ECN525_HW2

Rhodes Kirkpatrick

5/13/2022

1: A regression discontinuity looks at a boundary and provides an extremely reliable counter factual as the exploitation of the exogenous characteristic signals a causal effect between otherwise similar objects. In this case, a normal regression wouldn't work because the fixed boundary on the Huai will yield an effect that a regression will smooth over from the nature of the larger span of observations over the variables. If the Huai cutoff (but any discrete boundary) yields an effect by itself there should be a difference. We are trying to show something at a fixed point, not a continuous relationship, therefore we need an RDD. Latitude itself (without exogenous manipulation) shouldn't be correlated with higher pollution, but cities north of the Huai should have more pollution and there should be the biggest difference between groups because of the policy.

2: The assignment variable is "X", where the "c" is the Huai river. Cities to the North have an X>c and to the south, X<c, thus being north of the river when X>c is analogous to having a model where a treatment "D" is equal to 1. In this context the assignment variable is degrees north of the Huai. The outcome variable is the PM concentration.

3: A binned scatter plot aggregates similar data points into a "bin" (think a regular dot representing more than one observation) and then plots the summary of all those observations onto the 2D plot. The best part of it is that it adds a third, I call it a faux, dimension that can convey density, size, or concentration. The bin dots do this by changing size and thus you can project three dimensions onto a two dimensional scatter plot. Additionally, it looks a lot neater when given many data points to aggregate them into bins and allows a much clearer visualization of trend (Also 3D plots get too messy).

```
#Load Packages + Read Data
library(pacman)
p_load(haven, foreign, dplyr,tidyverse,tidyfast,modelsummary,rdrobust,ggplot2,rddtools)
#Read in Data
HR = read.dta('huairiver.dta')
```

```
#Binned ScatterPlot
rdplot(HR$pm10,HR$dist_huai,,c=0,p=2,nbins=15,binselect = "esmv",y.label = "PM10 Concentration",x.label = "Degree
Difference from Huai",title = "PM10 as a Function of Degrees North of the Huai River")
```
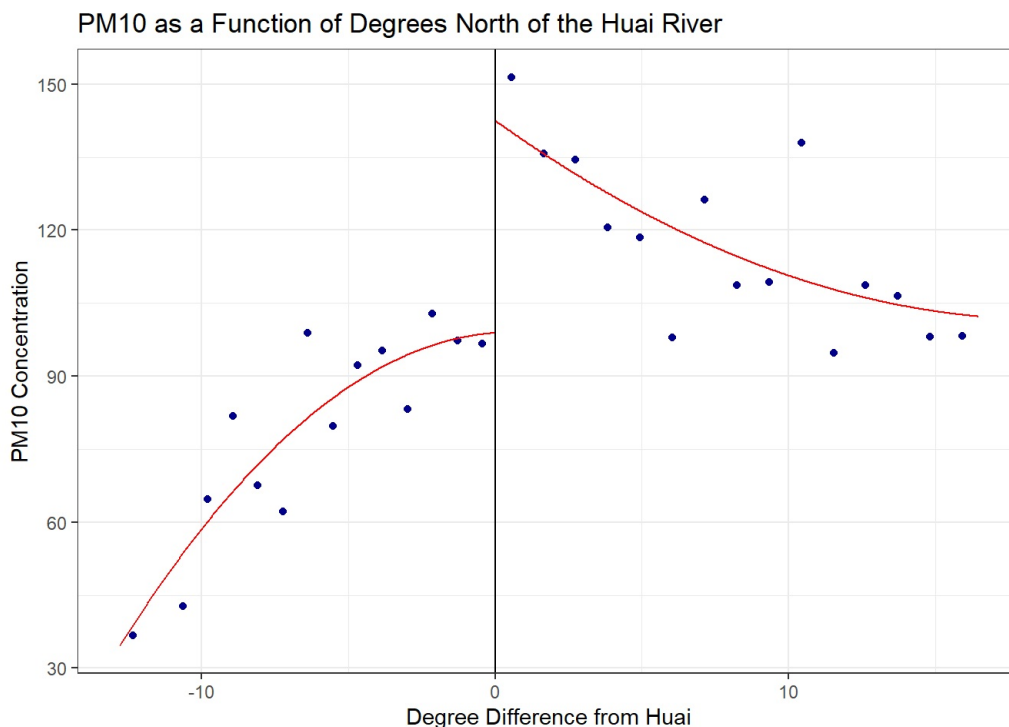

4b&5: Here, we will do the same RD design with wind speed, temperature, and precipitation as the dependent variables. I will group the graphs and regressions (questions 4b&5) together by dependent variable.

```
rdplot(HR$wspd,HR$dist_huai,,c=0,p=1,nbins=15,binselect = "esmv",y.label = "Wind Speed",x.label = "Degree Differe
nce from Huai",title = "Wind Speed as a Function of Degrees North of the Huai River")
```

Wind Speed as a Function of Degrees North of the Huai River
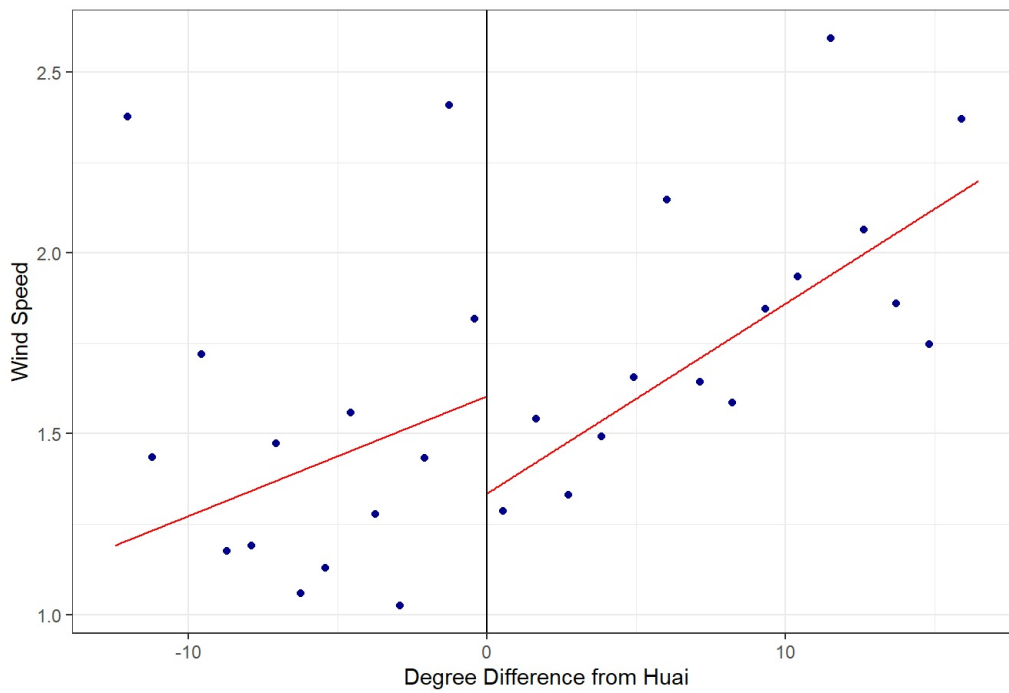

```
rdrobust(HR$wspd,HR$dist_huai,c=0,p=1)
```
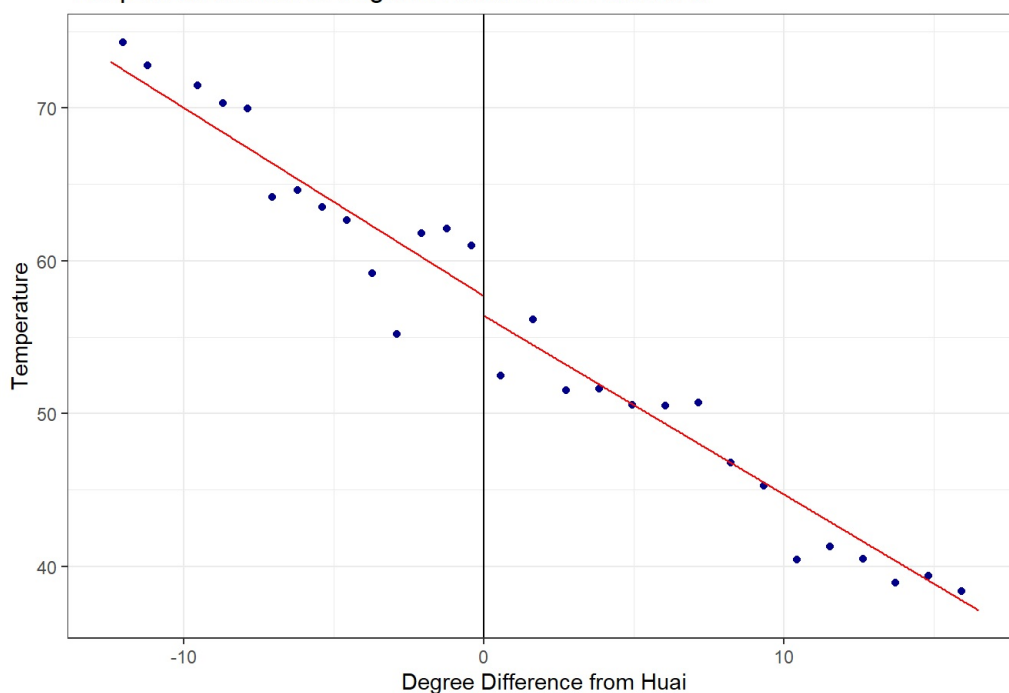
```
## Call: rdrobust
##
## Number of Obs.                    156
## BW type                         mserd
## Kernel                     Triangular
## VCE method                         NN
##
## Number of Obs.               78           78
## Eff. Number of Obs.          31           20
## Order est. (p)                1            1
## Order bias  (q)               2            2
## BW est. (h)               3.125        3.125
## BW bias (b)               4.655        4.655
## rho (h/b)                 0.671        0.671
## Unique Obs.                  78           78
```

```
rdplot(HR$temp,HR$dist_huai,,c=0,p=1,nbins=15,binselect = "esmv",y.label = "Temperature",x.label = "Degree Differ
ence from Huai",title = "Temp as a Function of Degrees North of the Huai River")
```

Temp as a Function of Degrees North of the Huai River

```
rdrobust(HR$temp,HR$dist_huai,,c=0,p=1)
```

```
## Call: rdrobust
##
## Number of Obs.                 153
## BW type                      mserd
## Kernel                  Triangular
## VCE method                      NN
##
## Number of Obs.            76          77
## Eff. Number of Obs.       38          29
## Order est.  (p)            1           1
## Order bias  (q)            2           2
## BW est. (h)            3.699       3.699
## BW bias (b)            5.970       5.970
## rho (h/b)             0.620       0.620
## Unique Obs.              76          77
```

```
rdplot(HR$prcp,HR$dist_huai,c=0,p=2,nbins=15,binselect = "esmv",y.label = "Precipitation",x.label = "Degree Diffe
rence from Huai",title = "Precipitation as a Function of Degrees North of the Huai River")
```


Precipitation as a Function of Degrees North of the Huai River
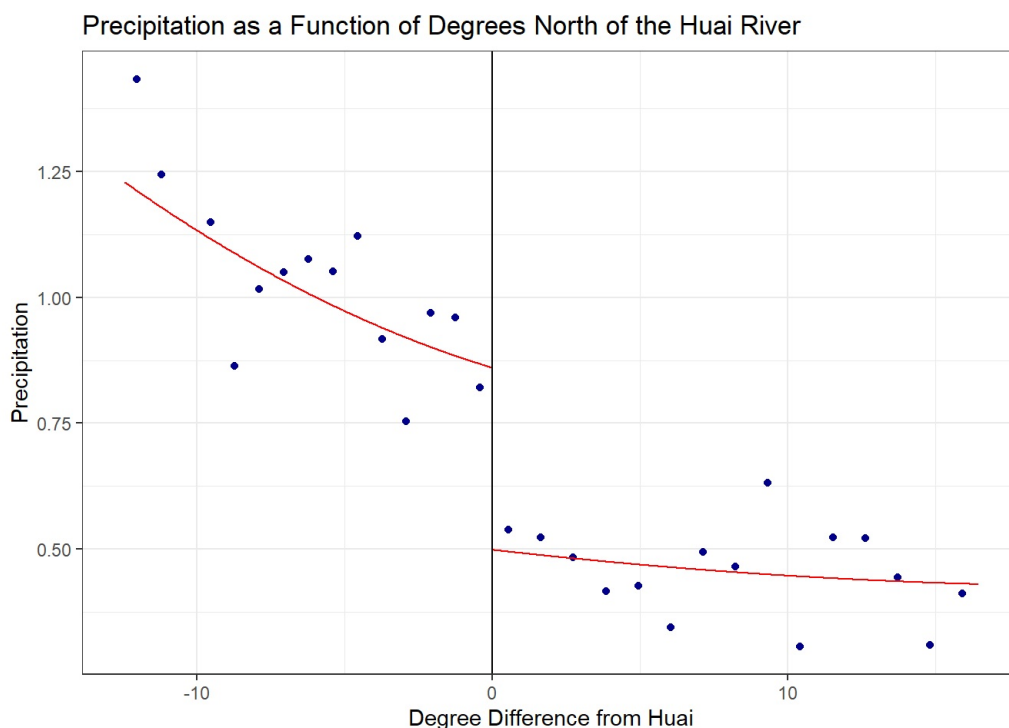
```
rdrobust(HR$prcp,HR$dist_huai,c=0,p=2)
```

```
## Call: rdrobust
##
## Number of Obs.                 153
## BW type                      mserd
## Kernel                  Triangular
## VCE method                      NN
##
## Number of Obs.            76          77
## Eff. Number of Obs.       28          19
## Order est.  (p)            2           2
## Order bias  (q)            3           3
## BW est. (h)            2.936       2.936
## BW bias (b)            5.101       5.101
## rho (h/b)             0.576       0.576
## Unique Obs.              76          77
```

6: Recall that we have an assignment variable "x" that relative to c, determines whether an object receives treatment. At c, the exogenous difference that supposedly represents causal discontinuity occurs, thus the identification is the jump. In the model, this takes form as tau which multiplied by treatment status gives the discrete difference in the data. The identification here is the pollution policy relative to the position of the Huai river, if we can causally state an affect on pollution, we correctly make the assumption and validate our RD design.

7: In this case, cities cannot physically move and because of that it is reasonable to assume that there is no mass migration across the Huai. In fact, if we did see movement of people for their health outcomes, we would see movement to just below the border and a dampening of the pollution north of the river, thus our identification would actually be correct but weaker from manipulation. From the histogram below, we see a

center just south of the river as possibly hypothesized, however even though skewed right, it is a smooth and normally distributed curve. I'm not worried about manipulation here.

```
Huai_Distance = HR$dist_huai
hist(Huai_Distance)
```

Histogram of Huai_Distance
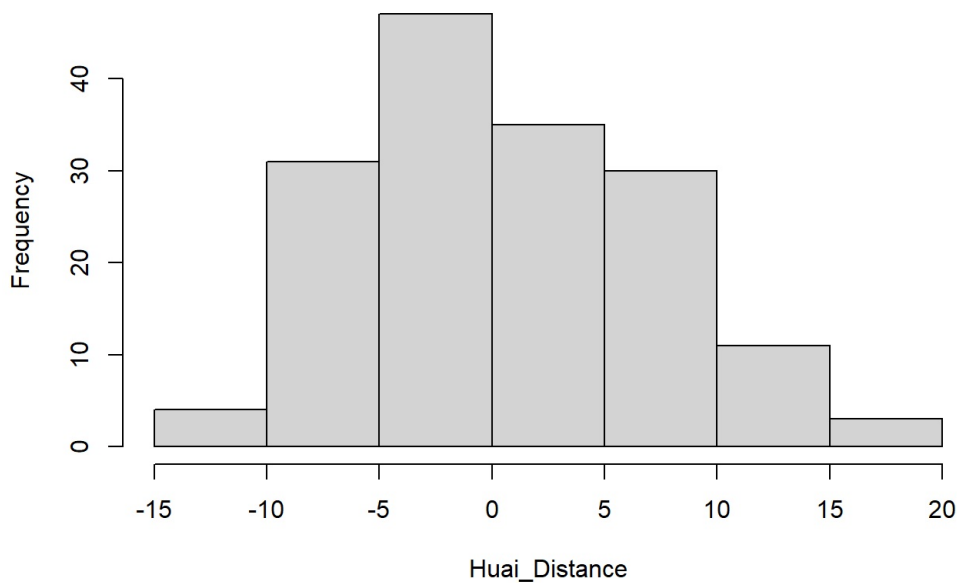
8a: Placebo tests allow researchers

to test for an association that would be present in a flawed design. By moving the the Huai river in the design, we extinguish the effects on treatment while being able to see potential flaws, such as the same discontinuity but at a different point. In the results, the only discontinuous point was at the actual Huai river, thus the test is informative in telling the reader that there are most likely minimal, if any, identification issues in the causal effect.

8b:

```
#Placebo Test Replication
#Mutate Distance from Huai Column
HR1 = HR %>%mutate(HR$dist_huai+1)
#Plot new RD
rdplot(HR1$pm10,HR1$dist_huai+1,,c=1,p=2,nbins=15,binselect = "esmv",y.label = "PM10 Concentration",x.label = "Degree Difference from Huai",title = "PM10 as a Function of Degrees North of the Huai River")
```
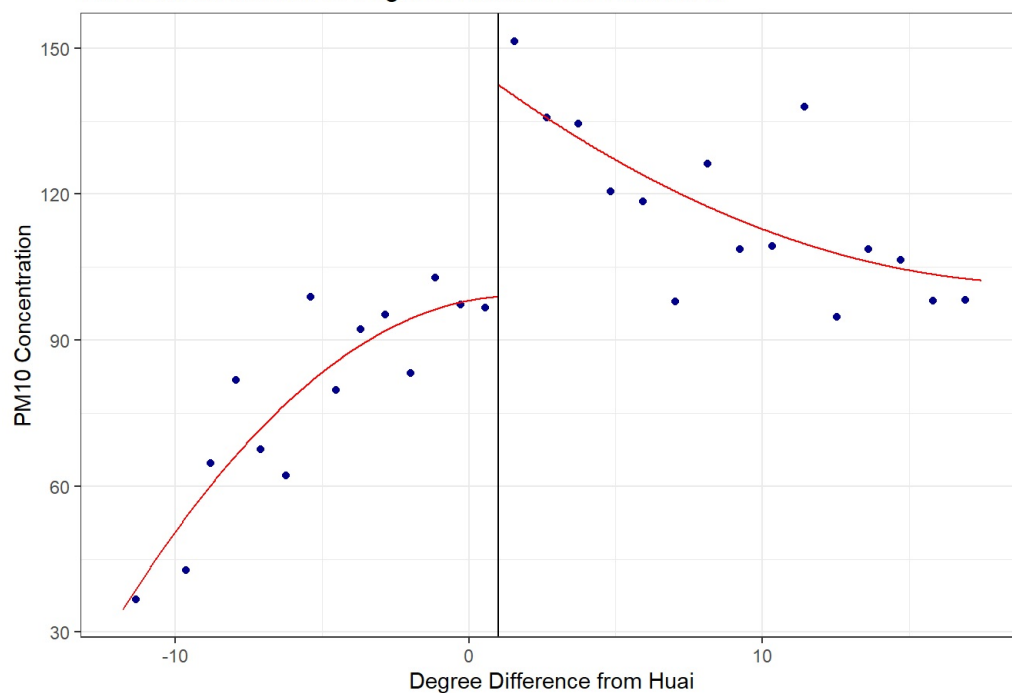
```
#Placebo Test
PlaceboRD <- rdd_data(y=HR1$pm10,x=HR1$`HR$dist_huai + 1`,cutpoint =0)

# Use with vcov function
reg_lminf <- rdd_reg_np(rdd_object=PlaceboRD, inference='lm')

# Apply Function to updated object!
vc <- function(x) vcovCluster(x, clusterVar=model.frame(x)$x)
plotPlacebo(reg_lminf, vcov. = vc)
```