THE MILTON FRIEDMAN INSTITUTE FOR RESEARCH IN ECONOMICS

MFI Working Paper Series
No. 2009-004

Unit Roots in White Noise

Harald Uhlig
University of Chicago, Department of Economics

Alexei Onatski
Columbia University, Department of Economics

Revised: March 2011

Original Version: March 2009

1126 East 59th Street
Chicago, Illinois 60637

T: 773.702.7587
F: 773.795.6891
mfi@uchicago.edu

Unit Roots in White Noise

Alexei Onatski and Harald Uhlig*

March 24, 2011

Abstract

We show that the empirical distribution of the roots of the vector auto-regression of order p fitted to T observations of a general stationary or non-stationary process, converges to the uniform distribution over the unit circle on the complex plane, when both T and p tend to infinity so that $(\ln T)/p \to 0$ and $p^3/T \to 0$. In particular, even if the process is a white noise, nearly all roots of the estimated vector auto-regression will converge by absolute value to unity. For fixed p, we derive an asymptotic approximation to the expected empirical distribution of the estimated roots as $T \to \infty$. The approximation is concentrated in a circular region in the complex plane for various data generating processes and sample sizes.

*Alexei Onatski, Dept. of Economics, University of Cambridge, ao319@cam.ac.uk and Harald Uhlig, Dept. of Economics, University of Chicago, huhlig@uchicago.edu

1 Introduction

The last two decades have witnessed the rapid development of econometric methods dealing with detecting and analyzing nonstationary or highly persistent features in time series: see e.g. Műller and Watson (2008) and the references therein for a recent leading example. Researchers are often inclined to interpret the presence of an estimated root with a near-unit absolute value as evidence for nonstationarity in the data. Should they? Recent studies suggest controversial answers. Johansen (2003) established $\sqrt{T}$ asymptotic normality of the estimated simple auto-regressive roots, which suggests that a large estimated root should indicate persistence. Granger and Jeon (2006) has found that the roots of auto-regressions (AR) fitted to US macroeconomic series when plotted on the complex plane "lie in an indistinct 'milky-way' band or 'halo', with modulus around 0.8". They speculate that such a strange pattern reflects over-fitting rather than the persistence of the underlying series. Nielsen and Nielsen (2008) point out that the usual $\sqrt{T}$ rate of convergence slows down to $T^{1/2k}$ for the roots of k-th order. They use this fact to provide a partial explanation of the 'halo phenomenon'.

In this paper, we shed light on these issues. We study the roots of the characteristic polynomials of vector auto-regressions (VAR) fitted either to stationary or to non-stationary data. We show that the empirical distribution of the roots converges to the uniform distribution over the unit circle when both the sample size T and the order p of the fitted VAR tend to infinity so

that $(\log T)/p \to 0$ and $p^3/T \to 0$. This convergence is independent from the covariance structure of the process approximated by the VAR. In particular, even if the process is a white noise, the absolute value of nearly all roots of the estimated VAR will converge to unity.

A simple piece of intuition explaining such a strange behavior of the absolute values of the roots is as follows. Imagine fitting a long AR $y_t + a_1 y_{t-1} + ... + a_p y_{t-p} = u_t$ to T observations of a white noise. Consider the estimated characteristic polynomial $z^p + \hat{a}_1 z^{p-1} + ... + \hat{a}_p = 0$ and denote its roots as z_i, $i = 1, ..., p$. By Vieta's theorem, $\sum_{i=1}^{p} \log|z_i| = \log|\hat{a}_p|$. But $\hat{a}_p$ should converge to zero at the rate $T^{-1/2}$ as $T \to \infty$. Hence, $\log|\hat{a}_p|$ must be of order $-\frac{1}{2}\log T$ for large T. On the other hand, with high probability, the estimated AR will be invertible so that $\log|z_i| < 0$ for all $i = 1, ..., p$. Therefore, with $p >> \log T$, the sum $\sum_{i=1}^{p} \log|z_i|$ can be of order $-\frac{1}{2}\log T$ only if the majority of $\log|z_i|$ are close to zero. In other words, most of the zeros z_i must be close to 1 by absolute value.

Our analysis builds on three results in particular. First, and for the econometric side, Saikkonen and Lütkepohl (1996) have analyzed the asymptotic properties of VAR estimates, when both the sample size T and the order p of the fitted VAR tend to infinity. Adopting their proofs allows us to derive helpful asymptotic properties in our context. Second and third, and for the algebraic side, we make use of theorems by Erdös and Turan (1950) and Hughes and Nikeghbali (2008), who have provided bounds for the number of roots of a polynomial lying in a segment and in an annulus of the complex

plane, respectively. The hard work in proving the main result then consists in "translating" the Saikkonen-Lütkepohl-inspired asymptotic results into the conditions needed for Erdös and Turan (1950) and Hughes and Nikeghbali (2008).

After obtaining our main result for $p, T \to \infty$, we consider the case of fixed p. For such a case, we focus on the roots of an AR(p) fitted to T observations of a univariate stationary AR(∞) process. Letting $T \to \infty$, we approximate the joint distribution of the coefficients of the fitted AR(p) by the asymptotic Gaussian distribution given in Bhansali (1981). Then, treating the approximation as the true distribution of the coefficients of the AR(p), we derive an analytic formula for the density of the expected empirical distribution of the corresponding characteristic roots. Our derivation is based on the results of Hammersley (1956). A numerical analysis shows that the derived asymptotic approximation of the density of the expected empirical distribution of the characteristic roots is accurate and is concentrated in a circular region in the complex plane for various data generating processes and sample sizes. For example, for AR(8) fitted to 500 observations of a white noise process, the characteristic roots are expected to lie in a halo-like region located around a circle of radius 0.7 in the complex plane.

The asymptotic joint distribution of the estimated roots of an AR(p) with $p = 2$ has been fully characterized in Pantula and Fuller (1993) and Nielsen and Nielsen (2008). In cases of fixed $p > 2$, the asymptotic joint distribution of the roots and the corresponding empirical distribution of the roots are so

complex that their direct analysis seems prohibitively difficult. Therefore, our fixed p analysis focuses on the expected empirical distribution of the roots as opposed to the empirical distribution of the roots. The obtained expected distribution provides us with an asymptotic approximation to the expected number of the roots inside any Borel measurable set in the complex plane.

2 The main result

Following Saikkonen and Lütkepohl (1996), we consider an n-dimensional process $y_t = (y_{1t}', y_{2t}')'$ such that its n_1-dimensional component y_{1t} and n_2-dimensional component y_{2t} satisfy:

$$\begin{aligned} y_{1t} &= C_1 y_{2t} + u_{1t}, \\ \Delta y_{2t} &= u_{2t}, \end{aligned} \tag{1}$$

where $n > 0, n_1 \geq 0, n_2 \geq 0$, and where $u_t = (u_{1t}', u_{2t}')'$ is a zero mean strictly stationary process.

Note that the triangular error correction model form of (1) is:

$$\Delta y_t = - \begin{bmatrix} I_{n_1} & -C_1 \\ 0 & 0 \end{bmatrix} y_{t-1} + v_t,$$

where $v_t = \begin{bmatrix} I_{n_1} & C_1 \\ 0 & I_{n_2} \end{bmatrix} u_t$. We assume that the process v_t (and hence also u_t) has a VAR(∞) representation:

$$\sum\nolimits_{j=0}^{\infty} G_j v_{t-j} = \varepsilon_t,\ G_0 = I_n. \tag{2}$$

Here $\{..., \varepsilon_{-1}, \varepsilon_0, \varepsilon_1, ...\}$ is a sequence of i.i.d. random $n \times 1$ vectors with mean $E\varepsilon_t = 0$, positive definite covariance matrix Σ_ε and finite fourth moments. Recall the definition of the Frobenius norm for a matrix $\|A\| = \sqrt{\sum_{ij} \mid A_{ij} \mid^2} = \sqrt{\operatorname{tr} AA'}$. We assume that the $n \times n$ coefficient matrices G_j are such that $\sum_{j=1}^{\infty} j \|G_j\| < \infty$ and that $G(z) \equiv I_n + G_1 z + G_2 z^2 + ...$ satisfies $\det G(z) \neq 0$ for $|z| \leq 1$. Note that the above DGP spans a wide range of processes from stationary invertible ARMA, when $n_2 = 0$, to general cointegrated processes.

Let $\hat{A}_1, ..., \hat{A}_p$ be the OLS estimates of the coefficient matrices of a VAR(p) fitted to T observations of y_t. Consider the estimated characteristic polynomial

$$\hat{P}_{p,T}(z) = \det\left(I_n z^p - \sum_{j=1}^{p} \hat{A}_j z^{p-j}\right) \tag{3}$$

Let us denote the number of the roots of $\hat{P}_{p,T}(z)$ that belong to a subset Ω of the complex plane as $N_{p,T}(\Omega)$. For any $0 < \delta < 1$ and $0 \leq \theta < \varphi \leq 2\pi$, let $C_\delta = \left\{z \in \mathbb{C} : 1 - \delta \leq |z| \leq (1-\delta)^{-1}\right\}$ be an annulus in the complex plane that contains the unit circle and let $D_{\theta,\varphi} = \{z \in \mathbb{C} : \theta \leq \operatorname{Arg}(z) \leq \varphi\}$ be a sector in the complex plane. Our result is as follows.

Theorem 1. *Let* $\{y_t\}$ *satisfy (1), and assume that* p *is chosen as a function of* T *so that* $p^3/T \to 0$, $(\log T)/p \to 0$, *and* $\sqrt{T}\left(\|G_p\| + \|G_{p+1}\| + ...\right) \to 0$ *as* $T \to \infty$. *Then, for any* $0 < \delta < 1$ *and any* $0 \leq \theta < \varphi \leq 2\pi$, *as* $T \to \infty$:
i) $\frac{1}{pn} N_{p,T}\left(D_{\theta,\varphi}\right) \xrightarrow{p} \frac{\varphi-\theta}{2\pi}$,
ii) $\frac{1}{pn} N_{p,T}\left(C_{\delta}\right) \xrightarrow{p} 1$.

Figure 1 illustrates the result. It shows the roots of $\hat{P}_{p,T}(z)$ for $T = 100, p = 12$ (100 MC replications) and for $T = 1000, p = 48$ (33 MC replications). The upper panel of the Figure corresponds to y_t which is a univariate white noise, the lower panel of the Figure corresponds to y_t which is a univariate random walk. As T and p become larger, nearly all roots stick to the unit circle in a uniform way for both the white noise and the random walk.

In practice, the order p of the approximating VAR would be chosen by an information criterion such as AIC or BIC. As is well-known (see, for example, section 7.4 of Hannan and Diestler, 1988), the rate of increase of the chosen p as $T \to \infty$ depends on how fast the difference between the smallest mean squared forecast error based on a VAR(r) approximation and the optimal mean squared forecast error decays when $r \to \infty$. The exponential decay would imply $\log(T)$ rate for the chosen p, whereas the polynomial decay would imply a rate faster than a positive power of T for the chosen p.

For example, if the true data generating process is a finite-order stationary VARMA with a non-degenerate MA part, then the difference between the VAR(r)-based and the optimal mean squared forecast errors decays as an exponent of r, and p chosen by AIC and BIC is of order $\log T$. Therefore,

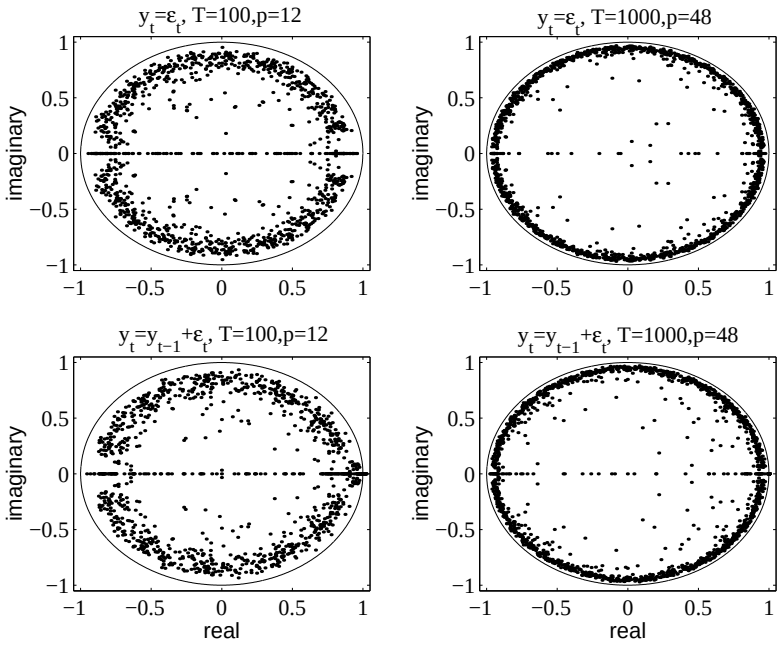

Figure 1: Characteristic roots of AR(p) fitted to T observations of white noise (top row) and a random walk (bottom row). Left panel: 100 MC replications, T=100, p=12. Right panel: 33 MC replications, T=1000, p=48. The pictures show the roots computed in all the replications so there are $12 \cdot 100 = 1,200$ roots in the left panel and $48 \cdot 33 = 1,584$ roots in the right panel.

in such a case, $(\log T)/p$ would not converge to zero as is required by our theorem, and the convergence of the empirical distribution of the roots to the uniform distribution over the unit circle would not take place. In contrast, if the true data generating process has an MA representation with slowly decaying coefficients so that the difference between the VAR(r)-based and the optimal mean squared forecast errors decays as a power of r, then the order p chosen by AIC and BIC would increase faster than a positive power

of T which would guarantee the condition $(\log T)/p \to 0$ of the theorem.

Note that $\hat{P}_{p,T}(z)$ is a polynomial with random coefficients. Shparo and Schur (1962) prove an equivalent of Theorem 1 for polynomials with i.i.d. coefficients ξ_j such that $E \max(0, \log|\xi_j|)^{1+\varepsilon} < \infty$ for some positive ε. Shmerling and Hochberg (2002) prove the theorem for polynomials with independent but not necessarily identically distributed coefficients ξ_j with continuous densities which are uniformly bounded in some neighborhood of the origin with finite means μ_j and standard deviations σ_j that satisfy the condition

$$\sup\left\{\limsup_{j\to\infty}|\mu_j|^{1/j}, \limsup_{j\to\infty}\sigma_j^{1/j}\right\} = 1.$$

Very recently, Hughes and Nikeghbali (2008) prove an equivalent of Theorem 1 for an array of polynomials $\sum_{j=0}^{N}\xi_{jN}z^j$ with not necessarily independent and not necessarily identically distributed random coefficients ξ_{jN} such that

$$E\left\{\log\sum_{j=0}^{N}|\xi_{jN}| - \frac{1}{2}\log|\xi_{0N}| - \frac{1}{2}\log|\xi_{NN}|\right\} = o(N) \text{ as } N\to\infty.$$

They show that sufficient conditions for the above asymptotics to take place is the existence of such $0 < \varepsilon \leq 1$ that $\sum_{j=0}^{N} E(|\xi_{jN}|^{\varepsilon}) = \exp(o(N))$, $E\log|\xi_{0N}| = o(N)$ and $E\log|\xi_{NN}| = o(N)$.

In contrast to the cases considered by Shparo and Schur (1962) and Shmerling and Hochberg (2002), our polynomial $\hat{P}_{p,T}(z)$ does not have inde-

pendent coefficients. Moreover checking the sufficient conditions spelled out by Hughes and Nikeghbali (2008) is a very non-trivial task for the case of $\hat{P}_{p,T}(z)$. We therefore establish our Theorem 1 directly by exploiting inequalities satisfied by the roots of deterministic polynomials which are described in Lemmas 2 and 3 below.

3 Providing a proof.

3.1 Four Lemmas

In order to prove Theorem 1, we need some asymptotic properties of $\hat{A}_1, ..., \hat{A}_p$. To that end, we draw on the analysis of (1), (2) in Saikkonen and Lütkepohl (1996), which needs to be adapted somewhat for our purposes. It is easy to see that y_t has the VAR representation $y_t = A_1 y_{t-1} + ... + A_p y_{t-p} + e_t$, where $e_t = \varepsilon_t - \sum_{j=p}^{\infty} G_j v_{t-j}$,

$$\begin{aligned} A_1 &= H - G_1, \\ A_j &= G_{j-1}H - G_j \text{ for } j = 2, 3, ..., p-1, \\ A_p &= G_{p-1}H, \end{aligned} \tag{4}$$

and $H \equiv \begin{pmatrix} 0 & C_1 \\ 0 & I_{n_2} \end{pmatrix}$.

Lemma 1. *Under the conditions of Theorem 1, we have:*

i) $\left\| \hat{A} - A \right\| = O_p\left(\sqrt{\frac{p}{T}}\right)$, *where* $\hat{A} \equiv [\hat{A}_1, ..., \hat{A}_p]$ and $A \equiv [A_1, ..., A_p]$,

ii) $\Pr\left(\sigma_n\left(\sqrt{T}\left(\hat{A}_p - A_p\right)\right) > \delta_T\right) \to 1$ *for any sequence* δ_T *such that* $\delta_T \to 0$ *as* $T \to \infty$. *Here* $\sigma_j(M)$ *denote the singular values of a matrix* M *in non-increasing order. Hence,* $\sigma_n(M)$ *is the smallest singular value of an* $n \times n$ *matrix* M, *that is the square root of the smallest eigenvalue of* MM'.

A proof of Lemma 1 is given in the Technical Appendix. It uses the same techniques as proofs in Saikkonen and Lütkepohl (1996). These authors have shown that any J linear combinations of $\hat{A} - A$ are asymptotically normal, for arbitrary values of J. With some work, this can be shown to imply the second statement in the Lemma. Furthermore, adapting their strategy delivers the first statement. Note that the length of the vector $\text{vec}\left(\hat{A} - A\right)$ is increasing with the sample size rather than being fixed at some length N. For stationary DGPs, the lemma follows from the proof of Theorem 1 and from Theorem 4 of Lewis and Reinsel (1985).

Additionally, we need the following lemmata:

Lemma 2. *(Erdös and Turan, 1950) Let* a_k, $k = 0, 1, ..., np$, *be arbitrary complex numbers not all of which are equal to zero, and let* $N(\theta, \varphi)$ *denote the number of zeros of* $F_{np}(z) = \sum_{k=0}^{np} a_k z^k$ *that lie in the sector* $0 \leq \theta \leq \arg z \leq \varphi$, *where* $\arg z$ *denotes the angular polar coordinate of the point* z *in the complex plane. Then, for* $a_0 a_{np} \neq 0$:

$$\left| N(\theta, \varphi) - \frac{(\varphi - \theta) np}{2\pi} \right| < 16 \left[np \log \frac{\sum_{k=0}^{np} |a_k|}{|a_0 a_{np}|^{1/2}} \right]^{1/2}. \tag{5}$$

Lemma 3. *(Hughes and Nikeghbali, 2008) Let* a_k, $k = 0, 1, ..., np$, *be arbitrary complex numbers not all of which are equal to zero, and let* $N(\delta)$ *with* $0 < \delta < 1$ *denote the number of zeros of* $F_{np}(z) = \sum_{k=0}^{np} a_k z^k$ *that lie in an annulus* $1 - \delta \leq |z| \leq (1 - \delta)^{-1}$. *Then, for* $a_0 a_{np} \neq 0$:

$$np - N(\delta) \leq 2\delta^{-1} \log \frac{\sum_{k=0}^{np} |a_k|}{|a_0 a_{np}|^{1/2}}. \tag{6}$$

Lemma 4. *Let* U, V *be two* $n \times n$ *matrices. Then*

$$|\det V|^{1/n} \geq \sigma_n(V + U) - \sigma_1(U) \geq \sigma_n(V + U) - \|U\|. \tag{7}$$

Proof. According to a singular value analog of Weyl's inequalities for eigenvalues (see Theorem 3.3.6 in Horn and Johnson, 1991), for any $n \times n$ matrices V and U and for any integers i and j such that $1 \leq i, j \leq n$ and $i + j \leq n + 1$, we have:

$$\sigma_{i+j-1}(V + U) \leq \sigma_i(V) + \sigma_j(U) \tag{8}$$

Inequality (8) implies that $\sigma_n(V + U) \leq \sigma_n(V) + \sigma_1(U)$ and therefore, $\sigma_n(V) \geq \sigma_n(V + U) - \sigma_1(U)$. The latter inequality and the fact that $|\det V| = \prod_{i=1}^{n} \sigma_i(V) \geq [\sigma_n(V)]^n$ implies the first inequality in (7). The second follows directly from $\sigma_1(U) \leq \|U\|$. Q.E.D. ■

3.2 The proof of Theorem 1

With these Lemmata, we are ready to state our proof for Theorem 1.

Proof. Taking $F_{np}(z) \equiv \sum_{k=0}^{np} a_k z^k = \det\left(z^p I_n - \sum_{j=1}^{p} \hat{A}_j z^{p-j}\right)$, we have: $a_0 a_{np} = \det\left(-\hat{A}_p\right)$. Taking $V = \sqrt{T}\hat{A}_p$ and $U = -\sqrt{T}A_p$ in (7) and noting that $\sigma_1(U) = \sigma_1(-U) \leq \|-U\|$, we get

$$\left|\det\left(\sqrt{T}\hat{A}_p\right)\right|^{1/n} \geq \sigma_n\left(\sqrt{T}\left(\hat{A}_p - A_p\right)\right) - \sqrt{T}\|A_p\|$$

The second term in the latter difference converges to zero by the assumption that $\sqrt{T}(\|G_p\| + \|G_{p+1}\| + ...) \to 0$. The first term satisfies Lemma 1ii) with, say, $\delta_T = p^{-1/2} + \sqrt{T}\|A_p\|$. Therefore,

$$\Pr\left(|a_0 a_{np}| > (pT)^{-n/2}\right) \to 1. \tag{9}$$

By definition of the determinant, $F_{np}(z) = \sum_{\tau} (-1)^{|\tau|} P_{1\tau(1)}(z) ... P_{n\tau(n)}(z)$, where the summation is over all permutations of $1, 2, ..., n$ and $P_{ij}(z) \equiv z^p - \hat{A}_{1,ij} z^{p-1} - ... - \hat{A}_{p,ij}$. Such a representation implies that

$$\sum_{k=0}^{pn} |a_k| \leq \sum_{\tau} \prod_{i=1}^{n} \left(1 + \sum_{j=1}^{p} \left|\hat{A}_{j,i\tau(i)}\right|\right) \leq \sum_{\tau} \prod_{i=1}^{n} \left(1 + \sqrt{p}\left\|\hat{A} - A\right\| + \sum_{j=1}^{p} \|A_j\|\right)$$

where the latter inequality uses the fact that for any vector $v = (v_1, ..., v_p)$, $\sum_{j=1}^{p} |v_j| \leq \sqrt{p}\|v\|$. But formulas (4) and the assumption that $\sum_{j=1}^{\infty} j\|G_j\| < \infty$ imply that $\sum_{j=1}^{p} \|A_j\|$ is uniformly bounded and by Lemma 1i) $\sqrt{p}\left\|\hat{A} - A\right\| = p^{-1/2} O_p\left(\sqrt{p^3/T}\right) \leq o_p(1)$. Therefore, there exists a constant M such that $\Pr\left(\sum_{k=0}^{pn} |a_k| \leq M\right) \to 1$. Combining the latter convergence with (9), we ob-

tain: $\Pr\left(\frac{\sum_{k=0}^{np}|a_k|}{|a_0 a_{np}|^{1/2}} < M\left(pT\right)^{n/4}\right) \to 1$. This fact and Lemmas 2 and 3 imply that

$$\Pr\left(\left|\frac{N\left(\theta,\varphi\right)}{np} - \frac{\left(\varphi-\theta\right)}{2\pi}\right| < 16\sqrt{\frac{\log M}{np} + \frac{\log T + \log p}{4p}}\right) \to 1 \text{ and}$$

$$\Pr\left(1 - \frac{N\left(\delta\right)}{np} \leq 2\delta^{-1}\left(\frac{\log M}{np} + \frac{\log T + \log p}{4p}\right)\right) \to 1$$

which proves Theorem 1 because $\left(\log T\right)/p \to 0$ by assumption. Q.E.D. ■

4 Fixed p analysis

Figure 1 suggests that the empirical distribution of the roots of estimated ARs converges to the uniform distribution on the unit circle in a very "regular" way. For relatively small p, the roots plotted on the complex plane form a 'halo' of the 'radius' smaller than one. As p increases, the 'radius' converges to one in accordance with our theorem. Such a suggestion is also supported by additional Monte Carlo exercises (not reported here) and by the empirical analysis of Granger and Jeon (2006). As was briefly mentioned in the Introduction, they observe a 'halo' formed by the roots of the ARs fitted to 215 different US macroeconomic indicators. Granger and Jeon (2006) report that, when determined by AIC criterion, the average number of lags in these ARs is 8.5.

In this section, we provide an analysis of the empirical distribution of the roots when p is fixed and small relative to the number of observations T.

Precisely, we consider univariate ARs fitted to stationary linear processes, and derive an asymptotic approximation to the expected empirical distribution of the corresponding roots as $T \to \infty$ but p remains fixed. Our analysis supports the above speculation that the convergence to the unit circle happens in the expanding-halos-way. It also reveals interesting root clumping regularities internal to each halo.

As is well known, the asymptotic distribution of the estimated autoregressive coefficients is Gaussian. Since the characteristic roots of multiplicity one are locally differentiable functions of the coefficients, we could have used the delta method to find an asymptotic Gaussian distribution for the distinct roots. However, this is not the way we proceed. One reason is that the roots are very non-linear functions of the autoregressive coefficients, and thus, the approximations based on the delta method work poorly in finite samples. Another reason is that we do not want to restrict attention to the roots of multiplicity one. Therefore, our approach is, first, to approximate the distribution of the autoregressive coefficients by the Gaussian distribution based on usual large-T asymptotics; and then, to derive the exact density of the expected empirical distribution of the roots conditional on the autoregressive coefficients having such a Gaussian distribution.

To perform such an analysis we need two additional Lemmas. Consider a polynomial

$$P(z) = c_0 + c_1 z + c_2 z^2 + ... + c_{p-1} z^{p-1} + z^p \tag{10}$$

with random coefficients. Let Ω denote Borel sets in the complex plane, and let $F_P(\Omega) \equiv \frac{1}{p}\{\text{the number of zeros of } P(z) \text{ in } \Omega\}$ be the empirical distribution of the zeros of polynomial (10). For any matrix M, let $\tilde{M}$ be defined as $\tilde{M} = \begin{pmatrix} \operatorname{Re} M & -\operatorname{Im} M \\ \operatorname{Im} M & \operatorname{Re} M \end{pmatrix}$.

Lemma 5. *Suppose that the coefficients $c_0, ..., c_{p-1}$ of the polynomial (10) are real Gaussian random variables with joint distribution $\mathcal{N}(\mu, \Sigma)$ with $\Sigma > 0$. Suppose further that $p > 1$. Then:*

i) *There exists a function $f(z)$ of complex variable $z = x + iy$, such that for any Borel set Ω in the complex plane which does not intersect with the real line, we have: $E[F_P(\Omega)] = \int_{z\in\Omega} f(z)dxdy$. Such a function is given by the following formula. Let us define two $p+1$-dimensional row vectors z_0 and z_1, a $2p+2$-dimensional column vector γ, a $2p+2$-dimensional square matrix V, and a $2 \times (p+1)$ matrix ξ as:*

$$z_0 = \left(1, z, ..., z^{p-1}, z^p\right), \; z_1 = \left(0, 1, 2z, ..., (p-1)z^{p-2}, pz^{p-1}\right),$$
$$\gamma = (\mu', 1, 0, ..., 0)', \; V = \begin{pmatrix} \Sigma & 0 \\ 0 & 0 \end{pmatrix}, \text{ and}$$
$$\xi = \tilde{z}_1 - \left(\tilde{z}_1 V \tilde{z}_0'\right)\left(\tilde{z}_0 V \tilde{z}_0'\right)^{-1} \tilde{z}_0.$$

Then

$$f(z) = \frac{\exp\left\{-\frac{1}{2}\left(\tilde{z}_0\gamma\right)'\left(\tilde{z}_0 V \tilde{z}_0'\right)^{-1}\left(\tilde{z}_0\gamma\right)\right\}}{2\pi p\sqrt{\det\left\{\tilde{z}_0 V \tilde{z}_0'\right\}}} \operatorname{tr}\left\{\xi\left(V + \gamma\gamma'\right)\xi'\right\}. \qquad (11)$$

ii) *There exists a function $g(x)$ of real variable x, such that for any interval (a,b) of real line, we have: $E\left[F_P\left((a,b)\right)\right] = \int_a^b g(x)dx$. Such a function is given by the following formula:*

$$g(x) = \frac{\sqrt{U_{00}U_{11} - U_{01}^2}\exp\left\{-\frac{1}{2}A_0^2/U_{00}\right\}}{\sqrt{2\pi}pU_{00}}\left[r + \sqrt{\frac{2}{\pi}}\int_r^\infty (t-r)\,e^{-t^2/2}dt\right], \tag{12}$$

where

$$A_0 = E\left[P(x)\right], A_1 = E\left[\frac{dP(x)}{dx}\right], U_{00} = Var\left(P(x)\right),$$
$$U_{01} = Cov\left(P(x), \frac{dP(x)}{dx}\right), U_{11} = Var\left(\frac{dP(x)}{dx}\right), \text{ and}$$
$$r = \frac{|U_{00}A_1 - U_{01}A_0|}{\sqrt{U_{00}\left(U_{00}U_{11} - U_{01}^2\right)}}.$$

Lemma 5 is a straightforward extension of the results of Hammersley (1956). He proves an equivalent of Lemma 5 for polynomials $c_0 + c_1 z + c_2 z^2 + ... + c_{p-1}z^{p-1} + c_p z^p$ with $c_0, c_1, ..., c_{p-1}, c_p$ having a non-degenerate joint normal distribution. In our case, c_p is identically equal to 1. Therefore, $c_0, c_1, ..., c_{p-1}, c_p$ have a degenerate joint normal distribution. We prove the extension of Hammersley's (1956) result to this particular degenerate case in the Technical Appendix.

Lemma 5 can be used together with an asymptotic normality result for the estimated autoregressive coefficients to approximate the expected empirical

distribution of the roots of the estimated characteristic equation for finite p and relatively large T. A useful asymptotic normality result is given by the following lemma. Consider a univariate stationary process having AR(∞) representation

$$y_t + a_1 y_{t-1} + a_2 y_{t-2} + a_3 y_{t-3} + ... = \varepsilon_t \tag{13}$$

such that $\sum_{j=1}^{\infty} |a_j| < \infty$, $a(z) = 1 + \sum_{j=1}^{\infty} a_j z^j$ is bounded away from zero for $|z| \leq 1$, and ε_t are i.i.d. with $E\varepsilon_t = 0$, $E\varepsilon_t^2 = \sigma^2$ and $E\varepsilon_t^4 < \infty$. Denote the $p \times p$ matrix with s,t-th element $Ey_s y_t$ as $R(p)$ and the p-dimensional vector with s-th entry $Ey_0 y_{p+1-s}$ as $r(p)$.

Lemma 6. (Bhansali, 1981) *Suppose that an AR(p) model is fit to T observations of process (13), and let $\hat{a}(p) = (\hat{a}_p, ..., \hat{a}_1)'$ denote the vector of OLS estimates of the autoregressive coefficients. Then, as $T \to \infty$,*

$$\sqrt{T}\left(\hat{a}(p) - \bar{a}(p)\right) \xrightarrow{d} \mathcal{N}(0, \Psi),$$

where $\bar{a}(p) = -R(p)^{-1} r(p)$ and Ψ is a positive definite covariance matrix.

Bhansali (1981) expresses the entries of matrix Ψ in the form of an integral which involves the spectral density of process (*13*) and two other functions which can be computed from the autocovariances and the coefficients of (*13*). To save space, we report the formula for Ψ in the Technical Appendix. For the special case when the AR(∞) representation (*13*) reduces to an AR(m) with $m \leq p$, the formula for Ψ considerably simplifies: $\Psi = \sigma^2 R(p)^{-1}$.

Taking the vector μ and the matrix Σ in Lemma 5 equal to $\bar{a}(p)$ and $\frac{1}{T}\Psi$, respectively, we obtain an approximation to the expected empirical distribution of the roots of the polynomial $z^p+\hat{a}_1 z^{p-1}+...+\hat{a}_p$. Such an approximation is straightforward to analyze numerically.[1] Below, we conduct such an analysis for three special cased of the true data generating process: a white noise, a relatively persistent AR(1) and a moving average of order one (MA(1)).

4.1 Numerical analysis

Case I: white noise. For white noise, we have: $\bar{a}(p) = 0$ and $\Psi = I_p$. Taking $\mu = 0$, $\Sigma = \frac{1}{T}I_p$, $p = 8$ and $T = 500$ and using Lemma 5, we numerically evaluate the density $f(z)$ of the expected empirical distribution of the estimated characteristic roots in the areas of the complex plain not intersecting with the real line, and the density $g(x)$ of that distribution on the real line. We take $p = 8$ and $T = 500$ because 8 was a typical length of the AR fitted to the macroeconomic data in Granger and Jeon's (2006) application and because 500 months is a reasonable time span over which monthly US macroeconomic data would be available.

The upper right panel of Figure 2 reports the plot of $g(x)$. We see that the real roots are expected to lie symmetrically around zero with a typical distance to zero equal to about 0.7. Numerically evaluating the integral $\int_{-\infty}^{\infty} pg(x)dx$, we find that the expected number of the real roots equals 1.31.

[1] We have also checked using Monte Carlo simulations that, for the processes and parameter values considered in the next subsection, such an approximation is very accurate. We do not report our Monte Carlo analysis to save space.

Hence, the expected number of the complex roots equals 8-1.31=6.69.

The upper left panel of Figure 2 reports the contour plot of $f(z)$. The outer contour represented by a relatively fat line is chosen so that the expected number of the roots inside this contour equals 95% of the expected total number of the complex roots. We see that the complex roots are expected to lie in a halo-like region in the complex plane (we show only the upper half of the complex plane because the picture is symmetric around the real axis). Moreover, the roots are more likely to occur at regularly spaced locations with the angular "distance" between the closest locations equal to about $\pi/8$.

The lower left panel of Figure 2 shows the plot of $\psi(r) \equiv \int_0^{2\pi} rf(re^{i\omega})d\omega$. Note that $rf(re^{i\omega})$ is the density of the expected empirical distribution of the complex roots in polar coordinates. Therefore, $\psi(r)$ can be interpreted as the marginal radial density of the expected empirical distribution of the complex roots. We see that the "radius" of the halo where most of the complex roots are expected to lie is close to 0.7. Note that the best AR(8) approximation of white noise is AR(8) with zero coefficients. Hence, as T tends to infinity, the estimated coefficients of AR(8) fitted to white noise data converge to zero and the corresponding roots must converge to zero too. Figure 2 shows that $T = 500$ is far from being enough to "observe" the convergence. We computed the marginal radial densities and the corresponding 5th, 25th, 50th, 75th and 95th quantiles for different values of T from $T = 10^2$ to $T = 10^6$ and plot these quantiles against $\log_{10} T$ in the right lower panel of

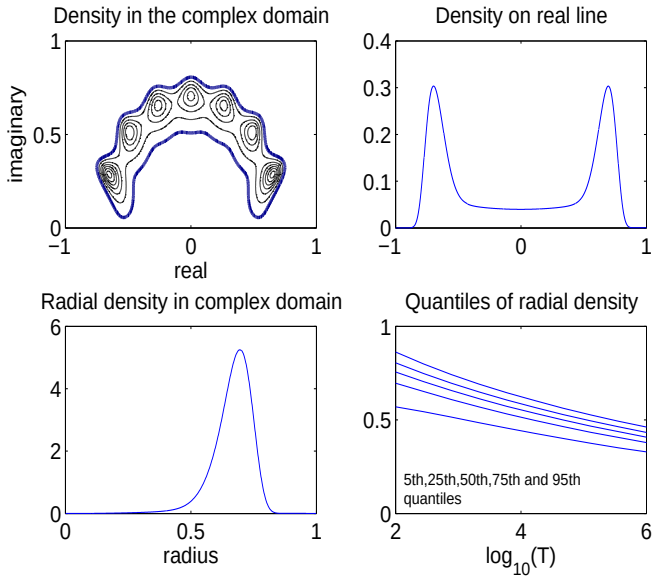

Figure 2: The density of the expected empirical distribution of the characteristic roots of AR(8) fitted to 500 observations of a white noise process.

Figure 2. The estimated complex roots are expected to lie at a substantial distance from zero even when T equals one million!

Case II: AR(1). Let us now consider the true data generating process of the form $y_t = 0.9y_{t-1} + \varepsilon_t$. Figure 3 is an equivalent of Figure 2 for such a process. There are several changes relative to Figure 2. First, the rightmost locus of the concentration of $f(z)$ shifted substantially towards the true 0.9 root on the real line. The remaining loci spread evenly in the angular range which was emptied after the shift. The density on the real line become concentrated around the true root. The modes at around ± 0.7 remain, but their magnitude substantially decreases. Moreover, the symmetry of these

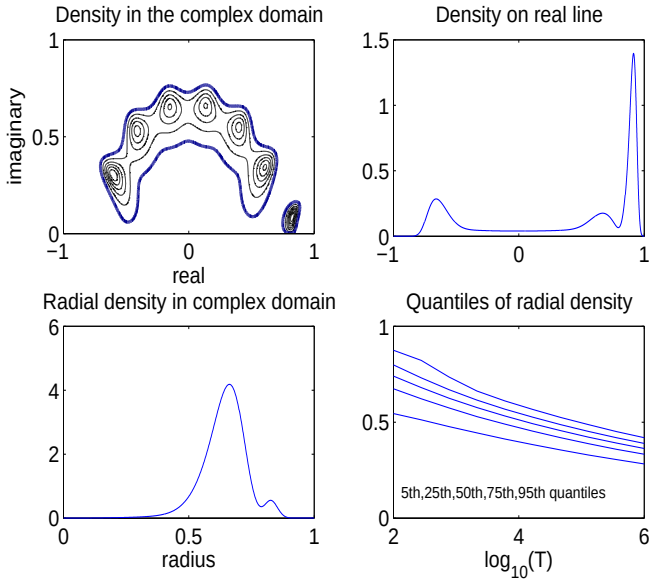

Figure 3: The density of the expected empirical distribution of the characteristic roots of AR(8) fitted to 500 observations of an AR(1) data generating process.

modes is broken: the new 0.9 location "absorbs" more mass from the +0.7 location than from the −0.7 location.

Case III: MA(1). Now we consider the true data generating process of the form $y_t = \varepsilon_t + 0.9\varepsilon_{t-1}$. It has an AR($\infty$) representation $\sum_{j=0}^{\infty} (-0.9)^j \, y_{t-j} = \varepsilon_t$. The corresponding results are presented in Figure 4. As before, the outer fat line contour shows the region in which the expected number of complex zeros equals to the 95% of the total number of complex zeros. However, the remaining contour lines now correspond to much higher levels (not reported) of the density than those on the previous two graphs. Further, instead of seven loci of the density concentration, we have only four. They are located

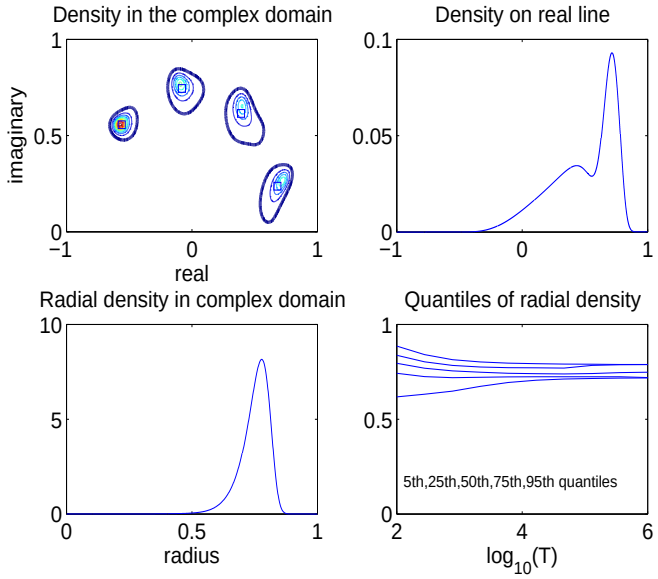

Figure 4: The density of the expected empirical distribution of the characteristic roots of AR(8) fitted to 500 observations of an MA(1) data generating process.

in the proximity of the roots of the best (in terms of minimizing expected squared error of prediction) AR(8) approximation to our MA(1) data generating process. These true roots are shown as squares at the contour plot.

At first sight, the fact that the true roots form a halo-like structure might seem spurious. However, there is a reason for such behavior. Jentzsch (1914) proved that if a deterministic series $1 + a_1 z + a_2 z^2 + a_3 z^3 + ...$ converges for all $|z| < 1$ but does not converge for some $|z| < 1 + \varepsilon$ for any $\varepsilon > 0$, that is, if it has the unit circle as the circle of convergence, then every point of the unit circle is a cluster point of zeros of the partial sums $S_p(z) = 1 + a_1 z + a_2 z^2 + ... + a_p z^p$. Szegö (1922) generalized this result by showing

that there are indices $n_1 < n_2 < ...$ such that the roots of $S_{n_i}(z)$ become uniformly dense in C_δ for arbitrarily small δ. In other words, a theorem similar to our Theorem 1 holds for the roots of the partial sums of a very large class of power series. Clearly, the power series $\sum_{j=0}^{\infty}(-0.9)^j z^j$ has the unit circle as the circle of convergence. Further, the characteristic equation (with z replaced by z^{-1}) for the AR(p) approximating our MA(1) process is close to the partial sum $\sum_{j=0}^{p}(-0.9)^j z^j$, at least for relatively large p. Hence, for relatively large p, we would expect the true zeros of the best AR(p) approximations to our MA(1) to be distributed in a halo-like fashion.

Minnesota prior. Our final numerical exercise is to compute the expected empirical distribution of the roots of AR(8) which corresponds to the Minnesota prior of Litterman (1986) about the auto-regressive coefficients. The prior is frequently used in economic research. It specifies the joint normal distribution for the coefficients of[2] $y_t = a_1 y_{t-1} + ... + a_p y_{t-p} + \varepsilon_t$, so that the different coefficients are independent, the coefficient a_1 has mean one and the standard deviation λ, and the coefficients a_j with $j > 1$ have mean zero and standard deviations λ/j. The usual choice for λ in economic forecasting applications is 0.2, and therefore, we set $\lambda = 0.2$. Figure 5 is an equivalent of Figures 2-4 for the Minnesota prior. The lower right panel is absent because the prior does not depend on T, whereas our approximations to the expected empirical distributions of the roots in the previous exercises do depend on T.

[2]The Minnesota prior was developed for vector auto regressions. In our numerical exercise, we are only interested in the special case of the prior for univariate autoregressions.

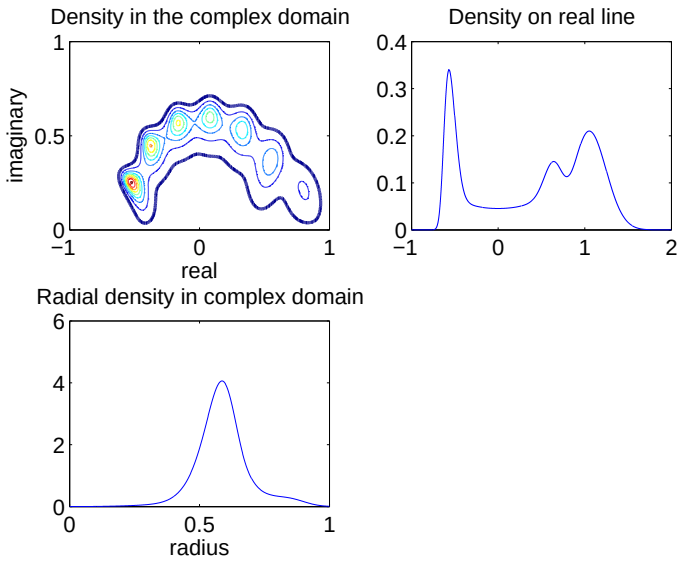

Figure 5: The density of the expected empirical distribution of the characteristic roots of AR(8) with the coefficients distributed according to the Minnesota prior.

Note that a researcher who has the Minnesota prior about the coefficients of AR(8) should expect the complex roots to form the halo-like patterns with "radius" equal to about 0.6. For that researcher, the halo is an a priori reasonable location of the roots of an AR(8) describing economic data. That researcher would therefore have to disagree with Granger and Jeon's (2006) speculation, that the halo-like pattern for the roots of the ARs fitted to the US macroeconomic data is due to the overfitting. Of course, it may well conversely be the case that the Minnesota prior has disagreeable properties. It is perhaps helpful to compare this prior to the standard deviation of the conditional MLE for an AR(8) and a sample of length $T+8$, when the DGP is white noise with unit variance: one obtains the same standard deviation for the estimator a_j with $j > 1$, if $T = 25j^2$. For a_8 for example, one may then think of the prior as similar to having observed a sample of $T = 1600$ and having fitted an AR(8) to a white noise process with an estimate of zero for a_8: it therefore appears to express considerable a priori confidence that the coefficient is indeed zero. Nonetheless, this a priori confidence is not sufficient to force the roots to cluster near zero. It may be good to bear this in mind, when applying this prior.

5 Conclusions

We have shown that the empirical distribution of the roots converges to the uniform distribution over the unit circle when both the sample size T and

the order p of a fitted VAR tends to infinity so that $(\log T)/p \to 0$ and $p^3/T \to 0$. In particular, even if the process is a white noise, nearly all roots of the estimated VAR will converge by absolute value to unity. Therefore, caution is recommended when finding a number of roots with absolute values near unity and drawing the conclusion that the data is highly persistent. Our fixed p analysis shows that the estimated roots are expected to form a halo-like pattern on the complex plane even if the true roots are zero and T is extremely large.

We would like to point out that the striking ubiquity of unit roots established by Theorem 1 does not have negative implications for the econometric procedures not directly based on the estimated roots. For example, univariate stationary processes that satisfy the conditions of Theorem 1 would satisfy Berk's (1974) conditions for the consistency and asymptotic normality of the auto-regressive spectral estimates. For another example, the critical coefficient in the "long" augmented Dickey-Fuller regression would not behave in an unusual way because it is related to the characteristic roots of the regression only through their sum. What Theorem 1 does imply, is that inference regarding the presence of unit roots and nonstationarity by looking at the largest roots in an AR can be highly misleading: with enough lags, one is bound to detect many roots near unity, even if the data is white noise. There may be deeper connections to the deterioration of the properties of cointegration tests, when the lag length is high, see Lütkepohl and Saikkonen (1999), and the debate on "power=size" for unrestricted unit root tests,

see Campbell and Perron (1991), that should prove fascinating to explore further.

References

[1] Berk, K. N. (1974) "Consistent autoregressive spectral estimates", Annals of Statistics 2, No3, 489-502.

[2] Bhansali, R.J. (1981) "Effects of Not knowing the Order of an Autoregressive Process on the Mean Squared Error of Prediction - I", Journal of the American Statistical Association 76, pp.588-597.

[3] Campbell, J.Y. and P. Perron (1991) "Pitfalls and Opportunities: What Macroeconomists Should Know About Unit Roots," in O.J. Blanchard and S. Fisher (eds.), NBER Macroeconomics Annual 1991, vol. 6,, MIT Press, 141-220.

[4] Erdös, P. and Turan, P. (1950) "On the distribution of roots of polynomials", Annals of Mathematics 51, 105-119.

[5] Granger, C.W.J. and Y. Jeon (2006) "Dynamics of Model Overfitting Measured in Terms of Autoregressive Roots", Journal of Time Series Analysis 27, 347-365.

[6] Hammersley, J.M. (1956) "The Zeros of Random polynomial", Proceedings of the Third Berkeley Symposium on probability and Statistics, Vol. II, pp.89-111.

[7] Hannan E.J. and M. Diestler (1988) The Statistical Theory of Linear Systems, John Wiley and Sons, New York.

[8] Horn, R.A. and Johnson C.R. (1991) Topics in Matrix Analysis, Cambridge University Press, Cambridge, New York, Port Chester, Melbourne, Sydney.

[9] Hughes, C.P. and A. Nikeghbali (2008) "The zeros of random polynomials cluster uniformly near the unit circle", Compositio Mathematica 144, 734-746.

[10] Jentzsch, R. (1914) Untersuchungen zur Theorie der Folgen analytischer Functionen. Dissertation. Berlin.

[11] Johansen, S. (2003) "The asymptotic variance of the estimated roots in a cointegrated vector autoregressive model", Journal of Time Series Analysis, 24, 663-678.

[12] Lewis, R. and G.C. Reinsel (1985) "Prediction of Multivariate Time Series by Autoregressive Model Fitting", Journal of Multivariate Analysis 16, 393-411.

[13] Litterman, R.B. (1986) "Forecasting with Bayesian Vector Autoregressions: Five Years of Experience", Journal of Business and Economic Statistics 4, 25-38.

[14] Lütkepohl, H. and Saikkonen, P. (1999) "Order selection in testing for the cointegrating rank of a VAR process", in R. F. Engle and H. White (eds), Cointegration, Causality, and Forecasting. A Festschrift in Honour of Clive W.J. Granger, Oxford University Press, Oxford, pp. 168-199.

[15] Müller, Ulrich K. and Mark W. Watson (2008), "Testing Models of Low-Frequency Variability", Econometrica 76 (5), 979-1016.

[16] Nielsen, B. and Nielsen, H. B. (2008) "Properties of Estimated Characteristic Roots", Univ. of Copenhagen Dept. of Economics Discussion Paper No. 08-13.

[17] Pantula, S.G. and W.A. Fuller (1993) "The Large Sample Distribution of the Roots of the Second Order Autoregressive Polynomial", Biometrika 80, pp.919-923.

[18] Saikkonen, P. (1992) "Estimation and Testing of Cointegrated Systems by an Autoregressive Approximation", Econometric Theory 8, 1-27.

[19] Saikkonen, P., and H. Lütkepohl (1996) "Infinite-Order Cointegrated Vector Autoregressive Processes: Estimation and Inference", Econometric Theory 12, 814-844.

[20] Shmerling, E. and K.J. Hochberg (2002) "Asymptotic behavior of roots of random polynomial equations", Proceedings of the American Mathematical Society 130 (9), 2761-70

[21] Shparo, D.I. and M.G. Schur (1962) "On the Distribution of Roots of Random Polynomials", Vestnik Moskovskogo Universiteta, Series 1: Mathematics and Mechanics, no.3, pp. 40-43.

[22] Szegö, G. (1922) "Über die Nullstellen der Polynomen, die in einem Kreise gleichmässig konvergieren", Sitzungsberichte der Berliner Mathematischen Gesellschaft 21, 59-64

Technical Appendix

6 Proof of Lemma 1

Let us reparametrize regression $y_t = A_1 y_{t-1} + ... + A_p y_{t-p} + e_t$ as $\Delta y_t = \sum_{j=1}^{p-1} \Xi_j u_{t-j} + \Xi_{p,1} u_{1,t-p} + \Psi_0 y_{2,t-1} + e_t$. Using equations of model (1), it is straightforward to check that the regressors in the first regression $X_t = \left[y'_{t-1}, ..., y'_{t-p}\right]'$ are linked to the regressors in the second regression $W_t = [u'_{t-1}, ..., u'_{t-p+1}, u'_{1,t-p}, y'_{2,t-1}]'$ by an invertible linear transformation $X_t = VW_t$, where $V = [S', Z']'$, $Z = \left[0_{n_2 \times n_1}, I_{n_2}, 0_{n_2 \times n(p-1)}\right]$,

$$S = \begin{pmatrix} I_{p-1} \otimes R & 0 \\ 0 & [I_{n_1}, -C_1] \end{pmatrix} + \begin{pmatrix} 0 & I_{p-1} \otimes Q \\ 0 & 0 \end{pmatrix}, \tag{14}$$

$$R = \begin{pmatrix} I_{n_1} & -C_1 \\ 0 & I_{n2} \end{pmatrix} \text{ and } Q = \begin{pmatrix} 0 & 0 \\ 0 & -I_{n2} \end{pmatrix}.$$

Let us denote the OLS estimates of the parameters of the second regression as $\hat{\Xi}_j$, $\hat{\Xi}_{p,1}$ and $\hat{\Psi}_0$,. Since the transformation V depends on the unknown parameter C_1, these estimators are infeasible. However, their asymptotic properties are relatively easy to analyze and, once obtained, can be used to derive the asymptotic properties of $\hat{A}_1, ..., \hat{A}_p$.

Let us denote $[\Xi_1, ..., \Xi_{p-1}, \Xi_{p,1}]$ as Ξ and the matrix of the corresponding OLS estimates as $\hat{\Xi}$. We have: $\hat{A} - A = \left(\hat{\Xi} - \Xi\right) S + \left(\hat{\Psi}_0 - \Psi_0\right) Z$. Now, $\Psi_0 =$

0 because both sides of the second regression must be $I(0)$. Further, $\left\|\hat{\Psi}_0\right\| = O_p\left(T^{-1}\right)$ (see equation (A14) in Saikkonen (1992)). Therefore, to establish part i) of Lemma 1, it is enough to show that $\left\|\left(\hat{\Xi} - \Xi\right) S\right\| = O_p\left(\sqrt{p/T}\right)$. Decomposition (14) implies that $\sigma_1(S) \leq \sigma_1(R) + \sigma_1([I_{n_1}, -C_1]) + \sigma_1(Q)$, and therefore, $\sigma_1(S)$ remains bounded when p rises. This fact implies that we only need to show that $\left\|\hat{\Xi} - \Xi\right\| = O_p\left(\sqrt{p/T}\right)$.

Let us define $e_{1t} = e_t - \varepsilon_t$, $U_t = [u'_{t-1}, ..., u'_{t-p+1}, u'_{1,t-p}]'$, $\Gamma_u = EU_tU'_t$, $N = T - p$, $\hat{\Gamma}_w = N^{-1}\sum_{t=p+1}^{T} W_tW'_t$, and let us partition the inverse $\hat{\Gamma}_w^{-1} = \left[\hat{\Gamma}_w^{(1)}, \hat{\Gamma}_w^{(2)}\right]$ conformably with $W_t = \left[U'_t, y'_{2,t-1}\right]'$. Consider the following decomposition: $\hat{\Xi} - \Xi = F_0 + F_1 + F_2 + F_3$, where $F_0 = N^{-1}\sum_{t=p+1}^{T} \varepsilon_t U'_t \Gamma_u^{-1}$, $F_1 = N^{-1}\sum_{t=p+1}^{T} e_{1t}U'_t\Gamma_u^{-1}$, $F_2 = N^{-1}\sum_{t=p+1}^{T} e_{1t}W'_t\left(\hat{\Gamma}_w^{(1)} - [\Gamma_u^{-1}, 0]'\right)$ and $F_3 = N^{-1}\sum_{t=p+1}^{T} \varepsilon_t W'_t\left(\hat{\Gamma}_w^{(1)} - [\Gamma_u^{-1}, 0]'\right)$. As in Saikkonen (1992, equation (A9)), it follows that $\|F_2\|, \|F_3\| = O_p\left(p^{3/2}/N\right) = o_p\left(\sqrt{p/T}\right)$. Thus, it remains to show that $\|F_0\|$ and $\|F_1\|$ are $O_p\left(\sqrt{p/T}\right)$.

First, let us establish a useful fact that $c_1 \leq \lambda_{\min}(\Gamma_u) \leq \lambda_{\max}(\Gamma_u) \leq c_2$ for some constants $0 < c_1 \leq c_2 < \infty$. Consider the spectral density matrix for u_t : $f_{uu}(\lambda) = \frac{1}{2\pi} RG(e^{i\lambda})\Sigma_\varepsilon\left(RG(e^{i\lambda})\right)^*$, where, by assumption, Σ_ε is positive definite, R is non-singular and $\det\left(G(e^{i\lambda})\right) \neq 0$ for $\lambda \in [0, 2\pi]$. Since $\det\left(G(e^{i\lambda})\right)$ is a continuous function of $\lambda \in [0, 2\pi]$ and since $\left\|G(e^{i\lambda})\right\| < \infty$ for $\lambda \in [0, 2\pi]$, we have: $\inf_{\lambda\in[0,2\pi]} \sigma_n(f_{uu}(\lambda)) = \gamma_1 > 0$ and $\sup_{\lambda\in[0,2\pi]} \sigma_1(f_{uu}(\lambda)) = \gamma_2 < \infty$. On the other hand, $\lambda_{\min}(\Gamma_u) \geq 2\pi \inf_{\lambda\in[0,2\pi]} \sigma_n(f_{uu}(\lambda))$ and $\lambda_{\max}(\Gamma_u) \leq 2\pi \inf_{\lambda\in[0,2\pi]} \sigma_n(f_{uu}(\lambda))$. Therefore, $c_1 \leq \lambda_{\min}(\Gamma_u) \leq \lambda_{\max}(\Gamma_u) \leq c_2$, where

$c_1 = 2\pi\gamma_1$ and $c_2 = 2\pi\gamma_2$. Now, for $\|F_1\|$, we have:

$$\begin{aligned} E\|F_1\| &= N^{-1}E\left\|\sum_{t=p+1}^{T} e_{1t}U_t'\Gamma_u^{-1}\right\| \leq E\left\|e_{1t}U_t'\Gamma_u^{-1}\right\| \\ &= E\left\|\left(\Gamma_u^{-1}U_t \otimes \Sigma_\varepsilon^{1/2}\right)\left(\Sigma_\varepsilon^{-1/2}e_{1t}\right)\right\| \\ &\leq \left(E\left\|\Gamma_u^{-1}U_t \otimes \Sigma_\varepsilon^{1/2}\right\|^2\right)^{1/2}\left(E\left\|\Sigma_\varepsilon^{-1/2}e_{1t}\right\|^2\right)^{1/2}. \end{aligned}$$

The latter square root is $o\left(T^{-1/2}\right)$ (see (A12) in Saikkonen, 1992). For the former, we have: $E\left\|\Gamma_u^{-1}U_t \otimes \Sigma_\varepsilon^{1/2}\right\|^2 = \operatorname{tr}\Gamma_u^{-1}\operatorname{tr}\Sigma_\varepsilon \leq np\lambda_{\min}^{-1}(\Gamma_u)\operatorname{tr}\Sigma_\varepsilon \leq npc_1^{-1}\operatorname{tr}\Sigma_\varepsilon = O(p)$. Therefore, $E\|F_1\| = O\left(\sqrt{p/T}\right)$ and $\|F_1\| = O_p\left(\sqrt{p/T}\right)$. For $\|F_0\|$, we have:

$$E\|F_0\| = N^{-1}E\left\|\sum_{t=p+1}^{T}\varepsilon_t U_t'\Gamma_u^{-1}\right\| \leq N^{-1}c_1^{-1}\left(E\left\|\sum_{t=p+1}^{T}\varepsilon_t U_t'\right\|^2\right)^{1/2}.$$

But $E\left\|\sum_{t=p+1}^{T}\varepsilon_t U_t'\right\|^2 = \sum_{t=p+1}^{T}E(\varepsilon_t'\varepsilon_t)E(U_t'U_t) \leq pN(\operatorname{tr}\Sigma_\varepsilon)E(u_t'u_t) = O(pN)$. Therefore, $E\|F_0\|$ is $O\left(\sqrt{p/T}\right)$, which implies that $\|F_0\|$ is $O_p\left(\sqrt{p/T}\right)$. Since both $\|F_0\|$ and $\|F_1\|$ are $O_p\left(\sqrt{p/T}\right)$, part i of Lemma 1 holds.

Turning to the proof of part ii), note from (14) that $\hat{A}_p - A_p = \left(\hat{\Xi} - \Xi\right)\Psi$, where Ψ is a matrix whose elements are zero except for the lower $n \times n$ block, which equals to $\Psi_p \equiv \begin{pmatrix} 0 & -I_{n_2} \\ I_{n_1} & -C_1 \end{pmatrix}$. Lemma A.3 of Saikkonen and Lütkepohl (1996) implies that for any sequence of n^2-dimensional vectors l_p, $\sqrt{T}\sigma_p^{-1}l_p'\operatorname{vec}\left(\hat{A}_p - A_p\right) \xrightarrow{d} \mathcal{N}(0,1)$, where $\sigma_p^2 = l_p'\left(\Psi'\Gamma_u^{-1}\Psi \otimes \Sigma_\varepsilon\right)l_p$. By Theorem 4.2.12 of Horn and Johnson (1991), which describes the eigen-

values of a Kronecker product as products of the eigenvalues of the components of the product, $\lambda_{\min}\left(\Psi'\Gamma_u^{-1}\Psi \otimes \Sigma_\varepsilon\right) \geq \lambda_{\min}\left(\Psi'\Gamma_u^{-1}\Psi\right)\lambda_{\min}\left(\Sigma_\varepsilon\right)$ and $\lambda_{\max}\left(\Psi'\Gamma_u^{-1}\Psi \otimes \Sigma_\varepsilon\right) \leq \lambda_{\max}\left(\Psi'\Gamma_u^{-1}\Psi\right)\lambda_{\max}\left(\Sigma_\varepsilon\right)$. On the other hand, $\lambda_{\min}\left(\Psi'\Gamma_u^{-1}\Psi\right) \geq \sigma_n^2\left(\Psi\right)\lambda_{\min}\left(\Gamma_u^{-1}\right) = \sigma_n^2\left(R\right)\lambda_{\max}^{-1}\left(\Gamma_u\right) \geq \sigma_n^2\left(R\right)c_2^{-1}$, where the middle equality follows from the fact that Ψ_p equals an orthogonal matrix times R. Similarly, $\lambda_{\max}\left(\Psi'\Gamma_u^{-1}\Psi\right)\lambda_{\max}\left(\Sigma_\varepsilon\right) \leq \sigma_1^2\left(\Psi\right)\lambda_{\max}\left(\Gamma_u^{-1}\right) = \sigma_1^2\left(R\right)\lambda_{\min}^{-1}\left(\Gamma_u\right) \leq \sigma_1^2\left(R\right)c_1^{-1}$. Summing up, $\sigma_n^2\left(R\right)c_2^{-1} \leq \sigma_p^2 \leq \sigma_1^2\left(R\right)c_1^{-1}$. But such inequalities imply that for any measurable subset Ω of n^2-dimensional Euclidean space such that $\Pr\left(\mathcal{N}\left(0, I_{n^2}\right) \in \Omega\right) \neq 0$, there exist constants $d_1, d_2 > 0$ such that $d_1 < \frac{\Pr\left(\sqrt{T}\operatorname{vec}\left(\hat{A}_p - A_p\right) \in \Omega\right)}{\Pr\left(\mathcal{N}\left(0, I_{n^2}\right) \in \Omega\right)} < d_2$ for large enough p. Now, had the statement ii) of Lemma 1 been false, there would have existed a sequence $\delta_T \to 0$ and $\varepsilon > 0$ such that for any T_0 there exists $T > T_0$ such that $\Pr\left(\sigma_n\left(\sqrt{T}\left(\hat{A}_p - A_p\right)\right) < \delta_T\right) > \varepsilon$. Given some $\delta > 0$, let $\Omega\left(\delta\right)$ be the set of all vectors $w \in \mathbb{R}^{n^2}$ such that the $n \times n$ matrix W defined by $\operatorname{vec} W \equiv w$ satisfies $\sigma_n\left(W\right) < \delta$. Choose δ small enough so that $\Pr\left(\mathcal{N}\left(0, I_{n^2}\right) \in \Omega\left(\delta\right)\right) < \varepsilon/\left(2d_2\right)$. Then, for large enough T, we have $\delta_T < \delta$ and therefore

$$\begin{aligned}
&\Pr\left(\sqrt{T}\sigma_n\left(\hat{A}_p - A_p\right) < \delta_T\right) \leq \Pr\left(\sqrt{T}\sigma_n\left(\hat{A}_p - A_p\right) < \delta\right) \\
&= \Pr\left(\sqrt{T}\operatorname{vec}\left(\hat{A}_p - A_p\right) \in \Omega(\delta)\right) < d_2 \Pr\left(\mathcal{N}\left(0, I_{n^2}\right) \in \Omega\left(\delta\right)\right) < \frac{\varepsilon}{2}.
\end{aligned}$$

We have got a contradiction, and therefore statement ii) of Lemma 1 is true. Q.E.D.

7 Proof of Lemma 5

Hammersley's (1956) Theorem 8.1 establishes an equivalent of Lemma 5 for polynomials $P_c(z) = \sum_{j=0}^{p} c_j z^j \equiv \sum_{j=0}^{p} (a_j + ib_j) z^j$, where $i = \sqrt{-1}$ and the components of $c \equiv (a_0, ..., a_p, b_0, ..., b_p)'$ have a joint distribution with continuous density $w(s)$. His Theorems 9.1 and 9.2 then use an extended Slutzky-Fréchet theorem (Theorem 3.1) to show that results for the real case when $b_0 = ... = b_p = 0$ can be obtained as limits of the corresponding results for the complex case. We need to modify Hammersley's arguments to further allow for $c_p \equiv 1$. Below, we give details of such a modification after sketching those arguments that remain unaltered.

Let Δ be a Borel measurable subset of $\mathbb{R}^2$ and let $D = \{z = x + iy : (x, y) \in \Delta\}$. Consider a mapping $\varphi : \Delta \times \mathbb{R}^{2p} \to \mathbb{R}^{2p+2}$ which sends $s \equiv (x, y, a_1, ..., a_p, b_1, ..., b_p)'$ to $c \equiv (a_0, ..., a_p, b_0, ..., b_p)'$, where a_0 and b_0 are the unique real numbers such that $P_c(x + iy) = 0$. The region $\Delta \times \mathbb{R}^{2p}$ can be represented as a union of regions Ω_j, $j = 1, ..., p$ such that φ is a j-to-1 mapping in Ω_j. Precisely, $s \in \Omega_j$ if and only if $P_{\varphi(s)}(z)$ has exactly j distinct roots in D. Since the Jacobian of φ equals $\frac{\partial a_0}{\partial x}\frac{\partial b_0}{\partial y} - \frac{\partial b_0}{\partial x}\frac{\partial a_0}{\partial y} = \left|P'_{\varphi(s)}(z)\right|^2$, where $z = x + iy$, we can write: $\frac{1}{p}\int_{\Delta\times\mathbb{R}^{2p}} w(\varphi(s)) \left|P'_{\varphi(s)}(z)\right|^2 ds = \frac{1}{p}\sum_{j=1}^{p} j \int_{\varphi(\Omega_j)} w(c)dc$. But the latter sum equals the expected number of zeros of $P_c(z)$ in D. Hence, if $\frac{1}{p}\int_{\Delta\times\mathbb{R}^{2p}} w(\varphi(s)) \left|P'_{\varphi(s)}(z)\right|^2 ds$ can be represented as $\int_{z\in D} f(z)dxdy$, then $f(z)$ is the density of the expected empirical distribution of the roots of $P_c(z)$. Note that $\frac{1}{p}\int_{\Delta\times\mathbb{R}^{2p}} w(\varphi(s)) \left|P'_{\varphi(s)}(z)\right|^2 ds = \int_{\Delta}\left[\frac{1}{p}\int_{P_c(z)=0} w(c) \left|P'_c(z)\right|^2 dc\right] dxdy$.

Denoting the joint density of $q \equiv (\operatorname{Re} P_c(z), \operatorname{Re} P_c'(z), \operatorname{Im} P_c(z), \operatorname{Im} P_c'(z))'$ as $\psi(q)$, we can rewrite the inner integral in the latter double integral as $\frac{1}{p}\int_{q_1=q_3=0} \psi(q)(q_2^2+q_4^2)\,dq_2 dq_4$, which provides us with a formula for $f(z)$. When c is Gaussian with mean γ and covariance V, q is also Gaussian with parameters that are simple functions of γ, V and z. Using this fact and some clever elementary algebra, Hammersley (1956) derives formula (11) from the above integral representation of $f(z)$.

In the above derivation, matrix V must be non-singular because c must have a continuous density, by assumption. However, as pointed out by Hammersley (1956), formula (11) remains valid for singular V as long as $\tilde{z}_0 V \tilde{z}_0'$ remains non-singular. Indeed, let $c(k) \sim \mathcal{N}(\gamma, V_k)$, where V_k are non-singular and such that $\lim_{k\to\infty} V_k = V$. Then, as has been shown above, the expected number of roots of $P_{c(k)}(z)$ in D equals $\int_{z\in D} f_k(z)dxdy$, where $f_k(z)$ is given by (11) with V replaced by V_k. Since $c(k) \xrightarrow{d} c$, the extended Slutzky-Fréchet theorem implies that $\int_{z\in D} f_k(z)dxdy$ converges to the expected number of roots of $P_c(z)$ in D. On the other hand, since $\tilde{z}_0 V_k \tilde{z}_0'$ converges to a non-singular matrix $\tilde{z}_0 V \tilde{z}_0'$, $\lim_{k\to\infty} f_k(z)$ exists and equals $f(z)$. Therefore, the expected number of roots of $P_c(z)$ in D is given by $\int_{z\in D} f(z)dxdy$, where $f(z)$ is as in (11).

Proof of part i) Theorem 9.1 of Hammersley (1956) considers a special case of singular V when $b_0 = ... = b_p = 0$ and shows that $\tilde{z}_0 V \tilde{z}_0'$ remains non-singular as long as $\operatorname{Im} z > 0$. Below, we will show that $\tilde{z}_0 V \tilde{z}_0'$ remains non-singular if, in addition to $b_0 = ... = b_p = 0$, we have: $a_p = 1$, so that

$c_p \equiv 1$. As discussed above, the proof of the non-singularity of $\tilde{z}_0 V \tilde{z}_0'$ will suffice to establish (11) in Lemma 5 i).

Let $z = x + iy$ with $y > 0$ and let $z_0 = (1, z, ..., z^{p-1}, z^p)$ as defined in Lemma 5. Consider a $(p+1) \times (p+1)$ matrix X with the element in the i-th row and j-th column equal to zero if $i > j$ and to $\frac{(j-1)!}{(i-1)!(j-i)!} x^{j-i}$ if $i \leq j$. The Taylor expansion of z_0 around $(1, x, ..., x^{p-1}, x^p)$ yields $z_0 = (1, iy, i^2y^2, ..., i^p y^p)\, X$. Therefore, $\tilde{z}_0 = \begin{pmatrix} y_0 X & -y_1 X \\ y_1 X & y_0 X \end{pmatrix}$, where y_0 and y_1 are the real and the imaginary parts of $(1, iy, i^2y^2, ..., i^p y^p)$.

Now, let X_1 be the sub-matrix of X which consists of the first p columns of X. Then, the above formula for $\tilde{z}_0$ and the definition of V in Lemma 5 imply that $\tilde{z}_0 V \tilde{z}_0' = \begin{pmatrix} y_0 X_1 \Sigma X_1' y_0' & y_0 X_1 \Sigma X_1' y_1' \\ y_1 X_1 \Sigma X_1' y_0' & y_1 X_1 \Sigma X_1' y_1' \end{pmatrix}$. Note that the $p+1$-th row of matrix X_1 consists of zeros. Therefore, $y_0 X_1 = \bar{y}_0 \bar{X}_1$ and $y_1 X_1 = \bar{y}_1 \bar{X}_1$, where $\bar{y}_0$ and $\bar{y}_1$ are the real and the imaginary parts of $(1, iy, i^2y^2, ..., i^{p-1}y^{p-1})$ and $\bar{X}_1$ is the sub-matrix of X_1, which consists of the first p rows of X_1. Hence, we have:

$$\tilde{z}_0 V \tilde{z}_0' = \begin{pmatrix} \bar{y}_0 \bar{X}_1 \Sigma \bar{X}_1' \bar{y}_0' & \bar{y}_0 \bar{X}_1 \Sigma \bar{X}_1' \bar{y}_1' \\ \bar{y}_1 \bar{X}_1 \Sigma \bar{X}_1' \bar{y}_0' & \bar{y}_1 \bar{X}_1 \Sigma \bar{X}_1' \bar{y}_1' \end{pmatrix}. \tag{15}$$

But $\bar{X}_1$ is an upper triangular matrix with all coefficients on the main diagonal equal to one. Therefore $\left|\bar{X}_1\right| = 1$ and $\bar{X}_1 \Sigma \bar{X}_1'$ is positive definite. Besides, since $p > 1$ and $y > 0$, neither $\bar{y}_0$ nor $\bar{y}_1$ is null and there exists a non-singular $p \times p$ matrix Y with the first row equal to $\bar{y}_0$ and the second row

equal to $\bar{y}_1$. It remains to note that, according to (15), $\tilde{z}_0 V \tilde{z}_0'$ is a minor on the main diagonal of the positive-definite matrix $Y \bar{X}_1 \Sigma \bar{X}_1' Y'$ and is therefore positive definite itself. Q.E.D.

Proof of part ii). In Theorem 9.2, Hammersley (1956) studies the case of real z, which corresponds to part ii) of Lemma 5. In such a case, $\tilde{z}_0 V \tilde{z}_0'$ is singular so (11) no longer holds. Hammersley (1956) proposes to derive the density of the expected empirical distribution of the real roots of $P_c(z)$ by evaluating the limit $g(x) \equiv \lim_{\delta \to +0} \int_{-\delta}^{\delta} f_\delta(z) dy$. Here $f_\delta(z)$ is obtained from (11) by replacing each of the zero diagonal elements of V by a small number $\delta^{16/7}$. Hammersley's (1956) derivations on pp. 107-109 show that $g(x)$ satisfies (12) as long as the joint distribution of $P_c(x)$ and $P_c'(x)$ is a non-degenerate normal distribution for each fixed real x. Let us check that this condition is satisfied under the assumptions of Lemma 5. Denote the row vector $(1, x, x^2, ..., x^{p-1})$ as x_0 and the row vector $(0, 1, 2x, ..., (p-1)\, x^{p-2})$ as x_1. Then, the covariance matrix of the vector $(P_c(x), P_c'(x))'$ equals $\begin{pmatrix} x_0' \Sigma x_0 & x_0' \Sigma x_1 \\ x_1' \Sigma x_0 & x_1' \Sigma x_1 \end{pmatrix}$. Since, by assumption, $p > 1$, neither x_0 nor x_1 is null, and there exists a non-singular matrix Q with the first row equal to x_0 and the second row equal to x_1. Hence, the covariance matrix of $(P_c(x), P_c'(x))'$ is a minor on the main diagonal of the positive definite matrix $Q \Sigma Q'$ and is therefore non-degenerate. Q.E.D.

8 Bhansali's formula for Ψ in Lemma 6

Let $f(\lambda) = \frac{1}{2\pi} \sum_{u=-\infty}^{\infty} E y_t y_{t+u} e^{-iu\lambda}$ be the spectral density of the process y_t described in (13). Further, let $R_p^{inv}(i,j)$ denote the i,j-th element of matrix $R(p)^{-1}$, let $A_p(\lambda) = 1 + \sum_{j=1}^{p} a_j z^j$, and let $H_u(\lambda) = \sum_{j=1}^{p} R_p^{inv}(u,j) e^{-ij\lambda}$. Then, the element in the u-th row and the v-th column of matrix Ψ from the statement of Lemma 6 equals:

$$\Psi_{uv} = 4\pi \int_{-\pi}^{\pi} \mathrm{Re}\left[A_p(\lambda)\,\overline{H_{p-u}(\lambda)}\right] \mathrm{Re}\left[A_p(\lambda)\,\overline{H_{p-v}(\lambda)}\right] f^2(\lambda)\, d\lambda.$$

This is a slightly reformulated version of formula (3.10) in Bhansali (1981). The difference with that formula is that our H has indices $p-u$ and $p-v$, whereas Bhansali's H has indices u and v. It is because we are interested in the asymptotic covariance matrix of vector $(\hat{a}_p, \hat{a}_{p-1}, ..., \hat{a}_1)'$, whereas Bhansali (1981) is interested in the asymptotic covariance matrix of $(\hat{a}_1, \hat{a}_2, ..., \hat{a}_p)'$.