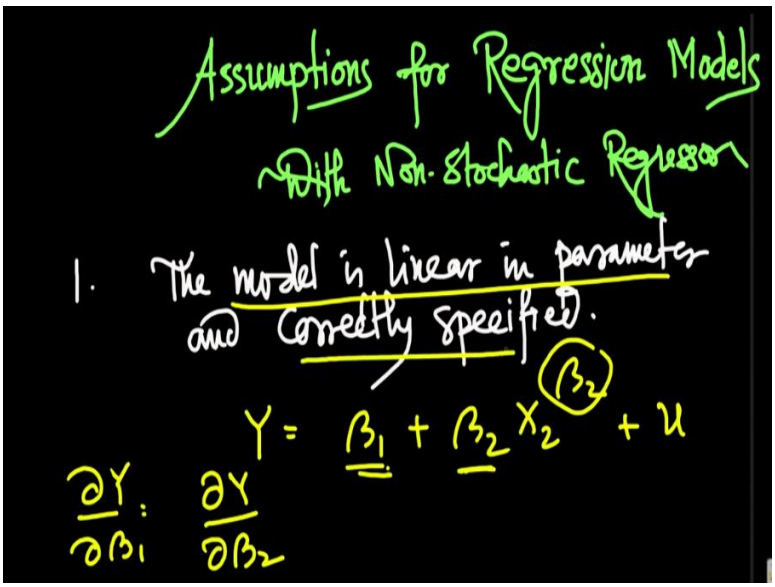
Applied Econometrics
Prof. Tutan Ahmed
Vinod Gupta School of Management
Indian Institute of Technology – Kharagpur

Lecture – 96
Assumptions for Regression Models with Non-Stochastic Regressor

(Refer Slide Time: 00:36)



Hello and welcome back to the lecture on applied econometrics. In this lecture, we are going to talk about the assumptions of regression models with non-stochastic regressor. Now, this is essentially a summary of all the different diagnostics that we have been doing all throughout in the past so many weeks. And we will first summarize these assumptions for a non-stochastic regressor.

So essentially, I told you in the previous lecture that the regressors that we have been using or the assumptions that we are making that our regressor is essentially non stochastic. And first we will talk about all the assumptions that we need to make for regression modelling when we are dealing with a non-stochastic regressor. So, let me write down the first assumption and you are actually familiar with all the assumptions.

Because you actually have done that when you have done with the functional specification when we talked about multicollinearity, when we talked about heteroscedasticity, when we talked about autocorrelation, when we talked about model specification and so forth. So, all

the different learnings that we had so far, it is going to be a summary of that. So, the first assumptions that you can write is that the model is linear in parameter and correctly specified.

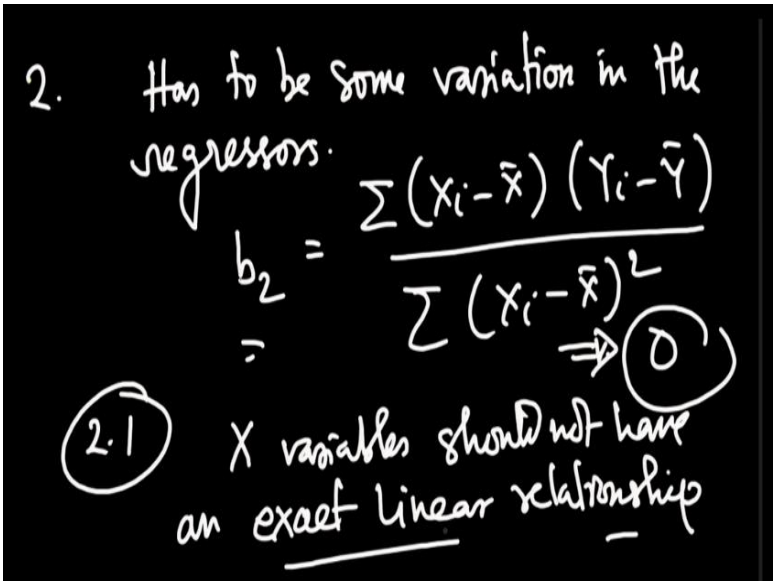
Now, you know about what I mean here, right. So, when I say the model is linear in parameter, so that means that the parameters should not have any power or it should not have any term that if I differentiate that it remains in the equation. So, let me actually again let us recap it. So, let us say Y is equal to we have an equation beta 1 + let us say beta 2 X 2 and let us say there is a power beta 2 and then we have our error term.

Now, in this regression equation now if I cannot have this here, because it is not linear in parameter and this is a problem because when I let us say differentiate, when I actually follow the least square method to estimate the parameter, so what I do basically I do this, I take partial differentiation of Y with respect to beta 1 and beta 2, these are my parameters. Now, if I have a linearity parameter, then what I will have?

In one equation, I will get my beta 1 vanished, in another equation I will get my beta 2 vanish. But that is if I have nonlinearity that is not going to happen. Then I will have even after taking the partial differential I will have this beta 1 and beta 2 present, both the beta 1 and beta 2 are present in the equation and that would be a problem, we cannot estimate the beta 1 and beta 2. So, it has to be linear in parameter. The second is it has to be correctly specified and we already know what it means.

By correctly specified, it means that the functional specific has to be correct. So if it is a log model in the Y variable or the other X variables, we have to ensure that we actually have the log very well, not the just X variable. So, you have to take a log of it. So, if it is a quadratic, we have to take a quadratic. So, whatever be the functional specification, we have to find that and that we have already dealt in detail how to actually ensure that we get the right functional specification.

(Refer Slide Time: 04:00)

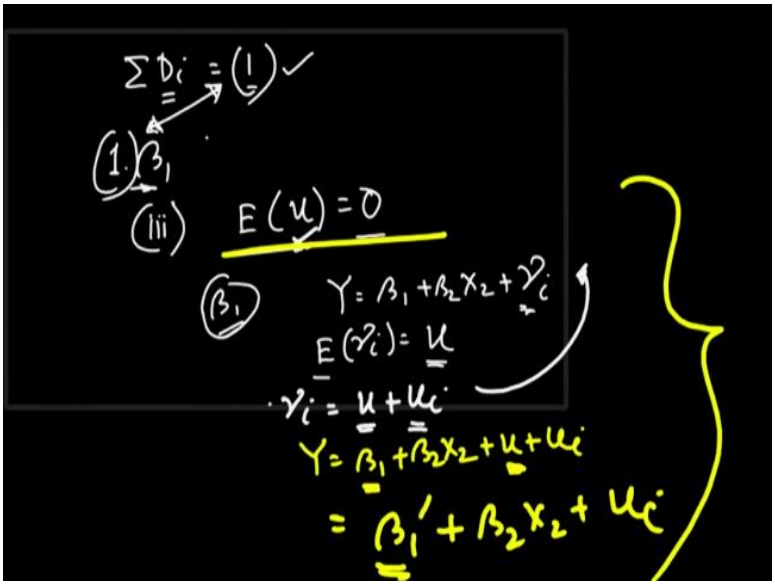


The second point is that there has to be some variation in the regressors. What I mean by that? So, if X is my regressor here and there X cannot have just the same value for all the different observations. Why is that because I know that my beta 2 the regression coefficient is summation X i – X bar and Y i – Y bar and then I have summation X i – X bar whole squared. Now, if all my X's are same, so the denominator is going to be what? Denominator is going to be 0.

The moment I have my denominator 0, my beta 2 is going to be undefined. So, my regression, my software will not actually give me a output because of this thing. So, we have to have some variation in the regression. Now, a follow up of let us say we can write 2.1, so, when we do a regression, we have to ensure that the X variables should not have an exact linear relationship and we have already seen it, we have already talked about in detail, recall dummy variable trap.

So, what we mean by X variables should not have an exact linear relationship? So, there can be some linear relationship, but it cannot have an exact linear relationship. So, at the moment you have exact linear the relationship, so you have a problem called collinearity, which we have talked in detail in multicollinearity.

(Refer Slide Time: 06:07)



So, again to recall what is the problem, the problem is let us say you have a dummy variable. And let us say you include the base dummy and when you sum up all the different values of the dummy variable, so essentially for all observations your D i is going to be 2. So there is no variation in the D i, so all your D i is going to be 1 and you remember that you have intercept and intercept let us say beta 1.

Now beta 1 the intercept and it has you can think that there is a variable multiplied with beta 1 and the variable is going to have a value of 1 that is constant one. Now, because for all the different values you will have this one constant here and you have this D i, subtotal of D i is constant, so essentially your program is unable to distinguish, it will think that the same variable is represented twice. So, it is called a problem of collinearity.

So, the moment you have exact collinearity, exact linear relationship, you will have the product collinearity because your software or basically your regression equation it will look like the same variable is represented twice. So you cannot have a regression with that. So, that is what we call there cannot have an exact linear relationship. So, we may allow some amount of linear relationship and that is what we have seen in the chapter of multicollinearity.

So, this is our second assumption. Now, coming to the third assumption, the third assumption is the expectation of the disturbance term has to be 0. So, the error term, whatever error term we are having, if we take an expected value of the error term, it has to be 0. So, there could be in some observation your u i could be positive, in some observation your u i could be

negative, but the moment you take expectation, you to basically take an average, the average is going to be 0 that is something you have to have.

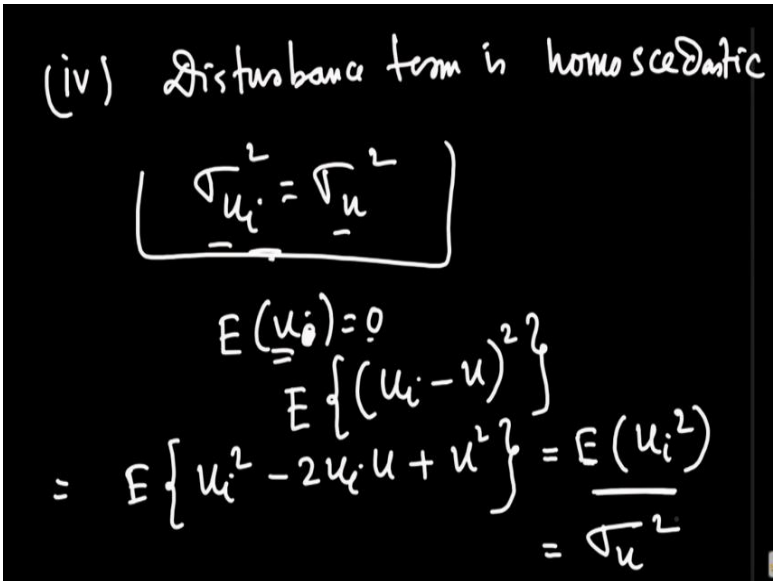
Now, let us assume that your error term has some run in that like some trend, some tendency in that. Usually, the intercept term that we have that takes into account that any tendency or trend in the error term. So, the beta 1 that we have that actually captures that tendency, so essentially we cannot have any run in the error term. So, let us say we have our equation Y = beta 1 + beta 2 X 2 and let us say I have an error gamma i.

And let us say gamma i actually it has some run, it has some trend. Now, I in order to ensure that my expectation of u is 0, so if I take expectation of gamma i, it is going to give me some u let us say, some value that is not zero, so which is not satisfying this condition. If in this particular case since gamma i has some trend, the moment you take expectation, it is not going to give you a value of 0, you are going to get a value of u.

Now, when you have this, essentially you can write that you can actually represent gamma i = some u + u i. So, the trend that you have along with an error term which is random for which; this condition is satisfied. Now, if you replace these in this equation, if you substitute, so then what you will have is, let me use a different colour, y = beta 1 + beta 2 X 2 + u + u i, alright. Now, you see your us is a constant term it can very well go with beta 1. So, essentially what you write is beta 1 + u, you have a beta 1 prime + beta 2 X 2 + u i.

So, essentially the reason I did this little bit of analysis is basically to show you that if you have any run, any trend in the error term, so it will be captured in the intercept term. And we really do not care so much about the intercept term when you do regression. So, it really does not matter if the beta 1 and beta 1 prime are different or whatever. So, essentially as long as this condition is satisfied, we are good. Expectation of error term has to be 0. So that is my third my assumption.

(Refer Slide Time: 10:30)



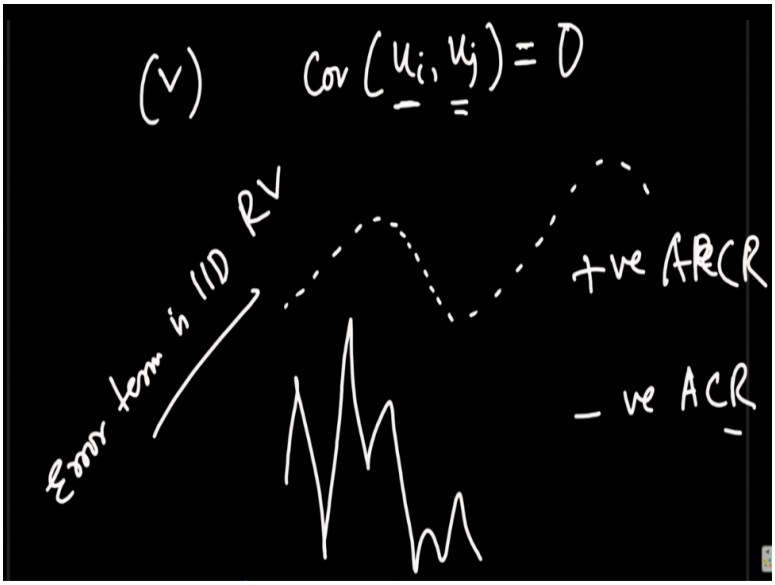
My fourth assumption is the disturbance term is homoscedastic. So, you know already know about what is homoscedastic, what is heteroscedastic. So, let me write it mathematically. So, if I write sigma u i squared = sigma u, so there is a constant variance. So, what means is that each observation is drawn from a constant population variance. So, how is that possible? How is your variance for each draw is going to be constant?

So, ideally it should be somewhere it is going to be a little more, somewhere it is going to be a little less and in very rare occasion you can have something like this, but you remember the concept of double structure variable. So, it is a case essentially of a double structure variable. So, essentially it is when you write this equation, we are writing it before actual realization. So, it is basically before you actually it is basically an expectation of the variance.

So, I can write it like so from the previous condition I know expectation of an error term, expectation of u i = 0 and I know I can represent variance as expectation of u i – u square and if I write it so expectation of u i square – 2 u i u + u square. Now, this is going to give us this expectation of u this being 0, I will actually have only, this will be this because of the previous condition I already have.

Now, I should actually write expectation of u, not the i. So with this, we can actually say that the variance of; it is a case of constant variance where this is going to be sigma u square. So, this expectation of the variance is going to be constant, alright.

(Refer Slide Time: 13:03)



Now, the next condition that we have already seen is a condition which we addressed in the case where we talked about autocorrelation and that is there should not be any relationship between the error terms for different observations. So, covariance of u i and u j has to be 0. So, there should not be any relationship among these error terms.

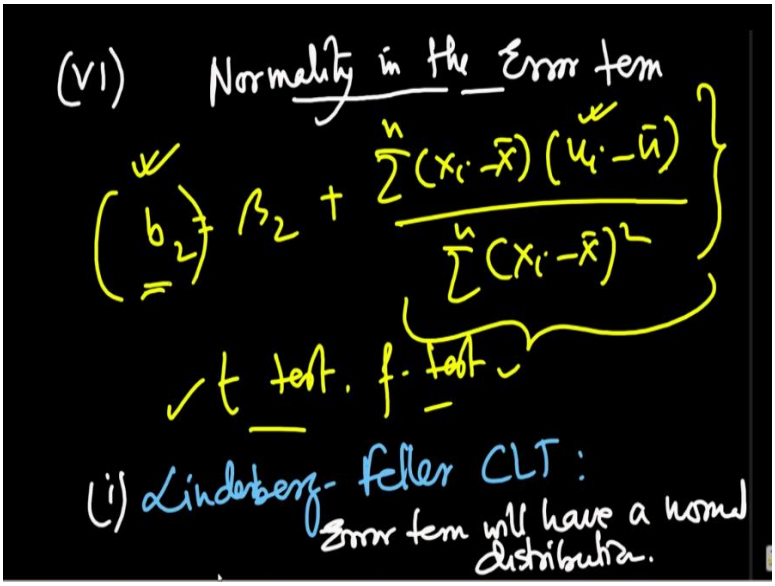
Now, sometimes it may happen when you look at your data or look at the error terms after you fit your models, you can see that the error terms might show some positive tendency or it can actually show positive, negative and positive and negative and so forth and you already know this is a case of positive autocorrelation or it can show some pattern like this, some zigzag patterns, and you know this is a case of negative autocorrelation, ACR negative autocorrelation.

So, usually, we do not want this kind of situation to exist and that is why you have to have and in all these cases, your covariance u i, u j is not equal to 0, but you want to ensure that covariance of u i and u j is equal to 0 because you do not want to have any pattern in your error term. So now in the previous assumption and in this assumption, essentially what we are talking about?

We are talking about if you can actually take care of this problem, you can take care of the problem omitted variable bias also, you can take care of the problem if there is any model specification or if there is the measurement error. So, by taking these problems into account, you actually take care of these problems. And when you have the assumption 4 and 5 right, so you can actually ensure that your error term is IID random variable.

So we know the importance of IID random variable. Our error term is IID random variable if I ensure homoscedasticity and no autocorrelation.

(Refer Slide Time: 15:03)



Now I will come to the sixth assumption. And the assumption 6 says is that there has to be normality in the error term. So, what I mean by normality in the error term and why I need my error term to be normal? Now, my normality of the error term it actually is important. First let me answer the why part. You remember the equation we have written when we wrote b 2 = beta 2 + summation X i – X bar into u i – u bar by sum total of X i – X bar whole square over all n.

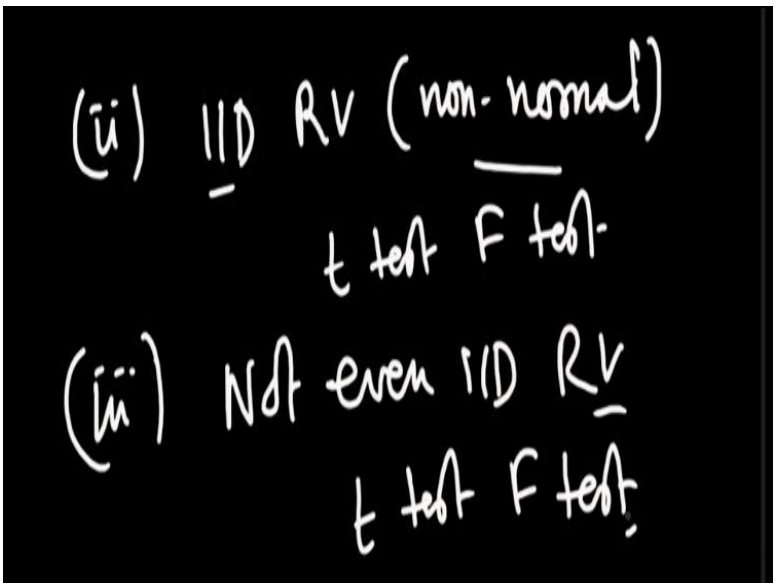
Now, you see that the moment you have this component this because of this error term here your b 2 is getting a distribution. At the moment your u has a normal distribution; your b 2 is also going to get a normal distribution. Now, in fact that is actually good for us and why it is good for us? Because the moment I have my b 2 it is having a normal distribution I can apply things like t test or f test which are essential.

Which are crucial to basically get your regression equation right, I mean you have to estimate you have to do the significance test, you have to do the confidence interval, you have to get the confidence interval and you can get that only if you can run this t test and f test. So, that is why the normality of error term is really functional because you can get a distribution for your b 2.

Now, even if usually the error term is normal, the reason is, actually we need to go back to our central limit theorem and it is a particular theorem by Linderberg, so Linderberg-Feller's central limit theorem. So, what it says is that if let us say there are a lot of influences on a particular variable. So, on my error term, there are a lot of variables of which I am not aware of, they are influencing; they are not present in the model, they are influenced by outside.

Now, as long as these outside influences are not very strong, my variable of interest is going to have a normal distribution. So, that means my error term will have a normal distribution, so usually that is why we actually get my error term to have a normal distribution. Even if let us say there are cases where your error term does not have a normal distribution, let us say this is a case 1.

(Refer Slide Time: 18:05)



Even if your error term does not have a normal distribution, if it is an IID random variable, which is non-normal, even then you will be able to perform your t test and F test. So, you will find t test and F test. And another case could be it is not even IID random variable, but if you have a very large number of observations, you can actually perform your t test and F test. So, essentially because your error term is either normal or it could be IID non-normal, so in most of the cases this is not a very big problem.

So, these are the assumptions that you need to ensure when you have your regression model with a non-stochastic regressor. And as you have seen, what I have done essentially just summarized all that I have been doing for the past few classes. So, with this I will end the

lecture here. And in the next lecture, I am going to talk about the same assumptions but for stochastic regressor. Thank you.