

Causal Inference Overview

Gov 2010 Section

April 1, 2010

Introduction.

Outline – Approaches to Causal Inference:

1. Randomized or Experimental Studies
2. Observational studies
 - ▶ Matching
 - ▶ Regression Discontinuity Design
 - ▶ IVs
3. (if we have time) Example from Research

Fundamental Problem of Causal Inference.

Most people use the “Potential Outcomes” framework developed by Don Rubin. Here’s the terminology:

- ▶ Let $Y_i(0)$ be the outcome of interest under control – i.e., no treatment.
- ▶ Let $Y_i(1)$ be the outcome of interest under treatment.
- ▶ The **causal effect** we are after is $Y_i(1) - Y_i(0)$.
- ▶ And $Y_i(0)$ and $Y_i(1)$ are the **potential outcomes**.

Let’s look at an example.

Fundamental Problem of Causal Inference (ctd).

In an ideal world, we would see this:

$Unit_i$	X_i^1	X_i^2	X_i^3	T_i	$Y_i(0)$	$Y_i(1)$	$Y_i(1) - Y_i(0)$
1	2	1	50	0	69	75	6
2	3	1	98	0	111	108	-3
3	2	2	80	1	92	102	10
4	3	1	98	1	112	111	-1

Fundamental Problem of Causal Inference (ctd).

But in the real world, we see this:

$Unit_i$	X_i^1	X_i^2	X_i^3	T_i	$Y_i(0)$	$Y_i(1)$	$Y_i(1) - Y_i(0)$
1	2	1	50	0	69	?	?
2	3	1	98	0	111	?	?
3	2	2	80	1	?	102	?
4	3	1	98	1	?	111	?

Fundamental Problem of Causal Inference (ctd).

- ▶ The **fundamental problem of causal inference** is that at most only one of the two potential outcomes $Y_i(0)$ or $Y_i(1)$ can be observed for each unit i .
- ▶ For control units, $Y_i(1)$ is the **counterfactual** (i.e., unobserved) potential outcome.
- ▶ For treatment units, $Y_i(0)$ is the **counterfactual**.
- ▶ For this reason, some people (including Don Rubin) call causal inference a *missing data* problem.

Quantities of Interest.

Before we talk about solutions, note that we might be interested in the following quantities of interest

- ▶ The individual treatment effect: $Y_i(1) - Y_i(0)$
- ▶ The average treatment effect (ATE):
$$E[Y(1) - Y(0)] = E[Y(1)] - E[Y(0)]$$
- ▶ The treatment effect on the treated (ATT):
$$E[Y(1|T = 1) - Y(0|T = 1)] = E[Y_t(1) - Y_t(0)]$$

At this point, all of them fail because of the fundamental problem.

Proposed Solutions.

So here are some possible solutions

1. Randomized or Experimental Studies
2. Observational studies
 - ▶ Matching
 - ▶ Regression Discontinuity Design
 - ▶ IVs

Assumptions.

Before we get too far, note that our solutions *must* satisfy two conditions:

1. Ignorability.

- ▶ Ignorability will hold $P(Y_i(1), Y_i(0)|X, T) = P(Y_i(1), Y_i(0)|X_i)$
- ▶ i.e., it holds if $Y_i(1)$ and $Y_i(0)$ are independent of T_i , conditional on X_i .
- ▶ In plain English: the treatment assignment is independent of the potential outcomes.

2. Stable Unit Treatment Variance Assumption (SUTVA).

- ▶ The treatment assignment for one unit should not affect the outcome for another.
- ▶ In other words, the treatment should not “seep” between units.

People generally focus on ignorability, but SUTVA is important too.

Proposed Solutions.

Ok, so, finally, let's go through some proposed solutions:

1. Randomized or Experimental Studies
2. Observational studies
 - ▶ Matching
 - ▶ Regression Discontinuity Design
 - ▶ IVs

Classical Randomized Experiments.

Since we cannot compare treatment and control outcomes for the same units, we compare treated units to controls and vice versa.

- ▶ The simplest way of doing this is through a randomized experiment.
- ▶ Units are randomly assigned to receive treatment and control.
- ▶ We still can't estimate individual-level causal effects, but we can estimate the population **average treatment effect**, $E[Y_i(1) - Y_i(0)]$.
- ▶ Key: To be set up correctly, the treatment assignment must be entirely random (i.e., ignorable).

Classical Randomized Experiments (ctd).

We have again made some strong assumptions:

1. **Ignorability.** There are no factors out there that affect both the probability of treatment and the outcome.
 - ▶ If the treatment was assigned in a truly random fashion, you are generally ok.
 - ▶ But if it wasn't, then you are in trouble!
2. **SUTVA.** We have assumed that assigning treatment to one unit doesn't affect the outcome for another unit.
 - ▶ Not always reasonable if there are potential peer effects!

Classical Randomized Experiments (ctd).

We also need to think about compliance issues.

- ▶ Did the units “take” the treatment as they were supposed to?
 - ▶ Never-taker: Unit never takes treatment
 - ▶ Always-taker: Unit always takes treatment
 - ▶ Complier: Unit takes the treatment like they are supposed to
 - ▶ Defier: Unit takes treatment when not assigned and control when assigned
- ▶ Naive solution: Focus on “intent to treat” rather than on actual treatment.
- ▶ Another solution: Use an IV approach with the intent to treat as an instrument.

What About Observational Data?

- ▶ That's all great, but, for the most part, we can't rely on randomization.
- ▶ We have to rely on observational data.
- ▶ There are particular problems with observational studies
 1. Even though we can still have a “treatment” and a “control,” there can be systematic differences between units receiving the treatment and other units.
 2. These differences can affect the outcome, Y_i .
 3. As a result, making causal inferences becomes really difficult.
 4. Too many people do it without thinking carefully about the assumptions.

Observational Studies (ctd).

Here's how people have thought about this problem:

- ▶ What if we condition on every variable that might affect both T_i and Y_i ?
- ▶ That way, it will appear for all intents and purposes that T_i has been randomly assigned.
- ▶ In plain English: We want to try to mimic a randomized study. We do so by conditioning on all possible confounding variables – those variables that are associated with both the treatment and the outcome.
- ▶ Note: We still have to satisfy SUTVA, the stable unit treatment value assumption.

Proposed Solutions.

So, with this in mind, what are some solutions for observational data?

1. Matching
2. Regression Discontinuity design
3. Instrumental Variables

Matching: A way to Ameliorate Model Dependence

- ▶ If we had pairs of observations that had the exact same covariate values (perfect **balance**) and differed only on treatment assignment, then we would have perfect conditional ignorability – i.e., conditional on the covariates, the treatment is independent of the outcome.
- ▶ Then we will get the same results regardless of the model.
- ▶ Matching is a method of trying to achieve better balance on covariates and reduce model dependence. Goal: Balance on covariates
- ▶ But remember: Don't match on post-treatment variables!

Matching Techniques: Propensity Scores Matching

- ▶ Suppose each observation has some true probability of receiving the treatment. The probability of receiving the treatment is the *propensity score*.
- ▶ We don't know the true propensity score but we can estimate it with a logit regression of T on X (assuming we have the right set of X that affected the treatment mechanism).
- ▶ Then we match an treated unit with control unit with a similar propensity score.
- ▶ Note: Since we don't have the true propensity scores, we need to check for balance on our covariates at the end.

Other Matching Techniques

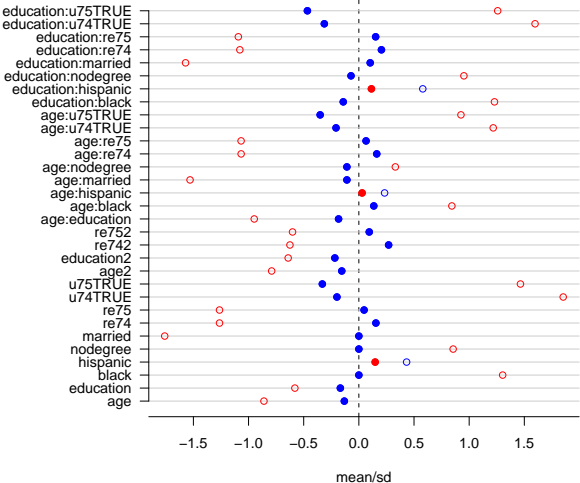
- ▶ **Exact Matching:** Match treated and control units with the exact same covariate profile. You generally need a lot of data.
- ▶ **Mahalanobis Matching:** Employs the Mahalanobis distance measure, which is basically a fancy way of measuring a distance between two vector columns. Nearest neighbor matching usually employs Mahalanobis distance matching.
- ▶ **CEM** (King et al.): Allows you to coarsen covariate values for a more substantive approach. R package: `cem`
- ▶ **Genetic Matching** (Diamond & Sekhon): Uses an evolutionary search algorithm to determine “weights” for each covariate. R package: `GenMatch`

The goal for all of these is to achieve balance.

Matching: Achieving Balance

Check balance with summary statistics. You can also see it visually:

Improved Balance from Caliper Matching



proposed solutions.

Other solutions?

1. Matching
2. **Regression Discontinuity Design**
3. Instrumental Variables

Regression Discontinuity Design

Sometimes called “natural experiments.” Here’s how it works:

- ▶ There is some pre-treatment variable x with a cutoff value – below the cut-off value the unit receives treatment and above the cutoff it receives control.
- ▶ The cutoff should be arbitrary and not related to the other covariates.
- ▶ There should be no real reason to think that the units below the cut-off are substantively that different from the units above the cut-off.
- ▶ We basically take care of ignorability. (Note: you still have to worry about SUTVA.)

Regression Discontinuity Design

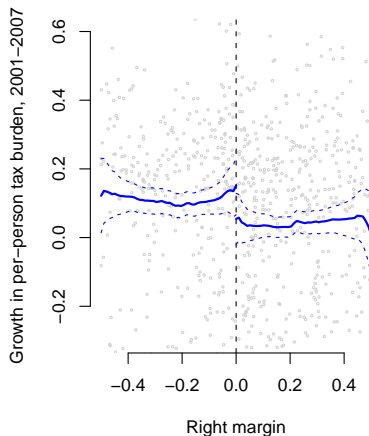
Here are some examples:

- ▶ Criminals who are 17 yrs old, 364 days are treated as juveniles, but those who are 18yrs old, 1 day are treated as adults. What is the effect of adult sentencing?
- ▶ European municipalities with populations of 3499 do not need to have dissenting party members sit on their councils, but those with populations of 3500 do. What is the effect of having dissenting party members on council decisions?
- ▶ HS standardized test takers who get a score of 90.5 get a small scholarship, but those that get 89.5 don't. What is the effect of the scholarship on college success?

It's sometimes easier to see this visually.

Regression Discontinuity Design (ctd.)

Example: In some French cities, the rightist party wins with 50.5% of the vote%; in other cities, they lose with 49.5% of the vote. What is the effect of rightist governments on tax policies?



Instrumental Variables

Most common in economics.

- ▶ The idea: Find an instrument Z that is randomly assigned (or assignment is ignorable) and that affects Y only through T .
- ▶ Useful for getting at compliance problems. Example: Y = post-Vietnam War civilian mortality, T = serving in the military during Vietnam War, Z = draft lottery
- ▶ Angrist, Joshua D., Guido W. Imbens and Donald B. Rubin. 1996. "Identification of Causal Effects Using Instrumental Variables." *Journal of the American Statistical Association* 91(434):444-455.

Appendix: Some of my own work.

I'm wrapping up a project using matching techniques. The questions guiding this research are:

1. Were minorities and women more likely to receive subprime loans during the mid-2000s?
2. Were they treated differently in other respects – for example in terms of the interest rate they got on loans, the loan amount, or the likelihood of rejection?
3. Can we make any causal claims?

The Data.

For *every* mortgage loan application filed in 2006, we have data on:

1. exact location of the property (state, county, census tract);
2. nature of the property (multiple dwelling, lien status);
3. applicant (and co-applicant) income, race, and gender;
4. loan purpose and type;
5. loan amount and interest rate.

In 2006 alone we have access to over 34 million observations from all 50 states plus D.C. and Puerto Rico.

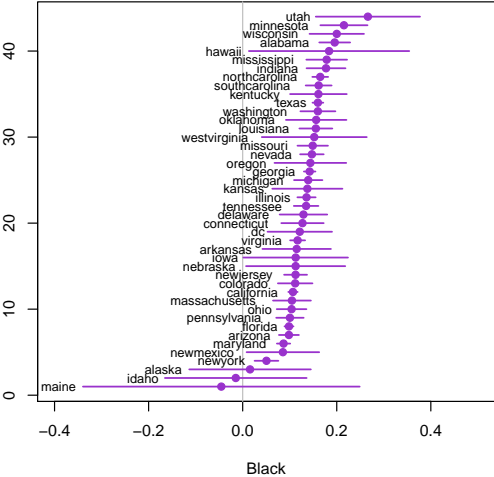
Methodology.

- ▶ Because of the sheer volume of the data, I can use exact matching on all of the pre-treatment covariates. No need to worry about balance.
- ▶ However, the potential outcomes framework operates with neatly defined manipulable treatment variables – like giving someone medicine, or exposing a mouse to radiation.
- ▶ It doesn't work as well with immutable characteristics (race, gender, and ethnicity – because everything is post-treatment).
- ▶ Solution: see developing literature of causal effects involving discriminatory agents (Greiner & Rubin 2010). The focus is not on actual race, but on perceptions of race made by the decision maker (in this case the lender).

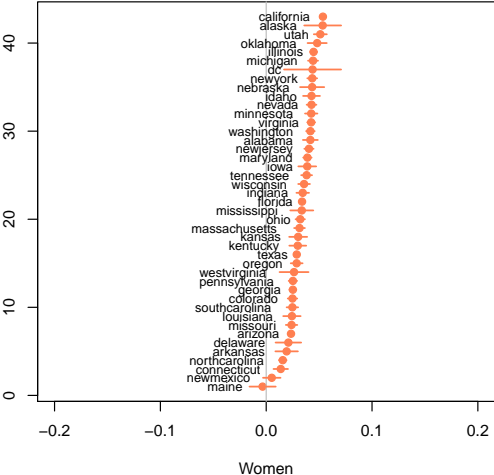
Methodology (ctd).

- ▶ The **treatments** are what the lender wrote down as being the borrower's race or gender: female, black, Hispanic, Asian.
- ▶ The **pre-treatment covariates** involve anything from the time period before an applicant walks in the door of a lending office (any attributes of the property, of the borrower);
- ▶ The **post-treatment covariates** are those that are realized after the person walks into a lending office (loan decision, interest rate, etc.).
- ▶ The **outcome of interest** is whether the person got a subprime (i.e., more expensive) loan versus a standard loan.

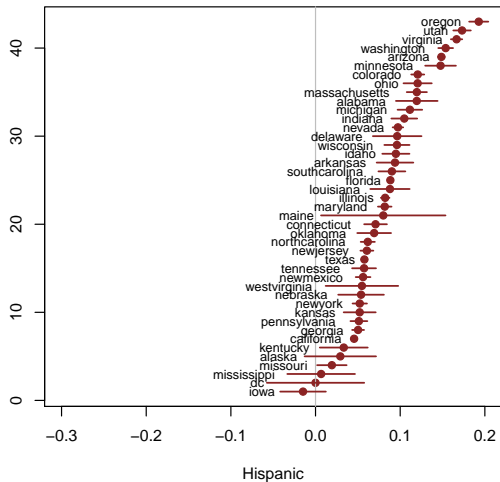
Result: Black borrowers.



Result: Women borrowers.



Result: Hispanic borrowers.



Result: Asian borrowers.

