Empirical Finance Lecture 9: Analysis of nonstationary processes II: Estimating and testing long-run relationships in systems of equations

Module Leader: Dr Stuart Fraser

stuart.fraser@wbs.ac.uk

Room D1.18 (Social Studies)



## Today

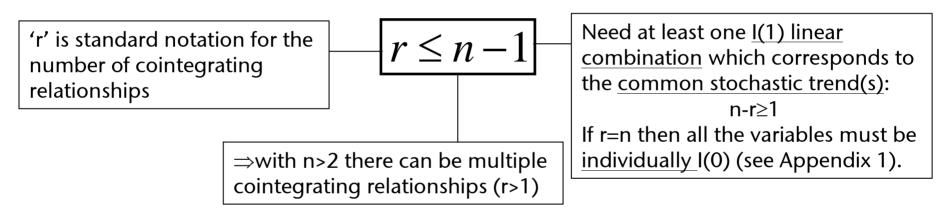
- 1. Key problems with Engle-Granger 2 Step Estimator.
- 2. VAR and VECM models.
- 3. 'Johansen': A systems approach to testing for cointegration.
  - Seminar 8: Testing PPP using Johansen

Problems with EG 2 step:

1. Presence of multiple cointegrating relationships. A key problem with EG 2 step is that it can only identify a <u>single cointegrating relationship</u>.

Single equation OLS is used to estimate the long-run (Step 1).

However with n variables there can be <u>up to n-1</u> cointegrating relationships (see Appendix 1):



If EG 2 Step is used in this context we may end up estimating some <u>unidentified linear combination</u> of all the cointegrating relationships.

## Example: Testing the Expectations Hypothesis of the term structure (see Brooks 7.12 and lecture 7).

The EH (+rational expectations) implies the yield spreads are cointegrated.

$$\Rightarrow R_t^{(j)} = R_t + T + \varepsilon_t^{(j)}, j = 2, \dots, n$$
$$\Rightarrow R_t^{(j)} - R_t^{(k)} \sim CI(1,1)$$

A test of the EH could be based on testing for cointegration amongst <u>pairs of yields</u> of different maturities.

However, a more comprehensive (and powerful) test could be based on a <u>vector of yields</u> from across the maturity spectrum

Normalizing the spreads on the one period spot rate, EH implies there are n-1 (linearly independent) spreads which are I(0):

$$\begin{array}{c|c}
R_{t} - R_{t}^{(j)} \sim CI(1,1), & j = 2, \dots, n \\
\hline R_{t} - R_{t}^{(j)} \sim CI(1,1), & j = 2, \dots, n \\
\hline \\
I & -1 & 0 & \dots & 0 \\
1 & 0 & -1 & \dots & 0 \\
\vdots & & & \vdots \\
1 & 0 & 0 & \dots & -1
\end{array}
\begin{pmatrix}
\beta \text{ is an } n \times (n-1) \text{ matrix.} \\
\text{The columns of } \beta (\text{rows of } \beta') \\
\text{are the cointegrating vectors.} \\
\sim CI(1,1) \Rightarrow \beta' y_{t} \sim CI(1,1)
\end{array}$$

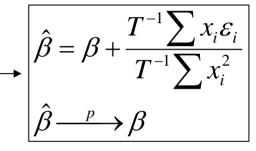
**Problems with EG 2 step:** 

2. EG 2 step is an inefficient estimator

On one level the distinction between exogenous and endogenous variables is unimportant with cointegrated variables.

Super-consistency still holds even if X is endogenous:

With cointegrated variables super-consistency still holds in the long run equation even if X is endogenous. Why? Because the sample variance of X tends to  $\infty$  (X is I(1)).



This is in sharp contrast to the CLRM where Y and X are assumed to be I(0). In that case endogeneity implies the OLS estimator is inconsistent (see Lecture 4).

So in effect it doesn't matter which variable (Y or X) we make the 'dependent variable':

$$Y_{t} = \mu + \beta X_{t} + \varepsilon_{t}$$
$$X_{t} = \mu^{*} + \beta^{*} Y_{t} + \varepsilon_{t}^{*}$$

Either regression will yield super-consistent estimates of the long run parameters.

For example, with PPP we 'normalized' the relationship on log(S) (log nominal exchange rate). However we might just as well have normalized on log(P) or log(P\*).

#### **Problems with EG 2 step: inefficiency**

However this also means we can write an ECM for X. Starting from the ADL for X...

$$X_{t} = \mu^{*} + \delta_{0}^{*}Y_{t} + \delta_{1}^{*}Y_{t-1} + \phi^{*}X_{t-1} + v_{t}^{*}$$

...the corresponding ECM for X is:

$$\Delta X_{t} = \delta_{0}^{*} \Delta Y_{t} + (\phi^{*} - 1)\varepsilon_{t-1}^{*} + v_{t}^{*}$$

Therefore Y and X form a system of ECMs:

$$\Delta Y_t = \delta_0 \Delta X_t + (\phi - 1)\varepsilon_{t-1} + v_t$$
  
$$\Delta X_t = \delta_0^* \Delta Y_t + (\phi^* - 1)\varepsilon_{t-1}^* + v_t^*$$

Both equations potentially contain information about the long-run parameters via the error correction term. An efficient information will use this information. Therefore a single equation approach (EG 2 Step) will in general be <u>inefficient</u>.

<u>There is an important exception</u>. If one of the adjustment parameters is zero then that equation contains <u>no information</u> about the long-run parameters. In that case the corresponding variable (e.g., X) is <u>weakly exogenous for the long-run</u> parameters – its equation can be ignored in estimating the long-run parameters.

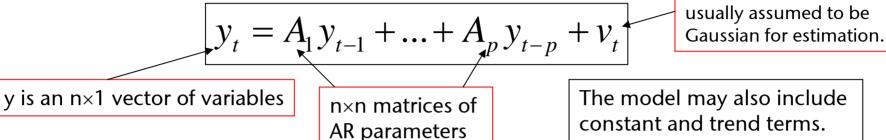
## **Problems with EG 2 Step**

Based on these problems with EG 2 Step, we would like an alternative estimator which

- 1. Can identify multiple cointegrating relationships.
- 2. Is an efficient estimator of the long-run parameters.
- This will lead us to look at the Johansen Systems Estimator shortly.
- But before that we need to look briefly at VAR models...

## From single to multiple equations: VAR

A popular time series (a-theoretical) model for systems of variables is a VECTOR AUTOREGRESSION (VAR)



- A VAR model provides a simple framework for estimating the dynamics of a system of endogenous variables. Specifically:
- All variables are treated as <u>endogenous</u> (each has its own equation) <u>no</u> assumptions of exogeneity are made.
- Since only lagged variables appear on the RHS there is no endogenous regressor problem in estimation – applying OLS equation by equation is a consistent (and efficient) estimator of the unrestricted coefficients (the A matrices).
- VARs are used widely in forecasting systems of variables (analogous to using AR models in univariate time series) and testing causality.

## Application of VAR models: testing Granger causality in a stationary VAR (Brooks 6.14).

stationary VAR (Brooks 6.14). X is said to Granger cause Y if lagged values of X affect Y. A VAR can be used to test this. For example:

$$\begin{array}{c} Y_{t} = \alpha_{11}Y_{t-1} + \ldots + \alpha_{1p}Y_{t-p} + \delta_{11}X_{t-1} + \ldots + \delta_{1p}X_{t-p} + v_{1t} \\ X_{t} = \alpha_{21}Y_{t-1} + \ldots + \alpha_{2p}Y_{t-p} + \delta_{21}X_{t-1} + \ldots + \delta_{2p}X_{t-p} + v_{2t} \end{array}$$
 Bivariate VAR(p) model

The null that X does not Granger cause Y is given by:

$$H_0: \delta_{11} = \ldots = \delta_{1p} = 0$$
 Test hypothesis with an F-test

Similarly the null that Y does not Granger cause X is given by:  $H_0: \alpha_{21} = \ldots = \alpha_{2p} = 0$  Test hypothesis with an F-test

The F tests are only valid in a stationary VAR (Y and X~I(0)).

- In a nonstationary VAR the parameters have non-standard distributions.
- In that case need to test causality in either:
  - A VAR in the I(0) differences of Y and X (if the variables are not cointegrated) or
  - A VECM (if the variables are cointegrated see below).

Warwick Business School

#### Application of VAR models: cointegration in a nonstationary VAR

The Granger Representation Theorem generalizes to systems of I(1) variables. If the system is cointegrated then there exists a: VECTOR ERROR CORRECTION MODEL (VECM)

From VAR(2) to VECM(1)  

$$y_{t} = A_{1}y_{t-1} + A_{2}y_{t-2} + v_{t}$$

$$\Rightarrow \Delta y_{t} = (A_{1} - I)y_{t-1} + A_{2}y_{t-2} + v_{t}$$

$$= (A_{1} - I)\Delta y_{t-1} + (A_{1} + A_{2} - I)y_{t-2} + v_{t} \quad \text{(subtract and add } (A_{1} - I)y_{t-2} \text{ on the RHS})$$

$$= \Gamma_{1}\Delta y_{t-1} + \Pi y_{t-2} + v_{t}$$

In general a cointegrated VAR(p) model has the following VECM(p-1) representation

$$\Delta y_{t} = \Gamma_{1} \Delta y_{t-1} + \dots + \Gamma_{p-1} \Delta y_{t-p+1} + \Pi y_{t-p} + v_{t},$$
  

$$\Gamma_{i} = (A_{1} + \dots + A_{i} - I)$$
  

$$\Pi = (A_{1} + \dots + A_{p} - I)$$

## Cointegrated VAR model: VECM

The matrix  $\Pi$  is central to the analysis. It is called the 'long-run matrix' because it defines the equilibrium solution to the system:

In equilibrium  $\Delta y_t = \Delta y_{t-1} = ... = 0$ . Therefore, in equilibrium,

the VECM yields the following solution for *y* 

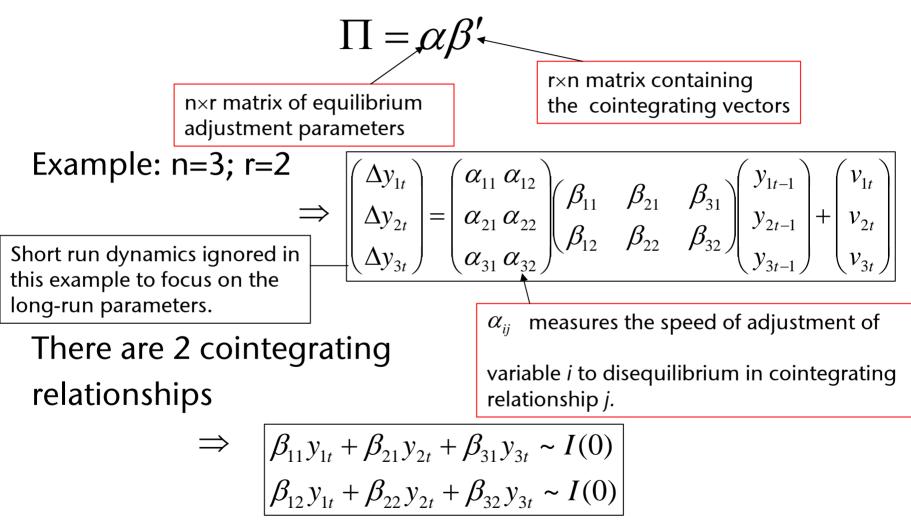
 $\Pi y^* = 0$ 

The <u>rank of  $\Pi$ </u> (number of linearly independent rows/columns) tells us how many (if any) cointegrating relationships there are in the system:

$rank(\Pi) = r$	Implications			
r = 0	$\Rightarrow \Pi$ is a null matrix (no long run relationships)			
	$\Rightarrow$ The model is a VAR in first differences			
0 < r < n	$\Rightarrow$ There are <i>r</i> cointegrating vectors			
r = n	$\Rightarrow$ The variables in the VAR are stationary in levels			
	(see Appendix 1: PROPOSITION 2)			

## **Cointegrated VAR model: VECM**

If  $rank(\Pi) = r$  then the long run matrix factorizes as



**VECM:** Weak exogeneity for the long run parameters

- Note that if the adjustment parameters  $\alpha$  are *all* different from zero then *each* equation contains information about the long-run parameters.
- However if row *i* of  $\alpha$  is null then the equation for variable *i* contains <u>no</u> information about the long-run parameters.
- In that case it is valid to drop equation *i* from the VECM when estimating the long run parameters  $\Rightarrow$ variable *i* is weakly exogenous for the longrun parameters (see slide 6 for a bivariate example).

## Weak exogeneity for $\beta$

Example: n=3; r=1

$$\begin{bmatrix} \Delta y_{1t} \\ \Delta y_{2t} \\ \Delta y_{3t} \end{bmatrix} = \begin{bmatrix} \alpha_{11} \\ 0 \\ 0 \end{bmatrix} (\beta_{11} \quad \beta_{21} \quad \beta_{31}) \begin{pmatrix} y_{1t-1} \\ y_{2t-1} \\ y_{3t-1} \end{pmatrix} + \begin{pmatrix} v_{1t} \\ v_{2t} \\ v_{3t} \end{pmatrix}$$
  
y<sub>2</sub> and y<sub>3</sub> are weakly exogenous  
for the long-run parameters.

For estimating  $\beta$  it is therefore valid to use a <u>single</u> equation estimator  $\langle v \rangle$ 

EG 2 step can be used to efficiently estimate this equation (given r=1).

$$\Delta y_{1t} = \alpha_{11} (\beta_{11} \quad \beta_{21} \quad \beta_{31}) \begin{pmatrix} y_{1t-1} \\ y_{2t-1} \\ y_{3t-1} \end{pmatrix} + \begin{pmatrix} v_{1t} \\ v_{2t} \\ v_{3t} \end{pmatrix}$$
$$\Rightarrow \Delta y_{1t} = \alpha_{11} \beta' y_{t-1} + v_{1t}$$

Note that if r>1 we would still need to use a systems estimator. EG 2 step cannot identify more than one cointegrating vector.

#### Systems estimator: Johansen Full Information Maximum Likelihood

The Johansen estimator provides a framework for:

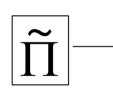
- 1. Estimating the cointegrating rank (rank of  $\Pi$ ) i.e., the number of long run relationships
- 2. Estimating the cointegrating vectors and adjustment parameters ( $\beta$  and  $\alpha$ )
- 3. Testing hypotheses about  $\beta$  and  $\alpha$  e.g.,
  - $\Rightarrow$  Testing PPP and EH restrictions on  $\beta$

 $\Rightarrow$ Testing weak exogeneity restrictions on  $\alpha$ 

In essence Johansen estimates all the distinct linear combinations of the <u>levels</u> y which produce <u>high correlations</u> with the <u>differences</u>  $\Delta y$ . These linear combinations are the cointegrating vectors.

## Johansen estimator: background

- In a sense the long-run matrix  $\Pi$  captures the correlation between linear combinations of the levels with the differences.
  - e.g., if  $\Pi$ =0 then there are <u>no</u> linear combinations of the levels which are correlated with the differences  $\Rightarrow$  no cointegration.
- To be precise the correlations are based on the matrix of squared correlations between the levels and differences:



This matrix is closely related to  $\Pi$  (see Appendix 2). Basically using this matrix (instead of  $\Pi$ ) ensures the correlations lie between 0 and 1.

Johansen uses the 'canonical correlations' which are given by the characteristic roots (eigenvalues) of  $\Pi$ 

These eigenvalues measure correlations between distinct (linearly independent) combinations of the levels with the differences.

The cointegrating vectors are given by the corresponding characteristic vectors (<u>eigenvectors</u>).

## Johansen estimator: background

Characteristic roots/eigenvalues of  $\tilde{\Pi}$ 

$$\widetilde{\Pi} = B \begin{pmatrix} \lambda_1 & \dots & 0 \\ \vdots & \dots & \vdots \\ 0 & \dots & \lambda_n \end{pmatrix} B' = B \Lambda B'$$
Rows of B' (eigenvectors) give the cointegrating vectors. These vectors are linearly independent. They capture distinct combinations of the levels which are I(0).
$$Diagonal \text{ matrix of eigenvalues} \\ (canonical correlations). These are ordered in descending value: \\ 1 \ge \lambda_1 \ge \lambda_2 \ge \dots \ge \lambda_n \ge 0$$

$$Rows of B' (eigenvectors) give the cointegrating vectors. These vectors are linearly independent. They capture distinct combinations of the levels which are I(0).
$$rank(\widetilde{\Pi}) = rank(\Pi) = number \text{ of non - zero eigenvalues}$$$$

If there are r linear combinations of the variables which are I(0) (r cointegrating vectors) then there are r positive eigenvalues; the remaining n-r equal zero.

### Johansen estimator: implementation

- Step 1: Ensure the variables in the system are individually I(1). Estimate a VAR of order p in the levels of the variables.
  - The Johansen estimator involves ML assuming Gaussian iid errors.
  - Therefore need to set p large enough to ensure a Gaussian iid error term in the VAR.
  - In practice the estimator is robust to non-normal errors.
  - But important that the errors are linearly independent (see Seminar 8).
- Step 2: In the VECM of order p–1 <u>estimate the cointegrating</u> <u>rank</u>, r, and the factorization

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

Step 3: Test hypotheses about the  $\alpha$  and  $\beta$  (see Seminar 8).

#### Johansen: estimating the cointegrating rank

- Tests of the cointegrating rank are based on the eigenvalues of  $\Pi$  (see slide 17). If rank( $\Pi$ )=r then:
  - The first r (largest) eigenvalues are non-zero.
  - The last n-r eigenvalues are zero:

$$\Rightarrow \log(1 - \lambda_j) = 0, \quad j = r + 1, \dots, n - \log(1) = 0$$

Johansen proposed two tests of cointegrating rank:

1. Maximum eigenvalue statistic:  $\begin{array}{c}
\text{Large test value } \Rightarrow \lambda_{r+1} \text{ is large} \\
\Rightarrow \text{ rejection of null} \\
\end{array}$   $\begin{array}{c}
H_0: rank(\Pi) \leq r, \\
H_1: rank(\Pi) \leq r, \\
H_1: rank(\Pi) > r
\end{array}$   $\begin{array}{c}
\text{Large test value } \Rightarrow \text{ the last one of the last n-r eigenvalues is large } \Rightarrow \text{ rejection of the null.} \\
\end{array}$   $\begin{array}{c}
\text{Large test value } \Rightarrow \text{ the last one of the last n-r eigenvalues is large } \Rightarrow \text{ rejection of the null.} \\
\end{array}$ 

Warwick Business School

#### Johansen: estimating the cointegrating rank

#### Test strategy

Start with  $H_0$  : r=0.

max-eigenvalue tests indicate r=1. This is in accord with PPP. r>1 would be hard to interpret. <u>Always</u> use economics/finance and stats to reach a 'sensible' conclusion about r.

Both the trace and

If this null is <u>not rejected</u> then  $\Rightarrow$  no cointegration.

If it is rejected then test  $H_0 : r \le 1$ .

If this null is not rejected then

 $\Rightarrow$ r=1 (since H<sub>0</sub> : r=0 was rejected)

If it is rejected then test  $H_0: r \le 2$ If this null is not rejected then  $\Rightarrow r=2$ 

If it is rejected then test  $H_0$ : r≤3...

Keep increasing the value of r until the null is not rejected.

#### Example: PPP - Denmark-US

Unrestricted Co	ointegration F	Rank Test (	Trace)			
Hypothesized		Trace	0.05			
No. of CE(s)	Eigenvalue	Statistic	Critical Va	Prob.**		
None *	0.053603	13 31718	29.79707	0.0008		
At most 1			15.49471			
At most 2	0.003877	2.272187	3.841466	0.1317		
Trace test indicates 1 cointegrating eqn(s) at the 0.05 level						
* denotes rejection of the hypothesis at the 0.05 level						
**MacKinnon-						
Unrestricted Cointegration Rank Test (Maximum Eigenvalue)						
Hypothesized		Max-Eigen	0.05			
No. of CE(s)	Eigenvalue	Statistic	Critical Va	Prob.**		
None *	0.053693	32.28499	21.13162	0.0009		
At most 1	0.014913	8.790006	14.2646	0.3041		
At most 2	0.003877	2.272187	3.841466	0.1317		
Max-eigenvalue test indicates 1 cointegrating eqn(s) at the 0						
* denotes rejection of the hypothesis at the 0.05 level						
**MacKinnon-Haug-Michelis (1999) p-values						

# Johansen: estimating the cointegrating vectors and adjustment parameters

The cointegrating vectors are estimated as the r eigenvectors of  $\Pi$  corresponding to the largest r

eigenvalues:

$$\tilde{\Pi} = B\Lambda B'$$

Estimates of cointegrating vectors correspond to the first r rows of B'

- These are the linear combinations of the levels of the variables which have the highest correlation with the differences.
- These linear combinations must be I(0) in order to be correlated with the I(0) differences (the correlation between I(1) and I(0) variables is zero ).
- The <u>adjustment parameters</u> ( $\alpha$ ) can then be estimated from a regression of  $\Delta y_t$  on  $\hat{\beta}' y_{t-p}$  (given  $\Delta y_{t-1},...,\Delta y_{t-p+1}$ )

#### Johansen: Issue of identification

The Johansen estimator does <u>not</u> identify the long-run parameters. Different combinations of  $\alpha$  and  $\beta$  give rise to the same  $\Pi$ :  $\Pi = \alpha \beta' = \alpha P P^{-1} \beta' - P^{\text{is any invertible rxr matrix}}$ 

Need to impose restrictions on the  $\beta$  for identification. If r=1 then only one restriction is required. For example:

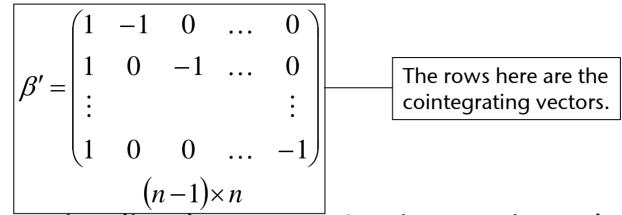
$$\beta' = (\beta_{11} \quad \beta_{21} \quad \beta_{31}), P = \beta_{11}$$
$$\Rightarrow P^{-1}\beta' = (1 \quad \beta_{21}/\beta_{11} \quad \beta_{31}/\beta_{11})$$

With r=1 it's sufficient to <u>normalize</u> the cointegrating vector on one of the variables for identification (e.g., normalize on log(S) in the PPP relationship)

However with r>1 then r linearly independent restrictions are required on <u>each</u> of the cointegrating vectors for identification:

- e.g., if r=2 a normalization <u>and</u> an exclusion restriction (setting one of the long-run coefficients to 0) would suffice in each vector.
- These restrictions should follow from economic/finance theory.

#### Johansen: example of identification The EH of the term structure implies $\beta$ is given by:



The theory implies there are n-1 cointegrating relationships (see slide 4).

The theory also provides the restrictions required to identify

- $\beta$ . For each cointegrating vector there is:
- A normalization on the one period spot yield (R). A homogeneity restriction (coefficient of -1 on  $R_t^{(j)}$ ).
- n-2 exclusion restrictions.

In total there are n restrictions on each cointegrating vector. In this case the long-run parameters are over-identified:

- Only n-1 restrictions are required for just identification of  $\beta$ .

## Conclusion

- Johansen's FIML estimator is a potentially powerful approach to estimating and testing cointegrated systems of I(1) variables.
- However, Johansen has numerous practical problems (aside from being quite hard to understand!)
  - The estimates of r are often sensitive to the choice of VAR order p.
  - Sometimes more cointegrating relationships are found than implied by economic/finance theory.
  - Also, the long run estimates cannot be interpreted without some underlying theory to help identify the parameters.
- You need <u>sound</u> economic/finance theory to help you decide on the cointegrating rank and to identify the long run parameters.
  - Without this basis in theory you will get lost applying Johansen (for sure).

## Remember this last point when applying Johansen in the project.

### References

Brooks (2002), Introductory econometrics for finance, CUP: Cambridge. Chps 6.14 & 7.9-7.13\*\*.

Verbeek (2004) A Guide to Modern Econometrics, 2nd Edition, Wiley: Chichester. Chps 9.4-9.5.

#### Appendix 1:

PROPOSITION 1: Amongst n I(1) variables there are up to n-1 stationary linear combinations.

Proof: Consider the case with n = 2. Suppose there are two linear combinations of I(1) variables which are I(0)

$$y - b_1 x = w \sim I(0)$$
$$y - b_2 x = v \sim I(0)$$

We can write w as

$$w = y - (b_1 - b_2)x - b_2 x$$
  
=  $v - (b_1 - b_2)x$ 

Since *w* is I(0) and *x* is I(1) therefore  $b_1 = b_2$  so there can be only n-1=1 I(0) linear combinations of x and y.

PROPOSITION 2: If there are n I(0) linear combinations amongst a group of n variables then all the variables must be *individually* I(0).

Proof: From the proof of PROPOSITION 1

$$w = y - (b_1 - b_2)x - b_2 x$$
  
=  $v - (b_1 - b_2)x$ 

Now if  $b_1 \neq b_2$  (so that there are n = 2 stationary linear combinations) then x must be I(0) since w is I(0). This also means that y must be I(0) since y is a linear combination of x and w.

#### Appendix 2:

#### Relationship between $\Pi$ (long-run matrix) and $\widetilde{\Pi}$ (squared correlation matrix)

An estimator of  $\Pi$  is given by

$$\hat{\Pi} = S_{0k} S_{kk}^{-1}$$

where:  $S_{0k}$  is the covariance matrix of the differences ('0') and levels ('k') conditional on the short run dynamics; and  $S_{kk}$  is the variance-covariance matrix of the levels conditional on the short-run dynamics.

The correlation matrix between the levels and differences is

$$P = S_{00}^{-0.5} S_{0k} S_{kk}^{-0.5}$$
$$= S_{00}^{-0.5} \hat{\Pi} S_{kk}^{0.5}$$

where  $S_{00}$  is the variance-covariance matrix of the differences conditional on the short run dynamics. The 'squared' correlations are given by

$$\widetilde{\Pi} = PP' = S_{00}^{-0.5} \widehat{\Pi} S_{kk} \widehat{\Pi}' S_{00}^{-0.5} = S_{00}^{-0.5} S_{0k} S_{kk}^{-1} S_{k0} S_{00}^{-0.5}$$

The Johansen estimator finds the eigenvalues of  $\Pi$  by finding the *n* characteristic roots of the determinant equation

$$\left|\lambda I - \widetilde{\Pi}\right| = 0$$

The cointegrating vectors are the r eigenvectors corresponding to the r non-zero solutions to this equation. These vectors give the r linear combinations of the levels which are correlated with the differences.