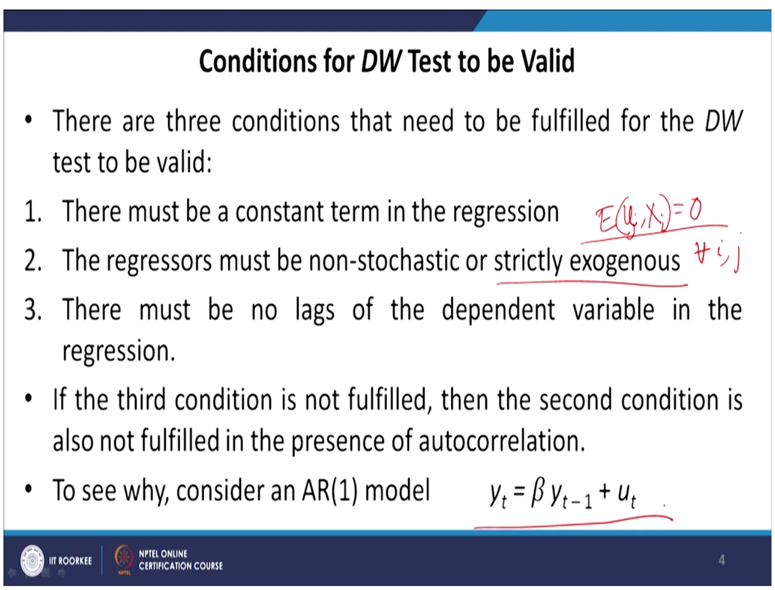
Econometric Modelling
Professor Sujata Kar
Department of Management Studies,
Indian Institute of Technology Roorkee
Lecture 25
Problem of Serial Correlation-II

This is Module 25 of Econometric Modelling. In 24, we have started discussing the problem of serial correlation. And there we defined a serial correlation, what kind of problem it poses to OLS estimation methods. And in this module 25 also we continue with the problem of serial correlation. We have already discussed one test for detecting serial correlation that is the DW statistic or DW test, Durbin-Watson test.

The test is designed to detect serial correlation only of first order. Now, we will discuss a few features or problems with the DW test and after that we will discuss alternative testing procedure and how to deal with autocorrelation or serial correlation that if there is, if the errors are serially correlated then how can we still obtain efficient OLS estimation or estimators.

(Refer Slide Time: 01:29)



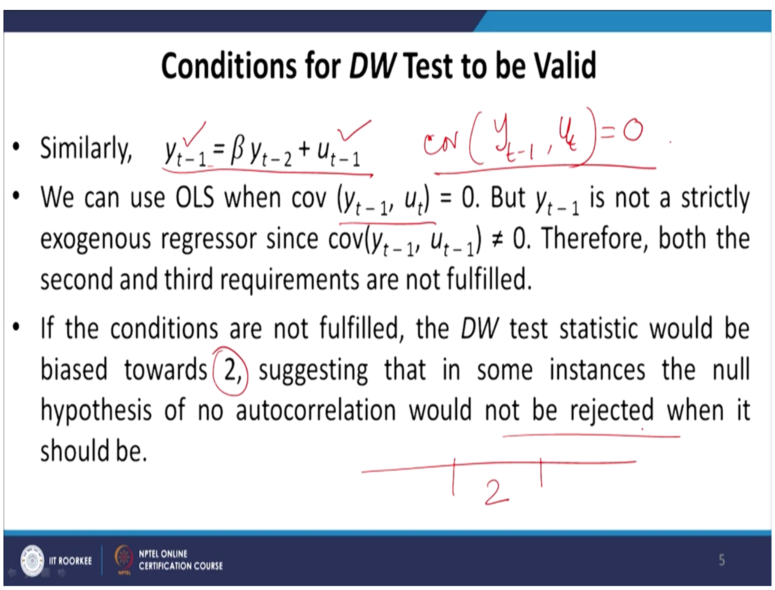
So, first of all, we discuss the conditions that are required for DW test to be valid. So, there are three conditions that need to be valid for the DW test. There must be a constant term in the regression. So, we cannot run a regression without a constant term. Second thing is that the regressors must be non-stochastic or strictly exogenous.

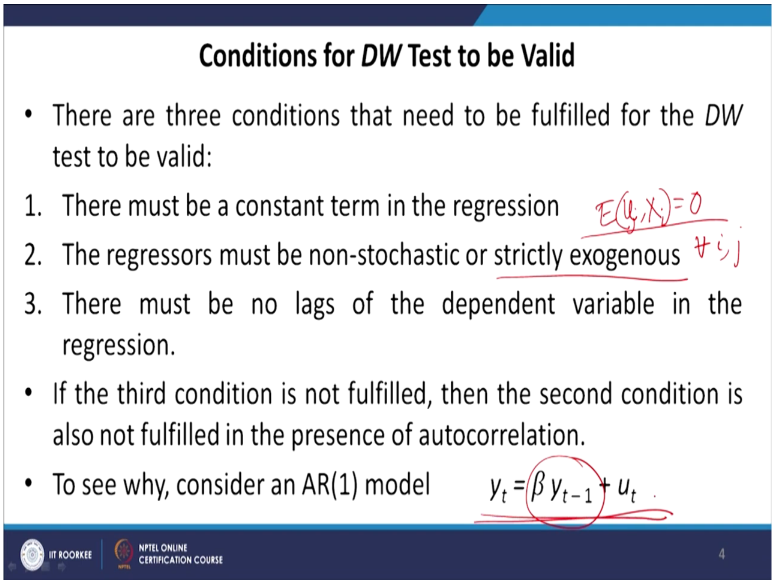
So, this actually refers to one of the CLRM a, where we assume that an alternative is to consider that the expected value of u and x must be 0, that is the covariance between the error terms and the independent variables as 0, strict exogenity means, they are not only contemporaneously not correlated, they are not correlated altogether.

So, i and j, uj, and xi are equal to 0 for all i and j. So, this is strict exogenity. Along with that, alternatively, we can assume non-stochastic which is non-stochastic independent variables alternatively which can be said that the independent variables are fixed in repeated samples. There must be no lags of the dependent variable present in the regression.

So, that is the third condition that is required for the validity of the DW test. Now, note that if the third condition is not fulfilled then the second condition is also not fulfilled in the presence of autocorrelation. To see why, consider an AR (1) model yt equals beta yt minus 1 plus ut.

(Refer Slide Time: 03:23)



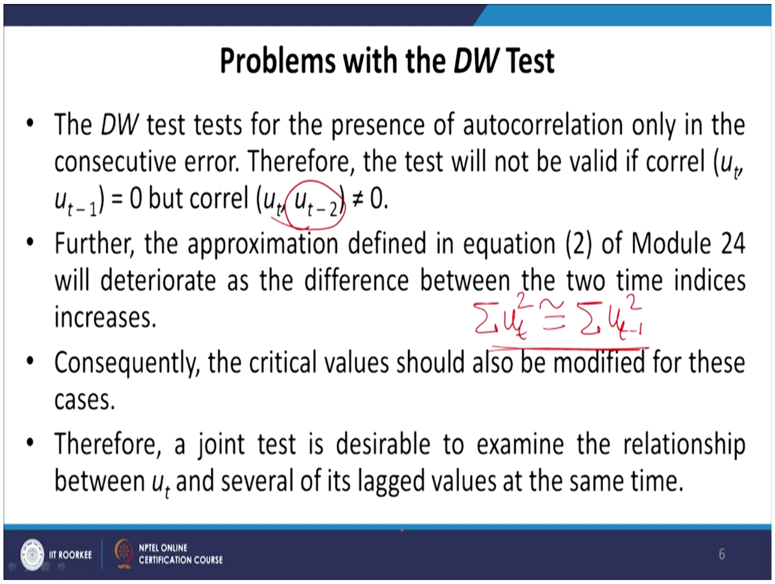


Similarly, we can also have the same structure like yt minus 1 equals beta yt minus 2 plus ut minus 1. We can use OLS when covariance between yt minus 1 and ut is equals to 0. But you can see that yt minus 1 is not strictly exogenous, because covariance between yt minus 1 and ut minus 1 is actually not equal to 0. yt minus 1 is explained by ut minus 1. So, it is quite possible that yt minus 1 and ut are not correlated with each other. So, this is 0 *(refer to slide time 03:23).*

But since this is not equal to 0, then there is contemporaneous endogeneity or yt minus 1 is not contemporaneously exogenous, and as a result of which we do not have strict exogenity. Therefore, both the second and the third requirements are not fulfilled when we have one lagged dependent variable term in the regression. So, we must not have one lagged dependent in the equation regression in order to have a valid DW statistic result.

If the conditions are not fulfilled, the DW test statistic would be biased towards 2. That is, in some instances the null hypothesis of no autocorrelation would not be rejected when it should be actually. If you remember, we had this diagram where around 2 used to be the area of non-rejection of the null hypothesis. So, the bias is towards 2 implies that it would be difficult for us to reject the null hypothesis, while it should be.

(Refer Slide Time: 05:14)

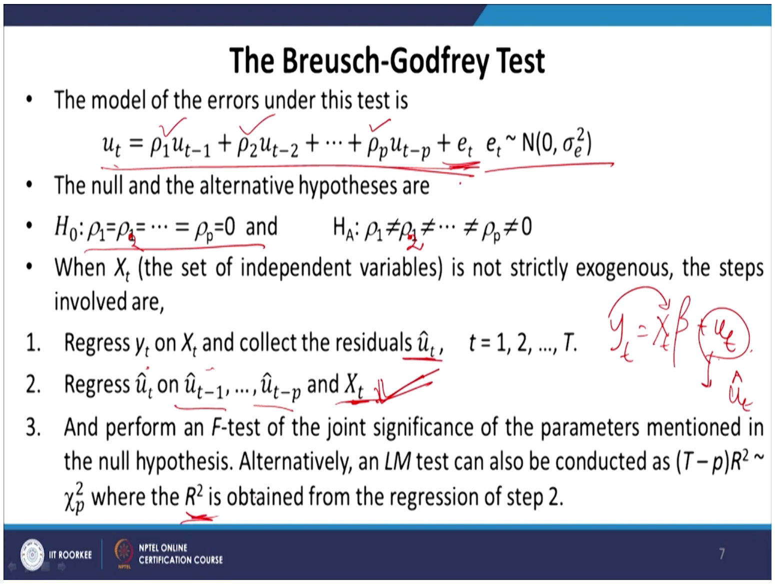


The DW tests for the presence of autocorrelation only in the consecutive error, which I have already just mentioned that, if it tests for only autocorrelation only in the first lag, the current one and the previous one. Therefore, the test will not be valid if the correlation between ut and ut minus 1 is equal to 0, but ut is correlated with ut minus 2 that is the error with two lags.

Further, the approximation is defined in equation 2 of module 24. So, the approximation that, if you remember, was like summation ut square was approximately equal to summation ut minus 1 squared. This approximation deteriorates as the difference between the two-time indices increase. So, this is approximately equal, but if I consider t minus 2, t minus 3, so the time periods are increasing and this approximation becomes gradually invalid or deteriorate.

Consequently, the critical values should also be modified for these cases. Therefore, a joint test is desirable to examine the relationship between ut and several of its lagged values at the same time. So, we should consider an alternative test that would be able to consider multiple autocorrelations between one current period and several alternative lags.

(Refer Slide Time: 06:53)



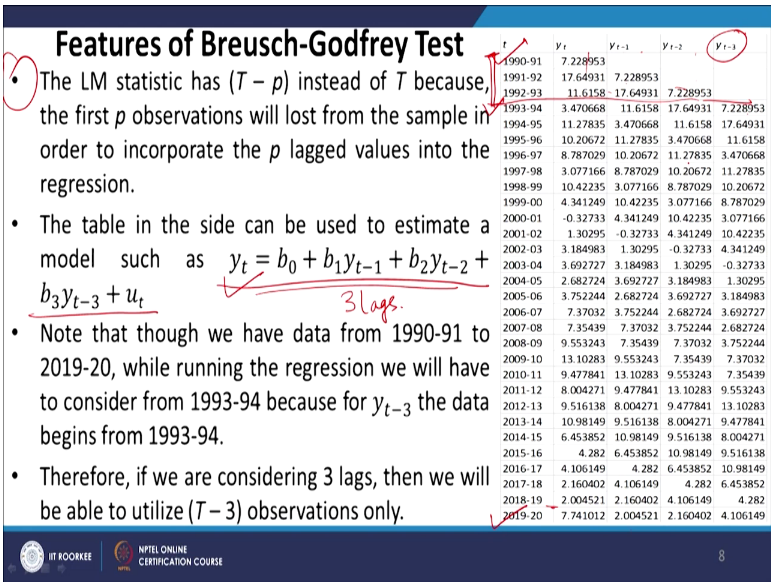
So, that kind of test is The Breusch-Godfrey test. The model of the errors under this test is actually p lag sequence, suppose it is a p lag sequence where ut equals rho 1 ut minus 1 rho 2 ut minus 2 and so on up to plus rho p ut minus p and we also have an error term et which is normally distributed with mean 0 and constant variance sigma e square.

The null and the alternative hypothesis are, we are actually going for joint significance of all these parameters. So, rho 1 equals to rho 2 equals to rho p and so on all equals 0. And similarly, we have then alternative hypothesis which is rho 1 is not equal to rho 2 not equal to rho 3, and so on up to rho p not equal to 0. When Xt, the set of independent variables, is not strictly exogenous, the steps involved are; first of all regress yt on Xt and collect the residuals.

So, this is my original model *(refer to slide time 06:53)*. If the original model is yt equals Xt beta plus ut then first we regress yt on Xt and collect ut, then regress ut hat that is the residual obtained from this regression is denoted by ut hat, we regress ut hat on ut hat ut minus 1 hat to up to ut minus p hat. Since we cannot see the population errors, this is actually replaced with their sample counterparts. And along with that, we also include Xt.

Now, here since Xt is not a strict exogenous that is why Xt is also incorporated into this framework. And perform an F-test of the joint significance of the parameters mentioned in the null hypothesis. Alternatively, an LM test can also be conducted as T minus p R square where it is distributed as chi-square distribution with p degrees of freedom. And this R square is obtained from the regression in step 2, that is from this regression this R square is obtained. So, this is actually an auxiliary R square.

(Refer Slide Time: 09:15)



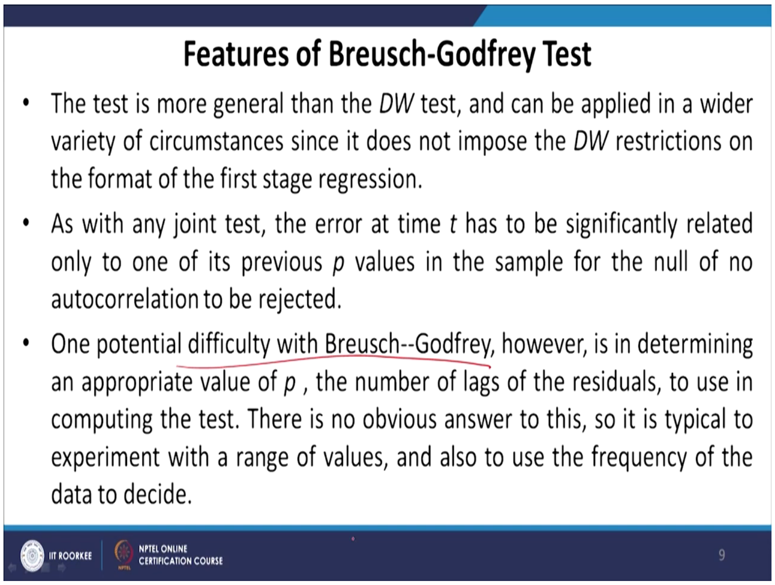
t	y_t	y_{t-1}	y_{t-2}	y_{t-3}
1990-91	7.228953			
1991-92	17.64931	7.228953		
1992-93	11.6158	17.64931	7.228953	
1993-94	3.470668	11.6158	17.64931	7.228953
1994-95	11.27835	3.470668	11.6158	17.64931
1995-96	10.20672	11.27835	3.470668	11.6158
1996-97	8.787029	10.20672	11.27835	3.470668
1997-98	3.077166	8.787029	10.20672	11.27835
1998-99	10.42235	3.077166	8.787029	10.20672
1999-00	4.341249	10.42235	3.077166	8.787029
2000-01	-0.32733	4.341249	10.42235	3.077166
2001-02	1.30295	-0.32733	4.341249	10.42235
2002-03	3.184983	1.30295	-0.32733	4.341249
2003-04	3.692727	3.184983	1.30295	-0.32733
2004-05	2.682724	3.692727	3.184983	1.30295
2005-06	3.752244	2.682724	3.692727	3.184983
2006-07	7.37032	3.752244	2.682724	3.692727
2007-08	7.35439	7.37032	3.752244	2.682724
2008-09	9.553243	7.35439	7.37032	3.752244
2009-10	13.10283	9.553243	7.35439	7.37032
2010-11	9.477841	13.10283	9.553243	7.35439
2011-12	8.004271	9.477841	13.10283	9.553243
2012-13	9.516138	8.004271	9.477841	13.10283
2013-14	10.98149	9.516138	8.004271	9.477841
2014-15	6.453852	10.98149	9.516138	8.004271
2015-16	4.282	6.453852	10.98149	9.516138
2016-17	4.106149	4.282	6.453852	10.98149
2017-18	2.160402	4.106149	4.282	6.453852
2018-19	2.004521	2.160402	4.106149	4.282
2019-20	7.741012	2.004521	2.160402	4.106149

Now, we talk about certain features of the Breusch–Godfrey test. The first thing is that the LM statistic has T minus p instead of T observations, that is we multiplied it by T minus p and not T, because the first p observations will be lost from the sample in order to incorporate the p lagged values into the regression. So, this is what is explained using the table *(refer to slide time 09:15)*.

The table can be used to estimate a model such as yt equals b naught plus b1 yt minus 1 plus b2 yt minus 2 b3 yt minus 3 plus ut. So, we are considering basically three lags of the dependent variable. Now, this is just an example that why we are going to lose three observations here or in case it is a ARP series then there are p lagged values then why we are going to lose p observations. This example has actually nothing to do with autocorrelation as such.

Note that though we have data from 1990-91 to 2019-20, while running the regression we will have to consider data from 1993 onwards. This is because for the variables yt minus 3 the data begins from 1993-94. And that is why for the other variable also these observations cannot be used *(refer to slide time 09:15*. Now, what we are doing that has already been shown to you earlier that this actually shifts down by one column. So, yt minus 1, the entire series up to this much shifts one row down, and similarly, yt minus 2 it shifts two rows down and things like that.

(Refer Slide Time: 11:08)



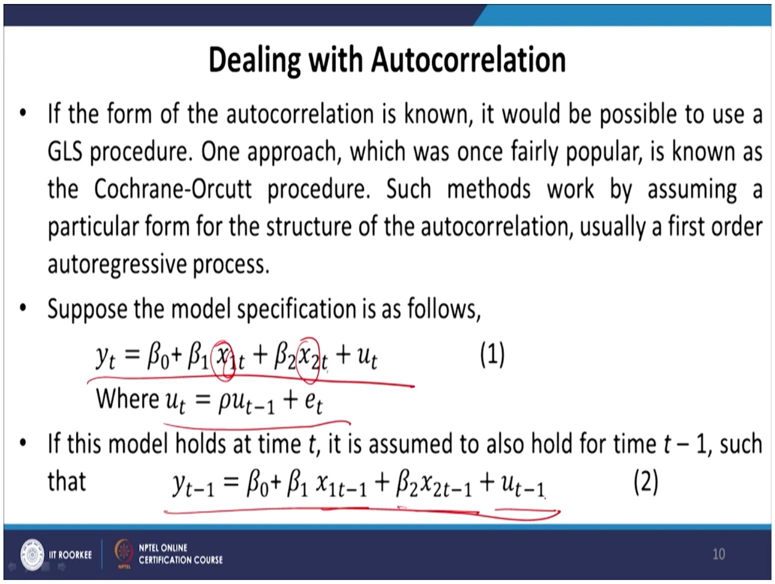
The second feature is that the test is more general than the DW test and can be applied in a wider variety of circumstances since it does not impose the DW restrictions on the format of the first-stage regression. As with any joint test, the error at time t has to be significantly related only to one of its previous p values in the sample for the null have no autocorrelation to be rejected, which basically implies that we really do not need to have all the rows equal to be 0 in order to not reject the null hypothesis.

Even if one of the rows is not equal to 0, we will not reject the null hypothesis. One potential difficulty with Breusch–Godfrey, however, is in determining an appropriate value of p that is the number of lags of the residuals to use in computing the test. There is no obvious answer to this. So, it is typical to experiment with that range of values and also to use the frequency of the data to decide how many lags to be included.

For example, if we are using monthly data, then we can go for 12 lags in order to see that how the months are correlated, whether there has been any autocorrelation between the months or so.

If we using quarterly data, then of course at least four lags to be included and so on. These, I am talking about the minimum number of lags to be included. So, this is one potential difficulty with Breusch–Godfrey test that how many lags to be included.

(Refer Slide Time: 12:51)



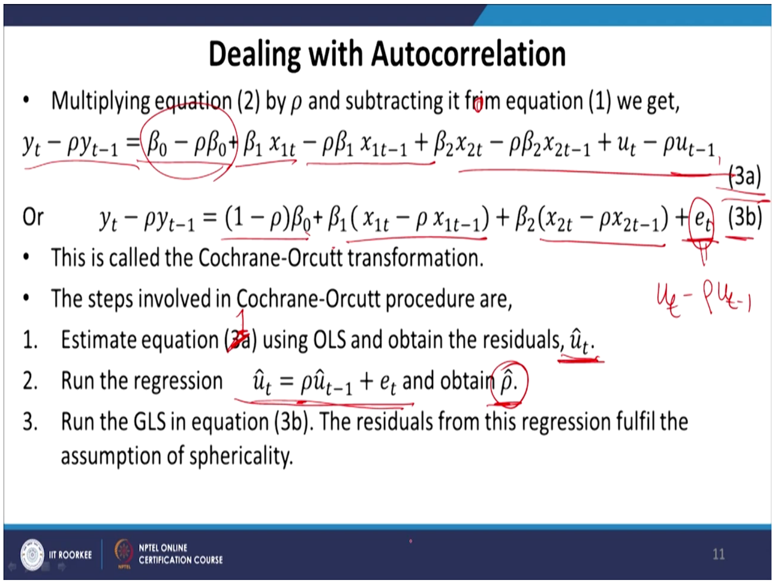
Now, we come to the question of how to deal with autocorrelation. So, if autocorrelation is detected, how do you problem this or how do you solve the problem because we know that autocorrelation or the presence of autocorrelation renders the OLS estimates inefficient. They are no more the best estimators. So, we should go for some corrections.

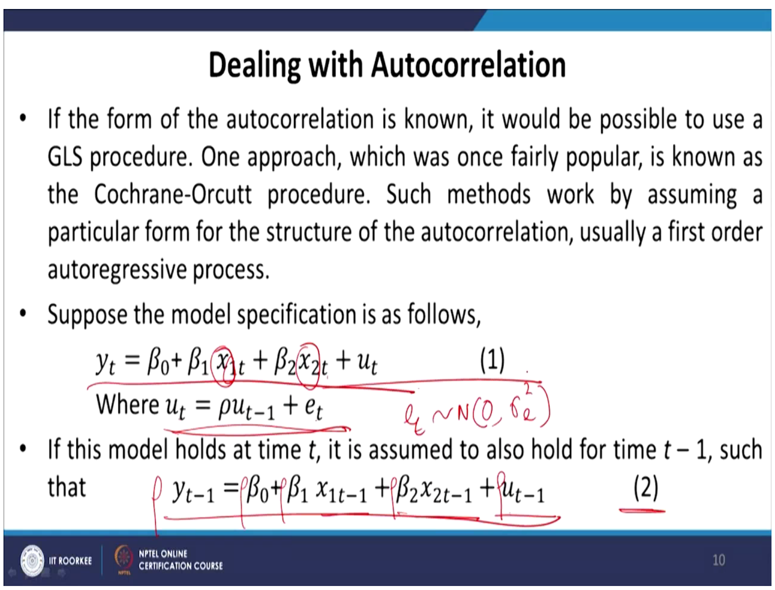
In this context, if you remember, the way we incorporated corrections or transformations in the original series or we call them either generalized least squares or GLS or feasible GLS that is FGLS similar procedure would be also followed here. So, if the form of the autocorrelation is known, it would be possible to use a GLS procedure. One approach which was once fairly popular is known as a Cochrane–Orcutt procedure or transformation.

Such methods work by assuming a particular form of the structure of the autocorrelation, usually a first-order autoregressive process. Now, suppose the model specification is as follows *(refer to slide time 12:51; equation 1)*. So, it is the original model, which has two independent variables. This is x1 reference to the first independent variable, x2 reference to the second independent variable and all of them have t observations.

Now, this ut is a first-order auto-correlated series, ut equals rho ut minus 1 plus et. If this model holds at time t, it is assumed to also hold for time t minus 1 such that we have yt minus 1 equals to this thing where we are replacing xt with x1 t minus 1 and x2 with x2 t minus 1, ut with ut minus 1.

(Refer Slide Time: 14:42)



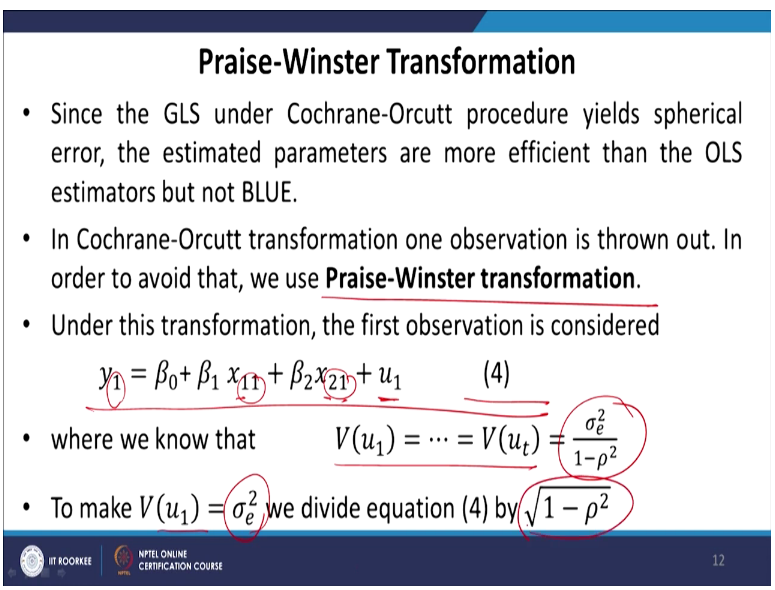


Now, multiplying equation 2, by rho and subtracting it from equation 1, what do we get? We are multiplying equation 2 with rho, then we are subtracting it from equation 1, we will be getting yt minus, rho yt minus 1 beta naught minus rho beta naught, beta 1 x1t, the original one, minus rho beta 1 x1t minus 1 and so on. We call it equation 3 *(refer to slide time 14:42)*. We rearrange the terms and write them as 1 minus rho multiplied by beta naught, this is what it is, beta 1 comes out and we have this thing.

Similarly, beta 2 multiplied by x2t minus x2t minus 1 and finally, et, et is equal to ut minus, rho ut minus 1 which actually follows directly from the relation. You can see et is ut minus rho ut minus 1 and ut, et is actually our white noise process or it follows a normal distribution with 0 mean and sigma square e, the variance. So, the moment we have et as the error term, then at least my errors are not correlated, auto-correlated anymore and I can have efficient estimators.

This is called the Cochrane–Orcutt transformation. The steps involved in Cochrane–Orcutt procedure are; first of all estimate equation 1, obtain the residuals ut hat, then we run the regression of ut hat equals rho ut minus 1 hat plus et and obtain rho hat and run the GLS in equation 3b or 3a. Residuals from this regression fulfill the assumptions of sphericality that is there is no homoscedasticity and the errors are also not uncorrelated.

(Refer Slide Time: 16:44)

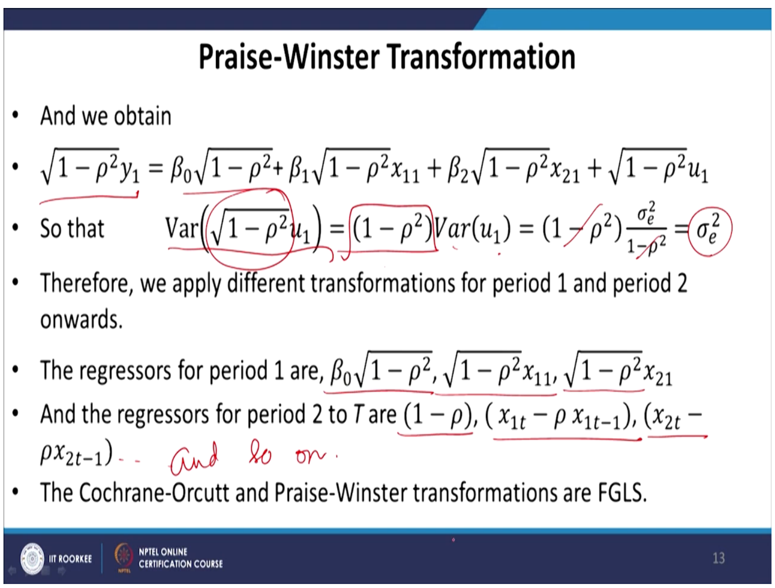


But there is one problem with this transformation. A problem is that since the GLS and a Cochrane–Orcutt procedure yields a spherical error, the estimated parameters are more efficient than the OLS estimators, but they are actually not BLUE indices. They are not the best linear unbiased estimators. In Cochrane–Orcutt transformation one observation is thrown out.

In order to avoid that we use another transformation only for the first observation and that is called the Praise-Winster transformation. Under this transformation, the first observation is considered which is, again we consider this the first observation so that is why we have y1, x11, and x21 that is corresponding to the first variable and the second variable the first observation, similarly u1, where we know that variance of u1 equals to variance of u2 equals to variance of ut all equals to sigma e square divided by 1 minus rho square.

To make the variance of u1 equals to sigma e square, that is it is actually a constant term and it is also the minimum variance possible, we divide equation 4 by root over 1 minus rho square.

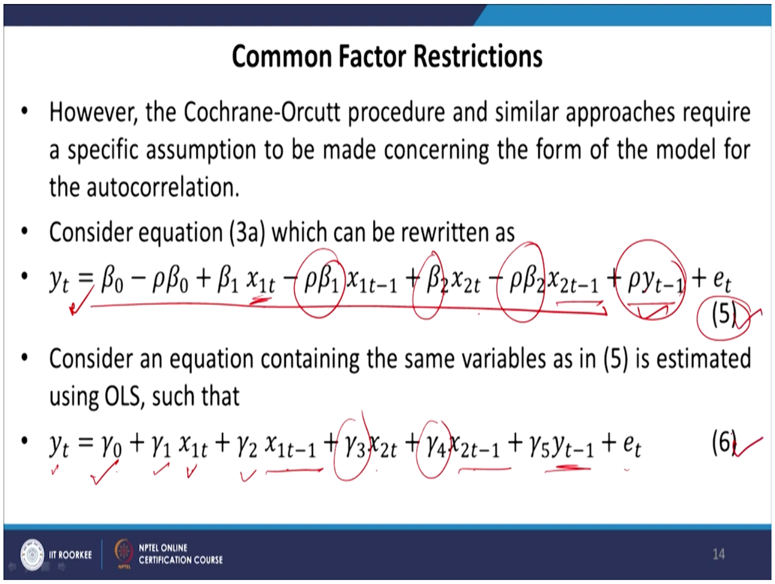
(Refer Slide Time: 18:04)

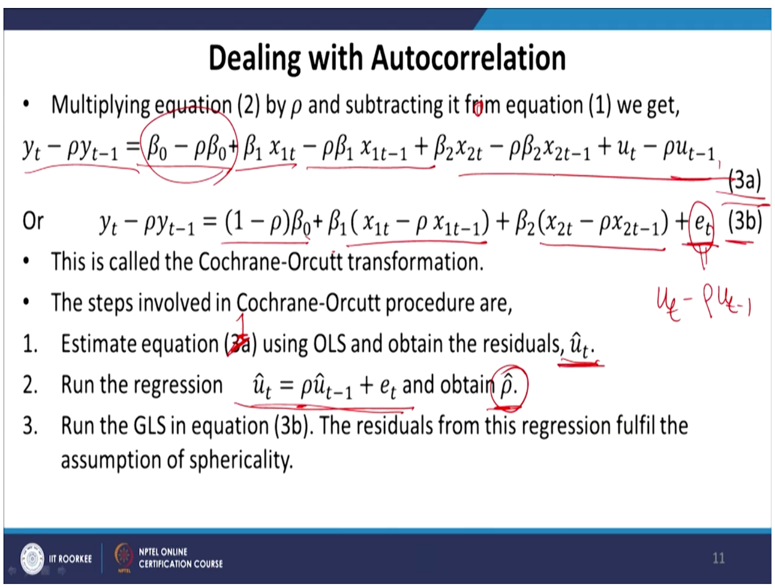


Then what happens. We have root over 1 minus rho square y1 into, equals to beta naught 1 minus rho square beta 1, 1 minus rho square, x11 and so on *(refer to slide time 18:04)*. So, that we have a variance of 1 minus rho square, root over 1 minus rho squared u1 equals 1 minus rho square, this thing is coming out as a square, then the variance of u1, so this cancels out and we have sigma e square. Therefore, we apply different transformations for period 1 and period 2 onwards.

The regressors for period 1 are beta naught multiplied by 1 minus rho square, 1 minus rho square multiplied by x11, and root over 1 minus rho square multiplied by x21 and the regressors for period 2 to T are 1 minus rho x1t minus rho x1t minus 1, x2t minus rho x2t minus 1 and so on. So, the Cochrane–Orcutt and Praise-Winster transformations are called feasible GLS or FGLS.

(Refer Slide Time: 19:18)



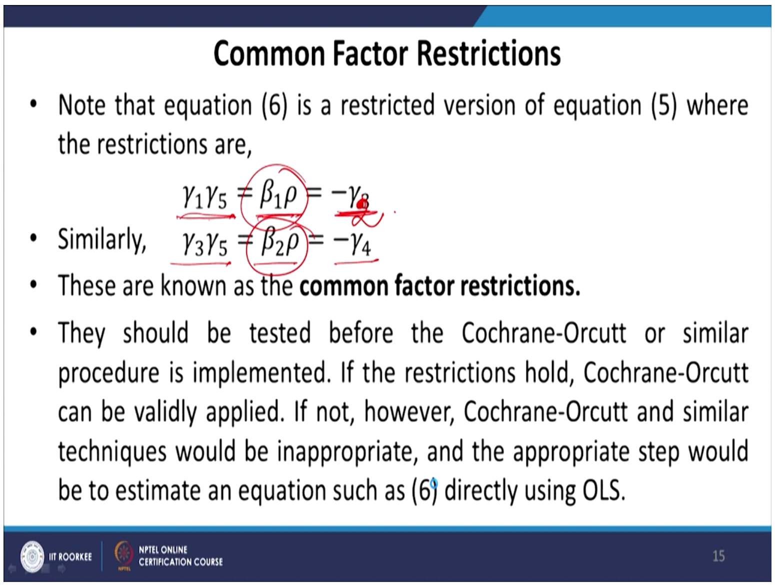


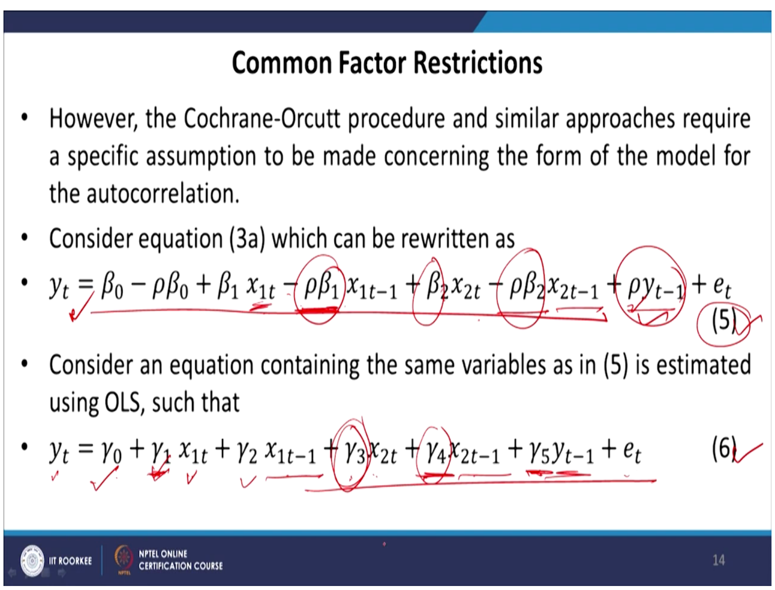
However, the Cochrane–Orcutt procedure and similar approaches require a specific assumption to be made concerning the form of the model for the autocorrelation. Consider equation 3a, this was my equation 3a, which can also be written as yt equals the everything remains the same, but only rho yt minus 1 has been shifted from the left-hand side to the right-hand side. So, it appears here as plus rho yt minus 1.

Now, consider an equation containing the same variables as in 5, this is equation 5, is estimated using OLS such that we have yt then we have gamma naught, gamma 1 x1t, gamma 2 x1t minus

1, gamma 3 x2t, gamma 4 x2t minus 1 and gamma 5 yt minus 1 plus et. So, if we are simply running a regression on 1, x1t, x1t minus 1, x2t, x2t minus 1, yt minus 1 and et then what are the equivalence we could draw between this equation 5 and equation 6 *(refer to slide time 19:18; first slide)*.

(Refer Slide Time: 20:42)





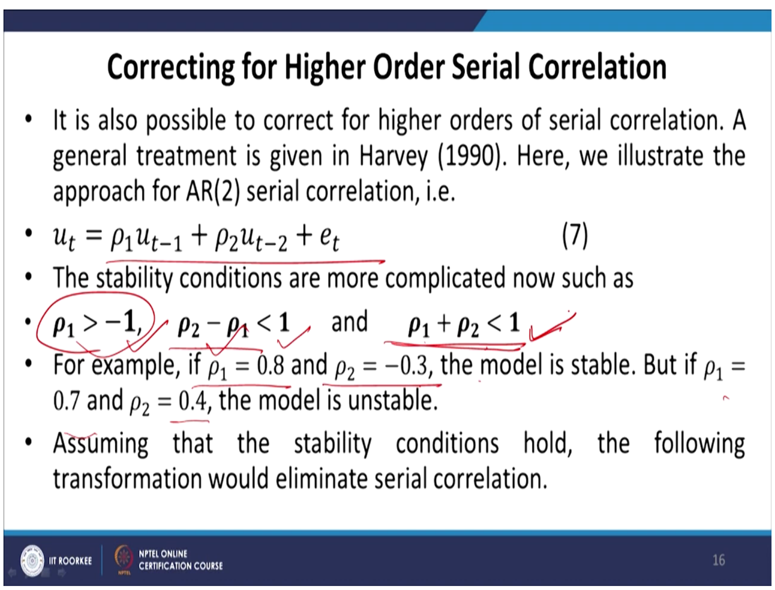
Note that equation 6 is a restricted version of equation 5 where the restrictions are rho 1 rho 5 is beta 1 rho. So, you can see that beta 1 rho, gamma 1, gamma 5 is beta 1 rho. So, beta 1 rho is gamma 1 and gamma 5, because gamma 5 is rho and gamma 1 is beta 1. So, rho beta 1 is gamma

1 and gamma 5. Now, there is a negative sign. So, we can write gamma 1 gamma 5 equals to beta 1 rho which is actually equal to minus gamma 3.

See this is my gamma 3 which corresponds to the, corresponds to this x2t, sorry, this is actually gamma 2, sorry, this is gamma 2. And similarly, gamma 2 gamma 3 is beta 2 rho and that equals to gamma 4. So, this is fine gamma 4, and this equals to minus rho beta 2 which is gamma 5 and gamma 3 *(refer to slide time 20:42)*. So, these are known as the common factor restrictions, where we are having basically certain restrictions on the parameter estimates.

They should be tested before the Cochrane–Orcutt or a similar procedure is implemented. If the restrictions hold, Cochrane–Orcutt can be validly applied. If not, however, Cochrane–Orcutt and similar techniques would be inappropriate and the appropriate step would be to estimate an equation such as 6 directly using OLS. So, we should go for OLS estimation of this kind of equation instead of applying Cochrane–Orcutt.

(Refer Slide Time: 22:37)

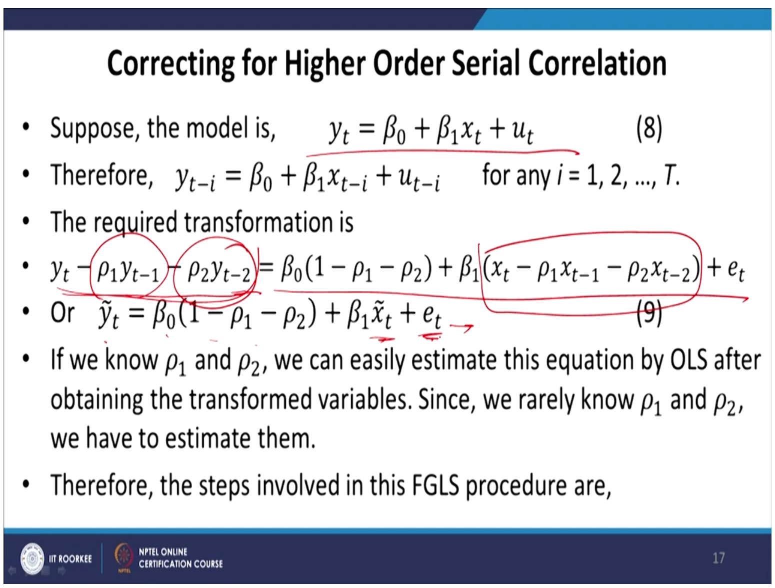


Now, how do we correct for higher-order serial correlations? It is also possible to correct for higher-order serial correlation. A general treatment is given in Harvey published in 1990. Here we illustrate the approach for AR(2) serial correlation for exposition, for simplicity of the exposition. So, this is an AR(2).

The stability conditions are more complicated now, such as rho 1 should be less than minus 1, rho 2 minus rho 1 is less than 1 and rho 1 plus rho 2 is also less than 1. For example, if rho 1 takes a value 0.8 and rho 2 takes a value minus 0.3 then you can see this condition is fulfilled and rho 2 minus rho 1 is also fulfilled and rho 1 plus rho 2 will also be fulfilled.

But if rho 1 equals 0.7 and rho 2 equals 0.5, then the model becomes unstable because this condition is not fulfilled, 0.7 plus 0.4 is now less than 1. Similarly, rho 2 is 0.4, this condition will be fulfilled, only this condition is not being fulfilled. Assuming that the stability conditions hold, the following transformation or eliminate serial correlation.

(Refer Slide Time: 24:01)

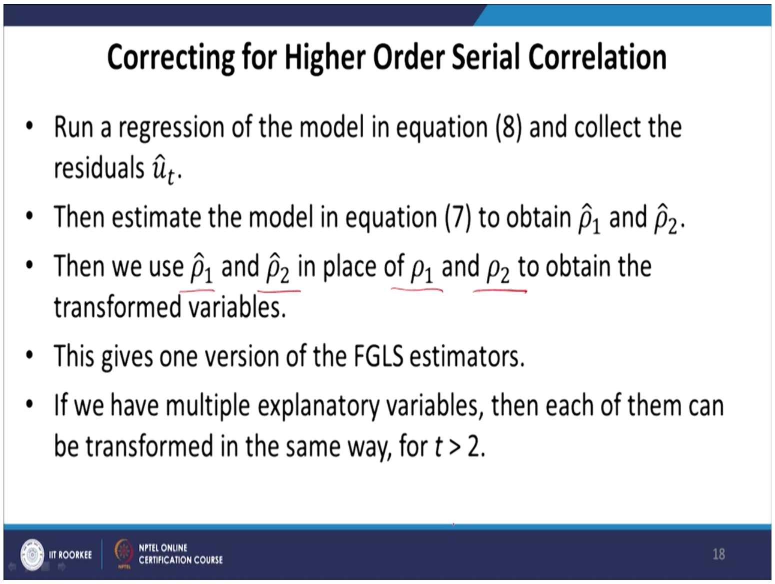


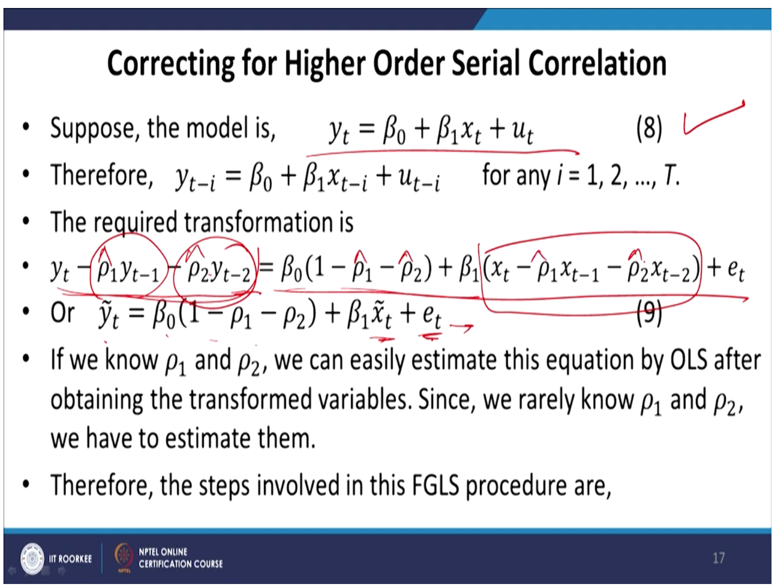
Suppose, this is my original model. Therefore, we have yt minus i equals beta naught plus beta 1 xt minus i plus ut minus i for any i, i equals 1 to capital T. The required transformation is, we simply subtract rho 1 yt minus 1 and rho 2 yt minus 1. Now, remember we are considering AR(2) processes and that is why up to 2 lag there are subtracted.

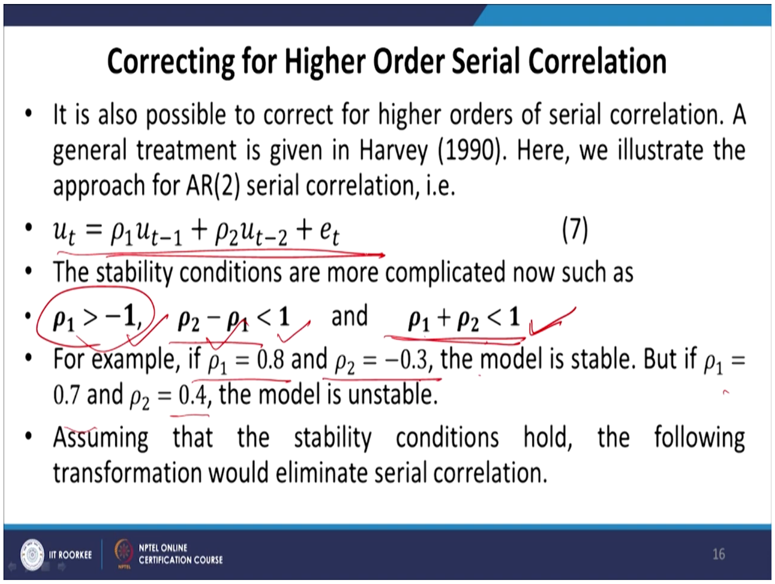
And also from the right-hand side similar subtractions are being made. And then rearranging terms we have yt tilde, so we are calling this expression yt tilde, beta naught 1 minus rho 1 minus rho 2, again we are calling this expression xt tilde, so we have beta 1 xt tilde plus et. The moment we have et on the, as the error actually is, errors are spherical and as a result of which we ensure minimum variance estimators.

If we know rho 1 and rho 2, we can easily estimate the equation by OLS after obtaining the transformed variables. Since we rarely know rho 1 and rho 2, we have to estimate them. So, basically, we follow a very similar procedure. So, the steps involved in the FGLS procedure are;

(Refer Slide Time: 25:22)



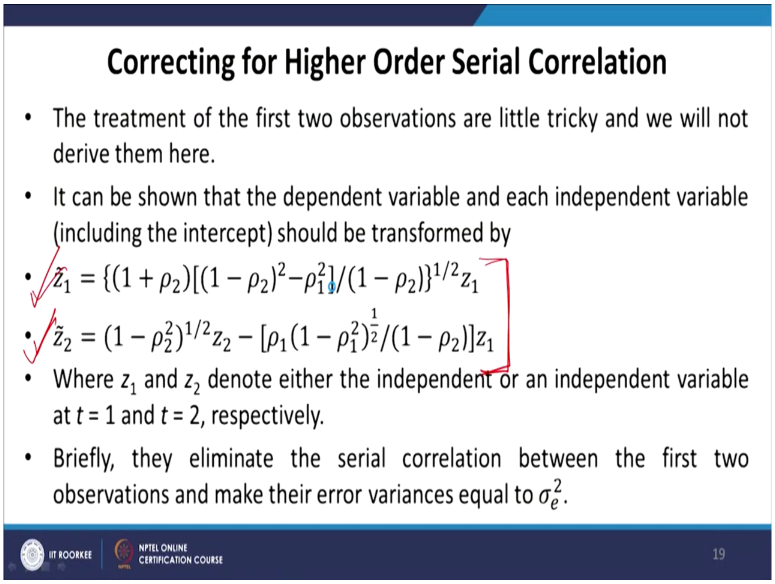




First of all, run a regression of the model in equation 8, the original model, and collect the residuals ut hat. Then estimate the model in equation 7 to obtain rho 1 hat and rho 2 hat. This was my equation 7, an AR(2) model, from where we obtain rho 1 hat and rho 2 hat. Then we use rho 1 hat and rho 2 hat in place of rho 1 and rho 2 obtained the transformed variables. So, for this transformation instead of rho 1 and rho 2, we have rho 1 hat and rho 2 hat.

This gives a version of the SGA f GLS estimators. We, if we have multiple explanatory variables, then each of them can be transformed in the same way for t greater than 2.

(Refer Slide Time: 26:12)

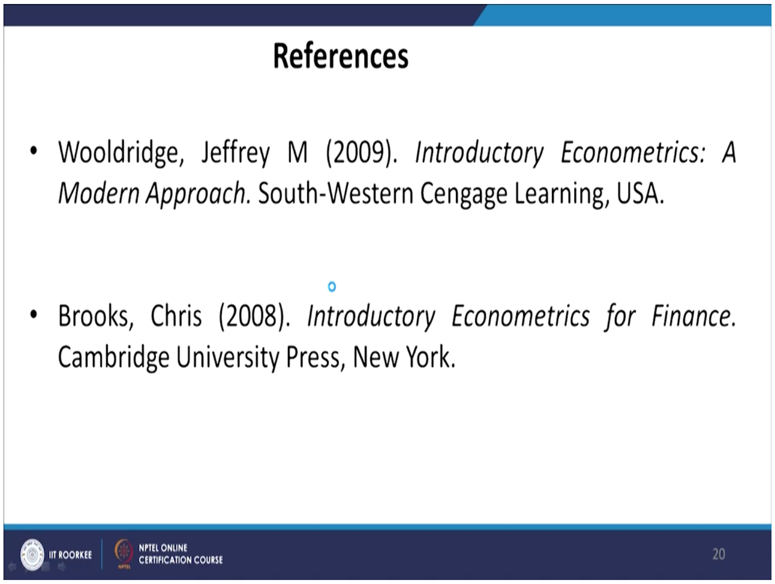


The treatment of the first two observations is a little tricky in this context and we will not derive them here. It can be shown that the dependent variable and each independent variable including the intercept should be transformed by this expression *(refer to slide time 26:12)*. Following these expressions, where z1 and z2 denote either the dependent or an independent variable at time t equals 1 and time t equals 2.

So, for the two times we have two transformations, this reference to the transformation for the first period, this reference to the transformation for the second period, they could be corresponding to an independent variable or the dependent variable *(refer to slide time 26:12)*. Briefly, they eliminate the serial correlation between the first two observations and make their error variances equal to sigma e square. So, we actually do not lose out on any observation.

For the first two observations, these transformations are considered. For the rest of the observation, we consider these transformations. And then what we obtain is basically the most efficient estimators.

(Refer Slide Time: 27:28)



So, these are the references I have considered. And that is all I wanted to discuss on the problem of serial correlation. Thank you.