Stat 209

WEEK 6

IV equations

week 6 math facts

omitted vars setting (fixing broken regressions) problem

$$y = \beta_0 + \beta_1 x + u \qquad cov(x,u) \neq 0$$

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

Angrist Krueger

instrument z s.t. $cov(z,u)=0$; $cov(z,x)\neq 0$

(e.g. birth qtr) (hope, untestable) (empirical association)

z exogenous: no partial effect on y

x on z: $x = \pi_0 + \pi_1 z + v$ ⇒ (for TSLS)

result $cov(z,y) = \beta_1 cov(z,x) + cov(z,u)$ ($cov(z,u) \to 0$ by assumption) wishin' and hopin'

$$\Rightarrow \beta_1 = \frac{cov(z,y)}{cov(z,x)} \qquad \text{in sample } \hat{\beta}_1 = \frac{s_{yz}}{s_{xz}}$$

cf meas error ex

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} \qquad Var(\hat{\beta}_1) = \frac{\sigma^2}{n\sigma_x^2 \rho_{xz}^2}$$

$\sigma^2 = Var(u)$ iv resids $\hat{u} = y_i - \hat{\beta}_0 - \hat{\beta}_1 x$

$$\hat{\sigma}^2 = \frac{1}{n-2}\sum \hat{u}^2 \qquad \hat{Var}(\hat{\beta}_1) = \hat{\sigma}^2/(SSX \cdot r_{xz}^2)$$

compare ancova using z as covariate ovr.

weak instrument inflates variance

mse worse even if bias = 0

Two-stage least squares (and IV) [PAF ch 8 proof in general]

predict x by z

$$\hat{x} = \bar{x} + \frac{\hat{cov}(x,z)}{\hat{var}(z)}(z - \bar{z})$$

slope of y on $\hat{x}$

$$\frac{\hat{cov}(y,\hat{x})}{\hat{var}(\hat{x})} = \frac{\left(\frac{\hat{cov}(x,z)}{\hat{var}(z)}\right)\hat{cov}(y,z)}{\hat{var}(z)\left(\frac{\hat{cov}(x,z)}{\hat{var}(z)}\right)^2}$$

(collect terms)

$$= \hat{cov}(y,z)/\hat{cov}(x,z) = \hat{\beta}_{yx}^{IV}$$

cf meas error ex

Stat209 2...

```
> #measurement error example, IV, TSLS
> w = rnorm (1000, 10,2)
> var(w)
[1] 4.016603
> y = w + rnorm(1000, 0, sqrt(2)) #make y corr with w sqrt(2/3)
> cor(w,y); sqrt(2/3)
[1] 0.825847
[1] 0.8164966
> x = w + rnorm(1000,0,sqrt(.5)) #observable x with rel .9
> var(w)/var(x)
[1] 0.9032945
> z = w + rnorm(1000,0,sqrt(.5)) #parallel observable z with rel .9
> var(w)/var(z)
[1] 0.8975333
> cor(x,z)  #parallel forms reliability estimate
[1] 0.8859663

> truereg = lm(y ~ w) > summary(truereg)
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.24118    0.22185  -1.087    0.277
w            1.02186    0.02209  46.267   <2e-16 ***
---

> obsreg = lm(y~x) #OLS with observable, biased downward
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.83746    0.23550   3.556 0.000394 ***
x            0.91378    0.02343  38.996  < 2e-16 ***
---

> cov(y,z)/cov(x,z) # IV estimator, z instrument for x
[1] 1.041717

> xonzreg = lm(x~z)
> tslsreg = lm(y ~ fitted(xonzreg))
> summary(tslsreg)
Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)     -0.41976    0.26046  -1.612    0.107
fitted(xonzreg)  1.04172    0.02604  40.005   <2e-16 ***
---

> install.packages("sem")
> library(sem)
> summary(tsls(y~x, ~z))
 2SLS Estimates   Model Formula: y ~ x   Instruments: ~z
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  -0.4198    0.26847  -1.564   0.1182
x             1.0417    0.02684  38.811   0.0000
Residual standard error: 1.5849 on 998 degrees of freedom
> sqrt(diag(vcov(tsls(y~x, ~z))))
(Intercept)           x
 0.26847157  0.02684083
>
```

should do 10,000

y outcome

predictor x

very good not perfect measurement

target param 1.0

biased downward

check props of z. as instrument

IV estimator using z "disattenuates"

S_{yz}/S_{xz}

imitate TSLS using instrument same estimate as IV

Lab 3 formal TSLS

S.E. IV, TSLS larger

MSE comp w/ OLS?

unbiased not necessarily best w/ weak instruments. see front

also Lab 3 package AER w/ ivreg STATA clone