

Econometric Modelling
Professor Sujata Kar
Department of Management Studies
Indian Institute of Technology Roorkee
Lecture 40
Basics of Cointegration

(Refer Slide Time: 0:31)

Part 1: Introduction to Econometrics Module 1: An Overview Module 2: Formulation of Econometric Modelling Module 3 & 4: Review of Basic Concepts Module 5: Types of Data	Part 5: Univariate Time Series Modeling Module 24 & 25: Problem of Serial Correlation Module 26 & Module 27: AR, MA & ARMA Processes Module 28 & 29: Modelling Trend & Seasonal Variations
Part 2: Overview of Classical Linear Regression Model Module 6 & 7: Simple Regression Module 8: Assumption of Classical Linear Regression Module 9: Properties of OLS Estimators Module 10: Hypothesis Testing	Part 6: Models with Binary Dependent and Independent Variables Module 30: Spline Function & Categorical Variables Module 31: Linear Probability Models Module 32: Probit and Logit Models Module 33: Tobit & Multinomial Logit Models
Part 3: Multiple Regression Analysis & Diagnostic Tests Module 11 & 12: Multiple Regression Module 13 & 14: Problems of Multicollinearity Module 15 & 16: Omitted Variables & Parameter Stability Module 17 & 18: Problem of Heteroscedasticity	Part 7: Multivariate Models Module 34: Panel Data Methods Module 35 & 36: Simultaneous Equations System Module 37: Introduction to VARs
Part 4: Statistical Inference Module 19: t-test Module 20 & 21: Wald test Module 22 & 23: F-test Module 24: Chow test	Part 8: Modelling Long Run Relationships Module 38 & 39: Stationarity & Unit Root Testing Module 40: Basics of Cointegration

 IIT ROORKEE  NPTEL ONLINE CERTIFICATION COURSE

Hello and I welcome you to the last module of this course on Econometric Modelling. In this last part that is part 8, we are discussing modeling long-run relationships and I have already discussed stationarity and unit testing procedures. The last thing that I am going to discuss is the basics of cointegration.

Now, before that, I need to tell you that how cointegration is actually related to stationarity and unit testing. Once we find a series non-stationary then so far, I have discussed how we can difference the series and make it a stationary one, because we cannot directly work with non-stationary series.

But then there is of course a possibility of working with non-stationary series as well because differencing leads to some kind of loss of information. So, in order to avoid that, what we can use is the concept of cointegration. Cointegration does not require us to difference the series but still, it can be used for estimation purposes. So, let us see what is cointegration all about.

(Refer Slide Time: 1:33)

Cointegration

- If a series has d unit roots then it is called an integrated process of order d and denoted by $I(d)$. If two time series X_t and y_t are $I(d)$, then in general any linear combination of them will also be $I(d)$; i.e. the residuals obtained from a regression of y_t on X_t will also be $I(d)$.
- If however, there exists a vector β such that $y_t - \beta x_t$ is of a lower order of integration, $I(d - b)$, where $b > 0$, then in 1987 Engel and Granger defined X_t and y_t as co-integrated of order (d, b) , denoted by $CI(d, b)$.
- This means that if y_t and X_t are $I(1)$ and u_t is $I(0)$, where $u_t = y_t - \beta x_t$, the two series would be cointegrated of order $CI(1, 1)$.
- This implies that if we wish to estimate the long-run relationship between y_t and x_t , it is only necessary to estimate the static model,
$$y_t = \beta x_t + u_t \quad (1)$$

IF ROORKEE NPTEL ONLINE CERTIFICATION COURSE

If a series has d unit-roots, then it is called an integrated process of order d and it is denoted by $I(d)$ that has already been discussed. So, a series having d unit roots is actually a non-stationary series. The way we used to call the non-stationary process of 1 unit root as $I(1)$ series and a stationary series can be denoted by $I(0)$ series with d unit-roots would be called an integrated series of order d and denoted by $I(d)$. If (refer slide time: 2:04- 3:30).

(Refer Slide Time: 3:30)

Engel-Granger (EG) Approach of Cointegration

- This regression achieves a consistent estimate of the long run relationship between the variables in the model. This arises because of the 'super-consistency' property of the OLS estimator when the series are co-integrated. According to the super-consistency property if y_t and x_t are co-integrated processes such that the error terms becomes an $I(0)$ process, then as sample size becomes larger, the OLS estimator of β converges to its true value at a much faster rate than the usual OLS estimator with stationary variables.
- The economic interpretation of co-integration is that if two or more series are linked to form an equilibrium relationship spanning the long run, then even though the series themselves may contain stochastic trend, they will nevertheless move closely together over time and the difference between them is constant or stationary.

IF ROORKEE NPTEL ONLINE CERTIFICATION COURSE

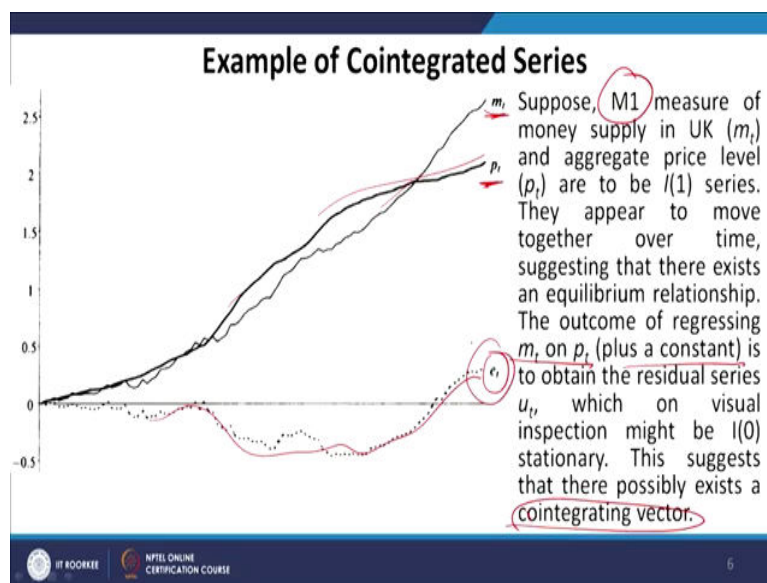
Now we are going to deal with primarily Engel-Granger's approach of cointegration. This regression achieves a consistent estimate of the long-run relationship between the variables in the model. This arises because of the super consistency property of the OLS estimator when the series are cointegrated.

So far, we discussed the concept of consistency, consistency was introduced in the very beginning itself, which is a property of the OLS estimators an asymptotic property or a large sample property. Now, there is something that comes is the concept of super consistency. So, according to the super consistency property, if y_t and x_t are cointegrated processes, such that the error term becomes an I (0) process. Then as the sample size becomes larger, the OLS estimator of β converges to the true value at a much faster rate than the usual OLS estimator with stationary variables.

So, first of all, the concept of consistency actually told us that β converges to its true value. So, when β converges to its true value with an increase in the sample size, then we call it a consistent estimator. But when β converges to its true value at a much faster rate than the usual OLS estimator, because of this cointegration properties, then we call it super consistency.

The economic interpretation of cointegration is that if two or more series are linked to form an equilibrium relationship spanning the long run, then even though the series themselves may contain a stochastic trends, they will nevertheless move closely together over time. And the difference between them is constant or stationary.

(Refer Slide Time: 5:21)



So, I will explain this concept with an example. Suppose the M1 measure of money supply, can be called the narrow definition of money in the UK which is denoted by m_t , and the aggregate price level denoted by p_t , are I (1) series. They appear to move together over time,

you can see that they are moving together over time suggesting that there exists an equilibrium relationship between them, they may exist.



The outcome of regressing m_t on p_t plus constant is to obtain the residual series u_t , which on visual inspection might be an $I(0)$ series. So, this is what is our, the estimated u_t , denoted by e_t , this is our estimated residuals. And this actually looks quite like a stationary series because it is hovering around 0 mean, or at least its variations are not increasing consistently over time. This suggests that there possibly exists a cointegrating vector between m_t and p_t which renders the series e_t , a stationary one.

(Refer Slide Time: 6:37)

Engel-Granger (EG) Approach of Cointegration

- To test the null hypothesis that y_t and x_t are not cointegrated amounts, in the EG framework, to directly testing whether $u_t \sim I(1)$ against the alternative that $u_t \sim I(0)$.
- There are several tests that can be used, including the Dickey-Fuller (DF) and the augmented Dickey-Fuller (ADF) tests. Engle and Granger (1987) advocated ADF tests of the following kind



$$\Delta \hat{u}_t = \theta^* \hat{u}_{t-1} + \sum_{i=1}^{p-1} \theta_i \Delta \hat{u}_{t-i} + \mu + \delta t + \omega_t \quad \omega_t \sim IID(0, \sigma^2) \quad (2)$$
- Where \hat{u}_t are obtained from estimating equation (1). The question of the inclusion of trend and/or constant terms in the test regression equation depends on whether a constant or trend term appears in (1). However, it is observed that irrespective of whether \hat{u}_t contains a deterministic trend or not, including a time trend in (2) results in a loss of power (i.e., leads to under-rejecting the null of no cointegration when it is false)



7

Cointegration

- If a series has d unit roots then it is called an integrated process of order d and denoted by $I(d)$. If two time series X_t and y_t are $I(d)$, then in general any linear combination of them will also be $I(d)$; i.e. the residuals obtained from a regression of y_t on X_t will also be $I(d)$.
- If however, there exists a vector β such that $y_t - \beta x_t$ is of a lower order of integration, $I(d - b)$, where $b > 0$, then in 1987 Engle and Granger defined X_t and y_t as co-integrated of order (d, b) , denoted by $CI(d, b)$.
- This means that if y_t and X_t are $I(1)$ and u_t is $I(0)$, where $u_t = y_t - \beta x_t$, the two series would be cointegrated of order $CI(1, 1)$.
- This implies that if we wish to estimate the long-run relationship between y_t and x_t , it is only necessary to estimate the static model,

$$y_t = \beta x_t + u_t \quad (1)$$



4

To test the null hypothesis that y_t and x_t are cointegrated amounts, in the EG framework that is Engel Granger framework to directly testing for whether u_t is $I(1)$ against the

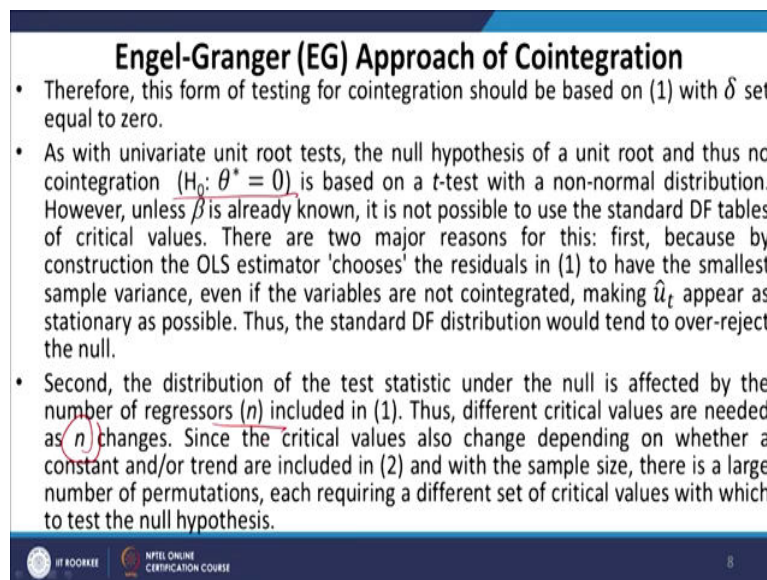
alternative that u_t follows an I (0) or u_t is an integrated series of orders 0, that is it is a stationary series.

There are several tests that can be used, including the Dickey-Fuller and the augmented Dickey-Fuller tests. Engel and Granger advocated ADF tests of the following kind. So, this is an ADF test of the estimated residuals to test that for whether the residuals are I (1) or I (0) and how we are doing it.

So, it actually follows the usual ADF structure that I had just discussed in the previous module, \hat{u}_t are obtained from estimating equation 1. The question of the inclusion of trend and or constant term, whether they should be included or not in the test regression equation depends on whether a constant or trend term appears in equation 1.

So, if the original equation includes a trend and a constant term then that should also be included in this specification otherwise not. However, it is observed that irrespective of whether \hat{u}_t contains a deterministic trend or not, including a time trend in 2, results in a loss of power that is it leads to under rejecting the null of no cointegration when it is actually false.

(Refer Slide Time: 8:17)



Engel-Granger (EG) Approach of Cointegration

- Therefore, this form of testing for cointegration should be based on (1) with δ set equal to zero.
- As with univariate unit root tests, the null hypothesis of a unit root and thus no cointegration ($H_0: \theta^* = 0$) is based on a t-test with a non-normal distribution. However, unless β is already known, it is not possible to use the standard DF tables of critical values. There are two major reasons for this: first, because by construction the OLS estimator 'chooses' the residuals in (1) to have the smallest sample variance, even if the variables are not cointegrated, making \hat{u}_t appear as stationary as possible. Thus, the standard DF distribution would tend to over-reject the null.
- Second, the distribution of the test statistic under the null is affected by the number of regressors (n) included in (1). Thus, different critical values are needed as n changes. Since the critical values also change depending on whether a constant and/or trend are included in (2) and with the sample size, there is a large number of permutations, each requiring a different set of critical values with which to test the null hypothesis.

IIIT KOOIKKEE NPTEL ONLINE CERTIFICATION COURSE 8

Engel-Granger (EG) Approach of Cointegration

- To test the null hypothesis that y_t and x_t are not cointegrated amounts, in the EG framework, to directly testing whether $u_t \sim I(1)$ against the alternative that $u_t \sim I(0)$.
- There are several tests that can be used, including the Dickey-Fuller (DF) and the augmented Dickey-Fuller (ADF) tests. Engle and Granger (1987) advocated ADF tests of the following kind

$$\Delta \hat{u}_t = \theta \hat{u}_{t-1} + \sum_{i=1}^{p-1} \theta_i \Delta \hat{u}_{t-i} + \mu + \delta t + \omega_t \quad \omega_t \sim IID(0, \sigma^2) \quad (2)$$

- Where \hat{u}_t are obtained from estimating equation (1). The question of the inclusion of trend and/or constant terms in the test regression equation depends on whether a constant or trend term appears in (1). However, it is observed that irrespective of whether \hat{u}_t contains a deterministic trend or not, including a time trend in (2) results in a loss of power (i.e., leads to under-rejecting the null of no cointegration when it is false)

Therefore, this form of testing for cointegration should be based on (refer slide time: 8:18-8:59). However, unless β is already known, it is not possible to use the standard DF tables of critical values. There are two major reasons for this, first, because by construction the OLS estimator chooses the residuals in 1 to have the smallest sample variance, even if the variables are not cointegrated making \hat{u}_t appear as stationary as possible.

Thus, the standard DF distribution would tend to over reject the null that is rejecting the null which is actually should not be rejected. Second, the distribution of the test statistic under the null is affected by the number of regressors n included in the very initial specification, the model involving y and x .

Thus, different critical values are needed as n changes. Now, here n refers to the number of regressors. Since the critical values also change depending on whether a constant and/or a trend is included in (2) and with a sample size, there is a large number of permutations, each requiring a different set of critical values with which to test the null hypothesis.

(Refer Slide Time: 10:16)

Engel-Granger (EG) Approach of Cointegration

- Fortunately, MacKinnon (1991) has linked the critical values for particular tests to a set of parameters of an equation of the response surfaces (φ s). That is, with the table of response surfaces, and the following relation,

$$C(p) = \varphi_{\infty} + \varphi_1 T^{-1} + \varphi_2 T^2$$

- where $C(p)$ is the p per cent critical value. It is possible to obtain the appropriate critical value for any test involving the residuals from an OLS equation where the number of regressors (excluding the constant and trend) lies between $1 \leq k \leq 6$. An extract from the table of response surfaces are given in the next slide. T is the number of observations.



So, a solution was given by MacKinnon in 1991, who linked the critical values for particular tests to a set of parameters of an equation of the response surfaces. So, we call these as response surfaces, (refer slide time: 10:30), for different sample sizes for different values of t , we would be calculating the critical values and then examine the validity of the test statistic, it is possible to obtain the appropriate critical value for any test involving the residuals from an OLS equation where the number of regressors excluding the constant and trend lies between 1 and 6. Here k refers to the number of regressors. An extract from the table of response surfaces is given in the next slide and T is the number of observations.

(Refer Slide Time: 11:30)

k	Model	p	φ_{∞}	φ_1	φ_2
1	No constant, no trend	1	-2.57	-1.96	-10.04
		5	-1.94	-0.40	0.00
1	Constant, no trend	1	-3.43	-6.00	-29.25
		5	-2.86	-2.74	-8.36
1	Constant + trend	1	-3.96	-8.35	-47.44
		5	-3.41	-4.04	-17.83
3	Constant, no trend	1	-4.30	-13.79	-46.37
		5	-3.74	-8.35	-13.41

For instance, the estimated 5% critical value for 105 observations when $k = 3$ in (1) and with a constant but no trend included in (2) is given by $(-3.7429 - 8.352/105 - 13.41/105^2) \approx -3.82$. Thus, reject the null of no cointegration at the 5% significance level if the t -value associated with θ^* is more negative than -3.82 .



Engel-Granger (EG) Approach of Cointegration

- Fortunately, MacKinnon (1991) has linked the critical values for particular tests to a set of parameters of an equation of the response surfaces (φ s). That is, with the table of response surfaces, and the following relation,

$$C(p) = \varphi_{\infty} + \varphi_1 T^{-1} + \varphi_2 T^2$$

- where $C(p)$ is the p per cent critical value. It is possible to obtain the appropriate critical value for any test involving the residuals from an OLS equation where the number of regressors (excluding the constant and trend) lies between $1 \leq k \leq 6$. An extract from the table of response surfaces are given in the next slide. T is the number of observations.



So, this is an extract from the table of response surfaces. So, first of all, we have considered several alternatives, like key values, are 1, 1, 1, and 3 and these are the models. The first model has no constant no trend, the second has one constant, but no trend. The third one has a constant plus trend. And the fourth one has 3 variables but no trend. So, constant and 2 other independent variables.

And then for different significance levels that is 1 percent and 5 percent. The φ_{∞} , φ_1 , and φ_2 values are given which were calculated by or given by MacKinnon. So, for instance, the estimated 5 percent critical values for 105 observations. So, here t is equal to 105 when k is equal to 3 in equation 1 with a constant but no trend included in equation 2 is given by this formula. So, this is actually obtained from simply like this (refer slide time: 12:30- 13:35).

(Refer Slide Time: 13:36)

Engel-Granger (EG) Approach of Cointegration

- The residual-based ADF test for cointegration just discussed, assumes that the variables in the OLS equation are all $I(1)$, such that the test for cointegration is whether $u_t \sim I(1)$ against the alternative that $u_t \sim I(0)$.
- If some of the variables are in fact $I(2)$, then cointegration is still possible if the $I(2)$ series cointegrates down to an $I(1)$ variable in order to potentially cointegrate with the other $I(1)$ variables.
- The critical values at the time of testing for cointegration when there is mix of $I(1)$ and $I(2)$ variables, are given by Haldrup (1994).



So, the residual-based ADF test for cointegration just discussed assumes that the variables in the OLS equation are all I (1) such that the test for cointegration is (refer slide time: 13:48-14:17).

(Refer Slide Time: 14:20)

Error Correction Models

- Let us consider a simple dynamic model

$$y_t = \alpha_0 + \gamma_0 x_t + \gamma_1 x_{t-1} + \alpha_1 y_{t-1} + u_t \quad u_t \sim IN(0, \sigma^2) \quad (3)$$
- The long run equilibrium between y_t and x_t is given as

$$y_t = \beta_0 + \beta_1 x_t \quad (4)$$
- Where $\beta_1 = \frac{\gamma_0 + \gamma_1}{1 - \alpha_1}$ and $\beta_0 = \frac{\alpha_0}{1 - \alpha_1}$
- The problems with this specification are, i) likely high level of correlation between current and lagged values of a variable, which will therefore result in problems of Multicollinearity. ii) Also, some (if not all) of the variables in a dynamic model of this kind are likely to be non-stationary, since they enter in levels. And iii) respecifying this model in terms of first differences removes any information about the long-run from the model and consequently is unlikely to be useful for forecasting purposes.

Handwritten notes: $x_t = x_{t-1}$, $y_t = y_{t-1}$

IPR KOOBEE | NPTEL ONLINE CERTIFICATION COURSE | 12

Now, we talk about a related concept that is error correction models. Let us consider a simple dynamic model. So, this is a dynamic model as you can see that we are also considering lag values of (refer slide time: 14:30- 15:10). The problems with this specification are first of all a likely high level of correlation between current and lagged values of a variable which will therefore result in problems of multicollinearity. So, we are, of course, talking about the first short-term dynamic model.



Second, also some if not all of the variables in a dynamic model of this kind are likely to be non-stationary. Say, since they enter in levels. And third, respecifying this model in terms of first differences, because they could be non-stationary series and removes any information about a long-run relationship from the model and consequently is unlikely to be useful for forecasting purposes.

(Refer Slide Time: 15:54)

Error Correction Models

- An alternative is to adopt the error-correction model or equilibrium correction model (ECM).
- Rearranging and reparameterizing (3) gives

$$\Delta y_t = \gamma_0 \Delta x_t - (1 - \alpha_1) [y_{t-1} - \beta_0 - \beta_1 x_{t-1}] + u_t \quad (5)$$
- The ECM in equation (5) has several advantages:
- First, and assuming that x and y are cointegrated, the ECM incorporates both short-run and long-run effects. Thus, if at any time the equilibrium holds then $[y_{t-1} - \beta_0 - \beta_1 x_{t-1}] = 0$. During periods of disequilibrium, this term is nonzero and measures the distance the system is away from equilibrium during time t . Thus, an estimate of $(1 - \alpha_1)$ will provide information on the speed of adjustment.





13

Error Correction Models

- Let us consider a simple dynamic model

$$y_t = \alpha_0 + \gamma_0 x_t + \gamma_1 x_{t-1} + \alpha_1 y_{t-1} + u_t \quad u_t \sim IN(0, \sigma^2) \quad (3)$$
 $x_t = x_{t-1}$
 $y_t = y_{t-1}$
- The long run equilibrium between y_t and x_t is given as

$$y_t = \beta_0 + \beta_1 x_t \quad (4)$$
- Where $\beta_1 = \frac{\gamma_0 + \gamma_1}{1 - \alpha_1}$ and $\beta_0 = \frac{\alpha_0}{1 - \alpha_1}$
- The problems with this specification are, i) likely high level of correlation between current and lagged values of a variable, which will therefore result in problems of Multicollinearity. ii) Also, some (if not all) of the variables in a dynamic model of this kind are likely to be non-stationary, since they enter in levels. And iii) respecifying this model in terms of first differences removes any information about the long-run from the model and consequently is unlikely to be useful for forecasting purposes.



12

So, an alternative is to adopt the error correction model or equilibrium correction model denoted by ECM. Rearranging and re-parameterizing 3, this was my initial model the short-term dynamic model 3, we rearrange the terms and write it like this. So, the ECM in equation 5 has several advantages, this is called an error correction model or equilibrium correction model or ECM in short.

So, the advantage with this kind of specification is that first of all assuming that x and y are cointegrated. The ECM incorporates both short-run and long-run effects. So, this is the short run specification in that you can also see the long run specification. Thus, if at any time the equilibrium holds then (refer slide time: 16:42).

During periods of disequilibrium, this term is nonzero and measures the distance, the system is away from the equilibrium during time t . Thus, an estimate of $1 - \alpha$ will provide

information on the speed of adjustment of the system from being disequilibrium to the equilibrium level in the long run.

(Refer Slide Time: 17:14)

Error Correction Models

- A second feature of the ECM is that all the terms in the model are stationary, so standard regression techniques are valid, assuming cointegration and that we have estimates of β_0 and β_1 .
- Third, Engle and Granger showed that if y_t and x_t are cointegrated CI(1, 1), then there must exist an ECM; and, conversely, that an ECM generates cointegrated series.
- The simple ECM depicted in (5) can be generalized to capture more complicated dynamic processes by increasing the lag length p and/or q .

$$A(L)\Delta y_t = B(L)\Delta x_t - (1 - \pi)[y_{t-p} - \beta_0 - \beta_1 x_{t-p}] + u_t$$
- Where $A(L) = 1 - \alpha_1 L - \dots - \alpha_p L^p$ and $B(L) = \gamma_0 + \gamma_1 L + \dots + \gamma_q L^q$ are the lag operators and $\pi = \alpha_1 + \alpha_2 + \dots + \alpha_p$. Further, it is also possible to specify the ECM in multivariate form, explicitly allowing for a set of cointegration vectors.

A second feature of the ECM is that all the terms in the model are stationary. So, standard regression techniques are valid, assuming cointegration and that we have estimated β_0 and β_1 . Third Engel and Granger showed that if y_t and x_t are cointegrated that is they are CI (1, 1) then there must exist an ECM, and conversely that an ECM generates cointegrated series.

The simple ECM depicted in 5 can be generalized to capture more complicated dynamic processes by increasing the lag length p and/or q . For example, we are having (refer slide time: 17:55- 18:38). This measures the speed of adjustment. Further, it is also possible to specify the ECM in multivariate form explicitly allowing for a set of cointegration vectors.

(Refer Slide Time: 18:51)

Methods of Parameter Estimation in Cointegrated Systems

- We will consider here **Engle-Granger 2-step method**.
- In *step 1* make sure that all the individual variables are $I(1)$. Then estimate the cointegrating regression using OLS. However, it is not possible to perform any inferences on the coefficient estimates in this regression. Save the residuals of the cointegrating regression, \hat{u}_t . Test these residuals to ensure that they are $I(0)$. If they are $I(0)$, proceed to Step 2; if they are $I(1)$, estimate a model containing only first differences.
- In *step 2* use the step 1 residuals as one variable in the error correction model $\Delta y_t = \beta_1 \Delta x_t - \beta_2 \hat{u}_t + v_t$
- It is now valid to perform inferences in the second-stage regression, i.e. concerning the parameters β_1 and β_2 (provided that there are no other forms of misspecification), since all variables in this regression are stationary.



Error Correction Models

- A second feature of the ECM is that all the terms in the model are stationary, so standard regression techniques are valid, assuming cointegration and that we have estimates of β_0 and β_1 .
- Third, Engle and Granger showed that if y_t and x_t are cointegrated $CI(1, 1)$, then there must exist an ECM; and, conversely, that an ECM generates cointegrated series.
- The simple ECM depicted in (5) can be generalized to capture more complicated dynamic processes by increasing the lag length p and/or q , u_{t-p} .

$$A(L)\Delta y_t = B(L)\Delta x_t - (1 - \pi)(y_{t-p} - \beta_0 - \beta_1 x_{t-p}) + u_t$$
- Where $A(L) = 1 - \alpha_1 L - \dots - \alpha_p L^p$ and $B(L) = \gamma_0 + \gamma_1 L + \dots + \gamma_q L^q$ are the lag operators and $\pi = \alpha_1 + \alpha_2 + \dots + \alpha_p$. Further, it is also possible to specify the ECM in multivariate form, explicitly allowing for a set of cointegration vectors.



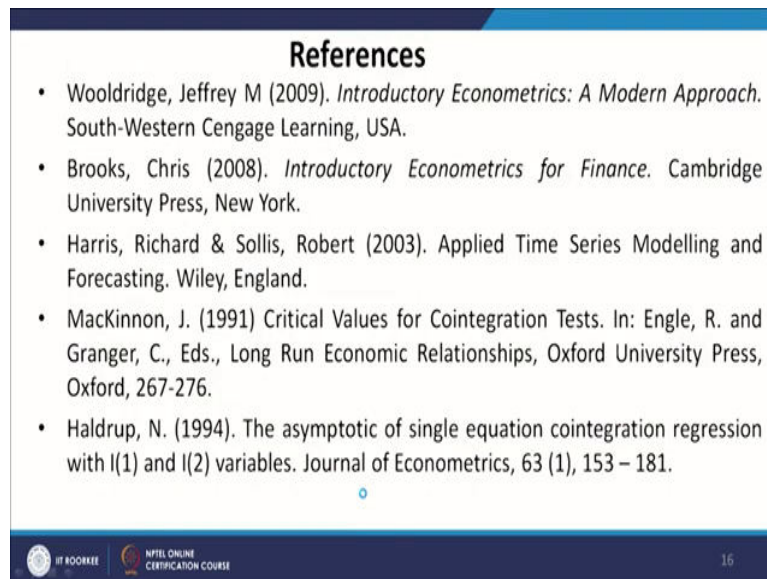
Finally, we talk about the methods of parameter estimation in cointegrated systems. So, we will consider here Engel-Granger 2 step method. In step 1, make sure that all individual variables are $I(1)$, then estimate the cointegrating regression using OLS. However, it is not possible to perform any inferences on the coefficient estimates in the regression because the error terms are not well behaved.

Then save the residuals of the cointegrating regression that is \hat{u}_t test these residuals to ensure that they are $I(0)$ if they are $I(0)$ then you proceed to step 2. If they are $I(1)$ then estimate a model containing only first differences, then we cannot do much. So, in step 2 use the step 1 residuals as one variable in the error correction model.

So, in the error correction model, this expression is actually equivalent to (refer slide time: 19:55- 20:26).

It is now valid to perform inferences in the second stage regression, that is concerning the parameters β_1 and β_2 provided that there are no other forms of misspecification since all variables in this regression are stationary. So, this is how once we have that parameter estimates of β_1 and β_2 , then, of course, they would give us the usual interpretations. And they are reliable estimates. So, that is all about the discussion on cointegration.

(Refer Slide Time: 20:58)



References

- Wooldridge, Jeffrey M (2009). *Introductory Econometrics: A Modern Approach*. South-Western Cengage Learning, USA.
- Brooks, Chris (2008). *Introductory Econometrics for Finance*. Cambridge University Press, New York.
- Harris, Richard & Sollis, Robert (2003). *Applied Time Series Modelling and Forecasting*. Wiley, England.
- MacKinnon, J. (1991) Critical Values for Cointegration Tests. In: Engle, R. and Granger, C., Eds., *Long Run Economic Relationships*, Oxford University Press, Oxford, 267-276.
- Haldrup, N. (1994). The asymptotic of single equation cointegration regression with I(1) and I(2) variables. *Journal of Econometrics*, 63 (1), 153 – 181.

IT KOOBEE NPTEL ONLINE CERTIFICATION COURSE 16

These are the references I have followed; some are mentioned here which are actually there in the discussion. And that is all about cointegration. Thank you.