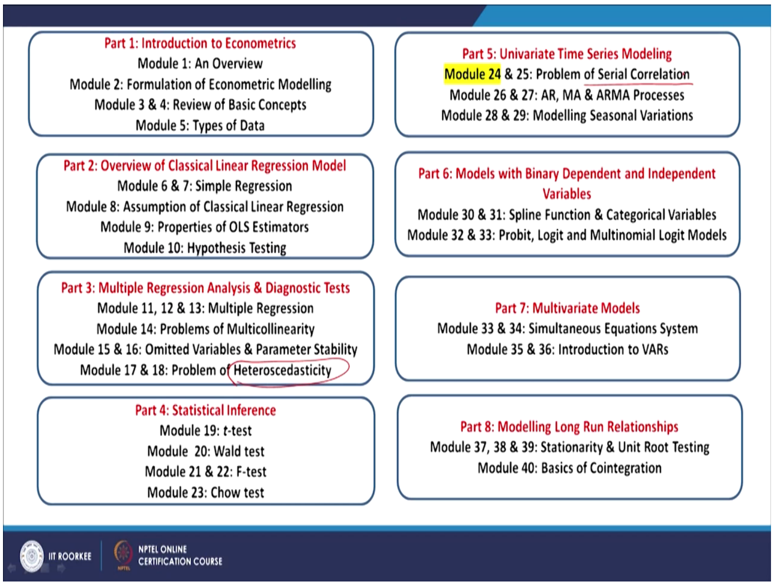
Econometric Modelling
Professor Sujata Kar
Department of Management Studies
Indian Institute of Technology Roorkee
Lecture 24
Problem of Serial Correlation-I

This is Module 24 of the course on Econometric Modelling.

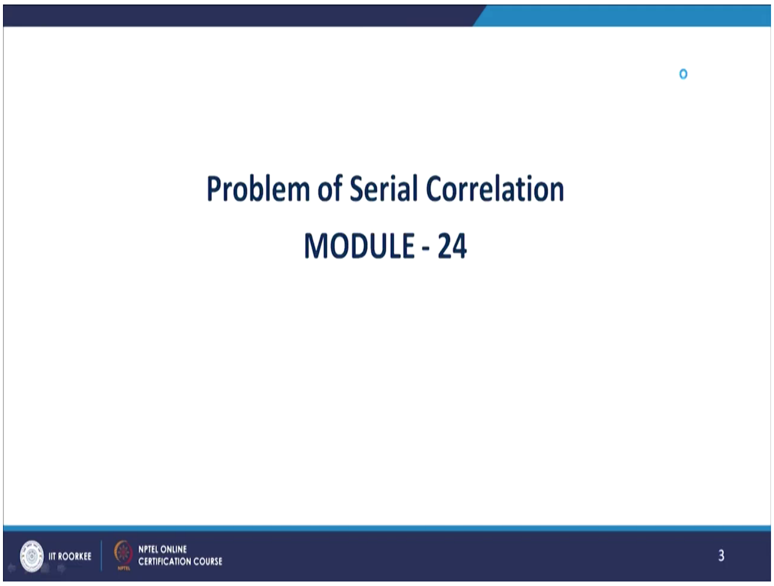
(Refer Slide Time: 00:32)



Module 24 begins Part 5 which deals with univariate time series modeling. So, univariate time series modeling what do they mean probably would be clearer in later modules, but module 24 to begin with deals with serial correlation. Serial correlation was one of the assumptions under CLRM.

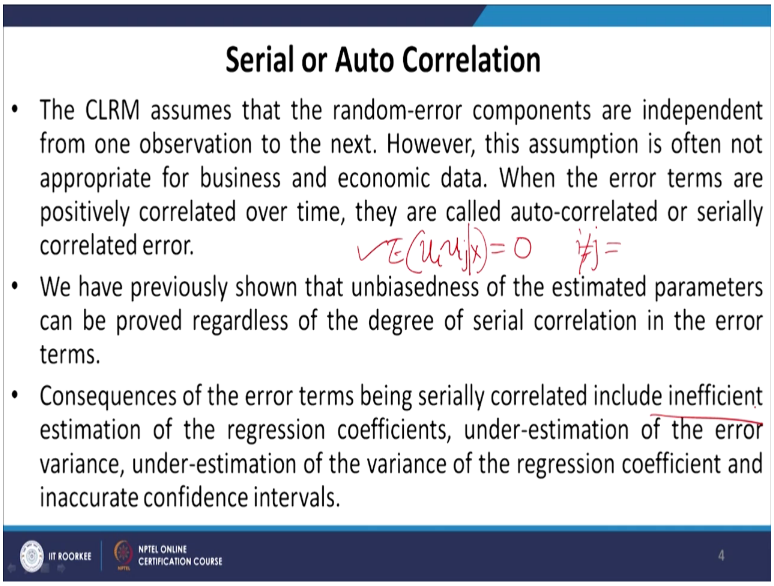
So, first of all, continuing with our multiple regression analysis, then inferences, we had earlier talked about heteroscedasticity, which is actually one violation of the CLRM assumptions or Gauss-Markov assumptions. Similarly, we also assume under Gauss-Markov theorem, that there is no serial correlation between the population errors. Now we will first explore that what happens if said that assumption is actually violated.

(Refer Slide Time: 01:30)



So, this module and the next module are on the problem of serial correlation.

(Refer Slide Time: 01:36)



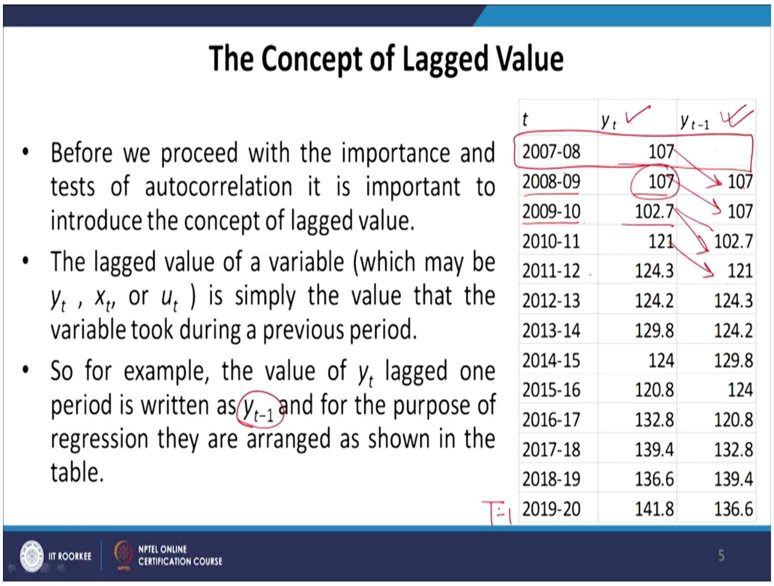
Serial correlation is also called autocorrelation. The CLRM is a classical linear regression model that assumes that the random error components are independent of one observation to the next. However, this assumption is often not appropriate for business and economic data. When the error terms are positively correlated over time, they are called auto-correlated or serially correlated errors. So, this actually has reference to the assumption of the expected value of ui, uj equals to 0 that could be conditional upon the values of x for all i, j, for all i not equal to j.

Now, when this assumption is not fulfilled, we have the problem of serial correlation. We have previously shown that the unbiasedness of the estimated parameters can be proved regardless of the degree of serial correlation in the error terms. Alternatively, when we proved unbiasedness of the estimated parameters, we did not need to or we did not consider whether the errors were serially correlated or not. It did not play any role in that process.

Consequences of the error terms being serially correlated include inefficient estimation. So, it has nothing to do with unbiasedness, but it has something to do with efficiency or the estimated parameters being the best parameters, so inefficient estimation of the regression coefficients, underestimation of the error variance, underestimation of the variance of the regression coefficient, and inaccurate confidence intervals levels.

So, whenever we would need error variance and covariance between the error terms, then in all those scenarios or situations we will land into trouble if the errors are serially correlated.

(Refer Slide Time: 03:48)

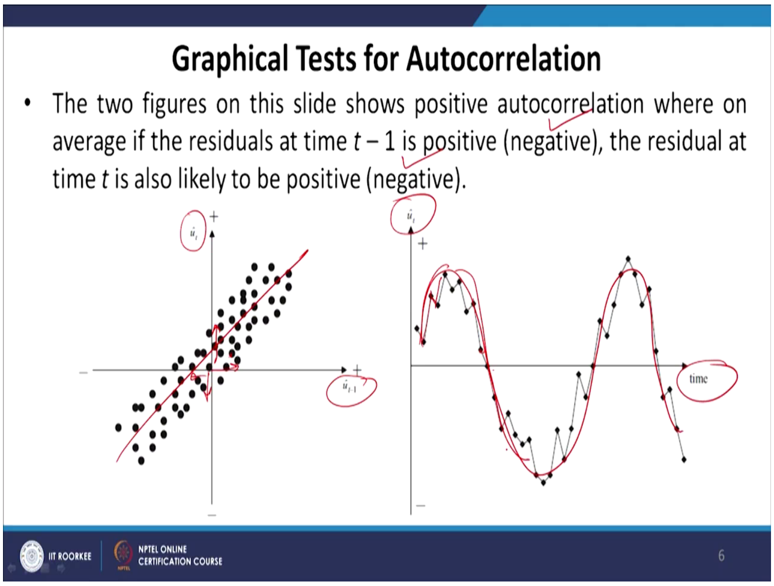


Before we proceed with the importance and tests of autocorrelation, it is important to introduce the concept of lagged value. The lagged value of a variable which may be ut, yt or xt is simply the value that the variable took during a previous period. So, for example, the value of yt lagged one period is written as yt minus 1. And for the purpose of regression, they are arranged as shown in this table *(refer to slide time 03:48).*

So, you can see that it is yt. And if t refers to the current period, then 2008-09 has an observation of 107, 2009-10 has an observation of 102.7. Now, one period lagged value with reference to 2009-10 is 2008-09 or the value associated with 2008-09 which is 107, so when we are measuring yt minus 1 we are bringing that 107, and so on. So, this is the concept of lagged value *(refer to slide time 03:48)*.

In the current period, we are considering the value of the previous period of the same variable and this is how they are stacked. Now, of course, you can see that when we run a regression of y2 on yt1, yt on yt minus 1, then we would miss out on one observation. The first observation is actually dropped from the regression because for the independent variable we do not have observations pertaining to 2007-08. So, if the total number of observations is t, then for regression, I will be able to use only t minus 1 observation.

(Refer Slide Time: 05:30)

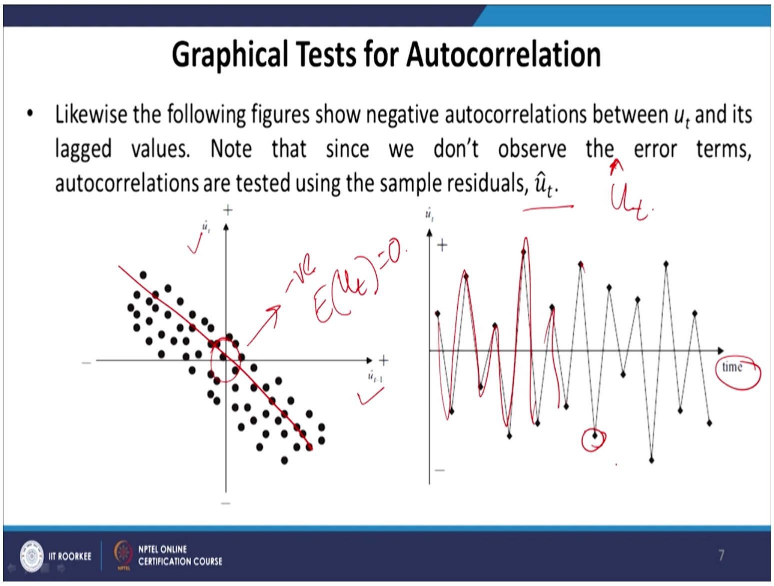


Now, we talk about graphical tests for autocorrelation. The two figures on this slide *(refer to slide time 05:30)* show positive autocorrelation where on average if the residuals at time t minus 1 is positive, the residual at time t is also likely to be positive, and similarly, the residuals at time t minus 1 is negative, the residual at time t is also likely to be negative. So, you can see that this is where we are measuring ut hat on the vertical axis and ut hat, ut minus 1 hat on the horizontal axis.

Now, you can see that we are measuring ut hat on the vertical axis and time on the horizontal axis. So, how ut hat is actually changing or moving across time? And it shows that with time it is increasing then if it is decreasing, but most often you can see that we are not crossing the horizontal axis very frequently *(refer to slide time 05:30)*. This implies that if I consider any two points, for example, this point and this point, then this roughly shows that whenever there was a decline, there has been a decline in the next period, then again there has been a consequent decline.

So, whenever there is an upward trend, generally it is followed. Whenever there is a downward trend that is also generally followed, which implies that overall when the values are decreasing, then they decrease for some period of time. When they start increasing, then again they start to keep on increasing for some period of time. This itself implies that the current value is positively associated with the previous value and vice versa.

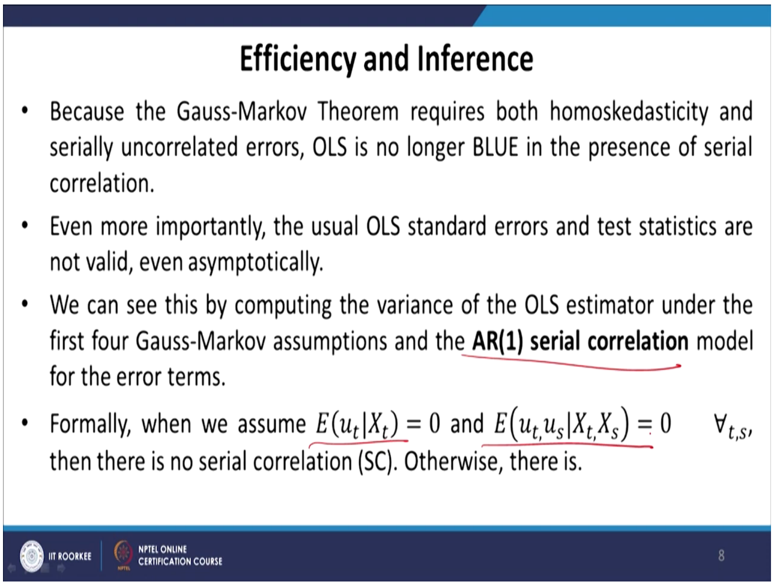
(Refer Slide Time: 07:38)



Likewise, the following figures *(refer to slide time 07:38)* show negative autocorrelation between ut and its lagged values. Note that since we do not observe the error terms, autocorrelations are tested using the sample residuals ut hat, since we cannot observe ut. So, that is why we work with ut hat. Now, you can see that we can plot a downward sloping line through these points. And as a result of which we can conceptualize a negative relationship between ut hat and ut minus 1 hat. You must also observe that all these observations are centered around 0, which implies that even though they are having serial correlation, the error term is actually having a mean of 0.

Now, if the error terms are plotted against time, then if they are negatively related, then we would first observe that they are changing, they are actually crossing the horizontal axis very frequently, because if there is an increase the next one is a sharp decrease, if the next one is currently it is a decrease then the next one is a sharp increase, so which implies that they are actually not following each other, and as a result of which we have this kind of fluctuations and they are negatively related.

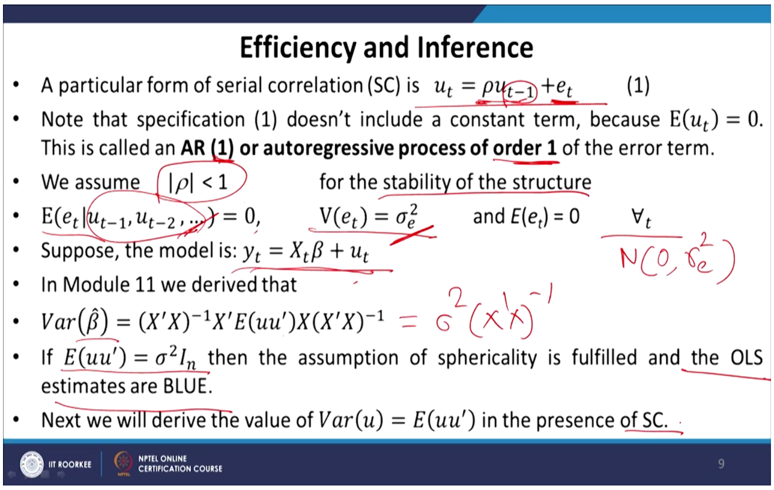
(Refer Slide Time: 09:02)



Now, we talk about efficiency and inference. What kind of problem do we actually land into if we have serial correlation or autocorrelation? Because the Gauss-Markov theorem requests both homoscedasticity and serially uncorrelated errors, OLS is no longer BLUE or best linear unbiased estimator in the presence of serial correlation. Even more importantly, the usual OLS standard errors and test statistics are not valid even asymptotically.

We can see this by computing the variance of the OLS estimator under the first four Gauss-Markov assumptions and the AR1 serial correlation model for the error terms. Now, AR1 serial correlation model I am going to define very soon in the next slide itself. Formally, when we assume the expected value of ut given xt is equals to 0 and the expected value of ut, st given our conditional upon the values of Xs is 0 for all t, s, then there is no serial correlation or autocorrelation, otherwise, there is.

(Refer Slide Time: 10:06)



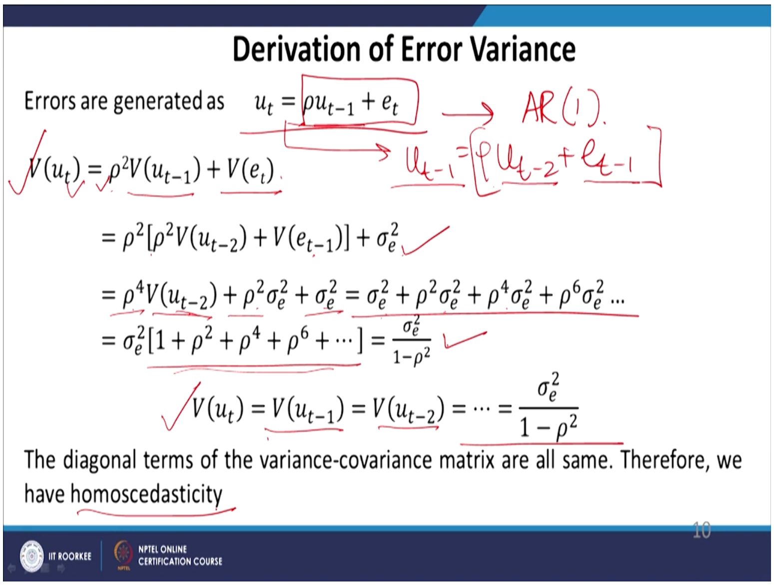
A particular form of serial correlation is ut equals rho ut minus 1 plus et. Note that specification 1 does not include a constant term, because the expected value of ut is equaled to 0. So, we do not need to include a constant term. This is called an AR1 or autoregressive process of order 1 of the error term. This is AR1 or order 1 because we are considering only one lag. If we would have considered two lags, then this would have been AR2. If we would have considered 10 lags, then this would have been AR10, and similarly, generalizing by considering P lags, we would have an ARP model *(refer to slide time 10:06)*.

We assume the mod value of rho is less than 1 because this is required for the stability of the structure and what it actually implies would be clarified much later towards the last units. Suppose, the model is yt equals to Xt beta plus ut, the original model, yes, we also assume that the expected value of the error term here is 0 conditional upon the values of the independent variables that are the previous values of ut and variance of et is equaled to sigma e square for all t. So, this is homoscedastic and this also has a mean of 0. So, obviously, it follows a normal distribution with mean 0 and sigma square e as the variance. Now, this is my original model *(refer to slide time 10:06)*.

In module 11 we derived that variance of the beta hat is equal to X prime X inverse X prime expected value of uu prime X, X prime X inverse. Now, if the expected value of uu prime is equal to sigma square In, where In is an identity matrix, then the assumption of sphericality is fulfilled because we have constant error variance and all the covariance is between the error terms are 0 and the OLS estimates are BLUE.

Also by incorporating sigma square In here *(refer to slide time 10:06)*, we can easily prove that we have already done that this becomes sigma square X prime X inverse which is the usual variance of the estimated parameter, the variance of the beta hat. Next, we will derive the value of the variance of u equals the expected value of uu prime in the presence of serial correlation.

(Refer Slide Time: 12:43)



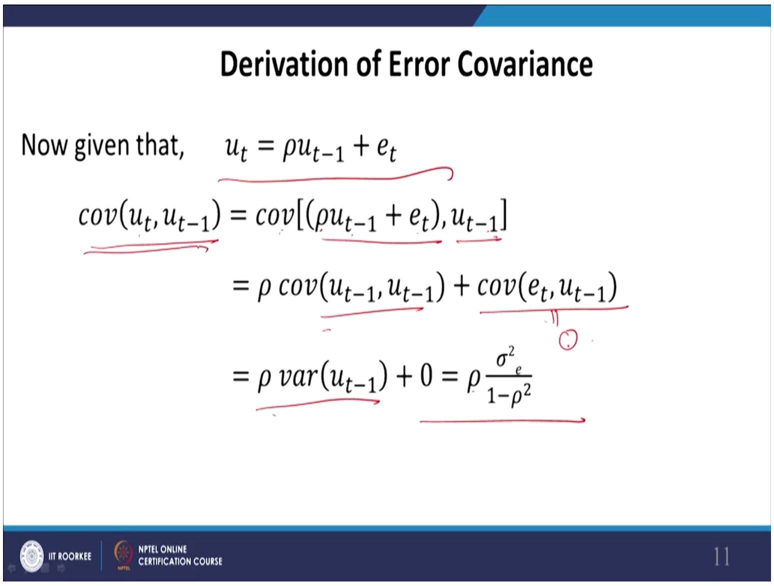
So, errors are generated as it is mentioned that it is an AR1 series variance of ut would be a variance of this expression. So, say rho being a constant it comes out, we have rho square, the variance of ut minus 1 plus the variance of et. Now, the variance of et comes here as sigma e square. The variance of ut, in place of ut here, I again plug in the value, the way I have written ut equals rho ut minus 1 plus et, ut minus 1 can also be written as rho ut minus 2 plus et minus 1. So, this expression is actually put in. Then what we have is the variance of ut minus 1 would be rho square variance of ut minus 2 plus the variance of et minus 1 *(refer to slide time 12:43)*.

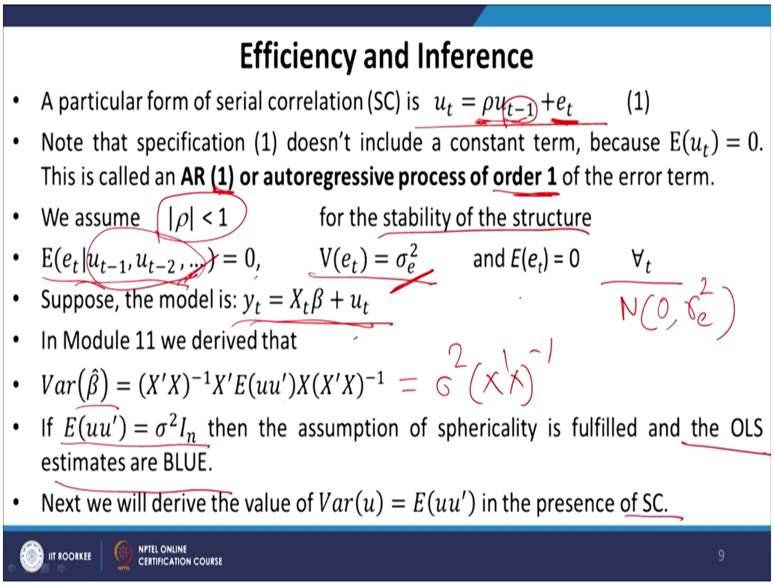
Now, when I open the bracket and multiply, I have rho raise to the power 4 variance of ut minus 2 plus rho square variance of et minus 1 is also sigma e square and we have sigma e square from here. So, in a similar fashion if I further substitute for the values of ut minus 1 here and then consequently ut minus 3 and so on, then I will have a series like this, where sigma e square plus rho square sigma e square plus rho raise to the power 4 sigma e square and so on.

By taking sigma e square common I have 1 plus rho square plus rho raise to the power 4 plus rho raise to the power 6 and so on. This is an infinite GP series. So, we can write it as sigma e square divided by 1 minus rho square. So, this is actually equal to the variance of ut. And since it is not dependent on t, that is the time subscript, so that would be also the case for the variance of ut minus 1, the variance of ut minus 2, and so on. So, the variance is constant.

The variance of ut, ut minus 1 or ui for all i equal to 1 to t are constant and that is equal to sigma e square divided by 1 minus rho square. So, what we have is the diagonal terms of the variance-covariance matrix are all same. Therefore, we have the assumption of homoscedasticity.

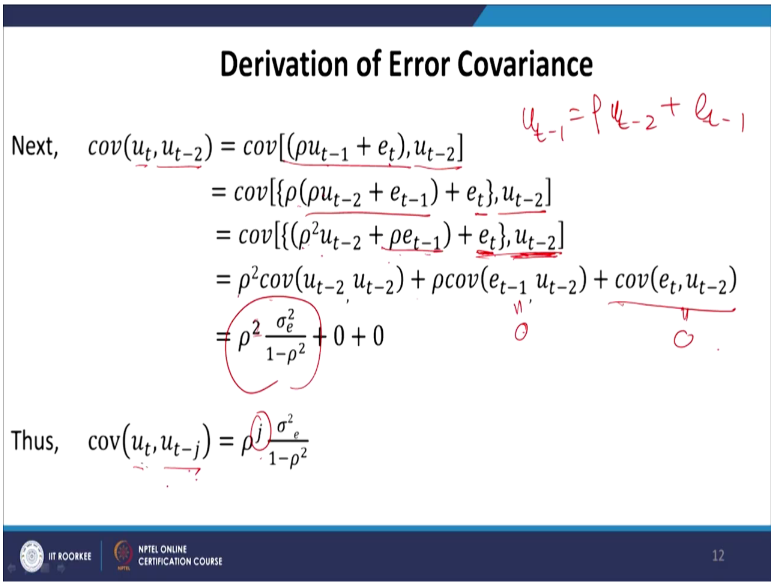
(Refer Slide Time: 15:07)

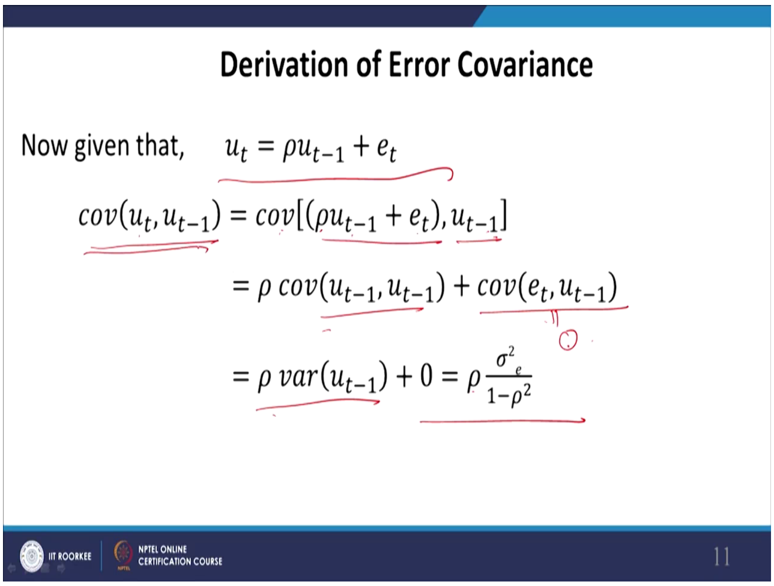




Now, given that ut equals rho ut minus 1 plus et, we derive the covariance between ut and ut minus 1. So, ut is rho ut minus 1 plus et and ut minus 1. Rho being constant comes out. We have covariance between ut minus 1, ut minus 1, the covariance between et and ut minus 1. This is basically variance of ut minus 1 and this is supposed to be 0 under the assumption that et is actually not correlated to the error terms and it has constant variance. So, this equals 0. Therefore, we have rho into sigma square e divided by 1 minus rho square as the covariance between ut and ut minus 1.

(Refer Slide Time: 15:58)



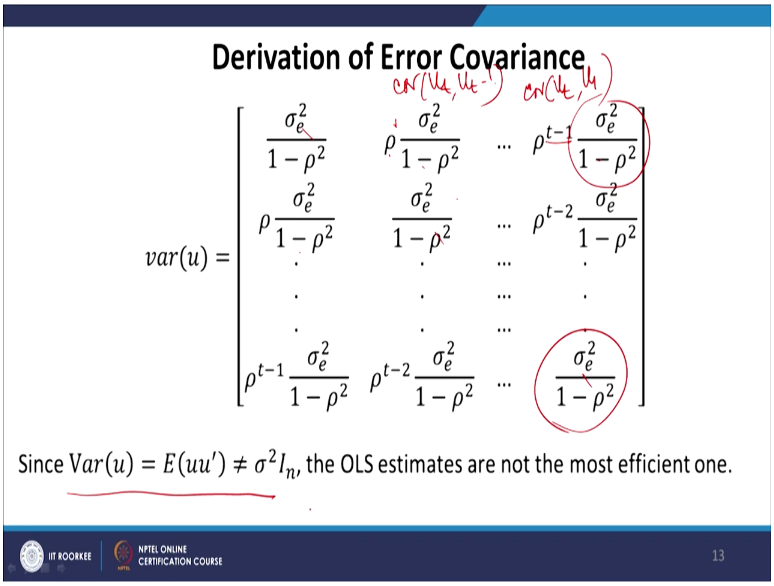


Next is we try to derive the covariance between ut and ut minus 2. Now, again in place of ut, I have rho ut minus 1 plus et and ut minus 2 as it is. Again, ut minus 2 can be written, as shown in the previous to previous slide, that ut minus 1 would be rho ut minus 2 plus et minus 1.

So, rho ut minus 2 plus et minus 1 and then et comes from here, and finally, et minus 2. So, what we are having covariance between rho square ut minus 2, rho into at minus 1 plus et and ut minus 2. So, rho square covariance ut minus 2, ut minus 2 plus rho et minus 1, ut minus 2 plus covariance between et and ut minus 2, the covariance between et and ut minus 2. So, what we have again this equals to 0, this equals to 0, we have rho square, sigma e square divided by 1 minus rho square.

And in a similar fashion, we can prove that when the covariance between ut and ut minus 2, then we have rho raise to the power 2. When it was ut and ut minus 1, we had rho raise to power 1. Similarly, when it is ut and ut minus j, it will be rho raise to the power j. So, now you can see that the covariance actually depends on the difference between the two time periods.

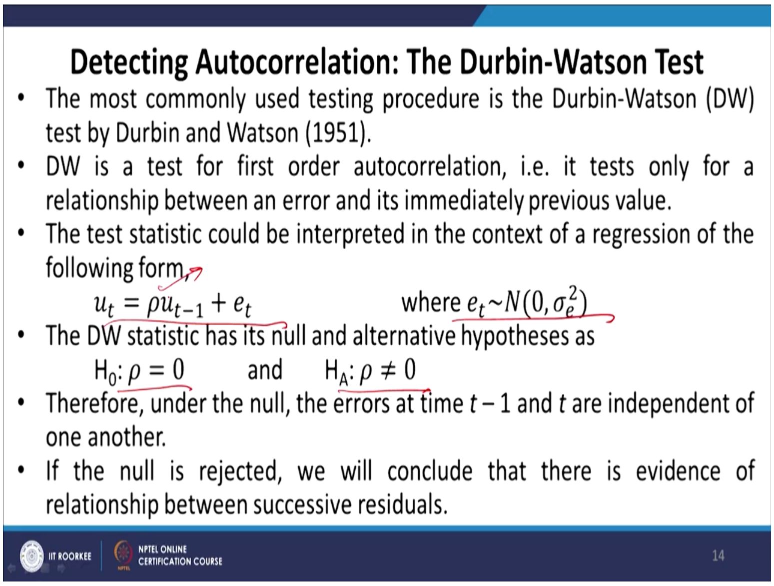
(Refer Slide Time: 17:39)



So, this is my variance-covariance matrix *(refer to slide time 17:39)* where all the diagonal terms are constant and equal to sigma e square divided by 1 minus rho square and the covariances are basically the variance term multiplied by rho raise to the power the difference between the two time period. So, this is actually covariance between ut and ut minus 1. So, the time period difference is 1. I have rho raise to power 1.

This will be covariance between ut and u1, the very first period, and that is why this is rho raise to the power t minus 1 and of course the variance term. So, that is how the entire variance-covariance matrix is derived. Since the variance of u equals to the expected value of uu prime is now not equal to sigma square In, the OLS estimates are not the most efficient one.

(Refer Slide Time: 18:32)

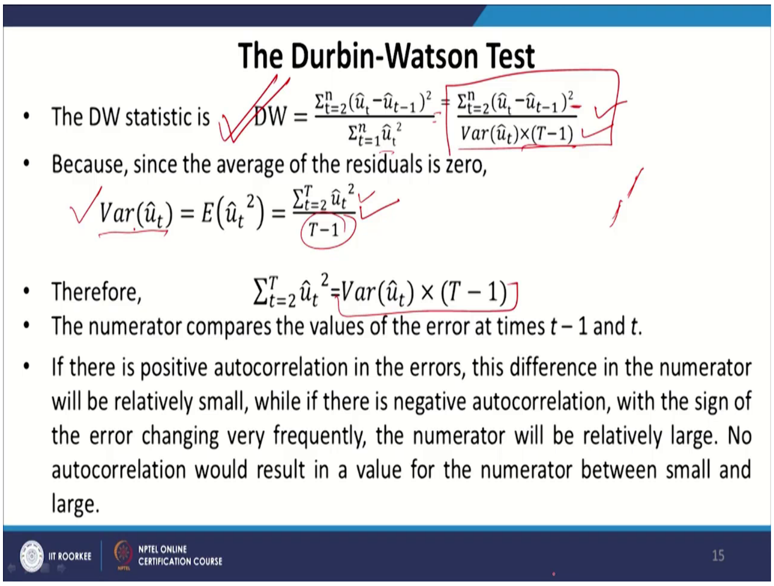


Now, how do I detect serial correlation or autocorrelation? The most popular or common test is the Durbin-Watson test. That was suggested by Durbin and Watson in 1951. DW is a test for first-order autocorrelation that is it tests only for a relationship between an error and its immediate previous value.

So, if there is higher-order autocorrelation, then that is not detectable by DW test. The test statistic could be interpreted in the context of a regression of the following form. So, this is our AR1 model of the error term, while et follows a normal distribution with 0 mean and sigma square e variance.

The DW statistic has its null and alternative hypothesis as rho equals to 0 and rho naught equals 0. So, the null hypothesis states that there is no autocorrelation between the error terms. Therefore, under the null, the errors at time t minus 1 and t are independent of one another. If the null is rejected, we will conclude that there is evidence of a relationship between successive residuals.

(Refer Slide Time: 19:43)



DW statistic looks like this *(refer to slide time 19:43)*. We do not get into the derivation of DW statistics, but then we will certainly get into the implications of the DW statistic. Now, this can also be written as something like this where you can see that the numerator remains the same, but the denominator is now slightly different. And this is obvious because the variance of ut hat is expected value of ut hat square which is equal to summation t equals 2 to capital T ut square divided by T minus 1. This is basically an unbiased estimator of the population error variance.

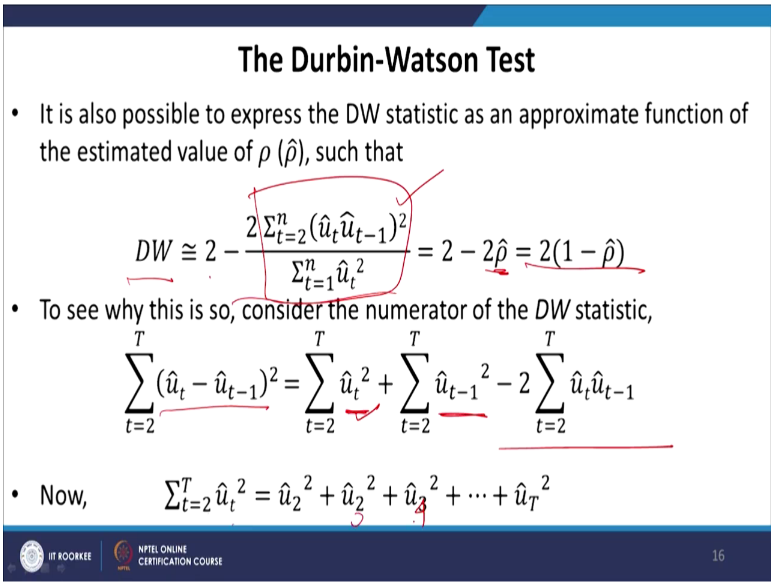
Therefore, we can always write variance of ut hat multiplied by t minus 1, which is the right-hand side here is equal to summation ut hat square and this is actually the denominator here. We are going to utilize this expression later and that is why I have mentioned it here. The numerator compares the values of the error at time t minus 1 and t.

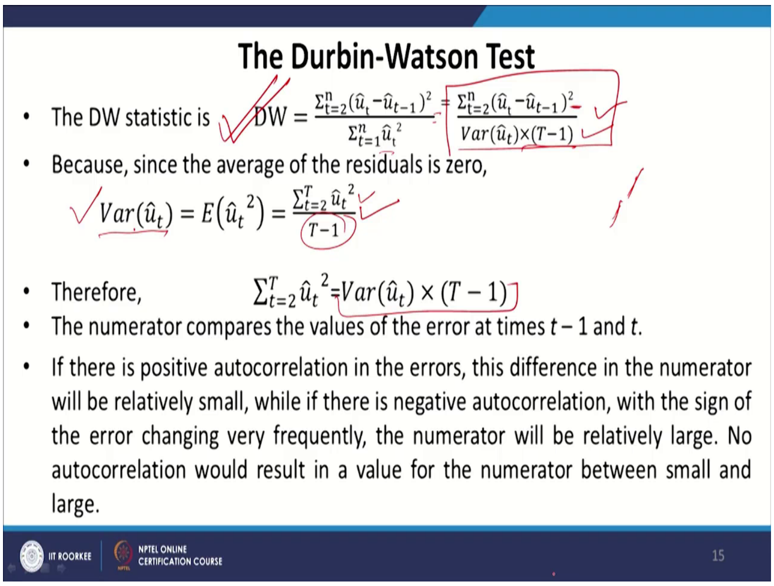
If there is positive autocorrelation in the errors, the difference in the numerator will be relatively small, because one is following the other. One is increasing, the other is also increasing. One is decreasing, the other is also decreasing. So, the difference between the consecutive numbers would be small.

While if there is negative autocorrelation with the sign of the error changing very frequently, the numerator will be relatively large. So, if one is coming down, the other one is going up, the one is going up, the other one is coming down. And as a result of which we expect specifically when the signs also change, the difference to be large.

No autocorrelation should result in a value for the numerator between small and large. So, for small values, we expect positive autocorrelation, for large values of the numerator we expect a negative autocorrelation. But note that the numerator will always be positive because it is actually a squared term. So, the difference could be positive or negative, but it is square that is why it is always positive and then they have summed up over n minus 2 observations.

(Refer Slide Time: 22:01)

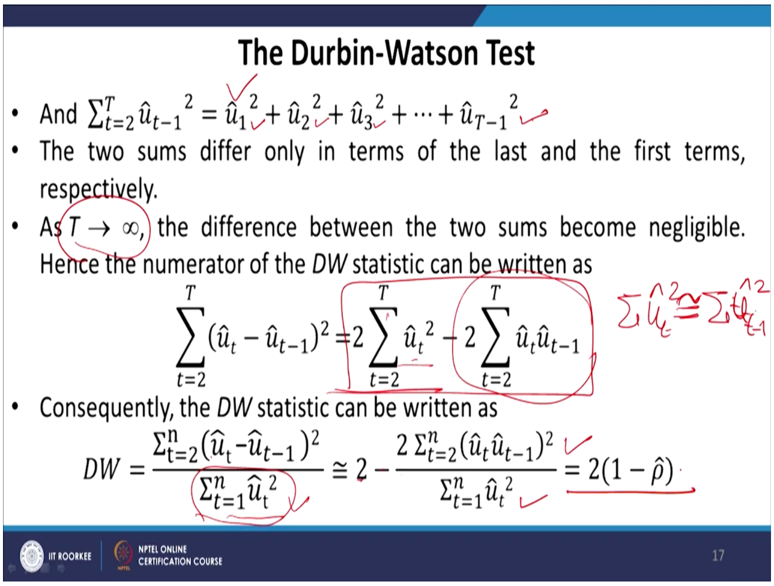


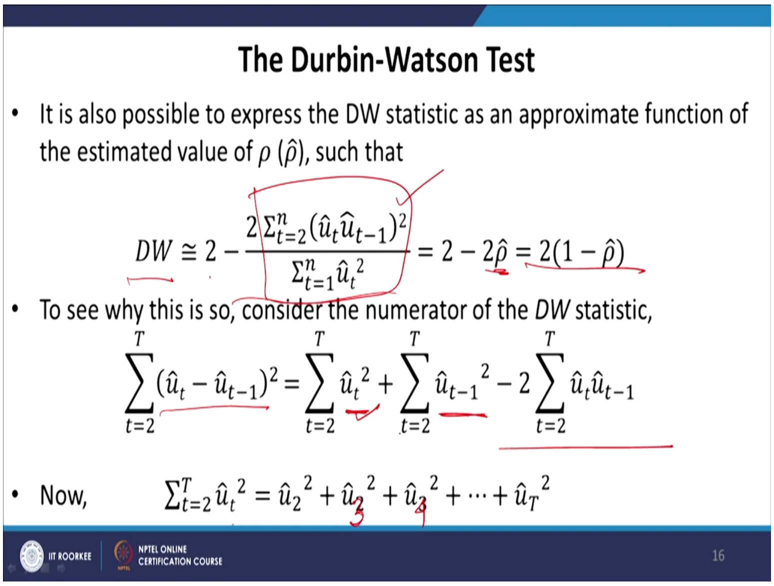


It is also possible to express the DW statistic as an approximate function of the estimated value of rho, which is rho hat, such that DW statistics is approximately equal to 2 minus this expression and this expression is equivalent to rho hat, the estimated parameter of the AR1, so which is equal to 2 into 1 minus rho hat.

So, we consider only the numerator of the DW statistic which is the original numerator. And what do you observe, we simply expand it like a minus b whole square. So, first of all, a square ut hat square, this is b square and then minus 2ab. Now, we consider the first component. This is actually the sum of the observations t running from 2 to capital T. So, u2 hat square u2, this should be u3 hat square then u4 hat square up to ut hat square.

(Refer Slide Time: 23:12)



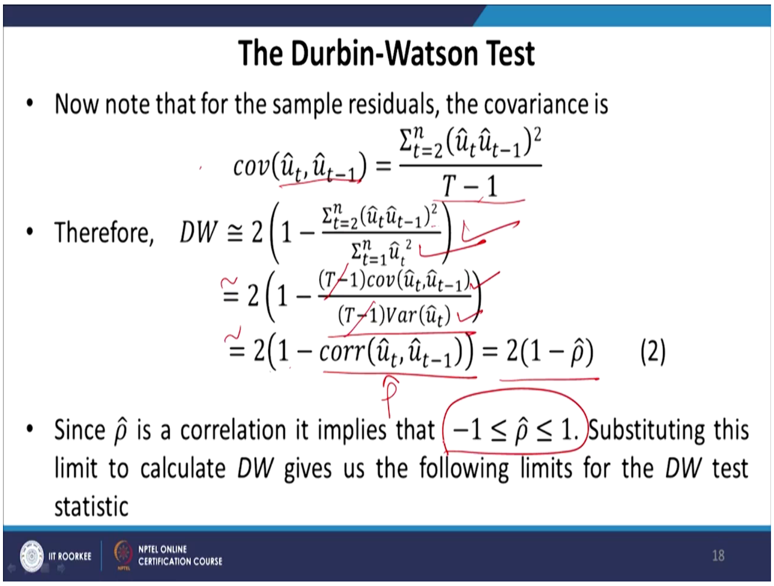


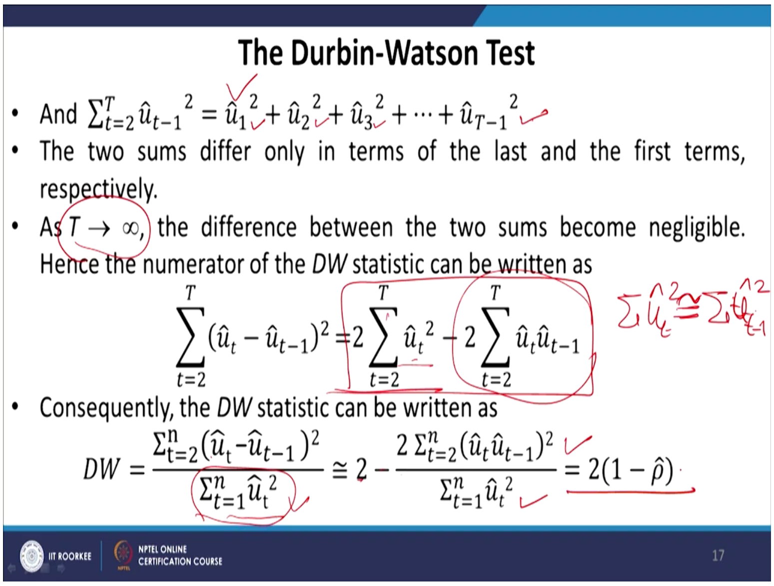
Similarly, when we consider the other term that is ut minus 1 again t running from 2 to capital T, we are having 2 minus 1. So, it starts from u1 hat square, u2 hat square, u3 hat square up to ut minus 1 hat square *(refer to slide time 23:12)*. So, you can see that the two sums differ only in terms of the last and the first terms.

So, only the last term of the first component and the first term of the second component is not there. So, as t tends to infinity that is the sample size becomes large and large, the difference between the two sums becomes negligible. Hence, the numerator of the DW statistic can be written as *(refer to slide time 23:12; the first slide)* 2 summation ut hat square, so we are assuming summation ut hat square and summation ut minus 1 hat square to be approximately equal. And that is why we are writing 2 summations ut hat square minus 2 ut hat ut minus 1 hat.

Consequently, the DW statistic can be written as *(refer to slide time 23:12; the first slide)* 2 minus, we are having this expression in the numerator, and in place of this, and in the denominator we have this. So, summation ut hat square and summation ut hat square roughly cancel out with each other. We have left with 2 minus this entire thing on the numerator divided by the usual denominator. And this basically is equal to 2 into 1 minus rho hat. Now, why this is so?

(Refer Slide Time: 25:01)



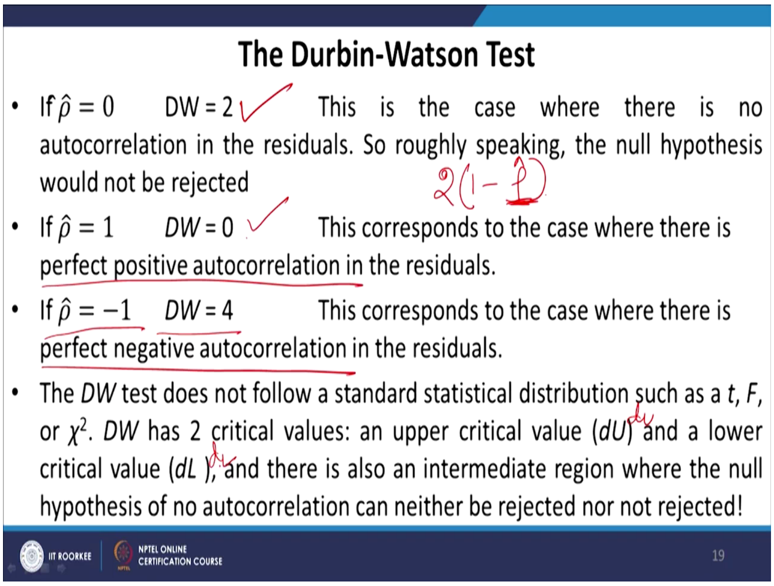


This is because we note that for the sample residuals, the covariance is the covariance between ut and ut minus 1 hat is summation ut hat ut minus 1 hat square divided by t minus 1. Now, this is the expression we actually derived here. Now, what we are doing is that in place of the numerator, we are writing t minus 1 multiplied by covariance ut hat ut minus 1, in place of the denominator, we are writing t minus 1 multiplied by the variance of ut hat, which I just derived a few a couple of slides ago.

So, T minus 1, T minus 1 cancels out *(refer to slide time 25:01; first slide)*. And covariance between ut hat and ut minus 1 hat divided by variance of ut hat gives us correlation between ut hat and ut minus 1 hat. So, this is rho hat. And that is how we have 2 into 1 minus rho hat. This is approximately equal to the initial expression.

Since rho hat is a correlation, it implies that rho hat would lie between minus 1 and plus 1, the usual correlation coefficients, as they always lie between minus 1 and plus 1. Substituting this limit to calculate DW gives us the limits for the DW test statistic.

(Refer Slide Time: 26:29)

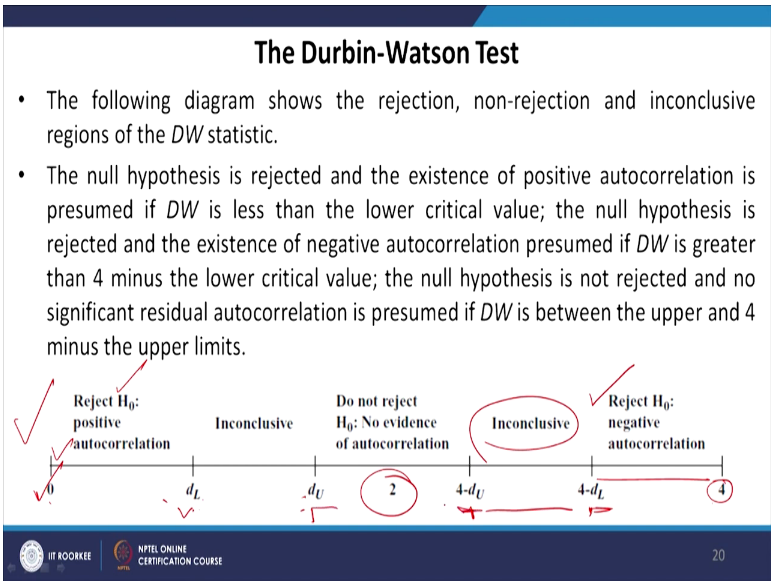


So, if rho is equal to 1, I just write for your convenience, what we observed is 2 into 1 minus rho hat. So, if rho hat is equal to 0, then DW equals to 2. This is the case where there is no autocorrelation in the residuals. So, roughly speaking, the null hypothesis will not be rejected. So, also rho hat implies no correlation or no autocorrelation. So, rho hat equals to 0 would give us a DW statistic, which is equal to 2, and this would correspond to no autocorrelation between the error terms.

When rho hat is equal to 1, we have DW statistic equal to 0. So, when there is perfect autocorrelation, perfect positive autocorrelation in the residuals, then the corresponding DW statistic is 0. And when rho hat is equal to minus 1, we have DW statistic equal to 4. This corresponds to the case when there is perfect negative autocorrelation.

DW test does not follow a standard statistical distribution such as t, F, or chi-square. DW has two critical values; an upper critical value denoted by dU and a lower critical value denoted by dL. We can also write it as d subscript U and d subscript L. And there is also an intermediate region where the null hypothesis of no autocorrelation can neither be rejected nor not rejected.

(Refer Slide Time: 28:05)

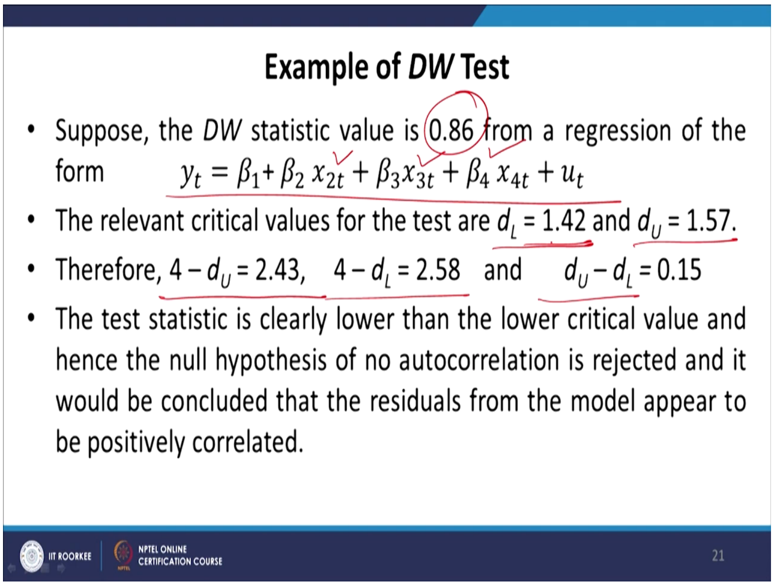


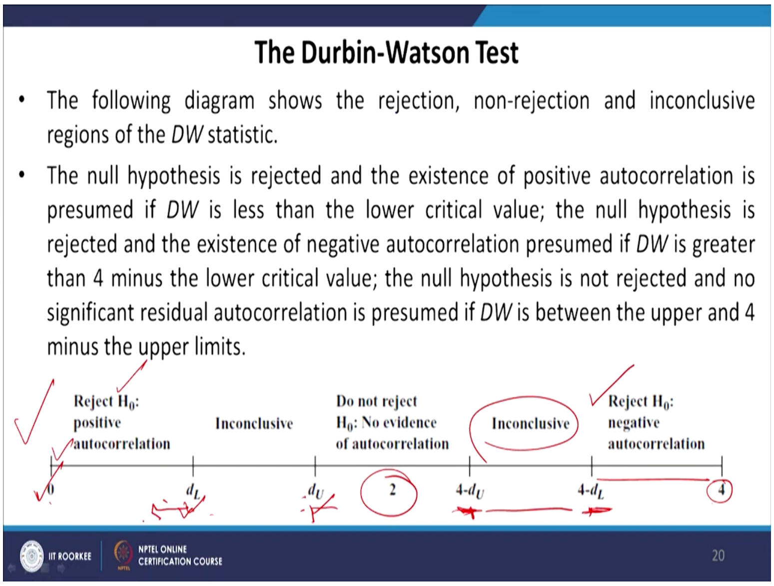
So, the diagram *(refer to slide time 28:05)*, shows the rejection, non-rejection, and inconclusive regions of the DW statistic. So, 2 is basically where we do not reject the null hypothesis, dU to 4 minus dU gives us this region. Then if we have 4, rho is equal to, dw equals 4 that is a perfect negative correlation and that is actually valid up to 4 minus the dL, lower DW statistic value.

So, for this region, we reject the null hypothesis. So, we have negative autocorrelation. And for this region, 0 to the DW lower limit, we reject again the null hypothesis and it shows positive autocorrelation with 0 being associated with perfect positive autocorrelation. Now, between dU and dL there is a region of inconclusion, and similarly, between 4 minus dL and 4 minus dU again we have an inconclusive region.

So, if the null hypothesis is rejected and the existence of positive autocorrelation is presumed, if DW is less than the lower critical value, the null hypothesis is rejected and the existence of negative autocorrelation is presumed if DW is greater than 4 minus the lower critical value. The null hypothesis is not rejected and no significant residual autocorrelation is presumed if DW is between the upper and 4 minus the upper critical limit.

(Refer Slide Time: 29:46)

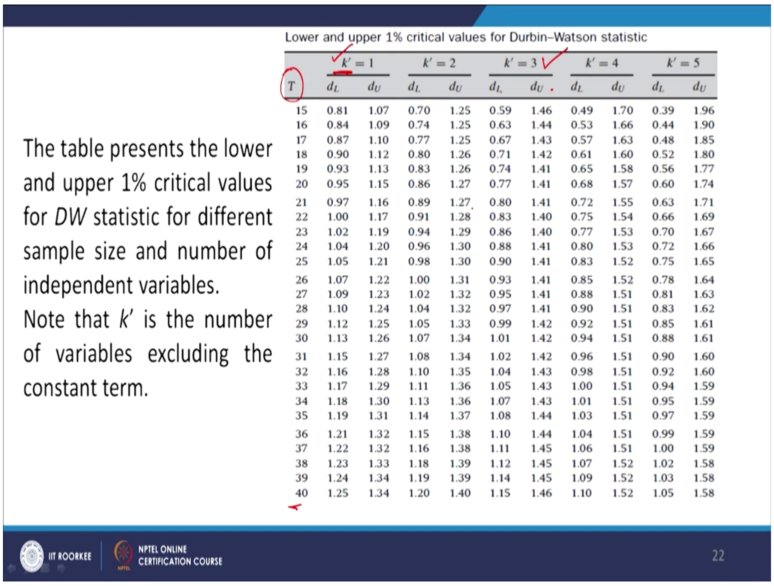




Now, suppose the DW statistic value is 0.86 from a regression of the form yt equals beta 1 plus beta 2. So, we have basically three explanatory variables. The relevant critical values for the test are dL at 1.42 and dU at 1.57. Therefore, 4 minus dU is 2.43, 4 minus dL is 2.58. So, 4 minus dU we calculate, 4 minus dL we calculate. We already have values of dL and dU. So, dU minus dL is 0.15. *(refer to slide time 29:46)*

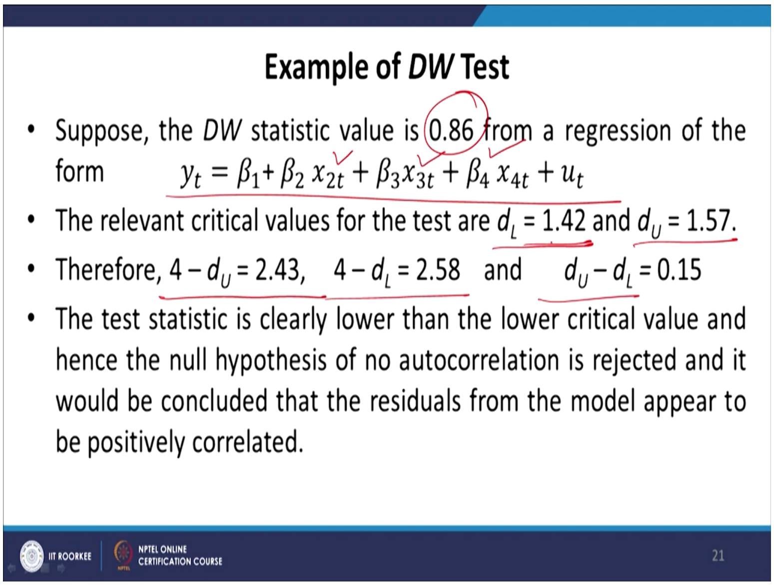
Now, our test statistic says that the value is 0.86. So, the test statistic is clearly lower than the lower critical value because dL is at 1.42. Hence, the null hypothesis of no autocorrelation is rejected and it would be concluded that the residuals from the model appear to be positively correlated.

(Refer Slide Time: 30: 49)



Lower and upper 1% critical values for Durbin–Watson statistic

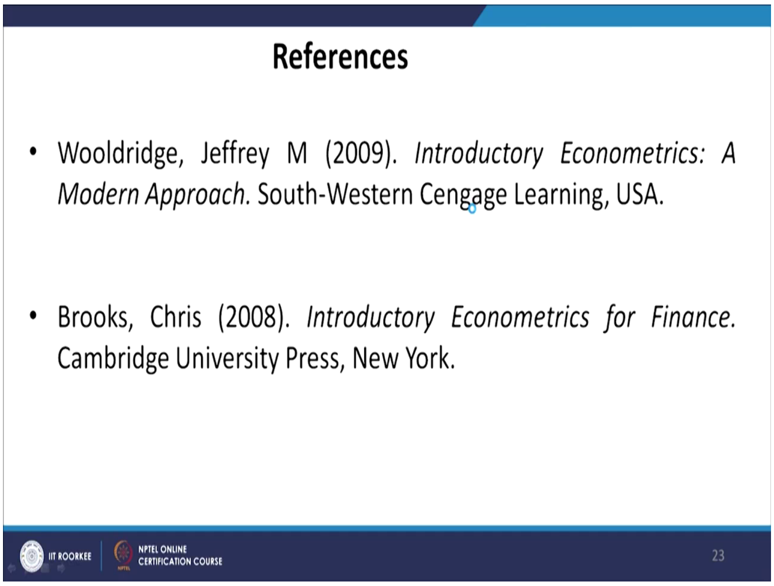
T	$k'=1$ d_L	$k'=1$ d_U	$k'=2$ d_L	$k'=2$ d_U	$k'=3$ d_L	$k'=3$ d_U	$k'=4$ d_L	$k'=4$ d_U	$k'=5$ d_L	$k'=5$ d_U
15	0.81	1.07	0.70	1.25	0.59	1.46	0.49	1.70	0.39	1.96
16	0.84	1.09	0.74	1.25	0.63	1.44	0.53	1.66	0.44	1.90
17	0.87	1.10	0.77	1.25	0.67	1.43	0.57	1.63	0.48	1.85
18	0.90	1.12	0.80	1.26	0.71	1.42	0.61	1.60	0.52	1.80
19	0.93	1.13	0.83	1.26	0.74	1.41	0.65	1.58	0.56	1.77
20	0.95	1.15	0.86	1.27	0.77	1.41	0.68	1.57	0.60	1.74
21	0.97	1.16	0.89	1.27	0.80	1.41	0.72	1.55	0.63	1.71
22	1.00	1.17	0.91	1.28	0.83	1.40	0.75	1.54	0.66	1.69
23	1.02	1.19	0.94	1.29	0.86	1.40	0.77	1.53	0.70	1.67
24	1.04	1.20	0.96	1.30	0.88	1.41	0.80	1.53	0.72	1.66
25	1.05	1.21	0.98	1.30	0.90	1.41	0.83	1.52	0.75	1.65
26	1.07	1.22	1.00	1.31	0.93	1.41	0.85	1.52	0.78	1.64
27	1.09	1.23	1.02	1.32	0.95	1.41	0.88	1.51	0.81	1.63
28	1.10	1.24	1.04	1.32	0.97	1.41	0.90	1.51	0.83	1.62
29	1.12	1.25	1.05	1.33	0.99	1.42	0.92	1.51	0.85	1.61
30	1.13	1.26	1.07	1.34	1.01	1.42	0.94	1.51	0.88	1.61
31	1.15	1.27	1.08	1.34	1.02	1.42	0.96	1.51	0.90	1.60
32	1.16	1.28	1.10	1.35	1.04	1.43	0.98	1.51	0.92	1.60
33	1.17	1.29	1.11	1.36	1.05	1.43	1.00	1.51	0.94	1.59
34	1.18	1.30	1.13	1.36	1.07	1.43	1.01	1.51	0.95	1.59
35	1.19	1.31	1.14	1.37	1.08	1.44	1.03	1.51	0.97	1.59
36	1.21	1.32	1.15	1.38	1.10	1.44	1.04	1.51	0.99	1.59
37	1.22	1.32	1.16	1.38	1.11	1.45	1.06	1.51	1.00	1.59
38	1.23	1.33	1.18	1.39	1.12	1.45	1.07	1.52	1.02	1.58
39	1.24	1.34	1.19	1.39	1.14	1.45	1.09	1.52	1.03	1.58
40	1.25	1.34	1.20	1.40	1.15	1.46	1.10	1.52	1.05	1.58



This is a truncated DW statistic model for lower and upper 1 percent critical values for the Durbin-Watson statistic. And this actually truncated because the observations T is here only up to 40 because this slide can accommodate only up to this many observations, but we can have DW statistics for larger observations as well. And here k prime actually refers to the number of independent variables. So, when we consider DW statistics, the k value actually does not include the intercept term.

So, for example, here I have three independent variables. So, that we would refer to this column *(refer to slide time 30:49; the first slide)*. Consider the dL and dU values corresponding to the number of observations we have and accordingly we can make decisions depending on the critical or the calculated DW statistic we have obtained.

(Refer Slide Time: 31:49)



So, these are the references. In the next module also, I will continue with the discussion of serial correlation while discussing another test and how to deal with the serial correlation that is how we can correct for the presence of serial correlation so that my estimated parameters are efficient. Thank you.