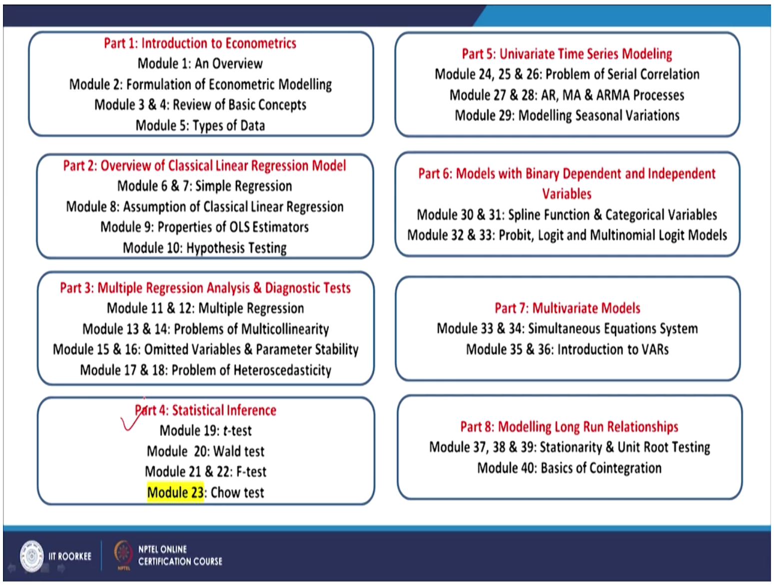
Econometric Modelling
Professor Sujata Kar
Department of Management Studies
Indian Institute of Technology Roorkee
Lecture 23
Chow Test

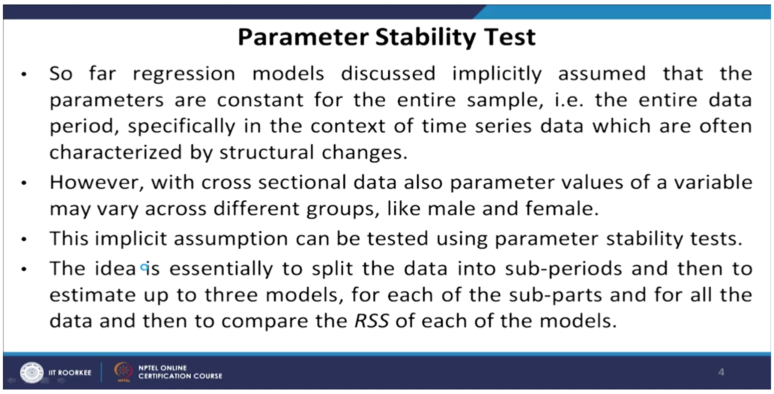
This Module 23 of the course of Econometric Modelling is on Chow test, which is basically an application of the F-test.

(Refer Slide Time: 00:37)



As you can see that we are in Part 4 and there we have been discussing statistical inferences. So, first we discussed a t-test, then Wald test and then F-tests in the last two modules. Basically, Wald test actually includes both t-test and F-test. F-test is a very important test and that is why that was discussed at length. And now Chow test, which is the topic of discussion in this module is basically is an application of F-test.

(Refer Slide Time: 01:15)



So, this Chow test is basically a kind of parameter stability test. There are other tests for parameter stability, but we would be focusing only on Chow test. So far, the regression models discussed, implicitly assumed that the parameters are constant for the entire sample, that is the entire data period specifically in the context of time series data, which are often characterized by structural changes.

Structural changes here is specifically a term used in the context of economics, where we say that because of some reason or other, there are certain changes that are inherent to some of the economic fundamentals or important measures. So, whenever such a situation happens, then we experience a structural change. Once there is a structural change possibly certain things change permanently or for quite a long period of time.

For example, in India, the structural adjustment program was introduced in 1990-91. So, that probably led to a structural change, though the change was not very sudden, that program was implemented and it was gradually being implemented and its impact was also felt over a long period of time. But then that is actually one policy or program that was introduced to bring in certain structural changes in the economy.

So, when such structural changes actually take place, then the economic fundamentals change permanently. So, if we assume that for a long period of time, the parameter estimates remain constant, then that may not be valid always, because if there are structural changes that have taken place in the economy, then that would lead to a change in those structural parameters or economic fundamentals permanently.

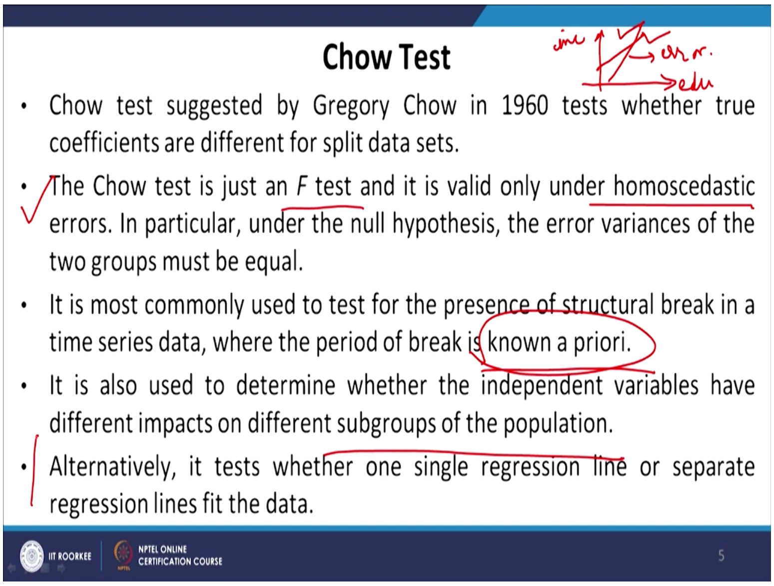
So, this is a test that is designed to capture whether the data is characterized by any structural change or not, or at least some of the parameters are characterized by structural changes or not. However, with cross-sectional data also parameter values of variables may vary across different groups like males and females.

For example, if we are considering the impact of education on income, then and we have considered both males and females in our sample, it is quite possible that the average level of income for male participants would be generally higher than the average income levels for women or female. So, as a result of which we may have some changes or parameters may not be stable for the entire sample.

Similarly, it may also vary from location to location, like in India we have Tier 1, Tier 2, Tier 3 cities, and we also have people from rural areas and average income from different cities or probably urban versus rural areas may vary, as a result of which if our sample consists of participants from a large number of locations or varied locations like including urban and rural areas, then their average income may vary despite having a very same or similar level of education. So, in that case, parameters are not stable across the entire sample.

So, whether parameters are stable or not is actually tested using the Chow test. This implicit assumption of parameter stability can be tested using parameter stability tests. The idea is essentially to split the data into sub-periods, and then to estimate up to three models for each of the sub-parts and for all the data and then to compare the RSS, that is the residual sum of the square, of each of the models.

(Refer Slide Time: 05:38)



So, the Chow test suggested by Gregory Chow in 1960 tests whether true coefficients are different for split data sets. So, when we are splitting the data set according to the characteristics of our cross-section the split could be for female and male or for different locations for time series the split could be depending on where we expect a structural change to take place. And then actually the Chow test tests whether the true coefficients, that is the population parameters, are different for the different data sets, the split data sets.

The Chow test is just an F-test and it is valid only under homoscedastic errors. In particular, under the null hypothesis, the error variances of the two groups must be equal. So, what we are contemplating here is that, suppose we have the data sets, if the data set is split and actually the parameters are not stable across two different groups, say male and female, so it is quite possible that we are having different intercepts and different slopes for two different groups, where we are measuring say income and education, so education is the independent variable, income is the dependent variable.

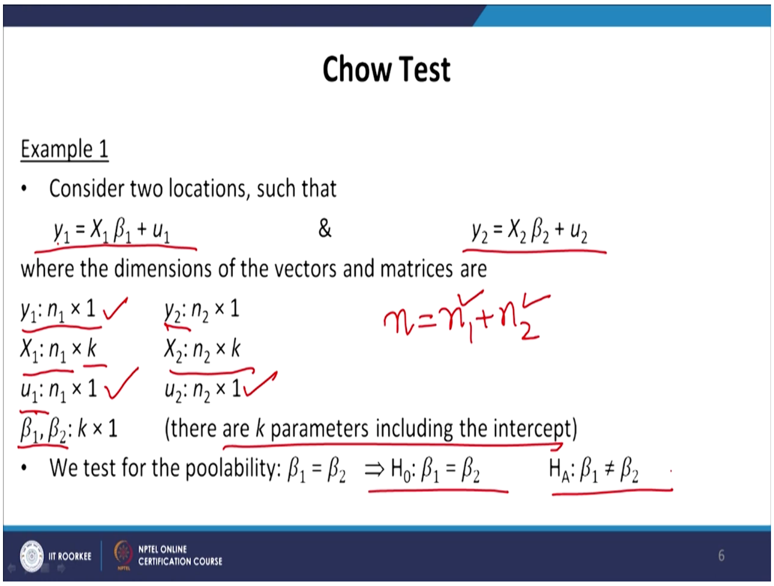
But what this paragraph says is that we can have different slopes and intercepts for different groups, but then the deviation of the estimated values from the actual ones measured by the error terms are actually not systematically, varying across observations or the dispersion in the error terms are not systematically connected to either the independent variables or time. So, the dispersions are still random, and as a result of which we have constant dispersions of the error terms across both groups.

So, this is the assumption of homoscedasticity which is required for the F-test, the Chow test to be valid, because as you can see that it is based on the F-test and the F-test also requires assumptions of CLRM to be fulfilled, that is classical linear regression model to be fulfilled. So, that is one of the classical linear regression model assumptions and that is why homoscedasticity must hold even if we do not have the same parameters across different groups.

It is most commonly used to test for the presence of a structural break in time-series data, where the period of breaks is known as a priori. So, this is an important characteristic of the Chow test that when we go for testing the presence of structural break, then we should know where actually the structure or the period from where the structural break is expected.

It is also used to determine whether the independent variables have different impacts on different subgroups of the population. So, this is actually a rather straightforward thing, like when we are working with cross-section and we expect different groups to have different slopes or intercept, then dividing the sample into two categories depending on the categories of the independent variable is actually not that difficult or it does not require any further assumptions, but this certainly does require. Alternatively, it tests whether one single regression line or separate regression lines fit the data. So, that is the basic precepts of the Chow test.

(Refer Slide Time: 09:28)



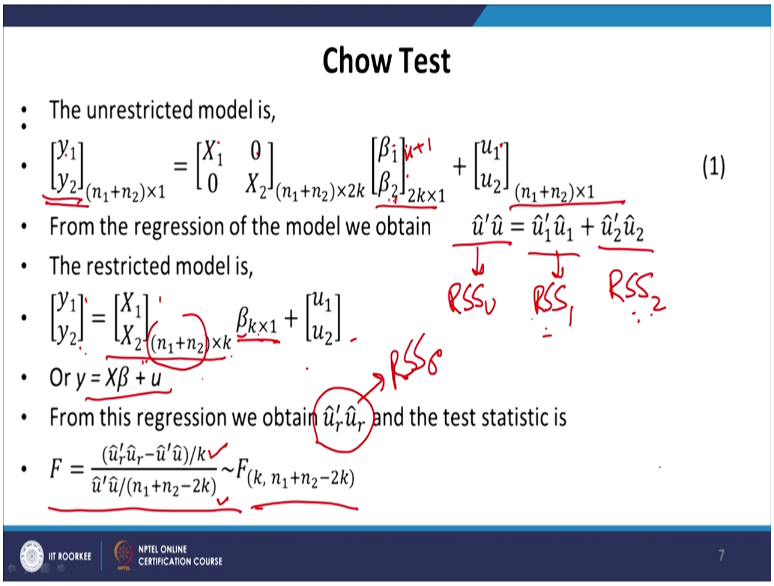
Now, we consider an example. The first example, suppose, considers two locations. So, for two locations, we have two different equations. And it is important to look at the dimensions of the dependent and independent variables along with the parameters and the error terms. *(Refer to slide time 09:58)*. So, here you can see that we have basically split that total observation which is given by n into two groups, one is n1 and the other one is n2.

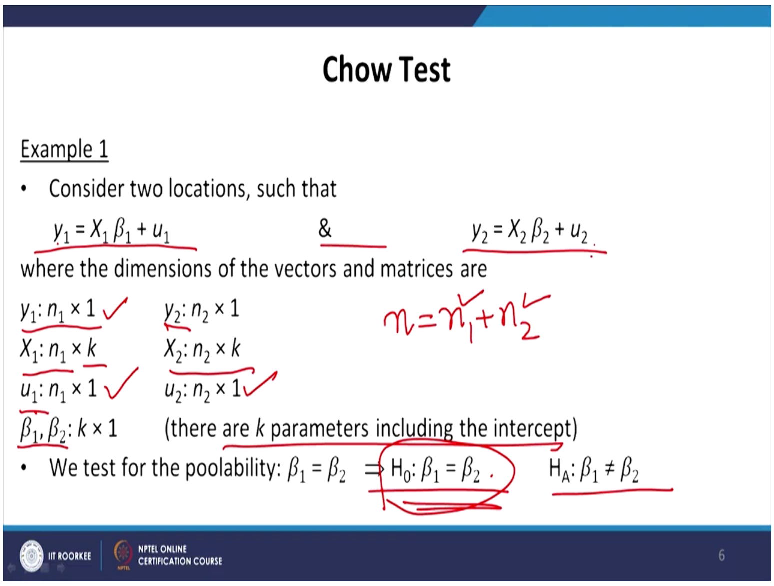
Now, it is not necessary that n1 and n2 have to be the same. Because if I am considering female and male, women or men, then it is not necessary that the number of male participants and the number of female participants should be the same. So, it is not at all necessary that n1 and n2 need to be the same.

Now, my y1 has n1 observations, my y2 has n2 observations. Similarly, corresponding to y1 we have x1 having n1 observations. The number of parameters is the same because for both the groups we are considering the same independent variables. So, that is k. Now, note here that k also includes the intercept term and x2 is n2 observations for all those k variables, including the intercept term *(Refer to slide time 09:58)*.

Of course, if y1 is n1 by 1, then u1 has to be n1 by 1, u2 is also n2 by 1, and beta 1, beta 2 are k by 1 vector. So, they basically consist of the parameter, population parameters for k variables including the intercept term, so basically k minus 1 independent variable, but we expect them to be different for the different groups and that is why they are being denoted as beta 1 and beta 2. So, we test for the probability, that is beta 1 equals beta 2. So, our null hypothesis says beta 1 equals beta 2. The alternative hypothesis is beta 1 is naught equal to beta 2.

(Refer Slide Time: 11:49)





So, first, we go for the unrestricted model. In the unrestricted model, you can see that the null hypothesis is beta 1 equals beta 2. So, the unrestricted model would treat them differently, while the restricted model would treat them the same. So, my unrestricted model has beta 1 and beta 2 stacked up. So, this is a vector of 2k by 1 dimension, this is k by 1 and this is k by 1, beta 2 is k by 1. So, in total, I have 2k plus, 2k by 1 dimension. Then y1 is, of course, y1 is n1 by 1, y2 is n2 by 1 *(Refer to slide time 11:49)*.

So, again they are stacked up vertically. So, my y1, y2 vector is n1 plus n2 by 1 dimension. You can see that the way this has been written would give us y1 equals x1 beta 1, 0 multiplied by beta 2 plus u1, which is actually y1 equals x1 beta 1 plus u1 and similarly, I have the equation for y2. So, we are having two different equations as you know shown initially for the two locations *(Refer to slide time 11:49)*.

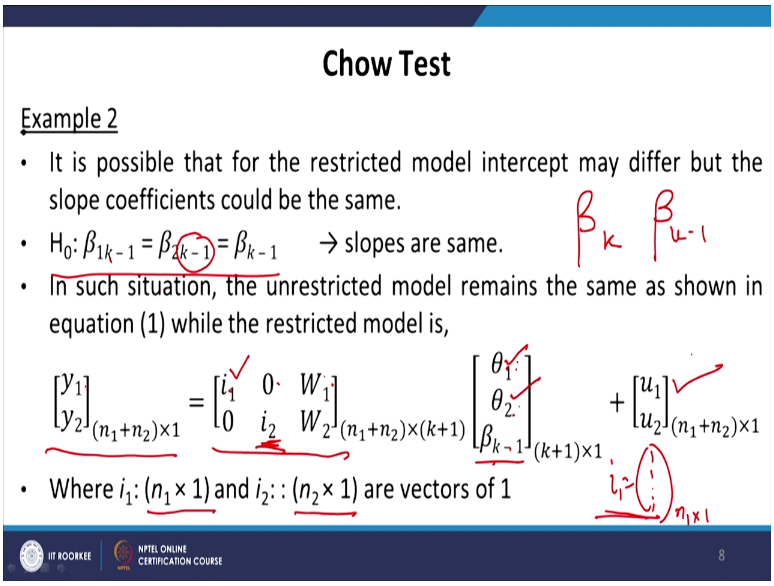
Now, from the regression, unrestricted regression model what we obtain is, u hat prime u hat which is basically u1 hat prime u1 hat, u2 hat prime u2 hat. So, this is RSS that is the residual sum of the square from the first model, this is the residual sum of the square from the second model, these are scalars or numbers and we add them up to get the residual sum of the square from the unrestricted model.

Now, we talk about the restricted model. This is y1, y2, x1, x2 and then since now we are assuming under the null beta 1 equals beta 2. So, we assume that the parameter coefficients or the estimated parameters, or the population parameters are actually not different across the two locations. So, we have the same parameter estimates. We do not need actually one beta 1 beta 2. We just need to have one beta, which consists of the k variables.

So, this is k by 1. This is n1 plus n2 by k. So, the entire observations or the entire sample and then k parameters. Similarly, u1, u2 would be n1 plus n2 by 1, y1, y2 is n1 plus n2 by 1. So, this model is actually the one if we collect a sample and just run a regression that is y equals x beta plus u and we do not go for any splitting of the sample, but this is my restricted model *(Refer to slide time 11:49)*.

So, from this regression, what we obtain is ur hat prime ur hat. So, this corresponds to RSS r that is the residual sum of the square from the restricted regression. And the test statistic is actually an F statistic which has its usual formula. You can see that there are k restrictions because beta 1 and beta 2 or beta consist of k independent variables. So, my numerator degrees of freedom is k and the denominator degrees of freedom in the remaining degrees of freedom which is n1 plus n2 minus 2k and it follows an F distribution with k and n1 plus n2 minus 2k degrees of freedom.

(Refer Slide Time: 15:25)



Next we consider another example. It is possible that for the restricted model intercept may differ but the slope coefficients could be the same. So, here we are only focusing on the slope coefficients excluding the intercept. So, we are allowing the intercept to vary. But testing whether the slope coefficients or the coefficients of the independent variables are the same or not. So, my null hypothesis is beta 1 k minus 1 equals beta 2 k minus 1 equals beta k minus 1*(Refer to slide time 15:25)*.

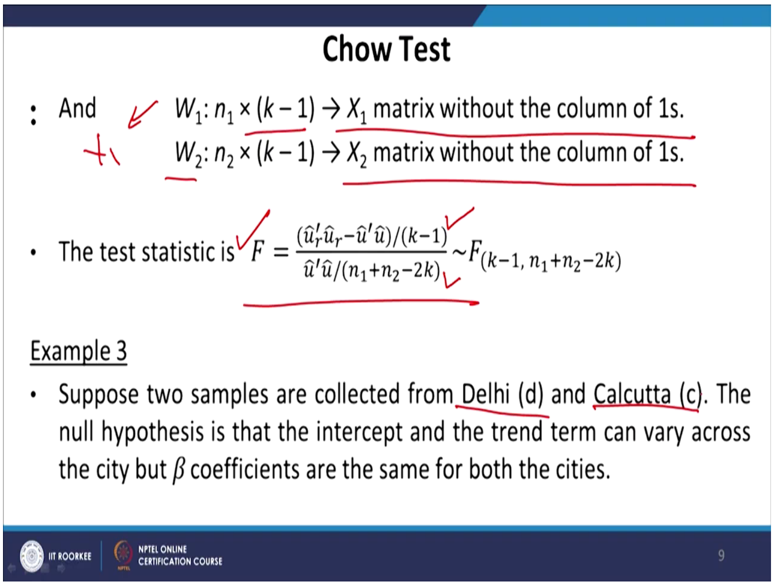
Here we are having k minus 1 notation in the subscript (*Refer to slide time 15:25)*. The reason is that if you remember beta consisted of k parameters. Now, the first parameter which is the intercept term is excluded from it. So, what we are trying to figure out is basically under the null hypothesis assuming that k minus 1 parameters are equal to each other. So, these k minus 1 parameters are associated with the k minus 1 explanatory or independent variables.

Now, in such a situation the unrestricted model remains the same as shown in equation 1 *Refer to slide time 15:25)*. So, our unrestricted model is the very same one, while a restricted model is now slightly different. And how does it look like? Now, since we are aligning the intercept to vary theta 1 and theta 2 refer to the intercepts. So, we are allowing intercepts to vary having separate intercepts from the two locations or two groups, but the same parameter estimates for the rest of the independent variables, k minus 1 independent variable. So, this is, of course, my usual dependent variable.

Now, the independent variable matrix has been actually split into two parts. This part i1 is actually an n1 by 1 vector and i2 is an n2 by 1 vector and they are all vectors of 1. So, which implies that i1 is a vector of 1, 1, 1, which is of n1 by 1 dimension and similarly the vector i2. Now, you can see that they represent the constant term because the constant term has the regressor as 1. So, this is actually the constant term for the first group, this is the constant term for the second group.

Now, if I multiply you would see that y1 equals i1 into theta 1, 0 theta 2, so no theta 2 actually, multiplied by w1, multiplied by, plus w1 multiplied by beta k minus 1 plus u1 and similarly for the y2 equation.

(Refer Slide Time: 18:28)

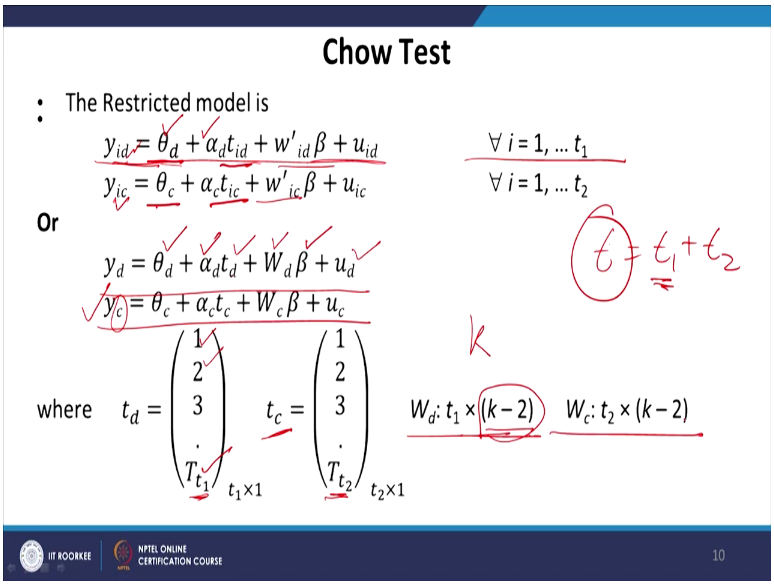


Now, my w1 is actually an n1 by 1 by k minus 1 matrix, because initially I had x1 and x1 consisted of all the variables plus the intercept term. Now, I have excluded the intercept term and that is why I am left with k minus 1 independent variable. The number of observations remains the same at n1. So, x1 this is actually x1 matrix without the columns of 1. Similarly, w2 is of dimension n2 by k minus 1 which is x2 matrix without the columns of 1. And the test statistic is the usual one.

Now, you can see that if I had a total number of parameters equal to k, then now the number of restrictions here, in this case, is k minus 1, because we are not putting any restriction on the intercept term. The restrictions are only with reference to the k minus 1 independent variables or parameter estimates. As a result of which my numerator degree of freedom is k minus 1. The denominator degrees of freedom remains the same because the unrestricted regression remains the same. So, this is what is the F statistic we calculate for the second example.

Now, we take another example to further clarify the process of the Chow test. Suppose two samples are collected again from two locations Delhi and Calcutta. The null hypothesis is that the intercept and the trend term can vary across the city, but beta coefficients are the same for both the cities.

(Refer Slide Time: 20:09)



So, this is actually very is similar to the second example, except for the fact that besides introducing or allowing the parameter, allowing the intercept terms to remain constant, we are also introducing another variable which is the trend term, and allowing its coefficient also to vary, but we are holding the coefficients of other independent variables constant.

Now, how does a trend term look like? First of all, this is the restricted model (*Refer to slide time 20:09*). Again, the unrestricted model may remain very similar, but, of course, here we have one more additional variable. Other than that, it is very similar to what we have already discussed. Our focus always remains on the unrestricted regression, because here we are actually incorporating the null hypothesis.

So, this is for an individual observation (*Refer to slide time 20:09)*, so for all i equals 1 to t1, now here we are denoting the observations by 1 to t1 and 1 to t2, where we can write that t equals to t1 plus t2. This is the number of observations. Now, this is why yid, d here reference to Delhi and c reference to Calcutta. So, for an individual observation, this is what we are, this is our equation, and then going by stacking all the observations for individual cities, this is our equations.

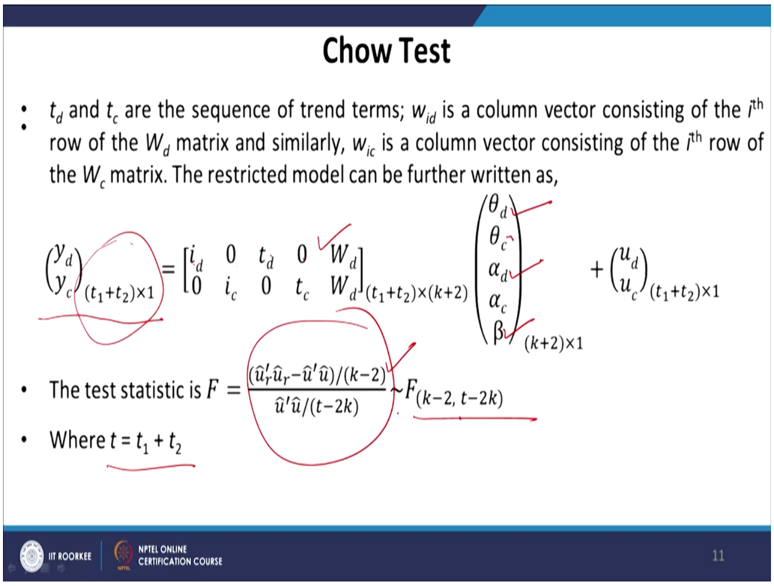
So, this is our intercept term, this is the trend term, this is corresponding, this is the parameter corresponding to the trend term, this is the parameter of all the explanatory variables and this is our matrix of all the explanatory variables, and finally, we have the error term and similar is the equation for the other location that is Calcutta (*Refer to slide time 20:09)*.

Now, how do we define the trend term? Trend term is basically, this is also called a deterministic trend when we include a deterministic trend when we expect the dependent variable to have some relationship with the time. So, here the time is measured as the first period is given a value 1, the second period is given a value 2 and so on that is how the last period here is given a value capital T which is associated with that t1 observation.

And similarly, this is capital T t2 which is associated with the last observation for the other city, Calcutta (*Refer to slide time 20:09)*. So, this is how we define the trend term. More about the trends and all will be discussed in a later module. And Wd is t1 by k minus 2 parameter estimates. This is the dimension.

Now, this is k minus 2 because we are assuming that k includes all the parameters including the intercept and the trend term or the parameter associated with the trend term. So, k minus 2 gives us the total number of parameter estimates minus the 2 which are allowed to vary. Similarly, Wc is t2 by k minus 2. These are the dimensions of these matrices.

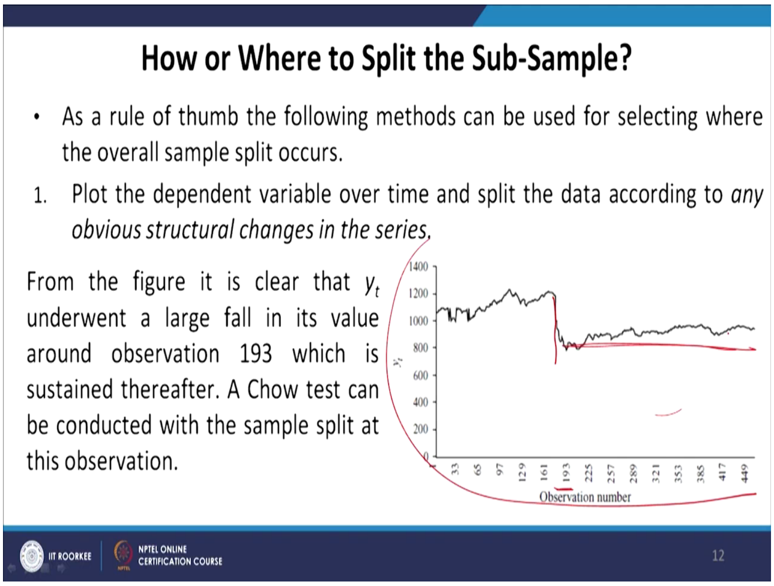
(Refer Slide Time: 23:32)



Td and tc are the sequences of trend terms, wid is a column vector consisting of the ith row of the Wd matrix, and similarly, wic is a column vector consisting of the ith row of the Wc matrix (*Refer to slide time 23:32)*. The restricted model can be further written as yd, yc which has a dimension of t1 plus t2 by 1. Then id is again the vector of 1s only. So, id multiplied by theta d, 0 multiplied by theta c implies that theta c does not exist in the first equation or the equation for Delhi.

Then td multiplied by alpha d and again 0 multiplied by alpha c means that alpha c does not appear in the equation for Delhi. And finally, we have beta Wd and similarly the equation for yc. Now, the test statistic is again the usual F statistic, but now the numerator degrees of freedom is k minus 2 because we have k minus 2 restrictions. So, it follows an F distribution with k minus 2 and t minus 2k degrees of freedom, where t is the sum of both the samples.

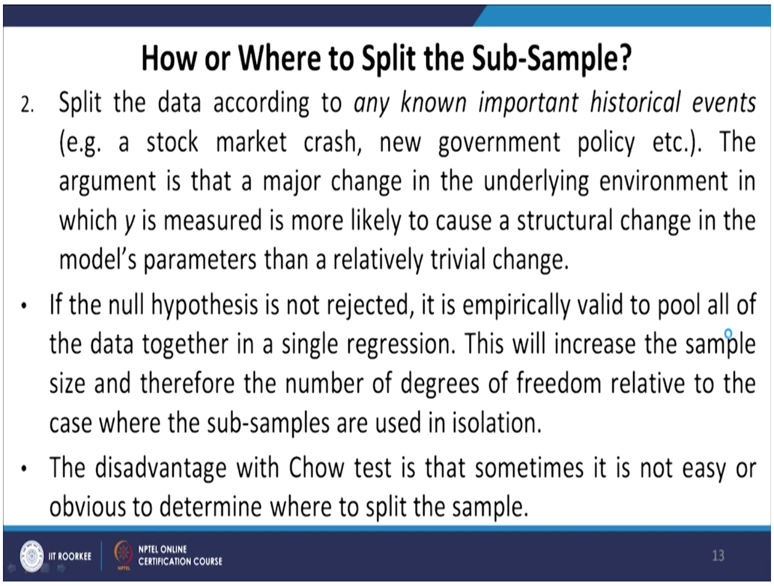
(Refer Slide Time: 24:52)



But the question that appears is or arises is how or where to split the sub-sample. As a rule of thumb, the following methods can be used for selecting whether an overall sample split occurs. Plot the independent variable over time and split the data according to any obvious structural changes in the series.

So, one example has been given here, where you can see that from the figure it is clear that yt underwent a large fall in its value around observation 193, which is sustained thereafter, which means it was not a temporary fall, this fall was sustained for the rest of the time period or observations. A Chow test can be conducted with the samples split at this observation. So, that is one procedure.

(Refer Slide Time: 25:44)

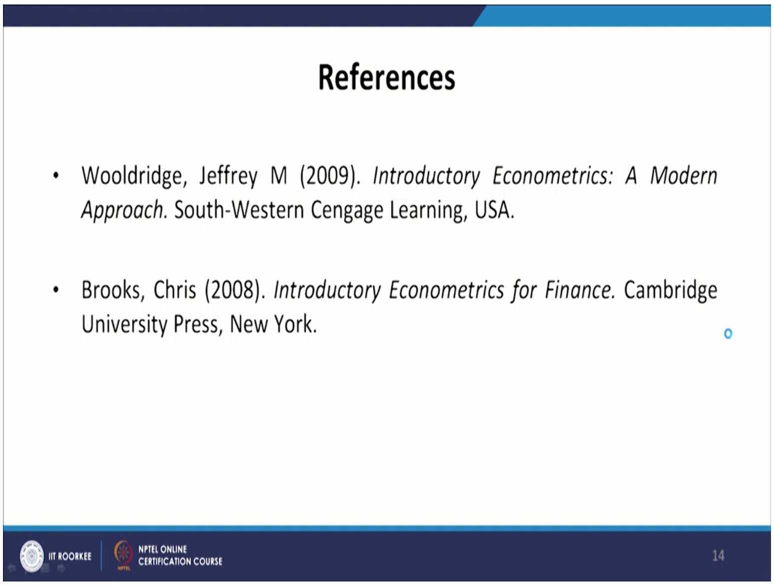


And the second procedure is to split the data according to any important known historical events. For example, a stock market crash, new government policy, etc. The argument is that a major change in the underlying environment in which y is measured is more likely to cause a structural change in the model's parameters than a relatively trivial change. So, structural change that is why I said telling you in the beginning implies more sustained an important change in the economic parameters and not a temporary and trivial change.

If the null hypothesis is not rejected, it is empirically valid to pull all of the data together in a single regression. This will increase the sample size and therefore the number of degrees of freedom relative to the case where the sub-samples are used in isolation.

The disadvantage with the Chow test is that sometimes it is not easy or obvious to determine where to split the sample. So, of course, there are alternative tests that give us clues about how to go about in those situations where we do not know what is like the sample, but that is for the time being is outside the purview of this course. So, that is all we discuss about the Chow test.

(Refer Slide Time: 27:06)



These are the references that I have followed. And this is where we complete the module or the part on statistical inference. In the next module, we will start with the discussion of basic time series models and their applications, their problems. Thank you.