

Instrumental Variables Estimation.

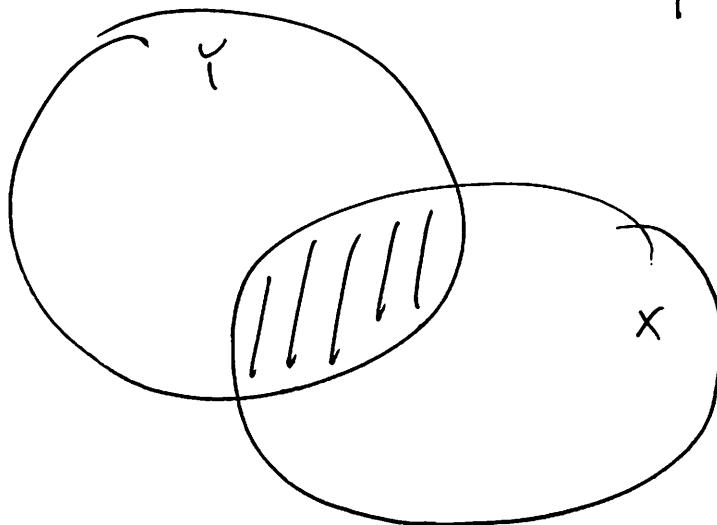
$$y_i = \beta_1 + \beta_2 x_i + u_i$$

$$i = 1, 2, \dots, n$$

If x_i and u_i correlated, then OLS
is inconsistent for β_1 & β_2 .

An instrumental variable, z_i ,
is ~~a variable~~ another variable that
helps isolate the portion of x_i
that is uncorrelated with u_i .

BACCentives (from Kennedy)

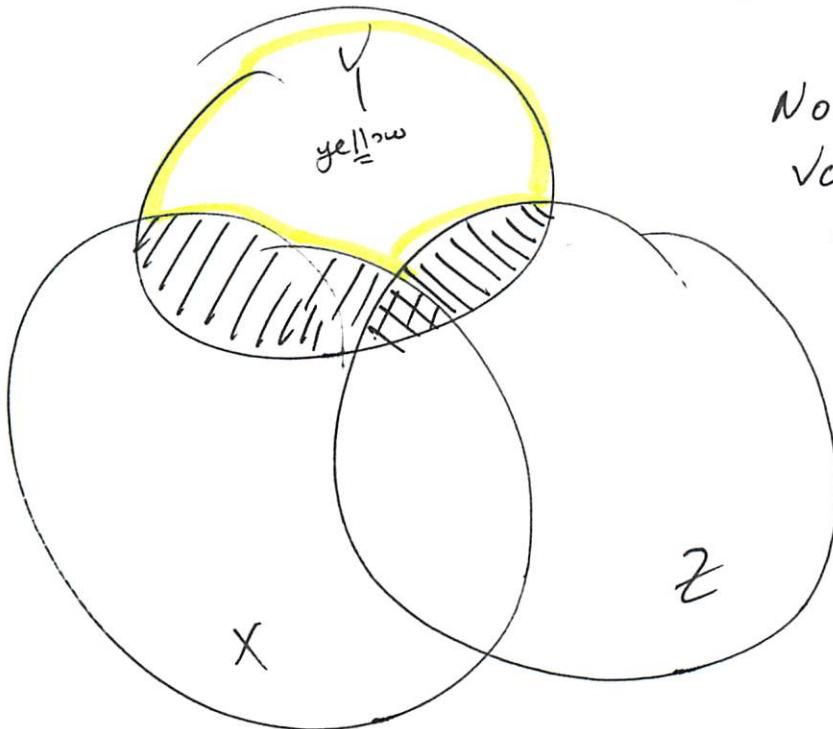


γ circle : Variation
in γ

X circle : Variation
in X .

The variables have common variation
||||| that is used to estimate
slope in the simple regression.
The larger the overlap, the
more information is available to
estimate the slope. \Rightarrow OLS more precise.

$$y = \beta_0 + \beta_1 x + \beta_2 z + \epsilon$$



Now, There is another variable, z , that is correlated with y (and x). The yellow = error term.

Independent variation in x
is used to estimate β_1

Independent variation in z is
used to estimate β_2

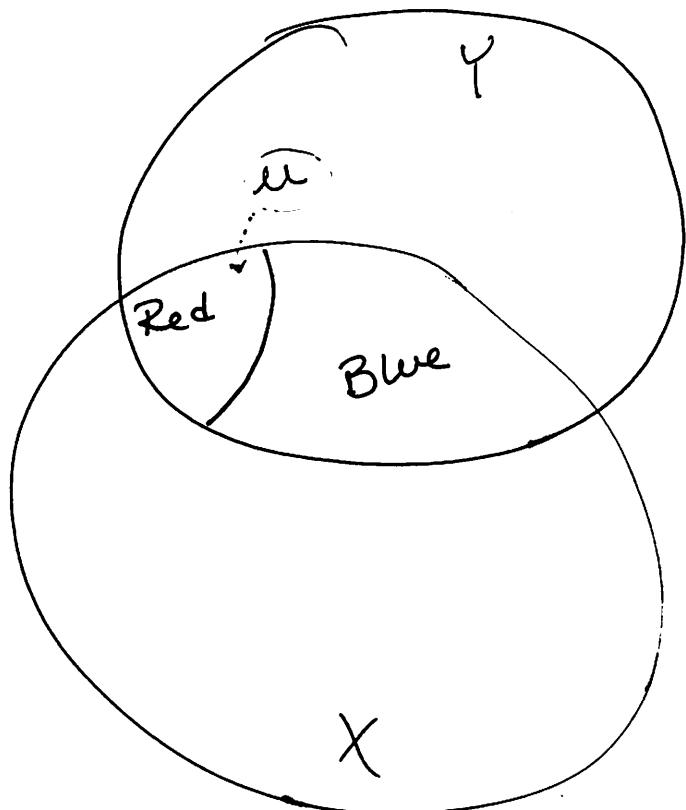
The overlap can't be assigned
to either (collinearity). The

longer this is, the less
precise OLS.

If z was omitted, then OLS
mistakenly assigns variation due to z (overlap)
all to x \therefore omitted variable bias

IV estimation

$$y = \beta_1 + \beta_2 x + u$$

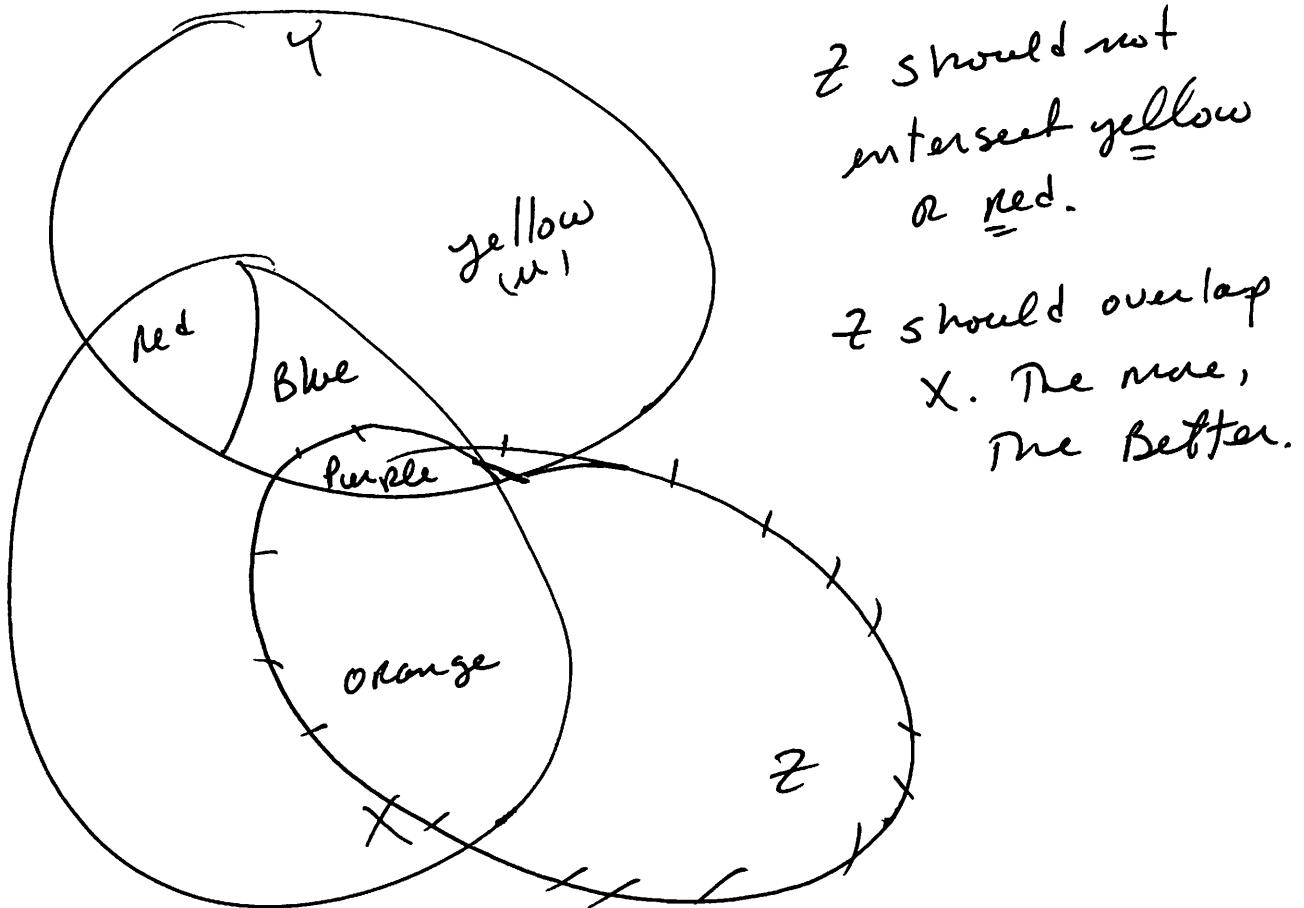


X is all not u .
 \therefore error overlaps with X .
 Variation in
 Red area is
 due to this overlap.
 Variation in Y in Red
 area is due to X
 and u .

Area in Red & Blue is
 used to estimate $\beta_2 : \beta_1$,
But, the Red area does
 not represent variation in Y
 arising from variation solely
 in X . \therefore OLS inconsistent.
 smaller the overlap (correlation
 between X & u) the less of A
 problem you'll have.

Now suppose there is another variable, z .

- (1) not correlated with x
- (2) highly correlated with y



(1) Regress x onto z

$$x = \pi_1 + \pi_2 z + \text{res.}$$

Purple + orange estimates π_i .

(2) Regress

\hat{x} onto y Purple area est. β_3 .

$\sigma = (\pi : z) \text{ non logically } \sigma \in \text{ (2)}$

$\sigma \neq (\pi : z) \text{ recursive } \sigma \in \text{ (1)}$

Instructions

$c = [:\pi | :n] \in \text{weakly}$

$c = [n .. i = j : \pi | :n] \in \text{hybrid}$

Recursive functions: non-logical

Recursive that are not connected with

our forms: endogenous.

Recursive that are connected with

Case when x_i are correlated

① Error-in-Variables.

$$\text{Saving} = \beta_1 + \beta_2 \text{Perm. Inc} + \mu.$$

(PI) Perm income is generally not observable. So, we measure it with current income (I)

$$\text{Income} = \text{PI.} + \text{error.} \quad \text{error} \sim (0, \sigma^2)$$

¶

$$S = \beta_1 + \beta_2 \text{PI} + \mu$$

$$S = \beta_1 + \beta_2 (I - \text{error}) + \mu$$

$$= \beta_1 + \beta_2 I + (\mu - \beta_2 \cdot \text{error})$$

$$= \beta_1 + \beta_2 I + \varepsilon$$

$$\text{where } \varepsilon = \mu - \beta_2 \cdot \text{error.}$$

$$\text{cov}(\varepsilon, I) = E[(\text{PI} + \text{error})(\mu - \beta_2 \cdot \text{error})]$$

$$= E[-\beta_2 \cdot \text{error}^2] = -\beta_2 \sigma^2 \neq 0$$

② Simultaneous Equations

$$Q_d = \beta_1 + \beta_2 P + u_d$$

$$Q_s = \alpha_1 + \alpha_2 P + u_s$$

$$Q_s = Q_d = Q$$

P & Q are jointly determined

causing ~~Q_s to be~~

P & u_s, u_d to be correlated.

We'll consider this in more detail later.

③ Omitted Variables.

If can be shown

$$\hat{\beta}_2 \stackrel{a}{\sim} N\left(\beta_2, \frac{\sigma^2}{r_{xz}^2 (\sum (x_i - \bar{x})^2)}\right)$$

r_{xz} is sample correlation
Between x & z . Note: The larger
this is, the more precise the
IV estimator.

Also, if X is exogenous, then
you can use X as instruments
for itself.

$$r_{xx} = 1 \quad \text{and} \quad \text{The}$$

Variance of $\hat{\beta}_2$ is minimized.
This is the OLS estimate and by
the Gauss-Markov Theorem, it
is efficient in homoskedastic
regressions.

MOM

Suppose Y is a random variable

$$Y \sim (\mu, \sigma^2)$$

$$\Rightarrow E(Y) = \mu$$

$$\Rightarrow \text{Var}(Y) = \sigma^2$$

Population
Moments.

$$= E[(Y - EY)^2]$$

MOM - Method-of-Moments - principle
says, "use sample moments as
estimates of population moments"

$$\text{Mom of } \mu = \frac{1}{n} \sum y_i = \bar{y}$$

$$\text{Mom of } \sigma^2 = \frac{1}{n} \sum (y_i - \bar{y})^2$$

Mom in Regression

$$\Rightarrow E[\tilde{u}_i] = 0$$

$$E[u_i | x_i] = 0 \Rightarrow E[x_i u_i] = 0$$

That is, x_i and u_i are uncorrelated.
This applies to constant as well

$$y_i = \beta_1 + \beta_2 x_i + u_i$$

$$u_i = y_i - \beta_1 - \beta_2 x_i$$

$$E[u_i] = E[y_i - \beta_1 - \beta_2 x_i] = 0$$

$$E[x_i u_i] = E[x_i (y_i - \beta_1 - \beta_2 x_i)] = 0$$

Two population Moments:

Estimate these with sample counterparts

$$\frac{1}{n} \sum (y_i - \hat{\beta}_1 - \hat{\beta}_2 x_i) = 0$$

$$\frac{1}{n} \sum (x_i y_i - \bar{x}_i \hat{\beta}_1 - \bar{x}_i \hat{\beta}_2) = 0$$

\Rightarrow 2 eqs, 2 unknowns

$$\hat{\beta}_1 = \bar{y} - \hat{\beta}_2 \bar{x}$$

$$\hat{\beta}_2 = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2}$$

These [^]_(MOMS) are the same as OCS.

Instruments

$$E(u_i | z_i) = 0$$

$$E(z_i u_i) = 0$$

$$E[(y_i - \beta_1 - \beta_2 x_i)] = 0$$

$$E[z_i(y_i - \beta_1 - \beta_2 x_i)] = 0$$

$$\frac{1}{n} \sum (z_i y_i - z_i \tilde{\beta}_1 - \tilde{\beta}_2 z_i x_i) = 0$$

$$\frac{1}{n} \sum (y_i - \tilde{\beta}_1 - \tilde{\beta}_2 x_i) = 0$$

$$\tilde{\beta}_2 = \frac{\sum (z_i - \bar{z})(y_i - \bar{y})}{\sum (z_i - \bar{z})(x_i - \bar{x})}$$

$$\tilde{\beta}_1 = \bar{y} - \tilde{\beta}_2 \bar{x}$$

Instruments

You'll need at least 1 unique instrument for each endogenous regressor.

If you have the same number of endog reg. as instruments then the model is exactly identified.

If you have more inst. than endog regressors, you are over identified.

~~exactly~~