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.

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? 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 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 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.

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:

$$y_{1t} = C_1 y_{2t} + u_{1t},$$
$$\Delta y_{2t} = u_{2t},$$

(1)

where $u_t = (u_{1t}', u_{2t}')'$ has a VAR($\infty$) representation:

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

(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 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 $\sqrt{\text{tr} H_j H_j}$, and $H(z) \equiv I_r + H_1 z + H_2 z^2 + \ldots$ 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 \( \hat{P}_{n,T}(z) = \det \left( I_{r} + \sum_{j=1}^{n} \hat{A}_j z^{n-j} \right) \).

Let us denote the number of the roots of \( \hat{P}_{n,T}(z) \) that belong to a subset \( \Omega \) 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 \( \sqrt{T} (\|H_n\| + \|H_{n+1}\| + ...) \to 0 \) as \( T \to \infty \). Then, for any \( 0 < \delta < 1 \) and any \( 0 \leq \theta < \varphi \leq 2\pi \), as \( T \to \infty \):

i) \( \frac{1}{m} N_{n,T}(D_{\theta, \varphi}) \overset{p}{\to} \frac{\varphi - \theta}{2\pi} \),

ii) \( \frac{1}{m} N_{n,T}(C_{\delta}) \overset{p}{\to} 1 \).

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 co-
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 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.

Let us now prove Theorem 1. First, we will describe useful asymptotic properties of $\hat{A}_1, \ldots, \hat{A}_n$. As shown, for example, in Saikkonen and Lütkepohl (1996), $y_t$ has the following VAR representation: $y_t = A_1 y_{t-1} + \ldots + A_n y_{t-n} + \epsilon_t$, \[ y_t = \eta_t, \quad T=100, n=12 \] \[ y_t = \eta_t, \quad T=1000, n=48 \] \[ y_t = y_{t-1} + \eta_t, \quad T=100, n=12 \] \[ y_t = y_{t-1} + \eta_t, \quad T=1000, n=48 \]
where $e_t = R \left( \eta_t - \sum_{j=n}^{\infty} H_j u_{t-j} \right)$,

\[ A_1 = R (Q - H_1) R^{-1}, \]
\[ A_j = R (-H_j + H_{j-1}Q) R^{-1} \text{ for } j = 2, 3, \ldots, n-1, \]
\[ A_n = -R H_{n-1} Q R^{-1}, \]

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) $\| \hat{A} - A \| = O_p \left( \sqrt{\frac{1}{T}} \right)$, where $\hat{A} \equiv [\hat{A}_1, \ldots, \hat{A}_n]$ and $A \equiv [A_1, \ldots, A_n]$,

ii) $\Pr \left( \sigma_r \left( \sqrt{T} (\hat{A}_n - A_n) \right) > \delta_T \right) \to 1$ for any sequence $\delta_T$ 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 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 $a_k$, $k = 0, 1, \ldots, rn$, 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. \text{ Then, for } a_0 a_{rn} \neq 0: \left| N(\theta, \varphi) - \frac{(0-\varphi)_{rn}}{2\pi} \right| < 16 \left[ \ln \frac{\sum_{k=0}^{rn} |a_k|}{|a_0 a_{rn}|^{1/2}} \right]^{1/2}.

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) , \text{ we have: } a_0 a_{rn} = \det \left( -\hat{A}_n \right) . \text{ Note that } \left| \det \left( \sqrt{T} \hat{A}_n \right) \right|^{1/r} \leq \sigma_T \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} \left( \| H_n \| + \| H_{n+1} \| + \ldots \right) \to 0. \text{ The first term satisfies Lemma 1ii) with, say, } \delta_T = n^{-1/2} + \sqrt{T} \| A_n \|. \text{ Therefore,}

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

By definition of the determinant, \( F_{rn}(z) = \sum_{\tau} (-1)^{|\tau|} P_{\tau(1)}(z) \ldots P_{\tau(r)}(z) \), where the summation is over all permutations of 1, 2, ..., \( r \) and \( P_{ij}(z) \equiv z^n - \hat{A}_1, ij z^{n-1} - \ldots - \hat{A}_n, ij \). \text{ Such a representation implies that } \sum_{k=0}^{nr} |a_k| \leq \sum_{\tau} \prod_{i=1}^{r} \left( 1 + \sum_{j=1}^{n} |\hat{A}_j, i\tau(i)| \right) \leq \sum_{\tau} \prod_{i=1}^{r} \left( 1 + \sqrt{n} \left\| \hat{A} - A \right\| + \sum_{j=1}^{n} \|A_j\| \right) , \text{ 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 \|. \text{ But formulas (3) and the assumption that } \sum_{j=1}^{\infty} j \| H_j \| < \infty \text{ imply that } \sum_{j=1}^{\infty} \| A_j \| < \infty, \text{ and by Lemma 1i) } \sqrt{n} \left\| \hat{A} - A \right\| = o_p(1). \text{ Therefore, there exists a constant } M \text{ such that } \Pr \left( \sum_{k=0}^{nr} |a_k| \leq M \right) \to 1. \text{ Combining the latter convergence with (4), we obtain: } \Pr \left( \sum_{k=0}^{nr} |a_k| < M \left( nT \right)^{r/4} \right) \to 1.

This fact and Lemma 2 imply that
\[
\Pr \left( \left| \frac{N(\theta, \varphi)}{rn} - \frac{(0-\varphi)_{rn}}{2\pi} \right| < 16 \sqrt{\frac{\ln M}{rn} + \frac{\ln T + \ln n}{4rn}} \right) \to 1 \text{ which proves statement i) of Theorem 1 because } \ln T/n \to 0 \text{ by assumption.}
\]

Turning to the proof of statement ii), define \( Z = [z^{-1} I_r, z^{-2} I_r, ..., z^{-n} I_r] \).
Then \( \hat{P}_{n,T}(z) = z^{rn} \det \left( I_r - AZ - (\hat{A} - A)Z \right) \), and therefore \( \left| \hat{P}_{n,T}(z) \right|^{1/r} \geq |z|^n \left( \sigma_r (I_r - AZ) - \sigma_1 \left( (\hat{A} - A)Z \right) \right) \). Using (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 \( \sigma_r (I_r - Qz^{-1}) \geq 1 - |z|^{-1} \), there exists a positive constant \( c \) such that for any \( |z| > 1 + \delta \) and large enough \( T \), \( \sigma_r (I_r - Qz^{-1}) > c \). Further, \( \sigma_1 \left( (\hat{A} - A)Z \right) \leq \left\| \hat{A} - A \right\| \sigma_1 (Z) = \left\| \hat{A} - A \right\| \sqrt{r^{1-|z|^{-2n}}} \leq \left\| \hat{A} - A \right\| \sqrt{\frac{r}{2r}} = o_p(1) \) uniformly over \( |z| > 1 + \delta \). Summing up, \( \min_{|z| > 1 + \delta} |z|^n (c - o_p(1)) > 0 \) with probability arbitrarily close to one for large enough \( T \). Hence, for any \( \delta > 0 \):

\[
\Pr \left( N_{n,T} (B_{1+\delta}) = rn \right) \rightarrow 1,
\]

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

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

\[
-\ln (1 + \tau) / \ln(1 - \delta) = \varepsilon / 2.
\]

Let \( z_1, \ldots, z_{rn} \) be the roots of \( \hat{P}_{n,T}(z) \) so that \( \hat{P}_{n,T}(z) = \prod_{i=1}^{rn} (z - z_i) \). Note that \( \det (-\hat{A}) \) equals \( (-1)^{rn} \prod_{i=1}^{rn} z_i \), and therefore, \( \left| \det (\hat{A}) \right| = \prod_{i=1}^{rn} |z_i| \). Replacing \( \delta \) by \( \tau \) 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 $N_{n,T}(B_1-\delta)$ of $|z_i|$ which are less than or equal to $1-\delta$. Thus, 

\[
\Pr \left( \left| \det \left( \sqrt{T} \hat{A}_n \right) \right| < T^{r/2} (1-\delta)^{N_{n,T}(B_1-\delta)} (1+\tau)^{rn} \right) \to 1.
\]

Using this convergence and (4), 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 (6), 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.

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. 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


