $\mathbb{T}_n$  is  $|\mathbb{T}_n| = 2^{n+1} - 1$ . Finally, we denote by  $\mathbb{T}_{n,p} = \{k \in \mathbb{T}_n, k \geq 2^p\}$  the sub-tree of all individuals up to the  $n$ th generation without  $\mathbb{T}_{p-1}$ . One can observe that, for all  $n \geq 1$ ,  $\mathbb{T}_{n,0} = \mathbb{T}_n$  and, for all  $p \geq 1$ ,  $\mathbb{T}_{p,p} = \mathbb{G}_p$ .

### 3 Least-squares estimation

The BAR( $p$ ) process (2.1) can be rewritten, for all  $n \geq 2^{p-1}$ , in the matrix form

$$Z_n = \theta^t Y_n + V_n \tag{3.1}$$

where

$$Z_n = \begin{pmatrix} X_{2n} \\ X_{2n+1} \end{pmatrix}, \quad Y_n = \begin{pmatrix} 1 \\ X_n \end{pmatrix}, \quad V_n = \begin{pmatrix} \varepsilon_{2n} \\ \varepsilon_{2n+1} \end{pmatrix},$$

and the  $(p + 1) \times 2$  matrix parameter  $\theta$  is given by

$$\theta = \begin{pmatrix} a_0 & b_0 \\ a_1 & b_1 \\ \vdots & \vdots \\ a_p & b_p \end{pmatrix}.$$

Our goal is to estimate  $\theta$  from the observation of all individuals up to the  $n$ th generation that is the complete sub-tree  $\mathbb{T}_n$ . Each new generation  $\mathbb{G}_n$  contains half the global available information. Consequently, we shall show that observing the whole tree  $\mathbb{T}_n$  or only generation  $\mathbb{G}_n$  is almost the same. We propose to make use of the standard LS estimator  $\hat{\theta}_n$  which minimizes

$$\Delta_n(\theta) = \frac{1}{2} \sum_{k \in \mathbb{T}_{n-1, p-1}} \|Z_k - \theta^t Y_k\|^2.$$

Consequently, we obviously have for all  $n \geq p$

$$\hat{\theta}_n = S_{n-1}^{-1} \sum_{k \in \mathbb{T}_{n-1, p-1}} Y_k Z_k^t, \quad (3.2)$$

where the  $(p + 1) \times (p + 1)$  matrix  $S_n$  is defined as

$$S_n = \sum_{k \in \mathbb{T}_{n, p-1}} Y_k Y_k^t = \sum_{k \in \mathbb{T}_{n, p-1}} \begin{pmatrix} 1 & \mathbb{X}_k^t \\ \mathbb{X}_k & \mathbb{X}_k \mathbb{X}_k^t \end{pmatrix}.$$

In the special case where  $p = 1$ ,  $S_n$  simply reduces to

$$S_n = \sum_{k \in \mathbb{T}_n} \begin{pmatrix} 1 & X_k \\ X_k & X_k^2 \end{pmatrix}.$$

In order to avoid useless invertibility assumption, we shall assume, without loss of generality, that for all  $n \geq p - 1$ ,  $S_n$  is invertible. Otherwise, we only have to add the identity matrix  $I_{p+1}$  to  $S_n$ . In all what follows, we shall make a slight abuse of notation by identifying  $\theta$  as well as  $\hat{\theta}_n$  to

$$\text{vec}(\theta) = \begin{pmatrix} a_0 \\ \vdots \\ a_p \\ b_0 \\ \vdots \\ b_p \end{pmatrix} \quad \text{and} \quad \text{vec}(\hat{\theta}_n) = \begin{pmatrix} \hat{a}_{0,n} \\ \vdots \\ \hat{a}_{p,n} \\ \hat{b}_{0,n} \\ \vdots \\ \hat{b}_{p,n} \end{pmatrix}.$$

The reason for this change will be explained in Section 4. Hence, we readily deduce from (3.2) that

$$\begin{aligned} \hat{\theta}_n &= (I_2 \otimes S_{n-1}^{-1}) \sum_{k \in \mathbb{T}_{n-1, p-1}} \text{vec}(Y_k Z_k^t) \\ &= (I_2 \otimes S_{n-1}^{-1}) \sum_{k \in \mathbb{T}_{n-1, p-1}} \begin{pmatrix} X_{2k} \\ X_k \mathbb{X}_{2k} \\ X_{2k+1} \\ X_k \mathbb{X}_{2k+1} \end{pmatrix}, \end{aligned}$$

where  $\otimes$  stands for the matrix Kronecker product. Consequently, it follows from (3.1) that

$$\begin{aligned}\widehat{\theta}_n - \theta &= (\mathbf{I}_2 \otimes S_{n-1}^{-1}) \sum_{k \in \mathbb{T}_{n-1, p-1}} \text{vec} (Y_k V_k^t) \\ &= (\mathbf{I}_2 \otimes S_{n-1}^{-1}) \sum_{k \in \mathbb{T}_{n-1, p-1}} \begin{pmatrix} \varepsilon_{2k} \\ \varepsilon_{2k} \mathbb{X}_k \\ \varepsilon_{2k+1} \\ \varepsilon_{2k+1} \mathbb{X}_k \end{pmatrix}.\end{aligned}\quad (3.3)$$

Denote by  $\mathbb{F} = (\mathcal{F}_n)$  the natural filtration associated with the BAR( $p$ ) process, which means that  $\mathcal{F}_n$  is the  $\sigma$ -algebra generated by all individuals up to the  $n$ th generation,  $\mathcal{F}_n = \sigma\{X_k, k \in \mathbb{T}_n\}$ . In all the sequel, we shall make use of the five following moment hypotheses.

**(H.1)** One can find  $\sigma^2 > 0$  such that, for all  $n \geq p - 1$  and for all  $k \in \mathbb{G}_{n+1}$ ,  $\varepsilon_k$  belongs to  $L^2$  with

$$\mathbb{E}[\varepsilon_k | \mathcal{F}_n] = 0 \quad \text{and} \quad \mathbb{E}[\varepsilon_k^2 | \mathcal{F}_n] = \sigma^2 \quad \text{a.s.}$$

**(H.2)** It exists  $|\rho| < \sigma^2$  such that, for all  $n \geq p - 1$  and for all different  $k, l \in \mathbb{G}_{n+1}$  with  $[k/2] = [l/2]$ ,

$$\mathbb{E}[\varepsilon_k \varepsilon_l | \mathcal{F}_n] = \rho \quad \text{a.s.}$$

Otherwise,  $\varepsilon_k$  and  $\varepsilon_l$  are conditionally independent given  $\mathcal{F}_n$ .

**(H.3)** For all  $n \geq p - 1$  and for all  $k \in \mathbb{G}_{n+1}$ ,  $\varepsilon_k$  belongs to  $L^4$  and

$$\sup_{n \geq p-1} \sup_{k \in \mathbb{G}_{n+1}} \mathbb{E}[\varepsilon_k^4 | \mathcal{F}_n] < \infty \quad \text{a.s.}$$

**(H.4)** One can find  $\tau^4 > 0$  such that, for all  $n \geq p - 1$  and for all  $k \in \mathbb{G}_{n+1}$ ,

$$\mathbb{E}[\varepsilon_k^4 | \mathcal{F}_n] = \tau^4 \quad \text{a.s.}$$

and, for  $\nu^2 < \tau^4$  and for all different  $k, l \in \mathbb{G}_{n+1}$  with  $[k/2] = [l/2]$

$$\mathbb{E}[\varepsilon_{2k}^2 \varepsilon_{2k+1}^2 | \mathcal{F}_n] = \nu^2 \quad \text{a.s.}$$

**(H.5)** For all  $n \geq p - 1$  and for all  $k \in \mathbb{G}_{n+1}$ ,  $\varepsilon_k$  belongs to  $L^8$  with

$$\sup_{n \geq p-1} \sup_{k \in \mathbb{G}_{n+1}} \mathbb{E}[\varepsilon_k^8 | \mathcal{F}_n] < \infty \quad \text{a.s.}$$

**Remark 3.1.** In contrast with [14], one can observe that we do not assume that  $(\varepsilon_{2n}, \varepsilon_{2n+1})$  is a sequence of independent and identically distributed bi-variate random vectors. The price to pay for giving up this iid assumption is higher moments, namely assumptions **(H.3)** and **(H.5)**. Indeed we need them to make use of the strong law of large numbers and the central limit theorem for martingales. However, we do not require any normality assumption on  $(\varepsilon_{2n}, \varepsilon_{2n+1})$ . Consequently, our assumptions are much weaker than the existing ones in previous literature.

We now turn to the estimation of the parameters  $\sigma^2$  and  $\rho$ . On the one hand, we propose to estimate the conditional variance  $\sigma^2$  by

$$\widehat{\sigma}_n^2 = \frac{1}{2|\mathbb{T}_{n-1}|} \sum_{k \in \mathbb{T}_{n-1, p-1}} \|\widehat{V}_k\|^2 = \frac{1}{2|\mathbb{T}_{n-1}|} \sum_{k \in \mathbb{T}_{n-1, p-1}} (\widehat{\varepsilon}_{2k}^2 + \widehat{\varepsilon}_{2k+1}^2) \quad (3.4)$$

where for all  $n \geq p - 1$  and for all  $k \in \mathbb{G}_n$ ,  $\widehat{V}_k^t = (\widehat{\varepsilon}_{2k}, \widehat{\varepsilon}_{2k+1})$  with

$$\begin{cases} \widehat{\varepsilon}_{2k} &= X_{2k} - \widehat{a}_{0,n} - \sum_{i=1}^p \widehat{a}_{i,n} X_{\lfloor \frac{k}{2^{i-1}} \rfloor}, \\ \widehat{\varepsilon}_{2k+1} &= X_{2k+1} - \widehat{b}_{0,n} - \sum_{i=1}^p \widehat{b}_{i,n} X_{\lfloor \frac{k}{2^{i-1}} \rfloor}. \end{cases}$$

One can observe that, on the above equations, we make use of only the past observations for the estimation of the parameters. This will be crucial in the asymptotic analysis. On the other hand, we estimate the conditional covariance  $\rho$  by

$$\widehat{\rho}_n = \frac{1}{|\mathbb{T}_{n-1}|} \sum_{k \in \mathbb{T}_{n-1, p-1}} \widehat{\varepsilon}_{2k} \widehat{\varepsilon}_{2k+1}. \quad (3.5)$$

## 4 Martingale approach

In order to establish all the asymptotic properties of our estimators, we shall make use of a martingale approach. It allows us to impose a very smooth restriction on the driven noise  $(\varepsilon_n)$  compared with the previous results in the literature. As a matter of fact, we only assume suitable moment conditions on  $(\varepsilon_n)$  and that  $(\varepsilon_{2n}, \varepsilon_{2n+1})$  are conditionally independent, while it is assumed in [14] that  $(\varepsilon_{2n}, \varepsilon_{2n+1})$  is a sequence of independent identically distributed random vectors. For all  $n \geq p$ , denote

$$M_n = \sum_{k \in \mathbb{T}_{n-1, p-1}} \begin{pmatrix} \varepsilon_{2k} \\ \varepsilon_{2k} \mathbb{X}_k \\ \varepsilon_{2k+1} \\ \varepsilon_{2k+1} \mathbb{X}_k \end{pmatrix} \in \mathbb{R}^{2(p+1)}.$$

Let  $\Sigma_n = I_2 \otimes S_n$ , and note that  $\Sigma_n^{-1} = I_2 \otimes S_n^{-1}$ . For all  $n \geq p$ , we can thus rewrite (3.3) as

$$\widehat{\theta}_n - \theta = \Sigma_{n-1}^{-1} M_n. \quad (4.1)$$

The key point of our approach is that  $(M_n)$  is a martingale. Most of all the asymptotic results for martingales were established for vector-valued martingales. That is the reason why we have chosen to make use of vector notation in Section 3. In order to show that  $(M_n)$  is a martingale adapted to the filtration  $\mathbb{F} = (\mathcal{F}_n)$ , we rewrite it in a compact form. Let  $\Psi_n = I_2 \otimes \Phi_n$ , where  $\Phi_n$  is the rectangular matrix of dimension  $(p+1) \times \delta_n$ , with  $\delta_n = 2^n$ , given by

$$\Phi_n = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ \mathbb{X}_{2^n} & \mathbb{X}_{2^{n+1}} & \cdots & \mathbb{X}_{2^{n+1}-1} \end{pmatrix}.$$

It contains the individuals of generations  $\mathbb{G}_{n-p+1}$  up to  $\mathbb{G}_n$  and is also the collection of all  $Y_k$ ,  $k \in \mathbb{G}_n$ . Let  $\xi_n$  be the random vector of dimension  $\delta_n$

$$\xi_n = \begin{pmatrix} \varepsilon_{2^n} \\ \varepsilon_{2^{n+2}} \\ \vdots \\ \varepsilon_{2^{n+1}-2} \\ \varepsilon_{2^{n+1}} \\ \varepsilon_{2^{n+3}} \\ \vdots \\ \varepsilon_{2^{n+1}-1} \end{pmatrix}.$$

The vector  $\xi_n$  gathers the noise variables of generation  $\mathbb{G}_n$ . The special ordering separating odd and even indices is tailor-made so that  $M_n$  can be written as

$$M_n = \sum_{k=p}^n \Psi_{k-1} \xi_k.$$

By the same token, one can observe that

$$S_n = \sum_{k=p-1}^n \Phi_k \Phi_k^t \quad \text{and} \quad \Sigma_n = \sum_{k=p-1}^n \Psi_k \Psi_k^t.$$

Under **(H.1)** and **(H.2)**, we clearly have for all  $n \geq 0$ ,  $\mathbb{E}[\xi_{n+1} | \mathcal{F}_n] = 0$  and  $\Psi_n$  is  $\mathcal{F}_n$ -measurable. In addition, it is not hard to see that for all  $n \geq 0$ ,  $\mathbb{E}[\xi_{n+1} \xi_{n+1}^t | \mathcal{F}_n] = \Gamma \otimes I_{\delta_n}$  where  $\Gamma$  is the covariance matrix associated with  $(\varepsilon_{2^n}, \varepsilon_{2^{n+1}})$

$$\Gamma = \begin{pmatrix} \sigma^2 & \rho \\ \rho & \sigma^2 \end{pmatrix}.$$

We shall also prove that  $(M_n)$  is a square integrable martingale. Its increasing process is given for all  $n \geq p + 1$  by

$$\langle M \rangle_n = \sum_{k=p-1}^{n-1} \Psi_k (\Gamma \otimes I_{\delta_k}) \Psi_k^t = \Gamma \otimes \sum_{k=p-1}^{n-1} \Phi_k \Phi_k^t = \Gamma \otimes S_{n-1}.$$

It is necessary to establish the convergence of  $S_n$ , properly normalized, in order to prove the asymptotic results for the BAR( $p$ ) estimators  $\hat{\theta}_n$ ,  $\hat{\sigma}_n^2$  and  $\hat{\rho}_n$ . One can observe that the sizes of  $\Psi_n$  and  $\xi_n$  are not fixed and double at each generation. This is why we have to adapt the proof of vector-valued martingale convergence given in [4] to our framework.

## 5 Main results

We now state our main results, first on the martingale  $(M_n)$  and then on our estimators.



**Proposition 5.1.** Assume that  $(\varepsilon_n)$  satisfies (H.1) to (H.3). Then, we have

$$\lim_{n \rightarrow \infty} \frac{S_n}{|\mathbb{T}_n|} = L \quad \text{a.s.} \quad (5.1)$$

where  $L$  is a positive definite matrix specified in Section 7.

This result is the keystone of our asymptotic analysis. It enables us to prove sharp asymptotic properties for  $(M_n)$ .

**Theorem 5.1.** Assume that  $(\varepsilon_n)$  satisfies (H.1) to (H.3). Then, we have

$$M_n^t \Sigma_{n-1}^{-1} M_n = \mathcal{O}(n) \quad \text{a.s.} \quad (5.2)$$

In addition, we also have

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=p}^n M_k^t \Sigma_{k-1}^{-1} M_k = 2(p+1)\sigma^2 \quad \text{a.s.} \quad (5.3)$$

Moreover, if  $(\varepsilon_n)$  satisfies (H.4) and (H.5), we have the central limit theorem

$$\frac{1}{\sqrt{|\mathbb{T}_{n-1}|}} M_n \xrightarrow{\mathcal{L}} \mathcal{N}(0, \Gamma \otimes L). \quad (5.4)$$

From the asymptotic properties of  $(M_n)$ , we deduce the asymptotic behavior of our estimators. Our first result deals with the almost sure asymptotic properties of the LS estimator  $\hat{\theta}_n$ .

**Theorem 5.2.** Assume that  $(\varepsilon_n)$  satisfies (H.1) to (H.3). Then,  $\hat{\theta}_n$  converges almost surely to  $\theta$  with the rate of convergence

$$\|\hat{\theta}_n - \theta\|^2 = \mathcal{O}\left(\frac{\log |\mathbb{T}_{n-1}|}{|\mathbb{T}_{n-1}|}\right) \quad \text{a.s.} \quad (5.5)$$

In addition, we also have the quadratic strong law

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n |\mathbb{T}_{k-1}| (\hat{\theta}_k - \theta)^t \Lambda (\hat{\theta}_k - \theta) = 2(p+1)\sigma^2 \quad \text{a.s.} \quad (5.6)$$

where  $\Lambda = I_2 \otimes L$ .

Our second result is devoted to the almost sure asymptotic properties of the variance and covariance estimators  $\hat{\sigma}_n^2$  and  $\hat{\rho}_n$ . Let

$$\sigma_n^2 = \frac{1}{2|\mathbb{T}_{n-1}|} \sum_{k \in \mathbb{T}_{n-1,p}} (\varepsilon_{2k}^2 + \varepsilon_{2k+1}^2) \quad \text{and} \quad \rho_n = \frac{1}{|\mathbb{T}_{n-1}|} \sum_{k \in \mathbb{T}_{n-1,p}} \varepsilon_{2k} \varepsilon_{2k+1}.$$

**Theorem 5.3.** Assume that  $(\varepsilon_n)$  satisfies (H.1) to (H.3). Then,  $\hat{\sigma}_n^2$  converges almost surely to  $\sigma^2$ . More precisely,

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k \in \mathbb{T}_{n-1,p}} (\hat{\varepsilon}_{2k} - \varepsilon_{2k})^2 + (\hat{\varepsilon}_{2k+1} - \varepsilon_{2k+1})^2 = 2(p+1)\sigma^2 \quad \text{a.s.} \quad (5.7)$$

$$\lim_{n \rightarrow \infty} \frac{|\mathbb{T}_n|}{n} (\widehat{\sigma}_n^2 - \sigma^2) = 2(p+1)\sigma^2 \quad a.s. \quad (5.8)$$

In addition,  $\widehat{\rho}_n$  converges almost surely to  $\rho$

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k \in \mathbb{T}_{n-1,p}} (\widehat{\varepsilon}_{2k} - \varepsilon_{2k})(\widehat{\varepsilon}_{2k+1} - \varepsilon_{2k+1}) = (p+1)\rho \quad a.s. \quad (5.9)$$

$$\lim_{n \rightarrow \infty} \frac{|\mathbb{T}_n|}{n} (\widehat{\rho}_n - \rho) = 2(p+1)\rho \quad a.s. \quad (5.10)$$

Our third result concerns the asymptotic normality for all our estimators  $\widehat{\theta}_n$ ,  $\widehat{\sigma}_n^2$  and  $\widehat{\rho}_n$ .

**Theorem 5.4.** Assume that  $(\varepsilon_n)$  satisfies (H.1) to (H.5). Then, we have the central limit theorem

$$\sqrt{|\mathbb{T}_{n-1}|}(\widehat{\theta}_n - \theta) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \Gamma \otimes L^{-1}). \quad (5.11)$$

In addition, we also have

$$\sqrt{|\mathbb{T}_{n-1}|}(\widehat{\sigma}_n^2 - \sigma^2) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \frac{\tau^4 - 2\sigma^4 + \nu^2}{2}\right) \quad (5.12)$$

and

$$\sqrt{|\mathbb{T}_{n-1}|}(\widehat{\rho}_n - \rho) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \nu^2 - \rho^2). \quad (5.13)$$

The rest of the paper is dedicated to the proof of our main results. We start by giving laws of large numbers for the noise sequence  $(\varepsilon_n)$  in Section 6. In Section 7, we give the proof of Proposition 5.1. Sections 8, 9 and 10 are devoted to the proofs of Theorems 5.2, 5.3 and 5.4, respectively. The more technical proofs, including that of Theorem 5.1, are postponed to the Appendices.

## 6 Laws of large numbers for the noise sequence

We first need to establish strong laws of large numbers for the noise sequence  $(\varepsilon_n)$ . These results will be useful in all the sequel. We will extensively use the strong law of large numbers for locally square integrable real martingales given in Theorem 1.3.15 of [4].

**Lemma 6.1.** Assume that  $(\varepsilon_n)$  satisfies (H.1) and (H.2). Then

$$\lim_{n \rightarrow +\infty} \frac{1}{|\mathbb{T}_n|} \sum_{k \in \mathbb{T}_{n,p}} \varepsilon_k = 0 \quad a.s. \quad (6.1)$$

In addition, if (H.3) holds, we also have

$$\lim_{n \rightarrow +\infty} \frac{1}{|\mathbb{T}_n|} \sum_{k \in \mathbb{T}_{n,p}} \varepsilon_k^2 = \sigma^2 \quad a.s. \quad (6.2)$$

and

$$\lim_{n \rightarrow +\infty} \frac{1}{|\mathbb{T}_{n-1}|} \sum_{k \in \mathbb{T}_{n-1,p-1}} \varepsilon_{2k} \varepsilon_{2k+1} = \rho \quad a.s. \quad (6.3)$$

**Proof:** On the one hand, let

$$P_n = \sum_{k \in \mathbb{T}_{n,p}} \varepsilon_k = \sum_{k=p}^n \sum_{i \in \mathbb{G}_k} \varepsilon_i.$$

We have

$$\Delta P_{n+1} = P_{n+1} - P_n = \sum_{k \in \mathbb{G}_{n+1}} \varepsilon_k.$$

Hence, it follows from **(H.1)** and **(H.2)** that  $(P_n)$  is a square integrable real martingale with increasing process

$$\langle P \rangle_n = (\sigma^2 + \rho) \sum_{k=p}^n |\mathbb{G}_k| = (\sigma^2 + \rho)(|\mathbb{T}_n| - |\mathbb{T}_{p-1}|).$$

Consequently, we deduce from Theorem 1.3.15 of [4] that  $P_n = o(\langle P \rangle_n)$  a.s. which implies (6.1). On the other hand, denote

$$Q_n = \sum_{k=p}^n \frac{1}{|\mathbb{G}_k|} \sum_{i \in \mathbb{G}_k} e_i,$$

where  $e_n = \varepsilon_n^2 - \sigma^2$ . We have

$$\Delta Q_{n+1} = Q_{n+1} - Q_n = \frac{1}{|\mathbb{G}_{n+1}|} \sum_{k \in \mathbb{G}_{n+1}} e_k.$$

First of all, it follows from **(H.1)** that for all  $k \in \mathbb{G}_{n+1}$ ,  $\mathbb{E}[e_k | \mathcal{F}_n] = 0$  a.s. In addition, for all different  $k, l \in \mathbb{G}_{n+1}$  with  $[k/2] \neq [l/2]$ ,

$$\mathbb{E}[e_k e_l | \mathcal{F}_n] = 0 \quad \text{a.s.}$$

thanks to the conditional independence given by **(H.2)**. Furthermore, we readily deduce from **(H.3)** that

$$\sup_{n \geq p-1} \sup_{k \in \mathbb{G}_{n+1}} \mathbb{E}[e_k^2 | \mathcal{F}_n] < \infty \quad \text{a.s.}$$

Therefore,  $(Q_n)$  is a square integrable real martingale with increasing process

$$\begin{aligned} \langle Q \rangle_n &\leq 2 \sup_{p-1 \leq k \leq n-1} \sup_{i \in \mathbb{G}_{k+1}} \mathbb{E}[e_i^2 | \mathcal{F}_k] \sum_{j=p}^n \frac{1}{|\mathbb{G}_j|} \quad \text{a.s.} \\ &\leq 2 \sup_{p-1 \leq k \leq n-1} \sup_{i \in \mathbb{G}_{k+1}} \mathbb{E}[e_i^2 | \mathcal{F}_k] \sum_{j=p}^n \left(\frac{1}{2}\right)^j \quad \text{a.s.} \\ &\leq 2 \sup_{p-1 \leq k \leq n-1} \sup_{i \in \mathbb{G}_{k+1}} \mathbb{E}[e_i^2 | \mathcal{F}_k] < \infty \quad \text{a.s.} \end{aligned}$$

Consequently, we obtain from the strong law of large numbers for martingales that  $(Q_n)$  converges almost surely. Finally, as  $(|\mathbb{G}_n|)$  is a positive real sequence which increases to infinity, we find from Lemma A.1 in Appendix A that

$$\sum_{k=p}^n \sum_{i \in \mathbb{G}_k} e_i = o(|\mathbb{G}_n|) \quad \text{a.s.}$$

leading to

$$\sum_{k=p}^n \sum_{i \in \mathbb{G}_k} e_i = o(|\mathbb{T}_n|) \quad \text{a.s.}$$

as  $|\mathbb{T}_n| - 1 = 2|\mathbb{G}_n|$ , which implies (6.2). We also establish (6.3) in a similar way. As a matter of fact, let

$$R_n = \sum_{k=p}^n \frac{1}{|\mathbb{G}_{k-1}|} \sum_{i \in \mathbb{G}_{k-1}} (\varepsilon_{2i} \varepsilon_{2i+1} - \rho).$$

Then,  $(R_n)$  is a square integrable real martingale which converges almost surely, leading to (6.3).  $\square$

**Remark 6.2.** Note that via Lemma A.2

$$\begin{aligned} \lim_{n \rightarrow +\infty} \frac{1}{|\mathbb{G}_n|} \sum_{k \in \mathbb{G}_n} \varepsilon_{2k} &= 0, & \lim_{n \rightarrow +\infty} \frac{1}{|\mathbb{G}_n|} \sum_{k \in \mathbb{G}_n} \varepsilon_{2k+1} &= 0 & \text{a.s.} \\ \lim_{n \rightarrow +\infty} \frac{1}{|\mathbb{G}_n|} \sum_{k \in \mathbb{G}_n} \varepsilon_{2k}^2 &= \sigma^2, & \lim_{n \rightarrow +\infty} \frac{1}{|\mathbb{G}_n|} \sum_{k \in \mathbb{G}_n} \varepsilon_{2k+1}^2 &= \sigma^2 & \text{a.s.} \end{aligned}$$

In fact, each new generation contains half the global available information, observing the whole tree  $\mathbb{T}_n$  or only generation  $\mathbb{G}_n$  is essentially the same.

For the CLT, we will also need the convergence of higher moments of the driven noise  $(\varepsilon_n)$ .

**Lemma 6.3.** Assume that  $(\varepsilon_n)$  satisfies (H.1) to (H.5). Then, we have

$$\lim_{n \rightarrow +\infty} \frac{1}{|\mathbb{T}_n|} \sum_{k \in \mathbb{T}_{n,p}} \varepsilon_k^4 = \tau^4 \quad \text{a.s.}$$

and

$$\lim_{n \rightarrow +\infty} \frac{1}{|\mathbb{T}_{n-1}|} \sum_{k \in \mathbb{T}_{n-1,p-1}} \varepsilon_{2k}^2 \varepsilon_{2k+1}^2 = \nu^2 \quad \text{a.s.}$$

**Proof :** The proof is left to the reader as it follows essentially the same lines as the proof of Lemma 6.1 using the square integrable real martingales

$$Q_n = \sum_{k=p}^n \frac{1}{|\mathbb{G}_k|} \sum_{i \in \mathbb{G}_k} (\varepsilon_i^4 - \tau^4)$$

and

$$R_n = \sum_{k=p}^n \frac{1}{|\mathbb{G}_{k-1}|} \sum_{i \in \mathbb{G}_{k-1}} (\varepsilon_{2i}^2 \varepsilon_{2i+1}^2 - \nu^2).$$

**Remark 6.4.** Note that again via Lemma A.2

$$\lim_{n \rightarrow +\infty} \frac{1}{|\mathbb{G}_n|} \sum_{k \in \mathbb{G}_n} \varepsilon_{2k}^4 = \tau^4 \quad \text{and} \quad \lim_{n \rightarrow +\infty} \frac{1}{|\mathbb{G}_n|} \sum_{k \in \mathbb{G}_n} \varepsilon_{2k+1}^4 = \tau^4 \quad \text{a.s.}$$

## 7 Proof of Proposition 5.1

Proposition 5.1 is a direct application of the two following lemmas which provide two strong laws of large numbers for the sequence of random vectors  $(\mathbb{X}_n)$ .

**Lemma 7.1.** *Assume that  $(\varepsilon_n)$  satisfies (H.1) and (H.2). Then, we have*

$$\lim_{n \rightarrow +\infty} \frac{1}{|\mathbb{T}_n|} \sum_{k \in \mathbb{T}_{n,p}} \mathbb{X}_k = \lambda = \bar{a}(I_p - \bar{A})^{-1}e_1 \quad a.s. \quad (7.1)$$

where  $\bar{a} = (a_0 + b_0)/2$  and  $\bar{A}$  is the mean of the companion matrices

$$\bar{A} = \frac{1}{2}(A + B).$$

**Lemma 7.2.** *Assume that  $(\varepsilon_n)$  satisfies (H.1) to (H.3). Then, we have*

$$\lim_{n \rightarrow +\infty} \frac{1}{|\mathbb{T}_n|} \sum_{k \in \mathbb{T}_{n,p}} \mathbb{X}_k \mathbb{X}_k^t = \ell, \quad a.s. \quad (7.2)$$

where the matrix  $\ell$  is the unique solution of the equation

$$\ell = T + \frac{1}{2}(A\ell A^t + B\ell B^t)$$

$$T = (\sigma^2 + \bar{a}^2)e_1 e_1^t + \frac{1}{2}(a_0(A\lambda e_1^t + e_1 \lambda^t A^t) + b_0(B\lambda e_1^t + e_1 \lambda^t B^t))$$

with  $\bar{a}^2 = (a_0^2 + b_0^2)/2$ .

**Proof :** The proofs are given in Appendix A. □

**Remark 7.3.** *We shall see in Appendix A that*

$$\ell = \sum_{k=0}^{\infty} \frac{1}{2^k} \sum_{C \in \{A;B\}^k} CTC^t$$

where the notation  $\{A;B\}^k$  means the set of all products of  $A$  and  $B$  with exactly  $k$  terms. For example, we have  $\{A;B\}^0 = \{I_p\}$ ,  $\{A;B\}^1 = \{A, B\}$ ,  $\{A;B\}^2 = \{A^2, AB, BA, B^2\}$  and so on. The cardinality of  $\{A;B\}^k$  is obviously  $2^k$ .

**Remark 7.4.** *One can observe that in the special case  $p = 1$ ,*

$$\begin{aligned} \lim_{n \rightarrow +\infty} \frac{1}{|\mathbb{T}_n|} \sum_{k \in \mathbb{T}_n} X_k &= \frac{\bar{a}}{1 - \bar{b}} \quad a.s. \\ \lim_{n \rightarrow +\infty} \frac{1}{|\mathbb{T}_n|} \sum_{k \in \mathbb{T}_n} X_k^2 &= \frac{\bar{a}^2 + \sigma^2 + 2\lambda \bar{a} \bar{b}}{1 - \bar{b}^2} \quad a.s. \end{aligned}$$

where

$$\bar{a} \bar{b} = \frac{a_0 a_1 + b_0 b_1}{2}, \quad \bar{b} = \frac{a_1 + b_1}{2}, \quad \bar{b}^2 = \frac{a_1^2 + b_1^2}{2}.$$

## 8 Proof of Theorems 5.1 and 5.2

Theorem 5.2 is a consequence of Theorem 5.1. The first result of Theorem 5.1 is a strong law of large numbers for the martingale  $(M_n)$ . We already mentioned that the standard strong law is useless here. This is due to the fact that the dimension of the random vector  $\xi_n$  grows exponentially fast as  $2^n$ . Consequently, we are led to propose a new strong law of large numbers for  $(M_n)$ , adapted to our framework.

**Proof of result (5.2) of Theorem 5.1:** For all  $n \geq p$ , let  $\mathcal{V}_n = M_n^t \Sigma_{n-1}^{-1} M_n$  where we recall that  $\Sigma_n = I_2 \otimes S_n$ , so that  $\Sigma_n^{-1} = I_2 \otimes S_n^{-1}$ . First of all, we have

$$\begin{aligned} \mathcal{V}_{n+1} &= M_{n+1}^t \Sigma_n^{-1} M_{n+1} = (M_n + \Delta M_{n+1})^t \Sigma_n^{-1} (M_n + \Delta M_{n+1}), \\ &= M_n^t \Sigma_n^{-1} M_n + 2M_n^t \Sigma_n^{-1} \Delta M_{n+1} + \Delta M_{n+1}^t \Sigma_n^{-1} \Delta M_{n+1}, \\ &= \mathcal{V}_n - M_n^t (\Sigma_{n-1}^{-1} - \Sigma_n^{-1}) M_n + 2M_n^t \Sigma_n^{-1} \Delta M_{n+1} + \Delta M_{n+1}^t \Sigma_n^{-1} \Delta M_{n+1}. \end{aligned}$$

By summing over this identity, we obtain the main decomposition

$$\mathcal{V}_{n+1} + \mathcal{A}_n = \mathcal{V}_p + \mathcal{B}_{n+1} + \mathcal{W}_{n+1}, \quad (8.1)$$

where

$$\begin{aligned} \mathcal{A}_n &= \sum_{k=p}^n M_k^t (\Sigma_{k-1}^{-1} - \Sigma_k^{-1}) M_k, \\ \mathcal{B}_{n+1} &= 2 \sum_{k=p}^n M_k^t \Sigma_k^{-1} \Delta M_{k+1} \quad \text{and} \quad \mathcal{W}_{n+1} = \sum_{k=p}^n \Delta M_{k+1}^t \Sigma_k^{-1} \Delta M_{k+1}. \end{aligned}$$

The asymptotic behavior of the left-hand side of (8.1) is as follows.

**Lemma 8.1.** *Assume that  $(\varepsilon_n)$  satisfies (H.1) to (H.3). Then, we have*

$$\lim_{n \rightarrow +\infty} \frac{\mathcal{V}_{n+1} + \mathcal{A}_n}{n} = (p+1)\sigma^2 \quad \text{a.s.} \quad (8.2)$$

**Proof:** The proof is given in Appendix B. It relies on the Riccations equation associated to  $(S_n)$  and the strong law of large numbers for  $(\mathcal{W}_n)$ .  $\square$

Since  $(\mathcal{V}_n)$  and  $(\mathcal{A}_n)$  are two sequences of positive real numbers, we infer from Lemma 8.1 that  $\mathcal{V}_{n+1} = \mathcal{O}(n)$  a.s. which ends the proof of (5.2).  $\square$

**Proof of result (5.5) of Theorem 5.2:** It clearly follows from (4.1) that

$$\mathcal{V}_n = (\widehat{\theta}_n - \theta)^t \Sigma_{n-1} (\widehat{\theta}_n - \theta).$$

Consequently, the asymptotic behavior of  $\widehat{\theta}_n - \theta$  is clearly related to the one of  $\mathcal{V}_n$ . More precisely, we can deduce from convergence (5.1) that

$$\lim_{n \rightarrow \infty} \frac{\lambda_{\min}(\Sigma_n)}{|\mathbb{T}_n|} = \lambda_{\min}(\Lambda) > 0 \quad \text{a.s.}$$

since  $L$  as well as  $\Lambda = I_2 \otimes L$  are definite positive matrices. Here  $\lambda_{\min}(\Lambda)$  stands for the smallest eigenvalue of the matrix  $\Lambda$ . Therefore, as

$$\|\hat{\theta}_n - \theta\|^2 \leq \frac{\mathcal{V}_n}{\lambda_{\min}(\Sigma_{n-1})},$$

we use (5.2) to conclude that

$$\|\hat{\theta}_n - \theta\|^2 = \mathcal{O}\left(\frac{n}{|\mathbb{T}_{n-1}|}\right) = \mathcal{O}\left(\frac{\log |\mathbb{T}_{n-1}|}{|\mathbb{T}_{n-1}|}\right) \quad \text{a.s.}$$

which completes the proof of (5.5).  $\square$

We now turn to the proof of the quadratic strong law. To this end, we need a sharper estimate of the asymptotic behavior of  $(\mathcal{V}_n)$ .

**Lemma 8.2.** *Assume that  $(\varepsilon_n)$  satisfies (H.1) to (H.3). Then, we have for all  $\delta > 1/2$ ,*

$$\|M_n\|^2 = o(|\mathbb{T}_{n-1}|n^\delta) \quad \text{a.s.} \quad (8.3)$$

**Proof:** The proof is given in Appendix C.  $\square$

A direct application of Lemma 8.2 ensures that  $\mathcal{V}_n = o(n^\delta)$  a.s. for all  $\delta > 1/2$ . Hence, Lemma 8.1 immediately leads to the following result.

**Corollary 8.3.** *Assume that  $(\varepsilon_n)$  satisfies (H.1) to (H.3). Then, we have*

$$\lim_{n \rightarrow +\infty} \frac{\mathcal{A}_n}{n} = (p+1)\sigma^2 \quad \text{a.s.} \quad (8.4)$$

**Proof of result (5.3) of Theorem 5.1:** First of all,  $\mathcal{A}_n$  may be rewritten as

$$\mathcal{A}_n = \sum_{k=p}^n M_k^t (\Sigma_{k-1}^{-1} - \Sigma_k^{-1}) M_k = \sum_{k=p}^n M_k^t \Sigma_{k-1}^{-1/2} \Delta_k \Sigma_{k-1}^{-1/2} M_k$$

where  $\Delta_n = I_{2(p+1)} - \Sigma_{n-1}^{1/2} \Sigma_n^{-1} \Sigma_{n-1}^{1/2}$ . In addition, via Proposition 5.1

$$\lim_{n \rightarrow \infty} \frac{\Sigma_n}{|\mathbb{T}_n|} = \Lambda \quad \text{a.s.} \quad (8.5)$$

which implies that

$$\lim_{n \rightarrow \infty} \Delta_n = \frac{1}{2} I_{2(p+1)} \quad \text{a.s.} \quad (8.6)$$

Furthermore, it follows from Corollary 8.3 that  $\mathcal{A}_n = \mathcal{O}(n)$  a.s. Hence, we deduce from (8.5) and (8.6) that

$$\frac{\mathcal{A}_n}{n} = \left( \frac{1}{2n} \sum_{k=p}^n M_k^t \Sigma_{k-1}^{-1} M_k \right) + o(1) \quad \text{a.s.} \quad (8.7)$$

and convergence (5.3) directly follows from Corollary 8.3.  $\square$

We are now in position to prove the QSL.

**Proof of result (5.6) of Theorem 5.2:** The QSL is a direct consequence of (5.3) together with the fact that  $\widehat{\theta}_n - \theta = \Sigma_{n-1}^{-1} M_n$ . Indeed, we have

$$\begin{aligned} \frac{1}{n} \sum_{k=p}^n M_k^t \Sigma_{k-1}^{-1} M_k &= \frac{1}{n} \sum_{k=p}^n (\widehat{\theta}_k - \theta)^t \Sigma_{k-1} (\widehat{\theta}_k - \theta) \\ &= \frac{1}{n} \sum_{k=p}^n |\mathbb{T}_{k-1}| (\widehat{\theta}_k - \theta)^t \frac{\Sigma_{k-1}}{|\mathbb{T}_{k-1}|} (\widehat{\theta}_k - \theta) \\ &= \frac{1}{n} \sum_{k=p}^n |\mathbb{T}_{k-1}| (\widehat{\theta}_k - \theta)^t \Lambda (\widehat{\theta}_k - \theta) + o(1) \quad \text{a.s.} \end{aligned}$$

which completes the proof of Theorem 5.2.  $\square$

## 9 Proof of Theorem 5.3

The almost sure convergence of  $\widehat{\sigma}_n^2$  and  $\widehat{\rho}_n$  is strongly related to that of  $\widehat{V}_n - V_n$ .

**Proof of result (5.7) of Theorem 5.3:** We need to prove that

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k \in \mathbb{T}_{n-1, p-1}} \|\widehat{V}_k - V_k\|^2 = 2(p+1)\sigma^2 \quad \text{a.s.} \quad (9.1)$$

Once again, we are searching for a link between the sum of  $\|\widehat{V}_n - V_n\|$  and the processes  $(\mathcal{A}_n)$  and  $(\mathcal{V}_n)$  whose convergence properties were previously investigated. For all  $n \geq p$ , we have

$$\begin{aligned} \sum_{k \in \mathbb{G}_n} \|\widehat{V}_k - V_k\|^2 &= \sum_{k \in \mathbb{G}_n} (\widehat{\varepsilon}_{2k} - \varepsilon_{2k})^2 + (\widehat{\varepsilon}_{2k+1} - \varepsilon_{2k+1})^2, \\ &= (\widehat{\theta}_n - \theta)^t \Psi_n \Psi_n^t (\widehat{\theta}_n - \theta), \\ &= M_n^t \Sigma_{n-1}^{-1} \Psi_n \Psi_n^t \Sigma_{n-1}^{-1} M_n, \\ &= M_n^t \Sigma_{n-1}^{-1/2} \Delta_n \Sigma_{n-1}^{-1/2} M_n, \end{aligned}$$

where

$$\Delta_n = \Sigma_{n-1}^{-1/2} \Psi_n \Psi_n^t \Sigma_{n-1}^{-1/2} = \Sigma_{n-1}^{-1/2} (\Sigma_n - \Sigma_{n-1}) \Sigma_{n-1}^{-1/2}.$$

Now, we can deduce from convergence (8.5) that

$$\lim_{n \rightarrow \infty} \Delta_n = I_{2(p+1)} \quad \text{a.s.}$$



which implies that

$$\sum_{k \in \mathbb{G}_n} \|\widehat{V}_k - V_k\|^2 = M_n^t \Sigma_{n-1}^{-1} M_n (1 + o(1)) \quad \text{a.s.}$$

Therefore, we can conclude via convergence (5.3) that

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k \in \mathbb{T}_{n-1, p-1}} \|\widehat{V}_k - V_k\|^2 = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=p}^n M_k^t \Sigma_{k-1}^{-1} M_k = 2(p+1)\sigma^2 \quad \text{a.s.}$$

**Proof of result (5.8) of Theorem 5.3:** First of all,

$$\begin{aligned} \widehat{\sigma}_n^2 - \sigma_n^2 &= \frac{1}{2|\mathbb{T}_{n-1}|} \sum_{k \in \mathbb{T}_{n-1, p-1}} (\|\widehat{V}_k\|^2 - \|V_k\|^2), \\ &= \frac{1}{2|\mathbb{T}_{n-1}|} \sum_{k \in \mathbb{T}_{n-1, p-1}} (\|\widehat{V}_k - V_k\|^2 + 2(\widehat{V}_k - V_k)^t V_k). \end{aligned}$$

Set

$$P_n = \sum_{k \in \mathbb{T}_{n-1, p-1}} (\widehat{V}_k - V_k)^t V_k = \sum_{k=p}^n \sum_{i \in \mathbb{G}_{k-1}} (\widehat{V}_i - V_i)^t V_i.$$

We clearly have

$$\Delta P_{n+1} = P_{n+1} - P_n = \sum_{k \in \mathbb{G}_n} (\widehat{V}_k - V_k)^t V_k.$$

One can observe that for all  $k \in \mathbb{G}_n$ ,  $\widehat{V}_k - V_k = (I_2 \otimes Y_k)^t (\theta - \widehat{\theta}_n)$  which implies that  $\widehat{V}_k - V_k$  is  $\mathcal{F}_n$ -measurable. Consequently,  $(P_n)$  is a real martingale transform. Hence, we can deduce from the strong law of large numbers for martingale transforms given in Theorem 1.3.24 of [4] together with (9.1) that

$$P_n = o\left(\sum_{k \in \mathbb{T}_{n-1, p-1}} \|\widehat{V}_k - V_k\|^2\right) = o(n) \quad \text{a.s.}$$

It ensures once again via convergence (9.1) that

$$\lim_{n \rightarrow \infty} \frac{|\mathbb{T}_n|}{n} (\widehat{\sigma}_n^2 - \sigma_n^2) = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k \in \mathbb{T}_{n-1, p-1}} \|\widehat{V}_k - V_k\|^2 = 2(p+1)\sigma^2 \quad \text{a.s.}$$

We now turn to the study of the covariance estimator  $\widehat{\rho}_n$ . We have

$$\begin{aligned} \widehat{\rho}_n - \rho_n &= \frac{1}{|\mathbb{T}_{n-1}|} \sum_{k \in \mathbb{T}_{n-1, p-1}} (\widehat{\varepsilon}_{2k} \widehat{\varepsilon}_{2k+1} - \varepsilon_{2k} \varepsilon_{2k+1}), \\ &= \frac{1}{|\mathbb{T}_{n-1}|} \sum_{k \in \mathbb{T}_{n-1, p-1}} (\widehat{\varepsilon}_{2k} - \varepsilon_{2k})(\widehat{\varepsilon}_{2k+1} - \varepsilon_{2k+1}) + \frac{1}{|\mathbb{T}_{n-1}|} Q_n, \end{aligned}$$

where

$$Q_n = \sum_{k \in \mathbb{T}_{n-1, p-1}} (\widehat{\varepsilon}_{2k} - \varepsilon_{2k}) \varepsilon_{2k+1} + (\widehat{\varepsilon}_{2k+1} - \varepsilon_{2k+1}) \varepsilon_{2k} = \sum_{k \in \mathbb{T}_{n-1, p-1}} (\widehat{V}_k - V_k)^t J_2 V_k$$

with

$$J_2 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

Moreover, one can observe that  $J_2 \Gamma J_2 = \Gamma$ . Hence, as before,  $(Q_n)$  is a real martingale transform satisfying

$$Q_n = o \left( \sum_{k \in \mathbb{T}_{n-1, p-1}} \|\widehat{V}_k - V_k\|^2 \right) = o(n) \quad \text{a.s.}$$

We will see in Appendix D that

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k \in \mathbb{T}_{n-1, p-1}} (\widehat{\varepsilon}_{2k} - \varepsilon_{2k})(\widehat{\varepsilon}_{2k+1} - \varepsilon_{2k+1}) = (p+1)\rho \quad \text{a.s.} \quad (9.2)$$

Finally, we find from (9.2) that

$$\lim_{n \rightarrow \infty} \frac{|\mathbb{T}_n|}{n} (\widehat{\rho}_n - \rho_n) = 2(p+1)\rho \quad \text{a.s.}$$

which completes the proof of Theorem 5.3.  $\square$

## 10 Proof of Theorem 5.4

In order to prove the CLT for the BAR( $p$ ) estimators, we will use the central limit theorem for martingale difference sequences given in Propositions 7.8 and 7.9 of Hamilton [8].

**Proposition 10.1.** *Assume that  $(W_n)$  is a vector martingale difference sequence satisfying*

(a) *For all  $n \geq 1$ ,  $\mathbb{E}[W_n W_n^t] = \Omega_n$  where  $\Omega_n$  is a positive definite matrix and*

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n \Omega_k = \Omega$$

*where  $\Omega$  is also a positive definite matrix.*

(b) *For all  $n \geq 1$  and for all  $i, j, k, l$ ,  $\mathbb{E}[W_{in} W_{jn} W_{kn} W_{ln}] < \infty$  where  $W_{in}$  is the  $i$ th element of the vector  $W_n$ .*

(c)

$$\frac{1}{n} \sum_{k=1}^n W_k W_k^t \xrightarrow{\mathcal{P}} \Omega.$$

Then, we have the central limit theorem

$$\frac{1}{\sqrt{n}} \sum_{k=1}^n W_k \xrightarrow{\mathcal{L}} \mathcal{N}(0, \Omega).$$

We wish to point out that for  $\text{BAR}(p)$  processes, it seems impossible to make use of the standard CLT for martingales. This is due to the fact that Lindeberg's condition is not satisfied in our framework. Moreover, as the size of  $(\xi_n)$  doubles at each generation, it is also impossible to check condition (c). To overcome this problem, we simply change the filtration. Instead of using the generation-wise filtration, we will use the sister pair-wise one. Let

$$\mathcal{G}_n = \sigma\{X_1, (X_{2k}, X_{2k+1}), 1 \leq k \leq n\}$$

be the  $\sigma$ -algebra generated by all pairs of individuals up to the offspring of individual  $n$ . Hence  $(\varepsilon_{2n}, \varepsilon_{2n+1})$  is  $\mathcal{G}_n$ -measurable. Note that  $\mathcal{G}_n$  is also the  $\sigma$ -algebra generated by, on the one hand, all the past generations up to that of individual  $n$ , i.e. the  $r_n$ th generation, and, on the other hand, all pairs of the  $(r_n + 1)$ th generation with ancestors less than or equal to  $n$ . In short,

$$\mathcal{G}_n = \sigma\left(\mathcal{F}_{r_n} \cup \{(X_{2k}, X_{2k+1}), k \in \mathbb{G}_{r_n}, k \leq n\}\right).$$

Therefore, **(H.2)** implies that the processes  $(\varepsilon_{2n}, \mathbb{X}_n \varepsilon_{2n}, \varepsilon_{2n+1}, \mathbb{X}_n \varepsilon_{2n+1})^t$ ,  $(\varepsilon_{2n}^2 + \varepsilon_{2n+1}^2 - 2\sigma^2)$  and  $(\varepsilon_{2n} \varepsilon_{2n+1} - \rho)$  are  $\mathcal{G}_n$ -martingales.

**Proof of result (5.4) of Theorem 5.1:** First, recall that  $Y_n = (1, \mathbb{X}_n)^t$ . We apply Propositions 10.1 to the  $\mathcal{G}_n$ -martingale difference sequence  $(D_n)$  given by

$$D_n = \text{vec}(Y_n V_n^t) = \begin{pmatrix} \varepsilon_{2n} \\ \mathbb{X}_n \varepsilon_{2n} \\ \varepsilon_{2n+1} \\ \mathbb{X}_n \varepsilon_{2n+1} \end{pmatrix}.$$

We clearly have

$$D_n D_n^t = \begin{pmatrix} \varepsilon_{2n}^2 & \varepsilon_{2n} \varepsilon_{2n+1} \\ \varepsilon_{2n+1} \varepsilon_{2n} & \varepsilon_{2n+1}^2 \end{pmatrix} \otimes Y_n Y_n^t.$$

Hence, it follows from **(H.1)** and **(H.2)** that

$$\mathbb{E}[D_n D_n^t] = \Gamma \otimes \mathbb{E}[Y_n Y_n^t].$$

Moreover, we can show by a slight change in the proof of Lemmas 7.1 and 7.2 that

$$\lim_{n \rightarrow \infty} \frac{1}{|\mathbb{T}_n|} \sum_{k \in \mathbb{T}_{n-1, p-1}} \mathbb{E}[D_k D_k^t] = \Gamma \otimes \lim_{n \rightarrow \infty} \frac{1}{|\mathbb{T}_n|} \mathbb{E}[S_n] = \Gamma \otimes L,$$

which is positive definite, so that condition (a) holds. Condition (b) also clearly holds under **(H.3)**. We now turn to condition (c). We have

$$\sum_{k \in \mathbb{T}_{n-1, p-1}} D_k D_k^t = \Gamma \otimes S_n + R_n$$

where

$$R_n = \sum_{k \in \mathbb{T}_{n-1, p-1}} \begin{pmatrix} \varepsilon_{2k}^2 - \sigma^2 & \varepsilon_{2k} \varepsilon_{2k+1} - \rho \\ \varepsilon_{2k+1} \varepsilon_{2k} - \rho & \varepsilon_{2k+1}^2 - \sigma^2 \end{pmatrix} \otimes Y_k Y_k^t.$$

Under (H.1) to (H.5), we can show that  $(R_n)$  is a martingale transform. Moreover, we can prove that  $R_n = o(n)$  a.s. using Lemma A.6 and similar calculations as in Appendix B where a more complicated martingale transform  $(K_n)$  is studied. Consequently, condition (c) also holds and we can conclude that

$$\frac{1}{\sqrt{|\mathbb{T}_{n-1}|}} \sum_{k \in \mathbb{T}_{n-1, p-1}} D_k = \frac{1}{\sqrt{|\mathbb{T}_{n-1}|}} M_n \xrightarrow{\mathcal{L}} \mathcal{N}(0, \Gamma \otimes L). \quad (10.1)$$

**Proof of result (5.11) of Theorem 5.4:** We deduce from (4.1) that

$$\sqrt{|\mathbb{T}_{n-1}|}(\hat{\theta}_n - \theta) = |\mathbb{T}_{n-1}| \Sigma_{n-1}^{-1} \frac{M_n}{\sqrt{|\mathbb{T}_{n-1}|}}.$$

Hence, (5.11) directly follows from (5.4) and convergence (8.5) together with Slutsky's Lemma.  $\square$

**Proof of results (5.12) and (5.13) of Theorem 5.4:** On the one hand, we apply Propositions 10.1 to the  $\mathcal{G}_n$ -martingale difference sequence  $(v_n)$  defined by

$$v_n = \varepsilon_{2n}^2 + \varepsilon_{2n+1}^2 - 2\sigma^2.$$

Under (H.4), one has  $\mathbb{E}[v_n^2] = 2\tau^4 - 4\sigma^4 + 2\nu^2$  which ensures that

$$\lim_{n \rightarrow \infty} \frac{1}{|\mathbb{T}_n|} \sum_{k \in \mathbb{T}_{n, p-1}} \mathbb{E}[v_k^2] = 2\tau^4 - 4\sigma^4 + 2\nu^2 > 0.$$

Hence, condition (a) holds. Once again, condition (b) clearly holds under (H.5), and Lemma 6.3 together with Remark 6.4 imply condition (c),

$$\lim_{n \rightarrow \infty} \frac{1}{|\mathbb{T}_n|} \sum_{k \in \mathbb{T}_{n, p-1}} v_k^2 = 2\tau^4 - 4\sigma^4 + 2\nu^2 \quad \text{a.s.}$$

Therefore, we obtain that

$$\frac{1}{\sqrt{|\mathbb{T}_{n-1}|}} \sum_{k \in \mathbb{T}_{n-1, p-1}} v_k = 2\sqrt{|\mathbb{T}_{n-1}|}(\sigma_n^2 - \sigma^2) \xrightarrow{\mathcal{L}} \mathcal{N}(0, 2\tau^4 - 4\sigma^4 + 2\nu^2). \quad (10.2)$$

Furthermore, we infer from (5.8) that

$$\lim_{n \rightarrow \infty} \sqrt{|\mathbb{T}_{n-1}|}(\hat{\sigma}_n^2 - \sigma_n^2) = 0 \quad \text{a.s.} \quad (10.3)$$

Finally, (10.2) and (10.3) imply (5.12). On the other hand, we apply again Proposition 10.1 to the  $\mathcal{G}_n$ -martingale difference sequence  $(w_n)$  given by

$$w_n = \varepsilon_{2n}\varepsilon_{2n+1} - \rho.$$

Under (H.4), one has  $\mathbb{E}[w_n^2] = \nu^2 - \rho^2$  which implies that condition (a) holds since

$$\lim_{n \rightarrow \infty} \frac{1}{|\mathbb{T}_n|} \sum_{k \in \mathbb{T}_{n, p-1}} \mathbb{E}[w_k^2] = \nu^2 - \rho^2 > 0.$$

Once again, condition (b) clearly holds under (H.5), and Lemmas 6.1 and 6.3 yield condition (c),

$$\lim_{n \rightarrow \infty} \frac{1}{|\mathbb{T}_n|} \sum_{k \in \mathbb{T}_{n,p-1}} w_k^2 = v^2 - \rho^2 \quad \text{a.s.}$$

Consequently, we obtain that

$$\frac{1}{\sqrt{|\mathbb{T}_{n-1}|}} \sum_{k \in \mathbb{T}_{n-1,p-1}} w_k = \sqrt{|\mathbb{T}_{n-1}|} (\rho_n - \rho) \xrightarrow{\mathcal{L}} \mathcal{N}(0, v^2 - \rho^2). \quad (10.4)$$

Furthermore, we infer from (5.10) that

$$\lim_{n \rightarrow \infty} \sqrt{|\mathbb{T}_{n-1}|} (\hat{\rho}_n - \rho_n) = 0 \quad \text{a.s.} \quad (10.5)$$

Finally, (5.13) follows from (10.4) and (10.5) which completes the proof of Theorem 5.4.  $\square$

## Appendices

### A Laws of large numbers for the BAR process

We start with some technical Lemmas we make repeatedly use of, the well-known Kronecker's Lemma given in Lemma 1.3.14 of [4] together with some related results.

**Lemma A.1.** *Let  $(\alpha_n)$  be a sequence of positive real numbers increasing to infinity. In addition, let  $(x_n)$  be a sequence of real numbers such that*

$$\sum_{n=0}^{\infty} \frac{|x_n|}{\alpha_n} < +\infty.$$

Then, one has

$$\lim_{n \rightarrow \infty} \frac{1}{\alpha_n} \sum_{k=0}^n x_k = 0.$$

**Lemma A.2.** *Let  $(x_n)$  be a sequence of real numbers. Then,*

$$\lim_{n \rightarrow \infty} \frac{1}{|\mathbb{T}_n|} \sum_{k \in \mathbb{T}_n} x_k = x \iff \lim_{n \rightarrow \infty} \frac{1}{|\mathbb{G}_n|} \sum_{k \in \mathbb{G}_n} x_k = x. \quad (\text{A.1})$$

**Proof:** First of all, recall that  $|\mathbb{T}_n| = 2^{n+1} - 1$  and  $|\mathbb{G}_n| = 2^n$ . Assume that

$$\lim_{n \rightarrow \infty} \frac{1}{|\mathbb{T}_n|} \sum_{k \in \mathbb{T}_n} x_k = x.$$

We have the decomposition,

$$\sum_{k \in \mathbb{T}_n} x_k = \sum_{k \in \mathbb{T}_{n-1}} x_k + \sum_{k \in \mathbb{G}_n} x_k.$$

Consequently,

$$\begin{aligned}\lim_{n \rightarrow \infty} \frac{1}{|\mathbb{G}_n|} \sum_{k \in \mathbb{G}_n} x_k &= \lim_{n \rightarrow \infty} \frac{2}{|\mathbb{T}_n| + 1} \sum_{k \in \mathbb{T}_n} x_k - \lim_{n \rightarrow \infty} \frac{1}{|\mathbb{T}_{n-1}| + 1} \sum_{k \in \mathbb{T}_{n-1}} x_k, \\ &= 2x - x = x.\end{aligned}$$

Conversely, suppose that

$$\lim_{n \rightarrow \infty} \frac{1}{|\mathbb{G}_n|} \sum_{k \in \mathbb{G}_n} x_k = x.$$

A direct application of Toeplitz Lemma given in Lemma 2.2.13 of [4]) yields

$$\begin{aligned}\lim_{n \rightarrow \infty} \frac{1}{|\mathbb{T}_n|} \sum_{k \in \mathbb{T}_n} x_k &= \lim_{n \rightarrow \infty} \frac{1}{|\mathbb{T}_n|} \sum_{k=0}^n \sum_{i \in \mathbb{G}_k} x_i, \\ &= \lim_{n \rightarrow \infty} \frac{1}{|\mathbb{T}_n|} \sum_{k=0}^n 2^k \frac{1}{|\mathbb{G}_k|} \sum_{i \in \mathbb{G}_k} x_i = x.\end{aligned}$$

**Lemma A.3.** Let  $(A_n)$  be a sequence of real-valued matrices such that  $\sum_{n=0}^{\infty} \|A_n\| < \infty$  and

$$\lim_{n \rightarrow \infty} \sum_{k=0}^n A_k = A.$$

In addition, let  $(X_n)$  be a sequence of real-valued vectors which converges to a limiting value  $X$ . Then,

$$\lim_{n \rightarrow \infty} \sum_{k=0}^n A_{n-k} X_k = AX. \quad (\text{A.2})$$

**Proof:** For all  $n \geq 0$ , let

$$U_n = \sum_{k=0}^n A_{n-k} X_k.$$

We clearly have for all integer  $n_0$  with  $1 \leq n_0 < n$ ,

$$\begin{aligned}\|U_n - AX\| &= \left\| \sum_{k=0}^n A_{n-k} X_k - \sum_{k=0}^n A_k X - \sum_{k=n+1}^{\infty} A_k X \right\|, \\ &\leq \sum_{k=0}^{n_0} \|A_{n-k}\| \|X_k - X\| + \sum_{k=n+1_0}^{\infty} \|A_k\| \|X\|, \\ &\leq \sum_{k=0}^{n_0} \|A_{n-k}\| \|X_k - X\| + \sum_{k=n_0+1}^n \|A_{n-k}\| \|X_k - X\| + \sum_{k=n+1}^{\infty} \|A_k\| \|X\|.\end{aligned}$$

We assume that  $(X_n)$  converges to a limiting value  $X$ . Consequently, we can choose  $n_0$  such that for all  $k > n_0$ ,  $\|X_k - X\| < \varepsilon$ . Moreover, one can find  $M > 0$  such that for all  $k \geq 0$ ,  $\|X_k - X\| \leq M$  and  $\|X\| \leq M$ . Therefore, we obtain that

$$\|U_n - AX\| \leq (n_0 + 1)M \sup_{k \geq n-n_0} \|A_k\| + \varepsilon \sum_{k=n_0+1}^n \|A_{n-k}\| + M \sum_{k=n+1}^{\infty} \|A_k\|.$$

On the one hand

$$\sup_{k \geq n-n_0} \|A_k\| \quad \text{and} \quad \sum_{k=n+1}^{\infty} \|A_k\|$$

both converge to 0 as  $n$  tends to infinity. On the other hand,

$$\sum_{k=n_0+1}^n \|A_{n-k}\| \leq \sum_{n=0}^{\infty} \|A_n\| < \infty.$$

Consequently,  $\|U_n - AX\|$  goes to 0 as  $n$  goes to infinity, as expected.  $\square$

**Lemma A.4.** *Let  $(T_n)$  be a convergent sequence of real-valued matrices with limiting value  $T$ . Then,*

$$\lim_{n \rightarrow \infty} \sum_{k=0}^n \frac{1}{2^k} \sum_{C \in \{A;B\}^k} C T_{n-k} C^t = \ell$$

where the matrix

$$\ell = \sum_{k=0}^{\infty} \frac{1}{2^k} \sum_{C \in \{A;B\}^k} C T C^t$$

is the unique solution of the equation

$$\ell = T + \frac{1}{2}(A\ell A^t + B\ell B^t). \quad (\text{A.3})$$

**Proof:** First of all, recall that  $\beta = \max\{\|A\|, \|B\|\} < 1$ . The cardinality of  $\{A;B\}^k$  is obviously  $2^k$ . Consequently, if

$$U_n = \sum_{k=0}^n \frac{1}{2^k} \sum_{C \in \{A;B\}^k} C(T_{n-k} - T)C^t,$$

it is not hard to see that

$$\|U_n\| \leq \sum_{k=0}^n \frac{1}{2^k} \times 2^k \beta^{2k} \|T_{n-k} - T\| = \sum_{k=0}^n \beta^{2(n-k)} \|T_k - T\|.$$

Hence,  $(U_n)$  converges to zero which completes the proof of Lemma A.4.  $\square$

We now return to the BAR process. We first need an estimate of the sum of the  $\|\mathbb{X}_n\|^2$  before being able to investigate the limits.

**Lemma A.5.** *Assume that  $(\varepsilon_n)$  satisfies (H.1) to (H.3). Then, we have*

$$\sum_{k \in \mathbb{T}_{n,p}} \|\mathbb{X}_k\|^2 = \mathcal{O}(|\mathbb{T}_n|) \quad \text{a.s.} \quad (\text{A.4})$$

**Proof:** In all the sequel, for all  $n \geq 2^{p-1}$ , denote  $A_{2n} = A$  and  $A_{2n+1} = B$ . It follows from a recursive application of relation (2.2) that for all  $n \geq 2^{p-1}$

$$\mathbb{X}_n = \left( \prod_{k=0}^{r_n-p} A_{\lfloor \frac{n}{2^k} \rfloor} \right) \mathbb{X}_{\lfloor \frac{n}{2^{r_n-p+1}} \rfloor} + \sum_{k=0}^{r_n-p} \left( \prod_{i=0}^{k-1} A_{\lfloor \frac{n}{2^i} \rfloor} \right) \eta_{\lfloor \frac{n}{2^k} \rfloor} \quad (\text{A.5})$$

with the convention that an empty product equals 1. Then, we can deduce from Cauchy-Schwarz inequality that for all  $n \geq 2^{p-1}$

$$\begin{aligned} \left\| \mathbb{X}_n - \left( \prod_{k=0}^{r_n-p} A_{\lfloor \frac{n}{2^k} \rfloor} \right) \mathbb{X}_{\lfloor \frac{n}{2^{r_n-p+1}} \rfloor} \right\|^2 &= \left\| \sum_{k=0}^{r_n-p} \left( \prod_{i=0}^{k-1} A_{\lfloor \frac{n}{2^i} \rfloor} \right) \eta_{\lfloor \frac{n}{2^k} \rfloor} \right\|^2 \\ &\leq \left( \sum_{k=0}^{r_n-p} \left( \prod_{i=0}^{k-1} \|A_{\lfloor \frac{n}{2^i} \rfloor}\| \right) \|\eta_{\lfloor \frac{n}{2^k} \rfloor}\| \right)^2 \\ &\leq \left( \sum_{k=0}^{r_n-p} \beta^k \|\eta_{\lfloor \frac{n}{2^k} \rfloor}\| \right)^2 \\ &\leq \left( \sum_{k=0}^{r_n-p} \beta^k \right) \left( \sum_{k=0}^{r_n-p} \beta^k \|\eta_{\lfloor \frac{n}{2^k} \rfloor}\|^2 \right) \\ &\leq \frac{1}{1-\beta} \left( \sum_{k=0}^{r_n-p} \beta^k \|\eta_{\lfloor \frac{n}{2^k} \rfloor}\|^2 \right). \end{aligned}$$

Hence, we obtain that for all  $n \geq 2^p$ ,

$$\begin{aligned} \|\mathbb{X}_n\|^2 &= \left\| \mathbb{X}_n - \left( \prod_{k=0}^{r_n-p} A_{\lfloor \frac{n}{2^k} \rfloor} \right) \mathbb{X}_{\lfloor \frac{n}{2^{r_n-p+1}} \rfloor} + \left( \prod_{k=0}^{r_n-p} A_{\lfloor \frac{n}{2^k} \rfloor} \right) \mathbb{X}_{\lfloor \frac{n}{2^{r_n-p+1}} \rfloor} \right\|^2 \\ &\leq \frac{2}{1-\beta} \left( \sum_{k=0}^{r_n-p} \beta^k \|\eta_{\lfloor \frac{n}{2^k} \rfloor}\|^2 \right) + 2\beta^{2(r_n-p+1)} \|\mathbb{X}_{\lfloor \frac{n}{2^{r_n-p+1}} \rfloor}\|^2. \end{aligned}$$

Denote  $\alpha = \max\{|a_0|, |b_0|\}$  and  $\bar{X}_1 = \max\{\|X_k\|, k \leq 2^{p-1}\}$ . Summing up over the sub-tree  $\mathbb{T}_{n,p}$ , we find that

$$\begin{aligned} \sum_{k \in \mathbb{T}_{n,p}} \|\mathbb{X}_k\|^2 &\leq \sum_{k \in \mathbb{T}_{n,p}} \frac{2}{1-\beta} \left( \sum_{i=0}^{r_k-p} \beta^i \|\eta_{\lfloor \frac{k}{2^i} \rfloor}\|^2 \right) + \sum_{k \in \mathbb{T}_{n,p}} 2\beta^{2(r_k-p+1)} \|\mathbb{X}_{\lfloor \frac{k}{2^{r_k-p+1}} \rfloor}\|^2 \\ &\leq \frac{4}{1-\beta} \sum_{k \in \mathbb{T}_{n,p}} \sum_{i=0}^{r_k-p} \beta^i (\alpha^2 + \varepsilon_{\lfloor \frac{k}{2^i} \rfloor}^2) + \sum_{k \in \mathbb{T}_{n,p}} 2\beta^{2(r_k-p+1)} \|X_{\lfloor \frac{k}{2^{r_k-p+1}} \rfloor}\|^2 \\ &\leq \frac{4}{1-\beta} \sum_{k \in \mathbb{T}_{n,p}} \sum_{i=0}^{r_k-p} \beta^i \varepsilon_{\lfloor \frac{k}{2^i} \rfloor}^2 + \frac{4\alpha^2}{1-\beta} \sum_{k \in \mathbb{T}_{n,p}} \sum_{i=0}^{r_k-p} \beta^i \\ &\quad + 2\bar{X}_1^{-2} \sum_{k \in \mathbb{T}_{n,p}} \beta^{2(r_k-p+1)}, \\ &\leq \frac{4P_n}{1-\beta} + \frac{4\alpha^2 Q_n}{1-\beta} + 2\bar{X}_1^{-2} R_n, \end{aligned} \quad (\text{A.6})$$



where

$$P_n = \sum_{k \in \mathbb{T}_{n,p}} \sum_{i=0}^{r_k-p} \beta^i \varepsilon_{\lfloor \frac{k}{2^i} \rfloor}^2, \quad Q_n = \sum_{k \in \mathbb{T}_{n,p}} \sum_{i=0}^{r_k-p} \beta^i, \quad R_n = \sum_{k \in \mathbb{T}_{n,p}} \beta^{2(r_k-p+1)}.$$

The last two terms of (A.6) are readily evaluated by splitting the sums generation-wise. As a matter of fact,

$$Q_n = \sum_{k=p}^n \sum_{i \in \mathbb{G}_k} \frac{1 - \beta^k}{1 - \beta} \leq \frac{1}{(1 - \beta)} \sum_{k=p}^n 2^k = \mathcal{O}(|\mathbb{T}_n|), \quad (\text{A.7})$$

and

$$R_n = \sum_{k=p}^n \sum_{i \in \mathbb{G}_k} \beta^{k-p+1} \leq \sum_{k=p}^n (2\beta)^k = \mathcal{O}(|\mathbb{T}_n|). \quad (\text{A.8})$$

It remains to control the first term  $P_n$ . One can observe that  $\varepsilon_k$  appears in  $P_n$  as many times as it has descendants up to the  $n$ th generation, and its multiplicative factor for its  $i$ th generation descendant is  $(2\beta)^i$ . Hence, one has

$$P_n = \sum_{k \in \mathbb{T}_{n,p}} \sum_{i=0}^{n-r_k} (2\beta)^i \varepsilon_k^2.$$

The evaluation of  $P_n$  depends on the value of  $0 < \beta < 1$ . On the one hand, if  $\beta = 1/2$ ,  $P_n$  reduces to

$$P_n = \sum_{k \in \mathbb{T}_{n,p}} (n+1-r_k) \varepsilon_k^2 = \sum_{k=p}^n (n+1-k) \sum_{i \in \mathbb{G}_k} \varepsilon_i^2.$$

Hence,

$$\frac{P_n}{|\mathbb{T}_n| + 1} = \sum_{k=p}^n \left( \frac{(n+1-k)}{2^{n+1-k}} \right) \left( \frac{1}{|\mathbb{G}_k|} \sum_{i \in \mathbb{G}_k} \varepsilon_i^2 \right).$$

However, it follows from Remark 6.2 that

$$\lim_{n \rightarrow +\infty} \frac{1}{|\mathbb{G}_n|} \sum_{k \in \mathbb{G}_n} \varepsilon_k^2 = \sigma^2 \quad \text{a.s.}$$

In addition, we also have

$$\lim_{n \rightarrow \infty} \sum_{k=1}^n \frac{k}{2^k} = 2.$$

Consequently, we infer from Lemma A.3 that

$$\lim_{n \rightarrow +\infty} \frac{P_n}{|\mathbb{T}_n|} = 2\sigma^2 \quad \text{a.s.} \quad (\text{A.9})$$

On the other hand, if  $\beta \neq 1/2$ , we have

$$P_n = \sum_{k \in \mathbb{T}_{n,p}} \frac{1 - (2\beta)^{n-r_k+1}}{1 - 2\beta} \varepsilon_k^2 = \frac{1}{1 - 2\beta} \sum_{k=p}^n (1 - (2\beta)^{n-k+1}) \sum_{i \in \mathbb{G}_k} \varepsilon_i^2.$$

Thus,

$$\frac{P_n}{|\mathbb{T}_n| + 1} = \frac{1}{1 - 2\beta} \sum_{k=p}^n \left( \left(\frac{1}{2}\right)^{n-k+1} - \beta^{n-k+1} \right) \left( \frac{1}{|\mathbb{G}_k|} \sum_{i \in \mathbb{G}_k} \varepsilon_i^2 \right).$$

Furthermore,

$$\lim_{n \rightarrow \infty} \frac{1}{1 - 2\beta} \sum_{k=1}^n \left( \left(\frac{1}{2}\right)^k - \beta^k \right) = \frac{1}{1 - \beta}.$$

As before, we deduce from Lemma A.3 that

$$\lim_{n \rightarrow +\infty} \frac{P_n}{|\mathbb{T}_n|} = \frac{\sigma^2}{1 - \beta}. \quad \text{a.s.} \quad (\text{A.10})$$

Finally, Lemma A.5 follows from the conjunction of (A.6), (A.7), (A.8) together with (A.9) and (A.10).  $\square$

**Proof of Lemma 7.1 :** First of all, denote

$$H_n = \sum_{k \in \mathbb{T}_{n,p-1}} \mathbb{X}_k \quad \text{and} \quad P_n = \sum_{k \in \mathbb{T}_{n,p}} \varepsilon_k,$$

As  $|\mathbb{T}_n| = 2^{n+1} - 1$ , we obtain from Equation (2.2) the recursive relation

$$\begin{aligned} H_n &= H_{p-1} + \sum_{k \in \mathbb{T}_{n,p}} \left( A_k \mathbb{X}_{\lfloor \frac{k}{2} \rfloor} + \eta_k \right), \\ &= H_{p-1} + 2\bar{A}H_{n-1} + 2\bar{a}(2^n - 2^{p-1})e_1 + P_n e_1 \end{aligned} \quad (\text{A.11})$$

where  $e_1 = (1, 0, \dots, 0)^t \in \mathbb{R}^p$ ,  $\bar{a} = (a_0 + b_0)/2$  and the matrix

$$\bar{A} = \frac{A+B}{2}.$$

By induction, we deduce from (A.11) that

$$\begin{aligned} \frac{H_n}{2^{n+1}} &= \frac{H_{p-1}}{2^{n+1}} + \bar{A} \frac{H_{n-1}}{2^n} + \bar{a} \left( \frac{2^n - 2^{p-1}}{2^n} \right) e_1 + \frac{P_n}{2^{n+1}} e_1, \\ &= (\bar{A})^{n-p+1} \frac{H_{p-1}}{2^p} + \sum_{k=p}^n (\bar{A})^{n-k} \left( \frac{H_{p-1}}{2^{k+1}} + \bar{a} \left( \frac{2^k - 2^{p-1}}{2^k} \right) e_1 + \frac{P_k}{2^{k+1}} e_1 \right). \end{aligned}$$

We have already seen via convergence (6.1) of Lemma 6.1 that

$$\lim_{n \rightarrow +\infty} \frac{P_n}{2^{n+1}} = 0 \quad \text{a.s.}$$

Finally, as  $\|\bar{A}\| < 1$ ,

$$\sum_{n=0}^{\infty} \|(\bar{A})^n\| < \infty \quad \text{and} \quad (I_p - \bar{A})^{-1} = \sum_{n=0}^{\infty} (\bar{A})^n,$$

it follows from Lemma A.3 that

$$\lim_{n \rightarrow \infty} \frac{H_n}{2^{n+1}} = \bar{a}(I_p - \bar{A})^{-1}e_1 \quad \text{a.s.}$$

which ends the proof of Lemma 7.1.  $\square$

**Proof of Lemma 7.2 :** We shall proceed as in the proof of Lemma 7.1 and use the same notation. Let

$$K_n = \sum_{k \in \mathbb{T}_{n,p-1}} \mathbb{X}_k \mathbb{X}_k^t \quad \text{and} \quad L_n = \sum_{k \in \mathbb{T}_{n,p}} \varepsilon_k^2.$$

We infer again from (2.2) that

$$\begin{aligned} K_n &= K_{p-1} + \sum_{k \in \mathbb{T}_{n,p}} \left( A_k \mathbb{X}_{\lfloor \frac{k}{2} \rfloor} + \eta_k \right) \left( A_k \mathbb{X}_{\lfloor \frac{k}{2} \rfloor} + \eta_k \right)^t \\ &= K_{p-1} + \sum_{k \in \mathbb{T}_{n,p}} \varepsilon_k^2 e_1 e_1^t + \sum_{k \in \mathbb{T}_{n-1,p-1}} \left( A \mathbb{X}_k \mathbb{X}_k^t A^t + B \mathbb{X}_k \mathbb{X}_k^t B^t \right) \\ &\quad + \sum_{k \in \mathbb{T}_{n-1,p-1}} \left( (a_0 + \varepsilon_{2k}) \mathbb{U}_k(A) + (b_0 + \varepsilon_{2k+1}) \mathbb{U}_k(B) + 2(\bar{a}^2 + \zeta_{2k}) e_1 e_1^t \right) \end{aligned}$$

where  $\mathbb{U}_k(A) = A \mathbb{X}_k e_1^t + e_1 \mathbb{X}_k^t A^t$  and  $\mathbb{U}_k(B) = B \mathbb{X}_k e_1^t + e_1 \mathbb{X}_k^t B^t$ . In addition,  $\bar{a}^2 = (a_0^2 + b_0^2)/2$  and  $\zeta_{2k} = (a_0 \varepsilon_{2k} + b_0 \varepsilon_{2k+1})$ . Therefore, we obtain that

$$\frac{K_n}{2^{n+1}} = \frac{1}{2} \left( A \frac{K_{n-1}}{2^n} A^t + B \frac{K_{n-1}}{2^n} B^t \right) + T_n \quad (\text{A.12})$$

where

$$\begin{aligned} T_n &= \left( \frac{L_n}{2^{n+1}} + \bar{a}^2 \left( \frac{2^n - 2^{p-1}}{2^n} \right) + \frac{1}{2^n} \sum_{k \in \mathbb{T}_{n-1,p-1}} \zeta_{2k} \right) e_1 e_1^t \\ &\quad + \frac{1}{2} \left( a_0 \left( A \frac{H_{n-1}}{2^n} e_1^t + e_1 \frac{H_{n-1}^t}{2^n} A^t \right) + b_0 \left( B \frac{H_{n-1}}{2^n} e_1^t + e_1 \frac{H_{n-1}^t}{2^n} B^t \right) \right) \\ &\quad + \frac{1}{2^{n+1}} \sum_{k \in \mathbb{T}_{n-1,p-1}} \left( \varepsilon_{2k} \mathbb{U}_k(A) + \varepsilon_{2k+1} \mathbb{U}_k(B) \right). \end{aligned}$$

The two first results (6.1) and (6.2) of Lemma 6.1 together with Remark 6.2 and Lemma A.2 readily imply that

$$\lim_{n \rightarrow +\infty} \frac{L_n}{2^{n+1}} = \sigma^2 \quad \text{a.s.}$$

and

$$\lim_{n \rightarrow +\infty} \frac{1}{2^n} \sum_{k \in \mathbb{T}_{n-1,p-1}} \zeta_{2k} = 0 \quad \text{a.s.}$$

In addition, Lemma 7.1 gives

$$\lim_{n \rightarrow +\infty} \frac{H_{n-1}}{2^n} = \lambda \quad \text{a.s.}$$

Furthermore, denote

$$U_n = \sum_{k \in \mathbb{T}_{n-1, p-1}} \left( \varepsilon_{2k} \mathbb{U}_k(A) + \varepsilon_{2k+1} \mathbb{U}_k(B) \right).$$

For all  $u \in \mathbb{R}^p$ , let  $U_n(u) = u^t U_n u$ . The sequence  $(U_n(u))$  is a real martingale transform. Moreover, it follows from Lemma A.5 that

$$\sum_{k \in \mathbb{T}_{n-1, p-1}} \left| u^t \mathbb{U}_k(A) u \right|^2 + \left| u^t \mathbb{U}_k(B) u \right|^2 = \mathcal{O}(|\mathbb{T}_n|) \quad \text{a.s.}$$

Consequently, we deduce from the strong law of large numbers for martingale transforms given in Theorem 1.3.24 of [4] that  $U_n(u) = o(|\mathbb{T}_n|)$  a.s. for all  $u \in \mathbb{R}^p$  which leads to  $U_n = o(|\mathbb{T}_n|)$  a.s. Therefore, we obtain that  $(T_n)$  converges a.s. to  $T$  given by

$$T = (\sigma^2 + \overline{a^2}) e_1 e_1^t + \frac{1}{2} \left( A \lambda a_0 e_1^t + a_0 e_1 \lambda^t A^t + B \lambda b_0 e_1^t + b_0 e_1 \lambda^t B^t \right).$$

Finally, iteration of the recursive relation (A.12) yields

$$\frac{K_n}{2^{n+1}} = \frac{1}{2^{n-p+1}} \sum_{C \in \{A; B\}^{n-p+1}} C \frac{K_{p-1}}{2^p} C^t + \sum_{k=0}^{n-p} \frac{1}{2^k} \sum_{C \in \{A; B\}^k} C T_{n-k} C^t.$$

On the one hand, the first term on the right-hand side converges a.s. to zero as its norm is bounded  $\beta^{2(n-p+1)} \|K_{p-1}\| / 2^p$ . On the other hand, thanks to Lemma A.4, the second term on the right-hand side converges to  $\ell$  given by (A.3), which completes the proof of Lemma 7.2.  $\square$

We now state a convergence result for the sum of  $\|\mathbb{X}_n\|^4$  which will be useful for the CLT.

**Lemma A.6.** *Assume that  $(\varepsilon_n)$  satisfies (H.1) to (H.5). Then, we have*

$$\sum_{k \in \mathbb{T}_{n, p}} \|\mathbb{X}_k\|^4 = \mathcal{O}(|\mathbb{T}_n|) \quad \text{a.s.} \quad (\text{A.13})$$

**Proof :** The proof is almost exactly the same as that of Lemma A.5. Instead of Equation (A.6), we have

$$\sum_{k \in \mathbb{T}_{n, p}} \|\mathbb{X}_k\|^4 \leq \frac{64 P_n}{(1 - \beta)^3} + \frac{64 \alpha^4 Q_n}{(1 - \beta)^3} + 8 \overline{X}_1^4 R_n$$

where

$$P_n = \sum_{k \in \mathbb{T}_{n, p}} \sum_{i=0}^{r_k - p} \beta^i \varepsilon_{\lfloor \frac{k}{2^i} \rfloor}^4, \quad Q_n = \sum_{k \in \mathbb{T}_{n, p}} \sum_{i=0}^{r_k - p} \beta^i, \quad R_n = \sum_{k \in \mathbb{T}_{n, p}} \beta^{4(r_k - p + 1)}.$$

We already saw that  $Q_n = \mathcal{O}(|\mathbb{T}_n|)$ . In addition, it is not hard to see that  $R_n = \mathcal{O}(|\mathbb{T}_n|)$ . Therefore, we only need a sharper estimate for  $u_n$ . Via the same lines as in the proof of Lemma A.5 together with the sharper results of Lemma 6.3, we can show that  $P_n = \mathcal{O}(|\mathbb{T}_n|)$  a.s. which leads to (A.13).  $\square$

## B On the quadratic strong law

We start with an auxiliary lemma closely related to the Riccacion Equation for the inverse of the matrix  $S_n$ .

**Lemma B.1.** *Let  $h_n$  and  $l_n$  be the two following symmetric square matrices of order  $\delta_n$*

$$h_n = \Phi_n^t S_n^{-1} \Phi_n \quad \text{and} \quad l_n = \Phi_n^t S_{n-1}^{-1} \Phi_n.$$

*Then, the inverse of  $S_n$  may be recursively calculated as*

$$S_n^{-1} = S_{n-1}^{-1} - S_{n-1}^{-1} \Phi_n (\mathbf{I}_{\delta_n} + l_n)^{-1} \Phi_n^t S_{n-1}^{-1}. \quad (\text{B.1})$$

*In addition, we also have  $(\mathbf{I}_{\delta_n} - h_n)(\mathbf{I}_{\delta_n} + l_n) = \mathbf{I}_{\delta_n}$ .*

**Remark B.2.** *If  $f_n = \Psi_n^t \Sigma_n^{-1} \Psi_n$ , it follows from Lemma B.1 that*

$$\Sigma_n^{-1} = \Sigma_{n-1}^{-1} - \Sigma_{n-1}^{-1} \Psi_n (\mathbf{I}_{2\delta_n} - f_n) \Psi_n^t \Sigma_{n-1}^{-1}. \quad (\text{B.2})$$

**Proof :** As  $S_n = S_{n-1} + \Phi_n \Phi_n^t$ , relation (B.1) immediately follows from Riccati Equation given e.g. in [4] page 96. By multiplying both side of (B.1) by  $\Phi_n$ , we obtain

$$\begin{aligned} S_n^{-1} \Phi_n &= S_{n-1}^{-1} \Phi_n - S_{n-1}^{-1} \Phi_n (\mathbf{I}_{\delta_n} + l_n)^{-1} l_n, \\ &= S_{n-1}^{-1} \Phi_n - S_{n-1}^{-1} \Phi_n (\mathbf{I}_{\delta_n} + l_n)^{-1} (\mathbf{I}_{\delta_n} + l_n - \mathbf{I}_{\delta_n}), \\ &= S_{n-1}^{-1} \Phi_n (\mathbf{I}_{\delta_n} + l_n)^{-1}. \end{aligned}$$

Consequently, multiplying this time on the left by  $\Phi_n^t$ , we obtain that

$$\begin{aligned} h_n &= l_n (\mathbf{I}_{\delta_n} + l_n)^{-1} = (l_n + \mathbf{I}_{\delta_n} - \mathbf{I}_{\delta_n}) (\mathbf{I}_{\delta_n} + l_n)^{-1}, \\ &= \mathbf{I}_{\delta_n} - (\mathbf{I}_{\delta_n} + l_n)^{-1} \end{aligned}$$

leading to  $(\mathbf{I}_{\delta_n} - h_n)(\mathbf{I}_{\delta_n} + l_n) = \mathbf{I}_{\delta_n}$ . □

In order to establish the quadratic strong law for  $(M_n)$ , we are going to study separately the asymptotic behaviour of  $(\mathcal{W}_n)$  and  $(\mathcal{B}_n)$  which appear in the main decomposition (8.1).

**Lemma B.3.** *Assume that  $(\varepsilon_n)$  satisfies (H.1) to (H.3). Then, we have*

$$\lim_{n \rightarrow +\infty} \frac{1}{n} \mathcal{W}_n = 2\sigma^2 \quad \text{a.s.} \quad (\text{B.3})$$

**Proof :** First of all, we have the decomposition  $\mathcal{W}_{n+1} = \mathcal{T}_{n+1} + \mathcal{R}_{n+1}$  where

$$\begin{aligned} \mathcal{T}_{n+1} &= \sum_{k=p}^n \frac{\Delta M_{k+1}^t \Lambda^{-1} \Delta M_{k+1}}{|\mathbb{T}_k|}, \\ \mathcal{R}_{n+1} &= \sum_{k=p}^n \frac{\Delta M_{k+1}^t (|\mathbb{T}_k| \Sigma_k^{-1} - \Lambda^{-1}) \Delta M_{k+1}}{|\mathbb{T}_k|}. \end{aligned}$$

We claim that

$$\lim_{n \rightarrow +\infty} \frac{1}{n} \mathcal{T}_n = (p+1)\sigma^2 \quad \text{a.s.}$$

It will ensure via (8.5) that  $\mathcal{R}_n = o(n)$  a.s. leading to (B.3). One can observe that  $\mathcal{T}_{n+1} = \text{tr}(\Lambda^{-1/2} H_{n+1} \Lambda^{-1/2})$  where

$$H_{n+1} = \sum_{k=p}^n \frac{\Delta M_{k+1} \Delta M_{k+1}^t}{|\mathbb{T}_k|}.$$

Our goal is to make use of the strong law of large numbers for martingale transforms, so we start by adding and subtracting a term involving the conditional expectation of  $\Delta H_{n+1}$  given  $\mathcal{F}_n$ . We have already seen in Section 4 that for all  $n \geq p-1$ ,  $\mathbb{E}[\Delta M_{n+1} \Delta M_{n+1}^t | \mathcal{F}_n] = \Gamma \otimes \Phi_n \Phi_n^t$ . Consequently, we can split  $H_{n+1}$  into two terms

$$H_{n+1} = \sum_{k=p}^n \frac{\Gamma \otimes \Phi_k \Phi_k^t}{|\mathbb{T}_k|} + K_{n+1}$$

where

$$K_{n+1} = \sum_{k=p}^n \frac{\Delta M_{k+1} \Delta M_{k+1}^t - \Gamma \otimes \Phi_k \Phi_k^t}{|\mathbb{T}_k|}.$$

On the one hand, it follows from convergence (5.1) and Lemma A.2 that

$$\lim_{n \rightarrow +\infty} \frac{\Phi_n \Phi_n^t}{|\mathbb{T}_n|} = \frac{1}{2} L \quad \text{a.s.}$$

Thus, Cesaro convergence yields

$$\lim_{n \rightarrow +\infty} \frac{1}{n} \sum_{k=p}^n \frac{\Gamma \otimes \Phi_k \Phi_k^t}{|\mathbb{T}_k|} = \frac{1}{2} (\Gamma \otimes L) \quad \text{a.s.} \quad (\text{B.4})$$

On the other hand, the sequence  $(K_n)$  is obviously a matrix martingale transform satisfying

$$\Delta K_{n+1} = K_{n+1} - K_n = \frac{1}{|\mathbb{T}_{n+1}|} \sum_{i,j \in \mathbb{G}_n} \Gamma_{ij} \otimes \begin{pmatrix} 1 & \mathbb{X}_j^t \\ \mathbb{X}_i & \mathbb{X}_i \mathbb{X}_j^t \end{pmatrix}$$

where

$$\Gamma_{ij} = \begin{pmatrix} \varepsilon_{2i} \varepsilon_{2j} - \mathbb{I}_{i=j} \sigma^2 & \varepsilon_{2i} \varepsilon_{2j+1} - \mathbb{I}_{i=j} \rho \\ \varepsilon_{2i+1} \varepsilon_{2j} - \mathbb{I}_{i=j} \rho & \varepsilon_{2i+1} \varepsilon_{2j+1} - \mathbb{I}_{i=j} \sigma^2 \end{pmatrix}.$$

For all  $u \in \mathbb{R}^{2(p+1)}$ , let  $K_n(u) = u^t K_n u$ . It follows from tedious but straightforward calculations, together with (A.4), (A.13) and the strong law of large numbers for martingale transforms given in Theorem 1.3.24 of [4] that  $K_n(u) = o(n)$  a.s. for all  $u \in \mathbb{R}^{2(p+1)}$  leading to  $K_n = o(n)$  a.s. Hence, we infer from (B.4) that

$$\lim_{n \rightarrow +\infty} \frac{1}{n} H_n = \frac{1}{2} (\Gamma \otimes L) \quad \text{a.s.} \quad (\text{B.5})$$

Finally, we find from (B.5) that

$$\begin{aligned} \lim_{n \rightarrow +\infty} \frac{1}{n} \mathcal{F}_n &= \frac{1}{2} \text{tr}(\Lambda^{-1/2}(\Gamma \otimes L)\Lambda^{-1/2}) \quad \text{a.s.} \\ &= \frac{1}{2} \text{tr}((\Gamma \otimes L)\Lambda^{-1}) \quad \text{a.s.} \\ &= \frac{1}{2} \text{tr}(\Gamma \otimes I_{p+1}) = (p+1)\sigma^2 \quad \text{a.s.} \end{aligned}$$

which completes the proof of Lemma B.3  $\square$

**Lemma B.4.** *Assume that  $(\varepsilon_n)$  satisfies (H.1) to (H.3). Then, we have*

$$\mathcal{B}_{n+1} = o(n) \quad \text{a.s.}$$

**Proof :** Recall that

$$\mathcal{B}_{n+1} = 2 \sum_{k=p}^n M_k^t \Sigma_k^{-1} \Delta M_{k+1} = 2 \sum_{k=p}^n M_k^t \Sigma_k^{-1} \Psi_k \xi_{k+1}.$$

The sequence  $(\mathcal{B}_n)$  is a real martingale transform satisfying

$$\Delta \mathcal{B}_{n+1} = \mathcal{B}_{n+1} - \mathcal{B}_n = 2M_n^t \Sigma_n^{-1} \Psi_n \xi_{n+1}.$$

Consequently, via the strong law of large numbers for martingale transforms [4], we find that either  $(\mathcal{B}_n)$  converges a.s. or  $\mathcal{B}_{n+1} = o(v_n)$  a.s. where

$$v_n = \sum_{k=p}^n M_k^t \Sigma_k^{-1} \Psi_k \Psi_k^t \Sigma_k^{-1} M_k.$$

However, for all  $n \geq 2^{p-1}$ ,  $\Psi_n \Psi_n^t = I_2 \otimes \Phi_n \Phi_n^t$  which implies that

$$v_n = \sum_{k=p}^n M_k^t \Sigma_k^{-1} (I_2 \otimes \Phi_k \Phi_k^t) \Sigma_k^{-1} M_k = \sum_{k=p}^n M_k^t (I_2 \otimes S_k^{-1} \Phi_k \Phi_k^t S_k^{-1}) M_k.$$

Furthermore, it follows from Lemma B.1 that

$$S_{n-1}^{-1} - S_n^{-1} = S_n^{-1} \Phi_n (I_{\delta_n} + l_n) \Phi_n^t S_n^{-1} \geq S_n^{-1} \Phi_n \Phi_n^t S_n^{-1}$$

as the matrix  $l_n$  is definite positive. Therefore, we obtain that

$$v_n \leq \sum_{k=p}^n M_k^t (\Sigma_{k-1}^{-1} - \Sigma_k^{-1}) M_k = \mathcal{A}_n.$$

Finally, we deduce from the main decomposition (8.1) that

$$\mathcal{V}_{n+1} + \mathcal{A}_n = o(\mathcal{A}_n) + \mathcal{O}(n) \quad \text{a.s.}$$

leading to  $\mathcal{V}_{n+1} = \mathcal{O}(n)$  and  $\mathcal{A}_n = \mathcal{O}(n)$  a.s. as  $\mathcal{V}_{n+1}$  and  $\mathcal{A}_n$  are non-negative, which implies in turn that  $\mathcal{B}_n = o(n)$  a.s. completing the proof of Lemma B.4.  $\square$

**Proof of Lemma 8.1 :** Convergence (8.2) immediately follows from (8.1) together with Lemmas B.3 and B.4.  $\square$

## C On Wei's Lemma

In order to prove (8.3), we shall apply Wei's Lemma given in [13] page 1672, to each entry of the vector-valued martingale

$$M_n = \sum_{k=p}^n \sum_{i \in \mathbb{G}_{k-1}} \begin{pmatrix} \varepsilon_{2i} \\ \mathbb{X}_i \varepsilon_{2i} \\ \varepsilon_{2i+1} \\ \mathbb{X}_i \varepsilon_{2i+1} \end{pmatrix}.$$

We shall only carry out the proof for the first  $(p+1)$  of  $M_n$  inasmuch as the proof for the  $(p+1)$  last components follows exactly the same lines. Denote

$$P_n = \sum_{k=p}^n \sum_{i \in \mathbb{G}_{k-1}} \varepsilon_{2i} \quad \text{and} \quad Q_n = \sum_{k=p}^n \sum_{i \in \mathbb{G}_{k-1}} \mathbb{X}_i \varepsilon_{2i}.$$

On the one hand,  $P_n$  can be rewritten as  $P_n = \sum_{k=p}^n \sqrt{|\mathbb{G}_{k-1}|} v_k$  where

$$v_n = \frac{1}{\sqrt{|\mathbb{G}_{n-1}|}} \sum_{i \in \mathbb{G}_{n-1}} \varepsilon_{2i}.$$

We clearly have  $\mathbb{E}[v_{n+1} | \mathcal{F}_n] = 0$ ,  $\mathbb{E}[v_{n+1}^2 | \mathcal{F}_n] = \sigma^2$  a.s. Moreover, it follows from **(H.1)** to **(H.3)** together with Cauchy-Schwarz inequality that

$$\begin{aligned} \mathbb{E}[v_{n+1}^4 | \mathcal{F}_n] &= \frac{1}{|\mathbb{G}_n|^2} \sum_{i \in \mathbb{G}_n} \mathbb{E}[\varepsilon_{2i}^4 | \mathcal{F}_n] + \frac{3}{|\mathbb{G}_n|^2} \sum_{i \in \mathbb{G}_n} \sum_{j \neq i} \mathbb{E}[\varepsilon_{2i}^2 | \mathcal{F}_n] \mathbb{E}[\varepsilon_{2j}^2 | \mathcal{F}_n] \\ &\leq 3 \sup_{i \in \mathbb{G}_n} \mathbb{E}[\varepsilon_{2i}^4 | \mathcal{F}_n] \quad \text{a.s.} \end{aligned}$$

which implies that  $\sup \mathbb{E}[v_{n+1}^4 | \mathcal{F}_n] < +\infty$  a.s. Consequently, we deduce from Wei's Lemma that for all  $\delta > 1/2$ ,

$$P_n^2 = o(|\mathbb{T}_{n-1}| n^\delta) \quad \text{a.s.}$$

On the other hand, we also have  $Q_n = \sum_{k=p}^n \sqrt{|\mathbb{G}_{k-1}|} w_k$  where

$$w_n = \frac{1}{\sqrt{|\mathbb{G}_{n-1}|}} \sum_{i \in \mathbb{G}_{n-1}} \mathbb{X}_i \varepsilon_{2i}.$$

It is not hard to see that  $\mathbb{E}[w_{n+1} | \mathcal{F}_n] = 0$  a.s. Moreover, for all  $1 \leq k \leq p$ , let  $w_n(k)$  be the  $k$ th coordinate of the vector  $w_n$ . It follows from **(H.1)** to **(H.3)** and Cauchy-Schwarz inequality that for all  $1 \leq k \leq p$ ,

$$\begin{aligned} \mathbb{E}[w_{n+1}(k)^4 | \mathcal{F}_n] &\leq \frac{1}{|\mathbb{G}_n|^2} \sum_{i \in \mathbb{G}_n} X_{\lfloor \frac{i}{2^{k-1}} \rfloor}^4 \mathbb{E}[\varepsilon_{2i}^4 | \mathcal{F}_n] + \frac{3\sigma^4}{|\mathbb{G}_n|^2} \sum_{i \in \mathbb{G}_n} \sum_{j \neq i} X_{\lfloor \frac{i}{2^{k-1}} \rfloor}^2 X_{\lfloor \frac{j}{2^{k-1}} \rfloor}^2 \\ &\leq 3 \sup_{i \in \mathbb{G}_n} \mathbb{E}[\varepsilon_{2i}^4 | \mathcal{F}_n] \left( \frac{1}{|\mathbb{G}_n|} \sum_{i \in \mathbb{G}_n} X_{\lfloor \frac{i}{2^{k-1}} \rfloor}^2 \right)^2 \quad \text{a.s.} \end{aligned}$$



Hence, we obtain from Lemma 7.2 that for all  $1 \leq k \leq p$ ,  $\sup \mathbb{E}[w_{n+1}(k)^4 | \mathcal{F}_n] < +\infty$  a.s. Once again, we deduce from Wei's Lemma applied to each component of  $Q_n$  that for all  $\delta > 1/2$ ,

$$\|Q_n\|^2 = o(|\mathbb{T}_{n-1}|n^\delta) \quad \text{a.s.}$$

which completes the proof of (8.3).  $\square$

## D On the convergence of the covariance estimator

It remains to prove that

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k \in \mathbb{T}_{n-1, p-1}} (\widehat{\varepsilon}_{2k} - \varepsilon_{2k})(\widehat{\varepsilon}_{2k+1} - \varepsilon_{2k+1}) = \lim_{n \rightarrow \infty} \frac{R_n}{2n} = (p+1)\rho \quad \text{a.s.}$$

where

$$R_n = \sum_{k \in \mathbb{T}_{n-1, p-1}} (\widehat{V}_k - V_k)^t J_2 (\widehat{V}_k - V_k).$$

It is not possible to make use of the previous convergence (9.1) because the matrix

$$J_2 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

is not positive definite. Hence, it is necessary to rewrite our proofs. Denote

$$\mathcal{V}'_n = M_n^t \Sigma_{n-1}^{-1/2} (J_2 \otimes I_{p+1}) \Sigma_{n-1}^{-1/2} M_n.$$

As in the proof of Theorem 5.2, we have the decomposition

$$\mathcal{V}'_{n+1} + \mathcal{A}'_n = \mathcal{V}'_1 + \mathcal{B}'_{n+1} + \mathcal{W}'_{n+1} \quad (\text{D.1})$$

where

$$\begin{aligned} \mathcal{A}'_n &= \sum_{k=p}^n M_k^t (J_2 \otimes (S_{k-1}^{-1} - S_k^{-1})) M_k, \\ \mathcal{B}'_{n+1} &= 2 \sum_{k=p}^n M_k^t (J_2 \otimes S_k^{-1}) \Delta M_{k+1}, \\ \mathcal{W}'_{n+1} &= \sum_{k=p}^n \Delta M_{k+1}^t (J_2 \otimes S_k^{-1}) \Delta M_{k+1}. \end{aligned}$$

First of all, via the same lines as in Appendix B, we obtain that

$$\begin{aligned} \lim_{n \rightarrow +\infty} \frac{1}{n} \mathcal{W}'_n &= \frac{1}{2} \text{tr}((J_2 \otimes L^{-1})^{1/2} (\Gamma \otimes L) (J_2 \otimes L^{-1})^{1/2}) \quad \text{a.s.} \\ &= \frac{1}{2} \text{tr}(\Gamma J_2 \otimes I_{p+1}) = (p+1)\rho \quad \text{a.s.} \end{aligned}$$

Next,  $(\mathcal{B}'_n)$  is a real martingale transform satisfying  $\mathcal{B}'_{n+1} = o(n)$  a.s. Hence, we find the analogous of convergence (8.2)

$$\lim_{n \rightarrow +\infty} \frac{\mathcal{V}'_{n+1} + \mathcal{A}'_n}{n} = (p+1)\rho \quad \text{a.s.} \quad (\text{D.2})$$

Furthermore, it follows from Wei's Lemma that for all  $\delta > 1/2$ ,

$$\mathcal{V}'_n = o(n^\delta) \quad \text{a.s.} \quad (\text{D.3})$$

Therefore, we infer (D.1), (D.2) and (D.3) that

$$\lim_{n \rightarrow +\infty} \frac{1}{n} \mathcal{A}'_n = (p+1)\rho \quad \text{a.s.} \quad (\text{D.4})$$

Finally, by the same lines as in the proof of the first part of Theorem 5.3, we find that

$$\lim_{n \rightarrow \infty} \frac{R_n}{n} = 2 \lim_{n \rightarrow \infty} \frac{\mathcal{A}'_n}{n} = 2(p+1)\rho \quad \text{a.s.}$$

which completes the proof of convergence (9.2).  $\square$

## Acknowledgements

The authors would like to thanks the anonymous referees for their very careful reading of the manuscript and for their suggestion to extend the first version of the paper to asymmetric BAR( $p$ ) processes.

## References

- [1] BASAWA, I. V., AND ZHOU, J. Non-Gaussian bifurcating models and quasi-likelihood estimation. *J. Appl. Probab.* 41A (2004), 55–64. MR2057565
- [2] COWAN, R., AND STAUDTE, R. G. The bifurcating autoregressive model in cell lineage studies. *Biometrics* 42 (1986), 769–783.
- [3] DELMAS, J.-F., AND MARSALLE, L. Detection of cellular aging in a Galton-Watson process. *arXiv*, 0807.0749 (2008).
- [4] DUFLO, M. *Random iterative models*, vol. 34 of *Applications of Mathematics*. Springer-Verlag, Berlin, 1997. MR1485774
- [5] GUYON, J. Limit theorems for bifurcating Markov chains. Application to the detection of cellular aging. *Ann. Appl. Probab.* 17, 5-6 (2007), 1538–1569. MR2358633
- [6] GUYON, J., BIZE, A., PAUL, G., STEWART, E., DELMAS, J.-F., AND TADDÉI, F. Statistical study of cellular aging. In *CEMRACS 2004—mathematics and applications to biology and medicine*, vol. 14 of *ESAIM Proc.* EDP Sci., Les Ulis, 2005, pp. 100–114 (electronic). MR2226805
- [7] HALL, P., AND HEYDE, C. C. *Martingale limit theory and its application*. Academic Press Inc., New York, 1980. Probability and Mathematical Statistics. MR0624435

- [8] HAMILTON, J. D. *Time series analysis*. Princeton University Press, Princeton, NJ, 1994. MR1278033
- [9] HUGGINS, R. M. Robust inference for variance components models for single trees of cell lineage data. *Ann. Statist.* 24, 3 (1996), 1145–1160. MR1401842
- [10] HUGGINS, R. M., AND BASAWA, I. V. Extensions of the bifurcating autoregressive model for cell lineage studies. *J. Appl. Probab.* 36, 4 (1999), 1225–1233. MR1746406
- [11] HUGGINS, R. M., AND BASAWA, I. V. Inference for the extended bifurcating autoregressive model for cell lineage studies. *Aust. N. Z. J. Stat.* 42, 4 (2000), 423–432. MR1802966
- [12] HWANG, S. Y., BASAWA, I. V., AND YEO, I. K. Local asymptotic normality for bifurcating autoregressive processes and related asymptotic inference. *Statistical Methodology* 6 (2009), 61–69.
- [13] WEI, C. Z. Adaptive prediction by least squares predictors in stochastic regression models with applications to time series. *Ann. Statist.* 15, 4 (1987), 1667–1682. MR0913581
- [14] ZHOU, J., AND BASAWA, I. V. Least-squares estimation for bifurcating autoregressive processes. *Statist. Probab. Lett.* 74, 1 (2005), 77–88. MR2189078
- [15] ZHOU, J., AND BASAWA, I. V. Maximum likelihood estimation for a first-order bifurcating autoregressive process with exponential errors. *J. Time Ser. Anal.* 26, 6 (2005), 825–842. MR2203513