Regression Discontinuity

Drew Dimmery drewd@nyu.edu

April 4, 2014

Structure

- RDD interpretation
- RDD estimation
- Placebo tests
- Sorting
- Other stuff

Interpretation

- It's a LATE!
- A different kind of LATE!
- It can be interpreted as a weighted average over all units (Lee & Lemieux 2010)
- (W, U) are observed and unobserved factors which explain all heterogeneity.
- $X = c$ is the cutpoint on the running variable, Y is the outcome
  $\lim_{\epsilon \downarrow 0} E[Y|X = c + \epsilon] - \lim_{\epsilon \uparrow 0} E[Y|X = c + \epsilon]$
  $= \sum_{w,u} \tau(w, u) p(W = w, U = u|X = c)$
  $= \sum_{w,u} \tau(w, u) \frac{f(c|W=w,U=u)}{f(c)} p(W = w, U = u)$
- What does this mean?
- It's a weight of individual treatment effects weighted by the likelihood that a unit will lie near the threshhold on the running variable.
- Keep this in mind as you interpret results.

Estimation

- If only someone wrote a package to do this. . .
- http://github.com/ddimmery/rdd

- The current best pracices is to use local polynomial regression.
- Typically linear
- There are also some interesting methods using randomization inference, though. (Cattaneo et al n.d.)

Replication

- I'll be replicating the recent Meyersson paper that's been making noise.
- Replication materials
- The paper shows a (local) result that when Islamic parties won elections in Turkey, this resulted in better outcomes for women.
- Running variable: vote margin (but not exclusively 2 party system as in Lee)
- Outcome that we'll look at: high school education

. . .

```
require(foreign, quietly = TRUE)
d <- read.dta("regdata0.dta")
summary(d$iwm94)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##    -1.0    -0.5    -0.3    -0.3    -0.1     1.0     544
```

Explore data

- Plot the raw data.

. . .

```
with(d, plot(iwm94, hischshr1520f, pch = 19, cex = 0.2, xlim = c(-0.5, 0.5)))
left.lm <- lm(hischshr1520f ~ iwm94, d, subset = iwm94 < 0)
right.lm <- lm(hischshr1520f ~ iwm94, d, subset = iwm94 >= 0)
left.x <- seq(-0.5, 0, 0.01)
right.x <- -left.x
lines(left.x, predict(left.lm, newd = data.frame(iwm94 = left.x)), col = "red")
lines(right.x, predict(right.lm, newd = data.frame(iwm94 = right.x)), col = "red")
```
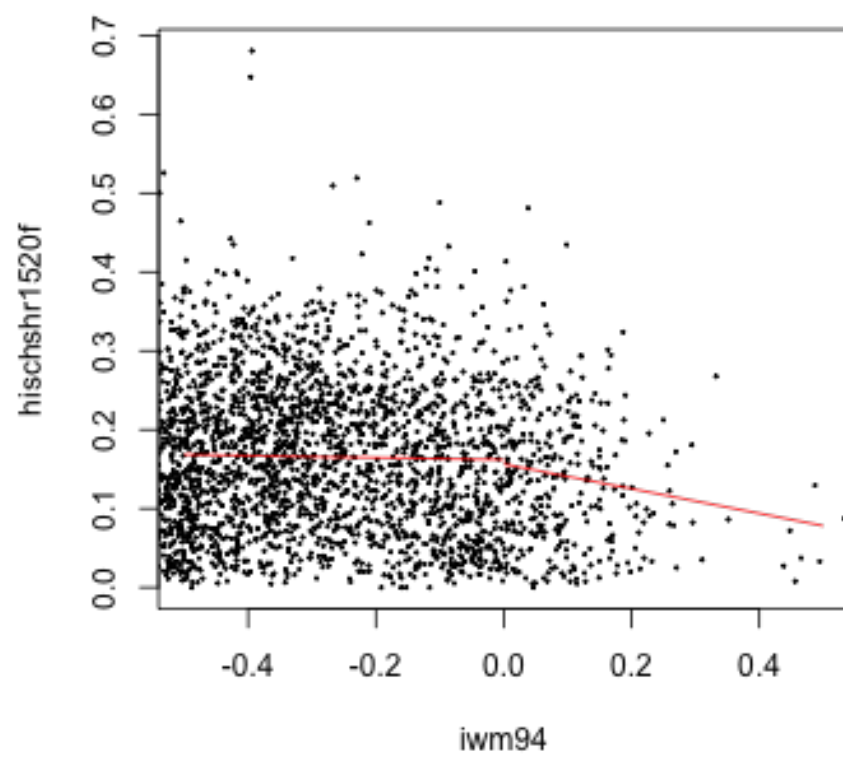
Estimation

- So the basic estimation would just take the difference of the intercepts from `left.lm` and `right.lm`.
- And there's an equivalency to just running a single regression as Cyrus showed in class.
- But I'm just going to use `rdd`

. . .

```
rd.out <- RDestimate(hischshr1520f ~ iwm94, d)
```

```
## Error: could not find function "RDestimate"
```

```
rd.out
```

```
## Error: object 'rd.out' not found
```

Full Results

```
summary(rd.out)
```

```
## Error: object 'rd.out' not found
```

Plot it

```
plot(rd.out, range = c(-0.4, 0.4))
```

```
## Error: object 'rd.out' not found
```

```
title(xlab = "Islamic Party Vote Margin", ylab = "Female High School Education Share")
```

```
## Error: plot.new has not been called yet
```

Placebo tests

- Do placebo tests on other covariates and other outcomes.
- They're "placebo" because there "shouldn't" be an effect on them (except occasionally by chance)

. . .

```r
# Age 19+
RDestimate(ageshr19 ~ iwm94, d)[c("est", "se")]
```

```
## Error: could not find function "RDestimate"
```

```r
# Log Population
RDestimate(lpop1994 ~ iwm94, d)[c("est", "se")]
```

```
## Error: could not find function "RDestimate"
```

```r
# Household Size
RDestimate(shhs ~ iwm94, d)[c("est", "se")]
```

```
## Error: could not find function "RDestimate"
```

More Placebos

```r
# Men in 2000
RDestimate(hischshr1520m ~ iwm94, d)[c("est", "se")]
```

```
## Error: could not find function "RDestimate"
```

```r
# Women in 1990 (pre-treatment)
RDestimate(c90hischshr1520f ~ iwm94, d)[c("est", "se")]
```

```
## Error: could not find function "RDestimate"
```

```r
# Men in 1990 (pre-treatment)
RDestimate(c90hischshr1520m ~ iwm94, d)[c("est", "se")]
```

```
## Error: could not find function "RDestimate"
```

Sorting

- As Cyrus discussed, density tests are also a good way to examine the possibility of sorting.

. . .

```
DCdensity(d$iwm94, verbose = TRUE, plot = FALSE)
```

```
## Error: could not find function "DCdensity"
```

Density Plot

```
DCdensity(d$iwm94)
```

```
## Error: could not find function "DCdensity"
```

Fuzzy designs

- I don't have an example for this, but it's quite easy.
- Do it the same way as before, but with `RDestimate(Y~runvar+treatment)`

Overall

- Some big things for RDD:
  - Lots of plots
  - Think about locality in interpretation
  - Use your covariates for robustness/placebo tests
  - Everything should be robust to different bandwidths, etc
  - If effects start disappearing as bw goes down, that's a bad sign.
  - Your bandwidth is probably to wide.
- If there's still more time, maybe I'll go through some high points of the `rdd` code.