Modeling Claim Sizes In Personal Line Non-Life Insurance

Ismail Adedeji Adeleke, University of Lagos, Nigeria
Ade Ibiwoye, University of Lagos, Nigeria

ABSTRACT

This paper uses claims data from the most prominent lines of non-life insurance business in Nigeria to determine appropriate models for claim amounts by fitting theoretical distributions to the various data. The risk premiums for each class of business are also estimated. The result of the study demonstrates that some lines of business are indeed better modeled with different distributions than had earlier been conjectured.

Keywords: non-life insurance; claim size; exploratory data analysis; risk premium; estimation; inflation

INTRODUCTION

An overview of the distributional models and diagnostics that can be used to represent large claims in non-life insurance was given by Beirlant and Teugels (1992). Antonio and Beirlant (2006) discussed both a priori and a posteriori rating systems in non-life insurance with examples from likelihood based and Bayesian estimations. Their main approach was to distinguish between low profile and high profile risks by splitting an insurance portfolio into classes that consist of similar profile and then determining a fair tariff for each. Keiding et al (1998) modified Cox regression model commonly used in survival analysis to study the hazard of occurrence of claims in auto, property and household insurance in a Danish county. Yip and Kelvin (2005) considered the problem of zero-inflated claim frequency data in general insurance. The majority of the other earlier works made use of regression. We observe that these models could be allowed for developed economies where data collection poses little problems.

In Nigeria, as well as other developing economies, there is paucity of data. Further, there is little analysis of the existing data beyond the simple descriptive statistics that organizations like the Nigerian Insurance Association annually conduct. This probably explains why insurance companies in such economies are not operating at their full potential. Claims, especially large ones, had been the Achilles' heels of the insurance industry as past incidents of repudiation of claims had created an image problem for the industry. An analysis of claim size is essential in determining actually fair premium. When this is not the case, an insurance company risks insolvency deriving mainly from adverse selection and moral hazard.

Incidentally, claim occurrence is not all that predictable. For an insurance company, neither the time, the size, nor the frequency of the claim is known with certainty. The degree of risk brought into the pool varies in terms of possibility of losses or hazard and the extent of potential loss. The insurer, therefore, needs to ensure that the premium charged to individual members of the pool is equitable, compared with the contribution of others bearing in mind the likely frequency and severity of claims that may be made by that individual. In the absence of knowledge about the occurrence of claims, premiums are only determined by guesswork, such as when the rates for other developed economies are adopted for developing economies. Unfortunately in many cases, the adaptation is inappropriate as the micro and macro economic conditions of the developed and developing economies vary widely. To continue with the practice of inappropriate premium pricing would do little to mitigate the poor image problems from which the insurance industry had been struggling to extricate itself (Osoka, 1992).

Although Hamadu (2010) rightly observed that each form of insurance has its background of occurrences, thereby creating contingencies that require specific protection, his investigation of the stochastic distribution of claim sizes of armed robbery cases in Nigeria was limited to only one line of business. From our experience, improper premium determination is not limited to the peril of armed robbery. The problem permeates other classes of insurance which poor management had constituted an impediment to the growth of the industry. The main lines include fire, motor, property, theft and armed robbery insurance. In this study, we determine appropriate models for claim amounts for different classes of non-life insurance in Nigeria by fitting theoretical distributions to the loss data in these lines of insurance business. We also explore the advantage of Pearson Chi-square statistic over Andersen and BIC criteria, to serve as our measure of goodness of fit. To the best of our knowledge, such a study has not been carried out for the Nigerian insurance market.

The study progresses as follows: Section 2 presents some theoretical distributions that can be fitted to the Nigerian claims data. Section 3 discusses the data. Section 4 analyzes the data and fits theoretical distributions to them. Section 5 estimates the risk premium for each line of business and examines the effect of inflation on the operation of the insurance company while Section 6 concludes.

CLAIM SIZE DISTRIBUTION

In the different classes of insurance business, it is not clear cut which distributions are suitable for modeling claim amounts as claims can take on large number of values (Promislow, 2006). In the following paragraphs, we examine some theoretical distributions.

Exponential Distribution

A natural place for starting the discourse on theoretical distributions should be the exponential distribution because it is one of the simplest and most basic distributions used in modeling. The exponential distributions are a class of continuous probability distributions that describe the times between events; that is, a process in which events occur continuously and independently at a constant average rate. Having observed a sample of *n* data points from an unknown exponential distribution, a frequent task is to use these samples to make predictions about future data from the same source. A common predictive distribution over future samples is formed by plugging a suitable estimate for the rate parameter λ into the exponential density function.

A random variable *X* is said to be exponentially distributed if it has a density function:

$f_X(x) = \lambda e^{-\lambda x}$ for $x > 0$, where λ is the parameter.

This distribution has a survival function given as:

$$\bar{F}_X(x) = e^{-\lambda x}$$

with mean given by $E(X) = 1/\lambda$ and variance $Var(X) = 1/\lambda^2$.

The moment generating function of *X* is given by:

$M_X(x) = \lambda/(\lambda - t)$, for any $t < \lambda$

The exponential distribution has a memory-less property, a property that is shared by no other continuous distribution. This unique property characterizes the family of exponential random variables. Since much of the theory of the generalized linear models is derived from this distribution, it is a very important distribution in modeling insurance claim counts (Boucher et al, 2008; Boland, 2007) as follows:

We define the sequence $\{X_i, i = 1,2, . . .\}$ as consecutive claim sizes. This sequence is assumed to be a renewal process generated by the claim size distribution defined as $B(x) = P(X \leq x)$ where *X* is a generic claim size. The exponential distribution is then given by $B(x) = 1 - \exp\left(-\frac{x}{\mu}\right)\ (\mu > 0)$ (see Beilant and Teugels

1992). Generally, the exponential distribution belonging to the super-exponential class which is characterized by $e(x) \equiv 1/\mu$ where $e(x) = \int_x^\infty \bar{F}(u)\,du/\bar{F}(x) = E(X - x|X > x), x > 0$.

As noted in McLaughlin (2001), the skewness and kurtosis for any exponential distribution are, respectively, 2 and 6. Comparing these values with the summaries in Table 1, we can infer that the claims data for the other classes of personal line insurance, besides fire, are more positively skewed and have a fatter tail than one would expect from an exponential distribution. The column for the fire claims summary in Figure 1 suggests that the exponential distribution might be a good fit for its data. The other lines of the business have fatter tails and the exponential distribution would not be a good fit.

Pareto Distribution

Another class of distribution that is commonly used to model income distribution in economics or claim size distribution in insurance is the Pareto distribution (Boland, 2007). A random variable *X* is said to have a Pareto distribution with positive parameters α and λ if it has density function given by $f_X(x) = \frac{\alpha\lambda^\alpha}{(\lambda+x)^{\alpha+1}}$, or equivalently, if its survival function is given by $\bar{F}_X(x) = \left(\frac{\lambda}{\lambda+x}\right)^\alpha$, for $x > 0$. Similar to the exponential family of random variables, the Pareto distribution has density and survival functions which are tractable. Pareto random variables have some preservative properties, for instance, if $X \sim Pareto(\alpha, \lambda)$ and $k > 0$, then $kX \sim Pareto(\alpha, k\lambda)$ since

$P(kX > x) = P\left(X > \frac{x}{k}\right) = \left(\frac{\lambda}{\lambda+x/k}\right)^\alpha = \left(\frac{k\lambda}{k\lambda+x}\right)^\alpha$. This property is useful in dealing with inflation in claims.

When $X \sim Pareto(\alpha, \lambda)$, one may readily determine the mean (when α > 1) and variance (when α > 2) by taking expectations. Thus;

$$E(X) = \int_0^\infty \left(\frac{\lambda}{\lambda+x}\right)^\alpha dx$$

$$= \frac{\lambda^\alpha}{(\alpha-1)(\lambda+x)^{\alpha-1}}\Big|_0^\infty$$

$= \frac{\lambda}{\alpha-1}$, and

$$E(X^2) = 2\int_0^\infty \left(\frac{\lambda}{\lambda+x}\right)^\alpha dx$$

$= \frac{2\lambda^2}{(\alpha-1)(\alpha-2)}$ and therefore

$$Var(X) = \frac{\alpha\lambda^2}{(\alpha-1)^2(\alpha-2)}$$

Using the method of moments to estimate the parameters α and λ of a Pareto distribution, we can solve the equations

$\frac{\lambda}{\alpha-1} = \bar{x}$ and $\frac{\alpha\lambda^2}{(\alpha-1)^2(\alpha-2)} = s^2$

to get

$\hat{\alpha} = \frac{2s^2}{s^2 - \bar{x}}$ and $\hat{\lambda} = (\hat{\alpha} - 1)\bar{x}$

Gamma Distribution

For risk analysis modeling, particularly, for claims size modeling, the Gamma distribution has been found to be extremely useful (Hogg, McKean and Craig, 2005). The Gamma function is defined for any $\alpha > 0$ by $\Gamma(\alpha) = \int_0^{+\infty} y^{\alpha-1} e^{-y}\, dy$.

It has the properties that

$\Gamma(n) = (n-1)\Gamma(n-1)$ and $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$.

X has a gamma distribution with parameters α and λ if *X* has a density function given by $f_X(x) = \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x} \quad for\ x > 0$ with mean and variance given respectively as $E(X) = \frac{\alpha}{\lambda}$ and $Var(X) = {}^{\alpha}/_{\lambda^2}$

Solving the above equations, the moment estimators of α and λ are obtained as:

$\hat{\alpha} = {}^{\bar{x}^2}/_{s^2}$ and $\hat{\lambda} = {}^{\bar{x}}/_{s^2}$

where $\bar{x}$ and s^2 are, respectively, the mean and variance of the claim sample.

Weibull Distribution

A random variable, *X*, has a Weibull distribution with parameters $c, \gamma > 0$ if it has density function $f_X(x) = c\gamma x^{\gamma-1} e^{-cx^\gamma}, x > 0$. The parameters $c\ and\ \gamma$ are often called the scale and shape parameters for the Weibull random variable, respectively. If the shape parameter $\gamma < 1$, then the tail of *X* is fatter than that of any exponential distribution, but not as heavy as that of a Pareto. When $\gamma = 1$, then *X* is exponential with parameter c. The Weibull distribution is one of the extreme value distributions in the sense that it is one of the possible limiting distributions of the minimum of independent random variables (Barlow and Proschan, 1965). A particular property of the Weibull distribution is the functional form of its survival function, which has led to its widespread use in modeling lifetimes. Another attractive aspect is that the failure or hazard rate function of the Weibull distribution is of polynomial form since

$r_X(x) = \frac{f_X(x)}{F_X(x)} = \frac{c\gamma x^{\gamma-1}}{e^{-cx^\gamma}} = c\gamma x^{\gamma-1}$

and the k^{th} moment of *X* is obtained, thus:

$E(X^k) = \frac{1}{c^{k/\gamma}} \Gamma\left(1 + \frac{k}{\gamma}\right)$

Apart from its wide application in life insurance, the Weibull distribution had been found to be particularly useful in non-life insurance for modeling the size of reinsurance claims (Beirlant and Teugels, 1992; Boland 2007).

Lognormal Distribution

A random variable, *X*, is said to have a lognormal distribution with parameters μ and σ^2 if

$Y = logX \sim N(\mu, \sigma^2)$.

Letting $g(Y) = e^Y = X$, the density function of *X* may be determined from that of *Y* as follows;

$$f_X(x) = f_Y(logx)|[g^{-1}(x)]'| = \left[\frac{1}{\sqrt{2\pi}\sigma} e^{-(log\,x-\mu)^2/2\sigma^2}\right]\frac{1}{x}\ for\ x > 0$$

Using the expression for the moment, generating function of a normal random variable, the mean and variance of *X* are determined as follows:

$E(X) = e^{\mu+\sigma^2/2}$ and $Var(X) = e^{2\mu+\sigma^2}\left[e^{\sigma^2} - 1\right]$

The lognormal distribution is skewed to the right and is often useful in modeling claim size (Boland, 2007, Hogg and Craig, 1978). In trying to fit a lognormal distribution to model a loss (or claim) distribution, one generally uses the method of moments to estimate the parameters μ and σ^2. The Lognormal distribution like its Weibull and Pareto counterparts have indeed been shown to belong to the class of sub-exponential distribution and satisfying the condition that

$\lim_{x\to\infty} \frac{1-F^{*2}(x)}{1-F(x)} = 2.$

Further, for any non-negative integer *n* as $n \to \infty$, $\frac{1-F^{*n}(x)}{1-F(x)} \to n$

where

$F^{*n}(x)$ is the n^{th} convolution of $F(x)$ with itself (Beirlant and Teugels, 1992).

DATA COLLECTION

The Insurance Decree number 2 of 1997 (the Act by which the military regime conduct the affairs of state) sub-divided the insurance business into two main classes - life and non-life. Some companies operate composite lines, but the majority specializes in only one class. We collected data on the following insurance categories: Fire Motor, Property, Theft, and Armed Robbery, because these are the classes that fit more into the personal lines. Others, such as Workmen's Compensation and Employer's Liability insurance, Oil and Gas insurance, and Contractors all Risk insurance, fit more into the corporate class. We presume that most corporate organizations know and appreciate the need for insurance, but individuals really do not and they create the greater part of the bad image for the insurance industry.

In addition to the five classes included in the study, we also have a sixth class which is composed of the five classes combined. This is to allow for the class commonly referred to in the industry as Commercial line. The major source of data is the Nigerian Insurer Association (NIA). The data contained a comparative report of premium, claim and management expenses of member companies under non-life business for a period of 20 years. The claims are those under policies relating to current years net of reinsurance recoveries and the data represent collated aggregate of claim size at regular intervals.

DATA ANALYSIS AND DISTRIBUTION FITTING

We found the Exploratory Data Analysis (EDA) techniques very useful in investigating the suitability of certain families of distributions for particular data in attempting to fit the various claims data (see Table 1).

Table 1: Descriptive Statistics for the Claims Data

	Fire	Motor	Property	Theft	Armed Robbery	Commercial
Mean	382096.024	415006.2445	605388.8257	320829.6683	508215.2453	409105.3857
Standard Error	131132.4101	105186.4152	240107.1034	107770.8091	139884.6119	56447.19793
Median	185000	141400	135000	98152.5	223500	128812.5
Standard Deviation	507873.6404	721121.731	1680749.724	580363.5684	541770.7723	913677.718
Sample Variance	2.57936E+11	5.20017E+11	2.82492E+12	3.36822E+11	2.93516E+11	8.34807E+11
Kurtosis	3.815963117	6.353829359	38.15293908	5.607076811	-1.264212655	83.39896508
Coefficient of variation	132.9178	173.76166	274.6565	180.89461	106.60262	223.33554
Skewness	1.987542595	2.608202381	5.911331308	2.595037759	0.723757549	7.606463987
Range	1776000	3033357	11461032.34	2181600.04	1419898.37	11473415.74
Minimum	24000	1350	12967.8	5399.96	1200	584.4
Maximum	1800000	3034707	11474000.14	2187000	1421098.37	11474000.14

These fitted distributions include exponential, Lognormal, Weibull and Gamma distributions. The Q-Q (quantile-quantile) plots and the plot of the fitted distributions give some support to the use of distributions for the data. Since these techniques for analyzing fit are exploratory, we needed to use one or more of the traditional classic methods to test the goodness of our fit. In our study, we made use of the chi-square goodness-of-fit, which has been found to be suitable for both discrete and continuous distributions to assess the fit. We chose the Chi-Squared goodness-of-fit test over the Kolmogorov-Smirnoff test because the latter is often not good at detecting tail discrepancies (Boland, 2007). Also we rejected the K-S test and its modification, the Anderson-Darling test, because they are non parametric.

The null and the alternative hypotheses for a Chi-Squared test are:

H_0: The data follow the specified distribution.

H_A: The data do not follow the specified distribution.

The hypothesis regarding the distributional form is rejected at the chosen significance level (α) if the test statistic is greater than the critical value.

Table 2: Observed and Expected Values for Fitting Classic Distribution to Property Data

Group	Claim Interval	observes frequency	Expected frequency (Exponential)	Expected frequency (Gamma)	Expected frequency (Weibull)	Expected frequency (Lognormal)
1	[0, 500000)	64	26.78	56.99	46.83	50.09
2	[500000, 1000000)	6	29.67	8.48	22.46	17.33
3	[1000000, 1500000)	2	13.13	3.92	6.53	5.16
4	[15000000, 2000000)	3	5.81	2.44	2.39	2.41
5	[20000000, 25000000)	1	2.57	1.69	0.98	1.37
6	[25000000, 30000000)	1	1.14	1.25	0.43	0.86
7	[30000000, ∞)	3	0.6	4.73	0.27	2.46
8	X^2 – statistic		91.96	6.62	50.43	13.59
	p – value		0.00	0.16	0.00	0.01

Table 2 shows that the Gamma distribution provides the best fit for the property data since it has the least Chi-square value of 6.62 and correspondingly highest p-value of 0.16 to confirm its suitability.

Table 3: Observed and Expected Values for Fitting Classic Distribution to Theft Data

Group	Claim Interval	observed frequency	Expected frequency (Exponential)	Expected frequency (Gamma)	Expected frequency (Weibull)	Expected frequency (Lognormal)
1	[0, 500000)	39	15.23	26.86	28.92	31.72
2	[500000, 1000000)	2	18.98	11.59	13.23	11.32
3	[1000000, 1500000)	2	9.92	5.65	5.28	3.93
4	[15000000, 2000000)	3	5.19	3.40	2.74	2.05
5	[20000000, 25000000)	3	2.71	2.21	1.60	1.27
6	[25000000, 30000000)	1	1.42	1.50	0.99	0.86
7	[30000000, 35000000)	1	0.74	1.04	0.65	0.62
8	[35000000, ∞)	4	0.59	2.34	1.35	2.99
	X^2 – statistic		79.61	17.45	21.69	13.71
	p – value		0.0000	0.0037	0.0006	0.0176

From Table 3, Lognormal model is the best fit for the theft data since it has the least Chi-square value of 13.71 and correspondingly highest p-value of 0.0176 to confirm its suitability.

Table 4: Observed and Expected Values for Fitting Classic Distribution to Armed Robbery Data

Group	Claim Interval	observes frequency	Expected frequency (Exponential)	Expected frequency (Gamma)	Expected frequency (Weibull)	Expected frequency (Lognormal)
1	[0, 500000)	17	10.89	12.11	15.45	16.70
2	[500000, 1000000)	4	11.05	9.84	6.28	5.58
3	[1000000, 1500000)	5	4.31	3.97	2.61	2.01
4	[15000000, 2000000)	1	1.68	1.71	1.43	1.09
5	[20000000, 25000000)	1	0.65	0.76	0.89	0.69
6	[25000000, ∞)	1	0.26	0.42	2.02	2.68
	X^2 – statistic		9.96	7.40	5.62	7.81
	p – value		0.04	0.06	0.013	0.05

Table 4 shows that the Weibull distribution provides the best fit for the armed robbery data since it has the least Chi-square value of 6.62 and correspondingly highest p-value of 0.16 to confirm its suitability.

Table 5: Observed and Expected Values for Fitting Classic Distribution to Fire Data

Group	Claim Interval	observes frequency	Expected frequency (Exponential)	Expected frequency (Gamma)	Expected frequency (Weibull)	Expected frequency (Lognormal)
1	[0, 500000)	11	6.1856548	11.0679945	10.57667448	10.99001088
2	[500000, 1000000)	4	7.6427229	4.498262188	6.148396083	5.685554801
3	[1000000, 1500000)	2	3.9491641	2.177481407	2.446469863	1.969317696
4	[15000000, 2000000)	1	2.0406205	1.311693119	1.200845138	0.992137388
5	[20000000, 25000000)	1	1.0544338	0.855977409	0.648123694	0.588897047
6	[25000000, 30000000)	1	0.5448493	0.582891265	0.370807624	0.384668266
7	[30000000, 35000000)	1	0.2815357	0.407626972	0.220981404	0.267801089
8	[35000000, ∞)	1	0.1555431	0.807767858	0.251895564	0.926425147
	X^2 – statistic		14.15953	3.304043	6.134896	5.369405
	p – value					

Table 5 shows that the Gamma distribution provides the best fit for the fire data since it has the least Chi-square value of 3.30.

Table 6: Observed and Expected Values for Fitting Distributions to Motor Insurance Claims

Group	Claim Interval	Observed frequency	Expected frequency (Exponential)	Expected frequency (Gamma)	Expected frequency (Weibull)	Expected frequency (Lognormal)
1	[0, 500000)	203	125.5260	168.2732	167.0382	174.8747
2	[500000, 1000000)	18	90.5428	40.8684	57.6762	41.8910
3	[1000000, 1500000)	6	21.3462	15.8405	13.0144	11.4898
4	[1500000, 2000000)	6	5.0325	7.9679	3.9633	5.2185
5	[2000000, 2500000)	3	1.1865	4.4043	1.3890	2.9037
6	[2500000, 3000000)	3	0.2797	2.5593	0.5318	1.8119
7	[3000000, 3500000)	4	0.0659	1.5343	0.2168	1.2183
8	[3500000, ∞)	1	0.0048	1.6124	0.0777	3.7283
	X^2 value		587.5827	31.28094	130	30.01892

Table 6 shows that the Log-normal distribution provides the best fit for the Motor Insurance data since it has the least Chi-square value of 30.01892 and suggests that the Log-normal distribution is the most suitable for Nigeria's motor accident claims (Table 6).

Table 7: Observed and Expected Values for Fitting Classic Distribution to Commercial Line Data

Group	Claim Interval	Observed frequency	Expected frequency (Exponential)	Expected frequency (Gamma)	Expected frequency (Weibull)	Expected frequency (Lognormal)
1	[0, 500000)	463	244.10	392.59	361.92	389.80
2	[500000, 1000000)	51	226.28	89.61	151.10	110.30
3	[1000000, 1500000)	20	77.41	38.82	43.99	33.56
4	[15000000, 2000000)	16	26.48	22.05	16.79	16.21
5	[20000000, 25000000)	14	9.06	13.82	7.26	9.44
6	[25000000, 30000000)	7	3.10	9.12	3.39	6.11
7	[30000000, 35000000)	10	1.06	6.22	1.67	4.24
8	[35000000, ∞)	7	0.32	13.40	1.40	16.67
	X^2 – statistic		600.19	45.91	181.75	66.90
	p – value		0.000	0.000	0.000	0.000

Table 7 shows that the Gamma distribution provides the best fit for the commercial line data since it has the least Chi-square value of 45.91.

Table 8: Estimated Parameters for the Fitted Models

	Exponential	Weibull		Gamma		Lognormal	
Parameter	**λ**	**A**	**B**	**α**	**β**	**Σ**	**μ**
Fire	1.32E-06	0.70895	4.54E+05	0.36483	2.08E+06	1.5673	12.431
Motor	2.89E-06	0.71744	2.05E+05	0.25971	1.33E+06	1.7035	11.453
Property	1.63E-06	0.75188	2.96E+05	0.1381	4.44E+06	1.6055	11.913
Theft	1.30E-06	0.60708	4.05E+05	0.38516	2.00E+06	1.8707	12.067
Armed Robbery	1.88E-06	0.54452	4.13E+05	0.8045	6.60E+06	2.029	12.041
Portfolio	1.88E-06	0.54452	4.13E+05	0.8045	6.60E+05	2.029	12.041

Table 8 summarizes the fitted distributions for various lines of insurance business and also displays their estimated parameters. Finally, we present the Histograms, fitted distributions and the Q-Q plots for the various classes of insurance business in Figures 1-6.

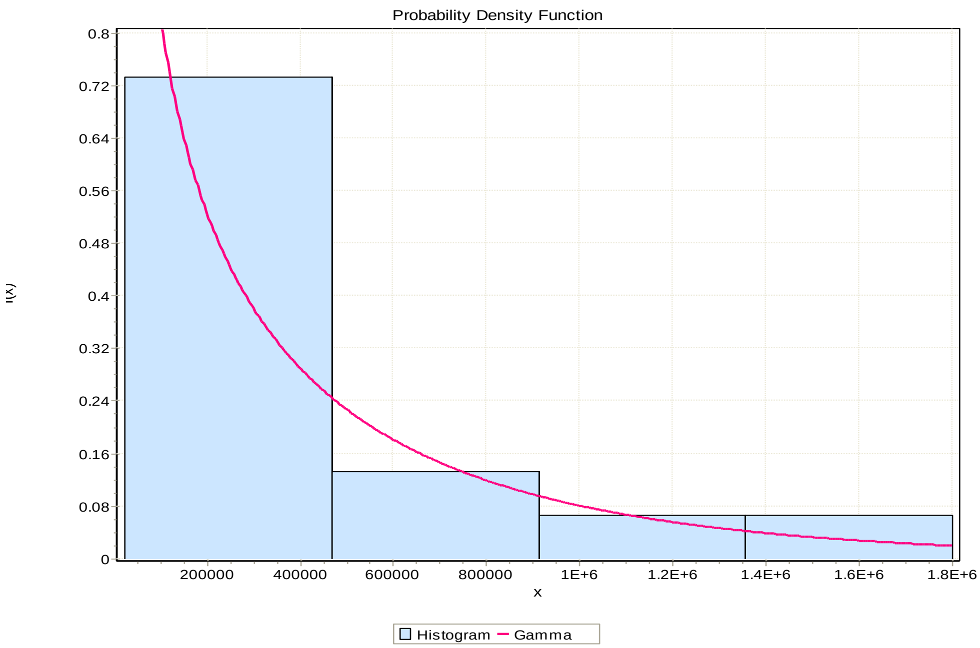

Figure 1a: Histogram and the Fitted Distribution for Fire

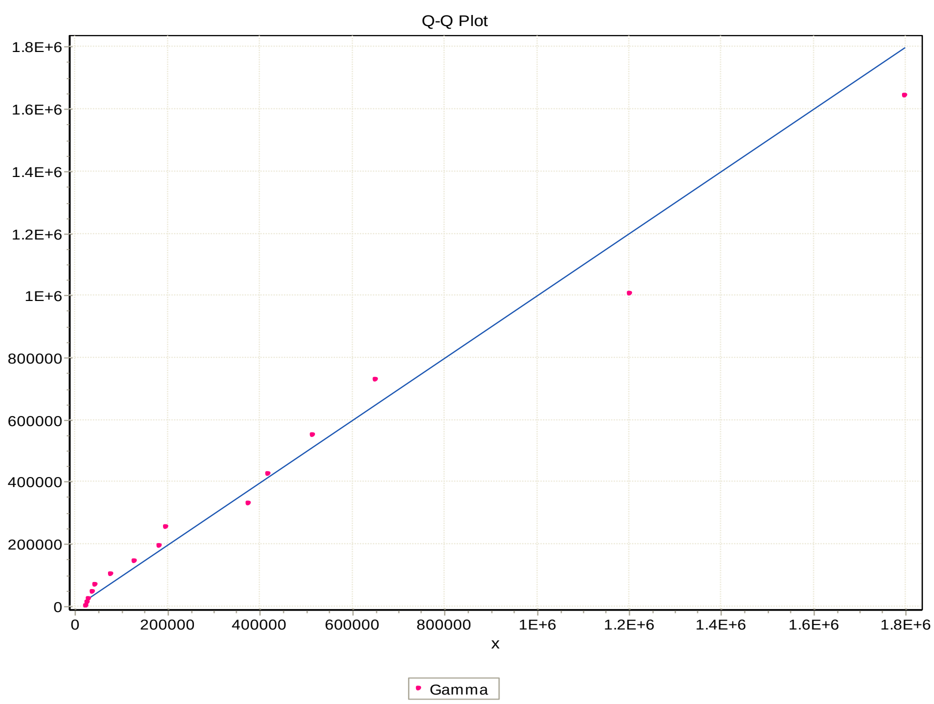

Figure 1b: Q-Q Plot for the Fitted Distribution for Fire

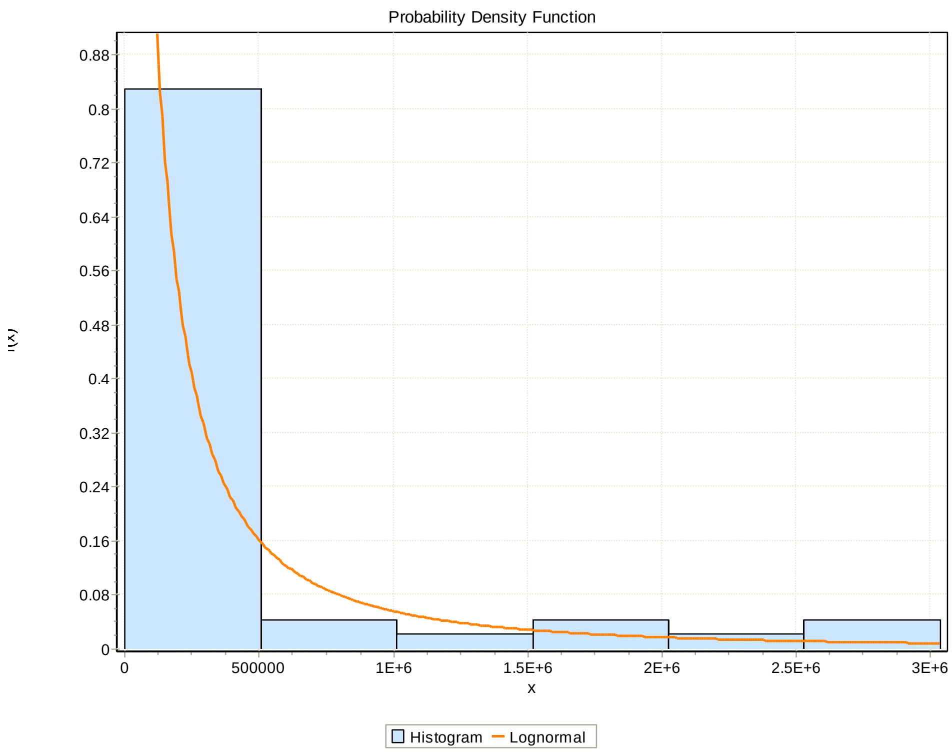

Figure 2a: Histogram and the Fitted Distribution for Motor

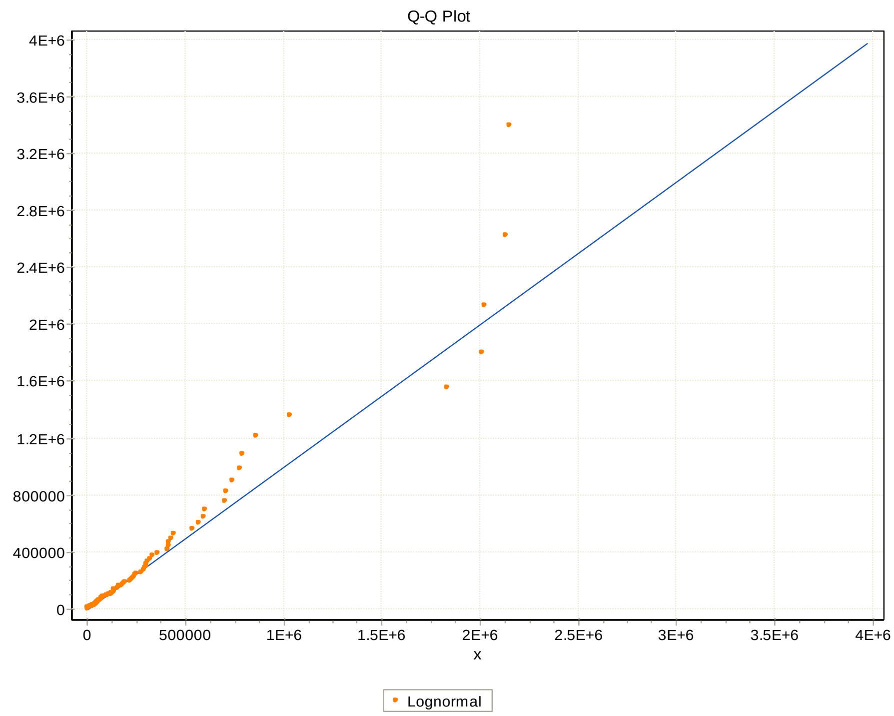

Figure 2b: Q-Q Plot for the Fitted Distribution for Motor

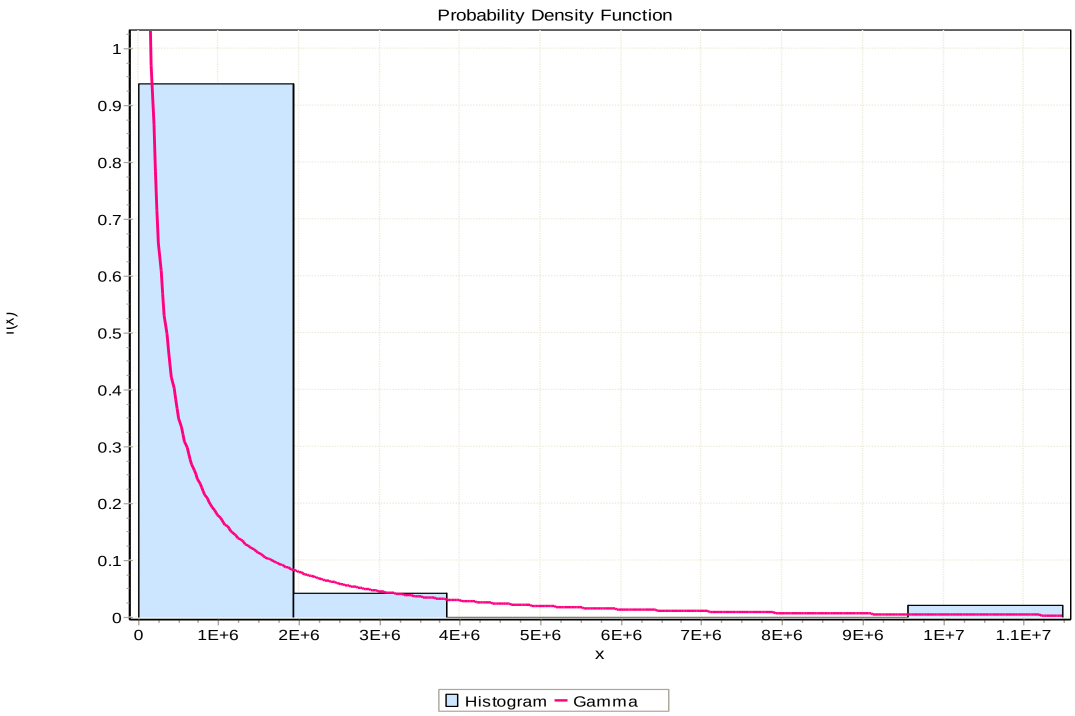

Figure 3a: Histogram and the Fitted Distribution for Property

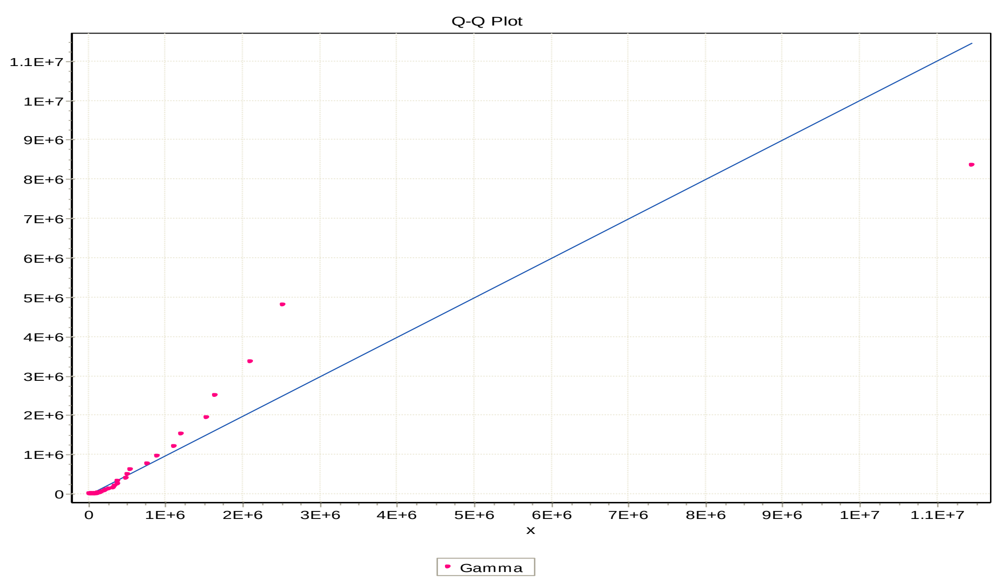

Figure 3b: Q-Q Plot for the Fitted Distribution for Property

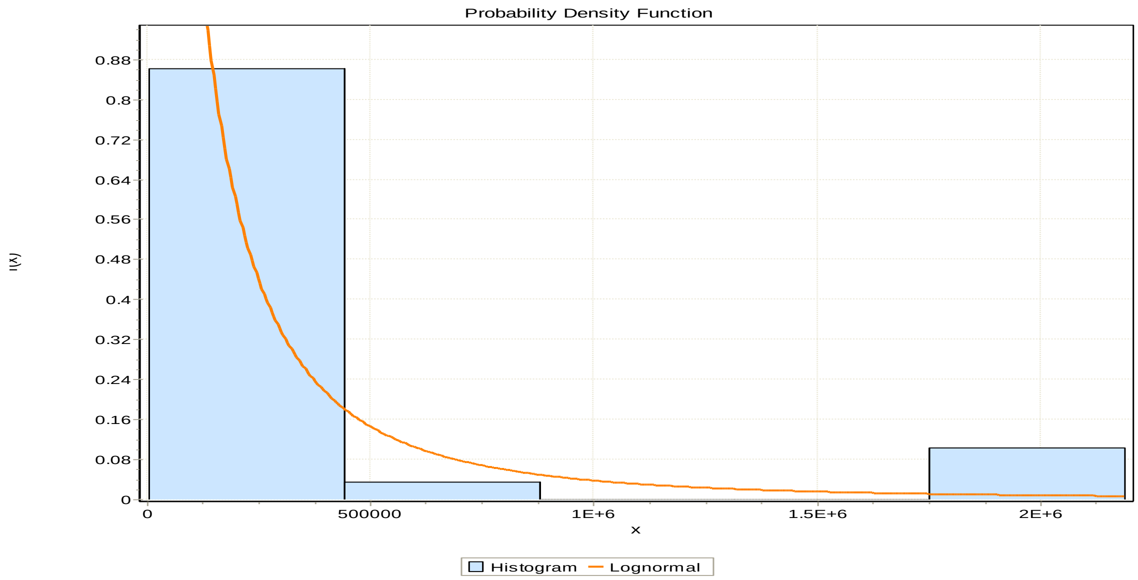

Figure 4a: Histogram and the Fitted Distribution for Theft

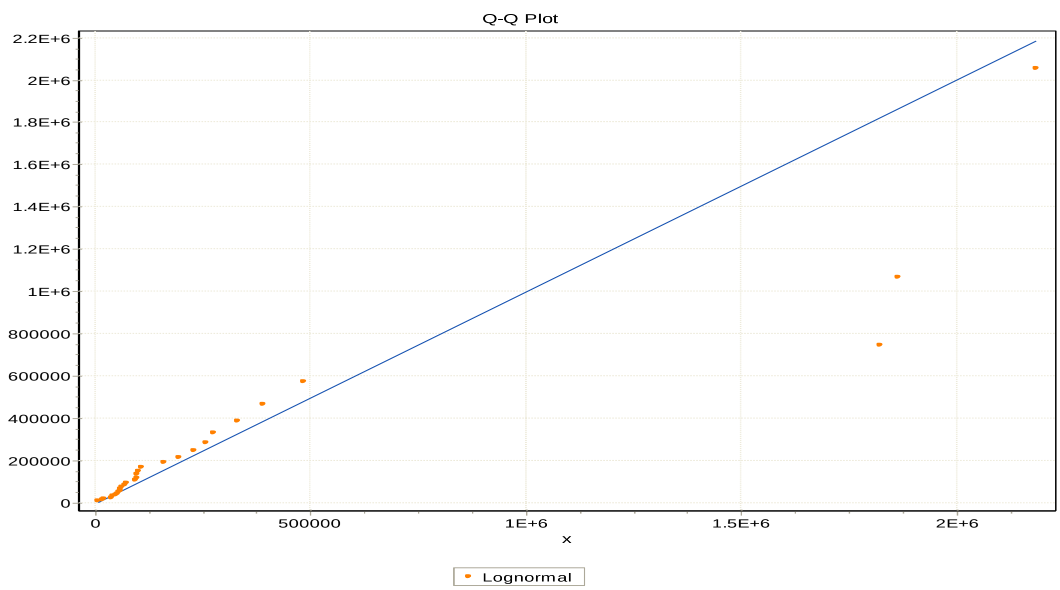

Figure 4b: Q-Q Plot for the Fitted Distribution for Theft

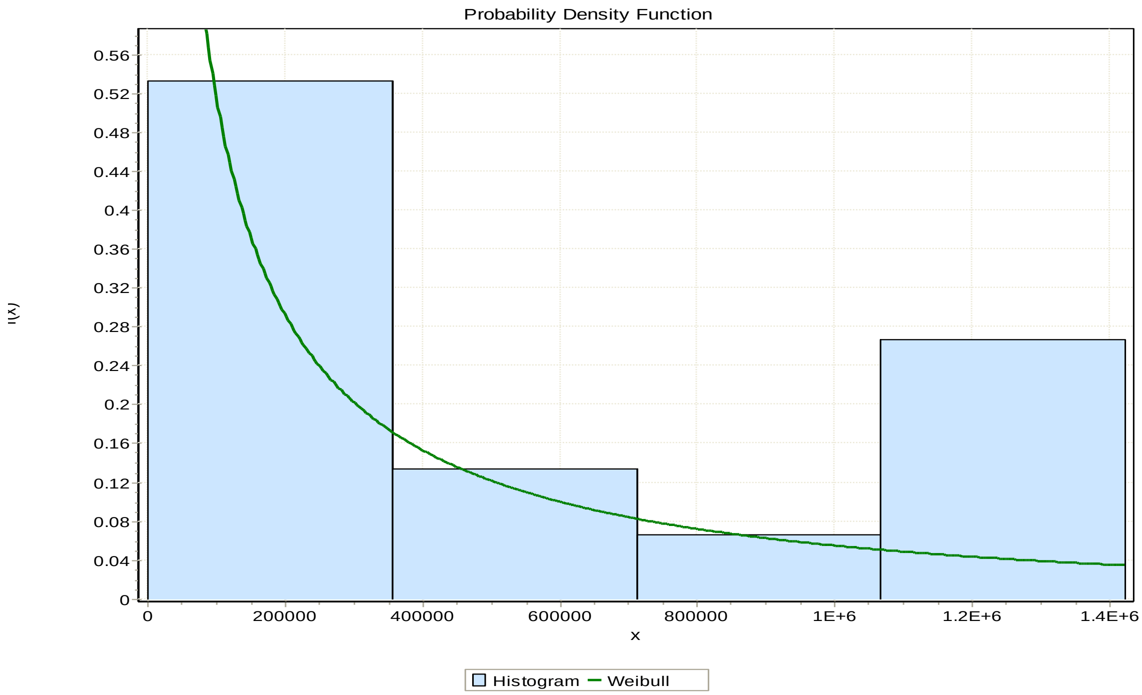

Figure 5a: Histogram and the Fitted Distribution for Armed Robbery

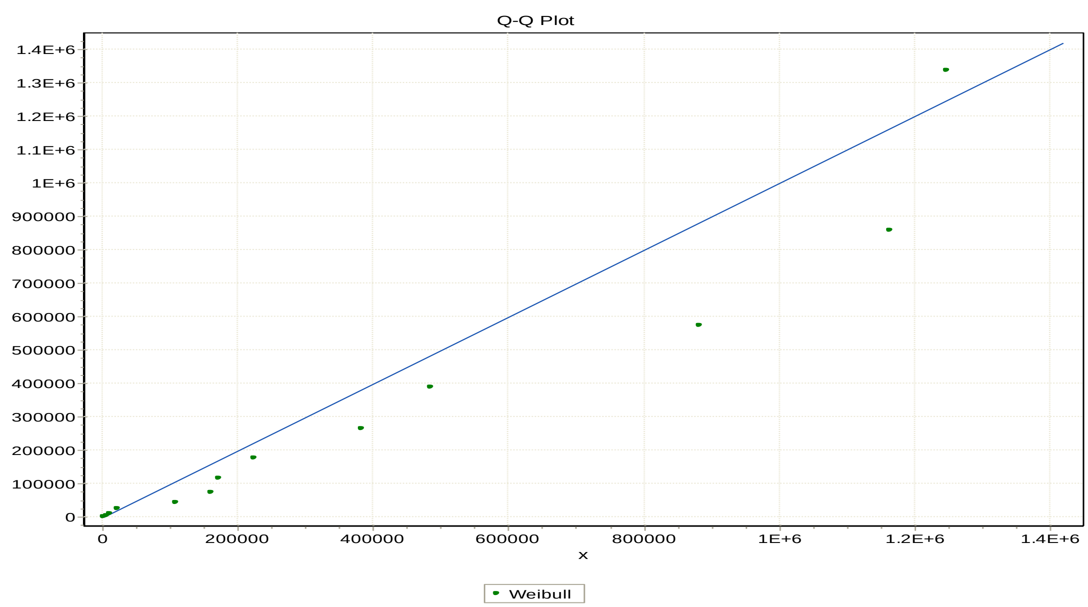

Figure 5b: Q-Q Plot for the Fitted Distribution for Armed Robbery

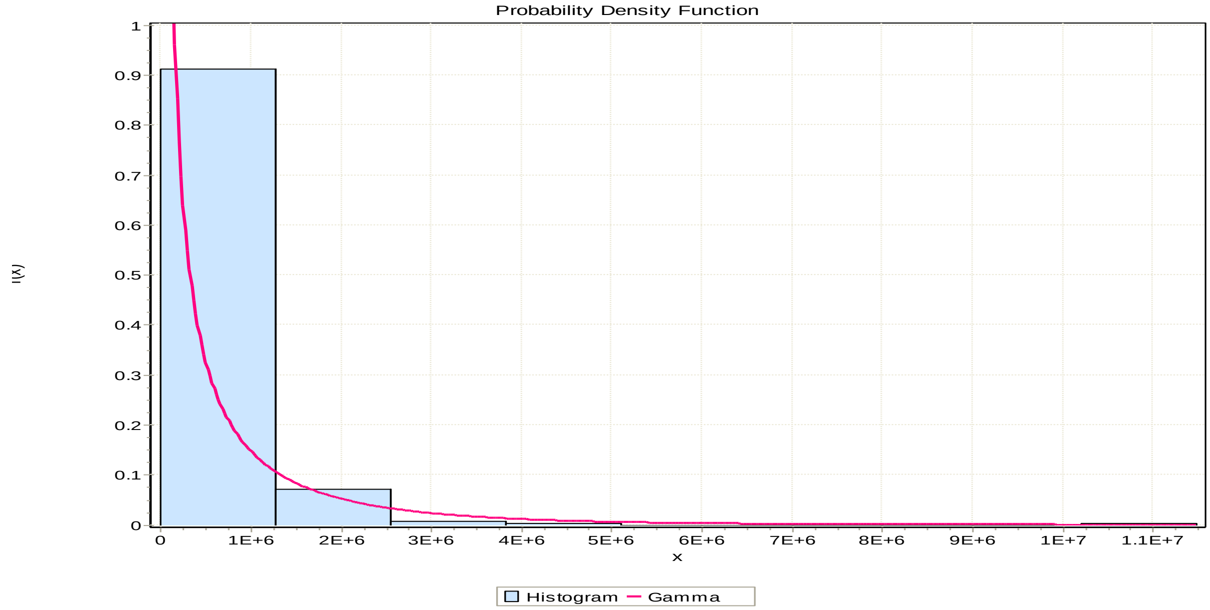

Figure 6a: Histogram and the Fitted Distribution for Overall claims

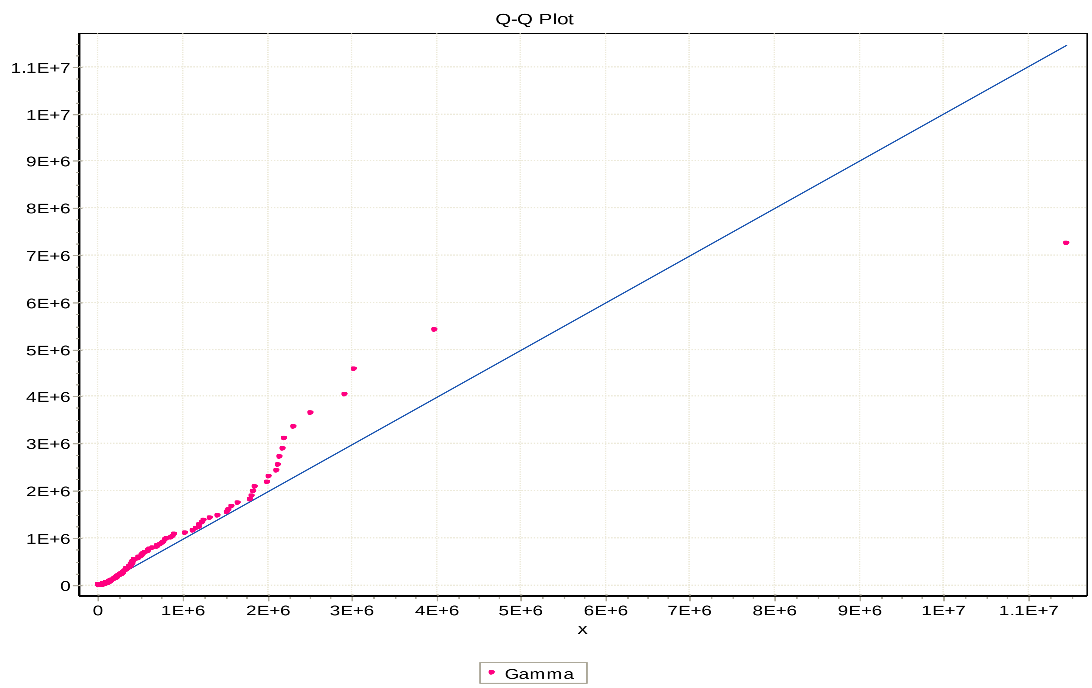

Figure 6b: Q-Q Plot for the Fitted Distribution for Overall Claims

Reinsurance

Claim size often increases over time due to inflation, and it is worth investigating how this affects typical payments for the ceding insurer and reinsurer if the same reinsurance treaty holds. Suppose that claims increase by a factor of k next year, but that the same excess level M is used in an excess of loss treaty between the insurer and reinsurer. Future cessions would be expected to increase as a result of inflation. Thus, the factor for the insurer would be greater than k since the total claim size, on the average, increases by k. Suppose that due to inflation next year, a typical claim X = Y+Z will have distribution X^*= kX, where k > 1. If Y is that part of the claim X handled by the (ceding) insurer this year, then next year it will be Y^* = g(X) defined by

$$Y^* = \begin{cases} kX & if \;\; X \leq M \\ M & if \; kX > M \end{cases}$$

The amount paid by the insurer next year on a typical claim X^* is given as

$$E(Y^*) = \int_0^{M/k} kxf_X(x)dx + \int_{M/k}^{\infty} Mf_X(x)dx$$

$$= k\left[\int_0^{M/k} xf_x(x)dx + \int_{M/k}^{M} (\frac{M}{k})f_{x(x)dx} + \int_M^{\infty}\left(\frac{M}{k}\right)f_{x(x)dx}\right] \qquad = kE(Y)$$

where *Y* is the amount paid by the baseline insurance company and *Z* is the amount paid by the reinsurer. But, it can be shown that

$$E(Y) = E(X) - \int_0^{\infty} yf_X(y+M)dy$$

In the case where *X* is exponentially distributed with parameter λ,

$$E(Y^*) = \left[E(X) - \int_0^{\infty} y\,\lambda e^{-\lambda(y+\frac{M}{K})}\right] = \frac{k}{\lambda}\left[1 - e^{-\lambda M/K}\right]$$

which yields

$$E(Y^*) = \frac{k}{\lambda}\left[1 - e^{-\lambda M/K}\right]$$

However, if *X* has a Pareto distribution with parameters α and λ, the $E(Y^*)$ can be shown to be

$$E(Y^*) = k \times \left[\frac{\lambda}{\alpha-1} - \frac{\lambda^{\alpha}}{\alpha-1}\left(\frac{1}{\lambda+M/k}\right)^{\alpha-1}\right]$$

For a Gamma distribution with parameters α and λ, the $E(Y^*)$ is

$$E(Y^*) = k \times \left[\frac{\alpha}{\lambda} - \left(\frac{\alpha}{\lambda} - \frac{M}{k}\right)e^{-\lambda M/k}\right]$$

Similarly, if X has a Weibull distribution with parameters α and λ, the $E(Y^*)$ can be expressed as

$$E(Y^*) = k \times \left[\frac{1}{c^{1/\gamma}}\Gamma\left(1+\frac{1}{\gamma}\right) - \left(\frac{1}{c^{1/\gamma}}\Gamma\left(1+\frac{1}{\gamma}\right) - \frac{M}{k}\right) e^{-c(M/k)^{\gamma}}\right]$$

Substituting the values of the estimated parameters of the various distributions, as obtained in Table 8, we obtain the expected cost per claim for the insurer and the corresponding expected payment per claim next year for the insurer given an interest rate of 11 percent (Business day, 2010), as forecasted by the apex bank.

ASSUMPTIONS

The discussions that follow require that we make some assumptions about the non-life insurance market in Nigeria. We base our assumption of the average cost of the purchase of a car on the cost of a fairly used car. The reason for this is that since the 1980s when the economic recession set in, the average Nigerian had resorted to the option of fairly used cars as the prices of new vehicles seem to have gone beyond his reach. From market survey, the average price of a fairly used car in Nigeria is about two million Naira (₦2,000,000). Also, the typical rate charged by member companies of the Nigeria Insurers Association is about 10%. Therefore, we have assumed that the average premium charged per vehicle is N200,000 (denoted M). We also conjecture that a realistic estimation would only be possible if allowance is made for inflation. The current rate of inflation, from published figures, is 11% (BusinessDay, 2010). On the foregoing basis, Table 9 is derived.

Table 9: Expected Loss at 11% Rate of Inflation

	k=11	**M= 200000**	
Exponential	**Gamma**	**Weibull**	**Lognormal**
197619.1	6.27E+07	1.44E+07	32130766
194836.3	5.63E+07	5.37E+07	18864785
197065.4	3.54E+08	6.91E+07	21606949
197654.9	5.71E+07	9.45E+09	63364464
196620.4	9.02E+07	9.24E+10	1.14E+08
196620.4	9.02E+06	9.24E+10	1.14E+08

Risk Premium Estimation

With the result obtained in the previous sections, we are now able to estimate the risk premium in each class of insurance business by taking the product of expected claim frequency and severity (Brockman & Wright 1992; Renshaw 1995; Haberman & Renshaw 1996). For motor insurance claims, for instance, the total loss, R, based on the fitted lognormal distribution is estimated as:

$$\hat{R} = N(\hat{\alpha} + \exp(\hat{\mu} + \hat{\sigma}^2 / 2)) = 31 \times 619319.8 = 18889254$$

Similar results hold for the calculated risk premium in the other classes of insurance business. The results are presented in Table 10.

Table 10: Estimates of Risk Premium

Class of Insurance	**Estimated Risk Premium**
Fire	12829591
Motor	18889254
Property	26525986
Theft	29036235
Armed Robbery	19924700
Commercial	357316278

CONCLUSION

This study has examined the main lines of non-life insurance business in Nigeria and has fitted appropriate theoretical distribution to each line. The study has established that a Gamma distribution would be best for the Property, Fire and the Commercial insurance products, lognormal for the Theft and Motor lines, while Weibull would best fit the Armed Robbery plan. With knowledge of these distributions and assuming the present rate of inflation is sustained, we were able to estimate the expected loss to an insurance company, as well as determine estimates of the premium. With these key figures, it becomes easier for an insurance company to allocate reserves or determine the level of cession to a reinsurance company or know when it is taking more than expected risk. To continue the previous approach where the uncertain business of insurance is run on the basis of guesses could hurt the industry in the long run.

AUTHOR INFORMATION

Adeleke, Ismaila Adedeji holds a B.Sc. (Hons.) degree in Statistics from the University of Ilorin and M.Sc. and PhD (Statistics) from the University of Lagos. He is a Lecturer in the department of actuarial science and insurance at the University of Lagos. His research interests include multivariate statistics, statistical modeling and survey techniques.

Ibiwoye, Ade holds a BSc. (Hons) degree in Mathematics from the University of Lagos, a Masters degree in Actuarial Science from the University of Manitoba and PhD in Operations Research from the University of Ilorin. Ade has cognate experience spanning over a decade in insurance underwriting as well as consulting in Pensions and Employee Benefits. Presently he teaches Actuarial and Financial Mathematics at the University of Lagos. His research interests include actuarial, optimization and pro.

REFERENCES

1. Antonio, Katrien and Jan Beirlant (2006), "Risk Classification in Non-life insurance," Proceedings of the 4th Actuarial and Financial Mathematics Day, Brussels
2. Barlow, R. E. and F. Proschan (1965), Mathematical Theory of Reliability and Life Testing, Probability Models, Holt, Rinehart and Winston, Inc., New York
3. Beirlant, Jan and Jozef L. Teugels (1992) "Modelling Large Claims in Non-Life Insurance," *Insurance: Mathematics and Economics*, Vol. 11, Issue 1, pp. 17-29.
4. Boland, Philip J. (2007), *Statistical and Probabilistic Methods in Actuarial Science*, Chapman & Hall/CRC, New York.
5. Bain, A. Stochastic Calculus, Updated 4/10/2008, available: chiark.greenend.org.uk/ alanb.
6. Boucher, Jean-Philippe; Michel Denuit, and Montserrat Guillen (2008) 'Models of Insurance Claim Counts with Time Dependence Based on Generalization of Poisson and Negative Binomial Distributions,' *Variance*, Vol. 2, Issue 1, pp. 135-162.
7. Brockman, M. J. and Wright, M. A. (1992) "Statistical Motor Rating: Making Effective Use of Your Data," *Journal of the Institute of Actuaries*, Vol. 119, No. 3, pp. 457-543.
8. Butt, Zoltan and Steven Haberman (2004) "Application of Frailty-based Mortality Models Using Generalized Linear Models," *Astin Bulletin*, Vol. 34, No. 1, pp. 175-197.
9. Copeland, T. E.; J. F. Weston and K. Shastri, *Financial Theory and Corporate Policy*, Fourth edition, Addison-Wesley, 2003.
10. Davis, M. (2000), Mathematics of Financial Markets in Mathematics Unlimited - 2001 and beyond
11. Hamadu, Dallah (2010) "Modelling Armed Robbery Crime Insurance Claims in Nigeria," *Global Journal of Management and Business Research*, Vol. 10, pp11-20
12. Hogg, Robert V.; Joseph W. McKean and Allen T. Craig (2005), *Introduction to Mathematical Statistics*, 6th Edition, Pearson Prentice Hall, New Jersey.
13. Keiding, Neils, Christian Andersen and Peter Fledelius (1998) "The Cox Regression Model for Claims Data in Non-Life Insurance," *Astin Bulletin*, Vol. 28, No. 1, pp. 95-118
14. McLaughlin, P. Micheal (2001), A Compendium of Common Probability Distribution, Regress+ Version 2.3, www.causascientia.org/math_stat/Dists/Compendium.pdf. sited on 4th, August, 2010.

15. NIA, Nigerian Insurers Association (2005), *Annual Digest*.
16. Osoka, Ogala (1992) *Insurance and the Nigerian Economy*, Panache Publications, Lagos.
17. Promislow, S. David (2006), *Fundamentals of Actuarial Mathematics*, John Wiley & Sons, Chichester.
18. Renshaw, A. E. (1995), "Graduation and Generalized Linear Models: An Overview," Actuarial Research Paper No. 73, Department of Actuarial Science & Statistics, City University, London.
19. Ross, S. (2006), *A First Course in Probability*, Seventh Edition, Prentice-Hall, New Jersey.
20. Yip, Karen C. H. and Kelvin, K. W. (2005), "On modeling claim frequency data in general insurance with extra zeros," *Insurance: Mathematics and Economics*, Vol. 36, Issue 2, pp. 153-163.

