Introduction to Econometrics

Professor Sabuj Kumar Mandal
Department of Humanities and Social Sciences
Indian Institute of Technology, Madras

Lecture 54

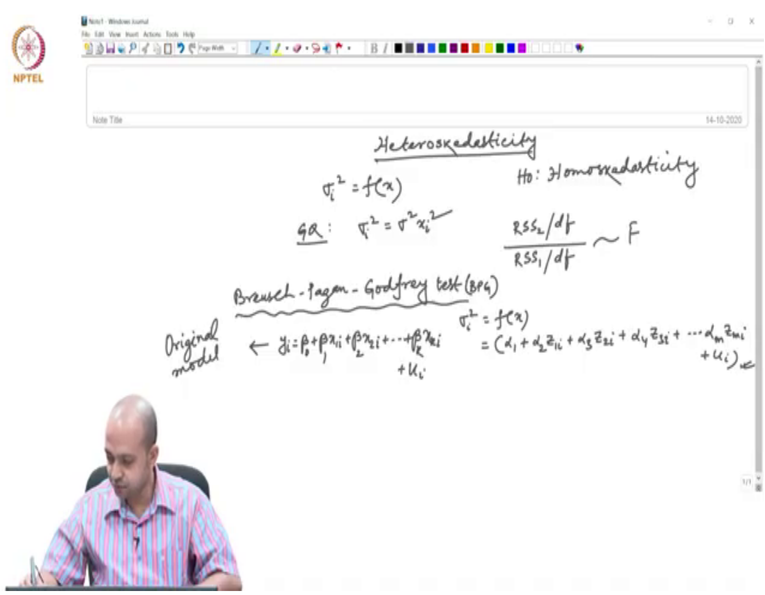
Relaxing the assumptions of CLRM-Autocorrelation and Heteroscedasticity Part - 5

So, welcome to our discussion of heteroscedasticity once again and if you recall yesterday in our last class we were discussing about Goldfeld and Quandt test for detecting heteroscedasticity. And then we also discussed about some of the major limitations of this heteroscedasticity detection by Goldfeld and Quandt test and the major limitation was that in Goldfeld and Quandt test you need to arrange your explanatory variables from smaller to the bigger ones.

That means if you have only one explanatory variable while it is quite easier to arrange your explanatory variable from small to big but when you have a large number of explanatory variables and you are not sure about which particular explanatory variable is creating heteroscedasticity problem then basically you need to repeat the same procedure of Goldfeld and Quandt test for all the explanatory variables which is quite time-consuming process.

Based on that difficulty econometricians have developed another test which is called Breusch-Pagan Godfrey test or in short BPG test. Before we go ahead with the BPG test if you recall yesterday, I said that all these tests based on which particular assumption you are making about your sigma square that means error variance.

(Refer Slide Time: 2:03)



In all these models we assume that sigma square i is actually a function of your explanatory variable where ax is basically explanatory variable, this is a function of x. And these different tests are developed based on which specific functional form you are assuming for f(x).

So, in Goldfeld-Quandt test GQ test, you assume that sigma square i equals to some sigma square into xi square. So, that means sigma square i is basically an increasing function of x it is xi square so that means as x increases sigma square i increases at an increasing rate.

How we have constructed the test statistic in Goldfeld and Quandt test if you recall? That was RSS2 by its corresponding degrees of freedom and divided by RSS1 by its corresponding degrees of freedom. What is RSS2? RSS2 is basically the RSS from the model wherein you are running a regression of y on x and x is from the second sample, you have arranged the data, you have arranged the x variables according to their magnitude.

So, second sample consists of the larger values of x and the first sample consists of the smaller values of x. So, obviously because of this particular assumption sigma square i equals to sigma square into xi square. RSS2 should be quite higher than RSS1.

Now, this particular test will follow the F statistic. Now, larger the difference between RSS2 and RSS1 greater would be the value of F and greater would be the probability of rejecting your null hypothesis where null is basically homoscedasticity. So, you can reject the homoscedasticity assumption that was the procedure we discussed about in Goldfeld and Quandt test.

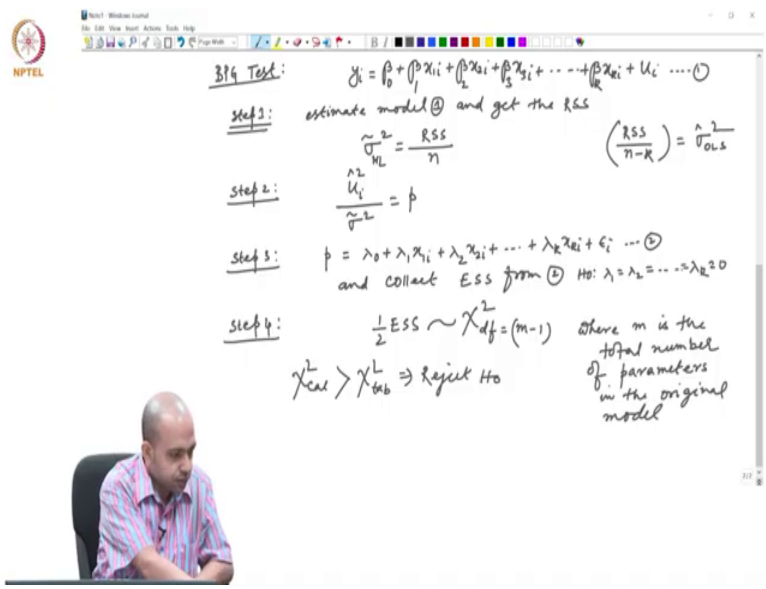
Now, today what we will do, we will discuss about Breusch-Pagan and Godfrey test, Breusch-Pagan and Godfrey, in short this is called BPG. So, here also what we assume? Sigma square i is a function of x and what is the specific functional form they have assumed? They have assumed sigma square i equals to some alpha 1 plus alpha 2 let us say z1i plus z2i, sorry, alpha 2 or alpha 3 z2i plus alpha 4 z3i plus alpha m zmi plus ui that is the specific functional form they have assumed in Breusch-Pagan and Godfrey test.

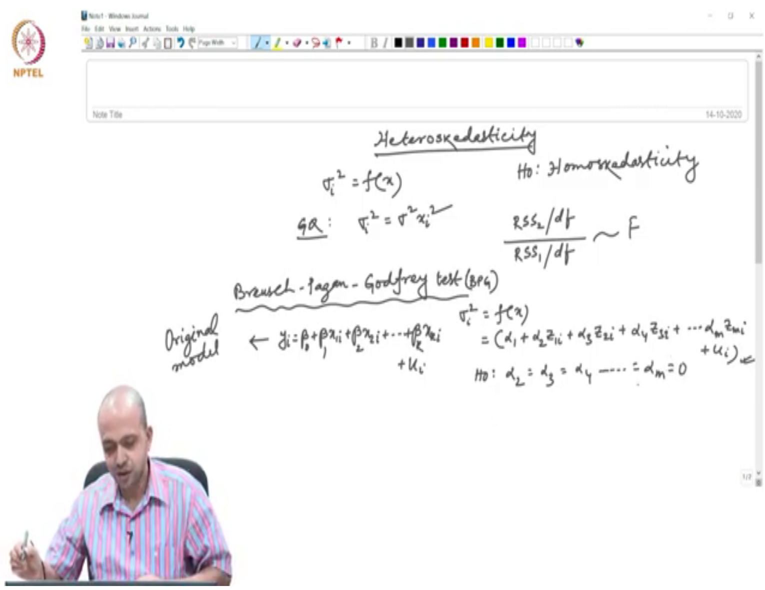
And what is this z actually, let us say that your original model is yi equals to beta 0 plus beta 1 x1i plus beta2 x2i plus betak xki plus ui this is your original model of interest. So, that means we are interested to see whether heteroscedasticity exists in this original model in this data set.

Now, this z basically, the z, the set of variable that consists z coming from all or some of your x variables, some of your x variables that is the assumption that you make. So, we assume that all or some of your x actually consist this z, z1 might be x1, z2 might be x2 like that, so all or some of your x is actually z, so you can write this equation as in terms of x also.

So, what you need to observe here, look at the way I have specifically mentioned the nature of f(x) here vis-a-vis in the context of Goldfeld and Quandt test, this is the procedure. Now, in Breusch-Pagan Godfrey test they have also given certain steps and I will now discuss about the tests.

(Refer Slide Time: 8:25)





So, this is BPG test. So, step 1 and your model is yi equals to beta 0 plus beta1 x1i plus beta 2 x2i plus beta3 x3i plus betak xki plus ui, this is your model. And what is your step 1? So, in step 1 they say that you estimate this model let us say from this is model 1, you estimate your model and then from this model estimate model 1 and get the RSS.

And from RSS what they are asking you to calculate some kind of let us say sigma tilde square which is basically nothing but RSS divided by n. Now, what is RSS by n? RSS by n is actually if you recall that RSS by n minus k equals to we said that this is basically sigma hat square from that OLS model, that means if you divide the RSS by n minus k you will get the sample statistic of the sigma square. sigma square is the error variance which is unknown population parameter and the sample counter part of sigma square i is basically sigma hat square which is coming from your OLS model that is RSS by n minus k.

Here when I am doing RSS by n that is basically the estimate of sigma square coming from the ml method or maximum likelihood method that is the difference. So, both sigma tilde square and sigma hat square, they are actually the sample statistic or sample counterpart of the true population parameter sigma square which is unknown.

So, we are, since in presence of heteroscedasticity sigma square i, sigma square i is actually biased, sigma square i does not reflect the true sigma square, we are trying to get sigma tilde square from the maximum likelihood method and how we are defining this is basically RSS by n.

Then in step 2 they say that once you rectify your sigma square by sigma tilde square then what you generally need to do is you just divide your ui hat square by sigma tilde square and define that as p. So, what I am doing? Since your error variance is disturbed in presence of heteroscedasticity look what I am doing, I am just correcting, I am just adjusting the ui hat square by this new estimate of sigma square which is sigma tilde square and I am defining that as p. That is your step 2.

Then in step 3, you regress this p equals to all your explanatory variables that means you regress p equals to sum lambda 0 plus lambda 1 x1i plus lambda 2 x2i plus lambda k xki plus epsilon i. So, you have to run a regression of p on this. And what you have to do, you have to and collect the ESS from let us say this is model 2.

Then in step 4 basically they say that half of ESS follows a chi square distribution where degrees of freedom equals to m minus 1, where m is the total number of parameters in this original model. So, if the calculated F is greater than chi square tabulated then you reject your H naught and what is our H naught, that there is no heteroscedasticity. So, that means if you think about the logic or philosophy of all these tests then things become very clear to you.

Goldfeld-Quandt test was also based on two RSS- RSS2 by RSS1. Why? Because we assumed that sigma square i equals to some sigma square multiplied by xi square, so sigma square i is basically a increasing function of x, so that is why we divided the entire sample into two, one consist of smaller value, another consist of higher value and then we are taking RSS2 divided by degrees of freedom divided by RSS1 by its degrees of freedom.

So, that means this RSS2 by degrees of freedom is nothing but sigma 2 hat square. And what is RSS1 by its degrees of freedom? That is sigma 1 hat square. Same thing we are doing here. Here we are assuming that sigma square is a linear function of sum of all your explanatory variable that is all.

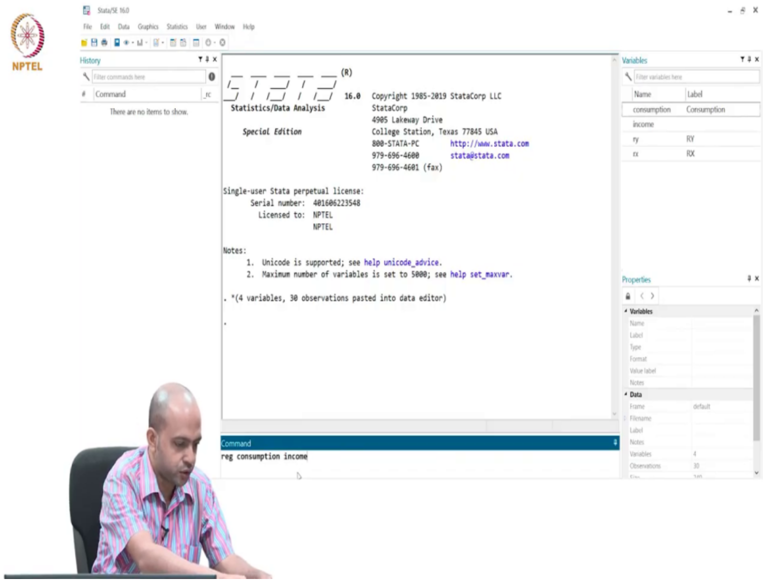
See if that is the case what we are trying to do, we are trying to adjust the error term of the original variable ui with a proper estimate of sigma square and how do you get a proper estimate of sigma square? This is the way the traditional definition RSS by n minus k which is the definition of sigma square from the OLS model is not working here in presence of heteroscedasticity.

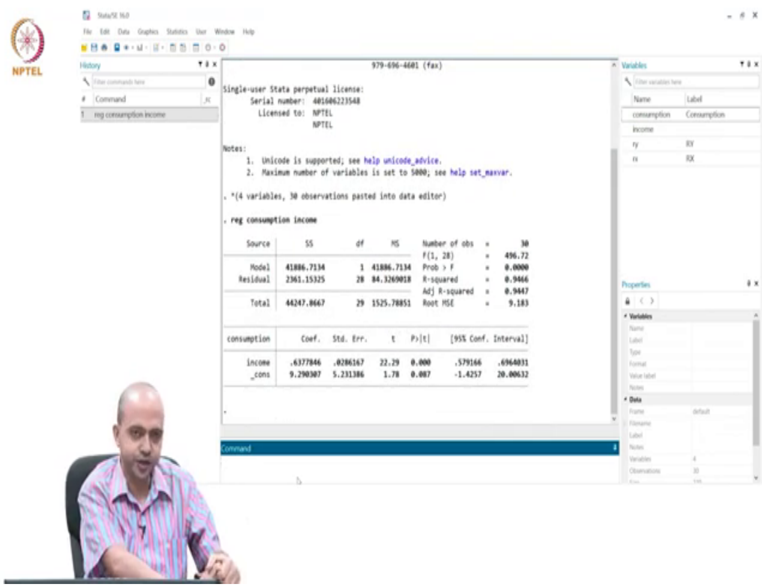
Instead of using that what we are doing, we are taking RSS by n, that is all. Then we are adjusting our ui hat square by this newly estimated sigma tilde square and defining that as p. So, if at all this adjusted error term has something to do with your explanatory variable then obviously this p defined in this way should be explained by this parameter.

So, that means basically from this model I am testing this hypothesis lambda 1 equals to lambda 2 equals to lambda k equals to 0 or not. So from here what I am testing this is my test alpha 2 equals to alpha 3 equals to alpha 4 equals to alpha m equals to 0. So, if this is rejected then that means all these parameters are actually not equal to 0 that means sigma square i is something to do with the explanatory variable, but it is not rejected that means these are all 0-sigma square i equals to only alpha you have constant error variance that is the logic.

So, we should understand the logic of this test and logic is basically we are trying to modify some way or the other the RSS term and then we are constructing the test statistic based on that, this is the RSS modified, then we are defining p, p we are regressing, then we are collecting the ESS dividing it by 2 and that follows this chi square distribution with m minus 1 degrees of freedom, where m is the total number of parameters to be estimated from the original model, this is the procedure of Breusch-Pagan and Godfrey test or BPG test. Now, what we will do, we will take the same data set what we are using yesterday and then we will see how to demonstrate this particular test in stata.

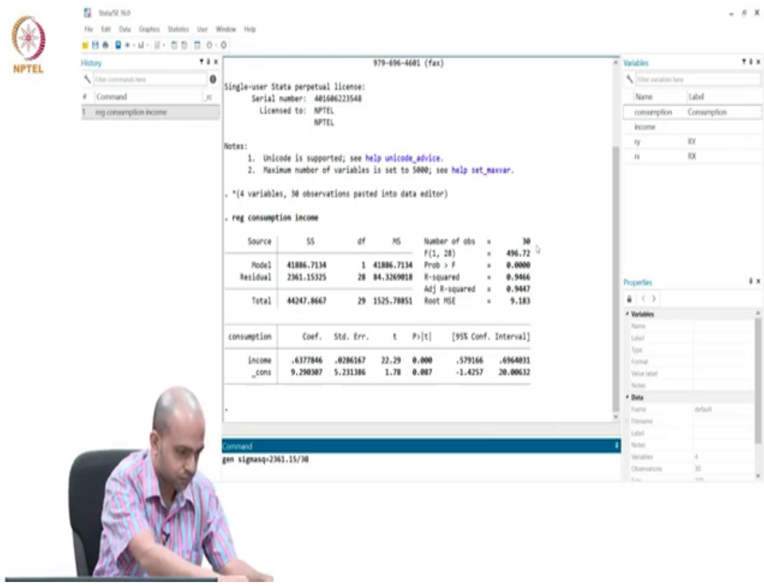
(Refer Slide Time: 20:29)





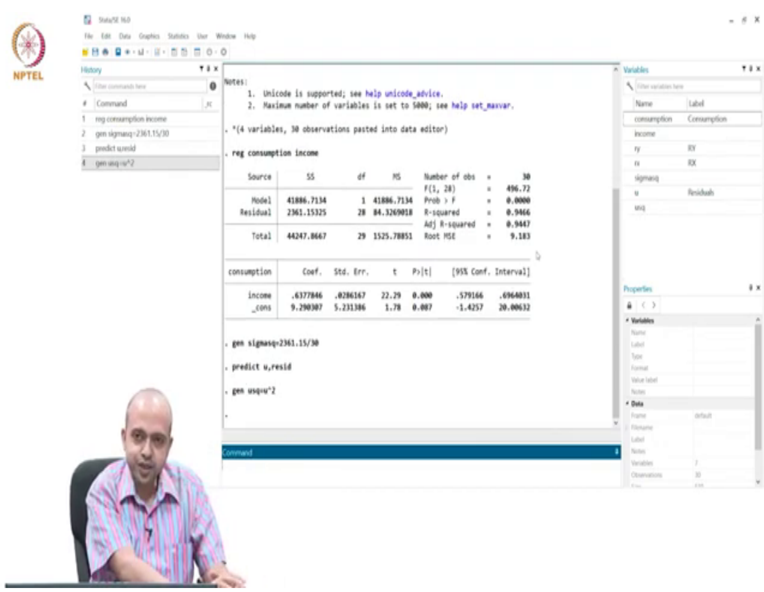
So, I have already imported the data, so first what you have to do, you have to run your original model where original model is reg consumption income, this is your original regression. And then from here what you have to do, you have to collect the RSS and you have to divide by RSS by n to get sigma tilde square.

(Refer Slide Time: 21:07)



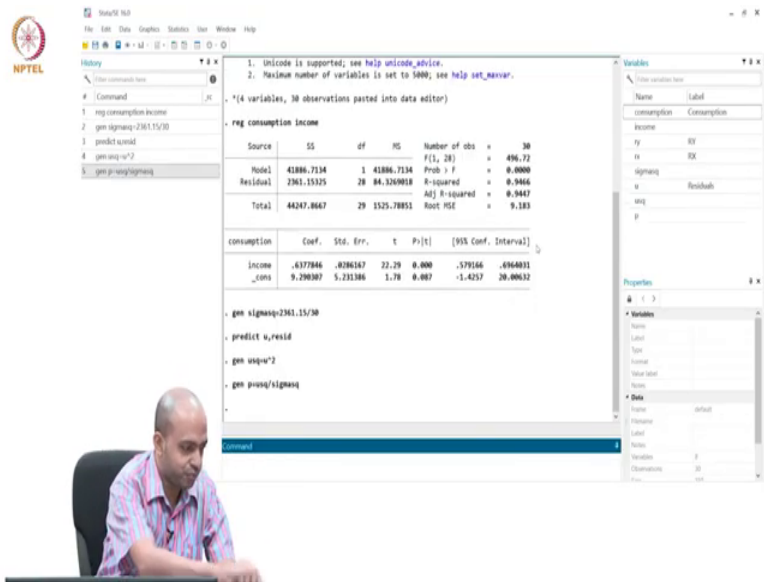
Now, what I will do I will define the sigma tilde i square I am simply writing sigma square sigma sq equals to, what is your RSS here? RSS look 2, 3, 6, 1 equals to 2, 3, 6, 1, 0.15 divided by n and what is your n here, 30.

(Refer Slide Time: 21:40)



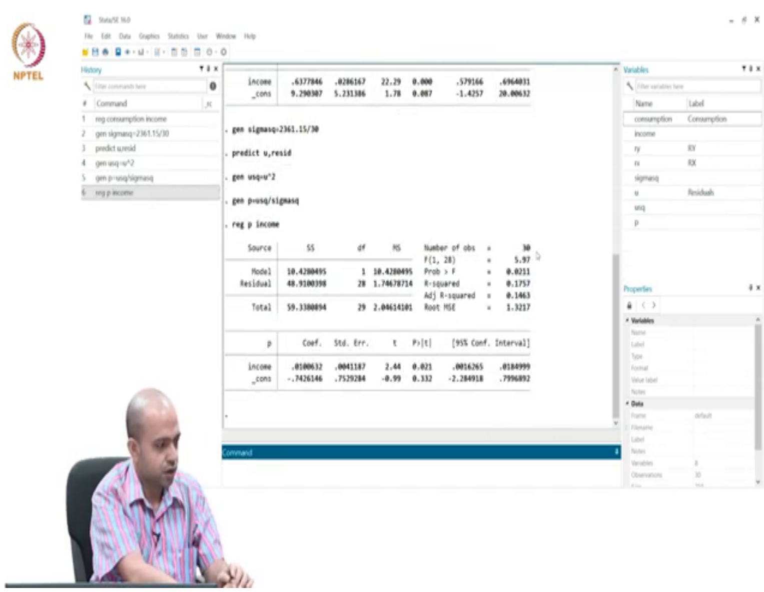
Then what you also do you predict, predict u comma residual and then you generate the square of u, gen u square usq equals to u, this is u square. So, I have generated sigma tilde square, I have generated u hat square. Then how do you define your p?

(Refer Slide Time: 22:21)



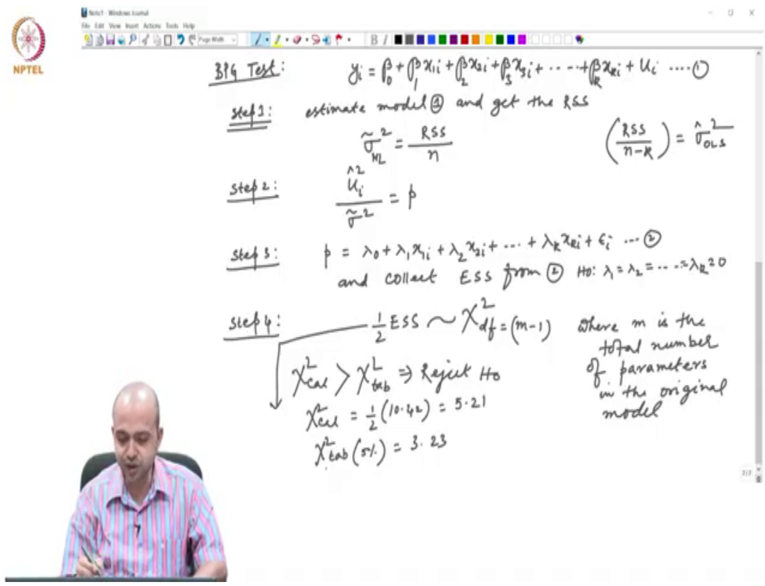
The p is basically gen p equals to this usq divided by usq divided by sigma sq that is how I have defined my p, so ui hat square divided by your sigma, sigma square, sigma tilde square so this is your p.

(Refer Slide Time: 22:49)



And then you regress p, p on your explanatory variable you have only one explanatory variable which is income. And from this model you have to take your ESS which is 10.42.

(Refer Slide Time: 23:29)



So, if it is 10.42 then if you look at from here if it is 10.42 that means yours chi square calculated should be half into 10.42. so which is nothing but 5.21 so this is your calculated chi square. Now, what is the tabulated value?

Chi square tabulated at 5 percent level of significance equals to something around 3.23 or something, you can check the chi square value with m degrees of freedom. What is m here? m equals to 2, so degrees of freedom would be 1, this chi square tabulated would be 1 degrees of freedom because in your original model you have only 2 parameters, income coefficient and the constant term, this is 3.2. So, you can reject your null that means this implies that chi square calculated is greater than chi square tabulated at 5 percent significance. So, at 5 percent level of significance you can say that your data is suffering from heteroscedasticity problem.