Inference for the Mean

Ulrich K. Müller
Princeton University

September 2017

preliminary

Motivation

- Key building block of econometrics: t-statistic for inference about mean

  $\Rightarrow$ White standard errors in regression reduce to inference about mean of $x_i e_i$ for suitably defined e_i, similar in GMM

- Potential challenge: inaccurate approximations by CLT in numerator and LLN in denominator

  $\Rightarrow$ induced by heavy-tailed population, especially asymmetry, in small samples

  $\Rightarrow$ effective sample size often not very large due to clustering

- Standard remedy: Bootstrap

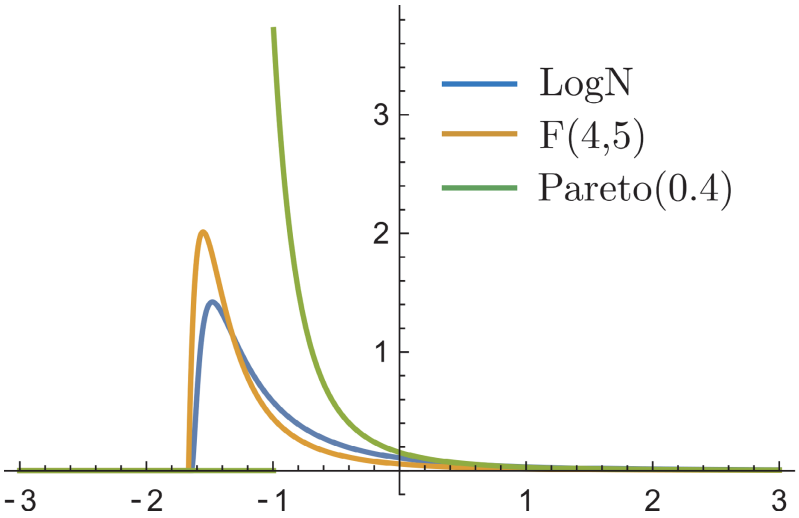
  $\Rightarrow$ provides refinement when at least three moments exist

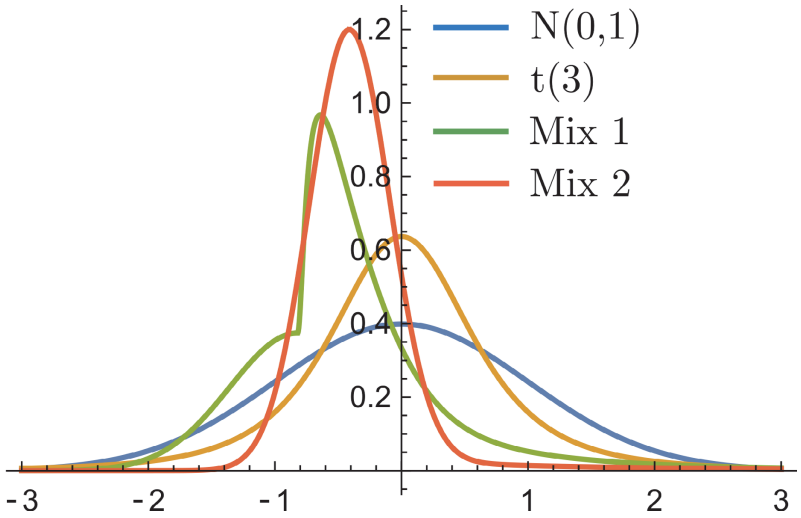
Small Sample Null Rejection Probabilities

	N(0,1)	LogN	F(4,5)	t(3)	P(0.4)	Mix 1	Mix 2
			$n = 50$				
t-stat	4.8	9.9	12.6	4.4	13.2	12.3	19.1
sym-boot	5.0	7.9	10.1	4.0	10.3	11.8	18.6
asym-boot	5.1	6.9	8.2	7.3	8.3	11.8	18.0
			$n = 100$				
t-stat	5.0	8.3	10.8	4.6	11.3	9.9	15.8
sym-boot	5.1	6.9	8.9	4.2	9.3	8.8	14.6
asym-boot	5.2	6.3	7.8	6.9	7.8	8.7	13.5
			$n = 500$				
t-stat	5.1	5.7	7.7	4.9	8.0	6.9	9.5
sym-boot	5.3	5.2	6.9	4.8	7.0	6.0	8.0
asym-boot	5.3	5.5	6.8	6.2	6.8	6.3	7.5

Note: Nominal 5% tests

Population Densities





Basic Idea

- W_i i.i.d. sample of size n with cdf F, $H_0 : E[W] = 0$, long right tail

- Divide and conquer: Largest k order statistics $\mathbf{W}^e = (W_1^e, \ldots, W_k^e)$, and remaining $n - k$ observations W_i^m

- Conditional on $\mathbf{W}^e$, W_i^m i.i.d. with cdf $F(w)/(1-F(W_k^e))$ for $w \leq W_k^e$

  $\Rightarrow$ Conditional mean under H_0:

  $(1 - P(W > w))E[W|W \leq w] + P(W > w)E[W|W > w] = 0$, so

$$m(w) = -E[W|W \leq w] = \frac{P(W > w)E[W|W > w]}{1 - P(W > w)}$$

Basic Idea, ctd.

- Asymptotic approximations:

  1. $\mathbf{W}^e$ has (joint) extreme value distribution

  2. Conditional on $\mathbf{W}^e$, $\sum_{i=1}^{n-k} W_i^m$ is approximately normal with mean $-(n-k)m(W_k^e)$

  Tail properties of F determine both asymptotic distribution of $\mathbf{W}^e$, and form of $m(\cdot)$

- Use computational methods to construct test that controls size under these asymptotic approximations

Contributions

- New asymptotic approximation for inference about mean

  $\Rightarrow$ exploits extreme value theory and CLT

- Theory: Generates refinement for population with more than two but less than three moments, while bootstrap does not

- Practice: Implementation that only requires few "tail observations"

Related Literature

- Inference for mean under heavy tails (less than two moments)

  Romano and Wolf (2000), Peng (2001, 2004), Johansson (2003)

- Higher order approximation to distribution of t-statistic

  Bentkus and Götze (1996), Bentkus, Bloznelis and Götze (1996), Bloznelis and Putter (2003), Hall and Wang (2004)

- Fixed-k inference about tail properties

  Müller and Wang (forthcoming)

- Nearly efficient tests and CIs in nonstandard problems

  Elliott, Müller and Watson (2015), Müller and Watson (2016), Müller and Norets (2016)

Outline of Talk

1. Introduction

2. Review of Extreme Value Theory

3. Review of distribution of t-statistic

4. Theory: Rates of errors in coverage probability

5. Implementation of new CI

6. Monte Carlo evidence

Review of Extreme Value Theory

- Sufficient for convergence of W_1^e to Fréchet extreme value distribution:

$$\lim_{w \to \infty} \frac{1 - F(w)}{w^{-1/\xi}} > 0$$

  $\Rightarrow$ convergence if upper tail is approximately Pareto

  $\Rightarrow$ more or less necessary

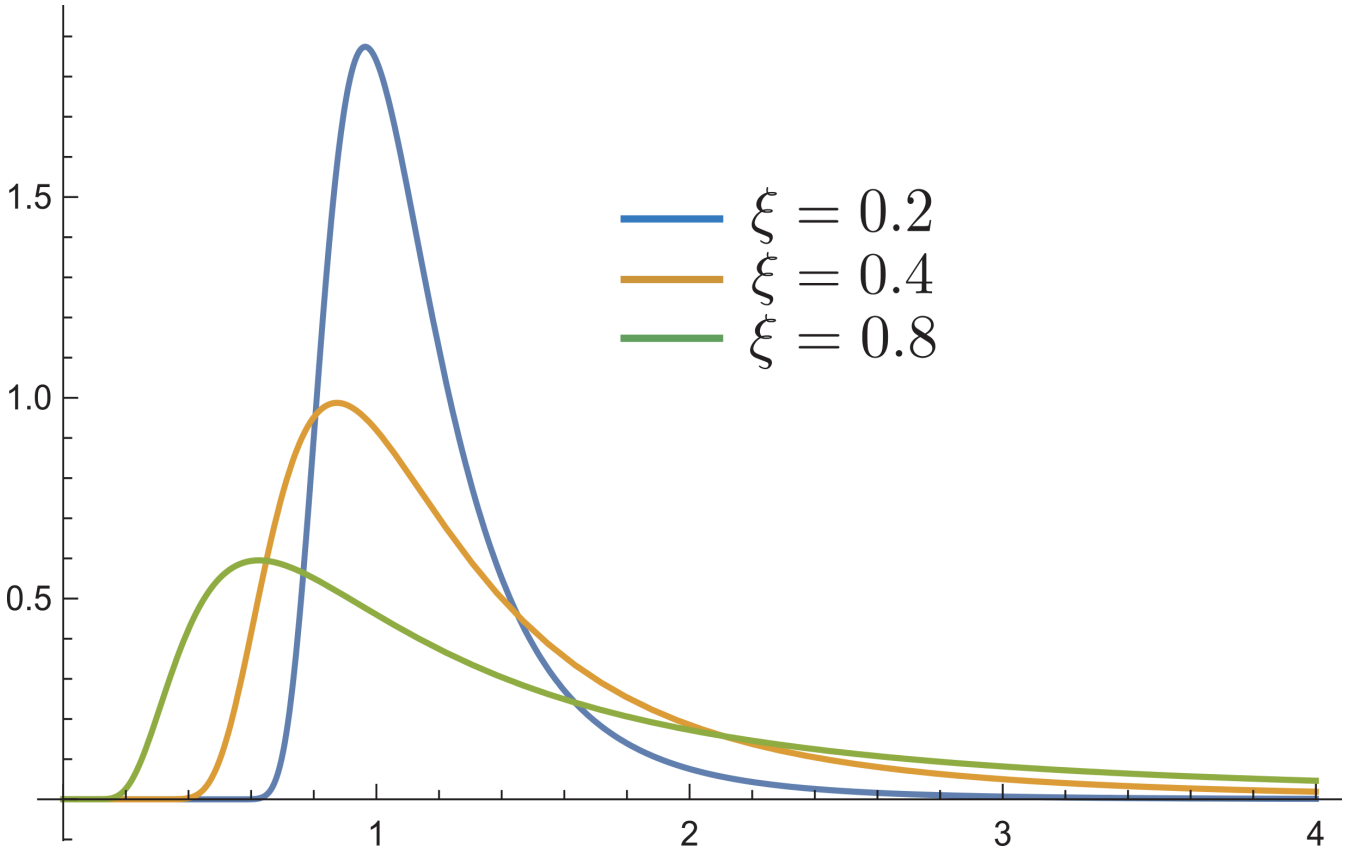
  $\Rightarrow$ student-t with df degrees of freedom induces convergence with $\xi = 1/\,\mathsf{df}$, etc.

- Then for some $\sigma > 0$

$$n^{-\xi} W_1^e \Rightarrow \sigma X_1$$

  where for $x > 0$, $P(X_1 \leq x) = G(x) = \exp(-x^{-1/\xi})$

Frechét Densities



Joint Convergence of Largest k Observations

- If $n^{-\xi}W_1^e \Rightarrow \sigma X_1$, then for any fixed k, also

$$n^{-\xi}\mathbf{W}^e = n^{-\xi}\begin{pmatrix} W_1^e \\ \vdots \\ W_k^e \end{pmatrix} \Rightarrow \sigma\mathbf{X} = \sigma\begin{pmatrix} X_1 \\ \vdots \\ X_k \end{pmatrix}$$

  where joint pdf of $\mathbf{X}$ is given by

$$G(x_k)\prod_{i=1}^{k} g(x_i)/G(x_i)$$

  with $g(x) = dG(x)/dx$

- $\mathbf{X}$ can be generated via $X_1 \sim G$, $X_2|X_1 = x_1 \sim G(x)/G(x_1)$, $X_3|X_2 = x_2, X_1 = x_1 \sim G(x)/G(x_2)$, etc.

Review: Rates of Convergence

- (Reiss, 1989): If for some w_0, F admits density for $w > w_0$ of the form

$$f(w) = \sigma^{-1}(w/\sigma)^{-1/\xi - 1}(1 + h(w))$$

with $h(w)$ uniformly bounded by $C_0 w^{-\delta/\xi}$, then

$$\sup_B |P(n^{-\xi}\mathbf{W}^e \in B) - P(\sigma\mathbf{X} \in B)| = O(n^{-\delta}).$$

Review: Distribution of t-statistic

- If $E[W] = 0$ and $E[W^2] < \infty$, then $T_n = \sqrt{n}\bar{W}_n/s_n \Rightarrow \mathcal{N}(0,1)$
- Let $T_n^*|\mathbf{W}$ be bootstrap draw of T_n conditional on $\mathbf{W} = \{W_i\}_{i=1}^n$
- Theorem (Bloznelis and Putter, 2003). If F is non-lattice and $E[|W|^3]$ exists, then

$$\sup_t |P(T_n^* < t|\mathbf{W}) - P(T_n < t)| = o(n^{-1/2}) \text{ a.s.}$$

while, for $E[W^3] \neq 0$, $\liminf_{n\to\infty} n^{1/2} \sup_t |P(T_n < t) - \Phi(t)| > 0$.

Review: Distribution of t-statistic

- Theorem (Bentkus and Götze, 1996): For some $C > 0$, and $E[W^2] = 1$,

$$\sup_t |P(T_n < t) - \Phi(t)| \leq CE[W^2 \mathbf{1}[|W| > \sqrt{n}]] + Cn^{-1/2} E[|W|^3 \mathbf{1}[|W| \leq \sqrt{n}]]$$

- Is sharp (Hall and Wang, 2004), holds uniformly in F (Bentkus, Bloznelis, Götze, 1996)

Theory Contributions

1. No bootstrap refinement if extreme value theory holds with $1/2 < \xi < 1/3$

2. Combining CLT for truncated sample with extreme value approximation for k largest observations yields refinement for $1/2 < \xi < 1/3$

Bootstrap under $1/2 < \xi < 1/3$

- Assume that for some $1/2 < \xi < 1/3$, $\lim_{w\to\infty} \frac{P(|W|>w)}{w^{-1/\xi}} > 0$, so that $|W|$ has Pareto tail with index ξ

- Theorem:

  (a) $\liminf_{n\to\infty} n^{-(1-1/(2\xi))} \sup_t |P(T_n < t) - \Phi(t)| > 0$

  (b) $n^{-3(\xi-1/2)} \sup_t |P(T_n^* < t|\mathbf{W}) - \Phi(t)| = O_p(1)$.

  $\Rightarrow$ Since $3(\xi - 1/2) > 1 - 1/(2\xi)$, also
  $\sup_t |P(T_n^* < t|\mathbf{W}) - P(T_n < t)| = O_p(n^{1-1/(2\xi)})$, so no refinement.

- Proof: (a) Follows from sharpness of Bentkus/Götze bound.

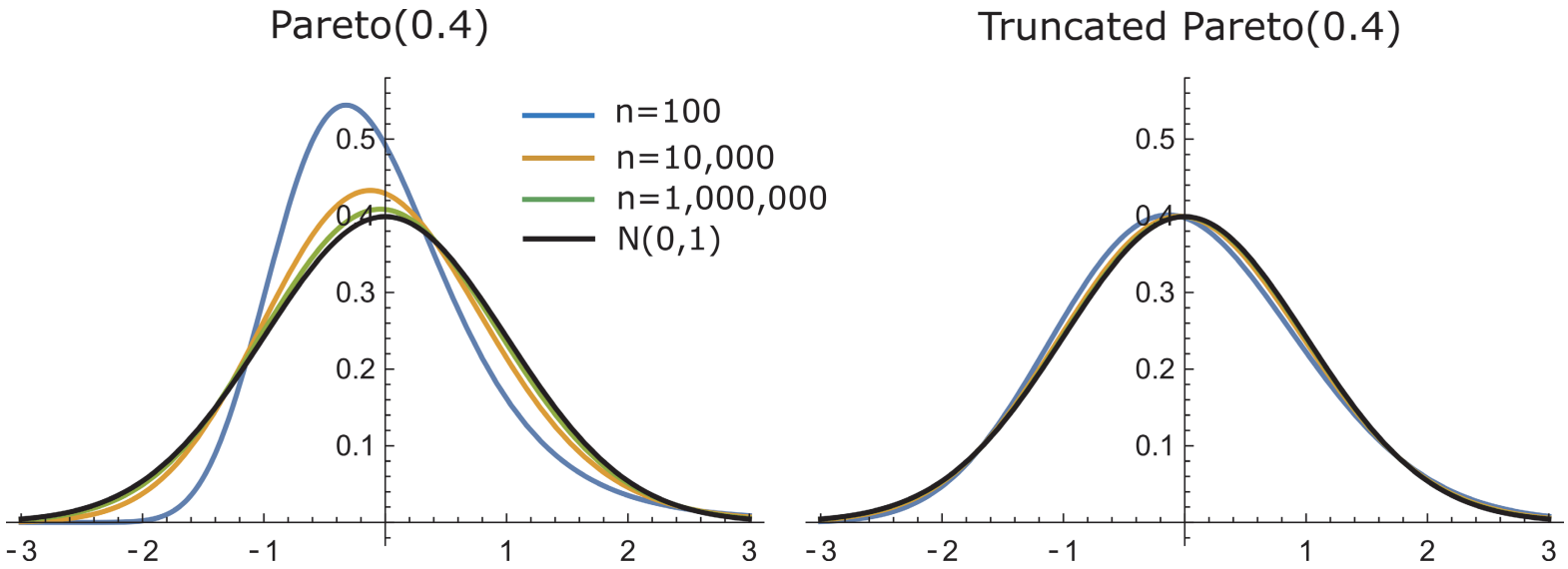
  (b) Apply Bentkus/Götze to bootstrap distribution: $n^{-1}\sum_{i=1}^n W_i^2 \mathbf{1}[|W_i| > \sqrt{n}] \xrightarrow{p} 0$ from $\max_i |W_i| = O_p(n^\xi)$, and $\sum_{i=1}^n |W_i|^3 = O_p(n^{3\xi})$.

Impact of Tail on CLT

- Roughly speaking, bootstrap distribution is generated by observing underlying distribution at $1/(n+1)$, $2/(n+1)$, $\ldots$, $n/(n+1)$ quantiles

  $\Rightarrow$ Bootstrap distribution has tail truncated at $\approx n/(n+1)$ quantile

- CLT approximations for location and scale normalized distributions:



New Asymptotic Approximation

- Suppose for $w > w_0$, F admits density of the form

$$f(w) = \sigma^{-1}(w/\sigma)^{-1/\xi-1}(1 + h(w)) \tag{1}$$

with $h(w)$ uniformly bounded by $C_0 w^{-\delta/\xi}$.

- Let $s_n^2 = (n-k)^{-1}\sum_{i=1}^{n-k}(W_i^m - \bar{W}^m)^2$. Then $s_n^2 \xrightarrow{p} E[W^2]$. By scale invariance of ultimate test, set $E[W^2] = 1$ wlog.

- Recall that from $E[W] = 0$,

$$m(w) = -E[W|W \leq w] = \frac{P(W > w)E[W|W > w]}{1 - P(W > w)}$$

$\Rightarrow$ Under (1), and for w large, $m(w) \approx M_0(w) = \sigma^{1/\xi}\frac{\xi}{1-\xi}w^{1-1/\xi}$

New Asymptotic Approximation

- Conditional on $\mathbf{W}^e$, $W_i^m \sim$iid$(-m(W_k^e), \sigma^2(W_k^e))$, where $\sigma^2(w) = V[W|W < w] \approx 1$. By CLT

$$\frac{\sum_{i=1}^{n-k} W_i^m}{\sqrt{(n-k)s_n^2}} | \mathbf{W}^e \overset{a}{\sim} \mathcal{N}\left(-\sqrt{n-k}\sigma^{1/\xi}\frac{\xi}{1-\xi}(W_k^e)^{1-1/\xi}, 1\right)$$

- From $n^{-\xi}\mathbf{W}^e \overset{a}{\sim} \sigma\mathbf{X}$,

$$\frac{\mathbf{W}^e}{\sqrt{(n-k)s_n^2}} \overset{a}{\sim} n^{-(1/2-\xi)}\sigma\mathbf{X} = \eta_n\mathbf{X}$$

- Joint approximation

$$\left(\frac{\sum_{i=1}^{n-k} W_i^m}{\sqrt{(n-k)s_n^2}}, \frac{\mathbf{W}^e}{\sqrt{(n-k)s_n^2}}\right) \overset{a}{\sim} \left(Z - \eta_n\frac{\xi}{1-\xi}X_k^{1-1/\xi}, \eta_n\mathbf{X}\right)$$

with $Z \sim \mathcal{N}(0,1)$

New Test

- Joint approximation

$$\left(\frac{\sum_{i=1}^{n-k} W_i^m}{\sqrt{(n-k)s_n^2}}, \frac{\mathbf{W}^e}{\sqrt{(n-k)s_n^2}} \right) \overset{a}{\sim} \left(Z - \eta_n \frac{\xi}{1-\xi} X_k^{1-1/\xi}, \eta_n \mathbf{X} \right)$$

- Construct interval valued function $\mathcal{I}$ such that for all $\eta > 0$ and $\xi < 1/2$

$$P\left(Z - \eta \frac{\xi}{1-\xi} X_k^{1-1/\xi} \notin \mathcal{I}(\eta \mathbf{X}) \right) \leq \alpha$$

  and then use $\mathcal{I}$ as test that rejects if

$$\frac{\sum_{i=1}^{n-k} W_i^m}{\sqrt{(n-k)s_n^2}} \notin \mathcal{I}\left(\frac{\mathbf{W}^e}{\sqrt{(n-k)s_n^2}} \right).$$

Theorem

Under (1) and some technical assumption on $\mathcal{I}$, for $k > 1$, $\frac{1+k}{1+3k} < \xi < 1/2$, $\delta \geq r_k(\xi)$ and all $\epsilon > 0$,

$$\left| P\left(\frac{\sum_{i=1}^{n-k} W_i^m}{\sqrt{(n-k)s_n^2}} \notin \mathcal{I}\left(\frac{\mathbf{W}^e}{\sqrt{(n-k)s_n^2}} \right) \right) - P\left(Z - \eta_n \frac{\xi}{1-\xi} X_k^{1-1/\xi} \notin \mathcal{I}(\eta_n \mathbf{X}) \right) \right| \leq C n^{-r_k(\xi)+\epsilon}$$

where $r_k(\xi) = \frac{3(1+k)(1-2\xi)}{2(1+k+2\xi)} > \frac{1}{2\xi} - 1$.

$\Rightarrow$ Recall $\liminf_{n\to\infty} n^{-(1-1/(2\xi))} \sup_t |P(T_n < t) - \Phi(t)| > 0$, so new approximation is refinement

$\Rightarrow$ Proof: Given Bentkus/Götze and TVD approximation of $n^{-\xi}\mathbf{W}^e \overset{a}{\sim} \sigma\mathbf{X}$, only hard part is to deal with s_n^2 (but that is very involved)

Determining $\mathcal{I}$

- Let $\theta = (\xi, \eta)$. For some given weighting function Π, consider program

$$\min_{\mathcal{I}} \int E_\theta[\text{length}(\mathcal{I}(\eta \mathbf{X}))] d\Pi(\theta)$$

$$\text{s.t. } P_\theta \left(Z - \eta \frac{\xi}{1-\xi} X_k^{1-1/\xi} \notin \mathcal{I}(\eta \mathbf{X}) \right) \leq \alpha \text{ for all } \theta \in \Theta$$

- Problem of the type considered in Elliott, Müller and Watson (2015), Müller and Watson (2016), Müller and Norets (2016)

  $\Rightarrow$ Apply numerical techniques developed there to obtain approximate solution

  $\Rightarrow$ Computationally intensive to determine $\mathcal{I}$, but not difficult to use in practice once obtained

Two Tailed Problem

- Theorem also holds for two fat tails, where now $\mathbf{W}^e = (\mathbf{W}^L, \mathbf{W}^R)$ and W_i^m are remaining $n - 2k$ middle observations

- Asymptotic approximations then become

$$\begin{pmatrix} \frac{\mathbf{W}^R}{\sqrt{(n-2k)s_n^2}} \\ \frac{-\mathbf{W}^L}{\sqrt{(n-2k)s_n^2}} \end{pmatrix} \overset{a}{\sim} \begin{pmatrix} n^{-(1/2-\xi^R)}\sigma^R\mathbf{X}^R \\ n^{-(1/2-\xi^L)}\sigma^L\mathbf{X}^L \end{pmatrix} = \begin{pmatrix} \eta_n^R\mathbf{X}^R \\ \eta_n^L\mathbf{X}^L \end{pmatrix}$$

and

$$\frac{\sum_{i=1}^{n-2k} W_i^m}{\sqrt{(n-2k)s_n^2}} | \mathbf{W}^e \overset{a}{\sim} Z - \eta_n^R \frac{\xi^R}{1-\xi^R}(X_k^R)^{1-1/\xi^R} + \eta_n^L \frac{\xi^L}{1-\xi^L}(X_k^L)^{1-1/\xi^L}$$

Small Sample Effects

1. Asymptotic approximations ignore translations of tail

   $\Rightarrow$ Unlikely to be very accurate in small samples

   $\Rightarrow$ Improve approximations via $\mathbf{W}^e \overset{a}{\sim} \eta_n(\mathbf{X} + \kappa_n)$ for suitably chosen offset κ_n

   $\Rightarrow$ Yields EVT approximations continuous in ξ also for $\xi \leq 0$ (bounded support)

   $\Rightarrow$ Leads to approximation of one tail with three parameters: mean κ, scale η and shape ξ, so total of 6 parameters

2. Better to use $m(w) \approx M_n(w)$ that doesn't impose $1 - P(W > w) \approx 1$

$\Rightarrow$ Construct $\mathcal{I} = \mathcal{I}_n$ for n moderate to small (say, smaller than 500)

Implementation

- $k = 8$, $\xi \leq 0.5$, κ_n such that

$$E[X_{20} + \kappa_n] \geq 0$$

for both tails (so that 20th order statistic has approximately non-negative expectation)

- For η and/or ξ small, bootstrap interval nearly controls size

$\Rightarrow \mathcal{I}_n$ is weighted average of nearly optimal solution and bootstrap interval, with weights determined by appropriate function of $(\mathbf{X}^L, \mathbf{X}^R)$

- For η large, uncertainty about mean of $\frac{\sum_{i=1}^{n-2k} W_i^m}{\sqrt{(n-2k)s_n^2}}$ dominates its randomness

$\Rightarrow$ transition to scale equivariant $\mathcal{I}_n$ for very large $\mathbf{X}^L$ or $\mathbf{X}^R$

Null Rejection Probabilities

	N(0,1)	LogN	F(4,5)	t(3)	P(0.4)	Mix 1	Mix 2
			$n = 50$				
t-stat	4.8	9.9	12.6	4.4	13.2	12.3	19.1
sym-boot	5.0	7.9	10.1	4.0	10.3	11.8	18.6
asym-boot	5.1	6.9	8.2	7.3	8.3	11.8	18.0
new	2.4	1.8	2.7	1.3	2.9	4.3	10.5
			$n = 100$				
t-stat	5.0	8.3	10.8	4.6	11.3	9.9	15.8
sym-boot	5.1	6.9	8.9	4.2	9.3	8.8	14.6
asym-boot	5.2	6.3	7.8	6.9	7.8	8.7	13.5
new	4.4	2.4	3.1	2.7	3.2	1.9	7.0
			$n = 500$				
t-stat	5.1	5.7	7.7	4.9	8.0	6.9	9.5
sym-boot	5.3	5.2	6.9	4.8	7.0	6.0	8.0
asym-boot	5.3	5.5	6.8	6.2	6.8	6.3	7.5
new	4.7	3.4	3.9	4.3	3.8	3.2	2.5

Normalized Average Lengths

	N(0,1)	LogN	F(4,5)	t(3)	P(0.4)	Mix 1	Mix 2
			$n = 50$				
t-stat	**1.01**	0.76	0.67	**1.02**	0.64	0.71	0.61
sym-boot	**1.00**	1.08	1.48	**1.11**	1.33	1.15	1.47
asym-boot	**1.00**	0.97	1.16	1.06	1.07	0.97	1.11
new	**1.22**	**1.63**	**1.69**	**2.02**	**1.78**	**1.50**	1.87
			$n = 100$				
t-stat	**1.00**	0.84	0.73	**1.02**	0.71	0.78	0.62
sym-boot	**0.99**	1.05	1.19	**1.09**	1.20	1.08	1.21
asym-boot	**0.99**	0.98	1.01	1.05	1.01	0.97	0.97
new	**1.04**	**1.44**	**1.31**	**1.42**	**1.27**	**1.52**	1.19
			$n = 500$				
t-stat	**1.00**	**0.97**	0.87	**1.00**	0.86	0.92	0.78
sym-boot	**0.99**	**1.02**	1.11	**1.02**	1.14	**1.01**	1.04
asym-boot	**0.99**	**1.01**	1.01	1.01	1.03	0.98	0.95
new	**1.01**	**1.25**	**1.32**	**1.15**	**1.30**	**1.31**	**1.39**

Note: Normalized by average length of size corrected t-stat

Empirical Monte Carlo

- Consider Compustat database: 495 series on large number of firms. Pooled panel 1992-2005, 59,300 observations, normalized by total assets if sensible

- At least 5000 non-missing observations, no more than 90% zeros

  $\Rightarrow$ 338 series

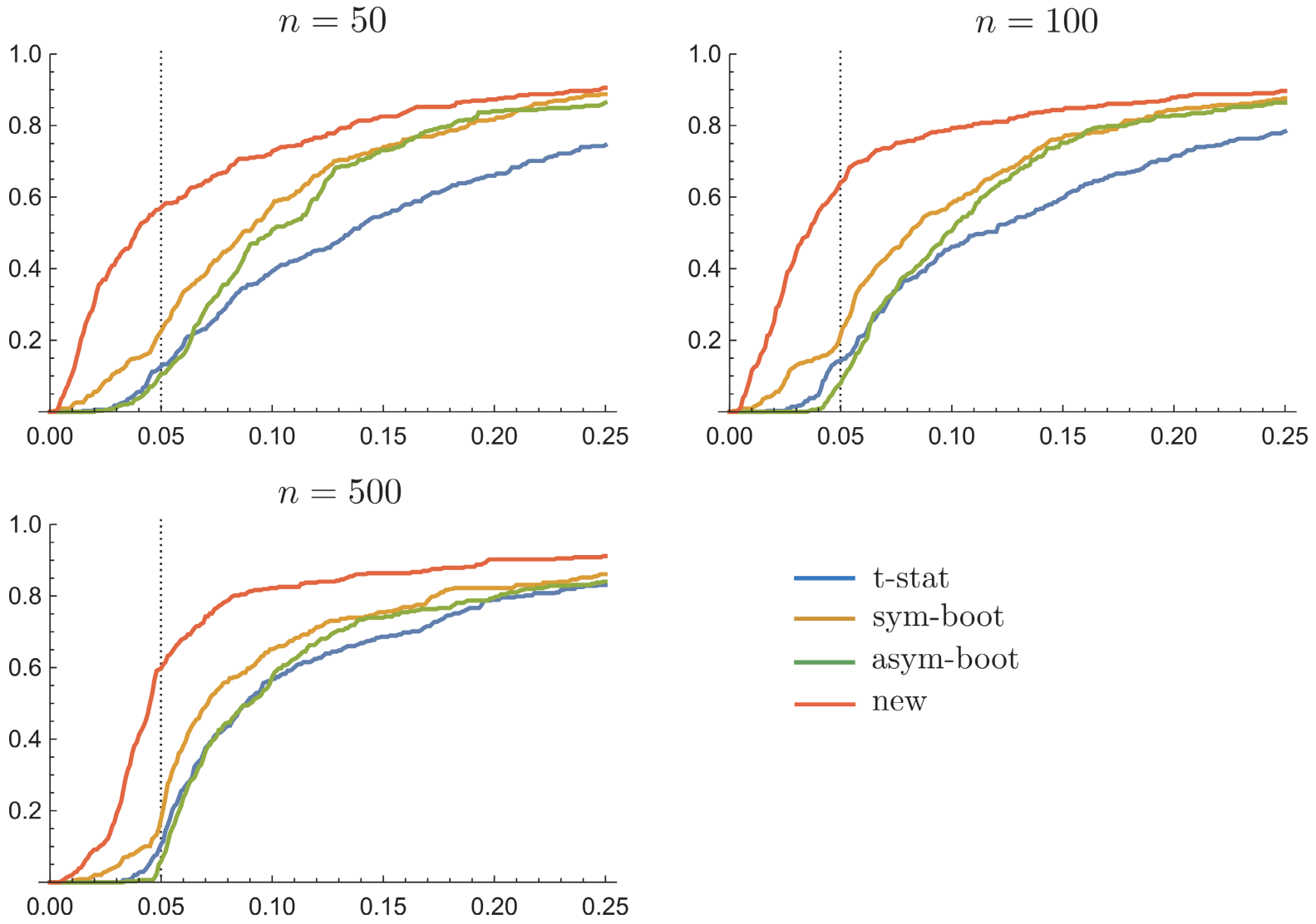
- For each series:

  - Compute mean of entire series, and treat as population mean

  - Repeat 20,000 times: Draw n data points at random, compute CIs and check whether it contains population mean

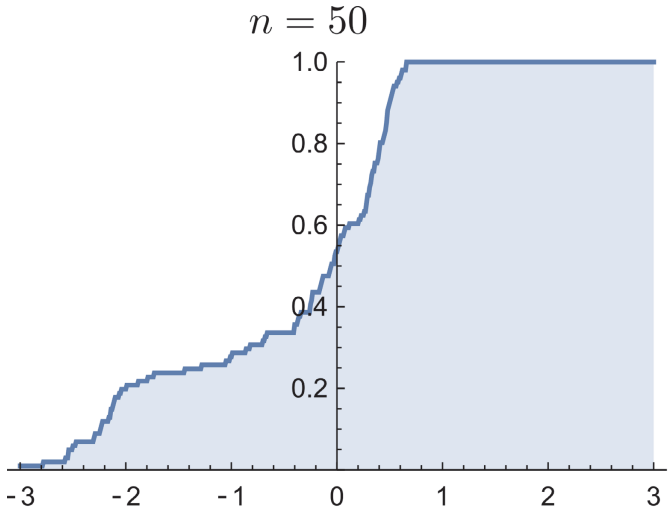
  $\Rightarrow$ Obtain 338 null rejection probabilities, and average lengths

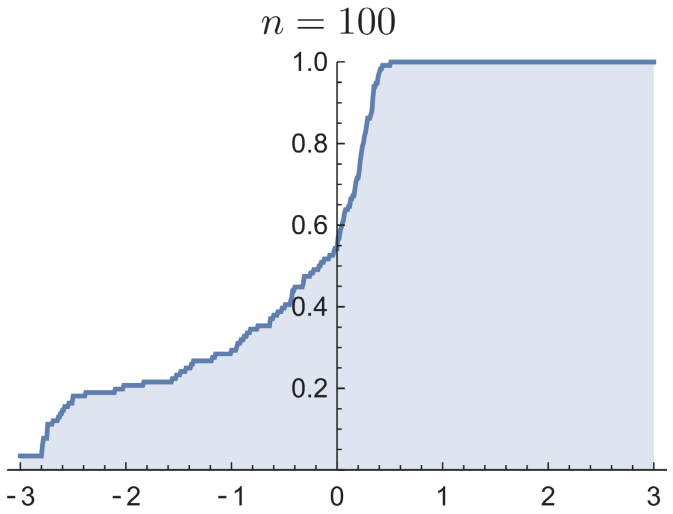
CDF of Null Rejection Probabilities

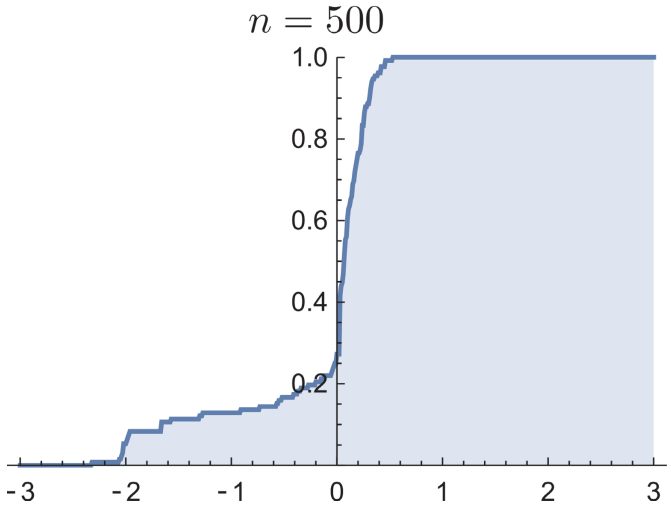


Log Ratios of Average Lengths

New relative to sym-boot, conditional on both rejection probabilities $\leq 6\%$

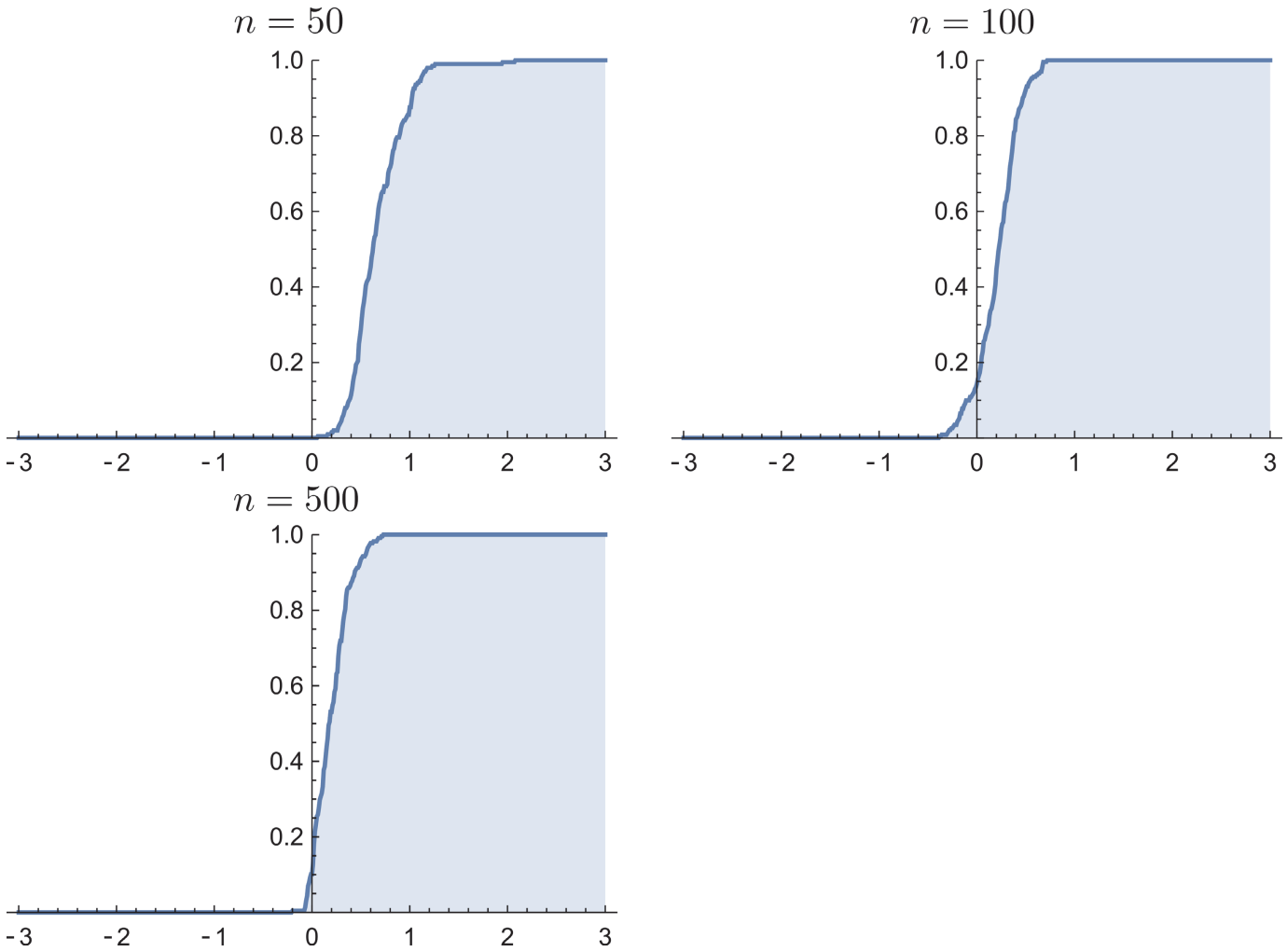






Log Ratios of Average Lengths

New relative to size-adjusted t-stat, conditional on rejection probability $\leq 6\%$



Conclusions

- New approach to inference for mean in presence of potentially fat tails
- Theory: Provides refinement under Pareto-like fat tails (more than two but less than three moments), while bootstrap does not
- Practice: Implementation that yields noticeably better size control in small to moderate samples from fat-tailed populations