International Journal of Mathematics and Physical Sciences Research ISSN 2348-5736 (Online)
Vol. 6, Issue 2, pp: (10-20), Month: October 2018 - March 2019, Available at: www.researchpublish.com

MODELLING OF LARGE INSURANCE CLAIMS USING EXTREME VALUE THEORY: A CASE STUDY OF KENINDIA ASSURANCE COMPANY LIMITED MOTOR BUSINESS

[1]SIMON KINYUA WERU, [2]PROF. A. WAITITU

[1]A student in Master in Applied Statistics of Jomo Kenyatta University of Agriculture and Technology
[2]Senior Lecturer in the College of Pure and Applied Science of Jomo Kenyatta University of Agriculture and Technology

Abstract: **Due to rare occurring of events in insurance companies in Kenya, this has directly affected the industry resulting to huge losses. Further, it has discouraged most insurance companies to offer cover to some listed risks that seem risky and likely to occur causing massive impacts. In particular, the motor cover which out of 47 insurance companies in Kenya, less than a half of this number offer the cover. These huge claims in insurance, finance and even metrological assume distribution with fat tails. This study concentrates on the right tails of the underlying distribution (extremely large observations), and specifically situations when EVT is assumed to be positive. A case study of Kenindia Assurance claims data illustrates how data sets can be analyzed in practice. It is shown to what extent discretion can/should be applied, as well as how different estimators can be used in a complementary fashion to give more insight into the nature of the data and the extreme tail of underlying distribution. The analysis is carried out from the point of raw data, to the construction of tables which can be used directly to gauge the risk of insurance portfolio over a given time frame. Using R, the obtained the mean excess plot was obtained which helped in measuring the shape of the distribution in the tail. The fitted a GPD over a chosen threshold to help us in measuring the shape of the underlying distribution.**

Keywords: **MODELLING, CLAIMS, EXTREME VALUE THEORY.**

1. INTRODUCTION

In one's lifetime, extra ordinary events arise that have never been experienced before. These events cause great impact to our lives, and are likely to occur again in future. For example, the Westgate terror attacks in 2013, China floods in 1931 and the 2004 Indian ocean tsunami. These events do not only cause great damage and loss of lives, but also make a great impact in other areas like finance and Insurance. These extreme events are widespread throughout all sectors, that is in weather, financial markets and also insurance companies.

Most Insurance and reinsurance companies have over the years tried to compile and analyze data to enable handle these catastrophic events, (for example, the1974 Munch Reinsurance Company (Munich Re). The data analyzed by Munich Re was clearly showing that the number of catastrophic was increasing over the years. It was clear that most companies had under estimated these risks and the potential financial effects of these extreme events hence facing massive losses when the claims occurred.

It is evident that these insurance companies cannot correctly assess the potential losses that can arise due to these extreme rare events, and hence put necessary reserves and purchases appropriate level of reinsurance contracts. Failure of these companies to do so lands them in major financial risks incase these rare events occur. The analysis and modeling extreme events, especially when assessing the nature of possible future extreme than those already observed, is therefore an important task.

Extreme value theory is one of the applied methods in mathematics to model these rare cases. This method aims at modeling the full behavior of a distribution explicitly, where extremes occur. (Embrechts, 1997) and (Beiriant, 2005), gives a good introduction to extreme values theory (EVT), they also go ahead to provide statistical aspects and applications of extreme value theory. Our main reference is (Embrechts, 1997) which emphasizes more on application of EVT to insure and finance.

The research problem:

Time series data of real world events like exchange rate or weather patterns are inherently non-stationary in nature that is, a random process whose statistical patterns vary with time. This may be due to changing business cycles in financial markets. The nature of these events is rare in business world and cause huge losses and even fall of financial institution and insurance companies.

When considering insurance companies, it is evident that large losses arising from natural catastrophes are becoming more frequent and more severe over time. In general, large losses in insurance may become more or less severe over time (Charez-Demoulin and Embrects, 2004). It is therefore important to account for non-stationary when attempting to model these extremes.

The standard practice of using normal distribution techniques to model non-stationary is increasingly becoming uncommon (Coles, 2001). This is because it does not incorporate smoothing methods. The use of non-parametric methods allows the smoothing methods and therefore it is seen as more preferable. Another advantage is that of assumptions are made on the data. In addition, it is the preferable technique when dealing with unexpected results.

The aim of this study was to try to approximate these rare events in the insurance industry. This in turn assists the company in reserving, product pricing, investments and product development. Hence, protecting the company against future collapse due to large losses.

Aim of the study:

The uncertain future is the core and only reason people, companies or even organizations take insurance cover. This is a way of spreading risk if the happens in future. Pure underwriting during undertaking of risk has resulted to not only large losses in insurance companies but also to their collapse. Lack or poor reserving in insurance industries has also been a contributing factor in the fall of these insurance and financial institutions.

In this proposal, the aim is to attempt to predict the future uncertain events that could result to large losses. This helped the company in reserving, premium determining, product pricing as well as the taking correct reinsurance policy. The task was to determine whether these events can occur and what would be the resulting losses

2. LITERATURE REVIEW

Introduction:

The theory of extreme dates back to 1709 when N. Bernoulli discussed the mean largest distance from the origin when n points lie at random on a straight line of length t (Johnson, 1995). Centuries later, Fourier in normal case studied the same using probabilities (Kinnson, 1985). The study of extreme events was first studied by the astronauts who were facing great challenges either utilizing or rejecting observation that appeared to differ greatly with the other set of data.

EVT was widely used in meteorological departments in the analysis of floods, rainfall and earthquakes. Most of the work in this field was mainly done focusing on the behavior changes in weather patterns.Major works in Extreme Value Theory (EVT) started in early 20's when (Tippet, 1925). presented tables of largest values and corresponding probabilities for various sample sizes from a Gaussian distribution (normal) and the mean range of such samples. This was later followed by publications concerning applications of extreme value statistics. (Gumbel, 1941).Being the first one to study the application of EVT.

The application of EVT in statistics started late in 1940, EVT provides a family of natural models for extreme phenomena, i.e. the assessment of catastrophic or extreme events. On the application front EVT finds wide utility. (Neftci, 2000). Compared VaR based on normal and extreme value distribution. (Gilli and Kellezi, 2003).also advocate EVT, Block maxima and Peak over Threshold (POT) to compute tail risk measures. (Embrechts, 1997).modeled rare events in insurance and other quantitative finance aspects using EVT. Extending the concept of EVT to insurance industry, (McNeil, 1996).used the Danish insurance data to highlight the relevance of Generalized Pareto Distribution (GPD), which is a sub class of GEV, for EVT. He also dealt with the parameter estimation and curve fitting for modeling rare

historical losses in non-insurance sector. Further, he dealt with the concept of loss severity and showed how to model the aggregate payments depending on the number of losses. He employed the method of maximum likelihood estimation (MLE) as well as probability weighted method (PWM) of moment for parameter and found that GPD is the best fit distribution for extreme values.

The most important aspect in modeling with GPD is the parameter estimation and curve fitting. (Jenkinson, 1955).And (Prescott and Walden, 1980).dealt with the estimation of GEV parameters. (H.W.Wainana and A.G.Waititu, 2014).dealt with the analysis of fire insurance claims in Kenyan companies using the Generalized Pareto Distribution. They found out this method unlike VaR did not allow for any assumption on the original distribution of all the observations. (Harmantzis*et al*, 2005). and (Marinelli et al, 2006). also discussed the performance of extreme value theory in VaR and expected shortfall estimation. They compared to the Gaussian and historical simulation models together with other heavy tailed approach. Their result was evident that fat tailed models can predict risk more accurately than nonfat tailed ones and there exists the benefits of EVT framework especially using GDP methods

Research Gaps:

Though major research focusing on domestic policies, there still remain notable gaps that up to now have not been fully captured or exploited. (H.W.Wainana and A.G.Waititu, 2014).studied the effects of these extreme events in fire insurance. It is worth noting that there are more claims from motor than fire which over the years which have been ignored in Kenya being termed as unpredictable, but it is evident that the losses caused by these claims are vital. The study attempted to find the amount of these losses and the effects the claims in motor can cause. This also tried to solve the question on the affordability of motor covers in insurance field.

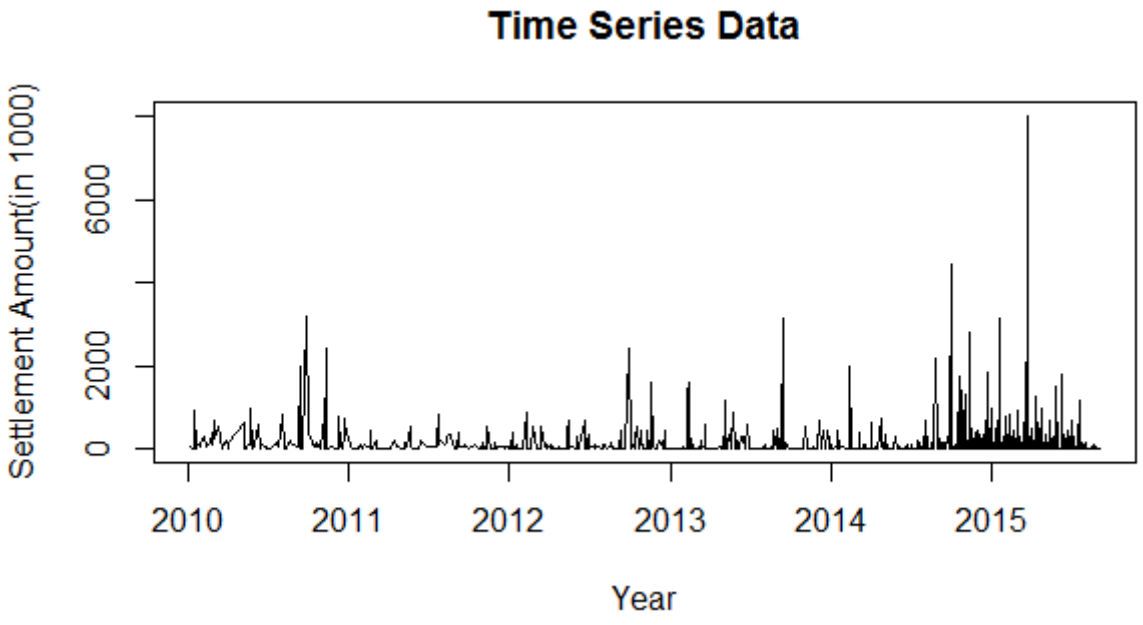
3. DATA ANALYSIS AND PRESENTATION

Introduction:

The data study generated both qualitative and quantitative data. Descriptive analysis method was applied to analyze both quantitative and qualitative data. Data obtained were processed through editing and encoding and then entering the data into a computer for analysis using descriptive statistics with help of statistical; packages R, which offers extensive data handling capabilities and numerous statistical analysis procedures those analyses small to large data statistics (bell 2007). Filling a generalized Pareto distribution and choosing the correct threshold, attempted to generate time series plots, mean excess plots and excess distribution plots.

Descriptive analyses are important since they provide the foundation upon which then correlation and experimental studies emerge. They also provide clues regarding the issues that should be focused on leading to further studies (Mugenda and Mugenda 2003). This summary is to be represented inform of pilots that are user friendly and easy to understand.

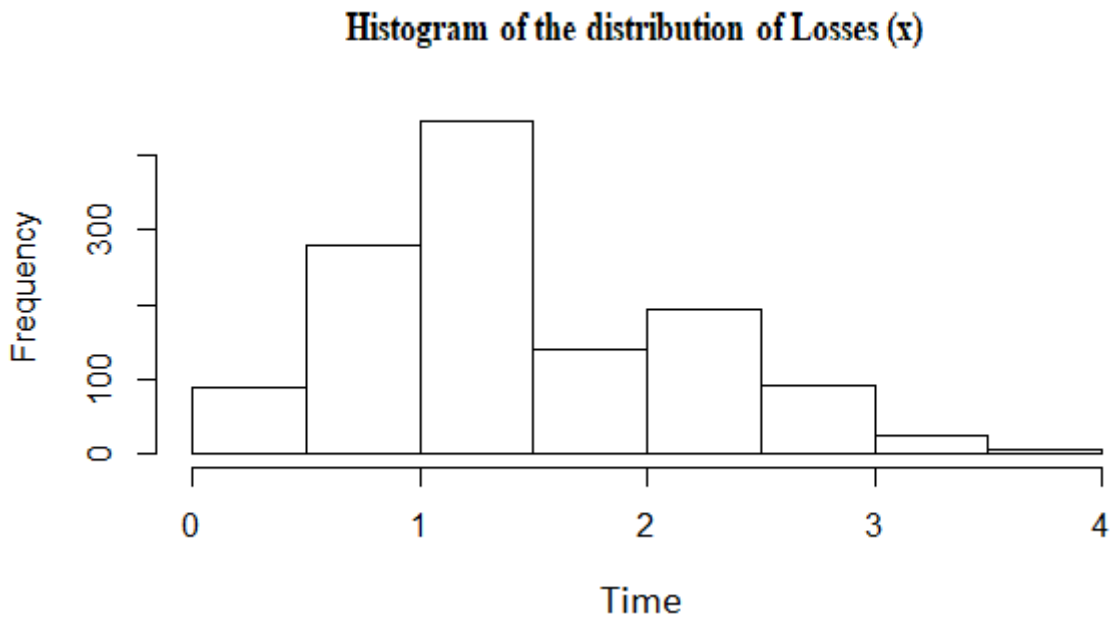
Time Series plot



This is a plot of the time a claim was made and the resettlement amount that was paid. The plot shows evidence of seasonal variations with high settlements at the end of each year. There is a general increase in trend in the amount that is paid for claims over the years, however the amount of settlement is unpredictable in nature. There is an upsurge for resettlement and number of claims in the year 2015.The year 2015 has an outlier whose settlement amount is over eight million shillings. There is a number of isolated extreme occurrences as evidenced in the plot.

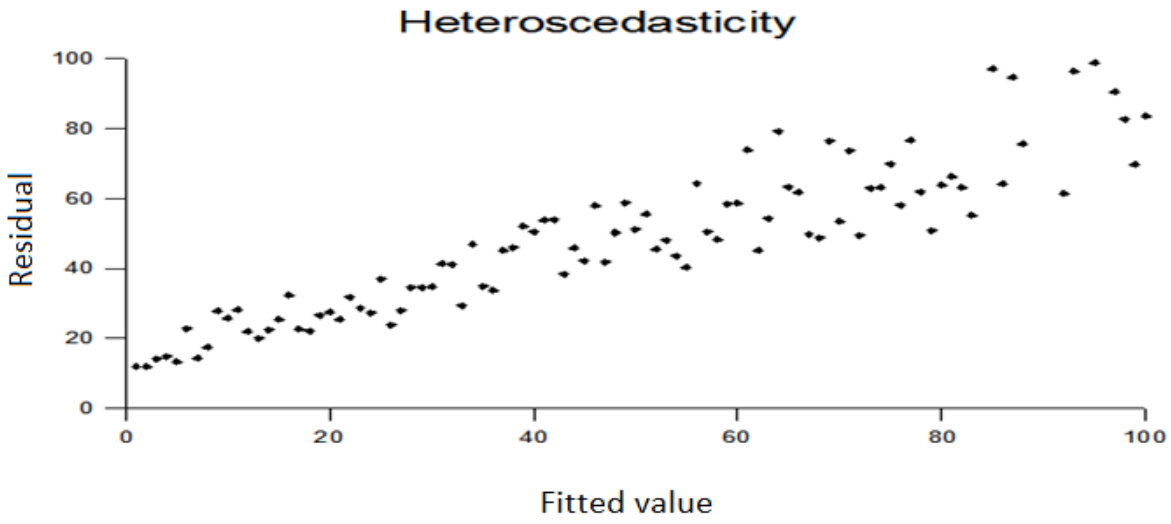
The claim is not stable and as we can see it graduates from 2010 to 2015 with notable years where there are extreme case of claims (outliers) being 2011,2013 and hugely 2015 this thus makes the insurance company have a hard time to actually predict the trend of these claims but it gives them room to tighten their policy claims payments and claims investigation.

Histogram of the distribution of Losses (x)

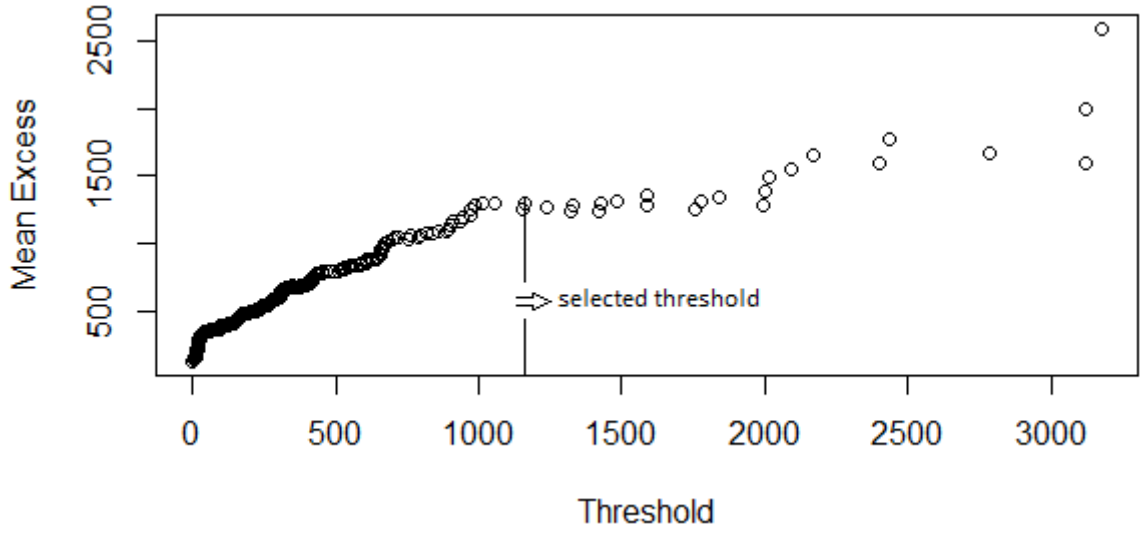


Standardizing the data we clearly see that the data is skewed to the right and the data is heteroscedastic in nature this gives the insurance a hard time to predict and simulate future expected losses and returns for the company for they operate on the extreme values of making a profit or a loss this thus is risky to the organization for in case of lose a huge loss they might not be able to meet their obligations as an insurance company.

Standardized simulation of extreme losses:



A plot of mean excess against the threshold:

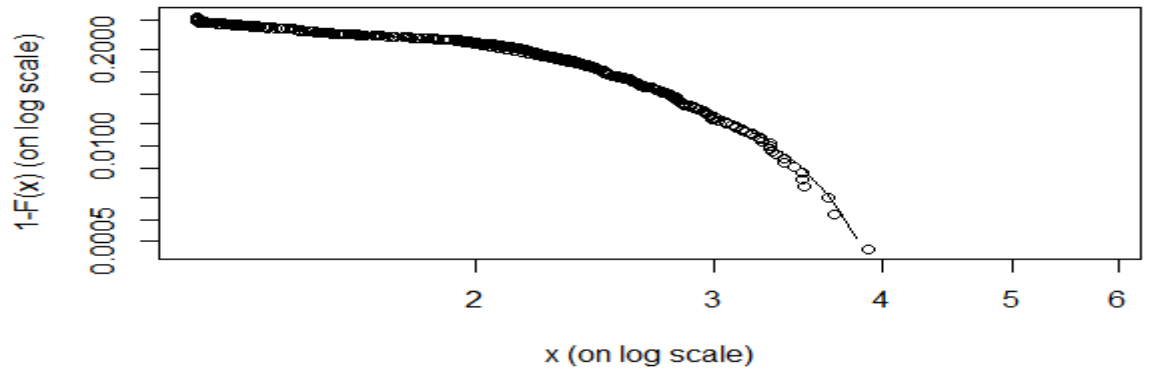


This is a mean excess plot of the monthly settlements. The tails indicate the excess settlements. We can therefore estimate the value at risk and expected shortfall from the plot.

A threshold of 1200 seems to be reasonable enough is chosen from the mean excess plot, and the data was fitted to a GPD model using Maximum Likelihood Estimate. The parameter estimates = 0.6814555 and =2.2753978. The shape parameter is greater than 0 implying 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 GDP fit was a good fit for the data.

Plot of Tail of Underlying Distribution:

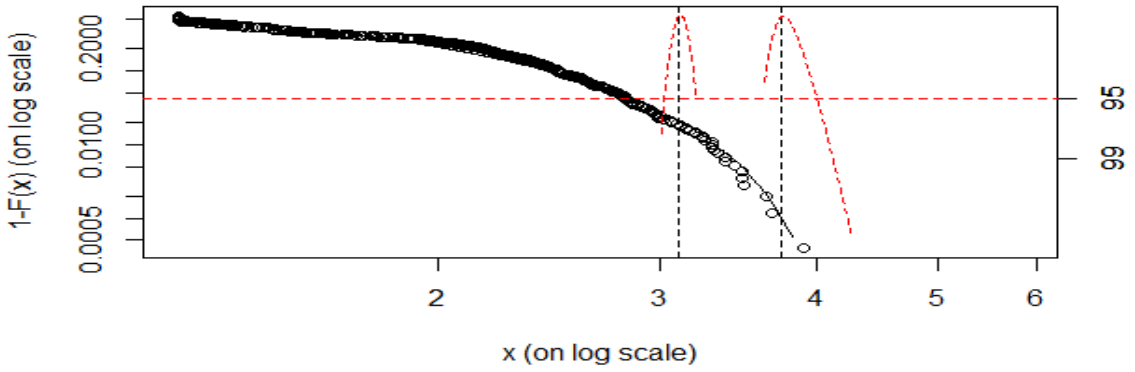
The plot shows empirical distribution of the data where (x) represents the frequency of losses per given epoch.



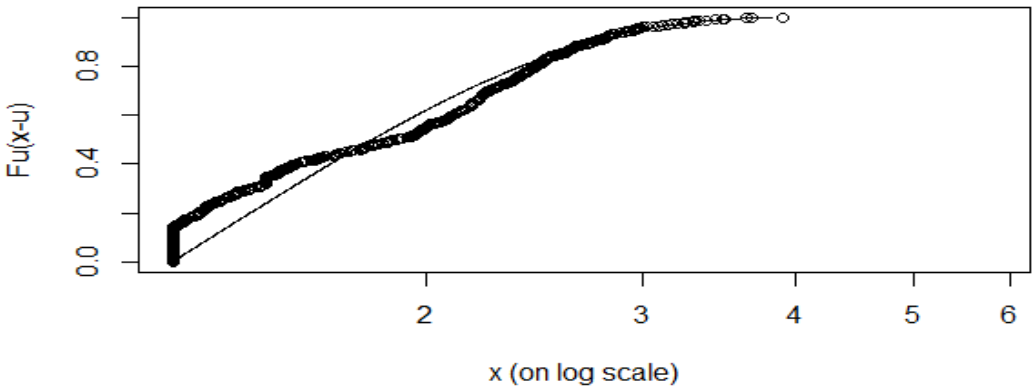
The plot shows the empirical distribution of settlements. The tails are nonlinear, implying the Pareto behavior. Extreme value theorem can therefore be confirmed. Some claims are small and some claims are large. There is no upper limit in the claim sizes. Mathematically speaking this means that the insurance data follows a heavy tailed distribution. The extreme value distributions are known to be heavy tailed. The choice of the priority limit depends only on the large claims. This means that the company is interested in the claims occurring in the tail of a given distribution. There is a possibility that large insurance claims follow the tail of an extreme value distribution.

Plot of the exceedances with confidence intervals:

For tail is approximately linear, implying the Pareto behavior, hence we are justified in fitting a generalized Pareto distribution to the tails.

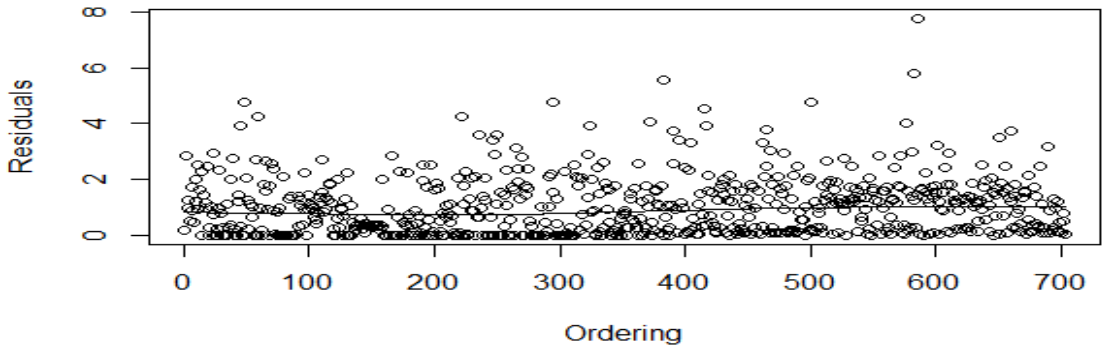


Plot of Excess Distribution:



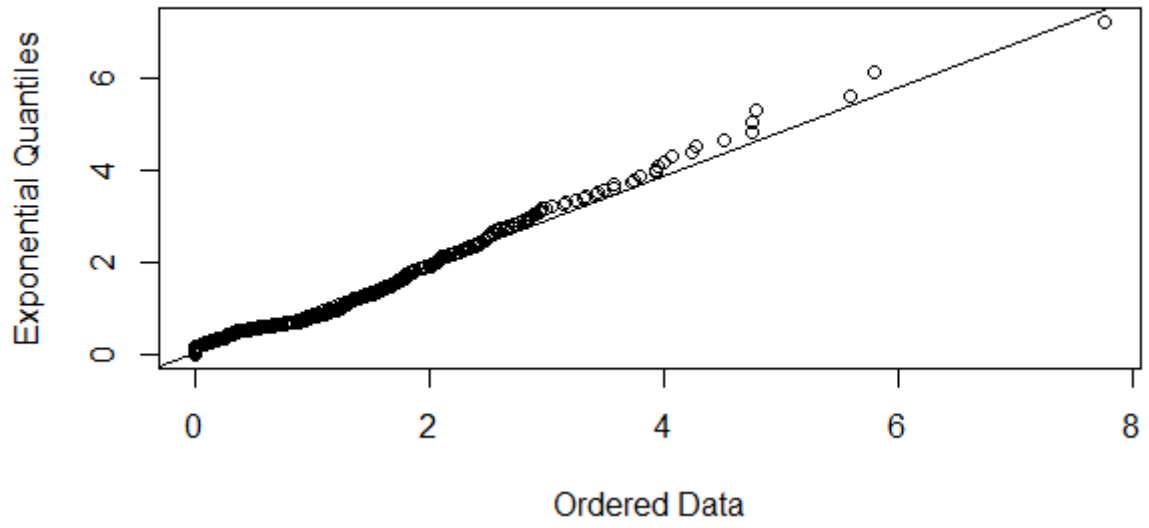
This is a plot of excess distribution, the excess distributions lie close to the theoretical curves and therefore a generalized Pareto linear distribution can describe the tails. With adjusted historical loss data. which we assume to be realizations of independent, identically distributed, truncated random variables, we attempt to find an estimate of the truncated severity distribution One way of doing this is by fitting parametric models to data and obtaining parameter estimates which optimize some fitting criterion - such as maximum Likelihood But problems arise when we have data that are to the extreme and just happen as a single instance like the huge loss that occurred in 2015.

Scatter plot of Residuals



The scatter plot of residuals does not depict any visible pattern to indicate independence of the exceeding distributions since they are randomly distributed. A residual distribution such as that in Figure above showing a trend to higher absolute residuals as the value of the response increases suggests that one should transform the response, perhaps by modeling its logarithm or square root, etc., (contractive transformations). Transforming a response in this fashion often simplifies its relationship with a predictor variable and leads to simpler models.

Q-Q plot of Residuals:



The Q-Q plot of residuals depict points that are approximately linear, we can therefore justify the extreme value theorem. In most of the VaR methods, the approximation by a normal distribution remains a basic assumption. However, most financial series are fat-tailed. The graph of the quantiles makes it possible to assess the goodness of fit of the series to the parametric model.

Given that the parametric model fits the data well, this graph must have a linear form. Thus, the graph makes it possible to compare various estimated models and choose the best. The more linear the Q-Q plots, the more appropriate the model in terms of goodness of fit. In addition, the original distribution of the data is more or less known, the Q-Q plots can help to detect outliers;

It makes it possible to assess how well the selected model fits the tail of the empirical distribution, which was to determine that if huge claims happen was the insurance company go under or stay afloat.

4. DISCUSSION OF RESULTS AGAINST THE OBJECTIVES

Maximum future loss:

A GDP model was fitted and further its quintiles above the threshold obtained. Since our main goal in this objective was to determine the future maximum loss as a result of huge claims. We obtained the boundaries of the Expected Shortfall (ES). The estimate was 3,105,800 with lower and upper limits as 3,026,980 and 3,206,190 respectively. These are the limits in which we can expect a huge future claim having selected a threshold of 1,200,000. Hence the company can be able to organize its reserves for a future loss that might arise.

Most probable time when these claims arise:

Having determined the expected shortfall or the expected future huge claim, we are tasked with determining or approximating what time of the year these claims will arise. Insurance companies need not to reserve huge amounts of money for the expected claims since they also require it for investments and day to day running of the organization. To achieve this, we plotted a time series plot of claims against time for the period under the study. The observation from the plot was that most of these claims arise towards the end of the year. It is therefore advisable for the insurance company to ensure it reserves adequately starting from the third quarter.

5. CONCLUSION AND RECOMMENDATIONS

Introduction:

Extreme value theory is widely used in insurance financial markets, it is a tool that can be used to model probabilities of extremely rare events, it describes the behavior of maxima and minima in a time series. This theorem fits our model since we are analyzing rare events, which are accidents. VaR can be defined as the maximum loss of a portfolio such that the likelihood of experiencing a loss exceeding that amount, over a specified risk horizon, is equal to a pre-specified tolerance level. ES measures the mean of losses that are equal to, or greater than, a corresponding VaR value.

Extreme Value Theorem:

Detailed analysis is carried out to model Motor industrial insurance returns of Kenindia Assurance Company Limited. The presence of extreme values was tested using Chi-square test, Anderson Darling, Cravon Misses test and the data sets were found to contain extreme values. The empirical findings in this paper reveal that Extreme Value Theory method of calculating VaR outweighs the other two methods of estimation, as EVT is known for its ability to model the tail area of the distribution much better. It is practically impossible to have negative VaR which is observed throughout in Historical Simulation method, The two other methods of estimation seem to perform well with high profile data, but EVT approach is able to model data irrespective of the number of observations.

Kenindia Assurance Company Limited companies is known to be in the business of investing premiums of many subscribers and only pay claims to those who have weighty losses from these premiums. A study such as this would help Kenindia Assurance Company Limited company to be able to estimate equitably reliably the amount of loss for a given number of customers within a specific time. It would also enable Kenindia Assurance Company Limited to put certain measures in place by employing adequate risk management measures so as to have minima loss and optimize profit.

The study would equally help subscribers/consumers of insurance products in Kenya to have a good understanding of the ability of insurance companies to pay claims, as most Kenyan citizen shun insurance products because of past defaults in paying claims. In the light of the conclusion, the following are hereby recommended:

- Insurance companies should always carry out thorough investigation about claims before they are paid; as some subscribers may be trying to "play smart".
- All Kenyan Insurance companies should install vehicle tracking device in the vehicle of their clients so as to curb vehicle theft as claims due to vehicle theft as captured in the report is outrageous.
- Consumers of insurance products should find out the integrity of any insurance company before subscribing into that company, particularly the newly established ones that are yet to have a proper registration.
- Policy makers should further enforce strict law guiding claims payment so as to further gain confidence of potential clients.
- Kenyan government should endeavor to further provide substantial support for insurance companies such as we have in America, United Kingdom and other developed nations so as to boost the state of the economy.

REFERENCES

[1] A.G.WaitituH.W.Wainana (2014). Modeling Insurance Returns with *Extreme Value Theory*.

[2] A.J. McNeil and R. Frey. Estimation of Tail-Related Risk Measures for Heteroscedastic Financial Time Series: An Extreme Value Approach. Journal of Empirical Finance, 7(3-4):271–300, 2000.

[3] Castillo, E., Hadi, A. S., Balakrishnan, N. and Sarabia, J. M. (2005).ExtremeValue and Related Models with Applications in Engineering and Science, John Wiley & Sons, Inc., Hoboken, New Jersey.

[4] Embrechts, P., S. Resnick, S. and G. Samorodnitsky (1999). Extreme Value Theory as a risk management tool.

[5] Fisher, R.A. &Tippett, L.H.C. 1928. On the estimation of the frequency distributions of the largest or smallest member of a sample. *Proceedings of the Cambridge Philosophical Society*, vol.24, pp.180-190.

[6] FragaAlves, M.I. 2002.*Estimation of first and second order parameters in heavy tails*. Technical report. Lisbon: University of Lisbon.

[7] FragaAlves, M.I., Gomes, M.I. & De Haan, L. 2003. A new class of semi-parametric estimators of the second order parameter. *PortugaliæMathematica*, vol.60.

[8] Geluk, J., De Haan, L., Resnick, S. &Starica, C. 1997. Second-order regular variation, convolution and the central limit theorem. *Stochastic Processes and their Applications*, vol.69, pp.139-159.

[9] Gnedenko, B.V. 1943. Sur la distribution limite du terme maximum d'unesériealéatoire.*Annals of Mathematics*, vol.44, no.3, pp.423-453.

[10] Hill, B.M. 1975. A simple general approach to inference about the tail of a distribution.*Annals of Statistics*, vol.3, pp.1163-1174.

[11] Karamata, J. 1930. Sur un mode de croissancerégulière des fonctions. *Mathematica (Cluj)*, vol.4, pp.38-53.

[12] Leadbetter, M.R., Lindgren, G. &Rootzén, H. 1983.*Extremes and related properties of random sequences and processes*. New York: Springer.

[13] Embrechts, P., Kluppelberg, C. and Mikosch, T. (1997). Modelling Extremal Events for Insurance and Finance, Springer.

[14] L. Yang and J. S. Marron. Iterated transformation-kernel density estimation.*J.Amer. Statist. Assoc.*, 94(446): 580{589, 1999. ISSN 0162-1459.

[15] S. Zhang, R. J. Karunamuni, and M. C. Jones. An improved estimator of the density function at the boundary. *J. Amer. Statist. Assoc.*, 94(448):1231{1241, 1999. ISSN0162-1459.

[16] Pickands, J. (1975). Statistical inference using extreme order statistics. Annals of Statistics, 3, 119-131.

APPENDIX – A

R CODES USED:

```
>library(ismev)
Loading required package: mgcv
Loading required package: nlme
This is mgcv 1.8-12. For overview type 'help("mgcv-package")'.
>library(mgcv)
>library(evir)
>library(dplyr)
Attaching package: ‘dplyr’
The following object is masked from ‘package:nlme’:
collapse
The following objects are masked from ‘package:stats’:
filter, lag
The following objects are masked from ‘package:base’:
intersect, setdiff, setequal, union
>library(stringr)
>library(doBy)
>options(scipen = 999)
>claims =read.csv("C:/Users/Symoh/Desktop/Project/project.csv",header = T)
```

```
>claims = slice(claims,2:1271)
>names(claims)
 [1] "PAID.CLAIMS.BORDEREAU" "X"                "X.1"
 [4] "X.2"                   "X.3"              "X.4"
 [7] "X.5"                   "X.6"              "X.7"
[10] "X.8"                   "X.9"              "X.10"
[13] "X.11"                  "X.12"             "X.13"
>claims$Settlement.Amt.. = as.character(claims$X.7)
>claims$Settlement.Amt.. = str_replace_all(claims$Settlement.Amt.., ",", "")
>claims$Settlement.Amt = as.numeric(claims$Settlement.Amt..)
>claims$Settlement.Amt = claims$Settlement.Amt/1000
>claims$Loss.Date = as.character(claims$X.4)
>claims$Loss.Date = as.Date(claims$Loss.Date,format='%d/%m/%Y')
>claims = orderBy(~Loss.Date, claims)
> w=plot(claims$Loss.Date,claims$Settlement.Amt ,type = "l",
+      xlab = "Year" ,ylab = "Settlement Amount(in 1000)", main = "Time Series Data")
> x = claims$Settlement.Amt
> x = na.omit(x)
> c = log10(x+1.05-min(x))
> b=log10(x)
>meplot(x)
>findthresh(b, 700)
[1] 1.238297
>findthresh(c, 700)
[1] 1.22037
>out<- gpd(b, nextremes = 700)
>plot(out)
Make a plot selection (or 0 to exit):
1: plot: Excess Distribution
2: plot: Tail of Underlying Distribution
3: plot: Scatterplot of Residuals
4: plot: QQplot of Residuals
Selection: 1
[1] "threshold = 1.24   xi = -0.309   scale = 0.907   location = 1.24"
Make a plot selection (or 0 to exit):
1: plot: Excess Distribution
2: plot: Tail of Underlying Distribution
```

3: plot: Scatterplot of Residuals

4: plot: QQplot of Residuals

Selection: 2

[1] "threshold = 1.24 xi = -0.309 scale = 1.09 location = 0.651"

Make a plot selection (or 0 to exit):

1: plot: Excess Distribution

2: plot: Tail of Underlying Distribution

3: plot: Scatterplot of Residuals

4: plot: QQplot of Residuals

Selection: 3

Make a plot selection (or 0 to exit):

1: plot: Excess Distribution

2: plot: Tail of Underlying Distribution

3: plot: Scatterplot of Residuals

4: plot: QQplot of Residuals

Selection: 4

Make a plot selection (or 0 to exit):

1: plot: Excess Distribution

2: plot: Tail of Underlying Distribution

3: plot: Scatterplot of Residuals

4: plot: QQplot of Residuals

Selection: 0

>emplot(b, alog="xy")

>out<- gpd(b, nextremes = 700)

>tp<- tailplot(out)

>gpd.q(tp, 0.999)

Lower CI Estimate Upper CI

3.639806 3.754832 4.004355

>gpd.sfall(tp, 0.95)

Lower CI Estimate Upper CI

3.026982 3.105897 3.206193

>hist(log10(x))