Regression discontinuity

Michal Kolesár

ECO539B, Spring 2025

April 14, 2025

Setting

- Observations $i = 1, \ldots, n$. Interested in effect of some treatment D_i on outcome Y_i.
- Key feature of regression discontinuity (RD) is that treatment is fully or partially determined by whether running variable X_i crosses a threshold.
- Normalizing threshold to zero:

$$D_i = \mathbb{1}\{X_i \geq 0\}, \qquad \text{(sharp RD)}$$

$$\lim_{x \downarrow 0} P(D_i = 1 \mid X_i = x) - \lim_{x \uparrow 0} P(D_i = 1 \mid X_i = x) > 0. \qquad \text{(fuzzy RD)}$$

- Running examples: Lee (2008), Haggag and Paci (2014), Klaauw (2002), and Bleemer and Mehta (2022).

Identification

Falsification

Estimation and Inference in sharp RD

Empirical illustration

Extensions

Sharp design: key assumption

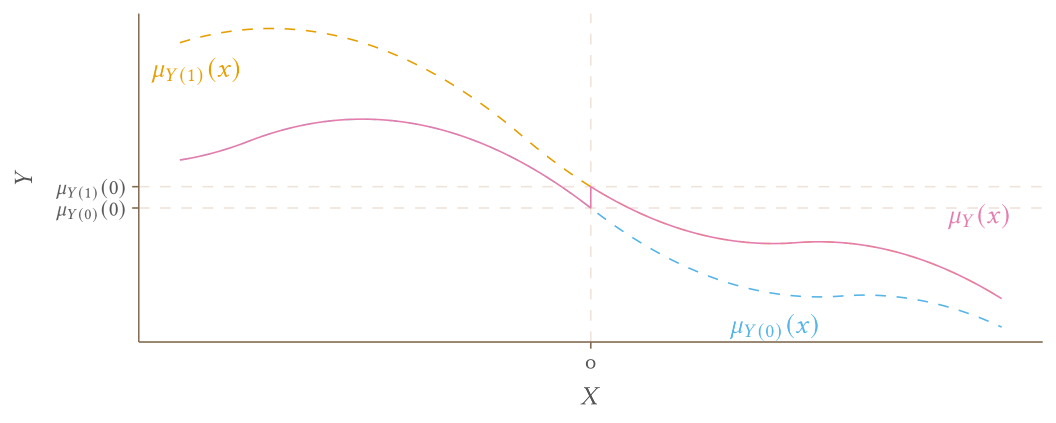
- Parameter of interest is jump in observed conditional mean function $\mu_Y(x) := E[Y_i \mid X_i = x]$ at cutoff:

$$\tau_Y = \lim_{x \downarrow 0} \mu_Y(x) - \lim_{x \uparrow 0} \mu_Y(x).$$

- Key assumption is continuity: $\mu_{Y(0)}(x) := E[Y_i(0) \mid X_i = x]$ and $\mu_{Y(1)}(x) := E[Y_i(1) \mid X_i = x]$ are both continuous in x at 0.
  - Rules out perfect manipulation of X_i (imperfect manipulation OK in theory). Taxi drivers in Haggag and Paci (2014) cannot keep driving until the fare is over $15. But it is fine if Mark Harris hires McCready to tamper with votes in Lee (2008), since McCready can't do it in a way that ensures Harris' victory.
  - Nothing else happens at the cutoff except for a change in treatment status. Strong assumption in geography-based or spatial RDs. Age-based cutoffs also need to be treated with care. In Battistin et al. (2009), things other than retirement may happen at retirement age cutoff.

Under continuity, τ_Y identifies average treatment effect (ATE) at the cutoff:

$$\tau_Y = \lim_{x \downarrow 0} E[Y_i(1) \mid X_i = x] - \lim_{x \uparrow 0} E[Y_i(0) \mid X_i = x] = E[Y_i(1) - Y_i(0) \mid X_i = 0].$$



Fuzzy design: key assumptions

- If jump in treatment probability at cutoff is smaller than 1, scale jump τ_Y by the size of the jump in treatment probability:

$$\theta = \frac{\lim_{x\downarrow 0}\mu_Y(x) - \lim_{x\uparrow 0}\mu_Y(x)}{\lim_{x\downarrow 0}\mu_D(x) - \lim_{x\uparrow 0}\mu_D(x)} = \frac{\tau_Y}{\tau_D}$$

- To interpret this, let $D_i(1), D_i(0)$ denote potential treatment if we make individual (in)eligible, by say, making exception to GPA cutoff requirement, or by changing the cutoff. Then $D_i = D_i(\mathbb{1}\{X_i \geq 0\})$
- Need two assumptions:

monotonicity $P(D_i(1) \geq D_i(0) \mid X_i) = 1$. Like in Imbens and Angrist (1994).

continuity $\mu_{Y(d)}(x)$, $\mu_{D(d)}(x)$, and $\mu_{D(d)Y(d')}(x)$ are continuous at $x = 0$ for $d, d' \in \{0, 1\}$. Again, allows for imperfect manipulation, like re-taking intro econ courses to improve GPA in Bleemer and Mehta (2022).

Under monotonicity and continuity, θ identifies local average treatment effect (LATE) at the cutoff

$$\theta = E[Y_i(1) - Y_i(0) \mid X_i = 0, D_i(1) > D_i(0)],$$

- Fuzzy RD is a local instrumental variables (IV) model: eligibility $\mathbb{1}\{X_i \geq 0\}$ is an instrument for D_i. Implicit in continuity assumption is exclusion restriction that eligibility itself doesn't affect potential outcomes.

Alternative frameworks

- Some papers propose a local randomization framework to formalize idea that sharp RD is like a localized random experiment, and fuzzy RD is like a localized experiment with imperfect compliance. But this would require $\mu_{Y(1)}$ and $\mu_{Y(0)}$ to be flat close to cutoff, which is not typically true. Also not clear how to pick right neighborhood.
- Ganong and Jäger (2018) propose randomization approach, thinking of cutoff as random. Allows for simple randomization inference, but would need to specify counterfactual cutoff distribution.
- See notes for more discussion

Identification

Falsification

Estimation and Inference in sharp RD

Empirical illustration

Extensions

Manipulation of running variable

Continuity assumption "questionable" if density of running variable not smooth around cutoff.

Formal tests described in the notes, in practice graphical evidence often convincing.

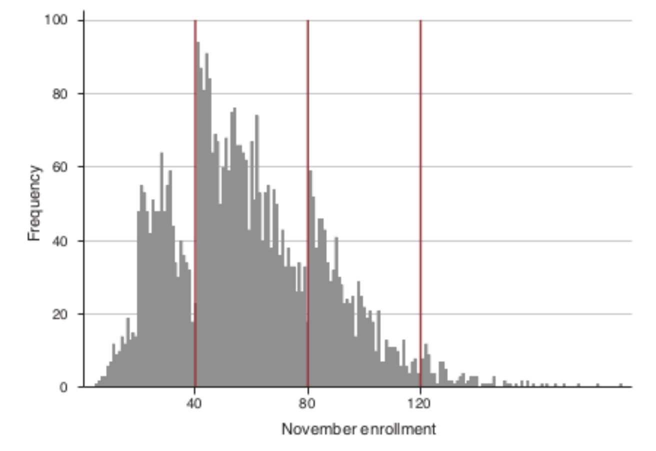


Figure 1 from Angrist et al. (2019). Fifth-grade enrollment distribution, as reported by school headmasters in November. Red reference lines indicate Maimonides' rule cutoffs at which an additional class is added.

Covariate balance checks

- Idea similar to "placebo" tests in other contexts: treatment should have no effect on pre-determined covariates
- Run RD, but with pre-determined covariate W_i as outcome. Can actually test the whole distribution, not just the mean, by comparing distribution of W_i just below and just above cutoff: Canay and Kamat (2018) propose a permutation test based on q closest observations to the cutoff.
    - Order the covariates W_i according to the running variable, obtaining $S = (W^-_{(q)}, \dots, W^-_{(1)}, W^+_{(1)}, \dots, W^+_{(q)})$.
    - Compare their empirical cumulative distribution functions (CDFs) $\hat{F}^+(w) = \frac{1}{q}\sum_j \mathbb{1}\{W^+_{(q)} \leq w\}$ using the Cramér-von Mises test statistic

$$T(S) = \frac{1}{2q}\sum_{j=1}^{2q}\left[\hat{F}^-(S_j) - \hat{F}^+(S_j)\right]^2,$$

    - Compute the critical value using a permutation test by permuting the elements of S.

Identification

Falsification

Estimation and Inference in sharp RD

Empirical illustration

Extensions

- Just need to estimate conditional mean at 0 separately for the treated and untreated subpopulations
- Key issue is that 0 is a boundary point in both regression problems, so extrapolation unavoidable
    - Parametric methods, such as specifying that $\mu_{Y(1)}$ and $\mu_{Y(0)}$ are exactly polynomial of order p, or using global nonparametric methods unattractive: observations far away from cutoff receive large weight
- Most estimators, including polynomial estimators can be written

$$\hat{\tau}_Y = \sum_i w(X_i) Y_i, \tag{1}$$

with $\sum_{i:\, X_i \geq 0} w(X_i) = -\sum_{i:\, X_i < 0} w(X_i) = 1$.

- average magnitude $|w(\cdot)|$ tends to increase with the order of polynomial p: if p large, some observations very influential.
- small misspecification can translate into large bias (Gelman and Imbens 2019)
- better to use local polynomial regression with $p = 1$ or 2.

- Local methods (e.g. local linear or local quadratic regression) only place weight on obs near cutoff
- Key distinction between "parametric" and "nonparametric" thinking: In "parametric" models, we don't worry about extrapolation bias. In "nonparametric" models, we both
    1. take into account the potential extrapolation bias when choosing between different estimators; don't just minimize variance
    2. should try to account for potential bias when conducting inference.
- How to operationalize this?

Standard approach

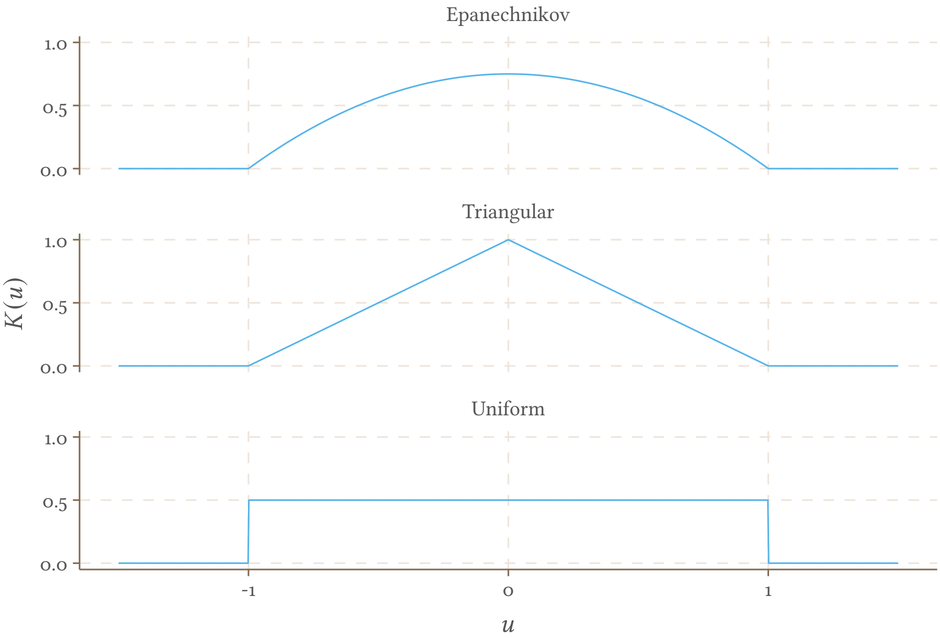
- Pick bandwidth h and polynomial order p. Keep only obs with distance h of cutoff
- Regress Y_i on powers of X_i above and below cutoff, difference in estimates is an estimate of τ_Y
- Can further downweight obs relatively further away from cutoff using kernel weights $K(x_i/h)$. Same as difference between weighted and ordinary least squares (OLS):

$$\hat{\mu}_{Y(1)}(0) = e_1' \left(\sum_i \mathbb{1}\{X_i \geq 0\} K(X_i/h) m(X_i) m(X_i)' \right)^{-1} \sum_i \mathbb{1}\{X_i \geq 0\} K(X_i/h) m(X_i) Y_i$$

$m(X) = (1, x, \dots, x^p)$, and $e_1 = (1, 0, \dots, 0)'$. Then

$$\hat{\tau}_{Y,h} = \hat{\mu}_{Y(1)}(0) - \hat{\mu}_{Y(0)}(0).$$

Can compute in one step by regressing Y_i onto D_i interacted with $m(X_i)$, with weights $K(X_i/h)$.


Epanechnikov
Triangular
Uniform
1.0
0.5
0.0
$K(u)$
-1
0
1
u

- How to pick polynomial order p? Formally depends on the amount of smoothness we assume. $p = 1$ optimal if we assume μ_Y twice differentiable with bounded second derivatives (Hölder class of order 2)
- How to pick K? Doesn't matter much, can use uniform for simplicity. Triangular and Epanechnikov slightly more efficient (Cheng, Fan, and Marron 1997; Armstrong and Kolesár 2020).
- How to pick h? More tricky and most consequential

Standard bandwidth selection argument i

- key tradeoff is between bias and variance:
    1. larger h lowers variance (we use more data)
    2. but also tends to increase bias, unless true regression function exactly polynomial of order p inside estimation window
- Estimator just weighted average of outcomes as in (1), so bias

$$\sum_i w(x_i; h)\mu_Y(x_i) - \tau_Y$$

  and variance as in regression conditional on X (remember OLS lecture!), $\operatorname{var}(\hat{\tau}_{Y,h} \mid X) = \sum_i w(X_i; h)^2 \sigma^2(X_i)$.
- Variance estimation easy (doesn't depend on μ_Y), but bias estimation tricky.

- Classic approach: approximate $\mu_{Y(d)}$ locally by Taylor expansion As $h \to 0$ and $n \to \infty$ (see Theorem 3.2 in Fan and Gijbels 1996):

$$\begin{aligned}\text{bias}(\hat{\tau}_{Y,h}) &= \left[C_B(p,K)\mu_{Y(1)}^{(p+1)}(0)h^{p+1} - C_B(p,K)\mu_{Y(0)}^{(p+1)}(0)h^{p+1}\right](1+o(1)),\\ &= C_B(p,K)h^{p+1}\left[\mu_{Y(1)}^{(p+1)}(0) - \mu_{Y(0)}^{(p+1)}(0)\right](1+o(1)),\end{aligned}$$

where $C_B(p,K)$ is a constant that depends only on the order of the polynomial and the kernel. One could similarly approximate the variance as $nh \to \infty$, and hence the mean squared error (MSE), which yields (pointwise) optimal bandwidth

$$h^*_{\text{PT}} = \left(\frac{C_V(p,K)}{2(p+1)C_B(p,K)^2}\frac{\sigma^2(0_+)+\sigma^2(0_-)}{2f_X(0)(\mu_{Y(1)}^{(p+1)}(0)-\mu_{Y(0)}^{(p+1)}(0))^2\cdot n}\right)^{\frac{1}{2p+3}}. \tag{2}$$

Standard bandwidth selection argument iii

- this bandwidth is not feasible, because we do not know the variances $\sigma^2(0_+)$, $\sigma^2(0_-)$, the derivatives $\mu_{Y(1)}^{(p+1)}(0)$, $\mu_{Y(0)}^{(p+1)}(0)$, or the density $f_X(0)$.
- Imbens and Kalyanaraman (2012) propose a feasible version of this bandwidth based on plugging in estimates of these unknown quantities: very popular in practice.
- So long as $\mu_{Y(1)}^{(p+1)}(0) \neq \mu_{Y(0)}^{(p+1)}(0)$, optimal bandwidth shrinks at rate $O(n^{-\frac{1}{2p+3}})$.
    - optimal if we assume $p+1$ derivatives.
    - resulting convergence rate of $\hat{\tau}_Y$ is $O_p(n^{-\frac{p+1}{2p+3}})$
    - Can get arbitrarily close to parametric rate by assuming enough derivatives...

Issues with standard bandwidth selection

1. Arbitrarily bad performance, even if we use infeasible h^*_{PT}.
    - Taylor-expansion method effectively assumes that we can approximate $\mu_{Y(d)}$ locally around zero by a polynomial of order $p+1$.
    - Fine if h^*_{PT} ends up small. But if $\mu^{(p+1)}_{Y(1)}(0) \approx \mu^{(p+1)}_{Y(0)}(0)$ h^*_{PT} large, and Taylor approximation can be very poor
    - Consider local linear regression and $-\mu_{Y(0)}(x) = \mu_{Y(1)}(x) = x^3$. $h^*_{\text{PT}} = \infty$, and we're not even consistent!
    - To address this problem, plug-in bandwidths such as the Imbens and Kalyanaraman (2012) bandwidth selector that estimate h^*_{PT} include tuning parameters to prevent bandwidth from getting too large. But method then driven by tuning parameter choice
2. To implement h^*_{PT}, need to estimate derivatives of order $p+1$:
    - much harder than our initial problem of estimating intercept
    - requires derivatives of order $p+2$ exist. But if that's the case, could have used polynomial of order $p+1$ instead!
    - estimator optimal in class of estimators (local polynomial estimators of order p), that is itself suboptimal.

Bias aware approach i

- Choose bandwidth to minimize worst-case MSE of $\hat{\tau}_Y$ over all possible μ_Y that have second derivatives bounded by M.
- If we use local linear regression, least favorable function has closed form: $\mu_{Y(1)}(x) = -Mx^2/2$ and $\mu_{Y(0)}(x) = Mx^2/2$
  - Intuition?
- Closed-form expression for worst-case MSE: no Taylor approximation!

$$\sum_i w(X_i;h)\sigma^2(X_i) - M\left[\sum_{i:\, X_i \geq 0} w(X_i;h)X_i^2 - \sum_{i:\, X_i<0} w(X_i;h)X_i^2\right]^2. \tag{3}$$

Can minimize numerically to obtain finite-sample optimal bandwidth h^*_{MSE}. In practice, need to estimate variance—can assume homoskedasticity to make that part easy.

Bias aware approach ii

- No assumptions on distribution of X_i: in particular, nothing changes if the distribution of the running variable is discrete
- Compare bandwidths in large samples:

$$h^*_{\text{PT}} = \left(\frac{C_V(p,K)}{2(p+1)C_B(p,K)^2} \frac{\sigma^2(0_+) + \sigma^2(0_-)}{2f_X(0)(\mu^{(p+1)}_{Y(1)}(0) - \mu^{(p+1)}_{Y(0)}(0))^2 \cdot n} \right)^{\frac{1}{2p+3}}$$

$$h^*_{\text{MSE}} = \left(\frac{C_V(p,K)}{2(p+1)\tilde{C}_B(p,K)^2} \cdot \frac{\sigma^2_+(0) + \sigma^2_-(0)}{2f_X(0) \cdot 4M^2 \cdot n} \right)^{\frac{1}{2p+3}} (1 + o_p(1)),$$

- To implement, need to figure out second derivative bound, curvature M (global problem) instead of second derivative at zero.

Curvature calibration i

- If M too large, we'll be unnecessarily conservative. Can we use data to estimate it well?
- No possible without further restrictions if goal is inference (Armstrong and Kolesár 2018).
  - Instance of the general issue with using pre-testing or using model selection rules: model selection distorts infernece
  - Here curvature parameter M indexes model size. Large M like saying we more of available covariates possible confounders in OLS, small M like saying we don't need to include them. Without restrictions on OLS coeffs, best we can do is include all of them! Otherwise need to use institutional knowledge
  - Ideally would use institutional knowledge to pick M: hard sell!
  - Analogous to reporting results based on different subsets of controls in columns of a table with regression results, vary choice of M by way of sensitivity analysis.
- Armstrong and Kolesár (2020) suggest rule of thumb for calibrating M, based on heuristic that local smoothness of μ_d is no smaller than its smoothness at large scales:

Curvature calibration ii

  - Fit a global polynomial on either side of the cutoff, and calculate the largest second derivative of the fitted polynomial. Set M to this value.
  - Formally "works" if second derivative of $\mu_{Y(1)}$ and $\mu_{Y(0)}$ near zero indeed bounded by the largest second derivative of a global polynomial approximation to $\mu_{Y(0)}$ and $\mu_{Y(1)}$
  - Question whether "better" calibration possible (alternatives have been proposed, e.g. Imbens and Wager (2019), but suffer form same issues). Recent work on this: https://arxiv.org/abs/2503.09907
- Good idea to plot approximation to μ that imposes this rule of thumb, by, say, fitting splines.

Inference

- Estimator just weights average of outcomes, so asymptotically normal under minimal assumptions, so long as weights $w(X_i)$ not too large:

$$\frac{\hat{\tau}_{Y,h} - \tau}{\mathrm{var}(\hat{\tau}_{Y,h})^{1/2}} \approx \mathcal{N}\left(\frac{\mathrm{bias}(\hat{\tau}_{Y,h})}{\mathrm{var}(\hat{\tau}_{Y.h})^{1/2}}, 1\right) + o_p(1), \tag{4}$$

- But if h optimally chosen, $b = \mathrm{bias}(\hat{\tau}_{Y,h})/\mathrm{var}(\hat{\tau}_{Y.h})^{1/2}$ not close to zero!

Standard solutions

1. Undersmooth: choose h smaller than optimal. But how small? In practice anything goes.
2. Bias-correct: try to estimate bias and subtract it off.
   - Like with h^*_{PT}, can only do this if have more smoothness than optimal for local linear to be optimal.
   - Even if feasible, bias estimate noise, and resulting confidence intervals (CIs) poor (Hall 1992)
   - Calonico, Cattaneo, and Titiunik (2014) propose adjusting variance estimator to take into account the variability of bias estimate, which they call robust bias correction (RBC).
   - Important special case: if bandwidth for bias estimation equals h, this reduces to local quadratic regression (but with original bandwidth, calibrated for local linear)
   - I think of RBC as particular (more principled) way of implementing undersmoothing.

Bias aware inference

- t-stat asymptotically normal, but don't know mean b. But we have bound on bias—already calculated it to compute optimal bandwidth, so use it to bound b
- leads to CI

$$\hat{\tau}_{Y,h} \pm \mathrm{cv}_{\alpha}(\bar{B}) \operatorname{var}(\hat{\tau}_{Y,h})^{1/2},$$

where $\mathrm{cv}_{\alpha}(b)$ is the α quantile of the $|\mathcal{N}(b, 1)|$ and $\bar{B}$ is given by ratio of bias bound to standard error.
- Advantages:
  1. honest: validity doesn't rely on undersmoothing, or any other asymptotic promises about how the bandwidth would shrink with the sample size
  2. valid uniformly over the whole parameter space of all functions μ_Y with second derivative bounded by M
  3. *bias-aware*: length reflects the potential finite-sample bias of the estimator.

Remarks

1. Can estimate variance using Eicker-Huber-White (EHW). But we're doing inference conditional on X_i, so this is conservative by lecture on OLS. Using nearest-neighbor variance estimator better
   - This is a case where we are misspecified and willing to admit it
2. Bias-aware inference works with discrete running variable. Discrete X_i formally ruled out in the undersmoothing and RBC approaches.
   Another popular proposal for handling discrete covariates: cluster the errors by the running variable (Lee and Card 2008).
   - Has a serious deficiency: it may lead to confidence intervals that are *shorter* than unclustered CIs. See Kolesár and Rothe (2018) for a detailed discussion of this point. Intuition: creates inference with a small number of clusters problem, but doesn't solve bias issue.
3. Since we're just doing OLS, main regularity check is to check leverage not too high

Identification

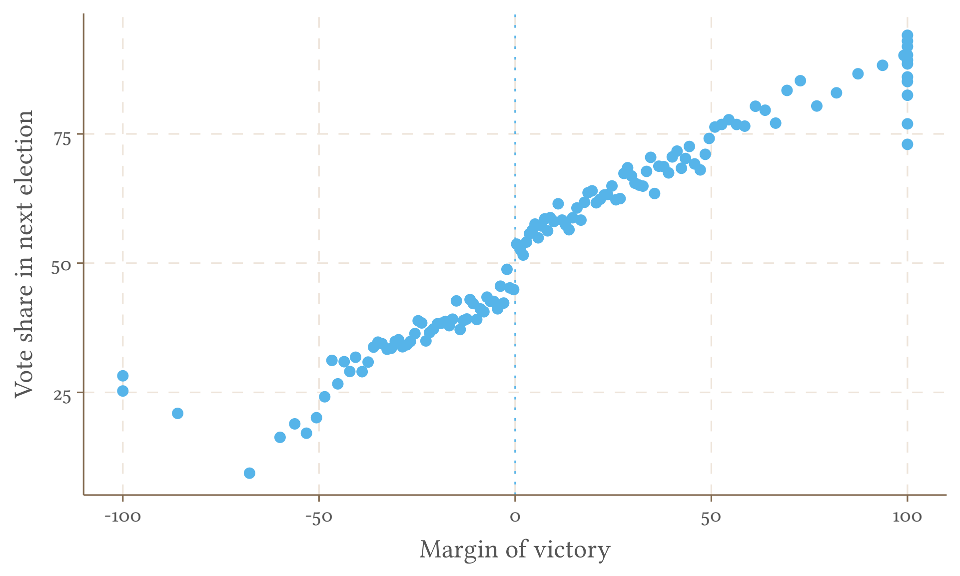
Falsification

Estimation and Inference in sharp RD

Empirical illustration

Extensions

- Use the dataset from Lee (2008) on 6,558 observations on elections to US House of Representatives between 1946 and 1998
- Running variable $X_i \in [-100, 100]$ is the Democratic margin of victory (in percentages) in election i. The outcome variable $y_i \in [0, 100]$ is the Democratic vote share (in percentages) in the next election.


Vote share in next election
75
50
25
-100
-50
0
50
100
Margin of victory

- For estimation, we use $p = 1$ (local linear regression), and the triangular kernel.
- Armstrong and Kolesár (2020) rule of thumb yields $M = 0.14$, which is driven by observatinons with $X \leq -50$ (can see from graph). If (somewhat arbitrarily) restrict attention to the 4,900 observations within distance 50 of the cutoff, we obtain $M = 0.04$.
- For comparison, IK bandwidth is about 30, due to small curvature near cutoff.

M	Estimate	Bias	SE	95% bias-aware CI	Effective obs.	h	$\bar{L}$
0.14	5.85	0.89	1.37	(2.69, 9.01)	764	7.7	0.01
0.04	6.24	0.71	1.12	(3.66, 8.81)	1250	12.8	0.01

See notes for discussion of first two extensions and references for the other extensions

- estimation and inference in fuzzy RD
- Incorporating covariates
- kink designs
- bunching designs
- multiple cutoffs, or multi-dimensional running variables
- extrapolating treatment effects away from cutoff

Angrist, Joshua D., Victor Lavy, Jetson Leder-Luis, and Adi Shany. 2019. "Maimonides' Rule Redux." *American Economic Review: Insights* 1, no. 3 (December): 309–324. https://doi.org/10.1257/aeri.20180120.

Armstrong, Timothy B., and Michal Kolesár. 2018. "Optimal Inference in a Class of Regression Models." *Econometrica* 86, no. 2 (March): 655–683. https://doi.org/10.3982/ECTA14434.

———. 2020. "Simple and Honest Confidence Intervals in Nonparametric Regression." *Quantitative Economics* 11, no. 1 (January): 1–39. https://doi.org/10.3982/QE1199.

Battistin, Erich, Agar Brugiavini, Enrico Rettore, and Guglielmo Weber. 2009. "The Retirement Consumption Puzzle: Evidence from a Regression Discontinuity Approach." *American Economic Review* 99, no. 5 (December): 2209–2226. https://doi.org/10.1257/aer.99.5.2209.

Bleemer, Zachary, and Aashish Mehta. 2022. "Will Studying Economics Make You Rich? A Regression Discontinuity Analysis of the Returns to College Major." *American Economic Journal: Applied Economics* 14, no. 2 (April): 1–22. https://doi.org/10.1257/app.20200447.

Calonico, Sebastian, Matias D. Cattaneo, and Rocío Titiunik. 2014. "Robust Nonparametric Confidence Intervals for Regression-Discontinuity Designs." *Econometrica* 82, no. 6 (November): 2295–2326. https://doi.org/10.3982/ECTA11757.

Canay, Ivan Alexis, and Vishal Kamat. 2018. "Approximate Permutation Tests and Induced Order Statistics in the Regression Discontinuity Design." *The Review of Economic Studies* 85, no. 3 (July): 1577–1608. https://doi.org/10.1093/restud/rdx062.

References ii

Cheng, Ming-Yen, Jianqing Fan, and J. S. Marron. 1997. "On Automatic Boundary Corrections." *The Annals of Statistics* 25, no. 4 (August): 1691–1708. https://doi.org/10.1214/aos/1031594737.

Fan, Jianqing, and Irène Gijbels. 1996. *Local Polynomial Modelling and Its Applications.* Monographs on Statistics and Applied Probability 66. New York, NY: Chapman & Hall/CRC. https://doi.org/10.1201/9780203748725.

Ganong, Peter, and Simon Jäger. 2018. "A Permutation Test for the Regression Kink Design." *Journal of the American Statistical Association* 113, no. 522 (April): 494–504. https://doi.org/10.1080/01621459.2017.1328356.

Gelman, Andrew, and Guido W. Imbens. 2019. "Why High-Order Polynomials Should Not Be Used in Regression Discontinuity Designs." *Journal of Business & Economic Statistics* 37, no. 3 (July): 447–456. https://doi.org/10.1080/07350015.2017.1366909.

Haggag, Kareem, and Giovanni Paci. 2014. "Default Tips." *American Economic Journal: Applied Economics* 6, no. 3 (July): 1–19. https://doi.org/10.1257/app.6.3.1.

Hall, Peter. 1992. "Effect of Bias Estimation on Coverage Accuracy of Bootstrap Confidence Intervals for a Probability Density." *The Annals of Statistics* 20, no. 2 (June): 675–694. https://doi.org/10.1214/aos/1176348651.

Imbens, Guido W., and Joshua D. Angrist. 1994. "Identification and Estimation of Local Average Treatment Effects." *Econometrica* 62, no. 2 (March): 467–475. https://doi.org/10.2307/2951620.

Imbens, Guido W., and Karthik Kalyanaraman. 2012. “Optimal Bandwidth Choice for the Regression Discontinuity Estimator.” *The Review of Economic Studies* 79, no. 3 (July): 933–959. https://doi.org/10.1093/restud/rdr043.

Imbens, Guido W., and Stefan Wager. 2019. “Optimized Regression Discontinuity Designs.” *The Review of Economics and Statistics* 101, no. 2 (May): 264–278. https://doi.org/10.1162/rest_a_00793.

Klaauw, Wilbert van der. 2002. “Estimating the Effect of Financial Aid Offers on College Enrollment: A Regression-Discontinuity Approach.” *International Economic Review* 43, no. 4 (November): 1249–1287. https://doi.org/10.1111/1468-2354.t01-1-00055.

Kolesár, Michal, and Christoph Rothe. 2018. “Inference in Regression Discontinuity Designs with a Discrete Running Variable.” *American Economic Review* 108, no. 8 (August): 2277–2304. https://doi.org/10.1257/aer.20160945.

Lee, David S. 2008. “Randomized Experiments from Non-Random Selection in U.S. House Elections.” *Journal of Econometrics* 142, no. 2 (February): 675–697. https://doi.org/10.1016/j.jeconom.2007.05.004.

Lee, David S., and David Card. 2008. “Regression Discontinuity Inference with Specification Error.” *Journal of Econometrics* 142, no. 2 (February): 655–674. https://doi.org/10.1016/j.jeconom.2007.05.003.