Section 7 : Matching II

Andrew Bertoli

16 October 2013

Roadmap

1. Regression Discontinuity (wrap-up)
2. Mahalanobis Distance
3. Matching
4. Balance Tests
5. Questions

Regression Discontinuity (wrap-up)

A common test to assess whether there was sorting in the data is the McCrary Sorting Test

You can run it using the DCdensity function in the 'rdd' package.

The paper is available here: `http://ideas.repec.org/a/eee/econom/v142y2008i2p698-714.html`

It is not appropriate for paired RDs or designs where you would expect scores around the cut-point to follow something other than the uniformly distributed.

It can miss sorting that is very close to the cut-point.

Regression Discontinuity (wrap-up)

Questions

Mahalanobis Distance

Basic Definition

$$md(\mathbf{X_i}, \mathbf{X_j}) = [(\mathbf{X_i} - \mathbf{X_j})^\top \mathbf{S}^{-1} (\mathbf{X_i} - \mathbf{X_j})]^{1/2}$$

where $\mathbf{X_i}$ and $\mathbf{X_j}$ are the covariates for Unit i and Unit j and $\mathbf{S}$ is the sample covariance matrix for $\mathbf{X}$.

Mahalanobis Distance

Example

Say we have the following dataset:

```
> data
                 Age Height
Kobe Bryant       35     78
Michelle Obama    49     71
Einstein          56     68
Billy              6     42
```

Mahalanobis Distance

Example

We can compute the Mahalanobis Distance for any two people as follows:

```
> S=var(data)
>
> S
             Age   Height
Age    489.6667 271.8333
Height 271.8333 247.5833
>
> K.M=as.numeric(data["Kobe",]-data["Michelle",])
>
> MD.Kobe.Michelle=sqrt(t(K.M)%*%solve(S)%*%(K.M))
>
> MD.Kobe.Michelle
         [,1]
[1,] 1.630141
```

Mahalanobis Distance

Kobe-Michelle: 1.630141

Kobe-Einstein: 2.398428

Kobe-Billy: 2.411238

Michelle-Einstein: 0.7684585

Michelle-Billy: 2.012

Einstein-Billy: 2.266593

Recall that these comparisons are based only on age and height.

Mahalanobis Distance

A faster function for getting the Mahalanobis distances is available in the R code. It returns the squares of the distances.

```
> fastPwMahal(as.matrix(data))
          1         2         3
2 2.6573612
3 5.7524553 0.5905284
4 5.8140665 4.0481441 5.1374444
```

Mahalanobis Distance

Matching on Mahalanobis distance is Affinely Invariant

This means that we would have gotten the same matches had we measured height in centimeters rather than inches.

Mahalanobis Distance

If your controls are all normally distributed (more precisely, follow and elliptic distribution) and your sample size is large enough, then matching on Mahalanobis distance has the Equal Percent Bias Reduction (EPBR) property.

This means that matching will not make balance on any covariate worse.

It does **not** mean that your estimated treatment effect will be less biased. It could be more biased if you do not have the right X's.

Mahalanobis Distance

Mahalanobis distance is just a distance metric.

We can use it to match on one dimension when we have a lot of covariates.

All the other matching decisions are still there. For instance, we have to decide if we will match with replacement, use a caliper, match one to one, ect.

Matching

Basic Facts

1. Matching does not mean that you will get better balance on the covariates that you match on.

2. Getting better balance on your controls does not mean that your bias will decrease

3. Matching is susceptible to attenuation bias

Matching

Mahalanobis Distance Example

Say we gave Einstein and Billy a treatment. Based on Mahalanobis distance, we would match both to Michelle Obama rather than Kobe. So we have

Mean difference in age before matching= $\frac{56+6}{2} - \frac{35+49}{2}$= -11

Mean difference in age after matching= $\frac{56+6}{2} - \frac{49+49}{2}$= -18

The tradeoff is that balance on height got better.

Mean difference in height before matching= $\frac{68+42}{2} - \frac{78+71}{2}$= -19.5

Mean difference in height after matching= $\frac{68+42}{2} - \frac{71+71}{2}$= -16

Matching

Kobe-Michelle: 1.630141

Kobe-Einstein: 2.398428

Kobe-Billy: 2.411238

Michelle-Einstein: 0.7684585

Michelle-Billy: 2.012

Einstein-Billy: 2.266593

Recall that these comparisons are based only on age and height.

Matching

Mahalanobis Distance Example

Say we gave Einstein and Billy a treatment. Based on Mahalanobis distance, we would match both to Michelle Obama rather than Kobe. So we have

Mean difference in age before matching= $\frac{56+6}{2} - \frac{35+49}{2}$= -11

Mean difference in age after matching= $\frac{56+6}{2} - \frac{49+49}{2}$= -18

The tradeoff is that balance on height got better.

Mean difference in height before matching= $\frac{68+42}{2} - \frac{78+71}{2}$= -19.5

Mean difference in height after matching= $\frac{68+42}{2} - \frac{71+71}{2}$= -16

Matching

Mahalanobis Distance Example

The problem here is that our controls do not each follow an elliptic distribution.

Matching

Propensity Score Example

Recall the hypothetical study about whether eating fast food every day causes heart disease.

We assumed that we knew the true model of the propensity score.

```
> pscore=glm(Treat  Age + Gender + Parents.Eaters, family=
binomial(link=logit),data=data)$fitted.values
```

Matching

Propensity Score Example

What happens if we have a misspecified model of the propensity score.

> pscore=glm(Treat Age + Gender + Parents.Eaters + Age * Gender + Age * Parents.Eaters + Age * Parents.Eaters , family= binomial(link=logit),data=data)$fitted.values

All we did here was control for interactions.

Matching

```
> t.test(t$Gender,c$Gender[Controls],paired=TRUE)

    Paired t-test

data:  t$Gender and c$Gender[Controls]
t = -2.846, df = 52, p-value = 0.006321
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.32171271 -0.05564578
sample estimates:
mean of the differences
             -0.1886792
```

Matching

```
> t.test(t$Parents.Eaters,c$Parents.Eaters[Controls],paired=TRUE)

    Paired t-test

data:  t$Parents.Eaters and c$Parents.Eaters[Controls]
t = 1.7664, df = 52, p-value = 0.08321
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.00770038  0.12090793
sample estimates:
mean of the differences
              0.05660377
```

Matching

Why does this happen?

1. You have decided to balance the propensity score rather than the individual covariates.

2. If you know the true propensity score, it should be random within each pair who got treated. So randomization should balance the groups on the control variables.

3. If you got the propensity score wrong, you do not know how matching on it will effect the covariates.

4. If the groups are not balanced after matching on the propensity score, than you probably have the wrong propensity score. If the groups are balanced, then you might or might not have the right propensity score, but your design passes this test.

Matching

Basic Facts

1. Matching does not mean that you will get better balance on the covariates that you match on.

2. Getting better balance on your controls does not mean that your bias will decrease

3. Matching is susceptible to attenuation bias

Matching

Example of Better Balance Increasing Bias

Imagine that there is a summer program that high school students can take. Enrollees tend to be (1) harder working and (2) younger.

At the end of the following school year, there are 20 awards given to students. Awards tend to be given to students who are (1) harder working and (2) older.

We want to estimate the effect of the summer program on the likelihood of winning an award. Imagine there is no effect. We match on age.

Before matching, the treatment group will have younger students and harder working students, so the bias will partly cancel.

After matching on age, the treatment group will tend to have harder workers, but not younger students. So the hard work bias will no longer be mitigated by the age bias.

Matching

Basic Facts

1. Matching does not mean that you will get better balance on the covariates that you match on.

2. Getting better balance on your controls does not mean that your bias will decrease

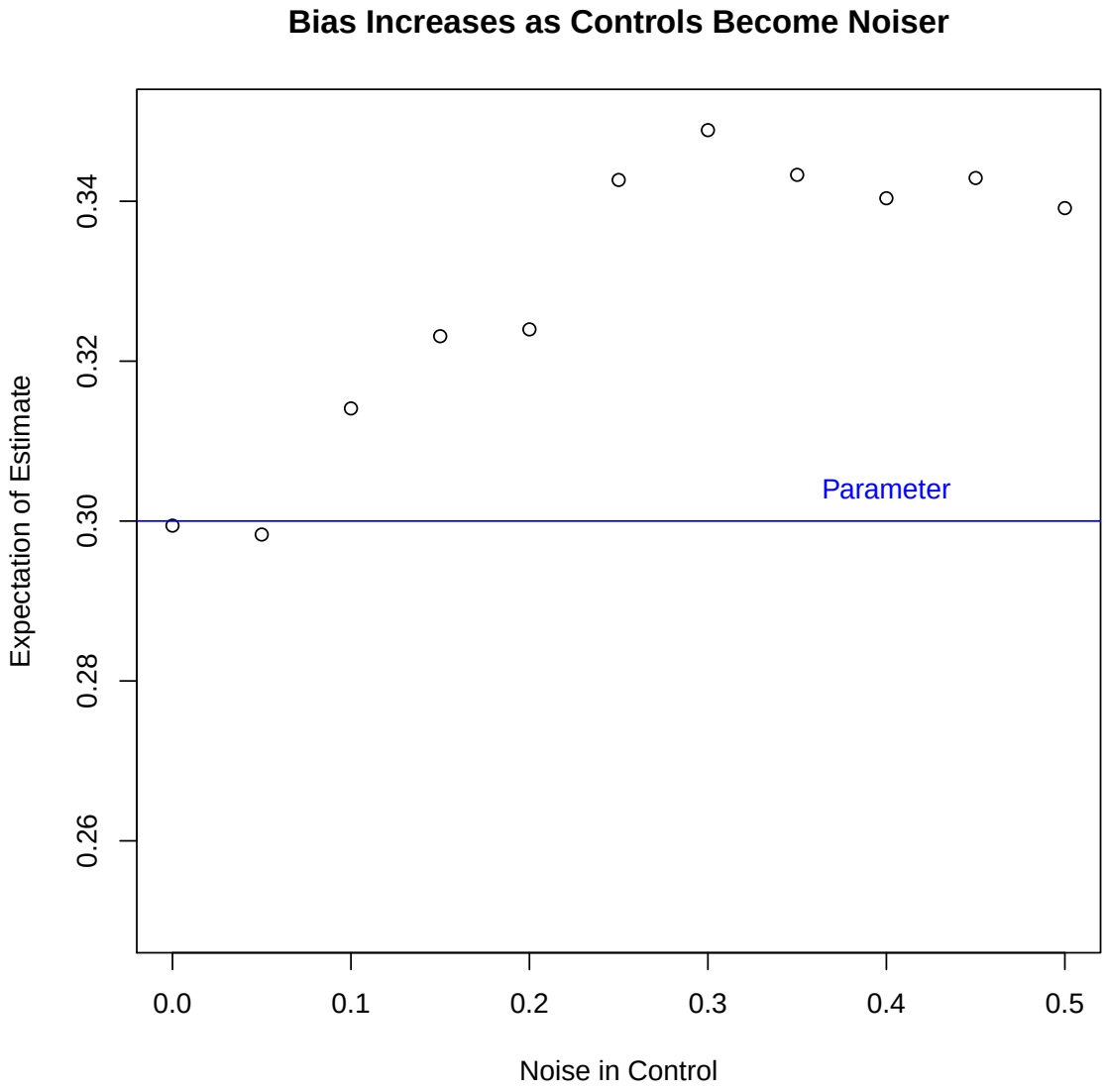
3. Matching is susceptible to attenuation bias

Matching

Question: Assuming that everything else with matching went right, what happens if there is random error in one of your controls?

Answer: It biases your results.

Matching


Bias Increases as Controls Become Noiser
Expectation of Estimate
0.26
0.28
0.30
0.32
0.34
Parameter
0.0
0.1
0.2
0.3
0.4
0.5
Noise in Control

Matching

Potential Problems with Matching

1. There might not be support in the data.

2. There might be support, but you chose the wrong X's.

3. You might have support and the right X's, but your formula for the propensity score is wrong (if you are doing propensity score matching) or your controls do not each follow an elliptic distribution (if you are doing Mahalanobis distance matching).

4. Everything else worked, but there was noise in your controls.

5. Everything worked perfectly, but people will still be skeptical or think that you p-hacked.

Matching

Genetic Matching solves one of these problems.

1. There might not be support in the data.

2. There might be support, but you chose the wrong X's.

3. You might have support and the right X's, but your formula for the propensity score is wrong (if you are doing propensity score matching) or your controls do not each follow an elliptic distribution (if you are doing Mahalanobis distance matching).

4. Everything else worked, but there was noise in your controls.

5. Everything worked perfectly, but people will still be skeptical or think that you p-hacked.

Matching

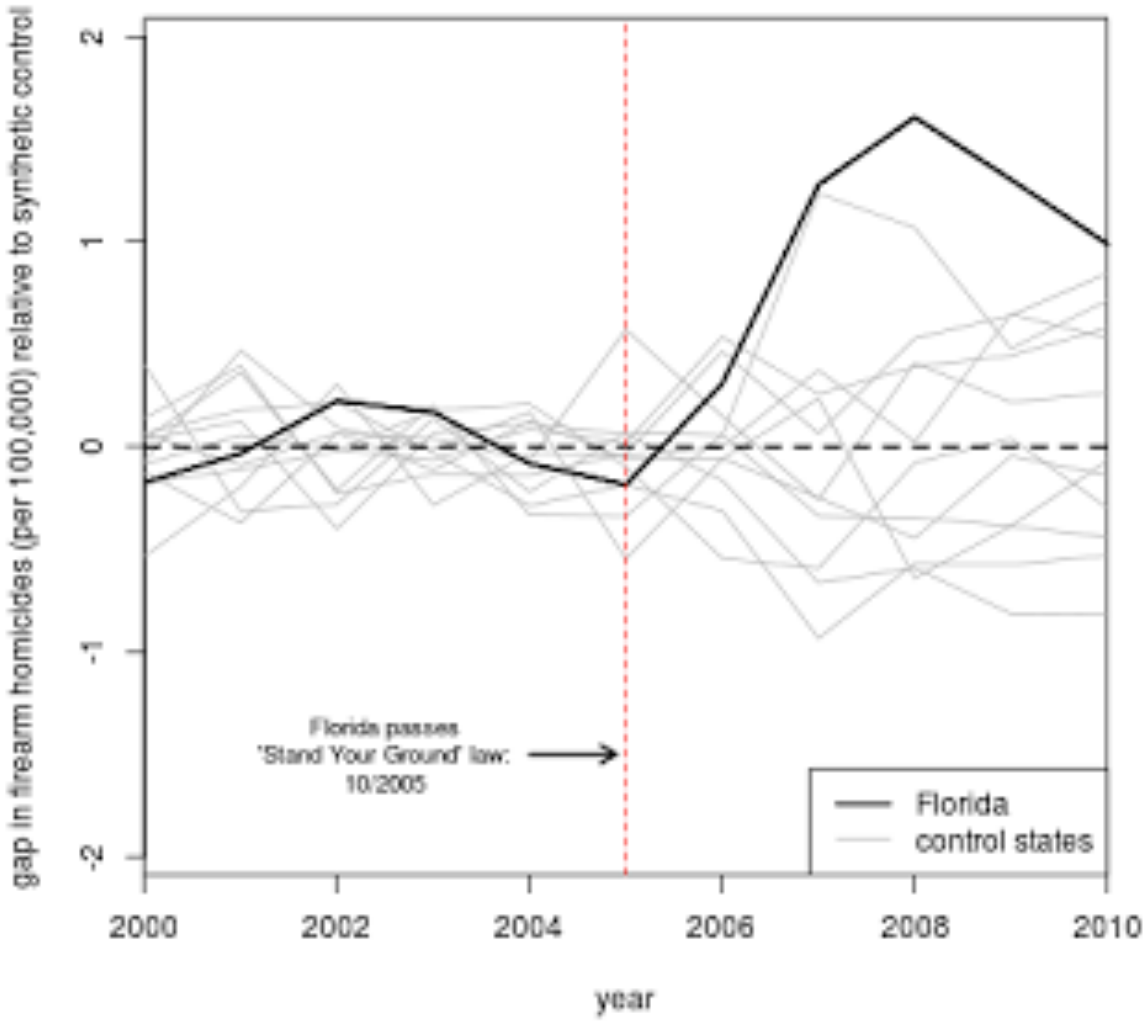
Suggestions

1. As a placebo test, redo the matching without the previous outcome and test the previous outcome. The two groups should be balanced on the previous outcome.

2. If possible, plot your outcome variable as a function of time before and after treatment. The treated and control units should look similar before treatment, but diverge afterwards (remember to also do this without matching on the previous outcome).

3. In general, matching studies that just show better balance and post-treatment tests should not be trusted.

Matching


gap in firearm homicides (per 100,000) relative to synthetic control
2
1
0
-1
-2
2000
2002
2004
2006
2008
2010
year
Florida passes
'Stand Your Ground' law:
10/2005
Florida
control states

Balance Tests

Common Types

1. T-tests
2. Equivalence tests
3. KS Tests

Balance Tests

T-tests

Advantage: Well known. Should alway be included.

Disadvantages: The null is that the groups are the same. Only test for similarity in the mean of the treatment and control group

Balance Tests

T-tests

```
library(stats)
```

If our data is not paired

```
t.test(x=treatment$Age, y=control$Age)
```

If our data is paired

```
differences=treatment$Age-control$Age
t.test(x=treatment$Age, y=control$Age, paired=TRUE)
```

Balance Tests

Equivalence Tests

Advantages: Use the null that the groups are different. Power of detecting similarity increases as sample size increases.

Disadvantages: Less well known. Make the tests to assess the covariates different from the tests to assess the outcome.

Balance Tests

Equivalence Tests

```
library(equivalence)
```

If our data is not paired

```
tost(x=treatment$Age, y=control$Age, alpha=0.05,
epsilon=sd(c(treatment$Age, control$Age)))
```

If our data is paired

```
differences=treatment$Age-control$Age
tost(x=differences, alpha=0.05, epsilon=0.2*sd(differences))
```

Balance Tests

KS-tests (Basic Idea)

The test statistic is the maximum distance between the empirical CDFs of the treatment and control distributions.

Balance Tests

KS-tests

Advantage: Compare the entire distribution of the treatment and control group.

Disadvantages: Less well known. The null is that the groups are the same. Make the tests to assess the covariates different from the tests to assess the outcome.

Balance Tests

KS-tests

```
library(stats)

ks.test(x=treatment$Age, y=control$Age)
```

Balance Tests

Suggestions

1. In general, you should check for balance using all of these tests.

2. Even if you have balance on your previous outcome, make sure that you check for balance on other important covariates.

3. Remember that achieving balance does not mean that you decreased the bias of your estimates.

Matching

Remember that good balance only solves Problem 3. It also suggests that we do not need to worry about Problem 1, at least for these control variables.

1. There might not be support in the data.

2. There might be support, but you chose the wrong X's.

3. You might have support and the right X's, but your formula for the propensity score is wrong (if you are doing propensity score matching) or your controls do not each follow an elliptic distribution (if you are doing Mahalanobis distance matching).

4. Everything else worked, but there was noise in your controls.

5. Everything worked perfectly, but people will still be skeptical or think that you p-hacked.

Matching

Questions

1. Lectures
2. Readings
3. Homework