Segment 5: Analyzing Observational Studies

Section 02: (Propensity Score) Matching Methods

Matching as "Design" or "Restructuring"

- *Exact Matching:* Match every $Z_i = 1$ unit to a $Z_j = 0$ unit with the exact same covariate values, $\mathbf{X}_i = \mathbf{X}_j$
  - Clearly infeasible for even modest dimension $\mathbf{X}$

- *Propensity Score Matching:* Match every $Z_i = 1$ unit to a $Z_j = 0$ unit with the same (or "close") value of $e_i(\mathbf{X}_i) \approx e_j(\mathbf{X}_j)$
  - Rosenbaum and Rubin (1983) says that this will ensure covariate balance among matched pairs
  - Relevant for a broader class of *balancing scores*

- Other forms of matching
  - Constrained optimization methods

Basic Setting

- Observational study with confounding $\Rightarrow$ covariate distribution among treatment groups are not the same
- Interest is in the Average Treatment Effect on the Treated (ATT)
  - $E[Y_i(1) - Y_i(0)|Z_i = 1]$
  - The average causal effect among those who received the treatment
- **Goal:** Use propensity scores to select an appropriate set of controls with $Z = 0$ to compare against units with $Z = 1$
  - Note that $Z = 0$ units could come from a different data source, as long as the same $\mathbf{X}$ are measured

Matching Problem Notation

- N units in the sample
- $\mathbb{I}_t = \{1, 2, \ldots, N_t\}$ indexes units with $Z = 1$
- $\mathbb{I}_c = \{N_{t+1}, N_{t+2}, \ldots, N_t + N_c\}$ indexes units with $Z = 0$
- $\mathcal{M}_i^c \subset \mathbb{I}_c$, set of controls matched to treated unit i
- In simplest case, $\mathcal{M}_i^c = \{m_i^c\}$
  - m_i^c indexes the unit with the closest covariate values among those with treatment different than Z_i
- Many different methods/algorithms for obtaining $\mathcal{M}_i^c$
  - Focus here is just a subset of methods that use propensity scores

Where We're Going...

After successful matching, can essentially use methods from paired randomized studies (under certain assumptions)

$$\hat{\tau}_i^{\text{match}} = Y_i^{obs} - Y_{m_i^c}^{obs}$$

is an unbiased estimator for the causal effect at $\mathbf{X} = \mathbf{x}_i$ for both units in the pair.

$$\hat{\tau}^{\text{match}} = \frac{1}{N_t} \sum_{i:Z_i=1} \hat{\tau}_i^{\text{match}} = \frac{1}{N_t} \sum_{i:Z_i=1} \left(Y_i^1 - Y_{m_i^c}^0 \right)$$

is an unbiased estimator for the average treatment effect for the ATT.

(see GHV Ch 20.8 "warning about matched pairs")

Algorithms for Inexact Matching

General Goal: Match the i^{th} treated unit to control unit m_i^c that solves:

$$m_i^c = \text{argmin}_{i' \in \mathbb{I}_c} \|\mathbf{X}_i - \mathbf{X}_{i'}\|$$

where $\|\cdot\|$ is a generic distance function.
Think: $\|\mathbf{X}_i - \mathbf{X}_{i'}\| = |e_i(\mathbf{X}_i) - e_{i'}(\mathbf{X}_{i'})|$

We will initially assume exactly one match for every treated unit

Issues:

- Distance metric?
- No "close" unit
- Multiple "close" units
- $j \in \mathbb{I}_c$ might be the best match for both i and i'
- Optimal vs. greedy algorithm

Distance Metric

- Propensity score distance
  - $\|x, x'\| = |e(x) - e(x')|$
  - $\|x, x'\| = |ln(\frac{e(x)}{(1-e(x))}) - ln(\frac{e(x')}{(1-e(x'))})|$
- Could define others not based on propensity score
  - Euclidean distance
  - Mahalanobis distance
  - ...

Optimal Algorithms

Simultaneously match all units to obtain an optimal allocation of matches for the full set $\mathbb{I}_t$. That is, find the N_t indices $m_1^c, \ldots, m_{N_t}^c$ that solve:

$$\text{argmin}_{m_i^c, \ldots, m_{N_t}^c \in \mathbb{I}_c} \sum_{i=1}^{N_t} \left\| \mathbf{X}_i - \mathbf{X}_{m_i^c} \right\|$$

subject to $m_i \neq m_{i'}$, for $i \neq i'$.

- Computationally demanding for even moderately large sample sizes
- Computational feasible variations of optimal matching methods

Greedy Algorithms

Sequentially obtain matches for the members of $\mathbb{I}_t$. For treated unit $i = 1$:

$$m_1^c = \text{argmin}_{m_1^c \in \mathbb{I}_c} \left\| \mathbf{X}_1 - \mathbf{X}_{m_1^c} \right\|$$

For treated unit $i = 2$:

$$m_2^c = \text{argmin}_{i' \in \mathbb{I}_c - \mathcal{M}_1^c} \left\| \mathbf{X}_2 - \mathbf{X}_{i'} \right\|$$

...

$$m_i^c = \text{argmin}_{i' \in \mathbb{I}_c - \cup_{j=1}^{j-1} \mathcal{M}_j^c} \left\| \mathbf{X}_i - \mathbf{X}_{i'} \right\|$$

- Match each unit to the closest unit that hasn't already been matched
- Order matters!

Tradeoffs: Lots of Matches vs. Good Matches

General Tradeoff: quality of matches vs. number of matched units

- "Good matches" are very close in $\mathbf{X}$
  - Closer in $\mathbf{X}$ (or in $e(\mathbf{X})$) $\rightarrow$ less bias
  - Does each treated unit have a close match in the data?
  - Implications for bias (closer matches $\rightarrow$ less bias)

- More matches $\rightarrow$ improves precision
  - More treated units with good matches
  - More matches for each treated unit
  - May have to settle for less good matches if we want more of them

Matching with Replacement

Key idea: Allow a control unit to be "the match" for > 1 treated unit.

Advantages:

- Ease computational burden
  - No tradeoff between "optimal" and "greedy"
  - Finding optimal set is straightforward
- May reduce bias
  - Less discrepancy between permitted matches

Disadvantages:

- Larger sampling variance
  - Estimator based on fewer controls
- More difficult variance estimation
  - Correlation across pairs

Matching with Replacement

Solve minimization problem (and find the optimal match) for each treated unit:

$$m_i^c = \text{argmin}_{i' \in \mathbb{I}_c} \|\mathbf{X}_i - \mathbf{X}_{i'}\|$$

(does not depend on ordering)

$$\hat{\tau}_t^{\text{repl}} = \frac{1}{N_t} \sum_{i=1}^{N} \left(Z_i Y_i^{obs} - (1 - Z_i) L(i) Y_i^{obs} \right)$$

where $L(i) = \sum_{j=1}^{N} \mathbf{1}_{j \in \mathcal{M}_i^c}$, the number of times each control unit is used as a match

Estimator is a weighted average of treated and control outcomes within the *full* sample where

- $Z_i = 1$ observations receive weight $1/N_t$
- $Z_i = 0$ observations receive varying weights (that sum to 1)

The Number of Matches

Rather than match each treated unit to a *single* unit (1:1 matching), possible to use **multiple matches** (1:k matching)

- Especially useful when pool of possible controls is large
- Could improve precision of the resulting estimator
  - Although this gain can be somewhat limited
- Could increase bias by including poorer matches

$$\hat{\tau}_t^{\text{match},M} = \frac{1}{N_t} \sum_{i=1}^{N_t} \left(Y_i^1 - \frac{1}{k} \sum_{j \in \mathcal{M}^c(i)} Y_j^0 \right)$$

where k is the number of control matches

Other Estimands

Methods presented here estimate the ATT (Average Treatment Effect in the Treated)

- The "default" estimand of matching methods is the Average Treatment Effect in the Treated (ATT)
  - Because we seek ≥ 1 match for each treated unit
  - (many untreated units may be left unmatched)
- Could "reverse" things and estimate the ATC (Average Treatment Effect in the Controls)
  - Seek ≥ 1 match for each control unit
  - (many treated units may be left unmatched)
- Could combine ATC and ATT to esitmate ATE in the entire sample
- Overlap issues $\rightarrow$ other estimands
  - E.g., poor overlap $\Rightarrow$ some treated units don't have a good match $\Rightarrow$ confine inference to subset of units that have good matches

Some Other Options for PS Matching

- 1:1 vs. 1:k vs. ratio matching
  - Match (multiple) treated to (multiple) controls with different ratios
- Calipers
  - Include all matches within a certain distance
  - Exclude matches beyond a certain distance
- Hybrid matching based on $\hat{e}(\mathbf{X}_i)$ and individual components of $\mathbf{X}_i$
- Stuart (2010) paper is a key review paper for different matching methods

Estimation of the Propensity Score

Implementing any of these matching procedures using the propensity score requires that we first *estimate* the propensity score!

- Essentially just binary regression for $E[Z|\mathbf{X}] = Pr(Z = 1|\mathbf{X})$
  - Many different strategies for this
- Objective is slightly different from "standard" prediction objectives
  - E.g., less concerned about *multicollinearity* of predictors
  - Less concerned about overfitting
- Not trying to predict $Z = 1$ as best as possible
  - Perfect prediction $\Leftrightarrow$ no overlap!
  - Perfect randomization $\Leftrightarrow$ poor predictive power
  - Really just trying to predict $Z = 1$ as a function of the confounders $\rightarrow$ balance
  - Very strong predictors of $Z = 1$ should be avoided (if they are not also confounders that predict Y)

Example: Child Care

Gelman, Hill, Vehtari textbook goes through 5 steps for restructuring with propensity sores in the context of an example evaluating a child care program

- Worth going through on your own!
- Illustrates using the `library(Matching)` R package to construct the matches
- Other good options in R too
  - `library(MatchIt)`
  - `library(optmatch)`
  - ...