Multivariate models and cointegration: the Johansen's procedure

Verbeek (2004), Ch. 9, Patterson, Ch. 14, Greene (2003), Ch. 20, section 20.4.

Here we look at the case where all variables are treated the same, that is no endogenous/exogenous differences are assumed. We need to look at the vector generalisation of AR processes. These are called VARs (Vector Auto Regressions). Let us consider a first order VAR where we have two variables y_1 and y_2. This is a **reduced form** in the sense that each equation will only have lagged values of the variables on the RHS.

$$\begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} + \begin{bmatrix} \pi_{11} & \pi_{12} \\ \pi_{21} & \pi_{22} \end{bmatrix} \begin{pmatrix} y_{1t-1} \\ y_{2t-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix} \tag{1}$$

which can be written as

$$y_t = \mu + \Pi_1 \, y_{t-1} + \varepsilon_t \tag{2}$$

Π_1 is a matrix, the other terms are vectors. μ is the drift term. Equilibrium stability conditions are generalisation of our earlier analysis of univariate autoregressive models. All the eigenvalues of Π_1 must have modulus less than one. Stationarity of the variables is also linked to this. Remember an eigenvalue is related to the roots of a polynomial.

[The eigenvalues of Π_1 are the roots of the nth order characteristic polynomial $|\Pi_1 - \lambda I_k|$ obtained by solving the characteristic equation $|\Pi_1 - \lambda I| = 0$. For stability, λ_1 and λ_2 must be less than one in modulus Alternatively, we can look at the roots of the determinant of the autoregressive polynomial $|A(L)| = |I - \Pi_1 L|$, and for stability these must have modulus greater than one].

Let us set $\mu = 0$ for the moment and <u>rewrite</u> (2) in error correction form:

$$\Delta \, y_t = \Pi \, y_{t-1} + \varepsilon_t \tag{3}$$

Note that this expression generalises with lagged difference terms on RHS, if the starting VAR is of order higher than one. For example, a VAR of order 2

$$y_t = \Pi_1 \, y_{t-1} + \Pi_2 \, y_{t-2} + \varepsilon_t \qquad\qquad (4)$$

can be rewritten in error correction form

$$\Delta \, y_t = \Pi \, y_{t-1} + \Gamma_1 \Delta \, y_{t-1} + \varepsilon_t \qquad\qquad (5)$$

where $\Pi = \Pi_1 + \Pi_2 - I$, and $\Gamma_1 = -\Pi_2$.

The Π matrix tells us how many linear combinations exist amongst the variables. This will be related to the rank of Π. If there are *n* variables in the system, then Π will be *n* x *n*. Therefore, the rank of Π will be between 0 and n. Rank of Π can only be 0 if Π is a null matrix. In this case we have a VAR in first differences, with no level terms. So, variables are I(1), but not cointegrated.

If Π has full rank, it implies that there are *n* independent linear combinations of the variables that are stationary. This in turn implies that we can solve uniquely for each of the *n* variables in terms of stationary variables. So, all the variables must be stationary in the first place, which implies a VAR in the levels.

If there are I(1) variables, then Π matrix will have a rank strictly between 0 and *n*. Supposing rank(Π) = r. This means that there are r linear combinations amongst the *n* variables. These are the cointegrating linear combinations since this must be I(0) for the equation to be balanced. Also in this case the matrix Π must have r non-zero eigenvalues. Hence a test for the number of cointegrating vectors is equivalent to a test of how many non-zero eigenvalues are there in the Π matrix.

Cointegrating rank (number of cointegrating vectors)	Implications
r = 0 (matrix Π is a zero matrix)	VAR can be written in first differences without any level terms. I.e. there is no long-run.
0 < r < n	There are r cointegrating vectors. I.e. r stationary linear combinations.
r = n (matrix Π is of full rank)	VAR is stationary in the levels.

Johansen (1988), Johansen and Juselius (1990) have tabulated critical values for testing the rank of the Π matrix. There are two tests: the **maximum eigenvalue test**, and the **trace test**. These tests are now provided by most of the software.

Let us denote the theoretical eigenvalues of the matrix Π in decreasing order as

$$\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_n$$

The idea is that if there are r cointegrating vectors (rank of Π=r), then $\log(1-\lambda_j)=0$ for the smallest n-r eigenvalues, that is for j=r+1, r+2,…,n.

The two tests described below use the estimated eigenvalues, say $\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq ... \geq \hat{\lambda}_n$, to test hypothesis about the rank of Π.

The <u>maximum eigenvalue</u> statistic: H_0: rank $\leq$ r, H_1: rank=r+1

$$\lambda_{max} = -T \ln\left(1 - \hat{\lambda}_{r+1}\right),\ r=0, 1, ..., n-1$$

tests whether the estimated (r+1)th largest eigenvalue is significantly different from zero.

The <u>Trace statistic</u> : H_0: rank $\leq$ r, H_1: rank $\geq$ r+1

$$\lambda_{trace} = -T \sum_{i=r+1}^{n} \ln\left(1 - \hat{\lambda}_i\right),\ \ r=0, 1, ..., n-1$$

tests whether the smallest n-r estimated eigenvalues are significantly different from zero.

These tests are carried out as a <u>sequence</u> of tests starting from $\lambda_0, \lambda_1, \ldots \lambda_{n-1}$. A large test statistic value implies a rejection of the null.

For the eigenvectors corresponding to the non-stationary part of the model, the eigenvalues $\lambda_i=0$ for $i=r+1, \ldots n$. So, the null hypothesis:

   H_0: there are at most r cointegrating vectors,   corresponds to

   H_0: $\lambda_i = 0$, $i=r+1,,\ldots,n$.

that is only the first r eigenvalues are non-zero $(\lambda_i \neq 0, i=1,\ldots,r)$

<u>Summary</u>

	λ_{max} **Test**		**Trace test**	
	H_0	**H_1**	**H_0**	**H_1**
(i)	r=0	r=1	r=0	r$\geq$1
(ii)	r$\leq$1	r=2	r$\leq$1	r$\geq$2
.	.	.	.	.
.	r$\leq$n-1	r=n	r$\leq$n-1	r=n

<u>Steps</u>: Start with (i). If H_0 is rejected then do (ii). If this is not rejected then it must imply that rank is either 0 or 1. But since we have already rejected the rank=0, conclude that rank must be 1. etc…

Tabulated critical values depend on the specifications of the deterministic terms (constants, dummies, trend etc.) in the equations.

Estimation of cointegration vectors

Having established the number of cointegrating vectors, say r, the next step is estimation of the cointegrating relationships. For more than one cointegrating relationship (r>1), OLS is inappropriate. Johansen develops maximum likelihood estimators.

We now go back to our equation (5) (from Handout 13):

$$\Delta y_t = \Pi\, y_{t-1} + \Gamma_1 \Delta y_{t-1} + \varepsilon_t \tag{5}$$

Let us assume that the above tests have established r cointegrating vectors. This means that the rank of matrix Π is r.

For r cointegrating vectors, Π can be decomposed as $\Pi = \alpha\beta'$, where α and β are both (nxr) matrices. β will contain the cointegrating vectors and α will contain the adjustment coefficients.

Supposing we have n=3 and r=2. We then have the following:

$$\Delta y_t = \alpha\beta' y_{t-1} + \Gamma_1 \Delta y_{t-1} + \varepsilon_t \tag{6}$$

y_t is a (3 x 1) vector. α and β are (3 x 2) matrices, Γ_1 is (3x3) matrix.

$$
\begin{bmatrix} \Delta y_{1t} \\ \Delta y_{2t} \\ \Delta y_{3t} \end{bmatrix}
=
\begin{bmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \\ \alpha_{31} & \alpha_{32} \end{bmatrix}
\begin{bmatrix} \beta_{11} & \beta_{21} & \beta_{31} \\ \beta_{12} & \beta_{22} & \beta_{32} \end{bmatrix}
\begin{bmatrix} y_{1t-1} \\ y_{2t-1} \\ y_{3t-1} \end{bmatrix}
+ \Gamma_1
\begin{bmatrix} \Delta y_{1t-1} \\ \Delta y_{2t-1} \\ \Delta y_{3t-1} \end{bmatrix}
+
\begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \\ \varepsilon_{3t} \end{bmatrix}
\tag{7}
$$

We have three equations with two error correction terms in each. Each of the r cointegration vectors in β is associated with a particular column in α that must have at least one non zero element. The tests for the (reduced) rank of Π (or for the number r of CVs) are equivalent to testing which columns of α are zero. These tests therefore require the factorization of Π, which cannot be done by standard regression techniques, but it is done by a procedure based on reduced rank regressions involving canonical correlations.

Johansen shows that the ML estimate for β is equal to the matrix $\hat{V}$ containing the r eigenvectors $(\hat{v}_1,...,\hat{v}_r)$ corresponding to the r largest eigenvalues $\hat{\lambda}_1, \hat{\lambda}_2,...,\hat{\lambda}_r$.

So, $(\hat{\beta}_1,...,\hat{\beta}_r) = (\hat{v}_1,...,\hat{v}_r)$. Estimates of α are then given for fixed β.

More specifically, the ML estimator of β is found by solving the equation

$$| \lambda S_{11} - S_{10}S_{00}^{-1}S_{01} | = 0 .$$

for the n eigenvalues $\hat{\lambda}_1 > \hat{\lambda}_1 > ... > \hat{\lambda}_n$ and corresponding eigenvectors $\hat{V} = (\hat{v}_1, \hat{v}_2,...\hat{v}_n)$. All calculations are based on the sample moment matrices

$$S_{ij} = T^{-1}\sum_{t=1}^{T}R_{it}R_{jt}^{'} \quad i,j=0,1$$

where $R_{0,t}$ are the residuals from the regressions of Δy_t on the lagged difference terms in (5). Note in our example we have only one lagged first difference of each of the three variables. $R_{1,t}$ are the residuals from the regressions of y_{t-1} on the lagged difference terms.

The n eigenvalues are a measure of correlation between the I(0) variables, Δy_t, and specific linear combinations of y_t. As only a stationary linear combination of the variables y_t will have a high correlation with Δy_t, large eigenvalues correspond to the existence of a stationary linear combination of y_t, that is, of a cointegrating relationship.

The eigenvectors corresponding to the large eigenvalues are then the cointegrating vectors $\hat{\beta} = (\hat{v}_1,...,\hat{v}_r)$.

For fixed β an estimate of α is given by

$$\hat{\alpha}(\beta) = S_{01}\beta(\beta'S_{11}\beta)^{-1}$$

Identification

It is important to note that for $r>1$ these parameter vectors are not uniquely defined as different combinations of α and β can give the same matrix $\Pi=\alpha\beta'$. For any invertible $r \times r$ matrix P,

$$\alpha\beta'=\alpha PP^{-1}\beta'=\alpha^*\beta^{*'}$$

and these yield the same Π matrix.

In order to obtain unique cointegrating relationships, we will need to impose restrictions on β. One obvious thing to do is to normalise so that in (7) we set $\beta_{11}=\beta_{22}=1$.

The above specification (7) implies the following:

$$\Delta y_{1t} = \alpha^*_{11}[1\ y_{1t} + \beta^*_{21}\ y_{2t} + \beta^*_{31}\ y_{3t}] + \alpha^*_{12}[\beta^*_{12}\ y_{1t} + 1\ y_{2t} + \beta^*_{32}\ y_{3t}] + \varepsilon_{1t}$$

and similarly for the other two variables. Stars indicate normalised coefficients, for example, $\alpha^*_{11} = \beta_{11}\alpha_{11}$, $\beta^*_{21}=\beta_{21}/\beta_{11}$, $\beta^*_{31}=\beta_{31}/\beta_{11}$, and so on.

In addition to this, for identification we need to impose another $r-1$ restrictions on each of the CV. In our example, with $n=3$ and $r=2$, one more restriction is needed in each CV. These restrictions can be equal and opposite signs and or zero (exclusion) restrictions. Can have more than the required restrictions and these are called over-identifying restrictions. These over-identifying restrictions can be tested.

Weak exogeneity

The adjustment coefficients αs can be tested to see whether they are zero. If they are zero then that particular error correction term does not enter that equation. More of this later in the money demand example.

If in equation (7) $\alpha_{31}=\alpha_{32}=0$

$$\begin{bmatrix} \Delta y_{1t} \\ \Delta y_{2t} \\ \Delta y_{3t} \end{bmatrix} = \begin{bmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \beta_{11} & \beta_{21} & \beta_{31} \\ \beta_{12} & \beta_{22} & \beta_{32} \end{bmatrix} \begin{bmatrix} y_{1t-1} \\ y_{2t-1} \\ y_{3t-1} \end{bmatrix} + \Gamma_1 \begin{bmatrix} \Delta y_{1t-1} \\ \Delta y_{2t-1} \\ \Delta y_{3t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \\ \varepsilon_{3t} \end{bmatrix} \qquad (8)$$

then the cointegrating vectors in $\boldsymbol{\beta}$ do not enter the equation for Δy_{3t}, and y_3 is said to be weakly exogenous with respect to $\boldsymbol{\beta}$. This means that when estimating $\boldsymbol{\beta}$ there is no loss of information from not modelling Δy_{3t}, and this variable can then enter at time t on the RHS of the VECM:

$$\begin{bmatrix} \Delta y_{1t} \\ \Delta y_{2t} \end{bmatrix} = \begin{bmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \end{bmatrix} \begin{bmatrix} \beta_{11} & \beta_{21} & \beta_{31} \\ \beta_{12} & \beta_{22} & \beta_{32} \end{bmatrix} \begin{bmatrix} y_{1t-1} \\ y_{2t-1} \\ y_{3t-1} \end{bmatrix} + \tilde{\Gamma}_1 \begin{bmatrix} \Delta y_{1t-1} \\ \Delta y_{2t-1} \\ \Delta y_{3t-1} \end{bmatrix} + \Gamma_0 \Delta y_{3t} + \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix} \qquad (9)$$

where $\tilde{\Gamma}_1$ is a (2x3) matrix, and Γ_0 a (2x1) vector.

The test for weak exogeneity is conducted by placing zero restrictions to the elements of the ith row of $\boldsymbol{\alpha}$ and by computing a likelihood ratio test involving the restricted and unrestricted model to test the null hypothesis that the restrictions are valid.

Notes:

(i)   We need to ensure that ε_ts are indeed iid. Will need enough lags in the VAR to eliminate the serial correlation.

(ii)   Cannot have too many lags in the VAR because of degrees of freedom problems.

(iii)   Main advantage is that we can test various hypotheses about the underlying coefficients which we could not do in the Engle-Granger methodology.