MASTER

ACTUARIAL SCIENCE

MASTER´S FINAL WORK

PROJECT REPORT

TRANSFORMING MOTOR INSURANCE PRICING IN THE SAUDI INSURANCE SECTOR: THE ROLE OF GLMS

AHMED MOHAMMAD J ALSADIQ

OCTOBER - 2024

MASTER
ACTUARIAL SCIENCE

MASTER´S FINAL WORK
PROJECT REPORT

TRANSFORMING MOTOR INSURANCE PRICING IN THE SAUDI INSURANCE SECTOR: THE ROLE OF GLMS

AHMED MOHAMMAD J ALSADIQ

SUPERVISION:
PROF. ALEXANDRA BUGALHO DE MOURA
SYED AZADAR HAIDER

OCTOBER - 2024

Glossary

AIC – Akaike Information Criterion

BADRI – BADRI Management Consultancy

BIC – Bayesian Information Criterion

GLM – Generalized Linear Model

GWP – Gross Written Premium

IA – Insurance Authority

IBNR – Incurred but Not Reported

IFoA – Institute and Faculty of Actuaries

LER – Loss Elimination Ratio

NCD – No-Claim Discount

OS – Outstanding

PRP – Pure Risk Premium

SAR – Saudi Riyal

UW – Underwriting

ABSTRACT

This report investigates the application of Generalized Linear Models (GLMs) pricing techniques in the Saudi Arabian motor insurance market which priorly relies on Burning Cost. In light of the market's recent volatility—most notably the significant losses incurred in 2021 due to a price war—this study explores how modern actuarial methodologies can enhance pricing accuracy and restore profitability. The research addresses the limitations of traditional pricing methods currently used in Saudi Arabia and examines the feasibility and potential advantages of implementing GLMs. Drawing on relevant literature from developed insurance markets, this report evaluates the challenges and opportunities for adopting GLMs in Saudi Arabia. The findings suggest that GLM pricing models could provide notable improvements over existing practices by supporting better risk segmentation, addressing pricing inefficiencies, and potentially enhancing financial stability. Through a detailed analysis of industry data, this report provides insights that are crucial for both insurers and regulators to foster a more sustainable and competitive insurance environment.

KEYWORDS: Generalized Linear Models (GLMs); Saudi Arabian Motor Insurance Market; Insurance Pricing Accuracy.

RESUMO

Este relatório investiga a aplicação de técnicas de precificação com Modelos Lineares Generalizados (GLMs) no mercado de seguro automóvel da Arábia Saudita, que anteriormente se baseava no método de Custo Queimado. À luz da recente volatilidade do mercado — mais notavelmente as perdas significativas sofridas em 2021 devido a uma guerra de preços — este estudo explora como as metodologias atuariais modernas podem melhorar a precisão da precificação e restaurar a rentabilidade. A pesquisa aborda as limitações dos métodos de precificação tradicionais atualmente utilizados na Arábia Saudita e examina a viabilidade e as possíveis vantagens da implementação de GLMs. Com base em literatura relevante de mercados de seguros desenvolvidos, este relatório avalia os desafios e as oportunidades para a adoção de GLMs na Arábia Saudita. As conclusões sugerem que os modelos de precificação GLM poderiam proporcionar melhorias significativas em relação às práticas existentes, apoiando uma melhor segmentação de riscos, enfrentando ineficiências de precificação e potencialmente aprimorando a estabilidade financeira. Por meio de uma análise detalhada de dados da indústria, este relatório oferece insights que são cruciais para seguradoras e reguladores promoverem um ambiente de seguros mais sustentável e competitivo.

PALAVRAS-CHAVE: Modelos Lineares Generalizados (GLMs); Mercado de Seguro Automóvel da Arábia Saudita; Precisão na Precificação de Seguros.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

I would like to express my deepest gratitude to everyone who has contributed to the completion of this work.

First and foremost, I extend my sincere thanks to Prof. Alexandra Bugalho De Moura, whose guidance, insights, and encouragement were invaluable. I am grateful for her expertise, patience, and unwavering support.

I would also like to acknowledge the contributions of my friend, flat mate and manager, Syed Azadar Haider, FSA, for his constructive feedback and continuous encouragement. His collaboration and exchange of ideas significantly enhanced the depth of my understanding. Without him, this project wouldn't have seen the light.

Special thanks to my family and friends for their emotional support and understanding during this journey. Without their belief in my abilities, this achievement would not have been possible.

Lastly, I would like to thank Instituto Superior de Economia e Gestão – ISEG, for giving me the opportunity to do my master's in the beautiful country, Portugal. This experience won't be forgettable.

1. INTRODUCTION

In today's dynamic world, risk management is a critical concern for individuals and organizations striving to safeguard their assets and future stability. Insurance plays a pivotal role in this process by transferring risk from the insured to the insurer in exchange for a premium. Determining the appropriate price for this risk transfer is essential for both the financial stability of insurers and maintaining a fair and competitive market.

This report focuses on the motor insurance industry in Saudi Arabia, a sector experiencing rapid growth as the country diversifies its economy beyond oil. Valued at approximately SAR 70 billion, with motor insurance holding the second-largest market share, the industry faces significant challenges. A major downturn in 2021, resulting in losses of SAR 400 million, highlighted the dangers of unhealthy competition and price wars. To address these issues, insurers must adopt innovative pricing techniques to ensure long-term profitability and sustainability.

The central research question of this report is: How can pricing techniques, particularly Generalized Linear Models (GLMs), improve pricing accuracy and profitability in Saudi Arabia's motor insurance market? Accurate pricing is crucial for maintaining both insurer solvency and market competitiveness. Traditional pricing methods such as Burning Cost, which have been prevalent in the Saudi market, are insufficient given the increasing complexity of risk factors. GLMs, commonly applied in mature insurance markets, provide a framework to model these factors, including demographics, vehicle characteristics, and geographic location.

This report applies GLM method to the Saudi motor insurance market and compares their effectiveness against existing pricing practices. The analysis utilizes data from Saudi insurance companies and considers risk factors like regulatory changes, demographic trends, and historical claims data. Additionally, the report assesses the feasibility and challenges of implementing GLMs in a market that is still developing. The implementation is carried out using Addactis Software.

Drawing on multiple sources, including financial statements from Saudi insurance companies, reports from the Insurance Authority (IA), and performance analyses by BADRI Management Consultancy, this study provides practical recommendations for insurers and regulators. The findings underscore the need for regulatory support and actuarial oversight to successfully implement GLMs and improve market practices.

The report is structured as follows: Chapter 2 examines the need for innovative pricing techniques and introduces GLMs, evaluating their potential compared to current pricing methods. Chapter 3 outlines the research methodology, discussing data preparation, model selection, and testing, and presents the results of applying GLMs in the Saudi motor insurance market. Finally, Chapter 4 summarizes the key findings, offers recommendations for insurers, and suggests areas for future research, with a focus on the importance of regulatory involvement for sustainable industry growth.

2. Contribution to the Industry and Pricing Models

2.1. Saudi Motor Insurance Market Performance and Profitability Trends

To provide a comprehensive understanding of the Saudi motor insurance market's recent trends, we shall begin by presenting the Gross Written Premium (GWP) and Loss Ratio figures for the period spanning 2020 to 2023. This data will serve as a foundational benchmark, illustrating the overall market performance and highlighting key shifts in premium volumes.

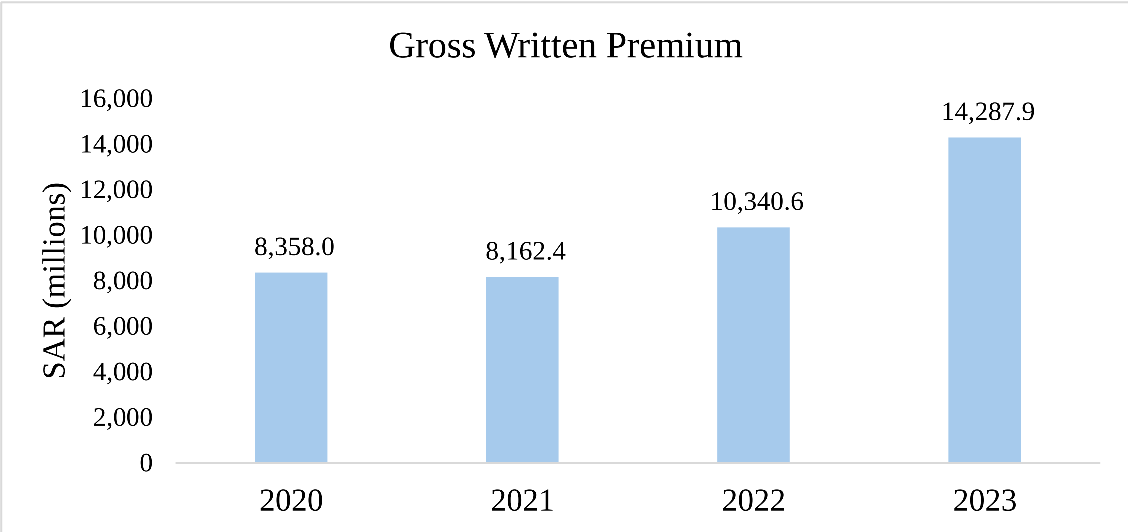

FIGURE 1.1 – Motor Insurance Gross Written Premium.[1]

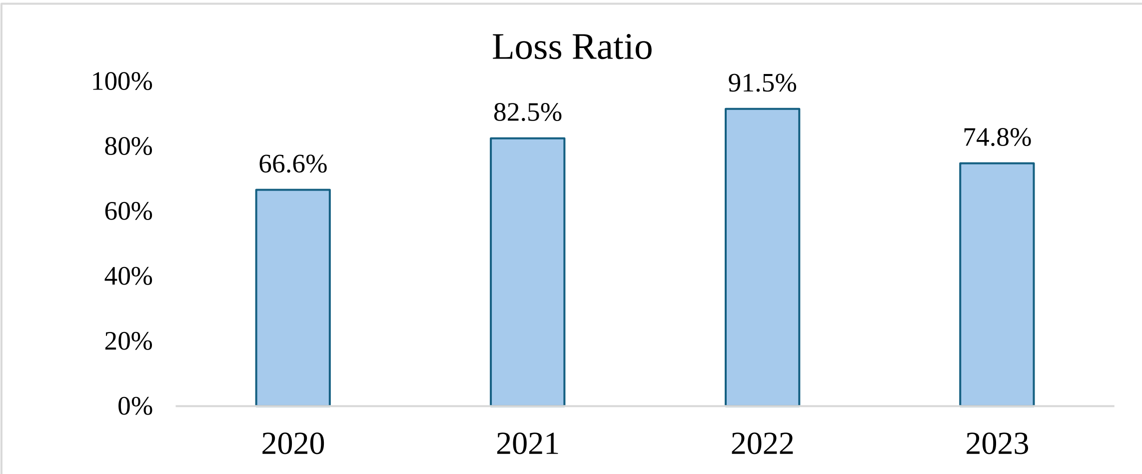

FIGURE 1.2 – Motor Insurance Loss Ratios.[2]

[1] IA's and BADRI's Insurance Industry Reports have been reviewed to verify this work.
[2] Same as Figure 1.1

At first glance, the motor insurance market appeared to be performing well in terms of GWP, with loss ratios remaining below 100%. Although the GWP in 2021 experienced a slight decline compared to the previous year, no immediate concerns were evident based on these figures alone. However, this assessment does not provide a comprehensive view. To fully understand the impact of the 2021 price war, it is essential to evaluate the market's profitability, as GWP and loss ratios alone do not capture the full financial picture.

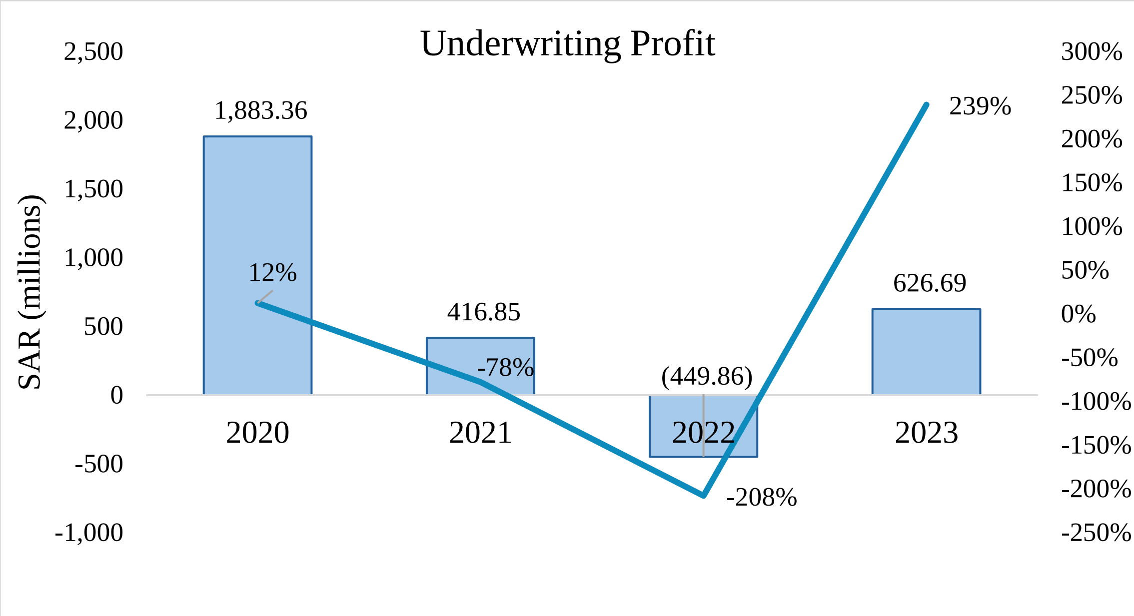

FIGURE 1.3 – Motor Insurance Underwriting Profit.[3]

As observed from the figure above, there was a significant decline in profitability in 2022, primarily driven by policies underwritten in 2021. The reported profit for 2022 was SAR -449.86 million, compared to SAR 416.85 million in the previous year, representing a sharp decline of approximately -208%. This substantial shift underscores the impact of the 2021 pricing strategies on the financial performance of the motor insurance market.

2.2. Challenges in the Sector, and Role of Regulators and Actuaries

Historically, insurance companies in Saudi Arabia enjoyed considerable freedom in determining the methods used to price their policies. For many years, they relied on relatively unsophisticated pricing techniques, largely to avoid the complexities associated with more innovative methods. A key reason for this was the limited statistical expertise

[3] The calculations were done by BADRI, and the files can't be shared.

among underwriters, who often lacked the necessary knowledge to conduct comprehensive risk assessments. Furthermore, Actuarial Science was still an emerging discipline in the Saudi market, with formal education in the field only beginning in 2011 when universities started offering programs.

The motor insurance sector, in particular, is characterized by volatility, where pricing changes can have immediate and visible impacts due to the short-term nature of policies, which typically last only one year. While the basic methods previously employed by insurers proved effective for some time, the situation deteriorated when companies began engaging in aggressive price competition, often offering unwarranted discounts in an attempt to undercut their competitors. This unsustainable practice ultimately led to significant financial strain across the industry.

In response to these challenges, IA introduced new regulations aimed at promoting greater rigor and oversight in the pricing of insurance products.

One of the key stipulations is that pricing must now be conducted by actuaries, and the pricing report must be signed by a certified Fellow Actuary from a recognized actuarial organization, such as the Institute and Faculty of Actuaries (IFoA). This requirement applies whether the actuary is employed internally by the insurer or serves as an external consultant.

Additionally, IA mandates that pricing methods for all lines of business must be sophisticated, such as GLM, utilizing advanced statistical techniques. However, IA does not prescribe specific methods, leaving companies the flexibility to choose their approach, provided it adheres to the requirement for advanced statistical rigor. These regulatory changes indicate a shift toward enhanced professionalism and technical precision in the Saudi insurance market, aiming for alignment with global actuarial standards.

2.3. Introduction to Generalized Linear Models (GLMs)

GLMs are a flexible extension of traditional regression models, particularly valuable in insurance pricing. Unlike traditional linear regression, which assumes normally distributed errors, GLMs allow for modelling non-normal data distributions, such as Poisson, Gamma, and Binomial, making them ideal for analysing insurance claims (Dobson & Barnett, 2008).

In the GLM models the relationship between μ_i (the model prediction) and the predictor variables (the independent variables) is as follows:

(1) $$g(\mu_i) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} \dots + \beta_p x_{ip} + \varepsilon_i$$

Where:

$g(\mu_i)$ *is the link function*

$\beta_0 = Intercept\ term$

$\beta_1, \beta_2, \dots, \beta_p = Estimated\ Coefficients$

$x_1, x_2, \dots, x_p$ = *Rating variables*

$\varepsilon_i = Error\ Term$

If the log-link function $g(\mu_i) = \ln(\mu_i)$ is considered, then the additive nature of the above equation transforms into a multiplicative equation, which is a desirable rating structure for pricing insurance (Ohlsson & Johansson, 2010).

In such case, the risk premium is determined using the following equation:

(2) $$\mu_i = e^{\beta_0} \times e^{\beta_1} \times e^{\beta_2} \times \dots e^{\beta_p}$$

In a GLM, the target variable—such as the frequency of claims, severity of claims, or the risk premium—is modeled as a random variable belonging to the exponential family of distributions. When modeling the severity of claims, the Gamma or Inverse Gaussian distribution is typically used. For modeling claim frequency, the Poisson or Negative Binomial distributions are commonly applied, while the Tweedie distribution is often employed to model the risk premium (De Jong & Heller, 2008).

To model the frequency and severity of claims, we utilized probability distributions that are commonly applied in insurance. For the frequency of claims, Poisson distribution was used. This distribution is suitable for modeling the number of events (in this case, claims) that occur over a fixed period, given that events happen independently of each other.
For the severity of claims, Gamma distribution was employed. The Gamma distribution is often used in insurance to model the size or cost of claims, particularly because it can capture a wide range of claim sizes, including those with significant variability.
The risk premium is calculated as the product of the expected claim frequency and the expected claim severity. This ensures that both the likelihood of a claim occurring, and the potential financial impact of the claim are taken into account when determining the appropriate premium. By combining these two components, the model reflects the total expected cost of claims, which forms the basis for setting insurance premiums.

To ensure the GLM is effective, the dataset is typically split into training and test sets (e.g., 70% training and 30% testing) to assess the model's predictive performance on unseen data (James et al., 2013). This process helps assess whether the model can generalize effectively to new cases and mitigate overfitting risks.

Model refinement is one of the most critical aspects of GLM construction. To evaluate how well the model fits the training data, several statistical measures are utilized. These diagnostics help assess the significance of each rating variable and the overall quality of the model. The commonly used diagnostic tools include:

- **Deviance Test:** The deviance test assesses whether the inclusion of a predictor variable significantly reduces the model's deviance, indicating an improvement in the model's fit. It is used to evaluate the goodness-of-fit in a GLM by comparing two nested models. The deviance test compares two models, one with and one without the predictor variable, and uses an F-distribution to determine whether the reduction in deviance is statistically significant, thereby indicating that the predictor variable improves model fit (James et al., 2013).

- **Residual Analysis**: Residuals are the differences between observed values and the predicted values from the model. Residual analysis helps evaluate the randomness in the model's errors, and a well-fitting model will display no discernible pattern in its residuals (Ohlsson & Johansson, 2010).

- **Akaike Information Criterion (AIC):** AIC is a criterion used for model selection, balancing goodness-of-fit with model complexity. It imposes a penalty for adding more parameters, aiming to find a model that is both accurate and simple. AIC evaluates models by balancing their goodness-of-fit against complexity. It is calculated using the log-likelihood of the model and adding a penalty for each additional parameter. The model with the lowest AIC is preferred, as it represents the best balance between accuracy and simplicity (James et al., 2013).

- **Bayesian Information Criterion (BIC):** BIC, like AIC, is used for model selection but applies a stricter penalty for models with more parameters. It is particularly useful with large datasets, where simpler models are generally preferred. BIC works similarly to AIC but imposes a larger penalty for model complexity, especially as the sample size increases, making it more likely to favor simpler models. The goal of BIC, like AIC, is to minimize its value when selecting the best model (James et al., 2013).

- **Pseudo-R^2:** Pseudo-R^2 is a measure of goodness-of-fit used for non-linear models, such as GLMs, where traditional R^2 is not applicable. It estimates how much of the variance in the data is explained by the model, offering a rough guide to model performance. Pseudo-R^2 compares the likelihood of the fitted model with a null model (a model with no predictors). Higher values indicate a better fit, but unlike AIC or BIC, Pseudo-R^2 does not penalize for additional parameters, making it less robust for model selection when overfitting is a concern (Hosmer & Lemeshow, 2000).

- **Consistency with Time and Common Sense**: These diagnostic checks if the predicted and observed values maintain logical consistency over time. This time consistency check is vital to ensure the model does not produce results that are at odds with industry knowledge or real-world trends (Ohlsson & Johansson, 2010).

- **Professional Judgment**: Beyond statistical measures, professional judgment and actuarial expertise are essential in evaluating whether the model outputs make sense in the context of real-world scenarios. This ensures that the model’s predictions are not only mathematically valid but also practically applicable.

GLMs have become a cornerstone of insurance pricing because they allow actuaries to integrate multiple risk factors and adjust models as conditions change. This flexibility supports transparency and fairness, ensuring alignment with regulatory requirements (Ohlsson & Johansson, 2010).

2.4. Comparison of GLMs with Existing Pricing Practices

Burning Cost is a simpler and more traditional approach to insurance pricing, especially in reinsurance and large commercial policies. The Burning Cost method calculates future premiums based on historical claims experience using the formula:

The Burning Cost or Pure Risk Premium (PRP) is calculated as:

$$Burning\ Cost = \frac{Ultimate\ Expected\ Incurred\ Claims}{Exposure} \tag{3}$$

Where:

Ultimate Expected Incurred Claims: The estimate of total claims liability, considering both reported and unreported claims, as well as reserves for case estimates and Incurred but Not Reported claims reserve, IBNR.

Exposure: The base measure of risk, such as the number of insured units (e.g., vehicles), premium volume, or total insured value.

This formula is commonly used in actuarial pricing and reserving to determine the appropriate premium rate based on historical claims experience and exposure (De Jong & Heller, 2008).

Burning Cost is ideal for scenarios where the risk profile is stable, and historical data provides a reliable forecast for future outcomes. It is frequently applied in policies with large deductibles or excess layers, where insurers cover only significant losses. However, its heavy reliance on past claims data reduces its effectiveness in more dynamic environments where risks evolve over time (Ohlsson & Johansson, 2010).

While straightforward and easy to apply, Burning Cost lacks the sophistication of models like GLMs, which can account for multiple risk factors and produce more granular pricing. Both Burning Cost and GLMs are used in actuarial pricing, but they differ significantly in complexity, flexibility, and applicability, as described in Table I.

TABLE I – COMPARISION BETWEEN GLM AND BURNING COST

Aspect	Generalized Linear Models (GLMs)	Burning Cost
Approach	Uses statistical models to incorporate multiple risk factors (Dobson & Barnett, 2008).	Relies purely on historical claims data for future premium calculations (Ohlsson & Johansson, 2010).
Data Requirements	Requires comprehensive datasets with multiple variables (James et al., 2013).	Needs only historical claims and exposure data, making it more applicable when data is limited (De Jong & Heller, 2008).
Complexity	High complexity, requiring statistical expertise and large datasets (Ohlsson & Johansson, 2010).	Simpler and easier to implement, relying mainly on historical data (James et al., 2013).
Flexibility	Highly flexible, allowing for complex modeling of risk environments (De Jong & Heller, 2008).	Less flexible, best used in stable environments where historical data predicts future outcomes (Ohlsson & Johansson, 2010).
Scalability	Easily scalable and adaptable as new data becomes available (Ohlsson & Johansson, 2010).	Limited scalability, primarily reliant on historical data, without easily incorporating new factors (De Jong & Heller, 2008).
Predictive Power	High predictive power, accounting for risk factors and changes over time (De Jong & Heller, 2008).	Lower predictive power, relying on historical data with limited ability to adapt to changing risks (James et al., 2013).
Accuracy and Fairness	More accurate and fair, adjusting premiums based on risk factors, reducing bias (Ohlsson & Johansson, 2010).	Can be less accurate if historical data does not adequately predict future risks (Ohlsson & Johansson, 2010).

While GLMs offer higher predictive accuracy and flexibility, they require substantial data and statistical expertise. Burning Cost, in contrast, is simpler and easier to apply, making it suitable for stable risk environments but less capable of handling dynamic changes in risk profiles (De Jong & Heller, 2008).

In practice, Burning Cost is typically used in reinsurance or excess layers, where historical claims provide a solid basis for pricing, whereas GLMs are more common in personal lines insurance, such as motor and home insurance, where detailed risk segmentation is essential (Ohlsson & Johansson, 2010).

3. GLM Implementation and Analysis

To demonstrate the superiority of the GLM for pricing motor insurance, we utilized market data representing approximately 60% of the total motor insurance market. We specifically focused on the motor comprehensive segment for comparison purposes, as comprehensive policies typically contain more detailed information about the driver and the vehicle than third-party liability policies. This allows for a more accurate and granular analysis.

3.1. Data Collection and Preparation

For the purpose of this analysis, we have collected the following data items:

1. Vehicle level premium data
2. Policy level claims data

The analysis is based on data from the most recent four years, spanning from 2020 to 2023. Table II below provides a summary of the premium and claims data for the policies included in the study, which comprise a total of 3,079,469 policies.

Table II – Motor Comprehensive – GWP & Gross Reserves Summary

(Amounts in SAR million)

UW Year	GWP	Paid	OS	Reported Claims	Reported Loss Ratio	IBNR	Earned Premium	Ultimate - Used	Ultimate Loss Ratio
2020	2,507	1,677	27	1,704	68%	3	2,507	1,707	68%
2021	2,449	1,613	6	1,619	66%	7	2,449	1,626	66%
2022	3,102	2,102	165	2,267	73%	61	3,102	2,327	75%
2023	4,286	882	1,136	2,017	78%	148	2,572	2,165	84%
Total	**12,344**	**6,273**	**1,333**	**7,606**	**72%**	**219**	**10,630**	**7,826**	**74%**

The data has been carefully reconciled with the respective financial statements to ensure the accuracy and integrity of the information used. Table III below presents the allocation of claims data by underwriting year. The formula used in calculating the exposure is given by:

$$\textbf{Exposure}\ (Vehicle\ Years) = \frac{(Vehicle\ End\ Date - Vehicle\ Start\ Date)}{365\ days} \tag{4}$$

Vehicle years earned are used as the exposure base for both the frequency and severity models. However, the final risk premium is projected as a percentage of the sum insured, aligning with industry practices where rates are typically expressed as a percentage of the sum insured. To calculate the final risk rate, the projected risk premium is divided by the sum insured. The unit of exposure is adjusted based on the standard 365-day coverage period typically provided by a policy.To execute the GLM, we utilized Addactis, a reputable and widely recognized software in the market, known for its specialization in GLM pricing. As mentioned earlier, within the GLM framework, we must assign costs to the predictive variable, which in this case is the exposure, and it has been calculated as follows:

TABLE III – EARNED EXPOSURE BY UW YEAR

UW Year	Vehicle Count	2021	2022	2023	Total	Earned Exposure/ Vehicle Count
2020	627,312	313,656	-	-	313,656	50%
2021	884,754	353,902	530,852	-	884,754	100%
2022	764,251	-	420,338	343,913	764,251	100%
2023	803,151	-		449,765	449,765	56%
Total	**3,079,469**	**667,558**	**951,191**	**793,678**	**2,412,426**	**78%**

In conducting the GLM analysis, we considered the rating variables presented in the Table VI and VII for the frequency and severity model respectively. However, it is important to note that the collection and availability of these variables can be challenging, as GLM is not yet widely adopted in the market, which presents a constraint in the pricing process. Despite this limitation, we have made every effort to utilize the variables available within the dataset.

3.2. Model Assumptions

3.2.1. Inflation Assumption

In order to understand the movement of market prices, we analyzed the inflation trend in rates over recent years. Figure 3.1 presents the inflation trend alongside the pure premium in the market.

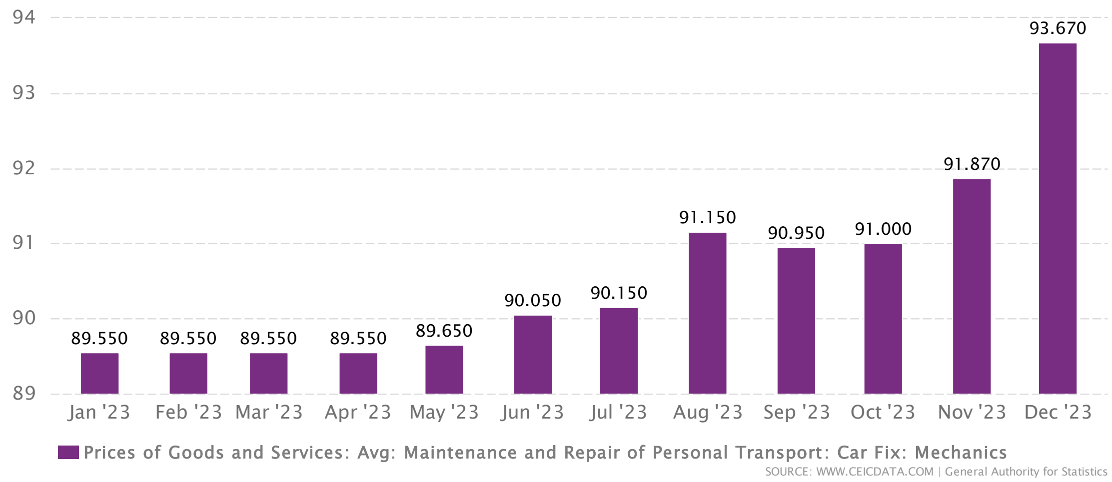

FIGURE 3.1 – Average Maintenance and Repair Services by Mechanics for Personal Cars in Saudi Arabia.[4]

As per Figure 3.1, 2% inflation has been considered. The following formula has been used to make the on-level adjustment for inflation in the considered period:

$$(5) \qquad Inflation = (1 + Trend)^{\frac{(Average(Policy\ Start\ Date,\ \ Policy\ End\ Date) - Pricing\ Date)}{Number\ of\ Average\ Days\ in\ a\ Month/Number\ of\ Months}} - 1$$

3.2.2. Claims Large Losses Adjustment

To simplify the analysis, the claims data has been reviewed, and outliers have been removed. However, to account for the impact of large losses, an additional loading of 2% has been added to reflect the threshold for large losses. The additional loading for large claim threshold has been added to the claims' data before uploading it to Addactis for pricing to incorporate the impact of large losses. In order to avoid distortions in the GLM results caused by large claims, we had to cap the large claims. The large claim threshold for Comprehensive is shown in Table IV.

TABLE IV – LARGE LOSS THRESHOLD

Segment	Large Loss Threshold
Comprehensive	100% of Sum Insured

[4] Source: https://www.ceicdata.com/en/saudi-arabia/average-prices-of-goods-and-services/prices-of-goods-and-services-avg-maintenance-and-repair-of-personal-transport-car-fix-mechanics

3.3. Model Selection

We evaluated AIC, BIC, Deviance, and Pseudo-R^2 for the GLM frequency and severity models, which offered varied insights into model performance and complexity in our insurance pricing analysis. By comparing these metrics, we aimed to find the model that best balances fitness, complexity, and predictive performance.

We chose AIC for model selection in our frequency and severity GLM models because it optimizes for predictive accuracy, which is essential for insurance pricing. Unlike BIC, which focuses more on parsimony, AIC allows for the selection of more complex models if they enhance predictive performance. This is crucial in our case, where capturing the complex relationships between variables improves the accuracy of future claims estimates. AIC's ability to handle model complexity while maintaining prediction accuracy makes it the most suitable criterion for our analysis.

Based on the model selection for both frequency and severity (see Tables A.I and A.II in the Appendix), Table V below presents the parameters and factors used for each of the models.

The detailed parameters and output are provided in the following Tables.

TABLE V – MODEL SETTING

	Frequency	**Severity**	**Risk Premium**
Probability Distribution	Poisson	Gamma	
Training/Validation Set	70% Training / 30% Validation	70% Training / 30% Validation	100% Training set
Factors:	Vehicle Make Sum Insured Repair Condition Driver Nationality Vehicle Age Driver Age Region	Vehicle Make Vehicle Body Type Sum Insured Repair Condition Vehicle Age Driver Age Region Year	Deductible Vehicle Make Vehicle Body type Sum Insured Repair Condition Driver Nationality Driver Age Vehicle Age Region NCD Year

For most of the rating variables, the data was readily available and of sufficient quality for use. However, information pertaining to Business Industry was not included in the dataset; instead, it was provided to us separately by the companies involved.

Driver Age was also not directly available in the data. We calculated it by using the driver's birth date in combination with the vehicle start date.

The No-Claim Discount (NCD) is determined by the IA. To incorporate the rate adjustment attributed to NCD loading, we manually input the relativity factors into the GLM during the calculation of the base rate and related factors because the NCD is fixed by the government as per the experience of the policyholder in Saudi Arabia. For example, an insured with having one year of no claim experience will get 10% discount on policy premium, which will be applied automatically through the system as per the instructions of the government. Similarly, 2 years of no claims will have 20% discounts and respectively. The complete details of the NCD relativities have been presented later Table XVII. This approach allowed us to integrate the impact of the NCD directly into the base rate, eliminating the need for a separate NCD loading mechanism.

For the deductibles offered under Comprehensive policies, the deductible bands have been selected based on actuarial judgement and market utilization. We applied the Loss Elimination Ratio (LER) method to determine the appropriate deductible relativities for each band. As these relativities are fixed, we employed a similar approach as with NCD, manually entering the relativity factors for deductibles into the GLM when determining the base rate and other related relativities.

3.4. GLM Model Results[5]

It should be noted that the multipliers presented in this section from Table IX to Table XVII are the GLM results. The multipliers that we have obtained from the GLM have been smoothed based on the market performance and experience of each factor within Saudi Arabia. For example, the relativity factors for the region, we know as per the claim experience that the country observes higher claims in the central region. Since the initial GLM had suggested otherwise, we had edited it to incorporate the market condition. The approach is based on the Actuarial judgement and understanding of the market behaviour.

[5] See Table A.III in the Appendix

Table VI – Motor Comprehensive Frequency Tested Models

Factor	Deviance	Pseudo-R^2	AIC	BIC
Null	10,091	0.0%	13,498	13,506
Vehicle Age	9,851	2.4%	13,282	13,384
Driver Age	9,799	2.9%	13,252	13,440
Sum Insured	9,763	3.2%	13,232	13,483
Vehicle Make	9,749	3.4%	13,224	13,498
Region	9,738	3.5%	13,217	13,507
Year	9,719	3.7%	13,202	13,507
Repair Condition	9,714	3.7%	13,199	13,513
Driver Nationality	9,711	3.8%	13,198	13,519
Body Type	9,707	3.8%	13,198	13,535
Body Color	9,705	3.8%	13,207	13,583

Table VII – Motor Comprehensive Severity Tested Models

Factor	Deviance	Pseudo-R^2	AIC	BIC
Null	2,224	0.0%	- 3,608	- 3,597
Sum Insured	2,016	9.3%	- 3,810	- 3,755
Year	1,943	12.6%	- 3,888	- 3,821
Vehicle Age	1,899	14.6%	- 3,915	- 3,781
Repair Condition	1,883	15.3%	- 3,932	- 3,792
Vehicle Make	1,887	14.4%	- 3,921	- 3,765
Driver Age	1,864	15.5%	- 3,926	- 3,709
Region	1,856	15.8%	- 3,932	- 3,703
Body Type	1,849	16.2%	- 3,936	- 3,696
Body Color	1,843	16.4%	- 3,933	- 3,665
Driver Nationality	1,843	16.4%	- 3,931	- 3,658

Table V1 and VII present the criterions for the model selection with the inclusion of each variable. For example, the first row of Table VI represents the value of the decision criterion considering the model has no variable but only the intercept term. Moving on, the values of the criteria after adding each variable in the model.

As mentioned earlier, we have selected AIC as our main criteria for model selection and, the model with all the variables having lowest AIC has been selected, and same has been highlighted in the Tables above.

TABLE VIII – MINIMUM RISK PREMIUM AND INTERCEPT

Risk Premium	Comprehensive
Intercept	6.30%
Minimum Risk Premium	SAR 1,100

It should be noted that this the intercept term in Table VIII is not an average rate since its value is entirely dependent upon the choice of which level of each factor is selected to be the base level.

TABLE IX – MOTOR COMPREHENSIVE - BODY TYPE

Body Type	Addactis Mapped	Multiplier	Smoothed Multiplier	Exposure
SUVs - 4x4/ JEEP	SUVs - 4x4/ JEEP	1.00	1.00	1,262,562
Sedans	Sedans	1.08	1.20	1,027,062
Big Buses	Others	1.17	0.70	
Equipment	Others	1.17	0.65	
Heavy Commercial	Others	1.17	0.70	
Light Commercial	Others	1.17	0.75	122,802
Motorcycles	Others	1.17	0.65	
Others	Others	1.17	0.70	
Pick-Ups	Others	1.17	1.15	
Small to Medium Buses	Others	1.17	1.17	

TABLE X – MOTOR COMPREHENSIVE - REGION

Region	Addactis Mapped	Multiplier	Smoothed Multiplier	Exposure
Eastern	Eastern	1.00	1.00	1,078,052
Blanks	Others	0.92	1.13	
Northern	Others	0.92	0.94	864,950
Southern	Others	0.92	0.90	
Western	Others	0.92	0.90	
Central	Central	1.16	1.13	469,424

TABLE XI – MOTOR COMPREHENSIVE – VEHICLE AGE

Vehicle Age	Addactis Mapped	Multiplier	Smoothed Multiplier	Exposure
0	[0; 1[	1.00	1.00	606,632
1	[1; 2[	0.95	1.05	283,538
2	[2; 3[	0.90	0.95	204,418
3	[3; 4[	0.78	0.85	165,476
4	[4; 5[	0.74	0.75	137,232
5	[5; 6[	0.68	0.70	157,026
6	[6; 7[	0.84	0.70	153,566
7	[7; 8[	0.57	0.65	156,436
8	[8; 9[	0.55	0.65	123,458
9	[9; 10[	0.69	0.65	92,814
10	[10; 11[	0.57	0.65	77,712
11	[11; ∞[	0.89	0.65	254,118

TABLE XII – MOTOR COMPREHENSIVE - NATIONALITY

Nationality	Addactis Mapped	Multiplier	Smoothed Multiplier	Exposure
Saudi	Saudi	1.00	1.00	1,986,726
Bangladeshi	Others	0.88	1.24	425,700
Yemeni	Others	0.88	0.79	
Syrian	Others	0.88	0.97	
Pakistani	Others	0.88	0.98	
Sudanese	Others	0.88	0.95	
Others	Others	0.88	0.99	
Palestinian	Others	0.88	0.99	
Afghan	Others	0.88	0.99	
Lebanese	Others	0.88	0.99	
Filipino	Others	0.88	0.99	
Kuwaiti	Others	0.88	0.99	
Europeans	Others	0.88	0.99	
American	Others	0.88	0.99	
Egyptian	Others	0.88	1.12	
Indian	Others	0.88	1.08	
Jordanian	Others	0.88	0.99	

TABLE XIII – MOTOR COMPREHENSIVE - SUM INSURED

Sum Insured	Addactis Mapped	Multiplier	Smoothed Multiplier	Exposure
0K to 25K	[0; 25001[	1.04	1.00	291,734
25K to 50K	[25001; 50001[	1.00	0.91	520,760
50K to 75K	[50001; 75001[	0.75	0.80	408,806
75K to 100K	[75001; 100001[	0.51	0.60	427,688
100K to 200K	[100001; 200001[	0.38	0.45	495,534
200K to 300K	[200001; 300001[	0.29	0.30	170,310
300K to 500K	[300001; 500001[	0.27	0.30	61,582
500K to 1M	[500001; 1000001[	0.14	0.30	25,520
Above 1M	[1000001; ∞[	0.10	0.30	10,492

TABLE XIV – MOTOR COMPREHENSIVE – REPAIR CONDITION

Repair Condition	Addactis Mapped	Multiplier	Smoothed Multiplier	Exposure
Agency	Agency	1.62	1.20	1,035,802
Non-Agency	Non-Agency	1.00	1.00	1,376,624

TABLE XV – MOTOR COMPREHENSIVE - DEDUCTIBLE

Deductible	Addactis Mapped	Multiplier	Smoothed Multiplier	Exposure
0 to 499	[0; 500[	1.00	1.00	75,514
500 to 749	[500; 750[	0.95	0.95	529,076
750 to 999	[750; 1000[	0.93	0.93	365,562
1,000 to 1,499	[1000; 1500[	0.90	0.90	944,102
1,500 to 1,999	[1500; 2000[	0.86	0.86	238,520
2,000 to 2,999	[2000; 3000[	0.81	0.81	171,168
3,000 to 4,999	[3000; 5000[	0.74	0.74	23,620
5,000 to 9,999	[5000; 10000[	0.65	0.65	49,064
10,000 Above	[10000; ∞[	0.50	0.50	15,800

TABLE XVI – MOTOR COMPREHENSIVE - DRIVER AGE

Driver Age	Addactis Mapped	Multiplier	Smoothed Multiplier	Exposure
Error	[-1; 16[	1.68	1.50	7,944
16-20	[16; 21[	1.03	1.27	58,144
21-24	[21; 25[	0.98	1.12	139,204
25-29	[25; 30[	1.00	1.00	335,432
30-34	[30; 35[	0.89	0.89	334,628
35-38	[35; 39[	0.87	0.89	264,852
39-42	[39; 43[	0.68	0.82	217,980
43-45	[43; 46[	0.80	0.90	154,636
46-50	[46; 51[	0.84	0.94	239,280
51-55	[51; 56[	0.82	0.97	206,190
56-59	[56; 60[	0.66	0.91	129,272
60+	[60; ∞[	0.68	0.92	324,778

TABLE XVII – MOTOR COMPREHENSIVE - NCD ELIGIBILITY

NCD Years	Multiplier	Smoothed Multiplier	Exposure
Missing	1.00	1.00	1,236,116
0	1.00	1.00	272,686
1	0.85	0.85	164,010
2	0.75	0.75	94,702
3	0.65	0.65	66,630
4	0.50	0.50	48,024
5	0.40	0.40	67,510
11	1.00	1.00	265,058
12	0.85	0.85	109,762
13	1.00	1.00	44,688
14	1.00	1.00	36,292
15	1.00	1.00	6,948

Table IX – XVII present the relativity factor that should be applied for the calculation of the premium based on the GLM pricing representing the characteristics of the policyholder. For examples, as per Table IX, we will be applying the relativity of the variable “Body Type” based on the vehicle of the policy holder, e.g. if the vehicle is an SUV, then the relativity of “1.00” should be applied.

3.5. GLM vs Burning Cost

In insurance pricing, the Burning Cost method and GLM might be used. These methods adjust premiums based on multiple risk factors, but they approach the problem differently. The Burning Cost method is a simpler approach that directly uses historical claims experience to set premiums, while GLM uses statistical modeling to account for correlations between risk factors.

This analysis focuses on comparing these two methods using a dataset with multiple variables. The analysis of the differences between Burning Cost and the GLM has been performed using two variables, Sum Insured and Repair Condition, for the simplification. However, practically it is not possible to calculate a unique Burning Cost for each of the combinations among the variable. This is one of the advantages of using GLM over Burning Cost. We will compare the premium calculations of both methods and highlight their respective strengths and weaknesses.

The dataset used in this comparison includes all the variables in Table XVIII

TABLE XVIII – OVERVIEW OF VARIABLES IN THE DATASET

Variable	Number of Levels
Vehicle Make	4
Body Type	3
Sum Insured	9
Repair Condition	2
Driver Nationality	2
Vehicle Age	12
Driver Age	12
Region	3
Year	3
Deductible	9
NCD	12
Total Combinations	**60,466,176**

There are over 60 million possible combinations of these variables, which makes the choice of pricing model highly important for ensuring accuracy and fairness in premium calculation. For this section, we focus on the Sum Insured and Repair Condition variables to explore the differences between Burning Cost and GLM.

The Burning Cost method calculates the premium based on historical claims data. It applies specific percentage adjustments for each combination of Sum Insured and Repair Condition without considering the underlying correlation between the variables. Below is Table XIX that represents the premium percentages for each combination of these two variables.

TABLE XIX – BURNING COST PREMIUM ADJUSTMENTS

Sum Insured	Repair Condition	Burning Cost Premium (%)
0K to 25K	Agency	3.72%
25K to 50K	Agency	0.96%
50K to 75K	Agency	3.54%
75K to 100K	Agency	3.25%
100K to 200K	Agency	1.95%
200K to 300K	Agency	1.94%
300K to 500K	Agency	2.54%
500K to 1M	Agency	3.34%
Above 1M	Agency	1.23%
0K to 25K	Non-Agency	4.22%
25K to 50K	Non-Agency	2.99%
50K to 75K	Non-Agency	0.08%
75K to 100K	Non-Agency	1.18%
100K to 200K	Non-Agency	3.28%
200K to 300K	Non-Agency	3.13%
300K to 500K	Non-Agency	3.57%
500K to 1M	Non-Agency	1.82%
Above 1M	Non-Agency	3.35%

In the Burning Cost method, the premium percentages are directly applied based on the combinations of Sum Insured and Repair Condition. There is no base premium in this approach; instead, each combination of factors is assigned a specific premium percentage.

The GLM approach takes a more sophisticated statistical approach. Instead of applying fixed percentages, the premium is calculated based on the formulaic approach by multiplying the base rate (intercept) with the respective multipliers that reflect the combined effect of the different variables. The GLM base is 6.3%, and this base premium is adjusted by applying multiplicative factors for each combination of Sum Insured and Repair Condition, as depicted in Table XX.

Table XX – GLM Premium Adjustments

Sum Insured	Multiplier	Repair Condition	Multiplier	GLM Premium (%)
0K to 25K	1.00	Agency	1.20	6.30%×1.00×1.20=7.56%
25K to 50K	0.91	Agency	1.20	6.30%×0.91×1.20=6.87%
50K to 75K	0.80	Agency	1.20	6.30%×0.80×1.20=6.05%
75K to 100K	0.60	Agency	1.20	6.30%×0.60×1.20=4.54%
100K to 200K	0.45	Agency	1.20	6.30%×0.45×1.20=3.40%
200K to 300K	0.30	Agency	1.20	6.30%×0.30×1.20=2.27%
300K to 500K	0.30	Agency	1.20	6.30%×0.30×1.20=2.27%
500K to 1M	0.30	Agency	1.20	6.30%×0.30×1.20=2.27%
Above 1M	0.30	Agency	1.20	6.30%×0.30×1.20=2.27%
0K to 25K	1.00	Non-Agency	1.00	6.30%×1.00×1.00=6.30%
25K to 50K	0.91	Non-Agency	1.00	6.30%×0.91×1.00=5.73%
50K to 75K	0.80	Non-Agency	1.00	6.30%×0.80×1.00=5.04%
75K to 100K	0.60	Non-Agency	1.00	6.30%×0.60×1.00=3.78%
100K to 200K	0.45	Non-Agency	1.00	6.30%×0.45×1.00=2.84%
200K to 300K	0.30	Non-Agency	1.00	6.30%×0.30×1.00=1.89%
300K to 500K	0.30	Non-Agency	1.00	6.30%×0.30×1.00=1.89%
500K to 1M	0.30	Non-Agency	1.00	6.30%×0.30×1.00=1.89%
Above 1M	0.30	Non-Agency	1.00	6.30%×0.30×1.00=1.89%

GLM allows for more refined adjustments by modeling the interaction between variables, capturing the combined effects of different risk factors.

To highlight the differences between the two methods, let's examine the scenario of a vehicle insured for 50K to 75K, as depicted in Table XXI

Table XXI – Comparison of Premiums: Burning Cost vs. GLM

Policy Characteristics	Burning Cost	GLM
Base Rate		6.30%
Toyota		0.86
Sedan		1.2
Agency	3.54%	1.2
Sum Insured: 50K to 75K		0.8
Central		1.13
Rate as % of Sum Insured	3.54%	7.05%
Actual Sum Insured	65,330	65,330
Final Premium	**2,313**	**4,608**

In this example, the GLM model assigns a relativity factor to each variable based on its influence on the risk. In contrast, the Burning Cost method relies directly on claims data for specific conditions, such as 'Agency' repair type and the 'Sum Insured: 50K to 75K' range. However, the Burning Cost approach often lacks sufficient data when additional policy characteristics need to be incorporated.

As a result, the GLM model suggested a higher premium for this specific policy compared to the Burning Cost method. This is because GLM leverages a more comprehensive dataset, allowing it to account for the combined impact of all variables, even when individual data points are sparse.

One of the key advantages of GLM is its ability to account for interactions between variables. In cases where two variables interact to increase risk (such as high Sum Insured and repairs handled through an Agency), GLM captures these effects through its multiplicative modeling. Burning Cost, on the other hand, treats variables independently, potentially missing important interactions and leading to either overpricing or underpricing for specific combinations.

For example, in high-dimensional data with 60 million combinations, GLM can identify trends and interactions across different variables even when historical data is sparse for specific combinations. By modeling these interactions, GLM may deliver more nuanced premiums and can be particularly useful for complex datasets.

In this comparison, GLM demonstrates advantages over Burning Cost in handling complex, high-dimensional datasets. While Burning Cost can be effective for simpler, well-populated combinations, it fails to capture the nuances of interactions between risk factors. In contrast, GLM effectively models these interactions, leading to more accurate premiums. Additionally, GLM handles sparse data more effectively by modeling interactions among multiple risk factors, which is essential for modern insurance pricing models dealing with large datasets.

4. Conclusion

This project provided a comparison between Generalized Linear Models (GLM) and the Burning Cost method in the context of motor insurance pricing for Saudi Arabia's evolving Motor insurance market. Through this analysis, it is clear that GLM offers a more accurate, flexible, and data-driven approach, especially when dealing with complex datasets that involve multiple interacting variables, making it the superior choice for modern insurers.

The Burning Cost method, while straightforward and simple to implement, reveals significant limitations when applied to large datasets with numerous combinations of risk factors. Its reliance on historical claims data, without accounting for correlations between variables, often results in inaccurate pricing. The method tends to either overprice or underprice certain risk combinations, particularly when data is sparse or when interactions between factors, such as Sum Insured and Repair Condition, have a significant influence on the risk.

The GLM approach provides a robust and statistically rigorous framework for integrating multiple risk factors while capturing the interactions between them. This allows insurers to assess the combined impact of different variables and produce more nuanced and accurate pricing models. For example, GLM can recognize how a high Sum Insured combined with Agency repairs increases risk, a complexity that Burning Cost fails to adequately model. By applying multiplicative adjustments, GLM ensures that each variable's contribution to overall risk is fully reflected in the premium calculation, leading to more accurate pricing decisions. This level of accuracy is particularly vital in fast-moving and competitive markets such as Saudi Arabia, where mispricing can quickly erode profitability.

One of the key advantages of GLM lies in its ability to handle high-dimensional data effectively. With over 60 million potential combinations of risk factors in the dataset, GLM allows insurers to navigate this complexity by generating reliable premium estimates, even in cases where historical data for certain combinations is limited. This makes GLM an ideal solution for the modern insurance landscape, where granular risk segmentation is crucial for maintaining a competitive edge and accurately pricing policies.

Through the course of this project, several key insights have emerged. First, GLM demonstrates superior accuracy by accounting for the interactions between various risk factors, resulting in more precise premium estimates. This level of precision is especially important in Saudi Arabia's dynamic market, where rapidly shifting trends and emerging risks demand responsive and accurate pricing. Second, GLM proves its capacity to manage the complexity of large datasets by modeling and understanding the relationship between multiple variables, an ability that the Burning Cost method lacks. Finally, GLM is highly adaptable to changing market conditions and emerging trends, making it a critical tool for insurers looking to maintain profitability while offering fair premiums in an evolving landscape.

Given these insights, it is clear that Saudi insurers should prioritize the adoption of GLM as their primary pricing tool. The method's flexibility, precision, and ability to handle vast amounts of data will allow insurers to align their pricing strategies more closely with actual risk. To fully capitalize on the benefits of GLM, insurers should also invest in building their actuarial expertise. Developing, maintaining, and refining GLM models will require a solid foundation of actuarial knowledge and technical skills. In addition, insurers must focus on continuously improving the quality and granularity of their data collection processes, ensuring that their models are built on robust, comprehensive datasets. This will allow for even greater precision and adaptability in their pricing strategies.

In conclusion, transitioning to GLM for pricing in Saudi Arabia's motor insurance market will not only lead to more accurate and equitable premiums but will also enhance the transparency and sustainability of the industry as a whole. GLM's capacity to handle complex data and model the interactions between risk factors can help insurers align premiums more closely with underlying risks, enhancing competitiveness and supporting profitability. Adopting GLM could enable Saudi insurers to better adapt to current market conditions and prepare for future risks, contributing to long-term stability in a competitive marketplace.

References

Badri Management Consultancy (2020-2023). *Performance analysis of KSA insurance companies*. https://badriconsultancy.com/performance-analysis-of-ksa-insurance-companies/

De Jong, P., & Heller, G. Z. (2008). *Generalized Linear Models for Insurance Data*. Cambridge University Press.

Dobson, A. J., & Barnett, A. G. (2008). *An Introduction to Generalized Linear Models* (3rd ed.). Chapman & Hall/CRC.

Hosmer, D. W., & Lemeshow, S. (2000). *Applied Logistic Regression* (2nd ed.). Wiley.

Insurance Authority (2020-2023). *The Saudi Insurance Market Report*. https://ia.gov.sa/en/Reports/sector-report.html

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An Introduction to Statistical Learning: With Applications in R*. Springer.

Ohlsson, E., & Johansson, B. (2010). *Non-life Insurance Pricing with Generalized Linear Models*. Springer.

APPENDIX

TABLE A.I – FREQUENCY MODEL SELECTION

Factor	Null	Vehicle Age	Driver Age	Vehicle Sum Insured	Vehicle Make	Region	Year	Repair	Driver Nationality	Body Type	Color
Vehicle_Make_Mapped					Yes	Yes	Yes	Yes	Yes	Yes	Yes
Vehicle_Body_Type_Mapped										Yes	Yes
Vehicle_Color_Mapped											Yes
Vehicle_Sum_Insured				Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Vehicle_Repair_Condition								Yes	Yes	Yes	Yes
Driver_Nationality_Mapped									Yes	Yes	Yes
Vehicle_age		Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Driver_age			Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Region_Mapped_New						Yes	Yes	Yes	Yes	Yes	Yes
Year							Yes	Yes	Yes	Yes	Yes
Status	Solved!	Solved!	Solved!	Solved!	Solved!	Solved!	Solved!	Solved!	Solved!	Solved!	Solved!
Probability distribution	Poisson	Poisson	Poisson	Poisson	Poisson	Poisson	Poisson	Poisson	Poisson	Poisson	Poisson
Exclusions											
Option	Training/ Validation sets	Training/ Validation sets	Training/ Validation sets	Training/ Validation sets	Training/ Validation sets	Training/ Validation sets	Training/ Validation sets	Training/ Validation sets	Training/ Validation sets	Training/ Validation sets	Training/ Validation sets
Deviance	10,091	9,851	9,799	9,763	9,749	9,738	9,719	9,714	9,711	9,707	9,705
Deviance based pseudo-R^2	0.0%	2.4%	2.9%	3.2%	3.4%	3.5%	3.7%	3.7%	3.8%	3.8%	3.8%
AIC (smaller is better)	13,498	13,282	13,252	13,232	13,224	13,217	13,202	13,199	13,198	13,198	13,207
BIC (smaller is better)	13,506	13,384	13,440	13,483	13,498	13,507	13,507	13,513	13,519	13,535	13,583

TABLE A.II – SEVERITY MODEL SELECTION

Factor	Null	Vehicle Sum Insured	Year	Vehicle Age	Repair	Make Mapped	Driver Age	Region	Body Type	Color Mapped	Nationality
Vehicle_Make_Mapped						Yes	Yes	Yes	Yes	Yes	Yes
Vehicle_Body_Type_Mapped									Yes	Yes	Yes
Vehicle_Color_Mapped										Yes	Yes
Vehicle_Sum_Insured		Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Vehicle_Repair_Condition					Yes	Yes	Yes	Yes	Yes	Yes	Yes
Driver_Nationality_Mapped											Yes
Vehicle_age				Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Driver_age							Yes	Yes	Yes	Yes	Yes
Region_Mapped_New								Yes	Yes	Yes	Yes
Year			Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Status	Solved!	Solved!	Solved!	Solved!	Solved!	Solved!	Solved!	Solved!	Solved!	Solved!	Solved!
Probability distribution	Gamma	Gamma	Gamma	Gamma	Gamma	Gamma	Gamma	Gamma	Gamma	Gamma	Gamma
Exclusions											
Option	Training/ Validation sets	Training/ Validation sets	Training/ Validation sets	Training/ Validation sets	Training/ Validation sets	Training/ Validation sets	Training/ Validation sets	Training/ Validation sets	Training/ Validation sets	Training/ Validation sets	Training/ Validation sets
Deviance	2,224	2,016	1,943	1,899	1,883	1,887	1,864	1,856	1,849	1,843	1,843
Deviance based pseudo-R^2	0.0%	9.3%	12.6%	14.6%	15.3%	14.4%	15.5%	15.8%	16.2%	16.4%	16.4%
AIC (smaller is better)	(3,608)	(3,810)	(3,888)	(3,915)	(3,932)	(3,921)	(3,926)	(3,932)	(3,936)	(3,933)	(3,931)
BIC (smaller is better)	(3,597)	(3,755)	(3,821)	(3,781)	(3,792)	(3,765)	(3,709)	(3,703)	(3,696)	(3,665)	(3,658)

TABLE A.III – MODEL RESULTS

Factor	Modality	Value	Standard error	Lower conf. limit	Upper conf. limit	Pr > Chi-2	Multiplier	Exposure
(constant)	(constant)	(2.389)	0.204	(2.789)	(1.988)	0.000	0.092	
Vehicle_Make_Mapped	Hyundai	0.184	0.141	(0.092)	0.460	0.192	1.202	241,660
	Luxury	0.225	0.187	(0.141)	0.591	0.229	1.252	250,056
	Others	-	-	-	-	-	1	1,504,688
	TOYOTA	(0.151)	0.134	(0.414)	0.113	0.262	0.860	416,022
Vehicle_Body_Type_Mapped	Others	0.159	0.225	(0.283)	0.601	0.481	1.172	122,802
	SUVs - 4x4/ JEEP	-	-	-	-	-	1	1,262,562
	Sedans	0.075	0.101	(0.123)	0.274	0.456	1.078	1,027,062
Vehicle_Sum_Insured	[0; 25001[	0.038	0.170	(0.296)	0.372	0.825	1.038	291,734
	[25001; 50001[	-	-	-	-	-	1	520,760
	[50001; 75001[	(0.282)	0.149	(0.574)	0.010	0.058	0.754	408,806
	[75001; 100001[	(0.664)	0.165	(0.988)	(0.341)	0.000	0.515	427,688
	[100001; 200001[	(0.958)	0.175	(1.301)	(0.614)	0.000	0.384	495,534
	[200001; 300001[	(1.253)	0.270	(1.781)	(0.724)	0.000	0.286	170,310
	[300001; 500001[	(1.306)	0.436	(2.159)	(0.452)	0.003	0.271	61,582
	[500001; 1000001[	(1.938)	0.878	(3.660)	(0.217)	0.027	0.144	25,520
	[1000001; ∞ [	(2.348)	1.779	(5.835)	1.139	0.187	0.096	10,492
Vehicle_Repair_Condition	Agency	0.485	0.128	0.234	0.736	0.000	1.624	1,035,802
	Non-Agency	-	-	-	-	-	1	1,376,624
Driver_Nationality_Mapped	Others	(0.126)	0.137	(0.395)	0.143	0.360	0.882	425,700
	Saudi	-	-	-	-	-	1	1,986,726
Vehicle_Age	[-3004; 0[	(0.090)	0.207	(0.497)	0.316	0.663	0.914	131,556
	[0; 1[	-	-	-	-	-	1	475,076
	[1; 2[	(0.050)	0.161	(0.365)	0.265	0.756	0.951	283,538
	[2; 3[	(0.104)	0.185	(0.467)	0.258	0.572	0.901	204,418
	[3; 4[	(0.251)	0.213	(0.669)	0.166	0.238	0.778	165,476
	[4; 5[	(0.300)	0.240	(0.770)	0.171	0.212	0.741	137,232
	[5; 6[	(0.387)	0.241	(0.859)	0.084	0.108	0.679	157,026
	[6; 7[	(0.172)	0.230	(0.622)	0.279	0.456	0.842	153,566
	[7; 8[	(0.569)	0.261	(1.080)	(0.058)	0.029	0.566	156,436
	[8; 9[	(0.603)	0.291	(1.172)	(0.034)	0.038	0.547	123,458
	[9; 10[	(0.378)	0.296	(0.959)	0.203	0.202	0.685	92,814
	[10; 11[	(0.563)	0.339	(1.226)	0.101	0.097	0.570	77,712
	[11; ∞ [	(0.115)	0.227	(0.561)	0.330	0.612	0.891	254,118
Driver_Age	[-1; 16[	0.518	0.603	(0.664)	1.701	0.390	1.679	7,944
	[16; 21[	0.032	0.271	(0.499)	0.563	0.906	1.032	58,144
	[21; 25[	(0.020)	0.200	(0.412)	0.372	0.920	0.980	139,204
	[25; 30[	-	-	-	-	-	1	335,432
	[30; 35[	(0.122)	0.162	(0.439)	0.196	0.452	0.885	334,628
	[35; 39[	(0.140)	0.177	(0.487)	0.206	0.427	0.869	264,852
	[39; 43[	(0.392)	0.205	(0.794)	0.010	0.056	0.675	217,980
	[43; 46[	(0.226)	0.219	(0.656)	0.204	0.303	0.798	154,636
	[46; 51[	(0.176)	0.189	(0.547)	0.196	0.354	0.839	239,280
	[51; 56[	(0.193)	0.198	(0.581)	0.195	0.329	0.824	206,190
	[56; 60[	(0.411)	0.259	(0.919)	0.096	0.112	0.663	129,272
	[60; ∞ [	(0.381)	0.190	(0.754)	(0.008)	0.045	0.683	324,778
Region_Mapped_New	Central	0.147	0.122	(0.091)	0.386	0.227	1.159	469,424
	Eastern	-	-	-	-	-	1	1,078,052
	Others	(0.081)	0.154	(0.384)	0.222	0.601	0.922	864,950
Year	[2021; 2022[	(0.272)	0.186	(0.636)	0.092	0.143	0.762	589,296
	[2022; 2023[	(0.170)	0.120	(0.404)	0.065	0.156	0.844	601,958
	[2023; ∞ [	-	-	-	-	-	1	1,221,172
Deductible	[0; 500[	-	-	-	-	1	1	75,514
	[500; 750[	(0.051)	-	(0.051)	(0.051)	1	0.950	529,076
	[750; 1000[	(0.073)	-	(0.073)	(0.073)	1	0.930	365,562
	[1000; 1500[	(0.105)	-	(0.105)	(0.105)	1	0.900	944,102
	[1500; 2000[	(0.151)	-	(0.151)	(0.151)	1	0.860	238,520
	[2000; 3000[	(0.211)	-	(0.211)	(0.211)	1	0.810	171,168
	[3000; 5000[	(0.301)	-	(0.301)	(0.301)	1	0.740	23,620
	[5000; 10000[	(0.431)	-	(0.431)	(0.431)	1	0.650	49,064
	[10000; ∞ [	(0.693)	-	(0.693)	(0.693)	1	0.500	15,800
Driver_NCD_Years		-	-	-	-	1	1	1,236,116
	0	-	-	-	-	1	1	272,686
	1	(0.163)	-	(0.163)	(0.163)	1	0.850	164,010
	2	(0.288)	-	(0.288)	(0.288)	1	0.750	94,702
	3	(0.431)	-	(0.431)	(0.431)	1	0.650	66,630
	4	(0.693)	-	(0.693)	(0.693)	1	0.500	48,024
	5	(0.916)	-	(0.916)	(0.916)	1	0.400	67,510
	11	-	-	-	-	1	1	265,058
	12	(0.163)	-	(0.163)	(0.163)	1	0.850	109,762
	13	-	-	-	-	1	1	44,688
	14	-	-	-	-	1	1	36,292
	15	-	-	-	-	1	1	6,948