Instrumental Variables

what if $E(\varepsilon | X) \neq 0$? Possible reasons: ① omitted variables
② measurement errors
③ reverse causality (simultaneity)

Ex. add more police ⇒ reduce crime

(reverse causality) ① Crime $= \alpha + \beta$ Police $+ \varepsilon$

Intuitively $\beta < 0$. But $\hat{\beta}_{OLS} > 0$ often

problem: cities w/ high crime have higher demand for police

In reality, it's

② $$\text{Police} = \gamma + \delta\,\text{Crime} + \eta \qquad \text{expect } \delta > 0$$

$$\text{Crime} = -\frac{\gamma}{\delta} + \frac{1}{\delta}\,\text{Police} - \frac{\eta}{\delta}$$

Sub. in eq. (1)

$$-\frac{\gamma}{\delta} + \frac{1}{\delta}\,\text{Police} - \frac{\eta}{\delta} = \alpha + \beta\,\text{Police} + \varepsilon$$

$$\left(\frac{1}{\delta} - \beta\right)\text{Police} = \frac{\eta}{\delta} + \frac{\gamma}{\delta} + \alpha + \varepsilon$$

$$\text{cov}(\text{Police}, \varepsilon) = \frac{\delta}{1-\beta\delta}\,\text{var}(\varepsilon) \neq 0$$

$$\hat{\beta}_{OLS} = \frac{\widehat{\text{cov}}(\text{Crime}, \text{Police})}{\widehat{\text{var}}(\text{Police})} = \beta + \frac{\widehat{\text{cov}}(\varepsilon, \text{Police})}{\widehat{\text{var}}(\text{Police})}$$

$$= \beta + \frac{\delta}{1-\beta\delta}\,\frac{\widehat{\text{var}}(\varepsilon)}{\widehat{\text{var}}(\text{Police})} \qquad \text{inconsistent}$$

Ex. Sample selection

$$\text{Health} = \alpha + \beta\,\text{New Medicine} + \varepsilon$$

intuition $\beta > 0$

reality: $\hat{\beta}_{OLS} < 0$

because the people who try a new medicine are often the sickest and most desperate (negative ε's)

So the people who choose a new medicine have negative ε's (correlation

omitted variable bias: what we need is a previous health condition var (we need to model the choice of new medicine) to remove the selection bias

Ex. measurement errors

Recall

True model $\quad y_i = \beta x_i + \varepsilon_i$

obs $\quad x_i^0 = x_i + u_i$

if we regress y on x^0, we get

$$y = \beta x^0 + \varepsilon - \beta u = \beta x^0 + \eta$$

and

$$\text{cov}(x^0, \eta) = \text{cov}(x + u, \varepsilon - \beta u) = -\beta \sigma_u^2 \neq 0$$

x's correlated w/errors

So

$\hat{\beta}$ is inconsistent

Sol'n. Suppose we have another ~~...~~ noisy measurement of x

$z_i = x_i + v_i \quad , \quad v_i$ uncorrelated w/ y_i, x_i, x_i^0

Then

$$\begin{aligned} E(z_i x_i^0) &= E[(x_i + v_i)(x_i + u_i)] \\ &= E[x_i x_i + \overset{0}{v_i x_i} + \overset{0}{x_i u_i} + \overset{0}{v_i u_i}] \\ &= E[x_i x_i] \end{aligned}$$

and

$$E(z_i y_i) = E\left[(x_i + v_i)\overset{y_i}{\cancel{(x + \varepsilon_i)}}\right] = E(x_i y_i)$$

So instead of $\hat\beta = S_{xy}/S_{xx}$ we can use $\hat\beta_{IV} = \dfrac{S_{zy}}{S_{zx}}$

instrument needs to be correlated w/regressors else disaster

and in more general case $\hat\beta_{IV} = (Z'X)^{-1} Z'Y$

General case:

$$\underset{k\times 1}{y_i = x_i'\beta + \varepsilon_i} \qquad E\varepsilon_i = 0 \qquad E(x_i\varepsilon_i) \neq \underset{\sim}{0}$$

Suppose $\exists$ instruments $\underset{L\times 1,\ L = k}{z_i}$ s.t. $E(z_i\varepsilon_i) = 0$, $E(z_i x_i')$ invertible

and X, Z $\overset{(iid)}{\text{well-behaved}}$. Consider $\hat\beta_{IV} = (Z'X)^{-1}Z'Y$

$$\text{plim}\left(\frac{Z'y}{n}\right) = \text{plim}\left(\frac{Z'X}{n}\right)\beta + \underbrace{\text{plim}\left(\frac{Z'\varepsilon}{n}\right)}_{0}$$

~~Thus~~ Since $Z'X$ is a square matrix, it's invertible, and

$$\left[\text{plim}\left(\frac{Z'X}{n}\right)\right]^{-1}\text{plim}\left(\frac{Z'y}{n}\right) = \beta \quad \Rightarrow \text{consistent}$$

$$\sqrt{n}(\hat\beta_{IV} - \beta) = \sqrt{n}\left[(Z'X)^{-1}Z'(X\beta + \varepsilon) - \beta\right]$$

$$= \sqrt{n}\left[(Z'X)^{-1}Z'\varepsilon\right]$$

$$= \left(\frac{Z'X}{n}\right)^{-1}\frac{Z'\varepsilon}{\sqrt{n}}$$

$$\Big\downarrow p \qquad\qquad \Big\downarrow d \text{ CLT}$$

$$\underset{\substack{\parallel\!\mid \\ D}}{E(z_i x_i)} \qquad N(0, \underset{\substack{\parallel\!\mid \\ V}}{Var(z_i\varepsilon_i)})$$

$$\sqrt{n}(\hat\beta_{IV} - \beta) \xrightarrow{d} N(0, D^{-1}VD^{-1\prime})$$

To estimate: $\hat{D} = \frac{Z'X}{n}$ $\quad V = E\left[z_i \varepsilon_i \varepsilon_i' z_i' \right]$
$= E\left(z_i \varepsilon_i^2 z_i' \right)$

$e_i = y_i - x_i' \hat{\beta}_{IV}$ $\quad \hat{V} = \frac{1}{n} \sum_{i=1}^{n} z_i z_i' e_i^2$ same logic as white estimator

in homoskedastic case this reduces to

$$\sqrt{n}\left(\hat{\beta}_{IV} - \beta\right) \xrightarrow{d} N\left(0, \ \sigma^2 D^{-1} C D^{-1\prime}\right)$$

$$\hat{C} = \frac{Z'Z}{n} \qquad \hat{\sigma}^2 = \frac{e'e}{n}$$

typically no d.f. correction because IV is a lg. sample result

Method of moments

fat tails example

2 moments $\longrightarrow$ 1 param

GMM

back to our IV setup $\quad y_i = \underset{K \times 1}{x_i'} \beta + \varepsilon_i$

$E(\varepsilon_i \mid \underset{L \times 1}{z_i}) = 0 \qquad L \geq K$

orthogonality condition

$$E(\varepsilon_i \mid z_i) = 0 \implies E\left(z_i(y_i - x_i'\beta)\right) = \underset{\sim}{0}$$

and summing over all i

$$E\left[\sum_{i=1}^{n} z_i(y_i - x_i'\beta)\right] = 0$$

$E[m(\beta)] = 0$, the population moment equation

basic idea of GMM: match sample moments to population moments

$$\sum_{i=1}^{n} z_i(y_i - x_i'\hat{\beta}) = m(\hat{\beta}) = 0$$

Note: if $z_i = x_i$, we are back to OLS.

So OLS is a special case of a GMM estimator.

Exactly identified : $L = K$

K o.c.'s , K parms to estimate in $\hat{\beta}$, can get $\underset{\text{sample}}{m}(\hat{\beta}) = \underline{0}$

single solution is   all moments exactly zero

$$\hat{\beta} = (Z'X)^{-1} Z'y$$

Overidentified : $L > K$

L o.c.'s , K parms to estimate in $\hat{\beta}$

let's use quadratic distance (least squares) as our loss function

$$\hat{\beta} = \underset{\beta}{\operatorname{argmin}} \; m(\beta)' m(\beta)$$

F.O.C. $\quad 2 \frac{\partial m'(\beta)}{\partial \beta} m(\hat{\beta}) = 0 \quad 2\, G(\hat{\beta})\, m(\hat{\beta}) = 0$

$$2\left[(X'Z)\, Z'(y - X\hat{\beta})\right] = 0$$

$$\underset{K \times 1}{\hat{\beta}_{GMM}} = \big[\underset{K \times L}{(X'Z)}\underset{L \times K}{(Z'X)}\big]^{-1} \underset{K \times L}{(X'Z)}\underset{L \times 1}{(Z'y)}$$

note that this reduces to $\hat{\beta}_{IV}$ iff $L = K$

to obtain consistency and asymp. dists, we'll need

GMM2 ergodic stationarity of $\{x_i, y_i, z_i\}$

GMM3 $E(z_i x_i')$ is of rank L (rank condition)

GMM5 $E(\varepsilon_i \mid \varepsilon_{i-1}, \dots, z_i, z_{i-1}, \dots, z_1) = 0$ innovation condition

and it can be shown that

$$\widehat{Avar}(\hat{\beta}_{GMM}) = n\left[(X'Z)(Z'X)\right]^{-1}(X'Z)\,\hat{V}\,(Z'X)\left[(X'Z)(Z'X)\right]^{-1}$$

$$\hat{V} = \frac{1}{n}\left[\sum_{i=1}^{n} e_i^2 z_i z_i' + \sum_{\ell=1}^{L}\sum_{t=\ell+1}^{n}\left(1 - \frac{\ell}{L+1}\right) e_t e_{t-\ell}\left(z_t z_{t-\ell}' + z_{t-\ell} z_t'\right)\right.$$

If there's no autocorrelation, the $\sum^{L}$ terms drop out

If there's no heterosk, then the 1^{st} term reduces to $\frac{e'e}{n}\ \frac{1}{n}\sum z_i z_i'$

or $\hat{V} = \frac{e'e}{n}\ \frac{Z'Z}{n}$

Is this efficient? No. Analogous to OLS

- Not clear that min sum of squares is at all optimal

Consider instead $\min\limits_{\beta}\ m(\beta)'\, W\, m(\beta)$ for some p.def. Weighting matrix W

F.o.c. $2\frac{\partial m'}{\partial\beta}(\hat{\beta})\, W\, m(\hat{\beta}) = 0$

$$2(X'Z)\,W\,Z'(y - X\hat{\beta}) = 0$$

$$\underset{}{\hat{\beta}} = \Big[\underset{K\times L}{(X'Z)}\,\underset{L\times L}{W}\,\underset{L\times K}{(Z'X)}\Big]^{-1}\underset{K\times L}{(X'Z)}\,\underset{L\times L}{W}\,\underset{L\times 1}{(Z'y)}$$

and

$$\text{Avar}(\hat{\beta}) = n\Big[(X'Z)W(Z'X)\Big]^{-1}(X'Z)\,WVW\,(Z'X)\Big[(X'Z)(Z'X)\Big]^{-1}$$

Does the choice of W matter? Maybe.

$L = K$ (exactly identified): doesn't matter, only 1 choice of $\hat{\beta}$ that exactly ~~matches~~ gets sample moments $= 0$ $\Rightarrow$ use $W = I$ for simplicity

$L > K$ (overidentified): turns out optimal $W = V^{-1}$ cf. GLS
in which case $\text{Avar}(\hat{\beta})$ reduces to

$$\text{Avar}(\hat{\beta}) = n\Big[(X'Z)\,V\,(Z'X)\Big]^{-1}$$

where do we get $\hat{V}$? use $W = I$ (1^{st} stage), then $\hat{V}$ as above, then use (GMM) $W = \hat{V}^{-1}$