A Pareto Tail Plot Without Moment Restrictions

Bernhard Klar

Published online: 19 Nov 2024.

Submit your article to this journal ⧉

Article views: 305

View related articles ⧉

View Crossmark data ⧉

Citing articles: 1 View citing articles ⧉

Taylor & Francis
Taylor & Francis Group

 OPEN ACCESS   Check for updates

A Pareto Tail Plot Without Moment Restrictions

Bernhard Klar

Institute of Stochastics, Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany

ABSTRACT

We propose a mean functional that exists for arbitrary probability distributions and characterizes the Pareto distribution within the set of distributions with finite left endpoint. This is in sharp contrast to the mean excess plot, which is meaningless for distributions without an existing mean and has nonstandard behavior when the mean is finite, but the second moment does not exist. The construction of the plot is based on the *principle of a single huge jump*, which differentiates between distributions with moderately heavy and super heavy tails. We present an estimator of the tail function based on *U*-statistics and study its large sample properties. Several loss datasets illustrate the use of the new plot.

1. Introduction

Pareto distributions are probably the most important and widely used class of heavy-tailed distributions. A possible parameterization is

$$F(x;\alpha) = 1 - (x_m/x)^{\alpha}, \quad x \geq x_m > 0, \qquad (1)$$

where $\alpha > 0$. The generalized Pareto distribution (GPD) for $\xi > 0$ is parameterized by

$$G(x;\beta,\xi) = 1 - (1 + \xi x/\beta)^{-1/\xi}, \quad x > 0,$$

where $\beta > 0$. For $\xi > 0$, a GPD can be converted to the form in (1) by a location-scale transformation. For $\alpha \leq 1$ (i.e., $\xi \geq 1$), the mean is infinite; if $1 < \alpha \leq 2$ (i.e., $1/2 \leq \xi < 1$), the mean exists, but the variance is infinite.

A main tool for determining the adequacy of the (generalized) Pareto distribution as a model for the tails of sample data is the mean excess (ME) function, also known as the mean residual life function in reliability theory. It is given by

$$M(u) = \mathbb{E}\left[X - u | X > u\right], \quad u > 0,$$

where X is a positive random variable with $\mathbb{E}X < \infty$. In fact, the GPD class with $\xi < 1$ is characterized by the linearity of the ME function (Embrechts, Klüppelberg, and Mikosch 1997). The tail behavior of sample data by visual means can then be explored by the ME plot, that is a plot of the empirical counterpart of the ME function. Given an iid sample $X_1, \ldots, X_n \sim X$, the empirical ME function is defined by

$$\hat{M}(u) = \frac{\sum_{i=1}^{n}(X_i - u)\mathbb{1}\left(X_i > u\right)}{\sum_{i=1}^{n}\mathbb{1}\left(X_i > u\right)}, \quad u > 0.$$

If $X_{(1)} \leq \cdots \leq X_{(n)}$ denote the order statistics of $X_1, \ldots, X_n$, one typically plots the pairs $(X_{(k)}, \hat{M}(X_{(k)}))$ for $1 < k \leq n$, see Ghosh and Resnick (2010) or Das and Ghosh (2016). An unfortunate feature of the ME plot is that it is well-defined only for distributions with finite expectation. Indeed, Ghosh and Resnick (2010) showed that the ME plot converges to a random curve in the case $\xi > 1$, which also holds in the case $\xi = 1$ after suitable rescaling. Thus, the ME plot is inconsistent when $\xi \geq 1$. Consequently, knowledge about the finiteness of the mean is required, a task that faces fundamental difficulties (Romano 2004), although there are some attempts to construct tests for or against the existence of a finite mean (Fedotenkov 2013; Trapani 2016). For statistical inference (e.g., confidence bounds), a finite second moment, that is $\xi < 1/2$ for the generalized Pareto distribution, is required to obtain a normal limit (see Das and Ghosh (2013) for a thorough discussion).

There are several other plotting tools used in connection with heavy tails and extreme values, for example the Pareto QQ plot, plots of estimators of the extreme value index like the Hill plot, see, for example, Drees (2012) or Das and Ghosh (2013).

The chasm between Pareto distributions with and without existing first moment (the latter are called extremely heavy-tailed) has been addressed in several papers over the past decades. For example, having exposures in multiple iid extremely heavy-tailed Pareto losses is worse than having only one Pareto loss of the same total exposure. More formally, it can be shown that a convex combination of such Pareto losses is larger in the usual stochastic order than a single infinite mean loss. This penalizes diversification, which is impossible if the expectation is finite. This phenomenon has been recognized by a number of authors over the years. It is found in the portfolio diversification literature (Samuelson 1967) and in the dependence and value-at-risk literature (Embrechts, McNeil, and Straumann 2002). The most comprehensive set of results is provided in a recent paper by Chen, Embrechts, and Wang (2023).

The importance of loss models without finite variance or even finite mean has been demonstrated by the thorough analysis of
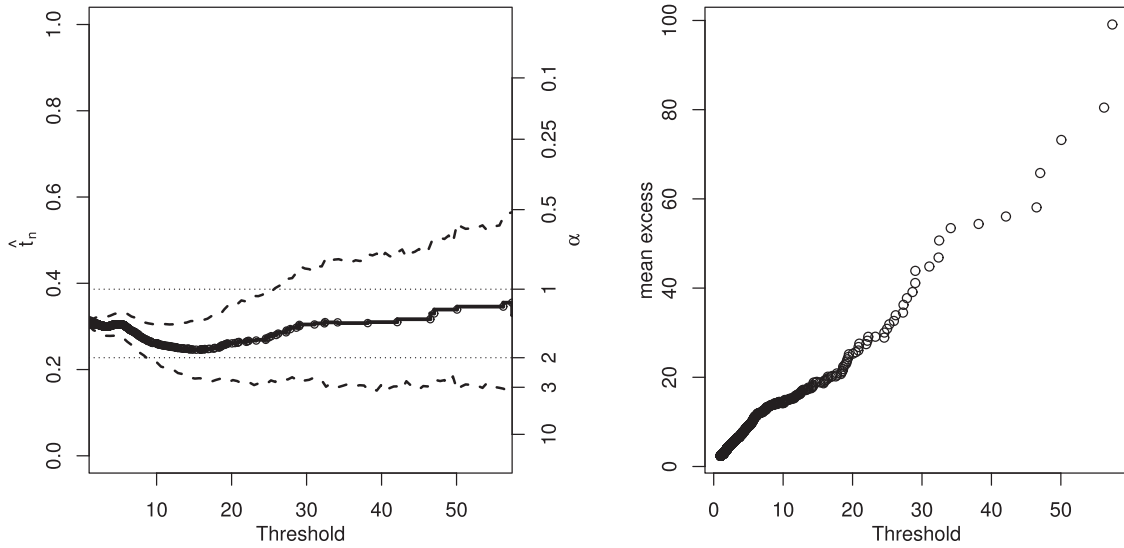
Figure 1. Plots for Danish fire insurance data. Left panel: Graph of the new Pareto tail function; the right scale indicates the corresponding shape parameter α under Pareto model. Right panel: Mean excess function.

various datasets. For an impressive list of such examples, see Chen, Embrechts, and Wang (2023). As another example, Cirillo and Taleb (2020) concluded that the sizes of pandemics, properly transformed into a distribution with unbounded support, have infinite mean. In view of these examples, it is extremely unsatisfactory that the ME plot breaks down in this important range of parameter values where variance or mean does not exist. To circumvent these problems, we propose an alternative mean functional that exists for arbitrary probability distributions and characterizes the Pareto distribution within the set of distributions with finite left endpoint. Specifically, assume that X, X_1, X_2 are iid random variables from an absolutely continuous distribution with support $[x_m, \infty)$. Then, the function

$$t_X(u) = t(u) = \mathbb{E}\left[\frac{|X_1 - X_2|}{X_1 + X_2}\,\middle|\,\min\{X_1, X_2\} \geq u\right] \quad (2)$$

is constant if and only if X has the distribution function $F(\cdot; \alpha)$ for some $\alpha > 0$.

Let us illustrate the use of the plots with two examples. First, we consider large fire insurance claims in Denmark, available as dataset `danish` in the R package `evir` (Pfaff and McNeil 2018). The ME plot of this dataset with sample size 2167 is shown in the right panel of Figure 1. The ME function shows an upward trend, indicating a heavy-tailed distribution. Since it follows a reasonably straight line, we may assume that the data follows a (generalized) Pareto distribution with positive shape parameter α. For this dataset, a detailed analysis using methods from extreme value theory (EVT) is available (McNeil 1997; Resnick 1997; McNeil, Frey, and Embrechts 2015), indicating a heavy-tailed distribution with a tail index α less than 2. Thus, the usual asymptotic results for the ME function assuming a finite second moment are not valid, and its interpretation is difficult. The left panel of Figure 1 shows our new Pareto tail plot. The graph is more or less horizontal with values between 0.25 and 0.35 (left axis), corresponding to values of α between 1.8 and 1.1 (right axis). Notice the bulge for thresholds u between 5 and 15. For the specific values $u = 5, 10, 15$, we get values of 0.30, 0.26, 0.25, corresponding to $\alpha = 1.40, 1.70, 1.82$, respectively. Looking at

the width of the (pointwise) confidence intervals given by the dashed lines, we see that the data are compatible with Pareto models at each of these values. These results are in good agreement with the cited literature: Table 1 in McNeil (1997) gives estimates for α between 1.4 and 2.0; Resnick (1997) concludes "based on an amalgam of the QQ, Hill and moment plots, we settle on an estimate of $\alpha = 1.4$".

Our second example considers a classical dataset of wind catastrophes taken from Hogg and Klugman (1984, p. 64). It represents 40 losses (in million U.S. dollars) due to wind-related disasters. Data are reported to the nearest million, including only losses of 2 million or more. Brazauskas and Serfling (2003) and Rizzo (2009) proposed goodness-of-fit tests for the Pareto model and applied them to the de-grouped wind catastrophes data, and concluded that there were no evidence against the model. In addition to the tests used in these articles, there are a variety of formal tests for Pareto models, see Chu, Dickin, and Nadarajah (2019) for an overview. Estimates for α under this model range from 0.605 to 0.791. The right panel of Figure 2 again shows the mean excess function, which, however, is meaningless for values of α less than 1. The left panel shows the new Pareto tail plot. The graph decreases, starting at 0.45 for the full dataset and decreasing to 0.33 for $u = 6$ and 0.16 for $u = 10$, using only the 15 and 10 largest claims, respectively. Under a Pareto model, these values would correspond to $\alpha = 0.79, 1.26$ and 2.89, but the plot clearly argues against the validity of this model. This is not a contradiction to the above remark: the fact that the goodness-of-fit tests do not reject the hypothesis of a Pareto distribution does not prove that the hypothesis holds.

The article is organized as follows. Section 2 formally presents the new characterization of the Pareto distribution and states some fundamental properties. Section 3 introduces and discusses the intimately connected *principle of a single huge jump*, which differentiates between distributions with moderately heavy and super heavy tails. In Section 4, an estimator of the tail function based on U-statistics is introduced, and its large sample properties are analyzed. Section 5 analyzes three additional datasets. Section 6 concludes the article.
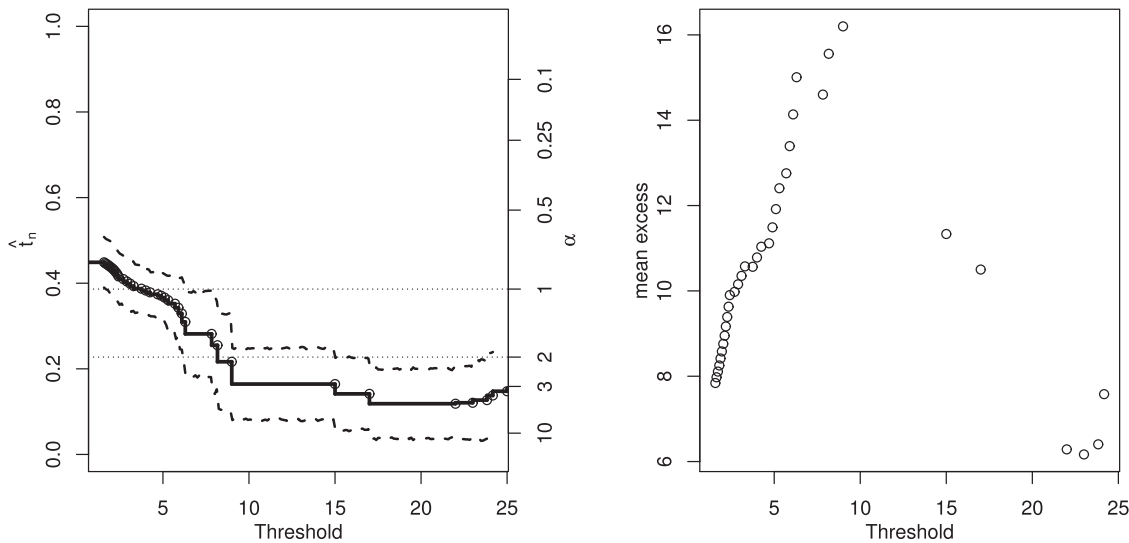
Figure 2. Plots for wind catastrophe losses. Left panel: Graph of the new Pareto tail function; the right scale indicates the corresponding shape parameter α under Pareto model. Right panel: Mean excess function (not meaningful for $\alpha \leq 1$).

2. A New Characterization of the Pareto Distribution

In an abstract, Srivastava (1965) indicated a characterization of the Pareto distribution with distribution function $F(x, \alpha) = 1 - (x_m/x)^\alpha$ for $x \geq x_m > 0$, where $\alpha > 0$, as follows. Let F be an absolutely continuous distribution function with $F(x_m) = 0$, and $X \sim F$. Let $X_{(1)} \leq \cdots \leq X_{(n)}$ denote the order statistic of a sample of size n from F. Then, $X_{(1)}$ and $(X_1 + \cdots + X_n)/X_{(1)}$ are independent if and only if X follows the Pareto distribution. This leads to the following result.

Theorem 1. Assume that $X, X_1, X_2 \sim F$ are iid random variables, where the distribution function F is absolutely continuous with $F(x_m) = 0$. Then, the function $t_X : [x_m, \infty) \to [0, 1]$ defined in (2) satisfies

$$t_X(u) = t_X(x_m), \quad \text{for all } u \geq x_m,$$

if and only if X is Pareto distributed, that is $X \sim F(\cdot, \alpha)$ for some $\alpha > 0$.

If $X \sim F(\cdot, \alpha)$, then, independent of x_m,

$$\widetilde{t}_\alpha = t_X(x_m) = \begin{cases} 2\alpha \sum_{k=1}^{\alpha-1} \frac{(-1)^{\alpha+1-k}}{k} \\ \quad + (-1)^{\alpha+1} 2\alpha \log 2 - 1, & \alpha \in \mathbb{N}, \quad (3) \\ \alpha \left(\Psi\left(\frac{\alpha+1}{2}\right) - \Psi\left(\frac{\alpha}{2}\right) \right) - 1, & \alpha > 0, \alpha \notin \mathbb{N}, \end{cases}$$

where $\Psi(z) = \frac{d}{dz} \log \Gamma(z)$ denotes the digamma function.

For Pareto distributed random variables with shape parameter $\alpha > 0$, the proof of Theorem 1 in Appendix A shows that $\widetilde{t}_\alpha = 2 \int_0^1 \alpha y^{\alpha-1}/(1 + y) dy - 1$. Using integration by parts, we obtain

$$\widetilde{t}_\alpha = 2 \int_0^1 \frac{y^\alpha}{(1 + y)^2} \, dy.$$

This representation reveals that $\widetilde{t}_\alpha$ is strictly decreasing as a function of α, with $\lim \widetilde{t}_\alpha = 1$ for $\alpha \to 0$, and $\lim \widetilde{t}_\alpha = 0$ for $\alpha \to \infty$. Specific values are given by $\widetilde{t}_{1/2} = \pi/2 - 1 \approx 0.571, \widetilde{t}_1 = 2 \log 2 - 1 \approx 0.368, \widetilde{t}_2 = 3 - 4 \log 2 \approx 0.227$ and $\widetilde{t}_3 = -4 + 6 \log 2 \approx 0.159$.

Since the conditional distribution of a Pareto-distributed random variable X, given $X > y$ (where $y > x_m$), is a Pareto distribution with the same α but with minimum y instead of x_m, we have $t_{X|X>y}(u) = t_X(u)$ for $u \geq y$. Furthermore, the function t satisfies $t_{cX}(u) = t_X(u/c)$ for $c > 0$ and $u \geq cx_m$. It has the following interpretation (see Asmussen and Lehtomaa 2017): if both X_1 and X_2 contribute equally to the sum $X_1 + X_2$, then t should eventually obtain values close to 0; if only one of the variables tends to be of the same magnitude as the whole sum, then t is close to 1 for large u. However, in contrast to the functional g considered in Asmussen and Lehtomaa (2017) and Iwashita and Klar (2023), this evaluation is performed only if both X_1 and X_2 are large, and not already if the sum is large. This interpretation will be explored further in the next section.

3. The Principle of a Single Huge Jump

Let X_1 and X_2 be independent random variables, having the same Weibull distribution with shape parameter k. To illustrate the *principle of a single big jump*, Foss, Korshunov, and Zachary (2013) considered the distribution of the random variable $X_1/X_1 + X_2$ conditional on the sum $X_1 + X_2 = u$ for increasing values of u. They showed that for $k < 1$, that is for heavy-tailed distributions, this distribution converges for $u \to \infty$ to $(\delta_0 + \delta_1)/2$, where δ_x denotes the Dirac measure in x. Here, the distribution (function) F is said to be *heavy-tailed* if

$$\int_{-\infty}^{\infty} e^{\lambda x} F(dx) = \infty \quad \text{for all } \lambda > 0,$$

otherwise F is said to be light-tailed (Foss, Korshunov, and Zachary 2013). For Weibull distributions with $k > 1$, that is light-tailed distributions, and as $u \to \infty$, the distribution converges to $\delta_{1/2}$. For $k = 1$, that is for the exponential distribution, there is no concentration of mass. Further results in this direction have been obtained by Lehtomaa (2015).

Instead of X_1/u, we can consider the distribution of the random variable $(X_{(2)} - X_{(1)})/u$ conditional on the sum $X_1 + X_2 = u$, which converges to δ_1 for $k < 1$, and to δ_0 for $k > 1$ as

$u \to \infty$. The first behavior is typical for heavy-tailed variables, and is an example of the principle of a single big jump: if the sum $X_1 + X_2$ is large, then one of the variables is large compared to the other. For light-tailed distributions, however, both of the variables X_1 and X_2 contribute equally, and the difference is small compared to the sum.

Now let X_1 and X_2 be nonnegative and iid random variables. Assume that X_1 has unbounded support (x_m, ∞) and density function f. Let us consider a similar setting as above, but where we condition on $\min\{X_1, X_2\} = u$ instead of $X_1 + X_2 = u$. We will see that this change leads to a formally similar partition within the distributions with heavy tails. Since the joint density of the order statistics $(X_{(1)}, X_{(2)})$ is given by $f_{()}(x_1, x_2) = 2f(x_1)f(x_2)$ for $x_1 < x_2$, the joint density of $(Y_1, Y_2) = (X_{(1)}, (X_{(2)} - X_{(1)})/(X_{(1)} + X_{(2)}))$ is

$$g(y_1, y_2) = 2f(y_1)f\left(\frac{y_1(1 + y_2)}{1 - y_2}\right)\frac{2y_1}{(1 - y_2)^2},$$
$$y_1 > x_m, \; 0 < y_2 < 1.$$

Division by the density of the first order statistics yields the density of the random variable $Z = (X_{(2)} - X_{(1)})/(X_{(1)} + X_{(2)})$ conditional on the minimum $X_{(1)} = u$ as

$$g(z|u) = \frac{2uf\left(\frac{u(1+z)}{1-z}\right)}{(1 - z)^2 \, \bar{F}(u)}, \quad 0 < z < 1, \, u > x_m. \tag{4}$$

Example 2. (a) For the Weibull distribution with shape parameter k, we obtain

$$g(z|u) = 2ku^k(1 + z)^{k-1}(1 - z)^{-(k+1)}$$
$$\exp\left\{-u^k\left[((1 + z)/(1 - z))^k - 1\right]\right\}$$

for $0 < z < 1, u > 0$. It follows that the conditional distribution converges to a one-point distribution in 0 as $u \to \infty$ for arbitrary k. Thus, for heavy-tailed Weibull distributions, the distribution of Z conditional on $X_{(1)} = u$ behaves completely different from conditioning on $X_1 + X_2 = u$ as $u \to \infty$.

(b) For the Pareto distribution with shape parameter α, the conditional density is

$$g(z|u) = \tilde{g}_\alpha(z) = 2\alpha \, (1 + z)^{-(\alpha+1)}(1 - z)^{\alpha-1},$$
$$0 < z < 1,$$

for all $u > x_m$. The conditional distribution does not depend on u, in agreement with Theorem 1. Thus, there is no concentration of mass and the family of Pareto distributions takes over the role of the exponential distributions in the first scenario.

(c) Denote by $\mathcal{R}_\alpha$ the class of regularly varying distributions, where $\bar{F}(x) = L(x)/x^\alpha$ with $\alpha > 0$ and $L(\cdot)$ is slowly varying, that is $\lim_{x\to\infty} L(tx)/L(x) = 1$ for $t > 0$. Assume that F has an ultimately decreasing density f, that is there exists c such that f is decreasing on $[c, \infty)$. This ensures that $f(x) = \alpha \tilde{L}(x)/x^{\alpha+1}$ with $\tilde{L}(x) \sim L(x)$ as $x \to \infty$ (Embrechts, Klüppelberg, and Mikosch 1997, Th. A.3.7). Here, $f(x) \sim g(x)$ as $x \to \infty$ means that $\lim_{x\to\infty} f(x)/g(x) = 1$. Then, the conditional density is

$$g(z|u) = 2\alpha \frac{(1 - z)^{\alpha-1}}{(1 + z)^{\alpha+1}} \frac{\tilde{L}\left(\frac{u(1+z)}{1-z}\right)}{L(u)}, \quad 0 < z < 1,$$

for all $u > x_m$. Therefore, $g(z|u) \sim \tilde{g}_\alpha(z)$ as $u \to \infty$. Thus, the limiting behavior of $g(z|u)$ for distributions in $\mathcal{R}_\alpha$ coincides with that of a Pareto distribution with the same shape parameter.

The left panel in Figure 3 shows graphs of $g(z|u)$ for the log-gamma distribution with density

$$f(x) = \frac{\alpha^\beta}{\Gamma(\beta)}(\log x)^{\beta-1}x^{-\alpha-1}, \quad x \geq 1 \quad (\alpha, \beta > 0),$$

for increasing values of u. The conditional density $g(z|u)$ of this regularly varying distribution converges to $\tilde{g}_\alpha(z)$ (plotted in red) as $u \to \infty$.

(d) The right panel in Figure 3 shows graphs of $g(z|u)$ for the (standard) log-Cauchy distribution, a distribution with heavier tails than Pareto distributions. Here, the conditional distribution of Z converges to a one point distribution in 1 as $u \to \infty$.
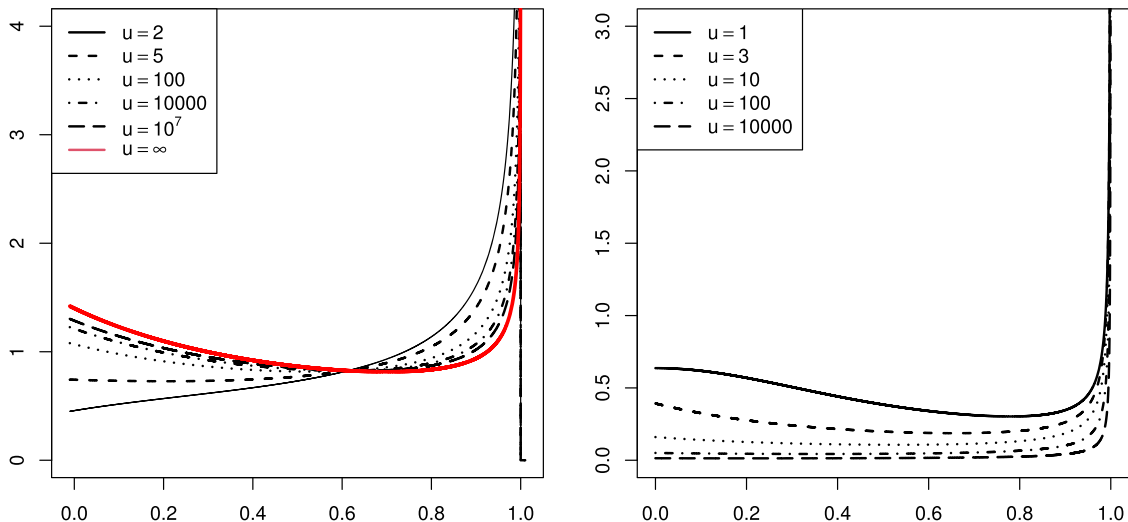
Figure 3. Plots of the conditional density $g(z|u)$ in (4) for log-gamma (with $\alpha = 0.7, \beta = 2$, left panel) and log-Cauchy distribution (right panel) for increasing values of the boundary. Left panel: $g(z|u)$ (in black) converges to $\tilde{g}_\alpha(z)$ (in red) as $u \to \infty$. Right panel: $g(z|u)$ converges to δ_1 as $u \to \infty$.

The last example of the log-Cauchy distribution shows an effect which can be called the *principle of a single huge jump*: if each of the variables X_1 and X_2 takes values over a large boundary, then one of the variables will still typically be much larger than the other. Distributions that follow this principle, and hence, $g(z|u) \to \delta_1$ as $u \to \infty$, could be called *super heavy-tailed*. There are other definitions of super heavy-tailed distributions in the literature (Alves, de Haan, and Neves 2009; Cormann and Reiss 2009); their connections to the definition given here require further investigation. On the contrary, if we consider observations X_1 and X_2 over a large boundary for Weibull distributions, then X_1 and X_2 contribute equally, and the difference is small compared to the sum. Thus, such distributions, where $g(z|u) \to \delta_0$ as $u \to \infty$, are (at most) moderately heavy-tailed. In between are the Pareto distributions for arbitrary shape parameters, where a limiting density of $g(z|u)$ exists without any concentration of probability mass.

The shape of the conditional density in Example 2 c) suggests that the limiting behavior of $t(u)$ as $u \to \infty$ for the class of regularly varying distributions is the same as for Pareto distributions. If L is slowly varying, then $L(ux)/L(x) \to 1$ as $x \to \infty$ uniformly over compact u sets (Embrechts, Klüppelberg, and Mikosch 1997, Th. A.3.2). Strengthening this property, we can indeed find the limiting behavior of t for a subset of regularly varying distributions; the proof is given in Appendix B.

Theorem 3. Assume that $X, X_1, X_2 \in \mathcal{R}_\alpha$ are iid random variables, where the distribution is absolutely continuous with an ultimately decreasing density. Assume further that $\lim_{x\to\infty} L(ux)/L(x) = 1$ uniformly in u on (x_m, ∞). Then,

$$\lim_{u\to\infty} t_X(u) = \widetilde{t}_\alpha,$$

where $\widetilde{t}_\alpha$ is defined in Theorem 1.

4. An Estimator of $t(u)$ based on U-statistics

To estimate t based on the order statistics $X_{(1)} \leq \cdots \leq X_{(n)}$, define U-statistics

$$U_n^{(1)}(u) = \frac{2}{n(n-1)} \sum_{1 \leq i < j \leq n} \frac{X_{(j)} - X_{(i)}}{X_{(i)} + X_{(j)}} \mathbb{1}\left(X_{(i)} \geq u\right),$$

$$U_n^{(2)}(u) = \frac{2}{n(n-1)} \sum_{1 \leq i < j \leq n} \mathbb{1}\left(X_{(i)} \geq u\right)$$

$$= \frac{2}{n(n-1)} \sum_{i=1}^{n} (n-i)\, \mathbb{1}\left(X_{(i)} \geq u\right),$$

where $\mathbb{1}(B)$ is the indicator function of the event B. A nonparametric estimator of $t(u)$ is then given by the ratio of these statistics:

$$\hat{t}_n(u) = \frac{U_n^{(1)}(u)}{U_n^{(2)}(u)}, \quad u \geq x_m.$$

Note that it suffices to evaluate $\hat{t}_n$ at the sample points, yielding

$$\hat{t}_n\left(X_{(k)}\right) = \frac{1}{\binom{n-k+1}{2}} \sum_{k \leq i < j} \frac{X_{(j)} - X_{(i)}}{X_{(i)} + X_{(j)}}, \quad k = 1, \ldots, n-1.$$

Hence, $\hat{t}_n$ itself can be seen as a U-statistic, applied to the $n-k+1$ largest observations. For computational purposes, note that

$$\hat{t}_n\left(X_{(k)}\right) = \frac{n-k+2}{n-k} \hat{t}_n\left(X_{(k-1)}\right)$$

$$- \frac{1}{\binom{n-k+1}{2}} \sum_{j=k}^{n} \frac{X_{(j)} - X_{(k-1)}}{X_{(k-1)} + X_{(j)}}, \quad k = 2, \ldots, n-1.$$

This can be evaluated very quickly, starting with $\hat{t}_n(X_{(n-1)}) = (X_{(n)} - X_{(n-1)})/(X_{(n)} + X_{(n-1)})$.

Figure 4 shows the graphs of $\hat{t}_n(u)$ for simulated samples with sample size 10,000 from different distributions on $(1, \infty)$. The bound u is in the range from 1 to the 0.995-quantile of the sample. The two dotted horizontal lines in the plots are given by $\widetilde{t}_1 \approx 0.39$ and $\widetilde{t}_2 \approx 0.23$ in (3). The top row shows the Pareto distribution with shape parameter 0.5 and the Pareto distributions of type 2 and 3 with scale parameter $\theta = 5$ and shape parameter $\alpha = 1.5$ and 3, respectively. Their densities are given by

$$f_2(x) = \frac{\alpha}{\theta(1 + (x-1)/\theta)^{\alpha+1}},$$

$$f_3(x) = \frac{\alpha((x-1)/\theta)^{\alpha-1}}{\theta\left(1 + ((x-1)/\theta)^\alpha\right)^2}, \quad x > 1,$$

and satisfy the assumptions of Theorem 3. As expected, $\hat{t}_n(u)$ is nearly constant after a short settling period. The middle row shows log-gamma distributions with $\alpha = 0.5, 1.5, 3$ and $\beta = 2$. These regularly varying distributions don't satisfy the uniform convergence assumption in Theorem 3. Nevertheless, $\hat{t}_n$ seems to converge to the limits corresponding to the shape parameter α. Finally, the bottom row shows shifted gamma distributions, whose conditional densities $g(z|u))$ converge to δ_0, as it is the case for Weibull distributions. Inspection of the proof of Theorem 3 suggests that in this case, $t(u)$ converges to 0 as $u \to \infty$, and the plots support this suggestion.

Remark 4. By the results in Section 2, $\widetilde{t}_\alpha$ is strictly decreasing in α. Thus, one can estimate the parameter α of the Pareto distribution using $\hat{t}_n$. More generally, one can estimate the tail index of a distribution with Pareto tails (for various possible definitions see Fedotenkov (2020), Assumptions A1–A8), based on the k largest observations of a sample. This adds another method to the more than one hundred Pareto tail index estimators listed in Fedotenkov (2020).

4.1. Large Sample Properties of $\hat{t}_n$

The U-statistic $U_n^{(l)}(u)$ are unbiased estimators of $\theta_u^{(l)} = \mathbb{E}[h^{(l)}(X_1, X_2; u)], l = 1, 2$, where the kernels $h^{(l)}$ of degree 2 are defined by

$$h^{(1)}(x_1, x_2; u) = \frac{|x_1 - x_2|}{x_1 + x_2} \mathbb{1}\left(\min\{x_1, x_2\} > u\right),$$

$$h^{(2)}(x_1, x_2; u) = \mathbb{1}\left(\min\{x_1, x_2\} > u\right).$$
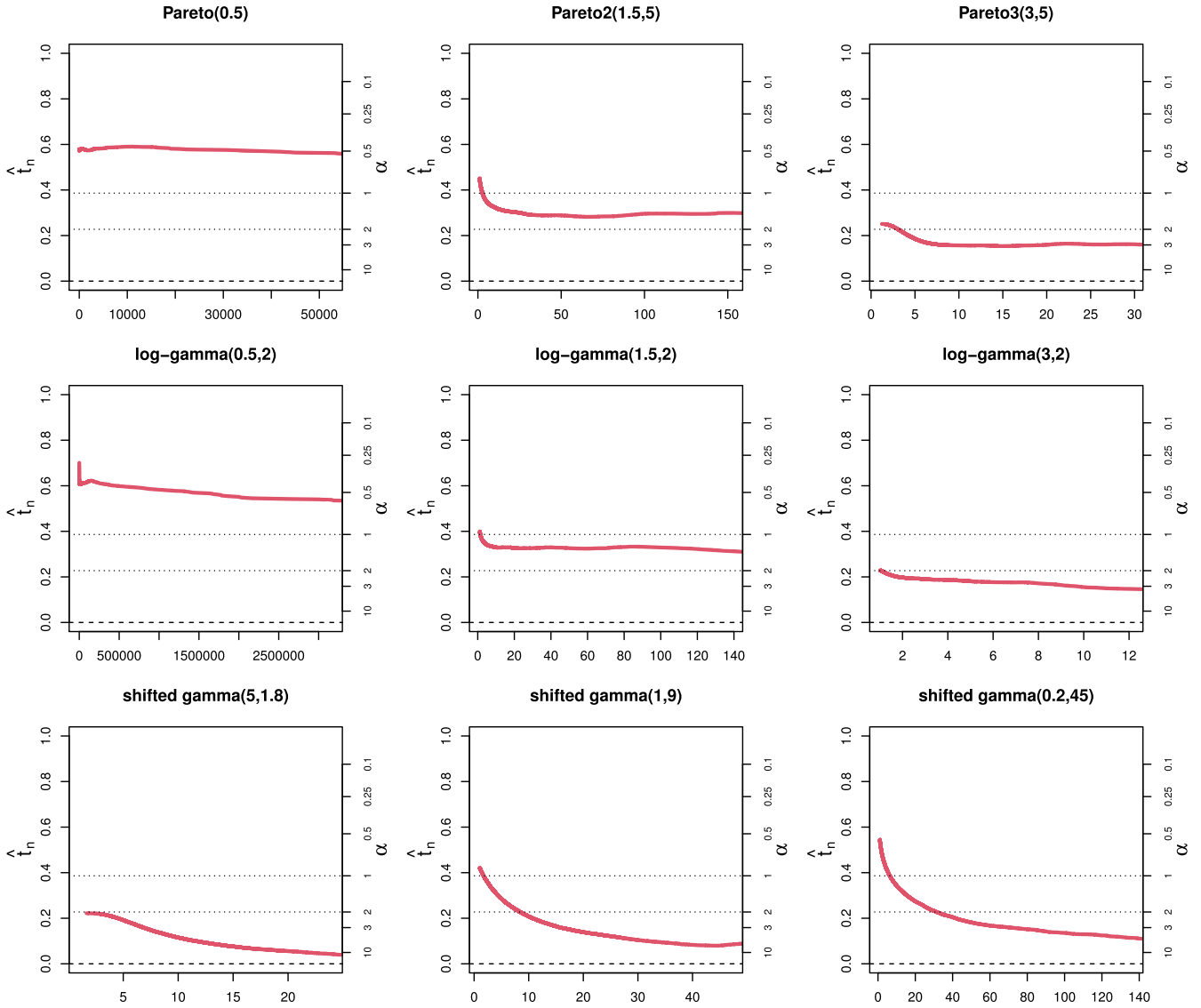
Figure 4. Graphs of $\hat{t}_n(u)$ for simulated samples from different distributions on $(1,\infty)$ with sample size 10,000. Upper row: Pareto distributions of type 1,2,3 with shape parameter 0.5, 1.5, 3. Middle row: Log-gamma distributions with $\alpha = 0.5, 1.5, 3$ and $\beta = 2$. Lower row: Shifted gamma distributions, each with expected value 10.

By the strong law of large numbers for U-statistics (Lee 1990, p. 122), $U_n^{(l)}(u), l = 1, 2$, and hence, $\hat{t}_n(u)$ are strongly consistent estimators for $\theta_u^{(l)}$ and $t(u)$, respectively. Using the general theory of U-statistics (see Lee (1990, p. 76) or Korolyuk and Borovskich (1994, p. 132)), we can derive the joint asymptotic distribution of $(U_n^{(1)}(u), U_n^{(2)}(u))$.

Proposition 5. For $l = 1, 2$, let

$$\psi^{(l)}(x_1, x_2; u) = h^{(l)}(x_1, x_2; u) - \theta_u^{(l)},$$

$$\psi_1^{(l)}(x_1; u) = \mathbb{E}[\psi^{(l)}(x_1, X_2; u)].$$

Further, define

$$\eta_1^{(l)}(u) = \mathbb{E}\left[\left(\psi_1^{(l)}(X_1; u)\right)^2\right] \quad (l = 1, 2),$$

$$\eta_1^{(1,2)}(u) = \mathbb{E}\left[\psi_1^{(1)}(X_1; u)\,\psi_1^{(2)}(X_1; u)\right].$$

If $\eta_1^{(l)}(u) > 0$ for $l = 1, 2$, then,

$$\sqrt{n}\left(\begin{pmatrix} U_n^{(1)}(d) \\ U_n^{(2)}(u) \end{pmatrix} - \begin{pmatrix} \theta_u^{(1)} \\ \theta_u^{(2)} \end{pmatrix}\right) \xrightarrow{\mathcal{D}} N_2(\mathbf{0}, 4\Sigma_u), \quad \text{where}$$

$$\Sigma_u = \begin{pmatrix} \eta_1^{(1)}(u) & \eta_1^{(1,2)}(u) \\ \eta_1^{(1,2)}(u) & \eta_1^{(2)}(u) \end{pmatrix}.$$

Using Proposition 5 and the delta method, we can derive the asymptotic behavior of $\hat{t}_n(u)$.

Theorem 6. Let $v_u = \theta_u^{(2)} > 0$, and $\eta_1^{(l)}(u) > 0$ for $l = 1, 2$. Then,

$$\sqrt{nv_u}\left(\hat{t}_n(u) - t(u)\right) \xrightarrow{\mathcal{D}} N\left(0, \sigma_u^2\right), \qquad (5)$$

where

$$\sigma_u^2 = \frac{4}{v_u}\left(\eta_1^{(1)}(u) - 2t(u)\,\eta_1^{(1,2)}(u) + t^2(u)\,\eta_1^{(2)}(u)\right). \qquad (6)$$

Table 1. Empirical coverage probability of 0.95-confidence intervals for $t(u)$ based on different estimators for effective sample size $n_{\text{eff}} = 20$, $u = 3$ and varying α.

α	$\hat{\sigma}_u^2$	$\hat{\sigma}_{B,u}^2$	$\hat{\sigma}_{J,u}^2$
0.2	90.0	93.2	93.2
0.5	91.6	92.3	93.4
1.0	92.0	92.1	93.3
2.0	93.0	93.0	93.7
3.0	93.5	93.4	93.8

Table 2. Empirical coverage probability of 0.95-confidence intervals for $t(u)$ based on different estimators for $\alpha = 1$, $u = 2$ and increasing effective sample size.

n_{eff}	$\hat{\sigma}_u^2$	$\hat{\sigma}_{B,u}^2$	$\hat{\sigma}_{J,u}^2$
10.0	87.9	89.4	91.1
20.0	91.8	91.7	93.0
40.0	93.2	93.0	93.7
80.0	94.5	94.5	94.7
160.0	94.9	94.8	95.0

Next, we want to construct confidence intervals for $t(u)$. Based on Theorem 6, a confidence interval with asymptotic coverage probability $1 - \gamma$ is given by

$$\left[\max\left\{ \hat{t}_n(u) - \frac{z_{1-\gamma/2}\,\hat{\sigma}_u}{(nU_n^{(2)}(u))^{1/2}}, 0 \right\},\right.$$
$$\left. \min\left\{ \hat{t}_n(u) + \frac{z_{1-\gamma/2}\,\hat{\sigma}_u}{(nU_n^{(2)}(u))^{1/2}}, 1 \right\} \right],$$

where $z_p = \Phi^{-1}(p)$, and $\hat{\sigma}_u^2$ is an estimator of the variance of $\sqrt{n v_d}\,\hat{t}_n(u)$, corresponding to σ_u^2 in (6). A possible approach is to derive unbiased estimators $\sigma_{U^{(l)}}^2 (l = 1, 2)$ and $\sigma_{U^{(1,2)}}^2$ of $\text{var}(U_n^{(l)}(u))$ and $\text{cov}(U_n^{(1)}(u), U_n^{(2)}(u))$, respectively. This can be done efficiently with complexity $O(n^2)$, see Appendix C. Then, an estimator of σ_u^2 is given by

$$\hat{\sigma}_u^2 = \frac{n}{U_n^{(2)}(d)} \left(\sigma_{U^{(1)}}^2 - 2\hat{t}_n(u)\,\sigma_{U^{(1,2)}}^2 + \hat{t}_n^2(u)\,\sigma_{U^{(2)}}^2 \right). \quad (7)$$

Further possible estimators are the bootstrap variance estimator $\hat{\sigma}_{B,u}^2$ and the jackknife estimator $\hat{\sigma}_{J,u}^2$. Details can be found in Shirahata and Sakamoto (1992) and Iwashita and Klar (2023).

4.2. Empirical Coverage Probability of Confidence Intervals for $t(u)$

In this section, we empirically study the coverage probabilities of confidence intervals for $t(u)$ based on the three variance estimators introduced in Section 5. As distributions, we choose Pareto distributions with $x_m = 1$ and different shape parameters α. In all simulations, we use effective sample sizes, defined as follows: for given values of α and u, the total sample size n was chosen such that $n v_u = n P(\min\{X_1, X_2\} > u) = n_{\text{eff}}$. All computations have been performed using the statistic program R (R Core Team 2023).

The results for confidence level $1 - \gamma = 0.95$, $n_{\text{eff}} = 20$, $u = 2$ and varying α are given in Table 1. For this small sample size, all intervals are anticonservative, that is they have coverage probability less than 0.95. The jackknife estimator $\hat{\sigma}_{J,u}^2$ performs slightly better than the competitors for $\alpha \geq 0.5$. For $\alpha = 0.2$, the coverage probability with $\hat{\sigma}_u^2$ is as low as 0.90, while the other two estimators work well.

Table 2 shows the results for $1 - \gamma = 0.95$, $\alpha = 1$, $u = 2$ and increasing sample sizes. For $n_{\text{eff}} = 40$, $\hat{\sigma}_{J,u}^2$ still has the edge over the competitors. For $n_{\text{eff}} = 80$ or larger, all intervals work very well. Since computing time of $\hat{\sigma}_{J,u}^2$ and of $\hat{\sigma}_{B,u}^2$, which was computed with 999 bootstrap replications, becomes a problem for large samples, we recommend to use $\hat{\sigma}_{J,u}^2$ for small to moderate samples, and to switch to $\hat{\sigma}_u^2$ for large samples sizes.

5. Empirical Illustrations

In this section, we describe and analyze three additional datasets: *French marine losses (2006), wildfire suppression costs (1995), costs of nuclear power accidents (2016)*. The first two datasets were also considered in Chen, Embrechts, and Wang (2023).
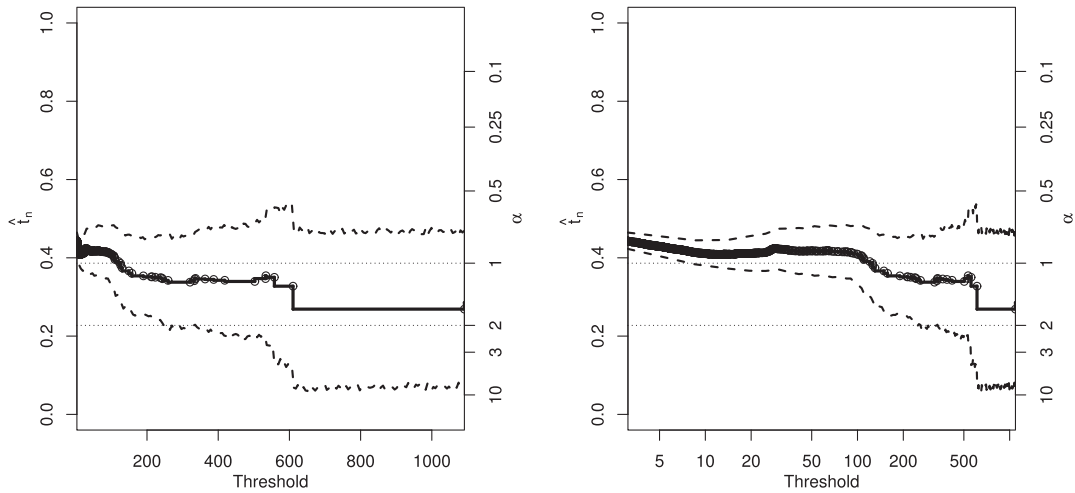
5.1. French Marine Losses

The marine losses dataset from the insurance data repository CASdatasets, available at *http://cas.uqam.ca/+*, comes from a French private insurer and includes 1274 marine losses between January 2003 and June 2006. We consider only the paid amount of the claims (which have been rescaled to mask the actual losses) over 3 monetary units. This results in a sample of size 657. The Pareto tail plot on the original scale is shown in the left panel of Figure 5; in the right panel, the horizontal axis is log-scaled for better visualization. For smaller thresholds up to 100, the function $\hat{t}_n$ takes values slightly above 0.4, corresponding to a shape parameter below 1. For example, for $u = 20, 50, 100$, which corresponds to using the $167, 72, 37$ largest data points, the values of $\hat{t}_n(u)$ are given by 0.411, 0.418, 0.408, in turn corresponding to values of α around 0.91, 0.89, 0.92, respectively. Then, $\hat{t}_n$ decreases to values around 0.35. For example, $\hat{t}_n(300) = 0.338$, which corresponds to $\alpha = 1.21$. However, in this region, the confidence intervals indicate a high uncertainty in the estimates, not contradicting the assumption of a Pareto model with a shape parameter below 1. While plotting the mean excess function makes no sense for this dataset, Chen, Embrechts, and Wang (2023) present a Hill plot of this dataset. Using a threshold around the top 5% of the order statistics of the data yields a Hill estimate of 0.916 for the parameter α, which is in good agreement with our analysis.
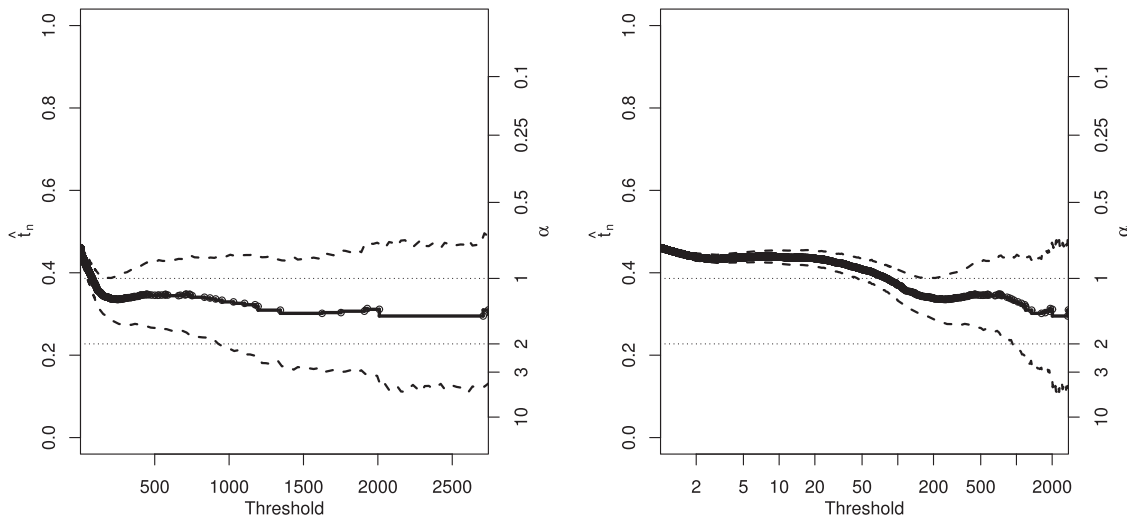
5.2. Wildfire Suppression Costs

The wildfire dataset contains 10,915 suppression costs (in Canadian dollars) collected in Alberta from 1983 to 1995. We only consider costs above 1000 dollars and divide them by 1000, resulting in a dataset of size 6599, with minimum value of 1. Therefore, the Pareto tail plot in the left panel in Figure 6 shows the costs in 1000 dollars on the x-axis, while it is log-scaled in the right panel. Specific values are given in Table 3. For small and medium thresholds, the shape parameter is estimated to be less than 1. For $u = 25$, which is approximately the upper 5% quantile of the complete dataset, the estimate of α is 0.85, essentially identical to the value given in Chen, Embrechts, and Wang (2023). This suggests a distribution with infinite mean. However, for larger thresholds, $u = 150$ and above, the func-

Figure 5. Plots for dataset on French marine losses, $n = 657$. Left panel: Graph of $\hat{t}_n(u)$; the right scale indicates the corresponding shape parameter α under the assumption that the data follows a Pareto distribution. Right panel: horizontal axis in log scale.

Figure 6. Plots for Wildfire suppression costs in Section 5.2, $n = 6599$. Left panel: Graph of $\hat{t}_n(u)$; the right scale indicates the corresponding shape parameter α under the assumption that the tail data follows a Pareto distribution. Right panel: horizontal axis in log scale.

Table 3. Values of $\hat{t}_n(u)$ for the wildfire suppression cost data for specific values of the threshold u and the corresponding shape parameter α under the assumption that the tail data follow a Pareto model.

Threshold u	# observations $\geq u$	$\hat{t}_n(u)$	Corresponding α
5	2197	0.44	0.82
25	571	0.43	0.85
60	293	0.40	0.94
150	142	0.35	1.17
400	50	0.34	1.18

tion $\hat{t}_n$ drops well below 0.368 and becomes stable, with values around 0.34, corresponding to $\alpha = 1.20$. For $u = 150$, the estimate is still based on 142 observations, leading to rather narrow confidence intervals. This raises doubts about an infinite mean model and rather supports a model with a finite mean, but an infinite second moment.

5.3. Costs of Nuclear Power Accidents

A dataset of 216 incidents and accidents occurring at nuclear power plants, with costs in millions of U.S. dollars, is available under *https://data.world/rebeccaclay/nuclear-power-accidents*. Considering only costs over 10 million dollars results in a dataset of size 125. The Pareto tail plot on the original scale is shown in the left panel of Figure 7; in the right panel, the horizontal axis is logarithmically scaled. The Pareto tail function $\hat{t}_n$ takes values between 0.45 and 0.63, corresponding to values of α between 0.78 and 0.40. For this dataset, the graph does not really show stability, even in the tail. On the other hand, even when considering the full dataset, it is quite small, so that a Pareto tail is not necessarily excluded. Consider the specific thresholds $u = 75, 500, 1640$, which roughly correspond to the $0.5, 0.75$, and 0.90 quantiles, using the $64, 31$, and 13 largest data points. The pertaining values of $\hat{t}_n(u)$ are given by $0.58, 0.46, 0.54$, corresponding to values of α given by $0.49, 0.76, 0.57$, respectively. Taking into account the width of the confidence intervals in Figure 7 shows that a Pareto distribution with tail function $t \equiv 0.5$, corresponding to a shape parameter around 0.65, is a possible model in the tail. In any case, it is very likely that the underlying distribution has an infinite mean value.
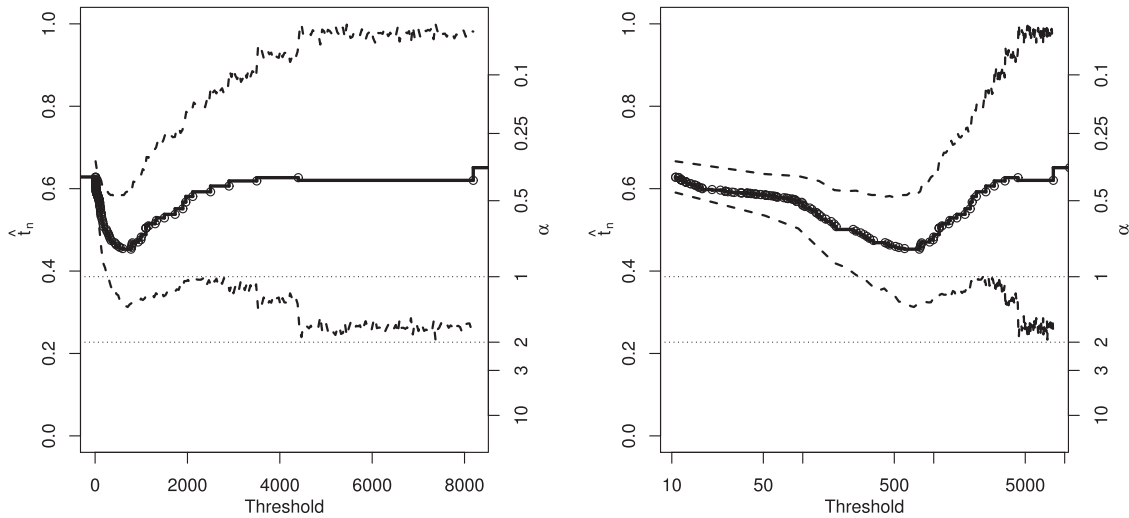
Figure 7. Plots for costs of nuclear power accidents in Section 5.3, $n = 125$. Left panel: Graph of $\hat{t}_n(u)$; the right scale indicates the corresponding shape parameter α under the assumption that the data follows a Pareto distribution. Right panel: horizontal axis in log scale.

6. Concluding Remarks

This article proposes a graphical method for assessing the validity of a Pareto model, especially with respect to the tails of the distribution. It is a fully nonparametric approach, applicable to all continuous loss distributions without any moment restrictions. This is in sharp contrast to the mean excess plot, which plays a fundamental role in many fields, but is of limited value when dealing with heavy tails, although it is available in virtually all EVT software (Belzile et al. 2023, Table 4). Furthermore, the proposed graphical approach can also be used as a *threshold stability plot*, where we choose the lowest threshold above which $\hat{t}_n$ is approximately constant, taking into account the uncertainty quantified by the interval estimates. In this respect, it would we worthwhile to construct simultaneous confidence bands; for the mean excess plot, such bands have been derived by Das and Ghosh (2013). The basic principle underlying the plot can also be interpreted as the *principle of a single huge jump*, distinguishing between distributions with moderate and super-heavy tails. Further research is needed to realize the full potential of this concept.

Appendix A. Proof of Theorem 1

Proof. For $n = 2$, Srivastava's characterization yields that $X_{(1)}$ and $W = X_{(1)}/(X_{(1)}+X_{(2)})$ as well as $X_{(1)}$ and $1-W = X_{(2)}/(X_{(1)}+X_{(2)})$ are independent if and only if X follows the Pareto distribution. Hence, the independence condition is equivalent to the condition that $X_{(1)}$ and $1-2W = (X_{(2)} - X_{(1)})/(X_{(1)} + X_{(2)})$ are independent, which yields the first assertion.

Now, assume that $X, X_1, X_2 \sim F(\cdot; \alpha)$ with density $f(\cdot; \alpha)$. The joint probability density function of $(X_{(1)}, X_{(2)})$ is given by

$$f(x_1, x_2) = 2f(x_1; \alpha)f(x_2; \alpha) = 2\alpha^2 x_m^{2\alpha} (x_1 x_2)^{-(\alpha+1)},$$

$$x_m \leq x_1 \leq x_2.$$

It follows that $Y = (X_{(2)} - X_{(1)})/(X_{(1)} + X_{(2)})$ has density function

$$g(y) = 2\alpha (1 + y)^{-(\alpha+1)}(1 - y)^{\alpha-1}, \quad 0 \leq y < 1.$$

Thus, the computation of $\widetilde{t}_\alpha$ boils down to the evaluation of

$$\widetilde{t}_\alpha = \mathbb{E}Y = \int_0^1 2\alpha (1 + y)^{-\alpha}(1 - y)^{\alpha-1} \, dy - 1 = \check{t}_\alpha - 1, \quad \text{(A1)}$$

say. First, assume that $\alpha \in \mathbb{N}, \alpha \geq 2$. Using integration by parts, we obtain

$$\check{t}_\alpha = \frac{\alpha}{\alpha - 1} \left(2 - \check{t}_{\alpha-1}\right),$$

and a recursive application results in

$$\check{t}_\alpha = 2\alpha \sum_{k=1}^{\alpha-1} \frac{(-1)^{\alpha+1-k}}{k} + (-1)^{\alpha+1} \alpha \check{t}_1.$$

Since $\check{t}_1 = 2 \log 2$, the first part of formula (3) holds. Next, assume $\alpha \notin \mathbb{N}$. In this case, (3) follows by noting that

$$\frac{\check{t}_\alpha}{2\alpha} = \int_0^1 \frac{y^{\alpha-1}}{1 + y} \, dy = \beta(\alpha),$$

where $\beta(x) = \sum_{k=0}^\infty (-1)^k/(x + k) = (\Psi((x + 1)/2) - \Psi(x/2))/2$ (Gradshteyn and Ryzhik 2000, pp. 292, 947). □

Appendix B. Proof of Theorem 3

With $Z = (X_{(2)} - X_{(1)})/(X_{(1)} + X_{(2)})$, we obtain

$$t_X(u) = \mathbb{E}\left[Z|X_{(1)} \geq u\right] = \frac{\mathbb{E}\left[Z \cdot \mathbb{1}\{X_{(1)} \geq u\}\right]}{P(X_{(1)} \geq u)},$$

and, using the notation of Section 3,

$$\mathbb{E}\left[Z \cdot \mathbb{1}\{X_{(1)} \geq u\}\right] = \int_u^\infty \mathbb{E}\left[Z|X_{(1)} = s\right]f_{X_{(1)}}(s) \, ds$$

$$= \int_u^\infty \int_0^1 zg(z|s)dz f_{X_{(1)}}(s) \, ds.$$

In the situation of Example 2c), we have

$$g(z|s) = \tilde{g}_\alpha(z) \frac{\tilde{L}\left(\frac{s(1+z)}{1-z}\right)}{L(s)}, \quad 0 < z < 1.$$

By assumption, for $\varepsilon > 0$ there exists $s_0 > x_m$ such that

$$\left|\tilde{L}\big(s(1+z)/(1-z)\big)/L(s) - 1\right| \leq \varepsilon \quad \forall s > s_0,$$

independently of z. It follows that

$$\left| \int_0^1 z g(z|s) dz - \int_0^1 z \tilde{g}_\alpha(z) dz \right| \le \varepsilon \int_0^1 z \tilde{g}_\alpha(z) dz \quad \forall s > s_0.$$

Thus, for $u > s_0$,

$$\left| t_X(u) - \tilde{t}_\alpha \right| = \left| \frac{\int_u^\infty \int_0^1 z g(z|s) dz f_{X_{(1)}}(s) \, ds}{P(X_{(1)} \ge u)} - \int_0^1 z \tilde{g}_\alpha(z) dz \right|$$

$$= \left| \frac{\int_u^\infty \int_0^1 z \left(g(z|s) - \tilde{g}_\alpha(z) \right) dz f_{X_{(1)}}(s) \, ds}{P(X_{(1)} \ge u)} \right|$$

$$\le \varepsilon \int_0^1 z \tilde{g}_\alpha(z) dz,$$

which proves the claim.

Appendix C. Unbiased Estimator for the Variance of a U-statistic

In this section, we discuss the estimation of $\mathrm{var}(U_n^{(l)}(d))(l = 1, 2)$ and $\mathrm{cov}(U_n^{(1)}(d), U_n^{(2)}(d))$ by estimators $\sigma^2_{U^{(l)}}$ and $\sigma^2_{U^{(1,2)}}$, respectively.

Let $U_n = 2/(n(n-1)) \sum_{i<j} h(X_i, X_j)$ be a general U-statistic of degree 2, estimating $\theta = \mathbb{E}h(X_1, X_2)$. Defining $h_1(x_1) = \mathbb{E}h(x_1, X_2)$ and

$$\zeta_0 = \theta^2, \quad \zeta_1 = \mathbb{E}\left[h_1^2(X_1) \right], \quad \zeta_2 = \mathbb{E}\left[h^2(X_1, X_2) \right],$$

the finite sample variance of U_n is given by $\mathrm{var}(U_n) = \frac{2}{n(n-1)} \{ 2(n-2)\zeta_1 + \zeta_2 - (2n-3)\zeta_0 \}$. One can estimate $\zeta_c, c = 0, 1, 2$, by

$$\hat{\zeta}_0 = \frac{1}{n^{\underline{4}}} \sum_{d(i,j,k,l)} h(X_i, X_j) h(X_k, X_l),$$

$$\hat{\zeta}_1 = \frac{1}{n^{\underline{3}}} \sum_{d(i,j,k)} h(X_i, X_j) h(X_i, X_k), \qquad \hat{\zeta}_2 = \binom{n}{2}^{-1} \sum_{i<j} h^2(X_i, X_j),$$

where $d(i_1, \ldots, i_m)$ denotes a set of distinct indices $1 \le i_1, \ldots, i_m \le n$, and $n^{\underline{m}} = n(n-1) \cdots (n-m+1)$. Then, the minimum variance unbiased estimator of $\mathrm{var}(U_n)$ is given by (Shirahata and Sakamoto 1992)

$$\hat{\sigma}_U^2 = \frac{2}{n(n-1)} \left\{ 2(n-2)\hat{\zeta}_1 + \hat{\zeta}_2 - (2n-3)\hat{\zeta}_0 \right\} = U_n^2 - \hat{\zeta}_0. \quad (C1)$$

The degree of the U-statistics in (C1) is 4, but it is possible to rewrite it as

$$\hat{\sigma}_U^2 = \frac{4C_1^2 - 2C_2^2}{n^{\underline{4}}} - \frac{4n-6}{(n-2)(n-3)} U_n^2, \quad (C2)$$

where

$$C_1^2 = \sum_{i=1}^n \left(\sum_{j \neq i} h(X_i, X_j) \right)^2, \quad C_2^2 = \sum_{i \neq j} h^2(X_i, X_j)$$

(Shirahata and Sakamoto 1992, p. 2972). To obtain a multivariate version of (C2), write hh^T and $U_n U_n^T$ instead of h^2 and U_n^2, and define

$$C_1^2 = \sum_{i=1}^n S_i S_i^T, \quad \text{where } S_i = \sum_{j \neq i} h(X_i, X_j).$$

References

Alves, I. F., de Haan, L., and Neves, C. (2009), "A Test Procedure for Detecting Super-Heavy Tails," *Journal of Statistical Planning and Inference*, 139, 213–227. [5]

Asmussen, S., and Lehtomaa, J. (2017), "Distinguishing Log-Concavity from Heavy Tails," *Risks*, 5, 10. [3]

Belzile, L. R., Dutang, C., Northrop, P. J., and Opitz, T. (2023), "A Modeler' Guide to Extreme Value Software," *Extremes*, 26, 595–638. DOI: 10.1037/a0028240. [9]

Brazauskas, V., and Serfling, R. (2003), "Favorable Estimators for Fitting Pareto Models: A Study Using Goodness-of-Fit Measures with Actual Data," *ASTIN Bulletin: The Journal of the IAA*, 33, 365–381. [2]

Chen, Y., Embrechts, P., and Wang, R. (2023), "An Unexpected Stochastic Dominance: Pareto Distributions, Catastrophes, and Risk Exchange," arXiv preprint arXiv:2208.08471. [1,2,7]

Chu, J., Dickin, O., and Nadarajah, S. (2019), "A Review of Goodness of Fit Tests for Pareto Distributions," *Journal of Computational and Applied Mathematics*, 361, 13–41. [2]

Cirillo, P., and Taleb, N. (2020), "Tail Risk of Contagious Diseases," *Nature Physics*, 16, 606–613. [2]

Cormann, U., and Reiss, R.-D. (2009), "Generalizing the Pareto to the Log-Pareto Model and Statistical Inference," *Extremes*, 12, 93–105. [5]

Das, B., and Ghosh, S. (2013), "Weak Limits for Exploratory Plots in the Analysis of Extremes," *Bernoulli*, 19, 308–343. [1,9]

——— (2016), "Detecting Tail Behavior: Mean Excess Plots with Confidence Bounds," *Extremes*, 19, 325–349. [1]

Drees, H. (2012), "Extreme Value Analysis of Actuarial Risks: Estimation and Model Validation," *AStA Advances in Statistical Analysis*, 96, 225–264. [1]

Embrechts, P., Klüppelberg, C., and Mikosch, T. (1997), *Modelling Extremal Events*, Berlin: Springer. [1,4,5]

Embrechts, P., McNeil, A. J., and Straumann, D. (2002), *Correlation and Dependence in Risk Management: Properties and Pitfalls*, pp. 176–223, Cambridge: Cambridge University Press. [1]

Fedotenkov, I. (2013), "A Bootstrap Method to Test for the Existence of Finite Moments," *Journal of Nonparametric Statistics*, 25, 315–322. [1]

——— (2020), "A Review of More than One Hundred Pareto-Tail Index Estimators," *Statistica*, 80, 245–299. [5]

Foss, S., Korshunov, D., and Zachary, S. (2013), *An Introduction to Heavy-Tailed and Subexponential Distributions* (2nd ed.), New York: Springer. [3]

Ghosh, S., and Resnick, S. (2010), "A Discussion on Mean Excess Plots," *Stochastic Processes and their Applications*, 120, 1492–1517. [1]

Gradshteyn, I., and Ryzhik, I. (2000), *Table of Integrals, Series, and Products* (6th ed.), New York: Academic Press. [9]

Hogg, R. V., and Klugman, S. A. (1984), *Loss Distributions*, New York: Wiley. [2]

Iwashita, T., and Klar, B. (2023), "A Gamma Tail Statistic and its Asymptotics," *Statistica Neerlandica*, 78, 264–280. https://onlinelibrary.wiley.com/doi/abs/10.1111/stan.12316. [3,7]

Korolyuk, V., and Borovskich, Y. (1994), *Theory of U-Statistics*, Dordrecht: Springer. [6]

Lee, A. (1990), *U-Statistics - Theory and Practice*, Boca Raton, FL: CRC Press. [6]

Lehtomaa, J. (2015), "Limiting Behaviour of Constrained Sums of Two Variables and the Principle of a Single Big Jump," *Statistics & Probability Letters*, 107, 157–163. [3]

McNeil, A. J. (1997), "Estimating the Tails of Loss Severity Distributions Using Extreme Value Theory," *ASTIN Bulletin: The Journal of the IAA*, 27, 117–137. [2]

McNeil, A. J., Frey, R., and Embrechts, P. (2015), *Quantitative Risk Management: Concepts, Techniques and Tools* (rev. ed.), Princeton: Princeton University Press. [2]

Pfaff, B., and McNeil, A. (2018), *evir: Extreme Values in R*. R package version 1.7-4. [2]

R Core Team. (2023), *R: A Language and Environment for Statistical Computing*, Vienna, Austria: R Foundation for Statistical Computing. [7]

Resnick, S. I. (1997), "Discussion of the Danish Data on Large Fire Insurance Losses," *ASTIN Bulletin: The Journal of the IAA*, 27, 139–151. [2]

Rizzo, M. L. (2009), "New Goodness-of-Fit Tests for Pareto Distributions," *ASTIN Bulletin: The Journal of the IAA*, 39, 691–715. [2]

Romano, J. P. (2004), "On Non-Parametric Testing, The Uniform Behaviour of the t-test, and Related Problems," *Scandinavian Journal of Statistics*, 31, 567–584. [1]

Samuelson, P. A. (1967), "General Proof that Diversification Pays," *The Journal of Financial and Quantitative Analysis*, 2, 1–13. [1]

Shirahata, S., and Sakamoto, Y. (1992), "Estimate of Variance of U-statistics," *Communications in Statistics - Theory and Methods*, 21, 2969–2981. [7,10]

Srivastava, M. S. (1965), "A Characterization of Pareto's Distribution and $(k+1)x^k/\theta^{(k+1)}$, (Abstract)," *The Annals of Mathematical Statistics*, 36, 361–362. [3]

Trapani, L. (2016), "Testing for (in)finite Moments," *Journal of Econometrics*, 191, 57–68. [1]