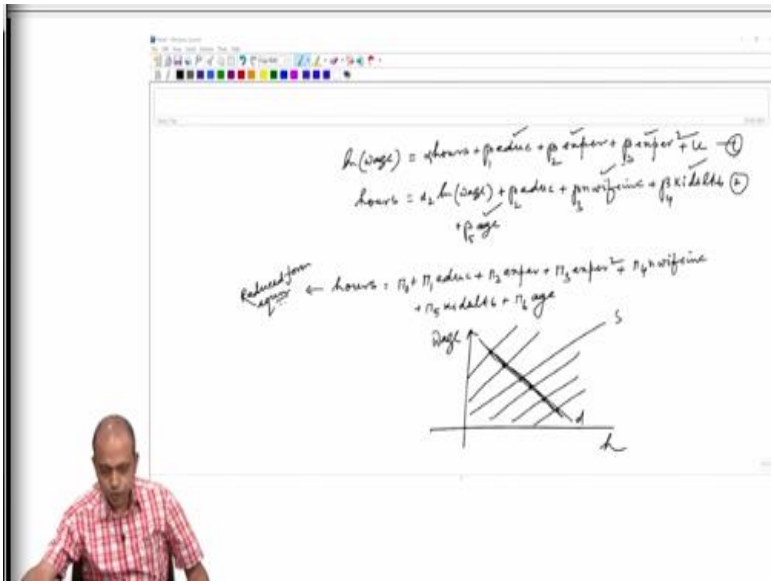
Applied Econometrics
Prof. Sabuj Kumar Mandal
Department of Humanities and Social Sciences
Indian Institute of Technology-Madras

Lecture-18
Simultaneous Equation Model-Part VII

So, we will come once again to our discussion on simultaneous equation model. And in our last class we were discussing about labour demand and labour supply function. And we also have learned how to estimate the equation using a data. We basically discussed about how to estimate the labour supply function wherein which was an endogenous variable. Now suppose our interest is just the opposite, that means we want to estimate a labour demand function or rather we said that this is a wage of our function.

(Refer Slide Time: 1:00)



Let us say that the function we want to estimate is: $\log(wage) = \alpha_1 hours + \beta_1 education + \beta_2 experience + \beta_3 experiencesquare + u$. This is the wage function we wanted to estimate. Now here hours is the endogenous variable because hours and wage are simultaneously determined. So we need to specify one equation for hours as well, so that would become:

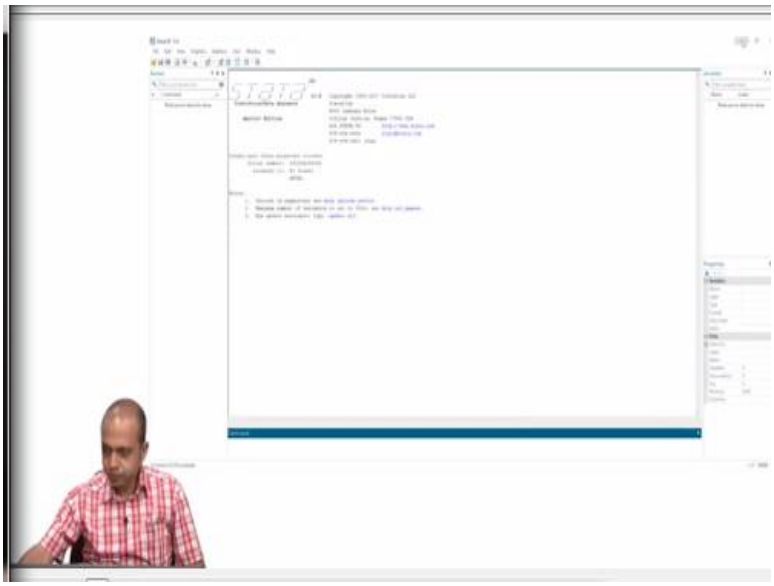
$hours = \alpha_2 \log(wage) + \beta_2 education + \beta_3 nonwifeincome + \beta_4 kidslt6 + \beta_5 ageofthemarriedwoman$

So, now we need to first write the reduced form equation for the endogenous variable which is hours. So, these 2 are the structural equations as I said. Now when you write the reduced form equation for hours as

$$hours = \pi_{i0} + \pi_{i1} education + \pi_{i2} experience + \pi_{i3} experiencesquare + \pi_{i4} nonwifeincome + \pi_{i5} kidslt6 + \pi_{i6} age$$

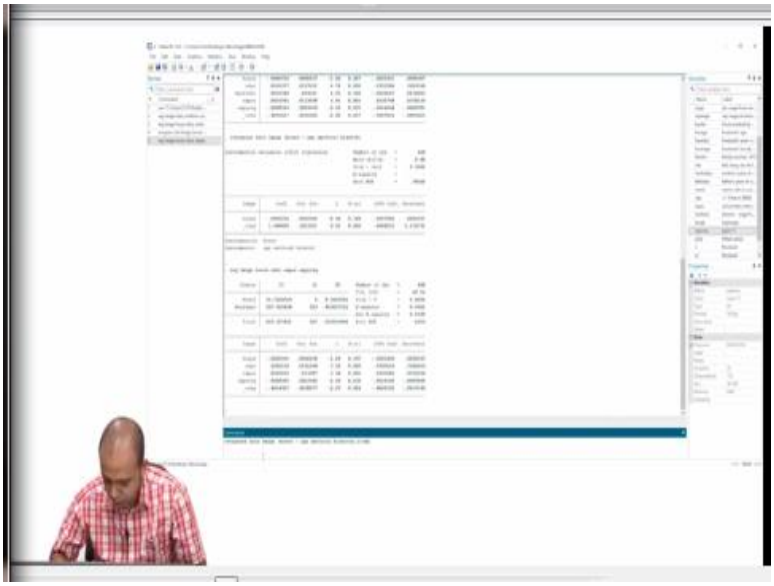
So, from this equation we will estimate hours and then we will put it in the structural equation following the 2-stage least square method. But as I said earlier, we will not do that manually because there would be some bias in the standard error. So we have to follow the STATA command for 2SLS to estimate this model. So, we will now go back to the same dataset.

(Refer Slide Time: 05:47)



First, again we will estimate the labour supply function without thinking of the simultaneity bias. So, reg hours, we are estimating the wage function reg of lwage, so that is basically here. So, log wage is a function of hours that is the main interest of our variable. Where is the wage? We are clicking this is the dependent variable and wage is a function of we are thinking of education. So, education is here. This is education and then you have experience and experience square. So, we are estimating this labour demand function without thinking of the simultaneity. So, what will happen if you estimate this?

(Refer Slide Time: 07:31)

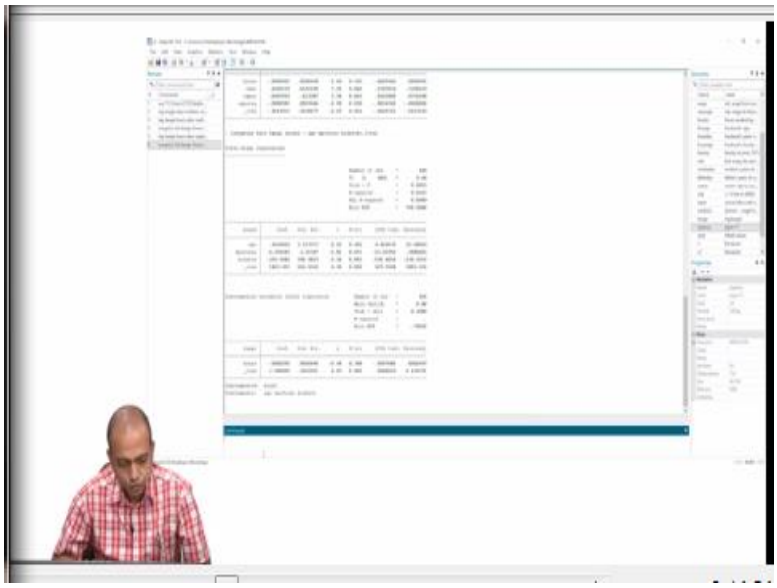


So, look at the coefficient 0.00000552, you have a negative relationship as labour supply increases, wage will decrease and that variable is not significant also. So, now what we have to do? We have to regress the 2 SLS method. Now in 2SLS we have to estimate the reduced form equation that means we have to use some instruments for this endogenous variable- hours. And what are the instruments? The variable which are excluded-non-wife income, kidslt6 and age are the 3 excluded variables from the wage offer function.

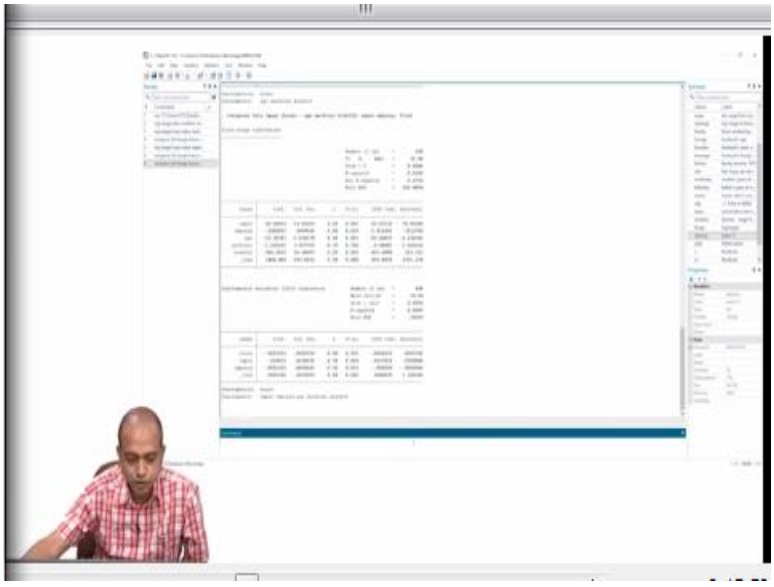
So, that means these 3 variables are actually working as instruments for the endogenous variable hours. You have to first identify what are the exogenous variables excluded from the wage offer function. Education variable is common for both the equations but non-wife income, kidslt6 and age variables are not appearing in the wage offer function. That means these 3 variables will work as observed supply shifter. So, that means for any change in non-wife income, kidslt6 and age, labour supply function will shift but labour demand function will not. As a result of which we would be able to identify the wage offer function. When there is a change in non-wife income or kidslt6 or age, the supply function will shift upward or downward. As a result, we will get new equilibrium and we will be able to identify the demand function. That is why I said that this variable non-wife income, kidslt6 and age are actually instruments. So, we will now estimate this equation 4. The dependent variable is log wage. Then for hours, which is our endogenous variable, we are using 3 instruments which are age, non-wife income and kidslt6**.** I will first regress log wage on hours and only exper and expersq.

This is the normal regression without thinking about the simultaneity. So, we have only 4 explanatory variables - hours, education, experience and experience square and this is the coefficient which is not significant.

(Refer Slide Time: 14:29)



(Refer Slide Time: 15:34)



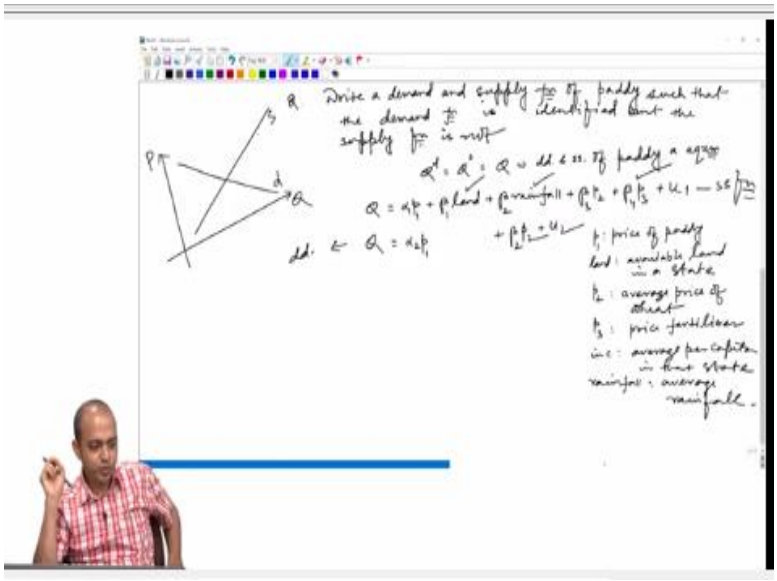
So this is the equation: IV regress 2sls, log wage then I have in the endogenous variable hours, these are the 3 instruments- age, non-wife income, and kids lt6, experience and experience square and then the other exogenous variables. So, in the first stage this is the reduced form equation for hours where you have experience, experience's square, age, non-wife income and kidslt6 which

are used as explanatory variables. And then finally I am getting log wage, it is dependent on hours, experience, this and that. Anyway, the hours variable is not significant even here also that is a different story altogether. But what I am saying this is actually the procedure if you want to estimate the wage or hour function as well. So, that means the logic is same, procedure is same, so for the endogenous variable we need to first construct the reduced form equation.

And from the reduced from equation, you need to get a predicted value of the endogenous variable. And that predicted endogenous variable should be used in the main structural equation to estimate the wage of our function, that is what I wanted to mean, is that fine? So, here as I said earlier we have used only these 3 instruments age, non-wife income, kids lt6 but STATA has used experience and experience's square also as instruments because these are the exogenous variables and exogenous variables themselves are the instruments for them. Now with this example at this stage you need to now understand very clearly how to write a simultaneous equation model of your own. And that too you should visualize very clearly when you are writing an equation whether the equation is identified or not. So, given a question you should be immediately able to write that question.

For example, let us say I am asking you to write a demand and supply of paddy, wherein let us assume that demand of paddy is identified but supply is not?

(Refer Slide Time: 18:38)



So, let us now assume that $Q_d = Q_s = Q$ is actually demand and supply of paddy at equilibrium. So, now I am writing this function let us say $Q = \alpha_1 p1 + \beta_1 land + \beta_2 rainfall + \beta_3 p2 + \beta_4 p3 + u$. And this is another equation equals to $\alpha_2 p1 + \beta_1 income + \beta_2 p2 + u_2$

Let us say p_1 is price of paddy, land indicates available land in a state let us say this is the state wise regression and rainfall is basically the average rainfall of the state, p_2 is actually the average price of wheat, p_3 is actually price of fertilizer, income is average per capita income in that state, rainfall is average rainfall. So, if I do not say anything let us say this is u_1, this is u_2, without specifying anything you can clearly understand which is your demand function, which is your supply function. Now since land, rainfall, price of fertilizer all those things are appearing in equation 1, obviously you will be understanding that the first equation is the supply function of paddy. Because rainfall, available land, price of fertilizer, if there is more rainfall that means paddy production or paddy supply would be more.

If the available land for paddy is more in that state, then paddy supply will go up. I have included average price of wheat also in the equation because more is the price of paddy, the farmers will have more incentive to grow paddy. So, we all know supply is an upward sloping function in a price quantity relationship. I have included price of a related crop, which is wheat, because cultivation of paddy has an opportunity cost since the same piece of land devoted for paddy cultivation will not be available for wheat cultivation. So, while deciding about whether to devote that piece of land for paddy cultivation and wheat cultivation, the farmer will think the relative price. And that is why the p_2 is also included in the function. If average price of wheat is more than the paddy, that means the opportunity cost of paddy cultivation for the farmer would be more. So, farmer may would like to then devote that piece of land for wheat cultivation which is more profitable because p_2 is higher than p_1.

That is the reason I have included p_2 in the supply function and p_3 is the average price of the fertilizer. If the price of fertilizer goes up, obviously the cultivation becomes less profitable. So, all these things they have a direct impact on the paddy production but they do not have any direct impact on paddy consumption. So, I can easily understand that the first equation is the supply function and second equation is the demand function where I have only price of paddy and average income. It is coming purely from the microeconomic theory of demand and also I have included

p_2, which is the average price of wheat. So, if we assume that wheat and rice are the 2 substitute products, that means they are related and substitute to each other. So, I can eat either rice or wheat depending on the relative price. So, if the price of p_2 goes up that means if wheat becomes costlier then the consumers will demand more rice.

Similarly, as p_2 goes down then the consumers will demand lesser paddy because wheat is now cheaper to eat than rice. So, that is why I have included price of the related product. So these are all coming from the microeconomic theory of demand and supply. Now the question is, in these 2 equations as I said price and quantity are the variables and other things are all parameters.

Because in price quantity relationship if this is price and this is quantity, so I can draw the demand and supply by keeping other factors constant. So, when there is a change in other factors, we will say that the demand or supply function will shift parametrically. But when there is a change in price there would be change along the demand curve or along the supply curve. Now which among these 2 equations is actually identified?

Now think about the first equation, I have one endogenous variable and then the income variable is excluded from the equation. So, that means order condition is satisfied and if β_1 is significant then I can say that the first equation is also identified. But my question was supply function is identified but demand function is not.

So, this is my supply function and this is my demand function let us make it ulta- demand function is identified but the supply function is not. Now look at the demand function- I have 1 endogenous variable and so many excluded variables from there so demand function is easily identified. I have land excluded, I have rainfall, I have p_3 which is the price of fertilizer. These 3 variables are excluded which is why demand function is actually over identified.

What about the supply function? For supply function I have the variable excluded income. So, if you do not want supply function to be identified, we need to remove this variable. Now look at the equation. So, p_2 is included in the supply function also, so that means supply function does not include any variable which is actually appearing in the demand function. That is why supply

function is not identified. So, if you want to identify an equation in a simultaneous equation model you must include some variable in the demand function which is actually not appearing in the supply function. So, average price or average income of the individual if you include in this equation then only you can actually identify the supply function otherwise it is not possible to identify the supply function. So just think that I need to specify the equation such that if you want that equation to be identified, in other equation we must include some variables which is not appearing in this equation. Then only for a change in those factors that function will shift and which will help this function to be identified. But this inclusion of variable in 1 equation it is not adhoc; that also you have to keep in mind.

When you are including land, rainfall, price of other product, fertilizer, so all these variables are actually coming from a theory, there is enough justification to put land, rainfall, price and all these things. Simply I cannot put a variable which has nothing to do with my paddy cultivation. If you include average intelligence of the people in that state, that has nothing to do with paddy production.

This variable what is included in this function that is how the other function is identified. I need to include the variable with a proper theoretical justification. So, we need to go back and think when you will be working with your project, so if the equation demands to be estimated in a simultaneous equation framework, we should carefully formulate the 2 equations in such a way that both the equations are actually identified.