

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

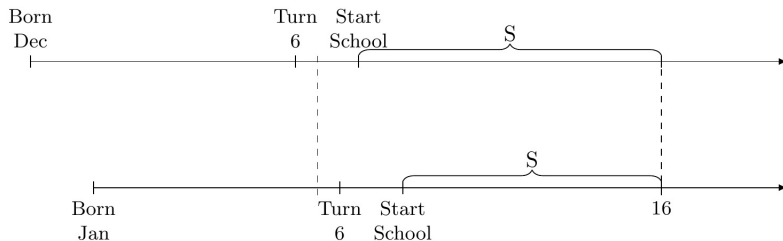
- Endogeneity
- One solution - <https://youtu.be/eoJUPd6104Q>
- Model

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

- Suppose z is a variable such that * see handwritten notes
(April 21
2 of 2)

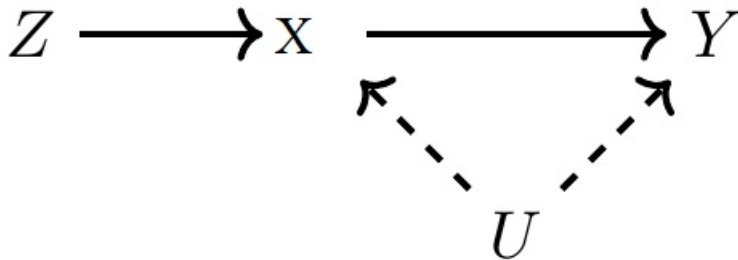
IV Estimation in a Simple Regression Model (cont.)

- 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.)

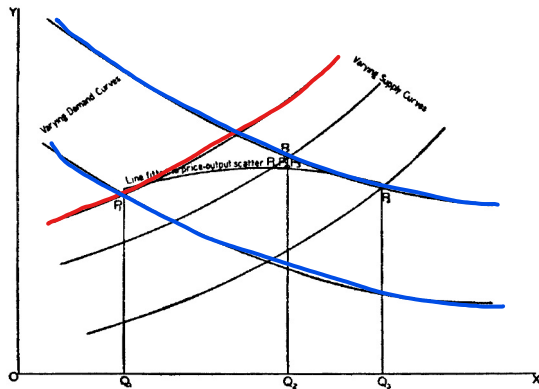
- z and x are correlated

- 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

$$y = \beta_0 + \beta_1 \overset{z}{\downarrow} x + u$$

- ▶ x and u : correlated
- ▶ z is an $\rightarrow$ instrument for x

- The instrumental variables (IV) estimator of β_1

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

when $z = x$ $\downarrow$ OLS formula

z :
e.g. charter
lottery

y : grade

x : charter
attendance

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*
- 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 $\hat{\beta}_1$

where SST_x is the total sum of squares of x

$R_{x,z}^2$: R-squared from the regression of x on z

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

$\hat{u}_i$: IV residuals

IV Estimation in a Simple Regression Model (cont.)

- Note

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

where SST_x is the total sum of squares of x

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

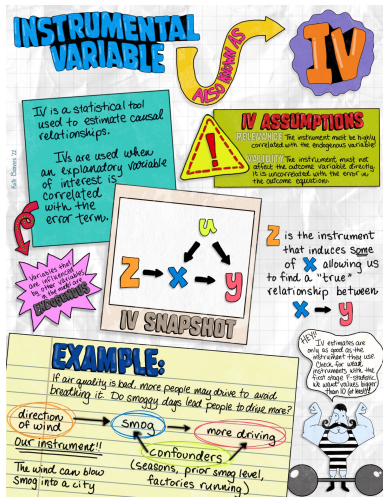
$\hat{u}_i$: OLS residuals

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

IV Estimation in a Simple Regression Model (cont.)

- 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

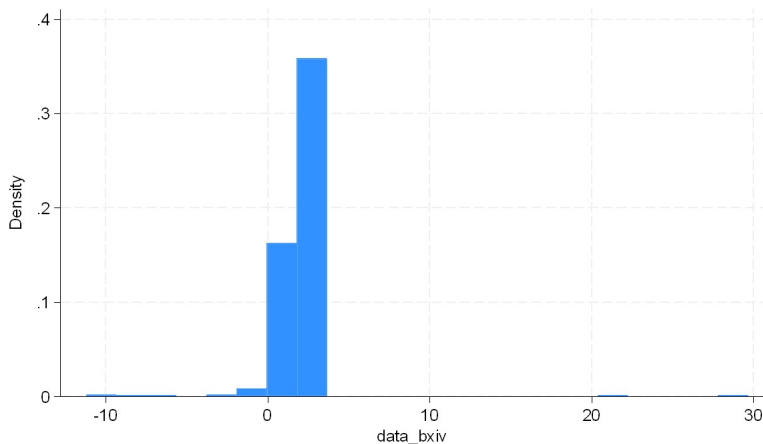
IV Estimation in a Simple Regression Model (cont.)



@Kate__Barnes

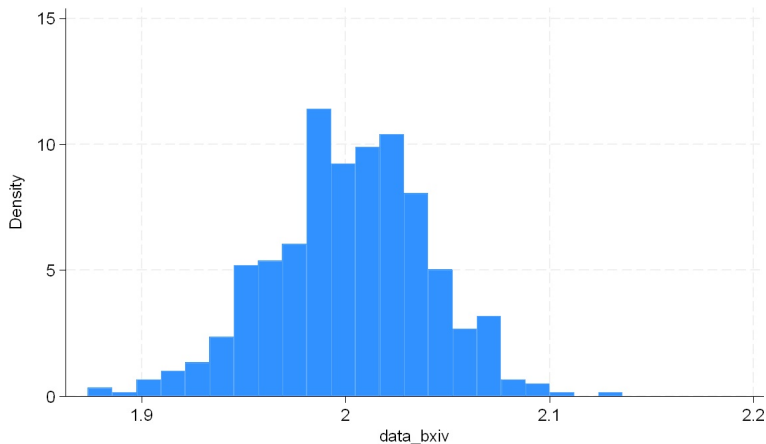
IV Estimation in a Simple Regression Model (cont.)

- $n = 10$, reps = 500, $\text{corr}(x, z) = 0.8$, $\text{corr}(x, u) = 0.5$, $\text{corr}(z, u) = 0$
- $y = 1 + 2x + u$



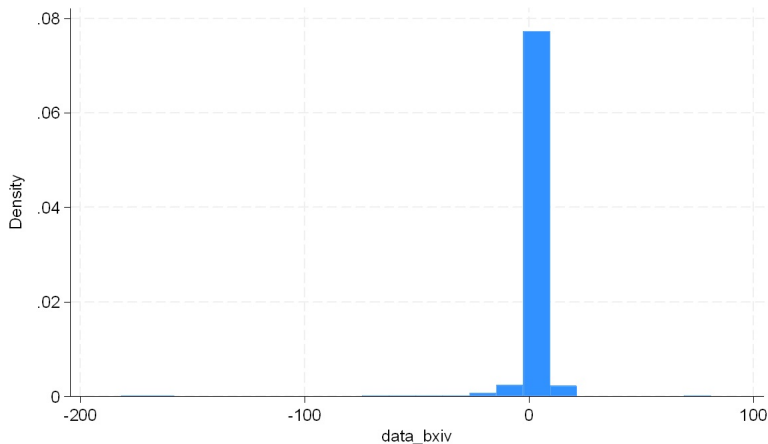
IV Estimation in a Simple Regression Model (cont.)

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



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.)

- 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