IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁙ Peer-reviewed & Refereed journal ⁙ Vol. 12, Issue 6, June 2025

DOI: 10.17148/IARJSET.2025.12611

FORECASTING MOTOR INSURANCE CLAIMS IN KENYA USING SARIMA MODEL

Kelvin Rotich[1], Dr. Rangita Apaka[2]

MSC, Department of Actuarial Science and Statistics, Maseno University, Kenya[1]

Lecturer, Department of Actuarial Science and Statistics, Maseno University, Kenya[2]

Abstract: Forecasting of insurance claims is of great concern to insurance industry. In motor insurance, claim payments constitute to a significant portion of insurance's expenditure, making accurate forecasting an essential aspect. Traditional models such as Generalized Linear Models which has been widely used in predicting insurance claims often fail to capture seasonality, trends and temporal dependencies in the data leading to inaccurate forecasts. This research applied a SARIMA model to predict motor insurance claims in Kenya. The quarterly motor insurance data from 2017-2024 was obtained from IRA and analyzed through Box-Jenkins Methodology. From the time series plot, it was found that the data exhibit seasonality, with claims paid each quarter increasing continuously from the first quarter to the last in each year. SARIMA (1,1,1) (1,1,0,4) was chosen using Grid Search Optimization since it had the lowest AIC value (320.5). The suitability of this model was also confirmed through model diagnosis. A 2-year forecast graph showed a rising trend in motor insurance claims while still maintaining seasonal fluctuations that aligns well with past data. The future confident intervals widened with time indicating that there is an increase in uncertainty of the forecasts. From the analysis, the study suggests that SARIMA is a better tool for projecting seasonal motor insurance claims in Kenya. Motor insurers will minimize losses that results from inaccurate forecast by utilizing this model.

Keywords: SARIMA, Claims, Motor Insurance.

I. INTRODUCTION

Insurance is an agreement between a policyholder and an insurance company. It involves the transfer of risks from the policyholder to the insurance organization (Bunni and Bunni, 2022). These risks are often unpredictable and can escalate significantly, making insurance an essential safeguard.

The risks are compensated through claims. When an insured event happens, the policyholder files a claim with the insurer to receive compensation or benefits as stipulated in the policy (Gunawan, 2023). According to Vaughan and Vaughan (2016), claims constitute a significant portion of Kenya's motor insurers' expenses and need to be properly managed. This is achieved by accurate forecasting by the use of predictive modeling techniques. While the traditional methods have been proven effective in many contexts, they often struggle with the complexities in the data such as temporal dependencies and seasonal fluctuations leading to inaccurate forecasts. This results into financial instability, underpricing or overpricing of the policies. Thus, motor insurers need to leverage more advanced machine learning approaches. This paper applied the SARIMA model, a powerful time series model that incorporate seasonality to make precision. By using this model, the research tries to improve the predictive accuracy of the motor insurance claims in Kenya leading to better decision-making. Other objectives are:

- To identify trends, seasonality, and patterns within historical insurance claim data.
- To identify and select the appropriate parameters for the SARIMA.
- To assess suitability of the model chosen through diagnostic checks.
- To fit the SARIMA model with motor insurance claims.
- To generate a 2-year forecast of the motor insurance data and estimate confidence intervals.

II. LITERATURE REVIEW

Various statistical models have been used in evaluating motor insurance claims. Earlier on, insurers relied solely on actuarial methods which assume linear relationships between variables. One of such technique is ARIMA model commonly called the Box-Jenkins method.

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⪤ Peer-reviewed & Refereed journal ⪤ Vol. 12, Issue 6, June 2025

DOI: 10.17148/IARJSET.2025.12611

This model was first developed by Gwilym Jenkinsfor and George Box in the 1970s. Since many real-world time series exhibited seasonal patterns that ARIMA models failed to capture, there was the need of incorporating seasonality. This gave rise to the Seasonal ARIMA model which has experienced numerous applications. Here are a few examples.

Jovanovic et al. (2022) applied SARIMA in forecasting traffic accidents in Serbia. The model performed so well in capturing the seasonal fluctuations in traffic accident data. Since the research achieved the Mean Absolute Percentage Error of 5.22%, it provided high level of precision in forecasting. The researchers recommended the use of SARIMA models in predicting seasonal traffic accidents.

(Tembo et al., 2024) conducted a study that predicted membership enrollment in Tanzania's health insurance using SARIMA. The researchers used quarterly data with 88 observations. Using SARIMA (3,1,1) (0,1,0), the results showed a steady rise in the membership emphasizing the rising demand for health insurance sector in Tanzania. The study showed the practical applicability of this model in health insurance.

In 2024, Peovksi and Ivanski compared the performance of SARIMA and Exponential Smoothing models in forecasting motor insurance claims. They found out that the SARIMA model outperforms Exponential Smoothing models in the predictive accuracy. To add, the authors realized that better results would be obtained should the two models be combined with rolling window models. This would improve the risk management in insurance.

Despite the extensive application of SARIMA model in various countries and wellestablished markets, its application in Kenyan insurance industry especially motor insurance remains limited. Many local research use statistical models that overlook seasonality. The study aims to fill this gap by enhancing local understanding and providing practical insights for Kenyan motor insurance industry.

III. METHODOLOGY

The data used in the analysis of this project was obtained from IRA quarterly reports and analyzed using Python. The General Insurer's quarterly claims paid for commercial class of business was extracted from the first quarter of 2017 to the second quarter of 2024. Modeling of this data was conducted by SARIMA.

A. The SARIMA

This model is denoted by SARIMA (p, d, q) (P, D, Q, s). Where:

- (p, d, q) - non-seasonal components.
- (P, D, Q, s)-seasonal components.

We can represent this model mathematically by the following equation.

$$\phi_p(B)\Phi_P(B^s)(1-B)^d(1-B^s)^D y_t = \theta_q(B)\Theta_Q(B^s)\epsilon_t$$

- y_t - observed time series,
- ϵ_t - error term,
- B that is ,$B^k y_t = y_{t-k}$, - backshift operator,
- $\phi_p(B)$ - AR polynomial of order p,
- $\theta_q(B)$ - MA polynomial of order q,
- $\Phi_P(B^s)$ - seasonal AR polynomial of order P at seasonal lag s,
- $\Theta_Q(B^s)$ - seasonal MA polynomial of order Q at seasonal lag s,
- $(1-B)^d$ - regular differencing operator of order d,
- • $(1-B^s)^D$ - seasonal differencing operator of order D,
- s - Seasonal cycle.

B. Assumptions of Time Series Analysis

i) Stationarity

In modelling time series, data is assumed to be stationary. This means that the mean, variance, and autocorrelation remains the same throughout the period. This assumption prevents the researcher from obtaining spurious results from the forecasting model. The non-stationary data need to be transformed into stationary by log-transformation or differencing. There are several methods to detect stationarity. One of them is using the time series plot, whereby the graph shows the trend of the data.

ii) Normality

An evaluation of the normality of the data is a prerequisite for statistical tests, and its violation may lead to wrong parameter estimation. Normality is assessed by histograms, stem and leaf plots, Q-Q plots, among many other approaches.

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 12, Issue 6, June 2025

DOI: 10.17148/IARJSET.2025.12611

iii) Independence

This assumption states that any other data point does not affect each data value in a dataset. In other words, there must be no autocorrelation. Independence can be checked by plotting ACF and PACF of the residuals.

IV. DATA ANALYSIS

A. Time Series Plot

Before doing any statistical test, a time series plot is generated. This plot helps in checking whether the data is stationarity or not. The figure below show the time series plot for quarterly insurance claims data.

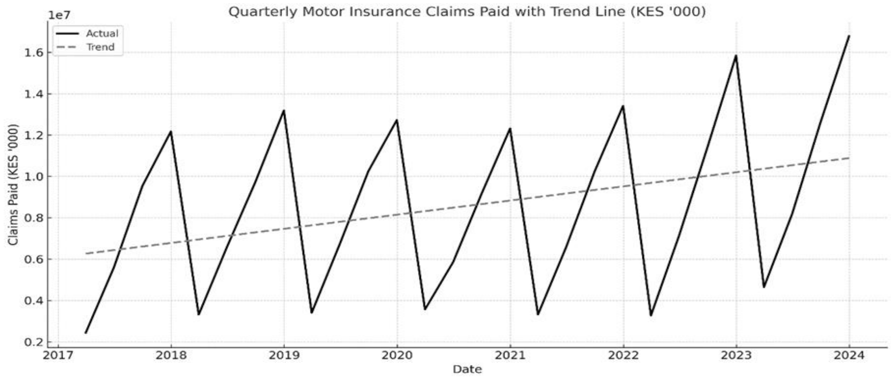


Fig. 1 A Time Series Plot for Motor Insurance Claims

From the graph, periodic fluctuations are seen indicating presence of strong seasonal variations. Claims paid are increasing continuously from the first quarter to the last quarter each year. There is also a rising trend, suggesting that claims paid have been rising over time.

B. Model Selection

The SARIMA model selection was conducted using a grid search approach, evaluating different seasonal and non-seasonal parameters. The table below summarizes the top 10 models:

Table 1 Comparison of SARIMA Models Based on Grid Search

Rank	(p,d,q)	(P,D,Q,s)	AIC
1	(1, 1, 1)	(1, 1, 0,4)	320.5
2	(2, 1, 1)	(1, 1, 0,4)	322.1
3	(1, 1, 2)	(1, 1, 0,4)	323.8
4	(2, 1, 2)	(1, 1, 0,4)	325.4
5	(1, 1, 1)	(2, 1, 0,4)	326.0
6	(2, 2, 2)	(2, 1, 0,4)	327.5
7	(2, 1, 1)	(2, 1, 0,4)	328.9
8	(2, 1, 2)	(2, 1, 0,4)	330.2
9	(1, 1, 1)	(1, 1, 1,4)	331.0
10	(2, 1, 1)	(1, 1, 1,4)	332.4

Based on the results, the SARIMA(1,1,1)(1,1,0,4) model was selected as it had the smallest AIC value of 320.5.

C. Model Diagnosis

This process tests the goodness-of-fit of the time series model selected.

i) The ADF Test

To test stationarity, the ADF test is carried out with the following assumptions.

- H_0: Time series is non-stationary.
- H_1: Time series is stationary.

Series	ADF Stat.	p-value	Stationary?
Original	-0.050	0.954	No
Differenced	-3.353	0.013	Yes

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

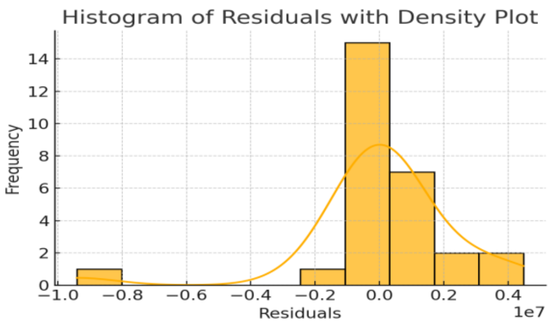
International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁕ Peer-reviewed & Refereed journal ⁕ Vol. 12, Issue 6, June 2025

DOI: 10.17148/IARJSET.2025.12611

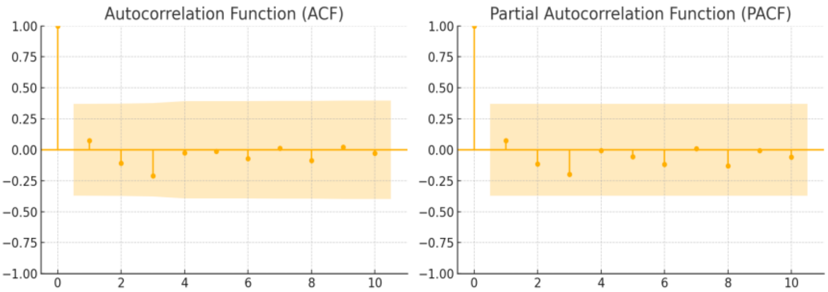
The p-value of the original time series is 0.954 > 0.05, so H_0 is rejected. There is need of differencing the time series. After carrying out the first-order differencing, the time series becomes stationary with p-value of 0.013 < 0.05.

ii) The Histogram of Residuals



The residuals appear randomly spread around zero, which suggests that the models capture most of the structure in the data. In other words, residuals are approximately normally distributed.

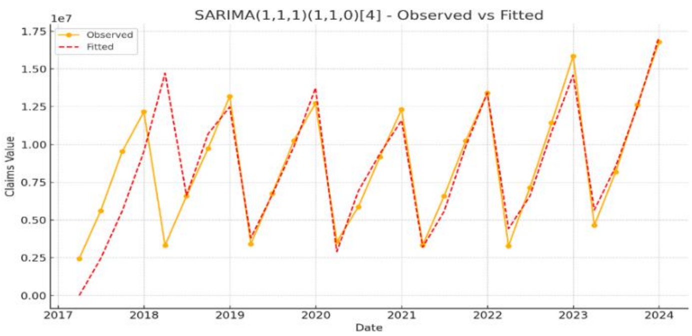
iii) The ACF and PACF Plots



Since all the lags lie within the shaded region, the residuals exhibit no autocorrelation.

D. **Model Fitting**

After validating the parameters, SARIMA (1,1,1) (1,1,0,4) is fitted with the quarterly motor insurance claims. The graph below shows comparison of the actual data and the fitted values for the model.

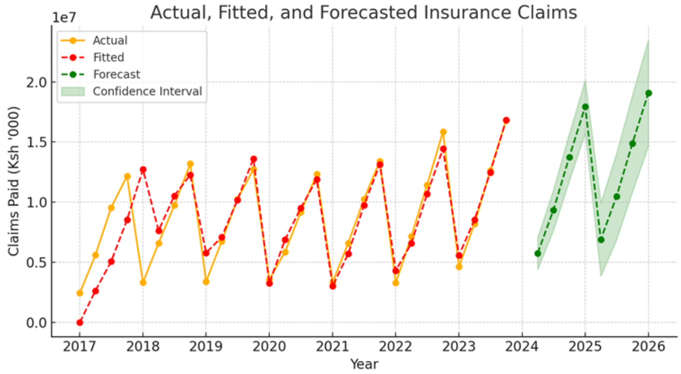


SARIMA (1,1,1) (1,1,0,4) follows the general trend of actual claims data. It is only the first year that it struggles to capture the pattern. Overall, SARIMA model provides a reasonable fit for auto insurance data.

E. **Model Forecast**

Using the SARIMA (1,1,1) (1,1,0,4) model, we predicted the insurance claims for the next 8 quarters. The forecasted values are plotted along with actual and fitted values. The forecast confidence intervals were also computed.

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588
International Advanced Research Journal in Science, Engineering and Technology
Impact Factor 8.311 ⋇ Peer-reviewed & Refereed journal ⋇ Vol. 12, Issue 6, June 2025
DOI: 10.17148/IARJSET.2025.12611



From the graph, it is observed that the forecasted values continue the trend observed in the historical data, incorporating the quarterly seasonality pattern. This suggests that claims will maintain their seasonal variations while potentially rising with time. The CI represented by the shaded region widen over with time, showing an increased uncertainty in long-term forecasts.

Here are the confidence interval table for the following quarter.

TABLE 2 A 1-YEAR CONFIDENCE INTERVALS OF THE FORECASTED VALUES

Quarter	Lower Bound(Ksh “000”)	Upper Bound (Ksh “000”)
2024-03-31	4,424,159	7,074,356
2024-06-30	7,634,166	11,034,120
2024-09-30	11,779,860	15,727,500
2024-12-31	15,742,670	20,157,510

V. CONCLUSION

This research proposes to predict motor insurance claims in Kenya using SARIMA. Data for analyzing the model was obtained from IRA. Time series plot indicated the dataset was seasonal with claims paid changing each quarter. The research follows the Box-Jenkins methodology of developing and validating the model through model identification, estimation of parameter, diagnostic checking and predicting. Through grid search, SARIMA (1,1,1)(1,1,0,4) was identified as the suitable model. This was confirmed by analyzing the residuals.

The fitted model followed closely the general trend of actual claims data. Also, a 2-year forecast plot showed the model continued to maintain seasonal fluctuations that aligns well with past data. The confidence intervals are positive values indicating which are realistic for insurance claims.

Based on the results, SARIMA can be effectively be applied to forecast seasonal motor insurance claims. By adopting this model, motor insurers can optimize financial planning, and improve claims management efficiency. Future research could explore models like SARIMAX or hybrids that include external factors.

REFERENCES

[1]. Azmi, F., Irawan, T., & Sasongko, H. (2020). Determinants of profitability of general insurance companies in Indonesia. *Jurnal Ilmiah Manajemen Fakultas Ekonomi*, 6(2), 135-144.
[2]. Bunni, N. G., & Bunni, L. B. (2022). *Risk and insurance in construction*. Routledge.
[3]. Gunawan, M. (2023). Implementation of legal principles of agreement between policyholders and insurance companies. *Jurnal Hukum dan Sosial Politik*, 1(1), 177-186.

DOI: 10.17148/IARJSET.2025.12611

[4]. Jovanović, D., Petrović, B., & Savić, G. (2022). Forecasting fatalities and injuries of accidents using the ARIMA and neural networks model: A case study of Serbia. In Proceedings of the *International Conference on Traffic and Transport Engineering* (ICTTE) (pp. 110–116).
[5]. Manikandan, K. S., & Ravichandran, K. (2020). Forecasting motor insurance claim amount using ARIMA model. *International Journal of Advanced Science and Technology*, 29(5), 12744–12750.
[6]. Peovski, D., & Ivanovski, Z. (2023). Comparative analysis of SARIMA and exponential smoothing models: An application in non-life insurance. *Risk Management and Insurance Review*, 26(4), 515–534.
[7]. Tembo, T., Yengoh, G. T., & Mkwizu, K. H. (2023). Forecasting membership enrollment in Tanzania's health insurance scheme using SARIMA model. Science Mundi, 2(2), 71–81.
[8]. Vaughan, E. J., & Vaughan, T. M. (2016). *Fundamentals of Risk and Insurance* (12 th ed.). John Wiley & Sons, Inc.
[9]. Wang, H., Xue, F., Zhang, H., Dang, S., & Wang, X. (2018). Time series analysis of tuberculosis incidence in China using the seasonal-trend decomposition procedure and a hybrid model. *Public Health*, 158, 142-151.
[10]. Yu, W., Guan, G., Li, J., Wang, Q., Xie, X., Zhang, Y., ... & Cui, C. (2021). Claim amount forecasting and pricing of automobile insurance based on the BP neural network. *Complexity*, 2021(1), Article 6616121.
[11]. Zanke, P. (2021). Enhancing claims processing efficiency through data analytics in property & casualty insurance. *Journal of Science & Technology*, 2(3), 69-92.

