Taylor & Francis
Taylor & Francis Group

Journal of the American Statistical Association

ISSN: 0162-1459 (Print) 1537-274X (Online) Journal homepage: http://www.tandfonline.com/loi/uasa20

Fixed-k Asymptotic Inference About Tail Properties

Ulrich K. Müller & Yulong Wang

To cite this article: Ulrich K. Müller & Yulong Wang (2017) Fixed-k Asymptotic Inference About Tail Properties, Journal of the American Statistical Association, 112:519, 1334-1343, DOI: 10.1080/01621459.2016.1215990

To link to this article: https://doi.org/10.1080/01621459.2016.1215990

Accepted author version posted online: 12 Aug 2016.
Published online: 13 Jun 2017.

JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION
2017, VOL. 112, NO. 519, 1334–1343, Theory and Methods
https://doi.org/10.1080/01621459.2016.1215990

Fixed-*k* Asymptotic Inference About Tail Properties

Ulrich K. Müller and Yulong Wang

Department of Economics, Princeton University, Princeton, NJ

ABSTRACT
We consider inference about tail properties of a distribution from an iid sample, based on extreme value theory. All of the numerous previous suggestions rely on asymptotics where eventually, an infinite number of observations from the tail behave as predicted by extreme value theory, enabling the consistent estimation of the key tail index, and the construction of confidence intervals using the delta method or other classic approaches. In small samples, however, extreme value theory might well provide good approximations for only a relatively small number of tail observations. To accommodate this concern, we develop asymptotically valid confidence intervals for high quantile and tail conditional expectations that only require extreme value theory to hold for the largest k observations, for a given and fixed k. Small-sample simulations show that these "fixed-k" intervals have excellent small-sample coverage properties, and we illustrate their use with mainland U.S. hurricane data. In addition, we provide an analytical result about the additional asymptotic robustness of the fixed-k approach compared to $k_n \to \infty$ inference.

ARTICLE HISTORY
Received February 2016
Accepted July 2016

KEYWORDS
Extreme quantiles; tail conditional expectations; extreme value distribution

1. Introduction

Tail properties of distributions have been an important and ongoing empirical and theoretical issue. A large literature is devoted to exploiting implications of extreme value theory and tail regularity conditions for purposes of estimation and inference: see Embrechts, Klüppelberg, and Mikosch (1997), Coles (2001), Reiss and Thomas (2007), Resnick (2007), de Haan and Ferreira (2007), Beirlant, Caeiro, and Gomes (2012) and Gomes and Guillou (2015) for reviews and references. As demonstrated by Balkema and de Haan (1974) and Pickands (1975), a distribution F is in the domain of attraction of an extreme value distribution with tail index ξ if and only if its tail is well approximated by a generalized Pareto distribution with shape parameter ξ. This regularity assumption about the tail implies that common tail properties of interest, such as tail probabilities, high quantiles, or tail conditional expectations are functions of ξ and the location and scale of the generalized Pareto distribution, at least approximately. Corresponding estimators of tail properties can hence be constructed by plugging in estimators of these three parameters. The literature contains numerous suggestions along those lines, reviewed in the surveys mentioned above.

A common feature of all of these approaches is that they model an increasing number $k_n \to \infty$, $k_n/n \to 0$ of tail observation in a sample of size n as stemming from the approximate generalized Pareto tail. Indeed, under additional regularity conditions, tail index estimators $\hat{\xi}$ can typically be shown to be consistent and asymptotically Gaussian, enabling the construction of confidence intervals about tail properties using variants of the delta method or similar approaches. For these asymptotics to provide a good approximation, however, k_n has to satisfy a delicate balance: on the one hand, picking k_n small invalidates the asymptotic approximation of consistency and/or Gaussianity of $\hat{\xi}$, and thus undermines the distribution theory that justifies the confidence interval construction. On the other hand, picking k_n large amounts to imposing the generalized Pareto approximation on a relatively large fraction of the distribution, especially if n is only moderately large. This can be a poor approximation, even for underlying F in the domain of attraction of an extreme value distribution, leading to substantial bias and undersized confidence intervals. As such, for some combinations of n and F, no choice of k_n leads to satisfactory inference derived under $k_n \to \infty$ asymptotics.

This article develops an alternative asymptotic embedding to address this issue. In particular, we derive asymptotically valid inference under the sole assumption that extreme value theory applies to the k largest observations, for given and fixed k. This should yield good small-sample approximations whenever (a little more than) a fraction k/n of the tail of F is well approximated by a generalized Pareto distribution. These "fixed-k" asymptotics reflect the small-sample nature of the inference problem in the sense that only relatively few observations are assumed to stem from the nicely behaved tail, while previous approaches crucially exploit that the number of such observations is (eventually) large. Our approach is close in spirit to recently developed alternative asymptotic embeddings in other contexts, such as the "fixed-b" asymptotics considered by Kiefer and Vogelsang (2005) in the context of heteroskedasticity and autocorrelation robust inference, or the small bandwidth asymptotics in nonparametric kernel estimation studied by Cattaneo, Crump and Jansson (2010, 2014).

More specifically, we focus in this article on the construction of confidence intervals for the $1 - h/n$ quantile, for given h, and corresponding tail conditional expectation from an iid sample. By restricting attention to scale and location equivariant

CONTACT Ulrich K. Müller umueller@princeton.edu Department of Economics, Princeton University, Princeton, NJ 08544.

intervals, the asymptotic problem under fixed- k extreme value theory becomes a reasonably transparent parametric small-sample problem: Given a single k-dimensional draw from the joint extreme value distribution (which is indexed only by the scalar ξ), construct an equivariant confidence interval for a specific deterministic function of ξ and h. One straightforward construction is obtained by inverting a generalized likelihood ratio statistic. A more systematic approach along the lines of Elliott, Müller, and Watson (2015) and Müller and Norets (2016) minimizes weighted expected length (over ξ) subject to the coverage constraint.

Section 2 contains the details of the corresponding derivations. Monte Carlo simulations in Section 3 show that these new confidence intervals have excellent small-sample coverage and length properties for moderately large k and n compared to previous confidence interval constructions. We illustrate the application of the new intervals with data on mainland U.S. Hurricane damage in Section 4. Finally, in Section 5, we conclude with a theorem that provides an analytical sense in which the fixed-k approach leads to more robust large sample inference compared to potentially more informative $k_n \to \infty$ approaches.

2. Derivation of Fixed-*k* Inference

2.1. High Quantile

We observe a random sample $Y_1, Y_2, \ldots, Y_n$ from some population with cumulative distribution function F. Write $U(F, p)$ for the $1 - 1/p$ quantile of F. We initially consider inference about the quantile $U(F, n/h)$ for given h. The objective is to construct an asymptotically valid confidence interval for $U(F, n/h)$ of level $1 - \alpha$, where h does not vary with n (so the quantile of interest is of the same order of magnitude as the sample maximum).

Let $Y_{n:1} \leq Y_{n:2} \leq \cdots \leq Y_{n:n}$ denote the order statistics, so that $Y_{n:n}$ is the sample maximum. The fundamental result in extreme value theory due to Fisher and Tippett (1928) and Gnedenko (1943) states that if there exist sequences a_n and b_n such that

$$\frac{Y_{n:n} - b_n}{a_n} \Rightarrow X_1 \tag{1}$$

for some nondegenerate random variable X_1, then the distribution of X_1 is, up to location and scale normalization, the generalized extreme value distribution with c.d.f.

$$G_\xi(x) = \begin{cases} \exp[-(1+\xi x)^{-1/\xi}],\ 1+\xi x \geq 0, \text{ for } \xi \neq 0 \\ \exp[-e^{-x}],\ x \in \mathbb{R}, \xi = 0 \end{cases} \tag{2}$$

where $\xi < 0$, $\xi = 0$ and $\xi > 0$ correspond to Weibull, Gumbel and Fréchet type extreme value distributions, respectively. Without loss of generality, assume that any location and scale normalization of X_1 is subsumed in a_n and b_n, so that the c.d.f. of X_1 in (1) is equal to G_ξ. It is well known (see, for instance, Theorem 3.5 of Coles (2001)) that if (1) holds, then extreme value theory also holds jointly for the first k-order statistics

$$\begin{pmatrix} \frac{Y_{n:n}-b_n}{a_n} \\ \vdots \\ \frac{Y_{n:n-k+1}-b_n}{a_n} \end{pmatrix} \Rightarrow \mathbf{X} = \begin{pmatrix} X_1 \\ \vdots \\ X_k \end{pmatrix} \tag{3}$$

for any fixed k, where the joint p.d.f. of $\mathbf{X}$ is given by $f_{\mathbf{X}|\xi}(x_1, \ldots, x_k) = G_\xi(x_k) \prod_{i=1}^k g_\xi(x_i)/G_\xi(x_i)$ on $x_k \leq x_{k-1} \leq \cdots \leq x_1$ with $g_\xi(x) = dG_\xi(x)/dx$, and zero otherwise.

We seek to construct an inference method that yields valid asymptotic inference whenever (3) holds, for some given finite k, to better approximate the small-sample effect of having a finite number of order statistics that are reasonably modeled as stemming from the generalized Pareto tail. Thus, the effective data becomes $\mathbf{Y}_k = (Y_{n:n}, \ldots, Y_{n:n-k+1})'$, and the objective is to construct a confidence set $S(\mathbf{Y}_k) \subset \mathbb{R}$ such that $P(U(F, n/h) \in S(\mathbf{Y}_k)) \geq 1 - \alpha$, at least as $n \to \infty$.

By definition of $U(F, n/h)$, $P(Y_{n:n} \leq U(F, n/h)) = (1 - h/n)^n \to e^{-h}$. Thus, under (3), $(U(F, n/h) - b_n)/a_n$ converges to the e^{-h} quantile of X_1, denoted $q(\xi, h)$ in the following. A straightforward calculation shows $q(\xi, h) = (h^{-\xi} - 1)/\xi$ for $\xi \neq 0$ and $q(0, h) = -\log(h)$. If a_n and b_n were known, the asymptotic problem would hence become inference about $q(\xi, h)$ based on the $k \times 1$ vector of observations $\mathbf{X}$. But a_n and b_n are not known, and they depend on tail properties of F themselves. To make further progress, we hence impose location and scale equivariance on the confidence set S. Specifically, we impose that for any constants $a > 0$ and b, $S(a\mathbf{Y}_k + b) = aS(\mathbf{Y}_k) + b$, where $aS(\mathbf{Y}_k) + b = \{y : (y - b)/a \in S(\mathbf{Y}_k)\}$. Under this equivariance constraint, we can write

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

where the notation P_ξ indicates that the probability of the event $q(\xi, h) \in S(\mathbf{X})$ for any given set valued function S only depends on ξ. Let $\Xi \subset \mathbb{R}$ be the set of tail indices for which we impose asymptotically valid inference. The asymptotic problem then is the construction of a location and scale equivariant S that satisfies

$$P_\xi\left(q(\xi, h) \in S(\mathbf{X})\right) \geq 1 - \alpha \text{ for all } \xi \in \Xi \tag{4}$$

since any S that satisfies (4) also satisfies $\liminf_{n\to\infty} P(U(F, n/h) \in S(\mathbf{Y}_k)) \geq 1 - \alpha$ under (3). This is a fairly transparent small-sample problem, involving a single observation $\mathbf{X} \in \mathbb{R}^k$ from a parametric distribution indexed only by the scalar parameter $\xi \in \Xi$.

One straightforward construction of S is based on the inversion of a generalized likelihood ratio statistic: let $L_{\tilde{\mathbf{X}}}(\mu, \sigma, \xi) = \log(f_{\mathbf{X}|\xi}(\frac{X_1-\mu}{\sigma}, \ldots, \frac{X_k-\mu}{\sigma})) - k\log(\sigma)$ be the log-likelihood of the $k \times 1$ random vector $\tilde{\mathbf{X}} = \mu + \sigma\mathbf{X}$. To test the null hypothesis $H_0 : q(\xi, h) = q_0$ based on $\mathbf{X}$, consider the test statistic

$$\begin{aligned} \mathrm{LR}(q_0, \mathbf{X}) = &\max_{\{(\mu,\sigma,\xi):\sigma>0,\xi\in\Xi\}} L_{\tilde{\mathbf{X}}}(\mu, \sigma, \xi) \\ &- \max_{\{(\mu,\sigma,\xi):\mu+\sigma q(\xi,h)=q_0,\sigma>0,\xi\in\Xi\}} L_{\tilde{\mathbf{X}}}(\mu, \sigma, \xi). \end{aligned}$$

Clearly, $\mathrm{LR}(aq_0 + b, a\mathbf{X} + b) = \mathrm{LR}(q_0, \mathbf{X})$ for all $a > 0$ and $b \in \mathbb{R}$, so the set

$$S_{\mathrm{LR}}(\mathbf{X}) = \{q_0 : \mathrm{LR}(q_0, \mathbf{X}) < \mathrm{cv}_{\mathrm{LR}}\} \tag{5}$$

is equivariant. To ensure (4) holds for $S = S_{\mathrm{LR}}$, $\mathrm{cv}_{\mathrm{LR}}$ here is chosen to solve $\sup_{\xi\in\Xi} P_\xi(\mathrm{LR}(q(\xi, h), \mathbf{X}) > \mathrm{cv}_{\mathrm{LR}}) = \alpha$.

A more systematic approach to confidence interval construction seeks to minimize the weighted average expected length criterion

$$\int E_\xi[\mathrm{lgth}(S(\mathbf{X}))]dW(\xi), \tag{6}$$

where W is positive measure with support on Ξ, and $\mathrm{lgth}(A) = \int \mathbf{1}[y \in A]dy$ for any Borel set $A \subset \mathbb{R}$. To solve the program of minimizing (6) subject to (4) among all equivariant set estimators S, introduce

$$\mathbf{X}^s = \left(\frac{X_1 - X_k}{X_1 - X_k}, \frac{X_2 - X_k}{X_1 - X_k}, \ldots, \frac{X_k - X_k}{X_1 - X_k}\right)$$

and

$$Y^s(\xi) = \frac{q(\xi, h) - X_k}{X_1 - X_k}.$$

In this notation, equivariance of S implies $E_\xi[\mathrm{lgth}(S(\mathbf{X}))] = E_\xi[(X_1 - X_k)\mathrm{lgth}(S(\mathbf{X}^s))] = E_\xi[\kappa_\xi(\mathbf{X}^s)\mathrm{lgth}(S(\mathbf{X}^s))]$ with $\kappa_\xi(\mathbf{X}^s) = E_\xi[X_1 - X_k|\mathbf{X}^s]$, and $P_\xi\left(q(\xi, h) \in S(\mathbf{X})\right) = P_\xi(Y^s(\xi) \in S(\mathbf{X}^s))$. Thus, the program of minimizing (6) subject to (4) among all equivariant set estimators S equivalently becomes

$$\min_{S(\cdot)} \int E_\xi[\kappa_\xi(\mathbf{X}^s)\,\mathrm{lgth}(S(\mathbf{X}^s))]dW(\xi)$$
$$\text{s.t.} \quad P_\xi(Y^s(\xi) \in S(\mathbf{X}^s)) \geq 1 - \alpha \text{ for all } \xi \in \Xi. \tag{7}$$

Note that (7) only involves S evaluated at $\mathbf{X}^s$, which is an element of a smaller dimensional subspace compared to $\mathbf{X}$ (since the first and last elements of $\mathbf{X}^s$ are always equal to 1 and 0, respectively). But any solution to (7) also provides the form of S for unconstrained $\mathbf{X}$ via equivariance, $S(\mathbf{X}) = (X_1 - X_k)S(\mathbf{X}^s) + X_k$.

By writing the expectations in (7) as integrals over the densities $f_{\mathbf{X}^s|\xi}$ of $\mathbf{X}^s$ and $f_{Y^s(\xi),\mathbf{X}^s|\xi}$ of $(Y^s(\xi), \mathbf{X}^s)$, it is not very difficult to see that its solution is

$$S_\Lambda(\mathbf{X}^s) = \left\{y : \int \kappa_\xi(\mathbf{X}^s)f_{\mathbf{X}^s|\xi}(\mathbf{X}^s)dW(\xi) < \int f_{Y^s(\xi),\mathbf{X}^s|\xi}(y, \mathbf{X}^s)d\Lambda(\xi)\right\} \tag{8}$$

provided the nonnegative measure Λ has support on $\{P_\xi(Y^s(\xi) \in S_\Lambda(\mathbf{X}^s)) = 1 - \alpha\} \subset \Xi$, and $S_\Lambda(\mathbf{X}^s)$ satisfies the constraint in (7).* The only remaining challenge is thus to identify suitable Lagrangian weights Λ, and to this end we resort to the numerical approach developed in Elliott, Müller, and Watson (2015) and Müller and Watson (2015). The output of these numerical approximations is a level $1 - \alpha$ equivariant set that has, by construction, W-weighted average expected length that is no more than 1% longer than any other level $1 - \alpha$ equivariant set (that is, any other equivariant set satisfying (4)). Further details are provided in the Appendix.

2.2. Tail Conditional Expectation

Now consider the problem of constructing an asymptotically valid fixed-k confidence interval for the tail conditional expectation $T_n = E[Y_i|Y_i \geq U(F, n/h)]$, for given h, where Y_i is iid with c.d.f. F. Assume F is in the domain of attraction with tail index $\xi < 1$ (otherwise, the tail conditional expectation does not exist). Recall that for a positive random variable Z with c.d.f. F_Z, $E[Z] = \int_0^\infty (1 - F_Z(z))dz$. Thus

$$\frac{T_n - b_n}{a_n} = \frac{U(F, n/h) - b_n}{a_n} + h^{-1}n \int_{\frac{U(F,n/h)-b_n}{a_n}}^{\infty} (1 - F(a_n y + b_n))dy.$$

As noted before, a necessary condition for F to be in the domain of attraction of an extreme value distribution with index ξ is that its tail is approximately generalized Pareto. This can be written in the form

$$n(1 - F(a_n y + b_n)) \to (1 + \xi y)^{-1/\xi} \text{ for all } y \text{ such that } 1 + \xi y > 0$$

as in Theorem 1.1.6 of de Haan and Ferreira (2007) (with $(1 + \xi y)^{-1/\xi}$ interpreted as e^{-y} for $\xi = 0$). Furthermore, as in Section 2.1 above, $(U(F, n/h) - b_n)/a_n \to q(\xi, h)$. Putting this together yields the convergence†

$$\frac{T_n - b_n}{a_n} \to q(\xi, h) + h^{-1} \int_{\{y: y > q(\xi,h) \text{ and } (1+\xi y)>0\}} (1 + \xi y)^{-1/\xi} dy$$
$$= \frac{h^{-\xi}}{\xi(1 - \xi)} - \frac{1}{\xi} \equiv \tau(\xi, h).$$

Thus, the only difference between fixed-k asymptotic equivariant inference about the quantile $U(F, n/h)$ and the tail conditional expectation T_n is that in the resulting small-sample problem involving $\mathbf{X}$, the object of interest is $\tau(\xi, h)$, rather than $q(\xi, h)$. The two approaches described in Section 2.1 to the construction of asymptotically valid confidence sets thus readily carry over to tail conditional expectations.

2.3. Implementation and Asymptotic Properties

The suggested confidence sets require a choice for the parameter space of the tail index Ξ, and, for S_Λ, a weight function W. We choose $\Xi = [-1/2, 1/2]$, although the methods could equally well be implemented for other Ξ (for S_Λ, all values in Ξ have to be smaller than one for the expected length to exist, though). Since $\xi < 1/2$ is necessary for F to have a finite second moment, $\Xi = [-1/2, 1/2]$ should be fairly agnostic for most applications.

For the weighting function W, we choose a density on Ξ that is inversely proportional to the minimal expected length of an equivariant confidence set under $\xi \in \Xi$ known. The idea is to use the known tail index case as a benchmark for the relative difficulty of obtaining informative inference. The inverse weighting then puts equal weight on all $\xi \in \Xi$ relative to this benchmark, so loosely speaking, we seek to minimize the average regret of not knowing ξ. This weighting scheme has the additional advantage that the (essentially arbitrary) scale normalization for different ξ of the extreme value distributions in (2) plays no role in the determination of S_Λ.

* Theorem 3 in Müller and Norets (2016) provides a corresponding formal result.

† The convergence formally follows from the dominated convergence theorem for $\xi < 0$ (since F then has bounded support), and by Karamata's Theorem for $\xi > 0$, as in Equation (1.4) of Zhu and Li (2012).

Table 1. Asymptotic lengths of fixed-k confidence intervals.

h	0.1					5				
ξ	−0.5	−0.25	0.0	0.25	0.5	−0.5	−0.25	0.0	0.25	0.5
					Quantile, $k=10$					
Fixed-k LR	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Fixed-k opt	0.89	0.91	0.95	1.04	1.16	1.02	1.00	0.98	0.97	0.97
Fixed-k env	0.80	0.90	0.92	0.92	0.89	0.98	0.98	0.98	0.96	0.95
					Quantile, $k=20$					
Fixed-k LR	0.68	0.90	0.95	0.84	0.69	1.00	0.98	0.97	0.90	0.78
Fixed-k opt	0.47	0.73	0.91	0.91	0.85	0.93	0.92	0.92	0.89	0.82
Fixed-k env	0.41	0.70	0.86	0.77	0.60	0.89	0.91	0.90	0.87	0.74
					Quantile, $k=50$					
Fixed-k LR	0.23	0.45	0.82	0.76	0.49	0.97	0.94	0.97	0.85	0.60
Fixed-k opt	0.17	0.30	0.67	0.76	0.57	0.88	0.89	0.88	0.81	0.65
Fixed-k env	0.16	0.29	0.63	0.71	0.43	0.86	0.87	0.84	0.78	0.56
					Tail conditional expectation, $k=10$					
Fixed-k LR	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Fixed-k opt	0.86	0.90	0.96	1.06	1.19	0.94	0.92	0.93	0.99	1.09
Fixed-k env	0.78	0.89	0.92	0.92	0.89	0.79	0.88	0.91	0.93	0.90
					Tail conditional expectation, $k=20$					
Fixed-k LR	0.60	0.87	0.94	0.84	0.69	0.82	0.95	0.97	0.86	0.71
Fixed-k opt	0.39	0.69	0.90	0.92	0.86	0.66	0.81	0.91	0.89	0.82
Fixed-k env	0.34	0.66	0.86	0.78	0.61	0.55	0.74	0.87	0.78	0.63
					Tail conditional expectation, $k=50$					
Fixed-k LR	0.13	0.34	0.78	0.76	0.50	0.52	0.64	0.88	0.79	0.51
Fixed-k opt	0.09	0.21	0.61	0.76	0.58	0.45	0.51	0.75	0.78	0.58
Fixed-k env	0.09	0.19	0.59	0.72	0.44	0.42	0.48	0.69	0.73	0.45

NOTES: Entries are asymptotic expected lengths of 95% confidence intervals about the $1-h/n$ quantile and tail conditional expectation relative to the expected length of the $k=10$ fixed-k LR interval. See the main text for a description of the three types of confidence intervals.

For any given k, h and confidence level α, the critical value $\mathrm{cv}_{\mathrm{LR}}$ and measure Λ only needs to be determined once. Given $\mathrm{cv}_{\mathrm{LR}}$ and Λ, the confidence sets S_{LR} and S_Λ are readily computed from (5) and (8). We provide corresponding Matlab code, and tables of $\mathrm{cv}_{\mathrm{LR}}$ and Λ distributions for $k \in \{5, 10, 15, 20, 30, 40, 50, 75, 100\}$, $\log h \in \{-5.0, -4.5, \dots, 3.0\}$ and $\alpha \in \{0.01, 0.05, 0.10, 0.20\}$. Over this range of parameters, the weighted expected length (6) of S_{LR} is up to 60% longer than S_Λ, with the largest differences occurring for small h and moderately large k. But for many other parameter configurations, the gains are more moderate: the median difference in weighted expected length is 12%. In this sense, S_{LR} is fairly close to weighted expected length optimal in many scenarios.

Table 1 reports the asymptotic expected lengths of 95% confidence intervals S_{LR} ("fixed-k LR") and the weighted expected length optimal interval S_Λ ("fixed-k opt") for $\xi \in \{-0.5, -0.25, 0, 0.25, 0.5\}$, $k \in \{10, 20, 50\}$ and $h \in \{0.1, 5\}$, relative to the expected lengths of S_{LR} for $k=10$ and the same value of ξ. As an additional comparison, we also report for each value of ξ the expected length of the length optimal confidence interval with degenerate weight function W with all mass at that value of ξ. These "fixed-k env" rows do not correspond to a feasible interval, but provide the lower envelope on the expected length of any 95% fixed-k confidence interval for each value of ξ.

As can be seen from Table 1, the S_{LR} intervals tend to be shorter than S_Λ for $\xi = 0.5$, but they have relatively less attractive properties for $\xi \le 0$. The S_Λ intervals are up to 40% longer than the envelope, although for $\xi \le 0.25$ the differences are much less pronounced. This demonstrates that a uniformly smallest expected length interval does not exist and correspondingly, the choice of W matters. At the same time, for $\xi \le 0.25$, the expected length of S_Λ comes reasonably close to the envelope, so no other choice of W could lead to much shorter intervals there. All these differences are more pronounced for $h = 0.1$ than for $h = 5$. Looking across different values of k, expected lengths tend to decrease at a rate slower than $k^{-1/2}$, with a pronounced exception for $h = 0.1$ and negative ξ , where larger k enable dramatically more informative inference.

3. Monte Carlo Results

This section reports some small-sample results for the resulting confidence set for $\alpha = 0.05$ and $n = 250$. We consider six data-generating processes: a standard normal and standard log-normal distribution (both in the domain of attraction of the Gumbel law, $\xi = 0$), a Student-t distribution with 3 degrees of freedom (in the domain of attraction of the Fréchet law with $\xi = 1/3$), an F-distribution with 4 degrees of freedom in both the numerator and denominator (in the domain of attraction of the Fréchet law with $\xi = 0.5$), a 0.8 / 0.2 mixture between a standard normal distribution and a Student-t distribution with 3 degrees of freedom (in the same domain of attraction as the pure Student-t distribution, since the tail is eventually dominated by the Student-t distribution), and a symmetric triangular distribution with support equal to $[-1, 1]$ (in the domain of attraction of the Weibull law with $\xi = -0.5$).

We compare the two intervals derived above ("fixed-k LR" and "fixed-k opt") with several alternatives developed in the literature under $k_n \to \infty$ asymptotics. The first method

("W-H") is the classic Weissman (1978) estimator that relies on a Pareto tail approximation and estimates the tail index by the Hill (1975) estimator. Denoting the Hill estimator by $\hat{\xi}^H$, the $1-h/n$ quantile estimator is $\hat{U}_{WH} = Y_{n:n-k}\,(h/k)^{-\hat{\xi}^H}$, and the corresponding tail conditional expectation estimator is $\hat{T}_{WH} = (Y_{n:n-k}\,(h/k)^{-\hat{\xi}^H})/(1-\hat{\xi}^H)$. Confidence intervals are obtained by exploiting the asymptotic normal limit of these estimators (after suitable scaling). Note that the method is asymptotically valid only for $\xi > 0$. For further discussion and small-sample simulation results, see Drees (2003). The second method ("dH-F") is described in Chapter 4 of de Haan and Ferreira's (2007) textbook and is based on the asymptotically normal estimator $\hat{U}_{\mathrm{dHF}} = Y_{n:n-k} + \hat{a}\,(n/k)\,((h/k)^{-\hat{\xi}^M} - 1)/\hat{\xi}^M$, where $\hat{a}\,(n/k)$ and $\hat{\xi}^M$ are estimators of the scale and the tail index, correspondingly (similar to what was derived by Dekkers and de Haan (1989) and de Haan and Rootzén (1993)). The analogous estimator for the tail conditional expectation is $\hat{T}_{dHF} = Y_{n:n-k} + \hat{a}\,(n/k)\,((h/k)^{-\hat{\xi}^M} - 1 + \hat{\xi}^M)/(\hat{\xi}^M(1-\hat{\xi}^M))$. See de Haan and Ferreira (2007) for further details on the construction of the corresponding confidence intervals. None of these estimators are location and scale equivariant, since $\hat{\xi}^H$ and $\hat{\xi}^M$ are not translation invariant. Invariant estimators were discussed by Santos, Alves, and Gomes (2006), but we found that the suggested confidence intervals based on these estimators do not perform well in our small sample. Finally, we consider intervals constructed from the profile likelihood based on a Poisson and generalized Pareto approximation ("profile"), as suggested and implemented by Davison and Smith (1990). This method assumes that the number of exceedances above a high threshold u follows a Poisson distribution with mean λ, and conditionally on the number of exceedances, the excess values are iid generalized Pareto. In our implementation, we choose $u = Y_{n:n-k}$, which renders the corresponding intervals scale and translation equivariant. For all three $k_n \to \infty$ methods, we impose the same parameter space restriction $\Xi = [-1/2, 1/2]$ on the tail index that we chose in the implementation of the fixed-k method.[‡]

Tables 2 and 3 report the coverage and length properties of these five procedures for $h \in \{0.1, 5\}$. We find that the fixed-k approaches have excellent small-sample coverage, especially for $k \leq 20$. In contrast, the W-H and dH-F intervals display very substantial undercoverage in many of the considered cases. In comparison, the profile likelihood intervals fare much better, corroborating corresponding remarks about the superior small-sample performance in Coles (2001) and Smith (2004). Still, the fixed-k approaches provides much more reliable inference, especially for $k \leq 20$.

The small-sample performance of the fixed-k intervals mirrors their asymptotic properties of Table 1: size control is very similar, and the LR intervals are shorter for underlying distributions with heavy tails, but at the expense of worse performance under thinner tails. In our view, the relative simplicity and transparency of the LR intervals give them the edge for applied work, unless there is good reason to be believe that the data might exhibit thin tails.

Larger k mostly leave coverage of the fixed-k intervals close to the nominal level, and average length decreases. This pattern reflects that most distributions considered in Tables 2 and 3 have nicely behaved tails, so that a relatively large fraction can be well approximated by a generalized Pareto distribution. But as demonstrated by the mixture distribution, it is easy to construct underlying distributions whose tail behavior is much less benign—most draws stem from the normal component, misleadingly indicating a thin tail, but the tail properties of interest are in fact determined by the Student-t component, especially for small h. In that case, choosing $k > 10$ leads to substantial undercoverage also for the fixed-k approaches.

4. Application to Hurricane Damage

Hurricanes cause significant damage to coastal communities in the U.S. It seems interesting to learn about tail features of the distribution of these damages, both from a public policy and insurance perspective. In its technical memorandum NWS NHC-6, the National Weather Service provides estimates of the damage of the 30 costliest mainland U.S. tropical cyclones from 1900–2010, measured in 2010 US\$. These estimates, however, are based on variable methodology, which is standardized only from the 1995 hurricane season onwards. We thus only rely on the data about the $k = 10$ costliest hurricanes in the 16 year window from 1995 to 2010. Panel A in Table 4 replicates these data points for convenience.

Note that in this example, the number n of total tropical cyclones in the period 1995–2010 is not known to us. Under the assumption that hurricane damage is iid and hurricane arrival is stationary, the $1-h/n$ quantiles nevertheless has a straightforward interpretation: a hurricane causing at least that amount of damage is expected every $16/h$ years. Similarly, the tail conditional expectation corresponds to the average damage of these extreme hurricanes.

Panel B in Table 4 provides 95% confidence intervals for these quantities for $h \in \{0.1, 1, 5\}$ using the two fixed-k approaches developed above. In all cases, the LR intervals are fully contained in the weighted expected length minimizing ones. It is interesting to note that the upper bound of the LR interval for $h = 1$ is \$116.3 billion, only slightly larger than the costliest hurricane that was observed in the sample (Katrina in 2005 with \$105.8 billion). As can be seen in the first panel, Katrina is an outlier relative to the other costliest hurricanes in 1995–2010, making it relatively implausible that the $1-1/n$ quantile is much bigger than the largest observation. The corresponding upper bound on the 95% LR confidence interval for the tail conditional expectation is \$266.4 billion. This is very large in absolute terms, corresponding to roughly 1.7% of current U.S. gross domestic product, and indicates substantial insurance needs even from a macroeconomic perspective.

5. Robustness of Fixed-k Asymptotic Inference

For a given distribution F in the domain of attraction of an extreme value distribution, there are eventually infinitely many observations from the generalized Pareto tail as $n \to \infty$. For this reason, we expect the fixed-k confidence intervals to be most useful for inference in small to moderately large samples, where a large k would amount to modeling a large fraction of F as being generalized Pareto. In this sense, the fixed-k approach can be

[‡] Not doing so leads to intervals that are very much longer on average, and with typically no better (but often) worse coverage properties.

Table 2. Small-sample performance for $1 - h/n$ quantile.

	10		20		50		10		20		50	
k	Cov	Lgth	Cov	Lgth	Cov	Lgth	Cov	Lgth	Cov	Lgth	Cov	Lgth
	$h = 0.1$											
	Normal						Student-t(3)					
Fixed-k LR	99.3	10.3	98.5	9.57	93.8	5.06	98.4	42.9	96.8	36.0	93.8	32.5
Fixed-k opt	95.3	9.98	94.5	8.35	91.4	3.45	97.5	46.8	94.6	39.8	91.2	32.2
W-H	99.2	5.83	77.2	8.08	0.00	25.4	94.3	50.5	98.8	48.3	100	41.5
dH-F	85.0	3.35	83.8	3.48	74.8	3.08	66.0	33.3	76.2	35.0	82.0	33.1
Profile	97.5	6.55	92.2	5.77	88.1	3.26	89.1	33.2	90.2	30.1	89.1	28.4
	Log-Normal						$F(4,4)$					
Fixed-k LR	98.7	97.7	97.9	83.0	97.2	72.7	94.8	207	94.9	146	95.3	104
Fixed-k opt	97.5	101	95.8	89.5	96.1	74.5	95.2	248	95.0	179	95.2	121
W-H	96.7	115	99.5	108	99.9	95.8	92.8	227	96.7	175	99.9	107
dH-F	70.9	72.2	82.2	79.1	91.5	82.7	65.1	219	75.7	214	87.3	192
Profile	90.6	73.7	91.9	67.7	94.7	65.1	84.4	162	89.2	126	91.9	95.7
	Mixture $\mathcal{N}(0,1)$/Student-t(3)						Triangular					
Fixed-k LR	97.2	20.0	93.3	17.4	79.1	14.4	99.6	1.89	99.8	1.32	99.6	0.44
Fixed-k opt	97.4	22.3	87.4	18.8	70.4	13.0	95.4	1.86	95.5	0.89	95.1	0.33
W-H	82.4	16.6	94.3	15.5	80.4	30.7	97.4	0.99	13.1	1.74	0.00	8.68
dH-F	57.9	16.0	60.1	14.3	50.8	9.99	99.0	0.47	99.8	0.51	99.9	0.56
Profile	85.5	15.5	82.9	14.1	68.3	11.0	100	1.90	100	2.41	100	1.09
	$h = 5$											
	Normal						Student-t(3)					
Fixed-k LR	95.9	0.78	96.5	0.76	95.7	0.72	95.5	2.95	95.8	2.64	95.8	2.47
Fixed-k opt	94.8	0.75	95.9	0.71	94.5	0.65	95.5	2.82	95.9	2.64	95.4	2.39
W-H	66.4	0.34	92.2	0.65	98.1	1.40	71.2	1.13	90.8	1.80	92.8	1.97
dH-F	24.6	0.11	66.7	0.31	89.2	0.58	26.7	0.40	66.7	1.04	90.5	1.86
Profile	93.5	0.70	93.5	0.95	93.6	1.18	90.6	2.57	93.0	2.61	93.7	2.23
	Log-Normal						$F(4,4)$					
Fixed-k LR	96.0	6.74	95.9	6.02	96.8	5.49	95.9	13.8	95.5	10.9	95.7	8.33
Fixed-k opt	95.2	6.49	95.4	5.85	95.5	5.28	95.9	13.5	95.1	11.5	95.4	9.09
W-H	70.2	2.55	90.0	4.01	98.7	4.61	61.9	3.97	79.8	5.56	42.4	5.06
dH-F	26.9	0.90	65.4	2.37	89.4	4.06	28.1	1.83	65.9	4.45	86.9	6.69
Profile	89.4	5.56	92.1	5.53	94.0	4.90	90.0	11.2	91.4	9.91	92.9	7.48
	Mixture $\mathcal{N}(0,1)$/Student-t(3)						Triangular					
Fixed-k LR	95.4	1.36	96.4	1.28	93.3	1.25	95.3	0.17	96.5	0.17	97.3	0.17
Fixed-k opt	96.3	1.31	96.1	1.27	92.3	1.17	95.2	0.17	96.0	0.16	95.2	0.14
W-H	70.5	0.56	91.4	0.91	98.7	1.57	62.7	0.08	94.4	0.19	92.7	0.53
dH-F	27.8	0.19	68.1	0.51	89.2	0.91	25.4	0.03	69.2	0.08	94.2	0.17
Profile	92.8	1.21	93.4	1.44	90.7	1.44	95.8	0.19	98.4	0.34	98.3	0.52

NOTES: Entries are coverage probabilities (in percent) and expected lengths of nominal 95% confidence intervals in a sample of size $n = 250$ about the $1 - h/n$ quantile of the underlying distribution, based on the largest k order statistics. See the main text for a description of the five types of confidence intervals. Based on 5000 Monte Carlo simulations.

thought of as a small-sample adjustment to $k_n \to \infty$ asymptotic approaches.

At the same time, for any sample size n, there exist F that are in the domain of attraction of the extreme value distribution, yet choosing k large yields poor inference (think of the mixture distribution case in the simulation section). This suggests studying the robustness of tail inference under triangular array asymptotics, where the iid sample of size n is drawn from a population indexed by n, so that $F = F_n$. The following theorem contrasts fixed-k asymptotic inference with potentially more informative $k_n \to \infty$ inference under such triangular array asymptotics. Its proof is in the Appendix.

Theorem 5.1.

(a) Let F_0 be in the domain of attraction of an extreme value distribution with tail index ξ, that is for some sequences a_n and b_n

$$[F_0(a_n x + b_n)]^n \to G_\xi(x). \tag{9}$$

Suppose the sequence of distribution functions F_n is such that for some $r_n \to \infty$, $F_n(y) = F_0(y)$ for all $y \geq U(F_0, n/r_n)$. Let $Y_{n:1}, \ldots, Y_{n:n}$ be the order statistics from an iid sample of size n from F_n. Then for any fixed $k \in \mathbb{N}$,

$$\left(\frac{Y_{n:n} - b_n}{a_n}, \ldots, \frac{Y_{n:n-k+1} - b_n}{a_n}\right) \Rightarrow (X_1, \ldots, X_k) \tag{10}$$

where $X_1, \ldots, X_k$ have joint extreme value distribution with tail index ξ.

(b) Write Pa_ξ for the c.d.f. of a Pareto distribution with support $[1, \infty)$ and shape parameter $\alpha = 1/\xi > 0$. Suppose $\hat{\xi}_n$ is a scale invariant estimator of ξ that, when applied to an iid sample from Pa_{ξ_0}, converges in probability to ξ_0 for

Table 3. Small-sample performance for $1 - h/n$ tail conditional expectation.

k	10		20		50		10		20		50	
	Cov	Lgth	Cov	Lgth	Cov	Lgth	Cov	Lgth	Cov	Lgth	Cov	Lgth
	$h = 0.1$											
	Normal						Student-t(3)					
Fixed-k LR	99.6	22.9	98.9	20.6	93.9	9.28	99.4	94.1	97.1	78.9	93.7	70.2
Fixed-k opt	94.4	20.9	95.2	17.7	90.5	5.32	95.9	102	94.1	87.5	89.8	68.4
W-H	84.9	7.62	24.0	11.4	0.00	48.0	92.8	89.2	98.2	89.0	89.4	82.9
dH-F	76.2	3.87	76.0	3.83	68.6	3.23	55.9	72.4	66.4	74.4	72.1	68.8
Profile	98.3	13.9	94.0	13.3	87.5	5.00	89.6	71.1	90.0	67.3	89.5	60.2
	Log-Normal						F(4,4)					
Fixed-k LR	99.5	217	98.1	180	97.4	159	94.9	448	94.7	316	95.1	229
Fixed-k opt	96.2	224	96.0	198	96.3	158	94.5	529	95.4	396	94.9	265
W-H	96.2	211	93.0	203	18.8	189	88.1	440	94.9	347	99.8	213
dH-F	63.2	138	76.1	168	87.9	184	56.0	562	68.3	572	82.7	531
Profile	91.4	162	92.2	149	94.5	141	85.0	363	88.8	274	92.0	214
	Mixture $\mathcal{N}(0, 1)$/Student-t(3)						Triangular					
Fixed-k LR	98.1	42.5	93.2	37.6	75.8	29.7	99.9	4.23	99.7	2.48	99.5	0.55
Fixed-k opt	96.6	48.2	87.7	41.4	68.3	26.4	95.6	3.75	95.5	1.61	95.4	0.39
W-H	74.4	26.7	88.8	24.8	21.6	59.5	37.5	1.16	0.26	2.22	0.00	15.9
dH-F	50.9	35.3	50.0	29.6	37.4	15.9	98.8	0.47	99.5	0.51	99.8	0.56
Profile	87.1	33.8	81.9	32.4	63.5	22.1	100	3.88	100	4.57	100	2.06
	$h = 5$											
	Normal						Student-t(3)					
Fixed-k LR	98.9	2.17	98.1	2.09	94.9	1.38	97.7	9.11	96.7	7.86	94.5	7.02
Fixed-k opt	96.1	1.99	95.4	1.82	92.9	1.09	97.8	9.22	96.0	8.15	93.2	6.92
W-H	56.4	0.44	65.7	0.91	0.01	2.61	48.8	1.93	76.1	3.28	97.1	3.94
dH-F	19.5	0.11	55.4	0.33	69.0	0.59	16.3	0.66	44.6	1.87	70.8	3.38
Profile	94.1	1.70	92.7	1.81	91.7	1.71	89.8	7.51	90.9	7.44	91.8	6.17
	Log-Normal						F(4,4)					
Fixed-k LR	97.9	20.6	97.8	17.8	96.7	15.5	95.8	44.5	95.1	32.0	95.5	22.8
Fixed-k opt	97.0	20.3	96.2	18.2	96.0	15.7	94.9	50.0	95.2	37.0	95.1	26.0
W-H	50.4	4.44	75.0	7.36	94.2	9.14	43.4	7.56	62.7	10.9	20.4	10.1
dH-F	17.5	1.42	48.2	4.19	77.9	8.14	15.9	3.91	48.5	10.5	75.6	17.8
Profile	90.1	15.9	92.1	15.5	94.9	13.9	85.6	34.5	89.2	27.2	92.0	20.8
	Mixture $\mathcal{N}(0, 1)$/Student-t(3)						Triangular					
Fixed-k LR	96.9	4.27	93.8	3.77	87.4	3.22	99.2	0.40	99.3	0.33	98.6	0.21
Fixed-k opt	97.4	4.35	92.7	3.83	84.1	3.00	94.9	0.36	95.3	0.25	95.2	0.17
W-H	42.1	0.84	69.1	1.42	17.8	3.02	65.4	0.20	51.2	0.24	0.08	0.95
dH-F	14.1	0.31	37.7	0.82	54.4	1.20	26.3	0.03	67.5	0.08	90.1	0.17
Profile	87.5	3.67	85.9	3.65	82.2	2.87	100	0.43	99.8	0.65	99.4	0.63

NOTES: Entries are coverage probabilities (in percent) and expected lengths of nominal 95% confidence intervals in a sample of size $n = 250$ about the tail conditional expectation $E[Y_i|Y_i > U(F, n/h)]$ of the underlying distribution F, based on the largest k order statistics. See the main text for a description of the five types of confidence intervals. Based on 5000 Monte Carlo simulations.

Table 4. Empirical results on damage of U.S. mainland hurricanes.

Panel A: Damage in US$ billion of 10 costliest hurricanes in 1995–2010									
105.8	27.8	20.6	19.8	15.8	11.8	11.0	10.0	9.2	8.1

Panel B: Endpoints of 95% confidence intervals for the $1 - h/n$ quantile and tail conditional expectation						
	$h = 0.1$		$h = 1$		$h = 5$	
	Fixed-k LR					
Quantile	40.2	439.2	16.5	116.3	7.1	32.8
TCE	54.9	914.6	27.7	266.4	14.9	99.9
	Fixed-k opt					
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

NOTE: Based on the $k = 10$ order statistics from Panel A.

some $\xi_0 > 0$. Then for any $\xi_1 > 0$ there exists a sequence of distributions F_n and a real sequence $r_n \to \infty$ such that $F_n(y) = \mathrm{Pa}_{\xi_1}(y)$ for all $y \geq U(\mathrm{Pa}_{\xi_1}, n/r_n)$, yet $\hat{\xi}_n$ applied to an iid sample of size n from F_n converges in probability to ξ_0.

(c) For some fixed $h > 0$, let $[\hat{l}_n, \hat{u}_n]$ be a scale equivariant confidence interval for $U(F, n/h)$. Suppose $\hat{l}_n/U(\mathrm{Pa}_{\xi_0}, n/h)$ and $\hat{u}_n/U(\mathrm{Pa}_{\xi_0}, n/h)$ converge in probability to unity when applied to an iid sample of size n from Pa_{ξ_0}. Then for any $\xi_1 \neq \xi_0$, $\xi_1 > 0$ there exists a sequence of distributions F_n and a real sequence $r_n \to \infty$ such that $F_n(y) = \mathrm{Pa}_{\xi_1}(y)$ for all $y \geq U(\mathrm{Pa}_{\xi_1}, n/r_n)$, yet $[\hat{l}_n, \hat{u}_n]$ applied to an iid sample of size n from F_n contains $U(F_n, n/h) = U(\mathrm{Pa}_{\xi_1}, n/h)$ with probability converging to zero.

Part (a) shows that it suffices for the sequence of distributions F_n to have a tail in the domain of attraction of extreme value distribution of mass equal to r_n/n to induce the corresponding joint extreme value distribution for the largest k order statistics, for any fixed k and arbitrarily slowly increasing sequence $r_n \to \infty$. Thus, to the extent that the object of interest concerns properties of this (extreme) tail, fixed k asymptotic inference is asymptotically valid.

In contrast, part (b) demonstrates that *any* approach to scale invariant consistent tail index estimation is necessarily nonrobust to some sequence F_n of this sort. All popular tail index estimators are scale invariant and consistent under iid data from a Pareto distribution. Thus, Theorem 1(b) applies and shows that for a large enough sample size, one can always construct an underlying distribution for which the extreme tail properties are Pareto with parameter ξ_1, yet the tail index estimator takes on values close to an entirely different value $\xi_0 \neq \xi_1$ with high probability. This typically implies very poor coverage properties of confidence intervals based on consistent tail index estimators.

The fixed-k intervals about $U(F, n/h)$ derived here have length $O_p(U(F, n/h))$ for data from a Pareto distribution. In contrast, intervals derived under $k_n \to \infty$ asymptotics usually are more informative and have length $o_p(U(F, n/h))$, so that their endpoints satisfy the assumption of part (c). The theorem shows that *any* such scale equivariant interval has zero asymptotic coverage for some underlying sequence of distributions F_n for which fixed-k inference is asymptotically valid, whether or not its construction involves a consistent tail index estimator $\hat{\xi}_n$. Taken together, Theorem 1 thus provides a precise sense in which fixed-k asymptotic inference is more robust than all previously suggested inference approaches we are aware of.

A key challenge for the empirical implementation of fixed-k based inference is the selection of k. Of course, also under $k_n \to \infty$, the determination of k_n in a given sample of size n is widely recognized as a difficult issue. But the problem is arguably even harder under fixed-k asymptotics, as there cannot exist a procedure based on the largest k order statistics that *consistently* determines whether, say, $k_1 \leq k$ or $k_2 < k_1$ is appropriate.

One useful way to think about the choice of k is to recall that the number K_r of exceedances above $U(F, n/r)$ in an iid sample from F has a Binomial distribution with parameters n and $p = r/n$, so it has mean r and a variance no larger than r. In fact, a calculation shows that if the upper tail of F of mass r_k/n with $r_k = k + 3 + 3\sqrt{k}$ is equal to a generalized Pareto distribution F_0, then with probability of at least 99.88%, the largest k-order statistics stem from F_0, for all $k \geq 5$ and $n \geq r_k$. The only remaining approximation of fixed-k asymptotic inference in small samples then involves the approximation of the distribution of K_{r_k} by a Poisson distribution with mean r_k. Unreported small-sample simulations with n as small as 25 show excellent coverage properties of the fixed-k intervals derived here for all F with upper tail of mass r_k/n equal to a generalized Pareto distribution, even if below the $U(F, n/r_k)$ quantile, the shape of F differs arbitrarily from the generalized Pareto tail distribution. In practice, the choice of k is thus directly interpretable as an assumption about the extent of the (approximate) generalized Pareto tail, and it makes sense to present consumers of tail inference with results under various choices for this key regularity assumption. We leave further consideration of this important issue to future research.

A. Appendix

A.1. Computational Details

The set S_Λ in (8) requires evaluation of $\kappa_\xi(\cdot) f_{\mathbf{X}^s|\xi}(\cdot)$ and $f_{Y^s(\xi),\mathbf{X}^s}(\cdot)$. Using the expression for $f_{\mathbf{X}|\xi}$ below (2), straightforward calculations yield

$$\kappa_\xi(\mathbf{X}^s) f_{\mathbf{X}^s|\xi}(\mathbf{X}^s) = \Gamma(k-\xi) \int_0^{b_0(\xi)} s^{k-1} \times \exp\left[-(1+\xi^{-1}) \sum_{i=1}^{k} \log(1+\xi X_i^s s)\right] ds,$$

where Γ is the Gamma function, $b_0(\xi) = -1/\xi$ for $\xi < 0$, and $b(\xi) = \infty$ otherwise, and

$$f_{Y^s(\xi),\mathbf{X}^s}(y, \mathbf{X}^s) = |y|^{-k} \int_{a_1(\xi)}^{b_1(\xi)} |q(\xi,h) - s|^{k-1} \times \exp\left[-(1+\xi s)^{-1/\xi} - (1+\xi^{-1}) \times \sum_{i=1}^{k} \log\left(1 + \xi s + X_i^s \xi \frac{q(\xi,h)-s}{y}\right)\right] ds,$$

where $a_1(\xi)$ and $b_1(\xi)$ are such that for all $s \in (a_1(\xi), b_1(\xi))$, $1 + \xi s > 0$, $1 + \xi s + \xi \frac{q(\xi,h)-s}{y} > 0$ and $(q(\xi, h) - s)/y > 0$. We evaluate these integrals by numerical quadrature.

The determination of Λ closely follows the approach developed in Elliott, Müller, and Watson (2015) and Müller and Watson (2015). As discussed there, consider $\Lambda = c\tilde{\Lambda}$, where $\tilde{\Lambda}$ is a given probability distribution with support on Ξ. Suppose the scalar $c > 0$ is such that if ξ is random and drawn from $\tilde{\Lambda}$, then $S_{c\tilde{\Lambda}}$ has coverage equal to the nominal level, that is c solves $\int P_\xi(Y^s(\xi) \in S_{c\tilde{\Lambda}}(\mathbf{X}^s)) d\tilde{\Lambda}(\xi) = 1 - \alpha$ (so c is a function of $\tilde{\Lambda}$). Then, the W-weighted average expected length of $S_{c\tilde{\Lambda}}$, $V_{\tilde{\Lambda}} = \int E_\xi[\kappa_\xi(\mathbf{X}^s) \mathrm{lgth}(S_{c\tilde{\Lambda}}(\mathbf{X}^s))] dW(\xi)$, provides a lower bound for the

W-weighted average expected length of *any* set S with uniform coverage $P_\xi(Y^s(\xi) \in S(\mathbf{X}^s)) \geq 1-\alpha$ for all $\xi \in \Xi$, since uniform coverage implies (at least) $\tilde{\Lambda}$-weighted average coverage for any probability distribution $\tilde{\Lambda}$ with support on Ξ, and by construction of S_Λ, $S_{c\tilde{\Lambda}}$ minimizes W-weighted average expected length among all sets with $\tilde{\Lambda}$-weighted coverage of at last $1-\alpha$.

So suppose we knew of some probability distribution $\tilde{\Lambda}^*$ with support on Ξ and constant $d^* > 0$ such that (i) $P_\xi(Y^s(\xi) \in S_{d^*\tilde{\Lambda}^*}(\mathbf{X}^s)) \geq 1-\alpha$ for all $\xi \in \Xi$ and (ii) $\int E_\xi[\kappa_\xi(\mathbf{X}^s)\mathrm{lgth}(S_{d^*\tilde{\Lambda}^*}(\mathbf{X}^s))]dW(\xi) \leq (1+\varepsilon)V_{\tilde{\Lambda}^*}$. Then we would have identified a level $1-\alpha$ confidence set, $S_{d^*\tilde{\Lambda}^*}(\mathbf{X}^s)$, that is demonstrably no more than $100\varepsilon\%$ longer in a W-weighted average expected sense than any other confidence set of the same level.

The remaining challenge is the determination of a suitable distribution $\tilde{\Lambda}^*$. To this end, we restrict $\tilde{\Lambda}^*$ to be a discrete distribution with support on $\Xi_a = \{-1/2, -1/2+1/59, \ldots, 1/2\}$, and determine the 60 point masses by fixed-point iterations based on importance sample Monte Carlo estimates of coverage, as suggested by Elliott, Müller, and Watson (2015) and Müller and Watson (2015). In particular, we simulate rejection probabilities with 100,000 iid draws from a proposal with ξ drawn uniformly from $\Xi_p = \{-0.5, -0.5+1/29, \ldots, 0.5\}$ (which yields Monte Carlo standard errors of approximately 0.1% for 95% level confidence sets), and iteratively increase or decrease the 60 point masses on Ξ_a as a function of whether the (estimated) inclusion probability under the corresponding $\xi \in \Xi_a$ is larger or smaller than the nominal level. After 500 iterations, the resulting discrete distribution is a candidate for $\tilde{\Lambda}^*$. We compute $V_{\tilde{\Lambda}^*}$, and then numerically determine d^* so that $\int E_\xi[\kappa_\xi(\mathbf{X}^s)\mathrm{lgth}(S_{d^*\tilde{\Lambda}^*}(\mathbf{X}^s))]dW(\xi) = (1+\varepsilon)V_{\tilde{\Lambda}^*}$ for $\varepsilon = 0.01$. In a last step, we check whether $S_{d^*\tilde{\Lambda}^*}$ indeed controls coverage uniformly by computing coverage probabilities over the fine grid $\Xi_f = \{-1/2, -1/2+\frac{1}{199}, \ldots, 1/2\}$.

The critical value $\mathrm{cv}_{\mathrm{LR}}$ of S_{LR} in (5) is computed from the same 100,000 importance sampling draws. In all cases, $P_\xi(\mathrm{LR}(q(\xi,h),\mathbf{X}) > \mathrm{cv}_{\mathrm{LR}}) = \alpha$ for $\xi = 1/2$.

For a given value of k and h, these computations take about one minute on a modern PC. Of course, in actual applications, there is no need to recompute $\mathrm{cv}_{\mathrm{LR}}$, d^*, and $\tilde{\Lambda}^*$; the determination of the confidence sets simply requires the numerical determination of (5) and (8) for the values of $\mathrm{cv}_{\mathrm{LR}}$ and $\Lambda = d^*\tilde{\Lambda}^*$ we already computed. For the latter, note that with $M_{\tilde{\mathbf{X}}} = \max_{\{(\mu,\sigma,\xi):\sigma>0,\xi\in\Xi\}} L_{\tilde{\mathbf{X}}}(\mu,\sigma,\xi)$, the upper endpoint of S_{LR} solves the nonlinearly constrained optimization problem

$$\max_{\{(\mu,\sigma,\xi):\sigma>0,\xi\in\Xi\}} \mu+\sigma q(\xi,h) \quad \text{s.t. } L_{\tilde{\mathbf{X}}}(\mu,\sigma,\xi) \geq M_{\tilde{\mathbf{X}}} - \mathrm{cv}_{\mathrm{LR}},$$

and the lower endpoint solves the corresponding minimization problem. We provide tables of $\mathrm{cv}_{\mathrm{LR}}$ and Λ and corresponding Matlab code on our website, www.princeton.edu/~umueller.

A.2. Proof of Theorem 1

(a) Let $U_i \sim \mathrm{iid}U[0,1]$ and $F_n^-(p) = \inf_y\{y \in \mathbb{R} : F_n(y) \geq p\}$. Generate the sample of size n from F_n via $Y_{n,i} = F_n^-(U_i)$, and denote the corresponding sample from F_0 by $\tilde{Y}_{n,i} = F_0^-(U_i)$, so that the order statistics are given by $Y_{n:i} = F_n^-(U_{n:i})$ and $\tilde{Y}_{n:i} = F_0^-(U_{n:i})$. As is well known (see Theorem 3.5 of Coles (2001), for instance), (9) implies that (10) holds for the sample $\tilde{Y}_{n,i}$, that is $a_n^{-1}(\tilde{Y}_{n:n} - b_n, \ldots, \tilde{Y}_{n:n-k+1} - b_n) \Rightarrow (X_1, \ldots, X_k)$. Let K_n be the number of $U_1, \ldots, U_n$ that exceed $1 - r_n/n$. The event $Y_{n:i} = \tilde{Y}_{n:i}$ for $i = n-k+1, \ldots, n$ is clearly implied by the event $K_n \geq k$. But K_n is binomially distributed with parameters r_n/n and n, so $r_n \to \infty$ implies $P(K_n \geq k) \to 1$, and the result follows.

(b) For $r > 0$, let $u_{r,n} = U(\mathrm{Pa}_{\xi_0}, n/r) = (n/r)^{\xi_0}$ and define the c.d.f. $H_{n,r}$ as

$$H_{n,r}(y) = \begin{cases} \mathrm{Pa}_{\xi_0}(y) \text{ for } y \leq u_{r,n} \\ \mathrm{Pa}_{\xi_1}(U(\mathrm{Pa}_{\xi_1}, n/r)y/u_{r,n}) \text{ for } y > u_{r,n} \end{cases}$$

$$= \begin{cases} 1 - y^{-1/\xi_0} \text{ for } y \leq u_{r,n} \\ 1 - u_{r,n}^{1/\xi_1 - 1/\xi_0} y^{-1/\xi_1} \text{ for } y > u_{r,n}. \end{cases}$$

We first show that for fixed r, the experiment of observing an iid sample of size n from $H_{n,r}$, $\{Y_{n,i}\}_{i=1}^n$, is contiguous to the experiment of observing an iid sample of size n from Pa_{ξ_0}. Write F_n^n for the product measure of an iid sample of size n from the distribution F_n. Note that the log likelihood ratio $\log(dH_{n,r}^n/d\mathrm{Pa}_{\xi_0}^n)$ is given by

$$\log(dH_{n,r}^n/d\mathrm{Pa}_{\xi_0}^n) = \sum_{i=1}^n \mathbf{1}[Y_{n,i} \geq u_{r,n}]((1/\xi_0 - 1/\xi_1) \times \log(Y_{n,i}/u_{r,n}) + \log(\xi_0/\xi_1)).$$

Let K_n the number of exceedances of $u_{r,n}$, $K_n = \sum_{i=1}^n \mathbf{1}[Y_{n,i} \geq u_{r,n}]$. Furthermore, let $W_{n,i}$, $i = 1, \ldots, K_n$ be a random permutation of $\{Y_{n:n-i+1} - u_{r,n}\}_{i=1}^{K_n}$. Then under $\mathrm{Pa}_{\xi_0}^n$, and conditional on K_n, $W_{n,i}$ is iid with c.d.f. $F_{W,n}(w) = (H_{n,r}(u_{r,n}+w) - H_{n,r}(u_{r,n}))/(1 - H_{n,r}(u_{r,n})) = 1 - (1 + w/u_{r,n})^{-1/\xi_0}$ for $w \geq 0$ as in Smith (1987), so that $Z_{n,i} = W_{n,i}/u_{r,n} + 1$ is iid with c.d.f. $1 - z^{-1/\xi_0}$ for $z \geq 1$. Furthermore, under $\mathrm{Pa}_{\xi_0}^n$, by Theorem 2.1.1 in Leadbetter (1983), $K_n \Rightarrow K$, where K is Poisson with parameter r. Thus, under $\mathrm{Pa}_{\xi_0}^n$,

$$\log\left(dH_{n,r}^n/d\mathrm{Pa}_{\xi_0}^n\right) \Rightarrow \log(L) = \sum_{i=1}^K ((1/\xi_0 - 1/\xi_1) \times \log Z_i + \log(\xi_0/\xi_1)),$$

where Z_i is iid with c.d.f. $1 - z^{-1/\xi_0}$ independent of K (and $\log(L) = 0$ if $K = 0$). Also

$$\begin{aligned} E[L] &= E\left[\exp\left[\sum_{i=1}^K ((1/\xi_0 - 1/\xi_1)\log Z_i + \log(\xi_0/\xi_1))\right]\right] \\ &= \sum_{s=0}^\infty P(K=s) E\left[\prod_{i=1}^s \frac{\xi_0}{\xi_1} Z_i^{1/\xi_0 - 1/\xi_1}\right] \\ &= \sum_{s=0}^\infty P(K=s) = 1, \end{aligned}$$

where the third equality uses $E[Z_i^{1/\xi_0 - 1/\xi_1}] = \xi_1/\xi_0$ from a direct calculation. But the convergence $dH_{n,r}^n/d\mathrm{Pa}_{\xi_0}^n \Rightarrow L$

under $\mathrm{Pa}_{\xi_0}^n$ and $E[L] = 1$ imply contiguity via LeCam's first lemma (see Lemma 6.4 in van der Vaart (1998), for instance).

By definition of contiguity, $\hat{\xi}_n \xrightarrow{p} \xi_0$ under $\mathrm{Pa}_{\xi_0}^n$ implies $\hat{\xi}_n \xrightarrow{p} \xi_0$ also under $H_{n,r}^n$, for any fixed r, where we write '$\xrightarrow{p}$' for convergence in probability. Now for any $r > 0$, let n_r be the smallest $n^* \geq 1$ such that $\sup_{n\geq n^*} P(|\hat{\xi}_n - \xi_0| > r^{-1}) < r^{-1}$ under $H_{n,r}^n$. Note that $\hat{\xi}_n \xrightarrow{p} \xi_0$ under $H_{n,r}^n$ for any fixed r implies $n_r < \infty$. Construct the inverse function r_n via $r_n = \sup\{r > 0 : n_r \leq n\}$. Since $n_r < \infty$ for all r, $r_n \to \infty$. Thus, under H_{n,r_n}^n, $P(|\hat{\xi}_n - \xi_0| > r_n^{-1}) < r_n^{-1}$ by construction of r_n, so that $\hat{\xi}_n \xrightarrow{p} \xi_0$ under H_{n,r_n}^n by definition of convergence in probability. By scale invariance, the same holds under F_n^n, where

$$F_n(y) = H_{n,r_n}\left(y u_{r_n,n}^{1-\xi_1/\xi_0}\right)$$

$$= \begin{cases} 1 - u_{r_n,n}^{(\xi_1-\xi_0)/\xi_0^2} y^{-1/\xi_0} \text{ for } y \leq u_{r_n,n}^{\xi_1/\xi_0} \\ 1 - y^{-1/\xi_1} \text{ for } y > u_{r_n,n}^{\xi_1/\xi_0} \end{cases}.$$

Finally, since $U(F_n, n/r_n) = u_{r_n,n}^{\xi_1/\xi_0} = U(\mathrm{Pa}_{\xi_1}, n/r_n)$, we have $F_n(y) = \mathrm{Pa}_{\xi_1}(y)$ for all $y \geq U(\mathrm{Pa}_{\xi_1}, n/r_n)$, as claimed.

(c) Proceed as in part (b) for the construction of $r_n \to \infty$, H_{n,r_n} and F_n, except that r_n is now chosen such that $P(|\hat{l}_n/U(\mathrm{Pa}_{\xi_0}, n/h) - 1| > r_n^{-1}) + P(|\hat{u}_n/U(\mathrm{Pa}_{\xi_0}, n/h) - 1| > r_n^{-1}) < r_n^{-1}$ under H_{n,r_n}^n. Scale equivariance of $[\hat{l}_n, \hat{u}_n]$ thus implies that under F_n^n, $u_{r_n,n}^{1-\xi_1/\xi_0}(\hat{l}_n, \hat{u}_n)/U(\mathrm{Pa}_{\xi_0}, n/h) \xrightarrow{p} (1, 1)$. Using $u_{r,n} = (n/r)^{\xi_0}$ and $U(\mathrm{Pa}_\xi, n/h) = (n/h)^\xi$, we thus find that

$$(r_n/h)^{\xi_1-\xi_0} \frac{(\hat{l}_n, \hat{u})}{U(\mathrm{Pa}_{\xi_1}, n/h)} \xrightarrow{p} (1, 1)$$

under F_n^n, so that from $r_n \to \infty$ implies $P(\hat{l}_n > U(\mathrm{Pa}_{\xi_1}, n/h)) \to 1$ if $\xi_1 < \xi_0$, and $P(\hat{u}_n < U(\mathrm{Pa}_{\xi_1}, n/h)) \to 1$ if $\xi_1 > \xi_0$.

Acknowledgment

We thank two anonymous referees, Marco del Negro and participants at workshops at Brown, Princeton and New York University for useful comments and advice.

Funding

Müller gratefully acknowledges financial support by the NSF via grant SES-1627660.

References

Balkema, A. A., and de Haan, L. (1974), "Residual Life Time at Great Age," *The Annals of Probability*, 2, 792–804. [1334]

Beirlant, J., Caeiro, F., and Gomes, M. I. (2012), "An Overview and Open Research Topics in Statistics of Univariate Extremes," *Revstat*, 10, 1–31. [1334]

Cattaneo, M. D., Crump, R. K., and Jansson, M. (2010), "Robust Data-Driven Inference for Density-Weighted Average Derivatives," *Journal of the American Statistical Association*, 105, 1070–1083. [1334]

Cattaneo, M. D., Crump, R. K., and Jansson, M. (2014), "Small Bandwidth Asymptotics for Density-Weighted Average Derivatives," *Econometric Theory*, 30, 176–200. [1334]

Coles, S. (2001), *An Introduction to Statistical Modeling of Extreme Values*, London: Springer. [1334,1335,1338]

Davison, A. C., and Smith, R. L. (1990), "Models for Exceedances over High Thresholds," *Journal of the Royal Statistical Society*, Series B, 393–442. [1338]

de Haan, L., and Ferreira, A. (2007), *Extreme Value Theory: An Introduction*, New York: Springer Science and Business Media. [1334,1336,1338]

de Haan, L., and Rootzén, H. (1993), "On the Estimation of High Quantiles," *Journal of Statistical Planning and Inference*, 35, 1–13. [1338]

Dekkers, A. L. M., and de Haan, L. (1989), "On the Estimation of the Extreme-Value Index and Large Quantile Estimation," *Annals of Statistics*, 17, 1795–1832. [1338]

Drees, H. (2003), "Extreme Quantile Estimation for Dependent Data, With Applications to Finance," *Bernoulli*, 9, 617–657. [1338]

Elliott, G., Müller, U. K., and Watson, M. W. (2015), "Nearly Optimal Tests When a Nuisance Parameter is Present Under the Null Hypothesis," *Econometrica*, 83, 771–811. [1335,1336,1341,1342]

Embrechts, P., Klüppelberg C., and Mikosch, T. (1997), *Modelling Extremal Events for Insurance and Finance*, New York: Springer. [1334]

Fisher, R. A., and Tippett, L. H. C. (1928), "Limiting Forms of the Frequency Distribution of the Largest or Smallest Member of a Sample," *Mathematical Proceedings of the Cambridge Philosophical Society*, 24, 180–190. [1335]

Gnedenko, B. V. (1943), "Sur la Distribution Limite du Terme Maximum d'une Serie aléatoire," *Annals of Mathematics*, 44, 423–453. [1335]

Gomes, M. I., and Guillou, A. (2015), "Extreme Value Theory and Statistics of Univariate Extremes: A Review," *International Statistical Review*, 83, 263–292. [1334]

Hill, B. M. (1975), "A Simple General Approach to Inference about the Tail of a Distribution," *Annals of Statistics*, 3, 1163–1174. [1338]

Kiefer, N., and Vogelsang, T. J. (2005), "A New Asymptotic Theory for Heteroskedasticity-Autocorrelation Robust Tests," *Econometric Theory*, 21, 1130–1164. [1334]

Leadbetter, M. R. (1983), "Extremes and local dependence in stationary sequences," *Probability Theory and Related Fields*, 652, 291–306. [1342]

Müller, U. K., and Norets, A. (2016), "Credibility of Confidence Sets in Nonstandard Econometric Problems," *Econometrica*, 84, 2183–2213. [1335]

Müller, U. K., and Watson, M. W. (2015), "Measuring Uncertainty about Long-Run Predictions," *Review of Economic Studies*, forthcoming. [1336,1341,1342]

Pickands, III, J. (1975), "Statistical Inference Using Extreme Order Statistics," *Annals of Statistics*, 31, 119–131. [1334]

Reiss, R.-D., and Thomas, M. (2007), *Statistical Analysis of Extreme Values*, Basel: Birkhäuser Verlag. [1334]

Resnick, S. I. (2007), *Heavy-tail Phenomena: Probabilistic and Statistical Modeling*, New York: Springer. [1334]

Santos, P. A., Alves, M., and Gomes, M. I. (2006), "Peaks Over Random Threshold Methodology for Tail Index and High Quantile Estimation," *Revstat*, 4, 227–247. [1338]

Smith, R. L. (1987), "Estimating Tails of Probability Distributions," *Annals of Statistics*, 15, 1174–1207. [1342]

Smith, R. L. (2004), "Statistics of Extremes, With Applications in Environment, Insurance and Finance," in *Extreme Values in Finance, Telecommunications, and the Environment*, eds. B. Finkenstädt, and H. Rootzén, London: Chapman and Hall, pp. 1–78. [1338]

van der Vaart, A. W. (1998), *Asymptotic Statistics*, Cambridge, UK: Cambridge University Press. [1343]

Weissman, I. (1978), "Estimation of Parameters and Large Quantiles Based on the *k* Largest Observations," *Journal of the American Statistical Association*, 73, 812–815. [1338]

Zhu, L., and Li, H. (2012), "Asymptotic Analysis of Multivariate Tail Conditional Expectations," *North American Actuarial Journal*, 163, 350–363. [1336]