ECON 3510: Poverty and Economic Development
Lecture 10: Matching

Instructor: Weizheng Lai

Bowdoin College

Fall 2025

Conditional Independence Assumption (CIA)

- We have been using the following regression

$$Y_i = \beta D_i + \mathbf{X}_i' \boldsymbol{\gamma} + \varepsilon_i.$$

- To interpret β as a causal effect, we need to assume:
  (i) $D_i \perp\!\!\!\perp \varepsilon_i \mid \mathbf{X}_i$;
  (ii) the linear relationship between Y_i and $\mathbf{X}_i$ is true.

- (ii) is often implicitly taken as granted. But it might be a strong assumption.
  - How do we know the correct form of $\mathbf{X}_i$? Linear, quadratic, cubic, log?
  - How much should we trust that $\mathbf{X}_i$ has constant effect $\boldsymbol{\gamma}$ across i?

- Ideally, we want to exploit the CIA to estimate causal effects without a strong functional assumption.

- **Fix: matching.** Basic idea: find comparable controls ($D_i = 0$) for the treated ($D_i = 1$) based upon $\mathbf{X}_i$.

Matching Basics

- **Step 1: Decide covariates X.**
  - Guided by economic theory.

- **Step 2: Match treated and control observations with similar values of X.**
  - How similar?
    - *Exact matching:* If $\mathbf{X}$ is binary or discrete, then it is possible to match observations with the same values.
    - *Nearest matching:* If $\mathbf{X}$ is continuous (so not possible to match exactly), can match the treated to a control with the closest value of $\mathbf{X}$.
    - *Radius matching:* can also match treated i to control j, as long as $|\mathbf{X}_i - \mathbf{X}_j| < r$, where radius/caliper r is chosen by the researcher.
  - With replacement or without replacement?
    - *With replacement:* After one time of matching, the control observation goes back to the pool for the next time of matching. Thus, it's possible for a control observation to be matched for multiple treated observations.
    - Either way is fine. With replacement may be preferred in small samples.
  - Thus, matching can be one-to-one or one-to-many. It's also likely that we can't find matches for some treated.

- **Step 3: Estimate Causal Effects.**
  - $\hat{\beta}_{\text{matching}} = \frac{1}{N_p}\sum_p (Y_p^T - Y_p^C)$, where p indexes matched pairs, and N_p is the number of matched pairs.
  - Run linear regression using the matched sample: $Y_i = \alpha + \beta D_i + \varepsilon_i$.
    - Variants: (i) control for pair FEs; (ii) control for covariates.

Remarks

- Basic matching can be done in Stata by `calipmatch` and other commands.
- Matching deals with selection on observables (i.e., CIA is assumed). It can't address selection on observables.
- In fact, we can never be sure which covariates are correct ones to match on.
- The state-of-the-art implementation of matching is to use it for selecting comparable controls, and implement some quasi-experimental methods in the matched sample.
  - E.g., the parallel trends assumption for the DiD might be more plausible in a matched sample.
- Important to check the validity of matching. Are covariates you match on indeed balanced between treatment and control groups? What about untargeted covariates?

Propensity Score Matching

- *Curse of Dimensionality:* One problem with the basic matching method is that if there are many variables in $\mathbf{X}_i$, it is difficult to find matches for all treated, yielding a matches sample that is too small to be useful.
  - But for a plausible CIA argument, we do want more variables in $\mathbf{X}_i$.

- One solution: **Propensity Score Matching (PSM)**.

- Rather than matching on $\mathbf{X}_i$, it's enough to match on the scalar **propensity score**

$$p(\mathbf{X}_i) = \Pr(D_i = 1 \mid \mathbf{X}_i).$$

  - **Theorem:** $D_i \perp\!\!\!\perp \varepsilon_i \mid \mathbf{X}_i$ implies $D_i \perp\!\!\!\perp \varepsilon_i \mid p(\mathbf{X}_i)$.
  - **Key Condition ("Overlap"):** $0 < p(\mathbf{X}_i) < 1$.

Procedures

- **Step 1: Decide covariates X.**

- **Step 2: Estimate propensity score $p(\mathbf{X}_i)$.**
  - Run a Probit regression
    $$\Pr(D_i = 1 \mid \mathbf{X}_i) = \Phi(\mathbf{X}_i'\boldsymbol{\delta}),$$
    where $\Phi(\cdot)$ is the cdf of $N(0,1)$.
  - Obtain estimated propensity score
    $$\widehat{p(\mathbf{X}_i)} = \Phi(\mathbf{X}_i'\hat{\boldsymbol{\delta}}).$$

- **Step 3: Match treated and control observations based upon $\widehat{p(\mathbf{X}_i)}$.**
  - Can do nearest matching or radius matching.
  - *Blocking:* block $\widehat{p(\mathbf{X}_i)}$ into several bins; treated and control observations in the same block are matched together.

- **Step 4: Estimation.**
  - Mean difference: $\hat{\beta}_{\text{matching}} = \frac{1}{N_p}\sum_p(\bar{Y}_p^T - \bar{Y}_p^C)$.
  - Regression with the matched sample: $Y_i = \alpha + \beta D_i + \varepsilon_i$.
    - Variants: (i) control for pair/block FEs; (ii) control for covariates.

Remarks

- PSM can be done in Stata by `psmatch2`.
- Again, PSM only addresses selection on observables, not selection on unobservables.
- Pros of matching:
  - Easy to tell what comparisons are used;
  - Does not rely on strong functional form assumptions.
- Cons of matching:
  - Low statistical power: samples are smaller;
  - Data greedy.