Simple instrumental variables regressions\textsuperscript{1}

Hedibert Freitas Lopes

May 2014

\textsuperscript{1}From Wooldridge (2013, Chapter 15)
Return to education (for women)

Estimating the return (log wages) to education for \( n = 428 \) married working women as

\[
l\text{wage}_i = \beta_0 + \beta_1 \text{educ}_i + u_i.
\]

**OLS:**

\[
\hat{l\text{wage}}_i = -0.1852 + 0.1086 \text{educ}_i; \quad R^2 = 0.1158.
\]

\([0.1852] \quad (0.0144)\)

95% C.I. for \( \beta_1 \): (0.0798, 0.1374).

**Conclusion:** Roughly 12% return for another year of education.
Fathers education as an instrument for education

1st requirement: $\text{cov}(\text{fathereduc}, u) = 0$.

2nd requirement: $\text{cov}(\text{fathereduc}, \text{educ}) \neq 0$.

$\hat{\text{educ}}_i = 10.2371 + 0.2694\text{fatheduc}_i \quad R^2 = 0.1706.$

(0.2759) (0.0286)

IV regression:

$\hat{\text{lwage}}_i = 0.441 + 0.059\text{educ}_i \quad R^2 = 0.09.$

(0.446) (0.035)

95% C.I. for $\beta_1 : (-0.011, 0.129)$.

Conclusion: About 6% return to education $\Rightarrow$ omitted ability bias.
Call: lm(formula = lwage ~ educ)

Coefficients:
                        Estimate Std. Error t value Pr(>|t|) 
(Intercept)            -0.1852    0.1852  -1.000  0.318 
educ                   0.1086     0.0144   7.545  2.76e-13 ***
---
Signif. codes: 0 *** 0.001 ** 0.01 * 0.05 . 0.1  1

Residual standard error: 0.68 on 426 degrees of freedom
Multiple R-squared:  0.1179, Adjusted R-squared:  0.1158
F-statistic: 56.93 on 1 and 426 DF,  p-value: 2.761e-13

Call: lm(formula = educ ~ fatheduc)

Residuals:    Min           1Q     Median           3Q       Max
-8.4704       -1.1231    -0.1231        0.9546      5.9546

Coefficients:
                    Estimate Std. Error  t value  Pr(>|t|) 
(Intercept)         10.23705   0.27594   37.099 <2e-16 ***
fatheduc            0.26944    0.02859    9.426 <2e-16 ***
---
Signif. codes: 0 *** 0.001 ** 0.01 * 0.05 . 0.1  1

Residual standard error: 2.081 on 426 degrees of freedom
Multiple R-squared:  0.1726, Adjusted R-squared:  0.1706
F-statistic: 88.84 on 1 and 426 DF,  p-value: < 2.2e-16
Return to education (men)

If the number of siblings is an instrument for education, ie.

\[ \text{educ}_i = \beta_0 + \beta_1 \text{sibs}_i + u_i, \]

so

\[ \hat{\text{educ}}_i = 14.1388 - 0.2279\text{sibs}_i \]

\[ R^2 = 0.05625. \]

Assuming that \( \text{cov}(\text{sibs}, u) = 0 \), then the IV fit is

\[ \hat{\text{lwage}}_i = 5.13 + 1.122\text{sibs}_i \]

\[ \text{OLS: } \hat{\beta}_1 = 0.0598 \text{ with a standard error of } 0.006 \text{ and } R^2 = 0.096. \]

\textbf{Conjecture:} Maybe more siblings means, on average, less parental attention, which could result in lower ability.
R output

Call:
  lm(formula = lwage ~ educ)

Coefficients:
            Estimate Std. Error   t value  Pr(>|t|)  
(Intercept)  5.973062   0.081374  73.4082 < 2e-16 ***
educ        0.059839   0.005963   10.0437 < 2e-16 *** 
---
Signif. codes: 0 *** 0.001 ** 0.01 * 0.05 . 0.1 1

Residual standard error: 0.4003 on 933 degrees of freedom
Multiple R-squared:  0.09742, Adjusted R-squared:  0.09645
F-statistic: 100.7 on 1 and 933 DF,  p-value: < 2.2e-16

Call:
  lm(formula = educ ~ sibs)

Coefficients:
            Estimate Std. Error   t value  Pr(>|t|)  
(Intercept) 14.138790   0.113138 124.9596 < 2e-16 ***
sibs      -0.227922   0.030281  -7.5277 1.22e-13 ***
---
Signif. codes: 0 *** 0.001 ** 0.01 * 0.05 . 0.1 1

Residual standard error: 2.134 on 933 degrees of freedom
Multiple R-squared:  0.05726, Adjusted R-squared:  0.05625
F-statistic: 56.67 on 1 and 933 DF,  p-value: 1.215e-13
Graphical summaries
Binary/categorical instrument

Angrist and Krueger (1991) proposed \texttt{frstqrt} (=1 if born in the 1st quarter of the year) as an instrumental variable for education.

\[
\text{cov}(\text{l wage} - \beta_0 - \beta_1 \text{educ}, \text{frstqrt}) = 0 \\
\text{cov}(\text{ability}, \text{frstqrt}) = 0
\]

Compulsory school attendance \implies \text{cov(educ, frstqrt)} \neq 0.

Years of education varies only slightly across quarter of birth. Based on \( n = 247,199 \) they found that

- OLS: \( \hat{\beta}_1 = 0.0801 \) (standard error 0.0004)
- IV: \( \hat{\beta}_1 = 0.0715 \) (0.0219).

\textbf{Headache:} Even a small amount of correlation between z and u can cause serious problems for the IV estimator.
Graphical summaries

- Education (highest grade completed)
- Log weekly wage

- 1st quarter birthday
- 2nd–4th quarter birthday