Matching Methods

Kosuke Imai

Harvard University

STAT186/GOV2002 CAUSAL INFERENCE

Fall 2018

Motivation

- Causal inference $\rightsquigarrow$ inference for counterfactuals
- Comparison between treated and control units
- Consider the Average Treatment Effect for the Treated (ATT):

$$\tau_{\text{ATT}} \;=\; \mathbb{E}(Y_i(1) - Y_i(0) \mid T_i = 1)$$

- Regression $\rightsquigarrow$ model-based imputation

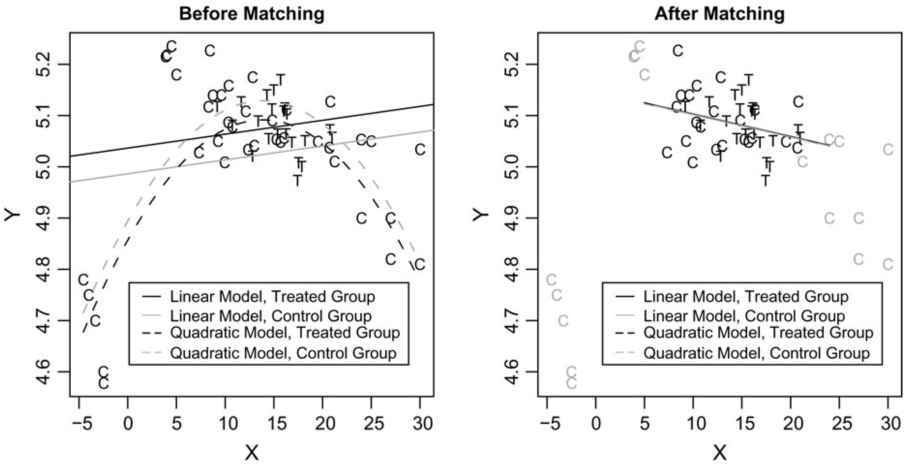
$$\hat{\tau}_{\text{reg}} \;=\; \frac{1}{n_1}\sum_{i=1}^{n} T_i \left(Y_i - \widehat{Y_i(0)}\right)$$

- Regression can be model-dependent
- Matching $\rightsquigarrow$ nonparametric imputation:

$$\hat{\tau}_{\text{match}} \;=\; \frac{1}{n_1}\sum_{i=1}^{n} T_i \left(Y_i - \frac{1}{|\mathcal{M}_i|}\sum_{i' \in \mathcal{M}_i} Y_{i'}\right)$$

Matching as Nonparametric Preprocessing

(Ho, et al. 2007. *Political Anal.*)



Sources of Bias for Regression Estimators

- Assumptions
  1. Overlap: $0 < \Pr(T_i = 1 \mid \mathbf{X}_i = \mathbf{x}) < 1$ for any $\mathbf{x}$
  2. Ignorability: $\{Y_i(1), Y_i(0)\} \perp\!\!\!\perp T_i \mid \mathbf{X}_i = \mathbf{x}$ for any $\mathbf{x}$
- Bias decomposition (Heckman et al. 1998. *Econometrica*):

$$
\begin{aligned}
& \mathbb{E}(Y_i(0) \mid T_i = 1) - \mathbb{E}\{\mu(0, \mathbf{X}_i)\} \\
= & \underbrace{\int_{S_1 \setminus S} \mathbb{E}(Y_i(0) \mid T_i = 1, \mathbf{X}_i = \mathbf{x}) dF_{\mathbf{X}_i \mid T_i = 1}(\mathbf{x}) - \int_{S_0 \setminus S} \mathbb{E}(Y_i(0) \mid T_i = 0, \mathbf{X}_i = \mathbf{x}) dF_{\mathbf{X}_i \mid T_i = 0}(\mathbf{x})}_{\text{bias due to lack of common support}} \\
& + \underbrace{\int_S \mathbb{E}(Y_i(0) \mid T_i = 0, \mathbf{X}_i = \mathbf{x}) d\{F_{\mathbf{X}_i \mid T_i = 1}(\mathbf{x}) - F_{\mathbf{X}_i \mid T_i = 0}(\mathbf{x})\}}_{\text{bias due to imbalance of observables}} \\
& + \underbrace{\int_S \{\mathbb{E}(Y_i(0) \mid T_i = 1, \mathbf{X}_i = \mathbf{x}) - \mathbb{E}(Y_i(0) \mid T_i = 0, \mathbf{X}_i = \mathbf{x})\} dF_{\mathbf{X}_i \mid T_i = 1}(\mathbf{x})}_{\text{bias due to unobservables}}
\end{aligned}
$$

Exact and Coarsened Exact Matching

- Exact Matching $\rightsquigarrow$ perfect covariate balance:

$$\widetilde{F}(\mathbf{X} \mid T_i = 1) \ = \ \widetilde{F}(\mathbf{X} \mid T_i = 0)$$

- No model dependence
- infeasible when
  - covariate is continuous
  - there are many covariates
- Coarsened Exact Matching (CEM) (Iacus et al. 2011 *Political Anal.*)
  - discretize covariates so that you can match
  - many covariates are discrete
  - discrete categories have substantive meanings
  - accounts for all interactions among coarsened variables
  - some treated units may have no matched controls (lack of overlap) $\rightsquigarrow$ changes estimand
  - bias variance tradeoff

Univariate Distance Measures

1. Mahalanobis distance:

$$D(\mathbf{X}_i, \mathbf{X}_j) \;=\; \sqrt{(\mathbf{X}_i - \mathbf{X}_j)^\top \widetilde{\Sigma}^{-1} (\mathbf{X}_i - \mathbf{X}_j)}$$

2. (Estimated) Propensity score distance:

$$D(\mathbf{X}_i, \mathbf{X}_j) \;=\; |\widehat{\pi(\mathbf{X}_i)} - \widehat{\pi(\mathbf{X}_j)}| \;=\; |\Pr(\widehat{T_i = 1} \mid \mathbf{X}_i) - \Pr(\widehat{T_j = 1} \mid \mathbf{X}_j)|$$

   or often with the linear predictor of logistic regression

$$D(\mathbf{X}_i, \mathbf{X}_j) \;=\; |\mathrm{logit}(\widehat{\pi(\mathbf{X}_i)}) - \mathrm{logit}(\widehat{\pi(\mathbf{X}_j)})|$$

- Classical matching techniques (Rubin. 2006. *Matched Sampling for Causal Effects*. Cambridge UP):
  - one-to-one, one-to-many
  - caliper
  - with and without replacement
  - subclassification instead of matching

Propensity Score as a Balancing Score (Rosenbaum and Rubin. 1983. *Biometrika*)

- Probability of receiving the treatment:

$$\pi(\mathbf{X}_i) \quad = \quad \Pr(T_i = 1 \mid \mathbf{X}_i)$$

- Balancing property:

$$T_i \quad \perp\!\!\!\perp \quad \mathbf{X}_i \mid \pi(\mathbf{X}_i)$$

- Exogeneity given the propensity score (under exogeneity given covariates):

$$(Y_i(1), Y_i(0)) \quad \perp\!\!\!\perp \quad T_i \quad \mid \quad \pi(\mathbf{X}_i)$$

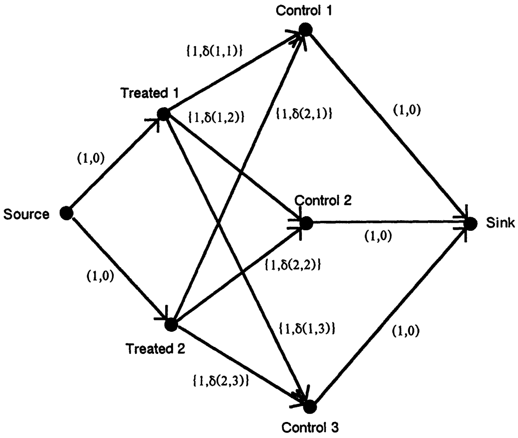
- Dimension reduction $\rightsquigarrow$ propensity score matching
- But, true propensity score is unknown: propensity score tautology

Optimal Matching (Rosenbaum. 1991. *J. Am. Stat. Assoc*)

- Greedy matching $\rightsquigarrow$ results depend on the order of matching
- Nearest neighbor matching $\rightsquigarrow$ not optimal
- Optimal matching to minimize the sum of pairwise distances
- Minimum cost flow
  - source $u_0 \longrightarrow$ sink v_0
  - edge $e(u_i, v_j)$
  - capacity $c(u_i, v_j) > 0$
  - flow $f(u_i, v_j) \geq 0$
  - cost $\delta(u_i, v_i)$
- minimize

$$\sum_{e(u_i, v_i) \in E} f(u_i, v_j) \cdot \delta(u_i, v_i)$$

  - all capacities $= 1$
  - $f(u_i, v_i) \leq c(u_i, v_i)$
  - $f(s, u_i) = 1$
  - $\sum_{j=1}^{n_0} f(u_i, v_j) = 1$
  - $\sum_{i=1}^{n_1} f(u_i, v_j) = f(v_j, t)$



Checking Covariate Balance

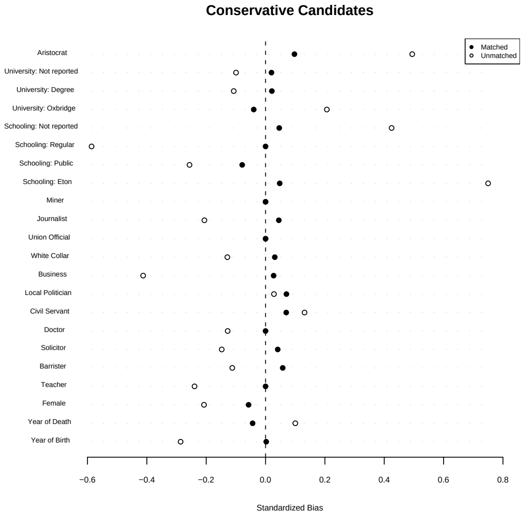
- Success of matching method depends on the resulting balance
  - Ideally, compare the joint distribution of all covariates
  - In practice, check lower-dimensional summaries (e.g., standardized mean difference, variance ratio, empirical CDF)

$$\frac{\frac{1}{n_1}\sum_{i=1}^{n} T_i \left(X_{ij} - \frac{1}{|\mathcal{M}_i|}\sum_{i' \in \mathcal{M}_i} X_{i'j}\right)}{\sqrt{\frac{1}{n_1-1}\sum_{i=1}^{n} T_i (X_{ij} - \overline{X}_{j1})^2}}$$

- Frequent use of balance test
  - failure to reject the null $\neq$ covariate balance
  - problematic especially because matching reduces the number of observations

- Estimating the financial benefits of political office

Figure 3: Covariate Balance Before and After Matching



Equivalence Tests (Wellek. 2010. *Testing Statistical Hypotheses of Equivalence and Noninferiority.* Chapman & Hall.)

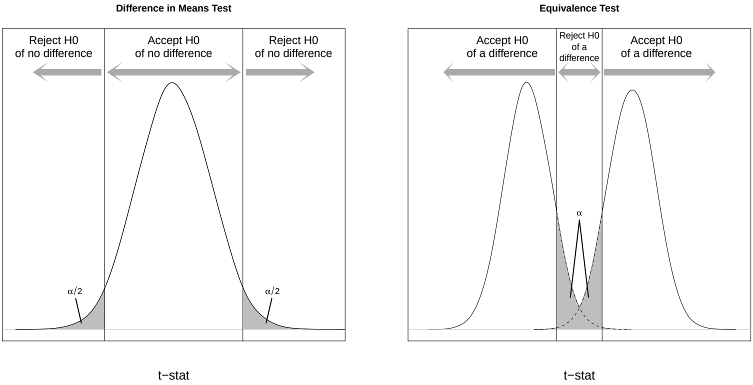
- Null hypothesis of usual balance tests: treatment and control groups are the same
- Problem: failure to reject the null $\neq$ the null is correct
- Shift the burden of proof $\rightsquigarrow$ reject the null hypothesis that treatment and control groups are different
- Use of Equivalence tests (Hartman and Hidalgo. In-press. *Am. J. Political Sci.*):

$$H_0 : \ |\tau| \geq \ \Delta \quad \text{and} \quad H_1 : \ |\tau| \ < \ \Delta$$

  for a pre-selected value of $\Delta > 0$
- Two one-sided test procedure (TOST: α level):

$$\frac{\hat{\tau} + \Delta}{\sqrt{\widehat{\mathbb{V}(\hat{\tau})}}} \ > \ z_{1-\alpha} \qquad \frac{\hat{\tau} - \Delta}{\sqrt{\widehat{\mathbb{V}(\hat{\tau})}}} \ < \ -z_{1-\alpha}$$

- Two groups are equivalent if and only if both are rejected
- $\alpha =$ probability of falsely concluding equivalence under the null

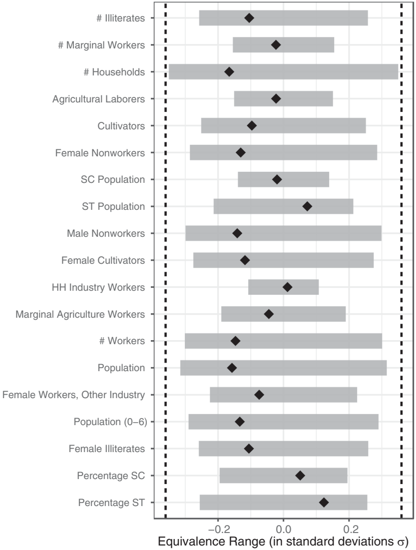


- Inverting the test $\rightsquigarrow$ the largest equivalence region which is consistent with the data at the $(1-\alpha)\times 100\%$ confidence level

Effect of Ethnic Quota on Redistribution

(Dunning and Nilekani. 2013. *Am. Political Sci. Rev.*)

- Ordered list determining villages whose council presidencies are reserved for scheduled castes
- balance tests $\rightsquigarrow$ number of households and mean female nonworkers have relatively small *p*-values



Bias of Matching

- Bias of matching arises because of imbalance:

$$B(\mathbf{X}_i, \mathcal{X}_{\mathcal{M}_i}) = \mathbb{E}(Y_i(0) \mid T_i = 1, \mathbf{X}_i) - \mathbb{E}\left\{ \frac{1}{|\mathcal{M}_i|} \sum_{i' \in \mathcal{M}_i} Y_{i'} \;\middle|\; \mathcal{X}_{\mathcal{M}_i} \right\}$$

  where $\mathcal{X}_{\mathcal{M}_i} = \{\mathbf{X}_{i'}\}_{i' \in \mathcal{M}_i}$

- Bias correction (Abadie and Imbens. 2011. *J Bus Econ Stat*):

$$\begin{aligned}
\widehat{Y_i(0)} &= \frac{1}{|\mathcal{M}_i|} \sum_{i' \in \mathcal{M}_i} Y_{i'} + B(\widehat{\mathbf{X}_i, \mathcal{X}_{\mathcal{M}_i}}) \\
&= \frac{1}{|\mathcal{M}_i|} \sum_{i' \in \mathcal{M}_i} \left\{ Y_{i'} + \hat{\beta}^\top (\mathbf{X}_i - \mathbf{X}_{i'}) \right\}
\end{aligned}$$

  where $\hat{\beta}$ is the estimated coefficient for the regression of $Y_{i'}$ on $\mathbf{X}_{i'}$ using all $i' \in \mathcal{M}_i$

Variance

- All matching estimators can be written as a weighting estimator:

$$
\begin{aligned}
\hat{\tau}_{\text{match}} &= \frac{1}{n_1}\sum_{i=1}^{n} T_i\left(Y_i - \frac{1}{|\mathcal{M}_i|}\sum_{i'\in\mathcal{M}_i} Y_{i'}\right) \\
&= \frac{1}{n_1}\sum_{i:T_i=1} Y_i - \frac{1}{n_0}\sum_{i:T_i=0} \underbrace{\left(\frac{n_0}{n_1}\sum_{i':T_{i'}=1}\frac{\mathbf{1}\{i\in\mathcal{M}_{i'}\}}{|\mathcal{M}_{i'}|}\right)}_{W_i} Y_i
\end{aligned}
$$

- Estimation error for CATT:

$$
\begin{aligned}
\hat{\tau}_{\text{match}} - \text{CATT} &= \underbrace{\frac{1}{n_1}\sum_{i:T_i=1}\mu(0,\mathbf{X}_i) - \frac{1}{n_0}\sum_{i:T_i=0} W_i\cdot\mu(0,\mathbf{X}_i)}_{\approx 0 \textit{ if matched well and in a large sample}} \\
&+\frac{1}{n_1}\sum_{i:T_i=1}(Y_i(1)-\mu(1,\mathbf{X}_i)) - \frac{1}{n_0}\sum_{i:T_i=0} W_i(Y_i(0)-\mu(0,\mathbf{X}_i))
\end{aligned}
$$

- Assume matching is done well and the sample is relatively large
- Conditional variance over sampling from a super-population,

$$\mathbb{V}(\hat{\tau}_{\text{match}} \mid \mathbf{X}, \mathbf{T}) \approx \frac{1}{n_1^2} \sum_{i: T_i = 1}^{n} \mathbb{V}(Y_i(1) \mid \mathbf{X}, \mathbf{T}) + \frac{1}{n_0^2} \sum_{i: T_i = 0}^{n} W_i^2 \mathbb{V}(Y_i(0) \mid \mathbf{X}, \mathbf{T})$$

- Estimation of $\mathbb{V}(Y_i(t) \mid \mathbf{X}, \mathbf{T})$ for each i

$$\mathbb{V}(\hat{\tau}_{\text{match}} \mid \mathbf{X}, \mathbf{T}) \approx \sum_{i=1}^{n} \widetilde{W}_i^2 \cdot \mathbb{V}(Y_i \mid \mathbf{X}, \mathbf{T})$$

where $\widetilde{W}_i = T_i / n_1 + (1 - T_i) W_i / n_0$

Sensitivity Analysis (Rosenbaum. 2002. *Observational Studies*. Springer)

- Consider a simple pair-matching of treated and control units
- Assumption: treatment assignment is random within each pair
- Question: How large a departure from the key (untestable) assumption must occur for the conclusions to no longer hold?
- Sensitivity analysis: for any pair j,

$$\frac{1}{\Gamma} \leq \frac{\Pr(T_i = 1 \mid S_i = j) / \Pr(T_i = 0 \mid S_i = j)}{\Pr(T_{i'} = 1 \mid S_{i'} = j) / \Pr(T_{i'} = 0 \mid S_{i'} = j)} \leq \Gamma \quad \text{where } \Gamma \geq 1$$

  - The model:

$$\Pr(T_i = 1 \mid \mathbf{X}_i, U_i) = \frac{\exp(f(\mathbf{X}_i) + \gamma U_i)}{1 + \exp(f(\mathbf{X}_i) + \gamma U_i)} \text{ where } \exp(\gamma) = \Gamma$$

  - Ratio of conditional treatment assignment probabilities can be bounded by $\Gamma/(1+\Gamma)$ and $1/(1+\Gamma)$ $\rightsquigarrow$ can bound p-value with Wilcoxon's signed rank sum test
  - How do the results change as you increase Γ?

Smoking and Lung Cancer

- Unobserved confounders:

  an error has been made, of an old kind, in arguing from correlation to causation, ... the possibility should be explored that the different smoking classes, non-smokers, cigarette smokers, cigar smokers, pipe smokers, etc., have adopted their habits partly by reason of their personal temperaments and dispositions, and are not lightly to be assumed to be equivalent in their genotypic composition. *(Fisher. 1958. Nature)*

- 36,975 heavy smokers paired with nonsomokers based on age, race, education, marital status, various health history measures, etc. (Hammond. 1964. *J. Natl. Cancer Inst.*)
  - Of these pairs, 122 pairs had exactly one person died of lung cancer – 110 heavy smokers
  - Sensitivity analysis based on McNemar's test (maximum p-value): < 0.0001 ($\Gamma = 3$), $0.004(\Gamma = 4)$, $0.03(\Gamma = 3)$, $0.1(\Gamma = 6)$

Summary

- Bias of regression due to covariate imbalance
- Matching reduces bias by improving covariate balance
- Various matching methods: propensity score, Mahalanobis, CEM, full matching, other optimal matching methods
- Importance of resulting balance $\rightsquigarrow$ equivalence test
- Matching does not eliminate bias due to unobservables
  $\rightsquigarrow$ sensitivity analysis

- Recommended readings:
  - Ho et al. 2007. "Matching as Nonparametric Preprocessing for Reducing Model Dependence in Parametric Causal Inference." *Political Analysis*
  - Stuart. 2010. "Matching methods for causal inference: A review and a look forward." *Statistical Science*
  - Imbens and Rubin. Chapters 12–15, 17–19, and 22.