Gov 2002: 1. Intro & Potential Outcomes

Matthew Blackwell

September 3, 2015

Welcome!

- Me: Matthew Blackwell, Assistant Professor in the Government Department
- What I study: causal inference, missing data, American politics, slavery, and so on.
- Your TF: Stephen Pettigrew, PhD Candidate in Gov.
- What he studies: Bayesian statistics, machine learnings, American politics, sports analytics.

Goals

1. Be able to understand and use recent advances in causal inference
2. Be able to diagnose problems and understand assumptions of causal inference
3. Be able to understand almost all causal inference in applied political science
4. Provide you with enough understanding to learn more on your own
5. Get you as excited about methods as we are

Prereqs

- Biggest: clear eyes, full hearts aka willingness to work hard.
- Working assumption is that you have taken Gov 2000 and 2001 or the equivalent.
- Basically, you vaguely still understand what this is:

$$(X'X)^{-1}X'y$$

- And these terms are familiar to you:
  - bias
  - consistency
  - null hypothesis
  - homoskedastic
  - parametric model
  - σ-algebras (just kidding)

R for computing

- It's free
- It's becoming the de facto standard in many applied statistical fields
- It's extremely powerful, but relatively simple to do basic stats
- Compared to other options (Stata, SPSS, etc) you'll be more free to implement what you need (as opposed to what Stata thinks is best)
- Will use it in lectures, much more help with it in sections

Teaching resources

- Lecture (where we will cover the broad topics)
- Sections (where you will get more specific, targeted help on assignments)
- Canvas site (where you'll find the syllabus, assignments, and where you can ask questions and discuss topics with us and your classmates)
- Office hours (where you can ask even more questions)

Textbook

- Angrist and Pischke, Mostly Harmless Econometrics:
  - Chatty, opinionated, but intuitive approach to causal inference
  - Very much from an econ perspective
- Hernan and Robins, Causal Inference.
  - Clear and basic introduction to foundational concepts
  - From a biostatistics/epidemiology perspective
  - Relies more on graphical approaches
- Other required readings are posted on the website.
- Lecture notes will be other main text.

Grading

1. biweekly homeworks (50%)
2. final project (40%)
3. participation/presentation (10%)

Final project

- Roughly 5-15 page research paper that either:
  - applies some methods of the course to an empirical problem, or
  - develops or expands a methodological approach.
- Co-authorship is encouraged, but comes with higher expectations.
- Fine to combine with another class paper.
- Focus on research design, data, methodology, and results.
- Milestones throughout the term, presentation on 12/10.

Broad outline

1. Primitives
   - Potential outcomes, confounding, DAGs
2. Experimental studies
   - Randomization, identification, estimation
3. Observational studies with no confounding
   - Regression, weighting, matching
4. Observational studies with confounding
   - Panel data, diff-in-diff, IV, RDD
5. Misc. Topics
   - Mechanisms/direct effects, dynamic causal inference, etc

What is causal inference?

- Causal inference is the study of counterfactuals:
  - what would happened if we were to change this aspect of the world?
- Social science theories are almost always causal in their nature.
  - H1: an increase in X causes Y to increase
- Knowing causal inference will help us:
  1. understand when we can answer these questions, and
  2. design better studies to provide answers

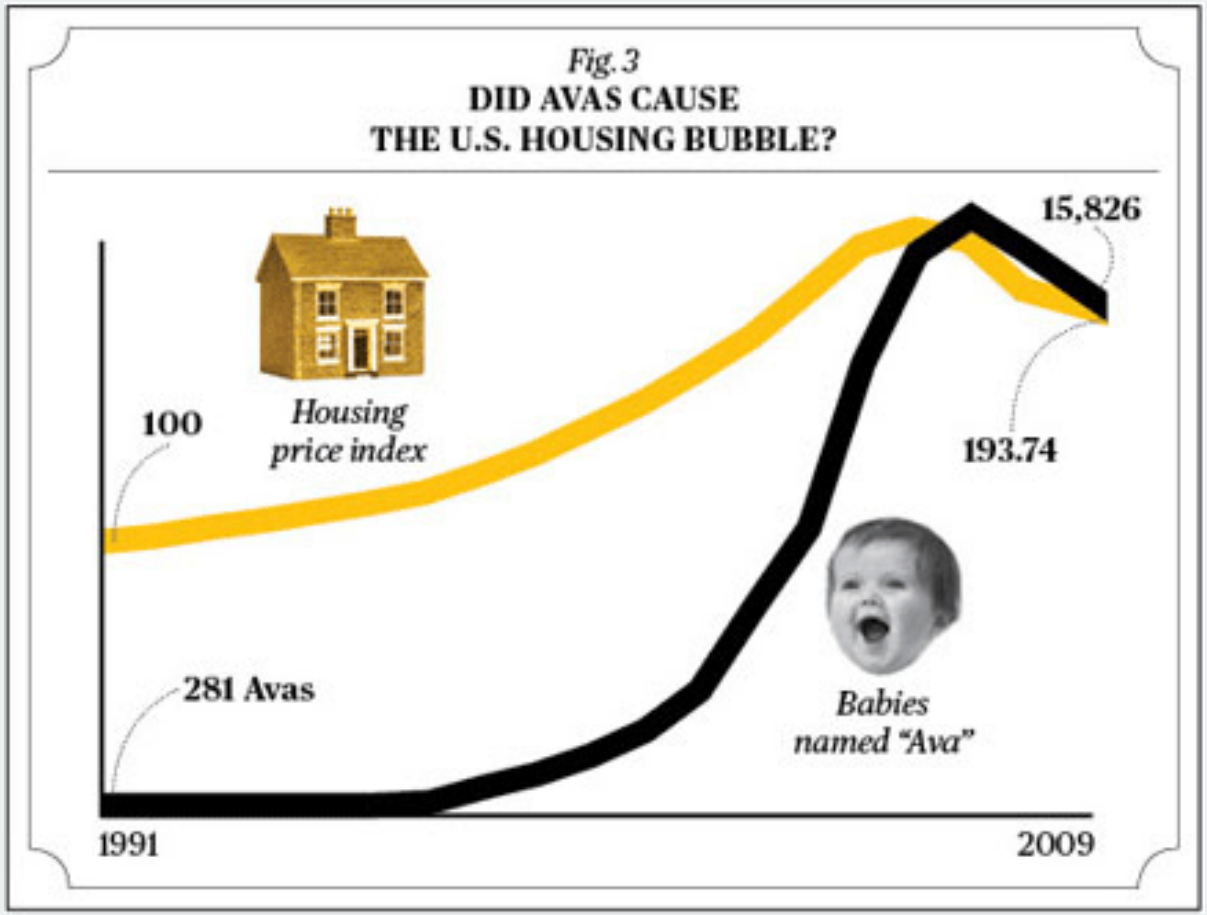
What is identification?

- Identification of a quantity of interest (mean, effect, etc) tells us what we can learn about that quantity from the type of data available.
- Would we know this quantity if we had access to unlimited data?
  - No worrying about estimation uncertainty here.
  - Standard errors on estimates are all 0.
- A quantity is identified if, with infinite data, it can only take on a single value.
- Statistical identification: not possible to estimate some coefficients in a linear model.
  - Dummy for incumbent candidate, $X_i = 1$ and dummy for challenger candidate, $Z_i = 1$.
  - Can't estimate the coefficient on both in the same model, no matter the sample size.

Causal identification

- Causal identification tells us what we can learn about a causal effect from the available data.
- Identification depends on assumptions, not on estimation strategies.
- If an effect is not identified, no estimation method will recover it.
- "What's your identification strategy?" = what are the assumptions that allow you to claim you've estimated a causal effect?
- Estimation method (regression, matching, weighting, 2SLS, 3SLS, SEM, GMM, GEE, dynamic panel, etc) are secondary to the identification assumptions.

Lack of identification, example



- High positive correlation.
- But without assumptions, we learn nothing about the causal effect.

Notation

- Population of units
  - Finite population: $U = \{1, 2, \dots, N\}$
  - Infinite (super)population: $U = \{1, 2, \dots, \infty\}$
- Observed outcomes: Y_i
- Binary treatment: $D_i = 1$ if treated, $D_i = 0$ if untreated (control)
- Pretreatment covariates: X_i, could be a matrix

What is association?

- Running example: effect of incumbent candidate negativity on the incumbent's share of the two party vote as the outcome.
- If Y_i and D_i are independent written $Y \perp\!\!\!\perp D$:

$$\Pr[Y = 1|D = 1] = \Pr[Y = 1|D = 0]$$

- If the variables are not independent, we say they are dependent or associated:

$$\Pr[Y = 1|D = 1] \neq \Pr[Y = 1|D = 0]$$

- Association: the distribution of the observed outcome depends on the value of the other variable.
- Nothing about counterfactuals or causality!

Potential outcomes

- We need someway to formally discuss counterfactuals. The Neyman-Rubin causal model of potential outcomes fills this role.
- $Y_i(d)$ is the value that the outcome would take if D_i were set to d.
  - $Y_i(1)$ is value that Y would take if the incumbent went negative.
  - $Y_i(0)$ is the outcome if the incumbent stays positive.
- Potential outcomes are fixed features of the units.
- Fundamental problem of causal inference: can only observe one potential outcome per unit.
- Easy to generalize when D_i is not binary.

Manipulation

- $Y_i(d)$ is the value that Y would take under D_i set to d.
- To be well-defined, D_i should be manipulable at least in principle.
- Leads to common motto: "No causation without manipulation" Holland (1986)
- Tricky causal problems:
  - Effect of race, sex, etc.

Consistency/SUTVA

- How do potential outcomes relate to observed outcomes?
- Need an assumption to make connection:
  - "Consistency" in epidemiology
  - "Stable unit treatment value assumption" (SUTVA) in econ and stats.
- Observed outcome is the potential outcome of the observed treatment:

$$Y_i(d) = Y_i \text{ if } D_i = d$$

- Also write this as:

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

- Two key points here:
  1. No interference between units: $Y_i(d_1, d_2, \ldots, d_N) = Y_i(d_i)$
  2. Variation in the treatment is irrelevant.

Causal inference = missing data

Negativity (Treatment) D_i	Observed Outomes Y_i	Potential Outcomes $Y_i(0)$	$Y_i(1)$
0	.63	.63	?
0	.52	.52	?
0	.55	.55	?
0	.47	.47	?
1	.49	?	.49
1	.51	?	.51
1	.43	?	.43
1	.52	?	.52

Estimands

- What are we trying to estimate? Differences between counterfactual worlds!
- Individual causal effect (ICE):

$$\tau_i = Y_i(1) - Y_i(0)$$

  - Difference between what would happen to me under treatment vs. control.
  - Within unit! $\rightsquigarrow$ FPOCI
  - Almost always unidentified without strong assumptions
- Average treatment effect (ATE):

$$\tau = \mathbb{E}[\tau_i] = \frac{1}{N} \sum_{i=1}^{N} [Y_i(1) - Y_i(0)]$$

  - Average of ICEs over the population.
  - We'll spend a lot time trying to identify this.

Other estimands

- Conditional average treatment effect (CATE) for a subpopulation:

$$\tau(x) = \mathbb{E}[\tau_i | X_i = x] = \frac{1}{N_x} \sum_{i:X_i=x} [Y_i(1) - Y_i(0)],$$

  - where N_x is the number of units in the subpopulation.
- Average treatment effect on the treated (ATT):

$$\tau_{ATT} = \mathbb{E}[\tau_i | D_i = 1] = \frac{1}{N_t} \sum_{i:D_i=1} [Y_i(1) - Y_i(0)],$$

where $N_t = \sum_i D_i$.

Samples versus Populations

- Estimands above all at the population level.
- Sometimes easier to make inferences about the sample actually observed.
- Sample $S \subset U$ of size $n < N$, with n_t treated and $n_c = n - n_t$ controls.
- **Sample average treatment effect (SATE)** is the average of ICEs in the sample:

$$SATE = \tau_S = \frac{1}{n} \sum_{i \in S} [Y_i(1) - Y_i(0)]$$

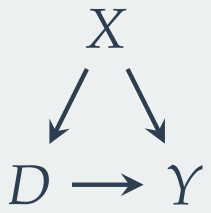
- Limit our inferences to the sample and don't generalize.
- In this context, usually refer to the ATE as the PATE.

Why focus on the sample?

- SATE is the in-sample versions of the PATE.
- SATE varies over samples from the population, whereas the PATE is fixed.
- SATE still unknown because we only observe $Y_i(1)$ or $Y_i(0)$ for unit i
- Estimators for the SATE have lower variance (less useful than it sounds).
- Useful when:
    1. We don't have a random sample from the population $\rightsquigarrow$ extrapolation bias
    2. The sample is the population (countries, states, etc)

Directed Acyclic Graphs

- We can encode assumptions about causal relationships in what are called causal Directed Acyclic Graphs or DAGs. Here is an example:

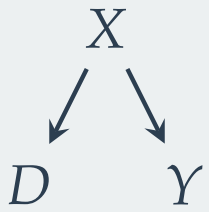


- Each arrow = a direct causal effect: $Y_i(d) \neq Y_i(d')$ for some i and d
- Lack of an arrow = no causal effect: $Y_i(d) = Y_i(d')$ for all i and d
- Directed: each arrow implies a direction
- Acyclic: no cycles: a variable cannot cause itself
- Causal Markov assumption: conditional on its direct causes, a variable V_j is independent of its non-descendents.

Causal DAGs and associations

- Can use DAGs to find potential associations between variables in the graph.
- A path between two variables (C and D) in a DAG is a route that connects the variables following nonintersecting edges.
- A path is causal if those edges all have their arrows pointed in the same direction.
  - Causal: $D \rightarrow X \rightarrow Y$
  - Noncausal: $D \leftarrow X \rightarrow Y$

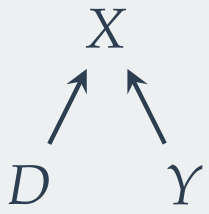
Confounders



- X here is a confounder (or common cause).
- Two variables connected by common causes will have a marginal associational relationship.
- That is, in this example:

$$\Pr[Y = 1|D = 1] \neq \Pr[Y = 1|D = 0]$$

Colliders



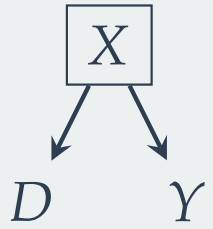
- Here, X is a collider: a node that two arrows point into.
- Are D and Y related? No. Why?
- The flow of association is blocked by a collider so that here:

$$\Pr[Y = 1|D = 1] = \Pr[Y = 1|D = 0]$$

- Example:
  - D is getting the flu and Y is getting hit by a bus.
  - X is being in the hospital
  - Knowing that I have the flu doesn't give me any information about whether or not I've been hit by a bus.

Conditioning on a confounder

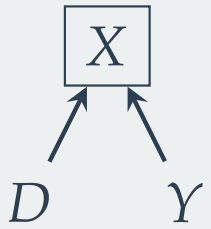
- What happens when we condition on a variable?
- We can represent conditioning on a variable by drawing a box around it.



- Can block the flow of association by:
  1. conditioning on a variable on a causal path, or
  2. conditioning on a confounder (above)

Conditioning on a collider

- Conditioning on a collider (a common consequence) actually opens the flow of association over that path, even though before there was none:



- Back to flu/bus example:
  - Conditional on being in the hospital, there is a negative relationship between the flu and getting hit by a bus.
- We'll talk more about these concepts in the next few weeks.

To sum up

- Causal inference is about comparing counterfactuals.
- Identification is figuring out what we can learn under a set of assumption with unlimited data.
- There are a number of potential causal quantities to identify and estimate.
- DAGs are a useful way to encode assumptions and assess potential associations.
- Next week: identifying causal effects in experiments.