ECON 256: Poverty, Growth & Inequality

Jack Rossbach

STATA Example for IV Regression

Suppose we observe a relationship between two variables. How do we know if it's causal?

- Lab experiments
- Randomized Assignment
- Natural Experiments
- Instrumental Variable

Example Setup

Suppose we have 1000 observations generated by the following system of equations

$$Z_i = \epsilon_i^Z$$

$$U_i = \epsilon_i^U$$

$$X_i = 0.75 \times Z_i + 2 \times U_i + \epsilon_i^X$$

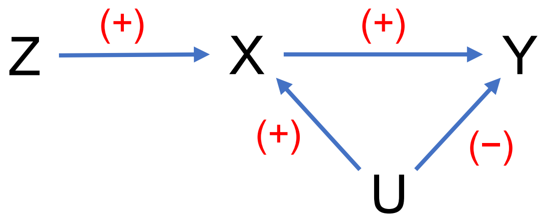
$$Y_i = 0.5 \times X_i - 4 \times U_i + \epsilon_i^Y$$

Where $\epsilon_i^Z, \epsilon_i^U, \epsilon_i^X, \epsilon_i^Y$ are all iid standard normal random variables.

STATA: [1] **set obs 1000** [2] **gen z = rnormal(0,1)** [3] **gen u = rnormal(0,1)** [4] **gen x = rnormal(0,1)+.75*z + 2*u** [5] **gen y = rnormal(0,1) -4*u + .5*x**

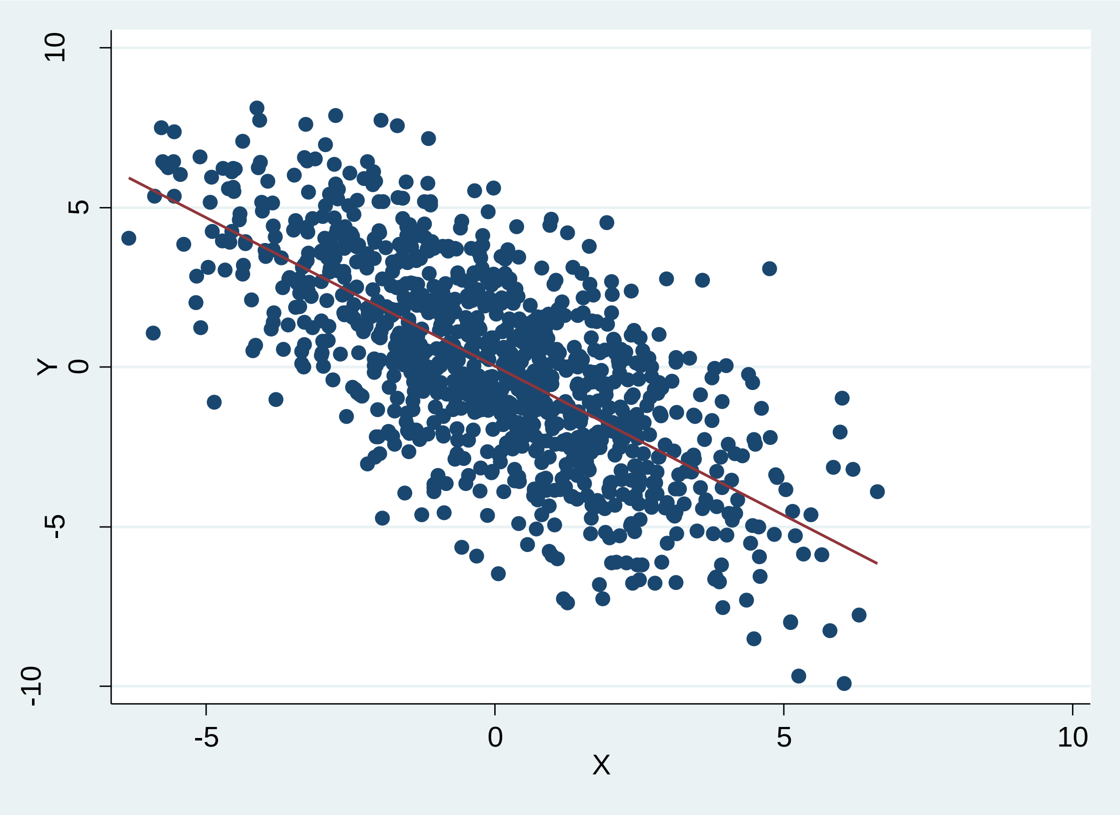
Challenge: Determining Relationship between X and Y

There is a positive causal relationship between X and Y by construction. Graphically:



- However, U affects both X and Y. It increases X while decreasing Y.
- U is a confounding factor. If we don't include U in our regression, it is an omitted variable
- If we don't have data on U, our estimated coefficient of X will be badly biased. In this case, it is biased enough to make it look like X decreases Y

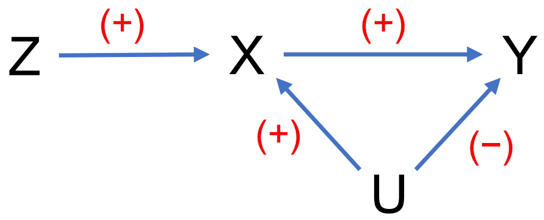
Scatter Plot between X and Y



STATA: **twoway scatter y x**

Simple Regression between X and Y

There is a positive causal relationship between X and Y by construction



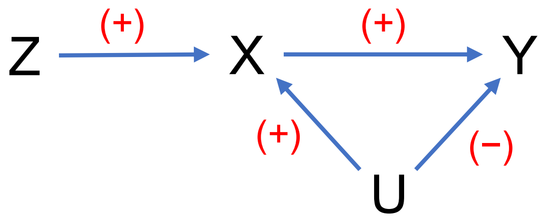
- If we run a simple regression, $Y_i = \alpha + \beta X + \epsilon_i$, our estimated β coefficient is negative and significant. STATA: **reg y x** (case sensitive, in code I stick with lower case)

- We know this is wrong in a causal sense, since β should be 0.5 by construction.

- If we include data on U in regression, everything works out fine. STATA: **reg y x u**

Instrumental Variable (IV) Approach

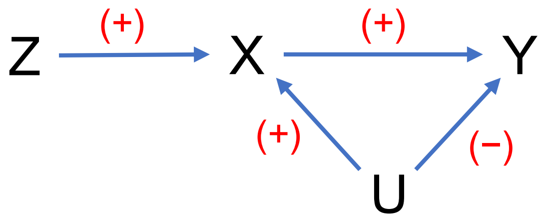
There is a positive causal relationship between X and Y by construction



- **What if we lack data on U?** Can still use Instrumental Variable approach.
- Z affects X and therefore affects Y only through its effect on X. We can Z as an IV for X.
- It is important Z does not have a direct effect on Y, this is called the exclusion restriction.

Instrumental Variable (IV) Approach

There is a positive causal relationship between X and Y by construction

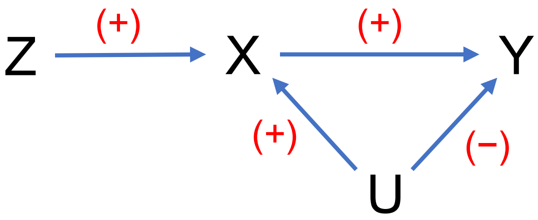


The IV Regression is three steps:

1. Regress X on Z (i.e. $X_i = \alpha + \beta Z_i + \epsilon_i$). STATA: **reg x z**
2. Use regression to predict values of X for each value of Z ($\hat{X}_i = \alpha + \beta Z_i$). STATA: **predict xpred if e(sample)** [have to have above be most recent regression run]
3. Regress Y on $\hat{X}$ (i.e. $Y_i = \alpha + \beta \hat{X}_i + \epsilon$) instead of on X. STATA: **reg Y X**

Instrumental Variable (IV) Approach

There is a positive causal relationship between X and Y by construction



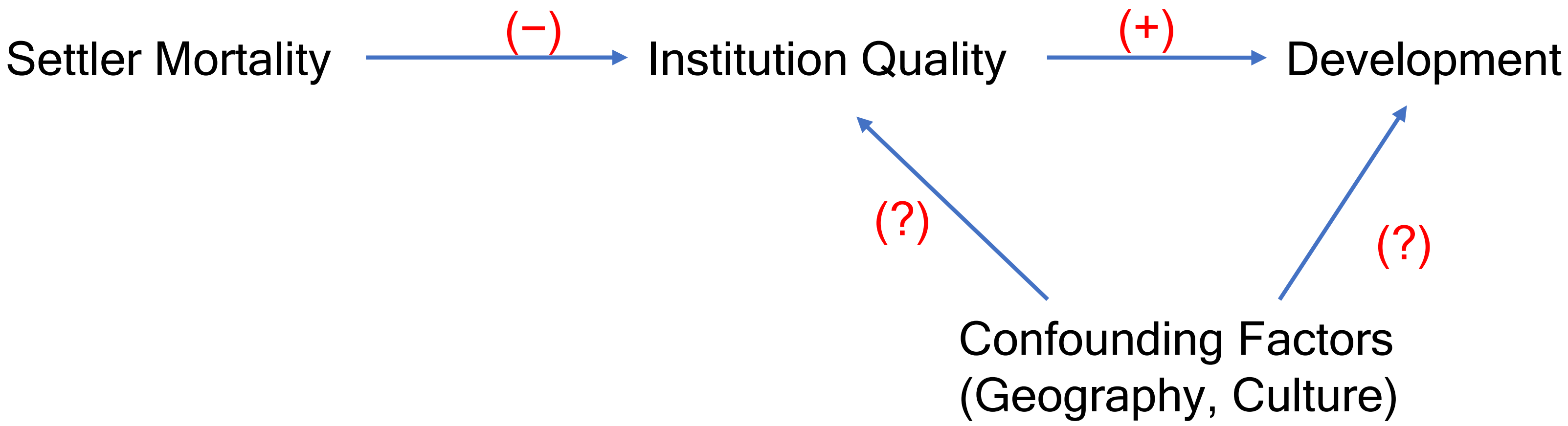
When we do the IV approach, we find a estimated coefficient near 0.5. So the IV approach successfully recovers the causal relationship between X and Y, even without data on U.

- It's not necessary to do all three steps by hand. What we did is called 2SLS (Two Stage Least Squares Regression). STATA: **ivregress 2sls y (x=z)**

Settler Mortality IV Approach

We use Settler Mortality as an IV for whether institutions are inclusive or extractive

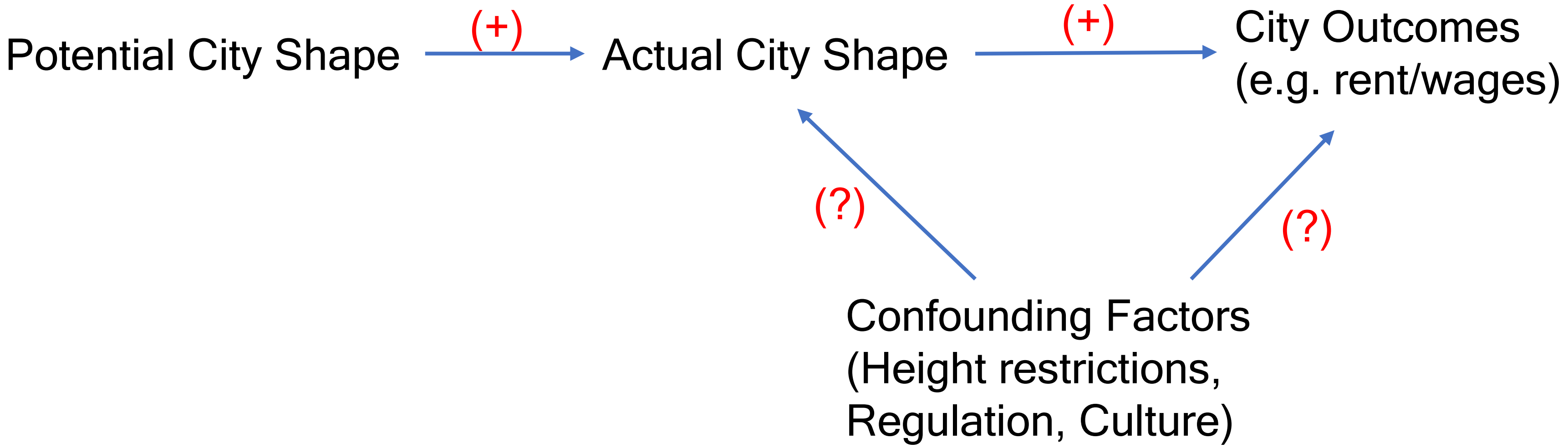
- High Settler mortality reduced the likelihood and feasibility of colonizers setting up inclusive institutions.



Examples of IV Approaches in Practice

We discussed one in context of geography and how city shape affects city outcomes.

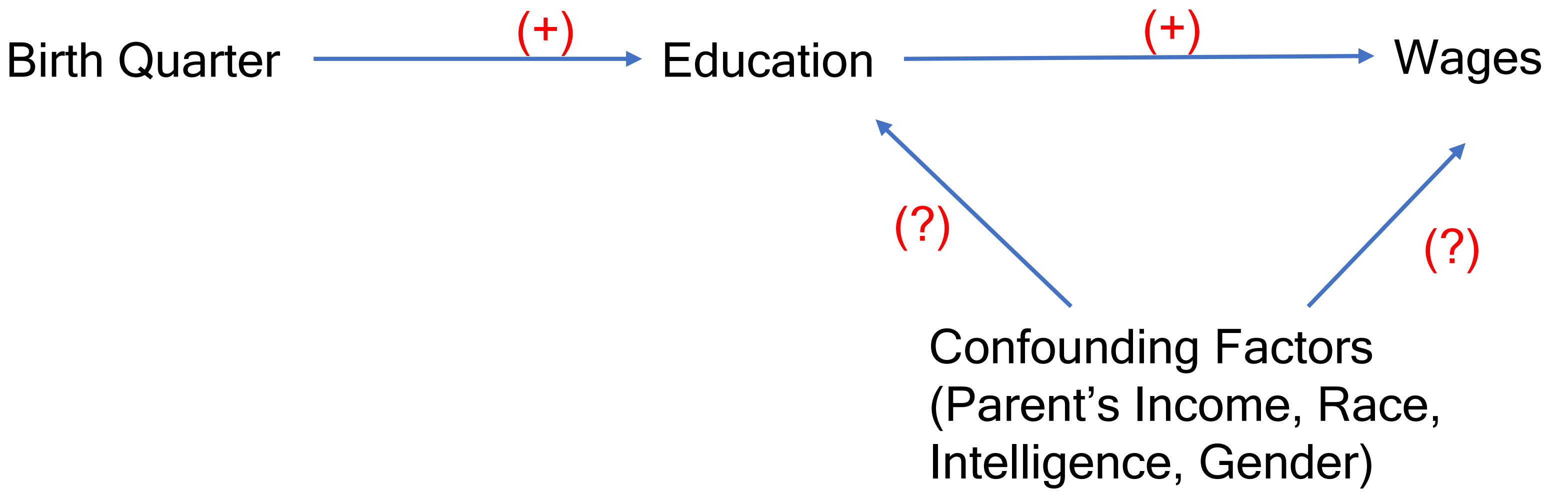
- Harari (2015) used Potential City Shape (water/mountains/etc affect whether city can be compact) as an IV for Actual City Shape.



Examples of IV Approaches in Practice

[Angrist and Krueger (1991)](#) used Birth Quarter to study affects of Education on Wages

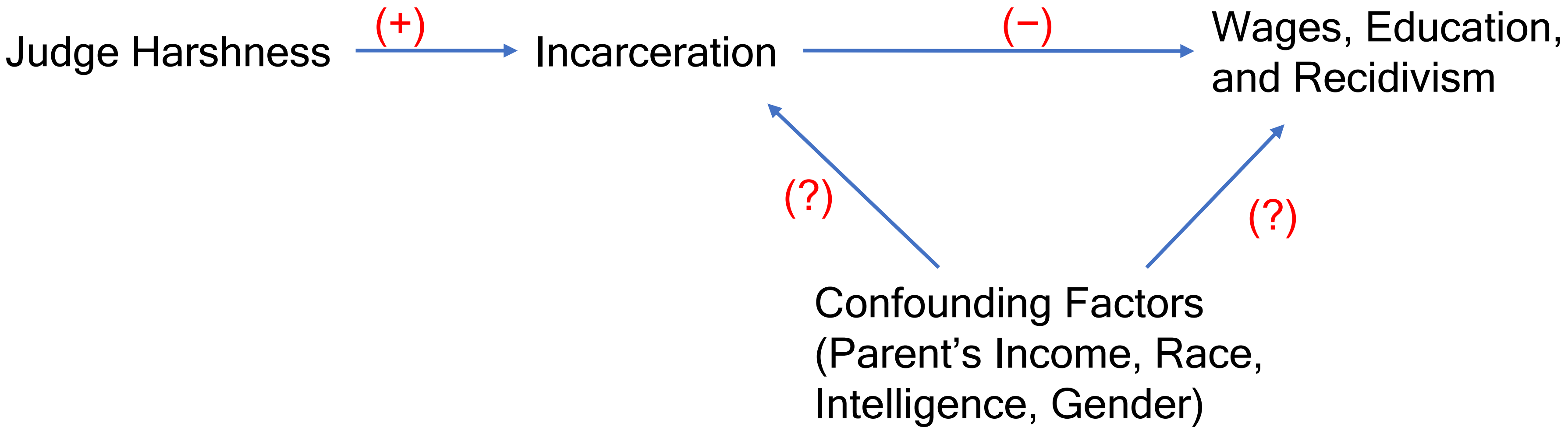
- Can’t drop out until 16 years old. School starts in August/September, depending when birthday is affects how many more months you need to stay in school.



Examples of IV Approaches in Practice

Aizer and Doyle (2015) use Judge Assignment as IV for Juvenile Incarceration

- Judge assignment is random, judges differ in how much they favor incarceration on average. That affects whether people are incarcerated.Find juvenile incarceration has long term negative impact on wages, education, and recidivism/adult incarceration.



Reading STATA Output

$Y_i = \alpha + \beta X_i + \epsilon_i$

```
. reg y x
```

Source	SS	df	MS
Model	4631.8089	1	4631.8089
Residual	5181.91594	998	5.19230055
Total	9813.72485	999	9.82354839

Number of Observations in Regression (Observations with missing data are excluded)

Number of obs =	1000
F(1, 998) =	892.05
Prob > F =	0.0000
R-squared =	0.4720
Adj R-squared =	0.4714
Root MSE =	2.2787

Coefficient of Determination
Wiki Link

Standard Error: Used to P Value and Confidence Interval

y	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
x	-.9328223	.0312323	-29.87	0.000	-.9941108	-.8715339
_cons	.0249781	.0720646	0.35	0.729	-.1164374	.1663937

Estimated Coefficients
$\hat{Y} = 0.024 - 0.932 \times X$

P Value (Statistically Significant if <0.05)

Confidence Interval (Statistically Significant if excludes zero)