Cointegration

In economics we usually think that there exist long-run relationships between many variables of interest. For example, although consumption and income may each follow random walks, it seem reasonable to expect that there is a long run relationship between these variables, or in other words that in the long run these variables move together. The alternative scenario will be that Income increase relative to consumption with time. This seems implausible. Other series that appear to move together are: short term - long term interest rates, imports and exports, prices and wages, stock prices and dividends ,etc.

Then the aim behind cointegration is the detection and analysis of long run relationships amongst economic time series variables. Given that most economic time series, appear to be non-stationary, they often require differencing or detrending to be transformed to stationarity. A problem with differencing or detrending is that we may remove relevant long run information. The cointegration analysis provides a way of retaining both short-run and long-run information.

Another reason why we are concerned with cointegration is that sometimes is thought to be a pre-requisite for the validity of some economic theory. For example if short term and long term interest rates (assuming there are I(1)) are not cointegrated, then, the term structure of interest rates cannot hold.

Definitions

A linear combination of two I(1) variables, say Y and X can be either I(1) or I(0). If this combination is I(1) the variables are said to be not-cointegrated. If there are I(0) such that $Y + \alpha X \tilde{} I(0)$ then the variables are said to be cointegrated.

Example 1

Consider the following model:

$$x_t + \beta y_t = u_t \tag{1}$$

$$x_t + \alpha y_t = e_t \tag{2}$$

$$u_t = u_{t-1} + \varepsilon_{1t} \tag{3}$$

$$e_t = \rho e_{t-1} + \varepsilon_{2t} \quad \text{with } |\rho| < 1 \tag{4}$$

$(\varepsilon_{1t}, \varepsilon_{2t})'$ is distributed identically and independently as a bivariate normal with

$$E(\varepsilon_{1t}) = E(\varepsilon_{2t}) = 0 \tag{5}$$

$$var(\varepsilon_{1t}) = \sigma_{11},\ var(\varepsilon_{2t}) = \sigma_{22},\ cov(\varepsilon_{1t}\varepsilon_{2t}) = \sigma_{12} \tag{6}$$

Solving for x_t and y_t form the above system with $\alpha \neq \beta$ gives

$$\begin{aligned} x_t &= \alpha(\alpha-\beta)^{-1}u_t - \beta(\alpha-\beta)^{-1}e_t, \\ y_t &= -(\alpha-\beta)^{-1}u_t + (\alpha-\beta)^{-1}e_t. \end{aligned}$$

Then we can conclude that both x_t and y_t are integrated of order one i.e., $x_t \tilde{} I(1)$, $y_t \tilde{} I(1)$, since u_t is integrated of order one. Nonetheless $x_t + \alpha y_t$ is I(0) because e_t is stationary. In this example the cointegrating vector is (1,α) and $x + \alpha y$ is the equilibrium relationship. In the long run the variables move towards the equilibrium $x + \alpha y = 0$ recognizing that this relationship need not to be realized exactly even as t tends to infinity.

In the bivariate case if the equilibrium condition exists, is unique.

Proof.

Suppose that there exist two distinct co-integrating parameters α and γ such that $x + \alpha y$ and $x + \gamma y$ are both ~I(0). This implies that $(\alpha - \gamma)y_t$ is also I(0) because a linear combination of two I(0) variable is also I(0). But we know that for $\alpha \neq \gamma$, $(\alpha - \gamma)y_t$ ~I(1) therefore we have a contradiction unless $\alpha = \gamma$.

Example

Consider the following example where cointegration of Prices and Dividends is a necessary condition for markets efficiency in the Fama sense.

Let us assume that stock prices might be written as

$$P_t = \sum_{i=1}^{\infty} (1/(1+r))^i E(D_{t+i}|I_t) + \varepsilon_t$$

where we may assume that ε_t is an I(0) process that might be, a white noise if it represents, say a measurement error, or it might be an autoregressive process if we assume agents are risk adverse. Let also assume D_t follows a random walk which is a special case of an I(1) variable.

$$D_t = D_{t-1} + \nu_t.$$

Then, we may express stock prices as

$$P_t = (1/r)D_t + \varepsilon_t$$

Given that D_t are integrated of order one, stock prices also are integrated of order one, since the sum of an I(1) process and an I(0) process is I(1).

We can easily see that if the theory is valid, $(1, -(1/r))$ is going to be a cointegrating vector since

$$(1, -(1/r)) \begin{bmatrix} P_t \\ D_t \end{bmatrix} = Z_t = \varepsilon_t$$

Notice that we assumed that ε_t was I(0), therefore if the theory holds dividends and prices should be cointegrated.

Different Representations for a cointegrating relationship

Consider the model described in equations (1) - (6). Take $|\rho| < 1$, then whenever x_t and y_t are cointegrated, we can show, for the simple two variables model, that the system of two equations has different representations, namely, vector autoregressive in levels, error-correction and moving-average representations.

VAR Representation

Let us reproduce for expositional reasons equations (1), (2), (3) and (4)

$$x_t + \beta y_t = u_t \tag{1}$$

$$x_t + \alpha y_t = e_t \tag{2}$$

$$u_t = u_{t-1} + \varepsilon_{1t} \tag{3}$$

$$e_t = \rho e_{t-1} + \varepsilon_{2t} \quad \text{with } |\rho| < 1 \tag{4}$$

If we lag one period equations (1) and (2) and subtract the lagged value from each expression we get

$$\Delta x_t + \beta \Delta y_t = \Delta u_t \tag{7}$$

$$\Delta x_t + \alpha \Delta y_t = \Delta e_t \tag{8}$$

Notice that

$\Delta u_t = \varepsilon_{1t}$

$\Delta e_t = -(1-\rho)e_{t-1} + \varepsilon_{2t}$ and using equation (2)

$= -(1-\rho)(x_{t-1} + \alpha y_{t-1}) + \varepsilon_{2t}$

then (7) and (8) might be rewritten as

$$\begin{bmatrix} 1 & \beta \\ 1 & \alpha \end{bmatrix} \begin{bmatrix} \Delta x_t \\ \Delta y_t \end{bmatrix} = \begin{bmatrix} \varepsilon_{1t} \\ -(1-\rho)(x_{t-1} + \alpha y_{t-1}) + \varepsilon_{2t} \end{bmatrix}$$

and inverting the matrix we have

$$\begin{bmatrix} \Delta x_t \\ \Delta y_t \end{bmatrix} = \frac{1}{\alpha - \beta} \begin{bmatrix} \alpha & -\beta \\ -1 & 1 \end{bmatrix} \begin{bmatrix} \varepsilon_{1t} \\ -(1-\rho)(x_{t-1} + \alpha y_{t-1}) + \varepsilon_{2t} \end{bmatrix}$$

or

$$\begin{bmatrix} \Delta x_t \\ \Delta y_t \end{bmatrix} = \frac{1}{\alpha - \beta} \begin{bmatrix} \alpha\varepsilon_{1t} + \beta(1-\rho)(x_{t-1} + \alpha y_{t-1}) - \beta\varepsilon_{2t} \\ -\varepsilon_{1t} - (1-\rho)(x_{t-1} + \alpha y_{t-1}) + \varepsilon_{2t} \end{bmatrix} \tag{9}$$

or

$$\begin{bmatrix} \Delta x_t \\ \Delta y_t \end{bmatrix} = \frac{1}{\alpha - \beta} \begin{bmatrix} \beta(1-\rho) & \beta(1-\rho)\alpha \\ -(1-\rho) & -(1-\rho)\alpha \end{bmatrix} \begin{bmatrix} x_{t-1} \\ y_{t-1} \end{bmatrix} + \begin{bmatrix} \zeta_{1t} \\ \zeta_{2t} \end{bmatrix}$$

where
$\zeta_{1t} = (\alpha - \beta)^{-1}(\alpha\varepsilon_{1t} - \beta\varepsilon_{2t})$
$\zeta_{2t} = (\alpha - \beta)^{-1}(\varepsilon_{2t} - \varepsilon_{1t})$

Notice that this VAR representation IS NOT a VAR in the first differences. We would see later on that a VAR in first differences is not a possible representation when the variables are I(0). Notice that the var can be written as

$$\begin{bmatrix} x_t \\ y_t \end{bmatrix} = \frac{1}{\alpha - \beta} \begin{bmatrix} \beta(1-\rho)+1 & \beta(1-\rho)\alpha \\ -(1-\rho) & -(1-\rho)\alpha + 1 \end{bmatrix} \begin{bmatrix} x_{t-1} \\ y_{t-1} \end{bmatrix} + \begin{bmatrix} \zeta_{1t} \\ \zeta_{2t} \end{bmatrix}$$

Error Correction Mechanism

In the 70's the way to proceed when the series under consideration where I(1) was just to take differences of these series and then regress the differenced series. It has been shown that this procedure is unsatisfactory mainly because it looses all long run information. (Also the interpretation of the coefficients is different from that one of the original regression)

In recent years one of the most popular ways of proceeding is to write our models as ECM (error correction mechanisms). The strong motivation for this models was mainly empirical since these models perform very well. In later years it have been shown that if two variables are cointegrated, there exist an ECM representation for these variables. This representation has the advantage that it keeps long run and short run information.

Consider a vector of $I(1)$ variables X_t, then if $X_t \tilde{} CI(1,1)$ there exist an error-correction representation for the data. A very general way of writing these type of models is

$$\phi(L)(1-L)X_t = -\alpha' X_{t-1} + \theta(L)\varepsilon_t,$$

where α is the cointegrating vector, $\theta(L)$ is a polynomial in the lag operator, $\phi(L)$ is a finite order lag polynomial with roots outside the unit circle and ε_t is a white noise. Also $\phi(0) = I$.

Alternatively we could write the model as

$$\Delta X_t = -\alpha' X_{t-1} + \Phi(L)\Delta X_t + \theta(L)\varepsilon_t,$$

where $\phi(L) = I + \Phi(L)$.

The intuition behind the error-correction model is that long run errors have to be corrected in the short run dynamics such that the process can move closer to its long run target.

Again let $X_t = (X_{1t}, X_{2t})$ and let Z_{t-1} be the cointegrating relationship.
Then the error correction equation for X_{2t} is

$$\Delta X_{2t} = \gamma Z_{t-1} + \Phi_1(L)\Delta X_{1t-1} + \Phi_2(L)\Delta X_{2t-1} + \theta(L)\nu_t$$

Then if the error correction theory is true we will expect that $\gamma < 0$ which implies that whenever $Z_{t-1} > 0$, *i.e.*, X_{2t} is above the long run equilibrium, then, ΔX_{2t} will be negative, *i.e.*, it will move in the direction of the equilibrium.

An important implication of the theory of cointegration is that asset prices cannot be cointerated if the weak efficiency market hypothesis holds. This is simply because if two asset prices are cointegrated, it is possible to forecast time t returns using information available a t time $t-1$. This goes against the simple no-arbitrage model.

ECM representation.

Consider once more the two variables model. It can be seen that the ECM representation follows directly from the VAR representation (equation (9)) .

$$\begin{bmatrix} \Delta x_t \\ \Delta y_t \end{bmatrix} = \frac{1}{\alpha - \beta} \begin{bmatrix} \beta(1-\rho)(e_{t-1}) \\ -(1-\rho)(e_{t-1}) \end{bmatrix} + \begin{bmatrix} \zeta_{1t} \\ \zeta_{2t} \end{bmatrix} \tag{9'}$$

Moving Average Representation
This follows directly from the VAR representation (1) and (2).
Equations (1) and (2) may be written in matrix notation as

$$\begin{bmatrix} 1 & \beta \\ 1 & \alpha \end{bmatrix} \begin{bmatrix} x_t \\ y_t \end{bmatrix} = \begin{bmatrix} u_t \\ e_t \end{bmatrix},$$

which can be also written as

$$\begin{bmatrix} x_t \\ y_t \end{bmatrix} = \frac{1}{\alpha - \beta} \begin{bmatrix} \alpha & -\beta \\ -1 & 1 \end{bmatrix} \begin{bmatrix} u_t \\ e_t \end{bmatrix},$$

or taking first diferences as

$$\begin{bmatrix} \Delta x_t \\ \Delta y_t \end{bmatrix} = \frac{1}{\alpha - \beta} \begin{bmatrix} \alpha & -\beta \\ -1 & 1 \end{bmatrix} \begin{bmatrix} \Delta u_t \\ \Delta e_t \end{bmatrix},$$

and noting that

$$\begin{bmatrix} \Delta u_t \\ \Delta e_t \end{bmatrix} = \begin{bmatrix} \varepsilon_{1t} \\ (1-L)(1-\rho L)^{-1}\varepsilon_{2t} \end{bmatrix},$$

we can obtain the MA representation as

$$\begin{bmatrix} \Delta x_t \\ \Delta y_t \end{bmatrix} = \frac{1}{\alpha - \beta} \begin{bmatrix} \alpha & -\beta \\ -1 & 1 \end{bmatrix} \begin{bmatrix} \varepsilon_{1t} \\ (1-L)(1-\rho L)^{-1}\varepsilon_{2t} \end{bmatrix}.$$

We have analyzed the possible representations assuming a cointegration relationship between two variables. The concept of cointegration can be easily extended to n-variables.

Cointegration between n variables

Definition

Consider a n×1 vector of stochastic variables $X_t = (X_{1t}, X_{2t}, .., X_{n_t})$. We say that the elements of the vector are cointegrated of order (d, b), which we denote $X_t \sim \text{CI}(d, b)$ if

i) each of the components of X_t are I(d).

ii) there exists (at least) a vector such that $Z_t = \alpha' X_t$ is I$(d-b)$ for $d \geq b > 0$.

Then, α is called the cointegrating vector

If $d = b = 0$, then $\alpha' X_t = 0$ defines a long-run equilibrium relationship.

Notice that α is not unique since for any non-zero scalar b, $b\ \alpha' X_t$ is also integrated of order zero. Then in speaking of the cointegrating vector, an arbitrary normalization must be made, such that the first element of is unity.

If $n > 2$, there may be $r \leq n-1$ linearly independent $(n \times 1)$ vectors $(\alpha_1, ..., \alpha_r)$ such that $A' X_t \sim \text{I}(0)$, where A is the $(n \times r)$ matrix $A = [\alpha_1, ..., \alpha_r]$, such that rank$(A)$=$r$, where r is the cointegrating rank.

The vectors $(\alpha_1, ..., \alpha_r)$ are not unique, since for any non-zero $(r \times 1)$ vector b, $b' A' X_t \sim \text{I}(0)$, so $b' A'$ could be described as a cointegrating matrix.

Granger's Representation Theorem

In the two variables case we have shown that when two variables are cointegrated they do have a VAR, an ECM and a MA representation. These were particular cases of the Granger's Representation Theorem which is valid for a N variables vector.

Consider an n - vector time series X_t which satisfies:

$$\Phi(L) X_t = c + u_t$$

where $\Phi(L) = I_n - \sum_{i=1}^{P} \Phi_i L^i$, and u_t is a white noise with positive definite covariance matrix. It is assumed that det$[\Phi(z)]$= 0 which implies $|z| \geq 1$. Suppose that there exist exactly r cointegrating relationships among the elements of X_t . Then:

(*i*) there exists an $(n \times r)$ matrix A, of rank $r < n$ such that $A' X_t \sim \text{I}(0)$.

(*ii*) ΔX_t has an MA representation given by $\Delta X_t = \mu + \Psi(L) u_t$ with

$$A' \Psi(1) = 0,$$

where

$$\Psi(L) = I_n + \sum_{i=1}^{\infty} \Psi_i L^i$$

To understand the meaning of the restriction $A' \Psi(1) = 0$ and its implications consider the MA

$$\Delta X_t = \mu + \Psi(L)u_t$$

Now re-write it as $X_t = X_{t-1} + \mu + \Psi(L)u_t$ and substitute backwards to obtain

$$X_t = X_0 + \mu t + \Psi(1)\sum_{i=1}^{t} u_i + \eta_t - \eta_0$$

where $\mu = E(\Delta X_t)$ and $\{\eta_t\}$ is a I(0) sequence: $\eta_t = \sum_{S=0}^{\infty} a_s u_{t-s}, \qquad a_s = -\sum_{i=1}^{\infty} \Psi_{s+i}$

Then, pre-multiplying X_t by the cointegrating matrix A', we get the cointegrating relationship. Therefore each term in the right hand side has to be I(0).

$$A'X_t = A'(X_0 - \eta_0) + A'\mu t + A'\Psi(1)\sum_{i=1}^{t} u_i + A'\eta_t$$

Then, it is easy to see that for $A'X_t$ ~I(0), it is necessary that $A'\Psi(1) = 0$. This is only a necessary condition. Stationarity of $A'X_t$ further requires that $A'\mu = 0$. If $\mu \neq 0$, then unless $A'\mu = 0$ is satisfied, the linear combination $A'X_t$ will grow deterministically at rate $A'\mu$.

Notice that $A'\Psi(1) = 0$ implies that $\Psi(1)$ is singular since A' is a matrix with r LI vectors. This means that the matrix operator $\Psi(L)$ is non invertible. Thus, a cointegrated system can never be represented by a finite-order VAR for ΔX_t.

(iii) $\Phi(1) = BA'$

Theorem 1

The $(n \times n)$ matrix $\Phi(1)$ has a reduced rank $r < n$, and there exists an $(n \times r)$ matrix B such that $\Phi(1) = BA'$.

Proof. Consider an n-vector time series X_t which satisfies:

$$\Phi(L)X_t = c + u_t$$

and the Wold representation

$$\Delta X_t = \mu + \Psi(L)u_t.$$

Then multiplying the Wold representation by $\Phi(L)$, we get

$$(1-L)\Phi(L)X_t = \Phi(1)\mu + \Phi(L)\Psi(L)u_t$$

and using the autoregressive representation (multiplied by$(1-L)$) we get

$$(1-L)\Phi(L)X_t = (1-L)(c + u_t)$$

and equating the last two expressions we get

$$(1-L)u_t = \Phi(1)\mu + \Phi(L)\Psi(L)u_t$$

(since $(1-L)c = 0$) then, the above expresion implies that:
1) $\Phi(1)\mu = 0$.
2) $(1-L) = \Phi(L)\Psi(L)$, in particular for $L = 1$, requires that $\Phi(1)\Psi(1) = 0$.

Let π' denote a row of $\Phi(1)$, then conditions 1) and 2) imply that π is a cointegrating vector.

Now if $a_1, a_2, ...a_r$ form a basis for the space of cointegrating vectors, then we can express, $\pi_{nx1} = (a_1, a_2, ...a_r)_{nxr} b_{r\times 1}$, since any linear combination of a cointegrating vector is also a cointegrating vector.or $\pi' = b'A'$ Applying the same reasoning for each of the rows of the rows we get

$$\Phi(1) = BA'$$

■

(*iv*) There exist a VAR representation in Levels and the determinant of the polynomial in Φ has a unit root. This can be shown by noticing that $\Phi(1)$ is singular.

(*v*) *Error Correction Representation.*

To show this point we first need to transform the original VAR

$$\Phi(L)X_t = c + u_t,$$

using the following relationship

$$I_n - \sum_{i=1}^{p} \Phi_i L^i = I_n - (\sum_{i=1}^{P} \Phi_i)L - (I-L)\sum_{i=1}^{p-1} \Gamma_i L^i$$

where
$\Gamma_i = -\sum_{i=j}^{p-1} \Phi_{j+1}$ for $j = 1, ..., p-1$
Then

$$(I_n - \sum_{i=1}^{p} \Phi_i L^i)X_t = \left(I_n - \left(\sum_{i=1}^{p} \Phi_i\right)L + (I-L)\sum_{i=1}^{p-1} \Gamma_i L^i\right)\right) X_t = c + u_t.$$

Then rearranging terms

$$\begin{aligned} X_t &= c + (\sum_{i=1}^{p} \Phi_i)X_{t-1} + \sum_{i=1}^{p-1} \Gamma_i \Delta X_{t-i} + u_t \\ &= c + \sum_{i=1}^{p-1} \Gamma_i \Delta X_{t-i} + (I_n - \Phi(1))\, X_{t-1} + u_t \end{aligned}$$

$$\Delta X_t = c + \sum_{i=1}^{p-1} \Gamma_i \Delta X_{t-i} - \Phi(1) X_{t-1} + u_t$$

From this expression we can derive the following results.

(a) If rank $[\Phi(1)] = 0$ then $\Phi(1) = 0$ and $X_t \tilde{} I(1)$ since we can write everything in terms of a VAR in diferences. This is only valid when the variables do not cointegrate.

(b) If rank $[\Phi(1)] = n$ then $det(\Phi(1)) \neq 0$ ($\Phi(L)$ does not have a unit root). This implies that $X_t \tilde{} I(0)$ and therefore it should have a MA representation since we can invert the original expression, *i.e.*,

$$X_t = \Phi(L)^{-1}(c + + u_t)$$

(c) If rank $[\Phi(1)] = r$, $0 < r < n$ then $\Phi(1) = BA'$, where B is an nxr matrix .

(See a explanation of these in the section about Johansen Procedure).

Restrictions in the parameters of the Error Correction Representation.

$$\Delta X_t = c + \sum_{i=1}^{p-1} \Gamma_i \Delta X_{t-i} - B Z_{t-1} + u_t, \; Z_t = A' X_t$$

In this error-correction representation all variables are $I(0)$. Then the term $A'X_t$ is viewed as the "error" from the long run equilibrium relationship, and B gives the "correction" to X_t caused by this error.

Notice that taking expected values in both sides we get;

$$[I - \sum_{i=1}^{p-1} \Gamma_i L^i] E(\Delta X_t) = c - B E(Z_{t-1})$$

Thus, in order to have a system in which there is no drift in any of the variables [i.e., $E(\Delta X_t) = 0$], we need to impose the restriction $c = BE(Z_{t-1})$ (this is equivalent to $A'\mu = 0$) In the absence of such a restriction, it is implied by the ECM that there are $n - r$ separate time trends that account for the trend X_t .

Then imposing this restriction we obtain

$$\begin{aligned} \Delta X_t &= BE(Z_{t-1}) + \sum_{i=1}^{p-1} \Gamma_i \Delta X_{t-i} + -B Z_{t-1} + u_t \\ &= -B(Z_{t-1} - E(Z_{t-1})) + \sum_{i=1}^{p-1} \Gamma_i \Delta X_{t-i} + u_t \end{aligned}$$

the intercept enters the system only via the error-correction term and there is no autonomous growth component.

Again notice that A is not unique since $\Phi(1) = BA' = BPP^{-1}A' = B^*A^{*\prime}$, for all $r \times r$ non-singular matrices P.

Tests For Cointegration.

Univariate Tests for Cointegration.

1) Two Stages Approaches

First stage: A static Regression
For the model

$$y_t = \beta x_t + \varepsilon_t$$

Regress y on x using OLS achieves a consistent estimate of the long-run steady -state relationship between the variables of the model, and all dynamics and endogeneity issues can be ignored asymptotically. This arises because of the super consistency property of the OLS estimator when the series are cointegrated.

Suppose that the model that dynamic model that captures both short adjustment and the long run relationship is

$$y_t = \gamma_o x_t + \gamma_1 x_{t-1} + \alpha y_{t-1} + \varepsilon_t$$

This can be re-written as

$$y_t = \lambda_0 x_t + \lambda_1 \Delta x_t + \lambda_2 \Delta y_t + \varepsilon_t$$

where $\lambda_0 = \frac{\gamma_0+\gamma_1}{1-\alpha}$, $\lambda_1 = \frac{-\gamma_1}{1-\alpha}$, $\lambda_2 = \frac{-\alpha}{1-\alpha}$.

Thus, estimating the static model to obtain the long-run parameter β is equivalent to estimating the dynamic model without the short run terms. According to the super consistency property if y_t and x_t are both non-stationary I(1) variables, and ε_t ~I(0), then as the sample size, T, becomes larger the OLS estimator converges to its true value at a much faster rate than the I(0) variables. Of course the omitted dynamic terms are captured by the residuals. This we will see later will be a problem in short samples.

As a second stage I can use alternative testing stragesies:

i) **Make an (ADF) test for unit roots for the residuals:***: The Engle - Granger Approach.*

If you do not reject the Hypothesis that the residuals have a unit root (that there are I(1)), then Y and X are not cointegrated. On the other hand if you do reject this hypothesis against the stationary alternative, then you conclude that the residuals (a linear combination of Y and X) are I(0), then Y and X are cointegrated.

Therefore we regress

$$\Delta\hat{\varepsilon}_t = \phi\hat{\varepsilon}_{t-1} + \sum_{i=1}^{p-1} \phi_i \Delta\hat{\varepsilon}_{t-i} + \mu + \delta t + \zeta_t$$

The question of including a trend and or a constant term in the test regression depends of whether this terms appear in the original regression. That is the deterministic component can be added either to the static regression or to the ADF regression, but not to both. Hansen (1992) has shown that including a deterministic trend results in loss of power.[1]

Note that is not possible to use standard Dickey - Fuller critical values. The standard table will tend to over reject the null of no cointegration. Note also that the distribution of the test statistic under the null is affected by the number of regressors included in the static model.

Fortunately MacKinon (1991) provides Table which has linked the critical values for particular tests to a set of parameters of an equation of response surfaces.

$$C(p) = \phi_\infty + \phi_1 T^{-1} + \phi_2 T^{-2}$$

Response surfaces for critical values of cointegration tests.

[1] Notice that if we consider the simple DF test for the residuals

$$\Delta\hat{\varepsilon}_t = \phi\hat{\varepsilon}_{t-1} + \zeta_t$$

This can be re-written (evaluating at $\beta = \hat{\beta}$)

$$\Delta(y_t - \beta x_t) = \phi(y_{t-1} - \beta x_{t-1}) + \zeta_t$$

or

$$\Delta y_t = \beta\Delta x_t + \phi(y_{t-1} - \beta x_{t-1}) + \zeta_t.$$

This model is not an unrestricted ECM and it imposes that the change in the short run is the same that the long run effect. This is unlikely to be true.

n	Model	%point	ϕ_∞	ϕ_1	ϕ_2
1	no C	1	-2.5658	-1.960	-10.04
	no t	5	-1.9393	-0.398	0.0
		10	-1.6156	-0.181	0.0
1	C	1	-3.4336	-5.999	-29.25
	no t	5	-2.8621	-2.738	-8.36
		10	-2.5671	-1.438	-4.48
1	C	1	-3.9638	-8.353	-47.44
	t	5	-3.4126	-4.039	-17.83
		10	-3.1279	-2.418	-7.58
3	C	1	-4.2981	-13.790	-46.37
	no t	5	-3.7429	-8.352	-13.41
		10	-3.4556	-6.241	-2.79

where $C(p)$ is the p per cent critical value.

The Cointegrating Durbin-Watson.

ii) **The null may be tested using the Sargan-Bhargava or CDW test**

This test is very simple and consist in comparing the DW statistic with tabulated values. The rationality of this procedure is as follows: Consider equations (2) and (4).

$$x_t + \alpha y_t = e_t \tag{2}$$

$$e_t = \rho e_{t-1} + \varepsilon_{2t} \quad \text{with } |\rho| < 1 \tag{4}$$

A regression of x_t on y_t will yield serially correlated residuals. We can use the DW statistic to get information about ρ since this statistic is approximately $2(1-\rho)$.

$$DW \cong 2(1-\rho)$$

Then if ρ is equal to 1 (a unit root), the DW statistic equals zero. Sargan and Bhargava provide tables to test this hypothesis. However, this critical value is only relevant when the disturbance follows a first order autoregressive process and there is there no higher serial correlation, which is unlikely. Thus the CRDW test is generally not a suitable test statistic.

2) Three Stages:The Engle-Granger-Yoo Approach.

Engle and Yoo propose a third step to the standard Engle-Granger procedure which seeks to overcome some of the problems inherent in using the static model which yields β generally biased in small samples. Assuming that there is a unique cointegration vector and weak exogeneity of the short run parameters, then the third step provides a correction of the first stage estimate of β and ensures it has a normal distribution. The methodology consists of correcting the long-run relationship by the small sample bias $(\gamma/(1-\alpha))$ and use the correct residuals to perform the ADF test.

Stg1 Regress $y_t = \gamma_o x_t + \gamma_1 x_{t-1} + \alpha y_{t-1} + \varepsilon_t$

Stg2 Construct $\widehat{u}_t = y_t - \frac{\widehat{\gamma}_0 + \widehat{\gamma}_1}{1-\widehat{\alpha}} x_t$

Stg3 Perform an ADF test for $\widehat{u}_t$.

Cointegration in Multivariate Systems- The Johansen Approach

Defining z_t , a vector of n potentially endogenous variables it is possible to specify the following DGP and the model z_t as an unrestricted VAR involving k-lags of z_t

$$z_t = A_1 z_{t-1} + ... + A_k z_{t-k} + u_t$$

where z_t is $n \times 1$ and each of the A_i is $n \times n$ matrix of parameters. As we show above this equation can be written as a vector error-correction model

$$\Delta z_t = \Gamma_1 \Delta z_{t-1} + ... + \Gamma_{k-1} \Delta z_{t-k-1} + \Pi z_{t-k} + u_t$$

where $\Gamma_i = -(I - A_1 - ... - A_i)$ $(i = 1, .., k-1)$, and $\Pi = -(I - A_1 - ... - A_k)$.[2] This way of specifying the system contains information on both the short run and the long run adjustments to changes in z_t , via the estimates of $\hat{\Gamma}_i$ and $\hat{\Pi}$ respectively. As it will be seen, $\Pi = \alpha\beta'$, where α represents the speed of adjustment to disequilibrium, while β is a matrix of long-run coefficients such that the term $\beta' z_{t-k}$ represents up to $n-1$ cointegration relationships. Assuming that z_t is a vector of non-stationary I(1) variables, then all the terms in differences are I(0) while Πz_{t-k} must also be I(0) for the error term to be a white noise.

There are two cases where this requirement is met:

1. When there is no cointegration at all, implying that there are no linear combinations of z_t that are I(0), and consequently Π is a matrix of $n \times n$ zeros. In this case the appropriate model is a VAR in differences.

[2] Notice that in this presentation the variable in levels is LAG k, while in the previous presentation the variable in levels was lag 1. The analysis can be carried out in both ways.

2. When there exist up to $n-1$ cointegration relationships $\beta' z_{t-k}$ ~I(0). In this instance there will exist a number $r \leq n-1$ cointegration vectors where r is the number of columns of which form r LI combinations of the variables in z_t , together with $(n-r)$ non-stationary vectors in β. Only the cointegration vectors in enter in the VECM, which implies that for the last $(n-r)$ columns, the are insignificantly small. Thus the typical problem faced, of determining how many r cointegration vectors exist amounts to equivalently testing which columns in Π are zero. Consequently testing for cointegration amounts to a consideration of the rank of Π, that is, finding the number of r linearly independent columns in Π.

$-If\ \Pi$ *is full rank the variables in* z_t *have to be* $I(0)$

$-If\ \Pi$ *is zero there is no cointegrating vector*

$-If\ \Pi$ *is reduced rank the number of cointegrating vectors is the* $RANK(\Pi)$

Canonical Correlations

Population Canonical Correlations

Let the $(n_1 \times 1)$ vector y_t and the $(n_2 \times 1)$ vector x_t denote stationary random variables. Typically this variables are measured in deviations from the population mean such that $E(y_t y_t')$ represents the variance-covariance matrix of y_t .

In general

$$\begin{bmatrix} E(y_t y_t') & E(y_t x_t') \\ E(x_t y_t') & E(x_t x_t') \end{bmatrix} = \begin{bmatrix} \Sigma_{YY} & \Sigma_{YX} \\ \Sigma_{XY} & \Sigma_{XX} \end{bmatrix}$$

We can often gain some insight into the nature of these correlations by defining two new $(n \times 1)$ random vectors φ_t and ξ_t , where n is the smaller of n_1 and n_2 .

$$\begin{aligned} \varphi_t &= K' y_t \\ \xi_t &= A' x_t \end{aligned}$$

The matrices K' and A' are chosen such that

$$\begin{bmatrix} E(y_t y_t') & E(y_t x_t') \\ E(x_t y_t') & E(x_t x_t') \end{bmatrix} = \begin{bmatrix} \Sigma_{YY} & \Sigma_{YX} \\ \Sigma_{XY} & \Sigma_{XX} \end{bmatrix}$$

$E(\varphi_t \varphi_t') = K' \Sigma_{YY} K = I,$

$E(\xi_t \xi_t') = A' \Sigma_{XX} A = I$ and,

$$E(\varphi_t \xi_t') = R = \begin{bmatrix} r_1 & & 0 \\ & r_2 & \\ & & \\ 0 & & r_n \end{bmatrix}$$

where the elements of φ_t and ξ_t are ordered in such a way that$1 \geq r_1 \geq r_2..... \geq r_n \geq 0$

The population r_i is known as the i^{th} population canonical correlation between y_t and x_t .

The canonical correlations can be calculated by calculating the eigen values of

$$\Sigma_{YY}^{-1}\Sigma_{YX}\Sigma_{XX}^{-1}\Sigma_{XY},$$

$\lambda_1 > \lambda_2 > ...\lambda_n$, and the canonical correlations turn out to be the square roots of these eigenvalues.

The Johansen Method of reduced rank regression

The procedure may be described as follows:

First rewrite the ECM Equation

$$\Delta z_t + \alpha\beta' z_{t-k} \quad = \Gamma_1 \Delta z_{t-1} + ... + \Gamma_{k-1}\Delta z_{t-(k-1)} + u_t$$

it is possible to correct for the short run dynamics (i.e., take out their effect) by regressing Δz_t and $'z_{t-k}$ separately on the right hand side of the previous equation, i.e.,

$$\begin{aligned} \Delta z_t &= P_1 \Delta z_{t-1} + ... + P_{k-1}\Delta z_{t-(k-1)} + R_{0t} \\ z_{t-k} &= T_1 \Delta z_{t-1} + ... + T_{k-1}\Delta z_{t-(k-1)} + R_{kt} \end{aligned}$$

Which can be used to form the residual (product moment) matrices

$$\widehat{S}_{ij} = T^{-1}\sum \widehat{R}_{it}\widehat{R}'_{jt} \quad i,j = 0,k$$

The maximum likelihood estimate of β is obtained as eigenvectors corresponding to the largest eigenvalues from solving the equation

$$|\lambda \widehat{S}_{kk} - \widehat{S}_{0k}\widehat{S}_{00}^{-1}\widehat{S}_{0k}| = 0$$

which gives the n eigenvalues $\hat{\lambda}_1 > \hat{\lambda}_2 > .....\hat{\lambda}_n$ and their corresponding eigen vectors, $\hat{\varphi} = (\hat{\varphi}_1 > \hat{\varphi}_2 > ... > \hat{\varphi}_n)$. Those r elements in $\hat{\varphi}$ which determine linear combinations of stationary relationships can be denoted $\beta = (\hat{\varphi}_1, \hat{\varphi}_2, ..., \hat{\varphi}_r)$, that is, these are the cointegration vectors. This is because the eigen values are the largest squared canonical correlations between the "levels" residuals R_{kt} and the "differences" residuals R_{0t}, that is, we obtain estimates of all the distinct

$\hat{\varphi}_i' z_t$ combinations of the I(1) levels of z_t which produce high correlations with the stationary Δz_t elements, such combinations being the cointegrating vectors by the virtue of the fact that they must themselves be I(0) to achieve a high correlation. Thus the magnitude of $\hat{\lambda}_i$ is a measure of how strongly the cointegration relations are correlated with the stationary part of the model. The last $(n-r)$ combinations indicate the non-stationary combinations and theoretically are uncorrelated with the stationary elements. Consequently, for the eigenvectors corresponding to the non-stationary part of the model, $\hat{\lambda}_i = 0$ for $i = r+1, ..., n$.

Testing for reduced rank

To find the number of cointegrating vectors we said that is equivalent to find the number of linearly independent columns in Π or the number of $n-r$ columns of significantly small.

The approach amounts to a reduced rank regression which provides n eigenvalues $\hat{\lambda}_1 > \hat{\lambda}_2 > .....\hat{\lambda}_n$ and their corresponding eigen vectors,$\hat{\varphi} = (\hat{\varphi}_1 > \hat{\varphi}_2 > ... > \hat{\varphi}_n)$. Those r elements in $\hat{\varphi}$ which determine linear combinations of stationary relationships can be denoted $\beta = (\hat{\varphi}_1, \hat{\varphi}_2, ..., \hat{\varphi}_r)$ that is, the distinct $\hat{\varphi}_i' z_t$, which we will denote $\beta_i' z_t$, are correlated with the stationary part of the model. The last $n-r$ combinations obtained from the Johansen approach indicate the non stationary combinations, and theoretically these are uncorrelated with the stationary elements in the ECM. Consequently, for the eigenvectors corresponding to the non-stationary part of the model, $\hat{\lambda}_i = 0$ for $i = r+1, ..., n$.

Thus to test the null hypothesis that there are at most r cointegration vectors amounts to test

$$H_0)\lambda_i = 0 \quad i = r+1, ..., n.$$

where only the first r eigenvalues are non-zero. (This is tested against the alternative of n cointegrating vectors).

It can be shown (see Hamilton) that the likelihood test that corresponds to the ECM under the null that there are only r cointegrating vectors is

$$L^*(H_0) = -(Tn/2)log(2\pi) - (Tn/2) - (T/2)log\left|\widehat{S}_{00}\right| - (T/2)\sum_{i=1}^{r}\log(1-\hat{\lambda}_i).$$

It can also be shown that the likelihood test that corresponds to the ECM without any restriction on the number of cointegrating vectors is

$$L^* = -(Tn/2)log(2\pi) - (Tn/2) - (T/2)log\left|\widehat{S}_{00}\right| - (T/2)\sum_{i=1}^{n}\log(1-\hat{\lambda}_i).$$

Then a likelihood ratio test, *using a non standard distribution*, can be constructed, using what is known as the **Trace statistic.**

$$\lambda_{trace} = -T \sum_{i=r+1}^{n} \log(1 - \hat{\lambda}_i) \qquad r = 0, 1, ..., n-2, n-1$$

Another test of the significance of the largest λ_r is the so called maximal-eigenvalue or $\boldsymbol{\lambda - \max}$ **statistic** :

$$\lambda_{\max} = -T \log(1 - \hat{\lambda}_{r+1}) \quad r = 0, 1, ..., n-2, n-1.$$

This tests the existence of r cointegrating vectors against the alternative that $r+1$ exist and is derived in exactly same way.

Testing restrictions on the cointegrating vector.

Many times we are interested to test restrictions on the cointegrating vector. For example I might be interested in some theoretical long run relationship which impose some restrictions on the values of the cointegrating relationship. We may be also interested in testing whether we should include or not a regressor in the cointegrating relationship.

The crucial point in deriving a LR test for these type of hypothesis is that both under the null and the alternative there are r cointegrating relationships and therefore the asymptotic theory is standard since the regressions only involve variables which are I(0)and the test would be distributed *chi-square*. The LR test will be

$$LR = -T \sum_{i=1}^{r} \log(1 - \hat{\lambda}_i) + T \sum_{i=1}^{r} \log(1 - \tilde{\lambda}_i)$$