A Comparison of SARIMA and Holt-Winters Models in Forecasting Motor Insurance Claims: A Case Study of Kenya

.

ABSTRACT

|  |
| --- |
| This paper compared two time series, the SARIMA and Holt-Winters multiplicative models, in the context of forecasting Kenya's motor insurance claims. Secondary data was sourced from the Insurance Regulatory Authority (IRA). The time series plot showed seasonality in the data, making it necessary to use models that capture seasonal components. The SARIMA (1, 1,0)(0,1,0)[4] model and the Holt-Winters multiplicative model with parameters α = 0.70, β = 0.10, and γ = 0.01 were selected as optimal models for forecasting. Upon fitting the two models with the underlying data, graphical results showed that both aligned closely with the actual claims, effectively capturing the seasonal and trend components present in the data. However, to supplement the visual analysis, further evaluation using RMSE, MAE, and MAPE was conducted. The results revealed that the Holt-Winters multiplicative model achieved lower values in all three metrics, indicating superior predictive performance. While the Holt-Winters model outperformed SARIMA, the SARIMA model nonetheless delivered satisfactory results and still can be considered as its suitable alternative. |

*Keywords: SARIMA, Holt-Winters, Forecasting, Insurance Claims.*

1. INTRODUCTION

Claims play one of the major roles in the insurance industry. An insurer is obliged to reimburse any valid claim. Thus, it is necessary that insurance companies have a good understanding of the amounts of claims they should anticipate. This is especially true for the motor insurance class of business which typically experiences a high frequency and volume of claims.

Without reliable forecasts, insurance providers may either under-reserve or over-reserve, leading to insufficient capital allocation. Insurers need to leverage more reliable forecasting methods. Generally, many forecasting techniques exist among which we get the SARIMA and Holt-Winters. These two machine algorithms are not only more accurate but also suitable for handling seasonal data. This article compares the two models in the context of the Kenyan motor industry. Mainly, the study tries to find out the most accurate model to anticipate seasonal claims.

2. literature review

The development of time series forecasting models has risen significantly over the last century, with key contributions from scholars aiming to improve prediction accuracy.

While SARIMA dates its origin in the 1970s, it gained popularity in the 1980s. George Box and Gwilym laid out structured and practical methods of implementing ARIMA models. Because of seasonality in many real-world data, the model morphed into SARIMA.

Holt-Winters on the other hand is a time series model that was developed in the mid-20th century to capture seasonality in the data. This technique is founded on the principles of exponential smoothing that incorporate components such as trend, seasonality, and level.

Since the two modeling techniques handle data with seasonality, various studies have been conducted in the insurance industry to compare their performances.

First, Pires et. at (2022) modelled monthly time series data using SARIMA and Holt-Winters techniques for claim frequencies of home insurance data. The study compared the performance of the two models using metrics such as MSE, MAPE, and RSME. The results demonstrated that SARIMA was more superior. Still, with good performance, the Holt-Winters could also be considered a possible alternative.

According to Kopani and Kopani (2025), SARIMA outperforms Holt-Winters in predicting Albania’s life insurance premiums for long-term accuracy while Holt-Winters did pretty well on short-term projections. The research highlighted the importance of aligning forecasting methods with the specific goals of the insurance companies.

Peovski & Ivanovksi (2024) carried out another research using non-life insurance data-including claims. They realized that SARIMA still outdoes Holt-Winters in several statistical metrics. Although the latter could occasionally produce better results, SARIMA demonstrated overall superiority in the test period.

In Kenya, only a few studies such as the one by Rotich and Apaka (2025) have applied seasonal time series models such as SARIMA to model motor insurance data; however, comparative examination with Holt-winters remains scanty.The research bridges this gap by evaluating the accuracy of the two models in forecasting Kenya’s motor insurance claims.

3. methodology

**3.1 Data description**

Secondary data from the Insurance Regulatory Authority of Kenya is used in the analysis of the study. In specific, motor insurance claims are extracted from 2017-2024 and modelled using SARIMA and Holt-Winters.

**3.2 SARIMA Model**

The SARIMA model is denoted as: ***SARIMA(p, d, q)(P, D, Q)s***

Where:

- p: AR terms

- d: Non-seasonal differences

- q: Lagged forecast errors

- P: Seasonal AR terms

- D: Seasonal differences

- Q: Seasonal MA terms

- s: Seasonal cycle

Its general form is written as:

Φ\_p(B)Φ\_P(B^s)(1−B)^d(1−B^s)^D y\_t = Θ\_q(B)Θ\_Q(B^s)ε\_t

Where:

- B: Backshift operator

- y\_t is the observed value at t

- ε\_t is white noise

- Φ and Θ represent polynomials for AR and MA components

**3.2.1 Assumptions of SARIMA model**

***3.2.1.1 Normality***

One of the most basic methods of assessing whether the time series is normal is by plotting the histogram. If the data is not normal, then there is a high chance of getting the wrong parameter.

***3.2.1.2 Stationarity***

To prevent the researcher from obtaining spurious results, data is assumed to be stationary. This implies that the time series has a constant mean, autocorrelation, and variance. This assumption can be verified by carrying out the ADF test. Usually, for data that is not stationary, either differencing or power transformation is applied. This stabilizes the variance and removes the trend.

***3.2.1.3. Independence***

There must be no autocorrelation, that is, the data values should not affect other data points of the dataset. This assumption is evaluated by ACF and PACF plots.

**3.2.2 Box-Jenkins Methodology**

This is an iterative process involving model identification, parameter estimation, and model verification. The objective of these steps is to obtain the most reliable model that captures patterns in the historical data while minimizing the risk of overfitting.

The initial step, model identification, involves determining the most appropriate parameters of the model. The traditional way of doing this is using ACF and PACF plots. However, this approach is sometimes subjective and time-consuming. Grid search offers an automated alternative that is free from human error.

Secondly, parameter estimation uses methods such as MLE and LSE to minimize the difference between the actual data and the model forecast.

Finally, model verification validates whether the model adequately captures the structure of the time series. Also, it helps in determining whether the assumptions are satisfied, ensuring the model is statistically valid. Once the model passes all these steps, it is now ready to be used in forecasting future values.

**3.3 The Holt-Winters Model**

The Holt-Winters model comes in two versions depending on the nature of the seasonality. The Additive Holt-Winters is usually applied for seasonal variations that do not grow or shrink with time while the Multiplicative Holt-Winters performs well on seasonal fluctuations that either increase or decrease with time. Since Kenya's motor insurance claims grow, the Multiplicative Holt-Winters model is the most appropriate. It is made up of the following three smoothing equations and forecasts.

1. Level:

 Lₜ = α \* (yₜ / Sₜ₋ₛ) + (1 - α) \* (Lₜ₋₁ + Tₜ₋₁)

2. Trend:

 Tₜ = β \* (Lₜ - Lₜ₋₁) + (1 - β) \* Tₜ₋₁

3. Seasonal:

 Sₜ = γ \* (yₜ / Lₜ) + (1 - γ) \* Sₜ₋ₛ

4. Forecast:

 ŷₜ₊ₘ = (Lₜ + m \* Tₜ) \* Sₜ₋ₛ₊ₘ

**3.3.1 Parameters**

α, β, γ: Smoothing parameters for level, trend, and seasonality respectively (0 < α, β, γ < 1).

These values are obtained by minimizing the sum of squared errors.

s: Length of the seasonal period

m: Number of periods ahead to forecast

yₜ is the actual value at time t,

4. results and discussion

**4.1 Time Series Plot**

First, we determine the seasonality of the data by generating the time series plot.



**Figure 1 Time Series Plot for Insurance Claims**

It is noted that Kenya’s motor insurance claims data exhibits a rising trend with clear seasonal peaks happening in the last quarter of each year. This implies that claims have generally been rising over time. Based on these, the SARIMA and Holt-Winters Multiplicative models are the most appropriate ones to use.

* 1. **The SARIMA Model**

**4.2.1 Model Selection and Identification**

In order to select the most appropriate model, grid search optimization was used. The process involves fitting SARIMA models across a range of parameter combinations for seasonal and non-seasonal components. That is, setting d=1 to remove the trend and D=1 to address quarterly seasonality. Other parameters, p,d, P, and Q are changed within a range of 0 and 2 during the search. After this, different combinations are evaluated using the Akaike Information Criterion(AIC). Always, the model with the lowest AIC is chosen.

**Table 1 The Comparison of Different SARIMA Models**

|  |  |  |  |
| --- | --- | --- | --- |
| 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 |

Based on the table above, the SARIMA(1,1,1)(1,1,0,4) model is selected.

**4.2.2 Model Diagnosis**

Before fitting the selected model to the data, diagnostic checks is done to assess its validity.

* + - 1. ***Stationarity***

The Augmented Dickey-Fuller test is used to determine the stationarity of the time series. The test has the following hypothesis:

* H0: Time series is non-stationary.
* H1: Time series is stationary.



With a p-value of 0.954, the original time series is not stationary. However, after first-order differencing, this value becomes less than 0.05, indicating the series is now stationary. This is noted on parameter d=1 on the model.

* + - 1. ***Normality***



**Figure 2 Histogram of Residuals**

The visual inspection of residuals appears approximately bell-shaped which is actually a near-normal distribution.

* + - 1. ***Independence***



**Figure 3 ACF and PACF plots**

The two plots clearly show there is no significant spike, that is, no autocorrelation exists. This further confirms the model is suitable to fit with the data.

**4.2.3 Model fitting**

The figure below shows the comparison of the fitted model versus actual data.



**Figure 4 The Graph of Fitted The Model**

While the model struggles to capture the actual trend in the first year, it progressively aligns more closely in the subsequent years. Generally, it is effective in tracking motor insurance patterns, making it suitable for predictions.

**4.2.4 Forecasting**

A 2-year forecast was developed to anticipate future claims patterns as shown below.



**Figure 5 A 2-Year Forecast**

The forecasts show continued trends observed in the historical data. There are also clear seasonal peaks in the fourth quarters that aligns with the actual data. The outcomes suggest the SARIMA model can be effectively applied in forecasting motor insurance claims. Having established this, it is also important to explore the Holt-Winters approach before comparing their performance.

* 1. **The Holt-Winters**

**4.3.1 Parameter Selection**

Before fitting the model with the data, it is crucial to find the appropriate values for smoothing parameters: level, trend, and seasonality. These values should be between 0 and 1. This testing is also done using a grid search procedure. The Root Mean Square Error(RMSE) is used as an evaluation metric.

The following table presents the candidate models from the Holt-Winters.

**Table 2 Candidate Models for Holt-Winters**

|  |  |  |  |
| --- | --- | --- | --- |
| Alpha | Beta | Gamma |  RMSE(KES) |
| 0.7 | 0.10 | 0.01 | 242,162.46 |
| 0.5 | 0.20 | 0.01 | 252,428.66 |
| 0.7 | 0.10 | 0.05 | 268,584.07 |
| 0.5 | 0.20 | 0.05 | 277,981.54 |
| 0.7 | 0.10 |  0.10 | 292,635.29 |
| 0.5 | 0.20 | 0.10 |  299,635.29 |
| 0.7 | 0.10 | 0.20 | 309,990.20 |

 From the table, it can be seen that the smallest values for RMSE is obtained when alpha=0.70, Beta=0.10, and Gamma=0.01. Next, with these parameters, is fitting the model.

**4.3.2 Model Fitting**



**Figure 6 Holt-Winters Fitted Model Graph**

Just like the first model, the Holt-Winters Multiplicative model still performs well in capturing the overall trend and quarterly seasonality of the data.

**4.3.3 Model Forecasting**



**Figure 7 A 2-year Forecast Using Holt-Winters**

Based on the figure above, it is noted that there is still an upward trend in the forecast which is consistent with the historical data.

* 1. **Comparison of SARIMA and Holt-Winters Models**

Because the two models appear to yield the same results when visualized on graphs, it is important to supplement visual representation with statistical metrics. Here, the RMSE, MAE, and MAPE are employed to compare further their performance.

**Table 3 Statistical Metrics of The Two models**

|  |  |  |
| --- | --- | --- |
| Metric | SARIMA(1,1,1)(1,1,0,4) |  Holt-Winters Multiplicative |
| RMSE | KSh. 785,107.04 | KSh. 639,881.77 |
| MAE | KSh. 655,829.66 | KSh. 553,740.31 |
| MAPE | 10.05% | 6.74% |

Based on these results, the Holt-Winters model outperforms the SARIMA across all three metrics. That is, it has lower RMSE, lower MAE, and lower MAPE. This suggests that the Hot-Winters multiplicative model captures motor insurance seasonal patterns more accurately.

5. Conclusion

The study aims to forecast motor insurance claims in Kenya using the SARIMA and the Holt-Winters Multiplicative Model. It focuses on capturing the rising trend and seasonality, which are present in the IRA’s motor insurance claims data. The initial time series plot confirms these patterns, justifying the application of the two models.

The SARIMA (1,1,1) (1,1,0,4) model was identified as the most appropriate parameter and fitted with the data. Its performance was satisfactory, capturing the upward trend and seasonality in the data. Like the SARIMA model, the Holt-Winters Multiplicative Model was fitted with the same data using parameters determined through the same approach. The model yielded similar outcomes.

While the graphical inspection portrayed that both models produced similar results/forecasts, further evaluation showed slight differences. By the use of RMSE, MAE, and MAPE, the quantitative comparison revealed that the Holt-Winters Multiplicative Model performs better than the SARIMA model in forecasting motor insurance. However, since there is not much difference in their performance, the SARIMA model can still be considered as an alternative to the Holt-Winters multiplicative model to forecast Kenya’s motor insurance claims.

**ABBREVIATIONS**

**RMSE**: Root Mean Square Error

**MAE**: Mean Absolute Error

**MAPE**: Mean Absolute Percentage Error

References

1. Box, G. E. P., & Jenkins, G. M. (1976). Time Series Analysis: Forecasting and Control. Holden-Day.
2. Hyndman, R. J., & Athanasopoulos, G. (2021). *Forecasting: Principles and Practice* (3rd ed.). OTexts. <https://otexts.com/fpp3/>
3. Insurance Regulatory Authority of Kenya. (2024). *Annual Insurance Industry Report 2024*. Nairobi: IRA Kenya. Retrieved from <https://www.ira.go.ke/>
4. Pires, L. F., Gonçalves, A. M., Ferreira, L. F., & Maranhão, L. (2022). Forecasting Models: An Application to Home Insurance. In Proceedings of the 18th International Conference on Computational and Financial Econometrics (CFE 2022). Universidade do Minho. Retrieved from <https://repositorium.sdum.uminho.pt/handle/1822/87989>
5. Peovski, F., & Ivanovski, I. (2022). *Forecasting models: An application to home insurance*.
6. Rotich, K., & Apaka, R. (2025). Forecasting motor insurance claims in Kenya using SARIMA model. *International Advanced Research Journal in Science, Engineering and Technology (IARJSET)*, 12(6), 45–52. <https://iarjset.com/wp-content/uploads/2025/06/IARJSET.2025.12611.pdf>