An introduction to Causal Inference using

Quasi Experiment & Observational Studies

Jessie Gao

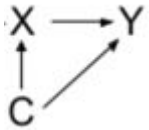
Content

Causal Inference Problem

Causal Inference Basics

- Unit: The person, place, or thing upon which a treatment will operate, at a particular time
- Treatment: An intervention, the effects of which (on some particular measurement on the units) the investigator wishes to assess relative to no intervention (i.e., the “control”)
- Potential Outcomes: The values of a unit’s measurement of interest after (a) application of the treatment and (b) non-application of the treatment (i.e., under control)
- Causal Effect: For each unit, the comparison of the potential outcome under treatment and the potential outcome under control
- The Fundamental Problem of Causal Inference: We can observe at most one of the potential outcomes for each unit (so we can only infer the effect from different units who received different treatments)

Problem Statement: Estimate the causal effect of X on Y, with existence some set of confounding factors, C, that affects both the input of interest (treatment) X, and the outcome of interest Y



X: input
Y: outcome
C: confounding variables

Example Use Cases

- *Tech: some product, feature, or initiative;*
- *Government: some policy regulation or laws*

Causal Inference Problem

Assumptions: In order to estimate any causal effect, assumptions that must hold

1. Positivity: Any individual has a positive probability of receiving all values of the treatment variable: $Pr(X=a) > 0$.
2. Replication: At least one unit receives treatment and at least one unit receives control
3. Consistency: The observed outcome for every treated individual equals her outcome if she had received treatment, and that the observed outcome for every untreated individual equals her outcome if she had remained untreated
4. No interference: Outcomes only depend on the treatment applied to the unit, and not on treatments applied to other units
5. Single Version of Treatment: Each Y corresponds to a single version of treatment
6. Exchangeability: the control and treatment groups are so similar that they could be exchanged

(4 & 5 combined is called Stable Unit Treatment Value Assumption, or SUTVA)

Two Types of Validity:

- Internal validity: Results are unbiased for the subpopulation studied
- External validity Results are unbiased for the full population

Causal Inference Methods Overview

Controlled Experiment

- Valid by Randomization
  - Internal Validity: confounding variables' effect can be ignored as they are same across control and treatment groups and the treatment and control samples are balanced.
  - External Validity: if the impact on the experimental group was representative of the overall population
- Gold standard if you can, but this type of experiment may be off the table due to:
  - Fundamentally unethical
  - Cost to implement
  - Losing trust: e.g., confusing users who have been used to the test function, or price

Quasi-Experiment or Observational Studies

- Valid utilizing the confounders (rarely they are all observable or measurable)
  - Controlled Regression: Adding proxies for unobservables (won't be sufficient if noticing that adding the proxy to the regression meaningfully impacts the coefficient of X)
- Valid by factoring out estimated confounders' effects
  - Regression Discontinuity Design (RDD)
  - Instrumental Variables (2-stage/IV) with/without Randomized Encouragement Trial
  - Difference in Differences (DID)

Content

Problem of Causal Inference

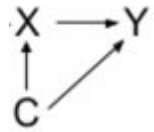
➔ Regression Discontinuity Design

Instrumental Variables

Difference in Differences

Regression Discontinuity Design

Solution: Taking advantage of randomness in the world that a very narrow range around a threshold can be seen as a locally randomised experiment. The differences of confounders among samples in the range are ignorable and considered randomized providing the large sample size around the threshold.



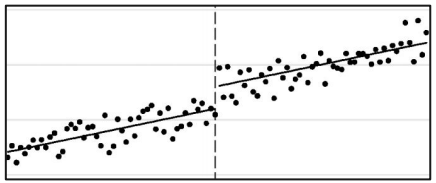
Example: Evaluation of impact of merit awards on future outcomes

- Treatment: The group with current test score above threshold (e.g., 60) would receive award, o.w., not
- Confounders: The groups who received awards tend to keep doing well in future and the outcome may not be from award but more from confounders (e.g., all of they reasons they are doing great now such as ability, income, motivation, etc) - cannot regress on y_outcome ~ x_award
- RDD: limiting to study the units around the threshold will give a natural randomized experiment
    - Confounding variables are similar for who got 59 or 60 and can be seen randomized
    - We can safely treat 59-ers as controlled group and 60-ers as treatment group

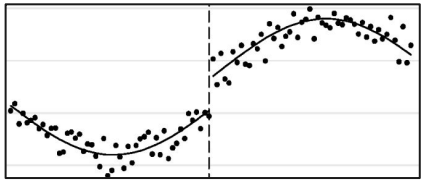
Regression Discontinuity Design (cont')

Modeling Formula: $$Yi = f_1(Xi)1(Xi < X_0) + f_2(Xi)1(Xi \geq X_0) + \beta Di + e_i$$

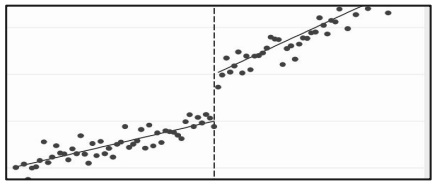
- By function type



Linear



Non-Linear



Varying (Unequal slopes)

- Sharp RDD vs Fuzzy RDD

$$Di = 1\ ,\ if\ X \geq X_0$$
$$Di = 0\ ,\ if\ X_i < X_0$$

Sharp RDD

$$Pr(Di = 1) = p(Xi)$$

Fuzzy RDD

Regression Discontinuity Design (cont')

Validity:

- Internal Validity:
  - Assumption 1: Unit cannot control whether they are just above or below the threshold (e.g., arguing with grader)
  - Assumption 2: No confounding discontinuities - being just above or below the threshold will not influence other features (e.g., no other actions are done other than the reward)
- External Validity:
  - In RDD, the need for measurement is only the effect of *local average treatment effect (LATE)* as the intervention we care is right on the margin (e.g., whether there is award depending on passing or not)

Content

Problem of Causal Inference

Regression Discontinuity Design

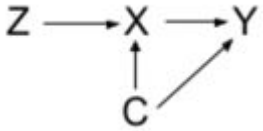
➔ Instrumental Variables

Difference in Differences

Instrumental Variables

Solution: Utilizing an Instrumental Variable Z to "factor out" confounder effects, which

- Meaningfully affects X (Strong first stage)
- Affects Y only through its effect on X (Exclusion restriction)



Without IV

Modeling Formula:

$$y = bx + e$$

e includes all errors not attributed including confounder bias

Validity:

- Internal

$$b = \frac{\mathrm{Cov}(x,y)}{\mathrm{Cov}(x,x)}$$

$$= \frac{\mathrm{Cov}(x, xb+e)}{\mathrm{Cov}(x,x)} = b + \frac{\mathrm{Cov}(x,e)}{\mathrm{Cov}(x,x)}$$

$\mathrm{Cov}(x,e) = 0$ *if there is no confounder effect*

$$= b$$

Without IV

First stage: $x_c = az_c + d_c$ (Instrument for X with Z)

Second stage: $y_c = bx_c + e_c$ (Estimate the effect of the (instrumented) X on Y

$$b^{iv} = \frac{\mathrm{Cov}(z,y)}{\mathrm{Cov}(z,x)}$$

$$= \frac{\mathrm{Cov}(z, bx+e)}{\mathrm{Cov}(z,x)}$$

$$= \frac{b\mathrm{Cov}(z,x) + \mathrm{Cov}(z,e)}{\mathrm{Cov}(z,x)}$$ *z,e are independent*

$= b$ *(Internal Validity from strong first stage, exclusion restriction)*

- External Valid if data is representative or if the regressor digested data of the whole population

Instrumental Variables (cont')

Example: The search for IVs is a critical step in using this methods. IVs are much easier to find in tech use cases

Policy Evaluation
Problem: Evaluating construction of highway systems (X) to economics growth (Y)
IV Selection: Geographical information such as Ruggedness as IV (Z)

Tech Product Analysis
Problem: Evaluating whether users who have more friends on platform are less likely to churn
IV Selection: treatment group -1, control group - 0

- Bucketing of a previous successful A/B test of referrals (e.g., email campaign test, very common
- Randomized Encouragement Trials: a test implemented to specifically facilitate IV, such as a prompt to add friends which aims to encourage treatment group to have a higher number of friends, while uncorrelated with anything else as it is caused by a randomly-assigned treatment (note that control groups also receive "blank" email just not mentioning referral problem in case that email instead of referral program is driving retention)

Validity:

- Strong first stage assumption: if treatment group has more friends
- Restriction Exclusion: Predicted number of friends captures only the variance in number of friends randomly generated through treatment (confounders distributions are same in both groups)

example

$$x_c = \overset{2}{a} z_c + \overset{15}{d_c}$$

$$y_c = \overset{0.00060}{b} x_c + \overset{0.2}{e_c}$$

Content

Problem of Causal Inference

Regression Discontinuity Design

Instrumental Variables

➔ Difference in Differences

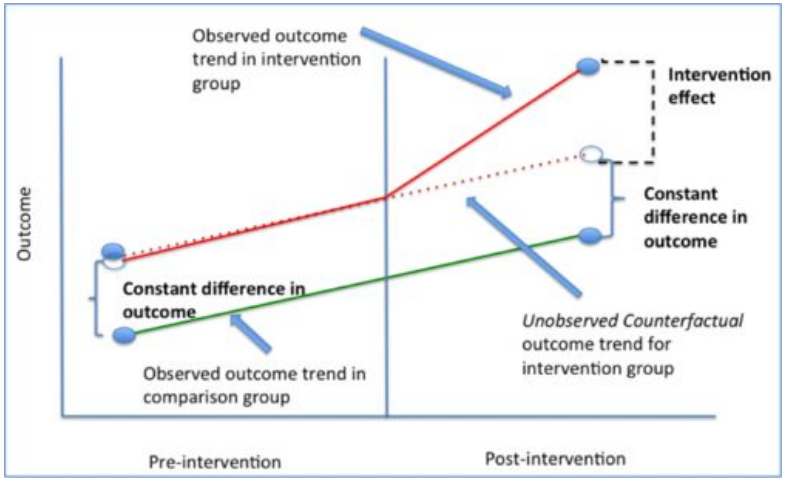
Difference in Differences

Solution: When exchangeability cannot be assumed between treatment and control group (randomization on individual level is not possible), but there is shared constant difference between the groups in absence of the treatment (*"parallel assumption"*), DID can be used to obtain some counterfactual $Y(X = 1) - Y(X = 0)$ to estimate causal effect

Example:

- **Tech Product Analysis:** Evaluation of revenue per session from campaign (treatment) launched in one region (treatment group) of which the counterfactual (this region's outcome absent the treatment) can only be achieved by estimating using some other region (control group), for which the outcome has co-occurred without the treatment
- **Policy Evaluation:** Estimation the effects of a passage of law, enactment of policy, or a program implementation (treatment) by comparing the changes in outcomes over time between a population that is enrolled (treatment group) and a population that is not (control group)

Illustration:



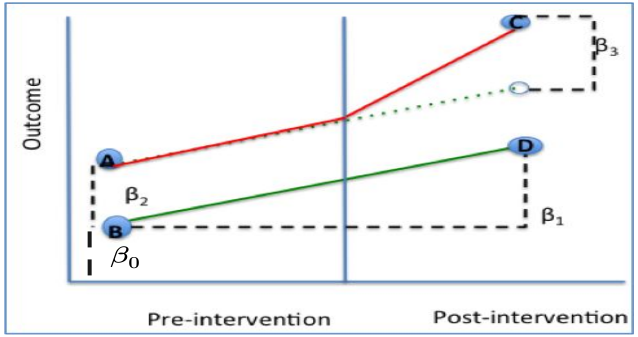
Difference in Differences (cont')

Modeling Formula:

$$Y = \beta_0 + \beta_1 * [Time] + \beta_2 * [Intervention] + \beta_3 * [Time * Intervention] + \varepsilon$$

The coefficient on the interaction term is the estimated effect

Coefficient	Calculation	Interpretation
β_0	B	Baseline average
β_1	D-B	Time trend in control group
β_2	A-B	Difference between two groups pre-intervention
β_3	(C-A)-(D-B)	Difference in changes over time



Validity:

- Internal: Treatment and control groups have Parallel Trends in outcome (hardest to fulfill)
  - Test 1) The trend before treatment are same for both groups, 2) The composition of population in both groups before and after is the same.
  - The result may only be valid by sub-group as Intervention may have similar/different effect on components. Also the smaller the time period tested, the more likely the assumption is to hold
- External: DID usually is used to estimate the treatment effect on the treated (causal effect in the exposed), although with stronger assumptions the technique can be used to estimate for the population

Summary

Controlled Experiment is always the gold standard and the most accurate.

When it's off the table, Quasi-experiment and Observational Studies can be used to achieve the same goal with less cost and also guaranteed accuracy if the assumptions made are realistic and tested.

Q&A

Re-work Conference Videos

Click here and login using my email address **gaojingxu.thu@gmail.com** and the password **REWORKSF20**

https://re-work.us16.list-manage.com/track/click?u=969b1987cadb0e680b9a47c09&id=6009cacf40&e=5d8d4efc01

Thank you!

Graveyard

Propensity Score Matching

WIP

Fixed Effect Regression

WIP

DD ->Score Propensity Score Matching PSM

Key assumption required for internal validity of the DD estimate is parallel trends: absent the treatment itself, the treatment and control markets would have followed the same trends. That is, any omitted variables affect treatment and control in the same way

Make the treatment and control groups as similar as possible. In the experimental set-up, consider implementing stratified randomization. Although generally unnecessary when samples are large (e.g., in user-level randomization), stratified randomization can be valuable when the number of units (here geos) is relatively small. Where feasible, we might even generate “matched pairs” — in this case markets that historically have followed similar trends and/or that we intuitively expect to respond similarly to any internal product changes and to external shocks.

After the stratified randomization (or matched pairing), check graphically and statistically that the trends are approximately parallel between the two groups pre-treatment

Also keep the monitoring: Confounders pop up across some subset of treatment and control markets

Fixed Effects Regression

Fixed effects is a particular type of controlled regression

Fixed effects regression is similar to RDD in that both take advantage of the fact that users are distributed quasi-randomly around some point. In RDD, there is a single point; with fixed effects regression, there are multiple points — in this case, one for each week of account opening. First, after conditioning on the fixed effects, users are as good as randomly assigned to to their X values — in this case, their portfolio returns. Second, there can be no confounding discontinuities, i.e., conditional on the fixed effects, users cannot otherwise be treated differently based on their X.