Instrumental variables

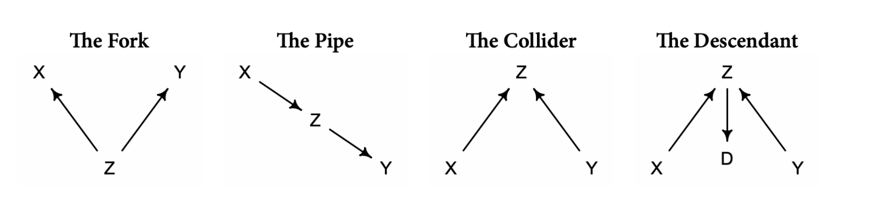
Identification and intuition

Sebastián Daza[1]

[1]Department of Sociology, UW-Madison

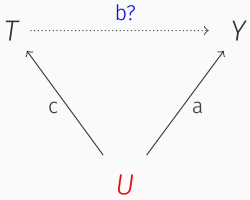
Causal Graphs: Directed Acyclic Graph (DAG)

- Type of causal graph (nodes and connections)
- To describe qualitative causal relationships among variables = not a full model description
- Directed = arrows indicating the direction of causal influence
- Acyclical = causes don't flow black on themselves



Instrumental variables

- **Problem**: treatment is not ignorable (omitted bias variable)



- Instrument could help us to recover *b*
- Quasi-experimental design (experimental analogy)
- No free, cost are additional assumptions (linearity, constant effects), higher standard errors
- If you have a great IV, do it, otherwise do OLS + formal sensitivity analysis, or better do IV + sensitivity analysis

Instrumental variables: Criteria

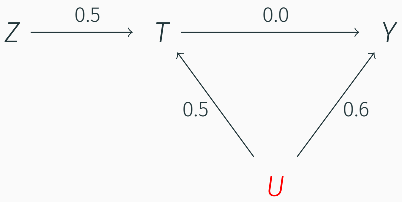
$$\begin{tikzcd} Z \arrow[r, "a"] & T \arrow[rr, "b"] & & Y \\ & & U \arrow[ul] \arrow[ur] & \end{tikzcd}$$

Z should be:

- Independent of U ($Z \perp\!\!\!\perp U$) = instrument ignorability
- Not independent of T ($Z \not\perp\!\!\!\perp T$) = first stage
- Z cannot influence Y (by any path) except through T = exclusion restriction
- Association between T and Z should be strong (otherwise close to dividing by 0)

$$\beta_{iv} = \frac{Cov(Y, Z)}{Cov(T, Z)} = \frac{ab}{a} = b + \frac{Cov(e, Z)}{Cov(T, Z)}$$

Instrumental variables: Simulation

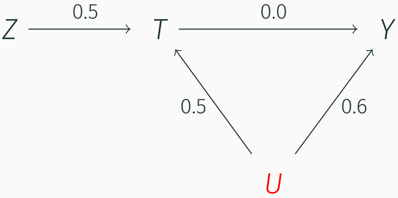



```
# SEM
Y ~ 0.6*U
T ~ 0.5*Z + 0.5*U

dat = simsem::simulateData(model, sample.nobs = 5000, model.type
↪  = "sem", orthogonal = TRUE)

m1 = lm(Y ~ T, data = dat) # naive
m2 = lm(Y ~ T + Z, data = dat) # bias amplifier!
# 2SLS by hand
dat$Th = fitted.values(lm(T ~ Z, data = dat))
m3 = lm(Y ~ Th, dat = dat)
m4 = ivreg(Y ~ T, ~ Z, data = dat) # 2SLS in R
m5 = lm(Y ~ T + U, data = dat) # God!
```

Instrumental variables: Simulation



$$\beta_{iv} = \frac{Cov(Y,Z)}{Cov(T,Z)}$$

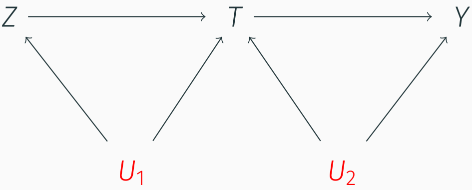
```
> cov(dat$Y, dat$Z) / cov(dat$T,dat$Z)
[1] 0.031
> (0.5 * 0.0) / 0.5
[1] 0
```

	Naive	Bias amplifier	2SLS by hand	2SLS	God
T	0.20***	0.24***		0.03	0.00
	(0.01)	(0.01)		(0.03)	(0.01)
Tf			0.03		
			(0.03)		
Z		-0.11***			
		(0.02)			
U					0.60***
					(0.02)

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Instrumental variables: Some Intuitions

No need to identify the causal effect of Z on T



Provided U_1 is not also a confounder of the relationship between T and Y

Instrumental variables: Some Intuitions

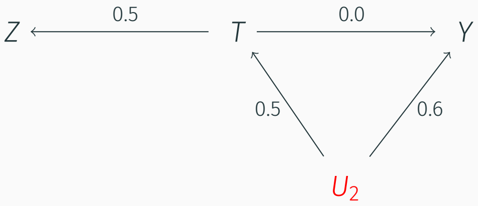
No need the instrument cause treatment at all!

$$
\begin{tikzcd}
Z & & T \arrow[rr] & & Y \\
& U_1 \arrow[ul] \arrow[ur] & & U_2 \arrow[ul] \arrow[ur] &
\end{tikzcd}
$$

Z could be just a proxy of the unobserved U_1

Instrumental variables: Some Intuitions

It's not true that Z just need to be associated with T in some arbitrary fashion

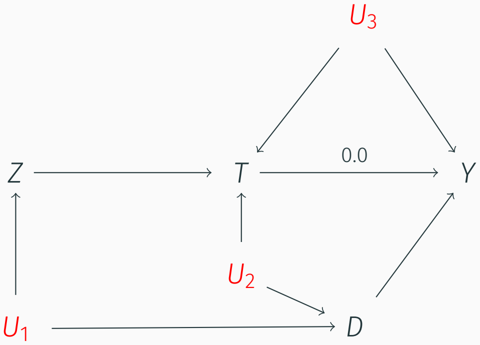


	Naive	2SLS	God
T	0.24***	0.26***	−0.00
	(0.01)	(0.03)	(0.01)
U			0.61***
			(0.02)

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Instrumental variables: Some Intuitions

Sometimes we need to adjust for covariates to heal exclusion violations...

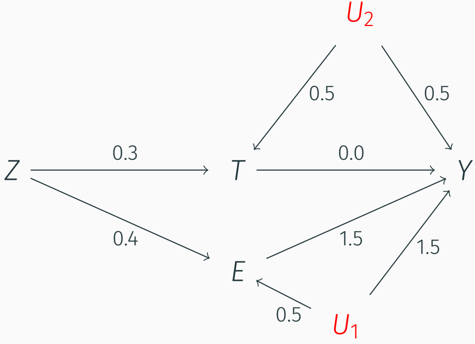


	Naive	Naive + D	2SLS	2SLS + D	God
T	0.27***	0.19***	0.10**	−0.00	0.00
	(0.01)	(0.01)	(0.03)	(0.03)	(0.01)
D		0.45***		0.50***	0.50***
		(0.01)		(0.02)	(0.01)

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Instrumental variables: Exclusion violations

Knowing and measuring the mechanism won't necessarily help

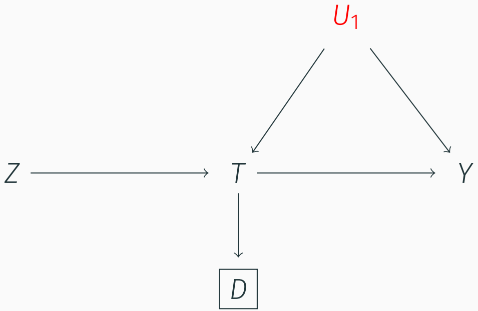


	Naive	Naive + E	2SLS	2SLS + E	God
T	0.31***	0.15***	2.04***	−0.78***	0.00
	(0.03)	(0.02)	(0.12)	(0.07)	(0.01)
E		2.02***		2.09***	1.51***
		(0.01)		(0.02)	(0.01)

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

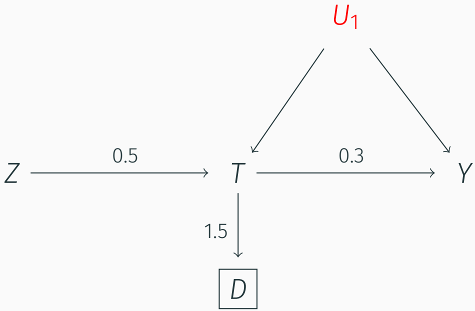
Instrumental variables: Exclusion violations

Post-treatment lost to follow-up (attrition)?



Instrumental variables: Exclusion violations

Post-treatment lost to follow-up (attrition)?



	Naive	2SLS	God
T	0.52***	0.20***	0.30***
	(0.01)	(0.04)	(0.01)

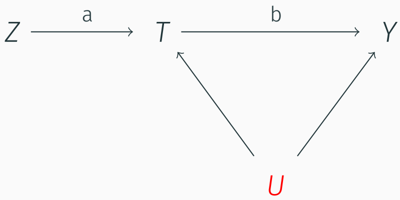
$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Instrumental variables: Homogeneity

$$\begin{tikzcd} Z \arrow[r, "a"] & T \arrow[rr, "b"] & & Y \\ & & U \arrow[ul] \arrow[ur] & \end{tikzcd}$$

- Effect of T on Y is the same for everybody
- Effect of Z on T is the same for everybody

Instrumental variables: Homogeneity



- IVs exploit only the portion of the variation in the treatment induced by the instrument
- If treatment effect for those who respond to the instrument differs from the one for those who don't response, we only can get the average causal effect of those who respond to the instrument (LATE)
- If the response to the instrument is so heterogeneous, and the treatment effect differs across groups, IV doesn't identify any causal effect (not even a weird average)

Instrumental variables: LATE

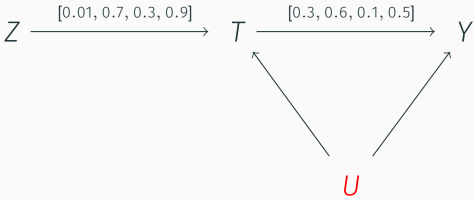
$$\begin{tikzcd}
Z \arrow[r, "{[0.01, 0.7]}"] & T \arrow[rr, "{[0.3, 0.6]}"] & & Y \\
& & U \arrow[ul] \arrow[ur] &
\end{tikzcd}$$

- Two groups of the same size
- Average treatment effect = $\frac{0.3+0.6}{2} = 0.45$

	Naive	2SLS	God using OLS
T	0.68***	0.60***	0.48***
	(0.01)	(0.02)	(0.01)

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Instrumental variables: LATE



- Four groups of the same size
- Average treatment effect = 0.375

	Naive	2SLS	God using OLS	God using ML
T	0.59***	0.47***	0.42***	0.38***
	(0.01)	(0.03)	(0.01)	(0.11)

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Instrumental variables: LATE + Monotonicity?

$$\begin{tikzcd} Z \arrow[rr, "{[-0.5, 0.7]}"] & & T \arrow[rr, "{[0.3, 0.6]}"] & & Y \\ & & & U \arrow[ul] \arrow[ur] & \end{tikzcd}$$

- Two groups of the same size
- Average treatment effect = 0.45

	Naive	2SLS	God using OLS
T	0.65***	1.34***	0.47***
	(0.01)	(0.10)	(0.01)

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Resources

- Chapter 9: Counterfactuals and Causal Inference (Morgan and Winship, 2nd edition)
- Chapter 4: Mostly Harmless Econometrics (Angrist and Pischke)
- Code simulations (R)
- DAGS: http://www.dagitty.net/