Determinants of Claims in Motor Insurance in Rwanda

Dr.Placide Aime Kwizera*

Steven Gakuru †

May 2024

Abstract

The paper examines the determinants of claims in the motor insurance sector in Rwanda. We utilised the adaptable Tweedie model alongside XGBoost to analyse and uncover key predictors influencing claim behaviour. Initial findings from the XGBoost analysis emphasised the significance of variables such as the vehicle's manufacturing year, policyholder, and coverage type in shaping claim probabilities. Subsequent Tweedie regression analysis provided deeper insights, revealing that institutional policyholders are more likely to file claims compared to individual policyholders. Moreover, comprehensive insurance coverage showed a notably higher claim probability than third-party coverage. Interestingly, larger cars exhibited a slightly lower likelihood of claims, while older vehicles and rainy seasons corresponded to increased claim probabilities. Furthermore, an increase in spare parts expenditure is associated with a marginal reduction in claim likelihood, possibly indicating prudent driving behaviour. Conversely, increased maintenance expenditure correlated with a higher probability of claims. Additionally, an increase in GDP was associated with a slight decrease in claim probability, potentially reflecting broader economic conditions influencing driving patterns, vehicle usage, or safety measures investment, thereby impacting claim behaviour. Empirical implications from the findings suggest that insurers could leverage insights from our analysis, such as vehicle manufacturing year, policyholder, and coverage type, to refine pricing and risk assessment strategies. Additionally, promoting proactive maintenance practices among policyholders could help reduce claim frequency and severity. On the regulatory side, the National Bank of Rwanda should foster a regulatory environment conducive to innovation and data-driven decision-making while ensuring compliance with standards and monitoring claims management processes for transparency and accountability. These findings offer valuable insights for sector stakeholders, guiding strategies to enhance industry stability and efficiency in Rwanda's evolving insurance landscape.

Keywords: Motor insurance Claims,Tweedie regression, XGBoost

JEL Classification: G22, G28, D81

*Principal Economist, Research Department, National Bank of Rwanda and corresponding author (pkwizera@bnr.rw).
†Senior Analyst, Financial Stability Monitoring Department, National Bank of Rwanda.

1 Introduction

The insurance industry has historically been growing and has played an important role in ensuring the country's economic well-being. Insurance claims are a significant and costly problem for insurance companies in all insurance industry sectors. The motor insurance is one of the most important branches of insurance, which has a direct impact on the life of the general public all over the world (Prinja, Kaur, & Kumar, 2012). Motor insurance had its first phase in the United Kingdom in the beginning part of the last century. The first motor car was introduced in England in 1894. The earliest motor insurance policy was established in the year 1895 to cover third-party liabilities(Bashir, Madhavaiah, & Naik, 2013).

However, by 1899, accidental damage to the car was added to the policy. In this way, it initiated a more comprehensive policy along the policies issued today. In 1903, General Insurance Corporation Ltd was initiated primarily to manage motor insurance, followed by other companies or organisations. After the First World War (World War I), there was a significant increase in the number of vehicles on the road, which led to an increase in the number of accidents. Many injured persons in road accidents could not recover losses since motor vehicles were not insured. This resulted in the establishment of mandatory third-party insurance by passing the Road Traffic Act of 1930 and 1934. The obligatory insurance provisions of these Acts have been consolidated by the Road Traffic Act 1960 (Ariff & Sirajuddin, 2016). Auto insurance protects vehicle owners against damages to their vehicles and pays for any third-party liability determined as per rules against the vehicle owner. Third Party Insurance is a legislative requirement. The vehicle's owner is lawfully liable for any injury or damage to third-party life or property caused by or arising from using the vehicle in public locations. Although Motor Vehicle Insurance is also considered an area that suffers the largest loss, it is also recognised as the major source of premium earnings for insurance companies (Mathur & Tripathi, 2014). Insurance companies typically employ a claims investigation unit to investigate the factors affecting claims.

The insurance market in Sub-Saharan Africa, particularly in motor insurance, has experienced significant advancements alongside challenges. Factors such as urbanization and improved infrastructure have spurred demand, prompting insurers to diversify products and utilize technology for broader outreach and operational efficiency (Signé & Johnson, 2020). Nevertheless, issues such as fraud and regulatory compliance remain prevalent. Meanwhile, Rwandan insurance companies are contending with a scenario where they are settling claims from their clients. Ongoing discussions revolve around whether the increase in car accidents involving motor vehicles in Rwanda has affected the industry's overall profit margins. With sustained economic growth in Rwanda, the number of vehicles on the roads tends to rise as individuals and businesses enjoy greater purchasing power. This surge in vehicle ownership directly contributes to heightened traffic volume and congestion, consequently increasing the likelihood of accidents occurring. Concurrently, the trend of motor claims has gained momentum in recent years, a trajectory that many experts anticipate will persist into the future. While most insurance companies concentrate on sales as the primary source of revenue, motor insurance will be the backbone of insurance market expansion as its underperformance significantly affects the profitability of individual insurance and the entire industry.

In addition, the increase in claims will have long-term implications not only on the investments and risk outlook of the sector but also for the players' operations. The National Bank of Rwanda reported steady growth in the insurance sector, with the industry's asset base increasing by 16.9 percent to FRW 963.2 billion in December 2023 from FRW 824 billion in December 2022. Meanwhile, motor insurance claims statistics in Rwanda have shown a notable impact on the sector, with a high claims ratio of 63 Percent in December

2023, indicating that 63 Percent of net earned premiums were paid out as claims. Given the importance of claims in driving operational excellence and shaping public perception, it is crucial to investigate the factors behind this increase. Policymakers need to understand these factors to implement necessary measures and guidance to sustain the market's development.

Against this background, this paper seeks to answer the following question: What are the key determinants influencing motor insurance claims in Rwanda, and how can this knowledge inform policy-making decisions aimed at improving the sustainability and efficiency of the insurance industry?. To address the above research questions, this study employs the Tweedie model, which has received significant attention in actuarial literature (Jørgensen & Paes De Souza, 1994) and further elaborated upon by Dunn and Smyth (2001),(Smyth & Jørgensen, 2002), and Hasan and Dunn (2011). Its beauty lies in its remarkable flexibility, seamlessly accommodating various data types, including zero-inflated datasets commonly encountered in insurance claims. Moreover, with easily interpretable parameters, such as regression coefficients, the Tweedie model offers valuable insights into the drivers of claim amounts, empowering insurers to make informed decisions in pricing and risk assessment. With this process, we first use the XGBoost to provide insights into predictors features important to enhance assurance of Tweedie's findings. All of these models could predict the likelihood of factors contributing to the increase in motor insurance claims.

To the best of our knowledge, this is the unique and first research to examine the determinants of claim motor insurance in Rwanda using the above techniques. The rest of the paper is as follows: Section 2 provides a review of the relevant literature, Section3 presents Rwanda's insurance sector landscape, Section 4 indicates the methodology and data, Section 5 shows the empirical results, and, finally, Section 6 concludes.

2 Literature Review

Theoretical literature on the determinants of claims in motor insurance encompasses various models and frameworks that seek to explain the factors influencing claim behavior. One prominent theoretical approach is the utility maximization theory, which posits that individuals make rational decisions based on maximising their utility or satisfaction (Friedman & Savage, 1948). Applied to motor insurance, this theory suggests that policyholders weigh the benefits of filing a claim (such as financial compensation for damages) against the costs (such as increased premiums or deductibles) to determine whether it is advantageous to file a claim, see the study of Dionne and Rothschild (2014).

Another theoretical perspective is the risk perception and risk aversion theory, which suggests that individuals' perceptions of risk and their aversion to risk influence their decision-making regarding insurance claims (Kunreuther, Pauly, et al., 2006). For example, policyholders who perceive a higher risk of accidents or damage may be more inclined to file claims to mitigate potential losses, while those who are risk-averse may be more cautious in filing claims to avoid potential increases in premiums.

Additionally, behavioural economics theories, such as prospect theory and loss aversion, highlight the role of cognitive biases and psychological factors in shaping individuals' decisions about insurance claims (Kahneman & Tversky, 1979; Tversky & Kahneman, 1992). Prospect theory suggests that individuals are more sensitive to potential losses than gains, leading them to weigh the potential outcomes of filing a claim differently.

Loss aversion theory posits that individuals are more averse to losses than they are motivated by equivalent gains, influencing their willingness to file claims based on perceived losses; see more the recent study by Einav, Finkelstein, Ryan, Schrimpf, and Cullen (2013). Furthermore, some contemporary studies applying agency theory have examined the dynamics between policyholders and insurance companies, focusing on issues of asymmetric information and moral hazard, suggesting that policyholders may engage in opportunistic behaviour, such as exaggerating claims, while insurance companies implement measures such as risk-based pricing and claims investigations to mitigate moral hazard and adverse selection issues.

However, on the empirical front, especially the auto-insurance business, is not only a fast-growing area that needs actuarial input (Haiss & Sümegi, 2008), but it also offers an exciting challenge since it manages a large number of scenarios involving different types of risks. The major aim of an insurance company is to compute a premium that correctly covers the type of risk an insured is insured against. The price should factor in the attributes of the customer because these attributes play a big role in classifying one's riskiness. The frequency of claims plays a vital role in the calculation of premiums; it is of prime importance to develop models that correctly describe the evolution of the frequency of claims, also called count data. Since no two persons have exactly similar attributes, this implies that models that incorporate risk factors should be considered to develop a suitable model that correctly estimates the frequency of claims. According to Boucher, Denuit, and Guillén (2008), regression analysis of count data allows the identification of the explanatory variables and the estimation of the expected number of claims conditional on the individual characteristics of the policyholder.

According to Mossin (1968), insurance increases welfare by transferring uncertainty from risk-averse individuals to risk-neutral insurers, which pool together a wide range of risks for better risk management. On the one hand, theoretical models of non-life insurance demand (Arrow et al., 1974; Mossin, 1968; Pratt, 1978) forecast increasing demand for higher levels of risk aversion (proxied by education or age structure of the population), probability of loss (proxied by urbanization or population density) and total wealth (proxied by income). On the other hand, other scholars like (Millo & Pasini, 2010) focus on non-life insurance demand, primarily from an empirical perspective. Overall, their models include a wide range of explanatory variables such as economic, socio-demographic, institutional, and/or behavioral factors, while the demand for non-life insurance is expressed mainly through insurance density or penetration. Moreover, the landmark study of (Feyen, Lester, & Rocha, 2011) targets the effect of four categories of explanatory factors (economic, demographic, sociocultural, and institutional/market structure determinants) on both life and non-life insurance for a broad sample of 90 countries during the period 2000-2008.

More recently,(Born & Bujakowski, 2022) examined the development of property-casualty and life health insurance markets in 18 post-communist European countries in 2008-2017, using the dynamic generalised method of moment(GMM) approach. First, their results show that the factors having the highest impact on motor hull insurance are as follows: GDP, domestic credit, size of the agricultural sector, and insurance market concentration. Second, for motor third-party liability insurance, the factors with higher impact seem to be education, urbanisation, foreign insurance penetration, and the size of the agricultural sector. On the relationship between urbanisation and automobile ownership, (Melia, Chatterjee, & Stokes, 2018) conclude that urban areas discourage car use. Thus, people living in urban areas perceive a higher risk of car accidents or theft, generating an increased demand for motor insurance (Sherden, 1984). Besides, (Browne & Hoyt, 2000), using urbanisation as a proxy of loss probability, shows that an increase in loss probability generates an increase in non-life insurance demand. (Esho, Kirievsky, Ward, & Zurbruegg, 2004) associate urbanisation with increased delinquency and conclude that in urban areas, additional sources of security are needed (i.e. security that can be obtained through property-casualty demand).

Overall, the literature targeting the determinants of non-life insurance consumption emphasises their different effects in high-income countries compared to emerging ones. Since the main component of non-life insurance is represented by motor insurance in emerging countries, we are particularly interested in the results obtained for these states. Dragos (2014) shows that in 10 emerging CEE countries, over the period 2001-2011, the non-life insurance demand is positively influenced by income, education and urbanisation, while it is negatively affected by income distribution. Likewise,Petkovski, Jordan, et al. (2014) examine the drivers of the non-life insurance market in 16 countries from Central and South-Eastern Europe over a period of twenty years (i.e. 1992-2011) and find a positive and significant relationship between the rule of law, population density, number of passenger cars per 1000 people, GDP per capita and the non-life insurance demand.

According to reports from the Faculty of Claims of the Chartered Insurance Institute, in recent years, there have been many changes in the way claims are handled, seen, and managed in Europe. These changes are incremental and cumulative rather than sudden and dramatic, creating an environment of constantly changing and evolving claims. It is agreed that claims are much closer to the heart of the industry than ever before. It is believed that many cases are the biggest trigger for the profits and losses of an organisation. In the study done by Petkovski et al. (2014) using linear country-specific fixed effects panel data regression model with common coefficients across all cross-section members of the pool, recommended that insurance companies strive to maintain healthy relationships with their clients to improve their overall performance by reducing risks. One of the primary areas to observe and promote such a healthy relationship is ensuring insurance companies can prudently observe the five key areas of claims management.

A study by Epetimehin and Odunaike (2002) emphasised that while a company's financial performance uses ratios, one of the key indicators to prefer in return on Assets was that anything below 5 percent is unsafe. Where return on equity and return on investment, anything within 10 percent and 14 percent are considered desirable. The term investment may refer to total assets or net assets. The funds employed in net assets are known as capital employed. Net assets equal net fixed assets plus current assets minus current liabilities, excluding bank loans. African Insurance Organizations (AIO), which controls about $69 billion in insurance markets, released its first Africa Insurance Barometer to improve transparency. The report pointed out the main driver behind African insurance companies' poor results is claims management practices. For instance, motor insurance has been rated the most frequently least profitable line of business due to the frequent competitiveness of motor insurance. This has led to insurance companies registering low levels of profitability and high claims inflation (Bilal et al., 2019).

From the reviewed literature, several variables have been identified as significant determinants of claims behaviour in the insurance industry. Premium amounts play a crucial role, influencing policyholders' behavior and their propensity to file claims. Additionally, policyholder characteristics such as age, gender, and driving history impact claim likelihood. The age and type of insured vehicles also contribute to claim frequencies, with older vehicles and certain types associated with higher probabilities of claims. Furthermore, the extent of policy coverage, whether comprehensive or third-party, affects claim behaviour, as does seasonality, with weather conditions influencing accident rates. Fluctuations in inflation rates and broader economic conditions also have implications for driving patterns and claim frequencies.

Therefore, the present study will investigate the significance of these variables that are scattered throughout the existing literature.

3 Insurance sector landscape in Rwanda

This section will present statistics on the insurance asset base, the share of assets in the total financial sector, product descriptions within the insurance sector, and the value of claims, along with the claims ratio, to describe how this sector has performed and provide stylized facts.

The insurance sector in Rwanda plays a vital role in the country's economy by offering financial protection to individuals and businesses against various risks. Despite the fact that it is still in the infant stage compared to peers in other countries, the sector has experienced significant development since its inception in the early 1990s.

Regulatory entities like the National Bank of Rwanda (NBR) and the Association of Rwandan Insurers (ASSAR) have played crucial roles in overseeing and regulating insurance activities within the country. However, the sector faces significant challenges, including low insurance penetration and awareness among the populace. This lack of understanding of insurance benefits has impeded sector growth and limited the adoption of insurance products and services. Furthermore, limited product diversity and distribution channels pose additional obstacles, with most offerings focused on conventional lines like motor, health, and property insurance. Despite these challenges, Rwanda's insurance sector holds promise for expansion and advancement. The country's robust economic performance, burgeoning middle class, and urbanisation trends provide fertile ground for market growth. Moreover, governmental efforts to promote financial inclusion and enhance regulatory frameworks signal positive prospects for the sector's future trajectory.

The private insurance sector in Rwanda is comprised of two main types: non-life insurance and life insurance. Additionally, it encompasses four special categories: microinsurance, captive insurance, health maintenance organisations (HMOs), and mutual insurance. Currently, there are 15 insurance companies operating in the private sector, with 9 offering non-life insurance products, 3 offering life insurance, and 1 each for microinsurance, captive insurance, HMO, and mutual insurance.

The sector has demonstrated consistent growth, evidenced by a 16.9 percent increase in the industry's asset base, rising from FRW 824 billion in December 2022 to FRW 963.2 billion in December 2023, as illustrated in the figure below:

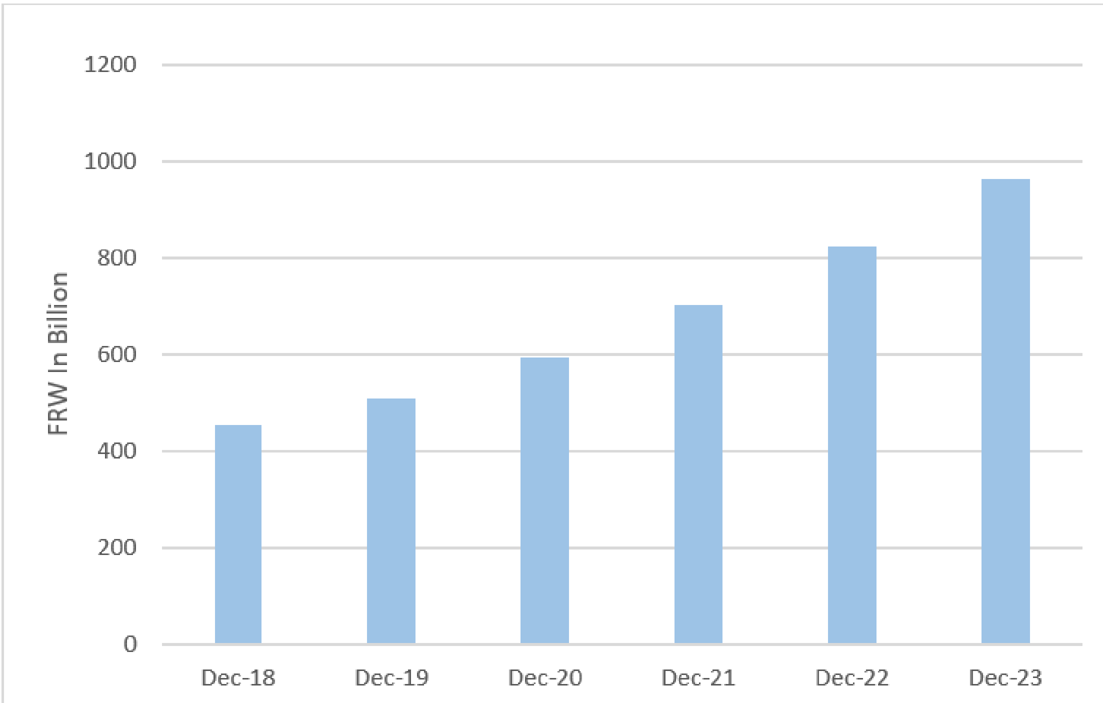

Figure 1: **Insurance asset base**
Source:National Bank of Rwanda,2023

In the last ten years, the insurance sector has experienced significant growth, with Gross Written Premiums (GWP) increasing by more than 300 percent. The Insurance Penetration Ratio, which measures the level of insurance coverage within an economy, has also seen growth, rising from 1.45 percent to 1.98 percent as insurance premiums compared to the Gross Domestic Product (GDP) of the country. Additionally, Insurance density is a financial metric that gauges the average amount of insurance premium paid per person within a specific population or geographical area. It offers valuable insights into the average level of insurance coverage among individuals or households in that area. In the last ten years, the insurance density in Rwanda increased from 8,110 RWF to 21,000 RWF in December 2023.

The financial sector in Rwanda continues to play a vital role in financing the economy, with ongoing deepening evident in its assets, which accounted for 66.7 percent of GDP as of December 2023. The insurance sector remains relatively small, ranking third in terms of asset size within the financial sector, following banking and pension. As of December 2023, its share of assets in the total financial sector was 9 percent, as illustrated in the figure below.

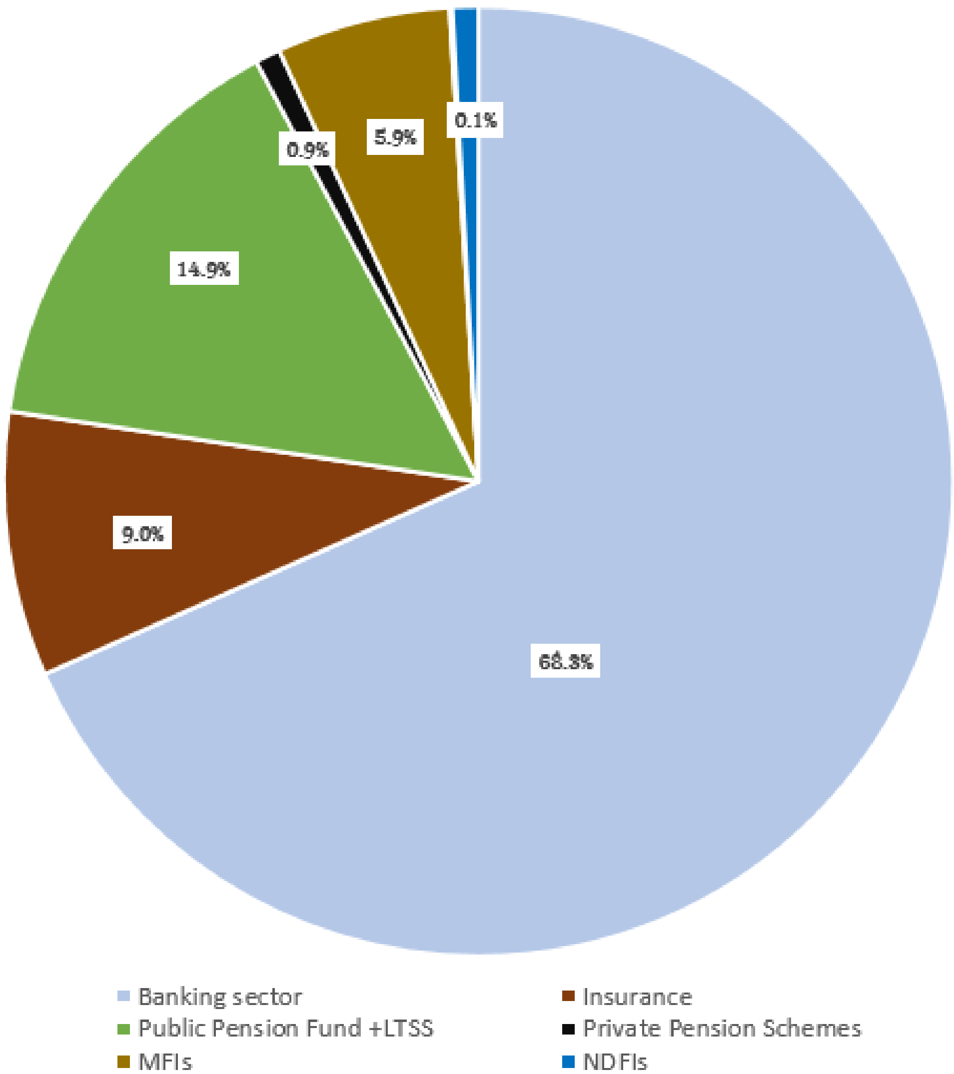

Figure 2: **Share of insurance asset financial sector**
Source:National Bank of Rwanda,2023

General insurance (Non-life) business, classified as short term, continued to be the largest contributor to private insurance premiums, accounting for 78.8 percent of private insurers' premiums and 46.8 percent of the sector's total Gross Written Premiums (GWP) as of December 2023. The general insurance business is predominantly led by motor and medical insurance classes, which together represent 63 percent of total gross premiums and 29 percent of the total premiums in the insurance industry.

Product	GWP	Net earned premium	Claims	Claims ratio
Motor	53,608	44,687	28,062	63%
Property	19,899	6,036	1,055	17%
Liability	4,425	1,649	540	33%
Transportation	1,549	659	73	11%
Accident & Health	3,627	1,836	404	22%
Engineering	11,611	2,294	508	22%
Guarantee	5,698	1,903	347	18%
Medical	32,847	26,363	16,593	63%
Miscellaneous	6,366	2,418	624	26%

Source:National Bank of Rwanda,2023

Table 1: **Product Description (In FRW Million) in Dec-23**

Claims in value in private insurance have been growing at an average rate of 12 percent every year since 2015, reaching 67.8 billion in December 2023 from 28.4 billion in 2015.

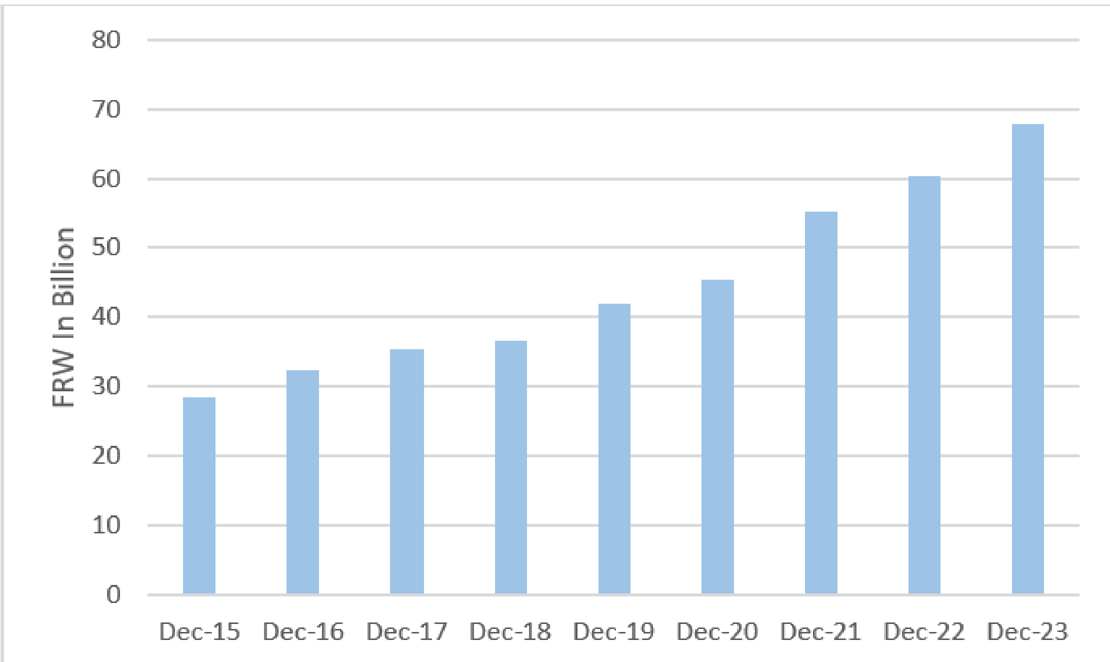

Figure 3: **Claims in value in private insurance**

Source:National Bank of Rwanda,2023

The claims ratio, which indicates the ratio of claims incurred to net premium earned, has averaged 64 percent over the last 9 years. This percentage has been primarily driven by the motor and medical insurance categories in non-life private insurance.

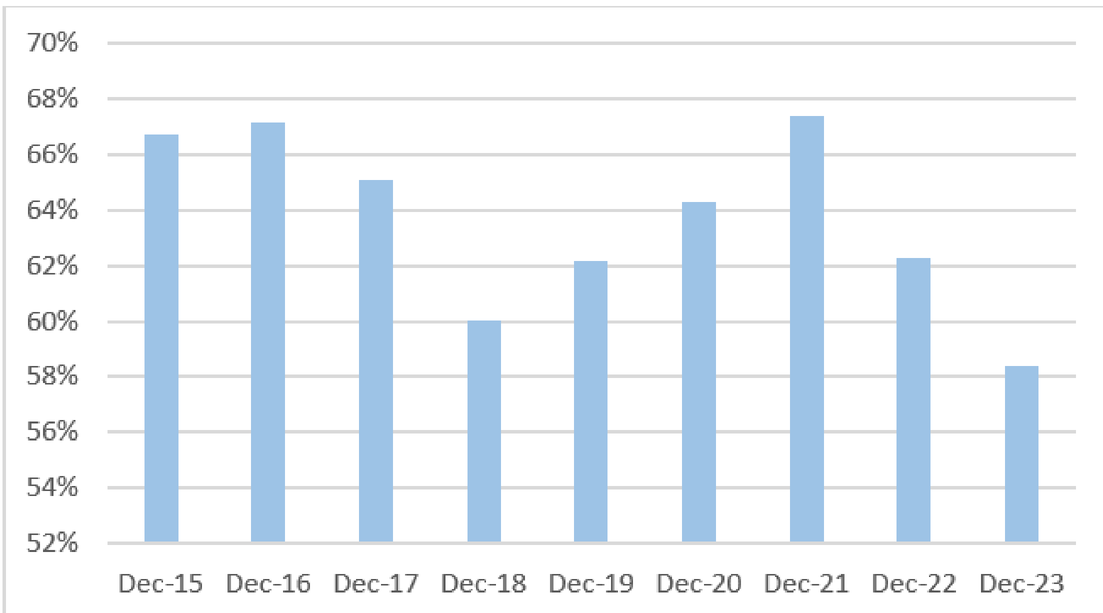

Figure 4: **Claims ratio in private insurance**
Source:National Bank of Rwanda,2023

4 Methodology and Data

In this section, we discuss the empirical strategy used to assess the determinants of claims in motor insurance. First, we describe the data used in the current study and define the practical model and the variables chosen.

4.1 Econometric approach

This study uses the Tweedie model, a generalised linear model (GLM) specifically designed to handle data with a Tweedie distribution, which is common in insurance claims data and other types of data with non-negative continuous values and excess zeros(Jørgensen & Paes De Souza, 1994; Jorgensen, 1997). It models the mean-variance relationship inherent in such data, making it suitable for empirical applications. Before Tweedie model representation, we need to present the functional form of the XGBoost feature importance:

$$\text{Importance}(f) = \sum_{t=1}^{T} \frac{\text{Gain}(f,t)}{\text{Total Gain}_t} \tag{1}$$

Where: Importance(f) is the importance score of feature f. Gain(f,t) is the improvement in the split criterion (e.g., mean squared error) brought by feature f at tree t.Total Gain$_t$ is the total improvement in the split criterion achieved by all features at tree t. By integrating XGBoost feature importance analysis into our Tweedie regression framework, we can enhance our understanding of the underlying factors driving claim variability and improve the accuracy of our empirical analysis.

This equation quantifies the contribution of each feature to the overall model performance and helps identify which features have the most significant impact on the regression task.

The Tweedie regression models were initially introduced by Jørgensen and Paes De Souza (1994) and further elaborated upon by Dunn and Smyth (2001) and Hasan and Dunn (2011), along with other researchers. These models consider independent responses $Y_1, Y_2, \ldots, Y_n$, which are observed such that

$$Y_{it} \sim \mathrm{Twp}(\mu_{it}, \phi) \tag{2}$$

where μ_{it} represents the mean response for observation i at time t, and ϕ denotes the dispersion parameter. The mean μ_{it} is linked to the linear predictor through a known link function g:

$$g(\mu_{it}) = x_{it}^T \beta$$

where x_{it} is a vector of covariates, and β is a vector of unknown regression parameters. Let q denote the dimension of β. Alternatively, we can represent the model using matrix notation. Let Y be a vector of response variables; then, the Tweedie regression model can be defined by

$$Y \sim \mathrm{Twp}(\mu, \phi I) \tag{3}$$

where I is an $n \times n$ dimensional identity matrix. In this case, we have $\mathrm{E}(Y) = \mu = g^{-1}(X\beta)$ and $\mathrm{Var}(Y) = C = \mathrm{diag}(\phi\mu^p)$. Here, the link function g is defined as the logarithm function. Note that the model can be equivalently defined by its joint distribution or by its first two moments (mean and variance). We denote the vector of parameters by $\theta = (\beta, \lambda = (\phi, p))$, where the parameter vector can be divided into two sets: the first comprises the regression parameters, and the second comprises parameters that describe the variance structure.

An empirical specification of this model can take the following form:

$$\begin{aligned}\mathrm{Claims}_{it} = &\alpha + \beta \mathrm{Prem}_{it} + \varpi \mathrm{Manuf\_D}_{it} + \delta \mathrm{V\_type}_{it} + \theta \mathrm{P\_holder}_{it} + \phi \mathrm{P\_cover}_{it} + \omega \mathrm{Season}_t + \vartheta \mathrm{Spare}_t \\ &+ \gamma \mathrm{Maint}_t + \psi \mathrm{GDP}_t + \varepsilon_{it}\end{aligned} \tag{4}$$

Where $\log(Claims_{it})$ represents the logarithm of the amount of the claim, the response variable. α, β, ϖ, δ, θ, ϕ, ω, ϑ, γ, ψ are the coefficients corresponding to the predictor variables Prem_{it}, $\mathrm{Manuf\_D}_{it}$, $\mathrm{V\_type}_{it}$, $\mathrm{P\_holder}_{it}$, $\mathrm{P\_cover}_{it}$, Season_t, Spare_t, Maint_t, GDP_t respectively. Finally, ε_{it} represents the error term.

4.2 Data structure

This study uses the official data from the data warehouse reported by insurance companies. The insurers are required to provide monthly cumulative reports to the National Bank of Rwanda. We have utilised data spanning from 2019 to 2023 for the four insurance companies to maintain a balanced panel dataset.

Variable	Definition	Expected relation	Type of the factor
Dependent variable			
Claims	Total claims	N/A	N/A
Independent variables			
Prem	Premium	Positive	Micro determinant
Manuf_D	Manufacturing year below 2010 taking value of 1 and 0 otherwise	Positive	Micro determinant
V_type	Classified large and small, taking value of 1 and 0 otherwise	Negative	Micro determinant
P_holder	Policyholder: institution and the individual taking the value of 1 and 0, otherwise	Positive	Micro determinant
P_cover	Policy cover: comprehensive and third-party taking the value of 1 and 0, otherwise	Positive	Micro determinant
Season	Season: rainy and dry, taking the value of 1 and 0 otherwise	Positive	Macro determinant
Spare	Spare part index from CPI	Positive	Macro determinant
Maint	Maintenance index from CPI	Positive	Macro determinant
GDP	Gross domestic product	Negative	Macro determinant

Source:Authors' literature survey

Table 2: **List of used variables**

5 Results and Discussion

5.1 Preliminary findings

Before delving into regression analysis, we first understand the dataset's characteristics and feature importance analysis to identify key predictors that significantly influence outcomes.

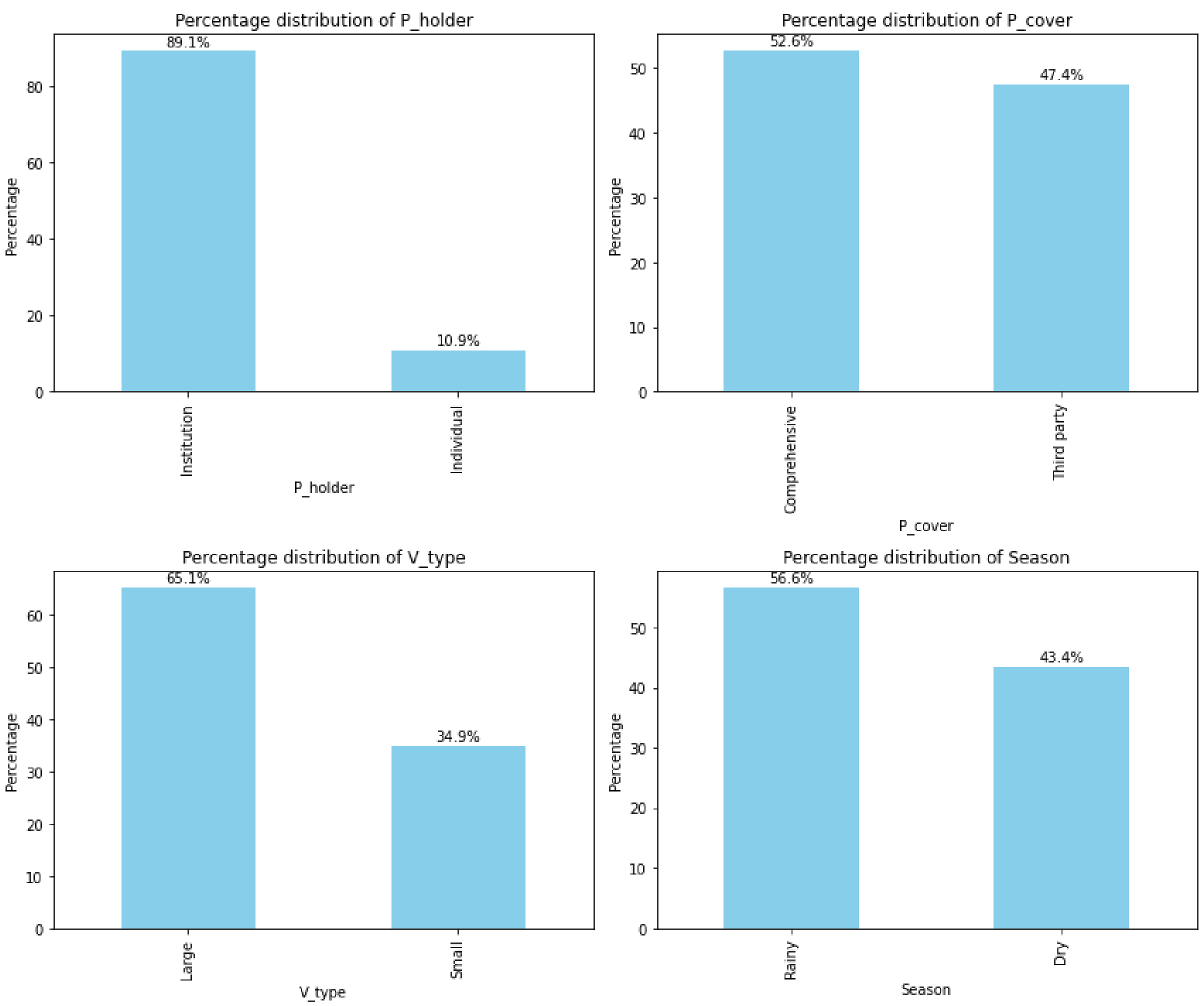

Figure 5: **Distribution of micro determinants**
Source: Authors computation

The preliminary descriptive analysis provides valuable insights into the composition of categorical variables within the dataset. Regarding policyholder, the majority (89.1 percent) are institutions, with individual policyholders accounting for a smaller proportion (10.9 percent). In terms of coverage type, comprehensive coverage is slightly more prevalent, representing 52.6 percent of observations, compared to 47.4 percent for third-party coverage. Moreover, the dataset is characterised by a higher proportion of large cars (65.1 percent) compared to small cars (34.9 percent). Finally, with regard to seasonality, 56.6 percent of observations occur during the rainy season, suggesting that the dataset may be skewed towards data collected during this period.

Before running regression, we first use XGBoost to capture the "feature importance" to know which features of the model is relied on the most to make accurate predictions.

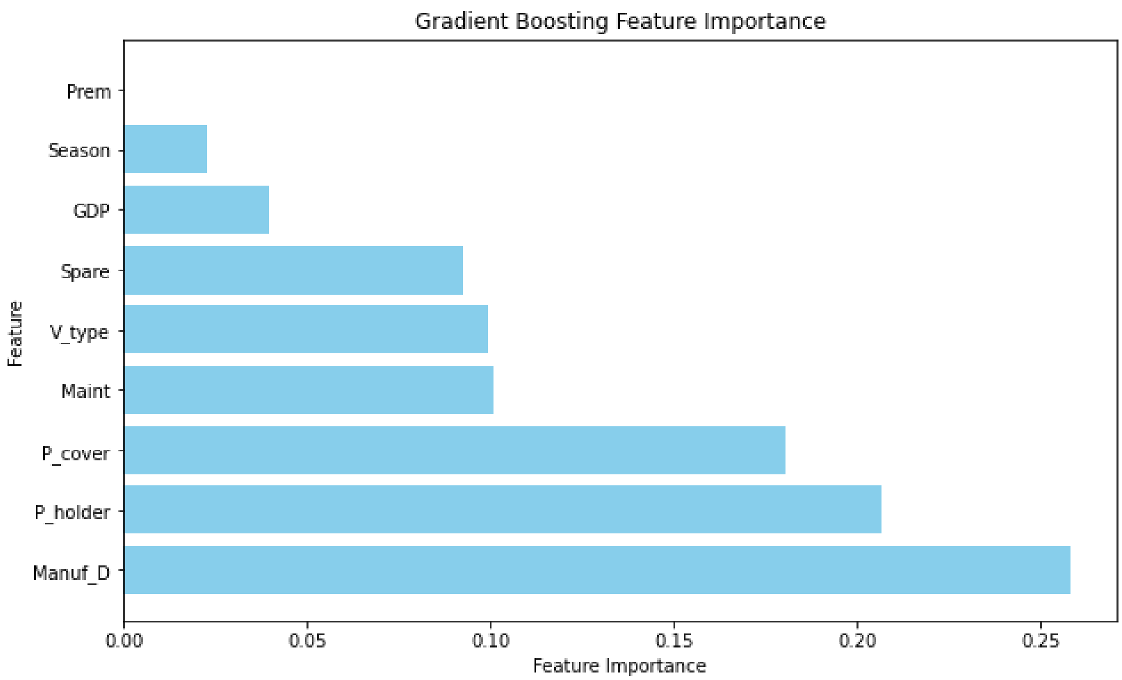

Figure 6: **Feature importance metrics**
Source: Authors computation

The figure for gradient boosting feature importance indicates that the variable with the highest importance is Manuf_D, meaning that the manufacturing year of the vehicle has the most significant impact on the predictive power of the model that we are going to estimate. The next important variable is P_holder, suggesting that whether the policyholder is an institution or an individual significantly influences the prediction of outcomes. P_cover, representing the type of coverage (comprehensive or third-party), indicates its substantial contribution to the model's predictive accuracy. Maintenance (Maint) comes next in importance, meaning that the level of maintenance performed on the vehicle plays a crucial role in predicting outcomes.

V_type, which denotes the type of vehicle (large or small), is ranked fifth in importance, suggesting its moderate influence on the model predictions. Spare parts expenditure (Spare) follows, indicating its relatively lower significance than the previous variables. GDP, representing the economic conditions, holds the second-last position in importance, suggesting a lesser impact on the model's predictive power. Lastly, season represents the seasonality factor, and prem, the premium amount, is ranked as the least important variable in predicting outcomes, indicating their minimal contribution to the model's predictive accuracy. Overall, this ranking provides valuable insights into the relative importance of different variables in predicting outcomes in the context of the gradient-boosting model, providing the basis for the ensuing regression analysis.

5.2 Main findings

Dep. Variable:	Logclaims	**No. Observations:**	43143
Model:	GLM	**Df Residuals:**	43134
Model Family:	Tweedie	**Df Model:**	8
Link Function:	log	**Scale:**	0.093156
Method:	IRLS	**Log-Likelihood:**	nan
		Deviance:	6113.0
		Pearson chi2:	4.02e+03
No. Iterations:	10		

	coef	**std err**	**z**	**P> |z|**	**[0.025**	**0.975]**
P_holder	0.0429	0.003	13.120	0.000	0.036	0.049
Prem	0.0646	0.201	0.322	0.747	-0.329	0.458
P_cover	0.0983	0.002	40.389	0.000	0.094	0.103
V_type	-0.0296	0.002	-14.023	0.000	-0.034	-0.025
Manuf_D	0.0697	0.002	28.347	0.000	0.065	0.074
Season	0.0319	0.002	16.625	0.000	0.028	0.036
Spare	-0.0011	0.000	-2.669	0.008	-0.002	-0.000
Maint	0.0170	0.001	31.852	0.000	0.016	0.018
GDP	-0.0045	0.000	-24.814	0.000	-0.005	-0.004

Source: Authors computation

Table 3: **Tweedie Model Result**

The findings indicate that the coefficient (0.0429) for P_holder suggests that institutions, compared to individual policyholders, exhibit a 4.29 percent higher probability of filing insurance claims. This may stem from the larger scale of operations or different risk profiles associated with institutional policyholders, potentially reflecting higher exposure to risks that lead to claims. Interestingly, the premium (0.0646) does not demonstrate a statistically significant influence on the probability of claims. This implies that variations in premium rates might not directly correlate with changes in the propensity to file claims. Vehicles covered by comprehensive insurance policies show a notably higher claim probability (0.0983) compared to those with third-party coverage. This finding underscores the broader coverage scope and potentially more extensive protection provided by comprehensive insurance policies, leading to a higher likelihood of claims being filed.

The negative coefficient (-0.0296) associated with larger cars suggests that they exhibit a slightly lower likelihood of filing claims compared to smaller cars. This could be attributed to factors such as enhanced safety features or driving behavior associated with larger vehicle types, resulting in a reduced probability of accidents or incidents leading to claims. Cars manufactured before 2010 display a higher claims likelihood (0.0697) compared to those manufactured after 2010. This may indicate that older vehicles are more prone to wear and tear, mechanical failures, or other issues that necessitate insurance claims, reflecting the impact of vehicle age on claim behaviour.

During the rainy season, claims probability increases by approximately 0.0319 compared to the dry season. This finding aligns with expectations, as adverse weather conditions during the rainy season can heighten the risk of accidents, vehicle damage, and subsequent insurance claims. Furthermore, an increase in spare parts expenditure was associated with a marginal reduction in claim likelihood, possibly indicating prudent driving behaviour or fewer incidents requiring repairs. Conversely, increased maintenance expenditure correlated with a higher probability of claims. Additionally, a one-percentage increase in GDP was associated with a slight decrease in claim probability, potentially reflecting broader economic conditions influencing driving patterns, vehicle usage, or safety measures investment, thereby impacting claim behaviour.

Dep. Variable:	Logclaims	**No. Observations:**	43143
Model:	GLM	**Df Residuals:**	43135
Model Family:	Tweedie	**Df Model:**	7
Link Function:	log	**Scale:**	0.093154
Method:	IRLS	**Log-Likelihood:**	nan
		Deviance:	6113.0
		Pearson chi2:	4.02e+03
No. Iterations:	10		

	coef	**std err**	**z**	**P> |z|**	**[0.025**	**0.975]**
P_holder	0.0428	0.003	13.117	0.000	0.036	0.049
P_cover	0.0983	0.002	40.389	0.000	0.094	0.103
V_type	-0.0296	0.002	-14.022	0.000	-0.034	-0.025
Manuf_D	0.0697	0.002	28.348	0.000	0.065	0.074
Season	0.0319	0.002	16.627	0.000	0.028	0.036
Spare	-0.0011	0.000	-2.671	0.008	-0.002	-0.000
Maint	0.0170	0.001	31.855	0.000	0.016	0.018
GDP	-0.0045	0.000	-24.813	0.000	-0.005	-0.004

Source: Authors computation

Table 4: **Tweedie Model Result**

The sensitivity analysis conducted by removing the insignificant variable, namely the premium, reaffirmed the stability and reliability of the regression model's findings regarding the determinants of claims in motor insurance. Despite excluding the premium variable, the estimated coefficients for other predictor variables remained consistent, indicating that their respective impacts on claim likelihood remained unchanged. Thus, our analysis enhances confidence in the validity of the model and the economic interpretations drawn from its results, providing valuable insights for insurers, policymakers, and stakeholders in the motor insurance sector.

6 Conclusion and policy implications

The insurance industry has long been a cornerstone of economic stability, playing a crucial role in safeguarding the financial well-being of individuals and businesses alike. The general insurance business is predominantly led by motor and medical insurance classes, which together represent big share of total gross premiums. However, insurance claims pose a significant challenge for insurance companies across all sectors, exerting considerable financial strain. Motor insurance, in particular, holds immense importance as it directly impacts the lives of individuals worldwide.

In this study, we investigate the determinants of claims in the motor insurance sector in Rwanda, utilizing the Tweedie model alongside XGBoost, renowned for its adaptability in handling diverse data types encountered in insurance claims. The model's transparency, facilitated by easily interpretable parameters like regression coefficients, offers valuable insights into the factors influencing claim amounts. This, in turn, aids insurers in making informed decisions in relation to pricing and risk assessment.

Preliminary findings from the XGBoost feature importance analysis reveal that the manufacturing year of vehicles emerges as the most influential variable, followed by the type of policyholder and the type of coverage. Maintenance level, vehicle type, and spare parts expenditure also exhibit notable importance, whereas economic conditions (GDP), seasonality (Season), and premium amount rank lower in importance.

The subsequent Tweedie regression analysis further elucidates the relationship between predictor variables and claim probability. Institutional policyholders demonstrate a higher probability of filing claims compared to individual policyholders, while comprehensive insurance coverage is associated with a notably higher claim probability than third-party coverage. Interestingly, larger cars exhibit a slightly lower likelihood of claims, while older vehicles and rainy seasons correspond to increased claim probabilities. Furthermore, an increase in spare parts expenditure was associated with a marginal reduction in claim likelihood, possibly indicating prudent driving behaviour or fewer incidents requiring repairs. Conversely, increased maintenance expenditure correlated with a higher probability of claims. Additionally, a one-percentage increase in GDP was associated with a slight decrease in claim probability, potentially reflecting broader economic conditions influencing driving patterns, vehicle usage, or safety measures investment, thereby impacting claim behaviour.

Overall, the findings underscore the complex interplay of various factors in shaping claim behavior in motor insurance, offering valuable insights for insurers and policymakers in Rwanda. For the insurers could leverage insights from our analysis, such as vehicle manufacturing year, policyholder type, and coverage type into their underwriting criteria, to refine pricing and risk assessment strategies. Additionally, promoting proactive maintenance practices among policyholders could also help reduce claim frequency and severity.

On the side of regulatory authority, the National Bank of Rwanda should foster a regulatory environment conducive to innovation and data-driven decision-making, while ensuring compliance with standards and monitoring claims management processes for transparency and accountability. Moreover, the BNR can enhance insurers' routine practices by facilitating the collection of data on drivers' characteristics, thereby enriching future analyses.

References

Ariff, U. T., & Sirajuddin, K. (2016). A study on policyholders perception towards motor vehicle insurance (with special reference to pollachi taluk). *International Journal of Multidisciplinary Research and Modern Education (IJMRME), Volume II*(2).

Arrow, K. J., et al. (1974). *Essays in the theory of risk-bearing* (Vol. 121). North-Holland Amsterdam.

Bashir, I., Madhavaiah, C., & Naik, J. R. K. (2013). Motor insurance frauds in india: Detection and control mechanisms. *Srusti Management Review*, *6*(2), 27.

Bilal, M., Oyedele, L. O., Kusimo, H. O., Owolabi, H. A., Akanbi, L. A., Ajayi, A. O., ... Delgado, J. M. D. (2019). Investigating profitability performance of construction projects using big data: A project analytics approach. *Journal of Building Engineering*, *26*, 100850.

Born, P., & Bujakowski, D. (2022). Economic transition and insurance market development: evidence from post-communist european countries. *The Geneva Risk and Insurance Review*, 1–37.

Boucher, J.-P., Denuit, M., & Guillén, M. (2008). Models of insurance claim counts with time dependence based on generalization of poisson and negative binomial distributions. *Variance*, *2*(1), 135–162.

Browne, M. J., & Hoyt, R. E. (2000). The demand for flood insurance: empirical evidence. *Journal of risk and uncertainty*, *20*, 291–306.

Dionne, G., & Rothschild, C. (2014). Economic effects of risk classification bans. *The Geneva Risk and Insurance Review*, *39*, 184–221.

Dragos, S. L. (2014). Life and non-life insurance demand: the different effects of influence factors in emerging countries from europe and asia. *Economic research-Ekonomska istraživanja*, *27*(1), 169–180.

Dunn, P. K., & Smyth, G. K. (2001). Tweedie family densities: methods of evaluation. In *Proceedings of the 16th international workshop on statistical modelling, odense, denmark* (pp. 2–6).

Einav, L., Finkelstein, A., Ryan, S. P., Schrimpf, P., & Cullen, M. R. (2013). Selection on moral hazard in health insurance. *American Economic Review*, *103*(1), 178–219.

Epetimehin, F., & Odunaike, O. (2002). Appraisal of the stakeholder management model in the insurance industry. *European Journal of Business and Management*, *4*(15), 49–45.

Esho, N., Kirievsky, A., Ward, D., & Zurbruegg, R. (2004). Law and the determinants of property-casualty insurance. *Journal of risk and Insurance*, *71*(2), 265–283.

Feyen, E., Lester, R. R., & Rocha, R. d. R. (2011). What drives the development of the insurance sector? an empirical analysis based on a panel of developed and developing countries. *An Empirical Analysis Based on a Panel of Developed and Developing Countries (February 1, 2011). World Bank Policy Research Working Paper*(5572).

Friedman, M., & Savage, L. J. (1948). The utility analysis of choices involving risk. *Journal of political Economy*, *56*(4), 279–304.

Haiss, P., & Sümegi, K. (2008). The relationship between insurance and economic growth in europe: a theoretical and empirical analysis. *Empirica*, *35*, 405–431.

Hasan, M. M., & Dunn, P. K. (2011). Two tweedie distributions that are near-optimal for modelling monthly rainfall in australia. *International Journal of Climatology*, *31*(9), 1389–1397.

Jorgensen, B. (1997). *The theory of dispersion models.* CRC Press.

Jørgensen, B., & Paes De Souza, M. C. (1994). Fitting tweedie's compound poisson model to insurance claims data. *Scandinavian Actuarial Journal*, *1994*(1), 69–93.

Kahneman, D., & Tversky, A. (1979). Kahneman. *Daniel, and Tversky, Amos.Prospect Theory: An Analysis of Decision Under Risk. Econometrica*, *47*, 263–91.

Kunreuther, H., Pauly, M., et al. (2006). Insurance decision-making and market behavior. *Foundations and Trends® in Microeconomics*, *1*(2), 63–127.

Mathur, D., & Tripathi, A. (2014). Factors influencing customers choice for insurance companies-a study of ajmer city. *IOSR Journal of Business Management*, *16*(2), 35–43.

Melia, S., Chatterjee, K., & Stokes, G. (2018). Is the urbanisation of young adults reducing their driving? *Transportation research part A: policy and practice*, *118*, 444–456.

Millo, G., & Pasini, G. (2010). Does social capital reduce moral hazard? a network model for non-life insurance demand. *Fiscal Studies*, *31*(3), 341–372.

Mossin, J. (1968). Aspects of rational insurance purchasing. *Journal of political economy*, *76*(4, Part 1), 553–568.

Petkovski, M., Jordan, K., et al. (2014). An analysis of non-life insurance determinants for selected countries in central and south eastern europe: a co-integration approach. *Romanian Journal of Economic Forecasting*, *17*(3), 160–178.

Pratt, J. W. (1978). Risk aversion in the small and in the large. In *Uncertainty in economics* (pp. 59–79). Elsevier.

Prinja, S., Kaur, M., & Kumar, R. (2012). Universal health insurance in india: ensuring equity, efficiency, and quality. *Indian journal of community medicine: official publication of Indian Association of Preventive & Social Medicine*, *37*(3), 142.

Sherden, W. A. (1984). An analysis of the determinants of the demand for automobile insurance. *Journal of Risk and Insurance*, 49–62.

Signé, L., & Johnson, C. (2020). Africa's insurance potential: Trends, drivers, opportunities and strategies.

Smyth, G. K., & Jørgensen, B. (2002). Fitting tweedie's compound poisson model to insurance claims data: dispersion modelling. *ASTIN Bulletin: The Journal of the IAA*, *32*(1), 143–157.

Tversky, A., & Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and uncertainty*, *5*, 297–323.