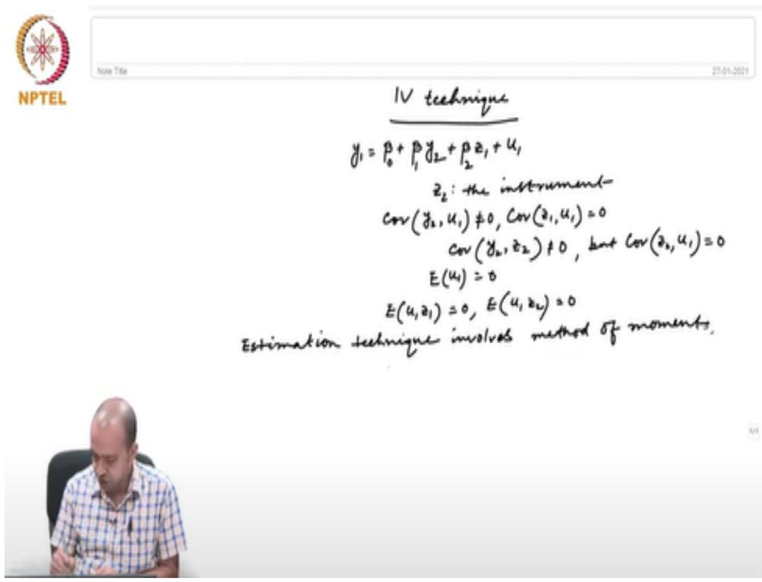
Applied Econometrics
Prof. Sabuj Kumar Mandal
Department of Humanities and Social Sciences
Indian Institute of Technology, Madras

Lecture - 03
Instrumental Variable Estimation – Part III

(Refer Slide Time: 00:16)



Welcome to our discussion on instrumental variable estimation technique. In our last class we were talking about instrumental variable estimation technique for a model where you have more than one explanatory variable. our model

$$y_1 = \beta_0 + \beta_1 y_2 + \beta_2 z_1 + u_1$$

as an example or you can think of that wage is actually a function of education y_2 indicates education. And then z_1 indicates let us say experience, z_2 is actually the instrument,

$Cov(y_2, u_1) \neq 0$; $Cov(z_1, u_1) = 0$

$Cov(y_2, z_2) \neq 0$; $Cov(z_2, u_1) = 0$

$E(u_1)=0$

Now in this context the estimation technique what we discussed in our previous class that we would like to make a modification.

Modifications in a sense that particular estimation technique where we are formulating a reduced form equation for y_2. And then we are getting the predicted value of y_2 from the reduced from equation and then we are replacing the predicted value in the structural equation. That technique we would like to use for some other model but here we will use a different technique. That means there is no problem as such using the technique what we discussed in your previous class for estimating this model.

But what I am saying that particular technique is applicable in a specific context and in this model if you apply that technique and the technique what I am going to discuss now both will result in similar estimates. It means the technique is not making any change in this particular model, that is why I said, we would preserve the estimation technique we discussed in our previous class for another model that I am going to discuss later on.

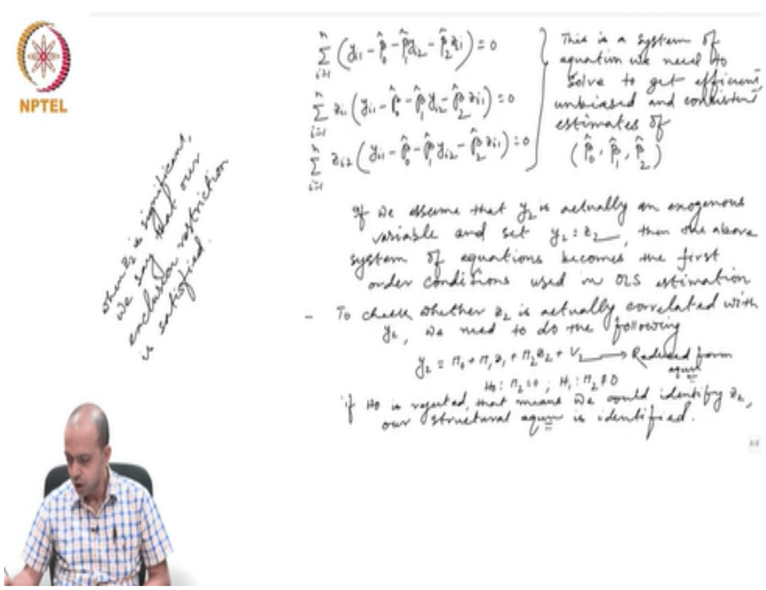
The above model that basically results in saying that expectation

$$E(u_1, z_1) = 0$$
$$E(u_1, z_2) = 0$$

Since this above condition is satisfied the estimation technique involves method of moments.

What is that method of moment condition?

(Refer Slide Time: 06:03)



The condition is

$$\sum_{i=1}^{n}(y_{i1} - \widehat{\beta_0} - \widehat{\beta_1}y_{i2} - \widehat{\beta_2}z_{i1} = 0$$

$$\sum_{i=1}^{n}z_{i1}(y_{i1} - \widehat{\beta_0} - \widehat{\beta_1}y_{i2} - \widehat{\beta_2}z_{i1} = 0$$

$$\sum_{i=1}^{n}z_{i2}(y_{i1} - \widehat{\beta_0} - \widehat{\beta_1}y_{i2} - \widehat{\beta_2}z_{i1} = 0$$

these are the three moment conditions that we need to use.

This is basically this is a system of equation we need to solve to get efficient unbiased and consistent estimates of $\widehat{\beta_0}, \widehat{\beta_1}, \widehat{\beta_2}$. it means this is a system of three equations and we are going to solve for three unknown parameters $\widehat{\beta_0}, \widehat{\beta_1}, \widehat{\beta_2}$. So, we can easily get it. Now if we closely look this system of equations one thing we can understand. If we assume that y_2 is actually an exogenous variable and set $y_2 = z_2$.

Because in our previous class we have already discussed that an exogenous variable is its own instrument. So, that means when we assume y_2 is exogenous it can be used for its own instruments, so if we set $y_2 = z_2$, then what will happen? Then the above system of equations actually becomes the three normal equations that means becomes the first order condition used in our OLS estimates.

So, that means this equation these are all nothing but actually what is this

$$\sum_{i=1}^{n}(y_{i1} - \widehat{\beta_0} - \widehat{\beta_1}y_{i2} - \widehat{\beta_2}z_{i1} = 0$$

that is basically u_1. So, we are trying to solve

$$\sum \widehat{u^2}$$

It means these three equations. If we solve so these three equations when we assume y_2 is actually exogenous variable then that would become if you go back and look at that not two normal equations of our ordinary list square technique estimation.

Then these equations are nothing but the three normal equations that we have used in OLS estimation. So, that means when y_2 becomes exogenous obviously this technique IV technique becomes the OLS is one. That is the estimation Technique we should use when your model involves more than one explanatory variable. So, that means here we said that z_2 is basically an instrument.

What we need to do we need to identify that instrument? That means we assumes it to an instrument but we have not yet checked whether the z_2 variable is actually correlated with the endogenous variable or not. How to check it? To check that condition, what we need to do to check whether z_2 is actually correlated with y_2 we need to do the following. we will write the reduced form equation for $y_2$which is nothing but

$$y_2 = \pi_0 + \pi_1 z_1 + \pi_2 z_2 + v_i$$

And what is the identification condition? Identification condition is the null hypothesis we have to check that

H_0: $\pi_2 = 0$

against the alternative

H_1: $\pi_2 \neq 0$

If we reject the null that means we could identify z_2 and that means our structural equation is actually identified. Of course, the exact meaning of identification you may not be able to understand at this context that when I will be discussing on simultaneous equation model.

For the timing you just understand that this is the identification condition where z_2 is significant and that means since we are trying to use the instrument for y_2 which is actually used in the structural equation. this I will say that the moments z_2 become significant it means our structural equation is also identified.

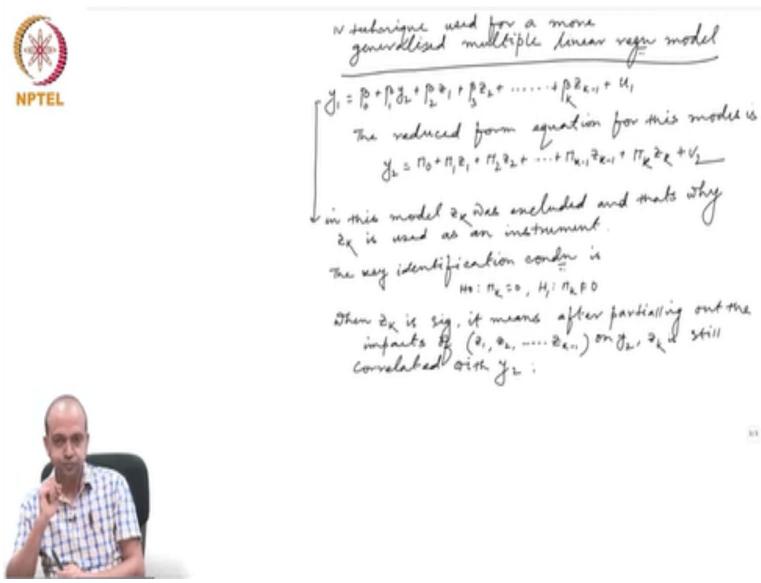
The original equation was structural form because this describes the structure of an economic theory where it says that wage is a function of your education as well as experience. But this reduced form equation as such is not coming from any economic theory rather it is simply an equation where an endogenous variable is a function of all exogenous variable included as well as excluded.

If you look at here

$$y_2 = \pi_0 + \pi_1 z_1 + \pi_2 z_2 + v_i$$

y_2 is a function of z_1 and z_2 where z_1 is an exogenous variable which is already included in the model. But z_2 is actually an excluded exogenous model, so that means the moment z_2 is significant we will say that our exclusion restriction is actually valid. When z_2 is significant we say that our exclusion restriction is satisfied.

(Refer Slide Time: 18:38)



Now suppose we would like to extend this multiple linear regression model into a K number of explanatory variable so what will happen when you have more two explanatory variable then our model will become the generalized. So, this is IV technique used for a generalized multiple linear regression model. So, what is that model? That model would become

$$y_1 = \beta_0 + \beta_1 y_2 + \beta_2 z_1 + \beta_3 z_2 + \cdots + \beta_k z_{k-1} + u_1$$

This is the more generalized model with k-1 number of explanatory variables and in that case what. for this model the key identification condition would become. before we discuss about identification condition let me write the reduced form equation. The reduced form equation for in the above model for this model, so your endogenous variable is

$$y_2 = \pi_0 + \pi_1 z_1 + \pi_2 z_2 + \cdots + \pi_{k-1} z_{k-1} + \pi_k z_k + v_2$$

Now if you look at in this generalized model z_k was actually excluded. In this model z_k was excluded and that is why z_k is used as and an instrument. The key identification condition is
H_0: $\pi_k = 0$ against the alternative
H_1: $\pi_k \neq 0$
That means the excluded variable should be significant, that is the key identification condition. When z_k not equals to zero this null hypothesis is rejected then we say that we could identify the instrument our structural equation is the identified. Actually, when I say that z_k is significant it means after partialling out the impact of impacts of z_1, z_2 , z_{k-1} on y_2 z_k is still correlated with y_2, that is the meaning, that we should understand. So, that means in the reduced form equation you have so many exogenous variable and all of them are correlated with the y_2.

So, what we are actually doing here in this reduced form equation after removing or partialling out the impact of all these exogenous variables we are trying to find out whether z_k is still correlated with y_2. That is the idea basically we learned from our multiple linear regression models. So, what we are checking here whether z_k is correlated with y_2 whether the impact of z_k is significant on y_2.

How will you do that? you have to first control for other explanatory variable, other exogenous variable which are already there in the model. That is why in this equation what I am saying after partialling out the impact of all this exogenous variable z_k is still correlated with y_2 that is the meaning.