Reinsurance Pricing of Large Motor Insurance Claims in Nigeria: An Extreme Value Analysis

Queensley C. Chukwudum

Correspondence: Pan African University Institute of Basic Sciences, Technology and Innovation, Kenya

Abstract

Reinsurance is of utmost importance to insurers because it enables insurance companies cover risks that they, under normal circumstances, would not be able to cover on their own. An insurer needs to be able to evaluate his solvency probability and consequently, adjust his retention levels appropriately because the insurer's retention level plays a vital role in determining the premiums he will pay to the reinsurer. To illustrate how Extreme Value theory can be applied, this study delves into modelling the probabilistic behaviour of the frequency and severity of large motor claims from the Nigerian insurance sector (2013-2016) using the Negative Binomial-Generalized Pareto distribution (NB-GPD). The annual loss distribution is simulated using the Monte Carlo method and it is used to predict the expected annual total claims and estimate the capital requirement for a year. Pricing of the Excess-of-loss (XL) reinsurance is also examined to aid insurers in optimizing their risk management decision in regards to the choice of their risk transfer position.

Keywords: extreme value theory, generalized Pareto distribution, risk management, XL reinsurance, negative binomial

1. Introduction

In the face of extreme risks, an insurer can suddenly fall into ruins. This is because of the low-frequency, high-severity nature of such occurrences. To manage these types of risks, an insurer needs to be well equipped beforehand, having the required knowledge and tools that will enable him to accurately gauge the magnitude of losses that may be experienced and to put appropriate measures in place in case such an event takes place. Extreme Value theory (EVT) provides us with a solid theoretical foundation upon which we can build statistical models that describe extreme events.

As highlighted by Embrechts (1999), modelling claim-size distributions in non-life insurance; stress testing of catastrophe insurance portfolios; pricing of multi-line, multi-year, rare event (high layer) reinsurance products and pricing of catastrophe triggered bonds are some of the gains that can be made when EVT is applied.

Extreme modelling-based inference is a very valuable source of information in the insurers' decision-aiding manual. As stated by Pérez-Fructuoso and García Pérez (2010), the application of EVT to modelling large claims will help the insurer to correctly measure the risks and select the optimal capital requirements. Furthermore, he is able to price the policy appropriately and choose the most favourable reinsurance layer. This important point is re-echoed by Paldynski (2015) who argued that for the company to remain solvent in cases of extremely large insurance claims a combination of capital buffers and reinsurance must be present. This will help to absorb the unexpected losses. Even more important is the fact that such capital shortfalls have a significant impact on the company's existence, thus the need to assess the probability of such occurrences. On reinsurance, Paldynski (2015) gives some reasons for the need for it. It can be used to hedge against say, catastrophic weather in a given geographical region or as a means of sharing risks especially for policies that incur large losses. Reinsurance is also a way of reducing the variance of underwriting results and can enable an insurance company to increase their geographical presence in a regulated industry. Chasseray et al. (2017) adds that it helps the insurer to better monitor the risks faced.

Embrechts (1999) give an introduction to EVT and showcase its use with the aid of examples involving financial and loss data. McNeil (1997) used EVT to model the tails of loss severity distributions while Rootzén and Tajvidi (1997) employed the EVT technique in a study of wind storm losses. Corradin and Verbrigghe (2001) estimated the loss frequency and loss severity of large fire claims. They also examined the impact of a quota and an XL reinsurance on the distribution of the total loss. An EV analysis of the Spanish motor liability insurance market data was carried out by Pérez-Fructuoso and García Pérez (2010). They show how the decision-making process, relating to the pricing of reinsurance, is optimized when considering a risk transfer position. Brazauskasa and Kleefeld (2016) modelled the severity and the tail risk associated with the Norwegian fire claims while that of Taiwan was fitted using the Generalized Pareto distribution (GPD) (Lee, 2012).

With reference to insurance companies or sectors within Africa, Wainaina and Watitu (2014) modelled Kenya's fire industrial insurance class of business using EVT. They estimated the VaR at different confidence intervals. Karobia (2015)

modelled extreme claims for a selected insurance company using GPD. In the assessment of the Nigeria's motor industrial class of business by Adesina et al. (2016) using EVT, the main purpose was to compare the different methods for estimating VaR. Their results reveal that extreme VaR is the most suitable.

This study contributes to literature specifically in the aspect of reinsurance pricing in which no study has been done so far using an African-based data, to the best of the author's knowledge. A simplified view of the concepts is given. The frequency and severity distributions used by Corradin and Verbrigghe (2001) are implemented in this paper and the XL rating under the NB-GP model for different quoted layers are computed. Quota reinsurance is not estimated. In contrast to the simulated dataset used by the authors, a real dataset is used in this paper to estimate the XL reinsurance. The Poisson-GP model was used by Pérez-Fructuoso and García Pérez (2010) and they compared the parametric and non-parametric estimation of the reinsurance risk premium. In this study, only the parmetric estimation procedure is shown.

The rest of this paper is structured accordingly: the theoretical framework is covered in section 2 and XL reinsurance in section 3. The data is analyzed and the results obtained are discussed in the 4th section. Finally, conclusions are made in the 5th section.

2. Theoretical Background

The foundations of EVT were developed by Fisher and Tippet (1928). Coles (2001) gives a very detailed presentation of the theoretical background of EVT. Classical EVT techniques consist of the Block maxima (BM) and the GPD approaches. This study makes use of the GPD method.

2.1 The Generalized Pareto Distribution and the Threshold Approach

This method considers more of the most extreme observations. This is one of the main advantage it possesses when compared to the BM approach. The threshold model is one of the methods for choosing what data to incorporate into the analysis for greater precision. This technique is due to Pickands (1975). Below the PickandsBalkemade Haan Theorem (here denoted as Theorem 1) is stated (see Embrechts, McNeil & Frey (2005) Theorem 7.20). This provides the theoretical justification that allows us to apply the GPD class of distributions when dealing with peaks over threshold modeling.

Theorem 1: Let $X_1, X_2, ...$ be a sequence of independent random variables with common distribution function F and let

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

Let us denote an arbitrary term in X_i sequence by X and suppose that X satisfies the Generalized extreme value conditions so that for large n

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

where convergence is in distribution. G is a non-degenerate distribution function given as

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

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

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

where

$$H(y) = 1 - \left(1 + \xi\frac{y - u}{\sigma_u}\right)_+^{\frac{-1}{\xi}} \qquad y > u \tag{2.1}$$

$H(y)$ is the GPD with the modified scale parameter $\sigma_u = \sigma + \xi(u - \mu)$ corresponding to the excess of the threshold u. $z_+ = max\{z, 0\}$. For $X > u$ we can set $X = X - u$ and thus rewrite the equation as

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

where

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

Again, $z_+ = max\{z, 0\}$ and $\sigma_u = \sigma + \xi(u - \mu)$

In general we have the GPD as

$$\mathbb{P}(X < y | X > u) = \begin{cases} 1 - \left\{ 1 + \xi \left(\frac{y-u}{\sigma_u} \right) \right\}_+^{\frac{-1}{\xi}} & \xi \neq 0 \\ 1 - exp\left\{ -\frac{y-u}{\sigma_u} \right\}_+ & \xi = 0 \end{cases} \tag{2.2}$$

2.2 Modeling Frequency and Severity: Negative Binomial (NB)-GPD Model

The GPD model alone accounts for the magnitudes of the extreme events but to obtain a more exact inference, it is important to take into account the times at which the events occur. Thus the large claims are modeled using a frequency-severity approach. The NB-GPD model which is one of the ways to achieve this, combines the peak-over-threshold (POT) with the NB distribution. This model is closely connected to the claim process in the CramérLundberg model. The CramérLundberg model is a very popular model in early ruin theory - a study of how insurance companies are exposed to insolvency risks.

The GPD model is used to model the severity of the losses while the NB model is fitted to the number of losses above the selected threshold to model the frequency. The NB is used because for the sample provided, the variance is greater than the mean which implies that overdispersion is evident. Thus, using the Poisson model, in this case, may lead to an underestimation of the standard errors. A mix of the Poisson and gamma distributions gives the NB distribution. In essence, we assume that the Poisson parameter λ is gamma distributed making it very useful especially when incorporating parameter uncertainty into the Poisson parameterization.

The frequency-severity model falls under the framework of the loss distribution approach (LDA) where the claim severity is assumed to be independent of the claim frequency and they are both modeled separately. We therefore have the aggregate loss S as:

$$S = \sum_{i=1}^{N} X_i \tag{2.3}$$

X_i is the ith claim (which is assumed to be strictly positive) and N is the number of claims. This is the collective risk model. The assumptions are as follows: the X_i's are independent and identically distributed (iid) and they are independent of N which is a discrete random variable representing the number of losses.

Quantities such as the distribution function, the expected value and the variance of the aggregate loss, S, are usually of interest to the insurer. Specifically in the case of reinsurance, the losses above a retention level R is considered. The distribution of losses above R is briefly discussed.

2.2.1 Distribution of Excess Losses

Drawing from Antal's (2009) study, one can determine the distribution for both the number of excess losses above R and the excess loss burden.

Proposition 1: Suppose the number of the excess losses be given by

$$N_R = \sum_{i=1}^{N} I_{[X_i > R]}$$

where $I_{[X_i > R]}$ is the indicator function and $N \sim NB(r, p)$ with $\pi = \mathbb{P}(X > R)$. Then the distribution of N_R is given as $NB(r, \frac{p}{p+\pi(1-p)})$.

Proof.

A convenient way to prove this is to use the moment generating function (mgf) technique, and the two important properties that will be made use of are stated below.

Property 1: Let X and Y be independent random variables. Let Z be equal to X with probability p and equal to Y with probability $1 - p$. Then the mgf of Z is
$$M_Z(s) = pM_X(s) + (1 - p)M_Y(s)$$

Property 2: Let $X_1, X_2, ...$ be a sequence of iid random variables. Let N be another independent random variable that takes nonnegative integer values. Let $Y = \sum_{i=1}^{N} X_i$. Then the mgf of Y is $M_Y(s) = M_N(log M_X(s))$.

The mgf of N can easily be shown to be

$$M_N(z) = \left(\frac{p}{1 - (1 - p)e^z} \right)^r$$

Let

$$Y = I_{[X>R]} = \begin{cases} 1 & X > R \\ 0 & X \le R \end{cases}$$

This means Y takes on the value 1 with probability π and 0 with probability $1 - \pi$. Therefore mgf of Y is

$$\begin{aligned} M_Y(z) &= E[e^{z.I_{[X_i>R]}}] \\ &= (1-\pi)E[e^{z.0}] + \pi E[e^{z.1}] \qquad (from\ property\ 1) \\ &= 1 - \pi + \pi e^z \end{aligned}$$

From property 2 , the mgf of $N_R = \sum_{i=1}^{N} Y_i$ will be

$$\begin{aligned} M_{N_R}(z) &= M_N[log(M_Y(z))] \\ &= \left(\frac{p}{1 - (1-p)e^{logM_Y(z)}} \right)^r = \left(\frac{p}{1 - (1-p)M_Y(z)} \right)^r \\ &= \left(\frac{p}{1 - (1-p)(1 - \pi + \pi e^z)} \right)^r = \left(\frac{p}{p + (1-p)\pi - (1-p)\pi e^z} \right)^r \\ &= \left(\frac{\frac{p}{p+(1-p)\pi}}{\frac{p+(1-p)\pi-(1-p)\pi e^z+pe^z-pe^z}{p+(1-p)\pi}} \right)^r = \left(\frac{\frac{p}{p+(1-p)\pi}}{1 - \left(1 - \frac{p}{p+(1-p)\pi}\right)e^z} \right)^r \qquad QED. \end{aligned}$$

$\Rightarrow N_R \sim NB(r, \frac{p}{p+\pi(1-p)})$

Next, let us look at the distribution of the excess loss burden $\sum_{i=0}^{N}(X_i - R)_+$.

Proposition 2: Given R and N_R (as defined in proposition 1) then

$$\sum_{i=0}^{N}(X_i - R)_+ \text{ has same distribution as } \sum_{i=0}^{N_R}(Y_i - R)_+.$$

where Y_i are iid random variables with distribution $\mathbb{P}[.|X > R]$ that are independent of N_R

Proof.

Let the probability measure of model (N, X) be $\mathcal{P}$ and that of model (Y, N_R) be $\mathbf{P}$.

$$\begin{aligned} \mathbf{P}\left[\sum_{i=1}^{N}(X_i - R)_+ \le s \right] &= \sum_{n=0}^{\infty} \mathbf{P}\left[\sum_{i=1}^{N}(X_i - R)_+ \le s | N_R = n \right] \mathbf{P}[N_R = n] \\ &= \sum_{n=0}^{\infty} \mathbf{P}\left[\sum_{i=1}^{n}(X_i - R)_+ \le s | X_1, ...X_n > R \right] \mathbf{P}[N_R = n] \\ &= \sum_{n=0}^{\infty} \mathcal{P}\left[\sum_{i=1}^{n}(Y_i - R)_+ \le s \right] \mathcal{P}[N_R = n] \\ &= \mathcal{P}\left[\sum_{i=1}^{N_R}(Y_i - R)_+ \le s \right] \qquad QED. \end{aligned}$$

2.3 Tail Risk Management Measures

A key issue in risk management is the definition of risk as well as the definition of relevant risk measures. The most commonly used risk measures are value at risk (VaR) and expected shortfall (ES). VaR describes the amount of extreme loss which is exceeded only with a certain small probability within a given time frame.

$$VaR_q = inf\{u : F(u) \ge q\}$$

To obtain VaR using GPD, we consider a threshold u. The distribution of the excesses above u is

$$F_u(y) = \mathbb{P}(X - u \le y | X > u) = \frac{F(y + u) - F(u)}{1 - F(u)}$$

for some underlying distribution F describing the series X_i. $0 \leq y < x_F - u$. x_F is the right end point of F. From Theorem 1, we know that $F_u(y)$ can be well approximated using the GPD distribution. That is, $F_u(y) \cong H_{(\xi,\sigma)}(y)$ as $u \to \infty$. Thus the underlying distribution becomes

$$F(x) = F(u) + [1 - F(u)]H_{(\xi,\sigma)}(x - u)$$

for $x > u$ and the empirical estimate for $F(u)$ is

$$\hat{F}(u) = \frac{n - N_u}{n}$$

where n is the total number of observations and N_u is the number of observations above the threshold u. Therefore, the tail estimate can be written as

$$\hat{F}(x) = \frac{N_u}{n}\left(1 - H_{(\xi,\sigma)}(x - u)\right) + \left(1 - \frac{N_u}{n}\right) = 1 - \frac{N_u}{n}\left(1 + \xi\left(\frac{x - u}{\sigma}\right)\right)_+^{\frac{-1}{\xi}}$$

By inverting the tail estimate formula we obtain VaR

$$VaR_q = u + \frac{\hat{\sigma}}{\hat{\xi}}\left(\frac{n}{N_u}(1 - q)^{-\hat{\xi}} - 1\right) \tag{2.4}$$

The ES which can be described as the expected value of the size of the loss exceeding the VaR with some level of probability conditioned on the fact the loss actually exceeds the VaR is

$$ES_q = VaR_q + E[X - VaR_q | X > VaR_q] = \frac{VaR_q}{1 - \hat{\xi}} + \frac{\hat{\sigma} - \hat{\xi}u}{1 - \hat{\xi}} \tag{2.5}$$

3. XL Reinsurance

Let X represent the loss incurred and suppose we have an insurance policy with retention level R, then, the payment made by the reinsurer will be $g(X)$ where

$$g(X) = \begin{cases} 0 & if\ X < R \\ X - R & if\ X \geq R \end{cases} \tag{3.6}$$

Hence the reinsurance premium (RP) will be the expected value of the claims, $E[g(X)]$.

When we factor in the frequency of claims into the computation, the expected total (aggregate) loss according to the classical risk theory hypothesis becomes

$$RP = E(S) = E(N) * E(g(X)) \tag{3.7}$$

Therefore under the NB-GPD model, it implies that:

$$E(N) = \frac{rq}{p} \text{ and } E((g(X)) = E(H_{\xi,\sigma,u}(x)) = E(H)$$

where $\frac{rq}{p}$ denotes the average number of claims and $E(H_{(\xi,\sigma,u)}(x))$ is the expected value of the GPD with parameters ξ, σ and retention u. We assume that the occurrence moments and the loss amounts fulfill the conditions of a compound NB process. That is, the collective risk model where N follows a NB distribution given by

$$S = \sum_{i=1}^{N} X_i \qquad N \sim NB(r, p)$$

Parametrically, we will estimate the reinsurance risk premium as

$$RP = \frac{rq}{p} * E(X | X > u) = \frac{rq}{p} * \int_u^\infty (x - u)dH^{(u)}(x) \tag{3.8}$$

with $dH^{(u)}(x) = h^{(u)}(x)dx$ being the density function of the GPD. To be more specific, the reinsurance layer may not coincide with the threshold u, of the optimized GPD, hence we have to compute $\mathbb{P}(X > R | X > u)$.

We recall that the expected random amount $(X - R)_+$ which concerns the reinsurer can also be written as

$$RP = E((X - R)_+ = \int_R^{x^*} (1 - F(y))dy$$

and since the mean excess function (MEF), e(R) is

$$e(R) = E(X|X > R) = \frac{\int_R^{x^*} (1 - F(y))dy}{1 - F(R)}$$

It implies that, in terms of the MEF

$$RP = e(R)\bar{F}(R) \tag{3.9}$$

where $\bar{F}(R) = 1 - F(R)$ is the survival function. In order to obtain a simpler expression for the integral of the reinsurace premium for an unlimited cover which is

$$RP = \frac{rq}{p} * \int_R^{\infty} (x - R)h^{(u)}(x - u)dx \tag{3.10}$$

we note that

$$\mathbb{P}(X > R|X > u) = \frac{\mathbb{P}(X > R)}{\mathbb{P}(X > u)}$$

Furthermore, the asymptotic behavior of the tail properties provides us with the formula to estimate the tail probability when $x \to \infty$ given that certain conditions hold.

$$\mathbb{P}(\bar{F}(x)) \sim E[N]\bar{F}(x) \quad as \ \ x \to \infty$$

Thus, we can approximate $\mathbb{P}(X > R) = \bar{F}(R)$ by

$$\bar{F}(R) \sim \bar{F}(u)\left(1 + \xi\frac{R - u}{\sigma}\right)^{-\frac{1}{\xi}} \qquad u \to \infty, R > u$$

when the tail is fitted with the GPD. From first principles, the formula of the MEF of the GPD is

$$e(u) = \frac{\sigma}{1 - \xi} + \xi\frac{x - u}{1 - \xi}$$

$\Rightarrow$ under the condition $R > u$, $\xi < 1$ and u is large

$$e(R) \sim \frac{\sigma}{1 - \xi} + \xi\frac{R - u}{1 - \xi} = \frac{\sigma}{1 - \xi} + \xi\frac{\sigma(R - u)}{\sigma(1 - \xi)}$$
$$= \frac{\sigma}{1 - \xi}\left(1 + \xi\frac{R - u}{\sigma}\right)$$

Therefore,

$$RP = e(R)\bar{F}(R)$$
$$= \bar{F}(u)\left(1 + \xi\frac{R - u}{\sigma}\right)^{-\frac{1}{\xi}}\left\{\frac{\sigma}{1 - \xi}\left(1 + \xi\frac{R - u}{\sigma}\right)\right\} \tag{3.11}$$
$$= \frac{rq}{p}\frac{\sigma}{1 - \xi}\left(1 + \xi\frac{R - u}{\sigma}\right)^{1 - \frac{1}{\xi}}$$

If we set $u = X_{n-\hat{k},n}$ where $\hat{k}$ is the suitable choice made for the number of upper statistics, we can then estimate the net RP for an unlimited XL layer above R as

$$RP = \frac{\hat{k}}{n}\frac{\hat{\sigma}}{1 - \hat{\xi}}\left(1 + \hat{\xi}\frac{R - X_{n-\hat{k},n}}{\hat{\sigma}}\right)^{1 - \frac{1}{\hat{\xi}}}$$

Here, the exceedance probability has been estimated empirically by $\frac{\hat{k}}{n}$.

However, if we have a bounded layer (R, R^*) with $u < R < R^*$ in excess of our threshold u, then

$$RP = \frac{rq}{p} * \int_R^{R^*} (x - R)h^{(u)}(x - u)dx + (R^* - R)(1 - H^{(u)}(R^* - u)) \tag{3.12}$$

and the explicit formula to compute the integral will be

$$I = \frac{rq}{p} \frac{\sigma}{\xi} \left[\left(1 + \frac{\xi}{\sigma}(R^* - u)\right)^{1-\frac{1}{\xi}} - \left(1 + \frac{\xi}{\sigma}(R - u)\right)^{1-\frac{1}{\xi}} \right] \qquad \xi \neq 0 \qquad (3.13)$$

4. Data Description and Analysis

The data used in this study are the large claims (3.5 million naira and above) of Nigeria's motor industrial class of business. It covers the years 2013 to 2016. The four years of data give a total of 1,324 claims. Analyzing these motor claims will provide the insurance industry with useful insights given that, according to Ademunigbohun and Oreshile (2014), it (specifically, third party motor insurance) is the major driver of the non-life insurance market in Nigeria. Data is obtained from Nigerian Insurance Commission (NAICOM) at *www.naicom.gov.ng*. It can be found in the Nigeria Insurance digest for the specified years.

The histogram of the claims (on the log scale) is strongly skewed towards the right (Figure 1). This suggests that there is a frequent occurrence of smaller motor claims while larger losses take place occasionally. From the plot we can observe that more of the losses fall between 3.5 and 6 million naira. Inflation is not taken into account in this analysis.
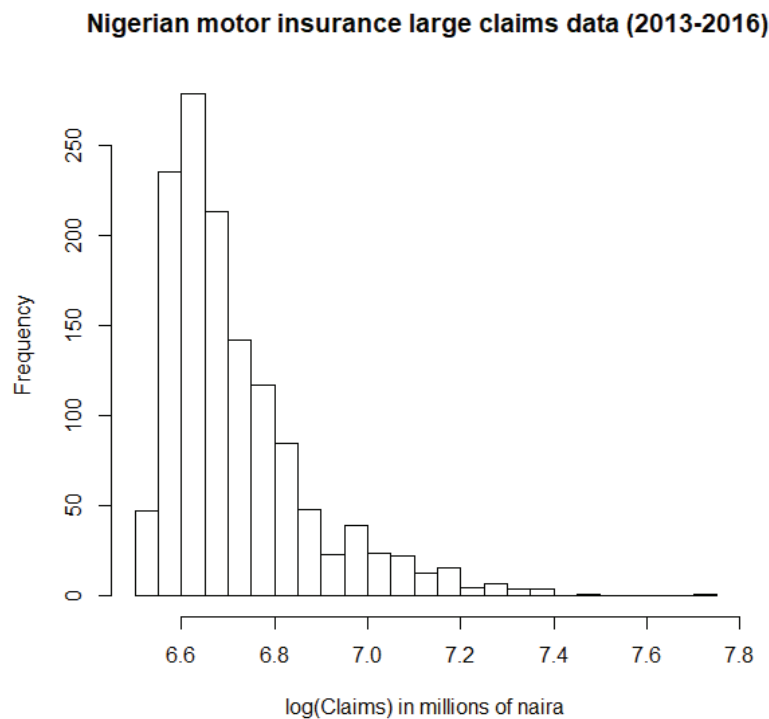

Figure 1. Histogram of claims on the log scale

In Tables 1 and 2, the summary statistics of the claims are shown. It reveals that more claims were paid in the year 2015. The general trend is encouraging. It shows that the industry is growing given that such large amounts of claims are being paid yearly.

Table 1. Summary statistics

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
Claims	3,247,500	4,050,000	4,680,000	5,708,460	6,075,000	51,810,400
Log claims	6.512	6.607	6.670	6.720	6.784	7.714

Table 2. The 3 largest claims and sample size in each year

Year	2013	2014	2015	2016
	16,245,000	21,765,289	19,000,000	21,802,500
	20,124,610	23,302,148	19,885,000	23,900,000
	23,479,720	24,752,550	51,810,400	28,215,000
Sample size	244	286	419	376

4.1 Fitting the NB-GPD Model

The shape of the plot (Figure 1) and the increasing MEF (Figure 2 right) indicate that the data is heavy tailed. Thus the sample is fitted to a GPD distribution to model the severity of claims. First we determine the threshold u. A useful tool for this is the mean excess plot (MEP) as seen in Figure 2.
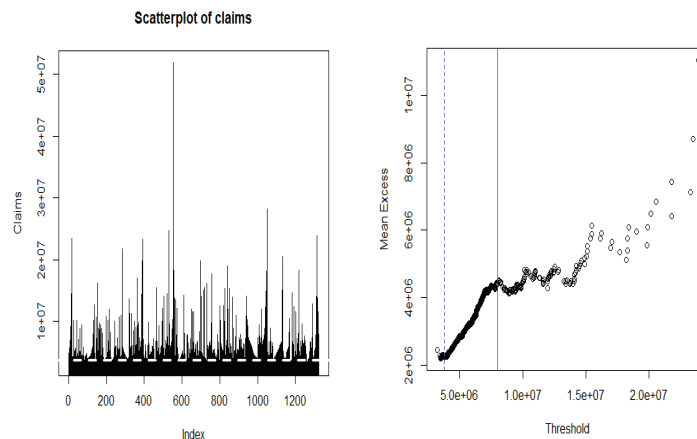

Figure 2. Scatterplot (left) and MEP (right) of motor claims

In the case of the GPD, the MEF of a random variable X is a linear function in u (Embrechts et al., 2005). In the Figure 2 (plot on the right) the plot is fairly linear (has a positive slope which is confirmed by the positive shape parameter, ξ in Table 3) up to the 8 million mark approximately (indicated by continous vertical line) and then, begins to waver. The wavering is due to the lack of data available as regards very huge losses. The lowest point from where the plot becomes linear, using the eyeball approach, is chosen as the threshold. This is set at 3.8 million as indicated by the dotted vertical line on the MEP and the dotted horizontal line on the scatterplot of the claims. This leaves 1128 losses as the number of excesses (nexc) to fit the GPD to (Table 3). The parameters are estimated using the maximum likelihood method with the respective standard errors given in the braces.

The diagnostic plots (Figure 3) show four different plots. The quantile (probability) plot has most of the points falling on the 45° diagonal line, the return level plot (a plot of the return level against the return period) has a central line representing the return level for the fitted model and two outer lines representing the pointwise 95% confidence limits and most of the points which are the empirical return levels (from the dataset) fall within the confidence limits. All these indicate that GPD is a good fit. The density plot is simply the fitted probabilty density function (GPD) superimposed on a histogram that has been generated using the data. The p-values for the statistical goodness of fit tests - Cramer-von Mises (CvM) and Anderson-Darling (AD) tests are 0.31 and 0.21 respectively (significance level, α= 0.05). This further confirms that GPD is a good fit.

Table 3. The estimated parameters based on the fitted NB-GPD

GPD parameters				NB parameters	
threshold	nexc	shape (ξ)	scale (σ)	size	mu
3,800,000	1128	0.3115441	1,578,755	24.24950	282.00000
		(0.03060106)		(18.53834)	(29.83875)

To model the frequency of the claims using the NB, the number of losses per year above the given threshold is taken. The estimated parameters are displayed on Table 3. We can easily obtain the probability of success, p, for the NB distribution from the size and mean parameters with this simple formula

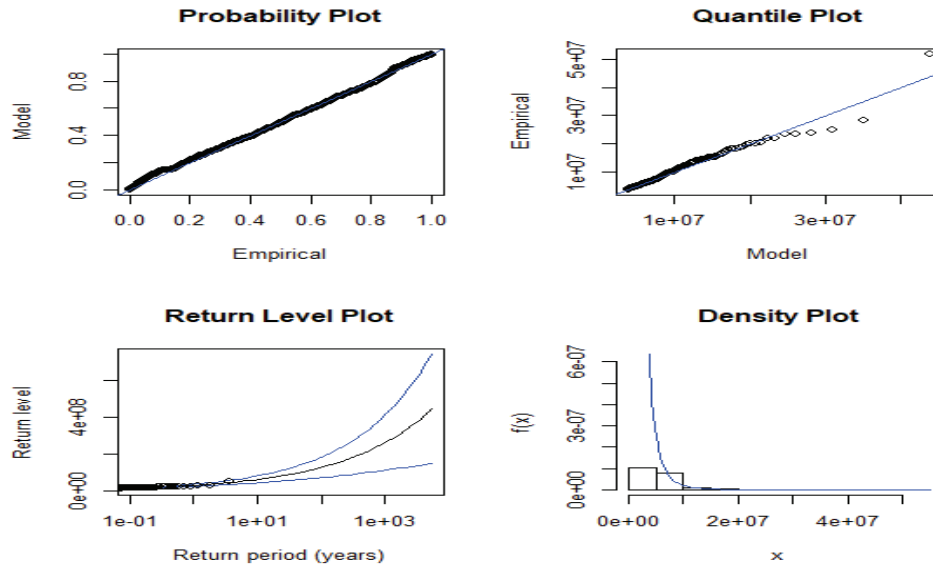$$p = \frac{size}{size + mu} = \frac{24.25}{24.25 + 282} = 0.079$$

Figure 3. Diagnostic plots to assess the GPD's goodness of fit for the historical motor claims (see brief explanation of these plots in Section 4.1, paragraph 3)

Risk Measures

After fitting the data to GPD, the VaR and ES values are estimated together with their respective confidence intervals (CI). It indicates that 99% of claims that occur result in a loss below 18,990,729 naira, and in the case where a claim results in a loss larger than this amount, the expected loss will be 28,180,720 naira (Table 4).

Table 4. VaR and ES at the 0.99 quantile

VaR			ES		
Lower CI	Estimate	Upper CI	Lower CI	Estimate	Upper CI
16850881	18990729	21979944	23291490	28180720	35956244

VaR and ES estimates for the historical claims with confidence intervals.

4.2 Simulating the Annual Loss Distribution

The annual aggregate loss distribution is simulated using the Monte Carlo technique. This involved making use of the parameter values that were obtained from the frequency and severity analysis above. The implemented algorithm follows that of Piacenza (2012) annual loss distribution R algorithm with the difference being that the NB is used here to obtain the frequency distribution. Random numbers of size n=100,000 are generated from the NB distribution with mean mu=282. Similarly, the severity distribution is generated by drawing the same number, n, from the GPD using the estimated parameters in Table 3. Then, for each random number extracted from the frequency distribution say, $n_i = 2$, two random numbers (in this case, losses) are also extracted from the severity distribution and the sum of the extracted losses is gotten.

That is, $S = \sum_{i=1}^{N} X_i$ where $N \sim NB(r, p)$. This process is repeated n times to obtain the annual loss distribution. To visualize how the distribution looks like, a histogram depicting this is shown in Figure 4b. Simulated individual losses are also shown (Figure 4a).
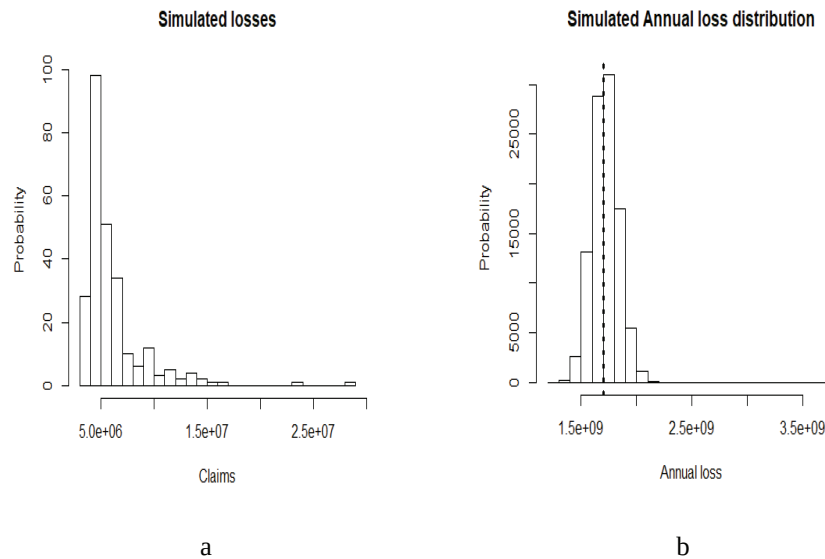

a

b

Figure 4. Simulated individual losses (a) and annual loss distribution (b) using the parameters obtained from the historical claims

From the annual loss distribution, the expected overall annual loss (EL) and its VaR can be computed. The EL is the quantile of the annual loss at the 0.5 mark, this gives 1,715,027,838. It is depicted by the dotted line in Figure 4b. This means that on average the motor insurance industry should expect a total of about 1.7 billion naira worth of claims in a year. The 99% VaR for the annual loss is 2,012,204,962. Thus the allocated or required capital will be gotten by deducting the EL from VaR resulting in 297,177,124 naira.

4.3 Computing the XL Reinsurance Risk Premium

After selecting a priority level R, an insurer may not want to insure all the risk that exceeds that level rather he may choose a layer of reinsurance corresponding to the interval $(R, R^*]$. Hence, to replicate this sort of scenario, 3 different layers of limited coverages are assumed. Also, one XL with an unlimited cover (20mill, ∞) is also considered. The 3 levels of limited-cover reinsurance are 2mill xs 10mill (10,12]mill, 3mill xs 12mill (12,15]mill and 5mill xs 15mill (15,20]mill. Here mill represents million.

In Table 5, the survival probabilities of exceeding the lower limit in each layer and their respective reinsurance net premium results are displayed.

Table 5. XL reinsurance layer and premium

Reinsurance layer (r,R)	Survival probability	Net Premium
(10,12]mill	0.0769	9,390,099
(12,15]mill	0.0455	4,641,891
(15,20]mill	0.0237	1,886,666
greater than 20mill	0.0100	964.7832

Assuming that only large claims above 3.5 million naira are in the portfolio, there is a 0.02 (2.4%) probability that the claims will exceed 15 million naira given that it has exceeded 3.8 million naira (table 5, third row). In other words, approximately 980 out of the next 1000 claims ($1 - 0.0237 = 0.9763$) that exceed the 3.8 million naira threshold will remain under 15 million naira. This implies that roughly about 20 of these claims will exceed 15 million naira costing

the motor class of business on average, about 1.9 million naira net premium to have a reinsurance cover in excess of 15 million naira with a limited cover of 20 million naira.

The trend indicates that the higher the layer of reinsurance, the lower the probability of survival and the lower the net premium the insurer will pay. This will imply less profit for the insurer if he picks the lowest layer and although the probability that the losses exceeding his retention limit will be higher, his probability to ruin will be smaller.

5. Conclusion

This paper showcases how powerful EVT is for modelling the tail of very large losses. Extreme losses were analyzed using a frequency and severity based approach because the survival of an insurance company critically depends on this. It was shown that the GPD served as a good fit to the tails of the motor claims as indicated by the AD and CvM tests. The expected annual total claims for the motor insurance industry was estimated based on the simulated annual loss distribution.

The importance of having good estimates of the tails of the claims was made evident in the pricing of the XL reinsurance from the perspective of the insurer. Solvency probabilities were calculated for the considered layers of reinsurance coverage to help the insurer in deciding whether to assume or cede certain risks and to have an idea of the risk premium he will be paying in the case of ceding. These analyses bring to the fore a major issue constantly faced by an insurer when deciding on his retention level, which is, balancing the trade-off between profit and security. This can serve as a scope for further study.

Another interesting direction for future research will include taking inflation into account and studying its effect on the pricing process. The pricing of different reinsurance structures and their various combinations based on the simulated loss distribution is also a research area one can examine.

Acknowledgements

References

Adesina, O., Adeleke, I., & Oladeji, T. (2016). Using Extreme Value Theory to Model Insurance Risk of Nigeria's Motor Industrial Class of Business. *The Journal of Risk Management and Insurance, 20*(1), 40-51.

Ademunigbohun, S. S., & Oreshile, S. A. (2014). Risk Attitude and Demand for Motor Insurance: An Examination of selected Motorist in Lagos State. *Developing Country Studies, 4*(21).

Antal, P. (2005). *Mathematical Methods in Reinsurance*. Swiss Reinsurance Company.

Brazauskasa, V., & Kleefeld, A. (2016). Modeling Severity and Measuring Tail Risk of Norwegian Fire Claims. *The North American Actuarial Journal, 20*(1). https://doi.org/10.1080/10920277.2015.1062784

Chasseray, P., Eldin, G., & Lefebvre, A. (2017). *P&C Reinsurance modeling: Pure Premium Estimation and Creation of a Reinsurance Program*. Euro-Institut dActuariat AXA Global P&C.

Coles, S. G., Heffernan, J., & Tawn, J. (2000). Dependence Measures for Extreme Value Analyses. *Extremes, 2*(4), 339-365. https://doi.org/10.1023/A:1009963131610

Corradin, S., & Verbrigghe, B. (2001). *Economic Risk Capital and Reinsurance: An Application to Fire Claims of an Insurance Company*.

Embrechts, P. (1999). *Extreme Value Theory in Finance and Insurance*. Dept. of Mathematics, ETH Zurich.

Embrechts, P., McNeil, A. J., & Frey, R. (2005). *Quantitative Risk Management: Concepts, Techniques, and Tools*. Princeton University Press, pp. 277.

Fisher, R., & Tippett, L. (1928). Limiting Forms of the Frequency Distribution of the Largest or Smallest Member of a Sample. *Mathematical Proceedings of the Cambridge Philosophical Society, 24*(02), 180-190. https://doi.org/10.1017/S0305004100015681

Karobia, R. J. (2015). *Modelling Extreme Claims using Generalized Pareto Distributions Family in Insurance Company*. MSc thesis, University of Nairobi.

Lee, W. (2012). Fitting the Generalized Pareto Distribution to Commercial Fire Loss Severity: Evidence from Taiwan. *Journal of Risk, 14*(3), 63-80. https://doi.org/10.21314/JOR.2012.244

McNeil, A. J. (1997). Estimating the Tails of Loss Severity Distributions Using Extreme Value Theory. *ASTIN Bulletin,*

27(1), 117-37. https://doi.org/10.2143/AST.27.1.563210

Paldynski, H. (2015). *Modelling Large Claims in Property and Home Insurance - Extreme Value Analysis*. MSc. Thesis, Lund University, Faculty of Engineering, Centre for Mathematical Sciences.

Pérez-Fructuoso, M. J., & García, P. A. (2010). Analyzing Solvency with Extreme Value Theory: An Application to the Spanish Motor Liability Insurance Market. *Innovar, 20*(36), 35-48.

Pickands, J. (1975). Statistical Inference using Extreme Order Statistics. *Annals of Statistics, 3*(1), 119-31. https://doi.org/10.1214/aos/1176343003

Rootzén, H., & Tajvidi, N. (1997). *Extreme Value Statistics and Wind Storm Losses: A Case Study*. *Scandinavian Actuarial Journal,* 70-94. https://doi.org/10.1080/03461238.1997.10413979

Wainaina, H. W., & Waititu, A. G. (2014). Modeling Insurance Returns with Extreme Value Theory (A Case Study for Kenya's Fire Industrial Insurance Class of Business). *Mathematical theory and modeling, 4.*

Copyrights