Matching

Ye Wang
University of North Carolina at Chapel Hill

Linear Methods in Causal Inference
POLI784

Review

- We discussed the difference between experiments and observational studies.

Review

- We discussed the difference between experiments and observational studies.
- Both require (conditional) randomization of the treatment for causal identification.

Review

- We discussed the difference between experiments and observational studies.
- Both require (conditional) randomization of the treatment for causal identification.
- The difference is the we do not know the probability of being treated in the latter.

Review

- We discussed the difference between experiments and observational studies.
- Both require (conditional) randomization of the treatment for causal identification.
- The difference is the we do not know the probability of being treated in the latter.
- We have to estimate nuisance parameters under imposed structural restrictions.

Review

- We discussed the difference between experiments and observational studies.
- Both require (conditional) randomization of the treatment for causal identification.
- The difference is the we do not know the probability of being treated in the latter.
- We have to estimate nuisance parameters under imposed structural restrictions.
- We can estimate either propensity scores or response surfaces.

Blocking

- Suppose our data are generated by a block randomization.

Blocking

- Suppose our data are generated by a block randomization.
- We know covariates that are used to form blocks but not the probability of being treated within each block.

Blocking

- Suppose our data are generated by a block randomization.
- We know covariates that are used to form blocks but not the probability of being treated within each block.
- A natural idea: allocate units with the same covariates values into the same group.

Blocking

- Suppose our data are generated by a block randomization.
- We know covariates that are used to form blocks but not the probability of being treated within each block.
- A natural idea: allocate units with the same covariates values into the same group.
- Next, we estimate the CATE in each group and aggregate them to estimate the SATE.

Blocking

- Suppose our data are generated by a block randomization.
- We know covariates that are used to form blocks but not the probability of being treated within each block.
- A natural idea: allocate units with the same covariates values into the same group.
- Next, we estimate the CATE in each group and aggregate them to estimate the SATE.
- This method is known as blocking in the literature.

Blocking

- Suppose our data are generated by a block randomization.
- We know covariates that are used to form blocks but not the probability of being treated within each block.
- A natural idea: allocate units with the same covariates values into the same group.
- Next, we estimate the CATE in each group and aggregate them to estimate the SATE.
- This method is known as blocking in the literature.
- It does not work when there are many confounders or some confounders are continuous.

Matching

- Instead, we can create a block for each unit using similar units (just like kernel regression).

Matching

- Instead, we can create a block for each unit using similar units (just like kernel regression).
- For any treated (untreated) unit i, we find M units from the control (treatment) group that are similar to it in covariates (nearest-neighbor matching or NN matching).

Matching

- Instead, we can create a block for each unit using similar units (just like kernel regression).
- For any treated (untreated) unit i, we find M units from the control (treatment) group that are similar to it in covariates (nearest-neighbor matching or NN matching).
- Or, we find M units from the control (treatment) group that are similar to it in the propensity score (propensity score matching or PS matching).

Matching

- Instead, we can create a block for each unit using similar units (just like kernel regression).
- For any treated (untreated) unit i, we find M units from the control (treatment) group that are similar to it in covariates (nearest-neighbor matching or NN matching).
- Or, we find M units from the control (treatment) group that are similar to it in the propensity score (propensity score matching or PS matching).
- We need to decide the number of neighbors (M) and whether replacement is allowed.

Matching

- Instead, we can create a block for each unit using similar units (just like kernel regression).
- For any treated (untreated) unit i, we find M units from the control (treatment) group that are similar to it in covariates (nearest-neighbor matching or NN matching).
- Or, we find M units from the control (treatment) group that are similar to it in the propensity score (propensity score matching or PS matching).
- We need to decide the number of neighbors (M) and whether replacement is allowed.
- If not, each unit from the control (treatment) group can only be matched to one treated (untreated) unit.

Matching

- Instead, we can create a block for each unit using similar units (just like kernel regression).
- For any treated (untreated) unit i, we find M units from the control (treatment) group that are similar to it in covariates (nearest-neighbor matching or NN matching).
- Or, we find M units from the control (treatment) group that are similar to it in the propensity score (propensity score matching or PS matching).
- We need to decide the number of neighbors (M) and whether replacement is allowed.
- If not, each unit from the control (treatment) group can only be matched to one treated (untreated) unit.
- Not all the units from the control (treatment) group will be matched to a treated (untreated) unit.

Matching

- Instead, we can create a block for each unit using similar units (just like kernel regression).
- For any treated (untreated) unit i, we find M units from the control (treatment) group that are similar to it in covariates (nearest-neighbor matching or NN matching).
- Or, we find M units from the control (treatment) group that are similar to it in the propensity score (propensity score matching or PS matching).
- We need to decide the number of neighbors (M) and whether replacement is allowed.
- If not, each unit from the control (treatment) group can only be matched to one treated (untreated) unit.
- Not all the units from the control (treatment) group will be matched to a treated (untreated) unit.
- The existence of such units usually suggests the violation of positivity.

Matching

- We also need a distance metric to measure how similar two units are.

Matching

- We also need a distance metric to measure how similar two units are.
- A natural choice is the Euclidean distance; but it is sensitive to the units of the variables.

Matching

- We also need a distance metric to measure how similar two units are.
- A natural choice is the Euclidean distance; but it is sensitive to the units of the variables.
- An alternative is the Mahalanobis distance:

$$d(\mathbf{X}_i, \mathbf{X}_j) = \sqrt{(\mathbf{X}_i - \bar{\mathbf{X}})' \mathbf{S}^{-1} (\mathbf{X}_j - \bar{\mathbf{X}})},$$

  where **S** is a positive-definite matrix such as the variance-covariance matrix of the variables.

Matching

- We also need a distance metric to measure how similar two units are.
- A natural choice is the Euclidean distance; but it is sensitive to the units of the variables.
- An alternative is the Mahalanobis distance:

$$d(\mathbf{X}_i, \mathbf{X}_j) = \sqrt{(\mathbf{X}_i - \bar{\mathbf{X}})' \mathbf{S}^{-1} (\mathbf{X}_j - \bar{\mathbf{X}})},$$

  where **S** is a positive-definite matrix such as the variance-covariance matrix of the variables.
- This metric standardizes all the variables such that units no longer matter.

Matching

- The basic steps are as follows

Matching

- The basic steps are as follows
    1. Decide your estimand (ATE or ATT),

Matching

- The basic steps are as follows
  1. Decide your estimand (ATE or ATT),
  2. Choose the matching method (what covariates to be matched on, NN or PS matching, number of neighbors, with or without replacement, etc.),

Matching

- The basic steps are as follows
  1. Decide your estimand (ATE or ATT),
  2. Choose the matching method (what covariates to be matched on, NN or PS matching, number of neighbors, with or without replacement, etc.),
  3. Choose a proper distance metric,

Matching

- The basic steps are as follows
  1. Decide your estimand (ATE or ATT),
  2. Choose the matching method (what covariates to be matched on, NN or PS matching, number of neighbors, with or without replacement, etc.),
  3. Choose a proper distance metric,
  4. Find matches on your set of covariates/propensity scores, and get rid of non-matches,

Matching

- The basic steps are as follows
  1. Decide your estimand (ATE or ATT),
  2. Choose the matching method (what covariates to be matched on, NN or PS matching, number of neighbors, with or without replacement, etc.),
  3. Choose a proper distance metric,
  4. Find matches on your set of covariates/propensity scores, and get rid of non-matches,
  5. Check balance in your matched data set,

Matching

- The basic steps are as follows
  1. Decide your estimand (ATE or ATT),
  2. Choose the matching method (what covariates to be matched on, NN or PS matching, number of neighbors, with or without replacement, etc.),
  3. Choose a proper distance metric,
  4. Find matches on your set of covariates/propensity scores, and get rid of non-matches,
  5. Check balance in your matched data set,
  6. Repeat these steps until your set exhibits acceptable balance,

Matching

- The basic steps are as follows
    1. Decide your estimand (ATE or ATT),
    2. Choose the matching method (what covariates to be matched on, NN or PS matching, number of neighbors, with or without replacement, etc.),
    3. Choose a proper distance metric,
    4. Find matches on your set of covariates/propensity scores, and get rid of non-matches,
    5. Check balance in your matched data set,
    6. Repeat these steps until your set exhibits acceptable balance,
    7. Calculate the estimate on your matched dataset,

Matching

- The basic steps are as follows

  1. Decide your estimand (ATE or ATT),
  2. Choose the matching method (what covariates to be matched on, NN or PS matching, number of neighbors, with or without replacement, etc.),
  3. Choose a proper distance metric,
  4. Find matches on your set of covariates/propensity scores, and get rid of non-matches,
  5. Check balance in your matched data set,
  6. Repeat these steps until your set exhibits acceptable balance,
  7. Calculate the estimate on your matched dataset,
  8. Derive standard errors and construct confidence intervals.

Nearest-neighbor matching

- Nearest-neighbor matching is similar to blocking if all the covariates only take discrete values (exact matching).

Nearest-neighbor matching

- Nearest-neighbor matching is similar to blocking if all the covariates only take discrete values (exact matching).
- But in practice, there may be many covariates, some of which are continuous.

Nearest-neighbor matching

- Nearest-neighbor matching is similar to blocking if all the covariates only take discrete values (exact matching).
- But in practice, there may be many covariates, some of which are continuous.
- We can only match each treated/untreated observation with M untreated/treated neighbors who are the nearest.

Nearest-neighbor matching

- Nearest-neighbor matching is similar to blocking if all the covariates only take discrete values (exact matching).
- But in practice, there may be many covariates, some of which are continuous.
- We can only match each treated/untreated observation with M untreated/treated neighbors who are the nearest.
- $\mathbf{X}_i$ is never identical to the average covariates of i's nearest neighbors even in large samples.

Nearest-neighbor matching

- Nearest-neighbor matching is similar to blocking if all the covariates only take discrete values (exact matching).
- But in practice, there may be many covariates, some of which are continuous.
- We can only match each treated/untreated observation with M untreated/treated neighbors who are the nearest.
- $\mathbf{X}_i$ is never identical to the average covariates of i's nearest neighbors even in large samples.
- This inaccuracy creates severe problems for statistical inference.

Nearest-neighbor matching

- Nearest-neighbor matching is similar to blocking if all the covariates only take discrete values (exact matching).
- But in practice, there may be many covariates, some of which are continuous.
- We can only match each treated/untreated observation with M untreated/treated neighbors who are the nearest.
- $\mathbf{X}_i$ is never identical to the average covariates of i's nearest neighbors even in large samples.
- This inaccuracy creates severe problems for statistical inference.
- Abadie and Imbens (2006) first derive the asymptotic distribution of the nearest-neighbor matching estimator.

Nearest-neighbor matching

- Nearest-neighbor matching is similar to blocking if all the covariates only take discrete values (exact matching).
- But in practice, there may be many covariates, some of which are continuous.
- We can only match each treated/untreated observation with M untreated/treated neighbors who are the nearest.
- $\mathbf{X}_i$ is never identical to the average covariates of i's nearest neighbors even in large samples.
- This inaccuracy creates severe problems for statistical inference.
- Abadie and Imbens (2006) first derive the asymptotic distribution of the nearest-neighbor matching estimator.
- It equals to a normal distribution plus an exponential distribution (the bias).

Nearest-neighbor matching: algorithm

- Match with M nearest neighbors; replacement is allowed; κ continuous covariates.

Nearest-neighbor matching: algorithm

- Match with M nearest neighbors; replacement is allowed; κ continuous covariates.
- We impute the counterfactual of each unit with the average of the matched units.

Nearest-neighbor matching: algorithm

- Match with M nearest neighbors; replacement is allowed; κ continuous covariates.
- We impute the counterfactual of each unit with the average of the matched units.
- For each treated observation i,

$$\hat{Y}_i(1) = \begin{cases} Y_i & D_i = 1 \\ \frac{1}{M}\sum_{j \in \mathcal{J}_M(i)} Y_j & D_i = 0, \end{cases}$$
$$\hat{Y}_i(0) = \begin{cases} \frac{1}{M}\sum_{j \in \mathcal{J}_M(i)} Y_j & D_i = 1 \\ Y_i & D_i = 0, \end{cases}$$

where $\mathcal{J}_M(i)$ is the set of units matched to i.

Nearest-neighbor matching: algorithm

- Match with M nearest neighbors; replacement is allowed; κ continuous covariates.
- We impute the counterfactual of each unit with the average of the matched units.
- For each treated observation i,

$$\hat{Y}_i(1) = \begin{cases} Y_i & D_i = 1 \\ \frac{1}{M}\sum_{j \in \mathcal{J}_M(i)} Y_j & D_i = 0, \end{cases}$$

$$\hat{Y}_i(0) = \begin{cases} \frac{1}{M}\sum_{j \in \mathcal{J}_M(i)} Y_j & D_i = 1 \\ Y_i & D_i = 0, \end{cases}$$

  where $\mathcal{J}_M(i)$ is the set of units matched to i.
- The ATE estimate using matching is

$$\hat{\tau}_M = \frac{1}{N}\sum_{i=1}^{N}(\widehat{Y}_i(1) - \widehat{Y}_i(0)).$$

Nearest-neighbor matching: inference (*)

- Denote $E\left[Y_i(D_i)|\mathbf{X}_i\right]$ as $m_{D_i}(\mathbf{X}_i)$ and $\varepsilon_i = Y_i - m_{D_i}(\mathbf{X}_i)$.

Nearest-neighbor matching: inference (*)

- Denote $E\left[Y_i(D_i)|\mathbf{X}_i\right]$ as $m_{D_i}(\mathbf{X}_i)$ and $\varepsilon_i = Y_i - m_{D_i}(\mathbf{X}_i)$.
- Each unit i might be matched to multiple other units and we denote the number as $K_M(i)$.

Nearest-neighbor matching: inference (*)

- Denote $E\left[Y_i(D_i)|\mathbf{X}_i\right]$ as $m_{D_i}(\mathbf{X}_i)$ and $\varepsilon_i = Y_i - m_{D_i}(\mathbf{X}_i)$.
- Each unit i might be matched to multiple other units and we denote the number as $K_M(i)$.
- Abadie and Imbens (2006) prove that

$$\sqrt{N}(\hat{\tau}_M - \tau) \to \mathcal{N}(B_M, V),$$

where

$$B_M = \frac{1}{N}\sum_{i=1}^{N}(2D_i - 1)\left[\frac{1}{M}\sum_{m=1}^{M}(m_{1-D_i}(\mathbf{X}_i) - m_{1-D_i}(\mathbf{X}_{j_m(i)}))\right]$$

is the bias caused by "mismatch" between $\mathbf{X}_i$ and $\mathbf{X}_{j_m(i)}$.

Nearest-neighbor matching: inference (*)

- Denote $E\left[Y_i(D_i)|\mathbf{X}_i\right]$ as $m_{D_i}(\mathbf{X}_i)$ and $\varepsilon_i = Y_i - m_{D_i}(\mathbf{X}_i)$.
- Each unit i might be matched to multiple other units and we denote the number as $K_M(i)$.
- Abadie and Imbens (2006) prove that

$$\sqrt{N}(\hat{\tau}_M - \tau) \rightarrow \mathcal{N}(B_M, V),$$

where

$$B_M = \frac{1}{N}\sum_{i=1}^{N}(2D_i - 1)\left[\frac{1}{M}\sum_{m=1}^{M}(m_{1-D_i}(\mathbf{X}_i) - m_{1-D_i}(\mathbf{X}_{j_m(i)}))\right]$$

is the bias caused by "mismatch" between $\mathbf{X}_i$ and $\mathbf{X}_{j_m(i)}$.
- It declines to zero slowly at the rate of $(O_P(N^{1/\kappa}))$.

Nearest-neighbor matching: bias

- The bias term is undesirable and makes matching inefficient.

Nearest-neighbor matching: bias

- The bias term is undesirable and makes matching inefficient.
- Because of it, bootstrap cannot be used to approximate the estimate's distribution (Abadie and Imbens 2008).

Nearest-neighbor matching: bias

- The bias term is undesirable and makes matching inefficient.
- Because of it, bootstrap cannot be used to approximate the estimate's distribution (Abadie and Imbens 2008).
- There is no need to worry about the bias if there is at most one continuous covariate.

Nearest-neighbor matching: bias

- The bias term is undesirable and makes matching inefficient.
- Because of it, bootstrap cannot be used to approximate the estimate's distribution (Abadie and Imbens 2008).
- There is no need to worry about the bias if there is at most one continuous covariate.
- Or if the following conditions hold:
  1. We are only interested in the ATT;
  2. Matching is conducted without replacement; and
  3. The number of untreated units is much larger than that of the treated ones (Abadie and Imbens 2012).

Nearest-neighbor matching: inference (*)

- Abadie and Imbens (2011): B_M can be estimated directly using nonparametric regression.

Nearest-neighbor matching: inference (*)

- Abadie and Imbens (2011): B_M can be estimated directly using nonparametric regression.
- We can replace our estimate with $\hat{\tau}_M - \hat{B}_M$.

Nearest-neighbor matching: inference (*)

- Abadie and Imbens (2011): B_M can be estimated directly using nonparametric regression.
- We can replace our estimate with $\hat{\tau}_M - \hat{B}_M$.
- This is known as the bias correction estimator for NN matching.

Nearest-neighbor matching: inference (*)

- Abadie and Imbens (2011): B_M can be estimated directly using nonparametric regression.
- We can replace our estimate with $\hat{\tau}_M - \hat{B}_M$.
- This is known as the bias correction estimator for NN matching.
- This estimator is root-N consistent and asymptotically normal.

Nearest-neighbor matching: inference (*)

- Abadie and Imbens (2011): B_M can be estimated directly using nonparametric regression.
- We can replace our estimate with $\hat{\tau}_M - \hat{B}_M$.
- This is known as the bias correction estimator for NN matching.
- This estimator is root-N consistent and asymptotically normal.
- They also provide an estimator for the variance based on jackknife.

Nearest-neighbor matching: inference (*)

- Abadie and Imbens (2011): B_M can be estimated directly using nonparametric regression.
- We can replace our estimate with $\hat{\tau}_M - \hat{B}_M$.
- This is known as the bias correction estimator for NN matching.
- This estimator is root-N consistent and asymptotically normal.
- They also provide an estimator for the variance based on jackknife.
- Otsu and Rai (2017) provided a variance estimator based on wild bootstrap.

Nearest-neighbor matching: inference (*)

- Abadie and Imbens (2011): B_M can be estimated directly using nonparametric regression.
- We can replace our estimate with $\hat{\tau}_M - \hat{B}_M$.
- This is known as the bias correction estimator for NN matching.
- This estimator is root-N consistent and asymptotically normal.
- They also provide an estimator for the variance based on jackknife.
- Otsu and Rai (2017) provided a variance estimator based on wild bootstrap.
- Lin, Ding, and Han (2023) showed that if we allow M to grow with N, the bias correction estimator is efficient.

Propensity score matching: asymptotics

- Note that the propensity score is a unidimensional continuous variable.

Propensity score matching: asymptotics

- Note that the propensity score is a unidimensional continuous variable.
- Matching on the propensity score does not induce the bias term.

Propensity score matching: asymptotics

- Note that the propensity score is a unidimensional continuous variable.
- Matching on the propensity score does not induce the bias term.
- Abadie and Imbens (2016) show that

$$\sqrt{N}(\hat{\tau}_M - \tau) \to \mathcal{N}(0, V)$$

Propensity score matching: asymptotics

- Note that the propensity score is a unidimensional continuous variable.
- Matching on the propensity score does not induce the bias term.
- Abadie and Imbens (2016) show that

$$\sqrt{N}(\hat{\tau}_M - \tau) \to \mathcal{N}(0, V)$$

- Since the propensity score is an estimate, the asymptotic variance is affected by its estimation.

Propensity score matching: asymptotics

- Note that the propensity score is a unidimensional continuous variable.
- Matching on the propensity score does not induce the bias term.
- Abadie and Imbens (2016) show that

$$\sqrt{N}(\hat{\tau}_M - \tau) \to \mathcal{N}(0, V)$$

- Since the propensity score is an estimate, the asymptotic variance is affected by its estimation.
- We must account for this extra uncertainty in variance estimation.

Propensity score matching: asymptotics

- Note that the propensity score is a unidimensional continuous variable.
- Matching on the propensity score does not induce the bias term.
- Abadie and Imbens (2016) show that

$$\sqrt{N}(\hat{\tau}_M - \tau) \to \mathcal{N}(0, V)$$

- Since the propensity score is an estimate, the asymptotic variance is affected by its estimation.
- We must account for this extra uncertainty in variance estimation.
- PS matching requires extra structural restrictions hence is less agnostic (Ho et al. 2007).

Propensity score matching: asymptotics

- Note that the propensity score is a unidimensional continuous variable.
- Matching on the propensity score does not induce the bias term.
- Abadie and Imbens (2016) show that

$$\sqrt{N}(\hat{\tau}_M - \tau) \rightarrow \mathcal{N}(0, V)$$

- Since the propensity score is an estimate, the asymptotic variance is affected by its estimation.
- We must account for this extra uncertainty in variance estimation.
- PS matching requires extra structural restrictions hence is less agnostic (Ho et al. 2007).
- But it does approximate a block randomization.

Development of the matching method

- Roadmap:
  - Abadie and Imbens (2006): asymptotic distribution for NN matching (with replacement)
  - Abadie and Imbens (2008): bootstrap doesn't work for NN matching
  - Abadie and Imbens (2011): bias correction matching estimator
  - Abadie and Imbens (2012): matching as a martingale (NN matching without replacement)
  - Diamond and Sekhon (2013): finding the optimal distance metric using evolutionary algorithm
  - Abadie and Imbens (2016): asymptotic distribution for PS matching
  - Huber et al. (2016): wild bootstrap for PS matching
  - Otsu and Rai (2017): wild bootstrap for NN matching
  - Abadie and Spiess (2021): regression after matching is valid only under the correct specification.
  - Armstrong and Kolesár (2021): NN matching is finite-sample optimal when the outcome is not smooth in covariates
  - Lin, Ding, and Han (2023): theoretical results for diverging M

Misconceptions about matching

- Matching makes your data look like a product of block randomization.

Misconceptions about matching

- Matching makes your data look like a product of block randomization.
- It does not guarantee the existence of such a randomization process.

Misconceptions about matching

- Matching makes your data look like a product of block randomization.
- It does not guarantee the existence of such a randomization process.
- It is an estimator rather than an identification assumption.

Misconceptions about matching

- Matching makes your data look like a product of block randomization.
- It does not guarantee the existence of such a randomization process.
- It is an estimator rather than an identification assumption.
- Strong ignorability is the pre-condition and cannot be made more plausible by using any estimator.

Misconceptions about matching

- Matching makes your data look like a product of block randomization.
- It does not guarantee the existence of such a randomization process.
- It is an estimator rather than an identification assumption.
- Strong ignorability is the pre-condition and cannot be made more plausible by using any estimator.
- Matching can be used just to ensure positivity by trimming units that cannot be matched (Imbens 2015).

Misconceptions about matching

- Matching makes your data look like a product of block randomization.
- It does not guarantee the existence of such a randomization process.
- It is an estimator rather than an identification assumption.
- Strong ignorability is the pre-condition and cannot be made more plausible by using any estimator.
- Matching can be used just to ensure positivity by trimming units that cannot be matched (Imbens 2015).
- But extra steps are needed to account for confounders.

Matching: application

- Let's revisit LaLonde (1986).
- We rely on the R-package *Matching* developed by Jas Sekhon.

```
## The OLS estimate is 1794.343

## The SE of OLS estimate is 670.9967

## The Lin regression estimate is 1583.468

## The SE of Lin regression estimate is 678.0574
```

Matching: application

```
##
## Estimate...  2050.5
## AI SE......  1727.8
## T-stat.....  1.1868
## p.val......  0.23532
##
## Original number of observations..............  2675
## Original number of treated obs...............  185
## Matched number of observations...............  185
## Matched number of observations  (unweighted).  201
```

Matching: application

```
##
## Estimate...  1468.7
## AI SE......  1385.5
## T-stat.....  1.06
## p.val......  0.28914
##
## Original number of observations..............  2675
## Original number of treated obs...............  185
## Matched number of observations...............  185
## Matched number of observations  (unweighted).  932
```

Matching: application

```
##              mean.Tr   mean.Co    sdiff T pval
## age           25.816    34.851 -126.266  0.000
## education     10.346    12.117  -88.077  0.000
## black          0.843     0.251  162.564  0.000
## hispanic       0.059     0.033   11.357  0.132
## married        0.189     0.866 -172.406  0.000
## nodegree       0.708     0.305   88.378  0.000
## re74        2095.574 19428.746 -354.707  0.000
## re75        1532.056 19063.338 -544.576  0.000
## u74            0.708     0.086  136.391  0.000
## u75            0.600     0.100  101.786  0.000
```

Matching: application

```
##               mean.Tr  mean.Co   sdiff T pval
## age            25.816   26.288  -6.598  0.368
## education      10.346   10.580 -11.650  0.037
## black           0.843    0.822   5.931  0.102
## hispanic        0.059    0.059   0.000  1.000
## married         0.189    0.195  -1.376  0.564
## nodegree        0.708    0.659  10.672  0.006
## re74         2095.574 2429.660  -6.837  0.004
## re75         1532.056 2251.461 -22.347  0.000
## u74             0.708    0.708   0.000  1.000
## u75             0.600    0.600   0.000  1.000
```

Matching: application

```
## Warning: glm.fit: fitted probabilities numerically 0 or
##
## Estimate...  1687.9
## AI SE......  1565.4
## T-stat.....  1.0783
## p.val......  0.28091
##
## Original number of observations..............  2675
## Original number of treated obs...............  185
## Matched number of observations...............  185
## Matched number of observations  (unweighted).  2739
```

Matching: application

```
##              mean.Tr  mean.Co   sdiff T pval
## age           25.816   24.676  15.940  0.092
## education     10.346   10.709 -18.058  0.101
## black          0.843    0.828   4.108  0.694
## hispanic       0.059    0.067  -3.269  0.767
## married        0.189    0.120  17.644  0.021
## nodegree       0.708    0.660  10.585  0.306
## re74        2095.574 2624.597 -10.826  0.233
## re75        1532.056 1862.146 -10.254  0.170
## u74            0.708    0.651  12.431  0.111
## u75            0.600    0.523  15.583  0.084
```

References I

Abadie, Alberto, and Guido W Imbens. 2006. "Large Sample Properties of Matching Estimators for Average Treatment Effects." *Econometrica* 74 (1): 235–67.

———. 2008. "On the Failure of the Bootstrap for Matching Estimators." *Econometrica* 76 (6): 1537–57.

———. 2011. "Bias-Corrected Matching Estimators for Average Treatment Effects." *Journal of Business & Economic Statistics* 29 (1): 1–11.

———. 2012. "A Martingale Representation for Matching Estimators." *Journal of the American Statistical Association* 107 (498): 833–43.

———. 2016. "Matching on the Estimated Propensity Score." *Econometrica* 84 (2): 781–807.

Abadie, Alberto, and Jann Spiess. 2021. "Robust Post-Matching Inference." *Journal of the American Statistical Association*, 1–13.

References II

Armstrong, Timothy B, and Michal Kolesár. 2021. "Finite-Sample Optimal Estimation and Inference on Average Treatment Effects Under Unconfoundedness." *Econometrica* 89 (3): 1141–77.

Diamond, Alexis, and Jasjeet S Sekhon. 2013. "Genetic Matching for Estimating Causal Effects: A General Multivariate Matching Method for Achieving Balance in Observational Studies." *Review of Economics and Statistics* 95 (3): 932–45.

Ho, Daniel E, Kosuke Imai, Gary King, and Elizabeth A Stuart. 2007. "Matching as Nonparametric Preprocessing for Reducing Model Dependence in Parametric Causal Inference." *Political Analysis* 15 (3): 199–236.

Huber, Martin, Lorenzo Camponovo, Hugo Bodory, and Michael Lechner. 2016. "A Wild Bootstrap Algorithm for Propensity Score Matching Estimators." Université de Fribourg.

Imbens, Guido W. 2015. "Matching Methods in Practice: Three Examples." *Journal of Human Resources* 50 (2): 373–419.

References III

LaLonde, Robert J. 1986. "Evaluating the Econometric Evaluations of Training Programs with Experimental Data." *The American Economic Review*, 604–20.

Lin, Zhexiao, Peng Ding, and Fang Han. 2023. "Estimation Based on Nearest Neighbor Matching: From Density Ratio to Average Treatment Effect." *Econometrica* 91 (6): 2187–217.

Otsu, Taisuke, and Yoshiyasu Rai. 2017. "Bootstrap Inference of Matching Estimators for Average Treatment Effects." *Journal of the American Statistical Association* 112 (520): 1720–32.