

Instrumental Variables

- 1 IV Estimation in a Simple Regression Model
- 2 IV Estimation in a Multiple Regression Model

IV Estimation in a Simple Regression Model

instrumental variable
or instrument for
 x

x and u are
correl.

$\Rightarrow x$ is endog.

- Endogeneity

- One solution - <https://youtu.be/eoJJPd6104Q>

- Model

grade
wage

attendance
charter educ.

$$y = \beta_0 + \beta_1 x + u$$

\Rightarrow OLS is
biased

- Suppose z is a variable such that

corr. w/ x and

uncorr. w/ u and

only affects y via x

then we may still be able to estimate β_1 .

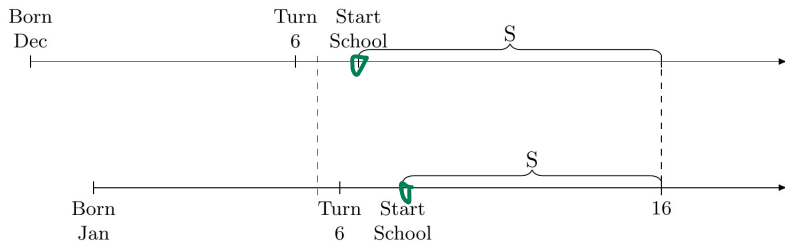
z : lottery

z : compulsory
schooling laws

z : dist. from
school

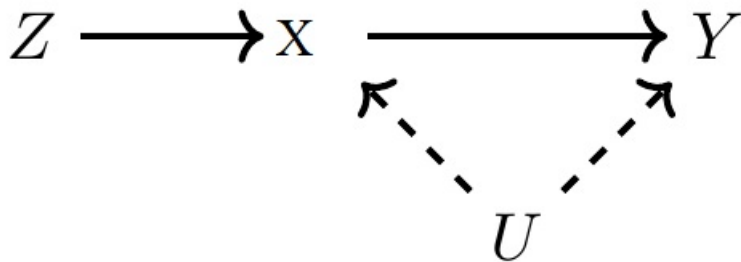
IV Estimation in a Simple Regression Model (cont.)

- Examples



Causal Inference: The Mixtape

IV Estimation in a Simple Regression Model (cont.)



Causal Inference: The Mixtape

IV Estimation in a Simple Regression Model (cont.)

- z and u are uncorrelated
- instrument exogeneity
- z has no direct effect on y (after controlling for x)
- z is uncorr. w/ omitted vars.
- often difficult to test

IV Estimation in a Simple Regression Model (cont.)

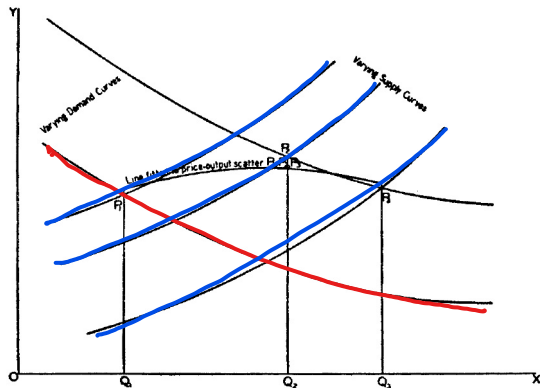
- z and x are correlated

- instrument relevance
- z is related to x
- can be tested by estimating

$$x = \pi_0 + \pi_1 z + v$$

IV Estimation in a Simple Regression Model (cont.)

“Movements in demand and supply can produce an arbitrary scatterplot ... which will trace out neither supply nor demand unless one of the curves is fixed.”



Stock and Trebbi (2003)

IV Estimation in a Simple Regression Model (cont.)

- Model

$$y = \beta_0 + \beta_1 x + u$$

- x and u : correlated
- z is an instrument for x

$z = x \rightarrow$ OLS formula

- The instrumental variables (IV) estimator of β_1

$$\hat{\beta}_1 = \frac{\sum_i (z_i - \bar{z})(y_i - \bar{y})}{\sum_i (z_i - \bar{z})(x_i - \bar{x})}$$

effect of z
on x

\times effect of x
on y

= effect of
 z on y

IV Estimation in a Simple Regression Model (cont.)

- If z and u are uncorrelated, and z and x are correlated, the IV estimator is *consistent*
- The IV estimator is never *unbiased (large samples preferred)*
- In small samples, the IV estimator can have *substantial bias*

IV Estimation in a Simple Regression Model (cont.)

- Statistical Inference

- ▶ Homoskedasticity assumption

$$E(u^2|z) = \sigma^2 = \text{Var}(u)$$

- ▶ Asymptotic standard error of $\hat{\beta}_1$

$$\sqrt{\frac{\hat{\sigma}^2}{SST_x R^2_{xz}}}$$

where SST_x is the total sum of squares of x

$R^2_{x,z}$: R-squared from the regression of x on z \rightarrow 1st stage

$$\hat{\sigma}^2 = \frac{1}{n-2} \sum_i \hat{u}_i^2$$

\hat{u}_i : IV residuals

IV Estimation in a Simple Regression Model (cont.)

- Note

- ▶ Standard error of $\hat{\beta}_1$ in case of OLS

$$\sqrt{\frac{\hat{\sigma}^2}{SST_x}}$$

where SST_x is the total sum of squares of x

$$\hat{\sigma}^2 = \frac{1}{n-2} \sum_i \hat{u}_i^2$$

\hat{u}_i : OLS residuals

- ▶ Typically $R_{x,z}^2 < 1$ and
- ▶ If x and z are only slightly correlated

the IV std. error is always larger than the OLS std. err.

↙
 $R_{x,z}^2$ is small and the IV std. error is large.

IV Estimation in a Simple Regression Model (cont.)

weak instruments

- Note (cont.)

- ▶ If the correlation between x and z is low, we have the problem of
 - ★ Inconsistency in the IV estimator related to $\frac{\text{Corr}(z,u)}{\text{Corr}(z,x)}$
 - ★ This inconsistency (asymptotic bias) in the IV estimator can be large

even if z and u are slightly correlated.



IV Estimation in a Simple Regression Model (cont.)

INSTRUMENTAL VARIABLE

IV is a statistical tool used to estimate causal relationships.

IVs are used when an explanatory variable of interest is correlated with the error term.

Variables that are correlated with the error term are called **EXPLANATORS**.

IV ASSUMPTIONS

- RELEVANCE:** The instrument must be highly correlated with the endogenous variable.
- EXCLUSION:** The instrument must not affect the outcome variable directly. It is uncorrelated with the error in the outcome equation.

IV SNAPSHOT

$Z \rightarrow X \rightarrow Y$

$Z \rightarrow Y$

$X \rightarrow Y$

u

Z is the instrument that induces some of X allowing us to find a "true" relationship between $X \rightarrow Y$.

EXAMPLE:

If air quality is bad, more people may drive to avoid breathing it. Do smoggy days lead people to drive more?

direction of wind \rightarrow Smog \rightarrow more driving

Our instrument!!

The wind can blow Smog into a city

confounders (seasons, prior smog level, factories running)

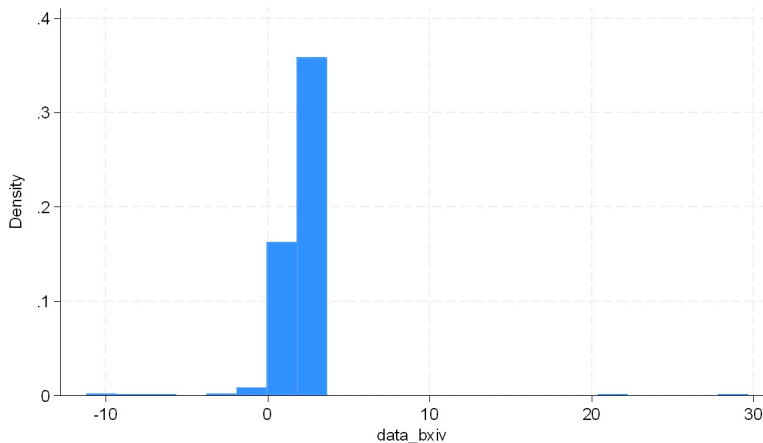
NOTE: IV estimates are only as good as the instrument they use. Check for valid instruments with the first stage F-statistic. We want values larger than 10 (or more!).

Cartoon character: A muscular man with a mustache and a striped shirt, holding a barbell.

@Kate__Barnes

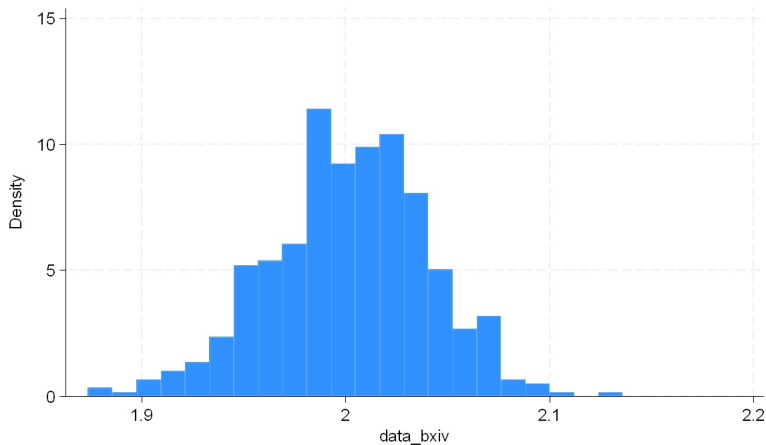
IV Estimation in a Simple Regression Model (cont.)

- $n = 107$ reps = 500, $\text{corr}(x, z) = 0.8$, $\text{corr}(x, u) = 0.5$, $\text{corr}(z, u) = 0$
- $y = 1 + 2x + u$



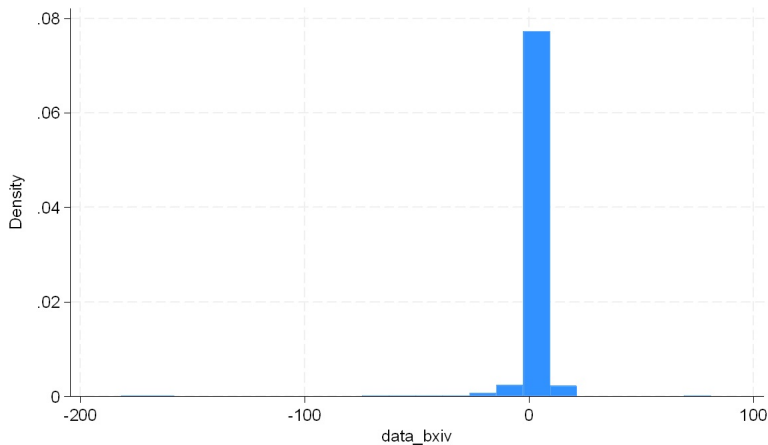
IV Estimation in a Simple Regression Model (cont.)

- $n = 1000$, $\text{reps} = 500$, $\text{corr}(x, z) = 0.8$, $\text{corr}(x, u) = 0.5$,
 $\text{corr}(z, u) = 0$
- $y = 1 + 2x + u$



IV Estimation in a Simple Regression Model (cont.)

- $n = 1000$, $\text{reps} = 500$, $\text{corr}(x, z = 0.01)$, $\text{corr}(x, u = 0.5)$,
 $\text{corr}(z, u = 0)$
- $y = 1 + 2x + u$



IV Estimation in a Multiple Regression Model

- Model

$$y_1 = \beta_0 + \beta_1 y_2 + \beta_2 z_1 + u_1$$

- ▶ Sometimes called
- ▶ y_1 correlated with $u_1 \Rightarrow$
- ▶ z_1 assumed to be uncorrelated with $u_1 \Rightarrow$
- ▶ y_2 correlated with $u_1 \Rightarrow$
- ▶ OLS estimators:
- ▶ z_2 instrumental variable for y_2

IV Estimation in a Multiple Regression Model (cont.)

- Assumptions

$$E(u_1) =$$

$$E(z_1 u_1) =$$

$$E(z_2 u_1) =$$

IV Estimation in a Multiple Regression Model (cont.)

- Equations

$$\begin{aligned} E(y_1 - \beta_0 - \beta_1 y_2 - \beta_2 z_1) &= \\ E[z_1(y_1 - \beta_0 - \beta_1 y_2 - \beta_2 z_1)] &= \\ E[z_2(y_1 - \beta_0 - \beta_1 y_2 - \beta_2 z_1)] &= \end{aligned}$$

IV Estimation in a Multiple Regression Model (cont.)

- Sample analogs

$$n^{-1} \sum_{i=1}^n \left(y_{i1} - \hat{\beta}_0 - \hat{\beta}_1 y_{i2} - \hat{\beta}_2 z_{i1} \right) =$$

$$n^{-1} \sum_{i=1}^n z_{i1} \left(y_{i1} - \hat{\beta}_0 - \hat{\beta}_1 y_{i2} - \hat{\beta}_2 z_{i1} \right) =$$

$$n^{-1} \sum_{i=1}^n z_{i2} \left(y_{i1} - \hat{\beta}_0 - \hat{\beta}_1 y_{i2} - \hat{\beta}_2 z_{i1} \right) =$$

- *Instrumental variables estimators*

IV Estimation in a Multiple Regression Model (cont.)

- Still need z_2 and y_2 to be
- *Reduced form equation*

$$y_2 = \pi_0 + \pi_1 z_1 + \pi_2 z_2 + v_2$$

- ▶ Key condition

IV Estimation in a Multiple Regression Model (cont.)

- Note

- ▶ Structural equation

$$y_1 = \beta_0 + \beta_1 y_2 + \beta_2 z_1 + u_1$$

- ▶ Substituting

$$y_2 = \pi_0 + \pi_1 z_1 + \pi_2 z_2 + v_2$$

- ▶ Reduced form

$$y_1 = \beta_0 + \beta_1 (\pi_0 + \pi_1 z_1 + \pi_2 z_2 + v_2) + \beta_2 z_1 + u_1$$

$$y_1 = (\beta_0 + \beta_1 \pi_0) + (\beta_1 \pi_1 + \beta_2) z_1 + \beta_1 \pi_2 z_2 + (u_1 + \beta_1 v_2)$$

- ▶ Need IV to estimate