TESTING FOR THE EXTENT OF INSTABILITY IN NEARLY-UNSTABLE PROCESSES

MARIE BADREAU AND FRÉDÉRIC PROÏA

ABSTRACT. This paper deals with unit root issues in time series analysis. It has been known for a long time that unit root tests may be flawed when a series although stationary has a root close to unity. That motivated recent papers dedicated to autoregressive processes where the bridge between stability and instability is expressed by means of time-varying coefficients. The process we consider has a companion matrix A_n with spectral radius $\rho(A_n) < 1$ satisfying $\rho(A_n) \to 1$, a situation described as 'nearly-unstable'. The question we investigate is: given an observed path supposed to come from a nearly-unstable process, is it possible to test for the 'extent of instability', *i.e.* to test how close we are to the unit root? In this regard, we develop a strategy to evaluate α and to test for $\mathcal{H}_0$: "$\alpha = \alpha_0$" against $\mathcal{H}_1$: "$\alpha > \alpha_0$" when $\rho(A_n)$ lies in an inner $O(n^{-\alpha})$-neighborhood of the unity, for some $0 < \alpha < 1$. Empirical evidence is given about the advantages of the flexibility induced by such a procedure compared to the common unit root tests. We also build a symmetric procedure for the usually left out situation where the dominant root lies around -1.

Keywords: Nearly unstable autoregressive process, Unit root test, Martingales.
Mathematics Subject Classification: 62M10 - 62F12 - 62F03 - 60G52 - 60G42.

1. INTRODUCTION AND MOTIVATION

This paper is devoted to the well-known unit root issue, still at the heart of many concerns in finance, econometrics, and more generally in time series analysis, at the origin, when undetected, of spurious regressions through unreliable estimates. In order to investigate the stochastic non-stationarity of an observed autoregressive path, two strategies are frequently encountered. The first one consists in estimating the coefficient behind the (supposed) unit root, and the second one directly looks for a hidden random walk in the disturbance. In a very particular case, Dickey and Fuller (1979) first tackled the unit root issue by considering an AR(1) model with a linear trend and concluded that, when there is a unit root in the generating process, the least-squares estimator of the autoregressive parameter has an asymmetric distribution that can be written asymptotically as a functional of a family of detrended Wiener processes, actually depending on the trend. Such a result had been anticipated in the Gaussian case by White (1958) in the absence of trend and, among numerous other studies, Chan and Wei (1988) improved the assumptions on the noise by considering a sequence of martingale differences (with sufficient moments). The behavior of the associated t-statistic has been tabulated by Dickey and Fuller (1979) for testing purposes, see also MacKinnon (1991). The *ADF test* (Augmented Dickey-Fuller) is probably the most commonly used nowadays to evaluate the presence of a unit root in a causal ARMA(p,q) process. It was formalized by Said and Dickey (1981) when p and q are known, and by Said and Dickey (1984) under a truncated approximation of the AR(∞) expression, therefore working for any p and q. At the basis of their study is the following differentiated form of

an AR(k) process with parameter θ,

$$(1.1) \qquad \forall\, n \geqslant 1, \quad \Delta X_n = (\theta_0 - 1) X_{n-1} + \sum_{i=1}^{k-1} \delta_i\, \Delta X_{n-i} + \varepsilon_n$$

where (ε_n) is a white noise sequence with finite variance, $\delta_i = -(\theta_{i+1} + \cdots + \theta_k)$ for $1 \leqslant i \leqslant k-1$ and $\theta_0 = \theta_1 + \cdots + \theta_k$, so that (ΔX_n) is stationary under the null hypothesis of unit root $\mathcal{H}_0$: "$\theta_0 = 1$" (actually trend-stationary if a trend is added in the model). That was an inspiration for us, and we will find a similar reasoning in this paper. The asymptotic behavior of the estimate is still the one of the seminal work of Dickey and Fuller (1979) as long as $k = O(n^{1/3})$ but the power of the test is impacted and, as it is explained by Schwert (1989) or by Ng and Perron (1995), some distortions may occur for badly truncated processes. Let us also mention the non-parametric approach suggested by Phillips (1987b) and deepened by Phillips and Perron (1988) in which all correlation phenomenon are postponed to the noise of the model, which is now supposed to be strongly mixing. Overall there is an abundant literature on unit roots tests, we have just summarized here some important topics. The interested reader will find in-depth studies for example in Bhargava (1986), Perron (1988), Schmidt and Phillips (1992), etc. From a practical point of view, Nelson and Plosser (1982) had already highlighted the presence of unit roots in a number of macroeconomic series *via* the Dickey-Fuller strategy. Although not directly related to this paper, we found it useful to mention the complementary approach consisting in testing for trend-stationarity instead of integration. The noise is now assumed to include a random walk whose variance is zero under the null hypothesis and among such procedures, the most famous seem to be the *KPSS test* of Kwiatkowski et al. (1992) and the quite similar *LMC test* of Leybourne and McCabe (1994). Here again there is an ample literature but too far from our scope to dwell on it, let us instead return to unit roots.

In the context of autoregressive processes, the substantial works of Lai and Wei (1983) and Chan and Wei (1988) helped to establish that the least-squares estimator is strongly consistent wherever the characteristic roots lie, but with very different convergence rates and limit distributions whether the process is stationary (stable), integrated (unstable), or even explosive. However, De Jong et al. (1992) observed that the unit root tests may be empirically less powerful than their stable counterparts when, although very close to it, the spectral radius remains less than 1, and some econometricians are not fully convinced by the compatibility between the theoretical impacts of this discontinuity – from $n^{1/2}$-normality to n–asymmetry while the spectral radius continuously moves towards the unit root – and real-word data. These are strong arguments in favor of attempts to bridge stability and instability that motivated numerous studies devoted to intermediate models, including ours. Most of them involve a first-order process with time-varying coefficients $(X_{n,k})$, generated by the triangular form

$$(1.2) \qquad \forall\, n \geqslant 1,\ \forall\, 1 \leqslant k \leqslant n, \quad X_{n,k} = \theta_n\, X_{n,k-1} + \varepsilon_k.$$

This is an easy-to-use relevant case to understand the underlying dynamics since to focus on the inner neighborhood of the unit root, we may only consider a sequence of coefficients satisfying $|\theta_n| < 1$ for all $n \geqslant 1$ but $|\theta_n| \to 1$, so that the model is always stable but increasingly close to instability. Chan and Wei (1987), Phillips (1987a), Giraitis and Phillips (2006) and later Phillips and Magdalinos (2007) were interested in such a model and, among other conclusions, it was established that the least-squares estimator is asymptotically normal

around θ_n with rate $(n v_n)^{1/2}$ when $\theta_n = 1 - c/v_n$, for any $1 \ll v_n \ll n$ and $c > 0$. The junction is not perfect – there is still a discontinuity in the asymptotic variance when $v_n = 1$ and in the asymptotic distribution when $v_n = n$ – but these are promising results. Indeed, by setting $v_n = n^\alpha$ for some $0 < \alpha < 1$, the gap between stability and instability is filled with rates of the form $n^{1/2} \ll n^{(1+\alpha)/2} \ll n$. The studies of Van der Meer et al. (1999), Park (2003), Buchmann and Chan (2013), or the more recent ones of Proïa (2020) and Badreau and Proïa (2023) aimed at finding a kind of generalization to nearly-unstable $\mathrm{AR}(p)$ processes. In the latter, the process $(\Phi_{n,k})$ is assumed to follow the $\mathrm{VAR}_p(1)$ triangular form

$$(1.3) \qquad \forall\, n \geqslant 1,\ \forall\, 1 \leqslant k \leqslant n, \quad \Phi_{n,k} = A_n\, \Phi_{n,k-1} + E_k$$

where (E_k) is a p-vectorial noise and

$$(1.4) \qquad A_n = \begin{pmatrix} \theta_{n,1} & \theta_{n,2} & \cdots & \theta_{n,p} \\ \ddots & & & 0 \\ & I_{p-1} & & \vdots \\ & & \ddots & 0 \end{pmatrix}$$

is the associated $p \times p$ companion matrix with spectral radius satisfying $\rho(A_n) < 1$ for all $n \geqslant 1$ and $\rho(A_n) \to 1$. Provided suitable assumptions of moments, the least-squares estimator is still shown to be consistent and asymptotically normal around θ_n with mixed rates (see Thm. 3 and 4), and it is deduced as a corollary that, when $\rho(A_n) = 1 - c/v_n$,

$$(1.5) \qquad \sqrt{n\, v_n}\, \big\langle L_n, \widehat{\theta}_n - \theta_n \big\rangle \ \xrightarrow{\ d\ }\ \mathcal{N}\big(0, h_0^2\big) \quad \text{with} \quad L_n^T = \big(\lambda_1^{i-1} \rho(A_n)^{1-i}\big)_{1 \leqslant i \leqslant p}$$

where $\lambda_1 = \pm 1$ (depending on the sign of the unit root in the unstable limit process) and the asymptotic variance $h_0^2 > 0$ is given as a function of c and the eigenvalues of A, the limit matrix of A_n. However, although this result and the ones that may be found in the references above are very useful for understanding the behavior of the process and the estimator in the inner neighborhood of the unit circle, they definitely lack practical application. Beyond the scope of this study, the concept of near-instability has also been defined for first-order INAR processes (AR with integer values), see *e.g.* Ispány et al. (2003). In that case, the mean value of the counting sequence is assumed to converge to one, that is to the unstable setting of an INAR(1) process, and some kind of bridges are built between the stable and unstable rates. An application to nearly-critical Galton-Watson processes can be found in the same reference or more deeply in Ispány et al. (2005). Let us also mention the study of Barreto-Souza and Chan (2024) related to nearly-unstable INARCH(1) processes, still dealing with counting time series but in a heteroscedastic context. The reader is referred to the large bibliography on the topic given in this very recent article.

The question we wish to investigate in this paper is the following: given an observed path supposed to come from a nearly-unstable process (*e.g.* doubtful ADF/KPSS diagnoses), is it possible to test for the 'extent of instability', *i.e.* to admit that the process is stable but to test how close we are to the unit root? Setting $\rho(A_n) = 1 - c/v_n$ with $v_n = n^\alpha$, the question is clearly related to the choice of α and may reduce to a test of $\mathcal{H}_0$: "$\alpha = \alpha_0$" against $\mathcal{H}_1$: "$\alpha > \alpha_0$" for some test value α_0. In addition, evaluating α is likely to provide adjustments in the confidence intervals associated with the autoregressive parameters. This is an innovative approach of the unit root issue (to the best of our knowledge), especially from a practical point of view. In Section 2, we give some specific features of the model and the hypotheses before developing the test procedure just mentioned. We will see that, as a

corollary of our reasonings, we obtain a version of the asymptotic normality for a corrected statistic closely related to the least-squares estimation that does not need a mixing of the rates matrix, actually deepening Badreau and Proïa (2023, Thm. 4). An empirical study is provided in Section 3, first on simulated data and then on real time series – the famous dataset coming from Nelson and Plosser (1982) – in order to assess the performances of our procedure and justify its utility. After a short conclusion in Section 4, we finally prove our results in Section 5.

2. A (Quasi-)Unit Root Test

Let us consider the triangular model (1.3), where

$$\forall\, n \geqslant 1,\ \forall\, 1 \leqslant k \leqslant n, \quad \Phi_{n,k}^T = \begin{pmatrix} X_{n,k} & \cdots & X_{n,k-p+1} \end{pmatrix} \quad \text{and} \quad E_k^T = \begin{pmatrix} \varepsilon_k & 0 & \cdots & 0 \end{pmatrix}.$$

The process (ε_k) is assumed to be a white noise with finite variance $\sigma^2 > 0$, and the arbitrary initial value $\Phi_{n,0}$ is square-integrable and independent of the noise. The companion matrix A_n of the process is given in (1.4) and has spectrum $\{\lambda_{n,i}\}_{1 \leqslant i \leqslant p}$ where $\lambda_{n,1}$ is the one with largest modulus (the complex eigenvalues are sorted according to their modulus, in descending order, with ties broken by lexicographic order, also in descending order). The following technical hypotheses are intended to reflect our understanding of 'near instability'.

(A1) *Convergence of the companion matrix.* There exists a $p \times p$ matrix A such that

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

with distinct eigenvalues $0 < |\lambda_p| \leqslant \ldots \leqslant |\lambda_2| \leqslant |\lambda_1| = \rho(A) = 1$, and the top-right element of A is non-zero ($\theta_p \neq 0$).

(A2) *Number of unit roots.* There is exactly one unit root in A ($\lambda_1 = \pm 1$ but $|\lambda_2| < 1$).

(A3) *Spectral radius of the companion matrix.* The spectral radius of A_n is given by

$$\forall\, n \geqslant 1, \quad \rho(A_n) = 1 - \frac{c}{n^\alpha}$$

for some $c > 0$ and $0 < \alpha < 1$.

For any $d \geqslant 1$, the usual $d \times d$ matrices

$$(2.1) \qquad I_d = \begin{pmatrix} 1 & 0 & \cdots \\ 0 & \ddots & \\ \vdots & & 1 \end{pmatrix} \quad \text{and} \quad K_d = \begin{pmatrix} 1 & 0 & \cdots \\ 0 & 0 & \\ \vdots & & \ddots \end{pmatrix}$$

will be frequently encountered and the first vector of the canonical basis of $\mathbb{R}^d$ will be denoted by $e_d^T = (1, 0, \ldots, 0)$. For readability purposes, we prove the results as if the limit process had a positive unit root (that is, $\lambda_1 = 1$), so that $\lambda_{n,1}$ is equal (to be rigorous, this is true only for n sufficiently large) to the spectral radius $\rho_n = \rho(A_n)$ of the companion matrix. Nevertheless they remain valid for $\lambda_1 = -1$, at the cost of some adjustments that will be clarified alongside (see Section 5.6). Technically, we should assume in addition that $0 < c < n^\alpha$ to ensure that for all n, $0 < \rho_n < 1$. Since our results are asymptotic w.r.t. n, we will allow ourselves a slight abuse of notation and simply set $c > 0$. The parameters c and α have a very different role to play: c is a *variance parameter* and is clearly less important than α, which is a *rate parameter*. In fact, the over-parametrization issued from the ratio c/n^α makes it impossible to dissociate c from α without additional constraints, which is the reason why we first assume in Section 2.1 that c is known to focus on α. However, a procedure will also be detailed in

Section 2.2 for a joint evaluation of c and α, according to a certain empirical protocol. First, like Park (2003), it is convenient to express the process under the hierarchical form

$$(2.2) \qquad \forall\, n \geqslant 1,\ \forall\, 1 \leqslant k \leqslant n, \qquad \begin{cases} X_{n,k} &= \lambda_{n,1}\, X_{n,k-1} + V_{n,k} \\ V_{n,k} &= \beta_{n,1}\, V_{n,k-1} + \ldots + \beta_{n,p-1}\, V_{n,k-p+1} + \varepsilon_k \end{cases}$$

where we merely set $V_{n,k} = \varepsilon_k$ if $p = 1$. As it can be easily seen by factoring the autoregressive polynomial of $(X_{n,k})$ by $(1 - \lambda_{n,1} z)$, the companion matrix $\bar{A}_n$ associated with $(V_{n,k})$ has spectrum $\{\lambda_{n,i}\}_{2 \leqslant i \leqslant p}$. Under (A1) and (A2), it is therefore a stable $\mathrm{AR}(p-1)$ process with time-varying coefficients $\beta_{n,i} \to \beta_i$, such that $\bar{A}_n \to \bar{A}$ and $\rho(\bar{A}_n) \to \rho(\bar{A}) = |\lambda_2| < 1$. So the limit process is itself stable and its (normalized) covariance matrix is given by

$$(2.3) \qquad \Sigma_{p-1} = \sum_{k \geqslant 0} \bar{A}^k\, K_{p-1}\, (\bar{A}^T)^k.$$

But that also leads to another representation of the process, closely related to (1.1) mentioned in the introduction,

$$(2.4) \qquad \forall\, n \geqslant 1,\ \forall\, 1 \leqslant k \leqslant n, \qquad X_{n,k} = \lambda_{n,1}\, X_{n,k-1} + \sum_{i=1}^{p-1} \beta_{n,i}\, \left(X_{n,k-i} - \lambda_{n,1}\, X_{n,k-i-1} \right) + \varepsilon_k$$

from which we will benefit in our following results. In consequence, it is straightforward to see that $\theta_n^T = (\theta_{n,1}, \ldots, \theta_{n,p})$ and $\beta_n^T = (\beta_{n,1}, \ldots, \beta_{n,p-1})$ are such that

$$(2.5) \qquad \forall n \geqslant 1, \qquad \theta_n = J_n\, \beta_n + \lambda_{n,1}\, e_p \qquad \text{where} \qquad J_n = \begin{pmatrix} 1 & & \\ -\lambda_{n,1} & \ddots & \\ & \ddots & 1 \\ & & -\lambda_{n,1} \end{pmatrix} \qquad (= 0 \text{ if } p = 1)$$

is a rectangular matrix of size $p \times (p-1)$. To emphasize that the latent variables $V_{n,k}$ depend on α, from now on we will use $V_{n,k}(\alpha)$ instead and we will do the same for all confusing notation. Accordingly, let

$$\forall\, n \geqslant 1,\ \forall\, 1 \leqslant k \leqslant n, \qquad \Psi_{n,k}^T(\alpha) = \begin{pmatrix} X_{n,k} & V_{n,k}(\alpha) & \cdots & V_{n,k-p+2}(\alpha) \end{pmatrix}$$

where the arbitrary initial value $\Psi_{n,0}(\alpha)$ is square-integrable, so that from (2.4),

$$(2.6) \qquad \forall\, n \geqslant 1,\ \forall\, 1 \leqslant k \leqslant n, \qquad X_{n,k} = \vartheta_n^T(\alpha)\, \Psi_{n,k-1}(\alpha) + \varepsilon_k$$

where $\vartheta_n^T(\alpha) = (\lambda_{n,1}, \beta_{n,1}, \ldots, \beta_{n,p-1})$. Let also

$$(2.7) \qquad S_{n,n} = \sum_{k=0}^{n} \Phi_{n,k}\, \Phi_{n,k}^T \qquad \text{and} \qquad T_{n,n}(\alpha) = \sum_{k=0}^{n} \Psi_{n,k}(\alpha)\, \Psi_{n,k}^T(\alpha)$$

together with the rate matrix of size $p \times p$, that shall help to control the convergence of $S_{n,n}$ and $T_{n,n}(\alpha)$, given by

$$(2.8) \qquad C_n(\alpha) = \mathrm{diag}\left(\frac{1}{1 - \rho_n(\alpha)}, 1, \ldots, 1 \right).$$

The sum of squares $\sum_k \|E_k\|^2$ in (1.3) is obviously minimized by

$$(2.9) \qquad \widehat{\theta}_n = S_{n,n-1}^{-1} \sum_{k=1}^{n} \Phi_{n,k-1}\, X_{n,k}$$

whereas in (2.6), it is minimized (as a function of α), by

$$(2.10) \qquad \widehat{\vartheta}_n(\alpha) = T_{n,n-1}^{-1}(\alpha) \sum_{k=1}^{n} \Psi_{n,k-1}(\alpha) X_{n,k}.$$

If $p = 1$, then $\widehat{\vartheta}_n(\alpha) = \widehat{\vartheta}_n$ coincides with $\widehat{\theta}_n$. But before coming to the study of the rate parameter α, it should be noted that as an interesting consequence of our study, it is now possible to build a consistent and asymptotically normal estimator of θ_n. On the basis of (2.5), let

$$(2.11) \qquad \widetilde{\theta}_n = \widehat{J}_n \, \widehat{\beta}_n + \widehat{v}_n \, e_p$$

where $\widehat{J}_n$ is J_n with $\lambda_{n,1}$ that can be estimated by $\widehat{v}_n = e_p^T \widehat{\vartheta}_n$, the first component of $\widehat{\vartheta}_n$, and $\widehat{\beta}_n$ is $\widehat{\vartheta}_n$ deprived of its first component (we drop the dependency on α in the notation of this paragraph, for readability). Then, we have the following result.

Proposition 2.1. *Assume that (A1), (A2) and (A3) hold, and that $\mathbb{E}[|\varepsilon_1|^{2+\nu}] = \eta_\nu < +\infty$ for some $\nu > 0$. Then,*

$$(2.12) \qquad \left\| \widetilde{\theta}_n - \theta_n \right\| \xrightarrow{p} 0 \quad and \quad \sqrt{n}\left(\widetilde{\theta}_n - \theta_n \right) \xrightarrow{d} \mathcal{N}\left(0, \, J\Sigma_{p-1}^{-1} J^T \right)$$

where Σ_{p-1} is given in (2.3) and

$$J = \begin{pmatrix} 1 & & & \\ -\lambda_1 & \ddots & & \\ & \ddots & 1 & \\ & & -\lambda_1 \end{pmatrix} \qquad (= 0 \ if \ p = 1).$$

Proof. See Section 5.3. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ $\square$

The asymptotic normality is degenerate in the sense that $J\Sigma_{p-1}^{-1} J^T$ is not invertible (as a $p \times p$ matrix of rank at most $p-1$). This may be illustrated by the relation $v_p^T J = 0$ where $v_p^T = (\lambda_1^{i-1})_{1 \leqslant i \leqslant p}$ so that (2.12) becomes

$$\sqrt{n}\, v_p^T \left(\widetilde{\theta}_n - \theta_n \right) \xrightarrow{p} 0$$

and a parallel can be made with (1.5). It is not surprising for $p = 1$ since it is well-known that in that case, $|\widetilde{\theta}_n - \theta_n| = O_p(n^{-(1+\alpha)/2})$, see *e.g.* Phillips and Magdalinos (2007).

Remark 2.1. Strictly speaking, $\widetilde{\theta}_n$ – as well as $\widehat{\theta}_n$ – is a consistent estimation of θ, the limit of θ_n, for obviously $\|\widetilde{\theta}_n - \theta\| \leqslant \|\widetilde{\theta}_n - \theta_n\| + \|\theta_n - \theta\|$. But it can only be asymptotically normal under the stronger hypothesis $\|\theta_n - \theta\| = o_p(n^{-1/2})$, which requires in particular that $\alpha > 1/2$. In this case, an immediate consequence of (2.12) is

$$(2.13) \qquad \sqrt{n}\left(\widetilde{\theta}_n - \theta \right) \xrightarrow{d} \mathcal{N}\left(0, \, J\Sigma_{p-1}^{-1} J^T \right).$$

2.1. **The variance parameter c is known.** After this contextualization, we are now ready to define a statistic that will be useful to achieve our objectives (recall that $\lambda_1 = \pm 1$ means that the limit process has a unit root located at ± 1). Suppose first that the variance parameter c is known and consider, provided that $\lambda_1 \widehat{v}_n(\alpha) < 1$, the statistic

$$(2.14) \qquad \forall n > 1, \quad \widehat{\alpha}_n(\alpha) = \frac{\ln c - \ln\left(1 - \lambda_1 \widehat{v}_n(\alpha) \right)}{\ln n}$$

where $\widehat{v}_n(\alpha) = e_p^T \widehat{\vartheta}_n(\alpha)$ is the first component of $\widehat{\vartheta}_n(\alpha)$. Despite appearances, $\widehat{\alpha}_n(\alpha)$ is *not* an estimator of α because it actually depends on α. It is a statistic whose behavior, as a function of α, will help to formalize the test.

Theorem 2.2. *Assume that (A1), (A2) and (A3) hold, and that $\mathbb{E}[|\varepsilon_1|^{2+\nu}] = \eta_\nu < +\infty$ for some $\nu > 0$. Then, for the 'true' values $0 < \alpha < 1$ and $c > 0$,*

$$(2.15) \qquad (\ln n)\sqrt{n^{1-\alpha}}\left(\widehat{\alpha}_n(\alpha) - \alpha\right) \xrightarrow{d} \mathcal{N}\left(0,\, 2\,c^{-1}\pi_{11}^{-2}\right)$$

where

$$(2.16) \qquad \pi_{11} = \frac{\lambda_1^{p-1}}{\prod_{j=2}^p (1 - \lambda_j)} \quad (= 1 \ if \ p = 1).$$

Proof. See Section 5.4. $\qquad\square$

Now choose some $\alpha/2 < \alpha_0 < 1$ (in particular, any $1/2 \leqslant \alpha_0 < 1$ is suitable), and consider the test statistic

$$(2.17) \qquad Z_n^2(\alpha_0) = \frac{c\,\widehat{\pi}_n^2}{2}\,(\ln n)^2\, n^{1-\alpha_0}\left(\widehat{\alpha}_n(\alpha_0) - \alpha_0\right)^2$$

where

$$\widehat{\pi}_n = \frac{\lambda_1^{p-1}}{\prod_{i=2}^p (1 - \widehat{\lambda}_{n,i})} \quad (= 1 \ if \ p = 1)$$

and $\widehat{\lambda}_{n,i}$ is the i-th eigenvalue of the estimated companion matrix $\widehat{A}_n$ built by plugging $\widehat{\theta}_n$ into A_n, ordered as indicated in the hypotheses. Whenever $\widehat{\alpha}_n(\alpha_0)$ is not defined for the reason mentioned above, we set $Z_n^2(\alpha_0) = +\infty$ by convention (so that $\mathcal{H}_0$ below is rejected). The following result is devoted to the asymptotic behavior of $Z_n^2(\alpha_0)$ under the null $\mathcal{H}_0$: "$\alpha = \alpha_0$" as well as the alternative $\mathcal{H}_1$: "$\alpha > \alpha_0$".

Theorem 2.3. *Assume that (A1), (A2) and (A3) hold, and that $\mathbb{E}[|\varepsilon_1|^{2+\nu}] = \eta_\nu < +\infty$ for some $\nu > 0$. Then, under $\mathcal{H}_0$: "$\alpha = \alpha_0$", the test statistic (2.17) satisfies*

$$(2.18) \qquad Z_n^2(\alpha_0) \xrightarrow{d} \chi_1^2$$

where χ_1^2 has a chi-square distribution with one degree of freedom. If we additionally assume that $\alpha < 2\alpha_0$, then under $\mathcal{H}_1$: "$\alpha > \alpha_0$",

$$(2.19) \qquad \inf_{\alpha/2 < \alpha_0 < \alpha} Z_n^2(\alpha_0) \xrightarrow{p} +\infty.$$

Proof. See Section 5.5. $\qquad\square$

From now on, we replace the assumption $\alpha < 2\alpha_0$ by the stronger one $\alpha_0 \geqslant 1/2$ (this will be justified in the empirical section). As a corollary, for some $\alpha_0 \geqslant 1/2$, the rejection of $\mathcal{H}_0$: "$\alpha = \alpha_0$" in favor of $\mathcal{H}_1$: "$\alpha > \alpha_0$" will happen as soon as $Z_n^2(\alpha_0) > z_{1-\epsilon}$ where $z_{1-\epsilon}$ is the quantile of order $1-\epsilon$ of the χ_1^2 distribution. All the empirical study of Section 3 is based on this rule (but note that instead of (2.17), we could as well use the first component of $\widehat{\vartheta}_n(\alpha_0)$ in (2.10) and reach the same conclusions, through the limiting distribution of Lemma 5.4). To evaluate the extent of instability, a natural choice may be, given c,

$$(2.20) \qquad \widehat{\alpha}_c = \inf\{\alpha_0 \in \mathcal{A} : \mathcal{H}_0 \text{ is not rejected}\}$$

where $\mathcal{A} \subset [1/2, 1[$ is a grid of test values (by convention $\inf \varnothing = +\infty$ so that our conclusion is that the series is integrated if $\mathcal{H}_0$ is always rejected and $\alpha \geqslant 1/2$). It remains to generalize the procedure to the less common situation where the limit process has a negative unit root, that is $\lambda_1 = -1$. Actually, Theorems 2.2 and 2.3 are still valid: λ_1 already appears in the first component of $\vartheta_n(\alpha)$ in (2.6), *i.e.* $\lambda_{n,1} = \lambda_1 \rho_n$, and in $\widehat{\alpha}_n(\alpha)$ given in (2.14). The reader is referred to Section 5.6 for more technical details.

2.2. **The variance parameter c is estimated.** As mentioned in the beginning of the section, the over-parametrization makes it impossible to dissociate c from α without additional constraints. We suggest instead a cross-validation with the following protocol:

- For each $c_0 \in \mathcal{C}$, compute $\widehat{\alpha}_{c_0}$ based on the first $(n - n_{val})$ values of the series.
- Choose $(\widehat{c}_0, \widehat{\alpha}_{\widehat{c}_0})$ as the pair that optimizes some criterion stemming from the one-step predictions of the n_{val} remaining values.

Clearly, the value of α selected by (2.20) has to be biased downwards (due to overlapping confidence intervals for increasing values of α_0). Thus, we must also expect $\widehat{c}_0$ to be biased downwards because of the ratio c/n^α. Paradoxically, despite unsatisfactory individual evaluations of c and α, the compensated result may prove interesting, as the empirical section suggests (see Figure 6). In terms of computational costs, proceeding like this is a great faster than testing all $(c, \alpha) \in \mathcal{C} \times \mathcal{A}$ since the cross-validation criterion is only computed once for each c. Overall, the procedure contains $O(1 + |\mathcal{C}||\mathcal{A}| + |\mathcal{C}|n_{val})$ steps, where in this context a 'step' requires at worst the inversion of a $p \times p$ matrix ($p \ll n$). Hence, this is not so large considering that n_{val} is usually a small percentage of n. Especially as from a practical point of view, it does not seem necessary for $\mathcal{C}$ to have a very fine mesh (unlike $\mathcal{A}$).

3. Empirical Study

This section is dedicated to the empirical performances of the test procedure described in Theorem 2.3 together with the selection rule (2.20), and also to the cross-validation process presented in Section 2.2.

3.1. **On simulated data (c is known).** Without loss of generality, let us consider $c = 1$ to produce the data, and set $\rho_n = 1 - n^{-\alpha} = \pm\lambda_{n,1}$ with $1/2 \leqslant \alpha < 1$. For each simulation, $\lambda_{n,2}, \dots, \lambda_{n,p}$ are randomly chosen in the interval $[-\rho_n + 0.1, \rho_n - 0.1]$, and a process of size n is generated with a standard Gaussian noise. The test of $\mathcal{H}_0 : ``\alpha = \alpha_0"$ against $\mathcal{H}_1 : ``\alpha > \alpha_0"$ is done *via* Theorem 2.3 for $\alpha_0 \in \mathcal{A} = \{1/2 + k/50, k = 0, \dots, 24\}$. The rejection frequencies are evaluated on the basis of 5000 replications for each α_0. The distribution of $Z_n^2(\alpha_0)$ and the empirical power of the test with 5% type I risk are represented on Figure 1 ($\alpha = 2/3$), Figure 2 ($\alpha = 3/4$) and Figure 3 ($\alpha = 4/5$), for $\lambda_1 = 1$, $n = 1000$ and $p \in \{1, 2, 3, 4\}$. To get an idea of the impact of smaller samples and a negative unit root, similar experiments are conducted on Figure 4 (with $\lambda_1 = 1$, $\alpha = 4/5$, $p = 1$, $n = 250$ on the top, and $\lambda_1 = -1$, $\alpha = 4/5$, $p = 1$, $n = 1000$ on the bottom). The light-gray items (at the right of the vertical line) correspond to $\alpha_0 > \alpha$, situations beyond the scope of our study that we decided anyway to incorporate in the simulations for informational purposes. Finally, the distribution of $\widehat{\alpha}_{c=1}$ over 5000 replications is outlined for $\alpha \in \{2/3, 3/4, 4/5\}$ with $n = 250$ and $n = 1000$, on Figure 5. For the latter simulations, we have adopted the rule (2.20) with 5% type I risk.

From the simulations of Figures 1–3, we observe that p has a negligible influence on the asymptotic behaviors when $\lambda_1 = 1$. Under $\mathcal{H}_0$, the rejection rate is $\approx 5\%$ as expected and
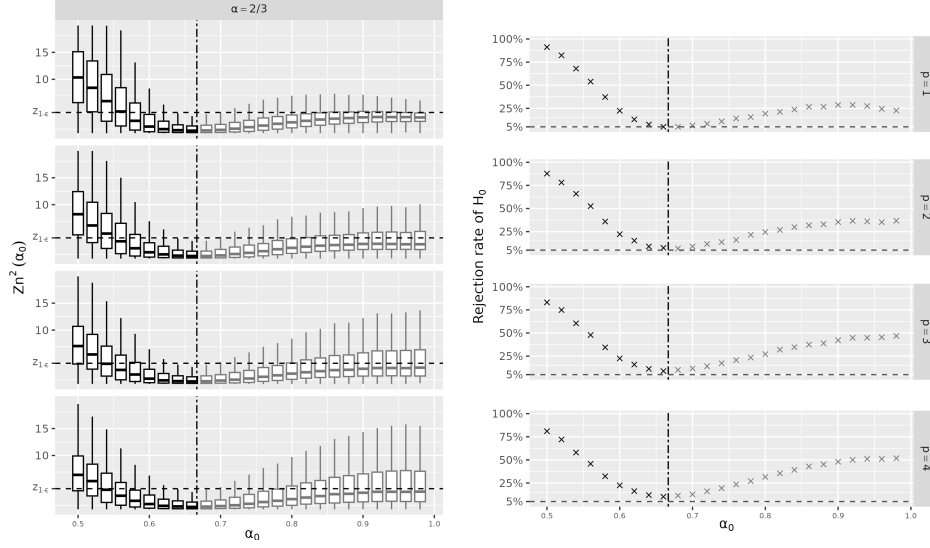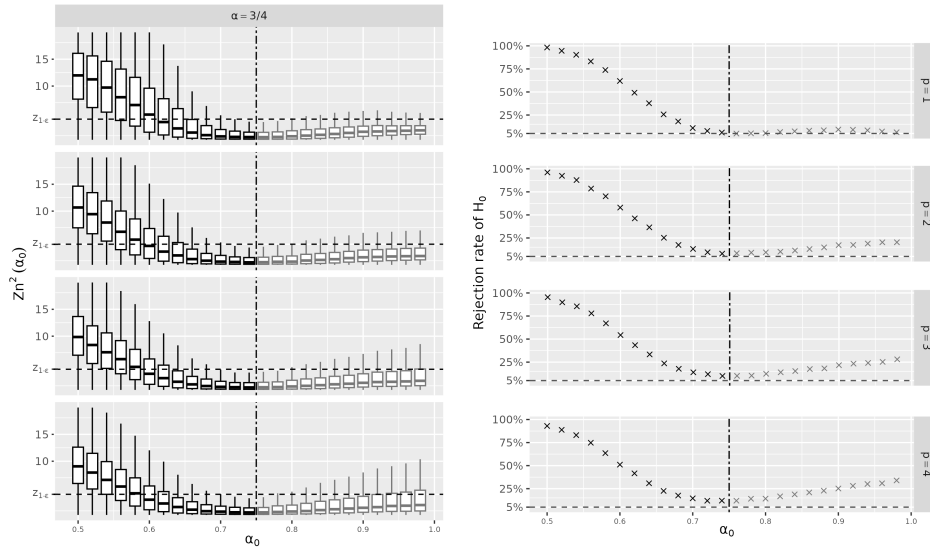
FIGURE 1. Empirical distribution of $Z_n^2(\alpha_0)$ for $n = 1000$, $\lambda_1 = 1$, $\alpha = 2/3$, $\alpha_0 \in \mathcal{A}$ and $p \in \{1, 2, 3, 4\}$ on the left. The corresponding rejection frequencies of $\mathcal{H}_0$: "$\alpha = \alpha_0$" against $\mathcal{H}_1$: "$\alpha > \alpha_0$" (5% level) are represented on the right.

FIGURE 2. Empirical distribution of $Z_n^2(\alpha_0)$ for $n = 1000$, $\lambda_1 = 1$, $\alpha = 3/4$, $\alpha_0 \in \mathcal{A}$ and $p \in \{1, 2, 3, 4\}$ on the left. The corresponding rejection frequencies of $\mathcal{H}_0$: "$\alpha = \alpha_0$" against $\mathcal{H}_1$: "$\alpha > \alpha_0$" (5% level) are represented on the right.

the power of the test is quite satisfying under $\mathcal{H}_1$ ($\geqslant 75\%$ for $\alpha_0 \leqslant \alpha - 0.2$, $\approx 100\%$ for $\alpha_0 \leqslant \alpha - 0.3$). The decrease in the rejection rates is rather similar for α_0 growing from $1/2$ to α, whether $\alpha = 2/3$, $\alpha = 3/4$ or $\alpha = 4/5$. However, we can see that the trickier alternatives where $\alpha_0 \leqslant \alpha - 0.1$ are underestimated ($\approx 25\%$): this is due to the fine mesh of $\mathcal{A}$, it is difficult to discern substantial differences in the statistic between values of α_0 that increases 0.02 by 0.02 and overall, the numerical results are encouraging.
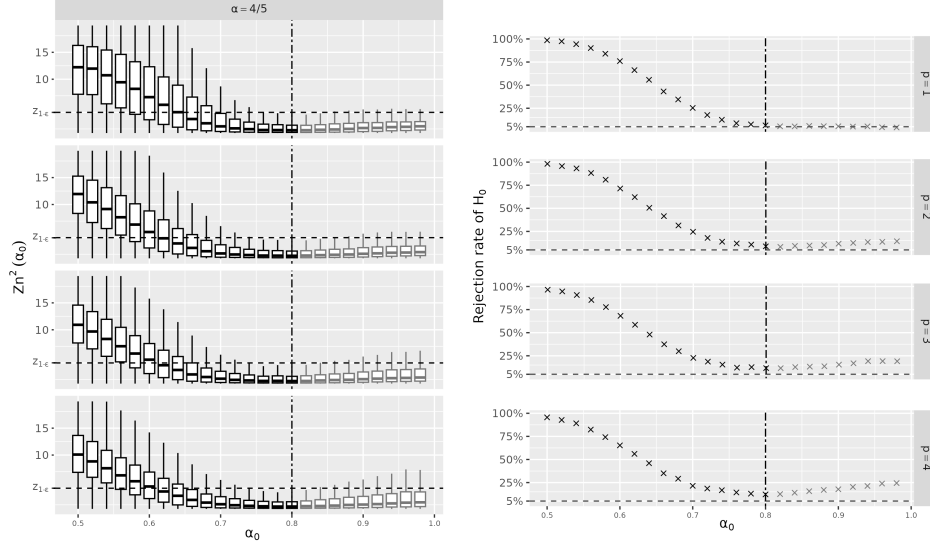
FIGURE 3. Empirical distribution of $Z_n^2(\alpha_0)$ for $n = 1000$, $\lambda_1 = 1$, $\alpha = 4/5$, $\alpha_0 \in \mathcal{A}$ and $p \in \{1, 2, 3, 4\}$ on the left. The corresponding rejection frequencies of $\mathcal{H}_0$: "$\alpha = \alpha_0$" against $\mathcal{H}_1$: "$\alpha > \alpha_0$" (5% level) are represented on the right.
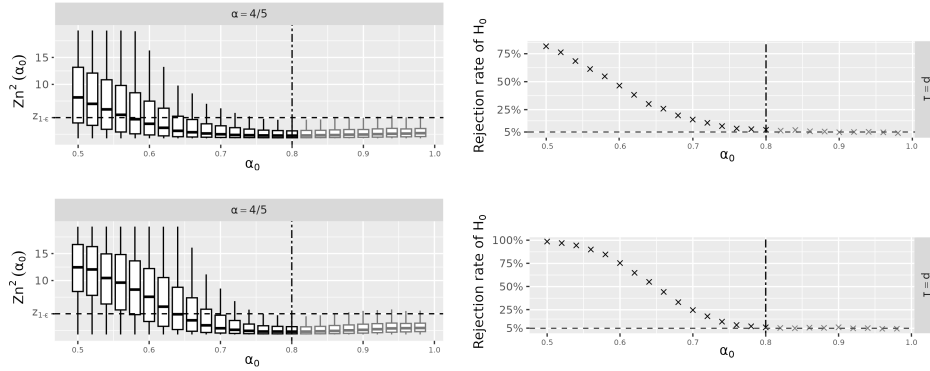

FIGURE 4. Empirical distribution of $Z_n^2(\alpha_0)$ for $n = 250$, $\lambda_1 = 1$, $\alpha = 4/5$, $\alpha_0 \in \mathcal{A}$ and $p = 1$ on the top left. The corresponding rejection frequencies of $\mathcal{H}_0$: "$\alpha = \alpha_0$" against $\mathcal{H}_1$: "$\alpha > \alpha_0$" (5% level) are represented on the right. Same experiments, but with $n = 1000$, $\lambda_1 = -1$, $\alpha = 4/5$, $\alpha_0 \in \mathcal{A}$ and $p = 1$ on the bottom.

In terms of sample sizes, the results are obviously impacted when we set $n = 250$ as can be seen on the upper part of Figure 4 with $p = 1$ and $\alpha = 4/5$ ($\geqslant 50\%$ for $\alpha_0 \leqslant \alpha - 0.2$), but the global behavior follows the same lines. The real time series of next subsection are even smaller ($n = 80$ to $n = 129$), that will mainly result in larger confidence intervals. Under the hypothesis that the limit process has a negative unit root, Figure 4 (on the bottom) also highlights the same efficiency but we must precise here that, unlike $\lambda_1 = 1$, more experiments than those presented here showed a notable degradation with p: the decrease is faster and the statistic falls below $z_{1-\epsilon}$ too quickly. However, despite the lack of power observed for $p \geqslant 2$, this remains an innovative approach since we recall that the ADF, KPSS, and all

usual tests for integration or trend-stationarity are unable to detect instability coming from negative unit roots.
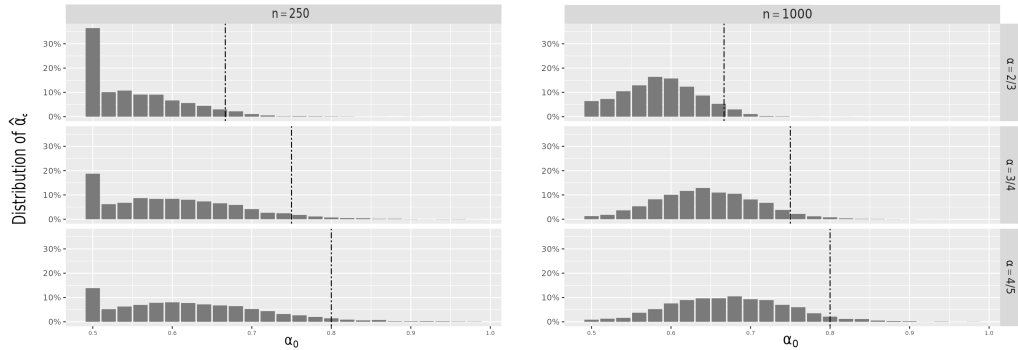

FIGURE 5. Empirical distribution of $\widehat{\alpha}_c$ (5% level) with known value $c = 1$, for $n = 250$ (left) and $n = 1000$ (right), $\lambda_1 = 1$, $\alpha \in \{2/3, 3/4, 4/5\}$, $\alpha_0 \in \mathcal{A}$ and $p = 1$.

As expected, the selected values of α on Figure 5 are underestimated, on average: for $n = 1000$, we can roughly say that $\widehat{\alpha}_{c=1}$ is close to $(\alpha - 0.1)$, yet greater than $(\alpha - 0.2)$ with an overwhelming probability. It is a direct consequence of the construction rule (2.20) and overlapping confidence intervals (in particular, together with $\inf \mathcal{A} = 1/2$, that explains the initial peak on smaller samples). So, the spectral radius must be underestimated as well. Fortunately the next section will show that, by evaluating c together with α instead of arbitrarily setting $c = 1$, the estimation is clearly improved. Of course a fully unbiased strategy may be an interesting trail for future studies (even if the procedure is asymptotically unbiased), see the conclusion for some comments on the topic.

3.2. **On simulated data (c is estimated).** The estimation of c according to the cross-validation described in Section 2.2 actually proves beneficial. Indeed, let us consider three groups of experiments generated as in the previous section, with $n = 1000$, $\lambda_1 = 1$, $p = 2$ and (α, c) uniformly chosen out of $\mathcal{A} \times \mathcal{C} = \{1/2 + k/50, k = 0, \ldots, 24\} \times \{1/2 + k/5, k = 0, \ldots, 22\}$. Figure 6 shows the scatterplot of $\widehat{c}/n^{\widehat{\alpha}}$ as a function of c/n^α together with a linear fit (with 95% confidence interval) for each of them. In the first one, we basically set $c = 1$ and estimate α using (2.20). This corresponds to the decreasing linear fit on the plot and reveals a significant deterioration, hence the inadequacy of making $\widehat{c} = 1$ independent of the data: the bias of $\widehat{\alpha}_{\widehat{c}=1}$ may strongly affect the results. The second group of experiments consists of giving $\widehat{c} = c$ the true value before computing $\widehat{\alpha}_{\widehat{c}=c}$ by the same strategy. The linear fit is clearly better, but the parallelism with the bisector of the coordinate axes unveils a systematic bias. This is a predictable consequence of overlapping confidence intervals of (2.20) already mentioned in the previous section: $\widehat{\alpha}_{\widehat{c}}$ is underestimated so that $c/n^{\widehat{\alpha}_{\widehat{c}}}$ overestimates the true value and the points are generally above the identity function. Finally, in the third group of experiments we estimate the pair (c, α) by the cross-validation of Section 2.2 with $c_0 \in \mathcal{C}$ defined above, $n_{val} = n/10$ and the MSE criterion. One can see that except for the small values of c/n^α (either c is small or α is large) corresponding to a spectral radius close to one and where all methods perform quite similarly, only the last one is able to counterbalance the bias of $\widehat{\alpha}_{\widehat{c}}$ and results in a more reliable estimation. Incidentally, the t-tests for equality

of slopes in the groups confirm the visual intuition through a clear rejection. The latter approach is now going to be used on a real dataset.
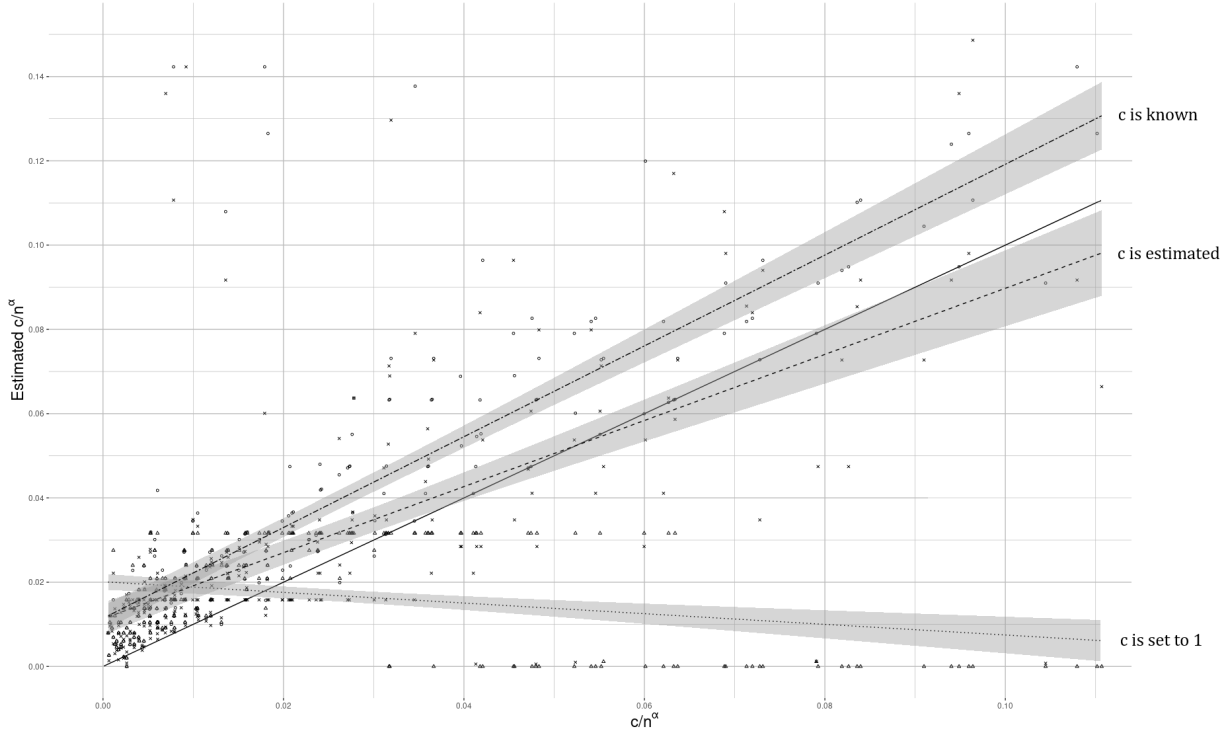

FIGURE 6. Scatterplot of $\widehat{c}/n^{\widehat{\alpha}}$ as a function of c/n^α in three different cases: c is set to 1 (triangles), c is known (circles) and c is estimated by cross-validation (crosses). Experiments are conducted according to the given protocol to simulate the data and estimate α. A linear fit (with 95% confidence interval) shows the mean evolution in each group and the solid line is the identity function.

3.3. **On a real dataset.** The extended dataset `NelPlo` is available as open-source `https://www.rdocumentation.org/packages/tseries/versions/0.9-9/topics/NelPlo` and contains 14 macroeconomic series ending in 1988. According to Nelson and Plosser (1982), those data are known to be either integrated (in the sense of the ADF test) or trend-stationary with spectral radius close to 1, so let us approach them from the nearly-unstable point of view, with positive unit root ($\lambda_1 = 1$). It is important to note that the data of `NelPlo` are longer than the original ones so that our conclusions may be slightly different. The statistical protocol itself is different: p is chosen by looking at the partial autocorrelations of the differentiated series and no trend is estimated. Then, the pair $(\widehat{c}, \widehat{\alpha}_{\widehat{c}})$ is selected through the cross-validation described in Section 2.2, with $n_{val} = n/10$ and the MSE criterion. An (approximate) 95%-confidence interval of $\widehat{\alpha}_{\widehat{c}}$ is also derived from Theorem 2.2 with $c = \widehat{c}$. Because n is usually small, in any case much smaller than in the simulation study, we have restrained the domain of c and only considered a search grid $\mathcal{C} \subset [0.8, 1.2]$, while as before $\mathcal{A} \subset [0.5, 1[$. See Table 1 below for the numerical outcomes.

Although the statistical context is different, our results are rather consistent with Table 5 of Nelson and Plosser (1982). According to these authors, all series except 'Unemployment

Series	n	p	$\widehat{c}$	$\widehat{\alpha}_{\widehat{c}}$	$\widehat{\rho}_n = 1 - \widehat{c}/n^{\widehat{\alpha}_{\widehat{c}}}$
Velocity	120	1	0.80	$\in [0.50, 0.66]$	$\in [0.93, 0.97]$
Industrial production	129	6	0.80	$\in [0.50, 0.72]$	$\in [0.93, 0.98]$
Nominal GNP	80	2	0.80	$\in [0.50, 0.57]$	$\in [0.91, 0.93]$
Consumer prices	129	4	1.08	$\in [0.73, 0.98]$	$\in [0.97, 0.99]$
Employment	99	5	0.80	$\in [0.50, 0.59]$	$\in [0.92, 0.95]$
Interest rate	89	4	0	$+\infty$	1
Wages	89	2	0.80	$\in [0.50, 0.56]$	$\in [0.92, 0.94]$
GNP deflator	100	6	1.20	$\in [0.50, 0.53]$	$\in [0.88, 0.90]$
Money stock	100	3	0.80	$\in [0.50, 0.53]$	$\in [0.92, 0.93]$
Real GNP	80	4	1.10	$\in [0.50, 0.74]$	$\in [0.88, 0.96]$
Common stock prices	118	6	0.82	$\in [0.50, 0.98]$	$\in [0.92, 0.99]$
Real per capita GNP	80	2	0.80	$\in [0.50, 0.62]$	$\in [0.91, 0.95]$
Real wages	89	2	1.04	$\in [0.50, 0.79]$	$\in [0.89, 0.97]$
Unemployment rate	99	3	1.08	$\in [0.50, 0.65]$	$\in [0.89, 0.95]$

TABLE 1. `NelPlo` dataset from the nearly-unstable point of view. For each series, $(\widehat{c}, \widehat{\alpha}_{\widehat{c}})$ coming from the cross-validation of Section 2.2 is given together with an (approximate) 95% confidence interval for $\widehat{\alpha}_{\widehat{c}}$. The estimated spectral radius is also precised.

rate' are integrated. In fact it seems that among them, only 'Interest rate' has a unit root whereas the nearly-unstable approach may be suitable for the others. It should be noted that the only stationary series of the authors (radius ≈ 0.71) corresponds to one of the lowest spectral radius given by our procedure. Similarly, the most integrated series of the authors is 'Interest rate' (radius ≈ 1.03), that is the only one being considered as integrated in our study. There are obviously points of divergence due to different protocols, to the fact that we did not try to detrend the series beforehand, and so on, but it is interesting to note that we share the same general ideas on these data. For illustrative purposes, Figures 7-10 contain some curves from the dataset and associated forecasts.

4. CONCLUSION AND PERSPECTIVES

Our motivation was to bring a fresh look to the unit root issues in time series analysis, through the concept of near-instability. On a dubious series in terms of stationarity, the 'extent of instability' may now be subject to a testing procedure, and many opportunities of improvements should follow from this innovative approach. For example, it would be interesting to unveil the limit behavior of the test statistic with no constraints on α. The light-gray boxplots of Figures 1–3 suggest that $\mathcal{H}_1 :$ "$\alpha < \alpha_0$" does not lead to a divergence but to some kind of stabilization away from zero (in distribution ?). But a spectral radius close to unity together with $\alpha < 1/2$ must correspond to high values of n, questioning the practical interest (a series with moderate size is more likely stable in that context). On the other hand, when $\alpha \geqslant 1/2$, a bilateral test would undoubtedly refine the selection of $\widehat{\alpha}_c$ and contribute to eliminate the observed bias, so this is an important issue. Also, the choice of c presented here is based on empirical foundations, which makes it difficult to obtain theoretical grounds related to consistency or normality. The authors are pretty convinced
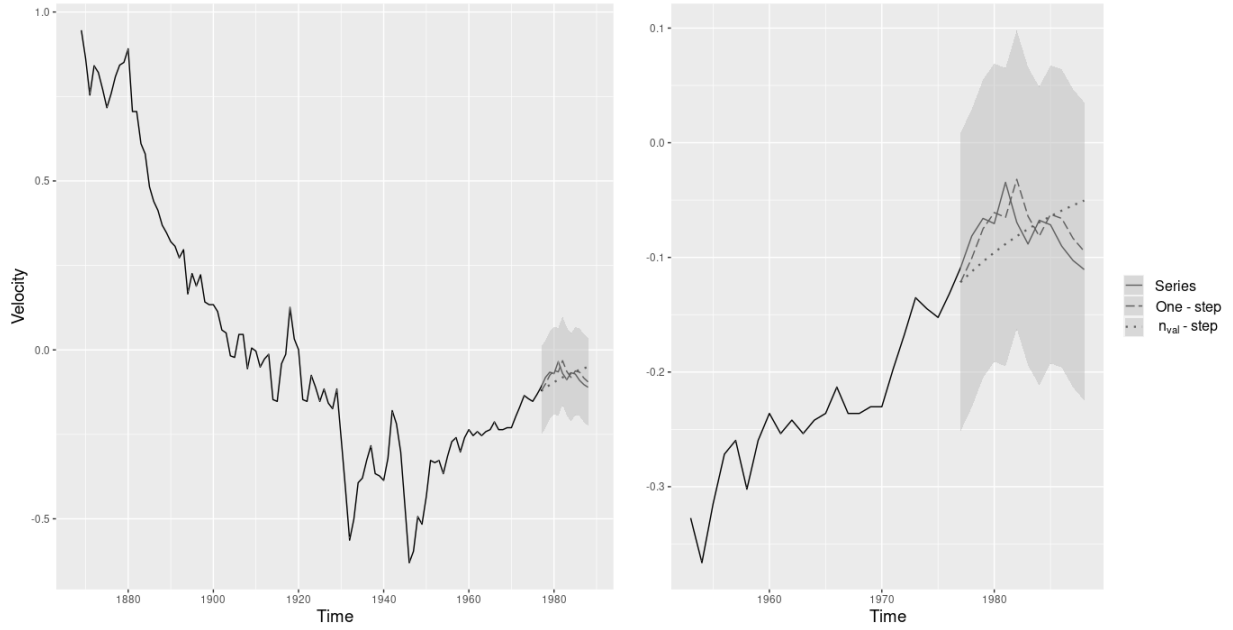
FIGURE 7. Series 'Velocity' and associated predictions for the last $n_{val} = 12$ values (one-step and n_{val}-step forecasts). The Gaussian 95% confidence interval for one-step predictions is added. The whole series is represented on the left whereas on the right a zoom focuses on the last decades.
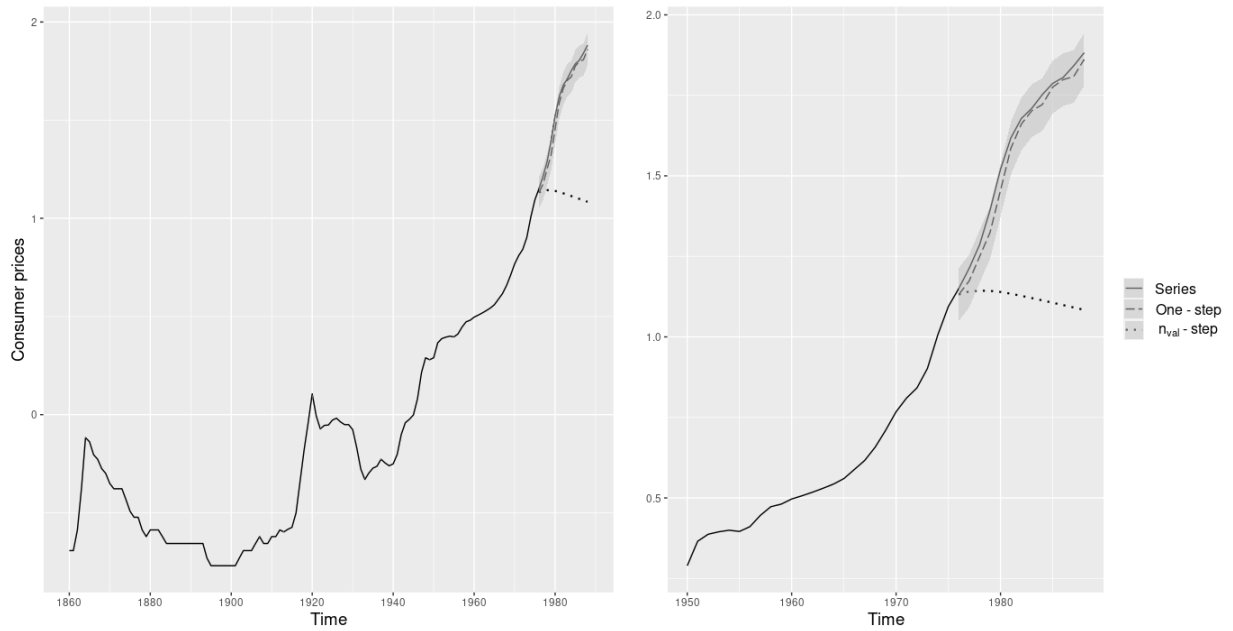

FIGURE 8. Same graphs as in Figure 7 with the series 'Consumer prices' and $n_{val} = 13$ values.

that a joint estimation of (c, α) through a penalized Gaussian likelihood may be a solution to improve the estimation, bypass over-parametrization through well-chosen constraints and
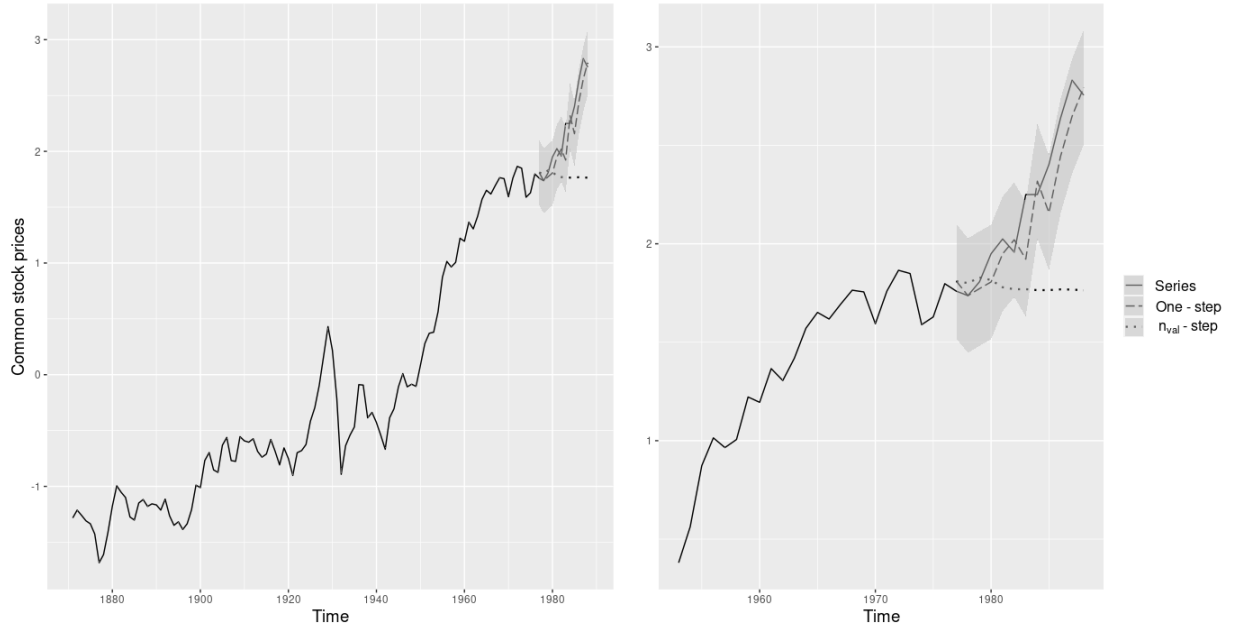
FIGURE 9. Same graphs as in Figure 7 with the series 'Common stock prices' and $n_{val} = 12$ values.
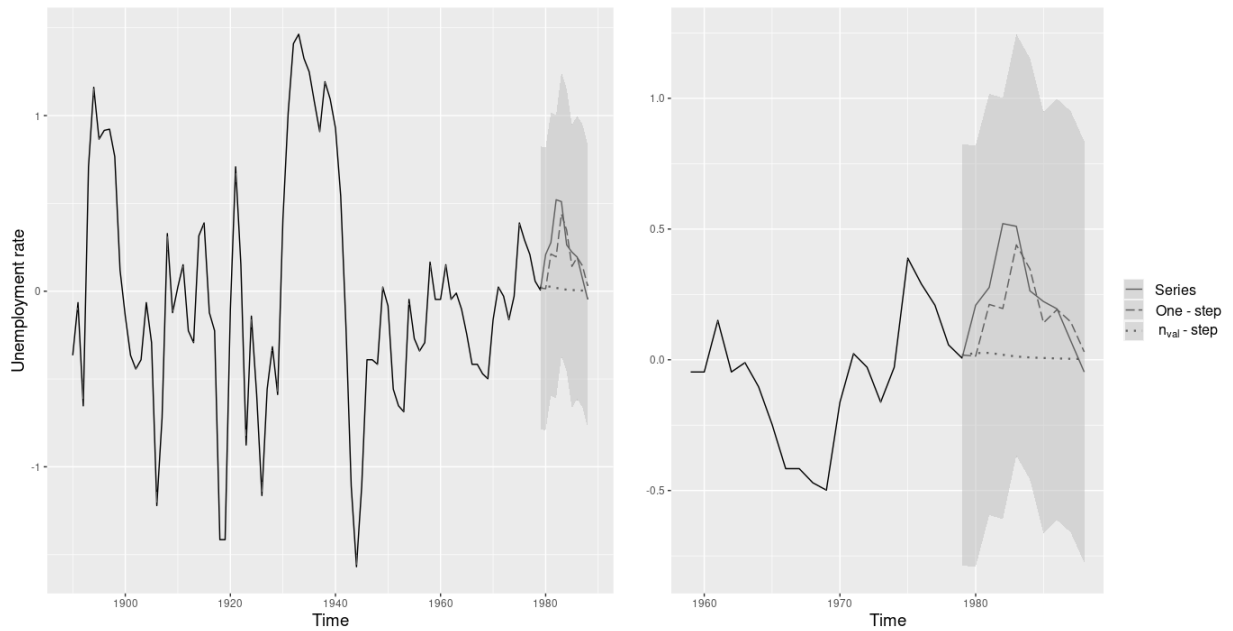
FIGURE 10. Same graphs as in Figure 7 with the series 'Unemployment rate' and $n_{val} = 10$ values.

overall, get guarantees. Authorizing $p = p_n$ to grow to infinity could lead to the formalization of nearly-unstable ARMA(p, q) processes under an hypothesis of invertibility at fixed n and some restrictions on p_n – like the strategy of ADF. The results of this paper shall then probably be generalized without much difficulty to that wider class of models. Especially

as, on a broader scale, any model with a critical parameter could be responsive to this kind of approach: the features of the reproduction law for counting and Galton-Watson processes (see Ispány et al. (2003), Ispány et al. (2005)), the mean value of the random coefficients in RCAR processes, the fractional order of integration in FARIMA processes, the Hurst exponent of long-term memory time series, and so on. For p-dimensional processes, the first step should be to express the stability/instability criterion through the spectral radius of a companion matrix. For example in Barczy et al. (2011), an INAR(p) process with $\rho(A) < 1$ is stable whereas it is unstable for $\rho(A) = 1$ (see A in (2.2), Prop. 2.6 and the definition that follows). Hence, through probably arduous demonstrations, it seems that our approach might be adaptable: let $\rho(A_n) < 1$ with $\rho(A_n) \to 1$, and consider near-unstability in such a count time series to generalize from first-order to p-order. Finally, a natural question arising here concerns explosive nearly-unstable processes, that is when the spectral radius tends to one from the *outer* neighborhood of the unit circle. Although less interesting from a practical point of view, this seems to be a challenging trail for future studies.

5. Technical Developments

To deal with martingale terms, we need to consider the triangular array of filtrations

(5.1) $$\forall\, n \geqslant 1,\ \forall\, 1 \leqslant k \leqslant n, \quad \mathcal{F}_{n,k} = \sigma(\Phi_{n,0}, \varepsilon_1, \ldots, \varepsilon_k)$$

obviously satisfying $\mathcal{F}_{n,0} \subseteq \mathcal{F}_{n,1} \subseteq \ldots \subseteq \mathcal{F}_{n,n}$ for each $n \geqslant 1$.

5.1. **Known results.** The two lemmas given in this section may be found with far more details in Badreau and Proïa (2023).

Lemma 5.1. *Assume that (A1) holds. Then, there exists $n_0 \geqslant 1$ such that, for all $n > n_0$, A_n is diagonalizable in the form $A_n = P_n D_n P_n^{-1}$ with $D_n = \mathrm{diag}(\lambda_{n,1}, \ldots, \lambda_{n,p})$ containing ordered distinct eigenvalues $1 > |\lambda_{n,1}| \geqslant \ldots \geqslant |\lambda_{n,p}| > 0$. In addition, $\|P_n\|$ and $\|P_n^{-1}\|$ are bounded and there exists an invertible matrix P such that $P_n \to P$ and $P_n^{-1} \to P^{-1}$.*

Note that because its spectrum is a subset of the one of A_n, the companion matrix $\bar{A}_n$ associated with $(V_{n,k})$ is also diagonalizable for all $n > n_0$ and the lemma still holds when replacing A_n by $\bar{A}_n$ and obviously p by $p-1$. This will be a key argument to justify (5.3) in the next section.

Lemma 5.2. *Assume that (A1), (A2) and (A3) hold and that $\mathbb{E}[\varepsilon_1^2] = \sigma^2 < +\infty$. Then, as n tends to infinity,*

$$(1 - \rho_n(\alpha)) \frac{S_{n,n}}{n} \overset{p}{\longrightarrow} \Gamma_p \triangleq \sigma^2 \ell \left(\lambda_1^{i+j}\right)_{1 \leqslant i,j \leqslant p} \quad with \quad \ell = \frac{\pi_{11}^2}{2} > 0,$$

where $S_{n,n}$ is given in (2.7) and π_{i1} ($i = 1, \ldots, p$) are the (non-zero) entries of the first column of P^{-1}. Moreover,

$$C_n^{-\frac{1}{2}}(\alpha)\, P_n^{-1} \frac{S_{n,n}}{n} (P_n^T)^{-1} C_n^{-\frac{1}{2}}(\alpha) \overset{p}{\longrightarrow} \sigma^2 \begin{pmatrix} \frac{\pi_{11}^2}{2} & 0 & \cdots & 0 \\ 0 & \frac{\pi_{21}^2}{1-\lambda_2^2} & \cdots & \frac{\pi_{21}\pi_{p1}}{1-\lambda_2\lambda_p} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \frac{\pi_{p1}\pi_{21}}{1-\lambda_p\lambda_2} & \cdots & \frac{\pi_{p1}^2}{1-\lambda_p^2} \end{pmatrix} \triangleq H_p$$

where $C_n(\alpha)$ is defined in (2.8). In addition, H_p is invertible (and positive definite if A has only real eigenvalues).

5.2. Technical lemmas.

Lemma 5.3. *Assume that (A1), (A2) and (A3) hold and that $\mathbb{E}[\varepsilon_1^2] = \sigma^2 < +\infty$. Then, as n tends to infinity, for any $\alpha/2 < \alpha_0 \leqslant \alpha$,*

$$C_n^{-\frac{1}{2}}(\alpha)\frac{T_{n,n}(\alpha_0)}{n}C_n^{-\frac{1}{2}}(\alpha) \xrightarrow{p} \Upsilon_p \overset{\Delta}{=} \sigma^2 \operatorname{diag}(\ell, \Sigma_{p-1})$$

where $T_{n,n}(\alpha)$ is defined in (2.7), $C_n(\alpha)$ in (2.8), and where ℓ and Σ_{p-1} are defined in Lemma 5.2 and (2.3), respectively. In addition, Υ_p is positive definite.

Proof. Note that $S_{n,n}$ and $T_{n,n}(\alpha)$ share the same top-left element (which actually does not depend on α), so we directly get

$$(5.2) \qquad \frac{1-\rho_n(\alpha)}{n}\sum_{k=0}^{n}X_{n,k}^2 = \frac{c}{n^{1+\alpha}}\sum_{k=0}^{n}X_{n,k}^2 \xrightarrow{p} \sigma^2\ell.$$

Moreover, coming back to the reasoning of Badreau and Proïa (2023), especially Lem. 13–15 and Prop. 2, and adapting it to the present case (see the remark after Lemma 5.1),

$$(5.3) \quad \frac{1}{n}\sum_{k=0}^{n}\Lambda_{n,k}(\alpha)\,\Lambda_{n,k}^T(\alpha) \xrightarrow{p} \sigma^2\Sigma_{p-1} \quad \text{where} \quad \Lambda_{n,k}^T(\alpha) = \begin{pmatrix} V_{n,k}(\alpha) & \cdots & V_{n,k-p+2}(\alpha) \end{pmatrix}.$$

Indeed, the equivalence

$$\frac{1}{n}\sum_{k=0}^{n}\Lambda_{n,k}(\alpha)\,\Lambda_{n,k}^T(\alpha) \overset{p}{\sim} \left(\sum_{k\geqslant 0}\bar{A}_n^k\,K_{p-1}\,(\bar{A}_n^T)^k\right)\frac{1}{n}\sum_{k=1}^{n}\varepsilon_k^2,$$

the law of large numbers together with the fact that $\rho(\bar{A}_n) < 1$ and $\rho(\bar{A}) < 1$ lead to (5.3). It remains to treat the (symmetrical) non-diagonal blocks. First,

$$(5.4) \qquad \sum_{k=0}^{n}\Lambda_{n,k}^{(p)}(\alpha)\,X_{n,k} = \Lambda_{n,0}^{(p)}(\alpha)X_{n,0} + \left[\sum_{k=1}^{n}(\Phi_{n,k} - \rho_n(\alpha)\,\Phi_{n,k-1})\,\Phi_{n,k}^T\right]e_p$$

$$= \left[S_{n,n}(I_p - \rho_n(\alpha)A_n^T) - \rho_n(\alpha)M_n + r_n\right]e_p$$

where $\Lambda_{n,k}^{(p)}$ is the p-dimensional version of $\Lambda_{n,k}$ (with last element $V_{n,k-p+1}(\alpha)$),

$$M_n = \sum_{k=1}^{n}\Phi_{n,k-1}\,E_k^T$$

and r_n is made of isolated terms. It is already proven (Lem. 14–15) that $\|M_n\|$ and $\|r_n\|$ are $o_p(n)$. As an intermediate result, we can see that for $n > n_0$,

$$(5.5) \qquad C_n^{\frac{1}{2}}(\alpha)\,P_n^T\left(I_p - \rho_n(\alpha)A_n^T\right) = C_n^{\frac{1}{2}}(\alpha)\left(I_p - \rho_n(\alpha)D_n\right)P_n^T$$

$$\xrightarrow{p} \operatorname{diag}\left(0, (1-\lambda_i)_{2\leqslant i\leqslant p}\right)P^T \overset{\Delta}{=} \Delta_p\,P^T.$$

Hence,

$$\frac{\sqrt{1-\rho_n(\alpha)}}{n}\,S_{n,n}\left(I_p - \rho_n(\alpha)A_n^T\right) = P_n\frac{C_n^{\frac{1}{2}}(\alpha)\sqrt{c}}{n^{\frac{\alpha}{2}}}\left[C_n^{-\frac{1}{2}}(\alpha)\,P_n^{-1}\frac{S_{n,n}}{n}\,(P_n^T)^{-1}C_n^{-\frac{1}{2}}(\alpha)\right]$$

$$\times\left[C_n^{\frac{1}{2}}(\alpha)\,P_n^T\left(I_p - \rho_n(\alpha)A_n^T\right)\right]$$

$$(5.6) \qquad\qquad\qquad \xrightarrow{p} P\,K_p\,H_p\,\Delta_p\,P^T$$

17

where the block diagonal matrix H_p is given in Lemma 5.2 and Δ is defined above. But K_p, H_p and Δ_p have particular structures consisting of zeros in key locations and implying the following pattern,

$$P K_p H_p \Delta_p = \begin{pmatrix} * & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ \vdots & \vdots & & \vdots \\ * & 0 & \cdots & 0 \end{pmatrix} \begin{pmatrix} 0 & \cdots & \cdots & 0 \\ \vdots & * & \cdots & * \\ \vdots & \vdots & & \vdots \\ 0 & * & \cdots & * \end{pmatrix} = 0.$$

The combination of (5.2), (5.3), (5.6) and the remarks that follow lead to the convergence

(5.7) $$C_n^{-\frac{1}{2}}(\alpha) \frac{T_{n,n}(\alpha)}{n} C_n^{-\frac{1}{2}}(\alpha) \xrightarrow{p} \Upsilon_p$$

that is, when α_0 takes the 'true' value α. We now wish to generalize the result to $T_{n,n}(\alpha_0)$ for any $\alpha/2 < \alpha_0 \leqslant \alpha$. Regarding this, note that

(5.8) $$\Psi_{n,k}(\alpha_0) = \Psi_{n,k}(\alpha) + c\left(n^{-\alpha_0} - n^{-\alpha}\right) \dot{\Phi}_{n,k}$$

where

$$\forall n \geqslant 1, \ \forall 1 \leqslant k \leqslant n, \quad \dot{\Phi}_{n,k}^T = \begin{pmatrix} 0 & X_{n,k-1} & \cdots & X_{n,k-p+1} \end{pmatrix}.$$

Thus, we have the decomposition

(5.9) $$T_{n,n}(\alpha_0) = T_{n,n}(\alpha) + R_n$$

with

$$R_n = c\left(n^{-\alpha_0} - n^{-\alpha}\right) \left[\sum_{k=0}^{n} \Psi_{n,k}(\alpha) \dot{\Phi}_{n,k}^T + \sum_{k=0}^{n} \dot{\Phi}_{n,k} \Psi_{n,k}^T(\alpha) \right]$$
$$+ c^2 \left(n^{-\alpha_0} - n^{-\alpha}\right)^2 \sum_{k=0}^{n} \dot{\Phi}_{n,k} \dot{\Phi}_{n,k}^T \ \triangleq \ R_{n,1} + R_{n,1}^T + R_{n,2}.$$

The direct computation shows that $\Psi_{n,k}(\alpha) \dot{\Phi}_{n,k}^T$ contains a column of zeros, a line made of $X_{n,k} X_{n,k-i}$ $(1 \leqslant i \leqslant p-1)$, and a $(p-1) \times (p-1)$ block of terms of the form $V_{n,k-j}(\alpha) X_{n,k-i}$ $(0 \leqslant j \leqslant p-2, 1 \leqslant i \leqslant p-1)$. For $2 \leqslant i \leqslant p-1$, according to the first expression in (2.2),

$$\rho_n^{2i}(\alpha) X_{n,k-i} = \rho_n^{2i-1}(\alpha) \left(X_{n,k-i+1} - V_{n,k-i+1}(\alpha) \right)$$
$$= \rho_n^{i+1}(\alpha) X_{n,k-1} - \sum_{j=1}^{i-1} \rho_n^{2i-j}(\alpha) V_{n,k-i+j}(\alpha)$$

which entails that, for any $0 \leqslant u \leqslant p-2$,

$$\rho_n^{2i}(\alpha) \sum_{k=1}^{n} V_{n,k-u}(\alpha) X_{n,k-i} = \rho_n^{i+1}(\alpha) \sum_{k=1}^{n} V_{n,k-u}(\alpha) X_{n,k-1}$$

(5.10) $$- \sum_{j=1}^{i-1} \rho_n^{2i-j}(\alpha) \sum_{k=1}^{n} V_{n,k-u}(\alpha) V_{n,k-i+j}(\alpha).$$

The first part of this equation is $o_p(n^{1+\alpha/2})$ and the second part is a finite sum of $O_p(n)$, see for instance (5.7). Thus, via Lemma 5.2,

(5.11) $$\left\| C_n^{-\frac{1}{2}}(\alpha) \frac{R_{n,1}}{n} C_n^{-\frac{1}{2}}(\alpha) \right\| = O_p\left(\frac{n^{-\alpha_0} - n^{-\alpha}}{n^{1+\frac{\alpha}{2}}} n^{1+\alpha} \right) = o_p(1)$$

18

as soon as $\alpha_0 > \alpha/2$. In the same way, $R_{n,2}$ is closely related to the covariance matrix $S_{n,n}$ but with zeros on the first line and column. So,

$$(5.12) \qquad \left\| C_n^{-\frac{1}{2}}(\alpha) \frac{R_{n,2}}{n} C_n^{-\frac{1}{2}}(\alpha) \right\| = O_p\left(\frac{(n^{-\alpha_0} - n^{-\alpha})^2}{n} n^{1+\alpha} \right) = o_p(1)$$

provided that $\alpha_0 > \alpha/2$. Using (5.7), (5.9), (5.11) and (5.12), the convergence is now established. Since $\ell > 0$ (Lemma 5.2), the positive-definiteness of Υ_p relies on the invertibility of Σ_{p-1}, which can be found $e.g.$ in Brockwell and Davis (1991) in the context of a stable $AR(p-1)$ process with $\sigma^2 > 0$. $\qquad \square$

Lemma 5.4. *Assume that (A1), (A2) and (A3) hold and that* $\mathbb{E}[|\varepsilon_1|^{2+\nu}] = \eta_\nu < +\infty$. *Then, as n tends to infinity, for the 'true' value of α,*

$$(5.13) \qquad \sqrt{n}\, C_n^{\frac{1}{2}}(\alpha) \left(\widehat{\vartheta}_n(\alpha) - \vartheta_n(\alpha) \right) \xrightarrow{d} \mathcal{N}\left(0, \sigma^2 \Upsilon_p^{-1} \right)$$

where the limit covariance Υ_p is given in Lemma 5.3. Moreover,

$$(5.14) \qquad \inf_{\alpha/2 < \alpha_0 < \alpha} \left\| \sqrt{n}\, C_n^{\frac{1}{2}}(\alpha_0) \left(\widehat{\vartheta}_n(\alpha_0) - \vartheta_n(\alpha_0) \right) \right\| \xrightarrow{p} +\infty.$$

Proof. Developing $X_{n,k} = \vartheta_n^T(\alpha) \Psi_{n,k-1}(\alpha) + \varepsilon_k$ in (2.10), the estimation error satisfies

$$(5.15) \qquad \sqrt{n}\, C_n^{\frac{1}{2}}(\alpha) \left(\widehat{\vartheta}_n(\alpha) - \vartheta_n(\alpha) \right) = \sqrt{n}\, C_n^{\frac{1}{2}}(\alpha) T_{n,n-1}^{-1}(\alpha) \sum_{k=1}^{n} \Psi_{n,k-1}(\alpha)\, \varepsilon_k.$$

The sequence $(\Psi_{n,k-1}(\alpha)\, \varepsilon_k)_{n,k}$ is a martingale-difference array w.r.t. $\mathcal{F}_{n,k}$ defined in (5.1) so, for all $u \in \mathbb{R}^p \backslash \{0\}$,

$$(5.16) \qquad M_{n,n}^{(u)} = u^T \frac{C_n^{-\frac{1}{2}}(\alpha)}{\sqrt{n}} \sum_{k=1}^{n} \Psi_{n,k-1}(\alpha)\, \varepsilon_k$$

is a martingale array having predictable quadratic variation

$$\langle M^{(u)} \rangle_{n,n} \quad = \quad \sigma^2 u^T C_n^{-\frac{1}{2}}(\alpha) \frac{T_{n,n-1}(\alpha)}{n} C_n^{-\frac{1}{2}}(\alpha)\, u$$

$$(5.17) \qquad \qquad \qquad \xrightarrow{p} \quad \sigma^2 u^T \Upsilon_p u > 0$$

by the convergence in Lemma 5.3 (with $\alpha_0 = \alpha$) and the positive definiteness of Υ_p. It is now necessary to establish the Lindeberg's condition in order to apply the central limit theorem for arrays of martingales (see $e.g.$ Thm. 1 of Sec. 8 of Pollard (1984)), namely that

$$(5.18) \qquad \forall\, \epsilon > 0, \quad \sum_{k=1}^{n} \mathbb{E}\left[(\Delta M_{n,k}^{(u)})^2\, \mathbb{1}_{\{|\Delta M_{n,k}^{(u)}| > \epsilon\}}\, \big|\, \mathcal{F}_{n,k-1} \right] \xrightarrow{p} 0$$

where $\Delta M_{n,k}^{(u)} = M_{n,k}^{(u)} - M_{n,k-1}^{(u)}$. But for any $1 \leqslant k \leqslant n$,

$$(5.19) \qquad \mathbb{E}\left[(\Delta M_{n,k}^{(u)})^2\, \mathbb{1}_{\{|\Delta M_{n,k}^{(u)}| > \epsilon\}}\, \big|\, \mathcal{F}_{n,k-1} \right] = \varphi_{n,k-1}^{(u)}(\alpha) \frac{\xi_{n,k}}{n} \leqslant \varphi_{n,k-1}^{(u)}(\alpha) \frac{\xi_n^*}{n}$$

where $\varphi_{n,k-1}^{(u)}(\alpha) = u^T C_n^{-1/2}(\alpha)\, \Psi_{n,k-1}(\alpha)\, \Psi_{n,k-1}^T(\alpha)\, C_n^{-1/2}(\alpha)\, u$,

$$\xi_{n,k} = \mathbb{E}\left[\varepsilon_k^2\, \mathbb{1}_{\{|\Delta M_{n,k}^{(u)}| > \epsilon\}}\, \big|\, \mathcal{F}_{n,k-1} \right] \quad \text{and} \quad \xi_n^* = \sup_{1 \leqslant k \leqslant n} \xi_{n,k}.$$

19

Applying Hölder's and Markov's inequalities,

$$\forall\, 1 \leqslant k \leqslant n, \quad \xi_{n,k} \;\leqslant\; \eta_\nu^{\frac{2}{2+\nu}} \, \mathbb{P}\big((\Delta M_{n,k}^{(u)})^2 > \epsilon^2 \,|\, \mathcal{F}_{n,k-1}\big)^{\frac{\nu}{2+\nu}}$$

$$\leqslant\; \eta_\nu^{\frac{2}{2+\nu}} \left(\frac{\sigma^2 \, \varphi_{n,k-1}^{(u)}(\alpha)}{\epsilon^2 \, n} \right)^{\frac{\nu}{2+\nu}} \;\xrightarrow{\;p\;}\; 0$$

from Lemma 5.3 (with $\alpha_0 = \alpha$), where η_ν is the moment of order $2 + \nu$ of the white noise (ε_k). Thus $\xi_n^* = o_p(1)$, see $e.g.$ Lem. 1.3.20 of Duflo (1997). Together with (5.19) and again Lemma 5.3, (5.18) follows. Consequently, the martingale defined in (5.16) satisfies

$$M_{n,n}^{(u)} \;\xrightarrow{\;d\;}\; \mathcal{N}(0, \sigma^2 \, u^T \Upsilon_p \, u)$$

and because that holds for any $u \neq 0$, by the Cramér-Wold device,

$$(5.20) \qquad \frac{C_n^{-\frac{1}{2}}(\alpha)}{\sqrt{n}} \sum_{k=1}^n \Psi_{n,k-1}(\alpha)\, \varepsilon_k \;\xrightarrow{\;d\;}\; \mathcal{N}\big(0, \sigma^2 \,\Upsilon_p\big).$$

Combined with (5.17), Slutsky's lemma and (5.15),

$$\sqrt{n}\, C_n^{\frac{1}{2}}(\alpha) \big(\widehat{\vartheta}_n(\alpha) - \vartheta_n(\alpha)\big) \;\xrightarrow{\;d\;}\; \mathcal{N}\big(0, \sigma^2 \,\Upsilon_p^{-1}\big).$$

That concludes the proof of the first part of the lemma. For the other part, note that using (2.6), (5.8) and (2.10), we have the decomposition

$$(5.21) \qquad \widehat{\vartheta}_n(\alpha_0) - \vartheta_n(\alpha_0) = \big[\vartheta_n(\alpha) - \vartheta_n(\alpha_0)\big] + T_{n,n-1}^{-1}(\alpha_0) \sum_{k=1}^n \Psi_{n,k-1}(\alpha_0)\, \varepsilon_k$$

$$+\; c\,(n^{-\alpha} - n^{-\alpha_0})\, T_{n,n-1}^{-1}(\alpha_0) \sum_{k=1}^n \Psi_{n,k-1}(\alpha_0)\, \dot{\Phi}_{n,k-1}^T \, \vartheta_n(\alpha)$$

$$\triangleq\; \tilde{R}_{n,1} + \tilde{R}_{n,2} + \tilde{R}_{n,3}.$$

The reasoning we have just developed also applies to $\tilde{R}_{n,2}$ via the same lines and tools, and as a result we obtain the convergence

$$(5.22) \qquad \sqrt{n}\, C_n^{\frac{1}{2}}(\alpha)\, \tilde{R}_{n,2} \;\xrightarrow{\;d\;}\; \mathcal{N}\big(0, \sigma^2 \,\Upsilon_p^{-1}\big).$$

That leads to the first intermediate step,

$$(5.23) \qquad \left| e_p^T \sqrt{n}\, C_n^{\frac{1}{2}}(\alpha_0)\, \tilde{R}_{n,2} \right| \;=\; \left| \sqrt{c\,n^{\alpha_0-\alpha}}\; e_p^T \sqrt{n}\, C_n^{\frac{1}{2}}(\alpha)\, \tilde{R}_{n,2} \right| \;=\; o_p(1)$$

as soon as $\alpha_0 < \alpha$. We can consider the rest of the decomposition similarly, that is

$$(5.24) \qquad \left| e_p^T \sqrt{n}\, C_n^{\frac{1}{2}}(\alpha_0) \big(\tilde{R}_{n,1} + \tilde{R}_{n,3}\big) \right| \;=\; \left| \sqrt{c\,n^{1-\alpha_0}}\,(1 - n^{\alpha_0-\alpha}) \right| \left| 1 - u_n \right|$$

where the first factor is clearly divergent when $\alpha_0 < \alpha$ and

$$u_n \;\triangleq\; e_p^T \, T_{n,n-1}^{-1}(\alpha_0) \left[\sum_{k=1}^n \Psi_{n,k-1}(\alpha_0)\, \dot{\Phi}_{n,k-1}^T \right] \vartheta_n(\alpha).$$

20

As in the proof of Lemma 5.3, it is easy to see that $\Psi_{n,k-1}(\alpha_0)\,\dot{\Phi}^T_{n,k-1}$ is a $p\times p$ matrix made of a first column of zeros, terms of the form $X_{n,k-1}\,X_{n,k-j}$ on the first line and $V_{n,k-i}(\alpha_0)\,X_{n,k-j}$ in the bottom-right block $(1 \leqslant i \leqslant p-1,\ 2 \leqslant j \leqslant p)$, broken down as

$$\rho_n^{2j}(\alpha)\sum_{k=1}^n V_{n,k-i}(\alpha_0)\,X_{n,k-j} \;=\; \rho_n^{j+1}(\alpha)\sum_{k=1}^n V_{n,k-i}(\alpha_0)\,X_{n,k-1}$$
$$-\sum_{s=1}^{j-1}\rho_n^{2j-s}(\alpha)\sum_{k=1}^n V_{n,k-i}(\alpha_0)\,V_{n,k-j+s}(\alpha)$$

as it is done in (5.10). As soon as $\alpha/2 < \alpha_0 < \alpha$, Lemma 5.3 entails that the first part of this equation is $o_p(n^{1+\alpha/2})$ and, using Cauchy–Schwarz inequality,

$$\left|\sum_{k=1}^n V_{n,k-i}(\alpha_0)\,V_{n,k-j+l}(\alpha)\right| \;\leqslant\; \sqrt{\sum_{k=1}^n V^2_{n,k-i}(\alpha_0)}\,\sqrt{\sum_{k=1}^n V^2_{n,k-j+l}(\alpha)}$$

where both sums are $O_p(n)$. Finally, it is straightforward to obtain the simplification

$$(5.25)\qquad u_n \xrightarrow{\;p\;} e_p^T\,\Upsilon_p^{-1}\,\sigma^2\,\ell\begin{pmatrix}0 & 1 & \cdots & 1\\ \vdots & 0 & \cdots & 0\\ \vdots & \vdots & & \vdots\\ 0 & 0 & \cdots & 0\end{pmatrix}\begin{pmatrix}1\\ \vdots\\ \beta_i\\ \vdots\end{pmatrix} = \sum_{i=1}^{p-1}\beta_i \triangleq s_\beta$$

where ℓ and Υ are defined in Lemmas 5.2 and 5.3, respectively. Note that $1-s_\beta$ is obviously non-zero since by stability, the polynomial $\Theta_\beta(z) = 1-\beta_1 z - \ldots - \beta_{p-1}z^{p-1}$ has no zero on the unit circle. Going back to (5.23) and (5.24), we then have, for any $\alpha/2 < \alpha_0 < \alpha$,

$$(5.26)\qquad \left|e_p^T\sqrt{n}\,C_n^{\frac{1}{2}}(\alpha_0)\left(\widehat{\vartheta}_n(\alpha_0)-\vartheta_n(\alpha_0)\right)\right| \xrightarrow{\;p\;} +\infty$$

which concludes the proof. $\qquad\square$

5.3. Proof of Proposition 2.1. First, note that

$$\sqrt{n}\left(\widetilde{\theta}_n-\theta_n\right) \;=\; \sqrt{n}\left(\widehat{J}_n-J_n\right)\left(\widehat{\beta}_n-\beta_n\right)+\sqrt{n}\left(\widehat{J}_n-J_n\right)\beta_n$$
$$+\sqrt{n}\,J_n\left(\widehat{\beta}_n-\beta_n\right)+\sqrt{n}\left(\widehat{v}_n-\lambda_{n,1}\right)e_p.$$

In fact, the proof rests entirely on Lemma 5.4. The first line of the asymptotic normality gives $|\widehat{v}_n-\lambda_{n,1}| = O_p(n^{-(1+\alpha)/2}) = o_p(n^{-1/2})$ and obviously $\|\widehat{J}_n-J_n\| = o_p(n^{-1/2})$. Now from the remaining part of (5.13),

$$\sqrt{n}\left(\widehat{\beta}_n-\beta_n\right) \xrightarrow{\;d\;} \mathcal{N}\left(0,\,\Sigma_{p-1}^{-1}\right)$$

where the (invertible) covariance Σ_{p-1} is given in (2.3). So $\|\widehat{J}_n-J_n\|\,\|\widehat{\beta}_n-\beta_n\| = o_p(n^{-1/2})$ and the result follows from (A1), that is from $J_n \to J$. $\qquad\square$

5.4. Proof of Theorem 2.2. Taking the first line of the convergence in Lemma 5.4,

$$(5.27)\qquad \sqrt{n^{1+\alpha}}\left(\widehat{v}_n(\alpha)-\rho_n(\alpha)\right) \xrightarrow{\;d\;} \mathcal{N}\left(0,\,c\,\ell^{-1}\right)$$

where we recall that $\widehat{v}_n(\alpha)$ is the first component of $\widehat{\vartheta}_n(\alpha)$, $\rho_n(\alpha) = 1-cn^{-\alpha}$ and $\ell > 0$ is given in Lemma 5.2. Now define the sequence of $\mathcal{C}^1$-mappings from $]0,1[$ to $\mathbb{R}$ given by

$$\forall\,n > 1,\qquad g_n(x) = \frac{\ln c - \ln(1-x)}{\ln n}$$

so that $\alpha = g_n(\rho_n(\alpha))$ and $\widehat{\alpha}_n(\alpha) = g_n(\widehat{v}_n(\alpha))$. Using Taylor's expansion, there is a sequence (ϑ_n^*) such that

$$\widehat{\alpha}_n(\alpha) = \alpha + \big(\widehat{v}_n(\alpha) - \rho_n(\alpha)\big) g_n'(\vartheta_n^*)$$

with, for all $n > 1$, $(\rho_n(\alpha) \wedge \widehat{v}_n(\alpha)) < \vartheta_n^* < (\rho_n(\alpha) \vee \widehat{v}_n(\alpha))$. So,

(5.28) $$(\ln n)\sqrt{n^{1-\alpha}}\,(\widehat{\alpha}_n(\alpha) - \alpha) = \sqrt{n^{1+\alpha}}\,\big(\widehat{v}_n(\alpha) - \rho_n(\alpha)\big)\frac{n^{-\alpha}}{1-\vartheta_n^*}.$$

We also note that

$$n^\alpha\big(1 - \rho_n(\alpha) \vee \widehat{v}_n(\alpha)\big) \;<\; n^\alpha\big(1 - \vartheta_n^*\big) \;<\; n^\alpha\big(1 - \rho_n(\alpha) \wedge \widehat{v}_n(\alpha)\big)$$

and

$$\begin{aligned} n^\alpha\big(1 - \widehat{v}_n(\alpha)\big) &= n^\alpha\big(1 - \rho_n(\alpha)\big) + n^\alpha\big(\rho_n(\alpha) - \widehat{v}_n(\alpha)\big) \\ &= c + O_p\big(\sqrt{n^{\alpha-1}}\big) \xrightarrow{p} c \end{aligned}$$

from Lemma 5.4 and since $\alpha < 1$. That clearly implies that

(5.29) $$n^\alpha\big(1 - \vartheta_n^*\big) \xrightarrow{p} c.$$

The result follows from (5.27), (5.28) and (5.29) together with Slutsky's lemma. $\qquad\square$

5.5. Proof of Theorem 2.3.

First, a direct computation of the entries of P^{-1} shows that

$$\pi_{11} = \frac{1}{\prod_{j=2}^{p}(1 - \lambda_j)} \quad (= 1 \text{ if } p = 1),$$

see (2.16) with $\lambda_1 = 1$. We know by Rem. 2.2 in Badreau and Proïa (2023) that the OLS of θ_n defined in (2.9) is consistent for θ, so obviously

$$\widehat{A}_n \xrightarrow{p} A$$

also holds. Since the eigenvalues of a square matrix are continuous functions of its entries, see e.g. Thm. 2.4.9.2 of Horn and Johnson (2012), it follows by the continuous mapping theorem that

$$\widehat{\pi}_n \xrightarrow{p} \pi_{11}.$$

Then, taking for α_0 the 'true' value α and using Slutsky's lemma, (2.18) is a straightforward consequence of Theorem 2.2. Now suppose that $\alpha/2 < \alpha_0 < \alpha$. Coming back to (5.28) with α_0, through the reasoning of the previous proof, there is a sequence (ϑ_n^{**}) such that

(5.30) $$(\ln n)\sqrt{n^{1-\alpha_0}}\,(\widehat{\alpha}_n(\alpha_0) - \alpha_0) = \sqrt{n^{1+\alpha_0}}\,\big(\widehat{v}_n(\alpha_0) - \rho_n(\alpha_0)\big)\frac{n^{-\alpha_0}}{1-\vartheta_n^{**}}$$

where $(\rho_n(\alpha_0) \wedge \widehat{v}_n(\alpha_0)) < \vartheta_n^{**} < (\rho_n(\alpha_0) \vee \widehat{v}_n(\alpha_0))$. Using again the decomposition (5.21), we first obtain with (5.22),

$$\big|n^{\alpha_0} e_p^T \tilde{R}_{n,2}\big| = O_p\big(\sqrt{n^{2\alpha_0-1-\alpha}}\big) \xrightarrow{p} 0$$

since $2\alpha_0 < \alpha + 1$. Then,

$$n^{\alpha_0} e_p^T\big(\tilde{R}_{n,1} + \tilde{R}_{n,3}\big) = c\,(1 - n^{\alpha_0-\alpha})\,(1 - u_n) \xrightarrow{p} c\,(1 - s_\beta)$$

where u_n has been introduced in (5.24) and its limit behavior established in (5.25). Hence,

$$n^{\alpha_0}\big(1 - \widehat{v}_n(\alpha_0)\big) = c + n^{\alpha_0}\big(\rho_n(\alpha_0) - \widehat{v}_n(\alpha_0)\big) \xrightarrow{p} c\,s_\beta.$$

This is sufficient to get that, for any $\alpha/2 < \alpha_0 < \alpha$,

$$(5.31) \qquad \liminf_{n \to +\infty} \frac{n^{-\alpha_0}}{|1 - \vartheta_n^{**}|} \geqslant \frac{1}{c\,(1 \vee |s_\beta|)} > 0.$$

Finally, (5.26) entails

$$(5.32) \qquad \left| \sqrt{n^{1+\alpha_0}} \left(\widehat{v}_n(\alpha_0) - \rho_n(\alpha_0) \right) \right| \xrightarrow{\ p\ } +\infty.$$

The combination of (5.30), (5.31) and (5.32) leads to (2.19), which concludes the proof. $\square$

5.6. **The case of a negative unit root.** Our results have all been established for $\lambda_1 = 1$ but they remain true in a context of a negative unit root, at the cost of some slight adjustments. In what follows, it is thus implied that $\lambda_1 = -1$ so that we may now consider, $\lambda_{n,1} = \lambda_1 \rho_n(\alpha) = -\rho_n(\alpha)$. The immediate consequence is that, in the definition of the latent variables $V_{n,k}$ in (2.2), $V_{n,k}(\alpha) = X_{n,k} + \rho_n(\alpha) X_{n,k-1}$. So, in the proof of Lemma 5.3, $\rho_n(\alpha)$ must be replaced by $-\rho_n(\alpha)$ in (5.4) and (5.10), and in particular that leads to

$$
\begin{aligned}
C_n^{\frac{1}{2}}(\alpha)\, P_n^T \left(I_p + \rho_n(\alpha) A_n^T \right) \quad &= \quad C_n^{\frac{1}{2}}(\alpha)\, (I_p + \rho_n(\alpha) D_n)\, P_n^T \\
&\xrightarrow{\ p\ } \quad \operatorname{diag}\left(0, (1+\lambda_i)_{2 \leqslant i \leqslant p}\right) P^T \triangleq \Delta_p\, P^T
\end{aligned}
$$

in place of (5.5). We may also note that c merely becomes $-c$ in (5.8), (5.9) and (5.21). The sequence (u_n) defined in (5.24) now satisfies

$$
u_n \xrightarrow{\ p\ } e_p^T \Upsilon_p^{-1} \sigma^2 \ell
\begin{pmatrix}
0 & -1 & \cdots & (-1)^{p-1} \\
\vdots & 0 & \cdots & 0 \\
\vdots & \vdots & & \vdots \\
0 & 0 & \cdots & 0
\end{pmatrix}
\begin{pmatrix}
-1 \\
\vdots \\
\beta_i \\
\vdots
\end{pmatrix}
= \sum_{i=1}^{p-1} (-1)^i \beta_i \triangleq s_\beta
$$

where $1 - s_\beta = \Theta_\beta(-1) \neq 0$. It is interesting to note that, for $\lambda_1 = \pm 1$, we have in general terms

$$1 - u_n \xrightarrow{\ p\ } \Theta_\beta(\lambda_1)$$

where $\Theta_\beta(z) = 1 - \beta_1 z - \ldots - \beta_{p-1} z^{p-1}$ has no zero on the unit circle, by stability: the sign of the asymptotic behavior of Z_n under $\mathcal{H}_1$ thus depends on the sign of $\Theta_\beta(\lambda_1) \neq 0$. Finally, by considering the $\mathcal{C}^1$-mappings from $]-1, 0[$ to $\mathbb{R}$ given by

$$\forall\, n > 1, \quad g_n(x) = \frac{\ln c - \ln(1+x)}{\ln n},$$

one can see that $\alpha = g_n(-\rho_n(\alpha))$ and $\widehat{\alpha}_n(\alpha) = g_n(\widehat{v}_n(\alpha))$ as defined in (2.14). Theorems 2.2 and 2.3 easily follow.

References

M. Badreau and F. Proïa. Consistency and asymptotic normality in a class of nearly unstable processes. *Stat. Inf. Stoch. Proc.*, 26:619–641, 2023.

M. Barczy, M. Ispány, and G. Pap. Asymptotic behavior of unstable INAR(p) processes. *Stoch. Proc. Appl.*, 121:563–608, 2011.

W. Barreto-Souza and N. H. Chan. Nearly unstable integer-valued ARCH process and unit root testing. *Scandin. J. Stat.*, 51:402–424, 2024.

A. Bhargava. On the theory of testing for unit roots in observed time series. *Rev. Econ. Stud.*, 53:369–384, 1986.

P. J. Brockwell and R. A. Davis. *Time series: Theory and Methods (Second Edition)*. Springer Series in Statistics. Springer, New York, 1991.

B. Buchmann and N. H. Chan. Unified asymptotic theory for nearly unstable AR(p) processes. *Stoch. Proc. Appl.*, 123:952–985, 2013.

N. H. Chan and C. Z. Wei. Asymptotic inference for nearly nonstationary AR(1) processes. *Ann. Stat.*, 15:1050–1063, 1987.

N. H. Chan and C. Z. Wei. Limiting distributions of least squares estimates of unstable autoregressive processes. *Ann. Stat.*, 16:367–401, 1988.

D. N. De Jong, J. C. Nankervis, N. E. Savin, and C. H. Whiteman. Integration versus trend stationarity in time series. *Econometrica.*, 60:423–433, 1992.

D. A. Dickey and W. A. Fuller. Distribution of the estimators for autoregressive time series with a unit root. *J. Am. Stat. Assoc.*, 74-366:427–431, 1979.

M. Duflo. *Random iterative models*. Applications of Mathematics (vol. 34), New York. Springer-Verlag, Berlin, 1997.

L. Giraitis and P. C. B. Phillips. Uniform limit theory for stationary autoregression. *J. Time Ser. Anal.*, 27:51–60, 2006.

R. A. Horn and C. R. Johnson. *Matrix Analysis (Second Edition)*. Cambridge University Press, Cambridge, New-York, 2012.

M. Ispány, G. Pap, and M. C. A. Van Zuijlen. Asymptotic inference for nearly unstable INAR(1) models. *J. Appl. Probab.*, 40:750–765, 2003.

M. Ispány, G. Pap, and M. C. A. Van Zuijlen. Fluctuation limit of branching processses with immigration and estimation of the means. *Adv. Appl. Prob.*, 37:523–538, 2005.

D. Kwiatkowski, P. C. B. Phillips, P. Schmidt, and Y. Shin. Testing the null hypothesis of stationarity against the alternative of a unit root : How sure are we that economic time series have a unit root? *J. Econometrics.*, 54:159–178, 1992.

T. L. Lai and C. Z. Wei. Asymptotic properties of general autoregressive models and strong consistency of least-squares estimates of their parameters. *J. Multivariate Anal.*, 13:1–23, 1983.

S. J. Leybourne and B. P. M. McCabe. A consistent test for a unit root. *J. Bus. Econ. Stat.*, 12:157–166, 1994.

J. G. MacKinnon. *Critical values for cointegration tests*. Long-Run Economic Relationships, ed. by R. F. Engle, and C. W. Granger. Oxford University Press, Oxford, 1991.

C. R. Nelson and C. I. Plosser. Trends and random walks in macroeconomic time series: Some evidence and implications. *J. Monet. Econ.*, 10:139–162, 1982.

S. Ng and P. Perron. Unit root tests in ARMA models with data-dependent methods for the selection of the truncation lag. *J. Am. Stat. Assoc.*, 90:268–281, 1995.

J. W. Park. *Weak unit roots.* Department of Economics, Rice University, 2003.

P. Perron. Trends and random walks in macroeconomic time series: Further evidence from a new approach. *J. Econ. Dyn. Control.*, 12:297–332, 1988.

P. C. B. Phillips. Towards a unified asymptotic theory for autoregression. *Biometrika.*, 74: 535–547, 1987a.

P. C. B. Phillips. Time series regression with a unit root. *Econometrica.*, 55:277–302, 1987b.

P. C. B. Phillips and T. Magdalinos. Limit theory for moderate deviations from a unit root. *J. Econometrics.*, 136:115–130, 2007.

P. C. B. Phillips and P. Perron. Testing for a unit root in time series regression. *Biometrika.*, 75:335–346, 1988.

D. Pollard. *Convergence of Stochastic Processes.* Springer, Berlin, 1984.

F. Proïa. Moderate deviations in a class of stable but nearly unstable processes. *J. Stat. Plan. Inference.*, 208:66–81, 2020.

E. S. Said and D. A. Dickey. Testing $\text{ARIMA}(p,1,q)$ versus $\text{ARMA}(p+1,q)$. *Proc. Bus. Econ. Statist. Sect., Am. Statist. Assoc.*, pages 318–322, 1981.

E. S. Said and D. A. Dickey. Testing for unit roots in autoregressive moving average models of unknown order. *Biometrika.*, 71-3:599–607, 1984.

P. Schmidt and P. Phillips. LM test for a unit root in the presence of deterministic trends. *Oxford B. Econ. Stat.*, 54:257–287, 1992.

G. Schwert. Tests for unit roots: a Monte Carlo investigation. *J. Bus. Econ. Stat.*, 7:147–160, 1989.

T. Van der Meer, G. Pap, and M. C. A. Van Zuijlen. Asymptotic inference for nearly unstable $\text{AR}(p)$ processes. *Econ. Theory.*, 15:184–217, 1999.

J. S. White. The limiting distribution of the serial correlation coefficient in the explosive case. *Ann. Math. Statist.*, 29:1188–1197, 1958.

Laboratoire Manceau de Mathématiques, Le Mans Université, Avenue Olivier Messiaen, 72085 LE MANS Cedex 09, France.
Email address: marie.badreau.etu@univ-lemans.fr

Univ Angers, CNRS, LAREMA, SFR MATHSTIC, F-49000 Angers, France.
Email address: frederic.proia@univ-angers.fr