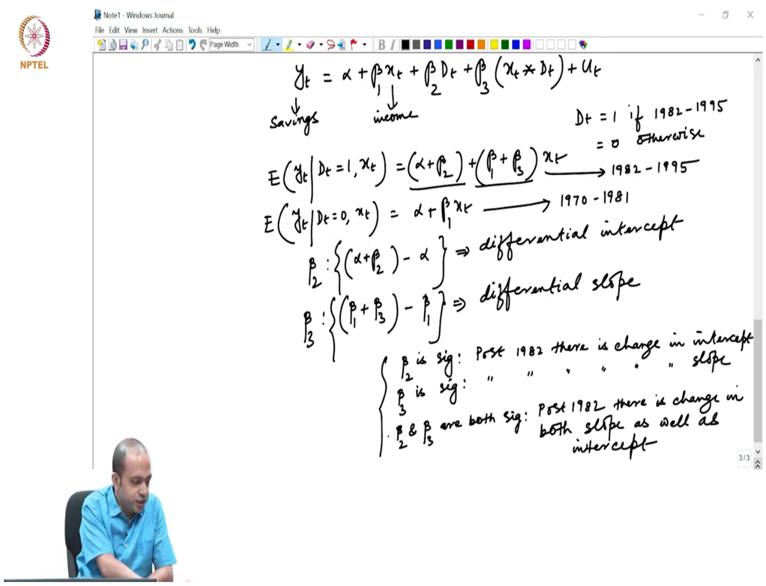
Statistical Analysis of Dummy Variable Models and Testing for Seasonal Fluctuations
Part-4
Professor Sabuj Kumar Mandal
Department of Humanities and Social Sciences
Lecture 41
Indian Institute of Technology, Madras

(Refer Slide Time: 0:15)



So, from this what you can understand that, what you can understand that beta 2 that how will you derive the, now you have to, what you have to do? Again, you have to derive the interpretation beta 2. How will you derive beta 2? Beta 2 is nothing but alpha plus beta 2 minus alpha. And what is alpha plus beta 2? That is the intercept of the post-recession period, an alpha is the intercept of the pre-recession period. So that means beta 2 indicates differential intercept.

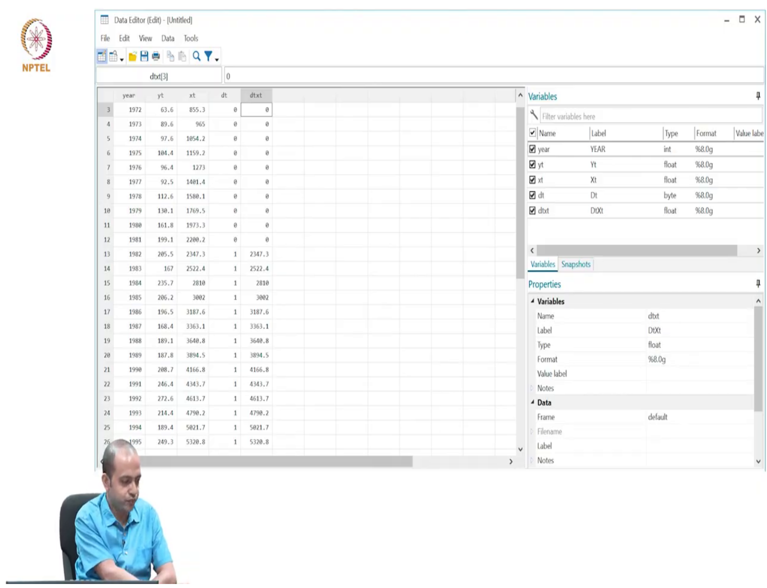
Since we have one more covariate xt present in this model, we cannot say that alpha is basically the value of the dependent variable in the base period. We can only tell that alpha is basically intercept, because this is an ANCOVA model. Apart from the dummy we have xt also present in the model. So that is why beta 2 also is defined as differential intercept.

And how do you define beta 3? Beta 3 is defined as beta 1 plus beta 3 minus beta 1. And if you define in this way what is beta 1 plus beta 3? Look at here beta 1 plus beta 3 is the slope of the saving income relationship post 1982 period, that means post-recession beta 1 is the slope for the pre-recession period. So that means beta 3 indicates differential slope, differential slope. So, you now understood how how the coefficients are interpreted when there is a dummy variable involved. So, you have to basically derive it.

So, depending on the significance of beta 2 and beta 3 we will conclude whether post-recession period there is a change in the intercept or change in the slope or both. So, when beta 2 is significant that means we will say that post 1982 there is change in intercept. When beta 3 is significant we will say that post 1982 there is change in slope.

When beta 2 and beta 3 both are significant, so post 1982 there is change in both slope as well as intercept. This is how you have to fix the model and then you have to estimate and this is how you can interpret. So, this would be the post estimation interpretation beta 2, beta 3 and significance of beta 2 and beta 3. So now what we will do? We will just estimate the model, I will show you the data set.

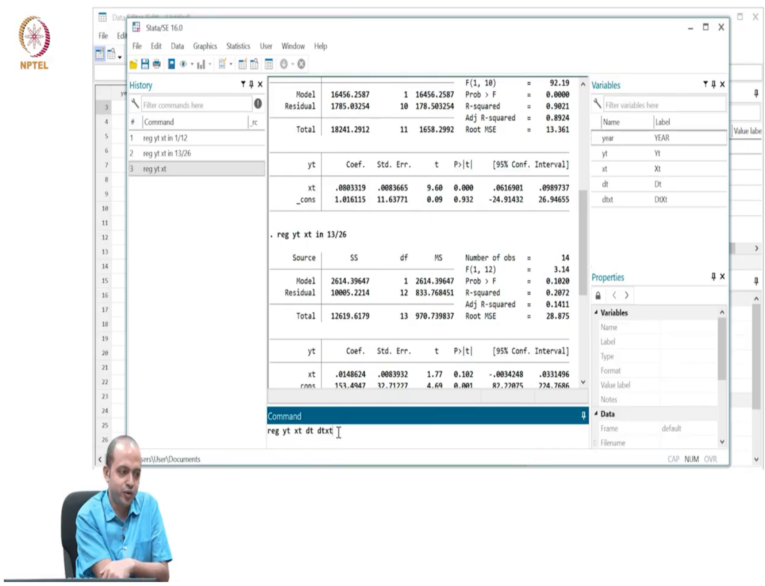
(Refer Slide Time: 5:30)



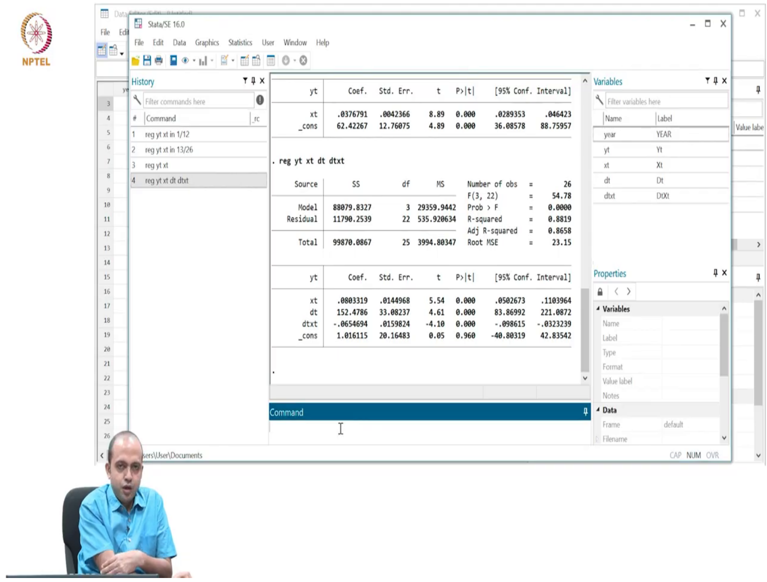
Look at the data set first see this is 1970 to 1995 data yt indicates saving, xt indicates income and this dt is basically the coefficient of the dummy and Dt xt is the interaction between the dummy and the quantitative covariate xt. So, we have assigned 0 value see, 0 starts from 1970 and it goes up to 1981. So, 1970 to 1981 it takes 0 value as we have defined earlier. This is the pre-recession period.

And starting from 1982 to 1995 in all these years this dummy variable takes the value 1 indicating post-recession starting from 1982 because that was the structural break point we have assumed. And this particular cell is simply created by the introduction that means the multiplication. So, this is how you have to create the variable.

(Refer Slide Time: 6:52)

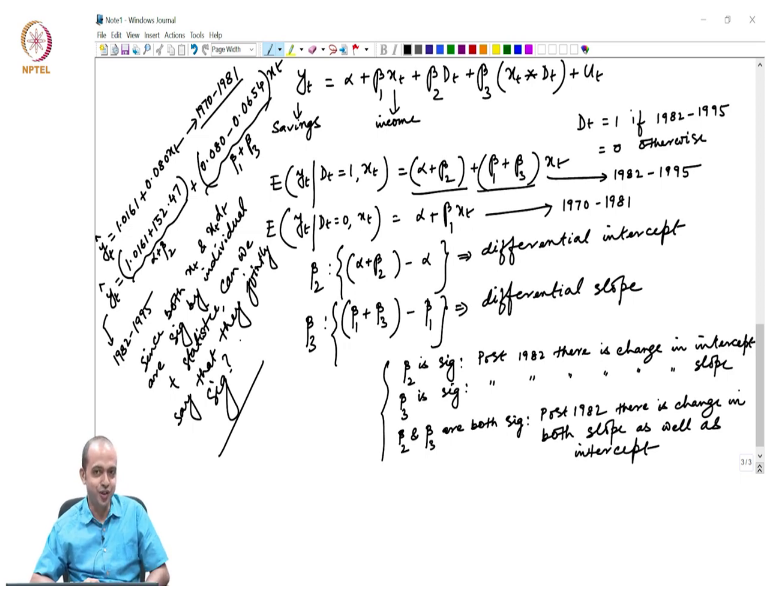


(Refer Slide Time: 7:18)



And now you go to your Stata and then what you have to do? You have to run this regression reg yt and then xt dt and dt xt. These are the three variables xt is for income, dt is the dummy variable, dt xt is actually your interaction term. So, we have estimated the model.

(Refer Slide Time: 8:39)



Now if you look at how we have defined, how we have explained the interpretation of the dummy and the interaction from this model there are several terms. I will just go back to the previous discussion see if dt equals to 0 then alpha plus beta 1 xt is your equation for the 1970 to

1981. So that means from this result what is your alpha? Alpha is this 1.0161. And what is your xt? xt is this.

So that means pre 1982 period what you have to do? That yt hat equals to what is the intercept 1.0161 plus what is the value of the slope? Slope coefficient is 0.080. So that means this would become 0.080 xt. So, this is your equation for the 1970 to 1981 pre-recession period. And what is the equation for the post-recession period? That would become yt hat equals to post-recession period 1992 to 1975 alpha plus beta 2 plus beta 1 plus beta 2 beta 3 xt.

So now from the result alpha is 1.01. And what is your differential intercept? Differential intercept is 152. So that means this would become 1.0161 plus 152 point and the value is 152.47. So, this is now your alpha plus beta 2. This is actually your alpha plus beta 2 plus, what is your beta 1 plus beta 3?

So that means you already know if you look at the beta 1 is actually is 0.080 minus. What is the differential slope? Differential slope here is 0.0654. So, 0.0654 minus 0654. So, this is actually your beta 1 plus beta 3 and xt. So, this is how you will get the regression equation, saving income relationship for the pre-recession period. This is for the post-recession period that means 1982 to 1995, 1995.

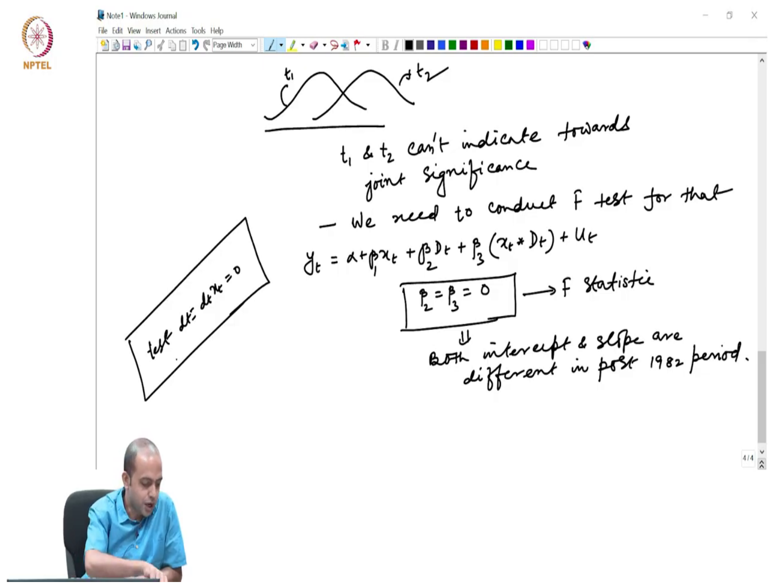
Now look at the level of significance. How do you know the differential slope? dt xt the coefficient of dt xt is basically differential slope that is minus 0654 is that significant? Look at the p value is 0.000. If you multiply that it is again 0.000. So that means differential slope is highly significant.

So post-recession period the differential slope is highly significant that means post-recession period slope is significantly different from the pre-recession period. What about differential intercept? Value is 152.47 look at the level of significance that is also 0.000. If you multiply that by 100 that is again highly significant. So that means post-recession period both are significant, both are significant.

Now one question that I would like to raise here. Since, since both xt and xt dt, xt dt are significant by individual, individual t statistic, t statistic can we say that they are jointly significant. So that means my question is can we use this individual t statistic to say that they are jointly significant without conducting any further test.

Individual test statistic shows that post-recession period slope is significantly different from the pre-recession period. Intercept is different from significantly different from pre-recession period. That means can we say that slope as well intercept they are jointly significantly different from the pre pre-recession period slope and intercept, can you say that?

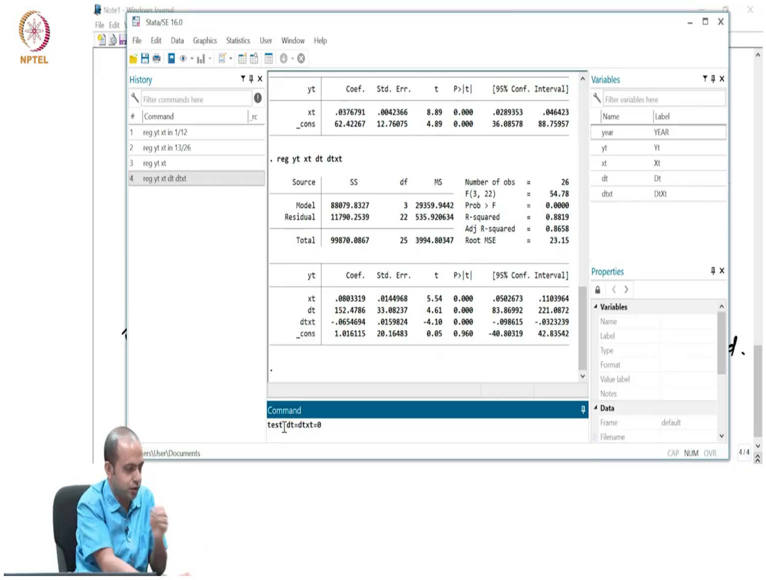
(Refer Slide Time: 15:40)

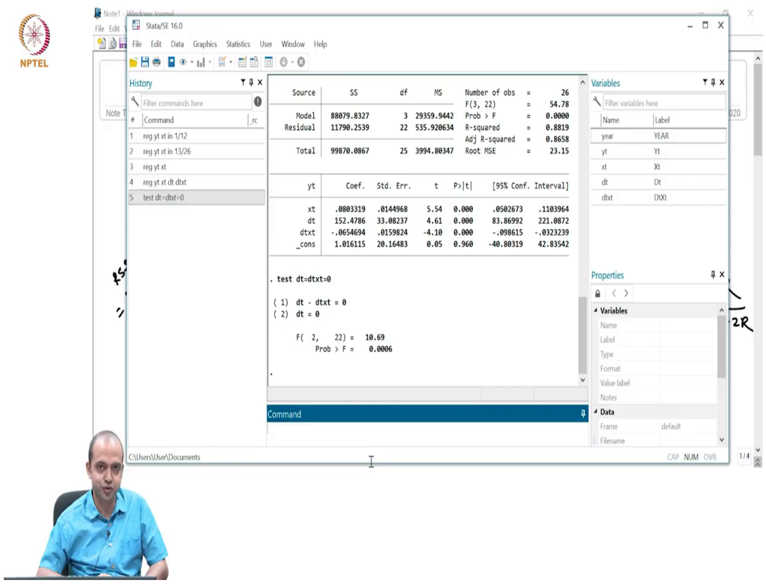


If you recall, we discussed previously that there is no one to one mapping between t and F when you have multiple linear regression model. Why this is so? I quickly say that you are collecting the data from the same sample so that is why if this is your t1 distribution and this is your t2 distribution they are actually not independent, there is some overlap. And because of this individual significance does not indicate the joints significance. So, t statistic t1 and t2 cannot, cannot indicate towards joint significance.

Then what we need? we need to conduct F test for that because F test basically shows the joint significance of the model. So that means in terms of this model when we are saying yt equals to alpha plus beta 1 xt plus beta 2 dt plus beta 3 xt multiplied by dt plus ut in this model what we are basically interested to check beta 2 equals to beta 3 equals to 0 or not. This is basically we are going to test using F statistic. And this would imply that both intercept and slope are different in post-1982 period. This we have test.

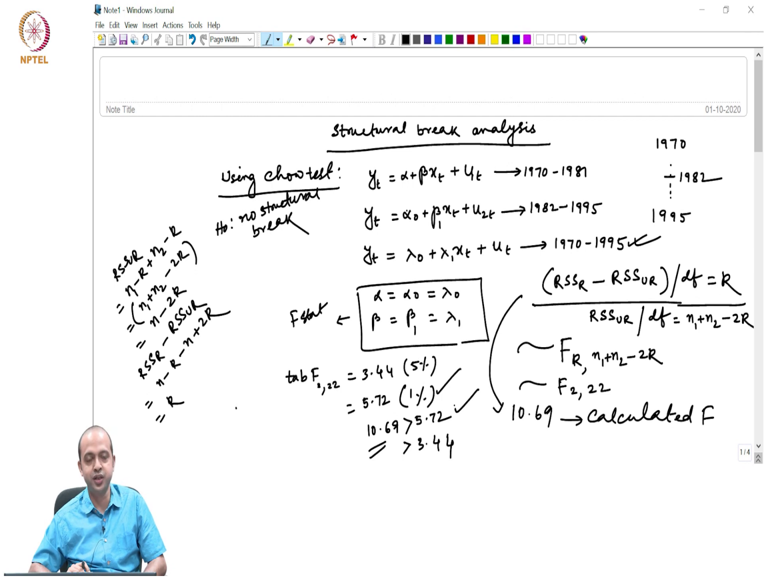
(Refer Slide Time: 19:26)





Now if you go back and as we said there is a specific command which says that you have to put the test command test you have to put dt and then space dt xt equals to 0. This is the test command in Stata. So, after estimating the model immediately after estimating the model, This is the full model. And now I want to check whether the differential intercept and differential slope they are jointly significant or not. And for that test dt equals to dt xt equals to 0 or not. F statistic is 10.69 which is exactly matching with the F statistic that we have calculated in the context of Chow test.

(Refer Slide Time: 20:25)

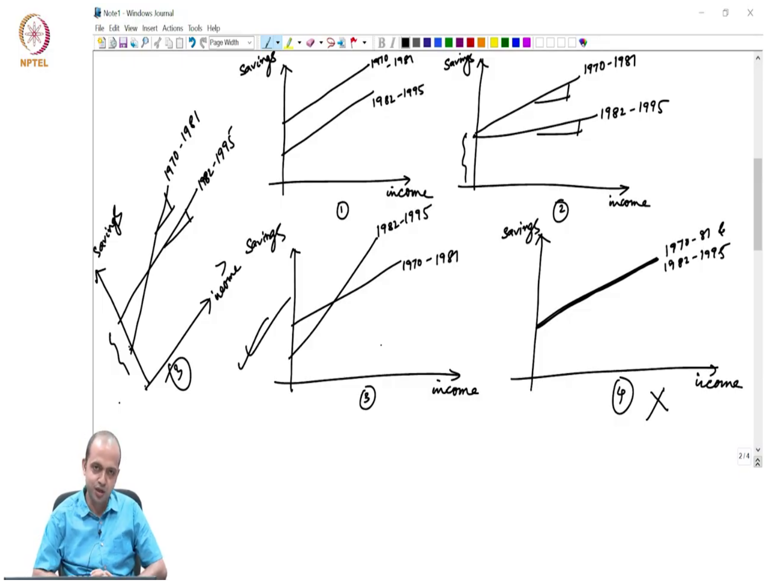


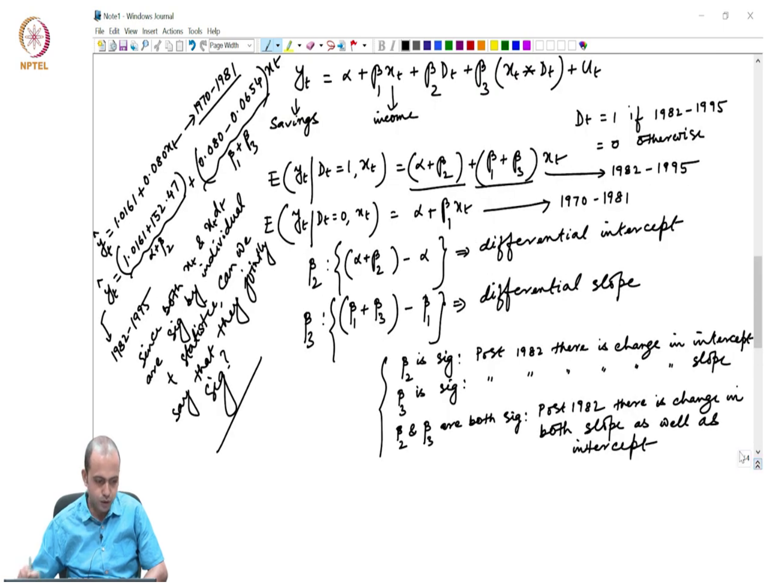
If you go back you can easily verify, you can easily verify that you get this 10.69 was the value, same value Stata is also reporting. Stata is also reporting the same value using the test command. So, this is a very very powerful command to check the equality between two regression coefficients and here it is not only checking the equality but also, they are equals towards 0. That is the test command we are applying, that is the test command.

So, F is following 2 and 22 degrees of freedom as we have discussed in the context of Chow. And the value is also matching and the probability greater than p value is 0.0006 which is highly significant. And in that context also we showed that it is significant at even at 1 percent level 10.69 is greater than 5.72, it is significant at 1 percent level. So, whatever we have constructed manually same thing we are doing using the dummy variable.

But the biggest advantage instead of running three different regression and then getting RSSr RSSur and then constructing F you can do it by applying a simple test command. And also, when you run one single equation using a dummy that is easier to check the structural break or stability analysis compared to the Chow wherein we were running three regressions.

(Refer Slide Time: 22:23)





So that is how the dummy variable model overcomes the limitations of Chow test. Now the dummy variable is saying that yes out of this three actually this is the case in our present context where both slope as well intercept are different, both slope as well as intercept are different.

So, we have to see carefully the diagram whether the slope is more or less depending on the sign here I have just arbitrarily drawn the diagram. So, depending on if you go to the previous one the value is actually, so, look at. So post-recession period your intercept actually is higher but slope is lower, intercept is higher but slope is lower. So, in this case intercept should be higher but slope is lower.

So, we have drawn it in a differently, so this three actually should be drawn in this way. This is income, this is saving. So, your intercept is higher but the slope is lower. So that means what you have to do? If this is your 1970 to 1981 your post-recession period should be this, intercept is higher. This is 1982 to 1995, intercept higher but the slope is lower, here the slope is small.

So that means marginal propensity to save actually came down in the post-recession period which is quite understandable, post-recession period people are not saving that much for every additional unit of income like they were doing earlier. So, this is actually the accurate case 3 instead of this. This I was drawing randomly. So, this is the case.

So, with this we have now learned how to do the structural break analysis using the Chow and using the dummy. But we do not conduct Chow test for conducting structural break analysis because that is time consuming as well as it has limitation. But I specifically and intentionally showed you both. So that you can understand the advantages of dummy variable over the Chow test in conducting the structural break analysis and you appreciate more about the dummy variable model.

That was the objective I had in my mind, that is the reason I conducted and I demonstrated both the tests Chow as well as dummy. So, with this we are closing our discussion today and in our next day we will see one more application of dummy variable model. We will see another application. And our discussion today is closed here onwards. Thank you very much.