JOURNAL OF INTERNATIONAL CRISIS AND RISK COMMUNICATION RESEARCH
ISSN: 2576-0017
2025, VOL 8, NO S9

A proposed actuarial model for estimating the risk premium for vehicle insurance using central moments theory of claims

Ahmed Mohammed Farahan[1], Raed Ali Alkhasawneh[2], Waheeb Hassan Yassin Gadour[3], Mostafa A. Radwan[4], M. Sh. Torky[5], Mohamed Ismail Abdulrahman Ismail[6], Alaa Fathi Soliman[7], Fatma Yousef Elshinawy[8], Samina Bashir[9], Khaled Alsaeed Qamar[10]

[1, 2, 3,4, 10] College of Applied Studies and Community Service, Imam Abdulrahman Bin Faisal University, P.O. Box 1982, Dammam 31441, Saudi Arabia.
[5, 7, 8, 9] College of Business Administration, Imam Abdulrahman Bin Faisal University, P.O. Box 1982, Dammam 31441, Saudi Arabia.
[6] Deanship of Graduate Studies, Imam Abdulrahman Bin Faisal University, P.O. Box 1982, Dammam 31441, Saudi Arabia.
**Author to whom correspondence should be addressed: amfmohamed@iau.edu.sa*

Abstract

This research aims to present an actuarial model for estimating the value at risk (VAR) for vehicle insurance, given the availability of a set of variables associated with determining this premium. These variables reflect both the driver's demographics and the vehicle model itself, and are based on an appropriate probability distribution for the number and value of vehicle insurance claims. Through the practical application of the proposed model, the article estimated the price limits for vehicle insurance rates at the company under study. It became clear that the rate used by the company deviates significantly from these limits, indicating a lack of consistency with the company's actual experience. The Gamma distribution is considered the appropriate distribution for claims values. The variables (age, vehicle age, policy term, and type) have a significant impact on claims values, while the remaining variables (vehicle value, education level, and marital status) have no significant impact on the study's response variable, which represents claims values. The results indicate that the total insurance rate at the company under study ranges between 7.22%, representing the minimum, and 7.48%, representing the maximum. Comparing this rate to the company's own rate of 9.87%, we find that it is higher than the company's experience. Therefore, it is necessary to reduce the insurance price to match its experience. The article recommended adopting the proposed pricing model, as practical application has proven that it reflects the actual experience of the company's data.

Keywords: Risk premium, Auto insurance, Actuarial pricing models, Saudi Insurance Market, Pricing, General insurance.

1. Introduction

Pricing is one of the most important technical functions of insurance. Pricing must be appropriate to each customer's risk level, sufficient to cover expected compensation, and achieve a profit margin that allows the company to continue its insurance business. The price of motor insurance refers to the expected value of future losses, based on past experience, which is often recent, so pricing is adequate and flexible, providing highly reliable results. Pricing in insurance relies on the prediction process, by arriving at an appropriate probability distribution that reflects the changes associated with this phenomenon, and thus relying on the proposed distribution in the pricing process. Determining the form of the probability distribution that governs the insured phenomenon is of great importance in the insurance field. Through the probability distribution, the value of the provisions required to address deviations between actual and expected claims values is estimated, as is the appropriate price for reinsurance. It also determines the probability of default, which reflects the possibility of claims exceeding premiums. Actuarial pricing principles are based on two basic approaches to estimating the risk premium. The first is based on estimating the average risk value, which is estimated based on the ratio of deductible claims to earned premiums. The second approach relies on estimating the net premium, which is estimated using the ratio of deductible claims to the number of risk units. The pricing process for insurance products must be characterized by adequacy, reasonableness, and fairness. The price should be commensurate with the degree of risk, reflecting the principle of fairness, and also achieving an appropriate profit margin for the insurer. Adequacy also indicates that the premium is sufficient to cover the costs incurred in concluding the policy (Atallah, 2013). The insurance market in the Kingdom of Saudi Arabia consists of thirty-three companies competing to attract customers, which may affect pricing. A company may resort to offering discounts that may affect the fair price of the risk, although competition must remain at a minimum in insurance pricing. What exacerbates the problem in this insurance market is that the pricing process relies on a fixed tariff for all vehicle types, regardless of the degree of risk. This also ignores the impact of other factors on pricing, which are assumed to have a significant impact on the degree of risk, such as the driver's demographic characteristics, loss experience, and vehicle characteristics. A fair price must reflect the expected costs of compensation incurred as a result of the insured risks, which leads to costs varying depending on the level of risk. Therefore, the research problem can be formulated as "the Saudi insurance market relies on a fixed tariff for pricing vehicle risks, regardless of the degree of risk, and does not take into account the factors affecting the degree of risk when setting the appropriate price." The overall objective of the research is "to build a model for estimating the risk premium in vehicle insurance." This is achieved by proposing an actuarial model for pricing vehicle insurance policies based on a set of elements assumed to affect the degree of risk, which reflect factors related to the driver or the vehicle model itself. This is intended to address the price competition that has prevailed in the insurance market in the Kingdom. Providing a quantitative model to estimate the appropriate risk premium for vehicle insurance, characterized by a sufficient and fair balance between the level of risk and the premiums paid, will achieve the principle of fairness and sufficiency, which will have a positive impact on all parties involved in the insurance process. The insurer needs to cover its obligations and achieve an appropriate profit margin, while the client will receive insurance coverage commensurate with the premiums paid. Relying on generalized linear models also helps identify more factors and variables that affect the degree of risk in vehicle insurance, contributing to a fair insurance price. This provides a scientific tool for insurance companies to classify risk in a more equitable manner. The research also presents a quantitative model that enables insurance companies in the Kingdom of Saudi Arabia to price their vehicle insurance based on a diverse set of factors, thus eliminating the scope of scientific and practical criticism, which often points to actuarial errors in this vast insurance market, which relies on a fixed tariff for vehicle insurance, regardless of the degree of risk. The remainder of this paper is organized as follows. Section 2 Literature Review, Section 3 theoretical Framework of the Proposed Actuarial Model, Section 4 practical application of the proposed model for estimating the net premium for vehicle insurance, Section 5 conclusions.

2. Literature Review

(Pandya & Shukla, 2023) proposed a new model that helps both customers obtain more realistic and compelling premiums and helps the sector better assess vehicle risk. In this proposed model, vehicle user location data is tracked, stored, and analyzed using artificial intelligence algorithms to extract a user profile, which also helps determine their reputation and risk profile in the system. He proposed a mobile application that helps calculate vehicle usage using GPS trackers installed on them, which in turn is used to generate vehicle-specific risk and usage factors. Stakeholders will use these factors to calculate the next insurance premium.

(Ieosanurak & Moumeesri,2023) presented new claims models that improve the accuracy of claims data fit, in terms of frequency and severity. Specifically, he proposed the Poisson-Gall distribution for claims frequency and the exponential Ga-Gall distribution for claims severity. He also proposed a new model for calculating claims/indemnity premiums. The results indicated that the proposed insurance premium model effectively solves the overpricing problem. Furthermore, the proposed model produces insurance premiums that are more closely aligned with policyholders' claims histories, benefiting both policyholders and insurers.

(Azaare & Ampaw, 2022) investigated the type of loss distribution function that best approximates policyholder claims in Ghana. They applied the Kullback-Leibler variance, Kolmogorov-Smirnov statistical tests, Anderson-Darling test, and maximum likelihood estimation (MLE) to estimate policyholder claims. The results indicate that motor policyholder claims are best approximated using a lognormal probability distribution. With a lognormal distribution, the industry can adequately assess policyholder claims to minimize potential losses.

(Yu, W & Cui,2021) used a genetic algorithm to optimize the BP neural network structure, which improves calculation speed. A total claim amount prediction model was then constructed. The results showed that the prediction accuracy of the BP neural network model for both Shandong Province and six cities exceeded 95%. The predicted total claim amount was then used to calculate the insurance premiums for five cities in Shandong Province based on the credibility theory. The results showed that the average premiums for the five cities were slightly higher than the actual claim amount. The combination of the BP neural network and credibility theory can accurately estimate the claim amount and pricing of auto insurance, effectively improving the current situation of the auto insurance industry and promoting the development of the insurance sector.

(Suwandani & Purwono, 2021) calculated the loss reserve by applying Gaussian process regression to estimate future claims. The modeling is performed on motor vehicle insurance data. This study uses the chained ladder method, which is the most common loss reserve method in theory and practice. The estimation results demonstrate that the Gaussian regression method is highly flexible and can be applied without significant modification. Motor insurance data has a short development period (when claims occur), so it falls under the short-term business category.

(Zibusiso, 2018) applied the GLM method to estimate the net premium for vehicle insurance. The Poisson distribution was used to fit the probability distribution curve for the number of claims, while the Gamma distribution was used to fit the claims values. The significance of a variety of variables expected to have an impact on insurance prices was tested, including (age, average daily mileage, vehicle value, policy term, vehicle age, number of claims, value of claims, marital status, gender, purpose of vehicle use (personal or commercial), geographic location, and level of education). The study concluded that all proposed variables had a significant impact on the probability distribution of the number of claims, except for (driver age and vehicle age). The researcher proved that all variables except average daily mileage were insignificant when fitting the probability distribution of claims value. Using the expected value of the different moments for each of the two distributions, the researcher arrived at the net price of vehicle insurance in Zimbabwe.

(Shahrazad, 2015) presented a vehicle accident pricing model for the National Insurance Company of Algeria. The Poisson distribution was used to fit the probability distribution of the number of claims, while the gamma distribution was used to fit the value of claims. The study yielded a set of results, perhaps the most important of which is that the pricing system in Algeria is governed by the gamma model for the distribution of loss amounts and the Poisson model for the distribution of the number of accidents. Based on these models, the researcher found that vehicle insurance pricing is based on a number of factors related to the driver (driver's age, gender, and driver-insurer compatibility), and factors related to the vehicle (vehicle age, use, and power, in addition to the bonus and penalty factor and the selected warranty). From this, it can be argued that the pricing system in Algeria is non-parametric, and that there are other variables on which the pricing system depends.

Through the previous review of previous studies, we have developed a vision for the set of variables proposed in the quantitative model for vehicle insurance pricing, which are assumed to have a significant impact on the degree of risk. Most previous studies indicated that these variables are a set of variables related to the driver and the vehicle model itself, such as (the driver's age - the value of the vehicle - the age of the vehicle - the duration of the policy - the marital status of the driver - the level of education - the gender). Previous studies also presented the mechanisms for constructing the probability distribution for both the value and number of claims and the most popular probability distributions used in the field of vehicle insurance. This enables the development of an integrated quantitative model for pricing risk in vehicle insurance, applied to the company under study.

3. Theoretical Framework of the Proposed Actuarial Model

3.1 Actuarial Models Addressing Pricing in General Insurance

Actuarial models used to price risk in general insurance can be divided into three models. The first of these models assumes that they deal with a closed insurance portfolio comprising a set of policies. It also assumes that the claims to which each insurance policy is exposed represent an independent random variable. This type of model is called "individual risk model." Models that deal with the total value of claims for the insurance portfolio, rather than for each policy individually, are called "aggregate risk model." They rely on two variables: the number and value of claims to which the insurance portfolio is exposed. Another type of model reflects the impact of various factors and their influence on the degree of risk according to the risk-influencing factors. These models include a variety of sub-models, such as flexible models, additive models, and multiplicative models.

3.2 Pricing characteristics in vehicle insurance

Pricing in vehicle insurance relies on several approaches, perhaps the most important of which is "pre-pricing," which relies on the vehicle's basic characteristics (power, value, usage, age, model), as well as the driver's characteristics (gender, age, place of residence). The second, equally common approach used in vehicle insurance risk pricing is "post-pricing," which relies on the insured's accident history. This is known as the discount, bonus, or incentive and penalty system. The premium depends to a large extent on the degree of risk, as the insurance amount is adjusted according to the driver's experience. The premium is reduced when no accidents are recorded, while the penalty is an increased premium in the event of an accident. There are a number of factors that influence vehicle insurance pricing, which are divided into technical factors that reflect the vehicle's demographics, such as the model, power, and age of the vehicle; and other human factors related to the insured, which reflect the type, age, and habits of the driver. Other factors unrelated to the vehicle or driver include the type of vehicle use, weather conditions, traffic and road conditions, and the geographic and residential nature of the vehicle's travel area (Salam, 2015).

There is also another classification of the risk pricing system in vehicle insurance, based on the record of losses incurred and the characteristics of the insured risk. This includes the individual pricing method, which relies on setting a specific price for each risk separately. This method ignores the existence of specific pricing approved by insurance companies. This approach is criticized for its difficulty in implementation, especially as the number of units exposed to risk increases. Risk can also be divided into categories or classes based on common characteristics for each category, such as the vehicle model or the characteristics of the vehicle owner. This is known as the stratified pricing method. This approach ignores the degree of homogeneity among the categories used to group similar risks. The third method relies on setting a specific price for the risk, which is then adjusted based on the insured's accident experience, whether by adding or subtracting. This pricing method is called the adjusted pricing method, and it has many advantages. Not only does it rely on the insured's past experience with pricing, but it also aims to encourage insureds to limit losses, or not report them if they can afford them.

3.3 Pricing Determinants in Vehicle Insurance

The vehicle insurance pricing system relies on several determinants, perhaps the most important of which is the existence of a pricing system that reflects changes associated with risk, and the updates and changes associated with them. These are factors that affect insurance prices, such as inflation rates and road accident rates. The pricing system must also be based on a long-term period that reflects loss experience, which supports a pricing system characterized by relative stability and consistency. There are several methods that can be used to achieve this goal, perhaps the most important of which is setting maximum limits for price increases and compensation. In addition, the vehicle insurance pricing system must rely on a margin of safety to reflect deviations in results resulting from inaccurate forecasting. The pricing system must also be applicable, meaning it is a realistic system that can be implemented by insurance companies and includes all the variables expected to affect the price of risk.

3.4 Evolution of Vehicle Insurance Pricing Models

In the past, vehicle insurance pricing relied on a statistical model that could adjust the price to suit any change, with a certain degree of confidence, to match the loss experience of any branch during any year. The insurance premium, P, depends on the minimum confidence period. This model can be represented by the

following equation:

$$P = \mu + Z_{1-\propto}\frac{\sigma}{\sqrt{n}}$$

That is, the premium equals the average of actual losses plus the allowances added to cover deviations between the actual and expected price. This represents the standard deviation of the loss distribution, while μ represents the initial risk premium. P represents the final net risk premium after adding the allowance for deviations between actual and expected losses to the initial net risk premium. Therefore, the final net risk premium equals the initial risk premium plus the allowances for deviations in losses. When estimating the value of the initial risk premium, it is preferable to rely on the number of years of experience and also to weight the results by the individual years of experience. This model is based on several assumptions, which are:

- The values of individual claims are independent and symmetrically distributed with mean $\overline{x}$ and variance $\frac{s^2}{\sqrt{n}}$.
- The values of the variable x, which reflect the individual claims, represent a random sample.
- The population is homogeneous.

On the other hand, Coutts pointed to the development of a quantitative pricing model based on a points system, which relies on data from UK insurance companies. He explained that it has many shortcomings, the most significant of which is its failure to take into account the personal characteristics of the risk holder. He advocated for pricing based on a careful analysis of the characteristics of all the different parties involved in the risk, including the risk holder and the risk-bearing entity. The statistician Tweedie presented a pricing model called the Tweedie Model, named after him. It is considered one of the best models for dealing with premiums, given that most of the data concentrations in the theoretical distribution are centered in the middle of the distribution and tend to be slightly skewed to the right, which fits the shape of the probability distribution of claims, which has roughly the same shape.

McCullagh and Nedler proposed a pricing approach using generalized linear models, which has become one of the most important methods used in pricing motor insurance policies. It represents one of the most important models with a high level of reliability, as it takes all pricing factors into account, enabling it to handle a large number of risk groups and their relationship to the size and experience of insurance companies' claims (Huang and Query, 2007).

In 1997, Nelder and Verrall incorporated credibility theory functions into generalized linear regression models. In 2004, Vein Schmitter developed a simple method for estimating the number of expected claims required to calculate the tariff derived from generalized linear regression models. Generalized regression models consist of two types: the first is additive models, which rely on the addition of covariates. The other model is the multiplier model. It is worth noting the shortcomings of additive models, as they can produce false results when a sample with sufficient values is not used. This may result in negative premium or claim values. For example, this type of model provides a fixed discount value, regardless of the premium value, while the multiplier model provides a fixed percentage, which is logical as it is proportional to the value of the premiums paid.

Silva and Afonso (2015) pointed to another group of models used in pricing this type of insurance, such as models that rely on estimating the net premium based on past experience of aggregate claims, as well as estimating the net premium based on expected aggregate claims, classical linear models, and generalized linear models. These models are based on a comparison between the increase in insurance prices and the amount of dispersion in net insurance premiums. The primary advantage of pricing based on net premiums is that it involves less dispersion, as it includes all insured data during the time period under study. Huang and Query (2007) also presented pricing models based on a combination of the Max Model and the GLM to improve the model's power and accuracy when the data is highly correlated with various risk factors.

The Max Model aims to resolve problems related to the common and complex correlations between the variables and factors involved in pricing. Another group of models developed from generalized linear models, perhaps the most important of which are (Goldburd, et al. (2016)), generalized linear mixed models (GLMMs), generalized linear dispersion models (DGLMs), generalized ensemble linear models (GAMs), multivariate regression models (MARS), and generalized linear models with elastic networks. By reviewing the evolution of vehicle insurance pricing models, it becomes clear that the practical application of these models still requires more attention and application, rather than relying on traditional methods, which are plagued by shortcomings. Therefore, insurance companies still have much to gain by shifting to applying statistical models in pricing and reserve estimation, compared to the deterministic methods they currently use.

4. Practical application of the proposed model for estimating the net premium for vehicle insurance

4.1 Introduction

The first stage of the proposed model involves preparing the data needed to estimate the net premium for vehicle insurance. This involves collecting data from both policies and claims. In most models, the pricing plan is based on the risk level derived from premiums paid, policy demographics, and claims data. The problem associated with this stage is the different timeframes for premiums and claims, as well as the fact that they are not stored in the same location. The timeline for each policy must be tracked separately to link all claims for the same policy for which premiums have been collected. Furthermore, data for each policy must be obtained and processed at the same point in time to avoid the impact of interest rates on the value of funds paid or collected. The researchers used the policy number to link the premiums and claims for each policy separately. The second stage focuses on modifying and refining the data, eliminating errors and outliers, ensuring the absence of duplications and illogical values, such as negative values in data not expected to contain negative values such as ages, premiums, or paid claims. This also ensures the absence of missing values. The third stage of the proposed model relies on attempting to fit an appropriate probability distribution for both the number and value of claims. The primary objective of this research is to arrive at the net premium for each insurance policy in the portfolio. This is achieved more fairly when dealing with two separate probability models for each of the number and value of claims separately. Therefore, the researchers relied on the Poisson distribution as a proposed probability distribution for the number of claims, as it represents one of the common distributions for the number of claims in the field of general insurance (David, 2015). The Gamma distribution was also used to model claim values at each level of exposure to risk. The Gamma distribution is considered a right-skewed distribution with a pointed peak and a long tail to the right. These characteristics make this distribution the most suitable distribution for claims values (zibusiso, 2018).

4.2 Net Premium Estimation

There are two methods for estimating the net premium. One relies on modeling the probability distribution of both the number of claims and the value of claims, then combining them. The other method uses the compound Poisson distribution to directly estimate the net premium. The researchers relied on the first method, which is based on estimating an independent probability distribution for both the number and value of claims, to capture all the parameters affecting the car pricing model. The factors affecting each are different (Denuit, 2004). The net premium reflects the expected value of claims, the various deductibles, and the expenses associated with settling claims. This value is converted to represent the actual pricing for each level of risk exposure in car insurance. The following equation illustrates the expected value of the net premium, which results from a probability distribution that represents the combination of two probability distributions. The first distribution relates to the expected value of the frequency of claims, while the second distribution relates to the value of claims or the severity of losses. Combining both distributions gives us the probability distribution of the expected values of the severity of claims at each level of risk:

$$Pure\ Premium = E[Claim\ Frequency]\ \times E[Claim\ Costs].$$

4.3 Data & statistical treatment

The database under study includes 46,354 observations. The data was collected from the records of a Saudi insurance company, covering the period from January 1, 2020 to December 31, 2024. For each observation, a set of data is available about the policy and the driver's demographic characteristics, as well as information about the vehicle and the geographic area. The proposed model relies on a set of explanatory variables, including the driver's age, vehicle value, vehicle age, policy duration, marital status, education level, and gender. Two variables representing the response (dependent variables) are also included: the number and value of claims. Claim frequency reflects whether the policyholder was involved in an accident during the study period, while claim value relates to the claim amounts paid to settle actual accidents. The number of claims (frequency) in the vehicle insurance portfolio of the Cooperative Insurance Company during the study period represents the first dependent variable included in the proposed model, as reflected in the data in the following table.

Table.1 Frequency distribution of claims values for the vehicle insurance branch*

Group	1 (0, 10000]	2 (10000, 20000]	3 (20000, 30000]	4 (30000, 40000]	5 (40000, 50000]	6 (50000, 60000]	7 (60000, 70000]	8 (70000, 80000]
Freq	29720	8488	3008	2424	1260	922	424	108

* Claims records of the company under study (vehicle branch).

By reviewing previous studies, a set of variables was considered, which are assumed to have an impact on the frequency and value of claims. These variables will then be subjected to study and analysis to determine the significance of these factors, which include (driver's age, vehicle value, vehicle age, policy duration, marital status, education level, and gender). The following table provides a statistical description of the proposed explanatory variables.

Table.2 Statistical Description of the Proposed Explanatory Variables

	Age	**Car value**	**Car age**	**Policy Length**	**Marital status**	**Education Level**	**Gender**
Mean	34	75655.751	6.45	2.68	2.34	2.82	1.17
Median	35	85008	7	2	2	3	1
Mode	28	13054	7	1	2	3	1
Standard Deviation	10.68	7706.882	3.22	1.062	0.657	0.725	0.542
Minimum	19	6286	0	1	1	1	1
Maximum	74	326828	30	6	3	4	2

The average age of the vehicles insured in the study sample was 6.45 years. The fact that the average value is close to the minimum can be explained by the company's keenness to attract newer vehicles. This may be due to several reasons, perhaps the most important of which is the increased safety factors in newer vehicles, in addition to the increasing tendency of customers to purchase newer vehicles, which offer greater capabilities and comfort factors than older models. The average age of the policies in the sample was 2.68 years, reflecting the company's high customer turnover rate. This may be due to the availability of better insurance offers than competing companies, as a result of the large number of insurance companies in the Saudi market, coupled with the lack of a fixed insurance tariff. The percentage of married individuals in the sample was approximately 78%. The percentage of those with a diploma or university degree was 72%, the largest segment of the sample. The percentage of males in the sample was 81%.

Table.3 Testing the validity of the statistical data for the variables of the proposed model

		Multicollinearity		Jarque-Bera Test	
		Tolerance	VIF	J-B	Prob.
1	Age $x1$	0.978	2.355	1.441	0.089
2	Car value $x2$	0.909	2.763	1.607	0.149
3	Car age $x3$	0.791	2.226	1.236	0.120
4	Policy length $x4$	0.788	1.748	1.120	0.114
5	Marital status $x5$	0.821	2.546	1.011	0.131
6	Education level $x6$	0.824	2.026	1.485	0.109
7	Gender $x7$	0.816	1.703	0.860	0.150
8	Claim freq $y1$	0.734	2.197	1.449	0.119
9	Claim sev $y2$	0.914	2.471	1.390	0.140
Autocorrelation					1.748
Heteroskedasticity (white test)					0.0168

The Durbin Watson Test was used to test the validity of the independent variables and their association with the dependent variable. Table 3 shows that the test result is 1.748, which falls within the appropriate range of $1.5 - 2.5$, indicating that there is no problem in testing the autocorrelation of the proposed model variables. The results of the non-stationarity test of the random error variance of the proposed model indicate that it reached 1.68%, which indicates that the standard error variance is stable. The proposed study models are valid for estimating the values of the independent and dependent variables. The parametric Jarque-Bera Test was used, and it was found that all model variables follow a normal distribution, as the P-value is greater than 5%. Collinearity was examined using Collinearity Diagnostics to calculate the tolerance coefficient for each of the independent variables to obtain the Variance Inflation Factor (VIF). If the VIF does not exceed five, this indicates the robustness of the study models in explaining the effect on the dependent variable. It is clear that all variables in the proposed models are less than five, indicating that the models do not suffer from collinearity problems. The presence of a moderate correlation between the variables included in generalized linear regression models is one of the most important requirements for applying this type of model (Goldburd et al., 2016). This type of model has the ability to determine the individual impact of each variable separately while excluding the impact of any other variables on the model. The following table shows the values of the correlation coefficients between the variables included in the proposed model. We find that the correlation coefficients are moderate, and no value falls outside the range $[-0.5, 0.5]$. These results provide strong evidence for the feasibility of using generalized linear regression models (Goldburd et al, 2016).

Table.4 Correlation matrix between the variables of the proposed model

	Age $\boldsymbol{x1}$	Car value $\boldsymbol{x2}$	Car age $\boldsymbol{x3}$	Policy length $\boldsymbol{x4}$	Marital status $\boldsymbol{x5}$	Education level $\boldsymbol{x6}$	Gender $\boldsymbol{x7}$	Claim freq $\boldsymbol{y1}$
Car value $\boldsymbol{x2}$	-0.013							
Car age $\boldsymbol{x3}$	-0.015	0.002						
Policy length $\boldsymbol{x4}$	-0.079	-0.019	0.448					
Marital status $\boldsymbol{x5}$	-0.028	0.003	-0.258	-0.455				
Education level $\boldsymbol{x6}$	0.009	-0.025	0.349	0.343	-0.459			
Gender $\boldsymbol{x7}$	-0.030	-0.013	0.397	0.394	-0.176	0.317		
Claim freq $\boldsymbol{y1}$	-0.048	-0.020	0.463	0.248	-0.403	0.479	0.309	
Claim sev $\boldsymbol{y2}$	-0.038	-0.025	0.455	0.346	-0.311	0.458	0.497	0.472

4.4 Proposed probability distribution of the number of claims

The Poisson distribution was used as a proposed distribution to estimate the expected number of claims. Using the statistical program (RStudio software package), the following results were obtained based on the Poisson distribution as a proposed probability distribution for the number of claims:

Table.5 Results of the generalized linear regression analysis model using the Poisson distribution as a proposed probability distribution for the number of claims

Coefficients:	Estimate	Std. Error	$\boldsymbol{z}$ value	$\boldsymbol{Pr}(>\lvert \boldsymbol{z}\rvert)$
(Intercept)	18.5400	120.365	0.184	0.919
Age $\boldsymbol{x1}$	0.0167	0.002	12.301	0.000
Car value $\boldsymbol{x2}$	0.0123	0.000	0.001	0.988
Car age $\boldsymbol{x3}$	0.0224	0.002	12.365	0.000
Policy length $\boldsymbol{x4}$	1.1043	0.021	75.968	0.000
Marital status $\boldsymbol{x5}$	-25.7377	115.200	-0.143	0.909
Education level $\boldsymbol{x6}$	-0.0436	0.032	-2.355	0.013
Gender $\boldsymbol{x7}$	-0.6989	0.021	-35.326	0.000
AIC		42658		
Number of Fisher Scoring iterations		22		
Dispersion parameter for poisson family taken to be $\mathbf{1}$				

The previous table reflects the coefficient values of the generalized linear model used to model the number of claims, based on the Poisson distribution as a linking equation. The AIC test statistic value is 42658. Examining the significance of the model coefficients reveals that the P-value of each of the variables (age, vehicle age, policy duration, education level, and gender) has a significant effect on the dependent variable (number of claims), while the variables (vehicle value and marital status) did not have a significant effect on the response variable. The over-dispersed Poisson distribution was proposed, and this model is being applied in an attempt to overcome the high dispersion of the previous model. The over-dispersed Poisson distribution differs from the Poisson distribution in that the variance does not equal the mean, but is a percentage of it. When estimating the probability distribution for the number of claims (C), it is assumed that they are distributed independently, taking the form of an over-dispersed Poisson distribution, with an expected value and variance as follows:

$$E[c_{ij}] = m_{ij} = x_i y_j \quad , \quad Var[c_{ij}] = m_{ij} = \emptyset x_i y_j \quad \text{[Where} \quad \textstyle\sum_{j=1}^{n} y_j = 1],$$

where x is the expected value of claims, y represents the ratio of claims to the total claims volume in the model, and represents the weight of claims. $\emptyset$ represents the distribution parameter, which represents an unknown value derived from the data under study. This model is considered a non-linear model with respect to the model's parameters, and therefore the function expressing the model's probability distribution must be reformulated to reflect the fact that the mean takes a linear form. This is done by relying on the logarithm equation to transform the expected value function of the distribution into a linear function, which takes the following form:

$$log(m_{ij}) = c + \alpha_i + \beta_j.$$

The previous function reflects the natural logarithm of the expected value of the probability distribution. It takes a linear form, reflecting the presence of a parameter for each variable and each risk unit separately. This is a form of the generalized linear model (GLM). Below are the outputs of the proposed model using the over-dispersed Poisson distribution for the number of claims (Amolo, 2011).

Table.6 Results of the generalized linear regression analysis model using the over-dispersed Poisson distribution as the proposed probability distribution for the number of claims

Coefficients:	Estimate	Std. Error	$\boldsymbol{t}$ value	$\boldsymbol{Pr}(>\lvert\boldsymbol{z}\rvert)$
(Intercept)	18.6511	58.213	0.326	0.793
Age $\boldsymbol{x1}$	0.1278	0.000	19.655	0.000
Car value $\boldsymbol{x2}$	0.1234	0.000	0.012	0.846
Car age $\boldsymbol{x3}$	0.1335	0.002	17.365	0.000
Policy length $\boldsymbol{x4}$	1.2154	0.005	181.266	0.000
Marital status $\boldsymbol{x5}$	-25.6266	52.481	-0.366	0.813
Education level $\boldsymbol{x6}$	0.0675	0.001	-11.658	0.000
Gender $\boldsymbol{x7}$	-0.5878	0.007	-65.330	0.000
AIC		24329		
Number of Fisher Scoring iterations		22		
Dispersion parameter for quasipoisson family taken to be $\mathbf{0.213586}$				

Table 6 summarizes the results of the Over-dispersed Poisson model. The low AIC value, which reached 24,329, provides significant evidence of the model's good fit. Using the RStudio statistical program, the values of the first four moments for the number of claims were determined based on the proposed model. The One-Sample Kolmogorov-Smirnov test, a non-parametric test that does not require data to follow a normal distribution, was used to test the goodness of fit of the data under study to the proposed Over-dispersed Poisson distribution. The test statistic value was 0.146, and therefore, we cannot reject the null hypothesis of the test, which states that the proposed model fits the Over-dispersed Poisson distribution well.

Table.7 Frequency Distribution and the First Four Moments for the Number of Claims

Number of claims	0		1	2	3	4	Sum
Number of cars	21174		11482	7612	5668	418	46354
Total number of claims	0		11482	15224	17004	1672	45382

	Moment	1st moment	2nd Moment	3rd moment	4th moment
	Individual Moment	1.4685	3.2235	8.16	22.632
	Moment for Total N of claims	34036.5	74722.5	189124.5	524542.5
	One-Sample Kolmogorov-Smirnov test	$D = 0.03548$		$p-value = 0.146$	

4.5 Proposed Probability Distribution of Claims Values

Claims values represent the product of a set of variables, the most important of which are compensation paid, claims settlement expenses, provision for outstanding compensation, and recoveries. Claims values are adjusted by the inverse of the price index to eliminate the effect of inflation. Pearson curves are one of the most important methods used to model claims values and estimate the values of the various moments, which represent the parameters of the proposed distribution. These curves are a set of curves called Pearson family curves, and the fit of these curves depends on the values of the skewness and kurtosis coefficients of the data. Pearson distributions are divided into two groups. The first includes the first, fourth, and sixth curves. The use of each of these types depends on the value of a specific coefficient calculated from the skewness and kurtosis coefficients. If the coefficient is less than zero—meaning the origin lies at the mode value—then the first curve is used. The fourth curve is used when the value of this coefficient lies between zero and one. If the coefficient value is greater than one, the sixth Pearson curve is used. The second group of Pearson curves includes ten curves and is called the transition curve group. They are typically used when the skewness coefficient value is equal to zero (Lahcene, 2013). Due to the many difficulties encountered when working with Pearson distribution functions, scientists have developed tabular methods to facilitate the use of these distributions, given the measure of skewness and kurtosis. Pearson distributions provide more accurate results. They also rely on the values of the first four moments of the probability distribution of claims values, providing sufficient information derived from the insurer's experience. They can also be used in cases where the number of years of experience is low (O. Podladchikova, 2003). Choosing an appropriate probability distribution for fitting claims values depends on the shape of the raw curve of the loss values, which is organized as shown in Table 8. Based on the values in the table, the values of the first four moments were estimated, which will be used to fit an appropriate probability distribution. The following table shows the values of the various moments of the frequency distribution of claims values for the company under study.

Table.8 The first four moments of claims values

	1st moment	**2nd Moment**	**3rd moment**	**4th moment**
Moment	$1.87E + 04$	5.02E+08	2.01E+13	9.89E+17

The gamma distribution is a common distribution for modeling claims values. Based on the same observations for the company under study, but after excluding observations with zero claims values, the data were subjected to the proposed distribution. Table 9 reflects the coefficient values of the generalized linear model used to model claims values using the gamma distribution as a linking equation. The AIC test statistic value is 65879. Examining the model's significance coefficients and the P-value of the test statistic shows that each of the variables (age, vehicle age, policy duration, and gender) has a significant effect on claims values, while each of the variables (vehicle value, education level, and marital status) had no significant effect on the response variable.

Table.9 Results of the generalized linear regression analysis model using the gamma distribution as a proposed probability distribution for claims values

Coefficients:	Estimate	Std. Error	$\boldsymbol{t}$ value	$\Pr(> \lvert \boldsymbol{z} \rvert)$
(Intercept)	7.9220	0.076	103.639	0.000
Age $\boldsymbol{x1}$	0.0021	0.000	5.802	0.000
Car value $\boldsymbol{x2}$	0.0000	0.000	0.430	0.667
Car age $\boldsymbol{x3}$	0.0047	0.001	3.858	0.000
Policy length $\boldsymbol{x4}$	0.6294	0.012	53.037	0.000
Marital status $\boldsymbol{x5}$	NA	NA	NA	NA
Education level $\boldsymbol{x6}$	-0.0012	0.027	-0.044	0.965
Gender $\boldsymbol{x7}$	-0.0799	0.023	-3.515	0.000
AIC		65879		
Number of Fisher Scoring iterations		7		
Dispersion parameter for Gamma family taken to be $\mathbf{0.4216591}$				

4.6 Determining the Net Insurance Price

The risk premium for motor insurance is determined based on estimating the moments of the total claims value distribution. This distribution represents a composite of two probability distributions: a discrete distribution, representing the number of claims, and a continuous distribution, reflecting the value of claims. The four moments of the composite distribution are estimated based on the moments of each of the two distributions. The Hon Shiang Lau method will be used to calculate the first four central moments of the total claims value, which are based on the following equations:

$$for \quad \partial = \sum_{i=1}^{n} x_i$$
$$mean = \bar{z}_{\partial} = \bar{x}\ \bar{y}$$
$$\bar{z}_2(\partial) = \bar{y}^2\ \bar{x}_2 + \bar{x}\ \bar{y}_{\ 2}$$
$$\bar{z}_3(\partial) = \bar{y}^3\ \bar{x}_3 + \bar{x}\ \bar{y}_3 + 3\,\bar{y}\ \bar{y}_3\,\bar{x}_2$$
$$\bar{z}_4(\partial) = \bar{y}^4\ \bar{x}_4 + \bar{x}\ \bar{y}_4 + 4\bar{y}\ \ \bar{y}_3\,\bar{x}_2 + 6\bar{y}^2\ \bar{y}_2\,[\bar{x}\,\bar{x}_2 + \ \bar{x}_3] + 3[\,\bar{y}_3]^2[\bar{x}^2 - \bar{x} + \bar{x}_2],$$

where $(\bar{x}, \bar{x}_2\,, \bar{x}_3\,, \bar{x}_4)$ represent the first four moments of the number of claims, while $(\bar{y}_1, \bar{y}_2\,, \bar{y}_3\,, \bar{y}_4)$ represent the first four moments of the claims values, and $(\bar{z}_{\partial}\,, \bar{z}_2(\partial), \bar{z}_3(\partial), \bar{z}_4(\partial)\,)$ represent the estimated values of the moments of the distribution of the total claims values. The previous equations express the first four moments of the total claims values, and these values depend on the four moments of the proposed probability distributions for both the number and values of claims. By using the RStudio program, the values of the first four central moments of the probability distribution of the total claims values were estimated, and they were as shown in the following table:

Table.10 The first four central moments of the probability distribution of the total claims values

	Moment for Total N of claims		Moment for claim amount		Moment for total claim amount	
1st moment	$\bar{x}$	34037	$\bar{y}$	1.90E+03	$\bar{z}_{\partial}$	6.47E+07
2nd Moment	$\bar{x}_2$	74723	$\bar{y}_2$	6.15E+07	$\bar{z}_2(\partial)$	4.60E+12
3rd moment	$\bar{x}_3$	189125	$\bar{y}_3$	2.66E+12	$\bar{z}_3(\partial)$	5.03E+17
4th moment	$\bar{x}_4$	524543	$\bar{y}_4$	1.36E+17	$\bar{z}_4(\partial)$	7.13E+22

Using the Bowman & Shenton method, the weighted standard deviation can be estimated at the significance level z_{α} used to estimate the value of the maximum possible loss. The functions used take the following form (Bowman, 1979) (Hong Shiang Lau, 1984).

$$z_x(Standard\ deviation) = z_{\alpha}\left(\sqrt{k_1}, k_2\right) = \frac{\sum_{i=1}^{4} r_i(\sqrt{k_1}^{-N_i})\ (k_2)^{D_i}}{\sum_{i=1}^{4} s_i(\sqrt{k_1}^{-N_i})\ (k_2)^{D_i}},$$

where $\sqrt{k_1(\partial)} = \frac{\bar{z}_3(\partial)}{[\bar{z}_2(\partial)]^{1.5}}$ $is\ the\ skewness\ measure$

$$\sqrt{k_1(\partial)} = \frac{5.03E+17}{[4.60E+12]^{1.5}} = 0.0510665$$

$$k_2(\partial) = \frac{\bar{z}_4(\partial)}{[\bar{z}_2(\partial)]^2}\ is\ the\ Kurtosis\ measure$$

$$k_2(\partial) = \frac{7.13E+22}{[4.60E+12]^2} = 0.0033780$$
$$z_x(Standard\ deviation) = 2.921732.$$

The estimation of the standard deviation value depends on the value of both the skewness and kurtosis coefficients, which are values that depend in their calculation on the different moments of the distribution of the total claims value. By calculating the value of the coefficient, the net insurance price can be estimated, which depends on the value of the maximum possible annual loss (MPY) and the total insurance amounts. The equation for estimating the maximum annual loss reflects the value of the average distribution of the total claims value, taking into account the value of the standard deviation. The net insurance price equation takes the following form:

$$Net\ insurance\ rate = \frac{MPY_\alpha}{Total\ premiums} \ ,$$

where $$MPY_\alpha = mean + z_x(Standard\ deviation)$$

$$= \bar{z}_\partial + z_\alpha\left(\sqrt{k_1}, k_2\right) \times \left(\bar{z}_2(\partial)\right)^{0.5} = 6.47E + 07 + 2.921732 \times (1.57521E + 12)^{0.5}$$

$$= 68366989$$

$$Net\ insurance\ rate = \frac{68366989}{1285532900} = 0.053181828$$

To arrive at the total insurance price, the following relationships are relied upon, taking into account that the profit margin of the company under study ranged between 2.5% and 5% during the study period. Therefore, both the minimum and maximum limits for the total price were determined based on the development of the profit margin value during the study period. The average expense ratio during the study period was 23.87% (Emmet, 1999) (Ahmed, 1991).

$$Total\ insurance\ price = \frac{Net\ insurance\ price}{1 - (Expenses\ rate + Profit\ margin)}$$

$$Total\ insurance\ price\ (Min) = \frac{0.053181828}{1 - (23.87\% + 2.5\%)} = 0.0722$$

$$Total\ insurance\ price\ (Max) = \frac{0.053181828}{1 - (23.87\% + 5\%)} = 0.0748$$

The above results show that the total insurance rate at the company under study ranges between 7.22%, which represents the minimum, and 7.48%, as the maximum. Comparing this rate with the company's rate of 9.87%, we find that it is higher than the company's experience. Therefore, it is necessary to reduce the insurance rate to match its experience.

5. Conclusion

This article aims to present an actuarial model for estimating the value at risk (VAR) for vehicle insurance, given the availability of a set of variables associated with determining this premium. These variables reflect both the driver's demographics and the vehicle model itself, and are based on an appropriate probability distribution for the number and value of vehicle insurance claims. The standard deviation was estimated using the values of both the skewness and kurtosis coefficients, which are calculated based on the different moments of the distribution of total claims values. By calculating the coefficient, the net insurance price can be estimated, which is based on the value of the maximum possible annual loss (MPY) and the total insurance sums. The total insurance price was estimated taking into account that the profit margin of the company under study ranged between 2.5% and 5% during the study period. Therefore, both the minimum and maximum limits for the total price were determined based on the evolution of the profit margin value during the study period. The average expense ratio during the study period was 23.87%. It turned out that the total insurance rate at the company under study ranged between 7.22%, representing the minimum, and 7.48%, as the maximum. Comparing this rate with the rate used by the company, which is 9.87%, we find that it is higher than the company's experience results. Therefore, it is necessary to reduce the insurance rate to match its experience results. Through the practical application of the proposed model, the article reached an estimate of the price limits for vehicle insurance rates at the company under study. It became clear that the rate used by the company is far from these limits, indicating its inconsistency with the company's actual experience results. The Gamma distribution is considered the appropriate distribution for claim values, and the variables (age, vehicle age, policy term, and type) have a significant impact on claim values. The remaining variables (vehicle value, education level, and marital status) had no significant impact on the study's response variable, which represents claim values. The results showed that the total insurance price at the company under study ranged between 7.22%, representing the minimum, and 7.48%, as the maximum. Comparing this rate to the company's own rate of 9.87%, we find that it is higher than the company's experience. Therefore, it is necessary to reduce the insurance price to match its experience. The article recommended the development of statistical models and tools used in pricing general insurance in general, and car insurance in particular. It is also important for general insurance companies to adopt a pricing strategy that ensures fairness and objectivity in the pricing process, using actual loss experience to reflect the credibility of the results.

6. References

1. Abdel-Baqi, Reda Saleh (2015), "An Actuarial Approach to Adjusting Prices and Limits of Civil Liability Arising from Vehicle Accidents in Egypt - Applied to Private Vehicles," Journal of Contemporary Commercial Research, Faculty of Commerce, Sohag University.
2. Ahmed, Mamdouh Hamza, (1990), "Using Probability Distributions in Insurance Pricing with Application to Burglary/Commercial Property Insurance," Unpublished PhD Thesis, Faculty of Commerce, Cairo University.
3. Amolo, Ogutu Julie,(2011), “Claims reserving using over dispersed Poisson model”, Master of science in Actuarial science, School of mathematics, University of Nairobi.
4. Azaare, J., Wu, Z., Zhu, Y., Armah, G., Engmann, G. M., Kwadwo, S. M., ... & Ampaw, E. M. (2022). Measuring the adequacy of loss distribution for the Ghanaian auto insurance risk exposure through maximum likelihood estimation. Open Journal of Business and Management, 10(2), 846-859.
5. Badawi, Maher Duraid, (2013), "Classifying Vehicle Risks in the Saudi Insurance Market Using Discriminant Function Analysis," Arab Journal of Management, Arab Administrative Development Organization.
6. Bakhit, Ali Sayed, (2004), "Developing an Advanced Model for Pricing General Insurance with Application to Industrial All-Risk Insurance Data," Journal of Contemporary Commercial Research, Faculty of Commerce, Sohag, South Valley University.
7. Bowman, K.O. And .Shenton , L .R, (1979)“Further Approximate Pearson Percentage Points and Cornish-Fisher” , Communications in Statistics , Vol 8(3).
8. David, M. (2015) Auto Insurance Premium Calculation Using Generalized Linear Models. Procedia Economics and Finance, 20 147–156.
9. Denuit, M. and Charpentier, A. (2004) Mathématiques de l’assurance non-vie. T. 1: Principes fondamentaux de théorie du risque. Collection économie et statistiques avancées. Paris: Economica.
10. Emmet J Vaughan, Therese M Vaughan, (1999), Fundamental Of Risk And Insurance, John Willey & Suns Inc.
11. Goldburd, M., Khare, A. and Tevet, D. (2016) Generalized Linear Models for Insurance Rating. Arlington, Virginia: Casualty Actuary Society.
12. Gonnet, Guillaume,(2010), "Etude de la tarification et de la segmentation en assurance automobile", Brief presented before the Institute of Financial Science and Insurance for the degree of Actuary of the University of Lyon , University Claude Bernard - Lyon 1, Institute of Financial Science and Insurance.
13. Hon Shiang Lau , (1984), “An Effective Approach For Estimating The Aggregate Loss Of An Insurance Portfolio” , The Journal of Risk and Insurance , Volume LI, No.1.
14. Ieosanurak, W., Khomkham, B., & Moumeesri, A. (2023). Claim modeling and insurance premium pricing under a bonus-malus system in motor insurance. International Journal of Applied Mathematics and Computer Science, 33(4).
15. Lahcene, Bachioua,(2013),”On Pearson families of distributions and its applications”, African Journal of Mathematics and Computer Science Research.

16. Lai, Shu-Fang,(2008), " The accident risk measuring model for urban arterials",Takming University of Science and Technology, Taiwan, (https://pdfs.semanticscholar.org/8c74/100be2b99d94ec3db92a06fc5af14515a684.pdf).
17. Masuku, Zibusiso Vusumuzi,(2018),"Application of generalized linear models in pricing comprehensive motor insurance", The requirements of the bachelor of commerce degree in Actuarial science, National university of science and technology, Bulawayo, Zimbabwe.
18. Mohsen, Sharif Mohamed (2006), "Pricing Supplementary Vehicle Insurance Applied to Microbuses," unpublished master's thesis, Faculty of Commerce, Menoufia University.
19. O.Podladchikova,(2003)," Classification of probability densities on the basis of Pearson's curves with application to coronal heating simulations", Nonlinear Processes in Geophysics, European Geosciences Union (EGU).
20. Olga A. Vasechko et al,(2009), " Modélisation de la fréquence des sinistres en assurance automobile ", French Actuarial Bulletin, Vol 9, n: 18 , pp 41-63.
21. Pandya, S., Vaidya, N. M., Undavia, J. N., Patel, A. M., Kant, K., & Shukla, A. (2023, January). Model for mobile app-based premium calculation for usage-based insurance (UBI) of vehicles. In Mobile Application Development: Practice and Experience: 12th Industry Symposium in Conjunction with 18th ICDCIT 2022 (pp. 141-152). Singapore: Springer Nature Singapore.
22. Sayed, Ahmed Abdel Rahman, (2014), "Using a Combination of Financial and Actuarial Models in Pricing Comprehensive Insurance for Private Vehicles in the Saudi Market," Journal of Financial and Commercial Research, Faculty of Commerce, Port Said University, Volume 5, Issue 4. Salam, Osama Azmi, and Musa, Shaqiri Nouri (2015), "Risk Management and Insurance," Dar Al-Hamed for Publishing and Distribution, Jordan.
23. Shahrazad, Salhi (2015), "Modeling Vehicle Accident Pricing: A Standard Study on the Algerian Insurance Company SAA," unpublished master's thesis, Faculty of Economics, Business, and Management Sciences, Ferhat Abbas University - Setif 1.
24. Suwandani, R. N., & Purwono, Y. (2021). Implementation of gaussian process regression in estimating motor vehicle insurance claims reserves. Journal of Asian Multicultural Research for Economy and Management Study, 2(1), 38-48.
25. Yu, W., Guan, G., Li, J., Wang, Q., Xie, X., Zhang, Y., ... & Cui, C. (2021). Claim amount forecasting and pricing of automobile insurance based on the BP neural network. Complexity, 2021(1), 6616121.