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

corr. w/ x

uncorr. w/ u

and only affects y via x

then we might still be able to estimate β_1.

x and u → correlated

x: endog.

⇒ OLS is biased

z: instrumental variable / instrument for x

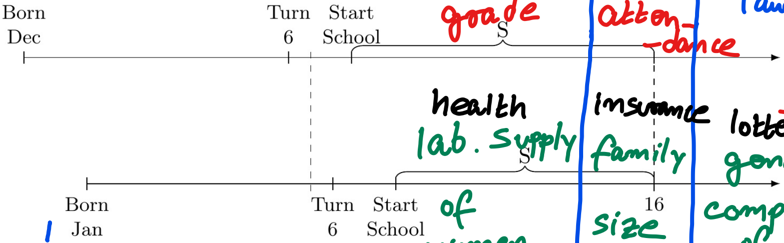
IV Estimation in a Simple Regression Model (cont.)

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

z

- Examples

y	x	z
~~wage~~ grade / wage	~~educ.~~ charter / educ.	compulsory educ. laws / lottery
grade	attendance	dist. from school
health	insurance	lottery expansion
lab. supply of women	family size	gender composition of 1st 2 children
earnings	military service	draft lottery no.
earnings	school	college in county

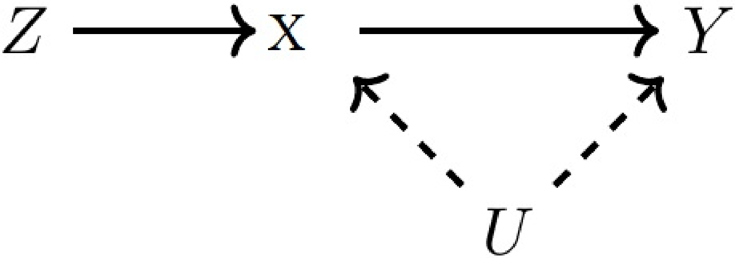


Causal Inference: The Mixtape

effect of z on x × effect of x on y ? = effect of z on y

earnings

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

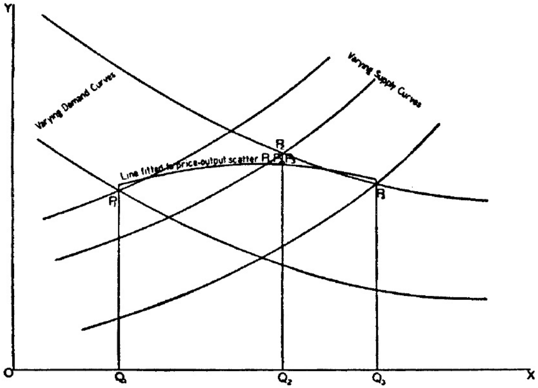
– z is related to x

– can be tested by

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

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.)

① reg x on z

↳ obtain $\widehat{x}$

- Model

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

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

② reg y on $\widehat{x}$

$\Rightarrow \widehat{\beta_1}$

- The *instrumental variables* (*IV*) *estimator* of β_1

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

e.g. y: lwage

x: educ

z: father's educ

when $z = x$

OLS formula

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

$$E(u^2|z) = \sigma^2 = Var\,(u)$$

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

$$\sqrt{\frac{\hat{\sigma}^2}{SST_x \; R^2_{x,z}}}$$

where SST_x is the total sum of squares of x
$R^2_{x,z}$: R-squared from the regression of x on z
$\hat{\sigma}^2 = \frac{1}{n-2}\sum_i \widehat{u}_i^2$
$\widehat{u}_i$: IV residuals

IV Estimation in a Simple Regression Model (cont.)

- Note
  - Standard error of $\hat{\beta}_1$ in case of OLS $\sqrt{\dfrac{\hat{\sigma}^2}{SST_x}}$

    where SST_x is the total sum of squares of x

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

    $\widehat{u}_i$: OLS residuals
  - Typically $R^2_{x,z} < 1$ and the IV std. error always > OLS std. error
  - If x and z are only slightly correlated

$\Rightarrow R^2_{x,z}$ is small

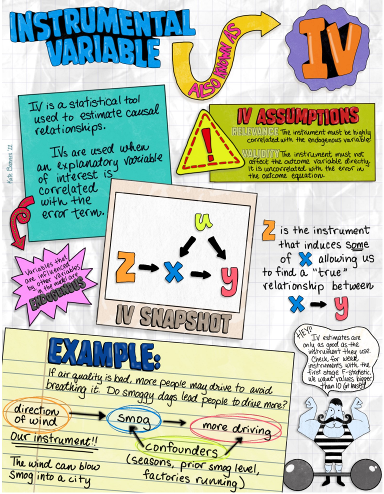
$\Rightarrow$ IV std. error

IV Estimation in a Simple Regression Model (cont.)

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

even if z & u are slightly correlated

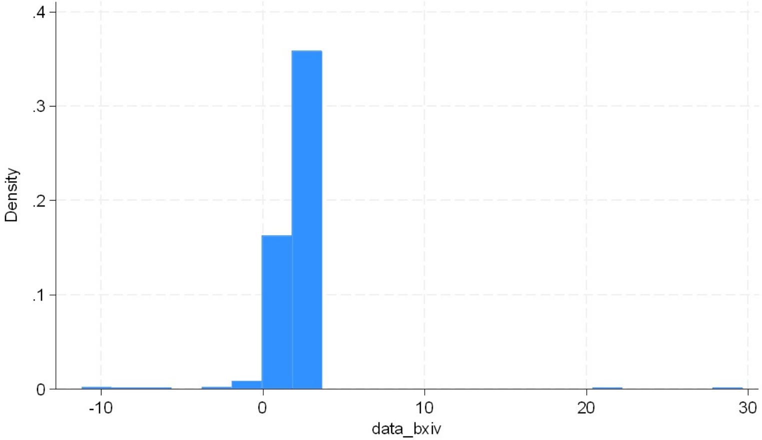
IV Estimation in a Simple Regression Model (cont.)



@Kate__Barnes

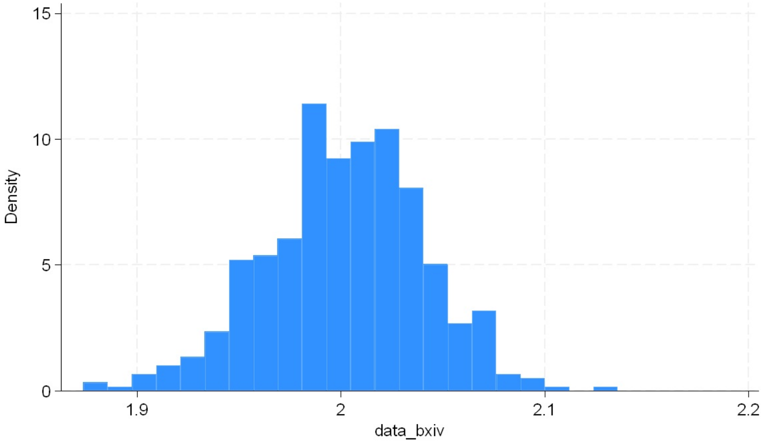
IV Estimation in a Simple Regression Model (cont.)

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



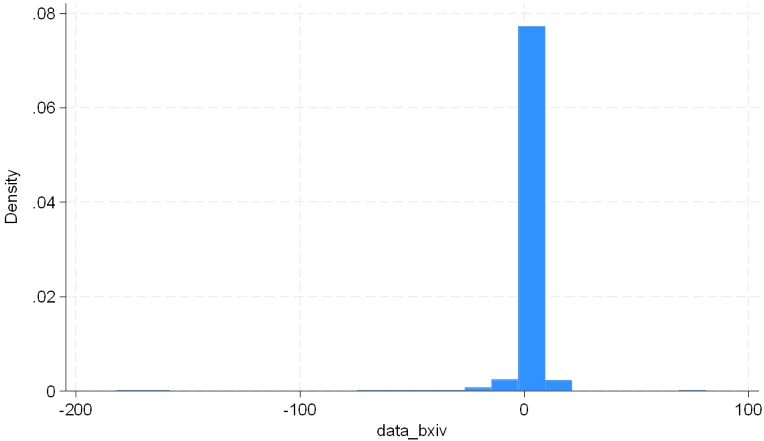
IV Estimation in a Simple Regression Model (cont.)

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



IV Estimation in a Simple Regression Model (cont.)

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



IV Estimation in a Multiple Regression Model

not in structural eq^n.

z_2

corr. w/ y_2

not corr. w/ u_1

- Model

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

  - Sometimes called structural eq^n.
  - y_1 correlated with u_1 $\Rightarrow$ endog.
  - z_1 assumed to be uncorrelated with u_1 $\Rightarrow$ exog.
  - y_2 correlated with u_1 $\Rightarrow$ endog.
  - OLS estimators: biased & inconsistent
  - z_2 instrumental variable for y_2

rainfall shocks

$$\text{conflict} = \beta_0 + \beta_1\,\text{growth} + \beta_2\,\text{rugged terrain} + u_1$$

IV Estimation in a Multiple Regression Model (cont.)

- Assumptions

$$
\begin{aligned}
E(u_1) &= 0 \\
E(z_1 u_1) &= 0 \\
E(z_2 u_1) &= 0
\end{aligned}
$$

IV Estimation in a Multiple Regression Model (cont.)

- Equations

$$
\begin{aligned}
E(y_1 - \beta_0 - \beta_1 y_2 - \beta_2 z_1) &= 0 \\
E\left[z_1(y_1 - \beta_0 - \beta_1 y_2 - \beta_2 z_1)\right] &= 0 \\
E\left[z_2(y_1 - \beta_0 - \beta_1 y_2 - \beta_2 z_1)\right] &= 0
\end{aligned}
$$

IV Estimation in a Multiple Regression Model (cont.)

- Sample analogs

$$
\begin{aligned}
n^{-1} \sum_{i=1}^{n} \left(y_{i1} - \widehat{\beta}_0 - \widehat{\beta}_1 y_{i2} - \widehat{\beta}_2 z_{i1} \right) &= 0 \\
n^{-1} \sum_{i=1}^{n} z_{i1} \left(y_{i1} - \widehat{\beta}_0 - \widehat{\beta}_1 y_{i2} - \widehat{\beta}_2 z_{i1} \right) &= 0 \\
n^{-1} \sum_{i=1}^{n} z_{i2} \left(y_{i1} - \widehat{\beta}_0 - \widehat{\beta}_1 y_{i2} - \widehat{\beta}_2 z_{i1} \right) &= 0
\end{aligned}
$$

- *Instrumental variables estimators*

IV Estimation in a Multiple Regression Model (cont.)

an endog. in terms of exog.

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

growth

terrain

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

rainfall

- Key condition

$E(v_2) = 0$

$Corr(z_1, v_2) = 0$

$Corr(z_2, v_2) = 0$

$\pi_2 \neq 0$

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
  $$\begin{aligned} y_1 &= \beta_0 + \beta_1 (\pi_0 + \pi_1 z_1 + \pi_2 z_2 + v_2) + \beta_2 z_1 + u_1 \\ y_1 &= (\beta_0 + \beta_1 \pi_0) + (\beta_1 \pi_1 + \beta_2) z_1 + \beta_1 \pi_2 z_2 + (u_1 + \beta_1 v_2) \end{aligned}$$
  - Need IV to estimate β_1 instead of $\beta_1 \pi_2$