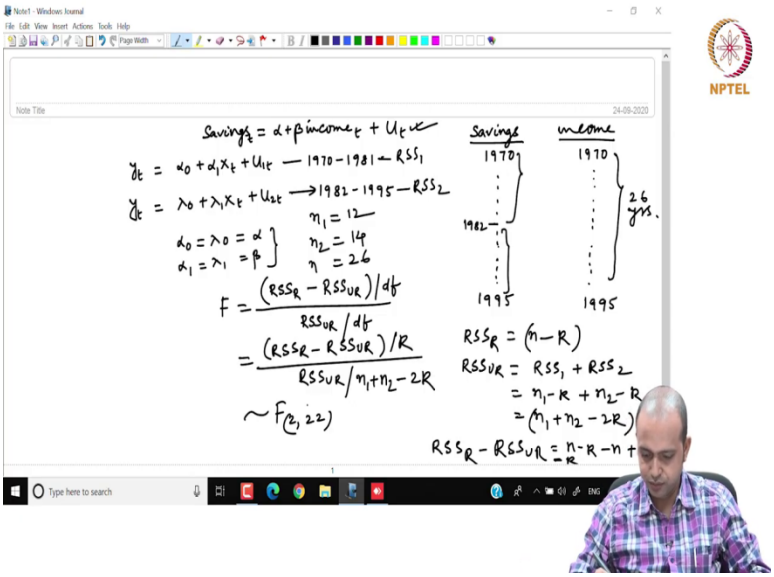
Introduction to Econometrics
Professor Sabuj Kumar Mandal
Department of Humanities and Social Sciences
Indian Institute of Technology, Madras
Lecture 31
Structural Break Analysis Using Chow Test Part - 4

So, we were discussing about structural brake analysis using chow test and we are using the F statistic for conducting that chow test or structural brake analysis. So, today what we will do, we will use one data set from US that is the savings income relationship and we have a time series data for twenty six years 1970 to 1995. Twenty six years data we have on savings and income. So, what we will do? We will first try to recap the F statistic that we are conducting that we are discussing yesterday.

(Refer Slide Time: 01:17)



So, our model is let me write my model, savings equals to alpha plus beta. This is savings t since this is a time series data beta income at time period t plus U t. This is our relationship savings income relationship in the context of US and we have data on savings and income, savings and income starting from 1970 to 1995, 1970 to 1975 sorry 1995. So, total we have twenty six years data, total twenty six years on savings and income.

Now, here what happened in 1982? In 1982, we hypothesize that there is some kind of structural break in the savings income relationship in 1982. What is the reason? Because in 1982 US economy suffered from the post peacetime recession; what was happening? The peacetime

recession was there in US economy and because of that peacetime recession, we hypothesize that there might be some kind of structural change in the savings income relationship. So, what we will do then?

We will specify two different type of savings income relationship, one for the period 1970 to 1981 and another from 1980 to 1995. So, these are the two sub samples we have created. So, we will as we have discussed, we have three regressions, actually three models, two unrestricted models and one restricted one. What are those unrestricted model? Unrestricted models are like this yt equals to, let us say alpha 0 plus alpha1 xt plus u1t, let us say. This is for the period 1970 to 1981. y denote savings, and x denotes income, then we will have another regression, which is yt equals to let us say, lambda 0 plus lambda1 xt plus u2t.

This is let us say, for the period 1982 to 1995 and the initial equation was this is the restricted equation and what is the assumption not we are making when you are running the on when we are running the restricted model that means, when you are assuming that there is no structural break, we are basically assuming that alpha 0 equals to lambda0 equals to alpha and alpha1 equal 2 lambda1 equals to beta.

This is that restriction that we are imposing and what is the F statistic that we have constructed? The F statistic that we constructed was we will run these two restricted, these two unrestricted models as well as the restricted one they not will do, we will have RSS, we will collect the RSS from the Restricted Model minus RSS Unrestricted Model divided by its degrees of freedom and denominator would be RSS UR.

So, F statistic was, sorry once again I will write, this is RSS restricted minus RSS unrestricted divided by it is degrees of freedom and what would be the denominator? Denominator is actually RSS unrestricted divided by its degrees of freedom and what is the degrees of freedom for the numerator? See, the numerator has RSS Restricted minus RSS Unrestricted and RSS Restricted; the Restricted RSS formulation had the degrees of freedom like this.

So, when I am writing RSS restricted it is basically that means for the entire model I am running single regression and that is in minus that is basically (n-k). Now, the question is how will you, how will you get the degrees of freedom for the unrestricted model? So, unrestricted model

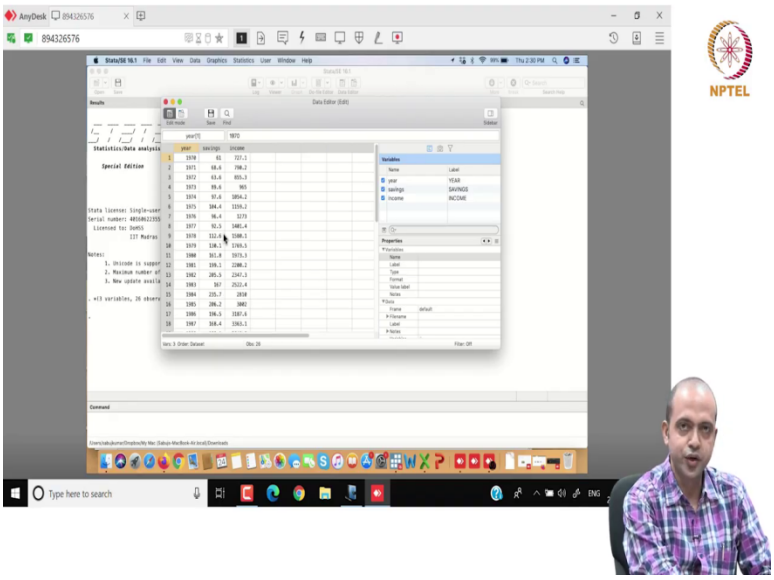
means we have two models here and both of them, both of them in both the models the RSS will have, so this is basically let us say for this period this is RSS1 and this is RSS2.

So, this is actually RSS1 plus RSS2 and RSS1 has degrees of freedom n1 minus k and this is plus n2 minus k. So, that means n1 plus n2 minus 2k. Now, if you calculate this RSS R minus RSS UR. So, RSS R minus RSS UR would become n-k minus n+2k - 2k. So, it will become k. So, that is why the degrees of freedom for the numerator would be k RSS R minus RSS UR divided by k and unrestricted model is basically RSS UR minus it will become n2 plus k or you can simply write (n1- k) + (n2 – k).

So, as I said (n1 + n2 - 2k). So, what do you can directly write sorry, you can directly write how you have derived unrestricted? Look at this unrestricted is n1 plus n2 minus 2k. So, I will write (n1 + n2 - 2k). This is the formula. So, in this particular case, we have twenty years data. So, that means (n1 + n2) is 26 and what is n1? n1 is 1970 to 1995, that means n1, n1 equals to 1970 to 1981, which is actually 12 and what is n2? Which is 1982 to 1995, that means 14 and what is k here?

k is the total number of parameters which is 2. So, n equals to 26, n1 plus n equals to n, which is actually 26. So, the degrees of freedom for the numerator would be 2 while degrees of freedom for the denominator would be (n-2k). So, that means 26 minus 4 equals to 22. So, this F will follow an F distribution with degrees of freedom 2 and 22, 2 for the numerator and 22 for the denominator. So, now what we will do? We will use that particular data in the context of US and we will estimate the model.

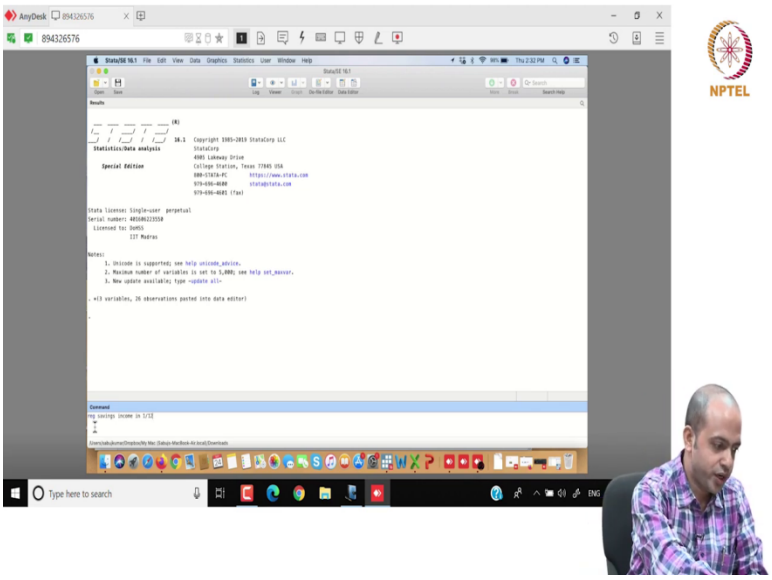
(Refer Slide Time: 12:49)

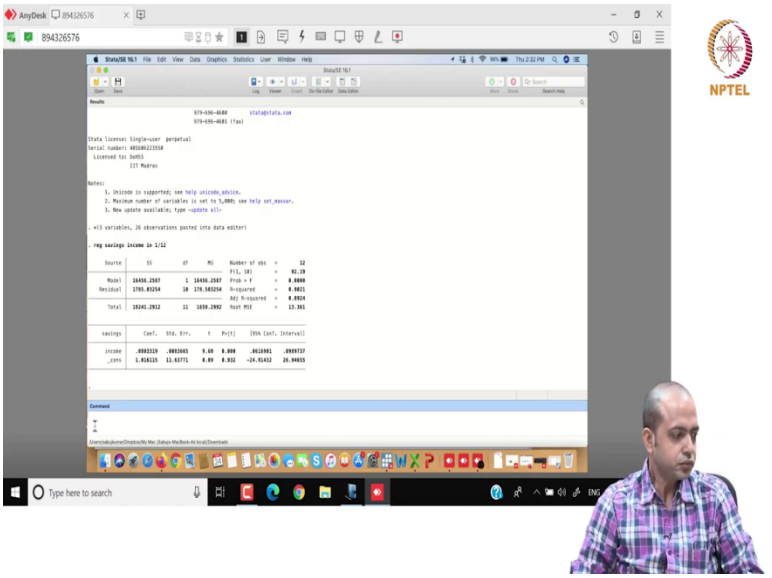


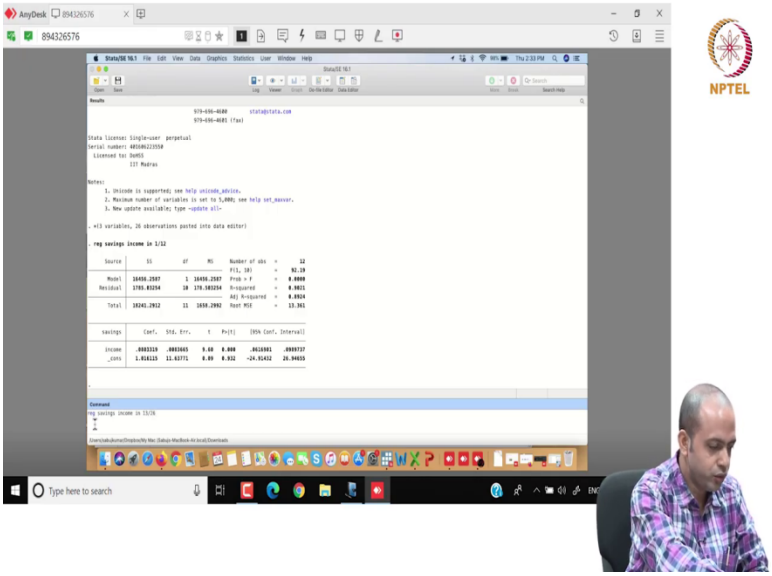
So now, what we will do? We will use the data from the US context and I will first show you what is the data, how does the data look like? If you click here, then you can see the data. See this is the data from 1970 to 1982, 1970 to 1995 we have savings income relationship. So, we have only two variable here the dependent variable is savings, which is a function of income. So, we have to run basically three regressions here, one for the period 1970 to 1991 and another for 1982 to 1995. These are the two unrestricted models and one for the entire time period, which is basically the restricted model.

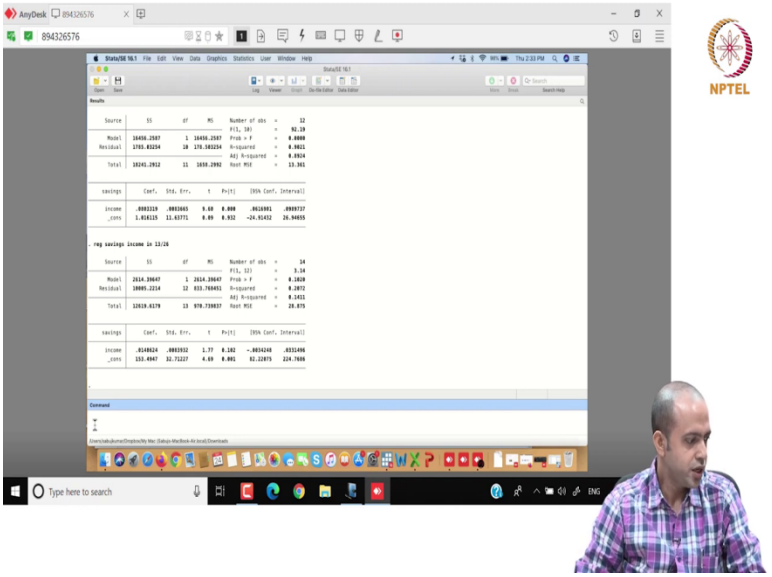
What is the restriction? We are assuming that there is no structural change in terms of the intercept as well as the slope coefficients of the savings income relationship. So, that means, when you have data on twenty six years, so you have to divide this sample into two periods. We should actually have two sub samples one for the period 1970 to 81 and another for the period 1982 to 1995. But in, you when you are using Stata actually you do not have to divide the sample into two sub samples because that is very very inconvenient and time consuming. Rather we will specify a simple command in Stata. So, that stata will run the regression, first considering the time period 1970 to 1981 that means status will first consider only 12 observations.

(Refer Slide Time: 14:49)









How to do that? The command is reg savings then income and this is you have to remember n1 by 1 by 12. So, that means by specifying 1 by 12. I am asking stata to run the regression using first 12 observations. Now if you put enter this is the regression, this is regression. So, from this regression you can easily look at the result and see the coefficient of income is 0 point 0.8. So, that means, pre - peacetime recession in the US economy, the marginal propensity to save was point 0.8.

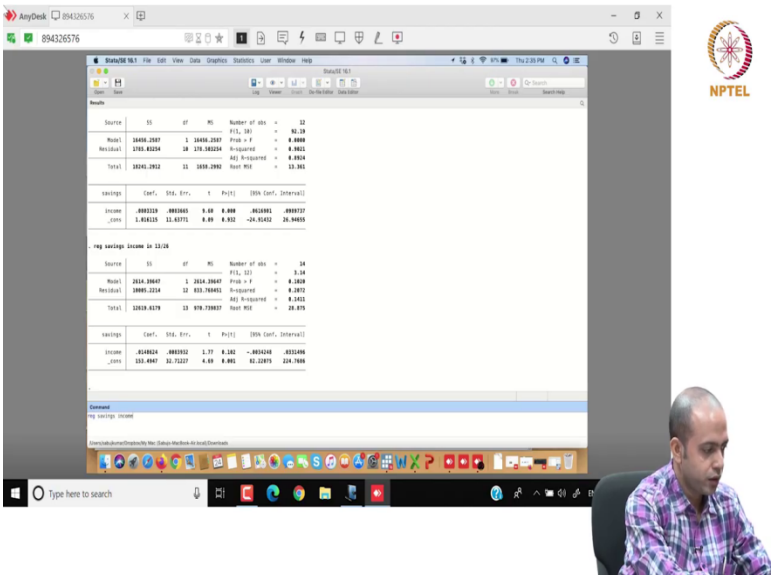
That is the interpretation. For unit change in income on an average savings increased by 0 point 0 8 unit; so, that is the MPS in the pre-recession period. Now what we will do, we will run another

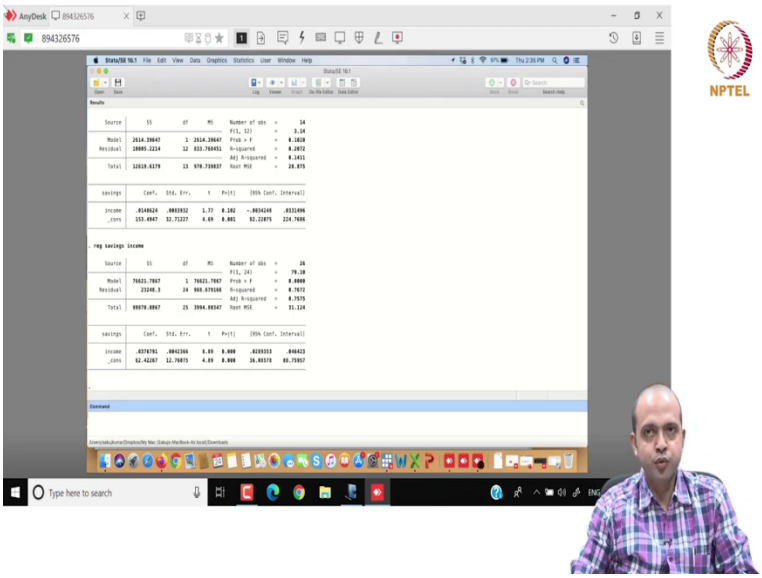
regression for the post-recession period which is 1982 to 1995 that means, we assume the structural break happened in the year 1982. So, similarly I will ask stata to run another regression taking the observation from 13 to 26, and look at what happened here? The marginal propensity to save which is point 0.8 in the pre-recession period, now it has gone down to 0 point 0.1 unit.

So, that means, just by looking at the magnitude of the coefficient what we can understand is that there is a drastic change in the individual saving behavior post-recession period. Similarly, there is a change in intercept also. In the pre-recession period it was 1 point 0 1 and in the post-recession it is 153 point 49. So, now, what we have to do actually? We have to conduct that F stat, F test or Chow test it is called Chow test because it was first introduced by the famous econometrician and Chow.

We have to conduct that test to verify whether the numerical difference between 0 point 0 8 and 0 point 0 1 is statistically significant or not. So, what we have to do? We have to take these two values that means, RSS from the two models and RSS from the combined model and then divide them, the numerator and the denominator by their respective degrees of freedom to arrive at the F statistic. So, what I will do? I will now go to so, here you have to remember two things that RSS, now another regression if I run for the entire period because that is the restricted model.

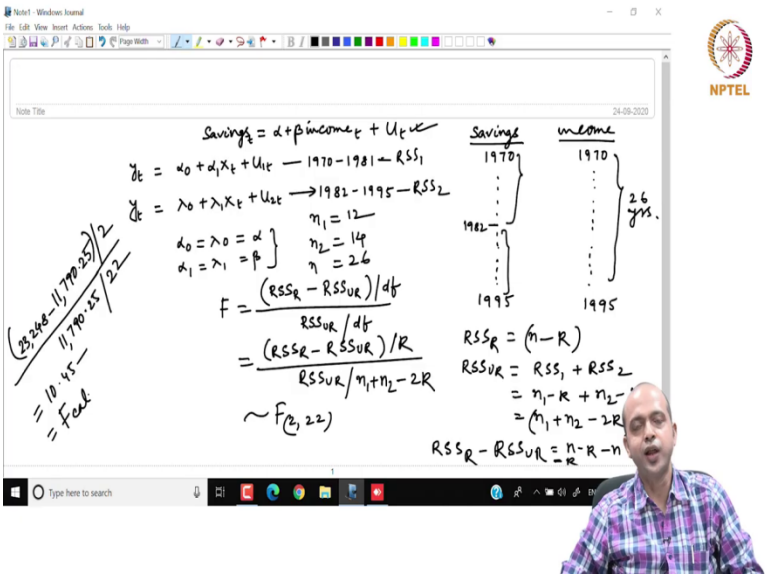
(Refer Slide Time: 18:33)

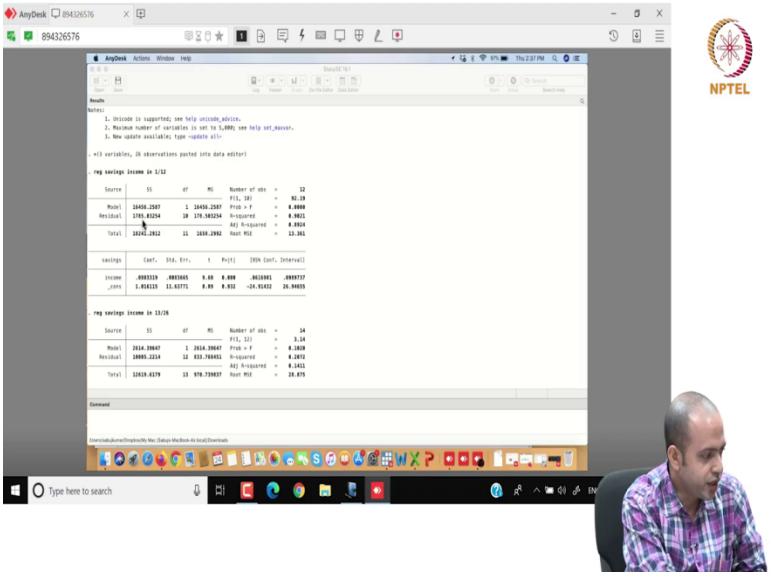




So, without putting any this thing reg savings income for the entire period and then you will get the marginal propensity to save for the entire period is 0 point 0 3. So, you have to carefully note it down the three RSS. So, what do you have to note it down the three RSS. So, from the combined one from the unrestricted model, the RSS is how much? Twenty three thousand two hundred and forty eight, twenty three thousand two hundred and forty eight that is the RSS from the restricted model.

(Refer Slide Time: 19:34)





So, we have to go to that formula twenty three thousand two hundred and forty eight, twenty three thousand. So, here the restricted RSS equals to twenty three thousand. So, I will use this one twenty three thousand two hundred and forty eight minus RSS UR. So, RSS UR is basically RSS1 plus RSS2. So, if you go there. So, RSS1 is basically, what is the RSS1? RSS1 is basically your this one. RSS 1 is you go to your model look at here this is thousand seven hundred and eighty five and here it is ten thousand five. So ten thousand five plus thousand seven hundred and eighty five, so it will come around eleven thousand seven hundred and ninety. Eleven thousand seven hundred and ninety, roughly eleven thousand seven hundred and ninety point something it will come.

So, what is the point? Point if you take, if you are interested in point also then it will become 0 point 0.3 plus here it is 0 point 2 2 so, it will come around 0 point 2 5. So, point 2 5 you can get. So, point 2 5 and this you have to divide by k where k is 2 and what is the numerator? Numerator is again RSS UR. RSS UR which is again your eleven thousand seven hundred and ninety, eleven thousand seven hundred and ninety, sorry eleven thousand seven hundred and ninety point 2 5 divided by 22. Divided by 22 and if you calculate these then that would become some 10 point something will come 10 point some value will come 4, 5 or something will come. You have to calculate this using these and these I would say that F calculated.