Wag the Tails

Jean-Jacques Forneron, Boston University

November 2022

Motivation

- Robustness checks commonly used, no systematic approach
  - outliers (e.g. 2020 recession)
  - misspecification for some obs. (e.g. zero market shares/lower bound)

Motivation

- Robustness checks commonly used, no systematic approach
    - outliers (e.g. 2020 recession)
    - misspecification for some obs. (e.g. zero market shares/lower bound)
- Solution: **robust estimation**

Motivation

- Robustness checks commonly used, no systematic approach
  - outliers (e.g. 2020 recession)
  - misspecification for some obs. (e.g. zero market shares/lower bound)
- Solution: **robust estimation**
  1. hard to compute (NP hard),
  2. inconsistent without symmetry assumptions

Motivation

- Robustness checks commonly used, no systematic approach
  - outliers (e.g. 2020 recession)
  - misspecification for some obs. (e.g. zero market shares/lower bound)
- Solution: **robust estimation**
  1. hard to compute (NP hard),
  2. inconsistent without symmetry assumptions
- Here, we'll try to find estimates that are:

Motivation

- Robustness checks commonly used, no systematic approach
  - outliers (e.g. 2020 recession)
  - misspecification for some obs. (e.g. zero market shares/lower bound)
- Solution: **robust estimation**
  1. hard to compute (NP hard),
  2. inconsistent without symmetry assumptions
- Here, we'll try to find estimates that are:
  1. robust to a small amounts of large contamination

Motivation

- Robustness checks commonly used, no systematic approach
  - outliers (e.g. 2020 recession)
  - misspecification for some obs. (e.g. zero market shares/lower bound)
- Solution: **robust estimation**
  1. hard to compute (NP hard),
  2. inconsistent without symmetry assumptions
- Here, we'll try to find estimates that are:
  1. robust to a small amounts of large contamination
  2. adjusted for asymmetry bias

Motivation

- Robustness checks commonly used, no systematic approach
  - outliers (e.g. 2020 recession)
  - misspecification for some obs. (e.g. zero market shares/lower bound)
- Solution: **robust estimation**
  1. hard to compute (NP hard),
  2. inconsistent without symmetry assumptions
- Here, we'll try to find estimates that are:
  1. robust to a small amounts of large contamination
  2. adjusted for asymmetry bias
  3. not too hard to compute

Introduction

- Focus on two aspects:

Introduction

- Focus on two aspects:
  - **Diagnostics:** useful to visualize issues

Introduction

- Focus on two aspects:
  - **Diagnostics:** useful to visualize issues
  - **Robust estimation:** systematic approach
    - Huber loss (Huber, 1964)
    - MM-estimator (Yohai, 1987)
    - ...

Introduction

- Focus on two aspects:
  - **Diagnostics:** useful to visualize issues
  - **Robust estimation:** systematic approach
    - Huber loss (Huber, 1964)
    - MM-estimator (Yohai, 1987)
    - ...
    - Jaeckel (1971): asymmetric contamination $\Rightarrow$ bias

Introduction

- Focus on two aspects:
  - **Diagnostics:** useful to visualize issues
  - **Robust estimation:** systematic approach
    - Huber loss (Huber, 1964)
    - MM-estimator (Yohai, 1987)
    - ...
    - Jaeckel (1971): asymmetric contamination $\Rightarrow$ bias
- Economic data:
  - often skewed (e.g. price, volume, income)

    $\Rightarrow$ robust estimates are biased, e.g. $|\text{median}(x_i) - \mathbb{E}(x_i)| \leq \sigma_x$

Introduction

- Focus on two aspects:
  - **Diagnostics:** useful to visualize issues
  - **Robust estimation:** systematic approach
    - Huber loss (Huber, 1964)
    - MM-estimator (Yohai, 1987)
    - ...
    - Jaeckel (1971): asymmetric contamination $\Rightarrow$ bias
- Economic data:
  - often skewed (e.g. price, volume, income)
    $\Rightarrow$ robust estimates are biased, e.g. $|\text{median}(x_i) - \mathbb{E}(x_i)| \leq \sigma_x$
  - few large outliers (e.g. Black Monday)

Introduction

- Focus on two aspects:
  - **Diagnostics:** useful to visualize issues
  - **Robust estimation:** systematic approach
    - Huber loss (Huber, 1964)
    - MM-estimator (Yohai, 1987)
    - ...
    - Jaeckel (1971): asymmetric contamination $\Rightarrow$ bias
- Economic data:
  - often skewed (e.g. price, volume, income)
    $\Rightarrow$ robust estimates are biased, e.g. $|\text{median}(x_i) - \mathbb{E}(x_i)| \leq \sigma_x$
  - few large outliers (e.g. Black Monday)
- **Approach:** "Kernel" to downweight outliers, derive bias expression, bias correction procedure

Motivating Example: Effect of trade openness on inflation

- Romer (1993): "Openness and Inflation: Theory and Evidence"
- Simple IV regression:

$$\text{inflation}_i = \beta_0 + \beta_1 \text{openness}_i + u_i$$

- Instrument using log-land area as an instrument

Motivating Example: Effect of trade openness on inflation

- Romer (1993): "Openness and Inflation: Theory and Evidence"
- Simple IV regression:

$$\text{inflation}_i = \beta_0 + \beta_1 \text{openness}_i + u_i$$

- Instrument using log-land area as an instrument
- Just like OLS, we have:

$$y = X\hat{\beta}_n + \hat{u}_n$$

Motivating Example: Effect of trade openness on inflation

- Romer (1993): "Openness and Inflation: Theory and Evidence"
- Simple IV regression:

$$\text{inflation}_i = \beta_0 + \beta_1 \text{openness}_i + u_i$$

- Instrument using log-land area as an instrument
- Just like OLS, we have:

$$y = X\hat{\beta}_n + \hat{u}_n$$

- ... and the following identity:

$$\hat{\beta}_n = \hat{\beta}_n + \underbrace{(Z'X)^{-1} \sum_{i=1}^{n} z_i \hat{u}_i}_{=0}$$

Motivating Example: Effect of trade openness on inflation

- Romer (1993): "Openness and Inflation: Theory and Evidence"
- Simple IV regression:

$$\text{inflation}_i = \beta_0 + \beta_1 \text{openness}_i + u_i$$

- Instrument using log-land area as an instrument
- Just like OLS, we have:

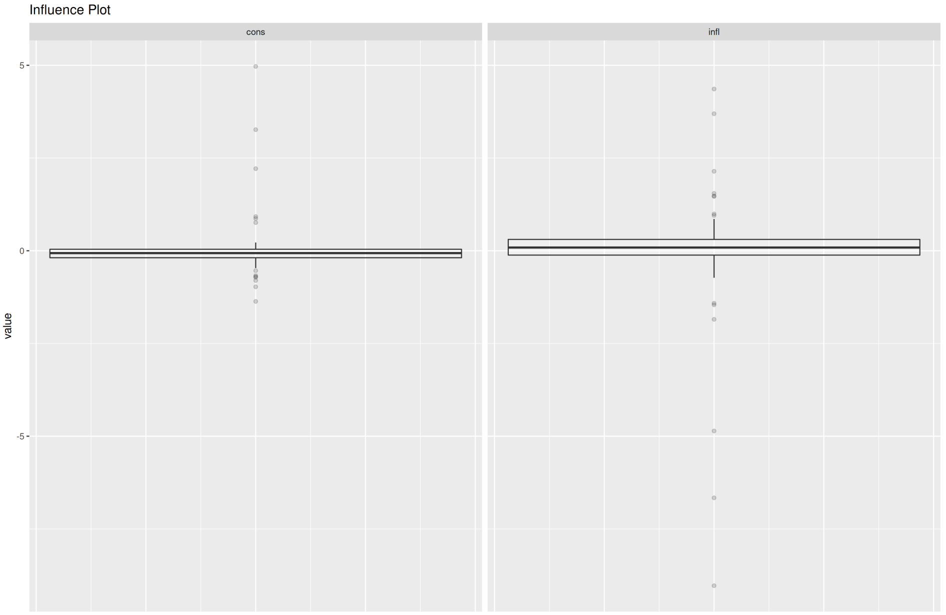
$$y = X\hat{\beta}_n + \hat{u}_n$$

- ...and the following identity:

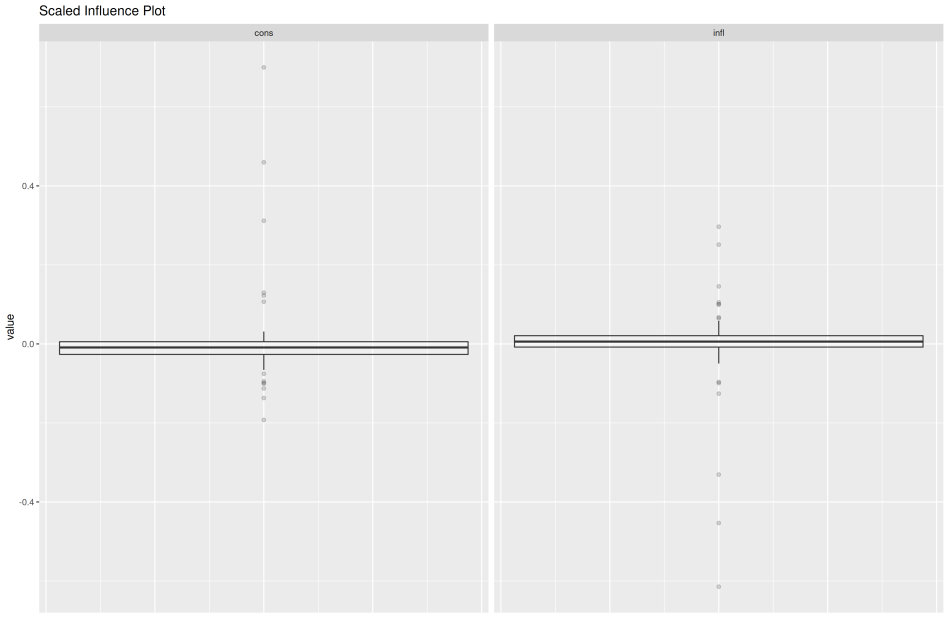
$$\hat{\beta}_n = \hat{\beta}_n + \underbrace{(Z'X)^{-1}\sum_{i=1}^{n} z_i \hat{u}_i}_{=0}$$

- Each $(Z'X)^{-1}\sum_{i=1}^{n} z_i \hat{u}_i$ measures contribution of i to $\hat{\beta}_0$ and $\hat{\beta}_1$

Diagnostics: coefficient-wise influence plot



Diagnostics: coefficient-wise influence plot



Effect of trade openness on inflation

- Drop observations with influence on β_1 greater than 1

	_cons	openness
	Full Sample	
$\hat{\beta}_n$	29.61	-33.29
SE	7.10	14.69
	Influence ≤ 1	
$\hat{\beta}_n$	17.63	-12.80
SE	1.78	3.72

- Estimates change by $\simeq$ 2SEs, standard errors more than quartered
- Remove 12 obs., $n = 114$

Estimation: A shrinkage estimator

- Suppose $x_i \sim F$ iid, want to estimate:

$$\mu = \mathbb{E}(x_i)$$

eventually we want θ s.t. $\mu(\theta) = 0$

Estimation: A shrinkage estimator

- Suppose $x_i \sim F$ iid, want to estimate:

$$\mu = \mathbb{E}(x_i)$$

  eventually we want θ s.t. $\mu(\theta) = 0$
- Start with a shrinkage estimator

Estimation: A shrinkage estimator

- Suppose $x_i \sim F$ iid, want to estimate:

$$\mu = \mathbb{E}(x_i)$$

  eventually we want θ s.t. $\mu(\theta) = 0$

- Start with a shrinkage estimator
- Pick any $\lambda > 0$:

$$\hat{\mu}_{n\lambda} = \frac{1}{n} \sum_{i=1}^{n} \frac{x_i}{1 + \|x_i\|^2/\lambda}$$

Estimation: A shrinkage estimator

- Suppose $x_i \sim F$ iid, want to estimate:

$$\mu = \mathbb{E}(x_i)$$

  eventually we want θ s.t. $\mu(\theta) = 0$

- Start with a shrinkage estimator
- Pick any $\lambda > 0$:

$$\hat{\mu}_{n\lambda} = \frac{1}{n} \sum_{i=1}^{n} \frac{x_i}{1 + \|x_i\|^2/\lambda}$$

- Notice:

$$x_i = \pm\infty \Rightarrow \frac{x_i}{1 + \|x_i\|^2/\lambda} = 0$$

  discards extreme observations

Estimation: A shrinkage estimator

- Suppose $x_i \sim F$ iid, want to estimate:

$$\mu = \mathbb{E}(x_i)$$

eventually we want θ s.t. $\mu(\theta) = 0$

- Start with a shrinkage estimator
- Pick any $\lambda > 0$:

$$\hat{\mu}_{n\lambda} = \frac{1}{n} \sum_{i=1}^{n} \frac{x_i}{1 + \|x_i\|^2/\lambda}$$

- Notice:

$$x_i = \pm\infty \Rightarrow \frac{x_i}{1 + \|x_i\|^2/\lambda} = 0$$

discards extreme observations

- $\hat{\mu}_{n\lambda}$ has some interesting properties

A shrinkage estimator: tails

- For any $x, \lambda > 0$:
$$\left\| \frac{x}{1 + \|x\|^2/\lambda} \right\|_\infty \leq \frac{\sqrt{\lambda}}{2}$$
the **influence function** is bounded by $\sqrt{\lambda}/2$

A shrinkage estimator: tails

- For any $x, \lambda > 0$:

$$\left\| \frac{x}{1 + \|x\|^2/\lambda} \right\|_\infty \leq \frac{\sqrt{\lambda}}{2}$$

  the **influence function** is bounded by $\sqrt{\lambda}/2$

- Bernstein's inequality (scalar):

$$\mathbb{P}\left(\sqrt{n}|\hat{\mu}_{n\lambda} - \mu_\lambda| \geq t\right) \leq 2\exp\left(-\frac{t^2}{2\sigma_\lambda^2 + \frac{t}{12}\sqrt{\lambda/n}}\right)$$

  where $\mu_\lambda = \mathbb{E}(\hat{\mu}_{n\lambda})$, $\sigma_\lambda^2 = n \cdot \mathrm{var}(\hat{\mu}_{n\lambda})$

A shrinkage estimator: tails

- For any $x, \lambda > 0$:

$$\left\| \frac{x}{1 + \|x\|^2/\lambda} \right\|_\infty \leq \frac{\sqrt{\lambda}}{2}$$

  the **influence function** is bounded by $\sqrt{\lambda}/2$

- Bernstein's inequality (scalar):

$$\mathbb{P}\left(\sqrt{n}|\hat{\mu}_{n\lambda} - \mu_\lambda| \geq t\right) \leq 2\exp\left(-\frac{t^2}{2\sigma_\lambda^2 + \frac{t}{12}\sqrt{\lambda/n}}\right)$$

  where $\mu_\lambda = \mathbb{E}(\hat{\mu}_{n\lambda})$, $\sigma_\lambda^2 = n \cdot \mathrm{var}(\hat{\mu}_{n\lambda})$

- In finite samples, behaves like:

$$2\exp\left(-\frac{t^2}{2\sigma_\lambda^2}\right) \text{ for } t \simeq 0, \quad 2\exp\left(-\frac{t}{12\sqrt{\lambda/n}}\right) \text{ for } t \to \infty$$

  sub-exponential for large deviations

A shrinkage estimator: bias

- If $\mathbb{E}\|x_i\| < \infty$

$$\mu_\lambda = \mu - \frac{1}{\lambda}\mathbb{E}\left(\frac{x_i\|x_i\|^2}{1 + \|x_i\|^2/\lambda}\right)$$

  bias $= 0$ if F symmetric around 0

A shrinkage estimator: bias

- If $\mathbb{E}\|x_i\| < \infty$

$$\mu_\lambda = \mu - \frac{1}{\lambda}\mathbb{E}\left(\frac{x_i\|x_i\|^2}{1 + \|x_i\|^2/\lambda}\right)$$

  bias $= 0$ if F symmetric around 0

- Finite + bounded moments imply:

$$|\text{bias}| = O(\lambda^{-1}) \text{ if } \mathbb{E}\|x_i\|^3 < \infty,$$

$$|\text{bias}| = O(\lambda^{-\frac{1}{2}}) \text{ if } \mathbb{E}\|x_i\|^2 < \infty$$

A shrinkage estimator: bias

- If $\mathbb{E}\|x_i\| < \infty$

$$\mu_\lambda = \mu - \frac{1}{\lambda}\mathbb{E}\left(\frac{x_i\|x_i\|^2}{1+\|x_i\|^2/\lambda}\right)$$

  bias $= 0$ if F symmetric around 0

- Finite + bounded moments imply:

$$|\text{bias}| = O(\lambda^{-1}) \text{ if } \mathbb{E}\|x_i\|^3 < \infty,$$
$$|\text{bias}| = O(\lambda^{-\frac{1}{2}}) \text{ if } \mathbb{E}\|x_i\|^2 < \infty$$

- Bias reduction:

$$2\mu_\lambda - \mu_{\lambda/2} = \mu - \frac{2}{\lambda^2}\mathbb{E}\left(\frac{x_i\|x_i\|^4}{(1+\|x_i\|^2/\lambda)(1+2\|x_i\|^2/\lambda)}\right),$$

  if $x_i \sim \delta_x$ (mass) then bias $\approx x$ as $x \to \pm\infty$

A shrinkage estimator: bias

- Repeat the reduction:

$$
\begin{aligned}
&2[2\mu_\lambda - \mu_{\lambda/2}] - [2\mu_{\lambda/2} - \mu_{\lambda/4}] \\
&= \mu - \frac{4}{\lambda^2}\mathbb{E}\left(\frac{xi\|x_i\|^4(2\|x_i\|^2/\lambda - 1)}{(1+\|x_i\|^2/\lambda)(1+2\|x_i\|^2/\lambda)(1+4\|x_i\|^2/\lambda)}\right)
\end{aligned}
$$

if $x_i \sim \delta_x$ (mass) then bias $= 0$ if $x \in \{0, \pm\sqrt{\lambda/2}\}$

A shrinkage estimator: outliers

- **Contamination**: $x_1, \ldots, x_N \sim (\mu, \sigma^2)$, $x_{N+1}, \ldots, x_{N+M}$ arbitrary

$$\sqrt{n}(\hat{\mu}_{n\lambda} - \mu_\lambda) = \sqrt{n}(\hat{\mu}_{N\mu} - \mu_{N\lambda}) + o(1),$$

if $n = N + M$ s.t. $M/n \to 0$ and $M = o(\sqrt{n/\lambda})$

A shrinkage estimator: outliers

- **Contamination**: $x_1, \ldots, x_N \sim (\mu, \sigma^2)$, $x_{N+1}, \ldots, x_{N+M}$ arbitrary

$$\sqrt{n}(\hat{\mu}_{n\lambda} - \mu_\lambda) = \sqrt{n}(\hat{\mu}_{N\mu} - \mu_{N\lambda}) + o(1),$$

  if $n = N + M$ s.t. $M/n \to 0$ and $M = o(\sqrt{n/\lambda})$
- Issues: $\hat{\mu}_{n\lambda}$ is not location, scale invariant

A non-shrinkage estimator

- Similar idea find μ s.t.:

$$\sum_{i=1}^{n} \frac{x_i - \hat{\mu}_{n,\lambda}}{1 + \|x_i - \hat{\mu}_{n,\lambda}\|^2_{\Sigma^{-1}}/\lambda} = 0$$

(FOC for multivariate Student with dof λ)

A non-shrinkage estimator

- Similar idea find μ s.t.:

$$\sum_{i=1}^{n} \frac{x_i - \hat{\mu}_{n,\lambda}}{1 + \|x_i - \hat{\mu}_{n,\lambda}\|_{\Sigma^{-1}}^2/\lambda} = 0$$

  (FOC for multivariate Student with dof λ)

- **Problem:** $= 0$ for $\|\hat{\mu}_{n,\lambda}\| = \infty$;

A non-shrinkage estimator

- Similar idea find μ s.t.:

$$\sum_{i=1}^{n} \frac{x_i - \hat{\mu}_{n,\lambda}}{1 + \|x_i - \hat{\mu}_{n,\lambda}\|^2_{\Sigma^{-1}}/\lambda} = 0$$

(FOC for multivariate Student with dof λ)

- **Problem:** $= 0$ for $\|\hat{\mu}_{n,\lambda}\| = \infty$; issue for computation and ULLN

A non-shrinkage estimator

- Similar idea find μ s.t.:

$$\sum_{i=1}^{n} \frac{x_i - \hat{\mu}_{n,\lambda}}{1 + \|x_i - \hat{\mu}_{n,\lambda}\|_{\Sigma^{-1}}^2/\lambda} = 0$$

(FOC for multivariate Student with dof λ)

- **Problem:** $= 0$ for $\|\hat{\mu}_{n,\lambda}\| = \infty$; issue for computation and ULLN
- Instead, minimize over (μ, Σ):

$$\log|\Sigma| + \frac{\lambda + p}{n} \sum_{i=1}^{n} \Big[\log\left(1 + \frac{\|x_i - \mu\|_{\Sigma^{-1}}^2}{\lambda}\right) \Big] + \frac{\kappa_1}{\lambda}\|\mu\|_{\Sigma^{-1}}^2 + \frac{\kappa_2}{\lambda}\mathrm{trace}(\Sigma)$$

A non-shrinkage estimator

- Similar idea find μ s.t.:

$$\sum_{i=1}^{n} \frac{x_i - \hat{\mu}_{n,\lambda}}{1 + \|x_i - \hat{\mu}_{n,\lambda}\|_{\Sigma^{-1}}^2/\lambda} = 0$$

(FOC for multivariate Student with dof λ)

- **Problem:** $= 0$ for $\|\hat{\mu}_{n,\lambda}\| = \infty$; issue for computation and ULLN
- Instead, minimize over (μ, Σ):

$$\log|\Sigma| + \frac{\lambda + p}{n} \sum_{i=1}^{n} \left[\log\left(1 + \frac{\|x_i - \mu\|_{\Sigma^{-1}}^2}{\lambda}\right) \right] + \frac{\kappa_1}{\lambda}\|\mu\|_{\Sigma^{-1}}^2 + \frac{\kappa_2}{\lambda}\mathrm{trace}(\Sigma)$$

- **Penalty + bounded influence function** imply:

$$\|\Sigma^{-1/2}\mu\|_\infty \leq \frac{\lambda^{3/2}}{2\kappa_1}, \quad \|\Sigma\|_2 \lesssim \lambda^{3/2}$$

with probability 1, for any dataset

Monte Carlo + Empirical Application

Monte Carlo

- Heteroskedastic design:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \sigma(x) u_i$$

  where $\sigma(x) = C\sqrt{1 + |x'\beta|}$ s.t. $\mathbb{E}(\sigma(x)^2) = 1$, $u_i \sim (\chi_1^2 - 1)$, $x_1, \ldots, x_3 \sim \chi_1^2$

- For $i = 1, \ldots, 10$: $x_1, x_2 = 4x_1, 4x_2$, $y = -3x'\beta + \sigma(x)u_i$
- Use $n = 500$
- Moment equation:

$$\mu(\beta) = \mathbb{E}(x_i[y_i - x_i'\beta]) = 0$$

- Compare different estimates:
  - OLS without $i \in \{1, \ldots, 10\}$ (clean),
  - OLS with all $i \in \{1, \ldots, n\}$ (full),
  - Huber estimates (RLM in R),
  - Bias (un)corrected estimates $\hat{\beta}_{n,\lambda}$, $\tilde{\beta}_{n,\lambda}$, $\tilde{\tilde{\beta}}_{n,\lambda}$

Monte Carlo ($n = 500$) - No outliers

	Stdev				Rejection rate (5%)			
	β_0	β_1	β_2	β_3	β_0	β_1	β_2	β_3
$\hat{\beta}_{clean}$	-	-	-	-	-	-	-	-
$\hat{\beta}_{full}$	0.099	0.066	0.058	0.059	0.051	0.082	0.079	0.095
$\hat{\beta}_{huber}$	0.065	0.034	0.031	0.032	0.649	0.534	0.315	0.352
$\hat{\beta}_{n\lambda}$	0.073	0.047	0.041	0.042	0.053	0.439	0.370	0.387
$\tilde{\beta}_{n\lambda}$	0.079	0.050	0.044	0.045	0.126	0.335	0.276	0.298
$\tilde{\tilde{\beta}}_{n\lambda}$	0.085	0.054	0.048	0.049	0.160	0.208	0.174	0.193

Legend: 1000 MC replications, $\hat{\beta}_{clean}$ without outliers, $\hat{\beta}_{full}$ full sample, $\hat{\beta}_{huber}$ Huber loss (*rlm*), $\hat{\beta}_{n\lambda}$, $\tilde{\beta}_{n\lambda}$, $\tilde{\tilde{\beta}}_{n\lambda}$ estimates using penalized Student mean estimates for $\overline{g}_n(\beta)$, corrected and twice corrected. $\lambda = n^{2/3}$

Monte Carlo ($n = 500$) - Contaminated

	Stdev				Rejection rate (5%)			
	β_0	β_1	β_2	β_3	β_0	β_1	β_2	β_3
$\hat{\beta}_{clean}$	0.100	0.066	0.061	0.058	0.049	0.086	0.097	0.077
$\hat{\beta}_{full}$	0.951	0.748	0.517	0.180	0.417	0.312	0.334	0.020
$\hat{\beta}_{huber}$	0.081	0.043	0.043	0.032	0.250	0.758	0.580	0.324
$\hat{\beta}_{n\lambda}$	0.077	0.045	0.043	0.043	0.025	0.220	0.179	0.266
$\tilde{\beta}_{n\lambda}$	0.086	0.048	0.046	0.047	0.038	0.130	0.103	0.181
$\tilde{\tilde{\beta}}_{n\lambda}$	0.096	0.053	0.052	0.052	0.041	0.061	0.049	0.101

Legend: 1000 MC replications, $\hat{\beta}_{clean}$ without outliers, $\hat{\beta}_{full}$ full sample, $\hat{\beta}_{huber}$ Huber loss (*rlm*), $\hat{\beta}_{n\lambda}$, $\tilde{\beta}_{n\lambda}$, $\tilde{\tilde{\beta}}_{n\lambda}$ estimates using penalized Student mean estimates for $\overline{g}_n(\beta)$, corrected and twice corrected. $\lambda = n^{2/3}$

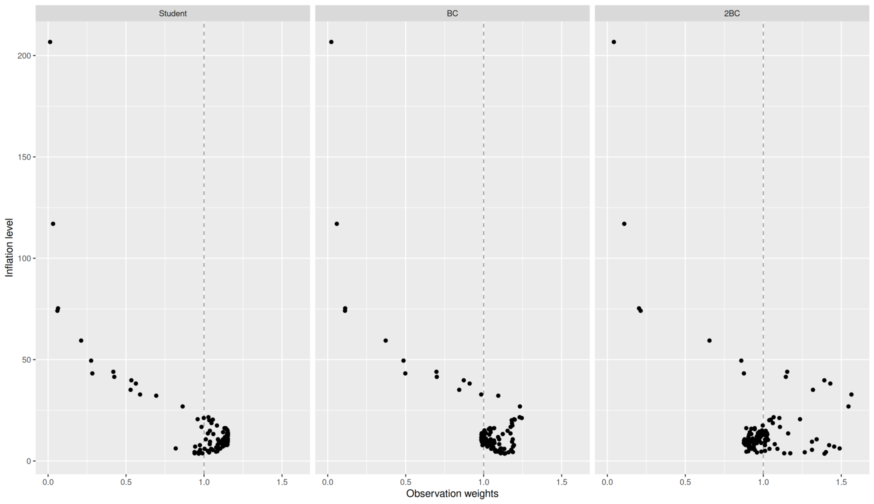
Effect of trade openness on inflation

	$\lambda = n^{2/3}$			$\lambda = \sqrt{2}n^{2/3}$		
	cons	β_{openness}	ESS	cons	β_{openness}	ESS
$\hat{\beta}_{2SLS}$	29.607	-33.287	114			
	(6.699)	(13.681)				
$\hat{\beta}_{n\lambda}$	15.800	-9.731	106.3	17.243	-11.977	108.6
	(2.105)	(4.321)		(2.612)	(5.413)	
$\tilde{\beta}_{n\lambda}$	17.259	-11.415	108.8	19.179	-14.483	110.5
	(2.456)	(4.971)		(3.093)	(6.410)	
$\approx\!\!\!\beta_{n\lambda}$	19.359	-13.804	107.9	21.839	-18.074	109.5
	(2.903)	(5.958)		(3.689)	(7.749)	

Legend: $n = 114$, $\lambda = n^{2/3} = 23.51$, $\lambda = \sqrt{2}n^{2/3} = 33.25$,

ESS = Effective Sample Size = $(\sum \omega_i)^2/(\sum \omega_i^2)$

Implied weights: $\tilde{\beta}_{n\lambda} = \left(\sum_i \omega_{i\lambda} z_i x_i'\right)^{-1} \sum_i \omega_{i\lambda} z_i y_i$



References

Huber, P. J. (1964). Robust Estimation of a Location Parameter. *The Annals of Mathematical Statistics*, 35(1):73 – 101.

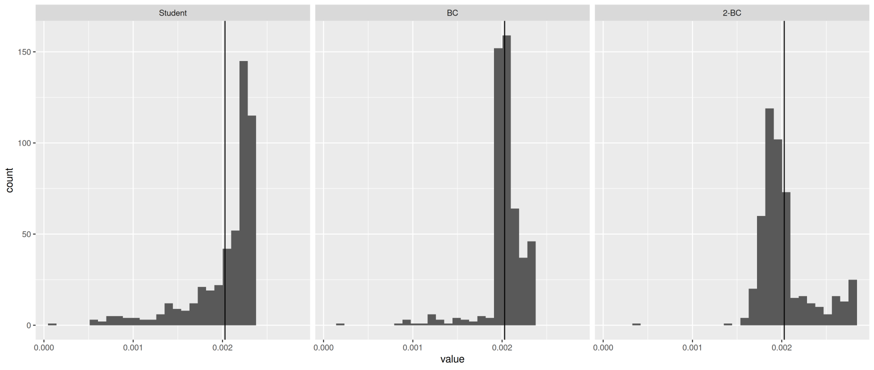
Jaeckel, L. A. (1971). Robust estimates of location: Symmetry and asymmetric contamination. *The Annals of Mathematical Statistics*, 42(3):1020–1034.

Yohai, V. J. (1987). High breakdown-point and high efficiency robust estimates for regression. *The Annals of statistics*, pages 642–656.

Conclusion

- Contamination with asymmetric data
  - robust to small amounts of outliers
  - bias = f(moments)
  - bias-correction
- Algorithm for minimizing the penalized t-distribution
- To do: detailed finite sample/asymptotic analysis for the t-estimator

Example 1: RDD



Diagnostics: effect of i on $\hat{\beta}_n$

- For OLS, we have:

$$Y = X\hat{\beta}_n + \hat{u}_n$$

- Which implies:

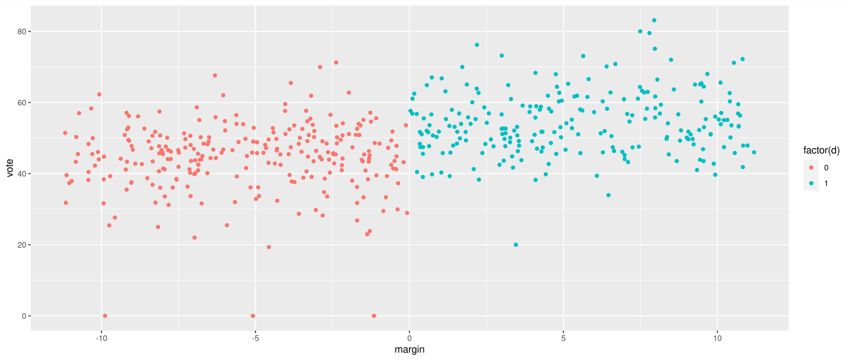
$$\hat{\beta}_n = \hat{\beta}_n + (X'X)^{-1}\sum_{i=1}^{n} x_i \hat{u}_i$$

- $(X'X)^{-1}\sum_{i=1}^{n} x_i \hat{u}_i$ measures relative contribution of i, **for each coefficient**

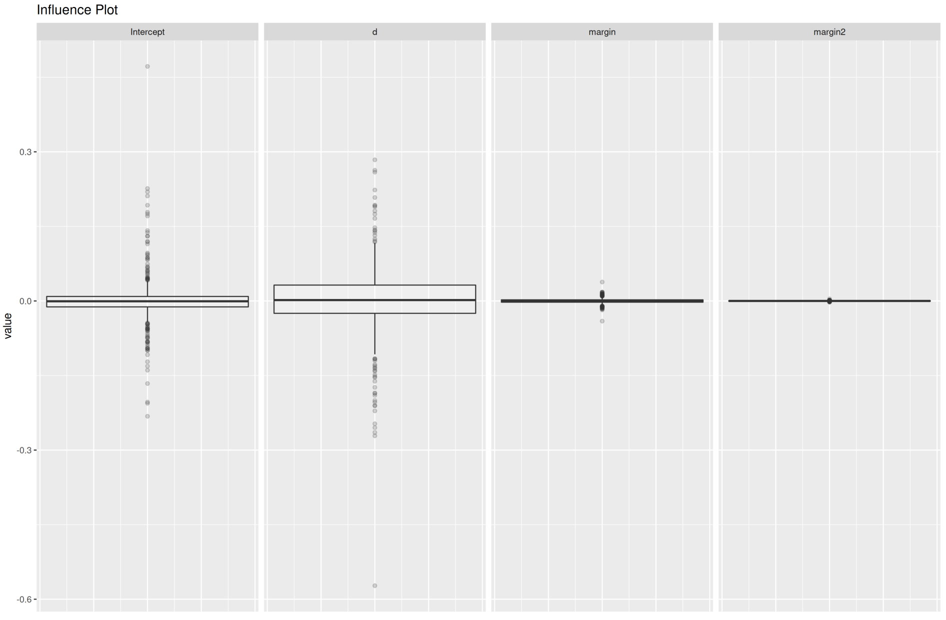
Example 1: RDD with senate election data

- Regression discontinuity design:

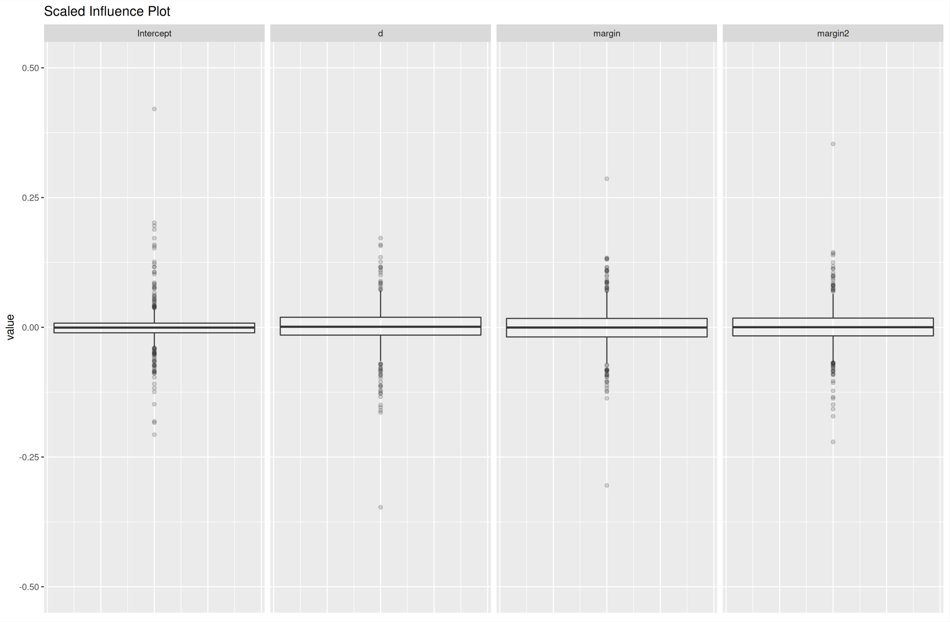
$$\text{vote}_i = \beta_0 + \beta_1 d_i + \beta_2 \text{margin}_i + \beta_3 \text{margin}_i^2 + u_i$$



Example 1: RDD with senate election data



Example 1: RDD with senate election data



Example 1: RDD with senate election data

	_cons	d	margin	margin^2
	Full Sample			
$\hat{\beta}_n$	45.4912	7.4005	0.2092	0.0048
SE	(1.1211)	(1.6514)	(0.1330)	(0.0118)
	Remove influential obs. on β_d			
$\hat{\beta}_n$	45.9675	6.8223	0.2476	0.0022
SE	(1.0241)	(1.5543)	(0.1275)	(0.0114)
	Remove influential obs. on β_{margin^2}			
$\hat{\beta}_n$	45.2824	7.6631	0.1681	0.0090
SE	(1.0993)	(1.6269)	(0.1266)	(0.0110)

Example 1: RDD

	$\lambda = n^{2/3}$				$\lambda = \sqrt{2}n^{2/3}$			
	cons	β_d	β_{margin}	β_{margin^2}	cons	β_d	β_{margin}	β_{margin^2}
$\hat{\beta}_{ols}$	45.491	7.400	0.209	0.005				
	(1.121)	(1.651)	(0.133)	(0.012)				
$\hat{\beta}_{huber}$	46.177	6.620	0.196	0.001				
	(0.960)	(1.533)	(0.125)	(0.011)				
$\hat{\beta}_{n\lambda}$	45.795	7.029	0.160	0.005	45.708	7.153	0.168	0.005
	(0.930)	(1.402)	(0.110)	(0.010)	(0.971)	(1.459)	(0.114)	(0.010)
$\tilde{\beta}_{n\lambda}$	45.518	7.476	0.155	0.005	45.474	7.512	0.168	0.006
	(1.042)	(1.562)	(0.119)	(0.010)	(1.063)	(1.586)	(0.122)	(0.011)
$\tilde{\tilde{\beta}}_{n\lambda}$	45.288	7.809	0.161	0.006	45.322	7.714	0.178	0.007
	(1.105)	(1.644)	(0.125)	(0.011)	(1.108)	(1.640)	(0.128)	(0.011)

Monte Carlo ($n = 500$) - No outliers

	Median Bias				Rejection rate (5%)			
	β_0	β_1	β_2	β_3	β_0	β_1	β_2	β_3
$\hat{\beta}_{clean}$	-	-	-	-	-	-	-	-
$\hat{\beta}_{full}$	-0.000	-0.003	-0.005	-0.007	0.051	0.083	0.078	0.095
$\hat{\beta}_{huber}$	-0.140	-0.068	-0.042	-0.044	0.649	0.534	0.315	0.352
$\hat{\beta}_{n\lambda}$	0.009	-0.071	-0.054	-0.057	0.053	0.439	0.370	0.387
$\tilde{\beta}_{n\lambda}$	0.063	-0.060	-0.048	-0.050	0.126	0.335	0.276	0.298
$\tilde{\tilde{\beta}}_{n\lambda}$	0.084	-0.043	-0.034	-0.036	0.160	0.208	0.174	0.193

Legend: 1000 MC replications, $\hat{\beta}_{clean}$ without outliers, $\hat{\beta}_{full}$ full sample, $\hat{\beta}_{huber}$ Huber loss (*rlm*), $\hat{\beta}_{n\lambda}$, $\tilde{\beta}_{n\lambda}$, $\tilde{\tilde{\beta}}_{n\lambda}$ estimates using penalized Student mean estimates for $\overline{g}_n(\beta)$, corrected and twice corrected. $\lambda = n^{2/3}$

Monte Carlo ($n = 500$) - Contaminated

	Median Bias				Rejection rate (5%)			
	β_0	β_1	β_2	β_3	β_0	β_1	β_2	β_3
$\hat{\beta}_{clean}$	0.008	-0.006	-0.006	-0.004	0.049	0.086	0.097	0.077
$\hat{\beta}_{full}$	1.633	-1.083	-0.853	-0.024	0.417	0.312	0.334	0.020
$\hat{\beta}_{huber}$	-0.075	-0.114	-0.091	-0.045	0.250	0.758	0.580	0.324
$\hat{\beta}_{n\lambda}$	-0.001	-0.067	-0.055	-0.051	0.025	0.220	0.179	0.266
$\tilde{\beta}_{n\lambda}$	0.045	-0.058	-0.050	-0.040	0.038	0.130	0.103	0.181
$\tilde{\tilde{\beta}}_{n\lambda}$	0.049	-0.042	-0.040	-0.023	0.041	0.061	0.049	0.101

Legend: 1000 MC replications, $\hat{\beta}_{clean}$ without outliers, $\hat{\beta}_{full}$ full sample, $\hat{\beta}_{huber}$ Huber loss (*rlm*), $\hat{\beta}_{n\lambda}$, $\tilde{\beta}_{n\lambda}$, $\tilde{\tilde{\beta}}_{n\lambda}$ estimates using penalized Student mean estimates for $\overline{g}_n(\beta)$, corrected and twice corrected. $\lambda = n^{2/3}$

Non-shrinkage estimator ($n = 125$)

	Stdev				Rejection rate (5%)			
	β_0	β_1	β_2	β_3	β_0	β_1	β_2	β_3
$\hat{\beta}_{clean}$	0.202	0.141	0.122	0.126	0.037	0.105	0.108	0.084
$\hat{\beta}_{full}$	1.274	1.023	0.698	0.478	0.652	0.564	0.558	0.032
$\hat{\beta}_{huber}$	1.037	0.993	0.661	0.142	0.117	0.175	0.151	0.069
$\hat{\beta}_{n\lambda}$	0.154	0.093	0.095	0.089	0.023	0.052	0.032	0.133
$\tilde{\beta}_{n\lambda}$	0.189	0.108	0.127	0.095	0.009	0.029	0.025	0.083
$\tilde{\tilde{\beta}}_{n\lambda}$	0.287	0.195	0.229	0.112	0.015	0.015	0.028	0.043

Non-shrinkage estimator ($n = 250$)

	Stdev				Rejection rate (5%)			
	β_0	β_1	β_2	β_3	β_0	β_1	β_2	β_3
$\hat{\beta}_{clean}$	0.144	0.093	0.087	0.085	0.057	0.110	0.108	0.096
$\hat{\beta}_{full}$	1.127	0.915	0.624	0.294	0.540	0.407	0.451	0.034
$\hat{\beta}_{huber}$	0.163	0.094	0.107	0.051	0.099	0.504	0.381	0.184
$\hat{\beta}_{n\lambda}$	0.109	0.065	0.058	0.061	0.036	0.155	0.116	0.233
$\tilde{\beta}_{n\lambda}$	0.123	0.070	0.063	0.065	0.020	0.090	0.078	0.178
$\tilde{\tilde{\beta}}_{n\lambda}$	0.144	0.079	0.074	0.073	0.018	0.048	0.034	0.105