Advanced Topics in Unit Roots and Cointegration

Outline:

- Unit roots
- Cointegration and common trends
- The single equation approach.
  - Engle-Granger procedure
  - Estimation of the cointegrating vector
- The system equation approach.
  - Johansen cointegration test and Error correction models
  - Hypothesis testing on the cointegrating vectors

Unit roots

- Consider the first-order autoregressive (AR(1)) model:

$$Y_t = \phi Y_{t-1} + \varepsilon_t,$$

  where ε_t is white noise (mean zero, constant variance σ_ε^2, zero autocovariances).
- This process is stationary if the roots of the polynomial $\Phi(z) = 1 - \phi z$ are outside the unit circle.
- We have the following cases:
- If $-1 < \phi < 1$, then Y_t is stationary.
- If $|\phi| > 1$, Y_t is explosive.
- If $\phi = 1$, Y_t has a unit root.

Unit roots

Recall that in a AR(p) process of the form

$$Y_t = \phi_1 Y_{t-1} + ... + \phi_p Y_{t-p} + \varepsilon_t$$

where ε_t is a white noise process.
To study stationarity we consider the roots of the polynomial $\Phi(z) = 1 - \phi_1 z - ... - \phi_p z^p$.

- If the roots lie outside the unit circle the process is stationary.
- If some roots are equal to one there are unit roots.
- If some roots are inside the unit circle the process is explosive.

Explosive series are not frequent in economics. Usually we have stationary series or series with unit roots.

Unit roots in the AR(1) model

- Consider the AR(1) process

$$Y_t = \phi Y_{t-1} + \varepsilon_t,$$

- If $\phi = 1$, Y_t has a unit root and the model becomes a *random walk*,

$$Y_t = Y_{t-1} + \varepsilon_t,$$

with mean Y_0 and variance $\sigma^2 t$.

- Often it is appropriate to include an intercept in the AR(1) model

$$Y_t = \mu + \phi Y_{t-1} + \varepsilon_t,$$

where ε_t is white noise process.

- If $\phi = 1$, Y_t has a unit root and the model becomes a *random walk with a drift*,

$$\Delta Y_t = \mu + \varepsilon_t,$$

and changes in Y_t are equal to a constant μ plus a stationary component ε_t.

Unit roots in the AR(1) model

- Sometimes a linear trend is also appropriate and the AR(1) model becomes

$$Y_t = \mu + \beta t + \phi Y_{t-1} + \varepsilon_t,$$

where ε_t is white noise process.

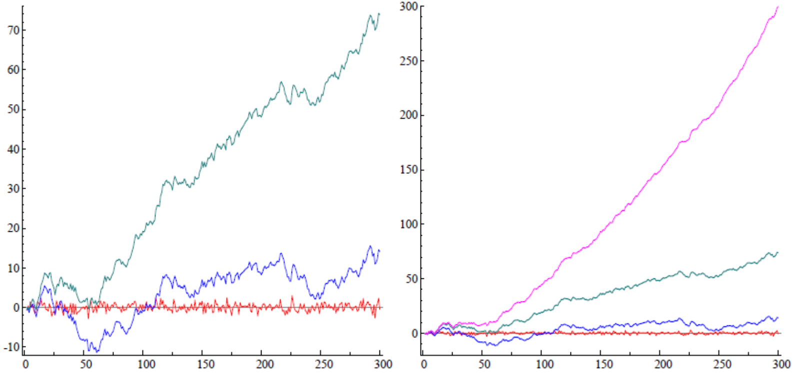
- If $\phi = 1$, Y_t has a unit root and the model becomes a *random walk with a drift and a trend*,

$$\Delta Y_t = \mu + \beta t + \varepsilon_t,$$

and changes in Y_t are equal to a linear trend plus a stationary component ε_t.

Unit roots in the AR(1) model

- **Examples:** The pictures below graph a white noise (in red), a random walk (in blue), a random walk with a drift (in green) and a random walk with a drift and a trend (in magenta)



Unit roots

- Inference using non-stationary processes (processes with a unit root) is problematic because the standard asymptotic results do not hold.
- For example, if *two independent random-walks* $\{x_t\}$ and $\{y_t\}$ are generated, regressing y_t on x_t leads to intriguing results:
  - Granger and Newbold (1974) showed that a t-test for the significance of the parameter associated with x_t often leads to the rejection of the null.
  - For example, for $T = 50$ the rejection frequency for a two-sided test at 5% is 66.2%.
  - With $T = 250$, the rejection frequency goes up to 84.7%.
  - The R^2 is often very high.
- These *spurious regressions* arise because, under the null, the model does not satisfy the usual assumptions.
- Regressions using non-stationary variables are only interesting in a particular case to be studied later.

- Consider the AR(1) process

$$Y_t = \phi Y_{t-1} + \varepsilon_t,$$

- Unit root test (*Dickey-Fuller*): t-test for $\gamma = \phi - 1 = 0$ against $\gamma < 0$ in least-squares regression

$$\Delta Y_t = \gamma Y_{t-1} + \varepsilon_t, \qquad t = 1, \ldots, T. \tag{1}$$

  Test for $H_0 : Y_t \sim I(1)$ ($\gamma = 0$) against $H_1 : Y_t$ is stationary ($\gamma < 0$).
- Testing this hypothesis can be done as t-ratio from the OLS regression in (1) which can be written as

$$t_\gamma = \frac{\hat{\gamma}}{se(\hat{\gamma})}$$

  This is known as Dickey-Fuller test.
- The Dickey-Fuller Test uses the t-statistic, but t_γ is not asymptotically normal. Its distribution is non-standard: One-sided 5% critical value is -1.95. **Rejection rule:** We reject H_0 if the observed value of t_γ is *smaller* than the critical value.

- Consider now the $AR(p)$ model

$$\begin{aligned} Y_t &= \sum_{i=1}^{p} \phi_i Y_{t-i} + \varepsilon_t, \\ \Phi(L) Y_t &= \varepsilon_t \end{aligned}$$

  where $\Phi(L) = 1 - \sum_{i=1}^{p} \phi_i L^i$.
- Test $H_0 : \Phi(z)$ has a unit root equal to 1, *i.e.* $\Phi(1) = 0$.
- The model can be written

$$\Delta Y_t = \gamma Y_{t-1} + \alpha_1 \Delta Y_{t-1} + \ldots + \alpha_p \Delta Y_{t-p+1} + \varepsilon_t.$$

  where $\gamma = -\Phi(1)$,and $\alpha_i = -\sum_{k=i+1}^{p} \phi_k$. We test $H_0 : Y_t \sim I(1)$ ($\gamma = 0$) against $H_1 : Y_t$ is stationary ($\gamma < 0$). We continue to use the t-statistic for H_0.
- Same (asymptotic) critical values as the DF statistic.
- Now it's called an *augmented Dickey-Fuller* (ADF) test, but still the same critical values

In practice we apply the DF/ADF test even if the true process is not a $AR(p)$ process:

- Lag length p should be chosen such that ε_t does not display any autocorrelation.
- The lags are intended to clear up any serial correlation, if too few, test won't be right.
- So we have to test if there is serial correlation in model (1). If there is, include lagged dependent variables.
- A popular approach is to base the lag length selection on the minimization of the Akaike or Schwarz information criteria.

In practice the model can be extended to include a constant or constant & linear trend:

- To allow for non-zero constant mean under null and alternative hypothesis, estimate:

  $$\Delta Y_t = \gamma Y_{t-1} + \mu + \varepsilon_t.$$

  The 5% critical value shifts to -2.86.
- To allow for drift under the null and trend-stationarity under the alternative hypothesis, estimate:

  $$\Delta Y_t = \gamma Y_{t-1} + \mu + \lambda t + \varepsilon_t.$$

  The 5% critical value shifts to -3.41.
- **Remark:** Usually, in practice we include always an intercept. The non-stochastic trend is included if the variable displays a trend. However, you could test this using a t-statistics for μ and λ. The asymptotic distributions of these t-statistics are also *non-standard*.

- The *Phillips-Perron* (PP) unit root test differ from the ADF tests mainly in how it deals with serial correlation and heteroskedasticity in the errors.
- In particular,where the ADF tests use a parametric autoregression to approximate the ARMA structure of the errors in the test regression, the PP test **corrects** the t statistics for any serial correlation and heteroskedasticity in the errors of the test.
- The asymptotic distribution of the PP statistic is the same asymptotic distributions of the ADF t-statistic

- The available evidence points towards the existence of many non-stationary macro-economic variables.
- As we have seen, estimating regressions with $I(1)$ variables is likely to lead to erroneous conclusions.
- A possible solution is to work with *differenced series* (this was the practice adopted by most people after the Granger and Newbold paper).
- However, models in differences are *mute* about the relation between the levels of the variables in a *steady state*.
- Moreover, economic theory suggests that there are *stable relations* between the levels of some of these variables.
- This is possible if some linear combination of non-stationary variables is stationary.
- That is, although the series have random trends, they *drift "together"*.

- Consider two time series y_t and x_t which are both $I(d)$:
    - If there exists a vector $\gamma = (1, -\beta)'$, such that the linear combination

$$u_t = y_t - \beta x_t \sim I(d-b),$$

    then, y_t and x_t are said to be *cointegrated* with γ being the **cointegrating vector** and $d \geq b > 0$.
- *In 2003, Clive Granger was awarded the Nobel prize for introducing "methods of analyzing economic time series with common trends (cointegration)".*

- Several points are worth noticing:
  1. Cointegration refers to a *linear* combination of nonstationary variables;
  2. The cointegrating vector is *not uniquely defined*;
  3. Both variables must be integrated of the *same order* to be candidates to form a cointegrating relationship.
  4. Like most of the literature, we will focus on the case that $d = b = 1$, since few economic variables prove to be integrated of higher order
  5. If y_t and x_t are cointegrated, they must share (up to a scalar) the same stochastic trend, called a *common trend*.
  6. When two series are cointegrated it suggests that even though both processes are nonstationary, there is some *long-run equilibrium relationship* linking both series so that relationship is stationary.

- Examples of possibly cointegrated pairs of time series:
  - exchange rates and relative prices (*purchasing power parity*);
  - spot and futures prices of assets or exchange rates;
  - short- and long-term interest rates (*term structure models*);
  - stock prices and dividends (*present value relations*).
- The concept is easily extended to more than two series: if $y_t, x_{1t}, \ldots, x_{kt}$ are all $I(1)$,

$$u_t = y_t - \beta_1 x_{1t} - \ldots - \beta_k x_{kt},$$

then there is a a cointegrating relation if u_t is stationary. $\delta = (1, -\beta_1, ..., -\beta_k)'$ is known as cointegrating vector.

There are two approaches to test for cointegration and estimate the cointegrating parameters:

- The single equation approach.
- The system equation approach.

The single equation approach.

Engle-Granger procedure

Engle and Granger proposed to analyse cointegration between time series $(y_t, x_{1t}, \ldots, x_{kt})$, as follows.

- Choose one of the variables as the dependent variable, e.g., y_t, and estimate the static long run equation

$$y_t = \beta_0 + \beta_1 x_{1t} + \ldots + \beta_k x_{kt} + u_t$$

  by ordinary least-squares.

- Apply an ADF unit root test to the residuals $\hat{u}_t$ from this regression.

$$\Delta \hat{u}_t = \gamma \hat{u}_{t-1} + \alpha_1 \Delta \hat{u}_{t-1} + \ldots + \alpha_p \Delta \hat{u}_{t-p} + v_t,$$

  (an intercept and a time trend can also be included) the t-statistic is given by

$$t_\gamma = \frac{\hat{\gamma}}{se(\hat{\gamma})}.$$

  where $\hat{\gamma}$ is the OLS estimator of γ.

The single equation approach.

Engle-Granger procedure

- The $t-$statistic has a non-standard distribution (different from the usual ADF tests). This yields a test for $H_0 : u_t \sim I(1)$ (spurious regression) against $H_1 : u_t \sim I(0)$ (cointegration). Critical values depend on number of regressors (k). They also change if an intercept and/or a time trend are included.
- **Rejection rule:** We reject H_0 if the observed value of t_γ is **smaller** than the critical value.

- Choice of the dependent variable is arbitrary; all variables could be endogenous. Selecting a variable x_{it} as dependent variable should not matter asymptotically (as number of observations $T \rightarrow \infty$), but will make a difference in practice.
- Method can only be used if there is only a single unique cointegrating relation, which involves y_t. Thus $(x_{1t}, \ldots, x_{kt})$ are not allowed to be cointegrated without y_t.

The single equation approach.

Estimation of the cointegrating vector

- The OLS estimators $\hat{\beta}$ are (super-)consistent estimators of the cointegrating vectors under cointegration but has some drawbacks.
  - It is can be considerably biased in finite samples.
  - we cannot apply standard t-tests or confidence intervals to them.
- One approach to this problem is to use the VAR approach described later.

- A simpler approach is to use the Saikkonen-Stock-Watson augmented least squares estimator.
- This estimator is obtained by running the regression

$$y_t = \alpha + \beta' x_t + \sum_{j=-p}^{p} \delta_j' x_{t+j} + v_t$$

  where p is allowed to increase with T at the appropriate rate $x_t = (x_{1t}, \ldots, x_{kt})'$. The resulting estimator is also super-consistent for β.
- In practice the choice of p is as in the ADF tests and the chosen value must ensure that the errors v_t are serially uncorrelated.
- It is also common practice to choose the value of p based on the Schwarz criterion.
- Although, the distribution of the estimator of β is not normal, Saikkonen (1991) showed that valid inference about the long-run parameters can be performed using t and Wald statistics. That is, the asymptotic distributions of the t-statistics and the Wald statistics are standard normal and chi-squared respectively

- Cointegration has implications for the short-run dynamics of the series.
- If the series are cointegrated the equilibrium error u_t contains information about Δy_t.
- Models incorporating this information were introduced by Davidson, Hendry, Srba and Yeo (1978) and are said to contain an *error correction mechanism*.
- **Granger representation theorem**: Let $y_t \sim I(1)$ and $x_t \sim I(1)$: y_t and x_t are cointegrated with cointegrating vector $(1, \beta')$ if and only if there is a (reduced rank) error correction model that explains the short run behaviour of y_t and x_t (more on this theorem later).

- The (reduced rank) error correction model describing the short run behaviour of y_t is given by

$$\Delta y_t = \theta_0 + \rho\left(y_{t-1} - \alpha - \beta' x_{t-1}\right) + \sum_{i=1}^{q_1} \lambda_i \Delta y_{t-i} + \sum_{i=1}^{q_2} \varphi_i' \Delta x_{t-i} + u_t$$

  where u_t is a stationary process and q_1 and q_2 are suitable values.
- The parameter ρ is known as the *adjustment coefficient:* $\rho\left(y_{t-1} - \alpha - \beta' x_{t-1}\right)$ corresponds to the adjustment in Δy_tin response to a disequilibrium $\left(y_{t-1} - \alpha - \beta' x_{t-1}\right)$.
- The vector error correction model can be estimated by running the regression

$$\Delta y_t = \theta_0 + \rho\left(y_{t-1} - \hat{\alpha} - \hat{\beta}' x_{t-1}\right) + \sum_{i=1}^{q_1} \lambda_i \Delta y_{t-i} + \sum_{i=1}^{q_2} \varphi_i' \Delta x_{t-i} + \eta_t$$

  where $\hat{\alpha}$ and $\hat{\beta}$ are super-consistent estimators of α and β and η_t is the error term.
- Stock (1987) showed that standard inference for the short-run parameters is valid.

An approach that allows us to overcome the problems of the Engle Granger test is the methodology introduced by Johansen. The starting point is a $VAR(p)$ process for $X_t = (X_{1t}, X_{2t}, \ldots, X_{kt})'$

$$\Phi(L)X_t = \varepsilon_t$$

where $\Phi(L) = I_k - \sum_{i=1}^{p} \Phi_i L^i$(no intercept for convenience) and ε_t is a multivariate white noise process. This process can be written as

$$\begin{aligned} \Delta X_t &= \Pi X_{t-1} + \sum_{i=1}^{p-1} \Phi_i^* \Delta X_{t-i} + \varepsilon_t, \\ \Phi_i^* &= -\sum_{j=i+i}^{p} \Phi_j, \\ \Pi &= \sum_{j=1}^{p} \Phi_j - I_k = -\Phi(1) \end{aligned}$$

Let us assume that X_t is $I(1)$ that is $(1-L)X_t$ is stationary.

We can only have two possible cases:

1. $\text{rank}(\Pi) = 0 \Leftrightarrow \Pi = 0 \Rightarrow \Phi(L)$ has at least k unit roots: $\Delta X_t = \sum_{i=1}^{p-1} \Phi_i^* \Delta X_{t-i}$, i.e. $VAR(p-1)$ model for ΔX_t.
2. $0 < \text{rank}(\Pi) = r < k$, *We have cointegration*.

Remark: The case $\text{rank}(\Pi) = k$ is not compatible with X_t being $I(1)$ as it implies that $|\Phi(1)| \neq 0$, so no unit roots: $X_t \sim I(0)$.

To see that case 2 implies cointegration note that if $0 < \text{rank}(\Pi) = r < k$ it can be shown that $\Pi = \alpha\beta'$ with α and β $(k \times r)$ matrices of rank r yielding the vector error correction model.

$$\Delta X_t = \alpha\beta' X_{t-1} + \sum_{i=1}^{p-1} \Phi_i^* \Delta X_{t-i} + \varepsilon_t \qquad (2)$$

Notice that since $\Delta X_t \sim I(0)$, we must have $\alpha\beta' X_{t-1} \sim I(0)$ otherwise (2) would be logically inconsistent.

- Notice that $\alpha\beta' X_{t-1} \sim I(0) \Leftrightarrow \beta' X_{t-1} \sim I(0)$.
- Let $\beta = [\beta_1, ..., \beta_r]$. That is, β_i is a column of β. Thus

$$\beta' X_{t-1} = \begin{bmatrix} \beta_1' X_{t-1} \\ \vdots \\ \beta_r' X_{t-1} \end{bmatrix}$$

Hence $\beta' X_{t-1} \sim I(0) \Rightarrow \beta_i' X_{t-1} \sim I(0)$. Thus the columns β_i of β are the *cointegrating vectors*, $\beta_i' X_{t-1}$ can be interpreted as long run equilibrium relations $\beta_i' X_{t-1}$ displays mean reversion, cannot drift too far from its mean.

- Let

$$\alpha = \begin{bmatrix} \alpha_{11} & \cdots & \alpha_{1r} \\ \vdots & \ddots & \vdots \\ \alpha_{k1} & \cdots & \alpha_{kr} \end{bmatrix}.$$

Thus

$$\alpha\beta' X_{t-1} = \begin{bmatrix} \sum_{j=1}^{r} \alpha_{1j}\beta_j' X_{t-1} \\ \vdots \\ \sum_{j=1}^{r} \alpha_{kj}\beta_j' X_{t-1} \end{bmatrix}.$$

Hence α contain the *adjustment coefficients*: $\alpha_{ij}\beta_j' X_{t-1}$ is the adjustment in ΔX_{it} in response to a disequilibrium in $\beta_j' X_t$.

Formally we have the following result:

Theorem (*Granger Representation Theorem*)

Let X_t be a $k \times 1$ and let

$$\Delta X_t = \sum_{\ell=0}^{\infty} \Psi_\ell \varepsilon_{t-j},$$

where ε_t is white noise with positive definite matrix $\Omega, \Psi_0 = I_k$, $\sum_{j=0}^{\infty} j \left| \Psi_{ij}(\ell) \right| < \infty$ $(i,j = 1, ..., k)$. If there are $0 < r < k$ cointegration relationships, then there is a $r \times k$ matrix β' such that $\beta' X_t$ is stationary. The matrix β' satisfies $\beta' [\sum_{\ell=0}^{\infty} \Psi_\ell] = 0$. Further, if X_t is a $VAR(p)$ process $\Phi(L) X_t = \varepsilon_t$, then there exists a $k \times r$ matrix α such that $\Phi(1) = -\alpha \beta'$ and

$$\Delta X_t = \alpha \beta' X_{t-1} + \sum_{i=1}^{p-1} \Phi_i^* \Delta X_{t-i} + \varepsilon_t$$

Remark: The *Granger's representation theorem* states that cointegration is equivalent to the existence of a reduced rank error correction model.

- α and β are not unique. Take any nonsingular $(r \times r)$ matrix Q and define

$$\alpha^* = \alpha Q', \ \beta^* = \beta Q^{-1} \Rightarrow \alpha^* \beta^{*\prime} = \alpha \beta' = \Pi$$

- Often useful class of identifying restrictions

$$\beta = \begin{bmatrix} I_r \\ -B \end{bmatrix}$$

where B is a $(k-r) \times r$. Thus if $X_t = (X'_{1t}, X'_{2t})'$, then $\beta' X_t = X_{1t} - B' X_{2t}$.

The system equation approach

Example:

- Bivariate VAR model X_t

$$X_t = \Phi X_{t-1} + \varepsilon_t,$$

$$\Phi = \begin{bmatrix} 0.8 & 0.2 \\ 0.4 & 0.6 \end{bmatrix}.$$

- Note that the model is equivalent to

$$\Delta X_t = \Pi X_{t-1} + \varepsilon_t$$

where

$$\begin{aligned} \Pi &= \Phi - I_2 \\ &= \begin{bmatrix} -0.2 & 0.2 \\ 0.4 & -0.4 \end{bmatrix}. \end{aligned}$$

- The roots of the characteristic equation $|I_2 - \Phi z| = 0$ are $z_1 = 1$ and $z_2 = 2.5 \Rightarrow$ The process has one unit root.
- Note that $\text{rank}(\Pi) = 1 \Rightarrow$ The process has one unit root and there is cointegration.

The system equation approach

Example:

- Writing the model in the reduced rank vector error correction form with $\beta = (1, b)'$ we have

$$\begin{aligned} \Delta X_t &= \Pi X_{t-1} + \varepsilon_t \\ &= \begin{bmatrix} -0.2 \\ 0.4 \end{bmatrix} \begin{bmatrix} 1 & -1 \end{bmatrix} X_{t-1} + \varepsilon_t \end{aligned}$$

Example:

$$\Delta X_{1t} = -0.2(X_{1,t-1} - X_{2,t-1}) + \varepsilon_{1t},$$
$$\Delta X_{2t} = 0.4(X_{1,t-1} - X_{2,t-1}) + \varepsilon_{2t},$$

Interpretation:

- If $X_{1,t-1} = X_{2t-1}$, ΔX_{1t} and ΔX_{2t} does not change much
- If $X_{1,t-1} > X_{2t-1}$, $\Delta X_{1t} \downarrow$ and $\Delta X_{2t} \uparrow$
- If $X_{1,t-1} < X_{2t-1}$, $\Delta X_{1t} \uparrow$ and $\Delta X_{2t} \downarrow$
- Deviation from the equilibrium level should be (partially) corrected in the next period, by X_{1t} and X_{2t}.

- General VECM for a vector $X_t = (X_{1t}, X_{2t}, \ldots, X_{kt})'$ of k time series:
  $$\Delta X_t = \Pi X_{t-1} + \sum_{i=1}^{p-1} \Phi_i^* \Delta X_{t-i} + \varepsilon_t$$
  where ε_t is vector white noise (mean zero, constant variance-covariance matrix, no (cross-)autocorrelation), and where Π and Γ_i are $k \times k$ matrices.
- Cointegration occurs if
  $$\operatorname{rank} \Pi = r < k, \qquad \Pi = \alpha\beta' = \alpha_1\beta_1' + \ldots + \alpha_r\beta_r',$$

- Johansen derived the likelihood ratio test for $H_r : \operatorname{rank} \Pi = r$ against the alternative $r < \operatorname{rank} \Pi \leq k$, in the VECM model with normally distributed errors ε_t.
- The test is known as the *trace test* (this is a LR test). The test statistics $\lambda_{trace}(r)$ can be expressed in terms of eigenvalues $\hat{\lambda}_i$ of a particular matrix. Its asymptotic distribution under the null hypothesis is a multivariate version of the Dickey-Fuller distribution.
- We reject for large positive values of the test statistic.

These tests may be used to estimate the cointegrating rank r in the following way:

1. Start with $r = 0$;
2. Test H_r with $\lambda_{trace}(r)$ (We reject H_r if the observed valued of $\lambda_{trace}(r)$ is larger than the critical value);
3. If H_r is not rejected, then $\hat{r} = r$; if it is rejected, replace r by $r + 1$ and go back to step 2;
4. If H_r is rejected for all $r = 0, 1, \ldots, k - 1$, then conclude $\hat{r} = k$ (this corresponds to a stationary system).

Remark: Just like with Dickey-Fuller test, we can to allow for deterministic terms (a constant and possibly a linear trend) in the VECM. The critical values of the test depend on the deterministic terms considered.

- Lag- Length (p) chosen using the BIC criterion for the VAR model in levels estimated by *Maximum Likelihood.*
- The estimators of the parameters of the VECM are obtained using *Maximum Likelihood* after identifying the number of cointegrating relationships.
- To construct the log-likelihood function it is assumed normality of the errors, though the asymptotic distributions of the test statistics do not depend on this assumption.
- Under the restriction $\beta' = [I_r, -B']$, the MLE for $\hat{B}$ is obtained from the unrestricted estimator.
- The asymptotic distribution of $\hat{B}$ is *not asymptotically normal*.
- However, resulting t-statistics for the individual elements of B are *asymptotically N(0,1)* under null hypothesis.
- and LR tests for restrictions on B have asymptotic χ^2 null distribution.
- Note that this approach is valid if the restriction $\beta' = [I_r, -B']$ makes sense.