Unit Roots in White Noise?!

A.Onatski and H. Uhlig

September 23, 2008

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 $n^3/T \to 0$.

Granger and Jeon (2006) have 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 is due to the over-fitting and suggest a heuristic partial explanation of the phenomenon. In this paper, 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 VAR. In particular, even if the process is a white noise, the roots of the estimated vector autoregression will converge by absolute value to unity.

We consider r-dimensional processes $y_t = (y_{1t}', y_{2t}')'$ 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 $u_t = (u_{1t}', u_{2t}')'$ has a VAR(∞) representation:

$$u_t + H_1 u_{t-1} + H_2 u_{t-2} + ... = \eta_t. \tag{2}$$

Here $\{..., \eta_{-1}, \eta_0, \eta_1, ...\}$ is a sequence of i.i.d. random $r \times 1$ vectors with mean $E\eta_t = 0$, positive definite covariance matrix $E\eta_t\eta_t' = \Sigma_\eta$ and finite fourth moments. We assume that the $r \times r$ coefficient matrices H_j are such that $\sum_{j=1}^{\infty} j \|H_j\| < \infty$, where $\|H_j\|$ is defined as $\sqrt{\operatorname{tr} H_j H_j'}$, and $H(z) \equiv I_r + H_1 z + H_2 z^2 + ...$ satisfies $\det H(z) \neq 0$ for $|z| \leq 1$. Note that the above DGP spans a wide range of processes from stationary invertible ARMA, when the dimensionality of y_{2t} is zero, 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 $\hat{P}_{n,T}(z) = \det\left(I_r z^n - \sum_{j=1}^{n} \hat{A}_j z^{n-j}\right)$.

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 $D_{\theta,\varphi} = \{z \in \mathbb{C} : \theta \leq \text{Arg}(z) \leq \varphi\}$ 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* $\sqrt{T}\left(\|H_n\| + \|H_{n+1}\| + ...\right) \to 0$ *as* $T \to \infty$. *Then, for any* $0 < \delta < 1$ *and any* $0 \leq \theta < \varphi \leq 2\pi$, *as* $T \to \infty$:
i) $\frac{1}{nr} N_{n,T}(D_{\theta,\varphi}) \xrightarrow{p} \frac{\varphi - \theta}{2\pi}$,
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.

Note that $\hat{P}_{n,T}(z)$ can be interpreted as a polynomial with random coefficients. Shparo and Shur (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 Shur's result see Edelman and Kostlan (1995). The contribution of this paper is to extend

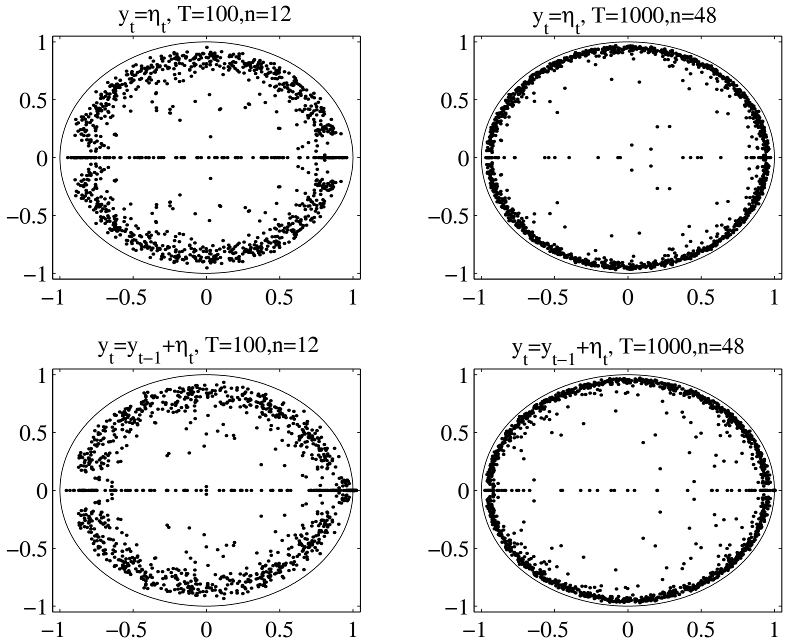

Figure 1: Characteristic roots of VAR(n) fitted to T observations of different DGPs. Left panel: 100 MC replications, right panel: 33 MC replications.

Shparo and Shur (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.

Let us now prove Theorem 1. First, we introduce the following notation. For any matrix M, we denote its j-th singular value, that is the square root of the j-th largest eigenvalue of MM', as $\sigma_j(M)$. We will now describe useful asymptotic properties of $\hat{A} \equiv [\hat{A}_1, ..., \hat{A}_n]$. As shown, for example, in Saikkonen and Lütkepohl (1996), y_t has the following VAR representation:

$y_t = A_1 y_{t-1} + ... + A_n y_{t-n} + e_t$, where

$$\begin{aligned} e_t &= R\left(\eta_t - \textstyle\sum_{j=n}^{\infty} H_j u_{t-j}\right), \\ A_1 &= R\left(Q - H_1\right) R^{-1}, \\ A_j &= R\left(-H_j + H_{j-1}Q\right) R^{-1} \text{ for } j = 2, 3, ..., n-1, \\ A_n &= -RH_{n-1}QR^{-1}, \end{aligned} \tag{3}$$

and $R \equiv \begin{pmatrix} I_{r_1} & C_1 \\ 0 & I_{r_2} \end{pmatrix}$, $Q \equiv \begin{pmatrix} 0 & 0 \\ 0 & I_{r_2} \end{pmatrix}$. We have the following:

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* $T \to \infty$.

A proof of Lemma 1 is available from us upon request. It uses the same techniques as proofs in Saikkonen and Lütkepohl (1996). For stationary DGP, the lemma follows from the proof of Theorem 1 and from Theorem 4 of Lewis and Reinsel (1985).

Now we are ready to prove statement i) of Theorem 1. Our main technical apparatus is the following lemma:

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

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

$$\Pr\left(|a_0 a_{rn}| > (nT)^{-r/2}\right) \to 1. \qquad (4)$$

By definition of the determinant, $F_{rn}\left(z\right) = \sum\limits_{\tau} \left(-1\right)^{|\tau|} P_{1\tau(1)}\left(z\right)...P_{r\tau(r)}\left(z\right)$, where the summation is over all permutations of $1, 2, ..., r$ and $P_{ij}\left(z\right) \equiv z^n - \hat{A}_{1,ij}z^{n-1} - ... - \hat{A}_{n,ij}$. Such a representation implies that $\sum\limits_{k=0}^{nr} |a_k| \leq \sum\limits_{\tau}\prod\limits_{i=1}^{r}\left(1 + \sum\limits_{j=1}^{n}\left|\hat{A}_{j,i\tau(i)}\right|\right) \leq \sum\limits_{\tau}\prod\limits_{i=1}^{r}\left(1 + \sqrt{n}\left\|\hat{A} - A\right\| + \sum\limits_{j=1}^{n}\left\|A_j\right\|\right)$, where the latter inequality uses the fact that for any vector $v = (v_1, ..., v_n)$, $\sum_{j=1}^{n}|v_j| \leq \sqrt{n}\left\|v\right\|$. But formulas (3) and the assumption that $\sum_{j=1}^{\infty} j\left\|H_j\right\| < \infty$ imply that $\sum_{j=1}^{\infty}\left\|A_j\right\| < \infty$, and by Lemma 1i) $\sqrt{n}\left\|\hat{A} - A\right\| = o_p(1)$. Therefore, there exists a constant M such that $\Pr\left(\sum\limits_{k=0}^{nr}|a_k| \leq M\right) \to 1$. Combining the latter convergence with (4), we obtain: $\Pr\left(\frac{\sum_{k=0}^{rn}|a_k|}{|a_0 a_{rn}|^{1/2}} < M\left(nT\right)^{r/4}\right) \to 1$. 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.

Turning to the proof of statement ii), define $Z = [z^{-1}I_r, z^{-2}I_r, ..., z^{-n}I_r]'$. Then $\hat{P}_{n,T}\left(z\right) = z^{rn}\det\left(I_r - AZ - (\hat{A} - A)Z\right)$, and therefore $\left|\hat{P}_{n,T}\left(z\right)\right|^{1/r} \geq$

$|z|^n \left(\sigma_r\left(I_r - AZ\right) - \sigma_1\left((\hat{A}-A)Z\right)\right)$. Using (3), we get: $I_r - AZ = (I_r - Qz^{-1})\left(I_r + \sum_{j=1}^{n-1} H_j z^{-j}\right)$. Note that the second term in the latter product converges to $H(z^{-1})$ uniformly outside the unit circle. Since $\det H(z) \neq 0$ for $|z| \leq 1$ and since $\sigma_r\left(I_r - Qz^{-1}\right) \geq 1 - |z|^{-1}$, there exists a positive constant c such that for any $|z| > 1+\delta$ and large enough T, $\sigma_r\left(I_r - AZ\right) > c$. Further, $\sigma_1\left((\hat{A}-A)Z\right) \leq \left\|\hat{A}-A\right\| \sigma_1(Z) = \left\|\hat{A}-A\right\| \sqrt{r\frac{1-|z|^{-2n}}{|z|^2-1}} \leq \left\|\hat{A}-A\right\| \sqrt{\frac{r}{2\delta}} = o_p(1)$ uniformly over $|z| > 1+\delta$. Summing up, $\min_{|z|>1+\delta} \left|\hat{P}_{n,T}(z)\right|^{1/r} \geq \min_{|z|>1+\delta} |z|^n \left(c - o_p(1)\right) > 0$ with probability arbitrarily close to one for large enough T. Hence, for any $\delta > 0$:

$$\Pr\left(N_{n,T}\left(B_{1+\delta}\right) = rn\right) \to 1, \tag{5}$$

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 $\varepsilon > 0$, $\Pr\left(\frac{1}{nr} N_{n,T}\left(B_{1-\delta}\right) < \varepsilon\right) \to 1$ as $T \to \infty$. Let us fix an $\varepsilon > 0$ and let $\tau > 0$ be such that

$$-\ln\left(1+\tau\right) / \ln(1-\delta) = \varepsilon/2. \tag{6}$$

Let $z_1, ..., z_{rn}$ be the roots of $\hat{P}_{n,T}(z)$ so that $\hat{P}_{n,T}(z) = \prod_{i=1}^{rn}\left(z - z_i\right)$. 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 (5), 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 $N_{n,T}\left(B_{1-\delta}\right)$ of $|z_i|$ which are less than or equal $1-\delta$. Thus, $\Pr\left(\left|\det\left(\sqrt{T}\hat{A}_n\right)\right| < T^{r/2}\left(1-\delta\right)^{N_{n,T}\left(B_{1-\delta}\right)}\left(1+\tau\right)^{rn}\right) \to 1$. Using this convergence and (4), we have: $\Pr\left(n^{-r/2} < T^{r/2}\left(1-\delta\right)^{N_{n,T}\left(B_{1-\delta}\right)}\left(1+\tau\right)^{rn}\right) \to 1$. Taking logarithms of the both sides of the latter inequality, rearranging and recalling (6), we get: $\Pr\left(\frac{1}{nr}N_{n,T}\left(B_{1-\delta}\right) < \frac{\varepsilon}{2} + \frac{1}{2n}\frac{\ln T + \ln n}{-\ln(1-\delta)}\right) \to 1$ which implies that $\Pr\left(\frac{1}{nr}N_{n,T}\left(B_{1-\delta}\right) < \varepsilon\right) \to 1$.

Q.E.D.

In conclusion, 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.

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

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

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