IJSA (2025), Vol. 9, No. 1, 145-156
DOI: https://doi.org/10.29244/ijsa.v9i1p145-156

Indonesian Journal of Statistics and Its Applications
Available at https://journal-stats.ipb.ac.id/index.php/ijsa
e-ISSN: 2599-0802

Forecasting Nonlinear Time Series with ARIMA, ANN, and Hybrid Models: A Case Study on Inflation Rate in Sri Lanka *

W. M. S. Bandara[1‡], W. A. R. De Mel[1]

[1] Department of Mathematics, Faculty of Science, University of Ruhuna, Sri Lanka
‡corresponding author: bandarasudarshana009@gmail.com

Abstract

In time series forecasting, hybrid models combining autoregressive integrated moving average (ARIMA) and artificial neural networks (ANNs) have gained prominence due to their ability to capture both linear and nonlinear patterns within data. ARIMA models are effective at modeling linear relationships, while ANNs are adept at handling complex nonlinear structures. However, each model has its limitations when used independently. This study presents a hybrid model that integrates the strengths of both ARIMA and ANN to forecast the monthly inflation rate in Sri Lanka using historical data from 1988 to 2018. Our findings demonstrate that the proposed hybrid model outperforms the standalone ARIMA and ANN models, particularly in terms of Mean Absolute Percentage Error (MAPE) and Root Mean Square Error (RMSE). By leveraging the complementary strengths of ARIMA and ANN, this hybrid approach provides a robust forecasting framework for handling the diverse structural complexities of time series data.

Keywords: ANN, ARIMA, Forecasting, Hybrid, Inflation.

* Received: Jul 2024; Reviewed: Dec 2024; Published: Jun 2025

1. Introduction

The rate of inflation is a crucial economic indicator that reflects the percentage change in the overall price level of goods and services within an economy over a given period, typically on a yearly basis. A positive inflation rate indicates an upward trend in prices, resulting in a decline in the purchasing power of money. Conversely, deflation, or a negative inflation rate, signifies falling prices (Mankiw, 2014) Monitoring inflation is essential for policymakers as it influences critical decisions regarding monetary policy, including adjustments in money supply and interest rates to maintain economic stability, encourage growth, and avoid recessions (Mishkin, 2010) .

From a microeconomic standpoint, inflation trends significantly affect a variety of financial decisions (Mankiw, 2014). For businesses, an understanding of inflation assists in forecasting future costs and setting appropriate pricing strategies for their products and services. For investors, inflation impacts returns on investments, bond yields, and stock market health. For individuals, inflation affects living costs and influences decisions on savings and consumption recessions (Mishkin, 2010) . Given these factors, accurate inflation forecasting, the primary objective of this study, provides valuable insights to a wide range of stakeholders.

Time series forecasting, a core area of data analysis, involves predicting future data points based on historical observations (Hyndman & Athanasopoulos, 2014). This technique has seen widespread application across various fields, including economics, finance, agriculture, meteorology, and biomedical research, owing to the availability of comprehensive datasets and advancements in information processing technology (Hyndman & Athanasopoulos, 2014). The key to successful forecasting lies in detecting and interpreting patterns within the data, which in turn supports well-informed predictions about future trends.

The Autoregressive Integrated Moving Average (ARIMA) model, introduced by (Box & Jenkins, 1976), has long been a preferred approach for time series forecasting. ARIMA models are popular due to their simplicity and flexibility, but they assume a linear relationship between time series data points (Hyndman & Athanasopoulos, 2014). This linearity assumption, however, often proves inadequate when dealing with complex, nonlinear data. To address this limitation, various nonlinear models have been proposed, including bilinear models (Lin et al., 2015) threshold autoregressive models, and autoregressive conditional heteroskedasticity (ARCH) models (Tong, 1990).

Artificial Neural Networks (ANNs) have also gained traction in the realm of time series forecasting, offering a flexible computational framework capable of modeling a wide range of nonlinear problems without the need for prior assumptions about the model's structure (Haykin, 2009), (Bandara & De Mel, 2024) ANNs are highly adaptable and can capture both linear and nonlinear dependencies in data (G. Zhang et al., 2008) However, neither ARIMA nor ANNs alone can consistently deliver accurate predictions for time series data that exhibit both linear and nonlinear characteristics (G. P. Zhang, 2003). This challenge has spurred researchers to explore hybrid models that integrate the strengths of ARIMA and ANN models (G. P. Zhang, 2003)

Several studies have demonstrated the efficacy of hybrid models in improving forecast accuracy. For instance, (G. P. Zhang, 2003)proposed a hybrid ARIMA-ANN model for forecasting nonlinear time series data, achieving superior results compared to standalone models. Similarly, (Cadenas & Rivera, 2009) applied a hybrid ARIMA-ANN model to wind speed forecasting in Mexico, finding that the hybrid model outperformed both ARIMA and ANN models in terms of accuracy. A recent study by (R. M. K. T. Rathnayaka., 2020)applied both ARIMA and ANN models to forecast daily stock price fluctuations of industry sectors in the Colombo Stock Exchange and reported that ANN significantly outperformed ARIMA, with a much lower Mean Absolute Percentage Error (MAPE), thereby supporting the effectiveness of hybrid or nonlinear models in financial time series forecasting.

In this study, we propose a hybrid model that integrates ARIMA and ANN to forecast the monthly inflation rate in Sri Lanka from 1988 to 2018. Our proposed model seeks to capitalize on the strengths of both ARIMA and ANN, offering a robust approach to accommodate both linear and nonlinear patterns within the data. The performance of the hybrid model will be compared with standalone ARIMA and ANN models, with accuracy assessed using MAPE and RMSE metrics.

2. Research Methodology

This section details the forecasting models utilized in this study, including ARIMA, ANN, and the hybrid ARIMA-ANN model. The models were evaluated using the monthly inflation rate data from Sri Lanka for the period January 1988 to August 2018. The data was obtained from the Trading View website (TradingView, 2023) The models were implemented using the R software (version 4.1.3) and the Keras library (Chollet & others, 2018) with accuracy evaluated based on Mean Absolute Percentage Error (MAPE) and Root Mean Square Error (RMSE).

2.1 Data

In this study, we consider the monthly inflation rate data in Sri Lanka from January 1988 to August 2018. 1 depicts these inflation rate data.

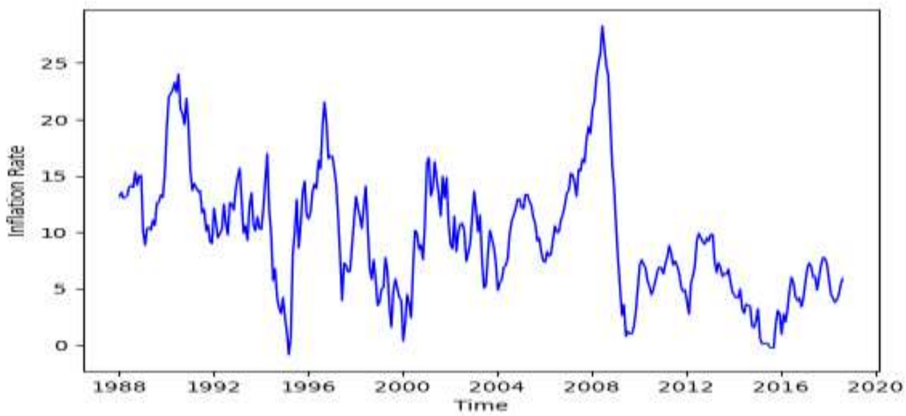

Figure 1: Monthly mean Inflation rate of Sri Lanka(1988-2018)

The data set used in this study contains a total of 368 observations, with a mean of 9.61, a variance of 29.91554, a minimum value of -0.890, and a maximum value of

28.310. The time series plot of the data exhibits a stochastic nature, with only a few unusual data points, particularly one in June 2008, likely due to the political situation in the country at the time. The data set was divided into three parts: training data from January 1988 to April 2017, used to test the accuracy of fitted models; test data from May 2017 to February 2018; and validation data from March 2018 to August 2018, used to assess the accuracy of forecasted values.

2.2 Study Area

The study focuses on forecasting inflation trends in Sri Lanka, an area of economic importance due to inflation's significant impact on purchasing power, policy decisions, and financial planning. Sri Lanka’s historical inflation patterns present both linear and nonlinear trends, making it a suitable case for testing hybrid forecasting techniques.

2.3 Model Formulation

2.3.1 ARIMA Model

The Autoregressive Integrated Moving Average (ARIMA) model is widely used for time series forecasting due to its simplicity and flexibility (Box & Jenkins, 1976)The ARIMA model is denoted as ARIMA(p,d,q), where p is the number of lag observations in the model (autoregressive part), d is the number of times the data has to be differenced to achieve stationarity (integrated part), q is the size of the moving average window. The mathematical formulation of the ARIMA model is given by:

$$(1-\sum_{i=1}^{p}\phi_i L^i)(1-L)^i y_t = (1+\sum_{j-1}^{q} q_i\, L^j)\varepsilon_t \qquad (1)$$

where L is the lag operator, ϕi are the autoregressive coefficients, θj are the moving average coefficients, and ϵt is white noise. ARIMA models assume linearity in the time series, which limits their application to nonlinear data (Hyndman & Athanasopoulos, 2014). This limitation led to the exploration of Artificial Neural Networks (ANN) and hybrid models.

2.3.2 Artificial Neural Networks (ANN)

Artificial Neural Networks (ANNs) have gained significant traction in time series forecasting due to their ability to capture complex nonlinear relationships without predefined assumptions about the underlying data structure (G. Zhang et al., 2008) The architecture of an ANN typically consists of three layers: (1) An input layer that receives the historical data points, (2) One or more hidden layers where nonlinear transformations are applied using activation functions, (3) An output layer that predicts the future values.

For this study, a single-hidden-layer feedforward network was used, where the relationship between the input and the output is expressed as:

$$x_t = \alpha_o + \sum_{j=1}^{p} \alpha_j\, G\left(\beta_o \sum_{i=1}^{q} \beta_{ij}\, X_{t-i}\right) + \varepsilon_t \qquad (2)$$

where G(X) is the activation function (logistic in our case), and α_j, β_{ij} are the parameters to be estimated. ANNs can model nonlinear patterns effectively but may struggle with linear patterns, hence the motivation for hybrid models (Haykin, 2009) (G. P. Zhang, 2003).

2.3.3 ARIMA-ANN Hybrid Model

To leverage the strengths of both ARIMA and ANN, we applied a hybrid model that combines the linearity of ARIMA and the nonlinearity of ANN. This approach, proposed by (G. P. Zhang, 2003)assumes that the time series consists of both linear and nonlinear components, which can be modeled separately.

The observed time series X_t can be decomposed as:

$$X_t = L_t + N_t \quad (3)$$

where Lt is the linear component captured by the ARIMA model, and Nt is the nonlinear component captured by the ANN model. First, the ARIMA model is applied to estimate Lt, and the residuals (nonlinear component) are used to train the ANN model:

$$\mathrm{N_t} = \mathrm{X_t} - \widehat{\mathrm{L}}_\mathrm{t} \quad (4)$$

The final forecast is obtained by combining the forecasts from the ARIMA and ANN models:

$$\widehat{X}_t = \widehat{L}_t + \widehat{N}_t \quad (5)$$

This hybrid model improves forecasting accuracy by accounting for both linear and nonlinear patterns in the data (G. P. Zhang, 2003, Cadenas & Rivera, 2009).

2.4 Algorithm for Optimal ARIMA Model Selection

The selection of the best ARIMA model for the time series data was based on minimizing the Akaike Information Criterion (AIC). The pseudo-code for selecting the optimal ARIMA order is provided in Algorithm 1.

```
Input: Data series X, max p, d, q
For p = 0 to p_max:
  For d = 0 to d_max:
    For q = 0 to q_max:
      Fit ARIMA(p,d,q)
      Compute AIC
      If AIC is minimum:
        Save model as optimal
Return: Optimal (p,d,q)
```

2.5 Model Evaluation Metrics

To evaluate the performance of the models, two common error metrics were used, namely Mean Absolute Percentage Error (MAPE) and Root Mean Square Error (RMSE). MAPE measures the accuracy of the forecast by calculating the average absolute percentage difference between actual and forecasted values:

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\frac{|Y_t - \widehat{Y}_t|}{Y_t} \tag{6}$$

RMSE measures the standard deviation of the residuals (forecast errors):

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(Y_t - \widehat{Y}_t)^2} \tag{7}$$

These metrics provide insight into the performance of each model, with lower values indicating better forecast accuracy (Hyndman & Athanasopoulos, 2014)

3. Results and Discussion

3.1 Pseudo-code for Finding Optimum ARIMA Order

In this study, the ARIMA model was first employed to capture the linear components of the inflation time series data. The order of the ARIMA model, denoted as (*p, d, q*), was selected by minimizing the Akaike Information Criterion (AIC) as described in Algorithm 2.

Algorithm 2: Pseudo-code for Finding Optimal ARIMA Order

Require: Data series *X*, maximum *p*, *d*, *q* values
Ensure: Optimal ARIMA Order **for** $p \leftarrow 0$ *to* p_{max} **do**
$d \leftarrow 0$ to d_{max} **for** $q \leftarrow 0$ *to* q_{max} **do**
Fit the ARIMA model with (*p, d, q*)
Calculate current AIC value for the model **if** *current AIC ¡ minimum AIC* **then**
Update minimum AIC
Update optimal ARIMA order
return Optimal ARIMA order

Algorithm 2 provides the pseudo-code used to determine the optimal ARIMA order by iteratingthrough different combinations of *p*, *d*, and *q* values to find the model that minimizes the AIC.This optimal ARIMA model was used in subsequent analyses.

3.2 ARIMA Model Errors

The residuals generated by the ARIMA model were analyzed to capture the nonlinear patternsthat the ARIMA model could not account for. Figure 2 illustrates the errors generated by the ARIMA model.

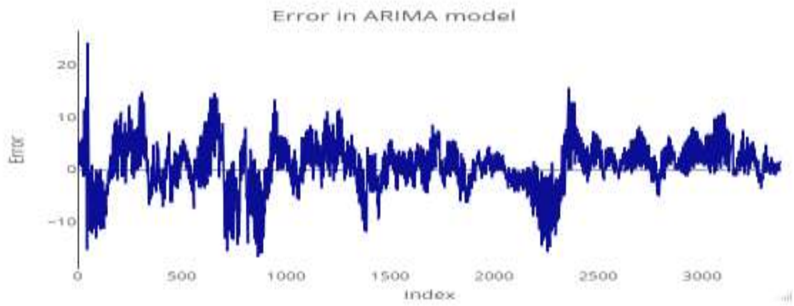

Figure 2: Residuals from the ARIMA model predictions

The minimum residual error was -16.61, the maximum was 24.1621, and the mean error was0.8166, with a variance of 20.5176. These residuals were used as input for the ANN model to capture the nonlinear components of the time series.

3.3 Predicted Test Data

The predicted test data, compared across ARIMA, ANN, and the hybrid model, is displayed inFigure 3. The hybrid model provides more accurate predictions than both standalone ARIMA and ANN models.

3.3.1 Pseudo-code for Finding Optimum ARIMA Order

In this study, the ARIMA model was first employed to capture the linear components of the inflation time series data. The order of the ARIMA model, denoted as (p, d, q), was selected by minimizing the Akaike Information Criterion (AIC) as described in Algorithm 2.

Algorithm 2: Pseudo-code for Finding Optimal ARIMA Order

Require: Data series X, maximum p, d, q values
Ensure: Optimal ARIMA Order for p ← 0 to pmax do
    d ← 0 to dmax for q ← 0 to qmax do
    Fit the ARIMA model with (p, d, q)
 Calculate current AIC value for the model if current AIC ¡ minimum AIC then

    Update minimum AIC
    Update optimal ARIMA order
return Optimal ARIMA order

Algorithm 2 provides the pseudo-code used to determine the optimal ARIMA order by iterating through different combinations of p, d, and q values to find the model that minimizes the AIC. This optimal ARIMA model was used in subsequent analyses.

3.3.2 ARIMA Model Errors

The residuals generated by the ARIMA model were analyzed to capture the nonlinear patterns that the ARIMA model could not account for. Figure 2 illustrates the errors generated by the ARIMA model.

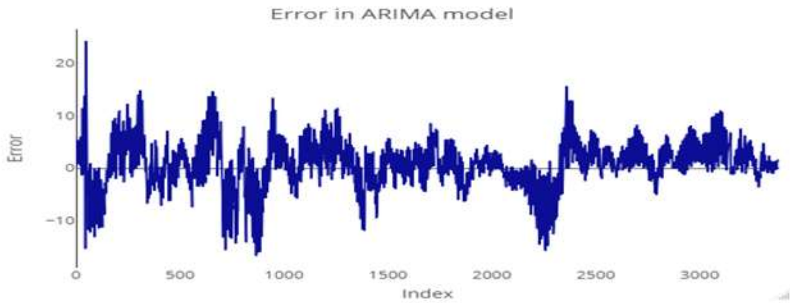

Figure 2: Residuals from the ARIMA model predictions

The minimum residual error was 16.61, the maximum was 24.1621, and the mean error was 0.8166, with a variance of 20.5176. These residuals were used as input for the ANN model to capture the nonlinear components of the time series.

3.3.3 Predicted Test Data

The predicted test data, compared across ARIMA, ANN, and the hybrid model, is displayed in Figure 3. The hybrid model provides more accurate predictions than both standalone ARIMA and ANN models.

3.3.4 RMSE Change with Forecast Horizon (Test Data)

Figure 4 shows the changes in RMSE values for the test data predictions over increasing forecast horizons.

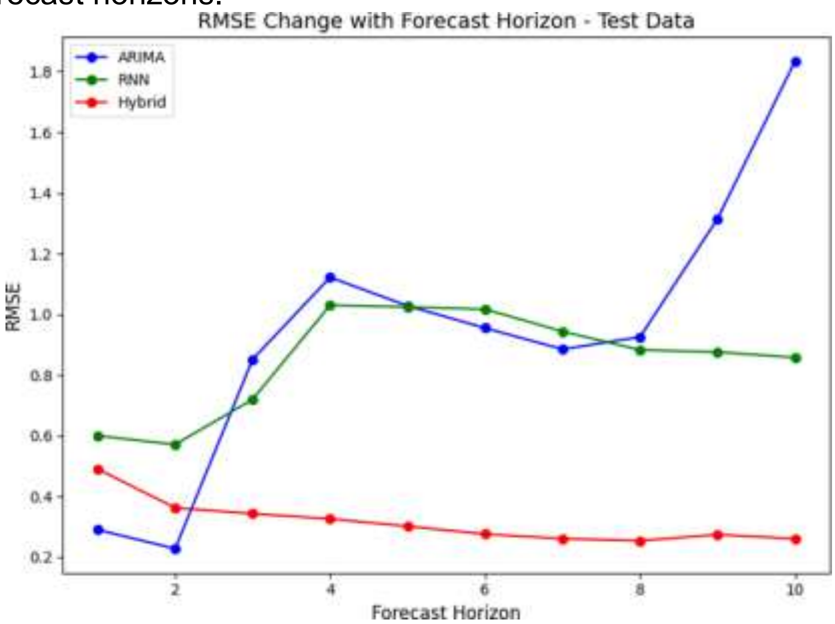

Figure 4: RMSE change with forecast horizon (Test Data)

As seen in Figure 4, while the ARIMA model performed reasonably well for short-term forecasts, its accuracy declined as the forecast horizon increased. In contrast, the hybrid model consistently outperformed both ARIMA and ANN across all forecast horizons due to its ability to model both linear and nonlinear components.

3.3.5 Forecasted Data Comparison

The validation data was used to evaluate the models' forecasting accuracy. Figure 5 compares the predicted values of the models with the actual values.

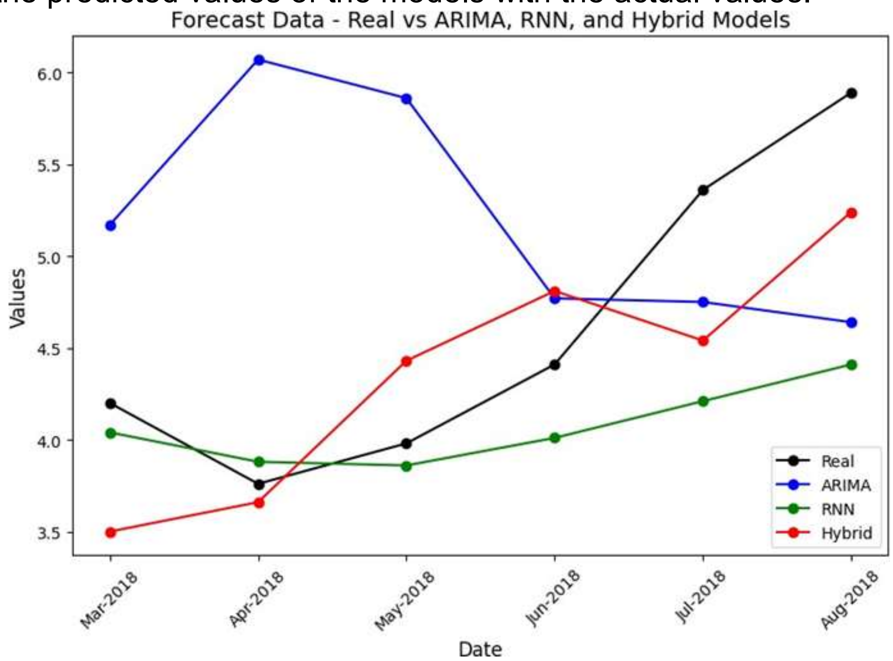

Figure 5: Comparison of Forecasted Data with Actual Values

In Figure 5, the hybrid model provided more accurate predictions compared to ARIMA and ANN models, particularly in the second half of the validation data, where the ARIMA model failed to capture the data's nonlinear structure.

3.3.6 Comparison of Forecasted Data

Table 2 provides a comparison of the forecasted data, highlighting the MAPE for each model.

Table 2: Comparison of Forecast Data (Validation Set)

Date	Real Data	ARIMA	ANN	Hybrid
Mar-2018	4.20	5.17	4.04	3.50
Apr-2018	3.76	6.07	3.88	3.66
May-2018	3.98	5.86	3.86	4.43
Jun-2018	4.41	4.77	4.01	4.81
Jul-2018	5.36	4.75	4.21	4.54
Aug-2018	5.89	4.64	4.41	5.24
MAPE		28.18	10.85	9.27

According to Table 2, the hybrid model achieved the lowest MAPE (9.27) compared to the ARIMA model (28.18) and ANN (10.85). This further supports the superior forecasting performance of the hybrid model.

3.3.7 RMSE Change with Forecast Horizon (Validation Data)

Figure 6 illustrates how RMSE values changed with increasing forecast horizons for the validation data

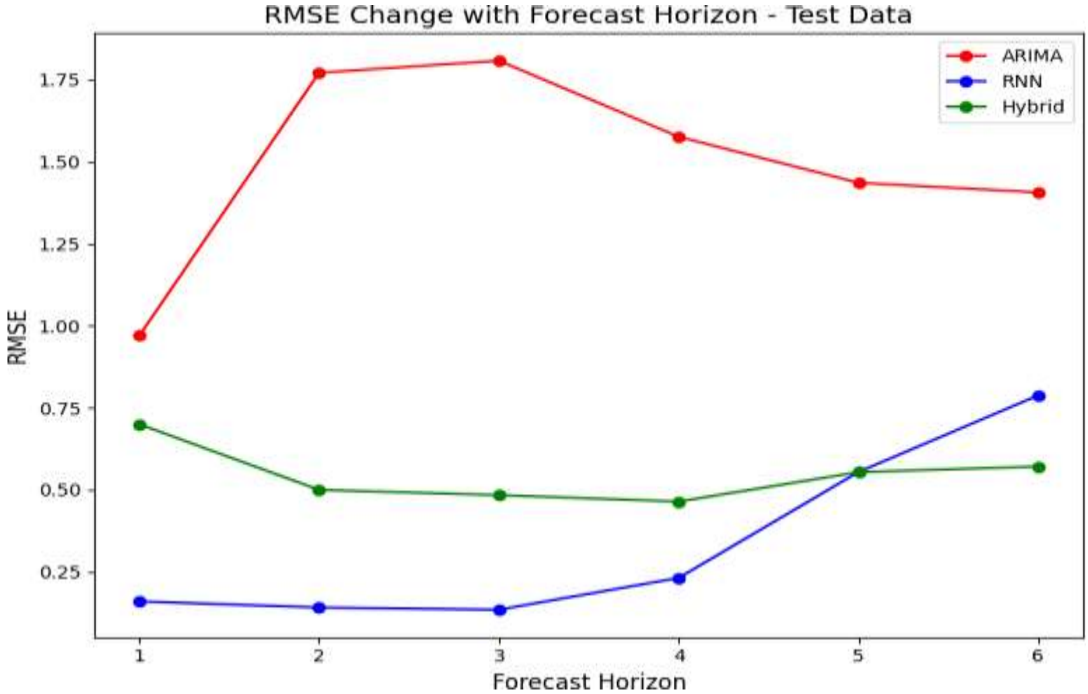

Figure 6: RMSE changes with forecast horizon (Validation Data)

Figure 6 demonstrates that, similar to the test data, the hybrid model outperformed both the ARIMA and ANN models for longer forecast horizons in the validation data set. The hybrid model’s ability to adapt to both linear and nonlinear components allowed it to achieve lower RMSE values overall.

The findings from this study validate the effectiveness of the Hybrid ARIMA-ANN model in capturing both the linear and nonlinear components of the inflation rate time series. While ARIMA alone effectively modeled the linear trends, it struggled with nonlinear variations. Conversely, the ANN model was more effective in identifying nonlinear dynamics but less robust in dealing with linear trends.

The Hybrid model, combining both techniques, showed significantly improved performance across all evaluation metrics, including MAPE and RMSE. This aligns with earlier research (Zhang, 2003; Cadenas & Rivera, 2009), confirming that a decomposition-based hybrid approach enhances forecast accuracy. Particularly for longer forecast horizons, where ARIMA performance degraded sharply, the Hybrid model maintained lower RMSE values, showcasing its adaptability and robustness.

Furthermore, residual analysis confirmed that ARIMA failed to account for nonlinearities that were later effectively captured by the ANN model. The final forecast, which combined ARIMA's linear forecast and ANN's nonlinear adjustments, led to a marked performance gain. This result suggests that policymakers and economic analysts in Sri Lanka can benefit significantly from using such hybrid models to forecast inflation and make data-driven decisions.

4. Conclusion

This study explored the use of a hybrid ARIMA-ANN model to forecast the monthly inflation rate in Sri Lanka from 1988 to 2018. The motivation for this approach stemmed from the limitations inherent in both the ARIMA and ANN models when applied separately. While ARIMA models are proficient in capturing linear relationships within time series data, they fall short when faced with nonlinear patterns. On the other hand, ANN models are highly capable of modeling nonlinear dependencies but may struggle with capturing linear trends. Recognizing these individual strengths and weaknesses, the study integrated both models into a hybrid framework, aiming to improve overall forecasting accuracy.

The evaluation of the models was conducted using both test and validation datasets. The results clearly demonstrated that the hybrid model outperformed the standalone ARIMA and ANN models in terms of Mean Absolute Percentage Error (MAPE) and Root Mean Square Error (RMSE). For instance, in the test dataset, the hybrid model achieved a significantly lower MAPE of 3.37 compared to 18.50 for ARIMA and 12.48 for ANN. Similar superiority was observed in the validation data, where the hybrid model maintained a MAPE of 9.27, outperforming both ARIMA and ANN once again. These findings reinforce the capability of hybrid models to handle complex time series data that embody both linear and nonlinear patterns.

In addition to superior predictive performance, the hybrid model also displayed robustness across varying forecast horizons. Whereas ARIMA's accuracy declined notably with increasing forecast horizon, the hybrid model consistently maintained lower RMSE values, indicating greater adaptability in long-term forecasting. This highlights its practical value in real-world applications, particularly for policymakers and financial analysts who rely on accurate inflation forecasts to make informed decisions.

The study suggests that hybrid models such as the ARIMA-ANN approach offer a more comprehensive and reliable framework for time series forecasting in complex economic contexts. Future research may consider incorporating more advanced neural network architectures, data preprocessing techniques like scaling, and smoothing enhancements to further improve forecasting performance and computational efficiency.

References

Bandara, W. M. S., & De Mel, W. A. R. (2024). Evaluating the Efficacy of Supervised Machine Learning Models in Inflation Forecasting in Sri Lanka. *American Journal of Applied Statistics and Economics*, *3*(1), 51–60. https://doi.org/10.54536/ajase.v3i1.2385

Box, G. E. P., & Jenkins, G. M. (1976). *Time series analysis: forecasting and control*. Holden-Day.

Cadenas, E., & Rivera, W. (2009). Wind speed forecasting in three different regions of Mexico, using a hybrid ARIMA-ANN model. *Renewable Energy*, *34*(12), 2732–2738.

Chollet, F., & others. (2018). *Keras: The Python Deep Learning library*. https://keras.io

Haykin, S. (2009). *Neural Networks and learning machines* (3rd ed.). Pearson Education.

Hyndman, R. J., & Athanasopoulos, G. (2014). *Forecasting: principles and practice*. OTexts.

Lin, T.-Y., RoyChowdhury, A., & Maji, S. (2015). Bilinear CNN Models for Fine-grained Visual Recognition. *Proceedings of the IEEE International Conference on Computer Vision*, 1449–1457.

Mankiw, N. G. (2014). *Principles of Economics*. South-Western Cengage Learning.

Mishkin, F. S. (2010). *The Economics of Money, Banking, and Financial Markets*. Pearson.

Rathnayaka, R. M. K. T., & Wijesinghe, G. W. R. I. (2020). ARIMA and ANN approach for forecasting daily stock price fluctuations of industries in Colombo Stock Exchange, Sri Lanka. *Proceedings of the 5th International Conference on Information Technology Research (ICITR 2020)*, 1–7. https://doi.org/10.1109/ICITR51448.2020.9310826

Tong, H. (1990). *Non-linear time series: a dynamical system approach*. Oxford University Press.

TradingView. (2023). *ECONOMICS-LKIRYY*. https://www.tradingview.com/symbols/ECONOMICS-LKIRYY/

Zhang, G. P. (2003). Time series forecasting using a hybrid ARIMA and neural network model. Neurocomputing, *50*, 159–175.

Zhang, G., Patuwo, B. E., & Hu, M. Y. (2008). Forecasting with artificial neural networks: The state of the art. *International Journal of Forecasting*, *14*(1), 35–62.