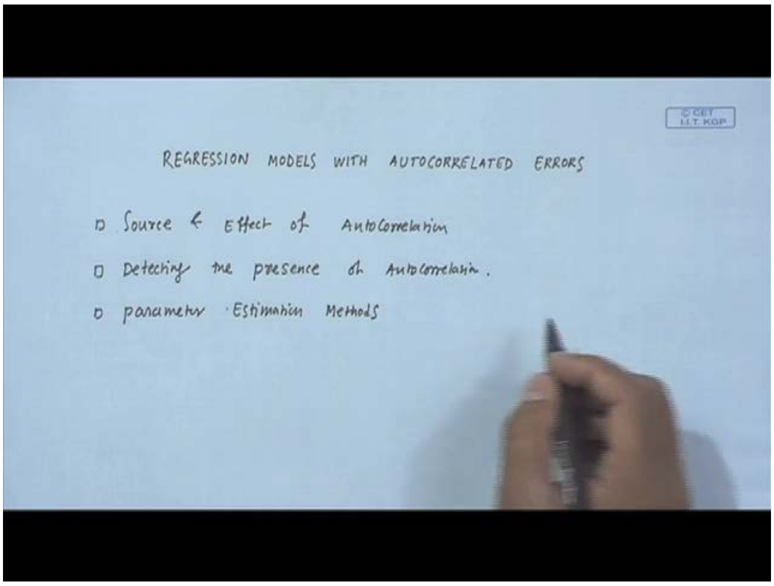
Regression Analysis
Prof. Soumen Maity
Department of Mathematics
Indian Institute of Technology, Kharagpur

Lecture - 34
Regression Models with Autocorrelated Errors (contd.)

Hi, this is my second lecture on regression models with auto correlated errors, and here is the contained of this topic.

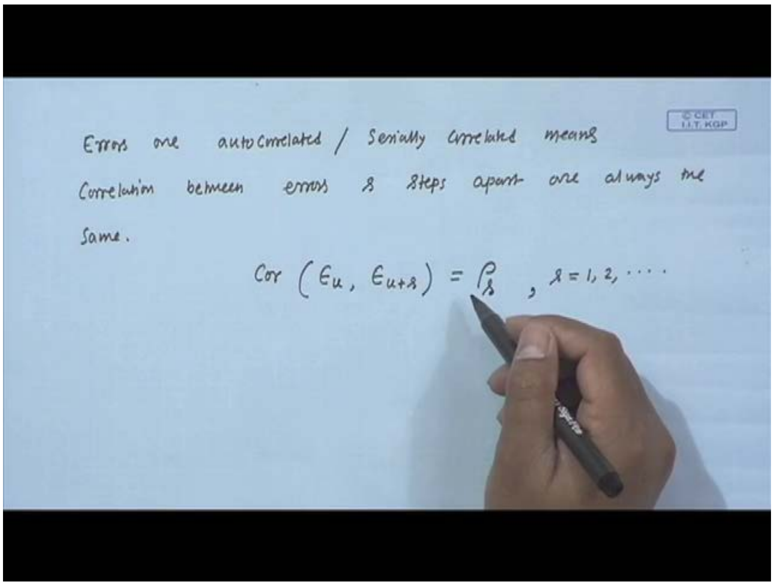
(Refer Slide Time: 00:37)



We already talked about sources and effect of autocorrelation in the regression model. And, in the previous class we started talking about how to detect presence of autocorrelation, and we will be talking about parameter estimation in the presence of autocorrelation model. Let me repeat the objective of this topic, is that in simple linear regression model or in the multiple linear regression model we make several assumption on the errors terms like expectation of epsilon is equal to 0, variants of epsilon is equal to sigma square and the error terms are correlated. And, also we make assumption on the normality of the error terms that epsilon i follows normal distribution with parameter 0 and sigma square.

These are the assumption that we make while fitting a simple linear regression model or multiple linear regression model using least square technique. But, the problem is that you know when the data are collected sequentially in time, the assumptions of independence of error term may not be true. That means while you are collecting the observation sequentially in time, the observations might not be independent. Which implies that the error terms epsilon i they are not independent, and there exist some sort of you know autocorrelation between the errors. So, let me write down the formal definition of autocorrelation that you know already.

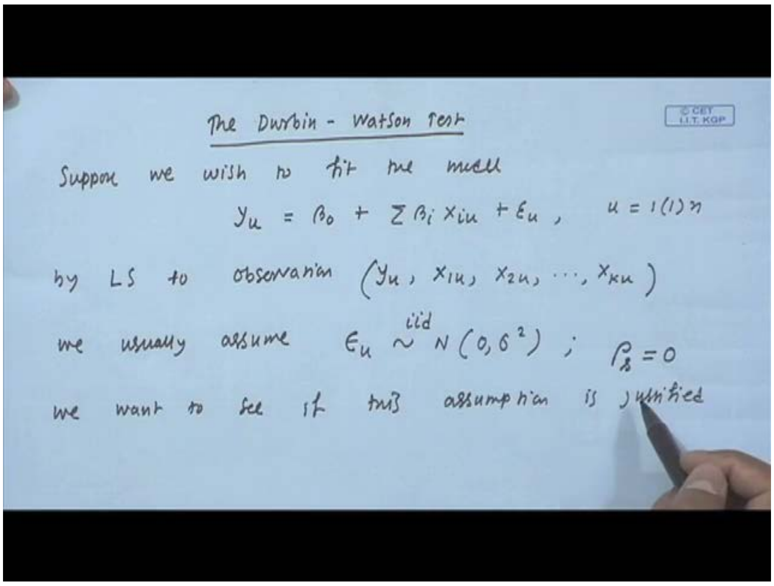
(Refer Slide Time: 03:03)



Errors are autocorrelated or some time we call serially correlated means correlation between errors s steps apart are always the same. That means, the correlation between epsilon u and epsilon u plus s is equal to, we denote this by, I mean this is not zero, and we denote this by rho s and this is for s equal to 1, 2 like this. This is what we mean by the autocorrelation in the error term, the error are correlated or serially correlated means correlation between errors s step apart are always the same.

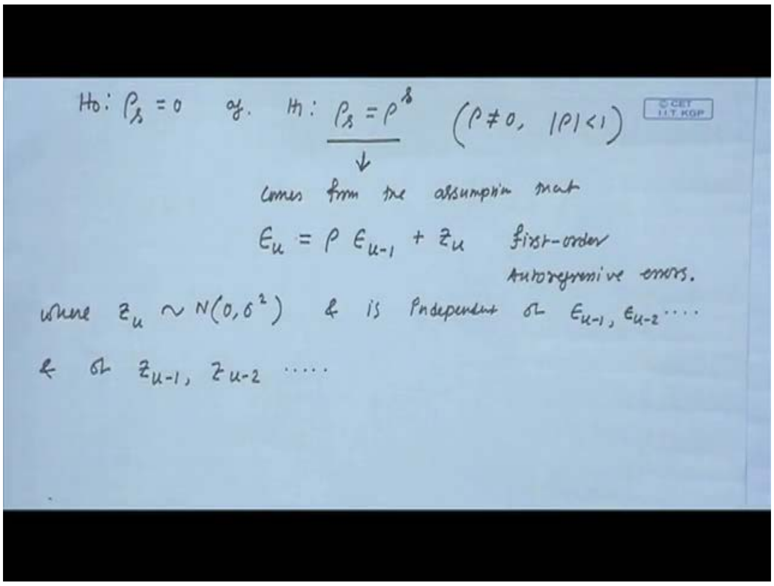
So, we talked about the source of this autocorrelation and the effect of autocorrelation in the previous class. And you are talking about the how to detect the presence of autocorrelation, once you have given a time data, we suspect that there exist autocorrelation but, you have to test whether autocorrelation exist or not. That means whether the error terms are correlated or not, for that we talked about statistical test called Durbin Watson test. We could not finish that in the last class so we sort of repeat that thing today again. So, here is the slide from my previous class, the Durbin Watson test.

(Refer Slide Time: 05:58)



So, you want to feat model. This is a multiple linear regression model by using the least square technique who the observation this. This could be a time series data that means the observations are taken sequentially in time. So, what we usually assume is that we assume that this epsilon the error term follows normal distribution with 0 sigma square, that means we assume that all the correlation rho s is the correlation between errors s step apart, that is zero that is what we assume. Now, here what we want to test is that we want to test whether this assumption is justified or not.

(Refer Slide Time: 06:46)

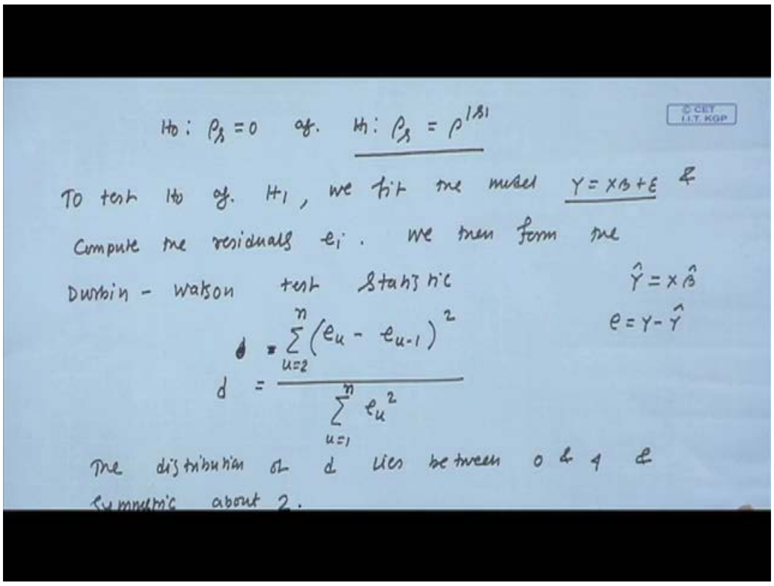


Here, is the hypothesis to test that, we test the hypothesis now, hypothesis that rho s is equal to 0 against the alternative hypothesis rho is equal to rho to the power s. Basically, you want to test whether this is equal to 0 against whether rho s is greater than 0 or less than 0 or not equal to 0. Greater than 0 means positive autocorrelation, rho s less than 0 means negative autocorrelation and not equal to 0 means autocorrelation exist.

Now, why this particular form rho to the power s that I explain this comes from the assumptions that the errors are having, I mean this is the error of first order autoregressive error. That means you can regress epsilon u 1 epsilon u minus 1. That is for a first order autoregressive error.

We assume that this is true for the errors terms and where z u is again the error term for this regression model which follows normal 0 sigma square, and z u is independent of epsilon u minus 1 epsilon u minus 2 of all the previous terms and it is independent of z u minus 1 z u minus 2. And I explained you know how to get this form in the previous class. Let’s refer my previous lecture for this one this is coming from the assumption of first order auto regressive assumption.

(Refer Slide Time: 08:54)

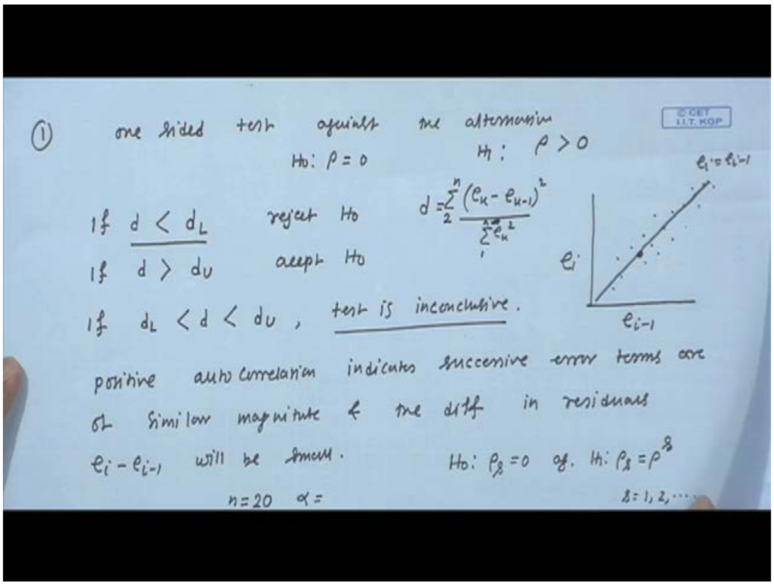


Now, to test this hypothesis we are trying to test the hypothesis whether the correlation between the errors which are a step apart is equal to zero or that is equal to rho to the power of s. To test this hypothesis what we do is that we consider this Durbin Watson test statistic.

Now, this involves the difference of residuals, successive residuals as you can see here. But, how to get them here is that you first fit regression model using ordinary least square technique assuming that all the assumptions are true on error term. And then compute the residual once once you fit this model, what you get is that you get y hat is equal to x beta hat. So, you have the fitted model once you have the fitted model you can compute e which is equal to y minus y hat, this is the observed and this is the estimated response and so you have all the residuals and then you form this base statistic.

And it is known that this distribution of d lies between 0 and 4, and it is symmetric about 2. Now, we are trying to test this hypothesis based on this, and the test statistic to test this hypothesis is this one, suggested by Durbin and Watson. Now, let me talk about the critical region you know how to decide whether to accept or reject this reasonal hypothesis based on this d value.

(Refer Slide Time: 11:05)



So, here is the first case. One sided test against the alternative, see we started with the hypothesis like h naught that rho s is equal to 0 against h 1, that rho s is equal to rho to the power of s. Now, we are testing the hypothesis which is rho equal to 0 against rho is greater than 0. Suppose this is true then rho s is also greater than 0, that means for all s, s equal to from s is 1 2 anything. So, testing this hypothesis is same are testing this hypothesis. Once rho is equal to 0 if null hypothesis is 0 then rho s is going to the 0. If the alternative hypothesis is true that means rho is greater than 0 then the original hypothesis says that rho s is greater than 0. That means the data has no positive autocorrelation.

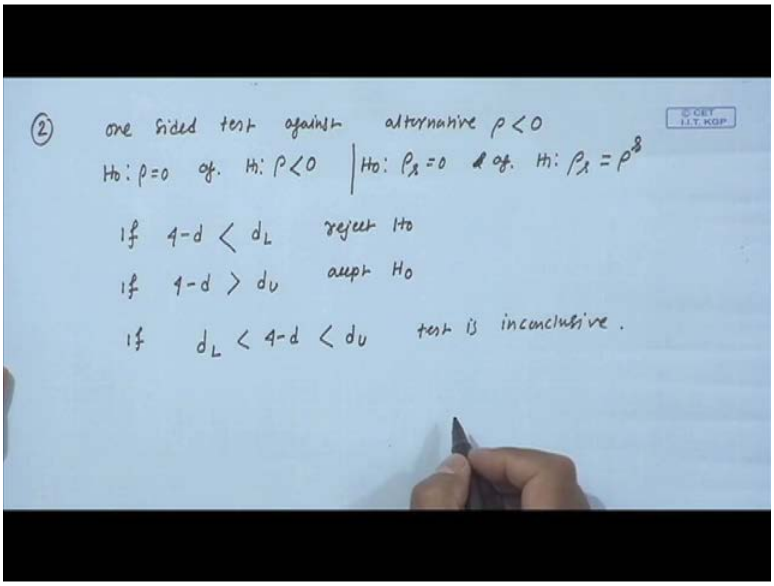
Well, so you have the Durbin Watson test statistic value d and if d is less than d L then it says that you reject h naught, d is greater than d u then you accept h naught and if it is between d L and d u the test is inconclusive. Now, let me talk little bit about what is this d L, this d lower value and d upper value. There is a table for d, d table. There you will get this d L and d u value for different n depending on how many observations are there. So, for different n and for different alpha the label of significance. Say for example, n equal to 20 in our previous example on soft drink concentration sale, you will get the d L value based on the different choices of alpha.

There exist a table for this d low and d up value. Well, if the observed d value is in between d L and d u, then the test is inconclusive. I talked about the significance of why we reject the null hypothesis h naught when the d value is small. So, the d value you know what is d right, d is e u minus e u minus 1 square by e u square 1 to n minus 1 perhaps and this is 2 to n, no this is till n. If d value is small we reject the null hypothesis that means we will accept all alternative hypothesis, that means we say there exist a positive autocorrelation.

When d is small there exist positive autocorrelation. Now, if you can recall the scatter plot for d i and against d i minus 1, so if you see the scatter plot like this we say there exist positive autocorrelation. That means e i increases with e i minus 1, so the ith observation depends on i minus 1 ith observations. There is a correlation between the successive observations.

So, if the data are centered about this line e i equal to e i minus 1, that means the successive error terms are of similar magnitude they are almost same. Here, it says the positive autocorrelation indicates successive error terms are of similar magnitude that means they are almost same. So, the difference in residual this difference will be small. That is why you know, I mean the small value of d indicates the existence of positive autocorrelation. I hope this is clear.

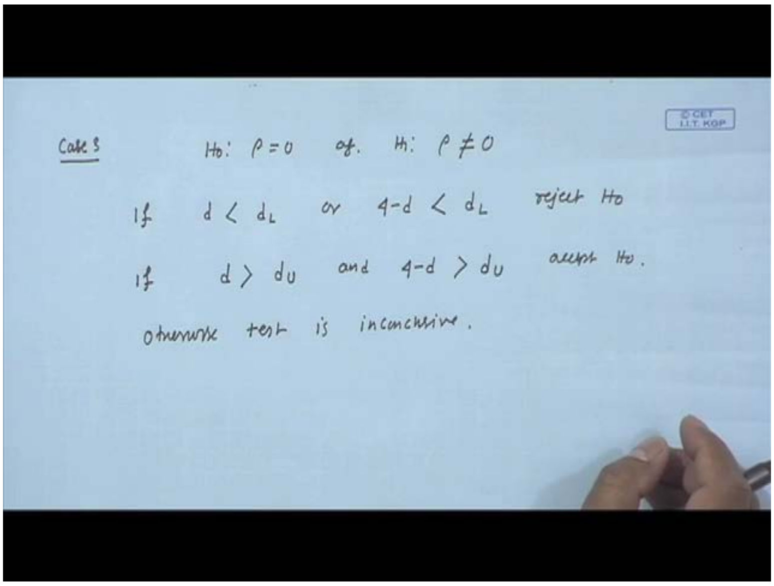
(Refer Slide Time: 16:48)



Now, let me go for the second case. One sided test against alternative rho less than 0. The meaning of this one is that original we stared with the hypothesis that rho s is equal to 0 against h 1 that rho s is equal to rho to the power of s. Now, if rho is negative then this rho s is going to be negative that means the alternative hypothesis says that there exist negative autocorrelation.

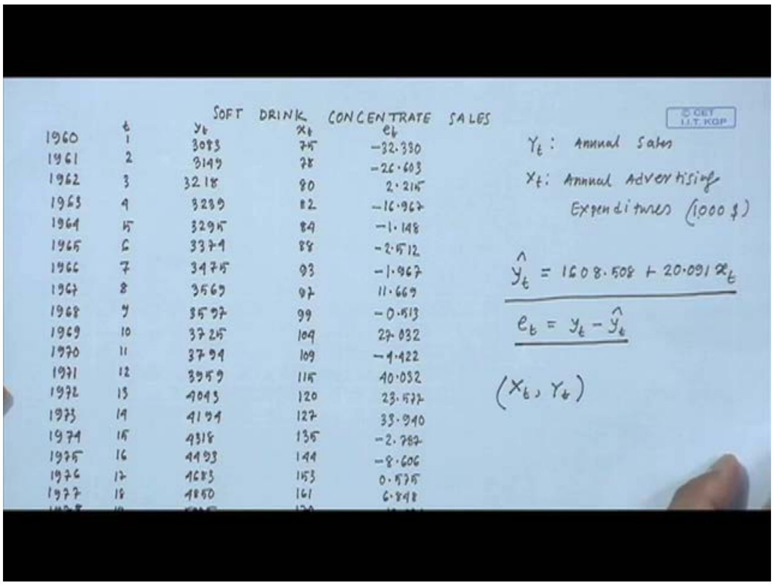
And this can be tested by testing the same statistic d. So, here it says that if 4 minus d is less than d L, you reject h naught. If 4 minus d is greater than d u, you accept h naught. And if 4 minus d value is in between d u and d L the test is inconclusive. So, similar argument. Here, basically we want to test hypothesis and finally, testing this hypothesis is same as testing rho equal to 0 against h 1 rho is less than 0. And the final case, the third case.

(Refer Slide Time: 19:31)



Case three, here we test h naught rho equal to 0 against the alternative hypothesis that rho is not equal to 0. So, it is a two sided alternative. And, here if d is less than d L or 4 minus d is less than d L we reject h naught so small value of d indicates that there exist a autocorrelation. And if d is greater than d u and 4 minus d is greater than d u, then you accept h naught, so the high value of d indicates that there is no autocorrelation in the error and otherwise the test is inconclusive. Now, let me consider an example.

(Refer Slide Time: 21:15)

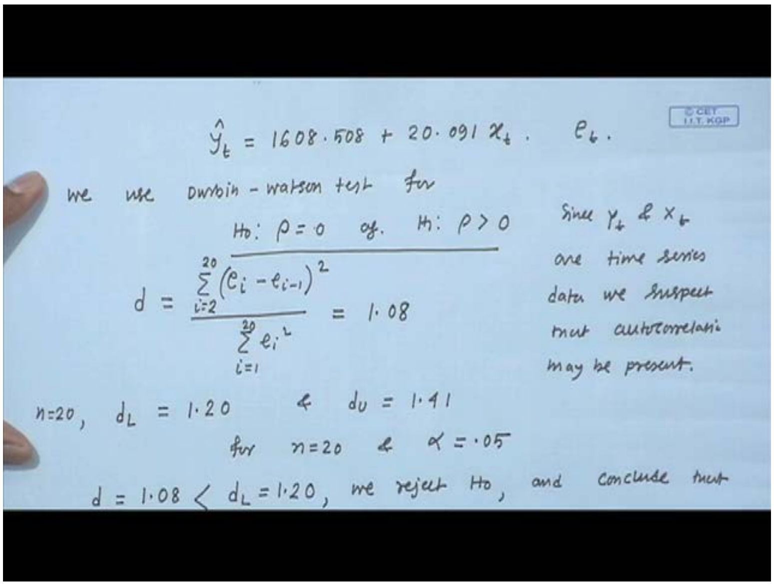


This is the example I took in the previous class also, this is called soft drink concentrate sales. Here, we have two one regressor variable that is annual advertising expenditure and y t is annual sales. This is the regressor variable sorry x t is the regressor variable and response variable is y t, so we have the data x t y t sequentially in time. So, we have the data for 20 years starting from say nineteen hundred sixty to nineteen hundred seventy nine.

That is of course, I mean this is what we call the time series data such data are called time series data. Initially, you know ignoring that whether the basic assumption while fitting a straight line model between y t and between y and x using the ordinary least square technique you forget about you ignore whether assumptions are true or not, you just fit a model between y and x.

This is the fitted straight line model between y and x. Once you have the fitted model you can compute the residuals, so e t is nothing but, the observed response value and the estimated respond value at e at say t. So, once you have this residuals now you might be interested to test whether because, since it is a time series data we suspect that you know there might be autocorrelation present in this data. Now we can go for Durbin Watson test to test whether autocorrelation exist or not.

(Refer Slide Time: 23:53)



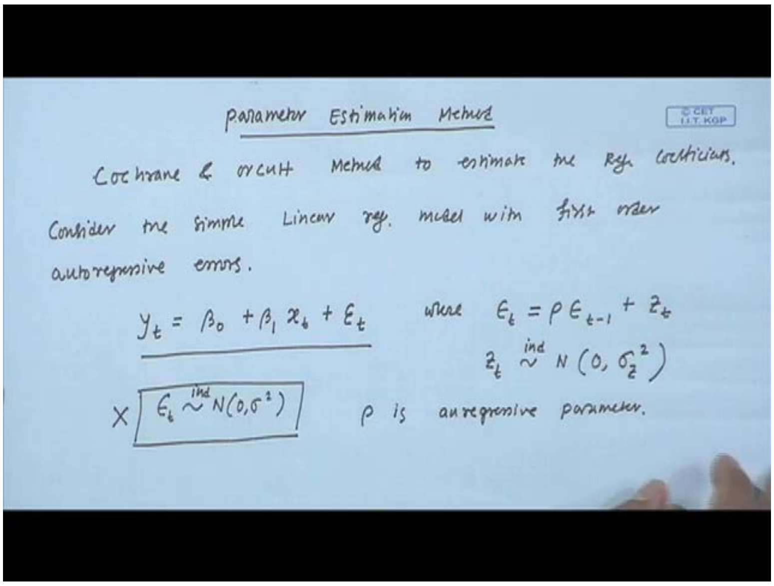
We have fitted this model say y hat y t is equal to 1608.508 plus 20.091 x t and then we can compute the residuals. Now, we use Durbin Watson test for this testing, that rho is equal to 0 against say h 1 rho greater than 0. Why we are doing this, because since the response variable y t and the regressor variable x t are time series data, they are taken over time series data so we suspect that autocorrelation may be present. If it is not a time series data, then we do not go for autocorrelation test. So, this is that hypothesis you want to test and we have the residuals. So, we computed the residuals here. Here you have the residuals.

And, then you compute the Durbin Watson test statistic d which is equal to e i minus e i minus 1 or e t minus e t minus 1 square i is from 2 to 20 by e i square i is from 1 to 20. So, you can check that this one is 1.08. Now, you have to find d L value from the table for n equal to 20, because we have 20 observations here, so we can check from the table that d L is equal to 1.20 and d u is equal to 1.41 for n equal to 20 and the level of significance alpha is equal to 0.05. So, what you see that the observed value this d which is equal to 1.08 is less than d L which is equal to1.20.

So, small d value indicates that there exist positive autocorrelation. Since, this is true we reject h naught, h naught says there is no autocorrelation. So reject h naught and conclude that the errors are positively auto correlated, this is one example to illustrate the Durbin Watson test. Now, given a time series data we suspect that there may exist autocorrelation, so what do we do is that we fit a simple straight line model using the ordinary least square technique.

And once you have the fitted model we find the residuals and using those residuals we compute the Durbin Watson test statistics and see whether autocorrelation is present in the data or not. Suppose the result of Durbin Watson test is that autocorrelation is present in the time series data you are given. Then the next issue is how to estimate the regression coefficients in the presence of autocorrelation, so we will talk about that now, the parameter estimation method.

(Refer Slide Time: 29:56)

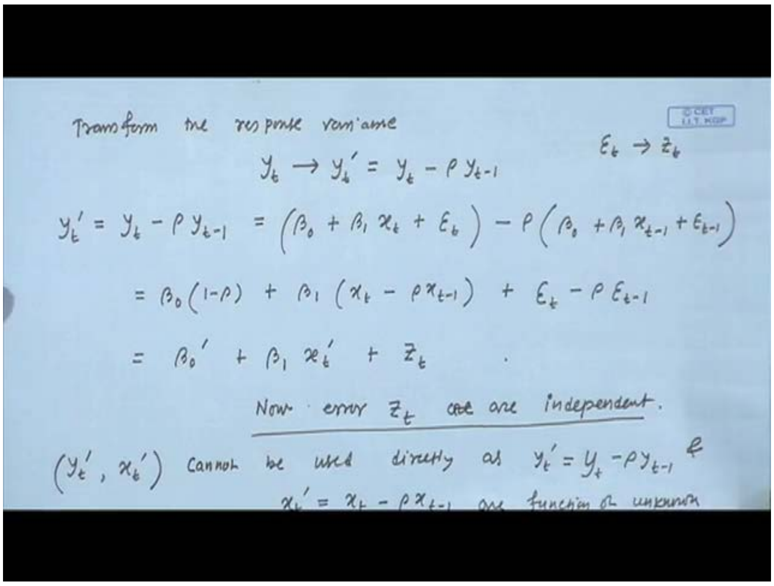


So, in the presence of parameter estimation method, in the presence of autocorrelation, in error, this is called Cochrane and Orcutt method to estimate the regression coefficients. It says that, considered the simple linear regression model with first order auto regressive error. That means, we are considering a model simple linear equation model y t is equal to beta naught plus beta 1 x t plus epsilon t. But, here epsilon t are not independent they are first order autoregressive error.

That means where epsilon t can be regress on epsilon t minus 1, so epsilon t is equal to rho epsilon t minus 1 plus z t. And, this z t is normal distributed with mean 0 and variants sigma z, it is a constant variants. This is just to distinguish from sigma square and this equal sigma z square and they are independent. And here this rho is called this rho is autoregressive parameter or autocorrelation parameter might be.

Now, how to fit this model because here you know you cannot apply ordinary least square technique, because the assumption on epsilon t that is this follows normal 0 sigma square with independent, this is not true, this is not true here. So, we cannot apply ordinary least square technique here. So, what we do is that we will transform this data y t to say y t dash.

(Refer Slide Time: 33:53)



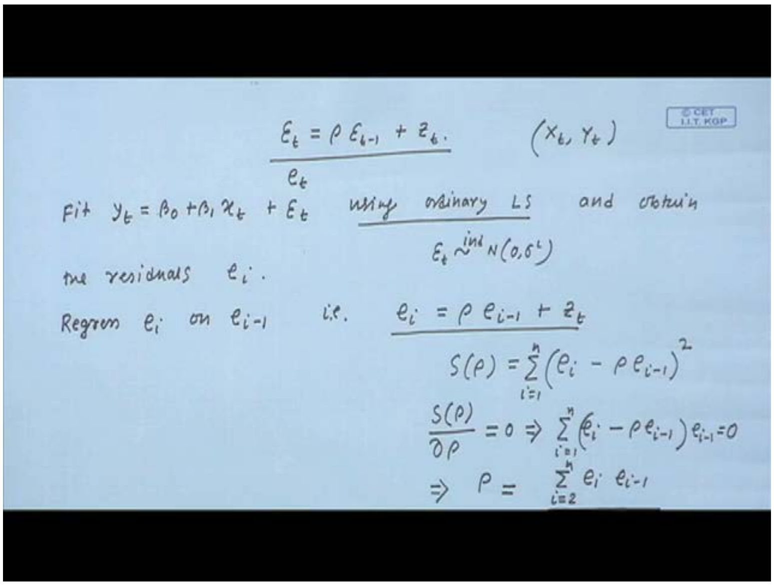
We transform the response variable y t to y t dash, which is equal to y t minus rho y t minus 1. Then let me check, what is this y t prime now. So, y t prime is equal to y t minus rho y t minus 1, which I can write, I know that y t is equal to beta naught plus beta 1 x t plus epsilon t, and this minus rho. What is y t minus 1 this is beta naught plus beta 1 x t minus 1 plus epsilon t minus 1.

So, this can be written as, beta naught into 1 minus rho plus beta 1 into x t minus rho x t minus 1 plus epsilon t minus rho epsilon t minus 1. Now, I can write this as beta naught dash plus beta 1 x t dash plus z t. If you can recall, we assume that this errors are first order autoregressive error. That means, epsilon t minus rho epsilon t minus 1 is equal to z t, where z t are independent with mean zero and variants sigma square z.

Now, we have transformed the error term epsilon t to z t, where z t now this transform error, now error z t are independent. Now, you can apply the ordinary least square technique to this transform data. But, the problem here is that this y t prime and x t prime this transform time series data cannot be used directly as these two things, this y t dash which is equal to y t minus rho y t minus 1. This involves an unknown parameter rho, we do not know the value of rho right.

And x t prime which is again x t minus rho x t minus 1, this two are function of unknown parameter rho. We cannot take this transformation right now but, let us see how to compute this unknown parameter, how to estimate this unknown parameter rho. So, this rho is called autocorrelation parameter or autoregressive parameter.

(Refer Slide Time: 39:08)



If you can recall, that this rho is basically epsilon t is equal to rho epsilon t minus 1 plus z t. Now, one way to compute or estimate this rho is that, we are given only the data x t and y t nothing else and it is known that they are time series data. So, what you have to do is that, you just fit simple linear regression model on x t y t, you compute residuals. And, then we know that the residuals are sort of observed e t, e t is observed value of epsilon t and then we can regress e t on e t minus 1 and from there we can compute the value of rho. So, let me explain that part now, first what you do is that you fit y t equal to beta naught plus beta 1 x t plus epsilon t, using ordinary least square technique.

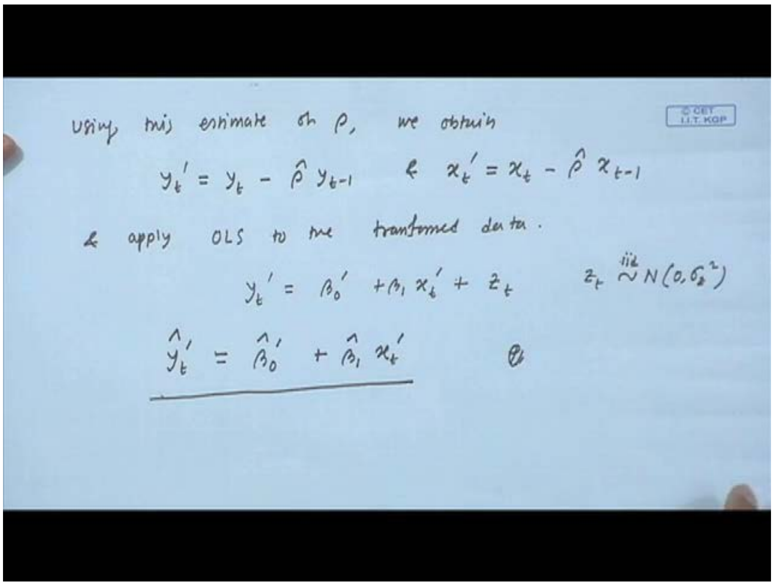
Using ordinary least square technique means you will assuming that this assumptions are true or you are ignoring all this assumption at this moment. So, you just fit the simple linear regression model between x t and y t and obtain the residuals e i. Once you have the residuals, then what you do is that you regress e i on e i minus 1, that is you fit a model like e i is equal to rho e i minus 1 plus z t. See, we do not know this epsilon t right and ith residual is sort of estimate of ith error term.

So, we are regressing e i on e i minus 1, so you fit this model. We know all this residuals. So, what is this rho value now, how to get estimate for rho. That can be obtained by minimizing this quantities so we will go for the least square estimate. We will minimize this quantities s rho, s rho is say e i minus rho e i minus 1 square and we estimate rho in such a way that this is minimum, which essentially says that you differentiate this s rho with respect to rho this equal to 0 implies that summation e i minus rho e i minus 1 into e i minus 1 is equal to 0.

I am differentiating with respect to rho and this gives me rho is from i is from 1 to n right. So, rho is equal to summation e i e i minus 1 i is from 2 to n now I have to take because of i minus 1 by summation e i minus 1 square i is from 2 to n and this is the estimated value of rho.

And, this can be written as finally, I can write so the least square estimate of rho which is, let me call it rho hat which is equal to e t e t minus 1 t is from 2 to n by summation e from 2 to n e i minus 1, I can replace this as e t square t is from 1 to n. So, this is how we estimate rho and now we use this rho to transform the data.

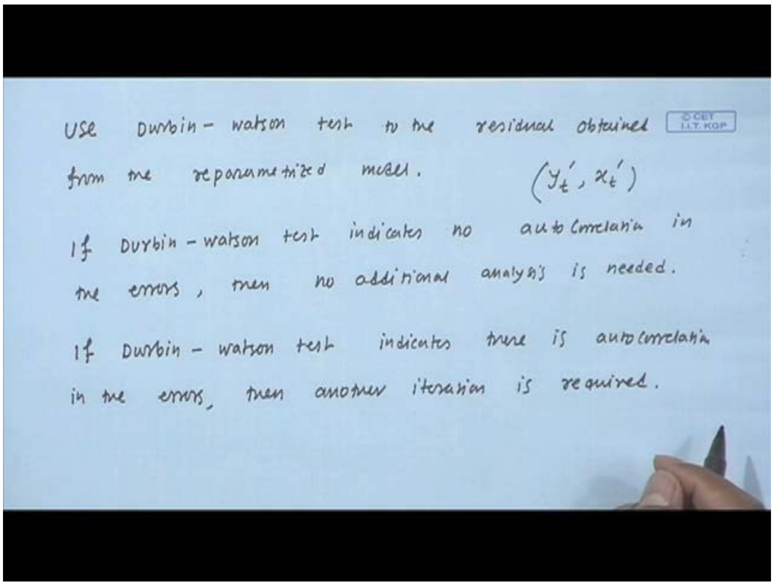
(Refer Slide Time: 45:06)



So, using this estimate of rho, we obtain y t prime is equal to y t minus rho hat y t minus 1 and x t prime is equal to x t minus rho hat x t minus 1. And, then you apply ordinary least square to the transform data, you can do this because you know that y t prime is equal to beta naught prime plus beta 1 x t prime plus z t. Where, z t follows all the conditions of Gauss Markov theorem. I mean, so this follows normal 0 z sigma z square and they are independent. So that is why you can apply ordinary least square technique to the transform data.

And once you are done with you know ordinary least square, I mean once you have the fitted model like y t dash hat is equal to beta naught dash hat plus beta 1 hat x t, so this is the fitted model. Once you have the fitted model, where this parameters are obtained using ordinary least square technique, again you compete what you do is that you compute the residual, e residual for this model.

(Refer Slide Time: 47:35)



Now, you use Durbin Watson test to the residual obtained from the reparamaterized model. To check that whether still you have applied ordinary least square technique to the transform data y t prime x t prime they are also time series data. Again you need to check whether autocorrelation still exist on the transform data. So, if your Durbin Watson test indicates no autocorrelation in the errors then no additional analysis is needed. But, if Durbin Watson test indicates there is autocorrelation in the errors, then another iteration is required.

That means you apply you need to check whether in the transform data you still have the autocorrelation using the Durbin Watson test. And, if you see in the transform data there is no autocorrelation in the errors for this transform data you stop there. There is no additional analysis required but, if you see that the Durbin Watson test indicated that there is exist what there is autocorrelation in the error for the transform time series data.

Then you have to repeat the same thing once more and you know you may go for two iteration maximum and there you have to stop. So, here we talked about the data which are collected sequentially over time and they are called time series data. And, in time series data generally we suspect that the observations are not independent that essentially same like errors are not independent they are correlated.

So, we need to test whether errors are auto correlated or not. For that, we have to learned Durbin Watson test and the residual plots and all this things. And, if you see that you know the Durbin Watson test results are indicate that autocorrelation exist in the data. Then we have learned a technique how to estimate the parameters in the presence of autocorrelation in the model. That is all for today.

Thank you.