11. Regression discontinuity

LPO 8852: Regression II

Sean P. Corcoran

RD - introduction

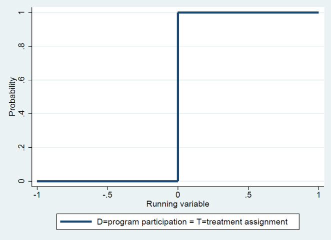
RD can be used when a **precise** rule based on a **continuous** characteristic determines treatment assignment. Examples:

- **Test scores**: can determine school admission, financial aid, summer school, remediation, graduation
- **Income or poverty score**: eligibility for income assistance or benefits, community eligibility for a means-tested anti-poverty program
- **Date**: age cutoff for retirement benefits, health insurance, school enrollment (KG or PK)
- **Elections**: fraction that voted for a particular candidate or initiative (e.g., school bond measure)

The continuous characteristic is typically called a **running variable** or **forcing variable**.

RD - introduction

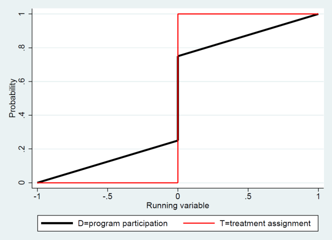
Sharp RD: treatment assignment goes from $0 \rightarrow 1$ at a threshold c. Program participation goes from 0% to 100% at c (full compliance).



We often re-center the running variable so that the threshold value is 0.

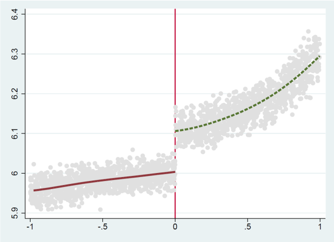
RD - introduction

Fuzzy RD: treatment assignment goes from $0 \rightarrow 1$ at a threshold c. Program participation increases sharply at c but there is non-compliance.



RD - introduction

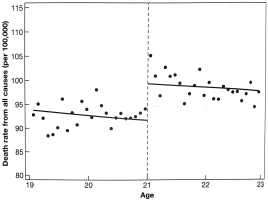
If there is a discrete jump in treatment and program participation at c (and the program has a treatment effect) one would expect to see a discrete jump in the mean outcome at c.



RD - introduction

In most cases there is a relationship between the running variable and the outcome, even in the absence of treatment. We need to properly account for this relationship, since our estimate of the treatment effect hinges on our willingness to extrapolate across the threshold.

FIGURE 4.2
A sharp RD estimate of MLDA mortality effects



Notes: This figure plots death rates from all causes against age in months. The lines in the figure show fitted values from a regression of death rates on an over-21 dummy and age in months (the vertical dashed line indicates the minimum legal drinking age (MLDA) cutoff).

Sharp RD

Treatment status D_i is a deterministic, discontinuous function of X_i:

$$D_i = \begin{cases} 1 & \text{if } X_i \geq c \\ 0 & \text{if } X_i < c \end{cases}$$

Assume a linear relationship between Y_i and X_i in the absence of treatment:

$$E[Y_{0i}|X_i] = \alpha + \beta X_i$$

Also assume constant treatment effect so that $Y_{1i} = Y_{0i} + \rho$. Then we can estimate ρ using the regression:

$$Y_i = \alpha + \beta X_i + \rho D_i + u_i$$

Sharp RD

Observations:

- The above is *very* easy to implement using OLS!
- D_i is completely determined by X_i. Having controlled for X_i, D_i is exogenous. There are no omitted variables correlated with D_i and u_i.
- Unlike regression and matching, there is *no common support* here between the treatment and control groups. We are not comparing outcomes of units with the same X. Rather, we are extrapolating.
- Because of this extrapolation, we have to rely heavily on functional form assumptions. The linear model may not be realistic, especially as we move away from the cutpoint.

Sharp RD

Alternatively, assume a smooth non-linear function:

$$E[Y_{0i}|X_i] = f(X_i)$$

Again assume constant treatment effect so that $Y_{1i} = Y_{0i} + \rho$. Then we can estimate ρ using the regression:

$$Y_i = f(X_i) + \rho D_i + u_i$$

The non-linear function could be a pth order polynomial, e.g.:

$$Y_i = \alpha + \beta_1 X_i + \beta_2 X_i^2 + ... + \beta_p X_i^p + \rho D_i + u_i$$

Sharp RD

One can also assume the linear or nonlinear relationship differs on either side of c. Let $\tilde{X}_i = X_i - c$ (distance from the cutoff).

$$E[Y_{0i}|X_i] = \alpha + \beta_{01}\tilde{X}_i + \beta_{02}\tilde{X}_i^2 + ... + \beta_{0p}\tilde{X}_i^p$$

$$E[Y_{1i}|X_i] = \alpha + \rho + \beta_{11}\tilde{X}_i + \beta_{12}\tilde{X}_i^2 + ... + \beta_{1p}\tilde{X}_i^p$$

So then:

$$E[Y_i|X_i] = E[Y_{0i}|X_i] + (E[Y_{1i}|X_i] - E[Y_{0i}|X_i])D_i$$

Sharp RD

We can then estimate ρ using the regression:

$$Y_i = \alpha + \beta_{01}\tilde{X}_i + \beta_{02}\tilde{X}_i^2 + ... + \beta_{0p}\tilde{X}_i^p$$
$$+\rho D_i + \beta_1^* D_i\tilde{X}_i + \beta_2^* D_i\tilde{X}_i^2 + \beta_p^* D_i\tilde{X}_i^p + u_i$$

Note $\beta_j^* = (\beta_{1j} - \beta_{0j})$, the difference in slope coefficients in the treated and untreated states. As an example, let $p = 2$ (quadratic):

$$Y_i = \alpha + \beta_{01}\tilde{X}_i + \beta_{02}\tilde{X}_i^2 + \rho D_i + \beta_1^* D_i\tilde{X}_i + \beta_2^* D_i\tilde{X}_i^2 + u_i$$

At the cutoff ($\tilde{X} = 0$), the treatment effect is ρ. Away from the cutoff, the treatment effect is $\rho + \beta_1^*\tilde{X}_i + \beta_2^*\tilde{X}_i^2$. This model is also easy to implement using OLS.

Example: Carpenter & Dobkin 2009

Increased mortality from legal access to alcohol: *Mastering Metrics* ch 4 example based on Carpenter & Dobkin (2009). See *AEJfigs.dta*

- Number of deaths in 50 equal-width age cells (from age 19-23)
- Recenter *age* at legal drinking age (let *age* = *agecell* - 21)
- Reaching legal drinking age is the "treatment." The discontinuity in treatment is sharp.
- We predict a discontinuity in deaths at 21. Let *over21* = *agecell* $\geq$ 21

Example: Carpenter & Dobkin 2009

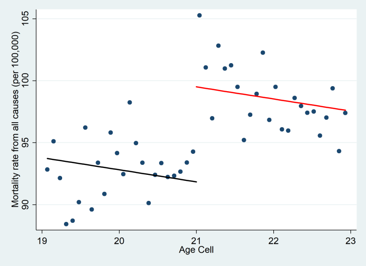
```
* recenter running variable at 21 and
* define treatment assignment variable
gen age = agecell - 21
gen over21 = agecell >= 21

* Regressions for Figure 4.2.
* linear trend, and linear on each side
* get predicted values for plotting figure below
reg all age over21
predict allfitlin
reg all c.age##i.over21
predict allfitlini
```


Example: Carpenter & Dobkin 2009

```
. * Regressions for Figure 4.2.
. * linear trend, and linear on each side
. * All = all deaths
. reg all age over21

      Source |       SS           df       MS      Number of obs   =        48
-------------+----------------------------------   F(2, 45)        =     32.99
       Model |  410.138151         2  205.069075   Prob > F        =    0.0000
    Residual |  279.682408        45  6.21516463   R-squared       =    0.5946
-------------+----------------------------------   Adj R-squared   =    0.5765
       Total |  689.820559        47  14.6770332   Root MSE        =     2.493

------------------------------------------------------------------------------
         all |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         age |  -.9746843   .6324613    -1.54   0.130    -2.248527    .2991581
      over21 |   7.662709   1.440286     5.32   0.000     4.761824    10.56359
       _cons |   91.84137   .8050394   114.08   0.000     90.21994     93.4628
------------------------------------------------------------------------------

. predict allfitlin
(option xb assumed; fitted values)

. reg all c.age##i.over21

      Source |       SS           df       MS      Number of obs   =        48
-------------+----------------------------------   F(3, 44)        =     29.47
       Model |  460.574058         3  153.524686   Prob > F        =    0.0000
    Residual |  229.246501        44  5.21014775   R-squared       =    0.6677
-------------+----------------------------------   Adj R-squared   =    0.6450
       Total |  689.820559        47  14.6770332   Root MSE        =    2.2826

------------------------------------------------------------------------------
         all |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         age |   .8269952   .8189316     1.01   0.318     -.823453    2.477443
    1.over21 |   7.662709   1.318704     5.81   0.000     5.005035    10.32038
             |
over21#c.age |
           1 |  -3.603359   1.158144    -3.11   0.003    -5.937445   -1.269273
             |
       _cons |   93.61837   .9324647   100.40   0.000     91.73911    95.49763
------------------------------------------------------------------------------

. predict allfitlini
(option xb assumed; fitted values)
```

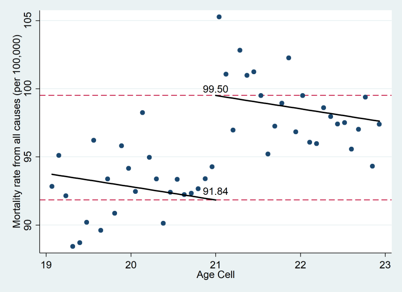
Example: Carpenter & Dobkin 2009

Note: uses *agecell* for x-axis instead of *age*. $agecell = 21$ where $age = 0$




```
twoway (scatter all agecell) (line allfitlin agecell if age < 0,
lcolor(black) lwidth(medthick)) (line allfitlin agecell if age >= 0,
lcolor(red) lwidth(medthick)), legend(off)
```

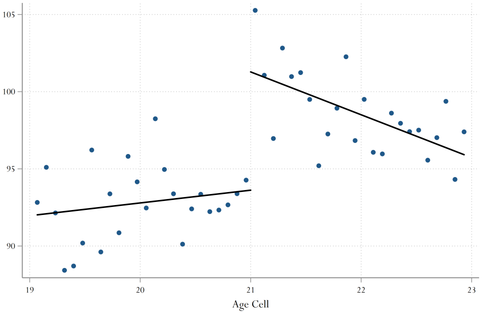

Example: Carpenter & Dobkin 2009



Intercept from the left: 91.84. From the right: 91.84 + **7.66** = 99.50

Example: Carpenter & Dobkin 2009

Version using different slopes on either side of c:



Example: Carpenter & Dobkin 2009

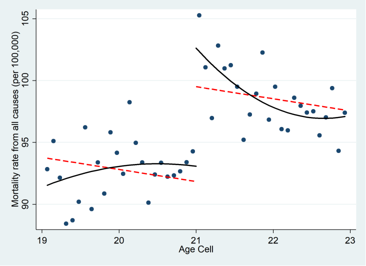
```
* Regressions for Figure 4.4.
* Quadratic, and quadratic on each side
* get predicted values for plotting figure below

reg all c.age##c.age over21
predict allfitq
reg all c.age##c.age##i.over21
predict allfitqi
```

Example: Carpenter & Dobkin 2009

```
. * Regressions for Figure 4.4.
. * Quadratic, and quadratic on each side
. reg all c.age##c.age over21

      Source |       SS           df       MS      Number of obs   =        48
-------------+----------------------------------   F(3, 44)        =     28.12
       Model |  453.339903         3  151.113301   Prob > F        =    0.0000
    Residual |  236.480656        44  5.37456036   R-squared       =    0.6572
-------------+----------------------------------   Adj R-squared   =    0.6338
       Total |  689.820559        47  14.6770332   Root MSE        =    2.3183

------------------------------------------------------------------------------
         all |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         age |  -.9746843   .5881378    -1.66   0.105    -2.159998    .2106296
             |
 c.age#c.age |  -.8186505   .2887482    -2.84   0.007    -1.400584   -.2367167
             |
      over21 |   7.662709   1.339349     5.72   0.000     4.963428    10.36199
       _cons |   92.90274   .8370061   110.99   0.000     91.21587    94.58962
------------------------------------------------------------------------------

. predict allfitq
(option xb assumed; fitted values)

. reg all c.age##c.age##i.over21

      Source |       SS           df       MS      Number of obs   =        48
-------------+----------------------------------   F(5, 42)        =     18.02
       Model |  470.512103         5  94.1024205   Prob > F        =    0.0000
    Residual |  219.308457        42  5.22162992   R-squared       =    0.6821
-------------+----------------------------------   Adj R-squared   =    0.6442
       Total |  689.820559        47  14.6770332   Root MSE        =    2.2851

------------------------------------------------------------------------------------
               all |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------------+----------------------------------------------------------------
               age |  -.8305827   3.290064    -0.25   0.802    -7.470202    5.809036
                   |
       c.age#c.age |  -.8402999   1.615268    -0.52   0.606    -4.100043    2.419443
                   |
          1.over21 |   9.547789   1.985277     4.81   0.000     5.541337    13.55424
                   |
      over21#c.age |
                1  |  -6.017014   4.652854    -1.29   0.203    -15.40685    3.372825
                   |
over21#c.age#c.age |
                1  |   2.904189   2.284334     1.27   0.211    -1.705784    7.514162
                   |
             _cons |   93.07294   1.403803    66.30   0.000     90.23995    95.90593
------------------------------------------------------------------------------------

. predict allfitqi
(option xb assumed; fitted values)
```


Example: Carpenter & Dobkin 2009

Note: uses *agecell* for x-axis instead of *age*. *agecell* $= 21$ is where *age* $= 0$




```
twoway (scatter all agecell) (line allfitlin allfitqi agecell if age <
0, lcolor(red black) lwidth(medthick medthick) lpattern(dash)) (line
allfitlin allfitqi agecell if age >= 0, lcolor(red black)
lwidth(medthick medthick) lpattern(dash)), legend(off)
```

Sharp RD

- The above example uses a restricted bandwidth of ages 19-23
  - Estimating a regression for observations near the cutoff is a *nonparametric* approach, e.g. local linear regression or local polynomial
  - Can expand bandwidth and explicitly try to model the relationship between Y_i and X_i, a *parametric* approach.
  - The choice of nonparametric vs. parametric approach involves a tradeoff of more bias for greater precision (re: larger N).
- When using polynomials it is good practice to discuss sensitivity to modeling choice. Current state of the art recommends *simplest* possible models (e.g., no more than quadratic).
- Must use caution when interpreting effects away from the cutoff.
- Can include other covariates for precision, although if assumption of continuity is correct, these covariates should not cause a discontinuity. Current state of the art recommends against using covariates.

Sharp RD

Addressing validity of the RD:

- Can estimate the same model for outcomes *not* expected to be affected by the discontinuous shift in treatment. See next slide where no effect is found of the "treatment" (reaching legal drinking age) on homicide, or deaths from all internal causes.
- Can estimate the same model using *pre-treatment* characteristics. Would not expect to see discontinuous shifts in these at the cutpoint.

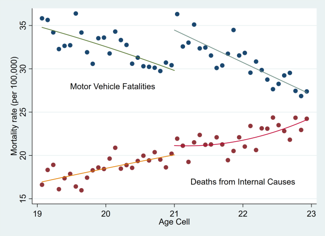
Example: Carpenter & Dobkin 2009

TABLE 4.1
Sharp RD estimates of MLDA effects on mortality

Dependent variable	Ages 19–22 (1)	Ages 19–22 (2)	Ages 20–21 (3)	Ages 20–21 (4)
All deaths	7.66 (1.51)	9.55 (1.83)	9.75 (2.06)	9.61 (2.29)
Motor vehicle accidents	4.53 (.72)	4.66 (1.09)	4.76 (1.08)	5.89 (1.33)
Suicide	1.79 (.50)	1.81 (.78)	1.72 (.73)	1.30 (1.14)
Homicide	.10 (.45)	.20 (.50)	.16 (.59)	−.45 (.93)
Other external causes	.84 (.42)	1.80 (.56)	1.41 (.59)	1.63 (.75)
All internal causes	.39 (.54)	1.07 (.80)	1.69 (.74)	1.25 (1.01)
Alcohol-related causes	.44 (.21)	.80 (.32)	.74 (.33)	1.03 (.41)
Controls	age	age, age^2, interacted with over-21	age	age, age^2, interacted with over-21
Sample size	48	48	24	24

Notes: This table reports coefficients on an over-21 dummy from regressions of month-of-age-specific death rates by cause on an over-21 dummy and linear or interacted quadratic age controls. Standard errors are reported in parentheses.

Example: Carpenter & Dobkin 2009




```
twoway (scatter mva internal agecell) (line exfitqi infitqi
agecell if agecell < 21) (line exfitqi infitqi agecell if
agecell >= 21), legend(off) text(28 20.1 "Motor Vehicle
Fatalities") text(17 22 "Deaths from Internal Causes")
```

Sharp RD

Ideally we would not have to rely on functional form assumptions at all. For small values of Δ:

$$E[Y_i|c-\Delta < X_i < c] \approx E[Y_{0i}|X_i = c]$$
$$E[Y_i|c \leq X_i < c+\Delta] \approx E[Y_{1i}|X_i = c]$$

In the limit:

$$\lim_{\Delta \to 0} E[Y_i|c \leq X_i < c+\Delta] - E[Y_i|c-\Delta < X_i < c] = E[Y_{1i} - Y_{0i}|X_i = c]$$

In other words, the difference in means in an extremely narrow band around c should represent the treatment effect *at that point* (a LATE or ATE_c). In practice researchers still use local linear regression to model relationship between Y and X around the cutoff.

Sharp RD

Bloom (2012) describes two characterizations of RD:

- *discontinuity at a cutpoint*, in which the expected outcome has some functional relationship with the running varaible X and then a discontinuity at the cutpoint c to be estimated, and
- *local randomization* (Lee, 2008) in which randomness in the neighborhood of c determines whether one is on the left or right side of this cutpoint.

Being above/below age 21 fits the first characterization; vote margin for/against a school bond measure might fit the second.

Estimation using `rd` command

Stata has a user-written ado file called `rd` that estimates treatment effects using local linear polynomials—see in-class exercise. For the sharp RD:

```
rd y x, z0(c)
```

where y is the outcome varaible, x is the running variable, and c is the cutpoint. `rd` includes lots of options. One option allows you to specify the bandwidth. `rd` will also produce estimates at 50% and 200% of your specified bandwidth. The default bandwidth is the "optimal bandwidth" of Imbens and Kalyanaraman (2009).

Deciding on a bandwidth

When using local linear regression, how wide should the bandwidth be? Larger bandwidths provide more observations (can improve *precision*) but also introduce *bias* if observations away from the cutpoint are systematically different (and/or the model you fit is imperfect).

Cross-validation is one method of finding a "good" bandwidth. Since the goal of RD is to get good estimates of the intercepts on either side of c, one could calculate "intercepts" from the left and right at points away from c using different bandwidths to see which provides the best predictions.

Logic of cross-validation

1. Choose a trial bandwidth b
2. Fit a local linear regression at each point in the dataset: for each X_c regress Y_i on $(X_i - X_c)$ using observations to the right of X_i within the bandwidth. Then do the same thing to the left of X_i.
3. Average the two intercepts (call them $a_i(b)$). These represent a prediction of Y_i
4. Mean squared prediction error is $(1/N)\sum_i (Y_i - a_i)^2$ (actual Y minus that predicted from the right and left).
5. Repeat for different choices of b. Choose b that provides lowest mean squared prediction error.
6. Can opt to exclude values of X_i far from c

`rdcv` is a user-written Stata command that estimates the sharp RD with control over the cross-validation options. See in-class exercise.

RD Assumptions

- The relationship between potential outcomes Y_i and X_i is **continuous** in the neighborhood of c. There is no reason to expect a sharp break in Y_i in the absence of treatment.
- X_i has **not been manipulated** to affect who receives treatment.
- There are no other programs or services with the **same eligibility rule** (to avoid confounding with some other treatment).
- Manipulation is not the same thing as non-compliance (the “fuzzy” RD). We return to this later.

Testing for continuity

One method for testing for discontinuities in $f(X_i)$ away from c:

- Regress Y on a high-order polynomial in X and include dummy variables for values of X above various quantiles of X (e.g., deciles).
- Conduct an F-test for significance of these dummy variables.
- If the relationship between Y and X is generally smooth, there should not be significance.

See in-class example.

Manipulation

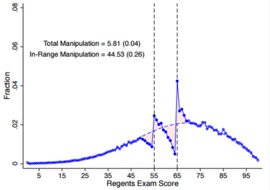
Manipulation occurs whenever cases have their value of X altered in order to affect their treatment status. For example, a teacher might adjust a test score in order to help a student pass or become eligible for a program.

- This may be visible in a histogram, or not if some with $X_i < c$ have their X_i increased but others with $X_i \geq c$ have their X_i reduced.
- If manipulation is random or uninformed, such that the expected value of Y_i in the absence of treatment is no different for those whose X has been manipulated, then manipulation will not pose a problem. Manipulation is not usually random, however.
- If the "best" of those on the margin are nudged into the treatment by manipulation of X, this will alter the equivalence of those below and above c. By removing these cases from the control group we may estimate an effect where there is none.

Testing for manipulation

Sometimes evidence for manipulation is clear from inspecting densities or histograms:

Figure 1
Test Score Distributions for Core Regents Exams, 2004-2010



Notes: This figure shows the test score distribution around the 55 and 65 score cutoffs for New York City high school test takers between 2004-2010. Core exams include English Language Arts, Global History, U.S. History, Math A/Integrated Algebra, and Living Environment. We include the first test in each subject for each student in our sample. Each point shows the fraction of test takers in a score bin with solid points indicating a manipulable score. The dotted line beneath the empirical distribution is a subject specific sixth-degree polynomial fitted to the empirical distribution excluding the manipulable scores near each cutoff. Total manipulation is the fraction of test takers with manipulated scores. In-range manipulation is the fraction of test takers with manipulated scores normalized by the average height of the counterfactual distribution to the left of each cutoff. Standard errors are calculated using the parametric bootstrap procedure described in the text. See the data appendix for additional details on the sample and variable definitions.

Testing for manipulation: McCrary (2008)

McCrary (2008) proposed a test for manipulation. The basic idea is:

1. Divide the observations into J equally-spaced bins
2. Calculate the fraction of observations in each bin, S_j
3. Assign each bin a value B_j equal to the midpoint of X in that bin. The fraction of observations in that bin is an approximation of the probability that $X = B_j$.
4. Construct a variable $(B_j - c)$ for each value of B_j. Run two regressions of S_j on $(B_j - c)$, one using bins to the left of c and the other using bins to the right. Weight the observations by the distance of B_j from c: bins further away from c get less weight, using triangular kernel: $K(t) = max[0, 1 - |(B_j - c)/h|]$ where h is a parameter you choose. (Bins "too far" away from the cutpoint get zero weight).

Testing for manipulation: McCrary

5. Let f^+ be the intercept from the regression using bins to the right of c and f^- be the intercept from the regression using bins to the left of c. Construct the test statistic $ln(f^+) - ln(f^-)$. The test statistic has an approximately normal distribution with standard error equal to $\sqrt{(1/nh)(24/5)(1/f^+ + 1/f^-)}$

The point of step #4 is to get a smooth linear approximation to the density, on either side of c.

- Default bin size: $b = 2s_x/\sqrt{n}$.
- Default bandwidth: see McCrary (2008)

Testing for manipulation: McCrary (2008)

Example from McCrary using his *DCdensity.ado* Stata command:

```
set seed 1234567
set obs 10000
gen Z=invnorm(uniform())
DCdensity Z, breakpoint(0) generate(Xj Yj r0 fhat se_fhat)
```

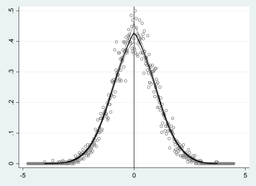
The above creates a dataset with 10,000 *N(0,1)* random draws. There is no manipulation, so we should not expect a discontinuity at the "breakpoint" (0). In this example:

- Default bin size = 0.0196
- Default bandwidth: 0.744
- Number of bins: 396 (range/0.0196)
- Test statistic (SE): -0.006 (0.055)

Testing for manipulation: McCrary (2008)

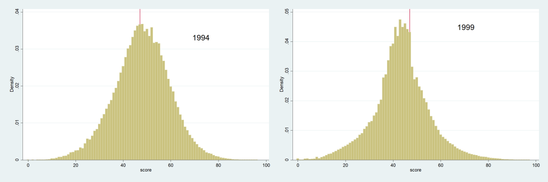
5 variables are created in the dataset (name using `generate` option):

- *Xj*: cell midpoint for the histogram
- *Yj*: cell height for the histogram
- *r0*: evaluation sequence for LLR loop
- *fhat*: local linear density estimate
- *se_fhat*: SE of local linear density estimate



Example: Camacho and Conover (2011)

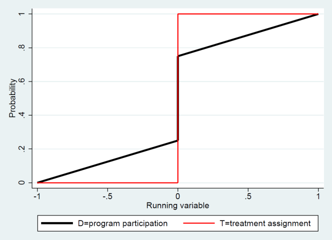
In 1998, Colombia set eligibility threshold for social welfare benefits at a poverty index of 47.



This is an example of a **discrete** running variable: an integer poverty index. Note the McCrary test finds evidence of manipulation in *1994*. Studies find the McCrary test over-rejects the null hypothesis of no manipulation when the running variable is discrete, and more so when N is large.

Fuzzy RD

Fuzzy RD: treatment *assignment* goes from $0 \rightarrow 1$ at a threshold c. Treatment *status* (e.g., program participation) increases sharply at c but there is non-compliance.



Fuzzy RD

Treatment status D_i is a discontinuous function of X_i, but X_i does not completely determine D_i:

$$Prob(D_i = 1|X_i) = \begin{cases} g_1(x_i) & \text{if } X_i \geq c \\ g_0(x_i) & \text{if } X_i < c \end{cases}$$

We need differing notation for treatment *assignment* and treatment *status* since these are no longer the same ting. Let $T_i = I(X_i \geq c)$ where $I()$ is an indicator function. Now we have two equations:

$$D_i = \gamma_0 + \gamma_1 X_i + \gamma_2 X_i^2 + ... + \gamma_p X_i^p + \pi T_i + \epsilon_{1i}$$

$$Y_i = \kappa_0 + \kappa_1 X_i + \kappa_2 X_i^2 + ... + \kappa_p X_i^p + \pi\rho T_i + \epsilon_{2i}$$

T is the treatment assignment, D is the treatment status.

Fuzzy RD

The first equation shows the relationship between treatment status D_i and treatment assignment T_i (whether or not i is below or above the threshold). D_i may have a relationship with X_i, approximated by the polynomial. The coefficient π captures the discontinuity at c.

This equation is a *first stage* regression.

In-class exercise example: recommendation for gifted program.

Fuzzy RD

The second equation shows the relationship between the outcome Y_i and T_i ($T_i = 1$ if i is above the threshold). As before, Y_i may have a relationship with X_i, approximated by the polynomial. The coefficient represented as $\pi\rho$ captures the discontinuity at c.

This equation is a *reduced form*. We expect to see a jump in Y at c, but this jump will be "diluted" because of non-compliance.

Note T_i does not present an OVB problem in the reduced form. It does not represent treatment status (participation) but rather treatment *assignment*. T_i is a deterministic function of X_i.

Fuzzy RD

The treatment effect ρ can be calculated as: $\pi\rho/\pi$. That is, the reduced form coefficient on T_i divided by the first stage coefficient on T_i. Known as a **local Wald estimate**.

There are some key assumptions baked into this approach.

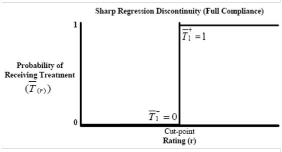
Bloom (2009, 2012)

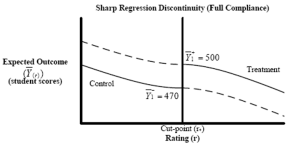
"Type I" fuzzy RD: the probability of treatment is < 1 for those above c and 0 for those below c.

- The *intent-to-treat* effect at the cutpoint: $ITT_c = (TOT_c \times \bar{T}^+) + (0 \times (1 - \bar{T}^+))$
- $\bar{T}^+$ is the proportion treated above c. TOT_c is the treatment effect for the treated, at the outcome. We assume a zero treatment effect for the "no-shows."
- So: $TOT_c = ITT_c/\bar{T}^+$
- We "scale up" the ITT by the proportion treated

Example from Bloom (2009, 2012)

Sharp RD:

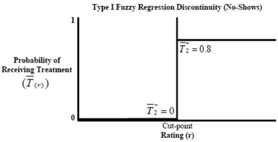


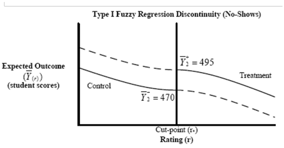


ATE_c (average treatment effect at the cutpoint) =
TOT_c (treatment effect on the treated at the cutpoint) =
$500 - 470 = 30$.

Example from Bloom (2009, 2012)

Fuzzy RD (Type I):





ITT_c (intent-to-treat effect at the cutpoint) $= 495 - 470 = 25$
TOT_c (treatment effect on the treated at the cutpoint)
$= ITT_c / \bar{T}^+ = (25/0.8) = 31.25$.
Note this is *not* the ATE_c. Not all who *should* be treated at the cutpoint are. We only have an estimate of the treatment effect for those who were induced to receive treatment.

Bloom (2009, 2012)

"Type II" fuzzy RD: the probability of treatment is < 1 for those above c and > 0 for those below c.

- We can still estimate *intent-to-treat* effect at the cutpoint. However, it is not as easy to partition this into "treated" and "no-shows".
- Problem: there are "crossovers," individuals below c that are treated. It is now harder to rationalize equivalence of those above/below c.

Bloom (2009, 2012)

Partition the population into four groups:

- **Compliers:** receive treatment if and only if they are assigned to it. Their treatment status changes at c, so they contribute to a treatment contrast.
- **Always-takers:** receive treatment regardless of assignment. Their treatment status is unchanged, so it is impossible to estimate a treatment effect for them.
- **Never-takers:** do not receive treatment regardless of assignment. Their treatment status is unchanged, so it is impossible to estimate a treatment effect for them.
- **Defiers:** receive treatment only when *not* assigned to it.

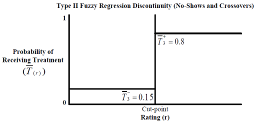
Bloom (2009, 2012)

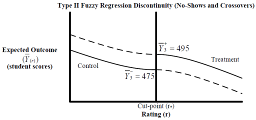
In practice, we can't conclusively determine which group individuals are in.

- It is often assumed, however, that there are no defiers.
- It is also assumed that the proportion of always-takers and never-takers is equal in the neighborhood of the cutoff. (The proportion may vary with the running variable X, but is continuous through c).
- Because compliers are the only ones whose treatment status is affected by the discontinuity in T at c, they are the only subgroup that contributes to the treatment contrast.
- We have a local average treatment effect at the cutpoint: $LATE_c = ITT_c/(\bar{T}^+ - \bar{T}^-)$

Example from Bloom (2009, 2012)

Fuzzy RD (Type II):





ITT_c (intent-to-treat effect at the cutpoint) $= 495 - 475 = 20$
$LATE_c$ (local average treatment effect at the cutpoint)
$= ITT_c/(\bar{T}^+ - \bar{T}^-) = (20/(0.80 - 0.15)) = 38.46.$
Note this is *not* the ATE_c or TOT_c, only $LATE_c$

Estimation using `rd` command

To estimate fuzzy rd using `rd`, include the treatment status variable t (the running variable x goes last)—see in-class exercise.

```
rd y t x, z0(c)
```

y is the outcome varaible, x is the running variable, t is the treatment status, and c is the cutpoint.

RD interpretation

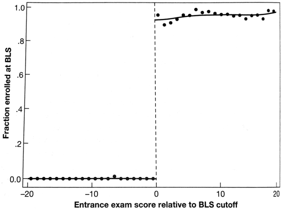
If the assumptions of RD hold, the design yields estimates with high internal validity: a local average treatment effect (LATE) at c.

If treatment effects are heterogeneous the RD may tell us little about impact away from c. The population near c may not be the one of greatest interest.

See Abdulkadiroğlu, Pathak, & Roth (2014), *The Elite Illusion: Achievement Effects at Boston and New York Exam Schools*.

Abdulkadiroğlu, Pathak, & Roth (2014)

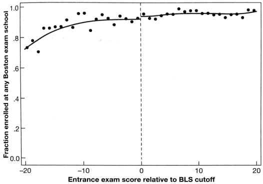
FIGURE 4.6
Enrollment at BLS



Notes: This figure plots enrollment rates at Boston Latin School (BLS), conditional on admissions test scores, for BLS applicants scoring near the BLS admissions cutoff. Solid lines show fitted values from a local linear regression estimated separately on either side of the cutoff (indicated by the vertical dashed line).

Abdulkadiroğlu, Pathak, & Roth (2014)

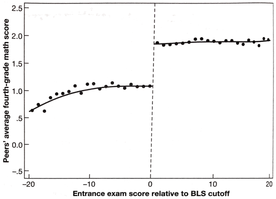
FIGURE 4.7
Enrollment at any Boston exam school



Notes: This figure plots enrollment rates at any Boston exam school, conditional on admissions test scores, for Boston Latin School (BLS) applicants scoring near the BLS admissions cutoff. Solid lines show fitted values from a local linear regression, estimated separately on either side of the cutoff (indicated by the vertical dashed line).

Abdulkadiroğlu, Pathak, & Roth (2014)

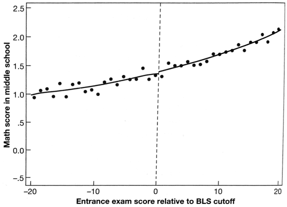
FIGURE 4.8
Peer quality around the BLS cutoff



Notes: This figure plots average seventh-grade peer quality for applicants to Boston Latin School (BLS), conditional on admissions test scores, for BLS applicants scoring near the admissions cutoff. Peer quality is measured by seventh-grade schoolmates' fourth-grade math scores. Solid lines show fitted values from a local linear regression, estimated separately on either side of the cutoff (indicated by the vertical dashed line).

Abdulkadiroğlu, Pathak, & Roth (2014)

FIGURE 4.9
Math scores around the BLS cutoff



Notes: This figure plots seventh- and eighth-grade math scores for applicants to the Boston Latin School (BLS), conditional on admissions test scores, for BLS applicants scoring near the admissions cutoff. Solid lines show fitted values from a local linear regression, estimated separately on either side of the cutoff (indicated by the vertical dashed line).

Further topics and references

- Nichols (2007) on quasi-experimental designs (see section on RD).
- Jacob et al. (2012) *A Practical Guide to Regression Discontinuity*
- Schochet et al. (2010) IES Standards for RD Designs
- Power calculations (Schochet, 2009)
- RD when the running variable is discrete (Kolesár & Rothe, 2018)
- RD with multiple cutpoints, and selection using more than one running variable (e.g., Reardon & Robinson, 2012)