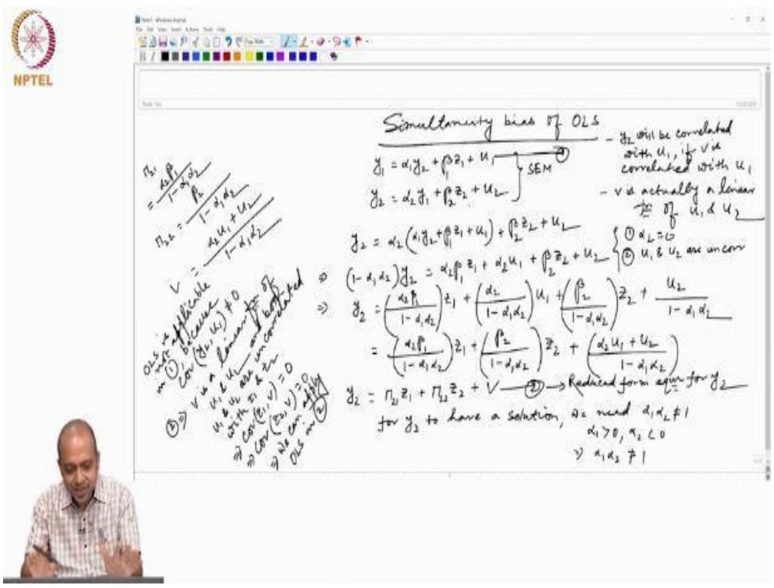
Applied Econometrics
Prof. Sabuj Kumar Mandal
Department of Humanities and Social Sciences
Indian Institute of Technology-Madras

Lecture-14
Simultaneous Equation Model-Part III

So, welcome to our discussion on simultaneous equation model once again. Yesterday we discussed about what exactly is a simultaneous education model and why do we need to understand simultaneous equation model particularly to solve the endogeneity bias due to simultaneous nature of the two variables. So, that is basically simultaneity bias. So, this is the second reason of endogeneity. So, we call it simultaneity bias.

So, that means in an equation if one of the explanatory variables is jointly determined with the dependent variable, then we will say that that the model is suffering from simultaneity bias and OLS cannot be applied because OLS will give biased and inefficient estimates. Now, today what we will do, we will take a simple example to understand mathematically the nature of that simultaneity bias.

(Refer Slide Time: 01:22)



So, we will take a simple example. What we are going to discuss is simultaneity bias of OLS and our model **is** $y_1 = \alpha_1 y_2 + \beta_1 z_1 + u_1$ and $y_2 = \alpha_2 y_1 + \beta_2 z_2 + u_2$**.** So, this is basically a simultaneous equation model, because the explanatory variable y_2 is jointly determined with the dependent variable y_1.

So, these two equations are actually constituting a simultaneous equation model and they satisfy the two conditions to be qualified as SEM. Number 1**:** y_1 and y_2 are jointly determined and both the equations have their own standalone or independent meaning. So, you can think of the first equation as the labour supply function, second equation as the labour demand function. So, that is the reason both the equations are stand alone or they have independent meaning.

Since, these two conditions are satisfied we call this is a simultaneous equation model. Now you will try to mathematically derive certain conditions. So, as to prove that first of all how $\mathbf{y_2}$ is actually correlated with $\mathbf{y_1}$ and then we will say that in that condition how OLS is actually not applicable and then we will discuss what is a remedial measure meaning what type of estimation procedure we need to follow in this particular context of simultaneous equation model.

So, from these you can write by substituting the value of y_1 from equation 1, $y_2 = \alpha_2(\alpha_1 y_2 + \beta_1 z_1 + u_1) + \beta_2 z_2 + u_2$**.** So, this is the case. So, that means I can write $(1-\alpha_1\alpha_2)y_2 = \alpha_2\beta_1 z_1 + \alpha_2 u_1 + \beta_2 z_2 + u_2$. Now, if dividing both sides of this equation by $(1-\alpha_1\alpha_2)$, you will get, $y_2 = \left(\frac{\alpha_2\beta_1}{1-\alpha_1\alpha_2}\right) z_1 + \left(\frac{\alpha_2}{1-\alpha_1\alpha_2}\right) u_1 + \left(\frac{\beta_2}{1-\alpha_1\alpha_2}\right) z_2 + \frac{u_2}{1-\alpha_1\alpha_2}$.

So, since u_1 and u_2 both have in the denominator $1-\alpha_1\alpha_2$, this again can be simplified as

$y_2 = \left(\frac{\alpha_2\beta_1}{1-\alpha_1\alpha_2}\right) z_1 + \left(\frac{\beta_2}{1-\alpha_1\alpha_2}\right) z_2 + \frac{\alpha_2 u_1 + u_2}{1-\alpha_1\alpha_2}$.

This I can write **:** $y_2 = \pi_{21} z_1 + \pi_{22} z_2 + V$

$$\pi_{21} = \frac{\alpha_{2\beta_1}}{1-\alpha_1\alpha_2}$$

$$\pi_{22} = \frac{\beta_2}{1-\alpha_1\alpha_2}$$

$$V = \frac{\alpha_2 u_1 + u_2}{1-\alpha_1\alpha_2}.$$

So, that means for y_2 to have a solution, the term $\alpha_1\alpha_2$ should be not equal to 1. Because, here if you look, if $\alpha_1\alpha_2$=1 then that becomes 0**.** Now, we need to think whether, this $\alpha_1\alpha_2$not being equal to 1 is a very very restrictive assumption. Actually it is not. Because if you go back to this original two equations of our simultaneous equation model and if you think that one is the supply function and another one is the demand function.

Then actually, if you go back and think about your labour supply function, then that is basically the y_2 is the wage rate. So, this is greater than 0 and alpha 2 is actually less than 0**.** So, that means $\alpha_1 and\ \alpha_2$ is not equal. One is negative and another one is positive. So, that means it is not a very restrictive assumption as far as we consider one of the two equations in the SEM is actually a demand function and another one is the supply function.

Now, let us assume that this is our equation 1 and let us say that this is our equation 2. Now, we said that in equation 1 OLS is not applicable because of correlation or covariance between y_2 and u_1 is actually not equal to 0. But, what is the proof? We assume that this is an endogenous variable, because of this reverse causality or y_2 and y_1 they are actually simultaneously determined.

But, we have not yet given any proof and that was the idea we are deriving this mathematical model just to show you how this y_2 is actually correlated with u_1. From equation 2, we can understand that v is actually a linear function of u 1 and u 2, because $V = \frac{\alpha_2 u_1 + u_2}{1-\alpha_1\alpha_2}$

So, v is the linear function of u_1 and u_2. And both u_1 and u_2 are uncorrelated with z_1 and z_2. So, v is a linear function of u_1 and u_2 and then u_1 and u_2 are uncorrelated with z_1 and z_2. That means, we can say that covariance between z_1 and V is actually 0 and also covariance between z_2 and V is also equals to 0.

So, that means in equation 2, the error term is uncorrelated with both z_1 and z_2 and that is the reason we can say that OLS is applicable in equation 2 but not in equation 1. So, that implies we can apply OLS in 2. So, in equation 1 we could not apply OLS because y_2 was simultaneously determined with y_1 and that is the reason we say that it is correlated with u_1.

We have not yet given any proof. We said only this much that y_2 and y_1 are simultaneously determined and that is the reason that will produce some kind of simultaneity bias, because y_2 will eventually get correlated with u_1, because simultaneity is the second reason of endogeneity. So, OLS is not applicable in equation 1. That is the reason we tried to derive an equation for y_2 the endogenous variable in the form of equation 2.

And then we proved that OLS is applicable to equation 2 because the error term in equation 2 is neither correlated with z_1 nor z_2. Now, if you think we are doing something new here in equation 2, it is actually not. We have actually experienced this type of equation in our first module itself where we were discussing about structural equation and reduced from equation.

If you think for a minute you will easily understand why this equation 2 is actually reduced from equation for y_2, because here y_2 is expressed only in terms of the 2 exogenous variables z_1 and z_2. So, this equation 2 is actually a reduced form equation for y_2. So, this is not new. So, in reduced form equation obviously our idea is that we can apply OLS here.

Then, we will take y_2 here and then we will plug in equation 1 and then we will estimate. So, that means the same two stages is two OLS procedure what we learned in first module. The same thing we are going to apply here. Now, what we have not proved so far is how y_2 is actually correlated with u_1. From this equation we can see that y_2 is a function of v. Now, y_2 will be correlated with u_1 if v is correlated with u_1, because v is the error term in this y_2 equation.

So, there is only one channel by which y_2 can be correlated with u_1 and that is through V. We have proved that V is actually a linear function of u_1 and u_2 where $V = \frac{\alpha_2 u_1 + u_2}{1 - \alpha_1 \alpha_2}$. If that is the case, V is a linear function of u_1, V is appearing here. So, obviously y_2 is correlated with u_1 y_2 is actually correlated with u_1 that is how we can prove.

Now, there are two cases where y_2 is actually not correlated with u_1. What are the two cases? If $\alpha_2 = 0$ is the first case and second one is that u_1 and u_2 are uncorrelated. These are the two conditions. If they are satisfied, then only we can say that y_2 is actually not correlated with u_1, otherwise y_2 is always correlated with u_1 as v is the error term appearing in y_2 equation and v is a linear function of u_1 and u_2. So, as long as u_1 and u_2 are correlated and α_2 is actually not equal to 0, then y_2 is correlated with u_1. Now, when α_2 is 0 from the y_2 equation, then this is also 0. So, that means when $\alpha_2 = 0$, then only we can say that from this equation y_2 equals to if we put 0 this is become 0, this is also become 0, so that means y_2 is not correlated to u_1.

But, if we assume that $\alpha_2 = 0$, that means y_1 is not appearing in y_2 equation. And if y_1 is not appearing in y_2 equation that means y_1 and y_2 are not simultaneously determined. So, if they are not simultaneously determined, then obviously there cannot be any kind of endogeneity.

You can easily estimate equation 1 by OLS. We assume that y_2 and y_1 are simultaneously determined. So, by assuming y_2 you are breaking the channel. You are saying that there is no simultaneity between y_1 and y_2 and that is the reason this is we can apply OLS there. What is the meaning of u_1 and u_2 not correlated?

So, when I am saying they are not correlated that means I am saying that y_2 is actually not correlated with u_1 due to any omitted variable or measurement error. That is the reason I am saying that when I am saying u_1 and u_2 are not correlated, because if at all y_2 has to correlate it with u_1, u_1 and u_2 must be correlated. When I am saying u_1 and u_2 are not correlated, there is no simultaneity also. That means, I am ruling out the possibility of two other channels of simultaneity. What are those? One is omitted variable bias and second one is measurement error. That is very clear. So, this $\alpha_2 = 0$ and α_1 and u_1 and u_2 are not are uncorrelated, they have specific meaning we need to understand that. I will repeat once again, when $\alpha_2 = 0$, that means y_1 is not appearing in y_2 equation.

And that means we are saying that these two variables are not simultaneously determined. Then you can obviously apply OLS to estimate equation 1. When I am saying u_1 and u_2 are uncorrelated, that means, I am actually ruling out the possibility of other two sources of endogeneity which are measurement error and omitted variable bias. So, this simple model can actually explain what exactly is the problem of simultaneity bias and how OLS cannot be applied in that simultaneous equation model.