



HAL
open science

Moderate deviations in a class of stable but nearly unstable processes

Frédéric Proïa

► **To cite this version:**

Frédéric Proïa. Moderate deviations in a class of stable but nearly unstable processes. Journal of Statistical Planning and Inference, 2020, 208, pp.66-81. 10.1016/j.jspi.2020.01.009 . hal-02123384

HAL Id: hal-02123384

<https://hal.science/hal-02123384v1>

Submitted on 22 Aug 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

MODERATE DEVIATIONS IN A CLASS OF STABLE BUT NEARLY UNSTABLE PROCESSES

FRÉDÉRIC PROÏA

ABSTRACT. We consider a stable but nearly unstable autoregressive process of any order. The bridge between stability and instability is expressed by a time-varying companion matrix A_n with spectral radius $\rho(A_n) < 1$ satisfying $\rho(A_n) \rightarrow 1$. In that framework, we establish a moderate deviation principle for the empirical covariance only relying on the elements of A_n through $1 - \rho(A_n)$ and, as a by-product, we establish a moderate deviation principle for the OLS estimator when Γ , the renormalized asymptotic variance of the process, is invertible. Finally, when Γ is singular, we also provide a compromise in the form of a moderate deviation principle for a penalized version of the estimator. Our proofs essentially rely on [truncations and deviations of \$m_n\$ -dependent sequences, with an unbounded rate \(\$m_n\$ \)](#).

2010 MATHEMATICS SUBJECT CLASSIFICATION. 60F10, 62M10, 62G05, 60G10.

1. INTRODUCTION AND ASSUMPTIONS

Unit root issues have long been crucial in time series econometrics and have therefore focused a great deal of research studies. This sudden demarcation between stability and instability is responsible for many inference problems in linear time series (see Brockwell and Davis [4] for a detailed overview of the linear stochastic processes). The remarkable works of Chan and Wei [7] encompass, in a much more general context, the now well-known fact that the least squares estimator is \sqrt{n} -consistent with Gaussian behavior when the underlying autoregressive process is stable, whereas it is n -consistent with asymmetrical distribution when the process is unstable. This rather abrupt change in the rate of convergence and in the asymptotic distribution certainly motivated the wide range of unit root testing procedures, but it also paved the way for studies based on time-varying coefficients. In a nearly unstable autoregressive process, we do not focus on a parameter θ satisfying $|\theta| < 1$ or $|\theta| = 1$ but, instead, the parameter is considered as [a sequence \$\(\theta_n\)\$ such that \$|\theta_n| < 1\$ and \$|\theta_n| \rightarrow 1\$ as \$n \rightarrow +\infty\$](#) . This sample size dependent structure allows a continuity between stability and instability. For example, Phillips and Magdalinos [20] treat the case where the coefficient is in a $O(\kappa_n^{-1})$ neighborhood of the unit root with $\kappa_n = n^\alpha = o(n)$. [Amongst other results, they prove a central limit theorem for the estimator at the rate \$\sqrt{n\kappa_n}\$, thereby making a bridge between the stable rate \$\sqrt{n}\$ and the unstable rate \$n\$](#) . In the same vein, let us also mention the work of Chan and Wei [6], natural generalizations like the study of Phillips and Lee [19] related to vector autoregressions, [or the recent unified theory of Buchmann and Chan \[5\]](#), focused on nearly unstable autoregressive processes. Our paper is precisely based on the latter topic, in a sense that will be precised in good time.

Key words and phrases. Nearly unstable autoregressive process, Moderate deviation principle, OLS estimation, Asymptotic behavior, Unit root.

Given a parametric generating process, the precision of the estimation is usually assessed by its rate of convergence and the deviations can be seen as a natural continuation after a central limit theorem or even a law of iterated logarithm. Roughly speaking, they may be used to estimate the exponential decline of the probability of tail events [associated with the precision of the estimation](#). We refer to [Dembo and Zeitouni \[8\]](#) regarding the [mathematical formalization](#). Since the 1980s, numerous authors have worked on large and/or moderate deviations in a time series context under many and varied hypotheses. Without claiming to be exhaustive, one can mention [the studies of Donsker and Varadhan \[10\]](#) and [Bercu *et al.* \[2\]](#) on stationary Gaussian processes and quadratic forms, [the paper of Worms \[21\]](#) on Markov chains and regression models [and the one of Bercu \[1\]](#) on first-order Gaussian stable, unstable and explosive processes. [One can also mention the works of Mas and Menneteau \[15\]](#) on Hilbertian processes, [Djellout *et al.* \[9\]](#) on non-linear functionals of moving average processes, [Wu and Zhao \[22\]](#) on stationary non-linear processes, [Miao and Shen \[16\]](#) on general autoregressive processes or, more recently, [Bitseki Penda *et al.* \[3\]](#) on first-order processes with correlated errors. [All the references inside may complete this concise list](#).

In [this](#) paper, we investigate the moderate deviations of the estimate in stable but nearly unstable autoregressions. This can be seen as a full generalization of the recent work of [Miao, Wang and Yang \[17\]](#), focused on the univariate case. [Our proofs essentially rely on truncations and deviations of \$m_n\$ -dependent sequences where the rate \$\(m_n\)\$ is unbounded](#). The main technical contributions are twofold. On the one hand, [expressing the nearly instability directly through the sequence of spectral radii of the companion matrix seems, to the best of our knowledge, a new approach having many advantages](#). For example the authors of the recent paper [\[5\]](#) introduce a perturbation in the Jordan canonical form of the model (see [Thm. 2.1](#)) which is a powerful idea to deal with the subject of their study, but somehow unnecessarily complex for ours. On the other hand, from a purely technical point of view, [unbounded truncations have already been used to get moderate deviations \(see *e.g.* \[18\] and \[17\]\)](#), but we will see [that the vector case treated here and the specific features of the model cannot be adapted as easily to the existing tools](#). As a consequence, we need to redevelop a full Gärtner-Ellis reasoning to establish the deviations of our unbounded vector truncations. [This quite general strategy might inspire future similar studies](#).

For a fixed $n \geq 1$, let the process [be](#) given for some $p \geq 1$ and $k \in \{1, \dots, n\}$ by

$$X_{n,k} = \sum_{i=1}^p \theta_{n,i} X_{n,k-i} + \varepsilon_k$$

where $(\varepsilon_k)_k$ is a sequence of zero-mean i.i.d. random variables. In an equivalent way, we can consider the vector expression

$$(1.1) \quad \Phi_{n,k} = A_n \Phi_{n,k-1} + E_k$$

where $E_k = (\varepsilon_k, 0, \dots, 0)^T$ is a p -vectorial noise, $\Phi_{n,k} = (X_{n,k}, \dots, X_{n,k-p+1})^T$ and

$$(1.2) \quad A_n = \begin{pmatrix} \theta_{n,1} & \theta_{n,2} & \cdots & \theta_{n,p} \\ & I_{p-1} & & 0 \\ & & & 2 \end{pmatrix}$$

is the $p \times p$ companion matrix of the autoregressive process. If $(E_k)_k$ has a finite variance, it is well-known that $(\Phi_{k,n})_k$ is a second-order stationary process having the causal form

$$(1.3) \quad \Phi_{n,k} = \sum_{\ell=0}^{+\infty} A_n^\ell E_{k-\ell}$$

when $\rho(A_n) < 1$, that is, when the largest modulus of its eigenvalues is less than 1 (see *e.g.* Thm. 11.3.1 of [4] and the fact that each eigenvalue of A_n is the inverse of a zero of the autoregressive polynomial of the process). Since $(\varepsilon_k)_k$ is an i.i.d. sequence, the process is strictly stationary with mean zero and variance given by

$$(1.4) \quad \Gamma_n = \sigma^2 \sum_{\ell=0}^{+\infty} A_n^\ell K_p (A_n^T)^\ell$$

where, for convenience, we will denote in the whole study

$$(1.5) \quad K_p = \begin{pmatrix} 1 & 0 \\ 0 & 0_{p-1} \end{pmatrix} \quad \text{and} \quad U_p = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

the $p \times p$ matrix with 1 at the top left and 0 elsewhere, and its first column standing for the first vector of the canonical basis of \mathbb{R}^p . As a consequence of the causal expression above, the initial vector $\Phi_{n,0}$ is not arbitrary and has to share the distribution of the process. This also implies the relation

$$(1.6) \quad \Gamma_n = A_n \Gamma_n A_n^T + \sigma^2 K_p.$$

As will be largely developed throughout the study, Γ_n is finite for all $n \geq 1$ but, as n increases, $\|\Gamma_n\| \rightarrow +\infty$. The keystone matrix Γ obtained after a correct standardization of Γ_n is the renormalized asymptotic variance of the process. Before we start, we define a matrix that will also prove to be crucial to our results,

$$(1.7) \quad B_n = I_{p^2} - A_n \otimes A_n.$$

We are now going to introduce and comment the hypotheses that will be needed, though not always simultaneously, in the whole paper. Section 2 is devoted to our main results : two [statements related to the moderate deviations of the empirical covariance and the OLS estimator](#), a set of explicit examples and some additional comments and conclusions. Finally, in Section 3 divided into numerous subsections, we will prove all our results, step by step. [For clarity purposes, the reader will be referred to a supplementary file for some of the technical proofs.](#)

Remark. We denote by $\|\cdot\|$ the Euclidean vector norm and by $\|\|\cdot\|\|$ the spectral matrix norm. Other norms may be used, in which case an appropriated subscript is added. Moreover, we will always denote by $\langle \cdot, \cdot \rangle$ the usual inner product of the Euclidean space \mathbb{R}^d for any $d \geq 1$. We write M^\dagger for the Moore-Penrose pseudo-inverse of any matrix M , whose definition and properties may be found in Sec. 0 of [12].

1.1. **Hypotheses.** First of all, we present the hypotheses that we retain.

(H₁) *Gaussian integrability condition.* There exists $\alpha > 0$ such that

$$\mathbb{E}[e^{\alpha \varepsilon_1^2}] < +\infty$$

where ε_1 represents the zero-mean i.i.d. sequence $(\varepsilon_k)_k$ of variance $\sigma^2 > 0$ and fourth-order moment $\tau^4 > 0$.

(H₂) *Convergence of the companion matrix.* There exists a $p \times p$ matrix A such that

$$\lim_{n \rightarrow +\infty} A_n = A$$

with distinct eigenvalues $0 < |\lambda_p| \leq \dots \leq |\lambda_1| = \rho(A)$, and the top right element of A is non-zero.

(H₃) *Spectral radius of the companion matrix.* For all $n \geq 1$, $\rho(A_n) < 1$. In addition,

$$\lim_{n \rightarrow +\infty} \rho(A_n) = \rho(A) = 1.$$

(H₄) *Renormalization.* We have the convergences

$$\lim_{n \rightarrow +\infty} \frac{B_n^{-1}}{\|B_n^{-1}\|_*} = H \quad \text{and} \quad \lim_{n \rightarrow +\infty} (1 - \rho(A_n)) \|B_n^{-1}\|_* = h$$

for some matrix norm, where H is a $p^2 \times p^2$ non-zero matrix and $h > 0$.

(H₅) *Moderate deviations.* The moderate deviations scale (b_n) satisfies

$$\lim_{n \rightarrow +\infty} b_n = +\infty \quad \text{and} \quad \lim_{n \rightarrow +\infty} \frac{\sqrt{n} (1 - \rho(A_n))^{\frac{3}{2} + \eta}}{b_n} = +\infty$$

for a small $\eta > 0$.

1.2. Comments on the hypotheses. First, conceding in (H₂) that the limiting matrix has distinct eigenvalues is a matter of simplification of the reasonings. Indeed, A_n turns out to be diagonalizable for a sufficiently large n , and, as a companion matrix, it is well-known that the change of basis is done *via* a Vandermonde matrix having numerous nice properties (more details are given in Section 3.1, and a discussion on the case of multiple eigenvalues is provided in Section 2.3). The top right element of A_n is $\theta_{n,p}$. So, assuming in (H₂) that $\theta_{n,p} \rightarrow 0$ ensures that the limit process is still of order p **and that 0 cannot be an eigenvalue of A , since $\det(A) = (-1)^{p+1} \theta_p$** . Moreover, note that, in (H₄), the invertibility of B_n for all n is guaranteed by (H₃). Indeed, $\rho(A_n \otimes A_n) = \rho^2(A_n) < 1$ (see *e.g.* Lem. 5.6.10 and Cor. 5.6.16 of [13]). In addition, we obviously have, for all $\ell \geq 0$,

$$\rho(A_n^\ell) = \rho^\ell(A_n) \leq \|A_n^\ell\|$$

so that we get

$$(1.8) \quad \frac{1}{1 - \rho(A_n)} \leq \sum_{\ell=0}^{+\infty} \|A_n^\ell\| = L_n$$

giving a lower bound for L_n . Similarly,

$$(1.9) \quad \frac{1}{(1 - \rho(A_n))^2} \leq \sum_{\ell=0}^{+\infty} (\ell + 1) \|A_n^\ell\| = M_n.$$

However, an exact upper bound for these sums may be difficult to reach and may require stringent conditions on the elements of A_n . We refer the reader to Lemma 3.1 where, under (H₂) and (H₃), some asymptotic upper bounds are established. We also refer to Section 2.2

where the explicit calculations in terms of some examples shall help to understand the rates involved in the hypotheses. Now for a fixed $n \geq 1$, let

$$\mu_n = \rho(A_n) + \frac{1 - \rho(A_n)}{2} = \frac{\rho(A_n) + 1}{2}.$$

Clearly, $\rho(A_n) < \mu_n < 1$. Hence, according to Prop. 2.3.15 of [11], for all $n \geq 0$, there exists a constant $c_n > 0$ such that, for all $\ell \geq 0$, $\|A_n^\ell\| \leq c_n \mu_n^\ell$ so that

$$L_n \leq \frac{c_n}{1 - \mu_n} < +\infty \quad \text{and} \quad M_n \leq \frac{c_n}{(1 - \mu_n)^2} < +\infty.$$

Letting n tend to infinity, it follows from (H₃) and (H₄) that

$$(1.10) \quad \lim_{n \rightarrow +\infty} \|B_n^{-1}\| = \lim_{n \rightarrow +\infty} L_n = \lim_{n \rightarrow +\infty} M_n = +\infty.$$

Finally, it will be established in good time that there is a limiting matrix Γ such that

$$(1.11) \quad \lim_{n \rightarrow +\infty} \frac{\Gamma_n}{\|B_n^{-1}\|_*} = \Gamma$$

where $\|\cdot\|_*$ is the matrix norm of (H₄).

Remark. To facilitate the reading, we consider from now on that the matrix norm $\|\cdot\|_*$ is identified in (H₄), and we will only note $\|\cdot\|$ in what follows.

2. MAIN RESULTS

This section contains two [statements](#) that constitute the main results of the paper. The first of them is quite long to establish and will need numerous technical lemmas, but the second one will essentially be deduced as a corollary of the first one. Subsequently, we provide some explicit examples for a better understanding and an easier interpretation of the hypotheses together with some graphics showing the evolution of the processes and the estimation of the autoregressive parameter. At the end of the section, we discuss the case of multiple eigenvalues. But, [first](#), let us recall the definition of the large and moderate deviation principles (see Sec. 1.2 of [8] for more details). [In what follows](#), a speed is considered as a positive sequence increasing to infinity.

Definition. A sequence of random variables $(U_n)_n$ on a topological space $(\mathcal{X}, \mathcal{B})$ satisfies a large deviation principle (LDP) with speed (a_n) and rate I if there is a lower semicontinuous mapping $I : \mathcal{X} \rightarrow \bar{\mathbb{R}}^+$ such that :

- for any closed set $F \in \mathcal{B}$,

$$\limsup_{n \rightarrow +\infty} \frac{1}{a_n} \ln \mathbb{P}(U_n \in F) \leq - \inf_{x \in F} I(x),$$

- for any open set $G \in \mathcal{B}$,

$$- \inf_{x \in G} I(x) \leq \liminf_{n \rightarrow +\infty} \frac{1}{a_n} \ln \mathbb{P}(U_n \in G).$$

In particular, if the infimum of I coincides on the interior H° and the closure \bar{H} of some $H \in \mathcal{B}$, then

$$\lim_{n \rightarrow +\infty} \frac{1}{a_n} \ln \mathbb{P}(U_n \in H) = - \inf_{x \in H} I(x).$$

Definition. A sequence of random variables $(V_n)_n$ on a topological space $(\mathcal{X}, \mathcal{B})$ satisfies a moderate deviation principle (MDP) with speed (b_n^2) and rate I if there is a speed (v_n) with $\frac{v_n}{b_n} \rightarrow +\infty$ such that $(\frac{v_n}{b_n} V_n)_n$ satisfies a large deviation principle of speed (b_n^2) and rate I .

2.1. Moderate deviations. We now consider an observable trajectory $X_{n,-p+1}, \dots, X_{n,n}$ for some fixed $n \geq 1$, and use it to provide an estimation of the parameter. It is well-known that the ordinary least squares (OLS) estimator of $\theta_n = (\theta_{n,1}, \dots, \theta_{n,p})^T$ is given by

$$(2.1) \quad \hat{\theta}_n = S_{n-1}^{-1} \sum_{k=1}^n \Phi_{n,k-1} X_{n,k} \quad \text{where} \quad S_{n-1} = \sum_{k=1}^n \Phi_{n,k-1} \Phi_{n,k-1}^T.$$

The first result is dedicated to the empirical variance $\frac{S_n}{n}$. In this regard, denote

$$(2.2) \quad \Upsilon = \sigma^2 H (K_p \otimes \Gamma^A + U_p \otimes \Gamma^A \otimes U_p^T + U_p^T \otimes \Gamma^A \otimes U_p + \Gamma^A \otimes K_p) H^T$$

where H comes from (H_4) , K_p and U_p are given in (1.5) and $\Gamma^A = A \Gamma A^T$ which, in turn, is built on Γ in (1.11) and A , the limit companion matrix introduced in (H_2) .

Theorem 2.1. *Under hypotheses (H_1) – (H_5) , the sequence*

$$\left(\frac{\sqrt{n} (1 - \rho(A_n))^{\frac{3}{2}}}{b_n} \text{vec} \left(\frac{1}{n} \sum_{k=1}^n (\Phi_{n,k} \Phi_{n,k}^T - \Gamma_n) \right) \right)_{n \geq 1}$$

satisfies an LDP with speed (b_n^2) and a rate function $I_\Gamma : \mathbb{R}^{p^2} \rightarrow \bar{\mathbb{R}}^+$ defined as

$$I_\Gamma(x) = \begin{cases} \frac{1}{2h^3} \langle x, \Upsilon^\dagger x \rangle & \text{for } x \in \text{Im}(\Upsilon) \\ +\infty & \text{otherwise} \end{cases}$$

where Υ is explicitly given above and h comes from (H_4) .

Proof. See Section 3.2.5. □

Remark. Through vectorization, this MDP is established on \mathbb{R}^{p^2} in order to avoid any confusion in the notations, but we might work in $\mathbb{R}^{p \times p}$ as well. The associated rate function would only require a slight modification of the proof. Also, to be punctilious, we may add a small $\epsilon > 0$ to the diagonal of S_{n-1} to ensure that it is non-singular for all $n \geq 1$ without disturbing the asymptotic behavior.

When the variance Γ given in (1.11) is invertible, we establish the MDP for the OLS in the [corollary](#) that follows. However, when it is not the case, there are some technical complications and, to reach an intermediate result, we need to introduce a penalized version of the OLS. For a small $\pi \geq 0$, define

$$(2.3) \quad \hat{\theta}_n^\pi = (S_{n-1}^\pi)^{-1} \sum_{k=1}^n \Phi_{n,k-1} X_{n,k} \quad \text{where} \quad S_{n-1}^\pi = S_{n-1} + \pi n \|\|B_n^{-1}\|\| I_p$$

with possibly $\pi = 0$ if Γ is invertible, in which case it is clearly the standard OLS given above, but necessarily $\pi > 0$ otherwise. Consider also the penalized version of the variance and the corrected parameter

$$(2.4) \quad \Gamma_\pi = \Gamma + \pi I_p \quad \text{and} \quad \theta_n^\pi = (S_{n-1}^\pi)^{-1} S_{n-1} \theta_n.$$

By construction, Γ is, at worst, non-negative definite and for $\pi > 0$, Γ_π turns out to be invertible. The same goes for S_{n-1}^π .

Corollary 2.2. *Under hypotheses (H₁)–(H₅), for all $\pi > 0$, the sequence*

$$\left(\frac{\sqrt{n}}{b_n (1 - \rho(A_n))^{\frac{1}{2}}} (\widehat{\theta}_n^\pi - \theta_n^\pi) \right)_{n \geq 1}$$

satisfies an LDP with speed (b_n^2) and a rate function $I_\theta^\pi : \mathbb{R}^p \rightarrow \bar{\mathbb{R}}^+$ defined as

$$I_\theta^\pi(x) = \begin{cases} \frac{h}{2\sigma^2} \langle x, \Gamma_\pi \Gamma^\dagger \Gamma_\pi x \rangle & \text{for } x \in \text{Im}(\Gamma_\pi^{-1} \Gamma) \\ +\infty & \text{otherwise} \end{cases}$$

where the variance Γ is given in (1.11), Γ_π is the penalized variance given in (2.4) and h comes from (H₄), respectively. If in addition Γ is invertible, then the sequence

$$\left(\frac{\sqrt{n}}{b_n (1 - \rho(A_n))^{\frac{1}{2}}} (\widehat{\theta}_n - \theta_n) \right)_{n \geq 1}$$

satisfies an LDP with speed (b_n^2) and a rate function $I_\theta : \mathbb{R}^p \rightarrow \mathbb{R}^+$ defined as

$$I_\theta(x) = \frac{h}{2\sigma^2} \langle x, \Gamma x \rangle.$$

Proof. See Section 3.2.6. □

To sum up, this result shows that, when Γ is invertible, the OLS satisfies an MDP, and even when Γ is singular, one may [reach a compromise by getting](#) an MDP for a penalized estimation. In the same vein, notice also that, [in the invertible case](#),

$$\lim_{\pi \rightarrow 0^+} I_\theta^\pi(x) = I_\theta(x).$$

Remark. In the stable case where $\rho(A_n) = \rho(A) < 1$, we simply have $(1 - \rho(A_n)) \|B_n^{-1}\| = h$ and $\Gamma_n \|B_n^{-1}\|^{-1} = \Gamma$ for all $n \geq 1$. By contraction, the MDP of [Corollary 2.2](#) coincides with the one of Thm. 3 of [21] when Γ is invertible.

2.2. Some explicit examples. Before giving some examples, we can already note that (H₅) implies $\sqrt{n}(1 - \rho(A_n)) \rightarrow +\infty$. Thus, necessarily, the convergence $1 - \rho(A_n) \rightarrow 0$ cannot occur with an exponential rate, this is the reason why we focus on polynomial rates of the form $1 - \rho(A_n) = cn^{-\alpha}$ for some $c > 0$ in this section. Accordingly, in all the examples below, (H₅) is only possible when $0 < \alpha < \frac{1}{3+2\eta} < \frac{1}{3}$. Thus, one cannot expect a sequence of coefficients moving too fast toward instability. The domain of validity of the speed of the MDP will be

$$1 \ll b_n \ll n^{\frac{1-(3+2\eta)\alpha}{2}} \ll \sqrt{n}.$$

2.2.1. Univariate case with one nearly unit root. Suppose that $p = 1$. Then, (H₂) and (H₃) imply that $|\theta_n| < 1$ and $\theta_n \rightarrow \pm 1$. We also have $B_n = 1 - \theta_n^2$ and (H₄) can be expressed like

$$\lim_{n \rightarrow +\infty} \frac{B_n^{-1}}{|B_n^{-1}|} = 1 \quad \text{and} \quad \lim_{n \rightarrow +\infty} (1 - |\theta_n|) |B_n^{-1}| = \frac{1}{2}.$$

A straightforward calculation shows that

$$\Gamma_n = \frac{\sigma^2}{1 - \theta_n^2} \quad \text{and} \quad \Gamma = \sigma^2 > 0$$

so that we can choose $\pi = 0$. The standard cases, illustrated on Figure 1, are $\theta_n = 1 - c_1 n^{-\alpha}$ for the positive unit root and $\theta_n = -1 + c_2 n^{-\alpha}$ for the negative unit root, with $c_1, c_2 > 0$ and $\alpha > 0$. The rate function associated with [Corollary 2.2](#) is $I_\theta(x) = \frac{x^2}{4}$, which corresponds to Prop. 2.1 of [17]. Indeed, their rate $x \mapsto \frac{x^2}{2}$ is associated to an LDP with the renormalization $(1 - \theta_n^2)^{\frac{1}{2}}$ whereas our normalization is $(1 - |\theta_n|)^{\frac{1}{2}}$. By contraction, the asymptotic factor $\sqrt{2}$ explains the difference.

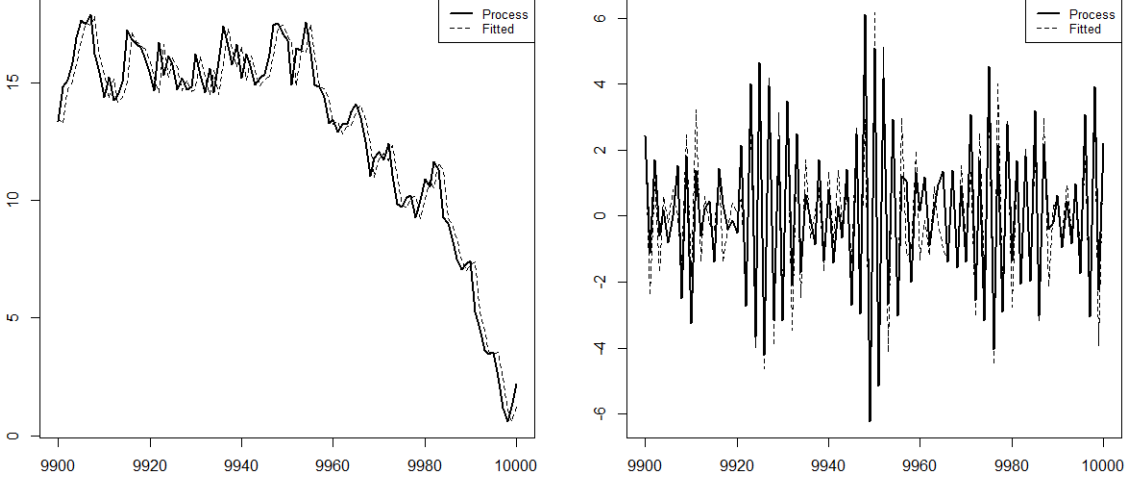


FIGURE 1. Simulation of the process (solid line) and fitted values (dotted line) for $n = 10^4$, $\pi = 0$ and $\mathcal{N}(0, 1)$ innovations. The setting is $c_1 = 0.1$ and $\alpha = 0.32$ on the left, $c_2 = 0.1$ and $\alpha = 0.32$ on the right.

2.2.2. *Bivariate case with one nearly unit root.* Suppose now that $p = 2$ and $\text{sp}(A) = \{\pm 1, \lambda\}$ with $|\lambda| < 1$. This situation occurs, for example, when

$$A_n = \begin{pmatrix} \lambda + 1 - cn^{-\alpha} & -\lambda(1 - cn^{-\alpha}) \\ 1 & 0 \end{pmatrix}$$

whose eigenvalues are $1 - cn^{-\alpha}$ and λ . This is illustrated on Figure 2. For $c > 0$ and $\alpha > 0$, (H_2) and (H_3) are satisfied. The direct calculation gives

$$B_n^{-1} = \frac{1}{2c(\lambda - 1)^2} \begin{pmatrix} 1 & -\lambda & -\lambda & \lambda^2 \\ 1 & -\lambda & -\lambda & \lambda^2 \\ 1 & -\lambda & -\lambda & \lambda^2 \\ 1 & -\lambda & -\lambda & \lambda^2 \end{pmatrix} (n^\alpha + O(1))$$

whence we obtain

$$\lim_{n \rightarrow +\infty} \frac{B_n^{-1}}{\|B_n^{-1}\|_1} = \frac{1}{4} \begin{pmatrix} 1 & -\lambda & -\lambda & \lambda^2 \\ 1 & -\lambda & -\lambda & \lambda^2 \\ 1 & -\lambda & -\lambda & \lambda^2 \\ 1 & -\lambda & -\lambda & \lambda^2 \end{pmatrix} \quad \text{and} \quad \lim_{n \rightarrow +\infty} n^{-\alpha} \|B_n^{-1}\|_1 = \frac{2}{c(\lambda - 1)^2}$$

so (H_4) is satisfied with the 1-norm. The choice $\pi = 0$ is impossible, and we finally find

$$\Gamma_\pi = \frac{\sigma^2}{4} \begin{pmatrix} 1 + \frac{4}{\sigma^2} \pi & 1 \\ 1 & 1 + \frac{4}{\sigma^2} \pi \end{pmatrix}.$$

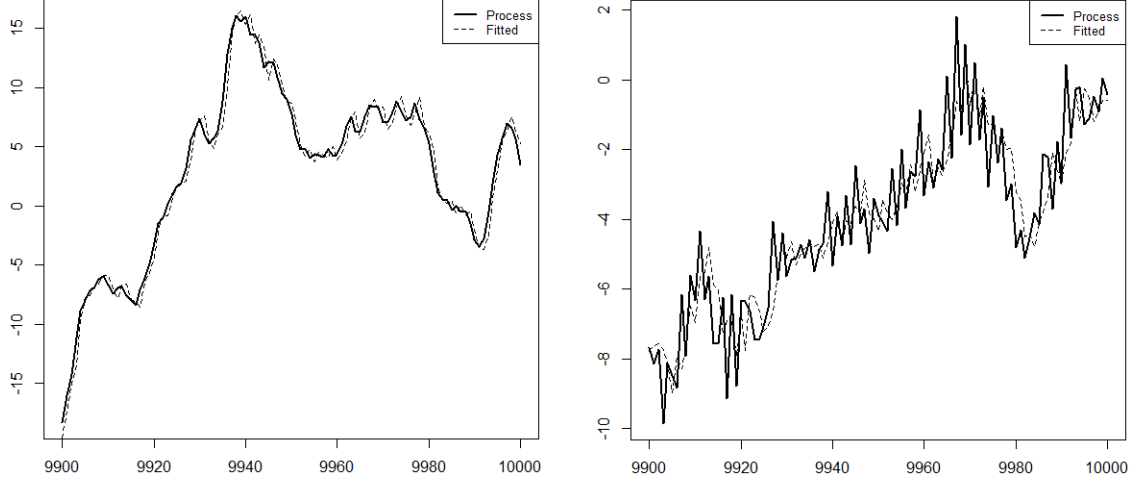


FIGURE 2. Simulation of the process (solid line) and fitted values (dotted line) for $n = 10^4$, $\pi = 10^{-5}$ and $\mathcal{N}(0, 1)$ innovations. The setting is $\lambda = 0.5$, $c = 0.1$ and $\alpha = 0.32$ on the left, $\lambda = -0.67$, $c = 0.2$ and $\alpha = 0.25$ on the right.

2.2.3. *Bivariate case with two nearly unit roots.* Following the same lines, suppose that $p = 2$ and $\text{sp}(A) = \{-1, 1\}$. This situation occurs, for example, when

$$A_n = \begin{pmatrix} (c_2 - c_1)n^{-\alpha} & (1 - c_1 n^{-\alpha})(1 - c_2 n^{-\alpha}) \\ 1 & 0 \end{pmatrix}$$

whose eigenvalues are $1 - c_1 n^{-\alpha}$ and $-1 + c_2 n^{-\alpha}$. This is illustrated on Figure 3. For $c_1, c_2 > 0$ and $\alpha > 0$, (H_2) and (H_3) are satisfied. The direct calculation gives

$$B_n^{-1} = \frac{1}{8c_1c_2} \begin{pmatrix} c_1 + c_2 & c_2 - c_1 & c_2 - c_1 & c_1 + c_2 \\ c_2 - c_1 & c_1 + c_2 & c_1 + c_2 & c_2 - c_1 \\ c_2 - c_1 & c_1 + c_2 & c_1 + c_2 & c_2 - c_1 \\ c_1 + c_2 & c_2 - c_1 & c_2 - c_1 & c_1 + c_2 \end{pmatrix} (n^\alpha + O(1))$$

whence we obtain

$$\lim_{n \rightarrow +\infty} \frac{B_n^{-1}}{\|B_n^{-1}\|_1} = \frac{1}{2(c_1 + c_2) + 2|c_2 - c_1|} \begin{pmatrix} c_1 + c_2 & c_2 - c_1 & c_2 - c_1 & c_1 + c_2 \\ c_2 - c_1 & c_1 + c_2 & c_1 + c_2 & c_2 - c_1 \\ c_2 - c_1 & c_1 + c_2 & c_1 + c_2 & c_2 - c_1 \\ c_1 + c_2 & c_2 - c_1 & c_2 - c_1 & c_1 + c_2 \end{pmatrix}.$$

Moreover,

$$\lim_{n \rightarrow +\infty} n^{-\alpha} \|B_n^{-1}\|_1 = \frac{(c_1 + c_2) + |c_2 - c_1|}{4c_1c_2}$$

so (H_4) is satisfied with the 1-norm. The choice $\pi = 0$ is possible and we finally find

$$\Gamma = \frac{\sigma^2}{2(c_1 + c_2) + 2|c_2 - c_1|} \begin{pmatrix} c_1 + c_2 & c_2 - c_1 \\ c_2 - c_1 & c_1 + c_2 \end{pmatrix}.$$

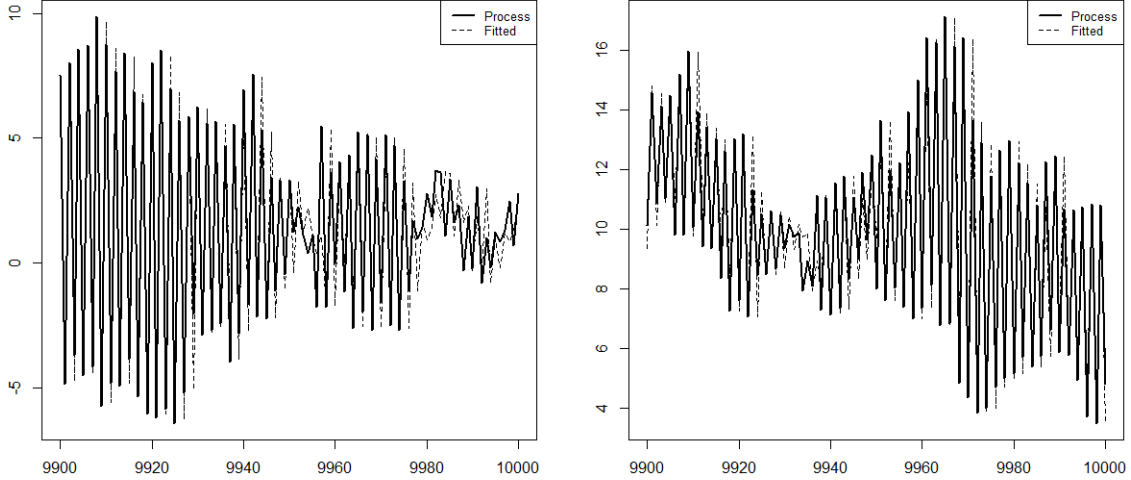


FIGURE 3. Simulation of the process (solid line) and fitted values (dotted line) for $n = 10^4$, $\pi = 0$ and $\mathcal{N}(0, 1)$ innovations. The setting is $c_1 = 0.1$, $c_2 = 0.2$ and $\alpha = 0.32$ on the left, $c_1 = 0.01$, $c_2 = 0.01$ and $\alpha = 0.25$ on the right.

2.3. Discussion on multiple eigenvalues and conclusion. As we will see in the proof of Lemma 3.1, the distinct eigenvalues assumption (H_2) is sufficient to reach our results. However, a less stringent formulation of (H_2) could be :

(H'_2) *Convergence of the companion matrix.* There exists a $p \times p$ matrix A such that

$$\lim_{n \rightarrow +\infty} A_n = A$$

and the top right element of A is non-zero. In addition, there exists a rank n_0 such that, for all $n > n_0$, A_n is diagonalizable and the change of basis matrix P_n satisfies $\|P_n\| \leq C_{st}$ and $\|P_n^{-1}\| \leq C_{st}$.

In general, multiple eigenvalues may not falsify our reasonings, except when the multiplicity concerns the eigenvalues whose modulus tends to 1. Indeed, the coefficients of $\|A_n^\ell\|$ may grow faster in that case. Consider the simple bivariate example where

$$A_n^\ell = \begin{pmatrix} a_{11,\ell} & a_{12,\ell} \\ a_{21,\ell} & a_{22,\ell} \end{pmatrix} = \begin{pmatrix} \theta_{n,1} a_{11,\ell-1} + \theta_{n,2} a_{11,\ell-2} & \theta_{n,2} a_{11,\ell-1} \\ \theta_{n,1} a_{21,\ell-1} + \theta_{n,2} a_{21,\ell-2} & \theta_{n,2} a_{21,\ell-1} \end{pmatrix}.$$

Then, it is not hard to solve this linear difference equation whose characteristic roots are the eigenvalues of A_n . In case of multiplicity, the top left term takes the form of

$$a_{11,\ell} = (c_n + d_n \ell) \rho^\ell(A_n)$$

and even if $|c_n| \leq C_{st}$ and $|d_n| \leq C_{st}$ for n large enough, it follows that

$$\sum_{\ell=0}^{+\infty} \|A_n^\ell\| \sim \frac{C_{st}}{(1 - \rho(A_n))^2}.$$

That invalidates all our reasonings and, in that case, new approaches are needed to potentially reach the moderate deviations. From our viewpoint, this is the main weakness of the [set of hypotheses](#). As it is already observed in [7], multiple unit roots located at 1 [influence](#) the rate of convergence of the OLS. We conjecture that the same phenomenon occurs here and that a larger power should come with $1 - \rho(A_n)$ in the renormalization.

To sum up, this study is a wide generalization of [17] and, although not complete in virtue of the latter remark, it covers most of the MDP issues for the estimation in the stable but nearly unstable case. Large deviations [would undoubtedly be](#) a very useful and challenging study to carry out, naturally extending this one. However, to the best of our knowledge, it is not even entirely treated in the stable time-invariant case $\rho(A_n) = \rho(A) < 1$, clearly revealing the complexity of the problem. [A complicated but stimulating trail for future studies could rely on the exponential, and not only polynomial, neighborhood of the unit root.](#) Along the same lines and even if it is of less practical interest, we might as well focus [on the explosive side of the unit roots, where new theoretical developments are necessary.](#)

3. TECHNICAL PROOFS

In all the proofs, C_{st} denotes a generic positive constant that is not necessarily identical from one line to another. We will frequently use the fact that $\|\text{vec}(\cdot)\| = \|\|\cdot\|\|_F \leq C_{st} \|\|\cdot\|\|$. For asymptotic equivalences, $f_n \asymp g_n$ means that both $f_n = O(g_n)$ and $g_n = O(f_n)$ whereas $f_n \sim g_n$ stands for $\frac{f_n}{g_n} \rightarrow 1$.

3.1. Some linear algebra tools. Thereafter, we denote by $\lambda_1, \dots, \lambda_p$ the (distinct) eigenvalues of A and $\lambda_{n,1}, \dots, \lambda_{n,p}$ those of A_n , in descending order of modulus. We start by establishing two lemmas that will prove to be very useful [in what follows](#).

Lemma 3.1. *Under hypotheses (H₂) and (H₃), as n tends to infinity,*

$$\sum_{\ell=0}^{+\infty} \|\|A_n^\ell\|\| \asymp \frac{1}{1 - \rho(A_n)} \quad \text{and} \quad \sum_{\ell=0}^{+\infty} (\ell + 1) \|\|A_n^\ell\|\| \asymp \frac{1}{(1 - \rho(A_n))^2}.$$

Proof. The lower bounds are established in Section 1.2, [in \(1.8\) and \(1.9\) precisely](#). For the upper bounds, fix

$$\delta = \frac{2}{|\lambda_p|}, \quad \epsilon_1 = \frac{1}{2} \min_{\substack{1 \leq i, j \leq p \\ i \neq j}} \left| \frac{1}{\lambda_i} - \frac{1}{\lambda_j} \right| \quad \text{and} \quad \epsilon_2 = 2 \max_{\substack{1 \leq i, j \leq p \\ i \neq j}} \left| \frac{1}{\lambda_i} - \frac{1}{\lambda_j} \right|.$$

According to Thm. 2.4.9.2 of [13], (H₂) implies the existence of a rank $n_0 = n_0(\delta, \epsilon_1, \epsilon_2)$ such that, for all $n > n_0$, the eigenvalues of A_n satisfy

$$(3.1) \quad 0 < \max_{1 \leq i \leq p} \left| \frac{1}{\lambda_{n,i}} \right| < \delta$$

and

$$(3.2) \quad \epsilon_1 < \min_{\substack{1 \leq i, j \leq p \\ i \neq j}} \left| \frac{1}{\lambda_{n,i}} - \frac{1}{\lambda_{n,j}} \right| < \max_{\substack{1 \leq i, j \leq p \\ i \neq j}} \left| \frac{1}{\lambda_{n,i}} - \frac{1}{\lambda_{n,j}} \right| < \epsilon_2.$$

Let P_n be a change of basis matrix in the diagonalization of A_n . Then, since A_n is a companion matrix, a standard choice would be

$$(3.3) \quad P_n = \begin{pmatrix} 1 & 1 & \dots & 1 \\ \frac{1}{\lambda_{n,1}} & \frac{1}{\lambda_{n,2}} & \dots & \frac{1}{\lambda_{n,p}} \\ \vdots & \vdots & & \vdots \\ \frac{1}{\lambda_{n,1}^{p-1}} & \frac{1}{\lambda_{n,2}^{p-1}} & \dots & \frac{1}{\lambda_{n,p}^{p-1}} \end{pmatrix}.$$

This Vandermonde matrix is invertible if and only if $\lambda_{n,i} \neq \lambda_{n,j}$ for all $i \neq j$ (see *e.g.* Sec. 0.9.11 of [13]). In that case, P_n^{-1} is closely related to the Lagrange interpolating polynomials given, for $i \in \{1, \dots, p\}$, by

$$L_i(X) = \frac{\prod_{j \neq i} (X - \frac{1}{\lambda_{n,j}})}{\prod_{j \neq i} (\frac{1}{\lambda_{n,i}} - \frac{1}{\lambda_{n,j}})}.$$

Precisely, the i -th row of P_n^{-1} contains the coefficients of $L_i(X)$ in the basis $\{1, X, \dots, X^{p-1}\}$ of $\mathbb{R}_{p-1}[X]$, *i.e.*

$$(3.4) \quad P_n^{-1} = \left(\frac{p_{n,i,j}}{\prod_{j \neq i} (\frac{1}{\lambda_{n,i}} - \frac{1}{\lambda_{n,j}})} \right)_{1 \leq i,j \leq p}$$

where the relation $\prod_{j \neq i} (X - \frac{1}{\lambda_{n,j}}) = p_{n,i,1} + p_{n,i,2} X + \dots + p_{n,i,p} X^{p-1}$ enables to identify each $p_{n,i,j}$. Combining (3.1) and (3.2), it follows that, for all $n > n_0$,

$$\| \| P_n \| \|_1 \leq p(1 + \delta + \dots + \delta^{p-1}) \leq C_{st}.$$

We also have $\| \| P_n^{-1} \| \|_1 \leq C_{st}$ since $\epsilon_1^{p-1} < \prod_{j \neq i} |\frac{1}{\lambda_{n,i}} - \frac{1}{\lambda_{n,j}}| < \epsilon_2^{p-1}$ and since $p_{n,i,j}$ is a finite combination of sums and products of $\frac{1}{\lambda_{n,1}}, \dots, \frac{1}{\lambda_{n,p}}$. To sum up, for all $\ell \geq 0$ and $n > n_0$,

$$A_n^\ell = P_n D_n^\ell P_n^{-1} \quad \text{where} \quad D_n = \text{diag}(\lambda_{n,1}, \dots, \lambda_{n,p}).$$

Consequently,

$$(3.5) \quad \begin{aligned} \| \| A_n^\ell \| \| &= \| \| A_n^\ell \| \| \mathbb{1}_{\{n \leq n_0\}} + \| \| P_n D_n^\ell P_n^{-1} \| \| \mathbb{1}_{\{n > n_0\}} \\ &\leq \| \| A_n^\ell \| \| \mathbb{1}_{\{n \leq n_0\}} + \| \| P_n \| \| \| \| P_n^{-1} \| \| \| \| D_n^\ell \| \| \mathbb{1}_{\{n > n_0\}} \\ &\leq \| \| A_n^\ell \| \| \mathbb{1}_{\{n \leq n_0\}} + C_{st} \rho^\ell(A_n) \mathbb{1}_{\{n > n_0\}}. \end{aligned}$$

It only remains to sum over ℓ and to let n tend to infinity to reach the first result. Similarly,

$$(\ell + 1) \| \| A_n^\ell \| \| \leq (\ell + 1) \| \| A_n^\ell \| \| \mathbb{1}_{\{n \leq n_0\}} + C_{st} (\ell + 1) \rho^\ell(A_n) \mathbb{1}_{\{n > n_0\}}$$

so we get the second result by following the same lines. \square

Lemma 3.2. *Under hypotheses (H_2) and (H_3) , we have the convergence*

$$\lim_{n \rightarrow +\infty} A_n^{w_n} = 0$$

for any rate (w_n) satisfying $w_n(1 - \rho(A_n)) \rightarrow +\infty$.

Proof. Consider the rank n_0 introduced in the proof of Lemma 3.1. Then, according to the inequality (3.5),

$$(3.6) \quad \| \| A_n^{w_n} \| \| \leq \| \| A_n^{w_n} \| \| \mathbb{1}_{\{n \leq n_0\}} + C_{st} \rho^{w_n}(A_n) \mathbb{1}_{\{n > n_0\}}$$

where the invertible and uniformly bounded matrices P_n and P_n^{-1} are given in (3.3) and (3.4), respectively. We also have

$$(3.7) \quad \lim_{n \rightarrow +\infty} \rho^{w_n}(A_n) = \lim_{n \rightarrow +\infty} e^{-w_n(1 - \rho(A_n))} = 0$$

from the hypothesis on (w_n) . It remains to let n tend to infinity in the above inequality. \square

3.2. Proofs of the main results. First of all, it is convenient to express the empirical variance of the process as

$$\begin{aligned}
\frac{1}{n} \sum_{k=1}^n (\Phi_{n,k} \Phi_{n,k}^T - \Gamma_n) &= \frac{1}{n} \sum_{k=1}^n A_n \Phi_{n,k-1} \Phi_{n,k-1}^T A_n^T + \frac{1}{n} \sum_{k=1}^n A_n \Phi_{n,k-1} E_k^T \\
&\quad + \frac{1}{n} \sum_{k=1}^n E_k \Phi_{n,k-1}^T A_n^T + \frac{1}{n} \sum_{k=1}^n E_k E_k^T - \Gamma_n \\
&= \frac{1}{n} \sum_{k=1}^n \Delta_{n,k} + \frac{1}{n} \sum_{k=1}^n A_n (\Phi_{n,k} \Phi_{n,k}^T - \Gamma_n) A_n^T - \frac{T_n}{n}
\end{aligned}$$

where the variance Γ_n is given in (1.4),

$$\begin{aligned}
(3.8) \quad \Delta_n &= \frac{1}{n} \sum_{k=1}^n \Delta_{n,k} \\
&= \frac{1}{n} \sum_{k=1}^n (A_n \Phi_{n,k-1} E_k^T + E_k \Phi_{n,k-1}^T A_n^T + E_k E_k^T + A_n \Gamma_n A_n^T - \Gamma_n)
\end{aligned}$$

and the residual term is

$$T_n = A_n (\Phi_{n,n} \Phi_{n,n}^T - \Phi_{n,0} \Phi_{n,0}^T) A_n^T.$$

Then, solving this generalized Sylvester equation (Lem. 2.1 of [14]) and considering the invertibility of B_n in (1.7) which is proved at the beginning of Section 1.2, we reach the decomposition

$$(3.9) \quad \text{vec} \left(\frac{1}{n} \sum_{k=1}^n (\Phi_{n,k} \Phi_{n,k}^T - \Gamma_n) \right) = B_n^{-1} \text{vec}(\Delta_n) - \frac{B_n^{-1} \text{vec}(T_n)}{n}.$$

Let us now reason step by step, *via* some intermediate results. The proofs of Lemmas 3.3, 3.4, 3.5, 3.6 and 3.7 are postponed to the supplementary file.

3.2.1. Exponential moments of the squared initial value. We recall that, from the causal form (1.3) of the process,

$$\Phi_{n,0} = \sum_{\ell=0}^{+\infty} A_n^\ell E_{-\ell}.$$

The following result gives an exponential moment for this correctly renormalized squared initial value.

Lemma 3.3. *Under hypothesis (H_1) ,*

$$\mathbb{E} \left[\exp \left(\frac{\alpha}{L_n^2} \|\Phi_{n,0} \Phi_{n,0}^T\| \right) \right] < +\infty$$

where L_n is given in (1.8).

3.2.2. *Exponential convergence of the residual term.* The residual term in the decomposition (3.9) is given by

$$(3.10) \quad R_n = \frac{B_n^{-1} \text{vec}(A_n (\Phi_{n,n} \Phi_{n,n}^T - \Phi_{n,0} \Phi_{n,0}^T) A_n^T)}{n}.$$

Our next objective is to prove the exponential negligibility of this residual.

Lemma 3.4. *Under hypotheses (H₁)–(H₅), for all $r > 0$,*

$$\lim_{n \rightarrow +\infty} \frac{1}{b_n^2} \ln \mathbb{P} \left(\frac{\sqrt{n} (1 - \rho(A_n))^{\frac{3}{2}}}{b_n} \|R_n\| \geq r \right) = -\infty.$$

3.2.3. *The truncated sequence.* In what follows, we define the rate

$$(3.11) \quad m_n = \left\lfloor \left(\frac{1}{1 - \rho(A_n)} \right)^{\frac{3+3\eta}{3+2\eta}} \right\rfloor$$

and we note from (H₃)–(H₅) that

$$(3.12) \quad \lim_{n \rightarrow +\infty} m_n (1 - \rho(A_n)) = +\infty \quad \text{and} \quad \lim_{n \rightarrow +\infty} \frac{b_n \|\|B_n^{-1}\|\|^{\frac{1}{2}} m_n^{1+\frac{2\eta}{3}}}{\sqrt{n}} = 0.$$

Following the idea of [17], we are going to use (m_n) as a truncation sequence. Consider

$$(3.13) \quad \Psi_{n,k} = \sum_{\ell=0}^{m_n-2} A_n^\ell E_{k-\ell}$$

as an approximation of $\Phi_{n,k}$ in its causal form (1.3). We also define the truncated version of the summands $\Delta_{n,k}$ in (3.8) as

$$(3.14) \quad \zeta_{n,k} = A_n \Psi_{n,k-1} E_k^T + E_k \Psi_{n,k-1}^T A_n^T + E_k E_k^T + A_n \Gamma_n A_n^T - \Gamma_n.$$

The process $(B_n^{-1} \text{vec}(\zeta_{n,k}))_k$ is strictly stationary and m_n –dependent, according to Def. 6.4.3 of [4]. Let us study some properties of this process.

Lemma 3.5. *Under hypotheses (H₁)–(H₄), we can find a constant $c_\alpha > 0$ such that, for a sufficiently large n ,*

$$\mathbb{E} \left[\exp \left(c_\alpha \|\|B_n^{-1}\|\|^{-1} \sum_{\ell=0}^{w_n} \|\|A_n^\ell E_{-\ell} E_1^T\|\| \right) \right] \leq \mathbb{E}[e^{\alpha \varepsilon_1^2}]$$

for any rate (w_n) satisfying $w_n (1 - \rho(A_n)) \rightarrow +\infty$.

Lemma 3.6. *Under hypotheses (H₂)–(H₄), for all $n \geq 1$ and $k \in \{1, \dots, n\}$,*

$$\mathbb{E}[\text{vec}(\zeta_{n,k})] = 0 \quad \text{and} \quad \text{Cov}(\text{vec}(\zeta_{n,k}), \text{vec}(\zeta_{n,j})) = \begin{cases} 0 & \text{for } k \neq j \\ \Upsilon_n & \text{for } k = j \end{cases}$$

where the $p^2 \times p^2$ covariance Υ_n can be explicitly built in terms of σ^2 , A_n and B_n . In addition,

$$\lim_{n \rightarrow +\infty} \frac{B_n^{-1} \Upsilon_n (B_n^{-1})^T}{\|\|B_n^{-1}\|\|^3} = \Upsilon$$

where the non-zero limiting matrix Υ is given in (2.2).

Remark. As a by-product, we also obtain, [going on with the reasoning of the proof](#),

$$\lim_{n \rightarrow +\infty} \frac{\Gamma_n}{\|B_n^{-1}\|} = \Gamma$$

where Γ_n is given in (1.4), which [proves](#) (1.11).

3.2.4. *The remainder of the truncation.* We denote by

$$(3.15) \quad \Lambda_n = \frac{1}{n} \sum_{k=1}^n (A_n (\Phi_{n,k-1} - \Psi_{n,k-1}) E_k^T + E_k (\Phi_{n,k-1} - \Psi_{n,k-1})^T A_n^T)$$

the remainder of the truncation of Δ_n in (3.8) made *via* (3.14). Our last preliminary objective is to establish the following lemma.

Lemma 3.7. *Under hypotheses (H_1) – (H_5) , for all $r > 0$,*

$$\lim_{n \rightarrow +\infty} \frac{1}{b_n^2} \ln \mathbb{P} \left(\frac{\sqrt{n} (1 - \rho(A_n))^{\frac{3}{2}}}{b_n} \|B_n^{-1} \text{vec}(\Lambda_n)\| \geq r \right) = -\infty.$$

We are now ready to prove [Theorem 2.1](#) and [Corollary 2.2](#).

3.2.5. *Proof of Theorem 2.1.* Consider the sequence

$$(3.16) \quad \xi_{n,k} = \frac{B_n^{-1} \text{vec}(\zeta_{n,k})}{\|B_n^{-1}\|^{\frac{3}{2}}}$$

where $\zeta_{n,k}$ is given in (3.14). The process $(\xi_{n,k})_k$ is also strictly stationary and m_n –dependent. Like in [18] or [17, suppl. mat.], let us extract an independent sequence from this process. For $j \in \{1, \dots, j_n\}$, define

$$\xi'_{n,j} = \xi_{n,(j-1)m_n+1} + \dots + \xi_{n,jm_n}$$

where $j_n = \lfloor \frac{n}{m_n} \rfloor$ and where (m_n) and its properties are given in (3.11). Then, $(\xi'_{n,j})_j$ is strictly stationary and 1–dependent. Next, for $t \in \{1, \dots, t_n\}$, define

$$\xi''_{n,t} = \xi'_{n,(t-1)u_n+1} + \dots + \xi'_{n,tu_n-1}$$

where $t_n = \lfloor \frac{j_n}{u_n} \rfloor$ and (u_n) is another rate satisfying

$$(3.17) \quad \lim_{n \rightarrow +\infty} u_n = +\infty \quad \text{and} \quad \lim_{n \rightarrow +\infty} \frac{b_n \|B_n^{-1}\|^{\frac{1}{2}} (m_n u_n)^{1+\frac{2\eta}{3}}}{\sqrt{n}} = 0.$$

To be convinced that such a rate exists, one can [use](#) (3.12) and the fact that $|\ln f_n| \rightarrow +\infty$ and $f_n |\ln f_n|^a \rightarrow 0$ when $f_n \rightarrow 0$. The process $(\xi''_{n,t})_t$ is now i.i.d. and the rates [satisfy](#)

$$(3.18) \quad \lim_{n \rightarrow +\infty} \frac{t_n u_n m_n}{n} = 1.$$

The reasoning of [17, suppl. mat.] does not suit us, so we need to reformulate the establishment of the MDP. First, by a Taylor-Lagrange expansion,

$$(3.19) \quad \exp \left(\left\langle \lambda, \frac{b_n}{\sqrt{n}} \xi''_{n,1} \right\rangle \right) = 1 + \frac{b_n}{\sqrt{n}} \langle \lambda, \xi''_{n,1} \rangle + \frac{b_n^2}{2n} \langle \lambda, \xi''_{n,1} \rangle^2 + \frac{b_n^3}{6n^{\frac{3}{2}}} \langle \lambda, \xi''_{n,1} \rangle^3 e^{\nu_n}$$

in which the **remainder term** satisfies, for any $\alpha > 0$,

$$\begin{aligned} e^{\alpha \nu_n} &< \exp\left(\frac{\alpha b_n}{\sqrt{n}} |\langle \lambda, \xi_{n,1}'' \rangle|\right) \leq \exp\left(\frac{\alpha b_n}{\sqrt{n}} \|\lambda\| \sum_{\ell=1}^{m_n u_n} \|\xi_{n,\ell}\|\right) \\ &\leq \exp\left(C_{st} \frac{b_n}{\sqrt{n}} \|\|B_n^{-1}\|\|^{-\frac{1}{2}} \sum_{\ell=1}^{m_n u_n} \|\|\zeta_{n,\ell}\|\|\right). \end{aligned}$$

Now, the random variables $\|\|\zeta_{n,\ell}\|\|$ sharing the same distribution for all $\ell \geq 0$, it follows from Hölder's inequality that,

$$\begin{aligned} \mathbb{E}[e^{\alpha \nu_n}] &< \mathbb{E}\left[\exp\left(C_{st} \frac{b_n m_n u_n}{\sqrt{n}} \|\|B_n^{-1}\|\|^{-\frac{1}{2}} \|\|\zeta_{n,1}\|\|\right)\right] \\ (3.20) \quad &= \mathbb{E}\left[\exp\left(C_{st} \frac{b_n \|\|B_n^{-1}\|\|^{\frac{1}{2}} m_n u_n}{\sqrt{n}} \|\|B_n^{-1}\|\|^{-1} \|\|\zeta_{n,1}\|\|\right)\right] < +\infty \end{aligned}$$

for n large enough, using Lemma 3.5 with $m_n(1 - \rho(A_n)) \rightarrow +\infty$ stemming from (3.12), the convergence of $\|\|A_n\|\|$, (H₁) and treating all the terms of (3.14) similarly. Taking the expectation in (3.19) and exploiting the independence of the zero-mean process $(\xi_{n,t}'')_t$, we obtain the decomposition

$$\begin{aligned} \frac{1}{b_n^2} \ln \mathbb{E}\left[\exp\left(\left\langle \lambda, \frac{b_n}{\sqrt{n}} \sum_{\ell=1}^n \xi_{n,\ell} \right\rangle\right)\right] &\sim \frac{t_n}{b_n^2} \ln \mathbb{E}\left[\exp\left(\left\langle \lambda, \frac{b_n}{\sqrt{n}} \xi_{n,1}'' \right\rangle\right)\right] \\ (3.21) \quad &= \frac{t_n}{2n} \mathbb{E}[\langle \lambda, \xi_{n,1}'' \rangle^2] + O\left(\frac{t_n b_n}{6 n^{\frac{3}{2}}} |\mathbb{E}[\langle \lambda, \xi_{n,1}'' \rangle^3 e^{\nu_n}]|\right) \end{aligned}$$

for we can see, as it is done in [18], that the residual term

$$\tau_n = \sum_{\ell=1}^n \xi_{n,\ell} - \sum_{\ell=1}^{t_n} \xi_{n,\ell}''$$

plays a negligible role in comparison to the main one. To eliminate the third-order term, we first look at the fourth-order moment of $\langle \lambda, \xi_{n,1}'' \rangle$, that is

$$\mathbb{E}[\langle \lambda, \xi_{n,1}'' \rangle^4] \leq \frac{C_{st} \|\lambda\|^4}{\|\|B_n^{-1}\|\|^2} \mathbb{E}\left[\left\|\sum_{\ell=1}^{m_n u_n} \zeta_{n,\ell}\right\|^4\right].$$

A long but standard calculation shows that

$$\begin{aligned} \mathbb{E}\left[\left\|\sum_{\ell=1}^{m_n u_n} A_n \Psi_{n,\ell-1} E_\ell^T\right\|^4\right] &\leq C_{st} \mathbb{E}\left[\left\|\sum_{\ell=1}^{m_n u_n} \Psi_{n,\ell-1} \varepsilon_\ell\right\|^4\right] \\ &= O((m_n u_n \|\|B_n^{-1}\|\|)^2) \end{aligned}$$

as n tends to infinity. This result is reached using the strict stationarity of the process, the explicit expression of $X_{n,0}^4$ in terms of A_n^ℓ , the inequality (3.6) and, finally, using (H₄) giving the equivalence between $(1 - \rho(A_n))^{-2}$ and $C_{st} \|\|B_n^{-1}\|\|^2$. So,

$$\mathbb{E}[\langle \lambda, \xi_{n,1}'' \rangle^4] = O(m_n^2 u_n^2).$$

By Lyapunov's inequality,

$$\mathbb{E}[|\langle \lambda, \xi''_{n,1} \rangle|^{3+\delta}] \leq (\mathbb{E}[\langle \lambda, \xi''_{n,1} \rangle^4])^{\frac{3+\delta}{4}} = O\left((m_n u_n)^{\frac{3+\delta}{2}}\right)$$

for a small $\delta > 0$. Now, combining this result with (3.20) and Hölder's inequality, for sufficiently large values of n ,

$$\begin{aligned} \frac{t_n b_n}{n^{\frac{3}{2}}} \mathbb{E}[|\langle \lambda, \xi''_{n,1} \rangle^3 e^{\nu_n}|] &\leq \frac{t_n b_n}{n^{\frac{3}{2}}} (\mathbb{E}[|\langle \lambda, \xi''_{n,1} \rangle|^{3+\delta}])^{\frac{3}{3+\delta}} (\mathbb{E}[e^{\frac{3+\delta}{\delta} \nu_n}])^{\frac{\delta}{3+\delta}} \\ (3.22) \qquad \qquad \qquad &\leq C_{st} \frac{t_n b_n}{n^{\frac{3}{2}}} (m_n u_n)^{\frac{3}{2}} \longrightarrow 0 \end{aligned}$$

by (3.20), (3.18) and the properties in (3.17). The second-order term in (3.21) satisfies

$$\begin{aligned} \frac{t_n}{2n} \mathbb{E}[\langle \lambda, \xi''_{n,1} \rangle^2] &= \frac{t_n}{2n} \lambda^T \mathbb{V}(\xi''_{n,1}) \lambda = \frac{t_n u_n m_n}{2n \|\|B_n^{-1}\|\|^3} \lambda^T B_n^{-1} \mathbb{V}(\text{vec}(\zeta_{n,1})) (B_n^{-1})^T \lambda \\ &= \frac{t_n u_n m_n}{2n} \lambda^T \frac{B_n^{-1} \Upsilon_n (B_n^{-1})^T}{\|\|B_n^{-1}\|\|^3} \lambda \\ (3.23) \qquad \qquad \qquad &\longrightarrow \frac{1}{2} \langle \lambda, \Upsilon \lambda \rangle \end{aligned}$$

where we used (3.18) and the results of Lemma 3.6. The combination of (3.21), (3.22) and (3.23) together with the Gärtner-Ellis theorem (see *e.g.* Sec. 2.3 of [8]) shows that the sequence

$$\left(\frac{1}{b_n \sqrt{n}} \sum_{\ell=1}^n \xi_{n,\ell} \right)_{n \geq 1}$$

satisfies an LDP with speed (b_n^2) and rate function given by the Fenchel-Legendre transform of the above logarithmic moment generating function, *i.e.*

$$I(x) = \sup_{\lambda \in \mathbb{R}^{p^2}} \left\{ \langle \lambda, x \rangle - \frac{1}{2} \langle \lambda, \Upsilon \lambda \rangle \right\}.$$

Note that, due to its particular structure, Υ is only non-negative definite as soon as $p > 1$ (by way of example, its last row and column are zero). In that case (see *e.g.* Ex. 1.1.4 of [12], page 212), the explicit expression of this quadratic rate function, strictly convex on its relative interior, is

$$I(x) = \begin{cases} \frac{1}{2} \langle x, \Upsilon^\dagger x \rangle & \text{for } x \in \text{Im}(\Upsilon) \\ +\infty & \text{otherwise.} \end{cases}$$

After the truncation introduced in (3.13), the decomposition (3.9) can be rewritten as

$$\begin{aligned} \frac{\sqrt{n} (1 - \rho(A_n))^{\frac{3}{2}}}{b_n} \text{vec} \left(\frac{1}{n} \sum_{k=1}^n (\Phi_{n,k} \Phi_{n,k}^T - \Gamma_n) \right) &= \frac{(1 - \rho(A_n))^{\frac{3}{2}} \|\|B_n^{-1}\|\|^{\frac{3}{2}}}{b_n \sqrt{n}} \sum_{k=1}^n \xi_{n,k} \\ &\quad + \frac{\sqrt{n} (1 - \rho(A_n))^{\frac{3}{2}}}{b_n} R_n^* \end{aligned}$$

where, in the remainder term $R_n^* = B_n^{-1} \text{vec}(\Lambda_n) - R_n$, the residual of the truncation is given in (3.15) and the main residual R_n is given in (3.10). Lemma 3.4 and Lemma 3.7 show that the first term in the right-hand side is an exponentially good approximation of the left-hand side and that, as a consequence, they share the same LDP (see Def. 4.2.10 and Thm. 4.2.13

of [8]). The contraction principle (see Thm. 4.2.1 of [8]) enables to compute the rate function associated with the LDP, namely

$$(3.24) \quad I_{\Gamma}(x) = I(h^{-\frac{3}{2}}x) = \begin{cases} \frac{1}{2h^3} \langle x, \Upsilon^{\dagger} x \rangle & \text{for } x \in \text{Im}(\Upsilon) \\ +\infty & \text{otherwise} \end{cases}$$

where the limiting value $h > 0$ comes from (\mathbf{H}_4) . \square

3.2.6. *Proof of Corollary 2.2.* Using (2.3) and (2.4),

$$\begin{aligned} \frac{\sqrt{n}}{b_n (1 - \rho(A_n))^{\frac{1}{2}}} (\widehat{\theta}_n^{\pi} - \theta_n^{\pi}) &= \frac{\sqrt{n} (S_{n-1}^{\pi})^{-1}}{b_n (1 - \rho(A_n))^{\frac{1}{2}}} \sum_{k=1}^n \Phi_{n,k-1} \varepsilon_k \\ &= \frac{n \|\|B_n^{-1}\|\| (S_{n-1}^{\pi})^{-1}}{b_n \sqrt{n} \|\|B_n^{-1}\|\|^{\frac{1}{2}} (1 - \rho(A_n))^{\frac{1}{2}} \|\|B_n^{-1}\|\|^{\frac{1}{2}}} \sum_{k=1}^n \Phi_{n,k-1} \varepsilon_k. \end{aligned}$$

Our objective is first to prove that, for all $r > 0$,

$$(3.25) \quad \lim_{n \rightarrow +\infty} \frac{1}{b_n^2} \ln \mathbb{P} \left(\|\|n \|\|B_n^{-1}\|\| (S_{n-1}^{\pi})^{-1} - \Gamma_{\pi}^{-1}\|\| \geq r \right) = -\infty$$

where Γ_{π} is the invertible penalized variance (2.4), and then to establish an LDP for the sequence

$$(3.26) \quad \left(\frac{1}{b_n \sqrt{n} \|\|B_n^{-1}\|\|^{\frac{1}{2}}} \sum_{k=1}^n \Phi_{n,k-1} \varepsilon_k \right)_{n \geq 1}$$

in order to obtain the announced result, *via* the contraction principle (Thm. 4.2.1 of [8]). On the one hand, we know from Theorem 2.1 and (3.24) that

$$\begin{aligned} \frac{1}{b_n^2} \ln \mathbb{P} \left(\left\| \left\| \frac{S_{n-1}}{n \|\|B_n^{-1}\|\|} - \frac{\Gamma_n}{\|\|B_n^{-1}\|\|} \right\| \right\| \geq r \right) &= \frac{1}{b_n^2} \ln \mathbb{P} \left(\frac{\sqrt{n}}{b_n \|\|B_n^{-1}\|\|^{\frac{3}{2}}} \left\| \left\| \frac{S_{n-1}}{n} - \Gamma_n \right\| \right\| \geq r_n \right) \\ &\rightarrow -\infty = - \lim_{\|x\| \rightarrow +\infty} I_{\Gamma}(x) \end{aligned}$$

since, by (\mathbf{H}_4) and (\mathbf{H}_5) ,

$$r_n = \frac{r \sqrt{n}}{b_n \|\|B_n^{-1}\|\|^{\frac{1}{2}}} \rightarrow +\infty$$

and $(1 - \rho(A_n))^{\frac{3}{2}} \sim h^{\frac{3}{2}} \|\|B_n^{-1}\|\|^{-\frac{3}{2}}$. So,

$$\lim_{n \rightarrow +\infty} \frac{1}{b_n^2} \ln \mathbb{P} \left(\left\| \left\| \frac{S_{n-1}^{\pi}}{n \|\|B_n^{-1}\|\|} - \Gamma_n^{\pi} \right\| \right\| \geq r \right) = -\infty \quad \text{for} \quad \Gamma_n^{\pi} = \frac{\Gamma_n}{\|\|B_n^{-1}\|\|} + \pi I_p.$$

It is also clear that

$$\left\{ \left\| \left\| \frac{S_{n-1}^{\pi}}{n \|\|B_n^{-1}\|\|} - \Gamma_n^{\pi} \right\| \right\| \geq r \right\} \subset \left\{ \left\| \left\| \frac{S_{n-1}^{\pi}}{n \|\|B_n^{-1}\|\|} - \Gamma_n^{\pi} \right\| \right\| \geq \frac{r}{2} \right\} \cup \left\{ \|\|\Gamma_n^{\pi} - \Gamma_{\pi}\|\| \geq \frac{r}{2} \right\}$$

and (1.11) shows that the second event in the right-hand side becomes impossible when n increases. Hence, from the reasoning above,

$$\lim_{n \rightarrow +\infty} \frac{1}{b_n^2} \ln \mathbb{P} \left(\left\| \left\| \frac{S_{n-1}^{\pi}}{n \|\|B_n^{-1}\|\|} - \Gamma_{\pi} \right\| \right\| \geq r \right) = -\infty.$$

Now we shall use Lem. 2 of [21] to get (3.25).

On the other hand, all the work consisting in proving that the sequence (3.26) satisfies an LDP with speed (b_n^2) has already been done in the proof of Theorem 2.1. Indeed, *via* the truncation (3.13),

$$\begin{aligned} \frac{1}{b_n \sqrt{n} \|\|B_n^{-1}\|\|^{1/2}} \sum_{k=1}^n \Psi_{n,k-1} \varepsilon_k &= \frac{1}{b_n \sqrt{n} \|\|B_n^{-1}\|\|^{1/2}} \sum_{k=1}^n \sum_{\ell=0}^{m_n-2} A_n^\ell E_{k-\ell-1} \varepsilon_k \\ &= \frac{1}{b_n \sqrt{n}} \sum_{k=1}^n Z_{n,k} \end{aligned}$$

where the process $(Z_{n,k})_k$ forms a strictly stationary and m_n -dependent sequence. However, apart from the renormalization, this is precisely the first column of the first term of (3.14). Thus, the calculations are perfectly similar and they lead to an LDP with rate function

$$J(x) = \sup_{\lambda \in \mathbb{R}^p} \left\{ \langle \lambda, x \rangle - \frac{\sigma^2}{2} \langle \lambda, \Gamma \lambda \rangle \right\} = \begin{cases} \frac{1}{2\sigma^2} \langle x, \Gamma^\dagger x \rangle & \text{for } x \in \text{Im}(\Gamma) \\ +\infty & \text{otherwise.} \end{cases}$$

The contraction principle finally gives

$$(3.27) \quad I_\theta(x) = J(\Gamma_\pi \sqrt{h} x) = \begin{cases} \frac{h}{2\sigma^2} \langle x, \Gamma_\pi \Gamma^\dagger \Gamma_\pi x \rangle & \text{for } x \in \text{Im}(\Gamma_\pi^{-1} \Gamma) \\ +\infty & \text{otherwise} \end{cases}$$

where the exponential convergence (3.25) has been combined to the LDP established on the sequence (3.26). \square

Acknowledgements. The author thanks the associate editor and the two anonymous reviewers for the numerous comments and suggestions that clearly helped to improve the paper. He also thanks R. Garbit for the constructive discussion about the link between Vandermonde matrices and Lagrange polynomials.

REFERENCES

- [1] BERCU, B. On large deviations in the Gaussian autoregressive process: stable, unstable and explosive cases. *Bernoulli*. 7 (2001), 299–316.
- [2] BERCU, B., GAMBOA, F., AND ROUAULT, A. Large deviations for quadratic forms of stationary Gaussian processes. *Stoch. Proc. Appl.* 71 (1997), 75–90.
- [3] BITSEKI PENDA, V., DJELLOUT, H., AND PROÏA, F. Moderate deviations for the Durbin-Watson statistic related to the first-order autoregressive process. *ESAIM Probab. Stat.* 18 (2014), 308–331.
- [4] BROCKWELL, P. J., AND DAVIS, R. A. *Time series: Theory and Methods (Second Edition)*. Springer Series in Statistics. Springer, New York, 1991.
- [5] BUCHMANN, B., AND CHAN, N. H. Unified asymptotic theory for nearly unstable AR(p) processes. *Stoch. Proc. Appl.* 123 (2013), 952–985.
- [6] CHAN, N. H., AND WEI, C. Z. Asymptotic inference for nearly nonstationary AR(1) processes. *Ann. Stat.* 15 (1987), 1050–1063.
- [7] CHAN, N. H., AND WEI, C. Z. Limiting distributions of least squares estimates of unstable autoregressive processes. *Ann. Statist.* 16 (1988), 367–401.
- [8] DEMBO, A., AND ZEITOUNI, O. *Large Deviations Techniques and Applications (Second Edition)*, vol. 38 of *Applications of Mathematics*. Springer, 1998.
- [9] DJELLOUT, H., GUILLIN, A., AND WU, L. Moderate deviations of empirical periodogram and non-linear functionals of moving average processes. *Ann. I. H. Poincaré*. 42 (2006), 393–416.
- [10] DONSKER, M. D., AND VARADHAN, S. R. S. Large deviations for stationary Gaussian processes. *Comm. Math. Phys.* 97 (1985), 187–210.
- [11] DUFLO, M. *Random iterative models*. Applications of Mathematics (vol. 34), New York. Springer-Verlag, Berlin, 1997.

- [12] HIRIART-URRUTY, J. B., AND LEMARÉCHAL, C. *Fundamentals of Convex Analysis*. Grundlehren Text Editions. Springer, 2012.
- [13] HORN, R. A., AND JOHNSON, C. R. *Matrix Analysis (Second Edition)*. Cambridge University Press, Cambridge, New-York, 2012.
- [14] JIANG, T., AND WEI, M. On solutions of the matrix equations $X - AXB = C$ and $X - A\bar{X}B = C$. *Linear Algebra Appl.* 367 (2003), 225–233.
- [15] MAS, A., AND MENNETEAU, L. Large and moderate deviations for infinite-dimensional autoregressive processes. *J. Multivariate Anal.* 87 (2003), 241–260.
- [16] MIAO, Y., AND SHEN, S. Moderate deviation principle for autoregressive processes. *J. Multivariate Anal.* 100 (2009), 1952–1961.
- [17] MIAO, Y., WANG, Y., AND YANG, G. Moderate deviation principles for empirical covariance in the neighbourhood of the unit root. *Scand. J. Stat.* 42 (2015), 234–255.
- [18] MIAO, Y., AND YANG, G. A moderate deviation principle for m -dependent random variables with unbounded m . *Acta Appl. Math.* 104 (2008), 191–199.
- [19] PHILLIPS, P. C. B., AND LEE, J. H. Limit theory for vars with mixed roots near unity. *Economet. Rev.* 34 (2015), 1034–1055.
- [20] PHILLIPS, P. C. B., AND MAGDALINOS, T. Limit theory for moderate deviations from a unit root. *J. Econometrics.* 136 (2007), 115–130.
- [21] WORMS, J. Moderate deviations for stable Markov chains and regression models. *Electron. J. Probab.* 4 (1999), 1–28.
- [22] WU, W. B., AND ZHAO, Z. Moderate deviations for stationary processes. *Stat. Sinica.* 18 (2008), 769–782.

LABORATOIRE ANGEVIN DE RECHERCHE EN MATHÉMATIQUES, LAREMA, UMR 6093, CNRS, UNIV ANGERS, SFR MATHSTIC, 2 BD LAVOISIER, 49045 ANGERS CEDEX 01, FRANCE.

Email address: frederic.proia@univ-angers.fr