Canonical Research Designs VII: Regression Discontinuity II: The Checklist

Paul Goldsmith-Pinkham

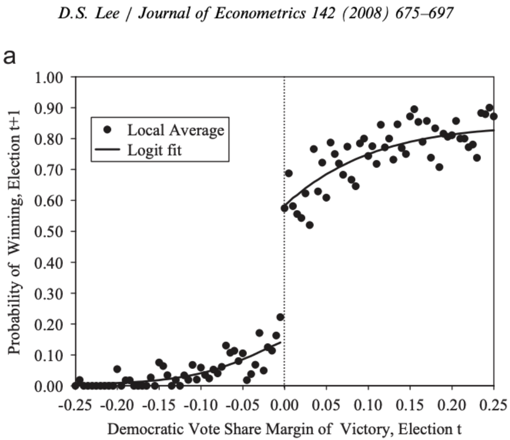
April 19, 2023

Roadmap for Today

- Last time: assumptions for RD, and estimation basics
- This time: how to implement RD, and checklist
    - E.g., if I'm writing a paper on RD, what would I need to show?

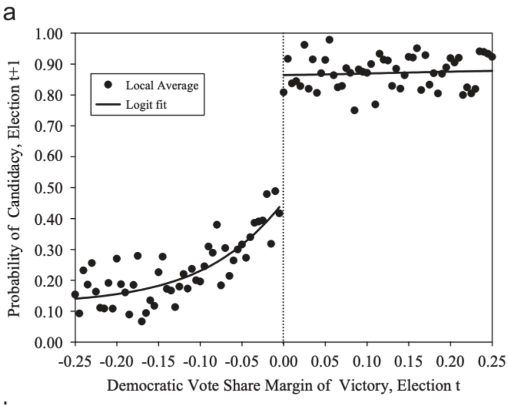
Running example

- Lee (2008) studies the impact of a Democrat winning on subsequent victory
- Running variable Z: vote share margin of victory
- D: winning election
- Y: Subsequent victory in an election



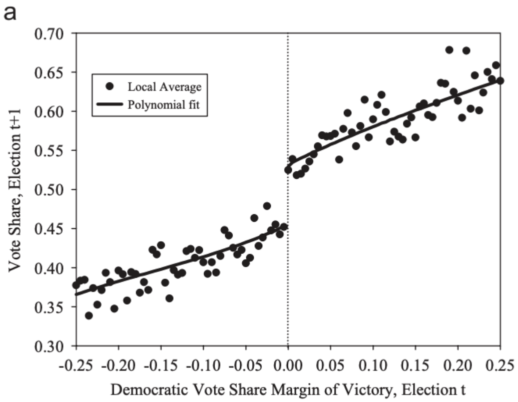
Running example

- Lee (2008) studies the impact of a Democrat winning on subsequent victory
- Running variable Z: vote share margin of victory
- D: winning election
- Y: Subsequent victory in an election
    - Y: Subsequent candidacy in an election



Running example

- Lee (2008) studies the impact of a Democrat winning on subsequent victory
- Running variable Z: vote share margin of victory
- D: winning election
- Y: Subsequent victory in an election
    - Y: Subsequent candidacy in an election
    - Y: Subsequent vote share in an election

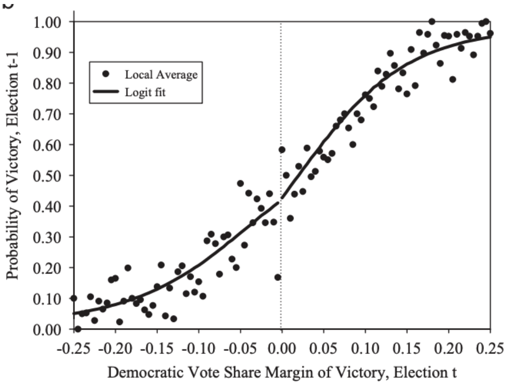


A checklist for how to support this analysis

- A graphical representation and test of "balance" and first stage (if fuzzy)
- Permutation test of characteristic at cutoff
- The density of the forcing variable (Mcrary test)
- Placebo checks
- A graphical representation of the outcomes (what we've already seen)
- Estimates based on optimal bandwidth choice and robust inference, using local linear analysis
  - These decisions vary depending on running variable. If discrete running variable, need to account for discreteness (Kolesar and Rothe (2018))
  - Should use local linear regression, and not global polynomials (Gelman and Imbens)
- Robustness analysis along bandwidth choice (and other tuning parameters)
  - Present this graphically

Checking for balance

- Our identification strategy, like in all settings, is not inherently testable
- But, there are things that we can look at whether they are consistent with our hypothesis
- In Lee (2007), the most natural test is whether the cutoff in period t affects the probability of victory in period $t-1$, the period before
- Other natural tests exist as well: looking for balance in outcomes that should not be affected by the treatment:
    - predermined covariates
    - things with no causal link

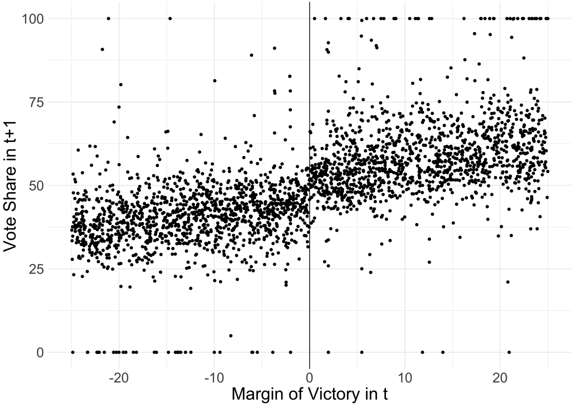


A quick aside on graphical construction

- One of the most powerful aspects of regression discontinuity is the ability to present the results graphically. So what's the right approach?

- First, worth noting that the raw data is rarely informative without some amount of grouping
  - Consider the main Lee (2008) result, with just raw data

- Remarkably, you fact see a jump in the distribution in the data
  - But the signal to noise ratio is low



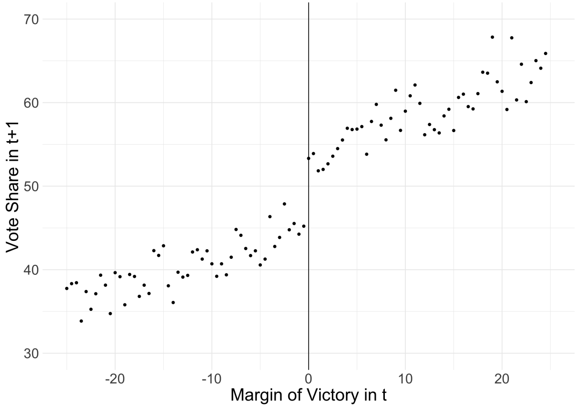
A quick aside on graphical construction

- Ideally, you would plot a version of the scatter plot, but focusing on means within binned areas
  - This is exactly the intuition from binscatter, and a similar statistical problem
  - How do we choose bins?
- Simple first approach - equally spaced bins
  - But how big?
  - Lee (2008) chooses 0.5 percent bins
  - But why does this look less compelling?

A quick aside on graphical construction

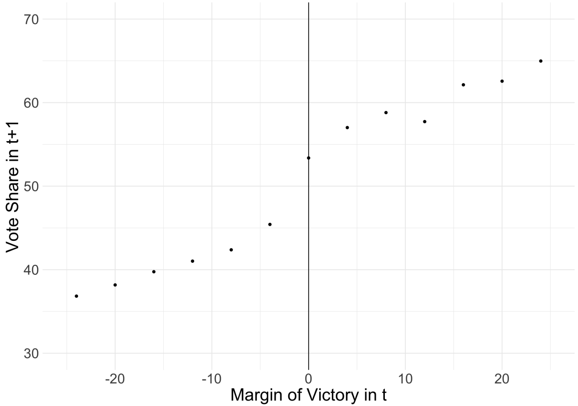
- Ideally, you would plot a version of the scatter plot, but focusing on means within binned areas
  - This is exactly the intuition from binscatter, and a similar statistical problem
  - How do we choose bins?

- Simple first approach - equally spaced bins
  - But how big?
  - Lee (2008) chooses 0.5 percent bins
  - But why does this look less compelling?



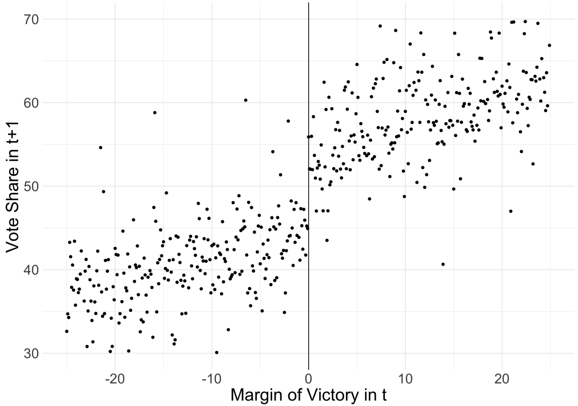
A quick aside on graphical construction

- The choice is bin is non-trivial
- What does this look like with bins of 4 percent? 0.1 percent?



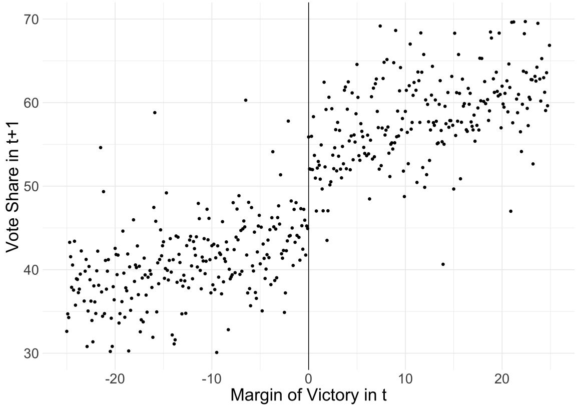
A quick aside on graphical construction

- The choice is bin is non-trivial
- What does this look like with bins of 4 percent? 0.1 percent?



A quick aside on graphical construction

- The choice is bin is non-trivial
- What does this look like with bins of 4 percent? 0.1 percent?
- As with our discussion of estimating non-parametric means, the trade-off in number of bins typically comes down to bias (more bins helps get closer to the "true" conditional means) and noise (less bins increases observations within bins, lowering the SE for a bin)

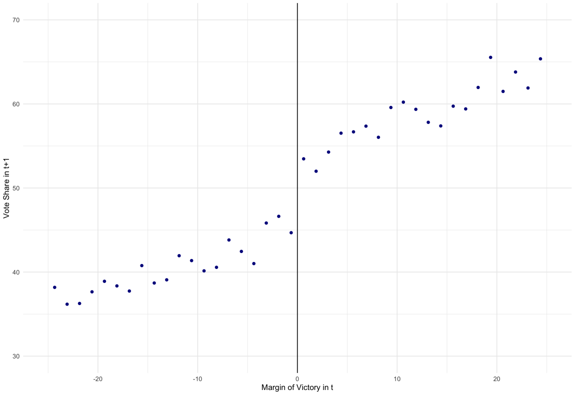


A quick aside on graphical construction

- Given how graphically important bin choice is, how should we choose it?

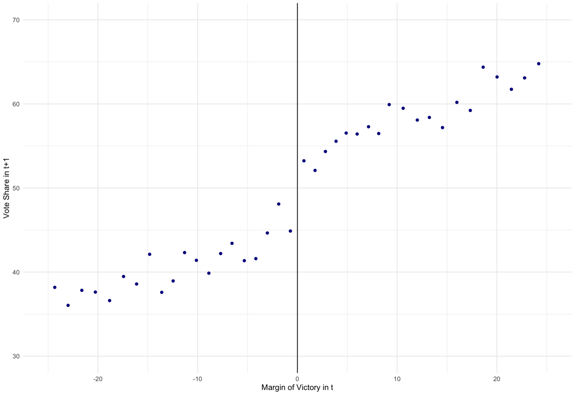
- Turns out there are two important decisions:
    - How to place the bins: equal-spaced, or quantile
    - How many bins?

- The equal-spaced vs. quantile choice is somewhat arbitrary, but quantile binning is more transparent
    - Choice of equally spaced bins can mask underlying density (not so much in this case)



A quick aside on graphical construction

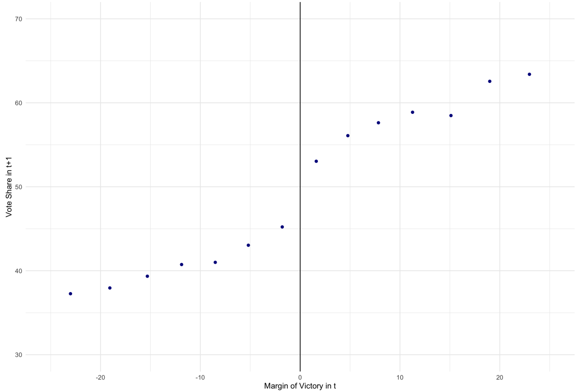
- Given how graphically important bin choice is, how should we choose it?
- Turns out there are two important decisions:
  - How to place the bins: equal-spaced, or quantile
  - How many bins?
- The equal-spaced vs. quantile choice is somewhat arbitrary, but quantile binning is more transparent
  - Choice of equally spaced bins can mask underlying density (not so much in this case)



A quick aside on graphical construction

- Once we choose how to do bins, how should we choose the number?
    - Can we choose "optimally"?

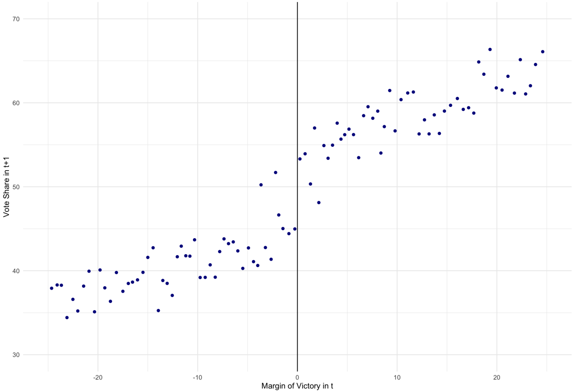
- Cattaneo et al. (2020) argue for two approaches (available in `rdrobust`'s `rdplot`): IMSE-minimizing, and mimicking variance.
    - IMSE-minimizing trades off between bias and variance in choice of bins, but does it over the whole range – proportional to $n^{1/3}$
    - Mimicking-variance tries to match the underlying variance of the raw data in the binned plots – proportional to $n/\log(n)^2$



A quick aside on graphical construction

- Once we choose how to do bins, how should we choose the number?
  - Can we choose “optimally”?

- Cattaneo et al. (2020) argue for two approaches (available in `rdrobust`’s `rdplot`): IMSE-minimizing, and mimicking variance.
  - IMSE-minimizing trades off between bias and variance in choice of bins, but does it over the whole range – proportional to $n^{1/3}$
  - Mimicking-variance tries to match the underlying variance of the raw data in the binned plots – proportional to $n/\log(n)^2$



A quick aside on graphical construction

- Obviously in the case of discrete random variables, this is not complicated! We would just bin directly on the discrete values
- The complicating issues will arise when imputing a smooth function on top of these discrete values
- See Kolesar and Rothe (2018) for details

Checking for balance

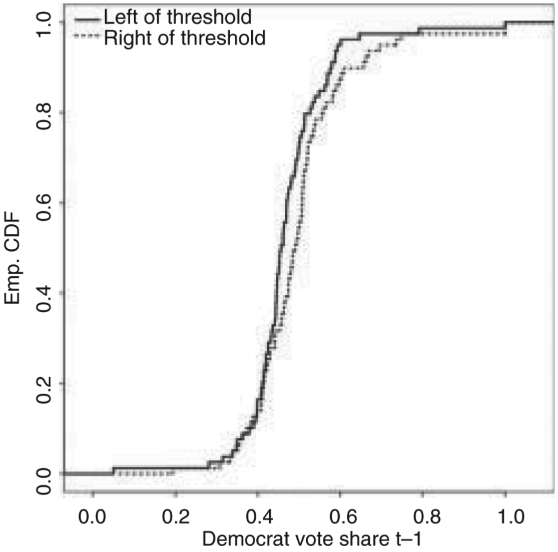
- As we discussed above, key test is to compare the averages of other variables at $Z = 0$, the cutoff.

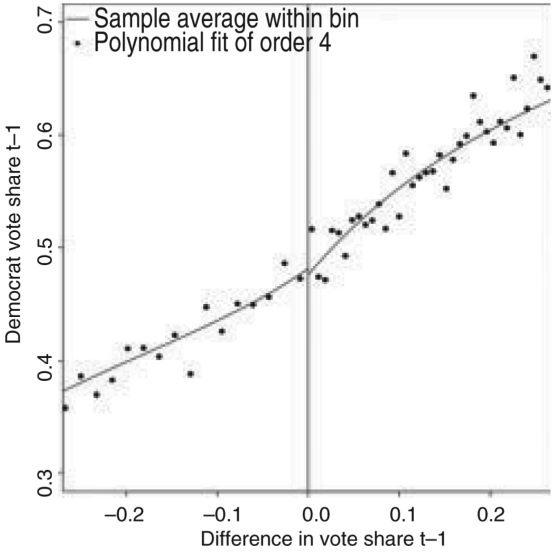
- Canay and Kamat (2017) show that if you are willing to assume a slightly stronger assumption – e.g. that choice of location around the cutoff is not fully deterministic – then you can do better

- Key points:
  - Testing just mean differences doesn't look at other parts of the distribution (which may more obviously violate this) and so may have low power
  - B/c the local sample size is effectively small, this can create problematic inference issues if the function is not particular smooth

- They propose a permutation test, which has better statistical properties
  - Key intuition – covariates should be *approximately* identically distributed on each side of the cutoff
  - This is an *asymptotic* argument, since it's not actually a random experiment!

Checking for balance- Canay and Kamat





Checking for balance- Canay and Kamat

- This approach requires a slightly stronger assumption than the necessary assumptions for identification
- E.g., this paper riffs off of Lee (2008) assumption that units are effectively permuted around the cutoff (somewhat randomly), such that the covariate distribution should be continuous at the cutoff
- Code is available in Stata and R: `rdperm` and `RATest`

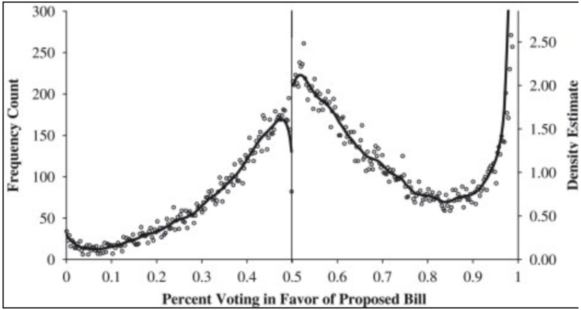
Testing for bunching in forcing variable

- Similar to the balance test, Mcrary (2008) proposed a test of the continuitiy in the density of the running variable

- In essence, is there “bunching” in the characteristic on one side or the other?
    - This intuition makes sense economically – if there’s a benefit of being on one side, why would you not “shift” yourself across the boundary?

- This is easily tested by comparing the values of an estimated density on the left and right of the cutoff

- Software is also available for this! `rddensity` in Stata and R

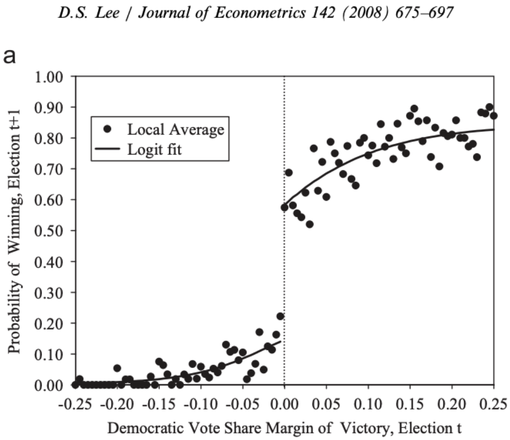


Testing for bunching in forcing variable

- Placebo checks are less formalized (or at least I know them less well)
- Ganong and Jager (2017) propose permutation tests for randomizing the location of the cutoff
  - This method presumably works as well in the RD setting
- Intuitively, one would pick cutoffs above and below the true cutoff, and test for jumps in the outcome. If these are insignificant, that gives credibility to the design
  - More formally, using a permutation test, one could permute the cutoff and look at the relative effect of the true effect compared to the null distribution
  - Effectively treats the choice of cutoff as the random design variable

Showing our outcome

- Finally, we plot our outcome.
- This involves both a plotting of the binned data, as well as our choice of the conditional mean function
- Notably, while we plot a large window around the cutoff, the window of plotted points is irrelevant for estimation
  - The choice of bandwidth will be smaller than the window
- This is really for the "eyeball test" (Korting et al. (2020))
  - A good graph is worth a lot! If you have a bad graph... maybe you have a bad experiment?

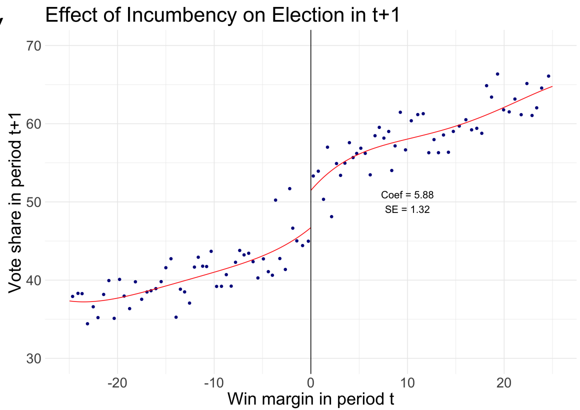


Estimating our outcomes

- The actual estimation, as you've seen, is subject to many tuning parameters:
  - Choice of estimation procedure, bandwidth, kernel
- Much of this is more automated now, but there is still discretion
- My suggestion: use defaults unless you have a very good reason not to
  - Defaults: local linear regression, optimized bandwidth from estimation procedures in packages (e.g. Cattaneo et al.'s `rdrobust` or Kolesar and Rothe's `RDRobust`), uniform kernel
  - Even in these categories, there is discretion, but important to be consistent and transparent
    - See Armstrong and Kolesar (2018) on bandwidth snooping

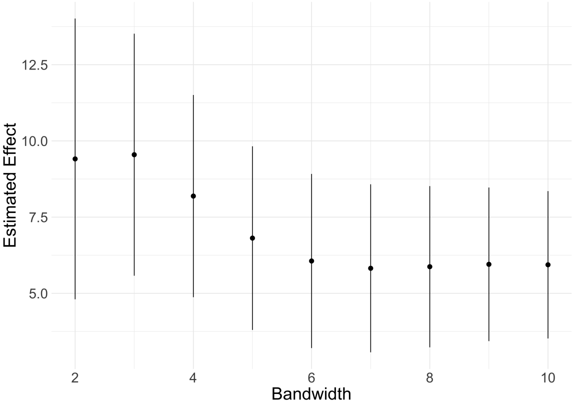
Estimating our outcomes

- Also, you can put your estimates directly in the graph! Why waste time with tables



Showing robustness

- Showing robustness is sometimes the only way to convince a reader or audience member
  - E.g. "I did this the optimal way!" is not a good excuse
- A simple and clean ways to present your robustness result
  - Consider many bandwidths / permutations, and plot how sensitive the estimates are



Next time

- Discrete RD
- Bias from RD estimation
- Regression Kink
- Bounds on treatment effects