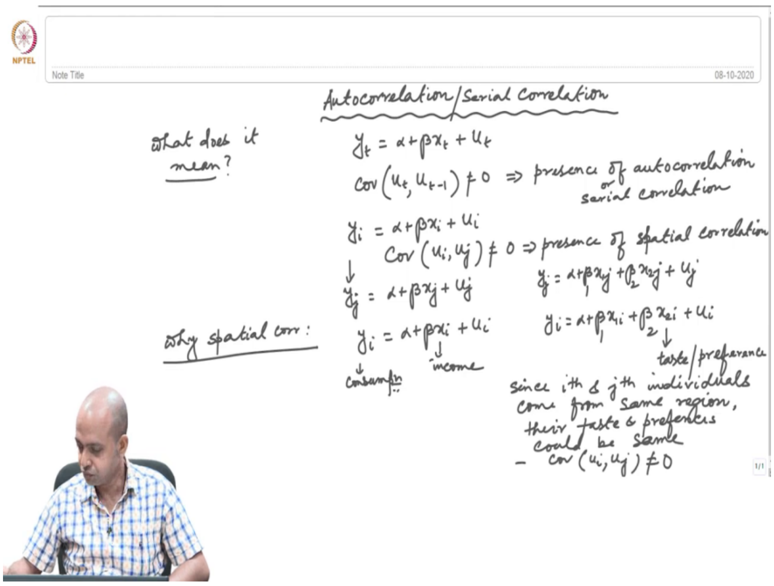
Introduction to Econometrics
Professor Sabuj Kumar Mandal
Department of Humanities and Social Sciences
Indian Institute of Technology, Madras
Lecture 48
Relaxing the assumptions of CLRM-Multicollinearity and Autocorrelation Part – 5

Welcome once again. As you know we are discussing about relaxing some of the important assumptions of classical regression model, what will happen if some of the assumptions are violated and firstly we discussed about multicollinearity. Basically if the no multicollinearity assumption is violated, then what would be the consequences, how to detect the problem of multi collinearity and what are the remedial measures? Today we will discuss about another assumption which is called autocorrelation or serial correlation.

(Refer Slide Time: 1:04)



So, let me write it down, this is called autocorrelation or sometime it is called serial correlation. So, first we will try to understand the meaning of this autocorrelation. What does it mean? Suppose we are dealing with a time series data and this is y t equals to alpha plus beta x t plus u t. This is a time series data, y t is let us say your consumption and x t is your income, at the country level, we have taken consumption and income over a period of time.

And what we assume in this model that covariance between u t and u t minus 1 is actually 0. u t and u t minus 1 is 0. So, that means the error terms for the ith, for the tth period and the error term of the t minus oneth period they are not at all correlated. That is why we make this assumption. This is the meaning and if this is not equals to 0, then only we say that there is a presence of autocorrelation.

Similarly, if you are dealing with a cross sectional data, let us say yi equals to alpha plus beta xi plus ui, then we assume covariance between ui and uj, it is equal to 0 and if it is not equal to 0, then we say that there is presence of no autocorrelation but spatial correlation. So, both autocorrelation and spatial correlation will give you similar type of adverse impact on your estimates, but meaning wise they are different.

While autocorrelation or serial correlation, this is also called as serial correlation is a time series phenomenon, serial correlation. Spatial correlation is more of a cross sectional phenomenon that means it happens in the context of a cross sectional data, so that is why it is called spatial correlation. So, that means error term for the ith person and the error term for the jth persons equation should not be correlated because this similar equation I write for the jth person as well. Instead of yi I can write let us say yj equals to alpha plus beta xj plus uj that is for the jth person and let us say we are still talking about the consumption function.

Now, the question here is why do we get autocorrelation problem? First we will talk about the cross sectional data, and then we will go back and discuss the reasons for autocorrelation in the context of time series data in detail. let us say first talk about the spatial correlation. So, when I am saying that yi equals to alpha plus beta xi plus ui, let us say that this is your consumption and this is your xi which is your income.

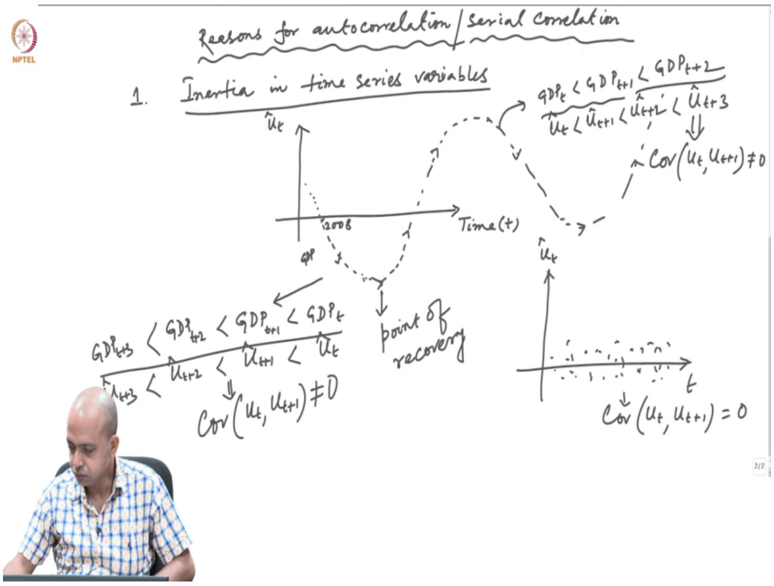
Now as you know that consumption depends not only on income, but also taste and preferences which we have not included in this model. So, that means, your true model could be yi equals to alpha plus beta1 x1i plus beta2 x2i plus ui and I have omitted the x2i variable and that is the reason x2i is now included in the error term itself. So, in this model, taste and preference is not included that is some kind of unobserved or omitted variable so that is why it is there in the error term.

Now when you are collecting the data, cross sectional data from the same region, it might so happen that the taste and preference for the i^{th} person and taste and preference for the j^{th} person they are similar, taste and preference; since they are coming from the same region. So, that means, what I saying the taste and preference for the i^{th} person and j^{th} person they are similar. As a result of which this ui and this uj they might be correlated, ui and uj might be correlated.

So, this is the similar equation if I write alpha plus beta1 x1j plus beta2 x2j plus uj. This x2j and x2i they are basically denoting taste and preferences for the i^{th} and the j^{th} person. So, what I am saying since i^{th} and j^{th} individuals come from same region, their taste and preferences could be same and that may result in covariance of ui and uj not equals to 0, so that means they are correlated.

So, we might have omitted a variable taste and preference from the equation; that is why the taste and preference is now there in the error term and since the taste and preference of the i^{th} and j^{th} individual are similar, because they are coming from the same region, it may so happen that the error term of the i^{th} person and error term of the jth person, they actually capture the same omitted variable which is taste and preference and that might be the reason for spatial correlation. Now let us talk about the reason for autocorrelation in detail.

(Refer Slide Time: 10:35)



So, this is reasons for autocorrelation or serial correlation. So, the first reason or the most important reason that we are talking about is the inertia. Inertia in time series variable. Inertia means that the variable gets similar type of pattern in over a period of time. I will give you an example of that. Let us say that we are talking about India's GDP and we have data from last 30 years and as you know that India's GDP, this macroeconomic variable, Indian economy experienced a shock which was given from outside, let us say in 2008 and 2016.

Let us say that we are talking about the 2008 financial crises and it may so happen that because of this financial crises the impact of this external shock in the form of financial crises, what is given in the economy that will therefore the economy for a longer period of time that means the impact of crises on the macroeconomic time series variable may not die down soon rather, the impact may persists for a longer period of time.

And as a result of which what will happen let us say that in a simple diagram I will represent this. Let us say in x axis I am measuring time t and y axis I measure ut. So, the shock here is treated as the error which is unobserved but that also explain how much GDP or growth the economy will achieve. We all know there is an impact of crises but we are not able to observe that particular crisis that is why we are not able to include that in the model and that is the reason it is there in the error term.

So, what will happen? Let us say in a particular time period this is 2008. So what will happen? Your GDP will show some kind of pattern and as a result of which at the point of 2008 when the financial crises is there so at each successive period, your GDP will go down. So, there would be a downward trend for your GDP and what does it indicate? This downward trend indicate that GDP at time period t minus 1, sorry this is let us say GDP at t plus 1 is less than GDP at t and GDP at t plus 1 is again less than GDP at t plus 2, it is again less than GDP at t plus 3 like that.

So, at each successive period, your GDP is going down. So, that means, there is some kind of pattern you see, there is some kind of pattern and if this is the case, then ut would be greater than ut plus 1, ut plus 1 would be greater than ut plus 2 and ut plus 2 would be greater than ut plus 3. Likewise, it will come down for a some, for some time period and then once the economy starts recovering, then what will happen, you will experience a down, an upward trend.

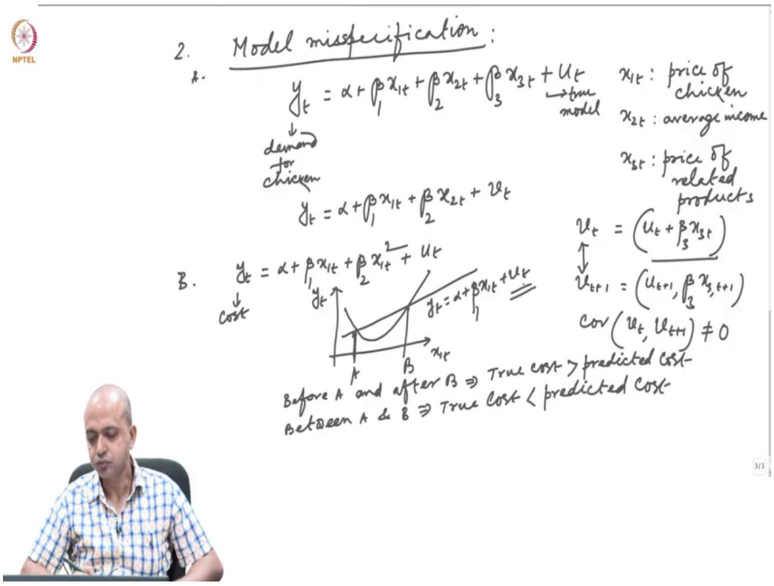
And then it will again come down and then it will again go up. So, this is coming down, this is going up, this is going up, this is called a business cycle basically and because of this business cycle, what I am saying when it is going up, then the reverse pattern what will happen? GDP at t, tth period would be less than GDP at t plus 1 and that is again less than GDP at t plus 2. So, the reverse pattern you will observe and as a result of which u t would be less than u t plus 1 and that would be again less than u t plus 2 and that would be again greater than u t plus 3. This is one pattern when the economy is going down.

That means when you are getting the negative shock and in every successive period, the growth is actually coming down and then this is the point of recovery; this is the point of recovery. It is going up and when it is going up, then this is the pattern and what does it mean? u t is less that u t plus 1 which is less than u t plus 2 which is less than u t plus 3. So, this is another pattern.

Whatever happens when the economy is going downward or upward, we see come kind of pattern in this u t and that is why we say that the in each successive period the error terms are actually correlated. Ideally what we should get if we plot time here and your u t, this is actually u t hat I would say u t. These are all hat. Ideally if you plot u t hat here, then I should not observe any pattern, there should not be any pattern. This basically indicates covariance between ut and u t plus 1 is actually equals to 0.

But both these, this equals to, this indicate covariance between u t and u t plus 1 actually not equals to 0, similarly this u t, this also indicates covariance between u t and u t plus 1 not equals to 0 and why this is happening? Because we observe some kind of inertia in the macroe6conomic time series variable; so, some kind of momentum gets involved in your time series macroeconomic variable, and as a result of which in every successive periods your error term gets correlated. So, that is one of the prime reason for autocorrelation to happen in time series data. Now let us say we are talking about another reason. This is reason two.

(Refer Slide Time: 20:16)



This is reason two for autocorrelation and that might be due to model misspecification, model misspecification. Suppose again I am talking about a demand function, demand for chicken which is a function of alpha plus beta 1 x 1 t plus beta 2 x 2 t plus beta 3 x 3 t plus u t, where x 1 t is basically what is x 1 t? x 1 t is let us say price of chicken, let us say this is demand for chicken, this is price of chicken, then x 2 t is basically average income demand for chicken at the city level, district level, state level whatever and then x 3 t is let us say price of related products may be mutton.

So, this is your true model. But by some mistake you are basically specified this type of model y t equals to alpha plus beta 1 x 1 t plus beta 2 x 2 t plus v t. That means, v t here is basically u t plus beta 3 x 3 t. So, since you are not including price of related product in the model, this is your true model, this is your true model, so the impact of the related product will be captured by v t that is why v t is specified in this way.

Now in every successive period, then again v t plus 1 will also capture u t plus 1 into beta 3, then x 3 t plus 1. So, same variable would be there for each and every successive time period and as a result of which we will get some kind of correlation between v t, v t plus 1 so on and so forth. Why this is happening? This is happening because you have miss-specified the model. So, autocorrelation is arising so that means here you will experience v t and v t plus 1 they are actually not equals to 0 which is arising due to model misspecification.

Model misspecification and how the model misspecification is arising here? Model misspecification is arising because you have omitted an important variable from the model. Let us say this is A. Similarly, model misspecification if you recall we have discussed earlier may arise due to improper imposition of improper functional form. For example, let us say you are estimating a cost function, you are estimating a cost function which is let us say a function of output plus output square; output and output square plus u t.

That means basically you are trying to estimate this type of crossed function, here it is output x 1 t, here it is the, this is cost, this is cost y t and this is output. So, the function will look like this as you all know. (())(25:41) and u shaped. It will first as production increases, output, cost per unit or average or marginal cost they will come down and they will go up. But instead of that suppose you have omitted this x 1 t square, so that means basically you are running a linear regression given by y t equals to alpha plus beta 1 x 1 t plus u t.

This is the model. Suppose this is your instead of u t let me write v t. So, that means, once again this v t captures this x 1 t square impact of x 1 t square would be there in v t and in each successive period this v t will capture the same variable x 1 t square. So, that means, what will happen here? For let us say this is the point A and this is point B. So, that means before A your true cost is actually lower than the predicted cost and beyond B your true cost is actually, before A and after B so at this side and that side what is happening, the true cost is higher than the predicted cost and between A and B the true cost is lower than the actual cost.

True cost is lower than the predicted cost. So, that means, at each successive period, this between A and B there you will observe some kind of pattern and beyond, before A and after

B also you will observe similar type of pattern because your true cost is greater than the predicted cost and as a result of which you will observe some kind of pattern in your error term and as a result of which you will observe some kind of autocorrelation problem which is happening due to improper functional form.

So, what I am saying, before A and after B true cost is greater than predicted cost and between A and B true cost is less than predicted cost. So, the error term will show some kind of pattern before A and B and another type of pattern between A and B. Whatever might be the case, you will definitely observe some kind of pattern. At each point the true is greater than this. So, that means, if this is your true model u t would at every successive period you will observe some kind of pattern because you are basically estimating this which is happening due to the model misspecification.