

Regression Discontinuity: Advanced Topics

NYU Wagner
Rajeev Dehejia



Summary of RD assumptions

- ✓ The treatment is determined at least in part by the assignment variable
- ✓ There is a discontinuity in the level of treatment at some cutoff value of the assignment variable (selection on observables at the cutpoint)
- ✓ Units cannot precisely manipulate the assignment variable to influence whether they receive the treatment or not
- ✓ Other variables that affect the treatment do not change discontinuously at the cutoff



Assessing the validity of an RD

- It is impossible to test the continuity assumption directly, but we can test some implications of it
- Namely, all observed predetermined characteristics (e.g. date of birth, parents' education level...) should have identical distributions on either side of the cutoff, in the limit, as we approach smaller and smaller bandwidths. That is, **there should be no discontinuities in the observables.**
- Again there is an analogy to an experiment: we cannot test whether unobserved characteristics are balanced, but we can test the observables. Rejection calls the randomization into question.



Internal & external validity

- The strength of the RD design is its internal validity, arguably the strongest of any quasi-experimental design
- External validity may be limited
- **Sharp RD** (SRD) provides estimates for the subpopulation with $X=c$, that is those right at the cutoff of the assignment variable.
 - The discontinuity is a weighted average treatment effect where weights are proportional to the ex ante likelihood that an individual's realization of X will be close to the threshold.
- **Fuzzy RD** (FRD) restricts the estimates further to compliers at the cutoff (more on this below)
- You need to justify extrapolation to other subpopulations (e.g., treatment homogeneity)



Threats to an RD analysis

There are **three general types of threats** to an RD design:

- 1. Other (i.e. non-outcome) variables change discontinuously at the cutoff**
 - Test for jumps in covariates, including pretreatment values of the outcome and the treatment
- 2. There are discontinuities at other values of the assignment variable**
- 3. Manipulation of the assignment variable**
 - Test for continuity in the density of the assignment variable at the cutoff



Specification checks

- A. Discontinuities in average covariates
- B. A discontinuity in the distribution of the forcing variable
- C. Discontinuities in average outcomes at other values
- D. Sensitivity to bandwidth choice
- E. Fuzzy RD design
- F. Extension: Regression kink design



A. Discontinuities in average covariates

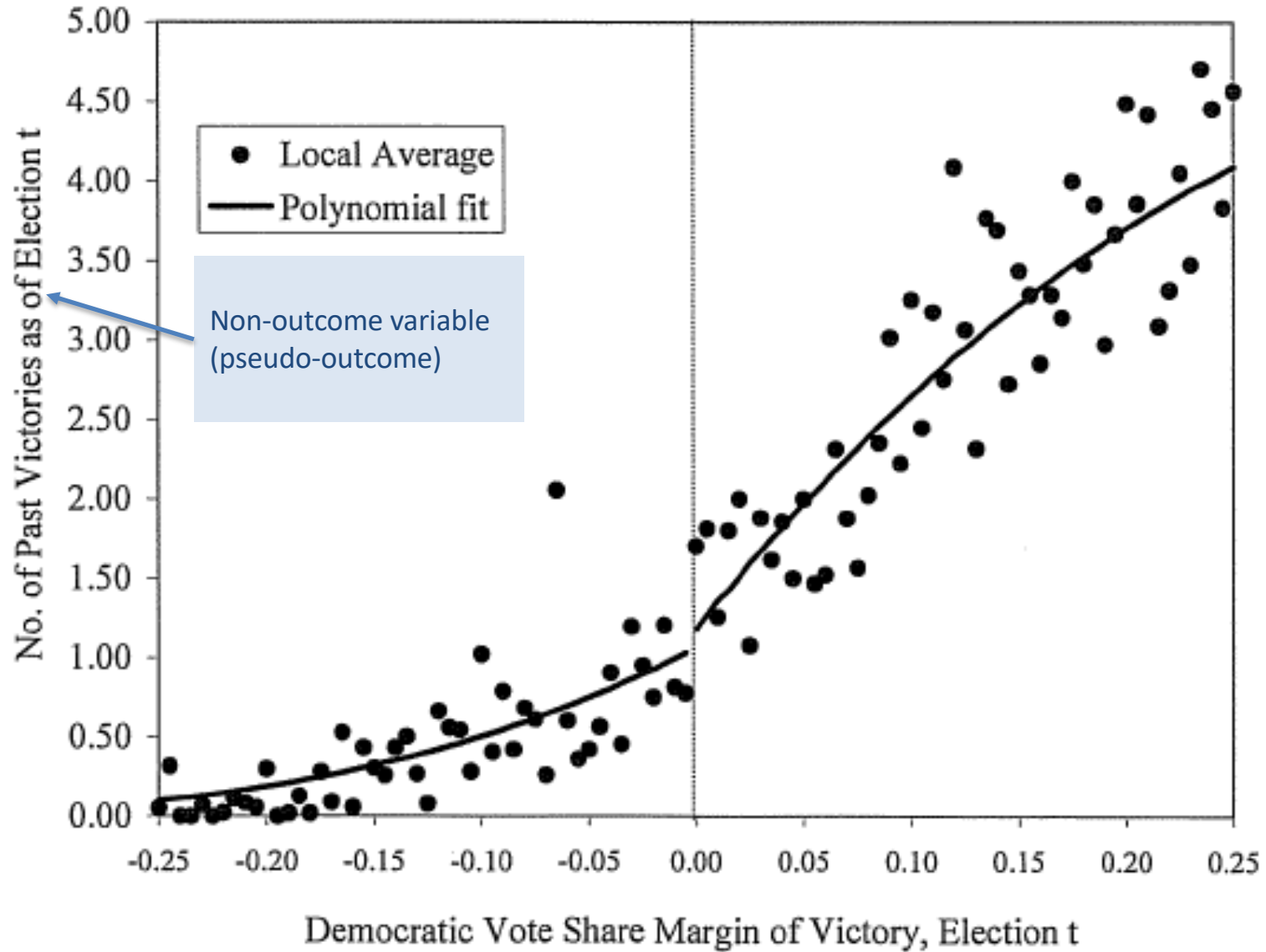
What this tests: Test the null hypothesis of a zero average effect on pseudo outcomes known not to be affected by the treatment.

Intuition behind pseudo outcomes: Such variables include covariates that are by definition not affected by the treatment. Such tests are familiar from settings with identification based on unconfoundedness assumptions.

Why it's worth testing: Although not required for the validity of the design, in most cases, the reason for the discontinuity in the probability of the treatment does not suggest a discontinuity in the average value of covariates. If we find such a discontinuity, it typically casts doubt on the assumptions underlying the RD design.



Figure IIb: Candidate's Accumulated Number of Past Election Victories, by Margin of Victory in Election t: local averages and parametric fit



A. Balance checks in average covariates

- Lee (2008) uses the regression discontinuity design to estimate party incumbency advantage in U.S. House elections.

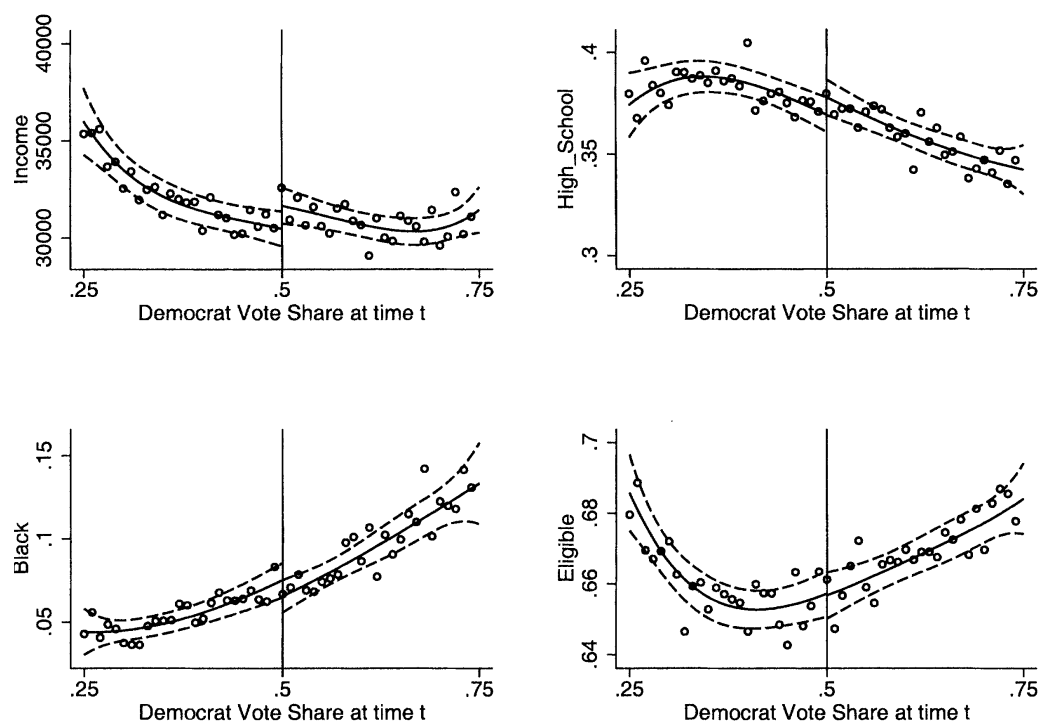


FIGURE III
Similarity of Constituents' Characteristics in Bare Democrat and Republican
Districts—Part 1



A. Balance checks in average covariates

- Lee (2008) uses the regression discontinuity design to estimate party incumbency advantage in U.S. House elections.

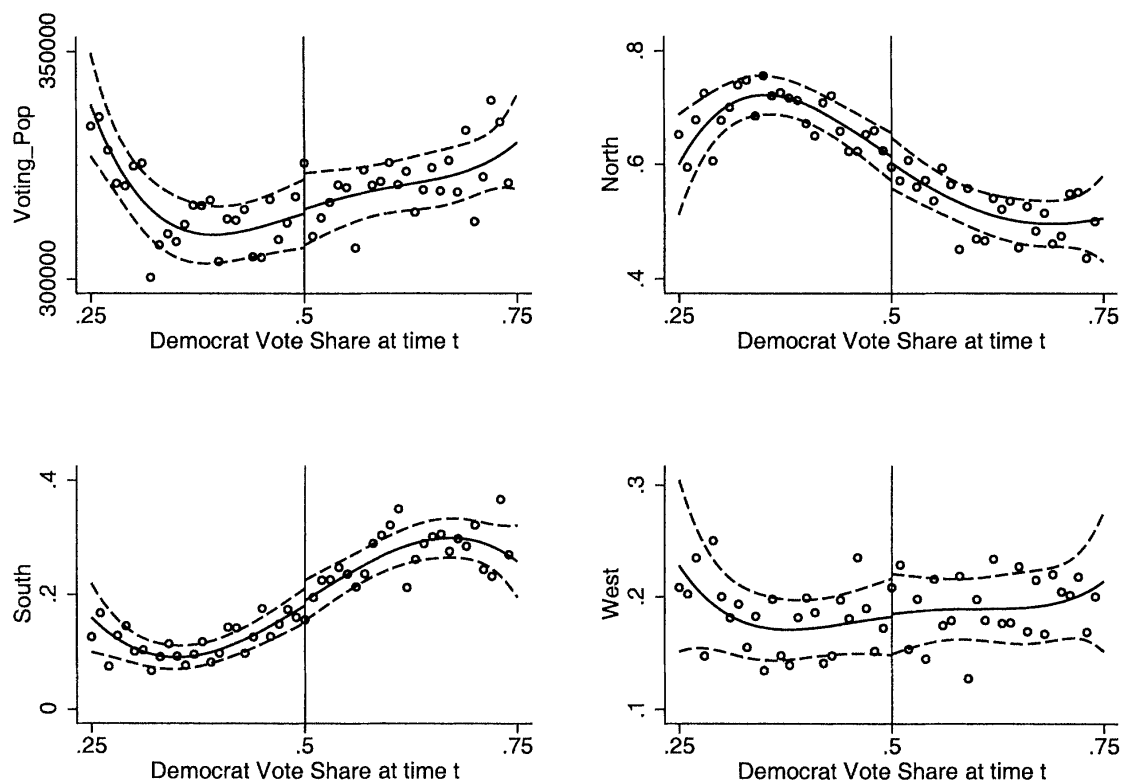


FIGURE IV
Similarity of Constituents' Characteristics in Bare Democrat and Republican
Districts—Part 2



A. Balance checks in average covariates

- Lee (2008) uses the regression discontinuity design to estimate party incumbency advantage in U.S. House elections.
- However ... more qualifications soon....



B. A discontinuity in the distribution of the forcing variable: Sorting/bunching/manipulation

- Subjects or program administrators may invalidate the continuity assumption if they strategically manipulate X to be just above or below the cutoff.
- This is a concern especially if the exact value of the cutoff is known to the subjects in advance.
- This type of behavior, if it exists, may create a discontinuity in the distribution of X at the cutoff (i.e., “bunching” to the right or to the left of the cutoff)



B. A discontinuity in the distribution of the forcing variable: Sorting/bunching/manipulation

- **Intuition behind manipulation:** If individuals have control over the assignment variable, then we should expect them to sort into (out of) treatment if treatment is desirable (undesirable)
 - Think of a means-tested income support program, or an election
 - Those just above the threshold will be a mixture of those who would have passed and those who barely failed without manipulation.
- **Expect discontinuity in density if there is sorting:** If individuals have precise control over the assignment variable, we would expect the density of X to be zero just below the threshold but positive just above the threshold (assuming the treatment is desirable).
 - McCrary (2008) provides a formal test for manipulation of the assignment variable in an RD. The idea is that the marginal density of X should be continuous without manipulation and hence we look for discontinuities in the density around the threshold.
 - How precise must the manipulation be in order to threaten the RD design? See Lee and Lemieux (2010).
- **Think about sorting mechanism:** This means that when you run an RD you must know something about the mechanism generating the assignment variable and how susceptible it could be to manipulation



B. A discontinuity in the distribution of the forcing variable

McCrary (2007) suggests testing the null hypothesis of the continuity of the density of the covariate that underlies the assignment at the discontinuity point, against the alternative of a jump in the density function at that point.

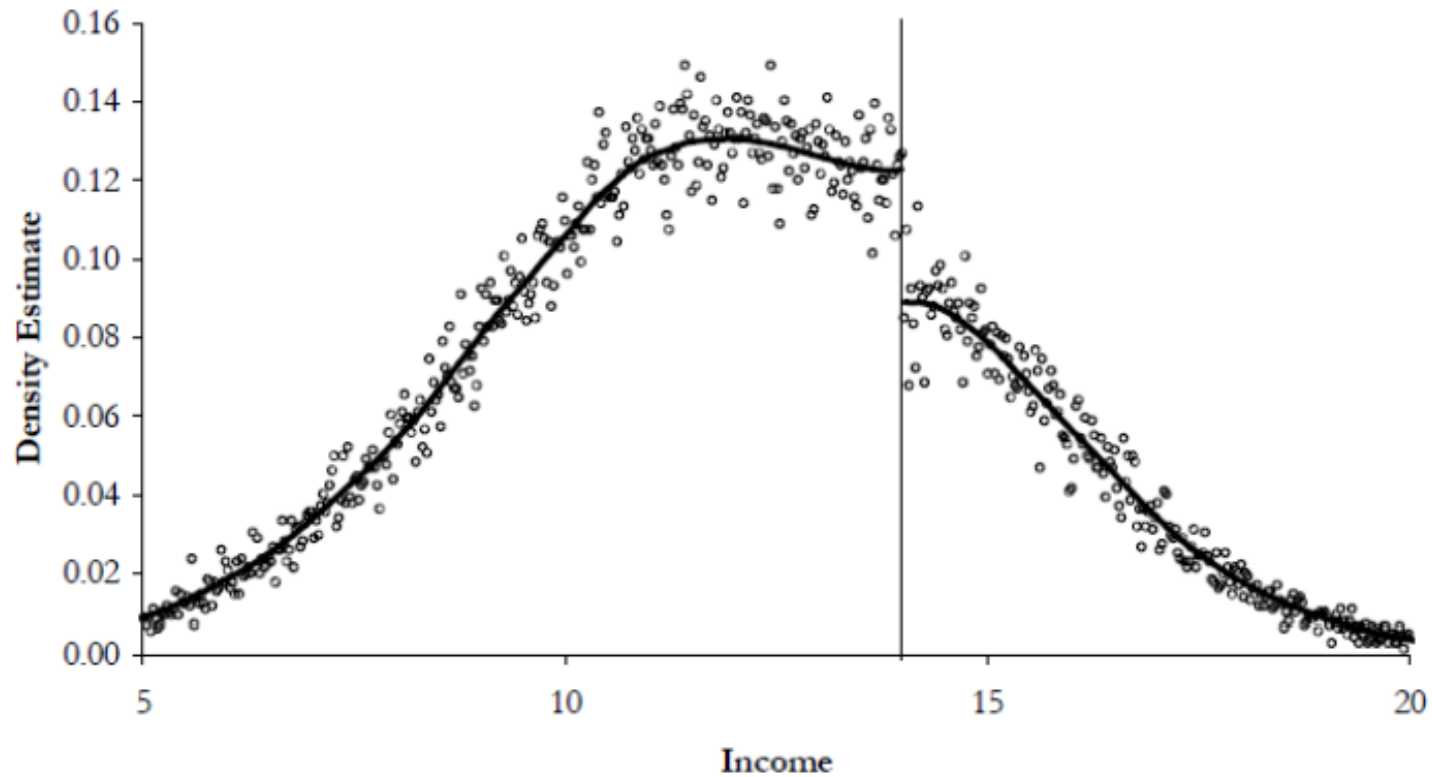
Again, in principle, the design does not require continuity of the density of X at c , but a discontinuity is suggestive of violations of the no-manipulation assumption.

If in fact individuals partly manage to manipulate the value of X in order to be on one side of the boundary rather than the other, one might expect to see a discontinuity in this density at the discontinuity point.



B. Example of manipulation

D. Density of Income
with Pre-Announcement and Manipulation



- An income support program in which those earning under \$14,000 qualify for support
- Simulated data from McCrary 2008



B. Discontinuity of the forcing variable (cont'd)

- 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 patients learn of the planned treatment assignment mechanism, we would expect them to proceed to waiting room A.
- We would expect for waiting room A to become crowded.
- In the regression discontinuity context, this is analogous to expecting the running variable to be discontinuous at the cutoff, with surprisingly many individuals just barely qualifying for a desirable treatment assignment and surprisingly few failing to qualify.



B. Discontinuity of the forcing variable (cont'd)

- Partial manipulation occurs when the running variable is under the agent's control, but also has an idiosyncratic element (e.g., can manipulate test score, but only imperfectly).
- Typically, partial manipulation of the running variable does not lead to identification problems (analogous to fuzzy RD).
- Complete manipulation occurs when the running variable is entirely under the agent's control.
- Typically, complete manipulation of the running variable does lead to identification problems.



B. Test of discontinuity of the forcing variable (cont'd)

- The density test may not be informative unless the existence of the program induces agents to adjust the running variable in one direction only.
 - Intuition: If you have sorting in both ways, it could cancel out.
- The density test could also fail, even when there is no failure of identification – but still often a useful test.

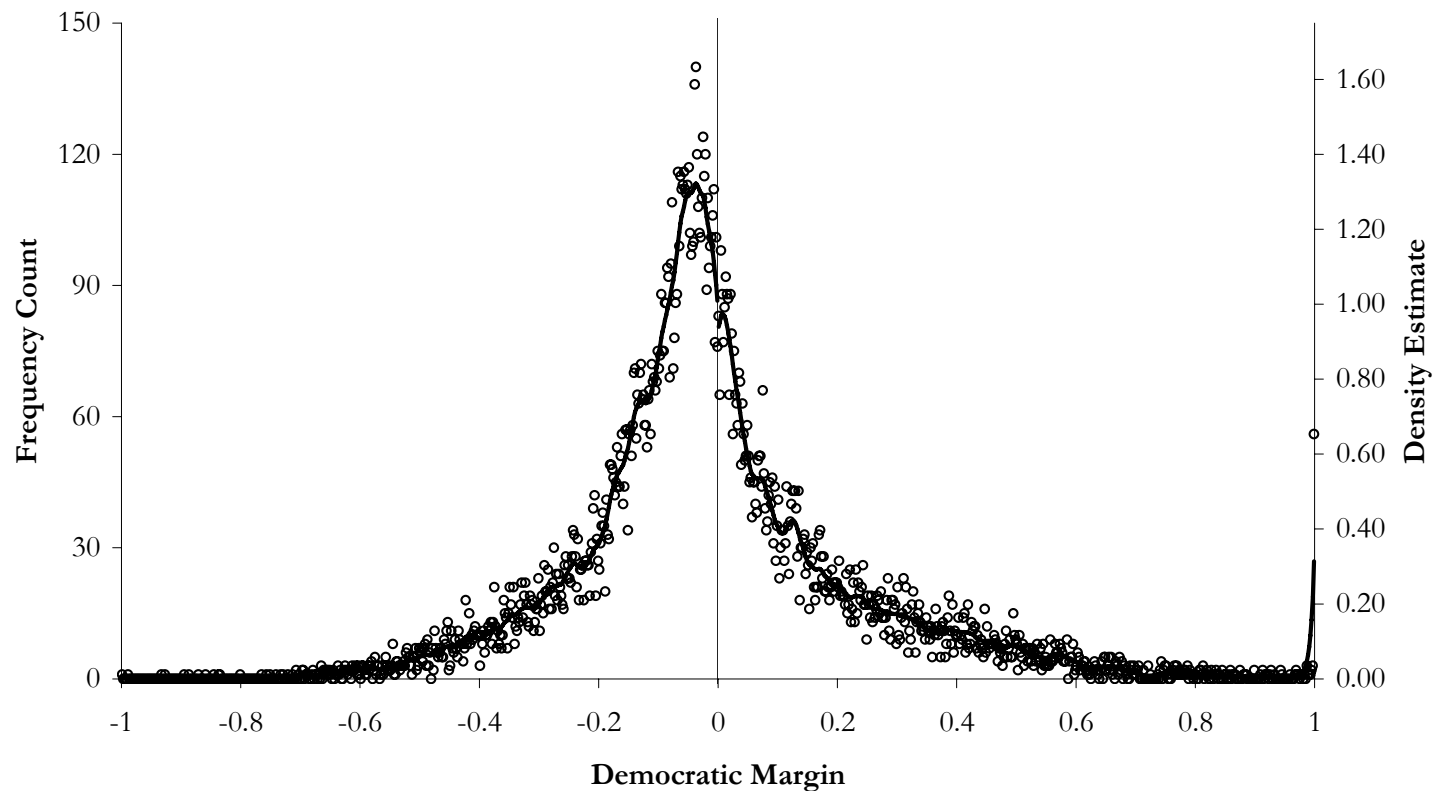
The test:

1. Estimate a very under-smoothed histogram. The bins for the histogram are defined carefully enough that no one histogram bin includes points both to the left and right of the point of discontinuity.
2. Estimate a local linear smoothing of the histogram. The midpoints of the histogram bins are treated as a regressor, and the normalized counts of the number of observations falling into the bins are treated as an outcome variable.
3. Use local linear regression estimates from step 2 to test for discontinuity.



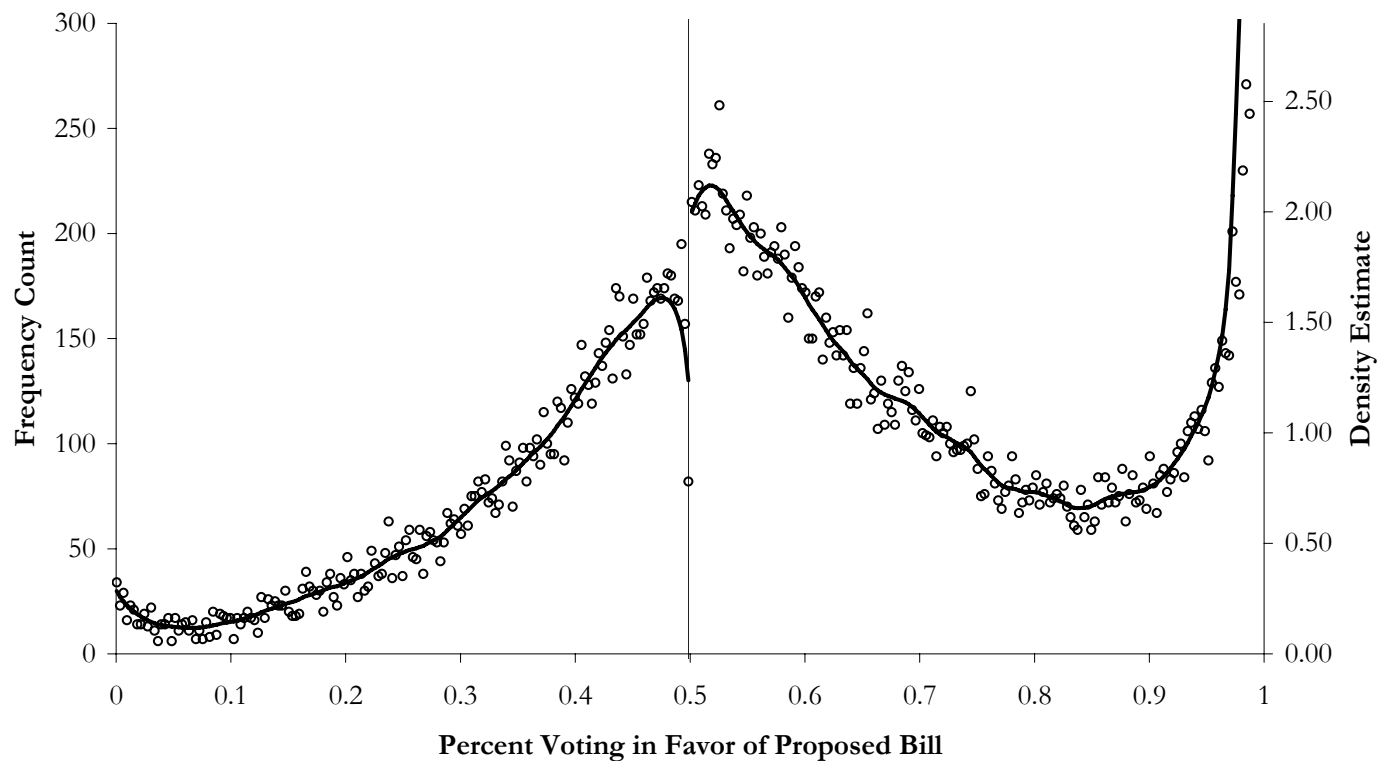
B. Test of discontinuity of the forcing variable (cont'd)

Figure 4. Democratic Vote Share Relative to Cutoff:
Popular Elections to the House of Representatives, 1900-1990



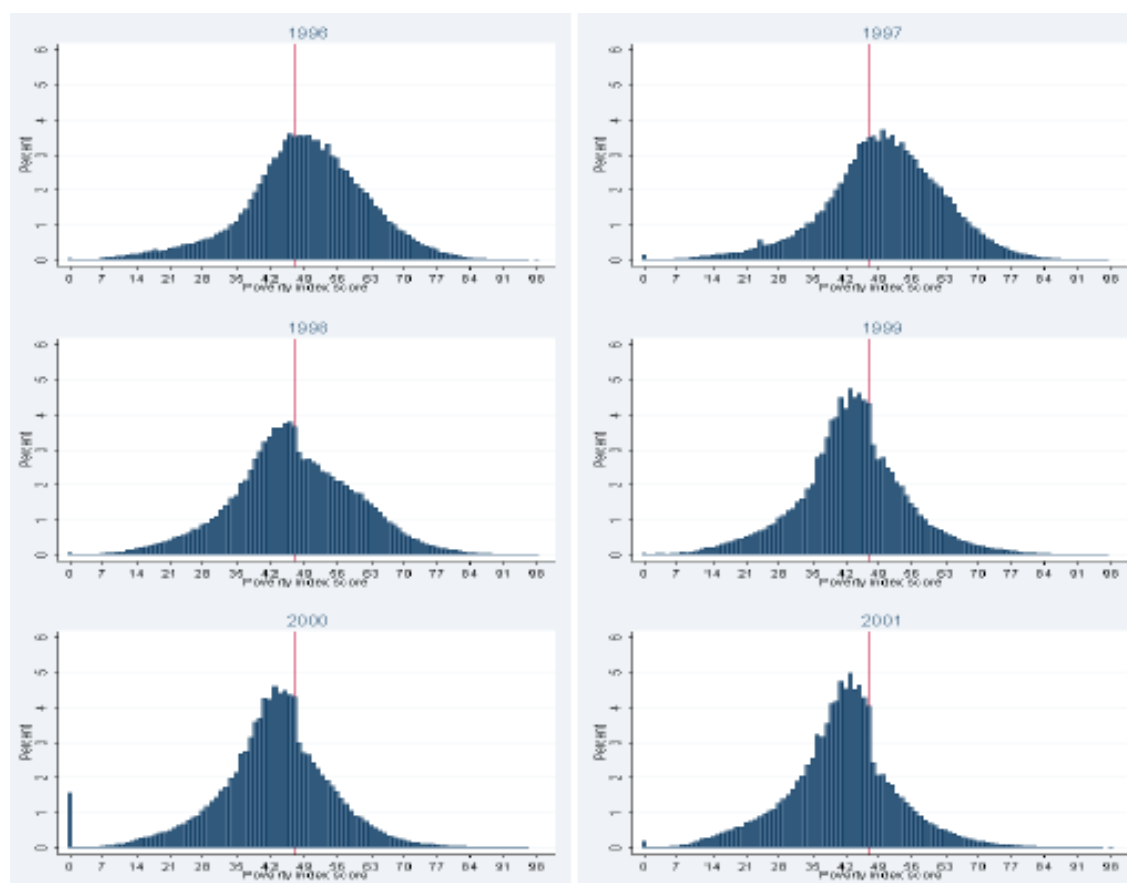
B. Test of discontinuity of the forcing variable (cont'd)

Figure 5. Percent Voting Yeay:
Roll Call Votes, U.S. House of Representatives, 1857-2004



RDD: Sorting/bunching (Camacho & Conover, 2010)

Example: Manipulation of a poverty index in Colombia. A poverty index is used to decide eligibility for social programs. The algorithm to create the poverty index becomes public during the second half of 1997.

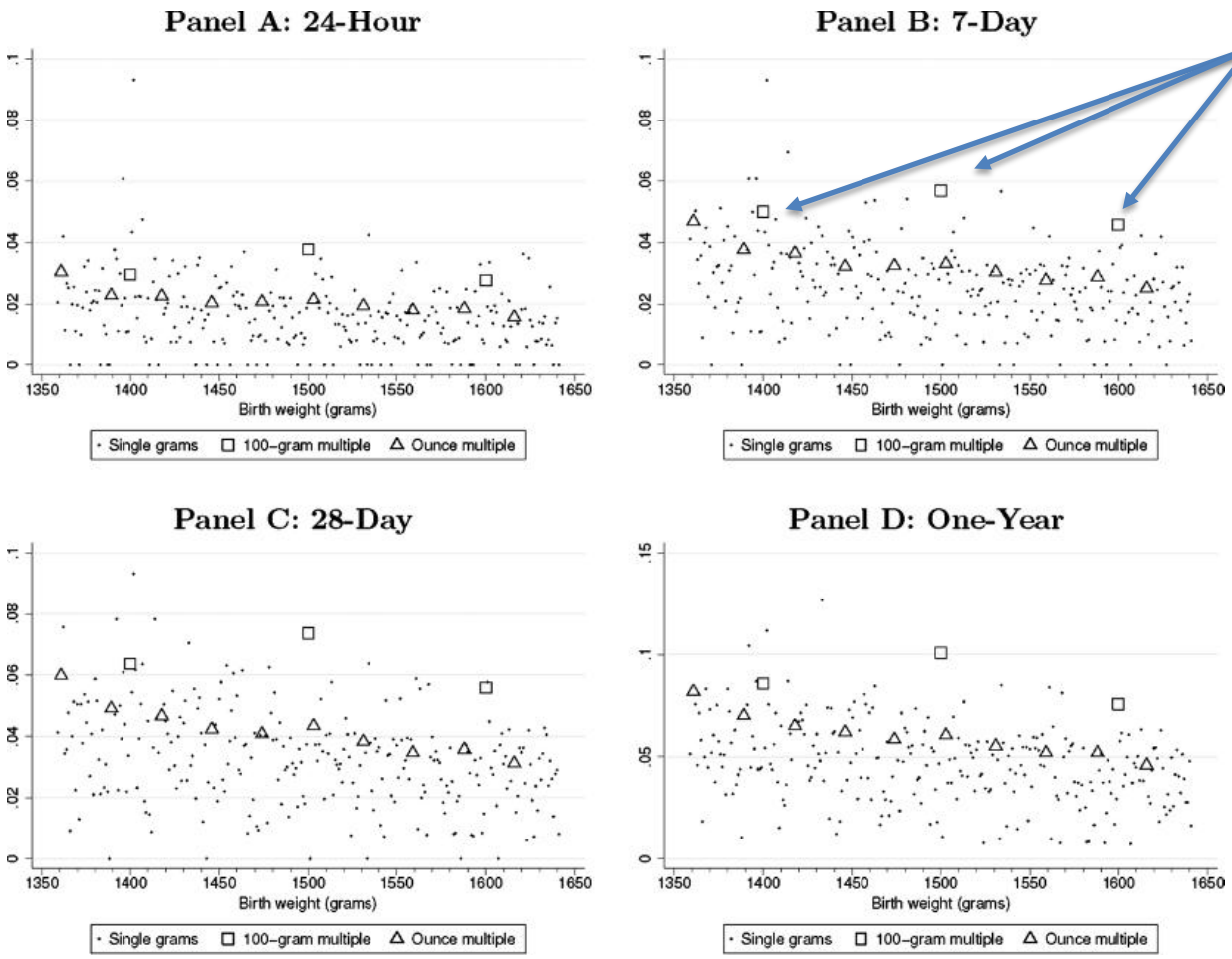


C. Placebo tests

- Almond et al. (2010) use a medical definition of “very low birth weight” < 1500 grams, to estimate the effect of additional medical care on newborns.
- They find that newborns just below the 1500 grams cutoff receive additional treatment and survive with higher probability than newborns just above the cutoff.
- However, Barreca et al. (2010) find evidence of non-random rounding at 100-gram multiples of birth weight.
- Newborns of low socioeconomic status, who tend to be less healthy, are disproportionately represented at 100-gram multiples (balance check).
- As a result, newborns with birth weights just below or above each 100-multiple have more-favorable mortality outcomes than newborns with birth weights at the cutoffs.



C. Placebo tests (Barreca et al., 2010)



This is a phenomenon called "heaping". Non-random spikes in mortality rate at the 100g interval rates. The implication is that less precise record-keeping is applied to these babies, which indicates a lower quality of care

FIGURE I



C. Discontinuities in average outcomes at other values

Taking the subsample with $X_i < c$ we can test for a jump in the conditional mean of the outcome at the median of the forcing variable.

To implement the test, use the same method for selecting the bandwidth as before. Also estimate the standard errors of the jump and use this to test the hypothesis of a zero jump.

Repeat this using the subsample to the right of the cutoff point with $X_i \geq c$. Now estimate the jump in the regression function and at $q_{X,1/2,r}$ and test whether it is equal to zero.



C. Discontinuities in average outcomes at other values

Example from DiNardo and Lee (2008)

FIGURE IIIa
Recognition, Subsequent Certification or Decertification, by Union Vote Share.

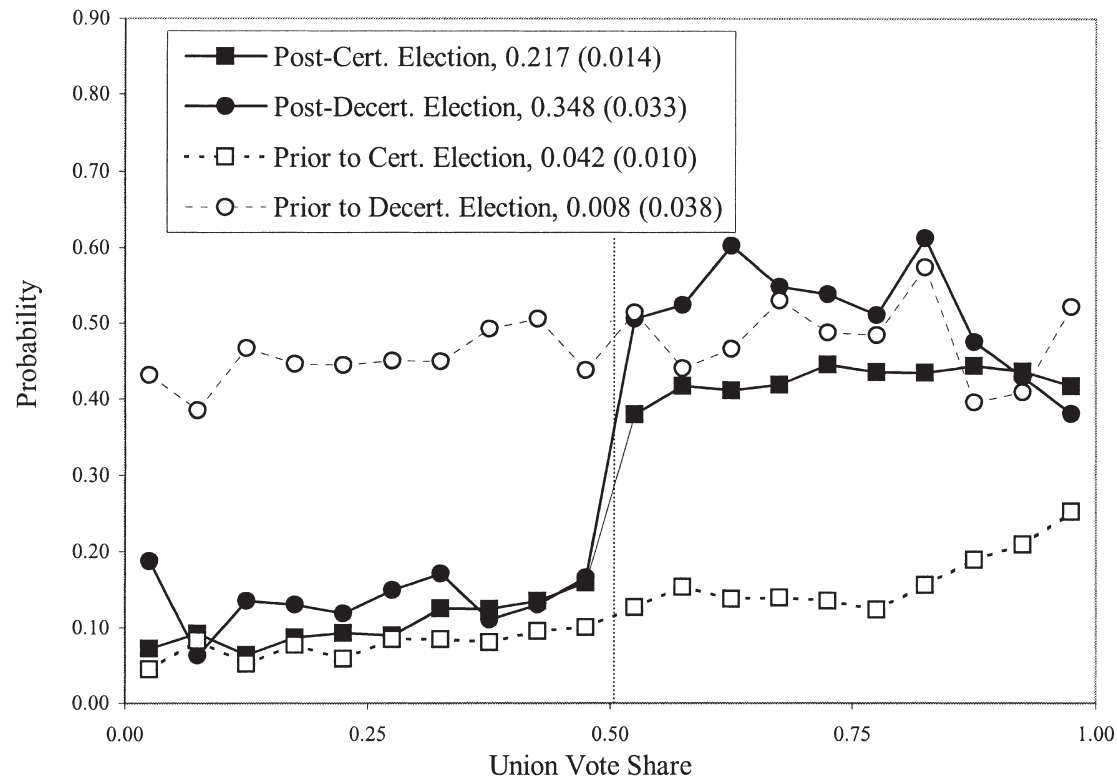


FIGURE IIIb



C. Discontinuities in average outcomes at other values

Example from DiNardo and Lee (2008)

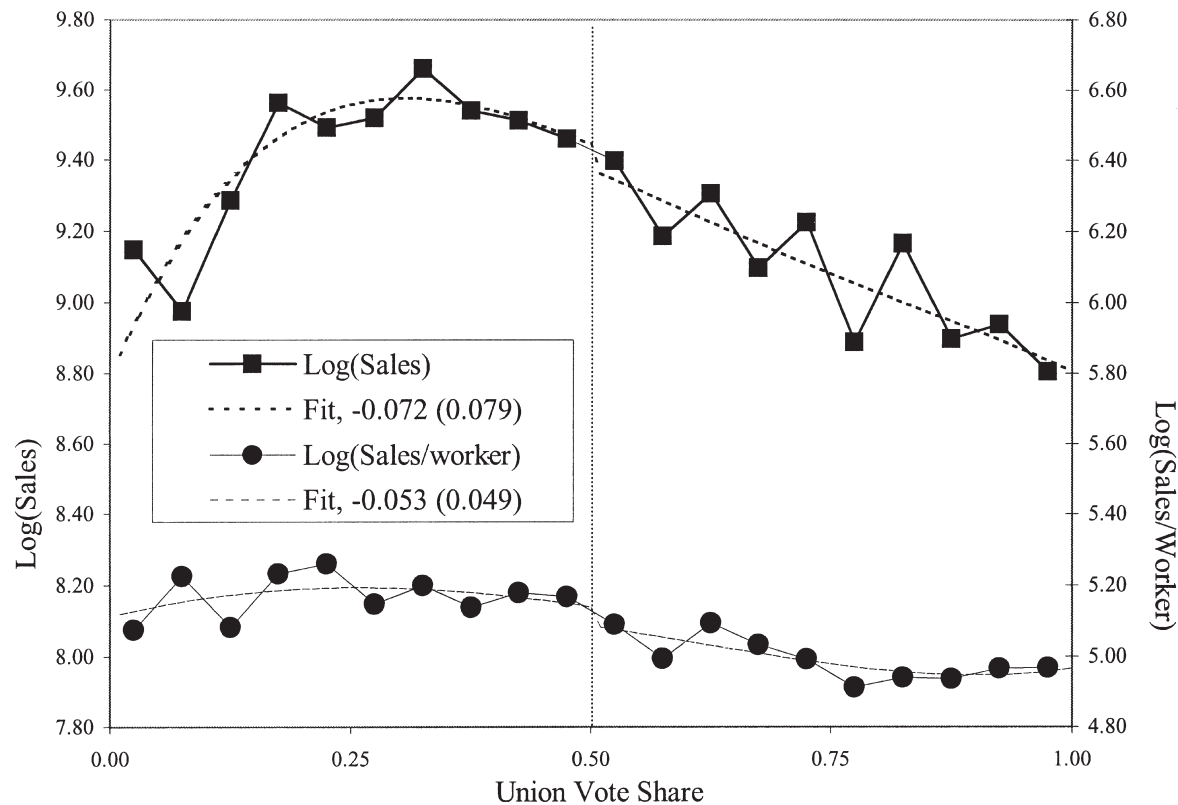


FIGURE V
Log(Sales) and Log(Sales/worker), by Union Vote Share



D. Bandwidth selection and sensitivity

- There are two general methods for selection bandwidth
 - Ad hoc, or substantively derived (e.g., elections between 48-52% are “close”)
 - Data driven
 - We discuss below



D. Bandwidth selection and sensitivity (cont'd)

For Polynomial Regression

- Choosing the order of the polynomial is analogous to the choice of bandwidth
- Informal approach: pick a reasonable number (e.g., 4th order)
- Two approaches
 - Use the Akaike information criterion (AIC) for model selection:
 $AIC = \ln(\hat{\sigma}^2) + 2p$, where $\hat{\sigma}^2$ is the mean squared error of the regression and p is the number of model parameters
 - Select a natural set of bins (as you would for an RD graph) and add bin dummies to the model and test their joint significance. Add higher order terms to the polynomial until the bin dummies are no longer jointly significant.
 - This also turns out to be a test for the presence of discontinuities in the regression function at points other than the cutoff, which you'll want to do anyway.



D. Bandwidth selection and sensitivity (cont'd)

For Local Linear Regression bandwidth selection represents the familiar tradeoff between bias and precision

- When the local regression function is more or less linear, there isn't much of a tradeoff so bandwidth can be larger
- So pick a “reasonable h ”.
- Optimal bandwidth choice:
 - The intuition is that you set up the objective of minimizing the (mean squared) error between the estimated treatment effect and actual treatment effect.
 - This gets very technical.
 - But the good news is that the two Stata commands do it for you.
 - Imbens' code at <http://users.nber.org/~rdehejia/!@SAEM/Topic%2007%20RD%20Advanced/rdob.ado> gives you bandwidth choice using `rdob`.
 - Also the program `rd (ssc install rd)` automatically uses the IK optimal bandwidth.



D. Bandwidth selection and sensitivity (cont'd)

In both cases

- In practice, you may want to focus on results for the “optimal” bandwidth, but **it’s important to test for lots of different bandwidths**. Think of the optimal bandwidth only as a starting point.
- If results critically depend on a particular bandwidth, they are less credible and choice of bandwidth requires a substantive justification.
- In principle, the optimal bandwidth for testing discontinuities in covariates may not be the same as the optimal bandwidth for the treatment. Again, follow the practice of testing robustness to variations in bandwidth.



E. Fuzzy RD design

- **Cutoff does not perfectly determine treatment but creates a discontinuity in the probability of receiving the treatment**
- For example:
 - The probability of being offered a scholarship may jump at a certain SAT score (above which the applications are given “special consideration”)
 - Incentives to participate in a program may change discontinuously at a threshold, but the change is not powerful enough to move all units from nonparticipation to participation
- For units close to the cutoff we can use

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

as an instrument for D_i .

- We estimate the effect of the treatment for compliers: those units (close to the discontinuity, $X_i \cong c$) whose treatment status, D_i , depends on Z_i .



E. Fuzzy RD design

- The idea is that for units that are very close to the discontinuity Z_i can act as an instrument
- The LATE parameter is:

$$\lim_{\substack{c-\varepsilon \leq X \leq c+\varepsilon \\ \varepsilon \rightarrow 0}} \left(\frac{E[Y|Z=1] - E[Y|Z=0]}{E[D|Z=1] - E[D|Z=0]} \right),$$

or

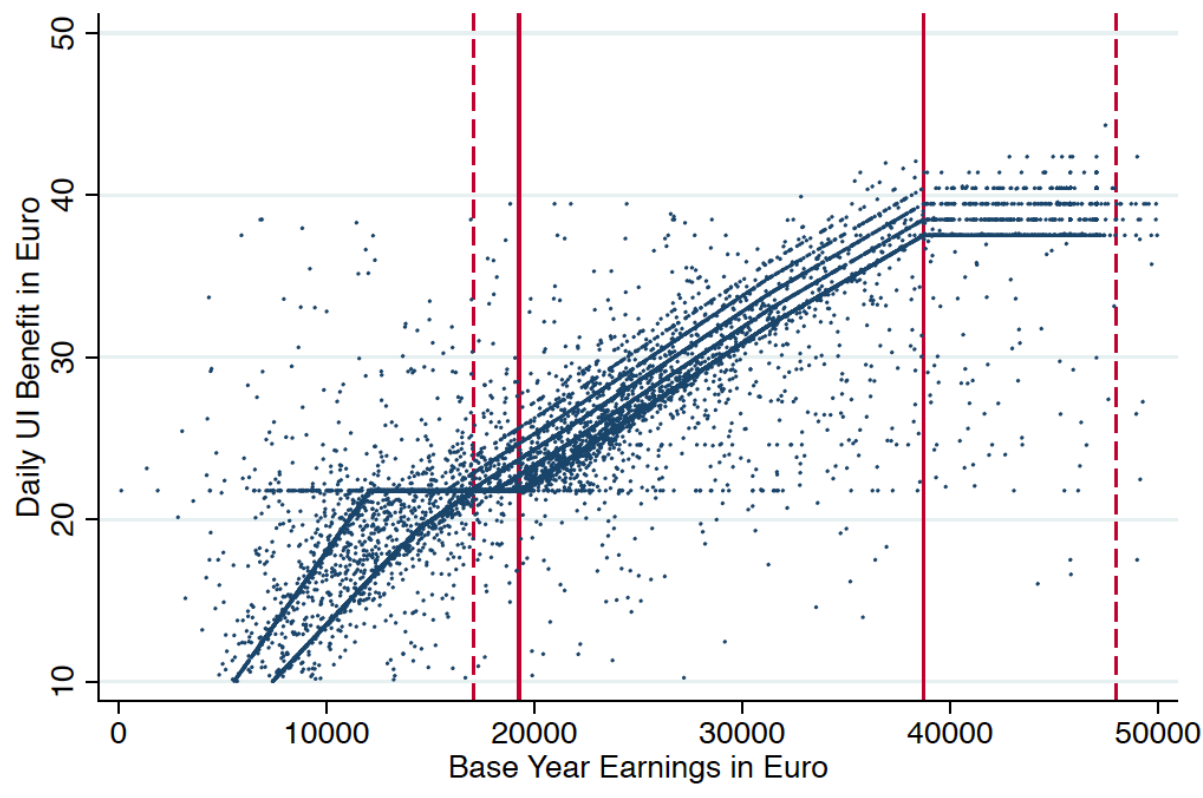
$$\frac{\lim_{x \downarrow c} E[Y|X=x] - \lim_{x \uparrow c} E[Y|X=x]}{\lim_{x \downarrow c} E[D|X=x] - \lim_{x \uparrow c} E[D|X=x]}$$

- This suggests:
 1. Run a sharp RDD for Y
 2. Run a sharp RDD for D
 3. Divide your estimate in step 1 by your estimate in step 2
- Alternatively, run instrumental variables for those units with $X \cong c$



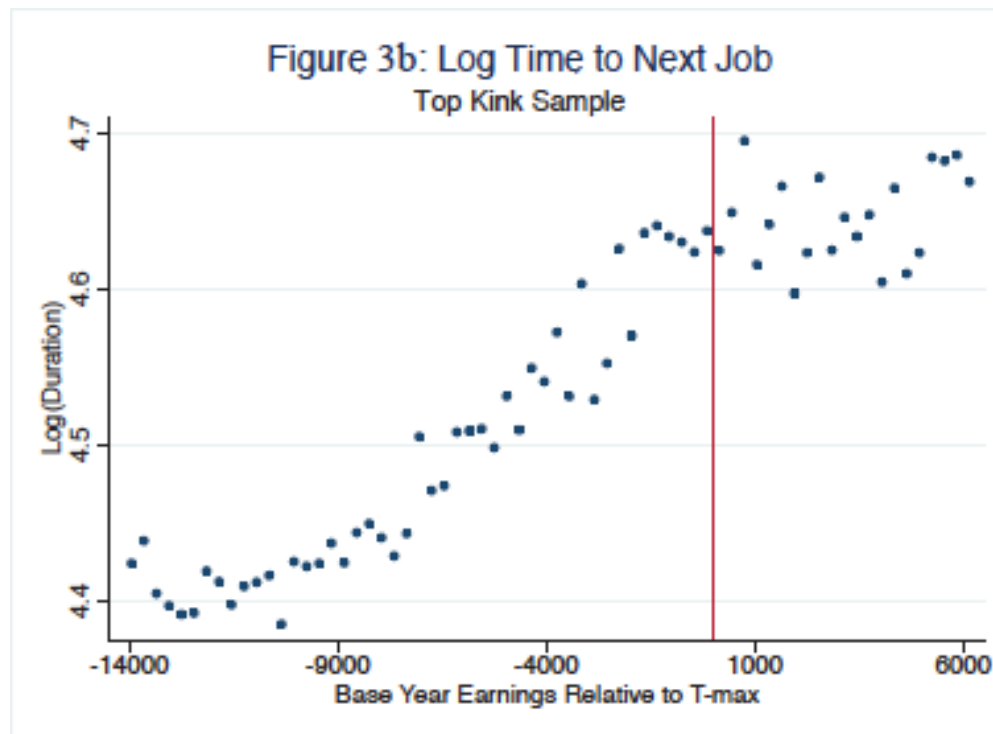
F. Regression kink design

- In some situations at the cutoff it is the slope of the treatment intensity that changes, not the level of treatment assignment (0 to 1).
- Classic example is unemployment benefits where your benefit is a function of prior earnings.



F. Regression kink design

- But then you expect that time to next job varies with base year earnings continuously (no jump), but with a change in slope.



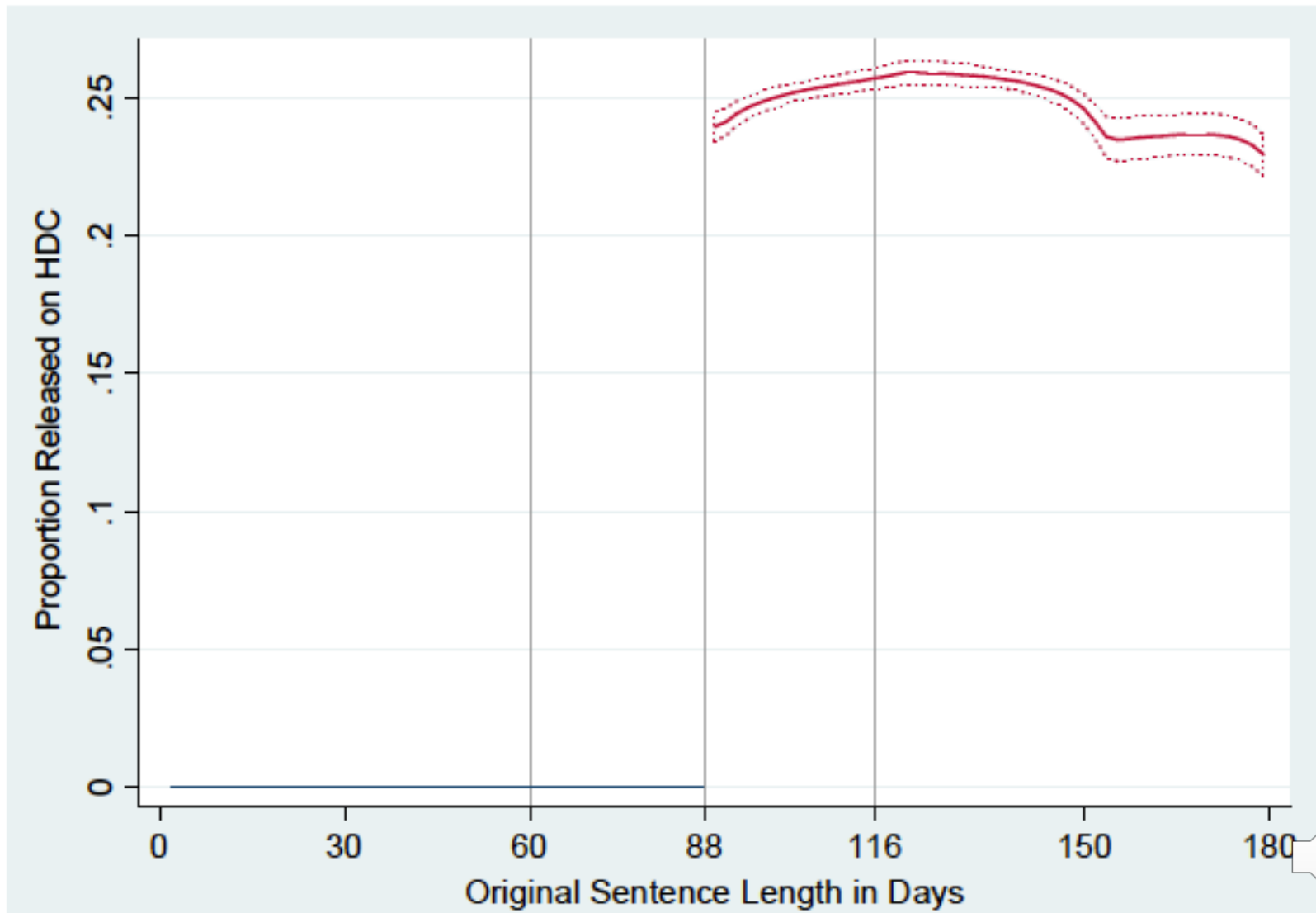
Early release program (Marie, 2009)

- Prison systems in many countries suffer from overcrowding and high recidivism rates after release
- Some countries use early discharge of prisoners on electronic monitoring
- Difficult to estimate impact of early release program on future criminal behavior: best behaved inmates are usually the ones to be released early
- Marie (2008) considers the Home Detention Curfew (HDC) program in England and Wales
- This is a fuzzy RDD: Only offenders sentenced to more than three month (88 days) in prison are eligible for HDC, but not all of those are offered HDC



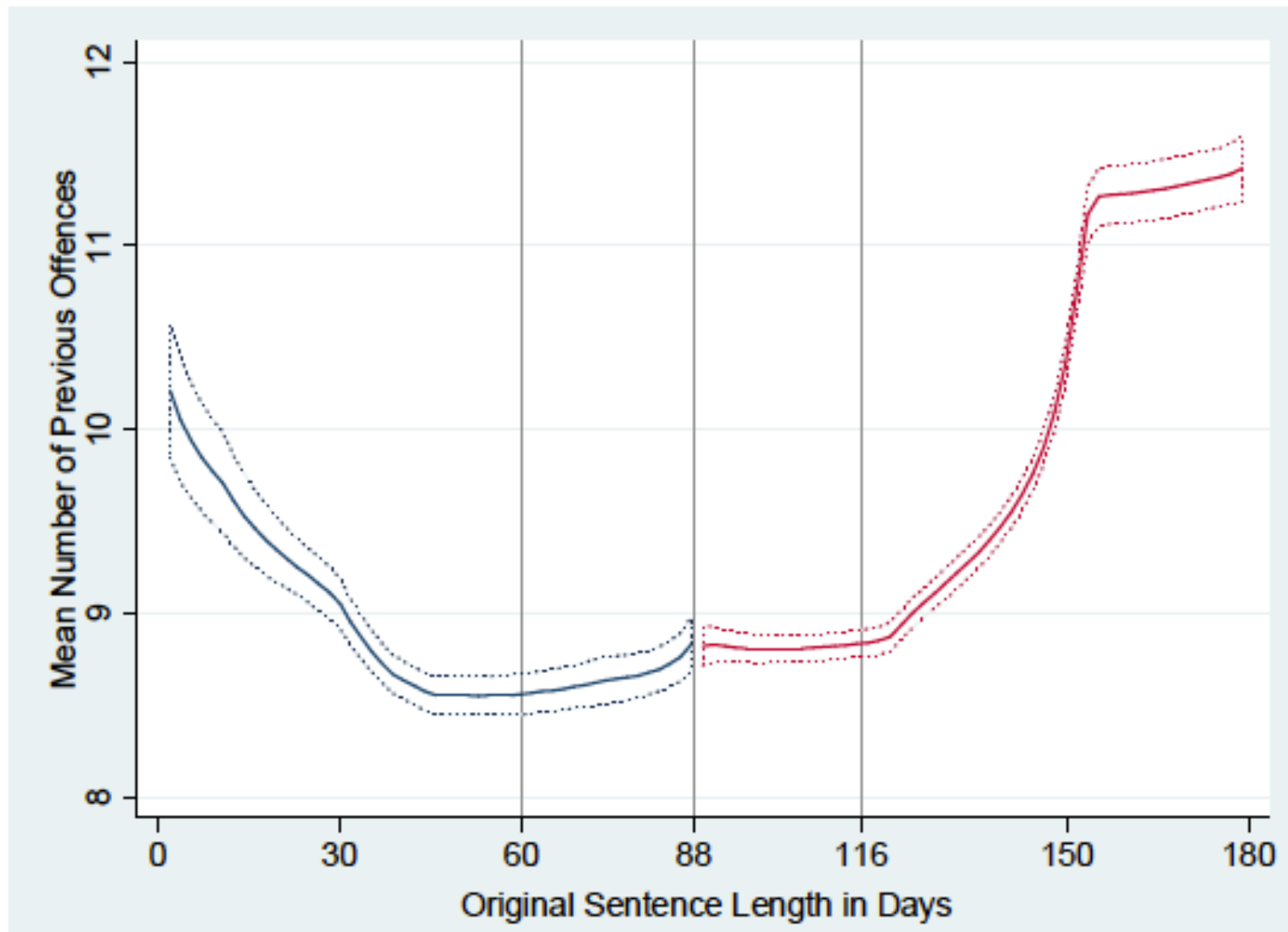
Early release program (Marie, 2009)

Figure 2 : Proportion Released on HDC by Original Sentence Length



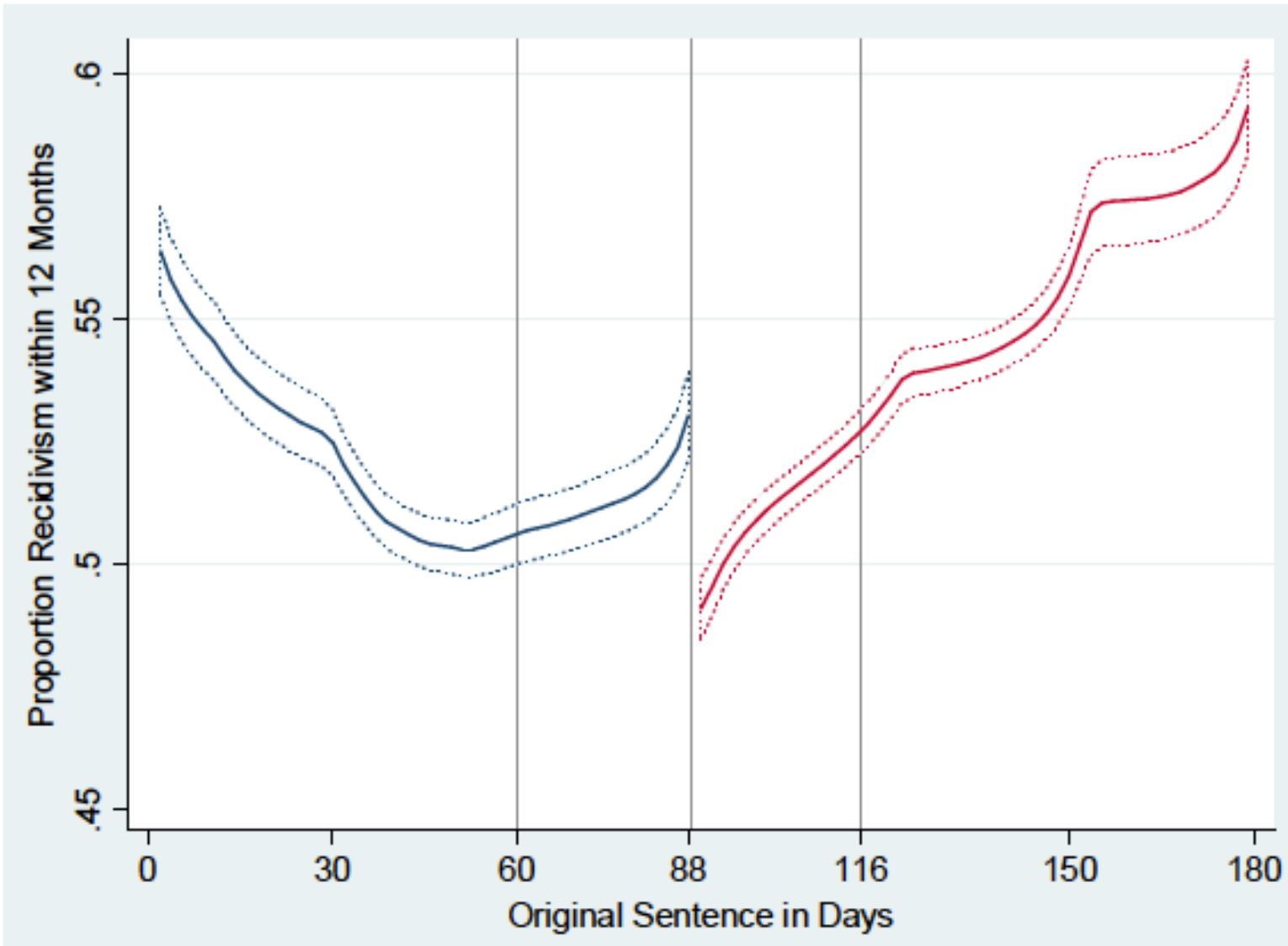
Early release program (Marie, 2009)

Figure 3 : Number of Previous Offences by Original Sentence Length



Early release program (Marie, 2009)

Figure 4: One Year Recidivism Rate by Original Sentence Length



Early release program (Marie, 2009)

Table 5: RD Estimates of HDC Impact on Recidivism

Panel A: Recidivism Within 12 Months of Release	Estimation on Individuals Sentenced to Between 58 and 118 Days: +/- 4 Weeks		
	(1)	(2)	(3)
Discontinuity of HDC Participation Around Threshold ($HDC^+ - HDC^-$)	.242 (.003)	.243 (.003)	.237 (.003)
Difference in Recidivism Around Threshold ($Rec^+ - Rec^-$)	-.023 (.005)	-.022 (.005)	-.016 (.005)
Estimated Effect of HDC on Recidivism Participation ($Rec^+ - Rec^-$) / ($HDC^+ - HDC^-$)	-.094 (.020)	-.090 (.018)	-.066 (.018)
Controls	No	Yes	Yes
Prison Fixed Effects	No	No	Yes
Sample Size	41,761	41,761	41,761

Panel B: Recidivism Within 24 Months of Release	Estimation on Individuals Sentenced to Between 58 and 118 Days: +/- 4 Weeks		
	(1)	(2)	(3)
Discontinuity of HDC Participation Around Threshold ($HDC^+ - HDC^-$)	.242 (.003)	.243 (.003)	.237 (.003)
Difference in Recidivism Around Threshold ($Rec^+ - Rec^-$)	-.019 (.005)	-.019 (.004)	-.013 (.005)
Estimated Effect of HDC on Recidivism Participation ($Rec^+ - Rec^-$) / ($HDC^+ - HDC^-$)	-.079 (.020)	-.077 (.018)	-.053 (.019)
Controls	No	Yes	Yes
Prison Fixed Effects	No	No	Yes
Sample Size	41,761	41,761	41,761

Note: Robust standard errors in parenthesis. The estimation is based on individuals sentenced to between 59 and 118 days. The controls included in column (2) are: gender, age, ethnic minority, breached in the past number previous offences, month and year of release dummies, and the type of crime incarcerated for (8 types). The same model with 126 prison establishment fixed effects is reported in column (3).



“Regression-Discontinuity Designs and Popular Elections: Implications of Pro-Incumbent Bias in Close U.S. House Races”

by Caughey and Sekhon (CS)



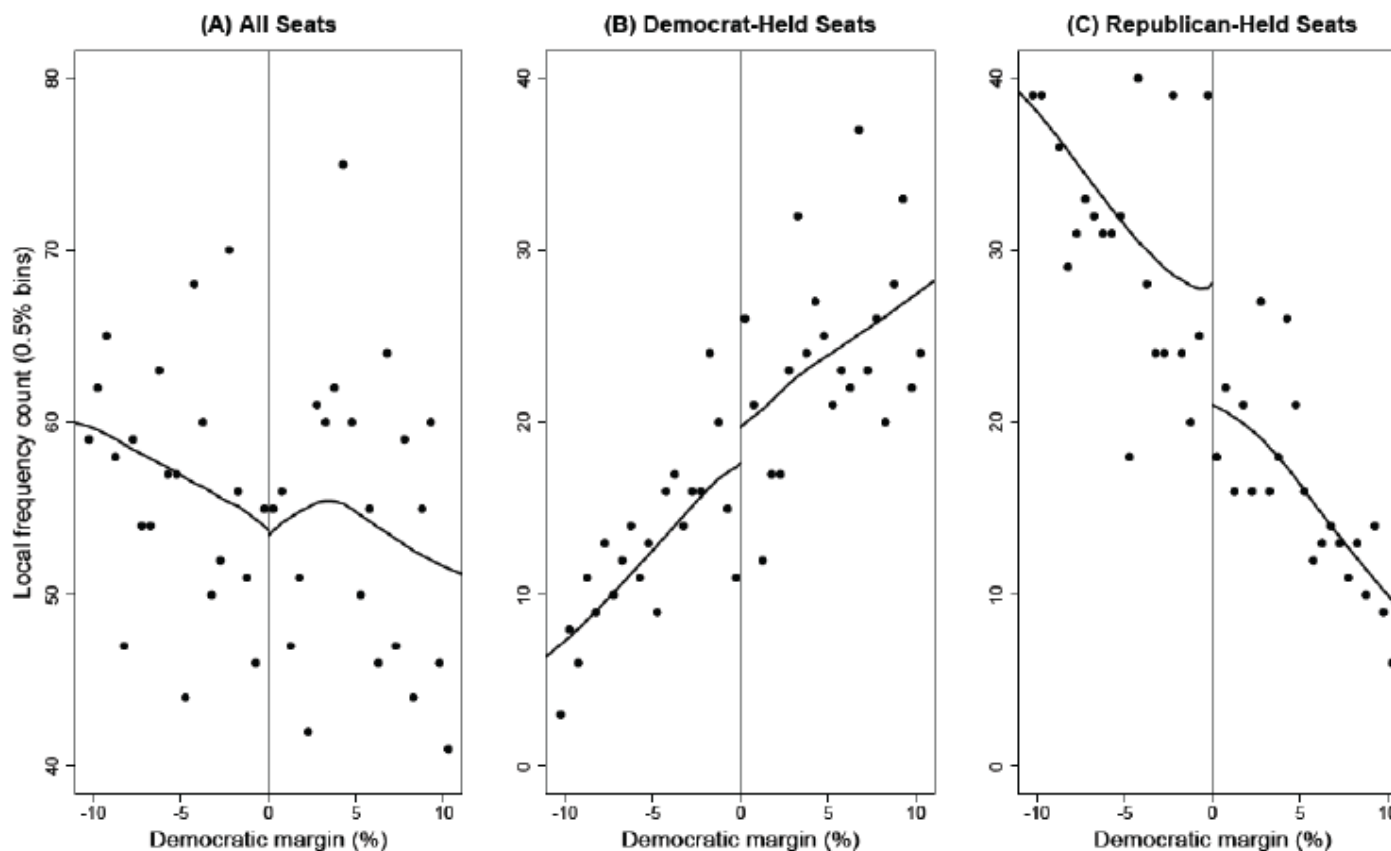
Basic argument

- “Close elections are not like other elections”
 - Strategic political actors have strong incentives to target their resources where they will have the greatest marginal impact
- There is an incumbency advantage even in very close elections
 - The incumbent wins disproportionately and has greater financial resources
- This finding, along with other covariate imbalances at the cut point, calls into question the LMB incumbency advantage results and, more generally, the assumption that outcomes in close elections are “as good as randomly assigned”
 - Note that CS critique Lee (2008), not LMB, but the implications for the incumbency advantage results in both papers are the same



The Lee (& McCrary) tests for manipulation

Figure 1: Local frequency counts of the Democratic margin in U.S. House elections, 1946-2008, with local linear estimate overlaid. Bandwidths were chosen by the algorithm proposed by Imbens and Kalyanaraman (2009).

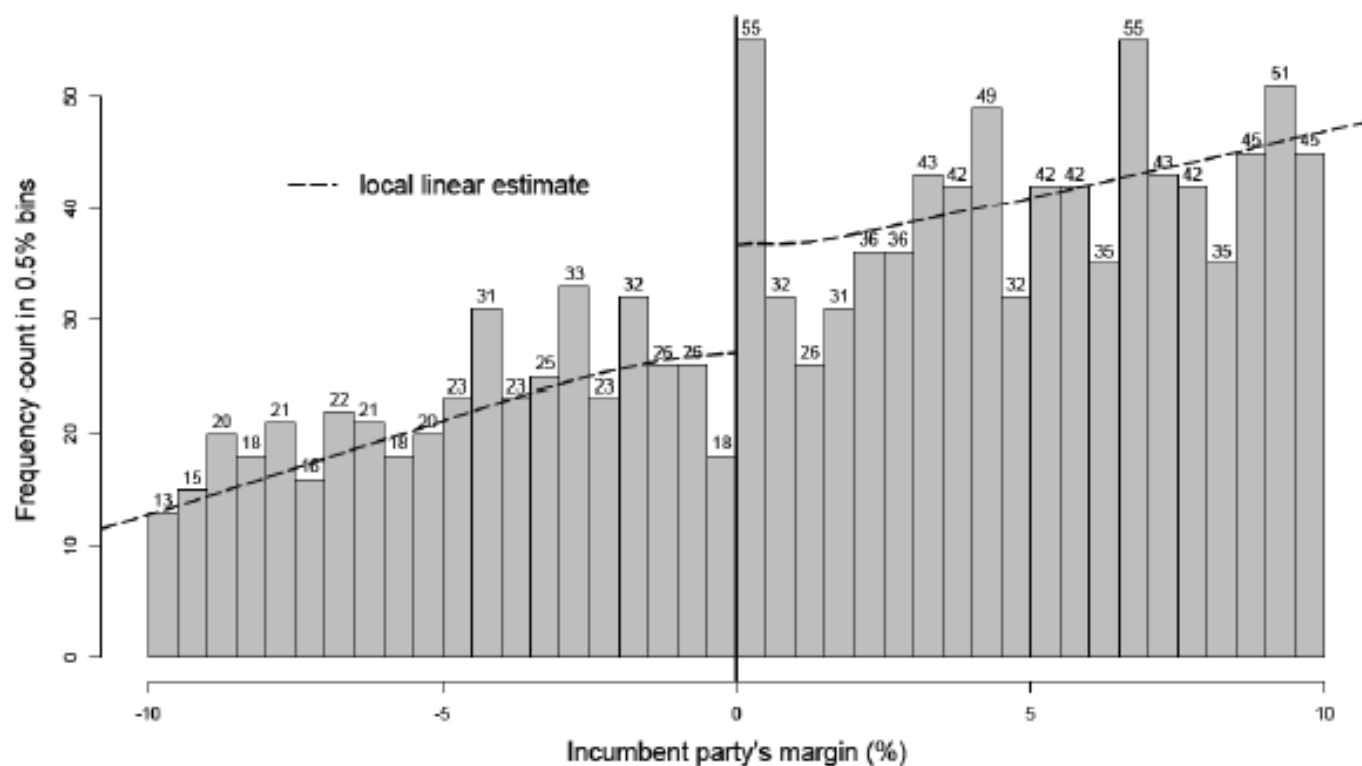


A graph like (A) led Lee, and separately McCrary, to conclude that there is no manipulation. However, (B) and (C) begin to suggest another story. Remember, the concern is with the incumbent party's vote share, not the Democratic vote share.



Density of the assignment variable

Figure 2: Histogram of the incumbent party's margin in U.S. House elections, 1946-2008. The local linear estimate is based on a triangular kernel with a bandwidth of 10.8, which is optimal according to the Imbens-Kalyanaraman algorithm. The estimated discontinuity at the cut-point is 9.5 ($SE = 3.7$).



Key Takeaway: The candidate of the incumbent party is about three times more likely to win election by half a percentage point or less than to lose by a similar margin. The density of this variable appears to diverge rather than converge in the neighborhood of the cut-point.



Covariate imbalance

Based on correcting some of Lee's data and adding some new variables, CS finds imbalance at the cutoff in the following:

- Democratic margin in the previous election
- The parties' relative campaign expenditures
- 1st dimension NOMINATE score of the current incumbent
- Whether the Democrat (Republican) candidate is the current incumbent
- Number of terms the Democrat (Republican) has served in the U.S. House of Representatives
- Whether the Democrat (Republican) has more political experience than the Republican (Democrat)
- Congressional Quarterly's October prediction of which candidate will win the race



Potential mechanisms

- Not likely to be outright fraud, because significance of lagged vote share is increasing over time and we believe potential for fraud has been decreasing.
- Control over recounts does not appear to be the key because they rarely happen and even more rarely change the outcome.
- But we don't need an explanation based on vote counting. Differences between winners and losers in incumbency, money, political experience, and other pre-election resources are evident far before any votes are cast, counted, or manipulated.
- These differences can be seen in elections expected to be close ex ante and in those that were in fact decided by a narrow margin.
- These facts contradict the idea that resources, expectations, and all else should be balanced in the closest elections.



Lessons from LMB & CS

- This is a cautionary tale
 - LMB are *very* good scholars.
 - They did almost everything right.
 - They do not do a lot to justify functional form or show robustness to different bandwidths.

Note: Michelle did her capstone research on a very similar concept, looking at how closeness of elections impacts voting behavior, using a difference-in-difference design.

Be in touch with her if this is a research area of interest!



Lessons from LMB & CS (cont'd)

- What can you learn from this exchange:
 - **Try to find problems in your design before someone else does it for you.** Address this throughout the paper, and in the limitations section. Discuss why yours was the best possible option!
 - Identify and collect accurate data on the observable covariates most likely to reveal sorting at the cut-point. This may not be the covariates that happen to be sitting in your data set.
 - Lagged values of the treatment variable are always a good idea. In elections, the party that currently controls the office.
 - **Automated bandwidth selection algorithms do not guarantee good results.** They are just a starting point.
 - For RD purposes, what constitutes a “close” election appears to be closer than the 48-52% bandwidth widely used up to now. CS get most of their results using 49.5-50.5%.



Guide to Practice



Steps for sharp RD analysis

- 1. Graph the data** by computing the average value of the outcome variable over a set of bins.
 - The bin width has to be large enough to have a sufficient amount of precision so that the plots looks smooth on either side of the cutoff value, but at the same time small enough to make the jump around the cutoff value clear.
- 2. Estimate the treatment effect by running linear regressions** on both sides of the cutoff point.
 - With a rectangular kernel, these are just standard regression estimated within a bin of width h on both sides of the cutoff point. Note that: Standard errors can be computed using standard least square methods (robust standard errors). The optimal bandwidth can be chosen using cross validation or other methods.



Steps for sharp RD analysis

3. Employ various specification tests to assess the robustness of the results.

- Looking at possible jumps in the value of other covariates at the cutoff point
- Testing for possible discontinuities in the conditional density of the forcing variable
- Looking whether the average outcome is discontinuous at other values of the forcing variable
- Using various values of the bandwidth, with and without other covariates that may be available.



Steps for fuzzy RD analysis

- 1. Graph the average outcomes** over a set of bins as in the case of SRD, but also graph the probability of treatment.
- 2. Estimate the treatment effect using 2SLS**, which is numerically equivalent to computing the ratio in the estimate of the jump (at the cutoff point) in the outcome variable over the jump in the treatment variable.
 - Standard errors can be computed using the usual (robust) 2SLS standard errors
 - The optimal bandwidth can again be chosen using one of the methods discussed above.
- 3. Assess the robustness of the results** using the various specification tests mentioned in the case of SRD designs.



Evaluating an RD Paper (possibly your own)

- Does the author show convincingly that
 - ✓ Treatment changes discontinuously at the cutpoint
 - ✓ Outcomes change discontinuously at the cutpoint
 - ✓ Other covariates do not change discontinuously at the cutpoint
 - ✓ Pre treatment outcomes do not change at the cutpoint
 - ✓ There is no manipulation of the assignment variable (bunching near the cutpoint)
- Are the basic results evident from a simple graph?
- Are the results robust to different functional form assumptions about the assignment variable?
 - For example, parametric and nonparametric fits, different bandwidths, etc.



Evaluating an RD Paper (possibly your own)

- Could other possibly unobserved treatments change discontinuously at the cutoff (bundling of institutions)?
 - For example, 18th birthday marks a discontinuous change in eligibility to vote, but also eligibility for draft, sentencing as an adult, and lots of other things, which may or may not be relevant depending on the outcome in question
- External validity
 - Are cases near the cutpoint different from cases far from the cutpoint in other ways? Do these differences make them more or less relevant from a theoretical or policy perspective?



Learn More about Evaluating Papers using Quasi-experimental Designs

- In these slides, we discussed a lot of “best practices” for evaluating papers employing RD. Many of the core principles, however, can and should be applied to reading and writing all research papers.
- Please book office hours with me or Michelle to discuss principles of critically producing and reading quality papers, especially if you are doing the quantitative research capstone.

Examples in STATA

