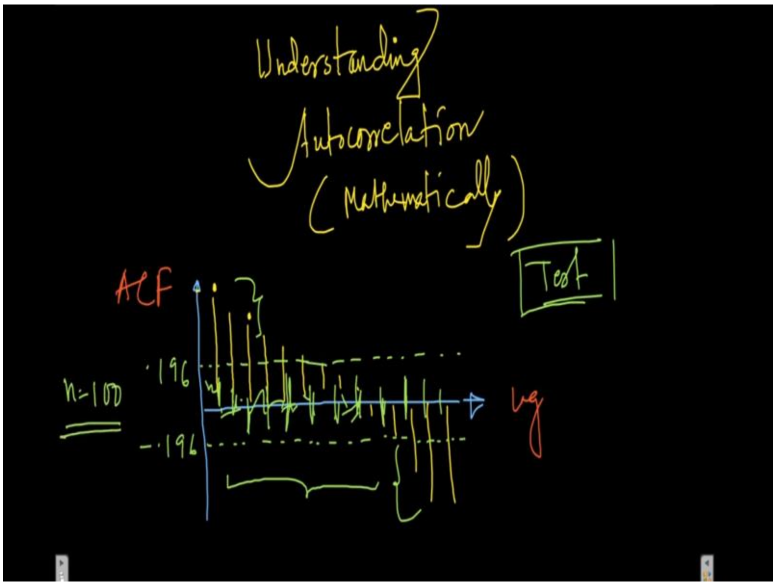
Applied Econometrics
Prof. Tutan Ahmed
Vinod Gupta School of Management
Indian Institute of Technology – Kharagpur

Lecture – 72
Autocorrelation (Contd.)

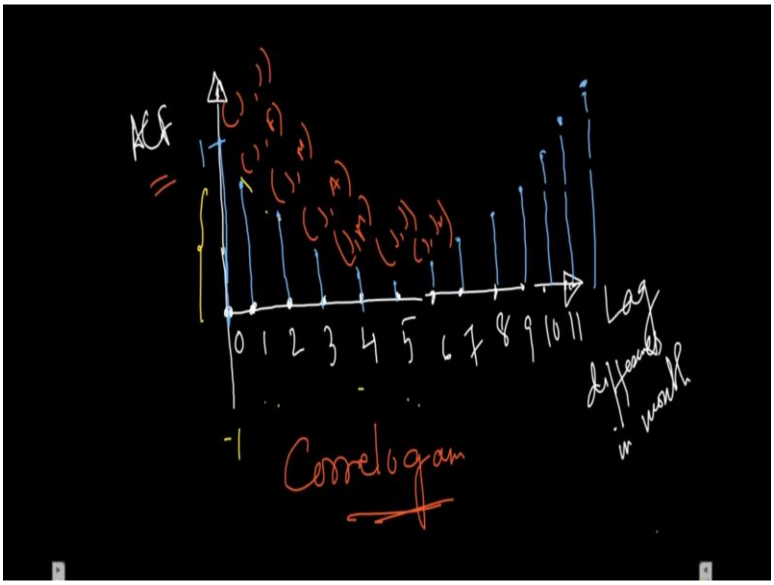
Hello and welcome back to the lecture on applied econometrics. So, in this lecture we are going to talk about autocorrelation.

(Refer Slide Time: 00:31)



And in this lecture, we will try to understand autocorrelation mathematically and it is going to help us in understanding the consequences autocorrelation. So, little bit of maths here.

(Refer Slide Time: 00:47)



Now, in the previous lecture, we have seen how, actually, if we have autocorrelation how it will look like? So, we will see some sort of pattern. So, if we actually plot this ACF with the lags, we can actually observe this sort of pattern here. Now, this pattern could have different shapes. And we could actually have some idea about these patterns. But one thing we need to remember is that we are actually plotting the correlation coefficient

And the correlation coefficient always varies between + 1 to - 1. So, we can expect the ACF to vary between + 1 to – 1. And therefore, the pattern that we will see, it will play in the range of + 1 to - 1. Now, if we think let us do an intuitive thinking and let us say that we just plot similarly, the ACF here, ACF on the y axis and the lag on the x axis, now I write lag and ACF.

A simple point, I want to make here is that if I have let us say, in one period, I have my ACF is close to 1, let us say but as I move in time as I actually progress in time. What will happen is that the impact? Or the correlation of what happened here? And what will happen here? They actually tend to decrease like as the time progresses. It is an intuitive thinking, as the time progresses whatever happened in the past, the impact of that is going to be less and less and less.

Now, so, that is basically usually we see these you know, sort of plot, it can also be it can actually oscillate like there could be a lot of cyclical, seasonal or cyclical pattern. So, where we will see that the ACF to vary between + 1 to - 1? So, these types of relationships, this types of plots, you will see time and again when we plot ACF with lag. Now, the point is of course with ACF, we can understand if there is autocorrelation.

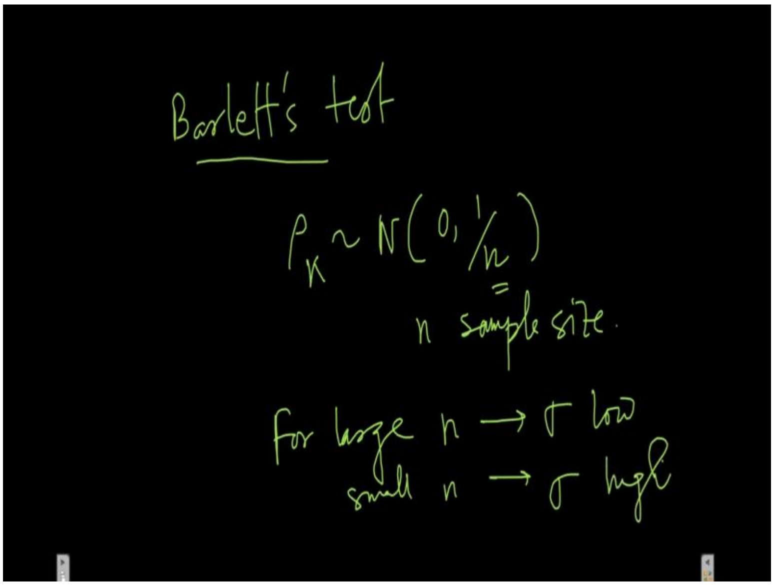
But, what more? Is there anything more? And there is something very interesting about this is that we can actually identify if the autocorrelation is significant. So essentially, it can also help us to test if this autocorrelation is significant or not. Now, how we; actually use this ACF and the lag plot or correlogram plot to actually identify if the autocorrelation is significant or not?

So, let us just think intuitively without getting into any complexity. Let us just think let us say if the correlation coefficient between 2 successive time period is actually less then perhaps autocorrelation is low. So, that means the error terms are not perhaps the correlation coefficient among the error terms are actually low. So, that they are not highly related.

So, then if the; correlation coefficients are low. Then perhaps I can say that well autocorrelation is not that much and to actually have some benchmark or some sort of threshold, we can actually have something like this. Let us say, if the correlation coefficients between 2 successive periods is actually lying or between this range. Let us say, all the ACA values are lying here it can go either side but it is a range is always constant is fixed.

So, as long as the ACL values are this low, so then I would say that well, perhaps autocorrelation is not that high but if it exceeds, if it goes in this zone in these regions or in this region. So, then I will say, well so the autocorrelation is actually a little you know, there are autocorrelation present. So, we need to really talk about that. Now, the test, one test that we do is called Barlett's test.

(Refer Slide Time: 04:56)



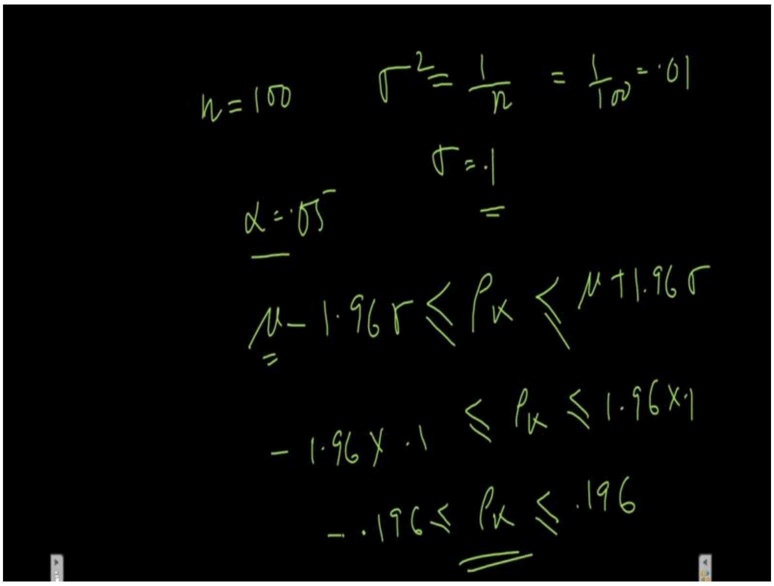
So, the test we do is called Barlett's test. So, it actually uses the same idea where we have some sort of threshold value for the correlation coefficient and if the correlation coefficient is more than the threshold value. Then we say that well the autocorrelation is significant and for Barlett's test, if we actually identify the auto ACF the correlation coefficient is rho and k being the lag and if we have rho k is the value of the ACF.

It will follow a normal distribution of 0 mean and 1 by n standard deviation variance. Now, were n is essentially the sample size, n is the sample size. Now, this is essentially if we can see that it is actually depending on the samples I then nothing else. And I just explained intuitively why this is the case, if we have large n, for large n what we will have is that the standard deviation.

The standard deviation is going to be low, sigma is low. For small n, the sigma is going to be high. Now, it means that for large n I am going to see very low variation. So, with large sample size I am going to see low variation now relate it to the previous diagram. So, with large n you will have more lags. And the more lags you have, it actually the impact of previous periods.

Actually, it is sort of coming down because you have a lot of time periods have elapsed and the effect has been reduced. Exactly the same here, in Barlett's test, with the large n we will have a lower standard deviation whereas for a small n I will have the standard deviation high.

(Refer Slide Time: 07:20)



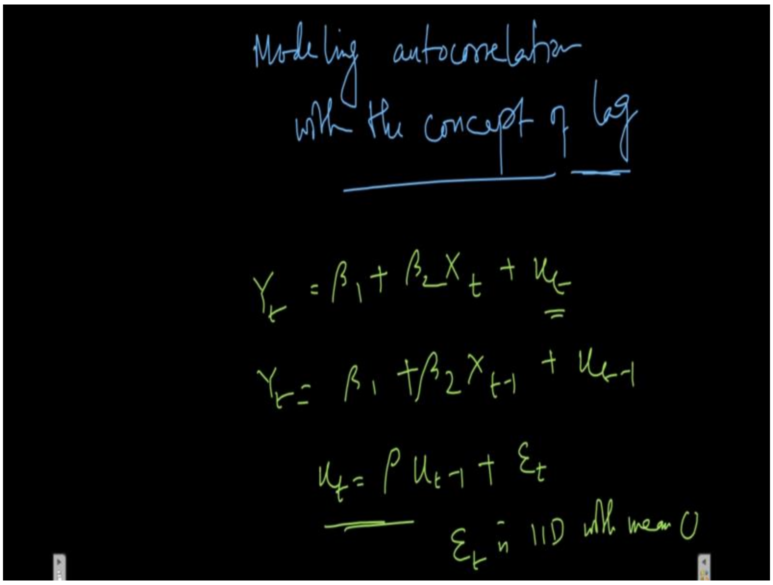
So, with this we can actually do some sort of estimation of whether, my autocorrelation is significant or not. Let us say I have an n of 100 and my 1 by n which is let us say sigma square is 1 by n. So, which is 1 by 100 so, which is 0.01 which means sigma is equal to 1 0.1. So, now, if I want to have a confidence interval if I want to have a you know confidence interval with 95% with the alpha = 0.05. So, then the confidence level is 95%.

So, if I want to construct a confidence interval. So then, how it will look like is that it will look like mu - 1.96 sigma the rho K value I am basically creating this confidence interval for rho K. So, it will be mu + 1.96 sigma. So, it means my sigma is equal to 0.1 and my mu is 0 but let us test follow 0 mean. So, that would mean it would be 1.96 into 0.1 rho K and here also 1.96 into 0.1. So, that would mean minus of 0.196 to + 0.196. So, as long as my rho K is in this range of minus of 0.196.

So, according to our diagram, let us say this will be 0.196 and this is minus of 0.196. As long as, it is in this range what I will have? I will have this threshold and as long as my rock is falling within this I would say there is no significant autocorrelation but if the autocorrelation is higher than this, say perhaps 2 or 0.2 or 2.5. So, then I would say yes autocorrelation is present and this is simply determined by n, n here is 100. So, depending on n, I can actually construct the confidence interval.

Of course, I have to have the confidence level or significance level. So, that is about that is what we understand by Barlett's test. Now, one thing we have to remember and that is very important to understand the autocorrelation process or we say you know, how the data is being generated in when the autocorrelation is actually determined? So, we can actually model that and we can model the autocorrelation using the concept of lag.

(Refer Slide Time: 10:26)



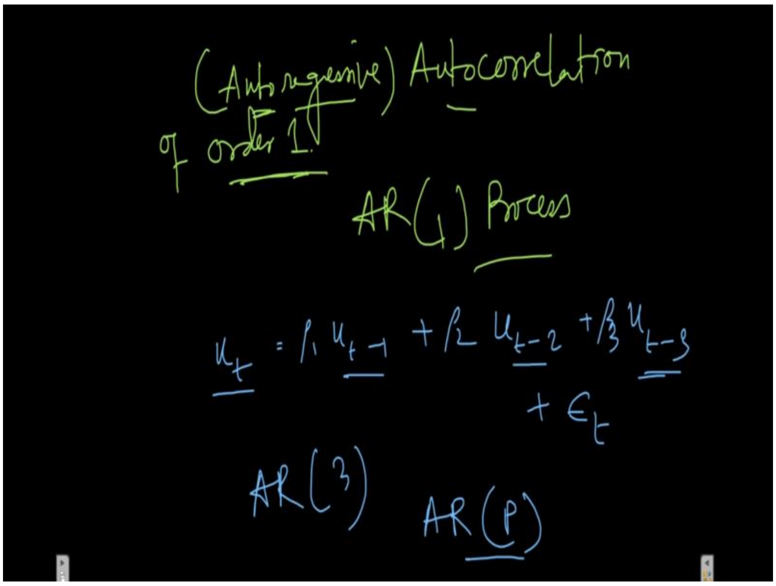
So, modelling autocorrelation with the concept of lag, now, why is that important? Why this lag thing is so important to model autocorrelation I will just talk about that. Now, if you are disturbed, so basically what we are talking about? When you talk about autocorrelation, what do you mean really is that? The error term of a th period is actually related to error term in t - 1 period.

So now, to what extent this relation extends? So, is it extending to only 1 period? So, can I write let us say, if Y t let me write it down Y t is equal to or actually I can use a different color if Y t is equal to beta 1 + beta 2 X t + U t, now I have my error term U t and I said that this error term is actually related. It is somehow related with the previous period error term which was U t - 1.

So, I can also write Y t is equal to beta 1, beta 2 X t - 1 + U t - 1. And I am claiming that U t and U t - 1 are actually related. So, then I can actually write down U t is equal to let us say I use some other coefficient. Let us say rho of course, because this is the correlation coefficient between U t and U t - 1 U t - 1 + some E t now, this E t has to be IID it has to be completely independent of the whole all the variables.

So, E t follows IID E t it is ID with mean 0. Now, with this I can actually explain the autocorrelation. I can actually say this autocorrelation is actually it is dependent on only the past period. Only one past period the just the past few days actually influencing the; present period error term.

(Refer Slide Time: 13:00)



So, I will say this is auto regressive autocorrelation of order 1. So, what I mean by that I will explain that so, the order 1 is understandable. Order 1 is just meaning that it is dependent on just the past period. Now, what is this autoregressive term? Now, regressive the moment we say the; word regressive. So, here what you see is that? We can actually predict U t depending on the past period.

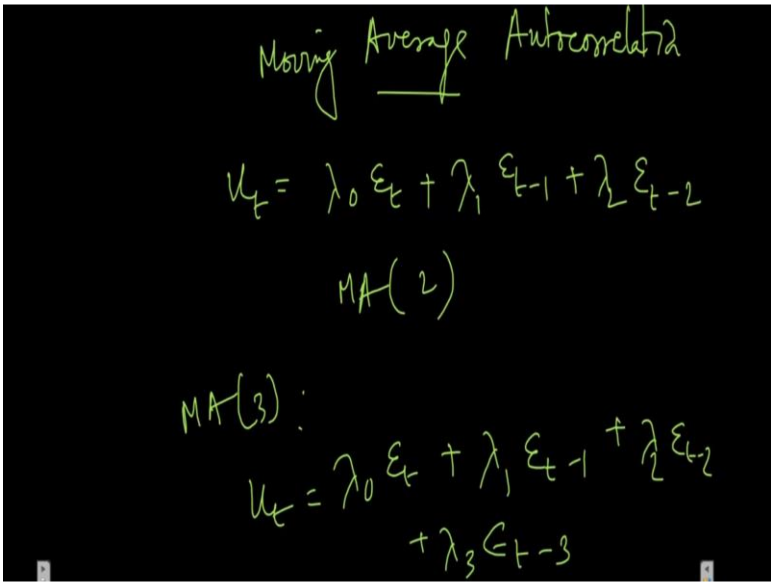
Value of the error term which is U t - 1 and we actually have an innovation term here, a random component here. So, essentially you need to regress to actually get from U t - 1 to U t. So, that is where the regressive term is coming in and why it is auto? Of course, we have we know what auto means you know, self so, it is regressing with oneself.

So, which is again making sense because U t is regressing with itself is but past period value. So, U t is regressing with U t - 1 and that is why we call it autoregressive, autocorrelation of order 1 or in short we write it AR 1 process. This is something we call AR one process. So, this is now if suppose instead of U t being determined just by 1 period. Let us say U t is determined by maybe you know 2 3 periods, 4 periods.

So, how do I write that? So, let us say I L is a different color here. Let us say U t is getting determined by rho 1, U t - 1, rho 2, U t - 2 and maybe parallel rho 3 u t - 3 and then you have an innovation done of course. So, this one you will write as AR 3 process because t is getting determined by 3 period errors. So, this way you can have 5 or n whatever you know order you want so, we usually write AR P.

So, autoregressive autocorrelation of order P. So, this is how we will understand the autocorrelation process. Now, another way of understanding it is that when we actually represent the autocorrelation in terms of the moving average.

(Refer Slide Time: 15:51)



So, let me write it down moving average autocorrelation so, how we represent a moving average autocorrelation? This is you know a simple process where we actually represent the error term in terms of the weighted sum of the past period and the present period error terms. So, we will write it as let me use a different colour or I will use the same colour just do it.

So, let us say it is represented as U t lambda 1, E t - 1 lambda 2, E t - 2. So, essentially it will be A is a moving average we represent as MA. It is MA 2 process whereas, the MA 3 process will be where I will present U t as equal to lambda 0 E t 1 t - 1 lambda 2 E t - 2 + lambda 3 E t - 3. So, this actually is a very brief representation of time series data how we construct a time series data we are not really discussing at length.

This AR process or MA process, there is plenty of discussion of how AR and MA process pans out and how AR and MA together pans out you know so on and so forth. So, we are not getting into that we are just introducing you this concept of AR and MA because we want to actually show you, in the next lecture the consequences of autocorrelation and to explain that consequences we need to use this mathematical formulation. So, with this way we end this lecture here. Thank you.