



Probability Modeling and Simulation of Insurance Claims in Ghana

Godfred Kwame Abledu¹, Elvis Dadey & Agbodah Kobina

1. Corresponding author: Applied Mathematics Deptment, Koforidua Polytechnic-Ghana

ABSTRACT

Probability modeling has a wide range of applications in the field of insurance and finance. Over the years, its applications have been modified to include simulation techniques which in most cases are engaged to enhanced or validate the desired estimates. This study investigates the probability distributions that best fit a number of insurance claims. In particular, it compares the Poisson distribution and the negative binomial distribution models to determine which distribution best fits insurance claim data obtained from two Insurance Companies in Ghana.

Data on number of claims of a funeral policy spanning from 2006 to 2010 were used for the study. Probability distribution models and the parametric bootstrap methods were employed in analyzing the data collected. The result of the study revealed that on the average, the number of claims tends to increase as the year go by. Also the Negative Binomial Distribution was found to be superior to the Poisson distribution in fitting the claims data. Finally, a comparison between the estimates obtained by the probability models and that of the parametric bootstrap estimates revealed no significant difference.

Keywords: parametric bootstrap, insurance claims, simulation techniques, negative binomial distribution, Poisson distribution.

Introduction

The risk involved in allocating insurance funds has boosted the need for statistical methods and principles. Insurance funds may be invested in several asset categories, which include bonds, equities and others with the aim of increasing investment real returns to meet claim payment demands and other financial obligations.

The puzzle of percentage allocation of insurance fund into investment portfolios is subject to insurance liabilities of which part is systematic and the other stochastic. Insurance liability relating to stochastic constraints, which is of much concern to this research work forms comparatively huge part of the total liability of insurance companies. It is solely linked to claim demands and uncertainties relating to it by which the policy contract mandates the insurer to forfeit. Insurance companies employ statistical methods of estimation to acquire concrete information about these uncertain liabilities with which decisions pertaining to assets allocation, expected monthly claims and payment targets as well as future insurance pricing can be ascertained. This is known in insurance administration as the expected claim liability estimation problem.

The roots of insurance might be traced to Babylonia, where traders were encouraged to assume the risks of the caravan trade through loans with interest that were rapid only after the goods had arrived safely. The Phoenicians and the Greeks applied a similar system to their seaborne commerce. The Romans used burial clubs as a form of life insurance, providing funeral expenses for members and later payments to the survivors. (<http://www.infoplease.com/business/insurance>)

Insurance developed rapidly with the growth of British commerce in the 17th and 18th centuries. Prior to the formation of corporations devoted solely to the business of writing insurance, policies were signed by a number of individuals, each of whom wrote his name and the amount of risk he was assuming underneath the insurance proposal. (<http://encyclopedia2.thefreedictionary.com/Insurance>).

The first stock companies to engage in insurance were chartered in England in 1720 and 1735. Fire insurance corporations were formed in New York City in 1787 and at Philadelphia in 1794. The Presbyterian synod of Philadelphia sponsored the first life insurance corporation in America for the benefits of Presbyterian ministers and their dependants. After 1840, with the decline of religious prejudice against the practice, life insurance entered a boom period. In the 1830s the practice of classifying risks began. (<http://www.infoplease.com/encyclopedia/business/insurance>).

In the 19th century, many friendly societies were founded to insure the life and health of their members. Similarly, fraternal orders and most labour organizations continued to provide insurance coverage for its members. Currently, many employers sponsor Group Insurance Policies for their employees. Such policies generally include not only life insurance but sickness, accident benefits and old-age pensions. The employees usually contribute a certain percentage of the premium. Since the 19th century, there has been a growing tendency for the state to enter the field of insurance, especially with respect to safeguarding workers against sickness and disability, either temporary or permanent, destitute old age and unemployment.

(<http://www.infoplease.com/encyclopedia/business/insurance>).

Classification

Insurance policy may be classified into two types namely, Life Insurance policy and Non-Life Insurance policy. Life insurance policy is originally conceived to protect a man's family when his death left them without income. This has developed into variety of policy plans. For example, in a Whole Life insurance policy, fixed premiums are paid throughout the insured's life time. This accumulated amount, augmented by compound interests, is paid to a beneficiary in a lump sum upon the insured's death. The benefits are paid even if the insured had terminated the policy.

Universal life insurance was created to provide more flexibility than whole life insurance by allowing the policy owner to shift money between the insurance and savings components of the policy. Premiums, which are variable, are broken down by the insurance company into insurance and savings, allowing the policy owner to make adjustments based on their individual circumstances. For example, if the savings portion is earning a low return, it can be used instead of external funds to pay the premiums. Unlike whole life insurance, universal life allows the cash value of investments to grow at a variable rate that is adjusted monthly.

(<http://www.investopedia.com/terms/u/universallife.asp>).

A typical example of a Non Life Insurance policy is Fire Insurance. This includes damage from lightning, tornado, flood and drought. Complete automobile insurance includes not only insurance against fire and theft but also against compensation for damage to a car and for personal injury to the victim of an accident (liability insurance); many car owners; however carry only partial insurance (<http://encyclopedia2.thefreedictionary.com/Insurance>).

In many countries, liability insurance is compulsory. Bonding or fidelity insurance is designed to protect an employer against dishonesty on the part of an employee. Title insurance is aimed at protecting purchasers of real estate from loss by reason of defective title. Credit insurance safeguards business against loss from the failure of customers to meet their obligations. Marine insurance protects shipping companies against the loss of a ship or its cargo, as well as many other items(<http://encyclopedia2.thefreedictionary.com/Insurance>).

In recent years, the insurance industry has broadened to guard against almost any conceivable risk; companies will insure a dancer's legs, a pianist fingers or an outdoor event against loss from rain on a specified day. A hybrid type of insurance is what we call composite insurance policy, which combines both life and non life insurance.

Insurance in Ghana

Insurance in Ghana began in the colonial era, 1924, with the establishment of the Royal Guardian Enterprise now known as the Enterprise Insurance Company Limited. The first indigenous private insurance company, the Gold Coast Insurance Company was established in 1955. In 1962, the State Insurance Company of Ghana (SIC) was set up. Eleven (11) more companies were established in 1971. Five years later seven (7) more insurance companies, one (1) reinsurance industry and insurance brokerage firms were established.

The 1989 insurance law, PNDC Law 229 led to the establishment of the National Insurance Commission. The National Insurance Commission (NIC) remains the sole institution that has been mandated to regulate and supervise insurance activities within the country. The insurance industry is thriving so much that the business is quite common across the globe. The major insurance companies in Ghana can be grouped into: (i) Life Insurance, (ii) Non-Life Insurance, and (iii) Composite Insurance (a combination of Life and Non-Life insurance).

The insurance industry as of 2010 statistics was made up of the following: 23 Non-Life companies, 17 Life companies, 2 Reinsurance companies with over 40 Broking companies, one reinsurance broking company, one loss adjusting company and about 4,000 insurance agents.

In Ghana, insurance is an accepted thing to do, as everyone must protect themselves against risks in life. The needs of insurance in Ghana for example cuts across travel insurance, fire insurance, motor insurance, marine insurance etc. Though quite competitive, the governing laws of insurance have also increased with time. For example, in Ghana, Insurance Act 58 stipulates that all car owners insure their vehicles against unintended risks and damages. Despite the many posing questions asked on insurance, Ghanaians are gradually accepting the fact of paying premiums to protect themselves and their loved ones in case of loss and damage.

This can be attributed to the rude awakening of the Great Fire of London in 1666 destroying over 13,200 houses and lives as well as other Ghanaians properties (houses and market buildings) that were damaged by fire and flood in the early years without the owners insuring them. To a greater extent the problems faced by the insurance industry lies in risk management in the wake of financial crisis, that is maintaining adequate capital and surplus in support of continuous operations. Poor public perception of insurance has also resulted in lack of confidence in insurance. Again changes in sales and distribution as a result of decline in business volumes and changes in customer focus and related operating modules poses as a threat to the industry. Ineffective international financial reporting standards and ineffective investment management procedures are just a few to highlight within the industry. (<http://mybusinessweekafrica.com/topheadlines-detail.php>).

Statement of the Problem

The statistics on industries in Ghana revealed that insurance companies have increased appreciably over the last five (5) years and the demand of Ghanaians to secure insurance policies for their business and households is overwhelming (<http://www.nicgh.org/live/en/?pg=104.pp>). To this end, Insurance Companies are repositioning themselves to meet the future challenges of demand for Insurance Policies, Risk Management and Investments as well as Claim Payment Demands. In view of the fact that every policy holder expects a cushion in the event of his/her economic loss as stipulated in an Insurance Contract, the challenge of meeting the payment terms becomes an issue of much concern to the Insurer.

The estimation of Expected Claim Liabilities of the Insurance Policies enables decisions on asset allocation and claim payment to be taken without much error thereby rising above the said challenges. The aim of this research is to explore probability models to model a number of Insurance Claims and to estimate the expected claim Liabilities of Insurance Policies.

Research Objectives

The main objective of the study was to explore probability models that would model adequately the number of claims occurring under funeral insurance policies in Ghana to estimate the expected number of claims.

The specific objectives of the study are to:

- (i) Identify and explain seasonal variation within the number of claims.
- (ii) Compare the Probability distribution estimates of the number of claims.
- (iii) Derive the Probability distribution model that best fits the number of Claims for the funeral policies.
- (iv) Construct bootstrap confidence intervals for the Expected Number of Claims and compare with that of the estimates obtained from the models.

Literature Review

Probability models of count data provide essential information for Insurance Pricing, Insurance Liability Estimation and Investment Decision Making. Claims and unexpired risks provisions in insurance companies are determined based on previous claims experience, knowledge of the events and the terms and conditions of the relevant policies and on interpretation of circumstances.

Most insurance companies set claim payment targets by considering the estimates obtained by predicting the number of claims for some unforeseen periods. Insurance claim liability estimation is a sub problem of insurance liability estimation. Most insurance companies prefer to tackle the problem holistically, but all the same the claim liability is estimated somehow. An example is the Bornhuetter- Ferguson loss reserve method of insurance liability estimation (Asmussen and Kliippelberg, 2008)

The product of the expected number of claim and the benefit paid under the policy is essentially the expected claim liability when the policy is a fixed benefit policy and all economic factors such as inflation, interest rates etc are all held constant. It is noteworthy that a good estimate of the probability distribution model may lead to a good estimate of the claim liability that will improve on other estimates and the overall insurance liability estimate.

In statistics, stochastic count data refers to data in which the observations are uncertain and can take only non-negative integer values, i.e. $\{0, 1, 2, 3, \dots\}$. These integers arise from counting rather than ranking and that the statistical treatment of such data is distinct from that of binary data, in which the observations can take only two values, usually represented by 0 and 1. The methodology of stochastic count data modeling also known as Probability Modeling evolved from just an assumption and parameter estimation to modeling and has grown considerably in the past years with a lot of new procedures and methods. This evolution has taken place in both discrete and continuous probability modeling in diverse field of the sciences with most effects felt in the Statistics and Engineering Sciences (Yaacob et al., 2010).

An appreciable amount of work has being done, making up theories and methods that have added to the modeling body of knowledge. The theories and methods invented dealt proficiently with the complexities and have rendered the resolutions mechanic. The Poisson distribution model of stochastic count data requires strong independence assumptions with regards to the underlining stochastic process and any violation of these assumptions invalidates the model. One important features of the Poisson distribution that has justified its inadequacy in stochastic count data modeling is the variance-to-mean ratio, also called Fisher dispersion index, which is equal to one (1) whatever the value of λ . This inappropriateness of modeling count data with the Poisson distribution has long been recognized, and collectively risk algorithms have being developed that works just as well as other frequency distributions, in particular the Negative Binomial Distribution (Asmussen, and Kliippelberg, 2008; Mikosch and Nagaev, 2007).

In this regards various test for detecting over dispersion have also been developed. Again more general models have been proposed that allow for over dispersion. Furthermore, its effect has been analyzed and corrected for within the maintained Poisson model (Schlegel, 1998). Other research works (Terceño, and Gil, 2012; Osgood, 2000) have ascribed this inappropriateness to the under dispersion and over dispersion in count data with the later being the specific case of the count data for insurance claims. The authors consequently defined over dispersion with respect to insurance claims modeling as the case where there is evidence that the observed random variation is greater than the expected random variation under the Poisson model while under dispersion means that the expected variation is greater than the observed variation.

Konstantinides et al., (2002) compared the adequacies of the Poisson model to each of the mixed Poisson models as well as the mixed Poisson models with the Belgian motor third party insurance portfolio observed during the year 1997 and concluded that the mixed Poisson models fits better than the Poisson model. Again the comparison amongst the mixed Poisson models, which include the Poisson-Normal, Poisson-Lognormal and the Poisson-Inverse Gaussian model revealed no significant difference.

Mikosch, and Nagaev (2007) in their quest to model count data employed the negative binomial distribution resulting from Poisson-Gamma mixture to investigate small sample properties of the likelihood-based test and compared their performances to those of the Wilcoxon and t-test. The authors established the inefficiencies of the Poisson model and illustrated procedures used to compute power and sample sizes to design studies with response variables that are over dispersed.

Konstantinides et al., (2002) and Hall and Yao (2003). modeled crush count data with the object to propose an approach to compare the techniques of count data modeling. The authors proposed a simulation approach that used Negative Binomial Regression Mean Estimate (NB), Negative Binomial Empirical Bayesian Estimate (NBE) and Pearson regression residuals to rank sites and to identify hot spots. The results deduced showed that a small fraction of the artificial site can be successfully identified regardless of the methodology and Pearson regression led to the best results.

Alternative model expressions in the family of the Mixed Poisson models were proposed by Asmussen and Albrecher (2010) by introducing an additive and multiplicative random effect Poisson model. The author modeled injected random effects with the Gamma and Inverse Gaussian distributions, which yielded both univariate and multivariate models in a form of mathematical expressions. The idea was that count data often exhibit dispersion and unobserved heterogeneity, which renders the usual Poisson model ineffective thereby soliciting for other appropriate models. This idea was also evident in the research paper by Asmussen and Kliippelberg (1996) when they intended to fit a

distribution to a data set on claim size in order to obtain rough estimate of risk premiums.

Random effect models, which provided the best fits were recommended by Ridout et al.(2001) and Migue and Olave (1999b) in their investigation into models of time dependent insurance claim count based on Generalized Poisson and Negative Binomial distributions. The recommendations were based on model fitting and comparisons by statistical specification test for nested and non nested models. Further deductions suggested that these random effect models can be applied on other claim count distributions such as Hurdle or Zero-inflated model that provided good fits for cross-sectional data.

Yue(2003),Thombs and Schucany (1990) used a nonparametric bootstrap method to obtain the prediction interval of the autoregressive model. Migue and Olave (1999a) applied this method to the ARCH model. Rodriguez and Ruiz (2008) proposed using the bootstrap method for state space models, to evaluate estimation errors. However, while developing prediction intervals, there were usually two limitations: (1) how the uncertainty of parametric estimation can be combined for the prediction interval and (2) the assumption of error may not be accurate.

To overcome these limitations, Pascual et al. (2006) proposed that, for the GARCH model, the bootstrap method used to create the prediction interval to combine with the uncertainty of parametric estimation and assumptions that are independent of e_t . However, when e_t had heavy-tailed distribution, this study expanded the method of Pascual et al.(2006) and added the concept of the sub-sample bootstrap method proposed by Hall and Yao (2003) and created the prediction interval in the GARCH model that had heavy-tailed distribution. Lee (2009) applied the bootstrap strategy proposed by Pascual et al. (2006) to study listed daily stock returns in Taiwan. The results showed that the volatility intervals created using the bootstrap method could appropriately manage the changes of volatility returns.

These and other studies are the many studies conducted in the field of probability modeling. Today, the results obtained from stochastic data modeling are rarely applied without some simulation due to unidentified stochastic shocks. In this regards, results obtained from these models are subjected to Monte Carlo simulation comprising bootstrapping, jackknifing etc in order to further perceive the results for adequate industrial decision to be made.

Methodology

The study focused on estimating the expected number of claims under funeral insurance policies sold by two insurance companies namely StarLife and MetLife insurance company. Data for the study included monthly number of claims filed under the policies spanning for a period of sixty (60) months (January, 2006 to December, 2010). Mathworks Inc.'s Technical programming Language MatLab, version 7.10.0.499 (R2010a) and the SAS Institute Inc.'s software, version 1.0.0.80 were the main statistical softwares used to perform the analyses. Microsoft Office Excel 2007 and R software version 2.13.0 were used to organize the data and produce graphs for the analysis. The final report was written using the Jonathan Kew, Stefan Loffer version 0.4.0 r. 759 (MikTeX) of TeXworks and TeXnicCenter version 1 Beta 7.50. It is a simple environment for editing, typesetting and previewing TeX documents.

Descriptive statistics of the data were generated to study the distribution of the number of claims for the entire period and on annual basis. Shapiro-Wilk test of normality and two sided Kolmogorov-Smirnov test were performed to draw conclusion on the nature of the distributions of the claims. Poisson probability distribution (PoisDis) and Negative Binomial distribution (NegBin) were used to fit the number of claims and the distribution which produced a better fit was subjected to an asymptotic chi-square test as a confirmatory analysis.

Finally the estimated distribution was simulated to obtain 100 re-samples of artificial data sets for monthly number of claims and the desired statistics recalculated. This was followed by comparative analysis to address the study objectives.

Data Analysis

The data used for the research were solely secondary data obtained from Star Life and Metropolitan life Insurance Companies. The data consisted of monthly recorded number of claims under a Family Funeral Insurance Policy for a period of Five years (2006 - 2010). The analyses were conducted in two folds, which included preliminary analyses and some further detailed analyses.

Exploratory and Confirmatory data analyses were the main techniques employed in the preliminary analysis. First of all, the number of claims of both StarLife and MetLife insurance companies were summarized with relevant summary statistics and compared to observe the nature of similarities or otherwise. The number of claims were analyzed on annual basis to measure the seasonal variations or otherwise that existed within and between the number of claims for both portfolios.

Again, visual representations in the form of Bar Graph, Histogram, Box Plots, QQ Plots and Density Plots were inspected in other to compare the underlining distribution of the number of claims to the empirical normal distribution. Finally, the results of the preliminary analysis were confirmed accordingly with the Shapiro Wilk test of normality and two sided Kolmogorov-smirnov test of similarity of distribution. The Shapiro-Wilk test was used to test the null hypothesis that a sample $x_1, x_2, x_3, \dots$ came from a normally distributed population.

Results and Discussions

The analysis focused on modeling the number of claims under the funeral policies sold by StarLife and MetLife insurance companies. The paramount purpose was to fit an appropriate probability distribution model to the sixty (60) data points of monthly recorded claims observed from the beginning of 2006 to the end of 2010 in order to estimate the expected number of claims to aid investment allocation.

The Poisson and Negative binomial distributions are considered for modeling whereas the simulation method employed was bootstrapping. The Pearson chi square test together with the scaled deviance test informed the adequacy of the models.

Summary Statistics

Figure 1 presents a bar plot of the number of claims for MetLife and StarLife Insurance companies on funeral policy for the period (2006 - 2010). A glance at Figure 1 reveals heavy-tailedness and skewness to the right of number of Claims for both companies. This distribution is typical of number of insurance claims. The number of claims fluctuate increasingly within the period for both claims with Metlife experiencing claims between zero (no claim) and sixty-four (64) claims and StarLife between zero (no claims) and ninety 90 claims. The mean number of claims for MetLife and StarLife over the period (2006 - 2010) was seventeen (17) and thirteen (13) claims respectively (see Table 1).

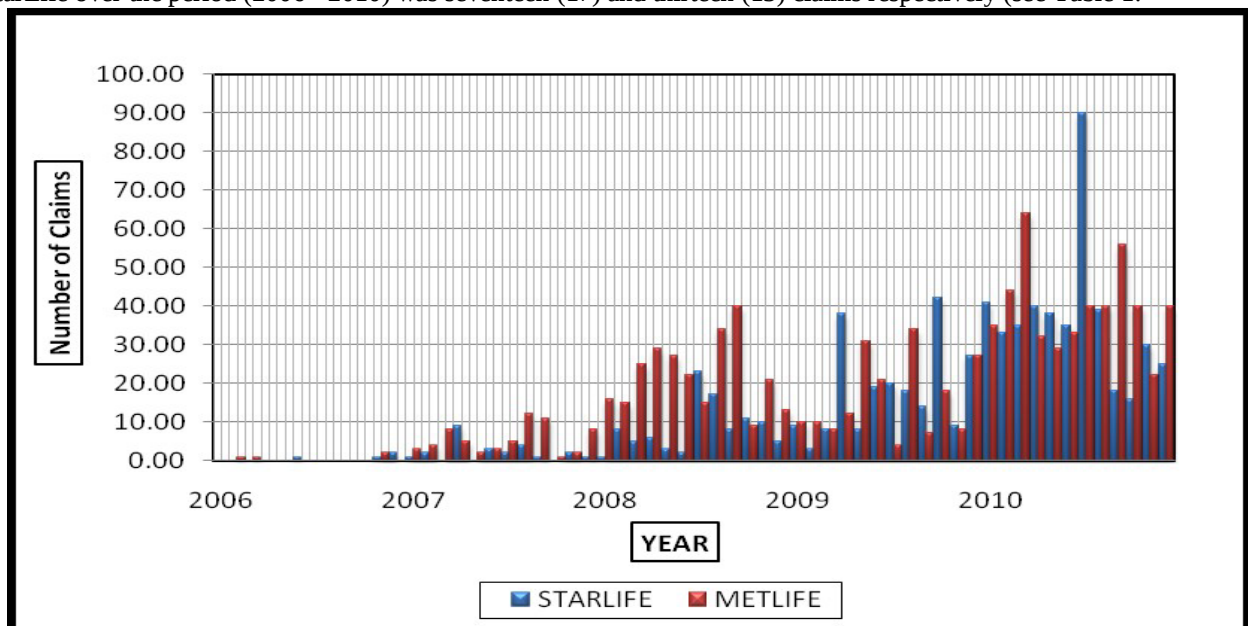


Figure 1: Bar plot of Number of Claims on Funeral Policy from StarLife and MetLife.

Table 1: Summary Statistics of Number of Claims of Funeral Policy from Starlife and MetLife.

Year	StarLife	MetLife
Mean	13.05	16.65
Median	7.00	11.50
SD	16.82	15.98
Skewness	2.07	0.89
Kurtosis	6.11	0.07

Seasonal Analysis

A study of the seasonal changes in the occurrence of the number of claims revealed an increasing pattern for both Portfolios along the period. The highest number of claims was observed in July and March, 2010 for StarLife and MetLife respectively. Several months between 2006 and 2007 recorded no claims (zero claims), that was linked to the reason that the policy was introduced in 2005 and as at the end of 2006, not much of the policies had being sold.

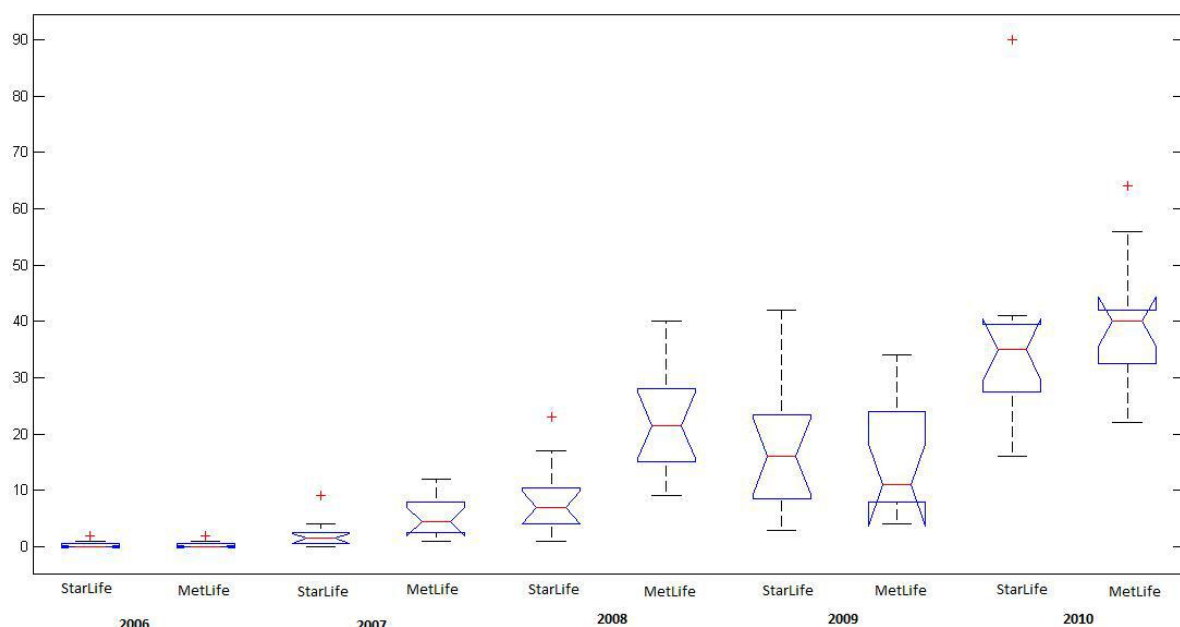


Figure 2: Boxplot of Yearly Number of Claims on Funeral Policy from StarLife and MetLife.

Figure 2 is a box plot of the monthly recorded number of claims on yearly basis for the underlying years (2006 - 2010). The yearly distributions of the number of claims were all tailed and skewed to either the left or right in the various years. The annual distributions of number of claims for MetLife were skewed to the right while those of StarLife were skewed to the left in 2010 but in the other years under review they were all skewed to the right (2006 - 2009). This revealed that the annual distributions were not conventionally bell shaped (normally distributed) in any of the years but were all tailed and skewed.

Considering, the annual distributions during the active years (2008 - 2010) where much of the policies had been sold, the year 2008 distribution of the number of claims from MetLife was heavily tailed, right skewed and flat arched (platykurtic) while that of StarLife was slightly tailed, right skewed and slender arched (leptokurtic). Furthermore, year 2009 distribution depicted a tailed, right skewed and slender arched (leptokurtic) for MetLife and tailed, right skewed and normally peaked (mesokurtic) for StarLife.

Finally in year 2010, the distributions for both portfolios were similar, tailed, asymmetric and slender arched (leptokurtic). Overall, the number of claims depicted increasing patterns across the years with rising annual averages (see Table 2).

Table 2: Trend of Annual Averages of Number of Claims from Starlife and MetLife

Year	2006	2007	2008	2009	2010
StarLife	0.33	2.08	8.25	17.72	36.67
MetLife	0.33	5.33	22.18	15.83	39.53

Comparison of the Distributions

The preliminary analysis on the number of claims suggests that the distributions of both portfolios were not normal. The null hypothesis of normality was rejected at a p-value of 1.28×10^{-8} for StarLife and 1.28×10^{-8} for MetLife according to Shapiro-Wilk test of normality.

To answer the question as to whether the distribution of number of claims for both insurance companies came from the same distribution, a QQ plot displayed in Figure 3 was constructed and the fairly linear trend preliminarily suggest that the distribution of the number of claims for both insurance companies belong to the same distribution. Moreover, the p-value of 2.13×10^{-1} obtained by the two sided Kolmogorov-Smirnov test was statistically significant to ascertain that the probability distribution of the number of claim were the same.

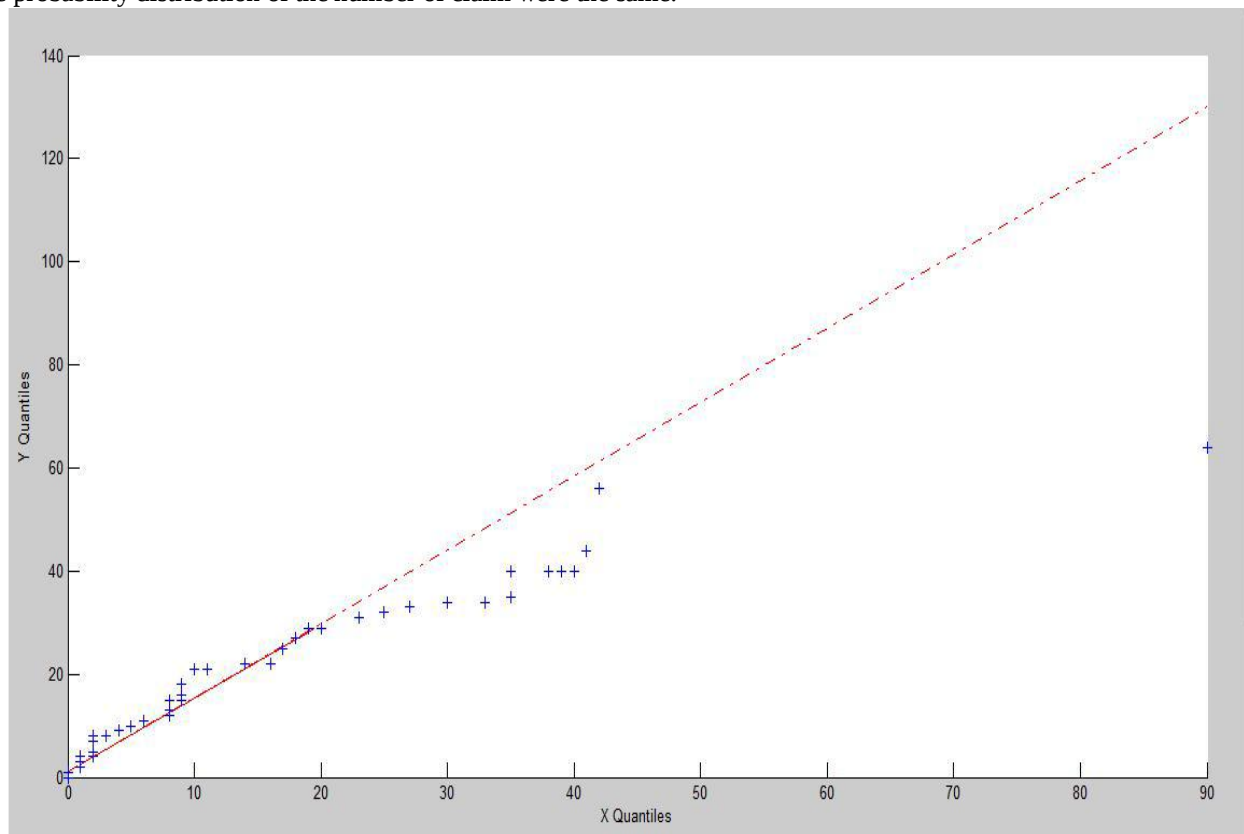


Figure 3: Quartile plot of Number of Claims on Funeral Policy from StarLife and MetLife.

Fitting Probability Distribution to Claims

The density estimates as displayed in Figure 4 for both portfolios are not significantly different. They both have a similar distribution but the StarLife distribution curve is steeper and slender than that of MetLife Insurance Company. Furthermore, it is visible that the sample distributions though unimodal, has several turning points which is not typical of the conventional probability distribution at a glance. However, ascribing this abnormality to the presence of outliers in the data set may warrant smoothening (fitting) to depict a true probability distribution. Empirical Distributions that can be fitted to the observed data included the Mixed Poisson Distributions starting with the Poisson distribution as discussed in the literature.

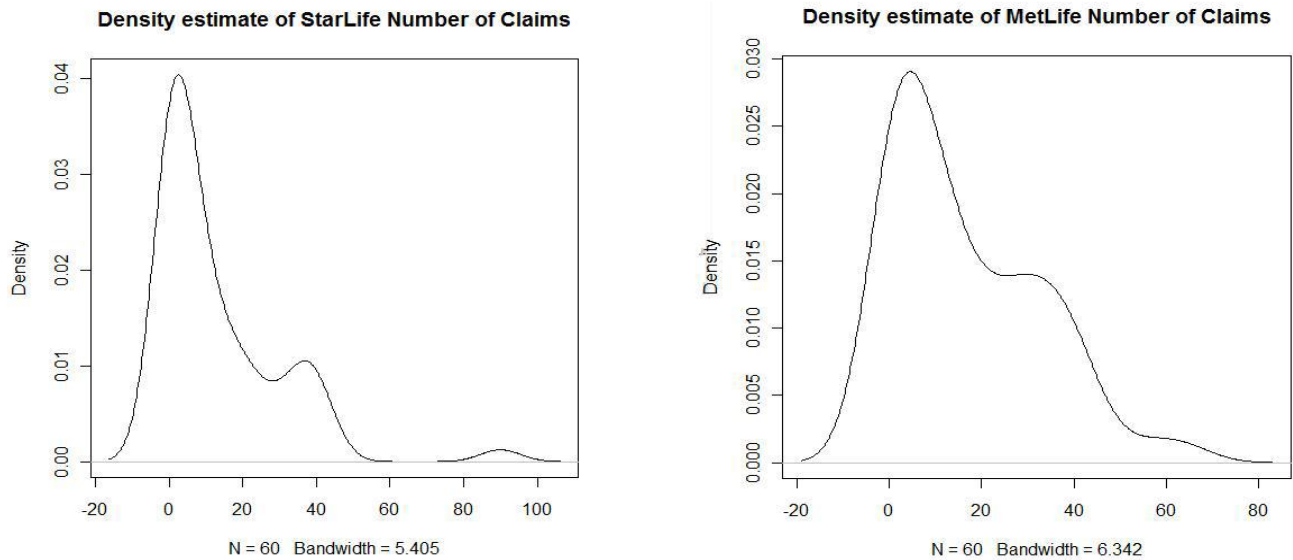


Figure 4: Density Estimate of Number of Claims on Funeral Policy from StarLife and MetLife.

Figures 5 and 6 were produced as a result of fitting the number of Claims for StarLife and MetLife respectively with the negative binomial and Poisson distribution. A critical look at the charts (Figures 5 and 6) reveals that the Poisson distribution does not fit well to the number of claims from both StarLife and MetLife Insurance Companies. However the Negative Binomial Distribution fits the Number of Claims reasonably well.

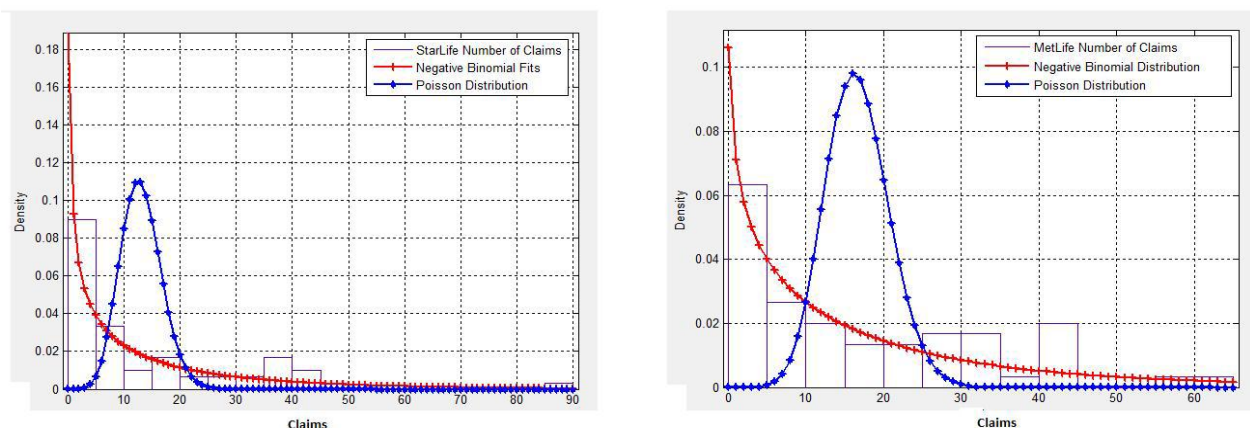


Figure 5: Negative Binomial and Poisson Distribution Fit to the Number of Claims on Funeral Policy from StarLife and MetLife.

The maximum likelihood estimates of the Poisson mean for StarLife and MetLife were $l_s = 13.05$ and $l_m = 16.06$ respectively. The 95% confidence interval for l_s and l_m were (12:1672; 13:9968) and (15:6489; 17:7152) and finally, the loglikelihood were 1228:3611 and 1810:5978. However, for the negative binomial distribution, the means were estimated to be $l_s = 13.0500$ and $l_s = 16.6500$ and the dispersion parameters to be $a_s = 2.0166$ and $a_m = 1.4323$ for the StarLife and MetLife number of claims.

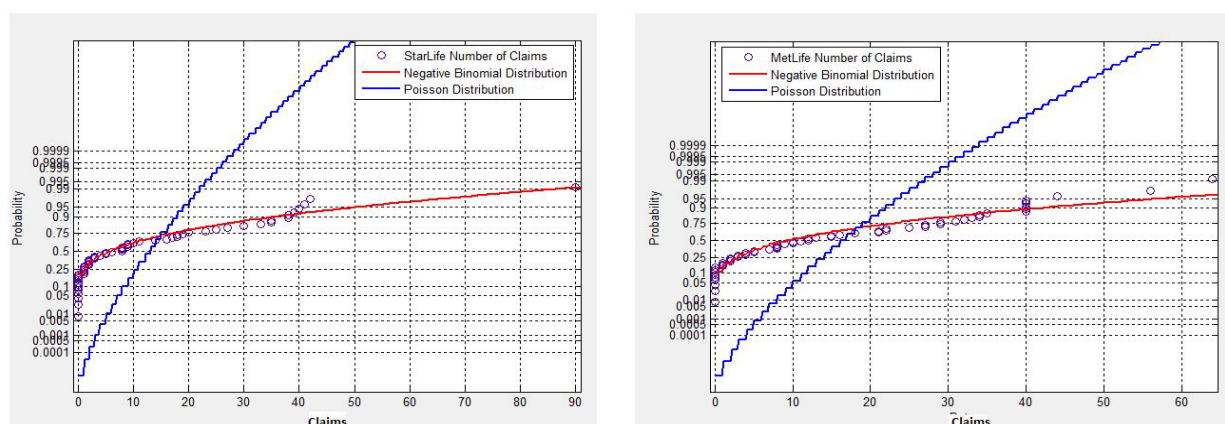


Figure 6: Percentile Plot of the Distribution of Number of Claims, Negative Binomial and Poisson distribution.

The variance of the random effect for both StarLife and MetLife were estimated to be $V_s(Q) = 0.49588$ and $V_m(Q) = 0.69818$ respectively. The 95% confidence interval for the dispersion parameters were (1.2633; 2.7699), (0.8912; 1.9734) and (0.3610; 0.7916), (0.5067; 1.1221) for their variance $V(Q)$.

The Log-likelihood of the Negative Binomial Distribution was 1667.5653 and 2175.5353 for StarLife and MetLife respectively which were far better than that of the Poisson distribution for both StarLife and MetLife respectively. Finally, the confidence interval estimates of the monthly expected number of claims estimated by the Negative Binomial Model were approximately (9; 19) and (12; 23) claims for StarLife and MetLife respectively (see Table 3).

Table 3: Interval Estimates of the Number of Claims from StarLife and MetLife

Company	Poisson	Estimate	Negative Binomial	Estimate
StarLife	12.1672	13.9968	9.0494	18.8196
MetLife	16.6500	16.7274	0.0774	16.5726

Table 4 provides summary statistics of the Goodness of fit of the negative binomial distribution for both StarLife and MetLife Insurance Companies. The scaled deviance of 69.3796 and 71.2286 for StarLife and MetLife compared to the asymptotic chi-square with 59 degrees of freedom yielded a p-value of about 0.15, which implies we cannot reject the null hypothesis that the specified negative binomial model is the correct model. Again, the dispersion parameter estimates were 2.0166 and 1.4323 for StarLife and MetLife respectively. The deductions are that the number of claims from StarLife and MetLife is not purely independent as required by the classical Poisson process.

The Number of Claims is duration and occurrence dependent because the composition of the number of portfolios are constantly varying (either increasing or decreasing) as a result of continuous sales of the policy.

Table 4: Criteria for assessing Goodness of Fit of the Negative Binomial Models

StarLife	Criteria	DF	Value	Value/DF
	Deviance	59	69.3796	1.1759
	Scaled Deviance	59	69.3796	1.1759
	Person Chi-Square	59	46.7985	0.7932
	Scaled Person Chi-Square	59	46.7985	0.7932
	Log Likelihood	1667.5653		
MetLife	Criteria	DF	Value	Value/DF
	Deviance	59	71.2286	1.2073
	Scaled Deviance	59	71.2286	1.2073
	Person Chi-Square	59	36.4253	0.6174
	Scaled Person Chi-Square	59	36.4253	0.6174
	Log Likelihood	2175.5353		

In effect the assumption of independence of the random process is being violated hence the inadequacy of the Poisson model. However, the Negative Binomial Distribution produced considerably good fits in the sense that it is the limiting form of the resulting distributions that arise in the situation of occurrence and duration dependence and is known as Pölya - Eggenberger distribution.

Again, the number of claim process is not homogeneous as required by the Poisson distribution. The occurrence of death which drives the number of claims varies with every policy holder as a result of unobservable social, moral, economic and health factors. Again, the Poisson distribution assumption of homogeneity is being violated and the Negative Binomial Distribution arises as the limiting distribution and for that matter is adequate.

Bootstrap Estimates

Table 5 shows the summary statistics of the bootstrap estimate of 100 resamples from the estimated probability distribution of the number of claims from StarLife and MetLife. A comparison between the expected monthly number of Claims $\hat{\theta}$ obtained by the probability models and that obtained from the bootstrap method $\hat{\theta}^c$ showed no significant variation as they were set at thirteen (13) and seventeen (17) claims for StarLife and MetLife insurance company respectively.

Table 5: Summary statistics of Bootstrap replicates of Number of Claims from StarLife and MetLife

Company	$\hat{\theta}$	$\hat{\theta}^*$	Bias	$\hat{\theta}^c$	$SE_B(\hat{\theta})$
StarLife	13.0500	12.9978	-0.0522	13.1022	1.9932
MetLife	16.6500	16.7274	0.0774	16.5726	1.7831

5. Conclusion and Recommendations

This research only focuses on choosing between the Poisson and the Negative Binomial distribution for fitting insurance claims and estimating the monthly expected number of claims. The Negative Binomial distribution appear to be superior to the Poisson distribution for fitting insurance claims and therefore, provides somewhat reliable estimates for planning, decision-making as well as estimation in insurance administration. The bootstrap estimates did not vary from the estimates obtained by the probability models.

We therefore recommend the negative binomial distribution over the Poisson distribution for fitting insurance claim. The bootstrap estimates should be obtained and compared with the estimates from the probability models to authenticate the estimates. Lastly, we recommend that further work should be conducted using other models including mixed Poisson probability models.

References

- Asmussen, S; Albrecher, H.(2010) Ruin Probabilities. In: Advanced series on statistical science & applied probability ; v. 14. Edition: 2nd ed. World Scientific Publishing Co.
- Asmussen, S., Kliippelberg, C, 2008. Large deviations results for sub-exponential tails, with applications to insurance risk. *Stochastic Process. Appl.* 64, no. 1, 103-125.
- Hall P. and Yao Q. (2003). "Inference in ARCH and GARCH Models with Heavy-Tailed Errors", *Econometrica* 71, pp. 285-317.
- Konstantinides, D., Tang, Q., Tsitsiashvili, G., 2002. Estimates for the ruin probability in the classical risk model with constant interest force in the presence of heavy tails. *Insurance Math. Econom.* 31, no. 3, 447-460.
- Mikosch, T., Nagaev, A.V., (2007). Large deviations of heavy-tailed sums with applications in insurance. *Extremes* 1, no. 1, 81-110.
- Miguel, J.A. and Olave, P. (1999a). "Forecast Intervals in ARCH Models: Bootstrap Versus Parametric Methods", *Applied Economics Letters* 6, pp. 323-327.
- Migue, J. A. and Olave, P. (1999b). "Bootstrapping Forecast Intervals in ARCH Models", *Test* 8, pp. 345-364.
- Lee Y. H. and Lee M. C. (2009). "Block Bootstrapping Prediction Intervals for Financial Time Series Models: Empirical Evidence from the Taiwan Stock Market", 2009 International Symposiums on Major Academic Disciplines, Taipei.
- Pascual, L., Romo, J., and Ruiz, E. (2006). "Bootstrap Prediction for Returns and Volatilities in GARCH Models", *Computational Statistics and Data Analysis* 50, pp. 2293-2312.
- Ridout, M.S., Hinde, J.P. and Demetrio, C.G.B. (2001) A score test for testing a zero-inflated Poisson regression model against zero-inflated negative binomial alternatives. *Biometrics*, 57, 219-223
- Schlegel, S., 1998. Ruin probabilities in perturbed risk models. The inter- play between insurance, finance and control (Aarhus, 1997). *Insurance Math. Econom.* 22, no. 1, 93-104.
- Tang, Q. and Tsitsiashvili, G., 2003. Precise estimates for the ruin probability in finite horizon in a discrete-time model with heavy-tailed insurance and financial risks. *Stochastic Process. Appl.* 108 (2003), no. 2, 299-325.
- Lambert, D. (1992) Zero-inflated Poisson regression, with an application to defects in manufacturing. *Technometrics*, 34, 1-14.
- Thombs, L. A. and Schucany, W. R. 1990. "Bootstrap Prediction Intervals for Autoregression", *Journal of the American Statistical Association* 85, pp. 486-492.
- Yue, F. (2003): Semi-parametric specification tests for discrete Probability models. *The Journal of Risk and Insurance*, , Vol. 70, No. 1, 73-84