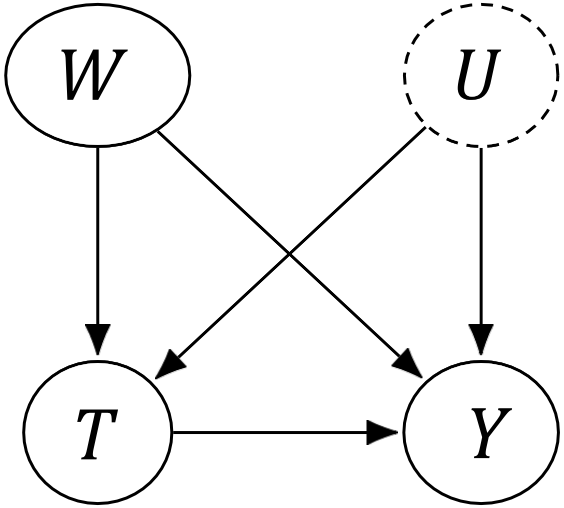
CSDS 452 Causality and Machine Learning

Lecture 7: Unobserved confounders (2)

Instructor: Jing Ma

Fall 2024, CDS@CWRU

Recap: Unobserved Confounders



Unbiased:

$$E[Y(1) - Y(0)] = E_{W,U}[E[Y|T=1,\ W, U] - E[Y|T=0,\ W, U]]$$

Biased:

$$E_W[E[Y|T=1,\ W] - E[Y|T=0,\ W]]$$

(Cannot adjust for unobserved confounders U)

Recap: Weaker assumption of Unconfoundedness?

- Unconfoundedness assumption

$$E[Y(1) - Y(0)] = E_W[E[Y|T = 1, W] - E[Y|T = 0, W]]$$

Identify a "Point"

- Weaker assumption
  - Allow the existence of some unobserved confounders
  - Instead of a point, identify an "interval"
  - "Partial identification" or "set identification"

Recap: Observational-Counterfactual Decomposition

$$
\begin{aligned}
&E[Y(1) - Y(0)] \\
&= E[Y(1)] - E[Y(0)] \\
&= P(T = 1)E[Y(1)|T = 1] + P(T = 0)E[Y(1)|T = 0] \\
&\quad -P(T = 1)E[Y(0)|T = 1] - P(T = 0)E[Y(0)|T = 0] \\
&= P(T = 1)\boxed{E[Y|T = 1]} + P(T = 0)\boxed{E[Y(1)|T = 0]} \\
&\quad -P(T = 1)\boxed{E[Y(0)|T = 1]} - P(T = 0)\boxed{E[Y|T = 0]}
\end{aligned}
$$

Observational

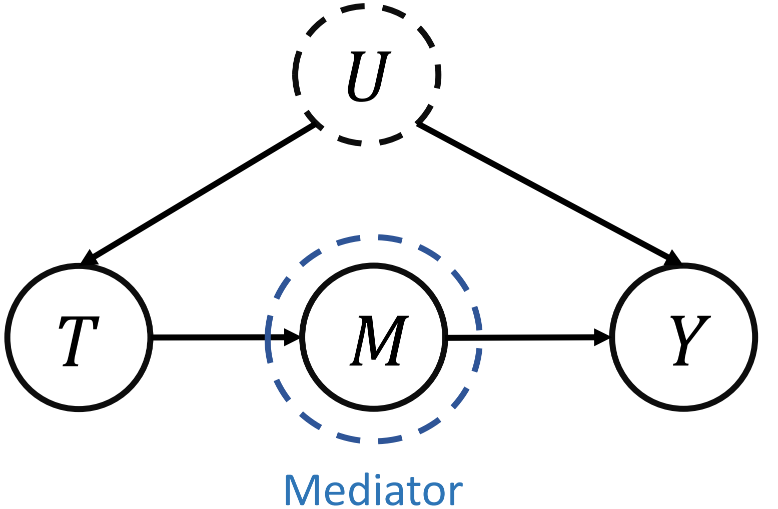
Counterfactual

Outline

- Front-door Adjustment (recap)
- Instrumental variables
  - What is Instrumental Variable
  - 3 Assumptions of Instrumental Variable
  - Linear setting
  - 2SLS
  - Non-parametric identification
- Proxy variables for unobserved confounders

Frontdoor Adjustment

- Step 1. Identify the causal effect of T on M
- Step 2. Identify the causal effect of M on Y
- Step 3. Based on the above two steps, identify the causal effect of T on Y



Outline

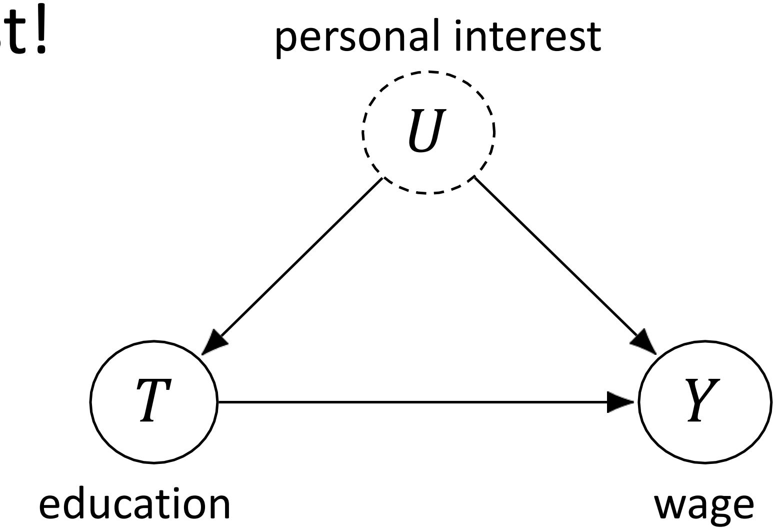
- Front-door Adjustment (recap)
- Instrumental variables
  - What is Instrumental Variable
  - 3 Assumptions of Instrumental Variable
  - Linear setting
  - 2SLS
  - Non-parametric identification
- Proxy variables for unobserved confounders

Instrumental Variable (IV)

- What is Instrumental Variable?
- Why do we need IVs?

What is Instrumental Variable?

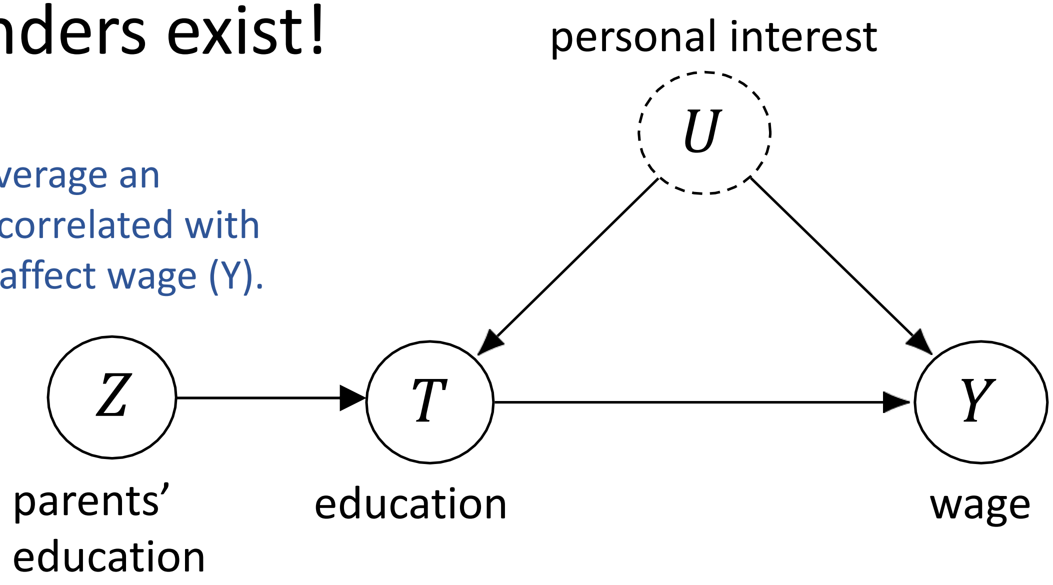
- Example: causal effect estimation of education on wage
- Hidden confounders exist!



What is Instrumental Variable?

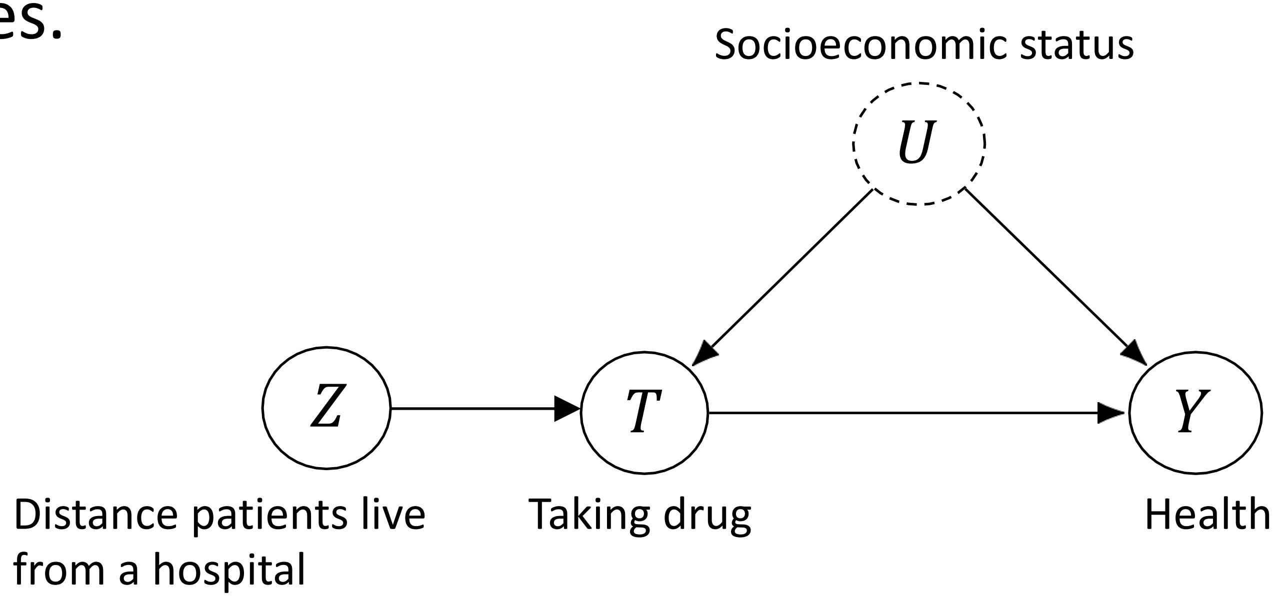
- Example: causal effect estimation of education on wage
- Hidden confounders exist!

To address this issue, you can leverage an instrumental variable (Z) that is correlated with education (X) but is not directly affect wage (Y).



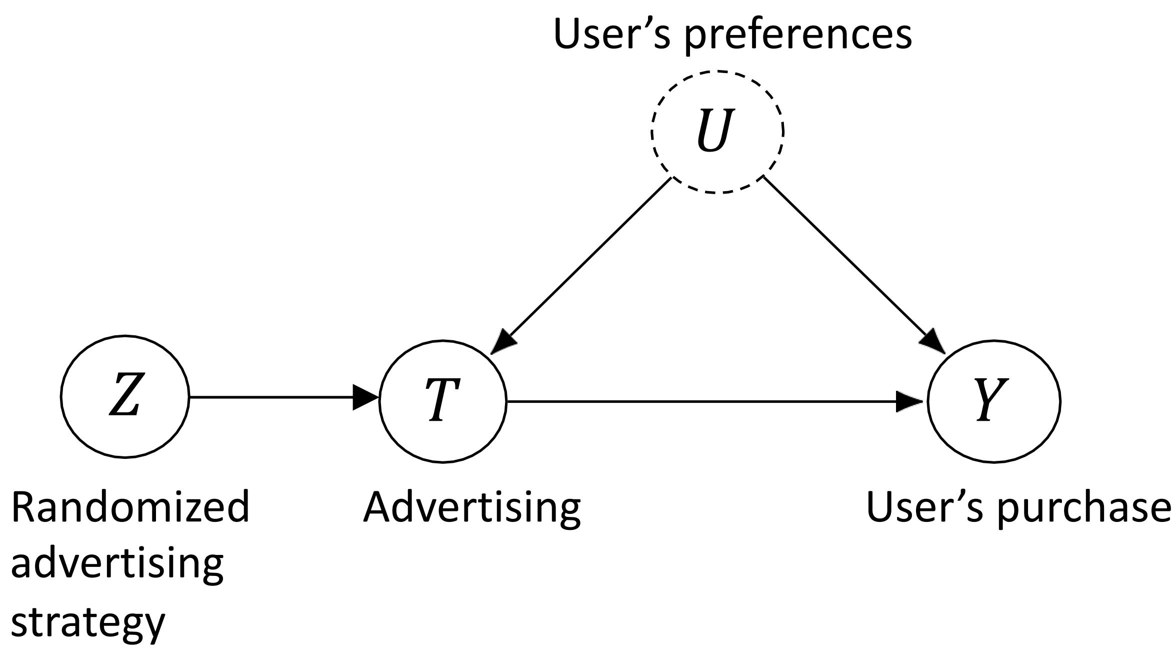
More Examples

- Example in Medicine: Estimating the Causal Effect of a new medication (Drug T) on patient's health outcomes.



More Examples

- Example in Economics: Estimating the Causal Effect of advertising on user's purchase.



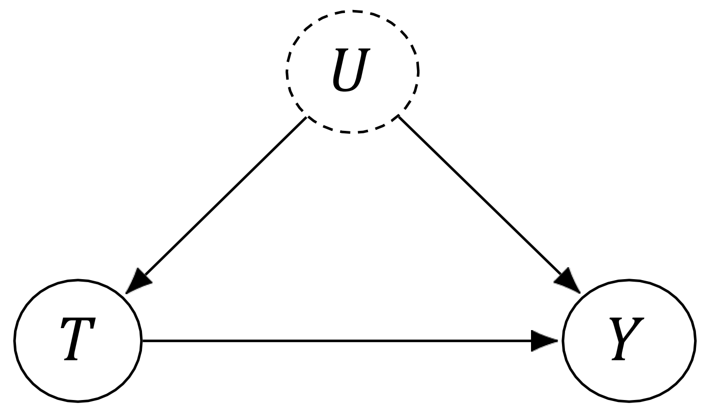
Why do we need IVs?

- Most causal inference methods rely on unconfoundedness assumption.
- Otherwise, there will be residual bias in causal estimates.

- Some methods can relax unconfoundedness assumption and enable causal effect estimation with unmeasured confounders.
  - IV-based estimation is one of them, and has been widely-used in real-world causal analysis!

Problem Setting

- Suppose we want to estimate the ATE of T on Y.
- Unobserved confounders U exist.



Outline

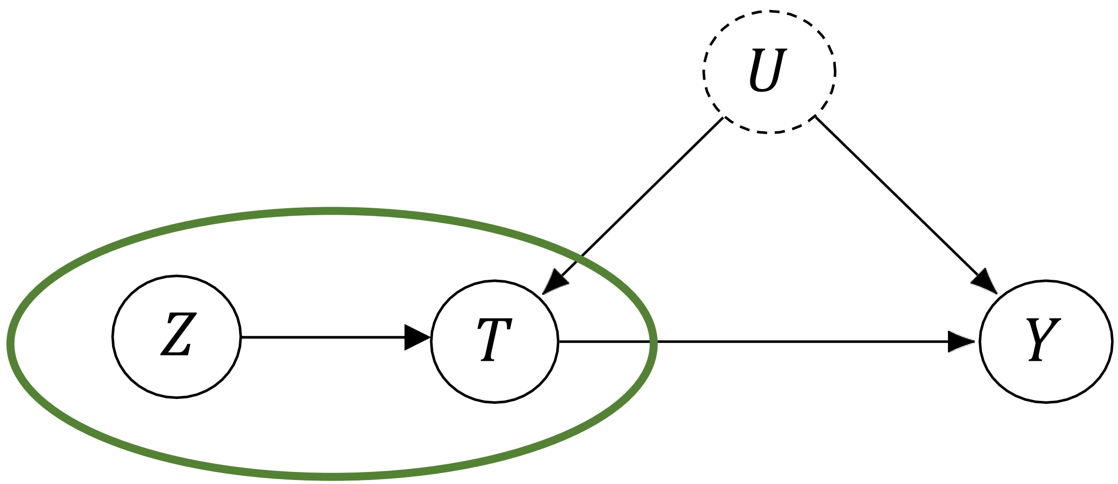
- Front-door Adjustment (recap)
- Instrumental variables
  - What is Instrumental Variable
  - 3 Assumptions of Instrumental Variable
  - Linear setting
  - 2SLS
  - Non-parametric identification
- Proxy variables for unobserved confounders

3 Assumptions for IVs

- A variable Z is an instrument if it meets three instrumental assumptions:
  - Relevance

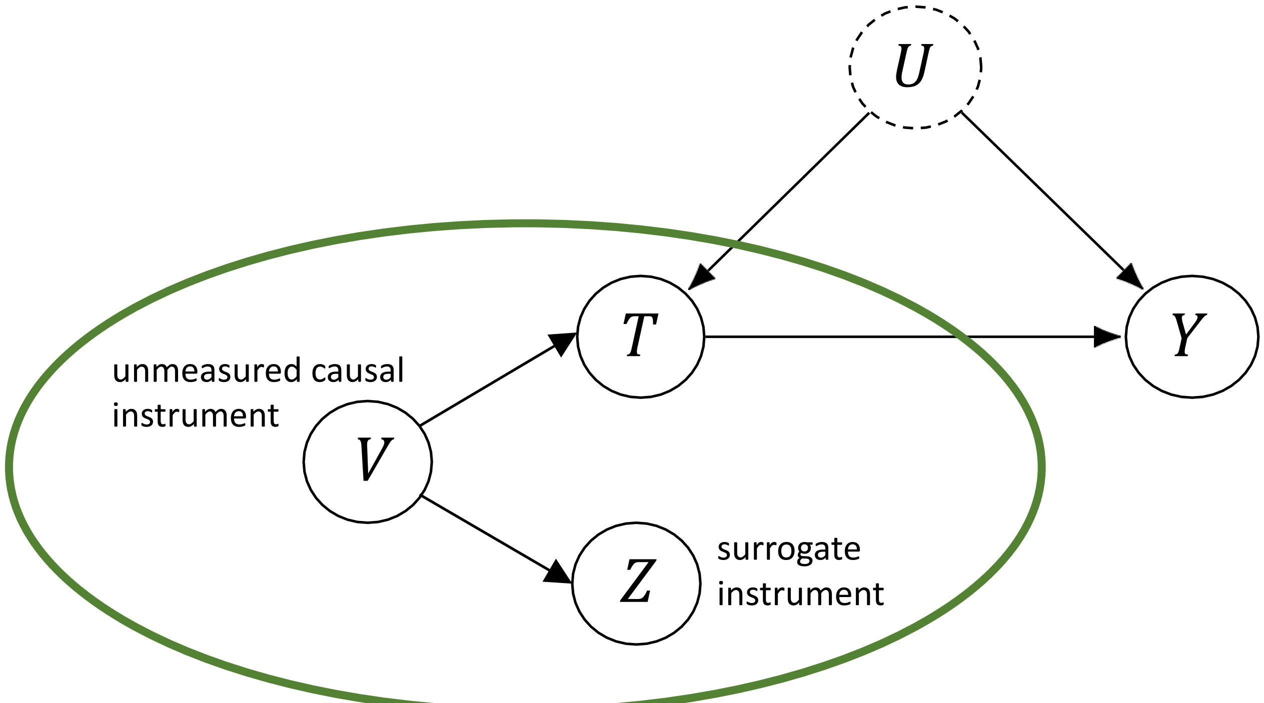
Assumption 1: Relevance

- Z is associated with T



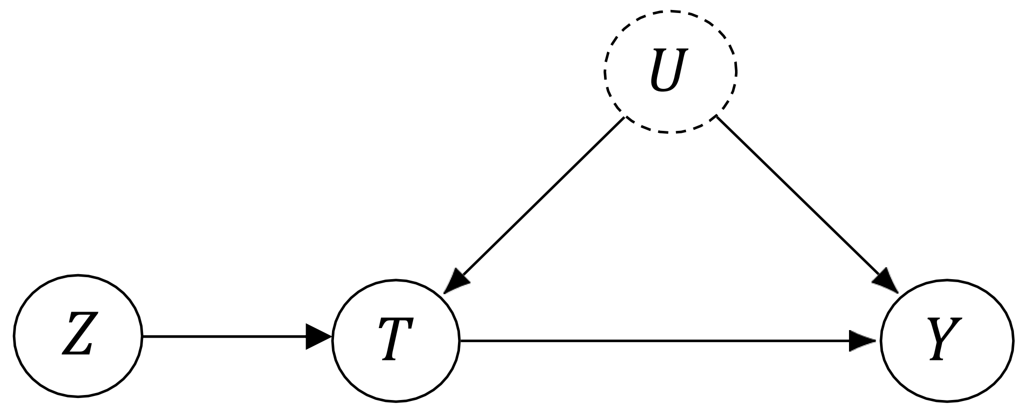
Assumption 1: Relevance

- Z is associated with T



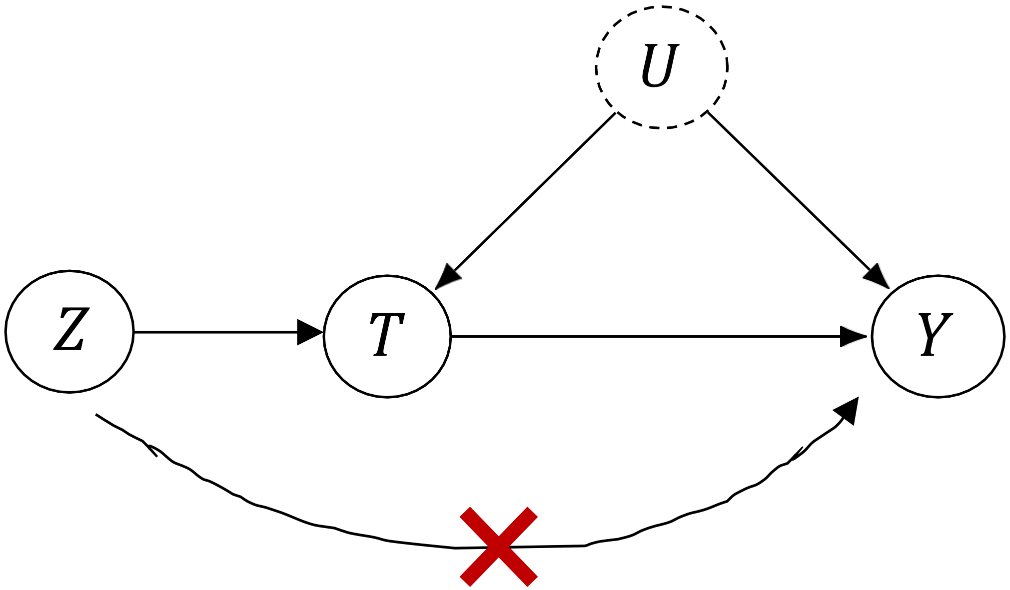
Assumption 2： Exclusion

- Z does not affect Y except through its potential effect on T. In other words, The causal effect of Z on Y is fully mediated by T



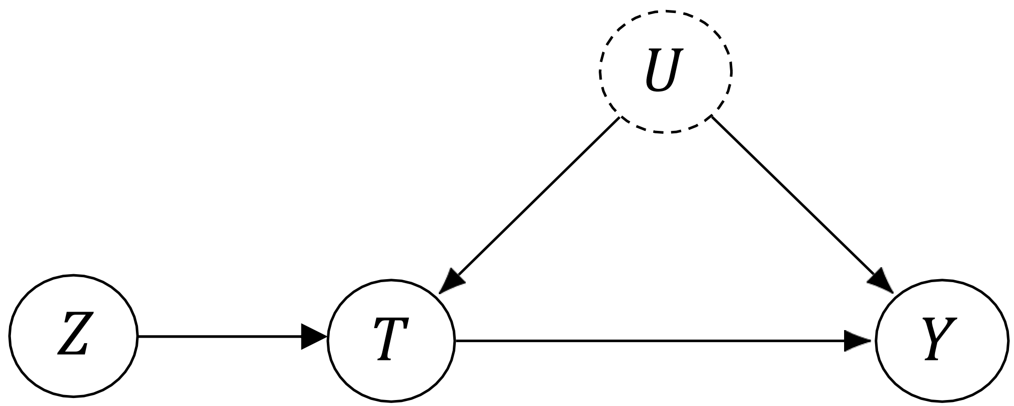
Assumption 2： Exclusion

- Z does not affect Y except through its potential effect on T. In other words, The causal effect of Z on Y is fully mediated by T



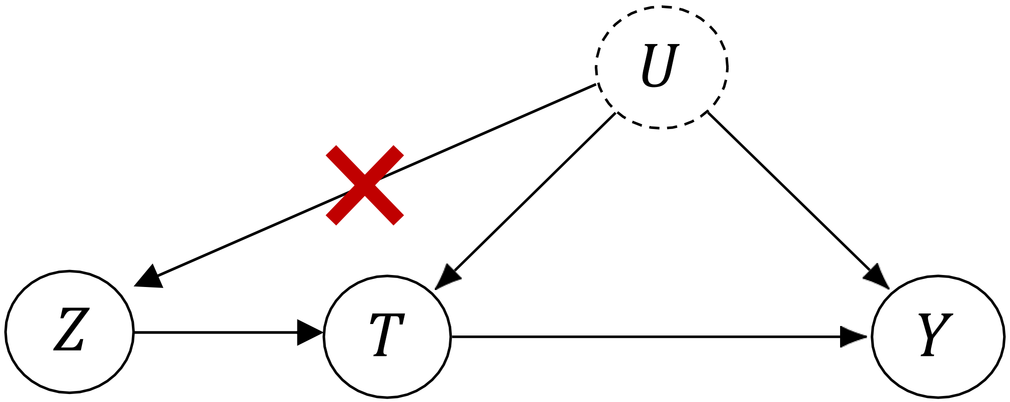
Assumption 3: Instrumental Unconfoundedness

- Z and Y do not share causes (i.e., Z is unconfounded)



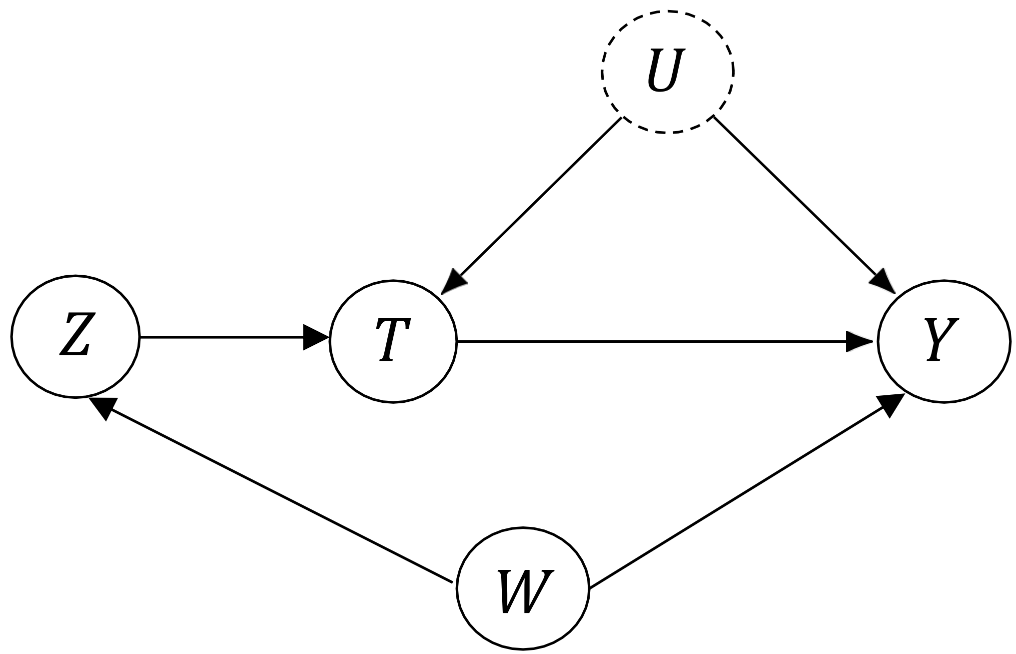
Assumption 3: Instrumental Unconfoundedness

- Z and Y do not share causes (i.e., Z is unconfounded)



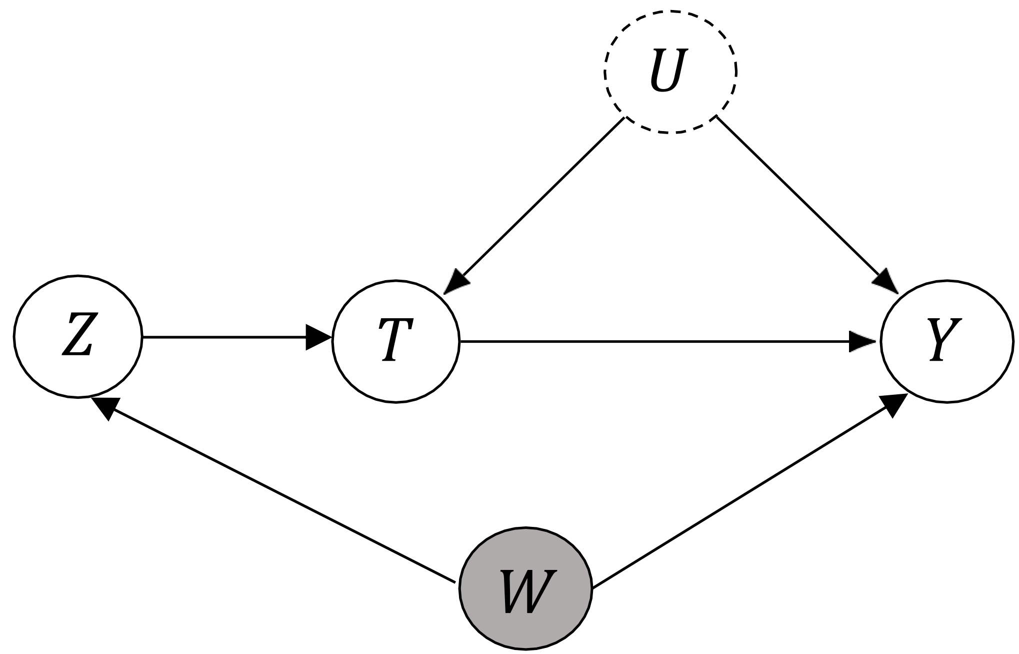
Weaker Assumption: Conditional Instruments

- A weaker version of Assumption 3:



Weaker Assumption: Conditional Instruments

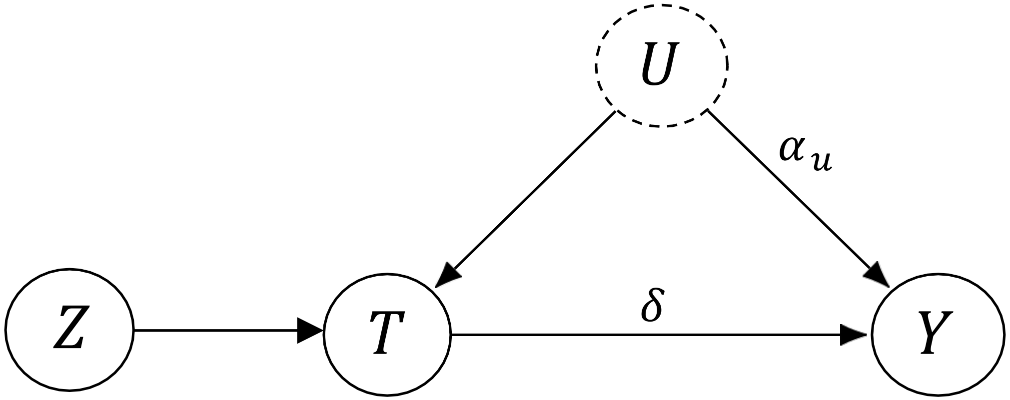
- A weaker version of Assumption 3: Unconfoundedness after conditioning on observed variables



Outline

- Front-door Adjustment (recap)
- Instrumental variables
    - What is Instrumental Variable
    - 3 Assumptions of Instrumental Variable
    - Linear setting
    - 2SLS
    - Non-parametric identification
- Proxy variables for unobserved confounders

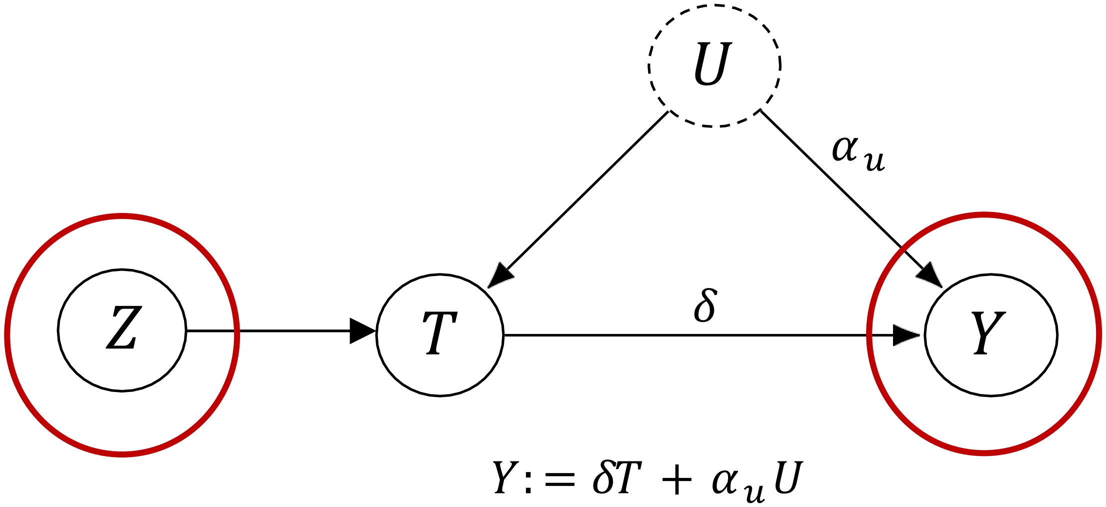
Start from a Linear Setting



$$Y := \delta T + \alpha_u U$$

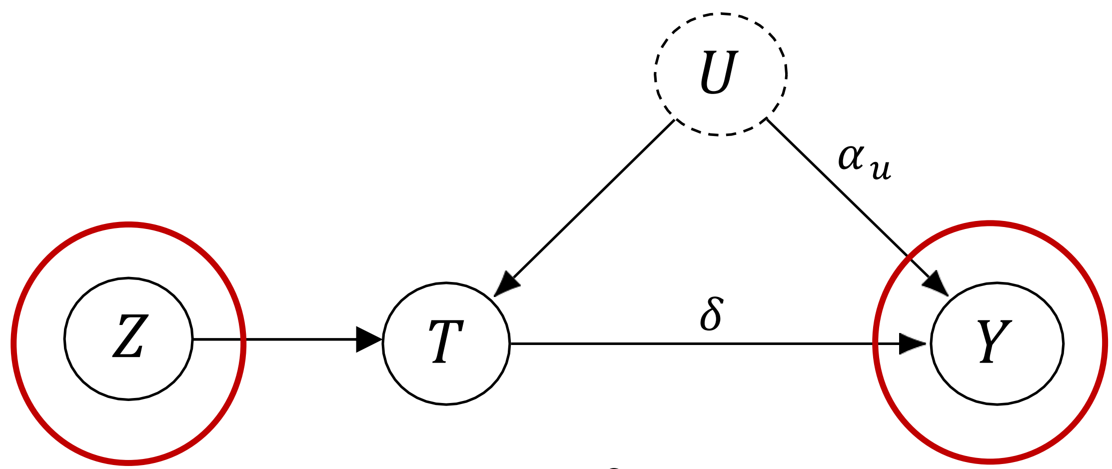
Start from a Linear Setting

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



Start from a Linear Setting

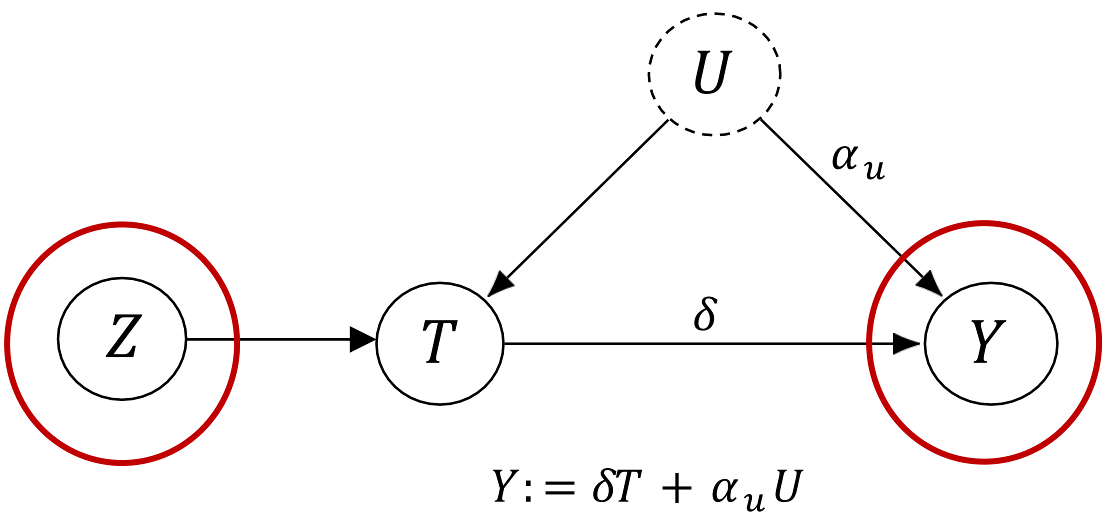
- $E[Y|Z=1] - E[Y|Z=0]$
  $= E[\underline{\delta T + \alpha_u U}|Z=1] - E[\underline{\delta T + \alpha_u U}|Z=0]$



$\underline{Y := \delta T + \alpha_u U}$

Start from a Linear Setting

- $E[Y|Z=1] - E[Y|Z=0]$
  $= E[\delta T + \alpha_u U|Z=1] - E[\delta T + \alpha_u U|Z=0]$
  $= \delta(E[T|Z=1] - E[T|Z=0]) + \alpha_u(E[U|Z=1] - E[U|Z=0])$



Start from a Linear Setting

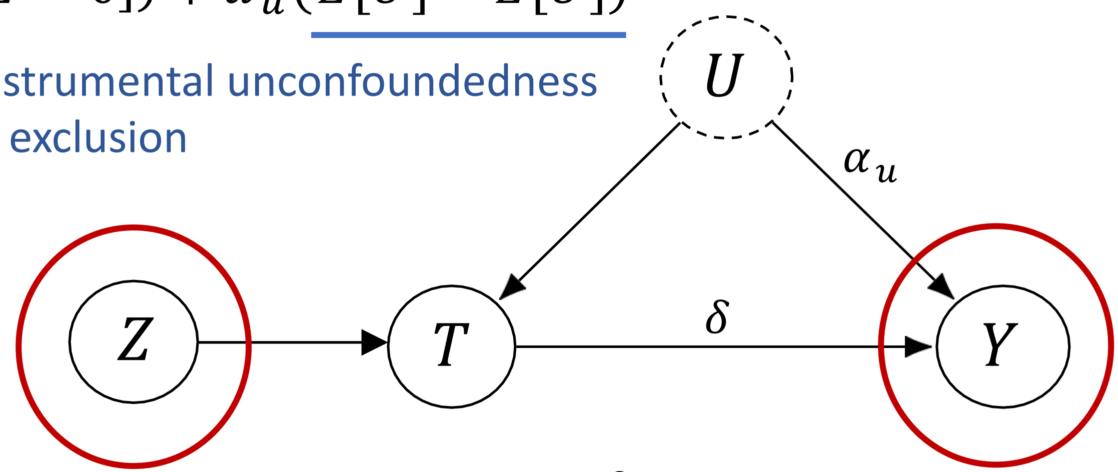
- $E[Y|Z=1] - E[Y|Z=0]$

  $= E[\delta T + \alpha_u U|Z=1] - E[\delta T + \alpha_u U|Z=0]$

  $= \delta(E[T|Z=1] - E[T|Z=0]) + \alpha_u(\underline{E[U|Z=1] - E[U|Z=0]})$

  $= \delta(E[T|Z=1] - E[T|Z=0]) + \alpha_u(\underline{E[U] - E[U]})$

Instrumental unconfoundedness
& exclusion



$Y := \delta T + \alpha_u U$

Start from a Linear Setting

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

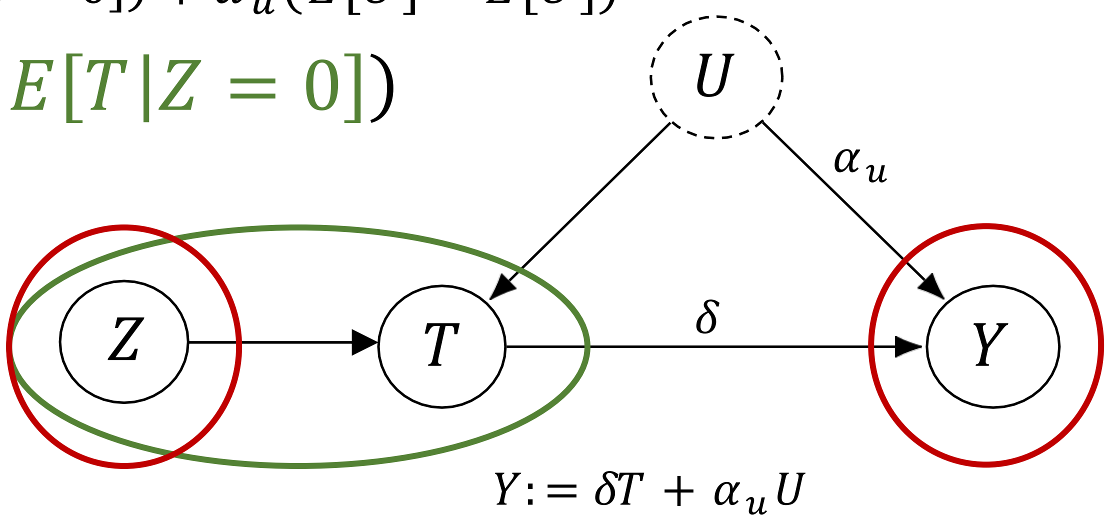
$= E[\delta T + \alpha_u U|Z=1] - E[\delta T + \alpha_u U|Z=0]$

$= \delta(E[T|Z=1]-E[T|Z=0]) + \alpha_u(E[U|Z=1]-E[U|Z=0])$

$= \delta(E[T|Z=1]-E[T|Z=0]) + \cancel{\alpha_u(E[U]-E[U])}$

$= \delta(E[T|Z=1]-E[T|Z=0])$

$\delta = ?$



Start from a Linear Setting

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

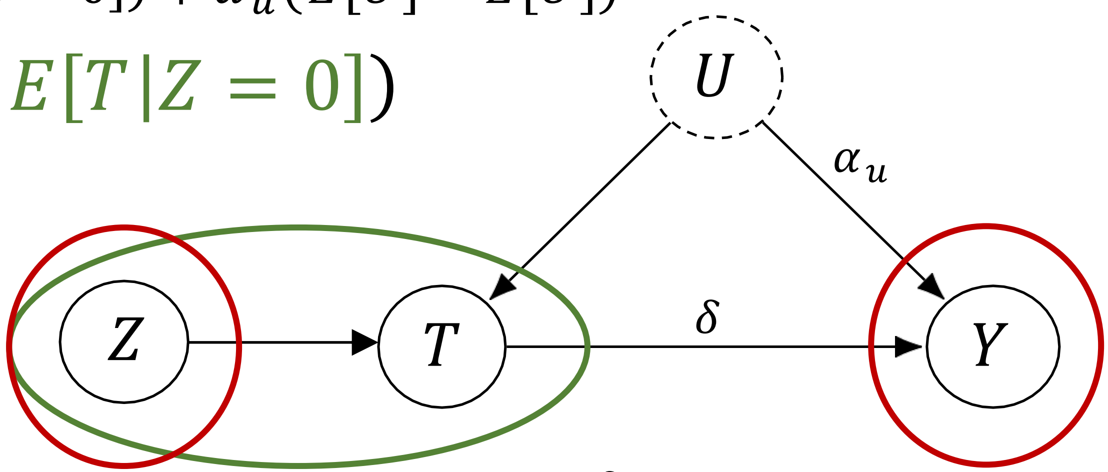
$= E[\delta T + \alpha_u U|Z=1] - E[\delta T + \alpha_u U|Z=0]$

$= \delta(E[T|Z=1] - E[T|Z=0]) + \alpha_u(E[U|Z=1] - E[U|Z=0])$

$= \delta(E[T|Z=1] - E[T|Z=0]) + \cancel{\alpha_u(E[U] - E[U])}$

$= \delta(E[T|Z=1] - E[T|Z=0])$

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



$Y := \delta T + \alpha_u U$

Start from a Linear Setting

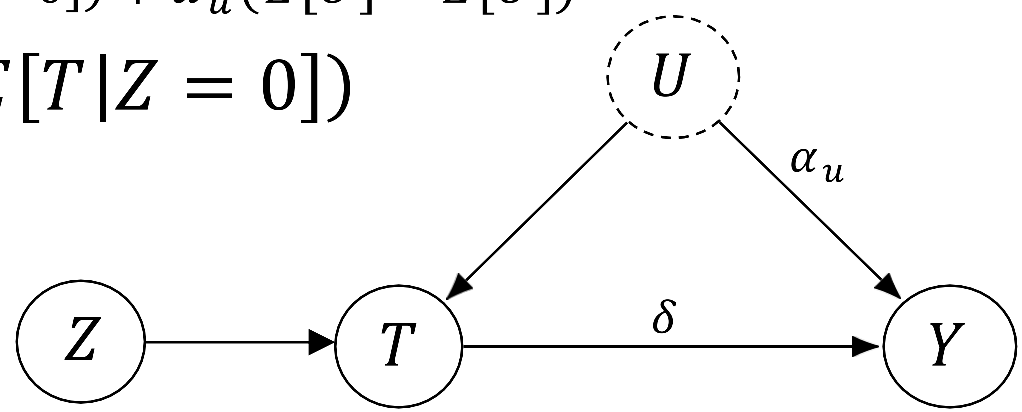
- $E[Y|Z=1] - E[Y|Z=0]$

  $= E[\delta T + \alpha_u U|Z=1] - E[\delta T + \alpha_u U|Z=0]$

  $= \delta(E[T|Z=1] - E[T|Z=0]) + \alpha_u(E[U|Z=1] - E[U|Z=0])$

  $= \delta(E[T|Z=1] - E[T|Z=0]) + \cancel{\alpha_u(E[U] - E[U])}$

  $= \delta(E[T|Z=1] - E[T|Z=0])$



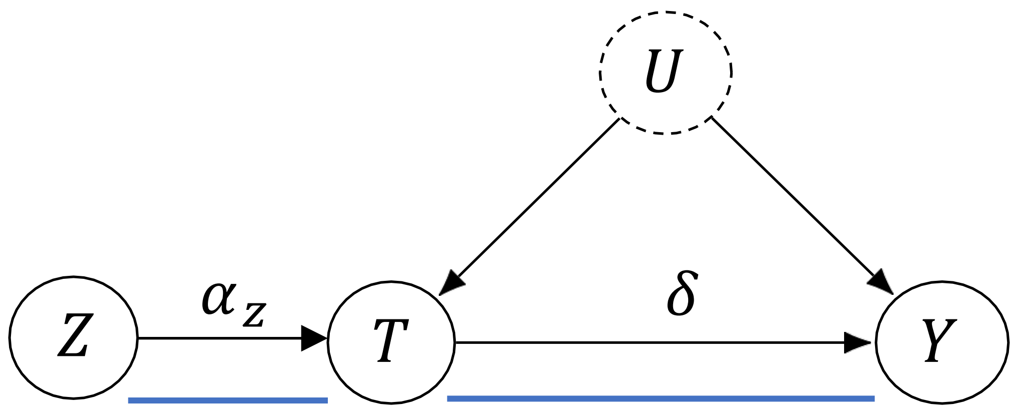
$Y := \delta T + \alpha_u U$

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

The denominator is not 0, based on relevance assumption

Multiplying Path Coefficients in Linear Setting

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

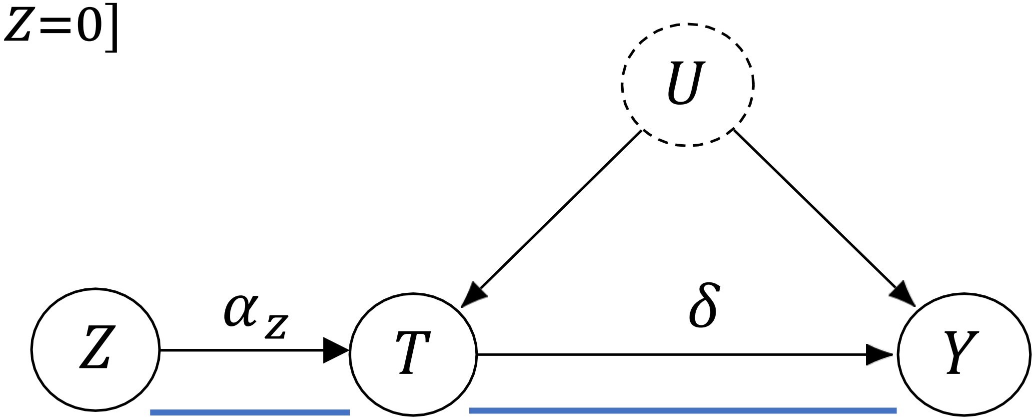


$$Y := \delta T + \alpha_u U$$

Multiplying Path Coefficients in Linear Setting

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

$$\delta = \frac{\alpha_Z \delta}{E[T|Z=1] - E[T|Z=0]}$$

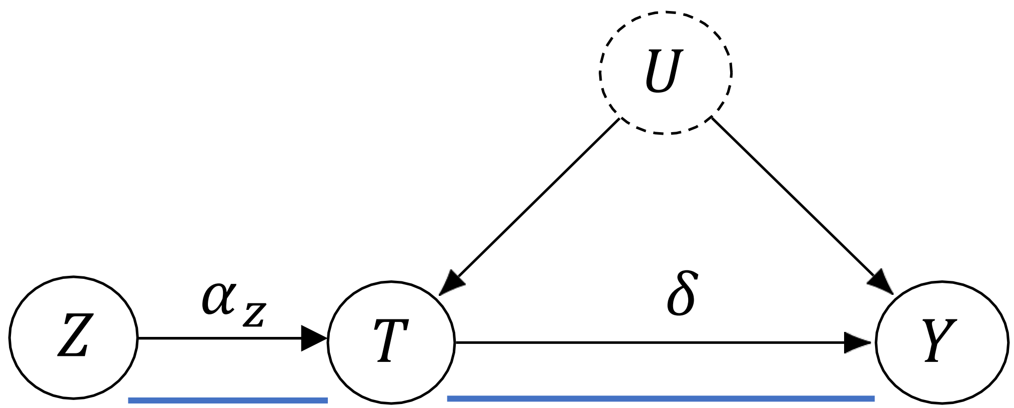


$$Y := \delta T + \alpha_u U$$

Multiplying Path Coefficients in Linear Setting

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

$$\delta = \frac{\alpha_Z \delta}{\alpha_Z}$$



$Y := \delta T + \alpha_u U$

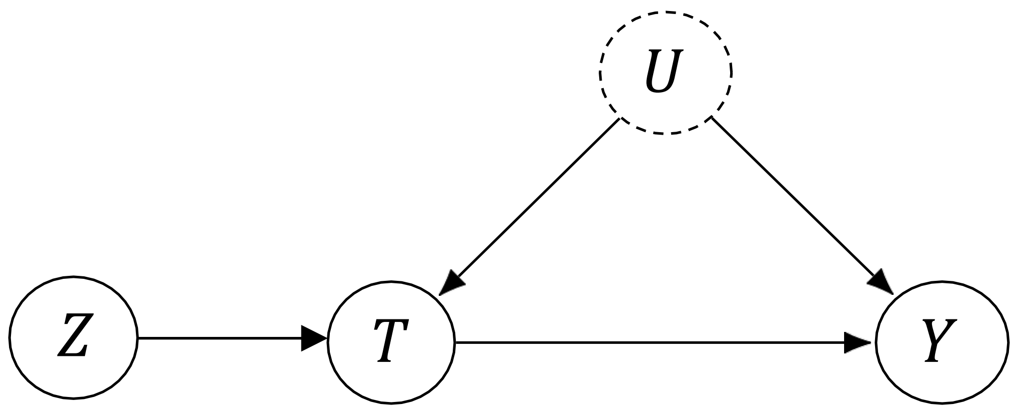
Wald Estimator

- Wald estimand: $\delta = \frac{E[Y|Z=1]-E[Y|Z=0]}{E[T|Z=1]-E[T|Z=0]}$
- Wald estimator: $\hat{\delta} = \frac{\frac{1}{n_1}\sum_{i:z_i=1} Y_i - \frac{1}{n_0}\sum_{i:z_i=0} Y_i}{\frac{1}{n_1}\sum_{i:z_i=1} T_i - \frac{1}{n_0}\sum_{i:z_i=0} T_i}$

Continuous Linear Setting

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

What if T and Z are continuous?



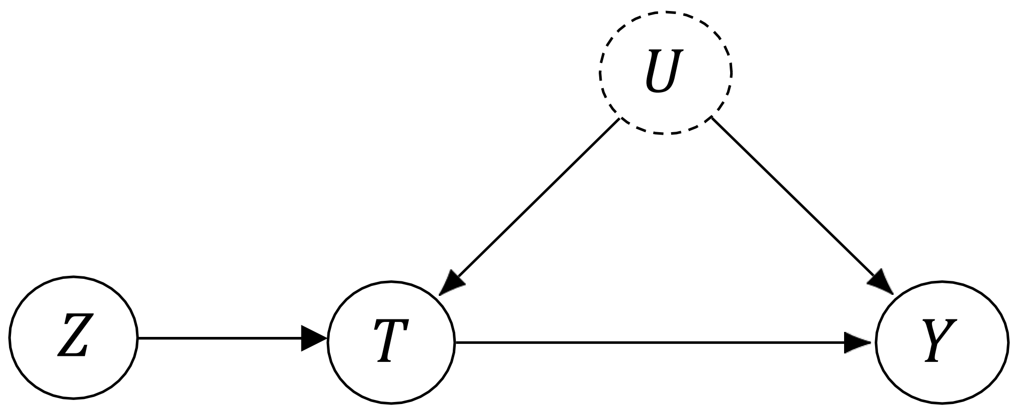
$$Y := \delta T + \alpha_u U$$

Continuous Linear Setting

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

What if T and Z are continuous?

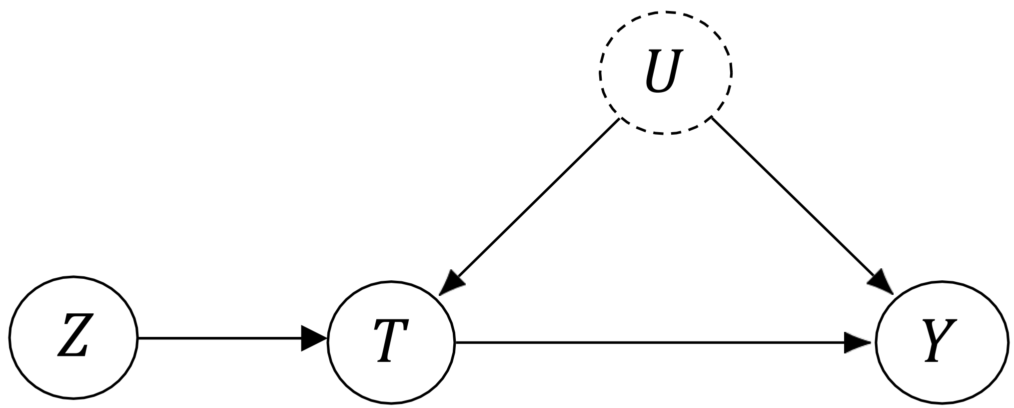
$$\delta = \frac{\mathrm{Cov}(Y,Z)}{\mathrm{Cov}(T,Z)}$$



$$Y := \delta T + \alpha_u U$$

Continuous Linear Setting

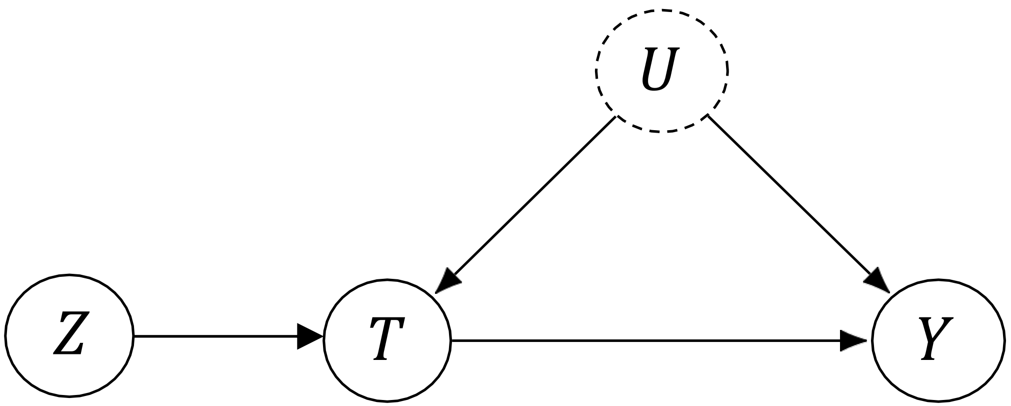
$\mathrm{Cov}(Y, Z)$



$Y := \delta T + \alpha_u U$

Continuous Linear Setting

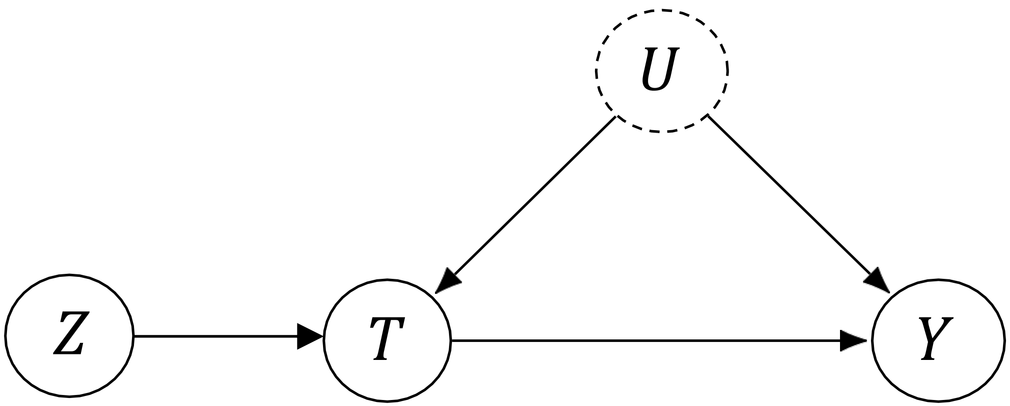
$\mathrm{Cov}(Y, Z) = E[YZ] - E[Y]E[Z]$



$Y := \delta T + \alpha_u U$

Continuous Linear Setting

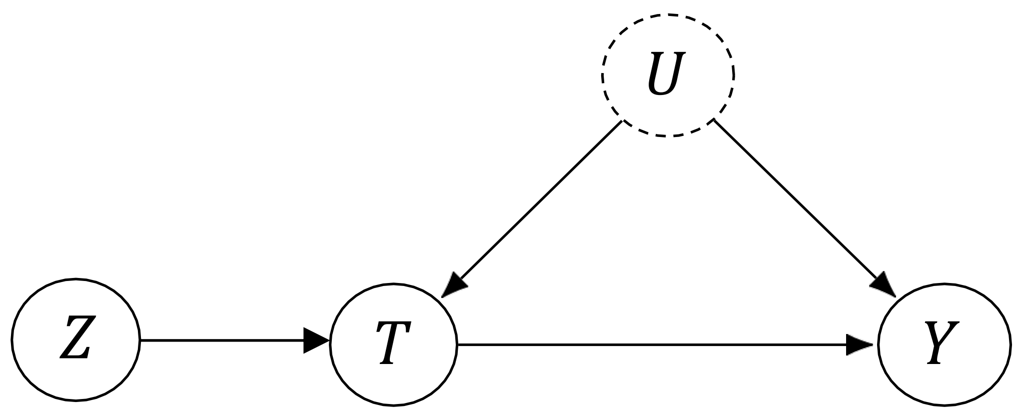
$$\begin{aligned}\mathrm{Cov}(Y, Z) &= E[YZ] - E[Y]E[Z] \\ &= E[(\delta T + \alpha_u U)Z] - E[\delta T + \alpha_u U]E[Z]\end{aligned}$$



$Y := \delta T + \alpha_u U$

Continuous Linear Setting

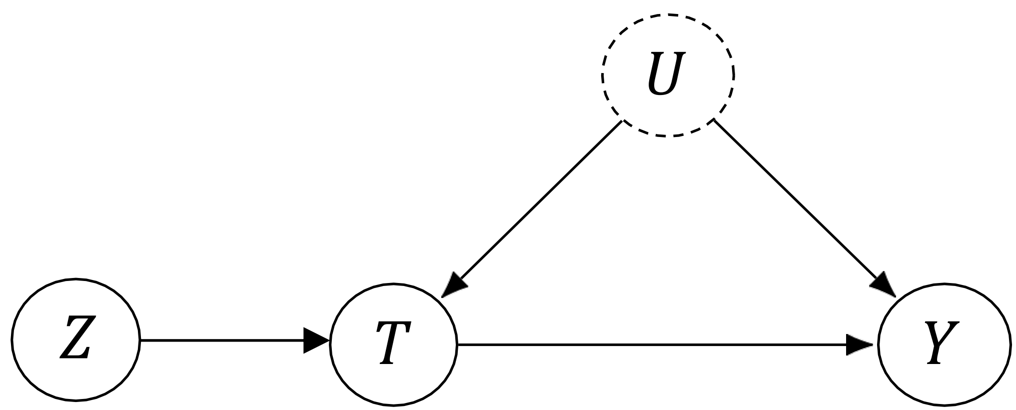
$$\begin{aligned}\mathrm{Cov}(Y,Z) &= E[YZ] - E[Y]E[Z] \\ &= E[(\delta T + \alpha_u U)Z] - E[\delta T + \alpha_u U]E[Z] \\ &= \delta E[TZ] + \alpha_u E[UZ] - \delta E[T]E[Z] - \alpha_u E[U]E[Z]\end{aligned}$$



$Y := \delta T + \alpha_u U$

Continuous Linear Setting

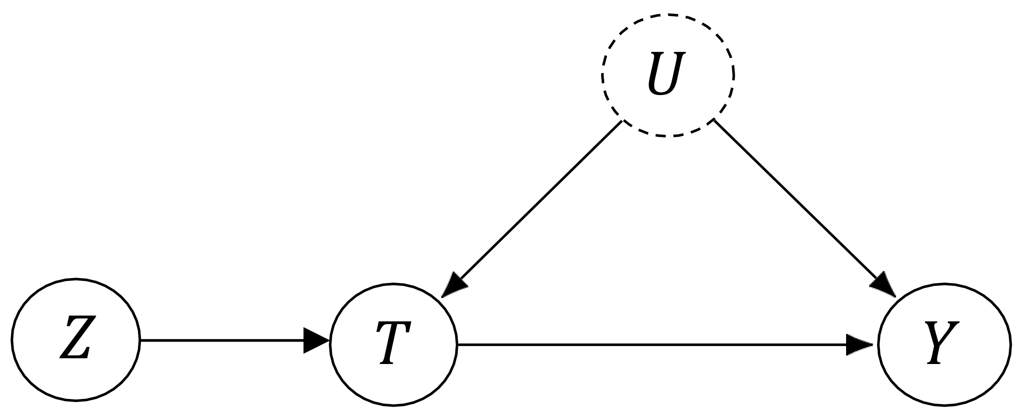
$$
\begin{aligned}
\mathrm{Cov}(Y, Z) &= E[YZ] - E[Y]E[Z] \\
&= E[(\delta T + \alpha_u U)Z] - E[\delta T + \alpha_u U]E[Z] \\
&= \delta E[TZ] + \alpha_u E[UZ] - \delta E[T]E[Z] - \alpha_u E[U]E[Z] \\
&= \delta(E[TZ] - E[T]E[Z]) + \alpha_u (E[UZ] - E[U]E[Z])
\end{aligned}
$$



$Y := \delta T + \alpha_u U$

Continuous Linear Setting

$$\begin{aligned}
\mathrm{Cov}(Y, Z) &= E[YZ] - E[Y]E[Z] \\
&= E[(\delta T + \alpha_u U)Z] - E[\delta T + \alpha_u U]E[Z] \\
&= \delta E[TZ] + \alpha_u E[UZ] - \delta E[T]E[Z] - \alpha_u E[U]E[Z] \\
&= \delta(E[TZ] - E[T]E[Z]) + \alpha_u(E[UZ] - E[U]E[Z]) \\
&= \delta \mathrm{Cov}(T, Z) + \alpha_u \mathrm{Cov}(U, Z)
\end{aligned}$$

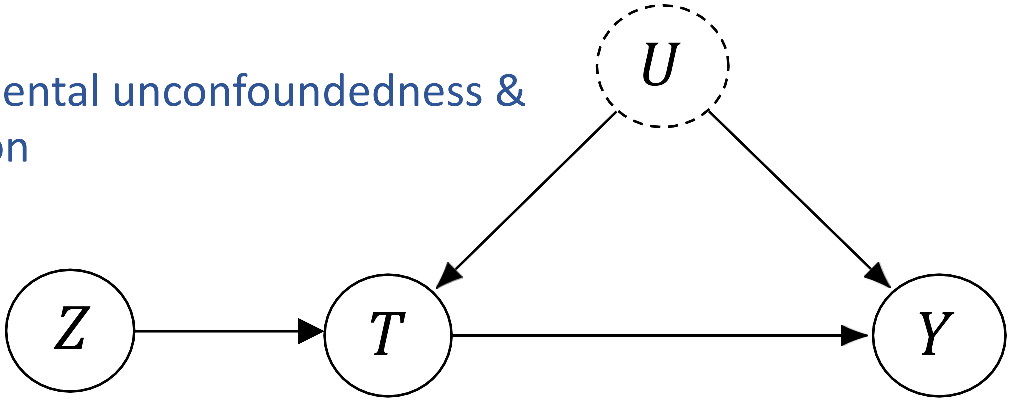


$$Y := \delta T + \alpha_u U$$

Continuous Linear Setting

$$\begin{aligned}
\mathrm{Cov}(Y,Z) &= E[YZ] - E[Y]E[Z] \\
&= E[(\delta T + \alpha_u U)Z] - E[\delta T + \alpha_u U]E[Z] \\
&= \delta E[TZ] + \alpha_u E[UZ] - \delta E[T]E[Z] - \alpha_u E[U]E[Z] \\
&= \delta(E[TZ] - E[T]E[Z]) + \alpha_u(E[UZ] - E[U]E[Z]) \\
&= \delta \mathrm{Cov}(T,Z) + \alpha_u \mathrm{Cov}(U,Z) \\
&= \delta \mathrm{Cov}(T,Z)
\end{aligned}$$

Instrumental unconfoundedness & exclusion

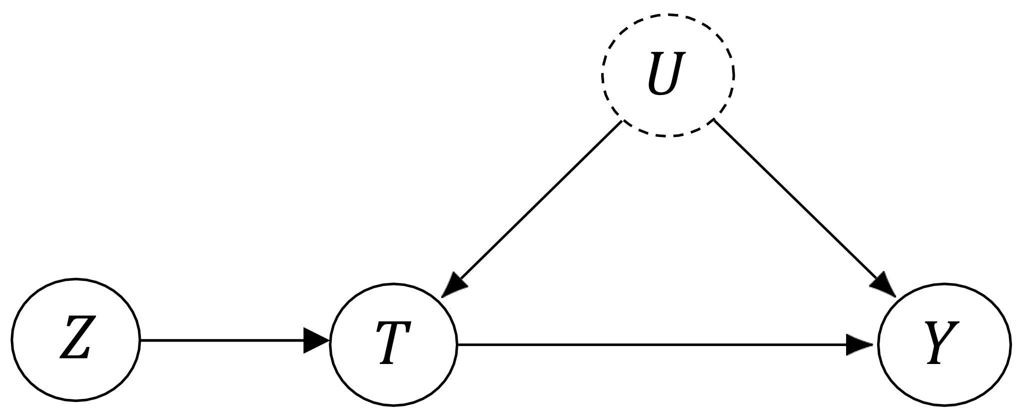


$$Y := \delta T + \alpha_u U$$

Continuous Linear Setting

$$
\begin{aligned}
\mathrm{Cov}(Y, Z) &= E[YZ] - E[Y]E[Z] \\
&= E[(\delta T + \alpha_u U)Z] - E[\delta T + \alpha_u U]E[Z] \\
&= \delta E[TZ] + \alpha_u E[UZ] - \delta E[T]E[Z] - \alpha_u E[U]E[Z] \\
&= \delta(E[TZ] - E[T]E[Z]) + \alpha_u(E[UZ] - E[U]E[Z]) \\
&= \delta \mathrm{Cov}(T, Z) + \alpha_u \mathrm{Cov}(U, Z) \\
&= \delta \mathrm{Cov}(T, Z)
\end{aligned}
$$

$$\delta = \frac{\mathrm{Cov}(Y,Z)}{\mathrm{Cov}(T,Z)}$$



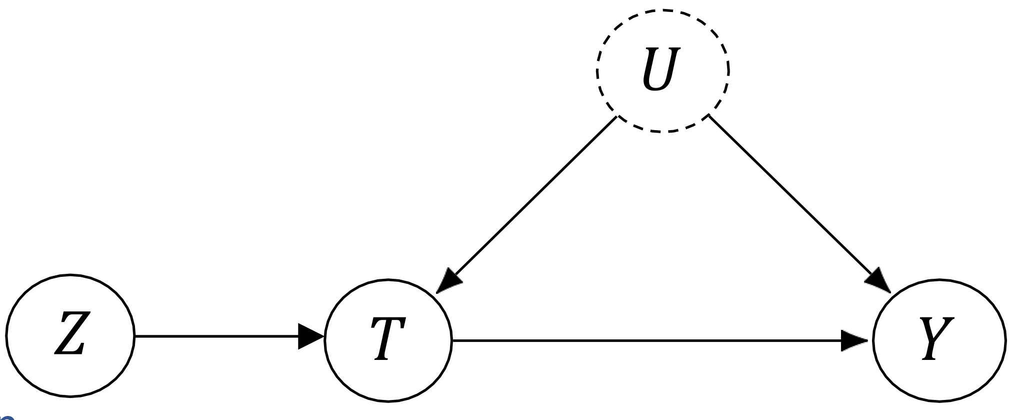
$Y := \delta T + \alpha_u U$

Continuous Linear Setting

$$
\begin{aligned}
\mathrm{Cov}(Y,Z) &= E[YZ] - E[Y]E[Z] \\
&= E[(\delta T + \alpha_u U)Z] - E[\delta T + \alpha_u U]E[Z] \\
&= \delta E[TZ] + \alpha_u E[UZ] - \delta E[T]E[Z] - \alpha_u E[U]E[Z] \\
&= \delta(E[TZ] - E[T]E[Z]) + \alpha_u(E[UZ] - E[U]E[Z]) \\
&= \delta \mathrm{Cov}(T,Z) + \alpha_u \mathrm{Cov}(U,Z) \\
&= \delta \mathrm{Cov}(T,Z)
\end{aligned}
$$

$$\delta = \frac{\mathrm{Cov}(Y,Z)}{\mathrm{Cov}(T,Z)}$$

The denominator is not 0, based on relevance assumption



$$Y := \delta T + \alpha_u U$$

Continuous Linear Setting: Estimator 1

- Estimand:

$$\delta = \frac{\mathrm{Cov}(Y,Z)}{\mathrm{Cov}(T,Z)}$$

- Estimator:

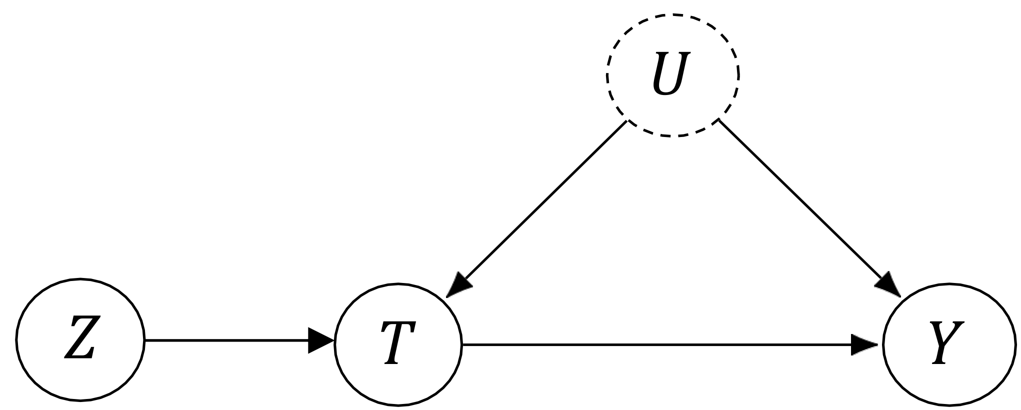
$$\hat{\delta} = \frac{\widehat{\mathrm{Cov}}\,(Y,Z)}{\widehat{\mathrm{Cov}}(T,Z)}$$

Outline

- Front-door Adjustment (recap)
- Instrumental variables
  - What is Instrumental Variable
  - 3 Assumptions of Instrumental Variable
  - Linear setting
  - 2SLS
  - Non-parametric identification
- Proxy variables for unobserved confounders

Two-Stage Least Squares Estimator (2SLS)

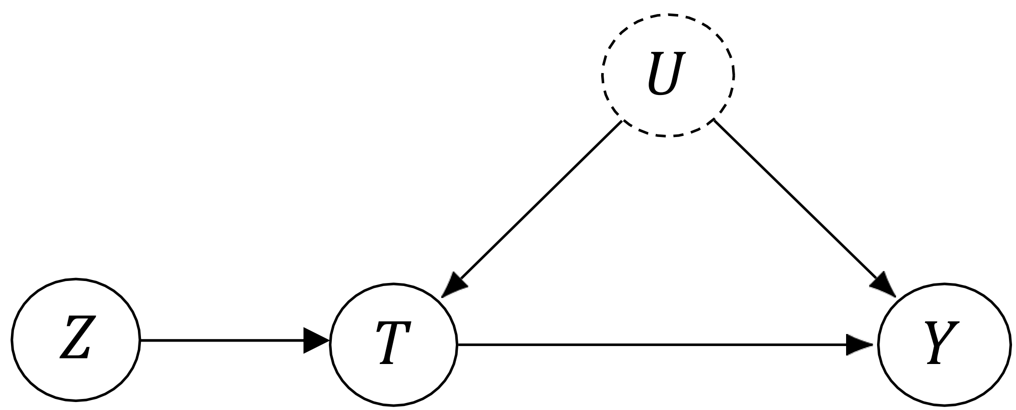
- 2SLS is one of the most widely-used IV-based causal effect estimation method
- 2 stages are included, which decompose the causal effect $Z \rightarrow Y$ into two parts:
  - $Z \rightarrow T$
  - $T \rightarrow Y$



Two-Stage Least Squares Estimator (2SLS)

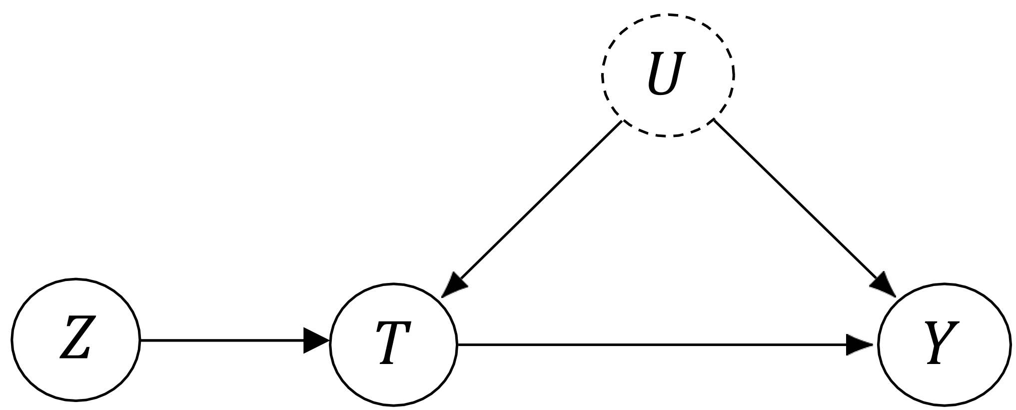
- Stage 1: Linearly regress T on Z to estimate $E[T|Z]$. This gives us the projection of T onto Z: $\hat{T}$

$$\hat{T} = \hat{E}[T|Z]$$



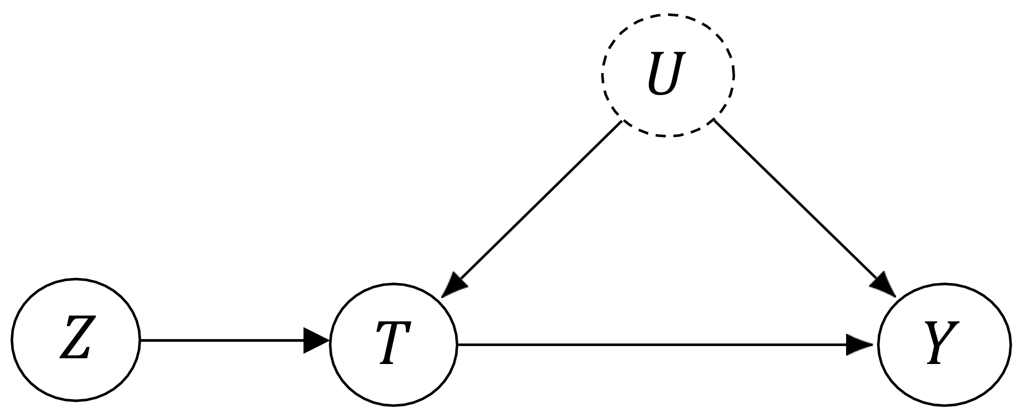
Two-Stage Least Squares Estimator (2SLS)

- Stage 1: Linearly regress T on Z to estimate $E[T|Z]$. This gives us the projection of T onto Z: $\hat{T}$
- Stage 2: Linearly regress Y on $\hat{T}$ to estimate $E[Y|\hat{T}]$. Obtain the estimate $\hat{\delta}$ as the fitted coefficient in front of $\hat{T}$.



Two-Stage Least Squares Estimator (2SLS)

- Stage 1: Linearly regress T on Z to estimate $E[T|Z]$. This gives us the projection of T onto Z: $\hat{T}$
- Stage 2: Linearly regress Y on $\hat{T}$ to estimate $E[Y|\hat{T}]$. Obtain the estimate $\hat{\delta}$ as the fitted coefficient in front of $\hat{T}$.

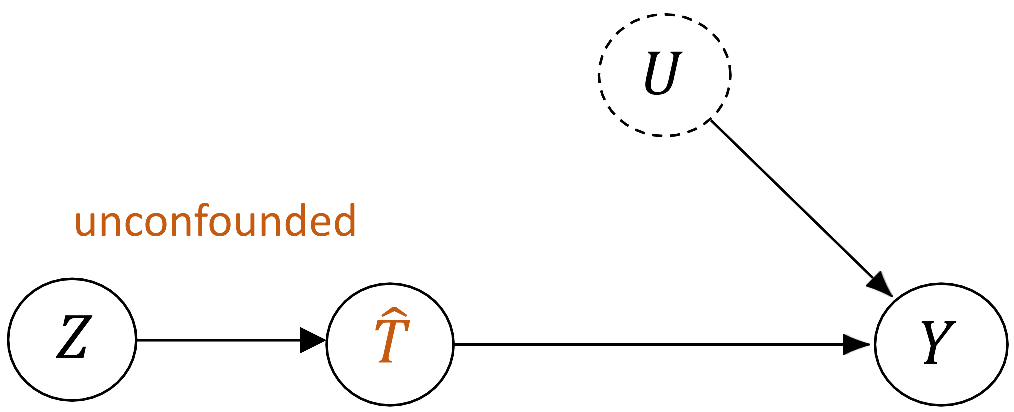


Also works in the binary setting

Intuition of 2SLS

- The key intuition: as the IVs are unconfounded, the predicted treatment from the first stage can provide more randomization, and thus it can help mitigate the confounding bias brought by hidden confounders in the second stage.

$\hat{T} = \hat{E}[T|Z]$



Outline

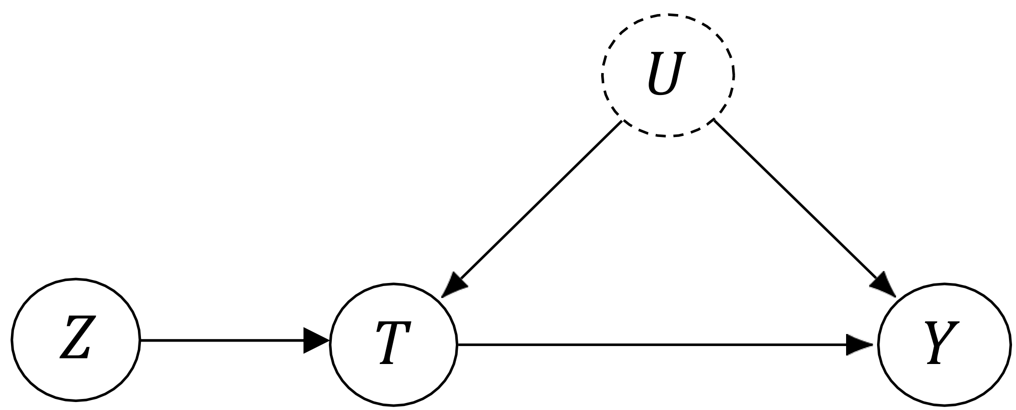
- Front-door Adjustment (recap)
- Instrumental variables
    - What is Instrumental Variable
    - 3 Assumptions of Instrumental Variable
    - Linear setting
    - 2SLS
    - Non-parametric identification
- Proxy variables for unobserved confounders

What about nonparametric case?

- We can identify ATE in linear case
- However, we cannot nonparametrically identify ATE.

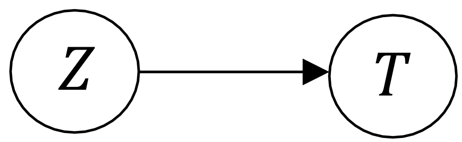
Non-parametric Identification

- Notations:
  - $Y(1)$ and $Y(0)$ represent $Y(T=1)$ and $Y(T=0)$
  - $T(1)$ and $T(0)$ represent $T(Z=1)$ and $T(Z=0)$



4 Strata

- The data can be divided into 4 strata (subgroups) based on how IV influence the treatment



- Compiles (always obey): $T(Z = 1) = 1, T(Z = 0) = 0$
- Defiers (always violate): $T(Z = 1) = 0, T(Z = 0) = 1$
- Always-takers: $T(Z = 1) = 1, T(Z = 0) = 1$
- Never-takers: $T(Z = 1) = 0, T(Z = 0) = 0$

4 Strata

- The data can be divided into 4 strata (subgroups) based on how IV influence the treatment

$$Z \rightarrow T$$

- Compiles (always obey): $T(Z=1)=1, T(Z=0)=0$
- Defiers (always violate): $T(Z=1)=0, T(Z=0)=1$
- Always-takers: $T(Z=1)=1, T(Z=0)=1$
- Never-takers: $T(Z=1)=0, T(Z=0)=0$

Z does not influence T

Monotonicity Assumption

$$\forall i, \qquad T_i(Z=1) \geq T_i(Z=0)$$

(No Defiers)

Local ATE Identification

- With Monotonicity Assumption, we can try to identify local ATE with instruments.

Local ATE Identification

- With Monotonicity Assumption, we can try to identify local ATE with instruments.

$$
\begin{aligned}
& E[Y(Z=1)-Y(Z=0)] \\
= \; & E[Y(Z=1)-Y(Z=0) \mid T(1)=1, T(0)=0]\, P(T(1)=1, T(0)=0) && \text{Compliers} \\
& + E[Y(Z=1)-Y(Z=0) \mid T(1)=0, T(0)=1]\, P(T(1)=0, T(0)=1) && \text{Defiers} \\
& + E[Y(Z=1)-Y(Z=0) \mid T(1)=1, T(0)=1]\, P(T(1)=1, T(0)=1) && \text{Always-takers} \\
& + E[Y(Z=1)-Y(Z=0) \mid T(1)=0, T(0)=0]\, P(T(1)=0, T(0)=0) && \text{Never-takers}
\end{aligned}
$$

Local ATE Identification

- With Monotonicity Assumption, we can try to identify local ATE with instruments.

$$
\begin{aligned}
& E[Y(Z=1)-Y(Z=0)] \\
= \ & E[Y(Z=1)-Y(Z=0) \mid T(1)=1, T(0)=0]\, P(T(1)=1, T(0)=0) && \text{Compliers} \\
& \cancel{+\, E[Y(Z=1)-Y(Z=0) \mid T(1)=0, T(0)=1]\, P(T(1)=0, T(0)=1)} && \cancel{\text{Defiers}} \\
& +\, E[Y(Z=1)-Y(Z=0) \mid T(1)=1, T(0)=1]\, P(T(1)=1, T(0)=1) && \text{Always-takers} \\
& +\, E[Y(Z=1)-Y(Z=0) \mid T(1)=0, T(0)=0]\, P(T(1)=0, T(0)=0) && \text{Never-takers}
\end{aligned}
$$

Local ATE Identification

- With Monotonicity Assumption, we can try to identify local ATE with instruments.

$$
\begin{aligned}
& E[Y(Z=1)-Y(Z=0)] \\
= \; & E[Y(Z=1)-Y(Z=0)|T(1)=1,T(0)=0]\, P(T(1)=1,T(0)=0) && \text{Compliers} \\
& \cancel{+E[Y(Z=1)-Y(Z=0)\,|\,T(1)=0,T(0)=1]\, P(T(1)=0,T(0)=1)} && \cancel{\text{Defiers}} \\
& \boxed{\begin{aligned} &+E[Y(Z=1)-Y(Z=0)\,|\,T(1)=1,T(0)=1]\, P(T(1)=1,T(0)=1) && \text{Always-takers} \\ &+E[Y(Z=1)-Y(Z=0)\,|\,T(1)=0,T(0)=0]\, P(T(1)=0,T(0)=0) && \text{Never-takers} \end{aligned}}
\end{aligned}
$$

Z does not causally influence T, so based on Exclusion Assumption, Z doe not causally influence Y

Local ATE Identification

- With Monotonicity Assumption, we can try to identify local ATE with instruments.

$$
\begin{aligned}
& E[Y(Z=1)-Y(Z=0)] \\
&= E[Y(Z=1)-Y(Z=0) \mid T(1)=1, T(0)=0]\, P(T(1)=1, T(0)=0) && \text{Compliers} \\
&\quad \cancel{+ E[Y(Z=1)-Y(Z=0) \mid T(1)=0, T(0)=1]\, P(T(1)=0, T(0)=1)} && \text{Defiers} \\
&\quad \boxed{\begin{array}{l} \cancel{+ E[Y(Z=1)-Y(Z=0) \mid T(1)=1, T(0)=1]\, P(T(1)=1, T(0)=1)} \quad \text{Always-takers} \\ \cancel{+ E[Y(Z=1)-Y(Z=0) \mid T(1)=0, T(0)=0]\, P(T(1)=0, T(0)=0)} \quad \text{Never-takers} \end{array}} \\
&= E[Y(Z=1)-Y(Z=0) \mid T(1)=1, T(0)=0]\, P(T(1)=1, T(0)=0)
\end{aligned}
$$

Z does not causally influence T, so based on Exclusion Assumption, Z doe not causally influence Y

Local ATE Identification

- With Monotonicity Assumption, we can try to identify local ATE with instruments.

$$
\begin{aligned}
& E[Y(Z=1)-Y(Z=0)] \\
&= E[Y(Z=1)-Y(Z=0) \mid T(1)=1, T(0)=0]\, P(T(1)=1, T(0)=0) \quad \text{Compliers} \\
&\quad \cancel{+ E[Y(Z=1)-Y(Z=0) \mid T(1)=0, T(0)=1]\, P(T(1)=0, T(0)=1)} \quad \cancel{\text{Defiers}} \\
&\quad \boxed{\cancel{+ E[Y(Z=1)-Y(Z=0) \mid T(1)=1, T(0)=1]\, P(T(1)=1, T(0)=1)} \quad \cancel{\text{Always-takers}} \\
&\quad \cancel{+ E[Y(Z=1)-Y(Z=0) \mid T(1)=0, T(0)=0]\, P(T(1)=0, T(0)=0)} \quad \cancel{\text{Never-takers}}} \\
&= E[Y(Z=1)-Y(Z=0) \mid T(1)=1, T(0)=0]\, P(T(1)=1, T(0)=0)
\end{aligned}
$$

Local ATE (LATE) or Complier average causal effect (CACE):

$$E[Y(Z=1)-Y(Z=0) \mid T(1)=1, T(0)=0] = \frac{E[Y(Z=1)-Y(Z=0)]}{P(T(1)=1, T(0)=0)}$$

Local ATE Identification

- With Monotonicity Assumption, we can try to identify local ATE with instruments.

$$
\begin{aligned}
& E[Y(Z=1)-Y(Z=0)] \\
= \; & E[Y(Z=1)-Y(Z=0) \mid T(1)=1, T(0)=0]\, P(T(1)=1, T(0)=0) && \text{Compliers} \\
& \cancel{+\, E[Y(Z=1)-Y(Z=0) \mid T(1)=0, T(0)=1]\, P(T(1)=0, T(0)=1)} && \cancel{\text{Defiers}} \\
& \boxed{\begin{aligned} & \cancel{+\, E[Y(Z=1)-Y(Z=0) \mid T(1)=1, T(0)=1]\, P(T(1)=1, T(0)=1)} && \cancel{\text{Always-takers}} \\ & \cancel{+\, E[Y(Z=1)-Y(Z=0) \mid T(1)=0, T(0)=0]\, P(T(1)=0, T(0)=0)} && \cancel{\text{Never-takers}} \end{aligned}} \\
= \; & E[Y(Z=1)-Y(Z=0) \mid T(1)=1, T(0)=0]\, P(T(1)=1, T(0)=0)
\end{aligned}
$$

Local ATE (LATE) or Complier average causal effect (CACE):

$$
E[Y(Z=1)-Y(Z=0) \mid T(1)=1, T(0)=0] = \frac{E[Y(Z=1)-Y(Z=0)]}{P(T(1)=1, T(0)=0)} = \frac{E[Y \mid Z=1]-E[Y \mid Z=0]}{P(T(1)=1, T(0)=0)}
$$

Instrumental unconfoundedness

Local ATE Identification

$$E[Y(Z=1)-Y(Z=0)|T(1)=1,T(0)=0] = \frac{E[Y(Z=1)-Y(Z=0)]}{P(T(1)=\ 1,T(0)=\ 0)} = \frac{E[Y|Z=1]-E[Y|Z=0]}{P(T(1)=\ 1,T(0)=\ 0)}$$

$$\begin{aligned}
&P(T(1)=1,T(0)=0)\\
&= 1 - P(T=0|\ Z=1) - P(T=1\ |Z=0)\\
&= 1 - \big(1 - P(T=1|Z=1)\big) - P(T=1|Z=0)\\
&= P(T=1|\ Z=1)\ - P(T=1|\ Z=0)\\
&=\ E[T\mid Z\ =\ 1]\ - E[T\mid Z\ =\ 0]
\end{aligned}$$

Local ATE Identification

$$E[Y(Z=1)-Y(Z=0)|T(1)=1,T(0)=0]=\frac{E[Y(Z=1)-Y(Z=0)]}{P(T(1)=\ 1,T(0)=\ 0)}=\frac{E[Y|Z=1]-E[Y|Z=0]}{P(T(1)=\ 1,T(0)=\ 0)}$$

$$\begin{aligned}
&P(T(1)=1,T(0)=0)\\
&=1-P(T=0|\ Z=1)-P(T=1\ |Z=0)\\
&=1-\big(1-P(T=1|Z=1)\big)-P(T=1|Z=0)\\
&=P(T=1|\ Z=1)\ -P(T=1|\ Z=0)\\
&=\ E[T\ |\ Z\ =\ 1]\ -E[T\ |\ Z\ =\ 0]
\end{aligned}$$

Wald estimand!

$$E[Y(Z=1)-Y(Z=0)|T(1)=1,T(0)=0]=\frac{E[Y|Z=1]-E[Y|Z=0]}{E[T|Z=1]-E[T|Z=0]}$$

Problems

- Monotonicity assumption may be violated in many occasions
- Even if Monotonicity assumption is satisfied, only CACE (for compliers) can be identified, not ATE for the whole population

More General Settings

- Nonparametric Outcome with Additive Noise

$$Y := f(T, W) + U$$

flexible model such as a deep neural network

Hartford J, Lewis G, Leyton-Brown K, et al. Deep IV: A flexible approach for counterfactual prediction[C]//International Conference on Machine Learning. PMLR, 2017: 1414-1423.

More General Settings

- Sometimes, even though we cannot point identify the ATE, we can still find bound it (set identification)

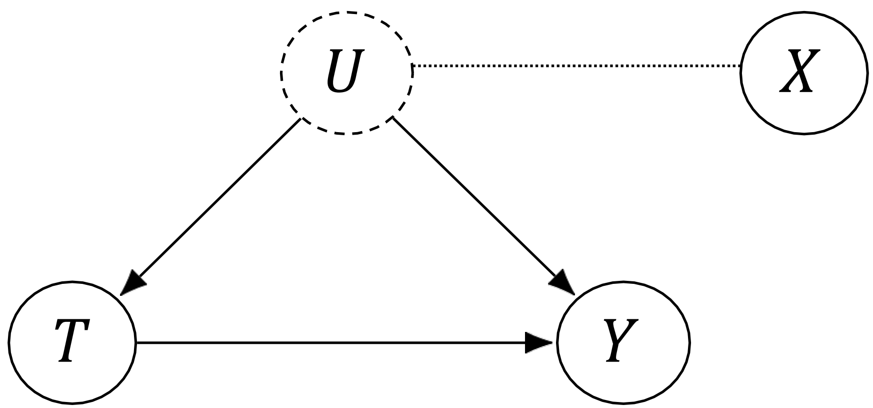
Kilbertus N, Kusner M J, Silva R. A class of algorithms for general instrumental variable models[J]. Advances in Neural Information Processing Systems, 2020, 33: 20108-20119.

Outline

- Front-door Adjustment (recap)
- Instrumental variables
  - What is Instrumental Variable
  - 3 Assumptions of Instrumental Variable
  - Linear setting
  - 2SLS
  - Non-parametric identification
- Proxy variables for unobserved confounders

Proxy Variables for Unobserved Confounders

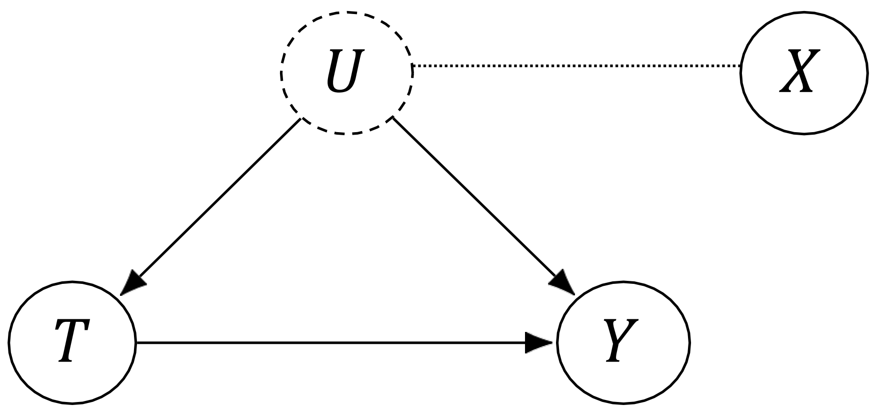
- Even though unobserved variables exist, there may be observed proxy variables X of the confounder available
- It has been proved that with at least two independent proxy variables satisfying a certain rank condition, the causal effect can be nonparametrically identified [1]



[1] Miao W, Geng Z, Tchetgen Tchetgen E J. Identifying causal effects with proxy variables of an unmeasured confounder[J]. Biometrika, 2018, 105(4): 987-993.

Intuition of Proxy Variables for Unobserved Confounders

- 1. Using proxy variables to infer the unobserved confounders
- 2. Adjusting for the inferred confounders, e.g., using backdoor adjustment



References

- Robins J, Hernan M A. Causal inference: what if[J]. Found Agnostic Stat, 2020: 235-281. Chapter 16
- Brady Neal. Introduction to Causal Inference. Chapter 9.
- Hartford J, Lewis G, Leyton-Brown K, et al. Deep IV: A flexible approach for counterfactual prediction[C]//International Conference on Machine Learning. PMLR, 2017: 1414-1423.
- Miao W, Geng Z, Tchetgen Tchetgen E J. Identifying causal effects with proxy variables of an unmeasured confounder[J]. Biometrika, 2018, 105(4): 987-993.

Reading Materials

- Hartford J, Lewis G, Leyton-Brown K, et al. Deep IV: A flexible approach for counterfactual prediction[C]//International Conference on Machine Learning. PMLR, 2017: 1414-1423.

- Robins J, Hernan M A. Causal inference: what if[J]. Found Agnostic Stat, 2020: 235-281. Chapter 16. Instrumental Variable Estimation.

Thank you!