

# Quant II

## Matching and Weighting

Ye Wang

2/28/2018

# Outline

- ▶ Last homework
- ▶ Matching
  - ▶ Why matching?
  - ▶ Various algorithms
  - ▶ Asymptotics of matching
- ▶ IPW
  - ▶ Why do we love/hate it?
  - ▶ CBPS and PW

## Why matching?

- ▶ 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 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 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  $\tau$  using Lin's approach, then,

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

(Imbens and Wooldridge, 2009)

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

# Why matching?

- ▶ To 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  $\tau$  using Lin's approach, then,

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

(Imbens and Wooldridge, 2009)

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

- ▶ Matching cannot 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. Calculate 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

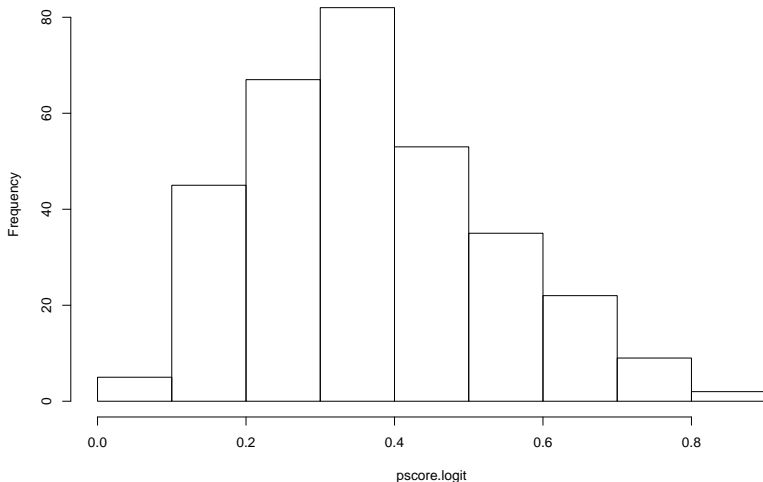
```
initial.balance <- round(t(rbind(means,t.p,ks.p)),digits=3)[c(2:8),]  
initial.balance
```

##	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

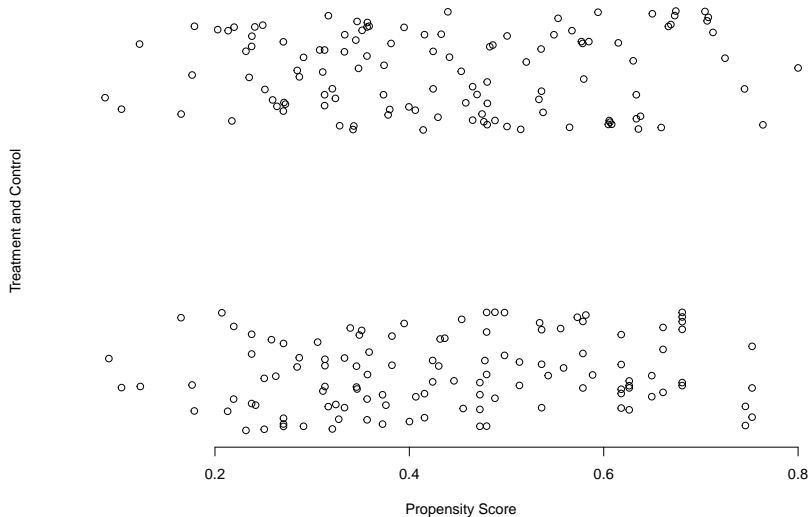
# Propensity score matching

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

Histogram of pscore.logit



# Propensity score matching





# Nearest Neighbor Matching

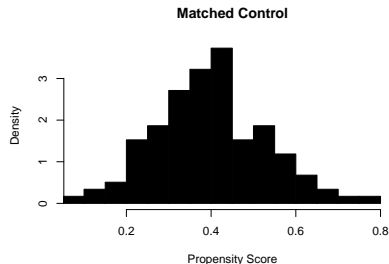
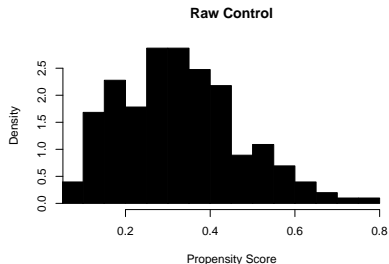
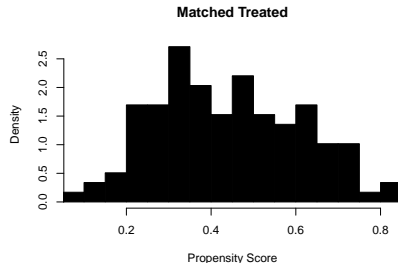
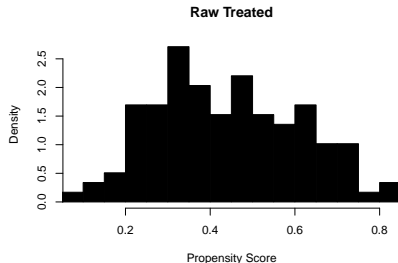
- ▶ Approximate a blocking experiment
- ▶ You can also use `MatchIt`

# 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="hist", col = "black", breaks = 20)
```



# 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 MatchIt or GenMatch

```
## Warning in ks.test(x[trt], x[!trt]): p-value will be approximate in  
## presence of ties
```

```
## Warning in ks.test(x[trt], x[!trt]): p-value will be approximate in  
## presence of ties
```

```
## Warning in ks.test(x[trt], x[!trt]): p-value will be approximate in  
## presence of ties
```

```
## Warning in ks.test(x[trt], x[!trt]): p-value will be approximate in  
## presence of ties
```

```
## Warning in ks.test(x[trt], x[!trt]): p-value will be approximate in
```

# 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

```
## Loading required package: tcltk
```

```
## Loading required package: lattice
```

```
##
```

```
## How to use CEM? Type vignette("cem")
```

```
##           GO  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 = cutpoints)  
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): bootstrap 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  $E[Y_i(D_i)|X_i]$  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

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

## Asymptotics of NN Matching

- ▶ Denote  $E[Y_i(D_i)|X_i]$  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

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

- ▶ 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

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

## 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,

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

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,

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

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!

# 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



# 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
- ▶ If we run regression with IPW: Doubly robust estimation (Robins et al., 1994)

# 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
- ▶ If we run regression with IPW: Doubly robust estimation (Robins et al., 1994)
- ▶ Can be extended to panel data (dynamic treatment regime)

## 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 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 drop data at the tail part
- ▶ Changes the quantity of interest
- ▶ Ma and Wang (2019): asymptotic distribution for both trimmed/untrimmed IPW
- ▶ They also provide a bias correction method based on resampling

## CBPS and PW

- ▶ What can you do when the treatment is continuous?

## CBPS and PW

- ▶ 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

$$\sum_i^N w_i (X_i^*, D_i^*, X_i^* * D_i^*) = 0, \quad \sum_i^N w_i = N$$

## CBPS and PW

- ▶ 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

$$\sum_i^N w_i(X_i^*, D_i^*, X_i^* * D_i^*) = 0, \quad \sum_i^N w_i = N$$

- ▶ Arbour and Dimmery (2019): use bootstrap to extract the information contained in propensity score

$$w_i = \frac{P(C_i = 1|D_i, X_i)}{P(C_i = 0|D_i, X_i)}$$

The two probabilities are obtained via machine learning



# An application of CBPS

- ▶ Wang and Wong (2018)

# An application of CBPS

- ▶ Wang and Wong (2018)
- ▶ TA's QP

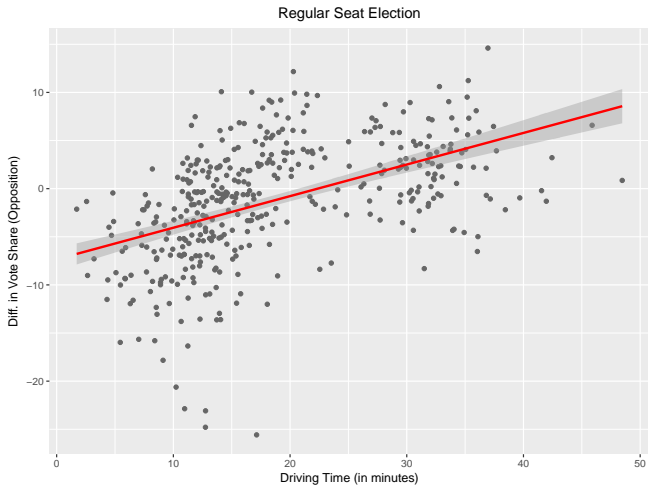
# An application of CBPS

- ▶ Wang and Wong (2018)
- ▶ TA's QP
- ▶ Does Hong Kong's Umbrella Movement reduce people's support for the opposition?

# An application of CBPS

- ▶ Wang and Wong (2018)
- ▶ TA's QP
- ▶ Does Hong Kong's Umbrella Movement reduce people's support for the opposition?

# An application of CBPS



Driven by protest exposure, or other features of the central city?

# An application of CBPS

