Quantitative Economics for the Evaluation of the European Policy

Department of Economics and Management

Irene Brunetti Davide Fiaschi Angela Parenti[1]

25/11/2016

[1] ireneb@ec.unipi.it, davide.fiaschi@unipi.it, and aparenti@ec.unipi.it.

Introduction

- The causal impact of a program/policy is the difference between what happens to recipients of the program and what would have happened to them if they had not received the program.

- The treatment effect on the outcome for unit i is thus the difference between two potential outcomes for each individual:

$$\Delta_i = Y_{1i} - Y_{0i}$$

- The **foundamental problem of causal inference**: we can not observe *both* what happens to an individual after taking the treatment (at a praticular point in time) *and* what happens to that same individual after not taking the treatment (at the same point in time).

- We can never measure a causal effect directly.

Introduction

Estimate of ATE

Y_{0i} and $(Y_{0i}|D = 1)$ are unobserved, then:

$$\begin{aligned} \hat{ATE} &= E[Y_{1t} - Y_{0c}] \\ &= E[Y_{1t}] - E[Y_{0c}] \end{aligned} \tag{1}$$

Assignment Mechanism

It is the procedure that determines which units are selected for the two different groups (treatment and control group). One technique:

- **Random assignment**

Randomized experiments

- **Basic idea**: we can not compare treatment and control outcomes for the same units $\Longrightarrow$ we try to compare them on similar units.
- Similarity is attained by using randomization to decide which unit is assigned to the treatment group and which unit is assigned to the control group.
- Under certain conditions, randomized experiments ensure that outcomes in the control group really do capture the **counterfactual** for a treatment group.

Types of randomizations/experiments

- **Bernoulli trials**: flip coins for each person in the experiment. Problematic because there could be very large or very small treated groups.

- **Completely randomized experiments**: randomly choose a number of treated units, N_t, from the N units in the population. All units have the same marginal probability of being treated. Problem: if there are covariates available, then you might get very unbalanced randomizations.

- **Stratified randomized experiments**: form J blocks, b_j, $j = 1...J$, based on the covariates and then use completely randomized assignment in each block. This eliminates the possibility of bad randomizations since the treatment is by design balanced within blocks. This type of experiment leads to conditional ignorability: $(Y_{1i}; Y_{0i}) \perp D_i | B_i$, where B_i is the blocking variable.

Types of randomizations/experiments

- **Pair randomized experiments**: a stratified randomized experiments where each block has 2 units, one of which receives the treatment. An extreme version of the stratified/blocked randomized experiment.

- **Natural experiments**: experiment where treatment is randomized in some fashion, but that randomization was not under the control of the researcher.

- Natural experiments obviously have lots of pitfalls, because we did not perform the randomization, it's more difficult to justify.

Completely randomized experiments: the Law of Large Numbers

- In randomized experiments, experimental sample is created by sampling from the population we want to study.
- The Law of Large Numbers promises that those in randomly assigned treatment and control samples will be similar if the samples are large enough.
- When you randomly draw a sample of units from a **large population**, and when the number of units you draw **gets large**,
- the average of any characteristic of your sample will tend to become closer to the expected value.

Completely of randomized experiments: the Law of Large Numbers (cont.)

- If the number of units in your sample grows, on average the sample will look like its original population.

- Because randomly assigned treatment and control groups come from the same underlyng population, they are the same in every way. In other words:

$$[\bar{Y}_1|D=1] = [\bar{Y}_1|D=0] \quad \text{and} \quad [\bar{Y}_0|D=1] = [\bar{Y}_0|D=0] \qquad (2)$$

Random assignment: How?

- If you want to assign 50% of the sample to treatment and control: flip a coin for each person.
- If you want to assign 40% of the sample to the treatment group, then roll a die for each person. A 1 or a 2 is treatment; a 3, 4, 5 or a 6 is control.
- Other percentages: Let Excel (or other programs) give each unit a random number. Decide how many units will be in the treatment group (call this T). Assign the T units that get the highest numbers to the treatment group.

Characteristics of randomized experiments

- Randomization ensures that treatment and control group are comparable in every respect (age, proportion of men/women, qualifications, motivation, experience, cognitive abilities, etc.).
- Analyzing data from randomized experimental, researchers almost always begin with a check on whether treatment and control groups look similar $\Longrightarrow$ *checking for balance*: comparison of sample average of covariates.

Covariates balance

- Can check random assignment with respect to observed covariates, $\mathbf{X}$, using so called balance tests (e.g., t-tests) to see if distributions of the covariates are the same in the treatment and control groups.
- $\mathbf{X}$ are pre-treatment variables that are measured prior to treatment assignment (i.e., at baseline).

Example: The RAND Health Insurance Experiment (HIE)

QUESTION: "Does healthcare decrease when its price goes up? and How much is this effect?"

- The experiment ran from 1974 to 1982.
- The HIE enrolled 3,958 people aged 14 to 61 from six area of United States.
- The sample excluded Medicare participants and most Medicaid and military insurance subscribers.
- HIE participants were randomly assigned to one of the 14 plans.
- **Outcome variable**: Price elasticity of demand for healt care.
- **Treatment variable**: Health Insurance plans.
- **Treatment group**: Any insurance plan; **Control group**: catastrophic plan (no-insurance).

The HIE: checking balance

- Demographic characteristics are unchanging. Health variables were measured before random assigment,
- We expect to see only **small differences** in these variables across the groups assigned to different plans.
- See Table for the results.

TABLE 1.3
Demographic characteristics and baseline health in the RAND HIE

	Means	Differences between plan groups			
	Catastrophic plan (1)	Deductible – catastrophic (2)	Coinsurance – catastrophic (3)	Free – catastrophic (4)	Any insurance – catastrophic (5)
A. Demographic characteristics					
Female	.560	−.023	−.025	−.038	−.030
		(.016)	(.015)	(.015)	(.013)
Nonwhite	.172	−.019	−.027	−.028	−.025
		(.027)	(.025)	(.025)	(.022)
Age	32.4	.56	.97	.43	.64
	[12.9]	(.68)	(.65)	(.61)	(.54)
Education	12.1	−.16	−.06	−.26	−.17
	[2.9]	(.19)	(.19)	(.18)	(.16)
Family income	31,603	−2,104	970	−976	−654
	[18,148]	(1,384)	(1,389)	(1,345)	(1,181)
Hospitalized last year	.115	.004	−.002	.001	.001
		(.016)	(.015)	(.015)	(.013)
B. Baseline health variables					
General health index	70.9	−1.44	.21	−1.31	−.93
	[14.9]	(.95)	(.92)	(.87)	(.77)
Cholesterol (mg/dl)	207	−1.42	−1.93	−5.25	−3.19
	[40]	(2.99)	(2.76)	(2.70)	(2.29)
Systolic blood pressure (mm Hg)	122	2.32	.91	1.12	1.39
	[17]	(1.15)	(1.08)	(1.01)	(.90)
Mental health index	73.8	−.12	1.19	.89	.71
	[14.3]	(.82)	(.81)	(.77)	(.68)
Number enrolled	759	881	1,022	1,295	3,198

Notes: This table describes the demographic characteristics and baseline health of subjects in the RAND Health Insurance Experiment (HIE). Column (1) shows the average for the group

Randomized experiments eliminates selection bias

The selection problem when comparing the mean outcomes for the treated and the untreated:

$$\begin{aligned}\mathbf{E}[Y|D=1]-\mathbf{E}[Y|D=0] &= \mathbf{E}[Y_1|D=1]-\mathbf{E}[Y_0|D=0] \\ &= \underbrace{\mathbf{E}[Y_1-Y_0|D=1]}_{\text{ATET}} + \underbrace{\mathbf{E}[Y_0|D=1]-\mathbf{E}[Y_0|D=0]}_{\text{BIAS}} \end{aligned} \tag{3}$$

- when D_i is randomly assigned: $\mathbf{E}[Y_0|D=1]=\mathbf{E}[Y_0|D=0]$ and the differences in expectations by treatement status capture the causal effect of treatment.

Randomized experiments eliminates selection bias

Randomization implies:

(Y_1, Y_0) independent of D, or $(Y_1, Y_0) \perp D$.

We have that $\mathbf{E}[Y_0|D=1] = \mathbf{E}[Y_0|D=0]$ and therefore

$$\alpha_{ATET} = \mathbf{E}[Y_1 - Y_0|D=1] = \mathbf{E}[Y|D=1] - E[Y|D=0]$$

Also, we have that

$$\alpha_{ATE} = \mathbf{E}[Y_1 - Y_0] = \mathbf{E}[Y_1 - Y_0|D=1] = \mathbf{E}[Y|D=1] - \mathbf{E}[Y|D=0]$$

As a result:

$$\underbrace{\mathbf{E}[Y|D=1] - \mathbf{E}[Y|D=0]}_{\text{Difference in means}} = \alpha_{ATE} = \alpha_{ATET}$$

Estimation and inference

Suppose to have a sample on N units. The estimator (proposed by Neyman (1923)):

$$\hat{\alpha} = \bar{Y}_t - \bar{Y}_c, \tag{4}$$

where:

$$\bar{Y}_t = \frac{1}{N_t} \sum_{D_i=1} Y_i;$$

$$\bar{Y}_c = \frac{1}{N_c} \sum_{D_i=0} Y_i$$

with $N_t = \sum_i D_i$ and $N_c = N - N_t$.

- It can be shown that $\hat{\alpha}$ is an **unbiased and consistent estimator** of α_{ATE}.

Estimation and inference

- The variability of the estimator can be attributed solely to the randomization of the treatment.
- Neyman shows that under the completely randomized design, the overall variance of the estimator is simply:

$$\hat{\mathbf{Var}}(\hat{\alpha}) = \frac{S_c^2}{N_c} + \frac{S_t^2}{N_t} \tag{5}$$

- This estimator is unbiased for the variance of the difference in means in the population OR a conservative estimate of the variance of the difference in means in the sample.

Estimation and inference

The randomization distribution of $\hat{\alpha}$ enables us to test the following null hypothesis:

$$H_0 : \alpha = 0$$

Testing in Large Samples

Let:

$$t = \frac{\hat{\alpha}}{\sqrt{\frac{S_c^2}{N_c} + \frac{S_t^2}{N_t}}} \tag{6}$$

- We reject the null hypothesis against the alternative $H_0 : \alpha \neq 0$ at the 5% significance level if $|t| > 1.96$

Can we use regression with experiments?

- We can just run a regression of the outcome on a dicotomic treatment variable.
- First, let's remember how we relate the potential outcomes to the observed outcome:

$$Y_i = Y_{1i}D_i + Y_{0i}(1 - D_i) = \beta + \rho D_i + \epsilon_i \tag{7}$$

  where $\beta = \mathbf{E}[Y_{0i}]$ and remember that $\rho = \mathbf{E}[Y_{1i} - Y_{0i}]$, and $\epsilon = Y_{0i} - \mathbf{E}[Y_{0i}]$.
- Let's check to see if the errors are independent of the treatment, which would imply that a regression estimator $\hat{\alpha}$ would be unbiased for α: $\mathbf{E}[\epsilon_i|D_i = 0] = 0$ and $\mathbf{E}[\epsilon_i|D_i = 1] = 0$.
- Thus, just using the randomization assumption, we have justified the use of regression.

Additional control variables?

Randomization implies that we do not have to adjust for any covariates when estimating causal effects. **BUT**

- To evaluate experimental data one may want to add additional **control variables** (**uncorrelated with the treatment**) in the regression. Instead of estimating Eq.(6) one would estimate:

$$Y_i = \beta + \rho D_i + X_i'\gamma + \varepsilon_i \tag{8}$$

  where X_i represents the vector of control variables.

- Two main reasons to include additional controls:
  - Conditional random assignment. Sometimes randomization is done conditional on some observables.
  - Additional controls increase precision of the estimates (residual variance decreases $\Longrightarrow$ lower standard errors).

Threads to the validity of randomized experiments

- **Internal validity**: *the estimated impact of the program is net of all other potential confounding factors*, or the control group represents the true counterfactual.
  $\Longrightarrow$ Fails when there are differences between treated and controls (other than the treatment itself) that affect the outcome and that we can not control for.

- **External validity**: *can we extrapolate our estimates to other populations? Can the impact estimated in the evaluation sample be generalized to the population of all eliglible units?*
  $\Longrightarrow$ Fails when the treatment effect is different outside the evaluation environment.

Threads to Internal Validity

- Non-compliance with experimental protocol;
- Attrition;
- "Hawthorne" Effects

Non-compliance with experimental protocol

- Not all units assigned to the treatment will actually receive the treatment;
- Some units assigned to control may still receive treatment.

Attrition

- 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.

"Hawthorne" Effects

- People behave differently because they are part of an experiment.
- If they operate differently on treatment and control groups they may introduce biases.
- If people from the control group behave differently these effects are sometimes called "John Henry" effects.

Most common threads to external validity

- Non-representative sample.
- Non-representative program:
  - The treatment differs in actual implementations,
  - Scale effects,
  - Actual implementations are not randomized.

Advantages

- Avoid inefficient expenses:
  - if training programs do not help unemployed to find jobs, stop training programs.
  - if reducing you prices does not have a big impact on demand, keep your prices high.
- Test and learn:
  - instead of generalizing policies which you do not know the effectiveness, test them before on a small sample.

Limits

- Not everything is testable with a randomized experiment:
  - you can not measure the impact of monetary policy on companies investment through a randomized experiment, because it is impossible to set a control group: central bank can have only one interest rate.
- These experiments are costly:
  - you must follow everyone both in the control and test group over a long period of time. When the percentage of people lost to follow-up is too important, this can threaten the validity of your results.