

Engineering Econometrics
Prof. Rudra P. Pradhan
Vinod Gupta School of Management
Indian Institute of Technology, Kharagpur

Lecture – 27
Autocorrelation Problem

Hello everybody, this is Rudra Pradhan here, welcome to Engineering Econometrics. Today we will continue with model diagnostics that to in the last lecture we have discussed something about the multicollinearity problem and then we have discussed some of the problems to address the multicollinearity issue and the kind of you know solution thereof. And in this lecture typically will discuss similar kind of you know problem and that to the concept of autocorrelation. First of all what is exactly this autocorrelation problem.

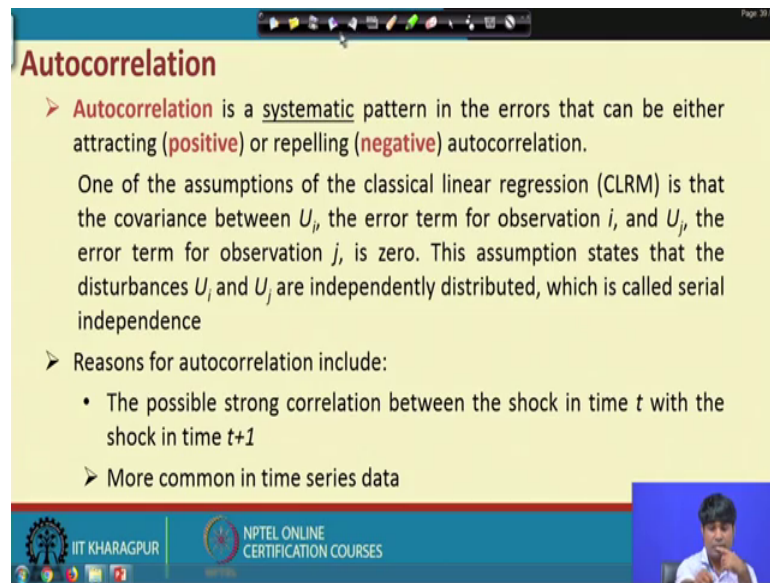
And a we will address certain issues behind this autocorrelation problem that to what is exactly the autocorrelation issue, how to detect this problem, what are the reasons behind this particular you know involvement and a what will we you know happen if that particular problem arise in the modelling. And then we really suggest certain you know guidelines or the kind of you know approach to the solution of autocorrelation problem.

By the way autocorrelation is a kind of you know problematic situation in the case of you know modelling, any kind of you know econometrics modelling. When we discuss certain problems and our usual approach is to fit the model and then we have we must have with the estimated model and that will be used for any kind of you know prediction so, far as a engineering problem is concerned.

And a before you start the prediction of a particular engineering problem, we like to you know justify that this the estimated models which we have from the fitted data should be free from all kind of you know virus, that is the what we use the term called as a model diagnostics.

And a so, here the issues is easy with respect to the error terms and a that to whether there is a any kind of you know relationship exist among the errors, which is usually derived through kind of you know estimation process.

(Refer Slide Time: 02:49)

The image shows a presentation slide titled "Autocorrelation" from an NPTEL course. The slide has a yellow background with a blue header and footer. The header contains the title "Autocorrelation" in red. The main content area is yellow and contains text and bullet points. The footer is blue and contains the IIT Kharagpur logo and the text "NPTEL ONLINE CERTIFICATION COURSES". A small video inset in the bottom right corner shows a man speaking.

Autocorrelation

➤ **Autocorrelation** is a systematic pattern in the errors that can be either attracting (**positive**) or repelling (**negative**) autocorrelation.

One of the assumptions of the classical linear regression (CLRM) is that the covariance between U_i , the error term for observation i , and U_j , the error term for observation j , is zero. This assumption states that the disturbances U_i and U_j are independently distributed, which is called serial independence

➤ Reasons for autocorrelation include:

- The possible strong correlation between the shock in time t with the shock in time $t+1$

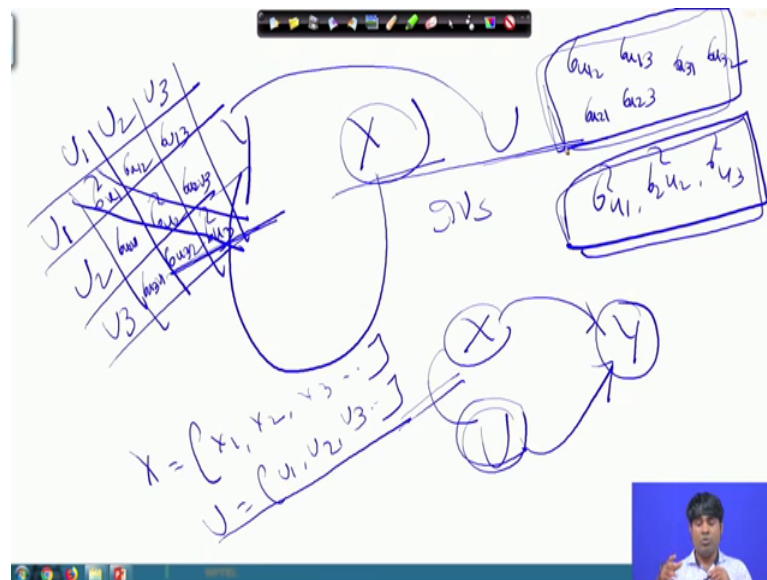
➤ More common in time series data

IIT KHARAGPUR | NPTEL ONLINE CERTIFICATION COURSES

The thing is that you know we really go through the particular you know process. What is all about the autocorrelation and a how is this particular you know concept and then we will think about the consequence, the kind of you know solutions and so, on. Usually autocorrelation is a systematic pattern in the errors that can be either positive or negative. So that means, the nature of the relationship is not so important. What is important? Whether there is a typical relationship exist among the error terms.

So; that means, when we a regress any kind of you know any kind of you know error terms so, we will find ok. So, usually a, for a you know simple modelling.

(Refer Slide Time: 03:53)



Let us say we have a Y and X and then U. So, the structure of modelling econometric modelling for any kind of you know engineering problem deals with the involvement of dependent variable, the involvement of independent variables and the involvement of you know error term. Usually X and U are called as a in independent variables that to one is explained in nature that is the X cluster and U is the unexplained independent variable that is the error cluster. And a both has you know a you know independent impact on the dependent variables.

So; that means, technically we if we have a dependent variable here let us say Y and then we have X variable here and then error term here. So, technically X must have a impact on U Y and a Y must U must have an impact on U know U and then there is a kind of you know relationship between X and U. So, what is the model requirement? There should not be any relationship between X and U.

Again within the X we have a you know multiple number of variables like X_1, X_2, X_3 and so, on. And then we under U we have also multiple number of variables say U_1, U_2, U_3 and so, on. So that means, we like to equally check whether there is any relationship among the regressors that to with respect X and that to with respect to U.

So, when we talk about the relationship among the X that 2 independent explained variable and that particular approach is called as a multicollinearity problem which we

have discussed in details in the last couple of lectures. And in these particular lectures we will be dealing with the error variables, which is nothing, but you know you clusters.

Now, having the U variables in the U vectors so, we can prepare a variance covariance matrix with respect to error terms. So, technically we can have error representation like this Y U_1 , U_2 , U_3 and so, on. But, in the mean times let us have a error covariance matrix with respect to U_1 , U_2 and U_3 .

So, ultimately we can have a variance covariance matrix a you know 3 into 3 variance covariance matrix and then. So, this is called as you know $\sigma^2_{U_1}$, this is called as a $\sigma^2_{U_2}$ and this is called as a $\sigma^2_{U_3}$ and then this is $\sigma_{U_1 U_2}$ and this is $\sigma_{U_1 U_3}$ and this is $\sigma_{U_2 U_1}$ and this is $\sigma_{U_2 U_3}$ and this is $\sigma_{U_3 U_1}$ and this is $\sigma_{U_3 U_2}$.

So; that means, we have a for 3 independent error variables so, we have variance covariance matrix of order 3. So, within this order 3 we have 3 diagonal elements that $\sigma^2_{U_1}$ $\sigma^2_{U_2}$ $\sigma^2_{U_3}$ so; that means, a $\sigma^2_{U_1}$ $\sigma^2_{U_2}$ and $\sigma^2_{U_3}$. So, this is one cluster and another cluster will be a will be a the error you know covariance that to $\sigma_{U_1 U_2}$ to $\sigma_{U_1 U_3}$, $\sigma_{U_2 U_1}$, $\sigma_{U_2 U_3}$ and $\sigma_{U_3 U_1}$ and $\sigma_{U_3 U_2}$.

So, these are all you know a 6 different you know covariance. Now, our discussion is with respect to autocorrelation problem. So, now, in the with respect to the variance covariance matrix of error terms we have 2 sets of problem that 2 with respect to variance of the error terms and then with respect to covariance of the error terms.

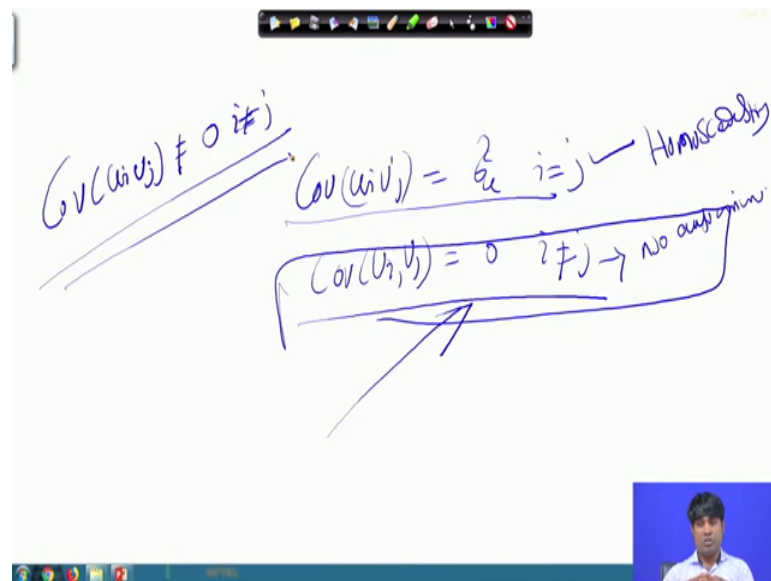
So, when we will be dealing with the covariance of the error terms that is the problem is called as a autocorrelation, when we will be dealing with the variance of the error term that is the problem called as you know heteroscedasticity problem. So, either heteroscedasticity or homoscedasticity so, what is heteroscedasticity? Or what is a homoscedasticity? We will be discuss in the next lecture. But in the mean times, we will be dealing with now autocorrelation problem and that to with respect to Variance covariance of the error terms.

So, in this case for you know having U_1 , U_2 , U_3 . So, we have couple of you know covariance like you know $\sigma_{U_1 U_2}$, $\sigma_{U_1 U_3}$, $\sigma_{U_2 U_1}$, $\sigma_{U_2 U_3}$ and then

$\sigma^2_{u_i}$ and $\sigma^2_{u_j}$. And with this you know covariance the requirement is that you know this covariance should be equal to 0. Because, our declaration is that the independent unexplained variable is you know independent one.

So, since it is a since the declaration is the independent variable so, there should not be any relationship you know between a between the two. So, if there is any such relationship, then that particular relationship is called as you know autocorrelation.

(Refer Slide Time: 09:57)



Handwritten notes on a whiteboard:

- $Cov(u_i, u_j) \neq 0 \text{ if } i \neq j$ (crossed out)
- $Cov(u_i, u_j) = \sigma^2_u \text{ if } i = j$ — Homoscedasticity
- $Cov(u_i, u_j) = 0 \text{ if } i \neq j \rightarrow \text{no autocorrelation}$

So, technically so, the issue is like you know so, with respect to error term. So, we can have $\sigma^2_{u_i}$; that means, technically it is better to write like you know covariance of u_i and u_j which is equal to σ^2_u for i equal to j and covariance of covariance of u_i and u_j , which is equal to actually a you know 0 for i not equal to j ok.

So, when this will appear then the problem is that you know the particular model is 3 from autocorrelation, but when this will happen so; that means, covariance of u_i equal to u_i and u_j is equal to σ^2_u where i equal to j if that will happen then this problem is called as a heteroscedasticity.

So, now so, this is the situation of you know homoscedasticity homoscedasticity homoscedasticity and this is the situation of you know no autocorrelation no autocorrelation ok. So, now, if this will not equal to σ^2_u so; that means,

covariance of $U_i U_j$ equal not equal to $\sigma^2 U$ when i equal to j then this problem is called as you know heteroscedasticity problem. And when covariance of $U_i U_j$ not equal to 0 for i not equal to j then it is called as you know autocorrelation problems.

So; that means, technically our discussion is on autocorrelation problem. So, now, so, the simple understanding of you know correlation is like this. So, covariance of $U_i U_j$ a equal to not equal to 0 for i not equal to j . So, this is the simple you know mathematical notation of you know autocorrelation. So that means, in the a simple language so, there should not be any relationship among the error terms. If there is any such a relationship or there is a correlation or covariance among these error terms then this particular problem is called as you know autocorrelation problem.

So, let us see how is this particular you know problem altogether and a we will discuss in details right. So, this is what the problem and a with simple understanding about the autocorrelation so, we like to see whether the particular model is it free from autocorrelation or is having autocorrelation. But in reality, most of the problems you know wherever you know kind of you know the estimation and the kind of analysis and when we are dealing with you know data to fit the model.

So, autocorrelation is a kind of in a genuine problem, but you know what is the kind of you know tolerance level, which you would like to keep so, that the model can be used for you know decision making process or the kind of you know forecasting requirement of you know any kind of an engineering problem. So, our job is to check what is the level of autocorrelation and whether that particular level is a absolutely for any kind of you know decision making process.

So, now, in between we like to discuss what are the you know criteria through which you can detect this you know autocorrelation problem and what are the you know regions through which it will be appear in the systems and what is the typical consequence and what is the kind of you know solutions.

So, these are the things we are supposed to target here's. So, in the simple way you know if you ask what is exactly autocorrelation so, we have already discussed. So, it is the relationship among the repressors. So, the requirement is that you know there should not

be any covariance or correlation among the error terms. If there is such a relationship then the simple understanding is that there is an autocorrelation problem.

(Refer Slide Time: 14:09)

Regression Model

$$Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + U_t$$

No autocorrelation: $\text{Cov}(U_i, U_j) = 0$
or $E(U_i, U_j) = 0$

Autocorrelation: $\text{Cov}(U_i, U_j) \neq 0$
or $E(U_i, U_j) \neq 0$

Note: $i \neq j$

In general $E(U_t, U_{t-s}) \neq 0$

U_{t-1} U_{t-2}

IIT KHARAGPUR | NPTEL ONLINE CERTIFICATION COURSES

So, now whatever we have discussed this is what the kind of you know cross checks. So, let us we have a model like this and this models give the indication that you know Y is equal to β_0 sorry β_1 plus $\beta_2 X_2$ plus $\beta_3 X_3$. So, that is called as you know 3 variable case. So, let us start with a simple models where y is a function of you know X_2 and X_3 , then the particular model will be a you know is like this Y equal to β_1 plus $\beta_2 X_2$ $\beta_3 X_3$ and the error term.

So, this is the intercept and this logs and this were the error terms. After the estimation process we can have the a magnitude of you know error term and then we like to check the kind of you know relationship you know between the these error terms.

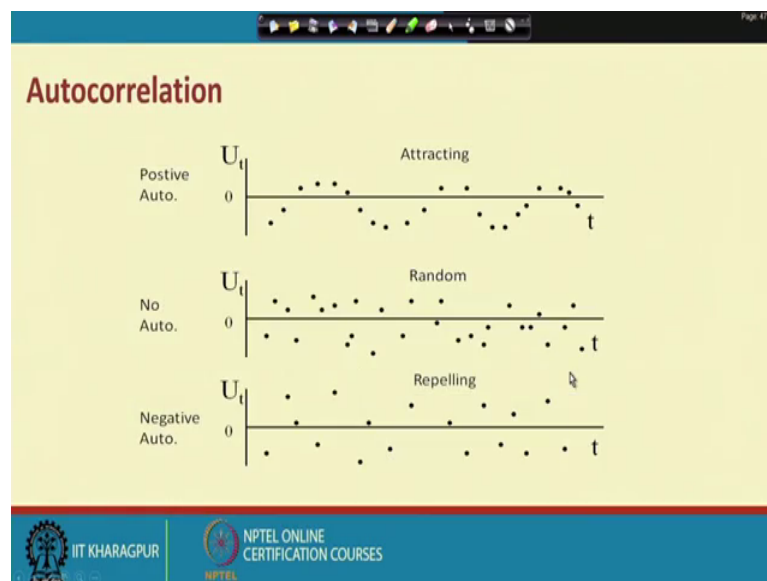
So, now, so, simple understanding what we have already discussed covariance of $U_i U_j$ equal to 0 that is the case of you know no autocorrelation. And when covariance of $U_i U_j$ not equal to 0 for i not equal to j that is the condition then there is a kind of you know autocorrelation. So, in general so, the structure of you know autocorrelation can be written like this because, for a particular you know error component we can create n number of error terms by using simply a you know time lags.

So, that is what the t minus s (Refer Time: 15:52) all about. So, for instance you can create U_t , U_{t-1} , U_{t-2} and so, on. So, every time we have to check whether there is a relationship between U_t , U_{t-1} , U_t , U_{t-2} and again U_{t-1} , U_{t-2} . So, there are there are you know n number of you know possible ways you can you know a check the relationship and then come with the conclusion whether there is a kind of an auto correlations or there is no autocorrelation in the systems.

So, a like this we will check and you know find out the kind of you know magnitude and the kind of you know solution there as for the particular you know requirement. So, now, what will you do? So, this is what the simple you know mathematical understanding about the presence of autocorrelation in any kind of you know econometric modelling, a that to when you process the you know the kind of you know solution for any kind of you know decision making.

So, we have to estimate the model and check the model diagnostics and here the diagnostic is with respect to error term. and the requirement is that the covariance among the error term should be equal to 0 and if that is not the case and that is the simple you know involvement of you know autocorrelation or that to the presence of autocorrelation in the system.

(Refer Slide Time: 17:18)



So, let us see how is this particular you know structuring. So, like what we have already discussed you know it is you know similar like you know multicollinearity problem. In

the case of multicollinearity it is the game among the independent variable that is explained in nature that is the X cluster and here it is the similar kind of you know structure. We like to check the relationship among the regressor that to unexplained independent variable and that to you know among the error terms.

So, like multicollinearity it relation you know there means there are 2, instances either there is a relationship or there is no relationship. If there is a relationship then there is autocorrelation, if there is a no relationship then there is no autocorrelation. Now, you know if there is autocorrelation either it will be negative or it can be a positive. So that means, we are looking for covariance or correlation to quantify the a relationship among the regressors.

So, whether it is a positive one or negative one, we do not have you know such kind of an issue. Ultimately it is a problem whether there negatively related or whether they are positively related, ultimately there is a relationship and there is a you know autocorrelation. And we like to check whether the particular relationship is a statistically significant or not or it is within a limit or you know beyond the limit. So, we have a kind of you know tolerance kind of you know structure through which you know autocorrelation can you know means may not have any problem.

But to when it will cross a particular limit it will have some kind of you know problem. It is exactly like you know you know multicollinearity case where you know we use VIF. If VIF is actually single digit and that too close to 0. So, the impact of multicollinearity is very less in the system.

But, here is also similar case and here like you know VIF so, we have a component called as a D W statistics that is called as Durbin-Watson d statistics which can give you the typical limit through which you can actually say whether there is the presence of autocorrelations or absence of auto correlations. And if it is a presence under what condition or what magnitude it will not create any problems after that it may create you know problem for any kind of you know management decision.

So; obviously, the simple understanding is that you know so, if the relationship is exactly 0 so, no problem. If the relationship is not 0 so, either it will be positive or negative it may not be also means it may not also problem. But, if the particular relationship is statistically significant then it will be a problem. So, we like to check whether it is

actually statistically significant or it going beyond the particular limit through which you know we cannot take a management decisions or we cannot use this particular model for any kind of you know engineering decision.

(Refer Slide Time: 20:24)

What is Autocorrelation

If this assumption is no longer valid, then the disturbances are not pairwise independent, but pairwise autocorrelated (or Serially Correlated).

This means that an error occurring at period t may be carried over to the next period $t+1$.

Autocorrelation is most likely to occur in time series data.

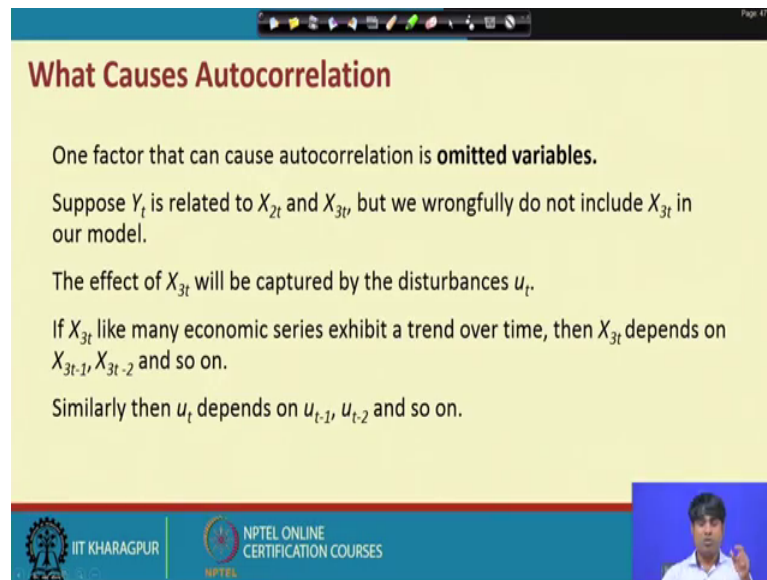
In cross-sectional we can change the arrangement of the data without altering the results.

The slide is part of an NPTEL presentation from IIT Kharagpur. It features a blue header with navigation icons and a footer with the IIT Kharagpur and NPTEL logos. The main content area has a yellow background.

So, likewise so, we have already discussed so, these are the kind of you know understanding through which you can you know project the autocorrelation. It may be a you know with respect to cross sectional you know structure; it may be also with respect to time series structure. But more frequent means a the particular problem is a very frequent in the kind of you know time series clustering. So, when we have a time series data and using the log variable, then autocorrelation usually a genuine problem.

But, when you are having cross sectional data the counterpart heteroscedasticity will be this genuine problem. But, by the way when you have a models and that to estimated model, which is actually a derived through the use of you know on data. So, whether it is a cross sectional data or time series data so, we have the estimated models. It is the mandatory requirement that we will check the model diagnostics and that to the autocorrelation issue and heteroscedasticity issue so, the.

(Refer Slide Time: 21:26)



What Causes Autocorrelation

One factor that can cause autocorrelation is **omitted variables**.

Suppose Y_t is related to X_{2t} and X_{3t} , but we wrongfully do not include X_{3t} in our model.

The effect of X_{3t} will be captured by the disturbances u_t .

If X_{3t} like many economic series exhibit a trend over time, then X_{3t} depends on X_{3t-1} , X_{3t-2} and so on.

Similarly then u_t depends on u_{t-1} , u_{t-2} and so on.

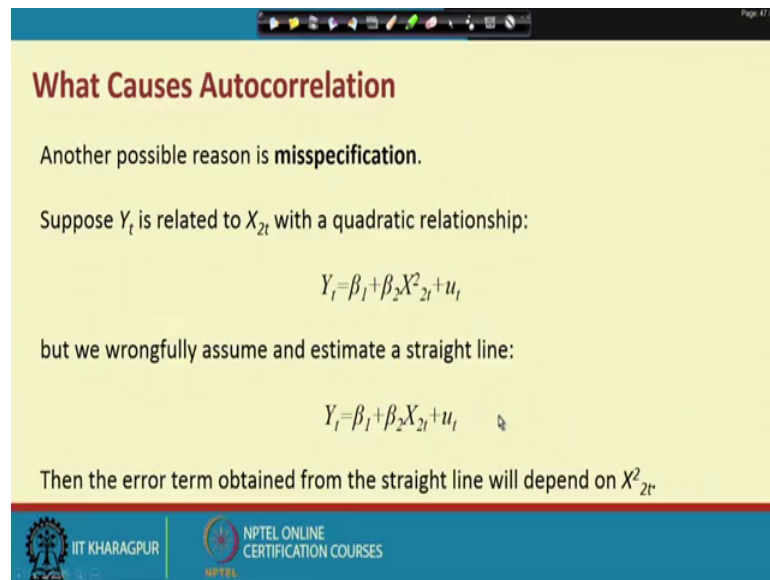
The slide is part of an NPTEL presentation from IIT Kharagpur. It includes a video feed of a male lecturer in the bottom right corner. The slide title is 'What Causes Autocorrelation'.

So, let us see how is the kind of you know structure through which you can you know understand the particular you know problem. And if you ask what is the reason for you know on the involvement of autocorrelation, there are many you know kind of you know regions. It is kind of you know multiple scenarios through which you can generate this problem. Like you know if the mathematical functionality is wrong and a if the variable specification is wrong and misspecification; that means, technically miss specification of the variable, omitted variable case.

And sometimes over identify models and the use of the log variables and sometimes when you manipulate the data and then increase the data size I mean say data structure through which actually look for the model accuracy. So, these are the instances or these are the, you know typical these instance through which actually autocorrelation may be appearing in the systems. There are many different ways you can say it will address or it will involve in the problem. So, our job is to check and they report what is the magnitude of in autocorrelation, whether it is a 0 or non-zero, if it is a non zero to what extent.

So, there are a couple of ways you can you know check and you know report the particular you know item.

(Refer Slide Time: 22:46)



What Causes Autocorrelation

Another possible reason is **misspecification**.

Suppose Y_t is related to X_{2t} with a quadratic relationship:

$$Y_t = \beta_1 + \beta_2 X_{2t}^2 + u_t$$

but we wrongfully assume and estimate a straight line:

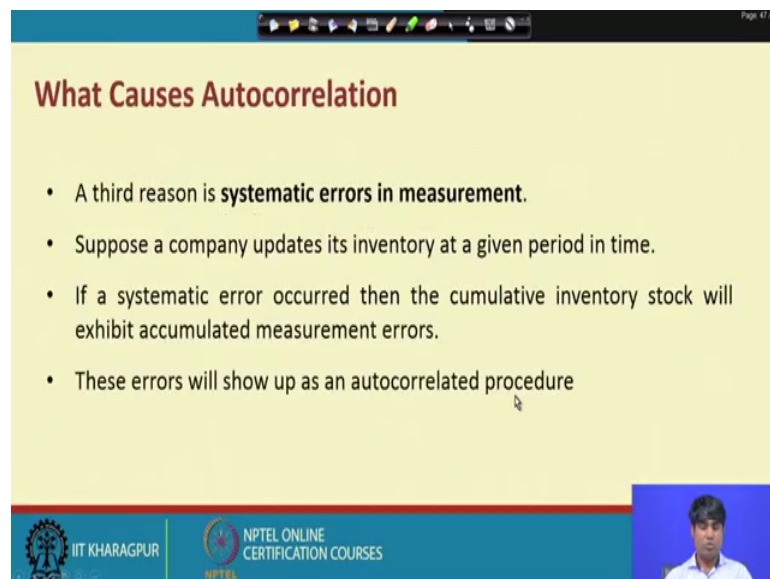
$$Y_t = \beta_1 + \beta_2 X_{2t} + u_t$$

Then the error term obtained from the straight line will depend on X_{2t}^2 .

IIT KHARAGPUR | NPTEL ONLINE CERTIFICATION COURSES

So, likewise these are the various you know a regions through which.

(Refer Slide Time: 22:52)



What Causes Autocorrelation

- A third reason is **systematic errors in measurement**.
- Suppose a company updates its inventory at a given period in time.
- If a systematic error occurred then the cumulative inventory stock will exhibit accumulated measurement errors.
- These errors will show up as an autocorrelated procedure

IIT KHARAGPUR | NPTEL ONLINE CERTIFICATION COURSES

You can autocorrelation can be a arise.

(Refer Slide Time: 22:55)

First-Order Autocorrelation

The simplest and most commonly observed is the **first-order** autocorrelation.

Consider the multiple regression model:

$$Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + \beta_4 X_{4t} + \dots + \beta_k X_{kt} + u_t$$

in which the current observation of the error term u_t is a function of the previous (lagged) observation of the error term:

$$u_t = \rho u_{t-1} + e_t$$

Handwritten notes on the slide:

- $H_0: \rho = 0$
- $H_A: \rho \neq 0$

The slide is from IIT Kharagpur NPTEL Online Certification Courses. A video inset shows a presenter.

So, now you know I will give you a little bit you know specification what exactly the autocorrelation and how we can actually simply detect and you know and commit a kind of you know conclusion. So, a basically it is a process outcome and that to it is just step by step process. So, we start with the simple theory and develop a models, then connect with the data, then have the estimated model, then with the help of estimated models we can derive the error term, then the game will start to check the autocorrelation.

So, like this you know so, let us say this is actually a multiple equation model you know model so, where we have one independent dependent variables and many independent variables that took k independent variables and then there is environment of error term. So, these are all called as a explained independent variable and these are all called as unexplained independent variables.

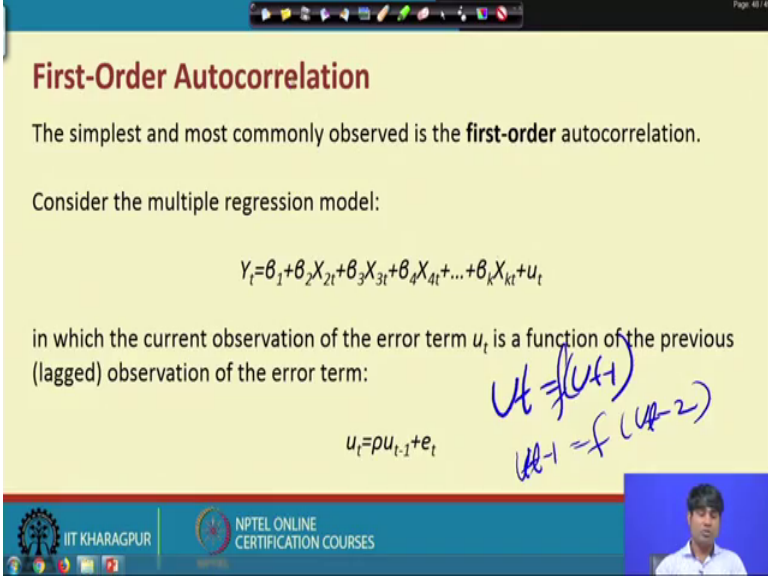
So, now, with the help of you know data and the use of you know less technique we will a get a estimated equation that is called as in U cap. So, by the way a error term will be removing the process. Now, the difference between actual Y and the you know estimated Y will give you the error terms. So, that is you know you can call as you know U_t or you know e_t . But, here's we will use the term U_t and like then we will check the kind of an autocorrelation.

So, mathematically if you say autocorrelation then you just you know link with the U_t with you know previous you know lag variables that is U_{t-1} and this is called as you know a correlation coefficient ρ .

So, here the null hypothesis that you know ρ should equal to 0 against the alternative are ρ not equal to 0. So that means, this ρ efficient that is the correlation between U_t and U_{t-1} and that value soon not be statistically significant. If that is significant, then there is a presence of autocorrelation and that is a; that too it is dangerous for any kind of you know decision making process.

So, now I will give you a little bit you know you know better understanding. if you like to know how this actually term related to each other. For instance a in this case simply U_t equal U_{t-1} .

(Refer Slide Time: 25:30)



First-Order Autocorrelation

The simplest and most commonly observed is the **first-order** autocorrelation.

Consider the multiple regression model:

$$Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + \beta_4 X_{4t} + \dots + \beta_k X_{kt} + u_t$$

in which the current observation of the error term u_t is a function of the previous (lagged) observation of the error term:

$$u_t = \rho u_{t-1} + e_t$$

Handwritten notes on the slide include:

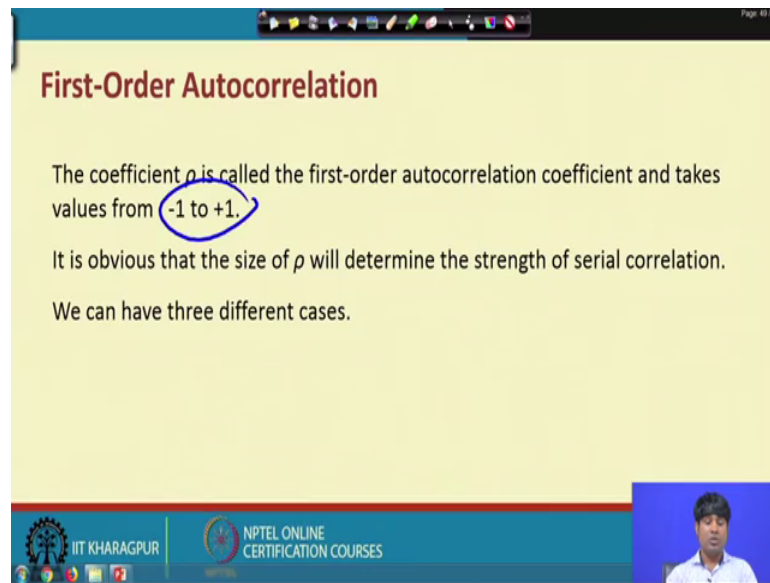
- $u_t = f(u_{t-1})$
- $u_{t-1} = f(u_{t-2})$

The slide footer includes the IIT Kharagpur logo and the text "NPTEL ONLINE CERTIFICATION COURSES". A small video inset of a presenter is visible in the bottom right corner.

So; that means, technically it is question of you know U_t and U_{t-1} that is that too the this case U_t equal to U_{t-1} the to function of U_{t-1} .

So, now against the U_{t-1} U_{t-2} may be function of U_{t-2} , again U_{t-2} can be a function of U_{t-3} so, it is like you know chain. So, obviously, there will be definitely you know presence of you know relationship among these error terms. So, our duty is to check and find out the magnitude whether it is actually under the tolerance or beyond the tolerance.

(Refer Slide Time: 26:11)



First-Order Autocorrelation

The coefficient ρ is called the first-order autocorrelation coefficient and takes values from -1 to $+1$.

It is obvious that the size of ρ will determine the strength of serial correlation.

We can have three different cases.

The slide is part of an NPTEL presentation from IIT Kharagpur. It features a title bar with navigation icons and a page number 'Page 43/50'. At the bottom, there are logos for IIT Kharagpur and NPTEL Online Certification Courses, along with a small video feed of the presenter.

So, likewise you will see how is the particular you know structure and ultimately like you know correlation coefficient, the autocorrelation coefficient range is also between you know minus 1 to ones like heres. So, what we have already discussed it may be the negative one or it may be the positive one. But, but ultimately there is a relationship and that relationship should not be so high and should not be statistically significant while using this estimator model for any kind of you know decision making process.

And ultimately rho is this symbol here to get to know, the kind of you know presence or absence of absence of an autocorrelation.

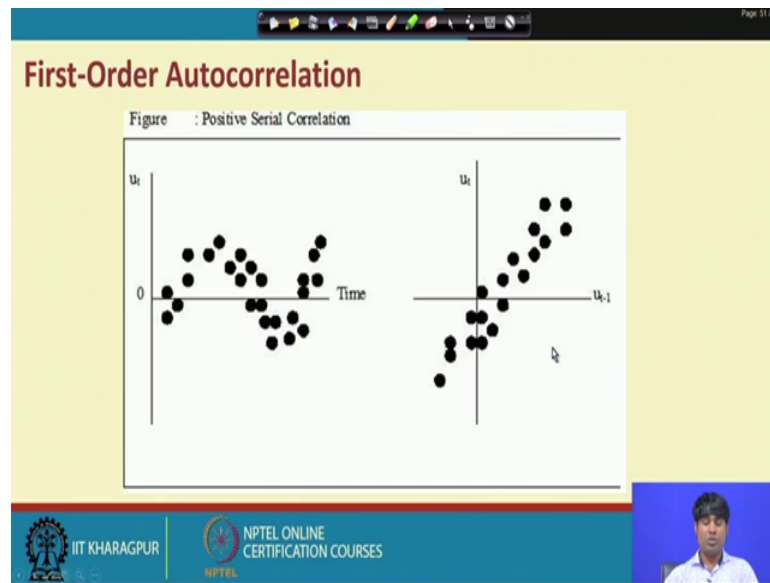
(Refer Slide Time: 26:50)

The slide is titled "First-Order Autocorrelation" in red. It contains three bullet points: (a) If ρ is zero, then we have no autocorrelation. (b) If ρ approaches unity, the value of the previous observation of the error becomes more important in determining the value of the current error and therefore high degree of autocorrelation exists. In this case we have positive autocorrelation. (c) If ρ approaches -1, we have high degree of negative autocorrelation. Handwritten notes in blue ink include the regression equation $Y = \beta_0 + \beta_1 X_t + u_t$ and the error term $u_t = \rho u_{t-1} + v_t$. A blue arrow points from the handwritten u_t equation to the word "autocorrelation" in point (b). The slide footer includes the IIT Kharagpur logo and the text "NPTEL ONLINE CERTIFICATION COURSES". A small video inset in the bottom right corner shows a man speaking.

So, now the simple you know modelling kind of you know structures; that means, you start with the simply Y equal to β_0 plus $\beta_1 X_1 t$ plus u_t . And then U_t is a function of you know, ρU_{t-1} plus U_t and other error terms and ρ is the kind of an autocorrelation coefficient. Then either ρ will be equal to 0 or not equal to 0, if it is a 0 then autocorrelation and if it is not equal to 0, it may be approaching towards you know minus 1.

And if that is the case it is called as a negative autocorrelation and if not then it will be approaching towards you know plus 1. Then this will be called as you know positive autocorrelation. If we exactly plus 1 then it is called as a perfect positive autocorrelation, if it is exactly minus 1, then it is called as a perfect negative autocorrelation. And a likewise so, the nature may be positive the nature may be negative, but do we have no issue whether it is a positive or negative, but ultimately will check what is the level.

(Refer Slide Time: 27:57)



So, this is how the kind of you know a graphical for plotting about the error terms and we like to check whether there is a kind of you know functional form among the error terms. If there is a functional form by default so, there is a kind of you know autocorrelation and that to it will give you a give you some kind of an indication about the a you know model kind of you know diagnostics.

And a ultimately so, we can graphically inspect but so, far as a quantification is concerned, we can actually find out the autocorrelation coefficient that is the rho here. And then we check whether you know it is actually problematic case or not probability case. Of course, there are a couple of methods you know technical methods are there to get to know the level of auto correlations and then we can check the kind of you know tolerance limit.

So, 1 such statistic which I have already highlighted is the Durbin-Watson d statistics.

(Refer Slide Time: 29:03)

Higher-Order Autocorrelation

Second-order when:
 $u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + e_t$

Third-order when:
 $u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + \rho_3 u_{t-3} + e_t$

p-th order when:
 $u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + \rho_3 u_{t-3} + \dots + \rho_p u_{t-p} + e_t$

Handwritten notes:
 $d = 2(1 - \rho)$
 $\rho = 0$, no Auto
 $\rho \neq 0$, Auto
 A graph showing a line with a positive slope, labeled with 1, 2, 3, 4, 5 on the x-axis.

And Durbin-Watson statistics the structure is that you know when rho equal to 0 D W statistic will exactly equal to you know 2.

(Refer Slide Time: 29:17)

Consequences of Autocorrelation

1. The OLS estimators are still unbiased and consistent. This is because both unbiasedness and consistency do not depend on assumption 6 which is in this case violated.
2. The OLS estimators will be inefficient and therefore no longer BLUE.
3. The estimated variances of the regression coefficients will be biased and inconsistent, and therefore hypothesis testing is no longer valid. In most of the cases, the R^2 will be overestimated and the t-statistics will tend to be higher.

So; that means, actually Durbin-Watson d statistic structure is like this. So that means, actually when I am connecting with it actually U_t upon U_t minus 1. So, then I can find out the auto correlate autocorrelation coefficient. Then with a particular you know setup so, we can calculate Durbin-Watson d statistics that is you know small d here.

So, it is simply the equation equal to $2 - \rho$. So, when ρ equal to 0 technically, then we have already declared there is a no autocorrelation and when ρ not equal to 0 there is a autocorrelation. So, now, when you put ρ equal to 0 where the signal is No autocorrelation, then by default in this case d exactly equal to 2.

So, now, when ρ not equal to 0 then the d value Durbin-Watson d value it may deviate actually above them too or you know less than 2. So, as a result so, DW statistic will be here to is the it kind of you know extreme. So, when exactly equal to 2 DW statistics exactly equal to 2 then there is an autocorrelation. Otherwise it may have auto correlation whether it is a less of it or whether it see more of more of it. But, technical is if you talk about the tolerance then we can have range here 1.5 and we can arrange here actually 2.5.

So, now in between if the Durbin-Watson statistic will lie, then this will not be serious problem so, for as you know decision making process is concerned. Otherwise it will be a serious problem. So, these are all different orders through which actually U_t is related and that means, it will give you some kind of you know kind of you know high order you know autocorrelation. When all these error; error terms are you know correlated each other over the time.

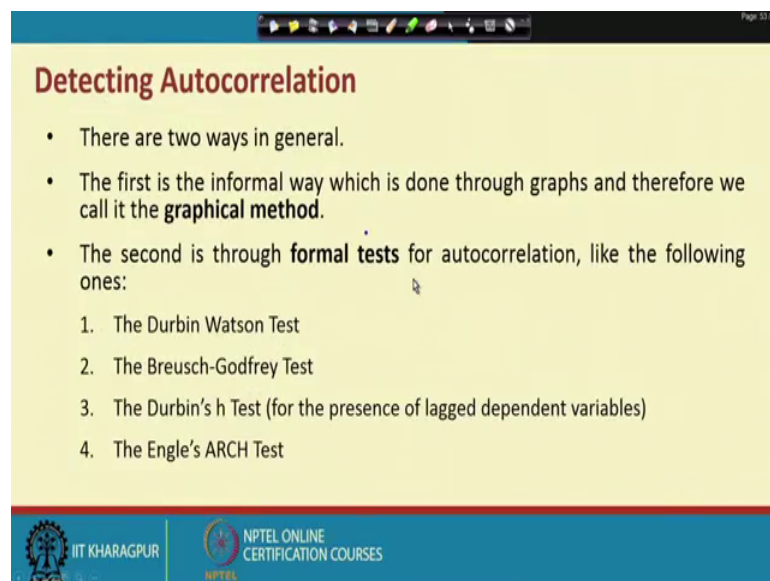
And then at the end of the day you defined the involvement of autocorrelation which at the highest level and that to it will be a danger signal for any kind of you know decision making process. So, likewise we will check and you know get to know what is the level of you know autocorrelation.

And by the way after knowing the a detection of you know autocorrelation we are very keen to know the kind of you know consequence and the kind of you know solution. So, but as a consequence is concerned so, the model simply cannot be used for prediction, forecasting or any kind of you know decision making.

And a because most of the instances the variables will be not significant because, this really a increased the standard error of the estimates and as a result the variables involved in the particular you know modelling will not be significant. So, not all the variables kind of you know requirement.

But, some of the variables maybe not significant and that may be because of you know the presence of in autocorrelation. And simple understanding is that you know when autocorrelation is there standard error standard error of the estimate should be very high and it imply actually effect the significance levels of the parameter. When it will affect the significance level of the parameter and as a result that particular variable will not significant for the you know model estimation or the kind of you know model forecasting.

(Refer Slide Time: 32:58)



The slide is titled "Detecting Autocorrelation" in a bold, dark red font. It contains a bulleted list of two general ways to detect autocorrelation, followed by a numbered list of four formal tests. The slide has a yellow background and is part of an NPTEL presentation from IIT Kharagpur.

- There are two ways in general.
- The first is the informal way which is done through graphs and therefore we call it the **graphical method**.
- The second is through **formal tests** for autocorrelation, like the following ones:
 1. The Durbin Watson Test
 2. The Breusch-Godfrey Test
 3. The Durbin's h Test (for the presence of lagged dependent variables)
 4. The Engle's ARCH Test

Page 33/34

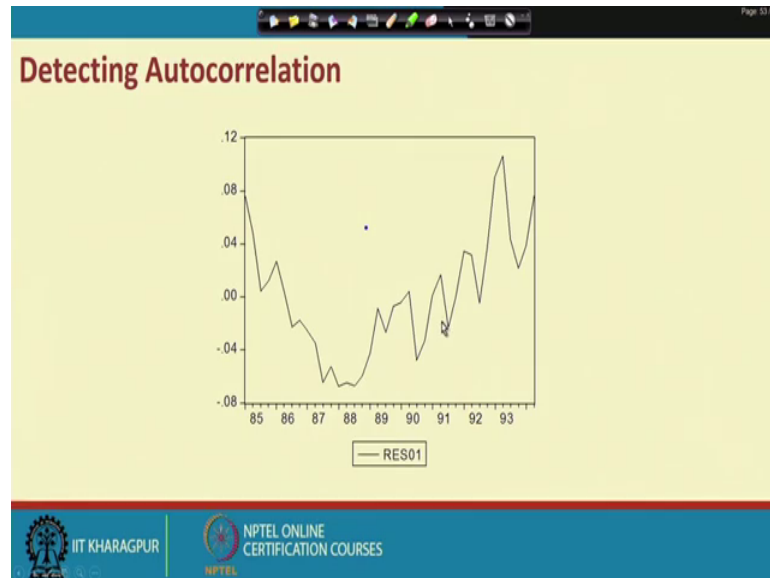
IIT KHARAGPUR | NPTEL ONLINE CERTIFICATION COURSES

So, now we will have a different kind of you know detection criteria, which I have already highlighted a graphical mechanism and I have also given you the technical mechanism or the best you know where to represent the autocorrelation is through Durbin-Watson d statistic. Because, it will give you a kind of you know quantitative structure through which you can actually exactly a you know identify the autocorrelation issue whether the kind of you know autocorrelation level is there or not there.

So; that means, when d Durbin-Watson d statistics exactly equal to you know 2 then there is not auto correlations if not then there is a autocorrelation. Again if it is within between 1.50, 2.5 that can be acceptable, otherwise you just you know restructure the model and have different output, again check the you know model diagnostics and that autocorrelation and a you know see whether the restructuring model is again free from

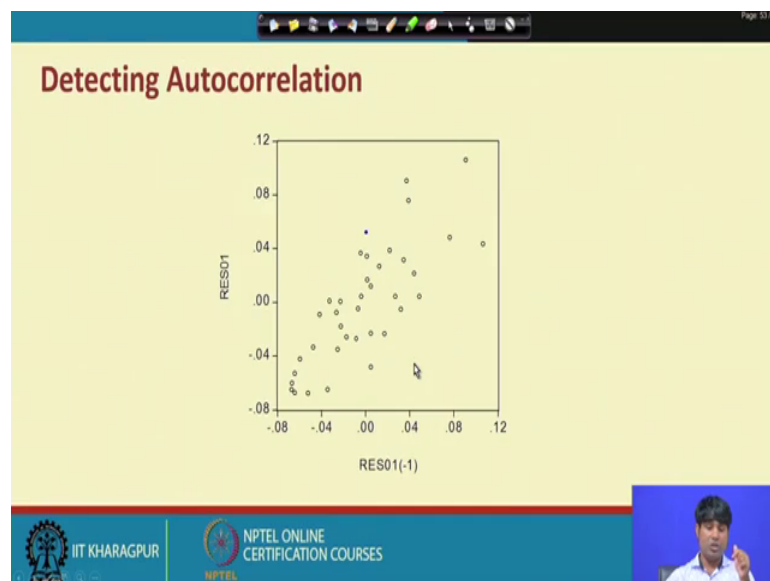
autocorrelation. Once you get a models which is actually under the control then you can use for you know decision making process.

(Refer Slide Time: 34:18)



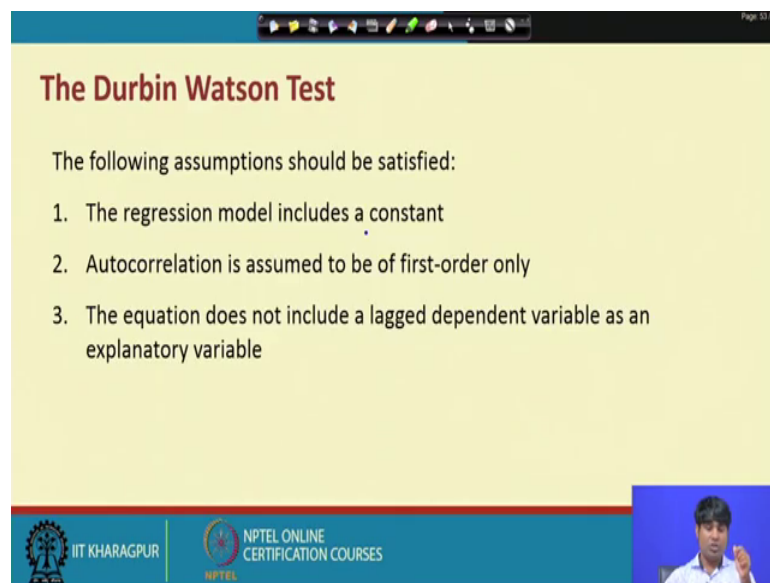
So, until unless you get the clean or you know green environment you should not actually use this model for any kind of you know decision making process. So, these are you know different graphical floating which you can check.

(Refer Slide Time: 34:24)



Whether there is a kind of in a relationship.

(Refer Slide Time: 34:28)



The Durbin Watson Test

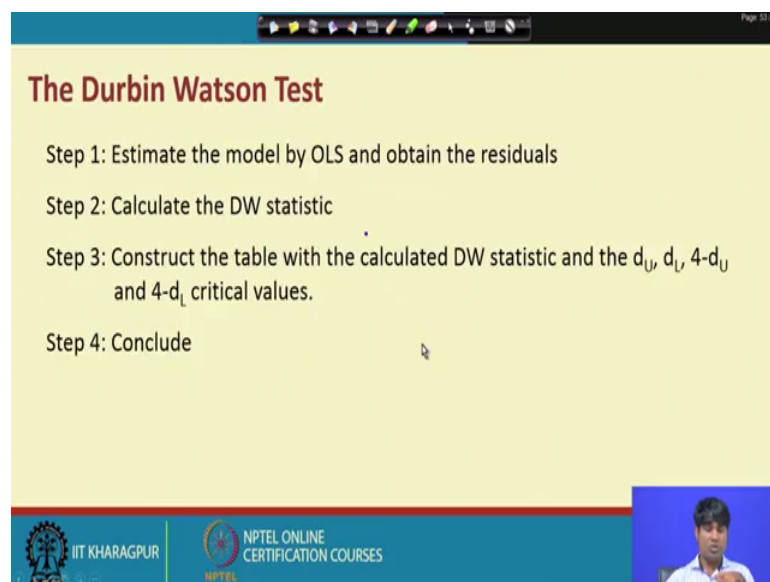
The following assumptions should be satisfied:

1. The regression model includes a constant
2. Autocorrelation is assumed to be of first-order only
3. The equation does not include a lagged dependent variable as an explanatory variable

IIT KHARAGPUR | NPTEL ONLINE CERTIFICATION COURSES

And this is another way of you know checking the kind of you know presence of autocorrelation. And Durbin this is what the Durbin-Watson d statistics.

(Refer Slide Time: 34:35)



The Durbin Watson Test

- Step 1: Estimate the model by OLS and obtain the residuals
- Step 2: Calculate the DW statistic
- Step 3: Construct the table with the calculated DW statistic and the d_U , d_L , $4-d_U$ and $4-d_L$ critical values.
- Step 4: Conclude

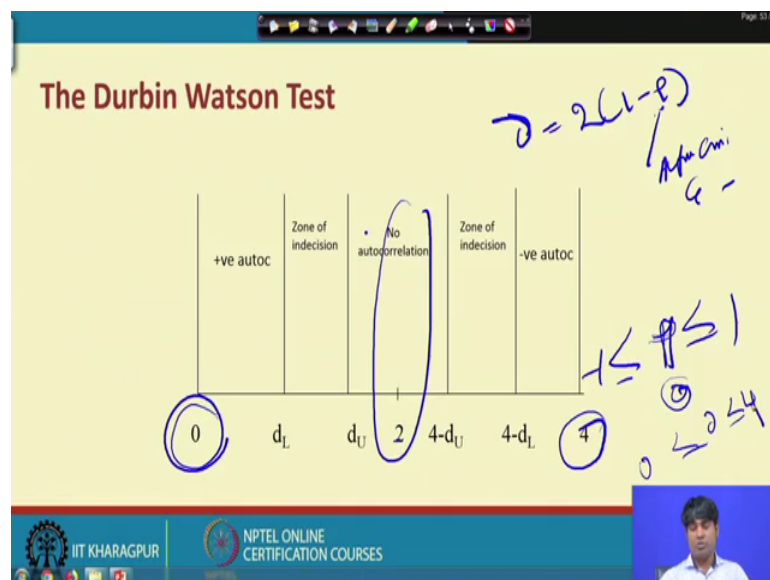
IIT KHARAGPUR | NPTEL ONLINE CERTIFICATION COURSES

And it is typically step by steps you know process, first to you know connect you know dependent variable to independent variable, help the estimated model, get the error terms and then connect error term with the previous error terms and get the autocorrelation coefficient. After knowing the autocorrelation coefficient you can get to know whether there is the presence of autocorrelation or you know absence of autocorrelation problem.

That means, technically if D W is exactly equal to 2, no autocorrelation if not then there is a autocorrelation. This is the simple you know understanding or the kind of you know check which you can do in the process of you know modelling. That is how among all the test Durbin -Watson statistic is very good ones and most of the subtests will provide the D W statistic directly without going through the step by step by step process.

So, what you can do? Just you run the model, have the detailed statistics and check what is the accuracy of the models. If the accuracy of model is then the D W statistic by default will support the kind of you know accuracy of the model by having actual D W statistic d equals to 2, if not then it you know at the problem. So, that we cannot actually use for any kind of you know decision making process.

(Refer Slide Time: 36:02)



So, that is the procedure through which actually you can go ahead the decisions and. While using Durbin-Watson d statistics for you know decision making process that to declare the presence of autocorrelation or absence of autocorrelation. So, the simple approach is D W must be exactly equal to 2, not a correlation if not it may be positive it may be negative

But there is a extremes, like you know autocorrelation coefficient can simply range between minus to minus 1 to plus 1 so; that means, technically if you would see here. So, D W statistic will be simply written like you know to equal 2 into 1 minus rho is the autocorrelation coefficient and autocorrelation coefficient. And then when rho equal to 0

then d equal to 2; that means, there is a no autocorrelation situation, when ρ equal; that means, ρ will be vary from minus 1 to 1.

So, in between the 0 will be coming into the picture so; that means, 0 is the middle one and when you report you know ρ equal to minus 1 then it becomes actually a D W statistical way for and when you put actually ρ equal to exactly 1 then and in D W statistics will be 0. So, when the autocorrelation coefficient is moving from minus 1 to 1 that is the usual limit so, the Durbin-Watson d statistic will vary from 0 to 4.

So; that means, technically a d will be ranging from 0 to 4 with you know high positive autocorrelation to high negative auto correlation. Ultimately ideal range is D W should be exactly equal to 2 and if not ideal, then it may be negative or positive. But, it is a dangerous when the D W will be exactly equal to 0 or D W exactly equal to 4.

So; that means, any D W statistically close towards 0 or close towards 4 is more danger signal and in that case you cannot just use the model for any kind of you know decision making process. So, you have to restructure, re estimate, redefine, then you look for the alternatives and again check then finally, you should accept the model which is exactly you know free from autocorrelation or else within a particularly you know tolerance range that to between 1.5 to 2.5.

So, this is how the kind of you know check process of you know Durbin-Watson d statistic through which you can take a decision. And with this we will stop here in the next class, I will take this problem and then calculate manually the D W statistic. And we can report what is the level of you know Durbin-Watson d statistic and whether there is in need of you know further restructuring to solve the autocorrelation problem, provided if the autocorrelation is coming very high.

Thank you very much have a nice day.