Inference about Extreme Quantiles and Tail Conditional Expectations

Ulrich K. Müller and Yulong Wang
Princeton University

March 2016

Motivation

- Interest in tail event features
  - Hydrology: How high does a dam have to be to withstand a 1 in a 100 years flood?
  - Value at risk, expected shortfall
  - Insurance losses
  - Income of the very rich

Tail Quantile and TCEs

- Consider sample of n i.i.d. draws Y_i from some population with cdf F

- Object of interest $q(F, 1 - h/n)$, the $1 - h/n$ quantile of F, for fixed h

$$P(Y_i \leq q(F, 1 - h/n)) = 1 - h/n$$

$\Rightarrow$ Think of h as taking values like 0.1, 1, or 5

$\Rightarrow$ Captures idea that for a given sample size n we are interested in a quantile for which sample information is scarce

- Tail conditional expectation (TCE):

$$T(1 - h/n, F) = E[Y_i | Y_i > q(F, 1 - h/n)]$$

$\Rightarrow$ Finance: expected shortfall, average value at risk

Extreme Value Theory

- Pickands (1975) and de Haan/Balkema (1974): For extreme value theory to hold, it is necessary and sufficient for tail of F to be generalized Pareto.

  $\Rightarrow$ Parameters of interests become functions of GP location, scale and tail index ξ (at least as $n \to \infty$)

- Largest k_n observations are approximately GP for $k_n/n \to 0$

  $\Rightarrow$ Enables consistent and Gaussian estimation of GP parameters based on $k_n \to \infty$ largest observations

  $\Rightarrow$ Subsequent confidence intervals about tail properties based on delta method or similar approaches

  $\Rightarrow$ Very many suggestions in literature along these lines

Asymptotics

- But: If n is not very large, or GP approximation only good far in the tail, then using $k_n \to \infty$ asymptotics won't work well

  - Choosing k_n large mistakenly imposes GP approximation over too much of the tail of F

  - Choosing k_n small invalidates distribution theory underlying confidence interval construction

- Our approach: Derive asymptotically valid inference about tail quantile and TCE based on k largest observations, for *fixed* and given k

Fixed-k Approach

- Rely on extreme value theory for k largest order statistics $Y_{n:n}, \ldots, Y_{n:n-k+1}$

  $\Rightarrow$ k embeds extent of regularity imposed on tail

  $\Rightarrow$ small k yields very robust, but less informative inference

- Under location and scale equivariance, asymptotic problem becomes conducting inference with a sample of size k from a joint extreme value distribution

- Nonstandard problem with tail index as single nuisance parameter

  $\Rightarrow$ Apply Elliott, Müller and Watson (2015) and Müller and Watson (2015)

Plan of Presentation

1. Motivation and introduction

2. Review of extreme value theory

3. Derivation of fixed-k confidence intervals

4. Monte Carlo simulations

5. Application: cyclone damage

Review of Extreme Value Theory

- **Extremal Types Theorem** (Fisher, Tipett, Gnedenko): If for some sequences of real numbers a_n and b_n,

$$\frac{Y_{n:n} - b_n}{a_n} \Rightarrow X_1$$

  for some nondegenerate X_1, then there exists constants a and b such that $aX_1 + b$ is generalized extreme value distributed with cdf

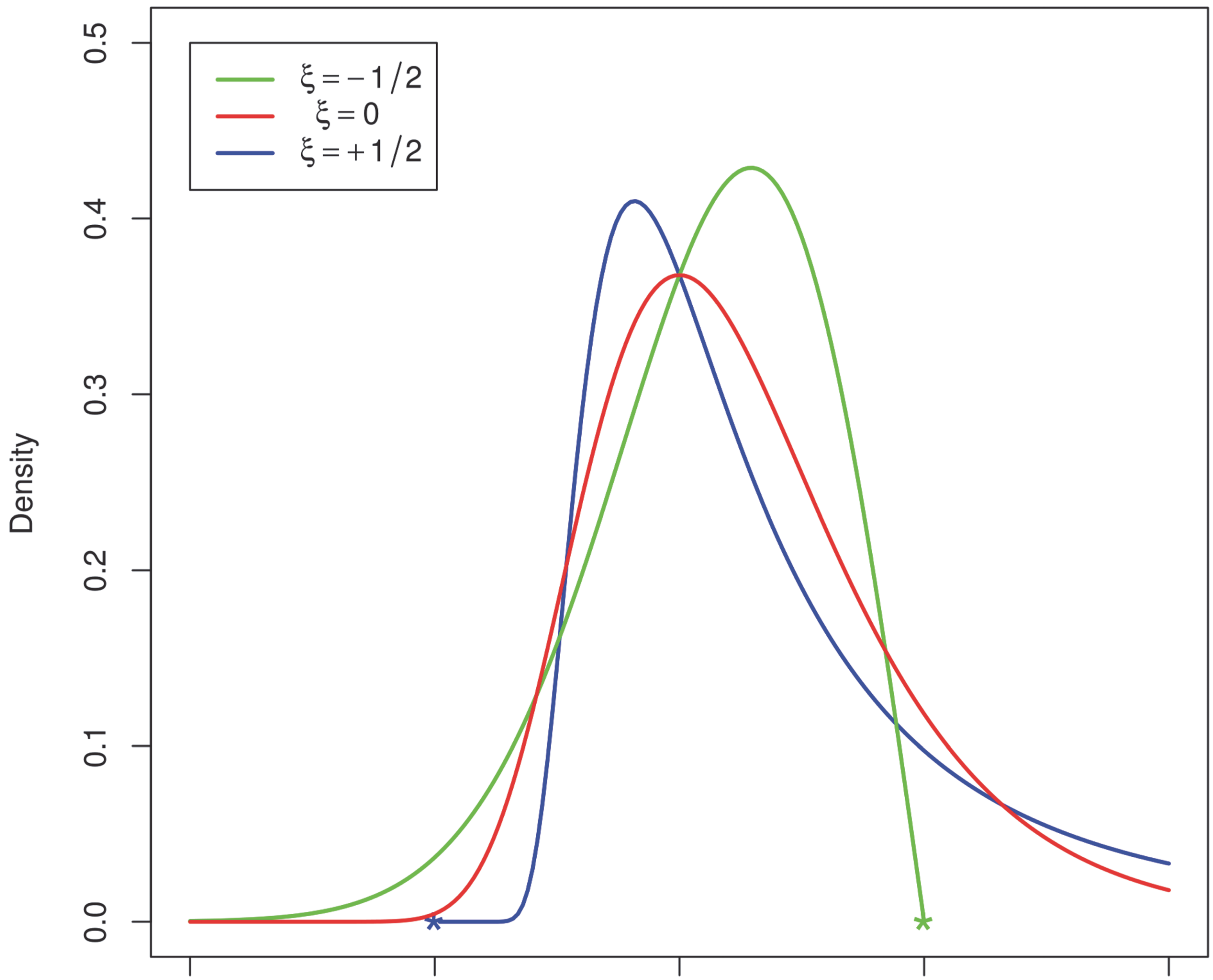
$$\begin{array}{ll} G_\xi(x) = \exp(-(1+\xi x)^{-1/\xi}), & 1+\xi x > 0,\ \xi \neq 0 \\ G_0(x) = \exp(-\exp(-x)) & x > 0 \end{array}$$

- $\xi > 0$: normalized Fréchet distribution

  $\xi = 0$: Gumbel distribution

  $\xi < 0$: normalized Weibull distribution

GEV Densities



Review of Extreme Value Theory

- Joint convergence of first k order statistics: If $\frac{Y_{n:n}-b_n}{a_n} \Rightarrow X_1$ where $X_1 \sim G_\xi$, then also

$$\begin{pmatrix} \frac{Y_{n:n}-b_n}{a_n} \\ \vdots \\ \frac{Y_{n:n-k+1}-b_n}{a_n} \end{pmatrix} \Rightarrow \mathbf{X} = \begin{pmatrix} X_1 \\ \vdots \\ X_k \end{pmatrix}$$

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

$$G_\xi(x_k) \prod_{i=1}^{k} g_\xi(x_i)/G_\xi(x_i)$$

with $g_\xi(x) = dG_\xi(x)/dx$

Our Approach: Extreme Quantile

- Object of interest is $q(1-h/n, F)$: $P(Y_i \leq q(1-h/n, F)) = 1-h/n$

- Clearly $P(Y_{n:n} \leq q(1-h/n, F)) = (1-h/n)^n \to e^{-h}$

- Assume

$$\begin{pmatrix} \frac{Y_{n:n}-b_n}{a_n} \\ \vdots \\ \frac{Y_{n:n-k+1}-b_n}{a_n} \end{pmatrix} \Rightarrow \mathbf{X} = \begin{pmatrix} X_1 \\ \vdots \\ X_k \end{pmatrix}$$

$\Rightarrow \frac{q(1-h/n,F)-b_n}{a_n}$ is approximately the e^{-h} quantile of X_1, $q_G(e^{-h}, \xi)$
$\Rightarrow$ With a_n and b_n known, asymptotic problem becomes inference about $q_G(e^{-h}, \xi)$, having observed $\mathbf{X}$ (whose distribution depends on ξ)

Location and Scale Equivariance

- Let $\mathbf{Y}_k = (Y_{n:n}, \ldots, Y_{n:n-k+1})'$. Aim: Construct confidence interval $S(\mathbf{Y}_k) \subset \mathbb{R}$ such that (at least appr. in large samples)

$$P(q(1 - h/n, F) \in S(\mathbf{Y}_k)) \geq 1 - \alpha$$

- Impose location and scale equivariance on S: $S(a\mathbf{Y}_k + b) = aS(\mathbf{Y}_k) + b$

$$\begin{aligned} P(q(1 - h/n, F) \in S(\mathbf{Y}_k)) &= P\left(\frac{q(1 - h/n, F) - b_n}{a_n} \in S\left(\frac{\mathbf{Y}_k - b_n}{a_n}\right)\right) \\ &\to P_\xi\left(q_G(\xi, h) \in S(\mathbf{X})\right) \end{aligned}$$

where $q_G(e^{-h}, \xi)$ is the e^{-h} quantile of X_1

- Asymptotic problem becomes constructing an (in some sense short) location and scale equivariant S that satisfies

$$P_\xi\left(q_G(\xi, h) \in S(\mathbf{X})\right) \geq 1 - \alpha \text{ for all } \xi \in \Xi$$

Asymptotic Small Sample Problem

- Optimization program

$$\min_{S(\cdot)} \int E_\xi[\text{lgth}(S(\mathbf{X}))]dW(\xi)$$

$$\text{s.t. } P_\xi(Y(\xi) \in S(\mathbf{X})) \geq 1 - \alpha \text{ for all } \xi \in \Xi$$

$$\text{and } S \text{ is equivariant}$$

- Essentially same problem as in Müller and Watson (2015), Elliott, Müller and Watson (2015)

  $\Rightarrow$ Use their algorithm to numerically determine appropriate S

Tail Conditional Expectation

- Can apply analogous arguments as for extreme quantiles
- Turns out: Only modification is substitution of $q_G(e^{-h}, \xi)$ by appropriate function $\tau_G(e^{-h}, \xi)$

Implementation

- $\Xi = [-1/2, 1/2]$

  $\Rightarrow$ For $\xi \geq 1/2$, variance of F ceases to exist, so relatively agnostic for most applications

- Density of weight function dW inversely proportional to expected length of optimal $S(\mathbf{X})$ for known $\xi \in \Xi$

  $\Rightarrow$ Minimize simple average of "regret" of not knowing ξ

- Derive optimal S for various values of h and k

Monte Carlo

- $n = 250$, $\alpha = 0.05$

- Six distributions: Normal, Log-Normal, Student-t(3), F(4,4), 0.8 / 0.2 mixture between Normal and Student-t(3), Triangular

- Four confidence intervals:

  1. Fixed-k
  2. Weissmann (1978) with Hill (1975) estimator of ξ
  3. de Haan and Ferreira (2007) textbook method
  4. Profile likelihood of Davison and Smith (1990)

Quantile Coverage for $h = 0.1$

k	10	20	50	10	20	50
	Normal			Student-t(3)		
Fixed-k	95.3	94.5	91.4	97.5	94.6	91.2
W-H	99.2	77.2	0.00	94.3	98.8	100
dH-F	85.0	83.8	74.8	66.0	76.2	82.0
profile	97.5	92.2	88.1	89.1	90.2	89.1
	Log-Normal			F(4,4)		
Fixed-k	97.5	95.8	96.1	95.2	95.0	95.2
W-H	96.7	99.5	99.9	92.8	96.7	99.9
dH-F	70.9	82.2	91.5	65.1	75.7	87.3
profile	90.6	91.9	94.7	84.4	89.2	91.9
	Mixture			Triangular		
Fixed-k	97.4	87.4	70.4	95.4	95.5	95.1
W-H	82.4	94.3	80.4	97.4	13.1	0.00
dH-F	57.9	60.1	50.8	99.0	99.8	99.9
profile	85.5	82.9	68.3	100	100	100

Quantile Coverage for $h = 5$

k	10	20	50	10	20	50
	Normal			Student-t(3)		
Fixed-k	94.8	95.9	94.5	95.5	95.9	95.4
W-H	66.4	92.2	98.1	71.2	90.8	92.8
dH-F	24.6	66.7	89.2	26.7	66.7	90.5
profile	93.5	93.5	93.6	90.6	93.0	93.7
	Log-Normal			F(4,4)		
Fixed-k	95.2	95.4	95.5	95.9	95.1	95.4
W-H	70.2	90.0	98.7	61.9	79.8	42.4
dH-F	26.9	65.4	89.4	28.1	65.9	86.9
profile	89.4	92.1	94.0	90.0	91.4	92.9
	Mixture			Triangular		
Fixed-k	96.3	96.1	92.3	95.2	96.0	95.2
W-H	70.5	91.4	98.7	62.7	94.4	92.7
dH-F	27.8	68.1	89.2	25.4	69.2	94.2
profile	92.8	93.4	90.7	95.8	98.4	98.3

TCE Coverage for $h = 0.1$

k	10	20	50	10	20	50
	Normal			Student-t(3)		
Fixed-k	94.4	95.2	90.5	95.9	94.1	89.8
W-H	84.9	24.0	0.00	92.8	98.2	89.4
dH-F	76.2	76.0	68.6	55.9	66.4	72.1
profile	98.3	94.0	87.5	89.6	90.0	89.5
	Log-Normal			F(4,4)		
Fixed-k	96.2	96.0	96.3	94.5	95.4	94.9
W-H	96.2	93.0	18.8	88.1	94.9	99.8
dH-F	63.2	76.1	87.9	56.0	68.3	82.7
profile	91.4	92.2	94.5	85.0	88.8	92.0
	Mixture			Triangular		
Fixed-k	96.6	87.7	68.3	95.6	95.5	95.4
W-H	74.4	88.8	21.6	37.5	0.26	0.00
dH-F	50.9	50.0	37.4	98.8	99.5	99.8
profile	87.1	81.9	63.5	100	100	100

TCE Coverage for $h = 5$

k	10	20	50	10	20	50
	Normal			Student-t(3)		
Fixed-k	96.1	95.4	92.9	97.8	96.0	93.9
W-H	56.4	65.7	0.01	48.8	76.1	97.1
dH-F	19.5	55.4	69.0	16.3	44.6	70.8
profile	94.1	92.7	91.7	89.8	90.9	91.8
	Log-Normal			F(4,4)		
Fixed-k	97.0	96.2	96.0	94.9	95.2	95.1
W-H	50.4	75.0	94.2	43.4	62.7	20.4
dH-F	17.5	48.2	77.9	15.9	48.5	75.6
profile	90.1	92.1	94.9	85.6	89.2	92.0
	Mixture			Triangular		
Fixed-k	97.4	92.7	84.1	94.9	95.3	95.2
W-H	42.1	69.1	17.8	65.4	51.2	0.08
dH-F	14.1	37.7	54.4	26.3	67.5	90.1
profile	87.5	85.9	82.2	100	99.8	99.4

Application

- Damage of mainland U.S. tropical cyclones 1995-2010

- Unknown n. But $q(F, 1 - h/n)$ still interpretable: hurricane as damaging occurs every $16/h$ years on average

- Worst 10 cyclones damage in 2010 $US billion (data from National Weather Service)

  105.8 27.8 20.6 19.8 15.8 11.8 11.0 10.0 9.2 8.1

- 95% Confidence Intervals ($k = 10$)

	$h = 0.1$		$h = 1$		$h = 5$	
Quantile	[33.9	589.7]	[13.6	139.1]	[8.1	30.0]
TCE	[49.2	1282.2]	[23.5	337.6]	[12.5	118.1]

Conclusions

- Method to conduct inference about tail properties under relatively weak assumptions

- Theorem in paper that formalizes additional robustness of fixed-k approach

- Extensions/further applications:
  - Confidence interval for tail index
  - Estimation of quantiles/TCE (VaR applications)
  - Stability of tail properties in time
  - Test of stochastic dominance in tail