REVSTAT – Statistical Journal
Volume 0, Number 0, Month 0000, 000-000

A COMPUTATIONAL APPROACH TO CONFIDENCE INTERVALS AND TESTING FOR GENERALIZED PARETO INDEX USING THE GREENWOOD STATISTIC

Authors: Marek Arendarczyk
– Mathematical Institute, University of Wroclaw, Poland (marendar@math.uni.wroc.pl)

Tomasz J. Kozubowski
– Department of Mathematics and Statistics, University of Nevada Reno, USA (tkozubow@unr.edu)

Anna K. Panorska ✉*
– Department of Mathematics and Statistics, University of Nevada Reno, USA (ania@unr.edu)

Received: Month 0000 Revised: Month 0000 Accepted: Month 0000

Abstract:

- The generalized Pareto distributions (GPDs) play an important role in the statistics of extremes. We point various problems with the likelihood-based inference for the index parameter α of the GPDs, and develop alternative testing strategies, which do not require parameter estimation. Our test statistic is the Greenwood statistic, which probability distribution is stochastically increasing with respect to α within the GPDs. We compare the performance of our test to a test with maximum-to-sum ratio test statistic R_n. New results on the properties of the R_n are also presented, as well as recommendations for calculating the p-values and illustrative data examples.

Key-Words:

- *Coefficient of variation; extremes; generalized Pareto distribution; heavy tailed distribution; power law; Peak-over-threshold*

AMS Subject Classification:

- Primary: 62F03, 62F25, 65C60; Secondary: 60E15, 62G32, 62P12, 65C05

*Corresponding author

1. Introduction

Generalized Pareto distributions (GPDs), along with the generalized extreme value distributions (GEV), play a central role in the theory and applications of the statistics of extremes. Important (monographic) references include Gumbel (1958), Leadbetter et al. (1983), Castillo (1988), Beirlant et al. (2004), Embrechts et al. (1997), Kotz and Nadarajah (2000), Reiss and Thomas (2001), Finkenstädt and Rootzén (2003), Coles (2004), Castillo et al. (2005), de Haan and Ferreira (2006), Chavez et al. (2016), and Dey and Yan (2016). In this work, we revisit the important issue of statistical inference for the tail index α within the class of the GPD. In particular, we develop sound, simulation-enabled testing and interval estimation procedures for α with the focus on small samples.

Recall that a GP random variable X can be described through the stochastic representation

$$X \stackrel{d}{=} \frac{1}{\beta}\frac{1}{\alpha}(e^{\alpha E}-1), \quad \alpha \in \mathbb{R}, \beta > 0, \tag{1.1}$$

where E is a standard exponential random variable, β is the scale parameter, and α is the index parameter (tail index for $\alpha > 0$). The corresponding survival function (SF) of X in (1.1) is of the form

$$S(x) = \mathbb{P}(X > x) = (1+\alpha\beta x)^{-1/\alpha}. \tag{1.2}$$

For $\alpha > 0$ we get Pareto II (Lomax) distributions with power law tails of order α while for $\alpha = 0$, understood in the limiting sense, the variable X in (1.1) reduces to an exponential random variable with mean $1/\beta$, and the probability density function (PDF)

$$f(x) = \beta e^{-\beta x}, \text{ for } x \in \mathbb{R}_+ = (0,\infty). \tag{1.3}$$

Both, Lomax and exponential distributions are supported on the positive half-line $\mathbb{R}_+$. For $\alpha < 0$, GPDs are re-scaled beta distributions with compact support on the interval $(0, -1/(\alpha\beta))$, and include, for instance, the uniform distribution for $\alpha = -1$. The importance of this family comes from the Peak Over Threshold (POT) theory (see, e.g., Balkema and de Haan, 1974; Pickands, 1975), where the GPDs provide natural approximations for the *excess* (or *exceedence*) random variables $X = Y - d|Y > d$ for large classes of random variables Y, where d is a high threshold. This approximation property, coupled with their power-law tail behavior for $\alpha > 0$, make GPDs very relevant and commonly used in insurance mathematics, hydrology, climate science and other areas where the observations over high thresholds are of primary importance.

Our main contribution is a mathematically rigorous procedure for testing and constructing confidence intervals (CIs) for the index α within the GPD family, with the focus on small samples. Our methodology is based on the *Greenwood* *statistic*,

$$T_n = \frac{\sum_{i=1}^n X_i^2}{(\sum_{i=1}^n X_i)^2}, \tag{1.4}$$

where the $\{X_i\}$ are the underlying data. Since its introduction in Greenwood (1946), this statistic appeared in many different contexts and application areas, and it is closely related to several other common statistics, such as the sample coefficient of variation

$$CV_n = \frac{\sqrt{\frac{1}{n}\sum_{i=1}^n (X_i - \overline{X})^2}}{\overline{X}} = \sqrt{nT_n - 1}, \tag{1.5}$$

the reciprocal of CV_n known as the *Sharpe's ratio* in finance and insurance applications, the *self-normalized sum* $S_n = T_n^{-1/2}$, (see, e.g., Albrecher et al., 2010 and references therein), and the *student t-statistic*, $ST_n = \sqrt{(n-1)/(nT_n - 1)}$. Since tests and confidence intervals are often based on estimates of parameters, the correctness and practical execution of the estimation procedures is of primary importance for many tests, including the commonly used likelihood ratio procedure. However, for the GPDs there are serious theoretical and computational problems with the standard, likelihood-based inference for α. Indeed, in general, the maximum likelihood estimates (MLEs) for α may not exist or may not be well defined, because without artificial restrictions on the parameter space, the likelihood function is infinite along one of its boundaries. In addition, even when an MLE does exist, the question of uniqueness is still open. The problems with the likelihood function lead to practical issues with numerical calculations of the MLEs and thus implementation of standard likelihood ratio tests for α (see, e.g., Neves et al., 2006). Further, we noted several errors and inaccuracies in the estimation literature, and we include a review and discussion of the selected key papers on estimation for the GPD family in Appendix A (in the Supplementary Material). The challenges of finding the MLE of α suggest a need for test procedures that do not require estimation of α. Our test based on the Greenwood statistic is an example of such a procedure.

Further, for reliable inference on small samples we need the test statistic to be stochastically monotone (increasing or decreasing) with respect to α in order to be useful for derivation of the critical regions and construction of confidence intervals (CIs) viz. "inversion of the test" method. Again, the Greenwood statistic satisfies this requirement, as it is stochastically increasing with respect to α within the GPD family (see, Arendarczyk et al. 2021).

The challenges with ML estimation and the need for stochastic order of the test statistic with respect to the parameter of interest are our main motivation for deriving a test based on the Greenwood statistic. In addition, there is a long history of using T_n in testing for exponentiality, where tests based on T_n (or related CV_n) have been shown to be *locally most powerful* within the GPD family (see, e.g., Marohn, 2000). Further, the statistic T_n comes up naturally in estimation within the GPD and Lomax (Pareto II) families, as shown in Appendix A.

While tests based on the coefficient of variation (or other statistics equivalent to T_n) have already been used for the GPD, most of them focused on testing exponentiality and had rejection regions based on the asymptotic distributions of the test statistics (see, for example Hasofer and Wang, 1992, Gomes and van Monfort, 1986, Marohn, 2000, Reiss and Thomas, 2001). In contrast, our approach uses the exact distribution of the test statistic, obtained by straightforward simulations, and is similar in spirit to that of Chaouche and Bacro (2004, 2006) and Tajvidi (2003). Chaouche and Bacro (2004, 2006) noticed that the population value of their statistic S used for testing exponentiality is increasing in α, and that its empirical distribution shifts to the right with increasing α. Our results formalize these observations and show that the probability distribution of the S statistic computed on a random sample from GPD is stochastically increasing in α *over the entire range* of its values. Castillo et al. (2014) considered testing exponentiality within GPD using a test statistic incorporating several sample CVs computed on different sets of exceedances of varying high thresholds. The power of the proposed test was compared with those of several other tests, including one based on the sample CV, for two alternatives: absolute value of student-t distribution as well as GPD with shape parameter larger than 0 (Pareto). It appeared that tests based on the CV performed best. The paper is not very clear however about the derivation of the critical values, whether these were done by simulations or using the asymptotic distribution of the test statistic. Castillo and Padilla (2015) extended these ideas to the full GPD case with a similar test statistic, based on the asymptotic distribution of the sample CV for GPD samples. Castillo and Serra (2015) focused on the MLEs (see also Castillo and Daoudi, 2009) and offered brief remarks about testing and interval estimation for α, but no details were provided in this regard. Tajvidi (2003) considered several methods for constructing confidence intervals for the index α in the GPD family using bootstrapping, likelihood ratio (LR) test, and profiling the likelihood function, concluding that the likelihood based methods perform better than the bootstrapping in small to moderate size samples. However, in view of the considerable theoretical and computational difficulties with the MLEs of the GPD parameters, likelihood based inference may not be effective for many data sets. In summary, although substantial work was done towards testing for the GPD tail index, to date, we have not found any test for α that works well on small samples without any restrictions of the values of α.

There is also a rich body of literature on the problem of testing for a GEV domain of attraction (DoA), which is equivalent to testing for a GPD domain of attraction. The importance of this problem is well understood in the extreme value literature. In particular, testing for Gumbel DoA ($\alpha = 0$) is a common need (see, e.g., Fraga Alves and Gomes, 1996; Gomes and Alpuin, 1986). When checking domain of attraction, the estimation (or testing) is usually carried out using the excesses of the sample values over a high threshold, taken to be either a predetermined value (see, e.g., Davison and Smith, 1990) or a particular order statistic (see, e.g., Neves and Fraga Alves, 2007 and the references therein). In the latter case the size of the resulting data set available for inference may be moderate to quite small. Neves and Fraga Alves (2007) considered tests for GEV domain of attraction, particularly the Gumbel DoA. Both of their tests are related to the

Greenwood statistic. The main result of the paper is the limiting distribution of (normalized) test statistics, assuming that $\alpha < 1/4$. Recently, Schluter and Trede (2018) used one of the test statistics of Neves and Fraga Alves (2007) for testing Gumbel domain of attraction ($\alpha = 0$) against heavy tailed GPD alternatives. Our results on the properties of the statistic T_n and its version computed on the exceedences (Section 2) show that T_n can be used for the DoA tests on the small samples, which are common in the problems considering exceedences.

Next, we note the tests based on the statistic R_n, involving the ratio of the maximum and the sum of the sample values,

$$R_n = \frac{\bigvee_{i=1}^{n} X_i}{\sum_{i=1}^{n} X_i}, \tag{1.6}$$

when the underlying sample $\{X_i\}$ is from a GPD or its domain of attraction (see Neves et al., 2006). Our interest in this statistic stems from its properties that are similar to those of T_n, which makes R_n a major competitor of T_n in testing. As discussed in Bryson (1974), a test based on R_n is the most powerful for testing exponentiality against uniformity (both special cases of GPDs). The statistic R_n is also mentioned in Chaouche and Bacro (2004) in connection with testing for α within the class of GPDs, and was proposed by Neves et al. (2006) for testing maximum DoA domain of attraction. A simulation study in Neves et al. (2006) showed that their test based on R_n, along with the test of Hasofer and Wang (1992), which in essence is based on the T_n, compared very favorably in terms of power with several other tests for $\alpha = 0$ within the GPDs. However, the rejection regions of both tests were based on the asymptotic distributions of the test statistics, which may not work well for small samples. We shall revisit the test based on R_n in Section 3, where we present new results on its properties.

We selected the test based on statistic R_n to perform power comparison with our test based on the Greenwood statistic. However, instead of using the asymptotic distribution of R_n, we use essentially the same numerical procedure to compute (simulated) p-values as in our test. We selected the test based on R_n for the power comparison for the following reasons: (1) a test based on R_n was shown as more powerful than several other tests (Neves et al., 2006), (2) the R_n statistic is stochastically increasing with respect to α which makes it appropriate for small sample testing, and (3) we did not find another test applicable to one- and two-sided hypotheses for α within the entire GPD family. All other tests have (or should have) some restriction on the range of α where they are applicable.

Our paper is organized as follows. We start with Section 2, where we review the key properties of the statistic T_n. New properties of the statistic R_n are presented in Section 3. The main contribution is Section 4, containing rigorous development of tests and confidence intervals for the index α within the GPD family. This is followed by a limited power comparison between tests based on T_n and R_n in Section 5. Illustrative data examples are presented in Section 6. Proofs are collected in Section 7. Appendix T contains power tables. A review and discussion of the main works on the estimation for GPD family is presented in Appendix A. Both appendixes are available in the Supplementary Material.

2. Fundamental properties of the Greenwood statistic T_n

Let $X_1, \ldots, X_n$ be a random sample from a probability distribution supported on the non-negative real line $\mathbb{R}_+ = [0, \infty)$. It is widely recognized that within the GPD family the distribution of T_n is rather complicated (even under exponentiality) with no closed form expressions for the PDF or the CDF for general n. However, its distribution is *scale-invariant* and bounded ($\frac{1}{n} \leq T_n \leq 1$ for all n), and so all the moments of the Greenwood statistic are finite, even when the distributional moments of the underlying sequence $\{X_i\}$ do not exist. The moments of T_n under exponentiality were derived by Moran (1947), who also established an asymptotic (normal) distribution of T_n for general distributions of $\{X_i\}$ with finite first four moments, noting that the convergence to the limiting normal distribution is rather slow. In particular, under exponentiality, for large n, the Greenwood statistic is approximately normal with mean $2/n$ and variance $4/n^3$. Going beyond light-tail distributions, Albrecher et al. (2010) provided exact asymptotic distributions of T_n as n goes to infinity for distributions of $\{X_i\}$ with regularly varying tail.

2.1. Stochastic ordering of T_n

Recall that if X and Y are two random variables with respective CDFs F_X and F_Y and quantile functions (QFs) $Q_X = F_X^{-1}$ and $Q_Y = F_Y^{-1}$, X is said to be stochastically smaller than Y, denoted by $X \leq_{\text{st}} Y$, whenever $F_Y(x) \leq F_X(x)$ for each $x \in \mathbb{R}$. This is the ordinary stochastic order (dominance). On the other hand, X is smaller than Y in the star-shaped order, denoted by $X \leq_* Y$, whenever $Q_Y(u_1)/Q_X(u_1) \leq Q_Y(u_2)/Q_X(u_2)$ for all $u_1 \leq u_2$, so that the function $Q_Y(u)/Q_X(u)$ is non-decreasing in u. For more information on stochastic orders, see, e.g., Belzunce et al. (2016).

An important result established in Arendarczyk et al. (2021), which is fundamental to this work, shows that:

1. When the underlying distribution of the $\{X_i\}$ is stochastically increasing with respected to the star-shaped order $\leq_*$ then the distribution of T_n is stochastically increasing with respect to the ordinary stochastic order $\leq_{\text{st}}$.
2. The GPDs given by the SF (1.2) are star-shaped ordered with respect to the parameter α.

Therefore, the Greenwood statistic T_n is stochastically increasing with respect to the parameter α within the GPD family. As discussed in Section 4, this key property of T_n plays crucial role in setting up testing and developing confidence intervals for α within this family.

2.2. Symmetry of T_n within the GPD family

In applications, the Greenwood statistic T_n and its functions are often applied to the exceedences $X_{(j)} - X_{(k)}$, where $X_{(1)} < \cdots < X_{(n)}$ are the (ascending) order statistics based on the random sample $X_1, \ldots, X_n$ and $j = k+1, \ldots, n$. This leads to the statistic

$$T_{n,k} = \frac{\sum_{i=1}^{n-k} \left(X_{(k+i)} - X_{(k)}\right)^2}{\left[\sum_{i=1}^{n-k} \left(X_{(k+i)} - X_{(k)}\right)\right]^2}, \quad k \in \{0, 1, \ldots, n-1\}, \tag{2.1}$$

where $X_{(0)} = 0$, so that $T_{n,0}$ reduces to T_n. By POT theory, when k is relatively large, then (under appropriate scaling) these $n-k$ exceedences behave as if they were $n-k$ order statistics (based on the sample of size $n-k$) from a GPD (see, e.g., Neves et al., 2007). This crucial property plays a fundamental role in testing for the extreme domain of attraction (see, e.g., Marohn, 2000; Neves et al. 2006; Neves and Fraga-Alves, 2007 and the references therein). Since the statistic $T_{n,k}$ is scale invariant, when the underlying distribution of the sample belongs to the domain of attraction of a GPD with index α given by (1.2), the distribution of the $T_{n,k}$ in (2.1) is approximately the same as that of T_{n-k} based on a sample of size $n-k$ from the GPD itself. In other words, we would have an approximate equality in distribution

$$T_{n,k} \stackrel{d}{\sim} \frac{\sum_{i=1}^{n-k} W_i^2}{(\sum_{i=1}^{n-k} W_i)^2}, \tag{2.2}$$

where the $\{W_i\}$ follow $GP(\alpha, \beta)$. We show below that if the data are generated by a GPD in the first place, then the the distributions of $T_{n,k}$ and T_{n-k} are *exactly* the same.

Proposition 2.1. *Let $n \in \mathbb{N}$ with $n \geq 2$ and let $k \in \{0, \ldots, n-1\}$. Suppose that $X_1, \ldots, X_n$ are IID and let $T_{n,k}$ be defined by (2.1). Let $Y_1, \ldots, Y_{n-k}$ be another random sample, and let T_{n-k} be the Greenwood statistic computed on the Y_i's. Suppose that both random samples are coming from a GPD with the same index α. Then, we have $T_{n,k} \stackrel{d}{=} T_{n-k}$.*

Remark 2.1. This new result complements Proposition 5 in Arendarczyk (2021), as it shows that the latter also holds with $k = 0$.

2.3. Limiting behavior of the Greenwood statistic within the GPD family

Another key property of T_n we use in this work is its limiting behavior as the parameter α approaches $\pm\infty$ within the GPD family while the sample size n

stays fixed (Arendarczyk et al., 2021). We include the result here for convenience of the reader.

Proposition 2.2. *Suppose that $n \in \mathbb{N}$ and $X_1, \ldots, X_n$ are IID and $GP(\alpha, \beta)$ distributed. Then*

$$T_n \xrightarrow{d} 1/n \text{ as } \alpha \to -\infty \text{ and } T_n \xrightarrow{d} 1 \text{ as } \alpha \to \infty. \tag{2.3}$$

In fact, the distribution of the Greenwood statistic T_n on a GPD sample changes continuously within the interval $(1/n, 1)$ as α increases within the interval $(-\infty, \infty)$.

Remark 2.2. It can be shown that as α increases within the range $(-\infty, \infty)$ then, for each $\gamma \in (0, 1)$, the $(1 - \gamma) \times 100\%$ percentiles of the distribution of T_n within the GPD family continuously increase from their limiting values of $t_\gamma = 1/n$ at $\alpha = -\infty$ to $t_\gamma = 1$ at $\alpha = \infty$.

This monotone behavior of the quantiles of T_n is important for constructing confidence intervals and testing for α, as discussed in Section 4.

3. Fundamental properties of the statistic R_n

Since we shall use the statistic R_n in the power comparisons in Section 5, we developed new results that facilitate R_n - based testing for α within the class of GPDs. As shown below, T_n and R_n share their key properties.

3.1. Stochastic ordering of R_n

It turns out that R_n computed on a sample from GPD is stochastically increasing with respect to α. This is due to the fact that the GPDs are star-shaped ordered with respect to α and the following new result concerning R_n, whose proof can be found in Section 7.

Theorem 3.1. *Let $\{\mathcal{P}_\theta,\ \theta \in \Theta \subset \mathbb{R}\}$ be a family of absolutely continuous probability distributions on $\mathbb{R}_+$, where for each $\theta_1 \le \theta_2$ we have $X^{(\theta_1)} \le_* X^{(\theta_2)}$, with $X^{(\theta_i)} \sim \mathcal{P}_{\theta_i}$, $i = 1, 2$. Then, for each $n \ge 2$, we have*

$$R_n^{(\theta_1)} \le_{st} R_n^{(\theta_2)} \text{ whenever } \theta_1 \le \theta_2,\ \theta_1, \theta_2 \in \Theta, \tag{3.1}$$

where $R_n^{(\theta)}$ is given by (1.6) with the $\{X_i\}$ having a common distribution $\mathcal{P}_\theta$.

3.2. Symmetry of R_n within the GPD family

Let $X_{(1)} \leq \cdots \leq X_{(n)}$ be the (ascending) order statistics based on a random sample of size n from a GPD (1.2). In analogy with $T_{n,k}$, define

$$R_{n,k} = \frac{X_{(n)} - X_{(k)}}{\sum_{i=1}^{n-k}(X_{(k+i)} - X_{(k)})}, \quad k \in \{0, 1, \ldots, n-1\}, \tag{3.2}$$

where for $k = 0$ we set $X_{(0)} = 0$. This is essentially the statistic R_{n-k} evaluated on the exceedences $X_{(j)} - X_{(k)}$, with $j = k+1, \ldots, n$, which has been used in this form for testing the extreme DoA (see Neves et al., 2006). In turn, when the statistic $R_{n-k+1,1}$ is evaluated on the set of $n-k+1$ observations $X_1, \ldots, X_{n-k+1}$, then we essentially get the statistic R_{n-k} computed on the $n-k$ exceedences $X_{(j)} - X_{(1)}$, with $j = 2, \ldots, n-k+1$. The following new result, whose proof is provided in Section7, shows that the statistics $R_{n,k}$ and $R_{n-k+1,1}$ have the same distributions.

Proposition 3.1. *Let $n \in \mathbb{N}$ with $n \geq 2$ and let $k \in \{0, 1, \ldots, n-1\}$. Suppose that $X_1, \ldots, X_n$ are IID and let $R_{n,k}$ be defined by (3.2). Let $Y_1, \ldots, Y_{n-k+1}$ be another random sample, and set*

$$R_{n-k+1,1} = \frac{Y_{(n-k+1)} - Y_{(1)}}{\sum_{i=1}^{n-k}\left(Y_{(1+i)} - Y_{(1)}\right)}. \tag{3.3}$$

Then, if the two samples are coming from a GPD with the same α, we have $R_{n,k} \stackrel{d}{=} R_{n-k+1,1}$.

Remark 3.1. The above result implies that if the data are generated by a GPD, then the distributions of $R_{n,k}$ and R_{n-k} are exactly the same, similarly to $T_{n,k}$ and T_{n-k} as shown in Proposition 2.1. Thus, if n and k are large and the statistic $R_{n,k}$ is evaluated on a set of order statistics $X_{(1)} \leq \cdots \leq X_{(n)}$ of an IID sample from a distribution in the GPD (with index α) domain of attraction then standard arguments from POT theory (see, e.g., Neves et al., 2006) show that this $R_{n,k}$ behaves as if computed on a sample from the GPD (with the same α).

3.3. Limiting behavior of R_n within the GPD family

As shown below, the statistic R_n behaves very similarly to the Greenwood statistic T_n as the parameter α approaches $\pm\infty$ within the GPD family. The proof of the following new result is included in Section 7.

Proposition 3.2. *Suppose that $n \in \mathbb{N}$ and $X_1, \ldots, X_n$ are IID and $GP(\alpha, \beta)$ distributed with the SF (1.2). Then*

$$R_n \stackrel{d}{\to} 1/n \ \text{ as } \alpha \to -\infty \ \text{ and } \ R_n \stackrel{d}{\to} 1 \ \text{ as } \alpha \to \infty. \tag{3.4}$$

4. Testing and interval estimation for α within the GPD family

In this section we develop exact tests and provide a rigorous derivation of confidence intervals for the parameter α within the GPD family based on the Greenwood statistic. The CIs are constructed using the standard "inversion of the test" method. In particular, our methodology is very convenient to test for exponential distribution (GPD with $\alpha = 0$) versus Pareto II distribution (GPD with $\alpha > 0$), and has essentially the same power as the likelihood ratio test developed for this special case in Kozubowski et al. (2009).

Let $X_1, \ldots, X_n$ be a random sample from the GPD model (1.2). Since the test statistic T_n is scale-invariant, we can assume for convenience that $\beta = 1$. We start with one-sided tests, followed by two-sided tests, and conclude with procedures for constructing confidence intervals for α.

4.1. One-sided tests for α within the GPD family

Consider the problem of testing

$$H_0 : \alpha \leq \alpha_0 \quad vs. \quad H_1 : \alpha > \alpha_0, \tag{4.1}$$

where $\alpha \in \mathbb{R}$ is the (unknown) index of the GPD and $\alpha_0 \in \mathbb{R}$ is a known constant. We denote the corresponding partition of the parameter space by $\Omega_0 = \{\alpha : -\infty < \alpha \leq \alpha_0\}$ and $\Omega_1 = \{\alpha : \alpha > \alpha_0\}$.

Our objective is a test δ of size $\gamma \in (0, 1)$ for the hypotheses specified in (4.1). Note that when $\alpha_0 = 0$ this test is a test of a light-tail versus a heavy-tail (Pareto II) distribution within the GPD class. Let T_n be the test statistic for δ. Since the statistic T_n is stochastically increasing with respect to α, the values of α larger than α_0 will be indicated by relatively large values of T_n computed from the sample. Consider the following decision rule for the test δ: Reject H_0 when $T_n > c_n$, where c_n is such that $\mathbb{P}(T_n(\alpha_0) > c_n) = \gamma$, and $\mathbb{P}(T_n(\alpha) \in A)$ denotes the probability of the event $\{T_n \in A\}$ assuming the true value of the parameter is α. That is, the critical number c_n is the $(1 - \gamma) \times 100\%$ percentile of the distribution of T_n when $\alpha = \alpha_0$.

Proposition 4.1. *The test δ described above has size γ and is unbiased for the hypotheses specified in (4.1).*

Remark 4.1. The same decision rule δ can also be used for testing the hypotheses

$$H_0 : \alpha = \alpha_0 \quad vs. \quad H_1 : \alpha > \alpha_0, \tag{4.2}$$

with the test being unbiased as well.

Next, we consider the problem of testing

$$H_0 : \alpha \geq \alpha_0 \quad vs. \quad H_1 : \alpha < \alpha_0 \quad \text{or} \quad H_0 : \alpha = \alpha_0 \quad vs. \quad H_1 : \alpha < \alpha_0 \tag{4.3}$$

with decision rule to reject H_0 when $T_n < d_n$, where d_n is such that

$$\mathbb{P}(T_n(\alpha_0) < d_n) = \gamma. \tag{4.4}$$

These tests are also of size γ and unbiased.

Proposition 4.2. *The test δ for the hypotheses in (4.3) that rejects H_0 whenever $T_n < d_n$ with d_n such that $\mathbb{P}(T_n(\alpha_0) < d_n) = \gamma$, has size γ and is unbiased.*

Since the computation of the p-values is straightforward, we chose to implement our tests using the p-value method. We note that the p-value approach we describe is equivalent to the critical number approach. For the hypotheses in (4.1) and (4.2), the p-value is given by

$$\text{p-value} = \mathbb{P}(T_n(\alpha_0) > t_n), \tag{4.5}$$

where t_n is the observed value of the test statistic T_n. This can be easily seen from the stochasticity of T_n. Similarly, the p-value for the hypotheses in (4.3) is given by

$$\text{p-value} = \mathbb{P}(T_n(\alpha_0) < t_n). \tag{4.6}$$

In practice, one can approximate the p-values for these tests viz. Monte-Carlo simulation of the probabilities on the right-hand-side in (4.5) or (4.6).

4.2. Two-sided test for α

We now consider the problem of testing

$$H_0 : \alpha = \alpha_0 \quad vs. \quad H_1 : \alpha \neq \alpha_0. \tag{4.7}$$

Because of the stochastic increasing of T_n with respect to α, the critical region CR for a test δ of the hypotheses in (4.7) should consist of two sections: $CR = [1/n, C_L) \cup (C_R, 1]$. To build a test of size γ, we have a choice of the portion of γ covered by each part of the CR. In general, we can choose any $0 < r < 1$ and assign the following probabilities to the two parts of the critical region:

$$\mathbb{P}(1/n < T_n(\alpha_0) < C_L) = (1-r)\gamma \quad \text{and} \quad \mathbb{P}(C_R < T_n(\alpha_0) < 1) = r\gamma. \tag{4.8}$$

Thus, the two critical numbers are: C_L equal to the $(1-r)\gamma 100\%$ percentile, and C_R equal to the $(1-r\gamma)100\%$ percentile of the distribution of T_n under the null hypothesis. To build the test, consider $R \in (1/n, 1)$ such that $\mathbb{P}(T_n(\alpha_0) < R) = 1 - r$ and $\mathbb{P}(T_n(\alpha_0) > R) = r$. We will consider two cases of the observed value t_n of the test statistic T_n in relation to R. Again, we use the p-value approach to make decisions.

Case 1: The observed value of T_n satisfies $1/n < t_n < R$. Then, $\mathbb{P}(T_n(\alpha_0) < t_n) = (1-r)\gamma_p$ for some $\gamma_p \in (0,1)$. We claim that this γ_p is actually the p-value, so that

$$\text{p-value} = \frac{\mathbb{P}(T_n(\alpha_0) < t_n)}{1-r}. \tag{4.9}$$

Indeed, if the right-hand-side in (4.9) is less than γ, then we must have $\mathbb{P}(T_n(\alpha_0) < t_n) < \gamma(1-r)$, so that by (4.8) we have $t_n < C_L$. Since the value of the test statistic is in the rejection region, the null hypothesis is rejected. On the other hand, if the right-hand-side in (4.9) is greater than γ, so that $\mathbb{P}(T_n(\alpha_0) < t_n) > \gamma(1-r)$, then we must have $t_n > C_L$. At the same time, since $t_n < R$ and $\mathbb{P}(T_n(\alpha_0) < R) = 1 - r$, we clearly have $t_n < C_R$. Thus, the observed value of the test statistic is not in the CR, and we fail to reject H_0. Consequently, the p-value is indeed given by (4.9).

Remark 4.2. Many standard tests use $r = 0.5$ in a similar setting, and in practice we recommend that standard choice of r.

Case 2: The observed value of T_n satisfies $R < t_n < 1$. Then, using an argument similar to that used in Case 1, we obtain the following expression for the p-value:

$$\text{p-value} = \frac{\mathbb{P}(T_n(\alpha_0) > t_n)}{r}. \tag{4.10}$$

Again, one can easily approximate the above p-values viz. Monte-Carlo simulation of the probabilities on the right-hand-side in (4.9) or (4.10). Sample R-codes that return p-values described above are available from the authors upon request.

4.3. Construction of the confidence intervals

We now turn to the derivation of confidence intervals for the index α of the GPD family. We use the classical procedure of "inverting the test" to derive confidence regions, see for example Casella and Berger (2002), Section 9.2.1. We start with one-sided confidence intervals, also known as upper and lower confidence bounds. First, consider the size-γ test δ for the hypotheses in (4.1). Then, a $(1-\gamma)100\%$ confidence set for α is the set of all α_0 for which the null hypothesis is *not rejected* for a given value t_n of the test statistic T_n. The null hypothesis is not rejected when the p-value given in (4.5) is greater than γ, so that

$$\mathbb{P}(T_n(\alpha_0) > t_n) > \gamma. \tag{4.11}$$

By the stochasticity of the test statistic T_n with respect to the parameter α, the set of all α_0 that satisfy this condition is an interval of the form $(\underline{\alpha}, \infty)$, where the quantity $\underline{\alpha} = \underline{\alpha}(t_n)$ satisfies the condition

$$\mathbb{P}(T_n(\underline{\alpha}) > t_n) = \gamma. \tag{4.12}$$

Note that in view of Proposition 2.2 and the remark following it, the quantity $\underline{\alpha}(t_n)$ can always be found for any $t_n \in (1/n, 1)$ and any $\gamma \in (0, 1)$. In fact, $\underline{\alpha}(\cdot)$ is a well defined function on the interval $(1/n, 1)$ onto the real line $\mathbb{R}$. This discussion leads to the following result, which provides a lower confidence bound (LCB) for the parameter α within the GPD family.

Proposition 4.3. *Let t_n be the observed value of the test statistic T_n based on the random sample $X_1, \ldots, X_n$ from a generalized Pareto distribution $GP(\alpha_0, \beta)$. Then we have*

$$\mathbb{P}(\alpha_0 > \underline{\alpha}(T_n(\alpha_0))) = 1 - \gamma, \tag{4.13}$$

so that $(\underline{\alpha}(t_n), \infty)$ is a $(1 - \gamma) \times 100\%$ LCB for the parameter α.

Next, consider a size-γ test δ for the hypotheses in (4.3). We can obtain the upper confidence bound (UCB) using similar methods to those employed to find the LCB.

Proposition 4.4. *Let t_n be the observed value of the test statistic T_n based on the random sample $X_1, \ldots, X_n$ from a generalized Pareto distribution $GP(\alpha_0, \beta)$. Then we have*

$$\mathbb{P}(\alpha_0 < \overline{\alpha}(T_n(\alpha_0))) = 1 - \gamma, \tag{4.14}$$

so that $(-\infty, \overline{\alpha}(t_n))$ is a $(1 - \gamma) \times 100\%$ UCB for the parameter α.

Finally, we derive a two-sided $(1-\gamma)100\%$ confidence set for the parameter α by inverting the two-tail test δ for the hypotheses in (4.7). To determine the p-value, we first find the value α^* such that $\mathbb{P}(T_n(\alpha^*) < t_n) = 1-r$. Then, by the stochasticity of T_n, whenever $\alpha_0 \geq \alpha^*$ we have $\mathbb{P}(T_n(\alpha_0) < t_n) \leq 1 - r$, so that the p-value is given by (4.9). Thus, the null hypothesis is not rejected whenever

$$\mathbb{P}(T_n(\alpha_0) < t_n) > (1 - r)\gamma. \tag{4.15}$$

Since $\mathbb{P}(T_n(\alpha^*) < t_n) = 1 - r$ and the probability on the left-hand-side of (4.15) is monotonically decreasing from $1 - r$ to zero as α_0 is increasing from α^* to ∞, we can find an $\overline{\alpha} \in (\alpha^*, \infty)$ such that

$$\mathbb{P}(T_n(\overline{\alpha}) < t_n) = (1 - r)\gamma. \tag{4.16}$$

Moreover, for all $\alpha_0 \in [\alpha^*, \overline{\alpha})$ the condition (4.15) will be fulfilled. Thus, for these values of α_0 the null hypothesis in (4.7) will not be rejected and consequently the interval $[\alpha^*, \overline{\alpha})$ is part of the confidence set. Similar analysis shows that the interval $(\underline{\alpha}, \alpha^*]$, where the quantity $\underline{\alpha}$ satisfies the condition

$$\mathbb{P}(T_n(\underline{\alpha}) < t_n) = 1 - r\gamma, \tag{4.17}$$

is part of the confidence set as well. Indeed, when $\alpha_0 \leq \alpha^*$ we have $\mathbb{P}_{\alpha_0}(T_n < t_n) \geq 1 - r$, so that the p-value is given by (4.10). Thus, the null hypothesis is not rejected whenever

$$\mathbb{P}(T_n(\alpha_0) > t_n) > r\gamma. \tag{4.18}$$

Since $\mathbb{P}(T_n(\alpha^*) > t_n) = r$ and the probability on the left-hand-side of (4.18) is monotonically increasing from zero to r as α_0 increases from $-\infty$ to α^*, we can find an $\underline{\alpha} \in (-\infty, \alpha^*)$ such that $\mathbb{P}(T_n(\underline{\alpha}) > t_n) = r\gamma$, which is equivalent to (4.17). Moreover, for all $\alpha_0 \in (\underline{\alpha}, \alpha^*]$ the condition (4.18) will be fulfilled. Thus, for these values of α_0 the null hypothesis in (4.7) will not be rejected. At the same time, the above analysis shows that for the values $\alpha_0 < \underline{\alpha}$ and $\alpha_0 > \overline{\alpha}$ the null hypothesis is rejected, so these are not part of the confidence set. In summary, the confidence set obtained by inverting the test is indeed an interval. The following result summarizes this discussion.

Proposition 4.5. *Let t_n be the observed value of the test statistic T_n based on the random sample $X_1, \ldots, X_n$ from a generalized Pareto distribution $GP(\alpha_0, \beta)$. Then we have*

$$\mathbb{P}(\underline{\alpha}(T_n(\alpha_0))) < \alpha_0 < \overline{\alpha}(T_n(\alpha_0))) = 1 - \gamma, \tag{4.19}$$

so that $(\underline{\alpha}(t_n), \overline{\alpha}(t_n))$ is a $(1-\gamma) \times 100\%$ confidence interval for the parameter α.

5. Simulation experiments

To assess the performance of our testing procedures discussed in Section 4, we performed two simulation experiments. First, we did power analysis of the testing procedure based on the statistic T_n in the context of testing $H_0 : \alpha = 0$ vs. $H_1 : \alpha > 0$ (exponentiality against Lomax), which is an important practical problem of detecting a power tail in the context of the GPDs (see, e.g., Kozubowski et al., 2009). The results are reported in subsection 5.1. In addition, we compared the power functions of tests for α based on T_n and its major competitor R_n within the GPD family. The results are discussed in subsection 5.2.

5.1. Power analysis: Exponential vs. Pareto test.

Kozubowski et al. (2009) discussed testing exponentiality vs. Pareto distribution, finding the likelihood ratio (LR) approach to be superior (in terms of power) to several other common tests. To compare the performance of the test based on the statistic T_n with that of the LR test, we ran power analysis for the Greenwood test using the same values of α and sample sizes as those used in Kozubowski et al. (2009). The results are shown in Table 1 in Appendix T, which should be compared with Table 4 in Kozubowski et al. (2009). As it turns out, the power of the test based on T_n is very similar as that of the LR test across all values of the parameters, with the LR test having a slight edge. However, it should be noted that the test based on T_n is easier to implement, as it avoids calculating the MLE of the parameter α. In addition, it leads to a confidence *interval* for α, which is not guaranteed when inverting the LR test.

Remark 5.1. The expected loss of power as the sample size goes down is clearly visible in Table 1 of Appendix T. This fact is important for practical consideration of formal test results as well as other measures of fit. In particular, when deciding whether a data set has exponential or Pareto-type tails, it is a common practice to consider a large threshold and perform the analysis on the resulting *exceedances* of the data over the threshold. As one increases the threshold, the number of exceedances used in model fitting and testing decreases, and the test looses power. This may lead to not rejecting exponentiality, when in fact the data have Pareto tails. While the "best" choice of the threshold remains one of the most difficult albeit important problems in analysis of data from the GPD domain of attraction, there are many methods already available for threshold selection, including automatic procedures. Excellent reviews of the existing methods and new methodologies can be found in the following works and the references therein: Kiran and Srinivas (2021), Schneider et al. (2021), Silva Lomba and Fraga Alves (2020), and Caeiro and Gomes (2016).

5.2. Power analysis: Tests based on T_n and R_n.

Below we provide the results of a limited simulation study, of the power of the tests based on T_n and R_n. Note that the process of calculating the p-values is essentially the same for tests based on the R_n and T_n, and it was described in Section 4.

Before presenting the results of the comparison, let us note one slight advantage of the test based on R_n: the probability density function of the test statistic R_n has an explicit form when sampling from the exponential distribution (see, e.g., Qeadan et al., 2012). In our numerical experiments, we considered four different sets of one-sided hypotheses as follows:

(1) $H_0 : \alpha \leq 0$ vs. $H_1 : \alpha > 0$ (2) $H_0 : \alpha \geq 0$ vs. $H_1 : \alpha < 0$

(3) $H_0 : \alpha \leq 1$ vs. $H_1 : \alpha > 1$ (4) $H_0 : \alpha \geq 1$ vs. $H_1 : \alpha < 1$

For the selected combinations of α and $n = 5, 10, 20, 50, 100$, we generated

1000 samples from the GPD(α, n) and tested the four sets of hypotheses using the statistics T_n and R_n. We then calculated the proportion of times the null hypothesis was rejected in each case, as an approximation of the value of the power function for that α. The results are presented in Tables 2, 3, 5, and 6 in Appendix T. While the two procedures have very similar power overall, the test based on the R_n performs slightly better when testing left-sided alternatives, and the test based on T_n performs a bit better for the right-sided alternatives. We conclude that in practice one may select R_n or T_n based on the alternative hypothesis.

6. Illustrative data examples

We applied our test and built confidence intervals for α for two commonly used data sets. The purpose of this data analysis is checking if the results expected from the literature are confirmed by our test. We did not study the fit of the GPD models to the data, as that was beyond the scope of this work. The first data set contains 154 exceedances over 65 m^3/s flow threshold of river Nidd at Hunsingore Weir between 1934 and 1969. This data set was analyzed by Hosking and Wallis (1987), Davison and Smith (1990), Papastahopoulos and Tawn (2012) and Castillo and Serra (2015). We obtained the data set "nidd.thresh" from R package "evir". The second data set contains 197 exceedances above 7s of zero-crossing hourly mean periods (in seconds) of the waves measured at Bilbao bay in Spain. This data set was also used in the literature. Notably, it was first used by Castillo and Hadi (1997), and then by Luceño (2006), and Zhang and Stephens (2009). One can obtain the data "bilbao" from the R package "ercv". Since the data sets were analyzed by many researchers in the past, we only report on the results of the new analysis using the exact methods presented in this paper: test results and CIs. For computation of the p-values we used 10,000 simulations of the values of T_n from data with the distribution specified in the null hypothesis.

6.1. Analysis of the Nidd data

In search for the answer to the question whether the Nidd exceedances are Pareto or exponential, we performed our exact test on exceedances over the same threshold as those reported on p. 126 of Castillo and Serra (2015), that is 65 m^3/s, 75 m^3/s, 85 m^3/s, 95 m^3/s, 100 m^3/s, 110 m^3/s, and 120 m^3/s. We tested the null hypothesis H_0: $\alpha \leq 0$ versus a one sided alternative of Pareto, H_1: $\alpha > 0$. We also computed exact 90% and 95% confidence intervals for α for exceedances over each threshold. We report the value of the number of exceedances used, the test statistic T_n, the p-value of the test, decision (P for Pareto, E for exponential), and the CIs in Table 6.1. Our conclusion is that the exponential model for the

Nidd exceedances can not be rejected.

T ($\mathbf{m^3/s}$)	**n**	T_n	**p-value**	**D**	**90% CI for** α	**95% CI for** α
65	154	0.0165	0.0041	P	(0.0679, 0.4446)	(0.0456, 0.4916)
75	117	0.0248	0.0006	P	(0.1217, 0.6212)	(0.0941, 0.6859)
85	72	0.0349	0.0255	P	(0.0330, 0.6035)	(0.0021, 0.6999)
95	49	0.0458	0.1282	E	(-0.0634, 0.5863)	(-0.0973, 0.6927)
100	39	0.0514	0.3539	E	(-0.1886, 0.4900)	(-0.2422, 0.6007)
110	31	0.0622	0.4388	E	(-0.2658, 0.5322)	(-0.3323, 0.6533)
120	24	0.0729	0.6509	E	(-0.4920, 0.4334)	(-0.5787, 0.5682)

Table 1: The table contains results of analysis of the Nidd data. Column labeled T contains the threshold, column labeled D contains the decision.

6.2. Analysis of Bilbao waves data

The question in the literature regarding Bilbao waves data was whether the exceedances are uniform or not. This is equivalent to testing null hypothesis of uniformity, H_0: $\alpha = -1$, versus a two sided alternative of Pareto, H_1: $\alpha \neq -1$, within the GPD family. We also computed exact 90% and 95% confidence intervals for α for exceedances over each threshold. We report the value of the number of exceedances used, test statistic T_n, p-value of the test, decision (P for Pareto, U for uniform), and the CIs in Table 6.2. Our conclusion is that the uniform model for the Bilbao waves' exceedances can not be rejected.

In summary, our results confirm the conclusions in Castillo and Serra (2015)

T (s)	**n**	T_n	**p-value**	**D**	**90% CI for** α	**95% CI for** α
7	179	0.0074	0.74	U	(-1.3813, -0.8014)	(-1.4534, -0.759)
7.5	154	0.0094	0.006	P	(-0.8304, -0.4208)	(-0.8788, -0.3890)
8	106	0.0135	0.054	U	(-0.9373, -0.4188)	(-1.0060, -0.3771)
8.5	69	0.0203	0.232	U	(-1.1346, -0.4274)	(-1.2420, -0.3755)
9	41	0.0333	0.632	U	(-1.4806, -0.4232)	(-1.6507, -0.3596)
9.5	17	0.0714	0.298	U	(-4.4557, -0.7056)	(-5.4287, -0.5671)

Table 2: The table contains results of analysis of the Bilbao waves data. Column labeled T contains the threshold, column labeled D contains the decision.

that the distribution of Nidd exceedances is likely exponential (for large thresholds) and the distribution of the Bilbao waves exceedances is likely uniform.

7. Proofs

7.1. Proof of Proposition 2.1

Since for $k = 0$ the result is trivial, we shall assume that $k \geq 1$. We assume further that the two random samples are from a GPD with the same index α. By proceeding as in the proof of Proposition 5 in Arendarczyk et al. (2021), we can express the statistic $T_{n,k}$ as

$$T_{n,k} \stackrel{d}{=} \frac{\sum_{j=1}^{n-k} \left[e^{\alpha Z_j} - 1\right]^2}{\left(\sum_{j=1}^{n-k} \left[e^{\alpha Z_j} - 1\right]\right)^2}, \tag{7.1}$$

where

$$Z_j = \frac{E_{n-k}}{n-k} + \frac{E_{n-k-1}}{n-k-1} + \cdots + \frac{E_{n-k-j+1}}{n-k-j+1}, \quad j = 1, \ldots, n-k, \tag{7.2}$$

and the $\{E_i\}$ are IID standard exponential variables. A similar calculation shows that the distribution of T_{n-k} coincides with that of the right-hand-side in (7.1) with the same $\{Z_j\}$. This proves the result.

7.2. Proof of Theorem 3.1

For $i = 1, 2$, let $Q_i(\cdot)$ be the quantile function of $X^{(\theta_i)}$. By standard probability transfer theorem, we have

$$R_n^{(\theta_i)} \stackrel{d}{=} \frac{Q_i(U_{(n)})}{\sum_{k=1}^{n} Q_i(U_{(k)})}, \quad n \geq 2, \quad i = 1, 2,$$

where the $\{U_{(k)}\}$ are the (ascending) standard uniform order statistics based on a sample of size n. To establish (3.1), we need to prove that $\mathbb{P}(R_n^{(\theta_1)} > x) \leq \mathbb{P}(R_n^{(\theta_2)} > x)$ for all $x > 0$. We establish this by showing that for each choice of $0 < u_1 \leq u_2 \leq \cdots \leq u_n < 1$ we have

$$r_n^{(1)}(u_1, \ldots, u_n) \leq r_n^{(2)}(u_1, \ldots, u_n), \tag{7.3}$$

where

$$r_n^{(i)}(u_1, \ldots, u_n) = \frac{Q_i(u_n)}{\sum_{k=1}^{n} Q_i(u_k)}, \quad n \geq 2, \quad i = 1, 2. \tag{7.4}$$

We proceed by induction to establish (7.3). First, assume that $n = 2$. Straightforward calculations show that in this case the inequality (7.3) is equivalent to

$$\frac{Q_1(u_2)}{Q_1(u_1)} \leq \frac{Q_2(u_2)}{Q_2(u_1)}, \quad 0 < u_1 \leq u_2 < 1.$$

However, the above is true by the assumption that $X^{(\theta_1)} \leq_* X^{(\theta_2)}$. Next, we assume that the inequality (7.3) holds for $n \geq 2$ and show its validity for $n+1$ where

$$0 < u_1 \leq u_2 \leq \cdots \leq u_{n+1} < 1. \tag{7.5}$$

To see this, write

$$r_{n+1}^{(i)}(u_1, \ldots, u_{n+1}) = H\left(r_n^{(i)}(u_2, \ldots, u_{n+1}), w_i(u_1, \ldots, u_{n+1})\right), \tag{7.6}$$

where $H(x, y) = x/(1 + y)$, $x, y \in \mathbb{R}_+$, and

$$w_i(u_1, \ldots, u_{n+1}) = \frac{Q_i(u_1)}{\sum_{k=2}^{n+1} Q_i(u_k)}, \quad i = 1, 2. \tag{7.7}$$

Since by the induction step we have $r_n^{(1)}(u_2, \ldots, u_{n+1}) \leq r_n^{(2)}(u_2, \ldots, u_{n+1})$ and the function $H(x, y)$ is increasing in x and decreasing in y, the inequality

$$r_{n+1}^{(1)}(u_1, \ldots, u_{n+1}) \leq r_{n+1}^{(2)}(u_1, \ldots, u_{n+1}) \tag{7.8}$$

would follow by (7.6) if we could show that

$$w_2(u_1, \ldots, u_{n+1}) \leq w_1(u_1, \ldots, u_{n+1}). \tag{7.9}$$

However, it is easy to see that the inequality in (7.9) is equivalent to

$$\sum_{k=2}^{n+1} \frac{Q_1(u_k)}{Q_1(u_1)} \leq \sum_{k=2}^{n+1} \frac{Q_2(u_k)}{Q_2(u_1)},$$

which holds in view of (7.5) since we have

$$\frac{Q_1(u_k)}{Q_1(u_1)} \leq \frac{Q_2(u_k)}{Q_2(u_1)}, \quad k = 2, \ldots n+1,$$

due to $X^{(\theta_1)} \leq_* X^{(\theta_2)}$. This completes the induction argument, and the proof.

7.3. Proof of Proposition 3.1

As in the proof of Proposition 5 in Arendarczyk et al. (2021), we express $R_{n,k}$ and $R_{n-k+1,1}$ in terms of exponential spacings using the stochastic representation (1.1). We first assume that $k > 0$ and start with $R_{n,k}$. By (1.1), we have

$$R_{n,k} \stackrel{d}{=} \frac{\frac{1}{\alpha}\left(e^{\alpha E_{n:n}} - 1\right) - \frac{1}{\alpha}\left(e^{\alpha E_{k:n}} - 1\right)}{\sum_{i=1}^{n-k}\left[\frac{1}{\alpha}\left(e^{\alpha E_{k+i:n}} - 1\right) - \frac{1}{\alpha}\left(e^{\alpha E_{k:n}} - 1\right)\right]}, \tag{7.10}$$

where the $E_{1:n} \le \cdots \le E_{n:n}$ are the order statistics based on a random sample of size n from standard exponential distribution. Further simplifications produce

$$R_{n,k} \stackrel{d}{=} \frac{e^{\alpha(E_{n:n}-E_{k:n})}-1}{\sum_{j=1}^{n-k}\left[e^{\alpha(E_{k+j:n}-E_{k:n})}-1\right]}. \tag{7.11}$$

We now write $E_{k+j:n} - E_{k:n} = \sum_{i=1}^{j} D_{k+i:n}$, where $D_{i:n} = E_{i:n} - E_{i-1:n}$, $i = 2, \ldots, n$ (with $D_{1:n} = E_{1:n}$) are the associated exponential spacings. Since these are independent and exponentially distributed with parameter $n-i+1$ (see, e.g., Rényi, 1953), we can express $R_{n,k}$ as

$$R_{n,k} \stackrel{d}{=} \frac{e^{\alpha Z_{n-k}}-1}{\sum_{j=1}^{n-k}\left[e^{\alpha Z_j}-1\right]}, \tag{7.12}$$

where

$$Z_j = \frac{E_{n-k}}{n-k} + \frac{E_{n-k-1}}{n-k-1} + \cdots + \frac{E_{n-k-j+1}}{n-k-j+1}, \quad j = 1, \ldots, n-k, \tag{7.13}$$

and the $\{E_i\}$, $i = 1, \ldots, n-k$, are independent standard exponential random variables. A similar approach shows that the $R_{n-k+1,1}$ in (3.3) has the same distribution as the right-hand-side in (7.12) with the $\{Z_j\}$ given by (7.13). This proves the result for $k > 0$. The case $k = 0$ can be established along the same lines.

7.4. Proof of Proposition 3.2

Since R_n does not depend on the scale parameter, we shall assume that $\beta = 1$. We start wit the limit at $-\infty$. Since $1/n \le R_n \le X_{(n)}/(nX_{(1)})$, it is enough to show that $X_{(n)}/X_{(1)} \stackrel{d}{\to} 1$ as $\alpha \to -\infty$. By (1.1), we have

$$\frac{X_{(n)}}{X_{(1)}} \stackrel{d}{=} \frac{e^{\alpha E_{(n)}}-1}{e^{\alpha E_{(1)}}-1}, \tag{7.14}$$

where the $E_{(1)} \le \cdots \le E_{(n)}$ are the order statistics based on a random sample of size n from standard exponential distribution. Since the two exponential terms on the right-hand-side in (7.14) both converge in distribution to zero as $\alpha \to -\infty$, the right-hand-side in (7.14) converges to 1 by continuous mapping and Slutsky's theorems.

Next, we consider the limit at ∞. Straightforward algebra shows that

$$\left(1 + (n-1)\frac{X_{(n-1)}}{X_{(n)}}\right)^{-1} \le R_n \le 1.$$

Thus, it is enough to show that $X_{(n-1)}/X_{(n)} \stackrel{d}{\to} 0$ as $\alpha \to \infty$. Again, by (1.1), we have

$$\frac{X_{(n-1)}}{X_{(n)}} \stackrel{d}{=} \frac{1-e^{-\alpha E_{(n-1)}}}{e^{\alpha(E_{(n)}-E_{(n-1)})}-e^{-\alpha E_{(n-1)}}}, \tag{7.15}$$

where the $\{E_{(i)}\}$ are as before. It is easy to see that, as $\alpha \to \infty$, the exponential term $e^{-\alpha E_{(n-1)}}$ in the expression above converges to zero while the term $e^{\alpha(E_{(n)}-E_{(n-1)})}$ converges to ∞. Consequently, the expression in (7.15) converges to 0 as desired.

7.5. Proof of Proposition 4.1

To show the test δ has size γ, note that

$$size(\delta) = \sup_{\alpha \le \alpha_0} \mathbb{P}(T_n(\alpha) > c_n) \le \mathbb{P}(T_n(\alpha_0) > c_n) = \gamma,$$

since T_n is stochastically increasing in α. Thus, the test δ has size γ. Next, we show that δ is also unbiased. Let $\pi(\alpha)$ be the power function of δ. We shall show that the power function is at least equal to the size for all $\alpha \in \Omega_1$. Indeed, for any $\alpha \in \Omega_1$ we have

$$\pi(\alpha) = \mathbb{P}(T_n(\alpha) > c_n) \ge \mathbb{P}(T_n(\alpha_0) > c_n) = \gamma,$$

since T_n is stochastically increasing in α. This shows that δ is unbiased, as desired.

7.6. Proof of Proposition 4.3

The probability in (4.13) concerns all the values t_n of the statistic T_n for which we have $\alpha_0 > \underline{\alpha}(t_n)$. By the definition of the function $\underline{\alpha}$ and the stochasticity of T_n, this is equivalent to the condition $S(t_n|\alpha_0) > \gamma$, where $S(\cdot|\alpha_0)$ is the SF of T_n when the true value of the parameter is α_0. Equivalently, the event $\alpha_0 > \underline{\alpha}(T_n)$ in (4.13) can be stated as $F(T_n|\alpha_0) \le 1-\gamma$, where $F(\cdot|\alpha_0)$ is the CDF of T_n when the true value of the parameter is α_0. However, the quantity $F(T_n|\alpha_0)$ has standard uniform distribution, so that

$$\mathbb{P}_{\alpha_0}(\alpha_0 > \underline{\alpha}(T_n)) = \mathbb{P}_{\alpha_0}(F(T_n|\alpha_0) \le 1-\gamma) = 1-\gamma,$$

as desired.

7.7. Proof of Proposition 4.5

The probability in (4.19) concerns all the values t_n of the statistic T_n for which we have $\underline{\alpha}(t_n) < \alpha_0 < \overline{\alpha}(t_n)$. By the definition of the quantities $\underline{\alpha}$ and $\overline{\alpha}$ and the stochasticity of T_n, this is equivalent to the condition $\gamma - r\gamma < F(t_n|\alpha_0) < 1 - r\gamma$, where $F(\cdot|\alpha_0)$ is the CDF of T_n when the true value of the parameter is α_0. Equivalently, the event $\underline{\alpha}(T_n)) < \alpha_0 < \overline{\alpha}(T_n)$ in (4.19) can be stated as $\gamma - r\gamma < F(T_n|\alpha_0) < 1 - r\gamma$. Since the quantity $F(T_n|\alpha_0)$ is standard uniform, we have

$$\mathbb{P}_{\alpha_0}(\underline{\alpha}(T_n) < \alpha_0 < \overline{\alpha}(T_n)) = \mathbb{P}_{\alpha_0}(\gamma - r\gamma < F(T_n|\alpha_0) < 1 - r\gamma) = 1 - \gamma,$$

as desired.

Acknowledgments We thank the editors and the reviewers for their comments and suggestions that helped us improve the manuscript. We also extend hearty thanks to Professor Joan del Castillo and Professor Ioannis Papasthopoulos for their kind help in obtaining the data sets. This project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no 318984 - RARE.

REFERENCES

[1] Albrecher, H., Ladoucette, S.A. and Teugels, J.L. (2010). Asymptotics of the sample coefficient of variation and the sample dispersion, *J. Statist. Plann. Inference* **140**(2), 358–368.

[2] Arendarczyk, M., Kozubowski, T.J. and Panorska, A.K. (2021). The Greenwood statistic, stochastic dominance, clustering and heavy tails, *Scand. J. Stat.* (in press).

[3] Balkema, A.A. and de Haan, L. (1974). Residual life time at high age, *Ann. Probab.* **2**, 792–804.

[4] Beirlant, J., Goegebeur, Y., Segers, J. and Teugels, J.L. (2004). *Statistics of Extremes: Theory and Applications*, Wiley, Hoboken.

[5] Belzunce, F., Martinez-Riquelme, C. and Mulero, J. (2016). *An Introduction to Stochastic Orders*, Academic Press, Amsterdam.

[6] Bryson, M.C. (1974). Heavy-tailed distributions: Properties and tests, *Technometrics* **16**(1), 61–68.

[7] Caeiro, F., and Gomes, M. I. (2016). Threshold selection in extreme value analysis. In D. K. Dey, & J. Yan (Eds.), Extreme Value Modeling And Risk Analysis: Methods And Applications (pp. 69 – 82). Chapman and Hall/CRC Press, Boca Raton.

[8] Casella, G., and Berger, R.L. (2002). *Statistical Inference*, 2nd ed., Duxbury Press/Thomson Learning, Pacific Grove.

[9] Castillo, E. (1988). *Extreme Value Theory in Engineering*, Academic Press, San Diego.

[10] Castillo, E., Hadi, A.S., Balakrishnan, N. and Sarabia, J.M. (2005). *Extreme Value and Related Models with Applications in Engineering and Science*, Wiley, Hoboken.

[11] Castillo, E. and Hadi, A.S. (1997). Fitting the generalized Pareto distribution to data, *J. Amer. Statist. Assoc.* **92**(440), 1609–1620.

[12] Castillo, J. and Daoudi, J. (2009). Estimation of the generalized Pareto distribution, *Statist. Probab. Lett.* **79**, 684–688.

[13] Castillo, J., Daoudi, J. and Lockhart, R. (2014). Methods to distinguish polynomial and exponential tails, *Scan. J. Statist.* **41**, 382–393.

[14] Castillo, J. and Padilla, M. (2015). Modeling extreme values by the residual coefficient of variation, arXiv: 1510.00179v1.

[15] Castillo, J. and Serra, I. (2015). Likelihood inference for generalized Pareto distribution, *Comput. Statist. Data Anal.* **83**, 116–128.

[16] Chaouche, A. and Bacro, J.-N. (2004). A statistical test procedure for the shape parameter of a generalized Pareto distribution, *Comput. Statist. Data Anal.* **45**, 787–803.

[17] Chaouche, A. and Bacro, J.-N. (2006). Statistical inference for the generalized Pareto distribution: Maximum likelihood revisited, *Comm. Statist. Theory Methods* **35**(5), 785–802.

[18] Chavez, M., Ghil, M. and Fucugauchi, U. (2016). *Extreme Events: Observations, Modeling, and Economics*, AGU/Wiley, Hoboken.

[19] Coles, S. (2004). *An introduction to Statistical Modeling of Extreme Values*, Springer, London.

[20] Davison, A.C. and and Smith, R.L. (1990). Models for exceedances over high thresholds, *J. Roy. Statist. Soc. Ser. B* **52**(3), 393–442.

[21] de Haan, J. and Ferreira, A. (2006). *Extreme Value Theory: An Introduction*, Springer, New York.

[22] Dey, D.K. and Yan, J. (2016). *Extreme Value Modeling and Risk Analysis: Methods and Applications*, Chapman and Hall/CRC Press, Boca Raton.

[23] Embrechts, P., Klüppelberg, C. and Mikosch, T. (1997). *Modelling Extremal Events for Insurance and Finance*, Springer-Verlag, Berlin.

[24] Fraga Alves, M.I. and Gomes, M.I. (1996). Statistical choice of extreme value domains of attraction - a comparative analysis, *Comm. Statist. Theory Methods* **25**(4), 789–811.

[25] Finkenstädt, B. and Rootzén, H. (edit) (2003). *Extreme Values in Finance, Telecommunications, and the Environment*, Chapman and Hall/CRC Press, Boca Raton.

[26] Gomes, M.I. and Alpuin, M.T. (1986). Inference in a multivariate generalized extreme value model: Asymptotic properties of two test statistics, *Scand. J. Stat.* **13**(4), 291–300.

[27] Gomes, M.I. and van Montfort, M.A.J. (1986). Exponentiality versus generalized Pareto, quick tests, in *Proceedings of the Third International Conference on Statistical Climatology*, pp. 185–195.

[28] Greenwood, M. (1946). The statistical study of infectious diseases, *J. Roy. Statist. Soc.* **109**(2), 85–110.

[29] Gumbel, E.J. (1958). *Statistics of Extremes*, Columbia University Press, New York.

[30] Hasofer, A.M. and Wang, Z. (1992). A test for extreme value domain of attraction, *J. Amer. Statist. Assoc.* **87**, 171–177.

[31] Hosking, J.R.M. and Wallis, J.R. (1987). Parameter and quantile estimation for the generalized Pareto Distribution, *Technometrics*, **29**(3), 339–349.

[32] Kiran, K. G., and Srinivas, V. V. (2021). A Mahalanobis distance-based automatic threshold selection method for Peaks Over Threshold model. *Water Resour. Research*, **57**, e2020WR027534.

[33] Kotz, S. and Nadarajah, S. (2000). *Extreme Value Distributions: Theory and Applications*, Imperial College Press, London.

[34] Kozubowski, T.J., Panorska, A.K., Qeadan, F., Gershunov, A. and Rominger, D. (2009). Testing exponentiality versus Pareto distribution via likelihood ratio, *Comm. Statist. Sim. Comput.* **38**(1), 118–139.

[35] Leadbetter, M.R., Lindgren, G. and Rootzén, H. (1983). *Extremes and Related Properties of Random Sequences and Processes*, Springer-Verlag, Berlin.

[36] Luceño, A. (2006). Fitting the generalized Pareto distribution to data using maximum goodness-of-fit estimators, *Comput. Statist. Data Anal.* **51**(2), 904–917.

[37] Marohn, F. (2000). Testing extreme value models, *Extremes* **3**(4), 363–384.

[38] Moran, P.A.P. (1947). The random division of an interval, *J. Roy. Statist. Soc. Suppl.* **9**(1), 92–98.

[39] Neves, C. and Fraga Alves, M.I. (2007). Semi-parametric approach to the Hasofer-Wang and Greenwood statistics in extremes, *Test* **16**, 297–313.

[40] Neves, C., Picek, J. and Fraga Alves, M.I. (2006). The contribution of the maximum to the sum of excesses for testing max-domains of attraction, *J. Statist. Plann. Inference* **136**, 1281–1301.

[41] Papastahopoulos, I. and Tawn, J.A. (2012). Extended generalized Pareto models for tail estimation, *J. Statist. Plann. Inference* **143**, 131–143.

[42] Pickands III, J. (1975). Statistical inference using extreme order statistics, *Ann. Statist.* **3**, 119–131.

[43] Qeadan, F., Kozubowski, T.J. and Panorska, A.K. (2012). The joint distribution of the sum and the maximum of n i.i.d. exponential random variables, *Comm. Statist. Theory Methods* **41**(3), 544–569.

[44] Reiss, R.-D. and Thomas, M. (2001). *Statistical Analysis of Extreme Values with Applications to Insurance, Finance, Hydrology, and Other Fields*, 2nd Ed., Birkhäuser, Basel.

[45] Rényi, A. (1953). On the theory of order statistics, *Acta Math. Acad. Sci. Hungar.* **4**, 191–232.

[46] Schluter, C. and Trede, M. (2018). Size distributions reconsidered. *Econometric Rev.*, **38**(4), 1 – 16.

[47] Schneider, L.F., Krajina, A. and Krivobokova, T. (2021). Threshold selection in univariate extreme value analysis, *Extremes* **24**, 881 – 913.

[48] Silva Lomba, J. and Fraga Alves, M.I. (2020). L-moments for automatic threshold selection in extreme value analysis. *Stoch. Environ. Res. Risk Assess.*, **34**, 465 – 491.

[49] Tajvidi, N. (2003). Confidence intervals and accuracy estimation for heavy-tailed generalized Pareto distributions, *Extremes* **6**, 111–123.

[50] Zhang, J. and Stephens, M.A. (2009). A new and efficient estimation method for the generalized Pareto distribution, *Technometrics* **51**, 316–325.