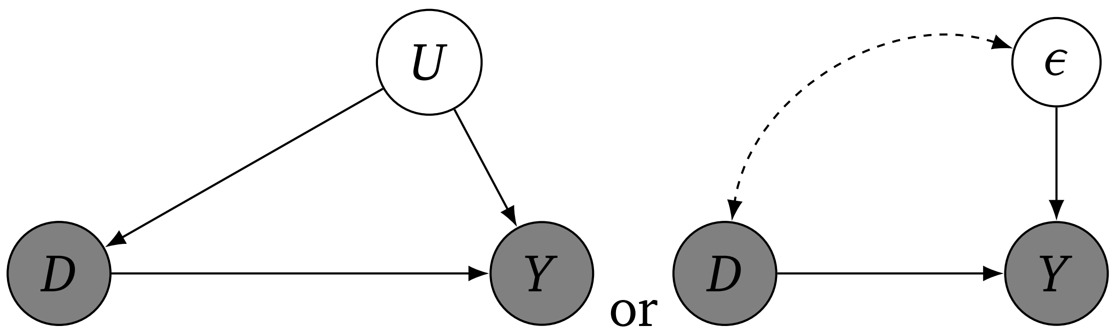
Causal inference,
Instrumental variables

Daniel Stegmueller

Duke University

The problem

- We are in the "selection on unobservables" world
- Expressed as DAG



The (classical) logic of instrumentation

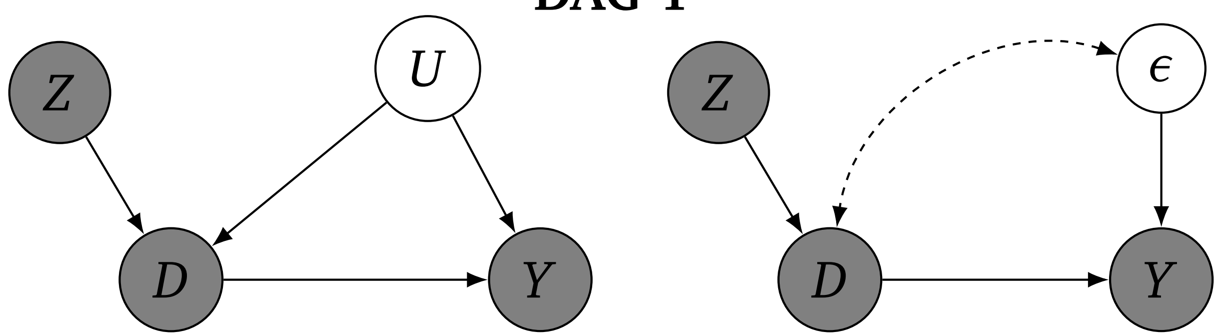
- Effect estimation with an instrumental variable
- We are interested in the following causal regression with a binary treatment and a constant causal effect δ

$$Y = \alpha + \delta D + \epsilon$$

(note the linearity here, we are saying that $f_Y(D, \epsilon_D)$ is given by $\alpha + \delta D + \epsilon$)

- Suppose there exists a binary variable Z that shapes the probability $P(D = 1)$

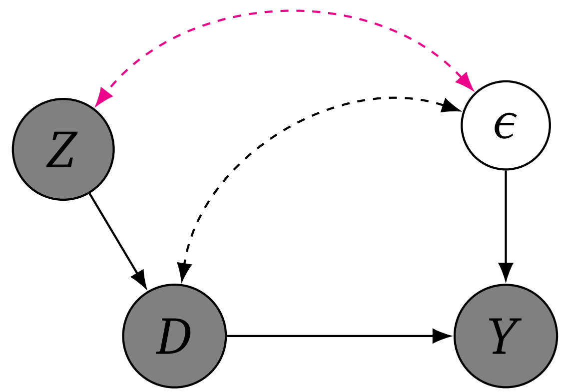
DAG 1



- Under both versions of **DAG 1** above, we can use variation in Z to examine covariation between D and Y that is causal

- However, under this DAG this is not the case

DAG 2



- The implication of these two different graphs on estimates becomes clear when we consider the population expectation of our causal regression equation

$$E[Y] = \alpha + \delta E[D] + E[\epsilon]$$

- We can rewrite this as a difference for the two levels of Z:

$$E[Y|Z=1] - E[Y|Z=0] =$$

$$\delta(E[D|Z=1] - E[D|Z=0]) + (E[\epsilon|Z=1] - E[\epsilon|Z=0])$$

- This is the variation in Y, D, and ϵ that exists due to changes in Z

- Divide both sides by $E[D|Z=1]-E[D|Z=0]$. This yields

$$\frac{E[Y|Z=1]-E[Y|Z=0]}{E[D|Z=1]-E[D|Z=0]} =$$

$$\frac{\delta(E[D|Z=1]-E[D|Z=0])+(E[\epsilon|Z=1]-E[\epsilon|Z=0])}{E[D|Z=1]-E[D|Z=0]}$$

- The difference between the two graphs is $E[\epsilon|Z=1]-E[\epsilon|Z=0]$
  - For DAG 1, $E[\epsilon|Z=1]-E[\epsilon|Z=0]=0$
  - For DAG 2, $E[\epsilon|Z=1]-E[\epsilon|Z=0]\neq 0$
- Thus, under DAG 1 we get

$$\frac{E[Y|Z=1]-E[Y|Z=0]}{E[D|Z=1]-E[D|Z=0]} = \delta$$

- That is, the ratio of the population-level linear association between Y and Z and D and Z is equal to the causal effect of $D \rightarrow Y$

- $E[\epsilon|Z=1]-E[\epsilon|Z=0]=0$ is an *exclusion restriction*.
- When it holds, the famous Wald estimator of the linear IV model estimates δ (using large samples)

$$\frac{E_N[y_i|z_i=1]-E_N[y_i|z_i=0]}{E_N[d_i|z_i=1]-E_N[d_i|z_i=0]}=\hat{\delta}_{IV}$$

written in the usual covariance ratio form, this is $Cov(y_i,z_i)/Cov(z_i,d_i)$ (and z can be non-binary)

- Remember that we assumed a constant δ !! (we will return to this shortly)
- If the exclusion restriction does *not* hold, IV estimates a *different quantity*, namely

$$\delta+\frac{E[\epsilon|Z=1]-E[\epsilon|Z=0]}{E[D|Z=1]-E[D|Z=0]}$$

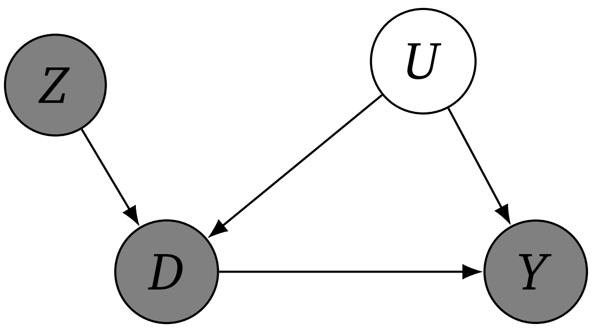
The exclusion is not testable

- It is important to remember that $E[\epsilon|Z=1] - E[\epsilon|Z=0] = 0$ involves an unobservable
- Another (intuitive) way to think about restriction is that Z has no relationship with Y except for the path $Z \rightarrow D \rightarrow Y$
- The exclusion restriction is not empirically testable!!
- Its validity depends on your defense of your (implied) DAG
- Often difficult to establish even when the instrument is generated by randomization Draft lottery example

The exclusion is not testable

An invalid test

- The belief, sometimes still encountered, that testing for an association between Z and Y conditional on D tests the exclusion (i.e.: "regress Y on Z and Y on Z controlling for D") is incorrect
- Consider our previous DAG:



  - Z influences D
  - Z influences D only via $Z \rightarrow D \rightarrow Y$
  - Thus, Z is a valid instrument

- D is a collider = "controlling" for it *creates* a dependency between Z and U (and Z Y for at least one level of D) even though Z is a valid instrument

IV estimators can have high variance

- The IV design uses only a fraction of the covariance between treatment and outcome
- Loss of statistical power leads to high variability of IV estimates
- This problem gets exacerbated when instruments are *weak*, i.e., Z predicts Y rather weakly
  - point estimates are obtained even for invalid IVs (that do not predict D)
  - standard errors are artificially small
  - There is an inverse relationship between IV strength and the bias induced by violated exclusion restrictions

$$\frac{E[\epsilon|Z=1]-E[\epsilon|Z=0]}{E[D|Z=1]-E[D|Z=0]}$$

- Common guidance
  - show “first stage” (i.e., $Z \rightarrow D$) estimates
  - show F statistic (“rules of thumb” for the size of F are in flux...)

The Local Average Treatment Effect

- So far, we have maintained that δ is constant over individuals
- What if individual-level causal effects are heterogeneous (as in our definition of the causal effects in previous weeks)? What parameter does classical IV estimation identify?
- Write the Wald estimator using potential outcomes (this goes back to the specification of Heckman and Robb 1985, 1986). Take $Y = DY_1 + (1-D)Y_0$ and rearrange and relabel the terms:

$$\begin{aligned} Y &= Y_0 + (Y_1 - Y_0)D \\ &= Y_0 + \delta D \\ &= \mu_0 + \delta D + \upsilon_0 \end{aligned}$$

- Note:
  - $\mu_0 \equiv E[Y_0]$ and $\upsilon_0 \equiv Y_0 - E[Y_0]$
  - $\delta \equiv Y_1 - Y_0$, i.e., it is *not* constant over individuals

- In an influential article, Imbens and Angrist propose a classification of individuals / cases based on their response to a shift in the instrument (say, from $Z = 0$ to $Z = 1$).
- To do so, define *potential* treatment states as function of the instrument, D_Z
- If both D and Z are binary, four logical groups arise

	Definition	
Name	$D_{Z=0}$	$D_{Z=1}$
Compliers (c)	0	1
Defiers (d)	1	0
Always takers (a)	1	1
Never taker (n)	0	0

- Define the observed treatment indicator as

$$D = D_{Z=0} + (D_{Z=1} - D_{Z=0})Z$$

or

$$D = D_{Z=0} + \kappa Z$$

- Here, κ is the individual-level (potentially heterogenous) effect of the instrument Z on D For compliers, $\kappa = 1$, for defiers it is $\kappa = -1$; else $\kappa = 0$

- If a valid instrument Z (with a non-zero effect of Z on D) for the causal effect $D \rightarrow Y$ satisfies the following assumptions it identifies a **L**ocal **A**verage **T**reatment **E**ffect
  - Independence
  $$(Y_1, Y_0, D_{Z=1}, D_{Z=0}) \perp Z$$
  This is the analog to the exclusion restriction discussed previously
  - Monotonicity
  $$\forall i : \kappa \geq 0 \text{ or } \forall i : \kappa \leq 0$$
  This states that the effect of Z on D is weakly monotonously positive or negative for all individuals

Some notes on LATE

- **The *LATE* identifies the average treatment effect for the subset of individuals whose treatment is shifted by the instrument.**
- It does not provide any information about the effect on always takers and never takers
- Classical IV estimates are hard/impossible to justify as estimating an average causal effect without a justification of monotonicity

  The estimate is a mixture of defiers and complieres
- The identified effect is defined by the instrument.
- The former has implications for how we think about many instruments:
  - 'Classical' literature: more instruments better
  - 'New' literature: different instruments define different average treatment effects for the same group of cases (mixture of LATEs)
- The compliance types (c,d,a,n) are not constant over instruments. Someone can be a never taker under Z_1 but a complier under Z_2

Extensions

- **Weak-instrument robust inference**. E.g., Moreira, Marcelo J. 2009. Tests with correct size when instruments can be arbitrarily weak. Journal of Econometrics 152(2)
- **Sensitivity analysis for violated exclusion restrictions**. E.g., Wang, Xuran, Yang Jiang, Nancy R Zhang, and Dylan S Small. 2018. Sensitivity analysis and power for instrumental variable studies. Biometrics 74(4)
- **Instrument validity tests using LATE-implied restrictions**. E.g., Huber, Martin and Mellace, Giovanni, 2015. Testing instrument validity for LATE identification based on inequality moment constraints. Review of Economics and Statistics 97(2). Kitagawa, Toru. 2015. A test for instrument validity. Econometrica 83(5)