Instrumental Variables

Econ 140 Spring 2025, Section 7

Jonathan Old

Syllabus/OH bCourses Website Feedback form (**Always open**) RStudio

Roadmap

1. Time for your questions
2. (Group) data projects
3. Introduction to Instrumental Variables
4. IV Conditions
5. IV Summary
6. Time for your questions
7. Group work
8. Time for your questions

Your questions

Any questions?

...Remember – this is a safe space! Every question is useful!

(Group) data projects

Your time for questions

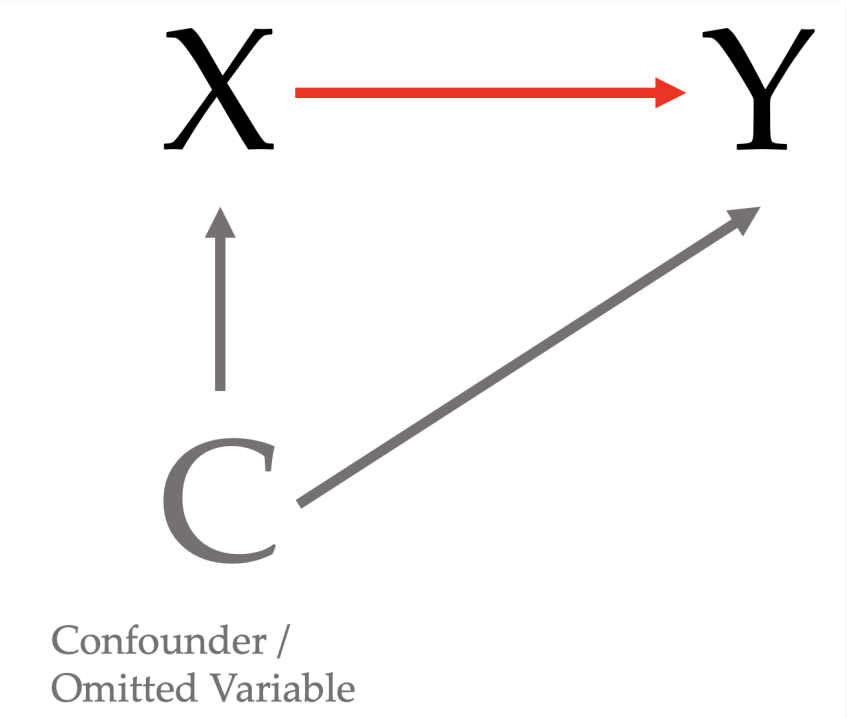
I prepared a document that can help you find data if you are lost. See it on bCourses.

The first deadline is on April 5.

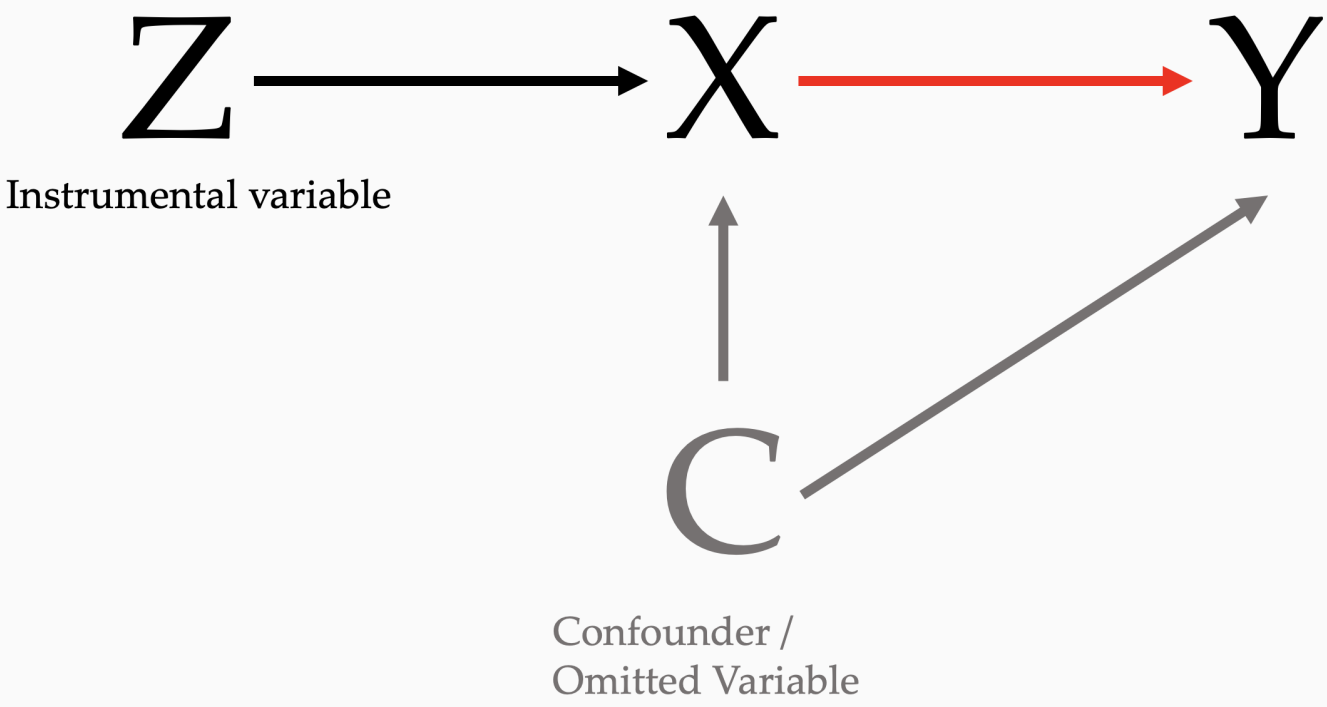
Group work encouraged!

Introduction to Instrumental Variables

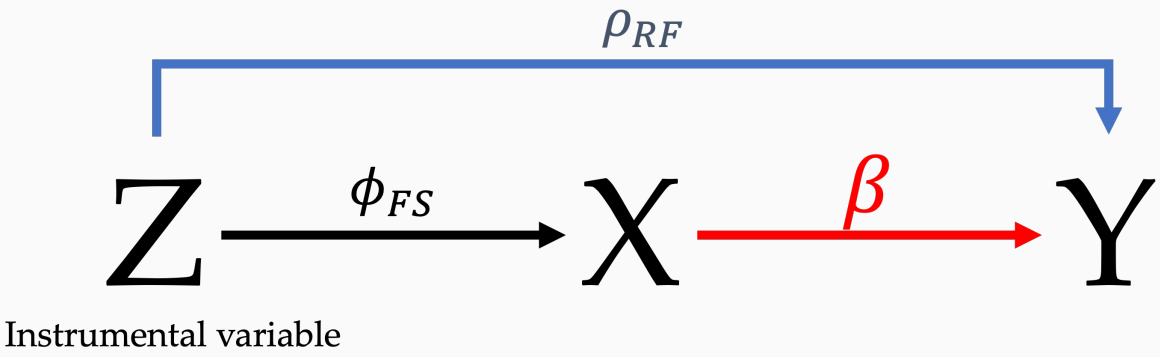
Recap: Omitted Variable Bias



Instrumental variables: The setup



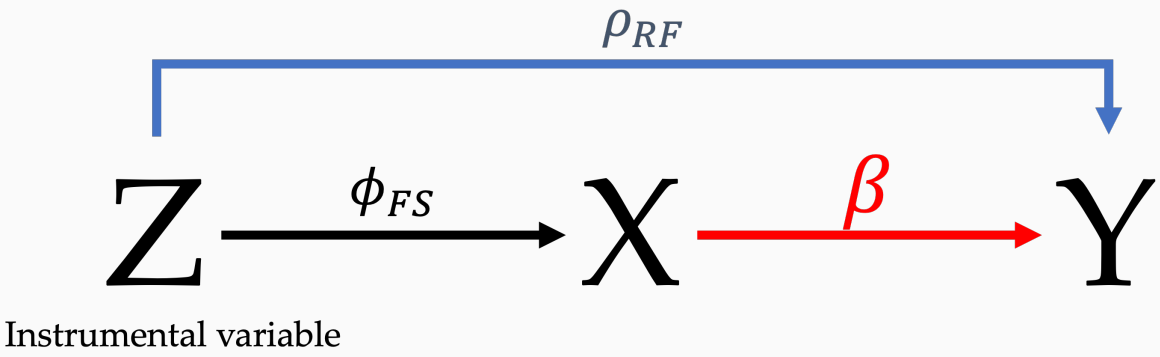
Recap: IV ”rescales” the effect



A simple example:

- We want to know the effect of chocolate (X) on happiness (Y), using a randomized voucher as instrument (Z).
- We find: people with voucher were 3 points more happy ($\rho = 3$), and ate 0.5 more chocolates ($\phi = 0.5$).
- Then, the effect of eating one more chocolate is:

Recap: IV "rescales" the effect



A simple example:

- We want to know the effect of chocolate (X) on happiness (Y), using a randomized voucher as instrument (Z).
- We find: people with voucher were 3 points more happy ($\rho = 3$), and ate 0.5 more chocolates ($\phi = 0.5$).
- Then, the effect of eating one more chocolate is: $\beta = \rho/\phi = 3/0.5 = 6$.

Calculating the IV coefficient

What is the effect of **eating chocolate** (D) on happiness (Y).

- Why not estimate: $Y_i = \alpha + \beta D_i + \varepsilon_i$?

Randomly give voucher to buy chocolate at 90% discount (Z).

- Why not estimate: $Y_i = \alpha + \beta Z_i + \varepsilon_i$?

Let us set up some regressions:

$$\text{Regression of interest: } Y_i = \alpha + \beta D_i + e_i$$

$$\text{First stage: } D_i = \alpha_1 + \phi_{FS} Z_i + u_i$$

$$\text{Reduced Form: } Y_i = \alpha_2 + \rho_{RF} Z_i + v_i$$

$$\text{Plug in regression of interest: } Y_i = \alpha + \beta(\alpha_1 + \phi_{FS} \cdot Z_i + u_i) + e_i$$

$$\text{Get back reduced form: } = \underbrace{(\alpha + \beta\alpha_1)}_{\alpha_2} + \underbrace{(\beta\gamma)}_{\rho_{RF}} Z_i + \underbrace{(\beta u_i + e_i)}_{v_i}$$

So we see that $\rho_{RF} = \beta\phi_{FS} \Leftrightarrow \beta = \rho_{RF}/\phi_{FS}$

Interpretation of the IV coefficient

- How do we interpret ϕ_{FS} ?

Interpretation of the IV coefficient

- How do we interpret ϕ_{FS} ? The average difference in chocolate consumption between those who got a voucher and those who didn't

Interpretation of the IV coefficient

- How do we interpret ϕ_{FS} ? The average difference in chocolate consumption between those who got a voucher and those who didn't
- How do we interpret ρ_{RF} ?

Interpretation of the IV coefficient

- How do we interpret ϕ_{FS} ? The average difference in chocolate consumption between those who got a voucher and those who didn't
- How do we interpret ρ_{RF} ? The average difference in happiness between those who got a voucher and those who didn't

Interpretation of the IV coefficient

- How do we interpret ϕ_{FS} ? The average difference in chocolate consumption between those who got a voucher and those who didn't
- How do we interpret ρ_{RF} ? The average difference in happiness between those who got a voucher and those who didn't

$$\beta = \frac{\rho_{RF}}{\phi_{FS}} = \frac{E\left[Y_i \mid Z_i = 1\right] - E\left[Y_i \mid Z_i = 0\right]}{E\left[D_i \mid Z_i = 1\right] - E\left[D_i \mid Z_i = 0\right]}$$

IV gives us the treatment effect for the compliers

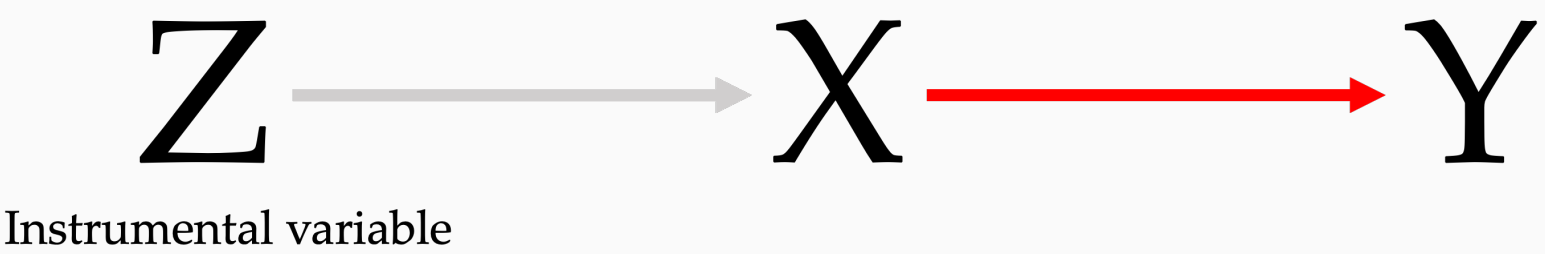
Potential outcomes! (unobserved)		***Does not get voucher (Z=0)***	
Gets voucher (Z=1)		*Eats chocolate (D=1)*	*Does not eat chocolate (D=0)*
	Eats chocolate (D=1)	Always-takers: E(D|Z=1)=E(D|Z=0)=1 ➔ E(Y|Z=1)=E(Y|Z=0)	Compliers
	Does not eat chocolate (D=0)	Defiers	Never-takers: E(D|Z=1)=E(D|Z=0)=0 ➔ E(Y|Z=1)=E(Y|Z=0)

IV Conditions

Does IV always work?

- No! It only works if we have a valid instrument
- For this, we need three conditions:

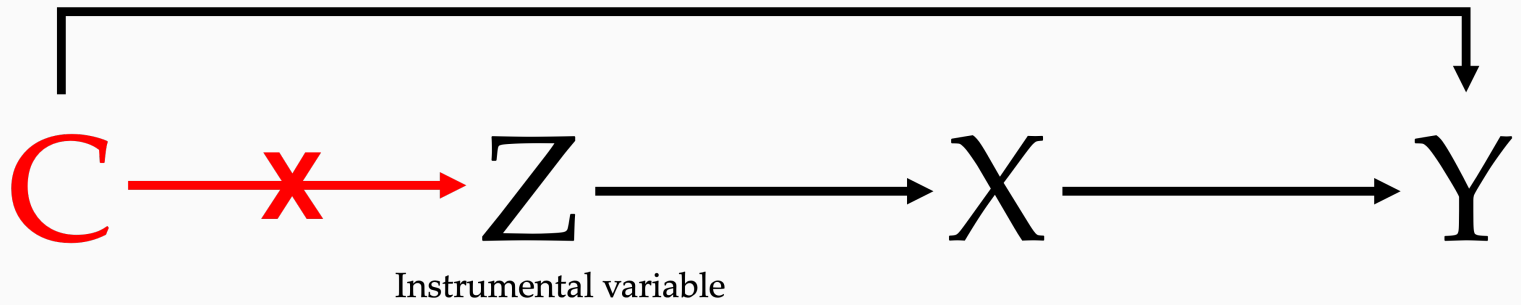
1. Relevance: Z must truly affect X



Does IV always work?

- No! It only works if we have a valid instrument
- For this, we need three conditions:

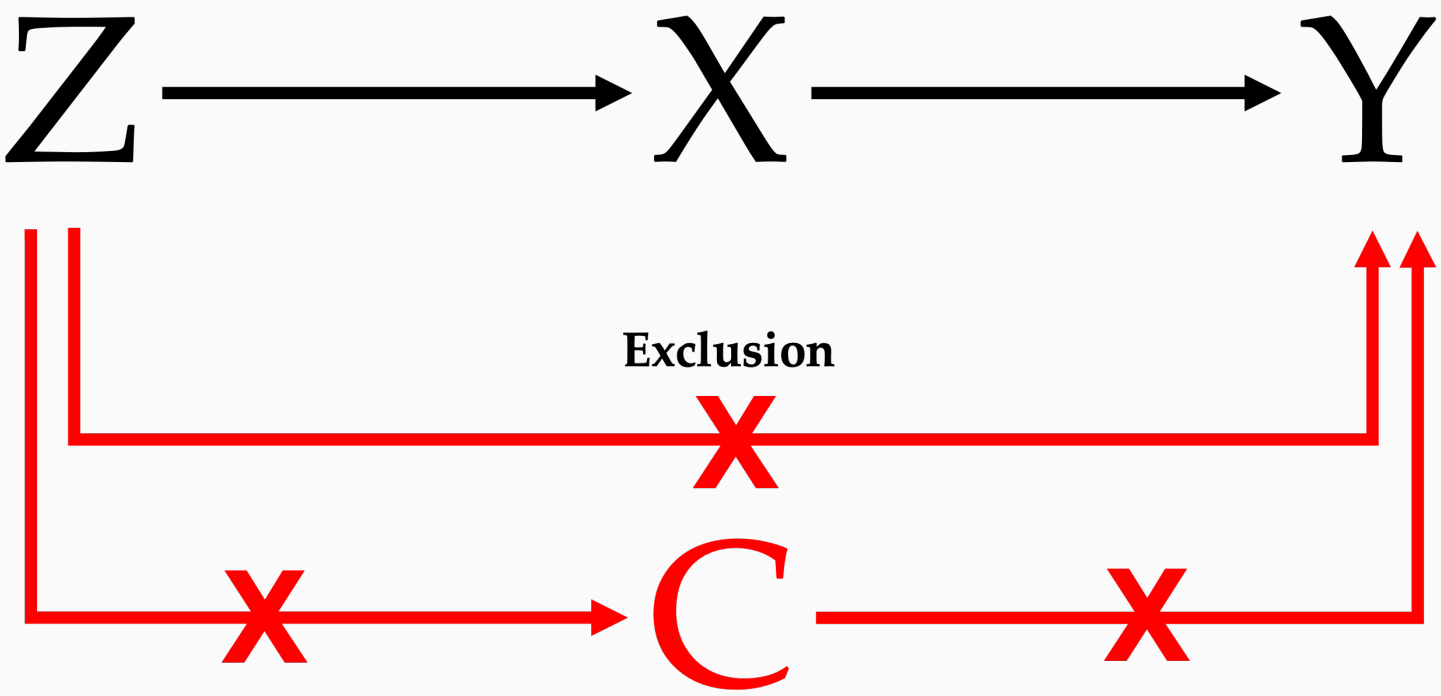
2. Independence/Exogeneity: Z is as good as randomly assigned



Does IV always work?

- No! It only works if we have a valid instrument
- For this, we need three conditions:

3. Exclusion: The **ONLY** way that *Z* affects *Y* is via *X*!

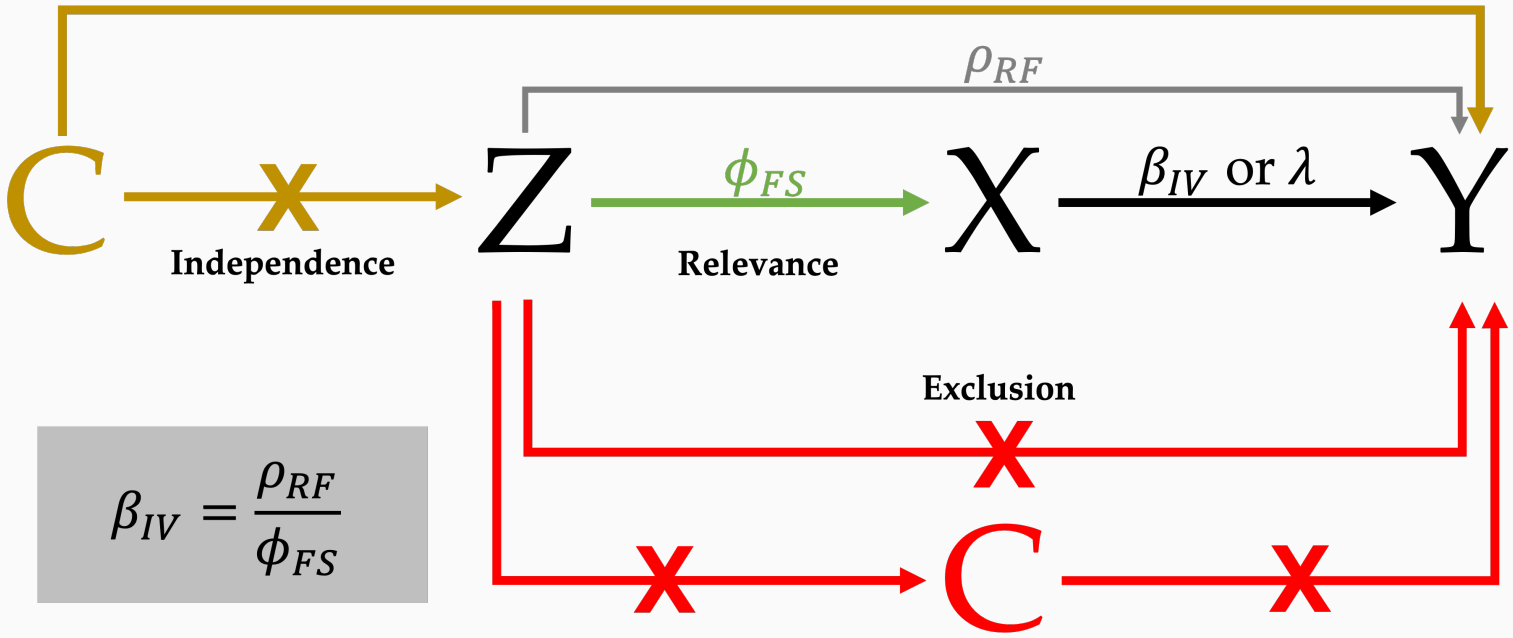


IV Summary

IV summary

We need the following three assumptions for IV to work:

1. **Relevance**: Z must truly affect X
2. **Independence/Exogeneity**: Z is as good as randomly assigned
3. **Exclusion Restriction**: The **only** way that Z affects Y is via X.



$$\beta_{IV} = \frac{\rho_{RF}}{\phi_{FS}}$$

How IV estimates can be different from OLS estimates

- **OVB:** If OLS had omitted variable bias, and our instrument is valid, then the IV estimate should be different. We can check whether this is plausible with the OVB formula
- **Measurement error:** If we have random measurement error in the independent variable (X), then we can use IV to overcome this. In that case, the IV coefficient will be larger (in absolute value) than the OLS coefficient
- **LATE:** OLS gives us ATT + Selection Bias, while IV gives us the treatment effect on the compliers (LATE). The ATT may be different from LATE, even without selection bias.
- **Invalid IV:** Hard to determine what exactly is going on
- **Sampling variation**: This can just happen by chance

IV and controls

- There is a catch in IV: independence (exogeneity) and exclusion restriction only need to hold **conditionally**
- This means: If we **know** C, we can control for it in the IV regression, and the concerns go away
- But then, we have the old OLS problems again (we maybe need to control for lots of things)
- So we should only do this in cases where we really know all the C's.

Your questions

Any questions?

...Remember – this is a safe space! Every question is useful!

Group work

Group work

Group 1: We are interested in the effect of being in the army on crime. We instrument being in the army with a lottery (paper)

Group 2: We are interested in the effect of income on conflict. We instrument income with rainfall (paper)

Group 3: We are interested in the effect of air pollution on mortality. We instrument local air pollution with wind direction (paper)

1. Relevance: Z must truly affect X
2. Independence/Exogeneity: Z is as good as randomly assigned
3. Exclusion restriction: The **only** way that Z affects Y is via X

Your job: Discuss whether these assumptions hold!

Your questions

Any questions?

...Remember – this is a safe space! Every question is useful!