Time Series Analysis

Cointegration

Binzhi Chen

Department of Economics
University of Birmingham

MSc in 2020 Fall

Introduction

In previous lectures, we go through time series analysis based on stationary multivariables and their lags.

However, in the reality, it is implausible to assume every time series is stationary.

Cointegration extends stationary VAR models to non-stationary time series.

Try to imagine a man with a dog.

Intriduction

Recall the definition of integration, we know that if:
$y_t \sim I(1)$, then $\Delta y_t = y_t - y_{t-1} \sim I(0)$

In most cases, if we combine two variables which are $\mathrm{I}(1)$, then the combination will also be $\mathrm{I}(1)$

More generally, if we combine variables with different orders of integration, the combination will have an order of integration equal to the largest:
If $X_{i,t} \sim I(d_i)$ for $i = 1, 2, \ldots, k$:
so we have k variables each integrated of order d_i

Let:

$$Z_t = \sum_{i=1}^{k} \alpha_i X_{i,t}$$

Then $Z_t \sim I(\mathrm{Max}\, d_i)$

Introduction

Assume $\mathbf{y_t}$ is a $k \times 1$ vector of variables, then the components of $\mathbf{y_t}$ are cointegrated as a linear combination if:

$$\mathbf{\Pi y_t} \sim I(0)$$

Many time series are non-stationary but "move together" over time.

If variables are cointegrated, it means that a linear combination of them will be stationary.

There may be up to r linearly independent cointegrating relationships (where $r \leq k-1$), also known as cointegrating vectors. r is also known as the cointegrating rank of $\mathbf{y_t}$.

A cointegrating relationship may also be seen as a long term relationship.

Introduction

Examples of possible Cointegrating Relationships in finance:

1. spot and futures prices
2. ratio of relative prices and an exchange rate
3. equity prices and dividends
4. Pairs Trading

No cointegration implies that series could wander apart without bound in the long run.

Introduction

When the concept of non-stationarity was first considered, a usual response was to independently take the first differences of a series of $I(1)$ variables.

The problem with this approach is that pure first differences models have no long run solution.

For example, consider y_t and x_t are both $I(1)$ In the long-run, the

difference terms would be zero. Hence:

$$\Delta y_t = \beta \Delta x_t + u_t$$

this collapses to nothing in the long run

Cointegration shows the "long-run equilibrium" relationship between variables.

ECM and VECM

What if we want to examine the "short-run dynamics"?

Phillips(1957) proposed the concept of error correction model. Assume x_t and y_t are two $I(1)$ series:

$$y_t = b_1 + b_2 y_{t-1} + b_3 x_t + b_4 x_{t-1} + \epsilon_t \qquad b_2 \neq 1$$

we can arrange it as error correction form:

$$\Delta y_t = b_1 + b_3 \Delta x_t + a(y_{t-1} + b x_{t-1}) + \epsilon_t$$

Where

$$a = b_2 - 1 \qquad b = \frac{b_3 + b_4}{b_2 - 1}$$

and

$$y_{t-1} + b x_{t-1} = z_t \sim I(0)$$

ECM and VECM

Apparently, the factors that influences the short-run change of y_t (which is Δy_t) are:

1. the short-run change of x_t (Which is Δx_t)
2. the equilibrium error from last period z_{t-1}

The advantages of ECM are:

1. First difference terms avoid spurious regression
2. The introduction of error correction terms ensures that the information of variables is not ignored
3. Since it is stationary, we can use OLS to estimate

ECM and VECM

Engle and Granger(1987) proposed Granger representaion theorem:

If a system is cointegrated then there exists a vector error correction model(VECM) with a reduced rank coefficient matrix and if there is a VECM with a reduced rank coefficient matrix then the system must be cointegrated.

The simplest VECM is:

$$\begin{bmatrix} \Delta Y_{1t} \\ \Delta Y_{2t} \end{bmatrix} = \begin{bmatrix} \pi_{11} & \pi_{12} \\ \pi_{21} & \pi_{22} \end{bmatrix} \begin{bmatrix} Y_{1t-1} \\ Y_{2t-1} \end{bmatrix} + \begin{bmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{bmatrix} \tag{1}$$

Or:

$$\mathbf{Y_t} = \mathbf{\Pi Y_{t-1}} + \boldsymbol{\epsilon_t}$$

Where

$$\mathbf{\Pi} = \boldsymbol{\alpha}\boldsymbol{\beta}'$$

$\boldsymbol{\alpha}$ is the adjustment coefficient and $\boldsymbol{\beta}$ is the cointegrating vectors

ECM and VECM

In the general multivariate case, a cointegrated VAR(P):

$$\mathbf{Y_t} = \mathbf{\Phi_0} + \mathbf{\Phi_1 Y_{t-1}} + \mathbf{\Phi_2 Y_{t-2}} + \cdots + \mathbf{\Phi_p Y_{t-p}} + \boldsymbol{\epsilon_t}$$

can be turned into an VECM by recursive substitution:

$$\Delta \mathbf{Y}_t = \mathbf{\Pi Y}_{t-1} + \boldsymbol{\pi_1} \Delta \mathbf{Y}_{t-1} + \boldsymbol{\pi_2} \Delta \mathbf{Y}_{t-2} + \ldots + \boldsymbol{\pi_{p-1}} \Delta \mathbf{Y}_{t-P+1} + \boldsymbol{\epsilon}_t$$

Where

$$\mathbf{\Pi} = -\mathbf{I_k} + \sum_{i=1}^{p} \mathbf{\Phi_i}$$

and

$$\boldsymbol{\pi_p} = -\sum_{i=p+1}^{p} \mathbf{\Phi_i}$$

Methods of Estimation in Cointegrated Systems

The procedure of estimation in cointegrated systems contains two parts: testing for cointegration and VECM parameter estimation for cointegration

There are (at least) 3 methods we could use:

1. Engle Granger
2. Engle and Yoo
3. Johansen

The Engle Granger 2 Step Method

This is a single equation technique which is conducted as follows: Step 1:

1. Make sure that all the individual variables are $I(1)$.
2. Then estimate the cointegrating regression using OLS.
3. Save the residuals of the cointegrating regression
4. Test these residuals to ensure that they are $I(0)$.

The model for the equilibrium correction term can be generalised to include more than two variables:

$$y_t = \beta_0 + \beta_1 x_{1t} + \beta_2 x_{2t} + \cdots + \beta_k x_{kt} + u_t$$

u_t should be $I(0)$ if the variables $y_t\, x_{1t}\, x_{2t}\, \ldots\, x_{kt}$ are cointegrated

So what we want to test is the residuals of this equation to see if they are non-stationary or stationary

The Engle Granger 2 Step Method

We can use the DF/ADF test on *ut*. So we have the regression:

$$\Delta u_t = \psi u_{t-1} + \epsilon_t$$

However, since this is a test on the residuals of an actual model $\hat{u}_t$, then the critical values are changed.

Engle and Granger (1987) have tabulated a new set of critical values and hence the test is known as the Engle Granger (E.G.) test.

The Engle Granger 2 Step Method

We can also use the Durbin Watson test statistic or the Phillips Perron approach to test for non-stationarity of $\hat{u}_t$

What are the null and alternative hypotheses for a test on the residuals of a potentially cointegrating regression?

$$H_0 : \hat{u}_t \sim I(1)$$
$$H_1 : \hat{u}_t \sim I(0)$$

The problem for the method is that, if we have k variables, we can only find 0 or only 1 cointegrating relationship by OLS.

So the Engle-Granger method is generally only applicable if there are two variables.

The Engle Granger 2 Step Method

Step 2:

Use the step 1 residuals as one variable in the error correction model:

$$\Delta y_t = \beta_1 \Delta x_t + \beta_2(\hat{u}_{t-1}) + u_t$$

Where

$$\hat{u}_{t-1} = y_{t-1} - \hat{\alpha} x_{t-1}$$

The Engle Granger 2 Step Method

This method suffers from a number of problems:

1. The Engle-Granger method is generally only applicable if there are two variables.
2. Unit root and cointegration tests have low power in finite samples.
3. We are forced to treat the variables asymmetrically and to specify one as the dependent and the other as independent variables.
4. Cannot perform any hypothesis tests about the actual cointegrating relationship estimated at stage 1.

Problem 1 is addressed by Johansen approach.
Problem 2 is a small sample problem that should disappear asymptotically.
Problem 3 is addressed by the Johansen approach.
Problem 4 is addressed by the Engle and Yoo approach or the Johansen approach.

The Engle & Yoo 3-Step Method:

One of the problems with the EG 2-step method is that we cannot make any inferences about the actual cointegrating regression.

The Engle Yoo (EY) 3-step procedure takes its first two steps from EG.

EY add a third step giving updated estimates of the cointegrating vector and its standard errors.

The most important problem with both these techniques is that in the general case above, where we have more than two variables which may be cointegrated, there could be more than one cointegrating relationship.

The Engle & Yoo 3-Step Method:

In fact there can be up to r linearly independent cointegrating vectors (where $r \leq k-1$), where k is the number of variables in total.

So, in the case where we just had y and x, then r can only be one or zero.

But in the general case there could be more cointegrating relationships.

And if there are others, how do we know how many there are or whether we have found the "best"?

The answer to this is to use a systems approach to cointegration which will allow determination of all r cointegrating relationships - Johansen's method.

Johansen Methodology

The Johansen methodology is the dominant technique used to determine whether a system of $I(1)$ variables is cointegrated and if so, to determine the number of cointegrating relationships.

One of the requirements for a set of integrated variables to be cointegrated is that $\mathbf{\Pi}$ has reduced rank in VECM:

$$\Delta \mathbf{Y}_t = \mathbf{\Pi Y}_{t-1} + \boldsymbol{\pi_1} \Delta \mathbf{Y}_{t-1} + \boldsymbol{\pi_2} \Delta \mathbf{Y}_{t-2} + \ldots + \boldsymbol{\pi_{p-1}} \Delta \mathbf{Y}_{t-P+1} + \boldsymbol{\epsilon}_t$$

The rank of a matrix is equal to the order of the largest square matrix we can obtain from $\mathbf{\Pi}$ which has a non-zero eigenvalues determinant.

Let's have quick review of linear algebra.

Johansen Methodology

Let denote $\mathbf{\Pi}$ a $k \times k$ square matrix and let $\mathbf{c}$ denote a $k \times 1$ non-zero vector, and let λ denote a set of scalars.

λ is called a characteristic root or set of roots of $\mathbf{\Pi}$ if we can write:

$$\mathbf{\Pi c} = \lambda \mathbf{c}$$

We can also write:

$$\mathbf{\Pi c} = \lambda \mathbf{I_k c}$$

and hence:

$$(\mathbf{\Pi} - \lambda \mathbf{I_k})\mathbf{c} = 0$$

Where $\mathbf{I_k}$ is an identity matrix.

Johansen Methodology

Since $c \neq 0$ by definition, then for this system to have zero solution, we require the matrix:

$$|\mathbf{\Pi} - \lambda \mathbf{I_k}| = 0$$

to be singular (i.e. to have zero determinant).

For example, let $\mathbf{\Pi}$ be the 2×2 matrix

$$\mathbf{\Pi} = \begin{bmatrix} 5 & 1 \\ 2 & 4 \end{bmatrix}$$

Then the characteristic equation is:

$$|\mathbf{\Pi} - \lambda \mathbf{I_2}| = \left| \begin{bmatrix} 5 & 1 \\ 2 & 4 \end{bmatrix} - \lambda \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \right| = 0$$

This gives the solutions $\lambda_1 = 6$ and $\lambda_2 = 3$

Johansen Methodology

The characteristic roots are also known as Eigenvalues.

We write Rank($\mathbf{\Pi}$) = r
The rank of a matrix is equal to the number of linearly independent rows or columns in the matrix.

What is the rank of $\mathbf{\Pi}$ in this example?

Some properties of the eigenvalues of any square matrix A:

1. the sum of the eigenvalues is the trace
2. the product of the eigenvalues is the determinant
3. the number of non-zero eigenvalues is the rank

Johansen Methodology

Now, let's go back to the VECM representaion:

$$\Delta \mathbf{Y}_t = \mathbf{\Pi Y}_{t-1} + \boldsymbol{\pi_1} \Delta \mathbf{Y}_{t-1} + \boldsymbol{\pi_2} \Delta \mathbf{Y}_{t-2} + \ldots + \boldsymbol{\pi_{p-1}} \Delta \mathbf{Y}_{t-P+1} + \boldsymbol{\epsilon}_t$$

The test for cointegration between the $\mathbf{Y}$ is calculated by looking at the rank of the $\mathbf{\Pi}$ matrix via its eigenvalues (to prove this requires some technical intermediate steps).

Please note that the rank of a matrix is equal to the number of its characteristic roots (eigenvalues) that are different from zero.

ECM and VECM

$$\mathbf{\Pi} = \boldsymbol{\alpha}\boldsymbol{\beta}'$$

is the decomposition of the matrix $\mathbf{\Pi}$ For any $1 < r < k$, $\mathbf{\Pi}$ is defined as the product of two matrices above.

For example, if $k = 4$ and $r = 1$, $\boldsymbol{\alpha}\boldsymbol{\beta}$ will be 4×1, and $\mathbf{\Pi}_{t-k}$ will be given by:

$$\mathbf{\Pi} = \begin{pmatrix} \alpha_{11} \\ \alpha_{12} \\ \alpha_{13} \\ \alpha_{14} \end{pmatrix} \begin{pmatrix} \beta_{11} & \beta_{12} & \beta_{13} & \beta_{14} \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{pmatrix}$$

Johansen Methodology

Two test statistics methods for Johansen techniques are the trace statistic and the maximum eigenvalue statistic.

The trace statistic is formulated as:

$$\lambda_{trace}(r) = -T \sum_{i=r+1}^{k} \ln\left(1 - \hat{\lambda}_i\right)$$

where $\hat{\lambda}_i$ is the estimated value for the ith ordered eigenvalue from the $\mathbf{\Pi}$ matrix.

λ_{trace} tests the null that the number of cointegrating vectors is less than equal to r against an unspecified alternative.

$\lambda_{trace} = 0$ when all the $\lambda_i = 0$, so it is a joint test.

Johansen Methodology

The maximum eigenvalue statistic is formulated as:

$$\lambda_{\max}(r, r+1) = -T \ln\left(1 - \hat{\lambda}_{r+1}\right)$$

$\lambda_{\max}$ tests the null that the number of cointegrating vectors is r against an alternative of $r+1$. The eigenvalues denoted are λ_i should put in order:

$$\lambda_1 > \lambda_2 > \cdots > \lambda_k$$

if the λ_i's are roots, they must be less than 1 in absolute value.

Johansen Methodology

If the variables are not cointegrated and the matrix is not a full rank, the rank of $\mathbf{\Pi}$ will not be significantly different from zero, hence:

$$\lambda_i = 0 \quad \text{for} \quad \forall\, i$$

Hence:

$$ln(1 - \lambda_i) \approx 0$$

If the rank of the matrix is equal to 1, then:

$$ln(1 - \lambda_1) < 0$$

and,

$$\lambda_{trace}(0) \approx ln(1 - \lambda_1)$$

which becomes arbitrarily large as T grows.

If the rank of the matrix $1 < r < k$, then:

$$ln(1 - \lambda_i) < 0 \quad \text{for} \quad \forall\, i > 1$$

Johansen Methodology

Johansen Juselius (1990) provide critical values for the 2 statistics. The distribution of the test statistics is non-standard. The critical values depend on:

1. the value of $k - r$, the number of non-stationary components
2. whether a constant and/or trend are included in the regressions.

If the test statistic is greater than the critical value from Johansen's tables, reject the null hypothesis that there are r cointegrating vectors in favour of the alternative that there are more than r.

Johansen Methodology

The testing sequence under the null is $r = 0, 1, \ldots, k-1$

so that the hypotheses for λ_{trace} are:

$$\begin{aligned} H_0 : r &= 0 \qquad H_1 : 0 < r \leq k \\ H_0 : r &= 1 \qquad H_1 : 1 < r \leq k \\ H_0 : r &= 2 \qquad H_1 : 2 < r \leq k \\ &\vdots \\ H_0 : r &= k-1 \qquad H_1 : r = k \end{aligned}$$

We keep increasing the value of r until we no longer reject the null.

Johansen Methodology

But how does this correspond to a test of the rank of the matrix?

$\mathbf{\Pi}$ cannot be of full rank(k) since this would correspond to the original y_t being stationary.

If $\mathbf{\Pi}$ has zero rank, then by analogy to the univariate case. There is no long run relationship between the elements of y_{t-1}. Hence there is no cointegration.

For $1 <$ rank(k) $<$ g, there are multiple cointegrating vectors.

Johansen Methodology

Recall that EG did not allow us to do hypothesis tests on the cointegrating relationship itself, but the Johansen approach does.

If there exist r cointegrating vectors, only these linear combinations will be stationary.

You can test a hypothesis about one or more coefficient in the cointegrating relationship by viewing the hypothesis as a restriction on the $\mathbf{\Pi}$ matrix.

All linear combinations of the cointegrating vectors are also cointegrating vectors.

Johansen Methodology

If the number of cointegrating vectors is large, and the hypothesis under consideration is simple, it may be possible to recombine the cointegrating vectors to satisfy the restrictions exactly.

As the restrictions become more complex or more numerous, it will eventually become impossible to satisfy them by renormalisation.

YAfter this point, if the restriction is not severe, then the cointegrating vectors will not change much upon imposing the restriction.

Johansen Methodology

A test statistic to test this hypothesis is given by:

$$-T\sum_{i=1}^{r}\left[\ln\left(1-\lambda_i\right)-\ln\left(1-\lambda_i^*\right)\right]\sim\chi^2(m)$$

where

λ_i^* are the characteristic roots of the restricted model,
λ_i are the characteristic roots of the unrestricted model,
r is the number of non-zero characteristic roots in the unrestricted model and m is the number of restrictions.