Partial Identification for Regression Discontinuity Donuts

Connor Dowd

University of Chicago Booth School of Business

Chicago Econometrics Lunch
November 12, 2019

Primary Question

Question

When we use a donut, how can we learn about the treatment effect?

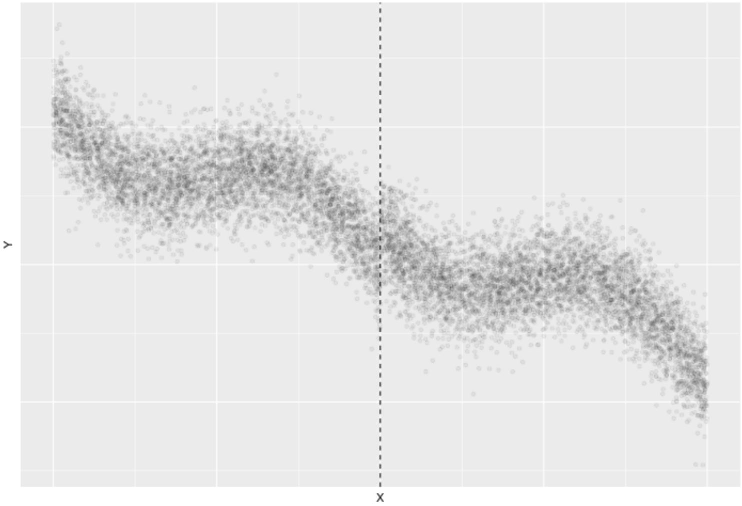
Primary Question

Question

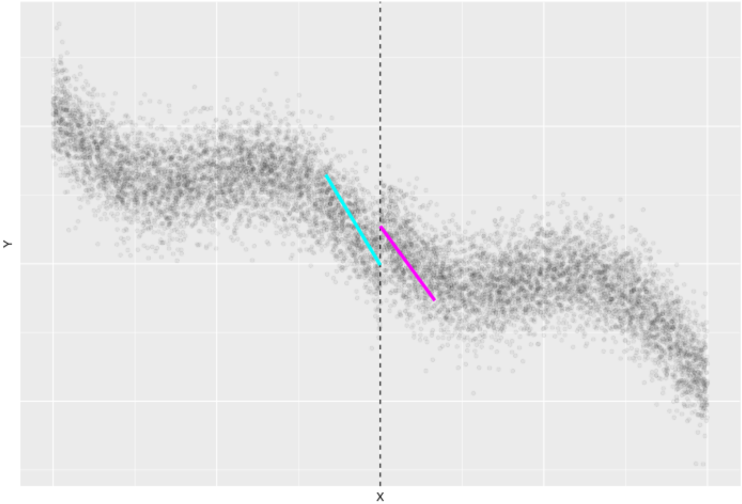
When we use a donut, how can we learn about the treatment effect?

Main result Under natural extensions of standard assumptions and data-driven derivative bounds, we get partial identification – and can conduct inference for the partially identified set.

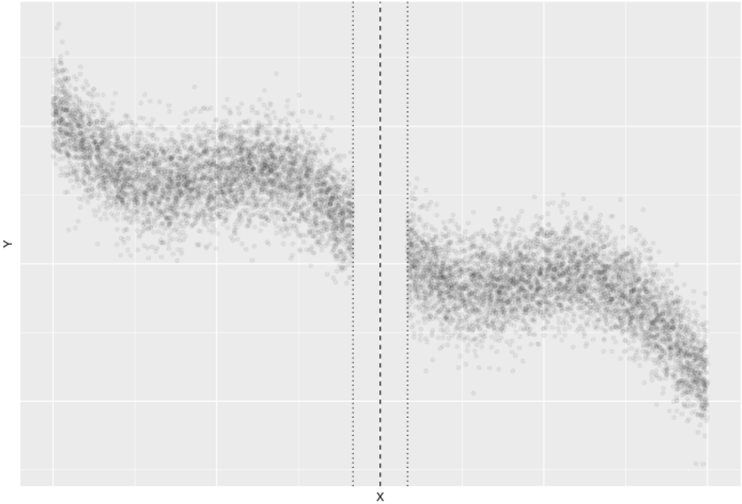
RD Example



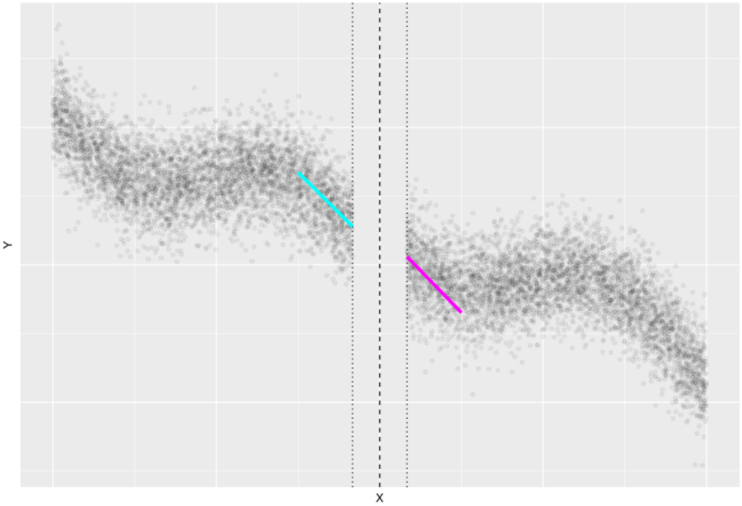
RD Example



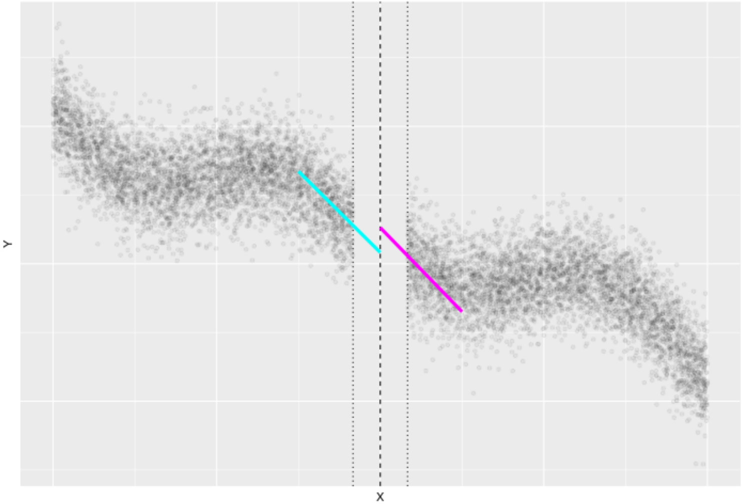
RD Donut Example



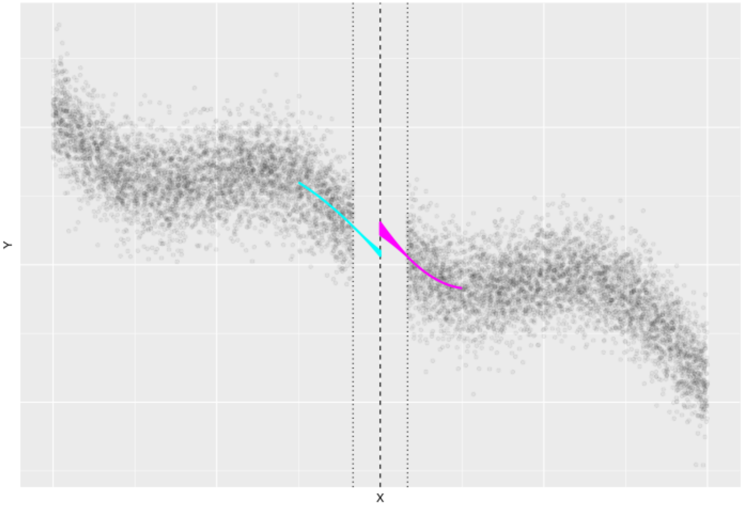
RD Donut Example



RD Donut Example



RD Donut Example



How Do Donuts Work?

- There is no extant theory despite widespread empirical use.
- Most papers make implicit functional form assumptions.
- Even under those assumptions, more is needed than for standard RD.
- Can we use weaker restrictions on DGP?

This paper...

I use smoothness assumptions, which are natural to the RD setting, to perform inference with a donut.

I focus on the Sharp RD case (full treatment), with an additive treatment effect.

Outline

- Donut Condition
- Standard RD Conditions
- Derivative Bounds

Outline

- Donut Condition
- Standard RD Conditions
- Derivative Bounds
- Results
- Probation Example
- Future Directions

Donut Exclusion

Condition 1: Donut Exclusion

(i) There is a known interval $\mathbb{D} = (d_-, d_+)$ such that all individuals who manipulate are contained to the interval, and would be contained in the counterfactual where they do not manipulate.

(ii) There is only one policy with a threshold relevant to the outcome variable inside the region $[d_- - \epsilon, d_+ + \epsilon]$ for some $\epsilon > 0$.

Donut Exclusion

Condition 1: Donut Exclusion

(i) There is a known interval $\mathbb{D} = (d_-, d_+)$ such that all individuals who manipulate are contained to the interval, and would be contained in the counterfactual where they do not manipulate.

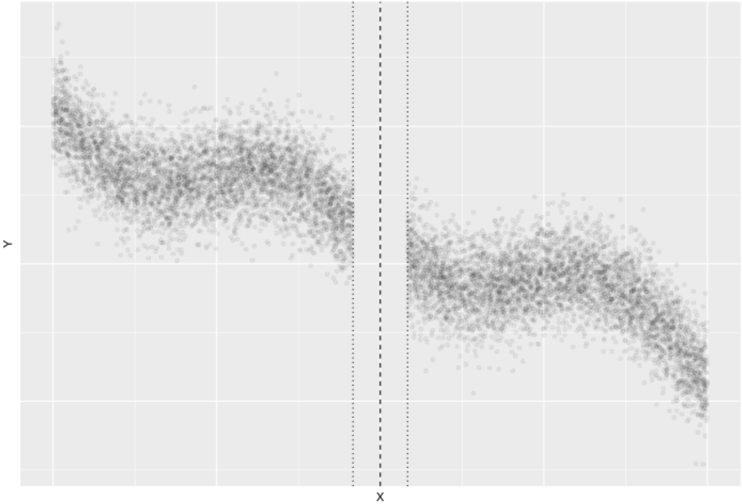
(ii) There is only one policy with a threshold relevant to the outcome variable inside the region $[d_- - \epsilon, d_+ + \epsilon]$ for some $\epsilon > 0$.

I define manipulation as the difference between the observed running variable, and the value in the counterfactual where all individuals treatment statuses were fixed in advance.

Comments

- (i) generalizes the standard RD assumption that there is no manipulation (i.e. $\mathbb{D} = (0, 0)$).
- (ii) generalizes the standard RD assumption that there are no co-located policies.
- We only care about manipulation which is caused by the treatment threshold.
- The Donut size is not shrinking asymptotically – it is a feature of people´s ability to manipulate, and so I take it as fixed.
- (i) insists we cannot just exclude the region where there is observable bunching.

Donut Example



Technical Conditions

Condition 2: DGP conditions

$\exists$ a known value $\eta > 0$ defining a set $\mathbb{C} = [d_- - \eta, d_-] \cup [d_+, d_+ + \eta]$ such that:

(i) The density of X, $f_x(\cdot)$ is positive over $\mathbb{C}$

(ii) $f_x^{(1)}(\cdot)$ is continuous over $\mathbb{C}$

(iii) $\mu_t^{(2)}(\cdot)$ are continuous over $\mathbb{C}$

(iv) $\upsilon_t^{(2)}(\cdot)$ are continuous over $\mathbb{C}$

(v) $\upsilon_t(\cdot)$ are positive and bounded over $\mathbb{C}$.

When $\mathbb{C} = [-\eta, \eta]$, these are standard conditions for RD.
See Porter [2003].
$\mu_t()$ is the CEF, and $\upsilon_t^2()$ is the conditional variance.

Local Polynomial Conditions

Condition 3: Kernel and Bandwidth

(i) The kernel function $K(\cdot)$ has support $(-1, 1)$, outside of which it takes value 0.

(ii) $K(\cdot)$ is symmetric, positive, bounded, and integrates to 1 over its support.

(iii) The bandwidth $h = h_n$ is set such that as $n \to \infty$, $h_n \to 0$ and $nh_n^3 \to \infty$.

(iv) $h \leq \eta \ \ \forall n$.

Local Polynomial Conditions

Condition 3: Kernel and Bandwidth

(i) The kernel function $K(\cdot)$ has support $(-1, 1)$, outside of which it takes value 0.

(ii) $K(\cdot)$ is symmetric, positive, bounded, and integrates to 1 over its support.

(iii) The bandwidth $h = h_n$ is set such that as $n \to \infty$, $h_n \to 0$ and $nh_n^3 \to \infty$.

(iv) $h \leq \eta \ \ \forall n$.

We use the kernel $K_h(x) = K(x/h)/h$.

These are standard conditions for local polynomial regressions.
See Fan, Heckman, and Wand [1995].

Derivative Bounds

Condition 4: Derivative Bounds

There is a known $k > 0$ such that

(i) $\mu_0^{(k)}(x) \in [l_0, u_0] \; \forall x \in [d_-, 0]$

(ii) $\mu_1^{(k)}(x) \in [l_1, u_1] \; \forall x \in [0, d_+]$.

Derivative Bounds

Condition 4: Derivative Bounds

There is a known $k > 0$ such that

(i) $\mu_0^{(k)}(x) \in [l_0, u_0]\ \forall x \in [d_-, 0]$

(ii) $\mu_1^{(k)}(x) \in [l_1, u_1]\ \forall x \in [0, d_+]$.

(iii) $\mu_t^{(k+2)}(\cdot)$ are continuous over $\mathbb{C} \cup \mathbb{D}$ for $\eta, \mathbb{C}$ from Condition 2.

We don't need to know l_1, l_0, u_1, or u_0, but we need to be able to estimate them 'well'.

Partial Identification

Lemma 1

Under conditions 1-4, there is some set $\phi = [\tau_l, \tau_u]$ such that

(a) $\tau \in \phi$

(b) $\tau_u - \tau_l < \infty$

Partial Identification

Lemma 1

Under conditions 1-4, there is some set $\phi = [\tau_l, \tau_u]$ such that

(a) $\tau \in \phi$

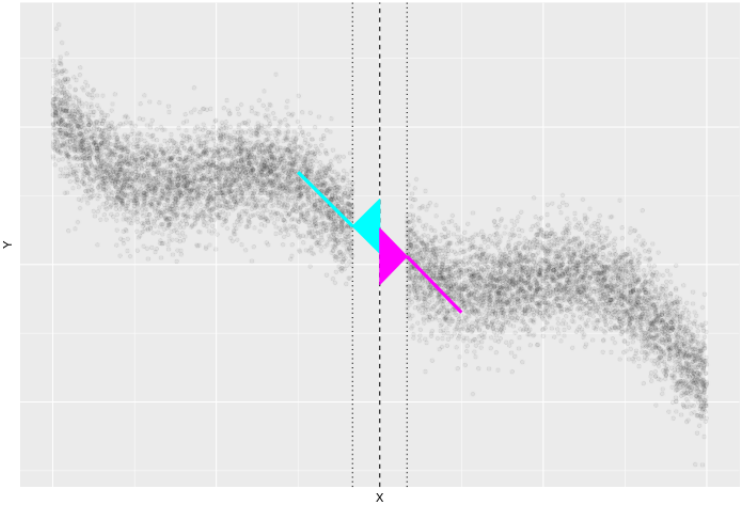
(b) $\tau_u - \tau_l < \infty$

Concretely:

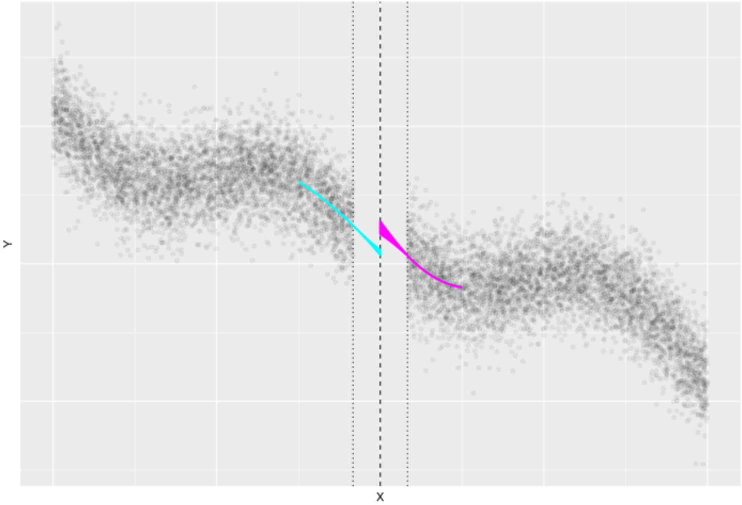
$$\mu_{0,l}(0) = \left(\sum_{j=0}^{k-1} \frac{d_-^j}{j!} \mu_0^{(j)}(d_-) \right) + \frac{d_-^k}{k!} l_0$$

$$\tau_u = \mu_{1,u}(0) - \mu_{0,l}(0)$$

Derivative Bound Example: $k = 1$



Derivative Bound Example $k = 2$



Normality

Lemma 2

Under conditions 1-4, local polynomial estimates of the stacked sequence of derivatives at d_-, d_+

$$\hat{\theta} = \left(\hat{\mu}_0^{(0)}(d_-), ..., \hat{\mu}_0^{(k-1)}(d_-), \quad \hat{\mu}_1^{(0)}(d_+), ..., \hat{\mu}_1^{(k-1)}(d_+)\right)^T$$

converge to a normal distribution with a block diagonal covariance and bias of the order nh^{2k+3}

Well Behaved bounds

Condition 5: Joint Distribution is Estimable

We have estimates of l_0, l_1, u_0, u_1 which are consistent and have a known (or easily estimated) joint sampling distribution when stacked with $\hat{\theta}$.

Examples:

- Bounds are known points.
- Bounds are given by differentiable functions of $\mu_t^{(k)}()$ at d_- or d_+. (Estimable covariance, joint normality)
- Bounds come from global structure estimated outside of bandwidth. (Independent from $\hat{\theta}$)
- Bounds come from pre-period with no treatment, post-period with full treatment.

Comments

- Moving forwards, I focus on inference for the region ϕ.
- In examples and discussion, I will use bounds of the form

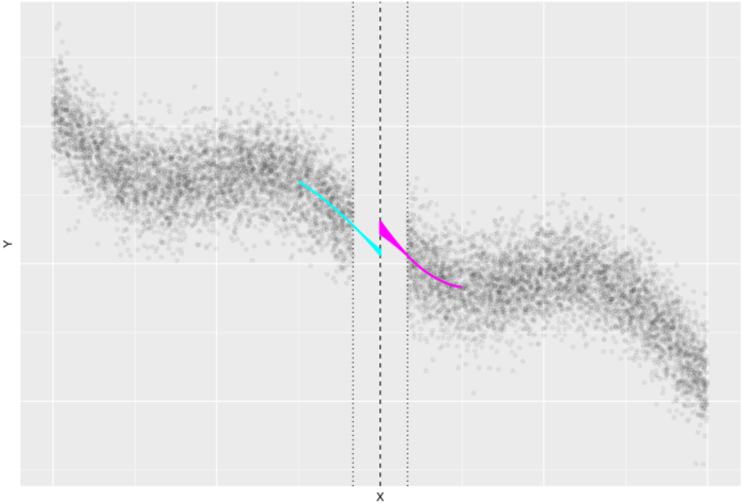
$$|\mu_0^{(k)}(x)| \leq |\mu_0^{(k)}(d_-)|$$

$$|\mu_1^{(k)}(x)| \leq |\mu_1^{(k)}(d_+)|$$

$\forall x \in \mathbb{D}$

This will give us joint normality of our Stacked vector. I'll highlight results leaning on joint normality today.

Derivative Bound Example



Coverage for ϕ

Define C such that $\Phi(C) - \Phi(-C) = 1 - \alpha$

$$\mathbb{S}_{1-\alpha} = \left[\hat{\tau}_l - C\hat{\sigma}_l/\sqrt{n},\ \hat{\tau}_u + C\hat{\sigma}_u/\sqrt{n}\right]$$

Theorem 1

Under conditions 1-4, and the condition that $nh^{2k+3} \to 0$, for all $\alpha \in (0, 1/2)$,

$$\lim_{n\to\infty} P[\phi \subseteq \mathbb{S}_{1-\alpha}] = 1 - \alpha$$

Comments on Theorem 1

Theorem 1 gives us asymptotic size control for the set ϕ.

Theorem 1 is dependent on the two bandwidth conditions:
$nh^3 \rightarrow \infty$ and $nh^{2k+3} \rightarrow 0$.

Theorem 1 is very conservative for each values of τ in ϕ. In order to cover the entire interval, each point must be covered with much higher probability.

$\mathbb{S}_{1-\alpha}$ Example

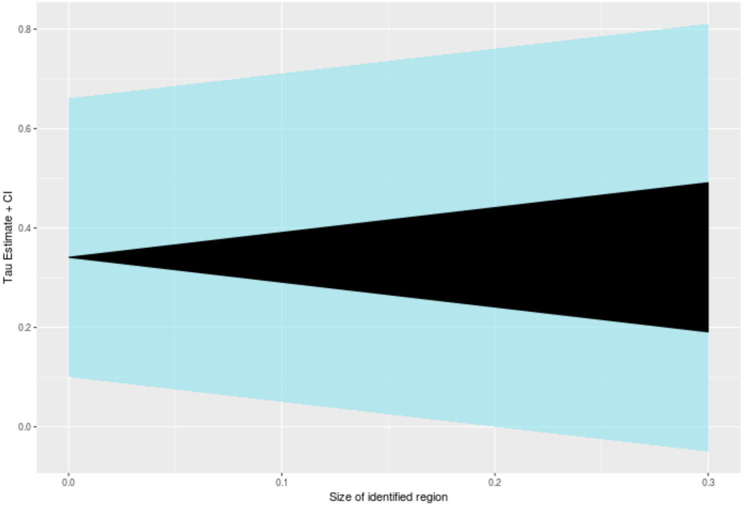


Figure 1: Confidence region across sizes of identified set.

Coverage for τ

Define $\hat{\Delta} = \hat{\tau}_u - \hat{\tau}_l$, and $\hat{\sigma}_m = max(\hat{\sigma}_l, \hat{\sigma}_u)$.

Define C_n such that $\Phi(C_n + \hat{\Delta}\sqrt{n}/\hat{\sigma}_m) - \Phi(-C_n) = 1 - \alpha$.

Define $\mathbb{Q}_{1-\alpha} = \left[\hat{\tau}_l - C_n\hat{\sigma}_l/\sqrt{n}, \;\; \hat{\tau}_u + C_n\hat{\sigma}_u/\sqrt{n}\right]$

Theorem 2

Under conditions 1-4, and the further conditions that $nh^{2k+3} \rightarrow 0$ and $d_+ - d_- > 0$, for all $\alpha \in (0, 1/2)$

$$\lim_{n\rightarrow\infty} \inf_{\theta\in\phi} P[\theta \in \mathbb{Q}_{1-\alpha}] = 1 - \alpha$$

$\mathbb{Q}_{1-\alpha}$ Example

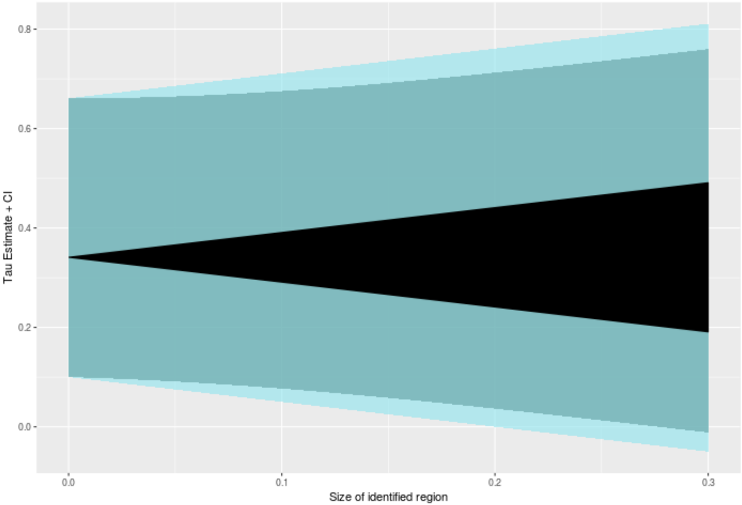


Figure 2: Confidence regions across size of identified set.

Academic Probation

A Canadian university system imposes academic probation for students who have a GPA less than 1.5 after their first year.

Lindo, Sanders, and Oreopoulos [2010] examine the data, test for discontinuity, and look at covariate smoothness. They perform inference for treatment effects on future GPA (among other things). Cattaneo, Idrobo, and Titiunik [2019] replicate and provide the data and code.

GPAs are very much under students' control. It is very possible (and extremely low cost) to ask professors to raise your grade.

Academic Probation – Binscatter

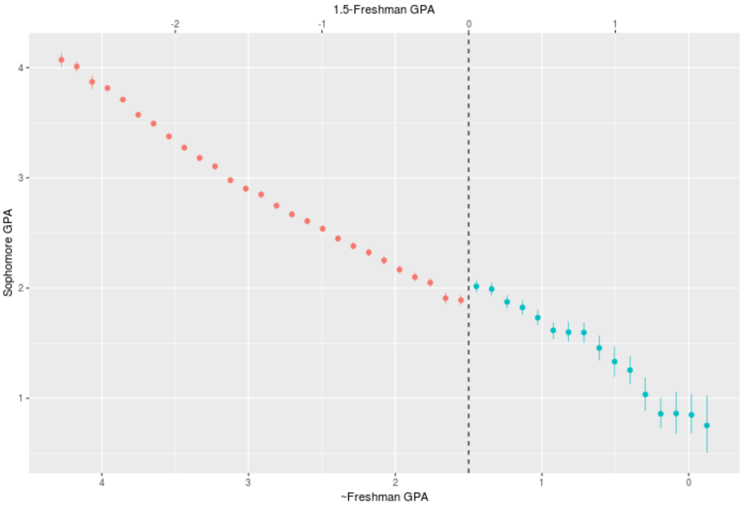


Figure 3: Dashed line indicates treatment.

Academic Probation - Methods

If 1/3 of grades are high enough that professors raise them one partial letter (e.g. C+ to B-) on being asked, that is a maximum GPA change of 0.2.

I use a donut of width 0.25.

I bound the 2nd derivative, and follow the original authors in using a uniform kernel and a bandwidth of 0.6.

Academic Probation - All Data

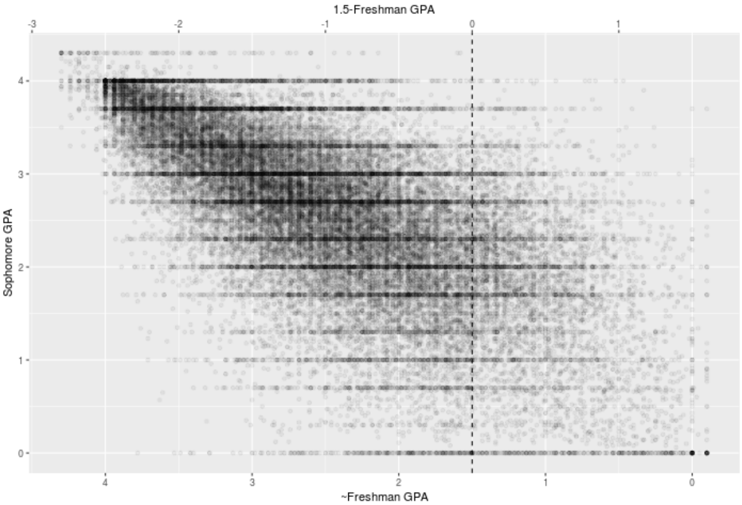


Figure 4: Dashed Line is the threshold.

Academic Probation - Drop Donut

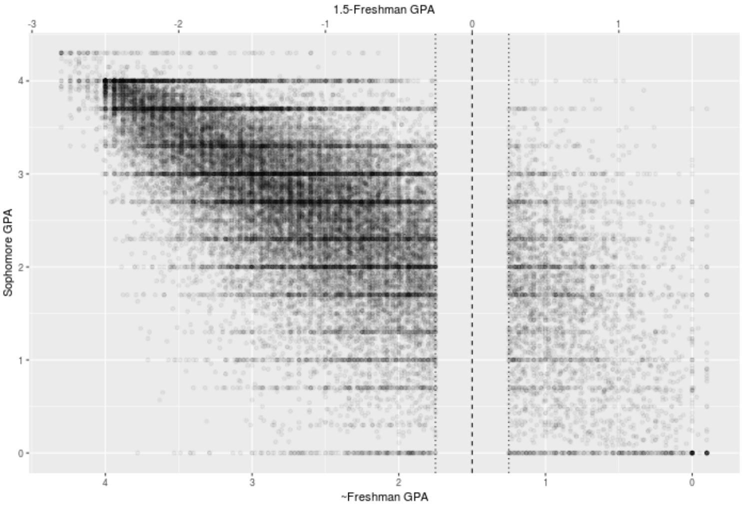


Figure 5: Dotted lines are donut boundaries.

Academic Probation - Inside Bandwidth

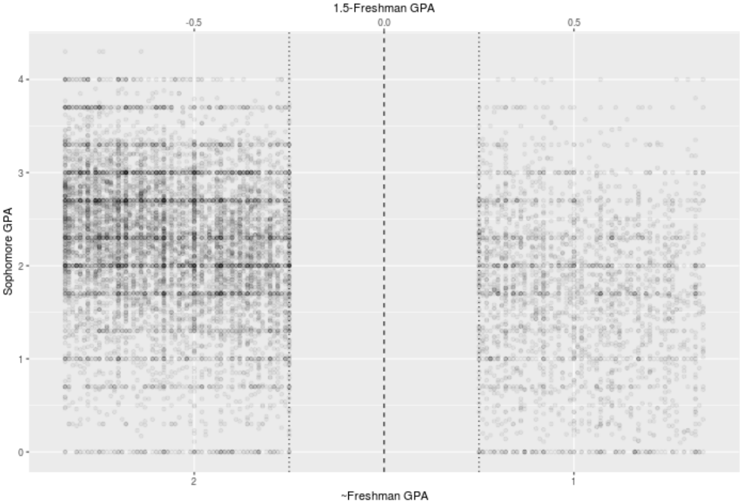


Figure 6: Only data within Bandwidths

Academic Probation - Fit Local Polynomials

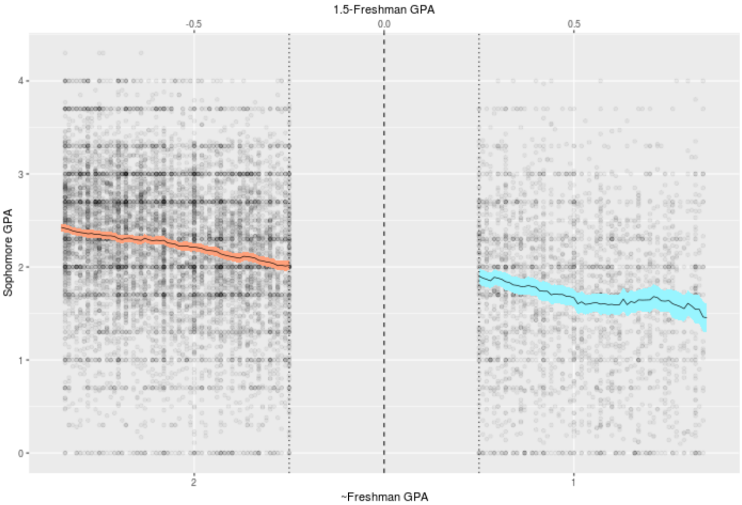


Figure 7: Lines are estimates of $\mu_t(x)$, with pointwise 95% CIs around them.

Academic Probation - Identified Region

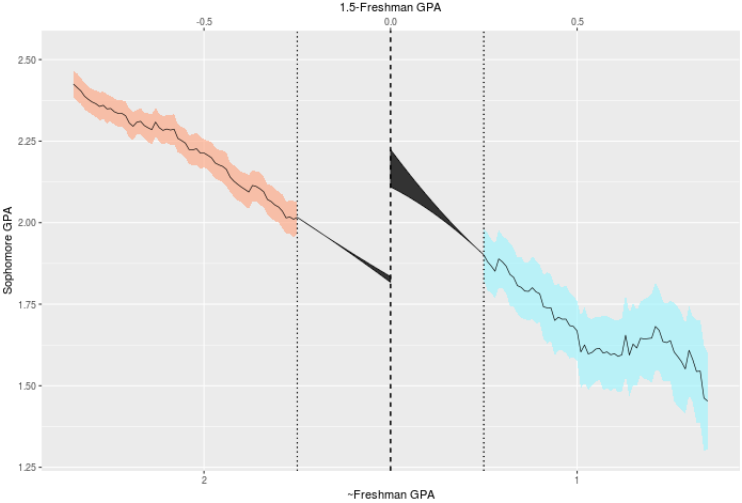


Figure 8: Black region is the estimate of the identified region. Y-axis changed.

Academic Probation - CR for Set

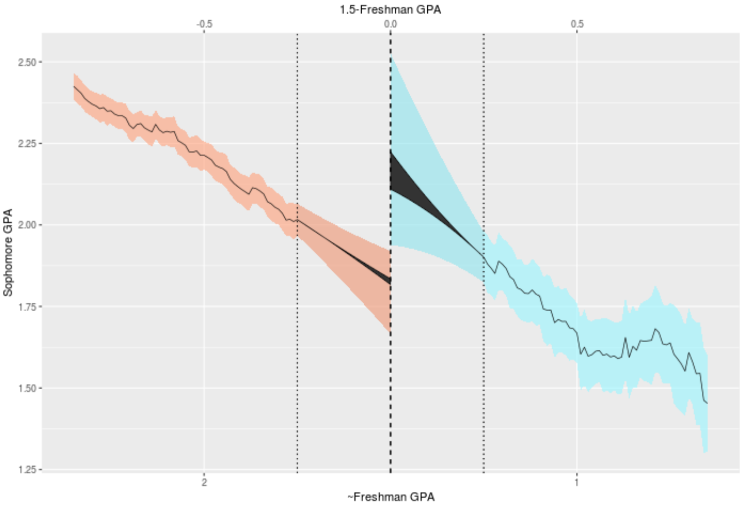


Figure 9: CR inside donut is built to contain entire identified set.

Academic Probation - CR for elements of set

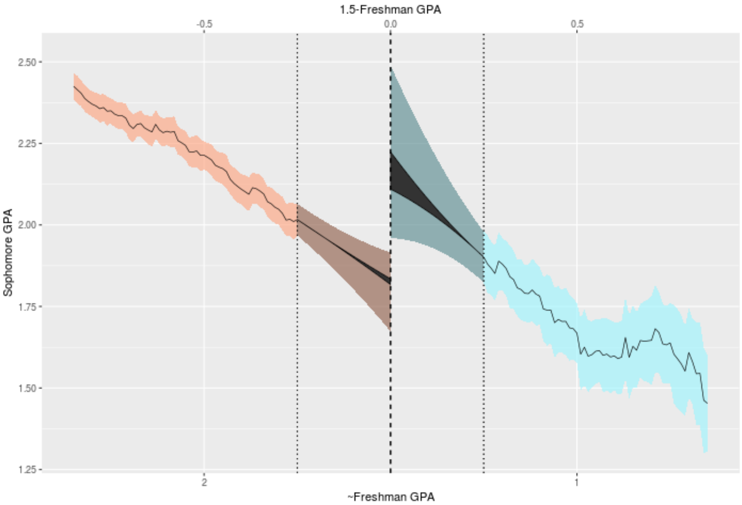


Figure 10: CR is built to cover elements in the identified set.

Tau Set

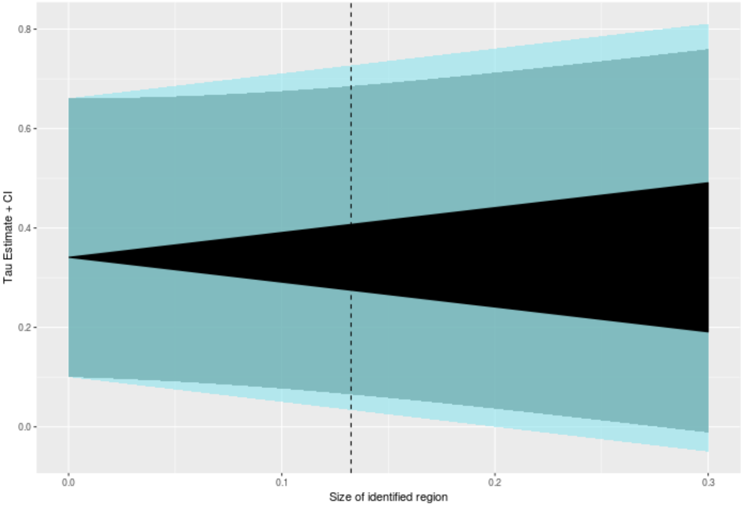


Figure 11: Dashed line gives the value of $\hat{\Delta}$.

Results

	Estimate	CR Lower	CR Upper
Original	0.233	0.181	0.285
Robust	0.213	0.122	0.304
$\hat{\tau} - DONUT$	$[0.275, 0.407]$	0.065	0.686
$\hat{\phi} - DONUT$	$[0.275, 0.407]$	0.034	0.727

Table 1: Comparison of Estimates from rdrobust and Donut routines

Note the improvement of the CR width between $\hat{\tau}$ intervals and $\hat{\phi}$ intervals.

Conclusion

We have discussed:

- Derivative based conditions under which a set is identified.
- Asymptotically Conservative inference for both parameters and the identified set.
- Application to Academic probation.

Future Work:

- Can we give more guidance on donut sizes?
- Efficiency in estimating ϕ?
- Guidance on polynomial order.

Thank you