Applied Empirical Economics

Liyousew G. Borga

University of Economics, Prague

January 23, 2020

The Evaluation Problem

“What if ...”: Answering Counterfactual Questions

Empirical methods in economics have been developed to try to answer “counterfactual” questions.

- What would have happened to this person’s behavior if she had been subjected to an alternative treatment?

“What if ...”: Answering Counterfactual Questions

Empirical methods in economics have been developed to try to answer “counterfactual” questions.

- What would have happened to this person’s behavior if she had been subjected to an alternative treatment?

We are concerned about identifying the effect of some action (e.g. a policy, an investment) on one or more outcomes of interest

“What if ...”: Answering Counterfactual Questions

Empirical methods in economics have been developed to try to answer “counterfactual” questions.

- What would have happened to this person’s behavior if she had been subjected to an alternative treatment?

We are concerned about identifying the effect of some action (e.g. a policy, an investment) on one or more outcomes of interest

The goal of the analysis is to “rule out” other possibilities /explanations for the observed effects (internal validity)

- The effect of counseling job search program for the unemployed youth
- The effect of education on wages
- The effect of migration influx on local labor market
- The effect of competition between schools on schooling quality

Learning Goal

- To be able to understand when we are being misinformed (intentionally or otherwise) or lied to.

[1] Source: https://www.bmj.com/content/335/7633/1288.full

Learning Goal

- To be able to understand when we are being misinformed (intentionally or otherwise) or lied to.
- Medical myths[1]
  - People should drink at least eight glasses of water a day
  - We use only 10% of our brains
  - Shaving hair causes it to grow back faster, darker, or coarser
  - Reading in dim light ruins your eyesight
  - 10,000 steps a day to keep healthy and fit

[1] Source: https://www.bmj.com/content/335/7633/1288.full

The Evaluation Problem

What is the impact of eating donuts on your cholesterol level?

The Evaluation Problem

What is the impact of eating donuts on your cholesterol level?

- In an ideal world (research-wise), we could clone each treated individual and observe the impacts of treatment on the outcomes of interest

vs.

The Evaluation Problem

What is the impact of eating donuts on your cholesterol level?

- In an ideal world (research-wise), we could clone each treated individual and observe the impacts of treatment on the outcomes of interest

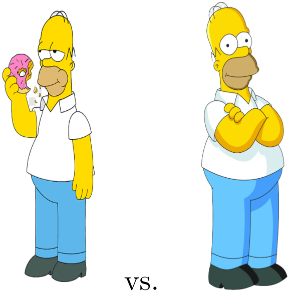
vs.

- What is the impact of giving Homer donuts on his cholesterol level?
  - Impact = Homer's test result with donuts - Homer's result without donuts

The Evaluation Problem

What is the impact of eating donuts on your cholesterol level?

- In an ideal world (research-wise), we could clone each treated individual and observe the impacts of treatment on the outcomes of interest



vs.

- What is the impact of giving Homer donuts on his cholesterol level?
  - Impact = Homer's test result with donuts - Homer's result without donuts
- In the real world, we either observe Homer with donuts or without
- We never observe the **counterfactual**

The Evaluation Problem

- To measure the causal impact of giving Homer donuts on his health, we need to find a similar individual that did not eat donuts

The Evaluation Problem

- To measure the causal impact of giving Homer donuts on his health, we need to find a similar individual that did not eat donuts

vs.

- Our estimate of the impact of the donuts is then the difference in test scores between the treatment group and the comparison group
  - Impact = Homer's test result with donuts - Marge's result without donuts
- As this example illustrates, finding a good comparison group is hard
- In applied economics, your research design is your counterfactual

Constructing the Counterfactual

Impact:

Impact is defined as a comparison between:

1. the outcome some time after the program has been introduced (the “factual)
2. the outcome at that same point in time had the program not been introduced (the “counterfactual”)

Problem

We never observe the outcome of the same individual with and without program at the same point in time

Constructing the Counterfactual

Solution

Compare before and after?

Constructing the Counterfactual

Solution

Compare before and after?

- Other things may have happened over time

Solution

Compare before and after?

Simply compare those who get the treatment with those who did not?

Constructing the Counterfactual

Solution

Compare before and after?

Simply compare those who get the treatment with those who did not?

- Some may choose not to participate/ those not offered somehow participate

Solution

Compare before and after?

Simply compare those who get the treatment with those who did not?

Find a valid Counterfactual

- Find a good proxy for what would have happened to the outcome in the absence of program
- Compare the treated with someone who is exactly like her but who was not exposed to the intervention
- Make sure that the **only reason** for different outcomes between treatment and counterfactual is the intervention

Evaluation Methods

Constructing the **counterfactual** in a convincing way is a key requirement of any serious evaluation method

- Social experiments methods (RCTs)
- Natural experiments
- Matching methods
- Instrumental methods
- Discontinuity design methods

All are an attempt to deal with endogenous selection (assignment)

Randomized Experiments: The “Gold Standard”?

- If properly designed and conducted, randomized evaluations provide the most credible method to estimate the impact of a program
- Because members of the groups (treatment and control) do not differ systematically at the outset of the evaluation,
- Any difference that subsequently arises between them can be attributed to the program rather than to other factors.

Randomized Experiments: Limitations

Attrition Bias

Randomization Bias

Hawthorne and John Henry Effects

Substitution Bias

Supply Side Changes

Cost, Ethics, Power, and Generalizability

Randomized Experiments: Limitations

Attrition Bias

- Attrition rates (i.e. leaving the sample between the baseline and the follow-up surveys) may be different in treatment and control groups
- The estimated treatment effect may therefore be biased

Randomization Bias

Hawthorne and John Henry Effects

Substitution Bias

Supply Side Changes

Cost, Ethics, Power, and Generalizability

Attrition Bias

Randomization Bias

- Can occur if treatment effects are heterogeneous
- The experimental sample may be different from the population of interest because of randomization
- People selecting to take part in the randomized trial may have different returns compared to the population average

Hawthorne and John Henry Effects

Substitution Bias

Supply Side Changes

Cost, Ethics, Power, and Generalizability

Randomized Experiments: Limitations

Attrition Bias

Randomization Bias

Hawthorne and John Henry Effects

- People behave differently because they are part of an experiment and cause bias ("Hawthorne" effects)
- If people from the control group behave differently ("John Henry" effects)

Substitution Bias

Supply Side Changes

Cost, Ethics, Power, and Generalizability

Attrition Bias

Randomization Bias

Hawthorne and John Henry Effects

Substitution Bias

- Control group members may seek substitutes for treatment. This would bias estimated treatment effects downwards
- Can also occur if the experiment frees up resources that can now be concentrated on the control group

Supply Side Changes

Cost, Ethics, Power, and Generalizability

Attrition Bias

Randomization Bias

Hawthorne and John Henry Effects

Substitution Bias

Supply Side Changes

- If programmes are scaled up the supply side implementing the treatment may be different
- In the trial phase the supply side may be more motivated than during the large scale roll-out of a programme

Cost, Ethics, Power, and Generalizability

Randomized Experiments: Limitations

Attrition Bias

Randomization Bias

Hawthorne and John Henry Effects

Substitution Bias

Supply Side Changes

Cost, Ethics, Power, and Generalizability

- Experiments are very costly and difficult to implement properly
- Substantial economic or social outcomes of the Treated
- Samples are often small (e.g. when unit of randomization is a group)
- Difficult to generalize the results of an experiment to the total population

The 2019 Nobel Prize in Economics goes to ...

Controlling for Observables

Basic Idea of Matching

- For each person who is enrolled in the program, match them with someone who is as similar as possible and not enrolled
- Compute the difference in outcomes for each match
- The treatment effect is the weighted average of these differences

$$ATT : \hat{D}_{ATT} = \frac{1}{N_T} \sum_{i=1}^{N_T} (Y_{i1}^T - Y_{j0}^C)$$

where N_T is the number of treated individuals, Y_{i1}^T is a treated observation, and Y_{j0}^C is the untreated observation that is matched with observation i

Identifying assumption

- Conditional on the set of observables X, the non-treated outcomes are independent of the participation status,
$$Y_{0i} \perp T_i | X_i$$
- The Conditional Independence Assumption (sometimes referred to as "unconfoundedness" or "selection on observables") requires that the common variables that affect treatment assignment and treatment-specific outcomes be observable
- The matching assumptions ensure that the only remaining difference between the two groups is programme participation

How to: Matching Methods

- In an ideal setup, the treatment effects are calculated by comparing individuals for whom the values of X are identical (i.e. exact matching on observables)
- In the absence of an exact match, the main alternatives of controlling for observable variables are:
    - Nearest Neighbor Matching
    - Kernel Matching
    - Propensity Score Matching

Selection on Unobservables

Natural or Quasi-Experiments

A popular method in empirical studies is exploiting naturally occurring exogenous variation to mimic a randomized experiment

- As random experiments are very rare, we rely on actual policy changes to identify the effects of policies on outcomes
- These are called "natural experiments" because we take advantage of changes that were not made explicitly to measure the effects of policies
- The key issue when analyzing a natural experiment is to divide the data into a control and treatment group
- The most obvious way to do that is to do a simple difference method using data before ($t = 0$) and after the change ($t = 1$); but it is difficult to distinguish the policy effect from a secular change

- Compare outcomes before and after a policy change for a group affected by the change (Treatment Group) to a group not affected by the change (Control Group)

	Pre	Post	ATE
Treatment	Y_1	Y_2	$(Y_2 - Y_1)$-$(Y_4 - Y_3)$
Control	Y_3	Y_4	

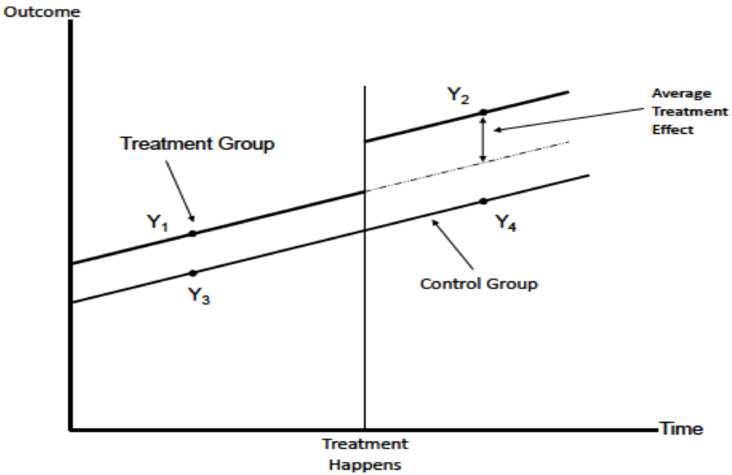
- The idea is to correct the simple difference before and after for the treatment group by subtracting the simple difference for the control group

$$DiD = [\hat{E}(Y_1|T) - \hat{E}(Y_0|T)] - [\hat{E}(Y_1|C) - \hat{E}(Y_0|C)]$$

Common (parallel) trends assumption

- In the absence of treatment, average outcome of the treated group would have changed in the same way as the average outcome of the control group

Dif-in-Dif: Identifying Assumption



Instrumental Variables

The Endogeneity Problem

Instrumental Variables methods are typically used to address the following problems encountered in OLS regression:

- Omitted variable bias.
- Measurement error.
- Simultaneity or reverse causality.

The IV solution

To overcome the endogeneity problem we can use the Instrumental Variables (IV) approach

A valid instrument, Z, needs to satisfy three conditions:

1. Z is as good as randomly assigned
2. Z satisfies the exclusion restriction, i.e. it does not appear as a separate regressor in the original regression we like to run
3. Z is relevant, i.e. affects the endogenous regressor

Of these, only condition 3 can be tested. Conditions 1 and 2 have to be argued based on knowledge from outside the data we have

Regression Discontinuity Design

RD Designs: Making Use of Arbitrary Rules

Regression discontinuity designs (RDD) exploit natural experiments generated by arbitrary rules

- Students receive a scholarship if their GPA is above 3.0
- Legislators are elected if they receive over 50% of the vote
- Children are allowed to start school if they turn 6 by 31 December that year
- Welfare relief is only given to those with less than 40 dollars per month

So, the idea is to estimate the treatment effect using individuals just below the threshold as a control for those just above. RDDs exploit the fact that:

- treatment assignment is based on the value of a continuous variable (the selection variable) X
- the participation rate is discontinuous at least at one known value of that selection variable
- On either side of the common threshold, individuals have very close characteristics, but some are treated and some are not

Tip of the Iceberg

- Many more methods
- Many more issues with each method