KYBERNETIKA — VOLUME *61* (2025), NUMBER 4, PAGES 467–480

A REGRESSION METHOD OF ESTIMATION FOR GENERALIZED EXTREME VALUE DISTRIBUTION

Anand R and Chandran C

This study focuses on parameter estimation for the generalized extreme value distribution (GEVD) using the regression method described by [27]. A regression equation is derived from the cumulative distribution function and the scale parameter is estimated by applying the iterative re-weighted least squares in this regression equation. For estimating the shape parameter, a profile likelihood is constructed based on this regression equation. A comparison study of the regression method with other existing estimators derived from the method of moments, maximum likelihood, probability-weighted moments, l-moments, and maximum product spacing is performed for the GEVD. Also, the left truncated GEVD is considered and the behaviour of its hazard function is studied. The parameter estimates of the left truncated GEVD is also derived using the regression method. An extensive simulation study is conducted and the efficiencies of the estimation techniques are analysed. The bootstrap confidence intervals for the estimators are also constructed. Finally, a real data analysis is carried out to illustrate the applicability of the models and estimation techniques.

Keywords: generalized extreme value distribution, regression method, Box–Cox transformation, profile likelihood

Classification: 60G70, 62F35

1. INTRODUCTION

The theory of extremes provides models which can be used in extreme situations such as flood, drought, earthquake, war, bush fire, stock market etc. The classical theory of extremes deals with the distributional properties of the statistics $M_n = \max(X_1, \ldots, X_n)$ and $m_n = \min(X_1, \ldots, X_n)$ of independent and identically distributed random variables $X_1, \ldots, X_n$. The pioneering works on the distributional aspects of extremes were carried out by [12] and [11]. It was in the works of [11], the versions of the well known three types of limiting distributions of extremes, also known as the type-I, type-II, and type-III extreme value distributions, were derived. These three models for extremes can be briefly described as follows. Let $\{X_n, n \geq 1\}$ be a sequence of independent and identically distributed random variables. If the asymptotic distribution of M_n under

DOI: 10.14736/kyb-2025-4-0467

proper normalization exits, *i.e.*, if

$$P\left(\frac{M_n - b_n}{a_n}\right) \quad \xrightarrow{w} \quad G(x), \tag{1.1}$$

for sequences of constants $\{a_n > 0\}$ and $\{b_n\}$ and a non-degenerate cumulative distribution function (CDF) $G(\cdot)$, then the CDF $G(\cdot)$ is one among the following three:

$$\begin{aligned}
&\text{Type-I} : G_1(x) = \exp\Bigg(-\exp\Bigg(-\frac{x-\mu}{\sigma}\Bigg)\Bigg), \qquad -\infty < x < \infty \\
&\text{Type-II} : G_2(x) = \begin{cases} 0, & x \le \mu \\ \exp\Bigg(-\Big(\frac{x-\mu}{\sigma}\Big)^{-\xi}\Bigg), & \text{for some } \xi > 0,\ x > \mu \end{cases} \\
&\text{Type-III} : G_3(x) = \begin{cases} \exp\Bigg(-\Bigg(-\Big(\frac{x-\mu}{\sigma}\Big)^{-\xi}\Bigg)\Bigg), & \text{for some } \xi < 0,\ x \le \mu \\ 1, & x > \mu. \end{cases}
\end{aligned} \tag{1.2}$$

When $\mu = 0$ and $\sigma = 1$, the type-I, type-II, and type-III distributions in (1.2) are the standard extreme value distributions. A complete theoretical framework of this result is done by [13]. For a given data we fit type-I, type-II, and type-III extreme value distributions and the best fitted distribution is used for further analysis of extremes. [28] and [17] unified these three families of distributions into a single family known as the generalized extreme value distribution (GEVD), by incorporating a new parameter in the model. The distribution function of GEVD proposed by [28] and [17] is of the form

$$G(x|\mu,\sigma,\xi) = \exp\Bigg(-\Bigg(1 - \frac{\xi(x-\mu)}{\sigma}\Bigg)^{\frac{1}{\xi}}\Bigg), \tag{1.3}$$

where $-\infty < \mu < \infty$ is the location parameter, $\sigma > 0$, is the scale parameter, and $-\infty < \xi < \infty$, is the shape parameter. Also, $x < \mu + \frac{\xi}{\sigma}$, when $\xi > 0$ and $x > \mu + \frac{\xi}{\sigma}$, when $\xi < 0$. The distribution function $G(x)$ given in equation (1.3) is the three parameter GEVD. The case $\xi = 0$, is interpreted as $\lim_{\xi \to 0} G(x|\mu,\sigma,\xi)$ which is the type-I extreme value distribution or Gumbel distribution $(G_1(x))$. For $\xi > 0$, $G(x)$ is the type-II extreme value distribution or Fréchet distribution $(G_2(x))$, and for $\xi < 0$, $G(x)$ is the type-III extreme value distribution or reverse Weibull distribution $(G_3(x))$. The practical advantage of model (1.3) is that by estimating the shape parameter ξ, the model for extreme is determined by the data itself. Detailed theoretical aspects of the above results, its consequences and practical applications can be seen in [18, 23] and [10]. Practical applications of model (1.3) in hydrology were discussed in [17] and a recent review on model (1.3) can be seen in [6]. Moreover, the national environment research council in 1975, recommended the use of GEVD in flood related studies in the United States.

The parameter estimation of the model (1.3) have been studied by various authors. An initial attempt to find the maximum likelihood (ML) estimates of GEVD can be seen in [22], and it was later developed by [14] and [25]. The Fisher information matrix for ML estimates of GEVD was also derived in [22]. It was [25], who described the existence of ML estimates for the parameters of GEVD in complete generality. [25] also showed that the regularity conditions for ML estimates holds only when the shape parameter ξ lies in the interval $\left(0, \frac{1}{2}\right)$. The asymptotic normality of ML estimates of the parameters of GEVD were established in [5]. Recently, [32] proved uniqueness and global optimality of the ML estimates. Also, [20], using simulation studies showed that for small samples, ML estimates of the parameters of GEVD leads to absurd results. [26] derived the moments method of estimation (MOM) of the parameters of GEVD. [16] gave a detailed note on how to calculate probability weighted moment (PWM) estimators of the parameters of GEVD. Similarly the l-moment (LM) estimates, which are equivalent to PWM estimates, were calculated by [15]. Moreover, in [16] it was shown that, for sample sizes from 15 to 100, the PWM or equivalent LM perform better in terms of bias and variance, than ML estimates. In [19], the MOM, PWM, and ML estimates are compared for sample sizes 10 to 50 and for $-0.25 < \xi < 0.3$, and it was inferred that the MOM estimates perform better than the ML and PWM estimates, in terms of root mean square error. [29] derived the maximum product spacing (MPS) estimators for the parameters of GEVD in situations where the ML estimates breaks down. Moreover, they observed that the MPS estimates are more efficient than the ML estimates for the parameters of GEVD. [29] also illustrated through simulation studies that for small samples, MPS technique provide more stable estimates than the ML and PWM estimates for the parameters of GEVD, in terms of bias and MSE. Moreover, [2] compared the MPS estimates of the parameters of GEVD with the corresponding LM estimates and it was observed that their efficiencies are relatively close. A detailed review on the various estimation techniques and their comparisons for the parameters of GEVD can be seen in [30] and [1]. The above mentioned estimation techniques have certain flaws in them. One notable limitation is the absence of ML estimates for $\xi < -\frac{1}{2}$. Additionally, traditional methods such as MOM, PWM, LM, and MPS often yield absurd results, particularly when confronted with small sample sizes. Due to these drawbacks, researchers are trying to develop better procedures to estimate the parameters of GEVD.

This served as a motivation for us to find new estimation techniques for the parameters of GEVD which rectifies the above mentioned limitations. [27] introduced one such technique called the regression method of estimation for the generalized Pareto distribution (GPD). [27] showed through simulation studies that the parameter estimates of GPD derived using the regression method exists throughout the parameter space and that it works well for small samples. [21] showed that for modelling extremes the GPD, is equivalent to the GEVD model whenever the maximum has a non-degenerate limit distribution $G(\cdot)$, as mentioned in equation (1.1). This connection is crucial for extreme value theory because it allows for modelling extremes using threshold excedences (GPD) or block maxima (GEVD). Therefore, we examine the regression method for estimating the parameters of the GEVD.

This paper aims to investigate the regression method for estimating the parameters of the GEVD model and to compare it with existing estimation techniques. The rest of the article is organised as follows: The main method is described in section 2. Also, a left truncated GEVD model is considered and the corresponding parameter estimation using the regression method is discussed. The behaviour of the hazard function of the truncated model is also analysed in this section. In section 3, an algorithm for constructing the bootstrap confidence intervals for the parameters of GEVD using the regression method is provided. A rigorous Monte Carlo simulation study is carried out in section 4 using simulated random samples to analyse and compare the regression method of estimation with the MOM, ML, PWM, LM, and MPS estimates. In section 5, a real life data of international crude oil prices during the initial stages of the Ukraine war is considered for illustration purposes. Finally, a brief conclusion of the study is given in section 6.

2. REGRESSION METHOD

In this section a regression method of estimation for the parameters of GEVD is discussed. The procedure is analogous to the one proposed by [27] to estimate the parameters of GPD. For computational ease, we assume $\mu = 0$. The method is not in general applicable to all models. However, if a linear regression equation can be obtained by inverting the CDF, then this method can be applied. Consider the CDF $G(\cdot)$ of GEVD given in (1.3). Now, by inverting $G(\cdot)$ we get

$$\log G(x) = -\left(1 - \frac{\xi x}{\sigma}\right)^{\frac{1}{\xi}}$$

$$\begin{aligned} \implies -1 + (-\log G(x))^{\xi} &= -\frac{\xi x}{\sigma} \\ \implies \frac{(-\log G(x))^{\xi} - 1}{\xi} &= -\frac{x}{\sigma}, \end{aligned} \tag{2.1}$$

which can be written in the form of a regression equation as described below. Let $(X_1, \ldots, X_n)$ be a random sample of size n from GEVD and $G_n(\cdot)$ be the empirical distribution function. Let $x_{(1)}, \ldots, x_{(n)}$ be the ordered statistics corresponding to the given sample. Since the empirical distribution function $G_n(\cdot)$, is an estimate of the $G(\cdot)$, by substituting $G_n(\cdot)$ in equation (2.1), we obtain

$$\frac{(-\log G_n(x))^{\xi} - 1}{\xi} = -\frac{x}{\sigma} + \epsilon_r, \tag{2.2}$$

where

$$G_n(x) = \begin{cases} 0, & x < x_{(1)} \\ \frac{r}{n}, & x_{(r)} \le x < x_{(r+1)}, \ r = 1, 2, \ldots, n \\ 1, & x > x_{(n)}. \end{cases} \tag{2.3}$$

For the given sample $x_{(1)}, \ldots, x_{(n)}$, we can write (2.2) as

$$G_{\xi}(x_{(r)}) = -\frac{x_{(r)}}{\sigma} + \epsilon_r, \quad r = 1, 2, \ldots, n, \tag{2.4}$$

where, $G_{\xi}(x_{(r)}) = \frac{(-\log G_n(x))^{\xi} - 1}{\xi}$, and ϵ_r is the error term, resulting from substituting $G(\cdot)$ with its estimate $G_n(\cdot)$. Note that, (2.4) is in the form of a regression equation which can be written in vector form as follows:

$$\boldsymbol{G}_{\xi} = \boldsymbol{X}\beta + \boldsymbol{\epsilon}, \tag{2.5}$$

where $\boldsymbol{G}_\xi = (G_\xi(x_{(1)}), \ldots, G_\xi(x_{(n)}))$, is an $n \times 1$ vector, $\boldsymbol{X} = (x_{(1)}, \ldots, x_{(n)})$, is an $n \times 1$ vector, $\beta = -\frac{1}{\sigma}$ and $\boldsymbol{\epsilon} = (\epsilon_1, \ldots, \epsilon_n)$ is an $n \times 1$ vector. Therefore, it is enough to estimate the parameters β and ξ of the regression equation (2.4). For this, we follow the Box–Cox procedure discussed in [31] where a Laplace assumption profile likelihood is constructed for a series of ξ's. Then, ξ is estimated as the value which maximizes the profile likelihood and β is estimated using iterative weighted least squares technique. The Box–Cox variance stabilizing transformation, suggested by [4] corresponding to the regression equation (2.2) is

$$G_\xi(x_{(r)}) = \begin{cases} \frac{(-\log G_n(x))^\xi - 1}{\xi}, & \xi \neq 0 \\ \log(-\log G_n(x)), & \xi = 0. \end{cases} \tag{2.6}$$

We now estimate the shape parameter ξ using profile likelihood method discussed in [31]. For this, we construct a profile likelihood for G_ξ as a function of ξ and β and obtain the estimate of ξ by assuming β as the nuisance parameter. As the left hand side of equation (2.2) is a function involving ξ, a Jacobian needs to be evaluated in order to obtain the profile likelihood. The Jacobian $J(\xi)$ is obtained as

$$J(\xi) = \prod_{r=1}^{n} \frac{\big(-\log G_\xi(x_{(r)})\big)^{(\xi-1)}}{G_\xi(x_{(r)}}. \tag{2.7}$$

The Laplace probability density with mean zero is given by

$$f(x) = \frac{1}{2\,\phi} \exp\left(-\frac{|u|}{\phi}\right), \quad -\infty < x < \infty,\ \phi > 0. \tag{2.8}$$

The variance of the above Laplace distribution is $2\,\phi$. For a given sample $(X_1, X_2, \ldots, X_n)$ of size n, the ML estimate of ϕ is the mean absolute deviation about the median, *i.e.*:

$$\hat{\phi} = \frac{1}{n} \sum_{i=1}^{n} \left|\mathbf{y} - \mathbf{X}\hat{\beta}\right|. \tag{2.9}$$

Following the procedure of [31], the Laplace profile likelihood for a given ξ denoted by $pL(\xi)$, is obtained as

$$pL(\xi) = C + \sum_{r=1}^{n} \log \left| \frac{\big(-\log G_\xi(x_{(r)})\big)^{(\xi-1)}}{G_\xi(x_{(r)}} \right| - n \log(\hat{\phi}(\xi)), \tag{2.10}$$

where C represents the proportionality constant. The estimate $\hat{\beta}$ of β is evaluated using the iterative weighted least squares technique. The iteration formula is given by

$$\hat{\beta}^{(j+1)} = (\mathbf{X}'W^{(j)}\mathbf{X})^{-1}\mathbf{X}'W^{(j)}G_\xi(x_{(r)}), \tag{2.11}$$

where $W^{(j)}$ is the diagonal matrix of weights with entries, $w_r^{(j)} = |G_\xi(x_{(r)}) - x_{(r)}\beta^{(j)}|^{-1}$, $r = 1, 2, \ldots, n$ and $j = 0, 1, \ldots$ denotes the number of iterations. For estimating the parameters β and ξ, we follow the method suggested by [31]. That is, we apply iterative weighted least squares estimate $\hat{\beta}$ for β given in (2.11), to the profile likelihood equation (2.10). Thus, the profile likelihood equation becomes a function solely depending on ξ. From this, the estimate $\hat{\xi}$ of ξ is obtained as the value of ξ which maximizes the profile likelihood. Now, to find $\hat{\beta}$, an initial value $\beta^{(0)}$ of β, is arbitrarily chosen and substituted in (2.11) to evaluate $\beta^{(1)}$. Once $\beta^{(1)}$ is obtained, $\beta^{(0)}$ is replaced by $\beta^{(1)}$ to obtain $\beta^{(2)}$. This process is continued out until $|\beta^{(j+1)} - \beta^{(j)}| < \delta$, where δ is a small quantity.

We note that the closed form expression for the above estimates do not exist. So, further evaluation of the estimates are carried out using simulation techniques and are discussed in section 4. Now, we consider a left truncated GEVD and its parameter estimation using the above method.

2.1. Truncated GEVD

To apply GEVD in reliability contexts, we consider the left truncated GEVD (TGEVD) at zero, as introduced by [3]. The TGEVD was employed to calibrate wind speed forecasts, thereby preventing negative wind speed predictions. The CDF of TGEVD is defined as

$$G_0(x|\mu_0,\sigma_0,\xi_0) = \begin{cases} 0, & x < 0 \\ \dfrac{\exp\left(-\left(1-\frac{\xi_0(x-\mu_0)}{\sigma_0}\right)^{\frac{1}{\xi_0}}\right) - \exp\left(-\left(1+\frac{\xi_0\mu_0}{\sigma_0}\right)^{\frac{1}{\xi_0}}\right)}{1-\exp\left(-\left(1+\frac{\xi_0\mu_0}{\sigma_0}\right)^{\frac{1}{\xi_0}}\right)}, & 0 < x < \frac{\sigma_0}{\xi_0} + \mu_0 \\ 1, & x > \frac{\sigma_0}{\xi_0} + \mu_0, \end{cases} \tag{2.12}$$

where μ_0 is the location parameter, σ_0 is the scale parameter and ξ_0 is the shape parameter of TGEVD. The corresponding probability density function is

$$g_0(x|\sigma_0,\xi_0) = \frac{\left(1-\frac{\xi_0(x-\mu_0)}{\sigma_0}\right)^{\frac{1}{\xi_0}-1}\exp\left(-\left(1-\frac{\xi_0(x-\mu_0)}{\sigma_0}\right)^{\frac{1}{\xi_0}}\right)}{\sigma_0\left(1-\exp\left(-\left(1+\frac{\xi_0\mu_0}{\sigma_0}\right)^{\frac{1}{\xi_0}}\right)\right)}, \qquad 0 < x < \frac{\sigma_0}{\xi_0} + \mu_0. \tag{2.13}$$

The hazard function of TGEVD, denoted by $h(x)$ is obtained as

$$\begin{aligned} h(x) &= \frac{g_0(x|\mu_0,\sigma_0,\xi_0)}{1-G_0(x|\mu_0,\sigma_0,\xi_0)} \\ &= \frac{\left(1-\frac{\xi_0(x-\mu_0)}{\sigma_0}\right)^{\frac{1}{\xi_0}}}{\sigma_0\left(1-\exp\left[\left(1-\frac{\xi_0(x-\mu_0)}{\sigma_0}\right)^{\frac{1}{\xi_0}} - \left(1+\frac{\xi_0\mu_0}{\sigma_0}\right)^{\frac{1}{\xi_0}}\right]\right)}. \end{aligned} \tag{2.14}$$

In the upcoming theorem, we analyse the behaviour of hazard function of the TGEVD model.

Theorem 2.1. Let X be a non-negative random variable following TGEVD. Then, the TGEVD has decreasing hazard rate.

Proof. Consider the hazard function $h(x)$ of TGEVD, given in equation (2.14). The first derivative of $h(x)$ is given by

$$\frac{d}{dx}(h(x)) = -\frac{\left(1-\frac{\xi_0(x-\mu_0)}{\sigma_0}\right)^{\frac{1}{\xi_0}-1}\left(1-\exp\left[\left(1-\frac{\xi_0(x-\mu_0)}{\sigma_0}\right)^{\frac{1}{\xi_0}} - \left(1+\frac{\xi_0\mu_0}{\sigma_0}\right)^{\frac{1}{\xi_0}}\left(1-\frac{\xi_0(x-\mu_0)}{\sigma_0}\right)^{\frac{1}{\xi_0}}\right]\right)}{\sigma_0^4\left(\left(1-\exp\left[\left(1-\frac{\xi_0(x-\mu_0)}{\sigma_0}\right)^{\frac{1}{\xi_0}} - \left(1+\frac{\xi_0\mu_0}{\sigma_0}\right)^{\frac{1}{\xi_0}}\right]\right)\right)^2}. \tag{2.15}$$

Since $0 < x < \frac{\sigma_0}{\xi_0} + \mu_0$ and $\frac{\sigma_0}{\xi_0} > -\mu_0$, $\frac{d}{dx}(h(x))$ is negative, which implies that, $h(x)$ is a decreasing function, which completes the proof. □

[3] has showed that, the mean of TGEVD exists only when $\xi_0 < 1$. Due to this constraint, the estimation of ξ_0 using MOM, PWM, and LM are restricted. So, we inspect the regression method of estimation discussed in section 2 for the TGEVD. As in section 2, we assume $\mu_0 = 0$. Then, the CDF (2.12) becomes

$$G_0(x|\sigma_0, \xi_0) = \frac{\exp\left(-\left(1-\frac{\xi_0(x-\mu_0)}{\sigma_0}\right)^{\frac{1}{\xi_0}}\right) - \exp(-1)}{1-\exp(-1)}. \tag{2.16}$$

The regression equation obtained from (2.16) is given as

$$\frac{\left(-\log\left(G_0(x)(1-\mathrm{e}^{-1}) + \mathrm{e}^{-1}\right)\right)^{\xi_0} - 1}{\xi_0} = -\frac{x}{\sigma_0}. \tag{2.17}$$

Following the procedure outlined in section 2, we derive the parameter estimates for the left TGEVD using the regression method. An illustration of this technique is also provided through simulation, as discussed in section 4.

For both GEVD and TGEVD, we note that the regression method provides estimates which can only be evaluated using iterative procedures. Thus the corresponding confidence intervals for the parameter estimates cannot be constructed. So as an alternative, bootstrap confidence intervals are considered.

3. BOOTSTRAP CONFIDENCE INTERVAL

The current section deals with the evaluation of bootstrap confidence intervals for the parameter estimates of GEVD and TGEVD using regression method. Bootstrap is a resampling technique introduced by [7], wherein the observed sample is considered as a finite population and random samples are generated with replacement from the observed sample. Detailed discussions on bootstrap are available in [8] and [9]. A practical oriented approach based on R programming can be seen in [24]. Here, the intervals are constructed using the bootstrap-p method, were p stands for percentile. Algorithm for evaluating the bootstrap confidence intervals for the parameters of GEVD is given below. Let $(X_1, \ldots, X_n)$ be a random sample from GEVD.

1. Generate a bootstrap sample $\{x_1^*, x_2^*, \ldots, x_n^*\}$ of size n from $(X_1, \ldots, X_n)$. Using this sample, the bootstrap estimates $\hat{\sigma}^*$ and $\hat{\xi}^*$ are evaluated using regression method discussed in section 3.
2. Step 1 is repeated B times.
3. Let $F_1^*(x)$ and $F_2^*(x)$ be the empirical distribution functions of bootstrap replicates for σ^* and ξ^* respectively. Define $\sigma^{*p}(x) = F_1^{*-1}(x)$ and $\xi^{*p}(x) = F_2^{*-1}(x)$. Then the approximate $100(1-\alpha)\%$ bootstrap-p asymptotic confidence intervals for $\hat{\sigma}^*$ and $\hat{\xi}^*$ are

$$\left(\sigma^{*p}(\alpha/2)\ ,\ \sigma^{*p}(1-\alpha/2)\right) \qquad \text{and} \qquad \left(\xi^{*p}(\alpha/2)\ ,\ \xi^{*p}(1-\alpha/2)\right).$$

respectively.

Similarly, the bootstrap confidence intervals for the parameter estimates of TGEVD can be obtained as

$$\left(\sigma_0^{*p}(\alpha/2)\ ,\ \sigma_0^{*p}(1-\alpha/2)\right) \qquad \text{and} \qquad \left(\xi_0^{*p}(\alpha/2)\ ,\ \xi_0^{*p}(1-\alpha/2)\right).$$

4. SIMULATION STUDY

In this section, we perform a rigorous Monte Carlo simulation to analyse the efficiencies of the regression method compared to other existing methods of estimation for the parameters of GEVD and TGEVD. For this purpose, the mean squared error (MSE) and absolute bias of the estimates are evaluated. For illustration purpose, 4 sets of parameter combinations ($\xi = 0.1$, $\sigma = 2$), ($\xi = 0.2$, $\sigma = 2$), ($\xi = -2$, $\sigma = 1$) and ($\xi = 2$, $\sigma = 1$) and five different sample sizes, 10, 20, 50, 100 and 500 are considered. R programming software is used for computation and the results are based on 1000 replications.

Sample Size	Shape Parameter		Scale Parameter		Sample Size	Shape Parameter		Scale Parameter	
	MSE	bias	MSE	bias		MSE	bias	MSE	bias
	$\xi_0 = 0.1$ and $\sigma_0 = 2$					$\xi_0 = 1$ and $\sigma_0 = 1$			
10	0.4215	0.6605	0.1790	0.2836	10	0.2943	0.5200	0.0150	0.1012
20	0.3343	0.4930	0.0992	0.2444	20	0.2636	0.4461	0.0055	0.0588
50	0.2564	0.4138	0.0252	0.1160	50	0.2499	0.4117	0.0012	0.0275
100	0.1943	0.3834	0.0090	0.0707	100	0.2048	0.3341	0.0006	0.0199
500	0.1852	0.3041	0.0015	0.0295	500	0.1168	0.2909	0.0002	0.0120
	$\xi_0 = 0.2$ and $\sigma_0 = 2$					$\xi_0 = 2$ and $\sigma_0 = 1$			
	MSE	bias	MSE	bias		MSE	bias	MSE	bias
10	0.2674	0.4366	0.3723	0.5680	10	0.5771	0.7083	0.0931	0.2441
20	0.2012	0.3824	0.1073	0.2782	20	0.2562	0.4373	0.0223	0.1256
50	0.1747	0.3252	0.0714	0.2036	50	0.2045	0.3936	0.0113	0.0887
100	0.1139	0.2823	0.0205	0.1144	100	0.1772	0.3585	0.0067	0.0645
500	0.0290	0.1273	0.0043	0.0512	500	0.0883	0.2541	0.0050	0.0199

Tab. 1: MSE and Absolute biases of the estimates of TGEVD using regression method.

Sample Size	Scale Parameter σ						Shape Parameter ξ					
	Regression	MOM	MLE	PWM	L-Moment	MPS	Regression	MOM	MLE	PWM	L-Moment	MPS
						$\xi = 0.1$, $\sigma = 2$						
10	0.0986	0.1368	0.4063	0.5141	0.4351	0.7325	0.0583	0.3496	0.3547	0.6213	0.4956	0.2260
20	0.0242	0.0739	0.2152	0.1616	0.1623	0.2468	0.0735	0.1317	0.2015	0.2344	0.2538	0.0506
50	0.0096	0.0455	0.1305	0.0934	0.1018	0.0661	0.0094	0.0507	0.1166	0.1228	0.1220	0.0109
100	0.0039	0.0340	0.0704	0.0619	0.0685	0.0284	0.0041	0.0356	0.0707	0.0752	0.0573	0.0049
500	0.0006	0.0178	0.0338	0.0086	0.0089	0.0053	0.0009	0.0142	0.0285	0.0110	0.0113	0.0007
						$\xi = 0.2$, $\sigma = 2$						
10	0.0707	0.1261	0.3427	0.4539	0.2236	0.7393	0.0299	0.2926	0.3789	0.6000	0.4693	0.2288
20	0.0219	0.1096	0.2272	0.1591	0.1598	0.2452	0.0303	0.1271	0.2103	0.2531	0.1721	0.0536
50	0.0084	0.0411	0.1206	0.0961	0.0882	0.0531	0.0085	0.0625	0.1078	0.1335	0.0636	0.0105
100	0.0036	0.0335	0.0841	0.0462	0.0563	0.0324	0.0072	0.0402	0.0749	0.0627	0.0309	0.0039
500	0.0005	0.0168	0.0339	0.0072	0.0072	0.0051	0.0009	0.0156	0.0272	0.0106	0.0074	0.0006
						$\xi = -2$, $\sigma = 1$						
10	0.1718	0.4200	0.3937	0.2360	0.1735	0.1905	0.0688	0.4215		0.6280	0.4985	0.2878
20	0.0596	0.1828	0.1930	0.1697	0.0857	0.0689	0.0651	0.2775		0.4990	0.2193	0.0888
50	0.0190	0.0429	0.1164	0.0253	0.0260	0.0200	0.0449	0.1725		0.1339	0.1162	0.0229
100	0.0075	0.0172	0.0795	0.0103	0.0107	0.0086	0.0231	0.0805		0.0576	0.0570	0.0094
500	0.0009	0.0018	0.0341	0.0016	0.0018	0.0012	0.0056	0.0115		0.0102	0.0112	0.0012
						$\xi = 2$ and $\sigma = 1$						
10	0.2114	0.4560	0.3581	0.3144	0.2521	0.1656	0.3498	0.3733		0.7933	0.7388	0.7046
20	0.1755	0.2647	0.1875	0.2307	0.2210	0.0528	0.2678	0.3621		0.6758	0.5710	0.2164
50	0.0895	0.2182	0.1171	0.0342	0.0316	0.0153	0.1709	0.1557		0.2156	0.2142	0.0614
100	0.0206	0.2289	0.0783	0.0128	0.0124	0.0070	0.0928	0.0918		0.0992	0.0950	0.0280
500	0.0051	0.1397	0.0341	0.0022	0.0022	0.0013	0.0255	0.0509		0.0190	0.0184	0.0052

Tab. 2: MSEs of the estimates of GEVD.

Sample Size	Scale Parameter σ						Shape Parameter ξ					
	Regression	MOM	MLE	PWM	L-Moment	MPS	Regression	MOM	MLE	PWM	L-Moment	MPS
						$\xi = 0.1$, $\sigma = 2$						
10	0.2374	0.2651	0.6173	1.1570	0.7136	0.7325	0.2082	0.3695	0.5847	1.2447	0.9174	0.2260
20	0.1213	0.2191	0.4571	0.3252	0.3227	0.2468	0.2489	0.2656	0.4448	0.4032	0.4086	0.0506
50	0.0778	0.1765	0.3471	0.2302	0.2418	0.0661	0.0758	0.1875	0.3362	0.2773	0.2745	0.0109
100	0.0506	0.1594	0.2654	0.1830	0.1747	0.0284	0.0506	0.1622	0.2658	0.2084	0.1997	0.0049
500	0.0199	0.1137	0.1838	0.0718	0.0425	0.0053	0.0254	0.1029	0.1687	0.0809	0.0825	0.0007
						$\xi = 0.2$ and $\sigma = 2$						
10	0.1805	0.2555	0.5710	0.9751	0.3875	0.7393	0.1530	0.3659	0.6052	1.1800	0.5027	0.2288
20	0.1193	0.2077	0.4647	0.3211	0.3181	0.2452	0.1334	0.2745	0.4484	0.4128	0.3163	0.0536
50	0.0740	0.1754	0.3456	0.2314	0.2222	0.0531	0.0736	0.2152	0.3272	0.2845	0.1954	0.0105
100	0.0483	0.1573	0.2877	0.1587	0.1718	0.0324	0.0666	0.1724	0.2714	0.1919	0.1350	0.0039
500	0.0177	0.1042	0.1841	0.0665	0.0651	0.0051	0.0248	0.1024	0.1649	0.0819	0.0672	0.0006
						$\xi = 1$ and $\sigma = 1$						
10	0.3132	0.6604	0.5977	0.3846	0.3175	0.1905	0.2308	0.6890		0.8344	0.5021	0.2878
20	0.1898	0.4152	0.4329	0.2271	0.2115	0.0689	0.2179	0.5068		0.6262	0.3738	0.0888
50	0.1082	0.1544	0.3400	0.1183	0.1190	0.0200	0.1839	0.3073		0.2806	0.2696	0.0229
100	0.0665	0.0952	0.2818	0.0772	0.0787	0.0086	0.1205	0.2080		0.1885	0.1876	0.0094
500	0.0238	0.0338	0.1846	0.0316	0.0339	0.0012	0.0610	0.0866		0.0810	0.0838	0.0012
						$\xi = 2$ and $\sigma = 1$						
10	0.3831	0.4797	0.5809	0.4085	0.4072	0.3124	0.5093	0.6220		1.0979	0.8745	0.6510
20	0.2411	0.3957	0.4641	0.2679	0.2461	0.177	0.4250	0.4626		0.8445	0.5895	0.3596
50	0.2329	0.3530	0.3407	0.1403	0.1366	0.0919	0.3361	0.3283		0.3458	0.3548	0.1845
100	0.1591	0.3415	0.2794	0.0900	0.0878	0.0623	0.2465	0.2669		0.2435	0.2402	0.1248
500	0.0551	0.2789	0.1847	0.0380	0.0377	0.0301	0.1271	0.2055		0.1093	0.1091	0.0603

Tab. 3: Absolute biases of the estimates of GEVD.

The MSE and absolute bias of estimates of GEVD evaluated using all the above mentioned techniques are tabulated in Table 2 and Table 3 respectively. Unlike MLE, the estimates obtained using regression method exists throughout the parameter space of GEVD. Also as sample size increases, the MSE and absolute bias decreases which indicates the efficiency of the regression method. We can infer that, for small samples the MSE and absolute bias of estimates evaluated using regression method are smaller when compared to that of other estimates, justifying that for small samples the regression method outperforms the other techniques. The PWM estimates and LM estimates are found to be least efficient for small samples. In the case of large samples, the estimates evaluated using regression method have smaller MSE and absolute bias than the other estimates.

In the case of TGEVD, from Table 1 it is evident that the MSE and absolute bias decreases as sample size increases. Also, the estimates calculated using regression method are found to be better for estimating scale parameter rather than shape parameter.

5. REAL DATA ANALYSIS

This section analyses a real dataset to demonstrate the computational method discussed in this study. Specifically, we examine the international price of crude oil during the Ukrainian war. The war in Ukraine has significantly impacted global economics, leading to a sharp increase in crude oil prices on the international market. Given the extreme nature of this scenario, we aim to model the price of crude oil (in US dollars) using the GEVD and TGEVD. The dataset comprises of the daily maximum prices of crude oil from January 1, 2022, to March 18, 2022, excluding weekends, taken from `https://in.investing.com/commodities/crude-oil-historical-data`. The data contains 52 values and it is given below in Table 4.

106.28 104.26, 99.21, 102.58, 109.94, 110.29, 114.87, 126.84, 129.44, 129.79, 116.10, 116.57, 112.50, 106.80, 99.13 100.53, 93.89, 94.93, 92.61, 93.35,95.01, 95.17, 95.82, 94.66, 91.74, 90.58, 91.68, 92.72, 93.17, 90.44,89.72, 88.87, 88.38, 88.84, 88.54, 87.94, 85.71, 86.08, 85.55, 87.10, 87.87, 86.62, 84.44, 82.93, 83.14, 81.58, 79.45, 80.48, 80.25, 78.73, 77.66, 76.46

Tab. 4: The daily maximum prices of crude oil in US dollars.

We now present the summary statistics of the considered data in the following Table 5.

Minimum	First Quartile	Median	Third	Maximum	Mean
76.46	86.46	92.17	101.04	129.79	95.14

Tab. 5: Descriptive statistics for the considered data.

To determine if the given dataset fits the GEVD, we performed the Kolmogorov-Smirnov (KS) goodness-of-fit test. The resulting p-value is 0.9211, and the KS statistic is 0.0736. Additionally, Figure 1 presents the empirical and theoretical cumulative distribution functions along with the P-P plot for the data fitted using GEVD. These results suggest that the GEVD provides an adequate fit for the dataset. The parameters of GEVD were estimated using the regression method, and bootstrap confidence intervals were calculated for each parameter. Moreover, for comparing the existing MLE method with the proposed method, we have evaluated the ML estimates of GEVD for the considered data. For these estimates, the p-value and KS statistic are obtained as 0.9086 and 0.0752 respectively. This shows that the estimates evaluated using the proposed regression method provides a better fit for the considered dataset. These results are summarized in Table 6.

Regression Method			Maximum Likelihood	
Parameters	Estimates	Bootstrap intervals	Parameters	Estimates
$\hat{\xi}$	-0.3002	$(-0.5210, -0.08016)$	$\hat{\xi}$	-0.1436
$\hat{\sigma}$	7.9853	$(5.8575, 9.0531)$	$\hat{\sigma}$	9.3690
$\hat{\mu}$	88.2761	$(86.3975, 89.7732)$	$\hat{\mu}$	88.6277

Tab. 6: Estimates and confidence intervals using regression method and the ML estimates for the crude oil data.

The TGEVD model is also fitted for the same data and the parameter estimates are obtained as $\hat{\xi} = -0.0649$, $\hat{\sigma} = 9.1977$ and $\hat{\mu} = 88.5821$. The corresponding p-value and KS statistic are 0.9726 and 0.0644 respectively. Comparing the goodness of fit measures of GEVD and TGEVD, it is observed that the TGEVD is a better fit for the crude oil price data. The plot between theoretical and empirical CDFs and P-P plot for the data fitted using TGEVD are given in Figure 2.

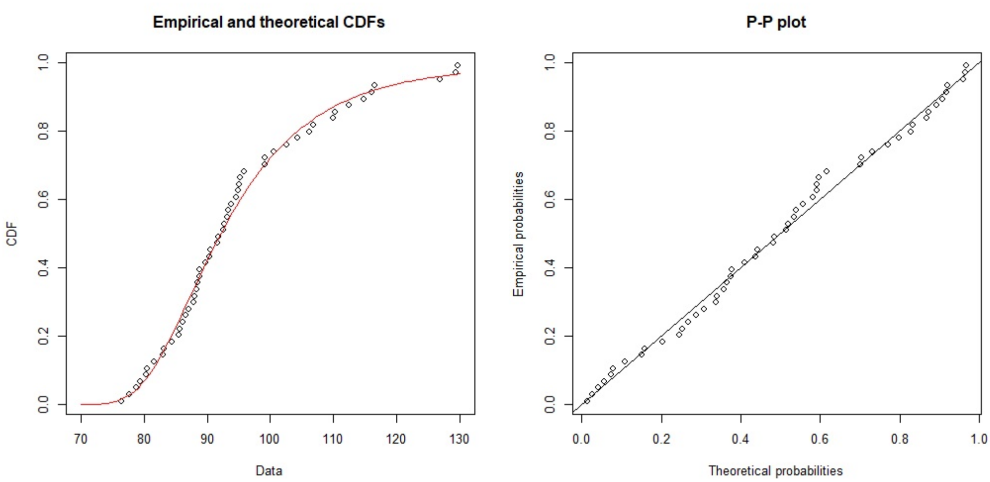

Fig. 1: Goodness of fit plots for GEVD using regression estimates.

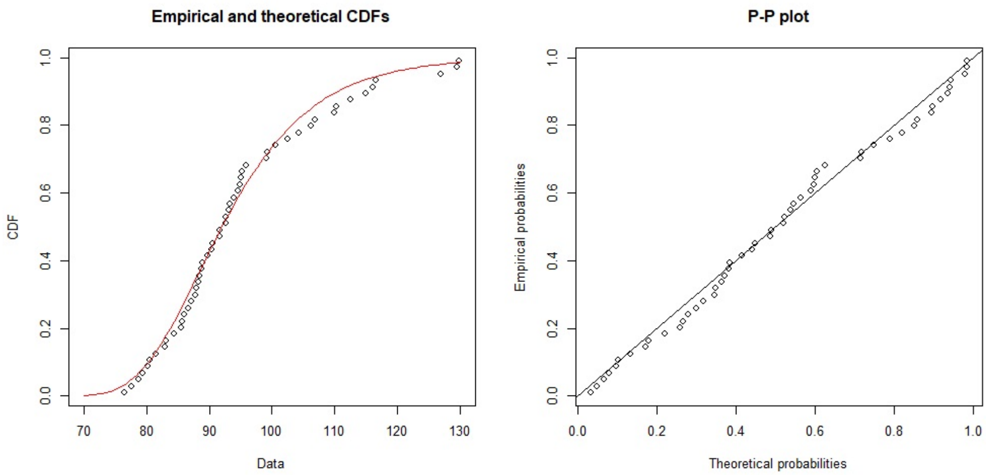

Fig. 2: Goodness of fit plots for TGEVD using regression estimates.

6. CONCLUSION

In this article, we discuss the derivation of parameter estimates for the GEVD using the regression method suggested by [27]. Initially, a regression equation is constructed from the CDF, and the scale parameter is estimated using the iterative re-weighted least squares technique. A profile likelihood is constructed using the regression equation to estimate the shape parameter. A comprehensive simulation study compares the regression method estimators with those obtained using the MOM, MLE, PWM, LM, and MPS methods. These comparisons are based on absolute bias and MSE for various sample sizes and parameter combinations. The results indicate that as the sample size increases, the absolute bias and MSE of the regression estimates for GEVD decrease, demonstrating the efficiency of this method. The regression method is found to exist throughout the parameter space. Moreover, the regression method consistently outperforms all other techniques across all considered cases, especially when the sample size is small. For large samples, the regression, LM, and MPS methods perform better than the other techniques for the considered parameter combinations. The left-truncated GEVD at zero is considered, and its hazard rate behaviour is studied, showing that the TGEVD has a decreasing hazard rate. Finally, a real dataset of crude oil prices during the Ukraine war is analysed to illustrate the applicability of the models and estimation techniques. The GEVD and TGEVD models are fitted to the dataset, and the parameter estimates are derived using the regression method. Bootstrap confidence intervals for the estimates are also constructed. Moreover, the TGEVD model is found to be a better fit for the data compared to the GEVD.

(Received June 23, 2024)

REFERENCES

[1] B. A. A. Abdulali, M. A. Abu Bakar, K. Ibrahim, and N. Mohd Ariff: Extreme value distributions: An overview of estimation and simulation. J. Probab. Statist. *1* (2022), 5449751. DOI:10.1155/2022/5449751

[2] M. A. A. Bakar, N. M. Ariff, and M. S. M. Nadzir: Comparative analysis between l-moments and maximum product spacing method for extreme pm concentration. In: International Conference on Mathematical Sciences and Statistics (ICMSS 2022), pp. 214–227.

[3] S. Baran, P. Szokol, and M. Szabó: Truncated generalized extreme value distribution based emos model for calibration of wind speed ensemble forecasts. Environmetrics *32* (2021), 6, e2678. DOI:10.1002/env.2678

[4] G. E. Box and D. R. Cox: An analysis of transformations. J. Royal Statist. Soc.: Series B (Methodological) *26* (1964), 2, 211–243. DOI:10.1111/j.2517-6161.1964.tb00553.x

[5] A. Bücher and J. Segers: On the maximum likelihood estimator for the generalized extreme-value distribution. Extremes *20* (2017), 4, 839–872. DOI:10.1007/s10687-017-0292-6

[6] S. Coles, J. Bawa, L. Trenner, and P. Dorazio: An Introduction to Statistical Modeling of Extreme Values. Springer, London 2001.

[7] B. Efron: Bootstrap methods: Another look at the jackknife. Ann. Statist. *1* (1979), 1, 1–26. DOI:10.1214/aos/1176344552

[8] B. Efron: Nonparametric standard errors and confidence intervals. Canadian J. Statist. *9* (1981), 2, 139–158. DOI:10.2307/3314608

[9] B. Efron: Better bootstrap confidence intervals. J. American Statis. Assoc. *82* (1987), 397, 171–185. DOI:10.1080/01621459.1987.10478410

[10] P. Embrechts, C. Klüppelberg, and T. Mikosch: Modelling Extremal Events for Insurance and Finance. Springer, Berlin 1997.

[11] R. A. Fisher and L. H. C. Tippett: Limiting forms of the frequency distribution of the largest or smallest member of a sample. Math. Proc. Cambridge Philosoph. Soc. *24* (1928), 2, 180–190. DOI:10.1017/S0305004100015681

[12] M. Fréchet: Sur la loi de probabilité de l'écart maximum. Ann. Soc. Math. Polon. *6* (1927), 93–116.

[13] B. Gnedenko: Sur la distribution limite du terme maximum d'une serie aleatoire. Ann. Math. *44* (1943), 2, 423–453.

[14] J. R. Hosking: Algorithm as 215: Maximum-likelihood estimation of the parameters of the generalized extreme-value distribution. J. Royal Statist. Soc.: Series C (Applied Statistics) *34* (1985), 3, 301–310. DOI:10.2307/2347484

[15] J. R. Hosking: L-moments: Analysis and estimation of distributions using linear combinations of order statistics. J. Royal Statist. Soc.: Series B (Methodological) *52* (1990), 1, 105–124. DOI:10.1111/j.2517-6161.1990.tb01775.x

[16] J. R. M. Hosking, J. R. Wallis, and E. F. Wood: Estimation of the generalized extreme-value distribution by the method of probability-weighted moments. Technometrics *27* (1985), 3, 251–261. DOI:10.1080/00401706.1985.10488049

[17] A. F. Jenkinson: The frequency distribution of the annual maximum (or minimum) values of meteorological elements. Quarterly J. Royal Meteorolog. Soc. *81* (1955), 348, 158–171. DOI:10.1002/qj.49708134804

[18] M. R. Leadbetter, G. Lindgren, and H. Rootzén: Extremes and Related Properties of Random Sequences and Processes. Springer, New York 1983.

[19] H. Madsen, P. F. Rasmussen, and D. Rosbjerg: Comparison of annual maximum series and partial duration series methods for modeling extreme hydrologic events: 1. At-site modeling. Water Resources Res. *33* (1997), 4, 747–757. DOI:10.1029/96WR03848

[20] E. S. Martins and J. R. Stedinger: Generalized maximum-likelihood generalized extreme-value quantile estimators for hydrologic data. Water Resources Res. *36* (2000), 3, 737–744. DOI:10.1029/1999WR900330

[21] J. Pickands: Statistical inference using extreme order statistics. Ann. Statist. *3* (1975), 1, 119–131. DOI:10.1214/aos/1176343003

[22] P. Prescott and A. Walden: Maximum likelihood estimation of the parameters of the generalized extreme-value distribution. Biometrika *67* (1980), 3, 723–724. DOI:10.1093/biomet/67.3.723

[23] S. I. Resnick: Extreme Values, Regular Variation, and Point Processes. Springer, New York 1987.

[24] M. L. Rizzo: Statistical Computing with R. Chapman and Hall/CRC, New York 2019.

[25] R. L. Smith: Maximum likelihood estimation in a class of nonregular cases. Biometrika *72* (1985), 1, 67–90. DOI:10.1093/biomet/72.1.67

[26] J. R. Stedinger, R. M. Vogel, and E. Foufoula-Georgiou: Frequency analysis of extreme events. In: Handbook of Hydrology (D. R. Maidment, ed.), *18* (1993), McGraw-Hill, New York, pp. 18.1–18.66.

[27] J. M. Van Zyl: A median regression model to estimate the parameters of the three-parameter generalized pareto distribution. Commun. Statist.-Simul. Comput. *41* (2012), 4, 544–553. DOI:10.1080/03610918.2011.595868

[28] R. Von Mises: La distribution de la plus grande de n valuers. Rev. Math. Union Interbalcanique *1* (1936), 141–160.

[29] T. Wong and W. Li: A note on the estimation of extreme value distributions using maximum product of spacings. Inst. Math. Statist. Lect. Notes – Monogr. Ser. *55* (2006), 272–283. DOI:10.3323/jcorr.55.283

[30] A. Yılmaz, M. Kara, and O. Özdemir: Comparison of different estimation methods for extreme value distribution. J. Appl. Statist. *48* (2021), 13–15, 2259–2284. DOI:10.1080/02664763.2021.1940109

[31] A. Zellner: An Introduction to Bayesian Inference in Econometrics. Wiley, New York 1996.

[32] L. Zhang and B. A. Shaby: Uniqueness and global optimality of the maximum likelihood estimator for the generalized extreme value distribution. Biometrika *109* (2022), 3, 853–864. DOI:10.1093/biomet/asab043

R. Anand, Research Scholar, Department of Statistics, University of Calicut, Thenhipalam, Kerala. India.
e-mail: anand94.r@gmail.com

C. Chandran, Professor, Department of Statistics, University of Calicut, Thenhipalam, Kerala. India.
e-mail: cheruvalathchandran@gmail.com