Asymptotic results for cointegration tests in non-stable cases

Bent Nielsen
Nuffield College, Oxford OX1 1NF, UK.
email: bent.nielsen@nuf.ox.ac.uk
http://www.nuff.ox.ac.uk/users/nielsen
14 May 1997

Abstract: Asymptotic analyses of unit root tests in autoregressive time series are usually based on the assumptions that the number of unit roots is known and that the remaining characteristic roots are stable. The last assumption seems not to be necessary. This is stated more precisely for two examples. One is a unit root test in a univariate second order model, the other a test for at most one cointegrating relation in a bivariate first order model.

1. Introduction

Asymptotic analysis of unit root tests is usually based on the assumptions that the number of unit root is given and that the remaining characteristic roots are stable. In many applications such assumptions may be quite reasonable and in particular if the estimation of a statistical models gives non-stable characteristic roots then the model would usually be reformulated. However, these assumptions imply a limitation in the theoretical interpretation of the test: If a different asymptotic distribution apply when the assumptions are not satisfied then the test is not similar. For some important examples it is shown that the number of unit roots is crucial for the asymptotic analysis whereas the assumption of stability of the remaining roots is redundant. These conclusions are useful for understanding small sample properties of unit root tests, see Nielsen (1997).

Various unit root test problems have been proposed and of these the following statistical model is considered. Let a p-dimensional time series X be described by a k-th order autoregressive equation,

$$\Delta X_t = \Pi X_{t-1} + \sum_{j=1}^{k-1} \Gamma_j \Delta X_{t-j} + \varepsilon_t, \tag{1}$$

for $t = 1, \dots, T$ and with fixed initial values, $X_0, \Delta X_0, \dots, \Delta X_{1-k}$ and independent normal distributed innovations with mean zero and variance Ω.

The parameters, $\Pi, \Gamma_1, \ldots \Gamma_{k-1}, \Omega$, vary freely. The hypothesis of interest is the hypothesis of at most r cointegrating relations, that is $rank\Pi \leq r$.

The likelihood criterion for the hypothesis is derived by Johansen (1988, 1995a) and given in terms of the $p-r$ smallest empirical canonical correlations of first differences and lagged levels of the time series where both are corrected for the remaining terms of the model.

For the distributional analysis consider the characteristic polynomial

$$A(z) = (1-z)I_p - z\Pi - \sum_{j=1}^{k-1}(1-z)z^j\Gamma_j.$$

Then under the assumptions

(A.1) $p-rank\Pi$ unit roots and $rank\Pi = r$,

(A.2) the remaining roots are numerically larger than one,

the distribution of the likelihood criterion can be approximated by

$$tr\left\{\int_0^1 dW_u W_u' \left(\int_0^1 W_u W_u' du\right)^{-1} \int_0^1 W_u dW_u'\right\} \tag{2}$$

where W is a Brownian motion of dimension $p-r$. It is conjectured that the assumption (A.2) is redundant so *for fixed parameters, then as* $T \to \infty$, *LR converges in distribution to the distribution given by* (2) *whenever only (*A.1) *is satisfied.*

The statistical analysis has two elements: a correction for additional lags and a canonical correlation analysis. The two simplest examples of this is the test for "no cointegration" in a univariate second order model, $p = 1$, $k = 2$, $r = 0$ and a test for at most one cointegrating relation in a bivariate first order model, $p = 2$, $k = 1$, $r = 1$. These examples will be analysed in detail. The proof for the case where (A.1) and (A.2) are satisfied is based on a well-established literature on weak convergence of random walks and stochastic integration. When (A.2) is not satisfied many of such results have to be invented. This is not trivial and thus only the two examples are analysed.

In Section 2 the univariate second order model is analysed. The bivariate second order model is analysed in Section 3. The proof is based on convergence results for the relevant product moment matrices. Thus in Section 4 a result on convergence of quadratic functions of a univariate explosive autoregressive time series with an intercept is proved. Next, in Section 5 it is shown that this convergence holds jointly with convergence of a random walk. Finally in Section 6 the necessary convergence results for product moment matrices are proved.

2. The univariate second order model

In this section the conjecture is proved for a univariate second order model: $p = 1$, $k = 2$.

The model (1) is reparametrised as

$$\Delta^2 X_t = \Pi X_{t-1} - \Gamma \Delta X_{t-1} + \varepsilon_t. \tag{3}$$

with fixed initial values, $X_0, \Delta X_0$. The hypothesis is that $\Pi = 0$. The characteristic polynomial has one unit root for $\Gamma \neq 0$ and two for $\Gamma = 0$. The likelihood criterion is given by $-T \log(1 - \hat{\lambda})$ where $\hat{\lambda}$ is given by

$$\frac{\left[\sum_{t=1}^T X_{t-1}\Delta^2 X_t - \frac{\sum_{t=1}^T X_{t-1}\Delta X_{t-1} \sum_{t=1}^T \Delta X_{t-1}\Delta^2 X_t}{\sum_{t=1}^T (\Delta X_{t-1})^2}\right]^2}{\left[\sum_{t=1}^T X_{t-1}^2 - \frac{\left(\sum_{t=1}^T X_{t-1}\Delta X_{t-1}\right)^2}{\sum_{t=1}^T (\Delta X_{t-1})^2}\right]\left[\sum_{t=1}^T (\Delta^2 X_t)^2 - \frac{\left(\sum_{t=1}^T \Delta X_{t-1}\Delta^2 X_t\right)^2}{\sum_{t=1}^T (\Delta X_{t-1})^2}\right]}$$

and it is asymptotically equivalent to $T\hat{\lambda}$. The asymptotic distribution of the likelihood criterion is given as

Theorem 2.1

For $\Gamma \neq 0$ then the likelihood criterion converges in distribution to (2) with one degree of freedom.

For $\Gamma = 0$ then the likelihood criterion converges in distribution to

$$\frac{\left(\int_0^1 \tilde{W}_u dW_u - \frac{\int_0^1 \tilde{W}_u W_u du \int_0^1 W_u dW_u}{\int_0^1 W_u^2 du}\right)^2}{\int_0^1 \tilde{W}_u^2 du - \frac{\left(\int_0^1 \tilde{W}_u W_u du\right)^2}{\int_0^1 W_u^2 du}} \tag{4}$$

where $\tilde{W}$ is the integral of the Brownian motion W.

The limit distribution (4) was previously found by Johansen (1995b) and only the case $\Gamma \neq 0$ has to be proved.

Proof. First, under the hypothesis the time series $\Delta^2 X_t$ corrected for ΔX_{t-1} can be rewritten as

$$\Delta^2 X_t - \Delta X_{t-1} \frac{\sum_{t=1}^T \Delta^2 X_t \Delta X_{t-1}}{\sum_{t=1}^T \left(\Delta X_{t-1}\right)^2} = \varepsilon_t - \Delta X_{t-1} \frac{\sum_{t=1}^T \Delta X_{t-1} \varepsilon_t}{\sum_{t=1}^T \left(\Delta X_{t-1}\right)^2}.$$

As for the levels of the time series, then summation of the equation (3) under the hypothesis gives

$$\Delta X_t - \Delta X_0 = -\Gamma \left(X_{t-1} - X_0 + \Delta X_0\right) + \sum_{j=1}^t \varepsilon_j.$$

Addition of the term ΓX_t and some rearrangements imply that for $\Gamma \neq 0$

$$X_t = \frac{1}{\Gamma} S_t - \frac{1-\Gamma}{\Gamma} \Delta X_t$$

with

$$S_t = \sum_{s=0}^t \varepsilon_t, \qquad \varepsilon_0 = \Gamma X_0 + (1-\Gamma) \Delta X_0.$$

Then X_{t-1} corrected for ΔX_{t-1} can be rewritten as

$$X_{t-1} - \Delta X_{t-1} \frac{\sum_{t=1}^T X_{t-1} \Delta X_{t-1}}{\sum_{t=1}^T \left(\Delta X_{t-1}\right)^2} = \frac{1}{\Gamma} \left(S_{t-1} - \Delta X_{t-1} \frac{\sum_{t=1}^T S_{t-1} \Delta X_{t-1}}{\sum_{t=1}^T \left(\Delta X_{t-1}\right)^2} \right).$$

As a consequence of the results above the empirical correlation, $\hat{\lambda}$, equals

$$\frac{\left[\sum_{t=1}^T S_{t-1} \varepsilon_t - \frac{\sum_{t=1}^T S_{t-1} \Delta X_{t-1} \sum_{t=1}^T \Delta X_{t-1} \varepsilon_t}{\sum_{t=1}^T (\Delta X_{t-1})^2}\right]^2}{\left[\sum_{t=1}^T S_{t-1}^2 - \frac{\left(\sum_{t=1}^T S_{t-1} \Delta X_{t-1}\right)^2}{\sum_{t=1}^T (\Delta X_{t-1})^2}\right] \left[\sum_{t=1}^T \varepsilon_t^2 - \frac{\left(\sum_{t=1}^T \Delta X_{t-1} \varepsilon_t\right)^2}{\sum_{t=1}^T (\Delta X_{t-1})^2}\right]}$$

It remains to conclude that this term equals

$$\frac{\sum_{t=2}^T S_{t-1} \varepsilon_t}{\sum_{t=1}^T \varepsilon_t^2 \sum_{t=2}^T S_{t-1}^2} + O_P\left(\frac{1}{T^2}\right)$$

in which case the Theorem follows. The initial term, ε_0, is asymptotically ignorable and thus it suffices that

$$\frac{\left(\sum_{t=1}^{T} \Delta X_{t-1}\varepsilon_t\right)^2}{\sum_{t=1}^{T} \left(\Delta X_{t-1}\right)^2}$$

converges in distribution, which follows from White (1958) for any Γ, and that

$$\frac{\left(\sum_{t=2}^{T} S_{t-1}\Delta X_{t-1}\right)^2}{T^2 \sum_{t=1}^{T} \left(\Delta X_{t-1}\right)^2} \tag{5}$$

converges in probability to zero. Now, (5) follows if

$$\left(\frac{1}{Tg}\sum_{t=2}^{T} S_{t-1}\Delta X_{t-1}, \frac{1}{g^2}\sum_{t=1}^{T} \left(\Delta X_{t-1}\right)^2\right) \xrightarrow{\mathcal{D}} (0, V) \tag{6}$$

where the random variable $V \neq 0$ almost surely and g is some normalisation constant. In case $|1-\Gamma| < 1$ then $\Delta X_t, \Delta S_t$ are stationary linear processes and (6) follows from Johansen (1995a, Theorem B.13) with $g = \sqrt{T}$. For the case $\Gamma = 2$ Chan and Wei (1988, Th. 2.2) show that S_t and ΔX_{t-1}, normalised by $\sqrt{T}$, converges to independent Brownian motions and the Theorems 2.4 and 3.4.1 of the same authors imply (6) with $g = T$. Finally, for the case $|1-\Gamma| > 1$, no reference has been found. However, from White (1958) it follows that

$$\frac{1}{\left(1-\Gamma\right)^{2T}}\sum_{t=1}^{T} \left(\Delta X_{t-1}\right)^2$$

is convergent in distribution. Further,

$$\begin{aligned}
\sum_{t=2}^{T} S_{t-1}\Delta X_{t-1} &= \sum_{t=2}^{T}\sum_{j=0}^{t-1} \varepsilon_j \sum_{k=1}^{t-1} \left(1-\Gamma\right)^{t-1-k} \varepsilon_k \\
&= \sum_{j=0}^{T-1}\sum_{k=1}^{T-1} \varepsilon_j\varepsilon_k \sum_{t=\max(j-k,0)}^{T-k-1} \left(1-\Gamma\right)^{t}.
\end{aligned}$$

By Chebychev's inequality it can be shown that this converges in probability to zero when normalised by $\left(1-\Gamma\right)^{T}\sqrt{T}$. ■

3. Bivariate first order model

The conjecture is stated for the bivariate first order model in Theorem 3.1. The proof is based on some joint convergence results for product moment matrices which are stated here and proved in the subsequent sections.

The model is given by (1) with $p = 2$, $k = 1$ and the hypothesis is that $rank\Pi \leq r = 1$. The distribution of empirical canonical correlations is invariant with respect to non-singular transformations, thus it can be assumed that the time series is of the form

$$\Delta X_t = \begin{pmatrix} A & B \\ 0 & 0 \end{pmatrix} X_{t-1} + \varepsilon_t, \tag{7}$$

where the innovations, ε_t, are standard normal distributed and have components $\varepsilon_{Y,t}$, $\varepsilon_{Z,t}$. The relation $\beta' X$, where $\beta' = (A, B)$, has special interest as it is the cointegrating relation in the stable case where $1+A$ is numerically smaller than one.

Define the sample product moments

$$\begin{pmatrix} S_{00} & S_{01} \\ S_{10} & S_{11} \end{pmatrix} = \frac{1}{T}\sum_{t=1}^{T} \begin{pmatrix} \Delta X_t \\ X_{t-1} \end{pmatrix} \begin{pmatrix} \Delta X_t \\ X_{t-1} \end{pmatrix}'$$

The squared empirical canonical correlations, $1 \geq \hat{\lambda}_1 \geq \hat{\lambda}_2 \geq 0$ are the solutions of the eigenvalue problem

$$\left|\lambda S_{11} - S_{10}\left(S_{00}\right)^{-1} S_{01}\right| = 0 \tag{8}$$

or of the dual problem

$$\left|\lambda S_{00} - S_{01}\left(S_{11}\right)^{-1} S_{10}\right| = 0. \tag{9}$$

The likelihood criterion is given by $-T\log(1-\hat{\lambda}_2)$ which is asymptotically equivalent to $T\hat{\lambda}_2$. For this model the conjectured result is a consequence of the following.

Theorem 3.1.
For $A \neq 0$. Define the constant $c = (A^2 + B^2)/(1-(1+A)^2)$. Then the largest eigenvalue, $\hat{\lambda}_1$, converges in probability to $c/(1+c)$ for $|1+A| < 1$ and to one otherwise. The normalised smallest eigenvalue, $T\hat{\lambda}_2$, converges in distribution to (2) *with one degree of freedom.*

For $A = B = 0$ the normalised eigenvalues, $T\hat{\lambda}_1, T\hat{\lambda}_2$, converge jointly in distribution to the eigenvalues of

$$\int_0^1 dW_u W_u' \left(\int_0^1 W_u W_u' du \right)^{-1} \int_0^1 W_u dW_u'$$

where W is a bivariate standard Brownian motion.

For $A = 0$, $B \neq 0$ then $\hat{\lambda}_1$ converges in probability to one and $T\hat{\lambda}_2$ converges in distribution to the I(2) distribution (4).

In the cases $A = B = 0$ and $|1 + A| < 1$ only I(0) and I(1) variables are involved and they are analysed by Johansen (1995a). A few additional comments are given before considering the non-stable cases.

Proof of Theorem 3.1 for $A = B = 0$.
This result can be deduced from Section 11.2 of Johansen (1995a).■

Proof of Theorem 3.1 for $|1 + A| < 1$.
The convergence of $T\hat{\lambda}_2$ follows from Theorem 11.1 of Johansen (1995a).

As for $\hat{\lambda}_1$ it follows from formula (11.16) of Johansen (1995a) that

$$\hat{\lambda}_1 \xrightarrow{\mathcal{P}} \Sigma_{\beta 0} \Sigma_{00}^{-1} \Sigma_{0\beta} \Sigma_{\beta\beta}^{-1}.$$

According to Lemma 10.3 the constants $\Sigma_{00}, \Sigma_{\beta 0}, \Sigma_{\beta\beta}$ are the limits in probability of the product moment matrices $S_{00}, \beta' S_{10}, \beta' S_{11} \beta$.

Now, multiplication of the first order equation for X by β' gives

$$\beta' \Delta X_t = A \beta' X_{t-1} + \beta' \varepsilon_{t.}$$

The time series $\beta' X_t$ can be given the stationary representation

$$\beta' X_t = \sum_{j=0}^{\infty} (1 + A)^j \beta' \varepsilon_{t-j}.$$

In particular

$$E\beta' X_t X_t' \beta = \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} (1 + A)^{j+k} \beta' E \varepsilon_{t-j} \varepsilon_{t-k}' \beta$$

$$
\begin{aligned}
&= \sum_{j=0}^{\infty} (1+A)^{2j} \beta'\beta \\
&= \frac{A^2+B^2}{1-(1+A)^2}.
\end{aligned}
$$

This is the constant c and hence

$$\beta' S_{11} \beta \xrightarrow{\mathcal{P}} \Sigma_{\beta\beta} = c.$$

By the identities of Lemma 10.1 it follows that

$$\Sigma_{\beta 0} \Sigma_{00}^{-1} \Sigma_{0\beta} \Sigma_{\beta\beta}^{-1} = \Sigma_{\beta\beta} \alpha' \left(\alpha \Sigma_{\beta\beta} \alpha' + \Omega\right)^{-1} \alpha \Sigma_{\beta\beta} \Sigma_{\beta\beta}^{-1} = \frac{c}{1+c}$$

and the Theorem 3.1 is proved for $|1+A| < 1$.■

An analysis of the convergence of product moment matrices is underlying Johansen's proof. For the remaining cases the joint convergence has to be stated.

Theorem 3.2. *The I(2) case, $A = 0, B \neq 0$. Let W be a standard Brownian motion and $\tilde{W}_{\cdot} = \int_0^{\cdot} W_u du$ the integrated process. Then the product moment matrices converges jointly as*

$$
\begin{pmatrix} \frac{1}{T^4} \sum_{t=1}^T Y_{t-1}^2 & \frac{1}{T^3} \sum_{t=1}^T Y_{t-1} Z_{t-1} \\ & \frac{1}{T^2} \sum_{t=1}^T Z_{t-1}^2 \end{pmatrix} \xrightarrow{\mathcal{D}} \begin{pmatrix} \int_0^1 \tilde{W}_u^2 du & \int_0^1 \tilde{W}_u W_u du \\ & \int_0^1 W_u^2 du \end{pmatrix},
$$

$$
\begin{pmatrix} \frac{1}{T^3} \sum_{t=1}^T \Delta Y_t Y_{t-1} & \frac{1}{T^2} \sum_{t=1}^T \Delta Y_t Z_{t-1} \\ \frac{1}{T^2} \sum_{t=1}^T \Delta Z_t Y_{t-1} & \frac{1}{T} \sum_{t=1}^T \Delta Z_t Z_{t-1} \end{pmatrix} \xrightarrow{\mathcal{D}} \begin{pmatrix} \int_0^1 \tilde{W}_u W_u du & \int_0^1 W_u^2 du \\ \int_0^1 \tilde{W}_u dW_u & \int_0^1 W_u dW_u \end{pmatrix},
$$

$$
\begin{pmatrix} \frac{1}{T^2} \sum_{t=1}^T (\Delta Y_t)^2 & \frac{1}{T} \sum_{t=1}^T \Delta Y_t \Delta Z_t \\ & \frac{1}{T} \sum_{t=1}^T (\Delta Z_t)^2 \end{pmatrix} \xrightarrow{\mathcal{D}} \begin{pmatrix} \int_0^1 W_u^2 du & \int_0^1 W_u dW_u \\ & 1 \end{pmatrix}.
$$

Proof of Theorem 3.2.
From (7) it follows that

$$Y_t = B\sum_{j=1}^{t-1} Z_j + \sum_{j=1}^{t} \varepsilon_{Y,j}, \qquad Z_t = \sum_{j=1}^{t} \varepsilon_{Z,j}.$$

The second component of Y is asymptotically negligible. Donsker's invariance principle describes the convergence of the random walk Z_t. Summation is a continuous function on the path space $D[0,1]$, thus by the Continuous Mapping Theorem (see Theorem 5.1 of Billingsley, 1968)

$$\left(\frac{1}{\sqrt{T}} Z_{[Tu]}, \frac{1}{T\sqrt{T}} Y_{[Tu]}\right) \xrightarrow{\mathcal{D}} \left(W_u, \tilde{W}_u\right) \qquad u \in [0,1]$$

on the path space $(D[0,1])^2$. The convergence of the product moment matrices follows by further application of the Continuous Mapping Theorem.■

Theorem 3.3. *The non-stable case,* $A = -2$, $|1+A| > 1$. *Define the normalisation*

$$g = \begin{cases} T & \text{for } A = -2, \\ \frac{|1+A|^T}{(1+A)^2 - 1} & \text{for } |1+A| > 1. \end{cases}$$

Let W be a standard Brownian motion and let V denote a random variable which is independent of W. Then the following results hold jointly for $\alpha > 1$,

$$\frac{1}{g^2}\sum_{t=1}^{T} \begin{pmatrix} \Delta Y_t \\ Y_{t-1} \end{pmatrix} \begin{pmatrix} \Delta Y_t \\ Y_{t-1} \end{pmatrix}' \xrightarrow{\mathcal{D}} \begin{pmatrix} A^2 & A \\ A & 1 \end{pmatrix} V, \tag{10}$$

$$\frac{1}{T^\alpha}\sum_{t=1}^{T} \begin{pmatrix} \Delta Y_t \\ Y_{t-1} \end{pmatrix} \begin{pmatrix} \Delta Z_t \\ Z_{t-1} \end{pmatrix}' \xrightarrow{\mathcal{P}} \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}, \tag{11}$$

$$\frac{1}{T}\sum_{t=1}^{T} (\Delta Z_t)^2 \xrightarrow{\mathcal{P}} 1, \tag{12}$$

$$\frac{1}{T}\sum_{t=1}^{T} Z_{t-1}\Delta Z_t \xrightarrow{\mathcal{D}} \int_0^1 W_u dW_u, \tag{13}$$

$$\frac{1}{T^2}\sum_{t=1}^{T} Z_{t-1}^2 \xrightarrow{\mathcal{D}} \int_0^1 W_u^2 du. \tag{14}$$

In case $A = -2$ then V has the distribution of the integral of a squared Brownian motion. For $|1+A| > 1$ the Laplace transform of V is given by (20)*. Moreover, for $|1+A| > 1$ and $\alpha = 1$ then the distribution of the left hand side of* (11) *is divergent.*

The proof of this theorem follows in the subsequent sections.

The Theorems 3.2 and 3.3 give sufficient results for proving the remaining part of Theorem 3.1. For the case $A = 0, B \neq 0$ the proof was given previously in Nielsen and Rahbek (1997).

Proof of Theorem 3.1 for remaining cases.
The argument follows those of Theorem 11.1 Johansen (1995a). However, whereas Johansen consider the eigenvalue problem (8) then for the case $A = 0, B \neq 0$ it seems necessary to consider the dual eigenvalue problem (9). The proof has two steps. First, consistency of the eigenvalues is proved, next, the convergence result for the normalised smallest eigenvalue is considered.

Some preliminary definitions are useful:

$$a = \int_0^1 \tilde{W}_u^2 du, \quad b = \int_0^1 \tilde{W}_u W_u du, \quad c = \int_0^1 W_u^2 du,$$

$$d = \int_0^1 \tilde{W}_u dW_u, \quad e = \int_0^1 W_u dW_u.$$

The consistency is proved by analysing the eigenvalues of (9). Define the diagonal matrices N and M with diagonal elements $(n, 1)$ and $(m, 1/\sqrt{T})$ respectively. In case $A = 0, B \neq 0$ then $n = 1/\sqrt{T}$ and $m = 1/T\sqrt{T}$ whereas for $A = -2$ and $|1+A| > 1$ then $n = m = \sqrt{T/g}$. Thus the eigenvalues of (9) are the same as the solutions of

$$\left| \lambda N S_{00} N - N S_{01} M \left(M S_{11} M \right)^{-1} M S_{10} N \right| = 0.$$

For $A = -2$ and $|1+A| > 1$ then by Theorem 3.2 the determinant converges to

$$\left| \lambda \begin{pmatrix} A^2 V & 0 \\ 0 & 1 \end{pmatrix} - \begin{pmatrix} AV & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} V & 0 \\ 0 & c \end{pmatrix}^{-1} \begin{pmatrix} AV & 0 \\ 0 & 0 \end{pmatrix} \right| = \lambda (\lambda - 1) A^2 V.$$

This shows that the eigenvalues converge to 1 and 0 respectively. For $A = 0, B \neq 0$ the determinant converges, using Theorem 3.3, to

$$\left| \lambda \begin{pmatrix} c & 0 \\ 0 & 1 \end{pmatrix} - \begin{pmatrix} b & c \\ 0 & 0 \end{pmatrix} \begin{pmatrix} a & b \\ b & c \end{pmatrix}^{-1} \begin{pmatrix} b & 0 \\ c & 0 \end{pmatrix} \right| = \lambda (\lambda - 1) c$$

and again the eigenvalues converge to 1 and 0 respectively.

The convergence of the smallest eigenvalues is proved by fixing $\rho = \lambda T$ in the asymptotic argument. Now, define the diagonal matrix Q with elements $(n, \sqrt{T})$. Then the eigenvalue problem of interest is

$$\left| \frac{\rho}{T} Q S_{00} Q - Q S_{01} M (M S_{11} M)^{-1} M S_{10} Q \right| = 0.$$

Using Theorem 3.2 then for $A = -2$ and $|1 + A| > 1$ the determinant converges to

$$\left| \rho \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} - \begin{pmatrix} AV & 0 \\ 0 & e \end{pmatrix} \begin{pmatrix} V & 0 \\ 0 & c \end{pmatrix}^{-1} \begin{pmatrix} AV & 0 \\ 0 & e \end{pmatrix} \right| = -A^2 V \left(\rho - \frac{e^2}{c} \right),$$

which shows that the normalised smallest eigenvalue, $T\hat{\lambda}_2$, converges to the distribution of (2). For $A = 0, B \neq 0$ then by Theorem 3.3 the determinant converges to

$$\left| \rho \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} - \begin{pmatrix} b & c \\ d & e \end{pmatrix} \begin{pmatrix} a & b \\ b & c \end{pmatrix}^{-1} \begin{pmatrix} b & d \\ c & e \end{pmatrix} \right| = -c \left(\rho - \frac{(d - be/c)^2}{a - b^2/c} \right)$$

so that $T\hat{\lambda}_2$ converges to the distribution of (4). ■

4. An explosive autoregression with an intercept.

For the joint convergence result in case $|1 + A| > 1$ some convergence results have to be proved. In this section the sum of the squared levels of an explosive autoregression with an intercept is analysed, where the intercept turns out to represent the parameter B of (7). In the next section it is proved that this convergence is joint with the convergence of the random walk given by Z.

Consider a first order autoregressive time series given by

$$\Delta Y_t = A (Y_{t-1} + C) + \varepsilon_t \tag{15}$$

where the initial value, Y_0, is fixed and the innovations are independently standard normal distributed. The parameters are chosen such that the time series has an explosive root, $|1+A|>1$, and $C \in \mathbf{R}$.

Then the following convergence result holds

Theorem 4.1. *The Laplace transform*

$$\varphi_T(\alpha) = E\exp\left(-\frac{\alpha}{g^2}\sum_{t=1}^{T} Y_{t-1}^2\right)$$

is convergent on the set given by $1+2\alpha>0$ *with the limit*

$$(1+2\alpha)^{-1/2}\exp\left\{-\frac{\left[(1+A)^2-1\right](U_0+C)^2\alpha}{1+2\alpha}\right\}. \tag{16}$$

This limit is a Laplace transform, it is well-defined for $1+2\alpha>0$ *and* $\sum Y^2/g^2$ *converges in distribution to the distribution given by this.*

The proof involves some algebraic result regarding Toeplitz-like matrices which will be stated first. Next, the proof of the Lemma follows.

Lemma 4.2. *(White, Larsson) Consider the* $T\times T$*-matrix given by*

$$\begin{pmatrix} p & q & & & \\ p & \ddots & \ddots & & \\ & & \ddots & p & q \\ & & & q & 1 \end{pmatrix}. \tag{17}$$

For $j\in\mathbf{N_0}$ *the lower right* $j\times j$*-block matrix has determinant*

$$D_j = \begin{cases} \frac{1-s}{r-s}r^j + \frac{1-r}{s-r}s^j & \text{for } p^2\neq 4q^2, \\ r^j + (1-r)\,j r^{j-1} & \text{for } p^2 = 4q^2. \end{cases}$$

where

$$r = \frac{p}{2}+\sqrt{\left(\frac{p}{2}\right)^2-q^2}, \qquad s = \frac{p}{2}-\sqrt{\left(\frac{p}{2}\right)^2-q^2}.$$

For $j\in\mathbf{N}$ *the top left* $(j-1)\times(j-1)$*-block matrix has determinant*

$$D^*_{j-1} = \begin{cases} \frac{1}{r-s}r^j + \frac{1}{s-r}s^j & \text{for } p^2\neq 4q^2, \\ (j+1)\,r^j & \text{for } p^2 = 4q^2. \end{cases}$$

The inverse of the matrix has (i,j)-th element

$$\begin{cases} (-q)^{j-i} D_{T-j} D^*_{i-1} / D_T & \text{for } i \leq j, \\ (D^{-1})_{ji} & \text{for } i > j. \end{cases}$$

Proof: The determinant, D, is found by White (1958): The determinant fulfil the second order difference equation

$$x_j = p x_{j-1} - q^2 x_{j-2}.$$

By solving the corresponding second order polynomial and applying the initial conditions

$$D_2 = p - q^2, \qquad D_1 = 1, \qquad D_0 = 1$$

the result follows. Larsson (1994, Lemma A.3) finds that the determinant D^* satisfies the same difference equation with initial conditions

$$D_2 = p^2 - q^2, \qquad D_1^* = p, \qquad D_0^* = 1.$$

The inverse of the matrix is found by Cramer's rule. First, eliminate the i-th row and the j-th column and then expand the new matrix along either the i-th column or the j-th row.■

Proof of Theorem 4.1.
The case where $C = 0$ is considered by White (1958). This proof, for a general C, follows his very closely.

From the equation (15) it follows that

$$\begin{aligned} Y_t &= \sum_{j=1}^{t} (1+A)^{t-j} (\varepsilon_j + AC) + (1+A)^t Y_0 \\ &= \sum_{j=1}^{t} (1+A)^{t-j} \varepsilon_j - C + (1+A)^t (Y_0 + C). \end{aligned}$$

Thus, $Y_t = U_t - C$, where U is an explosive autoregressive time series defined by

$$\Delta U_t = A U_{t-1} + \varepsilon_t$$

with initial value $U_0 = Y_0 + C$. Thus,

$$\begin{aligned}
\sum_{t=1}^{T} Y_{t-1}^2 &= \sum_{t=1}^{T} (U_{t-1} - C)^2 \\
&= \sum_{t=1}^{T-1} U_t^2 + U_0^2 - 2C \sum_{t=1}^{T-1} U_t - 2CU_0 + TC^2, \\
\sum_{t=1}^{T} \varepsilon_t^2 &= \sum_{t=1}^{T} (\Delta U_t - AU_{t-1})^2 \\
&= U_T^2 - 2\,(1+A) \sum_{t=2}^{T} U_t U_{t-1} + \left[1 + (1+A)^2\right] \sum_{t=2}^{T-1} U_{t-1}^2 \\
&\quad -2\,(1+A)\, U_0 U_1 + (1+A)^2\, U_0^2.
\end{aligned}$$

Now, define the matrix M_T as a matrix of the form (17) with

$$p = 1 + (1+A)^2 + \frac{2\alpha}{g^2}, \qquad q = -\,(1+A)\,,$$

and let $\mu = \mu_1 + \mu_2$ where

$$\begin{aligned}
\mu_1 &= (1+A)\, U_0 (1, 0, \ldots, 0)' M_T^{-1}, \\
\mu_2 &= \frac{-2\alpha C}{g^2} (1, \ldots, 1, 0)' M_T^{-1}.
\end{aligned}$$

Further, let $\underline{U}$ be the vector of the U's. Then the Laplace transform $\varphi_T(\alpha)$ equals

$$\begin{aligned}
&\exp\left[-\frac{\alpha}{g^2}\left(U_0^2 - 2CU_0 + TC^2\right) - \frac{1}{2}\,(1+A)^2\, U_0^2 + \frac{1}{2}\mu' M_T \mu\right] \\
&\times \int_{\mathbf{R^T}} (2\pi)^{-T/2} \exp\left[-\frac{1}{2}\,(\underline{U} - \mu)'\, M_T\, (\underline{U} - \mu)\right] d\underline{U}.
\end{aligned}$$

The integral is a normal integral and it equals the inverse square root of the determinant of M_T whenever M_T is positive definite. Note, that the vectors μ_1, μ_2 are well-defined in this case. Now, for $\alpha = 0$ the integrand is the density for the innovations and hence the matrix is positive definite with determinant one. Jensen and Nielsen (1997) argues that M_T is positive definite in the connected component around the origin where the determinant is positive. Moreover, the set of convergence points for the inverse of the determinant of M_T is given by the connected component around the origin where the limit of the determinant is positive. This determines the set of convergence points for the Laplace transform because the limit of the exponential function above turns out to be defined for all values of α.

Now, the cross products of the μ-vectors through M_T are

$$
\begin{aligned}
\mu_1' M_T \mu_1 &= (1+A)^2 U_0^2 \left(M_T^{-1}\right)_{11} \\
\mu_1' M_T \mu_2 &= -(1+A) U_0 \frac{2\alpha C}{g^2} \sum_{j=1}^{T-1} \left(M_T^{-1}\right)_{1j}, \\
\mu_2' M_T \mu_2 &= \frac{4\alpha^2 C^2}{g^4} \left[2 \sum_{k=1}^{T-1} \sum_{j=1}^{k-1} \left(M_T^{-1}\right)_{jk} + \sum_{j=1}^{T-1} \left(M_T^{-1}\right)_{jj} \right].
\end{aligned}
$$

In Lemma 4.2 the determinant of M_T is given in terms of quantities r, s. These are expanded in formula (4.7) of White (1958):

$$
\begin{aligned}
r &= 1 - \frac{2\alpha \left[(1+A)^2 - 1\right]}{(1+A)^{2T}} + \mathrm{O}\left(|1+A|^{-3T}\right), && (18) \\
s &= (1+A)^2 + \mathrm{O}\left(|1+A|^{-2T}\right). && (19)
\end{aligned}
$$

Thus from Lemma 4.2

$$
\begin{aligned}
D_j &= 1 + 2\alpha (1+A)^{2(j-T)} + \mathrm{O}\left(|1+A|^{-T}\right), \\
D_j^* &= \frac{1-(1+A)^{2j+2}}{1-(1+A)^2} + \mathrm{O}\left(|1+A|^{2(j-T)}\right),
\end{aligned}
$$

and

$$
\begin{aligned}
\left(M_T^{-1}\right)_{11} &= D_{T-1}/D_T \to \frac{1+2\alpha(1+A)^{-2}}{1+2\alpha}, \\
\sum_{j=1}^{T-1} \left(M_T^{-1}\right)_{1j} &= \sum_{j=1}^{T-1} (1+A)^{j-1} D_{T-j}/D_T = \mathrm{O}\left(|1+A|^{T}\right),
\end{aligned}
$$

$$\begin{aligned}
\sum_{j=1}^{T-1} \left(M_T^{-1}\right)_{jj} &= \sum_{j=1}^{T-1} D_{T-j} D_{j-1}^{*} / D_T = \mathrm{O}\left(|1+A|^{2T}\right), \\
\sum_{k=1}^{T-1} \sum_{j=1}^{k-1} \left(M_T^{-1}\right)_{jk} &= \sum_{k=1}^{T-1} \sum_{j=1}^{k-1} (1+A)^{k-j} D_{T-k} D_{j-1}^{*} / D_T = \mathrm{O}\left(|1+A|^{2T}\right).
\end{aligned}$$

As a consequence

$$\mu' M_T \mu \to U_0^2 \frac{(1+A)^2 + 2\alpha}{1+2\alpha}.$$

which finish the proof.■

5. Joint convergence of a random walk and an explosive root

From Theorem 4.1 it can be concluded that $\sum Y^2/g^2$ converges in distribution when Y is an explosive first order autoregressive time series. The limit distribution depends on the initial value. The next theorem states that this convergence is jointly with the convergence of the random walk Z.

Let $|1+A| > 1$ for a time series given by (7). Then the bivariate time series X with components Y, Z is given by

$$\begin{aligned}
\Delta Y_t &= A Y_{t-1} + \varepsilon_{Y,t} \\
\Delta Z_t &= \varepsilon_{Z,t}
\end{aligned}$$

with initial values Y_0, Z_0 and innovations which are standard normal distributed with correlation C. The random walk is defined on the path space $D\,[0,1]$. Since the limiting Brownian motion walk has continuous paths the uniform metric can be applied.

Theorem 5.1. *Consider the space* $\mathbf{R} \times \mathbf{D}\,[\mathbf{0}, \mathbf{1}]$ *with the metric*

$$d\left[(y', z'), (y'', z'')\right] = \sqrt{(y' - y'')^2 + \sup_{0 \le u \le 1} (z_u' - z_u'')^2}.$$

Then

$$\left(\frac{1}{g_T^2} \sum_{t=1}^{T} Y_{t-1}^2, \frac{1}{\sqrt{T}} Z_{[T\cdot]}\right) \xrightarrow{\mathcal{D}} (V, W.)$$

where the limiting distribution V is independent of the Brownian motion W. The distribution of V has Laplace transform

$$E\exp(-\alpha V) = (1+2\alpha)^{-1/2}\exp\left\{-\frac{\left[(1+A)^2-1\right]Y_0^2\alpha}{1+2\alpha}\right\}. \tag{20}$$

In particular, for $Y_0 = 0$ then V is χ^2-distributed with one degree of freedom.

Proof. The marginal convergence for each of the two components follows from Theorem 4.1 and Donsker's invariance principle (Billingsley, 1968, Theorem 10.1).

The arguments for the joint convergence resemble those leading to the invariance principle for the random walk. First, it has to be proved that the joint distribution of $\sum Y^2/g^2$ and all finite dimensional marginals of the random walk are convergent. Secondly, it has to be proved that the family of probability measures is tight.

To prove that the finite dimensional marginal distributions converge consider first the pair

$$\frac{1}{g^2}\sum_{t=1}^{T}Y_{t-1}^2, \frac{1}{\sqrt{T}}Z_T. \tag{21}$$

The convergence will be proved by analysing the Laplace transform of this statistic. The innovations of the time series Y and Z have correlation C. Independent innovations can be obtained by conditioning, so that

$$\begin{aligned}\Delta Y_t &= AY_{t-1} + C\varepsilon_{Z,t} + \eta_t, \\ \Delta Z_t &= \varepsilon_{Z,t},\end{aligned} \tag{22}$$

where $\eta_t = \varepsilon_{Y,t} - C\varepsilon_{Z,t}$ has variance $1 - C^2$ and it is independent of $\varepsilon_{Z,t}$. In the following the subscript, Z, is omitted and $\underline{\varepsilon}, \underline{\eta}$ denotes the vectors of all ε_t's and η_t's. The Laplace transform of $\sum Y^2/g^2, Z/\sqrt{T}$,

$$E\exp\left(-\frac{\alpha}{g^2}\sum_{t=1}^{T}Y_{t-1}^2 - \frac{\beta}{\sqrt{T}}Z_T\right),$$

then equals

$$\int_{\mathbf{R^{2T}}}\exp\left[-\frac{\alpha}{g^2}\sum_{t=1}^{T}Y_{t-1}^2 - \frac{\beta}{\sqrt{T}}\sum_{t=1}^{T}\varepsilon_t - \frac{1}{2}\sum_{t=1}^{T}\varepsilon_t^2 - \frac{1}{2(1-C^2)}\sum_{t=1}^{T}\eta_t^2\right]d\left(\underline{\varepsilon},\underline{\eta}\right)$$

normalised by $(2\pi\sqrt{1-C^2})^T$. The integrand equals

$$\exp\left[-\frac{\alpha}{g^2}\sum_{t=1}^{T}Y_{t-1}^2-\frac{1}{2}\sum_{t=1}^{T}\left(\varepsilon_t+\frac{\beta}{\sqrt{T}}\right)^2+\frac{\beta^2}{2}-\frac{1}{2\left(1-C^2\right)}\sum_{t=1}^{T}\eta_t^2\right].$$

Thus by the substitution $\tilde{\varepsilon}=\varepsilon+\beta/\sqrt{T}$ the Laplace transform equals

$$\frac{\exp\left(\beta^2/2\right)}{\left(2\pi\sqrt{1-C^2}\right)^T}\int_{\mathbf{R^{2T}}}\exp\left[-\frac{\alpha}{g_T^2}\sum_{t=1}^{T}Y_{t-1}^2-\frac{1}{2}\sum_{t=1}^{T}\tilde{\varepsilon}_t^2-\frac{1}{2\left(1-C^2\right)}\sum_{t=1}^{T}\eta_t^2\right]d\left(\underline{\tilde{\varepsilon}},\underline{\eta}\right). \tag{23}$$

The function $exp(\beta^2/2)$ is the Laplace transform of the standard normal distribution so it remains to prove that the normalised integral of (23) is related to the Laplace transform, φ_T of Theorem 4.1. Now, by (22)

$$\begin{aligned}
Y_t &= \sum_{j=1}^{t}(1+A)^{t-j}\left(C\varepsilon_j+\eta_j\right)+(1+A)^t Y_0 \\
&= \sum_{j=1}^{t}(1+A)^{t-j}\left(C\tilde{\varepsilon}_j+\eta_j-\frac{C\beta}{\sqrt{T}}\right)+(1+A)^t Y_0.
\end{aligned}$$

The variables $\tilde{\varepsilon}_t,\eta_t$ are transformed into the new variables $\nu_t=C\tilde{\varepsilon}_t+\eta_t$ and $\eta_t/(1-C^2)-\tilde{\varepsilon}_t/C$. These variables are independent, ν_t is standard normal distributed and integration with respect to the latter variable gives

$$(2\pi)^{-T/2}\int_{\mathbf{R^{2T}}}\exp\left(-\frac{\alpha}{g^2}\sum_{t=1}^{T}U_{t-1}^2-\frac{1}{2}\sum_{t=1}^{T}\nu_t\right)d\underline{\nu},$$

where the time series U_t is given by

$$\Delta U_t = AU_{t-1}-\frac{C\beta}{\sqrt{T}}+\nu_t$$

and has initial value $U_0=Y_0$ and standard normal innovations. This integral is a Laplace transform and converges to the Laplace transform (20) in a neighbourhood of the origin according to Theorem 4.1 and the convergence in distribution of (21) follows by Jensen and Nielsen (1997). Note that the joint Laplace transform is a product of the two marginal Laplace transform which ensures the asymptotic independence of (21).

Joint convergence of $\sum Y^2/g^2$ and any finite dimensional marginal distribution of the random walk Z can be proved along the same ideas.

As for the tightness argument the tightness argument for Donsker's invariance principle has to be mimicked. First, the spaces $\mathbf{R}$ and $C[0,1]$ are both separable and complete, thus the product space is separable and complete (Billingsley, 1968, p. 225). Secondly, the Arzelà-Ascoli Theorem has to be restated by adding the condition

$$\sup_{y \in A} |y| < \infty.$$

The proof follows by applying the suggested metric. Similarly the Theorems 8.1-8.4 of Billingsley (1968) can be restated.■

6. Convergence of the product moment matrices

In this section the proof of Theorem 3.3 is given.

This is based on some joint convergence results. For the explosive case, $|1+A| > 1$ Theorem 5.1 gives that the pair

$$\frac{1}{g_T^2} \sum_{t=1}^{T} Y_{t-1}^2, \qquad \frac{1}{\sqrt{T}} Z_{[T\cdot]} \tag{24}$$

converges jointly on $\mathbf{R} \times \mathbf{D[0,1]}$. This also holds for the case of negative unit root, $A = -2$: Then the marginal time series $(-1)^t Y_t$ is a random walk with innovations $(-1)^t \varepsilon_t$. Thus the limit of $\sum Y^2/g^2$, denoted V, has the same distribution as that of the integral of a squared Brownian motion. The alternating sign imply that this time series and the random walk Z are asymptotically independent, even if the innovations are correlated, see Theorem 2.2 of Chan and Wei (1988). In particular, by continuous mapping, the pair (24) converges jointly.

Proof of Theorem 3.3.

The convergence (12)-(14). It is well-known that this convergence hold. One argument is to rewrite the mixed product sum using partial summation in terms of the sum of the squared levels and the law of large numbers to the sum of the squared differences. The convergence then follows, jointly with that of (24), using the Law of Large Numbers and the Continuous Mapping Theorem.

The convergence (10). Using the equation for Y the left hand side of (10) can be rewritten as

$$\begin{pmatrix} A^2 & A \\ A & 1 \end{pmatrix} \frac{1}{g^2} \sum_{t=1}^{T} Y_{t-1}^2 + \begin{pmatrix} 2A & 1 \\ 1 & 0 \end{pmatrix} \frac{1}{g^2} \sum_{t=1}^{T} Y_{t-1}\varepsilon_{Y,t} + \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \frac{1}{g^2} \sum_{t=1}^{T} \varepsilon_{Y,t}^2$$

The second sum is actually of order g so the second term converges to zero, following White (1958). The last term converges to zero by the Law of Large Numbers.
The convergence (11).Using the equation for Y, the left hand side of (11) can be rewritten as

$$\begin{pmatrix} A \\ 1 \end{pmatrix} \frac{1}{T^\alpha} \sum_{t=1}^{T} Y_{t-1} \begin{pmatrix} \Delta Z_t \\ Z_{t-1} \end{pmatrix}' + \begin{pmatrix} 1 \\ 0 \end{pmatrix} \frac{1}{T^\alpha} \sum_{t=1}^{T} \varepsilon_{Y,t} \begin{pmatrix} \Delta Z_t \\ Z_{t-1} \end{pmatrix}'.$$

The four elements behave quite differently, however, each of them converge to zero when normalised with T^α, where $\alpha > 1$.

The last term converge to zero since

$$\frac{1}{T} \sum_{t=1}^{T} \varepsilon_{Y,t} \begin{pmatrix} \Delta Z_t \\ Z_{t-1} \end{pmatrix} \xrightarrow{\mathcal{D}} \begin{pmatrix} C \\ \int_0^1 W_u dB_u \end{pmatrix}$$

where B is a standard Brownian motion. This follows from the law of large numbers respectively Theorem 2.4 of Chan and Wei (1988).

The convergence of the first term,

$$\frac{1}{T^\alpha} \sum_{t=1}^{T} Y_{t-1} \begin{pmatrix} \Delta Z_t \\ Z_{t-1} \end{pmatrix}' \tag{25}$$

is now proved separately for the cases of a negative unit root and of an explosive root.

The case of a negative unit root, $A = -2$.

The first component of (25) can be rewritten as

$$\sum_{t=1}^{T} Y_{t-1}\Delta Z_t = -\sum_{t=1}^{T} \sum_{j=1}^{t-1} \left[(-1)^j \varepsilon_{Y,j} \right] \left[(-1)^t \varepsilon_{Z,t} \right].$$

Due to the symmetry of the normal distribution a variable with the same distribution is obtained by omitting the negative signs. Normalised by T this variable converges to the stochastic integral of B with respect to W by Theorem 2.4 of Chan and Wei (1988).

The last component of (25), the mixed product sum of the two levels is divergent when normalised by T: It can be rewritten as

$$\sum_{t=1}^{T} Y_{t-1} Z_{t-1} = \sum_{t=1}^{T} (-1)^{t-1} \sum_{j=1}^{t-1} \left[(-1)^j \varepsilon_{Y,j} \right] \sum_{k=1}^{t-1} \varepsilon_{Z,k}.$$

Change of summation index leads to the expression

$$\sum_{j=1}^{T-1} \sum_{k=1}^{T-1} \left[(-1)^j \varepsilon_{Y,j} \right] \varepsilon_{Z,k} \sum_{t=\max(j,k)}^{T-1} (-1)^t,$$

where

$$\sum_{t=\max(j,k)}^{T-1} (-1)^t = \begin{cases} 0 & \text{if } T - \max(j,k) \text{ is even,} \\ 1 & \text{if } T - \max(j,k) \text{ is odd and } T \text{ is odd,} \\ \text{-1} & \text{if } T - \max(j,k) \text{ is odd and } T \text{ is even.} \end{cases}$$

Now, introduce the partial sums

$$\begin{array}{ll} S_{Y,1}(t) = \sum_{j=1}^{t} \left[(-1)^{2j-1} \varepsilon_{Y,2j-1} \right], & S_{Z,1}(t) = \sum_{j=1}^{t} \varepsilon_{Z,2j-1}, \\ S_{Y,2}(t) = \sum_{j=1}^{t} \left[(-1)^{2j} \varepsilon_{Y,2j} \right], & S_{Z,2}(t) = \sum_{j=1}^{t} \varepsilon_{Z,2j}, \end{array}$$

where the pair $S_{Y,1}, S_{Z,1}$ is stochastically independent of $S_{Y,2}, S_{Z,2}$. Hence, the sum of the product of the levels equals

$$-S_{Y,1}\left(\frac{T}{2}\right) S_{Z,1}\left(\frac{T}{2}\right) - \sum_{j=1}^{T/2} S_{Z,2}(j-1) \Delta S_{Y,1}(j) - \sum_{j=1}^{T/2} S_{Y,2}(j-1) \Delta S_{Z,1}(j)$$

for T even, and

$$S_{Y,2}\left(\frac{T-1}{2}\right) S_{Z,2}\left(\frac{T-1}{2}\right) + \sum_{j=1}^{(T-1)/2} S_{Z,1}(j) \Delta S_{Y,2}(j) + \sum_{j=1}^{(T-1)/2} S_{Y,1}(j) \Delta S_{Z,2}(j)$$

for T odd. It follows that each of the even and odd sub-sequences,

$$\frac{1}{2T}\sum_{j=1}^{2T} Y_{t-1}Z_{t-1}, \qquad \frac{1}{2T-1}\sum_{j=1}^{2T-1} Y_{t-1}Z_{t-1}$$

converges in distribution, however, the entire sequence is divergent. Using Chebychev's inequality convergence can be proved for the sequence normalised by T^{α} rather than T.

The case of an explosive root, $|1+A|>1$.

For the first component of (25) note, that

$$\sum_{t=1}^{T} Y_{t-1}\Delta Z_t = \sum_{t=1}^{T}\sum_{j=1}^{t-1}(1+A)^{t-j-1}\varepsilon_{Y,j}\varepsilon_{Z,t},$$

whereas the second component of (25) can be rewritten by change of summation index and summation of a finite power series as

$$\begin{aligned}
\sum_{t=1}^{T} Y_{t-1}Z_{t-1} &= \sum_{t=1}^{T}\sum_{j=1}^{t-1}\sum_{k=1}^{t-1}(1+A)^{t-j-1}\varepsilon_{Y,j}\varepsilon_{Z,k}\\
&= \sum_{j=1}^{T-1}\sum_{k=1}^{T-1}\varepsilon_{Y,j}\varepsilon_{Z,k}\sum_{t=\max(j,k)}^{T-1}(1+A)^{t-j}\\
&= \sum_{j=1}^{T-1}\sum_{k=1}^{T-1}\varepsilon_{Y,j}\varepsilon_{Z,k}\frac{(1+A)^{T-j}-(1+A)^{\max(0,k-j)}}{A}.
\end{aligned}$$

In both case the second moment is of order $(1+A)^{2T}$ and thus the conclusion follows by Chebychev's inequality.■

References

Billingsley, P.(1968) *Convergence of probability measures.* New York: Wiley.

Chan, N.H. and C.Z. Wei (1988) Limiting distributions of least squares estimates of unstable autoregressive processes. *Annals of Statistics* 16, 367-401.

Johansen, S. (1988) Statistical analysis of cointegration vectors. *Journal of Economic Dynamics and Control* 12, 231-54.

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

Johansen, S. (1995b) A statistical analysis of cointegration for I(2) variables. *Econometric Theory* 11, 25-59.

Larsson, R. (1994) Bartlett corrections for unit root test statistics. *Preprint of Institute of Mathematical Statistics*, *University of Copenhagen.*

Nielsen, B. (1997) On the distribution of tests for cointegration rank. *Mimeo, Nuffield College.*

Nielsen, B. and A. Rahbek (1997) Strategies for cointegration testing based on inequalities. *Mimeo, University of Copenhagen.*

White, J.S. (1958) The limiting distribution of the serial correlation coefficient in the explosive case. *Annals of Mathematical Statistics* 29, 1188-1197.