Lecture 1: Causal Inference in Social Science

The 3rd Summer School in Advanced Econometrics by Nanjing University of Finance and Economics(NUFE)

Zhaopeng Qu

Business School,Nanjing University

August 14, 2020

Outlines

Course Overview

Course Overview

- Conceptually, the course is divided into three thematic blocks.
  1. **Causal inference in Social Science**
  2. **Oaxaca-Blinder decomposition**
  3. **Beyond the Mean: DFL decomposition**
- In practice, we also have two parts:
  - Theory: Introduction the basic ideas and rexmaples
  - Computer Labs: Using Stata(Learning by yourself)

Readings

1. **Causal inference in Social Science**
   - Joshua D. Angrist & Jorn-Steffen Pischke, (2014). *Mastering' metrics: The Path from Cause to Effect*. Princeton University Press.(中译本：精通计量：从原因到结果的探寻之旅，格致出版社出版)
2. **Wage Decomposition Methods**
   - Fortin, Nicole, Thomas Lemieux, Sergio Firpo (2011). *Decomposition Methods in Economics*. pp.1-102 in: O. Ashenfelter and D. Card (eds.). *Handbook of Labor Economics*. Amsterdam: Elsevier.
   - 郭继强、姜俪和陆利丽，“工资差异分解方法述评”，《经济学 (季刊)》，2011 年，第 10 卷，第 2 期。
   - Jann, Ben (2008). The Blinder-Oaxaca decomposition for linear regression models. The Stata Journal 8(4):453–479.

About Me

- My name is ***QU, Zhaopeng (曲兆鹏)***
  - Position and Affiliation: *Associate Professor, Institute of Population Studies, Business School.*
  - Research Fields: *Labor Economics and Applied Econometrics*
  - Email: qu@nju.edu.cn
  - Personal Website: https://byelenin.github.io/zh/index.html

Causal Inference in Social Science

The Purposes of Empirical Work

- To prove or disprove a theory(a relations)
  - "The objective of science is the discovery of the relations" —Lord Kelvin
- In most cases,we often want to explore the relationship between two variables in one paper.
  - eg. education and wage
- Then, in simplicity, there are two relationships between two variables.
  - Correlation(相关) V.S. Causality（因果）

A Classical Example: Hemline Index（裙边指数）

- George Taylor, an economist in the United States, made up the phrase it in the 1920s. The phrase is derived from the idea that hemlines on skirts are shorter or longer depending on the economy.
  - Before 1930s, fashion women favored middle skirts most.
  - In 1929, long skirts became popular. While the *Dow Jones Industrial Index(DJII)* plunged from about 400 to 200 and to 40 two years later.
  - In 1960s, DJII rushed to 1000. At the same time, short skirts showed up.
  - In 1970s, DJII fell to 590 and women began to wear long skirts again.
  - In 1990s, mini skirt debuted, DJII rushed to 10000.
  - In 2000s, bikini became a nice choice for girls, DJII was high up to 13000.
  - So what is about now? Long skirt is resorting?

Hemline Index:1920s-2010s

The Core of Empirical Studies: Causality v.s. Forecasting

- Some Big Data researchers think causality is not important any more in our times..
  - "Look at correlations. Look at the 'what' rather than the 'why', because that is often good enough." -Viktor Mayer-Schonberger(2013)
- Most empirical economists think that correlation only tell us the superficial, even false relationship while causal relationship can provide solid evidence to make interference to the real relationship.
  - Today, empirical economists care more about the causal relationship of their interests than ever before.
  - "the most interesting and challenging research in social science is about cause and effect" ——Angrist and Lavy(2008)

The Core of Empirical Studies: Causality v.s. Forecasting

- **Machine learning** is a set of data-driven algorithms that use data to predict or classify some variable Y as a function of other variables X.
  - There are many machine learning algorithm. The best methods vary with the particular data application
- Machine learning is mostly about **prediction**.
  - Having a good prediction does work sometimes but does **NOT** mean understanding causality.

The Core of Empirical Studies: Causality v.s. Forecasting

- Even though forecasting need not involve causal relationships, economic theory suggests patterns and relationships that might be useful for forecasting.
  - Econometric analysis(**times series**) allows us to quantify historical relationships suggested by economic theory, to check whether those relationships have been stable over time, to make quantitative forecasts about the future, and to assess the accuracy of those forecasts.
- The biggest difference between **machine learning** and **econometrics(or causal inference)**.

The Central Question of Causality(I)

- A simple example: **Do hospitals make people healthier?** (Q: Dependent variable and Independent variable?)
- A naive solution: compare the health status of those who have been to the hospital to the health of those who have not.
- Two key questions are documented by the questionnaires from *The National Health Interview Survey(NHIS)*
  1. "*During the past 12 months, was the respondent a patient in a hospital overnight*?"
  2. "*Would you say your health in general is excellent, very good, good ,fair and poor*" and scale it from the number "1" to "5" respectively.

The Central Question of Causality(II)

Hospital v.s. No Hospital

Group	Sample Size	Mean Health Status	Std.Dev
Hospital	7774	2.79	0.014
No Hospital	90049	2.07	0.003

- In favor of the non-hospitalized, WHY?
  - Hospitals not only cure but also hurt people.
    1. hospitals are full of other sick people who might infect us
    2. dangerous machines and chemicals that might hurt us.
  - More important : people having worse health tends to visit hospitals.
- This simple case exhibits that it is NOT easy to answer an causal question, so let us **formalize an model** to show where the problem is.

The Central Question of Causality(III)

- A right way to answer a causal questions is construct a counterfactual world, thus "What If ....then" , Such as
- An classical example: How much wage premium you can get from college attendance(上大学使工资增加多少 ?)
  - For any worker, we want to compare
    - Wage if he have a college degree
    - Wage if he had not a college degree
  - Then make a difference. This is the right answer to our question.

Difficulty in Identification

- Others are the same as
  - Military service
  - Migration
  - Public policies
  - Road building
  - Job training
  - Party membership
  - Others
- Difficulty: only one state can be observed

Formalization: Rubin Causal Model

- Treatment : $D_i = \{0, 1\}$; eg, go or not go to college

$$Potential\ Outcomes = \begin{cases} Y_{1i} & if\ D_i = 1 \\ Y_{0i} & if\ D_i = 0 \end{cases}$$

- To know the difference between Y_{1i} and Y_{0i}, thus $Y_{1i} - Y_{0i}$, which can be said to be the **causal effect** of going to college for individual i.

Definition

Causal inference is the process of estimating a **comparison of counterfactuals** under different treatment conditions on the same set of units.

Formalization: Treatment

- Treatment : D_i can be a **multiple valued**(countinuous) variable

$$D_i = s$$

- Examples:
  - Schooling years
  - Number of Children
  - Number of advertisements
  - Money Supply
- For simplicity, we assume treatment variable D_i is just a **dummy**.

Formalization: Potential Outcomes

- A potential outcome is the outcome that would be realized if the individual received a specific value of the treatment.
  - Annual earnings if attending to college
  - Annual earnings if not attending to college
- For each individual, we has two potential outcomes, Y_{1i} and Y_{0i}, one for each value of the treatment
  - Y_{1i} : Potential outcome for an individual i with treatment.
  - Y_{0i} : Potential outcome for an individual i with treatment.

$$Potential\ Outcomes = \begin{cases} Y_{1i} & if\ D_i = 1 \\ Y_{0i} & if\ D_i = 0 \end{cases}$$

Causal effect for an Individual

- To know the difference between Y_{1i} and Y_{0i}, which can be said to be the **causal effect** of going to college for individual i. (Do you agree with it?)

Definition

Causal inference is the process of estimating a comparison of counterfactuals under different treatment conditions on the same set of units. It also call Individual Treatment Effect(ICE)

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

Formalization: Estimate ICE

- Due to unobserved counterfactual outcome, we need to make strong assumptions to estimate ICE.
  - Rule out that the ICE differs across individuals ("*heterogeneity effect*")
- Knowing individual effect is not our final goal. As a social scientist, we would like more to know the *Average* effect as a **social pattern**.
- So it make us focus on the average wage for a group of people.
  - How can we get the average wage premium for college attendance?

Conditional Expectation:

- **Expectation:** We usually use $E[\,Y_i]$ (the expectation of a variable Y_i) to denote population average of Y_i
  - Suppose we have a population with N individuals

$$E[\,Y_i] = \frac{1}{N}\Sigma_{i=1}^{N}\, Y_i$$

- **Conditional Expectation:**
  - The average wage for those who attend college: $E[\,Y_i|\,D_i = 1]$
  - The average wage for those who did not attend college: $E[\,Y_i|\,D_i = 0]$

Average Causal Effects

Average Treatment Effect (ATE)

$$\alpha_{ATE} = E[\delta_i] = E[Y_{1i} - Y_{0i}]$$

- It is average of ICEs over **the population**.

Average treatment effect on the treated(ATT)

$$\alpha_{ATT} = E[\delta_i | D_i = 1] = E[Y_{1i} - Y_{0i} | D_i = 1]$$

- Average of ICEs over the **treated population**

Fundamental Problem of Causal Inference

- We can never directly observe causal effects (ICE, ATE or ATT)
- Because we can never observe both potential outcomes (Y_{0i}, Y_{1i}) for any individual.
- We need to compare **potential outcomes**, but we only have **observed outcomes**
- So by this view, causal inference is a **missing data** problem.

Observed Association and Selection Bias

- Causality is defined by **potential outcomes**, not by **realized (observed) outcomes**.
- In fact, we can not observe all potential outcomes .Therefore, we can not estimate the above causal effects without further assumptions.
- By using observed data, we can only establish **association (correlation)**, which is the observed difference in average outcome between those getting treatment and those not getting treatment.

$$\alpha_{corr} = E[\, Y_{1i} | D_i = 1] - E[\, Y_{0i} | D_i = 0]$$

College vs Non-College Wage Differentials:

- Comparing the average wage in labor market who went to college and did not go.

College vs Non-College Wage Differentials:

$$
\begin{aligned}
&= E[Y_{1i}|D_i = 1] - E[Y_{0i}|D_i = 0] \\
&= \{E[Y_{1i}|D_i = 1] - E[Y_{0i}|D_i = 1]\} + \{E[Y_{0i}|D_i = 1] - E[Y_{0i}|D_i = 0]\}
\end{aligned}
$$

- Question 1: Which one defines the causal effect of college attendance?

Formalization: Rubin Causal Model

- **Selection Bias(SB)**implies the potential outcomes of treatment and control groups are different even if both groups receive the same treatment

$$E[Y_{0i}|D_i = 1] - E[Y_{0i}|D_i = 0]$$

- Question 2: Selection Bias is positive or negative in the case?
- This means two groups could be quite different in other dimensions: other things are not equal.
- Observed association is *neither necessary nor sufficient for causality.*

Causal Effect and Identification Strategy

- Many Many Other examples
  - the effect of job training program on worker's earnings
  - the effect of class size on students performance
  - ....
- The main obstacle of identifying causal effects is **to eliminate the selection bias**.

Experimental Design as an Benchmark

Random Assignment Solves the Selection Problem

- Random assignment of treatment D_i can eliminates selection bias. It means that the treated group is a random sample from the population.
- Being a random sample, we know that those included in the sample are **the same, on average,** as those not included in the sample on any measure.
- Mathematically ,it makes D_i **independent** of potential outcomes, thus

$$D_i \perp (Y_{0i}, Y_{1i})$$

- **Independence**: Two variables are said to be independent if knowing the outcome of one provides no useful information about the outcome of the other.
  - Knowing outcome of $D_i(0, 1)$ does not help us understand what potential outcomes of (Y_{0i}, Y_{1i}) will be

Random Assignment Solves the Selection Problem

- So we have

$$E[Y_{0i}|D_i = 1] = E[Y_{0i}|D_i = 0]$$

- Thus the **Selection Bias** equals to **ZERO**.
- Then **ATT** equals **Observed Association** because the

$$\begin{aligned} E[Y_{1i}|D_i = 1] - E[Y_{0i}|D_i = 0] &= E[Y_{1i}|D_i = 1] - E[Y_{0i}|D_i = 1] \\ &= E[Y_{1i} - Y_{0i}|D_i = 1] \end{aligned}$$

- No matter what assumptions we make about the distribution of Y , we can always estimate it with the difference in means.

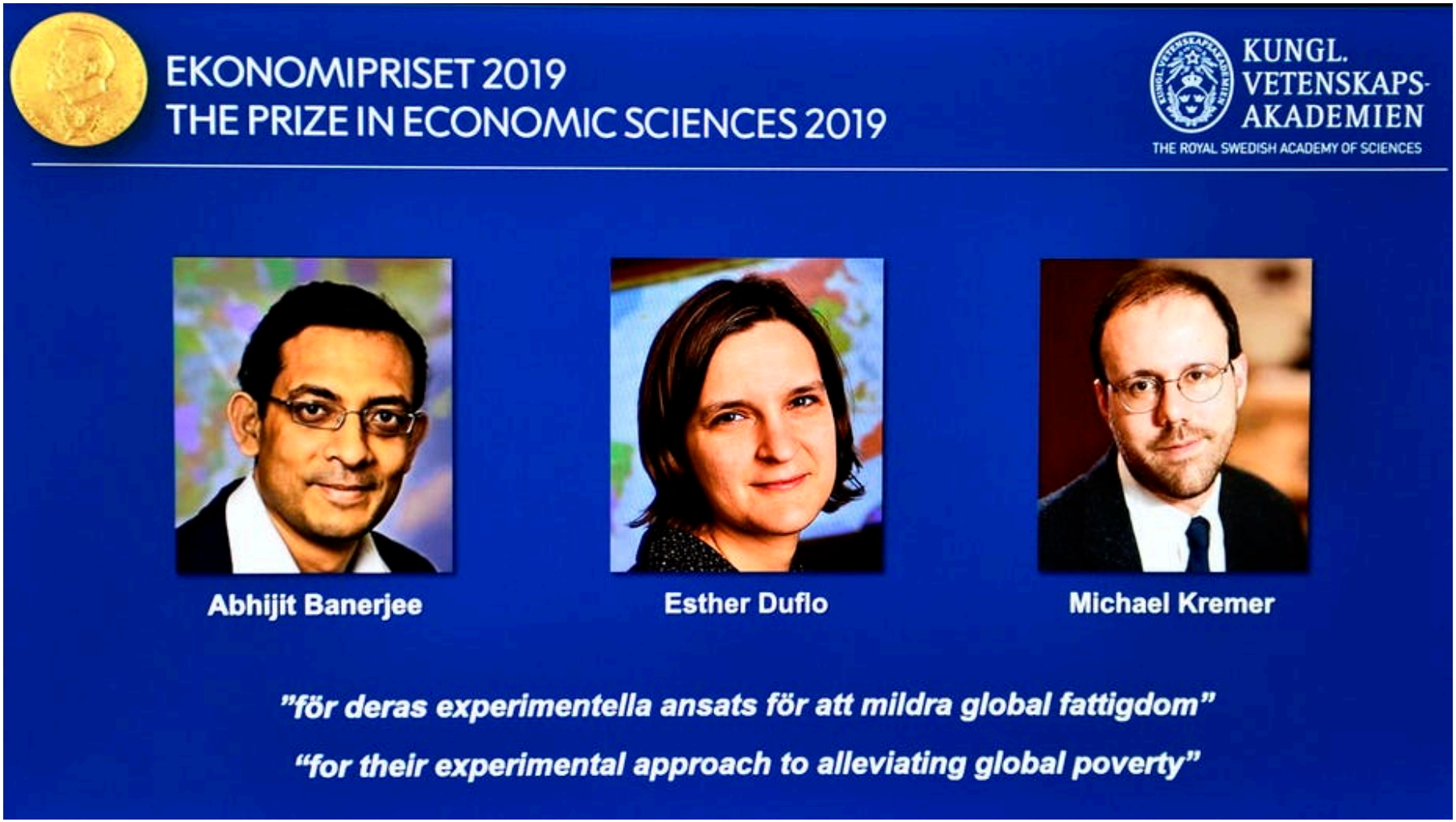
Our Benchmark: Randomized Experimental Methods

- Think of causal effects in terms of comparing counterfactuals or potential outcomes. However, we can never observe both counterfactuals —fundamental problem of causal inference.
- To overcome the difficulty, We could use **Random Controlled Trials/Experiments(RCT)**.
- RCT is a form of investigation in which units of observation (e.g. individuals, households, schools, states) are randomly assigned to treatment and control groups.

Our Benchmark: Randomized Experimental Methods

- RCT has two features that can help us hold "other things equal" and then eliminates selection bias
  - Random assign treatment:
    - Random assignment (such as a coin flip) ensures that every observation has the same probability of being assigned to the treatment group.
    - Therefore, the probability of receiving treatment is unrelated to any other confounding factors.
  - Sufficient large sample
    - Large sample size can ensure that the group differences in individual characteristics wash out
- If we could observe the counterfactual directly, then there is no evaluation problem, just **simply difference**.
- RCTs are considered the ***gold standard***for establishing a causal link between an intervention and change.

Randomized Experimental Methods: Noble Prize 2019



Program Evaluation Econometrics

RCTs are far from perfect!

- Potential Problems in Practice
  - Small sample: Student Effect
  - Hawthorne effect: The subjects are in an experiment can change their behaviors.
  - Attrition: It refers to subjects dropping out of the study after being randomly assigned to the treatment or control group.
  - Failure to randomize or failure to follow treatment protocol: People don' t always do what they are told.
    - Wearing glasses program in Western Rural China.
  - High Costs, Long Duration

RCTs are far from perfect!

- Potential Ethical Problems: "Parachutes reduce the risk of injury after gravitational challenge, but their effectiveness has not been proved with randomized controlled trials."
  - Milgram Experiment
  - Stanford Prison Experiment
  - Monkey Experiment
- Limited Generalizability
- RCTs allow us to gain knowledge about causal effects without knowing the mechanism.

Program Evaluation Econometrics(项目评估计量经济学)

- The various approaches using naturally-occurring data provide alternative methods of constructing the proper counterfactual.
  - **Econometrics or Program Evaluation Methods**
  - Congratuation! We are working and studying in a more tough and intractable area than many others disciplines including those often claim that they are more "scientific" .
- We should take the randomized experimental methods as our benchmark when we do empirical research whatever the methods we apply.

Program Evaluation Econometrics(项目评估计量经济学)

- Question: How to do empirical research scientifically when we can not do experiments? It means that we always have selection bias in our data, or in term of "endogeneity".
- Answer: Build a reasonable counterfactual world by naturally occurring data to find a proper control group is the core of econometrical methods.
- Here you Furious Seven Weapons in Applied Econometrics(应用计量的七种盖世武器)
  1. RCTs (随机实验)
  2. OLS (多元回归)
  3. Matching and Propensity Score (匹配与倾向得分)
  4. Wage Decomposition (分解)
  5. Instrumental Variable (工具变量)
  6. Regression Discontinuity (断点回归)
  7. Panel Data and Extensions: Fixed Effects (固定效应), Differences in Differences (双差分) ,Synthetic Control Methods (合成控制) ....

Intuition to All Methods: Mean Comparisons

- **Common Idea**: match silimar units or construct the proper counterfactuals for the actuals, then produce a mean comparison
  - RCT compares means directly between treatment and control group.
  - OLS gives conditional mean comparison.
  - Matching make a weighted conditional mean comparison.
  - IV compares means of instrumented and non-instrumented.
  - RD compares means around the cutoff.
  - DID compares difference in mean across locations or time.
  - SCM is a special type of DID.
- **Goal**: give a believable and reliable mean comparison with counterfactuals

Program Evaluation Econometrics

- These Furious Seven are the most basic and popular methods in applied econometrics and so powerful that
  - even if you just master one, you may finish your empirical paper and get a good score.
  - if you master several ones, you could have opportunity to publish your paper.
  - If you master all of them, you might to teach applied econometrics class just as what I am doing now.
- We will introduce a special one of these methods in the class: **Decomposition**. Let's start our journey together.

Wrap up

Summary

- The Core of Empirical Studies: *Causality v.s. Forecasting*
- The Central Question of Causality?
  - Rubin Causal Model: comparing counterfactuals or potential outcomes.
  - However, we can never observe both counterfactuals — fundamental problem of causal inference.
- To construct the counterfactuals, two broad categories of empirical strategies.
  - **Random Controlled Trials/Experiments** and **Program Evaluation Econometrics**
- Build a reasonable counterfactual world by naturally occurring data to find a proper control group is the core of econometrical methods.
- Furious Seven are amazing weapons in emprical studies of applied economics, and we had better learn and master them.

An Amazing Journey