J. Stat. Appl. Pro. **15**, No. 3, 579-597 (2026)

Journal of Statistics Applications & Probability
An International Journal

http://dx.doi.org/10.18576/jsap/150315

Evaluating Forecast Accuracy: A Comparative Error-Based Analysis of ANN and Markov-Switching Models in Inflation Prediction

Ntebogang Moroke[1,*], *Katleho Makatjane*[2] *and Claris Shoko* [3]

[1]Department of Business Statistics and Operations Research, North West University, North West, South Africa
[2]Department of Statistics and Population Studies, University of the Western Cape, Cape Town, South Africa
[3]Department of Statistics, Demographic and Population Sciences, University of Botswana, Gaborone, Botswana

Received: 12 Dec. 2025, Revised: 1 Feb. 2026, Accepted: 9 Mar. 2026

Abstract: In this study, we develop a hybrid modelling approach that integrates artificial neural networks with a Markov-switching autoregressive model to enhance the accuracy of inflation predictions, utilising monthly data from South Africa from 2009 to 2023. Ten statistical loss functions, including mean square error (MSE), mean absolute error (MAE), root mean square error (RMSE), mean absolute percentage error (MAPE), symmetric mean absolute percentage error (SMAPE), mean absolute scaled error (MASE), and Theil's U, are used to measure how well the forecast works. The MS(2)-AR(1)-ANN(5,3) performed the best overall out of all the competing specifications. The model has a mean squared error of 0.9897, a mean absolute error of 0.7573, a root mean squared error of 0.9948, and a mean absolute scaled error of 0.115. The model further yields Theil's U statistic of 3.568, where, out of the 100 loss functions used, it ranks first in eight, second in the mean prediction error, and third in the symmetric mean absolute percentage error, with an SMAPE score of 18.234. This score is slightly higher than some of the base learners, but scale-free measures like MASE are better at providing trustworthy advice when conditions change quickly. The findings show that percentage-based measures like MAPE have their limits and that MASE is a better predictor during times of structural change. In general, the results show that the hybrid MS-AR-ANN architecture produces inflation projections far more accurate than those from the separate base models. The suggested methodology offers valuable insights for policymakers and central banks seeking early-warning indicators and robust inflation-monitoring systems amid regime upheavals and economic turmoil.

Keywords: Artificial Neural Networks, Error Metrics, Forecast Accuracy, Inflation Forecasting, Markov-Switching Autoregressive Models. Nonlinear Time Series, South Africa

1 Introduction

Forecasting inflation is seen as a crucial component of monetary policy [1], and implementing this policy requires a thorough understanding of, and forecasting for, inflation above all other factors. Central banks acknowledge that predicted future inflation is a crucial factor influencing the formulation and implementation of monetary policy, where they have an informational advantage over the public regarding economic conditions, making inflation forecasting a comparative advantage of the institution [2]. Maintaining a low inflation rate promotes long-term economic development because high inflation erodes the real purchasing power of money holdings, making it harder for people to purchase goods and services unless nominal earnings change proportionately. After 1994, the South African government took responsibility for a persistently high inflation rate that exceeded 10% annually. Given these high inflation rates, the substantial rise in the prices of food and basic household necessities such as water and heating has had a regressive effect on the elderly and lower-income members of society. After the South African Reserve Bank (SARB) adopted an inflation-targeting policy, the country's inflation rate rose from 3% to 8.6% in early 2007, exceeding the target range of 3% to 6%. Instead of reducing inflation, the ongoing increase in the Repo rate has made it even worse, and [3] has noted that this period is marked by high production levels, a persistent decline in prices, and a reduction in business and

---

[*] Corresponding author e-mail: kmakatjane@uwc.ac.za

consumer confidence metrics. From 2009 to 2025, South Africa's inflation has mostly stayed in the single digits. However, the economy has had to contend with numerous structural shocks, including changes in global commodity prices, currency rate volatility, droughts that raise food costs, and the COVID-19 pandemic. These disturbances have significantly altered how inflation works, making South Africa a valuable place to test forecasting approaches that account for both stability and volatility. Also, the country's official inflation-targeting policy provides a clear policy benchmark against which to judge the accuracy of our forecasts, making this study more useful. Hence, our primary objective in this study is to evaluate the efficacy of various error measures for assessing forecasting models, notably Artificial Neural Networks (ANNs) and Markov-Switching Autoregressive (MS-AR) models, using monthly inflation data from South Africa. Unobserved components models (UCMs) and ARIMA (herein, Autoregressive Integrated Moving Average) are commonly employed and demonstrate efficacy in various contexts; however, their linear characteristics limit their ability to capture the discrete regime shifts and nonlinear trends observed in South African inflation during the study period.

The Markov-switching model enables modelling of structural breaks and non-linear behaviour in South African inflation by allowing parameter values to change across regimes. A key reason this type of model is appropriate for an emerging country like South Africa is that the country has undergone policy-driven structural changes, such as the South African Reserve Bank's implementation of formal inflation targeting at the beginning of 2000, as well as exogenous shocks. In this case, the Markov-Switching model does not require prior knowledge of when regime switches occur; instead, it identifies them on its own. Therefore, the Markov-Switching model can identify transitions from a low-inflation state to a high-inflation state. When in a low-inflation state, which would be characterised by inflation rates below the 3-6% upper bound of the South African Reserve Bank's inflation target, there is typically stability both within and outside the target range due to well-anchored expectations and relative volatility stability. Conversely, during a high-inflation regime, which would be characterised by higher volatility caused by various factors, such as currency devaluation, administered pricing, and global events, such as the 2008 global financial crisis [4], global events exert greater influence. Furthermore, empirical results suggest that, especially after the global financial crisis of 2007-2008, South Africa experienced smaller and less volatile episodes of high volatility, indicating greater credibility among policymakers as emphasised by [5]. As a result, the Markov-switching model enables us to capture the dynamic and state-dependent aspects of inflation dynamics with much greater success than the linear models, which are commonly used. Nonetheless, by identifying these regimes as latent states, the Markov-switching model captures the changing and state-dependent nature of inflationary dynamics. As a result, a two-regime MSM enables researchers to differentiate between periods of stable and turbulent economic conditions, to provide additional information on structural pressures, and to evaluate monetary policy under differing economic conditions. However, ANNs, on the other hand, enhance this by adeptly capturing intricate nonlinear correlations and seasonal patterns that other linear models, such as ARIMA and exponential smoothing, often overlook. The ANN concerning [6] also improve forecast quality compared with statistical or econometric methods. The first way this improvement occurs is by recognising and capturing the complexity of relationships in the data. For example, it has been shown that recurrent models such as Long Short-Term Memory networks and feed-forward architectures are able to learn much more complex and dynamic relationships than linear methods. Another way an ANN improves forecast quality is by managing multiple types of seasonality that other models find difficult to handle. For example, while models like Holt-Winters exponential smoothing are based on the assumption that there are few, consistent seasonal patterns, ANNs are able to handle complex, possibly irregular and changing, seasonal patterns. Also, an ANN learns to map inputs to outputs through a purely empirical process; therefore, ANNs do not require a previously specified theoretical model [7]. This makes them particularly useful when working empirically with data from processes whose theoretical model is either completely unknown or too complex. Because of their abilities listed above, numerous studies have demonstrated that ANNs achieve higher forecast accuracy and lower forecasting errors than linear and other nonlinear models, such as Generalised Autoregressive Conditional Heteroscedasticity (GARCH), in many areas of application, including but not limited to forecasting financial markets, energy demand, and macroeconomic forecasting.

The combination of Markov-switching and Artificial Neural Network methods enables the integration of regime identification into enhanced pattern recognition, improving predictive performance under rapidly changing economic conditions. Therefore, the Markov-switching model includes an Autoregressive structure and has been used as a two-regime MS-AR model. This is preferable to regime-switching models with a simple structure that focus on changes in the mean or variance, for example, basic Markov-switching mean models or pure Markov-switching volatility models, since the MS-AR charter not only captures short-run dependencies between observations within each regime but also accounts for regime-dependent shifts; therefore, it is well-suited to inflation processes in developing countries such as South Africa, where both persistent inflation and structural breaks in the dynamics determine the system's overall behaviour. In comparison to the Markov-switching GARCH models that focus primarily on modeling regime dependent volatility behaviour, the MS-AR model is better suited to this task because the main goal of our study is to identify and model regime-dependent changes in the conditional mean process and short-run inflation persistence, instead of trying to completely define a conditional heteroscedasticity structure. Although MS-GARCH models are very effective at modelling detailed volatility clusterings and tail-risk behaviours, [8] indicated that they add unnecessary parameters and computational burdens to the analysis, especially when analysing multiple hybrid models. Since other methodologies and evaluation measures address volatility behaviour, the use of the MS-AR model in this study offers a more parsimonious, easier-to-interpret structure while maintaining essential regime-switching properties.

Another gap in the literature relates to the choice of error metrics for model evaluation. Widely used measures such as Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Mean Absolute Percentage Error (MAPE) differ in their sensitivity to outliers (i.e., scale dependence) and in their behaviour under low inflation values [9]. Recent research highlights the advantages of scale-independent metrics like Mean Absolute Scaled Error (MASE), yet few empirical applications in emerging market contexts systematically compare its performance against conventional measures – particularly under structural breaks and high volatility. Moreover, it has been documented that the mean absolute percentage error (MAPE) is a popular error metric for quantifying prediction accuracy, as it does not rely on scale and is easy to understand. However, MAPE has the major problem that it yields infinite or undefined results when the actual values are zero or very close to zero. To address this problem in MAPE, [10] proposed a new procedure to assess prediction accuracy and called this error metric the mean arctangent absolute percentage error (MAAPE). In essence, [10] emphasised that MAAPE is a slope as an angle, while MAPE is a slope as a ratio, considering a triangle with adjacent and opposite sides that are equal to an actual value and the difference between the actual and forecast values, respectively. A significant amount of research has been conducted on predicting the short-term power output of wind farms, as they are becoming an increasing share of total power production. For power networks of varying sizes to perform successfully, precise plans must be established. The review and analytical study by [11] primarily focuses on a statistical analysis of forecasting error measures from more than 100 studies on wind generation estimates. The factors that affect the size of forecasting errors are listed and studied, with the main error measures being normalised root mean squared error (nRMSE) and normalised mean absolute error (nMAE). The authors have also proposed a new, unique error dispersion factor (EDF), defined as the ratio of nRMSE to nMAE. The variability of EDF, contingent on the selected wind farm criterion size, forecasting time frames, and the type of forecasting approach used, has been examined. This is a new study that examines error metrics in wind farm electricity output forecasts, which has never been done before.

A thorough and comparative study of predicting solar energy production in Morocco used six machine learning (ML) algorithms, namely Support Vector Regression (SVR), Artificial Neural Network (ANN), Decision Trees (DT), Random Forest (RF), Generalised Additive Model (GAM), and Extreme Gradient Boosting (XGBOOST). This study utilised daily data from a solar power plant in Benguerir, Morocco. And this work, [12] employed four metrics to assess the models' performance: root mean squared error, mean absolute error, mean absolute scaled error, and R-squared. The results indicated that the ANN model is the best predictive model for energy forecasting in similar situations, as it has the lowest RMSE and MSAE and the highest squared value, all of which are important performance requirements for the ANN model. These authors' findings not only show that the ANN algorithm works, but they also provide the correct settings for attaining the best results when predicting how much solar energy is to be produced, and they deliver relevant information that can be utilised in real life by figuring out the optimal approach to setting up the ANN algorithm. This improves solar energy systems and helps build a sustainable future, particularly by putting this data on an edge device to forecast when photovoltaic power plants need repairs. Despite this, performance measurements, particularly error measures, are crucial components of assessment systems in several fields. A study by [13] sought to offer a comprehensive overview of diverse performance metrics and their classification methodologies. The principal aim of this study is to develop a typology that enhances the understanding of metrics, facilitating their selection for machine learning regression, forecasting, and prognostics. Based on an examination of their structure, the authors suggest applying performance measurement in four groups: main metrics, extended metrics, composite metrics, and hybrid sets of metrics. The research revealed three principal components that affect the structure and characteristics of major metrics: the methodology for calculating point distance, the normalisation technique, and the aggregation approach for point distances within a dataset.

In a comparative analysis of statistical methods, machine learning, and deep learning algorithms for time series forecasting, [14] identified Exponential Smoothing, Prophet, a hybrid Autoregressive Integrated Moving Average-Generalised Autoregressive Conditional Heteroscedasticity (ARIMA-GARCH) model, K-Nearest Neighbours (KNN), and Long-Short Term Memory (LSTM) as the primary techniques. This comparison is based on a dataset of US inflation from 1965 to 2021, and examines univariate time series for short-term forecasting, and uses several metrics to measure model performance, such as Mean Squared Error, Mean Absolute Error, Median Absolute Error, and Root Mean Squared Error. According to the numerical comparison, Exponential Smoothing performed best, while KNN came in a close second. The results show that both models accurately forecast inflation in the United States. The ARIMA-GARCH, LSTM, and Prophet models, on the other hand, did not perform well. Nevertheless, to forecast global inflation rates, [15] evaluates two methodologies: gradient-boosted regression and the Autoregressive Integrated Moving Average models. These authors discovered that ARIMA performed better in cross-validation using the same approach, with RMSE, MAE, and standard deviation values of 2.53, 2.21, and 0.48, respectively. Gradient-boosted regression, on the other hand, performs better when evaluated on both the training and test datasets. It has a root-mean-square coefficient of 0.078, an MAE of 0.27, and a MSE of 0.22. This shows that it could be useful for making predictions. The study focuses on short-term forecasts of global inflation rates to mitigate susceptibility to prolonged political and economic disturbances. Because linear time series models such as ARIMA and more sophisticated nonlinear GARCH models rely on stationarity, they are limited in their ability to model nonlinear relationships and sudden regime changes. Furthermore, another study that has not evaluated the error measures while using the econometric models, deep learning and or the combination of the two is found in the empirical analysis of [16] who introduced a new combined forecasting model that uses the clear structure of an Exponential Smoothing Recurrent Neural Network and the creative features of a Variational Autoencoder to predict the risk of falling stock prices for Sasol Limited from 2010 to 2025. To assess the performance of the newly developed ensemble model, the authors used short-, medium-, and long-term forecasting horizons and further evaluated MSE, MAE, Forecast Error Proportion (FEP), and Theil Inequality Coefficient (TIC or U-statistic). As in other studies, [16] have never compared the efficiency of these metrics in selecting models.

1.1 Research Highlights and Key Findings

In this study, we investigate the modelling of inflation forecasting by combining MS-AR models with ANN and utilising South African inflation data as a test case. The main goal is to evaluate how various error measures affect the assessment of prediction performance across fundamentally heterogeneous models. We therefore capture both sudden structural changes and complex temporal dynamics in inflation behaviours by leveraging the regime-identification capabilities of MS-AR and the nonlinear pattern-recognition strengths of ANNs. The model provides point estimates and forecasts across various economic regimes, thereby analysing the stability and volatility of inflation over time. Importantly, we show that choosing an error metric, such as MSE, RMSE, MAE, MAPE, or MASE, significantly affects how forecasting models are ranked and understood. For instance, particularly under turbulent conditions, MASE is more resilient and scale-independent, whereas RMSE may benefit models that penalise significant deviations. These results are significant for legislators, as they enable more precise inflation forecasts and more informed monetary decisions. The main study highlights and summarises the fundamental results in Table 1 by drawing on comprehensive empirical studies.

Table 1: Research Highlights and Key Findings

Characteristic	**Highlights**	**Key Findings**
Forecasting Models	Evaluation of ANN and MS-AR models individually and in combination	Combined MS(2)-ANN(5,3) model delivers superior predictive accuracy compared to single models
Error Metric Comparison	Extensive analysis using MSE, RMSE, MAE, MAPE, MASE, Theil's U, and others	MASE found to be the most robust, scale-independent, and interpretable error metric
Inflation Case Study	Empirical application to South African inflation data (2009–2023)	Revealed structural breaks and volatile inflation patterns across economic regimes
Regime Detection	Captures economic regime shifts using Markov-Switching AR modelling	Regime one is characterised by lower volatility and higher stability than regime two
Nonlinear Pattern Learning	ANN models capture seasonal trends and complex data structures	ANN(5,3) improves forecast accuracy, especially when residuals and seasonal features are used
Combined Model	Incorporates regime probabilities from MS-AR into ANN inputs	Enhanced learning across regimes and improved performance in volatile conditions
Model Assessment	Treats error metrics as diagnostic tools, not just statistical summaries	Error metric selection significantly affects model comparison and interpretation
Policy Implications	Supports inflation-targeting in emerging markets through early warning tools	Provides policymakers with a more accurate forecasting system for volatile inflation periods

2 Methodology

The methodology section serves as the study's core structure, providing a comprehensive and methodical guide for analysing error-based evaluations of predictive accuracy in forecasting models. The connection between ANNs, error metrics, and Markov Switching Autoregressive models illustrates a dynamic interplay in predictive modelling, offering insights into model assessment and complexity. Artificial neural networks are well known for finding complex patterns in data [17], and they have been widely utilised across several forecasting disciplines. Error metrics are necessary to assess how well forecasting models perform, since they provide observable indicators of the difference between projected and actual values and help improve forecasting structures and model performance, as in [18]. At the same time, MS-AR models provide an alternative approach to predicting time series. These models, according to [19], account for regime transitions, which occur when the mechanism producing the data changes over time, and further accommodate dynamics that conventional linear models cannot capture, including non-stationarity and structural changes. The MS-AR models are therefore used in economic and financial forecasting to identify changes in market conditions, business cycles, and other factors that are not readily apparent.

The relationship among these three components becomes particularly interesting when considering complex forecasting scenarios, such as those involving financial time series data. The ANNs capture intricate patterns and nonlinear relationships within such data, potentially outperforming linear models. Error metrics then provide a quantitative charter for comparing the predictive accuracy of ANNs against other models, including MS-AR models. In this scenario, MS-AR models introduce a layer of complexity by considering regime shifts and non-stationarity. When evaluating their performance, error metrics tailored to time-series data analysis, such as MAPE and MASE, provide insights into how well the MS-AR model captures transitions between different regimes [20]. The synergistic use of ANNs and MS-AR models, along with error metrics, is particularly relevant in scenarios where data exhibits complex temporal patterns and dynamic changes. By comparing the performance of ANNs and MS-AR models using appropriate error metrics, researchers and practitioners can choose the most suitable model for a given forecasting task, considering factors such as model complexity, data characteristics, and the ability to capture regime shifts. In essence, the methodology section harmoniously integrates choices, processes, and procedures, guiding the investigation into error-based assessment of predictive accuracy in ANN forecasting models. This comprehensive approach ensures robustness, replicability, and a profound understanding of the potential and limitations of deep learning within forecasting, setting the stage for transformative insights. Finally, the preceding sections comprehensively detail the methodologies and steps undertaken for the meticulous analysis of the data.

2.1 Markov Switching Autoregressive (MS-AR) Model

We proceed and fit a Markov Switching Autoregressive model, which, according to [19], is a valuable tool for capturing regime changes in time series data. This model extends the Autoregressive (AR) model by allowing parameters to switch between regimes governed by an underlying Markov process. The general expression of this model, according to [21], is given by

$$X_t = C_{st} + \phi_1, \phi_{i,st} - 1 + \phi_2, \phi_{i,st} - 2 + \cdots + \phi_{p,i,st} - p + e_t, \tag{1}$$

where, C_{st} represents the intercept of the AR process in regime S_t, $\phi_{i,st}$ are the autoregressive coefficients of lag i in regime S_t. Furthermore, S_t is an unobserved Markov state variable at time t indicating the regime at that time, and e_t is the white noise error term at time t. The Markov state variable S_t follows a discrete-time Markov process with transition and initial probabilities that govern the switching behaviour between regimes. A two-regime transition probability matrix is then given by

$$P_{ij} = \begin{bmatrix} P_{11} & 1 - P_{11} \\ 1 - P_{22} & P_{22} \end{bmatrix}$$

where, [22,23] suggested that an underlying two-regime Autoregressive model denoted $MS(k) - AR(p)$ where k is the number of regime shifts and p is the order of AR parameters is given by

$$X_t = \begin{cases} \phi_1 + \sum_{i=1}^{p} \varphi_1 X_{t-i} e_{1,t}, S_t = 1 \\ \phi_2 + \sum_{i=1}^{p} \varphi_2 X_{t-i} e_{2,t}, S_t = 2 \end{cases}. \tag{2}$$

Finally, Figure 1 represents the schematic workflow of the MS-AR in this study. The upper section depicts the transitions of the latent Markov chain, while the lower boxes specify the Autoregressive equations and innovations ε_t unique to each regime.

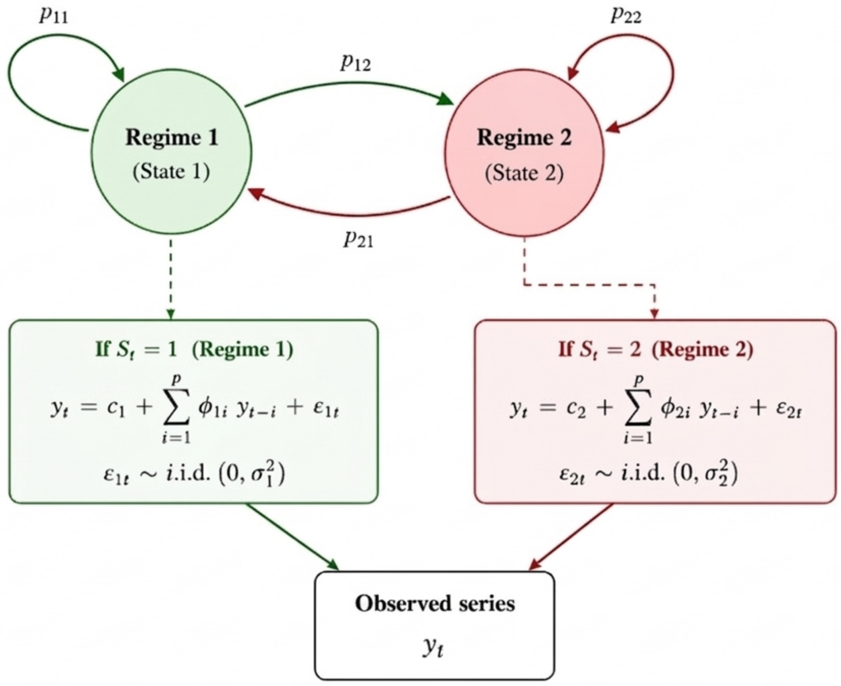

Figure 1: Schematic representation of a two-regime Markov-switching Autoregressive model. The diagram illustrates the transition dynamics between latent states (S_t) and the regime-dependent $AR(p)$ processes for the observed series y_t. **Source** Authors Computation

2.2 An artificial neural network (ANN) Framework

The structure and operation of biological (brain) neurons form the basis of the artificial neural network model, a novel computing technique. Like the human brain, an ANN can learn from examples; hence, it is a subset of artificial intelligence (AI), a field that encompasses computational models used to study intelligent learning. A learning process is used to program the ANN for a particular function, such as pattern recognition or data classification. Thus, a neural network is mathematically composed of a large number of straightforward, intricately connected processing units that are highly accurate. Any kind of parametric or non-parametric process is easily modelled using ANNs, which also automatically and effectively transform the input [24].

Let y_t be the set collecting the predictors at time t. In this study, we consider predictor sets of the form $y_t = \left(x_t, \cdots, x_{t-(B-1)}, z_t, \cdots, z_{t-(B-1)}\right)^T$, where the number of lags B may differ across sets. We now assume that the h-step-ahead inflation y_{t+1} evolves nonlinearly with respect to the predictors y_t. Mathematically, y_{t+1} is modelled as a nonlinear function of the predictors, G plus a non-predictable component ε_t that is assumed to be independently and identically distributed (i.i.d) with zero mean and variance σ^2 and it is independent of y_t, where

$$y_{t+h} = G(y_t; \Theta_h) + \varepsilon_{t+h} \tag{3}$$

and Θ_h represents the model's parameters. The underlying statistical problem consists of estimating the unknown function $G : y_t \to y_{t+h}$, where G takes the form of a neural network structure, in which case fitting the function to the data corresponds to estimating Θ_h. Nonetheless, the feed-forward architecture is the main part of a neural network, which is based on how the human brain works. The core concept behind this paradigm is rather simple: it comprises several basic components, called nodes, arranged into layers. The input layer is the first layer, and it contains the information that goes into the model. The output layer is therefore the last layer and provides the model's prediction; the layers in between

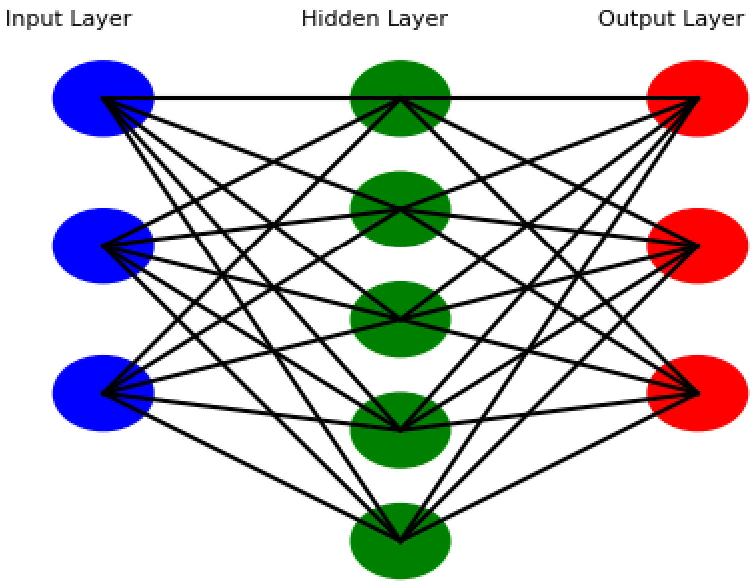

Figure 2: Feedforward Neural Network Architecture with 5 Hidden Layers. **Source** Authors Computation

are called hidden layers, and information flows in one direction, from inputs to outputs. This network, with respect to [25], is deep if it has many hidden layers, typically 2 to 8. Each node in the network, except those in the input layer, processes information from the preceding layer and transmits it to the following layer. The computation behind these units, according to [26], is a weighted average of the outputs from all nodes in the previous layer, followed by a nonlinear modification, the activation function. The weights assigned to each node are the model's parameters.

By considering a feed-forward neural network with Q hidden layers and letting y_t be the predictor set and $\alpha_t^i = \in \Re^{n\times 1}$ be vectors of hidden layers containing n nodes each, for $i = 1,2,3,\cdots,Q$. The constraint that every layer must have the same number of nodes, n, is imposed for simplicity, but it could be relaxed, in which case grid search would be used to choose the layer-specific number of nodes n as indicated by [27]. The feed-forward model takes the form of

$$G\left(y_t;\Theta_h\right) = gFF\left(y_t\right) \tag{4}$$

where, gFF denotes a feed-forward neural network with inputs y_t and can be expressed as follows

$$\begin{aligned} & gFF\left(y_t\right) = \omega_{Q+1}\alpha_t^Q + \beta_{Q+1} \\ & \alpha_t^i = ReLu\left(\omega_i\alpha_t^{i-1} + \beta_i\right), i = 1,\cdots,Q \\ & \alpha_t^0 = y_t \end{aligned} \tag{5}$$

where, $\Theta_h = \left(\{\omega_i\}_{i=1}^Q, \{\beta_i\}_{i=1}^Q\right)^T$ collects the parameters of the model. $\{\omega_i\}_{i=1}^Q$ is a vector of parameters relating the different layers of the network, $\{\beta_i\}_{i=1}^Q$ are intercept terms, and $ReLu:\Re\rightarrow\Re$ is the rectified linear unit activation function given by $ReLu\left(z\right) = \max\left(0,z\right)$, and it is applied element-wise. The hyperbolic tangent and sigmoid functions are other activation functions, but the $ReLu$ remains the best choice because its gradient is straightforward and does not cause estimation issues. In reality, the $ReLu$ function generally does better than other activation functions when it comes to both statistical performance and computing cost. A recent study by [28] demonstrates that the estimator of a deep network may attain convergence rates that are virtually optimum under various restrictions imposed on the target function, therefore, our suggested feedforward neural network design is shown in Figure 2.

2.3 Integration of MS and ANN

Analysing the outputs and predictions of the integrated Markov Switching (MS) and Artificial Neural Network model provides insights into the respective contributions of each component [29]. This analysis examines how the MS component captures regime changes and abrupt shifts in the data [19], while the ANN component captures complex patterns and relationships that traditional linear models may miss [30]. The regime probabilities or indicators extracted from the MS model provide valuable information on the likelihood that the data belong to specific regimes at each time point [31]. By combining these indicators with the ANN's predictions, analysts can discern how different regimes influence the model's overall forecasts [32]. This aids in identifying periods of stability and transitions between different states in the data, such as economic cycles, market regimes, or other hidden patterns [33].

2.3.1 Markov Switching Model

Let $Z_t$represent the latent regime at time t. The transition probabilities from regime j to regime k are denoted by π_{jk}.

$$\Pr\left(Z_t = k|Z_{t-1}\right) = \pi_{jk} \tag{6}$$

The observed data y_t can be expressed as

$$y_t = \mu_k + \varepsilon_t \tag{7}$$

where $\pi_{jk}$is the transition probability from regime j to regime k. $y_t$is the observed data at time t. μ_k is the mean of regime and finally $\varepsilon_t$is the error term.

2.3.2 Artificial neural network

In the context of a neural network, the predicted $\hat{y}_t$ at time t can be represented as

$$\hat{y}_t = f\left(W \bullet x_t + b\right) \tag{8}$$

where: $\hat{y}_t$ is the predicted value at time t, W is the weight matrix while $x_t$is the input vector. Moreover, b is the bias term and f is the activation function.

2.3.3 Integration

The integrated input I_t at time t combines regime indicators R_t from the MS model with additional features x_t from the ANN

$$I_t = [R_t, x_t] \tag{9}$$

where R_t represents regime indicators and x_t is the input vector from the ANN. Notes analysing the integrated model involve the following four steps: (1) examine how regime indicators influence predictions and transitions. (2) Identify if certain regimes have more impact during specific time periods. (3) Understand how the ANN captures intricate patterns and non-linear relationships. (4) Assess the integrated model's ability to adapt to varying regimes. By scrutinising the outputs and understanding the interplay between the MS and ANN components, analysts gain deeper insights into the data's behaviour and the model's ability to capture its nuances.

2.4 Selection of error metrics for assessment

The choice of suitable error metrics is essential for assessing the efficacy of forecasting models. Error metrics quantitatively assess prediction accuracy and evaluate the model's ability to capture underlying data patterns; hence, this study presents relevant error metrics about the topic. The Markov-Switching Autoregressive and Artificial Neural Network models employed in this study are trained by minimising squared-error loss to estimate the conditional mean of inflation rates in South Africa. Therefore, the primary error metrics most suitable for assessing forecast accuracy are those derived from squared errors, including Mean Squared Error (MSE) and Root Mean Squared Error (RMSE). Nevertheless, metrics like Mean Absolute Error (MAE) and Mean Absolute Scaled Error (MASE) serve as supplementary measures to offer further insight into model performance, although they do not align directly with the models' estimation objectives. Moreover, we recognise that metrics with minimal variation across models do not effectively differentiate performance; therefore, our analysis meticulously evaluates the relative informativeness of each metric in measuring forecasting quality.

Mean Squared Error (MSE) is a simple and common measure of error in predictive modelling. It serves as a reliable indicator of forecast precision by quantifying the magnitude of discrepancies between forecasts and actual observations. The MSE is the average of the squared differences between the predicted and actual values; squaring the differences makes larger errors stand out, highlighting how gaps between expectations and actual outcomes can have an impact. The Mean Squared Error, therefore, gives a broad picture of how accurate the model is by summing all squared errors and dividing by the total number of observations. It is calculated as follows.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_t)^2 . \tag{10}$$

In model 10, n is the number of observations, while y_i represents the actual observed value at time t and $\hat{y}_t$ represents the predicted value at time t.

Root Mean Squared Error (MSE) is a basic and widely used error statistic in predictive modelling. It serves as a reliable metric for evaluating overall prediction accuracy by measuring the extent of differences between projected values and actual observations. The MSE is the mean of the squared discrepancies between the expected and actual values. The squaring of differences amplifies errors, highlighting the impact of discrepancies between expectations and actual results. By aggregating these squared errors and dividing by the total number of observations, the Mean Squared Error provides a comprehensive assessment of the model's precision. Mathematically, the Root Mean Squared Error is given by

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_t)^2} \tag{11}$$

Mean Absolute Error (MAE) is the measure that tells how far off the projected values are from the real values on average. It is less affected by outliers than MSE and provides a more balanced picture of prediction accuracy [34]. This test is a basic error metric that measures the average absolute difference between projected values and actual observations. Unlike Mean Squared Error, MAE is robust to outliers and provides a fair assessment of forecast accuracy. The formula for MAE is given as follows.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_t| . \tag{12}$$

Mean Absolute Percentage Error (MAPE) is an important error measure that computes the average percentage difference between expected values and actual observations, based on the actual values themselves. This statistic is quite useful when one needs to know exactly how much the forecast errors differ from the actual numbers [9]. The mathematical formulation of MAPE is given as follows.

$$MAPE = \frac{100}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_t}{y_t}\right| \tag{13}$$

Symmetric Mean Absolute Percentage Error (SMAPE) is a valuable error metric that enhances percentage-based evaluation by balancing overestimation and underestimation. This symmetry enables SMAPE to provide a well-rounded and thorough evaluation of prediction accuracy [35], and mathematically, this error metric is expressed as

$$SMAPE = \frac{100}{n}\sum_{i=1}^{n}\frac{|y_i - \hat{y}_t|}{(|y_i| + |\hat{y}_i|)/_2} \tag{14}$$

The symmetric property of SMAPE ensures the absence of bias in overestimation or underestimation. This equitable evaluation renders SMAPE an optimal choice for assessing prediction accuracy without biasing one type of error over another. It is especially applicable in situations where overestimation and underestimation have equivalent repercussions, such as demand forecasting or resource allocation.

Percentage Error (PE) is a crucial notion in predictive modelling and forecasting. It describes the persistent and systematic differences between expected values and actual observations across a set of data points; hence, the forecast bias refers to a forecasting model's persistent tendency to overestimate or underestimate actual outcomes [**?**]. Forecast bias, therefore, extends beyond random error; it identifies a regular pattern in forecast disparities, indicating whether a model tends to be overly optimistic or extremely pessimistic. Nevertheless, identifying the presence and extent of forecast bias is critical for developing trustworthy models that appropriately reflect underlying data dynamics. Therefore,

$$PE = \frac{(y_i - \hat{y}_t)}{y_i} \times 100. \tag{15}$$

Forecast Bias: a critical concept in predictive modelling and forecasting. It refers to the consistent and systematic deviations between predicted values and actual observations across a range of data points. In essence, forecast bias refers to a forecasting model's consistent tendency to overestimate or underestimate the true outcomes [36]. Forecast bias goes beyond sporadic errors that could occur randomly. Instead, it uncovers the persistent pattern in prediction discrepancies, revealing whether a model consistently leans towards being overly optimistic or overly pessimistic. Identifying the existence and nature of forecast bias is paramount for building reliable models that accurately represent the underlying data dynamics. Mathematically;

$$FB = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_t). \tag{16}$$

Detecting forecast bias is crucial because it can have profound implications for decision-making. A model that consistently overestimates values can lead to suboptimal resource allocation, misguided strategic planning, and flawed risk assessment. Recognising forecast bias is the first step towards rectifying- and enhancing the accuracy of predictions.

Mean Forecast Error (MFE) represents the average discrepancy between predicted values and actual outcomes. This metric quantifies the average error irrespective of the error direction [37]. Mean Forecast Error is essential for evaluating the effectiveness of forecasting models; it measures the mean discrepancy between predicted values and actual observations, highlighting the average magnitude of error regardless of error direction. This metric differs from other error metrics by focusing exclusively on the magnitude of deviations, regardless of error direction (e.g., overestimation or underestimation). This characteristic, therefore, renders MFE a direction-agnostic metric, providing insights into the overall accuracy of predictions without distinguishing the nature of errors. The MFE statistic is articulated as follows

$$MFE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_t) \tag{17}$$

The interpretability of MFE is straightforward: a lower MFE indicates that the forecasts are, on average, closer to the actual values. In contrast, a higher MFE indicates a greater average discrepancy between forecasts and observations. The MFE is a useful statistic when the main goal is to understand the average error size without getting into the complexities of directional errors.

Mean Absolute Scaled Error (MASE) compares the accuracy of the forecasting model to that of a naive (simple) forecasting method. It accounts for error magnitudes and data seasonality [38]. It is a robust and intuitive evaluation metric designed to measure the accuracy of forecasting models relative to a naive baseline, as emphasised by [9]. Unlike some metrics that can be sensitive to scale and outliers, MASE offers a standardised way to assess forecast accuracy, making it suitable for various types of time series data, and it is computed by

$$MASE = \frac{MAE}{\frac{1}{n-1}\sum_{i=2}^{n}|y_t - y_{t-1}|}. \tag{18}$$

MASE's normalisation procedure makes it a scale-independent metric, rendering it highly valuable for comparing forecasting performance across different time series and domains [36]. A MASE value of 1 signifies that the model's predictions are, on average, as accurate as those of a simple naive model. Values below 1 indicate that the model outperforms the naive model, while values above 1 indicate less accurate predictions [39].

Theil's U Statistic combines measures of accuracy and bias to provide an overall assessment of prediction performance [40]. It is a versatile evaluation metric that goes beyond assessing accuracy alone; it also accounts for bias, offering a comprehensive view of predictive performance [40]. Developed by econometrician Henri Theil, this statistical measure has gained prominence as a powerful tool in the assessment of forecasting models. At its core, Theil's U-Statistic addresses two critical aspects of predictive evaluation, which are accuracy and bias. Accuracy refers to how close the predicted values are to the observed values, indicating how well the model captures the underlying patterns in the data. On the other hand, bias relates to any systematic deviation or tendency of the model's predictions from the actual values. Theil's U-statistic formula is given by

$$U = \sqrt{\frac{\frac{1}{n}\sum_{t=1}^{n}(y_t - \hat{y}_t)^2}{\frac{1}{n}\sum_{t=1}^{n}\left|\frac{y_t - \hat{y}_t}{y_t}\right|}} \tag{19}$$

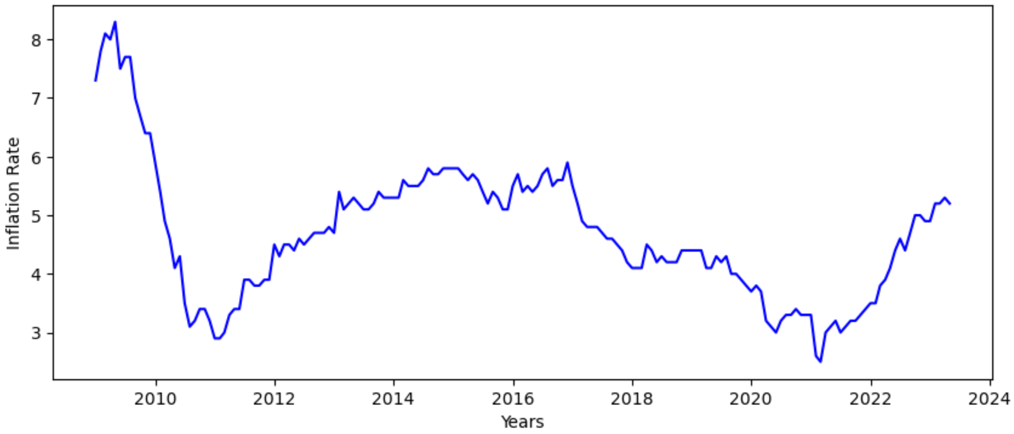

Figure 3: Time Series Plots of the Inflation Rates of South Africa

Theil's U-Statistic strikes a balance between accuracy and bias, making it a useful instrument for evaluating the effectiveness of forecasting models. A reduced U value indicates enhanced predictive performance, indicating a model with diminished relative root mean squared error and relative mean absolute error. Conversely, an elevated U-value indicates a greater discrepancy between the model's predictions and the actual values. In this study, a lower error signifies the model's proximity to actual values, reflecting a better fit [41]. Reduced error implies improved trend capturing, leading to accurate predictions [42]. Lower error enhances decision-making reliability, minimising incorrect choices [43]. Lower-error models yield valuable insights and successful outcomes, thereby benefiting business performance [36].

Descriptive statistics reveal the characteristics of time series data. The mean represents the central value, while the median handles extreme values robustly. Standard deviation quantifies variability [36], complemented by variance's portrayal of mean-squared deviations. The range highlights the spread of data, while the interquartile range (IQR) is resistant to outliers [44]. Skewness and kurtosis disclose distribution shape. Skewness shows tail asymmetry [41], and kurtosis indicates tail heaviness compared to the normal distribution. Amalgamating these statistics yields a multidimensional view that guides analyses, modelling, and informed decision-making. This synthesis empowers robust forecasting and data-driven choices.

3 Application to South African Inflation Rates

This portion of the research presents practical data analysis of inflation rates in South Africa. The dataset encompasses monthly inflation rates spanning from January 2009 to May 2025, resulting in a comprehensive sample size of 202 data points. To improve the accuracy of inflation rate predictions for South Africa, we used a Markov-Switching autoregressive model and an artificial neural network (ANN). According to the insights drawn from Figure 3, the time series visualisation reveals noteworthy oscillations in both upward and downward directions, accompanied by distinct seasonal patterns. Notably, a significant peak emerges between December 2009 and January 2010, a period when South Africa surpassed the upper threshold of the inflation target range for technical reasons. However, after this breach, inflation consistently re-established itself within the target range from February 2010 onwards [45].

[46] highlights six events that they think may have contributed to these periods of increased inflation, providing more background. The repeal of the Glass-Steagall Act by the Clinton administration in 1999, which transformed the regulatory environment for commercial and investment banking, is notable, as noted by Global Economist [47]. Other identified causes include the rise in subprime mortgages, changes to banking laws, the development of sophisticated financial products, changes in interest rates, and global financial imbalances [46]. Additionally, we found that the average inflation rate in South Africa (i.e., the mean) is around 4.67%. Therefore, on average, inflation has been somewhere between quite high and moderate during our sample period. Since the SARB uses an inflation target to guide monetary policy, its inflation-targeting regime appears to have performed fairly well. As the standard deviation of the inflation time series is roughly 1.13, there appears to have been some volatility in the inflation rate, though not excessive. That being said, three common tests for assessing whether data follow a normal probability distribution, namely the Jarque-Bera, the Kolmogorov-Smirnov, and the Shapiro-Wilk tests, all reject the assumption of normality at the 1 per cent confidence level. Thus, we now state with a very high degree of certainty that the distribution of inflation rates under study is non-normal,

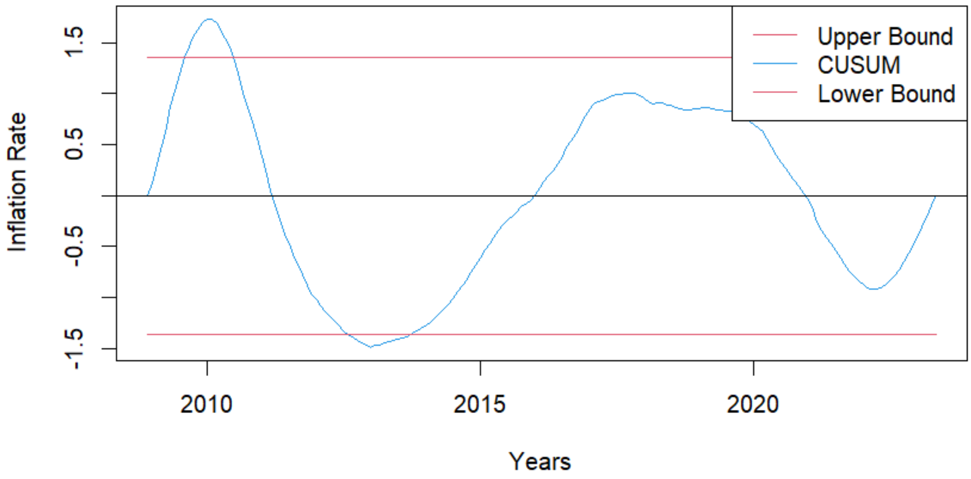

Figure 4: CUSUM of AR(1) model for Inflation rate

most likely reflecting both asymmetry and fatter tails. In econometric terms, this result means that any model used to forecast future inflation needs to account for potential deviations from normality, since normality-based models will fail to capture extreme inflation movements. Moreover, this result supports the use of more flexible modelling approaches designed to accommodate nonlinearities and or structural changes. Finally, from a policy perspective, non-normality in inflation implies that extreme inflation outcomes are more likely than they would be under a normal distribution. Our results of non-normality are further supported by the Cumulative Sum (CUSUM) in Figure 4, which reveals periods of volatility between 2010 and 2015 as shown by the sharp peak above the upper bound and a dip below the lower bound. This indicates structural instability in the inflation rate during this period. Visual examination highlights the intrinsic volatility of inflation rates in South Africa, as the plot exceeds the upper 95% limit band.

3.1 Fitting Markov-switching Autoregressive Model

The training and test datasets are separate from the dataset containing South Africa's inflation rates. In particular, the training dataset has 182 observations and covers the period from January 2009 to October 2023. On the other hand, the test dataset, which consists of 35 observations, spans the time period from November 1, 2023, to May 31, 2025. Nevertheless, our approach involves fitting an Autoregressive model within a dual-regime framework. Initially, we employed a flexible Fourier-based nonlinear unit root test introduced by [48]. This test was applied to the time series data to assess stationarity. The outcomes of the Fourier nonlinear unit root test provide substantial evidence of a nonlinear unit root in the South African inflation rate dataset. Consequently, this signifies the attainment of a stationary nonlinear time series [48]. Subsequently, we employed a set of six Autoregressive models with a lag length subject to two regimes, specifically ranging from $AR(1)$ to $AR(6)$ within the context of an $MS(2)-AR(1)$ framework. To determine the most parsimonious lag length, various criteria were utilised, including the final prediction error (FPE), predicted residual error sum of squares (PRESS) as proposed by [43], Bayesian information criterion (BIC) introduced by [49], and Akaike information criterion (AIC) as introduced by [50]. Based on the assessment of these criteria, a lag length of 1 was identified as the most parsimonious choice. Subsequent steps involved the application of a stationary $MS(2)-AR(1)$ model. The results of this modelling approach are presented in Table 2.

Table 2: Two-Regime MS-AR Model Estimates

Parameter	Coefficient	Std. Error	t-value	p-value
Regime 1				
$\hat{\mu}_1$	0.0190	0.0157	1.2108	0.0060
$\hat{\phi}_1$	0.0328	0.1078	0.3043	0.0009
$\hat{\sigma}_1$	0.0218	0.0057	3.8026	0.0001
Regime 2				
$\hat{\mu}_2$	-0.1260	0.0921	-1.3675	0.0005
$\hat{\phi}_2$	-0.0217	0.2165	-0.1003	0.0001
$\hat{\sigma}_2$	0.1342	0.0366	3.6623	0.0002
Transition Probabilities				
	$p_{11} = 0.7176$			$p_{12} = 0.1444$
	$p_{21} = 0.2824$			$p_{22} = 0.8556$

The two identified regimes possess distinct economic interpretations. The variance in regime one is 11.24% lower than that in regime two. The volatility in Regime 2 is six times higher than in Regime 1. Comparable findings are reported by [51] in their investigation of an early warning system for inflation utilising Markov-switching and logistic modelling approaches. The data indicate that South African inflation exhibits significant volatility and is prone to regime shifts, with a monthly rate of 3.66 per cent. However, during the first regime, the monthly average inflation rate decreased to 1.9 per cent, indicating a very low level. This means that when the series is in regime one, its probability to switch to regime two is $\Pr(S_t = 2|S_t = 1) = 28.244\%$. Table 3 further shows that there is a lower chance of maintaining the lower regime $p_{11} = 71.756\%$ than of maintaining the upper regime $p_{22} = 85.560\%$. Transitions to a lower regime are higher than transitions to a higher regime, $p_{21} > p_{12}$. The average duration of each regime further corroborates this behaviour, where regime two is expected to last 149 months, accounting for 76.02%, with an average duration of approximately 11 months. In contrast, regime one is projected to last 47 months, representing 23.98%, with an average duration of approximately 3 months. This demonstrates that regime 2 is more enduring than regime 1. In other words, the economy often stays in the negative growth stage longer than it does in the stable one. In conclusion, significant regime shifts in South African inflation rates are evident, as demonstrated by the filtered and smoothed probabilities presented in Figure 5.

Model diagnostics indicate that all parameter estimates are significant at the 5 per cent level. Moreover, the estimated Q-statistics provide substantial evidence that the estimated model conforms to a white-noise pattern with a nonlinear process, thereby yielding a root mean square error of 12.414 and a mean absolute error of 10.0257.

3.2 Combined Forecast Model

The ANN models utilise mean squared error (MSE) loss for training, which aligns with estimating the conditional mean of the inflation series. The initial configurations of the artificial neural network utilise solely the regime-specific residuals derived from the MS(2)-AR(1) model as input variables. Multiple architectures are evaluated, including ANN(2,1), ANN(3,1), ANN(3,2), ANN(4,2), ANN(4,3), ANN(5,2), and ANN(5,3). When exclusively trained on regime-residual inputs, all architectures yield validation RMSE values exceeding 0.232. Additional features that capture temporal and seasonal dynamics are incorporated to enhance predictive performance, where eleven monthly dummy variables represent seasonal effects, and a linear time index captures long-term temporal behaviour. The incorporation of these features significantly lowers validation RMSE values to under 0.150 across all architectures, and it is found that ANN(4,3), ANN(5,2), and ANN(5,3) demonstrate validation RMSE values of 0.138826, 0.139243, and 0.11015, respectively, reflecting notable enhancements compared to models that do not incorporate seasonal and temporal data. According to the minimum validation RMSE criterion, the final model selected is ANN(5,3). This architecture achieves an out-of-sample prediction accuracy of 94.98%, compared with 93.6% for the ANN(4,3) model. The findings indicate that incorporating seasonal structure and temporal dynamics improves the forecasting performance of neural network models for inflation prediction.

Table 3 presents forecasting performance results for three models: MS(2)-AR(1), ANN(5,3), and the hybrid MS(2)-AR(1)-ANN(5,3) model. The hybrid model outperforms the other models on most measures, demonstrating the effectiveness of integrating regime-switching processes with artificial neural network approaches. With regards to the squared errors, it is worth noting that the MS(2)-AR(1)-ANN(5,3) model has the lowest mean squared error of 0.9897 and RMSE of 0.9948, indicating that its predictions stray the least from actual observed values implying that the model efficiently captures both the central tendency and variance of the time series, reducing substantial deviations that hurt decision-making. Moreover, the MS(2)-AR(1)-ANN(5,3) furthermore, in terms of mean absolute percentage error, mean absolute error, percentage error, mean absolute scale error and Theil's U, the MS(2)-AR(1)-ANN(5,3) still outperforms

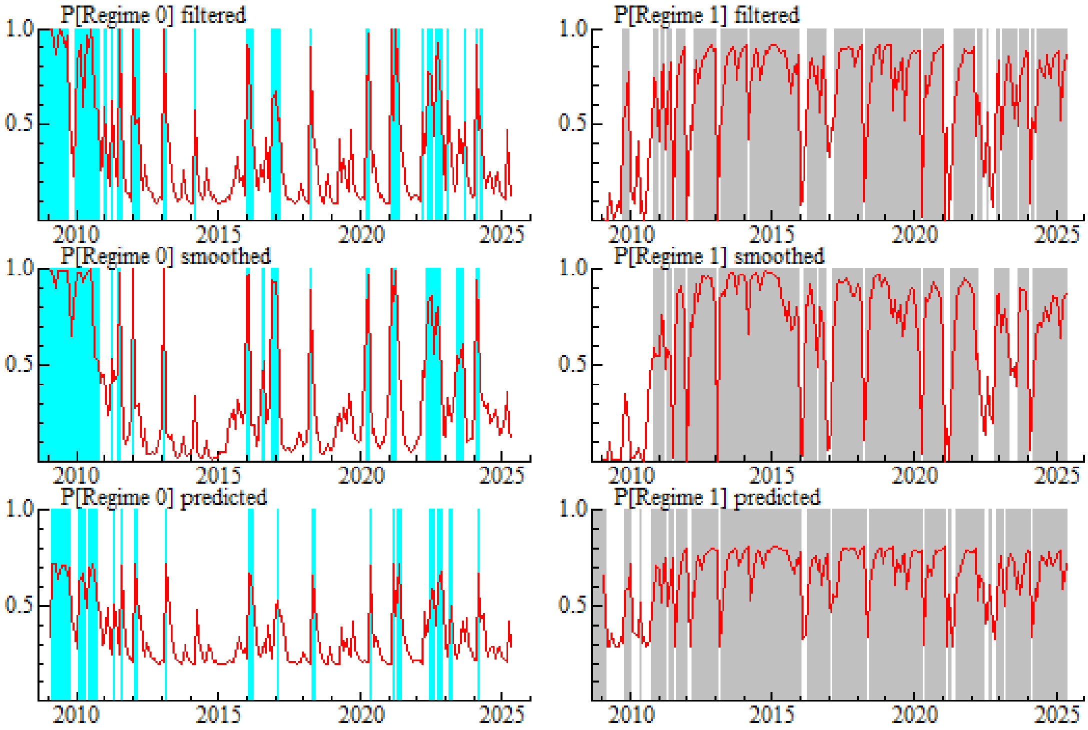

Figure 5: Filtered, smoothed, and predicted probabilities

the counter parts MS(2)-AR(1) and ANN(5,3) because it gives lower metrics in all of them. These metrics are especially useful because they account for the series's size and relative magnitude, making the model more resilient to heteroscedasticity and enabling meaningful comparisons across data magnitudes. Notably, the MASE value indicates that the model beats a naïve benchmark, validating its applicability in real-world forecasting scenarios. With regards to bias assessment, the MS(2)-AR(1)-ANN(5,3) model has minor forecast bias, as demonstrated by the Fractional Bias (0.0011), demonstrating that the model's predictions are symmetrically distributed around the observed values and do not routinely over- or under-predict. While the Mean Forecast Error of 0.050 is somewhat greater than ANN(5,3); this modest difference is offset by significant increases in other performance indicators. But with metrics that have relative weakness, the hybrid model has a larger symmetric mean absolute percentage error (SMAPE) of 18.234 than both MS(2)-AR(1) and ANN(5,3). This disparity is most likely caused by SMAPE's sensitivity to small denominators in the series or to excessive variation during volatile periods. Nonetheless, this shortcoming does not significantly detract from the model's overall excellence, especially when measured against strong, scale-independent measures. Finally, the comparative and practical implications emphasise that the integrated MS(2)-AR(1)-ANN(5,3) approach consistently outperforms the individual base learners in terms of absolute and relative metrics, demonstrating its ability to effectively capture both the structural regime changes inherent in the series and the complex nonlinear dynamics. The model's greater predictive power has important implications for policy and the economy. Accurate projections are critical for guiding inflation targeting, interest rate policy, and budgetary planning. The hybrid model's low error magnitudes and small bias reduce the risk of uninformed policy judgments, thereby increasing the credibility and effectiveness of economic planning.

Error metrics for evaluating forecasts are economically significant, as they influence modelling decisions and inform policy. The results of this study indicate that metrics produce substantially different evaluations of the same model. Therefore, if one prefer a model based upon RMSE or MAE, it does not necessarily mean that the model will be the best when utilising MAPE or Theil's U. These differences occur due to the unique statistical characteristics of each error metric, where RMSE is primarily biased toward large forecast errors and can create volatility that does not exist; whereas MAPE becomes unreliable as actual values decrease toward zero, which occurs frequently in South Africa with low inflation. On the other hand, MASE offers a scale-independent and statistically valid method for comparing model performance and creating comparable evaluations at varying time horizons. For an economist, these implications have direct influences on inflation targeting, determining interest rates and making budgetary plans, where credible use of

Table 3: Forecasting Performance of the Models

Metric	MS-AR	ANN(5,3)	MS-AR-ANN(5,3)	Rank (Best=1)
MSE	1.057	1.231	0.9897*	1: Hybrid
MAPE	18.726	18.010	17.9901*	1: Hybrid
MAE	0.813	0.973	0.7573*	1: Hybrid
SMAPE	15.444*	16.449	18.234	MS-AR
PE	4.992	5.250	4.7898*	1: Hybrid
FB	0.011	0.002	0.0011*	1: Hybrid
MFE	0.008*	0.075	0.050	MS-AR
MASE	0.174	0.169	0.115*	1: Hybrid
Theil U	4.434	4.0128	3.568*	1: Hybrid
RMSE	1.028	1.1095	0.9948*	1: Hybrid

forecasts by policymakers relies on the reliable evaluation of forecasts. Thus, the results support the need to employ a multi-error-metric charter, which increases transparency, reduces selection bias, and ensures that the chosen model reflects true predictive ability rather than artefacts introduced by selecting a single loss function. Additionally, because the MS-AR-ANN model consistently outperforms on multiple metrics, especially those that are scale-free, unbiased, and volatile, there is substantial evidence that it possesses better predictive capabilities than alternative models. As a result, the Hybrid Model's performance is not dependent on the specific error metric utilised, supporting its applicability to real-time macro-finance forecasting in environments sensitive to inflation.

Table 4: Sensitivity Analysis of Forecasting Metrics and Model Selection Influence

Metric	Scale Sensitivity	Outlier Sensitivity	Volatility Sensitivity	Bias Sensitivity	Low Denom. Sensitivity	Model Selection
MSE	High	High	High	Low	None	Hybrid
RMSE	Very High	Very High	High	Low	None	Hybrid
MAE	Medium	Medium	Medium	Low	None	Hybrid
MAPE	High	High	Medium	Low	Very High	Hybrid*
SMAPE	Medium	Medium	High	Low	High	MS-AR
PE	Medium	Medium	Medium	Low	None	Hybrid
FB	Low	Medium	Low	Very High	None	Hybrid
MFE	Low	Medium	Low	Very High	None	MS-AR
MASE	Low	Low	Low	Low	None	Hybrid**
Theil U	Medium	High	High	Medium	Medium	Hybrid

*Sensitive to values near zero. **Identified as the most robust specification.

The empirical evidence presented here illustrates how the choice of metrics for evaluating a model affects its performance. The differences in preferences among models, based upon the types of errors they are evaluated against, suggest that the loss function is important because it can emphasise different characteristics of predictive performance, e.g., the sensitivity of predictions to extreme values, scale-dependent behaviour, and bias. Consequently, using a single performance measure is likely to yield incorrect conclusions, especially if the data is characterised by significant tail heaviness or volatility clustering. Further, the stability of MASE relative to other error measures implies that evaluations should be based on measures that are insensitive to scale and or distributional characteristics of the data, particularly when evaluating multiple models with potentially differing characteristics. Finally, the instability of error measures such as MAPE due to low-denominator values highlights the need to consider the data's statistical properties when choosing appropriate error-measurement criteria. Moreover, Figures 6 depict the temporal response of the metrics across different conditions. Significant variations in model rankings across metrics reflect high sensitivity, whereas stable rankings indicate low sensitivity and thus enhanced reliability of the metric in differentiating model performance. The empirical results indicate that most metrics, especially those sensitive to scale, outliers, and volatility, consistently support the hybrid MS(2)-AR(1)-ANN(5,3) model. This model demonstrates a forecasting error distribution characterised by lower variance, reduced bias, and enhanced stability across regimes. In contrast, metrics like SMAPE and MFE, which are sensitive to volatility symmetry and directional bias, sometimes favour the MS-AR model, indicating nuanced variations in conditional mean behaviour during regime shifts.

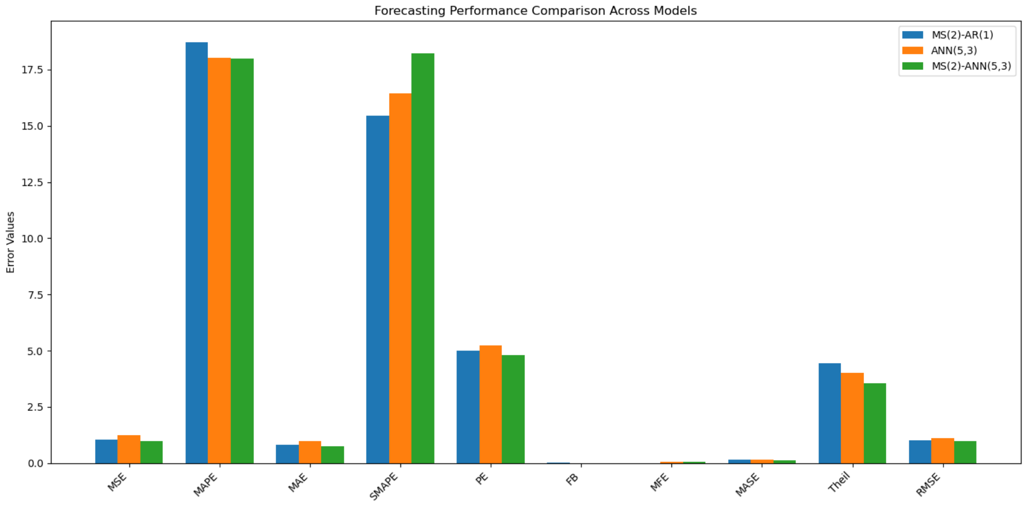

Figure 6: Sensitivity of Error Metrics across Forecasting Models

Figure 7 shows the fitted conditional mean that the MS(2)-AR(1)-ANN(5,3) model came up with. This visualisation shows that the hybrid specification closely tracks the observed inflation trend over the entire sample period. The fitted values accurately depict both the long-term trend and short-term variations in inflation, effectively capturing sudden shifts, pivotal moments, and transient anomalies that often pose difficulties for linear or solely regime-switching models. This tight fit shows that the hybrid model effectively combines structural regime dynamics with nonlinear learning, allowing it to better mimic the real-world behaviour of inflation than its parts. The visual evidence further backs up the results reported earlier in Table 3, showing that the hybrid MS(2)-AR(1)-ANN(5,3) model is a very good approach to show how inflation works in South Africa.

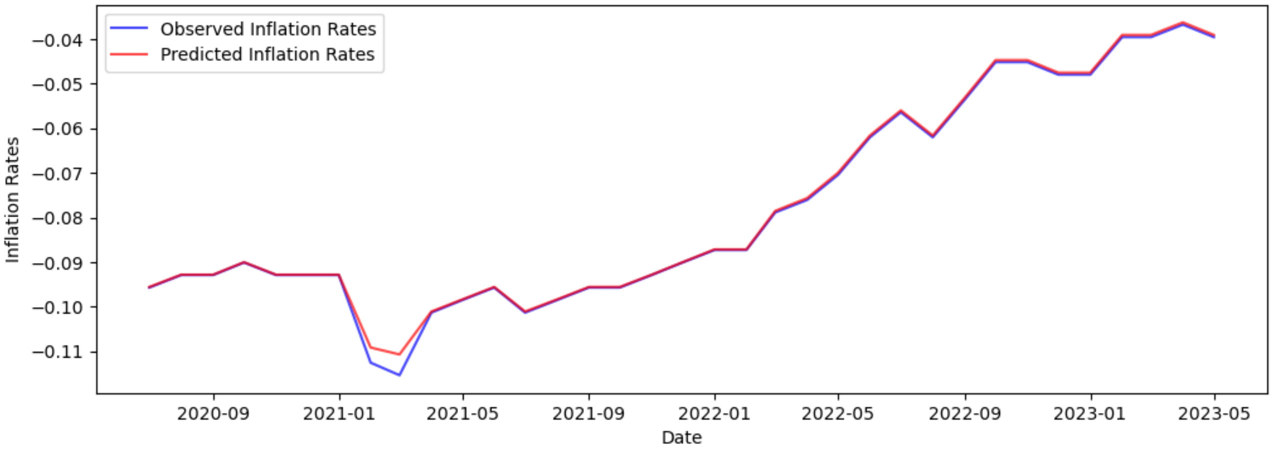

Figure 7: Observed Inflation Rates VS Predicted Inflation Rates in the testing set

From a macroeconomic standpoint, especially within the South African framework, the enhanced predictive efficacy of the hybrid MS(2)-AR(1)-ANN(5,3) model has significant implications. The economy of South Africa is marked by frequent changes in government, driven by currency rate fluctuations, commodity price changes, load-shedding events, and uncertainty about government policy. All of these make it hard to make short-term predictions. A model that accurately represents nonlinear dynamics and structural breakdowns improves the precision of forecasts for key macroeconomic variables, including inflation, interest rates, and monetary conditions. The South African Reserve Bank can do a better job of controlling inflation when it has more precise predictions for inflation and volatility. This lowers the risk of policy overshooting or delayed adjustments during times of high volatility. Also, stronger forecast stability during times of stress, such as when commodity prices drop, the rand loses value, or there are global risk-off events, helps the national treasury's fiscal authorities better plan for financing costs, debt-service risks, and budget execution limits. The model's lower bias and smaller errors mean that financial markets may make better decisions about how to allocate portfolios, manage risk, and price derivatives. This is especially true in fixed-income and currency markets, where volatility regimes are important. In general, the hybrid model is a statistically sound and economically useful way to address South Africa's structurally unstable environment, thereby improving the quality of macroeconomic planning and policy making.

4 Discussion, conclusion, and Recommendations

Forward-looking monetary policy is most successful when it is accompanied by precise and reliable inflation projections. Understanding inflation dynamics and forecasting their evolution across various macroeconomic scenarios is critical, since the South African Reserve Bank's core mandate remains the preservation of price stability. In this sense, inflation forecasting is critical in informing both monetary and fiscal policy choices. During periods of high uncertainty, such as the COVID-19 pandemic, reliable and timely predictions equip the SARB with the tools it needs to respond proactively to rising inflationary pressures and successfully anchor expectations. This study improves our knowledge of South African inflation dynamics by comparing the forecasting skills of three modelling methods: a Markov-Switching Autoregressive (MS-AR) model, an Artificial Neural Network (ANN), and a hybrid MS(2)-AR(1)-ANN(5,3) structure. The MS-AR model incorporates fundamental regime shifts in inflation behaviour, such as transitions from high- to low-volatility periods, which are common in countries exposed to global and local shocks. In contrast, the ANN model excels at detecting nonlinear correlations and seasonal patterns that standard time-series structures mostly miss. By incorporating MS-AR residuals into the ANN, the hybrid model leverages the strengths of both techniques, enabling deeper learning across diverse economic settings. Notably, it performed quite well during the tumultuous times associated with the COVID-19 shock, when single-model techniques often decline. None of the available published studies has combined the MS-AR model with the ANN architecture; to our knowledge, this is the first study to do so, specifically in the context of forecasting South African inflation rates. A study by [30] had combined a Markov Switching model for predicting stock trends with the Long Short-Term Memory (LSTM) model for predicting price fluctuations. The MS mode is said to be adept at adapting to changes across different market regimes, while the LSTM model can learn complex, long-term patterns in data due to its specialised neural network architecture.

The performance of the model is evaluated using a comprehensive set of forecasting measures, including MSE, RMSE, MAE, MAPE, and MASE. The investigation demonstrates that metric selection has a significant impact on model evaluation. Magnitude-based methods, such as MSE and RMSE, penalise significant deviations, but MASE enables a scale-free, robust assessment in both stable and dynamic environments. This emphasises the need not just to provide error measurements but also to critically understand them, since they affect model selection and policy implications [52]. Sensitivity analysis of the study further revealed that metric robustness varies with the economic situation, and consistent rankings across measures strengthen confidence in the combined model's supremacy. From a macroeconomic standpoint, the results have important policy implications for South Africa. The occurrence of regime-switching behaviour implies that inflation does not grow consistently across time, but rather alternates between periods of stability and instability; and [53] documented that several countries were hit by a series of macroeconomic shocks, most notably as a consequence of the COVID-19 pandemic and Russia's invasion of Ukraine, raising inflation rates to multi-decade highs and suspending well-documented macroeconomic relationships. Consequently, recognising these transitions helps policymakers prepare for abrupt shifts caused by external shocks, changes in global risk perception, domestic structural restrictions, or policy initiatives. But the ANN's ability to discover latent nonlinear patterns improves short-term prediction accuracy, enabling more flexible and responsive day-to-day policy choices. When combined, the hybrid model provides a comprehensive and robust framework for understanding inflationary dynamics, especially during periods of economic crisis.

The findings indicate that flexible modelling methodologies that incorporate both structural regimes and nonlinearities provide a more accurate and policy-relevant description of inflation in emerging market countries such as South Africa. Future research might expand on this work by incorporating multivariate applications that include interest rates, exchange rates, GDP growth, and global commodity prices, or by investigating extreme inflation concerns using tail-focused methodologies such as extreme value theory or copula-based dependent models. Such additions would provide a more comprehensive knowledge of joint macroeconomic behaviour and strengthen policy suggestions, particularly in a world characterised by global uncertainty.

Acknowledgments

The authors thank numerous people for their helpful comments on this paper.

Competing Interests

The authors declare that they have no competing interests

Declaration of generative AI

The authors hereby disclose that no generative AI tools were used to compile this research work.

References

[1] J. D. Fisher, C. T. Liu, and R. Zhou, When can we forecast inflation?, *Econ. Perspect. Fed. Reserve Bank Chicago* **26**(1), 32–44 (2002).

[2] S. Huseynov, V. Ahmadov, and S. Adigozalov, Beating a Random Walk: "Hard Times" for Forecasting Inflation in Post-Oil Boom Years? (2014). Available online: https://mpra.ub.uni-muenchen.de/id/eprint/63515 (Accessed on June 15 2025).

[3] K. Makatjane, N. Moroke, and D. Xaba, Prediction of inflation crises in South Africa using Markov-switching Bayesian VAR and logistic regression models, *J. Soc. Econ. Res.* **5**(1), 10–28 (2018). https://doi.org/10.18488/journal.35.2018.51.10.28

[4] P. Burger, Twenty-five years of inflation-targeting in South Africa: Going from 6% to 3%?, *WIDER Working Paper Series* No. 42/25, United Nations University World Institute for Development Economics Research (UNU-WIDER), Helsinki (2025). https://hdl.handle.net/10419/322173

[5] K. A. Sanusi and Z. Dickason-Koekemoer, Rand, rates, and returns: Unravelling the volatility nexus in South Africa's financial markets, *Journal of Risk and Financial Management* **19**(3), 230 (2026). https://doi.org/10.3390/jrfm19030230

[6] A. Kożuch, D. Cywicka, and K. Adamowicz, A comparison of artificial neural network and time series models for timber price forecasting, *Forests* **14**(2), 177 (2023). https://doi.org/10.3390/f14020177

[7] M. E. Bonilla Jr., J. McDonald, T. Toth, and B. Sadler, Conventional vs machine learning approaches: A comparison of time series modelling methods, *SMU Data Science Review* **7**(2), Article 2 (2023). https://scholar.smu.edu/datasciencereview/vol7/iss2/2

[8] E. Moyo, K. Makatjane, and R. Sivasamy, An improved volatility forecasting with GMLE-based MS-GARCH models: Evidence from BWP/USD, ZAR/USD, and BTC/USD, *Journal of Statistics Applications & Probability* **14**(6), 343–360 (2025). http://dx.doi.org/10.18576/jsap/140601

[9] R. J. Hyndman and A. B. Koehler, Another look at measures of forecast accuracy, *Int. J. Forecast.* **22**(4), 679–688 (2006). https://doi.org/10.1016/j.ijforecast.2006.03.001

[10] S. Kim and H. Kim, A new metric of absolute percentage error for intermittent demand forecasts. *International Journal of Forecasting*, **32**(3), 669-679 (2016). https://doi.org/10.1016/j.ijforecast.2015.12.003

[11] P. Piotrowski, I. Rutyna, D. Baczyński, and M. Kopyt, Evaluation metrics for wind power forecasts: A comprehensive review and statistical analysis of errors, *Energies* **15**(24), 9657 (2022). https://doi.org/10.3390/en15249657

[12] Y. Ledmaoui, A. El Maghraoui, M. El Aroussi, R. Saadane, A. Chebak, and A. Chehri, Forecasting solar energy production: A comparative study of machine learning algorithms, *Energy Reports* **10**, 1004–1012 (2023). https://doi.org/10.1016/j.egyr.2023.07.042

[13] A. Botchkarev, Performance metrics in machine learning regression, forecasting, and prognostics: Properties and typology, arXiv:1809.03006 (2018). https://doi.org/10.28945/4184

[14] Y. Jouilil and M. Iaousse, Comparing the accuracy of classical and machine learning methods in time series forecasting: A case study of USA inflation, *Stat. Optim. Inf. Comput.* **11**(4), 1041–1050 (2023). https://doi.org/10.19139/soic-2310-5070-1767

[15] G. Alomani, M. Kayid, and M. F. Abd El-Aal, Global inflation forecasting and uncertainty assessment: Comparing ARIMA with advanced machine learning, *J. Radiat. Res. Appl. Sci.* **18**(2), 101402 (2025). https://doi.org/10.1016/j.jrras.2025.101402

[16] C. Sigauke, N. Moroke, K. Makatjane, and C. Shoko. A deep learning forecasting of downside risk: application of a combined ESRNN-VAE. *Frontiers in Applied Mathematics and Statistics* **11** (2025): 1662252. https://doi.org/10.3389/fams.2025.1662252

[17] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, Learning representations by back-propagating errors, *Nature* **323**(6088), 533–536 (1986). https://doi.org/10.1038/323533a0

[18] R. Kieokaew, V. Haberle, A. Marchaudon, P. L. Blelly, and A. Chambodut, A novel neural network-based approach to derive a geomagnetic baseline for robust characterisation of geomagnetic indices at mid-latitude, *arXiv preprint* **2410.02311**(1)(2024). https://doi.org/10.48550/arXiv.2410.02311

[19] J. D. Hamilton, A new approach to the economic analysis of nonstationary time series and the business cycle, *Econometrica* **57**(2), 357–384 (1989). https://doi.org/10.2307/1912559

[20] S. Lima, A. M. Gonçalves, and M. Costa, Predictive accuracy of time series models applied to economic data: The European countries' retail trade, *Journal of Applied Statistics* **51**(9), 1818–1841 (2023). https://doi.org/10.1080/02664763.2023.2238249

[21] J. Durbin and S. J. Koopman, *Time Series Analysis by State Space Methods*, Oxford University Press (2012).

[22] M. T. Ismail and Z. Isa, Detecting regime shifts in Malaysian exchange rates, *Malays. J. Fundam. Appl. Sci.* **3**(2), 211–224 (2007). https://doi.org/10.11113/mjfas.v3n2.30

[23] C. J. F. Cruz and D. Mapa, An early warning system for inflation in the Philippines using Markov-switching and logistic regression models, *Theor. Pract. Res. Econ. Fields* **4**(2), 137–152 (2013). https://mpra.ub.uni-muenchen.de/id/eprint/50078

[24] S. H. Han, K. W. Kim, S. Kim, and Y. C. Youn, Artificial Neural Network: Understanding the Basic Concepts without Mathematics, *Dementia and Neurocognitive Disorders* **17**(3), 83–89 (2018). https://doi.org/10.12779/dnd.2018.17.3.83

[25] H. Z. Alemu, W. Wu, and J. Zhao, Feedforward Neural Networks with a Hidden Layer Regularisation Method, *Symmetry* **10**(10), 525 (2018). https://doi.org/10.3390/sym10100525

[26] S. Schmidgall, R. Ziaei, J. Achterberg, L. Kirsch, S. Hajiseyedrazi, and J. Eshraghian, Brain-inspired learning in artificial neural networks: a review, *APL Machine Learning* **2**(2) (2024). https://doi.org/10.1063/5.0186054

[27] L. Paranhos, Predicting inflation with neural networks, arXiv:2104.03757 (2021). https://doi.org/10.48550/arXiv.2104.03757

[28] C. Schmidt-Hieber, G. Toleikyte, L. Aitchison, A. Roth, B. Clark, A. Branco, T., and Häusser, M. Active dendritic integration as a mechanism for robust and precise grid cell firing, *Nat. Neurosci.* **20**(8), 1114–1121 (2017). https://doi.org/10.1038/nn.4582

[29] M. Bildirici and Ö. Ersin, Modelling Markov switching ARMA–GARCH neural networks models and an application to forecasting stock returns, *The Scientific World Journal* **2014**(1), 497941 (2014). https://doi.org/10.1155/2014/497941

[30] H. Hosseininesaz, *Master's Thesis*, University of Vaasa, School of Technology and Innovations, Industrial Systems Analytics, 67 (2024). Available online: https://osuva.uwasa.fi/items/4778ce7d-37e6-4322-88b4-1d10f375c3f9 (Accessed on November 23 2025).

[31] C. J. Kim and C. R. Nelson, State-Space Models with Regime Switching, *MIT Press* **1**(1) (1999). https://doi.org/10.7551/mitpress/6444.001.0001

[32] G. L. Hickey and D. Peña, The use of regime-switching models in time series analysis, *J. Time Ser. Anal.* **39**(4), 545–564 (2018).

[33] S. Hochreiter and J. Schmidhuber, Long short-term memory, *Neural Comput.* **9**(8), 1735–1780 (1997). https://doi.org/10.1162/neco.1997.9.8.1735

[34] C. J. Willmott and K. Matsuura, Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance, *Clim. Res.* **30**(1), 79–82 (2005). https://doi.org/10.3354/cr030079

[35] S. Makridakis, E. Spiliotis, and V. Assimakopoulos, The M4 Competition: 100,000 time series and 61 forecasting methods, *Int. J. Forecast.* **36**(1), 54–74 (2020). https://doi.org/10.1016/j.ijforecast.2019.04.014

[36] R. J. Hyndman and G. Athanasopoulos, *Forecasting: Principles and Practice*, OTexts (2018). Available online: https://otexts.com/fpp2/.

[37] L. J. Tashman, Out-of-sample tests of forecasting accuracy: An analysis and review, *Int. J. Forecast.* **16**(4), 437–450 (2000). https://doi.org/10.1016/S0169-2070(00)00065-0

[38] R. J. Hyndman and Y. Fan, Forecasting the global incidence of H1N1 influenza in 2009, arXiv:0912.5421 (2010). https://doi.org/10.48550/arXiv.0912.5421

[39] R. Koenker, *Quantile Regression*, *Wiley StatsRef* (2014). https://doi.org/10.1002/9781118445112.stat07557

[40] H. Theil, *Applied Economic Forecasting*, North-Holland Publishing (1966).

[41] D. C. Montgomery, E. A. Peck, and G. G. Vining, *Introduction to Linear Regression Analysis*, Wiley (2021).

[42] A. Gelman and J. Hill, *Data Analysis Using Regression and Multilevel/Hierarchical Models*, Cambridge University Press (2006).

[43] C. Chatfield, *Time-Series Forecasting*, Chapman and Hall/CRC (2000). https://doi.org/10.1201/9781420036206

[44] R. H. Shumway and D. S. Stoffer, *Time Series Analysis and Its Applications: With R Examples*, fifth edition Springer (2025). https://doi.org/10.1007/978-3-031-70584-7

[45] South African Reserve Bank, Monetary Policy Review (2010). Available online: https://www.resbank.co.za/content/dam/sarb/publications/reviews/mpr/2010/mprmay10.pdf (Accessed on 21 July 2025).

[46] V. Acharya, H. Khandwala, and T. Ramadorai, Causes of the Financial Crisis, *Crit. Rev.* **21**(2–3), 195–210 (2009). https://doi.org/10.1080/08913810902952903

[47] P. J. Wallison, Three narratives about the financial crisis, *Cato. J* **31**(3), 535–549 (2011). Available online: http://www.cato.org/sites/cato.org/files/serials/files/cato-journal/2011/9/cj31n3-8.pdf.

[48] B. Güriş, A new nonlinear unit root test with Fourier function, *Commun. Stat. Simul. Comput.* **48**(10), 3056–3062 (2019). https://doi.org/10.1080/03610918.2018.1473591

[49] G. Schwarz, Estimating the dimension of a model, *Ann. Stat.* **6**(2), 461–464 (1978). https://www.jstor.org/stable/2958889

[50] H. Akaike, A new look at the statistical model identification, *IEEE Transactions on Automatic Control* **19**(6), 716–723 (1974). https://doi.org/10.1109/TAC.1974.1100705

[51] K. Makatjane and D. Xaba, An early warning system for inflation using Markov-switching and logistic models, *Risk Gov. Control Financ. Mark. Inst.* **6**(4), 30–39 (2016). https://doi.org/10.22495/rcgv6i4art5

[52] A. Jierula, S. Wang, O. H. T-M. Wang, Study on Accuracy Metrics for Evaluating the Predictions of Damage Locations in Deep Piles Using Artificial Neural Networks with Acoustic Emission Data, *Applied Sciences* **11**(5), 2314 (2021). https://doi.org/10.3390/app11052314

[53] M. Ertl, I. Fortin, S. P. Koch, R. M. Kunst, and L. Sögner, Inflation forecasting in turbulent times, *Empirica* **52**(1), 5–37 (2025). https://doi.org/10.1007/s10663-024-09633-z