Annales de l'Institut Henri Poincaré - Probabilités et Statistiques
2014, Vol. 50, No. 3, 806–844
DOI: 10.1214/13-AIHP545

ANNALES
DE L'INSTITUT
HENRI
POINCARÉ
PROBABILITÉS
ET STATISTIQUES
www.imstat.org/aihp

Deviation inequalities and moderate deviations for estimators of parameters in bifurcating autoregressive models

S. Valère Bitseki Penda and Hacène Djellout

Laboratoire de Mathématiques, CNRS UMR 6620, Université Blaise Pascal, 24 Avenue des Landais, BP80026, 63177 Aubière, France.
E-mail: Valere.Bitsekipenda@math.univ-bpclermont.fr; Hacene.Djellout@math.univ-bpclermont.fr

Received 22 August 2012; revised 10 January 2013; accepted 31 January 2013

Abstract. The purpose of this paper is to investigate the deviation inequalities and the moderate deviation principle of the least squares estimators of the unknown parameters of general pth-order asymmetric bifurcating autoregressive processes, under suitable assumptions on the driven noise of the process. Our investigation relies on the moderate deviation principle for martingales.

Résumé. L'objetcif de ce papier est d'établir des inégalités de déviations et les principes de déviations modérées pour les estimateurs des moindres carrés des paramètres inconnus d'un processus bifurcant autorégressif asymétrique d'ordre p, sous certaines conditions sur la suite des bruits. Les preuves reposent sur les principes de déviations modérées des martingales.

MSC: 60F10; 62F12; 60G42; 62M10; 62G05

Keywords: Deviation inequalities; Moderate deviation principle; Bifurcating autoregressive process; Martingale; Limit theorems; Least squares estimation

1. Motivation and context

Bifurcating autoregressive processes (BAR, for short) are an adaptation of autoregressive processes, when the data has a binary tree structure. They were first introduced by Cowan and Staudte [6] for cell lineage data where each individual in one generation gives rise to two offspring in the next generation.

In their paper, the original BAR process is defined as follows. The initial cell is labelled 1, and the two offspring of cell k are labelled $2k$ and $2k+1$. If X_k denotes an observation of some characteristic of individual k then the first order BAR process is given, for all $k \geq 1$, by

$$\begin{cases} X_{2k} = a + bX_k + \varepsilon_{2k}, \\ X_{2k+1} = a + bX_k + \varepsilon_{2k+1}. \end{cases}$$

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

Several extensions of the model have been proposed and various estimators for the unknown parameters have been studied in the literature, see for instance [2,19–21,28,29]. See [3] for relevant references (although [3] deals with the asymmetric case unlike the above cited papers).

Recently, there have been many studies of the asymmetric BAR process, considering cases where the quantitative characteristics of the even and odd sisters are allowed to depend on their mother's through different sets of parameters.

In [18], Guyon proposes an interpretation of the asymmetric BAR process as a bifurcating Markov chain. This enables him to derive laws of large numbers and central limit theorems for the least squares estimators of the unknown parameters of the process. This Markov chain approach was further developed by Delmas and Marsalle [10], for cells which are allowed to die. They defined the genealogy of the cells through a Galton–Watson process, studying the same model on the Galton–Watson tree instead of a binary tree.

Another approach based on martingales theory was proposed by Bercu, de Saporta and Gégout-Petit [3], to sharpen the asymptotic analysis of Guyon, under weaker assumptions. It should be pointed out that missing data is not dealt with in this work. To take it into account in the estimation procedure, de Saporta et al. [8] and [9] use a two-type Galton–Watson process to model the genealogy.

Our objective in this paper is to go a step further by

- studying the moderate deviation principle (MDP, for short) of the least squares estimators of the unknown parameters of general asymmetric pth-order bifurcating autoregressive processes (BAR(p), for short). More precisely we are interested in the asymptotic estimations of

$$\mathbb{P}\left(\frac{\sqrt{n}}{v_n}(\Theta_n - \Theta) \in A\right),$$

where Θ_n denotes the estimator of the unknown parameter of interest Θ, A is a given domain of deviation, $(v_n > 0)$ is some sequence denoting the scale of deviation. When $v_n = 1$ this is exactly the estimation of the central limit theorem. When $v_n = \sqrt{n}$, it becomes the *large deviation*. And when $1 \ll v_n \ll \sqrt{n}$, this is the so called *moderate deviations*. Usually, MDP has a simpler rate function inherited from the approximated Gaussian process, and holds for a larger class of dependent random variables than the large deviation principle.

To prove our result on MDP, we use

(1) the work of Bercu et al. [3] on the almost sure convergence of the estimators with the quadratic strong law and the central limit theorem;
(2) the work of Dembo [11], and Worms [26,27] on the one hand, and the papers of Puhalskii [24] and Djellout [13] on the other hand, on the MDP for martingales.

- giving deviation inequalities for the estimator of bifurcating autoregressive processes, which are important for a rigorous nonasymptotic statistical study. We aim at obtaining estimates such as

$$\forall x > 0 \quad \mathbb{P}\big(\|\Theta_n - \Theta\| \geq x\big) \leq \mathrm{e}^{-C_n(x)},$$

where $C_n(x)$ will crucially depend on our set of assumptions. The upper bound in this inequality hold for arbitrary n and x (not a limit relation, unlike the MDP results), hence they are of much more practical use (in statistics). Deviation inequalities for estimators of the parameters associated with linear regression, autoregressive and branching processes were investigated by Bercu and Touati [4]. In the martingale case, deviation inequalities for a self normalized martingale have been developed by de la Peña et al. [7]. We also refer to the work of Ledoux [22] for precise credit and references. This type of inequalities is motivated by theoretical questions as well as numerous applications in different fields including the analysis of algorithms, mathematical physics and empirical processes. For some applications in nonasymptotic model selection problems we refer to Massart [23].

Let us emphasize that to our knowledge, there are no existing studies of the above questions, that is of the MDP and deviation inequalities for the least squares estimators of the unknown parameters of the general asymmetric BAR(p) process. These questions have been adressed recently by Bitseki Penda et al. [5], but for the BAR(1) processes. Moreover, in the latter, the authors have obtained their results under stronger assumptions than those made in this paper.

The main aspect of our contribution is that our results highlight the competition between the binary division and the speed of convergence in the MDP. Our MDP holds following three regimes, depending on the value of the ergodicity parameter of the BAR(p) compared with $1/2$. This new phenomenon is not seen in the case of the previously proved limit theorems: central limit theorem and law of large numbers. However, a similar phenomenon occurs for the central limit theorem of a branching particle system: see [1].

This paper is organized as follows. First of all, in Section 2, we introduce the BAR(p) model as well as the least squares estimators for the parameters of the observed BAR(p) process and some related notation and hypotheses. In Section 3, we state our main results on the deviation inequalities and MDP for our estimators. Section 4 is dedicated to the superexponential convergence of the quadratic variation of the martingale; this section contains exponential inequalities which are crucial for the proof of the deviation inequalities. The main results are proved in Section 5.

2. Notation and hypotheses

In all the sequel, let $p \in \mathbb{N}^*$. We consider the asymmetric BAR(p) process given, for all $n \geq 2^{p-1}$, by

$$\begin{cases} X_{2n} = a_0 + \sum_{k=1}^{p} a_k X_{[n/2^{k-1}]} + \varepsilon_{2n}, \\ X_{2n+1} = b_0 + \sum_{k=1}^{p} b_k X_{[n/2^{k-1}]} + \varepsilon_{2n+1}, \end{cases}$$

where the notation $[x]$ stands for the largest integer less than or equal to the real number 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 vectors.

For any matrix M the notation M^t, $\|M\|$ and $\mathrm{Tr}(M)$ stand for the transpose, the Euclidean norm and the trace of M respectively.

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

$$\begin{cases} \mathbb{X}_{2n} = A\mathbb{X}_n + \eta_{2n}, \\ \mathbb{X}_{2n+1} = B\mathbb{X}_n + \eta_{2n+1}, \end{cases} \tag{2.1}$$

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

$$A = \begin{pmatrix} a_1 & a_2 & \cdots & a_p \\ 1 & 0 & \cdots & 0 \\ 0 & \cdot & \cdot & \cdot \\ 0 & \cdot & 1 & \cdot \end{pmatrix} \quad \text{and} \quad B = \begin{pmatrix} b_1 & b_2 & \cdots & b_p \\ 1 & 0 & \cdots & 0 \\ 0 & \cdot & \cdot & \cdot \\ 0 & \cdot & 1 & \cdot \end{pmatrix}.$$

We shall assume that the matrices A and B satisfy the contraction property

$$\beta = \max\bigl(\|A\|, \|B\|\bigr) < 1. \tag{2.2}$$

One can view this BAR(p) process as a pth-order autoregressive process on a binary tree, where each vertex represents an individual or cell, vertex 1 being the original ancestor. For all $n \geq 1$, denote the nth generation by $\mathbb{G}_n = \{2^n, 2^n + 1, \dots, 2^{n+1} - 1\}$, see Figure 1.

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 the individual n is $[n/2]$. More generally, the ancestors of individual n are $[n/2], [n/2^2], \dots, [n/2^{r_n}]$. Furthermore, denote by

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

the subtree of all individuals from the original individual up to the nth generation. We denote by $\mathbb{T}_{n,p} = \{k \in \mathbb{T}_n, k \geq 2^p\}$ the subtree of all individuals between the pth and the nth generation ($\mathbb{T}_{p-1}$ removed). One can observe that, for all $n \geq 1$, $\mathbb{T}_{n,0} = \mathbb{T}_n$ and for all $p \geq 1$, $\mathbb{T}_{p,p} = \mathbb{G}_p$.

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

$$Z_n = \theta^t Y_n + V_n,$$

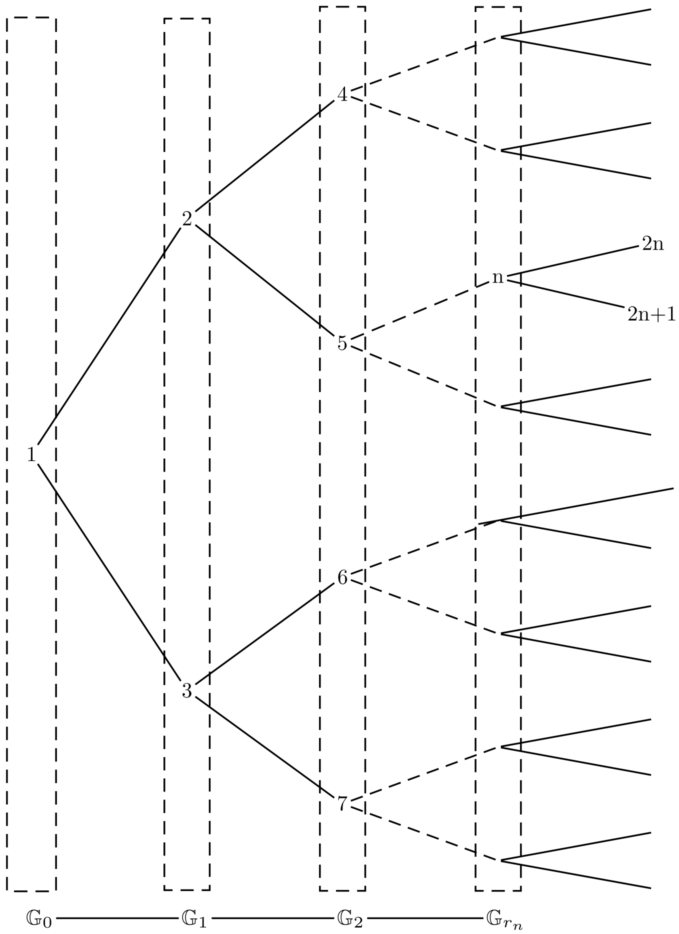

Fig. 1. The binary tree $\mathbb{T}$.

where

$$Z_n = \begin{pmatrix} X_{2n} \\ X_{2n+1} \end{pmatrix}, \qquad Y_n = \begin{pmatrix} 1 \\ \mathbb{X}_n \end{pmatrix}, \qquad V_n = \begin{pmatrix} \varepsilon_{2n} \\ \varepsilon_{2n+1} \end{pmatrix},$$

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

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

As in Bercu et al. [3], we introduce the least squares estimator $\hat{\theta}_n$ of θ for all $n \geq p$, from the observation of all individuals up to the nth generation (that is, the complete sub-tree $\mathbb{T}_n$)

$$\hat{\theta}_n = S_{n-1}^{-1} \sum_{k \in \mathbb{T}_{n-1,p-1}} Y_k Z_k^t, \tag{2.3}$$

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

$$S_n = \sum_{k\in\mathbb{T}_{n,p-1}} Y_k Y_k^t = \sum_{k\in\mathbb{T}_{n,p-1}} \begin{pmatrix} 1 & \mathbb{X}_k^t \\ \mathbb{X}_k & \mathbb{X}_k\mathbb{X}_k^t \end{pmatrix}. \tag{2.4}$$

We assume, without loss of generality, that for all $n \geq p-1$, S_n is invertible. From now on, we shall make a slight abuse of notation by identifying θ and $\hat{\theta}_n$ respectively to

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

Let $\Sigma_n = I_2 \otimes S_n$, where $\otimes$ stands for the matrix Kronecker product. We then deduce from (2.3) that

$$\hat{\theta}_n = \Sigma_{n-1}^{-1} \sum_{k\in\mathbb{T}_{n-1,p-1}} \operatorname{vec}\bigl(Y_k Z_k^t\bigr) = \Sigma_{n-1}^{-1} \sum_{k\in\mathbb{T}_{n-1,p-1}} \begin{pmatrix} X_{2k} \\ X_k\mathbb{X}_{2k} \\ X_{2k+1} \\ X_k\mathbb{X}_{2k+1} \end{pmatrix}.$$

Consequently, (2.1) yields

$$\hat{\theta}_n - \theta = \Sigma_{n-1}^{-1} \sum_{k\in\mathbb{T}_{n-1,p-1}} \begin{pmatrix} \varepsilon_{2k} \\ \varepsilon_{2k}\mathbb{X}_k \\ \varepsilon_{2k+1} \\ \varepsilon_{2k+1}\mathbb{X}_k \end{pmatrix}. \tag{2.5}$$

Denote by $\mathbb{F} = (\mathcal{F}_n)$ the natural filtration associated with the BAR(p) process, which means that $\mathcal{F}_n$ is the σ-algebra generated by the individuals up to the nth generation, in other words $\mathcal{F}_n = \sigma\{X_k, k \in \mathbb{T}_n\}$.

For the initial states, we set $\overline{X}_1 = \max\{\|\mathbb{X}_k\|, k \leq 2^{p-1}\}$ with the convention that $X_0 = 0$ and we introduce the following hypotheses:

(Xa) For some $a > 2$, there exists $\zeta > 0$ such that

$$\mathbb{E}\bigl[\exp\bigl(\zeta \overline{X}_1^a\bigr)\bigr] < \infty.$$

This assumption implies the weaker Gaussian integrability condition.

(X2) There is $\zeta > 0$ such that

$$\mathbb{E}\bigl[\exp\bigl(\zeta \overline{X}_1^2\bigr)\bigr] < \infty.$$

For the noise $(\varepsilon_{2n}, \varepsilon_{2n+1})$ the assumption may be of two types.

(1) In the first case we will assume the independence of the noise which allows us to impose less restrictive conditions on the exponential integrability of the noise.

Case 1: We shall assume that $((\varepsilon_{2n}, \varepsilon_{2n+1}), n \geq 1)$ forms a sequence of independent and identically distributed bi-variate centered random variables with covariance matrix Γ given by

$$\Gamma = \begin{pmatrix} \sigma^2 & \rho \\ \rho & \sigma^2 \end{pmatrix}, \quad \text{where } \sigma^2 > 0 \text{ and } |\rho| < \sigma^2. \tag{2.6}$$

For all $n \geq p-1$ and for all $k \in \mathbb{G}_n$, we set

$$\mathbb{E}\big[\varepsilon_k^2\big] = \sigma^2, \qquad \mathbb{E}\big[\varepsilon_k^4\big] = \tau^4, \qquad \mathbb{E}[\varepsilon_{2k}\varepsilon_{2k+1}] = \rho, \qquad \mathbb{E}\big[\varepsilon_{2k}^2\varepsilon_{2k+1}^2\big] = \nu^2, \quad \text{where } \tau^4 > 0, \nu^2 < \tau^4.$$

In addition, we assume that the condition (X2) on the initial state is satisfied and that

(G2) one can find $\gamma > 0$ and $c > 0$ such that for all $n \geq p-1$, for all $k \in \mathbb{G}_n$ and for all $|t| \leq c$

$$\mathbb{E}\big[\exp\big(t\big(\varepsilon_k^2 - \sigma^2\big)\big)\big] \leq \exp\bigg(\frac{\gamma t^2}{2}\bigg).$$

In this case, we impose the following hypotheses on the scale of the deviation

(V1) (v_n) will denote an increasing sequence of positive real numbers such that

$$v_n \longrightarrow +\infty$$

and for β given by (2.2)

- if $\beta \leq \frac{1}{2}$, the sequence (v_n) is such that $\frac{v_n \log n}{\sqrt{n}} \longrightarrow 0$,
- if $\beta > \frac{1}{2}$, the sequence (v_n) is such that $(v_n\sqrt{\log n})\beta^{(r_n+1)/2} \longrightarrow 0$.

(2) In contrast with the first case, in the second case we will not assume that the sequence $((\varepsilon_{2n}, \varepsilon_{2n+1}), n \geq 1)$ is i.i.d. The price to pay for giving up this i.i.d. assumption is to assume higher exponential moments. Indeed we need them to make use of the MDP for martingales, especially to prove the Lindeberg condition via the Lyapunov condition.

Case 2: We shall assume that for all $n \geq p-1$ and for all $j \in \mathbb{G}_{n+1}$ $\mathbb{E}[\varepsilon_j/\mathcal{F}_n] = 0$ and for all different $k, l \in \mathbb{G}_{n+1}$ with $[\frac{k}{2}] \neq [\frac{l}{2}]$, ε_k and ε_l are conditionally independent given $\mathcal{F}_n$. And we will use the same notation as in case 1: for all $n \geq p-1$ and for all $k \in \mathbb{G}_{n+1}$,

$$\mathbb{E}\big[\varepsilon_k^2/\mathcal{F}_n\big] = \sigma^2, \qquad \mathbb{E}\big[\varepsilon_k^4/\mathcal{F}_n\big] = \tau^4, \qquad \mathbb{E}[\varepsilon_{2k}\varepsilon_{2k+1}/\mathcal{F}_n] = \rho, \qquad \mathbb{E}\big[\varepsilon_{2k}^2\varepsilon_{2k+1}^2/\mathcal{F}_n\big] = \nu^2 \quad \text{a.s.}$$

where $\tau^4 > 0$, $\nu^2 < \tau^4$ and we use also Γ for the conditional covariance matrix associated with $(\varepsilon_{2n}, \varepsilon_{2n+1})$. In this case, we assume that the condition (Xa) on the initial state is satisfied, and we shall make the following hypotheses:

(Ea) for some $a > 2$, there exist $t > 0$ and $E > 0$ such that for all $n \geq p-1$ and for all $k \in \mathbb{G}_{n+1}$,

$$\mathbb{E}\big[\exp\big(t|\varepsilon_k|^{2a}\big)/\mathcal{F}_n\big] \leq E < \infty, \quad \text{a.s.}$$

Throughout this case, we introduce the following hypotheses on the scale of the deviation

(V2) (v_n) will denote an increasing sequence of positive real numbers such that

$$v_n \longrightarrow +\infty,$$

and for β given by (2.2)

- if $\beta^2 < \frac{1}{2}$, the sequence (v_n) is such that $\frac{v_n \log n}{\sqrt{n}} \longrightarrow 0$,
- if $\beta^2 = \frac{1}{2}$, the sequence (v_n) is such that $\frac{v_n (\log n)^{3/2}}{\sqrt{n}} \longrightarrow 0$,
- if $\beta^2 > \frac{1}{2}$, the sequence (v_n) is such that $(v_n \log n)\beta^{r_n+1} \longrightarrow 0$.

Remarks 2.1. *The condition on the scale of the deviation in case* 2, *is less restrictive than in case* 1, *since we assume a stronger integrability condition on the noise* (Ea). *This condition on the scale of the deviation naturally appears in the calculations. More precisely, the* log *term comes from the commutation of a probability and a sum.*

Remarks 2.2. *From* [14] *or* [22], *we deduce with* (Ea) *that*

(N1) *there is* $\phi > 0$ *such that for all* $n \geq p-1$, *for all* $k \in \mathbb{G}_{n+1}$ *and for all* $t \in \mathbb{R}$,

$$\mathbb{E}\big[\exp(t\varepsilon_k)/\mathcal{F}_n\big] < \exp\biggl(\frac{\phi t^2}{2}\biggr), \quad a.s.$$

We have the same conclusion in case 1, *without the conditioning*; *i.e.*

(G1) *there is* $\phi > 0$ *such that for all* $n \geq p-1$, *for all* $k \in \mathbb{G}_n$ *and for all* $t \in \mathbb{R}$,

$$\mathbb{E}\big[\exp(t\varepsilon_k)\big] < \exp\biggl(\frac{\phi t^2}{2}\biggr).$$

Remarks 2.3. *Armed with the recent development in the theory of transportation inequalities*, *exponential integrability and functional inequalities* (*see Ledoux* [22], *Gozlan* [16] *and Gozlan and Leonard* [17]), *we can prove that a sufficient condition for hypothesis* (G2) *to hold is the existence of* $t_0 > 0$ *such that for all* $n \geq p-1$ *and for all* $k \in \mathbb{G}_n$, $\mathbb{E}[\exp(t_0\varepsilon_k^2)] < \infty$.

We now turn to the estimation of the parameters σ^2 and ρ. On the one hand, we propose to estimate the conditional variance σ^2 by

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

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

$$\begin{cases} \hat{\varepsilon}_{2k} = X_{2k} - \hat{a}_{0,n} - \sum_{i=1}^p \hat{a}_{i,n} X_{[k/2^{i-1}]}, \\ \hat{\varepsilon}_{2k+1} = X_{2k+1} - \hat{b}_{0,n} - \sum_{i=1}^p \hat{b}_{i,n} X_{[k/2^{i-1}]}. \end{cases}$$

We also introduce

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

On the other hand, we estimate the conditional covariance ρ by

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

We also introduce

$$\rho_n = \frac{1}{|\mathbb{T}_{n-1}|} \sum_{k \in \mathbb{T}_{n-1,p}} \varepsilon_{2k}\varepsilon_{2k+1}.$$

In order to establish the MDP results of our estimators, we shall make use of a martingale approach. For all $n \geq p$, set

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

We can clearly rewrite (2.5) as

$$\hat{\theta}_n - \theta = \Sigma_{n-1}^{-1} M_n. \tag{2.7}$$

We know from Bercu et al. [3] that (M_n) is a square integrable martingale adapted to the filtration $\mathbb{F} = (\mathcal{F}_n)$. Its increasing process is given for all $n \geq p$ by

$$\langle M \rangle_n = \Gamma \otimes S_{n-1},$$

where S_n is given in (2.4) and Γ is given in (2.6).

Recall that for a sequence of random variables $(Z_n)_n$ on $\mathbb{R}^{d\times p}$, we say that $(Z_n)_n$ converges (v_n^2)-superexponentially fast in probability to some random variable Z if, for all $\delta > 0$,

$$\limsup_{n\to\infty} \frac{1}{v_n^2} \log \mathbb{P}\big(\|Z_n - Z\| > \delta\big) = -\infty.$$

This exponential convergence with speed v_n^2 will be abbreviated to

$$Z_n \underset{v_n^2}{\overset{\text{superexp}}{\Longrightarrow}}{}^{P} Z.$$

Remarks 2.4. *Note that for a determininistic sequence that converges to some limit ℓ, it also converges (v_n^2)-superexponentially fast to ℓ for any rate v_n.*

We follow Dembo and Zeitouni [12] for the language of the large deviations, throughout this paper. Before going further, let us recall the definition of a MDP: let (v_n) be an increasing sequence of positive real numbers such that

$$v_n \longrightarrow \infty \quad \text{and} \quad \frac{v_n}{\sqrt{n}} \longrightarrow 0. \tag{2.8}$$

We say that a sequence of centered random variables $(M_n)_n$ with topological state space $(S, \mathcal{S})$ satisfies a MDP with speed v_n^2 and rate function $I : S \to \mathbb{R}_+^*$ if for each $A \in \mathcal{S}$,

$$-\inf_{x\in A^o} I(x) \leq \liminf_{n\to\infty} \frac{1}{v_n^2} \log \mathbb{P}\left(\frac{\sqrt{n}}{v_n} M_n \in A\right) \leq \limsup_{n\to\infty} \frac{1}{v_n^2} \log \mathbb{P}\left(\frac{\sqrt{n}}{v_n} M_n \in A\right) \leq -\inf_{x\in \overline{A}} I(x),$$

where A^o and $\overline{A}$ denote the interior and closure of A respectively.

Before we present the main results, let us fix some more notation. Let

$$\overline{a} = \frac{a_0 + b_0}{2}, \qquad \overline{a^2} = \frac{a_0^2 + b_0^2}{2}, \qquad \overline{A} = \frac{A + B}{2}.$$

We set

$$\Xi = \overline{a}(I_p - \overline{A})^{-1} e_1, \tag{2.9}$$

and let Λ be the unique solution of the equation (see Lemma A.4 in [3])

$$\Lambda = T + \frac{1}{2}\big(A\Lambda A^t + B\Lambda B^t\big) \tag{2.10}$$

where

$$T = \big(\sigma^2 + \overline{a^2}\big) e_1 e_1^t + \frac{1}{2}\big(a_0\big(A\Xi e_1^t + e_1 \Xi^t A^t\big) + b_0\big(B\Xi e_1^t + e_1 \Xi^t B^t\big)\big). \tag{2.11}$$

We also introduce the following matrices L and Σ given by

$$L = \begin{pmatrix} 1 & \Xi \\ \Xi & \Lambda \end{pmatrix} \quad \text{and} \quad \Sigma = I_2 \otimes L. \tag{2.12}$$

3. Main results

Let us present now the main results of this paper. In the following theorem, we give the deviation inequalities of the estimator of the parameters.

Theorem 3.1.

(i) *In case* 1, *we have for all* $\delta > 0$ *and for all* $\ell > 0$ *such that* $\ell < \|\Sigma\|/(1+\delta)$

$$\mathbb{P}\big(\|\hat{\theta}_n - \theta\| > \delta\big) \leq \begin{cases} c_1 \exp\big(-\frac{c_2(\delta\ell)^2}{c_3+(\delta\ell)} \frac{2^n}{(n-1)^2}\big) & \text{if } \beta < \frac{1}{2}, \\ c_1(n-1)\exp\big(\frac{-c_2(\delta\ell)^2}{c_3+(\delta\ell)} \frac{2^n}{(n-1)^2}\big) & \text{if } \beta = \frac{1}{2}, \\ c_1(n-1)\exp\big(\frac{-c_2(\delta\ell)^2}{c_3+(\delta\ell)} \frac{1}{(n-1)\beta^n}\big) & \text{if } \beta > \frac{1}{2}, \end{cases} \tag{3.1}$$

where the constants c_1, c_2 *and* c_3 *depend on* σ^2, β, γ *and* ϕ, *may differ line by line and are such that* $c_1, c_2 > 0$, $c_3 \geq 0$.

(ii) *In case* 2, *we have for all* $\delta > 0$ *and for all* $\ell > 0$ *such that* $\ell < \|\Sigma\|/(1+\delta)$

$$\mathbb{P}\big(\|\hat{\theta}_n - \theta\| > \delta\big) \leq \begin{cases} c_1 \exp\big(-\frac{c_2(\delta\ell)^2}{c_3+c_4(\delta\ell)} \frac{2^n}{(n-1)^2}\big) & \text{if } \beta < \frac{\sqrt{2}}{2}, \\ c_1 \exp\big(-\frac{c_2(\delta\ell)^2}{c_3+c_4(\delta\ell)} \frac{2^n}{(n-1)^3}\big) & \text{if } \beta = \frac{\sqrt{2}}{2}, \\ c_1 \exp\big(-\frac{c_2(\delta\ell)^2}{c_3+c_4(\delta\ell)} \frac{1}{(n-1)^2\beta^{2n}}\big) & \text{if } \beta > \frac{\sqrt{2}}{2}, \end{cases} \tag{3.2}$$

where the constants c_1, c_2, c_3, *and* c_4 *depend on* σ^2, β, γ *and* ϕ, *may differ line by line and are such that* $c_1, c_2 > 0$, $c_3, c_4 \geq 0$, $(c_3, c_4) \neq (0,0)$.

Remarks 3.2. *Note that the estimate* (3.2) *is stronger than the estimate* (3.1). *This is due to the fact that the integrability condition* (Ea) *in case* 2 *is stronger than the integrability condition* (G2) *in case* 1.

Remarks 3.3. *Let us stress that by tedious but straightforward calculations, the constants which appear in the previous theorem can be well estimated.*

Remarks 3.4. *The upper bounds in previous theorem hold for arbitrary* $n \geq p-1$ (*not a limit relation, unlike the results below*), *hence they are very practical* (*in nonasymptotic statistics*) *when sample size does not allow the application of limit theorems.*

In the next result, we present the MDP of the estimator $\hat{\theta}_n$.

Theorem 3.5. *In case* 1 *or in case* 2, *the sequence* $(\sqrt{|\mathbb{T}_{n-1}|}(\hat{\theta}_n - \theta)/v_{|\mathbb{T}_{n-1}|})_{n\geq 1}$ *satisfies the MDP on* $\mathbb{R}^{2(p+1)}$ *with speed* $v^2_{|\mathbb{T}_{n-1}|}$ *and rate function*

$$I_\theta(x) = \sup_{\lambda \in \mathbb{R}^{2(p+1)}} \big\{\lambda^t x - \lambda\big(\Gamma \otimes L^{-1}\big)\lambda^t\big\} = \frac{1}{2} x^t \big(\Gamma \otimes L^{-1}\big)^{-1} x, \tag{3.3}$$

where L *and* Γ *are given in* (2.12) *and* (2.6) *respectively.*

Remarks 3.6. *Similar results about deviation inequalities and MDP have already been obtained in* [5], *in a restrictive case of bounded or Gaussian noise and when* $p = 1$, *but results therein also hold for general Markov models. Moreover in* [5], *when the noise is Gaussian, the range of speed of MDP is very restricted in comparison to the range of speed of MDP in case* 1 *of this paper. These improvements are due to the fact that in this paper, we take advantage of the autoregressive structure of the process while in* [5], *only its Markovian nature is used.*

Let us also mention that in case 2, *the Markovian nature of* BAR(p) *processes is lost and this case is not studied in* [5]. *However in case* 2, *for* $p=1$, *if we assume that the initial state* X_1 *and the noise take their values in a compact set*, *we can find the same results as in* [5]. *The results of this paper then allow to extend the results of the latter paper.*

Let us consider now the estimation of the parameter in the noise process.

Theorem 3.7. *Let* (v_n) *an increasing sequence of positive real numbers such that*

$$v_n \longrightarrow \infty \quad \textit{and} \quad \frac{v_n}{\sqrt{n}} \longrightarrow 0.$$

In case 1 *or in case* 2,

(1) *the sequence* $(\sqrt{|\mathbb{T}_{n-1}|}(\sigma_n^2-\sigma^2)/v_{|\mathbb{T}_{n-1}|})_{n\geq 1}$ *satisfies the MDP on* $\mathbb{R}$ *with speed* $v^2_{|\mathbb{T}_{n-1}|}$ *and rate function*

$$I_{\sigma^2}(x)=\frac{x^2}{\tau^4-2\sigma^4+\nu^2}; \tag{3.4}$$

(2) *the sequence* $(\sqrt{|\mathbb{T}_{n-1}|}(\rho_n-\rho)/v_{|\mathbb{T}_{n-1}|})_{n\geq 1}$ *satisfies the MDP on* $\mathbb{R}$ *with speed* $v^2_{|\mathbb{T}_{n-1}|}$ *and rate function*

$$I_{\rho}(x)=\frac{x^2}{2(\nu^2-\rho^2)}. \tag{3.5}$$

Remarks 3.8. *Note that in this case the MDP holds for all the scales* (v_n) *verifying* (2.8) *without other restriction.*

Remarks 3.9. *It would be more interesting to prove the MDP for* $(\sqrt{|\mathbb{T}_{n-1}|}(\hat{\sigma}_n^2-\sigma^2)/v_{|\mathbb{T}_{n-1}|})_{n\geq 1}$, *which will be the case if one proves for example that* $(\sqrt{|\mathbb{T}_{n-1}|}(\hat{\sigma}_n^2-\sigma^2)/v_{|\mathbb{T}_{n-1}|})_{n\geq 1}$ *and* $(\sqrt{|\mathbb{T}_{n-1}|}(\sigma_n^2-\sigma^2)/v_{|\mathbb{T}_{n-1}|})_{n\geq 1}$ *are exponentially equivalent in the sense of the MDP. This is described by the following convergence*

$$\frac{\sqrt{|\mathbb{T}_{n-1}|}}{v_{|\mathbb{T}_{n-1}|}}\left(\hat{\sigma}_n^2-\sigma_n^2\right) \underset{v^2_{|\mathbb{T}_{n-1}|}}{\overset{\text{superexp}}{\Longrightarrow}} 0.$$

The proof is very technical and very restrictive with respect to the scale (v_n) *of the deviation. Actually we are only able to prove that*

$$\hat{\sigma}_n^2-\sigma_n^2 \underset{v^2_{|\mathbb{T}_{n-1}|}}{\overset{\text{superexp}}{\Longrightarrow}} 0.$$

This superexponential convergence will be proved in Theorem 3.10.

In the following theorem we state the superexponential convergence.

Theorem 3.10. *In case* 1 *or in case* 2, *we have*

$$\hat{\sigma}_n^2 \underset{v^2_{|\mathbb{T}_{n-1}|}}{\overset{\text{superexp}}{\Longrightarrow}} \sigma^2.$$

In case 1, *if instead of* (G2), *we assume that*

(G2$'$) *one can find* $\gamma'>0$ *such that for all* $n\geq p-1$, *for all* $k,l\in\mathbb{G}_{n+1}$ *with* $[\frac{k}{2}]=[\frac{l}{2}]$ *and for all* $t\in]-c,c[$ *for some* $c>0$,

$$\mathbb{E}\left[\exp t(\varepsilon_k\varepsilon_l-\rho)\right]\leq \exp\left(\frac{\gamma' t^2}{2}\right),$$

and in case 2, *if instead of* (Ea), *we assume that*

(E2′) *one can find* $\gamma' > 0$ *such that for all* $n \geq p-1$, *for all* $k, l \in \mathbb{G}_{n+1}$ *with* $[\frac{k}{2}] = [\frac{l}{2}]$ *and for all* $t \in \mathbb{R}$

$$\mathbb{E}\bigl[\exp t(\varepsilon_k \varepsilon_l - \rho)/\mathcal{F}_n\bigr] \leq \exp\biggl(\frac{\gamma' t^2}{2}\biggr), \quad a.s.$$

Then in case 1 *or in case* 2, *we have*

$$\hat{\rho}_n \underset{v^2_{|\mathbb{T}_{n-1}|}}{\overset{\text{superexp}}{\Longrightarrow}} \rho.$$

Before going into the proofs, let us gather here for the convenience of the reader two theorems useful to establish MDP for martingales and used intensively in this paper. From these two theorems, we will be able to give a strategy for the proof.

The following proposition corresponds to the unidimensional case of Theorem 1 in [13].

Proposition 3.11. *Let* $M = (M_n, \mathcal{H}_n, n \geq 0)$ *be a centered square real valued integrable martingale defined on a probability space* $(\Omega, \mathcal{H}, \mathbb{P})$ *and let* $(\langle M\rangle_n)$ *be its bracket. Let* (v_n) *be an increasing sequence of real numbers satisfying* (2.8).

Let $c(n) := \frac{\sqrt{n}}{v_n}$ *be nondecreasing, and define the reciprocal function* $c^{-1}(t)$ *by*

$$c^{-1}(t) := \inf\bigl\{n \in \mathbb{N}: \ c(n) \geq t\bigr\}.$$

Under the following conditions

(D1) *there exists* $Q \in \mathbb{R}^*_+$ *such that* $\frac{\langle M\rangle_n}{n} \underset{v_n^2}{\overset{\text{superexp}}{\Longrightarrow}} Q$;

(D2) $\limsup_{n\to+\infty} \frac{n}{v_n^2} \log(n \operatorname{ess\,sup}_{1\leq k\leq c^{-1}(\sqrt{n+1}v_{n+1})} \mathbb{P}(|M_k - M_{k-1}| > v_n\sqrt{n}/\mathcal{H}_{k-1})) = -\infty$;

(D3) *for all* $a > 0$ $\frac{1}{n}\sum_{k=1}^n \mathbb{E}(|M_k - M_{k-1}|^2 \mathbf{1}_{\{|M_k - M_{k-1}| \geq a(\sqrt{n}/v_n)\}}/\mathcal{H}_{k-1}) \underset{v_n^2}{\overset{\text{superexp}}{\Longrightarrow}} 0$;

$(M_n/v_n\sqrt{n})_{n\geq 0}$ *satisfies the MDP in* $\mathbb{R}$ *with speed* v_n^2 *and rate function* $I(x) = \frac{x^2}{2Q}$.

Let us introduce a simplified version of Puhalskii's result [24] applied to a sequence of martingale differences.

Theorem 3.12. *Let* $(m_j^n)_{1\leq j\leq n}$ *be a triangular array of martingale differences with values in* $\mathbb{R}^d$, *with respect to some filtration* $(\mathcal{H}_n)_{n\geq 1}$. *Let* (v_n) *be an increasing sequence of real numbers satisfying* (2.8). *Under the following conditions*

(P1) *there exists a symmetric positive semi-definite matrix* Q *such that*

$$\frac{1}{n}\sum_{k=1}^n \mathbb{E}\bigl[m_k^n (m_k^n)^t | \mathcal{H}_{k-1}\bigr] \underset{v_n^2}{\overset{\text{superexp}}{\Longrightarrow}} Q,$$

(P2) *there exists a constant* $c > 0$ *such that, for each* $1 \leq k \leq n$, $|m_k^n| \leq c\frac{\sqrt{n}}{v_n}$ *a.s.*,

(P3) *for all* $a > 0$, *we have the exponential Lindeberg's condition*

$$\frac{1}{n}\sum_{k=1}^n \mathbb{E}\bigl[|m_k^n|^2 \mathbf{1}_{\{|m_k^n| \geq a(\sqrt{n}/v_n)\}} | \mathcal{H}_{k-1}\bigr] \underset{v_n^2}{\overset{\text{superexp}}{\Longrightarrow}} 0,$$

$(\sum_{k=1}^n m_k^n/(v_n\sqrt{n}))_{n\geq 1}$ *satisfies an MDP on* $\mathbb{R}^d$ *with speed* v_n^2 *and rate function*

$$\Lambda^*(v) = \sup_{\lambda\in\mathbb{R}^d}\biggl(\lambda^t v - \frac{1}{2}\lambda^t Q\lambda\biggr).$$

In particular, if Q is invertible, $\Lambda^*(v) = \frac{1}{2} v^t Q^{-1} v$.

As the reader can imagine naturally now, the strategy of the proof of the MDP consists in the following steps:

- the superexponential convergence of the quadratic variation of the martingale (M_n). This step is very crucial and the key for the rest of the paper. It will be realized by means of powerful exponential inequalities. This allows us to obtain the deviation inequalities for the estimator of the parameters,
- introduce a truncated martingale which satisfies the MDP, thanks to the classical Theorem 3.12,
- the truncated martingale is an exponentially good approximation of (M_n), in the sense of the moderate deviation.

4. Superexponential convergence of the quadratic variation of the martingale

First, it is necessary to establish the superexponential convergence of the quadratic variation of the martingale (M_n), properly normalized in order to prove the MDP of the estimators. Its proof is very technical, but crucial for the rest of the paper. This section contains also some deviation inequalities for some quantities needed in the proof later.

Proposition 4.1. *In case* 1 *or case* 2, *we have*

$$\frac{S_n}{|\mathbb{T}_n|} \underset{v^2_{|\mathbb{T}_n|}}{\overset{\text{superexp}}{\Longrightarrow}} L, \tag{4.1}$$

where S_n is given in (2.4) *and L is given in* (2.12).

For the proof we focus on case 2. Proposition 4.1 will follow from Proposition 4.3 and Proposition 4.9 below, where we assume that the sequence (v_n) satisfies the condition (V2). Proposition 4.10 gives some ideas of the proof in case 1.

Remarks 4.2. *Using* [14], *we infer from* (Ea) *that*

(N2) *one can find $\gamma > 0$ such that for all $n \geq p-1$, for all $k \in \mathbb{G}_{n+1}$ and for all $t \in \mathbb{R}$*

$$\mathbb{E}\big[\exp t\big(\varepsilon_k^2 - \sigma^2\big)/\mathcal{F}_n\big] \leq \exp\left(\frac{\gamma t^2}{2}\right) \quad a.s.$$

Proposition 4.3. *Assume that hypotheses* (N2) *and* (Xa) *are satisfied. Then we have*

$$\frac{1}{|\mathbb{T}_n|} \sum_{k \in \mathbb{T}_{n,p}} \mathbb{X}_k \mathbb{X}_k^t \underset{v^2_{|\mathbb{T}_n|}}{\overset{\text{superexp}}{\Longrightarrow}} \Lambda,$$

where Λ is given in (2.10).

Proof. Let

$$K_n = \sum_{k \in \mathbb{T}_{n,p-1}} \mathbb{X}_k \mathbb{X}_k^t \quad \text{and} \quad L_n = \sum_{k \in \mathbb{T}_{n,p}} \varepsilon_k^2. \tag{4.2}$$

Then from (2.1), and after straightforward calculations (see p. 2519 in [3] for more details), we get that

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

where the notation $\{A; B\}^k$ means the set of all products of A and B with exactly k terms. The cardinality of $\{A; B\}^k$ is obviously 2^k, and

$$T_k = \frac{L_k}{2^{k+1}} e_1 e_1^t + \overline{a^2}\left(\frac{2^k - 2^{p-1}}{2^k}\right) e_1 e_1^t + I_k^{(1)} + I_k^{(2)} + \frac{1}{2^{k+1}} U_k$$

with $\overline{a^2} = (a_0^2 + b_0^2)/2$ and

$$I_k^{(1)} = \frac{1}{2}\left(a_0\left(A\frac{H_{k-1}}{2^k}e_1^t + e_1\frac{H_{k-1}}{2^k}A^t\right) + b_0\left(B\frac{H_{k-1}}{2^k}e_1^t + e_1\frac{H_{k-1}}{2^k}B^t\right)\right), \tag{4.3}$$

$$I_k^{(2)} = \left(\frac{1}{2^k}\sum_{l \in \mathbb{T}_{k-1,p-1}} (a_0\varepsilon_{2l} + b_0\varepsilon_{2l+1})\right) e_1 e_1^t, \tag{4.4}$$

$$U_k = \sum_{l \in \mathbb{T}_{k-1,p-1}} \varepsilon_{2l}\left(A\mathbb{X}_l e_1^t + e_1 \mathbb{X}_l^t A^t\right) + \varepsilon_{2l+1}\left(B\mathbb{X}_l e_1^t + e_1\mathbb{X}_l^t B^t\right). \tag{4.5}$$

Then the proposition will follow if we prove Lemmas 4.4, 4.6, 4.7, 4.8 and 4.5. □

Lemma 4.4. *Assume that hypothesis* (Xa) *is satisfied. Then we have*

$$\frac{1}{2^{n-p+1}} \sum_{C \in \{A;B\}^{n-p+1}} C\frac{K_{p-1}}{2^p}C^t \underset{v^2_{|\mathbb{T}_n|}}{\overset{\text{superexp}}{\Longrightarrow}} 0, \tag{4.6}$$

where K_p is given in (4.2).

Proof. We get easily

$$\left\| \frac{1}{2^{n-p+1}} \sum_{C \in \{A;B\}^{n-p+1}} C\frac{K_{p-1}}{2^p}C^t \right\| \le c\beta^{2n}\overline{X}_1^2,$$

where β is given in (2.2), $\overline{X}_1$ is introduced in (Xa) and c is a positive constant which depends on p. Next, Chernoff inequality and hypothesis (X2) lead us easily to (4.6). □

Lemma 4.5. *Assume that hypotheses* (N2) *and* (Xa) *are satisfied. Then we have*

$$\sum_{k=0}^{n-p} \frac{1}{2^k} \sum_{C \in \{A;B\}^k} C\frac{U_{n-k}}{2^{n-k+1}}C^t \underset{v^2_{|\mathbb{T}_n|}}{\overset{\text{superexp}}{\Longrightarrow}} 0, \tag{4.7}$$

where U_k is given by (4.5).

Proof. Let $V_n = \sum_{k \in \mathbb{T}_{n-1,p-1}} \varepsilon_{2k} X_k$. Then (V_n) is an $\mathcal{F}_n$-martingale and its increasing process satisfies, for all $n \ge p$,

$$\langle V\rangle_n = \sigma^2 \sum_{k \in \mathbb{T}_{n-1,p}} X_k^2 \le \sigma^2 \sum_{k \in \mathbb{T}_{n-1,p-1}} X_k^2 \le \sigma^2 \sum_{k \in \mathbb{T}_{n-1,p-1}} \|\mathbb{X}_k\|^2.$$

For $\lambda > 0$, we infer from hypothesis (N1) that $(Y_k)_{p \le k \le n}$ given by

$$Y_n = \exp\left(\lambda V_n - \frac{\lambda^2 \phi}{2} \sum_{k \in \mathbb{T}_{n-1,p-1}} X_k^2\right),$$

is an $\mathcal{F}_k$-supermartingale and moreover $\mathbb{E}[Y_p] \le 1$. For $B > 0$ and $\delta > 0$, we have

$$\mathbb{P}\left(\frac{V_n}{|\mathbb{T}_n|+1} > \delta\right) \le \mathbb{P}\left(\frac{\phi}{|\mathbb{T}_n|+1}\sum_{k\in\mathbb{T}_{n-1,p-1}} X_k^2 > B\right) + \mathbb{P}\left(Y_n > \exp\left(\lambda\delta - \frac{\lambda^2 B}{2}\right)2^{n+1}\right)$$
$$\le \mathbb{P}\left(\frac{\phi}{|\mathbb{T}_n|+1}\sum_{k\in\mathbb{T}_{n-1,p-1}} X_k^2 > B\right) + \exp\left(\left(-\lambda\delta + \frac{\lambda^2 B}{2}\right)2^{n+1}\right).$$

Optimizing on λ, we get

$$\mathbb{P}\left(\frac{V_n}{|\mathbb{T}_n|+1} > \delta\right) \le \mathbb{P}\left(\frac{\phi}{|\mathbb{T}_n|+1}\sum_{k\in\mathbb{T}_{n-1,p-1}} X_k^2 > B\right) + \exp\left(-\frac{\delta^2}{B}2^{n+1}\right).$$

Since the same thing works for $-V_n$ instead of V_n and using the following inequality,

$$\sum_{k\in\mathbb{T}_{n-1,p-1}} X_k^2 \le \sum_{k\in\mathbb{T}_{n-1,p-1}} \|\mathbb{X}_k\|^2,$$

we get

$$\mathbb{P}\left(\frac{|V_n|}{|\mathbb{T}_n|+1} > \delta\right) \le \mathbb{P}\left(\frac{\phi}{|\mathbb{T}_n|+1}\sum_{k\in\mathbb{T}_{n-1,p-1}} \|\mathbb{X}_k\|^2 > B\right) + \exp\left(-\frac{\delta^2}{B}2^{n+1}\right). \tag{4.8}$$

From [3], with $\alpha = \max(|a_0|, |b_0|)$, we have

$$\sum_{k\in\mathbb{T}_{n-1,p-1}} \|\mathbb{X}_k\|^2 \le \frac{4}{1-\beta}P_{n-1} + \frac{4\alpha^2}{1-\beta}Q_{n-1} + 2\overline{X}_1^2 R_{n-1}, \tag{4.9}$$

where

$$P_n = \sum_{k\in\mathbb{T}_{n,p}}\sum_{i=0}^{r_k-p}\beta^i\varepsilon_{[k/2^i]}^2, \qquad Q_n = \sum_{k\in\mathbb{T}_{n,p}}\sum_{i=0}^{r_k-p}\beta^i, \qquad R_n = \sum_{k\in\mathbb{T}_{n,p-1}}\beta^{2(r_k-p+1)}.$$

Now, to control the first term in the right hand side of (4.8), we will use the decomposition given by (4.9). From the convergence of $\frac{4\phi}{(1-\beta)(|\mathbb{T}_n|+1)}P_n$ and $\frac{4\phi\alpha^2}{(1-\beta)(|\mathbb{T}_n|+1)}Q_n$ (see [3] for more details) let l_1 and l_2 be such that

$$\frac{4\phi P_{n-1}}{(1-\beta)(|\mathbb{T}_n|+1)} \to l_1 \quad \text{and} \quad \forall n \ge p-1 \quad \frac{4\phi\alpha^2 Q_{n-1}}{(1-\beta)(|\mathbb{T}_n|+1)} < l_2.$$

For $\delta > 0$, we choose $B = \delta + l_1 + l_2$, using (4.9), we then have

$$\mathbb{P}\left(\frac{\phi}{|\mathbb{T}_n|+1}\sum_{k\in\mathbb{T}_{n-1,p-1}} \|\mathbb{X}_k\|^2 > B\right)$$
$$\le \mathbb{P}\left(\frac{P_{n-1}}{|\mathbb{T}_n|+1} - l_1' > \delta_1\right) + \mathbb{P}\left(\frac{Q_{n-1}}{|\mathbb{T}_n|+1} - l_2' > \delta_2\right) + \mathbb{P}\left(\frac{R_{n-1}\overline{X}_1^2}{|\mathbb{T}_n|+1} > \delta_3\right), \tag{4.10}$$

where

$$\delta_1 = \frac{(1-\beta)\delta}{12\phi}, \qquad l_1' = \frac{(1-\beta)l_1}{4\phi}, \qquad \delta_2 = \frac{(1-\beta)\delta}{12\alpha^2\phi}, \qquad l_2' = \frac{(1-\beta)l_2}{4\alpha^2\phi} \quad \text{and} \quad \delta_3 = \frac{\delta}{6\phi}.$$

First, by the choice of l_2, we have

$$\mathbb{P}\left(\frac{Q_{n-1}}{|\mathbb{T}_n|+1} - l_2' > \delta_2\right) = 0. \tag{4.11}$$

Next, from Chernoff inequality and hypothesis (X2) we get easily

$$\mathbb{P}\left(\frac{R_{n-1}\overline{X}_1^2}{|\mathbb{T}_n|+1} > \delta_3\right) \leq \begin{cases} c_1 \exp\left(-c_2\delta 2^{n+1}\right) & \text{if } \beta < \frac{\sqrt{2}}{2}, \\ c_1 \exp\left(-c_2\delta \frac{2^{n+1}}{n+1}\right) & \text{if } \beta = \frac{\sqrt{2}}{2}, \\ c_1 \exp\left(-c_2\delta \left(\frac{1}{\beta^2}\right)^{n+1}\right) & \text{if } \beta > \frac{\sqrt{2}}{2} \end{cases} \tag{4.12}$$

for some positive constants c_1 and c_2. Let us now control the first term of the right hand side of (4.10).

First case. If $\beta = \frac{1}{2}$, from [3]

$$P_{n-1} = \sum_{k=p}^{n-1}(n-k)\sum_{i\in\mathbb{G}_k}\varepsilon_i^2 \quad \text{and} \quad l_1' = \sigma^2.$$

We thus have

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

In addition, we also have

$$\sigma^2\left(\sum_{k=p}^{n-1}\frac{n-k}{2^{n+1-k}} - 1\right) \leq 0.$$

We thus deduce that

$$\mathbb{P}\left(\frac{P_{n-1}}{|\mathbb{T}_n|+1} - l_1' > \delta_1\right) \leq \mathbb{P}\left(\frac{1}{|\mathbb{T}_n|+1}\sum_{k=p}^{n-1}(n-k)\sum_{i\in\mathbb{G}_k}\left(\varepsilon_i^2 - \sigma^2\right) > \delta_1\right).$$

On the one hand we have

$$\begin{aligned} &\mathbb{P}\left(\frac{1}{|\mathbb{T}_n|+1}\sum_{k=p}^{n-1}(n-k)\sum_{i\in\mathbb{G}_k}\left(\varepsilon_i^2 - \sigma^2\right) > \delta_1\right) \\ &\quad \leq \sum_{\eta=0}^{1}\mathbb{P}\left(\frac{1}{|\mathbb{T}_n|+1}\sum_{k=p-1}^{n-2}(n-k-1)\sum_{i\in\mathbb{G}_k}\left(\varepsilon_{2i+\eta}^2 - \sigma^2\right) > \delta_1/2\right). \end{aligned} \tag{4.13}$$

On the other hand, for all $\lambda > 0$, an application of Chernoff inequality yields

$$\begin{aligned} &\mathbb{P}\left(\frac{1}{|\mathbb{T}_n|+1}\sum_{k=p-1}^{n-2}(n-k-1)\sum_{i\in\mathbb{G}_k}\left(\varepsilon_{2i}^2 - \sigma^2\right) > \delta_1/2\right) \\ &\quad \leq \exp\left(\frac{-\delta_1\lambda 2^{n+1}}{2}\right) \times \mathbb{E}\left[\exp\left(\lambda\sum_{k=p-1}^{n-2}(n-k-1)\sum_{i\in\mathbb{G}_k}\left(\varepsilon_{2i}^2 - \sigma^2\right)\right)\right]. \end{aligned}$$

From hypothesis (N2) we get

$$\begin{aligned}
&\mathbb{E}\left[\exp\left(\lambda \sum_{k=p-1}^{n-2} (n-k-1) \sum_{i\in\mathbb{G}_k} \left(\varepsilon_{2i}^2-\sigma^2\right)\right)\right] \\
&\quad = \mathbb{E}\left[\mathbb{E}\left[\exp\left(\lambda \sum_{k=p-1}^{n-2} (n-k-1) \sum_{i\in\mathbb{G}_k} \left(\varepsilon_{2i}^2-\sigma^2\right)\right)\Big/\mathcal{F}_n\right]\right] \\
&\quad = \mathbb{E}\left[\exp\left(\lambda \sum_{k=p-1}^{n-3} (n-k-1) \sum_{i\in\mathbb{G}_k} \left(\varepsilon_{2i}^2-\sigma^2\right)\right) \prod_{i\in\mathbb{G}_{n-2}} \mathbb{E}\left[\exp\left(\lambda\left(\varepsilon_{2i}^2-\sigma^2\right)\right)/\mathcal{F}_n\right]\right] \\
&\quad \le \exp\left(\lambda^2\gamma|\mathbb{G}_{n-2}|\right)\mathbb{E}\left[\exp\left(\lambda \sum_{k=p-1}^{n-3} (n-k-1) \sum_{i\in\mathbb{G}_k} \left(\varepsilon_{2i}^2-\sigma^2\right)\right)\right].
\end{aligned}$$

Iterating this procedure, we obtain

$$\begin{aligned}
\mathbb{E}\left[\exp\left(\lambda \sum_{k=p-1}^{n-2} (n-k-1) \sum_{i\in\mathbb{G}_k} \left(\varepsilon_{2i}^2-\sigma^2\right)\right)\right] &\le \exp\left(\gamma\lambda^2 \sum_{k=2}^{n-p+1} k^2|\mathbb{G}_{n-k}|\right) \\
&\le \exp\left(c\gamma\lambda^2 2^{n+1}\right),
\end{aligned}$$

where $c = \sum_{k=1}^{\infty} \frac{k^2}{2^{k+2}} = \frac{3}{4}$. Optimizing on λ, we are led, for some positive constant c_1 to

$$\mathbb{P}\left(\frac{1}{|\mathbb{T}_n|+1} \sum_{k=p-1}^{n-2} (n-k-1) \sum_{i\in\mathbb{G}_k} \left(\varepsilon_{2i}^2-\sigma^2\right) > \delta_1/2\right) \le \exp\left(-c_1\delta^2|\mathbb{T}_n|\right).$$

Following the same lines, we obtain the same inequality for the second term in (4.13). It then follows that

$$\mathbb{P}\left(\frac{P_{n-1}}{|\mathbb{T}_n|+1} - l_1' > \delta_1\right) \le c_1 \exp\left(-c_2\delta^2|\mathbb{T}_n|\right) \tag{4.14}$$

for some positive constants c_1 and c_2.

Second case. If $\beta \neq \frac{1}{2}$, then from [3], we have $l_1' = \frac{\sigma^2}{2(1-\beta)}$. Since

$$\sigma^2\left(\sum_{k=p}^{n-1} \frac{1-(2\beta)^{n-k}}{(1-2\beta)2^{n-k+1}}\right) \le \frac{\sigma^2}{2(1-\beta)},$$

we deduce that

$$\mathbb{P}\left(\frac{P_{n-1}}{|\mathbb{T}_n|+1} - l_1' > \delta_1\right) \le \mathbb{P}\left(\frac{1}{|\mathbb{T}_n|+1} \sum_{k=p}^{n-1} \frac{1-(2\beta)^{n-k}}{1-2\beta} \sum_{i\in\mathbb{G}_k} \left(\varepsilon_i^2-\sigma^2\right) > \delta_1\right).$$

- If $\beta < \frac{1}{2}$, then for some positive constant c, we have

$$\mathbb{P}\left(\frac{P_{n-1}}{|\mathbb{T}_n|+1} - l_1' > \delta_1\right) \le \mathbb{P}\left(\frac{1}{|\mathbb{T}_n|+1} \sum_{k=p}^{n-1} \sum_{i\in\mathbb{G}_k} \left(\varepsilon_i^2-\sigma^2\right) > c\delta_1\right).$$

Proceeding now as in the proof of (4.21), we get

$$\mathbb{P}\left(\frac{P_{n-1}}{|\mathbb{T}_n|+1} - l_1' > \delta_1\right) \leq c_1 \exp\bigl(-c_2\delta^2|\mathbb{T}_n|\bigr) \tag{4.15}$$

for some positive constants c_1 and c_2.

- If $\beta > \frac{1}{2}$, then for some positive constant c, we have

$$\mathbb{P}\left(\frac{P_{n-1}}{|\mathbb{T}_n|+1} - l_1' > \delta_1\right) \leq \mathbb{P}\left(\frac{1}{|\mathbb{T}_n|+1}\sum_{k=p}^{n-1}(2\beta)^{n-k}\sum_{i\in\mathbb{G}_k}\bigl(\varepsilon_i^2-\sigma^2\bigr) > c\delta_1\right).$$

Now, from Chernoff inequality, hypothesis (N2) and after several successive conditioning, we get for all $\lambda > 0$

$$\begin{aligned}
&\mathbb{P}\left(\frac{1}{|\mathbb{T}_n|+1}\sum_{k=p}^{n-1}(2\beta)^{n-k}\sum_{i\in\mathbb{G}_k}\bigl(\varepsilon_i^2-\sigma^2\bigr) > c\delta_1\right)\\
&\quad\leq \exp\bigl(-c\delta_1\lambda 2^{n+1}\bigr)\exp\left(\gamma\lambda^2 2^{n+1}\sum_{k=2}^{n-p+1}\bigl(2\beta^2\bigr)^k\right).
\end{aligned}$$

Next, optimizing over λ, we are led, for some positive constant c to

$$\mathbb{P}\left(\frac{P_{n-1}}{|\mathbb{T}_n|+1} - l_1' > \delta_1\right) \leq \begin{cases} \exp\bigl(-c\delta^2|\mathbb{T}_n|\bigr) & \text{if } \frac{1}{2} < \beta < \frac{\sqrt{2}}{2},\\ \exp\bigl(-c\delta^2\frac{|\mathbb{T}_n|}{n}\bigr) & \text{if } \beta = \frac{\sqrt{2}}{2},\\ \exp\bigl(-c\delta^2\bigl(\frac{1}{\beta^2}\bigr)^{n+1}\bigr) & \text{if } \beta > \frac{\sqrt{2}}{2}. \end{cases} \tag{4.16}$$

Now combining (4.8), (4.10), (4.11), (4.12), (4.14), (4.15) and (4.16), we have thus showed that

$$\begin{aligned}
&\mathbb{P}\left(\frac{1}{|\mathbb{T}_n|+1}|V_n| > \delta\right)\\
&\quad\leq \begin{cases} c_1\exp\bigl(-c_2\delta^2 2^{n+1}\bigr) + c_1\exp\bigl(-c_2\delta 2^{n+1}\bigr) + \exp\bigl(\frac{-\delta^2}{\delta+l_1+l_2}2^{n+1}\bigr) & \text{if } \beta < \frac{\sqrt{2}}{2},\\ c_1\exp\bigl(-c_2\delta^2\frac{2^{n+1}}{n+1}\bigr) + c_1\exp\bigl(-c_2\delta\frac{2^{n+1}}{n+1}\bigr) + \exp\bigl(\frac{-\delta^2}{\delta+l_1+l_2}2^{n+1}\bigr) & \text{if } \beta = \frac{\sqrt{2}}{2},\\ c_1\exp\bigl(-c_2\delta^2\bigl(\frac{1}{\beta^2}\bigr)^{n+1}\bigr) + c_1\exp\bigl(-c_2\delta\bigl(\frac{1}{\beta^2}\bigr)^{n+1}\bigr) + \exp\bigl(\frac{-\delta^2}{\delta+l_1+l_2}2^{n+1}\bigr) & \text{if } \beta > \frac{\sqrt{2}}{2}, \end{cases}
\end{aligned} \tag{4.17}$$

where the positive constants c_1 and c_2 may differ term by term.

One can easily check that the coefficients of the matrix U_n are linear combinations of terms similar to V_n, so that performing calculations similar to the above for each of them, we deduce the same deviation inequalities for U_n as in (4.17).

Now we have

$$\begin{aligned}
\mathbb{P}\left(\sum_{k=0}^{n-p}\frac{1}{2^k}\left\|\sum_{C\in\{A;B\}^k} C\frac{U_{n-k}}{2^{n-k+1}}C^t\right\| > \delta\right) &\leq \mathbb{P}\left(\sum_{k=0}^{n-p}\frac{1}{2^k}\sum_{C\in\{A;B\}^k}\frac{1}{2^{n-k+1}}\bigl\|CU_{n-k}C^t\bigr\| > \delta\right)\\
&\leq \mathbb{P}\left(\sum_{k=p}^{n}\beta^{2(n-k)}\frac{1}{|\mathbb{T}_k|+1}\|U_k\| > \delta\right)\\
&\leq \sum_{k=p}^{n}\mathbb{P}\left(\frac{\|U_k\|}{|\mathbb{T}_k|+1} > \frac{\delta}{(n-p+1)\beta^{2(n-k)}}\right).
\end{aligned}$$

From (4.17), we infer the following

$$\mathbb{P}\left(\sum_{k=0}^{n-p}\frac{1}{2^k}\left\|\sum_{C\in\{A;B\}^k} C\frac{U_{n-k}}{2^{n-k+1}}C^t\right\|>\delta\right)$$

$$\leq\begin{cases} c_1\sum_{k=p}^n\exp\bigl(-c_2\frac{\delta^2(2\beta^4)^{k+1}}{n^2\beta^{4n}}\bigr)+c_1\sum_{k=p}^n\exp\bigl(-c_2\frac{\delta(2\beta^2)^{k+1}}{n\beta^{2n}}\bigr) \\ \quad +c_1\sum_{k=p}^n\exp\bigl(-c_2\frac{\delta^2 2^{k+1}}{(\delta+nl\beta^{2(n-k-1)})n\beta^{2(n-k-1)}}\bigr) \quad \text{if } \beta<\frac{\sqrt{2}}{2}, \\ c_1\sum_{k=p}^n\exp\bigl(-c_2\frac{\delta^2 4^n}{n^2(k+1)2^{k+1}}\bigr)+c_1\sum_{k=p}^n\exp\bigl(-c_2\frac{\delta 2^n}{(k+1)n}\bigr) \\ \quad +c_1\sum_{k=p}^n\exp\bigl(-c_2\frac{\delta^2 2^{k+1}}{(\delta+nl2^{-(n-k-1)})n2^{-(n-k-1)}}\bigr) \quad \text{if } \beta=\frac{\sqrt{2}}{2}, \\ c_1\sum_{k=p}^n\exp\bigl(-c_2\frac{\delta^2(2\beta^2)^{k+1}}{n^2\beta^{4n}}\bigr)+c_1\sum_{k=p}^n\exp\bigl(-c_2\frac{\delta}{n\beta^{2n}}\bigr) \\ \quad +c_1\sum_{k=p}^n\exp\bigl(-c_2\frac{\delta^2 2^{k+1}}{(\delta+nl\beta^{2(n-k-1)})n\beta^{2(n-k-1)}}\bigr) \quad \text{if } \beta>\frac{\sqrt{2}}{2}, \end{cases}$$

where $l=l_1+l_2$ and the positive constants c_1 and c_2 may differ term by term.

Now

- If $\beta<\frac{\sqrt{2}}{2}$, then on the one hand,

$$\begin{aligned}\sum_{k=p}^n\exp\left(-c\frac{\delta^2(2\beta^4)^{k+1}}{n^2\beta^{4n}}\right) &\\ &=\exp\left(-c\delta^2\beta^4\frac{2^{n+1}}{n^2}\right)\left(1+\sum_{k=p}^{n-1}\left(\exp\left(\frac{-c\delta^2}{n^2}\right)\right)^{(2\beta^4)^{k+1}\beta^{-4n}(1-(2\beta^4)^{n-k})}\right)\\ &\leq\exp\left(-c\delta^2\beta^4\frac{2^{n+1}}{n^2}\right)\bigl(1+o(1)\bigr),\end{aligned}$$

where the last inequality follows from the fact that for some positive constant c_1,

$$\bigl(2\beta^4\bigr)^{k+1}\beta^{-4n}\bigl(1-\bigl(2\beta^4\bigr)^{n-k}\bigr)\propto c_1\bigl(2\beta^4\bigr)^{k+1}\beta^{-4n}.$$

On the other hand, following the same lines as before, we obtain

$$\begin{aligned}\sum_{k=p}^n\exp\left(-\frac{\delta^2 2^{k+1}}{(\delta+nl\beta^{2(n-k-1)})n\beta^{2(n-k-1)}}\right)&\leq\sum_{k=p}^n\exp\left(-c\delta^2\frac{2^{k+1}}{n^2\beta^{2(n-k-1)}}\right)\\ &\leq\exp\left(-c\frac{\delta^2 2^{n+1}}{(\delta+l)n^2}\right)\bigl(1+o(1)\bigr),\end{aligned}$$

and

$$\begin{aligned}\sum_{k=p}^n\exp\left(-c\frac{\delta(2\beta^2)^{k+1}}{n\beta^{2n}}\right)\leq\sum_{k=p}^n\exp\left(-c\frac{\delta(2\beta^2)^{k+1}}{n^2\beta^{2n}}\right)&\\ \leq\exp\left(-c\delta\frac{2^{n+1}}{n^2}\right)\bigl(1+o(1)\bigr).&\end{aligned}$$

We thus deduce that

$$\mathbb{P}\left(\sum_{k=0}^{n-p}\frac{1}{2^k}\left\|\sum_{C\in\{A;B\}^k} C\frac{U_{n-k}}{2^{n-k+1}}C^t\right\|>\delta\right)\leq c_1\exp\left(-c_2\delta^2\frac{2^{n+1}}{n^2}\right)+c_1\exp\left(-c_2\delta\frac{2^{n+1}}{n^2}\right) \tag{4.18}$$

for some positive constants c_1 and c_2.

- If $\beta = \frac{\sqrt{2}}{2}$, then following the same lines as before, we show that

$$\sum_{k=p}^{n} \exp\left(-c\delta^2 \frac{4^n}{n^2(k+1)2^{k+1}}\right) \leq \exp\left(-c\delta^2 \frac{2^{n+1}}{n^3}\right)\bigl(1+\mathrm{o}(1)\bigr),$$

$$\sum_{k=p}^{n} \exp\left(-\frac{\delta^2 2^{k+1}}{(\delta + l n 2^{-(n-k-1)}) n 2^{-(n-k-1)}}\right) \leq \exp\left(-c \frac{\delta^2 2^{n+1}}{n^2(\delta + l)}\right)\bigl(1+\mathrm{o}(1)\bigr),$$

$$\sum_{k=p}^{n} \exp\left(-c\delta \frac{2^n}{n(k+1)}\right) \leq \exp\left(-c\delta \frac{2^{n+1}}{n^3}\right)\bigl(1+\mathrm{o}(1)\bigr).$$

It then follows that

$$\mathbb{P}\left(\sum_{k=0}^{n-p} \frac{1}{2^k} \left\| \sum_{C \in \{A;B\}^k} C \frac{U_{n-k}}{2^{n-k+1}} C^t \right\| > \delta\right)$$
$$\leq c_1 \exp\left(-c_2\delta^2 \frac{2^{n+1}}{n^3}\right) + c_1 \exp\left(-c_2 \frac{\delta^2 2^{n+1}}{n^2(\delta + l)}\right) + c_1 \exp\left(-c_2\delta \frac{2^{n+1}}{n^3}\right) \tag{4.19}$$

for some positive constants c_1 and c_2.

- If $\beta > \frac{\sqrt{2}}{2}$, once again following the previous lines, we get

$$\mathbb{P}\left(\sum_{k=0}^{n-p} \frac{1}{2^k} \left\| \sum_{C \in \{A;B\}^k} C \frac{U_{n-k}}{2^{n-k+1}} C^t \right\| > \delta\right)$$
$$\leq c_1 \exp\left(-c_2\delta^2 \frac{1}{n^2\beta^{2n}}\right) + c_1 \exp\left(-c_2 \frac{\delta^2}{(\delta + l) n^2 \beta^{2n}}\right) + c_1 n \exp\left(-c_2 \frac{\delta}{n^2\beta^{2n}}\right) \tag{4.20}$$

for some positive constants c_1 and c_2.

We infer from the inequalities (4.18), (4.19) and (4.20) that

$$\sum_{k=0}^{n-p} \frac{1}{2^k} \sum_{C \in \{A;B\}^k} C \frac{U_{n-k}}{2^{n-k+1}} C^t \overset{\text{superexp}}{\underset{v^2_{|\mathbb{T}_n|}}{\Longrightarrow}} 0.$$

□

Lemma 4.6. *Assume that hypotheses* (N2) *and* (Xa) *are satisfied. Then we have*

$$\sum_{k=0}^{n-p} \frac{1}{2^k} \sum_{C \in \{A;B\}^k} C \frac{L_{n-k}}{2^{n-k}} e_1 e_1^t C^t \overset{\text{superexp}}{\underset{v^2_{|\mathbb{T}_n|}}{\Longrightarrow}} l, \tag{4.21}$$

where L_k is given in the second part of (4.2) *and*

$$l = \sum_{k=0}^{+\infty} \frac{1}{2^k} \sum_{C \in \{A;B\}^k} C\bigl(\sigma^2 e_1 e_1^t\bigr) C^t$$

is the unique solution of the equation

$$l = \sigma^2 e_1 e_1^t + \frac{1}{2}\bigl(A l A^t + B l B^t\bigr).$$

Proof. First, since we have for all $k \geq p$ the following decomposition on odd and even part

$$\sum_{i\in\mathbb{T}_{k,p}} \left(\varepsilon_i^2 - \sigma^2\right) = \sum_{i\in\mathbb{T}_{k-1,p-1}} \left(\varepsilon_{2i}^2 - \sigma^2\right) + \left(\varepsilon_{2i+1}^2 - \sigma^2\right),$$

we obtain for all $\delta > 0$ that

$$\mathbb{P}\left(\frac{1}{|\mathbb{T}_k|+1} \sum_{i\in\mathbb{T}_{k,p}} \left(\varepsilon_i^2 - \sigma^2\right) > \delta\right) \leq \sum_{\eta=0}^{1} \mathbb{P}\left(\frac{1}{|\mathbb{T}_k|+1} \sum_{i\in\mathbb{T}_{k-1,p-1}} \left(\varepsilon_{2i+\eta}^2 - \sigma^2\right) > \frac{\delta}{2}\right).$$

We will treat only the case $\eta = 0$. Chernoff inequality gives us for all $\lambda > 0$

$$\mathbb{P}\left(\frac{1}{|\mathbb{T}_k|+1} \sum_{i\in\mathbb{T}_{k-1,p-1}} \left(\varepsilon_{2i}^2 - \sigma^2\right) > \frac{\delta}{2}\right) \leq \exp\left(-\lambda\frac{\delta}{2}2^{k+1}\right)\mathbb{E}\left[\exp\left(\lambda \sum_{i\in\mathbb{T}_{k-1,p-1}} \left(\varepsilon_{2i}^2 - \sigma^2\right)\right)\right].$$

We obtain from hypothesis (N2), after conditioning by $\mathcal{F}_{k-1}$

$$\mathbb{E}\left[\exp\left(\lambda \sum_{i\in\mathbb{T}_{k-1,p-1}} \left(\varepsilon_{2i}^2 - \sigma^2\right)\right)\right] \leq \exp\left(\lambda^2\gamma|\mathbb{G}_{k-1}|\right)\mathbb{E}\left[\exp\left(\lambda \sum_{i\in\mathbb{T}_{k-2,p-1}} \left(\varepsilon_{2i}^2 - \sigma^2\right)\right)\right].$$

Iterating this, we deduce that

$$\mathbb{E}\left[\exp\left(\lambda \sum_{i\in\mathbb{T}_{k-1,p-1}} \left(\varepsilon_{2i}^2 - \sigma^2\right)\right)\right] \leq \exp\left(\gamma\lambda^2 \sum_{l=p-1}^{k-1} |\mathbb{G}_l|\right) \leq \exp\left(\gamma\lambda^2 2^{k+1}\right).$$

Next, optimizing on λ, we get

$$\mathbb{P}\left(\frac{1}{|\mathbb{T}_k|+1} \sum_{i\in\mathbb{T}_{k-1,p-1}} \left(\varepsilon_{2i}^2 - \sigma^2\right) > \frac{\delta}{2}\right) \leq \exp\left(-c\delta^2|\mathbb{T}_k|\right)$$

for some positive constant c which depends on γ. Applying the foregoing to the random variables $-(\varepsilon_i^2 - \sigma^2)$, we obtain

$$\mathbb{P}\left(\frac{1}{|\mathbb{T}_k|+1}\left|\sum_{i\in\mathbb{T}_{k,p}} \left(\varepsilon_i^2 - \sigma^2\right)\right| > \delta\right) \leq 4\exp\left(-c\delta^2|\mathbb{T}_k|\right). \tag{4.22}$$

Now we have

$$\sum_{k=0}^{n-p} \frac{1}{2^k} \sum_{C\in\{A;B\}^k} C\frac{L_{n-k}}{2^{n-k}}e_1e_1^tC^t - l = \sum_{k=0}^{n-p} \frac{1}{2^k} \sum_{C\in\{A;B\}^k} C\left(\frac{L_{n-k}}{2^{n-k}} - \sigma^2\right)e_1e_1^tC^t$$
$$- \sum_{k=n-p+1}^{\infty} \frac{1}{2^k} \sum_{C\in\{A;B\}^k} C\left(\sigma^2e_1e_1^t\right)C^t$$

and since the second term of the right hand side of the last equality is deterministic and tends to 0, to prove Lemma 4.6, it suffices to show that

$$\sum_{k=0}^{n-p} \frac{1}{2^k} \sum_{C\in\{A;B\}^k} C\left(\frac{L_{n-k}}{2^{n-k}} - \sigma^2\right)e_1e_1^tC^t \underset{v_{|\mathbb{T}_n|}^2}{\overset{\text{superexp}}{\Longrightarrow}} 0.$$

From the following inequalities

$$\left\|\sum_{k=0}^{n-p}\frac{1}{2^k}\sum_{C\in\{A;B\}^k}C\left(\frac{L_{n-k}}{2^{n-k}}-\sigma^2\right)e_1e_1^tC^t\right\|\le\sum_{k=0}^{n-p}\frac{1}{2^k}\sum_{C\in\{A;B\}^k}\left|\frac{L_{n-k}}{2^{n-k}}-\sigma^2\right|\|Ce_1e_1^tC^t\|$$
$$\le\sum_{k=p}^{n}\beta^{2(n-k)}\left|\frac{L_k}{|\mathbb{T}_k|+1}-\sigma^2\right|$$

and from (4.22) applied with $\delta/((n-p+1)\beta^{2(n-k)})$ instead of δ, we get

$$\mathbb{P}\left(\left\|\sum_{k=0}^{n-p}\frac{1}{2^k}\sum_{C\in\{A;B\}^k}C\left(\frac{L_{n-k}}{2^{n-k}}-\sigma^2\right)e_1e_1^tC^t\right\|>\delta\right)\le\mathbb{P}\left(\sum_{k=p}^{n}\beta^{2(n-k)}\left|\frac{L_k}{|\mathbb{T}_k|+1}-\sigma^2\right|>\delta\right)$$
$$\le\sum_{k=p}^{n}\mathbb{P}\left(\left|\frac{L_k}{|\mathbb{T}_k|+1}-\sigma^2\right|>\frac{\delta}{(n-p+1)\beta^{2(n-k)}}\right)$$
$$\le c_1\sum_{k=p}^{n}\exp\left(-c_2\delta^2\frac{(2\beta^4)^{k+1}}{n^2\beta^{4n}}\right).$$

Now, following the same lines as in the proof of (4.7) we obtain

$$\mathbb{P}\left(\left\|\sum_{k=0}^{n-p}\frac{1}{2^k}\sum_{C\in\{A;B\}^k}C\left(\frac{L_{n-k}}{2^{n-k}}-\sigma^2\right)e_1e_1^tC^t\right\|>\delta\right)\le\begin{cases}c_1\exp\left(-c_2\delta^2\frac{2^{n+1}}{n^2}\right) & \text{if } \beta^4<\frac{1}{2},\\ c_1n\exp\left(-c_2\delta^2\frac{2^{n+1}}{n^2}\right) & \text{if } \beta^4=\frac{1}{2},\\ c_1\exp\left(-c_2\delta^2\frac{1}{n^2\beta^{4n}}\right) & \text{if } \beta^4>\frac{1}{2}\end{cases}\tag{4.23}$$

for some positive constants c_1 and c_2. From (4.23), we infer that (4.21) holds. □

Lemma 4.7. *Assume that hypothesis* (N1) *is satisfied. Then we have*

$$\sum_{k=0}^{n-p}\frac{1}{2^k}\sum_{C\in\{A;B\}^k}CI_{n-k}^{(2)}C^t\underset{v_{|\mathbb{T}_n|}^2}{\overset{\text{superexp}}{\Longrightarrow}}0,\tag{4.24}$$

where $I_k^{(2)}$ is given in (4.4).

Proof. This proof follows the same lines as that of (4.21). □

Lemma 4.8. *Assume that hypotheses* (N2) *and* (Xa) *are satisfied. Then we have*

$$\sum_{k=0}^{n-p}\frac{1}{2^k}\sum_{C\in\{A;B\}^k}CI_{n-k}^{(1)}C^t\underset{v_{|\mathbb{T}_n|}^2}{\overset{\text{superexp}}{\Longrightarrow}}\Lambda',\tag{4.25}$$

where

$$\Lambda'=\sum_{k=0}^{+\infty}\frac{1}{2^k}\sum_{C\in\{A;B\}^k}C\left(T-\left(\sigma^2+\overline{a^2}\right)e_1e_1^t\right)C^t,$$

is the unique solution of the equation

$$\Lambda'=T-\left(\sigma^2+\overline{a^2}\right)e_1e_1^t+\frac{1}{2}\left(A\Lambda'A^t+B\Lambda'B^t\right),$$

where T is given (2.11) *and $I_k^{(1)}$ is given in* (4.3).

Proof. Since in the definition of $I_n^{(1)}$ given by (4.3) there are four terms, we focus only on the first term

$$\frac{a_0}{2} A \frac{H_{k-1}}{2^k} e_1^t.$$

The other terms will be treated in the same way. Using (4.29), we obtain the following decomposition:

$$\frac{a_0}{2} \sum_{k=0}^{n-p} \frac{1}{2^k} \sum_{C \in \{A;B\}^k} C A \frac{H_{n-k-1}}{2^{n-k}} e_1^t C^t = T_n^{(1)} + T_n^{(2)} + T_n^{(3)},$$

where

$$T_n^{(1)} = \frac{a_0}{2} \sum_{k=0}^{n-p} \frac{1}{2^k} \sum_{C \in \{A;B\}^k} C A \left\{ \overline{A}^{n-k-p} \frac{H_{p-1}}{2^p} + \sum_{l=p}^{n-k-1} \overline{A}^{n-k-l-1} \frac{H_{p-1}}{2^{l+1}} \right\} e_1^t C^t,$$

$$T_n^{(2)} = \frac{a_0}{2} \sum_{k=0}^{n-p} \frac{1}{2^k} \sum_{C \in \{A;B\}^k} C A \left\{ \sum_{l=p}^{n-k-1} \overline{A}^{n-k-l-1} \overline{a} \left(\frac{2^l - 2^{p-1}}{2^l} \right) e_1 e_1^t \right\} C^t$$

and

$$T_n^{(3)} = \frac{a_0}{2} \sum_{k=0}^{n-p} \frac{1}{2^k} \sum_{C \in \{A;B\}^k} C A \sum_{l=p}^{n-k-1} \overline{A}^{n-k-l-1} \frac{P_l}{2^{l+1}} e_1 e_1^t C^t, \quad \text{with } P_n = \sum_{k \in \mathbb{T}_{n,p}} \varepsilon_k.$$

On the one hand, we have

$$\left\| T_n^{(3)} \right\| \le c \sum_{k=p}^{n} \beta^{n-k} \frac{|P_k|}{2^{k+1}},$$

where c is a positive constant such that $c > |a_0| \frac{1-\beta^{n-l}}{1-\beta}$ for all $n \ge l$, so that

$$\mathbb{P}\left(\left\| T_n^{(3)} \right\| > \delta \right) \le \sum_{k=p}^{n} \mathbb{P}\left(\frac{|P_k|}{|\mathbb{T}_k| + 1} > \frac{2\delta}{cn\beta^{n-k}} \right).$$

We deduce again from hypothesis (N1) and in the same way that we have obtained (4.22) that

$$\mathbb{P}\left(\frac{P_k}{|\mathbb{T}_k| + 1} > \frac{2\delta}{cn\beta^{n-k}} \right) \le \exp\left(-c_1 \delta^2 \frac{(2\beta^2)^{k+1}}{n^2 \beta^{2n}} \right) \quad \forall k \ge p$$

for some positive constant c_1. It then follows as in the proof of (4.7) that

$$\mathbb{P}\left(\left\| T_n^{(3)} \right\| > \delta \right) \le \begin{cases} \exp\left(-c_1 \delta^2 \frac{2^{n+1}}{n^2}\right) & \text{if } \beta^2 < \frac{1}{2}, \\ n \exp\left(-c_1 \delta^2 \frac{2^{n+1}}{n^2}\right) & \text{if } \beta^2 = \frac{1}{2}, \\ \exp\left(-c_1 \delta^2 \frac{1}{n^2 \beta^{2n}}\right) & \text{if } \beta^2 > \frac{1}{2}, \end{cases}$$

so that

$$T_n^{(3)} \underset{v_{|\mathbb{T}_n|}^2}{\overset{\text{superexp}}{\Longrightarrow}} 0. \tag{4.26}$$

On the other hand, we have after tedious calculations

$$\|T_n^{(1)}\| \le \begin{cases} c\frac{\overline{X}_1}{2^{n+1}} & \text{if } \beta < \frac{1}{2}, \\ c\frac{\overline{X}_1}{\sqrt{|\mathbb{T}_n|+1}} & \text{if } \beta = \frac{1}{2}, \\ c\beta^n \overline{X}_1 & \text{if } \beta > \frac{1}{2}, \end{cases}$$

where c is a positive constant which depends on p and $|a_0|$. Next, from hypothesis (X2) and Chernoff inequality we conclude that

$$T_n^{(1)} \underset{v_{|\mathbb{T}_n|}^2}{\overset{\text{superexp}}{\Longrightarrow}} 0. \tag{4.27}$$

Furthermore, since $(T_n^{(2)})$ is a deterministic sequence, we have (see [3], Lemma A.4)

$$T_n^{(2)} \underset{v_{|\mathbb{T}_n|}^2}{\overset{\text{superexp}}{\Longrightarrow}} \Lambda'', \tag{4.28}$$

where

$$\Lambda'' = \sum_{k=0}^{+\infty} \frac{1}{2^k} \sum_{C\in\{A;B\}^k} C\left(\frac{1}{2}a_0 A \varXi e_1^t\right)C^t$$

is the unique solution of

$$\Lambda'' = \frac{1}{2}a_0 A \varXi e_1^t + \frac{1}{2}\left(A\Lambda'' A^t + B\Lambda'' B^t\right).$$

It then follows that

$$\frac{a_0}{2}\sum_{k=0}^{n-p} \frac{1}{2^k} \sum_{C\in\{A;B\}^k} CA\frac{H_{n-k-1}}{2^{n-k}} e_1^t C^t \underset{v_{|\mathbb{T}_n|}^2}{\overset{\text{superexp}}{\Longrightarrow}} \Lambda''.$$

Doing the same for the three other terms of $I_k^{(1)}$, we end the proof of Lemma 4.8. □

Proposition 4.9. *Assume that hypotheses* (N2) *and* (Xa) *are satisfied. Then we have*

$$\frac{1}{|\mathbb{T}_n|}\sum_{k\in\mathbb{T}_{n,p}} \mathbb{X}_k \underset{v_{|\mathbb{T}_n|}^2}{\overset{\text{superexp}}{\Longrightarrow}} \varXi,$$

where $\varXi$ *is given in* (2.9).

Proof. Let

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

From p. 2517 in Bercu et al. [3], we have

$$\frac{H_n}{2^{n+1}} = \sum_{k=p-1}^{n} (\overline{A})^{n-k}\frac{H_{p-1}}{2^{k+1}} + \sum_{k=p}^{n} \overline{a}(\overline{A})^{n-k}\left(\frac{2^k - 2^{p-1}}{2^k}\right)e_1 + \sum_{k=p}^{n} \frac{P_k}{2^{k+1}}(\overline{A})^{n-k} e_1.$$

Since the second term in the right hand side of this equality is deterministic and converges to Ξ, this proposition will be proved if we show that

$$\sum_{k=p-1}^{n} \frac{(\overline{A})^{n-k}}{2^k} H_{p-1} \overset{\text{superexp}}{\underset{v^2_{|\mathbb{T}_n|}}{\Longrightarrow}} 0, \qquad \sum_{k=p}^{n} \frac{P_k}{2^{k+1}} (\overline{A})^{n-k} e_1 \overset{\text{superexp}}{\underset{v^2_{|\mathbb{T}_n|}}{\Longrightarrow}} 0, \tag{4.29}$$

which follows by reasoning as in the proof of Proposition 4.3 (see the proof of Proposition 4.3 for more details). □

We now explain the modification in the last proofs in case 1.

Proposition 4.10. *Within the framework of case* 1, *we have the same conclusions as Propositions* 4.9 *and* 4.3 *with the sequence* (v_n) *satisfying condition* (V1).

Proof. The proof follows exactly the same lines as the proof of Propositions 4.9 and 4.3, and uses the fact that if a superexponential convergence holds with a sequence (v_n) satisfying condition (V2), then it also holds with a sequence (v_n) satisfying condition (V1). We thus obtain the first convergence of (4.29), the convergences (4.6), (4.27), (4.28) and (4.24) within the framework of case 1 with (v_n) satisfying condition (V1). Next, following the same approach as which used to obtain (4.22), we get

$$\mathbb{P}\left(\frac{1}{|\mathbb{T}_k|+1}\left|\sum_{i\in\mathbb{T}_{k,p}}\left(\varepsilon_i^2-\sigma^2\right)\right|>\delta\right)\le\begin{cases} c_1\exp\left(-c_2\delta^2|\mathbb{T}_k|\right) & \text{if } \delta \text{ is small enough,}\\ c_1\exp\left(-c_2\delta|\mathbb{T}_k|\right) & \text{if } \delta \text{ is large enough,}\end{cases} \tag{4.30}$$

where c_1 and c_2 are positive constants which do not depend on δ. On the other hand, let n_0 such that for $n > n_0\delta/(n-p+1)\gamma\beta^{2(n-n_0)}$ is large enough. We have

$$\begin{aligned}
&\mathbb{P}\left(\left\|\sum_{k=0}^{n-p}\frac{1}{2^k}\sum_{C\in\{A;B\}^k} C\left(\frac{L_{n-k}}{2^{n-k}}-\sigma^2\right)e_1e_1^tC^t\right\|>\delta\right)\\
&\quad\le\sum_{k=p}^{n_0-1}\mathbb{P}\left(\left|\frac{L_k}{|\mathbb{T}_k|+1}-\sigma^2\right|>\frac{\delta}{(n-p+1)\beta^{2(n-k)}}\right)+\sum_{k=n_0}^{n}\mathbb{P}\left(\left|\frac{L_k}{|\mathbb{T}_k|+1}-\sigma^2\right|>\frac{\delta}{(n-p+1)\beta^{2(n-k)}}\right).
\end{aligned}$$

Now, using (4.30) with $\delta/(n-p+1)\beta^{2(n-k)}$ instead of δ and following the same approach used to obtain (4.18)–(4.20) in the two sums of the right hand side of the above inequality, we are led to

$$\begin{aligned}
&\mathbb{P}\left(\left\|\sum_{k=0}^{n-p}\frac{1}{2^k}\sum_{C\in\{A;B\}^k} C\left(\frac{L_{n-k}}{2^{n-k}}-\sigma^2\right)e_1e_1^tC^t\right\|>\delta\right)\\
&\quad\le\begin{cases} c_1\exp\left(-\frac{c_2\delta^2 2^{n+1}}{n^2}\right)+c_1\exp\left(-\frac{c_2\delta 2^{n+1}}{n}\right) & \text{if } \beta\le\frac{1}{2},\\ c_1 n\exp\left(-\frac{c_2\delta^2}{n^2\beta^{4n}}\right)+c_1\exp\left(-\frac{c_2\delta}{n\beta^{2n}}\right) & \text{if } \beta>\frac{1}{2},\end{cases}
\end{aligned}$$

and we thus obtain convergence (4.21) with (v_n) satisfying condition (V1). In the same way we obtain

$$\mathbb{P}\left(\left\|T_n^{(3)}\right\|>\delta\right)\le\begin{cases} c_1\exp\left(-\frac{c_2\delta^2 2^{n+1}}{n^2}\right)+c_1\exp\left(-\frac{c_2\delta 2^{n+1}}{n}\right) & \text{if } \beta<\frac{1}{2},\\ c_1 n\exp\left(-\frac{c_2\delta 2^{n+1}}{n}\right) & \text{if } \beta=\frac{1}{2},\\ c_1\exp\left(-\frac{c_2\delta^2}{n^2\beta^{2n}}\right)+c_1\exp\left(-\frac{c_2\delta}{n\beta^{n}}\right) & \text{if } \beta>\frac{1}{2},\end{cases}$$

so that (4.26) and then (4.25) hold for (v_n) satisfying condition (V1). To reach the convergence (4.7) and the second convergence of (4.29) with (v_n) satisfying condition (V1), we follow the same procedure as before and the proof of the proposition is then complete. □

Remark 4.11. *Let us note that we can actually prove that*

$$\frac{1}{n}\sum_{k=2^p}^{n}\mathbb{X}_k \overset{\text{superexp}}{\underset{v_n^2}{\Longrightarrow}} \Xi \quad \textit{and} \quad \frac{1}{n}\sum_{k=2^p}^{n}\mathbb{X}_k\mathbb{X}_k^t \overset{\text{superexp}}{\underset{v_n^2}{\Longrightarrow}} \Lambda.$$

Indeed, let $H_n = \sum_{k=2^{p-1}}^{n}\mathbb{X}_k$ *and* $P_l^{(n)} = \sum_{k=2^{r_n-l}}^{[n/2^l]}\varepsilon_k$. *We have the following decomposition*

$$\frac{H_n}{n} - \Xi = \frac{1}{n}\sum_{k\in\mathbb{T}_{r_n-1,p-1}}(\mathbb{X}_k - \Xi) + \frac{1}{n}\sum_{k=2^{r_n}}^{n}(\mathbb{X}_k - \Xi) + \frac{2^{p-1}-1}{n}\Xi.$$

On the one hand, observing that $v_n/v_{|\mathbb{T}_{r_n-1}|} < 2$, *we infer from Proposition* 4.9 *that*

$$\frac{1}{n}\sum_{k\in\mathbb{T}_{r_n-1,p-1}}(\mathbb{X}_k - \Xi) \overset{\text{superexp}}{\underset{v_n^2}{\Longrightarrow}} 0.$$

The sequence $(\frac{2^{p-1}-1}{n}\Xi)$ *being deterministic and converging to* 0, *we deduce that*

$$\frac{2^{p-1}-1}{n}\Xi \overset{\text{superexp}}{\underset{v_n^2}{\Longrightarrow}} 0.$$

On the other hand, from (2.1) *we deduce that*

$$\begin{aligned}\sum_{k=2^{r_n}}^{n}\mathbb{X}_k &= 2^{r_n-p+1}(\overline{A})^{r_n-p+1}\sum_{k=2^{p-1}}^{[n/(2^{r_n}-p+1)]}\mathbb{X}_k + 2\overline{a}\sum_{k=0}^{r_n-p}\left(\left[\frac{n}{2^k}\right] - 2^{r_n-k}+1\right)2^k(\overline{A})^k e_1\\ &\quad + \sum_{k=0}^{r_n-p}2^k(\overline{A})^k P_k^{(n)}e_1 - \sum_{k=1}^{r_n-p+1}s_k 2^{k-1}(\overline{A})^{k-1}(B\mathbb{X}_{[n/2^k]} + \eta_{[n/2^{k-1}]+1}),\end{aligned}$$

where

$$s_k = \begin{cases} 1 & \textit{if } \left[\frac{n}{2^{k-1}}\right] \textit{ is even},\\ 0 & \textit{if } \left[\frac{n}{2^{k-1}}\right] \textit{ is odd}.\end{cases}$$

Reasoning now as in the proof of Proposition 4.9, *tedious but straightforward calculations lead us to*

$$\frac{1}{n}\sum_{k=2^{r_n}}^{n}(\mathbb{X}_k - \Xi) \overset{\text{superexp}}{\underset{v_n^2}{\Longrightarrow}} 0.$$

It then follows that

$$\frac{1}{n}\sum_{k=2^p}^{n}\mathbb{X}_k \overset{\text{superexp}}{\underset{v_n^2}{\Longrightarrow}} \Xi.$$

The term $\frac{1}{n}\sum_{k=2^p}^{n}\mathbb{X}_k\mathbb{X}_k^t$ *can be dealt with in the same way.*

The rest of the paper is dedicated to the proof of our main results. We focus on the proof in case 2, and some explanations are given on how to obtain the results in case 1.

5. Proof of the main results

We start with the proof of the deviation inequalities.

5.1. *Proof of Theorem* 3.1

We begin the proof with case 2. Let $\delta > 0$ and $\ell > 0$ such that $\ell < \|\Sigma\|/(1+\delta)$. We have from (2.7)

$$\mathbb{P}\big(\|\hat{\theta}_n - \theta\| > \delta\big) = \mathbb{P}\bigg(\frac{\|M_n\|}{\|\Sigma_{n-1}\|} > \delta, \frac{\|\Sigma_{n-1}\|}{|\mathbb{T}_{n-1}|} \geq \ell\bigg) + \mathbb{P}\bigg(\frac{\|M_n\|}{\|\Sigma_{n-1}\|} > \delta, \frac{\|\Sigma_{n-1}\|}{|\mathbb{T}_{n-1}|} < \ell\bigg)$$
$$\leq \mathbb{P}\bigg(\frac{\|M_n\|}{|\mathbb{T}_{n-1}|} > \delta\ell\bigg) + \mathbb{P}\bigg(\bigg\|\frac{\Sigma_{n-1}}{|\mathbb{T}_{n-1}|} - \Sigma\bigg\| > \|\Sigma\| - \ell\bigg).$$

Since $\ell < \|\Sigma\|/(1+\delta)$, then

$$\mathbb{P}\bigg(\bigg\|\frac{\Sigma_{n-1}}{|\mathbb{T}_{n-1}|} - \Sigma\bigg\| > \|\Sigma\| - \ell\bigg) \leq \mathbb{P}\bigg(\bigg\|\frac{\Sigma_{n-1}}{|\mathbb{T}_{n-1}|} - \Sigma\bigg\| > \delta\ell\bigg).$$

It then follows that

$$\mathbb{P}\big(\|\hat{\theta}_n - \theta\| > \delta\big) \leq 2\max\bigg\{\mathbb{P}\bigg(\frac{\|M_n\|}{|\mathbb{T}_{n-1}|} > \delta\ell\bigg), \mathbb{P}\bigg(\bigg\|\frac{\Sigma_{n-1}}{|\mathbb{T}_{n-1}|} - \Sigma\bigg\| > \delta\ell\bigg)\bigg\}.$$

On the one hand, we have

$$\mathbb{P}\bigg(\frac{\|M_n\|}{|\mathbb{T}_{n-1}|} > \delta\ell\bigg) \leq \sum_{\eta=0}^{1}\bigg\{\mathbb{P}\bigg(\bigg|\frac{1}{|\mathbb{T}_{n-1}|}\sum_{k\in\mathbb{T}_{n-1,p-1}} \varepsilon_{2k+\eta}\bigg| > \frac{\delta\ell}{4}\bigg)$$
$$+ \mathbb{P}\bigg(\bigg\|\frac{1}{|\mathbb{T}_{n-1}|}\sum_{k\in\mathbb{T}_{n-1,p-1}} \varepsilon_{2k+\eta}\mathbb{X}_k\bigg\| > \frac{\delta\ell}{4}\bigg)\bigg\}.$$

Now, by carrying out the same calculations as those which have permitted us to obtain Lemma 4.7 and equation (4.17), we are led to

$$\mathbb{P}\bigg(\frac{\|M_n\|}{|\mathbb{T}_{n-1}|} > \delta\ell\bigg) \leq \begin{cases} c_1 \exp\big(-\frac{c_2(\delta\ell)^2}{c_3+c_4(\delta\ell)}2^n\big) & \text{if } \beta < \frac{\sqrt{2}}{2}, \\ c_1 \exp\big(-\frac{c_2(\delta\ell)^2}{c_3+c_4(\delta\ell)}\frac{2^n}{n}\big) & \text{if } \beta = \frac{\sqrt{2}}{2}, \\ c_1 \exp\big(-\frac{c_2(\delta\ell)^2}{c_3+c_4(\delta\ell)}\big(\frac{1}{\beta^2}\big)^n\big) & \text{if } \beta > \frac{\sqrt{2}}{2}, \end{cases} \tag{5.1}$$

where the positive constants c_1, c_2, c_3 and c_4 depend on σ, β, γ and ϕ and $(c_3, c_4) \neq (0,0)$.

On the other hand, noticing that $\Sigma_{n-1} = I_2 \otimes S_{n-1}$, we have

$$\mathbb{P}\bigg(\bigg\|\frac{\Sigma_{n-1}}{|\mathbb{T}_{n-1}|} - \Sigma\bigg\| > \delta\ell\bigg) \leq 2\mathbb{P}\bigg(\bigg\|\frac{S_{n-1}}{|\mathbb{T}_{n-1}|} - L\bigg\| > \frac{\delta\ell}{2}\bigg).$$

Next, from the proofs of Propositions 4.9 and 4.3, we deduce that

$$\mathbb{P}\bigg(\bigg\|\frac{\Sigma_{n-1}}{|\mathbb{T}_{n-1}|} - \Sigma\bigg\| > \frac{\ell}{2}\bigg) \leq \begin{cases} c_1 \exp\big(-\frac{c_2(\delta\ell)^2}{c_3+c_4(\delta\ell)}\frac{2^n}{(n-1)^2}\big) & \text{if } \beta < \frac{\sqrt{2}}{2}, \\ c_1 \exp\big(-\frac{c_2(\delta\ell)^2}{c_3+c_4(\delta\ell)}\frac{2^n}{(n-1)^3}\big) & \text{if } \beta = \frac{\sqrt{2}}{2}, \\ c_1 \exp\big(-\frac{c_2(\delta\ell)^2}{c_3+c_4(\delta\ell)}\big(\frac{1}{(n-1)^2\beta^{2n}}\big)\big) & \text{if } \beta > \frac{\sqrt{2}}{2}, \end{cases} \tag{5.2}$$

where the positive constants c_1, c_2, c_3 and c_4 depend on σ, β, γ and ϕ and $(c_3, c_4) \neq (0,0)$. Now, (3.1) follows from (5.1) and (5.2).

In case 1, the proof follows exactly the same lines as before and uses the same ideas as the proof of Proposition 4.10. In particular, we have in this case

$$\mathbb{P}\left(\left\|\frac{\Sigma_{n-1}}{|\mathbb{T}_{n-1}|}-\Sigma\right\|>\frac{\ell}{2}\right)\leq\begin{cases}c_1\exp\left(-\frac{c_2(\delta\ell)^2}{c_3+(\delta\ell)}\frac{2^n}{(n-1)^2}\right) & \text{if } \beta<\frac{1}{2},\\ c_1(n-1)\exp\left(-\frac{c_2(\delta\ell)^2}{c_3+(\delta\ell)}\frac{2^n}{(n-1)^2}\right) & \text{if } \beta=\frac{1}{2},\\ c_1(n-1)\exp\left(-\frac{c_2(\delta\ell)^2}{c_3+(\delta\ell)}\left(\frac{1}{(n-1)\beta^n}\right)\right) & \text{if } \beta>\frac{1}{2},\end{cases}$$

where the positive constants c_1, c_2 and c_3 depend on σ, β, γ and ϕ. (3.1) then follows in this case, and this ends the proof of Theorem 3.1.

5.2. *Proof of Theorem* 3.7

First we need to prove the following

Theorem 5.1. *In case* 1 *or in case* 2, *the sequence* $(M_n/(v_{|\mathbb{T}_{n-1}|}\sqrt{|\mathbb{T}_{n-1}|}))_{n\geq 1}$ *satisfies the MDP on* $\mathbb{R}^{2(p+1)}$ *with speed* $v^2_{|\mathbb{T}_{n-1}|}$ *and rate function*

$$I_M(x)=\sup_{\lambda\in\mathbb{R}^{2(p+1)}}\left\{\lambda^t x-\lambda^t(\Gamma\otimes L)\lambda\right\}=\frac{1}{2}x^t(\Gamma\otimes L)^{-1}x. \tag{5.3}$$

5.2.1. *Proof of Theorem* 5.1

Since the size of the data doubles at each generation, we are not able to verify the Lindeberg condition. To come over this problem, and as in Bercu et al. [3], p. 2510, we change the filtration and we will use the sister pair-wise one, that is, $(\mathcal{G}_n)_{n\geq 1}$ given by $\mathcal{G}_n=\sigma\{X_1,(X_{2k},X_{2k+1}),1\leq k\leq n\}$. We introduce the following $(\mathcal{G}_n)$ martingale difference sequence (D_n), given by

$$D_n=V_n\otimes Y_n=\begin{pmatrix}\varepsilon_{2n}\\ \varepsilon_{2n}\mathbb{X}_n\\ \varepsilon_{2n+1}\\ \varepsilon_{2n+1}\mathbb{X}_n\end{pmatrix}.$$

We clearly have

$$D_nD_n^t=V_nV_n^t\otimes Y_nY_n^t.$$

So we obtain that the quadratic variation of the $(\mathcal{G}_n)$ martingale $(N_n)_{n\geq 2^{p-1}}$ given by

$$N_n=\sum_{k=2^{p-1}}^{n}D_k$$

is

$$\langle N\rangle_n=\sum_{k=2^{p-1}}^{n}\mathbb{E}\left(D_kD_k^t/\mathcal{G}_{k-1}\right)=\Gamma\otimes\sum_{k=2^{p-1}}^{n}Y_kY_k^t.$$

Now we clearly have $M_n=N_{|\mathbb{T}_{n-1}|}$ and $\langle M\rangle_n=\langle N\rangle_{|\mathbb{T}_{n-1}|}=\Gamma\otimes S_{n-1}$. From Proposition 4.1, and since $\langle M\rangle_n=\Gamma\otimes S_{n-1}$, we have

$$\frac{\langle M\rangle_n}{|\mathbb{T}_n|}\underset{v^2_{|\mathbb{T}_{n-1}|}}{\overset{\text{superexp}}{\Longrightarrow}}\Gamma\otimes L. \tag{5.4}$$

Before going to the proof of the MDP results, we state the exponential Lyapounov condition for $(N_n)_{n\geq 2^{p-1}}$, which implies exponential Lindeberg condition, that is

$$\limsup \frac{1}{v_n^2}\log\mathbb{P}\left(\frac{1}{n}\sum_{k=2^{p-1}}^{n}\mathbb{E}\left[\|D_k\|^2\mathbf{1}_{\{\|D_k\|\geq r(\sqrt{n}/v_n)\}}\right]\geq\delta\right)=-\infty$$

(see Remark 3, p. 10, in [25] for more details on this implication).

Remarks 5.2. *By* [14], *we infer from the condition* (Ea) *that*

(Na) *one can find $\gamma_a>0$ such that for all $n\geq p-1$, for all $k\in\mathbb{G}_{n+1}$ and for all $t\in\mathbb{R}$, with $\mu_a=\mathbb{E}(|\varepsilon_k|^a/\mathcal{F}_n)$ a.s.*

$$\mathbb{E}\left[\exp t\left(|\varepsilon_k|^a-\mu_a\right)/\mathcal{F}_n\right]\leq\exp\left(\frac{\gamma_a t^2}{2}\right)\quad a.s.$$

Proposition 5.3. *Let (v_n) be a sequence satisfying assumption* (V2). *Assume that hypotheses* (Na) *and* (Xa) *are satisfied. Then there exists $B>0$ such that*

$$\limsup_{n\to\infty}\frac{1}{v_n^2}\log\mathbb{P}\left(\frac{1}{n}\sum_{j=2^{p-1}}^{n}\mathbb{E}\left[\|D_j\|^a/\mathcal{G}_{j-1}\right]>B\right)=-\infty.$$

Proof. We are going to prove that

$$\limsup_{n\to\infty}\frac{1}{v_{|\mathbb{T}_n|}^2}\log\mathbb{P}\left(\frac{1}{|\mathbb{T}_n|}\sum_{j=2^p}^{|\mathbb{T}_n|}\mathbb{E}\left[\|D_j\|^a/\mathcal{G}_{j-1}\right]>B\right)=-\infty,\tag{5.5}$$

and Proposition 5.3 will follow by proceeding as in Remark 4.11. We have

$$\sum_{j\in\mathbb{T}_{n,p}}\mathbb{E}\left[\|D_j\|^a/\mathcal{G}_{j-1}\right]\leq c\mu^a\sum_{j\in\mathbb{T}_{n,p}}\left(1+\|\mathbb{X}_j\|^a\right),$$

where c is a positive constant which depends on a. From (2.1), we deduce that

$$\sum_{j\in\mathbb{T}_{n,p}}\|\mathbb{X}_j\|^a\leq\frac{c^2}{(1-\beta)^{a-1}}P_n+\frac{c^2\alpha^a Q_n}{(1-\beta)^{a-1}}+2cR_n\overline{X}_1^a,$$

where

$$P_n=\sum_{j\in\mathbb{T}_{n,p}}\sum_{i=0}^{r_j-p}\beta^i|\varepsilon_{[j/2^i]}|^a,\qquad Q_n=\sum_{j\in\mathbb{T}_{n,p}}\sum_{i=0}^{r_j-p}\beta^i,\qquad R_n=\sum_{j\in\mathbb{T}_{n,p}}\beta^{a(r_j-p+1)},$$

and c is a positive constant. Now, proceeding as in the proof of Proposition 4.3, using hypotheses (Na) and (Xa) instead of (N2) and (X2), we get for B large enough

$$\limsup_{n\to\infty}\frac{1}{v_{|\mathbb{T}_n|}^2}\log\mathbb{P}\left(\frac{1}{|\mathbb{T}_n|}\sum_{j\in\mathbb{T}_{n,p}}\|\mathbb{X}_j\|^a>B\right)=-\infty.\tag{5.6}$$

Now (5.6) leads us to (5.5) and following the same approach as in Remark 4.11, we obtain Proposition 5.3. □

Remarks 5.4. *In case* 1, *we clearly have that* $(\mathbb{X}_n, n \in \mathbb{T}_{\cdot,p-1})$, *where*

$$\mathbb{T}_{\cdot,p-1} = \bigcup_{r=p-1}^{\infty} \mathbb{G}_r,$$

is a bifurcating Markov chain with initial state $\mathbb{X}_{2^{p-1}} = (X_{2^{p-1}}, X_{2^{p-2}}, \dots, X_1)^t$. *Let* ν *be the law of* $\mathbb{X}_{2^{p-1}}$. *From hypothesis* (X2), *we deduce that* ν *has finite moments of all orders. We denote by* P *the transition probability kernel associated to* $(\mathbb{X}_n, n \in \mathbb{T}_{\cdot,p-1})$. *Let* $(\mathbb{Y}_r, r \in \mathbb{N})$ *the ergodic stable Markov chain associated to* $(\mathbb{X}_n, n \in \mathbb{T}_{\cdot,p-1})$. *This Markov chain is defined as follows, starting from the root* $\mathbb{Y}_0 = \mathbb{X}_{2^{p-1}}$ *and if* $\mathbb{Y}_r = \mathbb{X}_n$ *then* $\mathbb{Y}_{r+1} = \mathbb{X}_{2n+\zeta_{r+1}}$ *for a sequence of independent Bernoulli r.v.* $(\zeta_q, q \in \mathbb{N}^*)$ *such that* $\mathbb{P}(\zeta_q = 0) = \mathbb{P}(\zeta_q = 1) = 1/2$.

Let μ *be the stationary distribution associated to* $(\mathbb{Y}_r, r \in \mathbb{N})$. *For more details on bifurcating Markov chain and the associated ergodic stable Markov chain, we refer to* [18] (*see also* [5]).

From [5], *we deduce that for all real bounded function* f *defined on* $(\mathbb{R}^p)^3$,

$$\frac{1}{v_{|\mathbb{T}_{n-1}|}\sqrt{|\mathbb{T}_{n-1}|}} \sum_{k\in\mathbb{T}_{n-1,p-1}} f(\mathbb{X}_k, \mathbb{X}_{2k}, \mathbb{X}_{2k+1})$$

satisfies a MDP on $\mathbb{R}$ *with speed* $v^2_{|\mathbb{T}_{n-1}|}$ *and the rate function* $I(x) = \frac{x^2}{2S^2(f)}$, *where* $S^2(f) = \langle \mu, P(f^2) - (Pf)^2\rangle$.

Now, let f *be the function defined on* $(\mathbb{R}^p)^3$ *by* $f(x,y,z) = \|x\|^2 + \|y\|^2 + \|z\|^2$. *Then, using the relation* (4.1) *in Proposition* 4.1, *the above MDP for real bounded functionals of the bifurcating Markov chain* $(\mathbb{X}_n, n \in \mathbb{T}_{\cdot,p-1})$ *and the truncation of the function* f, *we prove* (*in the same manner as the proof of Lemma* 3 *in Worms* [25]) *that for all* $r > 0$

$$\limsup_{R\to\infty}\limsup_{n\to\infty} \frac{1}{v_n^2}\log\mathbb{P}\Bigg(\frac{1}{n}\sum_{j=2^{p-1}}^{n}\big(\|X_j\|^2 + \|X_{2j}\|^2 + \|X_{2j+1}\|^2\big) \times \mathbf{1}_{\{\|\mathbb{X}_j\|+\|\mathbb{X}_{2j}\|+\|\mathbb{X}_{2j+1}\|>R\}} > r\Bigg) = -\infty,$$

which implies the following Lindeberg condition (*for more details, we refer to Proposition* 2 *in Worms* [25])

$$\limsup_{n\to\infty} \frac{1}{v_n^2}\log\mathbb{P}\Bigg(\frac{1}{n}\sum_{j=2^{p-1}}^{n}\big(\|X_j\|^2 + \|X_{2j}\|^2 + \|X_{2j+1}\|^2\big) \times \mathbf{1}_{\{\|\mathbb{X}_j\|+\|\mathbb{X}_{2j}\|+\|\mathbb{X}_{2j+1}\|>r(\sqrt{n}/v_n)\}} > \delta\Bigg) = -\infty$$

for all $\delta > 0$ *and for all* $r > 0$. *Notice that the above Lindeberg condition implies in particular the Lindeberg condition on the sequence* $(\mathbb{X}_n)$.

Now, we come back to the proof of Theorem 5.1. We divide the proof into four steps. In the first one, we introduce a truncation of the martingale $(M_n)_{n\geq 0}$ and prove that the truncated martingale satisfies some MDP thanks to Puhalskii's Theorem 3.12. In the second part, we show that the truncated martingale is an exponentially good approximation of (M_n), see e.g. Definition 4.2.14 in [12]. We conclude by the identification of the rate function.

Proof in case 2. *Step 1*. From now on, in order to apply Puhalskii's result [24] (Puhalskii's Theorem 3.12) for the MDP for martingales, we introduce the following truncation of the martingale $(M_n)_{n\geq 0}$. For $r > 0$ and $R > 0$,

$$M_n^{(r,R)} = \sum_{k\in\mathbb{T}_{n-1,p-1}} D_{k,n}^{(r,R)},$$

where, for all $1 \le k \le n$, $D_{k,n}^{(r,R)} = V_k^{(R)} \otimes Y_{k,n}^{(r)}$, with

$$V_n^{(R)} = \big(\varepsilon_{2n}^{(R)}, \varepsilon_{2n+1}^{(R)}\big)^t \quad \text{and} \quad Y_{k,n}^{(r)} = \big(1, \mathbb{X}_{k,n}^{(r)}\big)^t,$$

where

$$\varepsilon_k^{(R)} = \varepsilon_k \mathbf{1}_{\{|\varepsilon_k| \le R\}} - \mathbb{E}[\varepsilon_k \mathbf{1}_{\{|\varepsilon_k| \le R\}}], \qquad \mathbb{X}_{k,n}^{(r)} = \mathbb{X}_k \mathbf{1}_{\{\|\mathbb{X}_k\| \le r(\sqrt{|\mathbb{T}_{n-1}|}/v_{|\mathbb{T}_{n-1}|})\}}.$$

We introduce $\Gamma^{(R)}$ the conditional covariance matrix associated with $(\varepsilon_{2k}^{(R)}, \varepsilon_{2k+1}^{(R)})^t$ and the truncated matrix associated with S_n:

$$\Gamma^{(R)} = \begin{pmatrix} \sigma_R^2 & \rho_R \\ \rho_R & \sigma_R^2 \end{pmatrix} \quad \text{and} \quad S_n^{(r)} = \sum_{k \in \mathbb{T}_{n,p-1}} \begin{pmatrix} 1 & (\mathbb{X}_{k,n}^{(r)})^t \\ \mathbb{X}_{k,n}^{(r)} & \mathbb{X}_{k,n}^{(r)} (\mathbb{X}_{k,n}^{(r)})^t \end{pmatrix}.$$

The condition (P2) in Puhalskii's Theorem 3.12 is verified by the construction of the truncated martingale, that is for some positive constant c, we have that for all $k \in \mathbb{T}_{n-1}$

$$\big\| D_{k,n}^{(r,R)} \big\| \le c \frac{\sqrt{|\mathbb{T}_{n-1}|}}{v_{|\mathbb{T}_{n-1}|}}.$$

From Proposition 5.3, we also have for all $r > 0$,

$$\frac{1}{|\mathbb{T}_{n-1}|} \sum_{k \in \mathbb{T}_{n-1,p-1}} \mathbb{X}_k \mathbf{1}_{\{\|\mathbb{X}_k\| > r(\sqrt{|\mathbb{T}_{n-1}|}/v_{|\mathbb{T}_{n-1}|})\}} \underset{v_{|\mathbb{T}_{n-1}|}^2}{\overset{\text{superexp}}{\Longrightarrow}} 0; \tag{5.7}$$

and

$$\frac{1}{|\mathbb{T}_{n-1}|} \sum_{k \in \mathbb{T}_{n-1,p-1}} \mathbb{X}_k \mathbb{X}_k^t \mathbf{1}_{\{\|\mathbb{X}_k\| > r(\sqrt{|\mathbb{T}_{n-1}|}/v_{|\mathbb{T}_{n-1}|})\}} \underset{v_{|\mathbb{T}_{n-1}|}^2}{\overset{\text{superexp}}{\Longrightarrow}} 0. \tag{5.8}$$

From (5.7) and (5.8), we deduce that for all $r > 0$

$$\frac{1}{|\mathbb{T}_{n-1}|} \big(S_{n-1} - S_{n-1}^{(r)}\big) \underset{v_{|\mathbb{T}_{n-1}|}^2}{\overset{\text{superexp}}{\Longrightarrow}} 0. \tag{5.9}$$

Then, we easily transfer the properties (5.4) to the truncated martingale $(M_n^{(r,R)})_{n \ge 0}$. We have for all $R > 0$ and all $r > 0$,

$$\frac{\langle M^{(r,R)} \rangle_n}{|\mathbb{T}_{n-1}|} = \Gamma^{(R)} \otimes \frac{S_{n-1}^{(r)}}{|\mathbb{T}_{n-1}|} = -\Gamma^{(R)} \otimes \left(\frac{S_{n-1} - S_{n-1}^{(r)}}{|\mathbb{T}_{n-1}|} \right) + \Gamma^{(R)} \otimes \frac{S_{n-1}}{|\mathbb{T}_{n-1}|} \underset{v_{|\mathbb{T}_{n-1}|}^2}{\overset{\text{superexp}}{\Longrightarrow}} \Gamma^{(R)} \otimes L.$$

That is condition (P1) in Puhalskii's Theorem 3.12.

Note also that Proposition 5.3 works for the truncated martingale $(M_n^{(r,R)})_{n \ge 0}$, which ensures Lindeberg's condition and thus condition (P3) for $(M_n^{(r,R)})_{n \ge 0}$. By Theorem 3.12, we deduce that $(M_n^{(r,R)}/(v_{|\mathbb{T}_{n-1}|}\sqrt{|\mathbb{T}_{n-1}|}))_{n \ge 0}$ satisfies a MDP on $\mathbb{R}^{2(p+1)}$ with speed $v_{|\mathbb{T}_{n-1}|}^2$ and good rate function given by

$$I_R(x) = \frac{1}{2} x^t \big(\Gamma^{(R)} \otimes L\big)^{-1} x. \tag{5.10}$$

Step 2. First, we infer from the hypothesis (Ea) that:

(N1R) there is a sequence $(\kappa_R)_{R>0}$ with $\kappa_R \longrightarrow 0$ when R goes to infinity, such that for all $n \geq p-1$, for all $k \in \mathbb{G}_{n+1}$, for all $t \in \mathbb{R}$ and for R large enough

$$\mathbb{E}\big[\exp t\big(\varepsilon_k - \varepsilon_k^R\big)/\mathcal{F}_n\big] \leq \exp\left(\frac{\kappa_R t^2}{2}\right) \quad \text{a.s.}$$

Then, we have to prove that for all $r>0$ the sequence $(M_n^{(r,R)})_n$ is an exponentially good approximation of (M_n) as R goes to infinity, see e.g. Definition 4.2.14 in [12]. This approximation in the sense of the moderate deviation, is described by the following convergence, for all $r>0$ and all $\delta>0$,

$$\limsup_{R\to\infty}\limsup_{n\to\infty}\frac{1}{v_{|\mathbb{T}_{n-1}|}^2}\log\mathbb{P}\left(\frac{\|M_n - M_n^{(r,R)}\|}{\sqrt{|\mathbb{T}_{n-1}|}v_{|\mathbb{T}_{n-1}|}} > \delta\right) = -\infty.$$

For that, we shall prove that for $\eta \in \{0,1\}$

$$I_1 = \frac{1}{\sqrt{|\mathbb{T}_{n-1}|}v_{|\mathbb{T}_{n-1}|}}\sum_{k\in\mathbb{T}_{n-1,p-1}}\big(\varepsilon_{2k+\eta} - \varepsilon_{2k+\eta}^{(R)}\big) \underset{v_{|\mathbb{T}_{n-1}|}^2}{\overset{\text{superexp}}{\Longrightarrow}} 0, \tag{5.11}$$

$$I_2 = \frac{1}{\sqrt{|\mathbb{T}_{n-1}|}v_{|\mathbb{T}_{n-1}|}}\sum_{k\in\mathbb{T}_{n-1,p-1}}\big(\varepsilon_{2k+\eta}\mathbb{X}_k - \varepsilon_{2k+\eta}^{(R)}\mathbb{X}_{k,n}^{(r)}\big) \underset{v_{|\mathbb{T}_{n-1}|}^2}{\overset{\text{superexp}}{\Longrightarrow}} 0. \tag{5.12}$$

We need only prove (5.11) and (5.12) for $\eta = 0$, the same proof works for $\eta = 1$.

Proof of (5.11). We have, for all $\alpha > 0$ and R large enough

$$\begin{aligned}
&\mathbb{E}\left(\exp\left(\alpha\sum_{k\in\mathbb{T}_{n-1,p-1}}\big(\varepsilon_{2k} - \varepsilon_{2k}^{(R)}\big)\right)\right)\\
&\quad= \mathbb{E}\left[\prod_{k\in\mathbb{T}_{n-2,p-1}}\exp\big(\alpha\big(\varepsilon_{2k} - \varepsilon_{2k}^{(R)}\big)\big) \times \mathbb{E}\left[\prod_{k\in\mathbb{G}_{n-1}}\exp\big(\alpha\big(\varepsilon_{2k} - \varepsilon_{2k}^{(R)}\big)\big)\Big/\mathcal{F}_{n-1}\right]\right]\\
&\quad= \mathbb{E}\left[\prod_{k\in\mathbb{T}_{n-2,p-1}}\exp\big(\alpha\big(\varepsilon_{2k} - \varepsilon_{2k}^{(R)}\big)\big) \times \prod_{k\in\mathbb{G}_{n-1}}\mathbb{E}\big[\exp\big(\alpha\big(\varepsilon_{2k} - \varepsilon_{2k}^{(R)}\big)\big)/\mathcal{F}_{n-1}\big]\right]\\
&\quad\leq \mathbb{E}\left[\prod_{k\in\mathbb{T}_{n-2,p-1}}\exp\big(\alpha\big(\varepsilon_{2k} - \varepsilon_{2k}^{(R)}\big)\big)\exp\big(|\mathbb{G}_{n-1}|\alpha^2\kappa_R\big)\right]\\
&\quad\leq \exp\big(|\mathbb{T}_{n-1}|\alpha^2\kappa_R\big),
\end{aligned}$$

where hypothesis (N1R) was used to get the first inequality, and the second was obtained by induction. By Chebyshev inequality and the previous calculation applied to $\alpha = \lambda v_{|\mathbb{T}_{n-1}|}/|\mathbb{T}_{n-1}|$, we obtain for all $\delta > 0$

$$\mathbb{P}\left(\frac{1}{\sqrt{|\mathbb{T}_{n-1}|}v_{|\mathbb{T}_{n-1}|}}\sum_{k\in\mathbb{T}_{n-1,p-1}}\big(\varepsilon_{2k} - \varepsilon_{2k}^{(R)}\big) \geq \delta\right) \leq \exp\big(-v_{|\mathbb{T}_{n-1}|}^2\big(\delta\lambda - \kappa_R\lambda^2\big)\big).$$

Optimizing on λ, we obtain

$$\frac{1}{v_{|\mathbb{T}_{n-1}|}^2}\log\mathbb{P}\left(\frac{1}{\sqrt{|\mathbb{T}_{n-1}|}v_{|\mathbb{T}_{n-1}|}}\sum_{k\in\mathbb{T}_{n-1,p-1}}\big(\varepsilon_{2k} - \varepsilon_{2k}^{(R)}\big) \geq \delta\right) \leq -\frac{\delta^2}{4\kappa_R}.$$

Letting n go to infinity and then R go to infinity, we obtain the negligibility in (5.11).

Proof of (5.12). Now, since we have the decomposition

$$\varepsilon_{2k}\mathbb{X}_k - \varepsilon_{2k}^{(R)}\mathbb{X}_{k,n}^{(r)} = \big(\varepsilon_{2k} - \varepsilon_{2k}^{(R)}\big)\mathbb{X}_{k,n}^{(r)} + \varepsilon_{2k}\big(\mathbb{X}_k - \mathbb{X}_{k,n}^{(r)}\big),$$

we introduce the following notation

$$L_n^{(r)} = \sum_{k\in\mathbb{T}_{n-1,p-1}} \varepsilon_{2k}\big(\mathbb{X}_k - \mathbb{X}_{k,n}^{(r)}\big) \quad \text{and} \quad F_n^{(r,R)} = \sum_{k\in\mathbb{T}_{n-1,p-1}} \big(\varepsilon_{2k} - \varepsilon_{2k}^{(R)}\big)\mathbb{X}_{k,n}^{(r)}.$$

To prove (5.12), we will show that for all $r > 0$

$$\frac{L_n^{(r)}}{\sqrt{|\mathbb{T}_{n-1}|}v_{|\mathbb{T}_{n-1}|}} \xrightarrow[v^2_{|\mathbb{T}_{n-1}|}]{\text{superexp}} 0, \tag{5.13}$$

and for all $r > 0$ and all $\delta > 0$

$$\limsup_{R\to\infty}\limsup_{n\to\infty} \frac{1}{v^2_{|\mathbb{T}_{n-1}|}} \log\mathbb{P}\left(\frac{\|F_n^{(r,R)}\|}{v_{|\mathbb{T}_{n-1}|}\sqrt{|\mathbb{T}_{n-1}|}} > \delta\right) = -\infty. \tag{5.14}$$

Let us first deal with $(L_n^{(r)})$. Let its first component be

$$L_{n,1}^{(r)} = \sum_{k\in\mathbb{T}_{n-1,p-1}} \varepsilon_{2k}\big(X_k - X_{k,n}^{(r)}\big).$$

For $\lambda\in\mathbb{R}$, we consider the random sequence $(Z_{n,1}^{(r)})_{n\geq p-1}$ defined by

$$Z_{n,1}^{(r)} = \exp\left(\lambda L_{n,1}^{(r)} - \frac{\lambda^2\phi}{2}\sum_{k\in\mathbb{T}_{n-1,p-1}} X_k^2 \mathbf{1}_{\{\|\mathbb{X}_k\| > r(\sqrt{|\mathbb{T}_{n-1}|}/v_{|\mathbb{T}_{n-1}|})\}}\right)$$

where ϕ appears in (N1). For $h > 0$, we introduce the following event

$$A_{n,1}^{(r)}(h) = \left\{\frac{1}{|\mathbb{T}_{n-1}|}\sum_{k\in\mathbb{T}_{n-1,p-1}} X_k^2 \mathbf{1}_{\{\|\mathbb{X}_k\| > r(\sqrt{|\mathbb{T}_{n-1}|}/v_{|\mathbb{T}_{n-1}|})\}} > h\right\}.$$

Using (N1), we have for all $\delta > 0$

$$\begin{aligned}
&\mathbb{P}\left(\frac{1}{v_{|\mathbb{T}_{n-1}|}\sqrt{|\mathbb{T}_{n-1}|}} L_{n,1}^{(r)} > \delta\right)\\
&\quad\leq \mathbb{P}\big(A_{n,1}^{(r)}(h)\big) + \mathbb{P}\left(Z_{n,1}^{(r)} > \exp\left(\delta\lambda v_{|\mathbb{T}_{n-1}|}\sqrt{|\mathbb{T}_{n-1}|} - \frac{\lambda^2\phi}{2}h|\mathbb{T}_{n-1}|\right)\right)\\
&\quad\leq \mathbb{P}\big(A_{n,1}^{(r)}(h)\big) + \exp\left(-v_{|\mathbb{T}_{n-1}|}\sqrt{|\mathbb{T}_{n-1}|}\left(\delta\lambda - \frac{h\phi\sqrt{|\mathbb{T}_{n-1}|}}{2v_{|\mathbb{T}_{n-1}|}}\lambda^2\right)\right),
\end{aligned} \tag{5.15}$$

where the second term in (5.15) is obtained by conditioning successively on $(\mathcal{G}_i)_{2^{p-1}\leq i\leq |\mathbb{T}_{n-1}|-1}$ and using the fact that

$$\mathbb{E}\left[\exp\left(\lambda\varepsilon_{2^p}\big(X_{2^{p-1}} - X_{2^{p-1}}^{(r)}\big) - \frac{\lambda^2\phi}{2}X_{2^{p-1}}^2\mathbf{1}_{\{\|\mathbb{X}_{2^{p-1}}\| > r(\sqrt{2^{p-1}}/v_{2^{p-1}})\}}\right)\right] \leq 1,$$

which follows from (N1).

From Proposition 5.3, we have for all $h>0$

$$\limsup_{n\to\infty} \frac{1}{v^2_{|\mathbb{T}_{n-1}|}} \log \mathbb{P}\big(A^{(r)}_{n,1}(h)\big) = -\infty,$$

so that taking $\lambda = \delta v_{|\mathbb{T}_{n-1}|}/(h\phi\sqrt{|\mathbb{T}_{n-1}|})$ in (5.15), we are led to

$$\limsup_{n\to\infty} \frac{1}{v^2_{|\mathbb{T}_{n-1}|}} \log \mathbb{P}\left(\frac{L^{(r)}_{n,1}}{v_{|\mathbb{T}_{n-1}|}\sqrt{|\mathbb{T}_{n-1}|}} > \delta\right) \le -\frac{\delta^2}{2h\phi}.$$

Letting $h\to 0$, we obtain that the right hand side of the last inequality goes to $-\infty$.

Proceeding in the same way for $-L^{(r)}_{n,1}$, we deduce that for all $r>0$

$$\frac{L^{(r)}_{n,1}}{v_{|\mathbb{T}_{n-1}|}\sqrt{|\mathbb{T}_{n-1}|}} \overset{\text{superexp}}{\underset{v^2_{|\mathbb{T}_{n-1}|}}{\Longrightarrow}} 0.$$

Now, it is easy to check that the same proof works for the others components of $L^{(r)}_n$. We thus conclude the proof of (5.13).

Let us now consider the term $(F^{(r,R)}_n)$. We follow the same approach as in the proof of (5.13). Let its first component be

$$F^{(r,R)}_{n,1} = \sum_{k\in\mathbb{T}_{n-1,p-1}} \big(\varepsilon_{2k} - \varepsilon^{(R)}_{2k}\big) X^{(r)}_{k,n}.$$

For $\lambda\in\mathbb{R}$, we consider the random sequence $(W^{(r,R)}_{n,1})_{n\ge p-1}$ defined by

$$W^{(r,R)}_{n,1} = \exp\left(\lambda \sum_{k\in\mathbb{T}_{n-1,p-1}} \big(\varepsilon_{2k} - \varepsilon^{(R)}_{2k}\big) X^{(r)}_{k,n} - \frac{\lambda^2\kappa_R}{2} \sum_{k\in\mathbb{T}_{n-1,p-1}} \big(X^{(r)}_{k,n}\big)^2\right),$$

where κ_R appears in (N1R).

Let $h>0$. Consider the following event $B^{(r)}_{n,1}(h) = \{\frac{1}{|\mathbb{T}_{n-1}|}\sum_{k\in\mathbb{T}_{n-1,p-1}} (X^{(r)}_{k,n})^2 > h\}$.

We have for all $\delta>0$,

$$\begin{aligned}
&\mathbb{P}\left(\frac{F^{(r,R)}_{n,1}}{v_{|\mathbb{T}_{n-1}|}\sqrt{|\mathbb{T}_{n-1}|}} > \delta\right) \\
&\quad \le \mathbb{P}\big(B^{(r)}_{n,1}(h)\big) + \mathbb{P}\left(W^{(r,R)}_{n,1} > \exp\left(\delta\lambda v_{|\mathbb{T}_{n-1}|}\sqrt{|\mathbb{T}_{n-1}|} - \frac{\lambda^2\kappa_R}{2}|\mathbb{T}_{n-1}|h\right)\right) \\
&\quad \le \mathbb{P}\big(B^{(r)}_{n,1}(h)\big) + \exp\left(-v_{|\mathbb{T}_{n-1}|}\sqrt{|\mathbb{T}_{n-1}|}\left(\delta\lambda - \frac{h\kappa_R\sqrt{|\mathbb{T}_{n-1}|}}{2v_{|\mathbb{T}_{n-1}|}}\lambda^2\right)\right),
\end{aligned} \tag{5.16}$$

where the second term in (5.16) is obtained by conditioning successively on $(\mathcal{G}_i)_{2^{p-1}\le i\le|\mathbb{T}_{n-1}|-1}$ and using the fact that

$$\mathbb{E}\left[\exp\left(\lambda\big(\varepsilon_{2^p} - \varepsilon^{(R)}_{2^p}\big)X^{(r)}_{2^{p-1}} - \frac{\lambda^2\kappa_R}{2}\big(X^{(r)}_{2^{p-1}}\big)^2\right)\right] \le 1.$$

Since $B^{(r)}_{n,1}(h) \subset \{\frac{1}{|\mathbb{T}_{n-1}|}\sum_{k\in\mathbb{T}_{n-1,p-1}} X_k^2 > h\}$, from Proposition 4.3, we deduce that for h large enough

$$\limsup_{n\to\infty} \frac{1}{b^2_{|\mathbb{T}_{n-1}|}} \log \mathbb{P}\big(B^{(r)}_{n,1}(h)\big) = -\infty,$$

so that choosing $\lambda = \delta v_{|\mathbb{T}_{n-1}|}/(\kappa_R h \sqrt{|\mathbb{T}_{n-1}|})$, we get for all $\delta > 0$

$$\limsup_{n\to\infty} \frac{1}{v_{|\mathbb{T}_{n-1}|}^2} \log \mathbb{P}\left(\frac{F_{n,1}^{(r,R)}}{v_{|\mathbb{T}_{n-1}|}\sqrt{|\mathbb{T}_{n-1}|}} > \delta\right) \le -\frac{\delta^2}{2\kappa_R h}.$$

Letting R go to infinity, we obtain that

$$\limsup_{R\to\infty} \limsup_{n\to\infty} \frac{1}{v_{|\mathbb{T}_{n-1}|}^2} \log \mathbb{P}\left(\frac{F_{n,1}^{(r,R)}}{v_{|\mathbb{T}_{n-1}|}\sqrt{|\mathbb{T}_{n-1}|}} > \delta\right) = -\infty.$$

Now it is easy to check that the same works for $-F_{n,1}^{(r,R)}$ and for the others components of $F_n^{(r,R)}$. We thus conclude that (5.14) holds for all $r > 0$.

Step 3. By application of Theorem 4.2.16 in [12], we find that $(M_n/(v_{|\mathbb{T}_{n-1}|}\sqrt{|\mathbb{T}_{n-1}|}))$ satisfies an MDP on $\mathbb{R}^{2(p+1)}$ with speed $v_{|\mathbb{T}_{n-1}|}^2$ and rate function

$$\widetilde{I}(x) = \sup_{\delta>0} \liminf_{R\to\infty} \inf_{z\in B_{x,\delta}} I_R(z),$$

where I_R is given in (5.10) and $B_{x,\delta}$ denotes the ball $\{z\colon |z-x| < \delta\}$. The identification of the rate function $\widetilde{I} = I_M$, where I_M is given in (5.3) is done easily (see for example [15]), which concludes the proof of Theorem 5.1. □

Proof in case 1. For the proof in case 1, there are no changes in Step 1, and for Step 3, instead of (5.7), (5.8), and (N1), we use Remark 5.4 and (G1). In Step 2, the negligibility in (5.11) comes from the MDP of the i.i.d. sequences $(\varepsilon_{2k} - \varepsilon_{2k}^{(R)})$ since it satisfies the condition, for $\lambda > 0$ and all $R > 0$

$$\mathbb{E}\big(\exp\big(\lambda\big(\varepsilon_{2k} - \varepsilon_{2k}^{(R)}\big)\big)\big) < \infty.$$

The negligibility of $(L_n^{(r)})$ works in the same way. For $(F_n^{(r,R)})$ we will use the MDP for martingale, see Proposition 3.11. For R large enough, we have

$$\begin{aligned}\mathbb{P}\big(\big|X_{k,n}^{(r)}\big(\varepsilon_{2k} - \varepsilon_{2k}^{(R)}\big)\big| > v_{|\mathbb{T}_{n-1}|}\sqrt{|\mathbb{T}_{n-1}|}\big|\mathcal{F}_{k-1}\big) &\le \mathbb{P}\left(\big|\varepsilon_{2k} - \varepsilon_{2k}^{(R)}\big| > \frac{v_{|\mathbb{T}_{n-1}|}^2}{r}\right)\\ &= \mathbb{P}\left(\big|\varepsilon_{2} - \varepsilon_{2}^{(R)}\big| > \frac{v_{|\mathbb{T}_{n-1}|}^2}{r}\right) = 0.\end{aligned}$$

This implies that

$$\limsup_{n\to\infty} \frac{1}{v_{|\mathbb{T}_{n-1}|}^2} \log\Big(|\mathbb{T}_{n-1}| \operatorname*{ess\,sup}_{k\ge 1} \mathbb{P}\big(\big|X_{k,n}^{(r)}\big(\varepsilon_{2k} - \varepsilon_{2k}^{(R)}\big)\big| > v_{|\mathbb{T}_{n-1}|}\sqrt{|\mathbb{T}_{n-1}|}\big|\mathcal{F}_{k-1}\big)\Big) = -\infty.$$

That is condition (D2) in Proposition 3.11.

For all $\gamma > 0$ and all $\delta > 0$, we obtain from Remark 5.4, that

$$\begin{aligned}&\limsup_{n\to\infty} \frac{1}{v_{|\mathbb{T}_{n-1}|}^2} \log \mathbb{P}\left(\frac{1}{|\mathbb{T}_{n-1}|} \sum_{k\in\mathbb{T}_{n-1,p-1}} \big(X_{k,n}^{(r)}\big)^2 \mathbf{1}_{\{|X_{k,n}^{(r)}| > \gamma(\sqrt{|\mathbb{T}_{n-1}|}/v_{|\mathbb{T}_{n-1}|})\}} > \delta\right)\\ &\quad\le \limsup_{n\to\infty} \frac{1}{v_{|\mathbb{T}_{n-1}|}^2} \log \mathbb{P}\left(\frac{1}{|\mathbb{T}_{n-1}|} \sum_{k\in\mathbb{T}_{n-1,p-1}} X_k^2 \mathbf{1}_{\{|X_k| > \gamma(\sqrt{|\mathbb{T}_{n-1}|}/v_{|\mathbb{T}_{n-1}|})\}} > \delta\right) = -\infty.\end{aligned}$$

That is condition (D3) in Proposition 3.11. Finally, from Remark 5.4 and in the same way as in (5.9), it follows that

$$\frac{\langle F^{(r,R)}\rangle_{n,1}}{|\mathbb{T}_{n-1}|} = Q_R \frac{1}{|\mathbb{T}_{n-1}|} \sum_{k\in\mathbb{T}_{n-1,p-1}} \big(X_{k,n}^{(r)}\big)^2 \underset{v^2_{|\mathbb{T}_{n-1}|}}{\overset{\text{superexp}}{\Longrightarrow}} Q_R \ell$$

for some positive constant ℓ, where $Q_R = \mathbb{E}[(\varepsilon_2 - \varepsilon_2^{(R)})^2]$. That is condition (D1) in Proposition 3.11. Moreover, it is clear that Q_R converges to 0 as R goes to infinity. In light of above, we infer from Proposition 3.11 that $(F_{n,1}^{(r,R)}/(v_{|\mathbb{T}_{n-1}|}\sqrt{|\mathbb{T}_{n-1}|}))$ satisfies an MDP on $\mathbb{R}$ of speed $v^2_{|\mathbb{T}_{n-1}|}$ and rate function $I_R(x) = x^2/(2Q_R\ell)$. In particular, this implies that for all $\delta > 0$,

$$\limsup_{n\to\infty} \frac{1}{v^2_{|\mathbb{T}_{n-1}|}} \log \mathbb{P}\bigg(\frac{|F_{n,1}^{(r,R)}|}{v_{|\mathbb{T}_{n-1}|}\sqrt{|\mathbb{T}_{n-1}|}} > \delta\bigg) \le -\frac{\delta^2}{2Q_R\ell},$$

and letting R go to infinity clearly leads to the result. □

5.2.2. *Proof of Theorem* 3.5

The proof works in case 1 and in case 2. From (2.7), we have

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

From Proposition 4.1, we obtain that

$$\frac{\Sigma_n}{|\mathbb{T}_n|} = I_2 \otimes \frac{S_n}{|\mathbb{T}_n|} \underset{v^2_{|\mathbb{T}_n|}}{\overset{\text{superexp}}{\Longrightarrow}} I_2 \otimes L. \tag{5.17}$$

According to Lemma 4.1 of [26], together with (5.17), we deduce that

$$|\mathbb{T}_{n-1}|\Sigma_{n-1}^{-1} \underset{v^2_{|\mathbb{T}_{n-1}|}}{\overset{\text{superexp}}{\Longrightarrow}} I_2 \otimes L^{-1}. \tag{5.18}$$

From Theorem 5.1, (5.18) and the contraction principle [12], we deduce that the sequence $(\sqrt{|\mathbb{T}_{n-1}|}(\hat{\theta}_n - \theta)/v_{|\mathbb{T}_{n-1}|})_{n\ge1}$ satisfies the MDP with rate function I_θ given by (3.3).

5.3. *Proof of Theorem* 3.7

For the proof of Theorem 3.7, case 1 is an easy consequence of the classical MDP for i.i.d.r.v. applied to the sequence $(\varepsilon_{2k}^2 + \varepsilon_{2k+1}^2)$. For case 2, we will use Proposition 3.11, rather than Puhalskii's Theorem 3.12.

We will prove that the sequence$(\sqrt{|\mathbb{T}_{n-1}|}(\sigma_n^2 - \sigma^2)/v_{|\mathbb{T}_{n-1}|})$ satisfies the MDP. For that, we will prove that conditions (D1), (D2) and (D3) of Proposition 3.11 are verified. Let us consider the $\mathcal{G}_n$-martingale $(Q_n)_{n\ge 2^{p-1}}$ given by

$$Q_n = \sum_{k=2^{p-1}}^{n} \nu_k, \quad \text{where } \nu_k = \varepsilon_{2k}^2 + \varepsilon_{2k+1}^2 - 2\sigma^2.$$

It is easy to see that its predictable quadratic variation is given by

$$\langle Q\rangle_n = \sum_{k=2^{p-1}}^{n} \mathbb{E}\big[\nu_k^2/\mathcal{G}_{k-1}\big] = \big(n - 2^{p-1} + 1\big)\big(2\tau^4 - 4\sigma^4 + 2\nu^2\big),$$

which immediately implies that

$$\frac{\langle Q\rangle_n}{n} \underset{v_n^2}{\overset{\text{superexp}}{\Longrightarrow}} 2\tau^4 - 4\sigma^4 + 2\nu^2,$$

ensuring condition (D1) in Proposition 3.11.

Next, for $B > 0$ large enough, we have for $a > 2$ (in (Ea)), and some positive constant c

$$\mathbb{P}\left(\frac{1}{n}\sum_{k=2^{p-1}}^{n}|\nu_k|^a > B\right) \leq 3\max_{\eta\in\{0,1\}}\left\{\mathbb{P}\left(\frac{1}{n}\sum_{k=2^{p-1}}^{n}|\varepsilon_{2k+\eta}|^{2a} > \frac{B}{3c}\right)\right\}.$$

From hypothesis (Ea) and since B is large enough, we obtain for a suitable $t > 0$ via the Chernoff inequality and several successive conditionings on $(\mathcal{G}_n)$, for $\eta \in \{0,1\}$

$$\mathbb{P}\left(\frac{1}{n}\sum_{k=2^{p-1}}^{n}|\varepsilon_{2k+\eta}|^{2a} > \frac{B}{3c}\right) \leq \exp\left(-tn\left(\frac{B}{3c} - \log E\right)\right) \leq \exp\left(-tc'n\right),$$

where c, c' are positive generic constants. Therefore, for $B > 0$ large enough, we deduce that

$$\limsup_{n\to\infty}\frac{1}{n}\log\mathbb{P}\left(\frac{1}{n}\sum_{k=2^{p-1}}^{n}|\nu_k|^a > B\right) < 0,$$

and this implies (see e.g. [26]) exponential Lindeberg condition, that is for all $r > 0$

$$\frac{1}{n}\sum_{k=2^{p-1}}^{n}\nu_k^2\mathbf{1}_{\{|\nu_k|>r(\sqrt{n}/v_n)\}} \overset{\text{superexp}}{\underset{v_n^2}{\Longrightarrow}} 0.$$

That is condition (D3) in Proposition 3.11.

Now, for all $k \in \mathbb{N}$ and a suitable $t > 0$ we have

$$\begin{aligned}
\mathbb{P}\left(|\nu_k| > v_n\sqrt{n}/\mathcal{G}_{k-1}\right) &\leq \sum_{\eta=0}^{1}\mathbb{P}\left(\left|\varepsilon_{2k+\eta}^2 - \sigma^2\right| > \frac{v_n\sqrt{n}}{2}\Big/\mathcal{G}_{k-1}\right) \\
&\leq \exp\left(\frac{-tv_n\sqrt{n}}{2}\right)\sum_{\eta=0}^{1}\mathbb{E}\left[\exp\left(t\left|\varepsilon_{2k+\eta}^2 - \sigma^2\right|\right)/\mathcal{G}_{k-1}\right] \\
&\leq 2E'\exp\left(\frac{-tv_n\sqrt{n}}{2}\right),
\end{aligned}$$

where from hypothesis (Na), E' is finite and positive. We are thus led to

$$\frac{1}{v_n^2}\log\left(n\operatorname*{ess\,sup}_{k\in\mathbb{N}^*}\mathbb{P}\left(|\nu_k| > v_n\sqrt{n}/\mathcal{G}_{k-1}\right)\right) \leq \frac{\log(2E'n)}{v_n^2} - \frac{t\sqrt{n}}{v_n},$$

and consequently, letting n go to infinity, we get the condition (D2) in Proposition 3.11.

Now, applying Proposition 3.11, we conclude that $(Q_n/(v_n\sqrt{n}))_{n\geq 0}$ satisfies the MDP with speed v_n^2 and rate function

$$I_Q(x) = \frac{x^2}{4(\tau^4 - 2\sigma^4 + 2\nu^2)}.$$

Applying the above to $|\mathbb{T}_{n-1}|$ and using the contraction principle (see e.g. [12]), we deduce that the sequence

$$\frac{\sqrt{|\mathbb{T}_{n-1}|}}{v_{|\mathbb{T}_{n-1}|}}\left(\sigma_n^2 - \sigma^2\right) = \frac{Q_{|\mathbb{T}_{n-1}|}}{2v_{|\mathbb{T}_{n-1}|}\sqrt{|\mathbb{T}_{n-1}|}}$$

satisfies a MDP with speed $v_{|\mathbb{T}_{n-1}|}^2$ and rate function I_{σ^2} given by (3.4).

We obtain as in the proof of the first part, with a slight modification, that the sequence $(|\mathbb{T}_{n-1}|(\rho_n - \rho)/v_{|\mathbb{T}_{n-1}|})$ satisfies a MDP with speed $v^2_{|\mathbb{T}_{n-1}|}$ and rate function I_ρ given by (3.5).

5.4. *Proof of Theorem* 3.10

Here also the proof works for the two cases.

Let us first deal with $\hat{\sigma}_n$. We have

$$\hat{\sigma}_n^2 - \sigma^2 = \left(\hat{\sigma}_n^2 - \sigma_n^2\right) + \left(\sigma_n^2 - \sigma^2\right).$$

From (4.22) and (4.30), we easily deduce that $\sigma_n^2 \underset{v^2_{|\mathbb{T}_{n-1}|}}{\overset{\text{superexp}}{\Longrightarrow}} \sigma^2$ in case 1 and in case 2. Thus, it is enough to prove that $\hat{\sigma}_n^2 - \sigma_n^2 \underset{v^2_{|\mathbb{T}_{n-1}|}}{\overset{\text{superexp}}{\Longrightarrow}} 0$. Let $\theta^{(0)} = (a_0, a_1, \ldots, a_p)^t$, $\theta^{(1)} = (b_0, b_1, \ldots, b_p)^t$, $\hat{\theta}_n^{(0)} = (\hat{a}_{0,n}, \hat{a}_{1,n}, \ldots, \hat{a}_{p,n})^t$, $\hat{\theta}_n^{(1)} = (\hat{b}_{0,n}, \hat{b}_{1,n}, \ldots, \hat{b}_{p,n})^t$.

Let us introduce the following function f defined for x and z in $\mathbb{R}^{p+1}$ by

$$f(x, z) = \left(x_1 - z_1 - \sum_{i=2}^{p+1} z_i x_i\right)^2,$$

where x_i and z_i denote respectively the ith component of x and z. One can observe that

$$\begin{aligned}\hat{\sigma}_n^2 - \sigma_n^2 = {}& \frac{1}{2|\mathbb{T}_{n-1}|} \sum_{k \in \mathbb{T}_{n-1,p-1}} \left\{ f\left(\mathbb{X}_{2k}, \hat{\theta}_n^{(0)}\right) - f\left(\mathbb{X}_{2k}, \theta^{(0)}\right)\right\} \\ &+ \frac{1}{2|\mathbb{T}_{n-1}|} \sum_{k \in \mathbb{T}_{n-1,p-1}} \left\{ f\left(\mathbb{X}_{2k+1}, \hat{\theta}_n^{(1)}\right) - f\left(\mathbb{X}_{2k+1}, \theta^{(1)}\right)\right\}.\end{aligned}$$

By the Taylor–Lagrange formula, $\forall x \in \mathbb{R}^{p+1}$ and $\forall z, z' \in \mathbb{R}^{p+1}$, one can find $\lambda \in (0, 1)$ such that

$$f\left(x, z'\right) - f(x, z) = \sum_{j=1}^{p+1} \left(z'_j - z_j\right) \partial_{z_j} f\left(x, z + \lambda\left(z' - z\right)\right).$$

Let the function g be defined by

$$g(x, z) = x_1 - z_1 - \sum_{j=2}^{p+1} z_j x_j.$$

Observing that

$$\begin{cases} \frac{\partial f}{\partial z_1}(x, z) = -2g(x, z), \\ \frac{\partial f}{\partial z_j}(x, z) = -2x_j g(x, z) \quad \forall j \geq 2, \end{cases}$$

we get easily that $|\frac{\partial f}{\partial z_j}(x, z)| \leq 4(1 + \|z\|)(1 + \|x\|^2)$ for all $j \geq 1$, and this implies

$$\left|f\left(x, z'\right) - f(x, z)\right| \leq c\left\|z' - z\right\|\left(1 + \|z\| + \left\|z' - z\right\|\right)\left(1 + \|x\|^2\right)$$

for some positive constant c. Now, applying the above to $f(\mathbb{X}_{2k}, \hat{\theta}_n^{(0)}) - f(\mathbb{X}_{2k}, \theta^{(0)})$ and to $f(\mathbb{X}_{2k+1}, \hat{\theta}_n^{(1)}) - f(\mathbb{X}_{2k+1}, \theta^{(1)})$, we deduce easily that

$$\left|\hat{\sigma}_n^2 - \sigma_n^2\right| \leq c\|\hat{\theta}_n - \theta\|\big(1 + \|\theta\| + \|\hat{\theta}_n - \theta\|\big)\frac{1}{|\mathbb{T}_{n-1}|}\sum_{k\in\mathbb{T}_{n-1,p-1}}\big(1 + \|\mathbb{X}_k\|^2\big)$$

for some positive constant c. From the MDP of $\hat{\theta}_n - \theta$, we infer that

$$\|\hat{\theta}_n - \theta\| \underset{v^2_{|\mathbb{T}_{n-1}|}}{\overset{\text{superexp}}{\Longrightarrow}} 0. \tag{5.19}$$

Form Proposition 4.3 we deduce that

$$\frac{1}{|\mathbb{T}_{n-1}|}\sum_{k\in\mathbb{T}_{n-1,p-1}}\big(1 + \|\mathbb{X}_k\|^2\big) \underset{v^2_{|\mathbb{T}_{n-1}|}}{\overset{\text{superexp}}{\Longrightarrow}} 1 + \operatorname{Tr}(\Lambda). \tag{5.20}$$

We thus conclude via (5.19) and (5.20) that

$$\hat{\sigma}_n^2 - \sigma_n^2 \underset{v^2_{|\mathbb{T}_{n-1}|}}{\overset{\text{superexp}}{\Longrightarrow}} 0.$$

This ends the proof for $\hat{\sigma}_n$. The proof for $\hat{\rho}_n$ is very similar and uses hypotheses(G2′) and (N2′) to get inequalities similar to (4.22) and (4.30).

Acknowledgments

The authors thank Arnaud Guillin for all his advice and suggestions during the preparation of this work. We are grateful to the anonymous referee for her/his valuable comments which improve the presentation of this work.

References

[1] R. Adamczak and P. Milos. CLT for Ornstein–Uhlenbeck branching particle system. Preprint. Available at arXiv:1111.4559.
[2] I. V. Basawa and J. Zhou. Non-Gaussian bifurcating models and quasi-likelihood estimation. *J. Appl. Probab.* **41** (2004) 55–64. MR2057565
[3] B. Bercu, B. de Saporta and A. Gégout-Petit. Asymtotic analysis for bifurcating autoregressive processes via martingale approach. *Electron. J. Probab.* **14** (2009) 2492–2526. MR2563249
[4] B. Bercu and A. Touati. Exponential inequalities for self-normalized martingales with applications. *Ann. Appl. Probab.* **18** (2008) 1848–1869. MR2462551
[5] V. Bitseki Penda, H. Djellout and A. Guillin. Deviation inequalities, moderate deviations and some limit theorems for bifurcating Markov chains with application. *Ann. Appl. Probab.* **24** (2014) 235–291. MR3161647
[6] R. Cowan and R. G. Staudte. The bifurcating autoregressive model in cell lineage studies. *Biometrics* **42** (1986) 769–783.
[7] V. H. de la Peña, T. L. Lai and Q.-M. Shao. *Self-Normalized Processes. Limit Theory and Statistical Applications*. *Probability and Its Applications (New York)*. Springer-Verlag, Berlin, 2009. MR2488094
[8] B. de Saporta, A. Gégout-Petit and L. Marsalle. Parameters estimation for asymmetric bifurcating autoregressive processes with missing data. *Electron. J. Stat.* **5** (2011) 1313–1353. MR2842907
[9] B. de Saporta, A. Gégout-Petit and L. Marsalle. Asymmetry tests for bifurcating auto-regressive processes with missing data. *Statist. Probab. Lett.* **82** (2012) 1439–1444. MR2929798
[10] J. F. Delmas and L. Marsalle. Detection of cellular aging in a Galton–Watson process. *Stochastic Process. Appl.* **120** (2010) 2495–2519. MR2728175
[11] A. Dembo. Moderate deviations for martingales with bounded jumps. *Electron. Comm. Probab.* **1** (1996) 11–17. MR1386290
[12] A. Dembo and O. Zeitouni. *Large Deviations Techniques and Applications*, 2nd edition. Springer, New York, 1998. MR1619036
[13] H. Djellout. Moderate deviations for martingale differences and applications to ϕ-mixing sequences. *Stoch. Stoch. Rep.* **73** (2002) 37–63. MR1914978
[14] H. Djellout, A. Guillin and L. Wu. Moderate deviations of empirical periodogram and non-linear functionals of moving average processes. *Ann. Inst. Henri Poincaré Probab. Stat.* **42** (2006) 393–416. MR2242954
[15] H. Djellout and A. Guillin. Large and moderate deviations for moving average processes. *Ann. Fac. Sci. Toulouse Math.* (6) **10** (2001) 23–31. MR1928987

[16] N. Gozlan. Integral criteria for transportation-cost inequalities. *Electron. Comm. Probab.* **11** (2006) 64–77. MR2231734

[17] N. Gozlan and C. Léonard. A large deviation approach to some transportation cost inequalities. *Probab. Theory Related Fields* **139** (2007) 235–283. MR2322697

[18] J. Guyon. Limit theorems for bifurcating Markov chains. Application to the detection of cellular aging. *Ann. Appl. Probab.* **17** (2007) 1538–1569. MR2358633

[19] R. M. Huggins and I. V. Basawa. Extensions of the bifurcating autoregressive model for cell lineage studies. *J. Appl. Probab.* **36** (1999) 1225–1233. MR1746406

[20] R. M. Huggins and I. V. Basawa. Inference for the extended bifurcating autoregressive model for cell lineage studies. *Aust. N. Z. J. Stat.* **42** (2000) 423–432. MR1802966

[21] S. Y. Hwang, I. V. Basawa and I. K. Yeo. Local asymptotic normality for bifurcating autoregressive processes and related asymptotic inference. *Stat. Methodol.* **6** (2009) 61–69. MR2655539

[22] M. Ledoux. *The Concentration of Measure Phenomenon*. *Mathematical Surveys and Monographs* **89**. American Mathematical Society, Providence, RI, 2001. MR1849347

[23] P. Massart. *Concentration Inequalities and Model Selection*. *Lecture Notes in Mathematics* **1896**. Springer, Berlin, 2007. MR2319879

[24] A. Puhalskii. Large deviations of semimartingales: A maxingale problem approach. I. Limits as solutions to a maxingale problem. *Stoch. Stoch. Rep.* **61** (1997) 141–243. MR1488137

[25] J. Worms. Moderate deviations for stable Markov chains and regression models. *Electron. J. Probab.* **4** (1999) 28 pp. MR1684149

[26] J. Worms. Moderate deviations of some dependent variables. I. Martingales. *Math. Methods Statist.* **10** (2001) 38–72. MR1841808

[27] J. Worms. Moderate deviations of some dependent variables. II. Some kernel estimators. *Math. Methods Statist.* **10** (2001) 161–193. MR1851746

[28] J. Zhou and I. V. Basawa. Least-squares estimation for bifurcating autoregressive processes. *Statist. Probab. Lett.* **74** (2005) 77–88. MR2189078

[29] J. Zhou and I. V. Basawa. Maximum likelihood estimation for a first-order bifurcating autoregressive process with exponential errors. *J. Time Series Anal.* **26** (2005) 825–842. MR2203513