Causal inference from observational data

SICSS

University of Colorado Boulder

Institute of Behavioral Science

August 17, 2018

Amanda Jean Stevenson

Department of Sociology

University of Colorado Boulder

We can’t all be experimentalists

- Ethics
- Cost
- Omnipotence
- Time

We can’t all be experimentalists

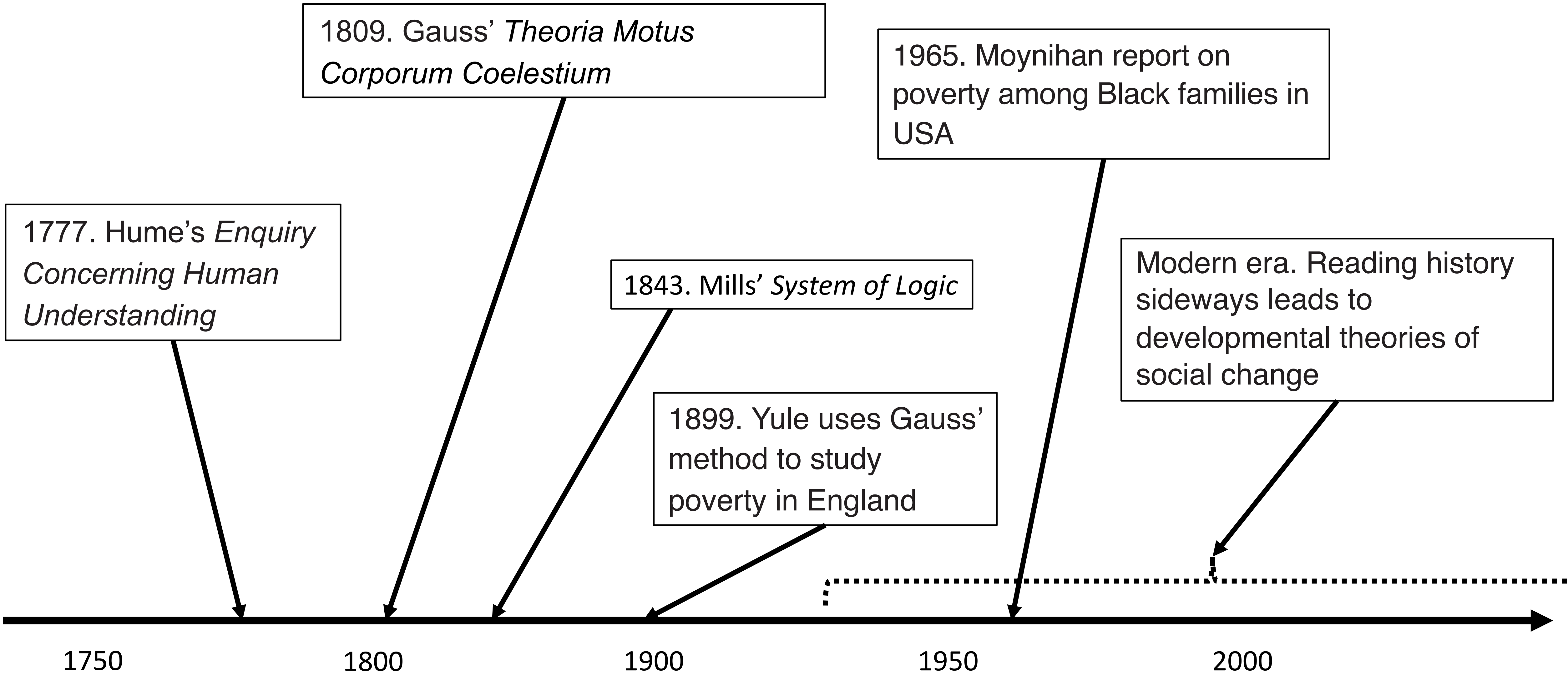
- Ethics
- Cost
- Omnipotence
- Time

But we can all analyze data

- Traces
- Registries
- Administrative
- Surveys

So many ways to go wrong

- Sample selection on dependent variable
- Conditioning on exogenous variable
- Correctly addressing colliders
- Confounding
- Selection
- Measurement error
- Simpson’s paradox
- Ecological fallacy



1809. Gauss' *Theoria Motus Corporum Coelestium*
1965. Moynihan report on poverty among Black families in USA
1777. Hume's *Enquiry Concerning Human Understanding*
1843. Mills' *System of Logic*
Modern era. Reading history sideways leads to developmental theories of social change
1899. Yule uses Gauss' method to study poverty in England
1750
1800
1900
1950
2000

A very brief survey of ways to avoid that stuff

- Counterfactual framework
- Directed acyclic graphs
- Empirical methods
  - Regression and conditioning
  - Regression discontinuity
  - Instrumental variables
  - Matching
  - Difference-in-differences

Counterfactuals and potential outcomes

Effect = (🌏|cause occurs) − (🌏|cause does not occur)

Counterfactuals and potential outcomes

Y_i^c outcome for person *i* if they **do not** receive the treatment

Y_i^t outcome for person *i* if they **do** receive the treatment (controls)

Then, for person *i*, the effect of the treatment on Y is

$$\delta_i = Y_i^t - Y_i^c$$

Counterfactuals and potential outcomes

For most outcomes Y and treatments T, we cannot observe both Y_i^t and Y_i^c for any individual *i*

Luckily, many research questions are about average or population-level effects, not the effect on an individual

Therefore, we estimate the average treatment effect by comparing the outcome among the treated and the untreated

$$\bar{\delta} \quad = \overline{Y^t} - \overline{Y^c}$$

Counterfactuals and potential outcomes

In reality, this is usually

$$\bar{\delta} \quad = \overline{Y_{i\in T}^{t}} - \overline{Y_{i\in C}^{c}}$$

Or even more realistically

$$\hat{\bar{\delta}} = \overline{\widehat{Y_{i\in T}^{t}}} - \overline{\widehat{Y_{i\in C}^{c}}}$$

Directed Acyclic Graphs

- A set of nodes (variables) and directional edges (relationships) in which no feedback loops or cycles are present

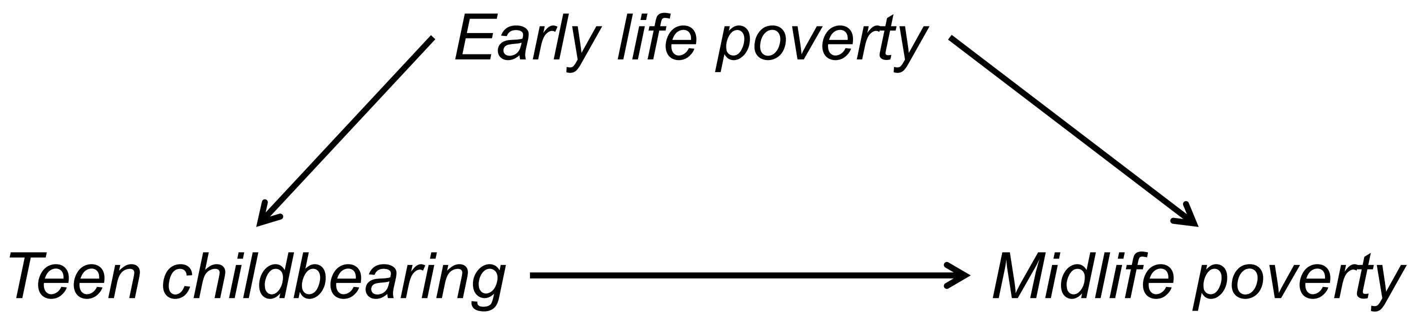
Directed Acyclic Graphs

- A set of nodes (variables) and directional edges (relationships) in which no feedback loops or cycles are present

Teen childbearing $\longrightarrow$ *Midlife poverty*

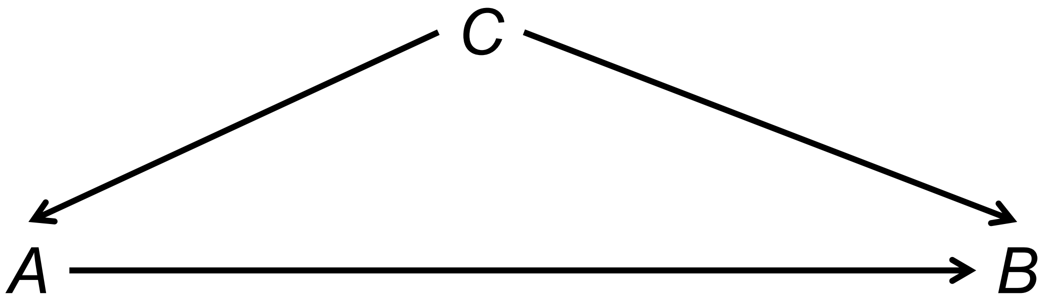
Directed Acyclic Graphs

- A set of nodes (variables) and directional edges (relationships) in which no feedback loops or cycles are present

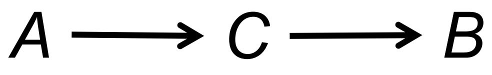


Directed Acyclic Graphs

- A set of nodes (variables) and directional edges (relationships) in which no feedback loops or cycles are present

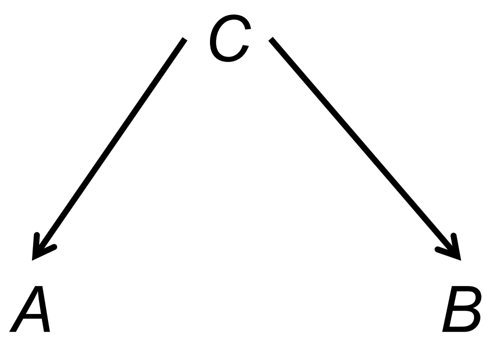


Basic patterns of causal relationships for 3 variables



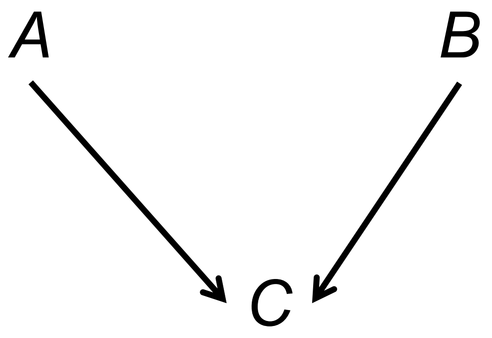
Mediation

chain



Mutual dependence

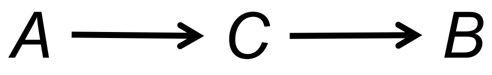
fork



Mutual causation

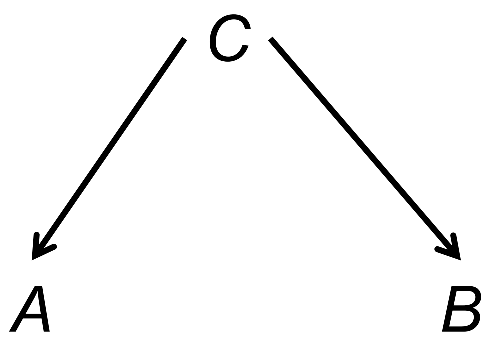
Inverted fork

Basic patterns of causal relationships for 3 variables



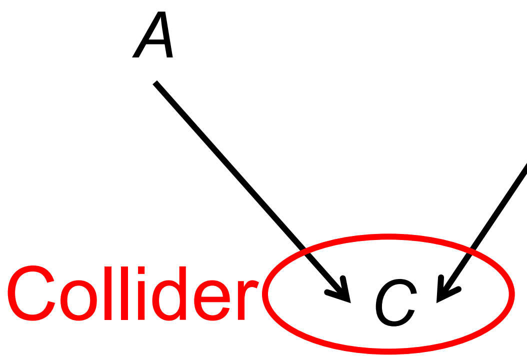
Mediation

chain



Mutual dependence

fork

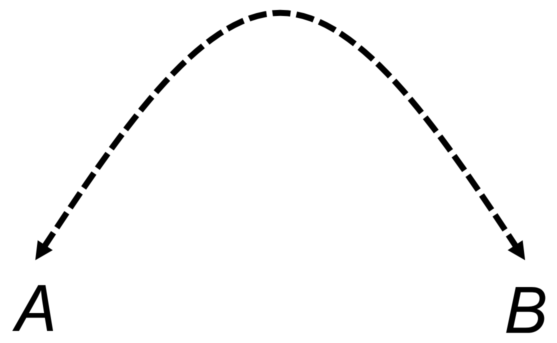


Mutual causation

Inverted fork

Joint dependence

DANGER! Almost all social phenomena share complex and unmeasured background causes.



Causal paths

$$A \longrightarrow C \longrightarrow D \longrightarrow B$$

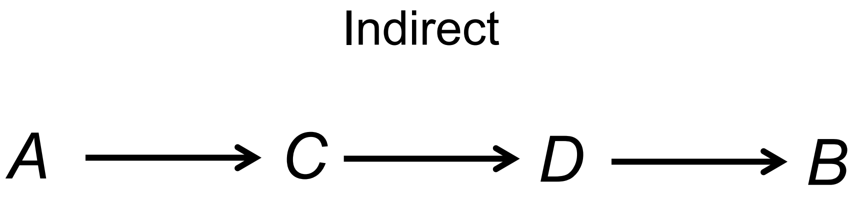
$$A \longrightarrow B$$

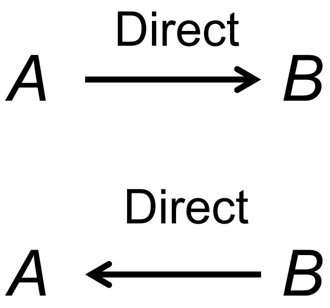
$$A \longleftarrow B$$

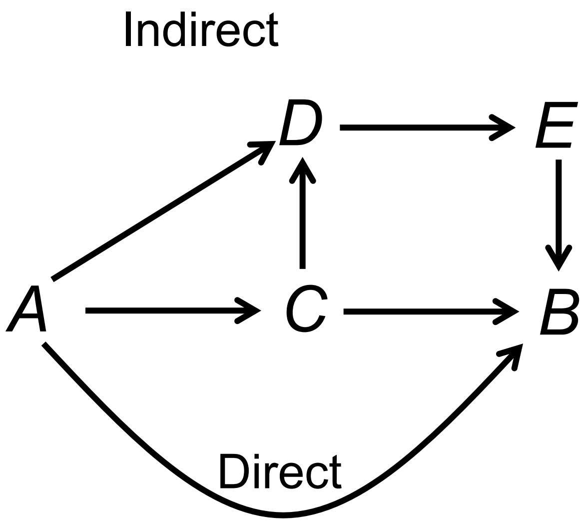
$$\begin{tikzcd}
 & D \arrow[r] & E \arrow[d] \\
A \arrow[ur] \arrow[r] \arrow[rr, bend right] & C \arrow[u] \arrow[r] & B
\end{tikzcd}$$

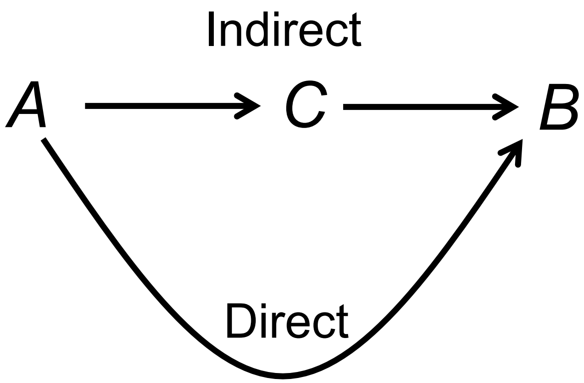
$$\begin{tikzcd}
A \arrow[r] \arrow[rr, bend right] & C \arrow[r] & B
\end{tikzcd}$$

Causal paths



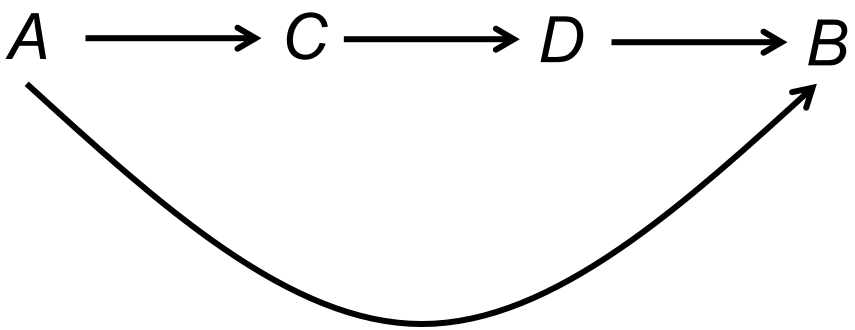


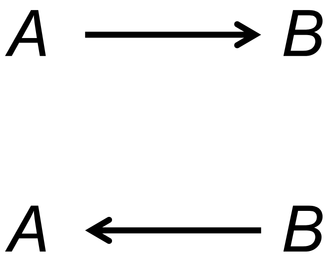




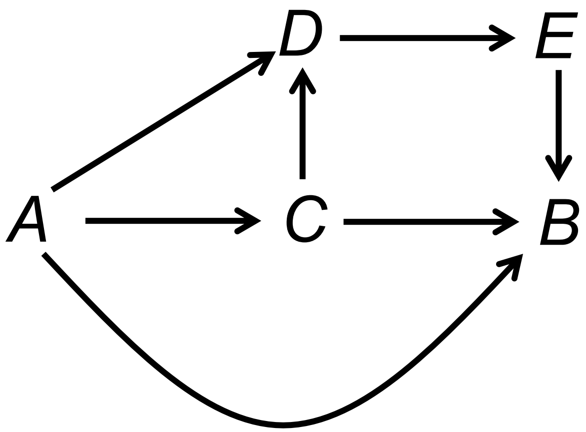
Causal paths

Backdoor

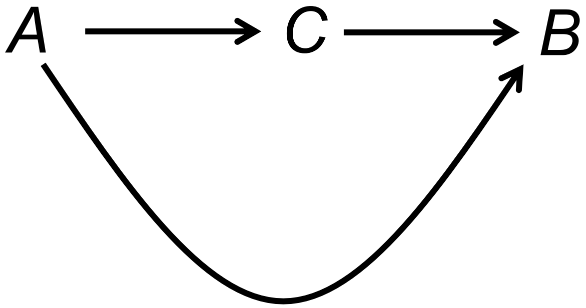




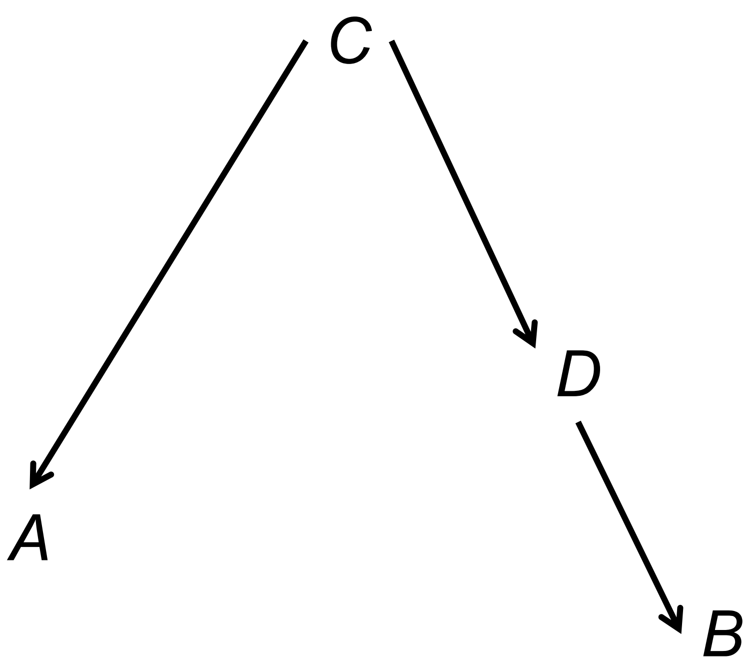
Backdoors

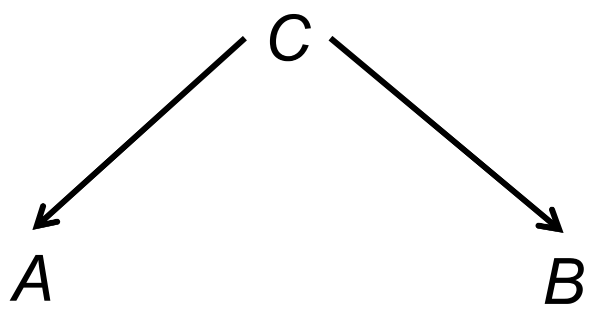


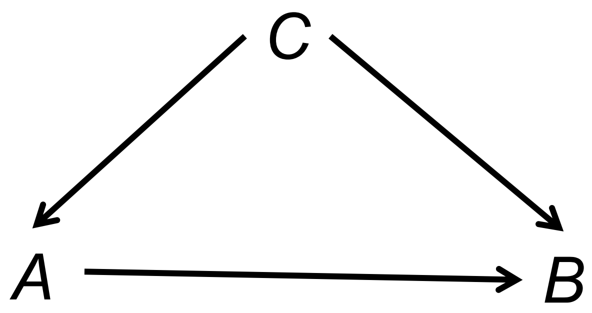
Back door



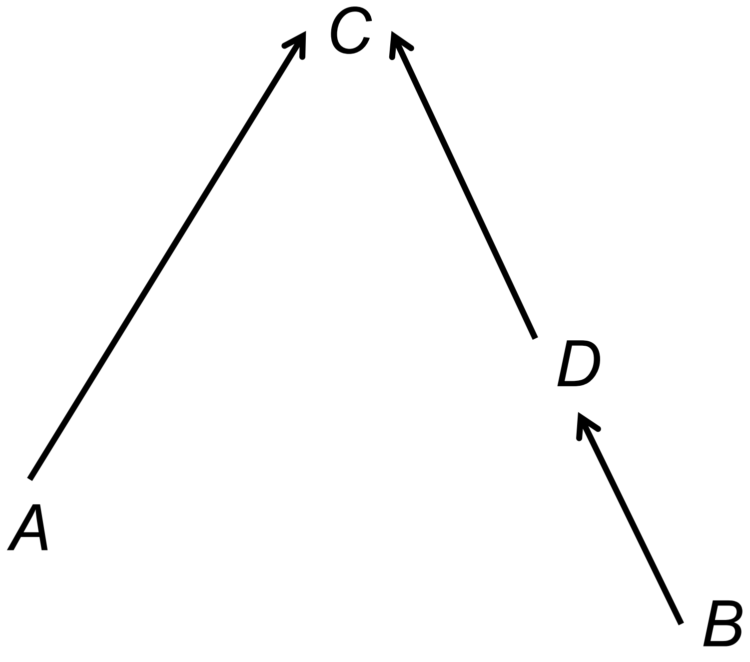
Confounded paths

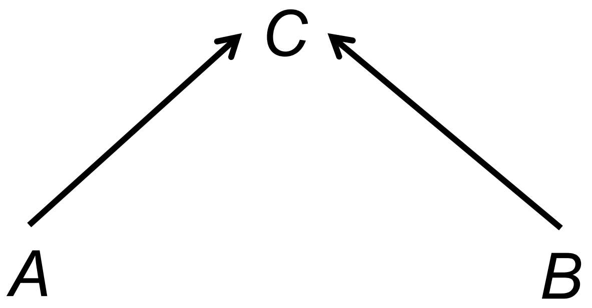


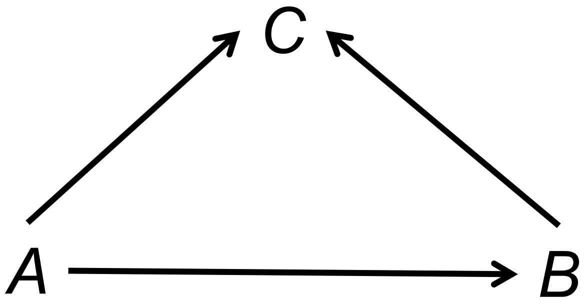




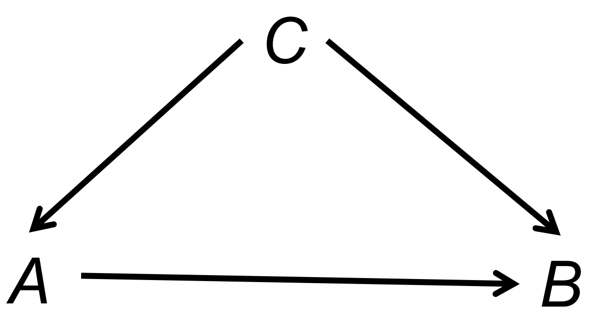
Colliding paths

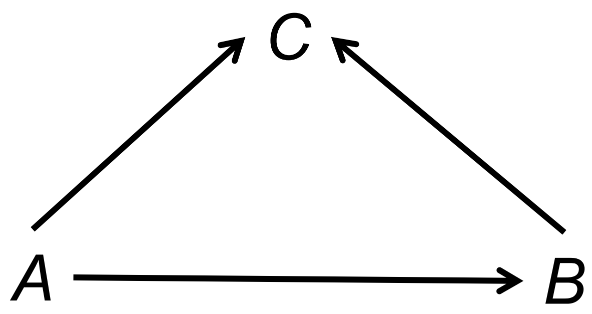


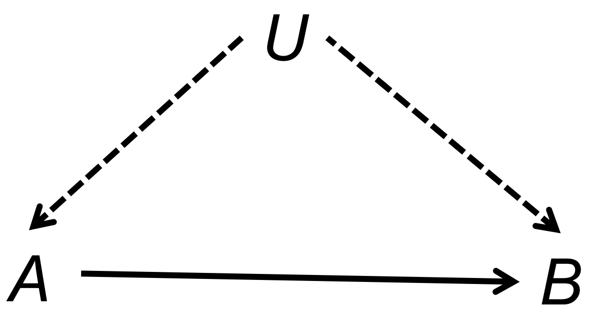




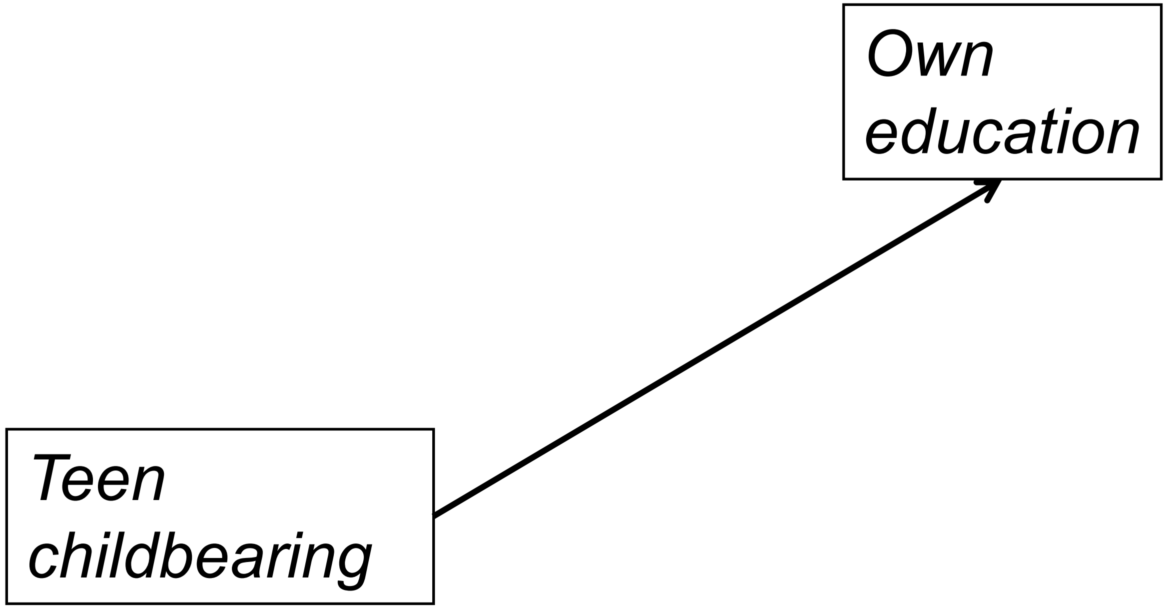
Open vs. closed backdoor paths

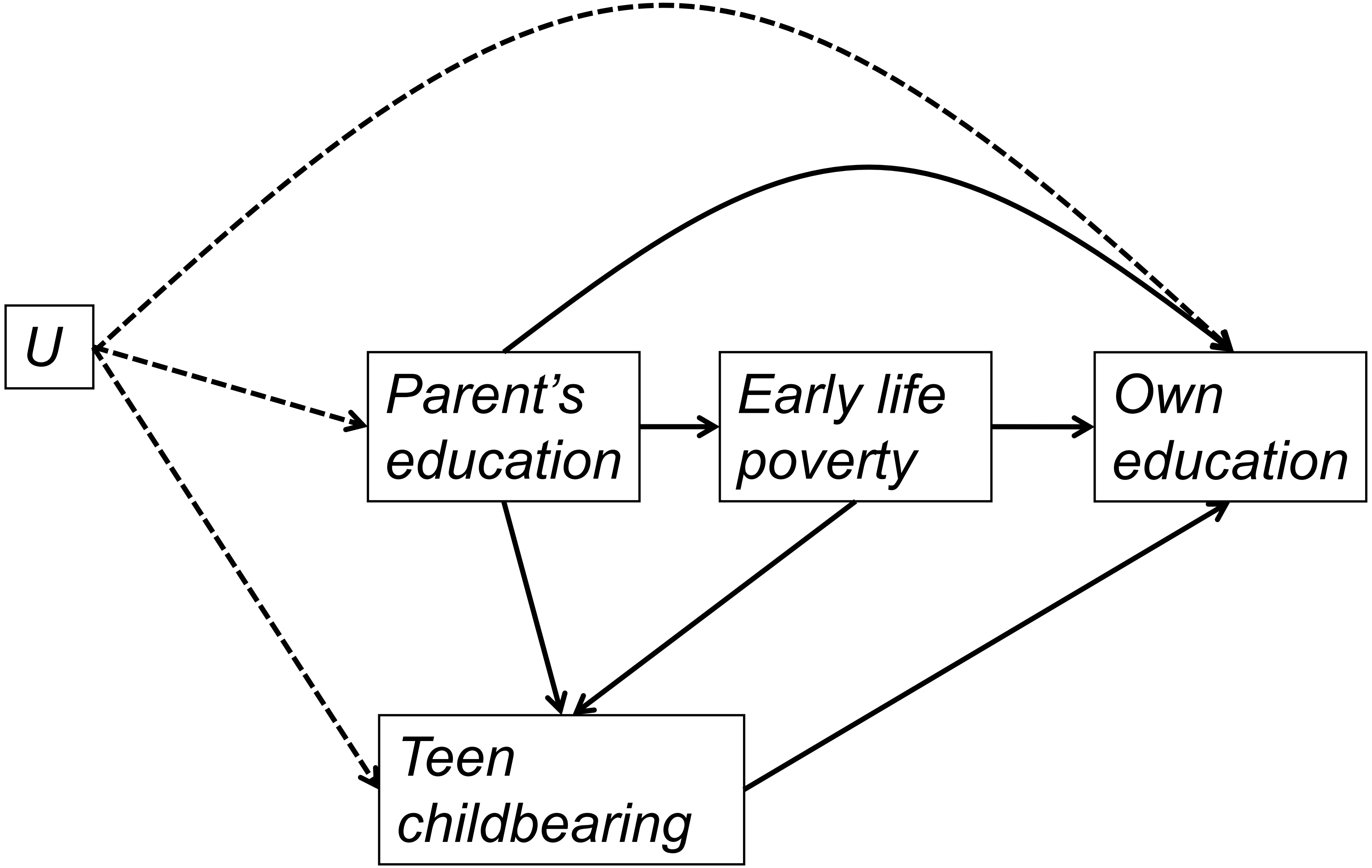






An example



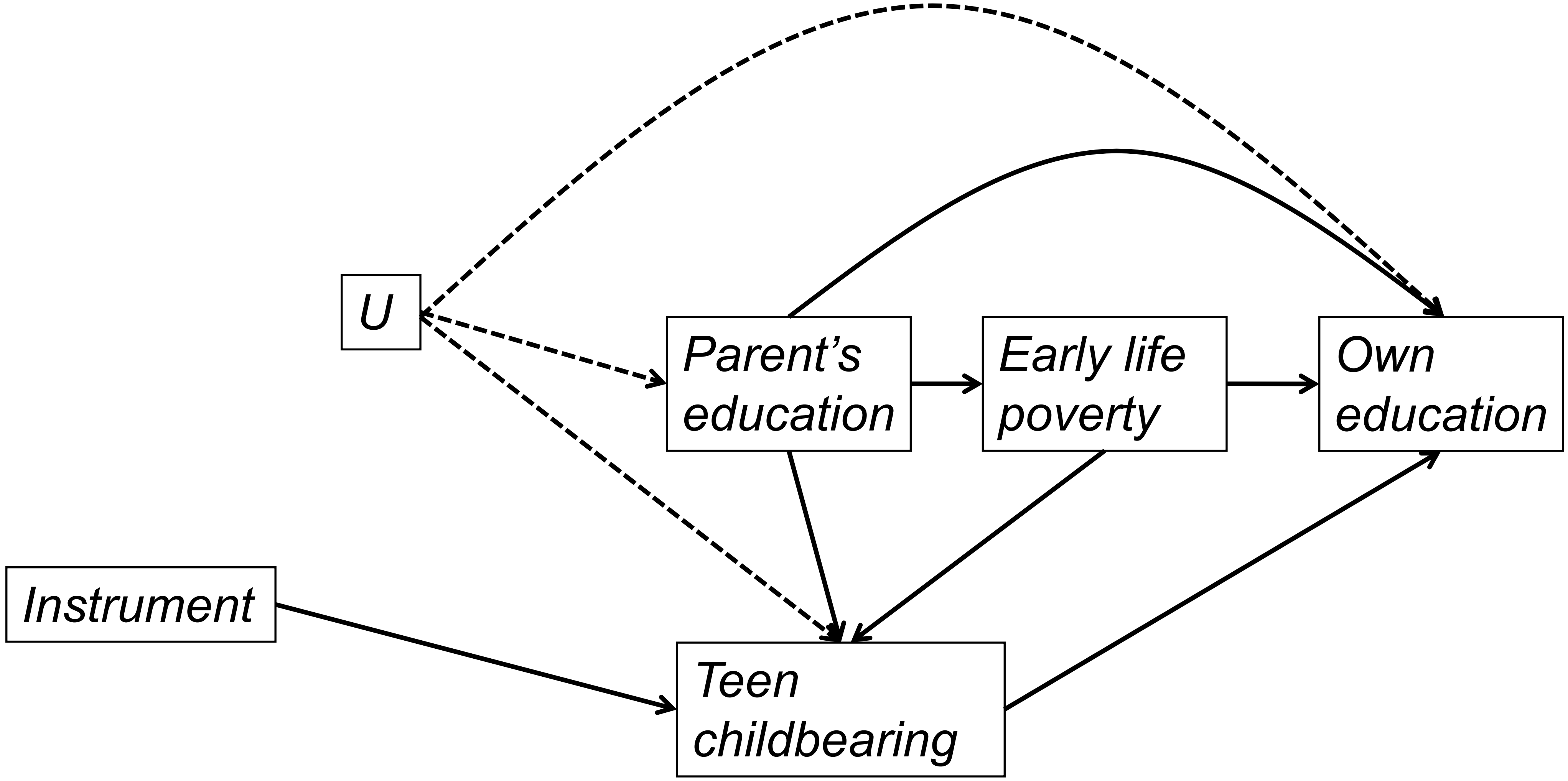

U
Parent's education
Early life poverty
Own education
Teen childbearing

Limitations of DAGs

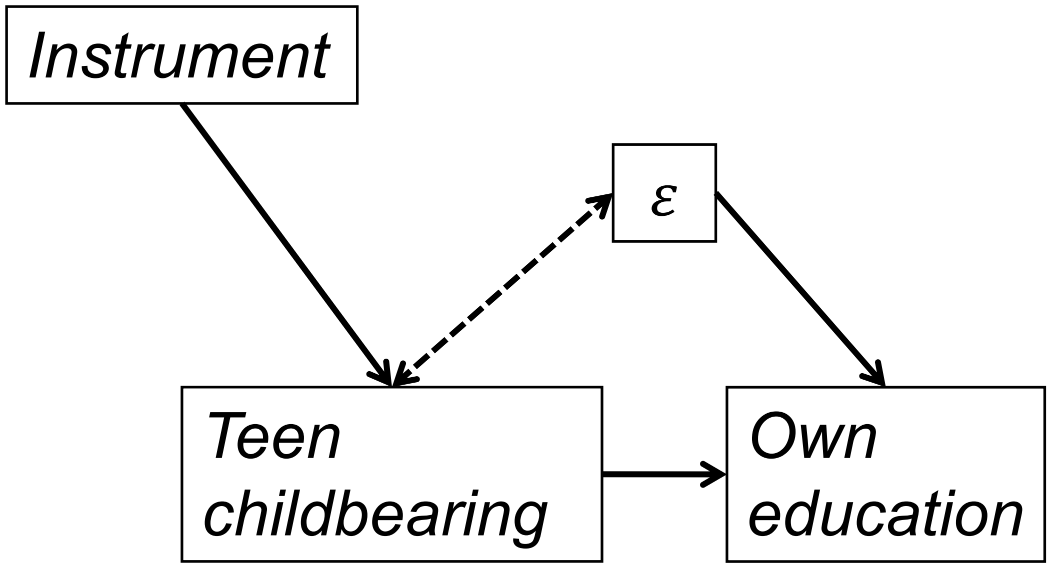
- You **must** have a theory and you're kind of betting the farm on it
- Not all potential outcome statements may be represented (time, for example is tough)
- Implicit focus on average treatment effects (ATEs)

Empirical methods

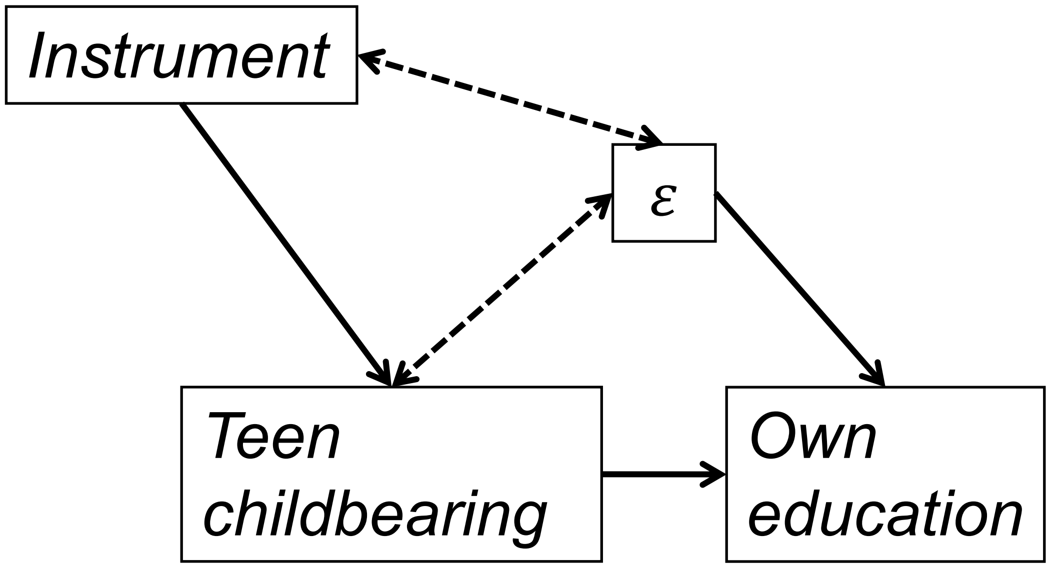
Instrumental variables



Instrumental variables



Valid instrument



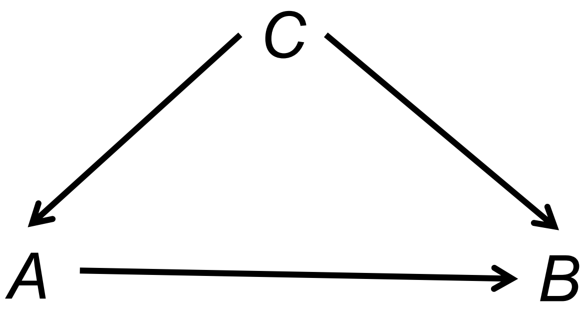
Invalid instrument

Instrumental variables

- Related to two-stage, latent variable selection, and control function models
  - Estimating the effect of a causal variable by estimating its effect through an exhaustive (closed) path

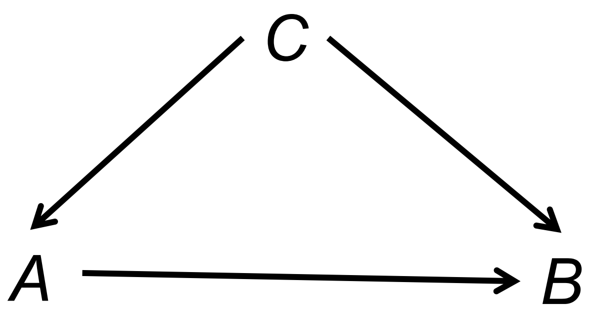
Regression and conditioning

- If your DAG tells you that the relationship between A and B is confounded by C and you can estimate $\hat{\bar{\delta}}$ separately by levels of C, then conditioning or regression may recover a valid estimate of ATE.
- Conditioning or stratification: Estimate $\hat{\bar{\delta}}$ separately by groups or strata of C and then weight the subgroup estimates to represent the population distribution.

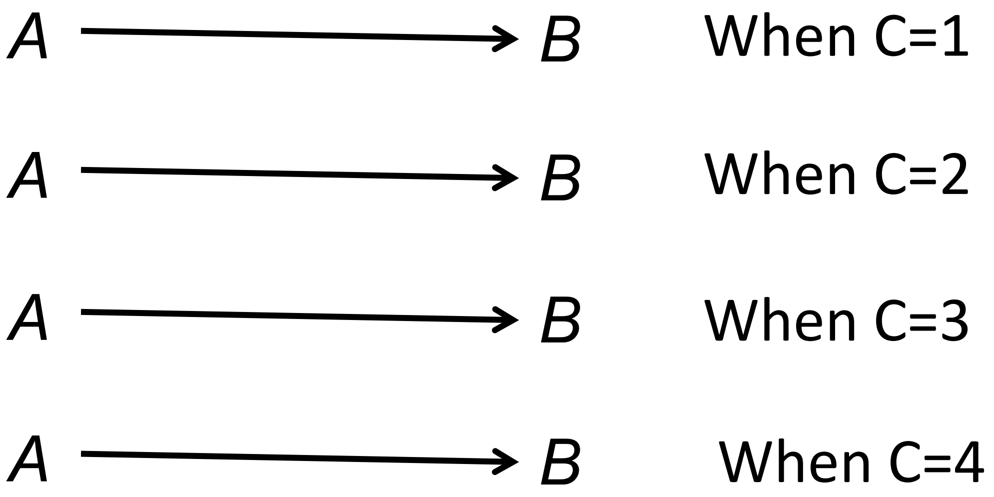


Regression and conditioning

- Conditioning removes nodes from your DAG, potentially simplifying your model and addressing confounding to allow the direct estimation of $\bar{\delta}$.



Becomes



Regression and conditioning

- But if you condition on a collider, your regression-based estimate may be wildly wrong.

- Regression is all about the error (what you can't measure)
  - Works when you can include $\overline{X}_i$ such that you minimize the correlation between treatment assignment T_i and the error term v_i

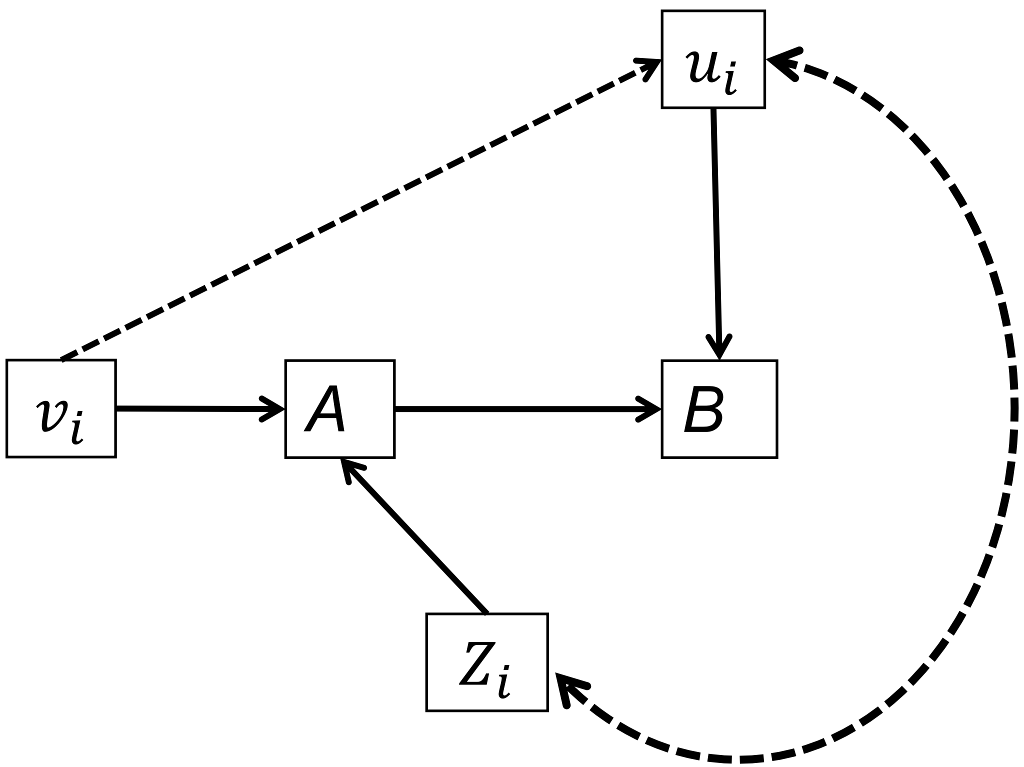
Matching

- Matching is a direct attempt to find counterfactuals at the individual level so that you can actually estimate $\delta_i = Y_i^t - Y_i^c$
- The basic approach is to find individuals *i* and *i** in the treatment and control group such that $Y_i^t = Y_{i*}^t$ and $Y_i^c = Y_{i*}^c$
- Then, $\delta_i = Y_i^t - Y_{i*}^c$ and $\delta_{i*} = Y_{i*}^t - Y_i^c$

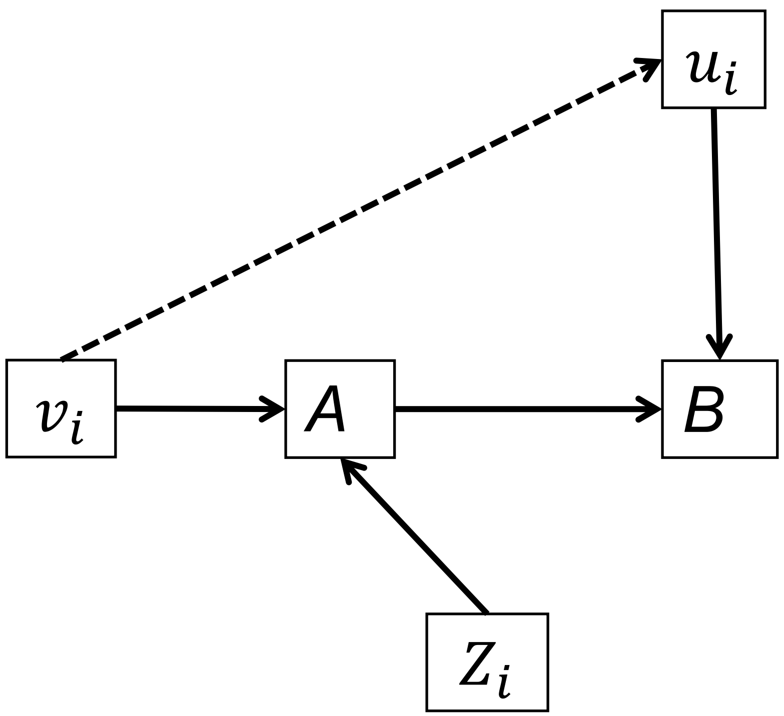
Matching

- Exact matching
  - i and i^* selected based on a full set of covariates $\bar{X}$ such that $\bar{X}_i = \bar{X}_{i^*}$
- Nearest neighbor matching
  - i and i^* selected such that $\bar{X}_i - \bar{X}_{i^*}$ is minimized
- Propensity score matching
  - $P(Z_i) = Prob(T = 1|Z_i)$

Matching



Most realities



The assumption made by matching

Matching

- Common support
  - Factors associated with treatment assignment needs to be relatively well-distributed across T and C
  - You usually need a lot of data

- No assumption about functional form of effect!
- Statistical efficiency

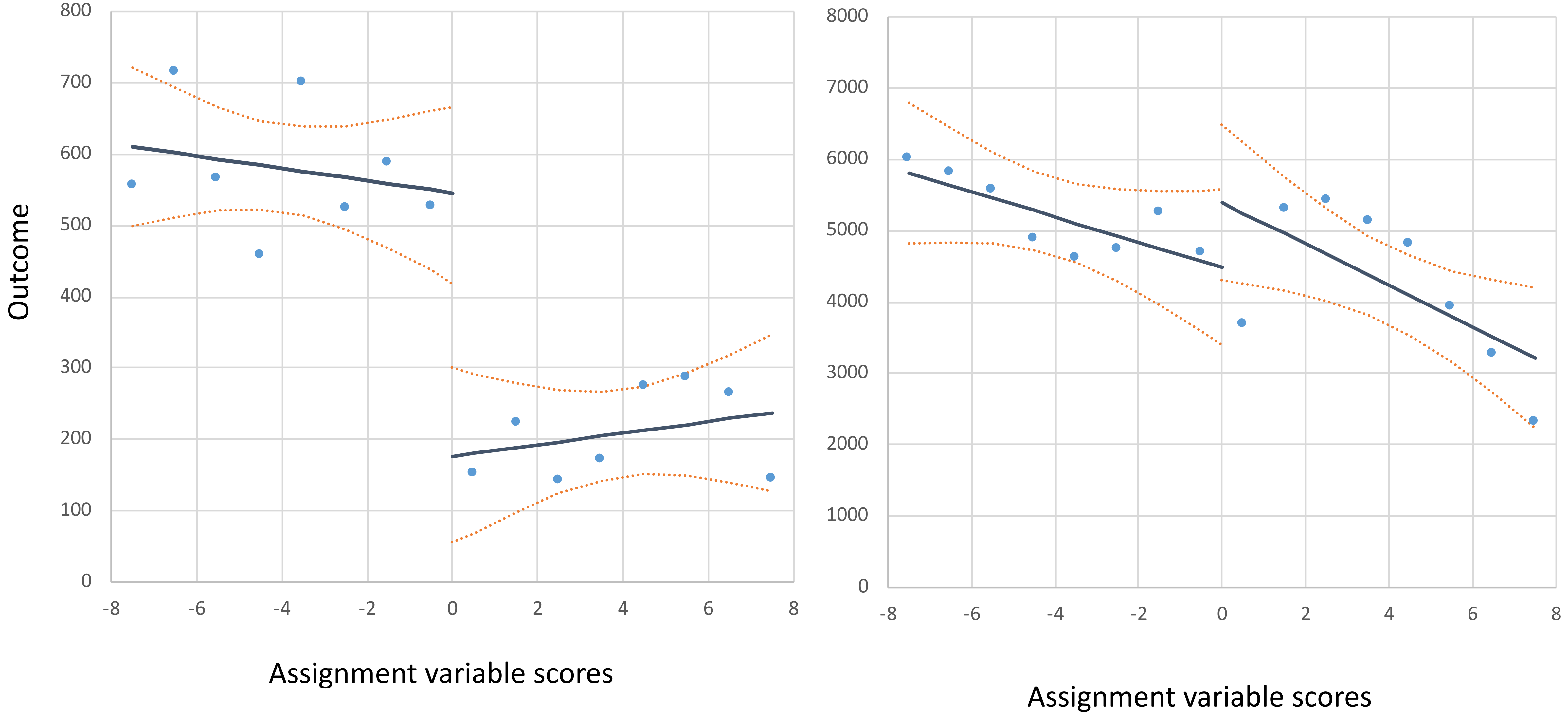
Regression discontinuity

- Requires an exogenous selection into treatment determined by an arbitrary cutoff in a running variable or assignment variable
  - Eligibility thresholds
  - Policy implementation on a specific date

- Must be able to model the running/assignment variable (poverty, time, etc.) on the outcome.

- Practically, easy to implement and provides a vivid illustration of effect

Regression discontinuity



Regression discontinuity

- Common problems:
  - Cutoffs aren’t actually strictly followed
  - Functional form of underlying relationship between running variable and outcome incorrectly specified
  - Only generates local area treatment effect (LATE)

Difference-in-differences

- Variation may function like an experiment if it is exogenous
- May be conducted as panel fixed effects, or repeated cross-sections

	Change over time		Difference-in-differences (DID)	P-value
	Intervention	No intervention		
Outcome	A=pre-post	B=pre-post	A-B	

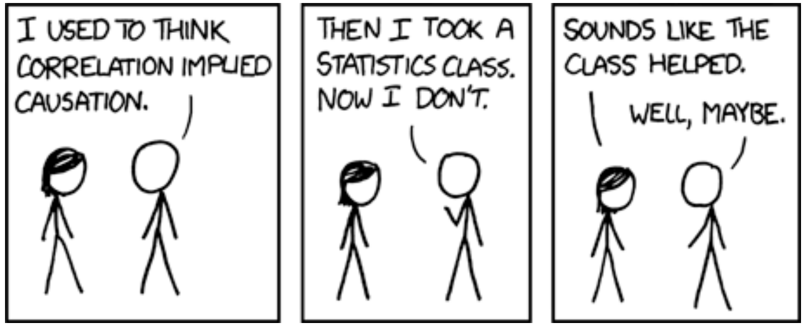
Some recommended further reading

Morgan, Stephen L., and Christopher Winship. *Counterfactuals and causal inference*. Cambridge University Press, 2015.

Pearl, Judea. *Causality*. Cambridge university press, 2009.

Cunninham, Scott. *Causal inference: the Mixtape*. V.1.7. http://scunning.com/cunningham_mixtape.pdf

Winship, Christopher, and Stephen L. Morgan. "The estimation of causal effects from observational data." *Annual review of sociology* 25.1 (1999): 659-706.



You already love xkcd, right?
https://xkcd.com/552/

Pearl’s three basic methods for identifying causal effects

- Condition on all backdoor paths
- Condition on variables in order to estimate by a mechanism
- Estimate effect using exogenous shock to cause