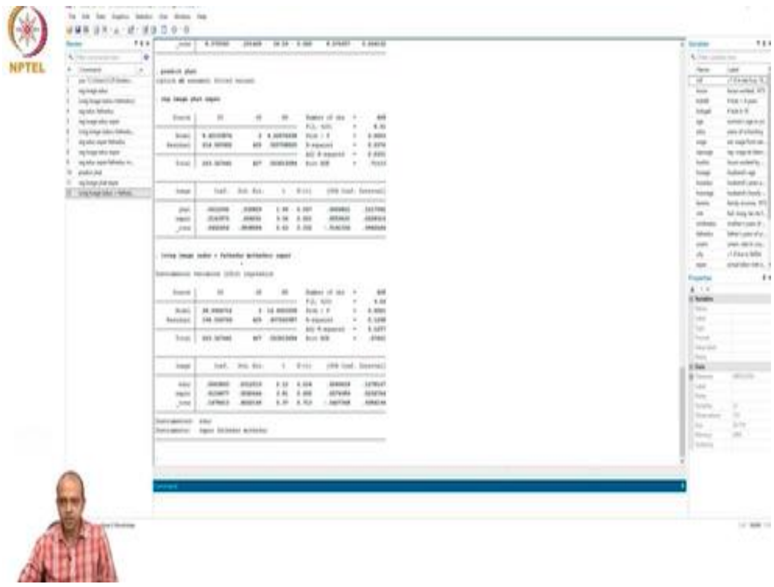
Applied Econometrics
Prof. Sabuj Kumar Mandal
Department of Humanities and Social Sciences
Indian Institute of Technology, Madras

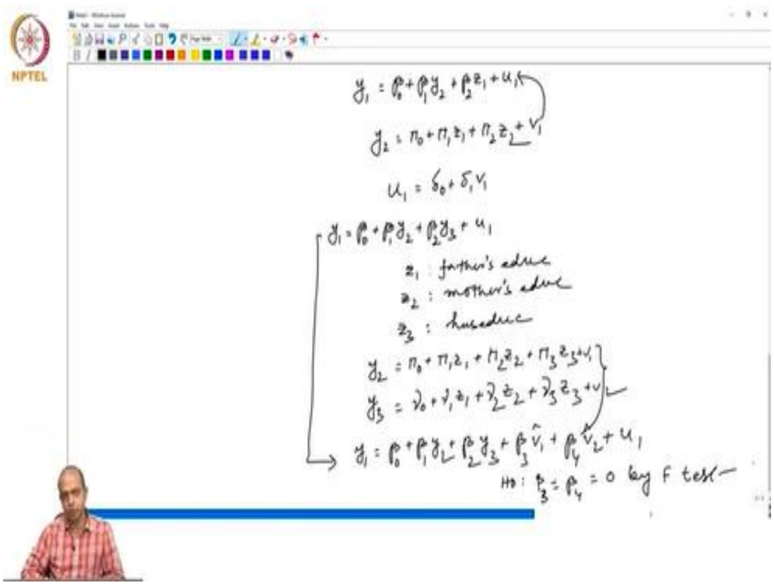
Lecture - 09
Instrumental Variable Estimation – Part IX

(Refer Slide Time: 00:16)



The next question is, if we want to test as we said that the model is if there is actual endogeneity problem then only, we should use the 2 SLS method. Now the question is, how to test endogeneity? So, to test endogeneity what we have to do? We have to estimate the model.

(Refer Slide Time: 00:55)



That means what we will do, if you recall the test of endogeneity $Y_1 = \beta_0 + \beta_1 Y_2 + \beta_2 Z_1 + U_1$. So, what we have to do? That means here we are saying, $y_2 = \pi_0 + \pi_1 Z_1 + \pi_2 Z_2$, Z_2 is the instrument + v_1. So, that means we are hype the only channel as we said by which, Y_2 can be correlated with u_1 if there is correlation between v_1 and u_1. So, that means, we are saying u_1 = $\partial_0 + \partial_1 v_1$. This is the relationship.

So, that means, from this reduced form we will collect the estimated value of the error term. And that we will put in the original structural form equation as an additional explanatory variable. Predicted value of the error term from the reduced form equation will be used, as an additional explanatory variable in the original structural form equation. And if the predicted value of the error term becomes significant, we will say that there is endogeneity. That is what we have discussed so, we will do that.

(**Video Starts: 02:31)**

So, here what we will do, we will first estimate the reduced form equation reg education = if a father's education. And this is the model, and from this what we have to do is to estimate the predicted value of the error term from the reduced form equation. How to do that? So, that means, from here we have to use this command, predict let us say, v, residual. This v will be used in the original equation as an additional explanatory variable.

So, reg 1 wage education and then v. And if we see here, v is actually not significant. That means there is endogeneity, there is no significant endogeneity problem in this particular data. This is how manually we can test endogeneity, manually. Now how will you test using the Hausman test command? The command is you use them you estimate the model using i v reg then education lwage, then education = fathers' education.

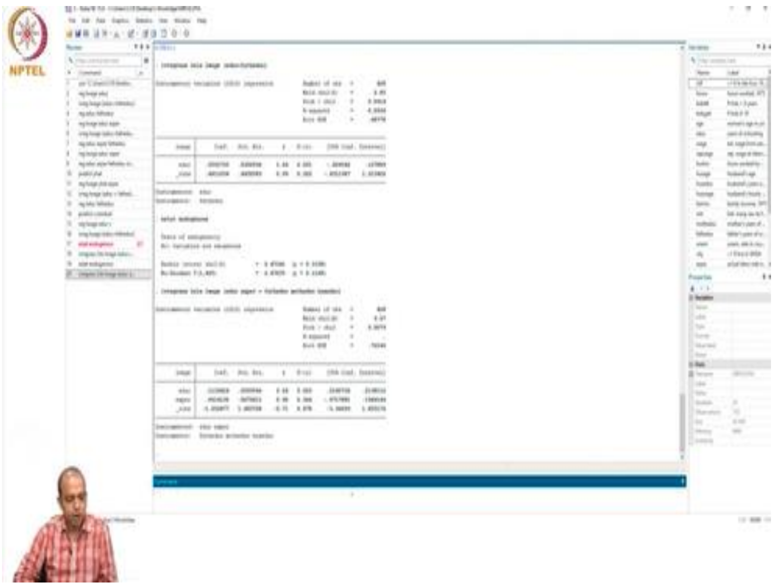
Then we have to use this command, estat endogenous IV 2 SLS command values lwage education = fathers' education and this is the result. So, instead of IV reg we have to use IV regress 2 SLS command to get this equation. Now look at, here what is the test we are doing? This is a test of endogeneity. What is the null hypothesis? Variables are exogenous. Then stator is reporting 2 test statistic.

For the timing you just keep this Durbin score aside. Just think about this Wu-Hausman F test. What is the value? 2.47 p value is 0.1168 which is not significant. So, that is why, the insignificance p value is showing there is no endogeneity. Now suppose, we have two endogenous variable in the model and we have more than two instruments that is the model we are going to estimate.

So, that means IV regress 2 SLS then lwage. Let us assume experience is also an endogenous variable. Let us assume that education and then experience both are endogenous variable and we have father's education, mother's education, as well as husband's education. This is the model.

(Video Ends: 07:39)

(Refer Slide Time: 07:40)



Now one thing again, I would like to mention here when we estimate this model, we have 2 endogenous variable and we have 3 instruments. And in this box data is showing instrumented or education and experience. That means these are the 2 endogenous variables, that is why we can easily identify 2 reduced form equations and these are the instruments. But one natural question comes to our mind.

When we mention these in the bracket, education and experience and then there are 3 instruments, which instruments are used for which particular endogenous variable? Because I am not mentioning anything here. I am simply mentioning education and experience as endogenous variable, then I am putting 3 instruments. So, obviously, the natural question will come to our mind that out of these 3 instruments, which among these 3 instruments are used to estimate education hat and which among these 3 are used for experience?

Now to answer this question, once again we will go back to our theory which says, let us now our model is$Y_1 = \beta_0 + \beta_1 Y_2 + \beta_2 Y_2 + U_1$. And we have 3 instruments let us say Z_1 which is father's education, Z_2 which is mother's education and z 3 which is husband's education. Now what would be the reduced form for Y_2? Y_2 would be $\pi_0 + \pi_1 Z_1 + \pi_2 Z_2 + \pi_3 Z_3$

And what would be the reduced form for Y_3 ? $\gamma_0 + \gamma_1 Z_1 + \gamma_2 Z_2 + \gamma_3 Z_3$. If you look at the reduced from equation for both Y_2 and Y_3, they are actually same. So, that means, all the 3 instruments are actually appearing in both the reduced form equation because, that is the definition of the reduced form equation. What is the definition of reduced form?

The endogenous variable is a function of all the exogenous variable included as well as excluded. In this model, we have not assumed any included exogenous variable. That is why it is a function of all the exogenous variables which are actually excluded from the model. So, since the reduced form equations for both Y_2 and Y_3 they are same. That is why we need not differentiate the reduced form equation of Y_2 and Y_3. That is why there is no need to specify anything here in this command.

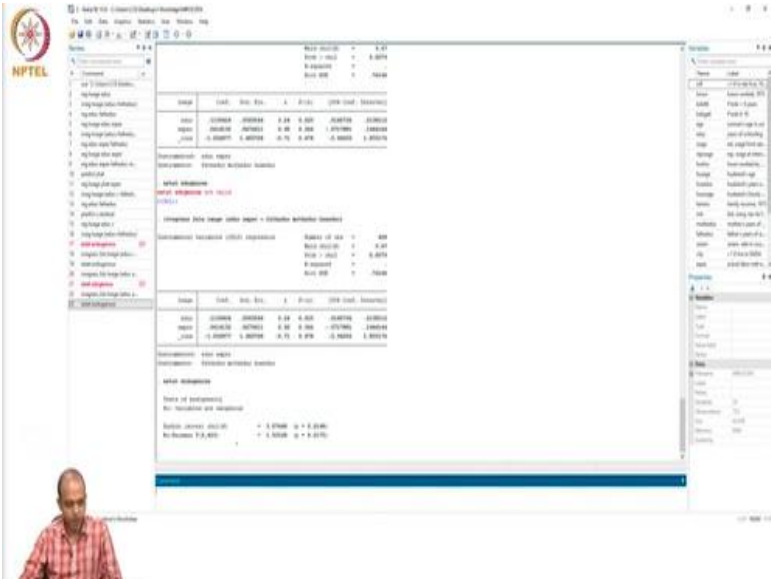
(**Video Starts: 11:42**)

That means here I am putting only education and experience to denote they are my endogenous variable. After that I am simply putting all the instruments. Because all the instruments will be used to estimate both the reduced form equation. Now after once we estimate, if we put estat endogenous. So, this is our model, IV regress 2 SLS lwage education experience, then we have 3 instruments, we have estimated the model using 2 SLS.

This is the command. So, once again we see the Wu-Hausman value is 1.53 which is actually not significant. Look at the p value not significant. That means there is no endogeneity.

(Video Ends: 14:06)

(Refer Slide Time: 14:06)



But as we discussed theoretically, there is a problem with this test because what this data is, what is this Hausman test doing if you recall when we have 2 instruments. So, that means from this reduced from equation we will get let us say this is v_1 and this is v_2. So, we will put v_1 hat and v_2 hat in the original equation. So, that means, our equation would become y 1 = beta 0 + beta 1 Y_2 + beta 2 Y_3 + l beta 3 v_1 hat + beta 4 v_2 hat + u_1. $Y_1 = \beta_0 + \beta_1 Y_2 + \beta_2 y_3 + \beta_3 \widehat{v_1} + \beta_4 \widehat{v_2} + U_1$

So, that means we are using the estimated value of the error term as a additional variable. And what we are doing here we will test this null hypothesis beta 3 = beta 4 = 0 by F test. So, that means we will test the joint significance of beta 3 and beta 4 by F test. And what is the alternative of F test? Alternative is at least one among them is significant. So, that means when it is rejected, we can only say that at least one among these two Y_2 or Y_3 are endogenous.

But we do not know which among these two is actually endogenous is it Y_2, or Y_3, or both. That is the problem of this test. The limitation of this test, when there is multiple endogenous variables. Multiple endogenous variable that is the problem. So, that is why here when I am saying estat endogenous, that means when I am putting that Hausman test we are we can only say that both of them are not.

That means the null hypothesis is at least one among them are exogenous alternative hypothesis. In this case it is not rejected. So, that is fine. In case it is rejected, we can say only one among them are endogenous. How to solve this problem? When we have multiple endogenous variables that means this simple test is not valid. So, we will apply another test and that will discuss in our next class. Thank you.