

# Point and exact interval estimation for the generalized Pareto distribution with small samples

JIAN HE, ZHUO SHENG, BING XING WANG, AND KEMING YU\*

In extreme value theory, the generalized Pareto distribution (GPD) is used to model another distribution on the tail. Since only a proportion of the data is used, the effective data size for fitting GPD is often small. As statistical properties, especially tail behaviour, of GPD largely depend on its shape parameter, performances of most existing methods are inconsistent when the value of the shape parameter varies. In this paper, we introduce a new method to fit GPD that improves the performance over existing methods for very small samples, in terms of bias and mean square error as well as confidence intervals. The numerical study in this paper also shows that better performance on parameter estimation does not necessarily lead to better performance on quantile estimation.

KEYWORDS AND PHRASES: Confidence interval, Extreme value theory, Generalized Pareto distribution, Quantile, Shape parameter, Small sample.

## 1. INTRODUCTION

The generalized Pareto distribution (GPD) has recently received a great deal of attention in the literature because of its application in the extreme value theory (EVT) (see [4], [28], [29] and [15], amongst others). The development of EVT is motivated by the need to measure extreme tail probabilities and tail quantiles, which, in fields such as the research of disasters, nature phenomena, insurance and finance, are often the major concern of risk-related problems. The essential idea of EVT is, as stated in [8], estimating extreme quantiles and probabilities by fitting models to a set of data using only the data on the tail. Consequently, since only a proportion of original data is used, small sample size is often confronted in tail modeling. Hence, we emphasize the aspect of small samples on fitting GPD in this paper.

Practically, GPD is often used to model exceedances over certain thresholds in the EVT. This method of modelling exceedances over high thresholds is often referred to as the peak over threshold (POT) method. Assume that  $F_Y(y)$  is the distribution function of random variable  $Y$  and that, the

definition of exceedances is the values of  $Y$  above a threshold  $u$ , namely  $X = Y - u$ . In the EVT, instead of assessing  $F_Y(y)$  directly, we are more interested in estimating the distribution of the exceedances

$$F_u(x) = Pr\{Y - u \leq x | Y \geq u\}, \quad 0 \leq x \leq y_F - u,$$

where  $x$  is the excess and  $y_F \leq \infty$  denotes the right end points of  $F_Y$ . It follows that

$$(1) \quad F_u(x) = \frac{F(u+x) - F(u)}{1 - F(u)} = \frac{F_Y(y) - F(u)}{1 - F(u)}.$$

[2] along with [17] showed that, when  $u$  is large, for a large class of underlying distribution functions, the conditional excess distribution function  $F_u$  is well approximated by GPD:

$$(2) \quad F_u(x) \approx F_{\text{GPD}}(x; \sigma, \xi), \text{ as } u \rightarrow \infty,$$

where  $F_{\text{GPD}}(x; \sigma, \xi)$  is the cumulative distribution function

$$(3) \quad F_{\text{GPD}}(x; \sigma, \xi) = \begin{cases} 1 - (1 + \frac{\xi}{\sigma}x)^{-1/\xi} & \text{if } \xi \neq 0 \\ 1 - e^{-x/\sigma} & \text{if } \xi = 0 \end{cases}$$

for  $x \in [0, y_F - u]$  if  $\xi \geq 0$  and  $x \in [0, -\frac{\sigma}{\xi}]$  if  $\xi < 0$ , where  $\xi$  is the shape parameter and  $\sigma$  is the scale parameter. If  $\xi > 0$  the GPD is heavy-tailed on the right side. If  $\xi = 0$  the distribution reduces to the exponential distribution and the right tail decays exponentially. If  $-1 \leq \xi < 0$ , the distribution has finite right end-points, and if  $\xi \leq -0.5$ , the finite right end-points  $y_F > 0$ . These are sometimes referred as ‘short-tailed’. Furthermore, when  $\xi = -1$ , the GPD actually becomes a uniform  $U(0, \sigma)$  distribution. The  $k$ th central moment of the GPD exists only if  $\xi < 1/k$ . For example, when  $\xi \geq 1/2$ ,  $var(X) = +\infty$  as the second central moment no longer exists. For more details about the properties of GPD and its parameters see [14].

From (3), when  $\xi \neq 0$ , the  $p$ -th quantile of GPD is given by

$$(4) \quad Q_{\text{GPD}}(p; \sigma, \xi) = F_{\text{GPD}}^{-1}(p; \sigma, \xi) = \frac{\sigma}{\xi} ((1-p)^{-\xi} - 1).$$

To simplify the notations, we will always assume that  $\xi \neq 0$  and write  $Q(p) = Q_{\text{GPD}}(p; \sigma, \xi)$  henceforth.

\*Corresponding author.

In practice, the final goal of employing the EVT is still estimating extreme probabilities or quantiles of the original distribution of  $Y$ . Hence, by substituting  $F_u$  in (1) with the GPD and approximating  $F(u)$  by  $(N - n)/N$ , where  $N$  is the total number of observations and  $n$  is the number of observations over the threshold  $u$ , the original distribution  $F_Y(y)$  is expressed in terms of GPD parameters as

$$(5) \quad F_Y(y) = 1 - \frac{n}{N} \left( 1 + \frac{\xi}{\sigma}(y - u) \right)^{-1/\xi}.$$

The  $p^*$ -th quantile of  $Y$ ,  $F_Y^{-1}(p^*)$  is then given by inverting (5):

$$(6) \quad F_Y^{-1}(p^*) = u + \frac{\sigma}{\xi} \left( \left( \frac{N}{n}(1 - p^*) \right)^{-\xi} - 1 \right).$$

Comparing (4) and (6) we will obtain that

$$(7) \quad F_Y^{-1}(p^*) = u + Q \left( \frac{N}{n} p^* \right).$$

From (6) when  $p^*$  is very extreme, instead of estimating the  $p^*$ -th quantile of  $F_Y$ , which could be very difficult, one could estimate less-extreme  $(\frac{N}{n} p^*)$ -th quantiles of  $F_{\text{GPD}}$ , assuming  $N \gg n$ .

## 1.1 A brief review of existing methods

For fitting the GPD, the typical method is the maximum likelihood estimation (MLE) ([6], [20], [21], [14], [12], *etc.* among others). However, Hosking and Wallis suggested that unless sample size is equal to or larger than to 500, the method of moment (MOM) and the probability weighted moment (PWM) are more reliable for  $\xi \in (-0.5, 0.5)$ . Then [28] proposed the likelihood moment estimation (LME) for fitting the GPD. And [29] also suggested an alternative method (referred as Z&S henceforth) similar to the Bayesian method for that aim. The Z&S method is fast to implement and their simulation tests show that the method is stable when  $\xi$  is in the domain of  $[-0.5, 1]$  for sample size  $n \in [50, 500]$ , and generally outperforms over MLE, MOM, PWM and LME in most cases.

Compared with point estimation, confidence interval (CI) estimation for the GPD has drawn less attention in the existing literature, due to the difficulty of constructing such interval estimates explicitly. The common strategy is finding approximated confidence intervals with coverage probabilities equal to the target confidence level, theoretically. One well-known approach is approximation based on the asymptotic normal distribution of estimates. The  $(1 - \beta)\%$  CI of parameter  $\theta$  is constructed in the form of

$$\hat{\theta} - z_{1-\beta/2} \sqrt{\hat{v}(\hat{\theta})} < \theta < \hat{\theta} + z_{1-\beta/2} \sqrt{\hat{v}(\hat{\theta})},$$

where  $\hat{\theta}$  is the estimator,  $z_\tau$  is the  $\tau$ -th quantile of the standard normal distribution, and  $\hat{v}(\hat{\theta})$  is the asymptotic variance of  $\hat{\theta}$ . Furthermore, the asymptotic variance can be obtained via either  $\hat{v}(\hat{\theta}) = 1/I(\hat{\theta})$  or  $\hat{v}(\hat{\theta}) = 1/J(\hat{\theta})$ , where  $I(\hat{\theta})$  and  $J(\hat{\theta})$  are the expected and the observed Fisher information, respectively. The asymptotic variance derived from the expected Fisher information can be found in [21] (for MLE), [14] (for MOM and PWM), and [28] (for LME), respectively. Example of CI estimation based on observed information can be found in [7] amongst others. CI for quantile estimators can also be obtained by using asymptotic variance via the Delta method ([18]). However, CI estimates based on asymptotic theory may require stronger conditions in assumption and suffer from small samples. For small samples, profile log-likelihood approach may be preferred (*e.g.*, [24], [7]). The profile log-likelihood approach is based on the fact that the relative likelihood  $\mathcal{L}(\theta)/\mathcal{L}(\hat{\theta})$  follows the  $\chi^2$  distribution with 1 degree of freedom. [11] gave examples of profile log-likelihood CI estimates for some financial criteria based on the GPD estimation. [23] compared sampling based approaches (*i.e.* jackknife and bootstrap) with profile likelihood approaches for heavy-tailed GPD and found the later is better for both small and large samples. In this paper, we adapt the inference in Wang, Yu and Jones (2010) and propose an approach that does not depend on asymptotic normality for calculating the exact and generalized confidence intervals for the GPD.

## 1.2 Challenging issues

Since in the POT approach only a proportion of the original sample is used for fitting the GPD, the sample size of the excess,  $n$ , is often small. Generally, the sample size  $n$  for modelling the GPD is determined by threshold  $u$ . According to (2), a high threshold generally provides better approximation of the data on the tail with smaller bias. However, higher  $u$  also leads to smaller sample size  $n$ , which may increase estimation variance and bring other problems to the GPD estimation. Furthermore, for some estimators, their asymptotic properties based on large samples may not be valid when  $n$  is too small. Consequently, interval estimations that are based on asymptotic theory or bootstrap approaches become less efficient and less accurate. In practice the GPD fitting is often confronted with small samples, although definitions of ‘small’ and ‘large’ samples are somehow vague. We illustrate this via a simple example: in finance, a typical one-year dataset of daily returns (or losses) normally consists of around 250 samples (as there are around 250 trading days per year), which might not be very large but at least not ‘small’. If 10% data on the tail is selected from the original dataset, there are only 25 samples available for fitting the GPD; even if 30% of data is selected there are still only 75 samples. It is possible to increase the sample size by using a longer sampling period, but it is not always wise as older data could be out-of-date. Data from other fields of research might contain larger samples, nonetheless,

as not the entire dataset is used, the effective sample size for estimating the GPD is often much smaller. In this paper, we focus on sample size  $n \leq 50$ .

As described in the previous section, some of the distributional properties of the GPD vary distinctively as the value of the shape parameter  $\xi$  changes, an ideal estimator for GPD should be reliable and consistent regardless of its ‘shape’, namely, the tail behaviour of the underlying distribution. However, performance of some existing methods are heavily impacted by the value of  $\xi$ , or even become invalid.

For example, the MLE method does not exist for  $\xi < -1$  since it can have no local maximum. The MOM estimates do not exist for  $\xi \geq 0.5$  and the PWM estimates do not exist for  $\xi \geq 1$ , since the second and first moments of the GPD become infinite in each case, respectively. Although [14] suggested that it should be sufficient to restrict to  $-0.5 < \xi < 0.5$  when considering reliability, [4] made a good line of argument that a good estimator for the GPD should cover a wider domain of  $\xi$  for both practical and theoretical reasons. The EPM method (elemental percentile method) they proposed for parameter estimations is reliable under a wider domain of  $\xi$  (from  $-2$  to  $2$ ), but it does not outperform classic methods as long as the value of  $\xi$  falls into  $(-0.5, 0.5)$ . The LME method proposed by [28] has high asymptotic efficiency when  $\xi$  is close to its optimal value. However, the LME is valid only if  $\xi > -0.5$ . Also, some methods such as MOM and PWM sometimes conflict with data (see, *e.g.*, [4], [9]). That is, for  $\xi < 0$ , the supported domain of  $x$  is  $0 < x < -\sigma/\xi$ , however, invalid estimation of  $\sigma$  and  $\xi$  may be given so that some observations can be out of this domain. This is referred to as a non-feasible solution or an invalid estimation issue in the literature.

Although some methods may be preferred for certain ranges of  $\xi$ , it is typical that the tail behaviour of the underlying data is unknown prior to modelling in real-world practice. In this paper, we elaborate on the reliability of different methods, with regard to  $\xi \in (-1, 1)$ , including the method we proposed.

In this paper, the proposed new approach to deal with these challenges in fitting GPD may not be always the best in all scenarios, but it is indeed the most versatile with overall the smallest bias for parameter estimations under very small sample size. The new method based confidence interval estimation is consistently reliable for moderately small sample sizes. Details of this method are described in Section 2. Results and discussions of simulation tests for different methods and for CI estimations are presented in Section 3. It is noticed that better performance on parameter estimation does not necessarily lead to better performance in estimating higher upper quantiles of the GPD. An example using real-world data will be discussed in Section 4, and finally summary and conclusions in the last section.

## 2. EXACT INFERENCE FOR THE GPD

### 2.1 Point estimation

Let  $\{X_i\}_{i=1}^n$  be the sample from the GPD, and let  $X_{1:n} \leq X_{2:n} \leq \dots \leq X_{n:n}$  be the associated order statistics from the GPD. Re-parametrize the GPD with  $\alpha = \xi/\sigma$ , for  $\xi \neq 0$ , the proposed estimator for  $\alpha$  is then given via some transformation from the c.d.f. (3) as follows. First, let

$$\begin{aligned} V_{(i)} &= -\log(1 - F_{\text{GPD}}(X_{i:n}; \alpha, \xi)) \\ &= \frac{1}{\xi} \log(1 + \alpha X_{i:n}), \quad i = 1, 2, \dots, n \end{aligned}$$

be a sequence of ascending order statistics of samples from the standard exponential distribution ([25]).  $V_{(i)}$  are not independent, so let

$$W_i = (n - i + 1)(V_{(i)} - V_{(i-1)}), \quad i = 1, 2, \dots, n;$$

where  $V_{(0)} = 0$ , then  $W_1, \dots, W_n$  are random variables from independent standard exponential distributions (see [22] as an example). Then for  $i = 1, 2, \dots, n$  let

$$\begin{aligned} (8) \quad D_i &= \sum_{j=1}^i W_j \\ &= \frac{1}{\xi} \left( \sum_{j=1}^i \log(1 + \alpha X_{j:n}) + (n - i) \log(1 + \alpha X_{i:n}) \right), \end{aligned}$$

and

$$\begin{aligned} (9) \quad U_{i:n} &= \frac{D_i}{D_n} \\ &= \frac{\sum_{j=1}^i \log(1 + \alpha X_{j:n}) + (n - i) \log(1 + \alpha X_{i:n})}{\sum_{j=1}^n \log(1 + \alpha X_{j:n})}, \end{aligned}$$

$\{U_{i:n}\}_{i=1}^{n-1}$  can be equivalently regarded as an induced order statistics of an i.i.d. sample which has a standard uniform  $U(0, 1)$  distribution ([22]). Note that  $U_{i:n}$  is also an ancillary statistic that only depends on parameter  $\alpha$ . The mean of  $\{U_{i:n}\}_{i=1}^{n-1}$  converges with probability one to  $1/2$ . Hence  $\hat{\alpha}$  is determined by solving

$$(10) \quad \bar{U}(\alpha) = \frac{1}{2},$$

where

$$\bar{U}(\alpha) = \frac{1}{n-1} \sum_{i=1}^{n-1} U_{i:n}.$$

Since  $V_{(i)}$  can be seen as order statistics from standard exponential distribution, it has sample mean  $\frac{1}{n} \sum_{i=1}^n V_{(i)} =$

1. This leads to an estimator for the shape parameter  $\xi$ , given estimator  $\hat{\alpha}$ :

$$(11) \quad \hat{\xi} = \frac{1}{n} \sum_{i=1}^n \log(1 + \hat{\alpha}X_i).$$

Note that (11) is identical with the log-likelihood function for  $\xi$  as in MLE. The method we proposed is referred as NEW henceforth.

We notice that the idea behind (10) is similar to [15], which assume that the probabilities of the GPD given data  $(X_1, X_2, \dots, X_n)$  can also be regarded as uniformly  $U(0, 1)$  distributed, and then the estimator  $\hat{\alpha}$  of  $\alpha$  can be obtained by solving the equation

$$\frac{1}{n} \sum_{i=1}^n F_{\text{GPD}}(X_i; \hat{\alpha}, \hat{\xi}) = \frac{1}{2},$$

along with the profile log-likelihood function (11). The essential difference between NEW method and the approach in Hüsler *et. al* is that the later requires to solve  $\hat{\xi}$  and  $\hat{\alpha}$  simultaneously. For the NEW method,  $\hat{\alpha}$  can be obtained directly, since  $U_{i:n}$  is an ancillary statistic that does not depend on any parameters other than  $\alpha$ . This provides the flexibility of employing different method of estimating the shape parameter  $\xi$ , given  $\hat{\alpha}$ . Through simulations we found that the profile log-likelihood estimate for  $\hat{\xi}$  as shown in (11) is the most accurate one.

Thus, our estimators for the GPD parameters are finally given by combining (10) and (11). And the  $p$ th quantile of the GPD is estimated by either  $\hat{Q}(p; \hat{\sigma}, \hat{\xi})$  or  $\hat{Q}(p; \hat{\alpha}, \hat{\xi})$ , using the inverse distribution function as shown in (4).

## 2.2 Exact and generalized confidence intervals estimation

Confidence interval estimation for  $\alpha$  and  $\xi$  can be derived from (10). Note that the mean of  $n$  independent uniformly  $U(0, 1)$  distributed random variables follows the Bates distribution ([3]), which has the p.d.f. as

$$f_X(x; n) = \frac{n}{2(n-1)!} \sum_{k=0}^n (-1)^k \binom{n}{k} (nx - k)^{n-1} \text{sgn}(nx - k),$$

for  $x \in (0, 1)$ , and

$$\text{sgn}(nx - k) = \begin{cases} -1 & nx < k \\ 0 & nx = k \\ 1 & nx > k. \end{cases}$$

Recall from the last section, given  $X = x$ , it is obvious that the sample distribution of  $\bar{U}(\alpha)$  is the Bates distribution with size  $n - 1$  (noted as  $\bar{U}(\alpha) \sim \text{Bates}(n - 1)$ ). Let  $\mu_L$

and  $\mu_U$  be lower and upper boundaries such that the probability of the value of  $\bar{U}(\alpha)$  falls between these boundaries are  $1 - \beta$ , i.e.

$$\mathbb{P}(\mu_L < \bar{U}(\alpha) < \mu_U) = 1 - \beta.$$

Let  $\text{Bates}_{\tau}^{-1}(n)$  represent the  $\tau$ -th quantile of the Bates distribution with sample size  $n$ , then  $\mu_L = \text{Bates}_{\beta/2}^{-1}(n - 1)$  and  $\mu_U = \text{Bates}_{1-\beta/2}^{-1}(n - 1)$ , respectively. Since  $\bar{U}(\alpha)$  is an pivot quantity and monotonic, it is straightforward to show that

$$\mathbb{P}(A(\mu_L; x) < \alpha < A(\mu_U; x)) = 1 - \beta,$$

where  $A(\mu; x) = \alpha$  is the inverse function of  $\bar{U}(\alpha) = \mu$ , given  $X = x$ . That is, the exact  $(1 - \beta)\%$  confidence interval for  $\alpha$  is

$$(12) \quad \left( A(\mu_L; x), \quad A(\mu_U; x) \right).$$

However,  $\mu_L$  and  $\mu_U$  are difficult to calculate analytically from the Bates distribution, especially for large  $n$ . It is more convenient to obtain these values via Monte Carlo simulation. For example, generate a large amount (say,  $m \geq 2000$ ) of random samples  $\boldsymbol{\mu}$  from  $\text{Bates}(n - 1)$  distribution, and the  $\tau$ -th population quantiles,  $\text{Bates}_{\tau}^{-1}(n - 1)$ , can be well approximated by the  $\tau$ -th sample quantile of  $\boldsymbol{\mu}$ . Then the confidence limits as in (12) can be obtained, using these simulated sample quantiles.

The generalized confidence interval for the shape parameter  $\xi$  can be derived accordingly. Recall  $D_i$  as in (8), and let

$$T = 2D_n = 2 \sum_{j=1}^n \log(1 + \alpha X_{j:n})^{\frac{1}{\xi}}.$$

As shown in Wang, Yu and Jones (2010),  $T$  has a  $\chi^2$  distribution with  $2n$  degrees of freedom, regardless of the values of  $\alpha$  and  $X$ . Then consider the following generalized pivot quantity for parameter  $\xi$

$$(13) \quad Z = \frac{2 \sum_{j=1}^n \log(1 + A(\mu; x) X_{j:n})}{T},$$

where  $A(\mu; x)$  is previously defined. Note that the value of  $Z$  will reduce to  $\xi$  if  $\mu = 1/2$ . As there are no unknown parameters in (13),  $Z$  is a generalized pivotal quantity for  $\xi$  as defined in [26] and [27], and its distribution can be approximated via Monte Carlo simulations, as follows:

1. generate random samples  $\mathbf{t} \sim \chi^2(2n)$ , from the  $\chi^2$  distribution with  $2n$  degrees of freedom, with the same sample size  $m$  as of random samples  $\boldsymbol{\mu}$ ;
2. replace  $\mu$  and  $T$  in (13) with  $\boldsymbol{\mu}$  and  $\mathbf{t}$ , respectively. This produces a simulated random sample set of  $Z$ , denoted as  $\mathbf{z}$ . Hence, the  $\tau$ -th sample quantile of  $\mathbf{z}$  can be used to approximate the  $\tau$ -th population quantile of  $Z$ , denoted as  $Z_{\tau}$ ;

- give the the generalized  $(1 - \beta)\%$  confidence interval of  $\xi$  by:

$$(Z_{\beta/2}, Z_{1-\beta/2}).$$

Similarly, the generalized confidence interval for the quantile of the GPD,  $Q_{\text{GPD}}(p; \alpha, \xi)$ , can be obtained via the following generalized pivotal quantity:

$$(14) \quad S = Q_{\text{GPD}}(p; A(\mu; x), Z).$$

The population quantiles of  $S$  can be derived along the same line: replacing  $\mu$  and  $Z$  in (14) with  $\boldsymbol{\mu}$  and  $\mathbf{z}$ , which are previously defined, to obtain the Monte Carlo samples of  $S$ . Then the  $\tau$ -th population quantile of  $S$ , denoted as  $S_\tau$ , can be approximated by the  $\tau$ -th sample quantile of the generated sample. Hence, the  $(1 - \beta)\%$  generalized confidence interval of  $Q_{\text{GPD}}(p; \alpha, \xi)$  is given by

$$(S_{\beta/2}, S_{1-\beta/2}).$$

In the next section, we conducted a series of simulation tests to compare the performance of the NEW method with other methods, for both point estimation and interval estimation.

### 3. SIMULATION STUDY

We conducted simulation experiments to compare parameter estimation from methods mentioned in previous sections. Random samples were generated from the GPD with different settings to cover a wide range of scenarios: sample sizes  $n = \{50, 30, 15\}$ , the scale parameter  $\sigma = 1$ , and, as arguably the most important variable in modelling the GPD, the shape parameter  $\xi$  ranging from  $-1$  to  $1$  by  $0.05$  interval, respectively.  $5000$  replicates were generated for each scenario. Since the scope of the study is small sample sizes, the MLE method is omitted for point estimation, as it has been proved inefficient when the sample size is small (see Section 1). Totally, five methods were adopted for point estimation, including method of moment (MOM), probability weighted moment (PWM), likelihood moment estimation (LME), Zhang and Stephens' method (Z&S), and the new method we proposed (NEW). MOM and PWM are listed as benchmarking methods and the POT package ([19]) in the statistical programming language R is used for the parameter and quantile estimation. For the implementation of the LME method, we need to know an auxiliary parameter  $r$  by taking account of any preliminary information about  $\xi$ , but  $r = -1/2$  is recommended by [28] if no information about  $\xi$  is available.

#### 3.1 Parameter estimation

First, we investigated the performance of the estimators  $\hat{\sigma}$  and  $\hat{\xi}$  from different models. Our main concern is how reliable the parameter estimators are under different tail behaviour, namely, different values of  $\xi$ . Estimation bias and root mean square errors (RMSE) of the five tested meth-

ods were calculated plotted against  $\xi$  in Figure 1 and 2, respectively. As for each sample size, there were 41 tests of different values of  $\xi$ . To save space, instead of presenting all results, we report several selected cases with different sample sizes and  $\xi = \{-1, -0.5, -0.25, 0, 0.25, 0.5, 0.75, 1\}$ , through Table 1 and 2. Both MOM and PWM methods have theoretical upper boundaries with regard to  $\xi$ , beyond which the methods no longer exist, hence their results are truncated in the figures and tables. The results can be summarized as follows.

From Figure 1 and 2, it is intuitive that the NEW estimators are the most versatile parameter estimators in terms of bias. The bias of the scale estimator  $\hat{\sigma}$  of NEW shows slight sensitivity to the tail behaviour and sample size, as the bias slightly increases when the value of  $\xi$  increases, or the sample size decreases; the bias of the shape estimator  $\hat{\xi}$  of NEW are consistently close to zero, through the tested range of  $\xi$  and different sample sizes. The performance of other tested methods are obviously more sensitive to the change of tail behaviour. The MOM and PWM methods tend to have significant estimation bias as the value of  $\xi$  closes to the theoretical boundaries ( $\xi = 0.5$  and  $1$ , respectively). This generally agrees with findings in previous research, such as in [29], Fig. 1&2 and [15], Fig. 2-5. LME, MOM and PWM methods overestimate the scale parameter  $\sigma$  and underestimate the shape parameter  $\xi$ , and become more biased as the tail of the GPD becomes heavier. But the bias from the LME are generally smaller than the other two. The Z&S estimators have an even smaller bias compared with MOM, PWM and LME estimators, but they are still more biased than the NEW and apparently more sensitive to  $\xi$ , as both parameter estimators become more biased when  $\xi$  closes to  $-1$  or  $1$ .

From the other aspect, the Z&S parameter estimators have the smallest overall estimation error. This can be identified from both graphical and numerical evidences. The RMSE of the LME, Z&S and NEW estimators are less sensitive to  $\xi$  and have smaller ranges, compared with the MOM and PWM methods. The second rows of Figure 1 and 2 suggest that, as for smaller sample size, the Z&S estimators' RMSE advantage becomes more obvious. The NEW estimators' performance in terms of RMSE is not far behind the Z&S and is comparable with LME estimators. Exceptions are observed for the scale estimator  $\hat{\sigma}$ : when the value of  $\xi$  closes to  $1$ , the NEW method actually has the smallest RMSE of all methods.

It is safe to draw conclusions that, for parameter estimation, the NEW method is the most versatile and consistent among all tested methods, with regard to varying tail behaviour and small sample sizes of the simulated GPD samples. In particular, the shape parameter  $\hat{\xi}$  of the NEW is unbiased, with small estimation variation.

#### 3.2 Quantile estimation

In practice, it is of great interest to assess the quantiles of the GPD. Thus we proceeded to analyze the per-

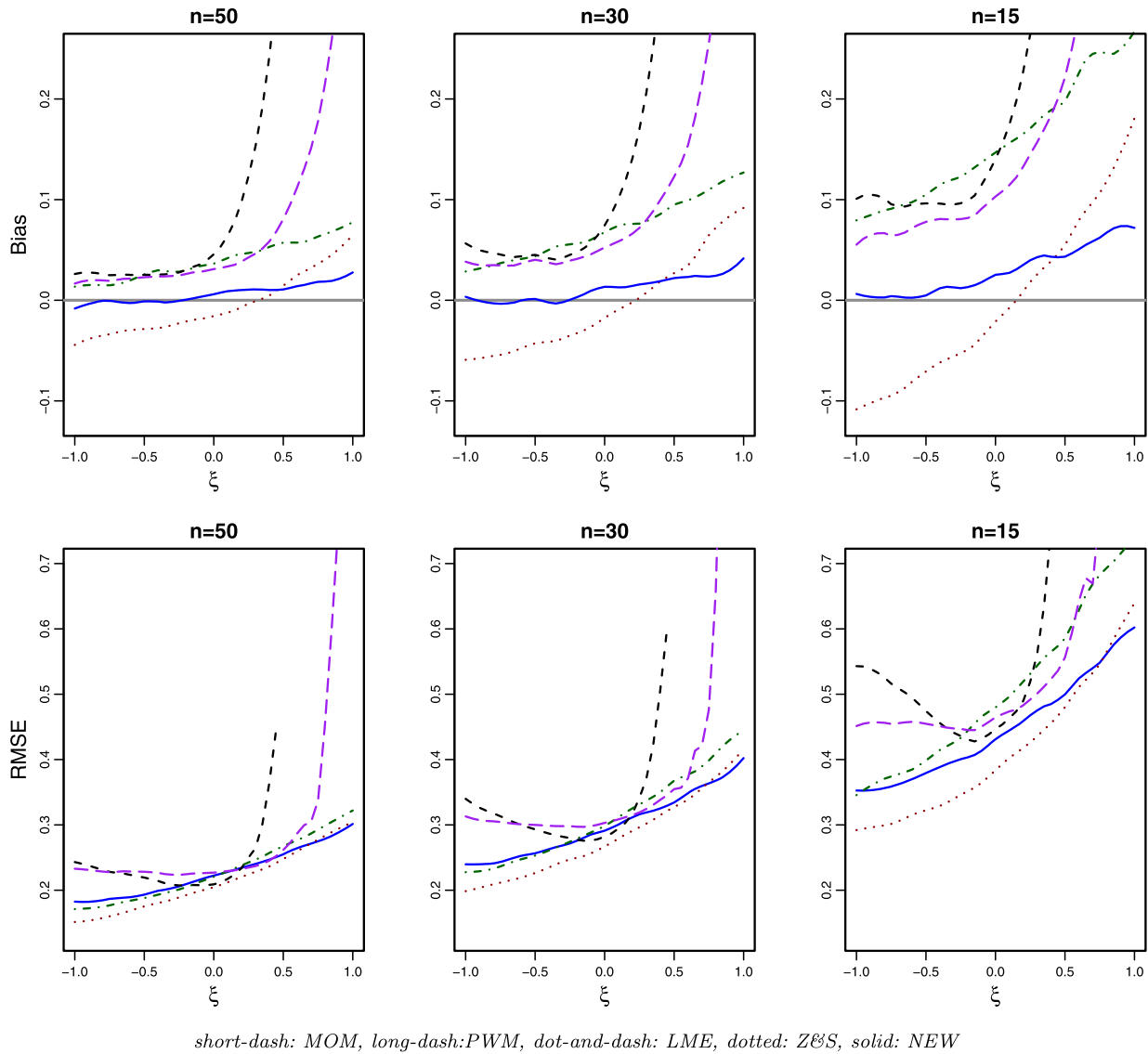


Figure 1. Bias and RMSE of the scale estimator  $\hat{\sigma}$  plotted against  $\xi$  (from  $-1$  to  $1$ ).

formance of different methods for quantile estimations using the same simulation data and results as in Section 3.1. In the literature, it is often assumed that good parameter estimators of the GPD, and *vice versa*. However, through simulation study we will show that this statement is arguably vague. The interpretation of the results of quantile estimation with different methods is somewhat more subtle. For all tested approaches, the performance of quantile estimators are poor for extreme upper quantiles of heavy-tailed GPD. In order to compare the results under various quantile levels ( $p = \{0.5, 0.75, 0.9\}$ ) and shape parameters ( $\xi = \{-1, -0.5, -0.25, 0.025, 0.5, 0.75, 1\}$ ), percentage-bias and percentage-RMSE for different methods are given in Tables 3 to 5. The percentage-bias and percentage-RMSE are defined as the ratio of estimation bias and RMSE over the

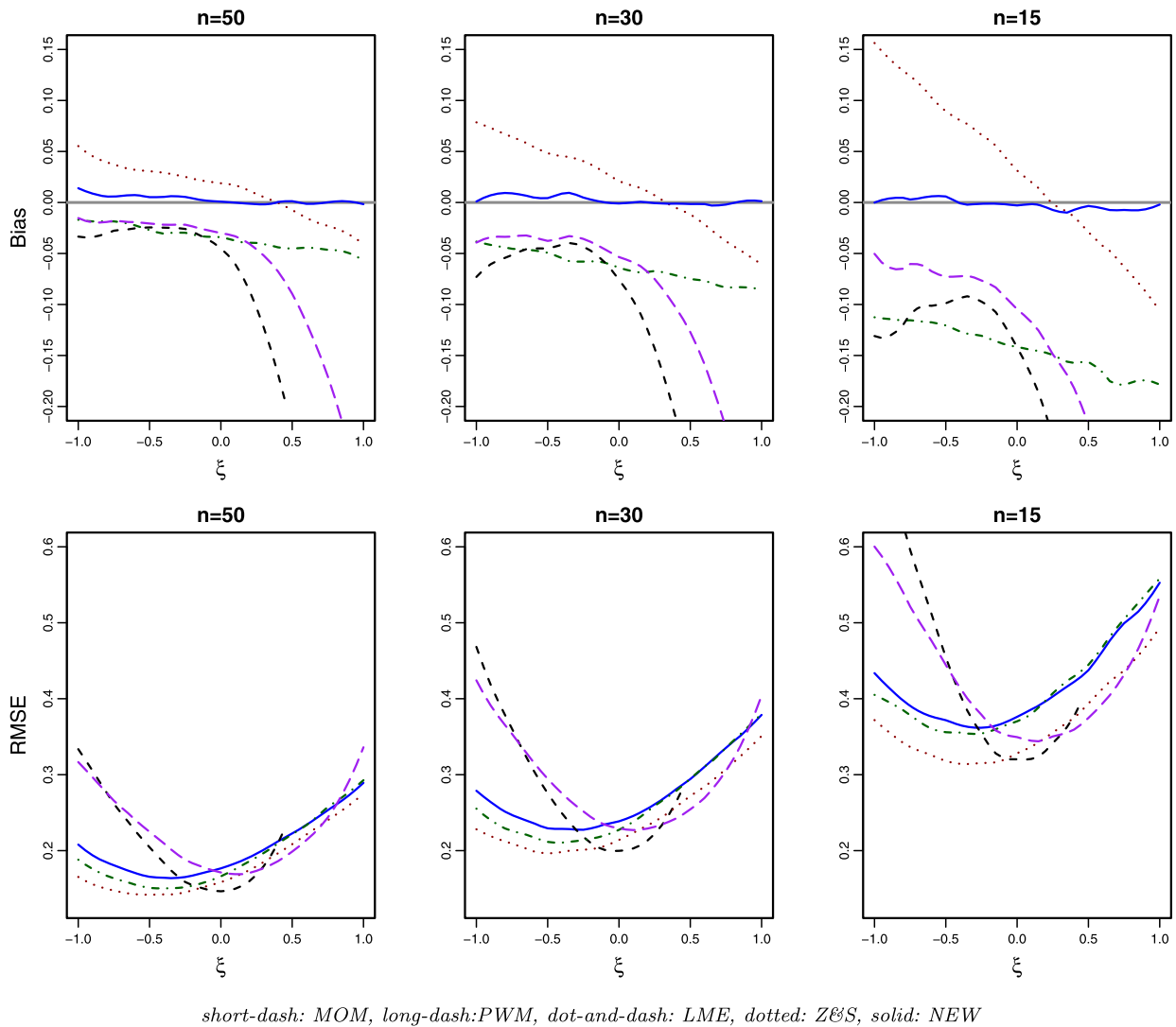
theoretical quantile value, respectively:

$$\text{Percentage-Bias} = \frac{\text{Bias}[\hat{Q}(p; \hat{\sigma}, \hat{\xi})]}{Q(p; \sigma, \xi)},$$

$$\text{Percentage-RMSE} = \frac{\text{RMSE}[\hat{Q}(p; \hat{\sigma}, \hat{\xi})]}{Q(p; \sigma, \xi)}.$$

We also plotted the percentage-bias and percentage-RMSE of 0.5, 0.75 and 0.9 quantiles of the MOM, PWM, LME, Z&S and NEW methods with sample size  $n = 30$  against the value of  $\xi$  in Figure 3.

The simulations seem to show that, when estimating less-extreme quantiles of the GPD given different  $\xi$  values, such as  $Q(0.5)$  and  $Q(0.75)$ , the NEW method is still the least sensitive to the changes of  $\xi$  and sample size, although the



short-dash: MOM, long-dash: PWM, dot-and-dash: LME, dotted: Z&S, solid: NEW

Figure 2. Bias and RMSE of the shape estimator  $\hat{\xi}$  plotted against  $\xi$  (from  $-1$  to  $1$ ).

NEW method does not always have the smallest percentage-bias. The MOM, PWM and LME methods have smaller percentage-bias when estimating quantiles of very ‘short-tailed’ GPD ( $-1 \leq \xi \leq -0.5$ ), but their bias substantially increase as the tail becomes heavier and the sample size becomes smaller. On the other hand, the Z&S method has the smallest overall RMSE of parameter estimation, but that does not give it the same advantage when estimating the quantiles. The differences in percentage-RMSE of estimating  $Q(0.5)$  and  $Q(0.75)$  among LME, Z&S and NEW are only marginal. For example, for  $\xi = 1$  and  $p = 0.5$  or  $p = 0.75$ , the NEW method has the smallest percentage-RMSE. For 0.5 and 0.75 quantiles, the quantile estimators of NEW are generally the most reliable.

However, for estimating the higher upper quantiles that close to 1, such as  $Q(0.9)$ , performance of the NEW method drops dramatically as its percentage-bias increases rapidly

as the tail becomes heavier. The performance of LME and Z&S’ quantile estimators also drop significantly. Their percentage-bias and percentage-RMSE almost doubled compared with the 0.75 quantile cases, as seen in Tables 4 and 5.

The numerical evidence and analysis seemed to suggest that the performance of different quantile estimators may not be consistent when estimating different quantiles. [14] show that, by Taylor expansion,  $Q(p) = \sigma p (1 + \frac{1}{2}(1 + \xi)p + O(p^2))$ , as  $p \rightarrow 0$ , thus the accuracy of  $\hat{Q}(p)$  for small  $p$  is mainly affected by the accuracy of  $\hat{\sigma}$ . For large  $p$ , the mechanic is rather unclear. Hence, we considered the following subsequent tests. Figure 4 presented the percentage-bias and percentage-RMSE of the quantile estimators of the five methods, given sample size  $n = 30$  and fixed values of  $\xi = \{0.25, 0.5, 0.75, 1\}$  (for each row of figures, respectively), plotted against varying values of  $p$ , ranged from 0.5 to 0.95 by 0.01 interval. Results were briefly

Table 1. Bias and RMSE of  $\hat{\sigma}$

n	$\xi$	Bias					RMSE				
		MOM	PWM	LME	Z&S	NEW	MOM	PWM	LME	Z&S	NEW
50	-1	0.024	0.015	0.012	-0.045	-0.008	0.242	0.232	0.171	0.152	0.183
	-0.5	0.020	0.017	0.026	-0.031	0.002	0.217	0.226	0.187	0.178	0.194
	-0.25	0.030	0.027	0.024	-0.020	-0.007	0.205	0.222	0.200	0.187	0.205
	0	0.046	0.032	0.033	-0.016	0.009	0.211	0.230	0.220	0.206	0.222
	0.25	0.121	0.040	0.048	-0.006	0.006	0.249	0.235	0.242	0.226	0.233
	0.5	—	0.083	0.056	0.016	0.009	—	0.264	0.267	0.248	0.255
	0.75	—	0.182	0.064	0.041	0.017	—	0.356	0.292	0.282	0.274
	1	—	—	0.079	0.065	0.029	—	—	0.323	0.304	0.301
30	-1	0.060	0.042	0.028	-0.058	0.003	0.346	0.318	0.227	0.200	0.239
	-0.5	0.050	0.045	0.042	-0.037	0.006	0.291	0.298	0.252	0.223	0.254
	-0.25	0.045	0.039	0.058	-0.035	0.002	0.277	0.295	0.275	0.246	0.270
	0	0.076	0.054	0.069	-0.017	0.014	0.285	0.306	0.301	0.270	0.292
	0.25	0.171	0.074	0.075	0.002	0.015	0.339	0.319	0.328	0.295	0.312
	0.5	—	0.124	0.089	0.029	0.030	—	0.357	0.361	0.331	0.337
	0.75	—	0.261	0.108	0.066	0.032	—	0.614	0.394	0.364	0.365
	1	—	—	0.129	0.091	0.042	—	—	0.442	0.415	0.402
15	-1	0.103	0.057	0.082	-0.107	0.006	0.543	0.451	0.348	0.290	0.352
	-0.5	0.098	0.079	0.101	-0.070	0.004	0.480	0.458	0.397	0.324	0.378
	-0.25	0.088	0.072	0.127	-0.059	0.014	0.429	0.440	0.431	0.346	0.395
	0	0.145	0.107	0.142	-0.017	0.023	0.456	0.473	0.479	0.386	0.429
	0.25	0.266	0.147	0.171	0.018	0.041	0.522	0.490	0.534	0.423	0.463
	0.5	—	0.223	0.192	0.058	0.035	—	0.564	0.570	0.479	0.489
	0.75	—	0.400	0.238	0.102	0.064	—	0.836	0.664	0.547	0.553
	1	—	—	0.267	0.181	0.070	—	—	0.745	0.646	0.602

Table 2. Bias and RMSE of  $\hat{\xi}$

n	$\xi$	Bias					RMSE				
		MOM	PWM	LME	Z&S	NEW	MOM	PWM	LME	Z&S	NEW
50	-1	-0.031	-0.013	-0.015	0.057	0.014	0.332	0.315	0.188	0.167	0.209
	-0.5	-0.019	-0.015	-0.029	0.033	0.005	0.201	0.222	0.150	0.144	0.166
	-0.25	-0.027	-0.022	-0.027	0.024	0.011	0.157	0.186	0.152	0.143	0.166
	0	-0.045	-0.031	-0.029	0.019	0.000	0.145	0.171	0.165	0.157	0.177
	0.25	-0.103	-0.046	-0.039	0.009	-0.003	0.168	0.173	0.189	0.181	0.196
	0.5	—	-0.089	-0.044	-0.007	0.008	—	0.202	0.223	0.210	0.224
	0.75	—	-0.173	-0.046	-0.027	0.003	—	0.247	0.255	0.241	0.254
	1	—	—	-0.057	-0.039	-0.004	—	—	0.295	0.277	0.290
30	-1	-0.077	-0.043	-0.038	0.077	0.001	0.475	0.429	0.255	0.229	0.278
	-0.5	-0.047	-0.040	-0.048	0.044	0.002	0.271	0.290	0.211	0.191	0.226
	-0.25	-0.041	-0.034	-0.062	0.042	0.005	0.215	0.249	0.217	0.199	0.225
	0	-0.079	-0.057	-0.063	0.018	-0.003	0.203	0.233	0.228	0.217	0.239
	0.25	-0.144	-0.077	-0.066	0.007	0.001	0.224	0.235	0.259	0.241	0.258
	0.5	—	-0.128	-0.068	-0.012	-0.005	—	0.250	0.298	0.271	0.295
	0.75	—	-0.220	-0.081	-0.038	-0.011	—	0.303	0.334	0.305	0.336
	1	—	—	-0.090	-0.060	-0.001	—	—	0.383	0.350	0.377
15	-1	-0.133	-0.052	-0.116	0.155	-0.001	0.742	0.601	0.408	0.370	0.431
	-0.5	-0.104	-0.078	-0.119	0.084	0.007	0.461	0.448	0.357	0.317	0.372
	-0.25	-0.092	-0.071	-0.137	0.072	-0.001	0.349	0.375	0.355	0.316	0.361
	0	-0.144	-0.106	-0.138	0.028	0.000	0.321	0.350	0.371	0.328	0.377
	0.25	-0.230	-0.150	-0.144	-0.001	-0.012	0.342	0.354	0.412	0.361	0.402
	0.5	—	-0.226	-0.148	-0.040	-0.003	—	0.381	0.436	0.391	0.433
	0.75	—	-0.320	-0.182	-0.063	-0.019	—	0.430	0.500	0.435	0.498
	1	—	—	-0.177	-0.109	0.001	—	—	0.557	0.492	0.555



Table 3. Percentage-bias and Percentage-RMSE of estimating  $Q(0.5; 1, \xi)$

n	$\xi$	Percentage-bias					Percentage-RMSE				
		MOM	PWM	LME	Z&S	NEW	MOM	PWM	LME	Z&S	NEW
50	-1	-0.001	-0.004	0.000	-0.034	-0.012	0.133	0.132	0.113	0.107	0.121
	-0.5	0.004	0.001	0.010	-0.027	-0.005	0.151	0.155	0.139	0.138	0.145
	-0.25	0.014	0.011	0.008	-0.018	-0.011	0.157	0.165	0.156	0.151	0.161
	0	0.025	0.013	0.016	-0.016	0.001	0.173	0.182	0.177	0.171	0.177
	0.25	0.078	0.017	0.027	-0.009	-0.002	0.209	0.196	0.198	0.191	0.192
	0.5	—	0.044	0.031	0.007	0.004	—	0.220	0.218	0.211	0.216
	0.75	—	0.107	0.038	0.023	0.010	—	0.303	0.241	0.240	0.234
	1	—	—	0.046	0.043	0.019	—	—	0.269	0.262	0.259
30	-1	0.007	0.003	0.004	-0.045	-0.011	0.171	0.170	0.146	0.140	0.154
	-0.5	0.017	0.013	0.014	-0.033	-0.007	0.192	0.196	0.183	0.171	0.186
	-0.25	0.018	0.012	0.024	-0.032	-0.010	0.205	0.213	0.207	0.196	0.208
	0	0.038	0.020	0.034	-0.021	-0.001	0.224	0.234	0.233	0.220	0.231
	0.25	0.109	0.035	0.038	-0.006	0.002	0.280	0.257	0.260	0.247	0.255
	0.5	—	0.066	0.049	0.012	0.014	—	0.297	0.287	0.281	0.279
	0.75	—	0.162	0.062	0.040	0.013	—	0.586	0.320	0.311	0.305
	1	—	—	0.076	0.056	0.026	—	—	0.358	0.359	0.343
15	-1	0.001	-0.007	0.018	-0.084	-0.026	0.230	0.230	0.206	0.204	0.219
	-0.5	0.022	0.013	0.032	-0.066	-0.025	0.275	0.278	0.266	0.246	0.264
	-0.25	0.029	0.016	0.050	-0.059	-0.017	0.290	0.299	0.301	0.271	0.289
	0	0.070	0.041	0.063	-0.030	-0.007	0.327	0.336	0.345	0.306	0.327
	0.25	0.161	0.068	0.084	-0.005	0.007	0.411	0.371	0.394	0.345	0.362
	0.5	—	0.114	0.104	0.019	0.003	—	0.443	0.440	0.395	0.392
	0.75	—	0.246	0.127	0.056	0.023	—	0.741	0.499	0.466	0.445
	1	—	—	0.154	0.114	0.035	—	—	0.574	0.549	0.490

Table 4. Percentage-bias and Percentage-RMSE of estimating  $Q(0.75; 1, \xi)$

n	$\xi$	Percentage-bias					Percentage-RMSE				
		MOM	PWM	LME	Z&S	NEW	MOM	PWM	LME	Z&S	NEW
50	-1	-0.009	-0.008	-0.005	-0.023	-0.011	0.073	0.074	0.072	0.072	0.076
	-0.5	-0.005	-0.006	-0.001	-0.020	-0.006	0.107	0.108	0.104	0.106	0.107
	-0.25	0.003	0.000	-0.003	-0.013	-0.011	0.125	0.126	0.124	0.123	0.127
	0	0.007	0.000	0.003	-0.011	-0.002	0.150	0.152	0.150	0.149	0.147
	0.25	0.038	-0.001	0.011	-0.007	-0.005	0.190	0.176	0.177	0.174	0.172
	0.5	—	0.008	0.014	0.005	0.009	—	0.207	0.202	0.204	0.208
	0.75	—	0.033	0.021	0.015	0.016	—	0.289	0.238	0.241	0.241
	1	—	—	0.028	0.034	0.026	—	—	0.283	0.281	0.282
30	-1	-0.010	-0.008	-0.008	-0.031	-0.015	0.093	0.094	0.093	0.093	0.096
	-0.5	-0.001	-0.003	-0.004	-0.025	-0.012	0.133	0.134	0.135	0.131	0.136
	-0.25	0.001	-0.004	0.000	-0.023	-0.014	0.161	0.163	0.163	0.159	0.163
	0	0.008	-0.003	0.008	-0.019	-0.006	0.190	0.193	0.192	0.188	0.191
	0.25	0.054	0.004	0.011	-0.005	0.000	0.253	0.229	0.227	0.227	0.229
	0.5	—	0.016	0.023	0.009	0.014	—	0.277	0.267	0.271	0.268
	0.75	—	0.069	0.034	0.031	0.015	—	0.606	0.316	0.317	0.314
	1	—	—	0.047	0.043	0.041	—	—	0.371	0.386	0.380
15	-1	-0.023	-0.019	-0.013	-0.056	-0.033	0.130	0.134	0.128	0.136	0.138
	-0.5	-0.014	-0.017	-0.010	-0.052	-0.034	0.187	0.190	0.190	0.186	0.188
	-0.25	-0.007	-0.015	-0.002	-0.045	-0.027	0.223	0.227	0.226	0.220	0.221
	0	0.016	-0.003	0.007	-0.028	-0.016	0.268	0.269	0.275	0.261	0.267
	0.25	0.072	0.009	0.026	-0.009	-0.002	0.361	0.320	0.332	0.315	0.319
	0.5	—	0.025	0.047	0.006	0.004	—	0.402	0.401	0.382	0.373
	0.75	—	0.111	0.059	0.043	0.030	—	0.720	0.468	0.487	0.465
	1	—	—	0.098	0.095	0.067	—	—	0.586	0.615	0.562

Table 5. Percentage-bias and Percentage-RMSE of estimating  $Q(0.9; 1, \xi)$

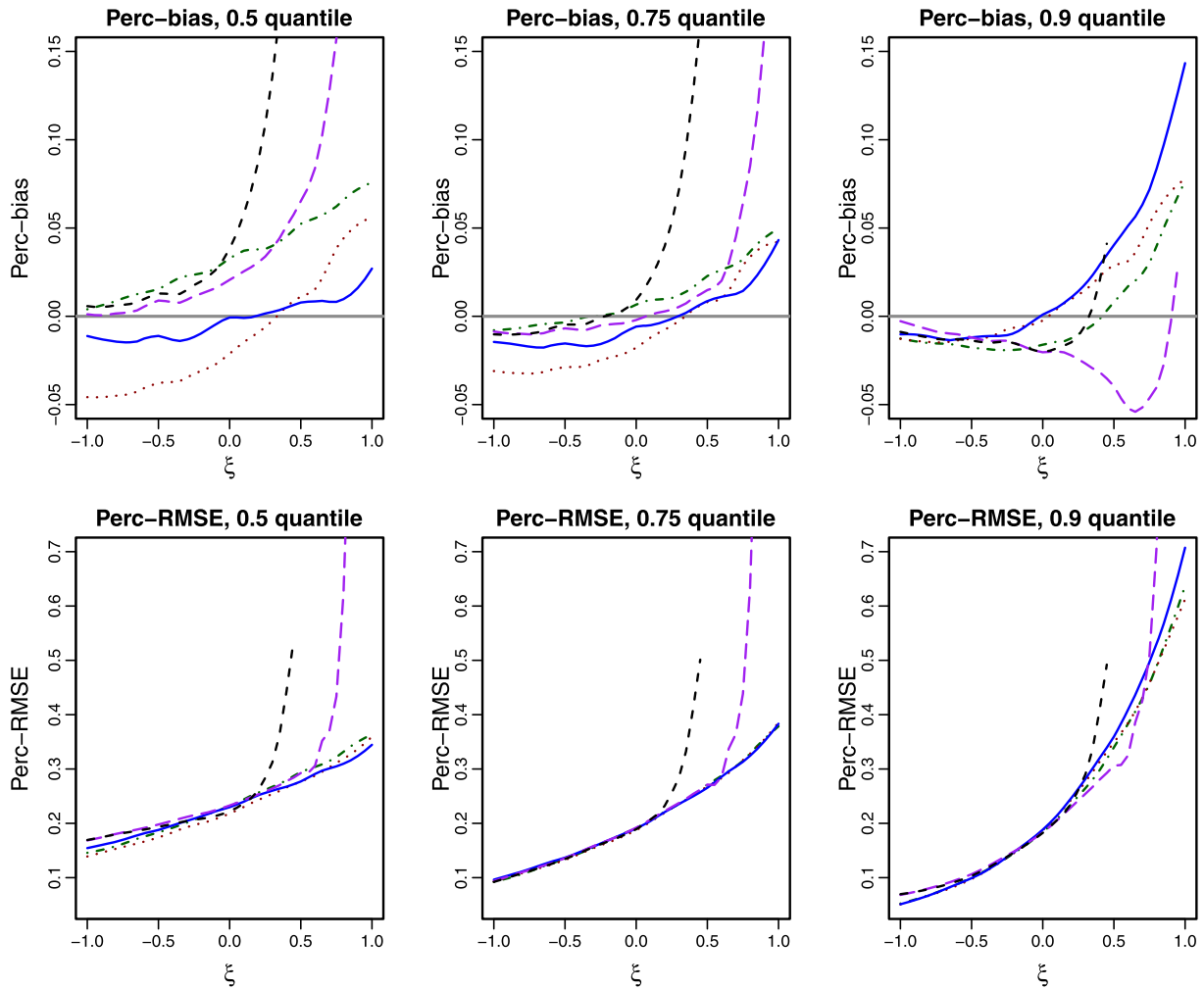
n	$\xi$	Percentage-bias					Percentage-RMSE				
		MOM	PWM	LME	Z&S	NEW	MOM	PWM	LME	Z&S	NEW
50	-1	-0.006	-0.003	-0.007	-0.010	-0.006	0.053	0.052	0.039	0.040	0.039
	-0.5	-0.008	-0.007	-0.010	-0.010	-0.004	0.081	0.084	0.076	0.077	0.076
	-0.25	-0.006	-0.006	-0.013	-0.003	-0.005	0.107	0.109	0.105	0.105	0.106
	0	-0.011	-0.011	-0.007	0.000	0.002	0.144	0.144	0.144	0.146	0.143
	0.25	-0.009	-0.018	-0.001	0.005	0.003	0.198	0.188	0.196	0.197	0.196
	0.5	—	-0.031	0.005	0.018	0.033	—	0.241	0.253	0.261	0.278
	0.75	—	-0.057	0.022	0.025	0.052	—	0.334	0.337	0.337	0.360
	1	—	—	0.038	0.055	0.077	—	—	0.447	0.444	0.465
30	-1	-0.009	-0.003	-0.013	-0.013	-0.010	0.070	0.069	0.052	0.051	0.051
	-0.5	-0.011	-0.009	-0.018	-0.010	-0.010	0.103	0.106	0.100	0.097	0.099
	-0.25	-0.012	-0.012	-0.021	-0.005	-0.009	0.138	0.141	0.139	0.136	0.138
	0	-0.022	-0.023	-0.014	-0.005	-0.001	0.184	0.184	0.184	0.188	0.188
	0.25	-0.010	-0.024	-0.009	0.013	0.017	0.259	0.243	0.249	0.262	0.266
	0.5	—	-0.040	0.014	0.029	0.043	—	0.311	0.346	0.350	0.363
	0.75	—	-0.041	0.034	0.054	0.067	—	0.695	0.462	0.458	0.498
	1	—	—	0.071	0.078	0.137	—	—	0.628	0.609	0.702
15	-1	-0.016	-0.005	-0.027	-0.017	-0.022	0.102	0.100	0.081	0.076	0.081
	-0.5	-0.030	-0.025	-0.040	-0.020	-0.023	0.152	0.155	0.146	0.144	0.142
	-0.25	-0.031	-0.030	-0.044	-0.010	-0.016	0.197	0.200	0.195	0.199	0.196
	0	-0.035	-0.036	-0.037	0.001	0.004	0.257	0.257	0.264	0.273	0.276
	0.25	-0.024	-0.044	-0.013	0.025	0.032	0.360	0.332	0.368	0.385	0.395
	0.5	—	-0.067	0.021	0.039	0.072	—	0.441	0.512	0.516	0.549
	0.75	—	-0.038	0.045	0.106	0.154	—	0.777	0.685	0.777	0.896
	1	—	—	0.160	0.192	0.301	—	—	1.141	1.191	1.328

summarised as follows.

First, for all tested methods, accurately estimating extreme upper quantiles is obviously more difficult beyond certain  $p$  thresholds. The slopes of estimation bias and error curves increase steeply when the  $p$  value is close to 1, even more so for heavier tails. A possible explanation for this is that for small sample sizes such as  $n \leq 50$ , the estimated extreme upper quantiles seem to be heavily affected by the largest sample point(s). Second, as pointed out by one of the reviewers, better performance on the estimation of  $\xi$  does not necessarily translate into better performance on the estimation of the upper quantiles, which is opposite to the default intuitive belief. For example, both of the Z&S and NEW methods have larger bias than the LME method when estimating higher upper quantiles with  $\xi = 0.25, 0.5, 0.75$ , although the LME method is obviously more biased on both parameter estimators than the former two. Furthermore, when estimating less extreme quantiles which are closer to the center of the distribution, the NEW estimator generally outperforms the others, with consistently smaller percentage-bias and RMSE. However, the performance of the NEW quantile estimator drops dramatically as  $p$  is close to or larger than 0.8.

The phenomenon stated above does raise some interesting questions that are worth further investigation. In literature, the POT approach is sometimes used for ‘extrapolating’ extreme quantiles that are beyond the sample range. Despite

general belief that the results under the EVT framework are usually better than results from conventional methods (*e.g.*, estimating the distribution of the entire dataset instead of fitting the GPD to the tail), extra caution should be paid when the sample size for fitting the GPD is small. As discussed in this section, the methods we tested all have significant performance loss when estimating higher upper quantiles of extremely heavy-tailed GPD, such as  $\xi \geq 0.5$  (namely, the GPD has infinite variance). Recall that  $p = (\frac{N}{n}p^*)$ , where  $p^*$  is the extreme probability or quantile level of the original distribution,  $F_Y$ , and  $N$  is the sample size of the original data.  $p^*$  and  $N$  are often predetermined, but the value of  $p$  can still be adjusted by changing the threshold value  $u$  (thus changing  $n$ ). From what we have observed, when  $p^*$  and  $N$  are fixed, raising the threshold  $u$  will lead to less-extreme  $p$ , which may reduce the bias and error of  $\hat{Q}(p)$ . However, this partially contradicts the well-known ‘bias-variance-tradeoff’ effect of the threshold selection, which, in general, suggests that lower thresholds will lead to larger bias but smaller variance, and higher thresholds will lead to smaller bias but larger variance. To answer these questions would probably require comparing the POT approach based on different methods with classic approaches and cross-examining different sample sizes and  $p$  values. Such problems are beyond the main scope of this paper. However, it can be concluded that choosing a threshold which leads to a moderate  $p$  value (*i.e.*,  $p \leq 0.8$ ) will benefit



short-dash: MOM, long-dash: PWM, dot-and-dash: LME, dotted: Z&S, solid: NEW

Figure 3. Percentage-bias and percentage-RMSE of estimating  $Q(p; 1, \xi)$ , plotted against  $\xi$ , for different value of  $p = 0.5, 0.75, 0.9$ , respectively; sample size  $n = 30$ .

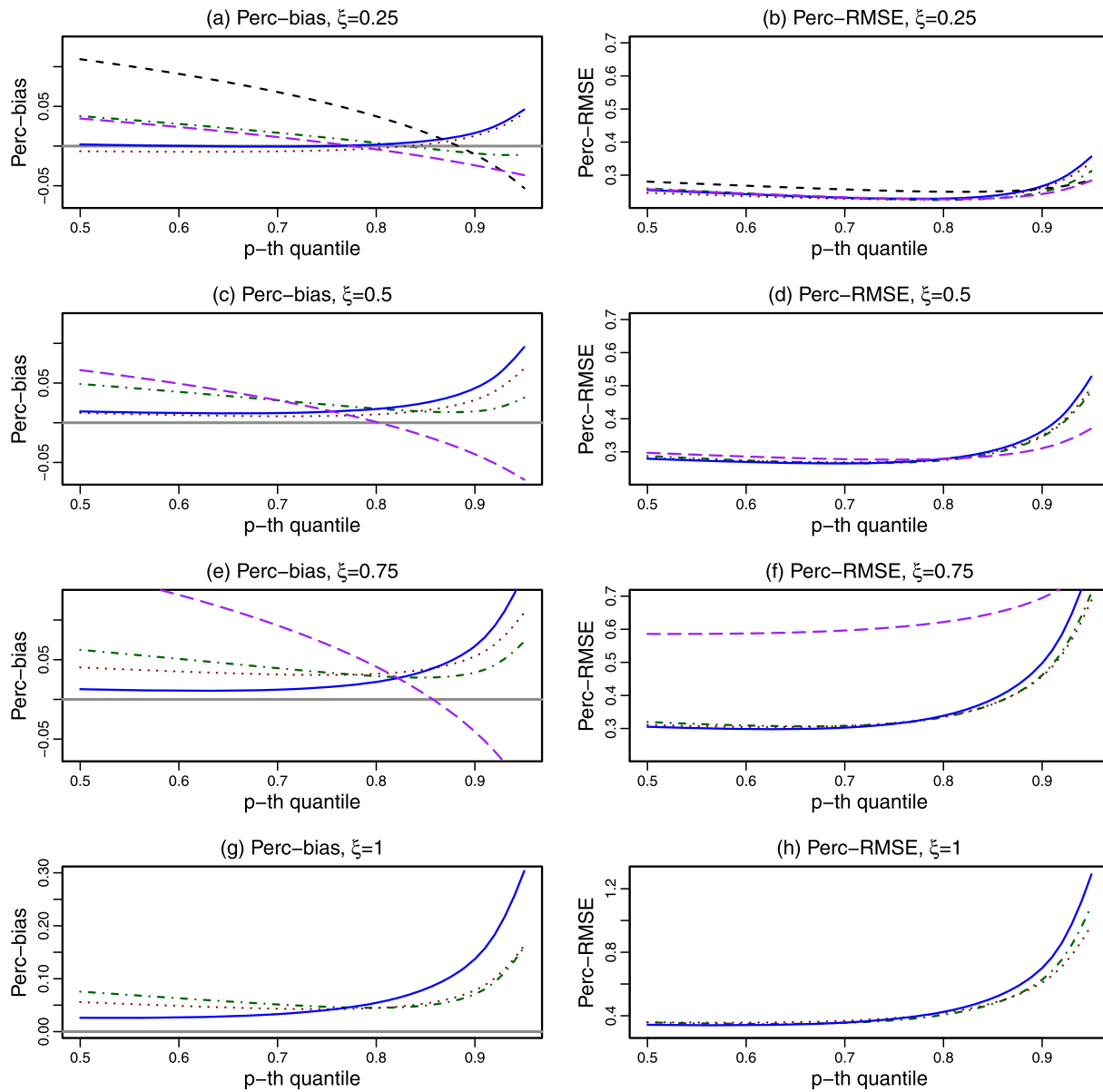
the overall performance of quantile estimation of the GPD. In particular, the NEW method we proposed has displayed promising properties such as versatility under these circumstances.

### 3.3 Confidence interval estimation

In practice, confidence interval not only measures the uncertainty of unknown quantities such as quantiles, but also provides valuable tail-information in modelling GPD. In this section we compare the performance of different confidence interval methods in terms of average length and coverage probability. These methods include the profile log-likelihood method based on MLE estimates (Profile), the asymptotic CI based on observed Fisher information of PWM and LME methods, the generalized CI of NEW method and two bootstrap methods: the percentile bootstrap confidence intervals (PERC) and the bias-corrected and accelerated ( $BC_a$ ) con-

fidence intervals. Details of both bootstrap methods can be found in [10] or seen in [23] for a more specific study of the bootstrap CI for heavy-tailed GPD.

We mainly focused on CI of quantile estimation of the GPD, which included 90% and 95% CI of  $Q(p)$ , where  $p = \{0.75, 0.9\}$ , respectively. CI for median  $Q(0.5)$  was not reported because similar conclusions can be drawn from the results. Random samples of different sizes ( $n = 50, 30$ ) were generated from  $\sigma = 1$  with different values of shape parameter  $\xi = \{-0.25, 0.25, 0.5, 0.75\}$ , under 1000 replicates. For the NEW estimates,  $\Theta_{(\tau)}$  was calculated based on 2000 Monte Carlo random samples; the bootstrap methods were calculated based on 1000 bootstrap samples. For each method, the average interval length and coverage probability are summarized in Tables 6 and 7. It is worth mentioning that CI estimation for sample size  $n = 15$  is not reported, due to tested methods that all failed to provide



short-dash: MOM, long-dash: PWM, dot-and-dash: LME, dotted: ZES, solid: NEW

Figure 4. Percentage-bias and percentage-RMSE of estimating  $Q(p; 1, \xi)$ , plotted against  $p$  ( $0.5 \leq p \leq 0.95$ ), for heavy-tailed GPD with different values of  $\xi = 0.25, 0.5, 0.75, 1$ , respectively; sample size  $n = 30$ .

reliable results in terms of coverage probability or average interval length.

For the same reasons stated in the previous section, we considered it important that the CI estimates should be consistently reliable for different tail behaviours (different  $\xi$  values). Also, ideally a good CI estimation should have the shortest interval length, combined with an accurate coverage probability (CP) close to the nominal  $(1 - \beta)\%$  confidence level, but as little above  $(1 - \beta)\%$  as possible ([16]).

Table 6 presents results for the less-extreme 75% quantile of the GPD. The average CI length of all tested methods

increases as sample size decreases. Overall, the generalized CIs provided by the NEW method have the most reliable good performance for different value of  $\xi$ s. The CP based on PWM estimates is always lower than the nominal level, decreasing as value of  $\xi$  increases. On the contrary, the CP based on LME estimates is obviously much higher than the nominal level when the value of  $\xi$  is smaller. But the CP of LME-based CI estimates decreases as value of  $\xi$  increases, and for the 90% and 95% CI of  $Q(0.9)$ , it coincidentally becomes very close to the nominal level. For the  $Q(0.75)$  cases, CIs based on the profile log-likelihood method have

Table 6. Average length and coverage probability (in parentheses) of confidence intervals of estimating  $Q(0.75; 1, \xi)$

n	$\xi$	90% CI						95% CI					
		Profile	PWM	LME	NEW	PERC	$BC_a$	Profile	PWM	LME	NEW	PERC	$BC_a$
50	-0.25	0.474 (0.893)	0.482 (0.897)	1.066 (0.997)	0.505 (0.904)	0.521 (0.886)	0.527 (0.894)	0.566 (0.922)	0.574 (0.926)	1.263 (0.997)	0.603 (0.95)	0.623 (0.947)	0.631 (0.951)
	0.25	0.961 (0.881)	0.926 (0.868)	1.703 (0.989)	1.016 (0.89)	0.963 (0.894)	0.982 (0.891)	1.170 (0.937)	1.110 (0.924)	2.038 (0.996)	1.235 (0.947)	1.165 (0.945)	1.193 (0.953)
	0.5	1.373 (0.883)	1.212 (0.859)	2.228 (0.98)	1.451 (0.897)	1.371 (0.896)	1.414 (0.894)	1.712 (0.93)	1.457 (0.905)	2.700 (0.986)	1.804 (0.938)	*** (0.943)	*** (0.948)
	0.75	2.030 (0.885)	1.644 (0.788)	3.054 (0.97)	2.154 (0.896)	*** (0.908)	*** (0.901)	2.490 (0.942)	1.864 (0.869)	3.621 (0.987)	2.633 (0.943)	*** (0.929)	*** (0.931)
30	-0.25	0.612 (0.868)	0.611 (0.874)	1.365 (0.998)	0.660 (0.913)	0.697 (0.909)	0.702 (0.897)	0.730 (0.935)	0.729 (0.933)	1.632 (0.999)	0.793 (0.957)	0.836 (0.948)	0.844 (0.943)
	0.25	1.245 (0.889)	1.169 (0.884)	2.184 (0.984)	1.371 (0.918)	1.303 (0.898)	1.348 (0.896)	1.569 (0.933)	1.410 (0.919)	2.644 (0.99)	1.732 (0.948)	1.576 (0.949)	1.640 (0.950)
	0.5	1.849 (0.900)	1.538 (0.861)	2.901 (0.980)	2.031 (0.917)	1.816 (0.890)	1.908 (0.886)	2.292 (0.947)	1.756 (0.885)	3.405 (0.986)	2.531 (0.958)	2.171 (0.941)	2.310 (0.932)
	0.75	2.796 (0.901)	1.842 (0.774)	3.989 (0.967)	3.107 (0.908)	*** (0.86)	*** (0.853)	3.400 (0.953)	2.292 (0.836)	4.718 (0.982)	3.911 (0.955)	*** (0.935)	*** (0.930)

\*\*\* denotes value larger than  $1 \times 10^5$

Table 7. Average length and coverage probability (in parentheses) of confidence intervals of estimating  $Q(0.9; 1, \xi)$

n	$\xi$	90% CI						95% CI					
		Profile	PWM	LME	NEW	PERC	$BC_a$	Profile	PWM	LME	NEW	PERC	$BC_a$
50	-0.25	0.594 (0.864)	0.616 (0.866)	1.920 (1.000)	0.713 (0.897)	0.572 (0.858)	0.582 (0.878)	0.723 (0.913)	0.736 (0.913)	2.283 (1.000)	0.897 (0.952)	0.684 (0.916)	0.699 (0.923)
	0.25	2.243 (0.876)	1.976 (0.845)	3.869 (0.994)	2.539 (0.898)	1.841 (0.84)	2.011 (0.843)	2.810 (0.936)	2.368 (0.897)	4.567 (0.998)	3.174 (0.953)	2.190 (0.885)	2.466 (0.899)
	0.5	3.643 (0.902)	3.413 (0.787)	5.881 (0.973)	4.700 (0.877)	3.601 (0.846)	4.202 (0.863)	4.416 (0.938)	4.377 (0.849)	7.153 (0.986)	6.431 (0.942)	4.121 (0.886)	5.027 (0.917)
	0.75	4.587 (0.847)	4.898 (0.613)	9.720 (0.947)	9.525 (0.881)	*** (0.842)	*** (0.849)	5.011 (0.833)	6.011 (0.663)	11.350 (0.957)	12.355 (0.943)	*** (0.882)	*** (0.901)
30	-0.25	0.764 (0.849)	0.780 (0.842)	2.430 (0.998)	1.031 (0.900)	0.708 (0.823)	0.718 (0.841)	0.989 (0.905)	0.937 (0.897)	2.927 (1.000)	1.389 (0.946)	0.855 (0.881)	0.872 (0.889)
	0.25	3.011 (0.874)	2.489 (0.821)	5.003 (0.982)	3.950 (0.892)	2.330 (0.81)	2.642 (0.827)	3.662 (0.93)	2.917 (0.868)	5.848 (0.991)	5.271 (0.944)	2.701 (0.833)	3.140 (0.865)
	0.5	4.151 (0.862)	4.271 (0.743)	7.792 (0.957)	8.143 (0.903)	24.477 (0.82)	2352.063 (0.838)	4.730 (0.911)	4.706 (0.820)	9.244 (0.972)	11.586 (0.956)	5.660 (0.862)	7.650 (0.876)
	0.75	4.725 (0.710)	5.873 (0.610)	13.305 (0.919)	18.083 (0.869)	84.915 (0.772)	*** (0.811)	5.328 (0.784)	7.032 (0.700)	15.706 (0.954)	26.350 (0.947)	16.986 (0.851)	93.397 (0.88)

\*\*\* denotes value larger than  $1 \times 10^5$

comparable performance with the NEW method, and are even slightly better in terms of average interval length. For the  $Q(0.9)$  cases, however, the profile log-likelihood based CI estimates have obvious insufficient CPs.

Regarding bootstrap-method-based (PERC and  $BC_a$ ) CIs for  $Q(0.75)$ , their CPs are close to the nominal level in most situations, but their interval lengths are very misleading when the GPD is extremely heavy-tailed, such as  $\xi = 0.5, 0.75$ . Extremely large average interval length ( $> 1 \times 10^5$ ) have been observed for both bootstrap methods, as denoted with '\*\*\*' in tables. For small sample sizes and extremely heavy-tailed distributions, estimations from naive bootstrap

methods are dominated by the largest sample point (or several extremely large sample points) and may not converge. For more detailed analysis of naive bootstrap failure potential in heavy-tailed cases, please refer to [1] and [13]. These extremely wide confidence intervals have basically no practical use.

For the CIs of higher quantiles  $Q(0.9)$ , the CPs from profile log-likelihood and PWM drop significantly as the tail of the GPD gets heavier. The LME-based CI is obviously overestimated, even CP=1 when  $\xi = -0.25$ . The bootstrap methods have the same convergence problem for heavy-tailed GPD, as shown in previous case, and their CPs

Table 8. Comparison of goodness-of-fit of Florida flood claims data for GPD with four-method

	$\hat{\sigma}$	$\hat{\xi}$	$W^2$	Asymptotic $p$ values	Bootstrap $p$ value	$A^2$	Asymptotic $p$ values	Bootstrap $p$ value
<i>Threshold</i>								
<i>u = 26.302</i>								
NEW	9.210	0.727	0.038	0.939	0.942	0.278	0.939	0.932
PWM	10.830	0.524	0.069	0.758	0.762	0.466	0.758	0.761
LME	9.950	0.649	0.049	0.879	0.875	0.341	0.879	0.883
Z&S	9.436	0.703	0.040	0.926	0.937	0.292	0.926	0.918
<i>Threshold</i>								
<i>u = 32.527</i>								
NEW	13.312	0.701	0.041	0.919	0.924	0.326	0.919	0.894
PWM	16.033	0.469	0.057	0.827	0.832	0.499	0.827	0.711
LME	14.986	0.581	0.049	0.873	0.876	0.418	0.873	0.794
Z&S	14.407	0.620	0.045	0.897	0.903	0.380	0.897	0.815
<i>Threshold</i>								
<i>u = 45.532</i>								
NEW	22.083	0.612	0.089	0.632	0.635	0.402	0.632	0.708
PWM	26.902	0.364	0.128	0.461	0.447	0.578	0.461	0.573
LME	25.930	0.446	0.120	0.492	0.494	0.536	0.492	0.602
Z&S	23.427	0.550	0.099	0.585	0.589	0.441	0.585	0.682

are significantly below the nominal level. But for all cases the NEW method has CPs close to the nominal level within  $\pm 0.03$  differences. Although the average interval lengths of CIs from the NEW method are large for extremely heavy-tailed situations (for  $\xi = 0.75$  the average length of the CI is 26.35, while the true value of  $Q(0.9; 1, 0.75) \approx 6.165$ ), we believe it is still reasonable since most other methods which have shorter intervals also have very low CPs. Generally speaking, except the extremely heavy-tailed situation where  $n = 15$  (which is not reported here), the confidence interval estimation based on the NEW method generally outperforms other tested methods for all tested scenarios and, most importantly, it is consistently reliable.

#### 4. EXAMPLE

We also examined a real-world example, using the flood insurance claim data from the U.S. National Flood Insurance Program (NFIP<sup>1</sup>). More details of this dataset are given in [5]. Average claim values per year (in thousands of dollars) of each county of the Florida state from 1998 to 2008 were gathered, with a total sample size of 504. We fit the GPD with this data by three different threshold values  $u = 26.302, 32.527, 45.532$  respectively. The three different threshold values result in three different sample sizes  $n = 50, 30, 15$  in the analysis. For each threshold, the Cramér-von Mises  $W^2$  statistic and the Anderson-Darling  $A^2$  statistic as well as Q-Q plot were used in the goodness-of-fit test.

The asymptotic  $p$  value and the bootstrap  $p$  value generated from 2000 samples for each statistic were calculated.

<sup>1</sup>The data are available on the webpage: <http://www.rff.org/Events/Pages/Data-Climate-Change-Extreme-Events.aspx>.

Results including estimated parameters, test statistics and their  $p$  values were reported in Table 8. For three sample sets, most methods gave shape parameter estimators  $\hat{\xi}$  close to or larger than 0.5, which implies that the underlying data follow a very heavy-tailed distribution which very likely has infinite second moments. The MOM method based estimates are known for having poor performance or even not existing, so MOM method is not included in this analysis. For the two lower thresholds, the NEW, LME and Z&S' estimation results do not show large differences, while the PWM method has notably smaller shape estimator and larger scale estimator. Both asymptotic and bootstrap  $p$  values are almost identical under both test statistics and generally lead to the same conclusion. In all tested cases, no method shows significant inadequate fitting. Judging by the  $p$  values, PWM estimates have the lowest goodness-of-fit level, while the NEW has the highest  $p$  values for all cases.

The Q-Q plots of the four tested methods with their threshold values are presented in Figure 5. Graphically, similar conclusions to the goodness-of-fit tests can be drawn from the Q-Q plots. Obviously the NEW and Z&S methods have better fitting than the other two methods, especially for the data on the upper tails.

#### 5. CONCLUSIONS

In this paper, we introduce a new method for point estimation and interval estimation of the GPD, specifically focusing on small sample cases. Numerical evidence has shown that the new method provides consistent point estimates against the varying tail behaviour of the GPD. At the same time it provides consistent, even accurate, confidence inter-

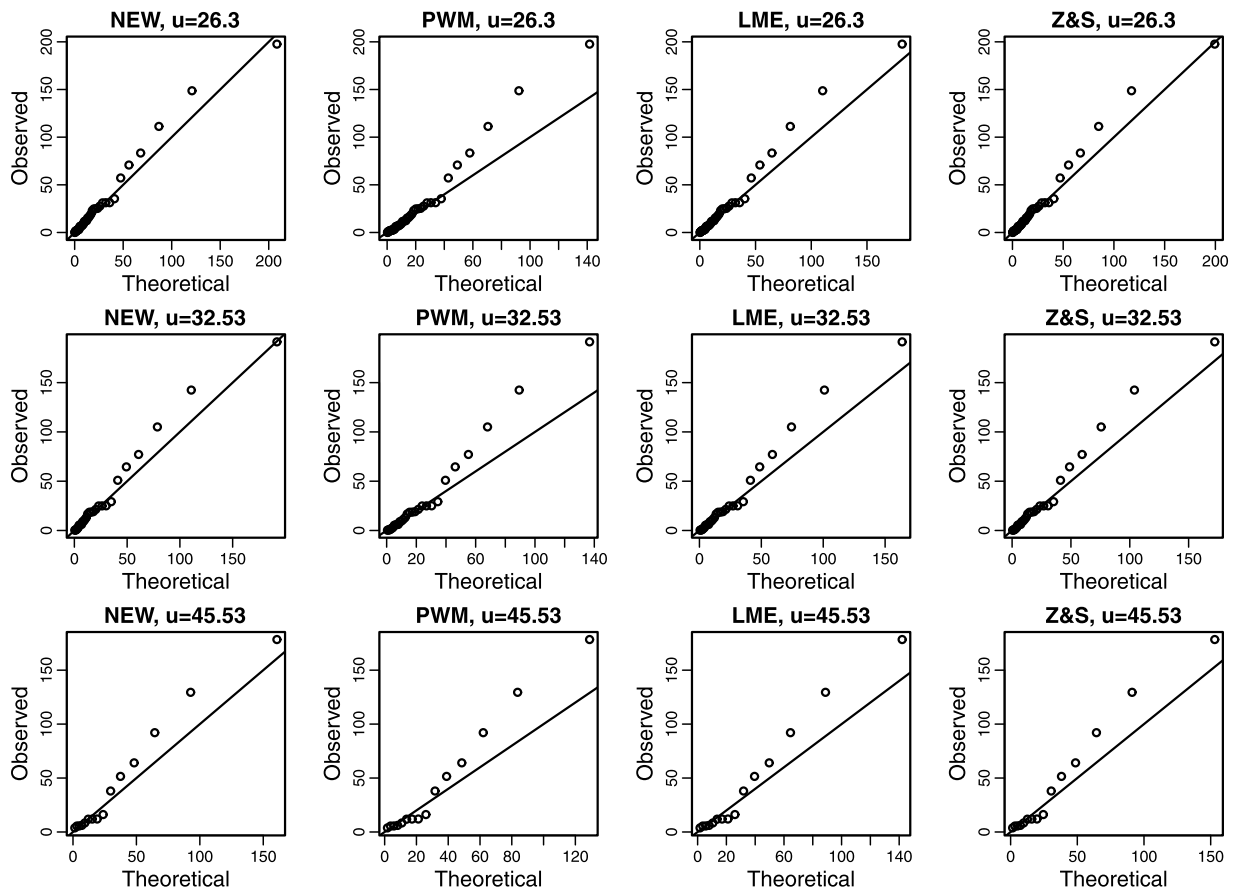


Figure 5. Q-Q plot for assessing the fitting of Florida flood claims data for GPD with four-method.

val estimates in terms of coverage probability and average interval length, for sample sizes as small as  $n = 30$ .

At the same time, our numerical study shows that all methods to fit GPD are sensitive to the value of shape parameter  $\xi$ . The new method, despite its reliability or versatility, may not always be best under some of values of  $\xi$ . In reality, however, it is not always possible to choose the ‘appropriate’ model according to  $\xi$ , as one cannot always know for sure what range the value of  $\xi$  falls into or what type of tail the underlying distribution may have. Thus, an estimation method like the proposed one in this paper that maintains reliable or versatile (not necessarily always the best) performance without any prior information about  $\xi$  can be a very useful in fitting GPD.

## ACKNOWLEDGEMENTS

We thank the reviewer, two Guest Editors/AEs and the Editor-in-Chief for their handling the manuscript review process. Their useful comments and suggestions significantly improved the quality of the paper. The research is supported in part by the National Sciences Foundation of China (NSFC) 11261048 and 11371322.

Received 27 May 2013

## REFERENCES

- [1] ATHREYA, K. (1987). Bootstrap of the mean in the infinite variance case. *The Annals of Statistics* **15** 724–731. [MR0888436](#)
- [2] BALKEMA, A. A. and DE HAAN, L. (1974). Residual life time at great age. *The Annals of Probability* **2** 792–804. [MR0359049](#)
- [3] BATES, G. E. (1955). Joint distributions of time intervals for the occurrence of successive accidents in a generalized Polya scheme. *The Annals of Mathematical Statistics* **26** 705–720. [MR0076265](#)
- [4] CASTILLO, E. and HADI, A. S. (1997). Fitting the Generalized Pareto Distribution to Data. *Journal of the American Statistical Association* **92**. [MR1615270](#)
- [5] COOKE, R. M. and NIEBOER, D. (2011). Heavy-Tailed Distributions: Data, Diagnostics, and New Developments. *Resources for the Future Discussion Paper* **11-19**.
- [6] DAVISON, A. C. (1984). Modelling excesses over high thresholds, with an application. *Statistical extremes and applications* **131** 461–482.
- [7] DAVISON, A. C. and SMITH, R. L. (1990). Models for exceedances over high thresholds. *Journal of the Royal Statistical Society. Series B (Methodological)* 393–442. [MR1086795](#)
- [8] DIEBOLD, F. X., SCHUERMANN, T. and STROUGHAIR, J. D. (2000). Pitfalls and Opportunities in the Use of Extreme Value Theory in Risk Management. *Journal of Risk Finance* **1** 30–35.
- [9] DUPUIS, D. J. and TSAO, M. (1998). A hybrid estimator for generalized Pareto and extreme-value distributions. *Communications in Statistics-Theory and Methods* **27** 925–941. [MR1613505](#)
- [10] EFRON, B. and TIBSHIRANI, R. (1986). Bootstrap Methods for Standard Errors, Confidence Intervals, and Other Measures of Statistical Accuracy. *Statistical Science* **1**. [MR0833275](#)

- [11] GILLI, M. and KELLEZI, E. (2006). An application of extreme value theory for measuring financial risk. *Computational Economics* **27** 207–228.
- [12] GRIMSHAW, S. D. (1993). Computing maximum likelihood estimates for the generalized Pareto distribution. *Technometrics* 185–191. [MR1225094](#)
- [13] HALL, P. (1990). Asymptotic properties of the bootstrap for heavy-tailed distributions. *The Annals of Probability* **18** 1342–1360. [MR1062071](#)
- [14] HOSKING, J. R. M. and WALLIS, J. R. (1987). Parameter and Quantile Estimation for the Generalized Pareto Distribution. *Technometrics* **29** 1609–1620. [MR0906643](#)
- [15] HÜSLER, J., LI, D. and RASCHKE, M. (2011). Estimation for the Generalized Pareto Distribution Using Maximum Likelihood and Goodness of Fit. *Communications in Statistics—Theory and Methods* **40** 2500–2510. [MR2863325](#)
- [16] NEWCOMBE, R. G. (1998). Two-sided confidence intervals for the single proportion: comparison of seven methods. *Statistics in medicine* **17** 857–872.
- [17] PICKANDS III, J. (1975). Statistical inference using extreme order statistics. *the Annals of Statistics* 119–131. [MR0423667](#)
- [18] RAO, C. R. (1973). *Linear statistical inference and its applications* **22**. Wiley-Interscience. [MR0346957](#)
- [19] RIBATEF, M. (2011). A User's Guide to the POT Package (Version 1.4) [Accessed 9 December 2011].
- [20] SMITH, R. L. (1985). Maximum likelihood estimation in a class of nonregular cases. *Biometrika* **72** 67–90. [MR0790201](#)
- [21] SMITH, R. L. (1987). Estimating tails of probability distributions. *The Annals of Statistics* **15** 1174–1207. [MR0902252](#)
- [22] STEPHENS, M. (1986). Tests for the exponential distribution. In *Goodness-of-fit Techniques* (R. B. D'Agostino and M. A. Stephens, eds.) 421–459. New York: Marcel Dekker.
- [23] TAJVIDI, N. (2003). Confidence intervals and accuracy estimation for heavy-tailed generalized Pareto distributions. *Extremes* **6** 111–123. [MR2076639](#)
- [24] VENZON, D. and MOOLGAVKAR, S. (1988). A method for computing profile-likelihood-based confidence intervals. *Applied Statistics* 87–94.
- [25] VIVEROS, R. and BALAKRISHNAN, N. (1994). Interval estimation of parameters of life from progressively censored data. *Technometrics* 84–91. [MR1256709](#)
- [26] WEERAHANDI, S. (1993). Generalized confidence intervals. *Journal of the American Statistical Association* 899–905. [MR1242940](#)
- [27] WEERAHANDI, S. (2004). *Generalized inference in repeated measures: Exact methods in MANOVA and mixed models* **500**. Wiley-Interscience. [MR2078699](#)
- [28] ZHANG, J. (2007). Likelihood moment estimation for the generalized Pareto distribution. *Australian & New Zealand Journal of Statistics* **49** 69–77. [MR2345411](#)
- [29] ZHANG, J. and STEPHENS, M. A. (2009). A new and efficient estimation method for the generalized Pareto distribution. *Technometrics* **51** 316–325. [MR2751074](#)

Jian He  
Department of Statistics and Finance  
Business School  
Shihezi University  
China

Zhuo Sheng  
Department of Mathematical Sciences  
Brunel University  
London  
UK

Bing Xing Wang  
Department of Mathematics  
Zhejiang Gongshang University  
China

Keming Yu  
Department of Mathematical Sciences  
Brunel University  
London  
UK  
Tel.: +44 (0)1895 266128  
Fax: +44 (0)1895 269732  
E-mail address: [keming.yu@brunel.ac.uk](mailto:keming.yu@brunel.ac.uk)