Journal of Statistics Applications & Probability
An International Journal

http://dx.doi.org/10.18576/jsap/150204

Composite Distributions and their Associated Risk Measures for Auto-mobile Insurance Claims Data

Williams Kumi[1,2,*], *Henry Otoo*[1], *Charles Kwofie*[2], *and Sampson Takyi Appiah*[2]

[1]Department of Mathematical Sciences, University of Mines and Technology, Tarkwa, Ghana
[2]Department of Mathematics and Statistics, University of Energy and Natural Resources, Sunyani, Ghana

Received: 12 Aug. 2025, Revised: 3 Oct. 2025, Accepted: 05 Nov. 2025
Published online: 1 Mar. 2026

Abstract: Insurance losses comprises of small and large claims rendering single distributions incapable of holistically capturing the different sizes together accurately. Risk measures associated with insurance losses are crucial for determining reserve levels and for assessing solvency. Hence, in estimating risk measures, the right probabilistic distributions has to be carefully fitted in order not to underestimate or over estimate associated parameters. In view of this, this paper employs a two component composite distribution to describe automobile insurances losses from Ghana using 11,879 data points. This research fitted 240 composite distributions and results of the top ten are selected and presented based on some goodness of fit criteria. Threshold values and mixing weights for each composite distribution are also estimated and presented. Value at Risk and Tail value at Risk are then estimated and presented for the top ten composite distributions at 95% and 99% security levels.

Keywords: Composite distribution, auto-mobile insurance, claims, risk measures, threshold, mixing weight.

1 Introduction

Risk modelling associated with insurance losses is an integral part of the work of an Actuary as it aids in estimating reserves and other crucial elements such as expected claims and variances. To this end, it is essential to employ the right probabilistic distribution that properly capture the dynamics in claims data in order to accurately estimate desired risk measures. While risk modelling help underwriters and actuaries to understand and predict claim frequency and severity, it also aid the insurer to estimate potential losses [1]. According to [2], risk modelling assist insurers to set adequate premiums by estimating expected losses to ensure that right premiums are charged to cover indemnities. More importantly, risk modelling guide the insurer to adequately strike a healthy balance between reserve and amount to be invested [3]. If the probabilistic model fitted to help estimate the risk measure mimic reality, it can be used to accurately predict losses which eventually help to minimize risk [4]. On a more general level, risk modelling also helps the insurer identify potential risk factors which assists in risk management practices through underwriting. Consequently, this helps the insurer adopt some risk mitigation strategies to ensure stability of the insurance industry in the face of potential losses. Furthermore, risk modelling provide insights for the insurance industry to strategise their business decisions especially on their product development, market expansion, and investment strategies.

When insurance losses are not modelled with the right probabilistic distributions, it leads to under reserving or over reserving. If an insurer reserves more than necessary, this subsequently affects other areas of the company as huge sum of money will be allocated to settle claims. This can consequently affect financial inflows of the insurance company such as investment income. On the other hand, if the insurer reserve less than necessary, it will be unable to settle claims as and when they arise. To this end, it is important to fit appropriate probabilistic distributions to claims data in order to mitigate issues that arise from over reserving or under reserving. This will in turn help the insurance company meet its obligation of paying claims. An insurance company inability to properly fit claims amount using an appropriate probabilistic distributions can plunge the company into insolvency if not properly handled [5].

* Corresponding author e-mail: williams.kumi@uenr.edu.gh

Single distributions have been used extensively to model loss data (see [6]). However, given that loss data comprise of small and large losses, single distributions cannot properly capture the underlying dynamics of the different sizes leading to under estimation or over estimation of the required metrics useful for making prudent decisions. In sequel, this can subsequently lead to under reserving and other related consequences. In view of this, although now gaining momentum, composite distributions have been employed to model losses (see [7], [8], [9]) due to its ability to piece together distributions that reflect different sizes of claims together in a novel manner. This in essence captures the tail behaviour of the relatively large losses in the data alongside the small and medium claims. In composite distributions, two or more distributions (normally called head and tail distributions) are pieced together at some threshold [7]. One peculiar question that however arises is how this threshold is optimally chosen in order to maximise the parameters of the distributions pieced together while ensuring that the composite distribution properly mimics the underlying data. In the case of insurance, composite models have been proven to model insurances loss well than single distributions.

On important reason for determining the probabilistic distribution of an insurance data is to compute associated risk measures. Risk measures have been used extensively in the insurance industry to know on average the possible loss an insurance company is expected to pay on any given day. These risk measures guide the insurance industry on reserve amount and also provide quantitative structure that can be used to assess and manage potential losses that can render the company insolvent. Reserving is a very key aspect of the insurance industry and as such it is expected of Actuaries and underwriters to ensure that right tools are employed in order to churn out precise estimates for reliable decision making [10]. This is achievable when risk measures at various percentiles of claims are well calculated as they inform the reserve amount in order not to render the company insolvent. Risk measures guide insurance companies to allocate enough capital to cover potential losses [11],[12]. This will ensure solvency of the insurance industry whilst protecting policyholders as well. Again, risk measures inform premium pricing decisions as it aids insurers to quantify risk of losses so that adequate premium is charged for the corresponding loss amounts [13]. Risk measures may also be used as insurance regulatory compliance metrics such as Solvency II [14]. Although there are several risk measures, this study considered two risk measures; Value at Risk (VaR) and Tail value at Risk (TVaR). Whereas VaR computes the average worst possible loss within any single day, TVaR on the other hand focuses on more severe losses which helps insurers to manage tail risk and catastrophic losses [15]. VaR and TVaR impacts the insurers capital planning decisions by looking at the potential impact of extreme events on insurer ability to pay claims [16]. TVaR aids to design a comprehensive enterprise risk management structure by addressing potential losses beyond expected levels which can impede the core activities of paying claims. In summary, VaR and TVaR are very vital risk measures in insurance reserving because they aid insurers manage potential losses, allocate capital, inform premium pricing, assists in reserve estimation and ensure regulatory compliance. The challenge is how ensuring that probabilistic distributions used in estimating these risk measures capture the dynamics in the data in a holistic manner.

This current study is motivated by the work of [7] where 256 composite models were evaluated for Danish fire loss data and South African taxi claims data. This study also considers 240 composite models for auto mobile insurance claims from Ghana. The study aims to discover composite models that have not been studied previously for auto mobile losses. The study also estimated the differentiability and continuity condition for composite distributions.

2 Methods

2.1 Composite Model

Composite models was first introduced by [17]. The probability density function of the composite model which was adapted by [18] is expressed in the form:

$$g(\alpha_1,\alpha_2,\gamma,\beta) = \begin{cases} \frac{1}{1+\beta} g_1^*(y|\alpha_1,\gamma), \; if\; 0<y\leq\gamma, \\ \frac{\beta}{1+\beta} g_2^*(y|\alpha_2,\gamma), \; if\; 0<y<\infty, \end{cases} \tag{1}$$

whereas the cdf can be written as;

$$G(\alpha_1,\alpha_2,\gamma,\beta) = \begin{cases} \frac{1}{1+\beta}\frac{G_1(y|\alpha_1)}{G_1(\gamma|\alpha_1)}, & if\; 0<y\leq\gamma, \\ \frac{1}{1+\beta}[1+\beta\frac{G_2(y|\alpha_2)-G_2(\gamma|\alpha_2)}{1-G_2(y|\alpha_2)}], & if\; \gamma<y<\infty \end{cases} \tag{2}$$

where γ represent the threshold and β the mixing weights. α_1 and α_2 represent the parameters for the first and second distributions. Also, $g_1^* = \frac{g_1(y)}{G_1(\gamma)}$ for $0 < y < \gamma$ and $g_2^* = \frac{g_2(y)}{G_2(\gamma)}$ for $\gamma < y < \infty$
To ensure a smooth composite distribution at the threshold γ, differentiability and continuity conditions are imposed. For continuity at the threshold, $f(\gamma^-) = f(\gamma^+)$ and consequently, must hold if the following condition is satisfied;

$$c = \frac{f_2(\gamma)F_1(\gamma)}{f_2(\gamma)F_1(\gamma) + f_1(\gamma)(1 - F_2(\gamma))} \tag{3}$$

For differentiability condition, the following condition must be satisfied;

$$c = \frac{f_2'(\gamma)F_1(\gamma)}{f_2'(\gamma)F_1(\gamma) + f_1'(\gamma)(1 - F_2(\gamma))} \tag{4}$$

Solving 3 and 4 ensures that the *pdf* of a composite distribution is always differentiable and continuous if $\frac{f_1(\gamma)}{f_2(\gamma)} = \frac{f_1'(\gamma)}{f_2'(\gamma)}$ is satisfied.

Next, we present some few parametric loss distributions known in literature and their corresponding density functions. This is in presented in Table 1 below

Table 1: Parametric loss distributions and their parameters

Distribution	Parameters	PDF
Burr	$\alpha > 0, \beta > 0,$ $\gamma > 0$	$\frac{\alpha\beta(\frac{x}{\gamma})^{\beta}}{x[1+(\frac{x}{\gamma})^{\beta}]^{\alpha+1}}$
Exponential	$\eta > 0$	$\frac{e^{-\frac{x}{\eta}}}{\eta}$
Gamma	$\xi > 0, \zeta > 0,$	$\frac{(\frac{x}{\zeta}^{\xi})e^{-\frac{x}{\zeta}}}{x\Gamma(\xi)}$
Generalised Pareto	$\psi > 0, \rho > 0,$ $\sigma > 0$	$\frac{\Gamma(\psi+\rho)}{\Gamma(\psi)\Gamma(\rho)} \frac{\sigma^{\psi} x^{\rho-1}}{(x+\sigma)^{\psi+\rho}}$
Inverse Burr	$\alpha > 0, \beta > 0,$ $\gamma > 0$	$\frac{\alpha\beta(\frac{x}{\gamma})^{\beta\alpha}}{x[1+(\frac{x}{\gamma})^{\beta}]^{\alpha+1}}$
Inverse Exponential	$\eta > 0$	$\frac{\eta e^{-\frac{\eta}{x}}}{x^2}$
Inverse Gamma	$\xi > 0, \zeta > 0,$	$\frac{(\frac{x}{\zeta})^{\xi} e^{-\frac{\zeta}{x}}}{x\Gamma(\xi)}$
Inverse Gaussian	$\mu > 0, \tau > 0,$	$(\frac{\tau}{2\pi x^3})^{\frac{1}{2}} e^{-\frac{\tau z^2}{2x}},$ $z = \frac{x-\mu}{\mu}$
Inverse Paralogistic	$\omega > 0, \kappa > 0,$	$\frac{\omega^2(\frac{x}{\kappa})^{\omega^2}}{x[1+(\frac{x}{\kappa})^{\omega}]^{\omega+1}}$
Inverse Pareto	$\lambda > 0, \delta > 0,$	$\frac{\lambda\delta x^{\lambda-1}}{(x+\delta)^{\lambda+1}}$
Inverse Weibull	$\varepsilon > 0, \nu > 0,$	$\frac{\varepsilon(\frac{\nu}{x})^{\varepsilon} e^{-(\frac{\nu}{x})}}{x}$
Loglogistic	$\iota > 0, \tau > 0,$	$\frac{\iota(\frac{x}{\tau})^{\iota}}{x[1+(\frac{x}{\tau})^{\iota}]}$
Lognormal	$\mu > 0, \sigma > 0,$	$\frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{z^2}{2}} = \frac{\phi(z)}{\sigma x}$ $z = \frac{lnx-\mu}{\sigma}$
Paralogistic	$\omega > 0, \kappa > 0,$	$\frac{\omega^2(\frac{x}{\kappa})^{\omega}}{x[1+(\frac{x}{\kappa})^{\omega}]^{\omega+1}}$
Pareto	$\lambda > 0, \delta > 0,$	$\frac{\lambda\delta^{\lambda}}{(x+\delta)^{\lambda+1}}$
weibull	$\varepsilon > 0, \nu > 0,$	$\frac{\varepsilon(\frac{x}{\nu})^{\varepsilon} e^{-(\frac{x}{\nu})^{\varepsilon}}}{x}$

2.2 Model Selection Criteria

We estimate 240 composite distributions and present results on only the top ten (10) distributions based on the three goodness of fit criteria below:

- Akaike Information Criterion (AIC)
- Bayesian Information Criterion (BIC)
- Negative Log-Likelihood (NLL)

The AIC is defined mathematically as:

$$AIC = -2\ell(\theta) + 2(G + K - 1), \tag{5}$$

where G defines the number of components and K is the length of the vector θ. The AIC takes into account the log likelihood function $\ell(\theta)$.

The BIC is also defined as:

$$BIC = -2\ln(\hat{l}) + k\ln(n) \tag{6}$$

Where $\hat{l}$ is the model maximum value of the likelihood function, k measures the number of parameters in the model and n represent the number of observation in the data set.

The NLL is defined by letting $\ell(\theta)$ represent the log-likelihood function for a particular model.

$$NLL = -\ell(\theta). \tag{7}$$

2.3 Risk Measures

Given the security level p, tractable risk measures for composite models for a given random loss X can be obtained. Value at risk (VaR) for this random loss X according to [17] and [18] is defined as:

$$VaR_p(X) = \begin{cases} G_1^{-1}(p(1+\beta)G_1(\alpha)), & if\ 0 < p \le \frac{1}{1+\beta}, \\ G_2^{-1}(G_2(\alpha) + (p(1+\beta) - 1(1 - G_2(\alpha))/\beta), & if\ \frac{1}{1+\beta} < p < 1. \end{cases} \tag{8}$$

while [18] also defined the theoretical estimates for the TVaR of X as:

$$TVaR_p(Y) = \begin{cases} \frac{1}{1-p}\left[\frac{\int_{\pi_p}^{\alpha} y g_1(y)dy}{G_1(\alpha)} + \frac{\int_{\alpha}^{\infty} y g_2(y)dy}{1-G_2(\alpha)}\right], & if\ 0 < p \le \frac{1}{1+\beta}, \\ \frac{1}{1-p}\frac{1}{1-G_2(\alpha)}\left[\int_{\pi_p}^{\infty} y g_2(y)dy\right], & if\ \frac{1}{1+\beta} < p < 1 \end{cases} \tag{9}$$

3 Results and Discussions

3.1 Preliminary analysis

The data used for the study consist of third party auto-mobile insurance losses from an insurance company in Ghana. It consist of 11,879 observations spanning two years. The descriptive statistics of the data is presented in Table 2 below:

Table 2: Descriptive statistics of the third party loss data

Statistic	value(in hundred Ghana Cedis)
Minimum	0.1
Mean	19.32651862
Maximum	435
Quantiles	(6.12,12.68,26.01)
Standard deviation	23.30407555
Skewness	5.241443498
Kurtosis	55.84664572

Figure 1 below shows a boxplot of the loss data used for the study. Clearly, the data is right skewed suggesting that the data comprises of relatively larger losses. This gives an idea of some plausible candidate distributions that can fit the data.

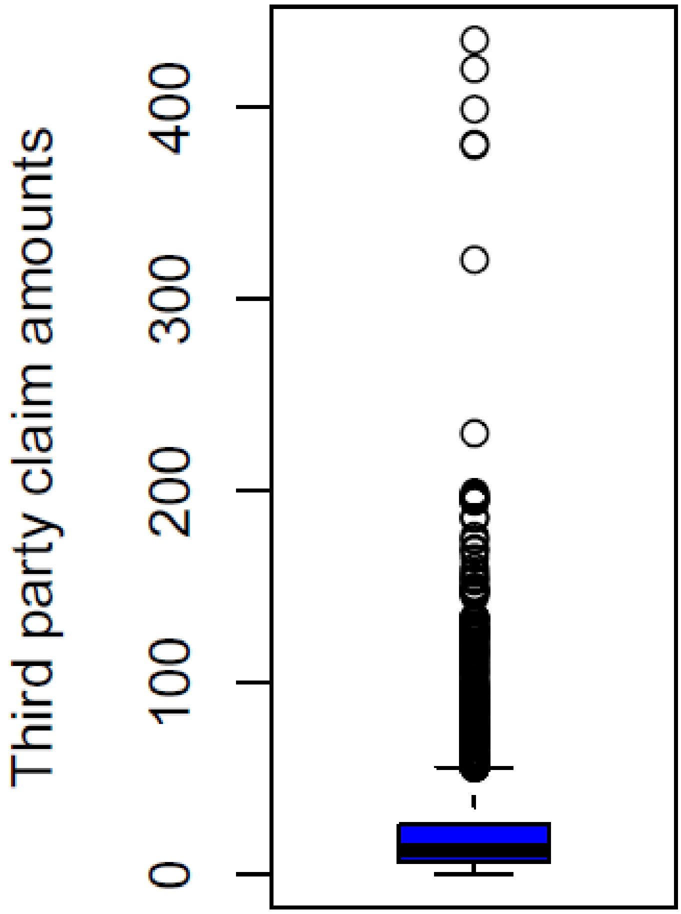

Fig. 1: Boxplot of auto-mobile loss

3.2 Fitting Composite distributions to the auto-mobile claims data

The 16 loss distributions in R-software package '*actuar*' outlined in Table 1 was used in fitting the 240 composite distributions. This was obtained by taking two distributions at a time making $^{16}C_2 \times 2 = 240$ in total. The model parameters are estimated using the the R-software package *'composite'*. Next we fit the auto-mobile claims data to the 240 distributions and present result of the top ten(10) based on the model selection criterion discussed in section 2.2.

This is presented in Table 3 below. It could be seen that distributions such us lognormal, burr and inverse burr fitted well to the body of data. This results is consistent with the work of [8] where lognormal was described as a good candidate for modelling smaller/moderate losses. Also, distributions such as Pareto, paralogistic and Weibull seem ideal for modelling large losses based on the results obtained. This result also side with the work of [8] where Pareto was described to have longer and thicker upper tails and therefore good for modelling larger losses. Also, using paralogistic and inverse paralogistic as the tail distribution is consistent with the work of [19] who proposed these distributions as good candidates for modelling large losses.

Table 3: Summary of the model selection criterion of the top 10 composite models and their parameters for the third party insurance loss data

Distribution	Model Sel. Criterion	Parameter (head)	Parameter(tail)
lognormal-Pareto	AIC=695.4123 BIC=695.6962 NLL=347.6662	μ=3.785412 σ=0.288956	λ =2.66793794 δ=0.04428109
Pareto-Paralogistic	AIC=701.4178 BIC=701.7016 NLL=350.6689	λ=2.496217 δ=0.062531	ω=2.600253 κ=0.482254
Burr-Inverse paralogistic	AIC=701.8400 BIC=702.1948 NLL=350.8700	α=0.5328918 β=1.32230835 γ=0.08777131	ω=2.58731399 κ=0.06027439
Burr-Inverse weibull	AIC=701.8400 BIC=702.1948 NLL=350.8700	α=0.5328918 β=1.32230835 γ=0.08777131	ε=2.58731399 ν=0.06027439
Inverse burr-Inverse Weibull	AIC=701.8413 BIC=702.1961 NLL=350.8707	α=1.1715960 β=1.1383744 γ=0.0627656	ε=2.5113566 ν=0.0406657
Inverse paralogistic-burr	AIC=701.8433 BIC=702.1933 NLL=350.8767	ω=1.14938086 κ=0.06123319	α=0.52644470 β=4.40846999 γ=0.04595472
Loglogistic-loglogistic	AIC=701.8855 BIC=702.1693 NLL=350.9027	ι=1.27827853 τ=0.05834516	ι=2.62845713 τ=0.04387424
Weibull-inverse burr	AIC=701.9740 BIC=702.3287 NLL=350.9370	ε=1.275757 ν=12.510394	α=0.11010811 β=2.87621009 γ=0.02695229
Inverse burr-Inverse burr	AIC=701.9740 BIC=702.3287 NLL=350.9370	α=1.275757 β=12.510394	α=0.11010811 β=2.87621009 γ=0.02695229
Inverse Weibull-burr	AIC=701.9841 BIC=702.3388 NLL=350.9420	ε=1032.058881 ν=9.999596	α=3.24130945 β=1.23005170 γ=0.02437112

The table shows that lognormal-pareto was the best composite distribution with an AIC of 695.4123. A critical look of the the Q-Q plots of the nine composite distributions in Figures 2, 3 and 4 show that most of them fitted quite well. However, Figure 5 which is the composite lognormal Pareto shows a very good fit as well. Clearly, the composite lognormal-Pareto fitted the data well than the other nine composite distributions. Further assessment of the top ten composite distributions clearly is needed.

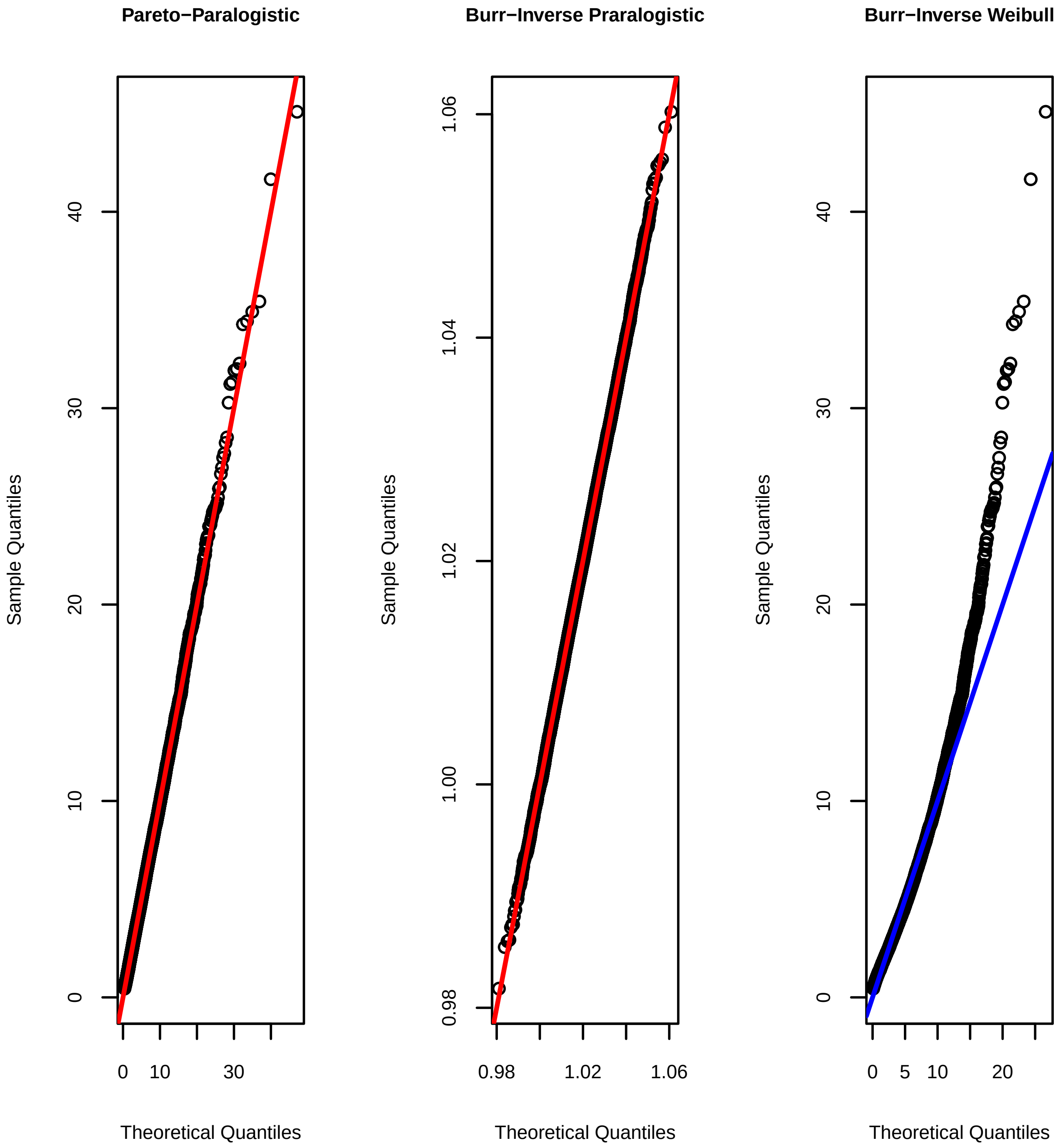

Fig. 2: Q - Q plots of composite Pareto-Paralogistic, Burr-Inverse Paralogistic and Burr-Inverse Weibull distributions respectively

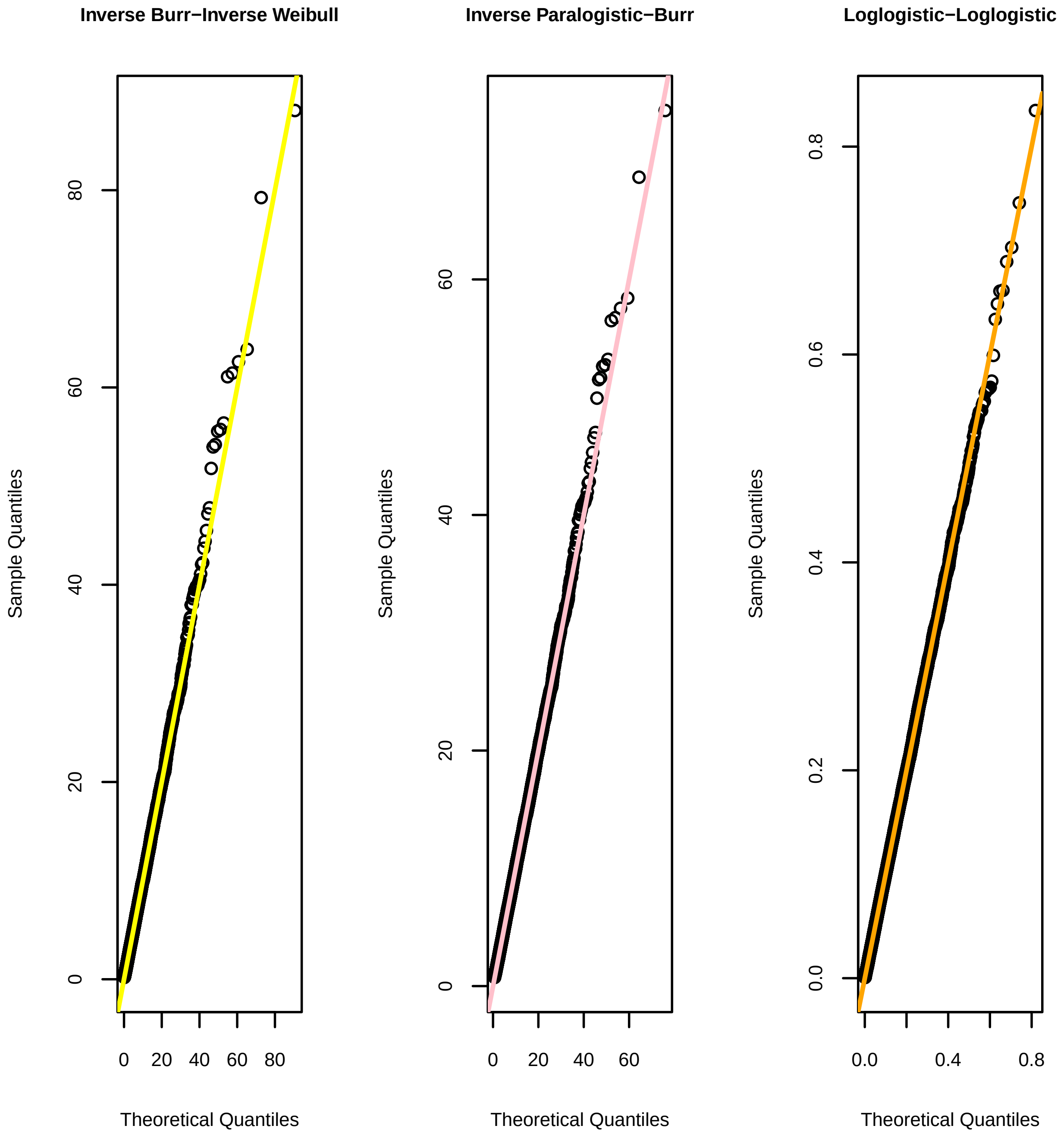

Fig. 3: Q - Q plots of composite Inverse Burr - Inverse Weibull, Inverse Paralogistic - Burr and Loglogistic - Loglogistic distributions respectively

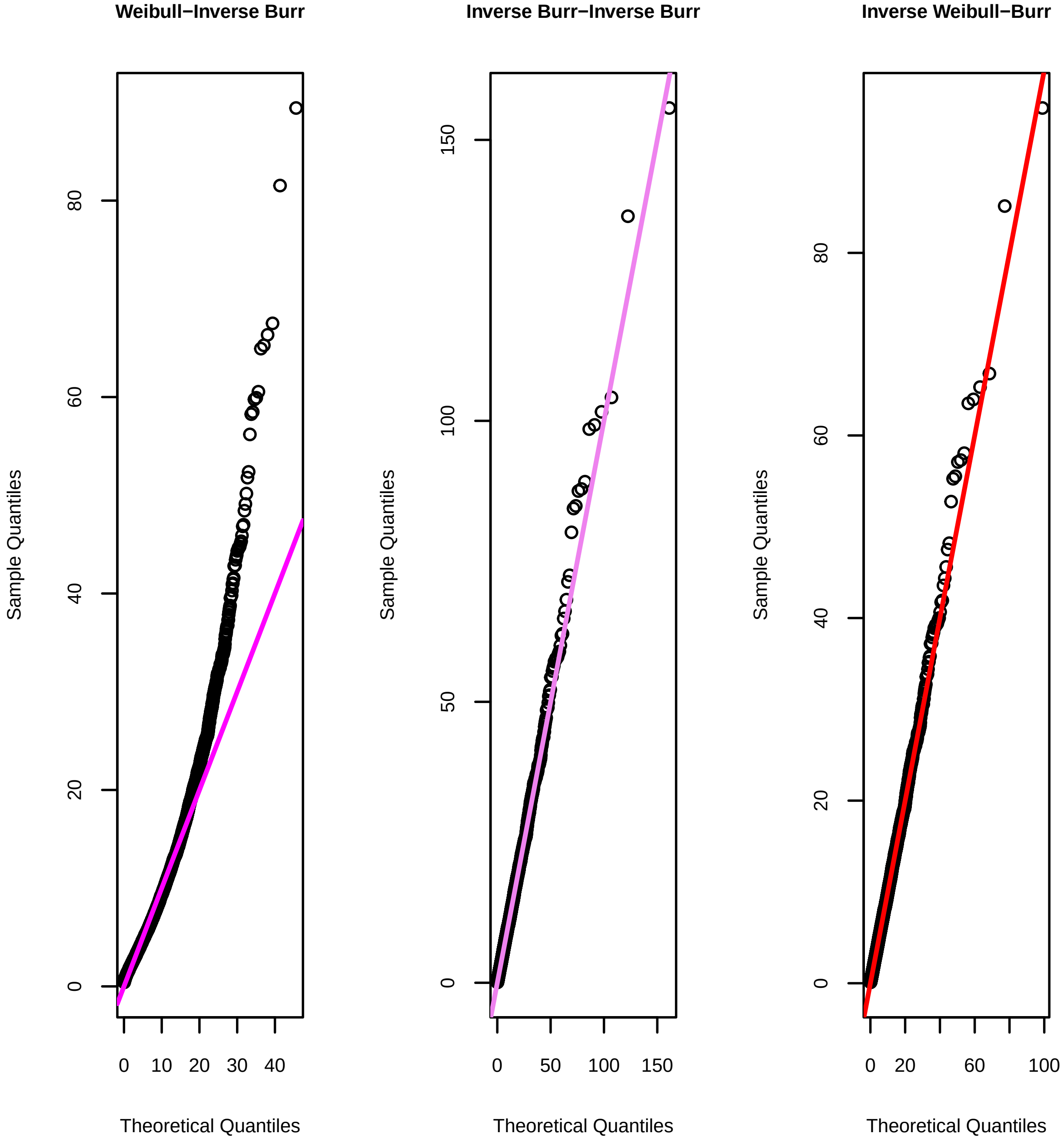

Fig. 4: Q-Q plots of composite Weibull - Inverse Burr, Inverse Burr - Inverse Burr and Inverse Weibull - Burr distributions respectively

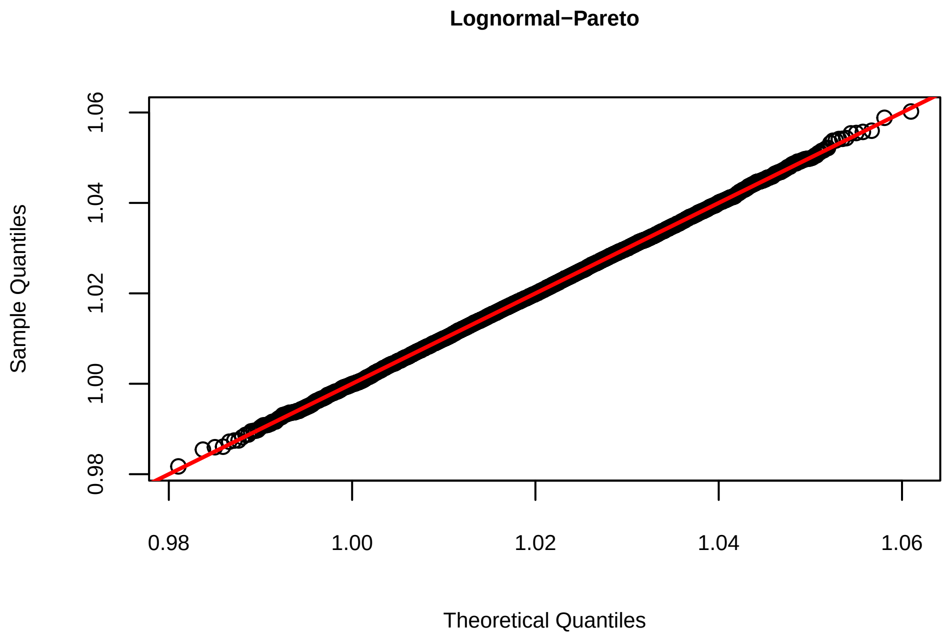

Fig. 5: Q-Q plot of Composite lognormal-Pareto distribution

Furthermore, the CDF plot of the lognormal Pareto in Figure 6 below has confirmed lognormal-Pareto as a good fit to the claims data. Below, further tests are performed to check the goodness of fit and to further cement the superiority of the preferred distribution that was used in the risk estimation.

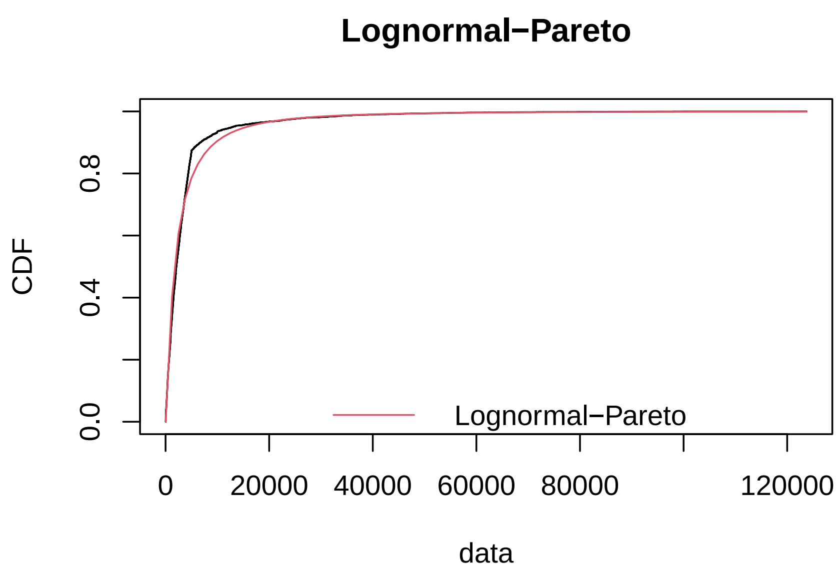

Fig. 6: Cumulative density plot for lognormal-Pareto distribution

Next in Table 4 below, we perform Kolmogorov- Smirnov (KS) test, and Cramer-von Mises (CVM) test to confirm whether the loss data follows the theoretical distributions specified. According to the K-S and CVM test, only the lognormal-Pareto had its p-value ($0.2040 > 0.01$) making it the most suitable distribution at 1% significance level. However at 5% significance level, the Pareto-Paralogistic distribution also provided a good fit to the data. The Vuong test was also conducted to compare the lognormal-Pareto distribution to the other nine composite distributions to confirm its superiority. The p-values affirm that the loss data is better fiited by the lognormal-Pareto comparatively. Clearly for all the tests conducted and from the plots, the lognormal-Pareto emerged as the best distribution and as such was used for the risk estimation. Next we describe the mixing weights and the threshold values. The threshold values where chosen in a way that continuity and differentiability conditions are satisfied.

Table 4: Goodness of Fit Test Statistics

Distribution	K-S Statistic	CVM Statistic	Vuong Test
Lognormal-Pareto	0.2040	0.3421	-
Pareto-Paralogistic	0.0520	0.0531	0.2341
Burr-Inverse Paralogistic	0.0001	0.0021	0.0041
Burr-Inverse Weibull	0.0002	0.0024	0.0020
Inverse Burr-Inverse Weibull	0.0054	0.0022	0.0201
Inverse Paralogistic-Burr	0.0032	0.0006	0.3421
Loglogistic-Loglogistic	0.0022	0.0018	0.5412
Weibull-Inverse Burr	0.0017	0.0067	0.1320
Inverse Burr-Inverse Burr	0.0034	0.0089	0.7890
Inverse Weibull-Burr	0.0022	0.0091	0.0020

Next, we present the top 10 models and their corresponding threshold and mixing weights for the composite distribution. This is presented in table 5 below. The mixing weights which is given by $\frac{1}{1+\beta}, \frac{\beta}{1+\beta}$ with $\beta > 0$ ensures that appropriate segment of the data set is fitted to the head and tail distribution respectively in the composite model to ensure good fit (that is by maximising the log-likelihood function of the joint distribution). From the table, all the top 10 composite distributions had greater percentage of the data points assigned to their body distributions with the remaining assigned to the tail distribution. Lognormal-Pareto, the best fitting composite distribution to our loss data revealed that about majority of the losses should be modelled with lognormal whilst the remaining is modelled by Pareto. The possible explanation is that the data at hand comprises of quite a number of smaller losses than larger losses. The threshold $\gamma > 0$ for the composite lognormal-pareto of 23.19613 is the point at which the data is divided for fitting the lognormal and pareto distribution since the composite model is a piecewise distribution. That is, losses up to 23.19613 was fitted with lognormal whilst losses above 23.19613 was fitted with Pareto. More generally, given that X_i are the losses, then $X_i < \gamma$ are fitted by the head distribution and $X_i \geq \gamma$ are fitted by the tail distribution. Surprisingly, Pareto-Paralogistic estimated that losses up to 64.37197 should be fitted with Pareto. This threshold is quite large as compared to that of lognormal in the composite lognormal-Pareto. This could possibly be explained that Pareto has been described in some literature to be capable of fitting relatively larger losses (see;[8]). Therefore using Pareto as head distribution will capture relatively larger loss as compared to lognormal. The 10 composite distributions estimated that on average, losses up to 23.5271109 should be fitted with the head distribution whilst losses above this threshold should be fitted with the tail.

Modelling claims data using lognormal-pareto distribution has many advantages and it truly reflects the distorted behaviour of insurance claims as there are some large claims that occurs at very small frequencies. The lognormal part fits the body of claims data which are moderate in size and highly skewed. This is consistent with several papers (See for example [6]; [20]; [21]). The pareto distribution on the other hand is ideal for extreme claims and used when there are large losses where heavy tailed behaviour is observed. This is also consistent with several papers (See for example, [22]; [23]; [24]). Hence our best composite distribution gives a good balance between small and large claims making it appropriate for estimating the underlying risk associated with the insurances losses. This gives furtherance to the Insurance Regulatory frameworks like solvency II and IFRS 17 which emphasizes correct modelling of tail risk. Also, since lognormal distribution is light tailed it usually underestimate risk while the Pareto distribution may overestimate solvency needs and these defects are cured by composite distributions as it gives a good balance between the two distributions. Cooray and Amanda (2005) also observed lognormal-Pareto distribution although the paper did not estimate any risk measure like in our paper and there was a vast difference in how the threshold was chosen. In this paper the thresholds are chosen in a way to ensure continuity and differentiability of the composite distributions. Some of the other 9 composite distributions although had a good fit, cannot be compared to the lognormal-distribution. These

Table 5: Summary of the top 10 composite models for the third party insurance claims and their corresponding threshold and mixing weight (in GH¢100)

Distribution	Threshold (γ)	Mixing weight (β)
lognormal-Pareto	23.19613	0.3983603
Pareto-Paralogistic	34.37197	0.3893017
Burr-Inverse paralogistic	24.84300	0.3393798
Burr-Inverse weibull	24.84300	0.3393798
Inverse burr-Inverse weibull	25.53389	0.3195328
Inverse paralogis-burr	26.11041	0.3035585
Log logistic-Log logistic	24.85599	0.3349225
Weibul-inverse burr	10.70834	0.3009890
Inverse burr-Inverse burr	10.70834	0.3009890
Inverse weibull-burr	10.00039	0.4247165

distributions have some advantages but their K-S statistics showed that they were not a good fit to our data hence their relevance to the actuarial modelling of claims are not discussed.
Table 6 below presents the risk measures of the top 10 composite distribution that best fit the third party auto-mobile insurance claims data. From the table, lognormal-Pareto shows that at the 99th percentile of claims, the insurance company can make an average loss of GH¢14,038.30 within any given day. This loss is substantially large especially for a typical Ghanaian third party. This loss is even much more higher for the same Lognormal-Pareto when looking at the tail value risk for the same 99th percentile of claims. This result is a bit critical for the Ghanaian insurance industry. Furthermore, Pareto-Paralogistic which was the second best fit distribution for the loss data shows that at the same 99th percentile of claims, the insurer can incur a loss of GH¢11,430.21 per any given day. This loss is a bit closer to Lognormal-Pareto which best fit the loss data. Again, inverse weibull-burr which was the least among the top 10 distributions that best fit the loss data revealed that at 99th percentile of claim, the insurer can record a loss of GH¢10, 409.65 for any given day. On average, the top 10 composite distribution that best fit the loss data revealed an average loss of GH¢9016.106 at 99th percentile of claims within any given day. This loss on several policies within a single day can be catastrophic for the ordinary insurance company in Ghana to settle.

Table 6: Risk Measures of the top 10 composite model of the third party insurance claim data (in GH¢100)

Distribution	$VaR_{0.95}$	$VaR_{0.99}$	$TVaR_{0.95}$	$TVaR_{0.99}$
Lognormal-Pareto	50.80211	140.38300	101.941.40	261.09420
Pareto-Paralogistic	87.87502	114.3021	101.0325	134.9203
Burr-Inverse paralogistic	46.5173	86.64975	45.75662	52.40518
Burr-Inverse weibull	46.5173	86.64974	45.75662	52.40517
Inverse burr-Inverse weibull	47.85505	27.07551	71.96752	148.5736
Inverse paralogistic-burr	54.89154	110.2185	93.91548	179.6023
Loglogistic-loglogistic	45.9139	84.69704	72.532	128.8246
Weibull-inverse burr	41.69403	79.65294	67.1178	117.4582
Inverse burr-Inverse burr	33.70246	67.88554	56.50806	100.7699
Inverse weibull-burr	57.71673	104.0965	87.6859	143.9634

Furthermore, the work of [6] where insurance losses from Ghana were modelled with several single loss distributions (like Burr, lognormal, Gamma, exponential Weibull etc.) projected lognormal as ideal for modelling the Ghanaian insurance dataset. Subsequently, they estimated risk measures at 95% and 99% security levels using lognormal as underlying distribution. However, given that loss data is righly skewed, the tail risk is likely to be underestimated and hence can affect amount reserved for indemnifying claims. Composite distributions which is adopted in this paper have shown good promise in modelling loss data properly than single than single distributions and consequently the risk estimates. This study estimated risk measures for loss data from Ghana using composite models as underlying probabilistic distributions. Similarly, [25] used single distributions with some parameter perturbations to estimate risk . In their work, they modelled the well - known South African Taxi claims data and Danish fire insurance loss data. They concluded that the transformed beta family of distributions provided a better fit to both datasets, they subsequently estimated risk measures using these single standard distributions. Our work is different form their work in the sense that we estimated risk measures for the losses using composite distributions. [20] argued that composite distributions provide good risk estimates. Hence the risk measure estimated in this this research is deeply rooted on the fact that composite distributions provide better risk estimates in comparisons to those mentioned above.

3.3 Implication of study for risk assessment and mitigation

This study provides valuable insights for planning reserves and investment components of premiums collected by insurance firms. It particularly provides estimates to inform policy on the purchase of reinsurance due to the dire consequences of observing higher claims by insurance firms. The core objective of this study which is to provide ways of quantifying risk ensures that insurers can offer appropriate coverage, set right premiums to ensure financial stability.
The main motive of purchasing reinsurance is to protect the insurer from large losses that can lead to insolvency. Also, one main challenge that the insurance companies faces is determining the right amount of reserve that can pay future losses. The estimates of the risk measures have provided a clue of how much the insurer should reserve in order to remain solvent. TVaR for Lognormal-Pareto gave the largest possible loss for the data at hand. We recommend that reinsurance is purchased by the insurance companies in Ghana. Due to the vast differences between estimates for the two risk measure, we recommend that several risk measures are computed and then an average of the estimates is then used for decision making. Again, the risk measures provided rough estimate of the nature of losses the company should expect in order to cover unforeseen losses. This therefore inform the insurance company how much they should reserve in order to remain solvent. That is risk estimates can be used as reserve benchmark whilst ensuring that enough is allocated for day to day activities of the industry and for investment to attract investment income.
The estimates of the two risk measures are substantially large and it is therefore essential that the insurance industry embark on some risk mitigation interventions that can reduce the frequency and severity of the losses.

3.4 Conclusion

Modelling insurance losses is an essential part of the work of an actuary as it aids in reserving and also create an awareness of future unexpected claims. Modelling insurance losses with single distributions cannot capture the tail behaviour of relatively large losses. In view of this, composite distributions have been used in several studies to model insurance losses (see; [7], [8], [9]). This study modelled third party motor insurance losses from an insurance company in Ghana using composite distribution. 240 composite distributions derived from 16 single loss distributions known in literature were fitted to the loss data. Using model selection criteria, top 10 composite distributions that best fit the third party motor insurance data were lognormal-Pareto, Pareto-Paralogistic, Burr-inverse Paralogistic, Burr-inverse Weibull, inverse burr-inverse Weibull, inverse paralogistic-Burr, loglogitic-loglogistic, Weibull-inverse Burr, inverse Burr-inverse Burr, and inverse weibull-Burr. This result is consistent with literature (see; [8], [19]). The mixing weights and the threshold for the top 10 composite distributions were estimated. Two risk measures; VaR and TVaR at 95th and 99th percentile of claims for the top 10 composite distributions were computed. Lognormal-Pareto which turned out to be the best estimated that the insurance company on any single day can pay a loss of Gh¢14,038.30 on a single policy from its risk estimate. This loss is relatively higher for a typical insurance company in Ghana. The top 10 composite distribution considered revealed an average loss of Gh¢9016.106 on single policy within any given day at a security level of 99%. This loss might look a bit smaller for the insurance industry but this could be catastrophic for the industry especially when looking at several policies together within any single day. The implication of this relatively large loss is that the industry should resort to reinsurance so that the losses could be shared to mitigate losses which could impact profitability. Also the insurance companies could engage policyholders on some risk management interventions so as to minimise the severity of the losses reported.

Acknowledgement

The authors express profound gratitude to the Editor-in-Chief and the entire editorial team for publishing this paper at no cost. We are also grateful to the anonymous reviewers for their constructive comments which have improved the paper.

Funding

The authors did not receive support from any organization for the submitted work.

Conflicts of Interest

All authors declare no conflict of interest.

References

[1] Cyprian, O.O., Shalyne G. N., & Joan M. W. M.Omari, C. O., Nyambura, S. G., & Mwangi, J. M. W. Modelling the frequency and severity of auto insurance claims using statistical distributions, 2018.
[2] Cummins, J.D. Statistical and financial models of insurance pricing and the insurance firm, *The Journal of Risk and Insurance*, 58(2):261?302, 1991.
[3] Lukasz, D., Mathias, L., & Mario, V W. Gamma mixture density networks and their application to modelling insurance claim amounts, *Insurance: Mathematics and Economics*, 101:240-261, 2021.
[4] Bjorn, W., Models, modelling and modellers: an application to risk analysis, *European Journal of Operational Research*, 75(3):477-487, 1994.
[5] Lukasz, D., Mathias, L, & Mario, V. W. *Gamma mixture density networks and their application to modelling insurance claim amounts*, Insurance: Mathematics and Economics, 101:240 - 261, 2021.
[6] Kwofie, C., Kumi, W., Otoo, H., Appiah, S.T. & Ocran, E., Risk measures associated with insurance losses in Ghana. *Journal of Operational Risk*, 2025.
[7] Marambakuyana, W. A., & Shongwe, S. C. . Quantifying risk of insurance claims data using various loss distributions. *Journal of Statistical Applications & Probability*, 13(3), 1031-1044,2024.
[8] Cooray, K., & Ananda, M. M. (2005). Modeling actuarial data with a composite lognormal-Pareto model. *Scandinavian Actuarial Journal*, 321-334,2005
[9] Pigeon, M., & Denuit, M. . Composite Lognormal?Pareto model with random threshold. *Scandinavian Actuarial Journal*, 2011(3), 177-192, 2011.
[10] Ochola, E. O. A stochastic analysis of claim reserving in General Insurance using bootstrapping technique (Doctoral dissertation, University of Nairobi).
[11] Myers, S. C., & Read Jr, J. A. (2001). Capital allocation for insurance companies. Journal of risk and insurance, 545-580, 2018.
[12] Bremermann, L., Rosa, M., Galvis, P., Nakasone, C., Carvalho, L., & Santos, F. Using VaR and CVaR techniques to calculate the long-term operational reserve. In 2016 International Conference on Probabilistic Methods Applied to Power Systems (PMAPS) (pp. 1-7). IEEE, 2016, October.
[13] Hitchcox, A. N., Hinder, I. A., Kaufman, A. M., Maynard, T. J., Smith, A. D., & White, M. G. . Assessment of target capital for general insurance firms. British Actuarial Journal, 13(1), 81-168, 2007.
[14] Burchi, A. Capital requirements for market risks: Value - at - risk models and stressed?VaR after the financial crisis. Journal of Financial Regulation and Compliance, 21(3), 284-304,2013.
[15] Xiong, Q., Peng, Z., & Nadarajah, S. . Optimal reinsurance under the linear combination of risk measures in the presence of reinsurance loss limit. Risks, 11(7), 125, 2023.
[16] Mubiri, G. W. Analysis of capital calculation models for a life insurer (Doctoral dissertation, University of Nairobi), 2015.
[17] Bakar, S. A., Hamzah, N. A., Maghsoudi, M., & Nadarajah, S. Modeling loss data using composite models. Insurance: Mathematics and Economics, 61, 146-154, 2015.
[18] Grun, B., & Miljkovic, T. Extending composite loss models using a general framework of advanced computational tools. Scandinavian Actuarial Journal, 2019(8), 642-660, 2019.
[19] Maghsoudi, M. Modeling Loss Data with Composite Models. University of Malaya (Malaysia), 2016.
[20] Nadarajah, S., & Kwofie, C. Heavy tailed modeling of automobile claim data from Ghana. *Journal of Computational and Applied Mathematics*, 405, 113947, 2022..
[21] Zuanetti, D. A., Diniz, C. A., & Leite, J. G. A lognormal model for insurance claims data. *REVSTAT-Statistical Journal*, 4(2), 131-142, 2006.
[22] Henry, J. B., & Hsieh, P. Extreme value analysis for partitioned insurance losses. Variance, 3 (2), 214-238, 2009.
[23] Dzupire, N. Modelling of Large Fire Insurance Claims: An Extreme Value Theory Approach. London Journal of Research In Science: Natural and Formal, 24(8), 47-67, 2024.
[24] Gomez-Deniz, E., & Calderin-Ojeda, E. Modelling insurance data with the Pareto ArcTan distribution. ASTIN Bulletin: The Journal of the IAA, 45(3), 639-660, 2015.
[25] C. Marambakuyana, W. A., & Shongwe, S. C. Quantifying risk of insurance claims data using various loss distributions. Journal of Statistical Applications & Probability, 13(3), 1031-1044, 2024.

Williams Kumi is a lecturer at the Department of Mathematics and Statistics, University of Energy and Natural Resources (UENR), Sunyani-Ghana. He is currently pursuing a PhD in Mathematics (Concentration in Financial Engineering) at the University of Mines and Technology (UMaT), Tarkwa-Ghana. His academic and research interests span financial mathematics, insurance mathematics, and probability theory. Through his work, he aims to contribute to the advancement of risk assessment and quantitative methods in finance and insurance.

Henry Otoo is an Associate Professor of applied Mathematics at the Department of Mathematical Sciences, University of Mines and Technology (UMaT), Tarkwa-Ghana. His research interests encompass Topological Dynamics, Dynamical Systems, Mathematical Modelling of Infectious Diseases, Operations Research, and Time Series Analysis and Forecasting. With extensive experience in both teaching and research, Professor Otoo's work integrates theoretical and applied mathematics to address complex real-world problems, particularly in public health and decision sciences.

Charles Kwofie is a senior Lecturer in the Department of Mathematics and statistics at the University of Energy and Natural Resources, Sunyani-Ghana. He Lectures Actuarial Science and statistics courses. He is actively involved in applying probability theory in insurance and in quantifying risk. Dr. Kwofie is also interested in Large deviations theory, statistical modeling, discrete choice experiments and in the development of new probabilistic distributions.

Sampson Takyi Appiah is an Associate Professor of Applied Mathematics at the Department of Mathematics and Statistics, University of Energy and Natural Resources (UENR), Sunyani - Ghana. His research interests include modelling with differential equations, statistics, and optimization techniques. Professor Appiah's work focuses on applying mathematical models and analytical methods to solve real-world problems across science, engineering, and industry, contributing significantly to the advancement of applied mathematical research and education in Ghana.