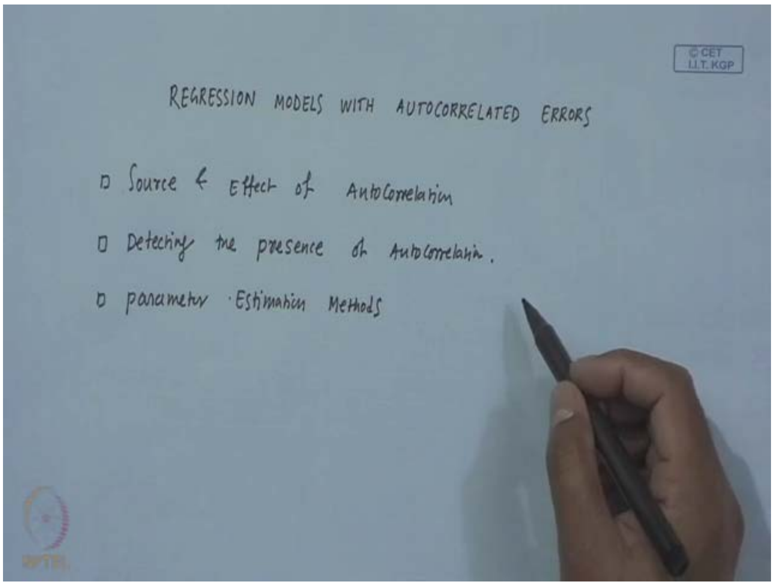
Regression Analysis
Prof. Soumen Maity
Department of Mathematics
Indian Institute of Technology, Kharagpur

Lecture - 33
Regression Models with Autocorrelated Errors

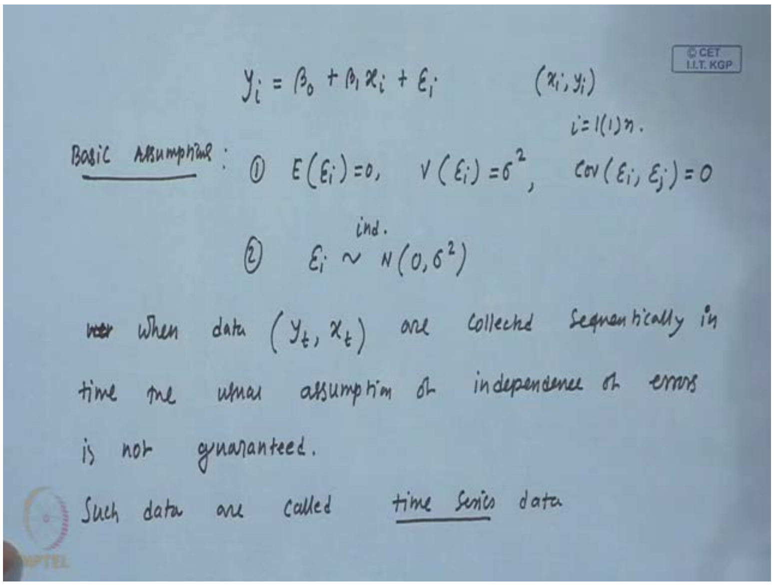
(Refer Slide Time: 00:37)



Today, we will start a new topic called a Regression Models with Auto Correlated Errors, here is the content of this module source and effect of autocorrelation and detecting the presence of autocorrelation. And then you know if you have autocorrelation in the model, I will explain what is this autocorrelation, and how to parameter, how to estimate the parameter of the model. So, let me give the idea what is the objective of this module given a set of data say x i, y i, while fitting a simple linear regression model say y equal to beta naught plus beta 1 x plus epsilon. So, we make several assumption on the error term, we assume that expectation of e equal to expectation of epsilon is equal to 0, variance of epsilon is equal to sigma square.

We assume that the errors are uncorrelated, and also we make a normality assumption on the error to for the testing of hypothesis, and you know for confidence and interval of a parameter. Now, if the data set let say x i and x i y i is a collected sequentially over time, then the assumption of this independent error is not guaranteed, so in that situation I mean when the data are collected over time, those type of data called you know time series data. And then how to deal with such situation, when the errors are correlated, let me you know write down the objective of this module clearly.

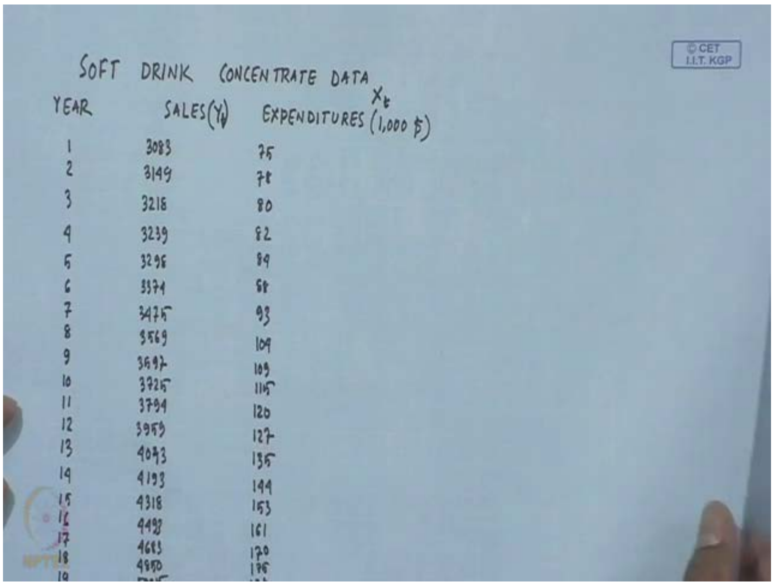
(Refer Slide Time: 03:08)



So, I am talking about model say very simple linear regression model y i equal to beta naught plus beta 1 x i plus epsilon i, and you are given the data set say x i y i, so i is from 1 to n. And the assumption we make that basic assumptions are we assume that expectation of epsilon i is equal to 0, variance of epsilon i is equal to sigma square, the constant variance and also we assume that you know the errors are uncorrelated. So, I can write in this form that covariance of epsilon i epsilon j is equal to 0, because the expectation is also 0.

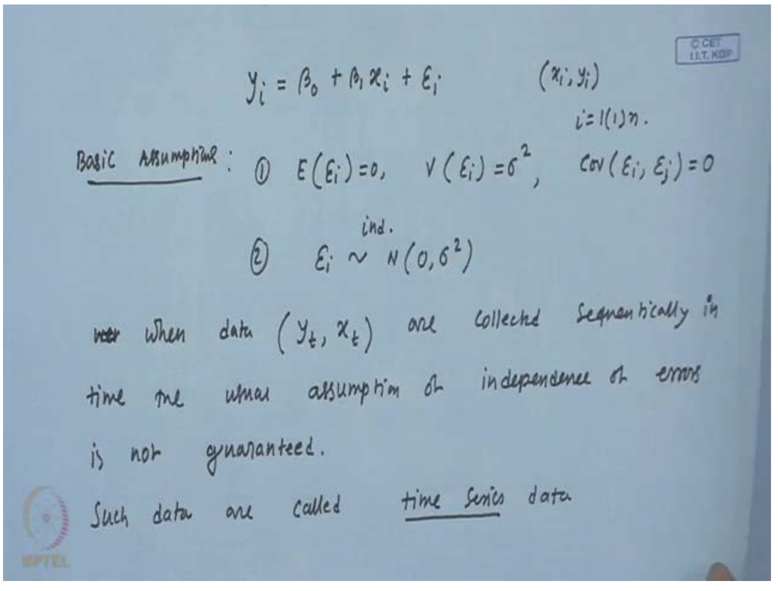
Well, and also we make the assumptions this is the first one, second one is that we assume that this epsilon i follows normal 0 sigma square, and they are independent. Now, what I said is that now, when this data say I will write y t instead of y i y t x t are collected sequentially in time the usual assumption of independence of errors is not guaranteed, anyway such data are called time series data. Let me just you know to make this part clear, what I mean by that data are collected sequentially in time, let me just give one example to clear your doubt.

(Refer Slide Time: 06:13)



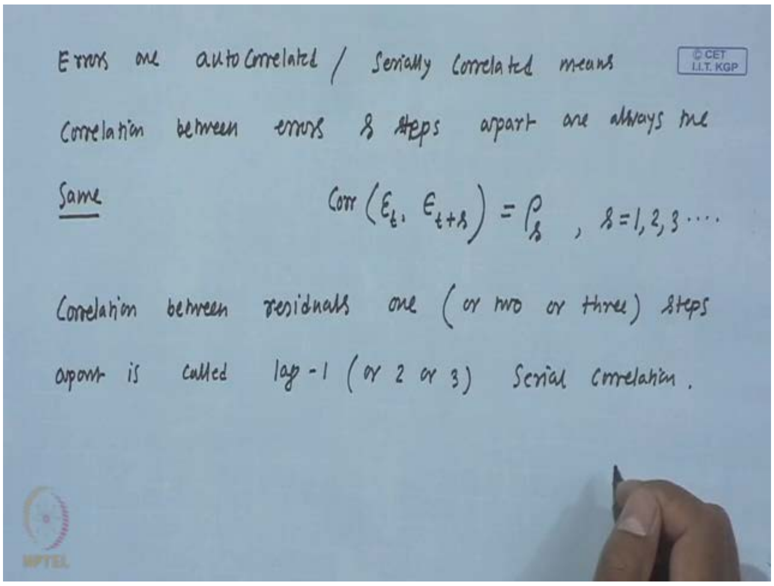
This is called you know soft drink concentrate data and this one is the regressor variable, this is the regressor variable x and this is the sales amount y, and this is the sales amount and this one is the x is expenditure on advertisement, so this is in 1000 dollar unit. And so we have data on the amount of money, you know used for the advertisement and sales amount for 20 years, so this data you know this is a let me call this is x t y t, so this x t and y t are collected over 20 years.

(Refer Slide Time: 07:31)



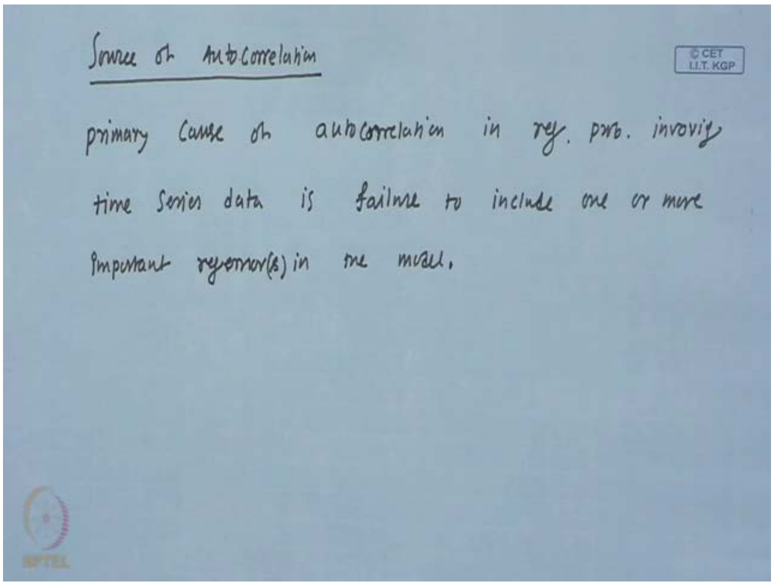
So, this is a time series data, so when the data are collected sequentially in time the usual assumption of independence of error is not guaranteed.

(Refer Slide Time: 07:50)



So, here we say that errors are autocorrelated or also you called a serially correlated, so that means, that errors are correlated or serially correlated means correlation between errors s steps apart are always the same. So, I am talking about the correlation between epsilon i or say epsilon t and epsilon t plus s, this is same for all t, and we usually this denote by rho s, for s equal to 1, 2, 3 like this. So, as the correlation between residuals say 1 or 2 or 3 steps apart is called a lag 1 or 2 or 3 serial correlation. So, if the s is equal to 1; that means, when you are constraint in the correlation between the errors, one step apart that is called lag one autocorrelation or serial correlation.

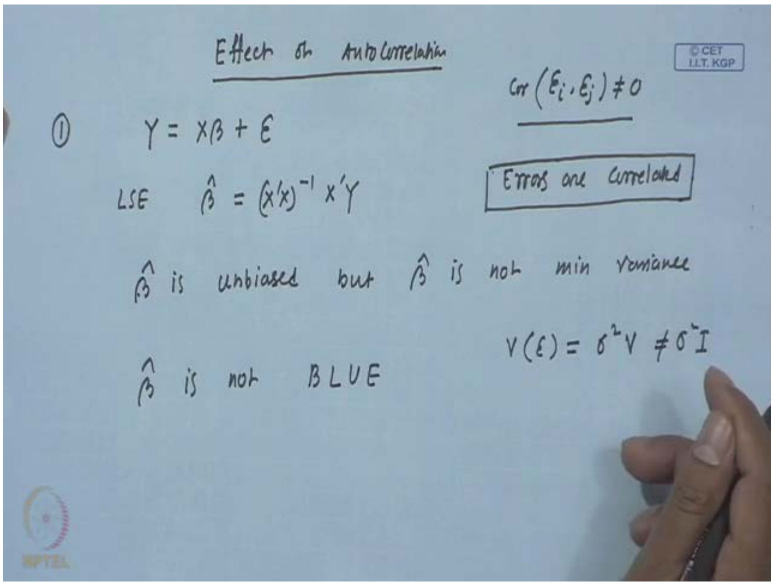
(Refer Slide Time: 11:26)



So, what is the source of this autocorrelation, so the primary source cause of autocorrelation in regression problem, involving a time series data is failure to include 1 or more important regressors in the model. So, what we mean by this one is that let me consider the example, of a soft drink concentrate data. So, there we are trying to regress the sales amount on the amount of expenditure for advertisement, but you know the growth, I mean the population increases over time and this growth in population has you know influence in the sales amount.

So, the population size is another important variable, which has influence on sales amount, so if you do not include the population size or increase in the population size you know that variable in the model, then you can expect autocorrelation in the time series data. So, the it says that the primary cause of autocorrelation in regression problem, involving time series data is you know is failure to include important regressor variable in the model, so we understood the source of autocorrelation, why it happens.

(Refer Slide Time: 14:54)

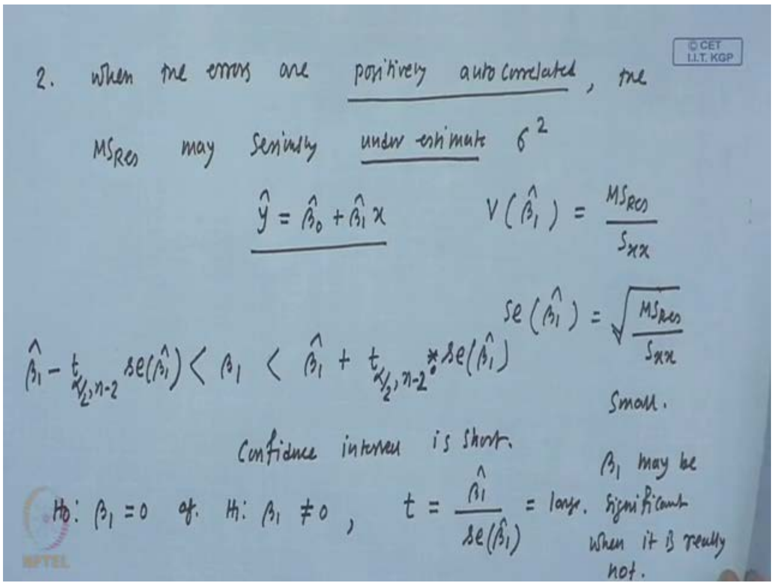


And then now let me talk about the effect of autocorrelation, so if effect of autocorrelation, so if autocorrelation is there in the model; that means, what is the meaning of this that epsilon i and epsilon j the correlation between them is not equal to 0. So, if this happen, if the errors are correlated and what is the effect of that while fitting a simple, say simple multiple linear regression model, so if you are fitting a multiple linear regression models, say y equal to x beta plus epsilon.

Then we know that the least square estimate is beta hat which is equal to x prime x inverse x prime y, so this is obtained using a least square estimate. And now, if you consider the basic assumptions on the model that you know this epsilon i follow normal distribution with mean 0, variance sigma square in the independent. Then the condition of the Gus Markov theorems are satisfied and so the beta hat, we get the beta hat is equal to x prime x inverse in x prime y that is the best linear unbiased estimate.

So, but here, once the condition that you know, here the errors are correlated, so here that condition is not true, so errors are correlated here, so because of the violence of this violation of this condition that you know errors are uncorrelated. Here, beta hat is unbiased, but beta hat is not minimum variance, so because of this problem like you know errors are correlated in this case, so the least square estimate beta hat is not the best linear unbiased estimator. Of course, you know if the variance of epsilon is equal to say sigma square v, where v can be written as sigma square I, then we know how to get the best linear unbiased estimator using the generalized least square technique.

(Refer Slide Time: 18:42)



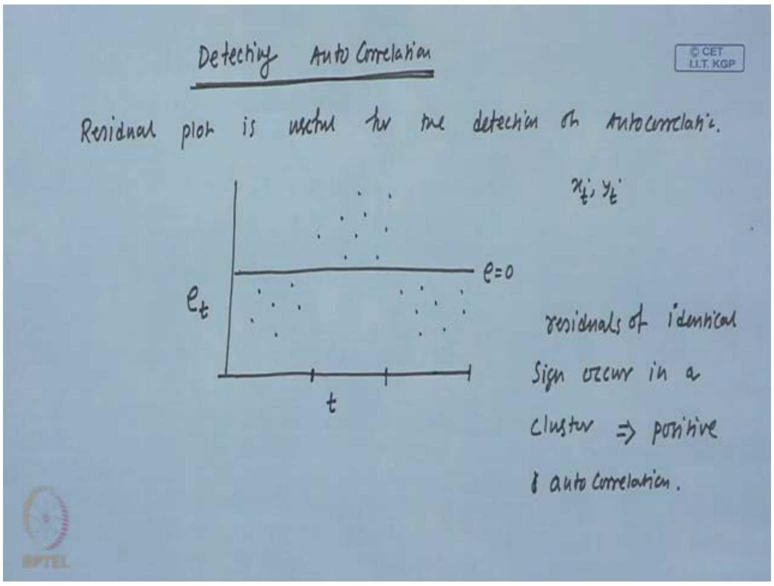
The second effect is when the errors are positively auto correlated I will say what is mean by this one, then the M S residual may seriously under estimate sigma square, because we know that M S residual is a unbiased estimator for sigma square. So, what is the consequence of this one that you know M S residual underestimate sigma square, the consequence of this one is that the variance of suppose you are fitting the model y hat a simple linear regression model beta naught hat plus beta 1 hat x is the fitted model.

And so we know that variance of beta 1 hat is equal to M S residual by S x x, so the standard error of beta 1 hat is equal to square root of this quantity M S residual by S x x. Now, since this one is small the standard error is going to be small, and the consequence of this one is that when we compute the confidence interval for say beta 1 hat. So, the confidence interval, if you refer my first module the confidence interval for beta 1 is beta 1 hat plus t alpha by 2 n minus 2 degree of freedom into standard error of beta 1 hat. And the lower boundaries beta 1 hat minus t alpha by 2 n minus 2 standard error of beta 1 hat.

Since, this one is small then this confidence interval is short, so you will get a narrow interval for the parameter, and which might not be the true interval for the parameter beta 1. And also in the regression of model, we test a hypothesis like to check the significance of beta 1 or to or the significance of the model or the linear term, we check the hypothesis like H naught is say beta 1 equal to 0 against the H 1 that a beta 1 is not equal to 0. And you know that the test statistic to test one is t equal to beta 1 hat by standard error of beta 1 hat.

So, see when there exist positive autocorrelation in error the M S residual underestimate sigma square, and the consequence of that is that the standard error of beta 1 hat is small. So, this one is small, so t is large, so; that means, that beta 1 may be significant, so you will get a result that beta 1 is significant, when it is really not, because of the positive autocorrelation in the data, these are the effect of autocorrelation.

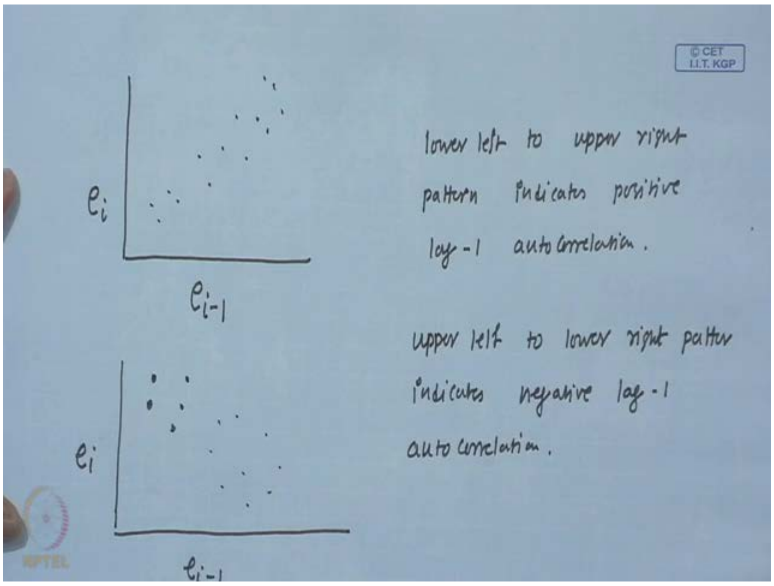
(Refer Slide Time: 24:11)



Now, let me talk about how to detect a autocorrelation, so detecting autocorrelation. So, the first technique is you know the residual plot is useful for the detection of autocorrelation. So, what we plot is that you plot, so given the data said x i y i or x t y t you fit a simple linear regression model, so once you have a fitted you can compute the residuals.

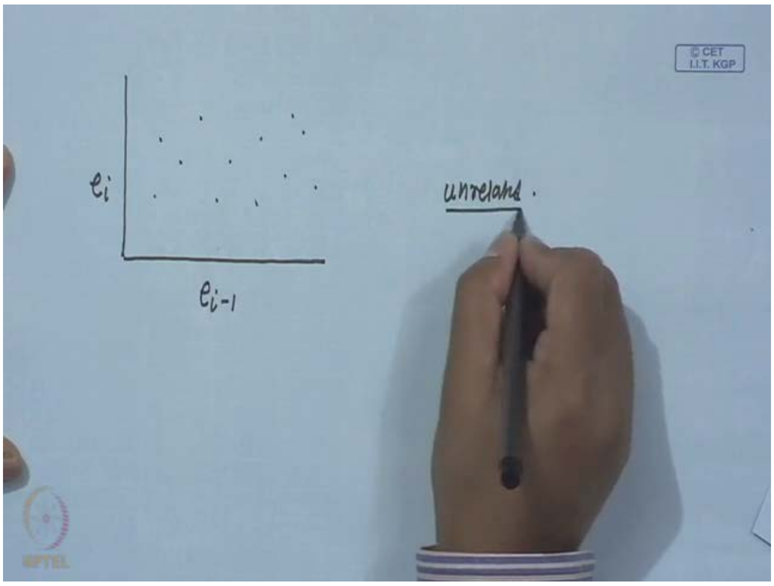
And then you plot let me call it x t y t and you plot residual e t against t, and if you see that your plot is like this say for example, this is the residual plot e t against t, here what happen is that till this point you see all the residuals, so this is the line e equal to 0. So, till this time point all the residuals are negative and from here to here you can see all the residuals are positive, and again the residuals are negative in this segment. So, residual of identical sign occur in a cluster, then this indicate the positive autocorrelation maybe I will explain why this is true, before that you know let me give some more residual plots of like you know.

(Refer Slide Time: 27:32)



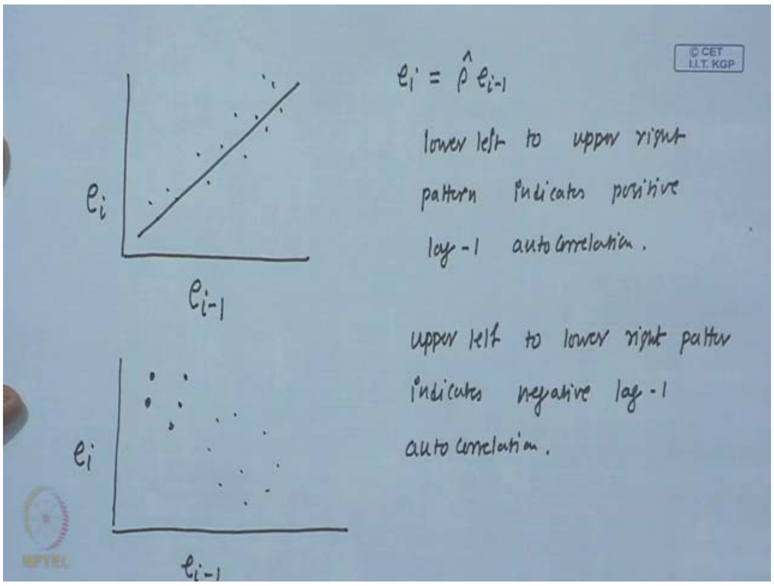
If the this residual plot is e i well e t against e i minus 1, so here we are short of trying to find the relation between e i and e i minus 1; that means, a lag one correlation. So, if you see the plot is like this see for example, then this one is sort of lower left to upper right pattern indicates positive lag one autocorrelation. And if you see the scatter plot of e i against e minus 1 is like this say e i, e i minus 1 it is say for example, like this; that means, this is upper left to lower right pattern, this indicates negative lag one autocorrelation.

(Refer Slide Time: 30:10)



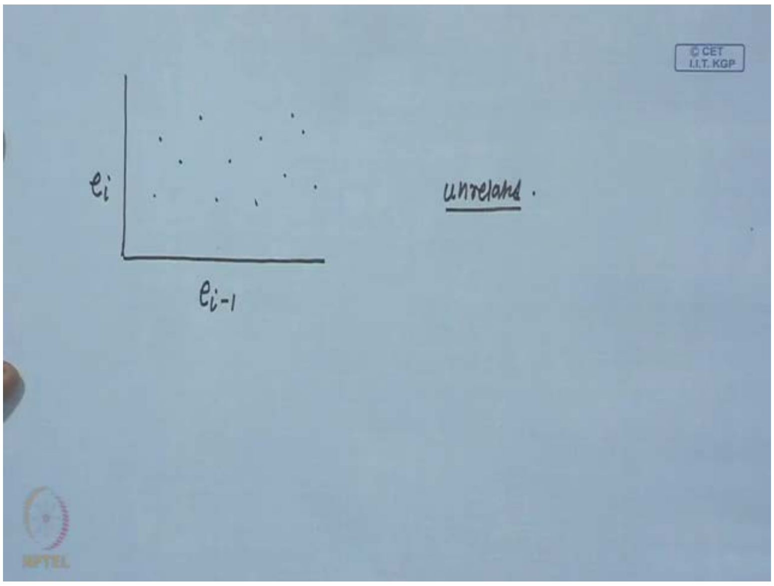
And other one is, so this plots are like to detect a lag 1 autocorrelation, similarly for lag 2, you have to plot e i against e i minus 2 and see how they are related. So, e i e i minus 1 if you see the pattern is like this, then this indicates that errors are uncorrelated or unrelated.

(Refer Slide Time: 30:51)



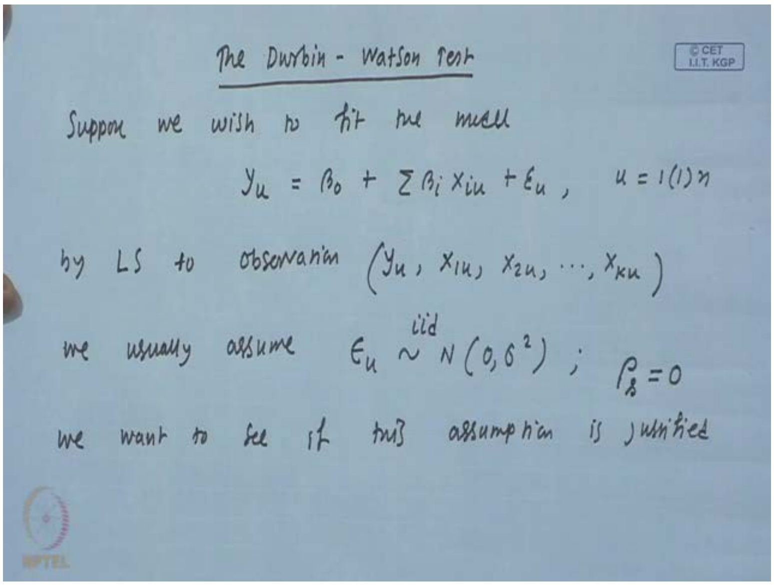
So, these are the graphical technique to identify the existence of autocorrelation and specially for lag one autocorrelation, you have to plot this is basically we are trying to find the relation between e i and the previous residual. And also you see you know e i now this sort of you know if you fit a straight line, if you fit a model between few regress e i on e i minus 1. You will get a straight line model like e i is equal to some rho into e i minus 1, and this clearly says here rho is positive and e i increases sort of as e i minus 1 increases, so they are very similar in magnitude. So, this is what the positive lag one autocorrelation, and negative lag one this indicates negative lag one autocorrelation.

(Refer Slide Time: 31:58)



And this one says that there are there is no correlation between e i and e i minus 1; that means, the errors are uncorrelated. Now, we will talk about one statistical test to test the presence of autocorrelation.

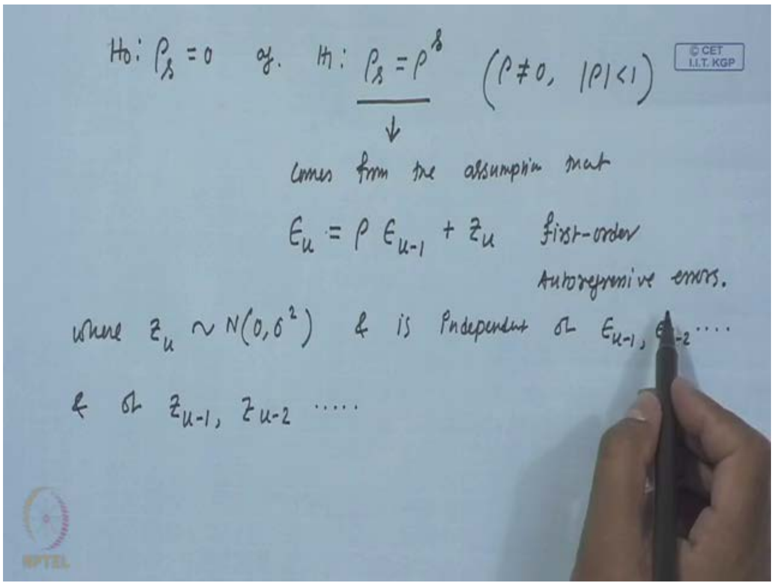
(Refer Slide Time: 32:23)



The test is called The Durbin Watson Test, suppose we wish to fit the model say y u equal to beta naught plus beta i x i u plus epsilon u, u is from 1 to n by least square technique, to observation say y u. And then I am talking about multiple linear regression model x 1 u x 2 u and something x k u, so what we do is that we usually assume that this epsilon u follows normal distribution with 0 sigma square, and there IID. That is what; that means, that we are assuming that the lag is autocorrelation is equal to 0; that means, correlation between the errors s step apart that is equal to 0.

So, if you want to use least square technique to fit this model you have to assume this; that means, we are assuming this. Now, we want to what we want to do is that, we want to see if this assumption is justified for the given data, for that what we will do is that we will test hypothesis.

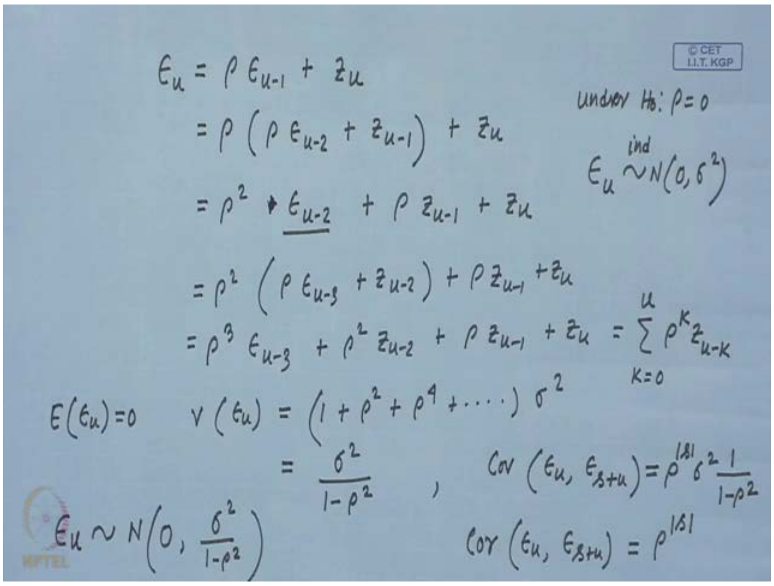
(Refer Slide Time: 35:23)



H naught that rho s is equal to 0, against the alternative hypothesis H 1 that rho s is equal to rho to the power of s, this rho is not equal to 0 and it is a modulus value is less than 1. Now, what I will do is that you know why particular, you are considering this alternative how this alternative comes, we will talk about that little bit. So, if null hypothesis accepted here in our test we will be talking about one test procedure using the Durbin Watson test, and if null hypothesis is accepted here that is rho s is equal to 0; that means, there is no autocorrelation in the error.

And here we wrote the alternative hypothesis rho is equal to rho to the power of s. Now, what I will do is, I will try to justify the choice of this alternative hypothesis. So, this alternative hypothesis comes from the assumption that the errors are error follow this model, epsilon u is equal to rho epsilon u minus 1 plus z u; that means, the errors are first order autoregressive errors. Where, this z u first order autoregressive means, there is a linear relation between epsilon u and epsilon u minus 1, where epsilon z u follows normal 0 sigma square. And this z u is independent of epsilon u minus 1 epsilon u minus 2 and of z u minus 1 z u minus 2 like this. So, if the errors are first order autoregressive error, then I can write this epsilon u in this form.

(Refer Slide Time: 39:00)



So, my epsilon u, I took this epsilon u is equal to rho epsilon u minus 1 plus z u I took. Now, I can write this one as rho rho epsilon u minus 2 plus z u minus 1, I am just replacing epsilon u minus 1 by this 1 plus z u. So, this can be written as rho square plus rho square epsilon u minus 2 plus rho z u minus 1 plus z u, this again if you replace this epsilon u minus 1 by this quantity as rho epsilon u minus 3 plus z u minus 2 plus rho z u minus 1 plus z u.

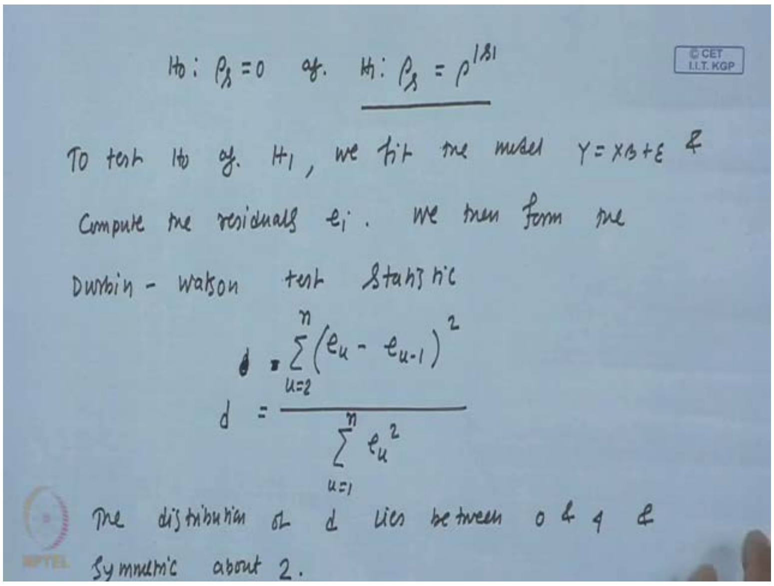
So, what we will get is that, you will get rho to the power of three epsilon u minus 3 plus rho square z u minus 2 plus rho z u minus 1 plus z u. So, ultimately, you can write this as again, you replace epsilon u 3 using this formula you can write this as rho to the power of k z u minus k, k is from 0 to u you can check that. So, this is what the epsilon u in terms of z u z, so expectation of epsilon u is equal to 0, because expectation of z is equal to 0, what about the variance of epsilon u, now the variance of they all are independent, z i is the independent.

So, you can write this the variance, this one as 1 plus rho square plus rho to the power of 4 like this into sigma square, because the variance of z u minus k is sigma square and they are independent. So, you can write in this form, and this can be written as sigma square by 1 minus rho square, and similarly you know you can check that the covariance of epsilon u and epsilon s plus u.

So, I am trying to find the correlation between if the errors are you know first order autoregressive, what is the correlation between epsilon u and epsilon s plus u. So, you can check the covariance is equal to rho to the power of s sigma square, 1 by 1 minus rho square, so this is the covariance. And since, this is the covariance, and then it is clearly the correlation between epsilon u and epsilon s plus u is equal to rho to the power of s. And here as you see now the epsilon u, which is first order autoregressive, they follow normal distribution with mean 0, and variance sigma square by 1 minus rho square and under the null hypothesis.

The null hypothesis is under H naught that rho equal to 0, under this null hypothesis this sigma epsilon u, they follow normal 0 sigma square, you put rho equal to 0 here normal sigma square. And correlation also become 0 because rho equal to 0 the correlation between them is 0, so they independent, so the under null hypothesis, epsilon u follow normal sigma square and they are independent.

(Refer Slide Time: 44:51)

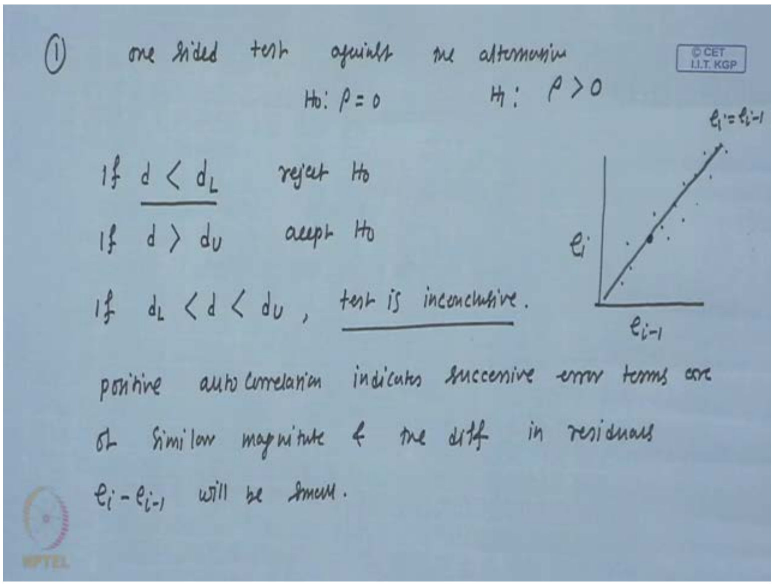


So, you understood the significance of this alternative hypothesis now, so we are testing the hypothesis that H naught is a rho s is equal to 0 against the alternative hypothesis H 1 that rho s is equal to rho to the power of s. So, and we have checked that this alternative hypothesis in on the assumption that the errors are first order autoregressive, then rho s is equal to rho to the power of s.

This is the correlation between sigma u epsilon u and epsilon u plus s, now to test this hypothesis to test H naught against H 1, so what we do is that we fit the model say y equal to x beta plus epsilon. Assuming, that all the basic assumptions are true and using the least square technique, and then once you have the fitted model you can compute the residual, and compute the residual e i, and then once you have the e i does not matter whether the basic assumptions are true or not.

Now, you can check whether there is a autocorrelation in the error term or not by using the test, we then form the Durbin Watson test statistic that d equal to e u minus e u minus 1 square, u is from 2 to n by e u square, u is from 1 to n. So, given a set of data you fit the model you find the residual, and then you compute the Durbin Watson test statistic, and then based on this test statistic, we will now test this hypothesis whether there exist autocorrelation in the data or not. So, the distribution of d lies between 0 and 4 and the distribution is symmetric about 2, so we know the test statistic to test this hypothesis, and now we will be talking about you know what are the critical reasons when to reject and accept.

(Refer Slide Time: 48:23)



Say first case, one sided test against the alternative that rho is greater than 0, so basically we are testing here of for a lag one that H naught rho equal to 0 against H 1 rho is greater than 0, so you compute the test statistic d. And then if d is less than d L, I will say what is this d L you reject H naught, so there is a table for this Durbin Watson test statistic.

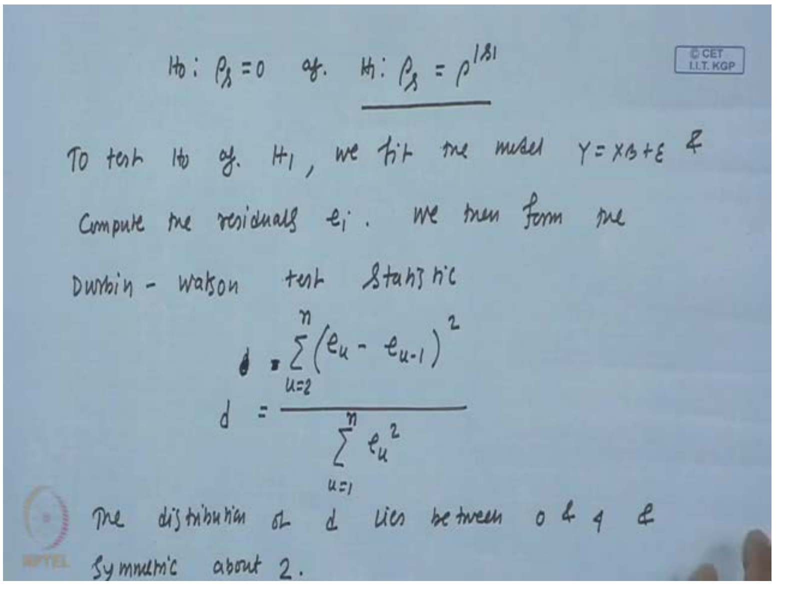
So, the table is given for it is only it requires the number of observations you have, so for that soft drink concentrate data, there we had 20 observations, and then from the table you have to see the d l and d u value corresponds to n equal to 20.

So, if d is less than d L, we reject a H naught, if d is greater than d U, we accept a H naught, I will explain you know why suddenly this a critical reason, and if d L is less than d and less than d U the test is inconclusive. Now, what happen is that, if the d is small if, so given a data you fit a model using the ordinary least square technique, and you get e i and once you have the residual you can compute the Durbin Watson test statistic.

So, small value of d implies rho is equal to 0; that means, small value of d indicates, there is no you reject this one; that means, you accept this 1, so small value of d indicates that auto autocorrelation exist in the model. So, if the d is small you are rejecting this; that means, you are accepting this, acceptance of rho is greater than 0 means the data has positive or the error has a positive autocorrelation, so let me just explain this part why this is true. So, the positive autocorrelation, when it is positive autocorrelation, you just recall the graph you are plotting e i against e i minus 1, so this is the case when it has positive autocorrelation.

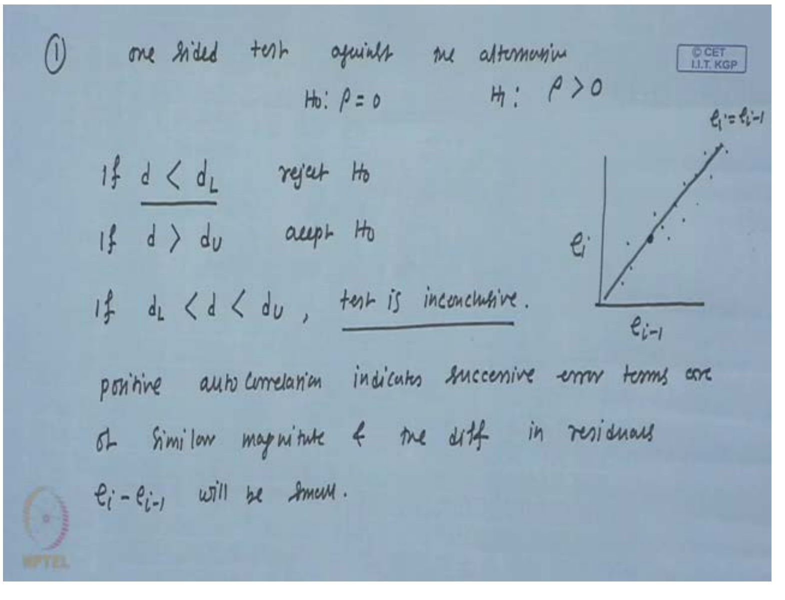
The positive autocorrelation indicates successive error terms are of similar magnitude, and the difference in residuals e i minus e i minus 1 will be small, so this is the case when it indicates the existence of positive autocorrelation. So, here you can see you take a point and this is the e i minus 1 value and this is, so the e i minus 1 and e i they are almost of similar magnitude, that is why you get a you know all the points are centered about the line x equal to y or the line this is centered about e i equal to e i minus 1. So, points are centered about this line means e i and e i minus 1 are very similar, and since they are similar the difference is small.

(Refer Slide Time: 54:32)



So, once the difference is small you now recall the Durbin Watson test statistic d, that involve the difference, so if the difference is small the d is going to be small.

(Refer Slide Time: 54:44)



And once d is small that implies the existence of positive lag one autocorrelation, and that is why we reject the null hypothesis and accept the alternative hypothesis, so I hope you know this will make clear, why this is a critical reason. As, I told you there is a table I will talk in the next class, this is a table for d L and d U value for different n and for different alpha, and I will talk about other cases also in the next class. So, we need to stop now, we will continue with the Durbin Watson test with the some example, to illustrate the Durbin Watsons test in the next class.

Thank you.