Hindawi

Research Article

Shrinkage Methods for Estimating the Shape Parameter of the Generalized Pareto Distribution

Wilhemina Adoma Pels ⓘ,[1] **Atinuke O. Adebanji** ⓘ,[1] **Sampson Twumasi-Ankrah** ⓘ,[1] **and Richard Minkah** ⓘ[2]

[1]*Department of Statistics and Actuarial Science, Kwame Nkrumah University of Science and Technology, Ghana*
[2]*Department of Statistics and Actuarial Science, University of Ghana, Ghana*

Correspondence should be addressed to Wilhemina Adoma Pels; wilheminapels2@gmail.com

The generalized Pareto distribution is one of the most important distributions in statistics of extremes as it has wide applications in fields such as finance, insurance, and hydrology. This study proposes two new methods for estimating the shape parameter of the generalized Pareto distribution (GPD). The proposed methods use the shrinkage principle to adapt the existing empirical Bayesian with data-based prior and the likelihood moment method to obtain two estimators. The performance of the proposed estimators is compared with the existing estimators (i.e., maximum likelihood, likelihood moment estimators, etc.) for the shape parameter of the generalized Pareto distribution in a simulation study. The results show that the proposed estimators perform better for small to moderate number of exceedances in estimating shape parameter of the light-tailed distributions and competitive when estimating heavy-tailed distributions. The proposed estimators are illustrated with practical datasets from climate and insurance studies.

1. Introduction

The generalized Pareto distribution (GPD) is one of the fundamental distributions used in extreme value theory. As reported in [1, 2], the GPD is the limiting distribution of excesses over a large threshold regardless of the underlying distribution of data, F. The significance of the GPD in extreme value theory cannot be overemphasized because of its extensive applicability: hydroelectric dam management [3], flood levels of river, insurance and finance [4, 5], waiting time problems [6], ecology [7], and climatology [8], among others.

Let $Y_1, Y_2, \cdots, Y_n$ be a random sample from Y with unknown underlying distribution function, F. An exceedance occurs over a deterministic threshold, u, if the value of Y is greater than u, i.e., $Y = Y_j | Y_i > u, j = 1, \cdots, n_u ; i = 1, \cdots, n$, where n_u is the number of observations exceeding u. Also, the excess over a threshold is defined as $Y - u$ with $Y > u$.

The excess distribution over threshold is given by

$$F_u(y) = P(Y - u \le y | Y > u) = \frac{F(y + u) - F(u)}{1 - F(u)}, 0 \le y \le y_0 - u,$$

(1)

where y_0 is the finite or infinite right endpoint of the distribution F. From [1, 2], the distribution of Y is well approximated by the GPD with the survival function defined as

$$1 - F(y) = \begin{cases} (1 + \varepsilon y)^{-1/\varepsilon}, & 1 + \varepsilon y > 0, y \ge 0, & \text{if} \quad \varepsilon \ne 0, \\ \exp(-y), & y \ge 0, & \text{if} \quad \varepsilon = 0. \end{cases}$$

(2)

If the location $(\mu(\mu \in \mathbb{R}))$, shape $(\varepsilon(\varepsilon \in \mathbb{R}))$, and scale $(\sigma(\sigma > 0))$ parameters are included (see, e.g., [9]), the resulting

three-parameter generalized Pareto distribution is given by

$$1 - F(y) = \begin{cases} \left(1 + \varepsilon\left(\dfrac{y-\mu}{\sigma}\right)\right)^{-1/\varepsilon}, & 1 + \varepsilon\left(\dfrac{y-\mu}{\sigma}\right), y \geq 0, \quad \text{if} \quad \varepsilon \neq 0, \\[2mm] \exp\left(-\dfrac{y-\mu}{\sigma}\right), & y \geq 0, \quad \text{if} \quad \varepsilon = 0. \end{cases}$$

$$(3)$$

The shape parameter, ε, is usually referred to as the tail index or the extreme value index [3, 10, 11]. The mean and the variance of a GPD random variable, Y, are given by

$$E(Y) = \frac{\sigma}{1 - \varepsilon},$$
$$\mathrm{Var}(Y) = \frac{\sigma^2}{\left[(1-\varepsilon)^2(1-2\varepsilon)\right]},$$

$$(4)$$

respectively. Here, the expected value of Y exists when $\varepsilon < 1$ and the variance when $\varepsilon < 1/2$.

The shape (tail index) parameter shows the tail heaviness of the underlying distribution and can be used to classify the distribution into three classes. These are short-tailed when $\varepsilon < 0$, medium-tailed when $\varepsilon = 0$, and heavy-tailed when $\varepsilon > 0$. In special cases, $\varepsilon = 0$ and $\varepsilon = 1$, the GPD reduces to exponential and uniform distributions, respectively.

Several estimation techniques have been proposed for estimating the parameters of the GPD such as maximum likelihood (ML), probability weighted moments, and elemental percentile. For large sample sizes, [12] showed that the ML estimator (MLE) is asymptotically the best. However, the MLE is known to pose computational difficulties and convergence problems [13, 14]. In addition, it does not exist when $\varepsilon > 1$ [15] and provides nonfeasible estimates if $\varepsilon < -0.5$ [13, 16].

To address these problems of the MLE, [15] proposed the likelihood moment estimators (LME). This estimator shows high efficiency for small sample sizes. However, [17] reports that it performs poorly if $\varepsilon \geq 2$ and tends to overestimate the shape parameter for larger values of the scale parameter. Also, the LME again is computationally intensive and slow. Hence, [18, 19] proposed an empirical Bayesian method to solve the computational intensiveness of the LME. The resulting estimators of the tail index from this method showed better performance for $\varepsilon \in [-0.5,1]$. Despite this advantage, the authors use all available data in the computation of the estimates of the parameters of the GPD. However, it is known that for GPD-based models, the inclusion of small to moderate observations introduces bias [11, 20].

In this paper, we propose two estimators of the tail index of the generalized Pareto distribution using the idea of the shrinkage principle. The estimators are a combination of the likelihood-based approach and that of the empirical Bayesian methods. The performance of the proposed method is evaluated in conjunction with the existing estimators in a simulation.

The rest of the paper is organized as follows. Review of the existing estimators and proposed two new estimators are in Section 2. The performances of the estimators are examined via an extensive simulation study in Section 3. A practical application of these estimators on two datasets is demonstrated. Finally, Section 5 provides concluding remarks.

2. Estimation of the Parameters of the Generalized Pareto Distribution

Assuming F is a GPD with scale and shape parameters σ and ε, respectively, the following methods can be used to obtain the estimators of the parameters, maximum likelihood [13], likelihood moment [15], profile likelihood with empirical Bayesian [18], and improved profile likelihood with empirical Bayesian [19] methods. In the next subsections, we present a brief description of each of these methods.

2.1. Maximum Likelihood Estimator (MLE). The maximum likelihood principle involves finding the set of parameters that maximize the likelihood function $L(Y, \theta)$ evaluated on the sample from the distribution, F, of the random variable, Y:

$$L(Y; \varepsilon, \sigma) = \prod_{i=1}^{n} f(y_i),$$

$$(5)$$

where f is the density function associated with F.

Many studies have shown that the estimators based on the maximum likelihood estimator are usually better in the case of large samples (see, e.g., [11, 12, 21, 22]), among others. In addition, its asymptotic properties including normality are easily found compared with other estimation methods.

In the case of statistics of extremes, and in particular the GPD, [11, 13, 21], among others, show that for small sample sizes, the maximum likelihood performs poorly. Also, the numerical algorithm may fail to converge to the maxima. [21] suggests reparametrization of (ε, σ) to (ε, θ) of which θ is of the form ε/σ and $\widehat{\sigma}$ is estimated as $\widehat{\sigma} = \widehat{\varepsilon}/\widehat{\theta}$, with

$$l^*(\theta, \varepsilon) = n \log\left(\frac{\theta}{\varepsilon}\right) + \left(\frac{1}{\varepsilon} - 1\right)\sum_{i=1}^{n} \log(1 - \theta Y_i),$$
$$1 - \varepsilon = \frac{n}{\sum_{i=1}^{n}(1 - \theta Y_i)^{-1}}.$$

$$(6)$$

Hence,

$$\varepsilon = n^{-1}\sum_{i=1}^{n} \log(1 - \theta Y_i),$$

$$(7)$$

$$l(\theta) = 1 - \left(\frac{n}{\sum_{i=1}^{n}(1 - \theta Y_i)^{-1}}\right) + n^{-1}\sum_{i=1}^{n} \log(1 - \theta Y_i) = 0, \quad \theta < 1/Y_{n,n},$$

$$(8)$$

where $Y_{1,n} \leq \cdots \leq Y_{n,n}$ are the order statistics associated with Y. Maximising (8) leads to the MLE estimator $\widehat{\theta}_{\mathrm{MLE}}$, and the corresponding MLE of σ and ε are obtained as

$$\widehat{\varepsilon}_{MLE} = -n^{-1} \sum_{i=1}^{n} \log \left(1 - \widehat{\theta}_{MLE} Y_i\right),$$

$$\widehat{\sigma}_{MLE} = \frac{\widehat{\varepsilon}_{MLE}}{\widehat{\theta}_{MLE}}. \tag{9}$$

With this reparametrization, the MLE becomes asymptotically efficient. However, the computational issues persist such as nonconvergence and, thus, the search for solutions including the likelihood moment estimation technique.

2.2. Likelihood Moment Estimator. The likelihood moment estimator (LME) was proposed by [15] to solve the computational complexities, efficiency, and convergence problems of the MLE. LME is reported to be efficient and robust, yet it is computationally intensive. The LME is estimated with the log-likelihood

$$l^*(\theta, \varepsilon) = n \log (\theta/\varepsilon) + (1/\varepsilon - 1) \sum_{i=1}^{n} \log (1 - \theta Y_i), \tag{10}$$

and the moment estimators as

$$\widehat{\varepsilon}_{mom} = \frac{\bar{Y}^2/s^2 - 1}{2},$$

$$\widehat{\sigma}_{mom} = \frac{\bar{Y}\left(\bar{Y}^2/s^2 + 1\right)}{2}. \tag{11}$$

It should be noted that for any constant r satisfying $1 + r\varepsilon > 0$, the $E((1 - \theta Y)^r) = (1 + r\varepsilon)^{-1}$ of which the sample version is equivalent to

$$\left(n^{-1} \sum_{i=1}^{n} (1 - \theta Y_i)^r - (1 + r\varepsilon)^{-1}\right)/r = 0. \tag{12}$$

Using some conditions on ε and r for the existence of the MLE estimator of the tail index, [15] obtains

$$\widehat{\varepsilon}_{LME} = -n^{-1} \sum_{i=1}^{n} \log \left(1 - \widehat{\theta}_{LME} Y_i\right),$$

$$\widehat{\sigma}_{LME} = \frac{\widehat{\varepsilon}_{LME}}{\widehat{\theta}_{LME}},$$

$$\widehat{\theta}_{LME} = \frac{1}{n} \sum_{i=1}^{n} (1 - Y_i)^p - (1 - r)^{-1}. \tag{13}$$

Here, $p = (rn)/(\sum_{i=1}^{n} \log (1 - \theta Y_i)), r < 1$ and $\theta = \varepsilon/\sigma$. [15] depicts that when $r = \varepsilon$, the asymptotic variances of LME and MLE are the same. Even though the choice of threshold is an important issue in the estimation of the GPD, it was not considered by the authors. Also, [23] states that the LME is less sensitive than other estimators to the choice of threshold. In general, the LME estimator performs

poorly if $\varepsilon \geq 2$ and tends to overestimate the shape parameter for large values of the scale parameter [17].

2.3. The Profile Likelihood with Empirical Bayesian Method (PB). This estimation method was proposed by [18] to improve the MLE and also solve the computational difficulties of the LME method. It borrows ideas from empirical Bayesian method where ε and σ are reparameterized as ε and θ, of which θ is of the form ε/σ.

The authors defined their estimator as

$$\widehat{\theta}_{PB} = \frac{\int \theta \cdot \pi(\theta) L(\theta) d(\theta)}{\int \pi(\theta) L(\theta) d(\theta)}, \tag{14}$$

where $l(\theta) = n[\log (\theta/\varepsilon) + \varepsilon - 1]$ is the profile log-likelihood function and $\pi(\cdot)$ is a data-driven "prior" density function for θ.

The integral in (14) is computationally difficult for most priors; hence, a simplified numerical version was proposed as

$$\widehat{\theta}_{PB} = \sum_{j=1}^{m} w(\theta_j) \cdot \theta_j, \tag{15}$$

where

$$\theta_j = \frac{1}{Y_{(n)}} + \frac{\left[1 - \left(\sqrt{m}/j - 0.5\right)\right]}{3Y^\star}, \tag{16}$$

with $Y^\star = Y_{[(n/4)+0.5]}$ as the first quartile of the sample data, $m = 20 + [\sqrt{n}], j = 1, \cdots, m$ and $w(\theta_j) = 1/\sum_{t=1}^{m} e^{l(\theta_t) - l(\theta_j)}$. The shape and scale parameters are obtained as

$$\widehat{\varepsilon} = -n^{-1} \sum_{i=1}^{n} \log \left(1 - \widehat{\theta}_{PB} Y_i\right),$$

$$\widehat{\sigma} = \frac{\widehat{\varepsilon}_{PB}}{\widehat{\theta}_{PB}}. \tag{17}$$

However, this method is reported to be sensitive to the shape of the prior distribution but performs well when $-1 \leq \varepsilon \leq 0.5$. Also, it is reported to have very poor performance for extremely heavy-tailed distributions and efficient for only small sample sizes [19].

Since this estimation method is in the peaks-over threshold framework, the selection of a suitable threshold is critical as it leads to bias-variance trade-offs. However, the authors did not consider this in their simulation study.

2.4. Improved Profile Likelihood with Empirical Bayesian Method (IPB). [19] proposed a refinement of the estimator in [18] with the aim of obtaining improved estimators of the tail index in the case of heavy-tailed distributions. In order to achieve this, a better data-driven prior distribution was proposed.
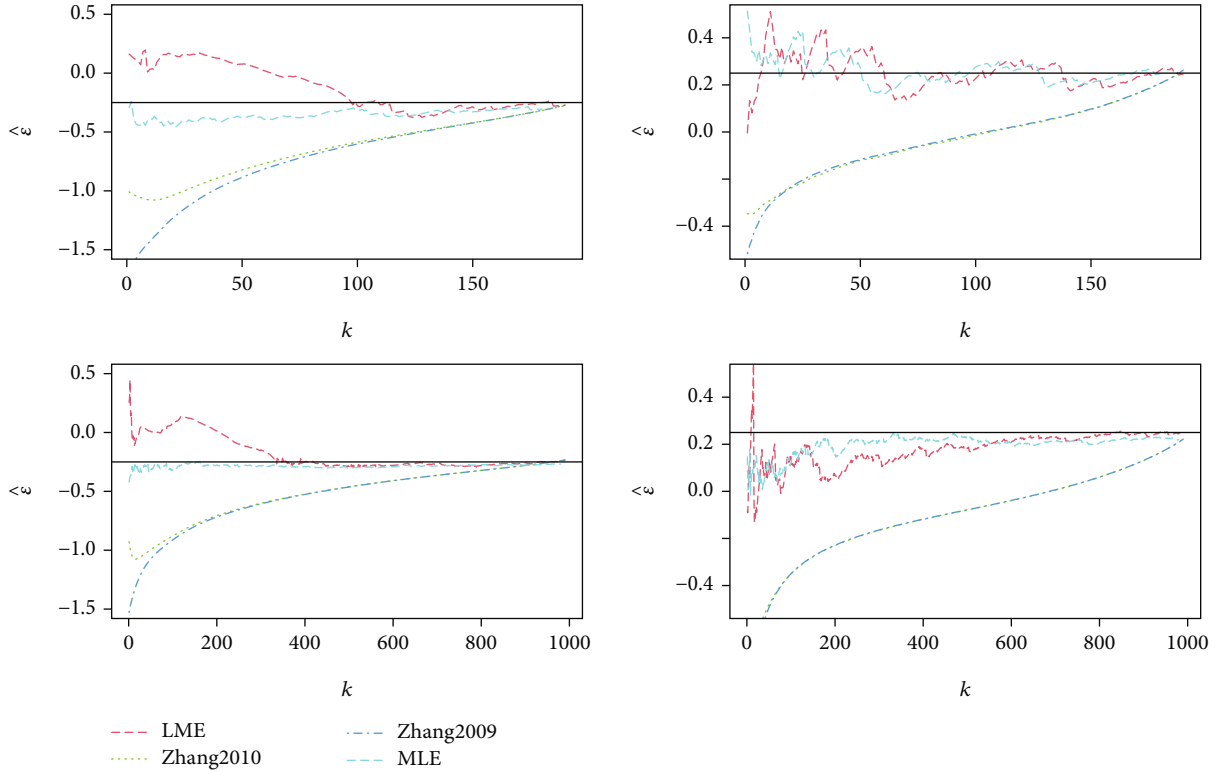
FIGURE 1: Estimators of ε. Leftmost column: $\varepsilon = -0.25$ and rightmost column: $\varepsilon = 0.25$. Topmost row: $n = 200$ and bottommost row: $n = 1000$. The solid line is the exact value of ε.

In this case, a GPD with scale $\sigma^\star$ and shape $\varepsilon^\star$ was chosen as the prior distribution for

$$\frac{n-1}{n+1} Y_{(n)}^{-1} - \theta. \tag{18}$$

Here, $\varepsilon^\star$ is chosen as negative to ensure that $Y_{(n)}^{-1}$ is positive and finite.

Similar to the estimator in Section 2.3, the modified estimators of ε and σ are obtained as

$$\widehat{\varepsilon}_{\text{IPB}} = -n^{-1} \sum_{i=1}^{n} \log\left(1 - \widehat{\theta}_{\text{IPB}} Y_i\right),$$
$$\widehat{\sigma}_{\text{IPB}} = \frac{\widehat{\varepsilon}_{\text{IPB}}}{\widehat{\theta}_{\text{IPB}}}. \tag{19}$$

Here,

$$\widehat{\theta}_{\text{IPB}} = \sum_{j=1}^{m} w\left(\theta_j^*\right) \cdot \theta_j^*, \tag{20}$$

with the computation of θ_j^* as

$$\frac{n-1}{n+1} Y_{(n)}^{-1} - \frac{\sigma^*}{\varepsilon^*}\left[1 - \left(\frac{j-0.5}{m}\right)^\varepsilon\right], j = 1, \cdots, m. \tag{21}$$

However, this method is reported to be sensitive to the

estimation of the scale parameter and has poor performance for light-tailed distributions [9, 24]. In addition, this estimator is not valid for $\varepsilon < 0.5$ [22]. Furthermore, the authors did not consider the effect of the number of exceedances, k, on their proposed estimator.

2.5. Empirical Review of Some Existing Estimation Methods. In [18, 19], it has been demonstrated numerically that their estimators of the tail index are less biased and efficient compared with existing estimators such as the maximum likelihood and likelihood moments. However, a closer look at the codes implemented for the simulation study reveals that the number of exceedances (k) is taken as the sample size (i.e., the whole sample is utilised in the estimation of the parameters). This is in contrast to the concepts in extreme value theory relating to the peaks-over threshold (POT) where emphasis is on the tails of the distributions. Therefore, in such a simulation study, estimators should be assessed on their sensitivity to the choice of top order statistics, i.e., number of exceedances. In view of the above, we present a simulation study to assess the performance of the proposals in [18, 19] with that of the MLE and the LME of tail index as a function of the number of top order statistics or exceedances.

We generate samples of size $n = 200$ and 1000 from the GPD with parameters, $\{\varepsilon = -0.25, \sigma = 1\}$ and $\{\varepsilon = 0.25, \sigma = 1\}$. Figure 1 shows the behaviour of the estimators arising from these estimators of the tail index of the underlying distributions of the generated datasets.

In the case of the estimation of $\varepsilon = -0.25$, LME tends to overestimate (show large deviations) the tail index when k is small. Also, PB and IPB methods underestimate the value of the tail index for small values of k and are only closer to the true value as $k \longrightarrow n$. In addition, for positive values of the shape parameter, the LME gives very good estimates, i.e., closer to the true value. However, the PB and IPB estimators give good estimates of ε only if $k \longrightarrow n$. Therefore, the estimators, PB [18] and IPB [19], do not perform well when the number of exceedances is small compared to the sample size.

Even though the [18, 19] estimators (i.e., PB and IPB) show attractive properties in the estimation of the extreme value index, this small simulation study gives an indication of their sensitivities to the choice of k. Therefore, in this study, we aim to provide estimators of the tail index that takes into account the advantages of the [18, 19] estimators but are less sensitive to the choice of the number of exceedances, k.

2.6. The Proposed Method. This estimation technique seeks to improve the methods proposed by [15] and that of [18, 19]. The proposed method uses the idea of shrinkage estimation which relies on a weighted combination of these methods.

For convenience, we reparametrize σ and ε of the GPD as θ and ε, where $\theta = \varepsilon\sigma^{-1}$ as done by [21]. This implies that the shape and scale parameter are estimated, respectively, by $\widehat{\varepsilon}$ and $\widehat{\sigma} = \widehat{\varepsilon}\widehat{\theta}^{-1}$ by maximizing the log-likelihood for the sample

$$l^{\star}(\theta, \varepsilon) = n \log \left(\frac{\theta}{\varepsilon} \right) + \left(\frac{1}{\varepsilon} - 1 \right) \sum_{i=1}^{n} \log \left(1 - \theta Y_i \right). \quad (22)$$

For $\nu \in [0, 1]$, the first and second proposed shrinkage estimators of θ are given by

$$\begin{aligned} \widehat{\theta}_1^{(s)} &= \nu\widehat{\theta}_{\mathrm{PB}} + (1 - \nu)\widehat{\theta}_{\mathrm{LME}}, \\ \widehat{\theta}_2^{(s)} &= \nu\widehat{\theta}_{\mathrm{IPB}} + (1 - \nu)\widehat{\theta}_{\mathrm{LME}}, \end{aligned} \quad (23)$$

respectively. The value of ν plays an important role in the shrinkage estimator: $\nu \longrightarrow 0$ implies that the LME dominates whereas with $\nu \longrightarrow 1$, the empirical Bayesian methods (PB and IPB) dominate. Also, when $\nu \longrightarrow 0.5$, equal proportion of both estimators is included in the proposed estimators of the tail index.

Therefore,

$$\begin{aligned} \widehat{\varepsilon} &= -n^{-1} \sum_{i=1}^{n} \log \left(1 - \widehat{\theta}_j^{(s)} Y_i \right), \quad j \in 1, 2, \\ \widehat{\sigma} &= \frac{\widehat{\varepsilon}}{\widehat{\theta}_j^{(s)}}. \end{aligned} \quad (24)$$

Now, the MSE of $\widehat{\theta}_j^{(s)}$ is computed as

$$\begin{aligned} \mathrm{MSE}\left(\widehat{\theta}_j^{(s)}, \theta \right) &= \mathbb{E}\left[\nu\widehat{\theta}_q + (1 - \nu)\widehat{\theta}_{\mathrm{LME}} - \theta \right]^2, \quad q \in \theta_{\mathrm{PB}}, \theta_{\mathrm{IPB}} \\ &= \mathbb{E}\left[\nu\left(\widehat{\theta}_q - \widehat{\theta}_{\mathrm{LME}} \right) - \left(\theta - \widehat{\theta}_{\mathrm{LME}} \right) \right]^2 \\ &= \mathbb{E}\left[\nu^2\left(\widehat{\theta}_q - \widehat{\theta}_{\mathrm{LME}} \right)^2 + \left(\theta - \widehat{\theta}_{\mathrm{LME}} \right)^2 - 2\nu\left(\widehat{\theta}_q - \widehat{\theta}_{\mathrm{LME}} \right)\left(\theta - \widehat{\theta}_{\mathrm{LME}} \right) \right] \\ &= \nu^2\mathbb{E}\left(\widehat{\theta}_q - \widehat{\theta}_{\mathrm{LME}} \right)^2 + \mathbb{E}\left(\theta - \widehat{\theta}_{\mathrm{LME}} \right)^2 - 2\nu\mathbb{E}\left(\widehat{\theta}_q - \widehat{\theta}_{\mathrm{LME}} \right)\left(\theta - \widehat{\theta}_{\mathrm{LME}} \right). \end{aligned} \quad (25)$$

For optimal weight, we minimise the $\mathrm{MSE}(\widehat{\theta}_j^{(s)}, \theta)$ with respect to ν:

$$\begin{aligned} \frac{\partial \mathrm{MSE}\left(\widehat{\theta}_q, \theta \right)}{\partial \nu} &= 2\nu\mathbb{E}\left(\widehat{\theta}_q - \widehat{\theta}_{\mathrm{LME}} \right)^2 - 2\mathbb{E}\left(\widehat{\theta}_q - \widehat{\theta}_{\mathrm{LME}} \right)\left(\theta - \widehat{\theta}_{\mathrm{LME}} \right) = 0, \\ \nu &= \frac{\mathbb{E}\left(\widehat{\theta}_q - \widehat{\theta}_{\mathrm{LME}} \right)\left(\theta - \widehat{\theta}_{\mathrm{LME}} \right)}{\mathbb{E}\left(\widehat{\theta}_q - \widehat{\theta}_{\mathrm{LME}} \right)^2}. \end{aligned} \quad (26)$$

The asymptotic properties of $\widehat{\theta}_q$ (i.e., $\widehat{\theta}_{\mathrm{PB}}$ and $\widehat{\theta}_{\mathrm{IPB}}$) remain an open problem, and hence, we resort to the use of simulation to find an approximate value for ν. An extensive simulation was done to choose the value of ν, and in practice, the findings show that for light-tailed situations, $\nu \in [0.4, 0.7]$ provides an appropriate or suitable estimates whereas for heavy-tailed cases, $\nu \in (0, 0.2]$ gives an optimal estimates. We provide a few examples as shown in Figure 2. The rest are available on request from the authors.

3. Simulation Study

In this section, a simulation study is conducted to compare the performance of the tail index estimation methods discussed in the previous section. Specifically, the two proposed estimators are compared with the MLE [13], LME [15], PB [18], and IPB [19]. We present the simulation design and a discussion of the results in the next subsubsections.

3.1. Simulation Design. Samples of different sizes are generated from the GPD with various values of the shape and scale parameters. The four existing estimators, MLE, LME, PB, and IPB, and the two proposed estimators, Proposed 1 and Proposed 2, are used to estimate the values of the parameters of the GPD. Two error metrics are used for comparison of the estimators, i.e., mean square error (MSE) and bias.

Algorithm 1 outlines the procedure for comparing the performance of estimators.

We remark that, from Step 2, the estimation ε leads to concurrent estimates of σ. However, our interest is the tail index (i.e., shape) parameter only.

3.2. Simulation Results and Discussion. The simulation is implemented in the statistical package R. The codes for IPB and PB are obtained from [18, 22], respectively. Also, the codes for LME can be found in the R package POT.
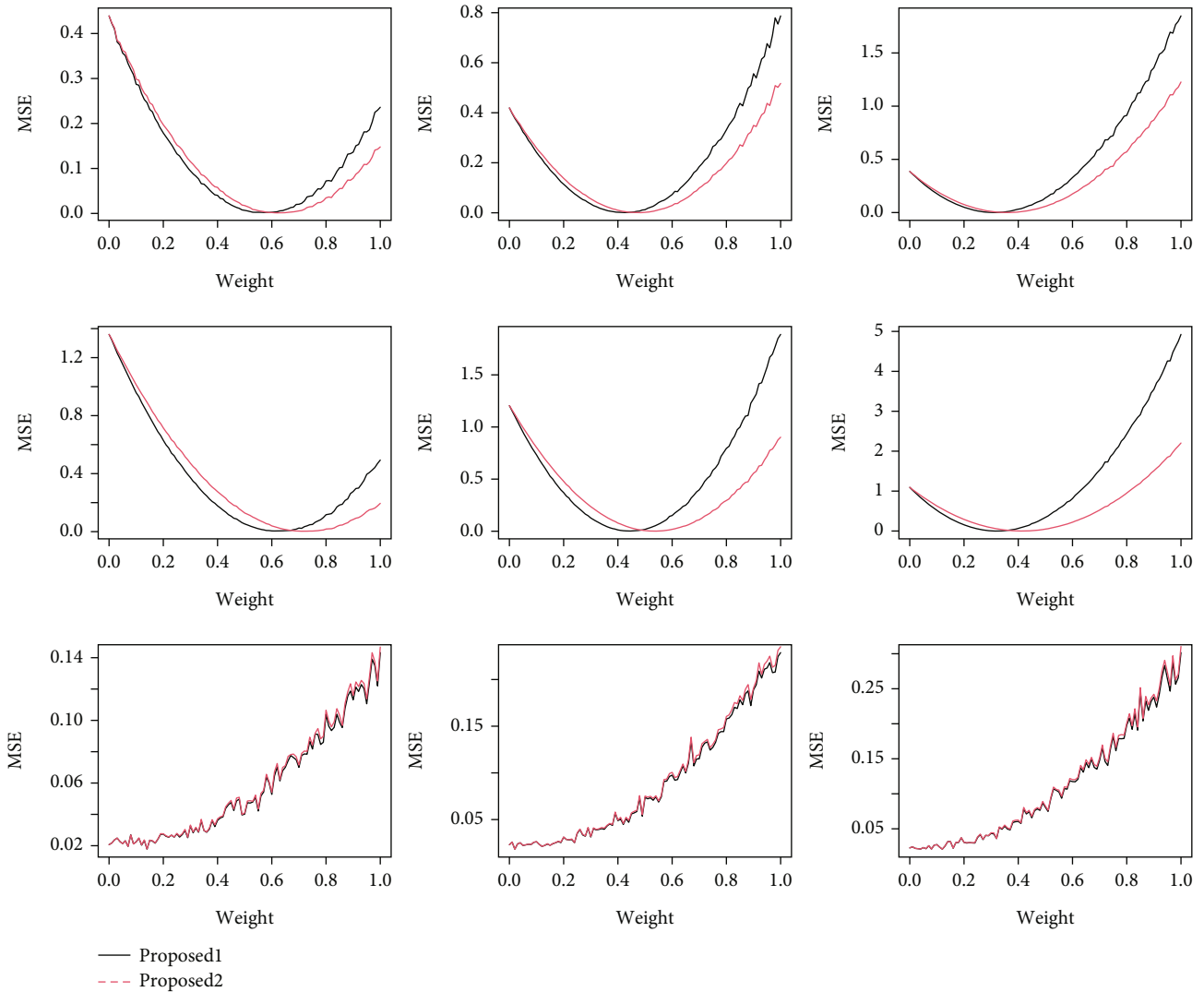
FIGURE 2: Weight with respect to MSE. First row: $\varepsilon = -0.5$; second row: $\varepsilon = -1$; and third row: $\varepsilon = 0.5$. Leftmost column: $n = 100$; middlemost column: $n = 200$; and rightmost column: $n = 500$.

Step 1. Generate a sample of size $n(n = 50,200,500)$ from a GPD with known shape (ε) and scale (σ) parameters such that both light and heavy-tailed distributions are realized.

Step 2. For each number of top order statistic k $(k = 5, \cdots, n - 1.)$ or equivalently threshold value u, compute estimates of $\varepsilon_{i,k}$ where $i \in$ {MLE, LME, PB, IPB, Proposed 1 and Proposed 2}.

Step 3. In order to compute the bias and mean square error for each estimator, repeat Steps 1 and 2 a large number of times, $R(R \geq 1000)$ to obtain the estimates, $\varepsilon_{i,k,r}$ with $r = 1, \cdots, R$.

Step 4. Compute the bias
$$bias\left(\widehat{\varepsilon}_{i,k,r}, \varepsilon\right) = 1/R \sum_{r=1}^{R} \widehat{\varepsilon}_{i,k,r} - \varepsilon$$
and Mean Square Error(MSE)
$$MSE\left(\widehat{\varepsilon}_{i,k,r}, \varepsilon\right) = 1/R \sum_{r=1}^{R} \left(\widehat{\varepsilon}_{i,k,r} - \varepsilon\right)^2$$
at each k.

ALGORITHM 1: Algorithm for comparing the performance of the GPD estimators.

The results of the simulation on the estimation of the shape parameter, ε, are presented in Figures 3 and 4 for light and heavy-tailed distributions, respectively.

First, the results show that the proposed methods generally have smaller MSE for small values of k (i.e., k less than 40% of the sample size). As $k \longrightarrow n$, PB and
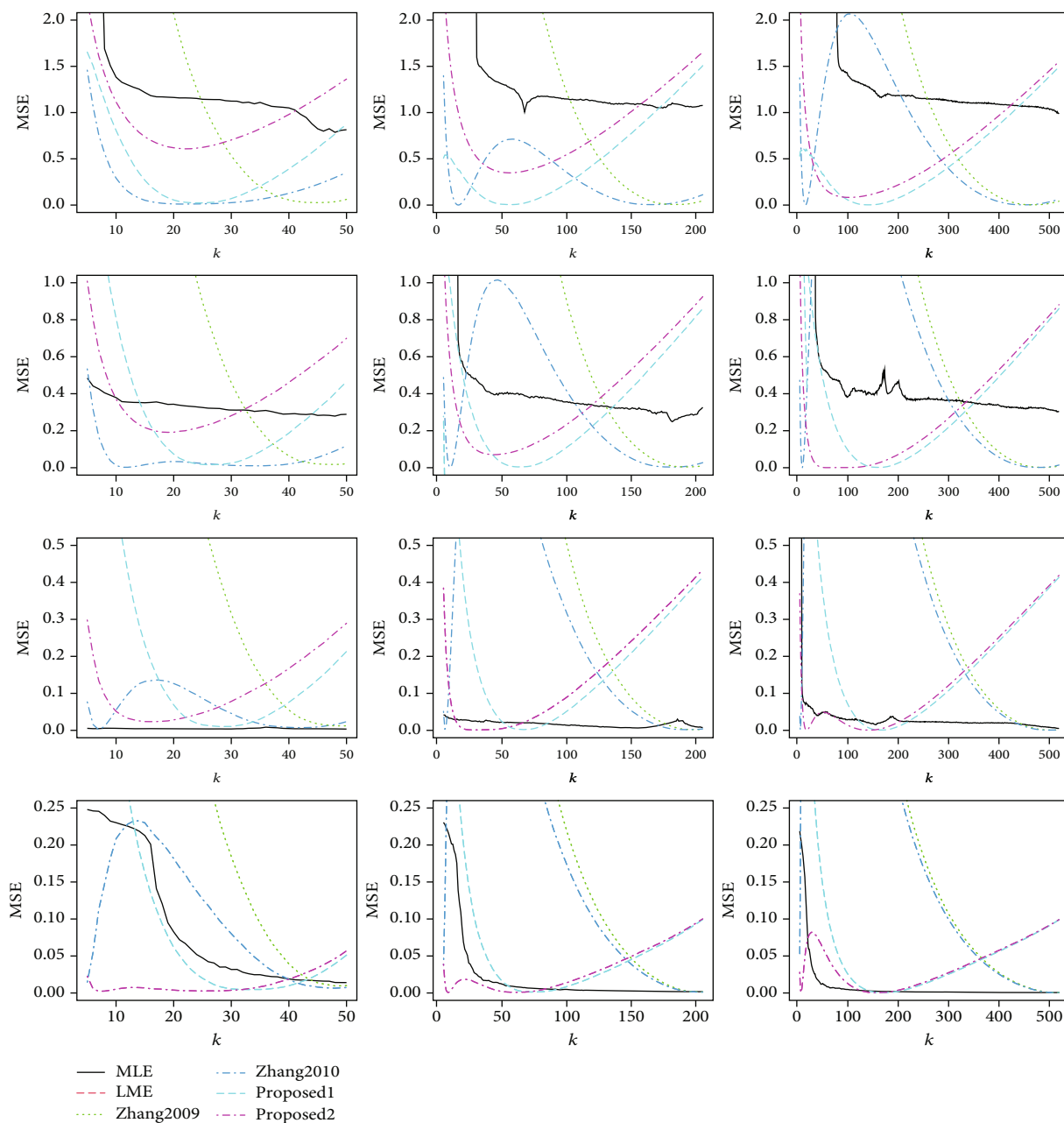
FIGURE 3: MSE($\hat{\varepsilon}$) with $\sigma = 1$. First row: $\varepsilon = -2$; second row: $\varepsilon = -1.5$; third row: $\varepsilon = -1$; and fourth row: $\varepsilon = -0.5$. Leftmost column: $n = 50$; middlemost column: $n = 200$; and rightmost column: $n = 500$.

IPB estimators appear to be the best as their MSEs approach 0. This is in conformity with what have been reported in [18, 19] and the results in Section 2.5. Also, the MLE estimator gives better performance than all other estimation methods when $\varepsilon = -1$ at $n = 50$. In general, the LME overestimates all values of $\varepsilon < 0$ and, hence, performs poorly.

In addition, Figure 4 shows the MSEs of the six estimators of $\varepsilon(\varepsilon \geq 0)$. The results indicate that the proposed methods are competitive with LME as they have smaller MSEs across most of the values of k. Closely following the sample path of both proposed methods and LME is

the MLE: it has better performance in some cases. Lastly, the MSEs of the estimators converge to 0 as $k \longrightarrow n$, and this shows a sign of empirical consistency.

Therefore, we conclude that the proposed estimators are competitive with the existing ones, and at worse, it has MSEs close to that of likelihood-based method (LME).

4. Practical Application

In order to demonstrate the use of the proposed methods, we consider the estimation of the tail indexes of the underlying distributions of two datasets. First, the monthly mean
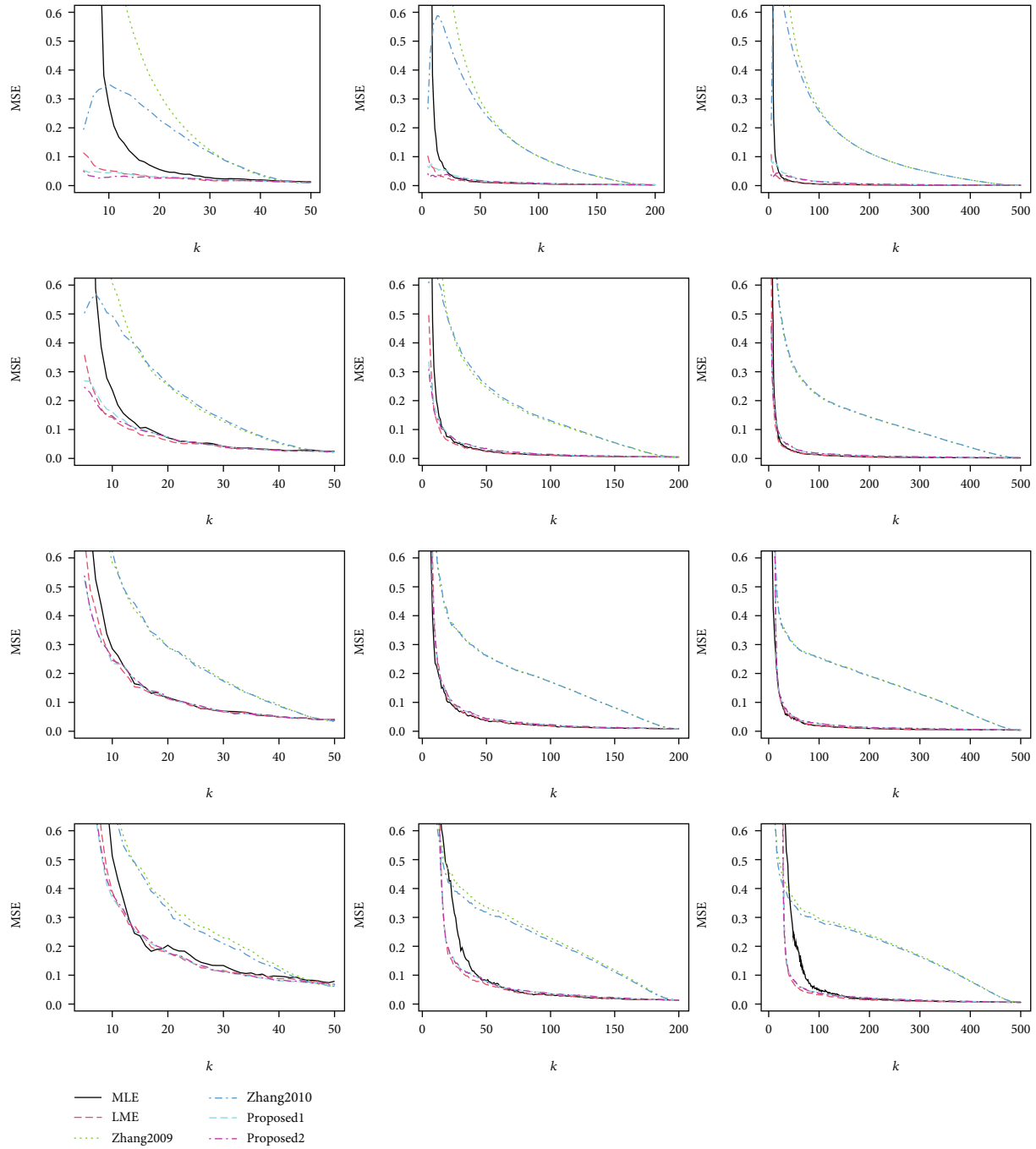
FIGURE 4: MSE($\widehat{\varepsilon}$) with $\sigma = 1$. First row: $\varepsilon = 0$; second row: $\varepsilon = 0.5$; third row: $\varepsilon = 1$; and fourth row: $\varepsilon = 1.5$. Leftmost column: $n = 50$; middlemost column: $n = 200$; and rightmost column: $n = 500$.

TABLE 1: Descriptive statistics of the monthly mean temperature of Ghana.

Data	n	Min.	Max.	Mean	St. dev.	Skewness	Shapiro-Wilk
Temperature	1417	24.033	31.134	27.260	1.492	0.350	0.971 (>0.01)
Fire claims	2167	1.000	263.250	3.385	8.507	18.750	0.197 (>0.01)

temperature series of Ghana for the period 1901 to 2019 obtained from the climate monitor unit hosted by the Climatic Research Unit, University of East Anglia. The second

data is the Danish fire insurance claims in Denmark for the period 3 January 1980 to 31 December 1990 and that was obtained from the EVIR package in R.
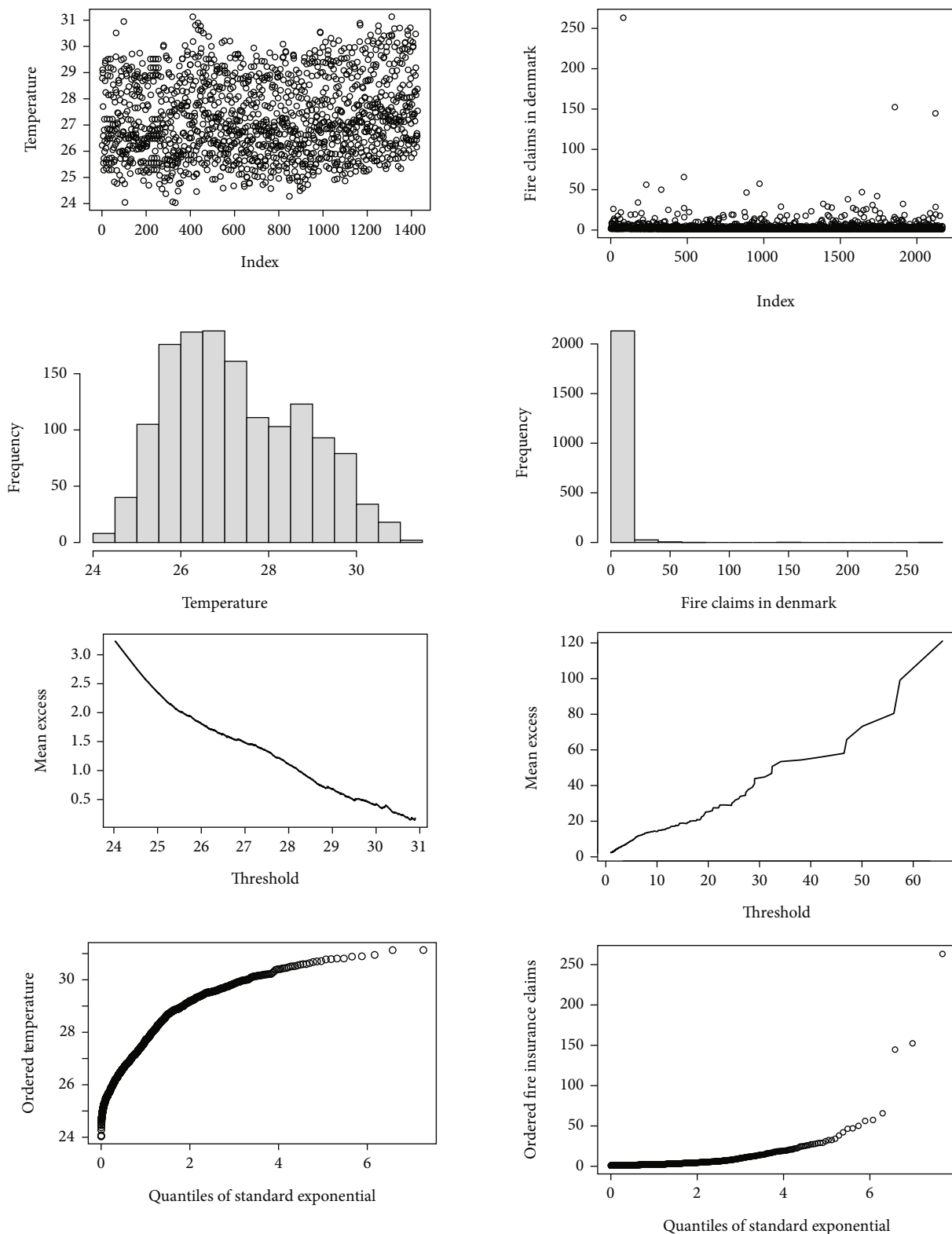
FIGURE 5: Plot of the monthly mean temperature of Ghana.

4.1. Descriptive Measures. Table 1 shows the main descriptive measures for both monthly mean temperature of Ghana with 1417 monthly observations and the fire insurance claims in Denmark with 2167 daily series. Whereas the fire claim data is heavy-tailed, the distribution of the temperature data shows a mild positive skewness.
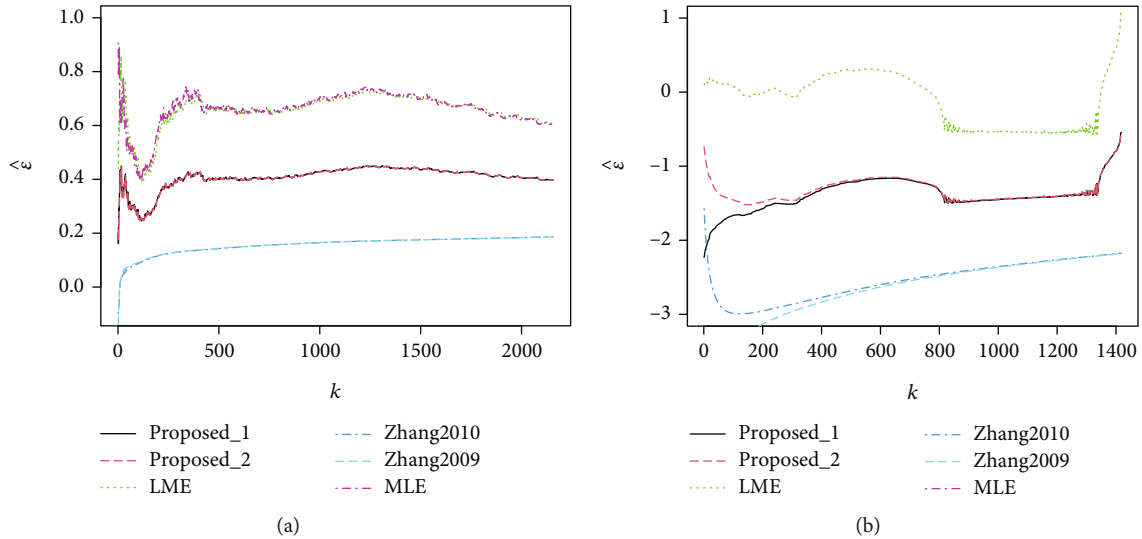
FIGURE 6: (a) Danish fire insurance claims; (b) mean temperature of Ghana.

The P values of the Shapiro-Wilk test for normality reject the null hypothesis of a normal distribution for both series, thereby making it appropriate for extreme value analysis.

From the plots in Figure 5, no obvious trend can be seen in the temperature data. However, the Danish fire insurance claim had few outliers. The temperature series is dense in the range of 25.5 to 27 while the Danish fire claims are denser from one million to 2 million Danish Krone. The histogram and the exponential quantile-quantile plots of the two datasets indicate that the fire insurance claims have a long tail (heavy tail) while the temperature series has a shorter tail.

In addition, the mean excess plots show that the temperature data is light-tailed and, hence, indicates a likely negative value of the tail index, ε, whereas the Danish shows more of heavy-tailed.

4.2. Estimation of the Tail Index. Figure 6 gives the sample paths of the estimators of the tail index of the underlying distribution of the Danish and temperature datasets. The sample paths of the estimators show that there exist large variations at smaller values of k and smoother for larger values of k. The former can be explained by the small number of observations involved in the estimation of the tail index and, hence, the large variations associated with the sample paths. On the other hand, the latter can be explained by the large number of observations involved in the estimation of the tail index and, thus, the less variations associated with the sample paths.

In addition, the proposed estimators are in between the sample paths of the likelihood-based and empirical Bayesian-based estimators as required.

5. Conclusion

The importance of the generalized Pareto distribution cannot be overemphasized as it has wide applications in finance, hydrology, and economics, among others. In this paper, we proposed two estimators of the parameters of the generalized Pareto distribution. The estimators are based on the existing likelihood moment and the empirical Bayesian-based estimators using the idea of shrinkage. The performance of the proposed estimators was assessed using a large-scale simulation study. The results show that the likelihood moment estimator is not appropriate for negative values of the shape parameter. Also, the empirical Bayesian-based estimators use all the available data and hence defeat the threshold characteristics of the GPD. In addition, it performs well for only in the case where the number of exceedances approaches the sample size. The proposed estimators have competitive MSE for small to moderate number of exceedances. This is appropriate as the inclusion of smaller observations (i.e., large number of exceedances) introduces large bias. Moreover, the proposed estimators are at worst close to the performance of the likelihood-based estimator. The estimators were illustrated using two real datasets from the climate sector representing light-tailed distributions and the insurance industry representing heavy-tailed distributions. The asymptotic properties of the extreme value index estimators in [18, 19] remain an open problem. Consequently, since our proposed shrinkage estimators of the extreme value index depend on these estimators, their asymptotic properties also remain an open problem. Therefore, these asymptotic properties, estimation of the scale parameter, and other problems such as estimation of high quantiles and exceedance probabilities are the subject of future research.

Data Availability

Data were obtained from the Climate Research Unit and the EVIR package in R. The temperature data used to support the findings of this study have been deposited in a git-hub repository (https://github.com/wilheminapels2/Shrinkage/blob/main/Temperature.csv).

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

[1] A. A. Balkema and L. De Haan, "Residual life time at great age," *The Annals of Probability*, vol. 2, no. 5, pp. 792–804, 1974.

[2] I. I. I. James Pickands, "Statistical inference using extreme order statistics," *The Annals of Statistics*, vol. 3, no. 1, pp. 119–131, 1975.

[3] R. Minkah, "An application of extreme value theory to the management of a hydroelectric dam," *Springer-Plus*, vol. 5, no. 1, 2016.

[4] J. Jockovic, "Quantile estimation for the generalized Pareto distribution with application to finance," *Yugoslav Journal of Operations Research*, vol. 22, no. 2, 2012.

[5] P. Embrechts, C. Kluppelberg, and T. Mikosch, *Modelling Extremal Events: For Insurance and Finance*, Springer, 1997.

[6] J. Jockovic and P. Mladenovic, "Coupon collector's problem and generalized Pareto distributions," *Journal of Statistical Planning and Inference*, vol. 141, no. 7, pp. 2348–2352, 2011.

[7] R. W. Katz, G. S. Brush, and M. B. Parlange, "Statistics of extremes: modeling ecological disturbances," *Ecology*, vol. 86, no. 5, pp. 1124–1134, 2005.

[8] J. Kysely, J. Picek, and R. Beranova, "Estimating extremes in climate change simulations using the peaks-over-threshold method with a non-stationary threshold," *Global and Planetary Change*, vol. 72, no. 1-2, pp. 55–68, 2010.

[9] W. A. Pels, A. O. Adebanji, and S. Twumasi-Ankrah, *Comparison of parameter estimators for generalized Pareto distribution under peak over threshold*, Horizon Research Publishing CorporationJournal: Mathematics and Statistics, 2020.

[10] S. Coles, *An Introduction to Statistical Modelling of Extreme Values*, Springer, London, 2001.

[11] R. Minkah, "Tail index estimation of the generalised Pareto distribution using a pivot from a transformed Pareto distribution," *Science and Development*, vol. 4, no. 1, pp. 1–19, 2020.

[12] R. Deidda and M. Puliga, "Performances of some parameter estimators of the generalized Pareto distribution over rounded-off samples," *Physics and Chemistry of the Earth, Parts A/B/C*, vol. 34, no. 10-12, pp. 626–634, 2009.

[13] J. R. M. Hosking and J. R. Wallis, "Parameter and quantile estimation for the generalized Pareto distribution," *Technometrics*, vol. 29, no. 3, pp. 339–349, 1987.

[14] E. Castillo and A. S. Hadi, "Fitting the generalized Pareto distribution to data," *Journal of the American Statistical Association*, vol. 92, no. 440, pp. 1609–1620, 1997.

[15] J. Zhang, "Likelihood moment estimation for the generalized Pareto distribution," *Australian&New Zealand Journal of Statistics*, vol. 49, no. 1, pp. 69–77, 2007.

[16] F. Ashkar and C. Nwentsa Tatsambon, "Revisiting some estimation methods for the generalized Pareto distribution," *Journal of Hydrology*, vol. 346, no. 3-4, pp. 136–143, 2007.

[17] J. Song and S. Song, "A quantile estimation for massive data with generalized Pareto distribution," *Computational Statistics & Data Analysis*, vol. 56, no. 1, pp. 143–150, 2012.

[18] J. Zhang and M. A. Stephens, "A new and efficient estimation method for the generalized Pareto distribution," *Technometrics*, vol. 51, no. 3, pp. 316–325, 2009.

[19] J. Zhang, "Improving on estimation for the generalized Pareto distribution," *Technometrics*, vol. 52, no. 3, pp. 335–339, 2010.

[20] J. Beirlant, Y. Goegebeur, J. Segers, and J. L. Teugels, *Statistics of Extremes: Theory and Applications*, vol. 558, John Wiley & Sons, 2004.

[21] J. Martin Van Zyl, "Estimation of the shape parameter of a generalized Pareto distribution based on a transformation to Pareto distributed variables," *Journal of Statistical Theory and Practice*, vol. 9, no. 1, pp. 171–183, 2015.

[22] S. Kang and J. Song, "Parameter and quantile estimation for the generalized Pareto distribution in peaks over threshold framework," *Journal of the Korean Statistical Society*, vol. 46, no. 4, pp. 487–501, 2017.

[23] E. B. L. Mackay, P. G. Challenor, and A. B. S. Bahaj, "A comparison of estimators for the generalised Pareto distribution," *Ocean Engineering*, vol. 38, no. 11-12, pp. 1338–1346, 2011.

[24] M. H. Park and J. H. T. Kim, "Estimating extreme tail risk measures with generalized Pareto distribution," *Computational Statistics & Data Analysis*, vol. 98, pp. 91–104, 2016.