Econometrics, Lecture 13A, Vector Autoregressions and cointegration.

Ron Smith
EMS, Birkbeck, University of London

Autumn 2020

Last time

- We looked at cointegration where there was a single long-run relationship.
- Modelled the adjustment to long-run equilibrium by error correction models
- Considered the issues in testing for a long-run relationship
- Showed how the ARDL/ECM framework was robust to different orders of integration.
- With two I(1) variables there can be only one cointegrating relationship: the drunk farmer following a random walk and his dog stay close together.
- With more than two I(1) variables there can be more than one cointegrating relationship: the farmer and two dogs, the distance between the farmer and either dog is I(0). Income, q_t, consumption, c_t, and investment, i_t, all I(1), but $c_t - q_t$ and $i_t - q_t$, I(0).

Vector Autoregressions VARs

- Now start thinking of relationships as systems of endogenous variables. No exogenous variables, initially. May be able to test for exogeneity.
- Consider y_t a $m \times 1$, vector, e.g. $y_t' = (q_t, c_t, i_t)$ is 3×1.
- Could model it as a VARMA or VARIMA. But these are very difficult to estimate and usual to use a vector autoregression, VAR, autoregression for a vector y_t.
- Distinguish them fromt VaR, value at risk.
- Zellner called them Very Awful Regression. We shall see why.

Estimating VARs

- Vector version of an AR2 is the VAR2:

$$y_t = a + A_1 y_{t-1} + A_2 y_{t-2} + \varepsilon_t$$

  y_t is $m \times 1$, a is $m \times 1$, A_1 and A_2 are $m \times m$, $\varepsilon_t \sim N(0, \Sigma)$, Σ is $m \times m$ with elements σ_{ij}.

- For $m = 2$, $y_t = (y_{1t}, y_{2t})'$ the VAR is:

$$\begin{aligned} y_{1t} &= a_1^0 + a_{11}^1 y_{1t-1} + a_{12}^1 y_{2t-1} + a_{11}^2 y_{1t-2} + a_{12}^2 y_{2t-2} + \varepsilon_{1t}, \\ y_{2t} &= a_2^0 + a_{21}^1 y_{1t-1} + a_{22}^1 y_{2t-1} + a_{21}^2 y_{1t-2} + a_{22}^2 y_{2t-2} + \varepsilon_{2t}. \end{aligned}$$

- Each equation of the VAR can be estimated by OLS and Σ can be estimated from the OLS residuals,

$$\widehat{\sigma}_{ij} = \frac{1}{T} \sum_{t=1}^{T} \widehat{\varepsilon}_{it} \widehat{\varepsilon}_{jt}$$

  $\widehat{\sigma}_{11}$ is the variance of ε_{1t}, $\widehat{\sigma}_{12}$ the covariance of ε_{1t} and ε_{2t}.

Granger Causality, GC

- A variable y_{2t} is said to Granger cause y_{1t} if knowing current values of y_{2t} helps you to predict future values of y_{1t} equivalently, current y_{1t} is explained by past y_{2t}.
- y_{2t} is GC for y_{1t} if either a_{12}^1 or a_{12}^2 are non zero in

$$y_{1t} = a_1^0 + a_{11}^1 y_{1t-1} + a_{12}^1 y_{2t-1} + a_{11}^2 y_{1t-2} + a_{12}^2 y_{2t-2} + \varepsilon_{1t},$$

- You can test that they are both zero with a standard F test of linear restrictions. The restricted model just excludes $y_{2,t-1}$ and $y_{2,t-2}$ from the equation for y_{1t}. The null hypothesis is no GC.
- GC can go in either, both or neither directions.
- GC is rarely the same as economic causality, particularly because expectations cause consequences to precede their cause: weather forecasts GC the weather.

Lag length

- Can decide lag length by Likelihood Ratio tests or model selection criteria like the AIC or BIC.
- Use the same sample for the restricted and unrestricted model; i.e. do not use the extra observation that becomes available when you shorten the lag length.
- The usual GC tests are no longer valid for I(1) variables. Toda and Yamamoto (1995) suggest dealing with the problem by adding extra lags, beyond the optimal number, which you do not use in the tests.
- With lag length p, each equation of the VAR with intercept has $1 + mp$ parameters. This can get large, 4 lags in a 4 variable VAR gives 17 parameters in each equation. You can easily run out of degrees of freedom.
- Be careful. Can get very bad small sample biases in VARs: hence very awful regressions.
- Bayesian VARs, used in forecasting, are one way of dealing with the problem of too many parameters.

Stability

- A pth order VAR

$$y_t = a + \sum_{i=1}^{p} A_i y_{t-i} + \varepsilon_t$$

is stationary if all the roots of the determinantal equation

$$\mid I - A_1 z - A_2 z^2 - ... - A_p z^p \mid = 0$$

lie outside the unit circle.

- Some programs will give you a graph of the inverse roots, which should lie inside the unit circle for the variables to all be stationary.
- If there are unit roots, some roots will lie on the unit circle.
- 4 inverse roots of VAR2 for log earnings and dividends: 1.008, $0.246 \pm 0.353i$, 0.097.

Impulse response functions, IRF

- Difficult to interpret the coefficients of the VAR so it is common to use Impulse Response Functions, IRFs.
- IRF gives the time profile of the effect of a typical size (one standard error) shock to the error of one variable on future values of that variable and other variables. It uses the MA representation.
- VAR1 $y_t = Ay_{t-1} + \varepsilon_t$ has moving average representation

$$y_t = \varepsilon_t + A\varepsilon_{t-1} + A^2\varepsilon_{t-2} + A^3\varepsilon_{t-3}\ldots\ldots$$

- Higher order VARs have more complicated ones, that can be easily calculated.

$$y_t = \varepsilon_t + \Phi_1\varepsilon_{t-1} + \Phi_2\varepsilon_{t-2} + \Phi_3\varepsilon_{t-3}\ldots\ldots = y_t = \sum_{i=0}^{\infty} \Phi_i\varepsilon_{t-i}$$

Generalised and orthogonalised IRFs

- The errors ε_t are not orthogonal (uncorrelated) since $E(\varepsilon_t \varepsilon_t') = \Sigma$, not diagonal. Generalised IRFs let the other errors respond to the shock according to the covariance matrix. Many think structural shocks should be orthogonal.
- Orthogonalised IRFs use a causal ordering to make the shock uncorrelated. A Choleski decomposition uses $\Sigma = PP'$, where P is lower triangular and

$$y_t = \sum_{i=0}^{\infty} (\Phi_i P)(P^{'-1} \varepsilon_{t-i}) = \sum_{i=0}^{\infty} \Psi_i \eta_i$$

- $E(\eta_i \eta_i') = E\left[(P^{-1}\varepsilon_{t-i})(P^{-1}\varepsilon_{t-i})'\right] = P^{-1}\Sigma P^{-1\prime} = I$. So the η_i are uncorrelated. But P is not unique, it depends on the ordering.
- We will return to these.

Vector Error Correction Models

- We can reparameterise the VAR2 as VECM:

$$y_t = a_0 + A_1 y_{t-1} + A_2 y_{t-2} + \varepsilon_t$$

$$\begin{aligned} y_t - y_{t-1} &= a_0 - (I - A_1 - A_2) y_{t-1} - A_2 (y_{t-1} - y_{t-2}) + \varepsilon_t \\ \Delta y_t &= a_0 - \Pi y_{t-1} + \Gamma \Delta y_{t-1} + \varepsilon_t \end{aligned}$$

- Reparameterizes the VAR(p) as VECM (p or p-1?):

$$\Delta y_t = a_0 - \Pi y_{t-1} + \sum_{i=1}^{p-1} \Gamma_i \Delta y_{t-i} + \varepsilon_t .$$

- This is the vector equivalent of the ADF regression used to test for unit roots.
- Exercise: Express the Γ_i in terms of the A_i.

Cointegration in VARs

- If all the variables, the m elements of y_t, are I(0), Π is a full rank matrix.
- If all the variables are I(1) and not cointegrated, $\Pi = 0$, and a VAR in first differences is appropriate.
- If the variables are I(1) and cointegrated, with r cointegrating vectors, then there are r cointegrating relations, combinations of y_t that are I(0),
$$z_t = \beta' y_t$$
where z_t is a $r \times 1$ vector and β' is a $r \times m$ matrix.

Various VECMs

$$\Delta y_t = a_0 - \Pi y_{t-1} + \sum_{i=1}^{p-1} \Gamma_i \Delta y_{t-i} + \varepsilon_t. \tag{1}$$

$$\Delta y_t = a_0 - \alpha z_{t-1} + \sum_{i=1}^{p-1} \Gamma_i \Delta y_{t-i} + \varepsilon_t, \tag{2}$$

$$\Delta y_t = a_0 - \alpha\beta' y_{t-1} + \sum_{i=1}^{p-1} \Gamma_i \Delta y_{t-i} + \varepsilon_t, \tag{3}$$

- (1) is the unrestricted relation.
- In (2) rhe I(0) dependent variable is only explained by I(0) variables. α is a $m \times r$ matrix of 'adjustment coefficients' which measure how the deviations from equilibrium (the r I(0) variables $z_{t-1} = \beta' y_{t-1}$) feed back on the changes.
- (3) shows the restrictions on Π.

Deterministic terms

- As with the Dickey Fuller regression, there is a problem with the treatment of the deterministic elements.
- If we have a linear trend in the VAR, and do not restrict the trends, the variables will be determined by $m - r$ quadratic trends.
- To avoid this (economic variables tend to show linear not quadratic trends), we enter the trends in the cointegrating vectors, so if elements of $\alpha = 0$, the trend drops out

$$\Delta y_t = a_0 - \alpha(\beta' y_{t-1} + ct) + \sum_{i=1}^{p-1} \Gamma_i \Delta y_{t-i} + \varepsilon_t,$$

.

- Programs ask you at the beginning to choose how you enter trends and intercepts; unrestricted intercepts and restricted trends, option 4 is a good choice for trended economic data.

Cointegrating vectors and stochastic trends

- $\Pi = \alpha\beta'$ has rank $r < m$ if there are r cointegrating vectors.
- If there are $r < m$ cointegrating vectors, then y_t will also be determined by $m - r$ stochastic trends, and will have $m - r$ roots on the unit circle and m roots outside the unit circle.
- If $m = r$, all the variables are $I(0)$
- If there is cointegration, some of the α must be non-zero, there must be some feedback on the y_t to keep them from diverging, i.e. there must be some Granger causality in the system.
- If there are r cointegrating vectors and Π has rank r, it will have r non-zero eigenvalues λ_i.

Testing for cointegration

- Johansen provided two tests for determining how many of the eigenvalues are different from zero.
- Suppose we have 3 variables. Both tests have the same sequence of nulls, they differ in the alternatives.

$$H_0^0 : r = 0; \quad H_0^1 : r = 1; \quad H_0^2 : r = 2;$$

  The maximal eigenvalue test has alternatives:

$$H_1^0 : r = 1; \quad H_1^1 : r = 2; \quad H_1^2 : r = 3.$$

  The trace test has alternatives:

$$H_1^0 : r \geq 1; \quad H_1^1 : r \geq 2; \quad H_1^2 : r = 3.$$

- These allow us to determine r. The two tests may give different answers. The trace test is usually better.
- The Johansen estimates of the cointegrating vectors β are the associated eigenvectors.

Identification 1

- The 'identification' problem is that α and β are not uniquely determined.
- Any non-singular $r \times r$ matrix P with $(\alpha P)(P^{-1}\beta) = \Pi$ gives new estimates $\alpha^* = (\alpha P)$ and $\beta^* = (P^{-1}\beta)$ which are observationally equivalent, though with different economic interpretations.
- If $z_{t-1} = \beta' y_{t-1}$ are I(0) so are $z^*_{t-1} = P^{-1}\beta' y_{t-1}$, since any linear combination of I(0) variables is I(0).
- We need to choose P to allow us to interpret the estimates.
- This requires r^2 restrictions, r on each cointegrating vector. One of these is provided by normalisation, we set the coefficient of the 'dependent variable' to unity, so if $r = 1$ this is straightforward (though it requires the coefficient set to unity to be non-zero).

Identification 2

- If there is more than one cointegrating vector, more prior economic assumptions than normalisation are required.
- The Johansen identification assumption, that the β are eigenvectors with unit length and orthogonal, do not allow an economic interpretation, but can be used for forecasting. All just identified models give the same forecasts.
- Programs allow you to specify the r^2 just identifying restrictions.
- You can test any extra 'over-identifying' restrictions.

Next time

- We will work through two examples to make the ideas more concrete.