Optimal Threshold Determination for Securities Exchange Volumes Using Improved Maximum Product of Spacing Methodology

Peter Murage[1], Joseph Mung'atu[2], Everlyne Odero[1]

[1]Department of Mathematics, Masinde Muliro University of Science and Technology, Kakamega, Kenya
[2]Department of Statistics and Actuarial Science, Jomo Kenyatta University of Agriculture and Technology, Nairobi, Kenya
Email: gitmurage@gmail.com, j.mungatu@fsc.jkuat.ac.ke, eodero@mmust.ac.ke

Abstract

To Statisticians, the structure of the extreme levels which exist in the tails of the ordinary distributions is very important in analyzing, predicting and forecasting the likelihood of an occurrence of extreme event. Extreme events are defined as values of the event below or above a certain value called threshold. A well chosen threshold helps to identify the extreme levels. Several methods have been used to determine threshold so as to analyze and model extreme events. One of the most successful methods is the maximum product of spacing (MPS). However, there is a problem encountered while modeling data through this method in that the method breaks down when there is a tie in the exceedances. This study offers a solution to model data even when it contains ties. In the study, a method that improved MPS method for determining an optimal threshold for extreme values in a data set containing ties was derived. The Generalized Pareto Distribution (GPD) parameters for the optimal threshold were derived and compared to GPD parameters determined through the standard MPS model. The study improved the standard MPS methodology by introducing the concept of frequency and used Generalized Pareto Distribution (GPD) and Peak over threshold (POT) methods as the basis of identifying extreme values. The improved MPS models and the standard models were applied to Nairobi Securities Exchange (NSE) trading volume data to determine the GPD parameters for different sectors registered in NSE market and their performance compared. It was realized that the improved MPS model performed better than the standard models. This study will help the Statisticians in different sectors of our economy to model extreme events involving ties.

1. Introduction

Certain values in the tails of any distribution, represent extreme events and they are pointers to eventuality. The values in the tails are rare, few, but can have a big impact on the conclusion arrived at by the analysts. Different sectors of our life experience Extreme events and here we mention just but a few. According to [1] and [2] Extreme low production in agriculture results to famine if the agriculture depends on rainfall. This means that the amount of rain experienced in that region was too low that crops dried up [3] or very high rainfall that it destroyed all crops that had been planted. [4], studying extreme rainfall in a mountainous region and [5] studying extreme rainfall in West Africa did observe that, how low or high the amount of rainfall depends on the threshold attached to the rainfall in that region. In insurance industries [6], while discussing tools in finance and insurance, noted that extreme high claims by the customers that can be very dangerous for the company while extreme low claims by the customers can be very beneficial for the company's profit. This means that there is a critical level that the insurance company would wish it is not surpassed and if it is, according to [7], it must be prepared for this eventuality. Very high emissions of the waste products from the manufacturing industries are harmful to the environment and ozone layer. However, countries must continue to industrialize or expand their industries for economic prosperity. Extreme value theory (EVT) is a tool which attempts to provide us with the best possible estimate of the tail area of the distribution [8]. While working on the importance of the tail dependence in Bivariate frequency analysis, noted that there are two principal kinds of model for extreme values. The oldest group of models is the block maxima models. According to [9] and [10] the block maxima/minima methods are fitted with the generalized extreme value (GEV) distribution. A more modern group of models is the peaks-over-threshold (POT) models; these are models for all large observations which exceed a high threshold. According to [11], theory of extreme value, the block of maxima of a sequence of identically and independently distributed (iid) random variables in the limits follows a generalized extreme value (GEV) distribution. At the same time, [10] showed that the excesses over a high threshold for these random variables, followed a Generalized Pareto distribution (GPD) using peak over threshold (POT) method. [12] proposed that choosing an optimal threshold is similar to choosing the number of upper order statistics and that a compromise between bias and variance has to be reached. The most successful method of determining the threshold of a GPD is the Maximum Product of Spacing (MPS). The Maximum Product of Spacing

(MPS) was introduced originally by [13] as one method of determining threshold. This general method is based on spacings (that is, the gaps between successive order statistics). A threshold approach for peaks over threshold using MPS was carried out by [14] and noted that the selection of a threshold is an important and challenging problem [15]. While studying traditional estimation methods and MPS in Generalized Inverted Exponential Distribution found out that MPS outperformed MLE and least square (LSE) methods on the basis of K-S distance and Akaike Information Criterion (AIC). While trying to compare the methods for parameter estimation for univariate continuous models [16], suggested that MPS is useful in estimating parameters for univariate continuous models with a shift at the origin. He also noted that MPS method would be an alternative method when MLE method encounters numerical difficulties in some parametric models [17]. While comparing parameter estimation for Generalized Power Weibull (GPW) proposed the use of MPS as compared to MLE and Bayesian to estimate the parameters of the GPW. However, the MPS method is sensitive to ties [18]. This study improved the MPS method so that it is able to handle any data even if it contains ties.

Maximum Product of Spacing (MPS) Methodology

According to [13], maximum spacing estimators are sensitive to closely spaced observations, and especially ties. In cases of ties, some scholars have suggested that one value of each tie is taken [18]. Let $x_1, x_2, \cdots, x_n$ be a random sample of independent observations from a continuous distribution F_{θ_0} belonging to $F_\theta, \theta \in \Theta$. Applying the probability transform $F_\theta(.)$ to the order statistics $x_{1,n} \le x_{2,n} \le \cdots \le x_{n,n}$ yields $0 \equiv F_\theta(x_{0,n}) \le F_\theta(x_{1,n}) \le \cdots \le F_\theta(x_{n+1,n}) \equiv 1$. We define the spacings as the gaps between the values of the distribution function at adjacent ordered points

$$D_i(\theta) = F_\theta(x_i) - F_\theta(x_{i-1}) \tag{1}$$

for $i = 1, 2, \cdots, n+1$. The maximum spacing estimator θ_0 was defined as value that maximizes the logarithm of the geometric mean of sample spacings.

$$\hat{\theta} = \arg\max_{\theta \in \Theta} S_n(\theta) \tag{2}$$

where

$$
\begin{aligned}
S_n(\theta) &= \ln \sqrt[n+1]{\left(D_1(\theta) \cdot D_2(\theta) \cdots D_{n+1}(\theta)\right)} \\
&= \frac{1}{n+1} \sum_{i=1}^{n+1} \ln D_i(\theta)
\end{aligned}
\tag{3}
$$

This maximum spacing estimator is sensitive to the ties. That is, for any

$$x_{i+m} = x_{i+m-1} = \cdots = x_i$$

Then

$$D_{i+m}(\theta) = D_{i+m-1}(\theta) = \cdots = D_i(\theta)$$

This, therefore, collapses the standard MPS method proposed by [13].

2. Methodology

2.1. Improved MPS Methodology

Let $x_1, x_2, \cdots, x_n$ be a random sample of independent observations from a continuous distribution F_{θ_0} belonging to $F_\theta, \theta \in \Theta$ [19]. Applying the probability transform $F_\theta(.)$ to the order statistics $x_{1,n} \le x_{2,n} \le \cdots \le x_{n,n}$ yields $0 \equiv F_\theta(x_{0,n}) \le F_\theta(x_{1,n}) \le \cdots \le F_\theta(x_{n+1,n}) \equiv 1$. We define the spacings as the gaps between the values of the distribution function at adjacent ordered points

$$D_i(\theta) = F_\theta(x_i) - F_\theta(x_{i-1}) \tag{4}$$

for $i = 1, 2, \cdots, n+1$. The maximum spacing estimator θ_0 was defined as value that maximizes the logarithm of the geometric mean of sample spacings.

$$\hat{\theta} = \arg\max_{\theta \in \Theta} S_n(\theta) \tag{5}$$

where

$$S_n(\theta) = \ln \sqrt[n+1]{\left(D_1(\theta) \cdot D_2(\theta) \cdots D_{n+1}(\theta)\right)}$$

The modified MPS method proposed here is to use grouped data frequency table. Let $x_1, x_2, \cdots, x_n$ occur $f_1, f_2, \cdots, f_n$ times respectively. The geometric mean is given by

$$G = \left(x_1^{f_1} \cdot x_2^{f_2} \cdots x_n^{f_n}\right)^{\frac{1}{N}} = \left[\prod_{i=1}^{n} x_i^{f_i}\right]^{\frac{1}{N}}$$

This implies that

$$\ln G = \frac{1}{N} \sum_{i=1}^{n} f_i \ln x_i \tag{6}$$

This leads to the modified MPS method as

$$S_n(\theta) = \ln \sqrt[n+1]{\left(D_1^{f_1}(\theta) \cdot D_2^{f_2}(\theta) \cdots D_{n+1}^{f_{n+1}}(\theta)\right)} = \frac{1}{n+1} \sum_{i=1}^{n+1} f_i \ln D_i(\theta) \tag{7}$$

In case $f_1 = f_2 = \cdots = f_{n+1} = 1$, then we go back to the standard MPS. The Spacings are such that $\sum_{i=1}^{n} D_i(\theta) = 1$. Under MPS, $D_i(\theta)$'s are defined as:

$$D_1(\theta) = F(x_{1:n}, \theta)$$

$$D_i(\theta) = F(x_{i:n}, \theta) - F(x_{i-1:n}, \theta)$$

$$D_{n+1}(\theta) = 1 - F(x_{n:n}, \theta)$$

Therefore, Equation (7) can be partitioned as:

$$S_n(x_i; \theta, \varepsilon, \sigma) = \frac{1}{n+1} \left\{ f_1 \ln D_1(\theta) + \sum_{i=2}^{n} f_i \ln D_i(\theta) + f_{n+1} \ln D_{n+1}(\theta) \right\} \tag{8}$$

2.2. Estimation of Generalized Pareto Distribution Using the Modified MPS Method

To estimate the parameters, we substitute the GPD

$$G(x;\varepsilon,\sigma) = \begin{cases} 1 - \left[1 + \varepsilon\left(\dfrac{x-u}{\sigma}\right)\right]^{-\frac{1}{\varepsilon}}, & \varepsilon \neq 0 \\[4mm] 1 - \exp\left[-\left(\dfrac{x-u}{\sigma}\right)\right], & \varepsilon = 0 \end{cases} \tag{9}$$

into the MPS method Equation (8)

2.2.1. Case 1: When $\varepsilon \neq 0$

Let

$$D_1 = 1 - \left[1 + \varepsilon\left(\frac{x_1 - \theta}{\sigma}\right)\right]^{-\frac{1}{\varepsilon}} \tag{10}$$

$$D_i = \left(1 - \left[1 + \varepsilon\left(\frac{x_i - \theta}{\sigma}\right)\right]^{-\frac{1}{\varepsilon}}\right) - \left(1 - \left[1 + \varepsilon\left(\frac{x_{i-1} - \theta}{\sigma}\right)\right]^{-\frac{1}{\varepsilon}}\right)$$

which leads to

$$D_i = \left[1 + \varepsilon\left(\frac{x_{i-1} - \theta}{\sigma}\right)\right]^{-\frac{1}{\varepsilon}} - \left[1 + \varepsilon\left(\frac{x_i - \theta}{\sigma}\right)\right]^{-\frac{1}{\varepsilon}} \tag{11}$$

and

$$D_{n+1} = 1 - \left(1 - \left[1 + \varepsilon\left(\frac{x_n - \theta}{\sigma}\right)\right]^{-\frac{1}{\varepsilon}}\right) \tag{12}$$

implying that

$$D_{n+1} = \left[1 + \varepsilon\left(\frac{x_n - \theta}{\sigma}\right)\right]^{-\frac{1}{\varepsilon}}$$

and

$$K_1 = \ln\left(1 - \left[1 + \varepsilon\left(\frac{x_1 - \theta}{\sigma}\right)\right]^{-\frac{1}{\varepsilon}}\right) \tag{13}$$

$$K_2 = \ln\left\{\left[1 + \varepsilon\left(\frac{x_{i-1} - \theta}{\sigma}\right)\right]^{-\frac{1}{\varepsilon}} - \left[1 + \varepsilon\left(\frac{x_i - \theta}{\sigma}\right)\right]^{-\frac{1}{\varepsilon}}\right\} \tag{14}$$

$$K_3 = \ln\left[1 + \varepsilon\left(\frac{x_n - \theta}{\sigma}\right)\right]^{-\frac{1}{\varepsilon}} \tag{15}$$

To optimize the function $S_n(\theta,\varepsilon,\sigma)$, we partially differentiate it with respect to θ,ε and σ to obtain;

$$S'_\varepsilon = \frac{1}{n+1}\left\{f_1\frac{\partial K_1}{\partial\varepsilon} + \sum_{i=2}^{n}f_i\frac{\partial K_2}{\partial\varepsilon} + f_{n+1}\frac{\partial K_3}{\partial\varepsilon}\right\} = 0 \tag{16}$$

$$S'_\sigma = \frac{1}{n+1}\left\{f_1\frac{\partial K_1}{\partial\sigma} + \sum_{i=2}^{n}f_i\frac{\partial K_2}{\partial\sigma} + f_{n+1}\frac{\partial K_3}{\partial\sigma}\right\} = 0 \tag{17}$$

$$S'_\theta = \frac{1}{n+1}\left\{ f_1 \frac{\partial K_1}{\partial \theta} + \sum_{i=2}^{n} f_i \frac{\partial K_2}{\partial \theta} + f_{n+1} \frac{\partial K_3}{\partial \theta} \right\} = 0 \tag{18}$$

2.2.2. Case 2: When $\varepsilon = 0$

Let

$$D_1 = 1 - \exp\left[-\left(\frac{x_1 - \theta}{\sigma} \right) \right] \tag{19}$$

$$D_i = \exp\left[-\left(\frac{x_{i-1} - \theta}{\sigma} \right) \right] - \exp\left[-\left(\frac{x_i - \theta}{\sigma} \right) \right] \tag{20}$$

$$D_{n+1} = \exp\left[-\left(\frac{x_n - \theta}{\sigma} \right) \right] \tag{21}$$

And

$$K_1^* = \ln D_1 \tag{22}$$

$$K_2^* = \ln P \tag{23}$$

where t $P = \exp\left[-\left(\frac{x_{i-1} - \theta}{\sigma} \right) \right] - \exp\left[-\left(\frac{x_i - \theta}{\sigma} \right) \right]$

$$K_3^* = \frac{(x_n - \theta)}{\sigma} \tag{24}$$

To optimize the function $S_n(\theta, \varepsilon, \sigma)$, we partially differentiate it with respect to θ, ε and σ to obtain;

$$S'_{\sigma^*} = \frac{1}{n+1}\left\{ f_1 \frac{\partial K_1^*}{\partial \sigma} + \sum_{i=2}^{n} f_i \frac{\partial K_2^*}{\partial \sigma} + f_{n+1} \frac{\partial K_3^*}{\partial \sigma} \right\} = 0 \tag{25}$$

$$S'_{\theta^*} = \frac{1}{n+1}\left\{ f_1 \frac{\partial K_1^*}{\partial \theta} + \sum_{i=2}^{n} f_i \frac{\partial K_2^*}{\partial \theta} + f_{n+1} \frac{\partial K_3^*}{\partial \theta} \right\} = 0 \tag{26}$$

3. Market Data

The models 16, 17, 18, 25 and 26 were coded in r-software and used to analyze the market volume data. Market data was sought from one company from each of the twelve (12) sectors trading in the Nairobi Securities Exchange (NSE), namely Agricultural (Sasini), Automobile and Accessories (Sameer Group), Banking (KCB), Commercial and Services (Kenya Airways), Construction and Allied (East African Cables), Energy and Petroleum (Kenol Kobil), Insurance (Kenya Re), Investment (Centum company), Investment Services (NSE), Manufacturing and Allied (East African Breweries), Telecommunication (Safaricom) and Real Estate Investment (Stanlib Fahari I-Reit). This was daily trading data for a period of three years (2016 to the end of 2018). Real market data contains ties and the first part of the analysis in this study was to check on this fact. This was done in Excel using pivotal tables by analyzing the volumes traded each day for the three years. The number of observations for volume traded in each company analyzed and the corresponding number of repetitions (ties) summarized as indicated in **Table 1**.

Table 1. Company data and number of repetitions.

Company	Number of Observations	Number of Repetitions
Centum	592	75
East African Breweries	699	131
East African Cables	711	378
Kenya Airways	634	83
Kenya Commercial Bank	739	21
KenolKobil	635	62
Kenya Reinsurance	729	182
Nairobi Securities Ltd	591	183
Safaricom	592	6
Sasini	711	450
Sameer	395	203
Stanlib	431	185
Claims Data	237	119

From Table 1, it was observed that all companies had some repetitions. These repetitions are synonymous to ties. Company's such as Sasini, Sameer, Stanlib, Kenya-Re and East African Cables had fairly many repetitions (ties). The densities of the data distribution in different sectors, were plotted as shown in Figure 1. This was done to assess the distribution of the volume data and assess whether the data had extreme values.

The y-axis of Figure 1 represent the density of the volume data while the x-axis represent the volume of the traded securities. The densities of all the companies are right skewed and indicate the tendencies of having extreme values in the right tail. These densities are very similar to the gamma density. Gamma density is right skewed and therefore contains extreme values in the right tail. All the sectors were therefore observed to contain extreme values in their right tail which served as a justification to subject our data to extreme value analysis using GPD and POT. The data were then subjected to both the standard MPS model and the improved models 16, 17, 18, 25 and 26, to determine the GPD parameters. Back testing techniques was used to assess the efficiency and consistency of these parameters and Akaike Information Criterion (AIC) was used to test the suitability of the derived model. The results are as indicated in Tables 2-13.

3.1. Investment Sector

The Company had 592 volume trading points. From Table 2, the exceedances above the threshold determined through the standard model, was 30 while the exceedances over the threshold determined through the improved model was 26. The standard two parameter model and standard three parameter model yielded the same number of exceedances above their respective thresholds. The proportion of the exceedances above the threshold was also the same. The improved two and three parameter models also yielded the same number of exceedances. The proportion of the exceedances above the threshold was also the same for the improved two and three parameter models. The scale parameter of the standard

models was 392,000 and that of improved model was 404,000. In this case, the scale parameter of the improved model was higher than that of the standard model. The shape parameter in the standard model was 0.0679 while that of the improved model was 0.0612. In this case, the shape parameter of the improved model was lower than that of the standard model.
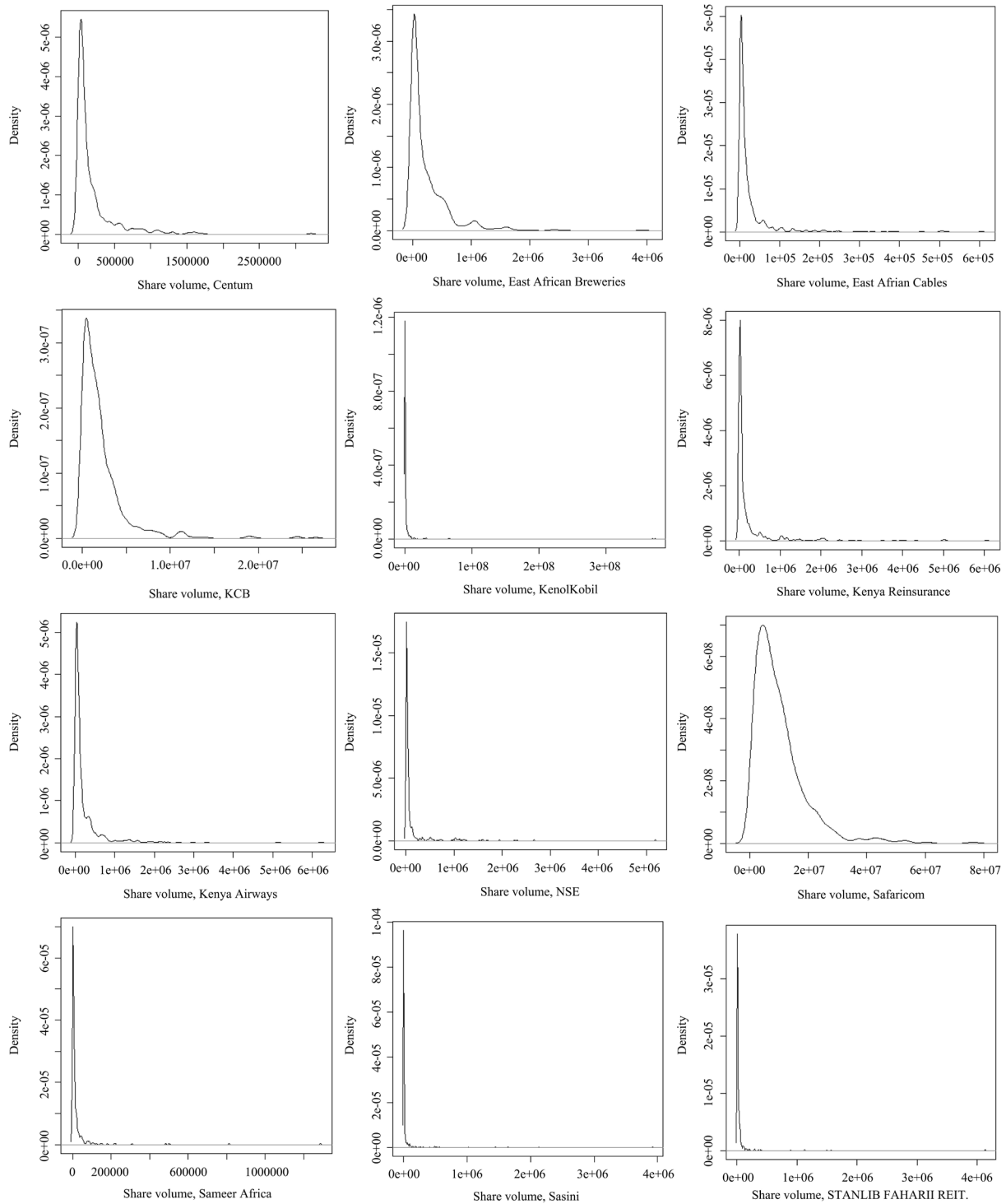

Figure 1. The densities for the market data from the NSE (Share volumes).

Table 2. MLE and MPLE Estimates-Centum.

	Two parameter		Three parameter	
Estimate	Standard	Improved	Standard	Improved
Threshold	820,065.217	864,608.8765	820,085.0487	864,598.9
No. above threshold	30	26	30	26
Proportion above	0.0507	0.0503	0.0507	0.0503
Scale Estimate	3.92E+05	4.04E+05	3.92E+05	4.04E+05
Scale standard error	1.05E+05	1.16E+05	1.05E+05	1.16E+05
Shape Estimate	6.79E−02	6.16E−02	6.79E−02	6.16E−02
Shape standard error	1.95E−01	2.08E−01	1.95E−01	2.08E−01
Asymptotic var-cov.scale	1.09E+10	1.33E+10	1.09E+10	1.33E+10
Asymptotic var-cov.shape	3.80E−02	4.33E−02	3.80E−02	−1.65E+04
Deviance	832.3728	723.0676	832.3695	723.0689
Penalized Deviance	832.4959	723.1747	832.4927	723.1761
AIC	836.3728	727.0676	836.3695	727.0689
Penalized AIC	836.4959	727.1747	836.4927	727.1761

Table 3. MLE and MPLE Estimates-East African Breweries (EABL).

	Two parameter		Three parameter	
Estimate	Standard	Improved	Standard	Improved
Threshold	1,025,741.62	1,058,458.097	1,025,775.349	1,058,460.419
No. above threshold	35	29	35	29
Proportion above	0.0501	0.0511	0.0501	0.0511
Scale Estimate	4.65E+05	5.23E+05	4.65E+05	5.23E+05
Scale standard error	5937.1688	8.40E+03	5.94E+03	8400.6126
Shape Estimate	1.31E−01	5.71E−02	1.31E−01	5.71E−02
Shape standard error	0.1567	0.1367	0.1567	0.1367
Asymptotic var- cov.scale	3.53E+07	7.06E+07	3.53E+07	7.06E+07
Asymptotic var- cov.shape	−4.01E+01	1.87E−02	2.46E−02	2.46E−02
Deviance	982.4578	821.3793	982.452	821.379
Penalized Deviance	982.7215	821.5269	982.7158	821.5266
AIC	986.4578	825.3793	986.452	825.379
Penalized AIC	986.7215	825.5269	986.7158	825.5266

Table 4. MLE and MPLE Estimates-East African Cables (EAC).

	Two parameter		Three parameter	
Estimate	Standard	Improved	Standard	Improved
Threshold	107,940.9	209,150.4	107,967.3	209,135.6
No. above threshold	36	17	36	17
Proportion above	0.0506	0.0511	0.0506	0.0511
Scale Estimate	1.42E+05	149300	1.42E+05	1.49E+05
Scale standard error	3.27E+04	4.65E+04	32691.52	4.65E+04
Shape Estimate	−4.60E−02	−0.175	−4.58E−02	−1.75E−01
Shape standard error	1.59E−01	2.00E−01	0.159	2.00E−01
Asymptotic var-cov.scale	1.07E+09	2.16E+09	1.07E+09	2.16E+09
Asymptotic var-cov.shape	2.53E−02	4.00E−02	2.53E−02	4.00E−02
Deviance	926.075	438.3711	926.0622	438.374
Penalized Deviance	926.075	438.3711	926.0622	438.374
AIC	930.075	442.3711	930.0622	442.374
Penalized AIC	930.075	442.3711	930.0622	442.374

Table 5. MLE and MPLE Estimates-Kenya Re-Insurance.

	Two parameter		Three parameter	
Estimate	Standard	Improved	Standard	Improved
Threshold	1,166,453.64	1,504,595.999	1,166,456.859	1,504,609.093
No. above threshold	37	28	37	28
Proportion above	0.0508	0.0512	0.0508	0.0512
Scale Estimate	1.35E+06	1.40E+06	1.35E+06	1.40E+06
Scale standard error	3.08E+05	3.58E+05	3.08E+05	3.58E+05
Shape Estimate	−4.15E−02	−8.24E−02	−4.15E−02	−8.24E−02
Shape standard error	1.58E−01	1.73E−01	1.58E−01	1.73E−01
Asymptotic var-cov.scale	9.49E+10	1.28E+11	9.49E+10	1.28E+11
Asymptotic var-cov.shape	2.48E−02	3.01E−02	2.48E−02	3.01E−02
Deviance	1.12E+03	848.127	1118.66	848.1265
Penalized Deviance	1118.661	848.127	1118.66	848.1265
AIC	1122.661	852.127	1122.66	852.1265
Penalized AIC	1122.661	852.127	1122.66	852.1265

Table 6. MLE and MPLE Estimates-Kenya Commercial Bank (KCB).

	Two parameter		Three parameter	
Estimate	Standard	Improved	Standard	Improved
Threshold	7,162,924	6,967,732	7,162,925	6,967,755
No. above threshold	37	36	37	36
Proportion above	0.0501	0.0501	0.0501	0.0501
Scale Estimate	4.50E+06	4.70E+06	4.50E+06	4.70E+06
Scale std error	1.08E+06	1.14E+06	1.08E+06	5931.723
Shape Estimate	7.23E−02	5.42E−02	7.23E−02	5.42E−02
Shape standard error	1.76E−01	1.76E−01	1.76E−01	0.1408
Asymptotic var-cov.scale	1.18E+12	1.29E+12	1.18E+12	3.52E+07
Asymptotic var-cov.shape	3.11E−02	3.09E−02	3.11E−02	1.98E−02
Deviance	1207.36	1177.986	1207.36	1177.986
Penalized Deviance	1207.491	1178.078	1207.491	1178.078
AIC	1211.36	1181.986	1211.36	1181.986
Penalized AIC	1211.491	1182.078	1211.491	1182.078

Table 7. MLE and MPLE Estimates-Kenol.

	Two parameter		Three parameter	
Estimate	Standard	Improved	Standard	Improved
Threshold	6,707,831.85	7,239,959.124	6,707,844.136	7,239,940.507
No. above threshold	32	29	32	29
Proportion above	0.0504	0.0506	0.0504	0.0506
Scale Estimate	3.22E+07	3.50E+07	3.22E+07	3.50E+07
Scale standard error	1.01E+07	1.15E+07	5931.6417	1.15E+07
Shape Estimate	5.57E−01	5.61E−01	5.57E−01	5.61E−01
Shape standard error	2.75E−01	2.90E−01	0.2923	2.90E−01
Asymptotic var- cov.scale	1.01E+14	1.32E+14	3.52E+07	1.32E+14
Asymptotic va-cov.shape	7.57E−02	8.40E−02	8.54E−02	8.40E−02
Deviance	1152.255	1049.732	1152.255	1049.732
Penalized Deviance	1153.961	1051.414	1153.961	1051.414
AIC	1156.255	1053.732	1156.255	1053.732
Penalized AIC	1157.961	1051.414	1157.961	1055.414

Table 8. MLE and MPLE Estimates-Kenya Airways (KQ).

	Two parameter		Three parameter	
Estimate	Standard	Improved	Standard	Improved
Threshold	998,460.871	1,140,889.615	998,448.5253	1,140,928.481
No. above threshold	32	28	32	28
Proportion above	0.0505	0.0508	0.0505	0.0508
Scale Estimate	9.27E+05	9.06E+05	9.27E+05	9.06E+05
Scale standard error	2.42E+05	2.57E+05	5932.8333	2.57E+05
Shape Estimate	9.13E−02	1.21E−01	9.13E−02	1.21E−01
Shape standard error	1.93E−01	2.12E−01	0.1418	2.12E−01
Asymptotic var- cov.scale	5.86E+10	6.58E+10	3.52E+07	6.58E+10
Asymptotic var-cov.shape	3.72E−02	4.49E−02	2.01E−02	4.49E−02
Deviance	942.79	823.3371	942.791	823.3344
Penalized Deviance	942.9646	823.5743	942.9655	823.5717
AIC	946.79	827.3371	946.791	827.3344
Penalized AIC	946.9646	827.5743	946.9655	827.5717

Table 9. MLE and MPLE Estimates-NSE.

	Two parameter		Three parameter	
Estimate	Standard	Improved	Standard	Improved
Threshold	573,942.146	1,025,881.322	573,933.6022	1,025,908.743
No. above threshold	30	21	30	21
Proportion above	0.0508	0.0515	0.0508	0.0515
Scale Estimate	8.14E+05	6.06E+05	8.14E+05	6.06E+05
Scale standard error	5933.39	2.07E+05	2.12E+05	2.07E+05
Shape Estimate	2.09E−02	2.24E−01	2.09E−02	2.24E−01
Shape standard error	0.186	2.67E−01	1.86E−01	2.67E−01
Asymptotic var-cov.scale	3.52E+07	4.28E+10	4.51E+10	4.28E+10
Asymptotic var-cov.shape	1.17E−02	7.14E−02	3.47E−02	7.14E−02
Deviance	876.5707	599.1953	876.5713	5.99E+02
Penalized Deviance	8.77E+02	599.668	876.5999	599.6657
AIC	880.5707	603.1953	8.81E+02	603.1929
Penalized AIC	603.1953	603.668	880.5999	603.6657

Table 10. MLE and MPLE Estimates-Safaricom.

Estimate	Two parameter		Three parameter	
	Standard	Improved	Standard	Improved
Threshold	26,193,609.9	26,206,498.57	26,193,619.03	26,206,476.71
No. above threshold	30	30	30	30
Proportion above	0.0507	0.0512	0.0507	0.0512
Scale Estimate	1.23E+07	1.23E+07	1.23E+07	1.23E+07
Scale standard error	5931.6563	3.13E+06	4194.3092	5931.6563
Shape Estimate	$-3.39E-02$	$-3.32E-02$	$-3.39E-02$	$-3.32E-02$
Shape standard error	0.1361	1.77E−01	0.1361	0.1363
Asymptotic var-cov.scale	3.52E+07	9.79E+12	1.76E+07	3.52E+07
Asymptotic var- cov.shape	1.85E−02	3.12E−02	1.85E−02	1.86E−02
Deviance	1039.625	1039.565	1039.625	1039.565
Penalized Deviance	1039.625	1039.565	1039.625	1039.565
AIC	1043.625	1043.565	1043.625	1043.565
Penalized AIC	1043.625	1043.565	1043.625	1043.565

Table 11. MLE and MPLE Estimates-Sameer.

Estimate	Two parameter		Three parameter	
	Standard	Improved	Standard	Improved
Threshold	77,951.3706	133,216.7968	77,957.31698	133,211.3589
No. above threshold	20	10	20	10
Proportion above	0.0506	0.0521	0.0506	0.0521
Scale Estimate	1.86E+05	2.98E+05	1.86E+05	2.98E+05
Scale standard error	6.73E+04	1.40E+05	6.73E+04	1.40E+05
Shape Estimate	3.07E−01	1.12E−01	3.07E−01	1.12E−01
Shape standard error	2.92E−01	3.52E−01	2.92E−01	3.52E−01
Asymptotic var-cov.scale	4.53E+09	1.97E+10	4.53E+09	1.97E+10
Asymptotic var-cov.shape	8.54E−02	1.24E−01	8.54E−02	1.24E−01
Deviance	522.5111	271.8956	522.5093	271.896
Penalized Deviance	523.1935	272.0439	523.1918	272.0443
AIC	526.5111	275.8956	526.5093	275.896
Penalized AIC	527.1935	276.0439	527.1918	276.0443

Table 12. MLE and MPLE Estimates-Sasini.

	Two parameter		Three parameter	
Estimate	Standard	Improved	Standard	Improved
Threshold	99,341.6073	276,597.6868	99,345.1023	276,579.6752
No. above threshold	36	14	36	14
Proportion above	0.0506	0.0536	0.0506	0.0536
Scale Estimate	3.97E+05	7.29E+05	3.97E+05	7.29E+05
Scale standard error	1.08E+05	3.01E+05	8396.617	1.19E+04
Shape Estimate	3.33E−01	1.96E−01	3.33E−01	1.96E−01
Shape standard error	2.22E−01	3.20E−01	0.204	2.87E−01
Asymptotic var-cov.scale	1.17E+10	9.07E+10	7.05E+07	1.41E+08
Asymptotic var-cov.shape	4.94E−02	1.02E−01	4.16E−02	8.21E−02
Deviance	992.7757	405.158	992.7747	405.1589
Penalized Deviance	993.624	405.5148	993.6231	405.5156
AIC	996.7757	409.158	996.7747	409.1589
Penalized AIC	997.624	409.5148	997.6231	409.5156

Table 13. MLE and MPLE Estimates-Stanlib.

	Two parameter		Three parameter	
Estimate	Standard	Improved	Standard	Improved
Threshold	102,597.003	163,127.5551	102,620.9575	163,119.5884
No. above threshold	22	13	22	13
Proportion above	0.051	0.0528	0.051	0.0528
Scale Estimate	4.64E+05	7.08E+05	4.64E+05	7.08E+05
Scale standard error	1.66E+05	3.12E+05	1.66E+05	3.12E+05
Shape Estimate	4.15E−01	2.64E−01	4.15E−01	2.64E−01
Shape standard error	3.02E−01	3.51E−01	3.02E−01	3.51E−01
Asymptotic var-cov.scale	2.76E+10	9.75E+10	2.76E+10	9.75E+10
Asymptotic var-cov.shape	9.09E−02	1.23E−01	9.10E−02	1.23E−01
Deviance	612.3207	374.8657	612.3172	374.8661
Penalized Deviance	613.3434	375.3727	613.3402	375.3731
AIC	616.3207	378.8657	616.3172	378.8661
Penalized AIC	617.3434	379.3727	617.3402	379.3731

3.2. Manufacturing Sector

In this sector, we considered East African Breweries Company because it was the most active desk in NSE in the manufacturing sector. The standard two parameter model and the standard three parameter model had 35 exceedances over their respective threshold (Table 3). The proportion of the exceedances above their thresholds was also the same. The improved two parameter model and the improved three parameter model had 29 volumes exceeding the respective thresholds. The proportion of these exceedances was 0.0511. The scale parameter of the standard models was 465,000 while that of the improved model was 523,000, implying that the scale parameter of the improved model was higher than that of the standard model. The shape parameter of the standard model was 0.131 and that of the improved model was 0.0571. The shape parameter of the improved model was lower than that of the standard model.

3.3. Construction Sector

The most active company in NSE from this sector was East African Cables Company. On this basis, we considered it in our analysis. The two parameter standard model in Table 4 had 36 volumes exceeding the threshold of 107,940.8638 while the three-parameter standard model had 36 volumes exceeding a threshold of 107,967.3084. The number exceeding the threshold represented a proportion of 0.0506. The improved two parameter model yielded a threshold of 209,150.4401 above which there were 17 exceedances as compared to the three parameter improved model which yielded a threshold of 209,135.6041 above which there were 17 exceedances. The proportion of the exceedances in the two parameter improved model and three parameters improved model were 0.0239. The scale parameter of the standard model 142,000 and that of the improved model was 149,300 meaning that the improved model had a higher scale parameter than the standard model. The shape parameter of the standard model was −0.046 and that of the improved model was −0.175 meaning that the shape parameter of the improved model was lower than that of the standard model.

3.4. Insurance Sector

We considered Kenya Reinsurance from this sector because it was the most active company in the NSE trading. In Table 5, the two parameter standard model yielded a threshold of 1,166,453.636 and above it, there were 37 excesses which represented a proportion of 0.0508. The three parameter standard model yielded a threshold of 1,166,456.859 with 37 excesses above it representing a proportion of 0.0508. The two parameter improved model yielded a threshold of 1,504,595.999 with 28 excesses above it representing a proportion of 0.0508 while the three parameter improved model yielded a threshold of 1,504,609.093 with 28 excesses above it. This represented a proportion of 0.0512. The scale parameter of the standard model was 1,350,000 and that of the improved model was 1,400,000. This means that the scale parameter of the improved model was higher than that of the

standard model. The shape parameter of the standard model was −0.0415 and that of the improved model was −0.0824. This means that shape parameter of the improved model was lower than that of the standard model.

3.5. Banking Sector

Kenya Commercial Bank was the most active company from the banking sector in the trading in NSE. We, therefore, considered it to help us visualize the behaviour of the trading in NSE concerning this sector. In Table 6, the threshold of the two parameter standard MPS model was 7,162,924.498 with 37 volumes above it. This represented a proportion of 0.0501. The threshold of the two parameter improved MPS model was 6,967,731.98 with 36 volumes above it. This represented a proportion of 0.0501. The two proportions were the same. The three parameter standard model yielded a threshold of 7,162,925.148 and 37 excesses over it. The three parameter improved MPS model yielded a threshold of 6,967,754.69 with 36 volumes above it. The proportion above the threshold in both models was 0.0501. The scale parameter of the two types of models were also the same. The shape parameter of the standard model was 0.0723 and that of the improved model was 0.0542. The shape parameter of the improved model was lower than of the standard model.

3.6. Energy and Petroleum

Kenol Kobil happens to be one of the most busy desks in NSE and therefore we considered it in this sector in order to visualize the happenings in this sector In Table 7, the two parameter standard MPS model yielded a threshold of 6,707,831.854 with 32 volume points exceeding it, constituting a proportion of 0.0504. The three parameter standard model yielded a threshold of 6,707,844.136 with 32 volume points exceeding it, making a proportion of 0.0504. The two parameter improved model yielded a threshold of 7,239,959.124 giving rise to 29 volume points above it. The three-parameter improved model yielded a threshold of 7,239,940.507 with 29 volume points above it. The two improved models yielded the same number of exceedances and had the same proportion of 0.0506 of exceedances over the threshold. The scale parameter of the standard model was 32,200,000 and that of the improved model was 35,000,000. The scale parameter of the improved model was higher than that of the standard model. The shape parameter of the standard model was 0.557 while that of the improved model was 0.561 meaning that the shape parameter of the improved model higher than that of the standard model.

3.7. Commercial and Services

In this sector, the most active company was Kenya Airways (KQ) and on this strength, we included it in our analysis. Table 8 indicates that the two parameter standard model had a threshold of 998,460.8711 while the three parameter standard model had a threshold of 998,448.5453. Both of these models had 32 exceed-

ances of the volume points contributing to a proportion of 0.0505. The two parameter improved model yielded a threshold of 1,140,889.615 while the three parameter improved model yielded a threshold of 1,140,928.481. These two improved models had 28 volume points of exceedances over the respective threshold which measured to a proportion of 0.0508. The scale parameter of the standard parameter was 927,000 and of the improved model was 906,000. The scale parameter of the standard model was higher than that of the improved model. The shape parameter of the standard model was 0.0913 and that of the improved model was 0.121. The shape parameter of the improved model was higher than that of the standard model.

3.8. Investments Service

In Table 9, the threshold of the two parameter standard model was 573,942.1563 with 30 exceedances while its counterpart three parameter standard model yielded a threshold of 573,933.6022 with 20 exceedances. These two models had exceedances contributing to a proportion of 0.0508. The two parameter improved model yielded a threshold of 1,025,881.322 with 21 exceedances. The three parameter improved model had a threshold of 1,025,908.743 with 21 exceedances. The exceedances over the threshold in the two improved models contributed to a proportion of 0.0515. The threshold of the improved models was higher than those of the standard models. The number above the threshold is lower in case of improved models compared to the standard models. The scale parameter of the standard model was 814,000 and that of the improved model was 606,000. The scale parameter of the improved model was lower than that of the standard model. The shape parameter of the standard model was 0.0209 while that of the improved model was 0.224. The shape parameter of the improved model was higher than that of the standard model.

3.9. Telecommunication Sector

The threshold of a two parameter standard model in Table 10 was 26,193,609.89 with 30 exceedances. The three parameter standard model had a threshold of 26,193,619.03 with 30 exceedances in both cases; the proportion of exceedances was 0.0507. The two parameter improved model had a threshold of 26,206,498.57 while that of the three-parameter improved model had a threshold of 26,206,476.71 with 30 exceedances. In both improved models, the proportion of the exceedances over the threshold was 0.0512. The threshold obtained in the improved models were higher than that obtained from the two standard models. The scale parameters were the same for the standard and improved models. Interestingly, the shape parameters were also the same for standard and improved models.

3.10. Automobiles and Accessories

The most active counter in this sector was Sameer group of companies. We therefore analyzed the Sameer Company's data to visualize the happenings within this sector. In Table 11, the threshold of the two parameter standard

models was 77,951.37064 with 20 exceedances. The three parameter standard model had a threshold of 77,957.31698 with 20 exceedances. The two models had exceedances contributing a proportion of 0.0506. The two parameter improved model had a threshold of 133,216.7968 with 10 exceedances. The three-parameter improved model had a threshold of 13,311.3589 with 10 exceedances. The two improved models had a proportion of 0.0521 in exceedances. The improved models gave higher thresholds than those of the standard models. The scale parameter of the standard model was 186,000 while that of the improved model was 298,000. The scale parameter of the improved model was higher than that of the standard model. The shape parameter of the standard model was 0.307 while that of the improved model was 0.112 meaning that the shape parameter of the improved model was lower than that of the standard model.

3.11. Agricultural Sector

In this sector, we considered using Sasini Company since it was the most busy company in the sector. In Table 12, the threshold of the two parameter standard model was 102,597.0028 with 22 exceedances while the three parameter standard model was 102,620.9575 with 22 exceedances. Both of these models had exceedances of a proportion of 0.051. The threshold of the two parameter improved model was 163,127.5551 with 13 exceedances. The threshold of the three parameter improved model was 163,119.5884 with 13 exceedances. The exceedances in the two models constituted a proportion of 0.0536. The threshold obtained through improved MPS model was higher than that obtained through the standard MPS model. The scale parameter of the standard model was 397,000 while that of the improved model was 729,000. This indicates that the scale parameter of the improved model was higher than that of the standard model. The shape parameter of the standard model was 0.333 while that of the improved model was 0.196 meaning that the shape parameter of the improved model was lower than that of the standard model.

3.12. Real Estate

The threshold of a two parameter standard model was 1,025,977.0028 with 22 exceedances Table 13. The threshold of the three parameter standard model was 102,620.9575 with 22 exceedances. The two models had a proportion of 0.051 in exceedances. The threshold of the two parameter improved model was 163,127.5551 with 13 exceedances. The threshold of the three parameter improved model was 163,119.5884 with 13 exceedances. The proportion of the exceedances in the two improved models was 0.0528. The threshold obtained through the improved MPS model was higher than that obtained through the standard model. The number of volume of the excesses was lower for improved model as compared to standard model. The scale parameter of the standard model was 464,000 while that of the improved model was 708,000, implying that the scale parameter of the improved model was higher than that of the standard model. The shape parameter of the standard model was 0.415 and that of the

improved model was 0.264 meaning that the shape parameter of the improved model was lower than that of the standard model.

4. Conclusions

This study helped to improve the MPS model by introducing the concept of *f* to both two parameter and three parameter MPS models [16] [17] [18]. An investigation was done to determine whether the NSE trading volume data contained ties 1. Interestingly, all companies in all the sectors listed in the NSE trading platform, contained ties. This fact reinforced the importance of this study. The improved two parameter and three parameter MPS models were developed to take care of data that would contain ties. These models were compared with their standard MPS models by comparing their performance in the NSE trade volume data Tables 2-13. In all the tables, the improved models yielded a higher threshold as compared to the threshold obtained through the standard MPS models. The number of exceedances was lower in the case of improved models as compared to the standard models. When the scale parameter was big for the improved model, the shape parameter was small as indicated in Tables 2-6, and Tables 10-13. When the scale parameter was small, the shape parameter was big as indicated in Table 7 and Table 8. For the telecommunication sector represented by Safaricom, Table 9 had scale parameters for the two types of models being the same and so the shape parameters were also the same. The deviance statistics of the improved models were lower than those of the standard models. The AIC criterion was lower in case of the improved models as compared to the standard MPS models. Whenever there are two or more competing models, the model with the lower deviance statistics and lower AIC criterion happens to be the best model. These two statistics, helps us to conclude that the improved model performs better than the standard models. Therefore the improved MPS model is the best method to model the data and to determine the threshold for different companies in different sectors because there is a high likelihood of company's data to contain ties. Improved MPS model has the advantage of modeling the data whether it contains ties or not because when $f = 1$, the improved model reduces to the standard MPS model.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1] Butterfield, R. (2009) DFID Economic Impacts of Climate Change in Kenya, Rwanda and Burundi. ICPAC Kenya and SEI, 1-45.

[2] Mario, H. (2011) Agricultural Management for Climate Change Adaptation. IRI.

[3] Katz, R. and Murphy, A. (1997) Economic Value of Weather and Climate Forecast. Cambridge University Press, London. https://doi.org/10.1017/CBO9780511608278

[4] Prudhome, C. and Duncan, W. (1999) Mapping Extreme Rainfall in Mountainous Region Using Geostatistical Techniques: A Case Study of Scotland. *International*

Journal of Climatology, **19**, 1337-1356.
https://doi.org/10.1002/(SICI)1097-0088(199910)19:12<1337::AID-JOC421>3.3.CO;2-7

[5] Panthou, G., Vischel, T., Lebel, T., Blanchet, J., Quantin, G. and Ali, A. (2012) Extreme Rainfall in West Africa: A Regional Modeling. *Water Resources Research*, **48**, W08501.

[6] Box, G.E.P. and Wilson, K.B. (1951) On the Experimental Attainment of Optimum Conditions. *Journal of the Royal Statistical Society*: *Series B* (*Methodological*), **13**, 1-38. https://doi.org/10.1111/j.2517-6161.1951.tb00067.x

[7] Embrechts, P., McNeil, A. and Straumann, D. (2002) Correlation and Dependence in Risk Management: Properties and Pitfalls. In: Dempster, M.A.H., Ed., *Risk Management*: *Value at Risk and Beyond*, Cambridge University Press, Cambridge, 176-223. https://doi.org/10.1017/CBO9780511615337.008

[8] Annie, P., David, H., Ann, C. and Stephene, P. (2007) Importance of Tail Dependence in Bivariate Frequency Analysis. *Journal of Hydrologic Engineering*, **12**, 394-403.

[9] Balkema, A. and de Haan, L. (1974) Residual Life Time at a Great Age. *The Annals of Probability*, **2**, 792-804. https://doi.org/10.1214/aop/1176996548

[10] Pickands, J. (1975) Statistical Inference Using Extreme Order Statistics. *The Annals of Statistics*, **3**, 119-131. https://doi.org/10.1214/aos/1176343003

[11] Fisher, R. and Tippet, L. (1928) Limiting Forms of the Frequency Distributions of the Largest or Smallest Member of a Sample. *Mathematical Proceedings of the Cambridge Philosophical Society*, **24**, 180-190. https://doi.org/10.1017/S0305004100015681

[12] Hill, B.M. (1975) A Simple General Approach to Inference about the Tail of a Distribution. *The Annals of Statistics*, **13**, 331-341.

[13] Cheng, R.C.H. and Amin, N.A.K. (1983) Estimating Parameters in Continuous Univariate Distribution with a Shifted Origin. *Journal of the Royal Statistical Society*: *Series B* (*Methodological*), **45**, 394-403. https://doi.org/10.1111/j.2517-6161.1983.tb01268.x

[14] Wong, S.T.W. and Wai, K.L. (2010) A Threshold Approach for Peaks over Threshold Modeling Using Maximum Product of Spacing. *Statistica Sinica*, **20**, 1257-1272.

[15] Umash, S., Kumar, S. and Singh, R.K. (2014) A Comparative Study of Traditional Estimation Methods and Maximum Product of Spacing Method in Generalized Inverted Exponential Distribution. *Journal of Statistics Applications and Probability*, **3**, 153-169. https://doi.org/10.12785/jsap/030206

[16] Loung, A. (2018) Unified Asymptotic Results for Maximum Spacing and Generalized Spacing Method for Continuous Models. *Open Journal of Statistics*, **8**, 614-639. https://doi.org/10.4236/ojs.2018.83040

[17] Almetwally, E.M. and Almongy, H.M. (2019) Maximum Product of Spacing and Bayesian Method for Parameter Estimation for Generalized Power Weibull Distribution under Censoring Scheme. *Journal of Data Science*, **17**, 407-444.

[18] Cheng, R.C.H. and Stephen, M.A. (1989) A Goodness of Fit Test Using Moran's with Estimated Parameters. *Biometrika*, **76**, 386-392. https://doi.org/10.1093/biomet/76.2.385

[19] Murage, P., Mung'atu, J. and Odero, E. (2019) Optimal Threshold Determination for the Maximum Product of Spacing Methodology with Ties for Extreme Events. *Open Journal of Modelling and Simulation*, **7**, 149-168. https://doi.org/10.4236/ojmsi.2019.73008