THE ASYMPTOTIC DISTRIBUTION OF UNIT ROOT TESTS OF UNSTABLE AUTOREGRESSIVE PROCESSES

By Bent Nielsen[1]
12 July 1999

1. introduction

Unit root testing has been developed through numerous papers since the work of Dickey and Fuller (1979). The idea is to test the hypothesis that the differences of an observed time series do not depend on its levels, or in other words, the levels of the time series has a unit root which can be removed by differencing. While it is in general possible to have multiple unit roots only the hypothesis of exactly one unit root is considered here. The available tests therefore hinge on two assumptions: (i) the levels of the time series has exactly one unit root which can be removed by differencing, and (ii) the remaining characteristic roots of the time series are stationary roots. In this paper it is proved that for the likelihood ratio test and a number of other likelihood based statistics the assumption (ii) is redundant whereas (i) is necessary. It is also shown that for some tests which are not likelihood based it is indeed necessary to assume that the differences have stationary roots.

The consequences of the result are perhaps best understood from the implications of condition (i). For autoregressive models of order two or higher that condition is not satisfied in the entire parameter space and the asymptotic distribution of the likelihood ratio test for a unit root depends on unknown nuisance parameters. In this situation the test statistic is not pivotal and the test is not similar and this complicates the testing. For non-likelihood based tests the necessity of condition (ii) implies an additional similarity problem. The practitioner is therefore faced with a trade off between likelihood based tests with fewer similarity problems and other tests which may have other advantageous properties. There are thus two empirical implications of the result. First, when analysing time series with stationary roots which have modulus close to one so that condition (ii) is nearly violated then the likelihood based tests are preferable and other tests should be used cautiously. Secondly, if explosive roots are found in an application

[1] Comments from K. Abadir, N. Shephard and the referees are gratefully acknowledged.

most of the statistical analysis is actually valid and should not necessarily be disregarded because of the presence of explosive roots.

Section 2 presents a Gaussian autoregressive model along with its statistical analysis and the result showing that condition (ii) is redundant for likelihood based tests. Robustness with respect to innovations which are martingale difference is also discussed. The results of Section 2 are given for a model without deterministic trends. In Section 3 these are generalised to models with deterministic terms. The mathematical proofs following in two appendices are based on the work of Lai and Wei (1983) and Chan and Wei (1988).

2. A MODEL WITHOUT DETERMINISTIC COMPONENTS

Consider the statistical model given by the autoregressive equation,

$$\Delta X_t = \alpha X_{t-1} + \sum_{j=1}^{p} \beta_j \Delta X_{t-j} + \varepsilon_t, \qquad (t = 1, \ldots, T), \tag{2.1}$$

conditional on X_0 and $\Delta X_0, \ldots, \Delta X_{1-p}$. The innovations are independently identically Gaussian distributed with zero mean and variance σ^2 and the parameters $\alpha, \beta_1, \ldots, \beta_p, \sigma^2$ vary freely. The unit root hypothesis is given by $\alpha = 0$.

The likelihood is maximised in two steps. First, ΔX_t and X_{t-1} are corrected for the remaining terms of equation (2.1) by least squares regression which gives the residuals $R_{0,t}$ and $R_{1,t}$, respectively:

$$(R_{0,t}, R_{1,t}) = (\Delta X_t, X_{t-1} | \Delta X_{t-1}, \ldots, \Delta X_{t-p}). \tag{2.2}$$

The likelihood ratio test statistic for the hypothesis is then computed as $LR = -T \log(1 - \hat{\lambda}^2)$ where $\hat{\lambda}$ is the sample correlation of $R_{0,t}$ and $R_{1,t}$, given by $\sum_{t=1}^{T} R_{0,t} R_{1,t} / (\sum_{t=1}^{T} R_{0,t}^2 \sum_{t=1}^{T} R_{1,t}^2)^{1/2}$. When the alternative is one-sided it is preferable to apply the signed version of the likelihood ratio test statistic which is given by $w = \text{sign}(\hat{\alpha})(LR)^{1/2}$ where $\hat{\alpha}$ is the maximum likelihood estimator for α given by $\sum_{t=1}^{T} R_{0,t} R_{1,t} / \sum_{t=1}^{T} R_{1,t}^2$.

The characteristic polynomial of the process X is important for the distributional analysis of the test statistic. Under the hypothesis this is given by $\psi(z) = (1 - z)(1 - \sum_{j=1}^{p} \beta_j z^j)$. Correspondingly, the characteristic polynomial for the differenced process, ΔX, is $\psi(z)/(1 - z)$. Usually two conditions are associated with the latter polynomial:

(i) The differenced process ΔX has no unit roots, or equivalently,

$$\sum_{j=1}^{p} \beta_j \neq 1. \tag{2.3}$$

If this is satisfied define $\delta = 1/(1 - \sum_{j=1}^{p} \beta_j)$.

(ii) The process ΔX can be given a stationary initial distribution, or equivalently, the characteristic polynomial for ΔX has its roots outside the complex unit circle.

In the following it is shown that the second condition is redundant for the distributional analysis of the test statistic.

THEOREM 1. *Suppose the model is given by* (2.1) *and that the hypothesis* $\alpha = 0$ *and the condition* (2.3) *are satisfied. Then, as* $T \to \infty$ *and with* W *being a standard Brownian Motion,*

$$w = \operatorname{sign}(\hat{\alpha})\sqrt{LR} \overset{\mathcal{D}}{\to} \frac{\int_0^1 W_u dW_u}{\left(\int_0^1 W_u^2 du\right)^{1/2}}. \tag{2.4}$$

The convergence result fails if the condition (2.3) *is not satisfied. The asymptotic density and distribution functions are given by Abadir (1995).*

The necessity of the assumption (2.3) was for instance encountered by Pantula (1989). Consequently, the test is not even asymptotically similar. This lack of similarity seems closely related to the problem that the asymptotic distribution can give poor approximations in finite samples, see Nielsen (1998) who also proved Theorem 1 for a second order model, $p = 1$.

Results like Theorem 1 have been discussed extensively in the literature under the additional assumption that ΔX only has stationary roots. White (1958) essentially proved the result for a first order model, $p = 0$. For the purely non-explosive case the result can be derived from Chan and Wei (1988). This is for instance done by Ghysels, Lee and Noh (1994) in connection with seasonal time series. They considered a model of order four or higher with one characteristic root at each of the quarterly frequencies, $1, i, -i, -1$, and assumed that the remaining roots are stationary.

The result of the Theorem 1 is shared by a number of test statistics. Examples are the Wald statistic $W = T\hat{\lambda}^2/(1-\hat{\lambda}^2)$ and the Lagrange multiplier statistic $LM = T\hat{\lambda}^2$ discussed by Evans and Savin (1981), as well as the t-type statistic $\hat{\tau} = \{(T-1)\hat{\lambda}^2/(1-\hat{\lambda}^2)\}^{1/2}$ suggested by Dickey and Fuller (1979) which is in one-one relation with w. Using the results of Appendix A it also follows that the maximum likelihood estimator for α is consistent and that $T\delta\hat{\alpha}$ converges in distribution to a variable which resembles that given in (2.4).

Other statistics, such as the first order estimator for α which is given by $\sum X_{t-1}\Delta X_t / \sum X_{t-1}^2$ is consistent for any lag length whenever ΔX has stationary roots, see Phillips (1987). That estimator is computed without regressing on the lagged differences and the non-stationary components of the time series are therefore not eliminated. Consequently, the estimator is not consistent in the generality described in Theorem 1. One of the simplest counter examples is when ΔX is a first order process with a root equal to minus one. This problem also applies to the t-type statistics constructed in the same way and is discussed in further detail by Perron (1996).

Robustness of the result with respect to innovations which are not independently, identically normal distributed has also been discussed extensively in the literature. For instance, Phillips (1987) discussed testing when the innovations are strongly mixing and Chan and Wei (1988) considered the case of innovations which are martingale differences. Here it will be proved that the result of Theorem 1 is robust in the case of martingale difference innovations.

ASSUMPTION. *Let $\{\varepsilon_t\}$ be a martingale difference and let $\mathcal{F}_t$ be the σ-field generated by the innovations, so that $E(\varepsilon_t|\mathcal{F}_{t-1}) = 0$ and $E(\varepsilon_t^2|\mathcal{F}_{t-1}) = \sigma^2$. Further, assume that the innovations have bounded moments of order $2+\gamma$ for some $\gamma > 0$, that is, with probability one, $\sup_t E(|\varepsilon_t|^{2+\gamma}|\mathcal{F}_{t-1}) < \infty$.*

THEOREM 2. *Suppose the process is given by (2.1) where $\alpha = 0$, the condition (2.3) is met and the innovations satisfy the martingale difference assumption. Then the statistic w converges in distribution as described in Theorem 1.*

The maximum likelihood estimators for the remaining parameters are consistent. This consistency is also robust with respect to innovations which satisfy martingale difference assumptions. For β_j this was discussed by Lai and Wei (1983) and for σ^2 this follows from the equation (4.6) below. Further, Chan and Wei (1988) discussed the asymptotic distribution of β_j in the purely non-explosive case. A recent account on the purely explosive case is given in Monsour and Mikulski (1998). Finally, asymptotic normality of the estimator for σ^2 can be proven under moment assumptions on the martingale difference sequence $\varepsilon_t^2 - 1$.

3. EXTENSIONS TO MODELS WITH DETERMINISTIC TRENDS

In most applications it would be convenient to include deterministic components in the model. There is a great variety of such augmented unit root tests in the literature. Two cases are considered, a model with a constant level and a model with a linear trend.

The model with a constant level is given by the autoregressive equation

$$\Delta X_t = (\alpha, \alpha_1) \begin{pmatrix} X_{t-1} \\ 1 \end{pmatrix} + \sum_{j=1}^{p} \beta_j \Delta X_{t-j} + \varepsilon_t. \tag{3.1}$$

In this model the unit root hypothesis can be formulated as either $\alpha = 0$ or $\alpha = \alpha_1 = 0$, see Dickey and Fuller (1979, 1981). For simplicity only the latter hypothesis is considered. The reason is two-fold. First, the latter hypothesis only questions the behaviour of the stochastic component of time series and not that of the deterministic component. Secondly, the asymptotic distribution of the likelihood ratio test statistics does not depend on the parameter related to the deterministic component.

The case of linear trend is correspondingly given by the equation

$$\Delta X_t = (\alpha, \alpha_1) \begin{pmatrix} X_{t-1} \\ t \end{pmatrix} + \mu + \sum_{j=1}^{p} \beta_j \Delta X_{t-j} + \varepsilon_t, \tag{3.2}$$

where the hypothesis of interest is given by $\alpha = \alpha_1 = 0$.

In both cases the statistical analysis is similar to that of the model without deterministic components. For notational convenience define X_t^* as the vectors $(X_{t-1}, 1)'$ and $(X_{t-1}, t)'$, respectively. The likelihood is then maximised by first correcting ΔX_t and X_{t-1}^* for the remaining components of the relevant model and then finding the sample correlation $\hat{\lambda}$, say, of the residuals. In both cases the levels of the process is corrected for a constant either through a linear transformation of X_t^* or by the initial regression on the remaining components of the model. This shows that X_t^* could equivalently be chosen as $(\sum_{s=1}^{t-1}\Delta X_s, 1)'$ or $(\sum_{s=1}^{t-1}\Delta X_s, t)'$ and hence the null distribution of the sample multiple correlation does not depend on the initial level X_0. The following result therefore applies.

THEOREM 3: *Suppose the model is given by either* (3.1) *or* (3.2) *and that the hypothesis* $\alpha = \alpha_1 = 0$ *is satisfied. Then the distribution of the likelihood ratio test statistic, LR, does not depend on* X_0. *If, in addition,* (2.3) *is satisfied, then, as* $T \to \infty$,

$$LR = -T\log\left(1-\hat{\lambda}^2\right) \xrightarrow{\mathcal{D}} \int_0^1 dW_u F_u' \left\{\int_0^1 F_u F_u' du\right\}^{-1} \int_0^1 F_u dW_u,$$

where F_u *is a two dimensional process given by* $(W_u, 1)$ *and* $(W_u - \int_0^1 W_v dv, u - 1/2)$, *respectively.*

The result is robust with respect to innovations which satisfy the martingale difference assumption.

This result is for instance proved by Johansen (1995, Theorem 6.1) under the additional assumption that ΔX has stationary roots. Corresponding results hold for other likelihood-based tests suggested in the literature. For the constant levels model (3.1) an example is the F-type statistic, $\Phi_1 = (T/2-1)\hat{\lambda}^2/(1-\hat{\lambda}^2)$, suggested by Dickey and Fuller (1981). When it comes to testing the hypothesis $\alpha = 0$ rather than $\alpha = \alpha_1 = 0$ the asymptotic distribution of the likelihood based tests depends on the value of α_1. Nonetheless, asymptotic results corresponding to that of Theorem 3 can be proven for tests such as the t-type statistic, $\hat{\tau}_\mu, \hat{\tau}_\tau$ suggested by Dickey and Fuller (1979). That proof would require a modification of the Lemma B1 in the Appendix due to the nuisance parameter α_1, see also Chan (1989).

Nuffield College, Oxford OX1 1NF, UK

APPENDIX A: Proofs of Theorems 1 and 2

Theorem 1 follows from Theorem 2. The proof of Theorem 2 has two parts. First, Lemma A1 describes the residuals in detail. Next, the asymptotic analysis follows in Lemmae A2-A4.

Lemma A1: *Suppose equation* (2.1)*, the hypothesis* $\alpha = 0$ *and the condition* (2.3) *are satisfied. Define*

$$\delta = \left(1 - \sum_{j=1}^{p} \beta_j\right)^{-1}, \qquad \varepsilon_0 = X_0/\delta + \sum_{j=0}^{p-1} \Delta X_{-j} \sum_{k=j+1}^{p} \beta_k, \qquad S_t = \sum_{j=0}^{t} \varepsilon_j.$$

Then, the residuals (2.2) *can be written as*

$$(R_{0,t}, R_{1,t}) = (\varepsilon_t, \delta S_{t-1} | \Delta X_{t-1}, \ldots, \Delta X_{t-p}).$$

Further, for any p-dimensional process Z_{t-1} *found by a non-singular linear transformation of* $(\Delta X_{t-1}, \ldots, \Delta X_{t-p})$

$$(R_{0,t}, R_{1,t}) = (\varepsilon_t, \delta S_{t-1} | Z_{t-1}). \tag{4.1}$$

Proof: Under the hypothesis the model equation (2.1) is given by $\Delta X_t = \sum_{j=1}^{p} \beta_j \Delta X_{t-j} + \varepsilon_t$, and the expression for $R_{0,t}$ follows immediately. This equation can be rewritten as $(1-\sum_{j=1}^{p} \beta_j)\Delta X_t = \varepsilon_t - \sum_{j=0}^{p-1} \Delta^2 X_{t-j} \sum_{k=j+1}^{p} \beta_k$. Under the assumption (2.3) the parameter δ is well-defined and cumulation of the latter equation gives $X_{t-1} = \delta S_{t-1} - \delta \sum_{j=1}^{p} \Delta X_{t-j} \sum_{k=j+1}^{p} \beta_k$. The expression for $R_{1,t}$ then follows. □

The sample correlation, $\hat{\lambda}$, is invariant with respect to scaling of $R_{0,t}$ or $R_{1,t}$ by σ and $\sigma\delta$ respectively. Thus for asymptotic purposes it suffices to assume $\sigma^2 = \delta = 1$.

Lemma A2. *Suppose* $\{\varepsilon_t\}$ *satisfy the martingale difference assumption and* $\sigma^2 = \delta = 1$*. Then*

$$\frac{1}{T} \sum_{t=1}^{T} \varepsilon_t^2 \xrightarrow{\mathcal{P}} 1, \tag{4.2}$$

and

$$\left(T^{-1/2}\sum_{t=1}^{T}\varepsilon_t, T^{-1}\sum_{t=1}^{T}S_{t-1}\varepsilon_t, T^{-3/2}\sum_{t=1}^{T}S_{t-1}, T^{-2}\sum_{t=1}^{T}S_{t-1}^2\right) \xrightarrow{\mathcal{D}} \left(W_1, \int_0^1 W_u dW_u, \int_0^1 W_u du, \int_0^1 W_u^2 du\right) \quad (4.3)$$

PROOF: The convergence (4.2) follows from Chan and Wei (1988, equation 2.13), whereas (4.3) follows from the Functional Central Limit Theorem for martingale difference sequences, see Chan and Wei (1988, Theorem 2.2), combined with the Continuous Mapping Theorem, see Billingsley (1968). □

The next lemmae show that for asymptotic purposes the residuals R_0 and R_1 can be replaced by ε_t and δS_{t-1}, respectively. The main idea is to choose the regressor Z conveniently. Follow Lai and Wei (1983, equation 4.2) and Chan and Wei (1988, equation 3.2) and decompose $\Delta X_{t-1}, \ldots, \Delta X_{t-p}$ into processes $A_t, B_{j,t}, C_t, D_t$ with characteristic roots at one, at $\exp(i\theta_j)$ and $\exp(-i\theta_j)$ but $\exp(i\theta_j) \neq 1$, outside the unit circle, and inside the unit circle, respectively. Further, the processes $A_t, B_{j,t}, C_t, D_t$ can be normalised so that the normalised process $Z_{t-1} = (a_t, b_{1,t}, \ldots, b_{l,t}, c_t, d_t)$, say, satisfies

$$\sum_{t=1}^{T} Z_{t-1}Z_{t-1}' \xrightarrow{\mathcal{D}} F, \quad (4.4)$$

where F is a symmetric block-diagonal random matrix which is positive definite with probability one. The Lemmae A3-A5 show that expressions of the type $T^y(\sum_{t=1}^T Z_{t-1}Z_{t-1}')^{-1/2}\sum_{t=1}^T Z_{t-1}Y_t$ converges to zero in probability. Due to (4.4) it suffices to prove that $T^y\sum_{t=1}^T Z_{t-1}Y_t$ converges to zero whenever Z_t is given by either of its components $a_t, b_{j,t}, c_t, d_t$.

LEMMA A3. *Suppose* $\{\varepsilon_t\}$ *satisfies the martingale difference assumption and that equation* (2.1) *and the hypothesis* $\alpha = 0$ *are satisfied. Then, for all* $\xi < \gamma/(2+\gamma)$,

$$T^{(\xi-1)/2}\left(\sum_{t=1}^{T} Z_{t-1}Z_{t-1}'\right)^{-1/2}\sum_{t=1}^{T} Z_{t-1}\varepsilon_t \xrightarrow{\mathcal{P}} 0.$$

Note, that the condition (2.3) *is not necessary for this result.*

PROOF: *For Z non-explosive*: see Chan and Wei (1988, Theorem 3.5.1).

For Z explosive: Lai and Wei (1983, equation 4.18) essentially prove the result for $\xi \leq 0$. For the general case note that $|| \sum_{t=1}^{T} d_{t-1}\varepsilon_t|| \leq \max_{t\leq T} ||\varepsilon_t|| \sum_{t=1}^{T} ||d_{t-1}||$. Using Lai and Wei (1983, Corollary 1) it follows that $\lim_{T\to\infty} \sum_{t=1}^{T} \|d_t\|$ is finite with probability one. Thus it suffices to show that ε_t is of order smaller than $T^{(1-\xi)/2}$ with probability one. Using an idea of Lai and Wei (1982) this follows from the conditional Borel-Cantelli Lemma, see Freedman (1973). By Chebychev's inequality

$$\sum_{t=1}^{\infty} P\left(|\varepsilon_t| > T^{(1-\xi)/2} \middle| \mathcal{F}_{t-1}\right) \leq \sum_{t=1}^{\infty} T^{-(1-\xi)(1+\gamma/2)} E\left(|\varepsilon_t|^{2+\gamma} \middle| \mathcal{F}_{t-1}\right),$$

where the latter series is convergent for $(1-\xi)(1+\gamma/2) > 1$ and consequently ε_t is of the postulated order. □

LEMMA A4. *Suppose* $\{\varepsilon_t\}$ *satisfies the martingale difference assumption and that equation* (2.1), *the hypothesis* $\alpha = 0$ *and the condition* (2.3) *are satisfied. Then, for all* $\eta > 0$,

$$T^{-(1+\eta)/2} \left(\sum_{t=1}^{T} Z_{t-1}Z'_{t-1}\right)^{-1/2} \sum_{t=1}^{T} Z_{t-1}S_{t-1} \xrightarrow{\mathcal{P}} 0.$$

PROOF: *For Z having unit roots only, but no roots at one*: For $\eta \geq 1$ the result follows from Chan and Wei (1988, Theorem 3.4.1). Their arguments can be sharpened. A typical element of the proof is the following. Suppose the process has one negative unit root. The product sum of interest is then

$$\sum_{t=1}^{T} b_{1,t-1}S_{t-1} = \frac{1}{T}\sum_{t=1}^{T} (-1)^{t-1} \sum_{j=1}^{t-1} (-1)^j \varepsilon_j \sum_{k=1}^{t-1} \varepsilon_k. \tag{4.5}$$

Define $X_n = \sum_{j=1}^{n} (-1)^j \varepsilon_j \sum_{k=1}^{n} \varepsilon_k$ and note that, for $n \geq m$

$$|X_n - X_m| \leq \left|\sum_{j=1}^{n} (-1)^j \varepsilon_j\right| \left|\sum_{k=m+1}^{n} \varepsilon_k\right| + \left|\sum_{j=m+1}^{n} (-1)^j \varepsilon_j\right| \left|\sum_{k=1}^{m} \varepsilon_k\right|.$$

The order of magnitude of the sum (4.5) follows from Theorem 2.1 of Chan and Wei (1988) saying: if $\{X_n\}$ is a sequence of random variables so (i) $E|X_n| = (n^\alpha)$ for some $\alpha > 0$, (ii) $A_j(n,m)$, $B_j(n,m)$ are random variables so $EA_j^2(n,m) = \mathrm{O}(n^{\varphi_j})$, $EB_j^2(n,m) = \mathrm{O}\{n^{\psi_j}(n-m)\}$ and $|X_n - X_m| \leq \sum_{j=1}^q A_j(n,m)B_j(n,m)$ for $\varphi_j, \psi_j \geq 0$, $n > m$, (iii) $\exp(i\theta) \neq 1$, (iv) $2\alpha > \varphi_j + \psi_j$ then $\sup_{1\leq j\leq n} |\sum_{t=1}^j \exp(it\theta)X_t| = \mathrm{o}_{\mathcal{P}}\left(n^{\alpha+1}\right)$. With the same proof and (iv) replaced by (iv') $2\alpha = \varphi_j + \psi_j + 1$ it actually follows that $\sup|\sum \exp(it\theta) X| = \mathrm{o}_{\mathcal{P}}\left(n^{\alpha+(1+\eta)/2}\right)$ for all $\eta > 0$. Now, choose X_n as above, $\alpha = 1$, and for instance $A_1(n,m) = \sum_{j=1}^n (-1)^j \varepsilon_j$, $B_1(n,m) = \sum_{k=m+1}^n \varepsilon_k$ with $\varphi_1 = 1, \psi_1 = 0$.

For Z stationary: For $\eta \geq 1$ the result follows from Chan and Wei (1988, Lemma 3.4.3). Their arguments can be sharpened. The Lemma says: if (i) z_t is a stationary autoregressive process, (ii) the process g_t satisfies $g_t = Mg_{t-1} + h_t$, (iii) $E\sum_{t=1}^n \|g_t\|^2 = \mathrm{O}(n^\alpha)$ for some $\alpha > 0$, (iv) $E\sum_{t=1}^n \|h_t\|^2 = \mathrm{o}(n^\alpha)$ then $E\|\sum_{t=1}^n g_t z_t'\| = \mathrm{o}(n^{(\alpha+1)/2})$. With the same proof and (iv) replaced by (iv') $E\sum_{t=1}^n \|h_t\|^2 = \mathrm{O}(n^{\alpha-1})$ it actually follows that $E\|\sum_{t=1}^n g_t z_t'\| = \mathrm{o}(n^{(\alpha+\eta)/2})$ for all $\eta > 0$. Now, let $h_t = \varepsilon_t$, $M = 1$ and hence $g_t = S_t$, let z_t be the non-normalised process C_t and choose $\alpha = 2$. Then it follows that $E||\sum_{t=1}^T C_{t-1}S_{t-1}|| = \mathrm{o}(T^{1+\eta/2})$ and $T^{-1-\eta/2}\sum_{t=1}^T C_{t-1}S_{t-1} \xrightarrow{\mathcal{P}} 0$. Since $\sum_{t=1}^T C_{t-1}C_{t-1}'/T$ converges in probability the desired result follows.

For Z explosive: As in the proof of Lemma A3 $||\sum_{t=1}^T d_{t-1}S_{t-1}||$ is of the same stochastic order as $\max_{t\leq T}||S_t||$. According to Chan and Wei (1988, Theorem 2.2) the process $T^{-1/2}S$ converges weakly to a Brownian motion on $D[0,1]$, the space of functions on [0,1] which are right continuous, have left hand limits, and is equipped with the Skorokhod topology. The supremum is a continuous mapping on $D[0,1]$ and, hence, by the Continuous Mapping Theorem, see Billingsley (1968), $\max\|T^{-1/2}S\|$ converges in distribution. □

Proof of Theorem 2: In Lemma A3 choose η, ξ such that $0 < \eta < \xi < \gamma/(2+\gamma) \leq 1$. Then by the expression (4.1)

$$\begin{aligned} \frac{1}{T}\sum_{t=1}^T R_{0,t}^2 &= \frac{1}{T}\sum_{t=1}^T \varepsilon_t^2 + \mathrm{o}_{\mathcal{P}}\left(T^{-\xi}\right), \qquad (4.6) \\ \frac{1}{T}\sum_{t=1}^T R_{1,t}R_{0,t} &= \frac{1}{T}\sum_{t=1}^T S_{t-1}\varepsilon_t + \mathrm{o}_{\mathcal{P}}(T^{(\eta-\xi)/2}), \end{aligned}$$

$$\frac{1}{T^2}\sum_{t=1}^{T} R_{1,t}^2 \;=\; \frac{1}{T^2}\sum_{t=1}^{T} S_{t-1}^2 + o_{\mathcal{P}}\left(T^{(\eta-1)}\right).$$

Note, that $\eta - \xi < 0$ and $\eta - 1 < 0$. Combine this with Lemma A2. □

APPENDIX B: Proof of Theorem 3

When analysing the asymptotic null distribution of the considered tests it suffices to consider the same probability measures as in Appendix A.1. For the constant level model (3.1) this is readily seen. For the linear trend model (3.2) the additional parameter μ turns out not to be important as long as the condition (2.3) is satisfied. First, by mimicking the proof of Lemma A1,

$$X_{t-1} = \delta S_{t-1} + \delta\mu t - \delta\sum_{j=1}^{p}\Delta X_{t-j}\sum_{k=j+1}^{p}\beta_k + \mathit{constant},$$

and the dependency of linear trend of X_{t-1} can be removed by linear transformation of X^*_{t-1}. Secondly, the regressors ΔX_{t-j} can be replaced by $\Delta X_{t-j}+\delta\mu$ which satisfies the equation for the model without deterministic terms (2.1) although the initial conditions are altered. Finally, the asymptotic result does not depend on the initial values and hence μ can be ignored.

The proof of Theorem 3 therefore follows by combining the arguments of Appendix A.1 with the following lemma.

Lemma B1. *Suppose $\{\varepsilon_t\}$ satisfy the martingale difference assumption and that equation (2.1), the hypothesis $\alpha = 0$ and the condition (2.3) are satisfied. Then, for all $\eta > 0$*

$$T^{-\eta}\left(\sum_{t=1}^{T} Z_{t-1}Z_{t-1}'\right)^{-1/2}\sum_{t=1}^{T} Z_{t-1} \xrightarrow{\mathcal{P}} 0, \qquad (4.7)$$

$$T^{-(1+\eta)}\left(\sum_{t=1}^{T} Z_{t-1}Z_{t-1}'\right)^{-1/2}\sum_{t=1}^{T} Z_{t-1}t \xrightarrow{\mathcal{P}} 0. \qquad (4.8)$$

If Z only has roots of length one and none equal to one then the result holds for $\eta > -1/2$.

Proof. *For Z having unit roots only, but no roots at one*: First, suppose the component b_1 has one root at -1 and $\varepsilon_0 = X_0$. By change of summation order

$$\begin{aligned}
\sum_{t=1}^{T} b_{1,t-1} &= T^{-1}\sum_{t=1}^{T}\sum_{k=1}^{t}(-1)^{t-k}\varepsilon_{k-1} = T^{-1}\sum_{k=1}^{T}\varepsilon_{k-1}\sum_{t=k}^{T}(-1)^{t-k} \\
&= T^{-1}\sum_{k=0}^{[(T-1)/2]}\varepsilon_{T-2k-1}, \qquad (4.9)
\end{aligned}$$

which is of order $T^{-1/2}$, see Chan and Wei (1988, Theorem 2.2), and (4.7) follows for $\eta > -1/2$. Similarly, (4.8) follows using the additional normalisation by T^{-1}. If the root multiplicity at -1 is higher than one, the result follows in a fashion similar to the proof of Chan and Wei (1988, Theorem 3.2.1). For the case where b_j has non-real roots at $\exp(i\theta_j)$ and $\exp(-i\theta_j)$ the argument is basically the same, albeit the notation is more complicated. A result like (4.9) is established using trigonometric identities as $\sum_{k=1}^{T}\sin(k2\theta_j) = \sin(T\theta_j)\sin\{(T+1)\theta_j\}/\sin(\theta_j)$ and the sharpened version of Chan and Wei (1988, Theorem 2.1) mentioned in the proof of Lemma A4.

For Z stationary: The vector process c satisfies $c_t = Cc_t + \tilde{\varepsilon}_t$ where $\tilde{\varepsilon}_t = (\varepsilon_t, 0, \ldots, 0)'$. Hence it is a linear process with exponentially decreasing coefficients. Therefore (4.7) follows from the Central Limit Theorem for linear processes with martingale difference innovations, see Phillips and Solo (1992, Theorem 3.16). Correspondingly, (4.8) follows by partial summation and the Central Limit Theorem together with the Invariance Principle also given in the above mentioned Theorem by Phillips and Solo.

For Z explosive: The result follows as in the proof of Lemma A3 with ε_t replaced by 1 and t respectively. □

REFERENCES

ABADIR, K.M. (1995): "The Limiting Distribution of the t Ratio Under a Unit Root," *Econometric Theory,* 11, 775-93

BILLINGSLEY, P. (1968): *Convergence of Probability Measures.* New York: Wiley.

CHAN, N.H. (1989): "Asymptotic inference for unstable autoregressive time series with drifts", *Journal of Statistical Planning and Inference*, 23, 301-312.

CHAN, N.H., AND C.Z. WEI (1988): "Limiting Distributions of Least Squares Estimates of Unstable Autoregressive Processes," *Annals of Statistics,* 16, 367-401.

DICKEY, D.A., AND W.A. FULLER (1979): "Distribution of the Estimators for Autoregressive Time Series with a Unit Root," *Journal of the American Statistical Association,* 74, 427-31.

DICKEY, D.A., AND W.A. FULLER (1981): "Likelihood Ratio Statistics for Autoregressive Time Series with a Unit Root," *Econometrica,* 49, 1057-72.

EVANS, G.B.A., AND N.E. SAVIN (1981): "Testing for Unit Roots: 1," *Econometrica,* 49, 753-79.

FREEDMAN, D. (1973): "Another Note on the Borel-Cantelli Lemma and the Strong Law, with the Poisson Approximation as a By-Product," *Annals of Probability,* 1, 910-25.

GHYSELS, E., H.S. LEE, AND J. NOH (1994): "Testing for unit roots in seasonal time series," *Journal of Econometrics*, 62, 415-42.

JOHANSEN, S. (1995): *Likelihood-based inference in cointegrated vector autoregressive models.* Oxford University Press.

LAI, T.L., AND C.Z. WEI (1982): "Least Squares Estimates in Stochastic Regression Models with Applications to Identification and Control of Dynamic Systems," *Annals of Statistics,* 10, 154-66.

——— (1983): "Asymptotic Properties of General Autoregressive Models and Strong Consistency of Least-Squares Estimates of Their Parameters," *Journal of Multivariate Analysis,* 13, 1-23.

MONSOUR, M.J., AND P.W. MIKULSKI (1998): "On Limiting Distributions in Explosive Autoregressive Processes," *Statistics and Probability Letters,* 37, 141-47.

NIELSEN, B. (1998): "On the Distribution of Tests for Cointegration Rank," Mimeo, Nuffield College.

PANTULA, S.G. (1989): "Testing for Unit Roots in Time Series Data," *Econometric Theory,* 5, 256-71.

PERRON, P. (1996): "The adequacy of asymptotic approximations in the near-integrated autoregressive model with dependent errors," *Journal of Econometrics*, 70, 317-350.
PHILLIPS, P.C.B. (1987): "Time Series Regression with a Unit Root," *Econometrica,* 55, 277-301.
PHILLIPS, P.C.B., AND V. SOLO (1992): "Asymptotics for Linear Processes," *Annals of Statistics,* 20, 971-1001.
WHITE, J.S. (1958): "The Limiting Distribution of the Serial Correlation Coefficient in the Explosive Case," *Annals of Mathematical Statistics,* 29, 1188-97.