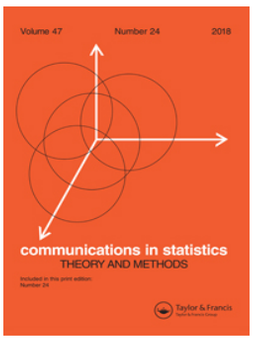

Taylor & Francis
Taylor & Francis Group

Communications in Statistics - Theory and Methods

ISSN: 0361-0926 (Print) 1532-415X (Online) Journal homepage: https://www.tandfonline.com/loi/lsta20

Optimal threshold determination based on the mean excess plot

Queensley C. Chukwudum, Peter Mwita & Joseph K. Mung'atu

To cite this article: Queensley C. Chukwudum, Peter Mwita & Joseph K. Mung'atu (2019): Optimal threshold determination based on the mean excess plot, Communications in Statistics - Theory and Methods, DOI: 10.1080/03610926.2019.1624772

To link to this article: https://doi.org/10.1080/03610926.2019.1624772

COMMUNICATIONS IN STATISTICS—THEORY AND METHODS
https://doi.org/10.1080/03610926.2019.1624772

Optimal threshold determination based on the mean excess plot

Queensley C. Chukwudum[a,b], Peter Mwita[c], and Joseph K. Mung'atu[d]

[a]Department of Mathematics, Pan African University Institute for Basic Sciences, Technology and Innovation, Juja, Kenya; [b]Department of Insurance, University of Uyo Faculty of Business Administration, Uyo, Akwa Ibom, Nigeria; [c]Deputy Vice Chancellor Research, Innovation and Linkages, Machakos University, Kenya., Machakos, Kenya; [d]Department of Statistics and Actuarial Science, Jomo Kenyatta University of Agriculture and Technology, Kenya, Juja, Kenya

ABSTRACT

Choosing a suitable threshold has been an issue in practice. Based on the mean excess plot (MEP), the eyeball inspection approach (EIA) is mainly used to determine the threshold. This involves fitting the threshold at the point the plot becomes approximately linear solely using one's sense of judgement in such a way that Generalized Pareto model is valid. This is a rather subjective choice. In this paper, we propose an alternative way of selecting the threshold where, instead of choosing individual thresholds in isolation and testing their fit, we make use of the bootstrap aggregate of these individual thresholds which are formulated in terms of quantiles.The method incorporates the visual technique and is aimed at reducing the subjectivity associated with solely using the EIA. The new approach is implemented using simulated datasets drawn from three different distributions. An application to the NSE All share Nigerian stock index is presented. The performance of the proposed model and the EIA are judged based on standard error, Negative log likelihood, the Akaike Information Criteria and the Bayesian Information Criteria. The results show that the new technique gives similar estimates as the EIA and in some cases it performs better. In comparison to other existing methods, the proposed model performs well.

ARTICLE HISTORY
Received 19 June 2018
Accepted 23 May 2019

KEYWORDS
Mean excess plot; Peaks-over-threshold model; Order statistics; Generalized Pareto distribution; quantile modeling; bootstrap aggregation

1. Introduction

In the field of Extreme value theory (EVT), threshold selection is critical especially when the threshold-based methods (such as the Peaks-over-threshold (POT) technique) are considered. This selection has to be done first, before we can fit the Generalized Pareto Distribution (GPD) to the given data. The choice of threshold comes at a cost. A high threshold value reduces bias while increasing variance for the estimators of the GPD parameters. A low threshold on the other hand, brings about an opposite effect. Thus, choosing a suitable threshold has been an issue in practice. Hence the importance of selecting an appropriate threshold that balances the two.

CONTACT Queensley C. Chukwudum queensleyv@yahoo.com Pan African University Institute for Basic Sciences, Technology and Innovation, P.O. Box 62000-00200, Nairobi, Kenya.
Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/lsta.

In literature, not much attention has been dedicated to the issue of selecting an optimal threshold which we denote by u. Some authors who have highlighted on this problem and proffered varied solutions include Lang et al. (1999), Dupuis (1999), Zoglat et al. (2014), and Thompson et al. (2009). Suggested approaches range from graphical (approaches based on visual inspection) to analytical methods (approaches based on goodness of fit tests) or a combination of both. The MEP is an example of the graphical approach which was proposed by Davison and Smith (1990) based on the linearity of the mean excess function for the GPD. When the Hill estimator is used to estimate the tail index, Guillou and Hall (2001) proposed a way of determining the optimal threshold by choosing the number of upper order statistics k such that the mean of the bias is significantly different from zero. Examples of the analytical approaches include the Multiple Threshold method (MTM), the Square Error method (SE) and the Likelihood Ratio Test method (LRT).

Focusing on the MEP, the data is ordered and the sample mean excess function of the ordered data $\hat{M}(X_{(i)})$ is plotted against the set of ordered data $X_{(i)}$. The plot becomes linear above u from where the GPD is valid for positive shape parameter. It is important to note that if the GP model is valid for excesses above u, then is also valid for $u* > u$. Judging from where the graph is approximately linear using only the eyeball inspection approach, is a rather subjective choice. Different thresholds may be selected by different viewers of the plot which in turn affects reproducibility and further inferences. This is a key drawback. We propose an alternative way of selecting the threshold as an improvement on the visual approach. This will provide a new insight on how to tackle the problem of selecting the best threshold from the MEP with the aim of reducing the subjective nature of such a choice.

What sets our proposed model apart from other threshold estimation models is the incorporation of the underlying idea from the bootstrap aggregation technique in improving the choice of the threshold (on an existing graphical procedure-MEP) which currently solely depends on visual inspection. Tancredi et al. (2006) used the Bayesian approach to overcome the difficulty of the fixed threshold technique, making use of both the extreme and non-extreme data. Bermudez et al. (2001) consider the probability cumulated up to the threshold and estimate it based on the data and the frequency of occurrence. A mixture of two distributions (GPD and Weibull) with data dependent weights are implemented in Frigessi et al.'s (2002) model. Here the threshold choice is indirectly performed. Most resampling approaches extensively reviewed in Scarrott and MacDonald (2012) were developed to obtain the optimal tail index estimate.

This paper is organized as follows. In Section 2, the EVT is briefly discussed and the underlying assumptions of the MEP is explained. Section 3 provides the basic concept of order statistics, quantile modeling and bootstrap aggregation. The new technique is also introduced here. Applying the proposed method, Section 4 presents the results of the simulation study and it is compared to other existing methods. We apply our model to a real dataset in Section 5. The paper ends with the concluding remarks in Section 6.

2. EVT background

The foundations of EVT were developed by Fisher and Tippett (1928). Coles (2001) gives a very detailed presentation of the theoretical background of EVT. The classical

EVT comprises mainly of two approaches - the block maxima (BM) and the POT approaches. In the former, the data is divided into blocks and the generalized extreme value (GEV) distribution is fitted to the maxima. In the latter, a threshold is determined above which the excesses are fitted with the GPD.

2.1. Generalized extreme value (GEV)

Let $X_1, ..., X_n$ be a sequence of independent and identically distributed (iid) random variables with distribution function F. This could be observations of insurance claims in a year. Define

$$M_n = \max\{X_1, ..., X_n\}$$

The distribution of M_n can thus be derived as:

$$\begin{aligned}\mathbb{P}\{M_n \leq z\} &= \mathbb{P}\{X_1 \leq z, ..., X_n \leq z\} \\ &= \mathbb{P}\{X_1 \leq z\} * ... * \mathbb{P}\{X_n \leq z\} = \{F(z)\}^n\end{aligned}$$

We state a modified version of the extremal types theorem, that is, the Fisher-Tippet theorem (Fisher and Tippett 1928) (referred to as Theorem 1 below). This theorem allows for the characteristics of an asymptotic distribution for the maxima to be defined. We can estimate F_n by the limiting distribution as $n \to \infty$.

Theorem 1. If there exist sequences of constants $(a_n)_{n\geq 0}>0$ and $(b_n)_{n\geq 0}$ such that for a non-degenerate distribution G,

$$\mathbb{P}\left[\frac{M_n - b_n}{a_n} \leq z\right] \to G \qquad as\ n \to \infty \tag{2.1}$$

then G is a member of the generalized extreme value family

$$G(z) = \exp\left\{-\left(1 + \xi\left(\frac{z-\mu}{\sigma}\right)\right)_+^{\frac{-1}{\xi}}\right\} \tag{2.2}$$

with μ, the location, $\sigma>0$, the scale and ξ, the shape parameters. $z_+ = \max\{z, 0\}$.

It is important to note that the shape parameter ξ, is the dominant factor in determining which particular distribution is obtained.

Definition 1.1. Domain of Attraction.

A distribution υ is said to be in the domain of attraction of an extreme value type distribution G (either Gumbel, Fréchet, or Weibull), represented as $\upsilon \in G(\upsilon)$, if there exist $a_n>0$ and $b_n \in R$ such that the distribution of $\frac{M_n - b_n}{a_n}$ converges weakly to G, where $M_n := max_{1\leq i\leq n} X_i$ for an i.i.d. sequence $(X_i)_{i\in N}$ with distribution υ.

2.2. The generalized pareto distribution (GPD) and the threshold approach

This method considers more of the most extreme observations. This is one of the main advantage it possesses when compared to the BM approach. This technique is due to Pickands (1975). The Pickands–Balkema–de Haan Theorem (here denoted as Theorem 2) is stated below (see Embrechts et al. (2005) Theorem 7.20)

Theorem 2. Let $X_1, X_2, \ldots$ be a sequence of iid random variables with common distribution function F and let

$$M_n = \max\{X_1, X_2, \ldots, X_n\}$$

Let us denote an arbitrary term in X_i sequence by X and suppose that X satisfies Theorem 1 so that for large n

$$\mathbb{P}[M_n \leq z] \leq z \to G(z) \qquad as\ n \to \infty$$

For a non-degenerate distribution function G given as

$$G(z) = \exp\left\{-\left(1+\xi\left(\frac{z-\mu}{\sigma}\right)\right)_+^{\frac{-1}{\xi}}\right\}$$

for some $\mu, \sigma > 0$ and ξ, then for large enough u the distribution function of X conditional on $X > u$ can be approximated as

$$\mathbb{P}\left[X \leq y | X > u\right] \to H(y) \qquad as\ u \to \infty$$

where

$$H(y) = 1 - \left(1+\xi\frac{y-u}{\sigma_u}\right)_+^{\frac{-1}{\xi}} \qquad y > u \tag{2.3}$$

$H(y)$ is the GPD with the modified scale parameter $\sigma_u = \sigma + \xi(u-\mu)$ corresponding to the excess of the threshold u.

2.3. Mean excess plot

A very popular diagnostic tool which aids in the selection of u is the mean excess (ME) plot. The ME function of a random variable X is defined as

$$M(u) := E[X-u | X > u] \qquad 0 \leq u \leq x_F \tag{2.4}$$

given that $EX_+ < \infty$. This is also called the mean residual life function, especially in reliabilty theory field. Its formula, given that X is a positive random variable is

$$M(u) = \frac{\int_u^{x_F} \bar{F}(x)dx}{\bar{F}(u)} \qquad 0 < u < x_F$$

where x_F is the endpoint of the distribution function F and $\bar{F}$ is the survivor function.

If a distribution function is subexponential the mean excess function tends to infinity, if it is an exponential distribution the mean excess function is a constant and for the normal distribution the mean excess function tends to zero.

When working with the data, the empirical ME is used as an estimate of $M(u)$. Given an i.i.d sample $X_1, \ldots, X_n$ from $F(x)$, the empirical ME function is defined as

$$\hat{M}(u) = \frac{\sum_{i=1}^n \left(X_{(i)}\right) I_{\left[X_{(i)} > u\right]}}{\sum_{i=1}^n I_{\left[X_{(i)} > u\right]}} - u = \frac{\sum_{i=1}^n \left(X_{(i)} - u\right) I_{\left[X_{(i)} > u\right]}}{\sum_{i=1}^n I_{\left[X_{(i)} > u\right]}} \qquad u \geq 0 \tag{2.5}$$

$X_{(i)}$ represents the order statistics of the data.

The ME function of the random variable X, is linear in u in the case of the GPD. The formula is given below.

$$M(u) = \frac{\sigma}{1-\xi} + \frac{\xi}{1-\xi}u \tag{2.6}$$

where $0 \leq u < \infty$ if $0 \leq \xi < 1$ and $0 \leq u \leq \frac{-\sigma}{\xi}$ if $\xi < 0$.

Based on the linearity property of the ME function in the case of the GPD, Davison and Smith (1990) used this property to develop the MEP in which they plot the points $(X_{(i,n)}, \hat{M}(X_{(i,n)}))$. For a detailed discussion on MEP see Ghosh and Resnick (2010).

3. Basic concept of quantiles and order statistics

Given that the MEP basically requires that the data be ordered, it is important to understand the basic properties of order statistics as it relates to our proposed method. First, the exact distribution theory of order statistics for a finite sample is discussed, this is then followed by the asymptotic distribution theory.

Definition 3.1. Order statistics.

Let $X_1, X_2, ..., X_n$ be independent and identically distributed random variables. Let $X_{1:n} \leq X_{2:n} \leq ... \leq X_{n:n}$ which we simply write as $X_{(1)} \leq X_{(2)} \leq ... \leq X_{(n)}$ denote the ordered values of $X_1, X_2, ..., X_n$. Then we call $X_{(1)} \leq X_{(2)} \leq ... \leq X_{(n)}$ the order statistics of $X_1, X_2, ..., X_n$. The k^{th} order statistic $X_{(k)}$ is the k^{th} observation in order. That is, the k^{th} smallest of $X_1, X_2, ..., X_n$

The asymptotic theory for order statistics simply has to do with the distribution of $X_{r:n}$, suitably standardized, as $n \to \infty$. Here are the three distinguishing cases:

1. Central Order Statistics (Central Percentiles). This corresponds to $\sqrt{n}(\frac{k_n}{n} - p) \to 0$, for $0 < p < 1$
2. Intermediate Order Statistics. Here $n - k_n \to \infty$ and $\frac{k_n}{n} \to 1$
3. Extreme Order Statistics having $n - k_n$ fixed

k_n represents the k upper order statistics. The subscript n stresses the dependence of k on the sample size

Definition 3.2. Quantile.

A quantile is the value that corresponds to a specified proportion of a sample or population.

Let $Q(p)$ be the quantile function (QF). This function gives the quantile values for all probabilities p, $0 \leq p \leq 1$. That is,

$$Q(p) = F^{-1}(p) = inf\{x : F(x) \geq p\} \qquad 0 \leq p \leq 1 \tag{3.1}$$

F^{-1} represents the inverse of F. On the other hand, the tail quantile is defined as

$$Q\left(1 - \frac{1}{x}\right) = inf\left\{y : F(y) \geq 1 - \frac{1}{x}\right\} \quad x > 1$$

Gilchrist (2000) states: The definitions of the QF and the CDF can be written for any pair of values (x, p) as $x = Q(p)$ and $p = F(x)$. These functions are thus simple inverses of each other, provided that they are both continuous increasing functions.

$$\Rightarrow p = Q^{-1}(x) \text{ and } x = F^{-1}(p)$$

Hence, $Q(p) = F^{-1}(p)$ and $F(x) = Q^{-1}(x)$

Given the sample $X_1, X_2, \ldots, X_n$ of observations on F, we define the sample *pth* quantile as the *pth* quantile of the sample (empirical) distribution function F_n which is denoted as $\hat{Q}(p) = F_n^{-1}(p)$

Some very interesting properties of the quantile function are

1. The sum of two quantile functions is also a quantile function.
2. The product of two positive quantile functions is a quantile function.

If we denote $L(p)$ as the basic form of our quantile, then the generalized form of the quantile model will include the position (λ) and the scale (η) prameters which we can represent as

$$Q(p) = \lambda + \eta L(p) \tag{3.2}$$

Other variables that affect these parameters can also be included depending on the problem at hand.

Recall that the expected value of X is

$$E[X] = \int_{-\infty}^{\infty} xf(x)dx$$

Noting that $x = Q(p)$ and $dp = f(x)dx$, it implies that in terms of the quantile function

$$E[X] = \int_0^1 Q(p)dp \tag{3.3}$$

From the following property of expectation $E[a + bX] = a + bE[X]$

$$\Rightarrow \qquad E\big[\lambda + \eta L(p)\big] = \lambda + \eta E\big[L(p)\big]$$

if λ and η are constants.

3.1. Relationship between order statistics and quantiles

The order statistics of a sample which is equivalent to the sample distribution function F_n, plays a major role in modeling with quantile-defined distributions. We can therefore express the *pth* sample quantile $\hat{Q}(p)$ as

$$\hat{Q}(p) = \begin{cases} X_{(np)} & \textit{if } np \textit{ is an integer} \\ X_{(np+1)} & \textit{if } np \textit{ is not an integer} \end{cases} \tag{3.4}$$

3.2. Bootstrap technique

Bootstrapping is a statistical technique that is used for estimating quantities about a given population. This is done by averaging estimates from several small data samples

that are drawn from a larger dataset. From literature, it is a well known fact that full sample bootstrap does not work for the extremes (see Politis and Romano (1994), Wu (1990), and Götze and Künsch (1990)), the conventional remedy for this problem is to use an m-out-of-n bootstrap or subsampling. This involves sampling m times without (or with) replacement from a sample size of n such that $m < n$ and the condition $m \to \infty$ and $\frac{m}{n} \to 0$ as $n \to \infty$ holds.

3.2.1. Bootstrap aggregation (bagging)

This is an ensemble method introduced by Breiman (1996). It consists of combining multiple predictors in order to get an aggregated predictor. In most cases the procedure has been shown to reduce the variance of the predictor whilst keeping the magnitude of the bias roughly the same. The basic idea of bagging which serves as one of the main motivation for the construction of our proposed model is bootstrapping and averaging.

Bootstrap is not appropriate when the dataset is small. This is because the original sample is no longer a good approximation of the population. Bagging can be used to overcome this difficulty. To apply bagging, the dataset is first divided into a training set and a test set. Bootstrap samples are taken from the training set, these can be referred to as bags. A model is trained on each of these bags and tested using the test set. The final model is then obtained through majority voting or aggregation. The basic algorithm is outlined below.

Repeating B number of times:

- Get bootstrap samples L_k from L(the original training data). A rule of thumb is to use two-thirds of the original sample.
- Fit a model using L_k.

Then, combine the B models by voting (for classification problem) or averaging (for estimation problem).

3.3. The quantile bootstrap aggregation procedure

In this section we lay out the assumptions of our model. The procedure is explained in an algorithmic format and then, the model is represented mathematically.

Assumptions

1. The observations are independent and identically distributed (*iid*).
2. The distribution is in the maximum domain of attraction (MDA) of the generalized extreme value distribution.

The algorithm

This is a two step procedure.

First step: Inspecting the MEP.

1a. Specify a suitable threshold u_j by visually inspecting the MEP of the data (D) of size n obtained at the point where the plot becomes positively linear.

i. Check model fit to confirm that GPD is valid at u_j. A simple graphical plot can be used to check for good fit, for example the QQ plot.

ii. Then, obtain an approximate corresponding quantile value $\hat{Q}_j$ of u_j.

1b. Next, a sequence of quantile values are selected that satisfies the range $L = Q_{j-1}<Q_j<Q_{j+1}$ We denote L as the quantile threshold sample (QTS). The size of L is three. That is, $n(L) = 3$ with range $Q_j \pm 0.01$.

Remark 1. In step 1a, we note that different threshold candidates are present. The eye inspection approach (EIA) will aid the practioner in choosing the first threshold and then our proposed method is applied to refine the threshold choice by making use of an interval of thresholds which includes the first threshold choice. Q_j, the first threshold choice, serves as a reference point to obtain the end points of the interval L. Moreover, L contains an infinite number of fractions. This is based on the fact that between any two real numbers numbers there are infinitely many rational numbers.

Second step: Bootstrapping and Aggregating.
Repeat B_m times:

2a. Sample randomly from L. Having selected N_s random values, we obtain the mean $E[L] = \frac{1}{N_s}\sum_{i=1}^{N_s} L_i$. This step provides random quantile estimates for each bootstrap sample in 2 b(i)

2b. i. Obtain bootstrap samples from D of size $n_2<n$ without replacement.

ii. Based on the value generated in step 2a, compute the initial quantile. That is, the mean obtained in 2a is the quantile that is fitted on 2b(i).

iii. Compute the probability of exceeding Q_{j-1} and multiply the result by that obtained in 2b(ii). This gives the final quantile.

iv. Obtain a threshold at Q_{j-1} by summing the threshold evaluated at quantile Q_{j-1} (i.e. $u_{j-1}(Q_{j-1})$) and the result in 2b(iii).

2c. Combine the B_m thresholds (in 2 b(iv)) by averaging. This gives us the aggregated threshold value to which we fit the GPD.

Remark 2. By making use of subsamples of the given data through bootstrapping and averaging over the bootstrapped samples, we incorporate the underlying idea of the bagging approach which is bootstrapping and aggregation. Step 2b(iv) can be viewed as B_m-point thresholds (or simply, bags of point estimate thresholds), obtained using the steps 2a -2b(iii) as opposed to just choosing a threshold using the EIA. The relationship between order statistics and quantiles makes it possible to obtain $u_{j-1}(Q_{j-1})$

Mathematical representation
Our model incorporates the general form of the quantile model described in Section 3.1.

1. The position vector is set at $\hat{Q}_{j-1} = Q_0$. This is the lower end of the range. This is the deterministic part of our model which does not allow for chance or variability. The threshold corresponding to this point is denoted by $u_0(Q_0)$.
2. The fixed quantile threshold sample is L. The *pth* quantile of each bootstrap subsample B_i will be fitted at point $E(L_i)$ the expected value of L which we denote as $\hat{Q}_{[E(L_i)]}$. This serves as the basic quantile form (the initial quantile).

3. The probability that X is greater than the lower end of our fixed range is taken into account since different subsamples from the given original data will be used. This condition allows us stay within the range of the tail and it allows for the influence of uncertainty. It is estimated using the formula

$$\mathbb{P}(X>Q_0) = \frac{n_u}{n}$$

where n_u is the number of variables exceeding Q_0 and n represents the size of sample.

Thus, our proposed threshold quantile model function is

$$\hat{U}_i = u_0(Q_0) + \mathbb{P}(X>Q_0)\hat{Q}_{[E(L_i)]}(B_i) \tag{3.5}$$

The empirical model is

$$\hat{U}_i^* = u_0(Q_0) + \frac{n_{u_i}}{n_i}\left\{Q_{\left(\frac{1}{N_s}\sum_{k=1}^{N_s} Q_k\right)}(B_i)\right\} \qquad Q_0<Q_k<Q_{j+1} \tag{3.6}$$

The optimal threshold quantile estimate is therefore computed as

$$U_{agg} = \frac{1}{B_m}\sum_{i=1}^{B_m} U_i^* \tag{3.7}$$

$\hat{U}_{agg}$ is the aggregate threshold, N_s is the sample size when sampling from L, the quantile threshold sample. With respect to the *ith* bootstrap sample from the original data, $\frac{n_{u_i}}{n_i}$ represents the probability of exceedence, $Q_{i(.)}$ is the quantile fitted at point $(.)$ and B_m is the number of bootstrap samples from the original data (D).

4. Simulation study

The three distributions used are the Skewed generalized t (simply called Skewed t), Pareto and Lognormal distributions. These are heavy-tailed distributions. For the Skewed t distribution we show results for simulations with sample sizes 10000 and 1000 while for the other two distributions we only display results for $n = 1000$.

4.1. Skewed generalized t distribution

Miljković and Radović (2006) have shown that asset returns are leptokurtic and skewed. To model asset returns we generate data from the Skewed generalized t distribution. A Skewed t distribution with between 3-5 degrees of freedom (df) is a fat-tailed distribution. It includes distributions that have both heavy tails and skewness (Arslan and Genç, 2009). Theodossiou (1998) showed that it is suited to fit asset returns in general. The MEP of the simulated data (for $n = 10000$) is displayed in Figure 1.

By the EIA a threshold u of 1.2 is quite appropraite as seen by the dotted vertical line and the graphs for the goodness of fit shows that the GPD is valid (Figure 2).

The approximate quantile to $u = 1.2$ is the 89th quantile. Next we choose the quantile threshold sample (QTS) which in this case will be the sequence (0.88,0.89,0.90). We set n_1 to 100, n_2 to $500(5\%\ of\ 10000)$ and perform steps already described in the algorithm in Section 3.3. $u_0(Q_0)$ in this simulation is 1.16 corresponding to the 88th

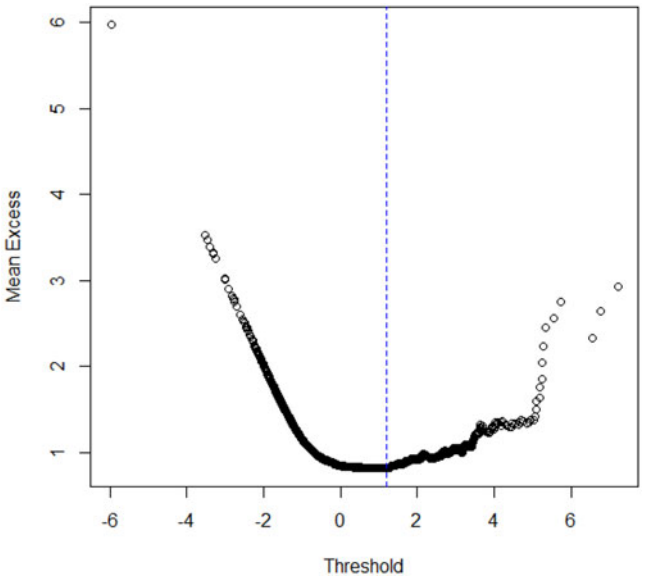

Figure 1. Mean excess plot fitted to the simulated Skewed t distribution.

quantile. This gives an aggregate threshold of $u_{agg} = 1.2482$. The quantile and probability plots in Figure 3 also confirm that it is a good fit.

A head-view of the aggregate bootstrap subsamples along with their respective thresholds for the Skewed t distribution is shown in Table 1. The choice of m is 5% of n. The fitted GPD parameters (estimated using MLE) are displayed in Table 2 with their respective standard errors enclosed in braces.

The EIA threshold is still $u = 1.2$ for $n = 1000$ and we get an aggregate threshold of 1.17 resulting in scale and shape parameters 0.72804576 (0.1033164) and 0.08106162 (0.1079181) respectively. Fitting the GPD based on the EIA gives 0.75102348 (0.1079566) for scale and 0.06615288 (0.1081323) for shape. We observe that in comparison with the EIA, the standard errors (in braces) are a little lower for the aggregate threshold with sample size 1000 than it is when the sample size is 10000.

4.2. Pareto and lognormal distributions

Similar analysis is carried out on the simulated datasets for the Pareto and Lognormal distributions (Figure 4).

In addition to the quantile threshold sample having width ±0.01, a wider threshold sample of width ±0.05, that is (0.45,0.50,0.55), is also tested for the Lognormal sample whose EIA is set at 7.5. The two vertical lines in Figure 4a correspond to the wider width, 6.9 and 8.1 being the respective thresholds at quantiles 0.45 and 0.55. The bootstrap samples are taken from within this range. It is noted that in this case, a lower aggregate threshold of 7.2779 is gotten which is much closer to the lower 6.9 threshold than the upper one. This can increase the uncertainty involved in the process hence, the need to use smaller quantile threshold sample widths. Table 3 displays the results for both the Pareto and Lognormal simulated distributions with sample size $n = 1000$ and m = 5%. The estimated GPD parameters are also provided. In order to ascertain the

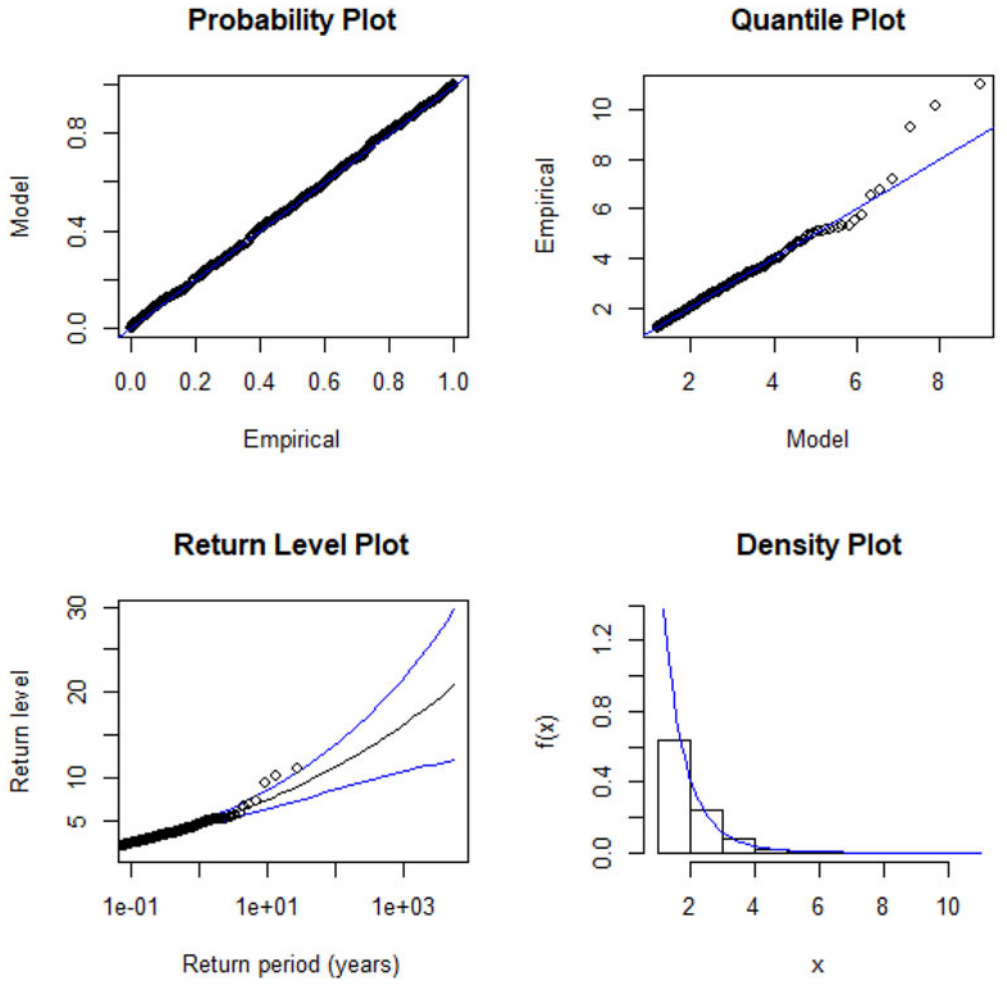

Figure 2. Diagnostic plots for single threshold.

performance of each fit, the Negative log likelihood (NLL) was applied. Also, the Akaike Information Criteria (AIC) and Bayesian Information Criteria (BIC) values were obtained for the three simulated datasets (Table 4).

Decision rule: The threshold having the lowest values for NLL, AIC and BIC gives the optimal threshold.

4.3. Discussion of findings

Although the standard errors for the aggregated and the single thresholds are roughly the same, the statistical tests (NLL, AIC and BIC values) for the aggregate threshold, u_{agg}, are noticeably lower than that of the single threshold for the Lognormal and Pareto cases. It is however a little higher for skewed t distribution. From Table 1 we note that the procedure makes it possible for every point within the selected interval to be sampled equally. It is noted that the subsample size does affect the threshold value obtained and consequently the goodness of fit tests. Generally, an increase in m leads to

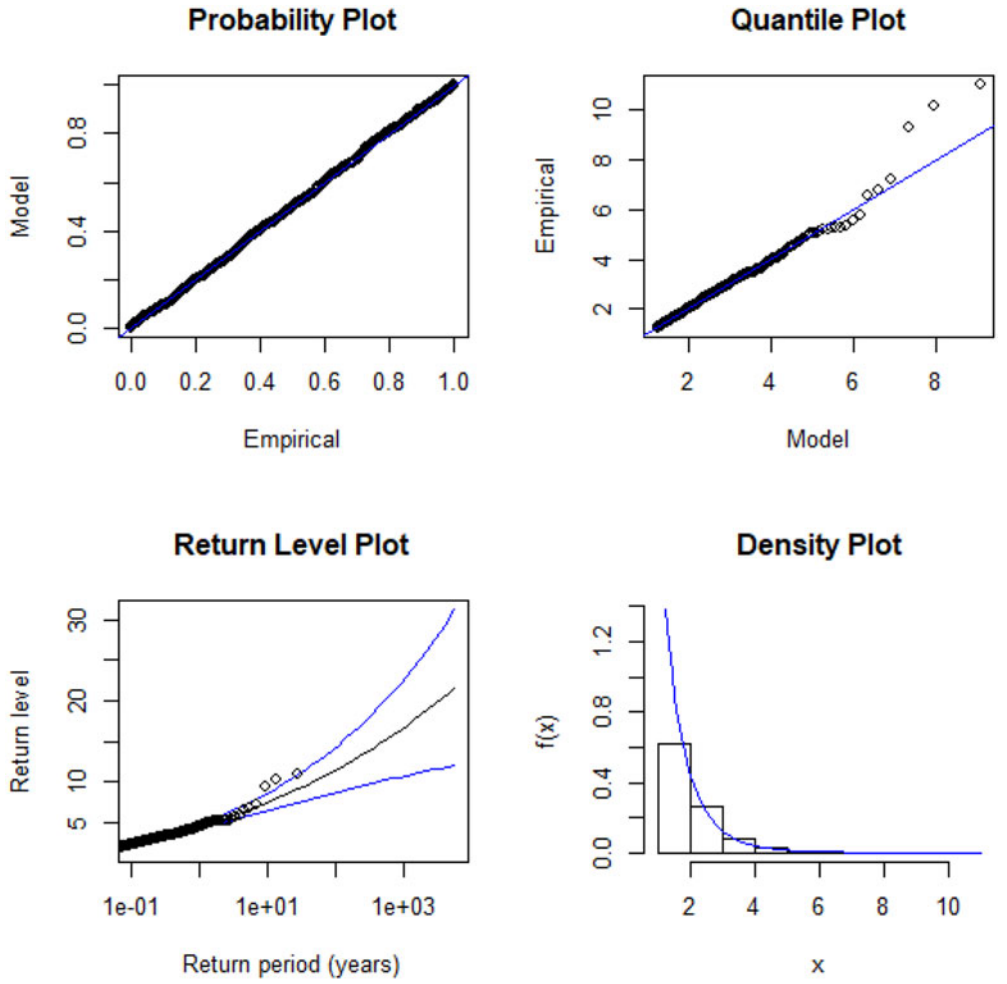

Figure 3. Diagnostic plots for aggregate threshold.

Table 1. View of aggregated bootstrap subsamples.

89.04%	88.98%	88.93%	89.02%	88.95%
1.248015	1.225361	1.262530	1.231192	1.234448
89.08%	88.91%	88.98%	88.98%	88.94%
1.238739	1.263966	1.261112	1.237068	1.222176

Table 2. Fitting the GPD (Skewed *t*).

	Single threshold (u_{sing})	Aggregate threshold (u_{agg})
threshold	1.2	1.2482
scale(σ)	0.7277(0.0334)	0.7234(0.0346)
shape(ξ)	0.1158(0.0340)	0.1228(0.0357)

a slight increase in the SE of the parameters but exhibiting lower NLL, AIC and BIC values. Thus the practitioner's discretion needs to be applied in this case in order to obtain a reasonable tradeoff between the standard errors and the other goodness of fit tests.

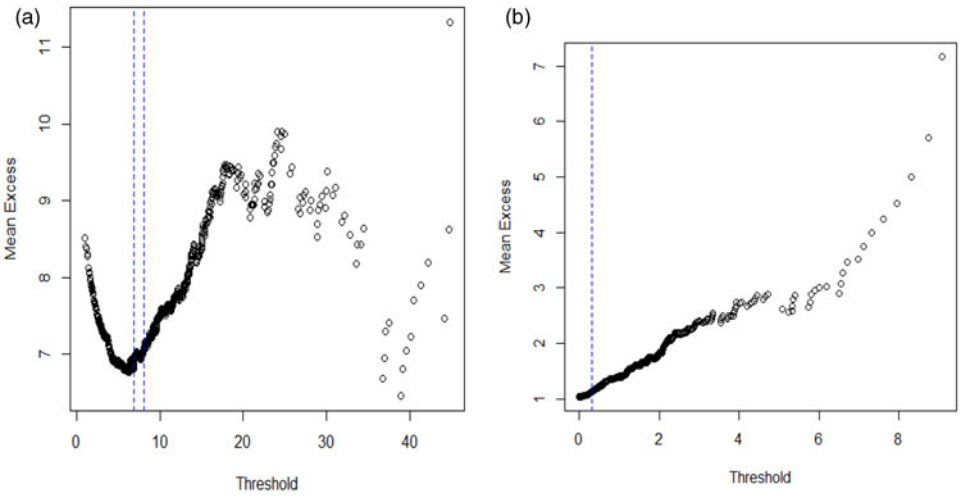

Figure 4. Mean excess plot of simulated data. lognormal (a), Pareto (b).

Table 3. Fitting the GPD (Lognormal and Pareto).

u_{sing}	Lognormal	Pareto
u	7.5	0.3
σ	5.9753(0.4187)	0.8175(0.0489)
ξ	0.1442(0.0541)	0.2590(0.0473)
u_{agg}		
u	7.6282	0.3094
σ	5.8538 (0.4184)	0.7785 (0.0486)
ξ	0.1571 (0.0559)	0.3115(0.0509)

Table 4. NLL, AIC and BIC values.

	u_{sing}	u_{agg}
Skewed t		
NLL	86.55253	89.0668
AIC	177.1051	182.1336
BIC	182.5241	187.675
Lognormal		
NLL	1454.187	1432.842
AIC	2912.375	2869.683
BIC	2920.788	2878.072
Pareto		
NLL	727.5075	721.0393
AIC	1459.015	1446.079
BIC	1468.083	1455.126

4.4. Comparing the result obtained with other threshold methods

The R package tea, by Ossberger (2017), describes different threshold estimation approaches that have been developed by various authors along with their specific details. Table 5 gives the thresholds (u) and their corresponding shape parameters (ξ) for sample size $n = 1000$. Only point estimates are provided. To have a fair base for comparison, similar thresholds to other existing methods are chosen for our model in order to align its result with the results obtained from other methods. Thresholds much higher

Table 5. Results from other threshold methods.

		Skewed t		Lognormal		Pareto	
	Technique	u	ξ	u	ξ	u	ξ
1	Eyeball	2.5	0.17	8.5	0.14	1.30	0.26
2	Proposed model	2.29	0.25	8.76	0.13	1.32	0.26
3	Gerstengarbe plot*	2.48	0.26	–	–	1.33	0.61
4	Developed by Danielsson et al. (2016)	2.78	0.33	9.18	0.50	1.40	0.60
5	Developed by Caeiro and Gomes (2016)	1.27	0.43	8.53	0.51	1.75	0.50
6	Developed by Guillou and Hall (2001)	2.03	0.32	–	–	2.02	0.49

*Developed by Gerstengarbe and Werner (1989).

Table 6. Parameters for different thresholds.

Threshold	shape	scale
1.1	0.3752	0.5234
1.2	0.3892	0.5418
1.3	0.4337	0.5248

than the needed range are not reflected on the table (i.e methods 3 and 6 for Lognormal). It is important to note that for each approach (methods 3 to 6), only the tail index is generated alongside the threshold hence, to obtain the shape parameter we use the relation

$$tail\ \ index = \frac{1}{shape\ \ parameter}$$

Thresholds from the range 1 to 3 for Skewed t, 8.5 to 9.5 for Logn and 1 to 2.5 for the Pareto distribution are chosen. Having close enough thresholds will enable us obtain comparable results for ξ. Close estimates of ξ are realized in the case of the Skewed t distribution. For the Lognormal and Pareto distributions, the values differ slightly. These differences however, are not significant. In general, the proposed method which we can term as the quantile bootstrap aggregation technique, performs reasonably well when compared to other methods especially the eyeball approach.

5. An illustration using the Nigerian stock exchange all share index (NSE ASI) (2005-2015)

In practice, the three thresholds on Table 6 represent thresholds that can be chosen by different practitioners. The data made use of here is the NSE ASI stock index obtained from the Nigerian Stock Exchange website www.nse.com.ng. The EIA threshold is 1.2 whose shape and scale parameters are 0.3892 (0.0686) and 0.5418 (0.04701) respectively. The corresponding approximate quantile is the 89th quantile (1.214075) and the QTS becomes (0.88,0.89,0.90). Applying the proposed model (with m = 5%) to the 2700 daily standardized residual returns of the NSE ASI (left tail) in order to obtain an aggregate threshold between 1.1 and 1.3, we get u_{agg}= 1.1783 (note that this is different from taking a simple arithmetic average of the three thresholds in Table 6). The GPD parameters set at u_{agg} result in $\xi = 0.3778671$ (0.06650394) and $\sigma = 0.5499974$ (0.04666013) which have slightly lower standard errors when compared to the EIA results.

6. Conclusion

In this paper we developed a quantile-based model that can serve as an alternative to selecting the threshold when the MEP is used. Simulated datasets and a real dataset were employed to test the model. In comparison to simply inspecting the MEP with the eye to obtain a single threshold, the use of the proposed method shows that a more optimal threshold can be estimated when we consider an aggregate of the thresholds.

Future research may take into account a parametric or semi-parametic approach in which the GPD distribution will be used.

Funding

Pan African University Institute for Basic Sciences, Technology and Innovation, Kenya.

References

Arslan, O., and A. İ. Genç. 2009. The skew generalized t distribution as the scale mixture of a skew exponential power distribution and its applications in robust estimation. *Statistics* 43 (5): 481–98. Taylor and Francis. doi:10.1080/02331880802401241.

Bermudez, P., M. A. Turkman, and K. F. Turkman. 2001. A predictive approach to tail probability estimation. *Extremes* 4:295–314. doi:10.1023/A:1016546027962.

Breiman, L. 1996. Bagging Predictors. *Machine Learning* 24 (2):123–40. doi:10.1007/BF00058655.

Caeiro, J., and M. I. Gomes. 2016. Threshold selection in Extreme Value Analysis. *Extreme Value Modeling and Risk Analysis: Methods and Applications* 1:69–86. doi:10.1201/b19721.

Coles, S. G. 2001. *An introduction to statistical modelling of extreme values*. London: Springer Series in Statistics.

Danielsson, J., L. M. Ergun, L. De Haan, and C. G. De Vries. 2016. *Tail index estimation: Quantile driven threshold selection*. Working paper, Systemic Risk Centre: The London School of Economics and Political Science.

Davison, A. C., and R. L. Smith. 1990. Models for exceedances over high thresholds. *Journal of the Royal Statistical Society. Series B, Statistical Methodology* 52:393–442. doi:10.1111/j.2517-6161.1990.tb01796.x.

Dupuis, D. J. 1999. Exceedances over high thresholds: A guide to threshold selection. *Extremes* 1 (3):251–61.

Embrechts, P., A. J. McNeil, and R. Frey. 2005. *Quantitative risk management: Concepts, techniques, and tools*. Princeton: Princeton University Press, pp. 277.

Fisher, R., and L. Tippett. 1928. Limiting forms of the frequency distribution of the largest or smallest member of a sample. *Mathematical Proceedings of the Cambridge Philosophical Society* 24 (2):180–90. doi:10.1017/S0305004100015681.

Frigessi, A., O. Haug, and H. Rue. 2002. A dynamic mixture model for unsupervised tail estimation without threshold selection. *Extremes* 5 (3):219–35.

Gerstengarbe, F. W., and P. C. Werner. 1989. A method for statistical definition of extreme value regions and their application to meteorological time series. *Zeitschrift Fuer Meteorologie* 39 (4): 224–6.

Ghosh, S., and S. Resnick. 2010. A discussion on Eean excess plots. *Stochastic Processes and Their Applications* 120 (8):1492–517. doi:10.1016/j.spa.2010.04.002.

Gilchrist, W. 2000. *Statistical modelling with quantile functions*. New York: Chapman and Hall/CRC Services.

Götze, F., and H. R. Künsch. 1990. *Blockwise bootstrap for dependent observations: higher order approximations for studentized statistics (abstract CP-401)*. IMS bulletin 19 443.

Guillou, A., and P. Hall. 2001. A diagnostic for selecting the threshold in extreme value analysis. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 63 (2):293–305. doi:10.1111/1467-9868.00286.

Lang, M., T. Ouarda, and B. Bobee. 1999. Towards operational guidelines for over-threshold modeling. *Journal of Hydrology* 225 (3–4):103–17. doi:10.1016/S0022-1694(99)00167-5.

Miljković, V., and O. Radović. 2006. Stylized facts of asset returns: Case of BELEX. *Facta Universitatis: Economics and Organization* 3 (2):189–201.

Ossberger, J. 2017. TEA: Threshold Estimation Approaches. R package version 1.0.

Pickands, J. 1975. Statistical inference using extreme order statistics. *Annals of Statistics* 3 (1): 119–31.

Politis, D. N., and J. P. Romano. 1994. A general theory for large sample confidence regions based on subsamples under minimal assumptions. *The Annals of Statistics* 22 (4):2031–50. doi:10.1214/aos/1176325770.

Scarrott, C., and A. MacDonald. 2012. A review of extreme value threshold estimation and uncertainty quantification. *REVSTAT-Statistical Journal* 10 (1):33–60.

Tancredi, A., C. Anderson, and A. O'Hagan. 2006. Accounting for threshold uncertainty in extreme value estimation. *Extremes* 9 (2):87–106. doi:10.1007/s10687-006-0009-8.

Theodossiou, P. 1998. Financial data and the skewed generalized t distribution. *Management Science* 44 (12-part-1):1650–61. doi:10.1287/mnsc.44.12.1650.

Thompson, P., Y. Cai, D. Reeve, and J. Stander. 2009. Automated threshold selection methods for extreme wave analysis. *Coastal Engineering* 56 (10):1013–21. doi:10.1016/j.coastaleng.2009.06.003.

Wu, C. F. J. 1990. On the asymptotic properties of the Jackknife histogram. *The Annals of Statistics* 18 (3):1438–52. doi:10.1214/aos/1176347759.

Zoglat, A., S. El Adlouni, F. Badaoui, A. Amar, and C. G. Okou. 2014. Managing hydrological risks with extreme modeling: Application of peaks over threshold model to the Loukkos Watershed, Morocco. *Journal of Hydrologic Engineering* 19 (9):05014010.