Microeconometrics Module

Lecture 7: Matching

Swapnil Singh

Lietuvos Bankas and KTU
Course Link

The Essence of Matching

Core Idea: When conducting causal inference, the conditional independence assumption holds that treatments are as random, dependent only on known covariates.

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

This assumption leads us to trust that matching estimators can accurately calculate treatment effects by conditioning on covariates.

Theoretical Basis:

$$\tau(x) = \mathbb{E}[Y_{1i} - Y_{0i} \mid X_i = x]$$

This equation illustrates how the conditional independence assumption underpins causal estimation within defined groups.

Goals of Matching

The fundamental problem of causal inference:

1. Discover the individual treatment effect, $\tau_i = Y_{1i} - Y_{0i}$.
2. Overcome the limitation of not observing both outcomes simultaneously.

Solution through Matching: *Match each treated observation with an untreated counterpart based on similarity in covariates, allowing estimation of the unobserved counterfactual outcome.*

A Formal Approach to Matching

Objective: Construct a valid counterfactual for each treated individual using matching techniques.

Strategy:

- Establish a set of weights for each treated individual, corresponding to matched controls.
- Ensure that these weights sum to one for each treated individual.

Notation: N_T and N_C treated and control individuals/units. For each i treated unit, associated N_C weights given as $(w_i(1), \cdots, w_i(N_C))$ such that $\sum_{j=1}^{N_C} w_i(j) = 1$.

Calculation of counterfactual:

$$\widehat{Y_{0i}} = \sum_{j \in (D=0)} w_i(j) Y_j$$

This formula calculates the estimated outcome for treated individuals, relying on weighted averages of control outcomes.

A Formal Approach to Matching

- Note that we have counterfactual for each treated individual/unit
- Treatment effect for individual i

$$\widehat{\tau}_i = \underbrace{Y_{1i}}_{observed} - \underbrace{\widehat{Y_{0i}}}_{constructed}$$

- We can write the matching estimator as

$$\widehat{\tau_m} = \frac{1}{N_T} \sum_{i \in (D=1)} \widehat{\tau}_i$$

A Formal Approach to Matching

- Note that we have counterfactual for each treated individual/unit
- Treatment effect for individual i

$$\widehat{\tau}_i = \underbrace{Y_{1i}}_{observed} - \underbrace{\widehat{Y_{0i}}}_{constructed}$$

- We can write the matching estimator as

$$\widehat{\tau_m} = \frac{1}{N_T} \sum_{i \in (D=1)} \widehat{\tau}_i$$

- Easy, huh! But wait. We have something missing here.

A Formal Approach to Matching

- Note that we have counterfactual for each treated individual/unit

- Treatment effect for individual i

$$\widehat{\tau}_i = \underbrace{Y_{1i}}_{observed} - \underbrace{\widehat{Y_{0i}}}_{constructed}$$

- We can write the matching estimator as

$$\widehat{\tau_m} = \frac{1}{N_T} \sum_{i \in (D=1)} \widehat{\tau}_i$$

- Easy, huh! But wait. We have something missing here.

- Weights i.e. $w_i(j)$'s. We don't have them and we need to calculate them

Matching Weights

Question: How do we choose/estimate weights?

Bad idea: $w_i(j) = \frac{1}{N_C}$. Why?

Matching Weights

Question: How do we choose/estimate weights?

Bad idea: $w_i(j) = \frac{1}{N_C}$. Why?

Answer: essentially you are calculating difference in means, discarding X and letting go of conditional independence assumption

Bad idea leading to good idea: use information contained in X to construct weights. Further, weights should signify how close X_i i.e. treated individuals covariate is close to X_j i.e. control group individual's covariate

Key takeaway: weights should tell us how close a treated unit is to control unit.

- Too low weight $\Rightarrow$ treated unit is far from control unit
- High weight $\Rightarrow$ treated unit is close to control unit

Matching Weights: Discrete Covariates

- If covariates are discrete, set weights equal to one if treatment and control group covariates are same, zero otherwise
- You will need to do scaling so that weights sum up to one

Example

Let's say there are two covariates: gender (male, female) and college education (yes, no). Then there will be four groups of covariates: (male, yes), (male, no), (female, yes), (female, no)

Question: How many groups will be created if there are 10 discrete covariates?

Matching Weights: Continuous Covariates

- With continuous variables, it is a folly to assume that two individuals will have same value of covariates
- So, we go with the notion of how near covariate values are
- Need to define distance between control and treatment units based on the value of their covariates
- Two examples of distances:
  1. Euclidean
  2. Mahalanobis

$$Euclidean: \quad d_{i,j} = (X_i - X_j)'(X_i - X_j)$$

$$Mahalanobis: \quad d_{i,j} = (X_i - X_j)'\Sigma_X^{-1}(X_i - X_j)$$

- For each treatment unit, pick a control unit which is closest to it in terms of distance

Note: We will get causal estimate under two conditions:
1. CIA holds
2. There is a common support (more on this in practical example)

Further Problems/Issues

Problem/Issue 1:

- We are pick one nearest neighbor and letting go of lot of information contained in other control units
- You solve this issue by using Kernel matching. What it does?
  - Give weights to each control unit
  - Closer is the control unit, larger is the weight

Problem/Issue 2:

- What if there are too many covariates $\Rightarrow$ curse of dimensionality
- Solution lies in propensity score matching
- Focus of the remaining lecture

Propensity Score Method: Theoretical Foundation

Propensity score theorem.

If $(Y_{0i}, Y_{1i}) \perp\!\!\!\perp D_i | X_i$, i.e. CIA holds, then $(Y_{0i}, Y_{1i}) \perp\!\!\!\perp D_i | p(X_i)$, where $p(X_i) = \mathbb{E}(D_i \mid X_i)$ is the propensity score, i.e. the probability of treatment given X_i.

Propensity score theorem.
If $(Y_{0i}, Y_{1i}) \perp\!\!\!\perp D_i | X_i$, i.e. CIA holds, then $(Y_{0i}, Y_{1i}) \perp\!\!\!\perp D_i | p(X_i)$, where $p(X_i) = \mathbb{E}(D_i \mid X_i)$ is the propensity score, i.e. the probability of treatment given X_i.

The theorem states that if the **Conditional Independence Assumption (CIA)** holds, meaning that the potential outcomes (Y_{0i}, Y_{1i}) are independent of the treatment assignment D_i once we control for covariates X_i, then a remarkable simplification occurs: the potential outcomes are also independent of the treatment assignment given the propensity score alone. This implies that the propensity score $p(X_i) = \mathbb{E}(D_i \mid X_i)$ effectively balances the treatment and control groups in terms of the distribution of covariates X_i.

Propensity Score Method: Theoretical Foundation

Propensity score theorem.
If $(Y_{0i}, Y_{1i}) \perp\!\!\!\perp D_i | X_i$, i.e. CIA holds, then $(Y_{0i}, Y_{1i}) \perp\!\!\!\perp D_i | p(X_i)$, where $p(X_i) = \mathbb{E}(D_i \mid X_i)$ is the propensity score, i.e. the probability of treatment given X_i.

The theorem states that if the **Conditional Independence Assumption (CIA)** holds, meaning that the potential outcomes (Y_{0i}, Y_{1i}) are independent of the treatment assignment D_i once we control for covariates X_i, then a remarkable simplification occurs: the potential outcomes are also independent of the treatment assignment given the propensity score alone. This implies that the propensity score $p(X_i) = \mathbb{E}(D_i \mid X_i)$ effectively balances the treatment and control groups in terms of the distribution of covariates X_i.

Life becomes simple if we know $p(X_i)$. How do we calculate them?

Propensity Score Estimation and Use

- We use logit method and regress treatment dummy on covariates and then predict the probability of treatment
- Once propensity score is estimated we can use it in the following way:
    - Directly use them as a covariate in regression

    $$Y_i = \alpha + \gamma p(X_i) + \beta D_i + u_i$$

    - Stratify on propensity scores (discussed next)
    - Do a matching based on propensity scores and then run regression
    - Run a simple regression but weight observations by the inverse of propensity scores (not discussed)

Stratification

1. Propensity scores are actually probability of treatment
2. Hence, they vary between 0 and 1
3. Divide the (0,1) interval, let's say in N groups
4. Within each group calculate the treatment effect based on treatment and control observations which fall within that group
5. Average out the treatment effect from each group to compute average treatment effect

Practical Example

More Experience

Follow these links:
Matching Methods
Balance Assessment
Estimating Effects After Matching