VAR SYSTEM PROPERTIES FROM THE JORDAN DECOMPOSITION; COINTEGRATION

1. The model

Suppose

$$x(t)_{n\times 1} = \sum_{t=1}^{k} B(s)x(t-s) + \varepsilon(t)\,, \tag{1}$$

where $\varepsilon(t)$ is the innovation in the $x(t)$ vector. We can always rewrite (1) as a first-order system in a longer data vector y as follows:

$$y(t) = \begin{bmatrix} x(t) \\ x(t-1) \\ \vdots \\ x(t-k+1) \end{bmatrix} \tag{2}$$

$$y(t) = \begin{bmatrix} B(1) & B(2) & \cdots & \vdots & B(k) \\ & I_{(k-1)\cdot n} & & \vdots & 0 \end{bmatrix} y(t-1) + \begin{bmatrix} \varepsilon(t) \\ 0 \end{bmatrix}\,. \tag{3}$$

We define Γ and $\eta(t)$ by rewriting (3) as

$$y(t) = \Gamma y(t-1) + \eta(t)\,. \tag{4}$$

2. Applying the Jordan decomposition

The Jordan decomposition of Γ is

$$\Gamma = P\Lambda P^{-1} \tag{5}$$

where Λ is diagonal except that there may be along its diagonal "Jordan blocks" of the form

$$\begin{bmatrix} \lambda & 1 & 0 & \dots\dots & 0 \\ 0 & \lambda & 1 & 0 & \dots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \dots & 0 & \lambda & 1 & 0 \\ 0 & \dots\dots & & 0 & \lambda & 1 \\ 0 & \dots\dots\dots & & & 0 & \lambda \end{bmatrix}, \tag{6}$$

Date: January 15, 2009.

i.e. constant down the main diagonal and equal to one on the first diagonal above the main diagonal. (More concisely, Λ is block diagonal with all the diagonal blocks Jordan blocks, though some or all may be trivial 1×1 Jordan blocks.) Any column of P corresponding to the first row of a Jordan block (or to a 1×1 Jordan block) is a right eigenvector of Γ. Corresponding rows of P^{-1} are left eigenvectors.

If we define $z(t) = P^{-1}y(t)$, then (5) implies

$$z(t) = \Lambda z(t-1) + \eta(t) \tag{7}$$

Every subvector $z_i$of z corresponding to a single Jordan block of Γ constitutes a separate subsystem of (7),

$$z_i(t) = \Lambda_i z_i(t-1) + \eta_i(t)\,. \tag{8}$$

In each of these subsystems, we can solve by recursive substitution to obtain

$$z_i(t) = \Lambda_i^t z_i(0) + \sum_{s=0}^{t-1} \Lambda_i^s \eta_i(t-1)\,. \tag{9}$$

For a Jordan block Λ_i with λ_i on the diagonal, Λ_i^p is an upper triangular matrix with λ_i^p on the main diagonal, $p\lambda_i^{p-1}$ on the first diagonal above the main, $p \cdot (p-1)\lambda^{p-2}/2$ on the next diagonal, etc. The general formula is that the q'th diagonal above the main contains

$$\lambda^{p-q}\binom{p}{q}$$

for $q \le p$, 0 for $q > p$.[1] Clearly if $|\lambda_i| < 1$, $\Lambda_i^p \to 0$ as $t \to \infty$. In this case, if η_i satisfies $E[\eta_i(t+1) \mid x(t-s),\ \text{all } s \ge 0] = 0$ for all t and η_i has constant, finite variance, we can let the date of the initial condition in (9) recede into the past and obtain the limiting result

$$z_i(t) = \sum_{s=0}^{\infty} \Lambda_i^s \eta_i(t-1)\,. \tag{10}$$

Of course for this result to hold, the model equations must be thought of as having been in force for indefinitely long into the past.

If all the η_i's are i.i.d. (for example — weaker assumptions would suffice), then $z_i(t)$ clearly has the same distribution for all t. This kind of z_i is called **stationary** or **stable**. If instead $|\lambda_i| = 1$, then the diagonal elements of Λ_i^p remain at one in absolute value for all p, and the above-diagonal elements grow at a polynomial rate. If $|\lambda_i| > 1$, then all the elements of the upper triangle of Λ_i^p explode at least as fast as λ_i^p in absolute value.

[1] The notation $\binom{p}{q}$ stands for the binomial coefficient $p!/(q!(p-q)!)$.

If any λ_i is complex, then (assuming Γ is itself real), λ_i^*, the complex conjugate of λ_i, also appears on the diagonal of Λ, exactly as may times as λ_i itself appears, and the corresponding columns of P and rows of P^{-1} are conjugates of each other. Complex roots λ_i generate oscillatory behavior in the corresponding $z_i(t)$.

But now from the definition of z we know that $y = Pz$, so we know that y is a linear combination of elements of z. Thus we can conclude that

(i) If all the λ_i are less than one in absolute value, y itself, and hence x, is stationary (being a sum of stationary z_i's).
(ii) If at least one of the λ_i's is equal to one in absolute value, and none exceed one in absolute value, the initial condition term in (9), $\Lambda^t z(0)$, contains components that eventually grow in absolute value at the polynomial rate t^{m-1}, where m is the order of the largest Jordan block Λ_i matrix corresponding to one of the unit-absolute-value λ_i's.
(iii) If any of the $|\lambda_i|$'s exceeds one in absolute value, $y(t)$ contains components that explode exponentially as $t \to \infty$.

Often it is useful in interpreting a model to examine the eigenvectors (columns of P and rows of P^{-1}) corresponding to various types of roots. For example, in data including several nominal variables (prices, wages, money stock, current-dollar GDP, etc.) in a country with high and variable inflation, we might expect one unstable root to correspond to the aggregate price level, contributing a non-stationary component to all the nominal variables. The ratios of nominal variables to each other, on the other hand, might be expected to be stationary. This implies that we should find one $|z_i| \geq 1$ and that the corresponding row of P^{-1} should put positive weight on a set of nominal variables. Also, if the variables are all measured in logs, the corresponding column of P should have the same number in every row corresponding to a nominal variable in y at a given lag. This would imply that nominal variables all move proportionately in response to the unstable component.

3. Cointegration

If the largest roots in absolute value are q in number and all equal to one another, and all of them correspond to trivial (1×1) Jordan blocks, then $q \leq n$. Furthermore, in this case there are exactly $n - q$ stationary linear combinations of x (not y). This is the situation known in the literature as **cointegration**. It is handy to know about, but the regularity condition required to deliver it — equality and non-repetition for the largest roots — is much more restrictive than reading the econometric theory literature might lead one to believe. These regularity conditions are widely and casually imposed without asking whether they have any foundation in economic reasoning. Nonetheless we proceed to discuss this case in detail and develop standard results.

We are always free to re-order the columns of P, the rows of P^{-1}, and the blocks on the diagonal of Λ, so long as all three are re-ordered in the same way. Thus we can always choose to have the diagonal elements of Λ sorted in order of decreasing absolute value, and we now assume that this has been done. If just q diagonal elements of Λ are greater than 1 in absolute value, then the first q elements of z are non-stationary, while the last $nk - q$ are stationary. This result is much like the standard co-integration result, but it is not the same thing (and indeed may be more useful). The standard co-integration result, which we derive below, gives conditions under which there are q non-stationary and $n - q$ stationary linear combination of $x(t)$, when there are q elements of absolute value equal to 1 on the diagonal of Λ. The result we have arrived at here shows instead that, under weaker conditions, there are q non-stationary and $nk - q$ stationary linear combinations of $y(t)$ (since the $z(t)$'s are linear combinations of the $y(t)$'s). Notice that since $y(t)$ consists of current and lagged x's, the stationary z's may involve current and lagged x's, not just current x's.

It is immediately clear that, if there are exactly $q < n$ diagonal elements of Λ that equal or exceed a value $\bar{\lambda}$ in absolute value, then there are *at least* $n - q$ linear combinations of $x(t)$ that grow more slowly than $\bar{\lambda}$. This follows because each element of $y(t)$, and hence each element of $x(t)$, is a linear combination of elements of $z(t)$. Since there are only q elements of $z(t)$ that correspond to roots (diagonal elements of Λ) that equal or exceed $\bar{\lambda}$, it must be possible to find $n - q$ linear combinations of $x(t)$ that include no component of these q overly explosive z's. It may be possible to find more stationary linear combinations than this, because it is not necessarily true that all q non-stationary z's receive weight in the linear combinations of z's that form $x(t)$. It is a corollary of the argument given below that if there are any roots that equal or exceed $\bar{\lambda}$ in absolute value, there are no more than $n - 1$ linear combinations of $x(t)$ that grow more slowly than $\bar{\lambda}$.

The eigenvectors of a matrix with the structure of the Γ defined by (3) and (4) have a special form. If we break the nk-length right eigenvector g into k components $g_1, \dots, g_k$ of length n each, then if λ is the corresponding eigenvalue, $g_j = \lambda g_{j+1}$, $j = 1, \dots, k-1$. This is easily verified from the definition of an eigenvector and the structure of the matrix. This means that if there are q distinct roots corresponding to the single largest eigenvalue $\bar{\lambda}$, each must consist of k scaled replicas of a unique n-dimensional vector h_j, $j = 1, \dots, q$. Consider the row vectors of the form $[f, 0, 0, \dots, 0]$, where f is an n-dimensional vector that is orthogonal to all q of the h_j's. Obviously there are $n - q$ of these. The first q rows of the matrix P^{-1} are the unit left-eigenvectors of Γ, while its $nk - q$ row are by construction of full row rank and orthogonal to the first q columns of P. They therefore span the space of nk-dimensional vectors that are orthogonal to the first q columns of P. But the vectors of the form $[f, 0, , \dots, 0]$ that we constructed above are orthogonal to all of the first q columns of P, and therefore lie in the space spanned by the last $nk - q$ rows of P^{-1}. Since these last $nk - q$ rows also define the coefficients of $nk - q$ stationary linear combinations of y. $[f, 0, \dots, 0]y_t = fx_t$ is stationary for each of the $n - q$ f's that are orthogonal to the q h_j's. The f's are called **cointegrating vectors** when $\bar{\lambda} = 1$.

Summarizing our results, we arrive at

Proposition 1. *If the q largest eigenvalues of Γ in* (3) *are all equal and non-repeating, then $q \leq n$ and there are $n - q$ linear combinations of $x(t)$ corresponding to eigenvalues less than $\bar{\lambda}$.*

To find the cointegrating vectors from an estimated VAR, one can form the system matrix Γ and take its eigenvalue decomposition. If there are q large roots, and if they correspond to distinct eigenvectors, determine the corresponding f vectors. They can be read off as the first n coefficients in each of the eigenvectors corresponding the large roots. Find $n - q$ vectors orthogonal to these f's, e.g. by finding $n - q$ linearly independent columns of $I - F(F'F)^{-1}F'$, where F is the $n \times q$ matrix whose columns are the f's. These $n - q$ vectors orthogonal to F are $n - q$ cointegrating vectors. More directly, if R is accessible, form `svdF <- svd(F, nu=n)`. Then the last $n - q$ columns of `svdF$u` span the space orthogonal to F and form the basis of the space of cointegrating vectors. If we call these last $n - q$ columns the matrix `v`, it may be convenient, depending on the application, to normalize to make the vectors more interpretable. The same computation is possible in Matlab or Octave, with the main difference that Matlab.Octave's `svd` function returns the "`nu=n`" version of the svd by default and returns R's default `svd(x)` via the Matlab/Octave command `[u,d,v] = svd(x,0)`.

It is not critical to this analysis that the large eigenvalues all be equal or nearly equal. It is critical that they are all disinct, meaning that each corresponds to a unique eigenvector, and that they are all large (or all small, for that matter) so that the distinction between them and the remaining roots has a clear interpretation.

4. EXAMPLES

(A) The simplest example of a system with q unit roots and more than $n - q$ stationary linear combinations is

$$\begin{aligned}(1-L)^2 y_1(t) &= \varepsilon_1(t)\\ y_2(t) &= \varepsilon_2(t)\,.\end{aligned}$$

There are two (repeating) unit roots in this 2×2 system, and nonetheless 1 stationary linear combination, y_2. Other simple examples can be constructed by taking linear transformations of this one, say

$$z_1(t) = 4z_1(t-1) - 4z_2(t-1) - 2z_1(t-2) + 2z_2(t-1) + \eta_1(t) \tag{11}$$

$$z_2(t) = 2z_1(t-1) - 2z_2(t-1) - z_1(t-2) + z_2(t-2) + \eta_2(t)\,. \tag{12}$$

This system is obtained by letting $z(t) = \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix} y(t)$, and it therefore also has two repeating unit roots and one stationary linear combination, which is here $2z_2(t) - z_1(t)$.

(B) The six-lag VAR in logged monthly ppi components that we estimated in an exercise delivers an 18×18 system matrix. If estimated by `rfvar3` with `mu=0` (i.e. with the prior pulling toward three distinct unit roots turned off, but the default `lambda=5` left in place still pulling toward at least one unit root), we find the largest three eigenvalues as 1.00152, 0.98816, 0.96703. (For this cointegration analysis results are nearly the same with `lambda=0,mu=0`.) The first root is very close to one and implies non-stationarity. The latter two correspond to mean lags ($\rho/(1-\rho)$ of 83 and 29 months, and to half-lives ($-\log 2/\log(\rho)$) of 58 and 21 months. In annual units, the mean lags are about 7 and about 2.5 years, and the half lives about 6 and 1.75 years. Relative to business cycle dynamics, these are far from non-stationary. The first three elements of the eigenpricevector corresponding to the non-stationary root are 0.2427, 0.2420, 0.2247. Note that this implies that the non-stationary component of the process makes the three series move by approximately the same amounts, as would be expected for a general inflationary factor. Taking `v <- c(.2427,.2420,.2247)` and forming `svdv <- svd(v, nu=3)`, we find `svdv$u[ , 2:3]` to be

$$\begin{bmatrix} -0.5904156 + 0i & -0.5483862 + 0i \\ 0.7810614 + 0i & -0.2033533 + 0i \\ -0.2033533 + 0i & 0.8111227 + 0i \end{bmatrix}$$

As expected from the fact that `v` itself is nearly a constant vector, when we normalize this matrix by post-multiplying by the inverse of its upper 2×2 submatrix, we get two co-integrating vectors that are close to representing simply the relative values of two of the pairs of series:

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \\ -1.08 & -1.08 \end{bmatrix} .$$

With the `mu=2` default left in place in `rfvar3`, we get rather different results. All three large roots are pushed closer to one, and if we calculate cointegrating vectors as above they are much farther from being simple (1,-1) vectors. If we needed to use this model, we might then calculate, by posterior simulation, error bands on the cointegration coefficients, or compute posterior odds on the unconstrained VAR vs one constrained to have [1, 0, -1] and [0, 1, -1] cointegrating vectors. The latter model can be estimated by fixing these vectors as the $n \times q$ γ in the VECM form $A(L)\Delta y_t = \alpha\gamma y_{t-1} + \varepsilon_t$, with $A(L)$ fifth order (so the system in levels is sixth order).