The intuition behind instrumental variables

Prof. Alex COAD

Instrumental Variables equations: The standard econometric approach

- $y = \alpha + \beta x + u$

- Endogeneity between y and x
  - Example: y = earnings, x = education
- β cannot be interpreted as the causal effect of x on y
  - Endogeneity: x correlated with u

- Instrumental variable: z
- Exclusion restriction: z is correlated with x but uncorrelated with u
  - NB: you can't actually check the exclusion restriction, because true u is unobserved!

2-stage least squares

- **First stage:** Regress x on z
- $x = \delta z + \epsilon$
- δ should be big and significant
- F-statistic should be large (>10)
- Generate $\hat{x}$

- **Second stage:** Use $\hat{x}$
- $y = \alpha_1 + \beta_1 \hat{x} + \varepsilon$

Problems discussing causality with equations

- $y = \beta x + u$
  - Therefore $\frac{dy}{dx} = \beta$?
  - Change x by 1 unit, and y changes by β?
- But we can rearrange to get $x = \frac{y-u}{\beta}$?
  - Therefore $\frac{dx}{dy} = \frac{1}{\beta}$?
  - But $\frac{dx}{dy} = 0$ if y does not cause x, right?

- Does "=" denote association or causality?..

Causal graphs are much clearer

(Pearl, 2009)

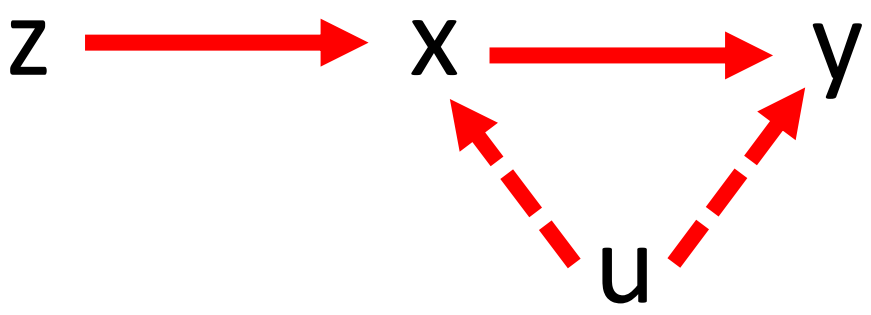
- x and y are endogenous
- z causes x
  - The first stage must be causal (this was not clear from the equations)

- REQUIREMENTS OF THE INSTRUMENTAL VARIABLE Z

1) Any correlation between z and x is a causal effect $z \rightarrow x$
2) Any correlation between z and y is a causal effect $z \rightarrow y$
   - It cannot be y that is causing z
   - Any causal effect of z on y operates only through x

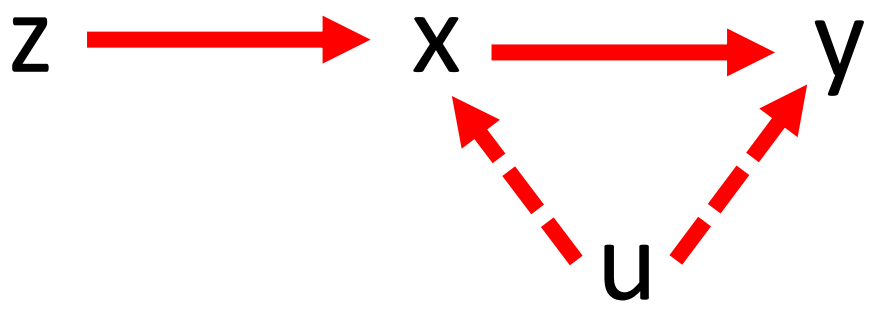
- Knowing 1) and 2), we can deduce the effect $x \rightarrow y$

- u is unobserved and causes both x and y



IV requires prior causal knowledge

- Prior causal knowledge
- New causal knowledge from IV



IV assumptions

- z is correlated with x but uncorrelated with u
  - Not a useful definition!
- Better: z causes x, and z only causes y via x

- IV only works if you bring your own causal knowledge
- “no causality in, no causality out”

- If z is correlated with x, this can only mean one of four things:

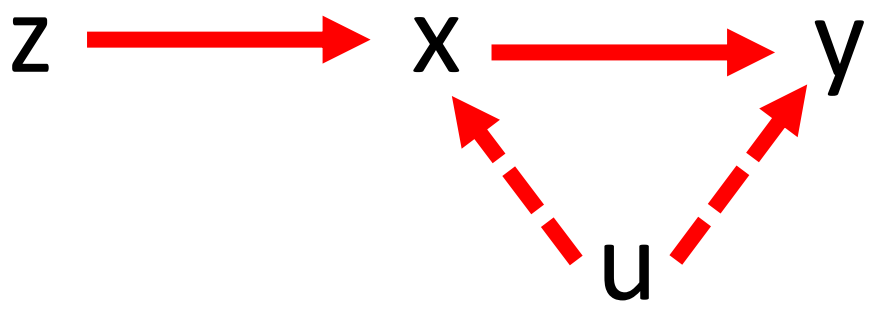
1) z→x
2) z←x
3) z←ω→x
4) z and x are related only due to selection bias

IV reasoning applied to an RCT

(Randomized Controlled Trial)

- Individuals are split between Treatment and Control groups on the basis of a random number

- z: random number
- x: group (treatment vs control)
- y: outcome



- z satisfies the two requirements of an IV

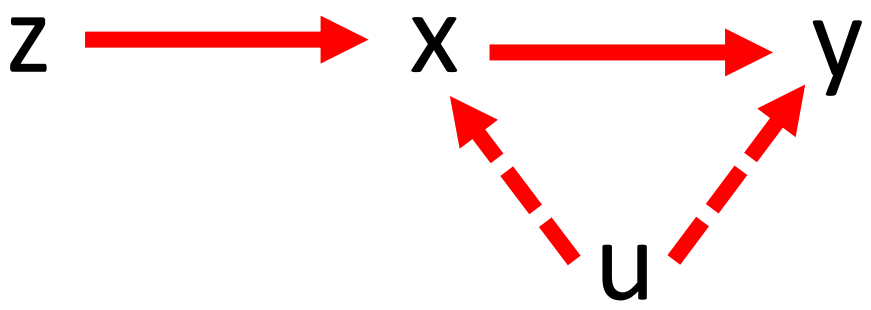
1) First stage: z causes x: z → x
2) Exclusion restriction: z causes y exclusively via x
   - z is just an administrative number and does not affect health outcomes except via x
   - Any causal effect of z on y operates only through x

IV reasoning applied to RDD

(Regression Discontinuity Design)

- Applicants are split between Treatment and Control groups on the basis of their rank score
  - There is negligible difference between applicants just above the threshold, and applicants just below

- z: ranking above the threshold (yes/no)
- x: group (grant recipients vs control)
- y: outcome (performance)



- z satisfies the two requirements of an IV

1) First stage: z causes x: z → x
2) Exclusion restriction: z causes y exclusively via x
   - z is just an internal bureaucratic score and does not affect performance except via x
   - Any causal effect of z on y operates only through x

References

- Coad, A. (2021). Econometrics and the growth of firms: perspectives from evolutionary economics. Strategy Science, 6 (4), 338–352. https://doi.org/10.1287/stsc.2021.0132
- Cunningham (2021). Causal inference: the mixtape. Yale University Press. Free to read online: https://mixtape.scunning.com/index.html
- Pearl, J. (2009) Causality: Models, Reasoning and Inference, 2nd edition. Cambridge: Cambridge University Press.
- Peters J., Janzing D., Scholkopf B., (2017). Elements of Causal Inference: Foundations and Learning Algorithms. MIT Press.