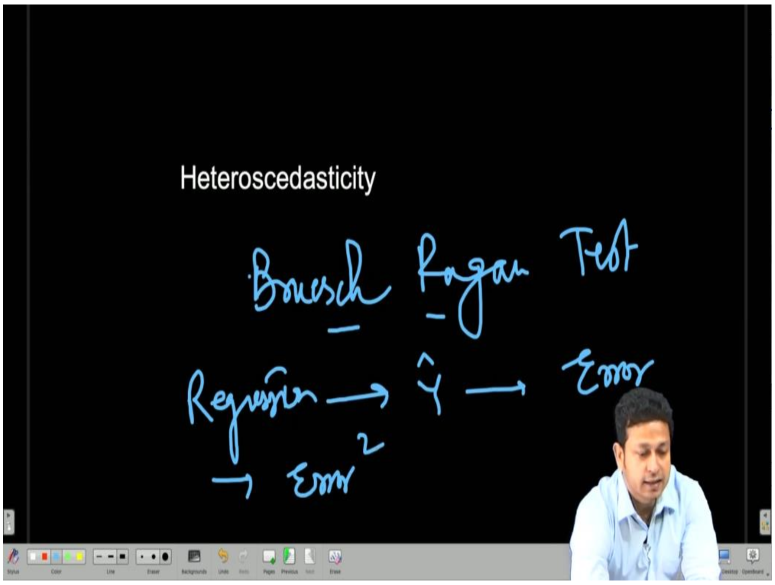
Applied Econometrics
Prof. Tutan Ahmed
Vinod Gupta School of Management
Indian Institute of Technology - Kharagpur

Module - 8
Lecture - 68
Heteroscedasticity (Contd.)

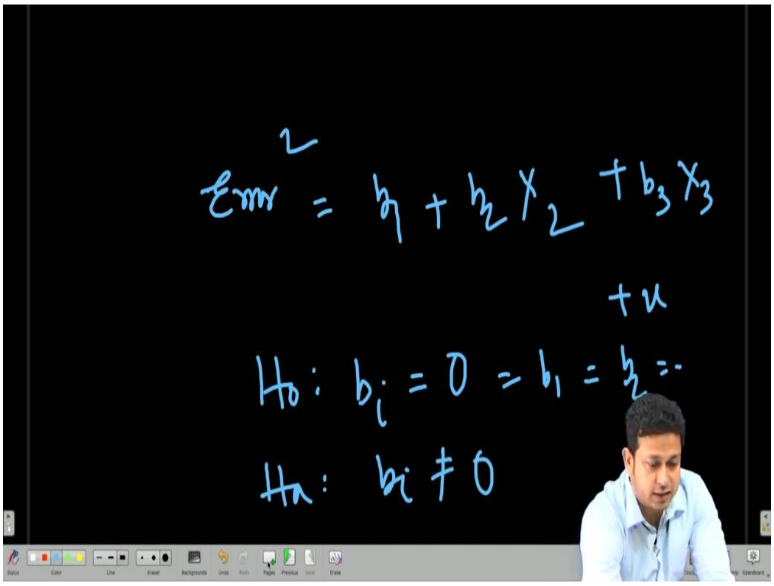
Hello and welcome back to the lecture on Applied Econometrics, and we have been talking about heteroscedasticity. Now, in the previous lecture, we have seen Goldfeld Quandt test. And we said that the test is performed under certain restriction, but we can also have tests where we do not have any such restriction; and one such test is called proof Breusch Pagan test.

(Refer Slide Time: 00:46)



So, Breusch Pagan test, it actually does not assume any such restriction like we had for Goldfeld Quandt test. So, how it is done? So, essentially, what we do is, we we run the OLS; we first create the regression equation. And once the regression equation is performed, we obtain the predicted Y of course. And from there, we actually get the error term. And when we get the error term, we actually square the error term.

(Refer Slide Time: 01:29)



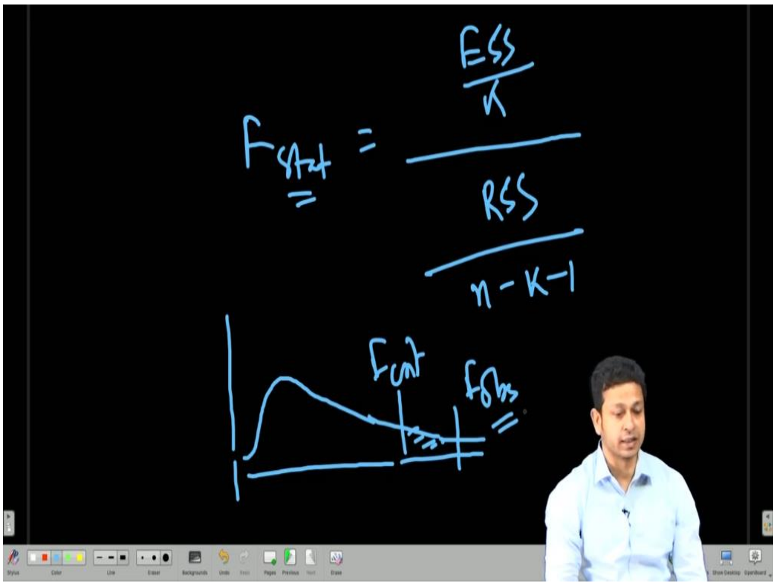
And then, we perform a regression where my Y variable is the error term square, and my X variable is all the X variables that I had in my original regression. So, the idea is this. I am trying to see here if my error term has any sort of relationship with the X variables. So, why I am squaring the error term? Because that is actually representing the variance of the error term. So, you really do not have the actual variance; so, this is something that you use as a proxy.

Now, when you do that and you basically use the X variables just to see if any of the X variables are going to be significant, if it is significantly influencing the error term. So, that means that there is a relationship between the X variables and the error terms. Then the null hypothesis will be, we can basically, looking at this equation, we can say, the null hypothesis is going to be, all the beta i's are going to be 0.

And basically, that means that beta 1 is equal to beta 2 and so forth. And all these are going to lead to a value of 0, whereas my H alternative is going to be, at least one of the beta i is not equal to 0. So, if at least I can show that 1 beta is not influencing the error term, 1 X is not influencing the error term; so, then, my error term is not correlated with any of these explanatory variables.

So, that is what we are going to see in the case of this Breusch Pagan test. So, that is as simple as that. Now, how do we do that? It is a very simple test; we are quite familiar with it; it is simply the F-test that we always perform. And what we do here is, we simply take the explained some of square; and then we divide the explained sum of square with the corresponding degrees of freedom; then you take the residual sum of square; you divide the residual sum of square with the corresponding degrees of freedom; and then you take the ratio between these two and simply get the F-statistics; as simple as that; what we do for a F-statistic.

(Refer Slide Time: 03:36)

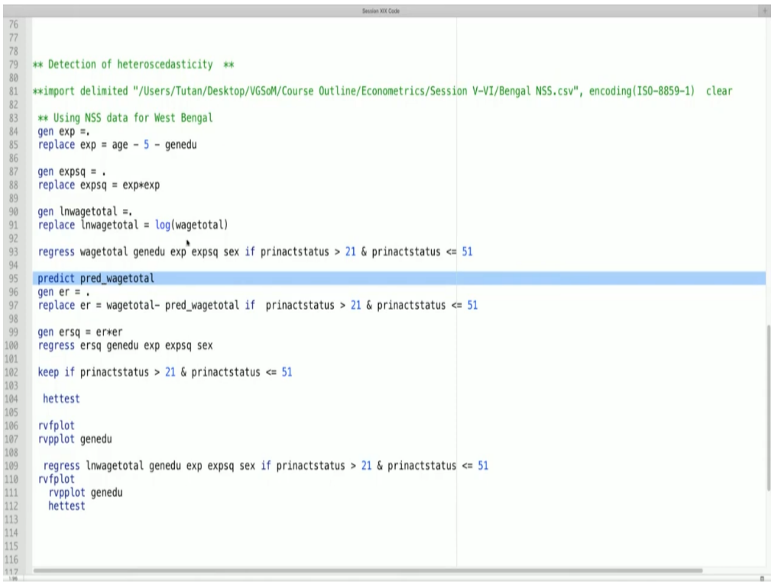


So, how we do it? So, let us write it down. F stat is nothing but my ESS, explained sum of square, and we divide the explained sum of square with K. So, K is the number of explanatory variable. And whereas, my residual sum of square, we divide that by n - K - 1. I mean, I have already explained why we use K and n - K - 1 as 2 degrees of freedom. And then, as usual, what we do is, we calculate the F stat.

My distribution is not really that great; let me draw it again. Now it is somewhat better. So, here I will have my F critical. And I will see where my F stat falls if it is here. I kind of reject the null that all the betas are actually equal and equal to 0. So, that means, if my F observed is here, so, then I will say at least 1 beta is not equal to 0. So, that is how we do it. It is as usual when we do for a simple F-test.

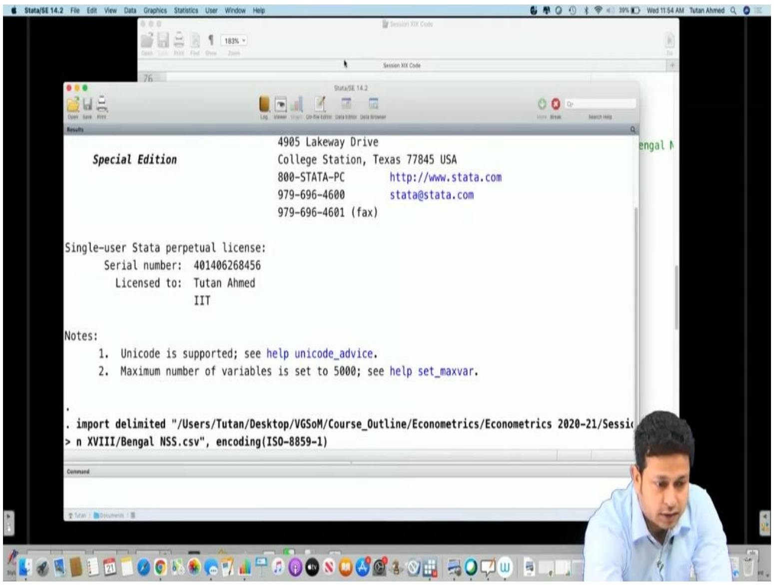
So, basically, when we fit a regression model, use the F-statistic to understand the joint significance of the model, and that is precisely what we do here again. So, let me actually show you how to do that in Stata.

(Refer Slide Time: 05:08)



So, let me check if I have imported the data; I need to import the data. So, I will import the data for West Bengal National Sample Survey data; let me check; this is the data set. So, data set is imported.

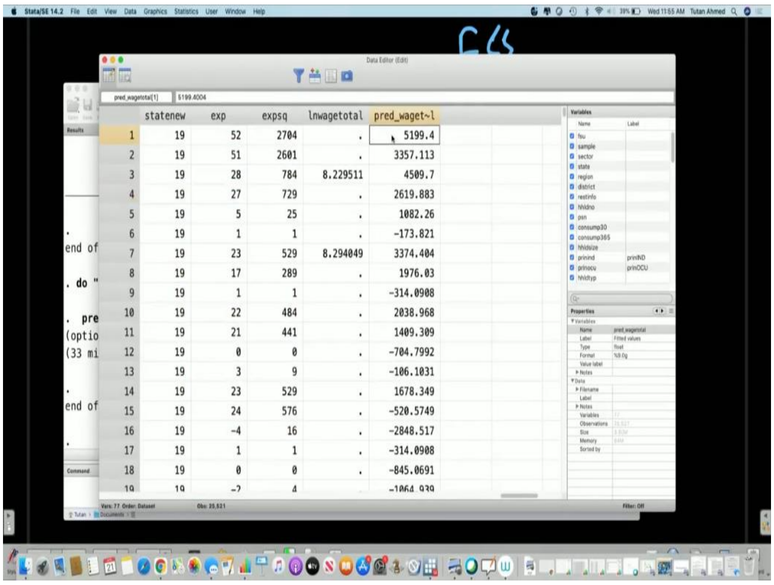
(Refer Slide Time: 05:32)



Now, I will create some variables. I have already created the code. So, it is very simple; experience, experience square and log of wage. So, I run this code. I have created the variables for experience, experience square and log of wage. Now I run the regression for certain groups; so, who are actually in the salaried or casual employment sector, because for them, I can get the wage details.

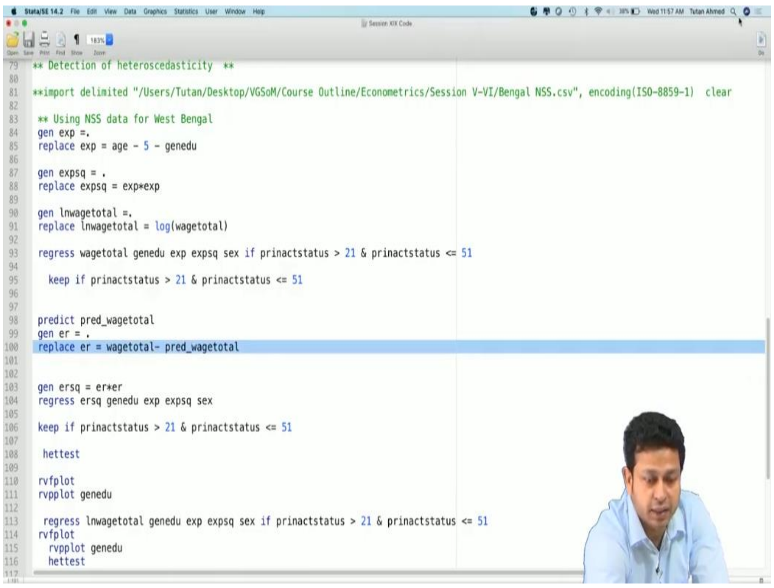
So, that is why I have created this filter; for this particular group, I want to see the wage. Now, instead, if you want to get the predicted value of Y, so, what you simply do is, you predict the wage total. So, using this explanatory variables, you want to see where the predicted value lies. So, you simply use the command predict. So, I created this variable pred wage total; you can write whatever you want; Y hat, anything you want. So, if I run it, I will get the predicted value of Y. And to give you a sense, how it looks like, we can actually go to the data editor and see how this variable is created.

(Refer Slide Time: 06:36)



So, here you see the predicted value. So, this is the predicted value of Y. So, this is the wage that my model is predicting for these people. Now, I will actually do one more thing, because the entire population is there, so, I can actually drop the people who are; let me actually drop these people first here.

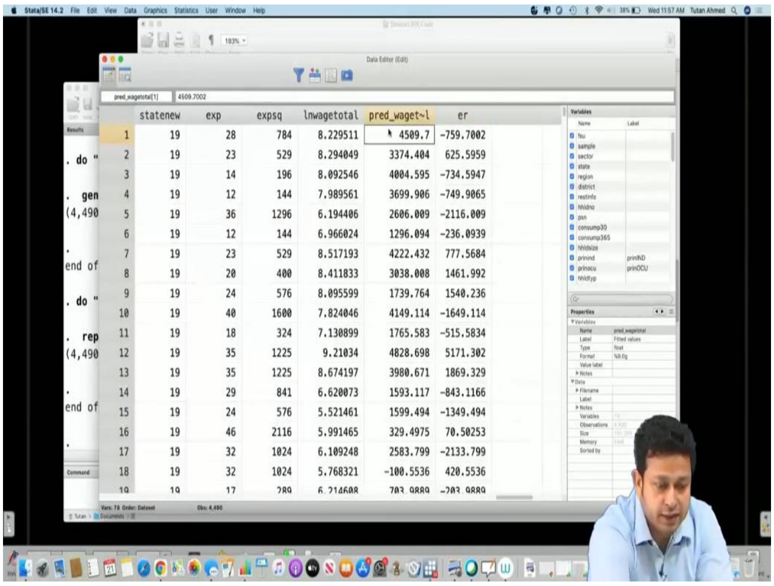
(Refer Slide Time: 07:07)



So, I am just keeping these people who are within this group, those who are in the salaried employment or casual employment. So, now, let me actually keep it here; I mean, does not matter, but just for our; so, I have just dropped these observations which are not in this range. So, that means I only have people who are in the salaried and casual sector. And now I do the prediction.

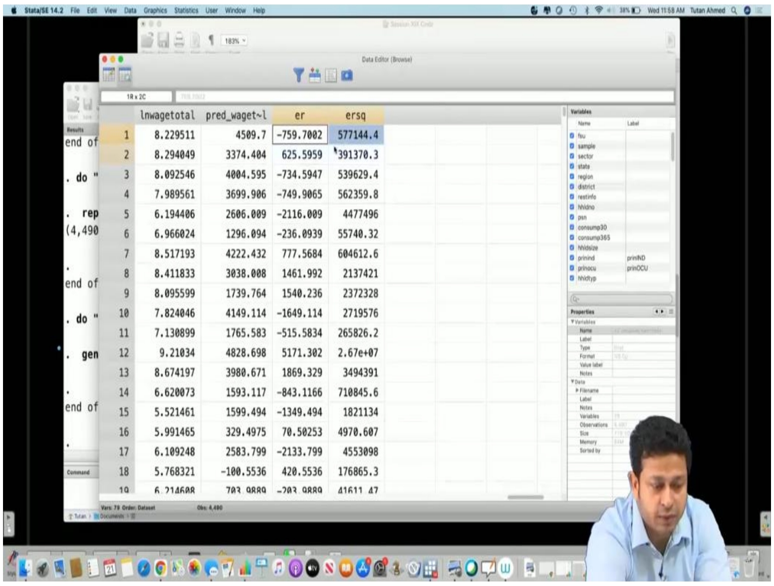
And then, we can actually generate a variable which is called er, which is going to be my error term. And how do I get the error term? I simply subtract the actual value of Y and predicted value of Y; as simple as that. So, let me just do that. So, I have generated the variable er; it is the error. And then I replace this er is equal to this Y variable, original Y variable wage total minus predicted wage total. So, I do not need to create this condition, because I have already dropped people who are not in this group. So, if I run it, what I get it is that; let us see.

(Refer Slide Time: 08:42)



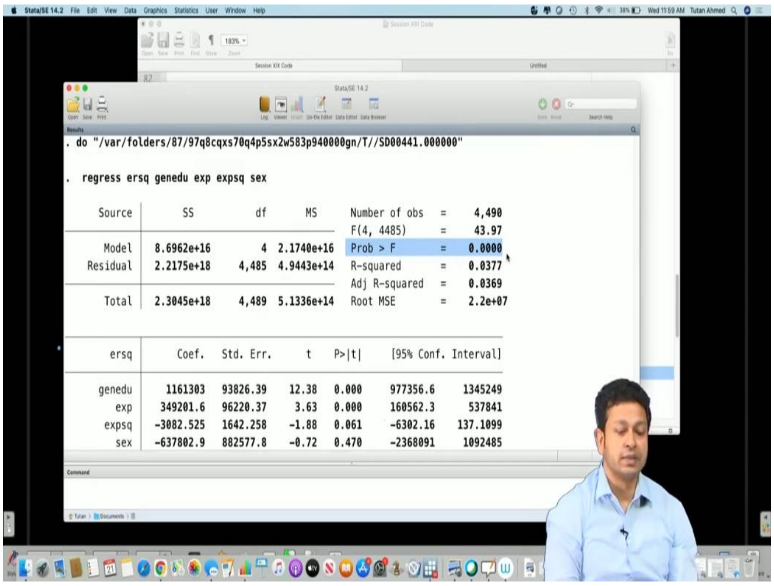
So, I will now see, going to the data editor, I will see how this error term looks like. So, now I have my predicted values of Y. And if I subtract from the original Y, so, I get these values. So, this is how I get the error term. Now, what I will do? I will basically square the error term. That is how Breusch Pagan test tells me to do. And I generate this variable error square and I define it as error into error. So, it will create the error square term. We can again see the error square, how it looks like.

(Refer Slide Time: 09:26)



It is going to be the square of the error term. Basically, this value, error value and its square. Now, if I run the regression, so, now, what I have to do? I have to regress the error square with all those explanatory variable which I had in my original model. So, that was my original model. I had education, experience, experience square and sex. So, I will just use these variables here, and I will regress. So, let me do that.

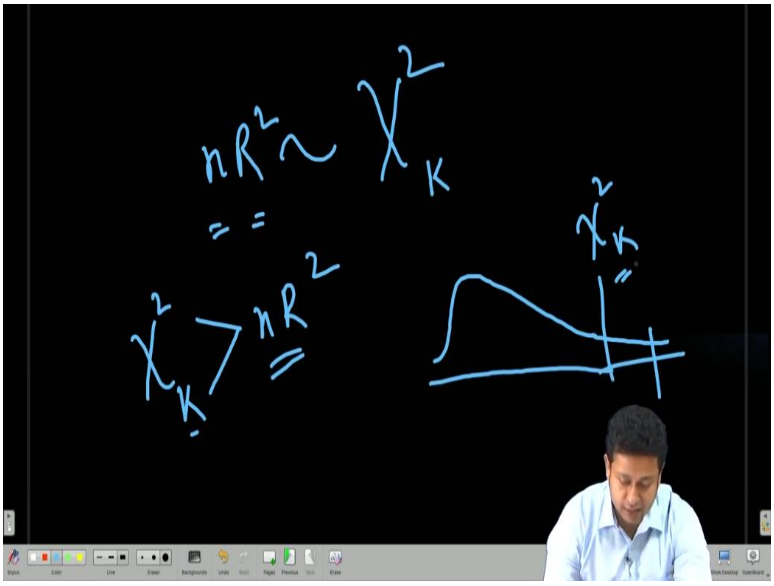
(Refer Slide Time: 09:58)



And now see what happens here. All I have to do is to see this F-statistic. So, what we have done previously? My model is already taken into account when I run a regression equation, and here it is. So, my F-statistic, the P-value is actually very small. So, that means, my null hypothesis of beta coefficient is equal to 0 is not correct. Essentially, it means that some of the betas are actually strongly or actually related with the error term.

So, basically, we have to check for the heteroscedasticity. So, we can assume that there is heteroscedasticity in the model. So, that is basically the idea of Breusch Pagan test. So, this is one way of conducting Breusch Pagan tests. So, there is another way of looking at Breusch Pagan test and that is basically, we take into account the R square value. So, what we do is, we assume that the R square value, and if it is multiplied with n; n is the number of observation; so, it follows a chi square distribution. So, how we do it?

(Refer Slide Time: 11:03)

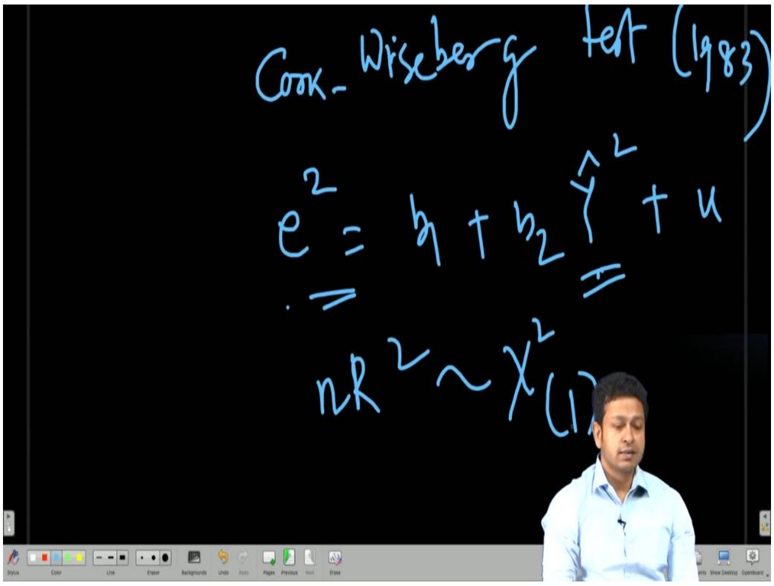


Basically, we take the r square from the regression and then multiply with n; and that follows a chi square distribution with degrees of freedom is equal to K. So, K is my number of explanatory variables in my model. So, this is also we can see. So, essentially, what we do is, we take the R square value; we multiply it with n; so, if let us say n R square is actually, say smaller than chi square critical, which is basically chi square value for the K degrees of freedom; if this small, so, then, what will happen is that, we will say that, similarly, like the F-test, we will say that; this is a very poor drawing; let me draw it again.

So, let us say n R square value is here and my chi square critical is here. So, I will say that we reject the null hypothesis. So, this is how we can also conduct the Breusch Pagan test. So, that is basically the idea of Breusch Pagan test. We also have something called; so, essentially, this model kept on improvising over time. And there was in 1983, it was Cook-Weisberg, they actually improvised the model further.

And what they did was, they ran the regression with Y predicted square as the independent variable. So, instead of taking so many other explanatory variable, I only take Y predicted square as my independent variable, explanatory variable to explain the error square term. The dependent variable remains same, we have the error square; and in the independent variable, I only have 1 variable, and which is Y hat square. So, that is it. And the regression, let me write down, Cook-Weisberg test.

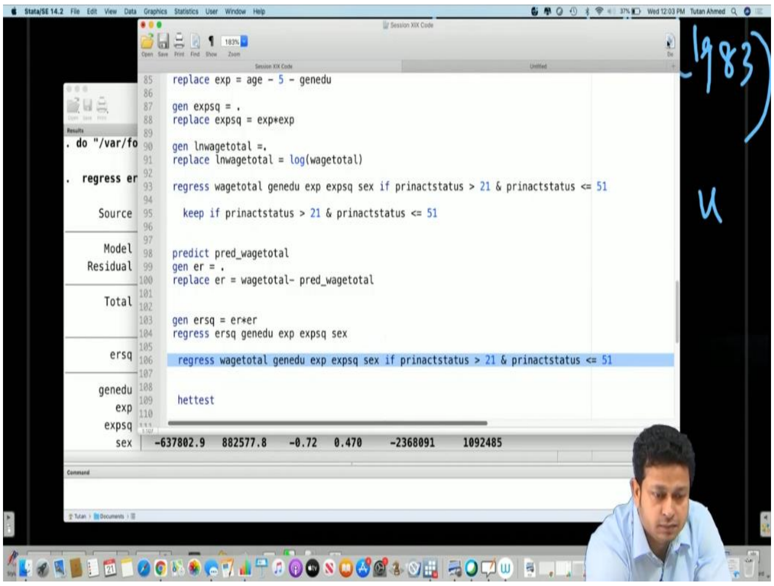
(Refer Slide Time: 13:11)



This was done in 1983. And how it was done? So, basically, the idea is that, the error square is regressed on Y hat square and u term. So, now here, I have like only 1 explanatory variable. And again, in the similar manner, the n into R square, the R square value that you get from this regression equation actually follows a chi square distribution. And because my independent variable is always 1, so, for Cook-Weisberg test, I will always have the chi square critical for 1 degrees of freedom.

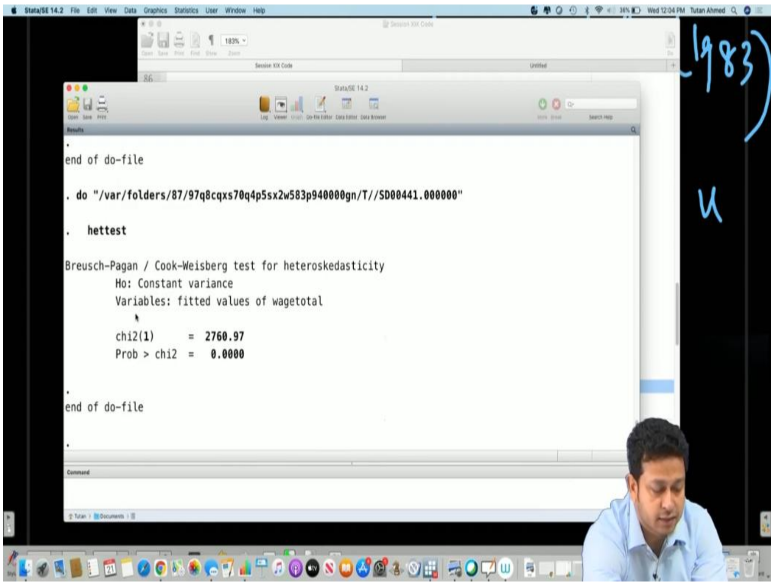
So, that is always constant, because I have only 1 explanatory variable. And this is a very simple command in Stata to perform it, and that is called hettest. And we are just going to see for our model, if we perform the Cook-Weisberg test, all I have to do is to write the command hettest.

(Refer Slide Time: 14:26)



So, let me delete this. So, if I run hettest, so, essentially, I have run the regression equation. So, let me run the original regression equation here. So, this is our original regression. And now, if I run the hettest, so, essentially, what it will do? It will do a whole lot of exercise. So, it will get the Y hat; it will square the Y hat; it will get the error term; it will square the error term; then it will run the regression. So, let us see the result here.

(Refer Slide Time: 15:03)



So, this is how it looks like when we perform a hettest. So, it says that the H nought is constant variance and variables are basically only 1 variable, which is fitted values of the wage total. So, this is Y hat and chi square is always having 1 degrees of freedom, and the value is 2760. And the P-value is going to be very small, and you kind of reject the null hypothesis.

So, this is how you basically perform the Cook-Weisberg test, which is a variation of the Breusch Pagan test. So, with this, we will end this lecture. And in the next lecture, we are going to talk about the White test. White test is again just one step ahead of the Breusch Pagan test. And that will give us some idea about all the different kinds of tests that we perform to check if there is any heteroscedasticity in the model. Thank you.