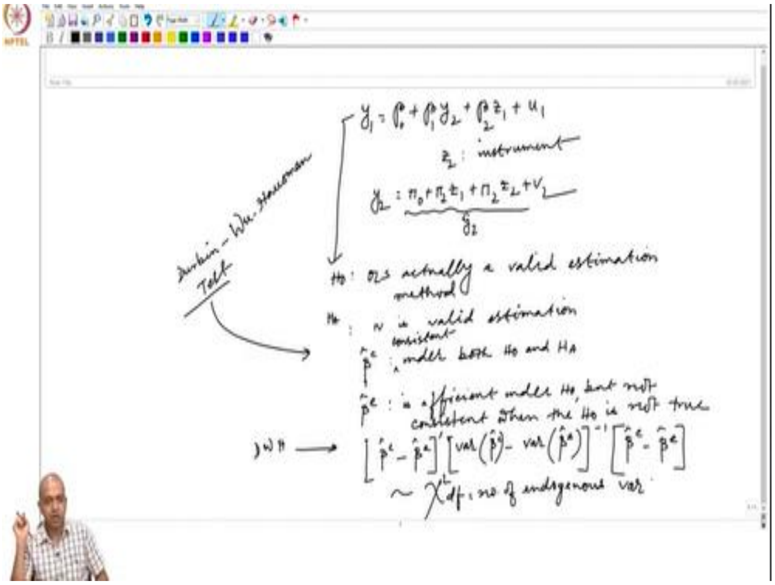
Applied Econometrics
Prof. Sabuj Kumar Mandal
Department of Humanities and Social Sciences
Indian Institute of Technology, Madras

Lecture - 10
Instrumental Variable Estimation – Part X

(Refer Slide Time: 00:15)



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

z_2 : instrument

$y_2 = \underbrace{\pi_0 + \pi_1 z_1 + \pi_2 z_2}_{\hat{y}_2} + v_2$

Durbin – Wu Hausman Test

H_0 : OLS actually a valid estimation method

H_A : IV is valid estimation

$\hat{\beta}^c$: consistent under both H_0 and H_A

$\hat{\beta}^e$: is efficient under H_0, but not consistent when the H_0 is not true

DWH $\longrightarrow$ $[\hat{\beta}^c - \hat{\beta}^e]' [var(\hat{\beta}^c) - var(\hat{\beta}^e)]^{-1} [\hat{\beta}^c - \hat{\beta}^e]$

$\sim \chi^2_{df = \text{no of endogenous var}}$

Welcome to our discussion on instrumental variable estimation technique. And yesterday we were talking about we were discussing how to detect endogeneity in a model before we actually implement instrumental variable estimation. And the test what we are discussing was the Hausman test which is also known as Durbin Wu Hausman test because the test was developed by these three econometrician Durbin, Wu and Hausman.

So, quickly we will recap what we are doing in this test basically let us $Y_1 = \beta_0 + \beta_1 Y_2 + \beta_2 Z_1 + U_1$, Y_2 is the endogenous variable and this is z 2 which is instrument that we are using. So, basically what we are doing we were running a reduced form equation $y_2 = \pi_0 + \pi_1 Z_1 + \pi_2 Z_2 + \pi_3 Z_3 + v_2$. So, we are trying to get the this is Y_2 hat and this is v_2 hat and we were, what we are doing?

We are estimating the predicted value of the error term from this reduced from equation and we are putting the predicted value of the term in the equation as an additional explanatory variable. And then we are checking whether the coefficient of that predicted error term is significant or not.

If significant then we say that there is endogeneity if not then we will say that there is no endogeneity. So, that is the mechanism we discussed.

(Video Starts: 02:20)

So, once again we will use the same data set, we will see. So, this is our data set and what we will do? We will first regress education on experience let us say experience and for father’s education or an mother's education. So, this is the model this is the reduced form equation for integration we are regressing on the excluded exogenous variables father's education and mother's education and included exogenous variable is experience.

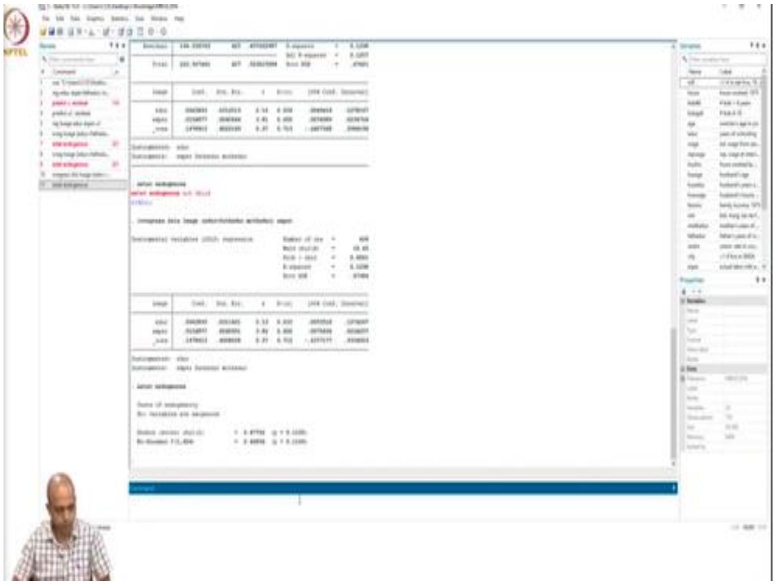
Then predict let us say V residual and then variable we already defined then what we will do? We will take predict some different name predict u 1 residual. So, then we will regress what we will do? Regress log wage on education, experience and this predicted error term. And if we look at the coefficient of this predicted error term the T value is 1.60 and the P value is 0.10 so that means this is not significant.

So, that means what would be our conclusion that education is actually not an endogenous variable in this particular context. Now the manual detection of instrumental variable estimation technique we can actually do by these using the statistical using the status command. In that case what we have to do? We have to first estimate the equation using IV reg then we will put education equals to father's education, mother's education and then we have experience so this is the model.

After this what you have to do estat endogenous that is what we discussed. So, this is our model IV reg then the command what we are using estat endogenous. Yesterday it was working, what is the problem of this command estat endogenous we will use instead of this IV regress 2 SLS.

(Video Ends: 07:10)

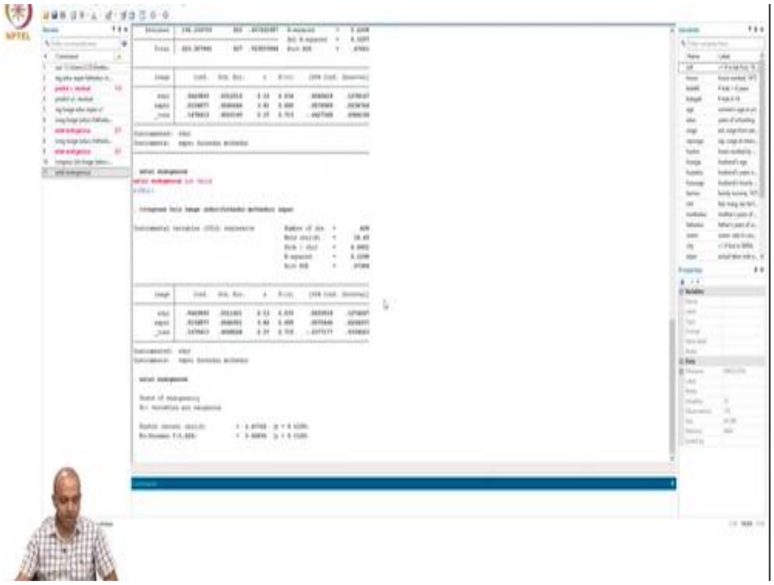
(Refer Slide Time: 07:10)



So, instead of IV reg we have used IV regress 2 SLS and then this estat endogenous command is working. Now what we have to take this Durbin Wu Hausman statistic which is the F value and corresponding T statistic is 2.46. Now this F value this Durbin this Wu Hausman statistic is basically the F value is nothing but if we go back then it is actually the square root of this T value.

So, that means when we are doing it manually, we are using T stage T statistic to test the individual significance of the predicted value of the error term.

(Refer Slide Time: 08:06)



And when we are using this Durbin Wu Hausman test that is the F statistic we are using where the value of this F is nothing but this square root of the previous T value. Now what exactly we are doing here in this Durbin Wu Hausman test that is something we need to discuss. Now in Durbin Wu Hausman test what they suggest basically if you recall what is the philosophy of Hausman test.

We need to estimate the model using OLS, we need to estimate the model using IV and then we need to check whether the coefficients estimated value of the coefficients derived from the OLS and that from the IV they are statistically different or not. If there is a significant difference in between OLS and IV estimates then we will say that OLS is actually not an appropriate method of estimation.

So, let us say that the same model this model we are using fast OLS or null hypothesis is OLS is actually a valid estimation method that is the null hypothesis. And what is the alternative? Alternative hypothesis is that IV is valid estimation. Let us assume that beta hat this is c that means consistent estimate of this beta 1 hat what we are doing let us say this is the endogenous variable. So, beta 1 hat which is consistent under both H_0 and H_A.

So, what we know that if OLS is actually a valid estimation technique that means if there is no endogeneity in the model then whether we apply OLS or IV both estimates would be consistent. So, there will not be any problem in consistency. So, beta hat c that means is consistent under both H_0 and H_A and beta hat e the efficiency property is efficient under H_0 but not consistent when the null is not true.

So, that means if OLS is actually not valid if there is endogeneity and if we still apply OLS then the resultant estimates will not be consistent, that is a problem of applying OLS. When there is endogeneity, we will lose consistency property. So, once again what I repeat that in presence of OLS if there is no endogeneity whatever we apply whether it is IV or OLS both the estimates would be consistent, consistency could be preserved.

But if there is no endogeneity and if we still apply IV then the efficiency property to some extent we have to sacrifice because IV estimates standard error are larger than the OLS standard error. And beta hat e which is actually efficient estimator that means it is fully efficient under H_0. That means when there is OLS if we apply OLS fully efficient when there is no endogeneity. But it is not consistent when H_0 is true that means when there is endogeneity this beta hat e is actually not consistent.

So, that is why the test statistic is constructed in Wu Hausman test this is Durbin Wu Hausman test we are talking about. So, it is beta hat c - beta hat e transpose, these are all in vector form. Variance of beta hat c - variance of beta hat e inverse beta hat c - beta hat e that is Durbin Wu Hausman test or in short DWH. And these follows X chi-square distribution with degrees of freedom equals to number of endogenous variable.

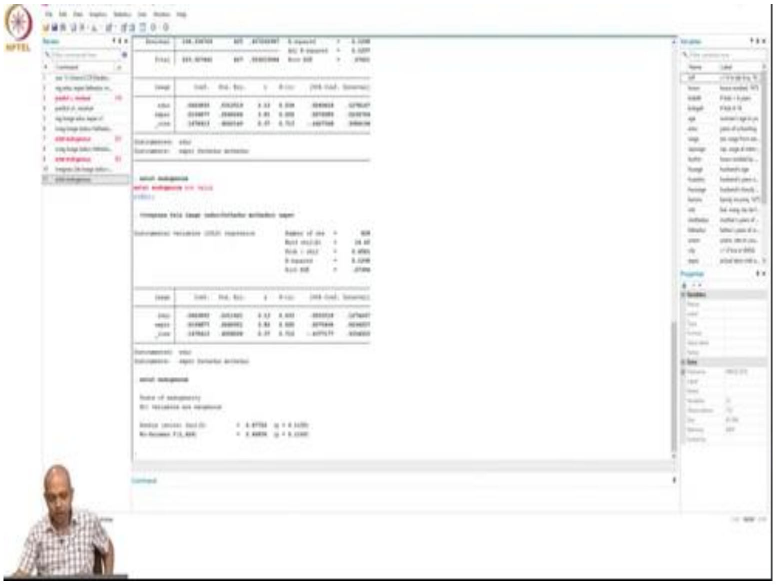
In this case we have only one endogenous variable that is why it is 1 here so this will follow a chi square distribution. So, these tests this Durbin Wu Hausman test is basically chi square test.

(Video Starts: 16:15)

So, that means here in this output we have one Durbin score and that Durbin score is actually coming from here. Because this is a chi square distribution and if you look at this is also a chi square distribution. But other Hausman test what we manually did that is basically F test which is basically the square root of this T. So, stata is reporting Durbin chi score as the Durbin Wu Hausman test.

(Video Ends: 16:57)

(Refer Slide Time: 17:08)



So, this test can also be implemented additionally what they say so that means we can say that this test is mode of a test for whether OLS is valid or not rather than testing endogeneity. So, the entire philosophy of Hausman test is based on two alternative methods of estimating the same equation, 1 by OLS, another by IV and then we test whether there is significant difference between the beta hat OLS and IV because here we are writing beta hat c and beta hat e, c is consistent e is efficient.

Consistency is preserved under IV and this is under OLS. So, this is IV - OLS again this is variance of beta hat IV - OLS. So, the entire DWH test statistic is based on how far the IV estimates are different from the OLS estimates that is the philosophy that is the logic of Hausman test. So, this logic also we can implement using stata command and we will now see how to do it.

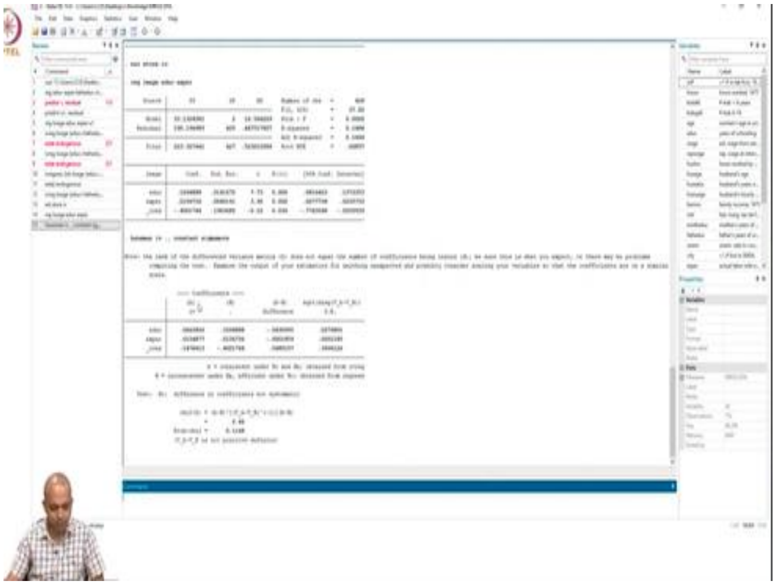
(Video Starts: 18:17)

So, we will say that let us say reg then education equals to father's education, mother’s education then experience so we are estimating the model using IV then we will store this result EST store IV. Then we will use the same equation estimate the same equation using OLS. So, using command take education, experience. Then what we will do we will put Hausman IV and the command is constant, sigma this is the command.

So, we will first estimate the model using so this is what is working here. So, what I am saying we will first estimate the model using IV that means two SLS method then we will use the same equation by OLS. And then by putting this command Hausman IV constant sigma more we are actually asking stata to report the DWH test statistic. How is it doing? Look at here B that means these IV estimates are consistent under H_0.

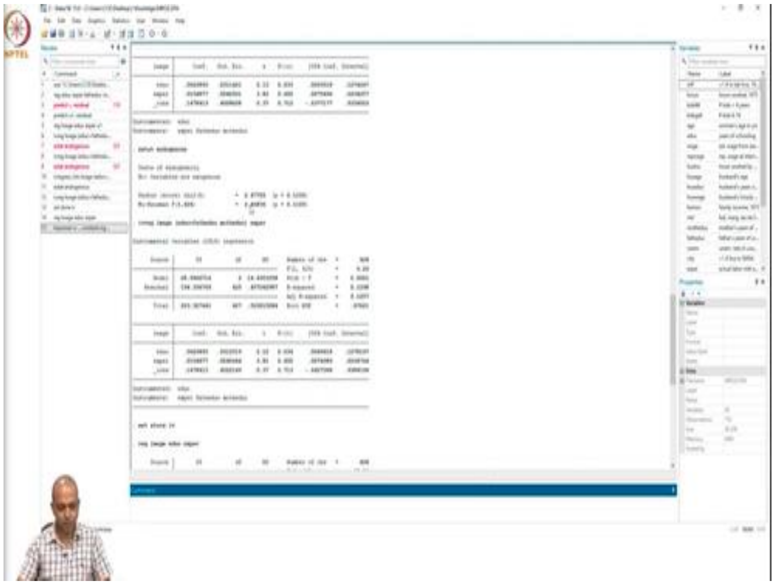
(Video Ends: 20:14)

(Refer Slide Time: 20:14)



So, what is H_0 OLS is appropriate. When is OLS appropriate? There is no endogeneity. So, when there is no endogeneity? IV is consistent OLS is also consistent. But this B is actually inconsistent under H_A. So, that means when there is endogeneity and if we apply OLS then that would be inconsistent. So, in our result this beta hat c and beta hat estat is denoting as b and capital B and then they are taking this and ultimately, they are arriving at the chi square value which is 2.46.

(Refer Slide Time: 21:17)

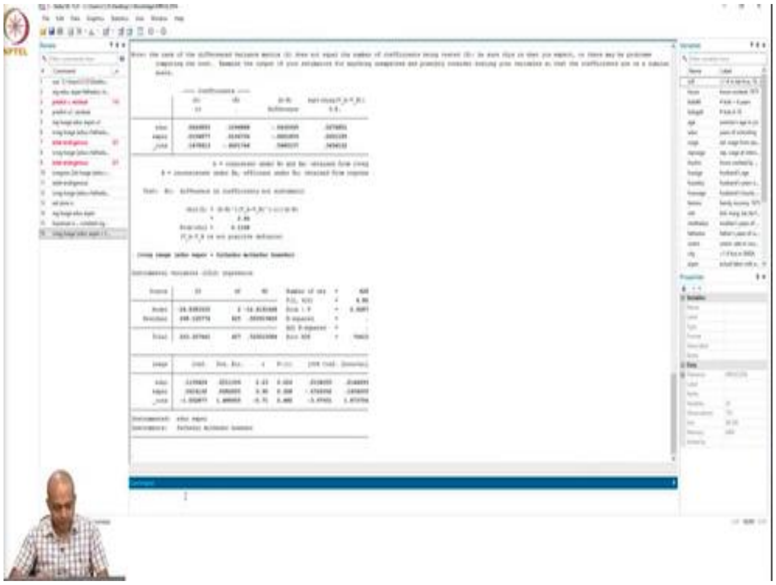


And if we look at the start endogenous look at the same thing 2.46 so chi square is 2.46. So, almost this chi square sorry this is the F test this is 2.47. So, almost same value we are getting at. So, this is a chi square test statistic that means whatever we are getting here and the same things data is reporting here in this chi square. So, this particular test is based on Durbin Wu Hausman test this is the stata command.

And test statistic once again this is the test command test statistic beta hat c - beta hat e variance of this. So, either we can do it manually or manually generally we do not do it. We will simply use this is stata endogenous command where is the state endogenous if you look at this is the command. So, this is how we have to test for endogeneity. But this test as I said it has a limitation. For example, what is the limitation we said?

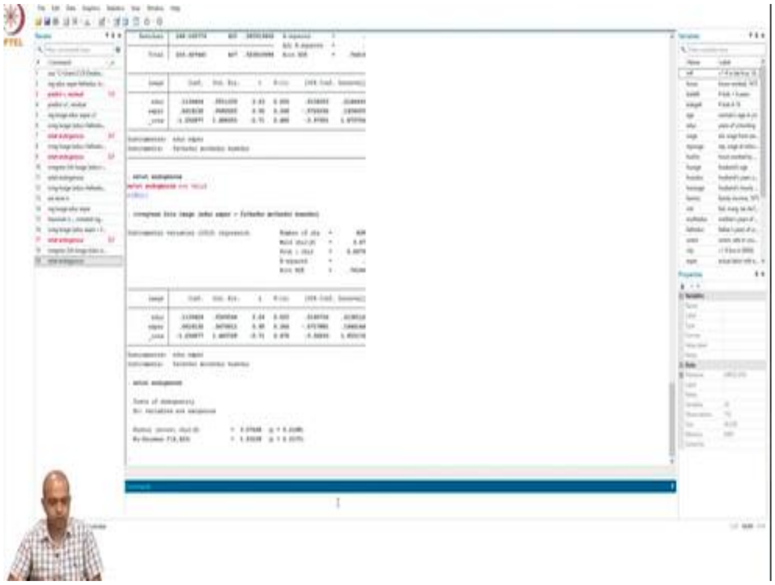
When there is more than one endogenous variable it will only say whether there is endogeneity or not but it cannot detect which particular variable is endogenous. So, for that I will give you an example. So, IV reg LH then education and let us say we have experience also as an another endogenous variable equals to and then if we put sorry this IV regress we have to use otherwise it will not work.

(Refer Slide Time: 23:48)



So, IV regress 2 SLS so whenever we are using this command, we have to use IV regress 2 SLS otherwise stata endogenous will not work and then is that endogenous will work.

(Refer Slide Time: 24:33)



Now if you look at what is the null hypothesis here, variables are exogenous and we cannot reject the null also so there is no endogeneity problem. But the difficulty is here they are saying all variables are exogenous but which particular variable is exogenous whether one variable is exogenous or all the variables are exogenous that we cannot say because it is based on an F statistic which is saying that whether the variables are significant jointly or not.

When there are two endogenous variable, we will get two reduce form equation from the two reduced form equation we will get two predicted value of the error term let us say v 1 hat and v_2 hat and these two predicted value of there are terms we will put it in the structural equation. And then we will Implement one F test to test whether v 1 hat and v_2 hat they are jointly significant. And what is the alternative of this F test at least one among them is significant.

So, if it is at least one then I do not know whether that education is endogenous what experience is endogenous or both of them are endogenous that is the limitation of this test. To overcome these what we have to do we have to use another command we have to keep in mind, here again we have to use IV reg.

(Video Starts: 26:08)

So, instead of this IV reg and then Ivendog we have to install this command. This is not connected I think we cannot install also. Anyway, after estimating this model if you put Ivendog and you have to put which particular variable to use for endogeneity. We have two endogenous variable so after IV n log if you put education then that should work. Here I am not able to use this Ivendog because this command I am not able to download here since it is not connected.

So, here I am not able to connect this SSC install Ivendog. So, if it does not work nothing to bear it.

(Video Ends: 29:03)

So, what I am saying using this Ivendog command you can specify which particular variable you want to use for endogeneity. So, you can first use Ivendog education and see whether education is endogenous or not then after that you can use Ivendog experience to test whether experience is endogenous or not. So, this way you can actually test the endogeneity of a specific variable which was the limitation of the Hausman test. Since Hausman test uses the F test that is the problem.