VAR System Properties from the Jordan Decomposition; cointegration

Chris Sims

April 17, 2018

The model

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

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

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

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.

Simplifying to 1 $\times$ 1 blocks

- Jordan blocks bigger than 1×1 complicate the algebra without much affecting the qualitative discussion that follows.
- They are also relatively rare in applied work. If the true model has a bigger block, the estimated model is likely not to have exactly equal roots, and hence all 1×1 blocks. However, in this case the estimated model's eigenvectors corresponding to the near-equal roots will be nearly the same.
- Henceforth in these notes we assume 1×1 blocks. Versions of these notes that handle the more general case are on, e.g., my 2017 ECO513 web site.

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 element z_i of z has its own single equation in (7),

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

In each of these equations, 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

$|\lambda_i| < 1 \Rightarrow \lambda_i^p \to 0$ as $t \to \infty$. In this case, if η_i satisfies

$$E[\eta_i(t+1) \mid x(t-s), \text{ all } s \geq 0] = 0$$

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

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

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 $\left|\lambda_i^p\right|$ remains at one in absolute value for all p. And if $|\lambda_i| > 1$ λ_i^p grows exponentially 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, $\lambda_i^t z(0)$ does not get smaller with increasing t. (A larger Jordan block with $|\lambda_i| = 1$ generates components that increase in size at polynomial rates)

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$\times$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.

Examples

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

$$(1-L)^2 y_1(t) = \varepsilon_1(t) \tag{11}$$

$$y_2(t) = \varepsilon_2(t)\,. \tag{12}$$

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

$$z_1(t) = 4z_1(t-1) - 4z_2(t-1) - 2z_1(t-2) + 2z_2(t-1) + \eta_1(t) \tag{13}$$

$$z_2(t) = 2z_1(t-1) - 2z_2(t-1) - z_1(t-2) + z_2(t-2) + \eta_2(t)\,. \tag{14}$$

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.

A practical approach

i. Estimate the system first in levels, conditioning on initial conditions.

ii. Look for roots within $1/T$ of 1 in absolute value and count them.

iii. Estimate a VECM model allowing for as many unit roots as you've counted, using the unconditional joint distribution of the stationary linear combinations as part of the likelihood.

A practical approach

- Note that this does not involve frequentist testing to find the number of unit roots — that is likely to fail to reject the unit root null for roots that are quite a bit farther from 1 than 1/T.

- In a system with k lags, the number of stationary linear combinations is $nk - q$, where n is the length of the y vector and q is the number of unit roots.

- If there are more than n unit roots, some of them repeat (i.e. correspond to non-trivial Jordan blocks), and the neat cointegration and VECM setup does not apply.