Spring 2023 Gregory A. Bruich, Ph.D.
Economics 1123 Harvard University

Section 8: IV Regression with Multiple Instruments and Intrinsic Heterogeneity

Estimating Returns to School w/ Two Instruments

Recall last week's IV example that studied returns to schooling through using distance to one's closest 4-year college as an instrument for education. Further nuance could come from including a second instrument of the distance to one's closest 2-year college.

```
. ivregress 2sls lwage (educ=nearc2 nearc4) exper expersq black south, r

Instrumental variables (2SLS) regression          Number of obs   =      3,010
                                                  Wald chi2(5)    =     405.49
                                                  Prob > chi2     =     0.0000
                                                  R-squared       =          .
                                                  Root MSE        =     .49548

------------------------------------------------------------------------------
             |               Robust
       lwage |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        educ |   .2403154   .0402514     5.97   0.000      .161424    .3192067
       exper |   .1517071   .0187201     8.10   0.000     .1150163    .1883978
     expersq |  -.0024099   .0004516    -5.34   0.000     -.003295   -.0015247
       black |  -.0182842   .0451119    -0.41   0.685     -.106702    .0701335
       south |  -.0807446   .0259706    -3.11   0.002     -.131646   -.0298432
       _cons |   1.998076   .6951721     2.87   0.004     .6355635    3.360588
------------------------------------------------------------------------------
Instrumented:  educ
Instruments:   exper expersq black south nearc2 nearc4
```

First Stage F^{MOP} statistic

```
. ssc install weakivtest
. weakivtest
(obs=3,010)

Montiel-Pflueger robust weak instrument test
--------------------------------------------
Effective F statistic:       20.103
Confidence level alpha:          5%
--------------------------------------------

--------------------------------------------
```

```
Critical Values                   TSLS      LIML
--------------------------------------------
% of Worst Case Bias
tau=5%                           4.445    19.766
tau=10%                          3.768    12.452
tau=20%                          3.396     8.362
tau=30%                          3.266     6.839
--------------------------------------------

.
. display "F MOP Critical Value = " r(c_TSLS_10)
F MOP Critical Value = 3.767954

. display "First Stage F MOP  statistic = "  r(F_eff)
First Stage F MOP  statistic = 20.103195
```

- What is the HR first stage F statistic?

- What conclusion can be drawn about relevance from this test?

J statistic

Terminology: with more instruments (two: nearc2 nearc4) than endogenous regressors (one: educ), we are "overidentified." In that case, we can calculate a heteroskedasticity robust (HR) or heteroskedasticity and autocorrelation robust (HAC) J -statistic known as a test of the overidentifying restrictions.

```
. estat overid, forcenonrobust

  Tests of overidentifying restrictions:

  Sargan chi2(1)          =   1.8588  (p = 0.1728)
  Basmann chi2(1)         =  1.85563  (p = 0.1731)
  Score chi2(1)           =  1.86471  (p = 0.1721)

. dis "HR J-test = " r(score) " HR p-value = " r(p_score)
HR J-test = 1.8647102 HR p-value = .17208216
```

```
. dis "Homoskedastic J-test = " r(basmann) " Homoskedastic p-value = " r(p_basmann)
Homoskedastic J-test = 1.8556251 Homoskedastic p-value = .17313058
```

- What is the null hypothesis of the J-test?

- What conclusion can be drawn about exogeneity from this test?

- What should we do if we reject the null? Are there any exceptions to this?

Intrinsic Heterogeneity

So far, we have thought about *constant* treatment effects. That is, for the regression function: $Y_i = \beta_0 + \beta_1 X_i + u_i$, we have assumed that $\beta_{1i} = Y_i(1) - Y_i(0)$ is the same for ALL individuals. Now, we no longer assume that - β_1 may vary by i - that is, X_i has a different effect on Y_i on different individuals and so we have:

$$Y_i = \beta_0 + \beta_{1i} X_i + u_i$$

Example: The return of schooling on income -

$$Income_i = \beta_{0i} + \beta_{1i} Education_i + u_i$$

- Why might individuals have different returns to education?

OLS and Average Treatment Effects

- If there are heterogenous treatment effects, X is randomly assigned, and we run OLS, we estimate the *Average Treatment Effect* (ATE): $E(\beta_{1i}) = E[Y_i(1) - Y_i(0)]$. How would we interpret $\hat{\beta}_1$ if Education is randomly assigned and we run OLS?

Introduction to LATE

- If X is not randomly assigned, and we decide to use IV, what happens?

In most cases we will no longer estimate the ATE and estimate the local average treatment effect ($LATE$) which in terms of math is:

$$LATE = \frac{E[\beta_{1i}\pi_{1i}]}{E[\pi_{1i}]}$$

where π_{1i} is the coefficient on the instrument from the first stage:

$$X_i = \pi_{0i} + \pi_{1i}Z_i + v_i.$$

What are the different terms? What do they mean?

In general, the $LATE$ tells us that the average treatment effect for the people who are *affected by the instrument* (sometimes these are called the "compliers").

What does it mean that people are "affected" by the instrument? What does it mean in terms of the first stage?

Example: Returns to Schooling

Let's go back to the schooling example before, where we decided to use instruments for schooling that exploit geographical differences in accessibility of college, specifically, the distance to one's closest two/four year college.

- Who do we think will be affected by the instruments?

- How do we think their effect of the treatment (i.e. schooling) is different than the general population?

Let's contrast OLS and 2SLS:

```
. ivregress 2sls lwage (educ=nearc2 nearc4) exper expersq black south, r

Instrumental variables (2SLS) regression          Number of obs   =      3,010
                                                  Wald chi2(5)    =     405.49
                                                  Prob > chi2     =     0.0000
                                                  R-squared       =          .
                                                  Root MSE        =     .49548

------------------------------------------------------------------------------
             |               Robust
       lwage |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        educ |   .2403154   .0402514     5.97   0.000      .161424    .3192067
       exper |   .1517071   .0187201     8.10   0.000     .1150163    .1883978
     expersq |  -.0024099   .0004516    -5.34   0.000     -.003295   -.0015247
       black |  -.0182842   .0451119    -0.41   0.685     -.106702    .0701335
       south |  -.0807446   .0259706    -3.11   0.002     -.131646   -.0298432
       _cons |   1.998076   .6951721     2.87   0.004     .6355635    3.360588
------------------------------------------------------------------------------
Instrumented:  educ
Instruments:   exper expersq black south nearc2 nearc4

. regress lwage educ exper expersq black south, r

Linear regression                               Number of obs     =      3,010
                                                F(5, 3004)        =     222.70
                                                Prob > F          =     0.0000
```

```
                                               R-squared          =     0.2651
                                               Root MSE           =     .38076

------------------------------------------------------------------------------
             |               Robust
       lwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        educ |    .078233   .0036886    21.21   0.000     .0710005    .0854655
       exper |   .0851268   .0068201    12.48   0.000     .0717543    .0984993
     expersq |  -.0023404    .000322    -7.27   0.000    -.0029718    -.001709
       black |  -.1780477    .017711   -10.05   0.000    -.2127745   -.1433209
       south |   -.150492   .0153809    -9.78   0.000    -.1806501   -.1203339
       _cons |   4.796325   .0716322    66.96   0.000     4.655872    4.936778
------------------------------------------------------------------------------
```

Last week we gave the explanation that OLS was *downward biased* since the IV $\hat{\beta}^{IV}$ *increased* compared to the OLS.

- What is another explanation?

When does $LATE = ATE$? There are three un-realistic conditions that are not likely satisfied in most applications which lead to $LATE = ATE$:

1. $\beta_{1i} = \beta_1$ for all i - that is no heterogeneity in the effect of the treatment on the outcome OR
2. $\pi_{1i} = \pi_1$ for all i – there is no heterogeneity in the effect of the instrument on the treatment OR
3. $cov(\beta_{1i}, \pi_{1i}) = 0$ – there is heterogeneity in both the effect of the treatment on the outcome and the effect of the instrument on the treatment, but these are not systematically related

- Always start by explaining why each of these conditions is not plausible. Explain why each of these conditions is unlikely to hold in the returns to schooling IV example.