Mathematical Theory and Modeling
ISSN 2224-5804 (Paper) ISSN 2225-0522 (Online)
Vol.4, No.2, 2014

Modeling Insurance Returns with Extreme Value Theory (A Case Study for Kenya's Fire Industrial Insurance Class of Business).

H.W. Wainaina*, A.G. Waititu

Department of Statistics and Actuarial Science, Jomo Kenyatta University of Agriculture and Technology, P.O. Box 62,000-00200, Nairobi, Kenya.

*hwaweru62@gmail.com

Abstract

Most General insurance companies have faced huge losses arising from fire industrial class of business .It is for this reason this study uses extreme value theory approach to model these returns. Traditionally normal distribution was applied and could not capture rare events which caused enormous losses. Kenya's Fire industrial insurance data for five insurance companies and average for entire industry was read into R program .The objective was to plot the time series data. The time series plots aimed to capture the trend and the behavior of the returns over a seven year period. The returns were then standardized in order to transform the negative returns. Using fExtremes in R, the mean excess plot was obtained which helped in measuring the shape of the distribution in the tail. The returns were fitted in a GPD Model in which the excess distribution and the tail of the underlying distribution were obtained over a chosen threshold. These were significant in capturing the values that exceeded the threshold. They were found to be a smooth curve which implied the GPD fit was a good for the data. Scatter plot was obtained and a solid line was observed in the scatter plot which was the smooth of the scattered residuals. QQ plots were also obtained and followed linear form which implied that the parametric model fitted the data well. VaR estimate was finally obtained using extreme value method. The log log empirical distribution was also obtained and indicated how the data points were distributed. After the excesses over a high threshold were fitted to the GPD, parameters were estimated which were used to estimate VaR at different confidence intervals.

Key words: Extreme Value Theory, Peak Over Threshold, Generalized Pareto Distribution, Value at Risk.

1.0: Introduction

Modeling large fires is becoming critical in the analysis of disturbances on ecosystems at the landscape and larger spatial scales. They have such a tremendous effect on ecosystems that is obvious that mathematical modeling of these events will help us understand ecosystem dynamics better. However, most of the current ecological models are based on measuring means and variances.

According to John P Hall (1982) estimating the value of property damaged as a result of fire is more an art than science, because there are no generally accepted procedures for loss estimation. Loss adjusters always have a difficult time determining what areas were damaged, how badly and estimating the total loss.

Lack of such procedures undermines consistency in loss estimation from one fire department to another, many who use fire loss data for planning and management are understandably nervous about its accuracy; horror stories continue to circulate about observers whose estimates of loss at the same fire differ 10 to 1 or more.(p.11)

In wild-land fires, the poison Model (Davis 1965) and the Truncated Pareto and Lognormal distributions (Strauss et al. 1989) have been suggested to model distributions that include large fires.

The risk of large losses and consequently large insurance claims, have been modeled with Pareto, Gamma, and Lognormal distributions for deciding on deductible and premium levels (Nigm et al.1987).

In large structural fires, a probabilistic approach to fire risk has been pursued to some extent by modeling fire growth, damage distribution, and fire spread (Ramachandran 1988). In contrast to wild land fire research, the studies on structural fires have concentrated on large fires rather than small ones.

A lack of adequate modeling of extreme fire causes decisions to be based solely on fire manager's experience and subjective assessment of the situation during large conflagrations (Thomas 1989).

When one parametric distribution is fitted for all the data, regardless of damage level, the probability of large fires will be underrepresented or not represented at all. Alvarado-Celestino (1992) shows that the characteristic largest value predicted by a probability distribution when fitted to the entire fire size distribution (e.g. the Weibull model).

In classical statistical modeling, large fires constitute outliers and models usually do not deal with them. One approach that has been suggested for outlier detection is to assume that outliers have different distribution from the rest of the observations (Davies and Gather 1993). This has also been suggested by several authors in the case of forest fires. In this study the extreme value distribution best describes the large fire distribution.

There have been previous studies for instance Alexander McNeil's (1996) study of the Danish data on large insurance losses provides us with a good example of use of EVT in this context. The goal of this reference is to show additional techniques and plotting strategies which can be employed in similar data in Kenya's insurance industry.

Until recently, the value-at-risk (VaR) approach was the standard for the risk management industry. VaR measures the worst anticipated loss over a period for a given probability and under normal market conditions. It can also be said to measure the minimal anticipated loss over a period with a given probability and under exceptional market conditions (Longuin 1999).

The VaR approach (see Jorion 1996) has been the subject of several criticisms. The most significant is that the majority of the parametric methods use a normal distribution approximation.

Using this approximation, the risk of the high quantiles is underestimated, especially for the fat tailed series, which are common in financial data.

Unlike VaR methods, no assumptions are made about the nature of the original distribution of all the observations. Some EVT techniques can be used to solve for very high quantiles, which is very useful for predicting crashes and extreme-loss situations.

The advantage of estimating VaR using GPD method is that this method can estimate VaR outside the sampling interval.

Harmantzis et.al. (2005) and Marinelli et.al.(2006) presented the performance of extreme value theory in VaR and expected shortfall estimation compared to the Gaussian and historical simulation models together with other heavy tailed approach. From their results it was found that fat tailed models can predict risk more accurately than non fat tailed ones and there exists the benefits of EVT framework especially method using GPD.

2.0: Methodology

Most statistical methods are concerned primarily with what goes on in the center of a statistical distribution, and do not pay particular attention to the tails of a distribution, or in other words, the most extreme values at either the high or low end. Extreme event risk is present in all areas of risk management – market, credit, day to day operation, and insurance. One of the greatest challenges to a risk manager is to implement risk management tools which allow for modeling rare but damaging events, and permit the measurement of their consequences. Extreme value theory (EVT) plays a vital role in these activities.

The standard mathematical approach to modeling risks uses the language of probability theory. Risks are random variables, mapping unforeseen future states of the world into values representing profits and losses. These risks may be considered individually, or seen as part of a stochastic process where present risks depend on previous risks. The potential values of a risk have a probability distribution which we will never observe exactly although past losses due to similar risks, where available, may provide partial information about that distribution. Extreme events occur when a risk takes values from the tail of its distribution.

We develop a model for risk by selecting a particular probability distribution. We may have estimated this distribution through statistical analysis of empirical data. In this case EVT is a tool which attempts to provide us with the best possible estimate of the tail area of the distribution. However, even in the absence of useful historical data, EVT provides guidance on the kind of distribution we should select so that extreme risks are handled conservatively. There are two principal kinds of model for extreme values. The oldest group of models is the block maxima models; these are models for the largest observations collected from large samples of identically distributed observations. For example, if we record daily or hourly losses and profits from trading a particular instrument or group of instruments, the block maxima/minima method provides a model which may be appropriate for the quarterly or annual maximum of such values. 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. The POT models are generally considered to be the most useful for practical applications, due to a number of reasons. First, by taking all exceedances over a suitably high threshold into account, they use the data more efficiently. Second, they are easily extended to situations where one wants to study how the extreme levels of a variable Y depend on some other variable X for instance, Y may be the level of tropospheric ozone on a particular day and X a vector of meteorological variables for that day. This kind of problem is almost impossible to handle through the annual maximum method. The POT methods are fitted with the generalized Pareto distribution (GPD).

2.1 Generalized pareto distribution

The GPD is a two-parameter distribution with distribution function

$$G_{\xi,\sigma}(x) = \begin{cases} 1-(1+\frac{\xi x}{\beta)^{-1/\xi}} & \xi \neq 0 \\ 1-\exp\left(-\frac{x}{\beta}\right) & \xi = 0 \end{cases}$$

Where β>0, and x≥0 when ξ≥0 and$0 \leq x \leq -\beta/\xi$when ξ<0. The parameters ξ and β are referred to respectively as the *shape* and *scale* parameters.

The GPD is generalized in the sense that it subsumes certain other distributions under a common parametric form. If ξ>0 then is a reparametrized version of the ordinary Pareto distribution ($with\ \alpha = 1/\xi\ and\ \kappa = \beta/\xi$), which has a long history in actuarial mathematics as a model for *large losses*; ξ=0 corresponds to the exponential distribution, i.e. a distribution with a *medium-sized* tail; and ξ<0 is a *short-tailed* Pareto type II distribution. The mean of the GPD is defined provided ξ<1 and is

$$E(X) = \frac{\beta}{1-\xi}$$

The first case is the most relevant for risk management purposes since the GPD is *heavy-tailed* when ξ>0. Whereas the normal distribution has *moments* of all orders, a heavy-tailed distribution does not possess a complete set of moments. In the case of the GPD with ξ >0 we find that $E[x^{\kappa}]$is infinite for $\kappa \geq 1/\xi$. When $\xi = 1/2$, the GPD is an infinite variance (second moment) distribution; when ξ=1/4, the GPD has an infinite fourth moment.

The role of the GPD in EVT is as a natural model for the excess distribution over a high threshold. Certain types of large claims data in insurance typically suggest an infinite second moment; similarly econometricians might claim that certain market returns indicate a distribution with infinite fourth moment. The normal distribution cannot model these phenomena but the GPD is used to capture precisely this kind of behavior.

2.2 Modeling excess distribution

Let *X* be a random variable with distribution function *F*. The distribution of *excesses* over a threshold *u* has distribution function

$$F_u(y) = P\{X - u \leq y | X > u\}$$

For $0 \leq y < x_0 - u\ where\ x_0 \leq \infty$ is the *right endpoint* of *F*. The excess distribution function $F_u$represents the probability that a loss exceeds the threshold u by at most an amount y, given the information that it exceeds the threshold. In survival analysis the excess distribution function is more commonly known as the residual life distribution function — it expresses the probability that, say, an electrical component which has functioned for *u* units of time fails in the time period $[u, u+y]$. It is very useful to observe that F_u can be written in terms of the underlying *F* as

$$F_u(y) = \frac{F(y+u) - F(u)}{1 - F(u)}$$

The mean excess function of a random variable *X* with finite mean is given by

$$e(u) = E(X - u | X > u)$$

The mean excess function $e(u)$ expresses the mean of F_u as a function of *u*. In survival analysis, the mean excess function is known as the mean residual life function and gives the expected residual lifetime for components with different ages.

Mostly we would assume our underlying F is a distribution with an infinite right endpoint, i.e. it allows the possibility of arbitrarily large losses, even if it attributes negligible probability to unreasonably large outcomes, e.g. the normal or t distributions. But it is also conceivable, in certain applications that$F$could have a finite right endpoint. An example is the beta distribution on the interval [0,1] which attributes zero probability to outcomes larger than 1 and which might be used, for example, as the distribution of credit losses expressed as a proportion of exposure.

2.4 The Pickands-Balkema-de Haans theorem

The Pickands-Balkema-de Haan limit theorem (Balkema and de Haan, 1974; Pickands, 1975) is a key result in EVT and explains the importance of the GPD. *We can find a (positive measurable function)* β(*u) such that*

$$\lim_{n\to x_0} \sup_{0\le y<x_0-u} \left|F_u(y) - G_{\xi,\beta(u)}(y)\right| = 0,\ \ \beta(u) = \beta + \xi u$$

$$if\ and\ only\ if\ F \in MDA\big(H_\xi\big).$$

The theorem shows that under MDA conditions the generalized Pareto distribution is the limiting distribution for the distribution of excesses as the threshold tends to the right endpoint. All the common continuous distributions of statistics and actuarial science (normal, lognormal, x^2, t, F, gamma, exponential, uniform, beta, etc.) are in MDA(H_ξ) for some ξ, so the above theorem proves to be a very widely applicable result that essentially says that the GPD is the natural model for the unknown excess distribution above sufficiently high thresholds.

2.5 Fitting a GPD model

Given loss data $X_1, X_2, \dots \dots, X_n$ from F, a random number $N_u$will exceed our threshold$u$; it will be convenient to re-label these data $\tilde{X}_1, \tilde{X}_2, \dots., \tilde{X}_N$. For each of these exceedances we calculate the amount $Y_j = \tilde{X}_j - u$ of the excess loss. We wish to estimate the parameters $\hat{\xi}\ and\ \hat{\beta}$ of a GPD model by fitting this distribution to the N_u excess losses. There are various ways of fitting the GPD including ML and PWM. The former method is more commonly used and easy to implement if the excess data can be assumed to be realizations of independent random variables, since the joint probability density of the observations will then be a product of marginal densities. This is the most general fitting method in statistics and it also allows us to give estimates of statistical error (standard errors) for the parameter estimates. Writing $G_{\xi,\beta}$ for the density of the GPD, the log-likelihood may be easily calculated to be

$$\ln L(\xi, \beta; Y_1, Y_2, \dots, Y_N) = \sum_{j=1}^{N_u} \ln g_{\xi,\beta}(Y_j) = -N_u \ln\beta - \left(1 + \frac{1}{\xi}\right) \sum_{j=1}^{N_u} \left(1 + \xi \frac{Y_j}{\beta}\right)$$

Which must be maximized subject to the parameter constraints that β>0 and $1 + \frac{\xi Y_j}{\beta} > 0\ \ for\ all\ j.$ Solving the maximization problem yields a GPD model $G_{\xi,\beta}$ for the excess distributionF_u.

Choice of the threshold is basically a compromise between choosing a sufficiently high threshold so that the asymptotic theorem can be considered to be essentially exact and choosing a sufficiently low threshold so that we have sufficient material for estimation of the parameters.

3.0 Research results and discussion

In this chapter an empirical analysis was carried out to model the fire industrial insurance returns of the Kenyan Insurance market. Five insurance companies' returns that were captured and modeled are discussed. The average returns for all the companies were calculated and modeled to give the results for the overall industry. The fire industrial insurance returns data was fitted in GPD model and discussed.

3.1 Fidelity Insurance Company

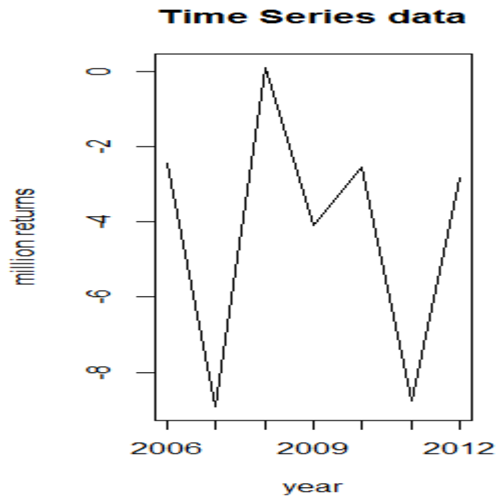


Fig.1(a)

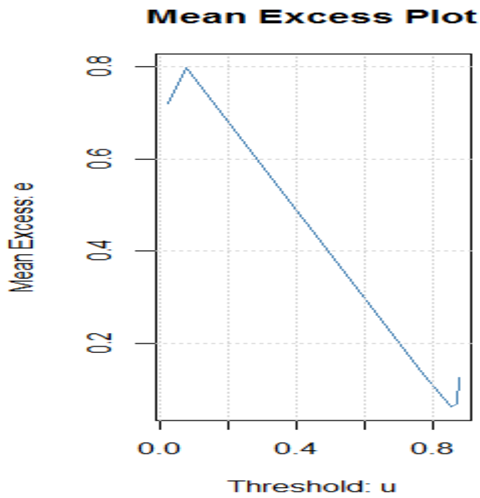


Fig.1(b)

In **Fig.1(a)** time series plot for fidelity insurance company shows presence of periodicity. Original data is then transformed to standardize it using the following function $\log_{10}$(X1+1.05-min(X1)) where X1 is the original data. The mean excess plot of Fidelity insurance company returns is obtained. It is observed that the graph declines and begins an upward trend as shown in **Fig.1b** which indicates the presence of a heavy tailed distribution.

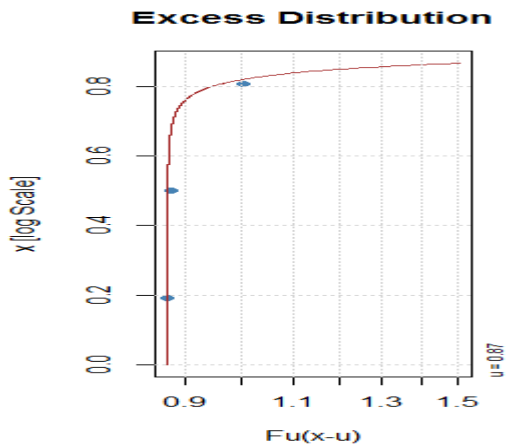


Fig.1(c)

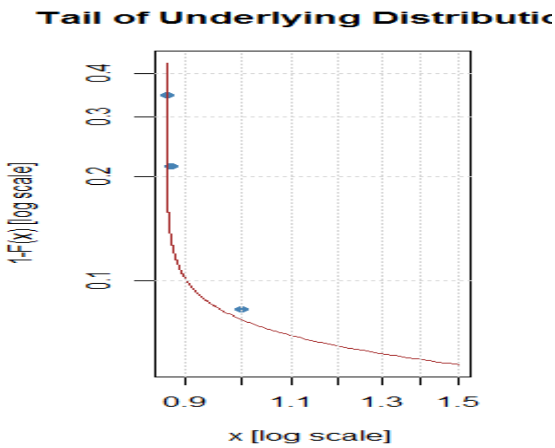
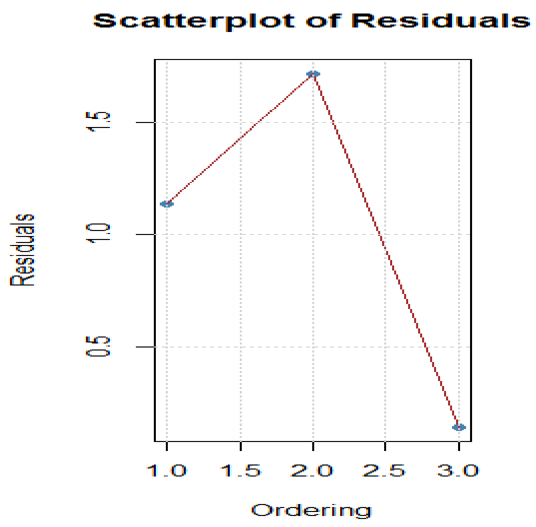


Fig.1(d)

A threshold of 0.87 is chosen, there are 3 out of 7 returns that exceed the threshold. The excesses were fitted to a GPD model using the MLE. The parameter estimates $\xi = 5.2195$ and$\beta = 0.0000875$. The shape parameter ξ is greater than 0 implying a heavy tailed distribution. This can be interpreted to mean that the higher the value of the shape parameter, the higher the derived return. The distribution for the excesses shows a smooth curve meaning GPD fit was a good fit for the data similar for the tail of the underlying distribution as shown in **Fig.1(c)** and **Fig.1(d)**



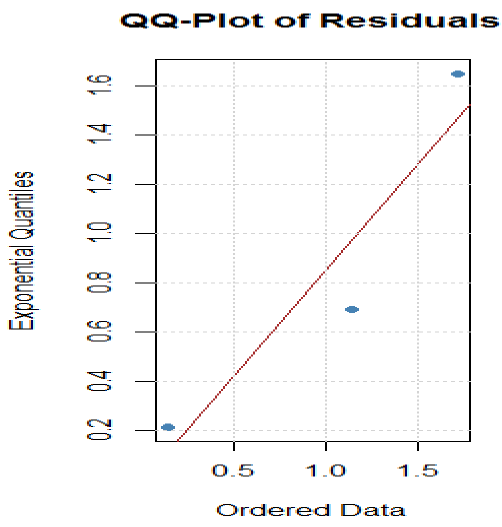


Fig.1(e) **Fig.1(f)**

The scatter plot of residuals from a GPD fitted to the data over a threshold of 0.87. The solid line observed is the smooth of the scattered residuals as shown in **Fig.1(e)**. The QQ plot follows a linear form as shown in **Fig.1(f)**; therefore the parametric model fits the data well.

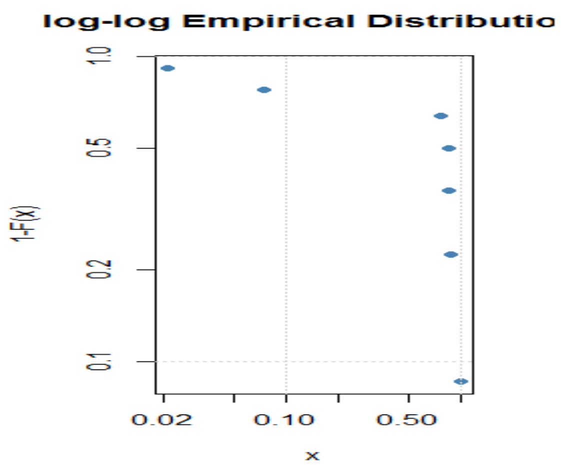


Fig.1(g)

In **Fig.1(g)** the empirical distribution of Fidelity Insurance Company fire returns data fitted defines a CDF consistent with data directly observed in the data set. In other words, it defines a CDF, F(x), such that F(x) is equal to the proportion of data points in the set less than or equal to *x*. The Value at Risk (VaR) with 5% level of confidence was Kshs.8.91 million. This implies that the coming year's loss for the entire industry would exceed is Kshs.8.91 million. Analogously the same interpretation holds for 1%. In practice when portfolio of loss is known, then precautions can be taken to mitigate against it.

3.2 Heritage Insurance Company

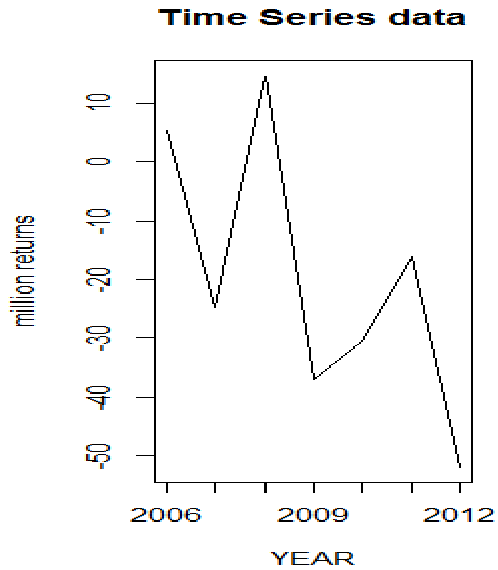


Fig.2(a)

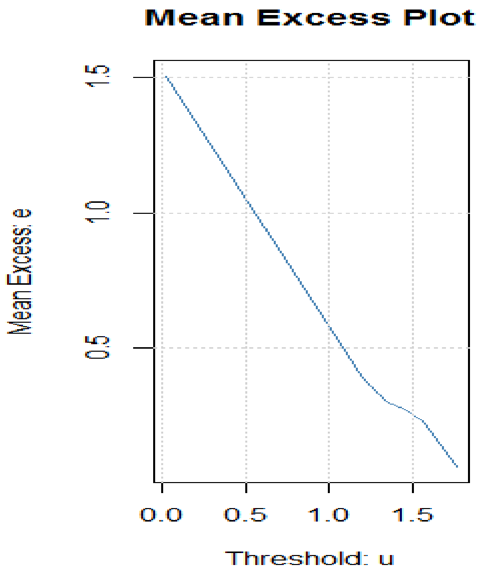


Fig.2(b)

In **Fig.2(a)** time series plot for Heritage Insurance Company shows presence of periodicity. Original data is then transformed to standardize it using the following function $\log_{10}$(X1+1.05-min(X1)) where X1 is the original data. The mean excess plot of Heritage Insurance Company returns is obtained. It is observed that the plot exhibits a downward trend as shown in **Fig.2(b)** which indicates a thin tailed distribution.

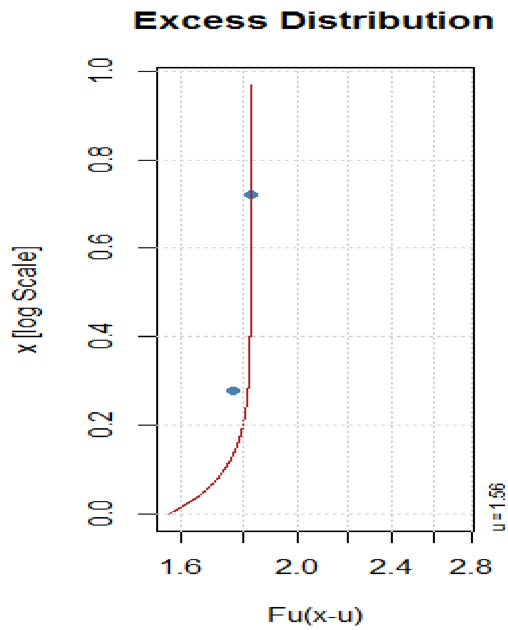


Fig.2(c)

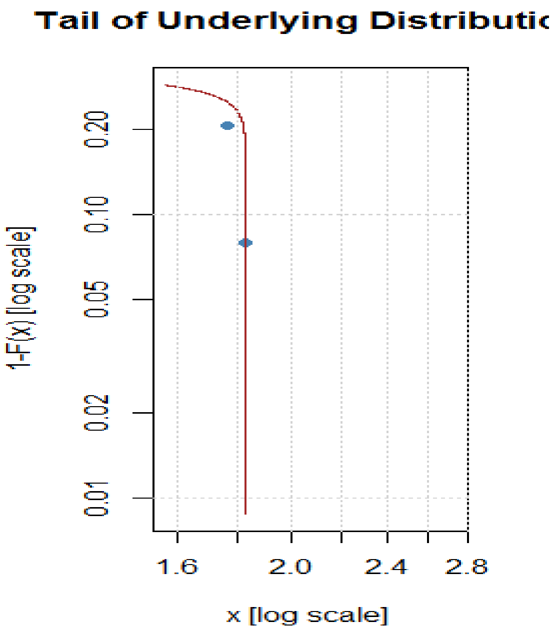


Fig.2(d)

A threshold of 1.56 is chosen, there are 2 out of 7 returns that exceed the threshold.The excesses were fitted to a GPD model using the MLE. The parameter estimates $\xi = -10.3646$ and$\beta = 2.7486$. The shape parameter ξ is less than 0 implying a thin tailed distribution; hence pareto type II is obtained. The distribution for the excesses shows a smooth curve meaning GPD fit was a good fit for the data similar for the tail of the underlying distribution as shown in **Fig.2(c)** and **Fig.2(d).**

.

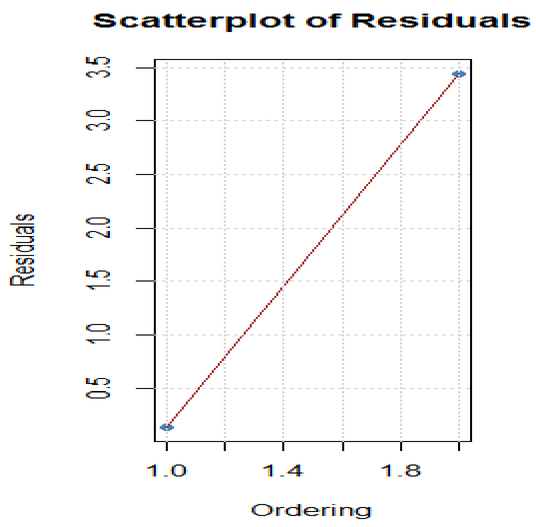


Fig.2(e)

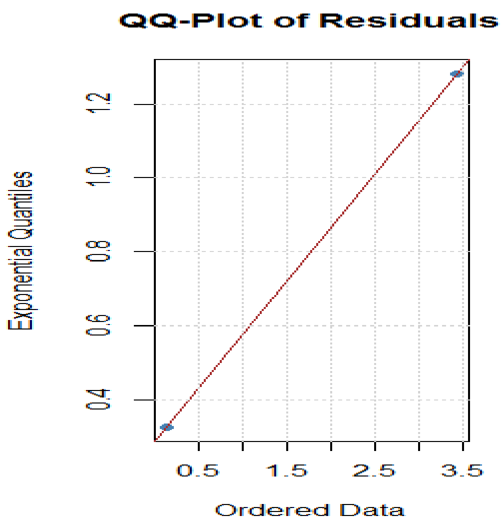


Fig.2(f)

The scatter plot of residuals from a GPD fitted to the data over a threshold of 1.56. The solid line observed is the smooth of the scattered residuals as shown in as shown in **Fig.2(e)**. The QQ plot follows a linear form as shown in **Fig.2(f)**.; therefore the parametric model fits the data well.

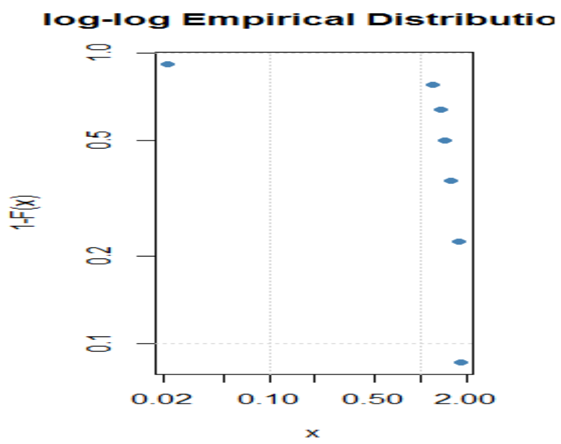


Fig.2(g)

In **Fig.2(g)** the empirical distribution of Heritage Insurance Company fire loss data fitted defines a CDF consistent with data directly observed in the data set. In other words, it defines a CDF, F(x), such that F(x) is equal to the proportion of data points in the set less than or equal to *x*.

The Value at Risk (VaR) with 5% level of confidence was Kshs.51.83 million. This implies that the coming year's loss for the entire industry would exceed is Kshs.51.83 million. Analogously the same interpretation holds for 1%. In practice when portfolio of loss is known, then precautions can be taken to mitigate against it.

3.3 Kenindia Insurance Company

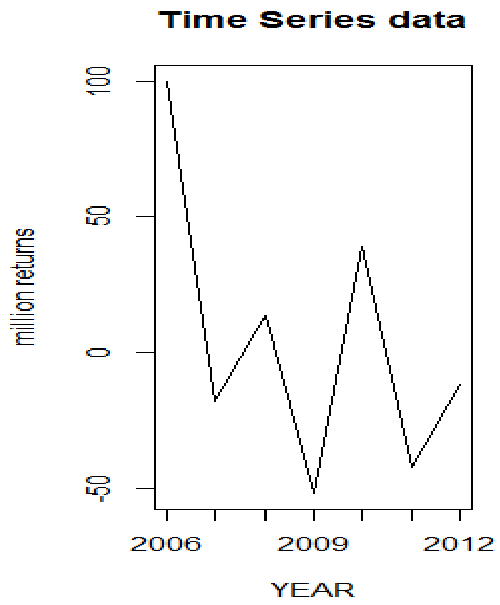


Fig.3(a)

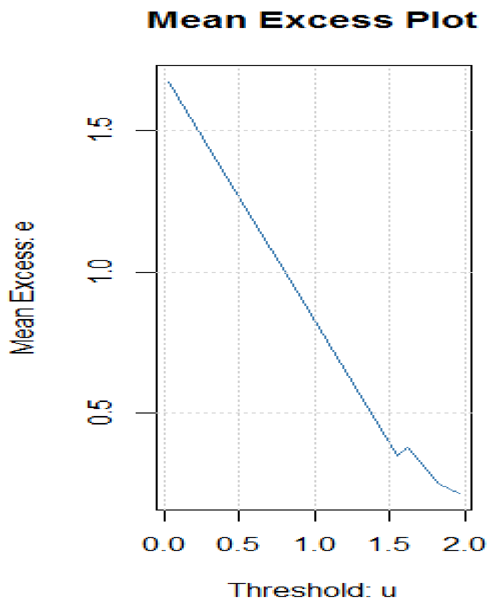


Fig.3(b)

In **Fig.3(a)** time series plot for Kenindia Insurance Company shows presence of periodicity. Original data is then transformed to standardize it using the following function $\log_{10}(X1+1.05-\min(X1))$ where X1 is the original data. The mean excess plot of Kenindia Insurance Company returns is obtained. It is observed that the plot exhibits a downward trend as shown in **Fig.3(b)** which indicates a thin tailed distribution.

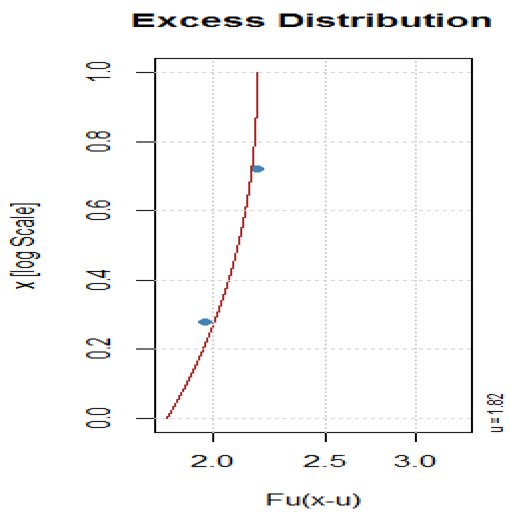


Fig.3(c)

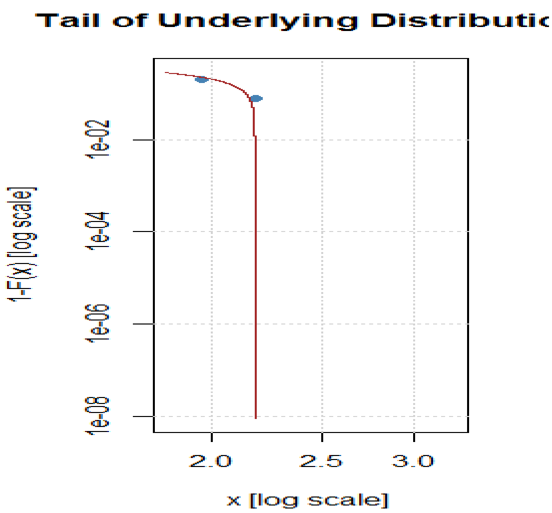


Fig.3(d)

A threshold of 1.82 is chosen, there are 2 out of 7 returns that exceed the threshold. The excesses were fitted to a GPD model using the MLE. The parameter estimates $\xi = -2.1302$ and$\beta = 0.7700$. The shape parameter ξ is less than 0 implying a thin tailed distribution; hence pareto type II is obtained. The distribution for the excesses shows a smooth curve meaning GPD fit was a good fit for the data similar for the tail of the underlying distribution as shown in **Fig.3(c)** and **Fig.3(d).**

.

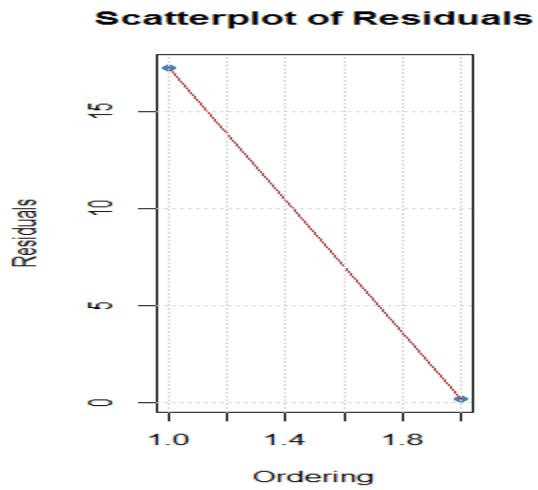


Fig.3(e)

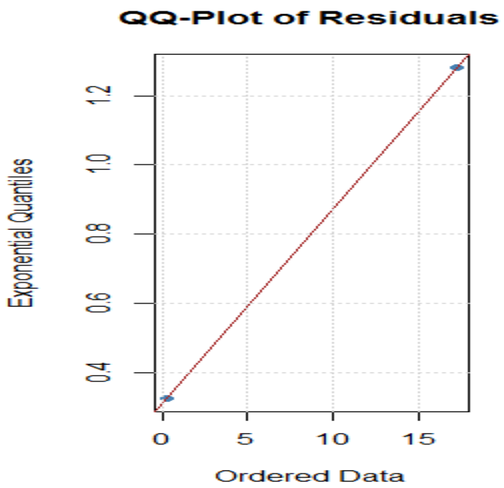


Fig.3(f)

The scatter plot of residuals from a GPD fitted to the data over a threshold of 1.82. The solid line observed is the smooth of the scattered residuals as shown in **Fig.3(e)**. The QQ plot follows a linear form as shown in **Fig.3(f)**; therefore the parametric model fits the data well.

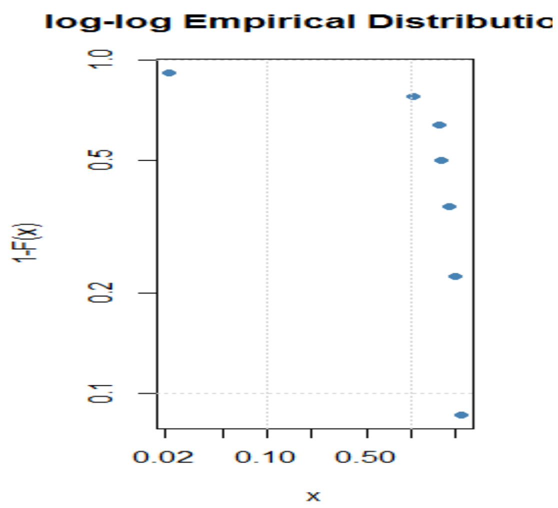


Fig.3(g)

In **Fig.3(g)** the empirical distribution of Kenindia Insurance Company fire loss data fitted defines a CDF consistent with data directly observed in the data set. In other words, it defines a CDF, F(x), such that F(x) is equal to the proportion of data points in the set less than or equal to *x*.

The Value at Risk (VaR) with 5% level of confidence was Kshs.51.7 million.This implies that the coming year's loss for the entire industry would exceed is Kshs.51.7 million. Analogously the same interpretation holds for 1%. In practice when portfolio of loss is known, then precautions can be taken to mitigate against it.

3.4 Mayfair Insurance Company

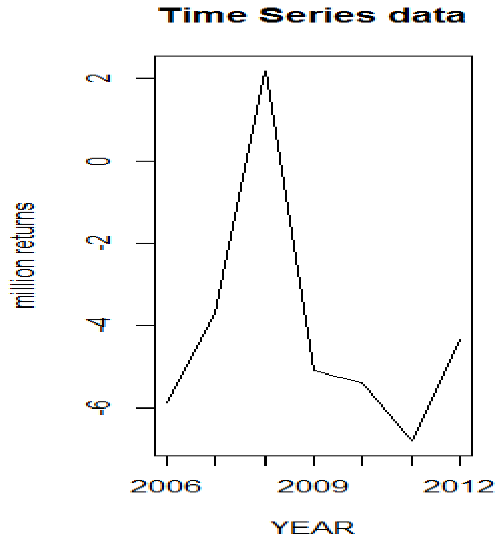


Fig.4(a)

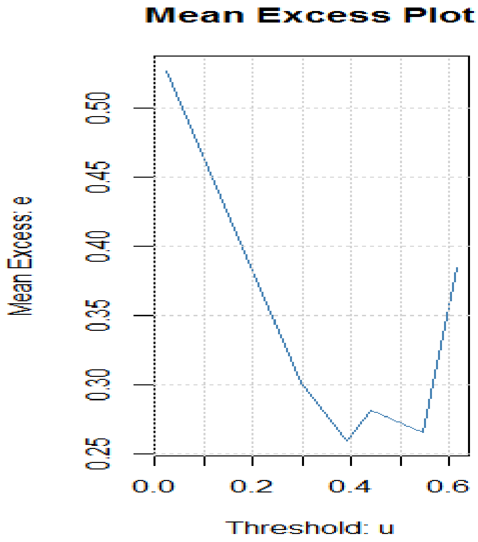


Fig.4(b)

In **Fig.4(a)** time series plot for Mayfair Insurance Company shows presence of periodicity. Original data is then transformed to standardize it using the following function $\log_{10}(X1+1.05-\min(X1))$ where X1 is the original data. The mean excess plot of Mayfair Insurance Company returns is obtained. It is observed that the graph declines and begins an upward trend as shown in **Fig.4(b)** which indicates the presence of a heavy tailed distribution.

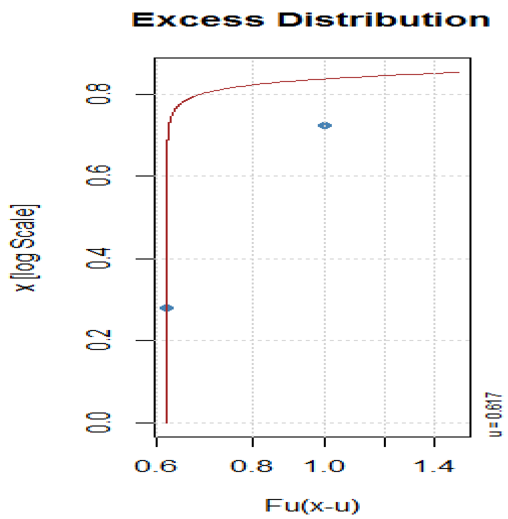


Fig.4(c)

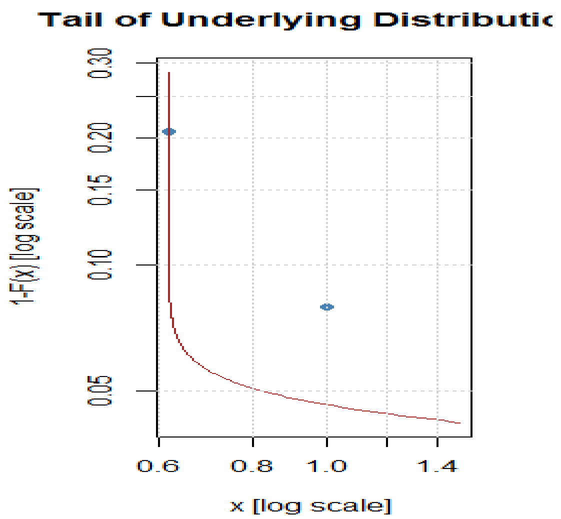


Fig.4(d)

A threshold of 0.617 is chosen, there are 2 out of 7 returns that exceed the threshold. The excesses were fitted to a GPD model using the MLE. The parameter estimates $\xi = 8.3958$ and $\beta = 0.00000075$. The shape parameter ξ is greater than 0 implying a heavy tailed distribution. This can be interpreted to mean that the higher the value of the shape parameter, the higher the derived return. The distribution for the excesses shows a smooth curve meaning GPD fit was a good fit for the data similar for the tail of the underlying distribution as shown in **Fig.4(c)** and **Fig.4(d).**

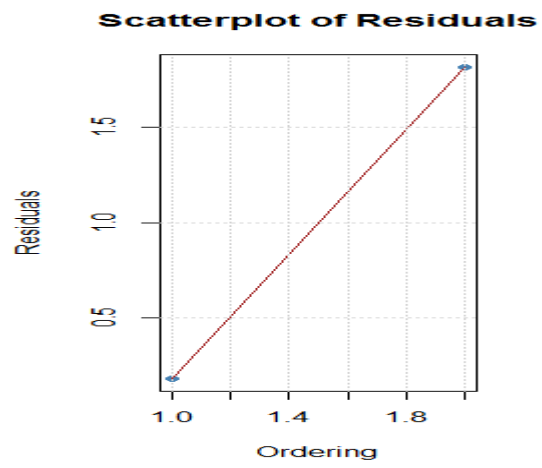


Fig.4(e)

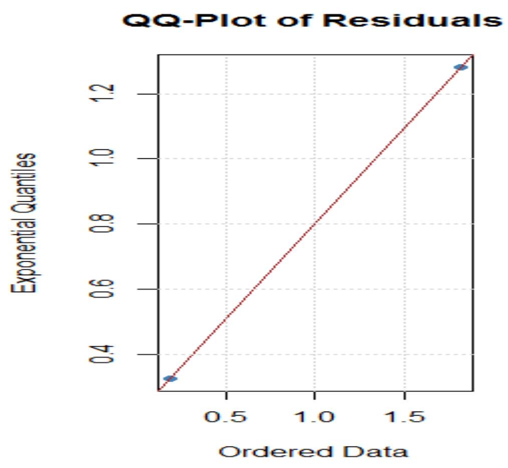


Fig.4(f)

The scatter plot of residuals from a GPD fitted to the data over a threshold of 0.617. The solid line observed is the smooth of the scattered residuals as shown in **Fig.4(e)**. The QQ plot follows a linear form as shown in **Fig.4(f)**; therefore the parametric model fits the data well.

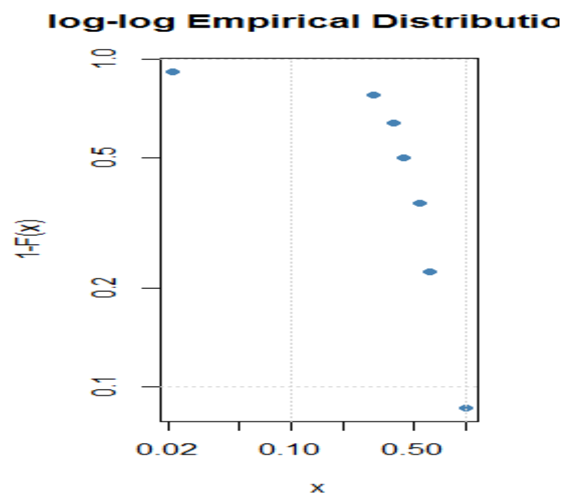


Fig.4(g)

In **Fig.4(g)** the empirical distribution of Mayfair Insurance Company fire loss data fitted defines a CDF consistent with data directly observed in the data set. In other words, it defines a CDF, F(x), such that F(x) is equal to the proportion of data points in the set less than or equal to *x*.

The Value at Risk (VaR) with 5% level of confidence was Kshs.6.8 million. This implies that the coming year's loss for the entire industry would exceed is Kshs.6.8 million. Analogously the same interpretation holds for 1%. In practice when portfolio of loss is known, then precautions can be taken to mitigate against it.

ISSN 2224-5804 (Paper) ISSN 2225-0522 (Online)
Vol.4, No.2, 2014

3.5 Real Insurance Company

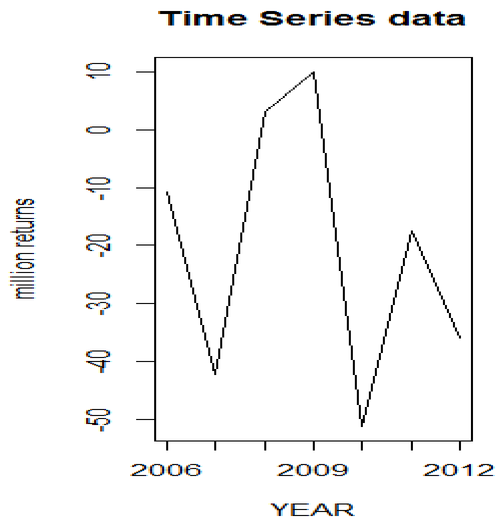


Fig.5(a)

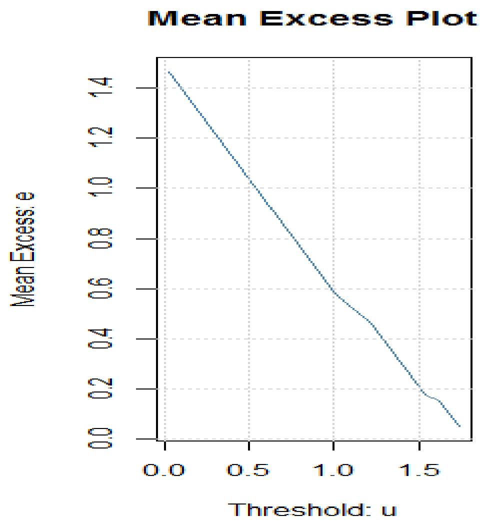


Fig.5(b)

In **Fig.5(a)** time series plot for Real Insurance Company shows presence of periodicity. Original data is then transformed to standardize it using the following function $\log_{10}(X1+1.05-\min(X1))$ where X1 is the original data. The mean excess plot of Real Insurance Company returns is obtained. It is observed that the graph declines and begins an upward trend as shown in **Fig.5(b)** which indicates the presence of a heavy tailed distribution.

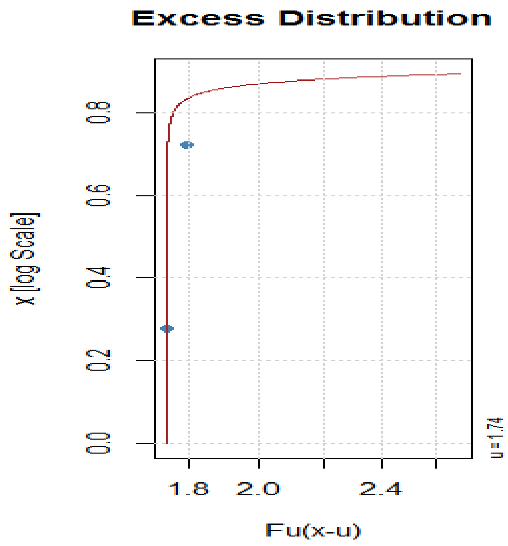


Fig.5(c)

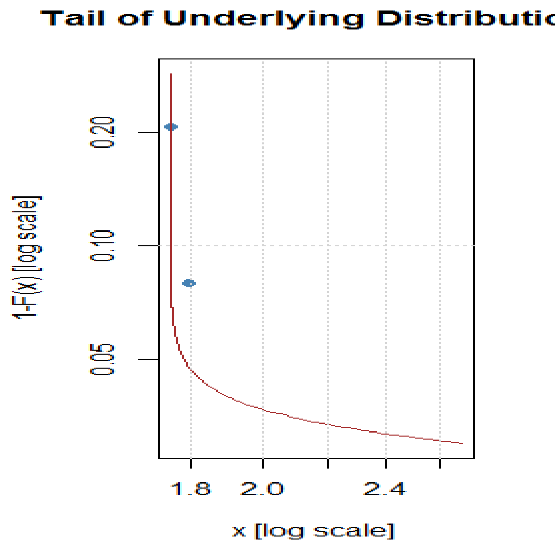


Fig.5(d)

A threshold of 1.74 is chosen, there are 2 out of 7 returns that exceed the threshold. The excesses were fitted to a GPD model using the MLE. The parameter estimates $\xi = 6.3401$ and$\beta = 0.0000037$. The shape parameter ξ is greater than 0 implying a heavy tailed distribution. This can be interpreted to mean that the higher the value of the shape parameter, the higher the derived return. The distribution for the excesses shows a smooth curve meaning GPD fit was a good fit for the data similar for the tail of the underlying distribution as shown in **Fig.5(c)** and **Fig.5(d).**

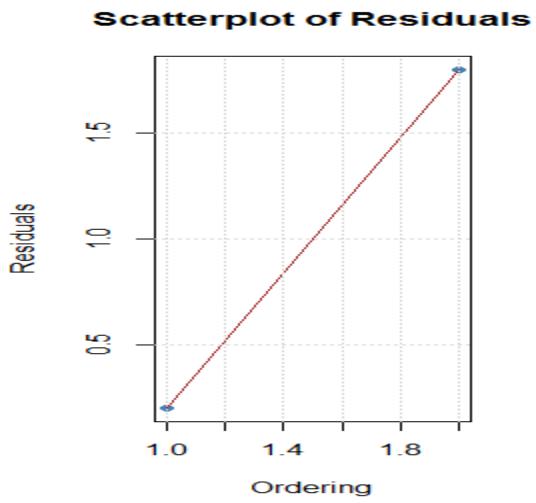


Fig.5(e)

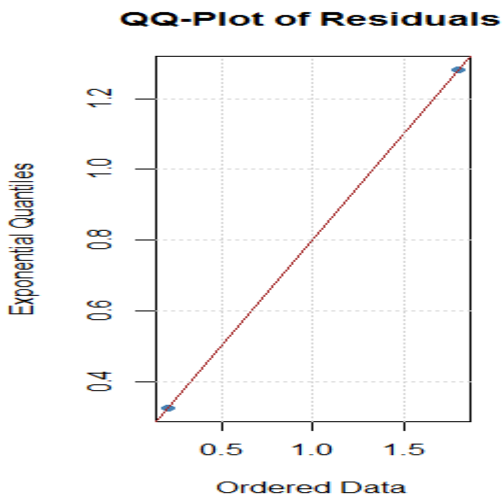


Fig.5(f)

The scatter plot of residuals from a GPD fitted to the data over a threshold of 1.74. The solid line observed is the smooth of the scattered residuals as shown in **Fig.5(e)**. The QQ plot follows a linear form as shown in **Fig.5(f)**; therefore the parametric model fits the data well.

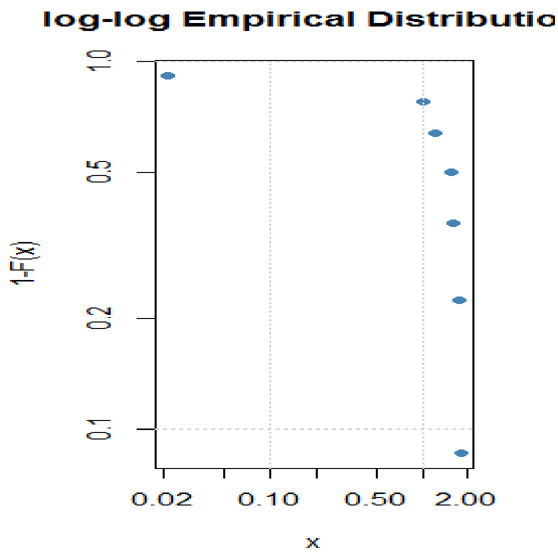


Fig.5(g)

In **Fig.5(g)** the empirical distribution of Real Insurance Company fire loss data fitted defines a CDF consistent with data directly observed in the data set. In other words, it defines a CDF, F(x), such that F(x) is equal to the proportion of data points in the set less than or equal to *x*.

The Value at Risk (VaR) with 5% level of confidence was Kshs.51.24 million. This implies that the coming year's loss for the entire industry would exceed is Kshs.51.24 million. Analogously the same interpretation holds for 1%. In practice when portfolio of loss is known, then precautions can be taken to mitigate against it.

3.6 Insurance Industry

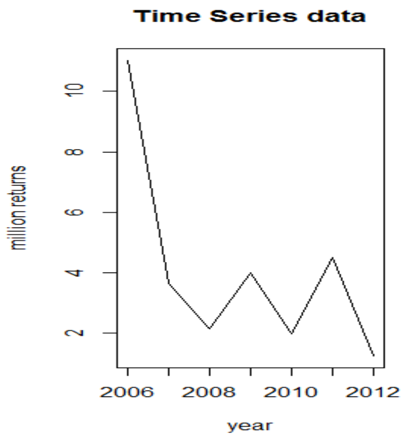


Fig.6(a)

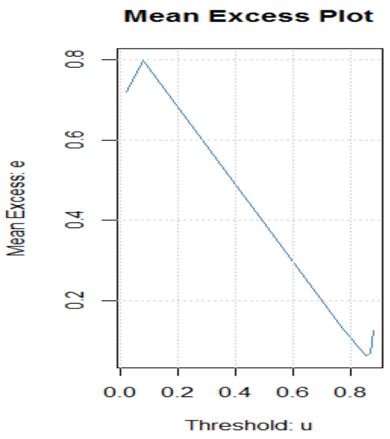


Fig.6(b)

In **Fig.6(a)** time series plot of the insurance industry on average shows presence of periodicity. Original data is then transformed to standardize it using the following function $\log_{10}(X1+1.05-\min(X1))$ where X1 is the original data. The mean excess plot of all insurance companies' average returns is obtained. It is observed that the plot exhibits a downward trend as shown in **Fig.6(b)** which indicates a thin tailed distribution.

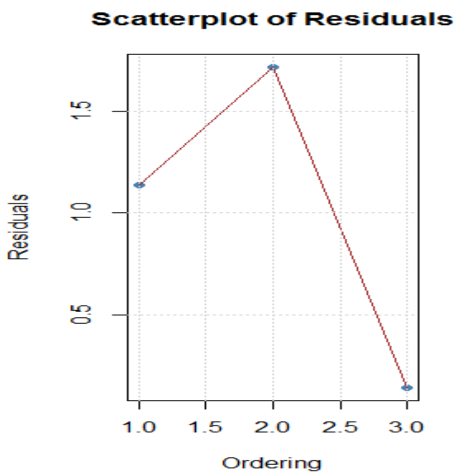


Fig.6(c)

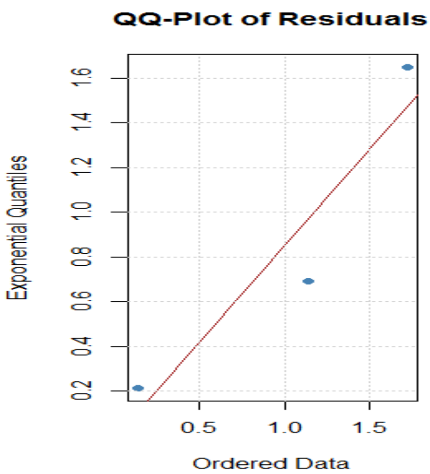


Fig.6(d)

The scatter plot of residuals from a GPD fitted to the data over a threshold of 2.0. The solid line observed is the smooth of the scattered residuals as shown in **Fig.6(c)**. The QQ plot follows a linear form as shown in **Fig.6(d)**; therefore the parametric model fits the data well.

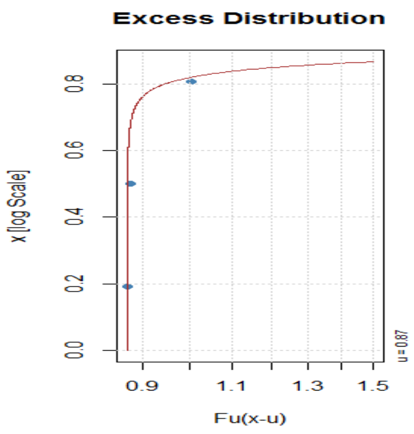


Fig.6(e)

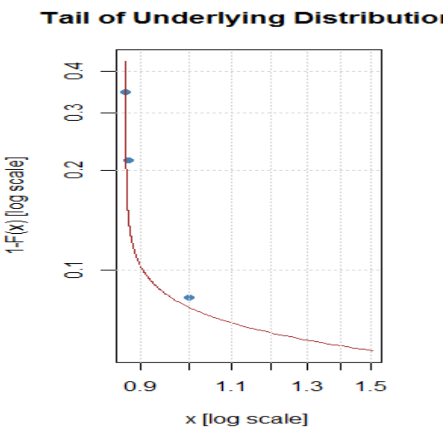


Fig.6(f)

A threshold of 0.87 is chosen, there are 3 out of 7 returns that exceed the threshold. The excesses were fitted to a GPD model using the MLE. The parameter estimates $\xi = 5.2195$ and$\beta = 8.7596$. The shape parameter ξ is greater than 0 implying a heavy tailed distribution. This can be interpreted to mean that the higher the value of the shape parameter, the higher the derived return. The distribution for the excesses shows a smooth curve meaning GPD fit was a good fit for the data similar for the tail of the underlying distribution as shown in **Fig.6(e)** and **Fig.6(f).**

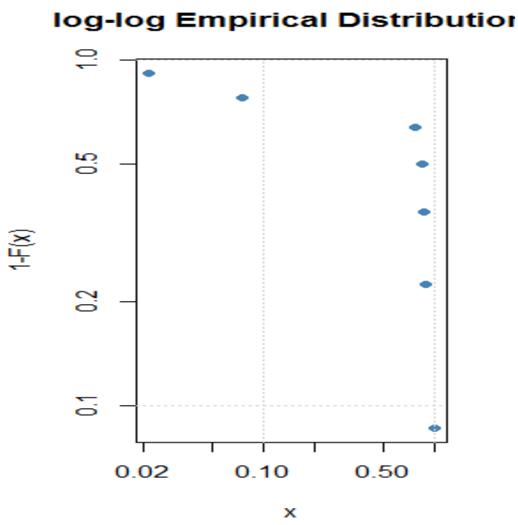


Fig.6(g)

In **Fig.6(g)** the empirical distribution of all the insurance companies average fire loss data fitted defines a CDF consistent with data directly observed in the data set. In other words, it defines a CDF, F(x), such that F(x) is equal to the proportion of data points in the set less than or equal to *x*.

The Value at Risk (VaR) with 5% level of confidence was Kshs.1.25 million. This implies that the coming year's loss for the entire industry would exceed is Kshs.1.25 million. Analogously the same interpretation holds for 1%. In practice when portfolio of loss is known, then precautions can be taken to mitigate against it.

4.0 Conclusion

The results obtained for the entire general insurance industry of Kenyan market show that the distribution for the excesses follows a smooth curve meaning GPD fit was a good fit for the data. Time series plot of the insurance industry on average shows presence of periodicity. Original data was then transformed to standardize it using the following function $\log_{10}(X1+1.05-\min(X1))$ where X1 was the original data. The mean excess plot of all insurance companies' average returns was obtained. It was also observed that the shape parameter is greater than 0 hence the distribution was found to be heavy tailed. The scatter plot of residuals from a GPD fitted to the data over a threshold of 2.0. The solid line observed is the smooth of the scattered residuals. The QQ plot follows a linear form; therefore the parametric model fits the data well. VaR estimate was finally obtained using extreme value method. After the excesses over a high threshold were fitted to the GPD, parameters were estimated which were used to estimate VaR.

References

Alvarado-Celestino, E 1992, *Large forest fires: An analysis using extreme value theory and robust statsistics*, University of Washington, USA.

Balkema, A & DeHaan, L 1974, Residual life at a great age, *Annals of Probability*, 2, 92-804.

Davies, L & Gather, U 1993, The identification of multiple outliers, *Journal of the American Statistical Association*, 88, 782-792.

Davis, L 1965, *The economics of wildfire protection with emphasis on fuel break system*, Division of Forestry, State of California.

Dr. Fotios, C 2005, Empirical Study of Fat –Tails in Maximum Drawdown: *The Stable Paretian Modeling Approach*.

Fisher, R, Tippett, C 1928, *Limiting forms of the frequency distribution of the largest*

Gnedenko, V 1943 Sur la distribution limite du terme d'une série aléatoire. *Annals of Mathematics,* 44, 423-453.

Hall, P 1982, 'On Some Simple Estimates of an Exponent of regular Variation', Journal *of royal Statistical Society London*, 44, 37-42.

Jorion, P 1996, 'Measuring the risk in value at risk', Financial *Analysts journal*; Nov/Dec 1996; 52, 6; ABI/INFORM Global pg.47.

Longin 1999, Optimal margin level in futures markets, Extreme price movements, *Journal of Futures Markets Volume 19 Issue 2*

Mckenzie, D, Peterson, D & Alvarado, E 1996. Extrapolation problems in modeling fire effects at large spatial scales. *International Journal of wildland Fire,* 6, 165-176**.**

Nigm, A, El-Habashi, M & Hamdy, H 1987, The effect of risk parameters on decision making, *Insurance, Mathematics and economics*, 6, 237-244.

Pickands, J 1975, Statistical inference using extreme value order statistics. *Annals of Statistics*, 3, 119-131.

Ramachandran, G 1988, Probabilistic approach to fire risk evaluation. *Fire Technology August*, 204-226.

Strauss, D, Bednar, L & Mees, R 1989. Do one percent of forest fires cause ninety-nine percent of the damage? *Forest Science,* 35, 319-328.

Thomas, D 1989, A fire behavior analysis in Yellowstone Park. In t. Walsh(ed.). *Wilderness and Wildfire*, University of Montana, Missoula, Montana, 50, 11-12.