VAR System Properties from the Jordan Decomposition; cointegration

Chris Sims

January 15, 2020

The model

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

where $\varepsilon(t)$ is the innovation in the $x(t)$ vector.

Stacking

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

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

The Jordan decomposition

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

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

$$\begin{bmatrix} \lambda & 1 & 0 & \multicolumn{2}{c}{\dots\dots\dots} & 0 \\ 0 & \lambda & 1 & 0 & \dots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \dots & 0 & \lambda & 1 & 0 \\ 0 & \multicolumn{2}{c}{\dots\dots\dots} & 0 & \lambda & 1 \\ 0 & \multicolumn{3}{c}{\dots\dots\dots\dots\dots} & 0 & \lambda \end{bmatrix}, \tag{6}$$

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.

Applying the Jordan decomposition

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

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

Root sizes

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 $q \leq p$, 0 for $q > p$.

$|\lambda_i| < 1 \Rightarrow \Lambda_i^p \to 0$ as $t \to \infty$. In this case, if η_i satisfies $E[\eta_i(t+1) \mid x(t-s)$, all $s \geq 0] = 0$ 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

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

Stability

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.

Complex roots

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 $z_i(t)$.

Implied properties of y

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, 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 $t \to \infty$.

Interpreting eigenvectors

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. This would imply that nominal variables all move proportionately in response to the unstable component.

Cointegration

- If the largest roots in absolute value are q in number and all equal to some $\lambda \geq 1$, and all of them correspond to trivial (1×1) Jordan blocks, then $q \leq n$.
- In this case there are exactly $n-q$ linear combinations of x (not y) that grow slower than λ^t.
- If $\lambda = 1$, these $n-q$ linear combinations are stationary, while the q linear combinations are not. This is the situation known in the literature as **cointegration**.

Skeptical remarks

- Cointegration is handy to know about, but the regularity condition required to deliver it — equality and non-repetition for the largest roots — is restrictive..

- The restrictions are widely and casually imposed without grounding them in any economic reasoning.

- The reason is probably mainly that if one pretends one knows that there are a given number of unit roots that don't repeat, frequentist distribution theory gets much easier.

- In practice, people look for models with q unit roots by experimenting — essentially estimating q.

- Frequentist distribution theory implies that doing this, and ignoring the estimation uncertainty in q, is “asymptotically justified”, because estimates of unit roots converge much faster than estimates of stationary roots.

- In practice, with real data, there are usually roots that are substantially different from one that are also statistically insignificantly different from one — so uncertainty about q is important

- Estimating q and ignoring uncertainty about it is better than differencing everything that looks non-stationary — the latter amounts to assuming $q = n$.

VECM models

- If there are exactly $q < n$ non-repeating unit roots, then the VAR can be written as
$$\Delta y_t = \underset{n\times(n-q)}{\alpha}\,\underset{(n-q)\times n}{\beta}\, y_{t-1} + \sum_{j=1}^{k-1} B_j \Delta y_{t-j}\,.$$
- The linear combinations given by βy_t will be stationary. Other linear combinations will be non-stationary.
- Estimating such a system for given q is fairly straightforward NLLS.

A more general "cointegration-like" result

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 $\bar{\lambda}$.

Suppose that it were possible to form an "LU" decomposition of P, i.e. a decomposition of the form $P = VU$, with V lower triangular with ones on the diagonal and U upper triangular. In that case we could write

$$y(t) = Pz(t) = VUz(t)\,, \qquad\qquad \therefore \quad V^{-1}y(t) = Uz(t)\,. \tag{11}$$

Because of the triangularity of V and U, this expression implies that only the first q elements of $y(t)$ (and hence the first q elements of $x(t)$, assuming $q < n$) put any weight on the first q elements of $z(t)$, which are the q "excessively explosive" components of the system. All the remaining elements of $V^{-1}y(t)$, and hence in particular the $q+1$'st through n'th, are stationary. But from the lower triangularity of V (which is preserved under inversion), we know that these $n-q$ linear combinations of $y(t)$ are actually just linear combinations of $x(t)$. This brings us to the conclusion that there are exactly $n-q$ linear combinations of $x(t)$ that explode more slowly than $\bar{\lambda}$.

This is not a proof that there are generally $n-q$ non-explosive linear combinations of $x(t)$, however, because even for a non-singular P, it is not always possible to calculate an LU decomposition without what as known as "pivoting". The condition that allows an LU decomposition without pivoting is that all the $j \times j$ matrices formed from the upper left submatrix of P, $j = 1, \ldots, nk$, be non-singular.[1] What is always possible is to find permutation matrices[2] Q and M such that QPM has an LU decomposition. To obtain the result that there are exactly $n-q$ stationary linear combinations, we need to assure ourselves that, having ordered P and Λ so that the unstable

[1] See the description of the LU decomposition in Golub and van Loan, *Matrix Computations* for a discussion of pivoting and how it relates to nonsingularity of the diagonal submatrices.

[2] A permutation matrix Q has all its elements 0 or 1 and satisfies $Q'Q = I$. Multiplication of a matrix by a permutation matrix simply reorders its rows or columns.

roots are in the upper left, we can choose Q lower block triangular so that its first n rows are zero except in the first n columns, and choose M upper block triangular so that its last $nk-q$ rows are zero in their first q columns.[3] This means the first n elements of $Qy(t)$ are still linear combinations of $x(t)$ alone, and the last $nk-q$ elements of $M^{-1}z(t)$ are linear combinations of the last $nk-q$ elements of $z(t)$ alone. (This latter depends on the fact that M's assumed block triangularity is preserved under inversion.)

A sufficient condition for our being able to choose Q and M this way is that the upper left $n \times q$ submatrix of P be of full column rank q, since in that case we can use row pivoting on the first n rows of P (i.e. choose

$$Q = \begin{bmatrix} \underset{n\times n}{Q_{11}} & 0 \\ 0 & I \end{bmatrix} \tag{12}$$

[3] Note that a block triangular permutation matrix must be block diagonal.

to make the upper left $q \times q$ matrix of QP non-singular, then use column pivoting on the last $nk - q$ columns (i.e. choose

$$M = \begin{bmatrix} I & 0 \\ 0 & \underset{nk-q\times nk-q}{M_{22}} \end{bmatrix} \tag{13}$$

to complete the LU decomposition. A sufficient condition for the upper left $n \times q$ submatrix of P to be of rank q is that Λ_{11}, the upper left $q \times q$ submatrix of Λ, is of the form λI. This implies that all non-stationary roots are of the same size and have unit multiplicity (i.e. correspond to 1×1 Jordan blocks.) The most common assumption is that all these roots are unit roots, i.e. $\lambda = 1$. Under this assumption we know that the upper left $n \times q$ submatrix of P is of full column rank. To see this, note that the first q columns of P, a $kn \times q$ matrix we label c, is under these conditions a set of right eigenvectors of Γ corresponding to the eigenvalue λ and thus, using the definition of Γ in terms of $B(L)$, satisfies

$$c_i = c_{i+1}\lambda\,, \quad i = 1,\ldots,k-1\,, \tag{14}$$

where c has been broken up into the k $n \times q$ blocks $c_1,\ldots,c_k$. Thus if c_1 is of less than full column rank, there is a $q \times 1$ vector γ such that not only $c_1\gamma$, but by (14) $c_i\gamma$ for every i, is zero. This would imply that c itself is less than full column rank, which is by construction not true. Thus c_1 must be of full column rank q. Since c_1 is $n \times q$, this lets us conclude that $q \leq n$, certainly. Also, as we have already observed, that c_1 is of rank q implies that there are exactly $n-q$ linearly independent stationary linear combinations of elements of the current x vector. 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 $n-q$ stationary linear combinations of $x(t)$.*

The coefficients of these linear combinations are what is known as

cointegrating vectors.

Now we can give a description of a fairly straightforward algorithm for locating cointegrating vectors:

i. Find the q left eigenvectors of Γ corresponding to the q equal, maximal roots.

ii. If necessary, re-order the variables in x so that the upper left $j \times j$ submatrix of the $q \times nk$ matrix formed by these eigenvectors is nonsingular, all $j \leq q$.

iii. Find $n - q$ additional left eigenvectors of Γ (corresponding to other roots) such that when these are placed below the first q to form the $n \times nk$ matrix $P^{1\cdot}$, the upper left $j \times j$ submatrix is nonsingular, $j = q + 1, \ldots, n$.

iv. Perform an LU decomposition of the resulting corresponding re-ordered P_{11} , the upper left $n \times n$ submatrix of P, so that $P_{11} = UV$ and $P^{11} = V^{-1}U^{-1}$, with V lower triangular and U upper triangular.

Then the $q+1$'st through n'th rows of V^{-1} contain the co-integrating vectors.

Examples

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

$$(1-L)^2 y_1(t) = \varepsilon_1(t) \tag{15}$$

$$y_2(t) = \varepsilon_2(t)\,. \tag{16}$$

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

$$z_1(t) = 4z_1(t-1) - 4z_2(t-1) - 2z_1(t-2) + 2z_2(t-1) + \eta_1(t) \tag{17}$$

$$z_2(t) = 2z_1(t-1) - 2z_2(t-1) - z_1(t-2) + z_2(t-2) + \eta_2(t)\,. \tag{18}$$

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 $2z_2(t) - z_1(t)$.

Unlikely initial conditions

- As we have discussed, estimated VAR's that do not enforce stationarity by assuming initial conditions have the unconditional joint distribution implied by the model often imply that initial conditions are many standard deviations away from their steady-state values.

- Letting $y^* = P^{-1}y$, the implications of "unusual initial conditions" can be quite different depending on whether the root associated with an unusual y^*_{i0} are very close to one or not.

- A root r_i with $|1 - r_i| < 1/T$, paired with a large y^*_{i0} value, produces a near-linear trend deterministic component, which is not unreasonable for many economic time series.

- With $|1 - r_i| \gg 1/T$ and $|r_i| < 1$, a corresponding y_{i0}^* multiple standard deviations from its mean implies initial mean-reverting behavior that will with high probability not be seen again over time spans several times T.

Diagnostic checks

- One can forecast the sample data from the initial conditions, using the posterior mean or mode parameter estimates, to see whether the deterministic component is showing unreasonable prescience.
- One can look at the system roots and corresponding y_{i0}^* values relative to their unconditional standard deviations.

Fixes

- One could use the unconditional joint distribution for the parts of y_0^* whose roots are more than $1/T$ from one, while not doing so for the roots near one.

- For model comparison, some prior not dependent on model parameters would be needed for the near-unit-root y_{i0}^*'s.

- If there are roots near the boundary, this straightforward strategy will have problems, as it generates discontinuities in the posterior as roots move across the boundary.