Instrumental Variable Methods

STAT 266 week 9

Observational Studies

Omitted Variables: Fixing Broken Regressions

$Y = \beta_0 + \underline{\beta_1} D + u$ Dose-response, (returns to schooling) $Y = \text{lwage}$ $D = \text{educ}$ but $Cov(D,u) \neq 0$

$Y = \beta_0 + \underline{\beta_1} G + u$ Group membership effects, t-test but $Cov(G,u) \neq 0$

"Broken" → D, G correlated with omitted variables in u

OLS fails for $Y = \beta_0 + \beta_1 X + u$ when $cov(X,u) \neq 0$

e.g. $\log(wage) = \beta_0 + \beta_1 X + u$ (ability omit, Angrist)

To the rescue? instrument Z such that $Cov(X,Z) \neq 0$ empirical assoc. strong, weak instr

AND $cov(Z,u) = 0$

cue Dusty S. hope, untestable

if true ancova also works

Z "exogenous" no partial effect on Y (even if Z random ass in RCT)

Properties: $Y = \beta_0 + \beta_1 X + u \Rightarrow$

$Cov(Z,Y) = \beta_1 cov(Z,X) + \cancel{cov(Z,u)} \to 0$ by wish upon a star

thus $\beta_1^{IV} = cov(Z,Y)/cov(Z,X)$ (Z replaces X)

$\hat{\beta}_1^{IV} = \dfrac{s_{YZ}}{s_{XZ}}$ $\quad \beta_0^{IV} = \bar{Y} - \hat{\beta}_1^{IV}\bar{X}$, $\widehat{Var}(\hat{\beta}_1^{IV}) = \dfrac{\hat{\sigma}^2}{SSX \cdot r_{XZ}^2}$ weak? instrument

($r_{XZ} = .1 \Rightarrow n/100$)

RCT IV Random Assignment $G = 1, 0$

Encouragement Designs

Clever, innovative designs for estimating Dose-Response

Dose, binary or measured

$G = 1, 0$ encourage or not (RCT)

D self-selected dose

Y outcome

IV assume assignment G no effect on Y

$\beta_{IV} = \dfrac{s_{YG}}{s_{DG}} = \dfrac{\bar{Y}_1 - \bar{Y}_0}{\bar{D}_1 - \bar{D}_0}$ wald estimator

salt, sesame st. in session

Compliance Adjustments

Desperate attempt to adjust, salvage broken protocols in RCT

Compliance seldom binary

Binary Compliance c (comply) n (not)

no crossover $\pi_c = P(T \mid G = 1)$

ITT $\mu_1 - \mu_0$

CACE $\mu_{c1} - \mu_{c0} = \dfrac{ITT}{\pi_c}$ iff

$\mu_{n1} - \mu_{n0} = 0$ ER, IV assumption

$\mu_1 - \mu_0 = \pi_c(\mu_{c1} - \mu_{c0}) + (1 - \pi_c)(\mu_{n1} - \mu_{n0}) \to 0$

examples in session

see full Rogosa session


```
A. Observational data (adapted from Lab 3 Stat209)    Mroz data, returns to schooling
##### Examples from Wooldridge, Introductory Econometrics #######  Chapters 15 and 16
## stata results available from
http://fmwww.bc.edu/gstat/examples/wooldridge/wooldridge15.html

> lm.posWage = lm(logWage ~ educ)  > summary(lm.posWage)#session has full output
Coefficients:  #year increase in educ, fit increases 11 cents hourly wage
            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 ***
Multiple R-Squared: 0.1179,     Adjusted R-squared: 0.1158
F-statistic: 56.93 on 1 and 426 DF,  p-value: 2.761e-13
> # we get a highly significant slope (but not big Rsq),

> install.packages("AER") > library(AER) > help(ivreg)
> ivreg1 = ivreg(logWage ~ educ| fatheduc)
####### use the diagnostics option of ivreg (recent)
>  summary(ivreg1, diagnostics = TRUE)
Call: ivreg(formula = logWage ~ educ | fatheduc)
Coefficients: # matches tsls Task1
            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 .

Diagnostic tests:
                 df1 df2 statistic p-value
Weak instruments   1 426     88.84  <2e-16 ***
Wu-Hausman         1 425      2.47   0.117
Sargan             0  NA        NA      NA
---
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
> confint(ivreg1)   # parameter confidence intervals as usual
                  2.5 %    97.5 %
(Intercept) -0.433239996 1.3154468
educ        -0.009703131 0.1280501

##Investigate the 'Weak instruments' entry, matches cor.test
> cor(educ, fatheduc)
[1] 0.415403

## just a test (like in the main lab of correlation between predictor and instrument

# the Hausman test is often used to show there's a difference between OLS and IV,
# but what does that tell you?
#  ivreg diagnostics finds discrepancy non-signicance even though IV est cuts the OLS
#  estimate in half  and IV estimate now non-significant
## to see the details of the Hausman test (and do it by hand) see pdf page 46 onward from
http://personal.rhul.ac.uk/uhte/006/ec2203/Lecture%2015_IVestimation.pdf
also the Basel class notes linked in week 6

## The Sargan test is a statistical test used for testing over-identifying restrictions
## relevant to simultaneous eqs, which we investigated in identifiability in lab script

# some package ivmodel display in week9 RQ

################################################################################################
in Rogosa session    B. Encouragement Design   C. Compliance Adjustments
```