Potential Outcomes and Directed Acyclic Graphs

Paul Goldsmith-Pinkham

January 13, 2026

Causality and counterfactuals

- Not every economics research paper is estimating a causal quantity
  - But, the implication or takeaway of papers is (almost) always a causal one

- Causality lies at the heart of every exercise

- Goal for today's class:
  1. Enumerate tools used to discuss causal questions
  2. Emphasize a *multimodal* approach
  3. Set terminology/definitions for future discussions

"We do not have knowledge of a thing until we have grasped its why, that is to say, its cause."

-Aristotle

Causality and counterfactuals - strong opinions

- The true underpinnings of causality are nearly philosphical in nature
  - If Aristotle didn't settle the question, neither will researchers in the 21t century

- I will avoid many of the discussions, but my biases will show up in one or two settings

- Key point: economics research is messy, and a careful discussion of causality entails two dimensions:
  1. A good framework to articulate your assumptions
  2. Readers that understand the framework

The problem of causal inference: a medical example

- Two variables:
  - $Y \in \{0, 1\}$: whether a person will get measles over their lifetime
  - $D \in \{0, 1\}$: whether a person gets a vaccine as a child

- Our question: does D causally affect Y?

- *Ignore the question of data for now* – this is purely a question of what is knowable.

- "The fundamental problem of causal inference" (Holland 1986) is that for a given individual, we can only observe one world – either they get the vaccine, or they do not

The problem of causal inference: a medical example

- What is knowable?
  - We need notation
  - Begin with the Neyman-Rubin Causal model

- There is a population of n individuals, indexed by i.

- Let $Y_i(D_i)$ denote the outcome given a particular vaccine treatment
  - $Y_i(1)$: they receive the vaccine
  - $Y_i(0)$: they do not receive the vaccine

- **What's a key assumption here**?

The problem of causal inference: a medical example

- What is knowable?
  - We need notation
  - Begin with the Neyman-Rubin Causal model

- There is a population of n individuals, indexed by i.

- Let $Y_i(D_i)$ denote the outcome given a particular vaccine treatment
  - $Y_i(1)$: they receive the vaccine
  - $Y_i(0)$: they do not receive the vaccine

- **What's a key assumption here**? person i's outcome is only affected by their own treatment. We will discuss relaxing this assumption later.
  - SUTVA - Stable Unit Treatment Variable Assignment

$$Y_i = D_i Y_i(1) + (1 - D_i) Y_i(0)$$

i	$Y_i(1)$	$Y_i(0)$	D_i	Y_i
1	1	0	1	1
2	0	0	1	0
3	1	0	0	0
		⋮		
n	0	1	0	1

Causal inference is a missing data problem

- In the potential outcomes framework, causal inference and missing data are tightly linked.
- Any causal answer uses assumptions to infer the “missing” counterfactual
- Goal of this course will be to discuss many ways to solve these types of problems
- Before diving into the many potential estimands, consider what the goal is.
  - A structural parameter? E.g. dInvestment/dTax Rate
  - Existence of an treatment effect?
  - A policy evaluation?

A brief aside: estimands, estimators and estimates

- Estimand: the quantity to be estimated
- Estimate: the approximation of the estimand using a finite data sample
- Estimator: the method or formula for arriving at the estimate for an estimand
- My way of remembering: `https://www.youtube.com/watch?v=dEOHfYvCibw`

Causal estimands

- We will start with the Average Treatment Effect:
  - $\tau_{ATE} = \mathbb{E}(\tau_i) = \mathbb{E}(Y_i(1) - Y_i(0)) = \mathbb{E}(Y_i(1)) - \mathbb{E}(Y_i(0))$
  - This expression is defined over the full population, and includes individuals who may never recieve the treatment.

- Average Treatment Effect on the Treated:

$$\begin{aligned}\tau_{ATT} = \mathbb{E}(\tau_i|D_i = 1) &= \mathbb{E}(Y_i(1) - Y_i(0)|D_i = 1) \\ &= \mathbb{E}(Y_i(1)|D_i = 1) - \mathbb{E}(Y_i(0)|D_i = 1)\end{aligned}$$

  - Estimated effect for individuals who *received* the treatment.
  - Note that $\mathbb{E}(Y_i(1)|D_i = 1)$ is observed data!

- Conditional Average Treatment Effect: For a characteristic X_i,

$$\tau_{CATE}(x) = \mathbb{E}(\tau_i|X_i = x) = \mathbb{E}(Y_i(1) - Y_i(0)|X_i = x)$$

A second brief aside: what is identification?

- What does (point) identification mean?

A second brief aside: what is identification?

- What does (point) identification mean?
- Intuitively, for an estimate of interest, τ_{ATE}, to be identified, it means that in a world with no uncertainty about data, can we always identify the value of τ from the data we observe?
  - In other words, it's an invertability condition

"Econometric identification really means just one thing: model parameters or features being uniquely determined from the observable population that generates the data"

-Lewbel (2019)

A second brief aside: what is identification?

- What does (point) identification mean?
- Intuitively, for an estimate of interest, τ_{ATE}, to be identified, it means that in a world with no uncertainty about data, can we always identify the value of τ from the data we observe?
  - In other words, it's an invertability condition

"Econometric identification really means just one thing: model parameters or features being uniquely determined from the observable population that generates the data"

-Lewbel (2019)

- Why would something not be identified if we only observe (Y_i, D_i)?
  - Consider τ_{ATT}. $\mathbb{E}(Y_i(1)|D_i = 1)$ is identified, mechanically. What about $\mathbb{E}(Y_i(0)|D_i = 1)$?
  - One approach: make an assumption on the relationship between D_i and $(Y_i(1), Y_i(0))$
    - This is known as a *design*-based approach (more later)

Under what conditions is the ATE identified?

Strong Ignorability: D_i is *strongly ignorable* conditional on a vector $\mathbf{X}_i$ if

1. $(Y_i(0), Y_i(1)) \perp\!\!\!\perp D_i | \mathbf{X}_i$
2. $\exists \epsilon > 0$ s.t. $\epsilon < \Pr(D_i = 1 | X_i) < 1 - \epsilon_i$

- The first condition asserts independence of the treatment from the "potential" outcomes
- The second condition asserts that there are both treated and untreated individuals
- N.B. The term "strong ignorability" is much more precise than exogeneous
  - But less commonly used in economics.
  - You might instead say "D_i is conditionally randomly assigned."
  - You *might* even say D_i is exogeneous.

When could we not identify the ATE?

- Intuitively, we understand why we typically can't estimate a treatment effect
- Consider an unobservable variable, $U_i \in \{0, 1\}$ where $(Y_i(0), Y_i(1), D_i) \not\perp\!\!\!\perp U_i$
- Simple example: when $E(D_i|U_i = 1) > E(D_i|U_i = 0)$ and $E(\tau_i|U_i = 1) > E(\tau_i|U_i = 0)$.
- In other word, there is a variable that influences both the potential outcomes and the choice of treatment.
  - In this case, estimating the counterfactual is contaminated by the variable U_i
- Many of the goals in this class will be to address this

Theorem: Identification of the ATE

Theorem: If D_i is strongly ignorable conditional on $\mathbf{X}_i$, then

$$\mathbb{E}(\tau_i) = \sum_{x \in \text{Supp } X_i} (\mathbb{E}(Y_i|D_i = 1, \mathbf{X}_i = x) - \mathbb{E}(Y_i|D_i = 0, \mathbf{X}_i = x))Pr(\mathbf{X}_i = x)$$

Proof: Note that $\mathbb{E}(Y_i(0)|\mathbf{X}_i) = \mathbb{E}(Y_i(0)|D_i = 0, \mathbf{X}_i) = \mathbb{E}(Y_i|D_i = 0, \mathbf{X}_i)$ by strong ignorability. In essence, independence of D_i and $(Y_i(0), Y_i(1))$ lets us interchange counterfactuals and realized data in conditionals. The rest follows by the law of iterated expectations. □

- Key implication – counterfactual can be generated by using the averages.

Identification of the ATE - Intuition

i	$Y_i(1)$	$Y_i(0)$	D_i	Y_i
1	1	-	1	1
2	0	-	1	0
3	1	-	1	1
4	1	-	1	1
5	-	0	0	0
6	-	0	0	0
7	-	0	0	0
8	-	1	0	1

- We can estimate $\mathbb{E}(Y_i|D_i = 1) = 0.75$ and $\mathbb{E}(Y_i|D_i = 0) = 0.25$.
- We are defining our counterfactual in the missing data as 0.25, or 0.75, respectively.
- If we had covariates, we would condition within those groups.
- Note that this is all *non-parametric* identification – we have made no model restriction on the data-generating process

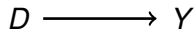
Identification through Directed Acyclic Graphs (DAGs)

- Above, we encoded random variables' relationships functionally, using potential outcomes

- An alternative approach does this graphically (with similar modeling under the hood – to be continued...)

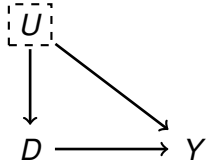
- We can encode the relationship between D and Y using an *arrow* in a graph. The direction emphasizes that D causes Y, and not vice versa.

- Substantially more *intuitive*



Identification through Directed Acyclic Graphs (DAGs)

- We can also allow for the unobservable U, which drove the identification concerns above



- In this case, U is termed a *confounder*. Why?

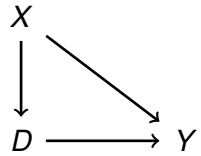
- Examine the paths by which D links to Y:
  - The standard direct effect $D \rightarrow Y$
  - The "Back-Door" path $D \leftarrow U \rightarrow Y$

- Note that the back-door is *not* causal

- Key point: effect of D on Y is not identified under this setup

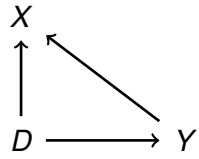
Identification through Directed Acyclic Graphs (DAGs)

- We replace U with an observable X identification concerns above
- X is still a confounder, but we could condition on it and identify our effect. Why?
- Examine the paths by which D links to Y:
  - The standard direct effect $D \rightarrow Y$
  - The "Back-Door" path $D \leftarrow X \rightarrow Y$
- Now, conditioning on a variable along the path "blocks" the path
  - E.g. D is independent of Y *conditional* on X (strong ignorability)



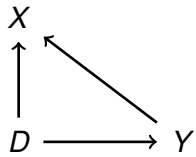
Identification through Directed Acyclic Graphs (DAGs)

- One more example before formalizing the goal
  - X is now a "collider" (note direction of arrows)
- Examine the paths by which D links to Y:
  - The standard direct effect $D \rightarrow Y$
  - The path $D \rightarrow X \leftarrow Y$
- Key difference: a collider is automatically blocked (if it or upstream variables are not conditioned on)
  - If you condition on X, you open the path!
  - Example: conditioning on an outcome variable



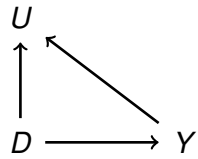
Identification through Directed Acyclic Graphs (DAGs)

- The graphs looked similar, but the order of true causal path mattered



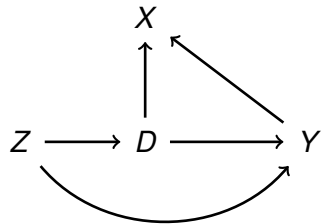
Identification through Directed Acyclic Graphs (DAGs)

- The graphs looked similar, but the order of true causal path mattered
- Identifying colliders is a crucial aspect of identifying whether an effect is identified



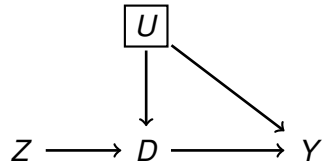
Identification through Directed Acyclic Graphs (DAGs)

- The graphs looked similar, but the order of true causal path mattered
- Identifying colliders is a crucial aspect of identifying whether an effect is identified
- Key value in a DAG (to me) is laying out a model of causality, and clarifying what effects need to be restricted, even in a complicated setting
  - For example, how is the effect of D on Y identified here?



Identification through Directed Acyclic Graphs (DAGs)

- The graphs looked similar, but the order of true causal path mattered
- Identifying colliders is a crucial aspect of identifying whether an effect is identified
- Key value in a DAG (to me) is laying out a model of causality, and clarifying what effects need to be restricted, even in a complicated setting
  - For example, how is the effect of D on Y identified here?
- What about now?



Key steps with a DAG

- Steps when using a DAG
  1. Write down the DAG, and identify what effect you want
  2. Write all paths between the two nodes
  3. What are the “causal” paths (e.g. the arrows all flow in the right direction)?
  4. How many backdoor paths are there? Are they blocked? Can they be?

- Crucial point: conditioning on colliders will cause more harm than good

- We will revisit this setup for some empirical settings
  - Let me know if you think there are good use cases!

Why PO or DAGs? Imbens (2020)

Arguments in favor of DAGs

1. Pedagogically very intuitive
2. Can be very systematic (easy to check identification).
    - Especially true with many covariates

Arguments in favor of PO

1. Some assumptions are more naturally encoded in PO (montonocity, shape)
2. Simultaneous equations and PO naturally align (see next slides) on supply and demand
3. Econ settings focus on just a few variables
4. PO works well with treatment effect heterogeneity
5. Ties directly to questions of survey and design of experiments

Structural equations and causal effects (Haile 2020)

- **Important**: do not lose sight of the fact that these should be estimates that inform our economic model

- (Haile 2020) The reduced form equation is one where the inputs are i) *exogeneous* (ed note: we have not defined this) and ii) unobservable ("structural errors") and the outputs are endogeneous variables. [E.g. $Y_i = f(D_i, X_i, \epsilon_i)$]

- The PO framework's key insight was considering the sets of counterfactuals for each individual. However, it is not magic; insights can typically map across different notations (DAGs, PO, structural econometric equations). Note that these are effectively equivalent:

$$Y_i = D_i Y_i(1) + (1 - D_i) Y_i(0)$$
$$Y_i = \alpha + D_i(\tau + v_i) + u_i$$

Concrete example: demand and supply

- Consider a demand and supply model: $P(Q)$ and $Q(P)$:

$$P = \alpha_0 + \alpha_1 Q + \alpha_2 W + \epsilon \quad (1)$$

$$Q = \beta_0 + \beta_1 P + \beta_2 V + \xi \quad (2)$$

- These are the “structural” equations
- The reduced form comes from plugging in the endogeneous variables and solving for only “exogeneous” variables on the RHS
- This will let us consider counterfactuals in the structural equations!