ECON 4160, Spring term 2014. Lecture 10

Co-integration

Ragnar Nymoen

Department of Economics

27 Nov 2014
Introduction I

So far we have considered:

- Stationary VAR, with deterministic extensions ("no unit roots")
  - Standard inference for dynamic models
- Non-stationary VAR with independent variables ("all unit-roots")
  - Danger of spurious relationships

We next consider \textit{cointegration}, the case of "some, but not only" unit-roots in the VAR.

- In such systems, there exist one or more linear combinations of $I(1)$ variables that are $I(0)$—they are called \textit{cointegration relationships}. 
We see already that cointegration is the “flip of the coin” of spurious regression: If we have two dependent $I(1)$ variables, they are cointegrated.

We can also guess that a test of the null hypothesis of no cointegration is going to be of the Dickey-Fuller type.

In this lecture we want to at least sketch the theory of cointegration more fully:

- The cointegrated VAR: The representation of VARs with some but not all unit-roots
- Testing the null-hypothesis of no cointegration
  - The cointegrating regression
  - The conditional ECM
  - VAR methods, testing hypotheses about rank reduction
- Estimating the cointegrated VAR.
The VAR with a unit root I

Consider the bi-variate first order VAR

\[ y_t = \Phi y_{t-1} + \varepsilon_t \]  \hspace{1cm} (1)

where \( y_t = (Y_t, X_t) \), \( \Phi \) is a 2 \( \times \) 2 matrix with coefficients and \( \varepsilon_t \) is a vector with Gaussian disturbances.

The characteristic equation of \( \Phi \):

\[ |\Phi - zI| = 0, \]

We consider the intermediate case of one unit-root and one stationary root. Specifically

\[ z_1 = 1, \text{ and } z_2 = \lambda, |\lambda| < 1. \]  \hspace{1cm} (2)
implying that both $X_t$ and $Y_t$ are $I(1)$. $\Phi$ has full rank, equal to 2. It can be diagonalized in terms of its eigenvalues and the corresponding eigenvectors:

$$\Phi = P \begin{bmatrix} 1 & 0 \\ 0 & \lambda \end{bmatrix} Q$$

(3)

$P$ has the eigenvectors as columns:

$$P = \begin{bmatrix} \alpha & \beta \\ \gamma & \delta \end{bmatrix}$$

(4)

and $Q = P^{-1}$. 
Cointegrated VAR I

(1) with (2) implies

\[
\begin{bmatrix}
W_t \\
-EC_t
\end{bmatrix} = 
\begin{bmatrix}
1 & 0 \\
0 & \lambda
\end{bmatrix} 
\begin{bmatrix}
W_{t-1} \\
-EC_{t-1}
\end{bmatrix} + \eta_t,
\]

(5)

where \( \eta_t \) contains linear combinations of the original VAR disturbances.

- \( W_t \sim I(1) \), is a stochastic trend, but
- \( EC_t \sim I(0) \)

The expression for \(-EC_t\) is

\[
EC_t = -\gamma Y_t + \alpha X_t.
\]

(6)
Cointegrated VAR II

- We say that there is **cointegration** between $X_t$ and $Y_t$, since $EC_t$ is a stationary variable, and it is a linear combination of $X_t$ and $Y_t$.

- $-\gamma$ and $\alpha$ are the **cointegrating parameters** in this example model.
The Common Trends representation I

The *Common Trends* representation for $Y_t$ and $X_t$ is:

$$Y_t = \alpha W_t - \beta EC_t \quad (7)$$

$$X_t = \gamma W_t - \delta EC_t. \quad (8)$$

- $X_t$ and $Y$ have one common stochastic trend, which is $W_t$. 
The Common Trends representation II

“Corollaries”

1. Forecasts for $X_{T+h|T}$ and $Y_{T+h|T}$ become dominated by the common stochastic trend.

2. Cointegration is maintained in the forecasts, so

$$EC_{T+h|T} = -\gamma X_{T+h|T} + \alpha Y_{T+h|T} = 0$$

for large $h$. 
The ECM representation I

We can always re-write (re-parameterixe) the VAR (1) as

\[ \Delta y_t = \Pi y_{t-1} + \varepsilon_t \]  \hspace{2cm} (9)

with

\[ \Pi = (\Phi - I) \]  \hspace{2cm} (10)

Next, define two \((2 \times 1)\) parameter vectors \(\alpha\) and \(\beta\) in such a way that the product \(\alpha \beta'\) gives \(\Pi\):

\[ \Pi = \alpha \beta' \]  \hspace{2cm} (11)

In our example, we can find \(\alpha\) and \(\beta\) by first expressing \(\Phi\) as

\[ \Phi = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + (1 - \lambda) \begin{bmatrix} \beta \\ \delta \end{bmatrix} \begin{bmatrix} \gamma & -\alpha \end{bmatrix} \]  \hspace{2cm} (12)
The ECM representation II

to give

\[ \Pi = (1 - \lambda) \begin{bmatrix} \beta \\ \delta \end{bmatrix} \begin{bmatrix} \gamma & -\alpha \end{bmatrix} = \begin{bmatrix} (1 - \lambda)\beta \\ (1 - \lambda)\delta \end{bmatrix} \begin{bmatrix} \gamma & -\alpha \end{bmatrix} \]

and then (9) can be expressed as:

\[
\begin{bmatrix} \Delta Y_t \\ \Delta X_t \end{bmatrix} = \alpha \beta' \begin{bmatrix} Y_{t-1} \\ X_{t-1} \end{bmatrix} + \varepsilon_t, \quad (13)
\]
The ECM representation III

- \( \alpha \) is known as the (matrix) of **equilibrium correction coefficients** (aka adjustment coefficients, or loadings),

\[
\alpha = \begin{bmatrix}
(1 - \lambda)\beta \\
(1 - \lambda)\delta
\end{bmatrix}
\] (14)

- \( \beta \) is the matrix of long-run cointegration coefficients

\[
\beta = \begin{bmatrix}
\gamma \\
-\alpha
\end{bmatrix}
\] (15)

In this formulation we see that

- \( \text{rank}(\Pi) = 0 \), reduced rank and no cointegration. Both eigenvalues are zero.
The ECM representation IV

- \( \text{rank}(\Pi) = 1 \), reduced rank and cointegration. One eigenvalue is different from zero.
- \( \text{rank}(\Pi) = 2 \), full rank, both eigenvalues are different from zero and the VAR (1) is stationary.

**Cointegration and Granger causality**
Since \( \lambda < 1 \) is equivalent with cointegration, we see from (14) that cointegration also implies Granger-causality in at least one direction: \((1 - \lambda)\beta \neq 0\) and/or \((1 - \lambda)\beta \neq 0\).
The ECM representation V

Cointegration and weak exogeneity

Assume $\delta = 0$, from (14). this implies

$$
\begin{bmatrix}
\Delta Y_t \\
\Delta X_t
\end{bmatrix} = (1 - \lambda) \begin{bmatrix}
\beta \\
0
\end{bmatrix} \left[ \gamma Y_{t-1} - \alpha X_{t-1} \right] + \varepsilon_t
$$

$$
\begin{bmatrix}
\Delta Y_t \\
\Delta X_t
\end{bmatrix} = \left(1 - \lambda\right) \beta \left[ \gamma Y_{t-1} - \alpha X_{t-1} \right] + \varepsilon_{y,t}
\begin{cases}
\varepsilon_{x,t}
\end{cases}
$$

The marginal model contains no information about the cointegration parameters $(\gamma, -\alpha)’$. $Y_t$ is WE for $(\gamma, -\alpha)’$. 

VAR(p) \rightarrow ECM general case I

If \( y_t \) is \( n \times 1 \) with \( I(1) \) variables. The VAR is:

\[
y_t = \Phi(L)y_{t-1} + \varepsilon_t
\]

where \( \varepsilon_t \) is multivariate Gaussian and

\[
\Phi(L) = \sum_{i=0}^{p} \Phi_{i+1}L^i
\]

In analogy to the scaler case, the matrix lag-polynomial can be written as

\[
\Phi(L) = \Phi(1) + \Delta \Phi^*(L)
\]

where the \( \Phi^*_i \) matrices

\[
\Phi^*(L) = \Phi_1^* + \Phi_2^*L + \ldots + \Phi_{p-1}^*L^{p-1}
\]
VAR(p) —> ECM general case II

are linear transformations of $\Phi_i \ (i = 1, \ldots, p)$. Substitution yields

$$y_t = \Phi^*(L) \Delta y_{t-1} + \Phi(1)y_{t-1} + \epsilon_t$$
$$\Delta y_t = \Phi^*(L) \Delta y_{t-1} + \Pi(1)y_{t-1} + \epsilon_t \quad (17)$$

where $\Pi(1) \equiv \Phi(1) - I_N = 0$ in the case of no cointegration but

$$\Pi(1) = \alpha \beta' \quad (18)$$

in the case of $r$ cointegrating-vectors.

$\beta_{n \times r}$ contains the CI-vectors as columns, while $\alpha_{n \times r}$ shows the strength of equilibrium correction in each of the equations for $\Delta Y_{1t}, \Delta Y_{2t}, \ldots, \Delta Y_{nt}$. In general $\text{rank}(\beta) = r$ and $\text{rank}(\Pi) = r < n$. 
VAR(p) —> ECM general case III

- If $\beta$ is known, the system

$$\Delta y_t = \Phi^*(L)\Delta y_{t-1} + \alpha [\beta' y]_{t-1} + \varepsilon_t$$  \hspace{1cm} (19)

contains only $I(0)$ variables and conventional asymptotic inference applies.

- Moreover: If $\beta$ is regarded as known, after first estimating $\beta$, conventional asymptotic inference also applies.

- $(19)$ is then a stationary VAR, called the VAR-ECM or the cointegrated VAR.

- This system can be identified and modelled with the concepts that we have developed for the stationary case
Restricted and unrestricted constant term

- Usually we include separate *Constants* in each row of the VAR.
- We call them unrestricted constant terms, and with a unit-root the implication is that each contains separate $Y_{jt}$ a deterministic trends (think of a Random Walk with drift)
- However if the constants are *restricted* to be in the $EC_{t-1}$ variables there are no drifts and therefore no trend in the levels variables. More on this in E 5101
- We mention it here because it reminds us that, in the same way as with DF-test, the role of deterministic terms is important when there are unit-roots.
- It also matters for the construction of the tests we use (again, the DF test is a parallel).
Conditional ECM I

Assume that $\alpha_{21} = 0$, i.e. $Y_{2t}$ is weakly exogenous for $\beta$. With Gaussian disturbances $\varepsilon_t = N(0, \Omega)$, where $\Omega$ has elements $\omega_{ij}$, we can derive the conditional model for $\Delta Y_{1t}$:

$$
\Delta Y_{1t} = \omega_{21} \omega_{22}^{-1} \Delta Y_{2t} + \alpha_{11} \beta' \begin{bmatrix} Y_{1t-1} \\ Y_{2t-1} \end{bmatrix} + \varepsilon_{1t} - \omega_{21} \omega_{22}^{-1} \varepsilon_{2t} \tag{20}
$$

the single equation ECM we have discussed before. (20) is an example of an open system, since $x_{t-1}$ is determined outside the model.

If we write it as

$$
\Delta Y_{1t} = b \Delta Y_{2t} + \alpha_{11} \beta_{11} Y_{1t-1} + \alpha_{11} \beta_{12} Y_{2t-1} + u_t
$$

we see that $\Pi = \alpha_{11} \beta_{11} \neq 0$, i.e., the $\Pi$ “matrix” has full rank.
Open systems

Conditional ECM II

- Open systems are often relevant, ideally after first testing $\alpha_{21} = 0$, but also when this is difficult.
- The common trend is now in the non-modelled variable $Y_{2t-1}$.
- Care must be taken: The relevant distribution for testing $\text{rank}(\Pi) = 0$ is (as we shall see) different from the distribution that applies for the closed system.
- Generalization: Open systems can of course contain $n_1$ endogenous $I(1)$ variables and $n_2$ non-modelled $I(1)$ variables. Cointegration is then consistent with

$$0 < \text{rank}(\Pi) \leq n_1$$
Identification 1

- When $n = 2$, cointegration implies $\text{rank}(\Pi) = 1$
  - There is one cointegration vector
    \[
    (\beta_{11}, \beta_{12})'
    \]
    which is uniquely identified after normalization. For example with $\beta_{11} = -1$ the ECM variable becomes
    \[
    ECM_{1t} = -Y_{1t} + \beta_{12} Y_{2t} \sim I(0)
    \]
- When $n > 2$, we can have $\text{rank}(\Pi) > 1$, and in these case the cointegrating vectors are not identified.
Assume that $\Pi$ is known (in practice, consistently estimated), and $\beta$ is a $n \times r$ cointegrating vector:

$$\Pi = \alpha \beta'$$

However for a $r \times r$ non-singular matrix $\Theta$:

$$\Pi = \alpha \Theta \Theta^{-1} \beta' = \alpha \Theta \beta'_{\Theta}$$

showing that $\beta'_{\Theta}$ is also a cointegrating vector.

This problem is equivalent to the identification problem in simultaneous equation models.
Identification III

- Assume \( \text{rank}(\Pi) = 2 \) for a \( n = 3 \) VAR

\[
- Y_{1t} + \beta_{12} Y_{2t} + \beta_{13} Y_{3t} = ECM_{1t} \\
\beta_{21} Y_{1t} - Y_{2t} + \beta_{13} Y_{3t} = ECM_{2t}
\]

- By simply viewing these as a pair of simultaneous equations, we see that they are not identified on the order-condition.

- Exact identification requires for example 1 linear restrictions on each of the equations.
  - For example \( \beta_{13} = 0 \) and \( \beta_{21} + \beta_{13} = 0 \) will result in exact identification
  - Identification = theory !!!

- Restrictions of the loading matrix can also help identification (then hypotheses about causation?)
Identification IV

- A very useful estimator of $\Pi$ is the Maximum-Likelihood estimator (OLS on each equation in the VAR). A natural test-statistic for any overidentifying restrictions is the LR test.

- The identification issue applies equally for open systems. Again, in direct analogy to the simultaneous equation model.
Cointegration: estimation and testing I

- Depends on how much we know about

\[ \Pi(1) \equiv \Phi(1) - I_N \]

apriori.

- A “typology” is (simplifying \( \Pi(1) = \Pi \)):

1. \textit{rank}(\( \Pi \)) is 1
   Estimating a unique cointegrating vector by means of:
   The cointegration regression
   The ECM estimator
Cointegration: estimation and testing II

2. $\text{rank}(\Pi)$ is 0 or 1
   Test $\text{rank}(\Pi) = 0$ against $\text{rank}(\Pi) = 1$, by
   Engle-Granger tests
   ECM test

3. Test and ML estimation based on VAR
   VAR based “Johansen test” for $\text{rank}(\Pi)$ (other than 0 or 1)
   ML estimation of $\beta$ for the case of’ $\text{rank}(\Pi) \geq 2$
   No assumptions about weak exogeneity of variables with respect
   to $\beta$. 
The cointegrating regression I

When \( \text{rank}(\Pi) = 1 \), the cointegration vector is unique (subject only to normalization).
Without loss of generality we set \( n = 1 \) and write \( y_t = (Y_t, X) \) as in a usual regression.
The cointegration parameter \( \beta \) can be estimated by OLS on

\[
Y_t = \beta X_t + u_t
\]

(21)

where \( u_t \sim I(0) \) by assumption.

\[
(\hat{\beta} - \beta) = \frac{\sum_{t=1}^T X_t u_t}{\sum_{t=1}^T X_t^2}.
\]

(22)

Since \( x_t \sim I(1) \) we are in a the same situation as with the first order AR case with autoregressive parameter equal to one.
The cointegrating regression II

In direct analogy, we need to multiply \((\hat{\beta} - \beta)\) by \(T\) in order to obtain a non-degenerate asymptotic distribution:

\[
T (\hat{\beta} - \beta) = \frac{1}{T} \sum_{t=1}^{T} X_t u_t,
\]

\[
\frac{1}{T^2} \sum_{t=1}^{T} X_t^2,
\]

\[
\rightarrow (\hat{\beta} - \beta) \text{ converges to zero at rate } T, \text{ instead of } \sqrt{T} \text{ as in the stationary case.}
\]

- This result is called the Engle-Granger super-consistency theorem.
- Remember that we assume \(r = 1\) so the cointegration vector is unique if it exists.
The distribution of the Engle-Granger (levels) estimator I

- Even with simple DGPs the E-G estimator is not normally distributed.
- The same applies to the $t$–value based on $\hat{\beta}$: It does not have a normal distribution
  $\implies$ Inference “in” the cointegration regression is generally impractical (because standard inference is not valid)
- This drawback is even more severe in DGPs with higher order dynamics, because the disturbance of the cointegrating equation is *autocorrelated* also in the case of cointegration.
Modified Engle-Granger estimator I

- Phillips and Hansen fully modified estimator:
  Subtract an estimate of the finite sample bias from $\hat{\beta}$ (i.e. keep the cointegration regression simple).
  The modified estimator has an asymptotic normal distribution, which allows inference on $\beta$.

- Saikonnen’s estimator,
  Is based on

$$Y_t = \beta X_t + \gamma_1 \Delta X_{t+1} + \gamma_2 \Delta X_{t-1} + u_t$$

or higher order lead/lags that “make” $u_t$ white-noise, see DM p 630.
ECM estimator I

The ECM represents a way of avoiding second order bias due to dynamic mis-specification.
This is because, under cointegration, the ECM is implied (the representation theorem).
With $n = 2$, $p = 1$ and weak exogeneity of $X_t (= Y_{2t})$ with respect to the cointegration parameter we have seen that the cointegrated VAR can be re-written as a conditional model and a marginal model

\[
\Delta Y_t = b \Delta X_t + \phi Y_{t-1} + \gamma X_{t-1} + \epsilon_t \quad (24)
\]

\[
\Delta X_t = \epsilon_{xt} \quad (25)
\]
where $b$ is the regression coefficient, and $\epsilon_t$ and $\epsilon_{xt}$ are uncorrelated normal variables (by regression).

\[
\Delta Y_t = b\Delta X_t + \phi(Y_{t-1} + \frac{\gamma}{\phi}X_{t-1}) + \epsilon_t
\]

\[
= b\Delta X + \phi(Y_{t-1} + \frac{\beta_{12}}{\beta_{11}}X_{t-1}) + \epsilon_t
\]

Normalization on $y_{t-1}$ by setting $\beta_{11} = -1$, and defining $\beta_{12} = \beta$, for comparison with E-G estimator, gives

\[
\Delta Y_t = b\Delta X_t + \phi(Y_{t-1} - \beta X_{t-1}) + \epsilon_t
\]
**ECM estimator III**

The ECM estimator $\hat{\beta}^{ECM}$, is obtained from OLS on (24)

$$\hat{\beta}^{ECM} = -\frac{\hat{\gamma}}{\hat{\phi}}$$

(26)

$\hat{\beta}^{ECM}$ is consistent if both $\hat{\gamma}$ and $\hat{\phi}$ are consistent.

OLS (by construction) chooses the $\hat{\gamma}$ and $\hat{\phi}$ that give the best predictor $y_{t-1} - \hat{\beta}^{ECM} x_{t-1}$ for $\Delta y_t$.

As $T$ grows towards infinity, the true parameters $\gamma$, $\phi$ and $\beta$ will therefore be found.

This is an example of *canonical correlation*, known from multivariate statistics.
ECM estimator IV

Therefore, by direct reasoning:

\[ \hat{\gamma} \xrightarrow{T \to \infty} \gamma, \hat{\phi} \xrightarrow{T \to \infty} \phi \text{ and } \hat{\beta}_{ECM} \xrightarrow{T \to \infty} \beta \]  

(27)

In fact:

- \( \hat{\beta}_{ECM} \) is super-consistent
- \( \hat{\beta}_{ECM} \) has better small sample properties than the E-G levels estimator, since it is based on a well specified econometric model (avoids the second-order bias problem).

Inference:

- The distributions of \( \hat{\gamma} \) and \( \hat{\phi} \) (under cointegration) can be shown to be so called “mixed normal” for large \( T \).
  - Their variances are stochastic variables rather than parameters.
ECM estimator V

- However, the OLS based t-values of $\hat{\gamma}$ and $\hat{\phi}$ are asymptotically $N(0, 1)$.
- $\hat{\beta}^{ECM}$ is also “mixed normal”, but

$$\left\{ \frac{\hat{\gamma}}{\phi} - \beta \right\} / \sqrt{\text{Var}(\hat{\beta}^{ECM})} \xrightarrow{T \to \infty} N(0, 1)$$

(28)

where, despite the change in notation, it is clear that $\text{Var}(\hat{\beta}^{ECM})$ can be by using the delta-method.

- The generalization to $n - 1$ explanatory variables, intercept and dummies is also unproblematic.
- Remember: The efficiency of the ECM estimator depends on the assumed weak exogeneity of $X_t$. 
Engle-Granger test

- The easiest approach is to use an ADF regression to the test null-hypothesis of a unit-root in the residuals $\hat{u}_t$ from the cointegrating regression (21).
- The motivation for the $\Delta\hat{u}_{t-j}$ is as before: to whiten the residuals of the ADF regression.
- The DF critical values are shifted to the left as deterministic terms, and/or more $I(1)$ variables in the regression are added.
The ECM test

- As we have seen, $r = 0$ corresponds to $\phi = 0$ in the ECM model in (24):

\[
\Delta Y_t = b\Delta X_t + \phi Y_{t-1} + \gamma X_{t-1} + \epsilon_t
\]

- It also comes as no surprise that the t-value $t_\phi$ have typical DF-like distributions under $H_0 : \phi = 0$.

- See DN and/or Ericsson and MacKinnon (2002) for critical values.
Why use ECM test instead of the Engle-Granger test? I

The size of the test (the probability of type 1 error) is more or less the same for the two tests. However, the power of the ECM test is generally larger than for the E-G test.

If \( t^{ECM}_\phi \) is the ECM test based on (24), it can be shown that

\[
\begin{align*}
t^{ECM}_\phi & \sim \frac{\sigma_e}{\sigma_{\epsilon}} t^{EG}_\tau, \\
& \quad \text{(29)}
\end{align*}
\]

where \( t^{EG} \) is the E-G test using

\[
\Delta \hat{u}_t = \tau \hat{u}_{t-1} + e_t \quad \text{(30)}
\]

The “t-values”, and therefore the power, will be equal when \( \sigma_e = \sigma_{\epsilon} \).
Why use ECM test instead of the Engle-Granger test? II

We can say something about when this will happen: Start with the ECM and bring it on ADL form:

\[ Y_t = bX_t + (1 + \phi)Y_{t-1} + (\gamma - b)X_{t-1} + \epsilon_t \]

\[ (1 - (1 + \phi)L)Y_t = (b + (\gamma - b)L)X_t + \epsilon_t \]

Assume next that the following *Common factor* restriction holds:

\[ \frac{(b + (\gamma - b)L)}{(1 - (1 + \phi)L)} = \beta \]  

so that

\[ b = \beta \]

\[ (\gamma - b) = -\beta(1 + \phi) \]
Why use ECM test instead of the Engle-Granger test? III

\[ Y_t = \beta X_t + (1 + \phi) Y_{t-1} - \beta (1 + \phi) X_{t-1} + \epsilon_t \quad (32) \]

\[ \Delta Y_t - \beta \Delta X_t = \phi (Y_{t-1} - \beta X_{t-1}) + \epsilon_t \]

If we replace $\beta$ by $\hat{\beta}$, we have

The ECM model (24) implies the Dickey-Fuller regression

\[ \Delta Y_t - \hat{\beta} \Delta X_t = \phi (y_{t-1} - \hat{\beta} X_{t-1}) + \epsilon_t \quad (33) \]

when the Common factor restriction in (31) is true.

- If the Common factor restriction is invalid, the E-G test is based on a mis-specified model.

- As a consequence $\sigma_e > \sigma_{\epsilon}$, and there is a loss of power relative to ECM test.
Testing cointegrating rank 1

For the vector $y_t$ consisting of $n \times 1$ variables, we have the Gaussian $VAR(p)$:

$$y_t = \Phi(L) y_{t-1} + \varepsilon_t \quad (34)$$

and use the re-parameterized equation:

$$\Delta y_t = \Phi^*(L) \Delta y_{t-1} + \Pi y_{t-1} + \varepsilon_t \quad (35)$$

We write the levels coefficient matrix $\Pi$ as the product of two matrices $\alpha_{n \times r}$ and $\beta'_{r \times n}$ where $r \equiv rank(\Pi)$:

$$\Pi = \alpha \beta' \quad (36)$$

We are interested in both the cointegrating case

$$0 < rank(\Pi) < n$$
Testing cointegrating rank II

and the case with no cointegration

\[ \text{rank}(\Pi) = 0 \]

- Since \( \text{rank}(\Pi) \) is given by the number of non-zero eigenvalues of \( \Pi \), one approach to testing is find the number of eigenvalues that are significantly different from zero.

- Fortunately, this problem has a solution because the eigenvalues has an interpretation as a special kind of squared correlation coefficients: \textit{canonical correlations}.

- This method has become known as the Johansen approach. It is likelihood based, see HN § 17.3.2
The Johansen method

Intuition I

- For concreteness, consider $n = 3$ so $r$ can be 0,1 or 2
- $r = 0$ corresponds to $\Pi = 0$ in the context of cointegration:
- From the representation theorem; with two unit-roots

$$\Pi = \Phi - I = P \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} P^{-1} - I = 0.$$  

- $r = 1$ corresponds to $\alpha_{3 \times 1} \neq 0$ for a single cointegration vector $\beta'_{1 \times 3}$.
- For this to make sense, $\beta'_{1 \times 3} y_{t-1}$ must be a $I(0)$ and it must be a significant predictor of $\Delta y_t$.  

The Johansen method

Intuition II

- The strength of the relationship can be estimated by the highest squared canonical correlation coefficient, call it $\hat{\rho}_1^2$, between $\Delta y_t$ and all the possible the linear combinations of the variables in $y_{t-1}$.
- If $\hat{\rho}_1^2 > 0$ is statistically significant, we reject that $r = 0$.
- $\hat{\rho}_1^2$ is the same as the highest eigenvalue of $\hat{\Pi}$, and $\hat{\beta}'_{1 \times 3}$ is the corresponding eigenvector.
- If $r = 0$ is rejected we can, continue, and test $r = 1$ against $r = 2$.
- If the second largest canonical correlation coefficient $\hat{\rho}_2^2$ is also significantly different from zero, we conclude that the number of cointegrating vectors is two. $\hat{\beta}'_{2 \times 3}$ is the corresponding eigenvector.
It can be shown that, for the Gaussian VAR, $\hat{\beta}_{1 \times 3}$ and $\hat{\beta}_{2 \times 3}$ are ML estimates.
Trace-test and max-eigenvalue test I

- We order the canonical correlations from largest to smallest and construct the so-called trace test:

\[
\text{Trace-test} = -T \sum_{i=r+1}^{3} \ln(1 - \hat{\rho}_i^2), \quad r = 0, 1, 2 \quad (37)
\]

- If \( \hat{\rho}_1^2 \) is close to zero, then clearly Trace-test will be close to zero, and we will not reject the \( H_0 \) of \( r = 0 \) against \( r \geq 1 \).

- And so on for \( H_0 \) of \( r = 1 \) against \( r \geq 2 \).

- Of course: to make this a formal testing procedure, we need the critical values from the distribution of the Trace-test for the sequence of null-hypotheses.
The distributions are non-standard, but at least the main cases are tabulated in PcGive

A closely related test is called the max-eigenvalue test, (but the trace test is today judged most reliable)

If there is a single cointegrating vector and there are $n - 1$ weakly-exogenous variables, the Johansen method reduces to the testing and estimation based on a single ECM equation (and OLS estimation as above)
Constant and other deterministic trends I

- It matters a great deal whether the constant is restricted to be in the cointegrating space or not.
- The advice for data with visible drift in levels:
  - include an deterministic trend as *restricted* together with an unrestricted constant.
  - After rank determination, can test significance of the restricted trend with standard inference.
- Shift in levels
  - Include restricted step dummy and a free impulse dummy.
- Exogenous I(1) variables, see table and program by MacKinnon, Haug and Michelis (1999).
I(0) variables in the VAR?

A misunderstanding that sometimes occurs is that “there can be no stationary variables in the cointegrating relationships”. Consider for example:

\[-Y_{1t} + \beta_{12} Y_{2t} + \beta_{13} Y_{3t} + \beta_{14} Y_{4t} = ecm_{1t} \quad (38)\]

\[\beta_{21} Y_{1t} - Y_{2t} + \beta_{23} Y_{3t} + \beta_{24} Y_{4t} = ecm_{2t} \quad (39)\]

If $Y_1$ is the log of real-wages, $Y_2$ productivity, $Y_3$ relative import prices, and $Y_4$ the rate of unemployment, then the first relationship may be a bargaining based wage and the second a mark-up equation.

$Y_{4t} \sim I(0)$, most sensibly, but we want to estimate and test the theory $\beta_{14} = 0$.

Hence: specify the VAR with $Y_{4t}$ included.
From $I(1)$ to $I(0)$

- When the rank has been determined, we are back in the stationary-case.
- The distribution of the identified cointegration coefficients are “mixed normal” so that conventional asymptotic inference can be performed on this $\hat{\beta}$.
- The determination of rank allows us to move from the $I(1)$ VAR, to the cointegrated VAR that contains only $I(0)$ variables.
- Another name for this $I(0)$ model is the *vector equilibrium correction model*, VECM.
- The VECM can be analysed further, using the tools of the stationary VAR!
- Hence, co-integration analysis is an important step in the analysis, but just one step.
Some important additional references

