Estimation and Inference in Cointegration Models
Economics 582

Eric Zivot

May 17, 2012

Tests for Cointegration

Let the $(n \times 1)$ vector $\mathbf{Y}_t$ be $I(1)$. Recall, $\mathbf{Y}_t$ is cointegrated with $0 < r < n$ cointegrating vectors if there exists an $(r \times n)$ matrix $\mathbf{B}'$ such that

$$\mathbf{B}'\mathbf{Y}_t = \begin{pmatrix} \boldsymbol{\beta}_1'\mathbf{Y}_t \\ \vdots \\ \boldsymbol{\beta}_r'\mathbf{Y}_t \end{pmatrix} = \begin{pmatrix} u_{1t} \\ \vdots \\ u_{rt} \end{pmatrix} \sim I(0)$$

Testing for cointegration may be thought of as testing for the existence of long-run equilibria among the elements of $\mathbf{Y}_t$.

Cointegration tests cover two situations:

- There is at most one cointegrating vector
  - Originally considered by Engle and Granger (1986), *Econometrica*. They developed a simple two-step residual-based testing procedure based on regression techniques.
- There are possibly $0 \leq r < n$ cointegrating vectors.
  - Originally considered by Johansen (1988), *Journal of Economics Dynamics and Control*. He developed a sophisticated sequential procedure for determining the existence of cointegration and for determining the number of cointegrating relationships based on maximum likelihood techniques (covered in econ 584)

Residual-Based Tests for Cointegration

Engle and Granger's two-step procedure for determining if the $(n \times 1)$ vector $\boldsymbol{\beta}$ is a cointegrating vector is as follows:

- Form the cointegrating residual $\boldsymbol{\beta}'\mathbf{Y}_t = u_t$
- Perform a unit root test on u_t to determine if it is $I(0)$.

The null hypothesis in the Engle-Granger procedure is no-cointegration and the alternative is cointegration.

There are two cases to consider.

1. The proposed cointegrating vector $\boldsymbol{\beta}$ is pre-specified (not estimated). For example, economic theory may imply specific values for the elements in $\boldsymbol{\beta}$ such as $\boldsymbol{\beta} = (1, -1)'$. The cointegrating residual is then readily constructed using the prespecified cointegrating vector.

2. The proposed cointegrating vector is estimated from the data and an estimate of the cointegrating residual $\hat{\boldsymbol{\beta}}'\mathbf{Y}_t = \hat{u}_t$ is formed.

Note: Tests for cointegration using a pre-specified cointegrating vector are generally much more powerful than tests employing an estimated vector.

Testing for Cointegration When the Cointegrating Vector Is Pre-specified

Let $\mathbf{Y}_t$ denote an $(n \times 1)$ vector of $I(1)$ time series, let $\boldsymbol{\beta}$ denote an $(n \times 1)$ prespecified cointegrating vector and let

$$u_t = \boldsymbol{\beta}'\mathbf{Y}_t = \text{cointegrating residual}$$

The hypotheses to be tested are

$$\begin{aligned} H_0 &: u_t = \boldsymbol{\beta}'\mathbf{Y}_t \sim I(1) \text{ (no cointegration)} \\ H_1 &: u_t = \boldsymbol{\beta}'\mathbf{Y}_t \sim I(0) \text{ (cointegration)} \end{aligned}$$

Remarks:

- Any unit root test statistic may be used to evaluate the above hypotheses.
- Cointegration is found if the unit root test rejects the no-cointegration null.
- It should be kept in mind, however, that the cointegrating residual may include deterministic terms (constant or trend) and the unit root tests should account for these terms accordingly.

Testing for Cointegration When the Cointegrating Vector Is Estimated

Let $\mathbf{Y}_t$ denote an $(n \times 1)$ vector of $I(1)$ time series and let $\boldsymbol{\beta}$ denote an $(n \times 1)$ unknown cointegrating vector. The hypotheses to be tested are

$$\begin{aligned} H_0 &: u_t = \boldsymbol{\beta}'\mathbf{Y}_t \sim I(1) \text{ (no cointegration)} \\ H_1 &: u_t = \boldsymbol{\beta}'\mathbf{Y}_t \sim I(0) \text{ (cointegration)} \end{aligned}$$

Remarks:

- Since $\boldsymbol{\beta}$ is unknown, to use the Engle-Granger procedure it must be first estimated from the data.

- Before $\boldsymbol{\beta}$ can be estimated some normalization assumption must be made to uniquely identify it.

- A common normalization is to specify $\mathbf{Y}_t = (y_{1t}, \mathbf{Y}_{2t}')'$ where $\mathbf{Y}_{2t} = (y_{2t}, \ldots, y_{nt})'$ is an $((n-1) \times 1)$ vector and the cointegrating vector is normalized as $\boldsymbol{\beta} = (1, -\boldsymbol{\beta}_2')'$.

Engle and Granger propose estimating the normalized cointegrating vector $\boldsymbol{\beta}_2$ by least squares from the regression

$$
\begin{aligned}
y_{1t} &= \gamma' \mathbf{D}_t + \boldsymbol{\beta}_2' \mathbf{Y}_{2t} + u_t \\
\mathbf{D}_t &= \text{deterministic terms}
\end{aligned}
$$

and testing the no-cointegration hypothesis with a unit root test using the estimated cointegrating residual

$$\hat{u}_t = y_{1t} - \hat{\gamma}' \mathbf{D}_t - \hat{\boldsymbol{\beta}}_2 \mathbf{Y}_{2t}$$

The unit root test regression in this case is without deterministic terms (constant or constant and trend).

For example, if one uses the ADF test, the test regression is

$$\Delta \hat{u}_t = \pi \hat{u}_{t-1} + \sum_{j=1}^{p} \xi \Delta \hat{u}_{t-j} + error$$

Distribution Theory

- Phillips and Ouliaris (PO) (1990) show that ADF unit root tests (t-tests and normalized bias) applied to the estimated cointegrating residual *do not* have the usual Dickey-Fuller distributions under the null hypothesis of no-cointegration.

- Due to the spurious regression phenomenon under the null hypothesis, the distribution of the ADF unit root tests have asymptotic distributions that are functions of Wiener processes that

  – Depend on the deterministic terms in the regression used to estimate $\boldsymbol{\beta}_2$

  – Depend on the number of variables, $n-1$, in $\mathbf{Y}_{2t}$.

PO Critical Values for Engle-Granger Cointegration Test. (Constant included in test regression)

n-1	**1%**	**5%**
1	-3.89	-3.36
2	-4.29	-3.74
3	-4.64	-4.09
4	-4.96	-4.41
5	-5.24	-4.71

To further complicate matters, Hansen (1992) showed the appropriate PO distributions of the ADF and PP unit root tests applied to the residuals also depend on the trend behavior of y_{1t} and $\mathbf{Y}_{2t}$ as follows:

Case I: $\mathbf{Y}_{2t}$ and y_{1t} are both $I(1)$ without drift and $\mathbf{D}_t = 1$. The ADF unit root test statistics follow the PO distributions, adjusted for a constant, with dimension parameter $n - 1$.

Case II: $\mathbf{Y}_{2t}$ is $I(1)$ with drift, y_{1t} may or may not be $I(1)$ with drift and $\mathbf{D}_t = 1$. The ADF and PP unit root test statistics follow the PO distributions, adjusted for a constant and trend, with dimension parameter $n - 2$. If $n - 2 = 0$ then the ADF unit root test statistics follow the DF distributions adjusted for a constant and trend.

Case III: $\mathbf{Y}_{2t}$ is $I(1)$ without drift, y_{1t} is $I(1)$ with drift and $\mathbf{D}_t = (1, t)'$. The resulting ADF unit root test statistics on the residuals follow the PO distributions, adjusted for a constant and trend, with dimension parameter $n - 1$.

Regression-Based Estimates of Cointegrating Vectors and Error Correction Models

Least Square Estimator

Least squares may be used to consistently estimate a normalized cointegrating vector. However, the asymptotic behavior of the least squares estimator is non-standard. The following results about the behavior of $\hat{\boldsymbol{\beta}}_2$ if $\mathbf{Y}_t$ is cointegrated are due to Stock (1987) and Phillips (1991):

- $T(\hat{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}_2)$ converges in distribution to a non-normal random variable not necessarily centered at zero.

- The least squares estimate $\hat{\boldsymbol{\beta}}_2$ is consistent for $\boldsymbol{\beta}_2$ and converges to $\boldsymbol{\beta}_2$ at rate T instead of the usual rate $T^{1/2}$. That is, $\hat{\boldsymbol{\beta}}_2$ is *super consistent*.

- $\hat{\boldsymbol{\beta}}_2$ is consistent even if $\mathbf{Y}_{2t}$ is correlated with u_t so that there is no asymptotic simultaneity bias.

- In general, the asymptotic distribution of $T(\hat{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}_2)$ is asymptotically biased and non-normal. The usual OLS formula for computing $\widehat{avar}(\hat{\boldsymbol{\beta}}_2)$ is incorrect and so the usual OLS standard errors are not correct.

- Even though the asymptotic bias goes to zero as T gets large $\hat{\boldsymbol{\beta}}_2$ may be substantially biased in small samples. The least squres estimator is also not efficient.

The above results indicate that the least squares estimator of the cointegrating vector $\boldsymbol{\beta}_2$ could be improved upon. A simple improvement is suggested by Stock and Watson (1993).

Stock and Watson's Efficient Lead/Lag Estimator

Stock and Watson (1993) provide a very simple method for obtaining an asymptotically efficient (equivalent to maximum likelihood) estimator for the normalized cointegrating vector $\boldsymbol{\beta}_2$ as well as a valid formula for computing its asymptotic variance. Let

$$\begin{aligned}\mathbf{Y}_t &= (y_{1t}, \mathbf{Y}_{2t}')' \\ \mathbf{Y}_{2t} &= (y_{2t}, \ldots, y_{nt})' \\ \boldsymbol{\beta} &= (1, -\boldsymbol{\beta}_2')'\end{aligned}$$

Stock and Watson's efficient estimation procedure is:

- Augment the cointegrating regression of y_{1t} on $\mathbf{Y}_{2t}$ with appropriate deterministic terms $\mathbf{D}_t$ with p leads and lags of $\Delta\mathbf{Y}_{2t}$

$$\begin{aligned}y_{1t} &= \boldsymbol{\gamma}'\mathbf{D}_t + \boldsymbol{\beta}_2'\mathbf{Y}_{2t} + \sum_{j=-p}^{p} \boldsymbol{\psi}_j'\Delta\mathbf{Y}_{2t-j} + u_t \\ &= \boldsymbol{\gamma}'\mathbf{D}_t + \boldsymbol{\beta}_2'\mathbf{Y}_{2t} + \boldsymbol{\psi}_0'\Delta\mathbf{Y}_{2t} \\ &\quad +\boldsymbol{\psi}_{j+p}'\Delta\mathbf{Y}_{2t+p} + \cdots + \boldsymbol{\psi}_{j+1}'\Delta\mathbf{Y}_{2t+1} \\ &\quad +\boldsymbol{\psi}_{j-1}'\Delta\mathbf{Y}_{2t-1} + \cdots + \boldsymbol{\psi}_{j-p}'\Delta\mathbf{Y}_{2t-p} + u_t\end{aligned}$$

- Estimate the augmented regression by least squares. The resulting estimator of $\boldsymbol{\beta}_2$ is called the *dynamic OLS* estimator and is denoted $\hat{\boldsymbol{\beta}}_{2,DOLS}$. It will be consistent, asymptotically normally distributed and efficient (equivalent to MLE) under certain assumptions (see Stock and Watson (1993))

- Asymptotically valid standard errors for the individual elements of $\hat{\boldsymbol{\beta}}_{2,DOLS}$ are computed using the Newey-West HAC standard errors.

Estimating Error Correction Models by Least Squares

Consider a bivariate $I(1)$ vector $\mathbf{Y}_t = (y_{1t}, y_{2t})'$ and assume that $\mathbf{Y}_t$ is cointegrated with cointegrating vector $\boldsymbol{\beta} = (1, -\beta_2)'$ so that $\boldsymbol{\beta}'\mathbf{Y}_t = y_{1t} - \beta_2 y_{2t}$ is $I(0)$. Suppose one has a consistent estimate $\hat{\beta}_2$ (by OLS or DOLS) of the cointegrating coefficient and is interested in estimating the corresponding error correction model for Δy_{1t} and Δy_{2t} using

$$
\begin{aligned}
\Delta y_{1t} &= c_1 + \alpha_1(y_{1t-1} - \hat{\beta}_2 y_{2t-1}) \\
&\quad + \sum_j \psi_{11}^j \Delta y_{1t-j} + \sum_j \psi_{12}^j \Delta y_{2t-j} + \varepsilon_{1t} \\
\Delta y_{2t} &= c_2 + \alpha_2(y_{1t-1} - \hat{\beta}_2 y_{2t-1}) \\
&\quad + \sum_j \psi_{21}^j \Delta y_{1t-j} + \sum_j \psi_{22}^2 \Delta y_{2t-j} + \varepsilon_{2t}
\end{aligned}
$$

- Because $\hat{\beta}_2$ is super consistent it may be treated as known in the ECM, so that the estimated disequilibrium error $y_{1t} - \hat{\beta}_2 y_{2t}$ may be treated like the known disequilibrium error $y_{1t} - \beta_2 y_{2t}$.

- Since all variables in the ECM are $I(0)$, the two regression equations may be consistently estimated using ordinary least squares (OLS).

- Alternatively, the ECM system may be estimated by seemingly unrelated regressions (SUR) to increase efficiency if the number of lags in the two equations are different.

VAR Models and Cointegration

- The Granger representation theorem links cointegration to error correction models.

- In a series of important papers and in a marvelous textbook, Soren Johansen firmly roots cointegration and error correction models in a vector autoregression framework.

- This section outlines Johansen's approach to cointegration modeling.

The Cointegrated VAR

Consider the levels VAR(p) for the $(n \times 1)$ vector $\mathbf{Y}_t$

$$\begin{aligned} \mathbf{Y}_t &= \mathbf{\Phi}\mathbf{D}_t + \mathbf{\Pi}_1\mathbf{Y}_{t-1} + \cdots + \mathbf{\Pi}_p\mathbf{Y}_{t-p} + \boldsymbol{\varepsilon}_t, \\ \mathbf{D}_t &= \text{deterministic terms} \end{aligned}$$

The VAR(p) model is stable if

$$\det(\mathbf{I}_n - \mathbf{\Pi}_1 z - \cdots - \mathbf{\Pi}_p z^p) = 0$$

has all roots outside the complex unit circle.

- If there are roots on the unit circle then some or all of the variables in $\mathbf{Y}_t$ are $I(1)$ and they may also be cointegrated.

VECM Representation

- If Y_t is cointegrated then the VAR representation is not the most suitable representation for analysis because the cointegrating relations are not explicitly apparent.

- The cointegrating relations become apparent if the levels VAR is transformed to the *vector error correction model* (VECM)

$$\begin{gathered} \Delta\mathbf{Y}_t = \mathbf{\Phi}\mathbf{D}_t + \mathbf{\Pi}\mathbf{Y}_{t-1} + \mathbf{\Gamma}_1\Delta\mathbf{Y}_{t-1} \\ + \cdots + \mathbf{\Gamma}_{p-1}\Delta\mathbf{Y}_{t-p+1} + \boldsymbol{\varepsilon}_t \\ \mathbf{\Pi} = \mathbf{\Pi}_1 + \cdots + \mathbf{\Pi}_p - \mathbf{I}_n \\ \mathbf{\Gamma}_k = -\sum_{j=k+1}^{p} \mathbf{\Pi}_j,\ k = 1, \ldots, p-1 \end{gathered}$$

Cointegration Restrictions

- In the VECM, $\mathbf{\Delta Y}_t$ and its lags are $I(0)$.

- The term $\mathbf{\Pi Y}_{t-1}$ is the only one which includes potential $I(1)$ variables and for $\mathbf{\Delta Y}_t$ to be $I(0)$ it must be the case that $\mathbf{\Pi Y}_{t-1}$ is also $I(0)$. Therefore, $\mathbf{\Pi Y}_{t-1}$ must contain the cointegrating relations if they exist.

- If the VAR(p) process has unit roots ($z = 1$) then

$$\begin{aligned} &\det(\mathbf{I}_n - \mathbf{\Pi}_1 - \cdots - \mathbf{\Pi}_p) = 0 \\ &\Rightarrow \det(\mathbf{\Pi}) = 0 \\ &\Rightarrow \mathbf{\Pi} \text{ is singular} \end{aligned}$$

If $\mathbf{\Pi}$ is singular then it has *reduced rank*; that is $rank(\mathbf{\Pi}) = r < n$.

There are two cases to consider:

1. $rank(\mathbf{\Pi}) = 0$. This implies that

$$\begin{aligned} &\mathbf{\Pi} = \mathbf{0} \\ &\mathbf{Y}_t \sim I(1) \text{ and not cointegrated} \end{aligned}$$

The VECM reduces to a VAR($p-1$) in first differences

$$\mathbf{\Delta Y}_t = \mathbf{\Phi D}_t + \mathbf{\Gamma}_1 \mathbf{\Delta Y}_{t-1} + \cdots + \mathbf{\Gamma}_{p-1} \mathbf{\Delta Y}_{t-p+1} + \varepsilon_t.$$

2. $0 < rank(\mathbf{\Pi}) = r < n$. This implies that $\mathbf{Y}_t$ is $I(1)$ with r linearly independent cointegrating vectors and $n - r$ common stochastic trends (unit roots). Since $\mathbf{\Pi}$ has rank r it can be written as the product

$$\underset{(n\times n)}{\mathbf{\Pi}} = \underset{(n\times r)}{\boldsymbol{\alpha}} \underset{(r\times n)}{\boldsymbol{\beta}}'$$

where $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are $(n \times r)$ matrices with $rank(\boldsymbol{\alpha}) = rank(\boldsymbol{\beta}) = r$. The rows of $\boldsymbol{\beta}'$ form a basis for the r cointegrating vectors and the elements of $\boldsymbol{\alpha}$ distribute the impact of the cointegrating vectors to the evolution of $\Delta\mathbf{Y}_t$. The VECM becomes

$$\begin{aligned} \Delta\mathbf{Y}_t &= \mathbf{\Phi}\mathbf{D}_t + \boldsymbol{\alpha}\boldsymbol{\beta}'\mathbf{Y}_{t-1} + \mathbf{\Gamma}_1\Delta\mathbf{Y}_{t-1} \\ &\quad + \cdots + \mathbf{\Gamma}_{p-1}\Delta\mathbf{Y}_{t-p+1} + \boldsymbol{\varepsilon}_t, \end{aligned}$$

where $\boldsymbol{\beta}'\mathbf{Y}_{t-1} \sim I(0)$ since $\boldsymbol{\beta}'$ is a matrix of cointegrating vectors.

Non-uniqueness

It is important to recognize that the factorization $\mathbf{\Pi} = \boldsymbol{\alpha}\boldsymbol{\beta}'$ is not unique since for any $r \times r$ nonsingular matrix $\mathbf{H}$ we have

$$\boldsymbol{\alpha}\boldsymbol{\beta}' = \boldsymbol{\alpha}\mathbf{H}\mathbf{H}^{-1}\boldsymbol{\beta}' = (\mathbf{aH})(\boldsymbol{\beta}\mathbf{H}^{-1\prime})' = \mathbf{a}^*\boldsymbol{\beta}^{*\prime}.$$

Hence the factorization $\mathbf{\Pi} = \boldsymbol{\alpha}\boldsymbol{\beta}'$ only identifies the space spanned by the cointegrating relations. To obtain unique values of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}'$ requires further restrictions on the model.

Example: A bivariate cointegrated VAR(1) model

Consider the bivariate VAR(1) model for $\mathbf{Y}_t = (y_{1t}, y_{2t})'$

$$\mathbf{Y}_t = \mathbf{\Pi}_1 \mathbf{Y}_{t-1} + \boldsymbol{\epsilon}_t.$$

The VECM is

$$\begin{gathered}\Delta \mathbf{Y}_t = \mathbf{\Pi} \mathbf{Y}_{t-1} + \varepsilon_t \\ \mathbf{\Pi} = \mathbf{\Pi}_1 - \mathbf{I}_2\end{gathered}$$

Assuming $\mathbf{Y}_t$ is cointegrated there exists a 2×1 vector $\boldsymbol{\beta} = (\beta_1, \beta_2)'$ such that

$$\boldsymbol{\beta}' \mathbf{Y}_t = \beta_1 y_{1t} + \beta_2 y_{2t} \sim I(0)$$

Using the normalization $\beta_1 = 1$ and $\beta_2 = -\beta$ the cointegrating relation becomes

$$\boldsymbol{\beta}' \mathbf{Y}_t = y_{1t} - \beta y_{2t}$$

This normalization suggests the stochastic long-run equilibrium relation

$$y_{1t} = \beta y_{2t} + u_t$$

Since $\mathbf{Y}_t$ is cointegrated with one cointegrating vector, $rank(\mathbf{\Pi}) = 1$ so that

$$\mathbf{\Pi} = \boldsymbol{\alpha} \boldsymbol{\beta}' = \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix} \begin{pmatrix} 1 & -\beta \end{pmatrix} = \begin{pmatrix} \alpha_1 & -\alpha_1 \beta \\ \alpha_2 & -\alpha_2 \beta \end{pmatrix}.$$

The elements in the vector $\boldsymbol{\alpha}$ are interpreted as *speed of adjustment* coefficients. The cointegrated VECM for $\Delta \mathbf{Y}_t$ may be rewritten as

$$\Delta \mathbf{Y}_t = \boldsymbol{\alpha} \boldsymbol{\beta}' \mathbf{Y}_{t-1} + \varepsilon_t.$$

Writing the VECM equation by equation gives

$$\begin{aligned} \Delta y_{1t} &= \alpha_1 (y_{1t-1} - \beta y_{2t-1}) + \varepsilon_{1t}, \\ \Delta y_{2t} &= \alpha_2 (y_{1t-1} - \beta y_{2t-1}) + \varepsilon_{2t}. \end{aligned}$$

Stability of the VECM

- The stability conditions for the bivariate VECM are related to the stability conditions for the disequilibrium error $\boldsymbol{\beta}'\mathbf{Y}_t$.

It is straightforward to show that $\boldsymbol{\beta}'\mathbf{Y}_t$ follows an AR(1) process

$$\boldsymbol{\beta}'\mathbf{Y}_t = (\mathbf{1} + \boldsymbol{\beta}'\boldsymbol{\alpha})\boldsymbol{\beta}'\mathbf{Y}_{t-1} + \boldsymbol{\beta}'\boldsymbol{\varepsilon}_t$$

or

$$\begin{aligned} u_t &= \phi u_{t-1} + v_t,\ u_t = \boldsymbol{\beta}'\mathbf{Y}_t \\ \phi &= 1 + \boldsymbol{\beta}'\boldsymbol{\alpha} = 1 + (\alpha_1 - \beta\alpha_2) \\ v_t &= \boldsymbol{\beta}'\boldsymbol{\varepsilon}_t = u_{1t} - \beta u_{2t} \end{aligned}$$

The AR(1) model for u_t is stable as long as

$$|\phi| = |1 + (\alpha_1 - \beta\alpha_2)| < 1$$

For example, suppose $\beta = 1$. Then the stability condition is

$$|\phi| = |1 + (\alpha_1 - \alpha_2)| < 1$$

which is satisfied if

$$\alpha_1 - \alpha_2 < 0 \text{ and } \alpha_1 - \alpha_2 > -2.$$

Johansen's Methodology for Modeling Cointegration

The basic steps in Johansen's methodology are:

- Specify and estimate a VAR(p) model for $\mathbf{Y}_t$.
- Construct likelihood ratio tests for the rank of $\mathbf{\Pi}$ to determine the number of cointegrating vectors.
- If necessary, impose normalization and identifying restrictions on the cointegrating vectors.
- Given the normalized cointegrating vectors estimate the resulting cointegrated VECM by maximum likelihood.

Likelihood Ratio Tests for the Number of Cointegrating Vectors

The unrestricted cointegrated VECM is denoted $H(r)$. The $I(1)$ model $H(r)$ can be formulated as the condition that the rank of $\mathbf{\Pi}$ is less than or equal to r. This creates a nested set of models

$$\begin{aligned} H(0) &\subset \cdots \subset H(r) \subset \cdots \subset H(n) \\ H(0) &= \text{non-cointegrated VAR} \\ H(n) &= \text{ stationary VAR(p)} \end{aligned}$$

This nested formulation is convenient for developing a sequential procedure to test for the number r of cointegrating relationships.

Remarks:

- Since the rank of the long-run impact matrix $\mathbf{\Pi}$ gives the number of cointegrating relationships in $\mathbf{Y}_t$, Johansen formulates likelihood ratio (LR) statistics for the number of cointegrating relationships as LR statistics for determining the rank of $\mathbf{\Pi}$.

- These LR tests are based on the estimated eigenvalues $\hat{\lambda}_1 > \hat{\lambda}_2 > \cdots > \hat{\lambda}_n$ of the matrix $\mathbf{\Pi}$. These eigenvalues also happen to equal the squared *canonical correlations* between $\mathbf{\Delta Y}_t$ and $\mathbf{Y}_{t-1}$ corrected for lagged $\mathbf{\Delta Y}_t$ and $\mathbf{D}_t$ and so lie between 0 and 1. Recall, the rank of $\mathbf{\Pi}$ is equal to the number of non-zero eigenvalues of $\mathbf{\Pi}$.

Johansen's Trace Statistic

Johansen's LR statistic tests the nested hypotheses

$$H_0(r) : r = r_0 \text{ vs. } H_1(r_0) : r > r_0$$

The LR statistic, called the *trace statistic*, is given by

$$LR_{trace}(r_0) = -T \sum_{i=r_0+1}^{n} \ln(1 - \hat{\lambda}_i)$$

- If $rank(\Pi) = r_0$ then $\hat{\lambda}_{r_0+1}, \ldots, \hat{\lambda}_n$ should all be close to zero and $LR_{trace}(r_0)$ should be small since $\ln(1 - \hat{\lambda}_i) \approx 0$ for $i > r_0$.

- In contrast, if $rank(\mathbf{\Pi}) > r_0$ then some of $\hat{\lambda}_{r_0+1}, \ldots, \hat{\lambda}_n$ will be nonzero (but less than 1) and $LR_{trace}(r_0)$ should be large since $\ln(1 - \hat{\lambda}_i) << 0$ for some $i > r_0$.

- The asymptotic null distribution of $LR_{trace}(r_0)$ is not chi-square but instead is a multivariate version of the Dickey-Fuller unit root distribution which depends on the dimension $n - r_0$ and the specification of the deterministic terms. Critical values for this distribution are tabulated in Osterwald-Lenum (1992) for $n - r_0 = 1, \ldots, 10$.

Sequential Procedure for Determining the Number of Cointegrating Vectors

1. First test $H_0(r_0 = 0)$ against $H_1(r_0 > 0)$. If this null is not rejected then it is concluded that there are no cointegrating vectors among the n variables in $\mathbf{Y}_t$.

2. If $H_0(r_0 = 0)$ is rejected then it is concluded that there is at least one cointegrating vector and proceed to test $H_0(r_0 = 1)$ against $H_1(r_0 > 1)$. If this null is not rejected then it is concluded that there is only one cointegrating vector.

3. If the $H_0(r_0 = 1)$ is rejected then it is concluded that there is at least two cointegrating vectors.

4. The sequential procedure is continued until the null is not rejected.

Johansen's Maximum Eigenvalue Statistic

Johansen also derives a LR statistic for the hypotheses

$$H_0(r_0) : r = r_0 \text{ vs. } H_1(r_0) : r_0 = r_0 + 1$$

The LR statistic, called the maximum eigenvalue statistic, is given by

$$LR_{\max}(r_0) = -T \ln(1 - \hat{\lambda}_{r_0+1})$$

As with the trace statistic, the asymptotic null distribution of $LR_{\max}(r_0)$ is not chi-square but instead is a complicated function of Brownian motion, which depends on the dimension $n - r_0$ and the specification of the deterministic terms. Critical values for this distribution are tabulated in Osterwald-Lenum (1992) for $n - r_0 = 1, \ldots, 10$.

Specification of Deterministic Terms

Following Johansen (1995), the deterministic terms in are restricted to the form

$$\boldsymbol{\Phi}\mathbf{D}_t = \boldsymbol{\mu}_t = \boldsymbol{\mu}_0 + \boldsymbol{\mu}_1 t$$

If the deterministic terms are unrestricted then the time series in $\mathbf{Y}_t$ may exhibit quadratic trends and there may be a linear trend term in the cointegrating relationships. Restricted versions of the trend parameters $\boldsymbol{\mu}_0$ and $\boldsymbol{\mu}_1$ limit the trending nature of the series in $\mathbf{Y}_t$. The trend behavior of $\mathbf{Y}_t$ can be classified into five cases:

1. Model $H_2(r)$: $\boldsymbol{\mu}_t = 0$ (no constant):

$$\Delta\mathbf{Y}_t = \boldsymbol{\alpha}\boldsymbol{\beta}'\mathbf{Y}_{t-1}$$
$$+\boldsymbol{\Gamma}_1\Delta\mathbf{Y}_{t-1} + \cdots + \boldsymbol{\Gamma}_{p-1}\Delta\mathbf{Y}_{t-p+1} + \boldsymbol{\varepsilon}_t,$$

and all the series in $\mathbf{Y}_t$ are $I(1)$ without drift and the cointegrating relations $\boldsymbol{\beta}'\mathbf{Y}_t$ have mean zero.

2. Model $H_1^*(r)$: $\boldsymbol{\mu}_t = \boldsymbol{\mu}_0 = \boldsymbol{\alpha}\boldsymbol{\rho}_0$ (restricted constant):

$$\Delta\mathbf{Y}_t = \boldsymbol{\alpha}(\boldsymbol{\beta}'\mathbf{Y}_{t-1} + \boldsymbol{\rho}_0)$$
$$+\boldsymbol{\Gamma}_1\Delta\mathbf{Y}_{t-1} + \cdots + \boldsymbol{\Gamma}_{p-1}\Delta\mathbf{Y}_{t-p+1} + \boldsymbol{\varepsilon}_t,$$

the series in $\mathbf{Y}_t$ are $I(1)$ without drift and the cointegrating relations $\boldsymbol{\beta}'\mathbf{Y}_t$ have non-zero means $\boldsymbol{\rho}_0$.

3. Model $H_1(r)$: $\boldsymbol{\mu}_t = \boldsymbol{\mu}_0$ (unrestricted constant):

$$\Delta\mathbf{Y}_t = \boldsymbol{\mu}_0 + \boldsymbol{\alpha}\boldsymbol{\beta}'\mathbf{Y}_{t-1}$$
$$+\boldsymbol{\Gamma}_1\Delta\mathbf{Y}_{t-1} + \cdots + \boldsymbol{\Gamma}_{p-1}\Delta\mathbf{Y}_{t-p+1} + \boldsymbol{\varepsilon}_t$$

the series in $\mathbf{Y}_t$ are $I(1)$ with drift vector $\boldsymbol{\mu}_0$ and the cointegrating relations $\boldsymbol{\beta}'\mathbf{Y}_t$ may have a non-zero mean.

4. Model $H^*(r)$: $\boldsymbol{\mu}_t = \boldsymbol{\mu}_0 + \boldsymbol{\alpha}\boldsymbol{\rho}_1 t$ (restricted trend). The restricted VECM is

$$\Delta\mathbf{Y}_t = \boldsymbol{\mu}_0 + \boldsymbol{\alpha}(\boldsymbol{\beta}'\mathbf{Y}_{t-1} + \boldsymbol{\rho}_1 t)$$
$$+\boldsymbol{\Gamma}_1\Delta\mathbf{Y}_{t-1} + \cdots + \boldsymbol{\Gamma}_{p-1}\Delta\mathbf{Y}_{t-p+1} + \boldsymbol{\varepsilon}_t$$

the series in $\mathbf{Y}_t$ are $I(1)$ with drift vector $\boldsymbol{\mu}_0$ and the cointegrating relations $\boldsymbol{\beta}'\mathbf{Y}_t$ have a linear trend term $\boldsymbol{\rho}_1 t$.

5. Model $H(r)$: $\boldsymbol{\mu}_t = \boldsymbol{\mu}_0 + \boldsymbol{\mu}_1 t$ (unrestricted constant and trend). The unrestricted VECM is

$$\begin{aligned}\Delta\mathbf{Y}_t &= \boldsymbol{\mu}_0 + \boldsymbol{\mu}_1 t + \boldsymbol{\alpha}\boldsymbol{\beta}'\mathbf{Y}_{t-1} \\ &\quad +\boldsymbol{\Gamma}_1\Delta\mathbf{Y}_{t-1} + \cdots + \boldsymbol{\Gamma}_{p-1}\Delta\mathbf{Y}_{t-p+1} + \boldsymbol{\varepsilon}_t,\end{aligned}$$

the series in $\mathbf{Y}_t$ are $I(1)$ with a linear trend (quadratic trend in levels) and the cointegrating relations $\boldsymbol{\beta}'\mathbf{Y}_t$ have a linear trend.

Maximum Likelihood Estimation of the Cointegrated VECM

If it is found that $rank(\mathbf{\Pi}) = r$, $0 < r < n$, then the cointegrated VECM

$$\begin{aligned}\Delta\mathbf{Y}_t &= \mathbf{\Phi}\mathbf{D}_t + \boldsymbol{\alpha}\boldsymbol{\beta}'\mathbf{Y}_{t-1} + \mathbf{\Gamma}_1\Delta\mathbf{Y}_{t-1} \\ &\quad + \cdots + \mathbf{\Gamma}_{p-1}\Delta\mathbf{Y}_{t-p+1} + \boldsymbol{\varepsilon}_t,\end{aligned}$$

becomes a reduced rank multivariate regression. Johansen derived the maximum likelihood estimation of the parametes under the reduced rank restriction $rank(\mathbf{\Pi}) = r$ (see Hamilton for details).

Johansen shows that

- $\hat{\boldsymbol{\beta}}_{mle} = (\hat{\mathbf{v}}_1, \ldots, \hat{\mathbf{v}}_r)$,where $\hat{\mathbf{v}}_i$ are the eigenvectors associated with the eigenvalues $\hat{\lambda}_i$,

- The MLEs of the remaining parameters are obtained by least squares estimation of

$$\begin{aligned}\Delta\mathbf{Y}_t &= \mathbf{\Phi}\mathbf{D}_t + \boldsymbol{\alpha}\hat{\boldsymbol{\beta}}'_{mle}\mathbf{Y}_{t-1} + \mathbf{\Gamma}_1\Delta\mathbf{Y}_{t-1} \\ &\quad + \cdots + \mathbf{\Gamma}_{p-1}\Delta\mathbf{Y}_{t-p+1} + \boldsymbol{\varepsilon}_t,\end{aligned}$$

Normalized Estimates of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$

- The factorization

$$\hat{\mathbf{\Pi}}_{mle} = \hat{\boldsymbol{\alpha}}_{mle}\hat{\boldsymbol{\beta}}'_{mle}$$

is not unique

- The columns of $\hat{\boldsymbol{\beta}}_{mle}$ may be interpreted as linear combinations of the underlying cointegrating relations.

- For interpretations, it is often convenient to normalize or identify the cointegrating vectors by choosing a specific coordinate system in which to express the variables.

Johansen's normalized MLE

- An arbitrary normalization, suggested by Johansen, is to solve for the triangular representation of the cointegrated system (default method in Eviews). The resulting normalized cointegrating vector is denoted $\hat{\boldsymbol{\beta}}_{c,mle}$. The normalization of the MLE for $\boldsymbol{\beta}$ to $\hat{\boldsymbol{\beta}}_{c,mle}$ will affect the MLE of $\boldsymbol{\alpha}$ but not the MLEs of the other parameters in the VECM.

- Let $\hat{\boldsymbol{\beta}}_{c,mle}$ denote the MLE of the normalized cointegrating matrix $\boldsymbol{\beta}_c$. Johansen (1995) showed that

$$T(vec(\hat{\boldsymbol{\beta}}_{c,mle}) - vec(\boldsymbol{\beta}_c))$$

is asymptotically (mixed) normally distributed

- $\hat{\boldsymbol{\beta}}_{c,mle}$ is super consistent

Testing Linear Restrictions on β

The Johansen MLE procedure only produces an estimate of the basis for the space of cointegrating vectors. It is often of interest to test if some hypothesized cointegrating vector lies in the space spanned by the estimated basis:

$$\begin{aligned} H_0 &: \underset{(r\times n)}{\boldsymbol{\beta}}' = \begin{pmatrix} \boldsymbol{\beta}_0' \\ \boldsymbol{\phi}' \end{pmatrix} \\ \boldsymbol{\beta}_0' &= s \times n \text{ matrix of hypothesized cv's} \\ \boldsymbol{\phi}' &= (r-s) \times n \text{ matrix of remaining unspecified cv's} \end{aligned}$$

Result: Johansen (1995) showed that a likelihood ratio statistic can be computed, which is asymptotically distributed as a χ^2 with $s(n-r)$ degrees of freedom.