Statistical Analysis of Dummy Variable Models and Testing for Seasonal Fluctuations Part-3

Professor Sabuj Kumar Mandal
Department of Humanities and Social Sciences

Lecture 40

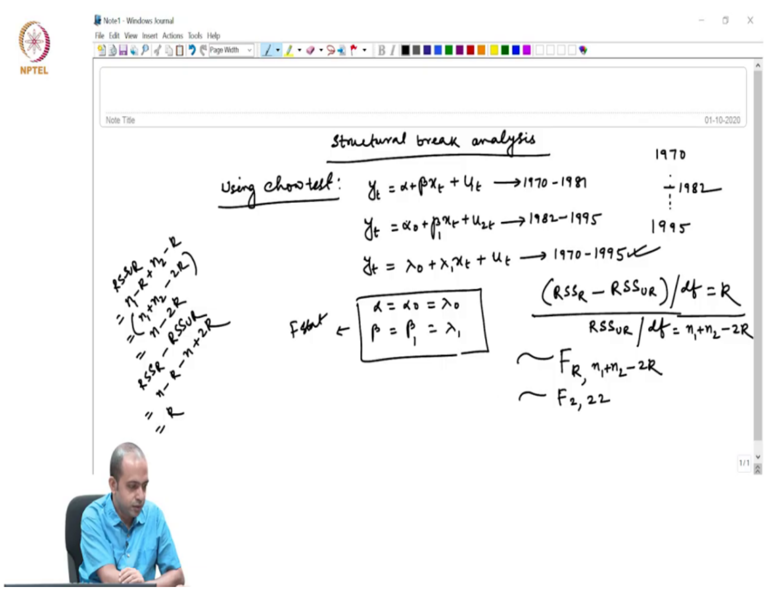
Indian Institute of Technology, Madras

Welcome to the class of discussion of Dummy Variable Model once again. So today what we will do? We will take another dummy variable model that actually we have discussed earlier in the context of Chow test. Do you remember that we were discussing structural break analysis in time series data particularly the savings income example in the context of U.S.? The same example, same data set will take here and what we will do?

We will do the same structural break analysis using dummy variable model. Why this is so? Because dummy variable model will overcome the short comings of Chow test. And what were the short comings of Chow test? Because Chow test cannot tell us if there is any structural break and what is the source of that.

So, what we will do? We will quickly recap the structural break using Chow test and then we will come back to the dummy variable model and we will see how to estimate the same thing using a dummy.

(Refer Slide Time: 1:36)



So, this is structural break analysis once again. So, what we are doing? So, we had some 1970 to 1995 data on saving and income and then in between we assume that in 1982 there was a structural break because that was the time U.S. economy was suffering from the peacetime recession.

So, what we did actually? We first ran three regressions one for the time period 1970 to 1981. So that was yt equals to alpha plus beta xt plus u1t, And then another one for yt equals to let us say some alpha 0 plus beta 1 xt plus u2t which is for 1982 to 1995. So, these are the unrestricted models and the restricted model was yt equals to some lambda 0 plus lambda1 xt plus ut, this is for the entire period 1970 to 1995.

So, one for the pre pre-recession and another for post-recession and the third regression is for the entire period. And we were checking the stability analysis whether the parameters alpha equals to alpha 0 equals to lambda 0 or beta equals to beta 1 equals to lambda 1, that is what we were testing.

Now these notations might be different earlier might you some other notations. So, please be careful about the notation, do not get misguided. So, I might be using different notation at different point of time. So basically, the idea is same, what we are doing, that we are testing alpha equals to alpha 0 plus equals to lambda 0 and then beta equals to beta 1 equals to lambda 1, this is the restriction what we are testing.

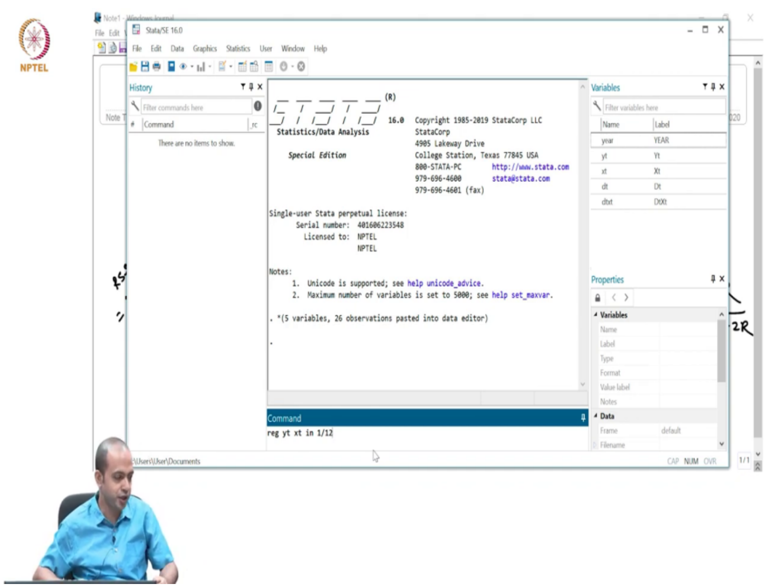
And we are using F statistic. How that is defined? What is the restricted model? This one is the restricted model. When I am specifying one single equation for the entire period, that is our restricted model. And then RSSur divided by its degrees of freedom.

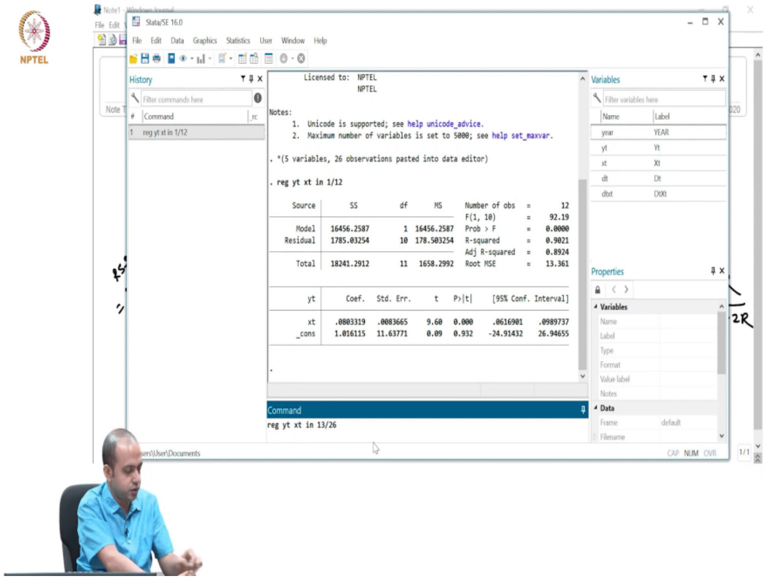
And if you recall what is a degrees of freedom for RSSr minus RSSur? Because that is RSSr what is the degrees of freedom? n minus k. And RSSur how you are getting RSSur? RSSur was n1 minus k plus n2 minus k. So it was n1 plus n2 minus 2k. And this is n minus k. So this is basically n1 plus n2 equals to n. So I can write n minus 2k.

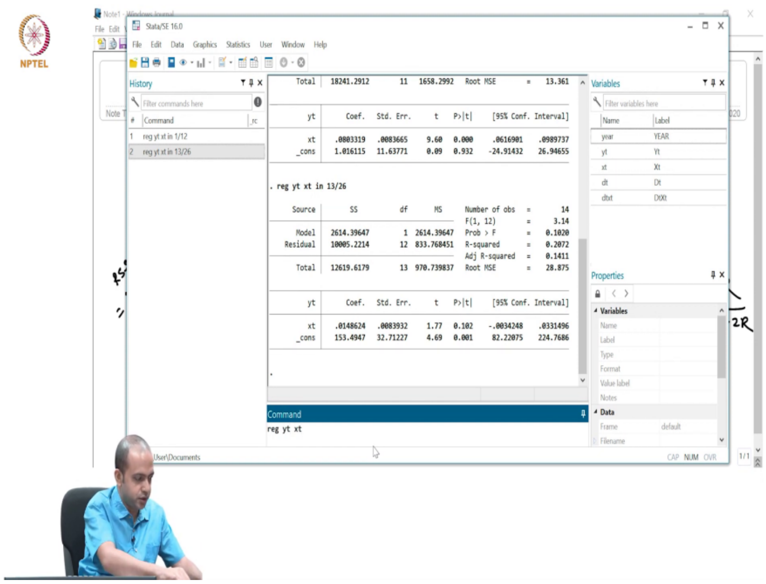
So RSSr minus RSSur equals to n minus k minus n plus 2k equals to k. So, degrees of freedom for the numerator is equal to k. And what is the degrees of freedom for this RSSur? Which is n1 plus n2 minus 2k, n1 plus n2 minus 2k. Because there are two unrestricted models. So now what you can do? This follows the F statistic with degrees of freedom k for the numerator and n1 plus n2 minus 2k for the denominator.

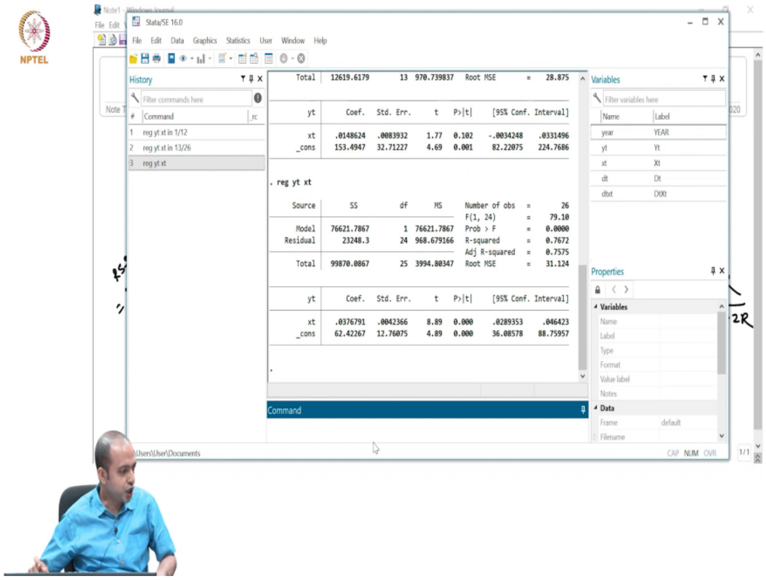
So that means here it is 2 and 22, n1 plus n2 equals to 26. There are total 26 periods of time and 2k means 2 into 2, 4. So 26 minus 4, 22. This follows F distribution with 2 and 22 degrees of freedom. So, this is the Chow test what we have discussed already. So we will quickly estimate the two models and we will just go to the data set once again.

(Refer Slide Time: 9:10)



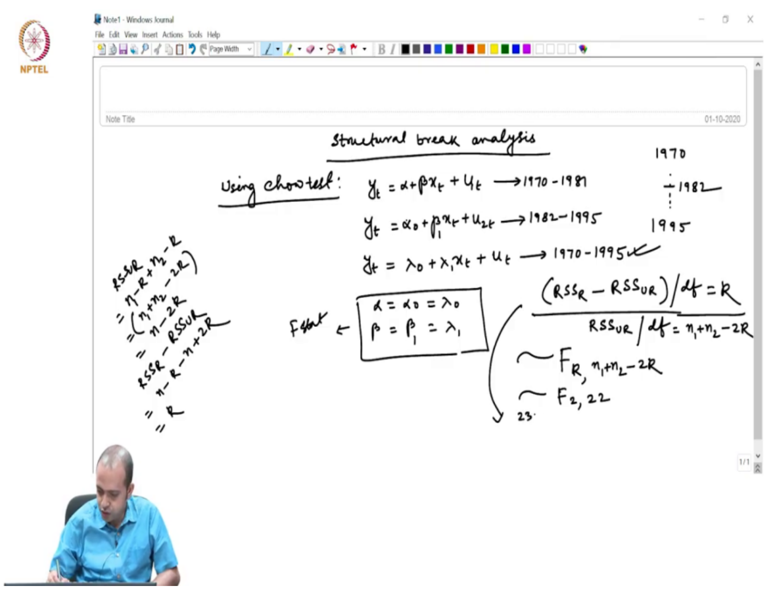

NPTEL
Stata/SE 16.0
File Edit Data Graphics Statistics User Window Help
History
Variables
Licensed to: NPTEL
NPTEL
Notes:
1. Unicode is supported; see help unicode_advice.
2. Maximum number of variables is set to 5000; see help set_maxvar.
. *(5 variables, 26 observations pasted into data editor)
. reg yt xt in 1/12
Source SS df MS Number of obs = 12
F(1, 10) = 92.19
Model 16456.2587 1 16456.2587 Prob > F = 0.0000
Residual 1785.03254 10 178.503254 R-squared = 0.9021
Adj R-squared = 0.8924
Total 18241.2912 11 1658.2992 Root MSE = 13.361
yt Coef. Std. Err. t P>|t| [95% Conf. Interval]
xt .0803319 .0083665 9.60 0.000 .0616901 .0989737
_cons 1.016115 11.63771 0.09 0.932 -24.91432 26.94655
Command
reg yt xt in 13/26
Properties


NPTEL
Stata/SE 16.0
File Edit Data Graphics Statistics User Window Help
History
Variables
Total 18241.2912 11 1658.2992 Root MSE = 13.361
yt Coef. Std. Err. t P>|t| [95% Conf. Interval]
xt .0803319 .0083665 9.60 0.000 .0616901 .0989737
_cons 1.016115 11.63771 0.09 0.932 -24.91432 26.94655
. reg yt xt in 13/26
Source SS df MS Number of obs = 14
F(1, 12) = 3.14
Model 2614.39647 1 2614.39647 Prob > F = 0.1020
Residual 10005.2214 12 833.768451 R-squared = 0.2072
Adj R-squared = 0.1411
Total 12619.6179 13 970.739837 Root MSE = 28.875
yt Coef. Std. Err. t P>|t| [95% Conf. Interval]
xt .0148624 .0083932 1.77 0.102 -.0034248 .0331496
_cons 153.4947 32.71227 4.69 0.001 82.22075 224.7686
Command
reg yt xt
Properties

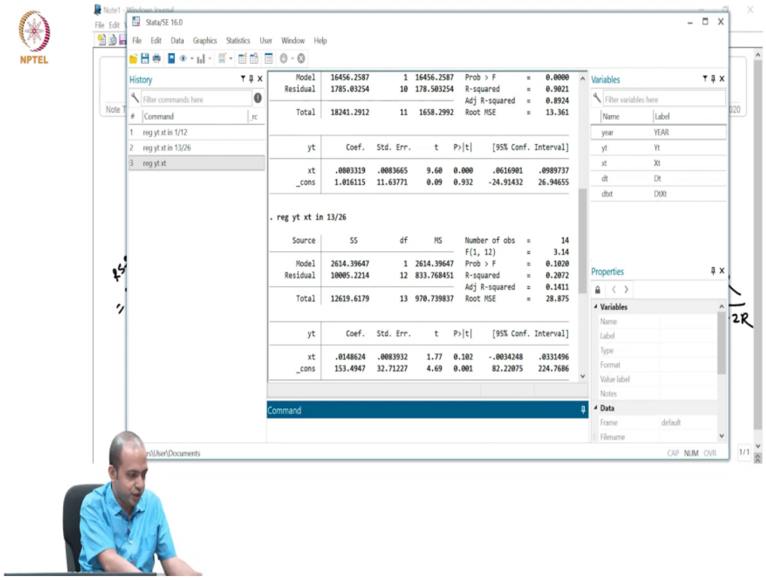


Now what we will do? We will estimate reg yt and then xt and we will first estimate for the first 12 observations 1970 to 1981, there are total 12 observations. And the second regression is reg yt then xt in 13 by 26, put enter. And then another for the entire period yt xt. This is what, so now you have to collect the RSS from all these models and calculate the F statistic. So, from the restricted model RSS is 23,248.3.

(Refer Slide Time: 10:36)

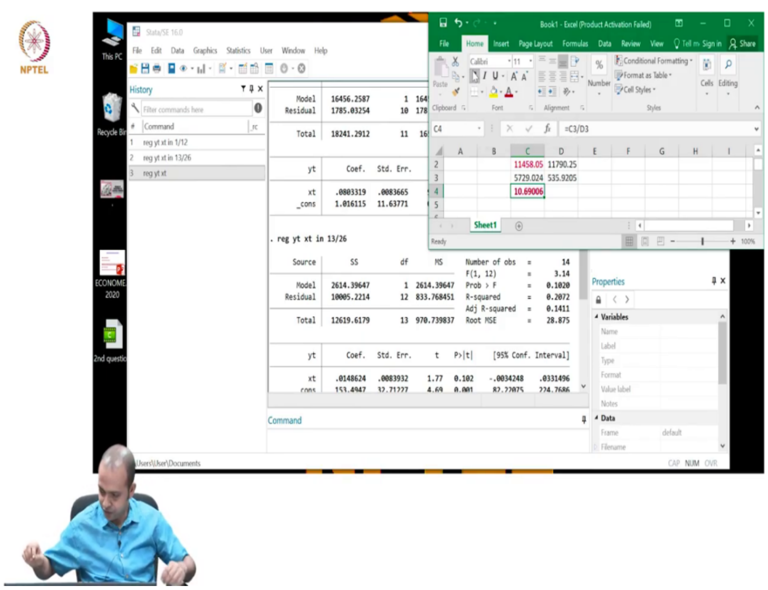


(Refer Slide Time: 11:01)



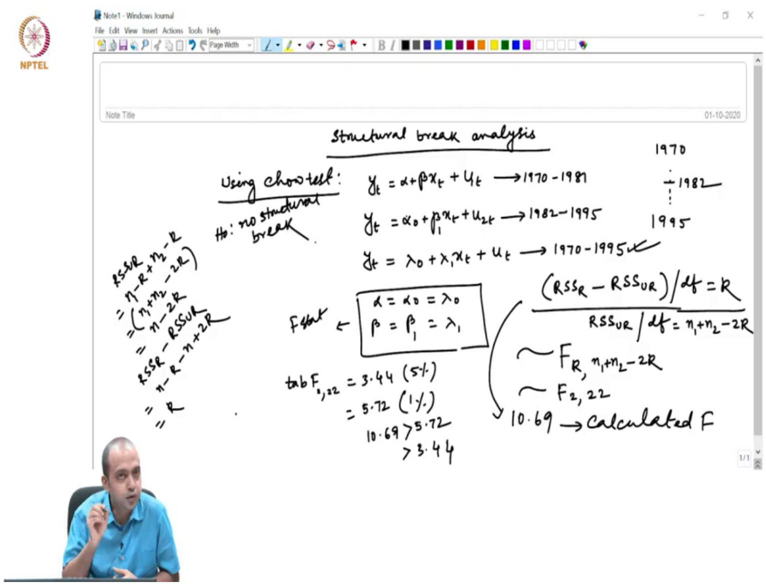
So, if you look at this formula RSSr is 20. So, from this what I can say that this is 23,248.3 minus, minus 23,248.3 minus what you have to do you have to take these two RSS, these two RSS 10,005. first one is this 1,785, 1,785. So basically, so I will calculate these in an excel sheet that would be easier, that would be easier for you to understand. So what I will do? I will just, I will just close this, close this and I will open an excel sheet.

(Refer Slide Time: 11:35)



This equals to 23,248 I think 23,248.3 minus minus you have 10,005, 10,005.2214 plus, plus the other one, other one is 1,785, 1,785.03. This is your RSSr minus RSSur. And this you have to divide by RSSur which is unrestricted which is again equals to 1785.03 plus 10,005, 10,005.2214. So, this should become, so the numerator this you have, this this divided by 2 and is equal to this and here it is RSSur you have to divide by 22 press enter. Now this divided by this some 10.69 that is your F value, 10.69.

(Refer Slide Time: 15:19)



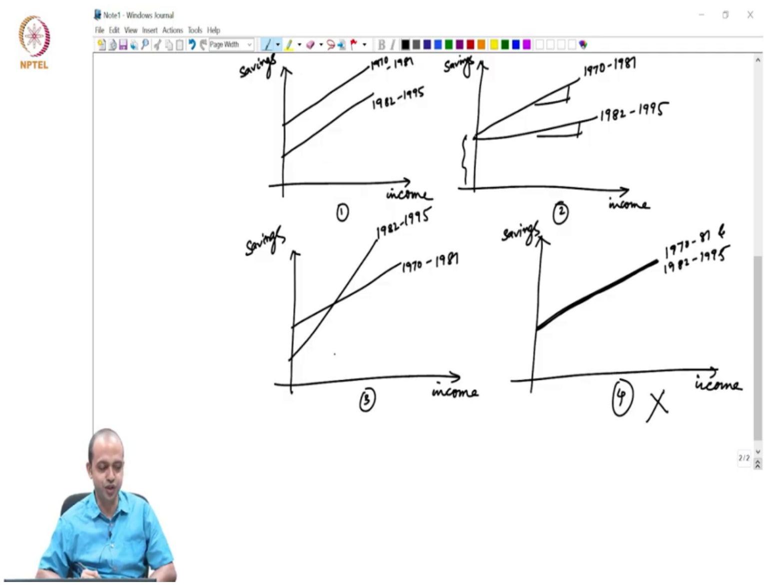
So ultimately your F value is, if you put all this, I am not doing all this again so this should become after putting the numerical values for RSSr, RSSur divided by k and again entire thing divided by RSSur divided by degrees of freedom what you will get, the value is 10.69. So, this is your calculated F.

Now if you compare this calculated F with the tabulated F at 5 percent, 10 percent and 1 percent level of significance. Then what we will get? If you compare the tabulated value from the F table what you will get is at at 2 and 22 degrees of freedom the tabulated F at 5 percent is 3.44 and 5.72 at 1 percent.

So that means 10.69 is greater than 5.72 as well as it is greater than 3.44. So, your calculated value is actually greater than the tabulated value. And then what we will do? If the calculated value is greater than the tabulated value we have to reject our null hypothesis. And what is our null hypothesis? There is no structural break.

And if you reject the null hypothesis, so null hypothesis is is this actually, the restriction what we have imposed that is actually the, that is actually our null hypothesis because if this conditions are satisfied that means there is no structural break. Pre and post finance recession period can be modeled by one single equation. So, both the intercept and the slope coefficients are same. So, no structural break is actually rejected.

(Refer Slide Time: 18:34)



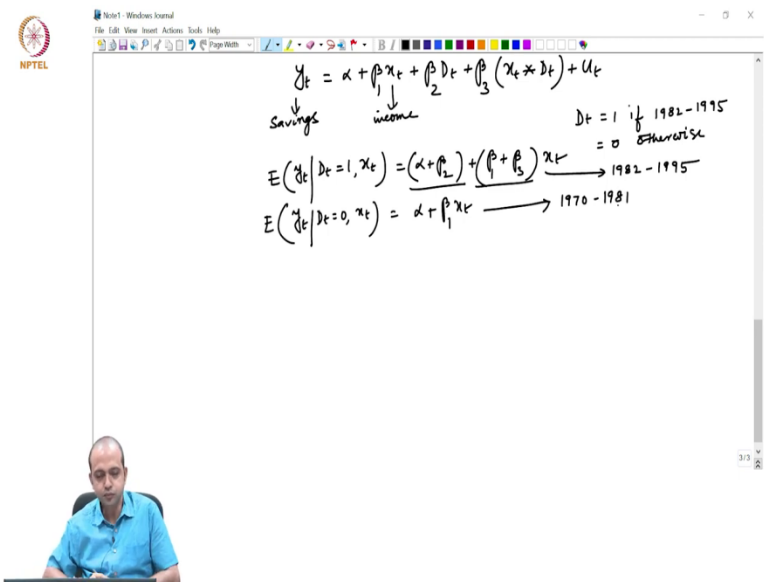
And what we said when this is rejected that means Chow test is indicating three situations. So here it is income and here it is saving. So, this is let us say 1970 to 1981 and this is 1982 to 1995, this is case 1. So, in this case post-recession period, the slope is same but the intercept is different.

In this case income-saving, so this is 1970 to 1981 this is 1982 to 1995. So intercept is same but the slope is different which is case 2. And this is case 3, income and saving, so it is 1970 to 1981, this is 1982 to 1995. So post-recession period both slope and intercept are different, this is case 3.

And this is actually case 4 saving and this is income. So this basically indicate 1970 to 81 also 1982 to 1995 for both the period neither the slope nor the intercept is different. So that means this is called concurrent relationship. So, both the time periods can be specified by one single equation.

So, Chow test can tell you that this case 4 is actually rejected. So, we are eliminating case 4 there are differences in the two time periods saving income relationship but we do not know which among this three are actually the relevant for this particular data set. So that means the difference in saving income relationship is coming due to intercept or slope or both that we do not know. And here comes the dummy variable model handy. Now we will see how to do the same analysis using a dummy variable model.

(Refer Slide Time: 22:22)



So, let us first try to construct the model. Let us say that this is our yt equals to alpha plus beta 1 xt plus beta 2 Dt plus beta 3 xt Dt plus ut this is our model. And how we have defined Dt? Dt equals to 1 if the time period is 1982 to 1995 and 0 otherwise.

Because when we say that the crisis or recession so presence of recessions in when the impact of recession is present that we are denoting by 1. Because in the dummy variable generally we say that presence of the attribute is 1 and absence is 0. So post-recession period that is 1 and 0 otherwise. So, this is how we have defined the dummy variable.

So now again here what is yt? yt is your actually savings and xt is actually income and if you take expectation, expectation of yt given Dt equals to 1 and xt equals to alpha plus beta 2 and then plus you have beta 1 xt here also you will get beta 3 xt. So, I am adding the two coefficients and then I am writing xt. So that means for the period 1982 to 1985 this is my saving income relationship with slope intercept equals to this and slope equals to this, this is for the period 1982 to 1995.

Post-recession period saving income relationship is defined as expectation of yt given Dt equals to 1 equals to alpha plus beta 2 plus beta 1 plus beta 3 xt. So that means this is the intercept and this is the slope. And pre-recession period how it is defined? Expectation of yt given Dt equals to 0 xt equals to alpha. So, this will vanish alpha plus beta 1 xt, beta 1 xt. So, this is for the period 1970 to 1981.