Part 8: Policy Evaluation (e): Regression Discontinuity

Chris Conlon

April 9, 2021

Applied Econometrics

Regression Discontinuity Design

- Another popular research design is the Regression Discontinuity Design.
- In some sense this is a special case of IV regression. (RDD estimates a LATE).
- Most of this is taken from the JEL Paper by Lee and Lemieux (2010).

RDD: Basics

- We have a running or forcing variable x such that

$$\lim_{x \to c^+} P(T_i | X_i = x) \neq \lim_{x \to c^-} P(T_i | X_i = x)$$

- The idea is that there is a discontinuous jump in the probability of being treated.
- For now we focus on the sharp discontinuity:
  $P(T_i | X_i \geq c) = 1$ and $P(T_i | X_i < c) = 0$
- There is no single x for which we observe treatment and control.
  (Compare to Propensity Score!).

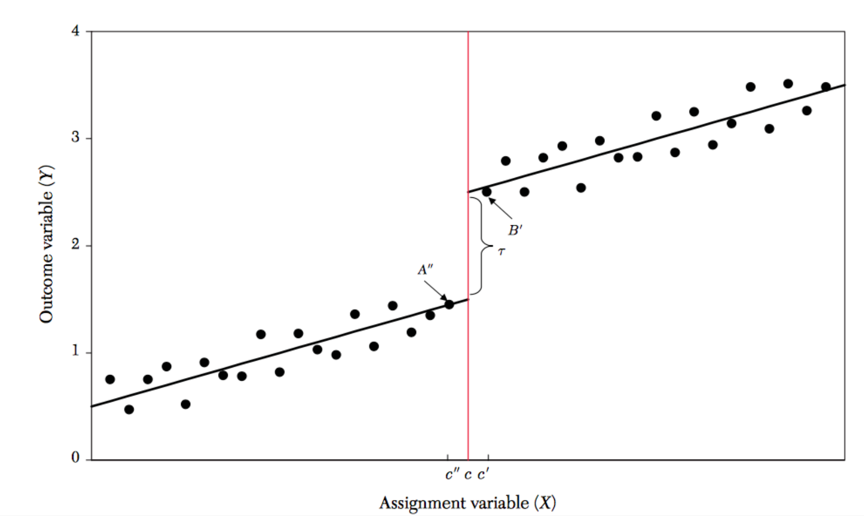
RDD: Basics

- Example: a social program is available to people who earned less than $25,000.
    - If we could compare people earning $24,999 to people earning $25,001 we would have as-if random assignment. (MAYBE)
    - But we might not have that many people...
- We are going to label the treatment effect τ_i.
  Note: my lack of precision here!
- The most important assumption is that of no manipulability $\tau_i \perp T_i$ in some neighborhood of c.
    - If agents can choose x_i we are in trouble: underreporting income, avoiding "possession with intent to distribute" for drugs, etc.

RDD: Continuity

- The central idea in RDD is that of continuity
- We need that $E[Y(1)|X]$ and $E[Y(0)|X]$ both be continuous at $X = c$.
  - We expect that $Y_i = f(x_i)$ to be a smooth, continuous function of x_i
  - The only departure from that is the treatment $\tau_i \cdot I(x_i \geq c)$.
- We want to be as agnostic as possible about functional form
  - Don't want to restrict ourselves to $f(x_i) = \beta_0 + \beta_1 x_i$.
  - The central idea: we know $f(x_i)$ absent the treatment!

RDD: In Pictures



RDD: Sharp RD Case

RDD uses a set of assumptions distinct from our LATE/IV assumptions. Instead it depends on continuity.

- People just to the left of c are a valid control for those just to the right of c.
- This is not a testable assumption $\rightarrow$ draw pictures!
- We could run the regression where $T_i = \mathbf{1}[X_i > c]$.

$$Y_i = \beta_0 + \tau_i \cdot T_i + X_i\beta + \epsilon_i$$

- This puts a lot of restrictions (linearity) on the relationship between Y and X.
- Also (without additional assumptions) we only learn about τ_i at the point $X = c$.

RDD: Nonlinearity

First thing to relax is assumption of linearity.

$$Y_i = f(x_i) + \tau T_i + \epsilon_i$$

This is known as partially linear model.

- Three options for $f(x_i)$:
  1. Kernels
  2. Polynomials: $Y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \cdots + \beta_p x^p + \tau T_i + \epsilon_i$.
     - Actually, people suggest different polynomials on each side of cutoff! (Interact everything with T_i).
  3. Local Linear/Polynomial Regression
- Same objective. Want to flexibly capture what happens on both sides of cutoff.
- Otherwise risk confusing nonlinearity with discontinuity!

RDD: Kernel Boundary Problem

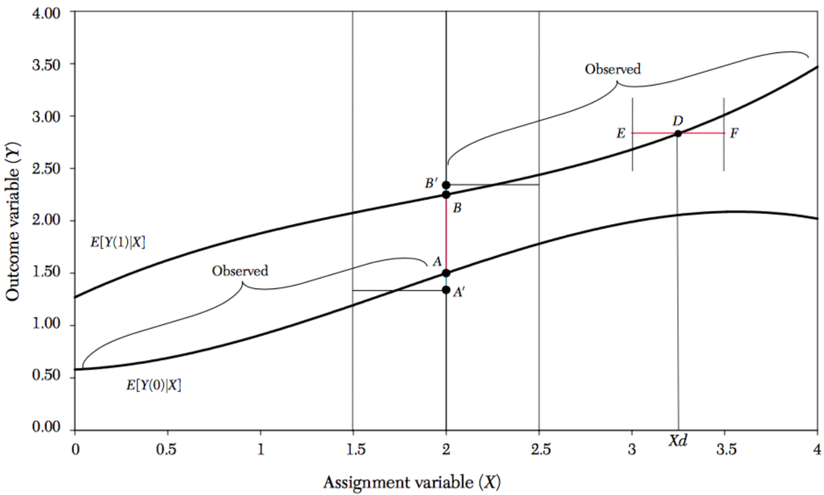


Figure 2. Nonlinear RD

RDD: Polynomial Implementation Details

To make life easier:

- replace $\tilde{x}_i = x_i - c$.
- Estimate coefficients β: $(1, \tilde{x}, \tilde{x}^2, \ldots, \tilde{x}^p)$ and $\tilde{\beta}$: $(T_i, T_i\tilde{x}, T_i\tilde{x}^2, \ldots, T_i\tilde{x}^p)$.
- Now treatment effect at c just the coefficient on T_i. (We can ignore the interaction terms).
- If we want treatment effect at $x_i > c$ then we have to account for interactions.
  - Identification away from c is somewhat dubious anyway.
- Lee and Lemieux (2010) suggest estimating a coefficient on a dummy for each bin in the polynomial regression $\sum_k \phi_k B_k$.
  - Add polynomials until you can satisfy the test that the joint hypothesis test that $\phi_1 = \cdots \phi_k = 0$.
  - There are better ways to choose polynomial order...

RDD: Checklist

Most RDD papers follow the same formula (so should yours)

- Plot of $P(T|X)$ so that we can see the discontinuity
- Plot of $E[Y|X]$ so that we see discontinuity there also
- Plot of $E[W|X]$ so that we don't see a discontinuity in controls.
- Density of X (check for manipulation).
- Show robustness to different "windows"
- The OLS RDD estimates
- The Local Linear RDD estimates
- The polynomial (from each side) RDD estimates
- An f-test of "bins" showing that the polynomial is flexible enough.

Read Lee and Lemieux (2010) before you get started.

The `rdd` package in R has many of these features.

Application: Lee (2008)

Looked at incumbency advantage in the US House of Representatives

- Running variable was vote share in previous election
  - Problem of naive approach: good candidates get lots of votes!
  - Compare outcomes of districts with barely D to barely R.
- First we plot bin-scatter plots and quartic (from each side) polynomials.
- Discussion about how to choose bin-scatter bandwidth (CV).

Lee (2008)

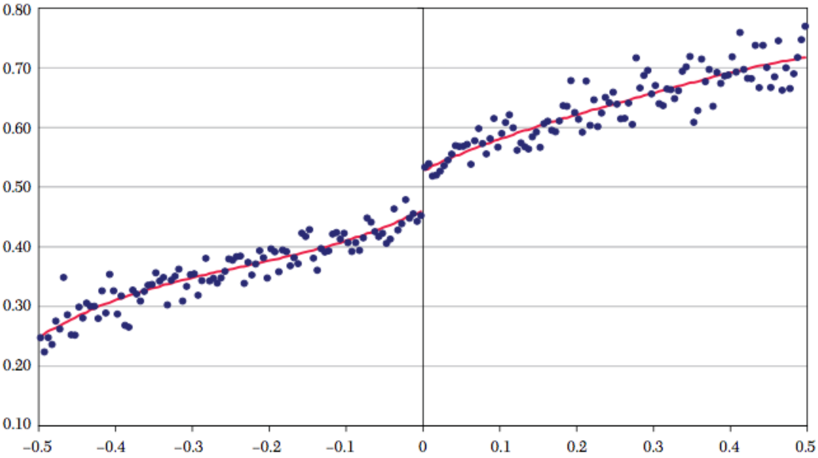


Figure 8. Share of Vote in Next Election, Bandwidth of 0.005 (200 bins)

Lee (2008)

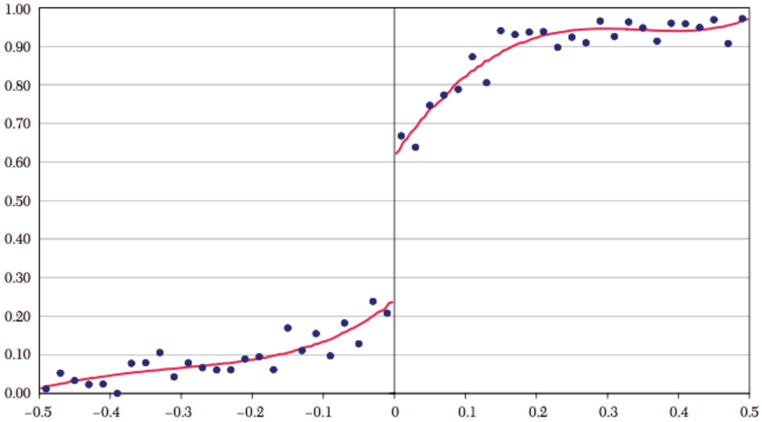


Figure 9. Winning the Next Election, Bandwidth of 0.02 (50 bins)

Lee (2008)

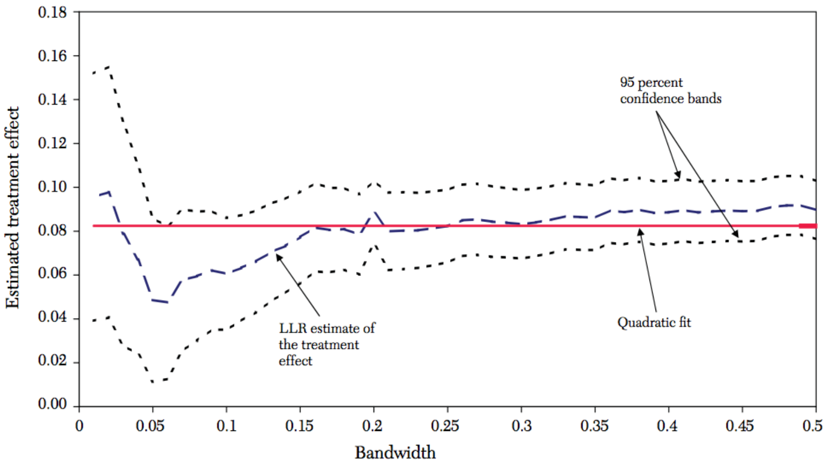


Figure 18. Local Linear Regression with Varying Bandwidth: Share of Vote at Next Election

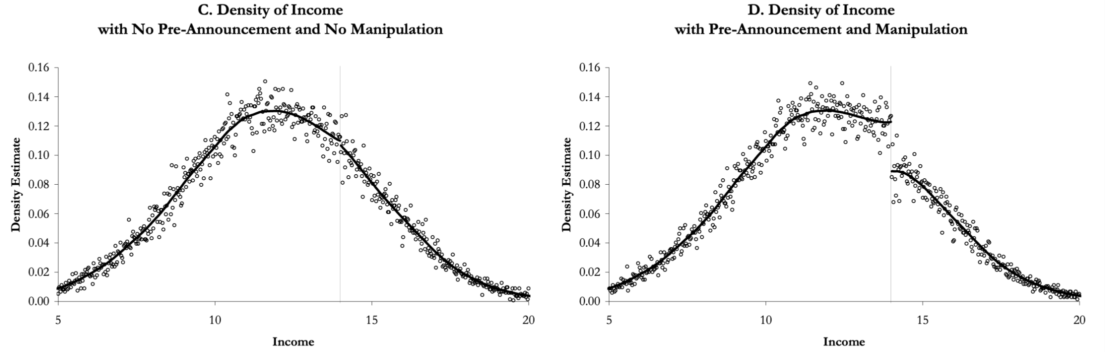
Other Examples

Luca on Yelp

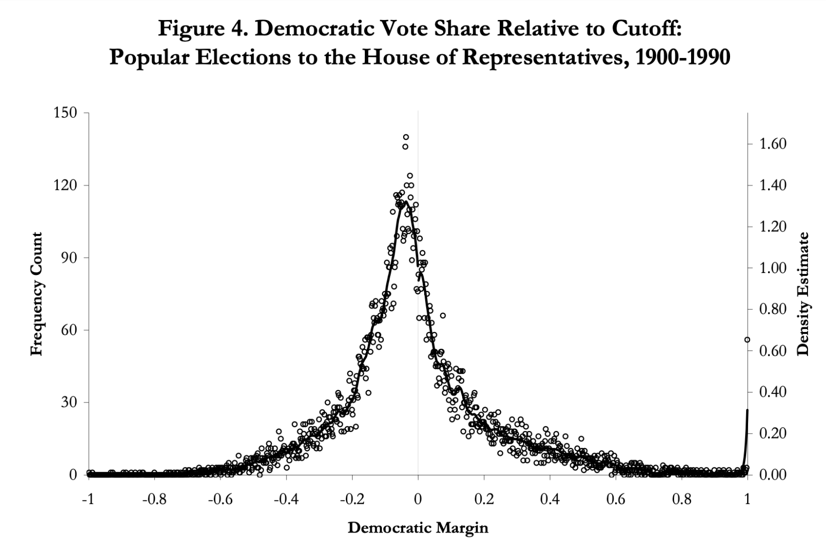
- Have data on restaurant revenues and yelp ratings.
- Yelp produces a yelp score (weighted average rating) to two decimals ie: 4.32.
- Score gets rounded to nearest half star
- Compare 4.24 to 4.26 to see the impact of an extra half star.
- Now there are multiple discontinuities: Pool them? Estimate multiple effects?

Manipulation

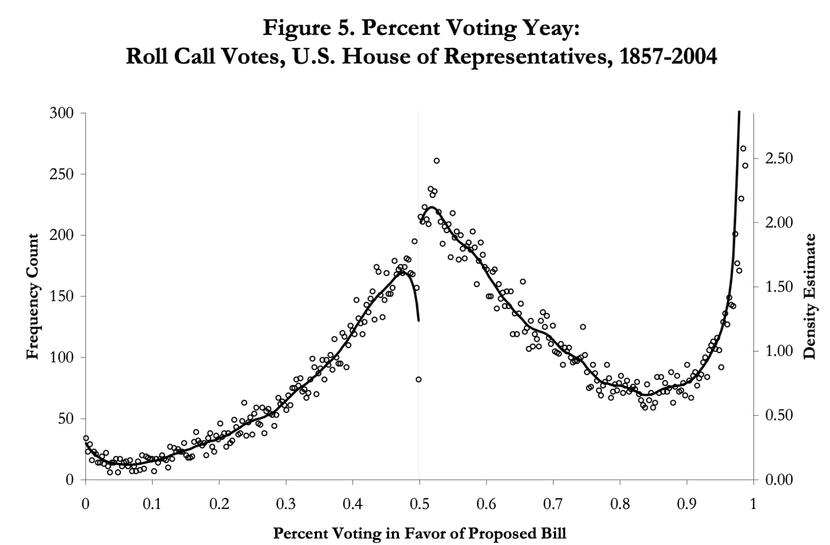
McCrary (2008) develops a test for manipulation in running variable x



Manipulation: Good



Manipulation: Bad



Extensions

Fuzzy RD

An important extension in the Fuzzy RD. Back to where we started:

$$\lim_{x \to c^+} P(T_i | X_i = x) \neq \lim_{x \to c^-} P(T_i | X_i = x)$$

- We need a discontinuous jump in probability of treatment, but it doesn't need to be $0 \to 1$.

$$\tau_i(c) = \frac{\lim_{x \to c^+} P(Y_i | X_i = x) - \lim_{x \to c^-} P(Y_i | X_i = x)}{\lim_{x \to c^+} P(T_i | X_i = x) - \lim_{x \to c^-} P(T_i | X_i = x)} \quad \text{Wald Estimator}$$

- Under sharp RD everyone was a complier, now we have some always takers and some never takers too.
- Now we are estimating the treatment effect only for the population of compliers at $x = c$.
- This should start to look familiar. We are going to do IV!

Related Idea: Kinks

A related idea is that of kinks.

- Instead of a discontinuous jump in the outcome there is a discontinuous jump in β_i on x_i.
- Often things like tax schedules or government benefits have a kinked pattern.

What about binscatter?

Binscatter: Easy Version

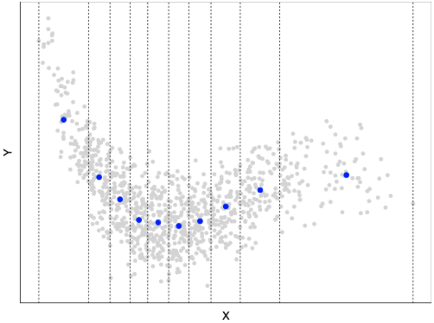
All of these RDD plots used something called binscatter

- For the running variable x_i break it up into $J = 20, 50, 100$ bins of equal density
- Use the quantiles of x_i.
  - If we have 10 bins, each contains 10% of the sample
  - Like a histogram but with bins of different width.
- Compute the average value: $b_x^j = E[x_i | x_i \in \text{bin}_j]$
- Compute the average value: $b_y^j = E[y_i | x_i \in \text{bin}_j]$
- Plot (b_y^j, b_x^j) for each bin j.
- Often draw a line of best fit through the points

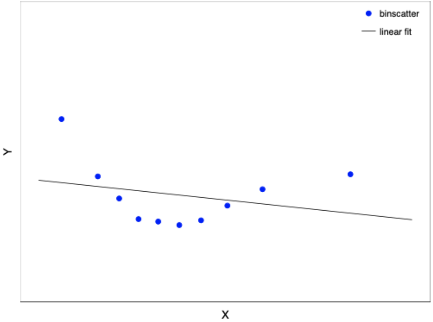
But often we have other covariates!

Binscatter: Under the Hood

Figure 1: The basic construction of a binned scatter plot.



(a) Scatter and Binscatter Plots

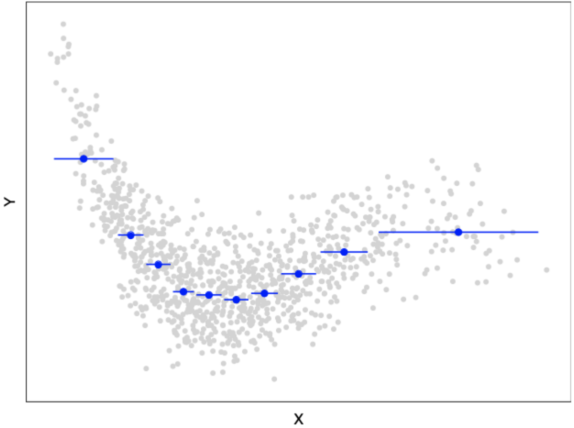


(b) Binscatter and Linear Fit

Notes. The data is divided into $J = 10$ bins according to the observed x. Within each bin a single dot is plotted at the mean of y for observations falling in the bin. The final plot (b) consists of only these J dots, and the fit from a least squares linear regression of y on x. Constructed using simulated data described in Section SA-5 of the supplemental appendix.

Binscatter: Under the Hood

Figure 3: The actual regressogram nonparametric estimator corresponding to a binned scatter plot.



(a) Binned Scatter Plot with Piecewise Constant Fit

Binscatter as Semiparametric Regression:
Cattaneo, Crump, Farrell, Feng (2019)

The right way to think of binscatter is as semiparametric or partially linear regression

$$y_i = \mu\left(x_i\right) + \mathbf{w}_i'\gamma + \epsilon_i, \quad \mathbb{E}\left[\epsilon_i | x_i, \mathbf{w}_i\right] = 0$$

- Unless $u(x_i)$ is linear, we can't partial out $y_i - \mathbf{w}_i'\gamma$ via Frisch-Lovell-Waugh
- Most binscatter software does this wrong. Use `binsreg` in R.

Thanks