

Matching Techniques

Enrijeta Shino

PhD Candidate
Department of Political Science
University of Florida

The Election Science Group Presentation

October 8, 2018

- 1 What is matching?
- 2 Why use matching?
- 3 Main assumptions
- 4 Similarities between matching and randomized experiments
- 5 Matching algorithms
- 6 Matching techniques
- 7 Implementation of matching methods in practice
- 8 Sensitivity analysis
- 9 Pros and Cons of matching
- 10 Example: CEM (R-code)

Definition: Matching

Matching methods constitute a growing collection of techniques that attempt to replicate, as closely as possible, the ideal of randomized experiments when using observational data (Stuart & Rubin 2007, 155).

- Many studies in social sciences aim to estimate the effects of a treatment on a subpopulation.
 - **Problem:** Treatment selection bias.
 - Where the units who receive the treatment may have different characteristics from those in the control condition.

Intervention group



Control group



Objective

- The main objective of matching is to **reduce bias**.
 - **Bias** is the tendency of a measurement process to over- or under-estimate the value of a population parameter.

Bias-Variance Trade-off

- Discarding units in the matching process will result in smaller sample sizes.
 - Might appear to lead to increases in sampling variance, this is not always the case because improved balance in the covariate distributions will decrease the variance of estimators (Snedecor & Cochran, 1980)

1 Strong Ignorability of Treatment.

- Assumes that the treatment assignment is independent of the potential outcome distributions, given observed covariates X :

$$(Y_i(1), Y_i(0)) \perp\!\!\!\perp W_i | X_i$$

(Rosenbaum & Rubin 1983)

2 Common Support

- A valid counterfactual must exist

3 Stable Unit Treatment Value Assumption (SUTVA)

1 Stable Treatment

- There is **only a single** version of the treatment for each unit.

2 Non-Interference

- One observation's treatment level does not affect another's outcome. "**No Spillover**".

Two Ways:

- 1 Matching aims to identify subsamples of treated and control units that are “balanced” with respect to observed covariates.
- 2 Study Design: Two-stages
 - 1 The units to be compared are selected, without use of the values of the outcome variables.
 - 2 Analyses of the outcome, estimating treatment effects using the matched sample.

1 Greedy matching (kNN)

- Cycles through the treated units one at a time, selecting for each the available control unit with the smallest distance to the treated unit.
- Seeks to minimize the distance between each pair, but does not minimize the total distance between all matched pairs.

2 Optimal matching (k2k)

- Used for one-to-one matching or full matching.
- Attempts to match all untreated individuals in the dataset to a treated counterpart, resulting in no loss of sample size as long as there is an adequate area of common support.
- Minimizes a global measure of balance.

3 Full Matching

- Creates strata where each stratum contains at least one treated individual and at least one untreated individual, minimizes both the within-strata and between-strata propensity score distances.

kNN Simple Algorithm Steps

- 1 Determine parameter K (K denotes the number of nearest neighbors)
- 2 Calculate the distance between observations in the treatment and control group for each observed covariate.
- 3 Sort the distance and determine the nearest neighbors based on the K-th minimum distance.
- 4 Gather the outcome variable of the nearest neighbors.
- 5 Compute **Average Treatment Effect (ATE)**:
 - The difference between the outcomes of the individuals in the treated and untreated groups.

$$ATE = \frac{1}{N} \sum_{i=1}^N \left\{ Y_i(1) - \frac{1}{|N_i|} \sum_{j \in N_i} Y_j(0) \right\}$$

Propensity Score Matching

- Finding close matches on each covariate is difficult.
- Rather than attempting to match on all of the covariates individually, propensity score matching matches on the **scalar** propensity score.
- The **propensity score** for unit i is defined as the probability of receiving the treatment given the observed covariates:

$$e_i(X) = P(W_i = 1|X)$$

- 1 **Matching with replacement.**
 - Selects one or multiple matches for each case (depending on the desired matching ratio), and then returns the matched cases to the pool of observations.
- 2 **Matching without replacement.**
 - Each observation can be used as a match only once.
- 3 **One-to-One matching.**
 - Each treated unit is matched to one control unit.
 - Sample reduction.
- 4 **Nearest Neighbor matching**
 - Consists of simply finding the untreated observation with closest propensity score to the propensity score of each treated observation.
- 5 **Within-Caliper matching**
 - **Caliper:** maximum distance within which matches are allowed.
 - The closest untreated unit to a treated unit is only considered an acceptable match if its propensity score lies within the caliper of the treated unit.

Implementation Steps

- 1 Choose the covariates to be used in the matching process.
- 2 Define a distance measure, used to assess whether units are “similar”.
- 3 Choose a specific matching algorithm to form matched sets.
- 4 Diagnose the matches obtained (and iterating between (2) and (3)).
- 5 Estimate the effect of the treatment on the outcome, using the matched sets found in (4) and possibly other adjustments.

- **Note:**

- No guidance on how many matches/caliper size
- No guidance on how much balance is “good enough”

- **Objective:** How large would an unobserved confounder have to be to switch the results?
 - Determines how strong the effect of an omitted covariate would have to be for the significance test of the treatment effect to change.
- Establishes the degree of **robustness of treatment effects** to hidden bias, which is the part of the selection bias due to **omitted confounders**.
- **Important** because propensity score methods only controls for selection bias due to observed confounders.
 - The *strong ignorability of treatment assignment* assumption is only strictly met if there are no omitted confounders.

Pros and Cons of Matching

Pros

- 1 Matching allows for the estimation of causal effects using observational data.
- 2 Matching allows the researcher to balance two problems that plague statistical estimation: bias and variance.
- 3 It's cheaper than running experiments.

Cons

- 1 Matching can be computationally intensive.
- 2 Matching relies on strong assumptions about the exogeneity of the treatment, which might make results less reliable.
- 3 Matching requires decisions at several steps of the process that may bias the estimates and limit their precision.
- 4 We only observe *one potential outcome* for each observation
- 5 Consequently: We do not observe the same observation both getting the treatment and not getting the treatment

Coarsened Exact Matching (CEM) (Iacus, King, & Porro 2009)

What is CEM?

- CEM is a **monotonic imbalance-reducing** matching method.
 - Reducing the maximum imbalance on one variable has no effect on others.
 - The balance between the treated and the control groups is chosen by ex-ante user choice.
- Coarsen the data, exact match on these coarsened data, and then run the analysis on the uncoarsened, matched data.
- Estimates SATT (sample average treatment of the treated).
- Coarsened exact matching is faster, is easier to use and understand.
- Example: Georgia Exact-Matching (R-code for CEM)

- 1 Stuart, Elizabeth A., and Donald B. Rubin. 2008. "Best Practices in Quasi-Experimental Designs." *Best Practices in Quantitative Methods*: 155-176***.
- 2 Iacus, Stefano, Gary King, and Giuseppe Porro. 2009. "CEM: Software for Coarsened Exact Matching." *Journal of Statistical Software* 30(13): 1-27.
- 3 Imbens, Guido W. 2004. "Nonparametric Estimation of Average Treatment Effects Under Exogeneity: A Review." *Review of Economics and Statistics* 1: 4-29.
- 4 Rosenbaum, Paul R., and Donald B. Rubin. 1983. "The Central Role of the Propensity Score in Observational Studies for Causal Effects." *Biometrika* 70(1): 41-55.
- 5 Rubin, Donald B. 1973. "Matching to Remove Bias in Observational Studies." *Biometrics*: 159-183.
- 6 Rubin, Donald B. 1980. "Bias Reduction Using Mahalanobis-Metric Matching." *Biometrics*: 293-298.