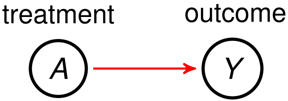
A Brief Introduction to Causal Inference

Linbo Wang

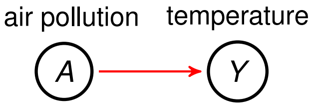
UNIVERSITY OF TORONTO

BIRS Workshop on Causality, Extremes and Climate
Kelowna, BC
June 27, 2022

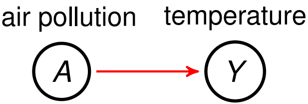
A Motivating Example



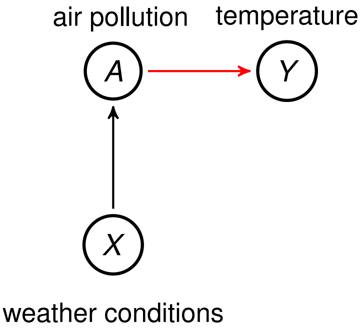
A Motivating Example



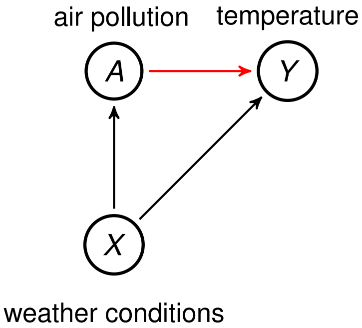
Association $\neq$ Causation!



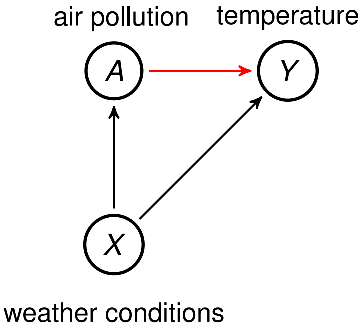
Association $\neq$ Causation!



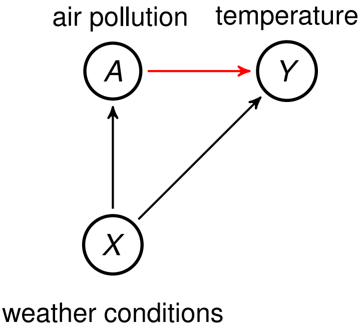
Association $\neq$ Causation!



Association = Causation + Confounding



Association = Causation + Confounding



Causation leads to **actionable insights**!

Canonical Approach in Statistics

Regression Modeling

$$Y = \beta_0 + \beta_1 A + \beta_2 X + \epsilon_Y$$

Canonical Approach in Statistics

Regression Modeling

$$Y = \beta_0 + \beta_1 A + \beta_2 X + \epsilon_Y$$

- When can we interpret the coefficient β_1 as causal?

- What can we do otherwise?

Canonical Approach in Statistics

Regression Modeling

$$Y = \beta_0 + \beta_1 A + \beta_2 X + \epsilon_Y$$

- When can we interpret the coefficient β_1 as causal?
  - What do we mean by a causal effect?
  - What are the conditions for the two quantities to equal?
- What can we do otherwise?

Frameworks in Causal Inference

Mechanistic versus agnostic approach to causal inference

Identification and estimation of causal effects

Causal Estimation Under No Unmeasured Confounding

Instrumental Variables

Frameworks in Causal Inference

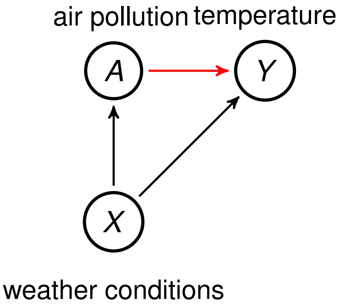
Mechanistic versus agnostic approach to causal inference

Identification and estimation of causal effects

Causal Estimation Under No Unmeasured Confounding

Instrumental Variables

Framework 1: Structural Equation Models

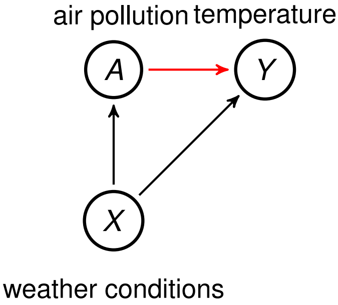


Non-Parametric Structural Equation Models with Independent Errors (NPSEM-IE)

$$X = \epsilon_X$$
$$A = f_A(X, \epsilon_A)$$
$$Y = f_Y(A, X, \epsilon_Y)$$

where $\epsilon_X \perp\!\!\!\perp \epsilon_A \perp\!\!\!\perp \epsilon_Y$

Framework 1: Structural Equation Models

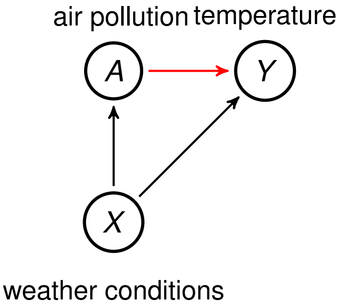


Example: Linear Structural Equation Models with Independent Errors

$$X = \epsilon_X$$
$$A = \alpha_0 + \alpha_1 X + \epsilon_A$$
$$Y = \beta_0 + \beta_1 A + \beta_2 X + \epsilon_Y$$

where $\epsilon_X \perp\!\!\!\perp \epsilon_A \perp\!\!\!\perp \epsilon_Y$

Framework 1: Structural Equation Models



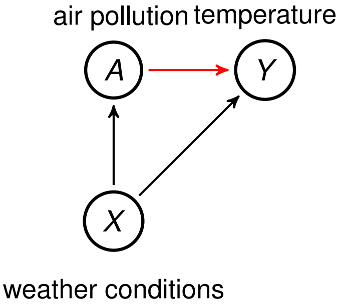
SEMs specify what happens in an observational world.

Coincides with parametric regression models:

$$X = \epsilon_X$$
$$A = \alpha_0 + \alpha_1 X + \epsilon_A$$
$$Y = \beta_0 + \beta_1 A + \beta_2 X + \epsilon_Y$$

where $\epsilon_X \perp\!\!\!\perp \epsilon_A \perp\!\!\!\perp \epsilon_Y$

Framework 1: Structural Equation Models



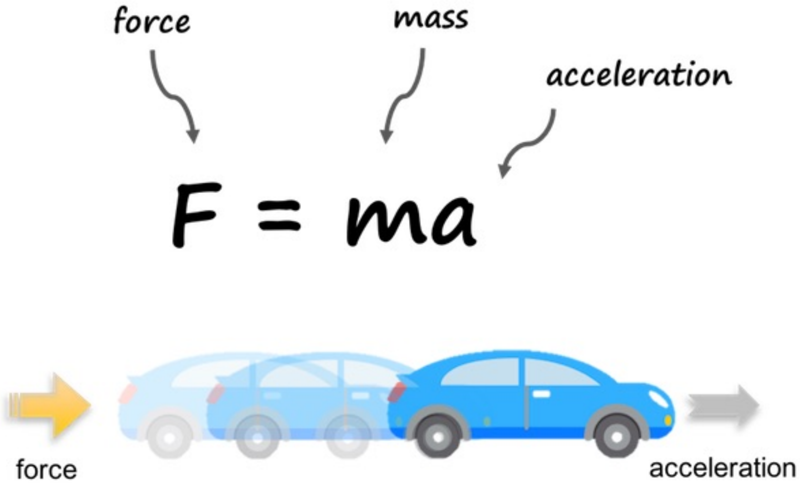
SEMs also specify what would happen in an experimental world!

Example: fixing A to 0

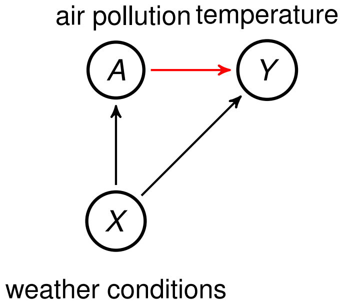
$$X = \epsilon_X$$
$$A = 0$$
$$Y = \beta_0 + \beta_1 \times 0 + \beta_2 X + \epsilon_Y$$

where $\epsilon_X \perp\!\!\!\perp \epsilon_A \perp\!\!\!\perp \epsilon_Y$

Analogy: Physics Law



Framework 1: Structural Equation Models



SEMs also specify what would happen in an experimental world!

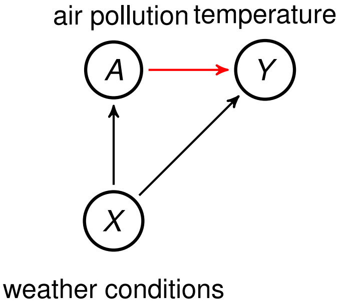
Example: fixing A to 0

$$X = \epsilon_X$$
$$A = 0$$
$$Y = \beta_0 + \beta_1 \times 0 + \beta_2 X + \epsilon_Y$$

where $\epsilon_X \perp\!\!\!\perp \epsilon_A \perp\!\!\!\perp \epsilon_Y$

Causal effect $\equiv \beta_1$

Framework 1: Structural Equation Models



Stability: The parameters of SEMs remain the same across different worlds

Regression coefficient = Causal effect!

Summary of Parametric SEMs

A mechanistic approach to causal inference

- Permits inferences with real-world interpretations and detailed predictions

Summary of Parametric SEMs

A mechanistic approach to causal inference

- Permits inferences with real-world interpretations and detailed predictions
- Rely on strong assumptions:
  - SEM: Assumes knowledge on the relationships among *all* the variables in the system
  - Linear SEM: Also makes parametric assumptions
  - Stability assumption: The coefficients remain constant across observational/experimental worlds

Summary of Parametric SEMs

A mechanistic approach to causal inference

- Permits inferences with real-world interpretations and detailed predictions
- Rely on strong assumptions:
  - SEM: Assumes knowledge on the relationships among *all* the variables in the system
  - Linear SEM: Also makes parametric assumptions
  - Stability assumption: The coefficients remain constant across observational/experimental worlds
- If these assumptions fail: how to even *define* the causal effects?

Framework 2: Potential Outcome (Agnostic Approach)

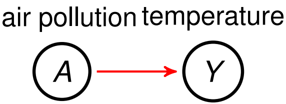
$Y(a) \equiv$ "the outcome Y that would have been observed if a subject had received treatment a"

Framework 2: Potential Outcome (Agnostic Approach)

$Y(a) \equiv$ "the outcome Y that would have been observed if a subject had received treatment a"

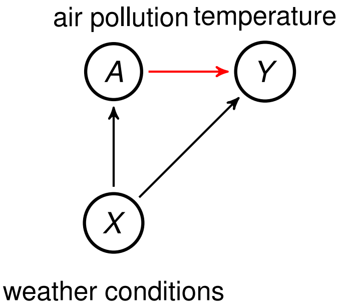
- Requires pre-specified treatment and outcome
- Does not require knowledge of the causal system: only A and Y but not X

Framework 2: Potential Outcome (Agnostic Approach)



$Y(1)$ = what the temperature in a location would be if there were air pollution in this location

NPSEM ⇒ Potential Outcomes

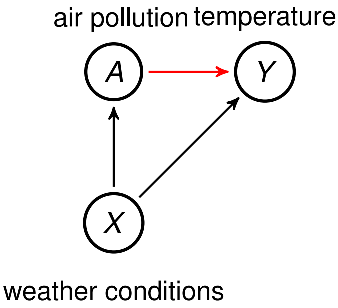


$Y(1)$ = what the temperature in a location would be if there were air pollution in this location

- Under a NPSEM

$$X = \epsilon_X$$
$$A = f_A(X, \epsilon_A)$$
$$Y = f_Y(A, X, \epsilon_Y)$$

We have $Y(1) = f_Y(1, X, \epsilon_Y)$



$Y(0)$ = what the temperature in a location would be if there were *no* air pollution in this location

- Under a NPSEM

$$X = \epsilon_X$$
$$A = f_A(X, \epsilon_A)$$
$$Y = f_Y(A, X, \epsilon_Y)$$

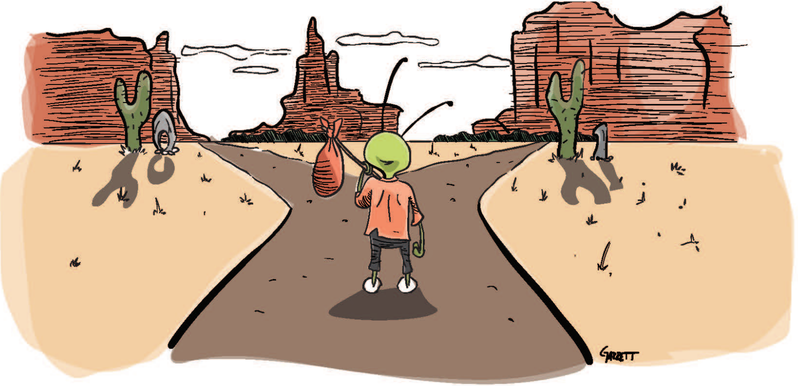
We have $Y(0) = f_Y(0, X, \epsilon_Y)$

Potential Outcome: Causal Contrast

$$
\begin{aligned}
CE &= Y(1) - Y(0) \\
&= f_Y(1, X, \epsilon_Y) - f_Y(0, X, \epsilon_Y) \quad \text{under the NPSEM} \\
&= \beta_1 \quad \text{under the linear SEM}
\end{aligned}
$$

- Does not depends on any parametric assumption
- Does not require knowledge of the full causal system

Relating Potential Outcomes to Observed Outcomes



The consistency assumption: $Y = AY(1) + (1 - A)Y(0) = Y(A)$

Potential Outcomes (Oracle's Table)

For every row, only see one outcome Y !

$Y_i(1)$	$Y_i(0)$	$Y_i(1) - Y_i(0)$	A_i
1.1	2.3	-1.2	1
1.8	0.3	1.5	0
2.0	2.1	-0.1	0
0.1	1.3	-1.2	1

Observed Outcomes (via Consistency)

For every row, only see one outcome Y !

$Y_i(1)$	$Y_i(0)$	$Y_i(1) - Y_i(0)$	A_i
1.1	?	-1.2	1
1.8	0.3	1.5	0
2.0	2.1	-0.1	0
0.1	1.3	-1.2	1

Observed Outcomes (via Consistency)

For every row, only see one outcome Y !

$Y_i(1)$	$Y_i(0)$	$Y_i(1) - Y_i(0)$	A_i
1.1	?	?	1
1.8	0.3	1.5	0
2.0	2.1	-0.1	0
0.1	1.3	-1.2	1

Observed Outcomes (via Consistency)

For every row, only see one outcome Y !

$Y_i(1)$	$Y_i(0)$	$Y_i(1) - Y_i(0)$	A_i
1.1	?	?	1
?	0.3	1.5	0
2.0	2.1	-0.1	0
0.1	1.3	-1.2	1

Observed Outcomes (via Consistency)

For every row, only see one outcome Y !

$Y_i(1)$	$Y_i(0)$	$Y_i(1) - Y_i(0)$	A_i
1.1	?	?	1
?	0.3	?	0
2.0	2.1	-0.1	0
0.1	1.3	-1.2	1

Observed Outcomes (via Consistency)

For every row, only see one outcome Y !

$Y_i(1)$	$Y_i(0)$	$Y_i(1) - Y_i(0)$	A_i
1.1	?	?	1
?	0.3	?	0
?	2.1	?	0
0.1	1.3	-1.2	1

Observed Outcomes (via Consistency)

For every row, only see one outcome Y !

$Y_i(1)$	$Y_i(0)$	$Y_i(1) - Y_i(0)$	A_i
1.1	?	?	1
?	0.3	?	0
?	2.1	?	0
0.1	?	?	1

The Fundamental Problem of Causal Inference

For every row, only see one outcome Y !

$Y_i(1)$	$Y_i(0)$	$Y_i(1) - Y_i(0)$	A_i
1.1	?	?	1
?	0.3	?	0
?	2.1	?	0
0.1	?	?	1

Fundamentally the potential outcome framework reduces causal inference to a missing data problem

- A is the missing data indicator

Individual vs Population Causal Effects

- Individual causal effects $Y_i(1) - Y_i(0)$ not identifiable
- Aim for Average Causal Effect (ACE) instead:

$$E[Y(1) - Y(0)]$$

  - Under our NPSEMs, this is $E[f_Y(1, X, \epsilon_Y) - f_Y(0, X, \epsilon_Y)]$

SEMs vs Potential Outcomes

Parametric Structural Equation Models

- ☺ Permits detailed predictions on what would be observed in an experimental setting

Potential Outcomes

- 😐 Often used for studying the effect of a particular cause on a particular outcome

SEMs vs Potential Outcomes

Parametric Structural Equation Models

- ☺ Permits detailed predictions on what would be observed in an experimental setting
- ☹ Definition of causal effects relies on correct specification of the Parametric SEMs (parametric assumption + knowledge of the whole system)

Potential Outcomes

- 😐 Often used for studying the effect of a particular cause on a particular outcome
- ☺ Causal effects defined non-parametrically

SEMs vs Potential Outcomes

Parametric Structural Equation Models

- ☺ Permits detailed predictions on what would be observed in an experimental setting
- ☹ Definition of causal effects relies on correct specification of the Parametric SEMs (parametric assumption + knowledge of the whole system)
- Relate the observational world to the (hypothetical) experimental world via the stability assumption

Potential Outcomes

- 😐 Often used for studying the effect of a particular cause on a particular outcome
- ☺ Causal effects defined non-parametrically
- Relate the observational world to the (hypothetical) experimental world via the consistency assumption

Frameworks in Causal Inference

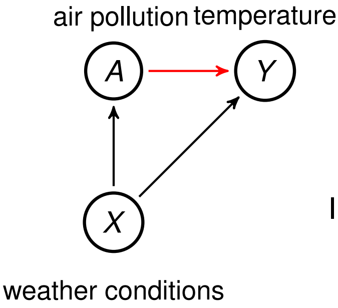
Mechanistic versus agnostic approach to causal inference

Identification and estimation of causal effects

Causal Estimation Under No Unmeasured Confounding

Instrumental Variables

Problem Description



Assume no unmeasured confounding

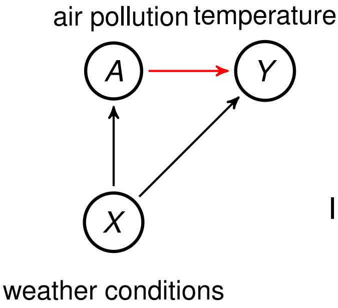
- SEM: no unmeasured variables in the system
- Potential outcome: $A \perp\!\!\!\perp Y(1), Y(0) \mid X$

Interested in estimating

$$E[Y(1) - Y(0)],$$

where $E[Y(a)] = E_X E[Y \mid A = a, X]$

Problem Description



Assume no unmeasured confounding

- SEM: no unmeasured variables in the system
- Potential outcome: $A \perp\!\!\!\perp Y(1), Y(0) \mid X$

Interested in estimating

$$E[Y(1) - Y(0)],$$

where $E[Y(a)] = E_X E[Y \mid A = a, X]$

A Statistical Problem!

Approach 1: Regression Adjustment

$$E[Y(1)] = E_X E[Y \mid A = 1, X]$$

Specify a regression model for $E[Y \mid A = 1, X]$

- Linear regression

$$E[Y \mid A = 1, X] = \beta_0' + \beta_2 X$$

- Non-parametric regression: spline, basis expansion, etc.
- Machine learning: random forest, neural networks, etc.

Approach 2: Inverse Probability Weighting (IPW)

Causal estimation is also a missing data problem

X	$Y(1)$	$Y(0)$	$Y(1) - Y(0)$	A
1	1.1	?	?	1
1	?	0.3	?	0
0	?	2.1	?	0
0	0.1	?	?	1

Approach 2: Inverse Probability Weighting (IPW)

Causal estimation is also a missing data problem

X	$Y(1)$			A
1	1.1			1
1	?			0
0	?			0
0	0.1			1

Approach 2: Inverse Probability Weighting (IPW)

Causal estimation is also a missing data problem

X	$Y(1)$		A
1	1.1		1
1	?		0
0	?		0
0	0.1		1

No unmeasured confounding = Missing at random

☞ Can use inverse probability weighting

$$E[Y(1)] = E\frac{AY}{P(A = 1 \mid X)}$$

- $P(A = 1 \mid X)$ is known as the propensity score

Regression vs Inverse probability weighting

$$f(Y, A, X) = f(Y \mid A, X)f(A \mid X)f(X)$$

Regression vs Inverse probability weighting

$f(Y, A, X) = f(Y \mid A, X) f(A \mid X) f(X)$

- Regression adjustment: model $E[Y \mid A = a, X]$

Regression vs Inverse probability weighting

$$f(Y, A, X) = f(Y \mid A, X) f(A \mid X) f(X)$$

- Regression adjustment: model $E[Y \mid A = a, X]$
- Propensity score: model $P(A = a \mid X)$

Approach 3: Doubly Robust (DR) Approach

$f(Y, A, X) = f(Y \mid A, X) f(A \mid X) f(X)$

- Regression adjustment: model $E[Y \mid A = a, X]$
- Propensity score: model $P(A = a \mid X)$

Doubly robust approach: model both $E[Y \mid A = a, X]$ and $P(A = a \mid X)$

Approach 3: Doubly Robust (DR) Approach

One canonical doubly robust estimator is

$$\hat{E}[Y(1)] = \mathbb{P}_n \left\{ \hat{B} + \frac{A}{\hat{\Pi}}(Y - \hat{B}) \right\}$$

where $\mathbb{P}_n$ = empirical mean, $B = E[Y \mid A = 1, X]$,
$\Pi = P(A = 1 \mid X)$

- **Double robustness:**

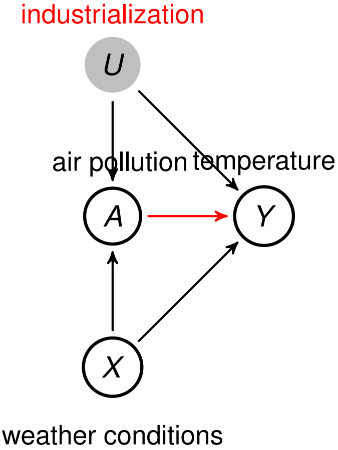
$$Bias(DR) \dot{\sim} Bias(Regression) \times Bias(Prop.\ Score)$$

Estimators for Average Causal Effect

- Regression
- Propensity score subclassification
- Propensity score weighting
- Doubly robust
- Many more...

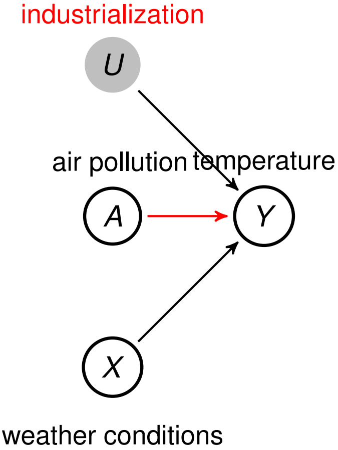
All assuming no unmeasured confounding...

Unmeasured Confounding

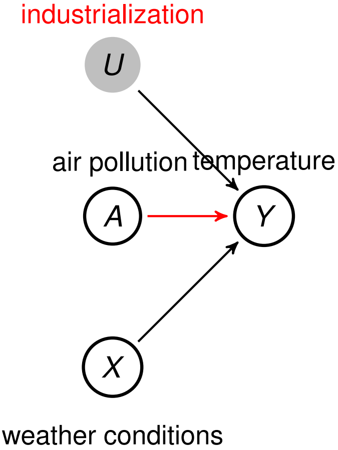


unmeasured
confounding!

Randomized Experiment

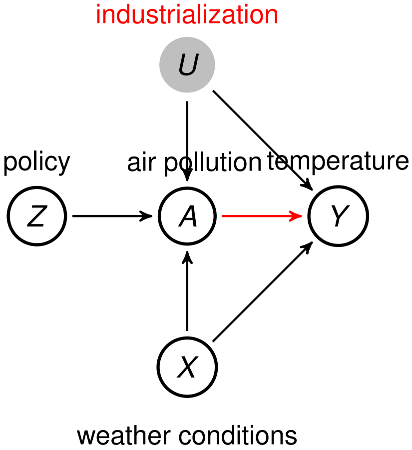


Randomized Experiment

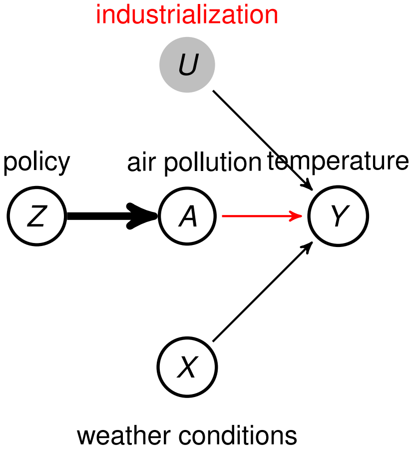


May not be feasible...

Instrumental Variable

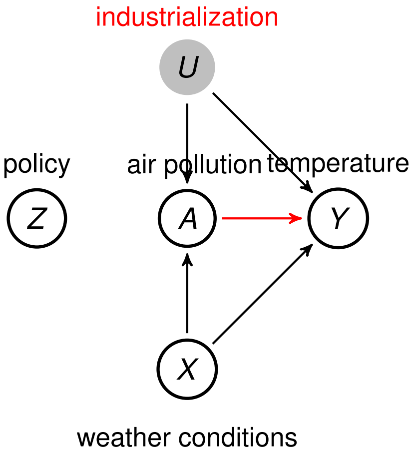


Instrumental Variable



$Z = A$: randomized trial

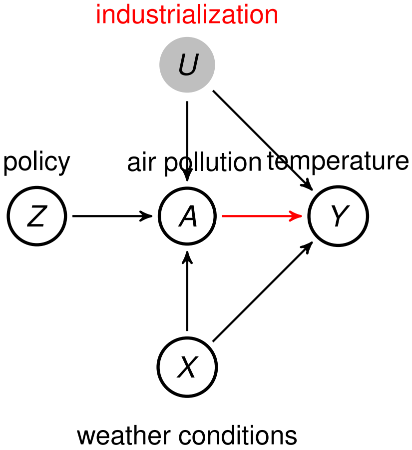
Instrumental Variable



$Z = A$: randomized trial

$Z \perp\!\!\!\perp A$: observational study

Instrumental Variable

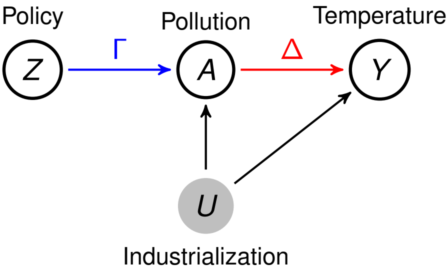


$Z = A$: randomized trial

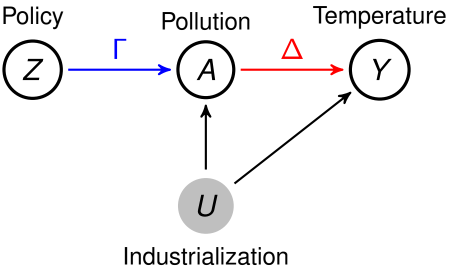
Reality: quasi-experiment

$Z \perp\!\!\!\perp A$: observational study

Identification of Causal Effects



Identification of Causal Effects

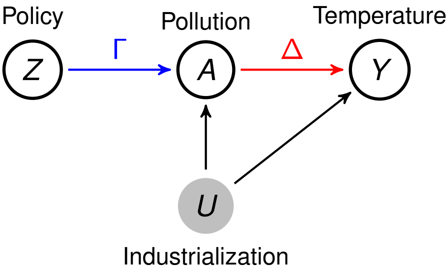


$ACE\,(Z \rightarrow Y)$ $ACE\,(Z \rightarrow A)$

↑ ↑

Observation: identifiable identifiable

Identification of Causal Effects



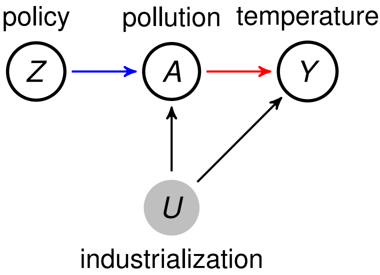
Key Result: Under additional assumptions,

$$ACE(Z \to Y) = ACE(Z \to A) \times ACE(A \to Y).$$

↑ identifiable ↑ identifiable

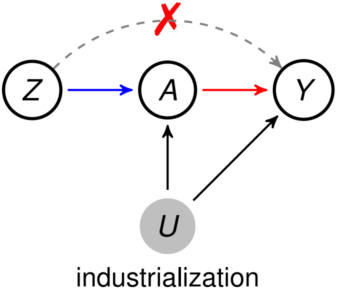
Instrumental Variable Model

Key Formula: $\textbf{ACE(Z} \rightarrow \textbf{Y)} = \textbf{ACE(Z} \rightarrow \textbf{A)} \times \textbf{ACE(A} \rightarrow \textbf{Y)}$



Instrumental Variable Model

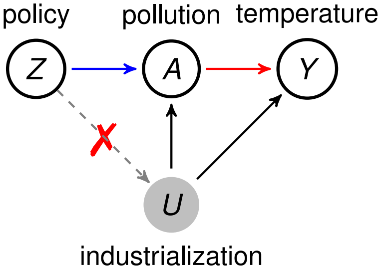
Key Formula: $\textbf{ACE(Z} \rightarrow \textbf{Y)} = \textbf{ACE(Z} \rightarrow \textbf{A)} \times \textbf{ACE(A} \rightarrow \textbf{Y)}$



- No direct effect on Y

Instrumental Variable Model

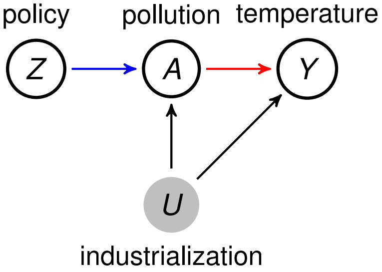
Key Formula: $\mathbf{ACE(Z \to Y) = ACE(Z \to A) \times ACE(A \to Y)}$



- No direct effect on Y
- $Z \perp\!\!\!\perp U$

Instrumental Variable Model

Key Formula: $\mathbf{ACE(Z \rightarrow Y) = ACE(Z \rightarrow A) \times ACE(A \rightarrow Y)}$



- No direct effect on Y
- $Z \perp\!\!\!\perp U$
- $ACE(Z \rightarrow A) \neq 0$

Summary

- Mechanistic (Parametric SEM) vs Agnostic (Potential outcome) approaches
  - Parametric SEM: More intuitive, permits detailed prediction
  - Potential outcome: Fewer assumptions, more robust to model misspecification

- Causal effect estimation
  - (Randomized experiment)
  - Assume no unmeasured confounding
  - Instrumental variable methods
  - Many more...

Thank you!

Contact: Linbo Wang (University of Toronto), linbo.wang@utoronto.ca