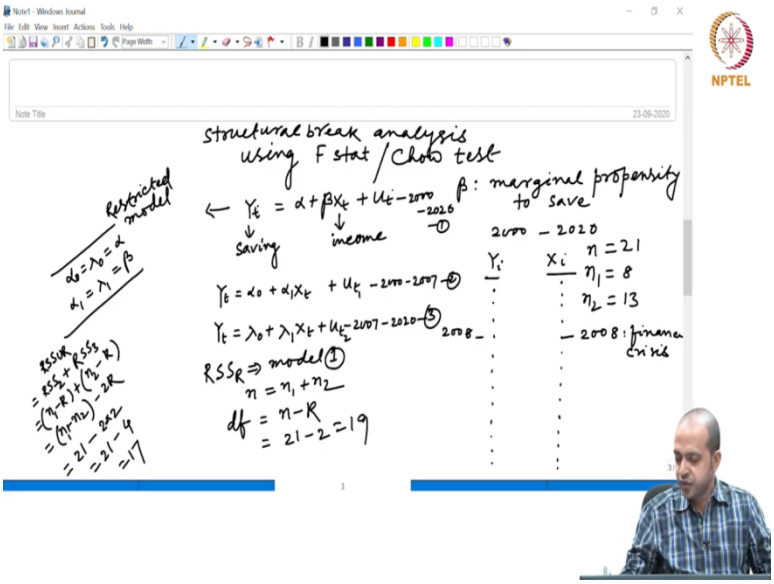
Introduction to Econometrics
Professor Sabuj Kumar Mandal
Department of Humanities and Social Sciences
Indian Institute of Technology, Madras
Lecture 30
Structural break analysis using Chow test Part - 3

(Refer Slide Time: 00:29)



And this particular test is called testing for structural break in time series data. So, what I will write, structural break analysis using F statistic. This is also known as Chow test. Please remember this particular test is applicable in a time series data. I will give you one example, suppose you are estimating a saving function, how does individual behave in the context of saving and we assume that Yi which is actually saving equals to alpha plus beta Xi plus ui where Yi is saving and X is income, excess income and what is beta? That means when income changes by 1 unit on an average saving changes by beta unit and that is called marginal propensity to save.

So, beta is called marginal propensity to save. And you have time series data and the data is available let us say from 2000 to 2020, for this last 21 years you have data on Yi, you have data on Xi for 21 years. So, here actually instead of i, since it is a time series phenomenon I should use t. So, this should be t. Yt equals to beta Xt plus Ut, this is a time series phenomenon. Now, when I am using this equation, what I am doing actually? I am implicitly assuming, let us say that when you have 2000 to 2020 data, if you recall in 2007 let us say this is 2007 or 2008.

Let us say this is 2008. In 2008 Indian economy or the global economy itself experienced with financial crisis. So, in 2008 what happened there was a financial crisis. So, because of this financial crisis which actually acted as an external shock to the global economy as well as Indian economy, it might so happen that the consumption and saving behavior of the people changed drastically post-financial crisis. It may so happen and as a researcher, that is my hypothesis or that is my guess. I am hypothesizing that the saving income relationship post-financial crisis is quite different from the saving income relationship pre-financial crisis.

So, that means from 2000 to 2007, the behavior of consumption and saving, consumption saving behavior from 2000 to 2007 is quite different from saving income relationship, sorry not consumption saving income relationship from 2008 onwards. That means post-financial crisis onwards. So, that means when I am estimating this relationship using a single equation for the entire period, my implicit assumption is that there is no structural break that means pre and post-financial crisis, there is no significant difference in individual's saving behavior but that is a very tall assumption to make, okay.

So what we will do now, let us formulate two different models, two different equation to represent saving-income relationship. One for 2000 to 2007 and another for 2008 to 2020 and how I am doing that? Let us say that Yt equals to let us say alpha 0 plus alpha 1 beta Xt plus Ut.

This is let us say 2000 to 2007, pre-financial crisis. Then post financial crisis, this is Yt equals to let us say some sorry this is alpha 1 there is no beta actually. Alpha 0 plus alpha 1 Xt plus Ut and this is post-financial crisis, that model I am representing as let us say lambda 0 plus lambda 1 Xt plus Ut. This is for 2007 to 2020 and this is for the entire period 2000 to 2020.

Let us say that this is model 1, this is model 2 and this is model 3. Now, out of these 3 models we are going to apply the F statistic or F test and F statistic will always require restricted RSS and unrestricted RSS. So, that means out of these three models, we have to identify which is our unrestricted model and which is our restricted model.

Now, if you think logically then you may easily understand that out of these three, model 1 is the restricted version. Why model one is restricted version? This model we are calling as restricted version. Why this is called restricted version? Because we are imposing a restriction and what is that restriction? The restriction is lambda sorry, we are imposing the restriction as alpha 0 equals

to lambda 0 equals to alpha. And alpha 1 equals to lambda 1 equals to beta. So, that means even post-financial crisis also there is no significant change in the intercept as well as slope of the saving-income relationship.

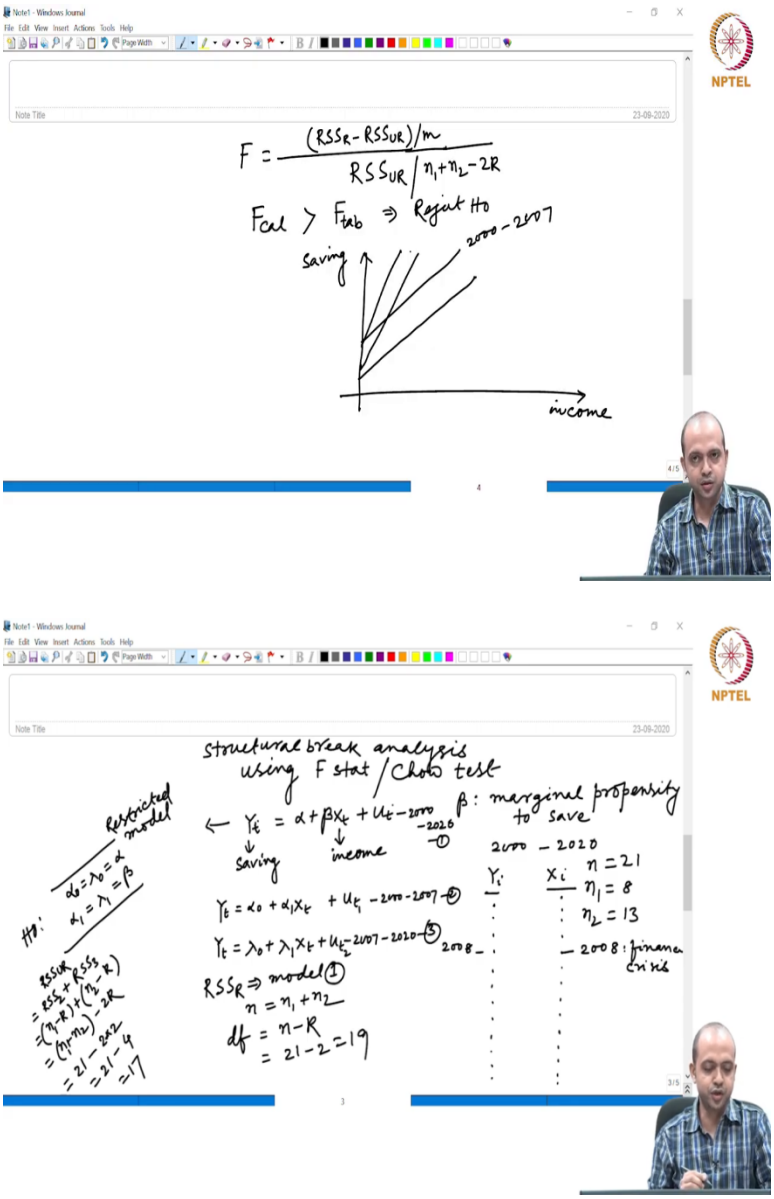
This is the restriction we have imposed in this model 1, that is why model 1 is called restricted model and we have 2 unrestricted versions. Model 2 and model 3, now what we have to do? We have to identify the RSS and their corresponding degrees of freedom. So, RSS restricted we will get from model 1 and what would be the degrees of freedom? Let us say that we have total number of observation in where n equals to let us say n1 plus n2.

So, total n equals to how much? 2000 to 2020 so n equals to 21. Total number of observation we have 21 years data and what is n1? n1 is basically 2000 to 2007, so 8 and in 2008 to 2020 which is basically 13. 13. So, that means RSS R, the restricted version what would be the degrees of freedom? For this, this would become n minus k, so that means the degrees of freedom for the restricted version equals to n minus k which is equals to total number of observation 21 minus total parameters to be estimated which is 2 equals to 19.

So, total so degrees of freedom for the restricted model is 19, and what would be the degrees of freedom for unrestricted? Since these 2 time periods are independent, please keep in mind what we are assuming. We are assuming that these two time periods are independent and as a result of which what will happen, that we will get RSS UR from RSS from model 2 plus RSS from model 3. We are assuming these 2 samples are independent to each other, as a result of which these two RSS are independent to each other.

So, that means basically we are saying that this if you say this is Ut1 and this is Ut2, they are independent. There is no relation. So, if you do so then RSS 2 degrees of freedom would be n1 minus k and this also would become in 2 minus k, so this would become n1 plus n2 minus 2k and what is n1 plus n2? That is n and what is n here? It is 21. So, 21 minus 2K and what is K? k is always 2, 1 beta and 1 alpha so that means one slope coefficient, one intercept coefficient and as a result of which 2 into 2, 4 so 21 minus 4 equals to 17. So, this is the degrees of freedom for the unrestricted model.

(Refer Slide Time: 14:34)



So, now how will you construct the test statistic, I will write the test statistic you have to construct in this way. So, your F would become RSS, this would become your RSS restricted minus RSS unrestricted divided by m number of restriction and this should become RSS UR and what is the degrees of freedom? This is n1 plus n2 minus 2k which is 17 in our case. So, this is how after estimating the restricted version and the unrestricted version that means you have to estimate 3 different models.

One for pre-financial crisis, one for post-financial crisis and one for the entire time period. So, these three equations you have to estimate to construct this F statistic. And this will become your F calculated and again, you have to compare the F with your tabulated value. So, that means from the F table if it is greater than this, then you reject your null. Reject h naught and what is h not? There is no significant difference in the saving-income relationship post-financial crisis.

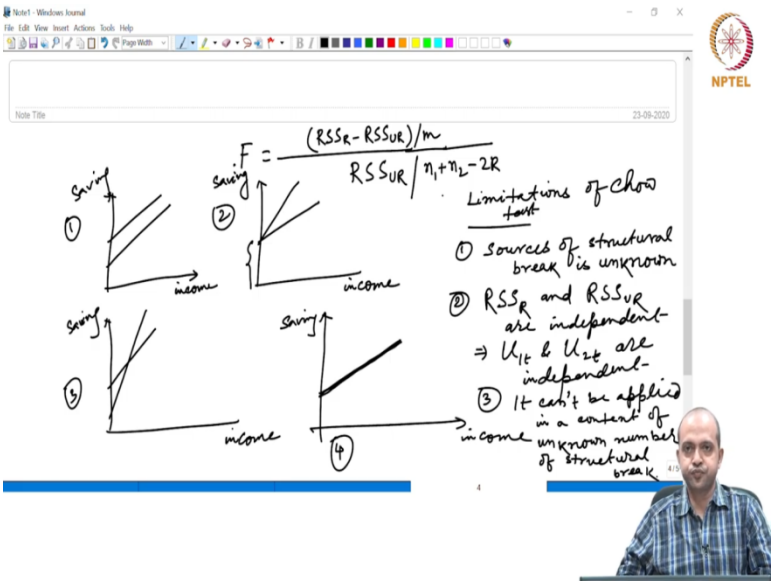
That means you are basically very, sorry there is significant relationship because you are rejecting the null. When you are rejecting the null that would be your this thing that means my null hypothesis is basically this one. You are rejecting this alpha 0 equals to lambda 0 equals to alpha that is rejected. So, post-financial crisis, there is significant difference in intercept and this. But you have to keep in mind that one thing, this is my null and what is my alternative hypothesis?

Alternative hypothesis is very tricky as always in the context of F statistic you have to very-very carefully keep in mind about the alternative of F statistic. It is not alpha 0 not equals to lambda 0 not equals to alpha 1 not equal to gamma lambda 1 not equals to beta. So, that means I cannot say that both intercept and slope are different post-financial crisis. That is not my alternative hypothesis. Alternative hypothesis should be at least one among these two parameters are different.

So, that means there would be either change in slope or change in intercept or change in both. If changes are there in both that is well and good but that is not my alternative. My alternative is always either there is a change in the slope or there is a change in the intercept or both, right. So, that means if you try to understand using a simple diagram, let us say, in the X-axis I am measuring income and in the Y-axis, we are measuring saving. So, let us say, this is my pre-financial crisis relationship 2000 to 2007.

Now, post-financial crisis relationship could be either this, so that means what is happening here? There is change in slope but there is no change in intercept. Or it could be this. There is change in intercept but there is no change in slope or it could be this, there is change in both slope and intercept.

(Refer Slide Time: 20:23)



Let me write 4 diagram here representing four possible cases. So, what will happen? Let us say this is case 1, where there is change in only intercept but there is no change in slope. This is income saving. Then case 2, there is income and then saving. So, here what is happening? There is change in slope but there is no change in intercept. Intercept remain same, then case 3 would be there is income saving. What is happening that means here? Here there is a change in both slope and intercept post-financial crisis and it may also happen in this way that means there is change neither in slope nor in intercept. So, post-financial crisis saving income relationship is exactly like the pre-financial crisis.

There is no significant difference, the saving income behavior of the people post financial crisis. This is case 1. This is case 2. This is case 3, this is case 4. Now, the problem the limitation with this Chow test is that Chow test will only tell you whether post-financial crisis saving-income relationship is significantly different from pre-financial crisis or not. But Chow test cannot tell you whether the change is due to the change in slope here sorry here or change in intercept or change in both. So, out of these three cases what is actually happening?

That means what is the source of this structural change in saving-income relationship? That the chow test cannot tell. So chow test has two major limitations. So, if I write limitation of Chow test, then first of all, sources of structural break is unknown. We don't know whether the change

is due to slope or intercept or both and what is the other limitation? If you recall that we have assumed two samples 2000 to 2007 and 2008 to 2020.

These two samples are independent that means RSS R and RSS UR, they are independent which implies that U1t and U2t, the error terms of the two different time periods are independent. There is another limitation, if you think see we say, in this data we are how we are dividing the sample, we are imposing the structural break in 2008 because we know that in 2008 only the financial crisis happened. So, we assume that is the point of break that is the structural break but it may so happen that if you have saving-income relationship for a particular country for the last 50 years.

Sometimes you may not be knowing that in which particular year there is a break and how many breaks are there here we have assumed only one break. It may so happen that there are multiple breaks. So, this Chow test cannot handle a situation where the data set exhibits unknown number of structural breaks. First of all, we do not know when exactly structural break happened and what is the number of such breaks? This is a simple test so that is why those things we cannot handle here.

For that we need to apply advanced econometric time series model. So, the third limitation of the Chow test is, we cannot, it cannot handle. So, that means, it cannot be applied in a situation context of unknown number of structural breaks. These limitations are there with the Chow test and we will try to overcome these limitations when we will be discussing about dummy variable model. So, we have our next module would be dummy variable model, there will try to overcome the limitation of Chow test. So, so far we have only discussed about the mechanism of F statistic or Chow test for the analysis of structural break.

So, what we will do next, we will take a data set and then we will try to implement this Chow test empirically. We will learn how to implement the Chow test. That will do in our next class using empirical data. So, application of chow test using empirical data, we will be using a particular data of saving an income in the context of US and then we will try to implement this particular Chow test. That we will discuss in our next class. So, today's class, we are closing here up to this level discussion of Chow test in the context of structural break analysis.