Computational Statistics and Data Analysis 98 (2016) 91–104

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

journal homepage: www.elsevier.com/locate/csda

Estimating extreme tail risk measures with generalized Pareto distribution

Myung Hyun Park, Joseph H.T. Kim *

Department of Applied Statistics, College of Business and Economics, Yonsei University, Seoul, 120-749, Republic of Korea

HIGHLIGHTS

- A new GPD parameter estimator is proposed.
- It is based on a nonlinear weighted least squares method.
- Under the POT framework, we estimate tail risk measures. Extensive simulation studies show the new method works well.

ARTICLE INFO

Article history:
Received 18 May 2015
Received in revised form 16 November 2015
Accepted 15 December 2015
Available online 4 January 2016

Keywords:
Generalized Pareto distribution
Value-at-Risk (VaR)
Conditional Tail Expectation (CTE)
Tail VaR
Peaks over threshold (POT)
Weighted nonlinear least squares

ABSTRACT

The generalized Pareto distribution (GPD) has been widely used in modelling heavy tail phenomena in many applications. The standard practice is to fit the tail region of the dataset to the GPD separately, a framework known as the peaks-over-threshold (POT) in the extreme value literature. In this paper we propose a new GPD parameter estimator, under the POT framework, to estimate common tail risk measures, the Value-at-Risk (VaR) and Conditional Tail Expectation (also known as Tail-VaR) for heavy-tailed losses. The proposed estimator is based on a nonlinear weighted least squares method that minimizes the sum of squared deviations between the empirical distribution function and the theoretical GPD for the data exceeding the tail threshold. The proposed method properly addresses a caveat of a similar estimator previously advocated, and further improves the performance by introducing appropriate weights in the optimization procedure. Using various simulation studies and a realistic heavy-tailed model, we compare alternative estimators and show that the new estimator is highly competitive, especially when the tail risk measures are concerned with extreme confidence levels.

1. Introduction

The last few decades have witnessed an unprecedented increase in the size of datasets available, a phenomenon of massive data, in various applications such as finance, insurance, computer science and communications. Such large datasets now allow various quantitative risk analyses which were once thought infeasible to implement, including the investigation on rare but huge loss events occurring in the tail of the distribution. For instance, for financial institutions, extreme quantiles of the loss distribution are of great interest for both internal and regulatory purposes. Accurate estimation of such tail-related quantities however generally requires generating considerably large loss samples (see Section 5 for details), which may be very time-consuming for large complicated financial portfolios, and calculating those quantities is also known to be

* Corresponding author.
E-mail address: jhtkim@yonsei.ac.kr (J.H.T. Kim).

http://dx.doi.org/10.1016/j.csda.2015.12.008
0167-9473/

expensive because of the computing time and memory storage, as observed in the computer science literature, e.g., Liechty et al. (2003), Chen et al. (2000) and Munro and Paterson (1980); see Song and Song (2012) for a more detailed account on the issue of estimating extreme quantiles from massive datasets.

In view of these difficulties, Extreme value theory (EVT) has received much attention as a modern tool to study tail quantities of the distribution, including the extreme quantiles. In the centre of EVT framework, the generalized Pareto distribution (GPD) emerged as the distribution of the exceedances above a sufficiently high threshold for arbitrary heavy-tailed loss data (Pickands, 1975). The standard procedure of fitting heavy-tailed data calls for a separate modelling of the tail region of the dataset using the GPD, a procedure commonly known as the peaks over threshold (POT); see, e.g., Embrechts et al. (1997). In finance and insurance applications, a main purpose of applying EVT under the POT is to determine common tail risk measures such as the Value-at-Risk (VaR) or Conditional Tail Expectation[1] (a.k.a. Tail-VaR) from the fitted GPD. However, residing in the tail region, these risk measure estimates are highly sensitive to the estimated GPD parameters, and their volatility becomes larger as the required quantile level gets extreme. For example, according to Basel II (BCBS, 2006), the operational risk capital a bank should hold must cover unexpected losses with at least a 99.9% probability under the Advanced Measurement Approach (AMA), equivalent to VaR at a confidence level of 99.9%. Similarly, the credit risk capital under the Internal Ratings Based (IRB) approach also requires VaR 99.9%; furthermore, VaR 99.99% is used in the regulator's backtesting analysis, assessing rare events occurring with probability 0.0001. Apparently, such extreme quantiles are highly sensitive to the estimation methods and small differences in the estimates could lead to a considerable impact on the financial position of banks, underscoring the importance of accurate GPD estimation.

Estimating the GPD parameters is a long-standing problem and various approaches have been investigated in the literature. For example, the traditional maximum likelihood estimation (MLE) is discussed in Grimshaw (1993), Davison (1984) and Smith (1985). Pickands (1975) proposed a method using the sample order statistics; Hosking and Wallis (1987) used the method of moment (MME) and probability weighted moments (PWM). A generalized version of the MME was discussed by Ashkar and Ouarda (1996) and generalized probability weighted moments (GPWM) was proposed by Rasmussen (2001); Dupuis and Tsao (1998) developed a hybrid PWM. Juárez and Schucany (2004) introduced a minimum density power divergence method, and Zhang (2007) proposed a likelihood moment estimator (LME). Zhang and Stephens (2009) and Zhang (2010) surveyed some relevant contributions to the literature of the GPD parameter estimation by means of the Bayesian methodology. The reader is also referred to de Zea Bermudez and Kotz (2010) for a survey of various GPD estimators. More recently, Song and Song (2012) introduced a new – yet computationally simple and fast – GPD parameter estimator for large samples based on a nonlinear least square method that minimizes the sum of squared deviations between the empirical distribution function (EDF) of the sample and the theoretical GPD, which was reported to outperform other existing methods they considered.

In this paper, we propose a new GPD parameter estimator under the POT framework to estimate tail risk measures at extreme quantiles. Our method is adapted from that of Song and Song (2012) and uses a nonlinear least squares method that minimizes the sum of squared deviations between the empirical distribution function and the theoretical GPD for the data exceeding the tail threshold. However, the proposed method uses a different object function and is better in its performance, and these are our contributions in the current paper. In particular, we first examine the method of Song and Song (2012) and point out its caveat, to show that their method is only applicable for the case where the tail of the loss sample is GPD distributed unconditionally. We address this issue and present a revised procedure. Second, in order to further improve the estimation, we introduce suitable weights in the revised optimization procedure using the weighted regression setup. Using the proposed procedure we estimate the VaR and CTE, the two popular tail risk measures, at extreme quantile levels. Using various simulation studies and a realistic heavy-tailed model, we compare alternative estimators and show that the performance of the proposed estimator is highly competitive compared to other existing estimators, especially for the risk measures with extreme confidence levels.

This paper is organized as follows. The POT approach in EVT is briefly reviewed in Section 2. In Section 3, after reviewing existing methods, we describe a new GPD parameter estimator under the POT. Section 4 is devoted to numerical exercises, where we compare the performances of different methods in estimating tail risk measures under various heavy-tailed common parametric distributions. In Section 5 a more realistic loss model is considered for a similar exercise. Section 6 concludes the paper.

2. EVT for extreme tail risk measures

2.1. Peaks over threshold

We start with a brief account for Extreme Value Theory (EVT) focused on the peaks over Threshold (POT) framework; comprehensive treatments of book length on EVT can be found in, e.g., Embrechts et al. (1997) and Beirlant et al. (2006). Let us denote the tail or survival function of a continuous random variable X by $\bar{F}(x) = 1 - F(x), 0 < x < \infty$. Then we say that

[1] In the literature, this is also known as the Conditional Value-at-Risk (CVaR) or Expected Shortfall (ES). The CTE will be formally defined later; see, e.g., McNeil et al. (2005) for a comprehensive discussion on risk measures.

$\bar{F}(x)$ is regularly varying with index $-\gamma < 0$, or simply $\bar{F} \in \mathcal{R}_{-\gamma}$, if

$$\lim_{x\to\infty} \frac{\bar{F}(x\lambda)}{\bar{F}(x)} = \lambda^{-\gamma}, \quad \lambda > 0. \tag{1}$$

When $\gamma = 0$ the tail is called slowly varying, or $\bar{F} \in \mathcal{R}_0$. Using this definition we can write a regularly varying distribution as $\bar{F}(x) \sim L(x)\, x^{-\gamma}$ where $L(\cdot) \in \mathcal{R}_0$. We note that $f(x) \sim g(x)$ means $\lim_{x\to\infty} f(x)/g(x) = 1$. Therefore the tail of regularly varying functions can be represented by power functions multiplied by slowly varying functions. Heavy-tailed distributions such as Pareto, Generalized Pareto, Log-gamma, Cauchy and Stable laws are examples of such functions. When $\bar{F} \in \mathcal{R}_{-\gamma}$ with $\gamma > 0$, the distribution is said to be in the maximum domain of attraction of the Fréchet distribution (MDA(Φ_γ) in short), a member of the generalized extreme value (GEV) distribution family, representing the class of distributions with heavy tails.

A fundamental result of EVT is the famous Pickands–Balkema–de Haan theorem (Balkema and De Haan, 1974 and Pickands, 1975) which states that, for $\bar{F} \in \mathcal{R}_{-\gamma}$, the excess loss $(X - u|X > u)$ from such a distribution with a large threshold $u > 0$ converges to the Generalized Pareto distribution (GPD) with positive Pareto parameter $\xi > 0$. That is,

$$\lim_{u\uparrow x_F} \sup_{0<x<x_F-u} |Pr(X - u < x|X > u) - G_{\xi,\sigma}(x)| = 0, \tag{2}$$

where $x_F \leq \infty$ is the right endpoint of F and $G_{\xi,\sigma}(x)$ is the GPD distribution function. This implies that the excess distribution $\Pr(X - u|X > u)$ converges to the GPD whenever X is heavy-tailed distribution in MDA(Φ_γ). This theorem, therefore, provides the basis of the POT framework which enables a separate modelling for the tail part of the dataset with the GPD, starting from u. Selecting the threshold u for a given sample however has been considered to be a nontrivial problem and one should explore different candidates in practice. The distribution function of the GPD is defined as

$$G_{\xi,\sigma}(x) = 1 - \left(1 + \frac{\xi}{\sigma}x\right)^{-1/\xi}, \quad x > 0,\ \xi > 0, \tag{3}$$

where $\sigma > 0$ is the scale parameter, ξ is the shape parameter and $1/\xi = \gamma$ is the tail index. Sometimes the location parameter $\mu \in (-\infty, \infty)$ is added, in which case the distribution function is denoted by $G_{\xi,\mu,\sigma}(x) = G_{\xi,\sigma}(x - \mu)$ with the support $x > \mu$. The value of ξ determines the shape of the GPD. Distributions with $\xi > 0$ are classified heavy-tailed with $E[X^k]$ being infinite for $k > 1/\xi$. Among other properties, we note that, if X is $G_{\xi,\sigma}(y)$ distributed, its excess loss is again GPD distributed with parameter $(\xi, \sigma + \xi u)$. That is,

$$Pr(X - u < x|X > u) = 1 - \left(1 + \frac{\xi}{\sigma + \xi u}x\right)^{-1/\xi}, \quad x > 0,\ \xi > 0. \tag{4}$$

In a sense, this so-called stability of the GPD resembles the memoryless property of the exponential distribution, a special case of the GPD with $\xi = 0$.

2.2. Tail risk measures

The VaR of a continuous random variable X at the $100p\%$ level is the $100p$ quantile of the distribution of X, denoted by

$$Q_p(X) = F_X{}^{-1}(p). \tag{5}$$

The VaR is a widely-accepted standard risk measure used in solvency and risk analyses in financial industry, with p close to 1. However, it has also been criticized for lacking a certain property that desirable risk measures should meet, the criteria known as the coherent risk measure axioms; see Artzner et al. (1999). In this regard, the CTE has received much attention as an alternative, coherent tail risk measure. The idea of the CTE is to measure the average severity of the loss when the extreme loss does occurs, where the extreme loss is represented by the VaR. Formally, the CTE of the random variable X at $100p\%$ level is defined as

$$CTE_p(X) = E[X|X > Q_p(X)]. \tag{6}$$

When the POT approach is employed, one can write the underlying loss distribution as a combination of the body and tail parts with the latter modelled by the GPD. For this, we let $F(x)$ be the distribution function of an arbitrary continuous distribution, and define the exceedance distribution of loss events over u by

$$F_u(y) = \Pr(X - u \leq y|X > u) = \frac{F(u + y) - F(u)}{1 - F(u)}, \tag{7}$$

which is assumed to be GPD(ξ, σ). Then we can rewrite $F(x)$ as

$$\begin{aligned} F(x) = P(X \leq x) &= (1 - F(u))F_u(x - u) + F(u) \\ &= (1 - F(u))G_{\xi,\sigma}(x - u) + F(u) \\ &= (1 - F(u))G_{\xi,u,\sigma}(x) + F(u), \end{aligned} \tag{8}$$

using the three-parameter GPD, and its estimated version by

$$\hat{F}(x) = (1 - F_n(u))G_{\hat{\xi},u,\hat{\sigma}}(x) + F_n(u), \tag{9}$$

where F_n is the EDF and $\hat{\xi}$, $\hat{\sigma}$ are the estimated GPD parameters from the sample. Assuming that the tail risk measures lie in the GPD realm, the VaR estimate is obtained from inverting (9) as

$$\widehat{Q}_p(X) = G^{-1}_{\hat{\xi},u,\hat{\sigma}}\left[1 - \frac{1-p}{1-F_n(u)}\right] = u + \frac{\hat{\sigma}}{\hat{\xi}}\left[\left(\frac{n}{n_u}(1-p)\right)^{-\hat{\xi}} - 1\right], \tag{10}$$

where n_u is the number of exceedances over the threshold u and n is the total number of observations. The CTE is determined from the fact that when $X - u|X > u$ is GPD(ξ, σ) distributed, $X - u'|X > u'$, with $u' > u$, is GPD$(\xi, \sigma + \xi(u' - u))$ distributed. Hence, subject to $\xi < 1$ for the mean to be finite, we obtain, for $Q_p(X) > u$,

$$\begin{aligned} CTE_p(X) &= E[X - Q_p(X)|X > Q_p(X)] + Q_p(X) \\ &= \frac{\sigma + \xi(Q_p(X) - u)}{1-\xi} + Q_p(X) = \frac{Q_p(X) + \sigma - \xi u}{1 - \xi}, \quad \xi < 1, \end{aligned} \tag{11}$$

which leads to its estimate

$$\widehat{CTE}_p(X) = \frac{\widehat{Q}_p(X) + \hat{\sigma} - \hat{\xi}u}{1-\hat{\xi}} = u + \frac{\hat{\sigma}}{\hat{\xi}}\left[\frac{1}{1-\hat{\xi}}\left(\frac{n}{n_u}(1-p)\right)^{-\hat{\xi}} - 1\right]. \tag{12}$$

3. GPD estimation

3.1. Existing methods

The traditional maximum likelihood estimation (MLE) was studied in Grimshaw (1993), Davison (1984) and Smith (1985). The likelihood function however approaches infinity when $\xi < -1$, so the MLE exists only for $\xi > -1$. Further, Smith (1984) showed that the resulting estimators for the GPD are consistent and asymptotically normal only for $\xi > -0.5$. The computational difficulty and convergence issue of the MLE, even when $\xi > -0.5$, was discussed in Grimshaw (1993) and Hosking and Wallis (1987). In order to improve the estimation, Hosking and Wallis (1987) proposed the Method of Moments Estimation (MME) using the first two moments, but its performance is poor for $\xi \geq 1/2$ where the second moment does not exist. In the same paper, they also derived a probability weighted moments (PWM). He and Fung (1993) proposed the method of median (MED), which is found by equating the sample median of the two partial derivatives to the corresponding population medians. The likelihood moment method (LME) was proposed by Zhang (2007) to solve the numerical problems of the MLE. Unlike the MLE, the LME always exists and is computationally simple. In a separate paper, Zhang (2010) also suggested a Bayesian estimator, improved over its previous one in Zhang and Stephens (2009), which is more efficient and adaptive for $\xi > 1$ by selecting a better prior distribution. Also see de Zea Bermudez and Kotz (2010) for a survey of various estimators.

Recently, Song and Song (2012) proposed a new GPD parameter estimator for massive heavy-tailed data based on the POT framework. Their procedure, called the nonlinear least squares (NLS) estimator, consists of two steps with the first step designed to stabilize the shape parameter ξ estimation with an interim estimate. We review the NLS procedure in detail here as it is closely related to our proposed one. Suppose that we have a sample $x_1, \ldots, x_n$ of size n, and $n_u < n$ observations that are greater than the selected GPD threshold u. Without loss of generality, it is assumed $x_1 > x_2 > \cdots > x_n$. In the first step, one finds the interim estimate $(\hat{\xi}_1, \hat{\sigma}_1)$ using a nonlinear minimization:

$$(\hat{\xi}_1, \hat{\sigma}_1) = \arg\min_{(\xi,\sigma)} \sum_{i=1}^{n_u}\Big[\log(1 - F_n(x_i)) - \log\big(1 - G_{\xi,u,\sigma}(x_i)\big)\Big]^2. \tag{13}$$

Here $F_n(x)$ is the EDF and $G_{\xi,u,\sigma}(x)$ is the distribution function of the three-parameter GPD. With $(\hat{\xi}_1, \hat{\sigma}_1)$ as the initial values, the following second step runs another optimization:

$$(\hat{\xi}_2, \hat{\sigma}_2) = \arg\min_{(\xi,\sigma)} \sum_{i=1}^{n_u}\left[F_n(x_i) - G_{\xi,u,\sigma}(x_i)\right]^2, \tag{14}$$

which is the least square estimation for the distribution function of the GPD. The resulting $(\hat{\xi}_2, \hat{\sigma}_2)$ is the final estimate of the GPD parameters. Song and Song (2012) commented that the direct application of (14) without the first step in (13) gives quite volatile estimates because, for high thresholds, the GPD distribution function values are concentrated on the interval very close to 1, making the global minimum hard to find in a stable manner. For various datasets simulated from standard heavy-tailed distributions, the two-step NLS procedure was reported to outperform other existing methods.

Table 1
RMSE and ARB (in parenthesis) of quantiles for GPD(ξ, σ); $n = 10{,}000$ with 1000 repetitions.

Quantile	0.95	0.99	0.999	0.9999
GPD($\xi = 1, \sigma = 10$)				
Sample quantiles	8.626(0.036)	100.865(0.080)	3216.99(0.246)	147 288.00(0.708)
MED	15.609(0.065)	150.776(0.120)	2686.93(0.207)	40 504.12(0.301)
Pickands	14.225(0.059)	141.117(0.111)	2522.91(0.193)	37 829.47(0.279)
LME	6.894(0.029)	64.148(0.051)	1108.47(0.087)	16 003.95(0.124)
Zhang	6.870(0.029)	63.662(0.051)	1096.26(0.087)	15 782.21(0.124)
NLS	9.049(0.038)	85.701(0.069)	1481.36(0.118)	21 396.81(0.169)
GPD($\xi = 2, \sigma = 5$)				
Sample quantiles	86.866(0.069)	5246.49(0.161)	2 002 801(0.520)	5 276 005 000(3.236)
MED	143.880(0.112)	6565.58(0.199)	1 172 964(0.338)	184 571 700(0.500)
Pickands	117.420(0.093)	5330.13(0.166)	921 705(0.275)	139 676 300(0.394)
LME	68.375(0.055)	2958.13(0.094)	488 195(0.152)	69 857 580(0.212)
Zhang	66.963(0.054)	2850.03(0.091)	463 349(0.147)	65 390 210(0.204)
NLS	81.429(0.066)	3493.13(0.111)	571 915(0.180)	81 499 940(0.251)

3.2. New method

We start this section by discussing a caveat of the NLS estimator of Song and Song (2012) as described in (13) and (14). In particular, while their idea of minimizing the squared deviations between $F_n(x)$ and the theoretical GPD over $x > u$ is valid in light of the POT framework, their actual formation is only applicable when the underlying model's tail fits GPD($\xi, 0, \sigma$), unconditionally. To elaborate this point, consider the EDF values in (13) and (14) for $x > u$, which is the range over which the optimization is carried out. In this range, $F_n(x)$ increases with the starting value of $F_n(u)$, which is already very close to 1 because u is the GPD threshold in the tail; for instance, the starting value is $F_n(u) = 0.98$ if u is set to be the 98th sample quantile. In contrast, the value of $G_{\xi,u,\sigma}(x)$ in (13) and (14) always starts from $G_{\xi,u,\sigma}(u) = 0$ as its support is always $[u, \infty)$. Hence the difference between $F_n(x)$ and $G_{\xi,u,\sigma}(x)$ is not measured consistently, and we see that the NLS procedure is only sensible and applicable when $u = 0$, the case where the distribution is unconditionally GPD from the very beginning of the data. This discrepancy can also be rephrased using the excess loss notation. Looking at the object function (14), $G_{\xi,u,\sigma}(x)$ represents the conditional distribution of $X|X > u$ whereas the EDF $F_n(x)$ is a distribution for the entire X. While the summation in (14) is carried out for the observations exceeding the threshold only, the issue remains unsolved in this second step because the EDF conditional on $X > u$ cannot be created by simply restricting the acceptable range of the observations; the conditional EDF should be constructed on its own basis using truncation, as will be shown shortly. In this regard, the limitation of NLS procedure is that it does not properly employ the POT framework despite their claim of doing so. In their numerical exercises, Song and Song (2012) report that the performance of the NLS is superior to alternative estimators. However the comparison was made unfairly, because all the other methods are based on the POT framework whereas the NLS is not. As $\hat{\sigma}$ needs to be obtained from conversion formula $\hat{\sigma} = \hat{\sigma}(u)\,(1-F_n(u))^{\hat{\xi}}$ under the POT framework, where $\sigma(u)$ is the scale parameter estimate from the truncated distribution (McNeil and Saladin, 1997), forgoing the POT element gives an advantage to the NLS method by unduly reducing parameter uncertainties.

One way to compare the alternative estimators from a proper perspective is to compare their performances assuming $u = 0$, which provides a level playing field for all estimators. For this, we carry out a simulation study where several estimators considered in Song and Song (2012) are compared for extreme VaRs when the underlying model is GPD($\xi, 0, \sigma$) with different (ξ, σ) values, of which the results are presented in Table 1. The performance for each estimator is evaluated using the root MSE (RMSE) and the absolute relative bias (ARB), where the latter is defined as the scaled difference between the estimator and true quantile, $E(|\hat{\theta} - \theta|/\theta)$ for an estimator $\hat{\theta}$. From the table, it is seen that, contrary to what (Song and Song, 2012) reported, the estimator of Zhang (2010) is superior to the NLS for VaR estimations, indicating that the performance of the NLS is not as good.

We now revise the NLS method so that it can properly account for the POT framework. The key element in this revision is to minimize the squared distance between $F_n(x)$ and $F(x)$ in (8), not the distance between $F_n(x)$ and $G_{\xi,u,\sigma}(x)$ as done in (13) and (14). The revised first step of the NLS then should be

$$
\begin{aligned}
(\hat{\xi}_1, \hat{\sigma}_1) &= \arg\min_{(\xi,\sigma)} \sum_{i=1}^{n_u} \Big[\log(1-F_n(x_i)) - \log(1-F(x_i))\Big]^2 \\
&= \arg\min_{(\xi,\sigma)} \sum_{i=1}^{n_u} \Big[\log(1-F_n(x_i)) - \log\big(1-(1-F_n(u))G_{\xi,u,\sigma}(x_i) - F_n(u)\big)\Big]^2 \\
&= \arg\min_{(\xi,\sigma)} \sum_{i=1}^{n_u} \left[\log\frac{1-F_n(x_i)}{1-F_n(u)} - \log(1-G_{\xi,u,\sigma}(x_i))\right]^2, \qquad (15)
\end{aligned}
$$

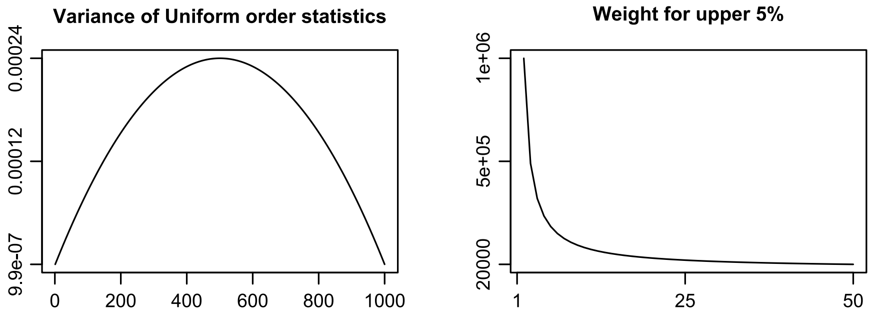

Fig. 1. Variance of $U_{n-i+1:n}$ (left) and the weights (right) for $n = 1000$.

followed by the revised second step:

$$
\begin{aligned}
(\hat{\xi}_2, \hat{\sigma}_2) &= \arg\min_{(\xi,\sigma)} \sum_{i=1}^{n_u} [F_n(x_i) - F(x_i)]^2 \\
&= \arg\min_{(\xi,\sigma)} \sum_{i=1}^{n_u} \left[F_n(x_i) - (1 - F_n(u))G_{\xi,u,\sigma}(x_i) - F_n(u)\right]^2 \\
&= \arg\min_{(\xi,\sigma)} \sum_{i=1}^{n_u} (1 - F_n(u))^2 \left[\frac{F_n(x_i) - F_n(u)}{1 - F_n(u)} - G_{\xi,u,\sigma}(x_i)\right]^2 \\
&= \arg\min_{(\xi,\sigma)} \sum_{i=1}^{n_u} \left[\frac{F_n(x_i) - F_n(u)}{1 - F_n(u)} - G_{\xi,u,\sigma}(x_i)\right]^2 .
\end{aligned} \tag{16}
$$

We call this revised two-step procedure "pot-NLS", as opposed to the original NLS. Compared to the NLS method, the pot-NLS addresses the tail part appropriately by using a truncated distribution, well-aligned with the POT framework.

Moreover, we observe that the second step optimization (16) can be further improved by adding suitably chosen weights for each squared deviance term, an idea from the weighted least squares regression. To find the suitable weight for each squared error, we recognize that the first line of the revised second step above is equivalent to applying the least squares method to the following regression setup:

$$F(x_i) = F_n(x_i) + \epsilon_i, \quad i = 1, \ldots, n, \tag{17}$$

where ϵ_i is the error term with $E(\epsilon_i) = 0$, which is justified from the fact that the EDF is an unbiased estimate of $F(x)$ for all x; see, e.g., Rohatgi (1984). Now noting that $F(X)$ is a uniform distributed random variable for any X, we can determine the weight for each response variable $F(X_i)$ as the reciprocal of its variance. That is, assuming without loss of generality that $X_1 > \cdots > X_n$, the distribution of $F(X_i)$ is that of $U_{n-i+1:n}$, the $(n-i+1)$th order statistic of the uniform random variable, of which the distribution is known to be $Beta(n-i+1, i)$ from the standard distribution theory. Because the first two moments of such beta variable are given by

$$E(U_{n-i+1:n}) = \frac{n-i+1}{n+1}, \qquad \mathrm{Var}(U_{n-i+1:n}) = \frac{i(n-i+1)}{(n+1)^2(n+2)}, \tag{18}$$

the weight for $F(X_i)$ should be $(\mathrm{Var}(U_{n-i+1:n}))^{-1}$, which yields the weighted version of the nonlinear optimization given in (16):

$$
\begin{aligned}
(\hat{\xi}_3, \hat{\sigma}_3) &= \arg\min_{(\xi,\sigma)} \sum_{i=1}^{n_u} [\mathrm{Var}(U_{n-i+1:n})]^{-1} [F_n(x_i) - F(x_i)]^2 \\
&= \arg\min_{(\xi,\sigma)} \sum_{i=1}^{n_u} \left[\frac{i(n-i+1)}{(n+1)^2(n+2)}\right]^{-1} \left[\frac{F_n(x_i) - F_n(u)}{1 - F_n(u)} - G_{\xi,u,\sigma}(x_i)\right]^2 .
\end{aligned} \tag{19}
$$

We will call $(\hat{\xi}_3, \hat{\sigma}_3)$, combined with the revised first step (15), the weighted NLS estimator under the POT (pot-WNLS in short). One advantage of the pot-WNLS estimator over the pot-NLS, as will be seen later, is that it estimates the extreme quantiles in a more stable manner because, as $\mathrm{Var}(U_{n-i+1:n})$ is increasing in i for $i \le n/2$, larger weights are given for $F(x_i)$ values as x_i moves towards the tail side. In Fig. 1, the values of $\mathrm{Var}(U_{n-i+1:n})$, $i = 1, \ldots, n$, and the weights in (19) are illustrated. We use the standard "optim" function in R, based on the default method of Nelder and Mead (1965), to implement the pot-WNLS estimator in our numerical exercises.

4. Simulation studies

We estimate the tail risk measures using different estimators from various parametric heavy-tailed distributions. Following the POT framework, we fit the GPD using the sample exceedances above a sufficiently high threshold to estimate the tail risk measures. For the risk measures, we compute the VaR and CTE at quantile levels ranging from 98% to 99.99%. We include 99.99%, despite its extremity, in light of its practical use as discussed in Introduction. As our interest is in risk measures rather than the parameter itself, we do not report the parameter results. The competing estimators we consider are the MED estimator by He and Fung (1993), the Pickands estimator by Pickands (1975), the LME by Zhang (2007), the Zhang's method of Zhang (2010), our NLS method (pot-NLS) and its weighted version (pot-WNLS). The MME, PWM and MLE are not considered; the first two estimators are known to perform poorly except for narrow ranges of ξ and may produce infeasible estimates (Zhang, 2007, 2010), and the MLE often cannot be obtained when $\xi > 0.5$ due to the singular observed information matrix as noted by Song and Song (2012), which is in agreement with Hosking and Wallis (1987).

4.1. VaR estimation

The simulated samples are generated from the GPD, Cauchy, Log-gamma and Pareto, respectively. For each sample we choose 97th, 98th and 99th sample quantiles for the thresholds u and the estimated quantiles are VaR 98%, 99%, 99.9%, and 99.99% of the distribution. We consider different threshold values because the exact GPD threshold is unknown in real applications. Practical experience suggests that the thresholds are relatively large with the number of exceedances less than 5% of the sample, justifying our candidate threshold choices; see Chapter 7 of McNeil et al. (2005). We describe the simulation procedure for each distribution as follows.

1. Generate a random sample of size 10,000 from the given distribution
2. Pick a threshold u, taken to be the 100pth sample quantile
3. Fit the GPD with the observations above u using alternative estimators to estimate ξ and σ, which in turn yield the VaR estimate using (10)
4. Repeat above steps 20,000 times to compute the RMSE and ARB. The true VaR for the distribution is obtained analytically or numerically.

Note that the sample size of 10,000 allows us to have 100 to 300 observations for the GPD fit, reasonable numbers of exceedances to investigate extreme tails. The simulation results are presented in Tables 2 and 3. We discuss the results for each distribution below, but overall we see that the "MED", "Pickands" are relatively poor estimators for all cases. The performance of the "pot-NLS" is about average. The "LME", "Zhang" and "pot-WNLS" form a better estimator class in most cases. In particular, in terms of the RMSE, the "pot-WNLS" performs best for VaR 99.9% and 99.99% under all models considered, with one exception. In terms of the ARB, the "pot-WNLS" always performs best for the VaR 99.99% and best in most cases for the VaR 99.9%.

Example 4.1. GPD(ξ, σ).

We consider three different parameter pairs for the VaR estimation under the GPD: $(\xi, \sigma) = (1, 10)$, $(2, 1)$ and $(4, 1)$. As shown in Table 2, there is no clear-cut winner in estimating the VaR 98% and 99%, but our "pot-WNLS" gives better results for all parameter choices in terms of both RMSE and ARB as we go further towards the extreme quantile levels, such as VaR 99.9% and 99.99%, with larger u values.

Example 4.2. Cauchy (α, β).

The Cauchy distribution is another heavy-tailed distribution with no finite mean. The distribution function of Cauchy is

$$F_{\alpha,\beta}(x) = \frac{1}{\pi}\arctan\left(\frac{x-\alpha}{\beta}\right), \quad -\infty < x < \infty, \tag{20}$$

where α is the location parameter and β is the scale parameter. We generated samples from Cauchy with $(\alpha, \beta) = (0, 1)$ and presented the results in Table 3. The numbers show that our "pot-WNLS" gives the smallest RMSE for the VaR 99.9% and 99.99%. In terms of the ARB, "pot-WNLS" is still the best choice for VaR 99.99%, but for VaR 99.9%, the "LME" and "Zhang" are sometimes better by a slight margin.

Example 4.3. Log-gamma (α, β).

The Log-gamma is derived from a transformed gamma variable. If X is a gamma random variable with parameter (α, β) where the mean is $\alpha\beta$, it is known that $Y = \exp(X)$ follows a Log-gamma distribution with density function

$$f_{\alpha,\beta}(y) = \frac{(\log y)^{\alpha-1}}{y\beta^{\alpha}\Gamma(\alpha)}\exp\left(-\log y/\beta\right), \quad y > 1, \tag{21}$$

Table 2
VaR estimation under GPD (u is threshold).

		RMSE				ARB			
		VaR 98%	VaR 99%	VaR 99.9%	VaR 99.99%	VaR 98%	VaR 99%	VaR 99.9%	VaR 99.99%
GPD(1,10)									
$u = 97\%$	MED	35.40	106.54	7946.11	5.61×10^5	0.0576	0.0849	0.4822	1.4913
	Pickands	36.64	94.13	6363.29	2.79×10^5	0.0594	0.0757	0.4304	1.2468
	LME	33.86	89.24	2665.95	6.23×10^4	0.0550	0.0715	0.2047	0.4299
	Zhang	33.67	88.95	2668.93	6.26×10^4	0.0548	0.0713	0.2046	0.4311
	pot-NLS	33.85	89.60	3470.96	9.36×10^4	0.0551	0.0721	0.2603	0.5864
	pot-WNLS	33.84	88.95	2635.37	5.87×10^4	0.0550	0.0716	0.2080	0.4195
$u = 98\%$	MED		99.71	8665.51	1.06×10^6		0.0797	0.5047	2.0788
	Pickands		99.03	6339.41	3.96×10^5		0.0795	0.4253	1.5032
	LME		94.64	2771.04	7.35×10^4		0.0758	0.2119	0.4898
	Zhang		93.88	2783.78	7.41×10^4		0.0752	0.2125	0.4943
	pot-NLS		93.41	3600.98	1.19×10^5		0.0752	0.2676	0.6857
	pot-WNLS		93.94	2705.26	6.62×10^4		0.0755	0.2143	0.4673
$u = 99\%$	MED			8780.79	1.37×10^7			0.4724	5.0633
	Pickands			5354.20	6.50×10^5			0.3707	2.0383
	LME			2822.65	9.82×10^4			0.2148	0.5963
	Zhang			2841.45	9.96×10^4			0.2158	0.6078
	pot-NLS			3338.88	1.55×10^5			0.2526	0.8233
	pot-WNLS			2686.33	8.18×10^4			0.2147	0.5449
GPD(2,1)									
$u = 97\%$	MED	182.16	1129.20	1.03×10^6	1.38×10^9	0.1151	0.1736	0.9768	4.3838
	Pickands	191.83	958.58	6.92×10^5	5.34×10^8	0.1207	0.1505	0.7711	2.8090
	LME	176.23	892.69	3.00×10^5	8.74×10^7	0.1113	0.1398	0.4179	0.9202
	Zhang	172.20	887.56	2.88×10^5	7.93×10^7	0.1089	0.1391	0.4060	0.8686
	pot-NLS	171.67	906.12	3.54×10^5	1.20×10^8	0.1088	0.1427	0.4778	1.1051
	pot-WNLS	173.86	882.42	2.70×10^5	7.11×10^7	0.1100	0.1396	0.3929	0.7910
$u = 98\%$	MED		1026.33	1.32×10^6	3.92×10^9		0.1595	1.0877	8.4030
	Pickands		1011.71	7.48×10^5	1.20×10^9		0.1581	0.7974	4.0328
	LME		968.95	3.25×10^5	1.20×10^8		0.1510	0.4445	1.1205
	Zhang		947.95	3.15×10^5	1.08×10^8		0.1481	0.4347	1.0562
	pot-NLS		937.69	3.91×10^5	2.04×10^8		0.1475	0.5104	1.3992
	pot-WNLS		946.59	2.85×10^5	8.94×10^7		0.1487	0.4131	0.9146
$u = 99\%$	MED			1.40×10^6	2.78×10^{10}			1.1213	31.0613
	Pickands			6.71×10^5	2.68×10^9			0.7461	7.2471
	LME			3.49×10^5	2.28×10^8			0.4666	1.6058
	Zhang			3.40×10^5	1.83×10^8			0.4607	1.4722
	pot-NLS			3.88×10^5	2.82×10^8			0.5128	1.9012
	pot-WNLS			2.92×10^5	1.32×10^8			0.4255	1.1398
GPD(4,1)									
$u = 97\%$	MED	4.85×10^5	1.31×10^7	2.76×10^{12}	1.50×10^{18}	0.2368	0.3697	2.9213	44.4741
	Pickands	1.06×10^6	1.76×10^7	1.53×10^{12}	9.33×10^{17}	0.5473	0.5767	1.7244	19.6077
	LME	1.04×10^6	1.72×10^7	4.27×10^{11}	3.21×10^{16}	0.5287	0.5604	1.0103	2.6578
	Zhang	1.03×10^6	1.72×10^7	3.81×10^{11}	2.16×10^{16}	0.5208	0.5596	0.9584	2.1473
	pot-NLS	1.05×10^6	1.77×10^7	4.84×10^{11}	5.34×10^{16}	0.5343	0.5851	1.0504	2.8428
	pot-WNLS	1.05×10^6	1.75×10^7	3.48×10^{11}	2.14×10^{16}	0.5371	0.5774	0.9116	1.8662
$u = 98\%$	MED		1.14×10^7	8.63×10^{12}	1.30×10^{20}		0.3322	3.8749	584.1382
	Pickands		1.78×10^7	2.22×10^{12}	9.08×10^{18}		0.5870	1.9234	76.3880
	LME		1.76×10^7	4.98×10^{11}	6.72×10^{16}		0.5750	1.0817	3.9995
	Zhang		1.74×10^7	4.63×10^{11}	6.33×10^{16}		0.5686	1.0281	3.1527
	pot-NLS		2.26×10^7	3.20×10^{11}	2.49×10^{16}		0.8686	1.0168	1.6699
	pot-WNLS		2.26×10^7	2.74×10^{11}	1.26×10^{16}		0.8687	0.9790	1.2858
$u = 99\%$	MED			4.80×10^{12}	9.99×10^{19}			4.0733	1313.3745
	Pickands			1.75×10^{12}	1.29×10^{19}			1.8664	221.1891
	LME			6.02×10^{11}	3.83×10^{17}			1.1691	10.6399
	Zhang			5.32×10^{11}	1.43×10^{17}			1.1174	6.2146
	pot-NLS			2.59×10^{11}	1.17×10^{16}			1.0024	1.0655
	pot-WNLS			2.53×10^{11}	3.92×10^{15}			1.0006	1.0210

where $\alpha > 0$ and $\beta > 0$. The Log-gamma is a member of MDA(Φ_β) so that $1/\beta$ represents the tail index. The simulation results for $\alpha = 2$ and $\beta = 3$ in the middle of Table 3 show us a similar conclusion, confirming the dominance of the "pot-WNLS" at VaR 99.9%, 99.99% in terms of the RMSE. Interestingly, the "pot-NLS" is ranked second in this specific model in the

Table 3
VaR estimation under Cauchy, Log-gamma and Pareto (u is threshold).

		RMSE				ARB			
		VaR 98%	VaR 99%	VaR 99.9%	VaR 99.99%	VaR 98%	VaR 99%	VaR 99.9%	VaR 99.99%
Cauchy(0,1)									
$u = 97\%$	MED	1.128	3.438	261.456	18 172.133	0.0563	0.0846	0.4933	1.5444
	Pickands	1.168	3.019	201.912	9186.911	0.0584	0.0752	0.4272	1.2296
	LME	1.084	2.854	85.269	2001.941	0.0541	0.0709	0.2046	0.4302
	Zhang	1.076	2.841	85.395	2005.070	0.0537	0.0706	0.2046	0.4313
	pot-NLS	1.078	2.868	110.968	2968.272	0.0538	0.0715	0.2611	0.5862
	pot-WNLS	1.082	2.843	84.143	1869.316	0.0539	0.0709	0.2075	0.4197
$u = 98\%$	MED		3.225	271.673	26 379.502		0.0801	0.4985	2.0222
	Pickands		3.202	198.101	11 515.164		0.0797	0.4231	1.4641
	LME		3.049	88.424	2356.160		0.0756	0.2112	0.4864
	Zhang		3.023	88.875	2390.725		0.0750	0.2116	0.4904
	pot-NLS		3.007	113.243	3670.554		0.0748	0.2645	0.6711
	pot-WNLS		3.026	86.571	2169.523		0.0753	0.2136	0.4684
$u = 99\%$	MED			288.634	285 515.096			0.4812	5.1274
	Pickands			171.118	21 683.572			0.3704	2.0642
	LME			90.144	3115.165			0.2147	0.5919
	Zhang			91.101	3287.340			0.2159	0.6076
	pot-NLS			107.808	5981.896			0.2532	0.8371
	pot-WNLS			86.194	2689.458			0.2142	0.5451
Log-gam(2,3)									
$u = 97\%$	MED	1.06×10^7	1.97×10^8	1.23×10^{13}	2.21×10^{18}	0.2043	0.3203	3.0252	65.9167
	Pickands	1.18×10^7	1.57×10^8	5.89×10^{12}	5.64×10^{17}	0.2244	0.2645	1.8811	24.7519
	LME	1.07×10^7	1.39×10^8	1.59×10^{12}	2.23×10^{16}	0.2038	0.2380	0.8108	3.1326
	Zhang	9.93×10^6	1.38×10^8	1.57×10^{12}	2.15×10^{16}	0.1913	0.2364	0.8132	3.0819
	pot-NLS	2.95×10^7	4.25×10^8	1.14×10^{12}	7.76×10^{15}	0.7258	0.9381	0.9984	1.1937
	pot-WNLS	2.95×10^7	4.25×10^8	1.08×10^{12}	3.93×10^{15}	0.7261	0.9375	0.9830	1.0751
$u = 98\%$	MED		1.74×10^8	1.50×10^{13}	1.05×10^{19}		0.2878	3.1638	168.8302
	Pickands		1.70×10^8	5.85×10^{12}	1.03×10^{18}		0.2829	1.8452	39.1523
	LME		1.58×10^8	1.76×10^{12}	3.85×10^{16}		0.2639	0.8593	4.0542
	Zhang		1.49×10^8	1.68×10^{12}	2.91×10^{16}		0.2523	0.8476	3.6611
	pot-NLS		4.03×10^8	1.09×10^{12}	5.89×10^{15}		0.9017	1.0011	1.0556
	pot-WNLS		4.03×10^8	1.07×10^{12}	2.51×10^{15}		0.9017	0.9979	1.0198
$u = 99\%$	MED			1.52×10^{13}	1.03×10^{20}			3.2460	1138.6472
	Pickands			6.36×10^{12}	2.23×10^{19}			1.7480	253.0998
	LME			2.01×10^{12}	2.25×10^{17}			0.9187	8.0632
	Zhang			1.89×10^{12}	6.96×10^{16}			0.9018	6.0299
	pot-NLS			1.07×10^{12}	2.09×10^{15}			0.9994	1.0011
	pot-WNLS			1.07×10^{12}	2.10×10^{15}			0.9994	1.0012
Pareto(2, $\frac{1}{2}$)									
$u = 97\%$	MED	7.29×10^2	4.52×10^3	4.14×10^6	5.50×10^9	0.1150	0.1736	0.9769	4.3848
	Pickands	7.67×10^2	3.83×10^3	2.77×10^6	2.14×10^9	0.1207	0.1505	0.7711	2.8090
	LME	7.05×10^2	3.57×10^3	1.20×10^6	3.50×10^8	0.1112	0.1398	0.4179	0.9202
	Zhang	6.89×10^2	3.55×10^3	1.15×10^6	3.17×10^8	0.1088	0.1391	0.4060	0.8686
	pot-NLS	6.87×10^2	3.62×10^3	1.42×10^6	4.78×10^8	0.1087	0.1427	0.4778	1.1051
	pot-WNLS	6.95×10^2	3.53×10^3	1.08×10^6	2.85×10^8	0.1099	0.1396	0.3929	0.7910
$u = 98\%$	MED		4.11×10^3	75.29×10^6	1.57×10^{10}		0.1594	1.0880	8.4095
	Pickands		4.05×10^3	2.99×10^6	4.79×10^9		0.1581	0.7974	4.0328
	LME		3.88×10^3	1.30×10^6	4.79×10^8		0.1510	0.4445	1.1205
	Zhang		3.79×10^3	1.26×10^6	4.32×10^8		0.1481	0.4347	1.0562
	pot-NLS		3.75×10^3	1.56×10^6	8.17×10^8		0.1475	0.5104	1.3992
	pot-WNLS		3.79×10^3	1.14×10^6	3.58×10^8		0.1487	0.4131	0.9146
$u = 99\%$	MED			5.63×10^6	1.13×10^{11}			1.1224	31.5062
	Pickands			2.68×10^6	1.07×10^{10}			0.7461	7.2471
	LME			1.40×10^6	9.14×10^8			0.4666	1.6058
	Zhang			1.36×10^6	7.32×10^8			0.4607	1.4722
	pot-NLS			1.55×10^6	1.13×10^9			0.5128	1.9012
	pot-WNLS			1.17×10^6	5.29×10^8			0.4255	1.1398

RMSE sense. In terms of the ARB, the "pot-WNLS" is still the best for VaR 99.99%, but not so for VaR 99.9%, where the "LME" and "Zhang" are better. At lower VaR levels at 98% and 99%, the "Zhang" estimator performs the best.

Table 4
CTE estimation for GPD ($\xi, \sigma = 1$).

GPD		CTE 97%		CTE 99%		CTE 99.9%	
Threshold u		$u = 95\%$		$u = 98\%$		$u = 99\%$	
		$\xi = 0.3$	$\xi = 0.75$	$\xi = 0.3$	$\xi = 0.75$	$\xi = 0.3$	$\xi = 0.75$
MED	RMSE	1.651	127.99	10.171	326.58	106.51	2503.01
	ARB	0.0823	0.7035	0.1885	0.7941	0.7749	1.0888
	Failure	0	1824	105	3772	535	4961
Pickands	RMSE	1.344	121.78	9.216	312.16	108.44	2554.34
	ARB	0.0836	0.6729	0.1863	0.7763	0.8391	1.1184
	Failure	0	1838	71	3818	494	5024
LME	RMSE	0.432	39.51	1.062	210.12	7.13	2420.01
	ARB	0.0334	0.2541	0.0535	0.4992	0.1509	0.9699
	Failure	0	23	0	470	0	1402
Zhang	RMSE	0.434	44.17	1.079	224.47	7.80	2480.09
	ARB	0.0335	0.2628	0.0541	0.5319	0.1581	1.0255
	Failure	0	24	0	513	0	1644
pot-NLS	RMSE	0.606	80.78	1.755	278.27	34.49	2517.18
	ARB	0.0457	0.4311	0.0784	0.6679	0.2920	1.0576
	Failure	0	256	0	1510	12	2755
pot-WNLS	RMSE	0.456	37.70	1.112	181.49	7.77	2052.42
	ARB	0.0355	0.2533	0.0569	0.4467	0.1640	0.8390
	Failure	0	26	0	391	0	1085

Example 4.4. Pareto (θ, α).

The distribution of the Pareto (type I) random variable is given by

$$F(x) = 1 - \left(\frac{x-\mu}{\theta}\right)^{-\alpha}, \quad x \geq \mu + \theta, \tag{22}$$

where μ is the location parameter, θ is the scale parameter and α is the shape parameter. Clearly, the Pareto distribution is in MDA($\Phi_{1/\alpha}$). We generated Pareto samples with the $\theta = 2$ and the $\alpha = \frac{1}{2}$, of which the results are presented at the bottom of Table 3. From the numbers, we see that the "pot-WNLS" outperforms other methods at 99%, 99.9% and 99.99% in terms of the RMSE; this pattern also repeats for the ARB except for 99%, in which case "Zhang" is slightly better.

4.2. CTE estimation

Now we turn to the CTE estimation and compare the same six estimators for simulated datasets from the GPD, Pareto and Log-gamma distributions, respectively, with varying shape parameters. The simulation procedure is similar. We first pick the 100pth sample quantile of the dataset as the threshold u and fit the exceedances to the GPD to obtain $\hat{\xi}$ and $\hat{\sigma}$. The fitted GPD then gives the CTE estimate (12), which can be compared against the true CTE of the given model. We note that the true CTE values for all models have been derived from the corresponding distribution function and available in closed-from. The sample size is again 10,000 with 20,000 repetitions for each model. The simulation results of CTE 97%, 99% and 99.9% are presented in Tables 4–6, for each model. We do not consider CTE 99.99% as it is, being way larger than the VaR 99.99%, too volatile for such heavy-tailed distributions. One distinct consideration in the CTE estimation is that the CTE does not exist for $\xi > 1$ due to the moment existence condition. Besides, as seen from (12), if one single estimate of $\hat{\xi}$ out of 20,000 repetitions turns out to be too close to 1, even though it is strictly less than 1, it produces a prohibitively large CTE estimate and yields a ruinous impact on the bias and variance of the estimate, which prevents us from any realistic comparison. To appropriately compare these estimates therefore we have excluded all occurrences that give $\hat{\xi} > 0.99$ in our RMSE and ARB computations. The number of occurrences excluded are separately recorded in the tables (coined as 'Failure') as this also provides valuable information about the estimate's quality.

Example 4.5. GPD(ξ, σ).

Using the 100pth quantile $Q_p(X) = \frac{\sigma}{\xi}[(1-p)^{-\xi} - 1]$, we obtain the CTE of the GPD, with a bit of calculus, as

$$CTE_p(X) = \frac{\sigma}{\xi}\left[\frac{(1-p)^{-\xi}}{1-\xi} - 1\right], \quad 0 < \xi < 1.$$

The samples were generated from the GPD with two different shape parameters: $(\xi, \sigma) = (0.3, 1)$ and (0.75,1). Recall that, for VaR estimations, our estimator "pot-WLNS" outperformed other estimator as the shape parameter ξ gets large. Similarly, for CTE estimations in Table 4, our estimator has the smallest RMSE and ARB when $\xi = 0.75$ at all CTE levels considered.

Table 5
CTE estimation for Log-gamma ($\alpha = 2, \beta$).

Log-gamma		CTE 97%		CTE 99%		CTE 99.9%	
Threshold u		$u = 95\%$		$u = 98\%$		$u = 99\%$	
		$\beta = 0.3$	$\beta = 0.6$	$\beta = 0.3$	$\beta = 0.6$	$\beta = 0.3$	$\beta = 0.6$
MED	RMSE	6.76	149.73	13.36	389.53	89.38	2838.44
	ARB	5.9760	9.9147	8.0009	12.5500	16.2393	19.4633
	Failure	1	796	118	2329	701	3680
Pickands	RMSE	6.71	147.59	12.76	372.59	83.27	2806.21
	ARB	5.9306	9.7568	7.8333	12.2460	15.9987	18.9272
	Failure	3	758	89	2263	601	3807
LME	RMSE	6.50	72.74	9.75	200.66	22.54	1759.73
	ARB	5.7937	7.2575	7.0926	9.3396	9.8530	15.5169
	Failure	0	0	0	52	0	357
Zhang	RMSE	6.51	73.60	9.78	210.60	23.23	1909.58
	ARB	5.8002	7.3223	7.1171	9.5938	10.1056	16.7392
	Failure	0	0	0	64	0	461
pot-NLS	RMSE	6.52	90.42	9.91	298.91	32.57	2490.85
	ARB	5.8104	7.9540	7.1673	10.9484	10.9653	18.8173
	Failure	0	25	0	491	27	1433
pot-WNLS	RMSE	6.46	72.23	9.59	190.98	21.55	1625.21
	ARB	5.7547	7.1391	6.9812	8.8679	9.3461	13.5634
	Failure	0	0	0	44	0	304

Table 6
CTE estimation for Pareto ($\theta = 2, \alpha$).

Pareto		CTE 97%		CTE 99%		CTE 99.9%	
Threshold u		$u = 95\%$		$u = 98\%$		$u = 99\%$	
		$\alpha = 1.5$	$\alpha = 2$	$\alpha = 1.5$	$\alpha = 2$	$\alpha = 1.5$	$\alpha = 2$
MED	RMSE	93.97	12.216	247.96	59.472	1979.30	492.14
	ARB	0.5205	0.1990	0.7072	0.4546	1.1917	1.1672
	Failure	677	41	2369	723	3667	1779
Pickands	RMSE	83.78	12.566	246.28	55.350	2004.44	481.18
	ARB	0.4965	0.1967	0.7177	0.4437	1.2287	1.1710
	Failure	688	45	2343	639	3719	1785
LME	RMSE	12.38	1.793	96.99	6.015	1309.07	95.65
	ARB	0.1358	0.0602	0.2862	0.1075	0.7516	0.3280
	Failure	0	0	60	0	443	17
Zhang	RMSE	12.94	1.819	106.39	6.304	1448.96	134.00
	ARB	0.1384	0.0607	0.3037	0.1106	0.8246	0.3628
	Failure	0	0	68	0	544	22
pot-NLS	RMSE	35.64	2.874	179.43	19.127	1745.67	266.49
	ARB	0.2292	0.0882	0.4848	0.1862	1.0219	0.6496
	Failure	23	0	503	26	1472	272
pot-WNLS	RMSE	13.13	1.879	87.36	6.007	1097.86	95.61
	ARB	0.1408	0.0637	0.2748	0.1112	0.6579	0.3298
	Failure	0	0	44	0	374	15

For $\xi = 0.3$, "LME" performs best in all cases. Thus a better performance of our estimate for heavier tails is reaffirmed. The number of failures shows that "MED" and "Pickands" methods produce the shape parameter over 0.99 quite frequently, almost 25% of the simulations at CTE 99.9%. In contrast, the other methods' failure occurrences are considerably smaller, with "pot-WLNS" having the smallest except for the CTE 97% case, indicating its competitive performance.

Example 4.6. Log-gamma (α, β).

We again let $Y = e^X$ be a Log-gamma variable where X is a gamma random variable with parameter (α, β). After $Q_p(Y)$ has been calculated numerically, the true CTE of Y with (α, β), $0 < \beta < 1$, is obtained from a variable transformation:

$$\begin{aligned} CTE_p(Y) &= E[e^X | e^X > Q_p(e^X)] = E[e^X | X > Q_p(X)] \\ &= \frac{1}{1-p} \int_{Q_p(X)}^{\infty} e^x f_X(x) dx = \frac{\exp\left(Q_p(X)\left(1 - \frac{1}{\beta}\right)\right)}{(1-p)(1-\beta)}. \end{aligned}$$

The second equality is valid because quantiles are preserved under an increasing monotone transformation. We generated samples from this distribution with $(\alpha, \beta) = (2, 0.3)$ and $(2, 0.6)$, and the results are presented in Table 5. From the table our estimator "pot-WNLS" performs best unanimously across all CTE levels considered and for all parameter choices. Furthermore, our estimator is least likely to fall in the failing range for all cases.

Example 4.7. Pareto(I) (θ, α).

Using the 100pth quantile of the Pareto, $Q_p(X) = \theta(1-p)^{-1/\alpha}$, its true CTE is obtained as

$$CTE_p(X) = \theta(1-p)^{-1/\alpha}\left(\frac{\alpha}{\alpha-1}\right), \quad 1 < \alpha.$$

Samples were generated from the Pareto with $(\theta, \alpha) = (2, 1.5)$ and $(2, 2)$, where the condition $\alpha > 1$ ensures a finite mean. As shown in Table 6, when high threshold values are used, i.e., $u = 98\%$ and 99%, the "pot-WNLS" method gives the best results on the CTE values, for both α choices. When $u = 95\%$, "LME" performs the best for both α choices. For the number of failures, "pot-WNLS" again shows the smallest among other estimators.

5. Application to a realistic dataset

One may wish to compare the performance of different estimators for actual datasets. However, such a task is not easy as the true value of the tail risk measure is generally unknown. It could be argued that, provided that the sample size is large enough, the empirical tail measure may serve as a feasible substitute of the true measure, but the variability of the empirical risk measure can actually be substantial even for large samples. In fact, recognizing that the VaR and CTE are special cases of L-statistics in the statistical literature, the asymptotic variance of the 100p% sample quantile and CTE are given by (see, e.g., Staudte and Sheather (1990) and Manistre and Hancock (2005))

$$\mathrm{Var}(\widehat{Q}_p(X)) \approx \frac{p(1-p)}{n\,(f(Q_p(X)))^2} \tag{23}$$

and

$$\mathrm{Var}(\widehat{CTE}_p(X)) \approx \frac{\mathrm{Var}(X|X > Q_p(X)) + p(CTE_p(X) - Q_p(X))^2}{n\,(1-p)}. \tag{24}$$

Though these variances decay at $O(n^{-1})$, when evaluated at finite samples, their values critically depend on confidence level p and the tail thickness of the underlying distribution through its density f or the CTE of the underlying distribution. Thus, for heavy-tailed datasets with $p \approx 1$, the variances of the estimates could easily produce unacceptably large values, invalidating further analyses. To avoid this situation, we consider in this section a realistic parametric model which models the entire range of some actual loss dataset, but still is in the MDA of the Fréchet distribution, so that the GPD under the POT framework can be properly applied. The model we consider is the Log Phase-type (LogPH) distribution class of Ahn et al. (2012), which is created via the exponential transformation of a Phase-type (PH) random variable, defined as the distribution of the time until absorption in a continuous time Markov chain with an absorbing state, as introduced by Neuts (1975). The class of PH distributions contains popular distributions such as the Exponential, Erlang, and their finite mixtures, and has attractive distributional properties. A notable property of the LogPH class is that it can approximate any non-negative distribution to any desired accuracy. The distribution function of the LogPH, Y, is

$$F_Y(y) = 1 - \alpha e^{T\log y}\mathbf{1}, \quad \text{for } y \geq 1, \tag{25}$$

where $\mathbf{1}$ is a p dimensional column vector consisting of ones, α is a p dimensional row vector and T is a $p \times p$ infinitesimal generator matrix. Ahn et al. (2012) showed that the LogPH distribution, being in the MDA of the Fréchet distribution, adequately models the entire range of the famous Danish fire insurance data[2] with $p = 2$. Their estimated parameters are $\hat{\alpha} = (0.622, 0.378)$ and

$$\hat{T} = \begin{bmatrix} -4.000 & 3.564 \\ 0.267 & -1.813 \end{bmatrix}.$$

Fig. 2 shows the LogPH fit of the dataset. The resulting model indicates that the moment exists only up to about 1.4, consistent with previous findings in the literature.

For our purpose, we generate samples of size 10,000 with 20,000 repetitions from this LogPH model and compare the performances of various tail risk measure estimates. Tables 7 and 8 present the results for the VaR and CTE estimation for this model, with the true tail risk measures which are obtained analytically as shown in the original paper. For the

[2] Retrieved from Dr. A. McNeil's webpage: http://www.ma.hw.ac.uk/mcneil/data.html; see McNeil et al. (2005) for further discussions and other references on the Danish fire data.

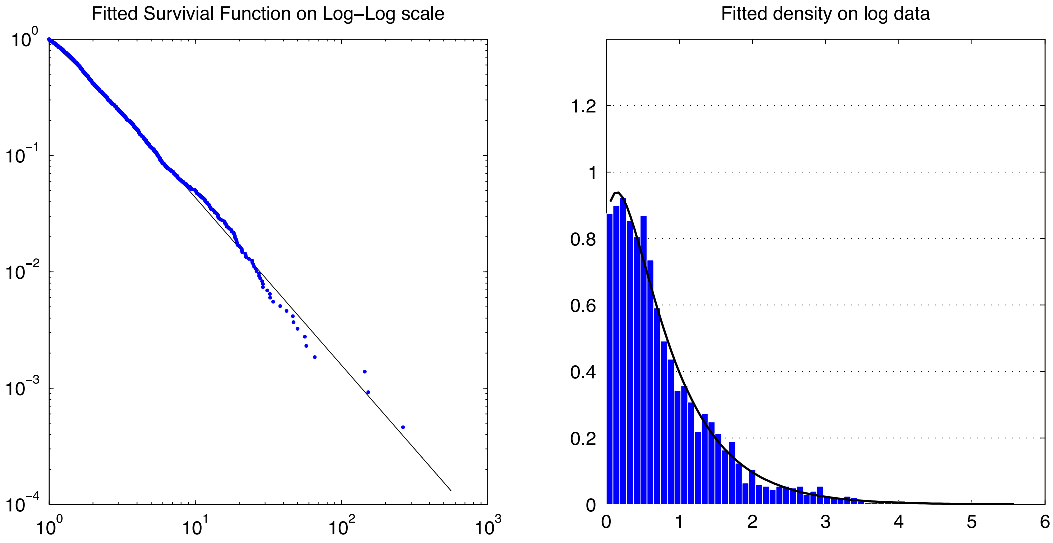

Fig. 2. Fitted LogPH model for Danish fire data.

Table 7
VaR estimation for Danish fire model.

		RMSE				ARB			
		VaR 98%	VaR 99%	VaR 99.9%	VaR 99.99%	VaR 98%	VaR 99%	VaR 99.9%	VaR 99.99%
$u = 97\%$	MED	0.829	2.016	75.178	1898.143	0.0385	0.0577	0.3535	1.0599
	Pickands	0.856	1.801	64.106	1269.995	0.0398	0.0519	0.3296	0.9588
	LME	0.792	1.720	25.276	294.219	0.0368	0.0497	0.1428	0.3121
	Zhang	0.790	1.716	25.514	301.350	0.0368	0.0496	0.1437	0.3171
	pot-NLS	0.798	1.723	34.837	470.379	0.0372	0.0499	0.1935	0.4613
	pot-WNLS	0.795	1.718	25.842	292.705	0.0370	0.0497	0.1497	0.3211
$u = 98\%$	MED		1.9391	76.457	2538.902		0.0557	0.3557	1.3248
	Pickands		1.9281	62.739	1843.811		0.0555	0.3180	1.1304
	LME		1.8075	25.869	334.630		0.0520	0.1459	0.3473
	Zhang		1.7993	26.162	345.848		0.0519	0.1469	0.3548
	pot-NLS		1.8028	34.885	562.576		0.0521	0.1929	0.5200
	pot-WNLS		1.8066	26.218	326.616		0.0521	0.1519	0.3525
$u = 99\%$	MED			72.298	7415.608			0.3240	2.2043
	Pickands			51.548	2795.047			0.2685	1.4795
	LME			26.005	409.610			0.1466	0.4057
	Zhang			26.294	433.278			0.1477	0.4216
	pot-NLS			31.884	746.959			0.1789	0.6249
	pot-WNLS			25.823	381.645			0.1509	0.4048

VaR estimation, 97th, 98th and 99th sample quantiles were selected as threshold u and we estimated the VaR 98%, 99%, 99.9%, and 99.99%. Our estimator "pot-WNLS" gives the smallest RMSE at VaR 99.99% for all threshold choices and has best results at VaR 99.9% and VaR 99.99% when the threshold u is 99th sample quantile. In other cases, "LME" and "Zhang" are slightly better. Overall, "pot-WNLS" again tends to outperform other estimators as we go towards the extreme quantile levels. Similarly, we estimated the CTE for the 97%, 99% and 99.9%, where the 95th, 98th and 99th sample quantiles were selected as the threshold. The numbers of Table 8 shows that our estimator "pot-WNLS" performs best at CTE 99% and 99.9% in terms of both RMSE and ARB, with the smallest numbers of failure at the same levels. Again, our estimator gives the best results as the threshold and quantiles go to extreme cases.

6. Concluding remarks

We proposed a new procedure to fit the GPD and estimate tail risk measures at extreme quantile levels under the POT framework. Our method is adapted from the recent NLS estimator of Song and Song (2012) which uses nonlinear minimization of the squared error between the GPD distribution function and the empirical distribution. We critically examine the NLS method and point out its caveat. After an appropriate revision on the NLS, we further introduce its improved version by adding a suitable weight for each empirical distribution point. From extensive numerical studies, using both simulated samples and a more realistic loss dataset, we found that the performance of the proposed estimator is highly

Table 8
CTE estimation for Danish fire model.

Actual data Threshold u		CTE 97% $u = 95\%$	CTE 99% $u = 98\%$	CTE 99.9% $u = 99\%$
MED	RMSE	68.28	179.66	1505.86
	ARB	0.5792	0.7435	1.2401
	Failure	978	2763	4160
Pickands	RMSE	64.91	177.34	1370.34
	ARB	0.5692	0.7536	1.1887
	Failure	982	2692	4190
LME	RMSE	12.53	82.97	1043.36
	ARB	0.1635	0.3452	0.8204
	Failure	1	137	634
Zhang	RMSE	12.37	89.27	1109.79
	ARB	0.1659	0.3643	0.8895
	Failure	2	167	814
pot-NLS	RMSE	31.39	130.56	1287.47
	ARB	0.2804	0.5264	1.0611
	Failure	60	793	1905
pot-WNLS	RMSE	12.42	78.83	957.27
	ARB	0.1665	0.3295	0.7546
	Failure	3	129	521

competitive in estimating extreme VaR and CTE for heavy-tailed data. In particular the proposed estimator performs better than alternative estimators as the shape parameter ξ of the GPD increases and the tail threshold u gets larger.

Acknowledgement

J.H.T. Kim is grateful for the support of the National Research Foundation of Korea (NRF-2015R1A1A1A05027336).

References

Ahn, S., Kim, J.H.T., Ramaswami, V., 2012. A new class of models for heavy tailed distributions in finance and insurance risk. Insurance Math. Econom. 51, 43–52.
Artzner, P., Delbaen, F., Eber, J.M., Heath, D., 1999. Coherent measures of risk. Math. Finance 9, 203–228.
Ashkar, F., Ouarda, T.B., 1996. On some methods of fitting the generalized pareto distribution. J. Hydrol. 177, 117–141.
Balkema, A.A., De Haan, L., 1974. Residual life time at great age. Ann. Probab. 792–804.
BCBS 2006. International convergence of capital measurement and capital standards: A revised framework, Comprehensive version. Basel Committee on Banking Supervision.
Beirlant, J., Goegebeur, Y., Segers, J., Teugels, J., 2006. Statistics of Extremes: Theory and Applications. John Wiley & Sons.
Chen, F., Lambert, D., Pinheiro, J.C., 2000. Incremental quantile estimation for massive tracking. In: Proceedings of the Sixth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, pp. 516–522.
Davison, A.C., 1984. Modelling excesses over high thresholds, with an application. In: Statistical Extremes and Applications. Springer, pp. 461–482.
de Zea Bermudez, P., Kotz, S., 2010. Parameter estimation of the generalized pareto distribution part i. J. Statist. Plann. Inference 140, 1353–1373.
Dupuis, D., Tsao, M., 1998. A hybrid estimator for generalized pareto and extreme-value distributions. Comm. Statist. Theory Methods 27, 925–941.
Embrechts, P., Klüppelberg, C., Mikosch, T., 1997. Modelling Extremal Events, Vol. 33. Springer Science & Business Media.
Grimshaw, S.D., 1993. Computing maximum likelihood estimates for the generalized pareto distribution. Technometrics 35, 185–191.
He, X., Fung, W., 1993. Method of medians for lifetime data with weibull models. Stat. Med. 18, 2009.
Hosking, J.R., Wallis, J.R., 1987. Parameter and quantile estimation for the generalized pareto distribution. Technometrics 29, 339–349.
Juárez, S.F., Schucany, W.R., 2004. Robust and efficient estimation for the generalized pareto distribution. Extremes 7, 237–251.
Liechty, J.C., Lin, D.K., McDermott, J.P., 2003. Single-pass low-storage arbitrary quantile estimation for massive datasets. Stat. Comput. 13, 91–100.
Manistre, B.J., Hancock, G.H., 2005. Variance of the CTE estimator. N. Am. Actuar. J. 9, 129–156.
McNeil, A.J., Frey, R., Embrechts, P., 2005. Quantitative Risk Management. Princeton University Press, New Jersey.
McNeil, A.J., Saladin, T., 1997. The peaks over thresholds method for estimating high quantiles of loss distributions. In: Proceedings of 28th International ASTIN Colloquium, pp. 23–43.
Munro, J.I., Paterson, M.S., 1980. Selection and sorting with limited storage. Theoret. Comput. Sci. 12, 315–323.
Nelder, J.A., Mead, R., 1965. A simplex method for function minimization. Comput. J. 7, 308–313.
Neuts, M., 1975. Probability distributions of phase type. In: Liber Amicorum Prof. Emeritus H. Florin. University of Louvain, Belgium, pp. 173–206.
Pickands III, J., 1975. Statistical inference using extreme order statistics. Ann. Statist. 119–131.
Rasmussen, P.F., 2001. Generalized probability weighted moments: application to the generalized pareto distribution. Water Resour. Res. 37, 1745–1751.
Rohatgi, V.K., 1984. Statistical Inference. John Wiley & Sons.
Smith, R.L., 1984. Threshold methods for sample extremes. In: Statistical Extremes and Applications. Springer, pp. 621–638.
Smith, R.L., 1985. Maximum likelihood estimation in a class of nonregular cases. Biometrika 72, 67–90.
Song, J., Song, S., 2012. A quantile estimation for massive data with generalized pareto distribution. Comput. Statist. Data Anal. 56, 143–150.
Staudte, R.G., Sheather, S.J., 1990. Robust Estimation and Testing. John Wiley & Sons, New York.
Zhang, J., 2007. Likelihood moment estimation for the generalized pareto distribution. Aust. N. Z. J. Stat. 49, 69–77.
Zhang, J., 2010. Improving on estimation for the generalized pareto distribution. Technometrics 52, 335–339.
Zhang, J., Stephens, M.A., 2009. A new and efficient estimation method for the generalized pareto distribution. Technometrics 51, 316–325.