مجلة التجارة والتمويل

https://caf.journals.ekb.eg/

كلية التجارة – جامعة طنطا

العدد : الرابع

ديسمبر 2023

(الجزء الثانى)

Parameter Estimation of the Generalized Pareto Distribution Using Robust Location and Scale Measures for Order Statistics

Mohammad Ibrahim Ahmmad Soliman Gaafar
Department of Statistics, Faculty of Commerce, Alexandria University, Alexandria, Egypt.
E-mail: Mohammad.solayman@alexu.edu.eg; aswanybakkar@yahoo.com

Parameter Estimation of the Generalized Pareto Distribution Using Robust Location and Scale Measures for Order Statistics

Abstract

In this paper, a new robust weighted two-step method is proposed for estimating the parameters of the generalized Pareto distribution (GPD). Through two simulation studies, the empirical performance of the proposed estimator is evaluated and compared with some competitor existing estimators. The methods were applied to a real-life data. Simulation and application results reflect the stability of the empirical performance of the proposed estimator and also show that, in most cases, the proposed estimator outperforms all other competitor estimators under investigation or at least equally likely with them.

Keywords: Generalized Pareto Distribution, Extreme Value Theory, The Peaks over Threshold Approach, Uniform and Exponential Order Statistics, The Inverse Regularized Incomplete Beta Function.

1. Introduction

When analyzing data, the main focus is not always on some population central characteristics (e.g., the average amount of air pollutants, the average amount of water flow, the median income, etc.) but it may also be on “extremes” (the minimum or maximum) of phenomena under study. For example, in designing a dam, the average flood helps in determining the amount of water to be stored, but engineers may also be interested in the maximum flood, the maximum earthquake intensity, or the minimum strength of the concrete used in building the dam seeking high degree of work perfection. Also, in designing offshore platforms, breakwaters, dikes, or any other harbor works, the wave height and the probability distribution of the highest waves is the main factor to be considered. Dry spells, earthquakes, snowfalls, hurricanes, floods, …etc. are all some examples of rare catastrophic events to be avoided or at least to be prepared for. Extreme value theory (EVT) is a statistical tool to be used in modelling extreme values and the risk of rare events. Applications of extreme value modeling involve the field of meteorology, hydrology, economics, material science, insurance, finance, and survival analysis.

There are two main approaches for modeling extreme values, the block maxima (BM) and the peaks over threshold (POT) approaches. The block maxima (BM) approach is based on the generalized extreme value (GEV) distribution through the Fisher-Tippett-Gnedenko theorem (Fisher and Tippett (1928) and Gnedenko (1943)). The peaks over threshold (POT) approach is based on the generalized Pareto (GP) distribution through the Pickands-Balkema-de Haan theorem (Balkema and de Haan (1974) and Pickands (1975)). In this paper, the POT approach is considered.

The statistical literature is rich in methods of estimating the parameters of the (GPD). Each method has its own advantages and disadvantages. Some of the most frequently used methods of estimating the parameters of the (GPD) are: the traditional method of moments applied by Hosking and Wallis (1987), the maximum likelihood method computed by Grimshaw (1993), the method of probability weighted moments proposed by Greenwood et al. (1979), the elemental percentile method proposed by Castillo and Hadi (1997), the method of L-moments introduced by Hosking (1990), the method of higher order L-moments introduced by Wang (1997), the method of generalized probability weighted moments introduced by Rasmussen (2001). Some likelihood-based methods of estimation were introduced and modified by Zhang (2007), Castillo and Serra (2015), Zhang and Stephens (2009), and Zhang (2010). Some least squares methods of estimation were introduced and modified by Song and Song (2012), Park and Kim (2016), Kang and Song (2017), and Zhao et al. (2019).

This paper is organized as follows. Section 2 introduces the extreme value theory. The (GPD) and some of its main properties are given in subsection 2.1, while the rationale behind the (POT) approach is explained in subsection 2.2. Section 3 is devoted to methods of estimation, competitor existing methods in subsection 3.1, while the new proposed estimation method in subsection 3.2. The two simulation studies are described in section 4. The parameters of the (GPD) are estimated under two settings: using all sample observations (Non-POT) in subsection 4.1 and in subsection 4.2 using only observations above a certain specific threshold (u) which is referred to as (POT). In section 5, the proposed and competitor methods of estimation are applied to a real-world data set, the Bilbao waves data, used in Castillo and Hadi (1997). Section 6 reports the main results and gives concise conclusions of the study.

2. Extreme Value Theory

2.1. Generalized Pareto distribution (GPD)

As shown in Coles (2001), the generalized Pareto distribution (GPD) with location parameter a ($a \in \mathbb{R}$), scale parameter λ ($\lambda > 0$), and shape parameter ν ($\nu \in \mathbb{R}$) is denoted by $GPD(a, \lambda, \nu)$. The cumulative distribution function (cdf) of the $GPD(a, \lambda, \nu)$ is defined as:

$$G(x; a, \lambda, \nu) = \begin{cases} 1 - \left(1 + \frac{\nu(x-a)}{\lambda}\right)^{\frac{-1}{\nu}}, & \left(1 + \frac{\nu(x-a)}{\lambda}\right) \geq 0, \lambda > 0, \nu \neq 0 \\ 1 - exp\left(-\frac{x-a}{\lambda}\right), & x \geq a, \lambda > 0, \nu = 0 \end{cases} \tag{1}$$

The probability density function (pdf) of the GPD is defined as:

$$g(x;a,\lambda,\nu)=\begin{cases}\frac{1}{\lambda}\left[1+\frac{\nu(x-a)}{\lambda}\right]^{\left(\frac{-1}{\nu}-1\right)}, & \left(1+\frac{\nu(x-a)}{\lambda}\right)\geq 0, \lambda>0, \nu\neq 0\\ \frac{1}{\lambda}exp\left(-\frac{x-a}{\lambda}\right), & x\geq a, \lambda>0, \nu=0\end{cases} \quad (2)$$

The quantile function is defined as:

$$G^{-1}(p;a,\lambda,\nu) = \begin{cases}a+\frac{\lambda}{\nu}[(1-p)^{-\nu}-1], & \nu\neq 0\\ a+\lambda(-Ln(1-p)), & \nu=0\end{cases} \quad (3)$$

- The mean, variance, skewness, and kurtosis are:
- $E(X) = a + \frac{\lambda}{1-\nu}$, $\quad \nu < 1$
- $Var\ (X) = \frac{\lambda^2}{(1-\nu)^2(1-2\nu)}$, $\quad \nu < \frac{1}{2}$
- $Skewness\ (X) = \frac{2(1+\nu)\sqrt{1-2\nu}}{1-3\nu}$, $\quad \nu < \frac{1}{3}$
- $Kurtosis\ (X) = \frac{3(1-2\nu)(3+\nu+2\nu^2)}{(1-3\nu)(1-4\nu)}$, $\quad \nu < \frac{1}{4}$

- When $a=0$, $GPD(a,\lambda,\nu)$ reduces to 2-parameter $GPD(\lambda,\nu)\equiv GPD(0,\lambda,\nu)$.
- When $\nu=0$, the GPD reduces to the exponential distribution with mean λ.
- When $\nu=-1$, the GPD reduces to the continuous uniform distribution $U(0,\lambda)$.
- The GPD is short-tailed for $\nu<0$, medium-tailed for $\nu=0$, and heavy-tailed for $\nu>0$.
- If the random variable X is $GPD(\lambda,\nu)$ then $(X-u\,|\,X>u)$ is $GPD(\lambda+u\nu,\nu)$, Coles (2001). That is to say, the GPD is stable with respect to truncations from the left. This property is called the threshold stability of GPD.

2.2. Peaks over threshold (POT)

Let F be the distribution function associated with a random variable X. The distribution function of the exceedances (or excesses) over a threshold (u) is defined as:

$$F_u(x) = P(X - u \leq x | X > u) = \frac{F(u + x) - F(u)}{1 - F(u)}, \qquad 0 \leq x \leq x_F - u$$

Wherex_F is the finite or infinite right endpoint of the distribution F.

According to the Fisher-Tippett-Gnedenko theorem (Fisher and Tippett (1928) and Gnedenko (1943)), the distribution of the excesses over a sufficiently high threshold (u) can be approximated by the generalized Pareto distribution (GPD) with positive scale parameter $\lambda(u)$, i.e.,

$$\lim_{u \to x_F} \sup_{0 \leq x \leq x_F - u} |F_u(x) - G(x; \lambda(u), \nu)| = 0,$$

if and only if the distribution function F belongs to a certain class of distributions in the maximum domain of attraction of the generalized extreme value distribution (GEV), i.e., $F \in MDA(H_\nu)$, with distribution function

$$H_\nu(x) = \begin{cases} \exp\left[-(1 + \nu x)^{\frac{-1}{\nu}}\right], & (1 + \nu x) \geq 0, \quad \nu \neq 0 \\ \exp[-\exp(-x)], & x \geq 0, \quad \nu = 0 \end{cases} \qquad (4)$$

The above (GEV) distribution involves three types of distributions: the heavy-tailed Fréchet distributions when $\nu > 0$, the medium-tailed Gumbel distributions when $\nu = 0$, and the short-tailed Weibull distributions when $\nu < 0$. The Pareto, Burr, Cauchy, log gamma, and t-distributions belong to the heavy-tailed Fréchet distributions. The normal, exponential, gamma, and lognormal distributions belong to the medium-tailed Gumbel distributions. The uniform and beta distributions belong to the short-tailed Weibull distributions.

3. Methods of Estimation

In this section, a review of four of the best existing least squares methods of estimation are considered, then the proposed method of estimation is introduced.

3.1. Some Existing Methods of Estimation

Four existing estimation methods are to be empirically evaluated and compared with the proposed one. Each method is based on a two-step minimization procedure, where the second minimization step takes the estimates of the first minimization step as initial values. The first method was proposed by Song and Song (2012) and then was corrected by Park and Kim (2016) and will be referred to, in this paper, as ($\boldsymbol{SSPK}$) method. The second method was proposed by Park and Kim (2016) and will be referred to as ($\boldsymbol{PK}$) method. The third and the fourth methods were proposed by Zhao et al. (2019) and will be referred to as ($\boldsymbol{ZZCZ\_1}$) and ($\boldsymbol{ZZCZ\_2}$) methods.

3.1.1. The ($\boldsymbol{SSPK}$) method

Suppose that $\boldsymbol{x_1, x_2, x_3, \cdots\cdots\cdots, x_n}$ is a random sample of size ($\boldsymbol{n}$) and let the corresponding sample order statistics be denoted as $\boldsymbol{x_{(1)}} \le \boldsymbol{x_{(2)}} \le \boldsymbol{x_{(3)}} \le \cdots\cdots\cdots \le \boldsymbol{x_{(n)}}$. According to the Pickands-Balkema-de Haan theorem (Balkema and de Haan (1974) and Pickands (1975)), under certain conditions with sufficiently large threshold ($\boldsymbol{u}$), the distribution of exceedances $(\boldsymbol{X_i - u \le x | X_i > u, i = n_u + 1, \cdots, n})$ can be approximated by the GPD, where $\boldsymbol{n\_u < n}$ is the number of observations that are less than the threshold ($\boldsymbol{u}$).

According to Song and Song (2012) and Park and Kim (2016), the parameters (λ, ν) of the GPD can be estimated through the following two steps:

(First step)

$$(\hat{\boldsymbol{\nu}}_1, \hat{\boldsymbol{\lambda}}_1) = \underset{(\nu,\lambda)}{\boldsymbol{arg\,min}} \sum_{i=n_u+1}^{n} \left[\boldsymbol{log}\left(1 - F_n(x_{(i)})\right) - \boldsymbol{log}\left(1 - F(x_{(i)})\right)\right]^2$$

$$(\hat{\boldsymbol{\nu}}_1, \hat{\boldsymbol{\lambda}}_1) = \underset{(\nu,\lambda)}{\boldsymbol{arg\,min}} \sum_{i=n_u+1}^{n} \left[\boldsymbol{log}\left(1 - \frac{i}{n}\right) - \boldsymbol{log}\{1 - [1 - F_n(u)]G_{\nu,\lambda}(x_{(i)} - u) - F_n(u)\}\right]^2$$

$$(\hat{\boldsymbol{\nu}}_1, \hat{\boldsymbol{\lambda}}_1) = \underset{(\nu,\lambda)}{\boldsymbol{arg\,min}} \sum_{i=n_u+1}^{n} \left[\boldsymbol{log}\left[\frac{1 - (i/n)}{1 - (n_u/n)}\right] - \boldsymbol{log}[1 - G_{\nu,\lambda}(x_{(i)} - u)]\right]^2$$

$$(\hat{\boldsymbol{\nu}}_1, \hat{\boldsymbol{\lambda}}_1) = \underset{(\nu,\lambda)}{\boldsymbol{arg\,min}} \sum_{i=n_u+1}^{n} \left[\boldsymbol{log}\left[\frac{n - i}{n - n_u}\right] - \boldsymbol{log}[1 - G_{\nu,\lambda}(x_{(i)} - u)]\right]^2 \qquad (5)$$

(Second step)

$$(\hat{\boldsymbol{\nu}}_2, \hat{\boldsymbol{\lambda}}_2) = \underset{(\nu,\lambda)}{\boldsymbol{arg\,min}} \sum_{i=n_u+1}^{n} \left[F_n(x_{(i)}) - F(x_{(i)})\right]^2$$

$$(\hat{v}_2,\hat{\lambda}_2)=\arg\min_{(v,\lambda)}\sum_{i=n_u+1}^{n}\left[F_n(x_{(i)})-[1-F_n(u)]G_{v,\lambda}(x_{(i)}-u)-F_n(u)\right]^2$$

$$(\hat{v}_2,\hat{\lambda}_2)=\arg\min_{(v,\lambda)}[1-F_n(u)]^2\sum_{i=n_u+1}^{n}\left[\frac{F_n(x_{(i)})-F_n(u)}{1-F_n(u)}-G_{v,\lambda}(x_{(i)}-u)\right]^2$$

$$(\hat{v}_2,\hat{\lambda}_2)=\arg\min_{(v,\lambda)}\sum_{i=n_u+1}^{n}\left[\frac{F_n(x_{(i)})-F_n(u)}{1-F_n(u)}-G_{v,\lambda}(x_{(i)}-u)\right]^2$$

$$(\hat{v}_2,\hat{\lambda}_2)=\arg\min_{(v,\lambda)}\sum_{i=n_u+1}^{n}\left[\frac{i-n_u}{n-n_u}-G_{v,\lambda}(x_{(i)}-u)\right]^2 \tag{6}$$

3.1.2. The (PK) method

According to Park and Kim (2016), the parameters (λ,v) of the GPD can be estimated through the following two steps:

(First step)

$$(\hat{v}_1,\hat{\lambda}_1)=\arg\min_{(v,\lambda)}\sum_{i=n_u+1}^{n}\left[\log\left(1-F_n(x_{(i)})\right)-\log\left(1-F(x_{(i)})\right)\right]^2$$

$$(\hat{v}_1,\hat{\lambda}_1)=\arg\min_{(v,\lambda)}\sum_{i=n_u+1}^{n}\left[\log\left(1-\frac{i}{n}\right)-\log\{1-[1-F_n(u)]G_{v,\lambda}(x_{(i)}-u)-F_n(u)\}\right]^2$$

$$(\hat{v}_1,\hat{\lambda}_1)=\arg\min_{(v,\lambda)}\sum_{i=n_u+1}^{n}\left[\log\left[\frac{1-(i/n)}{1-(n_u/n)}\right]-\log[1-G_{v,\lambda}(x_{(i)}-u)]\right]^2$$

$$(\hat{v}_1,\hat{\lambda}_1)=\arg\min_{(v,\lambda)}\sum_{i=n_u+1}^{n}\left[\log\left[\frac{n-i}{n-n_u}\right]-\log[1-G_{v,\lambda}(x_{(i)}-u)]\right]^2 \tag{7}$$

(Second step)

$$(\hat{v}_2^*,\hat{\lambda}_2^*)=\arg\min_{(v,\lambda)}\sum_{i=n_u+1}^{n}w_{(i)}\times[F_n(x_{(i)})-F(x_{(i)})]^2$$

$$(\hat{v}_2^*,\hat{\lambda}_2^*)=\arg\min_{(v,\lambda)}\sum_{i=n_u+1}^{n}\left[\frac{(n+1)^2(n+2)}{i(n-i+1)}\right]\times\left[\frac{i-n_u}{n-n_u}-G_{v,\lambda}(x_{(i)}-u)\right]^2 \tag{8}$$

One advantage of the PK method of estimation over the $SSPK$ method is that it estimates the extreme quantiles in a more stable manner as larger weights are given for $F(x_{(i)})$ values as $x_{(i)}$ moves towards the tail side.

3.1.3. The ($\boldsymbol{ZZCZ\_1}$ and $\boldsymbol{ZZCZ\_2}$) methods

Zhao et al. (2019) proposed two methods for estimating the parameters of the GPD. According to the first method, the parameters (λ, ν) of the GPD can be estimated through the following two steps:

(First step)

$$(\hat{\nu}_1, \hat{\lambda}_1) = \underset{(\nu,\lambda)}{arg\,min} \sum_{i=n_u+1}^{n} \left[-log\left(1 - F(x_{(i)})\right) - E\left\{-log\left(1 - F(X_{(i)})\right)\right\}\right]^2$$

$$(\hat{\nu}_1, \hat{\lambda}_1) = \underset{(\nu,\lambda)}{arg\,min} \sum_{i=n_u+1}^{n} \left[\sum_{j=1}^{i}\left(\frac{1}{n-j+1}\right) + log\{1 - [1 - F_n(u)]G_{\nu,\lambda}(x_{(i)} - u) - F_n(u)\}\right]^2$$

$$(\hat{\nu}_1, \hat{\lambda}_1) = \underset{(\nu,\lambda)}{arg\,min} \sum_{i=n_u+1}^{n} \left[\sum_{j=1}^{i}\left(\frac{1}{n-j+1}\right) + log[1 - F_n(u)] + log[1 - G_{\nu,\lambda}(x_{(i)} - u)]\right]^2 \quad (9)$$

(Second step)

$$(\hat{\nu}_2^*, \hat{\lambda}_2^*) = \underset{(\nu,\lambda)}{arg\,min} \sum_{i=n_u+1}^{n} w_{(i)} \times [F(x_{(i)}) - E\{F(X_{(i)})\}]^2$$

$$(\hat{\nu}_2^*, \hat{\lambda}_2^*) = \underset{(\nu,\lambda)}{arg\,min} \sum_{i=n_u+1}^{n} w_{(i)} \times \left[[1 - F_n(u)]G_{\nu,\lambda}(x_{(i)} - u) + F_n(u) - \frac{i}{n+1}\right]^2$$

$$(\hat{\nu}_2^*, \hat{\lambda}_2^*) = \underset{(\nu,\lambda)}{arg\,min}\ [1 - F_n(u)]^2 \sum_{i=n_u+1}^{n} w_{(i)} \times \left[\frac{(i/(n+1)) - F_n(u)}{1 - F_n(u)} - G_{\nu,\lambda}(x_{(i)} - u)\right]^2$$

$$(\hat{\nu}_2^*, \hat{\lambda}_2^*) = \underset{(\nu,\lambda)}{arg\,min} \sum_{i=n_u+1}^{n} \left[\frac{(n+1)^2(n+2)}{i(n-i+1)}\right] \times \left[\frac{(i/(n+1)) - F_n(u)}{1 - F_n(u)} - G_{\nu,\lambda}(x_{(i)} - u)\right]^2 \quad (10)$$

According to the second method, the GPD is reparametrized through a new parameter $(\boldsymbol{\theta} = (\boldsymbol{\nu}) / \boldsymbol{\lambda})$ and based on the maximum likelihood method of estimation, the shape parameter $(\boldsymbol{\nu})$ can be expressed as a function of the parameter $(\boldsymbol{\theta})$ as follows:

$$\nu(\theta) = \frac{1}{n - n_u} \sum_{i=n_u+1}^{n} log[1 + \theta(x_{(i)} - u)] \qquad \forall\ 1 + \theta(x_{(i)} - u) > 0 \quad (11)$$

The parameters (λ, ν) of the GPD can be estimated using $\boldsymbol{ZZCZ\_2}$ method through the following two steps:

(First step)

$$\hat{\theta}_1 = arg\min_{\theta} \sum_{i=n_u+1}^{n} \left[\sum_{j=1}^{i} \left(\frac{1}{n-j+1} \right) + log[1 - F_n(u)] + log[1 - G_{\nu(\theta),\lambda}(x_{(i)} - u)] \right]^2 \quad (12)$$

(Second step)

$$\hat{\theta}_2 = arg\min_{\theta} \sum_{i=n_u+1}^{n} \left[\frac{(n+1)^2(n+2)}{i(n-i+1)} \right] \times \left[\frac{(i/(n+1)) - F_n(u)}{1 - F_n(u)} - G_{\nu(\theta),\lambda}(x_{(i)} - u) \right]^2 \quad (13)$$

Now, the shape parameter $(\boldsymbol{\nu})$ and the scale parameter $(\boldsymbol{\lambda})$ can be estimated using the following relations:

$$\hat{\nu} = \frac{1}{n - n_u} \sum_{i=n_u+1}^{n} log[1 + \hat{\theta}_2(x_{(i)} - u)], \quad \hat{\lambda} = \hat{\nu}/\hat{\theta}_2 \quad (14)$$

*3.2. The Proposed Methods of Estimation (**Sol**)*

Going along the lines of Zhao et al. (2019), the proposed estimation method is a two-step weighted minimization procedure. The parameters (λ, ν) of the GPD can be estimated using the proposed method (Sol) through the following two steps:

(First step)

The target of the first step is to find the values of the parameters (ν, λ) that minimize the median absolute deviations between $(G_{(i)} = -\ln[1 - F(X_{(i)})])$ and the corresponding theoretical medians, i.e.,

$$\hat{\theta}_1 = arg\min_{\theta}\{Median|G_{(i)} - Median(G_{(i)})|\}, \quad (15)$$

$$\hat{\theta}_1 = arg\min_{\theta}\{Median|log[1 - I_{0.5}^{-1}(i, n-i+1)] - log[1 - F_n(u)] - log[1 - G_{\nu(\theta),\lambda}(x_{(i)} - u)]|\},$$

Where $G_{(i)} = -log[1 - U_{(i)}]$, $Median(G_{(i)}) = -log[1 - I_{0.5}^{-1}(i, n-i+1)]$

(Second step)

Taking the estimates in the first step as initial values, the second step give the values of the parameters (ν, λ) that minimize the median absolute deviations between $(U_{(i)} = F(X_{(i)}))$ and the corresponding theoretical medians weighted by the reciprocals of $\boldsymbol{MAD}(\boldsymbol{U}_{(i)})$. After reparametrizing the GPD through the parameter $(\theta = (\nu) / \lambda)$ and making good use of the maximum likelihood method of estimation, the shape parameter (ν) can be expressed as a function of the parameter (θ) as indicated in equation (11). Thus, the second step of the proposed estimation method is given as:

$$\hat{\boldsymbol{\theta}}_2 = \boldsymbol{arg}\,\underset{\boldsymbol{\theta}}{\boldsymbol{min}}\{\boldsymbol{Median}\left|\left[\boldsymbol{U}_{(i)} - \boldsymbol{Median}(\boldsymbol{U}_{(i)})\right]/\boldsymbol{MAD}(\boldsymbol{U}_{(i)})\right|\}, \qquad (16)$$

$$\hat{\boldsymbol{\theta}}_2 = \boldsymbol{arg}\,\underset{\boldsymbol{\theta}}{\boldsymbol{min}}\{\boldsymbol{Median}\left|\left[\boldsymbol{F}_n(\boldsymbol{u}) + [\boldsymbol{1} - \boldsymbol{F}_n(\boldsymbol{u})]\boldsymbol{G}_{\nu(\theta),\lambda}(\boldsymbol{x}_{(i)} - \boldsymbol{u}) - \boldsymbol{Median}(\boldsymbol{U}_{(i)})\right]/\boldsymbol{MAD}(\boldsymbol{U}_{(i)})\right|\},$$

Where $U_{(i)} = F(X_{(i)})$, $Median(U_{(i)}) = I_{0.5}^{-1}(i, n - i + 1)$, *and* $MAD(U_{(i)})$ *can be numerically calculated by solving the following equation for the value z:*

$$\boldsymbol{I}_{I_{0.5}^{-1}(i,n-i+1)+z}(\boldsymbol{i}, \boldsymbol{n} - \boldsymbol{i} + \boldsymbol{1}) - \boldsymbol{I}_{I_{0.5}^{-1}(i,n-i+1)-z}(\boldsymbol{i}, \boldsymbol{n} - \boldsymbol{i} + \boldsymbol{1}) = \boldsymbol{0.5} \qquad (17)$$

Where the regularized incomplete beta function $\boldsymbol{I}_{\boldsymbol{v}}(\boldsymbol{a}, \boldsymbol{b})$ *is defined as:*

$$\boldsymbol{I}_{\boldsymbol{v}}(\boldsymbol{a}, \boldsymbol{b}) = \left[\int_{\boldsymbol{0}}^{\boldsymbol{v}} \boldsymbol{t}^{\boldsymbol{a}-\boldsymbol{1}}(\boldsymbol{1} - \boldsymbol{t})^{\boldsymbol{b}-\boldsymbol{1}}.\,\boldsymbol{dt}\right] \Big/ \boldsymbol{\beta eta}(\boldsymbol{a}, \boldsymbol{b}) \qquad (18)$$

And the inverse regularized incomplete beta function $\boldsymbol{z} \equiv \boldsymbol{I}_{\boldsymbol{v}}^{-1}(\boldsymbol{a}, \boldsymbol{b})$ *is calculated by solving the following equation for the value z:* $\boldsymbol{I}_{\boldsymbol{z}}(\boldsymbol{a}, \boldsymbol{b}) \equiv \boldsymbol{v}$

As mentioned before, the shape and scale parameter (ν, λ) *can be estimated using equations in (14).*

4. Simulation Studies

This section includes two simulation studies to evaluate and compare the empirical performance of all above mentioned existing and proposed estimation methods. In section 4.1, the scale and shape parameters of the GPD are estimated using all observations in a random sample. In section 4.2, the scale and shape parameters of the GPD are estimated using only observations above threshold. Table 1 shows the skewness degree and direction and the tail-heaviness degree of each member of the family of GP distributions used in the two simulation studies. The degrees of skewness are measured by the well-known Bawley coefficient of skewness ($\Delta = [(Q(0.75) - Q(0.5)) - (Q(0.5) - Q(0.25))] / [Q(0.75) - Q(0.25)]$). The degrees of tail-heaviness are measured by the Moors measure ($\breve{\beta} = [(Q(0.875) - Q(0.125)) - (Q(0.625) - Q(0.375))] / [Q(0.75) - Q(0.25)]$) introduced by Moors (1988) and accompany all distributions used in this third simulation study, where $Q(p)$ refers to the p^{th} population quantile.

Table 1. Skewness and Tail-heaviness Degrees of the considered members of the family of GP Distributions

$\boldsymbol{\nu}$	-3	-2	-1.5	-1	-0.5	0	1	2
Δ	-0.4615	-0.25	-0.1285	0	0.1317	0.2619	0.5	0.6875
$\breve{\beta}$	1.1731	1	0.9719	1	1.1028	1.3063	2.1714	4.0882

It should be noted that the GPD can be negatively skewed, symmetric, or positively skewed when the shape parameter ($\boldsymbol{\nu}$) is less that (-1), equal to (-1), or greater than (-1) respectively. Also, the tail-heaviness of the GPD reaches its minimum when the shape parameter ($\boldsymbol{\nu}$) is equal to (-1.5) and starts to increase as the value of ($\boldsymbol{\nu}$) deviates from (-1.5). The rate of increasing the tail-heaviness of the GPD is larger for values of ($\boldsymbol{\nu}$) greater than (-1.5).

4.1. Parameter Estimation under the Non- POT Approach

The target of this section is to compare the performance of five estimators in estimating the scale and shape parameters of the generalized Pareto distribution under two settings: the (Non-POT), where the whole random sample is used and the (POT), where only observations above threshold are used. Using the whole random sample, the first simulation study is designed as follows:

1. Generate a random sample of size ($\boldsymbol{n}$) from the GPD with scale parameter ($\boldsymbol{\sigma}$) and shape parameter ($\boldsymbol{\nu}$).
2. based on the whole random sample, estimate the parameters ($\boldsymbol{\lambda}, \boldsymbol{\nu}$) using all methods under investigation.
3. Repeat above steps 10,000 times.
4. Compute the mean square error (MSE) of each estimator as follows:

$$\boldsymbol{MSE}(\hat{\boldsymbol{\sigma}}) = \frac{\mathbf{1}}{\mathbf{10000}} \sum_{i=1}^{10000} [\hat{\boldsymbol{\sigma}}_i - \boldsymbol{\sigma}]^2 = \frac{\mathbf{1}}{\mathbf{10000}} \sum_{i=1}^{10000} [\hat{\boldsymbol{\lambda}}_i - \boldsymbol{\sigma}]^2, \qquad (19)$$

$$\boldsymbol{MSE}(\hat{\boldsymbol{\nu}}) = \frac{\mathbf{1}}{\mathbf{10000}} \sum_{i=1}^{10000} [\hat{\boldsymbol{\nu}}_i - \boldsymbol{\nu}]^2 \qquad (20)$$

This first simulation study was conducted with sample sizes ($\boldsymbol{n} = 50{,}100, 200{,}1000$), a scale parameter ($\boldsymbol{\sigma} = \mathbf{1}$), and shape parameters ($\boldsymbol{\nu} = \mathbf{-3}, \mathbf{-2}, \mathbf{-1.5}, \mathbf{-1}, \mathbf{-0.5}, \mathbf{0}, \mathbf{1}, \mathbf{2}$).

For all methods of estimation and for each combination of the considered different levels of "$\boldsymbol{n}$ and $\boldsymbol{\nu}$", the MSEs are given in tables 2 and 3. The main results deducted from these first two tables can be summarized as follows:

1. As expected, and as can be seen form tables 2 and 3, increasing the sample size substantially reduces the MSEs of all estimators.
2. As the value of ($\boldsymbol{\nu}$) gets closer to ($\mathbf{-1}$), the degree of skewness decreases and the empirical performance of estimators improves in terms of decreasing MSEs of both the scale and shape parameters.
3. As the value of the shape parameter ($\boldsymbol{\nu}$) deviates from ($\mathbf{-1.5}$), the degree of tail-heaviness increases and the empirical performance of estimators gets worse.
4. The proposed estimator (Sol) shows the best performance in terms of least MSEs.

Table 2. Mean square errors (MSE) of estimators of (σ, ν) under Non-POT ($n = 50, n = 100, \boldsymbol{\sigma} = \mathbf{1}$).

	$n = 50$					$n = 100$				
	SSPK	PK	ZZCZ_1	ZZCZ_2	Sol	SSPK	PK	ZZCZ_1	ZZCZ_2	Sol
$\sigma = 1$	0.01166	0.01036	0.01120	0.01984	0.01056	0.01134	0.01018	0.01068	0.01900	0.01004
$\nu = -3$	0.10495	0.09320	0.10079	0.18048	0.09578	0.10219	0.09166	0.09614	0.17281	0.09137
$\sigma = 1$	0.01382	0.01390	0.01261	0.01014	0.00933	0.01316	0.01348	0.01217	0.00966	0.00888
$\nu = -2$	0.05726	0.05565	0.05075	0.04091	0.04059	0.054566	0.053996	0.048987	0.038990	0.03865
$\sigma = 1$	0.01760	0.01883	0.01563	0.01447	0.01429	0.016755	0.017907	0.014870	0.013772	0.013585
$\nu = -1.5$	0.02563	0.02149	0.01909	0.01741	0.01665	0.024390	0.020431	0.018154	0.016557	0.015800
$\sigma = 1$	0.02106	0.01813	0.01935	0.01985	0.01811	0.020060	0.017259	0.018424	0.018898	0.017193
$\nu = -1$	0.02156	0.01225	0.01463	0.01374	0.01131	0.020524	0.011659	0.013920	0.013065	0.010770
$\sigma = 1$	0.02944	0.02726	0.02785	0.02124	0.02030	0.028052	0.025960	0.026520	0.020239	0.019309
$\nu = -0.5$	0.03243	0.02164	0.02273	0.01404	0.01374	0.030908	0.020611	0.021646	0.013373	0.013066
$\sigma = 1$	0.05113	0.04970	0.04868	0.04756	0.04749	0.048682	0.047312	0.046343	0.045280	0.045156
$\nu = 0$	0.07216	0.05783	0.05592	0.04456	0.04400	0.068730	0.055056	0.053247	0.042440	0.041887
$\sigma = 1$	0.07561	0.07609	0.07359	0.07271	0.07267	0.071977	0.072418	0.070042	0.069207	0.069145
$\nu = 1$	0.13694	0.11791	0.11463	0.09722	0.09618	0.130406	0.112250	0.109141	0.092582	0.091592
$\sigma = 1$	0.18207	0.17749	0.16654	0.16910	0.16567	0.173204	0.168810	0.158433	0.160874	0.157601
$\nu = 2$	0.47872	0.43230	0.42568	0.38842	0.38777	0.455875	0.411595	0.405316	0.369898	0.369234

Table 3. Mean square errors (MSE) of estimators of (σ, ν) under Non-POT ($n = 200, n = 1000, \boldsymbol{\sigma} = \mathbf{1}$).

	$n = 200$					$n = 1000$				
	SSPK	PK	ZZCZ_1	ZZCZ_2	Sol	SSPK	PK	ZZCZ_1	ZZCZ_2	Sol
$\sigma = 1$	0.01074	0.00984	0.00965	0.01732	0.00902	0.0058268	0.0071136	0.0013864	0.0038432	0.0008644
$\nu = -3$	0.09668	0.08859	0.08685	0.15746	0.08255	0.0525624	0.0640227	0.0124776	0.0346836	0.0120029
$\sigma = 1$	0.01187	0.01266	0.01130	0.00871	0.00799	0.0015026	0.0060694	0.0042943	0.0011074	0.0008735
$\nu = -2$	0.04918	0.05070	0.04545	0.03515	0.03478	0.0060582	0.0242935	0.0171940	0.0044407	0.0037596
$\sigma = 1$	0.01507	0.01606	0.01335	0.01237	0.01217	0.0015564	0.0012976	0.0012169	0.0011688	0.0008162
$\nu = -1.5$	0.02191	0.01831	0.01628	0.01486	0.01409	0.0020932	0.0013497	0.0013103	0.0012578	0.0004405
$\sigma = 1$	0.01805	0.01552	0.01656	0.01699	0.01536	0.0020114	0.0016369	0.0016921	0.0016949	0.0006636
$\nu = -1$	0.01846	0.01047	0.01249	0.01172	0.00968	0.0019273	0.0009891	0.0010682	0.0009793	0.0009743
$\sigma = 1$	0.02527	0.02337	0.02386	0.01824	0.01732	0.0030296	0.0026173	0.0026150	0.0022416	0.0014315
$\nu = -0.5$	0.02787	0.01855	0.01947	0.01205	0.01173	0.0035799	0.0020660	0.0020491	0.0014487	0.0010147
$\sigma = 1$	0.04378	0.04253	0.04167	0.04072	0.04050	0.0046054	0.0042949	0.0042835	0.0041992	0.0032317
$\nu = 0$	0.06187	0.04951	0.04790	0.03820	0.03766	0.0070219	0.0051575	0.0051418	0.0042554	0.0038163
$\sigma = 1$	0.06470	0.06506	0.06295	0.06220	0.06209	0.0065277	0.0062343	0.0062112	0.0061478	0.0056769
$\nu = 1$	0.11734	0.10094	0.09817	0.08331	0.08242	0.0128351	0.0104158	0.0103886	0.0091170	0.0090222
$\sigma = 1$	0.15548	0.15146	0.14223	0.14441	0.14146	0.0136584	0.0126579	0.0125620	0.0127419	0.0123371
$\nu = 2$	0.41019	0.37018	0.36459	0.33286	0.33217	0.0447411	0.0388284	0.0387964	0.0365536	0.0356327

4..2. Parameter Estimation under the POT Approach

Using only observations above threshold, this second simulation study is designed as follows:

1. Generate a random sample of size ($\boldsymbol{n}$) from the GPD with scale parameter ($\boldsymbol{\sigma}$) and shape parameter ($\boldsymbol{\nu}$).
2. Take the $\mathbf{100}\boldsymbol{q}$th quantile as a threshold value ($\boldsymbol{u} = \boldsymbol{G}^{-1}(\boldsymbol{q}; \boldsymbol{\sigma}, \boldsymbol{\nu})$).
3. Based on only observations above threshold ($\boldsymbol{u}$), estimate the parameters ($\boldsymbol{\lambda}, \boldsymbol{\nu}$) using all methods under investigation.
4. Repeat above steps 10,000 times.
5. Compute the mean square error (MSE) of each estimator as follows:

$$\boldsymbol{MSE}(\hat{\boldsymbol{\sigma}}) = \sum_{\boldsymbol{i}=\mathbf{1}}^{\mathbf{10000}} [\hat{\boldsymbol{\sigma}}_{\boldsymbol{i}} - \boldsymbol{\sigma}]^2 \Big/ 10000 = \sum_{\boldsymbol{i}=\mathbf{1}}^{\mathbf{10000}} \left[\hat{\boldsymbol{\lambda}}_{\boldsymbol{i}} \left(\mathbf{1} - \frac{\boldsymbol{n}_{\boldsymbol{u}}}{\boldsymbol{n}}\right)^{\hat{\boldsymbol{\nu}}_{\boldsymbol{i}}} - \boldsymbol{\sigma}\right]^{\mathbf{2}} \Big/ 10000 \quad (21)$$

$$\boldsymbol{MSE}(\hat{\boldsymbol{\nu}}) = \sum_{\boldsymbol{i}=\mathbf{1}}^{\mathbf{10000}} [\hat{\boldsymbol{\nu}}_{\boldsymbol{i}} - \boldsymbol{\nu}]^2 \Big/ 10000 \quad (22)$$

This second simulation study was conducted with sample sizes ($\boldsymbol{n} = \mathbf{1000}, \mathbf{10000}$), a scale parameters ($\boldsymbol{\sigma} = \mathbf{1}$), shape parameters ($\boldsymbol{\nu} = -\mathbf{3}, -\mathbf{2}, -\mathbf{1.5}, -\mathbf{1}, -\mathbf{0.5}, \mathbf{0}, \mathbf{1}, \mathbf{2}$), and only one level of threshold ($\boldsymbol{u} = \boldsymbol{F}^{-\mathbf{1}}(\boldsymbol{q})$) equals the $\mathbf{100} \times \mathbf{0.90} = \mathbf{90}^{\boldsymbol{th}}$ quantile. It should be noted that the scale parameter ($\boldsymbol{\sigma}$) of the GPD of the random variable ($\boldsymbol{X}$) is related to the scale parameter ($\boldsymbol{\lambda}$) of the GPD of the random variable ($\boldsymbol{X} - \boldsymbol{u} \le \boldsymbol{x} \mid \boldsymbol{X} > \boldsymbol{u}$) by the relation ($\boldsymbol{\sigma} = \boldsymbol{\lambda}(\mathbf{1} - \boldsymbol{q})^{\boldsymbol{\nu}}$).

For all methods of estimation and for each combination of the considered different levels of "$\boldsymbol{n}$ and $\boldsymbol{\nu}$", the MSEs are given in table 4. The main results deducted from tables 2, 3, and 4 can be summarized as follows:

1. As expected, even if the effective sample size is equal, the MSEs accompanying estimators under the POT approach are greater than the MSEs accompanying estimators under the Non-POT approach.
2. As the sample size increases or the skewness degree decreases or the tail-heaviness degree decreases, the MSEs of all estimators decrease.

3. The proposed estimator (Sol) shows the best performance as indicated by the corresponding least MSEs.

Table 4. Mean square errors (MSE) of estimators of (σ, ν) under POT ($\boldsymbol{\sigma = 1}$).

	$n = 1000$					$n = 10000$				
	SSPK	PK	ZZCZ_1	ZZCZ_2	Sol	SSPK	PK	ZZCZ_1	ZZCZ_2	Sol
$\sigma = 1$	3.110	2.599	3.257	0.484	0.467	2.44585	2.53491	2.52650	0.068095	0.067963
$\nu = -3$	0.112	0.112	0.112	0.089	0.088	0.12127	0.12900	0.12849	0.013393	0.013193
$\sigma = 1$	0.869	0.496	0.902	0.473	0.433	0.04848	0.08099	0.07334	0.032202	0.032070
$\nu = -2$	0.049	0.034	0.044	0.038	0.034	0.00530	0.01151	0.00911	0.003808	0.003539
$\sigma = 1$	0.298	0.335	0.227	0.222	0.219	0.02141	0.01613	0.01415	0.013761	0.012864
$\nu = -1.5$	0.026	0.017	0.016	0.016	0.015	0.00218	0.00125	0.00121	0.001185	0.000215
$\sigma = 1$	0.183	0.142	0.160	0.164	0.139	0.01875	0.01059	0.01103	0.010733	0.010362
$\nu = -1$	0.021	0.010	0.011	0.011	0.009	0.00203	0.00082	0.00087	0.000816	0.000205
$\sigma = 1$	0.344	0.281	0.253	0.194	0.190	0.03671	0.02079	0.01993	0.015824	0.014863
$\nu = -0.5$	0.031	0.018	0.018	0.013	0.012	0.00355	0.00172	0.00168	0.001265	0.001001
$\sigma = 1$	1.510	1.158	0.807	0.753	0.749	0.06392	0.04429	0.04161	0.038951	0.038540
$\nu = 0$	0.071	0.051	0.049	0.043	0.038	0.00708	0.00448	0.00442	0.004021	0.003868
$\sigma = 1$	4.939	4.452	2.813	2.844	2.809	0.10387	0.09011	0.08290	0.082309	0.082041
$\nu = 1$	0.130	0.107	0.104	0.097	0.096	0.01216	0.00984	0.00985	0.009529	0.008586
$\sigma = 1$	357.396	397.112	196.404	220.334	101.403	0.48205	0.44758	0.38760	0.447380	0.387222
$\nu = 2$	0.459	0.424	0.419	0.410	0.391	0.04171	0.03567	0.03545	0.035479	0.034847

5. Application

In section 4, two simulation studies were conducted to evaluate and compare the empirical performance of the proposed estimator under two approaches, the Non-POT and POT. In this section, the proposed and competitor methods of estimation are applied to a real-world data set, the Bilbao waves data, used in Castillo and Hadi (1997). The data measures zero-crossing hourly mean periods (in seconds) of the sea waves in a Bilbao buoy in January 1997. This data is modelled by the GPD with different thresholds starting from 7 up to 9.5. The overall goodness-of-fit of each estimation method is assessed by the average scale absolute error (ASAE) defined as:

$$ASAE = \frac{1}{\mathrm{n} - n_u} \sum_{i=n_u+1}^{\mathrm{n}} \left[\left| x_{(i)} - \hat{x}_{(i)} \right| / \left(x_{(n)} - x_{(n_u+1)} \right) \right], \qquad (23)$$

where

$$\hat{x}_{(i)} = G^{-1}\left(\frac{i}{n+1}; u, \hat{\lambda}, \hat{\nu} \right) \qquad (24)$$

The values of ASAE for each method of estimation at various threshold levels are computed and summarized in table 5. The estimated parameters and their standard errors (in parentheses) for all estimation methods are given in table 6, for several threshold values u. The standard errors are computed based on 1000 bootstrap samples. According to tables 5 and 6, the following results can be deducted:

1. The best choice for the threshold value is ($u = 7.5$) as it corresponds to the least ASAE.
2. Based on ASAEs in table 5, the proposed estimator (Sol) outperforms all other competitor estimator as it has the least ASAE.
3. Based on standard errors in table 6, the proposed estimator (Sol) performs as well or better than all other competitor estimators as it has small standard errors.
4. At a threshold value of ($u = 7.5$), the proposed (Sol) estimates of (σ and $\boldsymbol{\nu}$) are (2.2568) and (-0.8512) with standard errors of (0.0020) and (0.0088) respectively.

Table 5. The Bilbao Waves Data: Average Scaled Absolute Errors (ASAE) for All Estimation Methods

Threshold u	$\mathbf{n} - n_u$	ASAE				
		SSPK	PK	ZZCZ_1	ZZCZ_2	Sol
7	179	0.0584	0.0388	0.0317	0.0314	0.0234
7.5	154	0.03587	0.0278	0.0234	0.0200	0.0138
8	106	0.0388	0.0292	0.0264	0.0261	0.0202
8.5	69	0.0413	0.0385	0.0298	0.0281	0.0283
9	41	0.0498	0.0452	0.0423	0.0411	0.0371
9.5	17	0.0935	0.0897	0.0811	0.0801	0.0779

Table 6. The Bilbao Waves Data: Estimated Parameters and the Corresponding Standard Errors

u	$\hat{\sigma}$					$\hat{\nu}$				
	SSPK	PK	ZZCZ_1	ZZCZ_2	Sol	SSPK	PK	ZZCZ_1	ZZCZ_2	Sol
7	3.9459 (0.0214)	2.5246 (0.0174)	2.4123 (0.012)	2.3621 (0.0105)	2.2568 (0.0088)	−0.8267 (0.0185)	−0.8612 (0.0127)	−0.8111 (0.0095)	−0.8437 (0.0078)	−0.8512 (0.0020)
7.5	1.0872 (0.0183)	1.8973 (0.0103)	1.7194 (0.0099)	1.6438 (0.0087)	1.5439 (0.0065)	−0.5979 (0.0136)	−0.5537 (0.0123)	−0.5134 (0.0010)	−0.4935 (0.0095)	−0.4436 (0.0021)
8	1.7832 (0.0224)	1.4903 (0.0131)	1.5378 (0.0113)	1.4961 (0.0095)	1.3782 (0.0070)	−0.7663 (0.0186)	−0.5849 (0.0155)	−0.7215 (0.0098)	−0.6123 (0.0083)	−0.5884 (0.0027)
8.5	1.2809 (0.0255)	1.2572 (0.0128)	1.2314 (0.0117)	1.1576 (0.0093)	0.9258 (0.0076)	−0.2978 (0.0305)	−0.4171 (0.0233)	−0.5382 (0.0153)	−0.8637 (0.0089)	−0.6783 (0.0034)
9	0.3296 (0.0266)	1.0179 (0.0139)	0.9783 (0.0112)	0.9138 (0.0083)	0.8155 (0.0065)	−0.7526 (0.0435)	−0.1233 (0.0305)	−0.6837 (0.0253)	−0.8134 (0.0201)	−0.4113 (0.0056)
9.5	0.3701 (0.0407)	0.3975 (0.0122)	0.4162 (0.0111)	0.4356 (0.0094)	0.4955 (0.0059)	−1.4828 (0.0655)	−1.1225 (0.0497)	−1.4851 (0.0321)	−1.8512 (0.0057)	−0.9631 (0.0089)

6. Conclusions

In this paper, an efficient new estimator of the GPD parameters is proposed. Through two simulation studies and an application to a real-life data, the empirical performance of the proposed (Sol) estimator is evaluated and compared with some existing estimators. Simulation results show that for the estimation of the scale and also the shape parameters, the (Sol) estimator performs as well or significantly better than all other competitor estimators. It is recommended to estimate the GPD parameters using the proposed new estimator using all sample observations or using only observations above a certain threshold.

References

[1] Balkema, A. A., & de Haan, L. (1974). Residual life time at great age. The Annals of Probability, 2, 792–804.

[2] Castillo, E., & Hadi, A. S. (1997). Fitting the generalized pareto distribution to data. Journal of the American Statistical Association, 92, 1609–1620.

[3] Coles, S. (2001). An Introduction to Statistical Modeling of Extreme Values. Springer, London.

[4] de Haan, L., & Ferreira, A. (2007). Extreme value theory: an introduction. Springer Science & Business Media.

[5] del Castillo, J., & Serra, I. (2015). Likelihood inference for generalized pareto distribution. Computational Statistics and Data Analysis, 83, 116–128.

[6] Ferreira, A., & de Haan, L. (2015). On the block maxima method in extreme value theory: PWM estimators. The Annals of Statistics, 43, 276–298.

[7] Fisher, R. A., & Tippett, L. H. C. (1928). Limiting forms of the frequency distribution of the largest or smallest member of a sample. Proceedings of the Cambridge Philosophical Society, 24, 180–290.

[8] Gnedenko, B. (1943). Sur la distribution limite du terme maximum d'une serie aleatoire. Annals of Mathematics, 44, 423–453.

[9] Greenwood, J.A., Landwehr, J.M., Matalas, N.C., Wallis, J.R. (1979). Probability weighted moments: definition and relation to parameters of several distributions expressible ininverseform.WaterResour.Res.15(5),1049–1054.

[10] Grimshaw, S.D. (1993). Computing maximum likelihood estimates for the generalized Pareto distribution. Technometrics 35,185–191.

[11] Hosking, J. R., & Wallis, J. R. (1987). Parameters and quantile estimation for the generalized pareto distribution. Technometrics, 29, 339–349.

[12] Hosking, J.R.M., 1990. L-Moments: analysis and estimation of distributions using linear combinations of order statistics. J. Roy. Statist. Soc. B52, 105–124.

[13] Kang, S. & Song, J. (2017). Parameter and quantile estimation for the generalized Pareto distribution in peaks over threshold framework. J. Korean Stat. Soc., 46, 487–501.

[14] McNeil, A.J., & Saladin, T. (1997). The peaks over thresholds method for estimating high quantiles of loss distributions. In Proceedings of 28th international ASTIN colloquium (pp. 23–43).

[15] Moors, J. J. A. (1988). A Quantile Alternative for Kurtosis. Journal of the Royal Statistical Society, series D (The Statistician); 37, pp. 25-32.

[16] Park, M., & Kim, J. H. T. (2016). Estimating extreme tail risk measures with generalized pareto distribution. Computational Statistics and Data Analysis, 98, 91–104.

[17] Pickands, J. (1975). Statistical inference using extreme order statistics. The Annals of Statistics, 3, 119–131.

[18] Rasmussen, P.F., 2001. Generalized probability weighted moments: application to the generalized Pareto distribution. WaterResour.Res.37(6),1745–1751.

[19] Song, J., & Song, S. (2012). A quantile estimation for massive data with generalized pareto distribution. Computational Statistics and Data Analysis, 56, 143–150.

[20] Wang, C., & Chen, G. (2016). A new hybrid estimation method for the generalized pareto distribution. Communications in Statistics—Theory and Methods, 45, 4285–4294.

[21] Zhang, J. (2007). Likelihood moment estimation for the generalized pareto distribution. Australian and New Zealand Journal of Statistics, 49, 69–77.

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

[23] Zhang, J. (2010). Improving on estimation for the generalized pareto distribution. Technometrics, 52, 335–339.

[24] Zhao, X., Zhang, Z., Cheng, W., Zhang, P. (2019). A New Parameter Estimator for the Generalized Pareto Distribution under the peaks over threshold framework. Mathematics, 7, 406.