An Introduction to Causal Inference

Cecilia Cavero Sanchez
Ph.D. student

Department of Political Science
Pennsylvania State University

August 1, 2024

Credibility Revolution: Model v. Design

- **Model-based:** control variables, functional forms, and parametric assumptions
- **Design-based:** identification strategy and OLS
  - Studying the effect of an intervention on an outcome
  - **T:** treatment
  - **Y:** outcome
  - **Y(1), Y(0):** potential outcomes
  - **X:** covariates
  - T_i determines which potential outcome is observed for subject i

The Fundamental Problem of Causal Inference I

- The best we can do to draw causal inferences is to make comparisons between similar groups
- **Example:** UN peacekeeping missions
  1. The first pieces of evidence associated their presence with higher levels of violence against civilians
  2. Later, better designed studies show that peacekeepers are sent to the most violent places, but that their presence actually reduces violence against civilians
- For comparisons to be valid, outcomes in comparison units have to look like what outcomes would have looked like in treatment units

The Fundamental Problem of Causal Inference II

- For each unit we assume that there are two post-intervention outcomes: $Y_i(1)$ and $Y_i(0)$
- $Y_i(1)$ is the outcome that we would obtain if the unit received the treatment
- The causal effect of treatment relative to control is:

$$t_i = Y_i(1) - Y_i(0)$$

- The fundamental problem of causal inference is that we only observe one of the two potential outcomes (Holland 1986)

Potential Outcomes Framework

- We can never observe both $Y_i(1)$ and $Y_i(0)$:

$$t_i = Y_i(1) - Y_i(0)$$

- This is what the ATE would look for unit i if we could observe both scenarios
- We still have the fundamental problem of causal inference
- Randomizing the treatment is the answer

$$ATE = E(Y_i(1)) - E(Y_i(0))$$
$$ATE = E(Y_i(1)|T_i = 1) - E(Y_i(0)|T_i = 0)$$

- Thanks to randomization, the two equations above are the same
- Every unit has two potential outcomes, and receiving treatment/control reveals the outcome

Randomization and independence

- For randomized experiments, the treatment indicator variable is forced by design to be independent of the potential outcome
- Independence is saying that all the facts about you (potential outcomes and covariates) do not affect your probability of treatment
- This means that the distributions of the potential outcomes tend to be the same for treated and control groups
- Thanks to the randomization of the treatment, we can use the observed outcome for the treatment and control group

$$ATE = \frac{1}{N} \sum_{i=1}^{n} Y_i(1) - Y_i(0)$$

Observational Data

- **Covariate balance:** both groups (treatment and control) should be similar (in expectation)
- In observational data, stronger assumptions are usually required to estimate causal effects
- Since the counterfactual units, $E(Y_i(0)|T_i = 1)$, are not observed, a control group must be constructed
- When we fulfill the assumption of (conditional) independence, we can say that treatment assignment is strongly ignorable

Causal Inference Methods

- The main causal inference designs are:

  1. Randomized experiments
  2. Survey and natural experiments
  3. Instrumental variables
  4. Regression discontinuity design
  5. Difference-in-differences
  6. Matching

- Today we are going to (briefly) (hopefully) go over instrumental variables, regression discontinuity design, and difference-in-differences

Instrumental variables

- Helpful to address endogeneity problems → the impact of economic conditions on civil conflict (Miguel 2004)

$$Y = \beta_0 + \beta_1 T$$

- No! Treatment is endogenous
- We need to find an instrument (a variable) that can only affect the outcome through the treatment: rainfall (Miguel 2004)

$$T = \beta_0 + \beta_1 Z$$

- Use predicted values of treatment based on instrument:

$$T = \hat{T}$$

$$Y = \beta_0 + \beta_1 \hat{T}$$

- Independence, the exclusion restriction, and monotonicity

Regression Discontinuity Design (I)

- We exploit intentional rules to create so-called "natural experiments" with the creation of a cutoff
- Treatment status is being above the cutoff
- Mostly we use it in the context of close electoral races $\rightarrow$ impact of drug enforcement policy on drug-related homicides (Dell 2018)

$$Y_{it} = \alpha + \beta_1 D_{it} + \beta_2 M_{it} + \beta_3 D_{it} \times M_{it} + \epsilon_{it}$$

- **D:** treatment (winning a close race $= 1$)
- **M:** running variable (margin of victory)
- Local effects that cannot be extrapolated to the population
- No sorting and continuity

Regression Discontinuity Design (II)

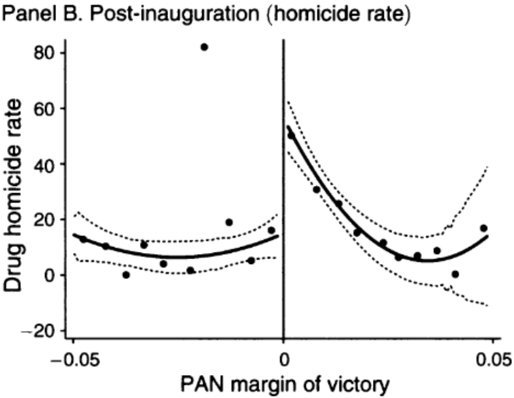

Figure 1: Dell (2018)

Difference-in-Differences (I)

Location	Cost	Per capita quantity
Chicago	$75	2
Indianapolis	$50	1
Milwaukee	$60	1.5
Madison	$55	0.8

Figure 2: Going to the movies

- Cost and quantity are positively correlated. Problems?
- We cannot say much about the role of price on going to the movies using this cross-section data
- With longitudinal data, we find ways to address hidden biases

Difference-in-Differences (II)

Location	Year	Cost	Per capity quantity
Chicago	2003	75	2
Chicago	2004	85	1.8
Indianapolis	2003	50	1
Indianapolis	2004	48	1.1
Milwaukee	2003	60	1.5
Milwaukee	2004	65	1.4
Madison	2003	55	0.8
Madison	2004	60	0.7

Figure 3: Longitudinal data helps!

- Treatment assignment is commonly correlated with group characteristics: we want to estimate its effect without being confounded
- A DiD design relies on the assumption that unmeasured covariates are either unit-specific but time-invariant or vice versa

Difference-in-Differences (III)

- These restrictions imply that the outcomes in each group should (i) differ by the same amount in every period and (ii) exhibit a common set of changes across periods
- Any divergence from these trends is due to the treatment
- By taking the difference between the treatment and control groups' outcomes before treatment (γ) and the difference between their outcomes after treatment ($\gamma + \delta$), we can obtain the final treatment effect (δ), which is the additional change in outcome for the treatment group compared to the control group after the treatment is implemented

Difference-in-Differences (IV)

	Before	After	After - Before
Control	α	$\alpha + \lambda$	
Treatment	$\alpha + \gamma$	$\alpha + \gamma + \lambda + \delta$	
Treatment - Control	γ	$\gamma + \delta$	δ

$$ATT = \underbrace{(E[Y_{i2}|D_i = 1] - E[Y_{i1}|D_i = 1])}_{\text{Change for treated}} - \underbrace{(E[Y_{i2}|D_i = 0] - E[Y_{i1}|D_i = 0])}_{\text{Change for control}}$$

- **D:** treatment status
- Parallel trends and no anticipation
- This is a very basic introduction! There's a lot more going on...

Dynamic Difference-in-Differences (I)

- When you have a lot of time periods, we might get biased results when the treatment is not homogenous across time (Goodman-Bacon 2021)
- Many solutions are being proposed, including the Callaway and Sant'anna estimator
- Deliver disaggregated group-time average treatment effects, treatment effects parameters corresponding to different lengths of exposure to the treatment, and overall treatment effect estimates
- **Groups:** based on when they were first exposed to the treatment
- Always treated are removed

Dynamic Difference-in-Differences (II)

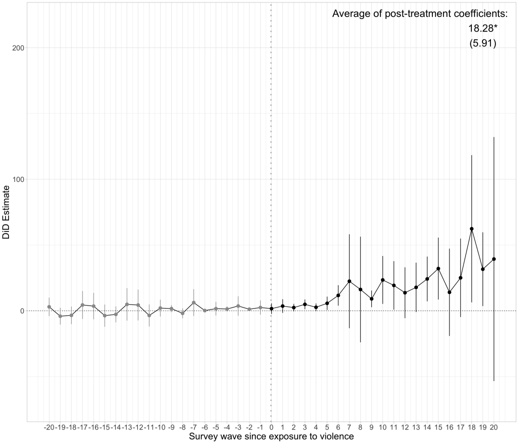

Figure 4: Average effect of exposure to violent actors on votes by length of exposure. N = 15,741 (municipality-year), working paper with Giancarlo Visconti

Importance of Ordinary Least Squares (OLS) for Causal Inference

- OLS is important in causal inference because it provides unbiased estimates of the causal effects under certain assumptions (independence!)
- As we know, the OLS estimator is given by $\hat{\beta}_{OLS} = (X'X)^{-1}X'Y$
- After coming up with a design (previous slides), we use OLS to estimate our treatment effects
- OLS is B.L.U.E.