Topics in Causal Inference

STAT41530

Jingshu Wang

Lecture 8

Topic:
Comparison between PO (potential outcome) and DAG

- This lecture is based on the paper:

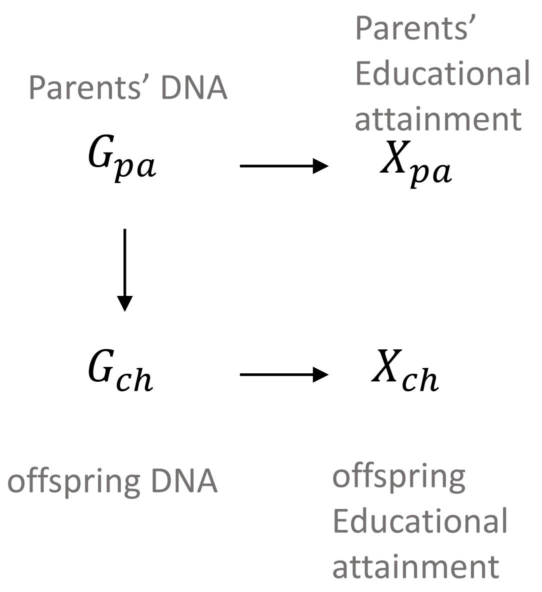
  Imbens, G. W. (2020). Potential outcome and directed acyclic graph approaches to causality: Relevance for empirical practice in economics. *Journal of Economic Literature*, *58*(4), 1129-79.

- Compare the weakness/strengths of PO and DAG from three aspects
  - Representation of complex causal structure
  - Representation of heterogeneous treatment effects
  - Treatment assignment mechanism
    - Regression discontinuity design (RDD)

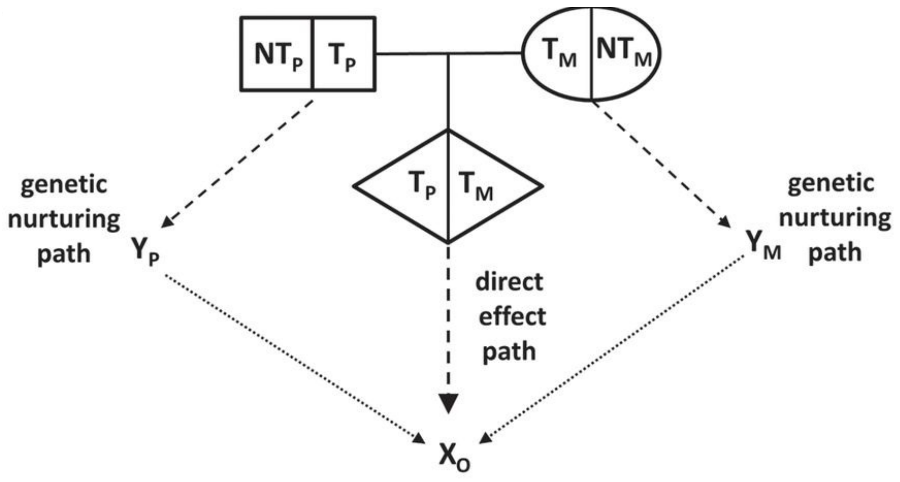
Representation of complex causal structure

- As a graphical approach, DAG is superior in illustrating the causal relationships in a complex model and in clarifying some key assumptions

A simple understanding of heritability

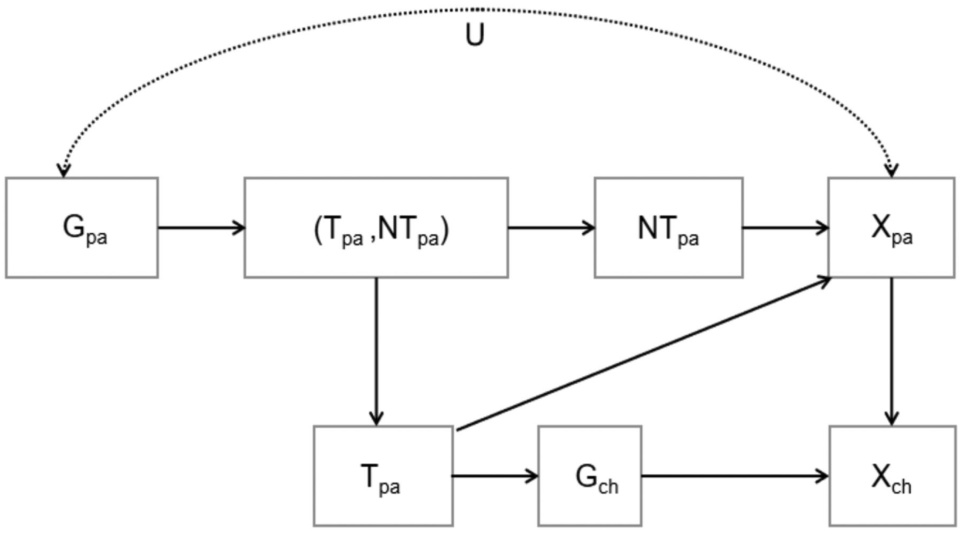


Genetic nurturing



Kong, A., Thorleifsson, G., Frigge, M. L., Vilhjalmsson, B. J., Young, A. I., Thorgeirsson, T. E., ... & Stefansson, K. (2018). The nature of nurture: Effects of parental genotypes. *Science*, *359*(6374), 424-428.

A more complicated relationship including Genetic nurturing



Shen, H., & Feldman, M. W. (2020). Genetic nurturing, missing heritability, and causal analysis in genetic statistics. *Proceedings of the national academy of sciences*, *117*(41), 25646-25654.

Representation of complex causal structure

- On the other hand, in empirical studies, we may want to avoid considering models with dozens or even a hundred variables and complex relations between them that do not reduce to simple identification strategies and the analyses would be totally impenetrable

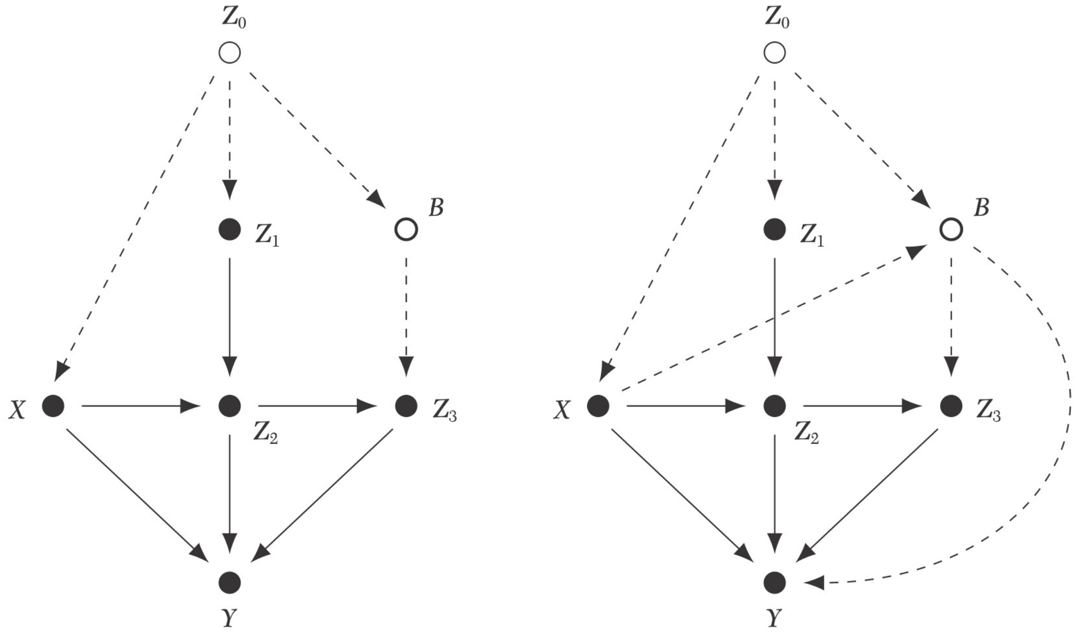


Figure 4. Two Examples of Complex DAGs

Sources: Based on figure 1 in Pearl (1995).

Example:
X: soil fumigation
Y: crop yield
B: bird population

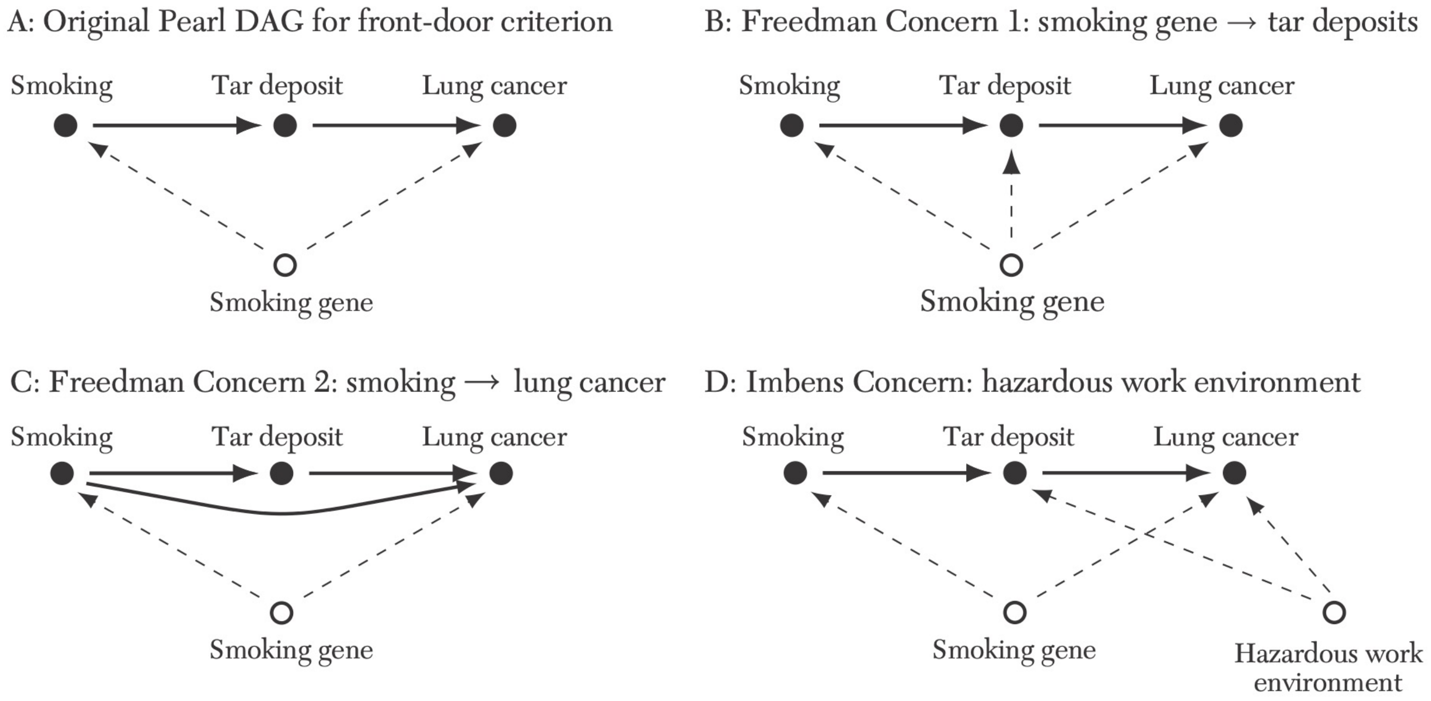
Z_0: eelworm population of last season
Z_1: eelworm population before the treatment
Z_2: eelworm population after the fumigation
Z_3: eelworm population at the end of the season

Representation of complex causal structure

- Two specific structures that can easily be discussed in DAG but not with potential outcome framework:

  the front-door criterion and the M-bias
- On the other hand, they are "toy models"

The front-door criterion



- Assumptions on the front-door criterion can be easily violated
- The mediators Z are not randomized
- Need real-world examples where the front-door assumptions are convincing

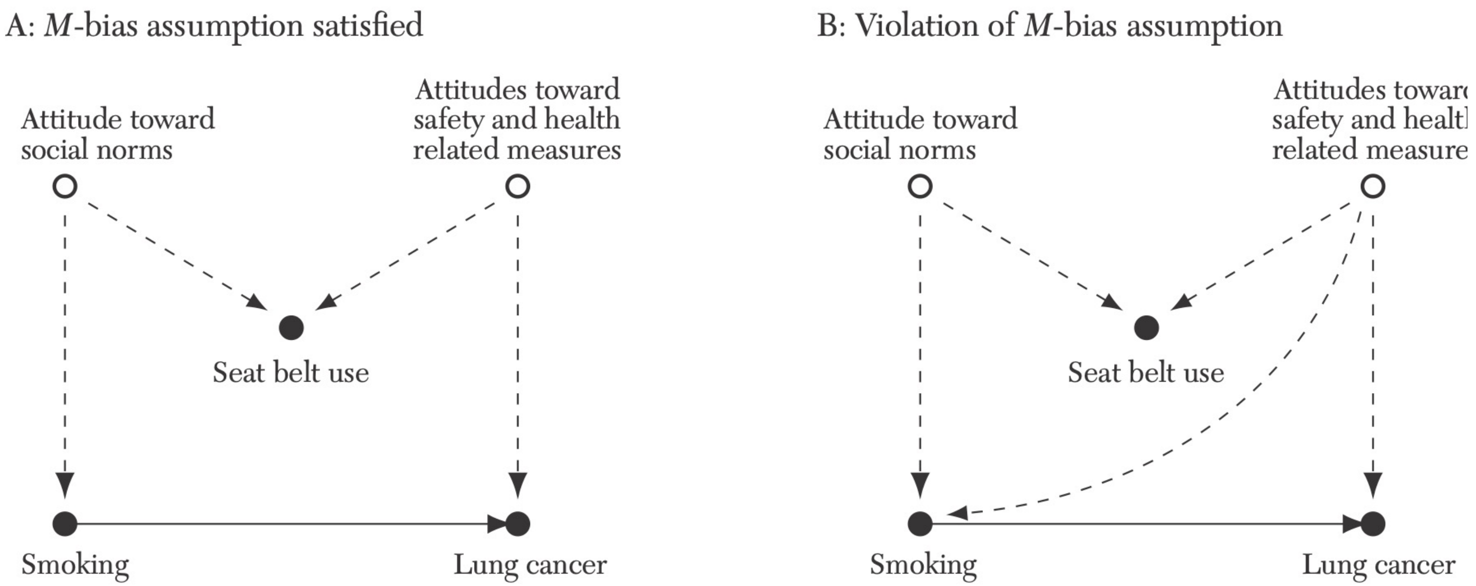
Representation of complex causal structure

- Two specific structures that can easily be discussed in DAG but not with potential outcome framework:

  the front-door criterion and the M-bias

- On the other hand, they are "toy models"

M-bias



- In potential outcome framework, we usually adjust for all pre-treatment covariates as possible confounding

- M-bias is a counter example that is clear in DAG

- However, Imbens questioned how likely is M-bias assumption satisfied in practice

Representation of heterogeneous treatment effects

- More general difficulties in the DAG framework to capture individual level heterogeneity
  - The use of population level random variables implicitly assumes i.i.d. assumptions over all individuals in the super population
  - For example, it cis not that natural to derive principal stratification and the monotonicity assumption in IV via DAG.

- There are implicit assumptions in the structural equations
  - For example, the additivity in the structural equation $Y = f(A, U) + E_Y$ implicitly assumes homogeneous treatment effect at each level of U: $Y(1) - Y(0) = f(1, U) - f(0, U)$
  - Formal identification results can be clearer using the potential outcome language

Treatment assignment mechanism

- In the potential outcome framework, randomized experiments has a special role.
  - In complete/conditional randomized experiments, the treatment assignment mechanism is known
  - We conceptualize observational studies as conditional randomized experiments
- In DAG, all variables are doable and the literature is silent about experiments
  - A variable may not indicate a particular intervention

    vague causal questions: "Causal effect of child poverty", "she did not get the position because she is a woman", "effect of obesity"
  - With DAG, we may ignore the important overlapping assumption
  - With the potential outcome language, it is more natural to discuss propensity scores, covariate balancing, doubly robust estimator, IPW, matching …
  - We can perform randomization inference that is purely based on the treatment assignment mechanism (not an assumption like i.i.d.)
  - We can discuss different design strategies (like RDD)

Regression discontinuity design (RDD)

Example

- An educational program where the eligibility of a student depends solely on whether his/her test score of an exam is above or below a threshold
- Students whose score are just above and students whose score are just below the threshold are comparable in their background (e.g., learning abilities and attitudes) [unmeasured confounders]
- Aim to identify the average treatment effect of the treatment at the threshold

- Pre-treatment variable (running variable): W
- Discontinuity assumption:

$$P(A = 1 \,|c_0^+) \neq P(A = 1 \,|c_0^-)$$

where $P(A = 1 \,|c_0^+) = \lim_{c \downarrow c_0} P(A = 1|w = c)$, $P(A = 1 \,|c_0^-) = \lim_{c \uparrow c_0} P(A = 1|w = c)$

- **Sharp RDD**: $A = 1_{W \geq c_0}$, $P(A = 1 \,|c_0^+) = 1$ and $P(A = 1 \,|c_0^-) = 0$ (or reversed)
- **Fuzzy RDD**: $P(A = 1 \,|c_0^+) < P(A = 1 \,|c_0^-)$ (or reversed) and $P(A = 1|w = c)$ is monotone in c

  The threshold is only an encouragement of treatment assignment

RDD: continuity-based approach

Sharp RDD

- Identification assumption: continuity of conditional regression functions: Both $E[Y(0) \mid W \; = c]$ and $E[Y(1) \mid W \; = c]$ are continuous in s
- Then $E[Y(0) \mid W \; = c] = \lim_{c \uparrow c_0} E(Y(0) | W = c) = \lim_{c \uparrow c_0} E(Y(0) | A = 0, W = c) =$

  $\lim_{c \uparrow c_0} E(Y | W = c)$

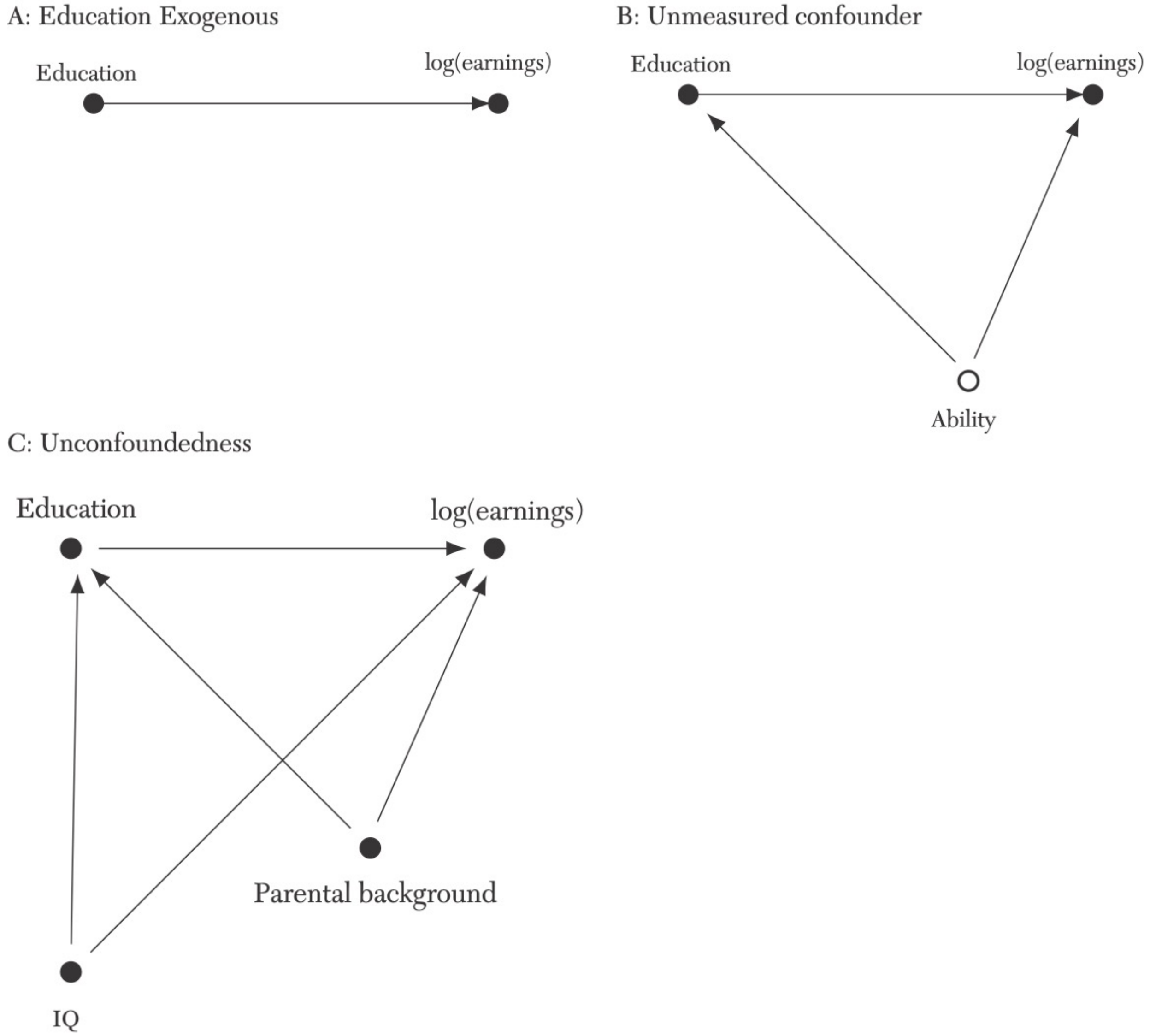
  Also, $E[Y(1) \mid W \; = c] = \lim_{c \downarrow c_0} E(Y | W = c)$
- Causal estimand: $\lim_{c \downarrow c_0} E(Y | W = c) - \lim_{c \uparrow c_0} E(Y | W = c)$

Fuzzy RDD

- Causal estimand

$$\frac{\lim_{c \downarrow c_0} E(Y | W = c) - \lim_{c \uparrow c_0} E(Y | W = c)}{\lim_{c \downarrow c_0} E(A | W = c) - \lim_{c \uparrow c_0} E(A | W = c)}$$

An example: identifying the returns to education

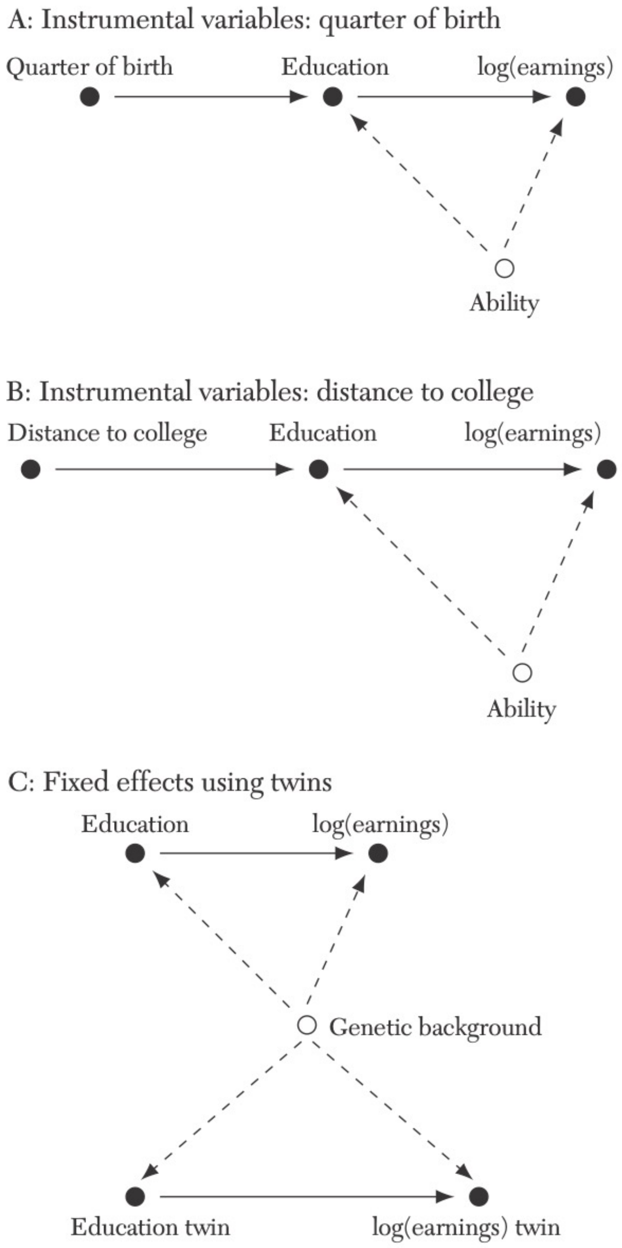


Treatment: years of education
Outcome: the logarithm of earnings

Unmeasured confounding:
an individual's ability

Strategy I:
find proxies for unmeasured ability, such as parental background and IQ

An example: identifying the returns to education



Strategy II:
quarter of birth, distance of college

Strategy III:
Look for siblings or twins
The genetic background of identical twins are the same, which controls for all unmeasured confounding

Other aspects

- With DAG, we can perform causal network discovery under additional assumptions
    - Identify the shape of DAG from observed data
    - Identify causal direction

- Potential outcome language is more natural to represent *equilibrium behavior*