Quant II

Lab 4: Conditioning: Matching, Weighing, and Sensitivity Analysis

Junlong Aaron Zhou

February 26, 2021

Outline

- Blocking and rerandomization
- Matching
  - Why matching?
  - Various algorithms
  - Asymptotics of matching
- IPW
  - Why do we love/hate it?
  - Covariate balancing
- What if confounders are unobservable?
  - Placebo
  - Sensitivity

Blocking

- Blocking: covariates adjustment before assignment
- Usually results in more efficient estimates
- Easier to get balance in covariates
- What is the optimal blocking algorithm?

Rerandomization

- What if your first draw leads to imbalance in covariates?

Rerandomization

- What if your first draw leads to imbalance in covariates?
- Rubin: draw the assignment again and do not tell anybody!

Rerandomization

- What if your first draw leads to imbalance in covariates?
- Rubin: draw the assignment again and do not tell anybody!
- But what is the distribution of the ATE estimates?
- Ding, Li and Rubin (2017): A truncated Gaussian distribution
- Rerandomization can be combined with regression adjustment

Why matching?

- To approximiate a blocking experiment
- To get rid of model dependence
  Matching is completely nonparametric: $\hat{\tau}_i = Y_i - \sum_{\mathcal{M}_i} Y_{i \in \mathcal{M}_i}$.

Why matching?

- To approximiate a blocking experiment
- To get rid of model dependence
  Matching is completely nonparametric: $\hat{\tau}_i = Y_i - \sum_{\mathcal{M}_i} Y_{i \in \mathcal{M}_i}$.
- To estimate heterogeneous treatment effects
  Straightforward.

Why matching?

- To approximiate a blocking experiment
- To get rid of model dependence
  Matching is completely nonparametric: $\hat{\tau}_i = Y_i - \sum_{\mathcal{M}_i} Y_{i \in \mathcal{M}_i}$.
- To estimate heterogeneous treatment effects
  Straightforward.
- To guarantee common support (positivity)
  Suppose we estimate τ using Lin's approach, then,

$$\hat{\tau} = \bar{Y}_1 - \bar{Y}_0 - (\frac{N_0}{N_0 + N_1} * \hat{\beta}_1 + \frac{N_1}{N_0 + N_1} * \hat{\beta}_0)'(\bar{X}_1 - \bar{X}_0)$$

(Imbens and Wooldridge, 2009)
Bias disappears only when $\bar{X}_1 = \bar{X}_0$ (LaLonde, 1986).

Why matching?

- To approximiate a blocking experiment
- To get rid of model dependence
  Matching is completely nonparametric: $\hat{\tau}_i = Y_i - \sum_{\mathcal{M}_i} Y_{i \in \mathcal{M}_i}$.
- To estimate heterogeneous treatment effects
  Straightforward.
- To guarantee common support (positivity)
  Suppose we estimate τ using Lin's approach, then,

$$\hat{\tau} = \bar{Y}_1 - \bar{Y}_0 - (\frac{N_0}{N_0 + N_1} * \hat{\beta}_1 + \frac{N_1}{N_0 + N_1} * \hat{\beta}_0)'(\bar{X}_1 - \bar{X}_0)$$

  (Imbens and Wooldridge, 2009)
  Bias disappears only when $\bar{X}_1 = \bar{X}_0$ (LaLonde, 1986).
- Matching cannnot help you get unconfoundedness.

Basic steps

1. Choose a distance metric

Basic steps

1. Choose a distance metric
2. Find matches on your set of covariates/propensity scores, and get rid of non-matches

Basic steps

1. Choose a distance metric
2. Find matches on your set of covariates/propensity scores, and get rid of non-matches (Warning!)

Basic steps

1. Choose a distance metric
2. Find matches on your set of covariates/propensity scores, and get rid of non-matches (Warning!)
3. Check balance in your matched data set

Basic steps

1. Choose a distance metric
2. Find matches on your set of covariates/propensity scores, and get rid of non-matches (Warning!)
3. Check balance in your matched data set
4. Repeat these steps until your set exhibits acceptable balance

Basic steps

1. Choose a distance metric
2. Find matches on your set of covariates/propensity scores, and get rid of non-matches (Warning!)
3. Check balance in your matched data set
4. Repeat these steps until your set exhibits acceptable balance
5. Calulate the ATT/ATE on your matched dataset

An example

- Boyd et al. (2010)
- The effect of gender on decision making
- Unit of analysis: the appellate court case
- Treatment: whether there is at least one female in the three judge panel
- Covariates: median ideology, median age, one racial minority, indicator for ideological direction of lower court's decision, indicator for whether a majority of the judges were nominated by Republicans, indicator for whether a majority of the judges on the panel had judicial experience prior to their nomination

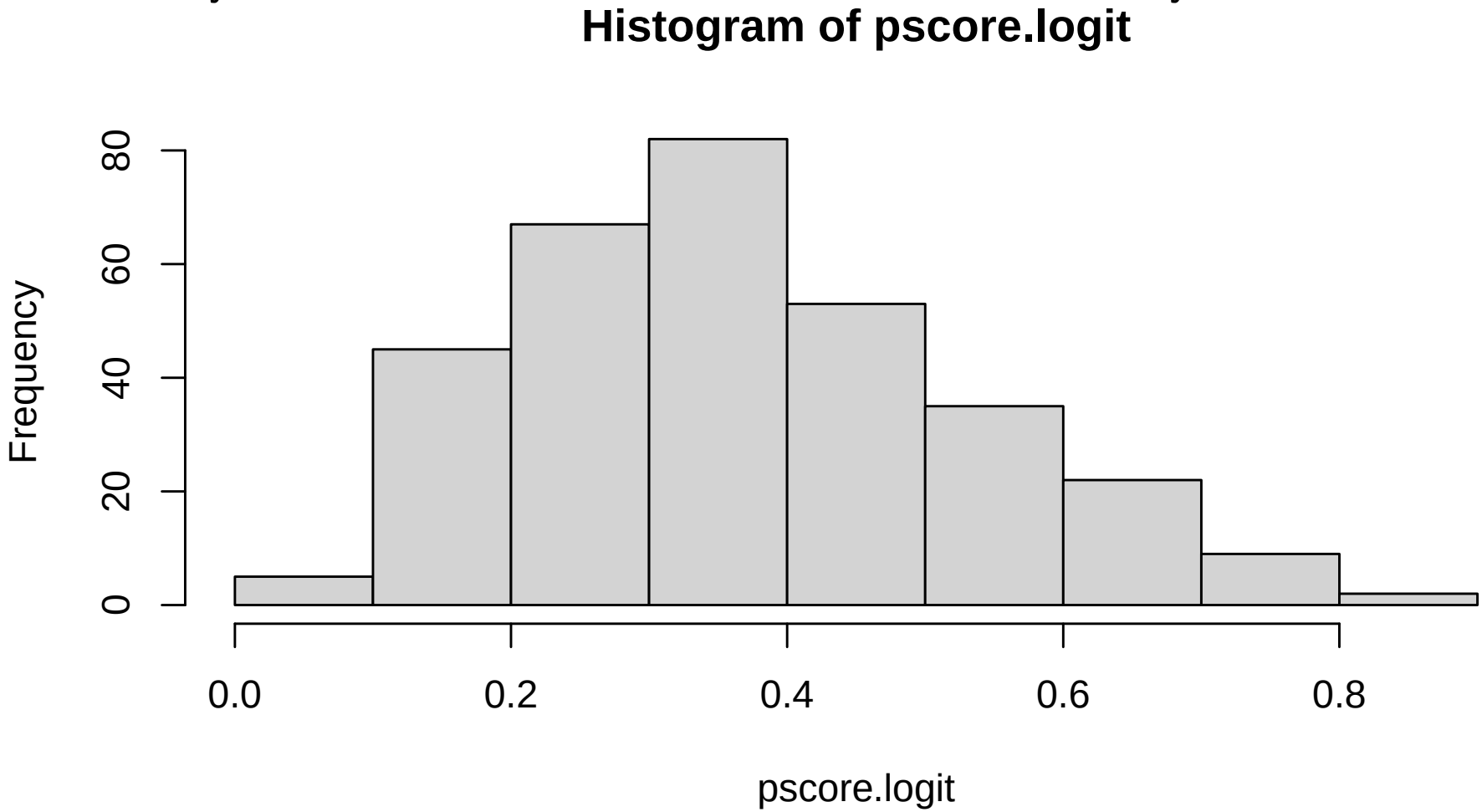
View Initial Balance

```
initial.balance <- round(t(rbind(means,t.p,ks.p)),digits=3)[c(2:8),]
initial.balance
```

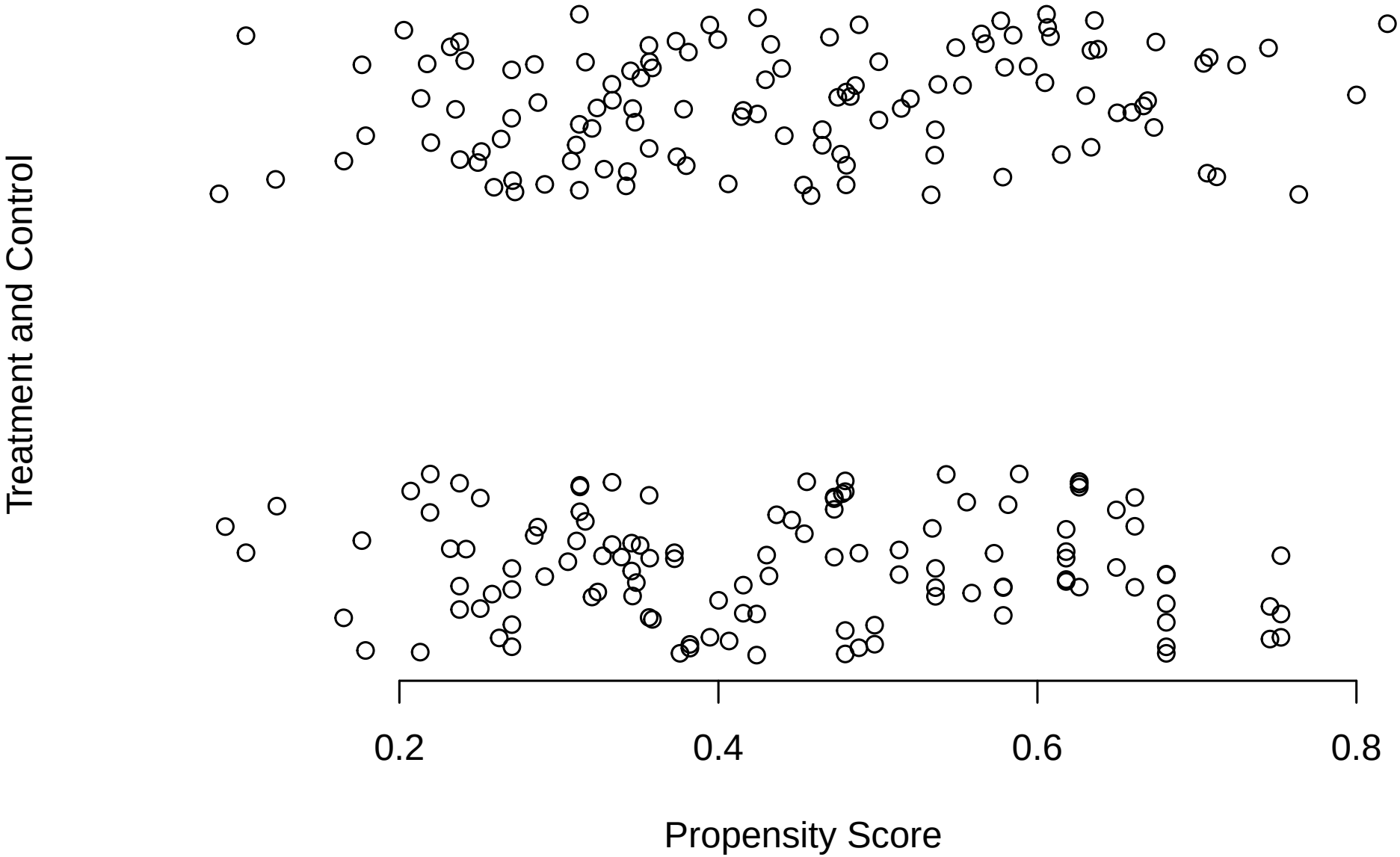
```
##                          Treated Control   t.p  ks.p
## median.ideo                0.900   0.802 0.008 0.006
## repub.majority             0.752   0.585 0.002 0.030
## has.minority               0.233   0.212 0.666 1.000
## maj.experienced            0.446   0.373 0.202 0.826
## median.age                63.416  60.483 0.000 0.001
## liberal.lower.direction    0.208   0.161 0.292 0.997
## liberalOutcome             0.366   0.424 0.314 0.967
```

Propensity score matching

- Pros: reduce the number of dimensions
- Cons: may not use information in the most efficient way



Propensity score matching



Nearest Neighbor Matching

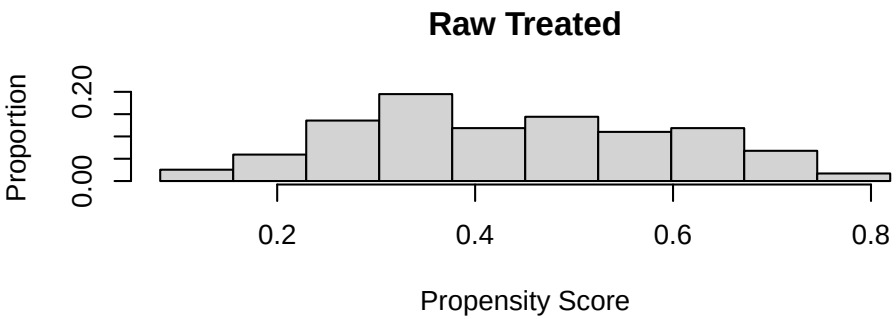
- Approximate a blocking experiment
- You can also use `MatchIt`

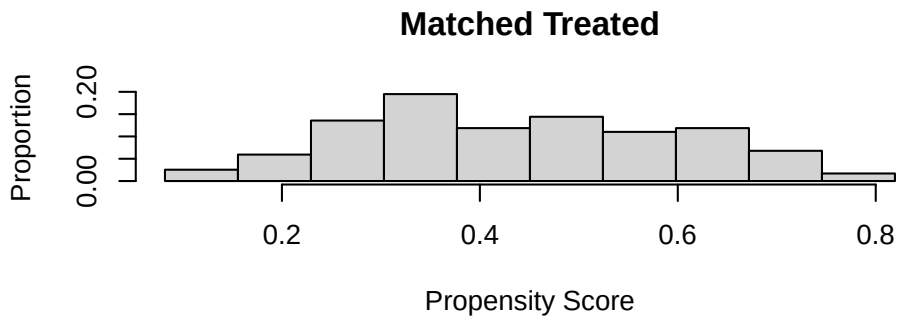
Nearest Neighbor Matching

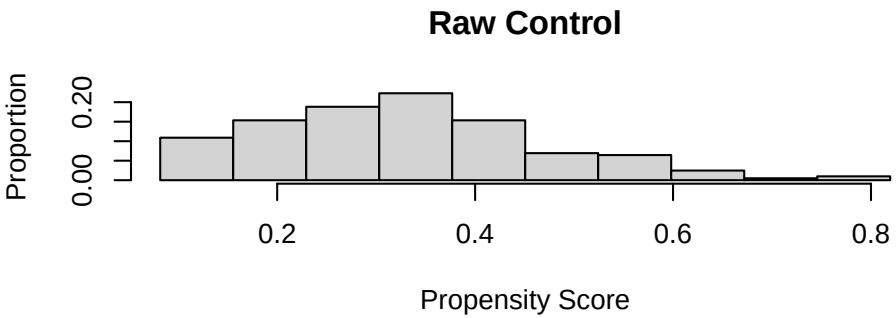
```
##                          Means.Treated Means.Control      t.NN
## median.ideo                  0.8024492     0.8995520 0.2161134
## median.age                  60.4830508    63.4158416 0.4253905
## repub.majority               0.5847458     0.7524752 1.0000000
## has.minority                 0.2118644     0.2326733 0.5952787
## maj.experienced              0.3728814     0.4455446 0.2003073
## liberal.lower.direction      0.1610169     0.2079208 0.8586550
```

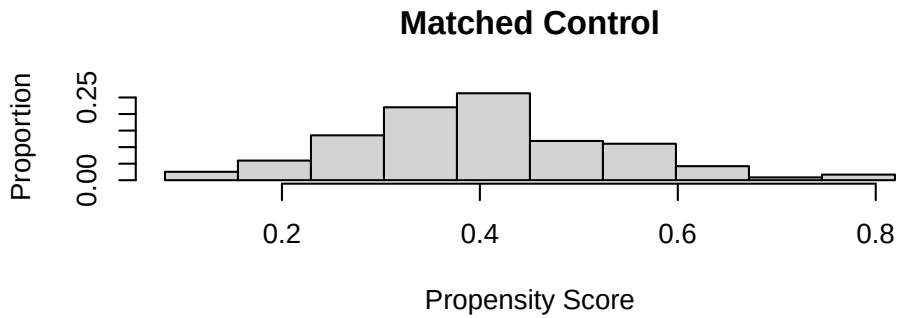
Nearest Neighbor Matching

```
plot(matched.NN, type="histogram")
```









Genetic Matching

- Set an objective function and update the distance metric iteratively

$$\sqrt{(X_i - X_j)'(S^{-1/2})'WS^{-1/2}(X_i - X_j)}$$

- Based upon evolutionary algorithm
- It is very slow (especially if you choose a reasonable `pop.size`)
- Can also do it with `Matching` or `GenMatch`

Genetic Matching

```
##                          Control  Treat   t.p
## median.ideo                0.825  0.802 0.422
## repub.majority             0.585  0.585 0.552
## has.minority               0.212  0.212 0.265
## maj.experienced            0.390  0.373 0.803
## median.age                60.822 60.483 0.581
## liberal.lower.direction    0.144  0.161 0.288
```

- And then you can calculate the effect of interest

CEM

- CEM creates bins along each covariate dimension (either pre-specified or automatic)
- Units lying in the same strata are then matched together
- Curse of dimensionality means that with lots of covariates, we'll only rarely have units in the same strata.

CEM

```
library(cem)
cem.match <- cem(treatment = "has.woman",
                 data = d.new,
                 drop = "liberalOutcome")
```

```
##
## Using 'has.woman'='1' as baseline group
```

```
cem.match
```

```
##              G0  G1
## All         202 118
## Matched      15  18
## Unmatched 187 100
```

CEM

- Hopefully you are lucky and you have more units
- If not, just tweak CEM

```
cutpoints <- list( median.ideo=c(0.3,0.5,0.7),
                   median.age= c(60,65))
cem.tweak.match <- cem(treatment = "has.woman",
                       data = d.new,
                       drop = "liberalOutcome", cutpoints = cutp
```

```
##
## Using 'has.woman'='1' as baseline group
```

```
cem.tweak.match
```

```
##                G0  G1
## All           202 118
## Matched        83  74
## Unmatched     119  44
```

Asymptotics of Matching

- Matching creates extra uncertainty (why?)
- What is the real standard error of $\hat{\tau}$?

Asymptotics of Matching

- Matching creates extra uncertainty (why?)
- What is the real standard error of $\hat{\tau}$?
- Roadmap:
  - Abadie and Imbens (2006): asymptotic distribution for NN matching (with replacement)
  - Abadie and Imbens (2011): debiased matching estimator
  - Abadie and Imbens (2008): boostrap doesn't work for matching
  - Abadie and Imbens (2012): matching as a martingale (NN without replacement)
  - Abadie and Imbens (2016): asymptotic distribution for PS matching
  - Otsu and Rai (2017): wild bootstrap for NN matching
  - Bodory et al. (2018): wild bootstrap for PS matching

Asymptotics of NN Matching

- Denote $E\left[Y_i(D_i)|X_i\right]$ as $\mu_{D_i}(X_i)$, then $Y_i = \mu_{D_i}(X_i) + \epsilon_i$
- Match with K nearest neighbors; replacement is allowed; covariates can be continuous

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

Asymptotics of NN Matching

- Denote $E\left[Y_i(D_i)|X_i\right]$ as $\mu_{D_i}(X_i)$, then $Y_i = \mu_{D_i}(X_i) + \epsilon_i$
- Match with K nearest neighbors; replacement is allowed; covariates can be continuous

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

- The bias from NN matching can be decomposed into three parts:

$$\hat{\tau}_M - \tau = \overline{\tau(X)} - \tau + E_M + B_M$$

where

$$\overline{\tau(X)} = \frac{1}{N}\sum_{i=1}^{N}(\mu_1(X_i) - \mu_0(X_i))$$

and

$$E_M = \frac{1}{N}\sum_{i=1}^{N}(2D_i - 1)(1 + \frac{K_M(i)}{M})\epsilon_i$$

Asymptotics of NN Matching

- Abadie and Imbens (2006) show that both $\overline{\tau(X)}$ (difference in conditional expectations) and E_M (sum of residuals) are asymptotically unbiased.
- However, the conditional bias relative to $\overline{\tau(X)}$,

$$B_M = \frac{1}{N}\sum_{i=1}^{N}(2D_i - 1)\left[\frac{1}{M}\sum_{m=1}^{M}\left(\mu_{1-D_i}(X_i) - \mu_{1-D_i}(X_{j_m(i)})\right)\right]$$

  is not.

Asymptotics of NN Matching

- Abadie and Imbens (2006) show that both $\overline{\tau(X)}$ (difference in conditional expectations) and E_M (sum of residuals) are asymptotically unbiased.
- However, the conditional bias relative to $\overline{\tau(X)}$,

$$B_M = \frac{1}{N}\sum_{i=1}^{N}(2D_i - 1)\left[\frac{1}{M}\sum_{m=1}^{M}\left(\mu_{1-D_i}(X_i) - \mu_{1-D_i}(X_{j_m(i)})\right)\right]$$

  is not.
- The bias caused by "mismatch"; it declines very slowly.
- The speed depends on the number of continuous covariates.
- B_M actually converges to an exponential distribution.
- We may estimate B_M directly using the serial estimator proposed by Newey (1995).
- Take-away: do not use bootstrap for NN matching!

About IPW

- It is actually the Horvitz-Thompson estimator.

About IPW

- It is actually the Horvitz-Thompson estimator.
- The duality of IPW and propensity score matching suggests two basic ways of conducting causal inference:
  - Adjust the response surface (matching, regression)
  - Adjust the assignment probability (weighting)
  - Either approach returns unbiased estimate
  - We can combine them to obtain doubly robustness

About IPW

- It is actually the Horvitz-Thompson estimator.
- The duality of IPW and propensity score matching suggests two basic ways of conducting causal inference:
  - Adjust the response surface (matching, regression)
  - Adjust the assignment probability (weighting)
  - Either approach returns unbiased estimate
  - We can combine them to obtain doubly robustness
- IPW can be extended to panel data (dynamic treatment regime).

The benefits of IPW

- Hirano et al. (2003): the variance of IPW estimators can reach the Cramer-Rao lower bound
- What if we use the real propensity score?

The benefits of IPW

- Hirano et al. (2003): the variance of IPW estimators can reach the Cramer-Rao lower bound
- What if we use the real propensity score?
- The variance will be larger! (Hahn, 1998)
- Empirical propensity scores take into account all the actual imbalances in the sample

Caveats for IPW

- It behaves poorly at the "tail" of the support

Caveats for IPW

- It behaves poorly at the "tail" of the support
- One solution is to stabilize it using the Hajek estimator

Caveats for IPW

- It behaves poorly at the "tail" of the support
- One solution is to stabilize it using the Hajek estimator
- Another solution is to drop data at the tail part
- Changes the quantity of interest

Caveats for IPW

- It behaves poorly at the "tail" of the support
- One solution is to stabilize it using the Hajek estimator
- Another solution is to drop data at the tail part
- Changes the quantity of interest
- Ma and Wang (2019): asymptotic distribution for both trimed/untrimed IPW
- They also provide a bias correction method based on resampling

```
##                                0          1
## median.ideo            0.8995520  0.8024492
## repub.majority         0.7524752  0.5847458
## has.minority           0.2326733  0.2118644
## maj.experienced        0.4455446  0.3728814
## median.age            63.4158416 60.4830508
## liberal.lower.direction 0.2079208  0.1610169
## has.woman              0.0000000  1.0000000
## liberalOutcome         0.3663366  0.4237288
```

CBPS and covariates balancing

- IPW balances the covariates, then why don't we direct seek for balance?
- What can you do when the treatment is continuous?

CBPS and covariates balancing

- IPW balances the covariates, then why don't we direct seek for balance?
- What can you do when the treatment is continuous?
- Imai and Ratkovic (2013); Fong, Hazlett and Imai (2018): Covariate Balancing Propensity Score
- Idea: find weights that are orthogonal to X, D, and their interaction

$$\sum_{i}^{N} w_i(X_i^*, D_i^*, X_i^* * D_i^*) = 0, \ \sum_{i}^{N} w_i = N$$

CBPS and covariates balancing

- IPW balances the covariates, then why don't we direct seek for balance?
- What can you do when the treatment is continuous?
- Imai and Ratkovic (2013); Fong, Hazlett and Imai (2018): Covariate Balancing Propensity Score
- Idea: find weights that are orthogonal to X, D, and their interaction

$$\sum_{i}^{N} w_i(X_i^*, D_i^*, X_i^* * D_i^*) = 0, \ \sum_{i}^{N} w_i = N$$

- Hainmueller (2012); Hazlett (2015); Arbour and Dimmery (2019)
- Various forms of convex optimization

CBPS and covariates balancing

```
## [1] "Finding ATT with T=1 as the treatment.  Set ATT=2 to f

##                                    0          1
## median.ideo                0.8995520  0.8024492
## repub.majority             0.7524752  0.5847458
## has.minority               0.2326733  0.2118644
## maj.experienced            0.4455446  0.3728814
## median.age                63.4158416 60.4830508
## liberal.lower.direction    0.2079208  0.1610169
## has.woman                  0.0000000  1.0000000
## liberalOutcome             0.3663366  0.4237288

## Converged within tolerance

##                                    0          1
## median.ideo                0.8995520  0.8024492
## repub.majority             0.7524752  0.5847458
## has.minority               0.2326733  0.2118644
```

Placebo test

- There is no fixed way of conducting placebo test
- Find some variable/observation that should not be affected by the treatment

Placebo test

- There is no fixed way of conducting placebo test
- Find some variable/observation that should not be affected by the treatment
- Will the weight of your friends affect yours? How about height?

Placebo test

- There is no fixed way of conducting placebo test
- Find some variable/observation that should not be affected by the treatment
- Will the weight of your friends affect yours? How about height?
- More common in panel data analysis

Sensitivity analysis

- The basic idea: How influential unobservable confounders have to be to make the estimate insignificant/zero?

Sensitivity analysis

- The basic idea: How influential unobservable confounders have to be to make the estimate insignificant/zero?
- Remember that confounders must be correlated with both D and Y
- Vary the two correlation coefficients and check how the estimate would change
- Compare the correlation coefficients against observable confounders
- Methods differ in their assumptions on the DGP

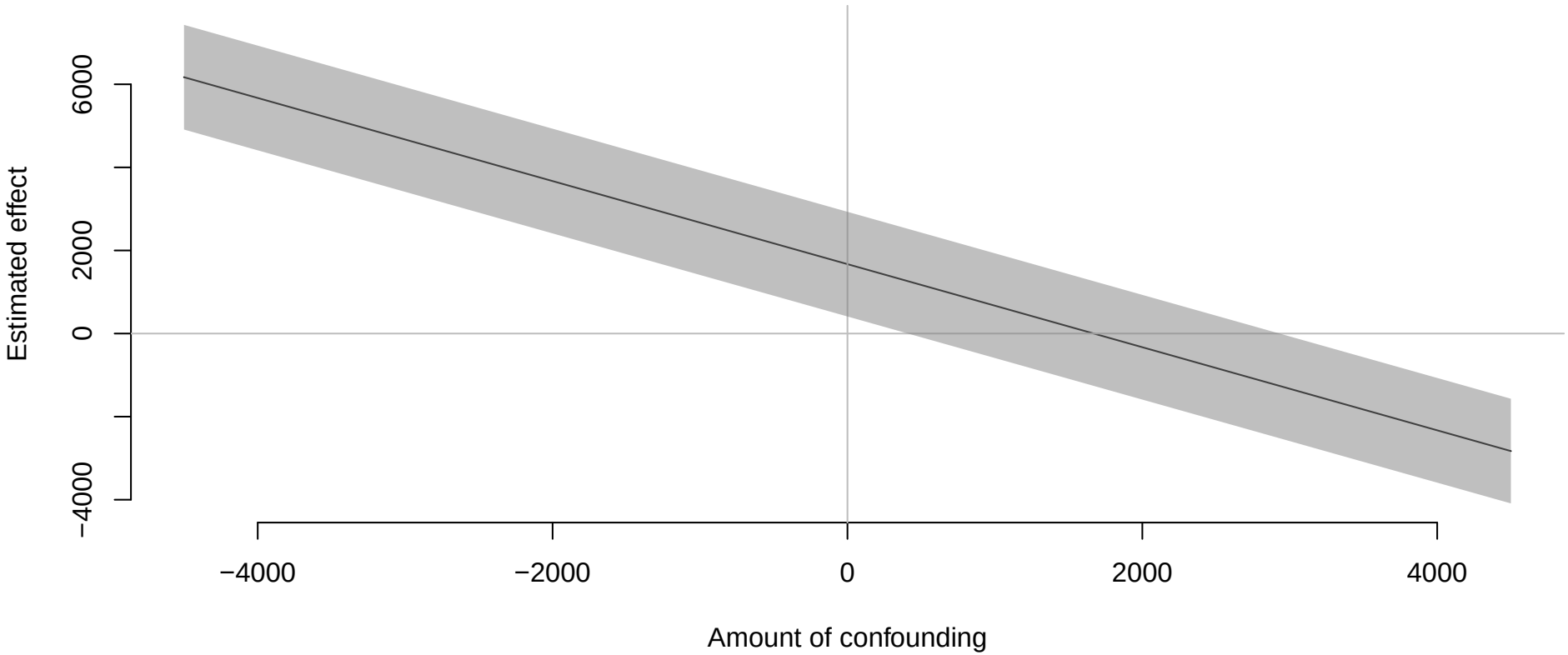
Sensitivity analysis

- First proposed by Rosenbaum and Rubin (1983)
- Imbens (2003): Full parametric model
- Blackwell (2013): Measure selection bias
- Dorie et al. (2016): Semi-parametric test using BART
- Cinelli and Hazlett (2020): Sensitivity from the OVB perspective

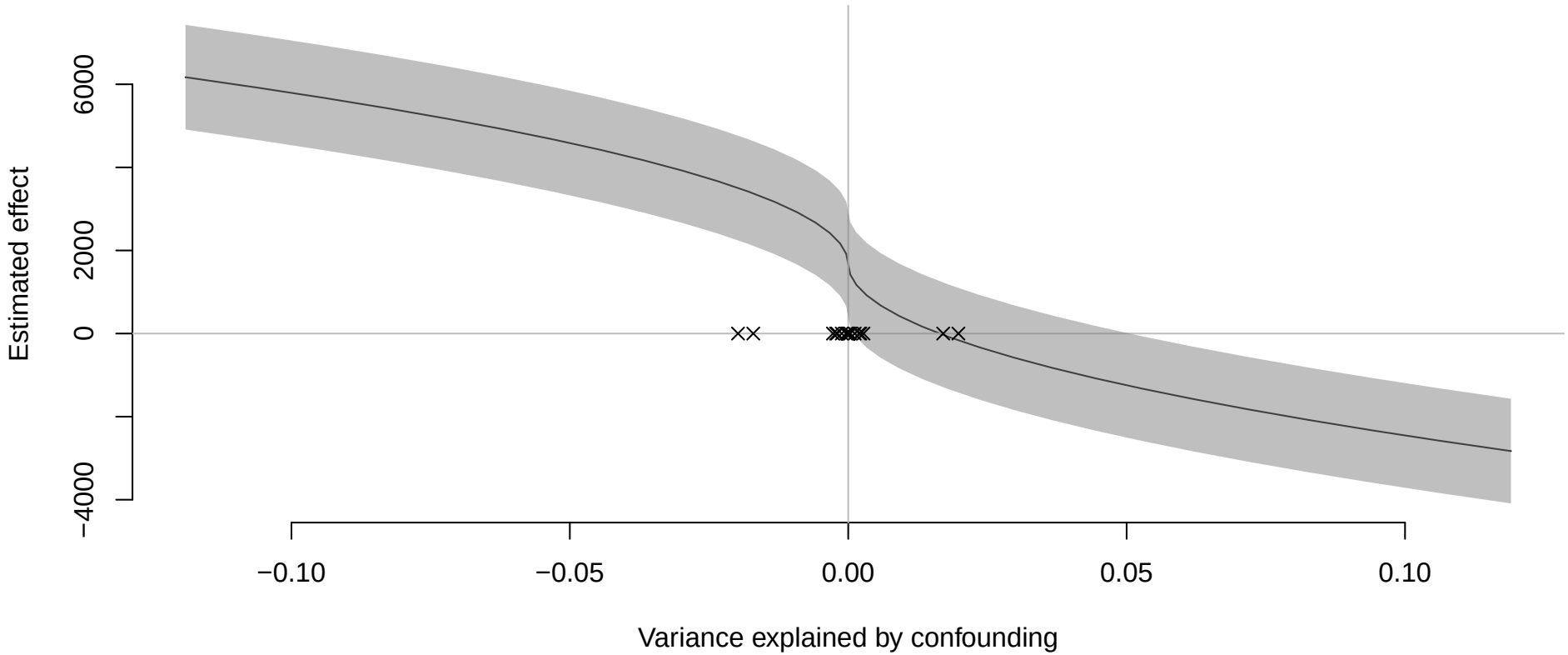
Blackwell (2013)

- Instead of imagining specific uni or multivariate omitted variable, imagine a function which defines the confounding.
- $q(d, x) = E[Y_i(d)|D_i = d, X_i = x] - E[Y_i(d)|D_i = 1 - d, X_i = x]$
- If q is positive units in group d have a higher mean potential outcome under d that those in group $1 - d$.
- So q encodes the selection bias of treatment assignment: it models violations of ignorability.
- After all, confounding means that potential outcomes vary by treatment status.
- Now, $Y_i^q = Y_i - qPr(1 - D_i|X_i)$ and we can redo the analysis.
- Package on CRAN: `causalsens`

Blackwell (2013)



Blackwell (2013)



Cinelli and Hazlett (2020)

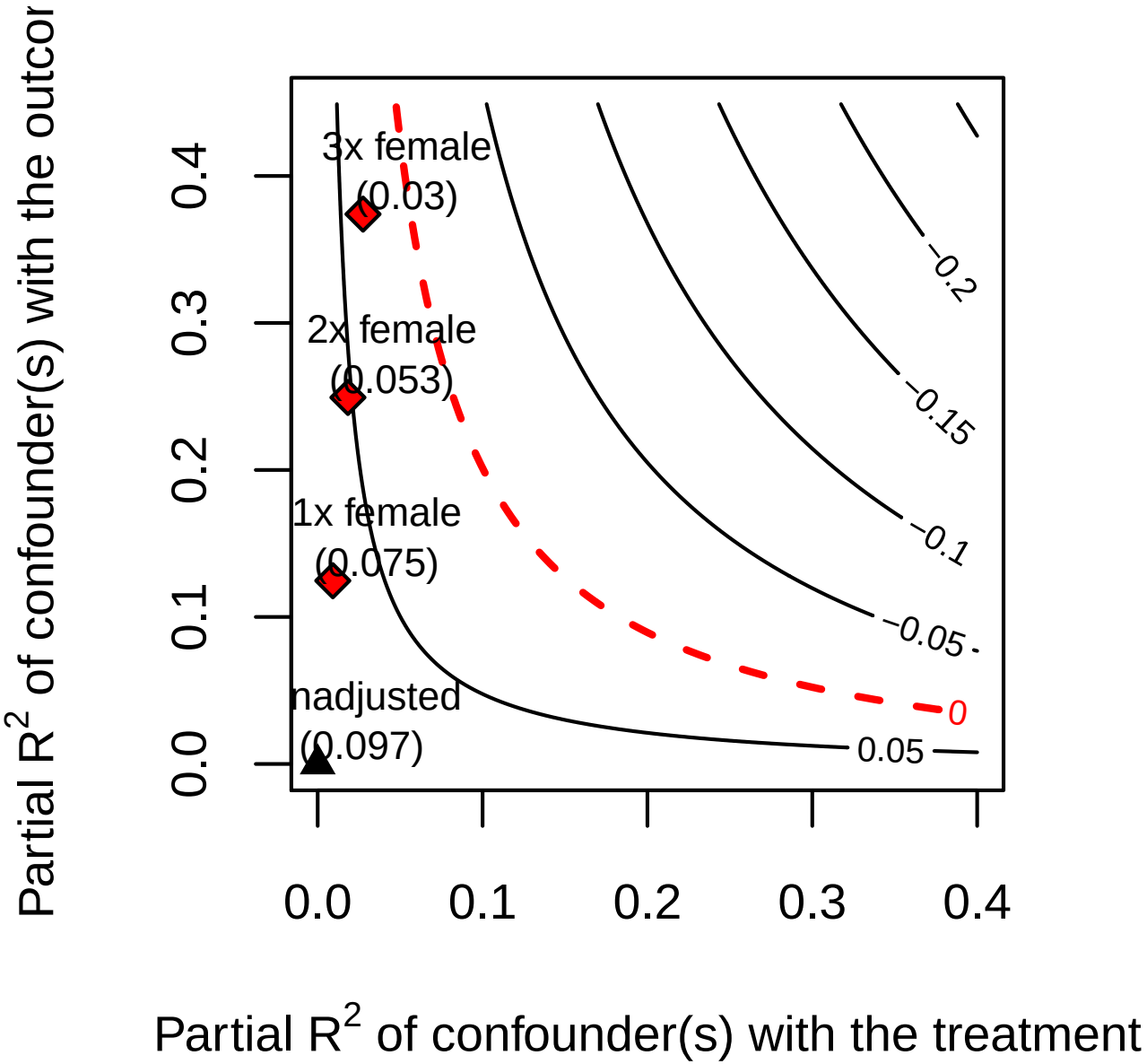
- Sensitivity from the omitted variable bias perspective
- Suppose the correct model is $Y = \hat{\tau} D + \mathbf{X}\hat{\beta} + \hat{\gamma} Z + \hat{\varepsilon}_{full}$
- But Z is unobservable
- So the real model is $Y = \hat{\tau}_{res} D + \mathbf{X}\hat{\beta}_{res} + \hat{\varepsilon}_{res}$
- It is easy to see:

$$
\begin{aligned}
\hat{\tau}_{res} =& \frac{Cov(D^{\perp \mathbf{X}}, Y^{\perp \mathbf{X}})}{Var(D^{\perp \mathbf{X}})} \\
=& \frac{Cov(D^{\perp \mathbf{X}}, \hat{\tau} D^{\perp \mathbf{X}} + \hat{\gamma} Z^{\perp \mathbf{X}})}{Var(D^{\perp \mathbf{X}})} \\
=& \hat{\tau} + \hat{\gamma} \frac{Cov(D^{\perp \mathbf{X}}, Z^{\perp \mathbf{X}})}{Var(D^{\perp \mathbf{X}})} \\
=& \hat{\tau} + \hat{\gamma}\hat{\delta}
\end{aligned}
$$

Cinelli and Hazlett (2020)

- The difference between the correct estimate $\hat{\tau}$ and the real estimate $\hat{\tau}_{res}$ has two parts:
  - $\hat{\gamma}$: the impact of the unobservable
  - $\hat{\delta}$: the imbalance of the unobservable
- Essentially, the estimate is robust to model misspecification when both Y and D can be largely explained by the observable covariates
- The idea could be extended to the nonlinear case where we use R^2 to measure the explanatory power of observable covariates
- Model misspecification is not dependent on the sample size

Cinelli and Hazlett (2020)



Cinelli and Hazlett (2020)