EC408 Topics in Applied Econometrics

B Fingleton, Dept of Economics, Strathclyde University
Applied Econometrics

- What is spurious regression?
- How do we check for stochastic trends?
- Cointegration and Error Correction Models
- Autoregressive distributed lag (ADL) models
- VAR models
Applied Econometrics

VAR models

– Vector error correction models
– Multiple cointegrating vectors
– Johansen’s procedure
Multiple cointegrating vectors

- g variables, it is convenient to collect these together and represent them as the g x 1 vector
- Given that we have g variables, we may be able to discover more than one linear combination of the g variables in Y that is stationary, with each linear combination, or cointegrating relationship, being uncorrelated with, or orthogonal to, the others
- Recognising this will give a better specified model
- First we write out the VECM in terms of differences and levels
- This can be shown to be mathematically equivalent to models in which the error correction term (the lagged residuals) is explicit
the vector error correction representation

\[ \Delta Y_t = \phi_1 \Delta Y_{t-1} + \phi_2 \Delta Y_{t-2} + \ldots + \phi_p \Delta Y_{t-p} + P_0 Y_{t-1} + U_t \]  

(72)

Where \( \Delta Y_t = Y_t - Y_{t-1} \) is a \( g \) by 1 vector of differences at time \( t \) of \( g \) endogenous variables. The terms \( \Delta Y_{t-1} = Y_{t-1} - Y_{t-2}, \Delta Y_{t-2} = Y_{t-2} - Y_{t-3}, \ldots, \Delta Y_{t-p} = Y_{t-p} - Y_{t-p-1} \) are the lagged differences. There are \( p \) lags, so \( \phi_j \) applies to the \( j \)'th lag. It is a \( g \times g \) matrix of coefficients to be estimated. Also \( U_t \) is a \( g \times 1 \) vector of error terms. \( P_0 \) is a \( g \times g \) matrix which is referred to as the (restricted) long-run matrix.
the vector error correction representation

let the number of lags $p = 1$, then equation (72) becomes

$$\Delta Y_t = \phi_1 \Delta Y_{t-1} + P_0 Y_{t-1} + U_t$$

for three variables, say $Y_{1t}, Y_{2t}, Y_{3t}$

$$\begin{bmatrix} \Delta Y_{1t} \\ \Delta Y_{2t} \\ \Delta Y_{3t} \end{bmatrix} = \begin{bmatrix} \phi_{11} & \phi_{12} & \phi_{13} \\ \phi_{21} & \phi_{22} & \phi_{23} \\ \phi_{31} & \phi_{32} & \phi_{33} \end{bmatrix} \begin{bmatrix} \Delta Y_{1t-1} \\ \Delta Y_{2t-1} \\ \Delta Y_{3t-1} \end{bmatrix} + \begin{bmatrix} P_{011} & P_{012} & P_{013} \\ P_{021} & P_{022} & P_{023} \\ P_{031} & P_{032} & P_{033} \end{bmatrix} \begin{bmatrix} Y_{1t-1} \\ Y_{2t-1} \\ Y_{3t-1} \end{bmatrix} + \begin{bmatrix} U_{1t} \\ U_{2t} \\ U_{3t} \end{bmatrix}$$

so that for variable $Y_1$

$$\Delta Y_{1t} = \phi_{11} \Delta Y_{1t-1} + \phi_{12} \Delta Y_{2t-1} + \phi_{13} \Delta Y_{3t-1} + P_{011} Y_{1t-1} + P_{012} Y_{2t-1} + P_{013} Y_{3t-1} + U_{1t}$$

and so on for $Y_2, Y_3$
the vector error correction representation

A mathematically equivalent way to write out this model is in terms of levels and lagged levels,

\[
\begin{bmatrix}
Y_{1t} \\
Y_{2t} \\
Y_{3t}
\end{bmatrix} =
\begin{bmatrix}
1 + P_{011} & P_{012} & P_{013} \\
P_{021} & 1 + P_{022} & P_{023} \\
P_{031} & P_{032} & 1 + P_{033}
\end{bmatrix}
\begin{bmatrix}
Y_{1t-1} \\
Y_{2t-1} \\
Y_{3t-1}
\end{bmatrix} +
\begin{bmatrix}
U_{1t} \\
U_{2t} \\
U_{3t}
\end{bmatrix} =
\begin{bmatrix}
\pi_{111} & \pi_{112} & \pi_{113} \\
\pi_{121} & \pi_{122} & \pi_{123} \\
\pi_{131} & \pi_{132} & \pi_{133}
\end{bmatrix}
\begin{bmatrix}
Y_{1t-1} \\
Y_{2t-1} \\
Y_{3t-1}
\end{bmatrix} +
\begin{bmatrix}
U_{1t} \\
U_{2t} \\
U_{3t}
\end{bmatrix}
\]

so that for variable \(Y_1\)

\[
Y_{1t} = (1 + P_{011})Y_{1t-1} + P_{012}Y_{2t-1} + P_{013}Y_{3t-1} + U_{1t} = \pi_{111}Y_{1t-1} + \pi_{112}Y_{2t-1} + \pi_{113}Y_{3t-1} + U_{1t}
\]

\[
Y_{2t} = P_{021}Y_{1t-1} + (1 + P_{022})Y_{2t-1} + P_{023}Y_{3t-1} + U_{2t} = \pi_{121}Y_{1t-1} + \pi_{122}Y_{2t-1} + \pi_{123}Y_{3t-1} + U_{2t}
\]

\[
Y_{3t} = P_{031}Y_{1t-1} + P_{032}Y_{2t-1} + (1 + P_{033})Y_{3t-1} + U_{3t} = \pi_{131}Y_{1t-1} + \pi_{132}Y_{2t-1} + \pi_{133}Y_{3t-1} + U_{3t}
\]

and \(\pi_1\) is a \(g \times g\) matrix of coefficients specific to lag 1
Additional variables in the ADL

we do reject Ho: $\delta = 0$ in favour of $\delta < 0$ for INFLAT, since $\text{ADF-INFLAT} = -6.197^{**}$ Critical values used in ADF test: 5%=-3.439, 1%=-4.019.

$\text{ADF-CONS} = -3.034$ which indicates that Ho: $\delta = 0$ should not be rejected in favour of $\delta < 0$ using critical values: 5%=-3.439, 1%=-4.019. We also find that $\delta = 0$ is not rejected for INC, since $\text{ADF-INC} = -3.14$ Critical values used in ADF test: 5%=-3.439, 1%=-4.019.
the vector error correction representation

As a numerical example, consider $P_0$ to be as follows,

\[
\begin{array}{c|c|c}
\text{CONS} & \text{INC} & \text{INFLAT} \\
\hline
\text{CONS} & -0.15042 & 0.14999 & -1.2089 \\
\text{INC} & 0.072210 & -0.070129 & -0.49009 \\
\text{INFLAT} & 0.019763 & -0.019382 & -0.026605 \\
\end{array}
\]

\[
\pi_1 = P_0 + I
\]

Where $I$ is a $g \times g$ identity matrix

\[
\begin{pmatrix}
1 & 0 & 0 \\
0 & 1 & 0 \\
0 & 0 & 1 \\
\end{pmatrix}
\]

$\pi_1$ is the following set of numerical coefficients

\[
\begin{align*}
\text{CONS} &= +0.8496\times\text{CONS}_1 + 0.15\times\text{INC}_1 - 1.209\times\text{INFLAT}_1 \\
\text{INC} &= +0.07221\times\text{CONS}_1 + 0.9299\times\text{INC}_1 - 0.4901\times\text{INFLAT}_1 \\
\text{INFLAT} &= +0.01976\times\text{CONS}_1 - 0.01938\times\text{INC}_1 + 0.9734\times\text{INFLAT}_1
\end{align*}
\]
the vector error correction representation

More generally, with \( p > 1 \)

\[
Y_t = \sum_{i=1}^{p} \pi_i Y_{t-i} + U_t
\]

in this \( \pi_i \) is a \( g \) by \( g \) matrix of coefficients specific to lag \( i \)  

(73)

And it follows that

\[
P_0 = \sum_{i=1}^{p} \pi_i - I
\]

(74)

\( I \) is a \( g \times g \) identity matrix, and \( P_0 \) is the \( g \times g \) (restricted) long-run matrix
the vector error correction representation

As a numerical example with $p = 2$, we have $P_0$ as follows

Restricted long-run matrix, rank 2

<table>
<thead>
<tr>
<th></th>
<th>CONS</th>
<th>INC</th>
<th>INFLAT</th>
</tr>
</thead>
<tbody>
<tr>
<td>CONS</td>
<td>-0.19280</td>
<td>0.19211</td>
<td>-1.4366</td>
</tr>
<tr>
<td>INC</td>
<td>-0.048286</td>
<td>0.049558</td>
<td>-1.1167</td>
</tr>
<tr>
<td>INFLAT</td>
<td>-0.0011498</td>
<td>0.0013482</td>
<td>-0.11461</td>
</tr>
</tbody>
</table>

Then $\pi_1$ is the set of numerical coefficients attached to the lag 1 terms, and $\pi_2$ is the set of numerical coefficients attached to the lag 2 terms

\[
\begin{align*}
\text{CONS} &= + 0.6958*\text{CONS}_1 + 0.1454*\text{INC}_1 - 1.182*\text{INFLAT}_1 + 0.1114*\text{CONS}_2 \\
&\quad + 0.04668*\text{INC}_2 - 0.2543*\text{INFLAT}_2 \\
\text{INC} &= - 0.3309*\text{CONS}_1 + 0.98*\text{INC}_1 + 0.2796*\text{INFLAT}_1 + 0.2826*\text{CONS}_2 \\
&\quad + 0.0696*\text{INC}_2 - 1.396*\text{INFLAT}_2 \\
\text{INFLAT} &= - 0.001926*\text{CONS}_1 + 0.01442*\text{INC}_1 + 1.535*\text{INFLAT}_1 + 0.0007761*\text{CONS}_2 \\
&\quad - 0.01307*\text{INC}_2 - 0.65*\text{INFLAT}_2
\end{align*}
\]
the vector error correction representation

\[ P_{011} = \pi_{111} + \pi_{211} - 1 \]
\[ P_{011} = 0.6958 + 0.1114 - 1 = -0.19280 \]
\[ P_{012} = \pi_{112} + \pi_{212} \]
\[ P_{012} = 0.1454 + 0.04668 = 0.19211 \]

Hence and so on.... for example

\[ P_{033} = \pi_{133} + \pi_{233} - 1 \]
\[ P_{033} = 1.535 + 0.65 - 1 = -0.11461 \]
We need $P_0 Y_{t-1}$ to be I(0) to balance the fact that $ΔY_t \sim I(0)$ in equation (72).

Given that we have $g$ variables, we may be able to discover more than one linear combination of the $g$ variables in $Y$ that is stationary, with each linear combination, or cointegrating relationship, being uncorrelated with, or orthogonal to, the others.

Although $P_0$ is a $g \times g$ matrix, it could be the result of $r < g$ cointegrating vectors.

The approach developed by Johansen (1988) is designed to seek the actual number $r$ of cointegrating relationships.
the vector error correction representation

There may also be some deterministic variables in the VAR, such as a time trend or a constant term, or some other non-modelled exogenous variables. We denote these by $X$. We also need these to be I(0) for our equation to balance, with I(0) variable throughout. Now the specification becomes

$$Y_t = \sum_{i=1}^{p} \pi_i Y_{t-1} + \sum_{j=0}^{r} \Gamma_j X_{t,j} + U_t$$

(75)
the vector error correction representation

$P_0$ is a $g \times g$ matrix of long-run responses

For the model to ‘work’ the left hand side and the right hand side must be $I(0)$

We know, since it assumed that $Y$ is $I(1)$, that $\Delta Y_t$ is a set of $I(0)$ variables

The rank of $P_0$ is the number of linearly independent rows of the matrix, and is given by the number of non-zero eigenvalues (characteristic roots).

Mathematically, since we assume $Y$ is $I(1)$, $P_0$ cannot be full rank (equal to $g$) and $P_0Y_{t-1} \sim I(0)$.

We have to place restrictions on the rank of $P_0$ so that the rank $r < g$.

If the rank of $P_0$ is $r$, this equals the number of independent cointegrating relationships between the $g$ variables.
the vector error correction representation

The long run matrix $P_0$ can be decomposed, it is the product of the $g \times r$ matrix of cointegrating vectors ($\beta$) and another $g \times r$ matrix ($\alpha$).
the vector error correction representation

\[ \alpha \ (3 \times 1) \]

\[
\begin{array}{c}
-0.19545 \\
-0.090935 \\
-0.0063286 \\
\end{array}
\]

\[ \alpha_{ij} \] is responsiveness of i'th variable to disequilibrium at t-1 given by j'th cointegrating vector here j = 1

\[ \beta \ (3 \times 1) \]

\[
\begin{array}{c}
1.0000 \\
-0.99413 \\
6.3027 \\
\end{array}
\]

Cointegrating vector

\[ P_0 = \alpha \beta' \ (3 \times 3) \]

\[
\begin{array}{ccc}
-0.19545 & 0.19431 & -1.2319 \\
-0.090935 & 0.090402 & -0.57314 \\
-0.0063286 & 0.0062915 & -0.039887 \\
\end{array}
\]

Rank = 1

Equation 1: d_CONS

<table>
<thead>
<tr>
<th>VARIABLE</th>
<th>COEFFICIENT</th>
<th>STDERROR</th>
<th>T STAT</th>
<th>P-VALUE</th>
</tr>
</thead>
<tbody>
<tr>
<td>EC1</td>
<td>-0.19542</td>
<td>0.0247129</td>
<td>-7.909</td>
<td>&lt;0.00001***</td>
</tr>
</tbody>
</table>

Equation 2: d_INC

<table>
<thead>
<tr>
<th>VARIABLE</th>
<th>COEFFICIENT</th>
<th>STDERROR</th>
<th>T STAT</th>
<th>P-VALUE</th>
</tr>
</thead>
<tbody>
<tr>
<td>EC1</td>
<td>-0.0909355</td>
<td>0.0436303</td>
<td>-2.084</td>
<td>0.03876**</td>
</tr>
</tbody>
</table>

Equation 3: d_INFLAT

<table>
<thead>
<tr>
<th>VARIABLE</th>
<th>COEFFICIENT</th>
<th>STDERROR</th>
<th>T STAT</th>
<th>P-VALUE</th>
</tr>
</thead>
<tbody>
<tr>
<td>EC1</td>
<td>-0.00632863</td>
<td>0.00648877</td>
<td>-0.975</td>
<td>0.33090</td>
</tr>
</tbody>
</table>
Johansen’s procedure: an informal introduction

designed to estimate the actual number of cointegrating linear combinations $r$

tests how many of the eigenvalues are significantly different from zero

The rank $r$ will range from 0 to $g$

If $r$ is zero, then that indicates there are no stationary linear combinations of the levels of the variables in $Y \sim I(1)$

If $r = 1$, then that means there is just one cointegrating vector

If $r = g$, then every linear combination of the variables in $Y$ is stationary

this implies that all the series in $Y \sim I(0)$

this contradicts the assumption that $Y \sim I(1)$ giving the left hand side variables $\Delta Y_t \sim I(0)$

we want $P_0$ to be less than full rank, or the number of columns in $\beta$ to be less than $g$
Johansen’s procedure: an informal introduction

First we fit unrestricted reduced forms (URFs) for each of the endogenous variables CONS, INC, INFLAT.

We then fit cointegrating equations, reducing the rank on $P_0$ from $r = g = 3$ to $r = 0$.

With $r = 3$ we see that the matrix $\beta$ has 3 columns and these are the 3 separate cointegrating vectors:

$\alpha$ (3 $\times$ 3)

\[
\begin{pmatrix}
-0.15038 & 0.14992 & -1.2053 \\
0.072278 & -0.070260 & -0.48341 \\
0.019754 & -0.019364 & -0.027522
\end{pmatrix}
\]

$\beta$ (3 $\times$ 3)

\[
\begin{pmatrix}
1 & 0 & 0 \\
0 & 1 & 0 \\
0 & 0 & 1
\end{pmatrix}
\]

$P_0 = \alpha \beta'$ (3 $\times$ 3)

\[
\begin{pmatrix}
-0.15038 & 0.14992 & -1.2053 \\
0.072278 & -0.070260 & -0.48341 \\
0.019754 & -0.019364 & -0.027522
\end{pmatrix}
\]

Log-likelihood $-748.758136$  This is exactly the same as for the unrestricted URF.

Thus the $r = 3$ specification entails no simplification of the URF model. They are identical.
Johansen’s procedure: an informal introduction

when we fit the model with \( r = 2 \), giving only 2 cointegrating vectors

\[
\begin{align*}
\alpha & \quad (3 \times 2) \\
-0.15042 & \quad 0.14999 \\
0.072210 & \quad -0.070129 \\
0.019763 & \quad -0.019382 \\
\beta & \quad (3 \times 2) \\
1.0000 & \quad 0.0000 \\
0.0000 & \quad 1.0000 \\
-560.33 & \quad -569.98 \\
P_0 = \alpha\beta' & \quad (3 \times 3) \\
-0.15042 & \quad 0.14999 & \quad -1.2089 \\
0.072210 & \quad -0.070129 & \quad -0.49009 \\
0.019763 & \quad -0.019382 & \quad -0.026605 \\
\end{align*}
\]

the log-likelihood is \( \log L_R = -748.797631 \), only marginally less than for \( r = 3 \)
Johansen’s procedure : an informal introduction

With \( r = 1 \), there is only 1 cointegrating vector

\[ \alpha \text{ (3 x 1)} \]

\[
\begin{array}{c}
-0.19545 \\
-0.090935 \\
-0.0063286 \\
\end{array}
\]

\[ \beta \text{ (3 x 1)} \]

\[
\begin{array}{c}
1.0000 \\
-0.99413 \\
6.3027 \\
\end{array}
\]

\[ P_0 = \alpha \beta' \text{ (3 x 3)} \]

\[
\begin{array}{ccc}
-0.19545 & 0.19431 & -1.2319 \\
-0.090935 & 0.090402 & -0.57314 \\
-0.0063286 & 0.0062915 & -0.039887 \\
\end{array}
\]

The log-likelihood is \(-756.56227\)
The results for $r = 1$ show a much bigger drop in log-likelihood, which is down to $L_{R} = -756.562268$

This suggests that the simplification involved in just having one cointegrating vector is too great.

Overall, the assumption that $r = 2$ appears to be the most acceptable, given the larger falls in the likelihood with lower $r$. However, we cannot really be sure about this until the significance of the changes in likelihood have been formally tested, which is the function of Johansen’s procedure for establishing the rank $r$ of $P_{0}$.

<table>
<thead>
<tr>
<th>Model</th>
<th>$T$</th>
<th>$p$</th>
<th>log-likelihood</th>
<th>SC</th>
<th>HQ</th>
<th>AIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>SYS(51)</td>
<td>158</td>
<td>5</td>
<td><strong>-756.56227</strong></td>
<td>9.7369</td>
<td>9.6794</td>
<td>9.6400</td>
</tr>
<tr>
<td>SYS(50)</td>
<td>158</td>
<td>8</td>
<td><strong>-748.79763</strong></td>
<td>9.7348</td>
<td>9.6427</td>
<td>9.5797</td>
</tr>
<tr>
<td>SYS(49)</td>
<td>158</td>
<td>9</td>
<td><strong>-748.75814</strong></td>
<td>9.7663</td>
<td>9.6627</td>
<td>9.5919</td>
</tr>
<tr>
<td>SYS(48)</td>
<td>158</td>
<td>9</td>
<td>OLS</td>
<td><strong>-748.75814</strong></td>
<td>9.7663</td>
<td>9.6627</td>
</tr>
</tbody>
</table>
**Johansen test**

- **Trace test**
- Compares likelihoods for rank \( r \) model and \( \text{var} \) model (full rank)
- If the difference is significant, we **cannot** assume rank is \( r \) and eliminate higher ranks
- If the difference is not significant, we can assume rank is \( r \)

- Null hypothesis: \( \text{rank} \leq r \)
- Alternative hypothesis: \( r < \text{rank} \leq \text{full rank} \)
Johansen test

• **Maximum eigenvalue test**
  • Compares likelihoods for rank r model and rank r+1 model
  • If the difference is significant, rank r + 1 improves likelihood
  • and we assume rank is r + 1

• If the difference is not significant, we can assume rank is r

• Null hypothesis : rank at most = r
• Alternative hypothesis : rank = r + 1
Johansen’s procedure

Johansen test:
Number of equations = 3
Lag order = 1
Estimation period: 1953:2 - 1992:3 (T = 158)

Case 1: No constant

<table>
<thead>
<tr>
<th>Rank</th>
<th>Eigenvalue</th>
<th>Trace test p-value</th>
<th>Lmax test p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.44574</td>
<td>108.85 [0.0000]</td>
<td>93.239 [0.0000]</td>
</tr>
<tr>
<td>1</td>
<td>0.093611</td>
<td>15.608 [0.0127]</td>
<td>15.529 [0.0071]</td>
</tr>
<tr>
<td>2</td>
<td>0.00049982</td>
<td>0.078991 [0.8414]</td>
<td>0.078991 [0.8327]</td>
</tr>
</tbody>
</table>
Johansen’s procedure

In terms of likelihood, this shows exactly the same results

The null hypothesis in this case is that $r \leq 1$, in other words columns 2 and 3 of $\beta$ are null (there is at most one cointegrating vector)

Comparing the log-likelihoods for $r = 3$ versus $r = 1$, we obtain a test statistic $15.608 = 2\{\log L_U - \log L_R\} = 2(-748.75814 + 756.56227)$

$$\text{prob}\{\chi^2_4 \geq 15.608\} = 0.0036$$

For reasons similar to those that produce the non-standard distributions for the Dickey-Fuller test statistic, $\chi^2_4$ is not the correct reference distribution

Both gretl and PcGIVE provide the appropriate p-value, equal to 0.013
Johansen’s procedure

Ho: r \leq 2

\[ 2\{\log L_U - \log L_R\} = 2(-748.75814 + 748.79763) = 0.078991 \]

p-value equal 0.841 (nb compare this with the p-value of 0.7787 given by the theoretical \( \chi_1^2 \) distribution)
Johansen’s procedure

Given that we have established the rank of $P_0$ and hence the number of cointegrating vectors, we can then move forward in the knowledge that we have a balanced model with stationary variables.

we can obtain estimates of the dependencies within the data that are not spurious, and we will ultimately be able to produce more credible forecasts and a richer and more informative picture of the interrelationships between the variables.