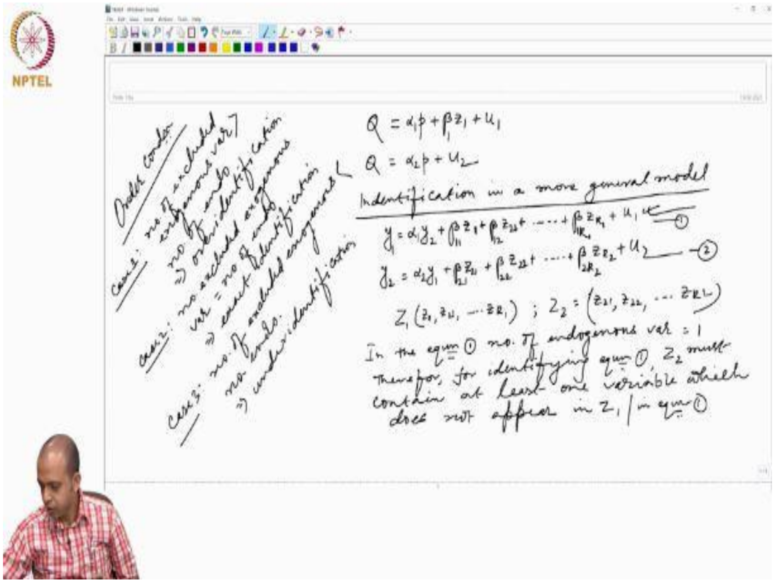
**Applied Econometrics**
**Prof. Sabuj Kumar Mandal**
**Department of Humanities and Social Sciences**
**Indian Institute of Technology-Madras**

**Lecture-16**
**Simultaneous Equation Model-Part V**

We were discussing with the demand and supply function of milk**.** And we said that we identify whether it is a demand function for milk or supply of milk by the way of the excluded variable**.**

**(Refer Slide Time: 00:53)**




So, the model what we are discussing yesterday was basically a simple one. The supply of milk equation is $\alpha_1 p + \beta_1 z_1 + u_1$and the demand equation is $\alpha_2 p + u_2$. The first one is the supply function and second one is the demand function because $z_1$ was basically the average price of cattle feed that we are discussing.

Since, average price of cattle feed has nothing to do with the demand for milk consumption, we said that the second equation was the demand function and the first equation is the supply because the average price of cattle feed actually affects the supply. This is so because price of cattle feed will decide how profitable milk production is**.** If the price of cattle feed increases, then cost of milk production will increase and milk production would become less profitable.

So, this is how there is some kind of relationship between milk supply and average price of cattle feed while, average price of cattle feed has nothing to do with milk consumption. While deciding how much milk I will consume, I generally do not care about the average price of the cattle feed. Now, this is a simple model and we have only one exogenous variable in the first equation and in the second equation we have only one endogenous explanatory variable.

We can always extend this model and we can discuss a more generalized case. So, I will write the identification in a more general model. So, let us say that this $y_1 = \alpha_1 y_2 + \beta_{11} z_{11} + \beta_{12} z_{12} + \cdots + \beta_{1k} z_{1k} + u_1$ and $y_2 = \alpha_2 y_1 + \beta_{21} z_1 + \beta_{22} z_2 + \cdots + \beta_{2k} z_{2k} + u_2$. So, that means from the first equation, I can say that $\mathbf{z_1}$ is actually a vector of explanatory variables, which includes $z_1$, $z_2$ or $z_{11}$, $z_{12}$, $z_{1k}$. And in the second equation $z_2$ also is a vector of explanatory variables, which is let us say $z_{21}$, $_{22}$….$z_{2k}$. So, $z_1$ and $z_2$ are basically the set of exogenous explanatory variables.

Now, regarding the identification of the first equation, some of the variable, which are appearing in $z_2$ are not actually appearing in $z_1$. So, that means we can say that z1 and $z_2$ consist of different set of variables, even though some kind of overlapping might be there. So, at least some difference should be there in the elements of $z_1$ and $z_2$. So that these two equations are identified.

For example, when we are trying to identify the first equation, then some of the exogenous variables should not appear here in the first equation, but must appear in the second equation with some non-zero coefficient. So, that means $z_2$ must contain some variables which are not appearing in the first equation. In the first equation there is only one endogenous variable, which is $y_2$.

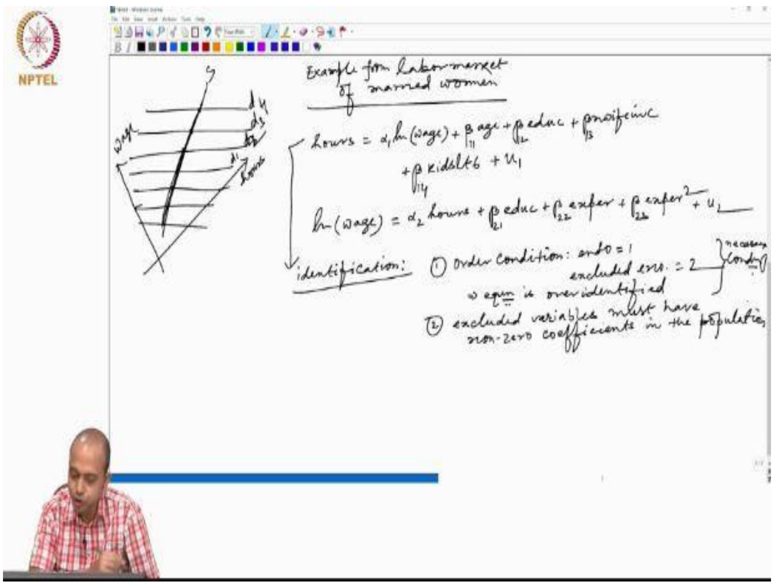
So, that means, from the first equation at least one exogenous variable must be excluded. In the equation 1, number of endogenous variable equals to 1. Therefore, for identifying equation 1, $z_2$ must contain at least one variable, which does not appear in $z_1$ or in equation 1 because equation 1 is constructed from by taking elements from $z_1$ only.

Similarly, for identifying equation 2, $z_1$ must contain at least one variable which is not appearing here. So, when the number of excluded exogenous variables is equal to the number of endogenous variables, we will say that the equation is exactly identified. If the number of excluded exogenous variables is greater than the number of endogenous variables we will say that the equation is over identified.

And if the number of excluded exogenous variables is less than the number of endogenous variables, we will say that the equation is under identified. These are the conditions of order condition. Order condition for identification has several cases. Case 1, number of excluded exogenous variables is actually greater than number of endogenous variables. Then this is called over identification. In case 2, number of excluded exogenous variable is actually equal to the number of endogenous variables, which is called exact identification and in case 3, we will say that number of excluded exogenous variables is actually less than the number of endogenous variables which is called under identification**.**

So, there should be at least as many excluded number of exogenous variables as the number of included endogenous variables**.** As many excluded variables as endogenous variables in that particular equation. That is called as identification. We will now take an example to understand this identification with more clarity and then we will try to use one data set to estimate that equation.

**(Refer Slide Time: 14:45)**




So, we will take an example from labour market of married women that means the woman who participates in the labour market even after marriage. We are taking that example. So, labour supply which is let us say measured by hours equal to $\alpha_1 logwage + \beta_{11} age + \beta_{12} education + \beta_{13} nonwifeincome$ **.**

Here, the first subscript stands for equation number and second one is for the variable. non-wife income means maybe husband's income or father's income. I am denoting non-wife income here because more is the husband's income less likely that a married woman will participate in the labour market. That is the assumption made which may or may not be true.

Because, the exact sign of $\beta_{13}$ will come out after estimation only. But, at least there is some kind of justification for non-wife income to be included in the model. Then plus $\beta_{14}$ how many kids you have less than 6 years of age, because if you have more number of kids, which are less than 6 years of age it would be very difficult for the married woman to participate in the labour market after taking care of those kids**.**

So, this is the labour supply equation of a married woman. We are going to estimate the labour supply function. Now, as we discussed earlier, wage is basically an endogenous variable because the observed market is actually an equilibrium which is determined by the demand and supply of labour.

That is why there is a two-way causal relationship between wage and labour supply and that is why wage is suspected to be endogenous here and we need to estimate this model using a simultaneous equation framework. Otherwise, the estimates will suffer from simultaneity bias. We will have biased estimation and inconsistent estimate if we apply OLS here.

All other variables are assumed to be exogenous, but you may ask how come education is actually exogenous. Education might also be indigenous, because we already discussed in our previous module that education might be correlated with a variable. Let us say ability factor, which will be there in the error term, let us say this is $u_1$. Ability cannot be observed and for the same reason education and the error term will get correlated through the ability factor.

But, we are ruling out the possibility of such omitted variable bias for the timing to make it to keep it simple. Try to understand that education might also be an endogenous variable but for the time being we are just keeping that aside. We are assuming all other variables are exogenous. So, since we cannot estimate this equation by a single equation method because of the simultaneity between wage and hours we will need to specify a demand for labour**.**

Now, those who are demanding labour will offer some kind of wage. So, that means the employer will offer some wage that is why the demand for labour is also known as wage offer function. And also here if you look at hours function, we have introduced wage in logarithmic form that is purely from the labour economics literature.

Generally, wage is represented in log form that is the idea. So, we need to specify another equation for wage which is called wedge offer function or level demand function**.** So, this is labour supply and this is wage offered function. Now, wage would be a function of again $\alpha_2 labour\ hours + \alpha_3 education + \beta_{22} experience + \beta_{23} experience\ square + u_2$**.**The coefficients of endogenous variables are denoted by alpha and coefficient of all exogenous variables are denoted by beta. We assume that this is also dependent on experience because how much wage the employer will offer to you depends on your previous experience and then we will also include experience square to check if there is any kind of non-linearity in the system**.** And then we will put $u_2$**.**

Now, here you look at that while experience is appearing in the wage offer function they are not appearing in the first equation because we assume how much labour is actually supplied in the labour market does not depend on your previous experience. That is the assumption. That is the assumption what I am making to make it simple. You can always challenge this assumption, because if you have some experience working in the labour market, probability of participating in the labour market would be higher, because you already know that where to work, what kind of work is available, whether it is suitable for you or not. So, experience makes people more likely to be participating in the labour market. That is why this labour supply will increase in that case. But, here we are assuming that experience and experience is square they do not appear in this**.**

Now, the question is whether the first equation which is the labour supply function is identified or not. Now, in the labour supply function we have to check order condition which says that you should have as many excluded exogenous variables as you have endogenous variables in that equation.

So, here I have only one endogenous variable. Number of exogenous variables is two- experience and experiences square which are not appearing here. So, that means since the number of excluded exogenous variable is more than number of endogenous variable in the equation we will say that the equation is over identified. But this is only necessary condition and not sufficient. What is the sufficient condition that we need? That at least one among these experience and experiences square must be significant**.** Because, if $\beta_{22,}$ $\beta_{23}$ both are 0 that means those variables are actually not appearing in the other equation. Yes, I understand they are excluded, but the significance of these two variables are required because they are called the observed shifter. So, that means if there is an increase in experience or experience square that will shift the labour demand function. So, the moment $\beta_{22}$ experience and experiences square change, the demand function will shift. If there is more experience, the employer will offer more, which so there will be right towards wage. If there is less experience, they will offer less wage.

So, this is supply, these are all demand functions. Let us say this is $d_1$, $d_2$, $d_3$, $d_4$. So, if you connect all these points then you will be able to identify the labour supply function**.** So, the two excluded variables should not only be excluded from the equation but must also be significant in the second equation. So, rank condition which is the second condition say that the excluded variable must have non-zero population coefficient.