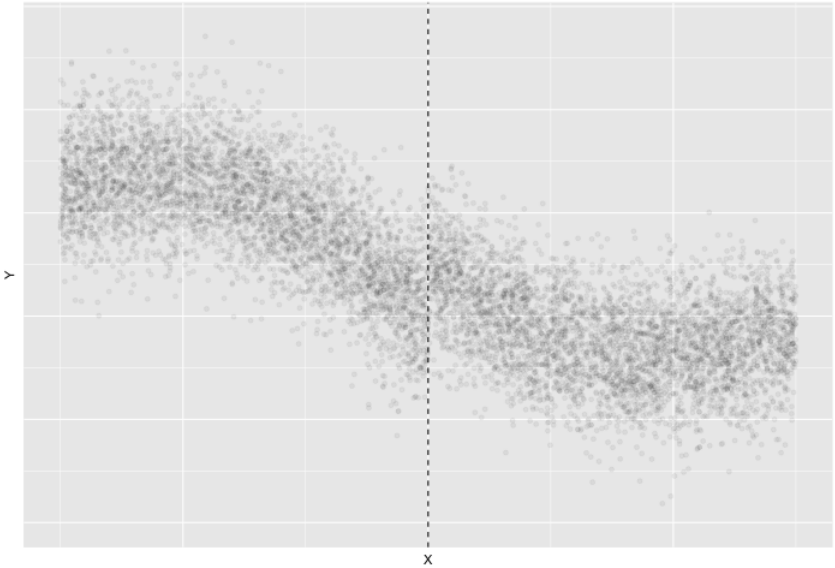
Regression Discontinuity Donuts

Connor Dowd

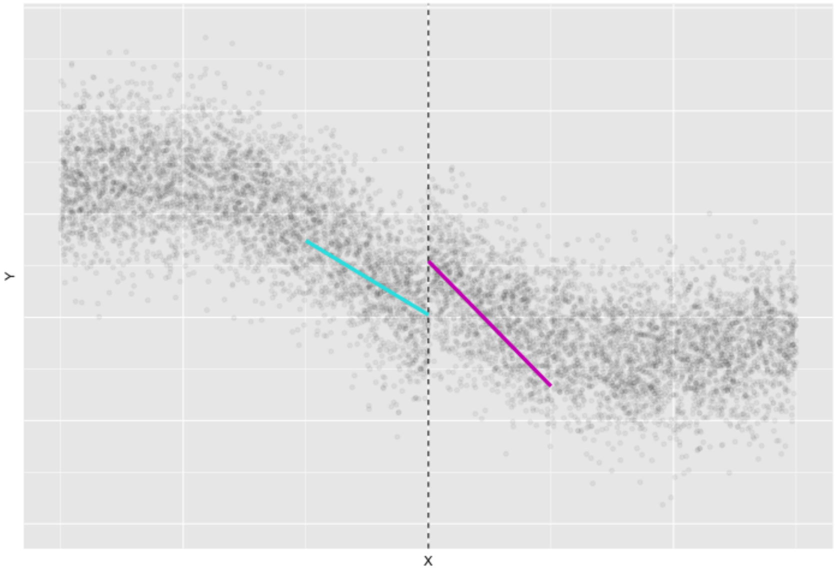
University of Chicago Booth School of Business

January 2021

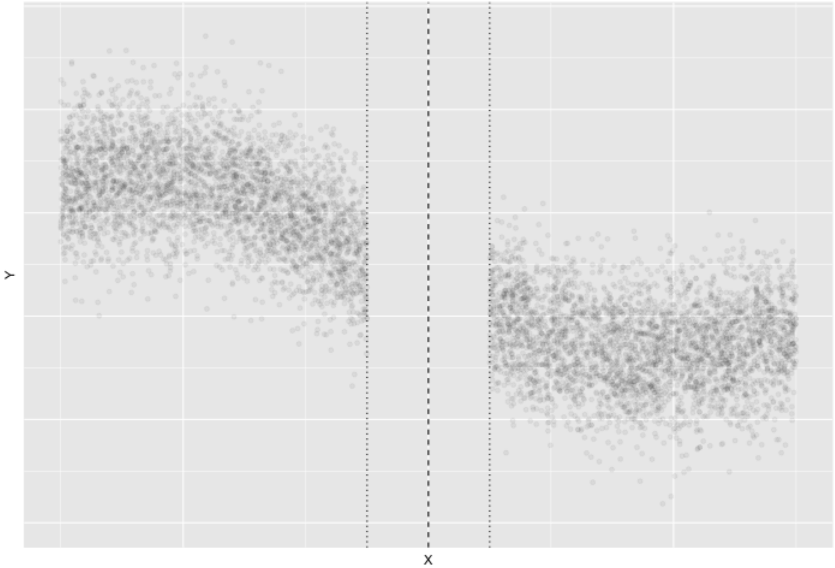
RD Example



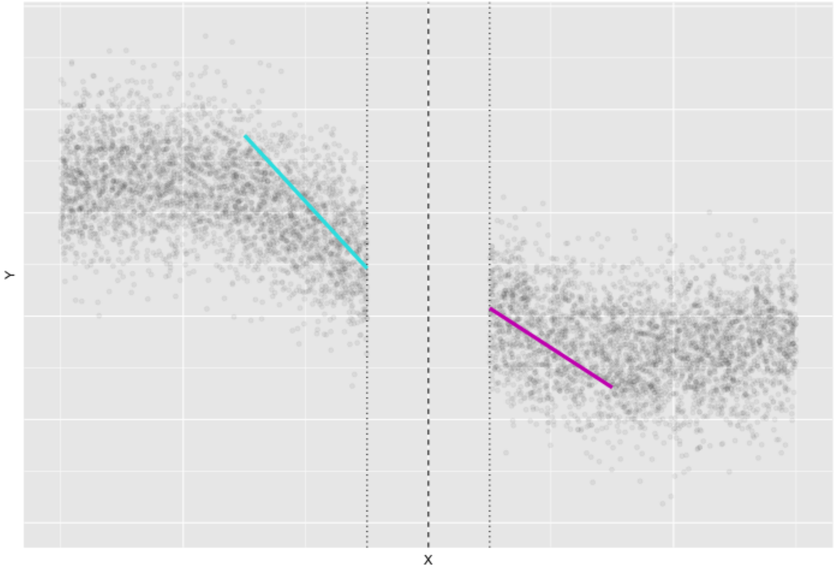
RD Example



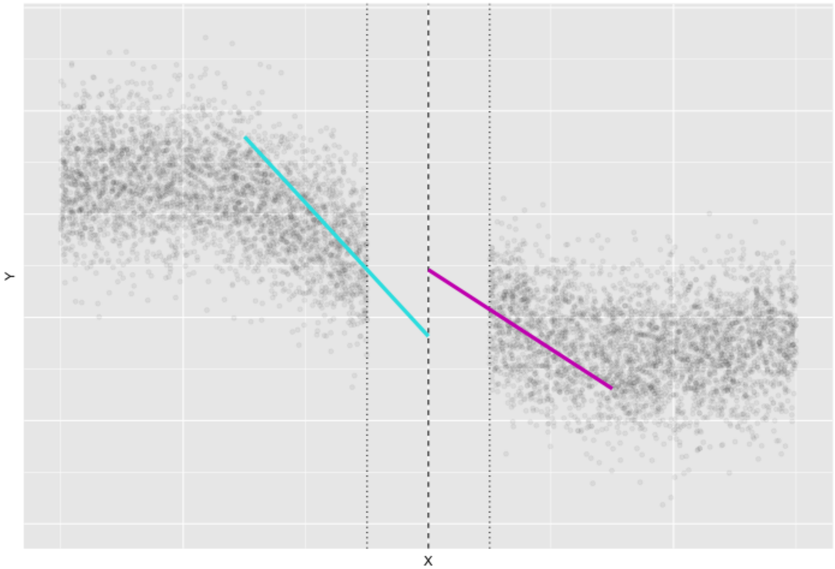
RD Donut Example



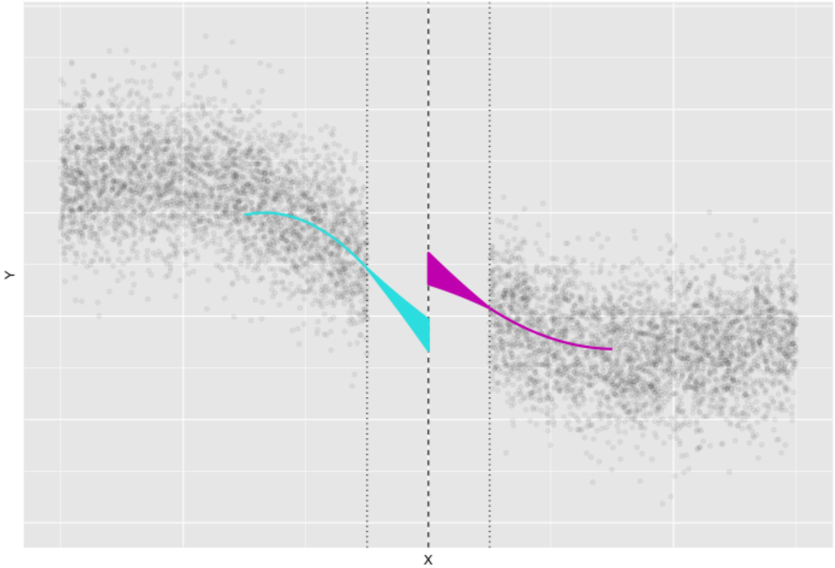
RD Donut Example



RD Donut Example



RD Donut Example



Motivating Question

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

Main result

Under:

1. natural extensions of standard assumptions,
2. known or data-determined derivative bounds,
3. and straightforward assumptions about selection

we get partial identification for causal effects – and validity while conducting inference for the partially identified set.

Outline

- Introduction to Shape Restrictions
- Assumptions/Conditions
- Results for ‘a priori‘ shape restrictions
- Different Confidence Intervals for Partial Identification

Outline

- Introduction to Shape Restrictions
- Assumptions/Conditions
- Results for ‘a priori‘ shape restrictions
- Different Confidence Intervals for Partial Identification
- How does data driven routine work?
- Combining Global restrictions with data
- Probation Example
- Future Directions

How Do Donuts Work?

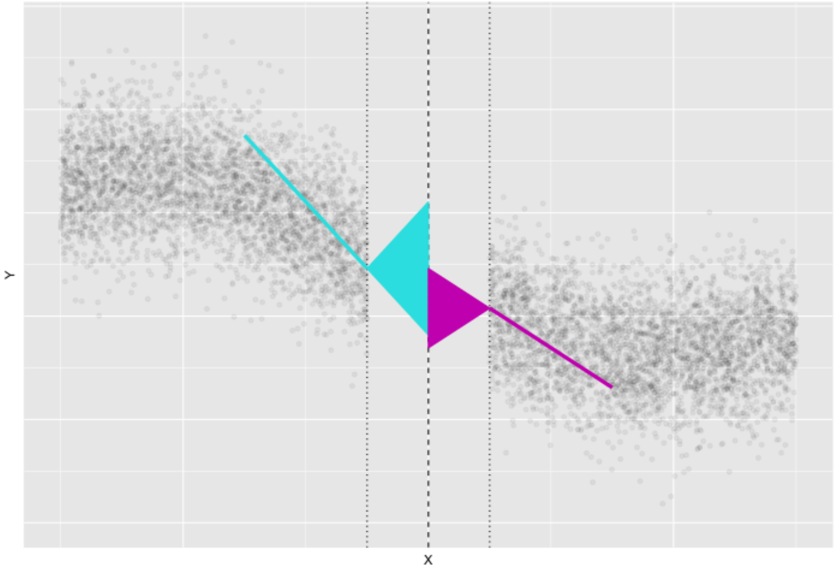
- There is no extant theory despite widespread empirical use.
- Most papers make implicit functional form assumptions.
- Even under those assumptions, more is needed than for standard RD.
- Can we use weaker restrictions on DGP?

This paper...

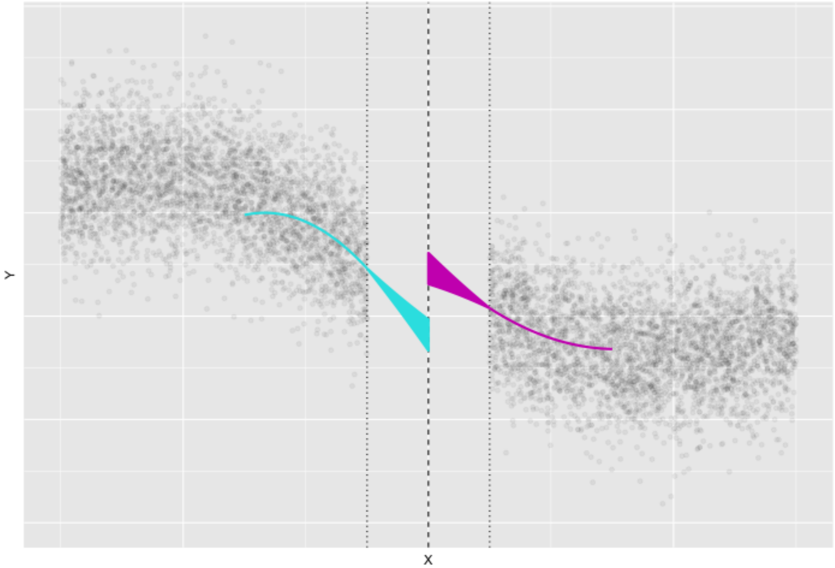
I use smoothness assumptions, which are natural to the RD setting, to perform inference with a donut.

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

Smoothness Example



Smoothness Example



Notation

1. We observe (Y,X,T)
2. $X \in \chi \subset \mathbb{R}$
3. $Y(T) = \mu_t(X) + \epsilon_t$
4. $Var(\epsilon_t | X, T) = \sigma_t^2(X)$

Most of our conditions will focus on the mean functions μ_t.

Denote the threshold c, and donut $\mathbb{D} = (d_-, d_+)$.

Outline of Procedure

1. Set a confidence level α, and $\kappa < \alpha$.
2. Estimate the $k-1$ derivatives of μ_t at the edge of the donut.
3. Predict μ_t at c, using its first $k-1$ derivatives and a Taylor projection.
4. Estimate $\tau(x_0) = \mu_1(x_0) - \mu_0(x_0)$ and build a $1-\alpha+\kappa$ CI.

Outline of Procedure

1. Set a confidence level α, and $\kappa < \alpha$.
2. Estimate the $k-1$ derivatives of μ_t at the edge of the donut.
3. Predict μ_t at c, using its first $k-1$ derivatives and a Taylor projection.
4. Estimate $\tau(x_0) = \mu_1(x_0) - \mu_0(x_0)$ and build a $1-\alpha+\kappa$ CI.
5. Find a set $\mathbb{C}_t$ that contains the $\mu_t^{(k)}$ with probability $1-\kappa/2$
6. Use the extreme values of $\mathbb{C}_t$ to find the maximal errors in the Taylor projection above.
7. Add those maximal errors for each side to the $1-\alpha+\kappa$ CI for τ.

Derivative Bounds

Condition 1: Derivative Bounds Exist

There is a known $k > 0$ such that

1. $\mu_t^{(k)}(x) \in [l_t, u_t] \; \forall x \in \chi$

Derivative Bounds

Condition 1: Derivative Bounds Exist

There is a known $k > 0$ such that

- (i) $\mu_t^{(k)}(x) \in [l_t, u_t] \; \forall x \in \chi$

For data-driven bounds, we also need to attain the bounds somewhere:

1. $\mu_t^{(k)}(x) = l_t$ for some $x \in \chi / \mathbb{D}$
2. $\mu_t^{(k)}(x) = u_t$ for some $x \in \chi / \mathbb{D}$

Derivative Bounds

Condition 1: Derivative Bounds Exist

There is a known $k > 0$ such that

(i) $\mu_t^{(k)}(x) \in [l_t, u_t] \ \forall x \in \chi$

For data-driven bounds, we also need to attain the bounds somewhere:

1. $\mu_t^{(k)}(x) = l_t$ for some $x \in \chi/\mathbb{D}$
2. $\mu_t^{(k)}(x) = u_t$ for some $x \in \chi/\mathbb{D}$

We don't need to know l_t or u_t, but we need to be able to estimate them 'well'.

Notice that this condition does not allow other treatment policies with a discontinuity in χ which affects Y.

Regularity Conditions

Condition 2: Regularity

1. (Y,X,T) are i.i.d. from a DGP as described above.
2. $\mu_t^{(k+2)}$ is continuous

Regularity Conditions

Condition 2: Regularity

1. (Y,X,T) are i.i.d. from a DGP as described above.
2. $\mu_t^{(k+2)}$ is continuous
3. The density of X, f_x is absolutely continuous and bounded away from zero over the region of interest χ.
4. $\sigma_t^2()$ is positive, bounded away from 0, and has two continuous derivatives.

Regularity Conditions

Condition 2: Regularity

1. (Y,X,T) are i.i.d. from a DGP as described above.
2. $\mu_t^{(k+2)}$ is continuous
3. The density of X, f_x is absolutely continuous and bounded away from zero over the region of interest χ.
4. $\sigma_t^2()$ is positive, bounded away from 0, and has two continuous derivatives.
5. $\underset{x \in \chi}{sup}\ \mathbb{E}\left[|\epsilon_i|^3 exp(|\epsilon_i|)|x_i = x\right] < \infty$

   which implies $\mathbb{E}\left[|\epsilon_i|^3 exp(|\epsilon_i|)\right] < \infty$.

Partial Identification

Lemma 1

Under conditions 1-2, 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 1-2, there is some set $\phi = [\tau_l, \tau_u]$ such that

(a) $\tau \in \phi$

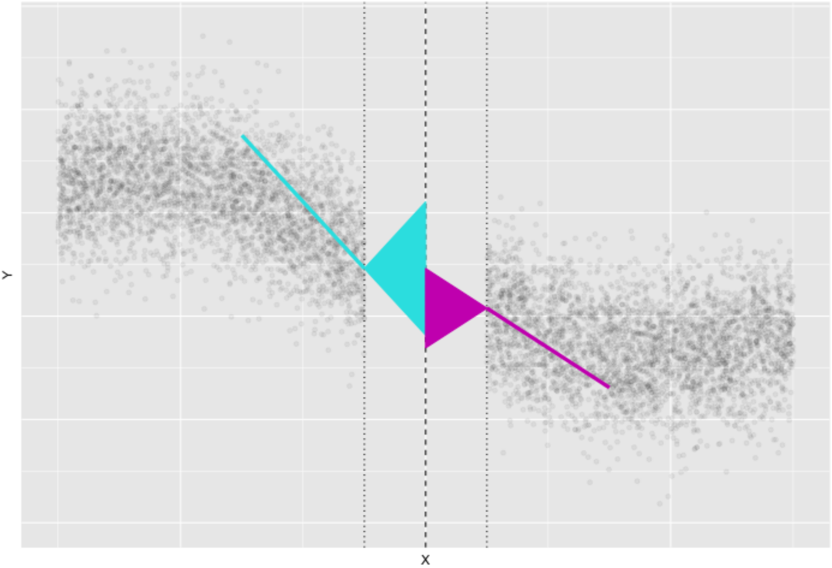
(b) $\tau_u - \tau_l < \infty$

Concretely:

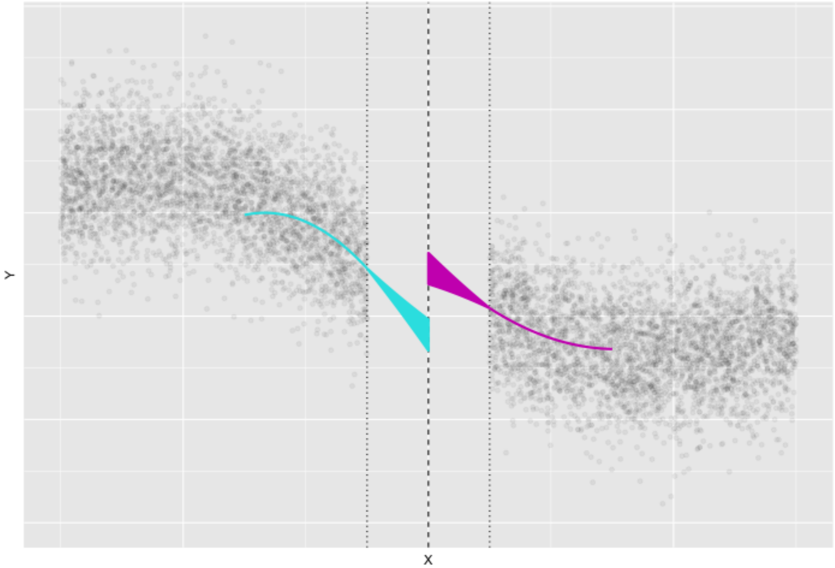
$$\mu_{0,l}(0) = \left(\sum_{j=0}^{k-1} \frac{d_-^j}{j!} \mu_0^{(j)}(d_-) \right) + \frac{d_-^k}{k!} l_0$$

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

Derivative Bound Example: $k = 1$



Derivative Bound Example $k = 2$



Local Polynomial Conditions

Condition 3: Kernel and Bandwidth for Local Polynomial

1. The kernel function $K(\cdot)$ has support $(-1, 1)$, outside of which it takes value 0.
2. $K(\cdot)$ is symmetric, positive, bounded, and integrates to 1 over its support.
3. The bandwidth $h = h_n$ is set such that as $n \to \infty$, $h_n \to 0$ and $nh_n^3 \to \infty$.
4. $\exists \quad \eta \geq h_n \quad \forall n$.

Local Polynomial Conditions

Condition 3: Kernel and Bandwidth for Local Polynomial

1. The kernel function $K(\cdot)$ has support $(-1, 1)$, outside of which it takes value 0.
2. $K(\cdot)$ is symmetric, positive, bounded, and integrates to 1 over its support.
3. The bandwidth $h = h_n$ is set such that as $n \to \infty$, $h_n \to 0$ and $nh_n^3 \to \infty$.
4. $\exists \quad \eta \geq h_n \quad \forall n$.

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].

Donut Exclusion

Condition D: Donut Exclusion

1. 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.
2. 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 D: Donut Exclusion

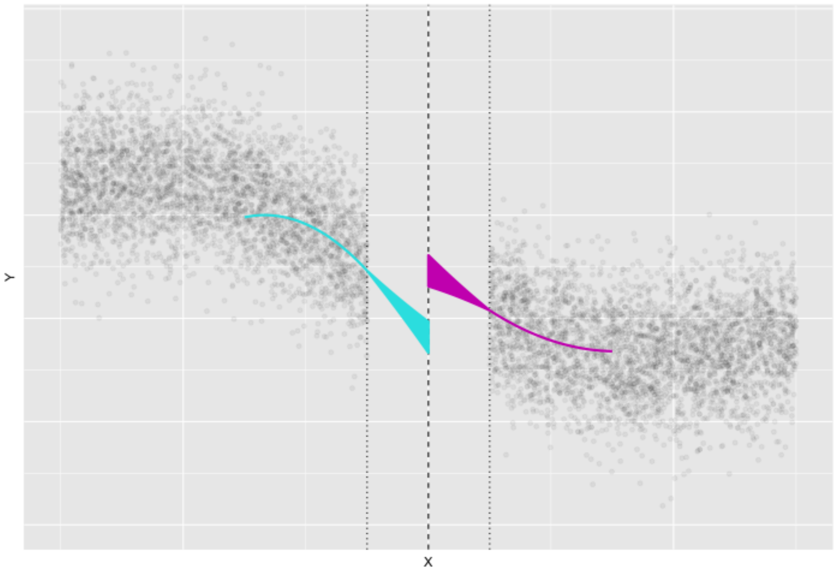
1. There is a known interval $\mathbb{D} = (d_-, d_+)$ such that all individuals who manipulate are contained to the interval, and would be contained in the counterfactual where they do not manipulate.
2. There is only one policy with a threshold relevant to the outcome variable inside the region $[d_- - \epsilon, d_+ + \epsilon]$ 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



Normality

Lemma 2

Under conditions 1-3, 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}

Outline of Procedure

1. Set a confidence level α, and $\kappa < \alpha$.
2. Estimate the $k-1$ derivatives of μ_t at the edge of the donut.
3. Predict μ_t at c, using its first $k-1$ derivatives and a Taylor projection.
4. Estimate $\tau(x_0) = \mu_1(x_0) - \mu_0(x_0)$ and build a $1-\alpha+\kappa$ CI.

Outline of Procedure

1. Set a confidence level α, and $\kappa < \alpha$.
2. Estimate the $k-1$ derivatives of μ_t at the edge of the donut.
3. Predict μ_t at c, using its first $k-1$ derivatives and a Taylor projection.
4. Estimate $\tau(x_0) = \mu_1(x_0) - \mu_0(x_0)$ and build a $1-\alpha+\kappa$ CI.
5. Find a set $\mathbb{C}_t$ that contains the $\mu_t^{(k)}$ with probability $1-\kappa/2$
6. Use the extreme values of $\mathbb{C}_t$ to find the maximal errors in the Taylor projection above.
7. Add those maximal errors for each side to the $1-\alpha+\kappa$ CI for τ.

Comments

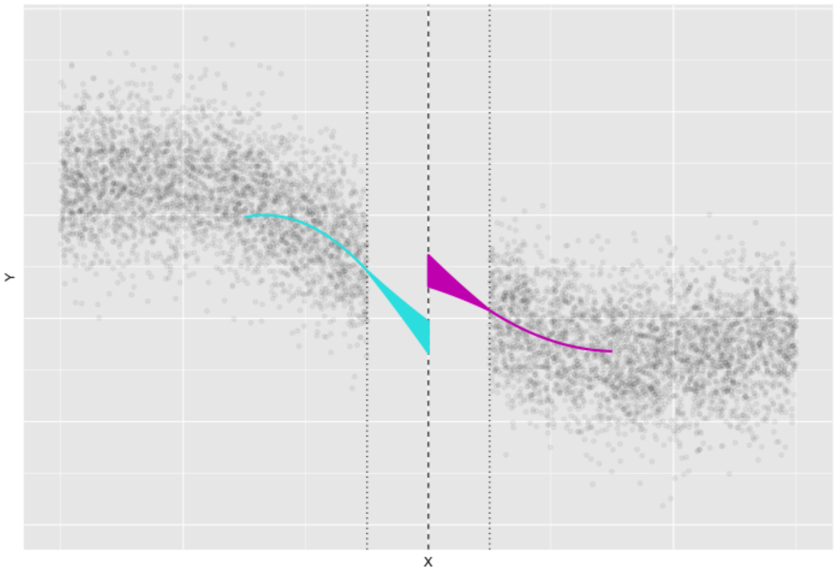
- Moving forwards, I focus on inference for the region ϕ.
- In examples and discussion, I will use bounds of the form

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

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

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

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 $\alpha \in (0, 1/2)$,

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

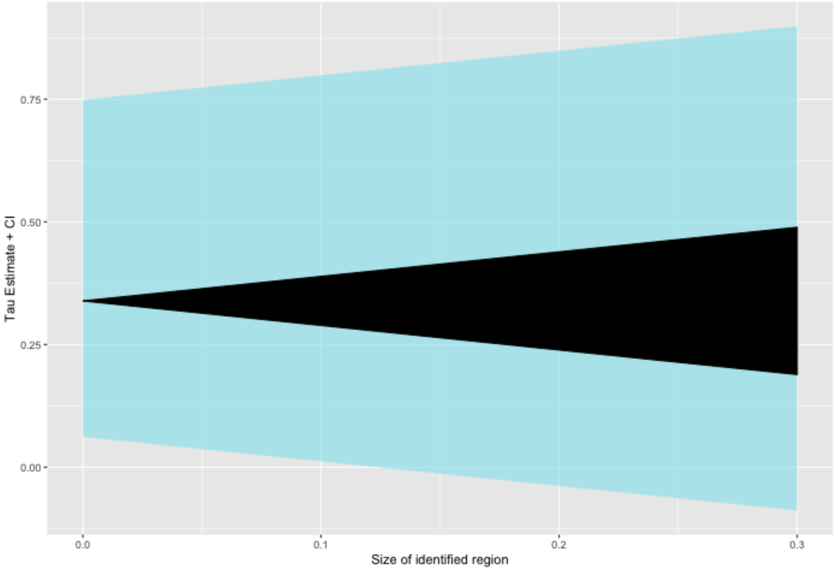
Comments on Theorem 1

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

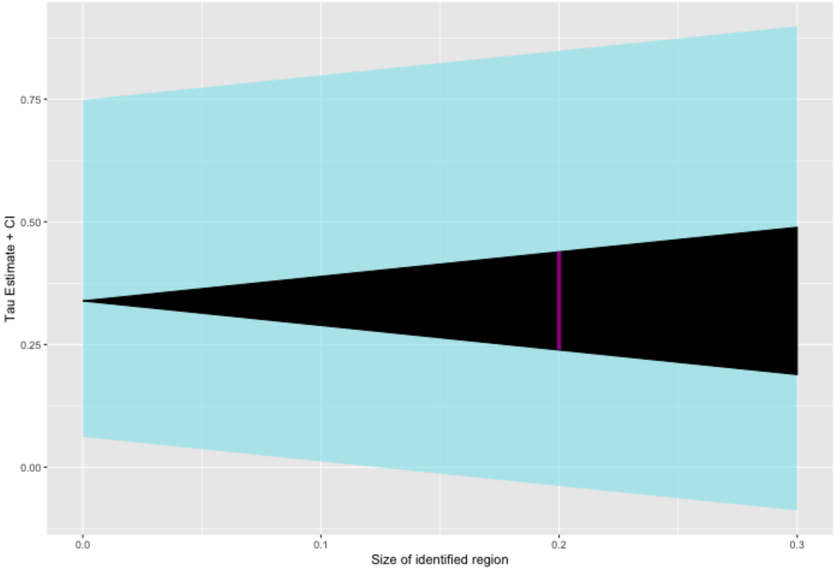
Theorem 1 is dependent on the two bandwidth conditions:
$nh^3 \to \infty$ and $nh^{2k+3} \to 0$.

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.

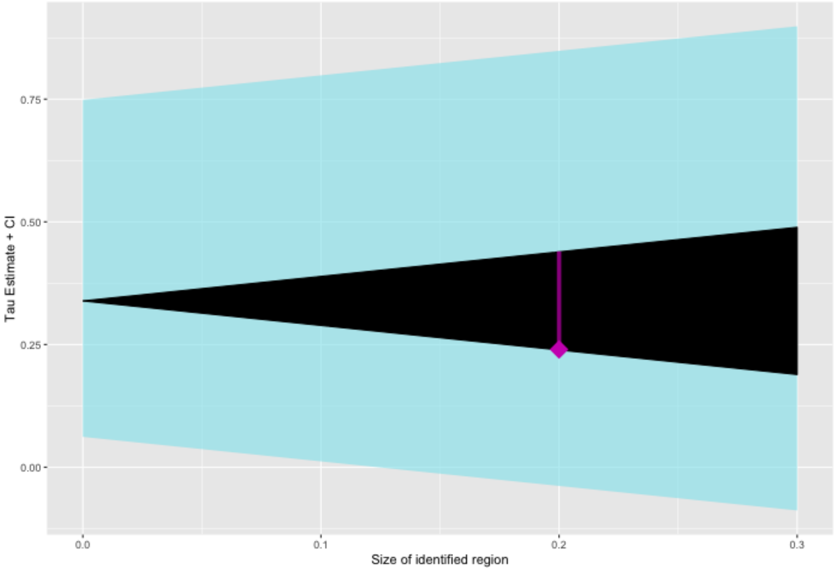
$\mathbb{S}_{1-\alpha}$ Example



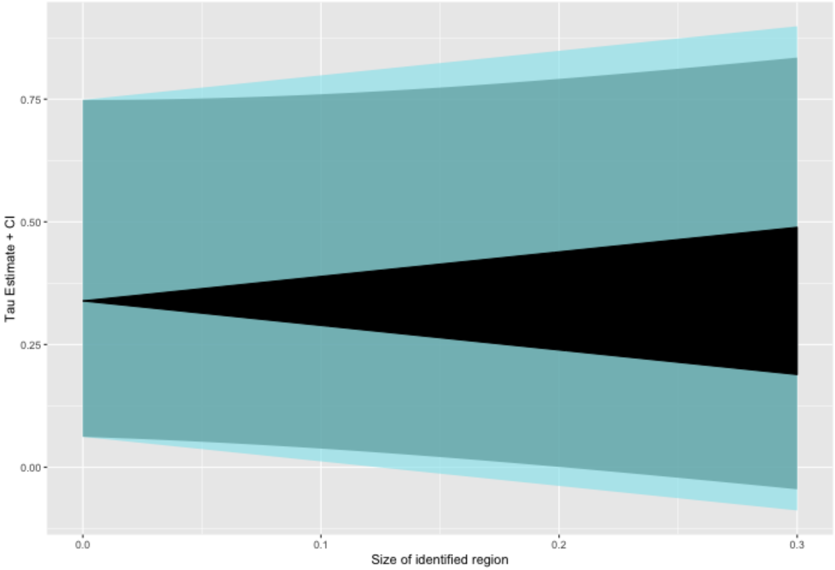
$\mathbb{S}_{1-\alpha}$ Example



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



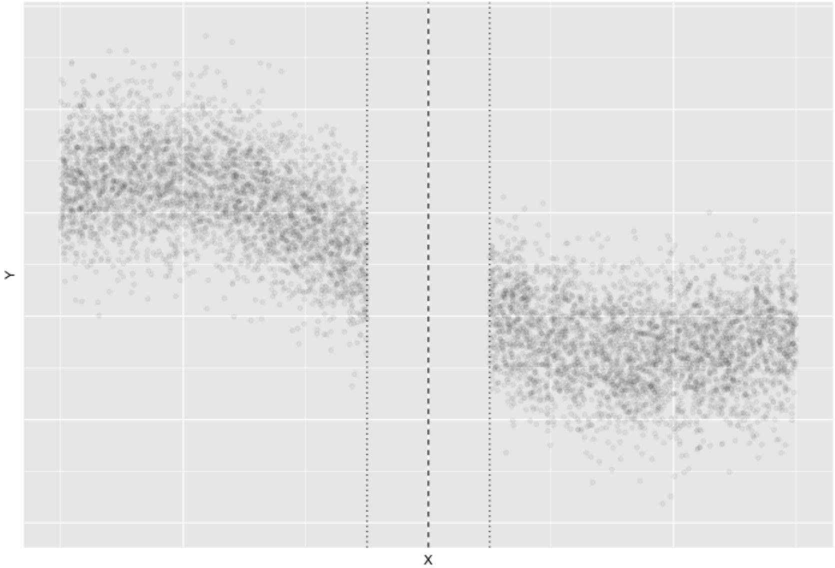
$\mathbb{Q}_{1-\alpha}$ Example



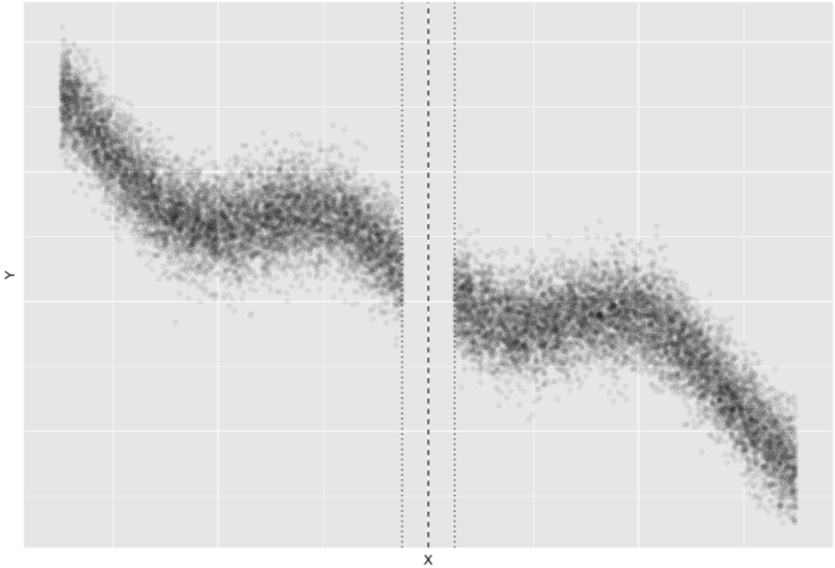
Global Smoothness

What if we want data driven smoothness conditions?

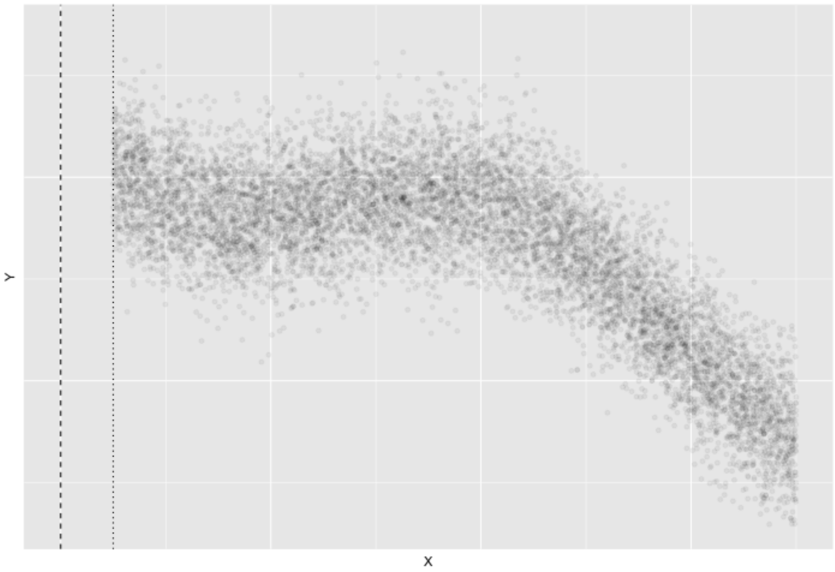
How can we make statements about derivative extrema?



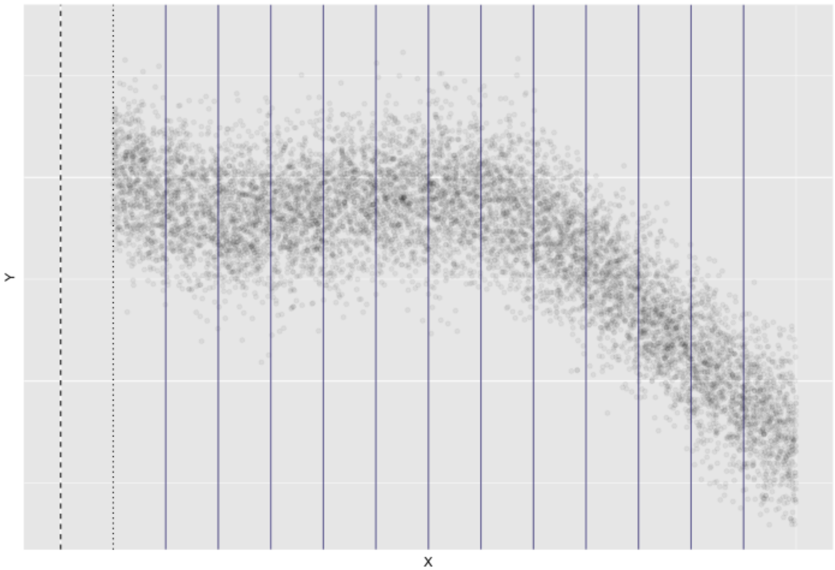
How can we make statements about derivative extrema?



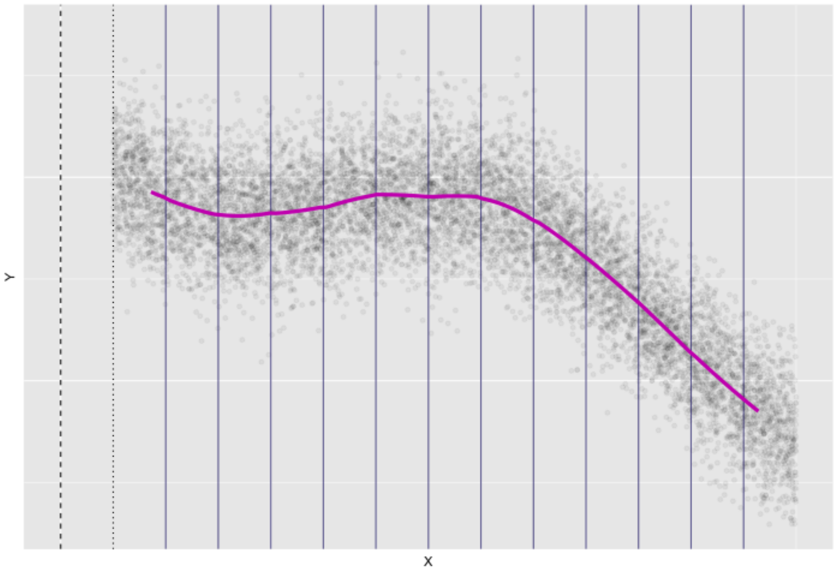
How can we make statements about derivative extrema?



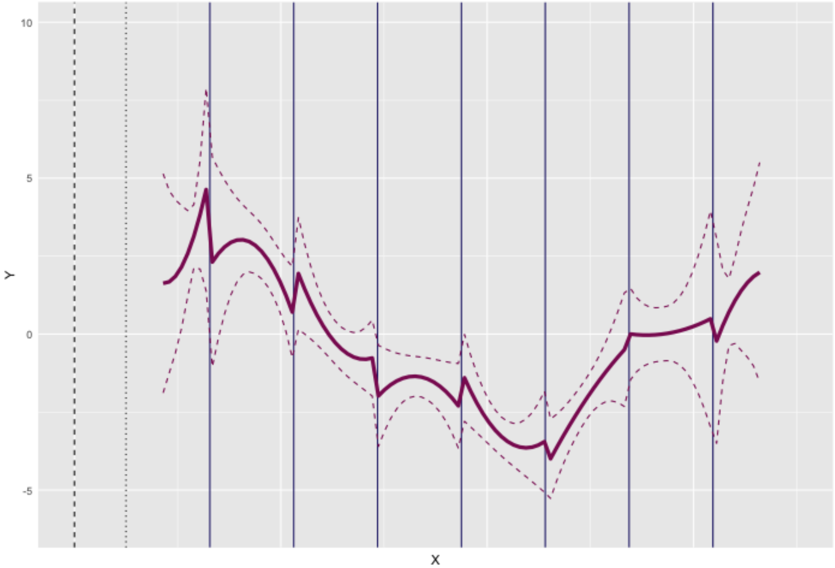
How can we make statements about derivative extrema?



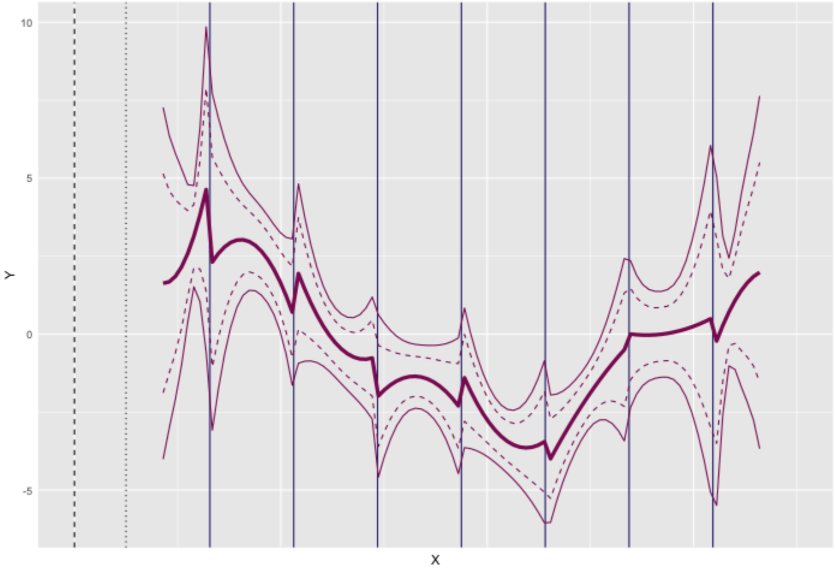
How can we make statements about derivative extrema?



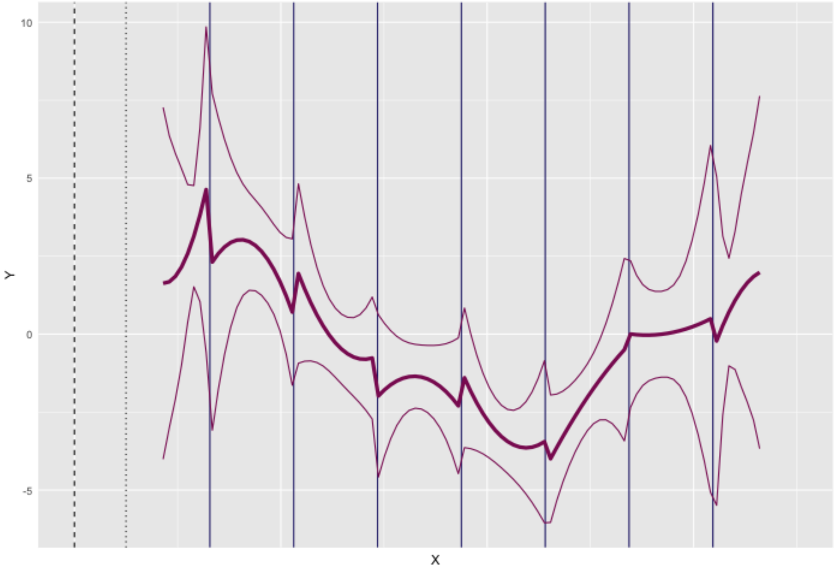
How can we make statements about derivative extrema?



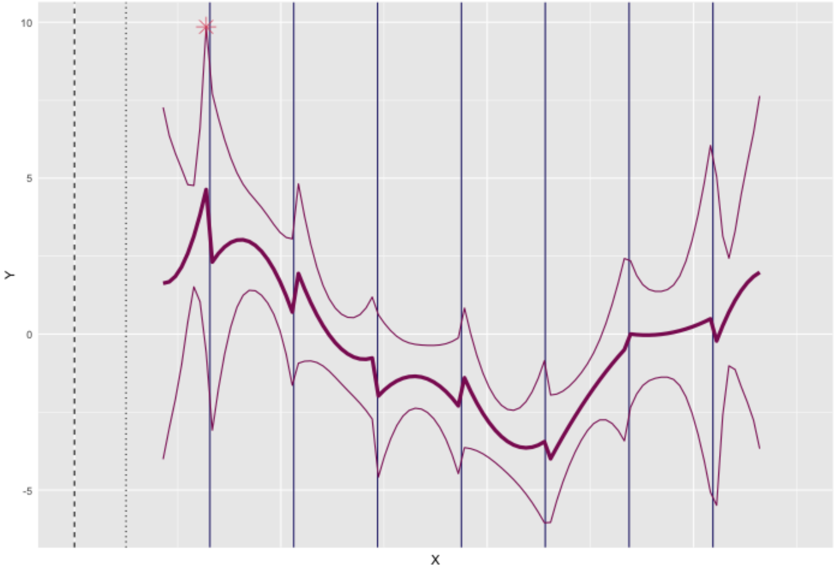
How can we make statements about derivative extrema?



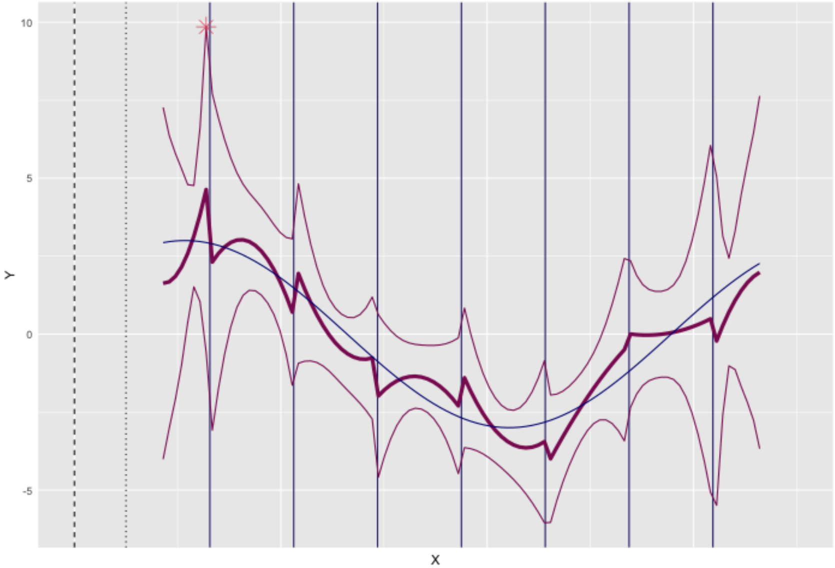
How can we make statements about derivative extrema?



How can we make statements about derivative extrema?



How can we make statements about derivative extrema?



Global Derivative Conditions

Condition 4: Bandwidth Conditions for Series Estimator

1. $\frac{log(n)^{3/2}}{\sqrt{nh_b}} = o_{\mathbb{P}}(1/log(n))$
2. $\frac{log(n)^4}{nh_b} = o(1/log(n))$
3. $nh_b^{1+2k} = o(1/log(n))$

These rate conditions apply to the *different* bandwidth controlling our inferential routine for the global smoothness.

Global Results

Theorem 3

Under conditions 1-4, using local polynomials to learn the 0,...,k-1 derivatives at 0 and using b-splines to learn the *sup* and *inf* of the kth derivative, we can build a $1-\alpha$ confidence region CR_g such that:

$$lim\Phi_t(x_0) \subset CR_g \geq 1-\alpha$$

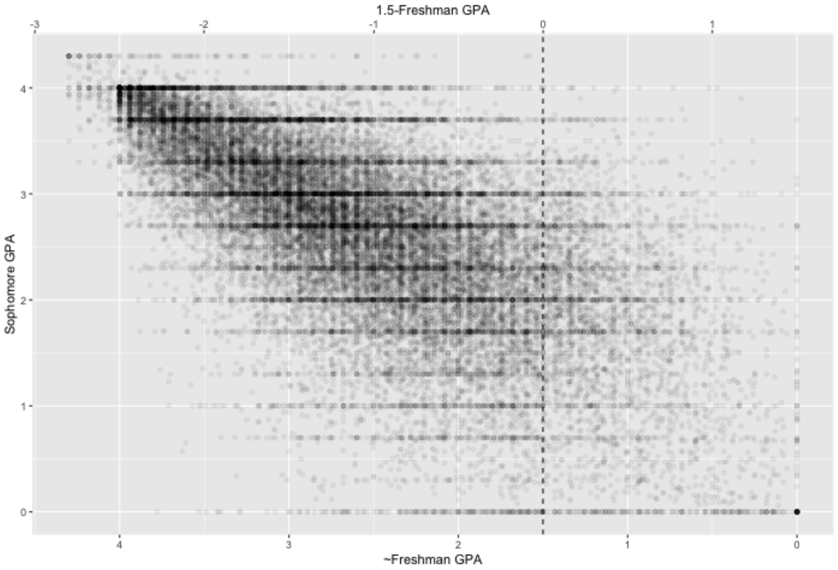
Academic Probation

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 – Data

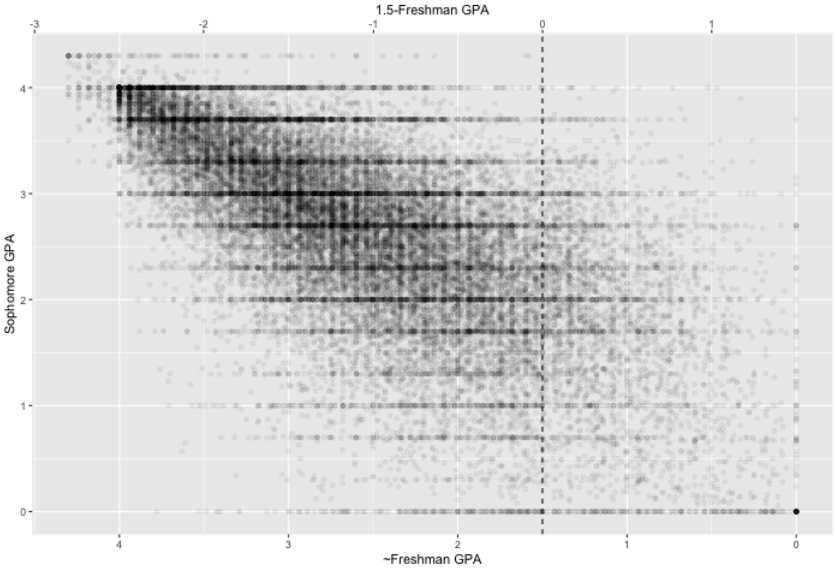


Academic Probation - Methods

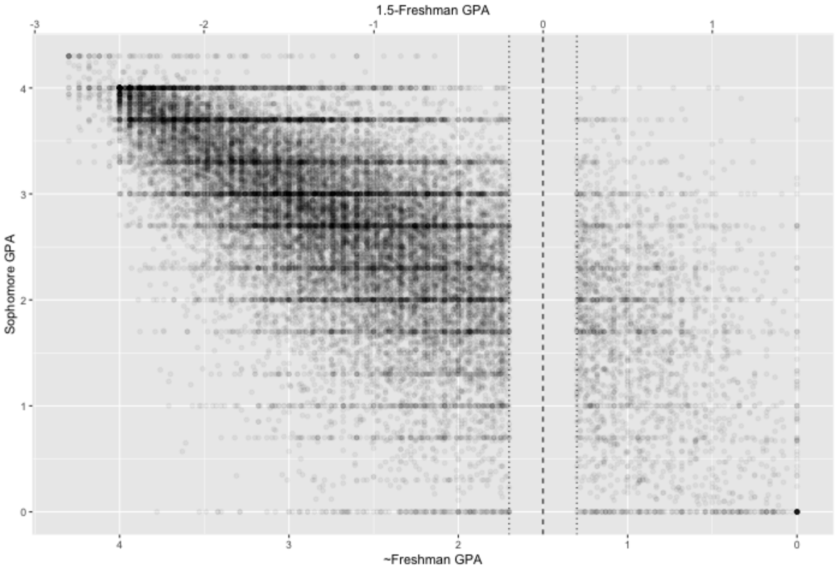
If 1/3 of grades are high enough that professors raise them one partial letter (e.g. C+ to B-) on being asked, that is a maximum GPA change of 0.2.

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

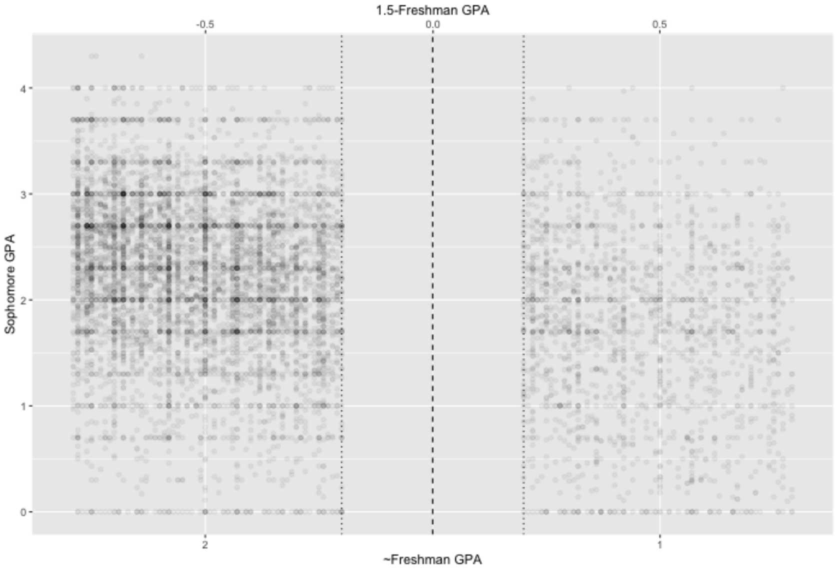
Academic Probation - All Data



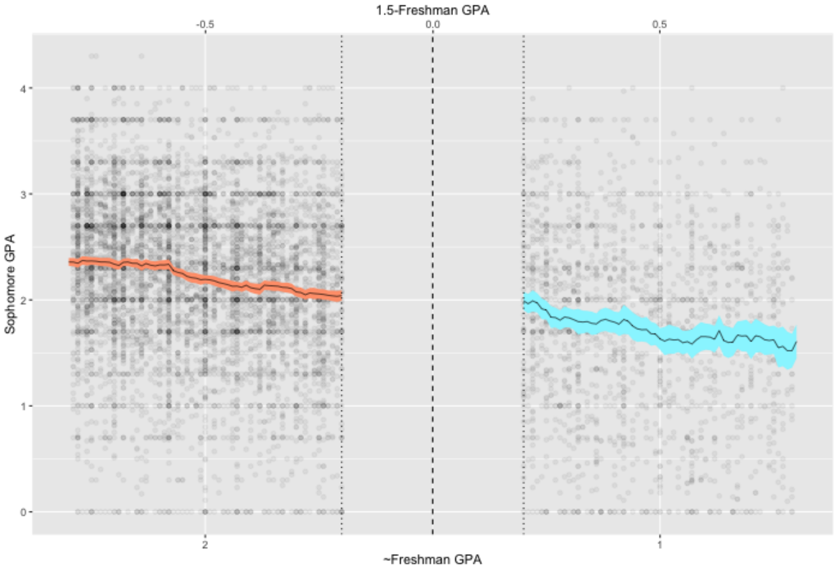
Academic Probation - Drop Donut



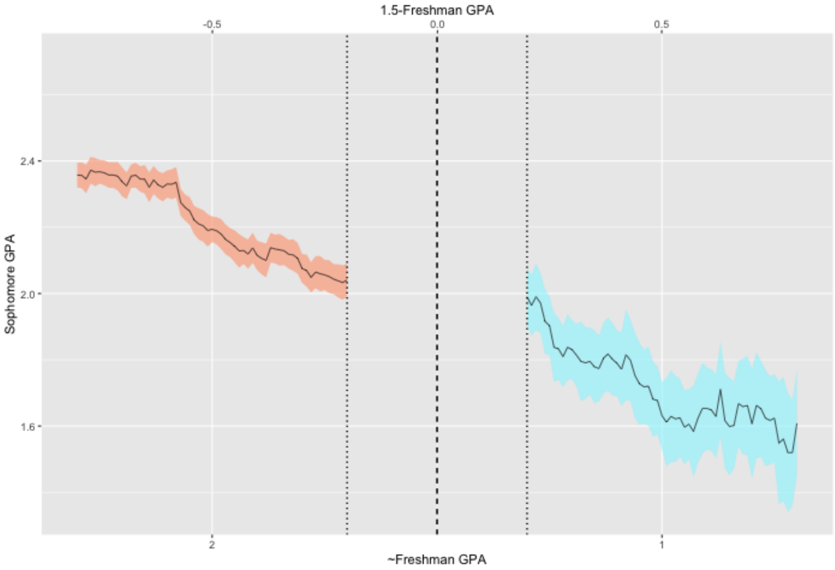
Academic Probation - Inside Bandwidth



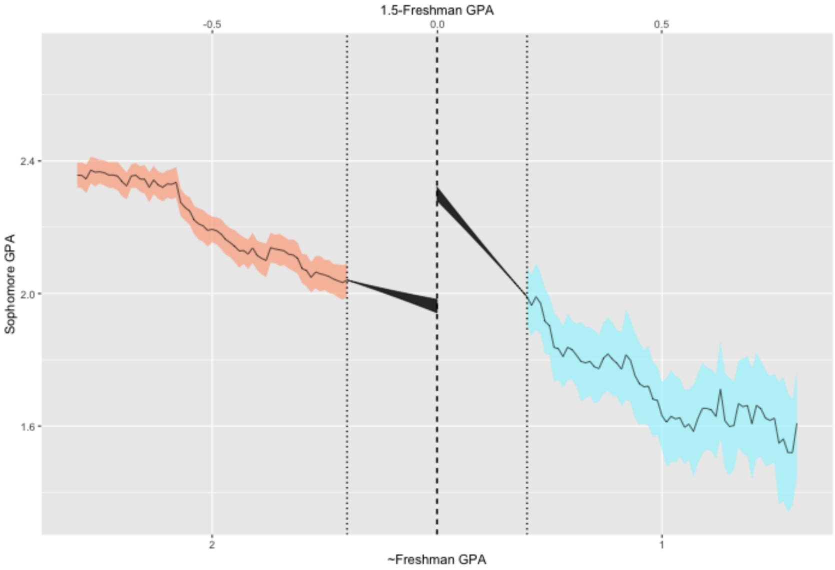
Academic Probation - Fit Local Polynomials



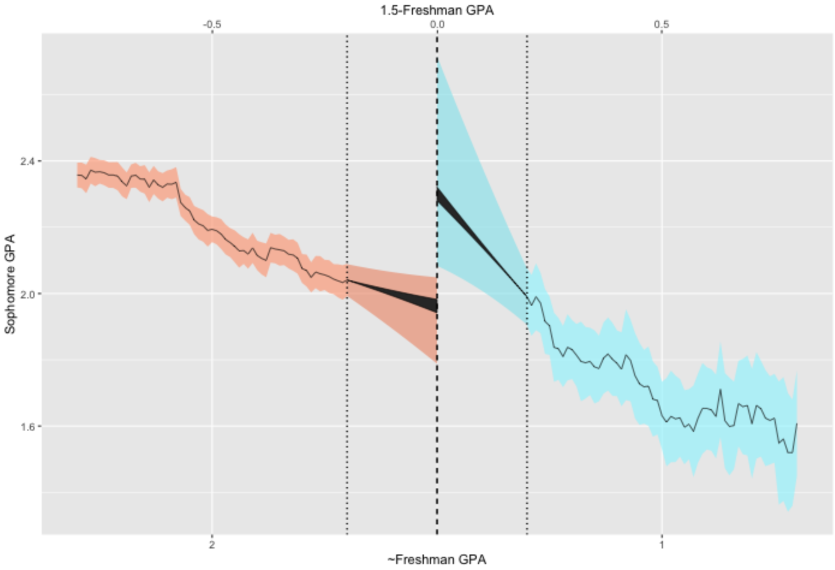
Academic Probation - Fit Local Polynomials



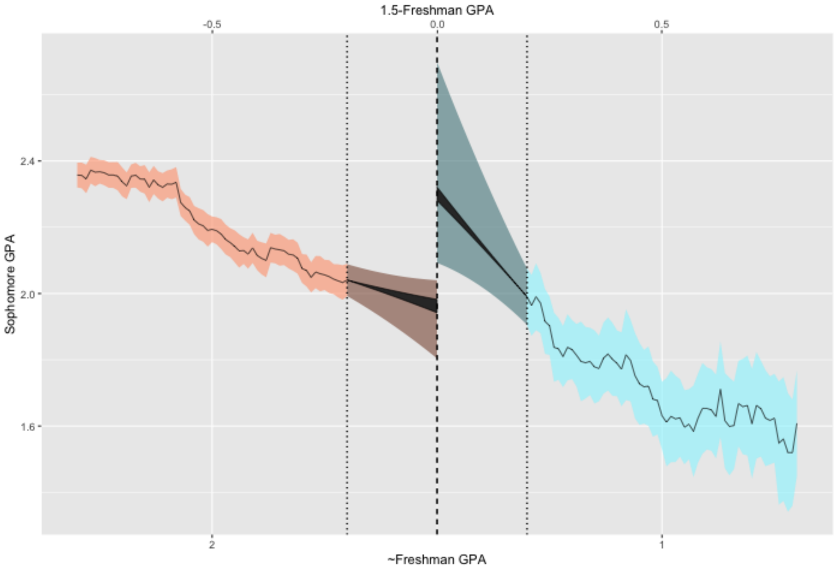
Academic Probation - Identified Region



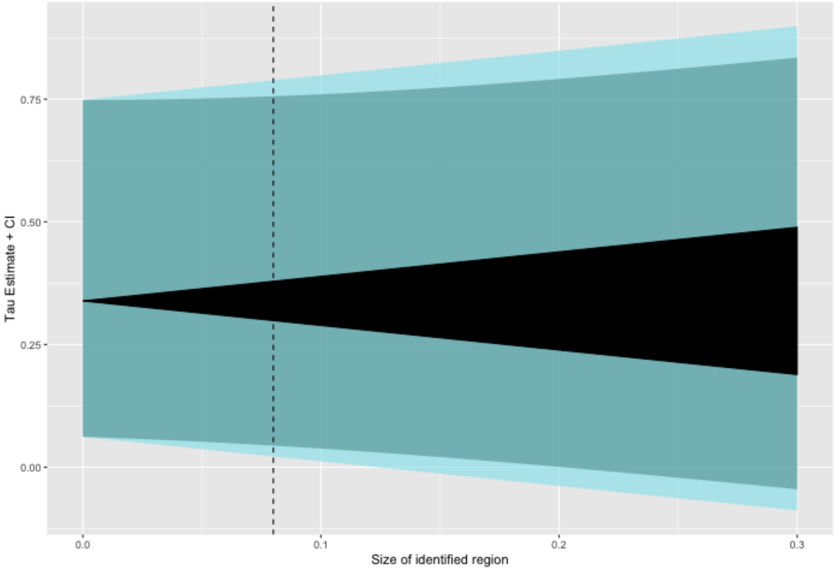
Academic Probation - CR for Set



Academic Probation - CR for elements of set



Tau Set



Tau Estimate + CI
0.75
0.50
0.25
0.00
0.0
0.1
0.2
0.3
Size of identified region

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