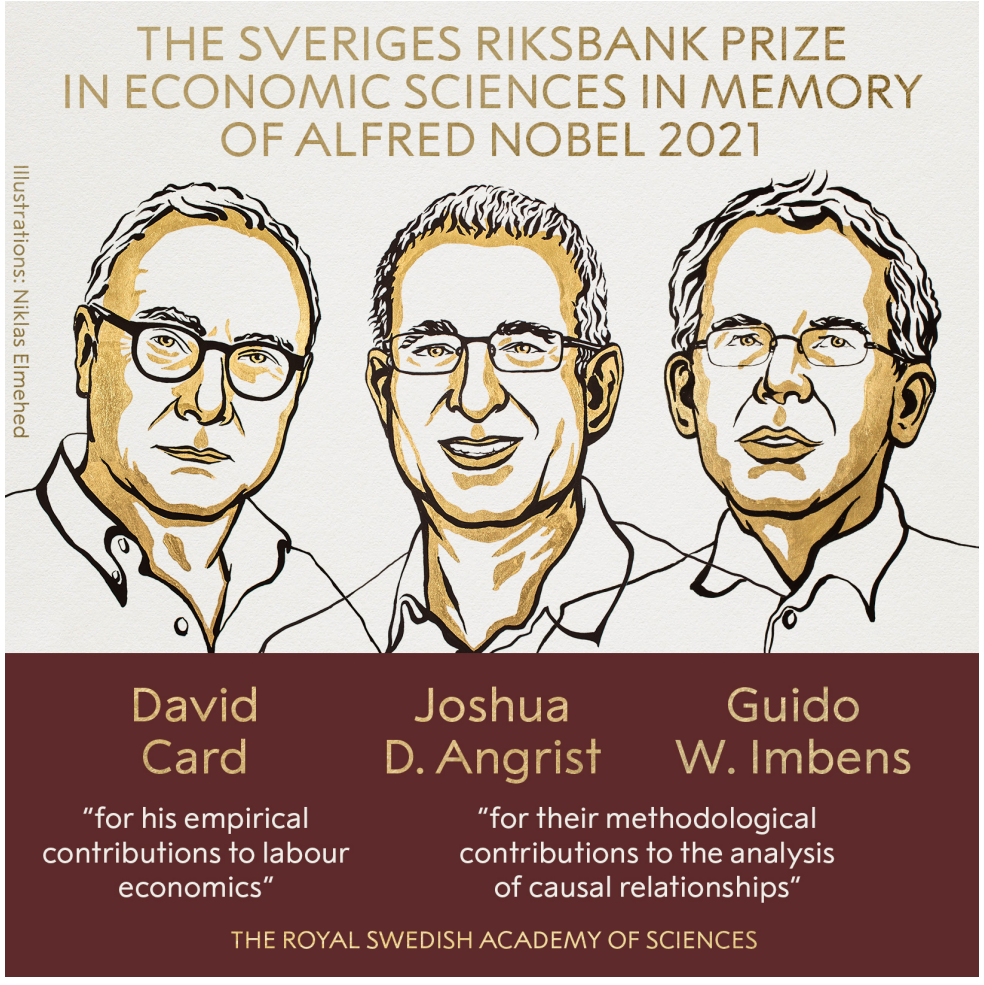
STA 235H - Natural Experiments

Fall 2021

McCombs School of Business, UT Austin

Amazing week for Causal Inference!



Announcements

Homework 3 is due on Thursday

Midterm will be posted that day

- Grades for homework 2 were posted:
  - Check the point assignment in the comments.
  - Any questions about your grade, send it to me.

About the midterm

- Remember: Midterm is an individual assignment
  - There is no collaboration allowed between students.
  - Starting next homework, there's no collaboration allowed between students of different groups either.
- Two review sessions for midterm:
  - Friday 10/15: 11:00am - 12:00pm
  - Monday 10/18: 4:00pm - 5:00pm
  - If attendance is < 10 students, no recording will be uploaded.
- Check out questions in the slides and in the R Code:
  - Feel free to check your answers with the instruction team.

Last class

- Selection on observables
  - Assumption of non-random selection, but selection on observables.
- Matching:
  - Use of other adjustment methods beyond regression.
  - Advantages and disadvantages of matching.

Today

- Natural Experiments:
  - Identifying random assignment* in observational studies
  - Use exogeneous variation to identify causal effects.
- Difference-in-Differences:
  - Using two dimensions for identification.
  - Assumptions and shortcomings.

Is there randomness out there?

Finding "RCTs" in the wild

- Given that we can't run RCTs for everything, the next best thing is finding a source of random variation that, for all practical purposes, would work as an RCT

Natural Experiments

You, as a researcher, did not assign units to treatment levels

1. Random: Assignment to an intervention is random (e.g. lottery).

2. As if random: Assignment to an intervention is not random, but it's not correlated with potential outcomes.

Context matters!

Examples of natural experiments

- Oregon Health experiment: Lotteries for Medicaid expansion
  - Random or "as if" random?
- Admission to charter schools
  - Random or "as if" random?
- GreatSchool ranking availability: Roll-out between states
  - Random or "as if" random?
- Glitch in information availability: Algorithmic tool used to assist child maltreatment hotline screening decisions.
  - Random or "as if" random?

Potential outcomes in Observational Studies

- The same potential outcomes framework that we reviewed for RCTs also work with observational studies.

Steps to identify a Natural Experiment:

1) Identify treatment groups: What is the control status?

2) Identify your estimand of interest: Write it down in terms of PO!

3) Identify potential threats to causality: Is this as good as random?

Let's talk about the JITT example

- A retailer provides a 15% discount to first 1,000 customers, 10% to customers 1,001-2,000 (and no discount after).

Is this a natural experiment?

Mixed answers

Let's think about this more carefully

An example: Timely discounts

- **Two treatments:** 10% discount (1) and 15% discount (2) (control is no discount).
- **Outcome:** Total sales ($)
- **Estimand:** Average Treatment Effects,

$$ATE_1 = E[Y(1) - Y(0)]$$

and

$$ATE_2 = E[Y(2) - Y(0)]$$

How are people assigned to treatment?

An example: Timely discounts

- **Two treatments:** 10% discount (1) and 15% discount (2) (control is no discount).
- **Outcome:** Total sales ($)
- **Estimand:** Average Treatment Effects,

$$ATE_1 = E[Y(1) - Y(0)]$$

and

$$ATE_2 = E[Y(2) - Y(0)]$$

Could there be confounding? Why?

An example: Timely discounts

- **Two treatments:** 10% discount (1) and 15% discount (2) (control is no discount).
- **Outcome:** Total sales ($)
- **Estimand:** Average Treatment Effects,

$$ATE_1 = E[Y(1) - Y(0)]$$

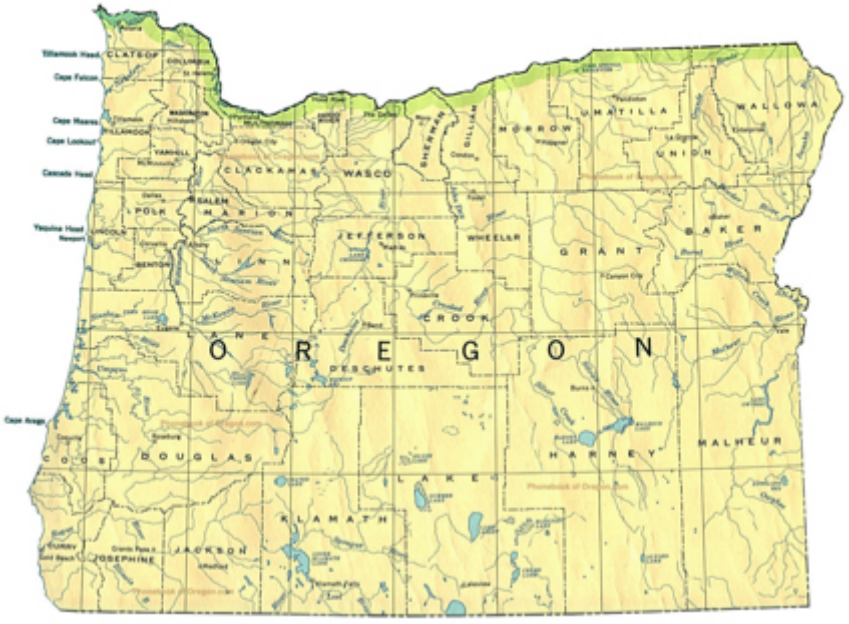
and

$$ATE_2 = E[Y(2) - Y(0)]$$

What if customers didn't know about the discount until they get there?

A true natural experiment: The Oregon Health Plan

- In 2008, Oregon implemented a limited expansion of Medicaid.
- Target population: Low-income adults.
- People selected through a lottery.



The Oregon Health Plan

What is the treatment in this case? What is *randomized?**

What is our estimand of interest?

What about external validity? For what population is this effect generalizable?

What is the first thing you would do with the data?

Things to look out for

- We can find **natural experiments** in many places!
- But it's important to **credibly** convince people they are not correlated with potential outcomes!
    - No confounding or CIA.
- What happens if we have two cities that have an equal probability to be hit by an hurracaine, and you want to analyze the effect of a natural disaster on employment.
    - City A: Large metropolitan city
    - City B: Smaller, urban city

Can you use this setting?

Takeaway points

- We don't always need to "randomize".
- We can exploit natural variation.
    - In natural experiments we are approximating an ideal RCT very straightforwardly.

References

- Angrist, J. and S. Pischke. (2015). "Mastering Metrics". *Chapter 2*.
- NBER (2021). "Oregon Health Insurance Experiment Background". *Summary of the policy*.
- Jensen, R. (2007). "The Digital Provide: Information (Technology), Market Performance, and Welfare in the South Indian Fisheries Sector". *Quarterly Journal of Economics*.