Applied Econometrics

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

Lecture-12
Simultaneous Equation Model-Part-I

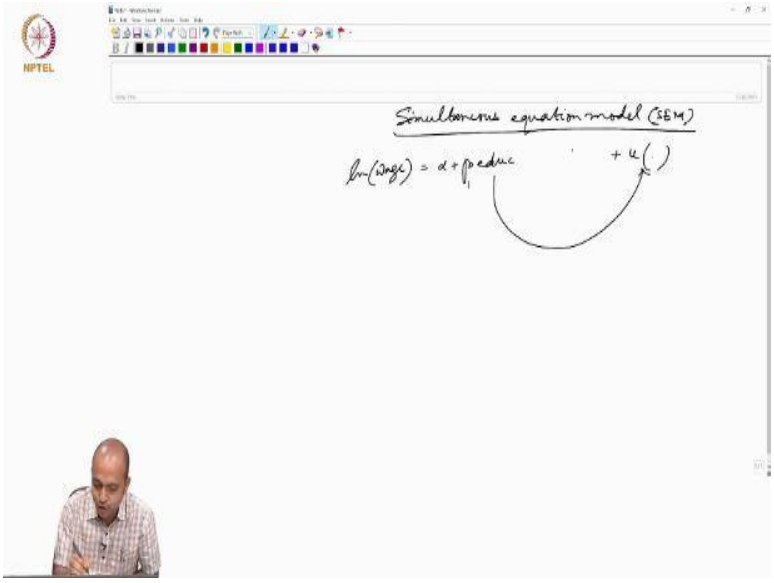
Welcome once again to our discussion on applied econometrics. Today we are going to discuss about Simultaneous Equation Model.

(Refer Slide Time: 00:34)

In short, I will be using the term SEM for Simultaneous Equation Model. Now, this simultaneous equation model is important to understand, because if you look in our previous section, we were talking about instrumental variable estimation technique to solve the endogeneity problem. And we said that there are three reasons because of which endogeneity may arise in a model.

The first reason for endogeneity was omitted variable. So, that means if you again think back while discussing about instrumental variable estimation technique, we said that let us think about our wage function, log of wage which is basically a function of $\alpha + \beta_1 education + \beta_2 ability + u$. Now, the variable ability, as said, is difficult to observe and difficult to measure as a result of which we generally do not include ability in the model and that ability goes into this error term.

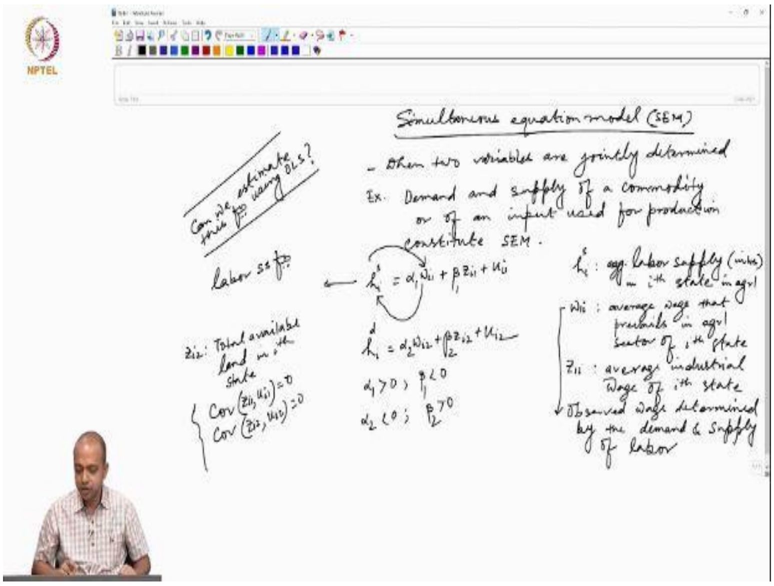
(Refer Slide Time: 02:22)



So, our model will only be $\beta_1 education + u$. And this u, now captures this ability. And if we assume that there is some kind of correlation between education and ability then this education variable gets correlated with the error term. So, that is the source of endogeneity we are talking about in our previous section. So, this is endogeneity due to omitted variable**.**

Now, the second reason for endogeneity is basically the simultaneous equation system or simultaneous relationship**.** So, that is the reason we are going to talk about simultaneous equation model. Otherwise, if the model demands SEM and we estimate a single equation method, then that model will suffer from simultaneity bias which will lead to the endogeneity problem. So, this is basically the second reason of endogeneity**.** Simultaneity is called simultaneous relationship meaning two way causal relationship. This is the second reason for endogeneity that is why we need to discuss in detail about this simultaneous equation model.

(Refer Slide Time: 04:12)



Now, the definition of Simultaneous Equation Model is as follows. When two variables are jointly determined then we say that it constitutes a simultaneous equation model. As an example, we can say that demand and supply of a commodity of an input used for production constitute SEM**.** Demand and supply example, if you can remember then it would be easy for us to understand simultaneous equation model. So, let us now talk about the demand and supply of an input, which is used for production and we will take labour demand and labour supply.

So, this is his which is basically the aggregate labour supply, let us say, in i^{th} state of a country. Let us say we are talking about in the Indian context and this is the i^{th} states' labour supply which is a function of let us say $\alpha_1 w_{i1} + \beta_1 z_{i1} + u_{i1}$. So, here h_i^s is the aggregate labour supply in hours, actually labour supply in i^{th} state.

Then w_{i1} is theaverage aggregate labour supply in i^{th} state, let us say in agriculture. So, we are talking about agricultural labour supply and agricultural labour demand, that is, labour supplied and demanded specifically in the agricultural sector. So, this is the average wages that prevails in the agricultural sector of i^{th} state. z_{i1} is the average industrial wage of i^{th} state and u_i is basically the error term.

This is the labour supply function who are supplying the labour, this is basically the farmers. The wage labourers, they are supplying labours. So, this is the labour supply function. So, the question that we are asking is whether we can estimate this labour supply function using OLS? Now, this equation is a very peculiar in nature and we need to clearly understand how and in which way this equation is different from the equations which we have discussed earlier.

When we applied OLS we assume the left hand side variable is purely endogenous, which is the labour supply and in the right hand side all variables are exogenous in nature**.** Now, this w_{i1}, which is the average**,** agricultural wage (assumed), is also an exogenous variable. What does it mean? It means, that we have a capacity to fix agricultural wage at different levels and then we can observe how the labourers are supplying the labour given that wage is fixed exogenously.

So, that means we need to conduct some kind of experiment by fixing w_{i1} and z_{i1} at different level, industry wage agricultural wage should be fixed at different level. And then we have to observe how labourers are supplying their labour, when the w_{i1} and z_{i1} are fixed at different level, but in reality that type of experiment is very hard to conduct. What we observe as w_{i1} and z_{i1} is actually the observed wage rate. So, w_{i1} is actually the observed wage rate determined by the demand and supply of labour.

So, what do we observe is actually the equilibrium wage rate, which is determined by the intersection of demand and supply assuming that labour market always clears. So, that means, there is no unemployment in the labour market which is the assumption**.** So, that means, in this equation w_{i1} is actually not an exogenous variable rather, this w_{i1} and this h_i^s -wage rate and labour supply are jointly determined. That is why they are forming actually simultaneous equation model.

So, causality runs from here to here and also from here to here**.** Average wage rate that prevails in the agricultural sector of the i^{th} state will determine the aggregate labour supply. At the same time depending on what is the aggregate supply, the wage rate will also be determined. That means, w_{i1} and h_i^s-this are jointly determined in a simultaneous equation model.

By specifying only the supply function, we cannot actually estimate this type of equation, because of this simultaneity bias. So, that is the reason I said this equation is quite different from the earlier equations, where you are using OLS. So, OLS is applicable only when causality is unidirectional - from independent variable to the dependent one.

Here is a case where we are getting bi-directional causality. That means, w_{i1} causes h_i^s - aggregate labour supply and aggregate labour supply also causes w_{i1}. That is why these two variables are jointly determined. That is why to estimate this equation we need to specify labour demand function as well. So, that means labour demand function- let us say, $h_i^d = \alpha_2 w_{i2} + \beta_2 z_{i2} + u_{i2}$.

w_{i2} is again the wage rate that prevails in the agricultural sector and is z_{i2} is basically the total available land in i^{th} state. Now, we will assume that covariance between z_{i1} and u_{i1} is 0, meaning z_{i1} is strictly exogenous in the first equation and covariance between z_{i2} and u_{i2} is also equal to 0. That means this labour demand, which is also a function of agricultural wage and also labour demand is a function of the total availability of agricultural land obviously.

If the available land is more, then there will be more demand for labour. Now, why we have included z_{i1} in the first equation? z_{i1} now if you talk about the signs of this so that means, α_1 is assumed to be positive because as wage rate increases obviously the labourers will supply more labour, that is why α_1 is assumed to be positive. Of course there are cases where α_1 can be negative as well but we are not discussing the unusual pattern of labour supply function.

The standard labour supply function comes from the optimization of the labourers that give α_1 to be positive. β_1 is actually negative because, obviously as the average industrial wage goes up, then labourers will move from agriculture to industry and obviously that is the reason labour supply in the agricultural sector comes down.

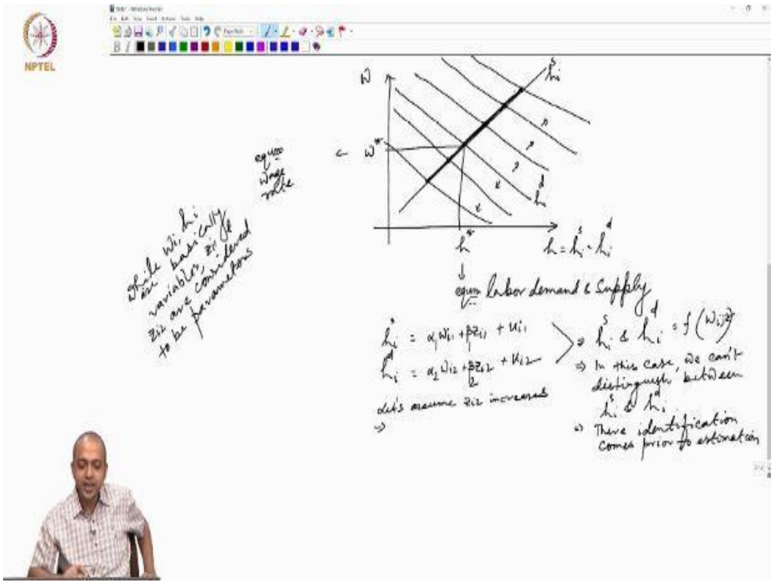
Why this is so? Alternatively, you can think that z_{i1} is basically the opportunity cost of supplying labour in agricultural sector. What is my next best alternative? Next best alternative is industrial sector. So, when I am supplying labour in the agricultural sector as a labourer that means I am sacrificing my opportunity to work in the industry.

So, that means working in agriculture has a cost in terms of the wage that I sacrifice from the industrial sector and that is the reason this β_1 is actually negative. If I have more industry wage then I will supply less labour in agricultural sector. That is the reason we have included z_{i1} in the labour supply function, without that the labour supply function will be misspecified. There are many other factors apart from these two and all those factors are captured by u_{i1}.

For simplicity, we have kept only two variables- agricultural wage and industry wage. While, agricultural wage directly affects labour supply, industry wage is included just to represent some kind of opportunity cost of labour supply in the agricultural sector. Similarly, in labour demand function what is the sign of α_2? Since it is labour demand, as wage rate increases obviously the farmers will demand less labour.

And what about β_2, since labour and land are complementary in nature neither labour alone nor land alone cannot produce anything. As a result of which, we assume that the sign of β_2 as land increases will also become positive and there would be more demand of labour from the farmers.

(Refer Slide Time: 21:43)



Now, if we represent these two: Let us say in the X axis we are measuring h, which is actually equal to $h_i^s = h_i^d$ **.** So basically we are measuring labour demand or labour supply whatever and here we are measuring h. So, this is our labour demand function which is downward sloping(h_i^d) and this is h_i^s and what we observe in the labour market is basically this**.**

So, this is aggregate labour supply and this is the equilibrium. So, this is the equilibrium labour demand and supply and this is the equilibrium. Now, in our previous equation we said that labour supply, $h_i^s = \alpha_1 w_{i1} + \beta_1 z_{i1} + u_{i1}$ and $h_i^d = \alpha_2 w_{i2} + \beta_2 z_{i2} + u_{i2}$

If we remove these two terms- z_{i1} and z_{i2} then we cannot actually distinguish we cannot identify the equations whether, we are estimating a demand function or a supply function. This is because both labour supply and labour demand is a function of w_i. So, that means, both labour supply and labour demand are functions of only one variable, which is w. So, in i^{th} state, aggregate labour supply and aggregate labour demand both are function of the average wage. Please try to understand the importance of z_i, i.e, the importance of the exogenous variable in the model. If we remove the exogenous variable, we keep only the wage rate in both the equations, then both labour demand and labour supply will become a function of h.

So, in this case we cannot distinguish between h_i^d and h_i^s. So, if we estimate what will happen? We will estimate the function, but we do not know whether we have estimated a demand function or a supply function. That is the reason why identification comes prior to estimation.

First, we have to identify and then we need to estimate. And if we need to identify, those exogenous variables will play an important role. Let us now introduce the z_i terms in equation 1 and 2. So, this is $\beta_1 z_{i1} and\ this\ is\ \beta_2 z_{i2}$. Let us now understand the identification problem conceptually and how these variables help identify these equations. Let us assume, that there is some kind of change in z_{i2}, meaning there is some kind of change in total labour with total availability of land.

So, in this equation what happens in the labour demand function- we assume that w_{i1} is basically a variable that means, if there is a change in w_{i1} there will be a change in the labour demand along the curve. And we assume z_i to be exogenous variable that means, we consider that as a parameter. So, while w_i is basically a variable h_i indicate labour demand and labour supply. They are basically variables. z_{i1} and z_{i2} considered to be parameters. If you go back to your principles of economics, we learned about shift along the curve and parametric shift. So, that means parameters are assumed to be constant in a two dimensional plane.

So, this is a two dimensional plane, in the y axis we are measuring w in the x axis we are measuring h. So, other factors affect this demand and supply. This w and h are considered to be fixed meaning they are parameters. If there is any change in the parameter then the entire curve will shift either left side or right side that is called parametric shift, because look at this h_i^d and h_i^s they are function of w_i only.

We assume w_i and then we will say z. So, that means they are considered to be parameters. So, if z_{i2} change, that will affect the h_i^d, because z_{i2} is not appearing in the supply function. So, labour demand will shift. Let us assume that z_{i2} increases, that is total availability of land, then h_i^d which is the labour demand will shift right toward like this.

If there is a decrease then that will shift will be leftward**.** And every time this labour demand function changes either upward or downward, it will intersect with the labour supply function because labour supply function is not changing. So, if we collect all these intersection points and connect that then we will identify labour supply function. This is going to be our labour supply function.

This is very conceptual please try to understand once again. What I am saying that let us assume that z_{i2} is changing. First we assumed that if z_{i2} is increasing, since z_{i2} is assumed to be a parameter in this equation, the labour demand function will shift upward, which is called a parametric shift - total availability of land in i^{th} sector is increasing.

Obviously, labour demand in i^{th} sector in aggregate will increase. So, this is a right ward shift. I have indicated h_i^d will shift upward. And every time it shifts that will give a new intersection with the labour supply, because labour supply is fixed at it is original point, because z_{i2} is not appearing in the supply function. So, if we collect all this equilibrium points and then connect them, that means, we are able to trace out the labour supply function. This is how we can identify the labour supply function, when z_{i2} is changing.