IDS 702: MODULE 6.5

STRATIFICATION AND MATCHING

DR. OLANREWAJU MICHAEL AKANDE

BALANCING COVARIATES: SMALL NUMBER OF COVARIATES

- When the number of covariates is small, the adjustment needed to get some balance can be achieved by matching or stratification.
- Exact matching: for each treated subject, get one control with the exact same value of the covariates (easier for categorical covariates).
- Exact matching ensures distributions of covariates in treatment and control groups are exactly the same, thus eliminating bias due to difference in X.
- After matching, compute treatment effect by using the matched data.
- However, exact matching is usually unfeasible, even with low dimensional covariates.

MATCHING

- Matching estimators impute the missing potential outcomes, using only the outcomes of nearest neighbors of the opposite treatment group.
- They have often (but not exclusively) been applied in settings where
    - the interest is in the ATT; and
    - there is a large reservoir of potential controls. This allows matching each treated unit to one or more distinct controls (nearest neighbors).
- More general settings: both treated and control units are (potentially) matched and matching is done with replacement.

MATCHING (FIXED NUMBER OF MATCHES)

- Let $\mathcal{M}_i$ be the set of the indices of M closest matches of unit i using a distance metric that depends on X.
- Let

$$\hat{Y}_i(0) = \sum_{i \in \mathcal{M}_i} \frac{Y_j}{M} \quad \text{if} \quad W_i = 1 \quad \text{and} \quad \hat{Y}_i(0) = Y_i^{\text{obs}} \quad \text{if} \quad W_i = 0;$$

$$\hat{Y}_i(1) = Y_i^{\text{obs}} \quad \text{if} \quad W_i = 1 \quad \text{and} \quad \hat{Y}_i(1) = \sum_{i \in \mathcal{M}_i} \frac{Y_j}{M} \quad \text{if} \quad W_i = 0.$$

- Then, the treatment effect within a pair is estimated as the difference in outcomes, and we can average these within-pair differences.
- That is,

$$\hat{\tau}^{\text{ATE}} = \sum_i \frac{\hat{Y}_i(1) - \hat{Y}_i(0)}{N};$$

$$\hat{\tau}^{\text{ATT}} = \sum_i \frac{\left(Y_i - \hat{Y}_i(0)\right) W_i}{N_t}.$$

MATCHING (FIXED NUMBER OF MATCHES)

- Pros: Matching estimators that ensure good balance in covariates between groups are generally robust.
- Cons: With fixed number of matches and matching with replacement, matching estimators can be biased.
- Matching estimators are generally not efficient.
- In fact, estimators combining matching and regression adjustment are usually more efficient.
- There can be residual imbalance in matching.
- Perform bias correction via regression on the matched sample.

MATCHING: TUNING

- Matching involves lots of tuning
    - distance metric
    - fixed or varying number of matches
    - for fixed M, number of matches
    - with or without replacement
- Tuning for matching is an art, with some theory and general guidelines available...

MATCHING: TUNING

- Distance metric: Mahalanobis distance, propensity score, tree-based.
- Fixed M or varying M? For varying M:
  - Matching with caliper: define a caliper (say 0.1) and all units within that caliper are matches
  - M increases with sample size.
- For fixed M, the choice of M (number of matches per unit) has a bias-variance trade-off:
  - smaller $M \Rightarrow$ smaller bias but larger variance
  - larger $M \Rightarrow$, larger bias but smaller variance.
- Also depends on the proportion of treatment versus control: when there is a much larger control group, we can use one-to-many matching.

MATCHING: TUNING

- Matching with replacement:
    - Pros:
        1. computationally easier
        2. both controls and treated can be matched, but with high variances
        3. not order-dependent
    - Cons: some units (especially ones with extreme propensity scores) can be matched many times and thus heavily influence overall estimates.
- What about matching ties? What should we do about them?
- Matching is a vast topic and there are so many matching methods.
- Implementation in R: `Matchit`, `Matching`, and many more.

STRATIFICATION

- Another option is stratification.
- Suppose we have a single covariate X with k levels (e.g. race).
- We will continue to assume unconfoundedness and overlap holds.
- Suppose we want to estimate ATE.
- Let
    - n_k be the number of observations with $X_i = k$; and
    - $\bar{Y}_{k,w}$ be the sample average of all Y_i values among observations in cell $X_i = k$ and $W_i = w$.
- Once again, recall that ATE is $\tau = \mathbb{E}[Y_i(1) - Y_i(0)]$.

STRATIFICATION

- Then we have

$$\mathbb{E}[Y_i(1)] = \sum_k \mathbb{E}[Y_i | X_i = k, W_i = 1] \cdot \mathbb{P}\mathrm{r}[X_i = k],$$

and

$$\mathbb{E}[Y_i(0)] = \sum_k \mathbb{E}[Y_i | X_i = k, W_i = 0] \cdot \mathbb{P}\mathrm{r}[X_i = k].$$

- We can estimate $\mathbb{E}[Y_i(1)]$ using a consistent estimator $\sum_k \bar{Y}_{k,1} \dfrac{n_k}{N}$. We can use a similar estimand for $\mathbb{E}[Y_i(0)]$.

- Therefore, the ATE τ can be estimated by

$$\hat{\tau} = \sum_k \left(\bar{Y}_{k,1} - \bar{Y}_{k,0}\right) \frac{n_k}{N}.$$

IDS 702

STRATIFICATION

- What if X is continuous?
- Stratification (subclassification): split X into k classes.
- Then, for class k, define n_k and $\bar{Y}_{k,w}$ as before.
- An estimator of τ is then once again

$$\hat{\tau}^k = \sum_k \left(\bar{Y}_{k,1} - \bar{Y}_{k,0}\right) \frac{n_k}{N}.$$

- $\hat{\tau}^k$ is generally biased for τ, however, stratification of over 5 blocks can remove 90% of the bias!
- **Overall, the key idea with stratification is this**: even though we may not have balance across the entire sample, we likely can get balance by focusing on subgroups, one at a time.

BALANCING COVARIATES: LARGE NUMBER OF COVARIATES

- What if we have a large number of covariates?
- With just 20 binary covariates, there are 2^{20} or about a million covariate patterns!
- Direct matching (exact of nearest neighbors) or stratification is nearly impossible.
- Need dimensional reduction to a single score which we can then use to match or stratify.
- The most popular option is the propensity score: $e(x) = \mathbb{P}\mathrm{r}[W_i = 1 | X_i = x]$.
- We will focus on propensity score methods over the next few modules and use them to analyze the minimum wage data.

ACKNOWLEDGEMENTS

These slides contain materials adapted from courses taught by Dr. Fan Li.

What's next?

Move on to the readings for the next module!