目录

Causal Relationship

DAG

- A graph that is both **directed** and **acylic** is called a **directed acyclic graph** (DAG).
- A back-door path from node A to node B is a path from A to B that starts with an incoming arrow into A and ends with an incoming arrow into B.

Basic patterns of causal relationships:

1. Chains (causal pathway)
2. Forks (confounding pathway)
3. Inverted forks (colliding pathway)

Strategies to estimate causal effects

1. condition on variables (that block all back-door paths from the causal variable to the outcome variable).
   - However, conditioning on a collider variable does not simplify the original graph but rather adds complications by creating new association.
2. use exogenous variation in an appropriate instrumental variable (to isolate covariation in the causal and outcome variables).
3. establish an isolated and exhaustive mechanism (that relates the causal variable to the outcome variable).

Back-door criterion

> The causal effect is identified by conditioning on a set of variables in Z if and only if all back-door paths between the causal variable and the outcome variable are blocked after conditioning on Z.
>
> All back-door paths are blocked by Z (i.e., d-separated by Z) if and only if each back-door path:
>
> 1. contains a chain of mediation $A \to C \to B$, where the middle variable C is in Z;
> 2. contains a fork of mutual dependence $A \leftarrow C \to B$, where the middle variable C is in Z;
> 3. contains an inverted fork of mutual causation $A \to C \leftarrow B$, where the middle variable C and all of C's descendants are not in Z.

Back-door path: A path between any causally ordered sequence of two variables that includes a directed edge $\rightarrow$ that points to the first variable.

Front-door criterion

> One can consistently estimate the effect of a causal variable on an outcome variable by estimating the effect as it propagates through an isolated and exhaustive mechanism.
>
> - Isolated: none of the components of unblockable back-door paths have direct effects on the mechanism.
> - Exhaustive: All front-door paths are identifiable.

Prove front-door criterion from back-door criterion:

- $D \rightarrow M$: There are no back-door paths between D and M.
- $M \rightarrow Y$: There are some back-door paths via D, however all of them can be blocked by D.

Causal Inference

Average treatment effect (ATE)

$$\tau_{ATE} = \mathrm{E}\left[Y_i\left(1\right) - Y_i\left(0\right)\right]$$

Average treatment effect on the treated (ATET)

$$\tau_{ATET} = \mathrm{E}\left[Y_i\left(1\right) - Y_i\left(0\right)|D_i = 1\right]$$

Average treatment effect on the untreated (ATENT)

$$\tau_{ATENT} = \mathrm{E}\left[Y_i\left(1\right) - Y_i\left(0\right)|D_i = 0\right]$$

> ATE is a **weighted average** of the ATET and the ATENT
>
> $$\tau_{ATE} = p\left(D_i = 1\right) \cdot \tau_{ATET} + p\left(D_i = 0\right) \cdot \tau_{ATENT}$$
>
> Conditional Independence Assumption: $Y_{(d)} \perp D_i|X_i$.
>
> Conditional Mean Independence: $\mathrm{E}\left[Y_i|x, D\right] = \mathrm{E}\left[Y_i|x\right], \mathrm{E}\left[Y_0|x, D\right] = \mathrm{E}\left[Y_0|x\right]$.

DID

Idea of DID: Compare the outcomes of the units that are affected by the policy change and those who are not affected before and after the policy was enacted.

DID with various forms

1. Different treatment time

$$y_{it} = \alpha_0 + \beta D_{it} + \mu_i + f_t + \varepsilon_{it}$$

2. Different treatment magnitude

$$y_{it} = \alpha_0 + \beta \left(x_i \cdot d \right) + \mu_i + f_t + \varepsilon_{it}$$

3. A general form

$$y_{it} = \alpha_0 + \beta_t \left(x_i \cdot d_{it} \right) + \mu_i + f_t + \varepsilon_{it}$$

DID with matching

- It can be seen as a non-parametric DID, reweighting observations according to a weighting function dependent on the specific matching approach adopted.
- Advantage: it does not require the imposition of the linear-in-parameters form of the outcome equation.

For panel data

$$\widehat{ATET}_{M-DID} = \frac{1}{N_1} \sum_{i \in T} \left[\left(Y_{i1}^T - Y_{i0^T} - \sum_{j \in C(i)} h\left(i, j\right) \left(Y_{j1}^C - y_{j0}^C \right) \right) \right]$$

Parallel trend assumption

- Test the assumption for periods before the treatment.
- Allows for heterogeneity in time trend.

IV

Problems of IV estimation

1. Inconsistency of IV: What if the instrument is not fully exogenous for the outcome. The bias increases as the covariance of Z and X decreases.
2. Lower efficiency of IV: The variance of IV is always greater than that of OLS.
3. Small-sample bias: The bias of 2SLS when one cannot invoke the usual asymptotic results.

Estimation with a binary IV

$$\hat{\beta} = \frac{\mathrm{E}\left[y|z=1\right] - \mathrm{E}\left[y|z=0\right]}{\mathrm{E}\left[d|z=1\right] - \mathrm{E}\left[d|z=0\right]}$$

- An IV estimator is a ratio that is a joint projection of y and d onto a third dimension of z.
- An IV estimator isolates a specific portion of the covariation in D and Y.
  - By using the fitted values of the endogenous regressor from the first-stage regression, our regression now uses only the exogenous variation in the regressor due to the instrumental variable itself.
  - Instrumental variable therefore reduces the variation in the data, but that variation is exogenous.
- A binary IV identifies only the average causal effect for compliers. (the subset of all students who would attend a private school if given a voucher but who would not attend a private school in the absence of a voucher)

LATE

The LATE theorem:

- Independece
- Exclusion
- First stage
- Monotonicity

LATE is not the same as ATT:

$$\begin{aligned}\underbrace{\mathrm{E}\left[Y_i^1 - Y_i^0|D_i=1\right]}_{ATT} =& \underbrace{\mathrm{E}\left[Y_i^1 - Y_i^0|D_i=1\right] \cdot p\left(D_i^0=1|D_i=1\right)}_{\text{Effect on Always-Takers}} \\ &+ \underbrace{\mathrm{E}\left[Y_i^1 - Y_i^0|D_i=1\right] \cdot p\left(D_i^1 > D_i^0, Z_i=1|D_i=1\right)}_{\text{Effect on Compliers}}\end{aligned}$$

ATT is a weighted average of the effects on always-takers and compliers.

PSM

Balancing scores

A balancing score $b\left(x\right)$ is a function of the covariates such that

$$W_i \perp X_i | b\left(X_i\right)$$

The idea is to find lower-dimensional functions of the covariates that suffice for the removing the bias associated with differences in the pre-treatment variables.

Property: If assignment to treatment is unconfounded given the full set of covariates, then assignment is also unconfounded conditioning only on a balancing score.

- The propensity score provides the biggest benefit in terms of reducing the number of variables we need to adjust for.
- Units stratified according to the propensity score should be indistinguishable in terms of their X.

Identification assumptions of matching

- Conditional independence (unconfoundedness): There exists a set of observable covariates such that after controlling for these covariates, treatment assignment is independent of potential outcomes.
- Common support: For each value of X, there is a positive probability of being both treated and untreated.

Decisions to make in matching

- Variables to be used for matching.
- Distances for measuring closeness in covariates.
- The number of controls.
- Caliper.
- Stratification: For covariates that should be matched exactly, forming the same strata depending on such covariates in the treatment and control groups.
- Greedy/Nongreedy. (Greedy matching: A control is matched to only one treated at most.)

Basic matching estimators

$$\hat{\tau} = \frac{1}{N_1} \sum_{t \in T} \left(Y_t - \sum_{c \in (C)} w_{ct} Y_c \right)$$

However, it is hard to find the asymptotic distribution of a matching estimator, because selecting a comparison group C_t for treated i involves all observations. This implies dependence across the paired differences to make the iid presumption is false. Ignoring this dependence is likely to underestimate the asymptotic variance.

Assessing overlap in covariate distribution

It is useful to assess the degree of overlap in the covariate distributions.

- The difference in locations: normalized difference
- The difference in dispersions: the logarithm of the ratio of standard deviations
- Direct measures of overlap: investigate the fraction of the treated units have covariates values that are in the tails of the distribution of the covariate values for the controls.

RCT

Definition of treatment effects

Individual treatment effect is the difference between the potential outcomes of it being treated minus those when it not being treated.

There are two important aspects of the definition of a causal effect:

- The definition of the causal effect depends on the **potential outcomes**, but it does not depend on which outcome is actually observed.
- The causal effect is the comparison of potential outcomes, for the same unit, at the same moment after the treatment.

In experimental and quasi-experimental designs, the treatment effect is generally estimated by the **counterfactual approach**. Counterfactual causality takes the form of a comparison between the outcome of a unit when the unit is treated in a certain way, and the outcome of the same unit when it is not treated.

Bias in estimation of treatment effects

We can use observed outcomes to estimate the treatment effect for unit i:

$$Y_i - Y_j = Y_i(1) - Y_j(0) = \underbrace{Y_i(1) - Y_i(0)}_{\gamma_i} + \underbrace{Y_i(0) - Y_j(0)}_{\text{Selection Bias}}$$

Bias in estimation of average treatment effect using observed outcomes:

- Bias in estimation of ATT using "naive"estimator:

$$T_1 - C_0 = \underbrace{T_1 - T_0}_{ATT} + \underbrace{T_0 - C_0}_{\text{bias}}$$

- Bias in estimation of ATU using “naive”estimator:

$$T_1 - C_0 = \underbrace{C_1 - C_0}_{ATU} + \underbrace{T_1 - C_1}_{\text{bias}}$$

- Bias in estimation of ATE using “naive”estimator:

$$T_1 - C_0 = \underbrace{\omega(T_1 - T_0) + (1-\omega)(C_1 - C_0)}_{ATE} + \underbrace{\omega(T_0 - C_0) + (1-\omega)(T_1 - C_1)}_{\text{bias}}$$

Properties of assignment mechanism

1. **Individualistic** assignment: It limits the dependence of the treatment assignment for unit i on the outcomes and assignments for other units.
2. **Probabilistic** assignment: Every unit has positive probability of being assigned to the treatment or the control group.
3. **Unconfounded** assignment: The assignment mechanism does not depend on the potential outcomes.

Under these assumptions, we can use the naive estimator to estimate treatment effects.

RDD

Identification at cutoff

Formally, RD refers to a regression model

$$\mathrm{E}\left[Y|S\right] = \beta_d \mathrm{E}\left[D|S\right] + m\left(S\right)$$

where $\mathrm{E}\left[D|S\right]$ is discontinuous at 0, and $m\left(S\right)$ is an unknown function continuous at 0.

$$\beta_d = \frac{\mathrm{E}\left[Y|0^+\right] - \mathrm{E}\left[Y|0^-\right]}{\mathrm{E}\left[D|0^+\right] - \mathrm{E}\left[D|0^-\right]}$$

The RD identification condition has two components:

- The discontinuity of $\mathrm{E}\left[D|S\right]$: $\mathrm{E}\left[D|S\right]$ has a break at 0, that is, $\mathrm{E}\left[D|0^+\right] - \mathrm{E}\left[D|0^-\right] \neq 0$.
- The continuity of $m\left(S\right)$ at 0: $m\left(S\right) = \mathrm{E}\left[Y|S\right] - \beta_d \mathrm{E}\left[D|S\right]$ is continuous at 0.

Main features of RD

In RD, the discontinuity of $\mathrm{E}\left[D|S\right]$ at 0 comes typically from an institutional feature and thus it is visible, but the continuity of $m\left(S\right)$ is an assumption.

This continuity leads to three main features of RD:

- **Local randomization**.
  - In SRD with $D = 1[S \geq 0]$
- **Robustness of RD to the endogeneity of D through S.**
  - Data generating process: $Y_i = \beta_d D_i + m(S_i) + U_i$, where $\mathrm{E}[U|S] = 0$.
  - RD allows $\mathrm{E}[U|S]$ to be a nontrivial function of S, as long as $\mathrm{E}[U|S]$ is continuous at 0; such a $\mathrm{E}[U|S]$ can be merged into $m(S)$.
  - The capacity to allow $\mathrm{E}[U|S]$ to be a function of S makes RD estimators robust to the endogeneity of D through S.
  - D in FRD affected by both S and ε can be still related to U through ε, and RD estimators are not robust to the D endogeneity through ε.
- **Ignorability of covariates other than S.**
  - As long as the covariates are continuous at the cutoff point, the covariates do not affect the procedure of identification.

Data requirement of RDD

1. Three basic variables:
   - Assignment variable (or called forcing variable, running variable): S
   - Cutoff point: C
   - Outcome variable: Y
2. Values of the assignment variable cannot be manipulated around the cutoff point (to reach local randomization).
3. Cutoff point is not affected by the assignment variable (one cannot choose a cutoff point to change the assignment).
4. Individual characteristics are unaffected around the cutoff point, other than the assignment status.

Specification tests of RDD

1. The breaks of $\mathrm{E}[Y|S]$ and $\mathrm{E}[D|S]$.
2. The continuity of $m(S)$.

Challenges to RDD

- Treatment is not as good as randomly assigned around the cutoff when agents are able to manipulate their running variable scores.
- Some other unobservable characteristic changes at the threshold, and this has a direct effect on the outcome.