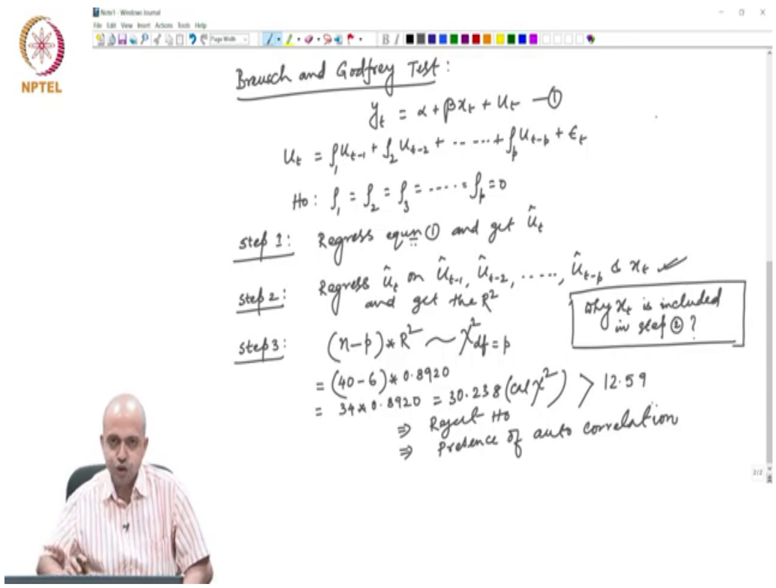
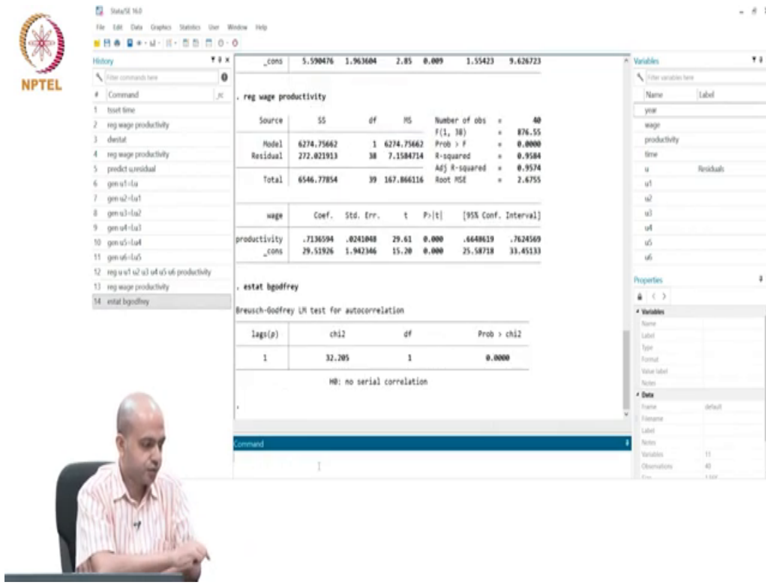
Introduction to Econometrics
Professor Sabuj Kumar Mandal
Department of Humanities and Social Sciences
Indian Institute of Technology, Madras
Lecture 51
Relaxing the assumptions of CLRM-Autocorrelation and Heteroscedasticity Part - 2

(Refer Slide Time: 0:16)



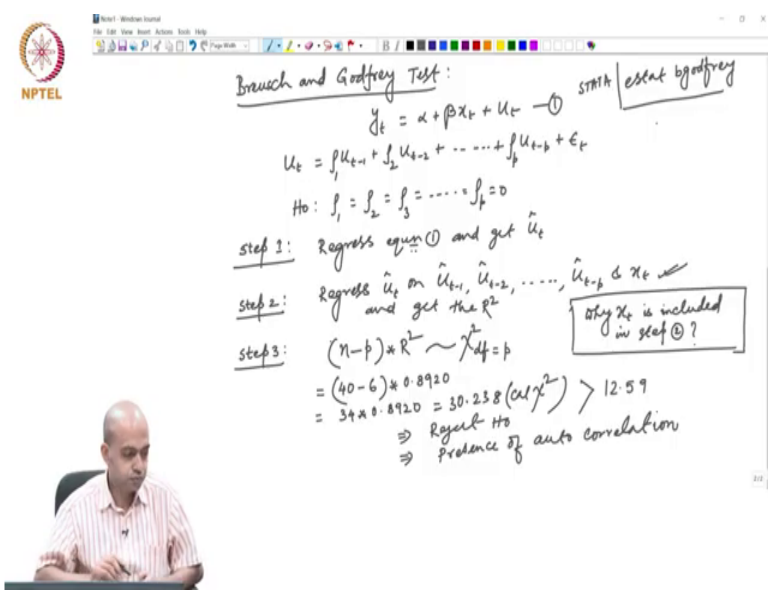
So, this is the procedure for the Breusch and Godfrey test, you have to estimate the model, you have to get the predicted value of the error term, then you have to also create the predicted value of the previous period lag error terms and then you need to regress the ut on its previous values and also on the explanatory variable, the explanatory variable. And then if you do so this is the result. This is the result.

(Refer Slide Time: 1:00)



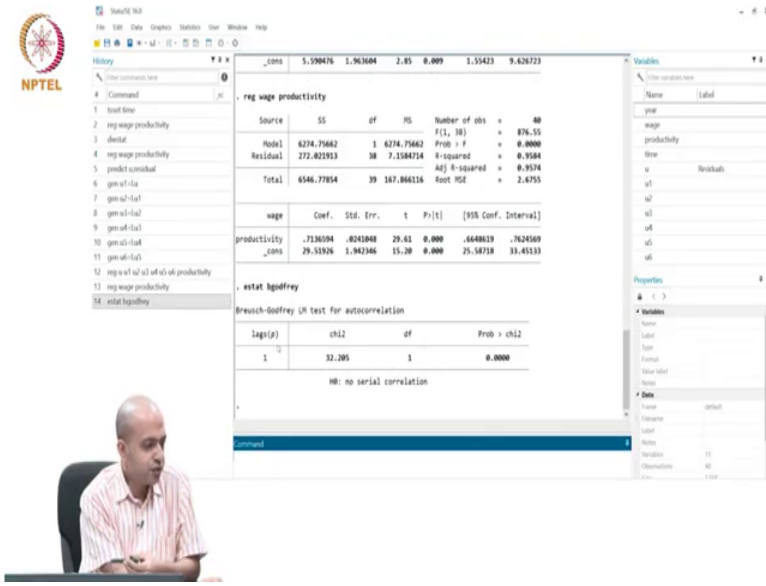
Now, in stata, you do not need to implement the test always manually. there is a specific command if you estimate the model reg wage and productivity then immediately after estimating the model what you need to do, you need to specifically mention a command for this Breusch and Godfrey test which is estat bgodfrey, so this is the command. I will first write the command here so that you should not forget.

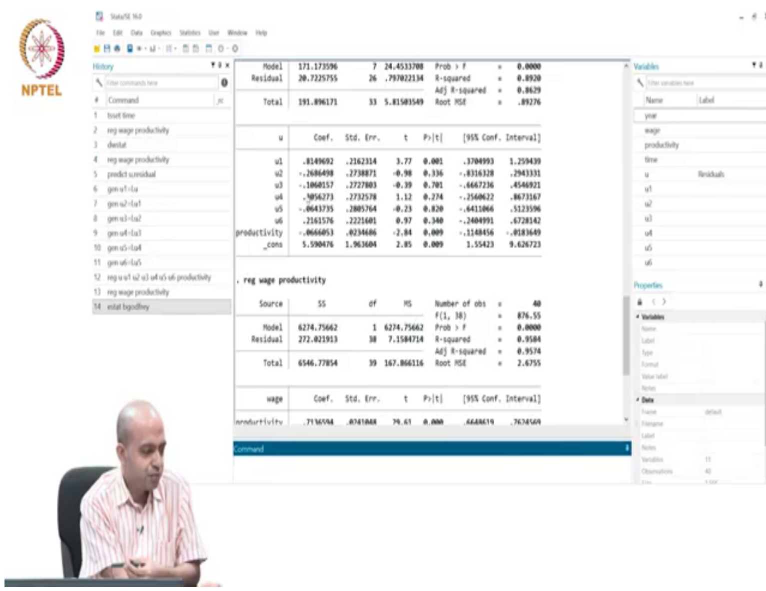
(Refer Slide Time: 1:49)



So, this is the stata command estat bgodfrey.

(Refer Slide Time: 2:17)





Now, see the chi square value is 32.20 and again highly significant even. But there is a difference between the test what we have conducted and slight difference what stata is reporting. See in our manual result what we did we have specified 6 lags but stata is reporting p equals to only 1, what is the number of lag? Only 1.

Now, your question is you are interested in testing higher order lags that is why we are going for Breusch and Godfrey test but stata is reporting only result with 1 period lag it is as good as the

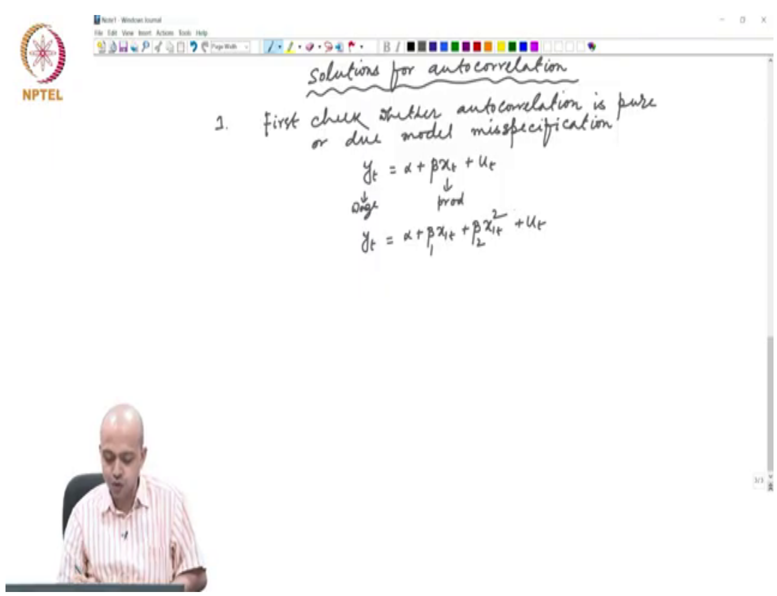
Durbin Watson test statistic, why this is so? Now, the reason is even though Breusch and Godfrey can detect higher order autocorrelation if you look at your result.

See apart from u1 that means apart from the first order lag u2, u3, u4, u5, u6 all these coefficients are actually insignificant, that is the reason stata is reporting the Breusch and Godfrey test with only 1 period lag. Had there been higher order lag, stata would have considered those number of lags. Let us say in the data you have lag up to 3 periods then stata would have reported the result with lag equals to 3.

But here, only u1 is significant that means ut minus 1 hat is significant, ut minus 2 hat, u3 minus, ut minus 3 hat all are insignificant. Since all our insignificant stata has not reported the result with higher period lags, stata has reported only this, this should be your conclusion.

So now, what we have learned? We have learned Durbin Watson test, we have also learned the Breusch and Godfrey test to test for higher order autocorrelation. Now, the next question that comes to our mind, that once you detect autocorrelation, what is the solution?

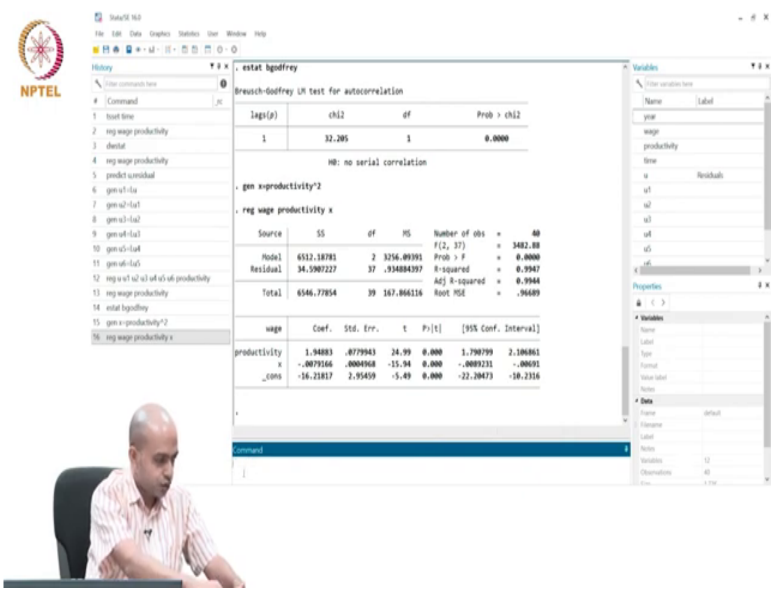
(Refer Slide Time: 5:19)



So, solutions for autocorrelation. Now, the first solution is what you need to check is whether autocorrelation is pure or due to model misspecification. So, that means when we were running this model yt equals to alpha plus beta xt plus ut where yt is wage and xt is productivity, it may so happen that in the true model there is non-linearity between wage and productivity and that is actually leading to this autocorrelation problem.

So, that means we have to check whether yt equals to alpha plus beta1 x1t plus beta2 x2t plus ut, sorry this is not 2t I will say this is beta2 x1t square, so that means I am including productivity as well as square in the model to check whether there is any model misspecification. if the autocorrelation is due to model misspecification, then once you correct that by a square term your autocorrelation problem will get automatically resolved, you do not need to do anything else. That is why econometrician first say that before applying any other medicine you just see whether autocorrelation is pure or due to model misspecification. So, how will you do that? You need to generate the productivity square.

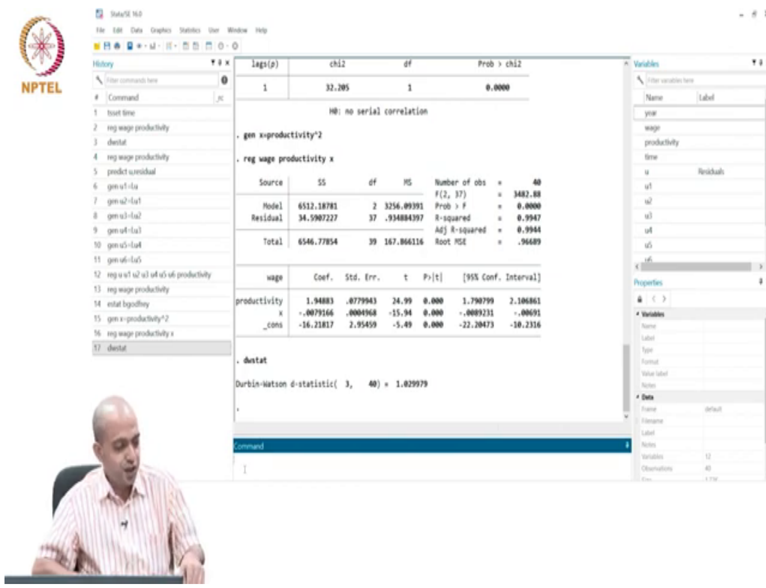
(Refer Slide Time: 8:15)



So, gen, let us say I am giving a different name which is let us say x and how I am defining x? This is nothing but productivity, productivity square, productivity square and then automatically my x variable is included. So now, what you need to do reg wage then productivity and you also include your x.

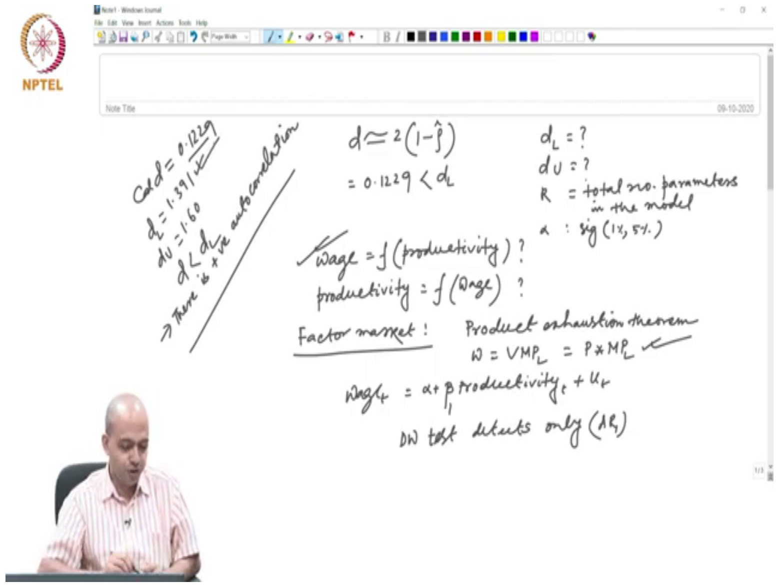
Now, interestingly you see the productivity is positive, coefficient of productivity is positive but productivity square is actually negative and significant that means as productivity increases wage first increases but after some point of time it is coming down, it is not increasing that much, so there might be some reason behind this that means some kind of non-linearity might exist between wage and productivity. you might explore the reason I am not talking into that part here. So, what I will do now after this model once you estimate this modified model let us see what is the value of the Durbin Watson test statistic.

(Refer Slide Time: 9:50)



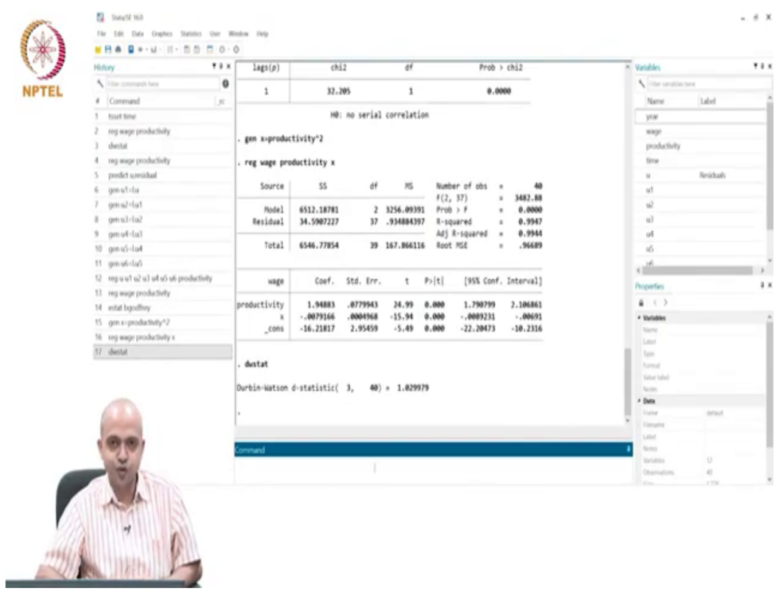
So, dwstat. Now see, the productivity, after improving productivity square the Durbin Watson test statistic has improved a lot from 0.12292 it is now 1.02. So it has improved but still it is lower than the Durbin Watson lower value. What was the lower value we were discussing?

(Refer Slide Time: 10:31)



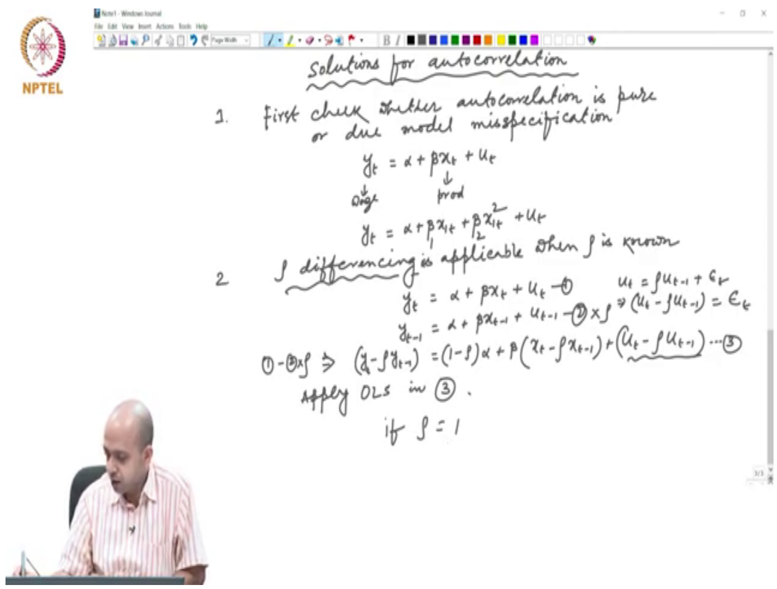
If you remember it is 1.39 and how much value you are getting here?

(Refer Slide Time: 10:40)



It is 1.02, so it is still lower than the lower limit of Durbin Watson test so that means the autocorrelation is not due to model misspecification, it is actually a pure autocorrelation meaning the other reasons that inertia and other things that means the time stage macro-economic variables they exhibit business cycle, so on and so forth and that is why this autocorrelation is actually existing in the data. So, that means by modifying the model we cannot solve autocorrelation problem, we need to think about other method. And now we will discuss about other method of solving autocorrelation problem.

(Refer Slide Time: 11:46)



So, the second method is called rho differencing, where rho is the autocorrelation coefficient. So, your model was yt equals to alpha plus beta xt plus ut and ut follows ut equals to rho ut minus 1 plus epsilon t that is your data generating process.

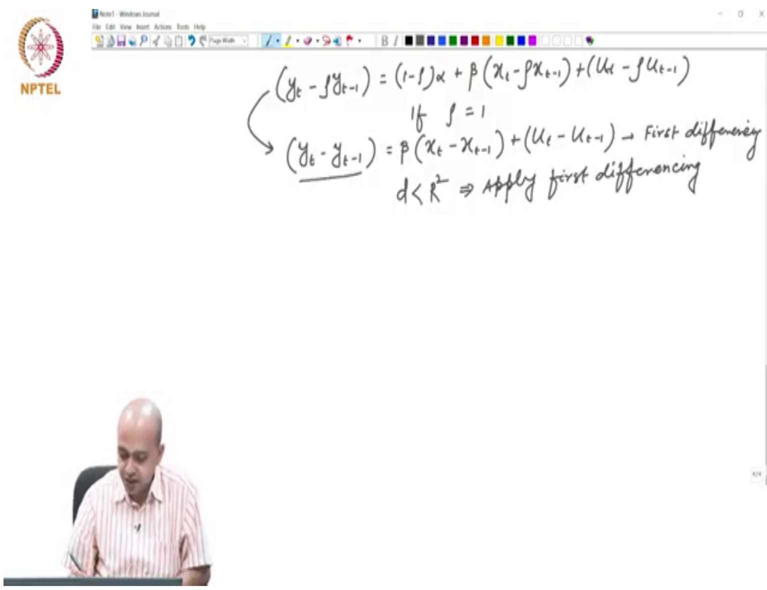
So, you need to first take the difference of this, so that means yt minus 1 should be equal to alpha plus beta xt minus 1 plus ut minus 1. Then what you need to do, yt minus rho yt minus 1 equals to, so this is alpha minus rho alpha, alpha minus rho alpha, so I am multiplying let us say this is equation 1, this is equation 2 and I am multiplying equation 2 by rho and then I am taking 1 minus 2 into rho that implies yt minus rho yt minus 1, this would become alpha minus rho alpha so that means if you take 1 minus rho into alpha, alpha plus beta into xt minus rho xt minus 1 plus ut minus rho ut minus 1. Let us say this is 3.

And then apply OLS in 3 that is the procedure. So, that means what I am saying while OLS was not applicable in equation 1, how come OLS is still applicable in this modified equation after rho differencing? Because if you look at the data generating process for the ut this says that ut minus rho ut minus 1 is actually epsilon t and epsilon t follows all the assumption of classical linear regression model.

That is the reason this error term is actually the classical error term, it does not show any autocorrelation problem. So, this is the rho differencing method so that means you need to transform your dependent as well as independent variable. How?

If your dependent variable is y then the modified dependent variable would be yt minus rho yt minus 1 and your independent variable would be transformed as xt minus rho xt minus 1. But the point here is this rho differencing method is applicable only when a rho is known to you but if you do not know the rho what will happen? If you know then you put the specific value of rho, if you do not know then what we need to do, we need to assume a specific value for rho. Now, if we assume, rho equals to 1 then what will happen?

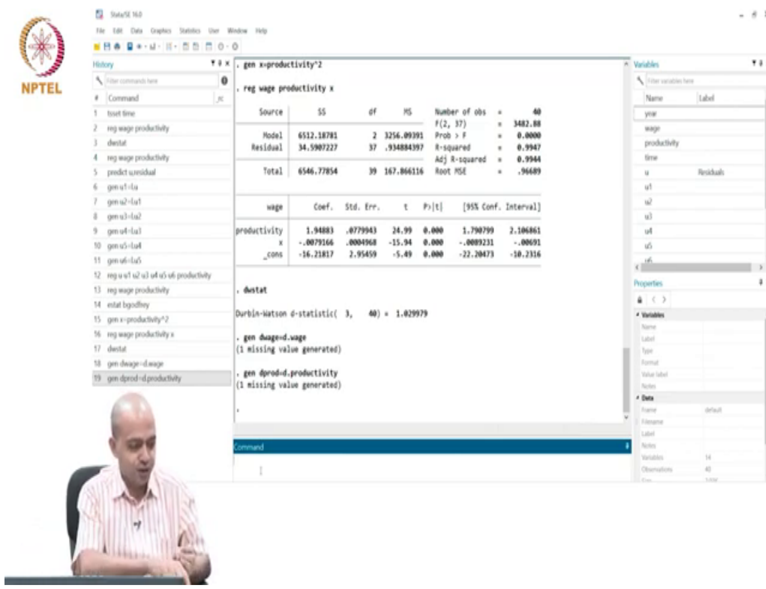
(Refer Slide Time: 16:50)



So, that means your rho transforming equation was yt minus 1 equals to 1 minus rho into alpha plus beta xt minus rho xt minus 1 plus ut minus ut minus 1, sorry rho ut minus 1, ut minus rho ut minus 1. So, if you put rho equals to 1 that means if we assume there is perfect first order, perfect positive autocorrelation then this equation will become yt minus yt minus 1 equals to beta xt minus xt minus 1 plus ut minus ut minus 1, just by putting rho equals to 1 and this equation is called first differencing.

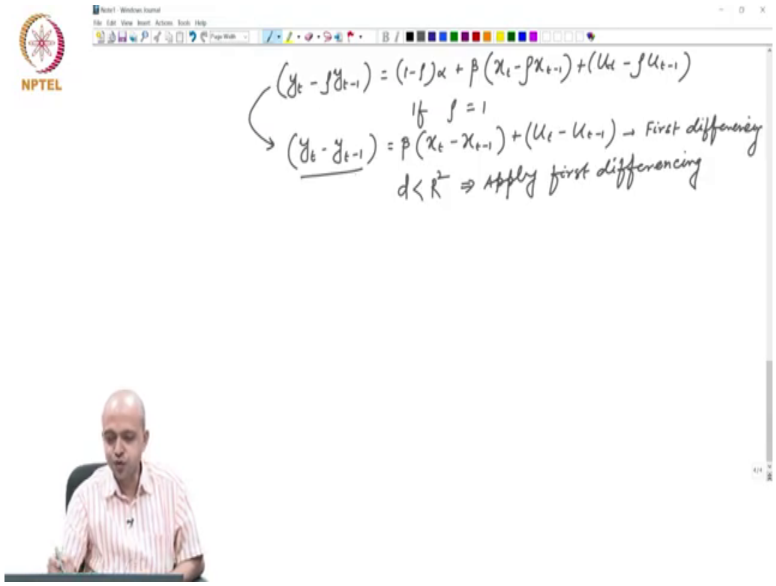
So, first differencing is a solution to solve autocorrelation problem when rho equals to 1. So, generally econometrician say that when d is less than R square apply first differencing. Now, what we need to do, we need to transform the variable as yt minus yt minus 1 and xt as xt minus 1.

(Refer Slide Time: 19:19)



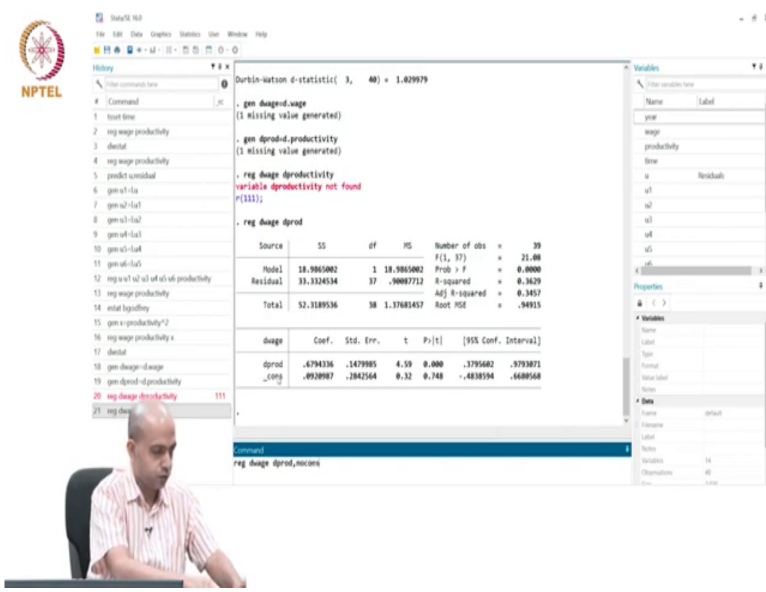
So, what we will do, we will transform the variable, so we will say that gen d wage, this is d wage I am writing equals to d dot wage. So, what is the command I am using? Earlier I was using l, l for lag and d for difference. So, the moment I say gen d wage, d wage is basically a name and d dot wage means wage t minus 1 that means basically yt minus yt minus 1. I have created the first difference. Similarly, gen d prod equals to d dot productivity. So, I have created, I have created the two, first difference of wage as well as productivity. And now if you regress d wage on d productivity then we will see.

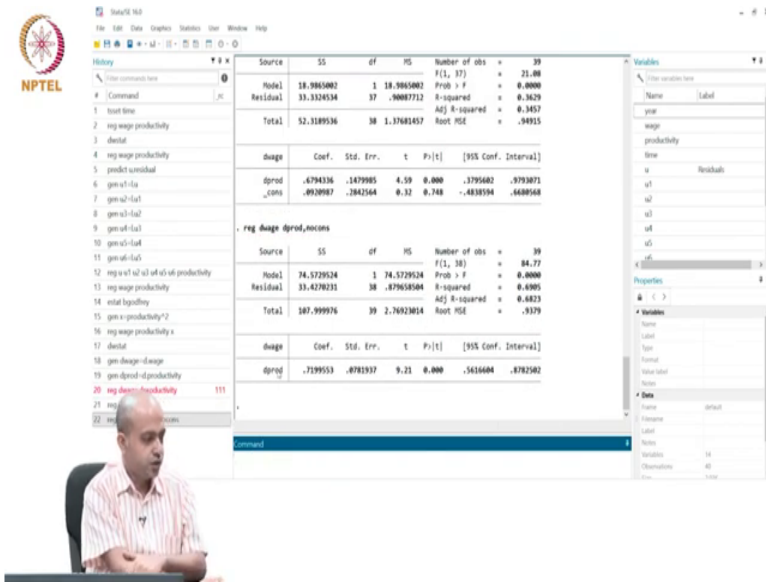
(Refer Slide Time: 20:55)

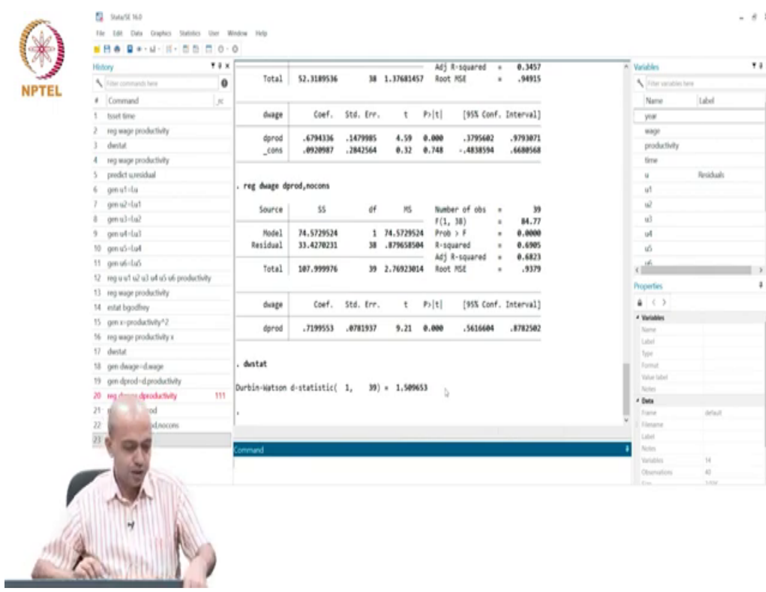


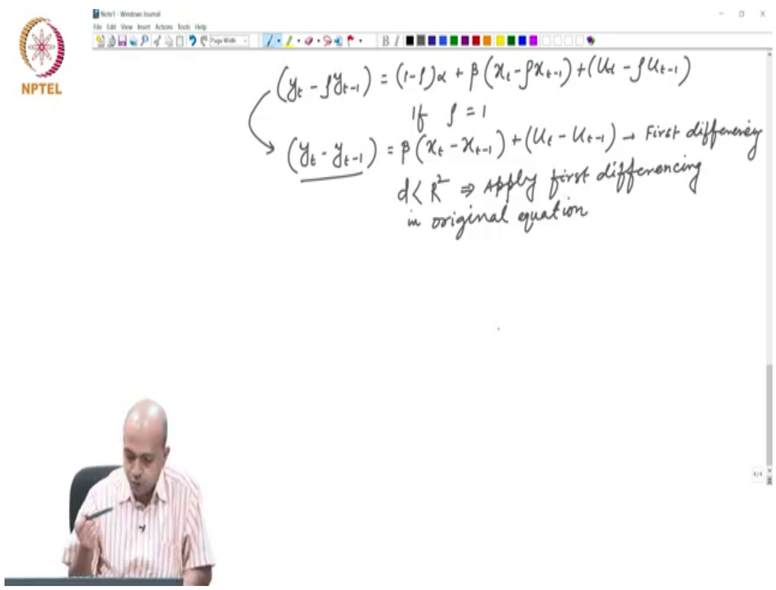
But before we do that you, we have to again look back to our model see in the first difference equation there is no constant term. The constant term gets cancelled when you put rho equals to 1, so in any first difference equation you have to keep in mind whenever you run first difference equation you should not include constant term in the model.

(Refer Slide Time: 21:15)




NPTEL
Source SS df MS Number of obs = 39
F(1, 37) = 21.08
Model 18.9865002 1 18.9865002 Prob > F = 0.0000
Residual 33.3324534 37 .90087712 R-squared = 0.3629
Adj R-squared = 0.3457
Total 52.3189536 38 1.37681457 Root MSE = .94915
dwage Coef. Std. Err. t P>|t| [95% Conf. Interval]
dprod .6794336 .1479985 4.59 0.000 .3795602 .9793071
_cons .0920987 .2842564 0.32 0.748 -.4838594 .6680568
. reg dwage dprod,nocons
Source SS df MS Number of obs = 39
F(1, 38) = 84.77
Model 74.5729524 1 74.5729524 Prob > F = 0.0000
Residual 33.4270231 38 .879658504 R-squared = 0.6905
Adj R-squared = 0.6823
Total 107.999976 39 2.76923014 Root MSE = .9379
dwage Coef. Std. Err. t P>|t| [95% Conf. Interval]
dprod .7199553 .0781937 9.21 0.000 .5616604 .8782502
Command

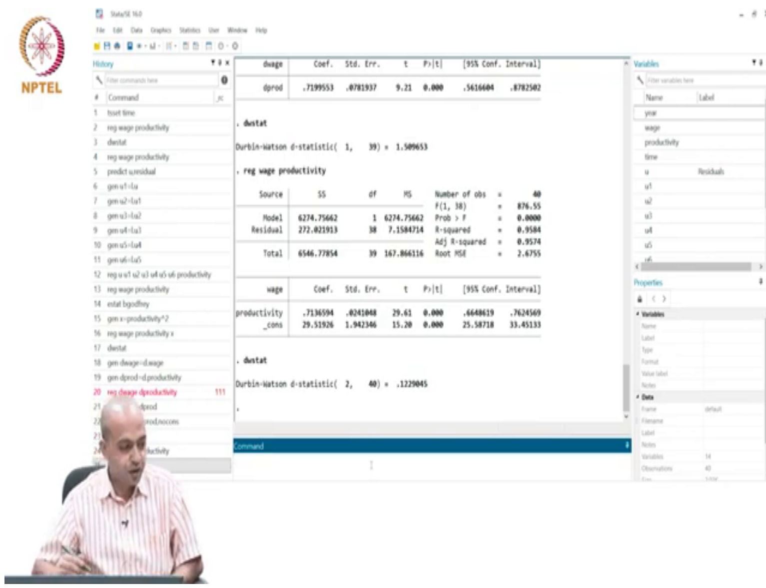

NPTEL
Total 52.3189536 38 1.37681457 Root MSE = .94915
dwage Coef. Std. Err. t P>|t| [95% Conf. Interval]
dprod .6794336 .1479985 4.59 0.000 .3795602 .9793071
_cons .0920987 .2842564 0.32 0.748 -.4838594 .6680568
. reg dwage dprod,nocons
Source SS df MS Number of obs = 39
F(1, 38) = 84.77
Model 74.5729524 1 74.5729524 Prob > F = 0.0000
Residual 33.4270231 38 .879658504 R-squared = 0.6905
Adj R-squared = 0.6823
Total 107.999976 39 2.76923014 Root MSE = .9379
dwage Coef. Std. Err. t P>|t| [95% Conf. Interval]
dprod .7199553 .0781937 9.21 0.000 .5616604 .8782502
. dwstat
Durbin-Watson d-statistic(1, 39) = 1.509653
Command



So, now you reg d wage d prod then no cons. this is the command. Now, here again after running this first difference equation what is the Durbin Watson value?

Durbin Watson value has improved a lot 1.5096 in the first difference, so the Durbin Watson value is 1.50 which is actually now if you go back which is if you go back then you will see that your value was Durbin Watson lower value was something around 1.34 or something, so that means it has now improved a lot.

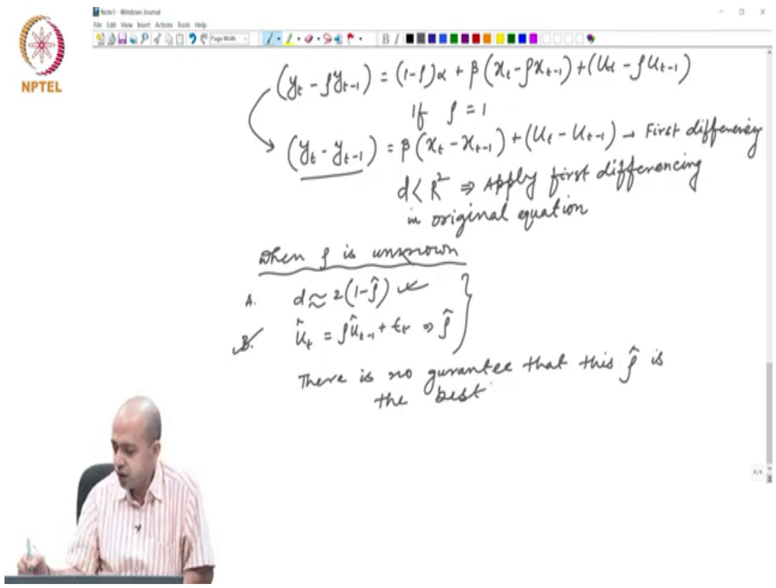
(Refer Slide Time: 24:18)



If you look at the original equation that means this equation you have to see reg wage and productivity where we have run that equation, if you look at the reg wage productivity, reg wage and productivity and then dw stat yeah, so this is lower than the R square and first difference is applicable.

But the only problem with the first difference is that we are specifying a specific value for rho hat which is 1 and also the moment you take first difference, the interpretation of the var, of the coefficient we were interested to see the impact of xt on yt but ultimately what we are getting the impact of delta yt on delta xt that is the problem. But anyway, the severity of autocorrelation is largely reduced.

(Refer Slide Time: 25:26)

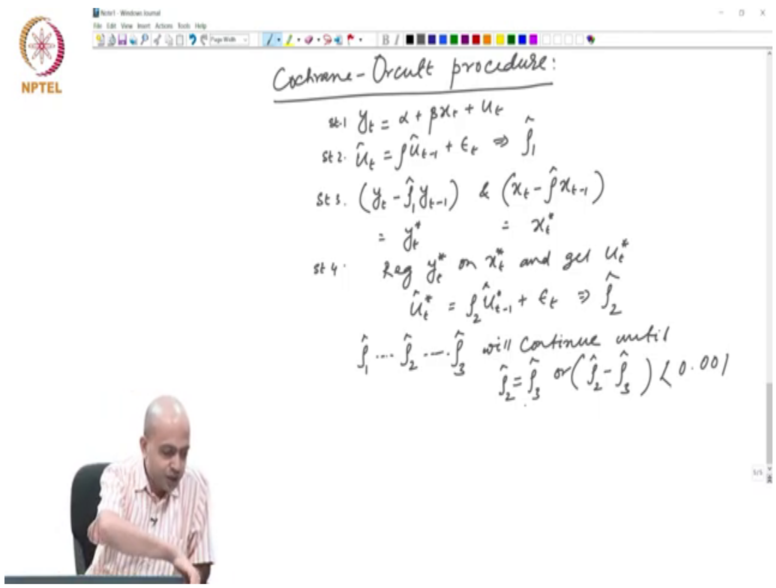


The next question what we need to do is if we do not know the value of when this rho is unknown what actually you can do instead of putting rho equals to 1 there are two ways by which you can get rho. See if you recall we have a relationship like this is 2 into 1 minus rho hat, so the moment you know Durbin Watson test statistic using this relationship you can get rho hat, so you estimate the model, get your Durbin Watson test statistic and then put the Durbin Watson value here and get your rho hat, that is method number 1.

Method number 2 is you estimate the model, get your ut hat and then run this regression reg this equals to rho ut minus 1 plus epsilon t, so that means if you regress ut hat on ut minus 1 hat from there also you will get rho hat that is also possible, so when rho is unknown these are the two ways by which you can get rho hat but please keep in mind the relationship d is almost equals to 2 into 1 minus rho hat is applicable when you have sufficiently large number of time periods.

Here the number of time periods is 40 which is large but if you have 10, 15 years data then this is not actually applicable this method you have to use the method number B this one, you can simply regress ut on ut minus 1 and get your rho hat. Now, econometrician say sometimes that in this method how do you know that the rho hat is the best method? Because you estimate the model regress ut hat on ut minus 1 hat, collect that rho hat and using that rho hat to transform the variable, so there is no guarantee that this rho hat is the best.

(Refer Slide Time: 28:29)



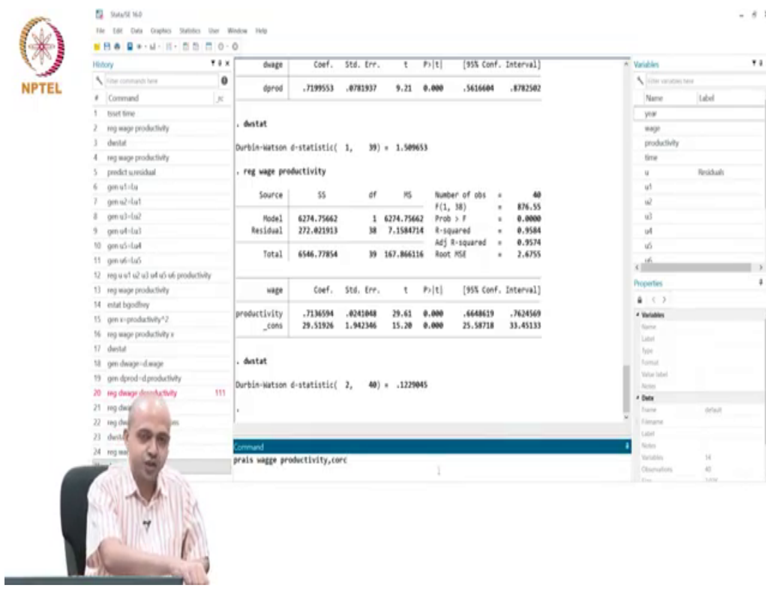
That is why they sometimes suggest to iterate this procedure for one or two couple of more rounds so and that procedure is called Cochrane Orcutt procedure. What they say? You run this regression yt equals to alpha plus beta xt plus ut and then you regress ut on rho ut minus 1 plus epsilon t and then from there you collect your rho 1 ut minus 1 rho 1 hat.

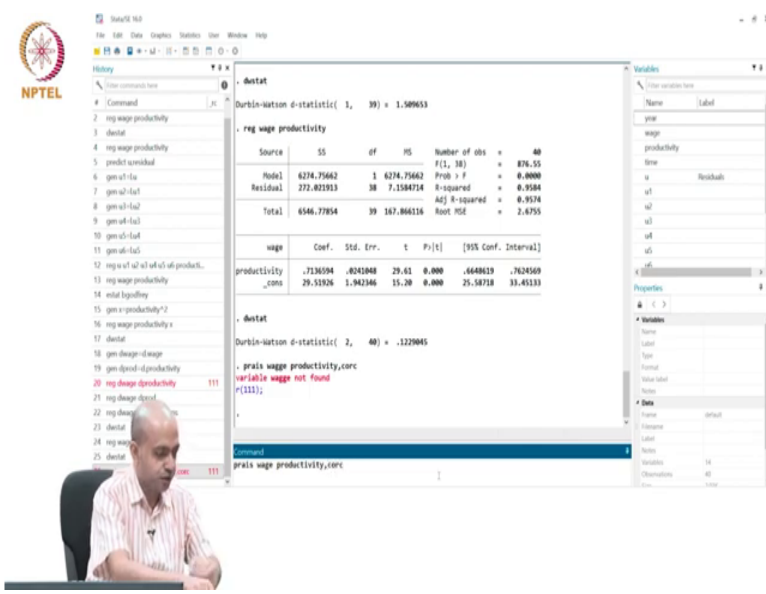
Then what you do this is step 1, this is step 2, then in step 3 what you do you transform your dependent and independent variable with this rho 1 hat xt minus rho xt minus 1 so I have changed my this and let us say this is called yt star and let us say this is called xt star. Then what I need to do? I need to regress this yt in step 4, reg yt star on xt star and get ut star.

Then what you need to do? Again, run ut star hat equals to rho 2 ut minus 1 plus epsilon t and get rho 2 hat. Again, you take this rho 2 hat to transform your dependent and independent variable. So, that means you will get rho 1 hat, rho 2 hat, rho 3 hat in every successive iteration and I will stop when the two successive periods rho hat is actually equal and this will continue until rho 2 hat minus rho 3 hat equals to rho 3 hat or rho 2 hat and the rho 3 hat this difference is less than 0.001. And I will take that rho hat to transform the variable, I will take that rho hat to transform the variable instead of using this rho 1 hat which is coming in the first stage. Is this clear?

So, this procedure Cochrane Orcutt Procedure is nothing but repeating the procedure for couple of more rounds, more iteration and then take that particular rho when two successive periods rho hat actually converges to each other or the difference between the two rho hat is less than 0.001. And in the process how will you do that?

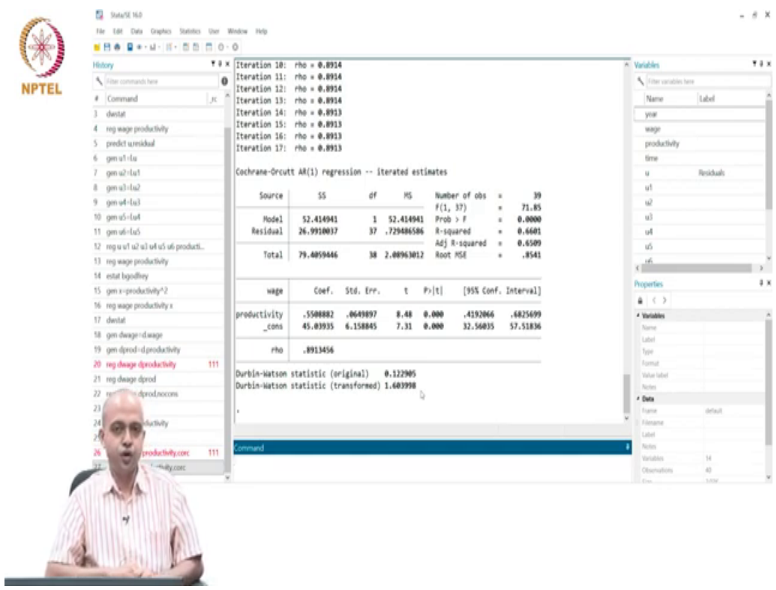
(Refer Slide Time: 33:01)





In the data if you want to apply this Cochrane Orcutt Procedure to correct your yt and xt, transform your yt and xt and then run the regression, problem of autocorrelation will come down a lot, I will show you the command, this is again a specific command, prais wage productivity in corc, this is the command.

(Refer Slide Time: 34:09)



Now, look at what stata is doing? Stata is running 17 rounds of iteration and then lastly stata is taking 0.8913 as the best rho. And that rho is used to transform your yt that means yt is now yt minus rho yt minus 1 xt is now xt minus rho xt minus 1 and then the regression is run using this rho this is the regression output and you see the Durbin Watson original value which was 0.122 now it is 1.60.

So, you have solved actually the autocorrelation problem and you have got the best result, this is how you can solve your autocorrelation problem by a generalized method of rho differencing suggested by Cochrane and Orcutt Procedure. So, with this we are closing our discussion on autocorrelation, we have discussed about what is autocorrelation, what are the consequences, consequence is basically the standard error gets disturbed, efficiency property get disturbed and as a result of which you have problem in your hypothesis testing.

And then we have discussed about how to detect autocorrelation using Durbin Watson as well as Breusch and Godfrey test and lastly, we discussed about solution, rho differencing and first differencing. And then we lastly discussed about Cochrane Orcutt Procedure which is an iterative procedure to get the best rho to transform your dependent and independent variables. With this the discussion of autocorrelation is over. Thank you.