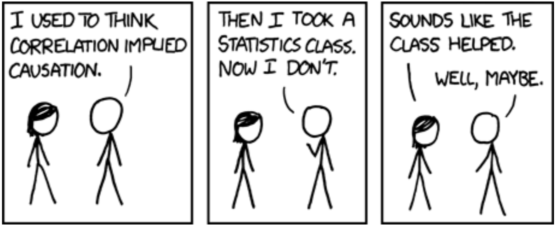
Advanced Applied Econometrics

Felix Weinhardt, BSE



RDD Visual Example

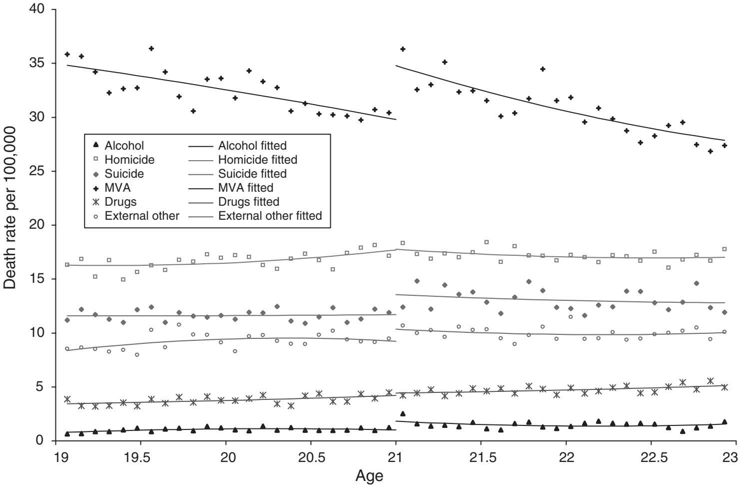


FIGURE 4. AGE PROFILES FOR DEATH RATES BY EXTERNAL CAUSE

Notes: See notes to Figure 3. The categories are mutally exclusive. The order of precedence is homicide, suicide, MVA, deaths with a mention of alcohol, and deaths with a mention of drugs. The ICD-9 and ICD-10 Codes are

RDD Data Requirements

- Question: So, where can I find these "jumps"? Answer: Humans are embedding "jumps" in their rules all the time. Dumb rules – while usually bad policy – are *great* for research.
- Validity doesn't require the assignment rule be arbitrary, only that it is known, precise and free of manipulation. The most effective RDD studies involve programs where X has a "hair trigger" that is not tightly related to the outcome being studied. Examples:
  - Probability of being tried as an adult (higher penalties for a given crime) "jumps" at age 18
  - Probability of being arrested for DWI "jumps" at blood alcohol content >0.08
  - Probability of receiving universal healthcare insurance "jumps" at age 65
  - Probability of receiving medical attention "jumps" when birthweight falls below 1,500 grams
  - Probability of having to attend summer school "jumps" when grade falls below 60
- Data requirements can be substantial. Large sample sizes are characteristic features of the RDD
  - If there are strong trends, one typically needs a lot of data
  - Researchers are typically using administrative data or settings such as birth records where there are **many** observations

Sharp vs. Fuzzy RDD

1. Sharp RDD: Treatment is a deterministic function of running variable, X. Example: Medicare benefits.
2. Fuzzy RDD: Discontinuous "jump" in the *probability* of treatment when $X > c_0$. Cutoff is used as an instrumental variable for treatment. Example: Maimonides Rule (Angrist and Lavy 1999)

Treatment assignment in the sharp RDD

Deterministic treatment assignment ("sharp RDD")

In Sharp RDD, treatment status is a deterministic and discontinuous function of a covariate, X_i:

$$D_i = \begin{cases} 1 \text{ if } & X_i \geq c_0 \\ 0 \text{ if } & X_i < c_0 \end{cases}$$

where c_0 is a known threshold or cutoff. In other words, if you know the value of X_i for a unit i, you know treatment assignment for unit i with certainty.

Example: Let X be age. Americans aged 64 are not eligible for Medicare, but Americans aged 65 ($X \geq c_{65}$) are eligible for Medicare (ignoring disability exemptions)

Implication: Sharp RDD relies on extrapolation across covariate values for its causal inference because there is no value of X_i (e.g., age) for which you observe both treatment and control observations. Therefore, we cannot be agnostic about regression functional form.

Treatment effect: definition and estimation

Definition of treatment effect

Assume constant treatment effects potential outcomes model linear in X:

$$\begin{aligned} Y_i^0 &= \alpha + \beta X_i \\ Y_i^1 &= Y_i^0 + \delta \end{aligned}$$

Use the switching equation and write in terms of Y_i:

$$\begin{aligned} Y_i &= Y_i^0 + (Y_i^1 - Y_i^0)D_i \\ Y_i &= \alpha + \beta X_i + \delta D_i + \varepsilon_i \end{aligned}$$

The treatment effect parameter, δ, is the discontinuity in the conditional expectation function:

$$\begin{aligned} \delta &= lim_{X_i \rightarrow c_0} E[Y_i^1 | X_i = c_0] - lim_{c_0 \leftarrow X_i} E[Y_i^0 | X_i = c_0] \\ &= lim_{X_i \rightarrow c_0} E[Y_i | X_i = c_0] - lim_{c_0 \leftarrow X_i} E[Y_i | X_i = c_0] \end{aligned}$$

The sharp RDD estimation is interpreted as an average causal effect of the treatment at the discontinuity

$$\delta_{SRD} = E[Y_i^1 - Y_i^0 | X_i = c_0]$$

Notice the role of *extrapolation* in estimating treatment effects when sharp RDD

- Left of cutoff, only non-treated observations, $D_i = 0$ for $X < c_0$
- Right of cutoff, only treated observations, $D_i = 1$ for $X \geq c_0$

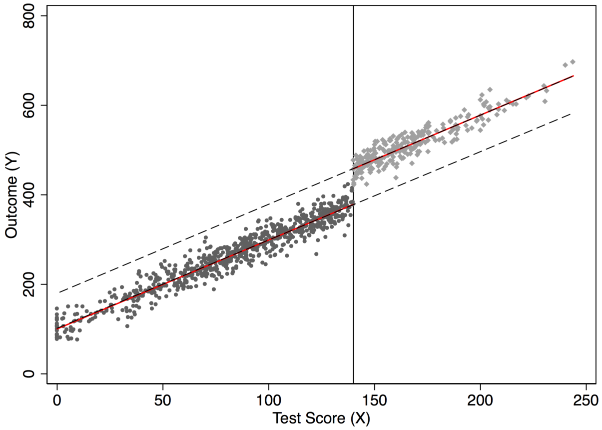


Figure: Dashed lines are extrapolations

Extrapolation and Continuous

- Notice the role of *extrapolation* in estimating treatment effects when sharp RDD
  - Left of cutoff, only non-treated observations, $D_i = 0$ for $X < c_0$
  - Right of cutoff, only treated observations, $D_i = 1$ for $X \geq c_0$
- The extrapolation is to a counterfactual. Recall the Rubin model:
  - Y_i^1 denotes outcome for person i if treated, Y_i^0 denotes *same* person i outcome if not treated
  - Causal effect for person i is $\delta_i = Y_i^1 - Y_i^0$ and the average treatment effect is $E[\delta_i] = E[Y_i^1 - Y_i^0]$
- In RDD, the counterfactuals are conditional on X. We are interested in the treatment effect at $X = c_0$:

$$E[Y_i^1 - Y_i^0 | X_i = c_0]$$

Key identifying assumption

Continuity of conditional regression functions (Hahn, Todd and Van der Klaauw 2001; Lee 2008)

$E[Y_i^0|X=c_0]$ and $E[Y_i^1|X=c_0]$ are continuous (smooth) in X at c_0.

Remark Alfred Marshall quotes Darwin in Principles of Economics (1890), "*Natura non fact saltum*", which means"nature does not make jumps". Jumps are unnatural, so when they occur, they require an explanation

Meaning If population average *potential outcomes*, Y^1 and Y^0, are continuous functions of X at the cutoff, c_0, then potential average outcomes *do not* jump at c_0. All other unobserved determinants of Y are continuously related to the running variable, X

Implication This assumption allows us to use average outcome of units right below the cutoff as a valid counterfactual for units right above the cutoff. The causal effect of the treatment will be based on extrapolation from the trend, $E[Y_i^0|X<c_0]$, to those values of $X>c_0$ for the $E[Y_i^0|X>c_0]$.

Conditional Independence

CIA

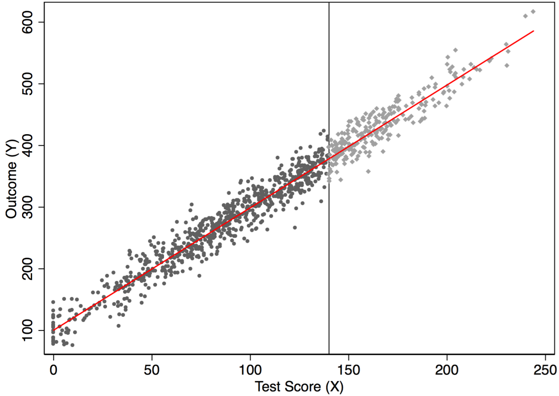
$Y^0, Y^1 \perp\!\!\!\perp D|X$. Once I fix on X, the running variable, D is independent of potential outcomes

- Trivial assumption (for the first time)
- There is no variation in the treatment variable, D, conditional on the running variable.
  - Once I condition on the running variable, there is no longer any variation left in treatment
- No common support – RDD is a local extrapolation outside the support of the data to predict mean treated and untreated potential outcomes at c_0

```
/// --- Examples using simulated data
set obs 1000
set seed 1234567

* Generate running variable
gen x = rnormal(100, 50)
replace x=0 if x < 0
drop if x > 280
sum x, det

* Set the cutoff at X=140. Treated if X > 140
gen D = 0
replace D = 1 if x > 140
```

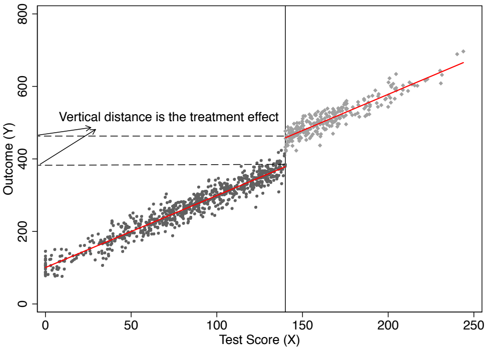
Graphical example of continuous assumption

```
gen y1 = 100 + 0*D + 2*x + rnormal(0, 20)

scatter y1 x if D==0, msize(vsmall) || scatter y1 x if D==1, msize(vsmall) le
xline(140, lstyle(foreground)) || lfit y1 x if D ==0, color(red) || lfit y1
x if D ==1, color(red) ytitle("Outcome (Y)") xtitle("Test Score (X)")
```



Graphical example with treatment effect, cutoff=140

```
gen y = 100 + 80*D + 2*x + rnormal(0, 20)

scatter y x if D==0, msize(vsmall) || scatter y x if D==1, ///
msize(vsmall) legend(off) xline(140, lstyle(foreground)) || /// lfit y x if
D ==0, color(red) || lfit y x if D ==1, color(red) /// ytitle("Outcome (Y)")
xtitle("Test Score (X)")
```



Tangent: centering at the cutoff point

- It is common for authors to transform X by "centering" at c_0:

$$Y_i = \alpha + \beta(X_i - c_0) + \delta D_i + \varepsilon_i$$

  This doesn't change the interpretation of the treatment effect – only the interpretation of the intercept.
- Example: Medicare and age 65. Center the running variable (age) by subtracting 65:

$$\begin{aligned} Y &= \beta_0 + \beta_1(Age - 65) + \beta_2 Edu \\ &= \beta_0 + \beta_1 Age - \beta_1 65 + \beta_2 Edu \\ &= \alpha + \beta_1 Age + \beta_2 Edu \end{aligned}$$

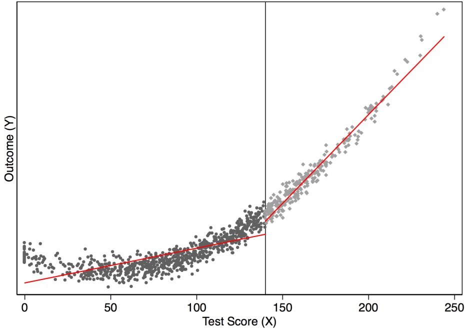
  where $\alpha = \beta_0 - \beta_1 65$. All other coefficients, notice, have the same interpretation, except for the intercept.

```
. gen x_c=x-140

. reg y D x

      Source |       SS       df       MS              Number of obs =     999
-------------+------------------------------           F(  2,   996) =19988.47
       Model |  15842893.9     2  7921446.97           Prob > F      =  0.0000
    Residual |  394715.557   996   396.30076           R-squared     =  0.9757
-------------+------------------------------           Adj R-squared =  0.9756
       Total |  16237609.5   998  16270.1498           Root MSE      =  19.907

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           D |   80.01418   2.144779    37.31   0.000     75.80537    84.22298
           x |   1.986975   .0186779   106.38   0.000     1.950322    2.023627
       _cons |   100.3885    1.70944    58.73   0.000     97.03397     103.743
------------------------------------------------------------------------------

. reg y D x_c

      Source |       SS       df       MS              Number of obs =     999
-------------+------------------------------           F(  2,   996) =19988.47
       Model |  15842893.9     2  7921446.97           Prob > F      =  0.0000
    Residual |  394715.554   996  396.300757           R-squared     =  0.9757
-------------+------------------------------           Adj R-squared =  0.9756
       Total |  16237609.5   998  16270.1498           Root MSE      =  19.907

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           D |   80.01418   2.144779    37.31   0.000     75.80537    84.22298
         x_c |   1.986975   .0186779   106.38   0.000     1.950322    2.023627
       _cons |    378.565   1.290755   293.29   0.000      376.032    381.0979
------------------------------------------------------------------------------
```

Nonlinearity bias

Smoothness in $E[Y_i^0|X_i]$ and linearity are different things. What if the trend relation $E[Y_i^0|X_i]$ does not jump at c_0 but rather is simply nonlinear? False positive problem.

```
gen x2 = x*x
gen x3 = x*x*x
gen y = 10000 + 0*D - 100*x +x2 + rnormal(0, 1000)

scatter y x if D==0, msize(vsmall) || scatter y x if D==1, msize(vsmall) legend(off) xline(140, lstyle(f
ylabel(none) || lfit y x if D ==0, color(red) || lfit y x if D ==1, color(red) xtitle("Test Score (X)")
ytitle("Outcome (Y)")
```



Sharp RDD: Nonlinear Case

- Suppose the nonlinear relationship is $E[Y_i^0|X_i] = f(X_i)$ for some reasonably smooth function $f(X_i)$. In that case we'd fit the regression model:

$$Y_i = f(X_i) + \delta D_i + \eta_i$$

- Since $f(X_i)$ is counterfactual for values of $X_i > c_0$, how will we model the nonlinearity? There are 2 ways of approximating $f(X_i)$:
  1. Let $f(X_i)$ equal a p^{th} order polynomial:

$$Y_i = \alpha + \beta_1 x_i + \beta_2 x_i^2 + \cdots + \beta_p x_i^p + \delta D_i + \eta_i$$

  2. Use a nonparametric kernel method (later)

Different polynomials on the 2 sides of the discontinuity

- We can generalize the function, $f(x_i)$, by allowing the x_i terms to differ on both sides of the cutoff by including them both individually and interacting them with D_i. In that case we have:

$$\begin{aligned} E[Y_i^0|X_i] &= \alpha + \beta_{01}\tilde{X}_i + \beta_{02}\tilde{X}_i^2 + \cdots + \beta_{0p}\tilde{X}_i^p \\ E[Y_i^1|X_i] &= \alpha + \delta + \beta_{11}\tilde{X}_i + \beta_{12}\tilde{X}_i^2 + \cdots + \beta_{1p}\tilde{X}_i^p \end{aligned}$$

  where $\tilde{X}_i$ is the centered running variable (i.e., $X_i - c_0$). Centering at c_0 ensures that the treatment effect at $X_i = c_0$ is the coefficient on D_i in a regression model with interaction terms
- As Lee and Lemieux (2010) note, allowing different functions on both sides of the discontinuity should be the main results in an RDD paper as otherwise we use values from both sides of the cutoff to estimation the function on each side

Different polynomials on the 2 sides of the discontinuity

- To derive a regression model, first note that the observed values must be used in place of the potential outcomes:

$$E[Y|X] = E[Y^0|X] + \left(E[Y^1|X] - E[Y^0|X]\right) D$$

- Regression model you estimate is:

$$\begin{aligned} Y_i &= \alpha + \beta_{01}\tilde{x}_i + \beta_{02}\tilde{x}_i^2 + \cdots + \beta_{0p}\tilde{x}_i^p \\ &\quad + \delta D_i + \beta_1^* D_i \tilde{x}_i + \beta_2^* D_i \tilde{x}_i^2 + \cdots + \beta_p^* D_i \tilde{x}_i^p + \varepsilon_i \end{aligned}$$

  where $\beta_1^* = \beta_{11} - \beta_{01}$, $\beta_2^* = \beta_{21} - \beta_{21}$ and $\beta_p^* = \beta_{1p} - \beta_{0p}$
- The equation we looked at earlier a few slides back was just a special case of the above equation with $\beta_1^* = \beta_2^* = \beta_p^* = 0$
- The treatment effect at c_0 is δ
- The treatment effect at $X_i - c_0 = c > 0$ is: $\delta + \beta_1^* c + \beta_2^* c^2 + \cdots + \beta_p^* c^p$

Polynomial simulation example

```
capture drop y x2 x3
gen x2 = x*x
gen x3 = x*x*x
gen y = 10000 + 0*D - 100*x +x2 + rnormal(0, 1000)

reg y D x x2 x3

predict yhat

scatter y x if D==0, msize(vsmall) || scatter y x if D==1, msize(vsmall) legend(off)
xline(140, lstyle(foreground)) ylabel(none) || line yhat x if D ==0, color(red)
sort || line yhat x if D ==1, sort color(red) xtitle("Test Score (X)") ytitle("Outcome
(Y)")
```

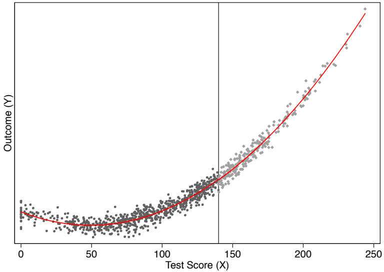


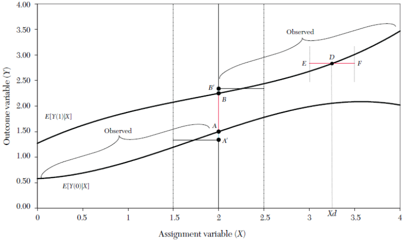
Figure: Third degree polynomial. Actual model second degree polynomial.

```
. gen x_c = x - 140

. gen x2_c = x2-140

. gen x3_c = x3-140

.
. reg y D x x2

      Source |       SS       df       MS              Number of obs =     999
-------------+------------------------------           F(  3,   995) =13115.22
       Model |  3.7863e+10     3  1.2621e+10           Prob > F      =  0.0000
    Residual |   957507024   995  962318.617           R-squared     =  0.9753
-------------+------------------------------           Adj R-squared =  0.9753
       Total |  3.8821e+10   998  38898361.8           Root MSE      =  980.98

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           D |  -115.5381   127.4967    -0.91   0.365    -365.7314    134.6552
           x |  -98.57582   2.285769   -43.13   0.000    -103.0613   -94.09034
          x2 |   1.000001   .0122767    81.45   0.000     .9759098    1.024092
       _cons |   9864.218   111.1206    88.77   0.000      9646.16    10082.28
------------------------------------------------------------------------------

. reg y D x_c x2_c

      Source |       SS       df       MS              Number of obs =     999
-------------+------------------------------           F(  3,   995) =13115.22
       Model |  3.7863e+10     3  1.2621e+10           Prob > F      =  0.0000
    Residual |   957507020   995  962318.613           R-squared     =  0.9753
-------------+------------------------------           Adj R-squared =  0.9753
       Total |  3.8821e+10   998  38898361.8           Root MSE      =  980.98

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           D |  -115.5381   127.4967    -0.91   0.365    -365.7315    134.6552
         x_c |  -98.57582   2.285769   -43.13   0.000    -103.0613   -94.09034
        x2_c |   1.000001   .0122767    81.45   0.000     .9759098    1.024092
       _cons |  -3796.397   227.7894   -16.67   0.000      -4243.4   -3349.394
------------------------------------------------------------------------------
```

Figure: P-order polynomial regressions with and without centering. Only intercept changes.

Kernel regression

- In addition to using p^{th} order polynomials to model the nonlinearities, we can use kernel regression. Neither is right or wrong – they have advantages and disadvantages.
- The nonparametric kernel method has problems because you are trying to estimate regressions at the cutoff point which results in a "boundary problem" (Hahn, Todd and Van der Klaauw 2001)



- While the "true" effect is AB, with a certain bandwidth a rectangular kernel would estimate the effect as $A'B'$
- There is therefore systematic bias with the kernel method if the $f(X)$ is upwards or downwards sloping

Kernel Method - Local linear regression

- The standard solution to this problem is to run local linear nonparametric regression (Hahn, Todd and Van der Klaauw 2001) – substantially reduces the bias
- Think of it as a weighted regression restricted to a window – kernel provides the weights to that regression.

$$(\widehat{a}, \widehat{b}) \equiv_{a,b} \sum_{i=1}^{n} (y_i - a - b(x_i - c_0))^2 K\left(\frac{x_i - c_0}{h}\right) 1(x_i > c_0)$$

where x_i is the value of the running variable, c_0 is the cutoff, K is a kernel function and $h > 0$ is a suitable bandwidth

Optimal bandwidths

- A rectangular kernel would give the same result as taking $E[Y]$ at a given bin on X. The triangular kernel gives more importance to the observations closer to the center.
- While estimating this in a given window of width h around the cutoff is straightforward, it's more difficult to choose this bandwidth (or window), and the method is sensitive to the choice of bandwidth. There is essentially a tradeoff between bias and efficiency. See Lee and Lemieux (2010) for two methods to choose on the bandwidth. This is an active area of research in econometrics.
- Imbens and Kalyanaraman (2012) propose a method for estimating optimal bandwidths which may differ on either side of the cutoff. This can be implemented now with `rdrobust`

Sharp vs. Fuzzy RDD

RDD estimates the causal effect by distinguishing the discontinuous function, $1(X \geq c_0)$, from the smooth selection function, $f(X)$ (though Van der Klaauw uses S instead of X notation)

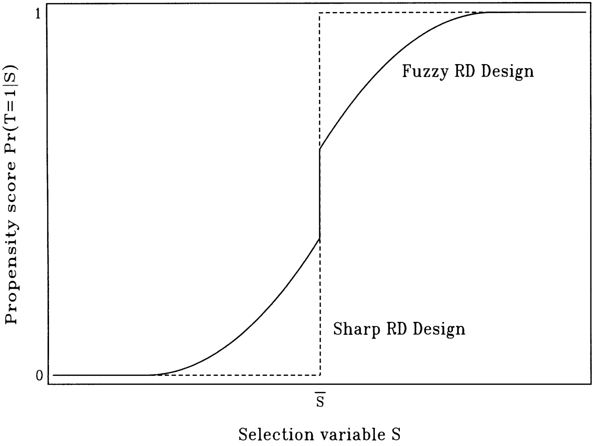


FIGURE 2

ASSIGNMENT IN THE SHARP (DASHED) AND FUZZY (SOLID) RD DESIGN

Probabilistic treatment assignment (i.e. "fuzzy RDD")

The probability of receiving treatment changes discontinuously at the cutoff, c_0, but need not go from 0 to 1

$$lim_{X_i \rightarrow c_0} Pr(D_i = 1 | X_i = c_0) \neq lim_{c_0 \leftarrow X_i} Pr(D_i = 1 | X_i = c_0)$$

Examples: Incentives to participate in some program may change discontinuously at the cutoff but are not powerful enough to move everyone from non participation to participation.

- In the sharp RDD, D_i was *determined* by $X_i \geq c_0$; in the fuzzy RDD, the conditional probability of treatment *jumps* at c_0.
- The relationship between the probability of treatment and X_i can be written as:

$$P[D_i = 1 | X_i] = g_0(X_i) + [g_1(X_i) - g_0(X_i)] T_i$$

where $T_i = 1$ if $(X_i \geq c_0)$ and 0 otherwise.

Use the discontinuity as IV

Wald estimator of treatment effect under Fuzzy RDD

Average causal effect of the treatment is the Wald IV parameter

$$\delta_{\text{Fuzzy RDD}} = \frac{lim_{X\rightarrow c_0} E[Y|X = c_0] - lim_{c_0 \leftarrow X} E[Y|X = c_0]}{lim_{X\rightarrow c_0} E[D|X = c_0] - lim_{c_0 \leftarrow X} E[D|X = c_0]}$$

- Fuzzy RDD is numerically equivalent and conceptually similar to instrumental variables
  - Numerator: "jump" in the regression of the outcome on the running variable, X. "Reduced form".
  - Denominator: "jump" in the regression of the treatment indicator on the running variable X. "First stage".
- Use software package to estimate (e.g., `ivregress 2sls` in STATA).
- Same IV assumptions, caveats about compliers vs. defiers, and statistical tests that we discussed with instrumental variables apply here – e.g., check for weak instruments using F test on instrument in first stage, etc.

RD's Relationship to IV

- RDD is really IV with control variables
- Redefine the measurement of R so that R is equal to zero at r_0
- Define $Z = \mathbf{1}(R \geq 0)$
- Then the coefficient on Z in a regression like

```
. reg Y Z R R2 R3
```

  is the reduced form discontinuity, and

```
. reg X Z R R2 R3
```

  is the first stage discontinuity
- The ratio of discontinuities is what is known as "fuzzy RDD" – this is an RD design where you are interested in the effect of X on Y
- A simple way to implement this is using IV

```
. ivregress 2sls Y (X=Z) R R2 R3
```

What does Fuzzy RDD Estimate?

- As Hahn, Todd and van der Klaauw (2001) point out, one needs the same assumptions as in the standard IV framework
- As with other binary IVs, the fuzzy RDD is estimating LATE: the average treatment effect for the compliers
- In RDD, the compliers are those whose treatment status changed as we moved the value of x_i from just to the left of c_0 to just to the right of c_0

Challenges to RDD

- Treatment is not as good as randomly assigned around the cutoff, c_0, when agents are able to manipulate their running variable scores. This happens when:
  1. the assignment rule is known in advance
  2. agents are interested in adjusting
  3. agents have time to adjust
  4. Examples: re-take an exam, self-reported income, etc.
- Some other unobservable characteristic changes at the threshold, and this has a direct effect on the outcome.
  - In other words, the cutoff is endogenous
  - Example: Age thresholds used for policy (i.e., person turns 18, and faces more severe penalties for crime) is correlated with other variables that affect the outcome (i.e., graduation, voting rights, etc.)

Econometricians and applied social scientists have developed several formalized tests to evaluate the severity of these problems which we now discuss.

Test 1: Manipulation of the running variable

Sorting on the running variable (i.e., Manipulation)

Assume a desirable treatment, D, and an assignment rule $X \geq c_0$. If individuals sort into D by choosing X such that $X \geq c_0$, then we say individuals are sorting on the running variable.

- Motivating example: Suppose a doctor plans to randomly assign heart patients to a statin and a placebo to study the effect of the statin on heart attacks within 10 years. The doctor randomly assigns patients to two different waiting rooms, A and B, and plans to give those in A the statin and those in B the placebo. If some of the patents learn of the planned treatment assignment mechanism, what would we expect to happen? And how would you check for it?

McCrary Density Test

We would expect waiting room A to become *crowded*. In the RDD context, sorting on the running variable implies heaping on the "good side" of c_0

- McCrary (2008) suggests a formal test. Under the null the density should be continuous at the cutoff point. Under the alternative hypothesis, the density should increase at the kink (where D is viewed as good)
  1. Partition the assignment variable into bins and calculate frequencies (i.e., number of observations) in each bin
  2. Treat those frequency counts as dependent variable in a local linear regression
- The McCrary Density Test has become **mandatory** for every analysis using RDD.
  - If you can estimate the conditional expectations, you evidently have data on the running variable. So in principle you can always do a density test
  - You can download the (no longer supported) STATA ado package, `DCdensity`, to implement McCrary's density test (`http://eml.berkeley.edu/~jmccrary/DCdensity/`)
  - You can install `rdd` for R too (`http://cran.r-project.org/web/packages/rdd/rdd.pdf`)

Caveats about McCrary Density Test

- For RDD to be useful, you already need to know something about the mechanism generating the assignment variable and how susceptible it could be to manipulation. Note the rationality of economic actors that this test is built on.
- A discontinuity in the density is "suspicious" – it *suggests* manipulation of X around the cutoff is probably going on. In principle one doesn't need continuity.
- This is a high-powered test. You need a lot of observations at c_0 to distinguish a discontinuity in the density from noise.

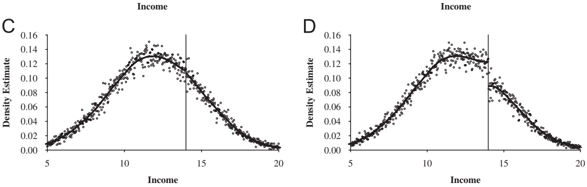


Figure: Panel C is density of income when there is no pre-announcement and no manipulation. Panel D is the density of income when there is pre-announcement and manipulation. From McCrary (2008).

Test 2: Balance test on covariates

- This is a type of placebo test. For RDD to be valid in your study, you do not want to observe a discontinuity around the cutoff, c_0, for average values of covariates that should **not** be affected by the treatment – e.g., pretreatment characteristics
- Question: What does a jump in the average values of pre-treatment characteristics have to do with the continuous (smoothness) assumption?
  - Choose other pre-treatment covariates, Z and do a similar graphical plot as you did for Y
  - You do **not** want to see a jump around the cutoff, c_0
  - A formal balance test involves the same procedure used to estimate the treatment effect, only use Z instead of Y as a LHS variable
  - Can combine test on multiple covariates into a single test statistic using seemingly unrelated regression (`SUR`) or a single "stacked" regression

Visualizing Placebo Test

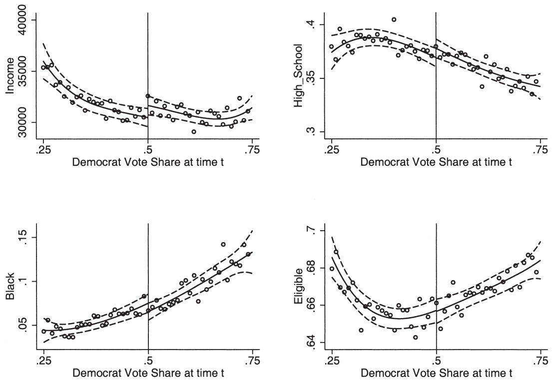


Figure: Figure 3 from Lee, Moretti and Butler (2004), "Do Voters Affect or Elect Policies?" *Quarterly Journal of Economcis*. Panels refer to (top left to bottom right) the following district characteristics: real income, percentage with high-school degree, percentage black, percentage eligible to vote. Circles represent the average characteristic within intervals of 0.01 in Democratic vote share. The continuous line represents the predicted values from a fourth-order polynomial in vote share fitted separately for points above and below the 50 percent threshold. The dotted line represents the 95 percent confidence interval.

Test 3: Jumps at non-discontinuous points

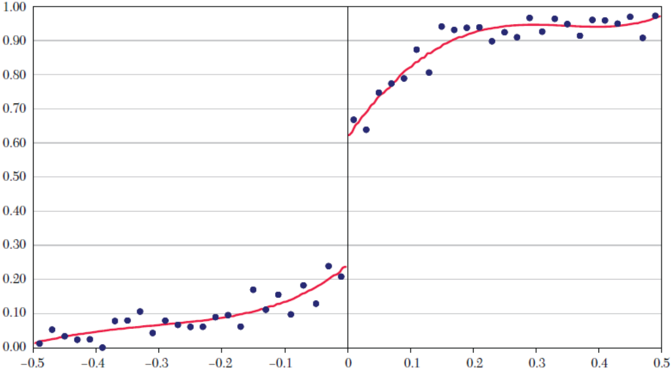
- Imbens and Lemieux (2008) suggest to look at one side of the discontinuity (e.g., $X < c_0$), take the median value of the running variable in that section, and pretend it was a discontinuity, c_0'
- Then test whether in reality there is a discontinuity at c_0'. You do **not** want to find anything.
- Another kind of placebo test. Similar to synthetic control falsification exercises where you look for non-effects in non-treatment periods. (See Abadie, Diamond and Hainmueller 2014)

RDD graphs

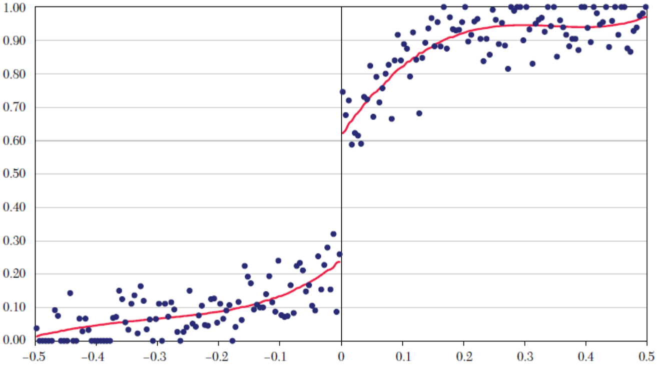
A graphical analysis should be an integral part of any RDD.

1. **Outcome by running variable, (X_i):**
    - The standard graph showing the discontinuity in the outcome variable
    - Construct bins and average the outcome within bins on both sides of the cutoff
    - You should also look at different bin sizes when constructing these graphs (see Lee and Lemieux, 2010, for details)
    - Plot the running variables, X_i, on the horizontal axis and the average of Y_i for each bin on the vertical axis
    - You may also want to plot a relatively flexible regression line on top of the bin means
    - Inspect whether there is a discontinuity at c_0
    - Inspect whether there are other unexpected discontinuities

Example: Outcomes by Running Variables



Example: Outcomes by Running Variables with smaller bins



More RDD Graphs

2. **Probability of treatment by running variable if fuzzy RDD**
   - In a fuzzy RDD, you also want to see that the treatment variable jumps at c_0
   - This tells you whether you have a first stage
3. **Covariates by a running variable**
   - Construct a similar graph to the one before but using a covariate as the "outcome"
   - There should be **no jump** in other covariates at the discontinuity, c_0.
   - If the covariates jump at the discontinuity, one would doubt the identifying assumption

Example: Covariates by Running Variable

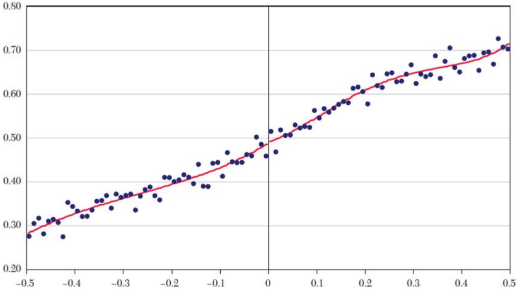


Figure 17. Discontinuity in Baseline Covariate (Share of Vote in Prior Election)

More RDD graphs!

4. **The density of the running variable**
   - One should plot the number of observations in each bin.
   - This plot allows to investigate whether there is a discontinuity in the distribution of the running variable at the threshold
   - This would suggest that people can manipulate the running variable around the threshold.
   - This is an indirect test of the identifying assumption that each individual has imprecise control over the assignment variable

Density of the running variable

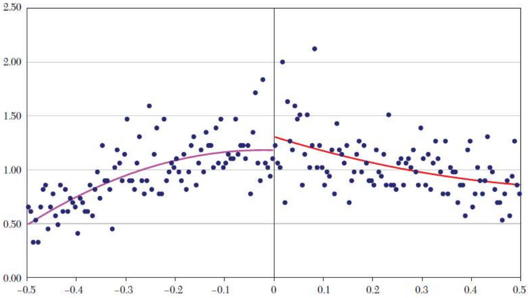


Figure 16. Density of the Forcing Variable (Vote Share in Previous Election)

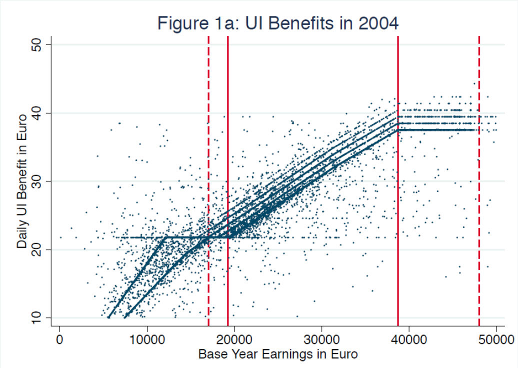
Regression Kink Design

- Card, Lee, Pei and Weber (2012) introduces a variant of the RDD which they call regression kink design (RKD)
- They essentially use a kink in some policy rule to identify the causal effect of the policy
- Instead of a jump in the outcome you now expect a jump in the first derivative

Unemployment benefits in Austria

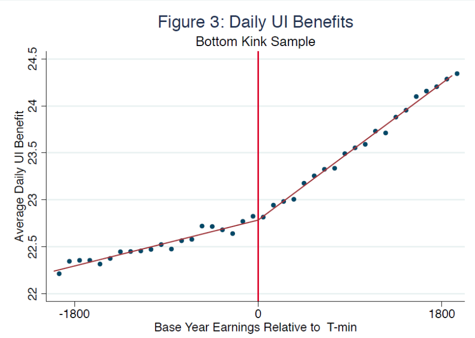
- They apply their design to answer the question whether the level of unemployment benefits affects the length of unemployment in Austria
- Unemployment benefits are based on income in a base period
- The benefit formula for unemployment exhibits 2 kinks
  - There is a minimum benefit level that isn't binding for people with low earnings
  - Then benefits are 55% of the earnings in the base period
  - There is a maximum benefit level that is adjusted every year
- People with dependents get small supplements (which is the reason why one can distinguish five “solid” lines in the following graph)
- Not everyone receives benefits that correspond one to one to the formula because of mistakes in the administrative data

Base Year Earnings and Unemployment Benefits



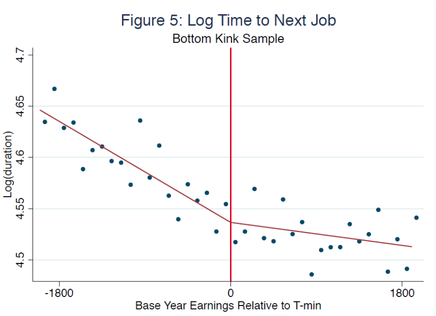
- The graph shows unemployment benefits (vertical axis) as a function of pre-unemployment earnings (horizontal axis)

Base Year Earnings and Benefits for Single Individuals



- Bin-size: 100 euros
- For single individuals UI benefits are flat below the cutoff. The relationship is still upward sloping because of family benefits.

Time to Next Job for Single Individuals



- People with higher base earnings have less trouble finding a job (negative slope)
- There is a kink: the relationship becomes shallower once benefits increase more.