Unit Roots in White Noise

Onatski, Alexei and Uhlig, Harald

Dept. of Econ., Columbia University, USA, Dept. of Econ., University of Chicago, USA

8 March 2009

Unit Roots in White Noise

Alexei Onatski and Harald Uhlig*

March 3, 2009

Abstract

We show that the empirical distribution of the roots of the vector auto-regression of order n 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 n tend to infinity so that $(\ln T)/n \to 0$ and $n^3/T \to 0$. In particular, even if the process is a white noise, the roots of the estimated vector auto-regression will converge by absolute value to unity.

*Alexei Onatski, Dept. of Economics, Columbia University, ao2027@columbia.edu 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 large estimated root should indicate persistence. Granger and Jeon (2006) has found that the roots of auto-regressions 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 the 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 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 n of the fitted VAR tend to infinity so that $(\ln T)/n \to 0$ and

$n^3/T \to 0$. 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 roots of the estimated vector auto-regression will converge by absolute value to unity.

Our analysis builds on two 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 n of the fitted VAR tend to infinity. Adopting their proofs allows us to derive helpful asymptotic properties in our context. Second, and for the algebraic side, we make use of a theorem by Erdös and Turan (1950), who have provided a bound for the number of roots of a polynomial lying in a segment of the complex plane. 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 to derive near-unity lower and upper bounds for the absolute value of the roots.

2 The main result

Following Saikkonen and Lütkepohl (1996), we consider a r-dimensional process $y_t = (y'_{1t}, y'_{2t})'$ such that its r_1-dimensional component y_{1t} and r_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 $r > 0, r_1 \geq 0, r_2 \geq 0$, 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:

$$\Delta y_t = - \begin{bmatrix} I_{r_1} & -C_1 \\ 0 & 0 \end{bmatrix} y_{t-1} + v_t,$$

where $v_t = \begin{bmatrix} I_{r_1} & C_1 \\ 0 & I_{r_2} \end{bmatrix} u_t$. We assume that the process v_t (and hence also u_t) has a VAR(∞) representation:

$$\sum_{j=0}^{\infty} G_j v_{t-j} = \varepsilon_t, \ G_0 = I_r. \tag{2}$$

Here $\{..., \varepsilon_{-1}, \varepsilon_0, \varepsilon_1, ...\}$ is a sequence of i.i.d. random $r \times 1$ 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 $r \times r$ coefficient matrices G_j are such that $\sum_{j=1}^{\infty} j \|G_j\| < \infty$ and that $G(z) \equiv I_r + G_1 z + G_2 z^2 + ...$ 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 $r_2 = 0$, to general cointegrated processes.

Let $\hat{A}_1, ..., \hat{A}_n$ be the OLS estimates of the coefficient matrices of a vector auto-regression of n-th order fitted to T observations of y_t. Consider the estimated characteristic polynomial

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

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

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

Figure 1 illustrates the result. It shows the roots of $\hat{P}_{n,T}(z)$ for $T = 100, n = 12$ (100 MC replications) and for $T = 1000, n = 48$ (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 n become larger, the roots stick to the unit circle in a uniform way for both the white noise and the random walk.

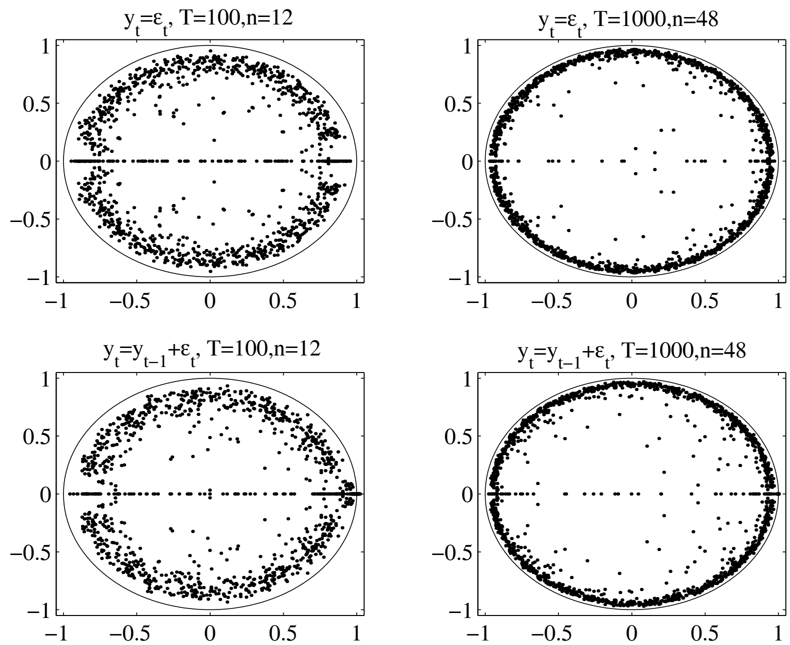

Figure 1: Characteristic roots of VAR(n) fitted to T observations of white noise (top row) and a random walk (bottom row). Left panel: 100 MC replications, T=100, n=12. Right panel: 33 MC replications, T=1000, n=48.

Note that $\hat{P}_{n,T}(z)$ can be interpreted as a polynomial with random coefficients. Shparo and Schur (1962) prove an equivalent of Theorem 1 for polynomials with i.i.d. coefficients under very general assumptions. For a beautiful geometric discussion of the properties of the roots of random polynomials which provides a piece of intuition for the Shparo and Schur's result see Edelman and Kostlan (1995). The contribution of this paper is to extend Shparo and Schur (1962) to $\hat{P}_{n,T}(z)$ whose coefficients are functions of OLS estimates of the auto-regressive parameters, and therefore not i.i.d.

3 Providing a proof.

3.1 Three Lemmas

In order to prove Theorem 1, we need some asymptotic properties of $\hat{A}_1, ..., \hat{A}_n$. 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 $y_t = A_1 y_{t-1} + ... + A_n y_{t-n} + e_t$, where $e_t = \varepsilon_t - \sum_{j=n}^{\infty} G_j v_{t-j}$,

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

and $H \equiv \begin{pmatrix} 0 & C_1 \\ 0 & I_{r_2} \end{pmatrix}$.

Lemma 1. *Under the conditions of Theorem 1, we have:*
i) $\left\| \hat{A} - A \right\| = O_p\left(\sqrt{\frac{n}{T}}\right)$, *where* $\hat{A} \equiv [\hat{A}_1, ..., \hat{A}_n]$ and $A \equiv [A_1, ..., A_n]$,
ii) $\Pr\left(\sigma_r\left(\sqrt{T}\left(\hat{A}_n - A_n\right)\right) > \delta_T\right) \to 1$ *for any sequence* δ_T *such that* $\delta_T \to 0$ *as* $T \to \infty$. *Here* $\sigma_r(M)$ *denotes the* r*-th singular value of a matrix* M*, that is the square root of the* r*-th largest 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 $\mathrm{vec}\left(\hat{A} - A\right)$ is increasing with the sample size rather than being fixed at some length N. For stationary DGP, 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, $k = 0, 1, ..., rn$, be arbitrary complex numbers not all of which are equal to zero, and let $N(\theta, \varphi)$ denote the number of zeros of $F_{rn}(z) = \sum_{k=0}^{rn} a_k z^k$ that lie in the sector $0 \leq \theta \leq \arg z \leq \varphi$. Then, for $a_0 a_{rn} \neq 0$:*

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

Lemma 3. *Let U, V be two $r \times r$ matrices. Then*

$$|\det V|^{1/r} \geq \sigma_r(V + U) - \sigma_1(U) \geq \sigma_r(V + U) - \|U\|. \tag{6}$$

If U and V are nonsingular, then

$$\sigma_r(VU) \geq \sigma_r(V)\, \sigma_r(U). \tag{7}$$

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 $r \times r$ matrices V and U and for any integers i and j such that $1 \leq i, j \leq r$ and $i+j \leq r+1$, we have:

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

$$\sigma_{i+j-1}(VU) \leq \sigma_i(V)\,\sigma_j(U). \tag{9}$$

Inequality (8) implies that $\sigma_r(V+U) \leq \sigma_r(V) + \sigma_1(U)$ and therefore, $\sigma_r(V) \geq \sigma_r(V+U) - \sigma_1(U)$. The latter inequality and the fact that $|\det V| = \prod_{i=1}^{r} \sigma_i(V) \geq [\sigma_r(V)]^r$ implies the first inequality in (6). The second follows directly from $\sigma_1(U) \leq \|U\|$. Noting that $\sigma_r(U) = \sigma_1^{-1}(U^{-1})$ for any non-singular $r \times r$ matrix U and using inequality (9), we get (7) per

$$\sigma_r(VU) = \sigma_1^{-1}\left(U^{-1}V^{-1}\right) \geq \sigma_1^{-1}\left(U^{-1}\right)\sigma_1^{-1}\left(V^{-1}\right) = \sigma_r(V)\,\sigma_r(U) \tag{10}$$

Q.E.D. ■

3.2 The proof of Theorem 1

With these Lemmata, we are ready to state our proof for Theorem 1. It may be useful to provide a road map first.

The key is that the determinant of $\hat{A}_n$ is the product of the roots of the characteristic polynomial (3). The first part of Lemma 3 allows us to bound the determinant of $\hat{A}_n$ from below with the difference of $\sigma_r\left(\hat{A}_n - A_n\right)$ and $\|A_n\|$. While the second term converges to zero faster than $1/\sqrt{T}$, the first term converges slower than $\delta_T/\sqrt{T}$ for any sequence $\delta_T \to 0$ per the first statement of Lemma 1.

We thereby obtain a lower bound for the denominator of the right hand side of (5). Furthermore, the first statement of Lemma 1 provides us with an upper bound of the numerator. Taken together, they turn out to imply the right upper bound in (5) to establish the first part of Theorem 1.

For the second part, we first provide an upper bound and then a lower bound on the roots. The upper bound, i.e. that the roots of the estimated VAR are not too explosive, is plausible intuitively, but requires some work, since the number of VAR lags increases to infinity. Lemma 3 together with $\det G(z) \neq 0$ for $|z| \leq 1$ allows us to bound (3), i.e. the determinant of a lag polynomial, above zero for any complex number exceeding $1 + \delta$, where $\delta > 0$. The first statement of Lemma 1 establishes convergence of the bound for the estimate to the true upper bound.

The more surprising result surely is the lower bound. Suppose, though, that a positive fraction of the roots violates that lower bound. Since the other roots are bounded above by the previous argument, the product of all roots and therefore the determinant of $\hat{A}_n$ can be shown to shrink faster than is allowed by the second statement of Lemma 1, a contradiction.

The sketch above necessarily leaves away some crucial calculations which are needed to show that the bounds work out exactly as desired for Theorem

1 to hold. It is time to provide the details.

Proof.

for *i)* Taking $F_{rn}(z) \equiv \sum_{k=0}^{rn} a_k z^k = \det\left(z^n I_r - \sum_{j=1}^{n} \hat{A}_j z^{n-j}\right)$, we have: $a_0 a_{rn} = \det\left(-\hat{A}_n\right)$. Taking $V = \sqrt{T}\hat{A}_n$ and $U = -\sqrt{T}A_n$ in (6) and noting that $\sigma_1(U) = \sigma_1(-U) \leq \|-U\|$, we get: $\left|\det\left(\sqrt{T}\hat{A}_n\right)\right|^{1/r} \geq \sigma_r\left(\sqrt{T}\left(\hat{A}_n - A_n\right)\right) - \sqrt{T}\|A_n\|$. The second term in the latter difference converges to zero by the assumption that $\sqrt{T}(\|G_n\| + \|G_{n+1}\| + ...) \to 0$. The first term satisfies Lemma 1ii) with, say, $\delta_T = n^{-1/2} + \sqrt{T}\|A_n\|$. Therefore,

$$\Pr\left(|a_0 a_{rn}| > (nT)^{-r/2}\right) \to 1. \tag{11}$$

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

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

where the latter inequality uses the fact that for any vector $v = (v_1, ..., v_n)$, $\sum_{j=1}^{n} |v_j| \leq \sqrt{n}\|v\|$. But formulas (4) and the assumption that $\sum_{j=1}^{\infty} j\|G_j\| < \infty$ imply that $\sum_{j=1}^{n} \|A_j\|$ is uniformly bounded and by Lemma 1i) $\sqrt{n}\left\|\hat{A} - A\right\| = n^{-1/2} O_p\left(\sqrt{n^3/T}\right) \leq o_p(1)$. Therefore, there exists a constant M such that $\Pr\left(\sum_{k=0}^{nr} |a_k| \leq M\right) \to 1$. Combining the latter convergence with (11), we obtain: $\Pr\left(\frac{\sum_{k=0}^{rn} |a_k|}{|a_0 a_{rn}|^{1/2}} < M(nT)^{r/4}\right) \to 1$.

This fact and Lemma 2 imply that

$$\Pr\left(\left|\frac{N(\theta,\varphi)}{rn}-\frac{(\varphi-\theta)}{2\pi}\right|<16\sqrt{\frac{\ln M}{rn}+\frac{\ln T+\ln n}{4n}}\right)\to 1$$

which proves statement i) of Theorem 1 because $\ln T/n\to 0$ by assumption.

for *ii)* Define $Z=\left[z^{-1}I_r,z^{-2}I_r,...,z^{-n}I_r\right]'$. Then $\hat{P}_{n,T}(z)=z^{rn}\det\left(I_r-\hat{A}Z\right)$ and therefore, using (6) with $V=z^n\left(I_r-\hat{A}Z\right)$ and $U=z^n(\hat{A}-A)Z$, we have: $\left|\hat{P}_{n,T}(z)\right|^{1/r}\geq|z|^n\left(\sigma_r(I_r-AZ)-\sigma_1\left((\hat{A}-A)Z\right)\right)$. Using (4), we get:

$$I_r-AZ=\left(\sum_{j=0}^{n-1}G_jz^{-j}\right)(I_r-Hz^{-1}). \tag{12}$$

Let us prove that there exists a constant $c>0$ such that

$$\inf_{|z|>1+\delta}\sigma_r(I_r-AZ)>c \text{ for large enough } n. \tag{13}$$

Inequality (7) implies that

$$\begin{aligned}\sigma_r(I_r-AZ) &= \sigma_r\left(\left(\sum\nolimits_{j=0}^{n-1}G_jz^{-j}\right)(I_r-Hz^{-1})\right)\\ &\geq \sigma_r\left(I_r-Hz^{-1}\right)\sigma_r\left(\sum\nolimits_{j=0}^{n-1}G_jz^{-j}\right)\end{aligned} \tag{14}$$

We see that to establish (13), it is enough to show that there exists $c>0$ such that $\inf_{|z|>1+\delta}\sigma_r\left(I_r-Hz^{-1}\right)>\sqrt{c}$ and $\inf_{|z|>1+\delta}\sigma_r\left(\sum_{j=0}^{n-1}G_jz^{-j}\right)>\sqrt{c}$ for large enough n.

For $\sigma_r\left(I_r - Hz^{-1}\right)$, consider a decomposition $H = RQ$, where

$$R \equiv \begin{pmatrix} I_{r_1} & C_1 \\ 0 & I_{r_2} \end{pmatrix} \text{ and } Q \equiv \begin{pmatrix} 0 & 0 \\ 0 & I_{r_2} \end{pmatrix}.$$

Note that $I_r - Hz^{-1} = R\left(I_r - Qz^{-1}\right)R^{-1}$ because $QR^{-1} = Q$. Using inequality (7) twice, we get:

$$\begin{aligned} \sigma_r\left(I_r - Hz^{-1}\right) &\geq \sigma_r(R)\,\sigma_r\left[\left(I_r - Qz^{-1}\right)R^{-1}\right] \\ &\geq \sigma_r(R)\,\sigma_r\left(I_r - Qz^{-1}\right)\sigma_r\left(R^{-1}\right). \end{aligned} \tag{15}$$

But by definition of Q, $\sigma_r\left(I_r - Qz^{-1}\right) = \min\left\{1, \left|1 - z^{-1}\right|\right\} \geq 1 - |z|^{-1}$ and thus, $\inf_{|z|>1+\delta}\sigma_r\left(I_r - Qz^{-1}\right) > 1 - (1+\delta)^{-1}$. Since R is a fixed non-singular matrix, the latter inequality and (15) imply that there exists $c_1 > 0$ such that $\inf_{|z|>1+\delta}\sigma_r\left(I_r - Hz^{-1}\right) > c_1$.

As to $\sigma_r\left(\sum_{j=0}^{n-1} G_j z^{-j}\right)$, note that, since $\sum_{j=0}^{n-1} G_j z^{-j}$ converges to $G(z^{-1})$ in Frobenius norm uniformly on $|z| \geq 1$, we have:

$$\inf_{|z|\geq 1+\delta}\sigma_r\left(\sum_{j=0}^{n-1} G_j z^{-j}\right) \geq c_1 \inf_{|z|\geq 1+\delta}\sigma_r\left(G(z^{-1})\right) \tag{16}$$

for some $c_1 > 0$ and large enough n. On the other hand, since $\det G(z^{-1}) = \prod_{i=1}^{r}\sigma_i\left(G(z^{-1})\right)$ and since $\sigma_i\left(G(z^{-1})\right) \leq \left\|G(z^{-1})\right\|$, we have:

$$\sigma_r\left(G(z^{-1})\right) = \frac{\det G(z^{-1})}{\prod_{i=1}^{r-1}\sigma_i\left(G(z^{-1})\right)} \geq \frac{\det G(z^{-1})}{\left\|G(z^{-1})\right\|^{r-1}}. \tag{17}$$

But by assumption, $\sum_{j=1}^{\infty}\|G_j\|<\infty$. Therefore, $\sup_{|z|\geq 1}\|G(z^{-1})\|^{r-1}<\infty$. Finally, since $\det G\left(z^{-1}\right)\neq 0$ is a continuous function of z for $|z|\geq 1$, we have: $\inf_{|z|\geq 1}\det G(z^{-1})>0$. Hence, by (17), there exists $c_2>0$ such that $\inf_{|z|\geq 1}\sigma_r\left(G(z^{-1})\right)>c_2$. This fact together with (16) implies that $\inf_{|z|>1+\delta}\sigma_r\left(\sum_{j=0}^{n-1}G_jz^{-j}\right)\geq c_1c_2$ for large enough n, which completes the proof of (13).

Now, using (9) and the fact that for any matrix U, $\sigma_1\left(U\right)\leq\|U\|$, we obtain: $\sigma_1\left((\hat{A}-A)Z\right)\leq\left\|\hat{A}-A\right\|\sigma_1\left(Z\right)=\left\|\hat{A}-A\right\|\sqrt{\frac{1-|z|^{-2n}}{|z|^2-1}}\leq\left\|\hat{A}-A\right\|\sqrt{\frac{1}{2\delta}}=o_p\left(1\right)$ uniformly over $|z|>1+\delta$. Summing up, $\min_{|z|>1+\delta}\left|\hat{P}_{n,T}\left(z\right)\right|^{1/r}\geq\min_{|z|>1+\delta}|z|^n\left(c-o_p(1)\right)>0$ with probability arbitrarily close to one for large enough T and n. Hence, for any $\delta>0:$

$$\Pr\left(N_{n,T}\left(B_{1+\delta}\right)=rn\right)\rightarrow 1, \tag{18}$$

where $B_{1+\delta}$ is the ball of radius $1+\delta$ in the complex plane.

It remains to be shown that $\frac{1}{nr}N_{n,T}\left(B_{1-\delta}\right)\xrightarrow{p}0$, or, in other words, that for any $\varepsilon>0$, $\Pr\left(\frac{1}{nr}N_{n,T}\left(B_{1-\delta}\right)<\varepsilon\right)\rightarrow 1$ as $T\rightarrow\infty$. Let us fix an $\varepsilon>0$ and let $\tau>0$ be such that

$$-\ln\left(1+\tau\right)/\ln(1-\delta)=\varepsilon/2. \tag{19}$$

Let $z_1,...,z_{rn}$ be the roots of $\hat{P}_{n,T}\left(z\right)$ so that $\hat{P}_{n,T}\left(z\right)=\prod_{i=1}^{rn}\left(z-z_i\right).$ Note that $\det\left(-\hat{A}_n\right)$ equals $(-1)^{rn}\prod_{i=1}^{rn}z_i$, and therefore, $\left|\det\left(\hat{A}_n\right)\right|=\prod_{i=1}^{rn}|z_i|$. Replacing δ by τ in (18), we see that all of $|z_i|$ are no larger than $1+\tau$ with probability arbitrarily close to one for large enough T. Furthermore, by definition, there are $N_{n,T}(B_{1-\delta})$ of $|z_i|$ which are less than or equal $1-\delta$. Thus,

$$\Pr\left(\left|\det\left(\sqrt{T}\hat{A}_n\right)\right| < T^{r/2}(1-\delta)^{N_{n,T}(B_{1-\delta})}(1+\tau)^{rn}\right) \to 1 \tag{20}$$

Using this convergence and (11), we have: $\Pr\left(n^{-r/2} < T^{r/2}(1-\delta)^{N_{n,T}(B_{1-\delta})}(1+\tau)^{rn}\right) \to 1$. Taking logarithms of the both sides of the latter inequality, rearranging and recalling (19), we get: $\Pr\left(\frac{1}{nr}N_{n,T}(B_{1-\delta}) < \frac{\varepsilon}{2} + \frac{1}{2n}\frac{\ln T + \ln n}{-\ln(1-\delta)}\right) \to 1$ which implies that $\Pr\left(\frac{1}{nr}N_{n,T}(B_{1-\delta}) < \varepsilon\right) \to 1$.

Q.E.D. ■

4 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 n of a fitted VAR tends to infinity so that $(\ln T)/n \to 0$ and $n^3/T \to 0$. In particular, even if the process is a white noise, the roots of the estimated vector auto-regression 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.

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 peculiarly 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 autoregression can be highly misleading: with enough lags, one is bound to detect many roots near unity, even if the data is white noise.

References

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

[2] Edelman, A. and Kostlan, E. (1995) "How many zeros of a random polynomial are real?", Bulletin of the American Mathematical Society 32, pp. 1-37.

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

[4] 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.

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

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

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

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

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

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

[11] 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.

Technical Appendix

5 Proof of Lemma 1

In the proof, we will use both notations of our paper and those of Saikkonen and Lütkepohl (1996), SU in what follows. Note that n here corresponds to $h+1$ there and $T-n$ here corresponds to N there. Our assumptions imply $T/N = T/(T-n) \to 1$.

First, let us prove that

$$\left\|\tilde{\Xi} - \Xi\right\| = O_p\left(\sqrt{\frac{n}{T}}\right), \tag{21}$$

where Ξ and $\tilde{\Xi}$ are defined on p.830 of SU. Equation (A.4) in SU implies

$$\sqrt{\frac{T}{n}}\left\|\tilde{\Xi} - \Xi\right\| \leq \sqrt{\frac{1}{n}}\left(\frac{T}{T-n}T^{-1/2}\left\|\sum\nolimits_{t=n+1}^{T}\varepsilon_t U_t' \Gamma_u^{-1}\right\| + \sqrt{\frac{T}{T-n}}\sum\nolimits_{j=1}^{3}\|F_j\|\right)$$

Since SU show in the proof of Lemma A.3, that $\|F_2\|$ and $\|F_3\|$ are $O_p\left(\sqrt{n^3/T}\right) = o_p(1)$, it remains to show that $\|F_1\|$ and $T^{-1/2}\left\|\sum_{t=n+1}^{T}\varepsilon_t U_t' \Gamma_u^{-1}\right\|$ are both $O_p\left(\sqrt{n}\right)$. As in the proof of Lemma A.3 except for dropping M_h, we have that

$$E\left\|F_1\right\| \leq \left(E\left\|\Gamma_u^{-1}U_t \otimes \Sigma_\varepsilon^{1/2}\right\|^2\right)^{1/2}\left(\left(T-n\right)E\left\|\Sigma_\varepsilon^{-1/2}e_{1t}\right\|^2\right)^{1/2}$$

and the latter square root is $o(1)$. For the former, we have: $E\left\|\Gamma_u^{-1}U_t \otimes \Sigma_\varepsilon^{1/2}\right\|^2 = \operatorname{tr}\Gamma_u^{-1}\operatorname{tr}\Sigma_\varepsilon \leq nr\lambda_{\min}^{-1}\left(\Gamma_u\right)\operatorname{tr}\Sigma_\varepsilon$. Since, as we show below, $\lambda_{\min}\left(\Gamma_u\right) > c$ for some constant $c > 0$, we have: $E\left\|F_1\right\| = O\left(\sqrt{n}\right)$ and $\left\|F_1\right\| = O_p\left(\sqrt{n}\right)$.

The lower bound on $\lambda_{\min}\left(\Gamma_u\right)$ can be established as follows. By definition, $\Gamma_u = EU_tU_t'$, where $U_t' = \left[u_{t-1}', ..., u_{t-n+1}', u_{1,t-n}'\right]$. Let $x' = \left[x_1', ..., x_{n-1}', x_{n,1}'\right]$ be any unit-length vector with r-dimensional subvectors $x_j' = \left[x_{j,1}', x_{j,2}'\right]$, where $x_{j,1}$ is $r_1 \times 1$, $x_{j,2}$ is $r_2 \times 1$ and $x_{n,2} \equiv 0$. We have:

$$\begin{aligned}
x'\Gamma_u x &= \operatorname{Var}\left[\left(\sum\nolimits_{j=1}^n x_j'L^j\right)u_t\right] = \int_0^{2\pi}\left(\sum\nolimits_{j=1}^n x_j'e^{ij\lambda}\right)f_{uu}\left(\lambda\right)\left(\sum\nolimits_{j=1}^n x_je^{-ij\lambda}\right)d\lambda \\
&\geq \inf_{\lambda\in[0,2\pi]}\sigma_r\left(f_{uu}\left(\lambda\right)\right)\int_0^{2\pi}\left\|\sum\nolimits_{j=1}^n x_j'e^{ij\lambda}\right\|^2 d\lambda = 2\pi\inf_{\lambda\in[0,2\pi]}\sigma_r\left(f_{uu}\left(\lambda\right)\right),
\end{aligned}$$

where $f_{uu}\left(\lambda\right)$ is the spectral density matrix for u_t. Therefore, $\lambda_{\min}\left(\Gamma_u\right) > 2\pi\inf_{\lambda\in[0,2\pi]}\sigma_r\left(f_{uu}\left(\lambda\right)\right)$. But $f_{uu}\left(\lambda\right) = \frac{1}{2\pi}R^{-1}G(e^{i\lambda})\Sigma_\varepsilon\left(R^{-1}G(e^{i\lambda})\right)^*$, where, by assumption, Σ_ε is positive definite, R is non-singular, $\det\left(G(e^{i\lambda})\right) \neq 0$ for $\lambda \in [0, 2\pi]$ and $\left\|G(e^{i\lambda})\right\| < \infty$ for $\lambda \in [0, 2\pi]$. Since $\det\left(G(e^{i\lambda})\right)$ is continuous function of $\lambda \in [0, 2\pi]$, we have: $\inf_{\lambda\in[0,2\pi]}\sigma_r\left(f_{uu}\left(\lambda\right)\right) > 0$, and therefore, $\lambda_{\min}\left(\Gamma_u\right) > c$ for some constant $c > 0$.

To establish (21), it remains to show that $T^{-1/2}\left\|\sum_{t=n+1}^T \varepsilon_tU_t'\Gamma_u^{-1}\right\|$ is $O_p\left(\sqrt{n}\right)$. We have: $E\left\|\sum_{t=n+1}^T \varepsilon_tU_t'\Gamma_u^{-1}\right\| \leq c^{-1}E\left\|\sum_{t=n+1}^T \varepsilon_tU_t'\right\|$ and

$E\left\|\sum_{t=n+1}^{T}\varepsilon_t U_t'\right\|^2 = \sum_{t=n+1}^{T} E\varepsilon_t'\varepsilon_t E U_t' U_t \leq n(T-n)\,\mathrm{tr}\,\Sigma_\varepsilon\,\mathrm{tr}\,E u_t u_t'$. Therefore, $E\left\|\sum_{t=n+1}^{T}\varepsilon_t U_t'\Gamma_u^{-1}\right\|$ is $O\left(\sqrt{nT}\right)$ which implies that $T^{-1/2}\left\|\sum_{t=n+1}^{T}\varepsilon_t U_t'\Gamma_u^{-1}\right\|$ is $O_p\left(\sqrt{n}\right)$.

Now, using definitions of Ξ_j, $\tilde{\Xi}_j$ and $\hat{A}_j$, we obtain:

$$
\begin{aligned}
&A_1 = \Xi_1 R^{-1} + I_r && \text{and} \quad \hat{A}_1 = \tilde{\Xi}_1 R^{-1} + I_r + \left[0_{r\times r_1} : \tilde{\Psi}_2 + \tilde{\Psi}_1 C_1\right];\\
&A_j = \Xi_j R^{-1} - \Xi_{j-1} Q && \text{and} \quad \hat{A}_j = \tilde{\Xi}_j R^{-1} - \tilde{\Xi}_{j-1} Q, \text{ for } j = 2,3,...,n-1;\\
&A_n = [\Xi_{n,1} : -\Xi_{n,1}C_1] - \Xi_{n-1}Q && \text{and} \quad \hat{A}_n = \left[\tilde{\Xi}_{n,1} : -\tilde{\Xi}_{n,1}C_1\right] - \tilde{\Xi}_{n-1}Q
\end{aligned}
$$

and therefore, $\hat{A} - A = \left(\tilde{\Xi} - \Xi\right) S_h + [Z_1 : 0_{r\times hr}]$, where

$$
\begin{aligned}
S_h &= \begin{pmatrix} I_h \otimes R^{-1} & 0_{hr\times r} \\ 0_{r_1\times hr} & [I_{r_1}, -C_1] \end{pmatrix} + \begin{pmatrix} 0_{hr\times r} & -I_h \otimes Q \\ 0_{r_1\times r} & 0_{r_1\times hr} \end{pmatrix} \text{ and} \qquad (22)\\
Z_1 &= \left[0_{r\times r_1} : \tilde{\Psi}_2 + \tilde{\Psi}_1 C_1\right].
\end{aligned}
$$

Decomposition (22) implies that $\sigma_1(S_h) \leq \sigma_1(R^{-1}) + \sigma_1([I_{r_1}, -C_1]) + \sigma_1(Q)$, and therefore, $\sigma_1(S_h)$ is bounded. Further, as stated on page 832 of SU, $\|Z_1\| = O_p(T^{-1})$. The latter two facts together with (21) imply that $\left\|\hat{A} - A\right\| = O_p\left(\sqrt{\frac{n}{T}}\right)$ and thus, the first statement of Lemma 1 holds.

Turning to the proof of the second statement, note from (22) that $\hat{A}_n - A_n = \left(\tilde{\Xi} - \Xi\right)\Psi$, where Ψ is an $(hr + r_1)\times r$ matrix with the upper $(hr - r_2)\times r$ block zero and the lower $r \times r$ block equal to $\begin{pmatrix} 0 & -I_{r_2} \\ I_{r_1} & -C_1 \end{pmatrix}$.

Lemma A.3 of SU implies that for any sequence of r^2-dimensional vector l_n, $\sqrt{T}\sigma_n^{-1}l_n' \operatorname{vec}\left(\hat{A}_n - A_n\right) \xrightarrow{d} N(0,1)$, where $\sigma_n^2 = l_n'\left(\Psi'\Gamma_u^{-1}\Psi \otimes \Sigma_\varepsilon\right) l_n$. By Theorem 4.2.12 of Horn and Johnson (1991), which describes the eigenvalues 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, $\lambda_{\min}\left(\Psi'\Gamma_u^{-1}\Psi\right) \geq \sigma_r^2(\Psi)\lambda_{\min}\left(\Gamma_u^{-1}\right) = \sigma_r^2\left(R^{-1}\right)\lambda_{\min}\left(\Gamma_u^{-1}\right) > b_1$ for some $b_1 > 0$, where the middle equality follows from the fact that $\begin{pmatrix} 0 & -I_{r_2} \\ I_{r_1} & -C_1 \end{pmatrix}$ equals an orthogonal matrix times R^{-1}. Similarly, $\lambda_{\max}\left(\Psi'\Gamma_u^{-1}\Psi\right)\lambda_{\max}\left(\Sigma_\varepsilon\right) \leq \sigma_1^2(\Psi)\lambda_{\max}\left(\Gamma_u^{-1}\right) = \sigma_1^2\left(R^{-1}\right)\lambda_{\max}\left(\Gamma_u^{-1}\right) < b_2$ for some $b_2 > 0$. Summing up, $c_1 < \sigma_n^2 < c_2$ for some $c_1, c_2 > 0$. But such inequalities imply that for any measurable subset Ω of r^2-dimensional Euclidean space such that $\Pr\left(N\left(0, I_{r^2}\right) \in \Omega\right) \neq 0$, there exist $d_1, d_2 > 0$ such that $d_1 < \frac{\Pr\left(\sqrt{T}\operatorname{vec}\left(\hat{A}_n - A_n\right) \in \Omega\right)}{\Pr\left(N\left(0, I_{r^2}\right) \in \Omega\right)} < d_2$ for large enough n.

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 $\Pr\left(\sigma_r\left(\sqrt{T}\left(\hat{A}_n - A_n\right)\right) < \delta_T\right) > \varepsilon$. Given some $\delta > 0$, let $\Omega(\delta)$ be the subset of r^2-dimensional Euclidean space such that for any vector $w \in \Omega(\delta)$, the $r \times r$ matrix W defined by $\operatorname{vec} W \equiv w$ satisfies $\sigma_r(W) < \delta$. Choose δ small enough so that $\Pr\left(N\left(0, I_{r^2}\right) \in \Omega(\delta)\right) < \frac{\varepsilon}{2d_2}$. Then, for large enough T, we have $\delta_T < \delta$ and therefore $\Pr\left(\sqrt{T}\sigma_r\left(\hat{A}_n - A_n\right) < \delta_T\right) \leq \Pr\left(\sqrt{T}\sigma_r\left(\hat{A}_n - A_n\right) < \delta\right) = \Pr\left(\sqrt{T}\operatorname{vec}\left(\hat{A}_n - A_n\right) \in \Omega(\delta)\right) < d_2 \Pr\left(N\left(0, I_{r^2}\right) \in \Omega(\delta)\right) < \frac{\varepsilon}{2}$. We have got a contradiction, and therefore statement ii) of Lemma 1 is true. Q.E.D.