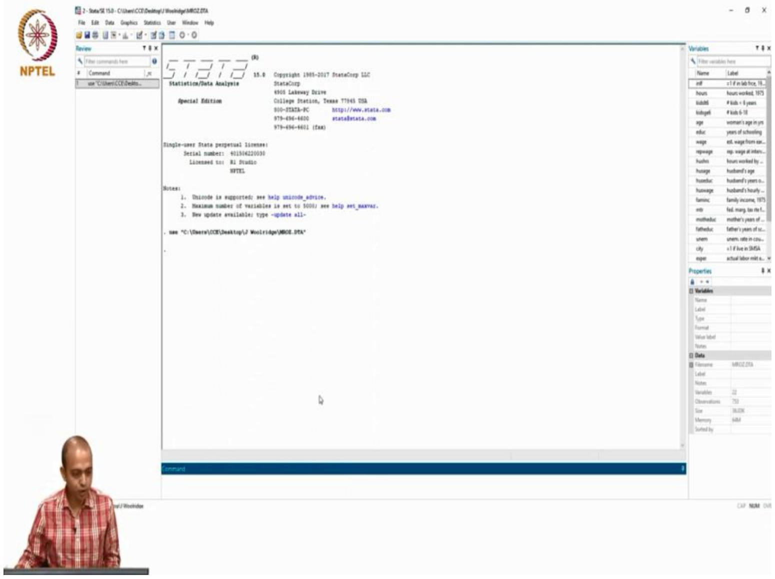
Applied Econometrics
Prof. Sabuj Kumar Mandal
Department of Humanities and Social Sciences
Indian Institute of Technology, Madras

Lecture - 08
Instrumental Variable Estimation – Part VIII

(Refer Slide Time: 00:15)



```
  _____  _____  _____  _____  _____   (R)
 /__    /__   /  ___/ /__   /  ___/
 ___/  /  /  /__   /  /  /  /___/    15.0   Copyright 1985-2017 StataCorp LLC
  Statistics/Data Analysis                  StataCorp
                                            4905 Lakeway Drive
     Special Edition                        College Station, Texas 77845 USA
                                            800-STATA-PC        http://www.stata.com
                                            979-696-4600        stata@stata.com
                                            979-696-4601 (fax)

Single-user Stata perpetual license:
       Serial number:  401506220030
         Licensed to:  Hi Studio
                       NPTEL

Notes:
      1.  Unicode is supported; see help unicode_advice.
      2.  Maximum number of variables is set to 5000; see help set_maxvar.
      3.  New update available; type -update all-

. use "C:\Users\CCE\Desktop\J Woolridge\MROZ.DTA"

.
```

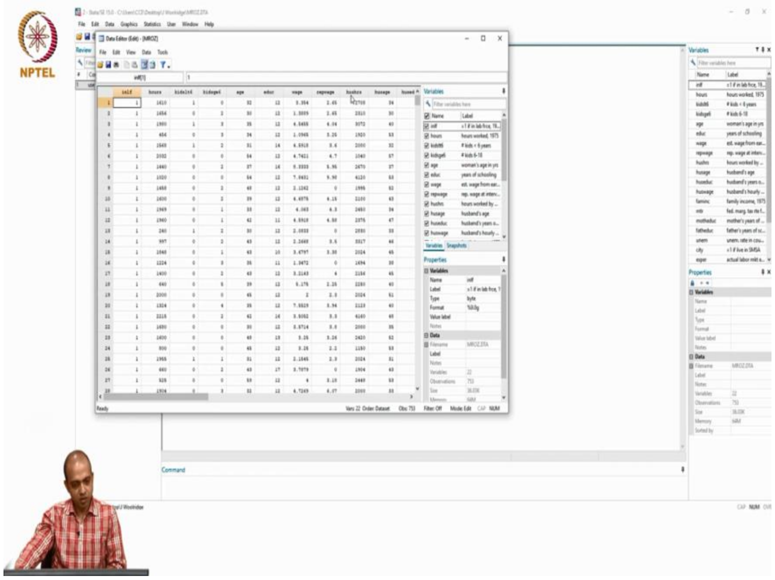
Name	Label
inlf	=1 if in lab frce, 19...
hours	hours worked, 1975
kidslt6	# kids < 6 years
kidsge6	# kids 6-18
age	woman's age in yrs
educ	years of schooling
wage	est. wage from ear...
repwage	rep. wage at interv...
hushrs	hours worked by ...
husage	husband's age
huseduc	husband's years o...
huswage	husband's hourly ...
faminc	family income, 1975
mtr	fed. marg. tax rte f...
motheduc	mother's years of ...
fatheduc	father's years of sc...
unem	unem. rate in cou...
city	=1 if live in SMSA
exper	actual labor mkt e...

Filename: MROZ.DTA

Once again to our discussion on instrumental variable estimation technique so far, we have discussed mainly the theoretical portion of this topic. So, by now we know what is instrumental variable, what is the instrumental variable estimation technique, when to apply I mean how to check endogeneity is there in the model or not, then how to extend the simple instrumental variable estimation in the context of multiple linear regression model where we have more than one explanatory variables.

And then how to test endogeneity when there are more number of endogenous variables, how to test over identification so on and so forth. Now today what we will do we will take one data set and then we will learn how to estimate all those models using the statistical software stator. So, this is the data what we are going to use this is a data on I will look I will show you the data source how does the data look like.

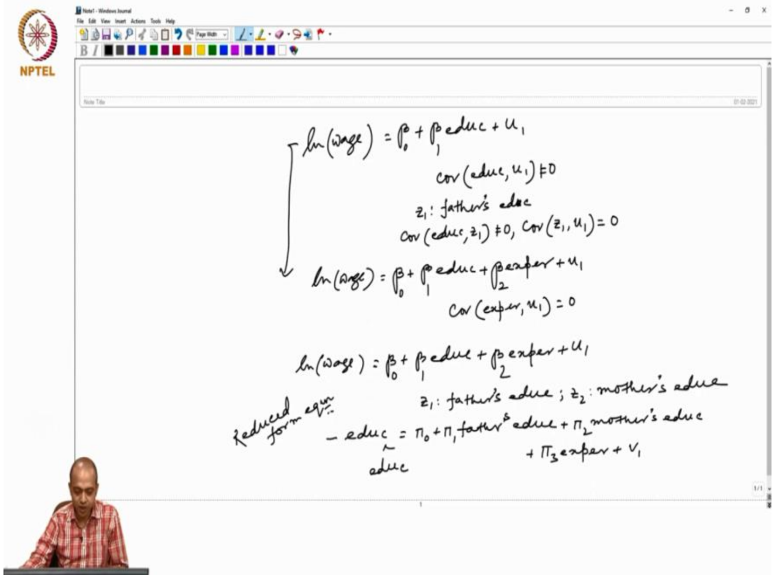
(Refer Slide Time: 01:36)



So, this is a data on married women's labour force participation where then we have this is a qualitative response kind of thing where in labour force is the dependent variable which takes the value one if the woman participate in the labour force otherwise no, otherwise 0. Then what is the labour supply measured by hours, then what is the number of kids less than six years of age, then age education, then wage, then husbands labour supply husband switch so on and so forth.

So, this is basically the data. Now what we will do we will try to estimate a wage function, the same example what we have discussed in our class.

(Refer Slide Time: 02:32)



$$\ln(wage) = \beta_0 + \beta_1 educ + u_1$$

$$cov(educ, u_1) \neq 0$$

z_1: father's educ

$$cov(educ, z_1) \neq 0, \; cov(z_1, u_1) = 0$$

$$\ln(wage) = \beta_0 + \beta_1 educ + \beta_2 exper + u_1$$

$$cov(exper, u_1) = 0$$

$$\ln(wage) = \beta_0 + \beta_1 educ + \beta_2 exper + u_1$$

z_1: father's educ ; z_2: mother's educ

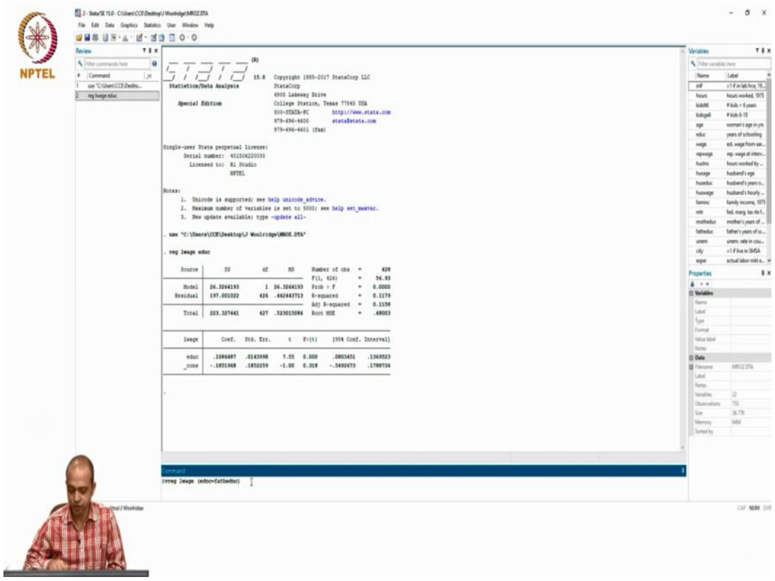
Reduced form eqn:

$$- \underset{\widehat{educ}}{educ} = \pi_0 + \pi_1 father's\ educ + \pi_2 mother's\ educ + \pi_3 exper + v_1$$

So, that means let us say that this is the wage function log of wage is actually a function of $\beta_0 + \beta_1\, education + u_1$ where we assume that covariance between education and u_1 that is actually not equals to 0. So, education is an endogenous variable why this is so because we are assuming that error term captures an omitted variable. Let us say ability where ability is correlated with the education and that is why there is a covariance between education and u_1.

And we also assume that we have two instruments let us say Z_1 which is father's education for the timing we will work with father's education. This is the instrument where we assume that covariance between education and Z_1 is actually not equals to 0 but covariance between Z_1 and u_1 that is actually equals to 0. So, this is our model and now we will estimate this using the statistical software.

(Refer Slide Time: 04:07)



2 - Stata/SE 15.0 - C:\Users\CCE\Desktop\J Woolridge\MROZ.DTA
File Edit Data Graphics Statistics User Window Help
Review
Command
1 use "C:\Users\CCE\Deskto...
2 reg lwage educ
Variables
Name
Label
Command
NPTEL

```
  ___  ____  ____  ____  ____ (R)
 /__    /   ____/   /   ____/
___/   /   /___/   /   /___/   15.0   Copyright 1985-2017 StataCorp LLC
  Statistics/Data Analysis            StataCorp
                                      4905 Lakeway Drive
     Special Edition                  College Station, Texas 77845 USA
                                      800-STATA-PC        http://www.stata.com
                                      979-696-4600        stata@stata.com
                                      979-696-4601 (fax)

Single-user Stata perpetual license:
       Serial number:  401506220030
         Licensed to:  R1 Studio
                       NPTEL

Notes:
      1.  Unicode is supported; see help unicode_advice.
      2.  Maximum number of variables is set to 5000; see help set_maxvar.
      3.  New update available; type -update all-

. use "C:\Users\CCE\Desktop\J Woolridge\MROZ.DTA"

. reg lwage educ

      Source |       SS           df       MS      Number of obs   =       428
-------------+----------------------------------   F(1, 426)       =     56.93
       Model |  26.3264193         1  26.3264193   Prob > F        =    0.0000
    Residual |  197.001022       426  .462443713   R-squared       =    0.1179
-------------+----------------------------------   Adj R-squared   =    0.1158
       Total |  223.327441       427  .523015084   Root MSE        =    .68003

------------------------------------------------------------------------------
       lwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        educ |   .1086487   .0143998     7.55   0.000     .0803451    .1369523
       _cons |  -.1851968   .1852259    -1.00   0.318    -.5492673    .1788736
------------------------------------------------------------------------------

.
```

```
ivreg lwage (educ=fathedu)
```

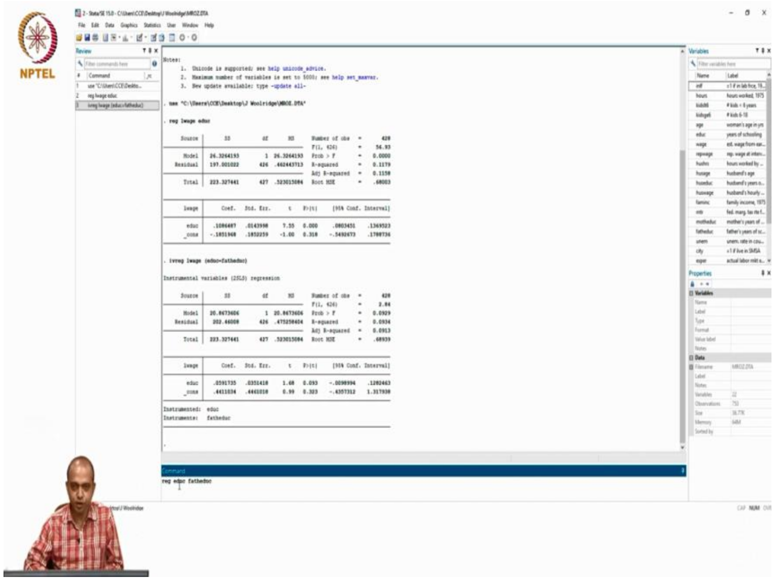
So, let us first try to estimate the model using ordinary list square technique and how to estimate the model, we all know. So, the command would be reg then l wage and education and if we this is the command in stata we have to put and if we put enter this is the model would, which is estimated. Now from this model, how to interpret the coefficient? See the coefficient of education is 0.1086 what does it mean how to interpret this coefficient and here the dependent variable is log wage.

So, that means we can say that for a unit change in education the wave changes by 10% that is what is the interpretation of this. So, that means returns to education is basically 10 around 10% 10.86% that is the interpretation. Now since we have estimated this model using ordinary list square and we suspect that education to be endogenous because ability factor is not included in the model which is there in the error term, so we suspect education to be endogenous.

And now let us estimate the model using the instrumental variable estimation technique. To estimate the model using instrumental variable estimation technique that means what is the command, the command is very simple the command is IV reg L wage the dependent variable then in the bracket we will put education and that will put equals to fatheduc. So, that means instrument we are putting in the bracket.

That means we are asking status to estimate the model by replacing education that means by using the father's education as an instrument for the endogenous variable.

(Refer Slide Time: 06:30)



```
Notes:
      1.  Unicode is supported; see help unicode_advice.
      2.  Maximum number of variables is set to 5000; see help set_maxvar.
      3.  New update available; type -update all-

. use "C:\Users\CCE\Desktop\J Woolridge\MROZ.DTA"

. reg lwage educ

      Source |       SS           df       MS      Number of obs   =       428
-------------+----------------------------------   F(1, 426)       =     56.93
       Model |  26.3264193         1  26.3264193   Prob > F        =    0.0000
    Residual |  197.001022       426  .462443713   R-squared       =    0.1179
-------------+----------------------------------   Adj R-squared   =    0.1158
       Total |  223.327441       427  .523015084   Root MSE        =    .68003

------------------------------------------------------------------------------
       lwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        educ |   .1086487   .0143998     7.55   0.000     .0803451    .1369523
       _cons |  -.1851968   .1852259    -1.00   0.318    -.5492673    .1788736
------------------------------------------------------------------------------

. ivreg lwage (educ=fatheduc)

Instrumental variables (2SLS) regression

      Source |       SS           df       MS      Number of obs   =       428
-------------+----------------------------------   F(1, 426)       =      2.84
       Model |  20.8673606         1  20.8673606   Prob > F        =    0.0929
    Residual |   202.46009       426  .475258404   R-squared       =    0.0934
-------------+----------------------------------   Adj R-squared   =    0.0913
       Total |  223.327441       427  .523015084   Root MSE        =    .68939

------------------------------------------------------------------------------
       lwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        educ |   .0591735   .0351418     1.68   0.093    -.0098994    .1282463
       _cons |   .4411034   .4461018     0.99   0.323    -.4357312    1.317938
------------------------------------------------------------------------------
Instrumented:  educ
Instruments:   fatheduc
------------------------------------------------------------------------------

.
```

```
reg educ fatheduc
```

Now look at this. So, in stata is reporting this result look at this result, the result is instrumental variable then in the bracket 2 SLS regression. Now if you recall we said that 2 SLS is a specific type of IV regression technique particularly applicable in the context of two in two instruments. So, in this case since we have only one instrument fathers' education so that means IV estimation which is done generally by method of moments which we have discussed and these two stages list square they will give you the same result.

So, that is why stata will always report the 2 SLS output only even in the context of one instrument that we have to keep in mind, so that means even though there is a distinction between IV and 2 SLS, 2 SLS is a specific case of IV which is applicable only when there are two instruments instead of one because why 2 SLS when we have two instruments we do not know which one to use.

And that is why econometrician this suggests that when you have more than one instrument, we should use the linear projection. So, that means linear combination of all those instruments as the instrument, so we have to take linear combination. How to take linear combination? In a second stage reduced form equation where the endogenous variable is actually regressed on all the exogenous variable included as well as excluded as instruments.

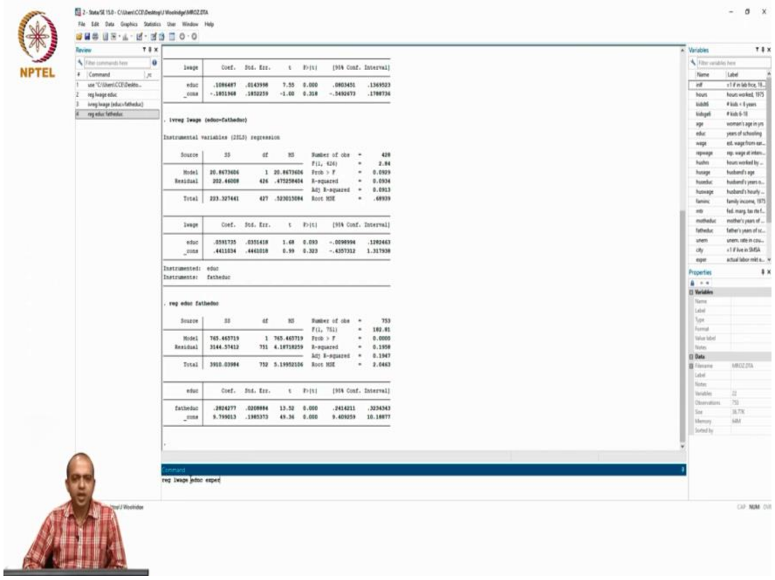
So, in this case stata is reporting two-stage list square technique where the dependent variable is log wage and independent variable is education but we have instrumented education by father's education. So, that means in this output stata is reporting what is instrumented and what are the instruments. What is instrumented is actually education that means education is the endogenous variable.

And father's education which is actually Z_1 excluded from the model is actually the instruments. Now if you look at the returns to education which was earlier 10.86 percent, now it is 5.91 percent so that means the returns to education has become half, so with this result we can now understand the severity of endogeneity problem. But one thing we have to keep in mind we do not know actually what is the true returns to equation whether 5 percent 5.291 percent is too low of true returns to education or not because we have collected only one sample.

And we are estimating returns to education from this sample but only one thing we have to keep in mind that there is a drastic change in the coefficient from 10.86 percent now it has become 5.91 percent that is the change in returns to education, so this is our first model. Now we have used father’s education as an instrument but as you know that instrument must exhibit two conditions, two conditions must be satisfied for the instruments to be a valid instrument.

So, that means the first condition is the instrument must be correlated with the endogenous variable and it should be known correlated with the error term. Now how will you check whether the instrument is actually correlated with the error term with the endogenous variable or not? Simply we will write one regression where education is the dependent variable and for others education is the independent one.

(Refer Slide Time: 10:50)



```
       lwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        educ |   .1086487   .0143998     7.55   0.000     .0803451    .1369523
       _cons |  -.1851968   .1852259    -1.00   0.318    -.5492673    .1788736
------------------------------------------------------------------------------

. ivreg lwage (educ=fatheduc)

Instrumental variables (2SLS) regression

      Source |       SS           df       MS      Number of obs   =       428
-------------+----------------------------------   F(1, 426)       =      2.84
       Model |  20.8673606         1  20.8673606   Prob > F        =    0.0929
    Residual |   202.46008       426  .475258404   R-squared       =    0.0934
-------------+----------------------------------   Adj R-squared   =    0.0913
       Total |  223.327441       427  .523015084   Root MSE        =    .68939

------------------------------------------------------------------------------
       lwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        educ |   .0591735   .0351418     1.68   0.093    -.0098994    .1282463
       _cons |   .4411034   .4461018     0.99   0.323    -.4357312    1.317938
------------------------------------------------------------------------------
Instrumented:  educ
Instruments:   fatheduc
------------------------------------------------------------------------------

. reg educ fatheduc

      Source |       SS           df       MS      Number of obs   =       753
-------------+----------------------------------   F(1, 751)       =    182.81
       Model |  765.463719         1  765.463719   Prob > F        =    0.0000
    Residual |  3144.37412       751  4.18718259   R-squared       =    0.1958
-------------+----------------------------------   Adj R-squared   =    0.1947
       Total |  3910.03984       752  5.19952106   Root MSE        =    2.0463

------------------------------------------------------------------------------
        educ |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    fatheduc |   .2824277   .0208884    13.52   0.000     .2414211    .3234343
       _cons |   9.799013   .1985373    49.36   0.000     9.409259    10.18877
------------------------------------------------------------------------------

.
```

```
reg lwage educ exper
```

And from this if you see that father's education is highly significant and R square of this model is around 20 percent. So, 20 percent of the variation in education can be explained by father's education itself. So, that means we can say that father's education is highly correlated with the individual's education, so the first condition is satisfied. What about the second condition? Second condition as we said when we have only one instrument, we cannot really check whether father's education is uncorrelated with the error term or not, that we cannot check.

So, for the timing we have to rely only our theoretical understanding that since father's education is not correlated with the wage of an individual that means father's education does not qualify to be an explanatory variable. Since it is not an explanatory variable, we will say that u i does not capture father's education. Because if at all it is captured then some relationship would be there with education sorry error term.

For example, ability was having some impact on father's individuals which that is why ability factor if I am not included in the model that means it is going and sitting inside the error term. Here we assume by our own common sense that somebody's wage is not dependent on father's education. But it is not though as simple as we are saying others education might also be an indirect impact on education individuals wage as well and in that context father’s education and error term would be related.

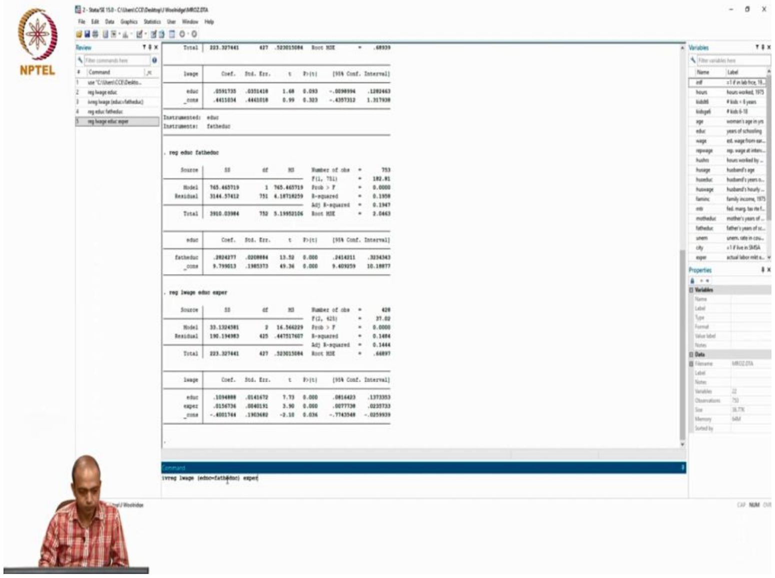
So, to make the things simple for the timing we say that father's education does not have any impact on individuals wage and that is the reason it is not going and sitting in the error term. So, we say that father's education is not correlated with the error term by our theoretical understanding but we cannot test this statistically using this data. So, this is how we have to estimate instrumental variable estimation.

And the command is IV reg l wage in the bracket we will first put the endogenous variable we will put equality sign and then this. So, that means once you estimate this box instrumented and instruments will give will help you identify what is the reduced form you have estimated. What is the reduced form equation? That means education is actually the dependent variable in the reduced form equation and father's education in the independent world.

Now we will simply extend this model when we have multiple explanatory variable. So, that means now we are estimating this type of model reg l wage education and let us say we have experience also as an additional explanatory variable but experience we are assuming that it is an exogenous variable. So, that means we are extending this model form $lwage = \beta_0 + \beta_1\, education + \beta_2\, experience + u_1$ where we assume that covariance between experience and the error term is actually 0.

So, that means experience is a purely exogenous variable included in the model. But one thing is we have to keep in mind that since experience is already included in the model as an explanatory variable experience cannot be used as an instrument. So, an exogenous variable which is included in the model cannot be used as an instrument, the instrument must be excluded from the model. So, we need to look for some other variable like father's education, so this is the model we are estimating.

(Refer Slide Time: 15:42)



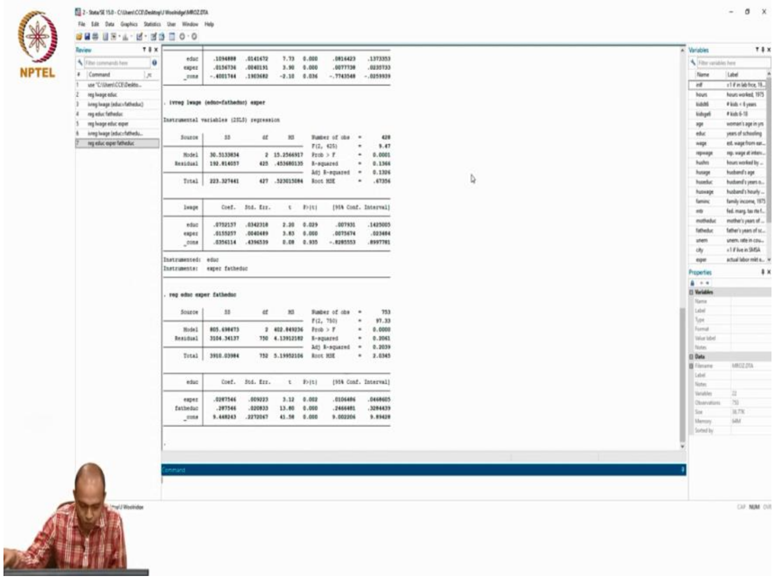
```
       Total   223.327441       427  .523015084   Root MSE        =    .68939

       lwage        Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]

        educ     .0591735   .0351418     1.68   0.093    -.0098994    .1282463
       _cons     .4411034   .4461018     0.99   0.323    -.4357312    1.317938

Instrumented:  educ
Instruments:   fatheduc

. reg educ fatheduc

      Source         SS           df       MS      Number of obs   =       753
                                                   F(1, 751)       =    182.81
       Model   765.465719         1  765.465719   Prob > F        =    0.0000
    Residual   3144.57412       751  4.18718259   R-squared       =    0.1958
                                                   Adj R-squared   =    0.1947
       Total   3910.03984       752  5.19952106   Root MSE        =    2.0463

        educ        Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]

    fatheduc     .2824277   .0208884    13.52   0.000     .2414211    .3234343
       _cons     9.799013   .1985373    49.36   0.000     9.409259    10.18877

. reg lwage educ exper

      Source         SS           df       MS      Number of obs   =       428
                                                   F(2, 425)       =     37.02
       Model   33.1324581         2  16.566229    Prob > F        =    0.0000
    Residual   190.194983       425  .447517607   R-squared       =    0.1484
                                                   Adj R-squared   =    0.1444
       Total   223.327441       427  .523015084   Root MSE        =    .66897

       lwage        Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]

        educ     .1094888   .0141672     7.73   0.000     .0816423    .1373353
       exper     .0156736   .0040191     3.90   0.000     .0077738    .0235733
       _cons    -.4001744   .1903682    -2.10   0.036    -.7743548   -.0259939

.
```

```
ivreg lwage (educ=fatheduc) exper
```

And then so here this is the model we are estimating so returns to education is again 10.94 percent when we are using OLS but when you are using instrumental variable IV reg l wage and then we will put education equals to father's education and then experience. So, now returns to education has gone down from 10.94 percent to 7.52 percent and again if you want to trace your reduced form equation this education = β_0 + β_1 experience + β_2 father’s education.

So, that means reduced form equation as I said earlier it is a function of the endogenous variable with all exogenous variable included as well as excluded experiences included exogenous variable and father's education is the excluded exogenous period. And once again if you want to test how best the father's education is related reg education with experience with father's education.

(Refer Slide Time: 17:07)



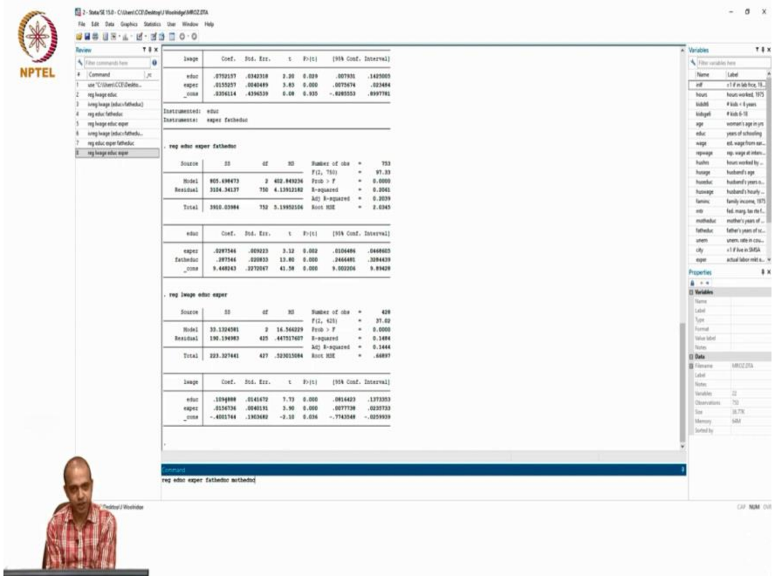
```
        educ |   .1094888   .0141472     7.73   0.000     .0816423    .1373353
       exper |   .0156736   .0040191     3.90   0.000     .0077738    .0235733
       _cons |  -.4001744   .1903682    -2.10   0.036    -.7743548   -.0259939
------------------------------------------------------------------------------

. ivreg lwage (educ=fathedu) exper

Instrumental variables (2SLS) regression

      Source |       SS           df       MS      Number of obs   =       428
-------------+----------------------------------   F(2, 425)       =      9.47
       Model |  30.5133834         2  15.2566917   Prob > F        =    0.0001
    Residual |  192.814057       425  .453680135   R-squared       =    0.1366
-------------+----------------------------------   Adj R-squared   =    0.1326
       Total |  223.327441       427  .523015084   Root MSE        =    .67356

------------------------------------------------------------------------------
       lwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        educ |   .0752137   .0342318     2.20   0.029      .007931    .1425005
       exper |   .0155257   .0040489     3.83   0.000     .0075674     .023484
       _cons |   .0356114   .4396539     0.08   0.935    -.8285553    .8997781
------------------------------------------------------------------------------
Instrumented:  educ
Instruments:   exper fathedu
------------------------------------------------------------------------------

. reg educ exper fathedu

      Source |       SS           df       MS      Number of obs   =       753
-------------+----------------------------------   F(2, 750)       =     97.33
       Model |  805.698473         2  402.849236   Prob > F        =    0.0000
    Residual |  3104.34137       750  4.13912182   R-squared       =    0.2061
-------------+----------------------------------   Adj R-squared   =    0.2039
       Total |  3910.03984       752  5.19952106   Root MSE        =    2.0345

------------------------------------------------------------------------------
        educ |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       exper |   .0287546    .009223     3.12   0.002     .0106486    .0468605
     fathedu |    .287546    .020833    13.80   0.000     .2466481    .3284439
       _cons |   9.448243   .2272067    41.58   0.000     9.002206     9.89428
------------------------------------------------------------------------------
```

See father's education is again highly correlated with the education because the P value is 0.00. So, that means we can extend the simple IV estimation technique in the context of multiple linear regression model as well. Now let us assume that we will estimate a model where we have more than one instrument that means we are let us say running this type of model log wage = β_0 + β_1 education + β_2 experience + u_1 and we have two instruments.

Let us say Z_1 which is father's education and Z_2 which is let us say mother's education. So, when we have two instruments then what we have this discussed in our class that instead of using one instrument we have to take a linear combination of these two. And how to take linear combination of these two? It is basically running a reduced form equation that means education equals to let us say $\pi_0 + \pi_1\, father's\ education\ + \pi_2\, mother's\ education\ +\ \pi_3\, experience\ +\ \ v_1$.

This is the model, so that means the estimation involves firstly estimating this reduced form equation this is the reduced form equation. From here what I have to get? I have to get the education hat, estimated value of this. And then I need to substitute education by education hat to remove the endogeneity, that is what we have discussed. Now let us see how to go about this.

(Refer Slide Time: 20:08)



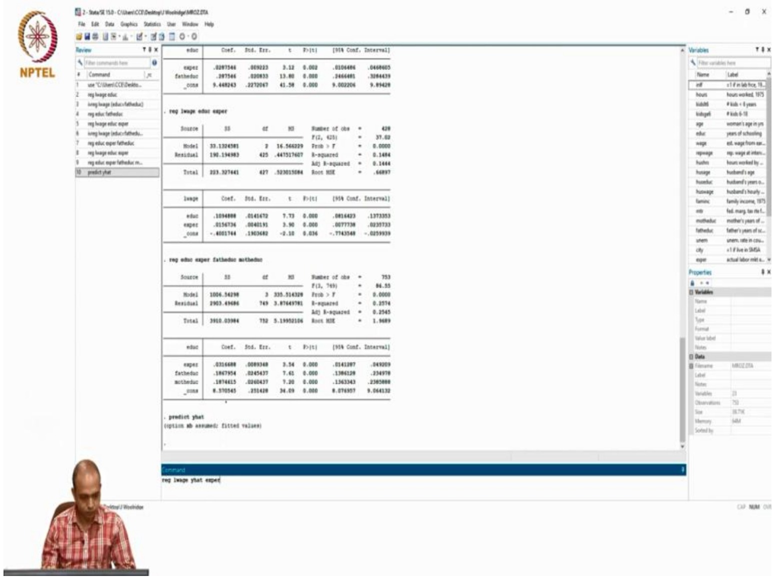
```
       lwage │      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
─────────────┼────────────────────────────────────────────────────────────────
        educ │   .0752157   .0342318     2.20   0.029      .007931    .1425005
       exper │   .0155257   .0040489     3.83   0.000     .0075674     .023484
       _cons │   .0356114   .4396539     0.08   0.935    -.8285553    .8997781
──────────────────────────────────────────────────────────────────────────────
Instrumented:  educ
Instruments:   exper fatheduc
──────────────────────────────────────────────────────────────────────────────

. reg educ exper fatheduc

      Source │       SS           df       MS      Number of obs   =       753
─────────────┼──────────────────────────────────   F(2, 750)       =     97.33
       Model │  805.698473         2  402.849236   Prob > F        =    0.0000
    Residual │  3104.34137       750  4.13912182   R-squared       =    0.2061
─────────────┼──────────────────────────────────   Adj R-squared   =    0.2039
       Total │  3910.03984       752  5.19952106   Root MSE        =    2.0345

──────────────────────────────────────────────────────────────────────────────
        educ │      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
─────────────┼────────────────────────────────────────────────────────────────
       exper │   .0287546    .009223     3.12   0.002     .0106486    .0468605
    fatheduc │    .287546    .020833    13.80   0.000     .2466481    .3284439
       _cons │   9.448243   .2272067    41.58   0.000     9.002206     9.89428
──────────────────────────────────────────────────────────────────────────────

. reg lwage educ exper

      Source │       SS           df       MS      Number of obs   =       428
─────────────┼──────────────────────────────────   F(2, 425)       =     37.02
       Model │  33.1324581         2  16.566229    Prob > F        =    0.0000
    Residual │  190.194983       425  .447517607   R-squared       =    0.1484
─────────────┼──────────────────────────────────   Adj R-squared   =    0.1444
       Total │  223.327441       427  .523015084   Root MSE        =    .66897

──────────────────────────────────────────────────────────────────────────────
       lwage │      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
─────────────┼────────────────────────────────────────────────────────────────
        educ │   .1094888   .0141672     7.73   0.000     .0816423    .1373353
       exper │   .0156736   .0040191     3.90   0.000     .0077738    .0235733
       _cons │  -.4001744   .1903682    -2.10   0.036    -.7743548   -.0259939
──────────────────────────────────────────────────────────────────────────────

.
```

```
reg educ exper fatheduc motheduc
```

So, that means we will estimate the model, let us say here this is our model reg log education and we have experience, this is our model, we have one endogenous variable, one exogenous variable this is the regression. And from here we can understand that returns to education is 10.94 when we estimate the model using OLS without considering the endogeneity. And now since we have two instruments what to do now in the first step what will do reg education equals to experience then father's education and mother's education.

This is the reduced form equation we are running. So, we will estimate education hat from this regression. Education is a function of experience father's education and mother's education.

(Refer Slide Time: 21:23)



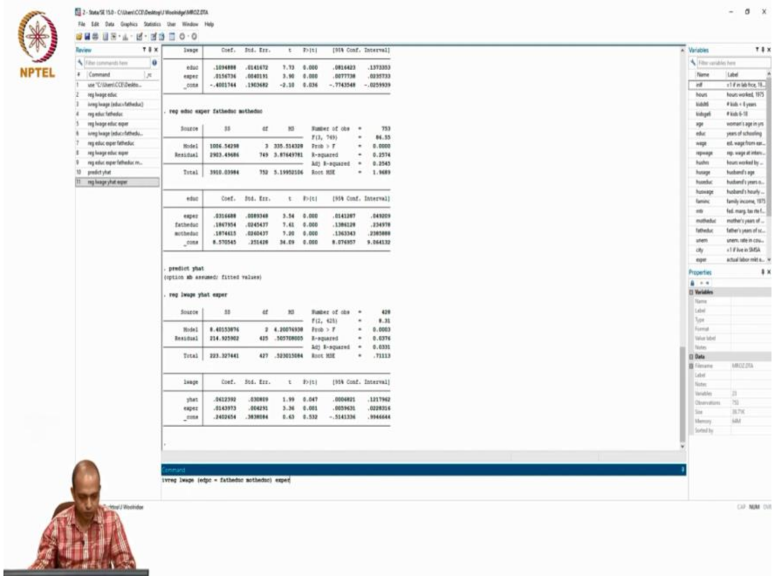
```
        educ │      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
─────────────┼────────────────────────────────────────────────────────────────
       exper │   .0287546    .009223     3.12   0.002     .0106486    .0468605
    fatheduc │    .287546    .020833    13.80   0.000     .2466481    .3284439
       _cons │   9.448243   .2272067    41.58   0.000     9.002206     9.89428
──────────────────────────────────────────────────────────────────────────────

. reg lwage educ exper

      Source │       SS           df       MS      Number of obs   =       428
─────────────┼──────────────────────────────────   F(2, 425)       =     37.02
       Model │  33.1324581         2  16.566229    Prob > F        =    0.0000
    Residual │  190.194983       425  .447517607   R-squared       =    0.1484
─────────────┼──────────────────────────────────   Adj R-squared   =    0.1444
       Total │  223.327441       427  .523015084   Root MSE        =    .66897

──────────────────────────────────────────────────────────────────────────────
       lwage │      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
─────────────┼────────────────────────────────────────────────────────────────
        educ │   .1094888   .0141672     7.73   0.000     .0816423    .1373353
       exper │   .0156736   .0040191     3.90   0.000     .0077738    .0235733
       _cons │  -.4001744   .1903682    -2.10   0.036    -.7743548   -.0259939
──────────────────────────────────────────────────────────────────────────────

. reg educ exper fatheduc motheduc

      Source │       SS           df       MS      Number of obs   =       753
─────────────┼──────────────────────────────────   F(3, 749)       =     86.55
       Model │  1006.54298         3  335.514328   Prob > F        =    0.0000
    Residual │  2903.49686       749  3.87649781   R-squared       =    0.2574
─────────────┼──────────────────────────────────   Adj R-squared   =    0.2545
       Total │  3910.03984       752  5.19952106   Root MSE        =    1.9689

──────────────────────────────────────────────────────────────────────────────
        educ │      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
─────────────┼────────────────────────────────────────────────────────────────
       exper │   .0316688   .0089348     3.54   0.000     .0141287     .049209
    fatheduc │   .1867954   .0245437     7.61   0.000     .1386128     .234978
    motheduc │   .1874615   .0260437     7.20   0.000     .1363343    .2385888
       _cons │   8.370545    .251428    34.09   0.000     8.076957    9.064132
──────────────────────────────────────────────────────────────────────────────

. predict yhat
(option xb assumed; fitted values)

.
```

```
reg lwage yhat exper
```

And from these what you have to get? We have to get the predicted value of the dependent variable which is education. So, what is the command for that predict y hat or you can put education hat as well y hat, if you put then stata will predict the estimated value of the dependent variable from the reduced form equation and that reduced form equal that estimated value from the reduced form equation we will replace into the original structural form equation.

So, if we do that our the equation would become reg lwage now instead of education we are putting y hat then experience that is the only additional explanatory variable we have.

(Refer Slide Time: 22:43)



2 - Stata/SE 15.0 - C:\Users\CCE\Desktop\J Wooldridge\MROZ.DTA
File Edit Data Graphics Statistics User Window Help
Review
Filter commands here
NPTEL
Command

```
1   use "C:\Users\CCE\Deskto...
2   reg lwage educ
3   ivreg lwage (educ=fatheduc)
4   reg educ fatheduc
5   reg lwage educ exper
6   ivreg lwage (educ=fathedu...
7   reg educ exper fatheduc
8   reg lwage educ exper
9   reg educ exper fatheduc m...
10  predict yhat
11  reg lwage yhat exper
```

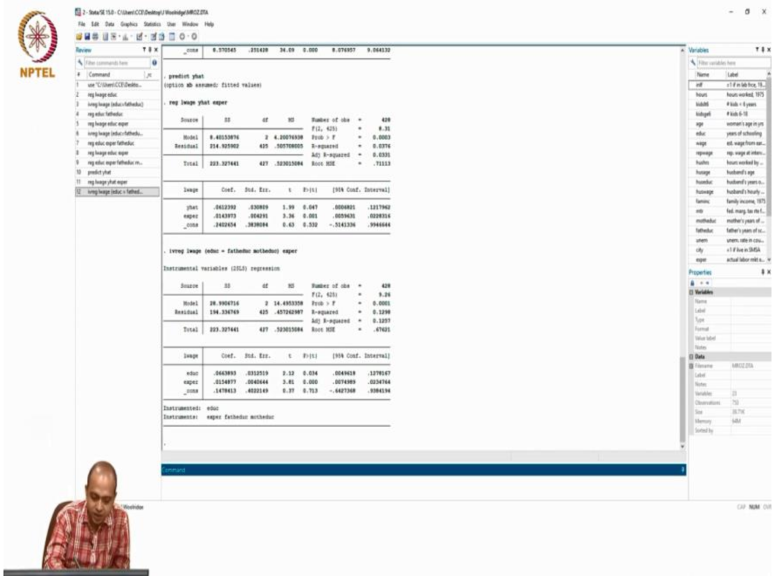
```
       lwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        educ |   .1094888   .0141672     7.73   0.000     .0816423    .1373353
       exper |   .0156736   .0040191     3.90   0.000     .0077738    .0235733
       _cons |  -.4001744   .1903682    -2.10   0.036    -.7743548   -.0259939
------------------------------------------------------------------------------

. reg educ exper fatheduc motheduc

      Source |       SS           df       MS      Number of obs   =       753
-------------+----------------------------------   F(3, 749)       =     86.55
       Model |  1006.54298         3  335.514328   Prob > F        =    0.0000
    Residual |  2903.49686       749  3.87649781   R-squared       =    0.2574
-------------+----------------------------------   Adj R-squared   =    0.2545
       Total |  3910.03984       752  5.19950106   Root MSE        =    1.9689

------------------------------------------------------------------------------
        educ |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       exper |   .0316688   .0089348     3.54   0.000     .0141287     .049209
    fatheduc |   .1867954   .0245437     7.61   0.000     .1386128     .234978
    motheduc |   .1874615   .0260437     7.20   0.000     .1363343    .2385888
       _cons |   8.570545    .251428    34.09   0.000     8.076957    9.064132
------------------------------------------------------------------------------

. predict yhat
(option xb assumed; fitted values)

. reg lwage yhat exper

      Source |       SS           df       MS      Number of obs   =       428
-------------+----------------------------------   F(2, 425)       =      8.31
       Model |  8.40153876         2  4.20076938   Prob > F        =    0.0003
    Residual |  214.925902       425  .505708005   R-squared       =    0.0376
-------------+----------------------------------   Adj R-squared   =    0.0331
       Total |  223.327441       427  .523015084   Root MSE        =    .71113

------------------------------------------------------------------------------
       lwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        yhat |   .0612392    .030809     1.99   0.047     .0006821    .1217962
       exper |   .0143973    .004291     3.36   0.001     .0059631    .0228316
       _cons |   .2402654   .3838084     0.63   0.532    -.5141336    .9946644
------------------------------------------------------------------------------

.
```

```
ivreg lwage (edpc = fatheduc motheduc) exper
```

And now this y hat is nothing but education hat, so the returns to education is 6.12 percent. So, this is what we did? We have basically estimated the model using 2 SLS but we conducted 2 SLS manually. In the first step we have estimated the value of education the endogenous variable we have predicted the value of the endogenous variable and then we replace education by education hat to remove the endogeneity.

Now we have to note down this is 6.12 percent when we run the model by conducting 2 SLS manually. Now if I regress the model by IV regression, let us see what is happening so IV reg l wage, then for education I am putting education equals to father's education mother's education also and then experience.

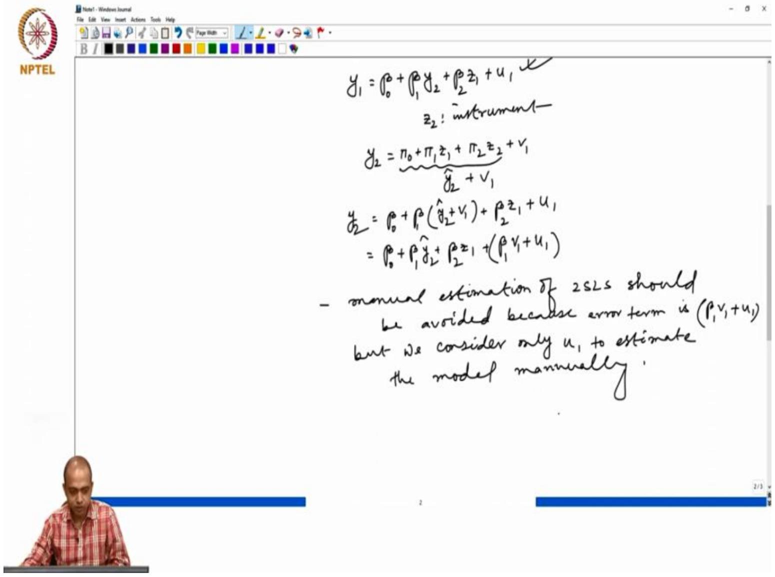
(Refer Slide Time: 24:25)



```
      _cons |   8.570545    .251428    34.09   0.000     8.076957    9.064132
------------------------------------------------------------------------------

. predict yhat
(option xb assumed; fitted values)

. reg lwage yhat exper

      Source |       SS           df       MS      Number of obs   =       428
-------------+----------------------------------   F(2, 425)       =      8.31
       Model |  8.40153876         2  4.20076938   Prob > F        =    0.0003
    Residual |  214.925902       425  .505708005   R-squared       =    0.0376
-------------+----------------------------------   Adj R-squared   =    0.0331
       Total |  223.327441       427  .523015084   Root MSE        =    .71113

------------------------------------------------------------------------------
       lwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        yhat |   .0612392    .030809     1.99   0.047     .0006821    .1217962
       exper |   .0143973    .004291     3.36   0.001     .0059631    .0228316
       _cons |   .2402654   .3830084     0.63   0.532    -.5141336    .9946644
------------------------------------------------------------------------------

. ivreg lwage (educ = fatheduc motheduc) exper

Instrumental variables (2SLS) regression

      Source |       SS           df       MS      Number of obs   =       428
-------------+----------------------------------   F(2, 425)       =      9.26
       Model |  28.9906716         2  14.4953358   Prob > F        =    0.0001
    Residual |  194.336769       425  .457262987   R-squared       =    0.1298
-------------+----------------------------------   Adj R-squared   =    0.1257
       Total |  223.327441       427  .523015084   Root MSE        =    .67621

------------------------------------------------------------------------------
       lwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        educ |   .0663893   .0312519     2.12   0.034     .0049618    .1278167
       exper |   .0154877   .0040644     3.81   0.000     .0074989    .0234764
       _cons |   .1478413   .4022149     0.37   0.713    -.6427368    .9384194
------------------------------------------------------------------------------
Instrumented:  educ
Instruments:   exper fatheduc motheduc
------------------------------------------------------------------------------
```

So, this is the running 2 SLS estimates using the statistical software. Here what is the return education? 6.63 it was 6.12 and the standard error 0.030 this is 0.0312. So, that means there is some kind of change in the coefficient as well as in the standard error. Now the question is which among these two manual or this IV reg command to be used for estimation. Now econometrician they suggest that we should actually avoid the manual estimation of 2 SLS because of the following reason.

(Refer Slide Time: 25:12)



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

z_2 : instrument

$y_2 = \underbrace{\pi_0 + \pi_1 z_1 + \pi_2 z_2}_{\hat{y}_2} + v_1$

$\hat{y}_2 + v_1$

$y_2 = \beta_0 + \beta_1(\hat{y}_2 + v_1) + \beta_2 z_1 + u_1$

$= \beta_0 + \beta_1 \hat{y}_2 + \beta_2 z_1 + (\beta_1 v_1 + u_1)$

- manual estimation of 2SLS should be avoided because error term is $(\beta_1 v_1 + u_1)$ but we consider only u_1 to estimate the model mannually.

What is the reason? Here this is our model y 1 = β_0 + β_1 y_2+ β_2 Z_1 + u_1 this is the model and Z_2 . Let us say is the instrument so in the reduced form equation what we do y_2 we run this regression $\pi_0 + \pi_1 Z_1 + \pi_2 Z_2 + v_1$. And we said this is basically that means y_2 equals to this is $\widehat{y_2} + v_1$.

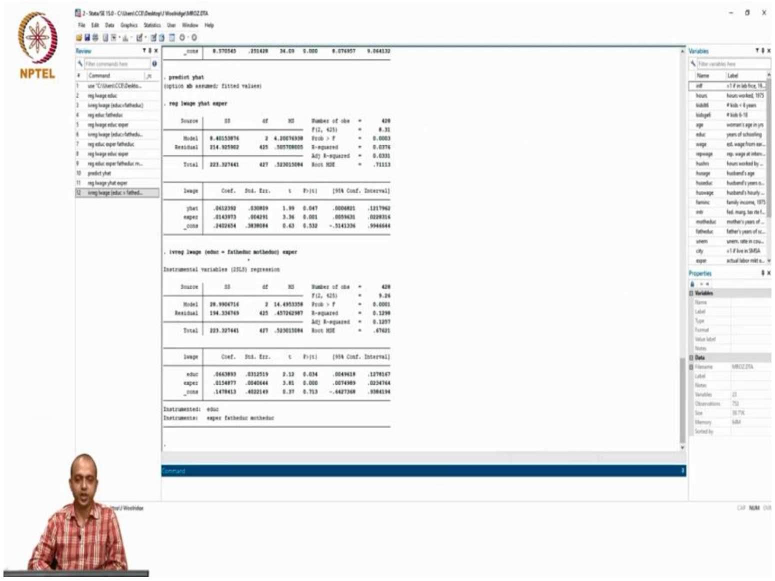
Now if we replace this into the original equation then what will happen y 1 will become β_0 + β_1 $\widehat{y_2}$ + v_1 + β_2 Z_1 + u_1.

That means this would become β_0 + β_1 $\widehat{y_2}$+ β_2 Z_1 + β_1 v_1 + u_1. So, that means in the original structural form equation the error term becomes a composite one with additional component v_1 v_1. This is actual error term in the structural equation but when we are manually estimating this, we are only replacing $\widehat{y_2}$ here and trying to calculate the standard error of this β_1 hat by using this error v_1.

Whereas the actual error term should be u_1 + β_1 v_1. Please try to understand what I am saying manual so basically, I am writing here manual estimation of 2 SLS should be avoided because error term is β_1 v_1 + u_1 but we consider only u_1 to estimate the model manually. So, the original estimates is actually so the original error term is β_1 v_1 + u_1 which is a composite in nature but when we replace the estimated value of y_2, this is y_2 not y 1 y_2 we take only the first component.

So, that means only this much $\pi_0 + \pi_1 Z_1 + \pi_2 Z_2$ and we ignore the v_1 hat version that is why this manual thing is not suggested we have to always go by this.

(Refer Slide Time: 29:48)



```
       _cons    8.570543    .251428    34.09   0.000     8.076957    9.064132

. predict yhat
(option xb assumed; fitted values)

. reg lwage yhat exper

      Source |       SS           df       MS      Number of obs   =       428
-------------+----------------------------------   F(2, 425)       =      8.31
       Model |  8.40153876         2  4.20076938   Prob > F        =    0.0003
    Residual |  214.925902       425  .505708005   R-squared       =    0.0376
-------------+----------------------------------   Adj R-squared   =    0.0331
       Total |  223.327441       427  .523015084   Root MSE        =    .71113

------------------------------------------------------------------------------
       lwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        yhat |   .0612392    .030809     1.99   0.047     .0006821    .1217962
       exper |   .0143973    .004291     3.36   0.001     .0059631    .0228316
       _cons |   .2402654   .3839084     0.63   0.532    -.5141336    .9946644
------------------------------------------------------------------------------

. ivreg lwage (educ = fatheduc motheduc) exper

Instrumental variables (2SLS) regression

      Source |       SS           df       MS      Number of obs   =       428
-------------+----------------------------------   F(2, 425)       =      9.26
       Model |  28.9904716         2  14.4953358   Prob > F        =    0.0001
    Residual |  194.336769       425  .457262987   R-squared       =    0.1298
-------------+----------------------------------   Adj R-squared   =    0.1257
       Total |  223.327441       427  .523015084   Root MSE        =    .67621

------------------------------------------------------------------------------
       lwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        educ |   .0663893   .0312519     2.12   0.034     .0049618    .1278167
       exper |   .0154877   .0040644     3.81   0.000     .0074989    .0234764
       _cons |   .1478413   .4022149     0.37   0.713    -.6427368    .9384194
------------------------------------------------------------------------------
Instrumented:  educ
Instruments:   exper fatheduc motheduc
------------------------------------------------------------------------------
```

We have to always go by this IV reg method only that is what I wanted to make.