www.iiste.org

Modeling and Estimation of Market Risk Using Extreme Value Theory

Sheillah Wanjala[1*], George Orwa[1,2], Joel Koima[1,3] and Joyce Jeptoo[1,4]

1. Department of statistics and actuarial sciences, Jomo Kenyatta University of Agriculture and Technology, P.O BOX 62000-00200 , Nairobi, Kenya.
2. Department of statistics and actuarial sciences, Jomo Kenyatta University of Agriculture and Technology, P.O BOX 62000-00200 , Nairobi, Kenya.
3. Department of statistics and actuarial sciences, Jomo Kenyatta University of Agriculture and Technology, P.O BOX 62000-00200 , Nairobi, Kenya.
4. Department of statistics and actuarial sciences, Jomo Kenyatta University of Agriculture and Technology, P.O BOX 62000-00200 , Nairobi, Kenya.

* hswanjala@gmail.com

Abstract

Most commercial airline agencies have suffered losses due to extreme conditions like climatic conditions, price fluctuations and airline problems. One of these airline agencies that have experienced large losses is the Kenya Airways. This study assesses the market risk in the Kenyan stock market using Kenya Airways (KQ) share prices. This paper helps an investor to weigh whether to invest in the company at a particular time or not. Due to the fluctuations and variations of share prices, the Extreme Value Theory has been used to capture the extreme and rare events. Extreme Value Theory provides a well-established statistical model for the computation of extreme risk measures which include; Value at Risk and Expected Shortfall. In this paper, Univariate Extreme Value Theory is applied to model extreme prices of shares for the Kenya airways in Kenyan market. This paper demonstrates how successful Extreme value theory can be applied in predicting future Value at Risk to the share prices. This provides solutions to the problems faced by the investors and the owners of the commercial agency in the market. This paper concentrates on the Peak over Threshold (POT) as the method of estimating Value at Risk. This technique models the distribution of extremes and the distribution of exceedances over a certain high threshold rather than the individual observations. It concentrates on observations that exceed central limits, focusing on the tail of the distribution. Extreme value theory is also applied to compute the tail risk measures at given confidence interval. An overview of the Extreme Value Theory and Peaks Over Threshold Method are also given. This paper also shows that POT is the best method since the modeling the exceedances, they follow a generalized pareto distribution (GPD) which fits the sample losses very well. This technique of extreme value theory helps us to model and estimate the stock market risk of the Kenya airways since it helps us to assess the goodness of fit of the series and detect the outliers and also in estimation of the extreme quantiles.This study used the stock data which was obtained from the Nairobi stock exchange (NSE). With Generalized Pareto Distribution the estimates of value at risk and the expected shortfall for share prices indicate that with probability 5% the daily loss for the stock market prices exceeds 2.5105%, for 1% the daily loss exceeds 2.5330, for 0.5% the daily loss exceeds 2.5358 and for 0.1% the daily loss exceeds 2.5380%. These results can be used to estimate risk measures in the stock market as well as providing insights to the managers of the airline and also as a reference for actual or potential investors in the airline industry.

Keywords: Extreme Value Theory, Peak -Over Threshold (POT), Generalized Pareto Distribution (GPD), Value at Risk (VaR), Expected Shortfall (ES).

1. Introduction

Value at Risk answers the question about, how much one can lose over a certain time horizon with a given probability level. It also summarizes in a single number the overall market risk faced by a financial institution, according to (Jorion, 2007). Expected Shortfall (coherent risk measure) is an expected value of the loss, given that a VaR violation occurred. In other words, Expected shortfall estimates the potential size of the loss exceeding VaR at φ probability level, (Freddy D., 2002). These risk measures are achieved efficiently by use of POT method an approach of extreme value theory.

The foundations of the theory were set by (Fisher and Tippett, 1928) and (Gnedenko,1943), who proved that the distribution of the extreme values of an independent and identically distributed (iid) sample, can converge to one out of only three possible distributions (Frechet, Weibull or Gumbel), (Acerbi et al, 2001).

Many researchers have oriented their work towards more efficient tail-oriented models of risk, namely EVT approach. The superiority of EVT has been extensively demonstrated by many researchers, in fields like insurance or financial risk management. (Embrechts et al,1997) applied EVT approach to assess fat tails of different time series, like hydrologic, insurance and financial data, supported by a very detailed and complex mathematical framework. Similar work is found in (Resnick,2007), who studies extreme events in data networks, finance and insurance. (Ana-Maria and Moisa, 2009) applied exchange rate returns of four currencies against the Euro to analyze the relative performance of several VaR models and Extreme Value Theory. They revealed that in extreme market conditions, extreme measures are needed. (McNeil and Frey, 2000) suggested EVT to estimate the tail related risk measures for heteroscedastic financial time series and they compared this approach to the various other methods for extreme value at risk estimation for financial time series data. They found that the EVT outperform other models for extreme value at risk estimation.

A considerable amount of research has also been dedicated to more specific issues of EVT, for example, tail index and graphical tools of the framework, like mean excess function plot, QQ plots etc. Starting with the work of (Hill, 1975) and (Pickands, 1975), many studies have tried to establish a measure of the tail thickness of fat-tailed distributions.

(Morgan, 1990) started the Value at Risk method as a project in order to assist practitioners to summarize their risk in a single number. He showed that typical Value at Risk approach would add a figure on a probability of losing on more than a certain amount of money over a time horizon. Value at Risk may be estimated using various methods depending on the assumptions made and methodologies used. Extreme Value Theory is a well developed theory in the field of probability that studies the distribution of extreme realizations of a given distribution function, or of a stochastic process, satisfying suitable assumptions.

(Dekkers et al., 1989) improved the Hill estimator and proved consistency as well as asymptotic normality.
(Koima et al., 2013) studied the equities of Barclays bank and used POT and GPD to capture the rare events. In the estimation of extreme quantiles, the distribution of excesses over a certain high threshold was based on Peak-Over-Threshold method which identified the starting of the tail. After the excesses over a high threshold were fitted to the GPD, parameters were estimated which were used to estimate Value at Risk. The point estimates and interval estimates of VaR at 99.9% and 95% confidence intervals for the equities were clearly shown and found to capture the financial risks significantly since this method can estimate VaR outside the sampling interval.

Note that, if by visual judgment it is observed that the required quantiles does not fall in the peak area, we use other appropriate method. That is, a series of iid random variables may be generated from normal or t distribution if it is observed that the required level of α (in this case φ) does fall in appropriate region. A set of quantiles is then fitted on the excesses over the quantiles adjusted random variables (Mwita, 2003).

2. METHODOLOGY

Extreme Value theory is the most efficient method in the estimation of the tail behavior of a distribution. EVT models the tail returns and estimates measures of risks of financial returns.
The research model is a time series of stock prices over a period of more than 10 years. This shows the relationship between the current price and the previous prices of the stock prices. This time series is represented by the following model;

$$Y_t = q(X_t) + e(t) \tag{1}$$

Where Yt is the current price, q(Xt) is the extreme VaR and e(t) is the random error.

2.1 *Generalized pareto distribution (GPD)*

This describes the behavior of large observations which exceed high thresholds. The GPD is usually expressed as a two parameter distribution with a distribution function;

$$G_{\varepsilon,\beta(u)} = \begin{cases} 1-\left(1+\frac{\varepsilon x}{\sigma}\right)^{\frac{1}{\varepsilon}}, if\ \varepsilon \neq 0 \\ 1-\exp\left(\frac{-x}{\sigma}\right), if\ \varepsilon = 0 \end{cases} \quad (2)$$

Where $\varepsilon > 0$ and $x \geq 0$ when $\varepsilon \geq 0$ and $0 \leq x \leq -\sigma/\varepsilon$ when $\varepsilon < 0$.

The GPD again sub sums other distributions under its parameterization. When $\varepsilon > 0$, we have a reparameterized version of the usual pareto distribution with shape parameter $\alpha = 1/\varepsilon$; if $\varepsilon < 0$, we have a type II pareto distribution; if $\varepsilon = 0$, we have the exponential distribution. Adding a location parameter, μ, we extend this family of GPD to $G_{\varepsilon,\mu,\beta(x)}$ which is defined to be $G_{\varepsilon,\beta(x-\mu)}$.

The Pickands Balkema de Haan Theorem

Consider a certain high threshold u which might for instance be the lower attachment point of a high excess loss layer. At any event u will be greater than any possible displacement δ associated with the data. We are interested in excesses above this threshold or in other words the amount by which observations overshoot this level. Let x_0 be the finite or infinite right endpoint of the distribution F , that is to say

$$x_0 = \sup\{x \in \Re : F(x) < 1\} \leq \infty$$

We define the distribution function of the excesses over the high threshold u by

$$F_u(x) = P\{X - u \leq x \mid X > u\} = \frac{F(x+u) - F(u)}{1 - F(u)} \quad (3)$$

For $0 \leq x < x_0 - u$

The (Balkema-de Haan, 1974) and (Pickands, 1975) theorem shows that under the maximum domain of attraction conditions, the generalized pareto distribution is the limiting distribution for the distribution of the excesses as the threshold tends to the right end point, that is we can find a positive measurable function $\beta(u)$ such that

$$\lim_{u \to x} \sup_{0 \leq x < x_0 - u} \left| F_u(x) - G_{\varepsilon,\beta(u)}(x) \right| = 0 iff F \in MDA(H_\varepsilon)$$

This theorem suggests that for sufficiently high threshold u, the distribution function of the excesses may be approximated by $G_{\varepsilon,\beta(x)}$ for some values of ε and β. Equivalently, for $x - u > 0$ the distribution function of the ground-up exceedances $F_u(x-u)$ (excesses plus u) may be approximated by

$$G_{\varepsilon,\beta(x-u)} = G_{\varepsilon,u,\beta(x)} \quad (4)$$

In this case, the choice of a threshold should be large enough to satisfy the conditions to permit its applications as $u \to \infty$ while at the same time leaving sufficient observations for the estimation.

2.2 *Estimation of Value at Risk*

By definition, VaR is the p-quantile of the distribution of the log change in price. EVT makes it possible to model the empirical distribution of the extreme observations. Extreme VaR is defined as the p-quantile estimated from the extreme distribution. when extreme observations follow exactly a GPD distribution, the approximation of excesses over a threshold by a GPD leads to the following estimator:

$$VaR_{extreme} = u + \frac{\hat{\beta}}{\hat{\varepsilon}}\left[\left(\frac{n}{N_u}(1-p)\right)^{-\hat{\varepsilon}} - 1\right] \quad (5)$$

For a given probability $q > F(u)$, the VaR estimate is calculated by inverting the tail estimation formula 5 to get

$$\hat{VaR}_q = u + \frac{\hat{\beta}}{\hat{\varepsilon}}\left[\left(\frac{n}{N_u}(1-q)\right)^{-\hat{\varepsilon}} - 1 \right] \quad (6)$$

In standard statistical language this is a quantile estimate, where the quantile is an unknown parameter of an unknown underlying distribution. It is possible to give a confidence interval for VaRq using a method known as profile likelihood; this yields an asymptotic interval in which we have confidence that VaR lies. The asymmetric interval reflects a fundamental asymmetry in the problem of estimating a high quantile for heavy-tailed data: it is easier to bound the interval below than to bound it above.

2.3 Estimation of ES

Expected shortfall is related to VaR by

$$ES_q = VaR_q + E\left[X - VaR_q \mid X > VaR_q\right] \quad (7)$$

where the second term is simply the mean of the excess distribution $F_{VaR_q}(y)$ over the threshold VaR_q. Our model for the excess distribution above the threshold u has a nice stability property. If we take any higher threshold, such as VaR_q for $q > F(u)$, then the excess distribution above the higher threshold is also GPD with the same shape parameter, but a different scaling. It is easily shown that a consequence of the excess distribution model becomes;

$$F_{VaR_q}(y) = G_{\varepsilon,\beta+\varepsilon(VaR_q-u)}(y) \quad (8)$$

The beauty of eqn. 7 is that we have a simple explicit model for the excess losses above the VaR. With this model we can calculate many characteristics of the losses beyond VaR. By noting that the mean of the distribution in eqn.8 is $(\beta + \varepsilon(VaR_q - u))/(1-\varepsilon)$, we can calculate the expected shortfall.
We find that

$$\frac{ES}{VaR_q} = \frac{1}{1-\varepsilon} + \frac{\beta - \varepsilon u}{(1-\varepsilon)VaR_q} \quad (9)$$

It is worth examining this ratio a little more closely in the case where the underlying distribution has an infinite right endpoint. In this case, the ratio is largely determined by the factor $1/(1-\varepsilon)$. The second term on the right hand side of eqn.9 becomes negligibly small as the probability q gets nearer and nearer to 1. This asymptotic observation underlines the importance of the shape parameter ε in tail estimation. It determines how our two risk measures differ in the extreme regions of the loss distribution.
Expected shortfall is estimated by substituting data-based estimates for everything which is unknown in eqn.9 to obtain

$$\hat{ES} = \frac{\hat{VaR}_q}{1-\hat{\varepsilon}} + \frac{\hat{\beta} - \hat{\varepsilon}u}{1-\hat{\varepsilon}} \quad (10)$$

3 RESULTS AND DISCUSSION

3.1 *Estimation of extreme Quantiles and Value at Risk*

In this section MATLAB is used in data analysis. The estimation of the distribution of excesses over a certain high threshold using the POT approach which actually identifies the starting point of the tail is demonstrated. The data used is stock market data in particular the share prices of the Kenya Airways consisting of 2863 observations.
Logarithmic transformation was performed on the raw data to obtain logarithmic returns which then was applied in data analysis. Figure 1 below shows logarithmic returns plotted against time.

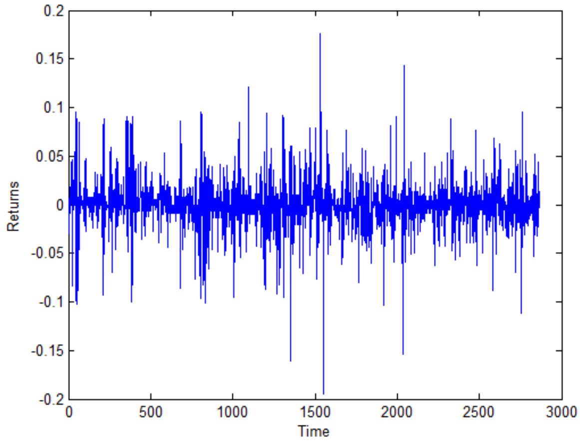

Figure 1: Daily Negative logarithmic Returns

After logarithmic transformation, the mean of the data became stationary. From figure 1 it is observed that there is presence of values clustering. This actually means that small values follows small values and similarly large values follows large values which indicates presence of short-range dependence, resulting into doubting the assumption independent and identically distributed returns that may be violated due to the observed clustering. The presence of clustering of returns indicates the presence of stochastic volatility.

Modeling the tails of a distribution with GPD requires observations to be approximately independent and identically distributed (iid). To produce a series of iid observations, GARCH-type stochastic volatility model is fitted to the returns and the filtered residuals and volatilities are extracted from the returns of the share prices. The residuals obtained then are standardized to have a Zero mean, unit-variance and also iid of which are now used in EVT estimation of the sample CDF tails. Value at Risk can now be obtained using equation (6).

First, the QQ-plots and the distribution of mean of excesses are examined. Most financial data series are fat tailed and therefore the graph makes it possible to assess their goodness of fit to the parametric model. The fat-tailness of a distribution is confirmed by the use of QQ-plot, which should be concave in nature to indicate a fat-tailed distribution. This can be shown in figure 2.

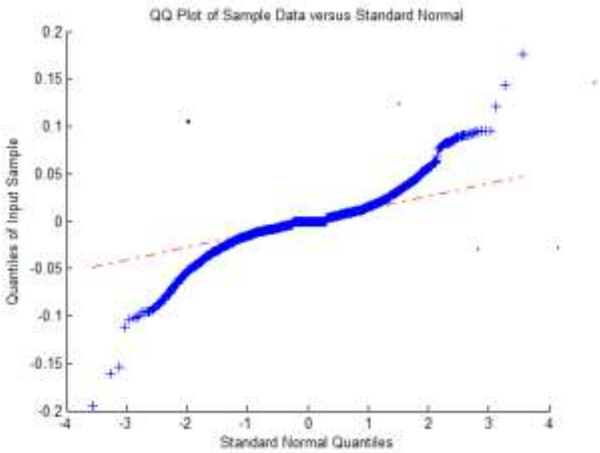

Figure 2: QQ-plots

From figure 2, it is shown that the graph is curved to the top at the right end or to the bottom at the left end indicating that the empirical data is fat tailed.

Figure 3 shows a plot of sample mean excesses of a stock data against different thresholds. This is a plot of some few sample mean excesses. This is because plotting all the excesses affect and finally distorts the plotting. A threshold is chosen observing the area with a linear shape on the graph straightening upwards. It is observed that the graph begins to straighten upward around threshold 2 and also the function tends to infinity like a GPD, which provides a reasonable fit to the whole data set. The chosen threshold is 2.00 meaning that from the actual data, 70 out of 2863 data points exceed the threshold.

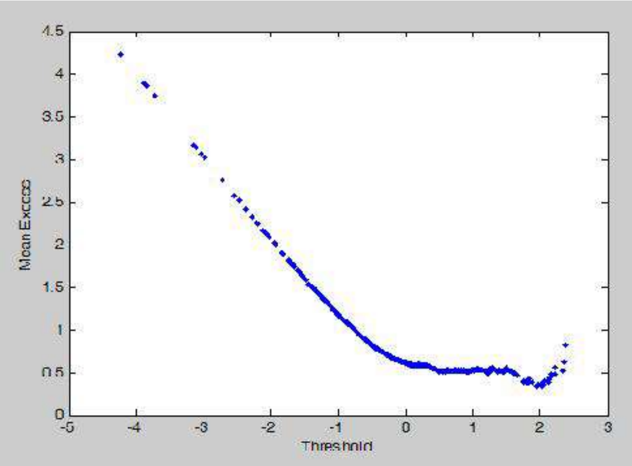

Figure 3: Mean excess plot

3.2 ***Estimations at the Tail of the Distribution***

The 70 exceedances over high threshold are fitted to the GPD using Maximum Likelihood Estimation (MLE). The quantile (percentile) value at the tail from estimated parameters of different distributions is estimated. Basing on these data, the parameter estimates are ε = 0.3582 and β =0.2212. The shape parameter ε > 0 is an indication of a heavy-tailed distribution. This means that, the higher the value of the shape parameter the heavier is the tail and the higher the derived quantile estimates and the corresponding prices. At this point quantile estimates and confidence intervals for a high quantiles above the threshold in a GPD are calculated.

Figure 4 shows a smooth curve which is the estimated GPD Model for the excess distribution. The points representing the emperical distribution of 70 extreme values shows that GPD model fits the excess losses well.

In figure 5, y-axis indicates the tail probabilities $1-F(x)$. The Threshold of 2.00 corresponds to a tail probability of approximately 0.05 which can be observed from the top left corner of the graph. The solid curve indicates the tail estimation which can be extrapolated to areas of sparse data and they also show the 70 large losses.

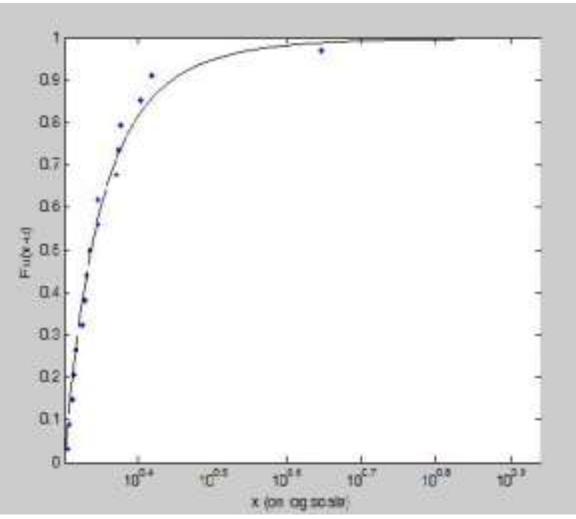

Figure 4: Distribution of exceedances

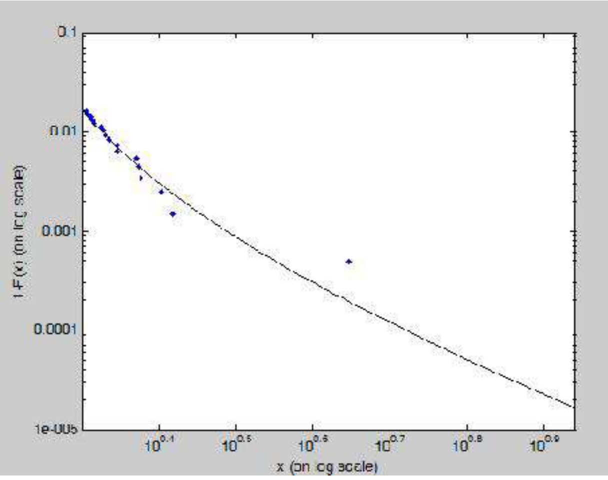

Figure 5: Tail estimate

Figure 6 shows the scatter plot of residuals from a Generalized Pareto Distribution fitted to stock market data over a high threshold 2.The solid line observed is the smooth curve of the scattered residuals.

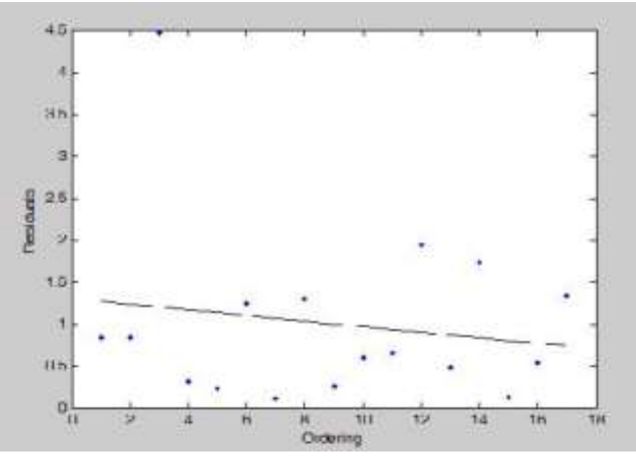

Figure 6: Scatter plot of residuals

In Figure 7, Value at Risk is esimated to be VaR0:999=3.0726 which is calculated using equation 6 and is statistically referred to as a quantile estimate. The figure shows the point and interval estimation of VaR at the 0.999^{th} quantile and 95 percent confidence intervals for the stock data. Confidence interval of VaR yields an asymptotic interval where VaR lies and these demonstrate a fundamental asymmetry in the estimation of a high quantile for heavy-tailed data. Where the vertical red dotted line intersects with the tail estimate gives the point Value at Risk of 3.0726. The left vertical dotted blue line shows the lower confidence level which is 2.5822 and the right vertical dotted blue line shows the upper confidence level which is 6.6548 for the Value at Risk and the Horizontal red thick line corresponds to the 99.9% confidence level. The advantage of estimating VaR using GPD method is that, it can estimate VaR outside the sampling interval.

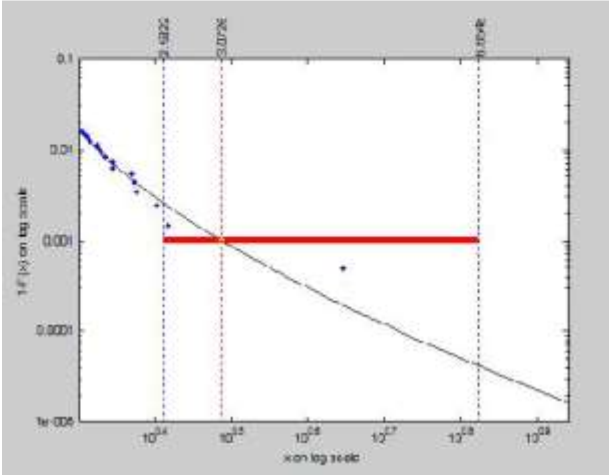

Figure 7: Point estimate at the tail

3.3 *Computing VaR and ES using Extreme Value Theory Method*

Using the threshold of 2, number of observations above the threshold as 70, total number of observations as 2863, Shape parameter as 0.3582, Scaling parameter as 0.2212 and varying confidence level, the results may be obtained as follows.

Level of confidence	0.95	0.99	0.995	0.999
ε	0.3582	0.3582	0.3582	0.3582
β	0.2212	0.2212	0.2212	0.2212
VaR_q	2.5105	2.5330	2.5358	2.5380

Table 1: Point Estimates for VaR and ES

Looking at the VaR from table 1, with 5% confidence level the tomorrow's loss (left-tail) for the stock market prices to exceeds 2.5105. Analogously the same interpretation holds for 1%, 0.5% and 0.1%. Practically, when the stock price loss is known, then precautions can be taken to diminish it.

4.Conclusion

VaR estimates were obtained using Extreme Value Theory method. In the estimation of extreme quantiles, the distribution of excesses over a certain high threshold was based on Peak-Over-Threshold method which identified the starting point of the tail. After the excesses over a high threshold were fitted to the GPD, parameters were estimated which were used to estimate Value at Risk. The point estimates and interval estimates of VaR at 99.9% and 95% confidence intervals for the share prices were clearly shown and found to capture the financial Risks significantly since this method can estimate VaR outside the sampling interval.

References

Acerbi,C., Nordio C. and Sirtori C.(2001).Expected shortfall as a tool for financial risk Management. Working Paper.

Dekkers, M., Einmahl J., and de Haan L. (1989), A Moment Estimator for the Index of an Extreme-Value Distribution, *The Annals of Statistics*, 17, 1833-1855.

Embrechts, P., Kluppelberg C., and Mikosch T. (1997), Modelling Extremal Events for Insurance and Finance, Springer-Verlag, Berlin.

Fisher, A and Tippett (1928) Limiting forms of the frequency distribution of the largest or smallest member of a sample. Proceedings of the Cambridge Philosophical Society 24:18090.

Freddy Delbaens (2002) Probability and Finance. *Journal of the American Statistical Association* Vol (97), no. 459, p. 923)

Gnedenko, V. (1943). Sur la distribution limite du terme maximum dune serie aleatoire, *Annals of Mathematics* 44, 42353.

Hill, M. (1975), A Simple General Approach to Inference about the Tail of a Distribution, *the Annals of Statistics*, 3, 1163-1174.

Jorion. P (2007), Value at Risk, the new benchmark for managing financial risk, third edition, McGraw-Hill.

Koima J., Mwita P. and Nassiuma D. (2013), Application of Extreme Value Theory in the estimation of Value at Risk in Kenyan stock market, Working paper, Jomo Kenyatta University of Agriculture and Technology.

McNeil, J and Frey, R (2000) Estimation of the tail related risk measures for heteroscedastic Financial time series: An extreme value approach, *journal of Empirical Finance*, 7(3-4), 271-300

Mwita, P. (2003). Semi-parametric Estimation of conditional quantiles for time series with application in finance. Ph.D Thesis, University of Kaiserlautern.

Pickands, J. (1975), Statistical Inference Using Extreme Order Statistics, *The Annals of Statistics*, 3, 119-131.

Resnick, S. (2007), Heavy-Tail Phenomena Probabilistic and Statistical Modeling, Springer, New York, 73-114.

Morgan,M. and Henrion,M (1990), Uncertainity: A guide to dealing with uncertainity in Quantitative Risk and Policy Analysis, *Cambridge University Press*, New York.