Essays in Econometrics

Connor Dowd

University of Chicago Booth School of Business

Dissertation Defense [1]
April 19, 2021

[1] Committee: Chris Hansen (chair), Max Farrell, Panos Toulis, Constantine Yannelis

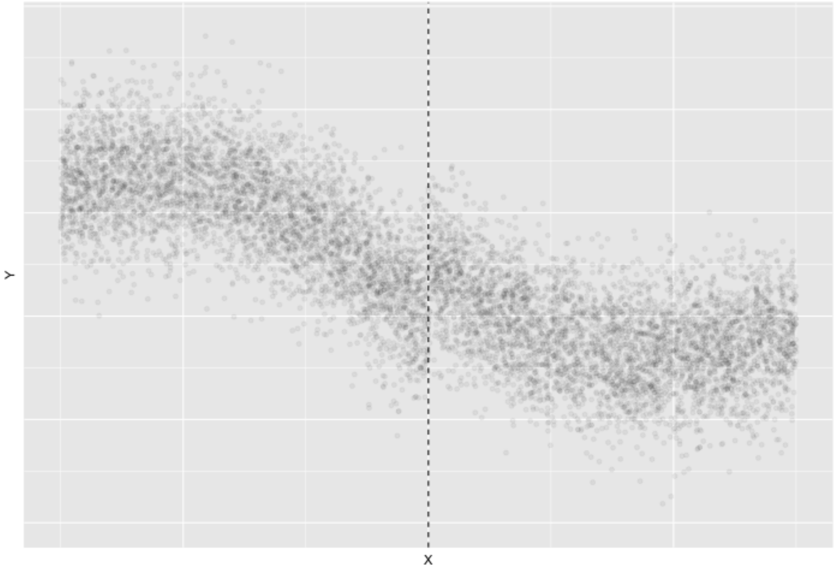
Outline

3 Essays

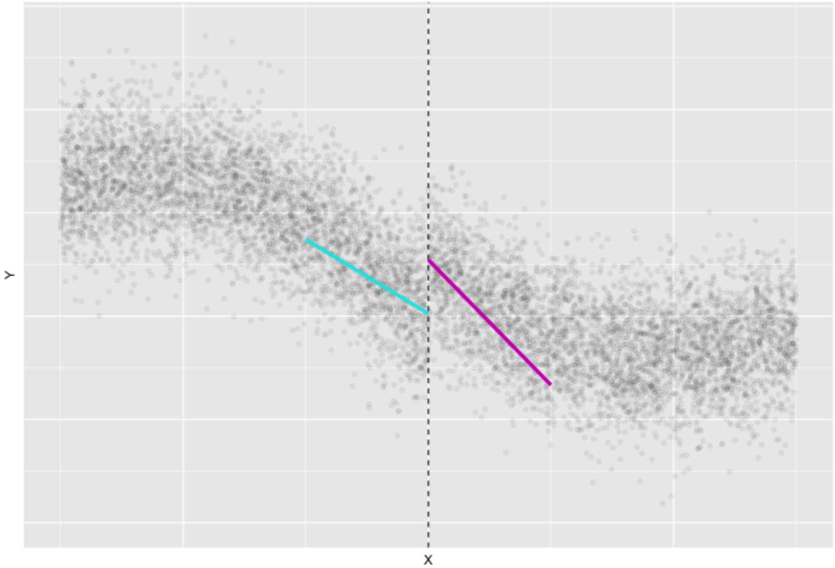
1. Regression Discontinuity Donuts:
2. Synthetic Controls with Spillovers: Examples and Simulations
3. ECDF two-sample package

Regression Discontinuity Donuts

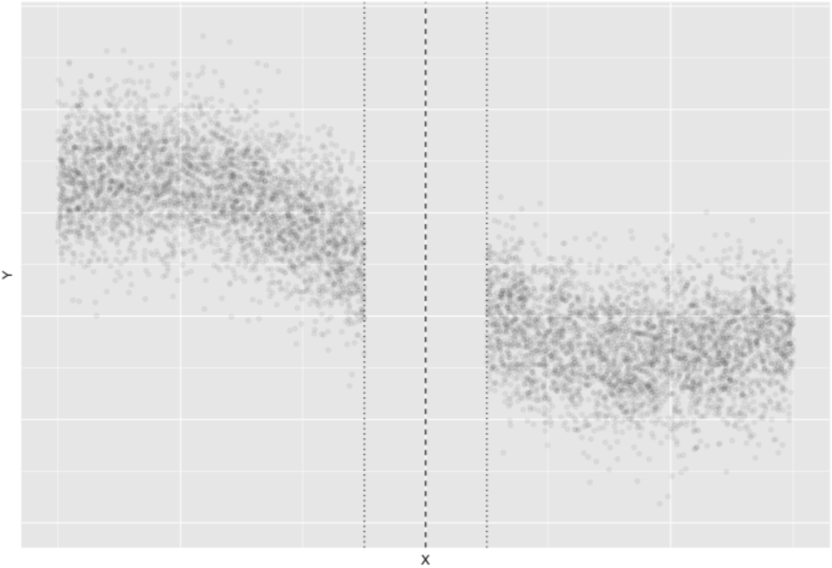
RD Example



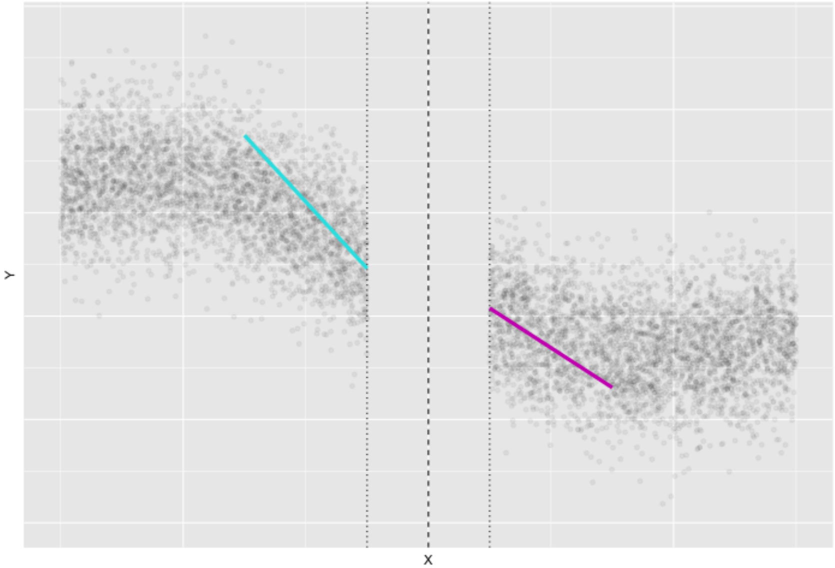
RD Example



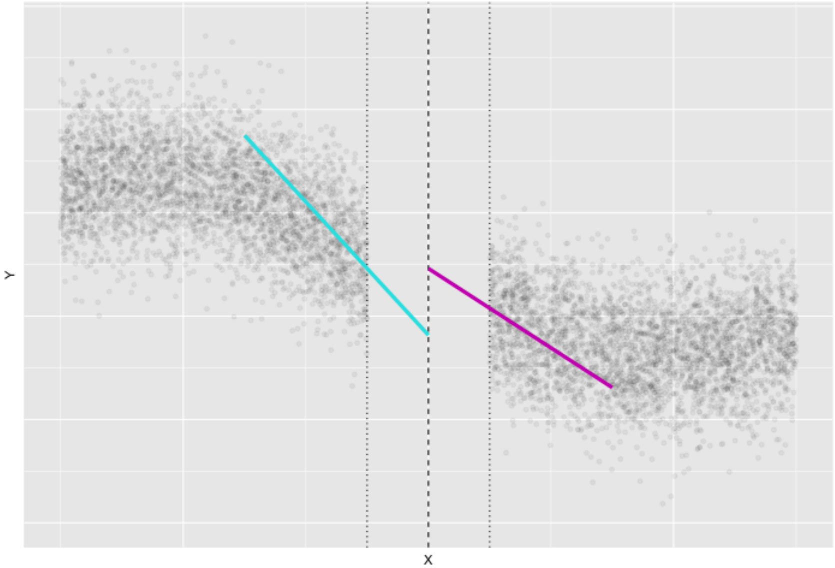
RD Donut Example



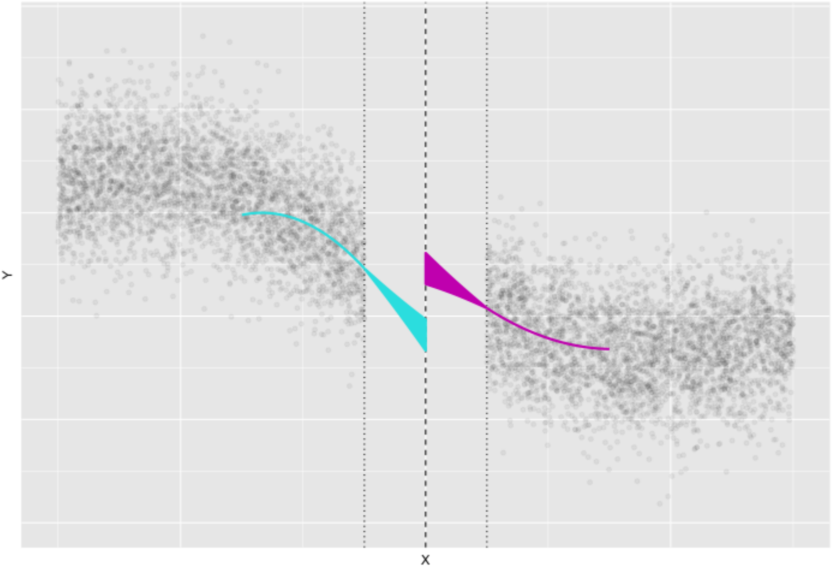
RD Donut Example



RD Donut Example



RD Donut Example



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

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. $\exists \quad \eta \geq h_n \ \ \forall n.$

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.

Coverage for ϕ

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

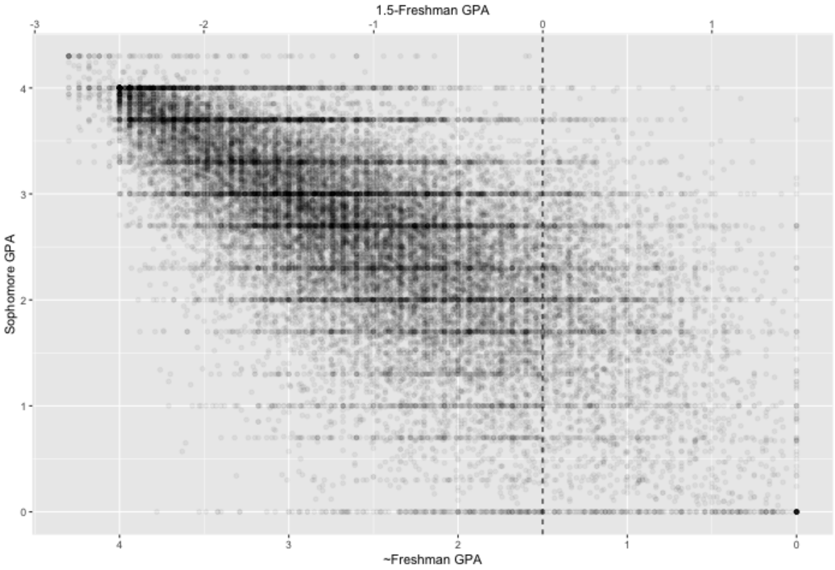
$$\mathbb{S}_{1-\alpha} = \left[\hat{\tau}_l - C\hat{\sigma}_l/\sqrt{n},\ \hat{\tau}_u + C\hat{\sigma}_u/\sqrt{n}\right]$$

Theorem 1

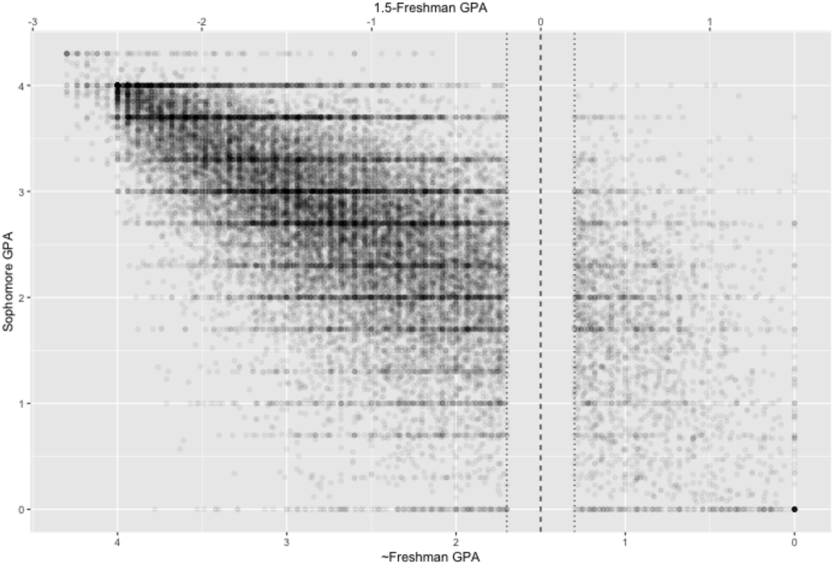
Under conditions 1-4, and the condition that $nh^{2k+3} \to 0$, for all $\alpha \in (0, 1/2)$,

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

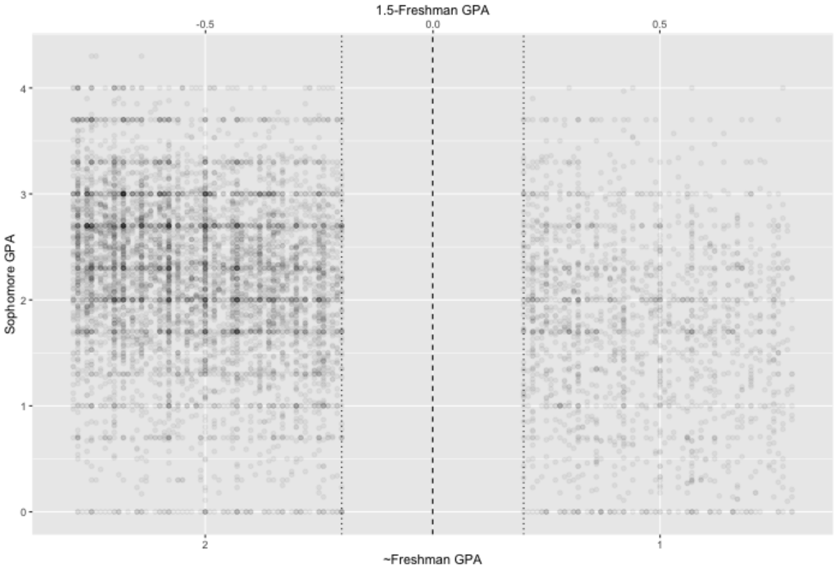
Academic Probation - All Data



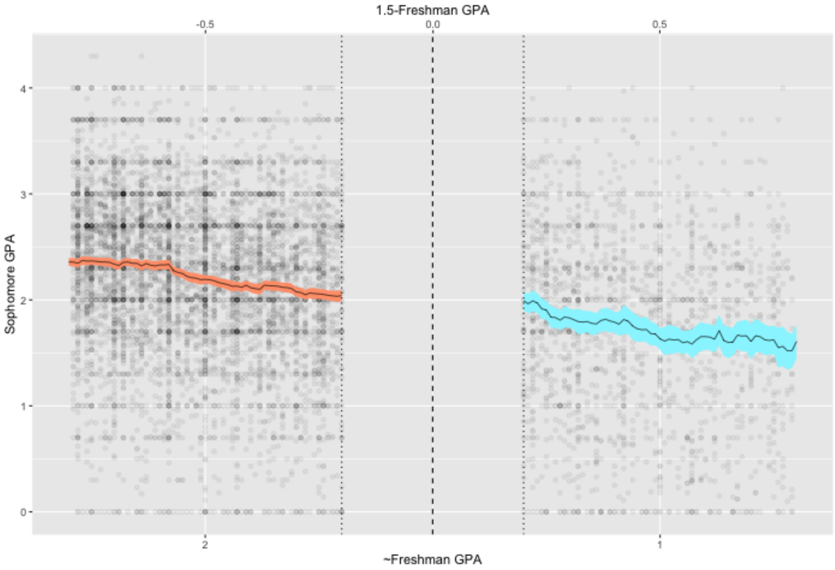
Academic Probation - Drop Donut



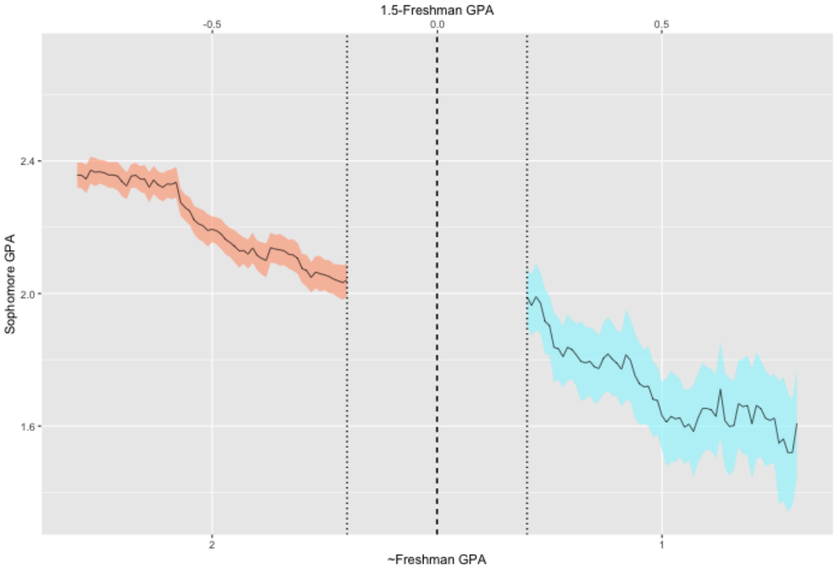
Academic Probation - Inside Bandwidth



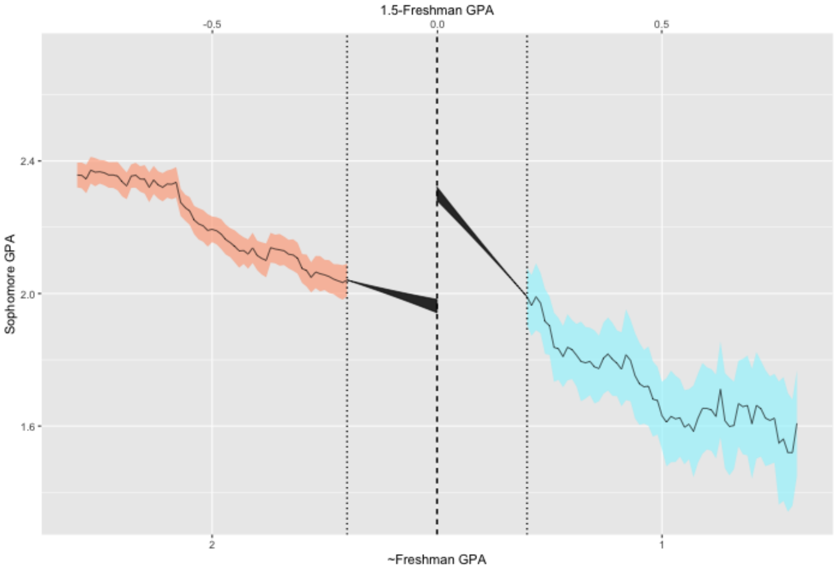
Academic Probation - Fit Local Polynomials



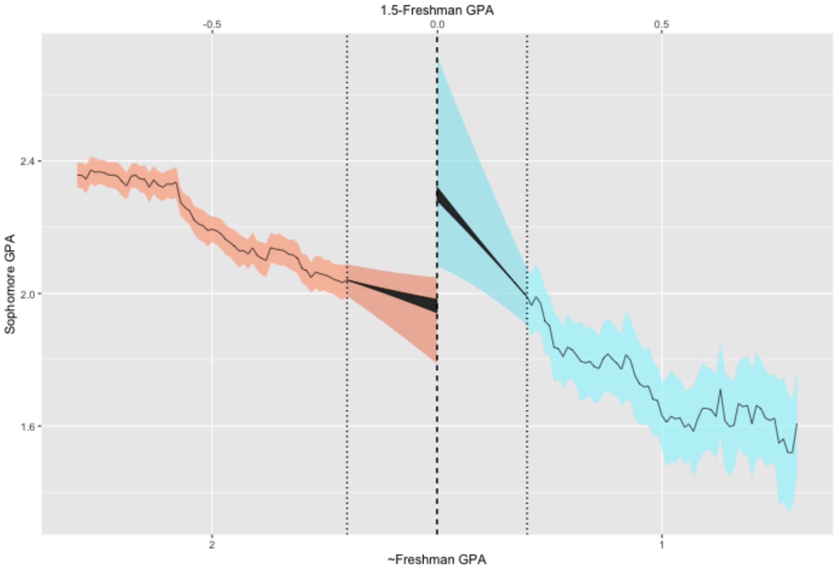
Academic Probation - Fit Local Polynomials



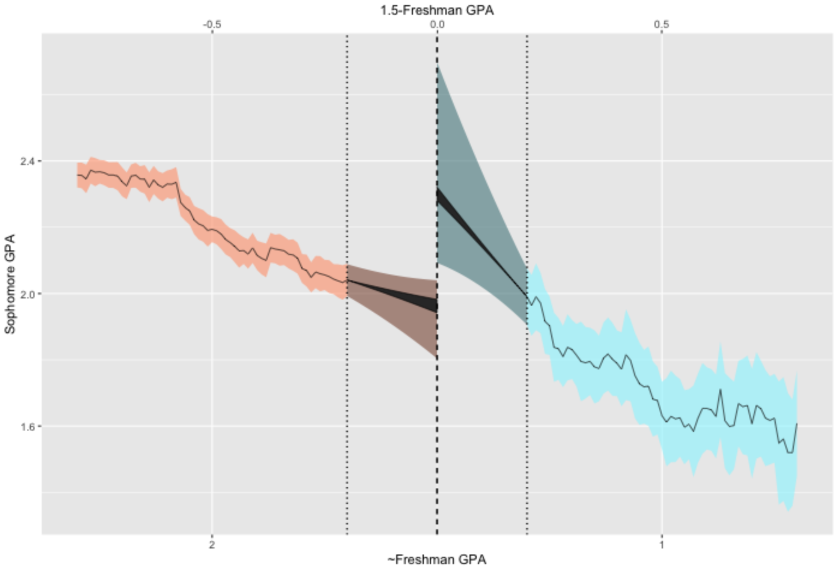
Academic Probation - Identified Region



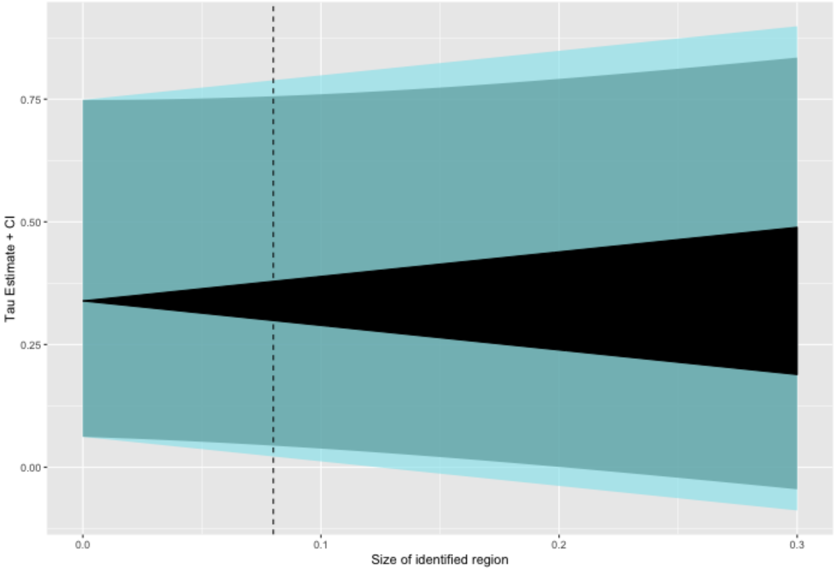
Academic Probation - CR for Set



Academic Probation - CR for elements of set



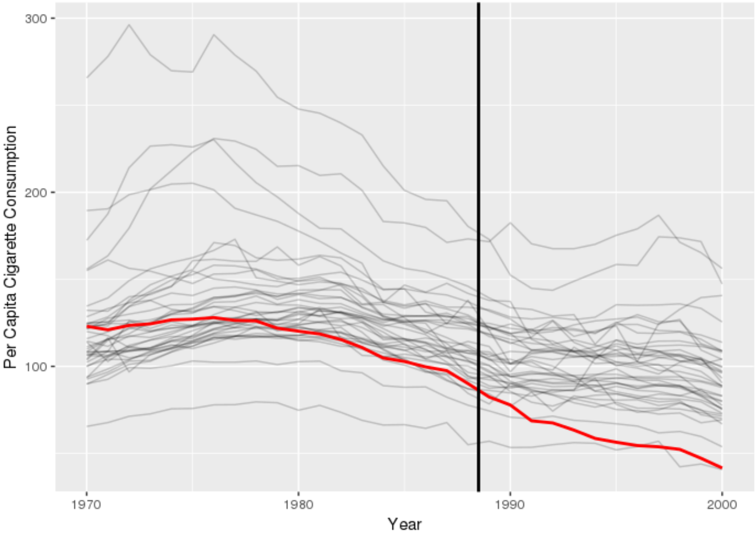
Tau Set



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

Synthetic Controls with Spillovers: Examples and Simulations

Example: California Prop 9



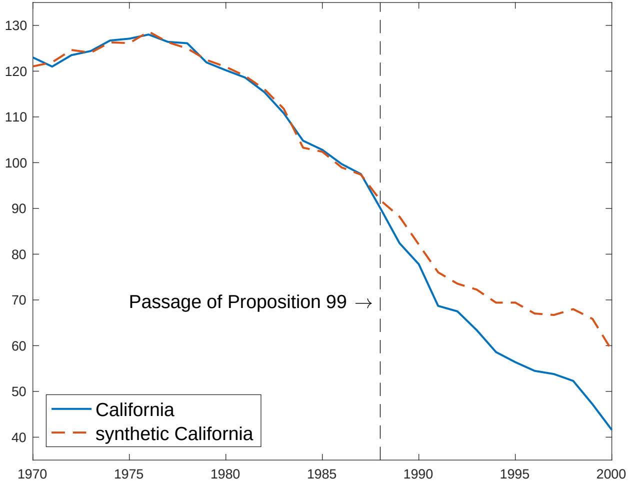
Synthetic Controls Setting

$$
\begin{array}{cccc|c}
y_{1,1} & y_{1,2} & \cdots & y_{1,T} & y_{1,T+1} \\
\hline
y_{2,1} & y_{2,2} & \cdots & y_{2,T} & y_{2,T+1} \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
y_{N,1} & y_{N,2} & \cdots & y_{N,T} & y_{N,T+1}
\end{array}
\begin{array}{l}
\} \text{ treated unit} \\
\left.\begin{array}{l} \\ \\ \\ \end{array}\right\} \text{control units}
\end{array}
$$

$\uparrow$ treatment

$$y_{1,T+1}(1) = y_{1,T+1}(0) + \alpha$$

Abadie, Diamond, and Hainmueller (2010)



Synthetic Control Estimator

Synthetic control weights:

$$\begin{pmatrix} \hat{z}_1 \\ \hat{b}_1 \end{pmatrix} = \underset{(z,b')' \in W}{\arg\min} \sum_{t=1}^{T} (y_{1,t} - z - Y_t' b)^2,$$

where $Y_t = (y_{1,t}, y_{2,t}, \ldots, y_{N,t})'$ and $W = \mathbb{R} \times \{0\} \times \Delta_{N-1}$

i.e. $\hat{b}_1$ are normalized weights – they sum to one, are non-negative, and we force a weight of 0 on the 'own' observation.

z is included because of the work in Ferman and Pinto (2017) showing an intercept is necessary for unbiased procedures.

Spillover effects in synthetic control

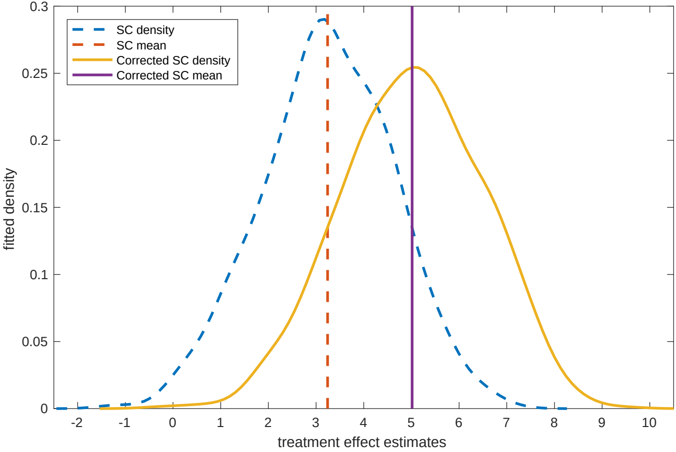
The synthetic control estimator

$$\hat{\alpha} = y_{1,T+1}(1) - \hat{y}_{1,T+1}(0) = y_{1,T+1} - Y'_{T+1}\hat{b}_1 - \hat{z}$$

can be severely biased in the presence of spillover effects:

- As in a diff-in-diff setting, if treatment can affect your control group - your ATE may be biased.
- SCM is particularly vulnerable as it may put extra weight on the same units that are 'contaminated'.
- This can be bad luck – but more worryingly, similar units may actually be prone to spillovers.
- Regularization properties of simplex are suddenly a potential downside.

True treatment effect = 5



Linear Spillover effects

Remember: $\alpha_i = y_{i,T+1}(1) - y_{i,T+1}(0)$
i.e. In the counterfactual, had treatment not occurred anywhere, how different was unit i?

Let

$$\alpha = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_N \end{bmatrix}$$

Assume linear spillover effects: $\alpha = A\gamma$

- A known
- γ unknown parameters

Synthetic control weights for all units

We estimate the weight vectors and intercepts for each unit.
For each i,

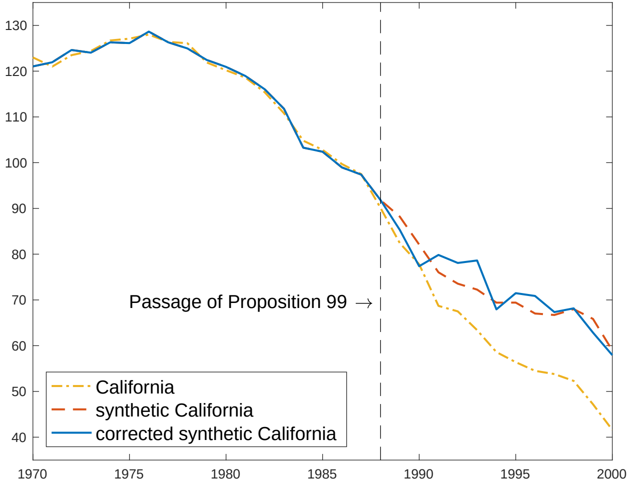
$$\begin{pmatrix} \hat{z}_i \\ \hat{b}_i \end{pmatrix} = \underset{(z,b)\in W_i}{\arg\min} \sum_{t=1}^{T} (y_{i,t} - z - Y_t b')^2,$$

with i-th entry of $\hat{b}_i$ being 0, and $\hat{b}_i \in \Delta_N$ (non-negative, sum to one).
$Y_t = (y_{1,t}, y_{2,t}, \ldots, y_{N,t})'$

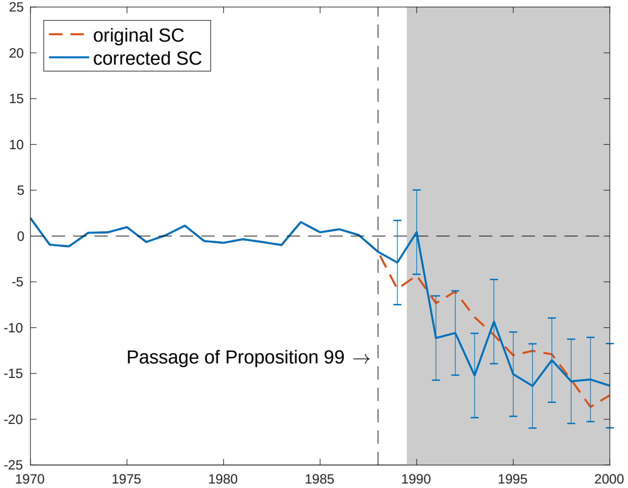
Define $z_i = \text{plim } \hat{z}_i$ and $b_i = \text{plim } \hat{b}_i$. Let

$$Z = \begin{bmatrix} z_1 \\ \vdots \\ z_N \end{bmatrix}, \ B = \begin{bmatrix} b_1' \\ \vdots \\ b_N' \end{bmatrix}$$

Abadie, Diamond, and Hainmueller (2010)



Abadie, Diamond, and Hainmueller (2010)



Dark Grey Indicates test for presence of Spillovers rejects Null.
CIs are 90%.

Two-Sample Tests
New test + R package

Comparing two samples

- Frequently we want to compare two samples, and see if they come from the same distribution.
- Many well known tests are in use for this purpose. T-test, Anova, etc.
- Non-parametric tests, which require few assumptions to work, often are very conservative, or only test for difference in one of a few moments of a distribution.

ECDF test statistics

- ECDF based tests are in this class – they are defined by taking some norm on the two empirical cumulative distributions.
- Randomization versions of these tests avoid need to be conservative – but are only as powerful as the test statistic allows.

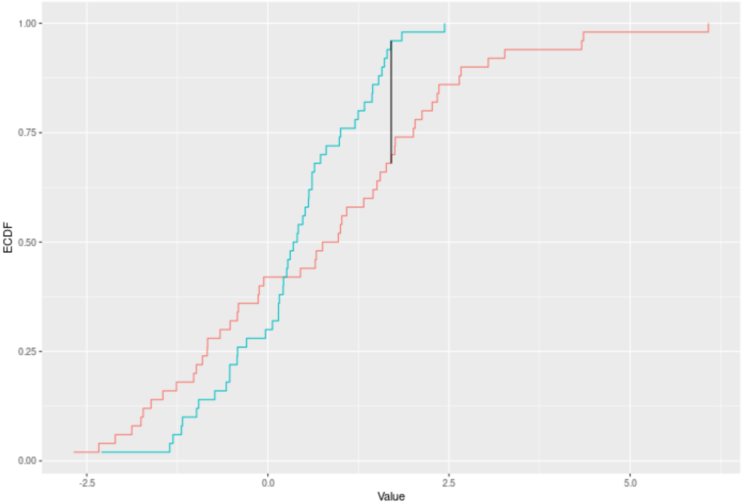
ECDF test statistics

- ECDF based tests are in this class – they are defined by taking some norm on the two empirical cumulative distributions.
- Randomization versions of these tests avoid need to be conservative – but are only as powerful as the test statistic allows.
- I discuss a new ECDF statistic which performs exceptionally well and is motivated by theory.

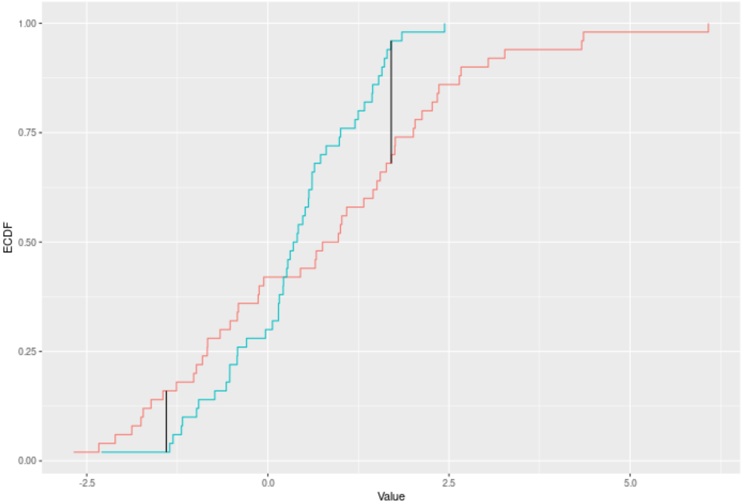
ECDF test statistics

- ECDF based tests are in this class – they are defined by taking some norm on the two empirical cumulative distributions.
- Randomization versions of these tests avoid need to be conservative – but are only as powerful as the test statistic allows.
- I discuss a new ECDF statistic which performs exceptionally well and is motivated by theory.
- I implement that statistic (and other ECDF tests) in an R package – `twosamples`.

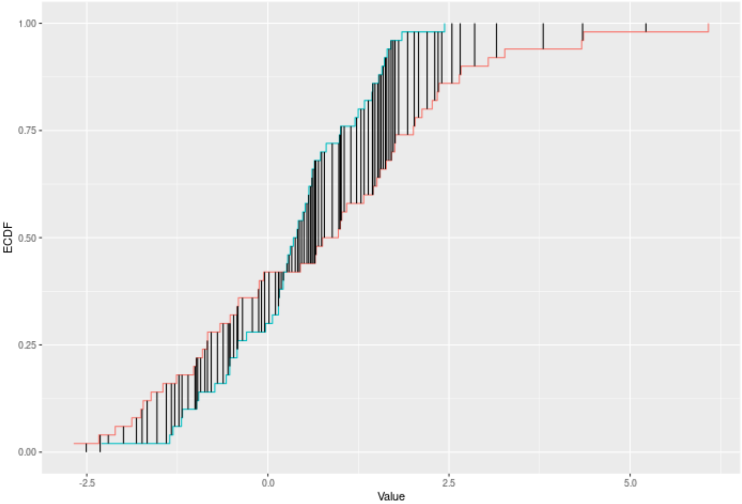
ECDF Statistics - Kolmogorov-Smirnov



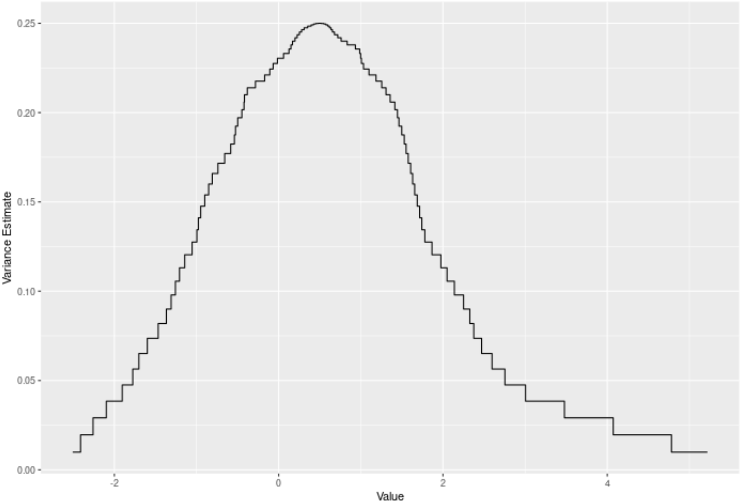
ECDF Statistics - Kuiper



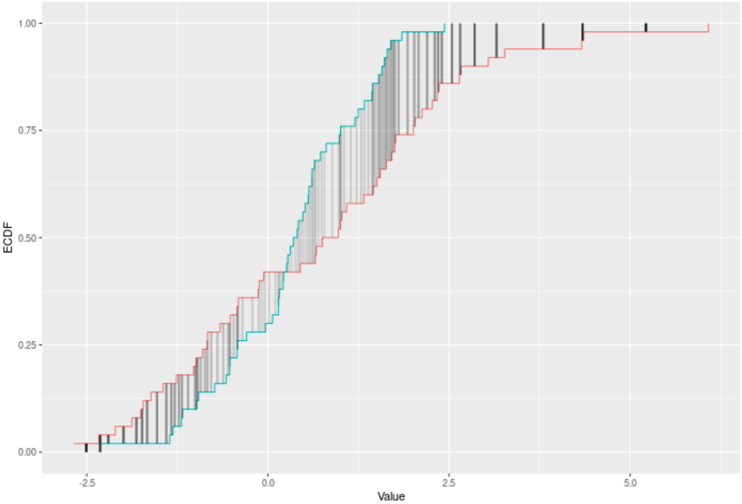
ECDF Statistics - Cramer-Von Mises



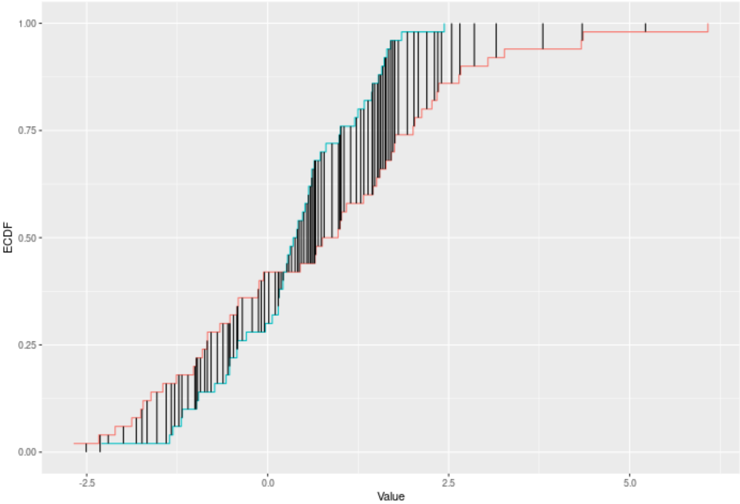
ECDF Statistics - Variance of ECDF



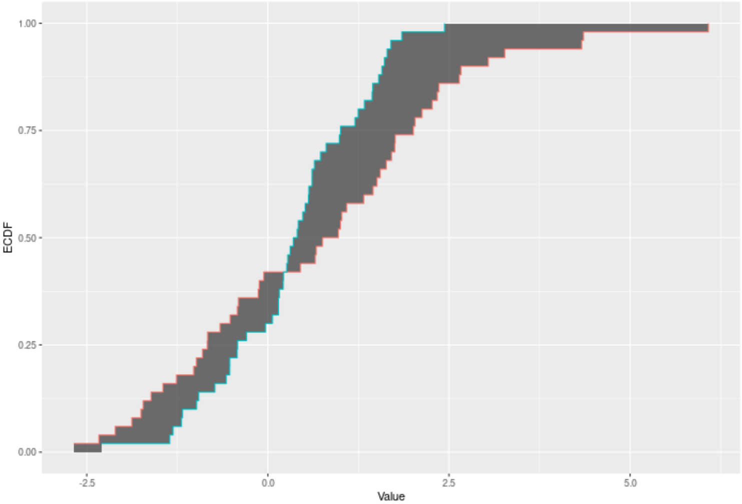
ECDF Statistics - Anderson-Darling



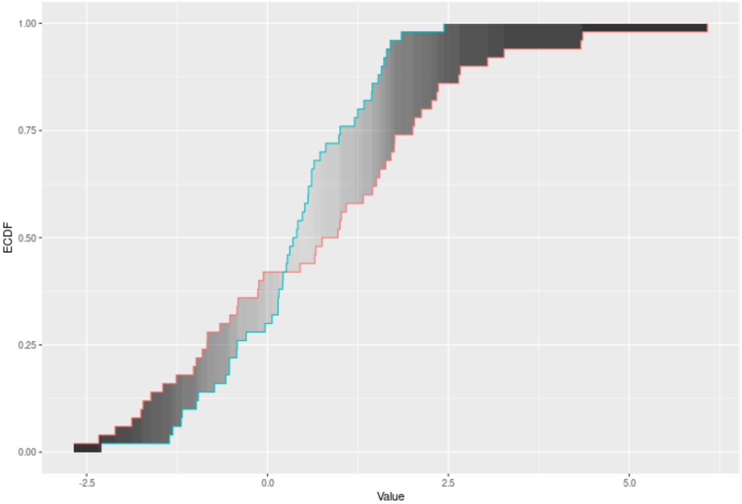
ECDF Statistics - Cramer-Von Mises



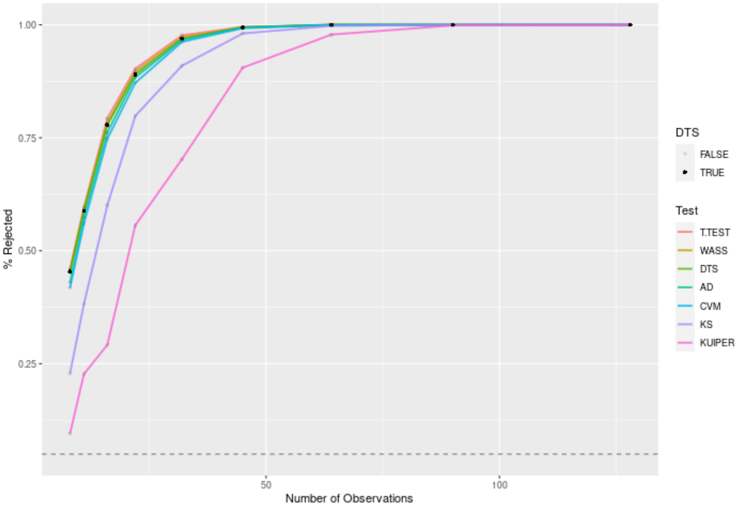
ECDF Statistics - Wasserstein



ECDF Statistics - Proposal

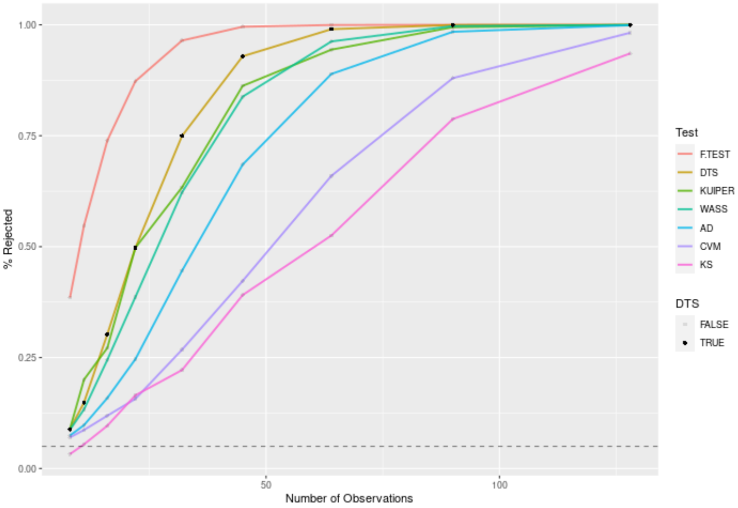


Power Simulations
Normal w/ Mean Shift

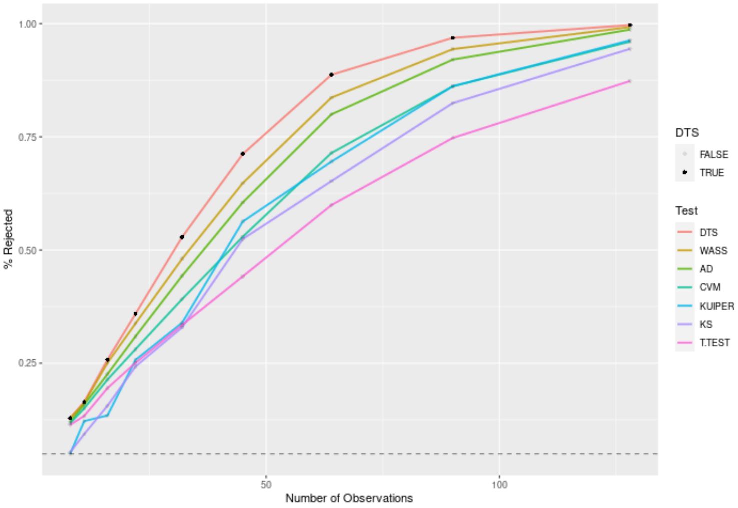


Power Simulations

Normal w/ Variance Inflation

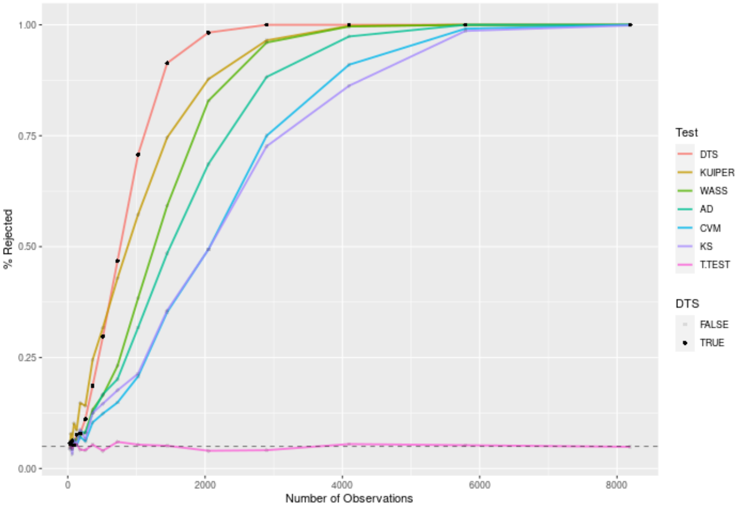


Power Simulations
Normal w/ Mean and Variance Shift



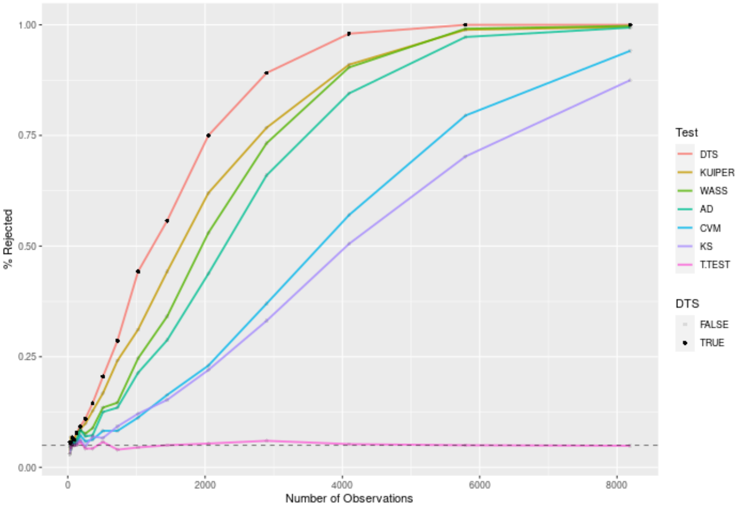
Power Simulations

Mixture of Normals – Mean/Variance Constant – Kurtosis changed



Power Simulations

Mixture of Normals – Bimodal, Constant mean



twosamples

- R package on CRAN
- Publicly available since 2018.
- Randomization tests, with test statistics in C++
- Includes KS, Kuiper, CVM, AD, WASS, and proposed test statistics.
- $\approx$ 600 downloads/month

Thank you all