Module 7a: Matching Estimators

Fall 2021

Matthew Blackwell

Gov 2003 (Harvard)

The problem with regression

- Causal inference is all about comparing **counterfactuals**, like the ATT:

$$\tau_{ATT} = \mathbb{E}[Y_i(1) - Y_i(0) \mid D_i = 1]$$

The problem with regression

- Causal inference is all about comparing **counterfactuals**, like the ATT:

$$\tau_{ATT} = \mathbb{E}[Y_i(1) - Y_i(0) \mid D_i = 1]$$

- Recall the **imputation** estimators with regression.

$$\widehat{\tau}_{\text{reg}} = \frac{1}{n_1} \sum_{i=1}^{n} D_i (Y_i - \hat{\mu}_0(\mathbf{X}_i))$$

The problem with regression

- Causal inference is all about comparing **counterfactuals**, like the ATT:

$$\tau_{ATT} = \mathbb{E}[Y_i(1) - Y_i(0) \mid D_i = 1]$$

- Recall the **imputation** estimators with regression.

$$\widehat{\tau}_{\text{reg}} = \frac{1}{n_1} \sum_{i=1}^{n} D_i(Y_i - \hat{\mu}_0(\mathbf{X}_i))$$

- Common solution: use a parametric model for $\hat{\mu}_0(\mathbf{X}_i)$

The problem with regression

- Causal inference is all about comparing **counterfactuals**, like the ATT:

$$\tau_{ATT} = \mathbb{E}[Y_i(1) - Y_i(0) \mid D_i = 1]$$

- Recall the **imputation** estimators with regression.

$$\widehat{\tau}_{\text{reg}} = \frac{1}{n_1} \sum_{i=1}^{n} D_i(Y_i - \widehat{\mu}_0(\mathbf{X}_i))$$

- Common solution: use a parametric model for $\widehat{\mu}_0(\mathbf{X}_i)$
    - For example, could assume it is linear: $\mu_0(\mathbf{x}) = \mathbf{x}'\beta$

The problem with regression

- Causal inference is all about comparing **counterfactuals**, like the ATT:

$$\tau_{ATT} = \mathbb{E}[Y_i(1) - Y_i(0) \mid D_i = 1]$$

- Recall the **imputation** estimators with regression.

$$\widehat{\tau}_{\text{reg}} = \frac{1}{n_1} \sum_{i=1}^{n} D_i(Y_i - \hat{\mu}_0(\mathbf{X}_i))$$

- Common solution: use a parametric model for $\hat{\mu}_0(\mathbf{X}_i)$
    - For example, could assume it is linear: $\mu_0(\mathbf{x}) = \mathbf{x}'\beta$
    - Regression, MLE, Bayes, etc.

The problem with regression

- Causal inference is all about comparing **counterfactuals**, like the ATT:

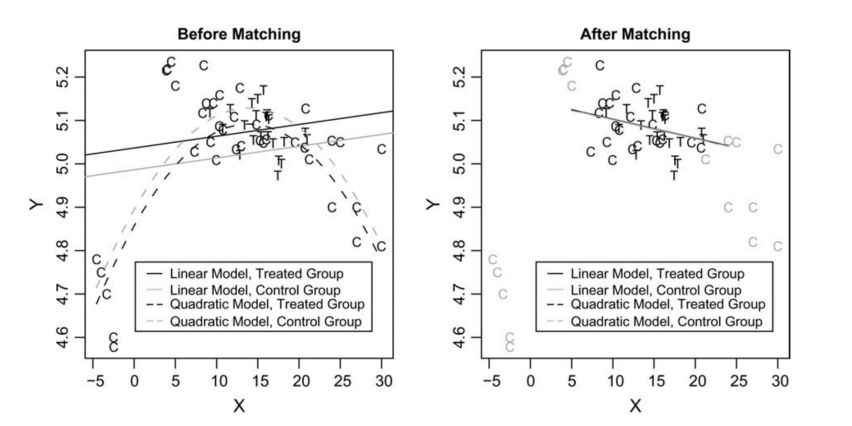
$$\tau_{ATT} = \mathbb{E}[Y_i(1) - Y_i(0) \mid D_i = 1]$$

- Recall the **imputation** estimators with regression.

$$\widehat{\tau}_{\text{reg}} = \frac{1}{n_1} \sum_{i=1}^{n} D_i (Y_i - \hat{\mu}_0(\mathbf{X}_i))$$

- Common solution: use a parametric model for $\hat{\mu}_0(\mathbf{X}_i)$
    - For example, could assume it is linear: $\mu_0(\mathbf{x}) = \mathbf{x}'\beta$
    - Regression, MLE, Bayes, etc.
    - But this model might be wrong $\rightsquigarrow$ wrong causal estimates.

Model dependence



What is matching?

- **Matching** is a nonparametric imputation estimator:

$$\widehat{\tau}_m = \frac{1}{n_1} \sum_{i=1}^{n} D_i \left(Y_i - \frac{1}{M} \sum_{j \in \mathcal{J}_M(i)} Y_j \right)$$

What is matching?

- **Matching** is a nonparametric imputation estimator:

$$\widehat{\tau}_m = \frac{1}{n_1} \sum_{i=1}^{n} D_i \left(Y_i - \frac{1}{M} \sum_{j \in \mathcal{J}_M(i)} Y_j \right)$$

  - $\mathcal{J}(i)$ are the set of M closest control units to i in terms of $\mathbf{X}_i$

What is matching?

- **Matching** is a nonparametric imputation estimator:

$$\widehat{\tau}_m = \frac{1}{n_1} \sum_{i=1}^{n} D_i \left(Y_i - \frac{1}{M} \sum_{j \in \mathcal{J}_M(i)} Y_j \right)$$

  - $\mathcal{J}(i)$ are the set of M closest control units to i in terms of $\mathbf{X}_i$
- Matching has strong advantages:

What is matching?

- **Matching** is a nonparametric imputation estimator:

$$\widehat{\tau}_m = \frac{1}{n_1} \sum_{i=1}^{n} D_i \left(Y_i - \frac{1}{M} \sum_{j \in \mathcal{J}_M(i)} Y_j \right)$$

  - $\mathcal{J}(i)$ are the set of M closest control units to i in terms of $\mathbf{X}_i$
- Matching has strong advantages:
  1. Reduces dependence of estimates on parametric models.

What is matching?

- **Matching** is a nonparametric imputation estimator:

$$\widehat{\tau}_m = \frac{1}{n_1} \sum_{i=1}^{n} D_i \left(Y_i - \frac{1}{M} \sum_{j \in \mathcal{J}_M(i)} Y_j \right)$$

  - $\mathcal{J}(i)$ are the set of M closest control units to i in terms of $\mathbf{X}_i$

- Matching has strong advantages:

  1. Reduces dependence of estimates on parametric models.
  2. Reduces model-based extrapolation.

What is matching?

- **Matching** is a nonparametric imputation estimator:

$$\widehat{\tau}_m = \frac{1}{n_1} \sum_{i=1}^{n} D_i \left(Y_i - \frac{1}{M} \sum_{j \in \mathcal{J}_M(i)} Y_j \right)$$

  - $\mathcal{J}(i)$ are the set of M closest control units to i in terms of $\mathbf{X}_i$

- Matching has strong advantages:

  1. Reduces dependence of estimates on parametric models.
  2. Reduces model-based extrapolation.
  3. Makes counterfactual comparisons more transparent.

What is matching?

- **Matching** is a nonparametric imputation estimator:

$$\widehat{\tau}_m = \frac{1}{n_1} \sum_{i=1}^{n} D_i \left(Y_i - \frac{1}{M} \sum_{j \in \mathcal{J}_M(i)} Y_j \right)$$

  - $\mathcal{J}(i)$ are the set of M closest control units to i in terms of $\mathbf{X}_i$

- Matching has strong advantages:

  1. Reduces dependence of estimates on parametric models.
  2. Reduces model-based extrapolation.
  3. Makes counterfactual comparisons more transparent.

- What matching isn't: a solution for selection on unobservables.

What is matching?

- **Matching** is a nonparametric imputation estimator:

$$\widehat{\tau}_m = \frac{1}{n_1} \sum_{i=1}^{n} D_i \left(Y_i - \frac{1}{M} \sum_{j \in \mathcal{J}_M(i)} Y_j \right)$$

  - $\mathcal{J}(i)$ are the set of M closest control units to i in terms of $\mathbf{X}_i$

- Matching has strong advantages:
  1. Reduces dependence of estimates on parametric models.
  2. Reduces model-based extrapolation.
  3. Makes counterfactual comparisons more transparent.
- What matching isn't: a solution for selection on unobservables.
  - Matching is an **estimation** technique, not an identification strategy.

Types of matching

- Assumptions:

Types of matching

- Assumptions:
  - No unmeasured confounders: $D_i \perp\!\!\!\perp (Y_i(0), Y_i(1)) \mid \mathbf{X}_i$

Types of matching

- Assumptions:
    - No unmeasured confounders: $D_i \perp\!\!\!\perp (Y_i(0), Y_i(1)) \mid \mathbf{X}_i$
    - Overlap/positivity: $0 < \mathbb{P}(D_i = 1 \mid \mathbf{X}_i = x) < 1$

Types of matching

- Assumptions:
  - No unmeasured confounders: $D_i \perp\!\!\!\perp (Y_i(0), Y_i(1)) \mid \mathbf{X}_i$
  - Overlap/positivity: $0 < \mathbb{P}(D_i = 1 \mid \mathbf{X}_i = x) < 1$
- **Exact matching**: choose matches that have the same value of $\mathbf{X}_i$

Types of matching

- Assumptions:
  - No unmeasured confounders: $D_i \perp\!\!\!\perp (Y_i(0), Y_i(1)) \mid \mathbf{X}_i$
  - Overlap/positivity: $0 < \mathbb{P}(D_i = 1 \mid \mathbf{X}_i = x) < 1$
- **Exact matching**: choose matches that have the same value of $\mathbf{X}_i$
  - $\mathcal{J}_M(i)$ is a random set of M control units with $\mathbf{X}_j = \mathbf{X}_i$

Types of matching

- Assumptions:
  - No unmeasured confounders: $D_i \perp\!\!\!\perp (Y_i(0), Y_i(1)) \mid \mathbf{X}_i$
  - Overlap/positivity: $0 < \mathbb{P}(D_i = 1 \mid \mathbf{X}_i = x) < 1$
- **Exact matching**: choose matches that have the same value of $\mathbf{X}_i$
  - $\mathcal{J}_M(i)$ is a random set of M control units with $\mathbf{X}_j = \mathbf{X}_i$
  - Covariate distribution is treated and matched controls exactly the same:

$$\widehat{\mathbb{P}}(\mathbf{X}_i = \mathbf{x} \mid D_i = 1) = \widehat{\mathbb{P}}(\mathbf{X}_j = \mathbf{x} \mid D_j = 0, j \text{ is matched})$$
$$\rightsquigarrow \mathbb{E}[Y_i(0) \mid D_i = 1] = \mathbb{E}[Y_j \mid D_j = 0, j \text{ is matched}]$$

Types of matching

- Assumptions:
  - No unmeasured confounders: $D_i \perp\!\!\!\perp (Y_i(0), Y_i(1)) \mid \mathbf{X}_i$
  - Overlap/positivity: $0 < \mathbb{P}(D_i = 1 \mid \mathbf{X}_i = x) < 1$
- **Exact matching**: choose matches that have the same value of $\mathbf{X}_i$
  - $\mathcal{J}_M(i)$ is a random set of M control units with $\mathbf{X}_j = \mathbf{X}_i$
  - Covariate distribution is treated and matched controls exactly the same:

$$\widehat{\mathbb{P}}(\mathbf{X}_i = \mathbf{x} \mid D_i = 1) = \widehat{\mathbb{P}}(\mathbf{X}_j = \mathbf{x} \mid D_j = 0, j \text{ is matched})$$
$$\rightsquigarrow \mathbb{E}[Y_i(0) \mid D_i = 1] = \mathbb{E}[Y_j \mid D_j = 0, j \text{ is matched}]$$

  - Problem: not feasible with high-dimesional or continuous $\mathbf{X}_i$

Types of matching

- Assumptions:
  - No unmeasured confounders: $D_i \perp\!\!\!\perp (Y_i(0), Y_i(1)) \mid \mathbf{X}_i$
  - Overlap/positivity: $0 < \mathbb{P}(D_i = 1 \mid \mathbf{X}_i = x) < 1$
- **Exact matching**: choose matches that have the same value of $\mathbf{X}_i$
  - $\mathcal{J}_M(i)$ is a random set of M control units with $\mathbf{X}_j = \mathbf{X}_i$
  - Covariate distribution is treated and matched controls exactly the same:

$$\widehat{\mathbb{P}}(\mathbf{X}_i = \mathbf{x} \mid D_i = 1) = \widehat{\mathbb{P}}(\mathbf{X}_j = \mathbf{x} \mid D_j = 0, j \text{ is matched})$$
$$\rightsquigarrow \mathbb{E}[Y_i(0) \mid D_i = 1] = \mathbb{E}[Y_j \mid D_j = 0, j \text{ is matched}]$$

  - Problem: not feasible with high-dimesional or continuous $\mathbf{X}_i$
- **Coarsened exact matching** (Iacus et al, 2011)

Types of matching

- Assumptions:
  - No unmeasured confounders: $D_i \perp\!\!\!\perp (Y_i(0), Y_i(1)) \mid \mathbf{X}_i$
  - Overlap/positivity: $0 < \mathbb{P}(D_i = 1 \mid \mathbf{X}_i = x) < 1$
- **Exact matching**: choose matches that have the same value of $\mathbf{X}_i$
  - $\mathcal{J}_M(i)$ is a random set of M control units with $\mathbf{X}_j = \mathbf{X}_i$
  - Covariate distribution is treated and matched controls exactly the same:

$$\widehat{\mathbb{P}}(\mathbf{X}_i = \mathbf{x} \mid D_i = 1) = \widehat{\mathbb{P}}(\mathbf{X}_j = \mathbf{x} \mid D_j = 0, j \text{ is matched})$$
$$\rightsquigarrow \mathbb{E}[Y_i(0) \mid D_i = 1] = \mathbb{E}[Y_j \mid D_j = 0, j \text{ is matched}]$$

  - Problem: not feasible with high-dimesional or continuous $\mathbf{X}_i$
- **Coarsened exact matching** (Iacus et al, 2011)
  - Discretize and group covariates into substantively meaningful bins

Types of matching

- Assumptions:
  - No unmeasured confounders: $D_i \perp\!\!\!\perp (Y_i(0), Y_i(1)) \mid \mathbf{X}_i$
  - Overlap/positivity: $0 < \mathbb{P}(D_i = 1 \mid \mathbf{X}_i = x) < 1$
- **Exact matching**: choose matches that have the same value of $\mathbf{X}_i$
  - $\mathcal{J}_M(i)$ is a random set of M control units with $\mathbf{X}_j = \mathbf{X}_i$
  - Covariate distribution is treated and matched controls exactly the same:

$$\widehat{\mathbb{P}}(\mathbf{X}_i = \mathbf{x} \mid D_i = 1) = \widehat{\mathbb{P}}(\mathbf{X}_j = \mathbf{x} \mid D_j = 0, j \text{ is matched})$$
$$\rightsquigarrow \mathbb{E}[Y_i(0) \mid D_i = 1] = \mathbb{E}[Y_j \mid D_j = 0, j \text{ is matched}]$$

  - Problem: not feasible with high-dimesional or continuous $\mathbf{X}_i$
- **Coarsened exact matching** (Iacus et al, 2011)
  - Discretize and group covariates into substantively meaningful bins
  - Exact match on these bins $\rightsquigarrow$ accounts for interactions

Types of matching

- Assumptions:
  - No unmeasured confounders: $D_i \perp\!\!\!\perp (Y_i(0), Y_i(1)) \mid \mathbf{X}_i$
  - Overlap/positivity: $0 < \mathbb{P}(D_i = 1 \mid \mathbf{X}_i = x) < 1$
- **Exact matching**: choose matches that have the same value of $\mathbf{X}_i$
  - $\mathcal{J}_M(i)$ is a random set of M control units with $\mathbf{X}_j = \mathbf{X}_i$
  - Covariate distribution is treated and matched controls exactly the same:

$$\widehat{\mathbb{P}}(\mathbf{X}_i = \mathbf{x} \mid D_i = 1) = \widehat{\mathbb{P}}(\mathbf{X}_j = \mathbf{x} \mid D_j = 0, j \text{ is matched})$$
$$\rightsquigarrow \mathbb{E}[Y_i(0) \mid D_i = 1] = \mathbb{E}[Y_j \mid D_j = 0, j \text{ is matched}]$$

  - Problem: not feasible with high-dimesional or continuous $\mathbf{X}_i$
- **Coarsened exact matching** (Iacus et al, 2011)
  - Discretize and group covariates into substantively meaningful bins
  - Exact match on these bins $\rightsquigarrow$ accounts for interactions
  - Have to drop treated units in bins with no controls $\rightsquigarrow$ changes estimand.

Types of matching

- Assumptions:
  - No unmeasured confounders: $D_i \perp\!\!\!\perp (Y_i(0), Y_i(1)) \mid \mathbf{X}_i$
  - Overlap/positivity: $0 < \mathbb{P}(D_i = 1 \mid \mathbf{X}_i = x) < 1$
- **Exact matching**: choose matches that have the same value of $\mathbf{X}_i$
  - $\mathcal{J}_M(i)$ is a random set of M control units with $\mathbf{X}_j = \mathbf{X}_i$
  - Covariate distribution is treated and matched controls exactly the same:

$$\widehat{\mathbb{P}}(\mathbf{X}_i = \mathbf{x} \mid D_i = 1) = \widehat{\mathbb{P}}(\mathbf{X}_j = \mathbf{x} \mid D_j = 0, j \text{ is matched})$$
$$\rightsquigarrow \mathbb{E}[Y_i(0) \mid D_i = 1] = \mathbb{E}[Y_j \mid D_j = 0, j \text{ is matched}]$$

  - Problem: not feasible with high-dimesional or continuous $\mathbf{X}_i$
- **Coarsened exact matching** (Iacus et al, 2011)
  - Discretize and group covariates into substantively meaningful bins
  - Exact match on these bins $\rightsquigarrow$ accounts for interactions
  - Have to drop treated units in bins with no controls $\rightsquigarrow$ changes estimand.
  - Allows you to control bias/variance tradeoff through coarsening.

Matching in high dimensions

- Even CEM can break down with high dimensional $\mathbf{X}_i$.

Matching in high dimensions

- Even CEM can break down with high dimensional $\mathbf{X}_i$.
- We can define closeness using lower dimensional **distance metrics**

Matching in high dimensions

- Even CEM can break down with high dimensional $\mathbf{X}_i$.
- We can define closeness using lower dimensional **distance metrics**
    - Reduces dimensionality: maps two vectors to a single number

Matching in high dimensions

- Even CEM can break down with high dimensional $\mathbf{X}_i$.
- We can define closeness using lower dimensional **distance metrics**
    - Reduces dimensionality: maps two vectors to a single number
- **Mahalanobis distance**:

$$D(\mathbf{X}_j, \mathbf{X}_j) = \sqrt{(\mathbf{X}_i - \mathbf{X}_j)'\widehat{\Sigma}^{-1}(\mathbf{X}_i - \mathbf{X}_j)}$$

Matching in high dimensions

- Even CEM can break down with high dimensional $\mathbf{X}_i$.
- We can define closeness using lower dimensional **distance metrics**
  - Reduces dimensionality: maps two vectors to a single number
- **Mahalanobis distance**:

$$D(\mathbf{X}_j, \mathbf{X}_j) = \sqrt{(\mathbf{X}_i - \mathbf{X}_j)' \widehat{\Sigma}^{-1} (\mathbf{X}_i - \mathbf{X}_j)}$$

  - $\widehat{\Sigma}$ is the estimated variance-covariance matrix of the observations:

$$\widehat{\Sigma} = \frac{1}{n} \sum_{i=1}^{n} (\mathbf{X}_i - \bar{\mathbf{X}})(\mathbf{X}_i - \bar{\mathbf{X}})^T$$

Matching in high dimensions

- Even CEM can break down with high dimensional $\mathbf{X}_i$.
- We can define closeness using lower dimensional **distance metrics**
  - Reduces dimensionality: maps two vectors to a single number
- **Mahalanobis distance**:

$$D(\mathbf{X}_j, \mathbf{X}_j) = \sqrt{(\mathbf{X}_i - \mathbf{X}_j)'\widehat{\Sigma}^{-1}(\mathbf{X}_i - \mathbf{X}_j)}$$

  - $\widehat{\Sigma}$ is the estimated variance-covariance matrix of the observations:

$$\widehat{\Sigma} = \frac{1}{n}\sum_{i=1}^{n}(\mathbf{X}_i - \bar{\mathbf{X}})(\mathbf{X}_i - \bar{\mathbf{X}})^T$$

- Estimated propensity score:

$$D(\mathbf{X}_i, \mathbf{X}_j) = |\widehat{\pi}(\mathbf{X}_i) - \widehat{\pi}(\mathbf{X}_j)| = \left|\widehat{\mathbb{P}}(D_i = 1 \mid \mathbf{X}_i) - \widehat{\mathbb{P}}(D_i = 1 \mid \mathbf{X}_j)\right|$$

Matching in high dimensions

- Even CEM can break down with high dimensional $\mathbf{X}_i$.
- We can define closeness using lower dimensional **distance metrics**
    - Reduces dimensionality: maps two vectors to a single number
- **Mahalanobis distance**:

$$D(\mathbf{X}_j, \mathbf{X}_j) = \sqrt{(\mathbf{X}_i - \mathbf{X}_j)'\widehat{\Sigma}^{-1}(\mathbf{X}_i - \mathbf{X}_j)}$$

    - $\widehat{\Sigma}$ is the estimated variance-covariance matrix of the observations:

$$\widehat{\Sigma} = \frac{1}{n}\sum_{i=1}^{n}(\mathbf{X}_i - \bar{\mathbf{X}})(\mathbf{X}_i - \bar{\mathbf{X}})^T$$

- Estimated propensity score:

$$D(\mathbf{X}_i, \mathbf{X}_j) = |\widehat{\pi}(\mathbf{X}_i) - \widehat{\pi}(\mathbf{X}_j)| = \left|\widehat{\mathbb{P}}(D_i = 1 \mid \mathbf{X}_i) - \widehat{\mathbb{P}}(D_i = 1 \mid \mathbf{X}_j)\right|$$

    - Some use the linear predictor: $D(\mathbf{X}_i, \mathbf{X}_j) = |\text{logit}(\widehat{\pi}(\mathbf{X}_i)) - \text{logit}(\widehat{\pi}(\mathbf{X}_j))|$

Other matching choices

- **Matching ratio** how many control units per treated?

Other matching choices

- **Matching ratio** how many control units per treated?
    - Lower reduces bias (only use the closest matches)

Other matching choices

- **Matching ratio** how many control units per treated?
  - Lower reduces bias (only use the closest matches)
  - Lower increases variance

Other matching choices

- **Matching ratio** how many control units per treated?
  - Lower reduces bias (only use the closest matches)
  - Lower increases variance
- **With or without replacement** same control matched to multiple treated?

Other matching choices

- **Matching ratio** how many control units per treated?
  - Lower reduces bias (only use the closest matches)
  - Lower increases variance
- **With or without replacement** same control matched to multiple treated?
  - With replacement gives better matches & matching order doesn't matter.

Other matching choices

- **Matching ratio** how many control units per treated?
    - Lower reduces bias (only use the closest matches)
    - Lower increases variance
- **With or without replacement** same control matched to multiple treated?
    - With replacement gives better matches & matching order doesn't matter.
    - Without replacement simplifies variance estimation.

Other matching choices

- **Matching ratio** how many control units per treated?
  - Lower reduces bias (only use the closest matches)
  - Lower increases variance
- **With or without replacement** same control matched to multiple treated?
  - With replacement gives better matches & matching order doesn't matter.
  - Without replacement simplifies variance estimation.
- **Caliper**: drop poor matches?

Other matching choices

- **Matching ratio** how many control units per treated?
  - Lower reduces bias (only use the closest matches)
  - Lower increases variance
- **With or without replacement** same control matched to multiple treated?
  - With replacement gives better matches & matching order doesn't matter.
  - Without replacement simplifies variance estimation.
- **Caliper**: drop poor matches?
  - Only keep matches below a distance threshold, $D(\mathbf{X}_i, \mathbf{X}_j) \leq c$

Other matching choices

- **Matching ratio** how many control units per treated?
    - Lower reduces bias (only use the closest matches)
    - Lower increases variance
- **With or without replacement** same control matched to multiple treated?
    - With replacement gives better matches & matching order doesn't matter.
    - Without replacement simplifies variance estimation.
- **Caliper**: drop poor matches?
    - Only keep matches below a distance threshold, $D(\mathbf{X}_i, \mathbf{X}_j) \leq c$
    - Reduces imbalance, but if you drop treated units, estimand changes.

Propensity scores, redux

- Covariates are balanced conditional on true propensity score:

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

Propensity scores, redux

- Covariates are balanced conditional on true propensity score:

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

- Implies we only need to match/balance on $\pi(\mathbf{x})$:

$$\big(Y_i(0), Y_i(1)\big) \perp\!\!\!\perp D_i \mid \mathbf{X}_i \iff \big(Y_i(0), Y_i(1)\big) \perp\!\!\!\perp D_i \mid \pi(\mathbf{X}_i)$$

Propensity scores, redux

- Covariates are balanced conditional on true propensity score:

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

- Implies we only need to match/balance on $\pi(\mathbf{x})$:

$$\big(Y_i(0), Y_i(1)\big) \perp\!\!\!\perp D_i \mid \mathbf{X}_i \iff \big(Y_i(0), Y_i(1)\big) \perp\!\!\!\perp D_i \mid \pi(\mathbf{X}_i)$$

- In observational data we never know the true $\pi(\mathbf{x}) \rightsquigarrow$ estimate it $\widehat{\pi}(\mathbf{x})$.

Propensity scores, redux

- Covariates are balanced conditional on true propensity score:

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

- Implies we only need to match/balance on $\pi(\mathbf{x})$:

$$\left(Y_i(0), Y_i(1)\right) \perp\!\!\!\perp D_i \mid \mathbf{X}_i \iff \left(Y_i(0), Y_i(1)\right) \perp\!\!\!\perp D_i \mid \pi(\mathbf{X}_i)$$

- In observational data we never know the true $\pi(\mathbf{x}) \rightsquigarrow$ estimate it $\widehat{\pi}(\mathbf{x})$.
- Is balancing on $\widehat{\pi}(\mathbf{x})$ sufficient?

Propensity scores, redux

- Covariates are balanced conditional on true propensity score:

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

- Implies we only need to match/balance on $\pi(\mathbf{x})$:

$$\left(Y_i(0), Y_i(1)\right) \perp\!\!\!\perp D_i \mid \mathbf{X}_i \iff \left(Y_i(0), Y_i(1)\right) \perp\!\!\!\perp D_i \mid \pi(\mathbf{X}_i)$$

- In observational data we never know the true $\pi(\mathbf{x}) \rightsquigarrow$ estimate it $\widehat{\pi}(\mathbf{x})$.
- Is balancing on $\widehat{\pi}(\mathbf{x})$ sufficient? **No idea!!**

Propensity scores, redux

- Covariates are balanced conditional on true propensity score:

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

- Implies we only need to match/balance on $\pi(\mathbf{x})$:

$$\left(Y_i(0), Y_i(1)\right) \perp\!\!\!\perp D_i \mid \mathbf{X}_i \iff \left(Y_i(0), Y_i(1)\right) \perp\!\!\!\perp D_i \mid \pi(\mathbf{X}_i)$$

- In observational data we never know the true $\pi(\mathbf{x}) \rightsquigarrow$ estimate it $\widehat{\pi}(\mathbf{x})$.
- Is balancing on $\widehat{\pi}(\mathbf{x})$ sufficient? **No idea!!**
    - Have to check if $\mathbf{X}_i$ is actually balanced.

Propensity scores, redux

- Covariates are balanced conditional on true propensity score:

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

- Implies we only need to match/balance on $\pi(\mathbf{x})$:

$$\big(Y_i(0), Y_i(1)\big) \perp\!\!\!\perp D_i \mid \mathbf{X}_i \iff \big(Y_i(0), Y_i(1)\big) \perp\!\!\!\perp D_i \mid \pi(\mathbf{X}_i)$$

- In observational data we never know the true $\pi(\mathbf{x}) \rightsquigarrow$ estimate it $\widehat{\pi}(\mathbf{x})$.
- Is balancing on $\widehat{\pi}(\mathbf{x})$ sufficient? **No idea!!**
    - Have to check if $\mathbf{X}_i$ is actually balanced.
    - Somewhat deflates the benefits of PS matching/balancing.

Propensity scores, redux

- Covariates are balanced conditional on true propensity score:

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

- Implies we only need to match/balance on $\pi(\mathbf{x})$:

$$\left(Y_i(0), Y_i(1)\right) \perp\!\!\!\perp D_i \mid \mathbf{X}_i \iff \left(Y_i(0), Y_i(1)\right) \perp\!\!\!\perp D_i \mid \pi(\mathbf{X}_i)$$

- In observational data we never know the true $\pi(\mathbf{x}) \rightsquigarrow$ estimate it $\widehat{\pi}(\mathbf{x})$.
- Is balancing on $\widehat{\pi}(\mathbf{x})$ sufficient? **No idea!!**
    - Have to check if $\mathbf{X}_i$ is actually balanced.
    - Somewhat deflates the benefits of PS matching/balancing.
- $\rightsquigarrow$ "propensity score tautology"

Assessing balance

- Goal of matching is to maximize balance: $\widehat{F}_1(\mathbf{x}) \approx \widehat{F}_{0,\mathcal{J}}(\mathbf{x})$

Assessing balance

- Goal of matching is to maximize balance: $\widehat{F}_1(\mathbf{x}) \approx \widehat{F}_{0,\mathcal{J}}(\mathbf{x})$
    - Joint distribution of $\mathbf{X}_i$ is similar between treated and matched controls.

Assessing balance

- Goal of matching is to maximize balance: $\widehat{F}_1(\mathbf{x}) \approx \widehat{F}_{0,\mathcal{J}}(\mathbf{x})$
    - Joint distribution of $\mathbf{X}_i$ is similar between treated and matched controls.
    - Difficult to asses balance across many dimensions $\rightsquigarrow$ summaries.

Assessing balance

- Goal of matching is to maximize balance: $\widehat{F}_1(\mathbf{x}) \approx \widehat{F}_{0,\mathcal{J}}(\mathbf{x})$
    - Joint distribution of $\mathbf{X}_i$ is similar between treated and matched controls.
    - Difficult to asses balance across many dimensions $\rightsquigarrow$ summaries.
- Options:

Assessing balance

- Goal of matching is to maximize balance: $\widehat{F}_1(\mathbf{x}) \approx \widehat{F}_{0,\mathcal{J}}(\mathbf{x})$
    - Joint distribution of $\mathbf{X}_i$ is similar between treated and matched controls.
    - Difficult to asses balance across many dimensions $\rightsquigarrow$ summaries.
- Options:
    - Differences-in-means/medians, standardized.

Assessing balance

- Goal of matching is to maximize balance: $\widehat{F}_1(\mathbf{x}) \approx \widehat{F}_{0,\mathcal{J}}(\mathbf{x})$
    - Joint distribution of $\mathbf{X}_i$ is similar between treated and matched controls.
    - Difficult to asses balance across many dimensions $\rightsquigarrow$ summaries.
- Options:
    - Differences-in-means/medians, standardized.
    - QQ plots/KS statistics for comparing the entire distribution of X_i.

Assessing balance

- Goal of matching is to maximize balance: $\widehat{F}_1(\mathbf{x}) \approx \widehat{F}_{0,\mathcal{J}}(\mathbf{x})$
  - Joint distribution of $\mathbf{X}_i$ is similar between treated and matched controls.
  - Difficult to asses balance across many dimensions $\rightsquigarrow$ summaries.
- Options:
  - Differences-in-means/medians, standardized.
  - QQ plots/KS statistics for comparing the entire distribution of X_i.
  - L_1: multivariate histogram (for CEM)

Assessing balance

- Goal of matching is to maximize balance: $\widehat{F}_1(\mathbf{x}) \approx \widehat{F}_{0,\mathcal{J}}(\mathbf{x})$
    - Joint distribution of $\mathbf{X}_i$ is similar between treated and matched controls.
    - Difficult to asses balance across many dimensions $\rightsquigarrow$ summaries.
- Options:
    - Differences-in-means/medians, standardized.
    - QQ plots/KS statistics for comparing the entire distribution of X_i.
    - L_1: multivariate histogram (for CEM)
    - Choice of metric can change what matching method works best.

Assessing balance

- Goal of matching is to maximize balance: $\widehat{F}_1(\mathbf{x}) \approx \widehat{F}_{0,\mathcal{J}}(\mathbf{x})$
    - Joint distribution of $\mathbf{X}_i$ is similar between treated and matched controls.
    - Difficult to asses balance across many dimensions $\rightsquigarrow$ summaries.
- Options:
    - Differences-in-means/medians, standardized.
    - QQ plots/KS statistics for comparing the entire distribution of X_i.
    - L_1: multivariate histogram (for CEM)
    - Choice of metric can change what matching method works best.
- Hypothesis tests for balance are problematic:

Assessing balance

- Goal of matching is to maximize balance: $\widehat{F}_1(\mathbf{x}) \approx \widehat{F}_{0,\mathcal{J}}(\mathbf{x})$
  - Joint distribution of $\mathbf{X}_i$ is similar between treated and matched controls.
  - Difficult to asses balance across many dimensions $\rightsquigarrow$ summaries.
- Options:
  - Differences-in-means/medians, standardized.
  - QQ plots/KS statistics for comparing the entire distribution of X_i.
  - L_1: multivariate histogram (for CEM)
  - Choice of metric can change what matching method works best.
- Hypothesis tests for balance are problematic:
  - Dropping units can lower power ($\uparrow$ p-values) without a change in balance.

Bias of inexact matching

- To show the bias on matching, focus on finding a single control match.

Bias of inexact matching

- To show the bias on matching, focus on finding a single control match.
- Let $j(i)$ be the matched control for unit i, the bias is:

$$\mathbb{E}\left[Y_j \mid D_i = 1, \mathbf{X}_i, \mathbf{X}_j\right] - \mathbb{E}[Y_i(0) \mid D_i = 1, \mathbf{X}_i] = \underbrace{(\mu_0(\mathbf{X}_i) - \mu_0(\mathbf{X}_{j(i)}))}_{\text{unit-level bias}}$$

Bias of inexact matching

- To show the bias on matching, focus on finding a single control match.
- Let $j(i)$ be the matched control for unit i, the bias is:

$$\mathbb{E}\left[Y_j \mid D_i = 1, \mathbf{X}_i, \mathbf{X}_j\right] - \mathbb{E}[Y_i(0) \mid D_i = 1, \mathbf{X}_i] = \underbrace{(\mu_0(\mathbf{X}_i) - \mu_0(\mathbf{X}_{j(i)}))}_{\text{unit-level bias}}$$

  - Bias is 0 if matching is exact since $\mathbf{X}_i = \mathbf{X}_{j(i)}$

Bias of inexact matching

- To show the bias on matching, focus on finding a single control match.
- Let $j(i)$ be the matched control for unit i, the bias is:

$$\mathbb{E}\left[Y_j \mid D_i = 1, \mathbf{X}_i, \mathbf{X}_j\right] - \mathbb{E}[Y_i(0) \mid D_i = 1, \mathbf{X}_i] = \underbrace{(\mu_0(\mathbf{X}_i) - \mu_0(\mathbf{X}_{j(i)}))}_{\text{unit-level bias}}$$

  - Bias is 0 if matching is exact since $\mathbf{X}_i = \mathbf{X}_{j(i)}$
  - Bias grows with **matching discrepancy**/imbalance.

Bias of inexact matching

- To show the bias on matching, focus on finding a single control match.
- Let $j(i)$ be the matched control for unit i, the bias is:

$$\mathbb{E}\left[Y_j \mid D_i = 1, \mathbf{X}_i, \mathbf{X}_j\right] - \mathbb{E}[Y_i(0) \mid D_i = 1, \mathbf{X}_i] = \underbrace{(\mu_0(\mathbf{X}_i) - \mu_0(\mathbf{X}_{j(i)}))}_{\text{unit-level bias}}$$

  - Bias is 0 if matching is exact since $\mathbf{X}_i = \mathbf{X}_{j(i)}$
  - Bias grows with **matching discrepancy**/imbalance.

- **Bias correction**: estimate $\hat{\mu}_0(\mathbf{x})$ with regression and estimate bias.

$$\widehat{Y}_i(0) = Y_{j(i)} - (\hat{\mu}_0(\mathbf{X}_i) - \hat{\mu}_0(\mathbf{X}_{j(i)}))$$

Bias of inexact matching

- To show the bias on matching, focus on finding a single control match.
- Let $j(i)$ be the matched control for unit i, the bias is:

$$\mathbb{E}\left[Y_j \mid D_i = 1, \mathbf{X}_i, \mathbf{X}_j\right] - \mathbb{E}[Y_i(0) \mid D_i = 1, \mathbf{X}_i] = \underbrace{(\mu_0(\mathbf{X}_i) - \mu_0(\mathbf{X}_{j(i)}))}_{\text{unit-level bias}}$$

  - Bias is 0 if matching is exact since $\mathbf{X}_i = \mathbf{X}_{j(i)}$
  - Bias grows with **matching discrepancy**/imbalance.
- **Bias correction**: estimate $\hat{\mu}_0(\mathbf{x})$ with regression and estimate bias.

$$\widehat{Y_i}(0) = Y_{j(i)} - (\hat{\mu}_0(\mathbf{X}_i) - \hat{\mu}_0(\mathbf{X}_{j(i)}))$$

  - Imputation of missing potential outcome now matching + regression.

Bias of inexact matching

- To show the bias on matching, focus on finding a single control match.
- Let $j(i)$ be the matched control for unit i, the bias is:

$$\mathbb{E}\left[Y_j \mid D_i = 1, \mathbf{X}_i, \mathbf{X}_j\right] - \mathbb{E}[Y_i(0) \mid D_i = 1, \mathbf{X}_i] = \underbrace{(\mu_0(\mathbf{X}_i) - \mu_0(\mathbf{X}_{j(i)}))}_{\text{unit-level bias}}$$

  - Bias is 0 if matching is exact since $\mathbf{X}_i = \mathbf{X}_{j(i)}$
  - Bias grows with **matching discrepancy**/imbalance.

- **Bias correction**: estimate $\hat{\mu}_0(\mathbf{x})$ with regression and estimate bias.

$$\widehat{Y_i}(0) = Y_{j(i)} - (\hat{\mu}_0(\mathbf{X}_i) - \hat{\mu}_0(\mathbf{X}_{j(i)}))$$

  - Imputation of missing potential outcome now matching + regression.
  - Generalizes easily to any number of matches.

Variance

- Matching with replacement: cluster on the match.

Variance

- Matching with replacement: cluster on the match.
    - Can either use clustered SEs or cluster bootstrap.

Variance

- Matching with replacement: cluster on the match.
    - Can either use clustered SEs or cluster bootstrap.
    - Valid for post-matching regression (Abadie and Spiess, 2021)

Variance

- Matching with replacement: cluster on the match.
    - Can either use clustered SEs or cluster bootstrap.
    - Valid for post-matching regression (Abadie and Spiess, 2021)
- Matching without replacement: more complicated.

Variance

- Matching with replacement: cluster on the match.
  - Can either use clustered SEs or cluster bootstrap.
  - Valid for post-matching regression (Abadie and Spiess, 2021)
- Matching without replacement: more complicated.
  - Same control unit matched to multiple treated: no easy clustering.

Variance

- Matching with replacement: cluster on the match.
  - Can either use clustered SEs or cluster bootstrap.
  - Valid for post-matching regression (Abadie and Spiess, 2021)
- Matching without replacement: more complicated.
  - Same control unit matched to multiple treated: no easy clustering.
  - $K_M(i)$ is the number of times a unit is used as a match

Variance

- Matching with replacement: cluster on the match.
  - Can either use clustered SEs or cluster bootstrap.
  - Valid for post-matching regression (Abadie and Spiess, 2021)
- Matching without replacement: more complicated.
  - Same control unit matched to multiple treated: no easy clustering.
  - $K_M(i)$ is the number of times a unit is used as a match
- Assuming units are well-matched so bias can be ignored,

$$\mathbb{V}(\widehat{\tau}_m) = \frac{1}{n_1}\left(\underbrace{\mathbb{E}\left[(\tau(\mathbf{X}_i) - \tau_{ATT})^2 \mid D_i = 1\right]}_{\text{variance of CATE on treated}} + \underbrace{\mathbb{V}[\widehat{\tau}_m \mid \mathbb{X}, \mathbf{D}]}_{\text{conditional variance}}\right)$$

Variance

- Matching with replacement: cluster on the match.
  - Can either use clustered SEs or cluster bootstrap.
  - Valid for post-matching regression (Abadie and Spiess, 2021)
- Matching without replacement: more complicated.
  - Same control unit matched to multiple treated: no easy clustering.
  - $K_M(i)$ is the number of times a unit is used as a match
- Assuming units are well-matched so bias can be ignored,

$$\mathbb{V}(\widehat{\tau}_m) = \frac{1}{n_1}\left(\underbrace{\mathbb{E}\left[(\tau(\mathbf{X}_i) - \tau_{ATT})^2 \mid D_i = 1\right]}_{\text{variance of CATE on treated}} + \underbrace{\mathbb{V}[\widehat{\tau}_m \mid \mathbb{X}, \mathbf{D}]}_{\text{conditional variance}}\right)$$

  - Abadie and Imbens (2006) provides matching-based variance estimators.