RETURN TO EDUCATION

Estimating the return to education for married women

mroz.csv: 753 observations and 22 variables

1. inlf =1 if in labor force, 1975
2. hours hours worked, 1975
3. kids1t6 # kids < 6 years
4. kidsge6 # kids 6-18
5. age woman’s age in yrs
6. educ years of schooling
7. wage estimated wage from earns., hours
8. repwage reported wage at interview in 1976
9. hushrs hours worked by husband, 1975
10. husage husband’s age
11. huseduc husband’s years of schooling
12. huswage husband’s hourly wage, 1975
13. faminc family income, 1975
14. mtr fed. marginal tax rate facing woman
15. motheduc mother’s years of schooling
16. fatheduc father’s years of schooling
17. unem unem. rate in county of resid.
18. city =1 if live in SMSA
19. exper actual labor mkt exper
20. nwifeinc (faminc - wage*hours)/1000
21. lwage log(wage)
22. expersq exper^2
We use the data on married working women to estimate the return to education in the simple regression model

\[ \log(\text{wage}) = \beta_0 + \beta_1 \text{educ} + u. \]

OLS estimates (for comparison):

\[ \hat{\log(\text{wage})} = -0.185 + 0.109 \text{ educ} \]

\[ (0.185) \quad (0.014) \]

where \( n = 428 \) and \( R^2 = 0.118. \)

The estimate for \( \beta_1 \) implies an almost 11% return for another year of education.
We have to maintain that:

- \textbf{fatheduc} is uncorrelated with \textit{u}, and
- \textbf{educ} and \textbf{fatheduc} are correlated.

Simple regression of \textbf{educ} on \textbf{fatheduc}:

\[
\hat{\text{educ}} = 10.24 + 0.269 \text{ fatheduc}
\]

\[
(0.28) \quad (0.029)
\]

where \( n = 428 \) and \( R^2 = 0.173 \).
IV regression

Using fatheduc as an IV for educ gives

$$\hat{\log(wage)} = 0.441 + 0.059 \text{ educ}$$

(0.446) (0.035)

where $n = 428$ and $R^2 = 0.093$.

The IV estimate of the return to education is 5.9%, which is barely more than one-half of the OLS estimate.

This suggests that the OLS estimate is too high and is consistent with omitted ability bias.
install.packages("ivpack")
library(ivpack)

data = read.csv("mroz.csv",header=TRUE)
attach(data)
n = nrow(data)
reg1 = lm(lwage~educ)
reg2 = lm(educ~fatheduc)
reg3 = ivreg(lwage ~ educ | fatheduc)
IV regression

Call:
ivreg(formula = lwage ~ educ | fatheduc)

Residuals:

<table>
<thead>
<tr>
<th></th>
<th>Min</th>
<th>1Q</th>
<th>Median</th>
<th>3Q</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>-3.0870</td>
<td>-0.3393</td>
<td>0.0525</td>
<td>0.4042</td>
<td>2.0677</td>
</tr>
</tbody>
</table>

Coefficients:

|                | Estimate | Std. Error | t value | Pr(>|t|) |
|----------------|----------|------------|---------|---------|
| (Intercept)    | 0.44110  | 0.44610    | 0.989   | 0.3233  |
| educ           | 0.05917  | 0.03514    | 1.684   | 0.0929  .|

---

Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1 1

Residual standard error: 0.6894 on 426 degrees of freedom
Multiple R-Squared: 0.09344, Adjusted R-squared: 0.09131
Wald test: 2.835 on 1 and 426 DF, p-value: 0.09294

\[1\] This is done by two-stage least squares.
Suppose now we entertain the following structural model

\[
\log(\text{wage}) = \beta_0 + \beta_1 \text{educ} + \beta_2 \text{exper} + \beta_3 \text{expersq} + u,
\]

with motheduc, fatheduc, and huseduc as instrumental variables for educ.

The reduced form is

\[
\text{educ} = \delta_0 + \delta_1 \text{motheduc} + \delta_2 \text{fatheduc} + \delta_3 \text{huseduc} + \delta_4 \text{exper} + \delta_5 \text{expersq} + \epsilon.
\]
install.packages("ivpack")

library(ivpack)

data = read.csv("mroz.csv",header=TRUE)

attach(data)

n = nrow(data)

reg1 = lm(lwage ~ educ + exper + expersq)

reg2 = lm(educ ~ motheduc + fatheduc + huseduc + exper + expersq)

reg3 = ivreg(lwage ~ educ + exper + expersq | motheduc + fatheduc + huseduc + exper + expersq)
> summary(reg1)

Call:
  lm(formula = lwage ~ educ + exper + expersq)

Residuals:
   Min     1Q  Median     3Q    Max 
-3.08404 -0.30627  0.04952  0.37498  2.37115 

Coefficients: 
                         Estimate Std. Error t value   Pr(>|t|)
(Intercept)          -0.5220407  0.1986321 -2.628  0.00890 **
educ                 0.1074896  0.0141465  7.598 1.94e-13 ***
exper                0.0415665  0.0131752  3.155 0.00172 **
expersq              -0.0008112  0.0003932 -2.063 0.03974 *
---
Signif. codes: 0 *** 0.001 ** 0.01 * 0.05 . 0.1 1

Residual standard error: 0.6664 on 424 degrees of freedom 
Multiple R-squared: 0.1568, Adjusted R-squared: 0.1509 
F-statistic: 26.29 on 3 and 424 DF,  p-value: 1.302e-15
> summary(reg2)

Call:
  lm(formula = educ ~ motheduc + fatheduc + huseduc + exper + expersq)

Residuals:
   Min     1Q Median     3Q    Max
-6.6882 -1.1519  0.0097  1.0640  5.7302

Coefficients:  Estimate  Std. Error   t value  Pr(>|t|)
(Intercept)  5.5383110  0.4597824  12.046 < 2e-16 ***
motheduc    0.1141532  0.0307835   3.708  0.000237 ***
fatheduc    0.1060801  0.0295153   3.594  0.000364 ***
huseduc     0.3752548  0.0296347  12.663 < 2e-16 ***
exper       0.0374977  0.0343102   1.093  0.275059
expersq    -0.0006002  0.0010261  -0.585  0.558899
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1 1

Residual standard error: 1.738 on 422 degrees of freedom
Multiple R-squared:  0.4286, Adjusted R-squared:  0.4218
F-statistic: 63.3 on 5 and 422 DF,  p-value: < 2.2e-16
> summary(reg3)

Call:
ivreg(formula = lwage ~ educ + exper + expersq |
      motheduc + fatheduc + huseduc + exper + expersq)

Residuals:
    Min     1Q    Median     3Q    Max
-3.0838 -0.3214  0.0354  0.3693  2.3583

Coefficients:
                 Estimate  Std. Error t value Pr(>|t|)    
(Intercept)   -0.186857  0.285396  -0.655  0.513006    
educ          0.080392  0.021774   3.692  0.000251 ***
exper         0.043097  0.013265   3.249  0.001250 **
expersq      -0.000863  0.000396  -2.178  0.029976     

---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1 1

Residual standard error: 0.6693 on 424 degrees of freedom
Multiple R-Squared: 0.1495, Adjusted R-squared: 0.1435
Wald test: 11.52 on 3 and 424 DF, p-value: 2.817e-07
Hausman’s test

```r
reg2 = lm(educ ~ motheduc + fatheduc + huseduc + exper + expersq)
error = reg2$res
reg4 = lm(lwage ~ educ + exper + expersq + error)
summary(reg4)
```

Coefficients:

| Estimate  | Std. Error | t value | Pr(>|t|) |
|-----------|------------|---------|---------|
| (Intercept) | -0.1868573 | 0.2835905 | -0.659 | 0.51032 |
| educ       | 0.0803918  | 0.0216362 | 3.716  | 0.00023 *** |
| exper      | 0.0430973  | 0.0131810 | 3.270  | 0.00116 **  |
| expersq    | -0.0008628 | 0.0003937 | -2.192 | 0.02895 *   |
| error      | 0.0471890  | 0.0285519 | 1.653  | 0.09912 .   |

---

Signif. codes: 0 *** 0.001 ** 0.01 * 0.05 . 0.1 1

Since the estimated coefficient of `error`, 0.0471890, is statistically significant at the 10% level, then `educ` is endogenous. Not at the 5% levels though!
Sargan’s test

\[
\text{reg3} = \text{ivreg(lwage} \sim \text{educ + exper + expersq} | \\
\text{motheduc + fatheduc + huseduc + exper + expersq)}
\]

\[
\text{error} = \text{reg3}\$\text{res}
\]

\[
\text{reg5} = \text{lm(error} \sim \text{motheduc + fatheduc + huseduc + exper + expersq)}
\]

\[
p = 3
\]

\[
q = 1
\]

\[
\text{sarg} = (n-6)*\text{summary(reg5)}\$\text{r.sq}
\]

\[
\text{p.val} = 1-\text{pchisq(sarg,p-q)}
\]

\[
\text{p.val}
\]

[1] 0.5771194

Conclusion: Fail to reject $H_0$ : instruments are valid.
With or without huseduc?

```
reg6 = ivreg(lwage ~ educ + exper + expersq | motheduc + fatheduc + exper + expersq)
summary(reg6)

Coefficients:
                Estimate Std. Error  t value Pr(>|t|)
(Intercept)   0.0481003  0.4003281   0.120   0.90442
educ         0.0613966  0.0314367   1.953   0.05147 .
exper        0.0441704  0.0134325   3.288  0.00109 **
expersq     -0.0008990  0.0004017  -2.238  0.02574 *
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1 1

Residual standard error: 0.6747 on 424 degrees of freedom
Multiple R-Squared: 0.1357, Adjusted R-squared: 0.1296
Wald test: 8.141 on 3 and 424 DF, p-value: 2.787e-05

Notice that educ becomes statistically irrelevant!
```
WITH OR WITHOUT huseduc?

educ is endogenous at the 10% significance level. Fail to reject $H_0$: instruments are valid.

# Hausman’s test
reg2 = lm(educ ~ motheduc + fatheduc + exper + expersq)
error = reg2$res
reg4 = lm(lwage ~ educ + exper + expersq + error)
summary(reg4)
Coefficients:

|            | Estimate | Std. Error | t value | Pr(>|t|) |
|------------|----------|------------|---------|----------|
| (Intercept)| 0.0481003| 0.3945753  | 0.122   | 0.903033 |
| educ       | 0.0613966| 0.0309849  | 1.981   | 0.048182 *|
| exper      | 0.0441704| 0.0132394  | 3.336   | 0.000924 ***|
| expersq    | -0.0008990| 0.0003959  | -2.271  | 0.023672 *|
| error      | 0.0581666| 0.0348073  | 1.671   | 0.095440 .|

---
Signif. codes: 0 *** 0.001 ** 0.01 * 0.05 . 0.1 1

# Sargan’s test
reg3 = ivreg(lwage ~ educ + exper + expersq |
            motheduc + fatheduc + exper + expersq)
error = reg3$res
reg5 = lm(error ~ motheduc + fatheduc + exper + expersq)
p = 2
q = 1
sarg = (n-5)*summary(reg5)$r.sq
p.val = 1-pchisq(sarg,p-q)
p.val
[1] 0.541019
Summary

Model 1
\[ \log(\text{wage}) = \beta_0 + \beta_1 \text{educ} + u \]
IV for educ: fatheduc
\[ \hat{\beta}_1 = 0.05917, \text{s.e.}(\hat{\beta}_1) = 0.03514, \text{p-value} = 0.0929. \]

Model 2
\[ \log(\text{wage}) = \beta_0 + \beta_1 \text{educ} + \beta_2 \text{exper} + \beta_3 \text{expersq} + u \]
IV for educ: motheduc, fatheduc, and huseduc
\[ \hat{\beta}_1 = 0.08039, \text{s.e.}(\hat{\beta}_1) = 0.02177, \text{p-value} = 0.0003. \]

Model 3
\[ \log(\text{wage}) = \beta_0 + \beta_1 \text{educ} + \beta_2 \text{exper} + \beta_3 \text{expersq} + u \]
IV for educ: motheduc and fatheduc
\[ \hat{\beta}_1 = 0.06140, \text{s.e.}(\hat{\beta}_1) = 0.03143, \text{p-value} = 0.0515. \]