Modelling of Large Fire Insurance Claims: An Extreme Value Theory Approach

Nelson Christopher Dzupire, Dalitso Kachusa & Samuel Gyamerah

University of Malawi

ABSTRACT

This research paper aims to develop a mathematical model that employs Extreme Value Theory (EVT) and Risk Measures to estimate and forecast significant fire insurance claims. The primary goal is to provide insurance companies with a more accurate understanding of the potential risks associated with substantial fire-related losses. The study incorporates a three-parameter Generalized Pareto Distribution (GPD) within the EVT framework to assess insurer risk concerning catastrophic fire events. The importance of evaluating fire-related financial losses for insurers is emphasized, especially given the impact of infrequent yet impactful extreme events on overall loss trends. By applying EVT techniques, including the GPD and Peaks Over Threshold (POT) method, to a historical dataset of fire insurance claims, the study effectively models the tail behavior of large losses. Parameters obtained from these models facilitate the calculation of probabilities for extreme loss occurrences, thereby enhancing risk management and pricing strategies for insurance firms. The results demonstrate the EVT approach's effectiveness in accurately modeling and estimating the risk associated with significant fire insurance claims. This research contributes to the insurance domain by presenting an enhanced mathematical and statistical framework for modeling substantial fire insurance claims. Such an approach enables insurers to better comprehend the potential financial implications of rare fire incidents, leading to more informed risk evaluation and resource allocatio.

Keywords: extreme value theory (EVT), generalized pareto distribution (GPD), fire insurance claims, risk management, 1 catastrophic events.

Classification: LCC Code: HG9970.3

Language: English

Great Britain Journals Press

LJP Copyright ID: 925695
Print ISSN: 2631-8490
Online ISSN: 2631-8504

London Journal of Research in Science: Natural and Formal

Volume 24 | Issue 8 | Compilation 1.0

Modelling of Large Fire Insurance Claims: An Extreme Value Theory Approach

Nelson Christopher Dzupire[α *], Dalitso Kachusa[σ] & Samuel Gyamerah[ρ]

ABSTRACT

This research paper aims to develop a mathematical model that employs Extreme Value Theory (EVT) and Risk Measures to estimate and forecast significant fire insurance claims. The primary goal is to provide insurance companies with a more accurate understanding of the potential risks associated with substantial fire-related losses. The study incorporates a three-parameter Generalized Pareto Distribution (GPD) within the EVT framework to assess insurer risk concerning catastrophic fire events. The importance of evaluating fire-related financial losses for insurers is emphasized, especially given the impact of infrequent yet impactful extreme events on overall loss trends. By applying EVT techniques, including the GPD and Peaks Over Threshold (POT) method, to a historical dataset of fire insurance claims, the study effectively models the tail behavior of large losses. Parameters obtained from these models facilitate the calculation of probabilities for extreme loss occurrences, thereby enhancing risk management and pricing strategies for insurance firms. The results demonstrate the EVT approach's effectiveness in accurately modeling and estimating the risk associated with significant fire insurance claims. This research contributes to the insurance domain by presenting an enhanced mathematical and statistical framework for modeling substantial fire insurance claims. Such an approach enables insurers to better comprehend the potential financial implications of rare fire incidents, leading to more informed risk evaluation and resource allocation.

Keywords: extreme value theory (EVT), generalized pareto distribution (GPD), fire insurance claims, risk management, 1 catastrophic events.

Author α σ : University of Malawi, Zomba, Malawi.
ρ: Kwame Nkrumah University of Science and Technology Kumasi-Ghana.

I. INTRODUCTION

In the course of a lifetime, exceptional occurrences can emerge that are entirely unprecedented. These occurrences are unexpected and are likely to repeat in the future. For instance, we can consider the historical China floods of 1931 or the 2004 Indian Ocean disaster, along with the recurring catastrophic floods in the lower Shire region of Malawi and the devastating effects of Cyclone Fredd in 2023 where lives and properties where lost. Such events lead to significant infrastructural harm, including damage to buildings, roads, bridges, and power lines. This brings about disruptions in essential services such as transportation, communication, and utilities, coupled with the unfortunate loss of numerous lives. These rare events do not only cause great damage and loss of lives but also their effects are widespread throughout all sectors of the economy including the financial sector of which insurance companies are of great concern [19]. Hence, modelling the behaviour and the distribution pattern of these extreme and unusual events is of great importance in electing early warning systems and risk management applications. Extreme events such as catastrophic floods, fires and disease outbreaks do not occur frequently. However, their occurrence result in huge losses and since they rarely occur, many people do not take time to insure against them. A couple of studies and insurance companies have over the years

London Journal of Research in Science: Natural and Formal

tried to compile and analyze data to enable them handle these catastrophic events [3, 16, 7, 18]. Yet, they have challenges in putting necessary reserves and appropriate level of reinsurance which could later on lead many insurers insolvent when catastrophic losses reoccur. Inadequate loss reserves often stand out as a frequent cause of insolvency [5]. A number of researchers have modelled the impact of surgency of insurance claims on the operation of insurance companies. For instance, the work of [2] modelled catastrophe claims with left truncated severity distributions and modelled event frequency with a non-homogeneous poissoin process and later used the model for forecasting. Similarly, [10] studied claim severity of extreme fire insurance claims for tariffication using extreme value theory (EVT) and generalised linear models (GLM). They then suggested that GLM do not work adequately for extreme claims and revealed that peaks over threshold methodology from extreme value theory (EVT) using a two-parameter generalised pareto distribution (GPD) as an accurate distribution for large insurance claims. On modelling rare and catastrophic events and their impacts in finance and insurance applications, many studies have achieved promising results by using different approaches. For instance, in a study by [20], large motor insurance claims in Kenya was modelled using the extreme value theory. Their study concentrated on the right tails of the underlying distribution (extremely large observations) and they fitted a generalised pareto distribution (GPD) which is a family distribution in EVT. Their empirical findings revealed that modelling extreme outcomes under EVT theory outweighs other methods of estimation such as econometric methods, as EVT is known for its ability to model the tail area of the distribution where extreme outcomes are located much better. Similar results were obtained by [11] who tried to investigate the tail behaviour of the extreme outcomes in the US stock market using EVT. The study concluded that, S&P 500 daily return data can also be characterized by GPD. However, the commonality in these above studies is that they all used two parameter GPD in modelling the extreme outcomes. Likewise, [13] modelled large flood insurance claims in Zimbabwe using frequency and severity models, their study presented a framework for choosing the most suitable probability distribution which they later fitted it to the past claims data and the parameters were estimated using maximum likelihood method (MLE). According to their findings, Pareto and Negative Binomial model provided the best fit to claims severity and frequency respectively. In the same pursuit, [15] estimated the risks of extremely large fire insurance claims using a Markov Chain monte Carlo approach. They proposed a Bayesian method using Markov Chain Monte Carlo techniques to calculate probabilities of large fire losses and demonstrated its potential advantages for actuarial and risk evaluations. Amongst emerging researchers, [14] modelled the frequency and severity of auto insurance claims using statistical distributions. Their paper presented a methodical framework for choosing a suitable probability model that best describes automobile claim frequency and loss severity as well as their application in risk management. Their findings from empirical analysis indicated that claims severity distribution is more accurately modeled by a skewed and heavy-tailed distribution. In a recent investigation conducted by [8], an examination was carried out on the behavior of extreme returns within the South African Industrial Index (J520) spanning the years 1995 to 2018. The study leveraged the Generalized Extreme Value Distribution (GEVD) to assess and estimate extreme risk measures. Their research outcomes revealed that across distinct quarterly return periods (8, 20, and 40 quarters), the approximated values for extreme losses were 9.28%, 13.65%, and 17.03% respectively, whereas the potential levels of extreme gains during these same periods were noted as 9.81%, 11.63%, and 12.68%.

Due to increasingly severe world catastrophes in the last two decades the property insurance industry has paid out over $125 billion in losses [17]. In 2004, property insured losses resulting from natural catastrophes and man made disasters, excluding the tragic tsunami of December 26, amounted to $42 billion, of which 95% was caused by natural disasters and 5% by man-made incidents such as fire. These huge billion-dollar figures call for very accurate models of catastrophe losses. Even small discrepancies in model parameters can result in underestimation of risk leading to billion-dollar losses to either insurer or the reinsurer. Therefore, employing suitable mathematical and statistical models, along with a thorough analysis of catastrophic data and precise estimation of claim frequency and severity distributions, holds the essential solution for accurately assessing costs and the probability of financial distress for insurers [9]. Even though the methods used in existing literature are helpful, there has been no studies on linking a three-parameter GPD under EVT and measures of risk such as Value at Risk (VaR) and expected shortfall (ES) for estimating probabilities and magnitude of large event outcomes in Malawian Insurance industry. Hence, the motivation behind this investigation lies in the need to comprehend and analyze the patterns underlying the severity of fire-related losses arising from significant insurance claims. The primary objective is to develop a robust model that captures the intricate dynamics of such losses and further assesses the potential for encountering exceedingly rare and severe fire-related losses. The pursuit of this research stems from its potential to equip insurance firms with valuable insights. Specifically, this inquiry can empower insurance companies to predict the potential magnitude of losses that might be incurred across a specific customer base within a defined timeframe. These predictions hold substantial value in guiding the processes of pricing and risk management for insurance firms operating in the non-life sector.

In this paper we present a modelling framework for insurance losses resulting from catastrophic fires in the property insurance industry, which will aid in forecasting insurance risk and liabilities. By doing this, we also seek to derive distribution model of fire loss under EVT, fit fire claims to the loss distribution model for estimating probabilities and magnitude of large fire loss and then forecast expected insurance loss due to risk of fire loss. The results of this study will provide academicians and other researchers with a strong foundation in a wide range of mathematical and probabilistic methods for risk modelling in general insurance, model-based pricing, risk sharing, ruin theory and credibility. Insurance regulators around the world, like the Reserve Bank of Malawi through Insurance Act 2017, requires that every commercial building be insured against collapse, fire, earthquake, storm and flood. Henceforth, Modelling fire loss severity will help insurance companies to make a framework for new product development and make data-driven projections of future fire losses. The rest of the study is organized as follows: chapter two explains the Methodology adopted in the study. Chapter three gives the Empirical Results and Discussions. Finally chapter four gives Conclusions and Recommendations of the study.

II. MATERIALS AND METHODS

The research relied on fire loss data, encompassing claim amounts, extracted from Reserve Bank of Malawi (RBM) and NICO General Insurance. NICO was chosen due to its prominence as the largest General Insurer in Malawi based on market capitalization. The Reserve Bank of Malawi (www.rbm.mw) served as the regulatory authority for the country's insurance sector. Our focus centered on fire insurance claim data, specifically examining claim amounts disbursed (referred to as claim-size). The data collection period spanned from January 2005 to December 2021. The modeling process centered on a 10-year dataset spanning 2011 to 2021. Data organization and analysis were conducted using Microsoft Excel and the R programming package.

2.1 Model Specification

[20] used Extreme Value Theory(peaks over threshold modelling) for modelling large motor insurance losses due to accident claims of Kenindia Insurance Company in Kenya. They fitted a two-parameter GPD under EVT and revealed that it gives a more satisfactory fit to large motor insurance claims. The purpose of this study is to model the risk of large losses in an insurance portfolio due fire damage in commercial property. In light with this, the study adopted and refined the Extreme Values Model used by [20]) by extending the model to a three-parameter GPD. The two-parameter GPD model used by [20] is specified as the function $G_{\xi,\beta}(x)$ defined as follows:

where:

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

$\beta > 0$ is the scale parameter

$x \geq 0$ when $\xi \geq 0$ and $0 \leq x \leq -\frac{\beta}{\xi}$

ξ is the shape parameter

X represents the motor claim amount.

In EVT we consider a random variable say X and fix a threshold u and focus on the positive part of $X - u$ (since we are focusing on the upper tail, where large outcomes are located). This distribution—that is, $F_u(x)$ in EVT theory is given by;

$$F_u(x) = Pr(X - u \leq x)|X > u) = \frac{F_x(X) - F_x(u)}{1 - F_x(u)}\ , for\ all\ X > u \quad (2)$$

The key result in EVT is that as the threshold u $\rightarrow \infty$ $F_u(x)$ converges to GPD, $G_{\xi,\beta}(y)$;

$$G_{\xi,\beta}(x) = \begin{cases} 1-\left(1+\frac{\xi x}{\beta}\right)^{-\frac{1}{\xi}}, & \text{if } \xi \neq 0 \\ 1-\exp\left(-\frac{x}{\beta}\right), & \text{if } \xi = 0 \end{cases} \quad (3)$$

where $\beta > 0$ is the scale parameter; and $x \geq 0$ when $\xi \geq 0$ and $0 \leq x \leq -\frac{}{\frac{\xi}{\beta}}$

In the present study the model was re-specified by using a **three-parameter GPD under EVT framework** and incorporating risk measures such as value at Risk and Expected Shortfall for estimating risk of extreme fire loss in an insurance portfolio. This will enhance the accurancy of the estimates as compared to 2 parameter GPD, with the added location parameter. The three-parameter GPD in this study is specified as follows;

$$G(X; u, \xi, \beta) = \begin{cases} 1-(1+\frac{\xi(x-u)}{\beta})^{-\frac{1}{\xi}}, & \xi \neq 0 \\ 1-exp(\frac{-x-u}{\beta}), & \xi = 0 \end{cases} \quad (4)$$

where X = claim amount or fire loss random variable, u = location parameter β= scale parameter ξ = shape parameter

The generalized Pareto distribution (GPD) approach is based on the idea that EVT holds sufficiently far out in the tails such that we can obtain the distribution not only of the maxima but also of other extremely large observations [4].Nevertheless, this theorem fits our study since we are attempting to analyze rare and large event outcomes, which in this case are catastrophic fires leading to huge insurance loss. Extreme Value Theory (EVT) is concerned with the mathematical and statistical analysis of extreme events hence a perfect fit to the present study.

2.2 Parameter Estimation and Derivation of Three-Parameter GPD

In this study the derivation of the parameters of GDP loss distribution will be done using the maximum likelihood estimation (MLE). MLE is commonly applied for estimation in a variety of problems. According to [1], MLE yields better estimates as compared to other methods like, least squares technique and methods of moments. They argued further that MLE method fully utilizes all information about parameters contained in the data.

Suppose random variables (fire claim amounts) $X_1, X_2, ..., X_n$ form a random sample from a pdf $f(x|\theta)$. We define the joint density function $f(x_1, x_2, ..., x_n|\theta)$ as the likelihood function. The likelihood function depends on the unkown parameter θ (or a vector of parameters), is always denoted as L(θ);

$$L(\theta) = f(x_1, x_2, ..., x_n|\theta)$$

$$l(\theta) = logL(\theta) = log\prod_{i=1}^{n} f(x_i|\theta)$$

$$l(\theta) = \sum_{i=1}^{n} logf(x_i|\theta) \tag{5}$$

The goal of maximum likelihood estimation (MLE) is to find the values of the model parameters that maximizes log likelihood function over the parameter space. Thus, estimating parameter θ with MLE principle gives;

$$\theta_{MLE} = argmax_{\theta} \sum_{i=1}^{n} logf(x_i|\theta) \tag{6}$$

The density function for the three-parameter Generalized Pareto Distribution (GPD) is given by:

$$f(y; u, \xi, \beta) = \begin{cases} \frac{1}{\beta}\left(1 + \frac{\xi(x-u)}{\beta}\right)^{-\frac{1}{\xi}-1}, & \text{if } \xi \neq 0 \end{cases} \tag{7}$$

Here, the tail index is denoted as $\alpha = \frac{1}{\beta}$, and the study focuses on the case where $0 < \xi \leq 1$.

The transformation used is $\frac{z}{u} = 1 + \frac{1}{\beta}(x-u)$ or $Z = \frac{\xi u}{\beta}x + u\left(1 - \frac{\xi u}{\beta}\right)$. This leads to the expression $x = \frac{\beta}{\xi u}z + u - \frac{\beta}{\xi}$, with a Jacobian of $(\frac{\beta}{\xi u})$ for $\xi > 0$. Consequently, the density of the transformed variable follows a Pareto distribution with density $f(z) = \alpha\left(\frac{u^{\alpha}}{z^{\alpha+1}}\right)$ for $z \geq u$ and $\alpha > 0$.

The Type I Pareto distribution is defined as $f(x:\alpha,\lambda) = \alpha\frac{\lambda^{\alpha}}{(x+\lambda)^{\alpha+1}}$ for $z \geq u$ and $x \geq 0$, where α is the shape parameter and λ is the scale parameter.

The log-likelihood function for Maximum Likelihood Estimation (MLE) is given by:

$$L(\alpha,\lambda) = \prod_{i=1}^{n} \alpha \frac{\lambda^{\alpha}}{(x_i)^{\alpha+1}} \tag{8}$$

Maximizing the log-likelihood function entails adjusting λ to $\lambda = \min\{x_i\}$, ensuring that λ is not larger than the smallest value of x in the dataset.

The parameter estimate for α is determined by equating the derivative of the log-likelihood function to zero, yielding:

$$\frac{1}{\alpha} + \log(\lambda) - \frac{1}{n}\sum_{i=1}^{n}\log(x_i) = 0 \tag{9}$$

This results in the expression:

$$\alpha = \frac{n}{\sum_{i=1}^{n}\log\left(\frac{x_i}{\lambda}\right)} \tag{10}$$

For a Generalized Pareto Distribution estimated from a Type I Pareto distribution, the Maximum Likelihood Estimators (MLE) for ξ and u are:

$$\xi = \frac{1}{n}\sum_{j=1}^{n}\log\left(\frac{z_j}{u}\right)$$

$$\beta = z_{(1)}$$

Whare the mean and variance of GPD estimated from these parameters are:

Mean $= u + \frac{\beta}{1-\xi}$ for $\xi < 1$

Variance $= \frac{\beta^2}{(1-\xi)^2(1-2\xi)}$

2.3 Determination of Threshold (u)

Several methods have been proposed to determine the optimal threshold. The most common approach is the eyeball method where we look for a region where the tail index seems to be stable. More formal methods are based on minimizing the mean squared error (MSE) of the Hill estimator (i.e., finding the optimal point), but such methods are not easy to implement [4]. Hence, a more easier and reflective method to determine u, is the mean excess loss function e(u), defied as the average excess of the random variable X over the threshold u [9]. Hence, this study will employ mean excess plot to determine threshold. Here we shall plot the mean excess function (MEF) defined as;

$$e(u) = E[X - u | X > u] \tag{11}$$

$$e(u) = \frac{1}{n_u} \sum_{i=1}^{n_u} (x_i - u) \tag{12}$$

where u is the threshold value, and n_u denotes total number of values that which exceed the threshold in the same line of thought, [6] argued that the mean excess function of the GPD is a linear function of threshold u, they further claimed that the reasonable way to determine the threshold u, is to find values over which the sample mean excess function is approximately linear.

2.4 Generalized Pareto Distribution (GPD) and Risk Measures Estimation

The distribution function of the three-parameter GPD is expressed as:

$$G(y; u, \xi, \beta) = \begin{cases} 1 - \left(1 + \frac{\xi(x-u)}{\beta}\right)^{-\frac{1}{\xi}}, & \text{if } \xi \neq 0 \\ 1 - \exp\left(\frac{-x-u}{\beta}\right), & \text{if } \xi = 0 \end{cases} \tag{13}$$

Let $V = X - u$, which represents the excess loss over the threshold u. The probability that the random variable X exceeds the threshold is given by $1 - F(u)$, and the probability that $X > u + v$ given that $X > u$ is $1 - G_u(V)$.

Therefore, the unconditional probability that $X > u + v$ will be obtained by:

$$F(X > u + v) = [1 - F(u)].[1 - G_u(V)]$$

According to Lee (2012), $[1 - F(u)]$ can be estimated by the empirical estimator $\left(\frac{k}{n}\right)$, where n is the total number of observations and k is the number of observations exceeding the threshold value u. Hence, the probability $Pr(x > u + v)$ will be given by:

$$\frac{n_u}{n} = \frac{k}{n}\left(1 + \xi\left(\frac{x-u}{\beta}\right)\right)^{-\frac{1}{\xi}}$$

This can be further simplified to:

$$= \frac{k}{n}\left(1 + \xi\left(\frac{x-u}{\beta}\right)\right)^{-\frac{1}{\xi}}$$

Such that the tail estimator for the distribution function will be given by:

$$F(x; u, \beta, \xi) = 1 - \frac{k}{n}\left(1 + \xi\left(\frac{x-u}{\beta}\right)\right)^{-\frac{1}{\xi}} \tag{14}$$

2.4.1 Value at Risk (VaR) of GPD

Value at Risk (VaR) is the 100th percentile of a of an insurance portfolio due to the risk of fire loss distribution (Klugman, 2013). It is defined as the will be calculated as follows:

$$F(VaR) = p$$

$$p = 1 - \frac{k}{n}\left(1 + \xi\frac{(VaR - u)}{\beta}\right)^{-\frac{1}{\xi}}$$

$$\frac{k}{n}\left(1 + \xi\frac{(VaR - u)}{\beta}\right)^{-\frac{1}{\xi}} = 1 - p$$

$$\left(1 + \xi\frac{(VaR - u)}{\beta}\right)^{-\frac{1}{\xi}} = \frac{n}{k}(1 - p)$$

$$1 + \xi\frac{(VaR - u)}{\beta} = \left(\frac{n}{k}(1 - p)\right)^{-\xi}$$

$$\xi\frac{(VaR - u)}{\beta} = \left(\frac{n}{k}(1 - p)\right)^{-\xi} - 1$$

Therefore,

$$VaR = u + \frac{\beta}{\xi}\left(\frac{n}{k}(1 - p)\right)^{-\xi} - 1 \quad (15)$$

2.4.2 Expected Shortfall (ES) on GPD

Expected Shortfall (ES), also known as Tail Value at Risk (TVaR), is an extension of VaR. It captures the average loss given that the loss is greater than VaR (Klugman, 2013). At $p\%$ level, it may be defined as the expected loss in the worst $p\%$ cases. For instance, $ES(0.1)$ is the expectation of the worst 10 cases out of 100 cases (Lee, 2012). Mathematically, the expected shortfall of an insurance portfolio due to the risk of fire loss in this study will be calculated as follows:

$$ES_p = E[\text{loss given that loss } > VaR_p]$$

$$= \frac{\int_p^1 VaR_u(X)du}{1 - p}$$

If X is continuous at $VaR_p(X)$, then:

$$ES_p = E[X/X > VaR_p(X)]$$

$$= \pi_p + \frac{\int_{VaR_p(X)}^{\infty}(X - \pi_p)f(x)dx}{1-p}$$

For the GPD case, the expected shortfall will be estimated as follows:

$$ES_p = VaR_p + \frac{\beta + \xi(VaR_p - u)}{1-\xi}$$

$$= \frac{VaR_p(1-\xi) + \beta + \xi(VaR_p - u)}{1-\xi}$$

$$= \frac{VaR_p - \xi VaR_p + \beta + \xi VaR_p - \xi u}{1-\xi}$$

$$= \frac{VaR_p + \beta - \xi u}{1-\xi}$$

Therefore,

$$ES_p = \frac{VaR_p}{1-\xi} + \frac{\beta - \xi u}{1-\xi} \tag{16}$$

2.4.3 Study Assumptions

The study assumptions encompass three key aspects: firstly, the continuous and independently identically dis tributed nature of claim amounts (X_s, $s > 0$); secondly, the absence of claim handling expenses in the data, focusing solely on paid fire losses; and thirdly, the adherence to extreme value theory (EVT) principles, indicating that the data is not normally distributed. In accordance with this theory, we therefore tested normality of data, and we found out that our data was indeed not normally distributed hence consistent for modelling.

III. EMPIRICAL RESULTS AND DISCUSSIONS

This section presents and discusses the empirical findings of the analysis and modelling of large fire insurance claims and estimation of risk measures based on three parameter GPD under extreme value theory. For analysis of the modelling procedure the study used quantitative claim size data collected from RBM and NICO ranging from 2011 to 2021. Some basic data cleaning was applied using Ms Excel where claim records with missing values and duplicates were removed. The data obtained were exported to R statistical package for further analysis.

3.1 Exploratory Data Analysis

Descriptive analyses are important since they provide the foundation upon which correlation and experimental studies emerge. They also provide clues regarding the issues that should be focused on to help for further studies [12]. Therefore, the descriptive statistics of the data used in this study is calculated to get a fair idea about the data before the analysis. Table 1 presents the Number of Observations, Mean, Kurtosis, Skewness, Minimum and Maximum Values of claim amounts paid. There are 846 observations of the variable claim amount. The average claim payment is MK2,168,100. Minimum and Maximum observations are MK25,000 and MK59,113,780 respectively. The value of the Skewness is greater than zero, this indicate the existence of fat tails/heavy tails in the data. There exist large observations (outliers) in the right tail of the distribution curve which is consistent with the underlying assumption of Extreme Value Theory (EVT). Likewise, Kurtosis is greater than 3, this means data is not normally distributed hence we can proceed for analysis based on the assumptions of the EVT. 13

Mean	2168.10
Standard error	117.95
Median	1060.46
Mode	2500.00
Standard deviation	3430.77
Sample variance	11770205.49
Kurtosis	96.94
Skewness	7.31
Range	59088.78
Minimum	25.00
Maximum	59113.78
Count	846.00

Table 1: Descriptive Statistics

Besides noting that our data is not normally distributed from descriptive statistics, we further tested for normality using a diagnostic test, Pearson chi-square test for normality at 5% significance level. We therefore obtained a p-value less than the significance level leading to rejecting null hypothesis. The test revealed that our data was indeed not normally distributed and inline with EVT assumption as shown in Table 2 below.

Table 2: Normality Test

Test	Hypothesis	Alternative Hypothesis	Test Statistic	P-Values
Pearson Chi-square	Data is Normal	Data not Normal	1623.9	2.2×10^{-16}

In Figure 1, we observe that fire loss data is skewed to the right and uneven variability of claim observations as seen on x-axis,(that is, data being heteroskedastic). This gives the insurance company a hard time to predict expected losses and returns as they are operating on extreme values of making a profit or loss hence more risky in case of huge fire loss. We also observe that the distribution of claims is positively skewed, suggesting that small fire losses occur quite frequently and very large losses occur less frequently but they are catastrophic.

According to [4], a more useful visualization of data can be obtained using logarithmic scale for the x-axis (or even both axes). This is performed by plotting the Empirical Complementary Cumulative Distribution Function (ccdf), that is, the empirical probability of the claims exceeding any given threshold, sometimes also referred to as the Survival function as shown in Figure 2. This (Figure 2) shows the empirical distribution of claim settlements. The tails are nonlinear implying the Pareto behaviou (power law) and extreme value theory can be confirmed. Extreme Value Models are known to be heavy tailed. According to EVT theory, extreme observations in a given sample data follow the tail of an EVT distribution called Generalized Pareto Distribution GPD. Hence, there is a possibility that our data of large fire insurance claims follow the tail of a GPD as evidenced by the Pareto behaviour in the data.

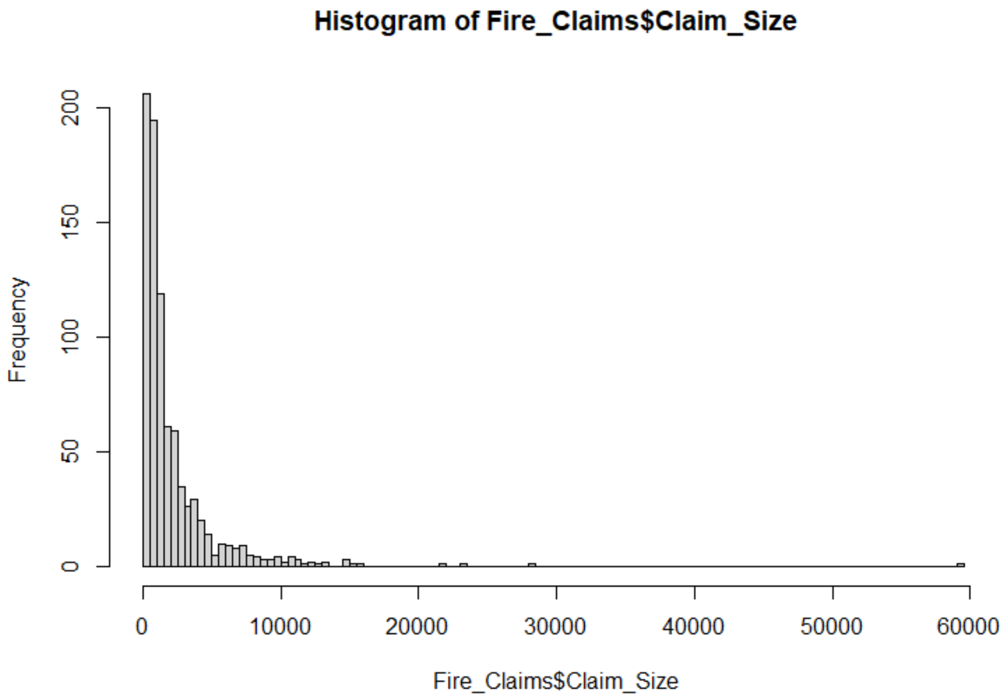

Figure 1: Histogram of Fire Losses

3.2 The Mean Excess Function and the Determination of the Threshold

Now that we established that the data is fat-tailed and follows a power law, we turn to fitting a GPD distribution to the threshold exceedances. However, before performing that we need to determine an appropriate threshold (a starting point where we shall assume any claim beyond it as being extreme/large). In the figure 3 below, we plotted Mean Excess Function to confirm convergence of GPD to any given threshold. The resulting plot looks fairly linear across the whole spectrum of losses. Nonetheless, a small kink just below MK5, 000,000 is observed indicating that smaller losses follow a somewhat different law/distribution. A fairly linear region can be observed between 0 and MK5,000,000; above MK5,000,000 the data becomes sparse. Therefore, a threshold of MK5,000,000 was considered as a reasonable choice that is consistent with Mean Excess Function under EVT.

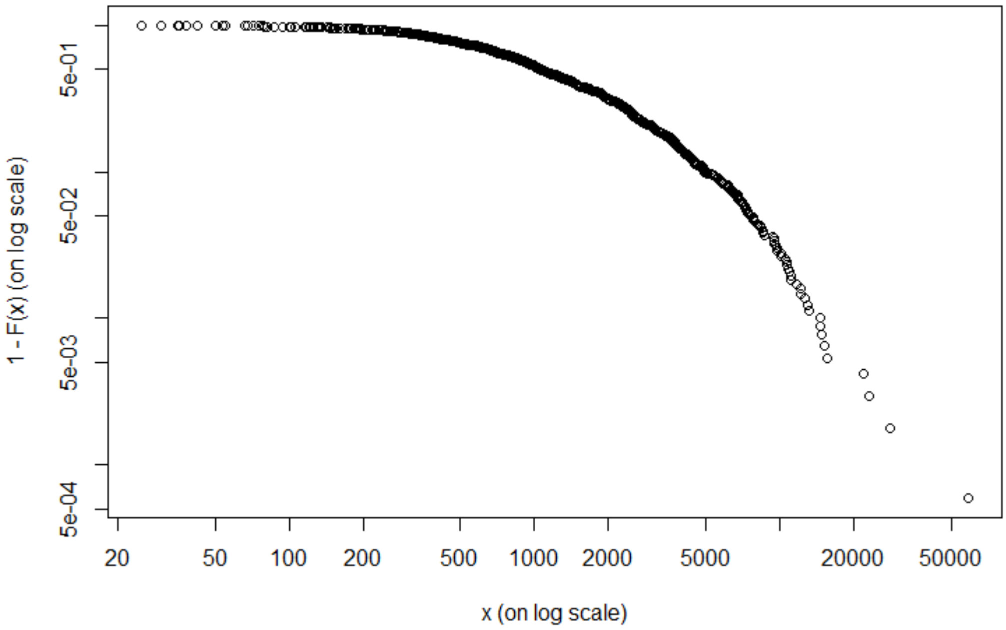

Figure 2: Complementary Cumulative Distribution Function

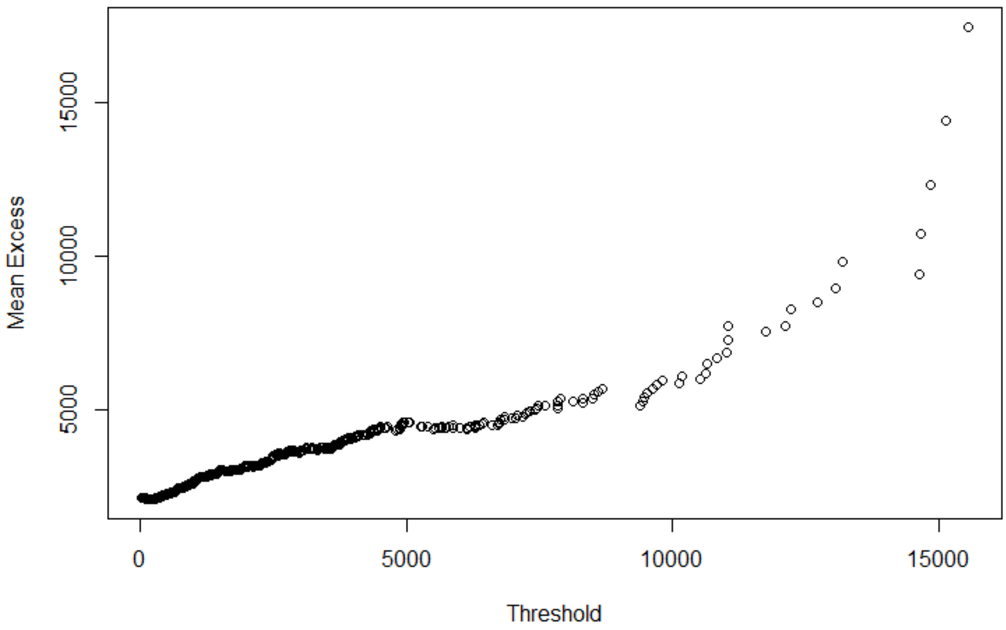

Figure 3: Mean Excess Function Plot against Thresholds

Since the empirical distribution was non-linear implying the Pareto behaviour, we were justified to fit a Generalised Pareto Distribution (GPD) to the tails of fire loss data. The data was fitted to a GPD model using Maximum Likelihood Estimate as shown in Figure 4. The parameter estimates are U = 2167, ξ = 0.1938 and β = 3612. 76 as shown in Table 3. The shape parameter ξ is greater than 0 implying heavy tailed distribution followed by the data. The distribution for the excesses shows a smooth curve (see

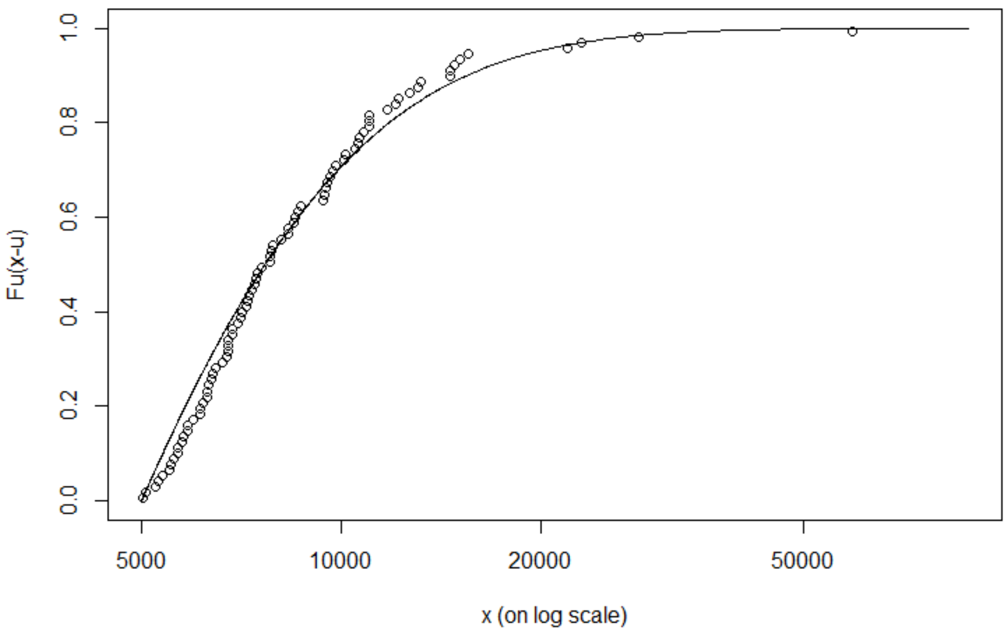

Figure 4: Excess Distribution and Fitted GPD

Table 3: Model Parameters Estimation

Model	Location (U)	Shape (ξ)	Scale (β)
G(X: U, ξ, β)	2167	0.1938	3612.76

Therefore, the estimated three-parameter GPD model using MLE method becomes;

$$G(X; u, \xi, \beta) = \begin{cases} 1 - \left(1 + \frac{0.1938(x-2167)}{3612.76}\right)^{-\frac{1}{0.1938}}, & \text{if } \xi \neq 0 \\ 1 - \exp\left(\frac{-(x-2167)}{3612.76}\right), & \text{if } \xi = 0 \end{cases} \quad (17)$$

To measure the goodness of fit between the observed data and the theoretical distribution. The quantile-quantile plot (QQ-plot) serves as a valuable tool for initially evaluating the suitability of a parametric distribution's fit. In the context of Financial and Insurance scenarios, where data sets often exhibit fat-tailed behavior, the QQ-plot proves particularly useful [4]. A well-suited parametric distribution should lead to a linear graph in the QQ-plot, as demonstrated in Figure 5. This visualization verifies the excellent fit of the Generalized Pareto Distribution (GPD) to the data, substantiating its efficacy for predictions. Furthermore, the QQ-plot aids in the identification of outliers within the dataset.

Additionally, an Anderson-Darling (AD) test was performed to assess the goodness of fit as shown in Table 4. The resulting p-value exceeded the 5% significance level, leading to the non-rejection of the null hypothesis. This outcome supports the conclusion that the data indeed adheres to the GPD within the context of Extreme Value Theory (EVT). The establishment of a good fit is pivotal; it empowers us to infer that the insurer is equipped to manage substantial claims by leveraging risk estimates, thus ensuring financial stability even in the face of significant challenges.

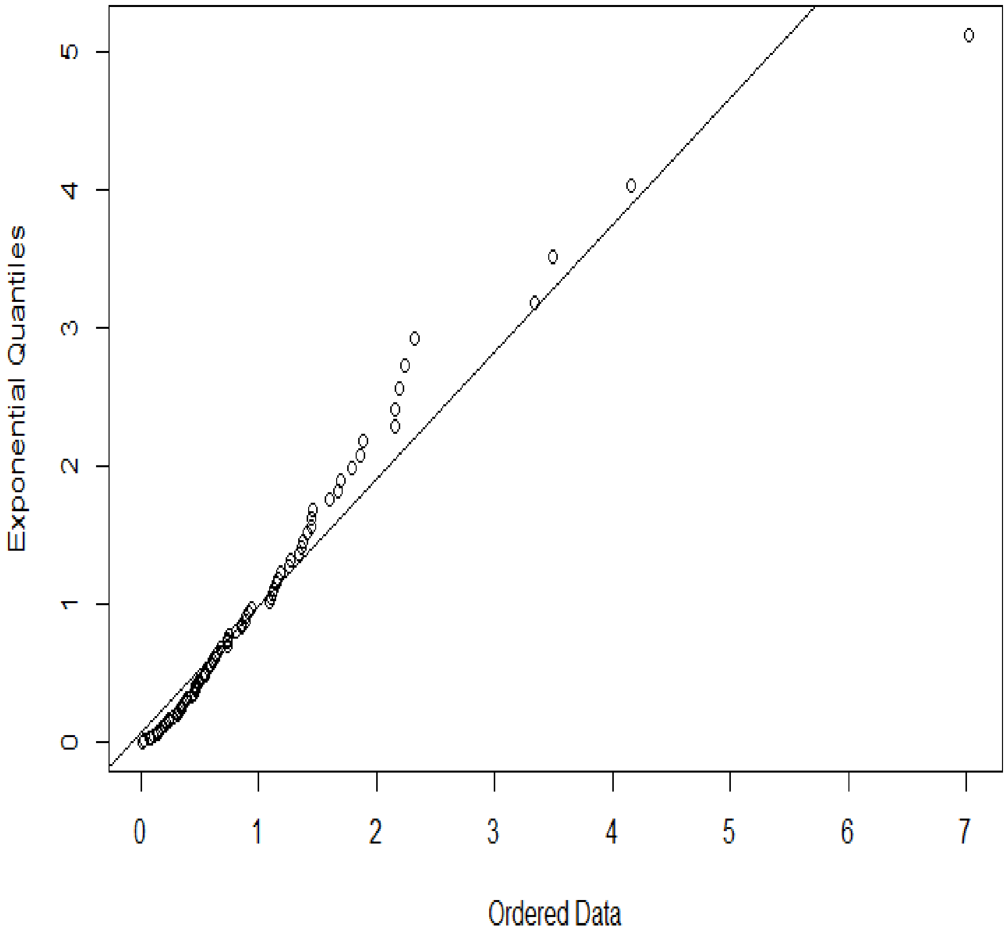

Figure 5: Quantile-Quantile Plot (QQ-Plot) of Residuals

Table 4: Goodness of Fit Test

Test	Hypothesis	Alternative Hypothesis	Test Statistic	P-Values
A-D	Data follow GPD	Data does not follow GPD	4.02	0.659

3.3 Risk Measures

With the GPD model successfully fitted to fire loss data, we can now employ it to assess the risk associated with higher quantiles, as well as insurers' capital value at risk, and the expected losses within a specified time frame. For instance, our estimated 99% Value at Risk stands at MK15,437,440 in the most unfavorable scenario, accompanied by a lower interval of MK13,156,630 and an upper interval of MK19,201,590. This prediction encompasses the projected loss that the company should anticipate in their claims handling for the upcoming fiscal year.

This holds notable significance within the realm of operational risk management, particularly in adherence to regulatory requirements. Calculating extremely high quantiles (99%) aids in evaluating potential losses and gauging the company's solvency under exceedingly adverse conditions. Consequently, such assessments help in determining the company's resilience in the face of catastrophic scenarios, allowing for early detection and necessary action. As depicted in Figure 6, the Tail Loss alongside the Estimated 99% VaR exemplifies these insights effectively.

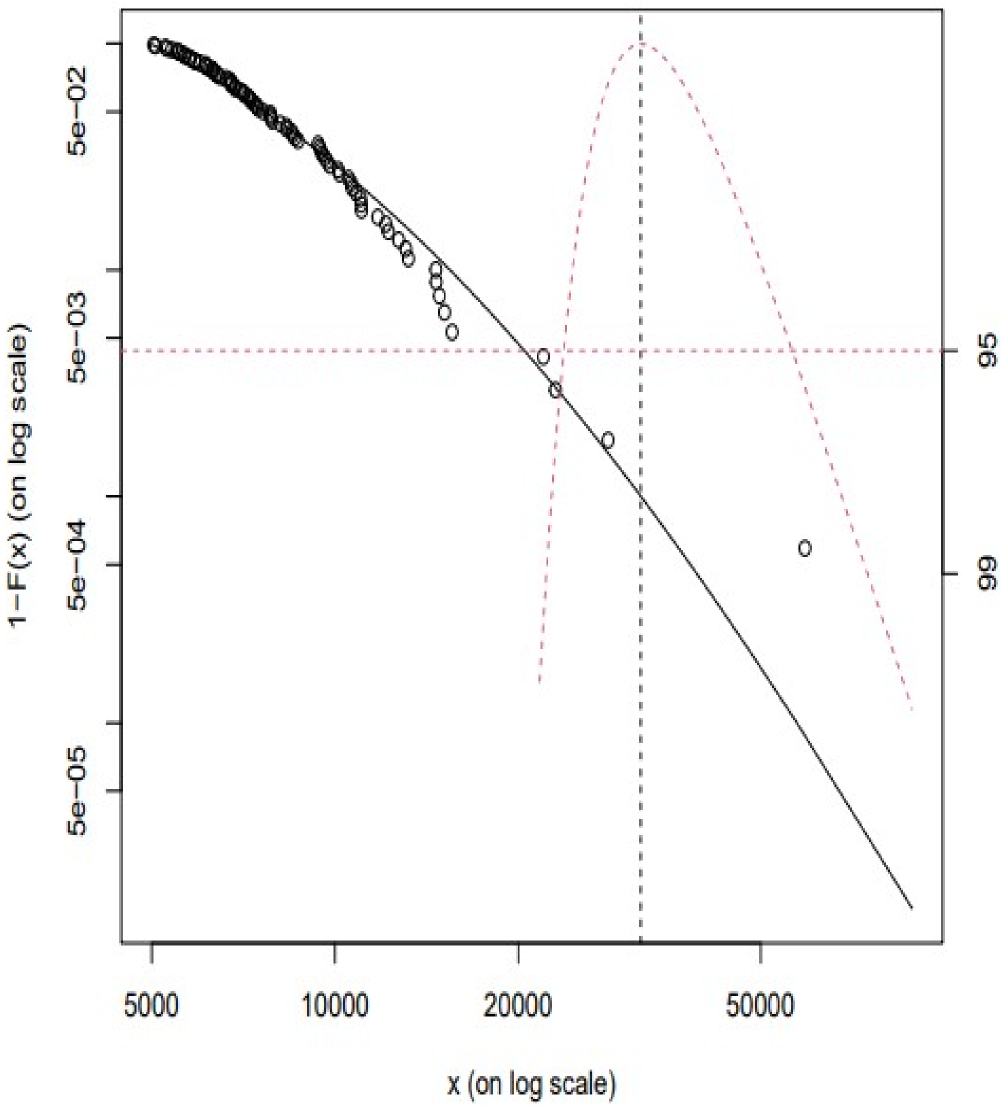

Figure 6: Tail Loss with Estimated 99% Value at Risk (VaR)

Even though Value at Risk (VaR) offers an estimation of the highest potential loss at the designated confidence level, i.e. 99%, it lacks insight into the scale of losses that could surpass the VaR threshold. Expected Shortfall (ES) overcomes this drawback by assessing the average value of all losses surpassing the VaR threshold. ES tell us about the average expected loss given that the loss amount is greater than VaR with certain probability level. For example, ES(0.05) with confidence level of 95% is defined as the expectation of the worst 5 cases out of 100 cases provided the loss is greater than VaR(0.05). Using a confidence level of 95% the ES is MK12,762,850 revealing that there is 5% probability that minimum loss would be equal to MK12,762,850 or greater or we are 95% confident that the maximum loss would be equal MK12,762,850 or less if loss exceeds VaR calculation in our next financial year.Table 5 below, reports the estimates of VaR and expected losses at different confidence levels as estimated by a three-parameter GPD.

Table 5: Value at Risk and Expected Loss Estimates (Interval Estimate (MK'000))

VaR		Lower CI	Estimate	Upper CI
	90%	4853.09	4970.04	5302.19
	95%	7038.21	7645.84	8439.94
	99%	13156.63	15437.44	19201.59
ES	90%	8469.46	9443.91	11202.59
	95%	11065.51	12762.85	16145.44
	99%	17923.64	22427.06	35467.68

The estimated 95% VaR is MK7,645,840 and the estimated 95% Expected shortfall is MK12,762,850. This means that assuming that 95% VaR level of MK7,645,840 is exceeded then the expected loss is MK12,762,850. The resulting graph displays both the 95% VaR (first vertical dashed line and its profile likelihood curve) and 95% Expected shortfall.

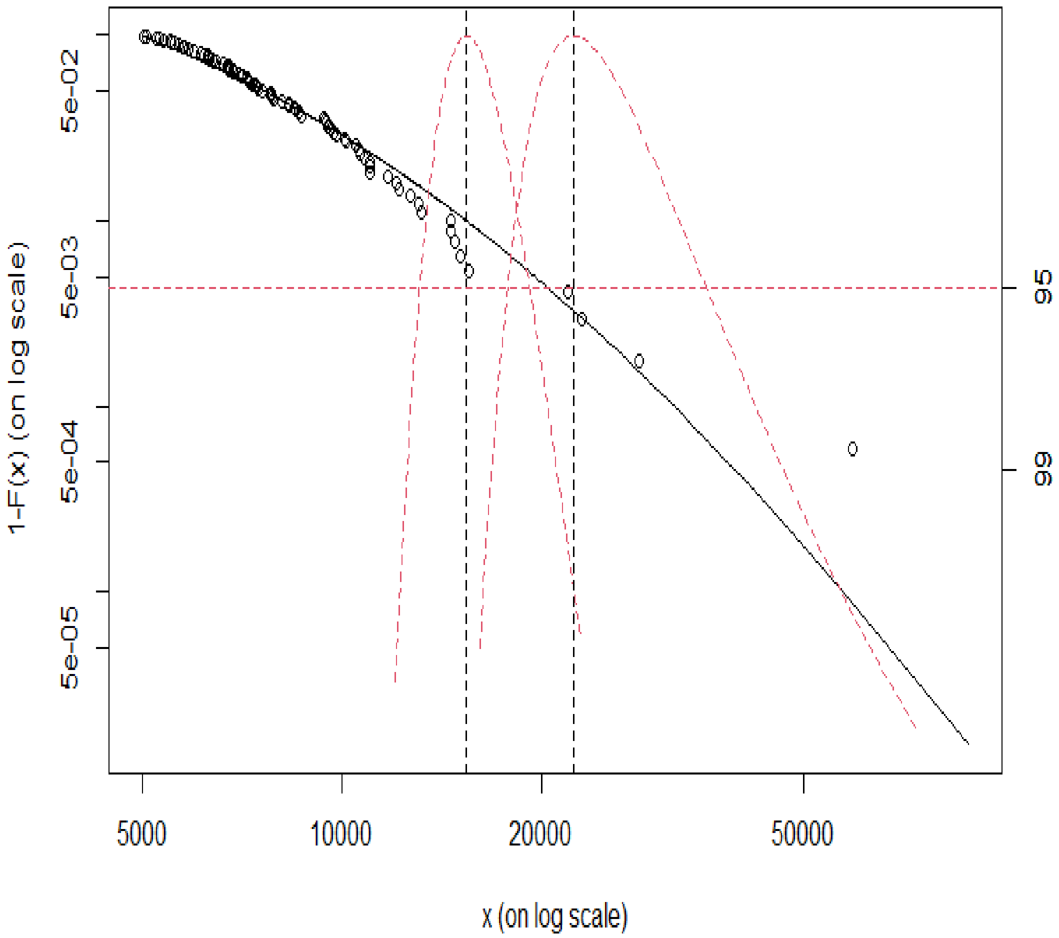

Figure 7: Estimate of 95% Value at Risk and Expected Shortfall

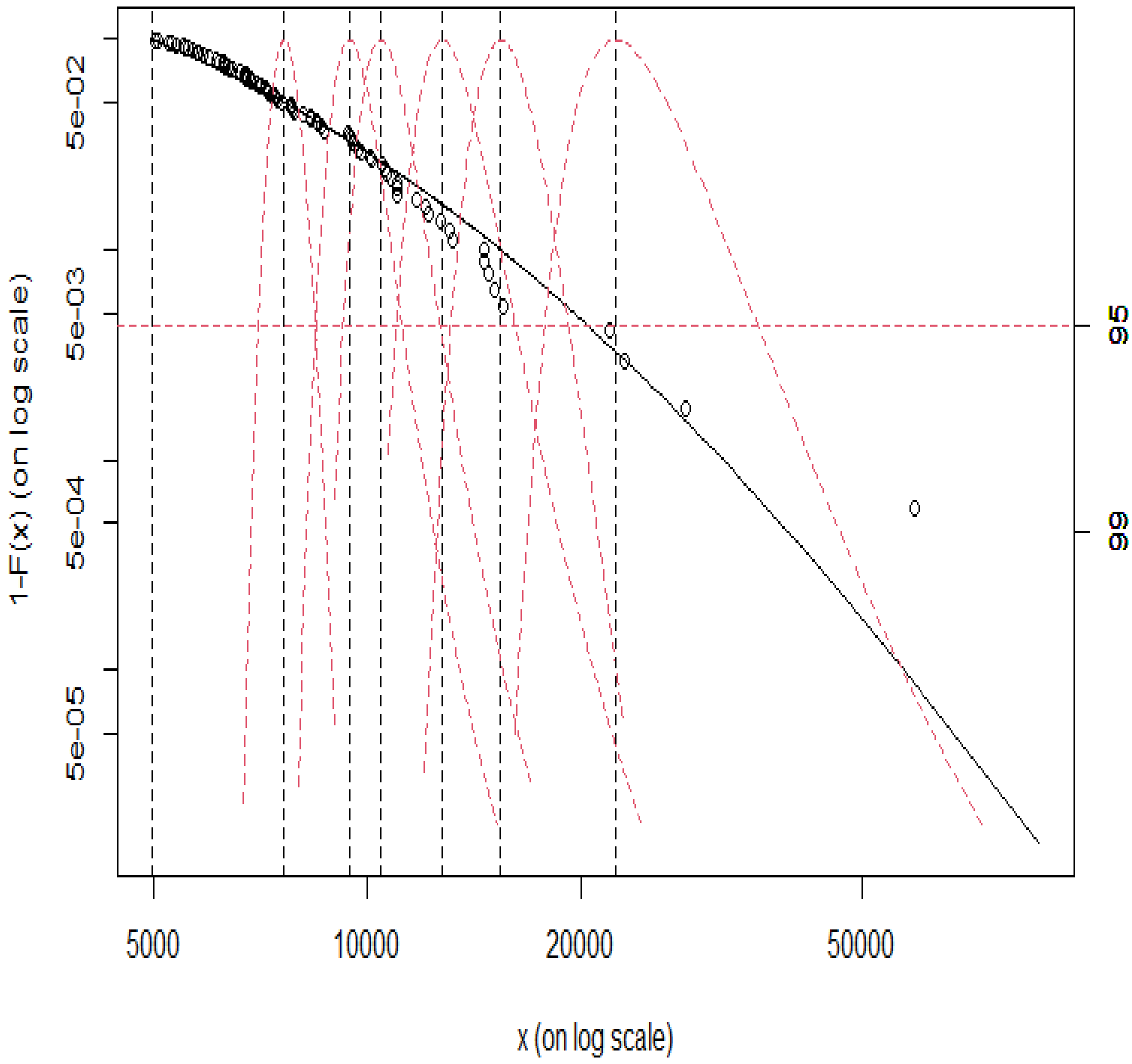

Figure 8: Estimates of 90%, 95%, and 99% Value at Risk and Expected Shortfalls

IV. CONCLUSION AND RECOMMENDATIONS

4.1 Conclusion

The study's primary objective was to develop a mathematical model employing Extreme Value Theory (EVT) and Risk Measures to estimate and predict significant fire insurance claims. This aimed to offer insurance companies a more precise grasp of potential risks linked to substantial fire-related losses. The study established a three parameter Generalized Pareto Distribution (GPD) within the framework of EVT for estimating insurer risk due to catastrophic fire incidents. The significance of assessing fire-related financial losses for insurers was highlighted, es pecially considering the infrequent but influential extreme events that can distort overall loss patterns. By applying EVT techniques, including the GPD and Peaks Over Threshold (POT) method, to a dataset of historical fire insur ance claims, the study accurately modeled the tail behavior of large losses. Parameters derived from these models enabled the calculation of probabilities for extreme loss events, thereby supporting enhanced risk management and pricing strategies for insurance companies. The findings showcased the efficacy of the EVT approach in accurately modeling and estimating the risk associated with substantial fire insurance claims. The research contributes to the insurance field by offering an improved mathematical and statistical framework for modeling significant fire insurance claims. Through this approach, insurers can better comprehend the potential financial consequences of rare fire incidents, facilitating more informed risk assessment and resource allocation. In summary, this study's methodology and findings have significant implications for managing and mitigating the risks associated with fire related losses in the insurance sector.

In terms of modeling significant fire insurance claims through the extreme value approach, our recommendations encompass several key points. It is advisable for insurers to employ the three-parameter Generalized Pareto Distri bution (GPD) within this approach, enabling a robust assessment of risk associated with substantial fire insurance claims. Moreover, insurance companies are encouraged to allocate capital for potential large losses, guided by the Value at Risk (VaR) and Expected Shortfall metrics calculated from the three-parameter GPD under the Ex treme Value Theory framework. Leveraging insights derived from this methodology, companies can then formulate well-informed strategies for pricing and capital reserving that align with the distinct risk profile of fire insurance policies. In addition, we recommend that, future research should include other methods for estimating VaR of fire loss such as Historical method and compare results to Extreme VaR and also, should extend univariate EVT models to multivariate EVT models that may capture a broad spectrum of covariates of fire occurence influencing fire losses to commercial property.

ACKNOWLEDGEMENTS

The authors would like to thank the Mathematical Sciences department of the University of Malawi.

List of Abbreviations

The following table below lists abbreviations in the manuscript:

Abbreviation	Meaning
EVT	Extreme Value Theory
GPD	Generalised Pareto Distributions
POT	Peaks Over Threshold
GLM	Generalised Linear Models
GEVD	Generalised Extreme Value Distribution
ES	Expected Shortfalls
VaR	Value at Risk
RBM	Reserve Bank of Malawi
MLE	Maximum Likelihood Estimation

MSE	Mean Square Error
MEF	Mean Excess Function
ccdf	Complementary Cumulative distribution function
AD	Anderson Darling test
Q-Q plot	Quantile-Quantile plot

Declarations

Availability of Data

Data will be made available upon request.

Competing interests

The authors declare that there are no competing interests.

Funding

The authors received no funding from any any agency.

Author's contributions

The manuscript concept design and supervision was done by Nelson Dzupire, coding and analysis was done by Dan Kachusa, and finally the manuscript development was done by Samuel Gyamerah.

Ethical Approval section

This is not applicable.

REFERENCES

1. Jean-Philippe Boucher, Michel Denuit, and Montserrat Guill´en. Risk classification for claim counts: a com parative analysis of various zeroinflated mixed poisson and hurdle models. *North American Actuarial Journal*, 11(4):110–131, 2007.
2. Anna Chernobai, Krzysztof Burnecki, Svetlozar Rachev, Stefan Truck, and Rafał Weron. Modelling catastrophe ¨ claims with left-truncated severity distributions. *Computational Statistics*, 21:537–555, 2006.
3. Louis Anthony Cox, Jr. Community resilience and decision theory challenges for catastrophic events. *Risk Analysis: An International Journal*, 32(11):1919–1934, 2012.
4. Jon Danielsson. *Financial risk forecasting: The theory and practice of forecasting market risk with implementa tion in R and Matlab*. John Wiley & Sons, 2011.
5. Georges Dionne and Denise Desjardins. A re-examination of the us insurance market's capacity to pay catas trophe losses. *Risk Management and Insurance Review*, 25(4):515–549, 2022.
6. Souvik Ghosh and Sidney Resnick. A discussion on mean excess plots. *Stochastic Processes and their Applica tions*, 120(8):1492–1517, 2010.
7. Sarah B Henderson, Kathleen E McLean, Michael J Lee, and Tom Kosatsky. Analysis of community deaths dur ing the catastrophic 2021 heat dome: Early evidence to inform the public health

response during subsequent events in greater vancouver, canada. *Environmental Epidemiology*, 6(1), 2022.
8. Owen Jakata and Delson Chikobvu. Extreme value modelling of the south african industrial index (j520) returns using the generalised extreme value distribution. *International Journal of Applied Management Science*, 14(4):299–315, 2022.
9. Stuart A Klugman, Harry H Panjer, and Gordon E Willmot. *Loss models: from data to decisions*, volume 715. John Wiley & Sons, 2012.
10. Christian Laudag´e, Sascha Desmettre, and Jorg Wenzel. Severity modeling of extreme insurance claims for ¨ tariffication. *Insurance: Mathematics and Economics*, 88:77–92, 2019.
11. Franc¸ois M Longin. The asymptotic distribution of extreme stock market returns. *Journal of business*, pages 383–408, 1996.
12. Olive Mwihaki Mugenda and Abel Gitau Mugenda. *Research methods: Quantitative & qualitative apporaches*, volume 2. Acts press Nairobi, 2003.
13. James Kiprotich Ng'elechei, Joel Cheruiyot Chelule, Herbert Imboga Orango, and Ayubu Okango Anapapa. Modeling frequency and severity of insurance claims in an insurance portfolio. *American Journal of Applied Mathematics and Statistics*, 8(3):103–111, 2020.
14. Cyprian Ondieki Omari, Shalyne Gathoni Nyambura, and Joan Martha Wairimu Mwangi. Modeling the fre quency and severity of auto insurance claims using statistical distributions. 2018.
15. Wan-Kai Pang, Shui-Hung Hou, Marvin D Troutt, Wing-Tong Yu, and Ken WK Li. A markov chain monte carlo approach to estimate the risks of extremely large insurance claims. *International Journal of Business and Economics*, 6(3):225, 2007.
16. Gabriel Raviv, Barak Fishbain, and Aviad Shapira. Analyzing risk factors in crane-related near-miss and accident reports. *Safety science*, 91:192–205, 2017.
17. Swiss Re. World insurance: Riding out the 2020 pandemic storm. *sigma*, 4:1–34, 2020.
18. Lukasz Scislo. High activity earthquake swarm event monitoring and impact analysis on underground high energy physics research facilities. *Energies*, 15(10):3705, 2022.
19. Lin Wang and Ali M Kutan. The impact of natural disasters on stock markets: Evidence from japan and the us. *Comparative Economic Studies*, 55:672–686, 2013.
20. Simon Kinyua Weru, Antony Waititu, and Antony Ngunyi. Modelling energy market volatility using garch models and estimating value-at-risk. *Journal of Statistics and Actuarial Research*, 2(1):1–32, 2019.