Engineering Econometrics
Prof. Rudra P. Pradhan
Vinod Gupta School of Management
Indian Institute of Technology, Kharagpur

Lecture – 30
Heteroskedasticity Problem (Contd.)

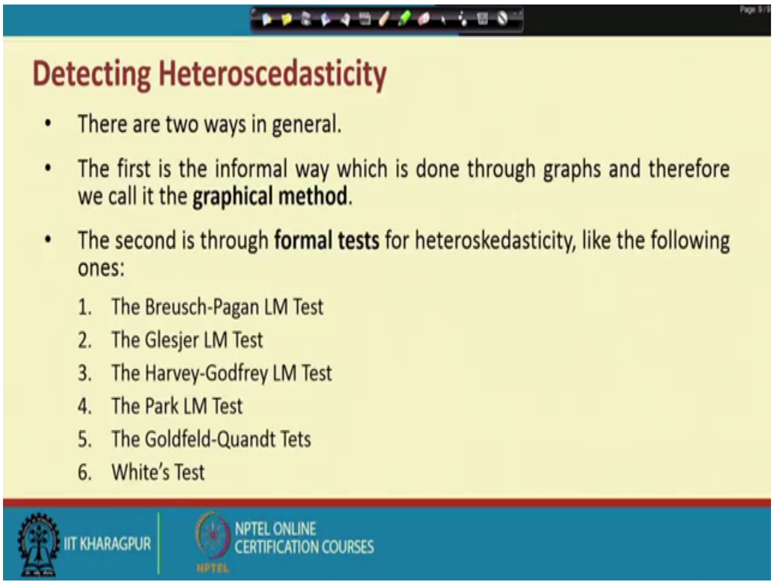
Hello everybody. This is Rudra Pradhan here. Welcome to Engineering Econometrics. Today we will continue with modeling diagnostics and that to again with respect to heteroskedasticity problem. In the last lecture we have specifically highlighted this particular component it is a kind of you know virus in the modeling environment, and that to with respect to error term and again that with respect to error variance.

The requirement of classical linear regression modeling or the requirement of you know the use of ordinary least square mechanism is that error variance should be similar over the time or over the cross sectional unit. But in reality you will be find error variance cannot be similar in most of the instances because of you know various reasons, because one of these strongest region is that you know because of you know the presence of outliers or over the time committing error will be at the at the minimum. So, as a result, so there is a high chance the error variance may not actually same over the time. So, if you go by error learning modeling approach so, the committing of errors usually in a kind of you know declining trend. So, as a result there is a high chance and the error variance cannot you know same.

So, there is a little bit you know unequal and we like to check whether it is a kind of you know you know equal variance or whether there is a kind of you know unequal variance and we have discussed in details in the last lecture. And against what we will do here econometrically we like to check because in the last we have given you the structure about the graphical you know inspection and the kind of you know visualization process.

So, here we have standard you know techniques or econometrics technique through which actually you can detect the a heteroskedasticity problem and after detections we should look for the kind of you know solution. So, in this lectures we like to use different techniques to detect the heteroskedasticity and then look for the kind of you know solution as per the particular you know a requirement.

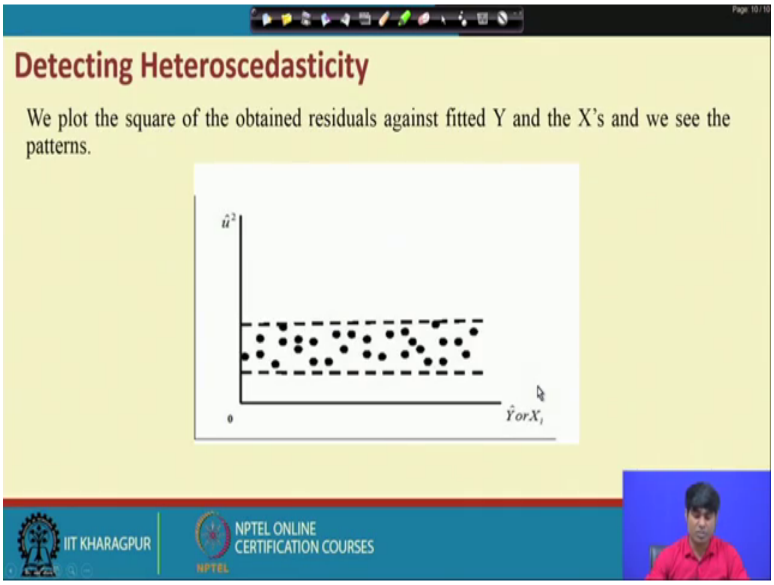
(Refer Slide Time: 02:59)



So, now coming to the detection of you know heteroskedasticity. So, there are two ways in general first you know it is the informal way which is drawn through graphs and therefore, we call as you know graphical methods or you know visualization mechanism which you have already discussed in the last lecture. In the second case, so that is what the first mechanism through graphical structure and then second case so we have actually very much standardized you know test structure that is what called as you know formal test to deter the heteroskedasticity. And we have actually number of you know such kind of you know tests formal tests are there and to check the level of you know heteroskedasticity. Like you know Breusch-Pagan LM test, Glesjer LM test, Harvey or Godfrey test, Park test, White test, Spearman rank correlation test, GQ test.

So, there are so many tests are there which can help you to detect the heteroskedasticity and then like to quantify whether it is actually statistically together significant or not. So, there may be chance of you know heteroskedasticity if it is not starts together significant. So, we will not you know bother about it, but if it is coming statistically significant then we should actually see how is the kind of you know solutions, until unless you solve that particular items you should not you know use this model for the prediction and the kind of you know decision making process. So, let us have the kind of you know you know on this context.

(Refer Slide Time: 04:46)

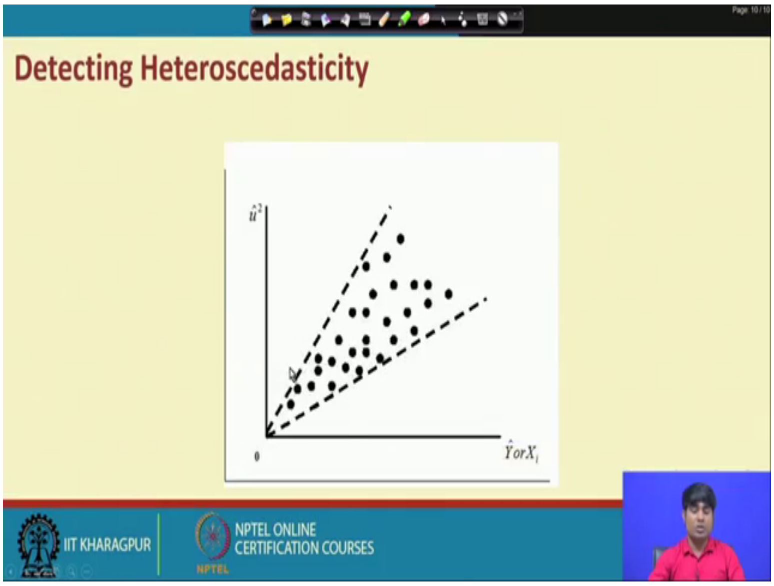


So, first we will start with you know all more or less you know or you know all these tests are you know a very similar kind of you know things say, because it is the game of you know deriving error term and checking the pattern of the error term and graphically we have just you know visualize. And when you are using graphical methods then you need not require to connect with any test or you know connect with any of the kind of you know mechanisms.

But when you use formal test, so there is a lots of you know mechanisms are there to process and then see the kind of you know results about the heteroskedasticity. So, this is what the plotting which you have already discussed, and plotting of you know error term with respect to either Y or you know X so that means, technically a firsthand process is to have the estimated line to get the error terms. So, now, look the behavior of the error term that is the first requirement.

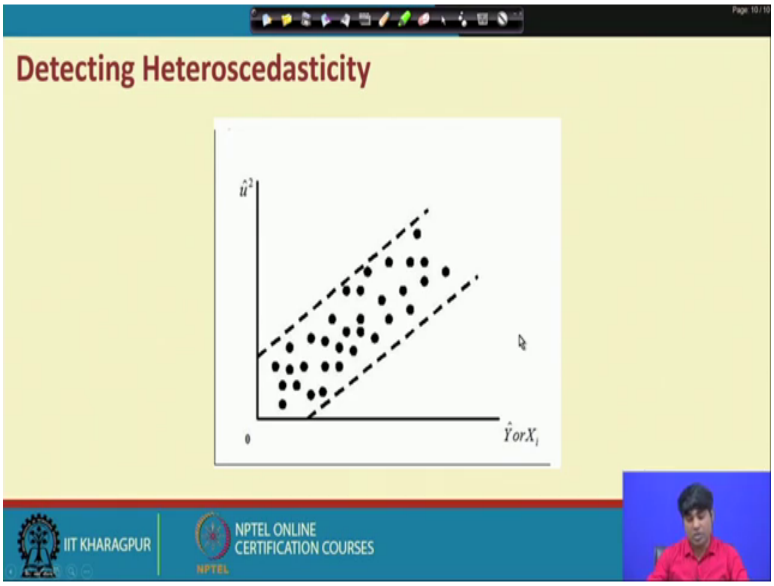
In the second requirement if you go by formal mechanism, then you can connect the movement of the error term with respect to dependent variable or with respect to independent variable. So that means, we like to check whether the error term is having some kind of you know typical relationship with the independent variable or dependent variable that too if the formal mechanism is concerned to detect the heteroskedasticity. So, that is what here actually. So, we are just plotting the behavior of the error term that is the behavior of the error term, means it is the behavior of the variance of the error terms not simply the error term it is a the because we are looking for you know a variance component to just declare the homoscedasticity or the heteroskedasticity. That is why we have to plot the error variance instead of simply error terms, right.

(Refer Slide Time: 06:36)



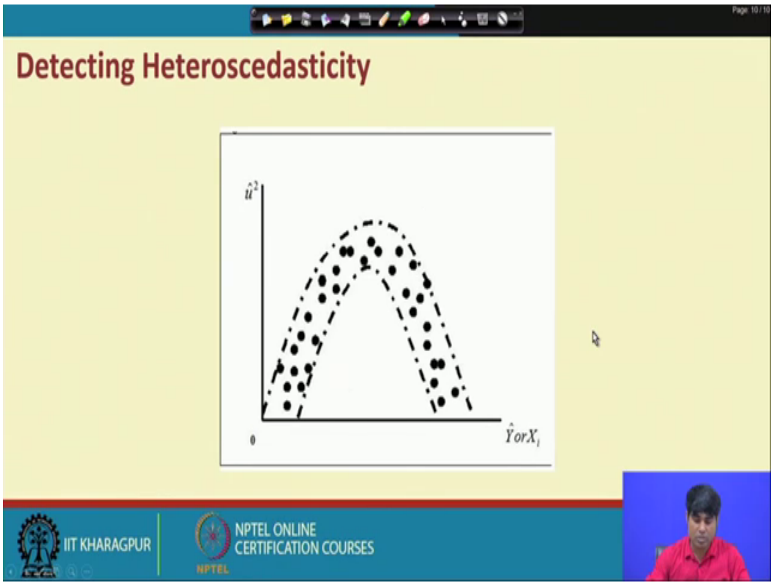
And this is what the equal spread you know diversifying. We have already checked this in the last class.

(Refer Slide Time: 06:42)



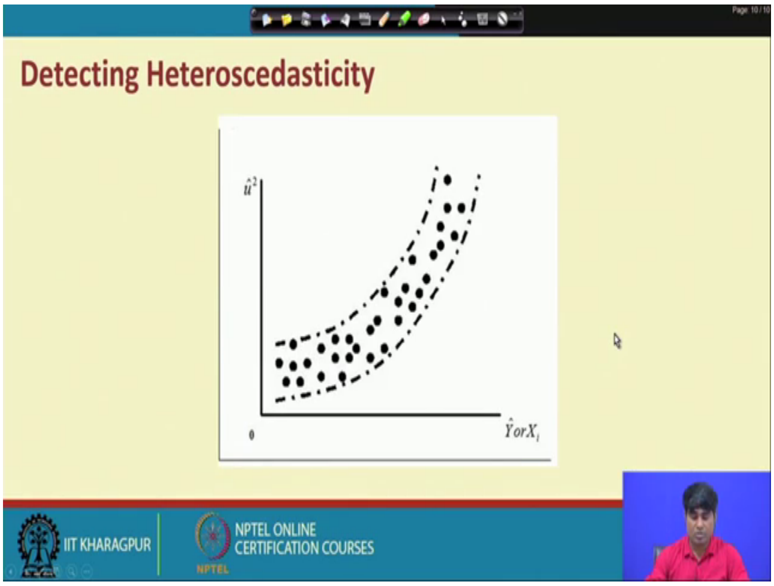
And this is against not actually very much diversifying, but it is it going little bit volatile and then it is in a kind of you know standard you know confidence interval. But still there is a heteroskedasticity because the confidence interval for this you know movement is very high technically. So, that is what the kind of you know.

(Refer Slide Time: 07:02)



So, it is also a presence of in a heteroskedasticity, here the moments are actually not so high it is in a kind of you know similar confidence interval, but ultimately what is happening. So, the variance of these floatings, you know following a particular you know pattern. And if that is the case then this is a clear cut signal about the heteroskedasticity involvement.

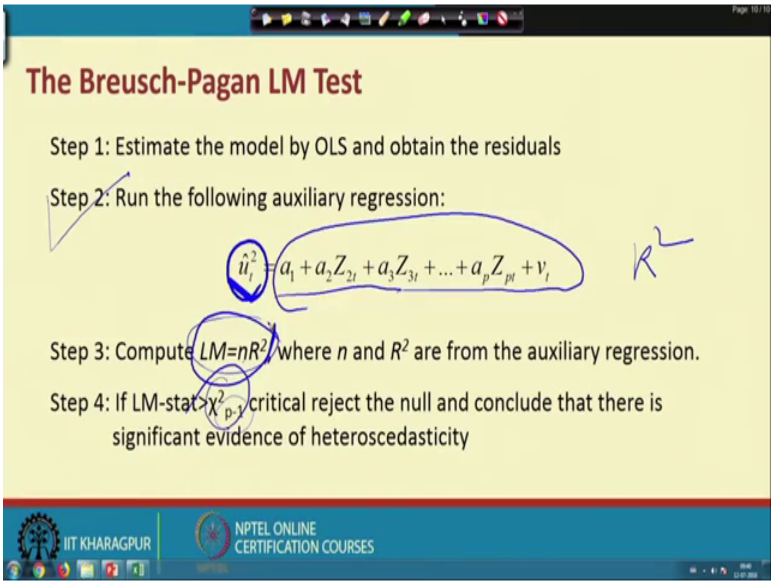
(Refer Slide Time: 07:30)



Similarly, this is also case this brings the functionality of you know exponential function so that means, it follows a particular you know patent. So, the moment error variance follows a particular patents if it is a constant then it would be near line only either in horizontally or you know vertically. So, it will be in a kind of you know line that is what the kind of you know requirement of all.

Ultimately when these variants will be behaving in some kind of you know patents like you know in a kind of you know declining trend or you know increasing trend and a different loop hole together like this case. So, then by default this is the clear cut signal that you know there is a heteroskedasticity problem.

(Refer Slide Time: 08:11)



And if you go by formal processing and then the first method is you know VP test and here is what we are doing actually the step 1 is to apply wireless get the estimated equation and use the estimated equation to have the residuals, then square the residuals and have the kind of you know modeling further.

So, what we are doing actually? It is a very simple mechanism here. So, get the error term and square the error term and then connect these error terms with all the kind of you know independent variables that is the case here actually. So, ultimately, so this is the error variance and then square of the error term that is you know taking you know that is where we are usually using for the you know further processing to detect the heteroskedasticity.

And these are all actually different variables. That is what the step to process. After running this models, so you will get actually R square as usual and then we will use lagrangian multiplier tests that is technically called as a LM test and which is nothing, but actually n multiplied by R square that is coefficient of determination. And n is the sample size and this particular test follows the pattern called as you know chi squared distribution. So, you can go to the chi square table with a p minus 1 degrees of freedom, then you will get the tabulated value. So, now, here we have actually calculated value just compare you know tabulated value is a calculated value at a particular probability level. If the calculated (Refer Time: 09:50) tabulated value and then by default we are rejecting the true null hypothesis and declaring that you know there is a heteroskedasticity. So, our null hypothesis to test the particular process is that you know there is no heteroskedasticity in the estimated model.

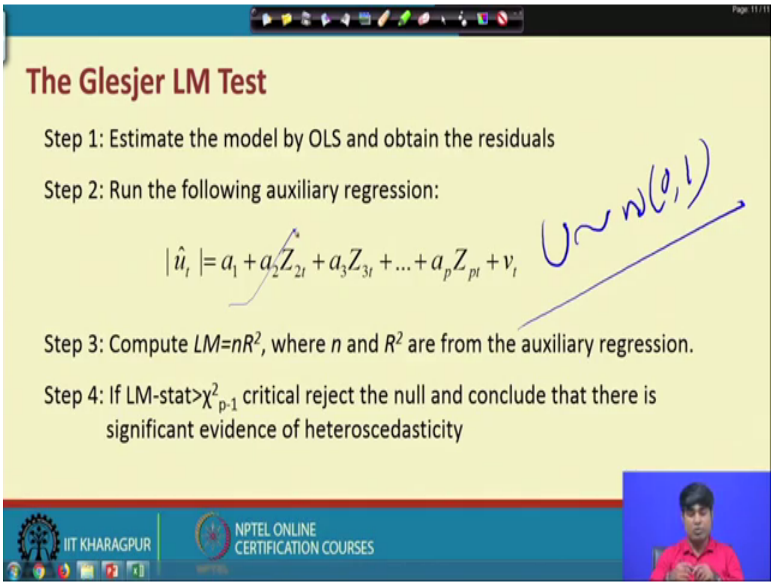
But ultimately we will go by this particular rule and rejecting the particular for and by default the declaration is that you know there is a heteroskedasticity and that to the heteroskedasticity to be statistically you know significant. So, the moment the level of heteroskedasticity signal is statistically significant so we by default will need actually solution.

So, that may be a restructuring of the models the entire process or the kind of you know data whatever may be this. So, means we will go by trial and error mechanism then we will continue the process because it is a very much iterative process and continuous process. So, every process you can you know restructure then re-estimate get the output check the heteroskedasticity level of heteroskedasticity and check the significance level. Then you just start at a particular point of times where there may be heteroskedasticity completely removed or if not heteroskedasticity will be there at a minimum level, but it should not be statistically significant that is what the requirement is all about, ok.

So, with this you know this is VP test likewise you we have a couple of tests this is a Glesjer test. So, this is also similar mechanism and more or less seems what I have already mentioned. So, here what is happening we are actually connecting square of the error terms. Now, we are here actually modeling modulus of the error terms that means, when you derive error term by default some of the error term will be positive some of the error term will be negative that is why mean of the error term is actually equal to 0.

So, and if it is actually normal case homoscedasticity case then by default 50 percent of the error terms will be a left of the left of the estimated line and 50 percent of the error term will be right up the you know estimated lines or 50 percent will be the up and percent would be the down. So, then by default they will tie and as a result the final result will be 0.

(Refer Slide Time: 12:10)

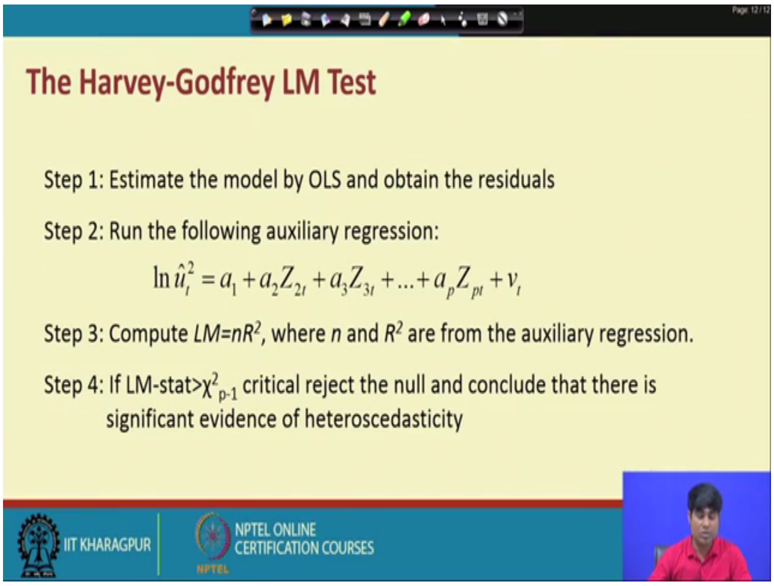


So, if that is the case, so then yeah you know the declaration is that you know u follows normally distributed with you know 0 mean and unit variance. If that is not the case and by default there will be heteroskedasticity problem so that means, technically here what we are doing, we are having the error term then take the modulus because if you do not take the modulus and there is a chance the sum will be a get or equal to 0.

So, as a result, so you will take the modulus standardized then you know go for the estimation and again it follows the LM test and it is more or less same. So, you calculate with you know n multiplied by coefficient determination then connect with the chi square tabulated below and then look for whether you know rejecting the true null hypothesis where it was the formulation of h 0 is that there is no heteroskedasticity and the moment you are rejecting then there is a heteroskedasticity. So, that is what the kind of you know check process about the LM test.

So, likewise you can actually look for you know further processing. Here also this hg LM test that is so again more or less same; so what we have done earlier. So, one case its u square of the error term u square and this is the modulus of the error terms to avoid the positive and negative signs.

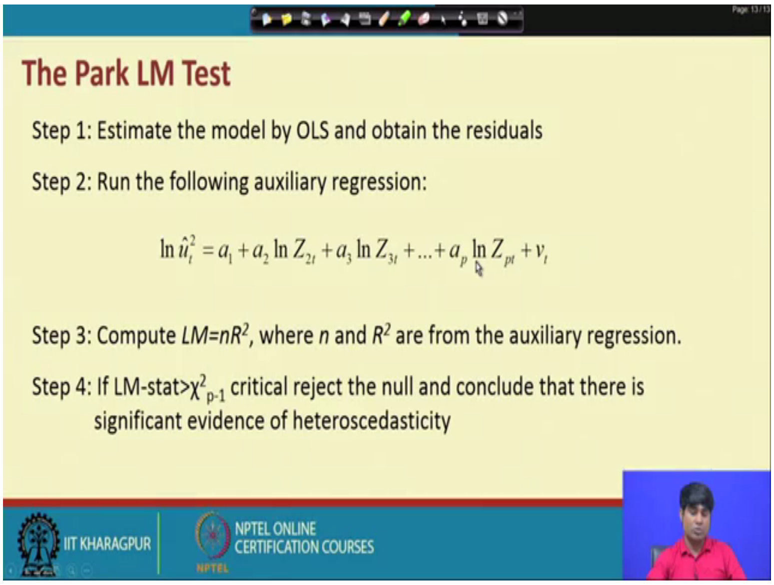
(Refer Slide Time: 13:29)



And again here we are going for you know square of the error term and then we are taking the log transformations. Sometimes you need to go for a log transformation and by default the level of you know volatility or you know the kind of you know heteroskedasticity will be minimized, and your model accuracy will be very high and this may give you this digital signal about the heteroskedasticity and homoscedasticity.

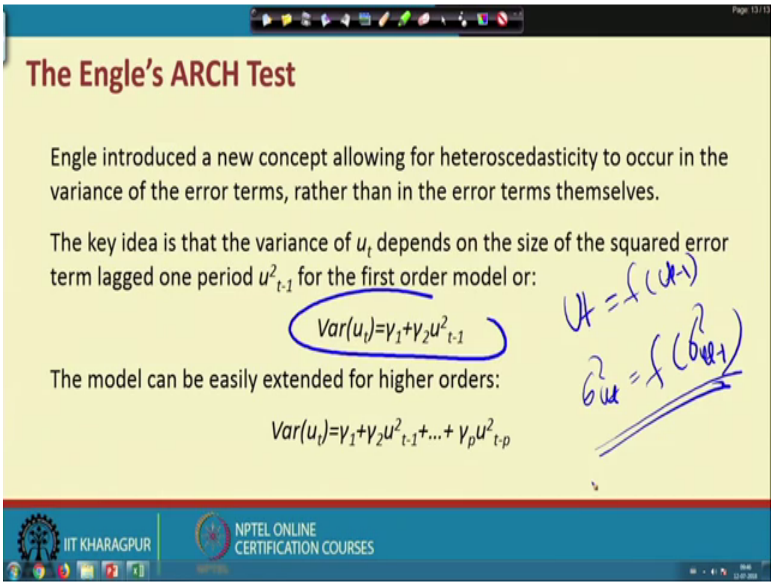
So, by default here also same you know like in last case I have mentioned that you know they are more or less same. So, again it is a LM test case and this gives the signal that you know you know the process through which you can you know detect the heteroskedasticity.

(Refer Slide Time: 14:11)



So, like you know every test is a actually a similar kind of you know structure and this is a park test here actually again we are connecting a log u square with you know other independent variables and again it follows the Lagrangian multiplier test to check the significance level.

(Refer Slide Time: 14:27)



And there is also spearman rank correlation test where what you can do we observe the error term and then we use actually called as you know rank correlation mechanisms. So, we first you know take the modulus and then rank the error terms, and then we can compare with the any of the independent variable and then rank the rank the data of the X variables, then connect the rank of the error terms and rank of the independent variable and calculate the rank correlation coefficient.

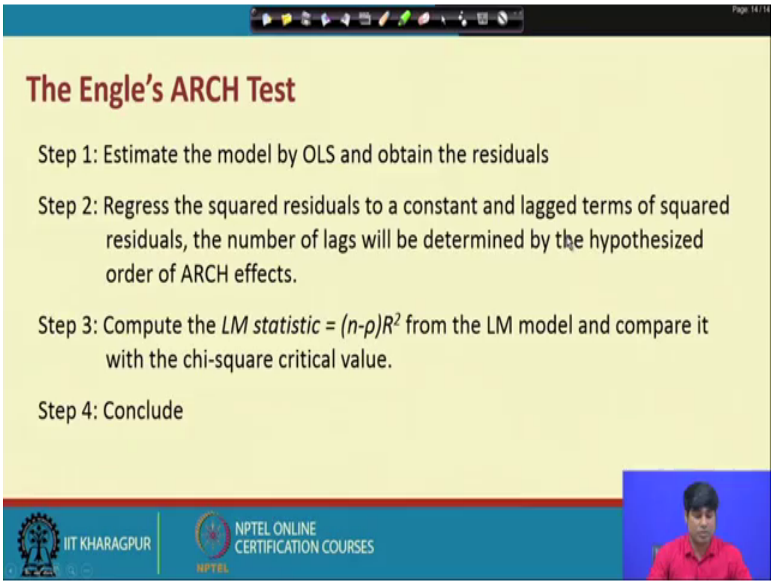
And that rank correlation coefficient need to be tested to see whether there is a presence of heteroskedasticity or absence of you know heteroskedasticity. And now this is much better test called as an Engle's arch test autoregressive conditional heteroskedasticity test. Simply variance of u t here is actually depends of you know u square t minus 1's, like you know completely you know like autocorrelation, where we are connecting u t upon u t minus 1 means like in the case of autocorrelation we are connecting u t as a function of you know u t minus 1.

So, here we are connecting sigma square u t is equal to constants of sigma square u t minus 1 that is the case actually and a if you find there is a kind of you know relationship. Then by default the declaration is that you know there is a heteroskedasticity, if there is you know such you know kind of you know significance in the systems then the question is that you know it is a case of you know heteroskedasticity and homoscedasticity.

So that means, technically it is the game of the error term. Sometimes we are connecting error term to error terms we are connecting error term with the independent variables, dependent variables then variance of the error term with its cellular component like you know this case. Then finally, a you know we want actually whether the particular model is if free from this particular virus that is what the heteroskedasticity, if not then the simple the requirement or you know good requirement is that there is a presence of you know heteroskedasticity that is what the a need of you know equal variance.

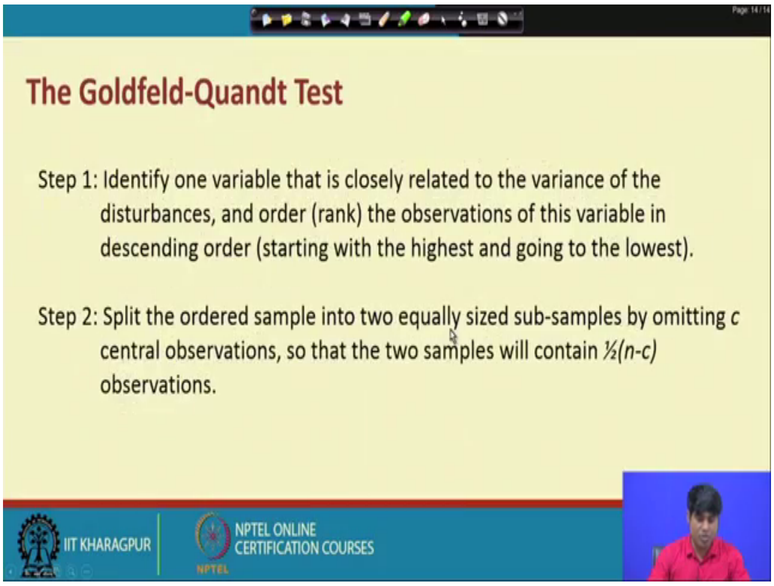
So, they are very close to each other and there is no you know kind of you know spread at all or you know very minimal levels of this spread that is what called as you know closely equal variance over the time and over the cross sectional unit.

(Refer Slide Time: 17:08)



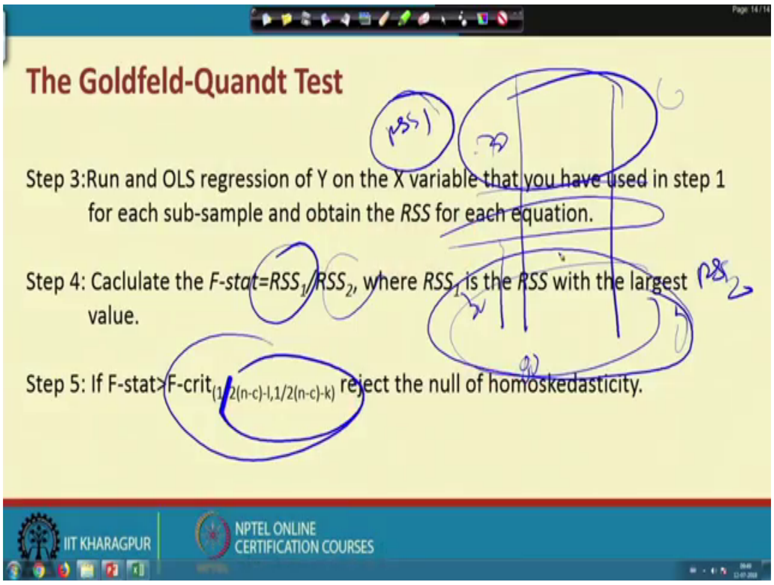
So, now the case is like this. Similarly, this step by step process again same LM test can be used and check it.

(Refer Slide Time: 17:14)



So, this is another way of you know GQ test to check the particular process.

(Refer Slide Time: 17:17)



Here what is happening? Then after getting the error terms you can divide the level of you know sampling into two parts and in betweens you can remove few of these samples like you know here the case like critical value which I have mentioned here. So, and this is what actually.

So, you know c means actually the item which you are actually removing in between process. For instance let us say this is your data set and this is the data set and what we can do we can put in a kind of you know ascending to descending order with respect to one particular variable by default other data will be adjusted. And then you divide the data into two parts let us say it is a 80 data points then we can take actually first 30 data points and last 30 data points, and remove the 20 data points and you can build the model separately here and we can build the model separately here.

Then in this models we will have RSS 1 that is what here and in this case we feed the models and you get the RSS 2, ok. So, there is you know means it is not necessarily that you know this sample and this sample will be uniform it can be have a different sample, but how many points you can actually omit that you have to declare properly.

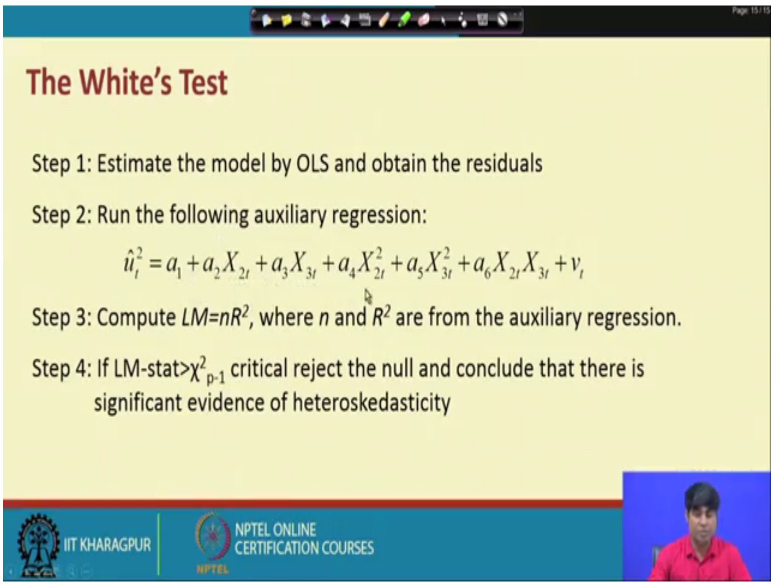
So, as a result you know in this first you know clustering you will get residual sum of square and then we declare RSS 1 and here you will get residual sum of square we declare RSS 2. And then we will create a F statistic which is the ratio between RSS 1 by

RSS 2 and then connect with you know F criticals with the adjustment of the you know removal of this you know data points.

And, if the calculator will overtake the tabulated then the declaration is that you know there is a you know heteroskedasticity and if not then there is a kind of you know homoscedasticity. This again very interesting test which you can actually take and you know validate the kind of you know results. So that means, there are lots of formal tests are there to check the heteroskedasticity that means, from the types of you know test you can just guess that you know this could be the you know very dangerous virus, while using any kind of you know models with heteroskedasticity for any kind of you know decision making process. Otherwise there may not be you know lots and lots of you know informal and formal tests are there to check the you know level of heteroskedasticity or to declare the kind of you know heteroskedasticity.

So, obviously, this is a problem case, we like to check, we like to detect, we like to you know quantify, then check the statistical levels and then finally, if the requirement means, if there is any kind of you know requirement. Then you can restructure the entire process and then a transfer or have the model in such a way that even that model can give you know good results and that too as per the kind of you know decision making requirement or any kind of an engineering requirement. So, likewise, so this is another test, White test is also there.

(Refer Slide Time: 20:37)



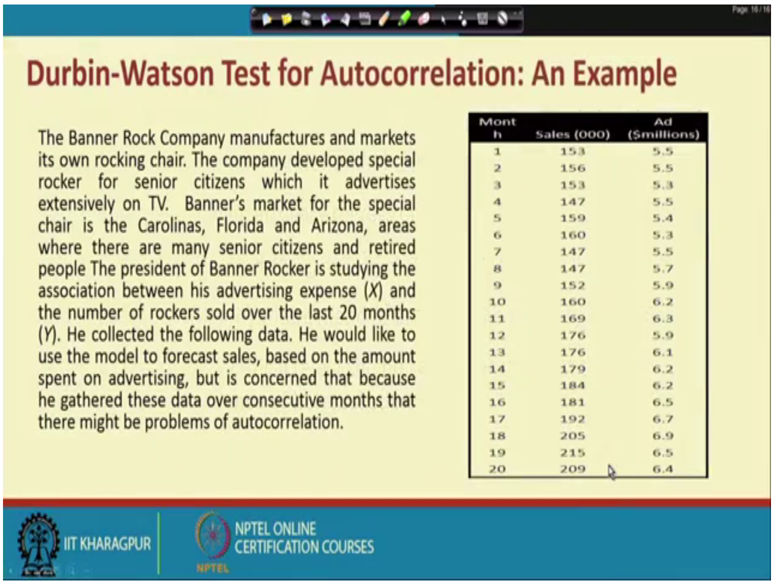
This is another kind of you know test where you know we are using actually a non-linear structures compared to all these tests where we use actually linear structure by just connecting you with you know other variables, but here the interesting part of this you know white test is there. So, what we can do here you see here we are just taking the square of the error term then there are two independent variables X 2 and X 3. And we are again a you know including the kind of you know what we can call as you know dot product that is square of this particular cost variables, square of this second variables. And then they are cross product.

So that means, X square 2 and X square 3 and then X 2, X 3. So, (Refer Time: 21:31) these 3 this where the kind of you know structure. So, in this case we are using the non-linearity set up to check the particular you know issue that to declaration of the homoscedasticity or you know heteroskedasticity. And, compared to other models this is more interesting because you know other models we use actually linearity structure and in this model we are using actually non-linearity structure.

But by the way the overall conclusion is that you know whether there is a kind of you know heteroskedasticity if not what is the level of heteroskedasticity and whether it is in at the level of tolerance or against the tolerance. So, really check and accordingly we can restructure as per the particular requirement and as per the decision making requirement.

So, likewise really you know use a different test to not the level of heteroskedasticity and the kind of you know solutions as per you know the particular you know engineering problem requirement. So, what we can do here is whatever test we have discussed till now to check this particular heteroskedasticity and the kind of you know modeling requirement, I can connect here with a particular you know problem which we have already discussed in the case of you know autocorrelation.

(Refer Slide Time: 22:53)

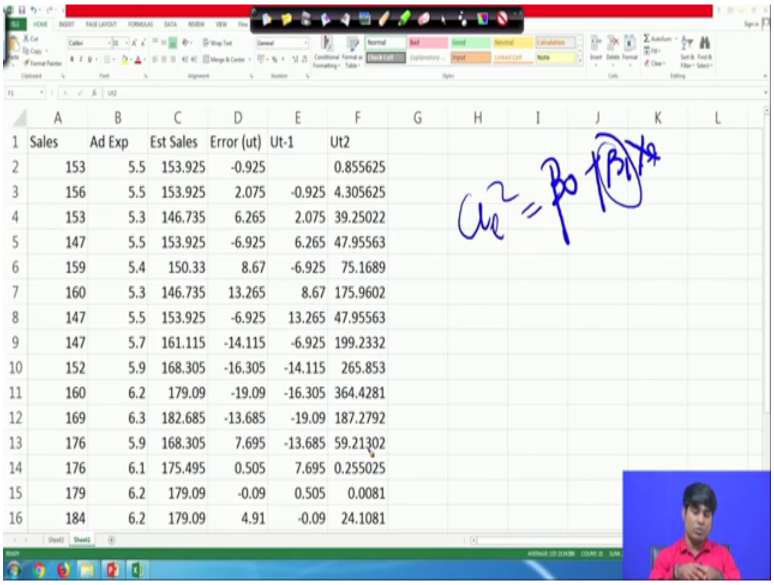


Month	Sales (000)	Ad ($millions)
1	153	5.5
2	156	5.5
3	153	5.3
4	147	5.5
5	159	5.4
6	160	5.3
7	147	5.5
8	147	5.7
9	152	5.9
10	160	6.2
11	169	6.3
12	176	5.9
13	176	6.1
14	179	6.2
15	184	6.2
16	181	6.5
17	192	6.7
18	205	6.9
19	215	6.5
20	209	6.4

The same problem here, what you can do here is we can connect with you know cells and you know advertising expenditures and first we estimate the model and have the error terms and after getting the error terms. So, you can create you know log of the error terms and variance of this error term.

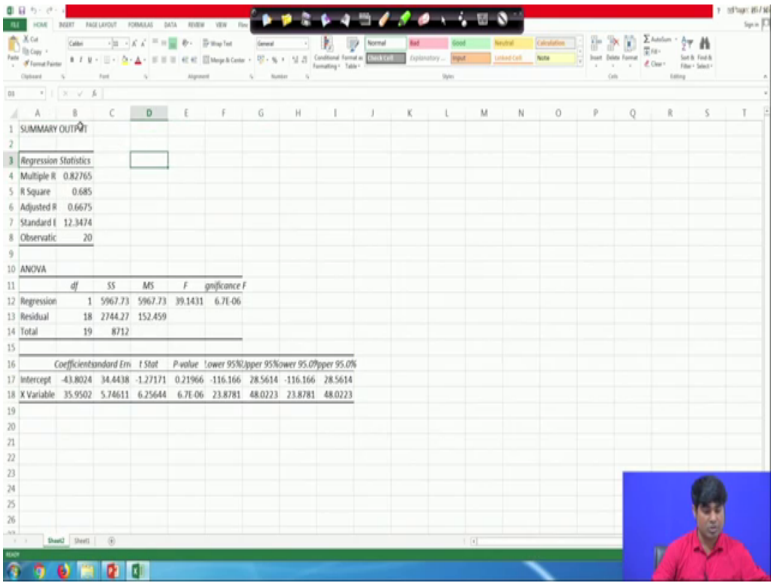
So, when you are connecting error term with log error terms then this that will give you the signal about the autocorrelation. Now, what is actually happening here we are just squaring the error term and checking the behavior of that particular error terms, and if the error term behavior is actually coming actually closer you know convergent to each other of a different point of time of a different cross sectional unit. Then the declaration is that you know there is a homoscedasticity if not then the declaration is actually heteroskedasticity.

(Refer Slide Time: 23:50)



To simplify so, I will connect the you know last class problems. And this were the actually first hand data inputs and dependent variables, independent variables and we have here estimated results.

(Refer Slide Time: 24:01)



And this were the estimated results. And the model is actually very goods which you have already discussed yesterday that means, the specification test you know gives the green signal, then the gf t is also giving good signal. And what is we are actually checking here the diagnostic part and that to the previous case we have seen there is a kind of you know autocorrelation and to do that. So, what you will do? We just derived the coefficients and then we take to this you know data set and using these coefficients we will have the estimated equation. And after putting this estimated equation you can actually just scroll it and then get the kind of you know estimated you know data set.

Now, we have actual cells and we have estimated cells and the difference will give you the error components what we have used earlier. So, with the error terms we create a log error term and then trying to check whether there is a relationship between u t and U t minus 1. Now, in this particular case particularly you know heteroskedasticity and homoscedasticity case, so we just create you know here is the U t squares and then the only thing is that you know this need to be actually simply square of these terms or you can simply take the modulus and again same. So, what you can do? Again you can create a spreadsheet here. So, just you just use do it ok. So, this will be give you,. So, they just this will give you actually you U t squares and just scroll it.

So, you will get actually U t square datasets. So, now, with respect to autocorrelation we are connecting u t with you know created U t minus 1, now in this heteroskedasticity we are connecting u t then we are deriving u t squares now the connection will be U t square with any of the independent variable let us say advertising expenditures.

So, now, we can connect advertising expenditure with error terms and then check whether the you know particular parameters are statistically significant that means, technically. So, what will you do here? So, this is a square of the error terms. So, our square of the error terms that is U t squares should be connected with the a you know advertising expenditure that is the X variable. So, the model will be beta 0 plus beta 1 X t. So, where X t represents the advertising expenditure now, or the null hypothesis that you know beta ones should not be should be statistically significant. If that is the case then we can reject the null hypothesis and declaration is that you know the heteroskedasticity problem.

So, likewise you know different test can be applied sometimes we take the modulus of you know error terms then connect with the independent variables, then sometimes you can take the square of the error terms then connect with independent variable, sometimes you can use the square of the error term then create the log of the error terms. For instance, u t minus 1 you can create here yeah u t square t minus 1s and then you can connect with independent variable and check the significance levels. If that is the case then there is a presence of heteroskedasticity and if heteroskedasticity is there. So, we simply cannot use this model for the decision making process.

So, against there is a need of you know restructure and restructuring of the entire process. And you continue like you know iterative process till you get the model which is free from all these virus and that to specifically the heteroskedasticity. And with this we will stop here.

Thank you very much. Have a nice day.