-

Forecasting Exchange Rate Dynamics: A Comparative Study of Traditional Econometric Models and Machine Learning Models.

A Research Proposal presented for the degree of Master of Economic Science

in the School of Economics and Finance,

University of the Witwatersrand

By

Teyven Ndlovu

Student No: 2088089

Supervisor: Prof. Gregory Farrell

ABSTRACT

Accurate exchange rate forecasting is crucial for economic policy, risk management, and financial decision-making. This study compares traditional hybrid econometric models, specifically the Autoregressive Integrated Moving Average with Exogenous Variables and Generalized Autoregressive Conditional Heteroskedasticity (ARIMAX-GARCH), with machine learning approaches, including Random Forest, Multi-Layer Perceptron, and Long Short-Term Memory networks, to predict the South African Rand (ZAR) against major global currencies: the United States Dollar (USD), Euro (EUR), British Pound (GBP), Japanese Yen (JPY), and Chinese Yuan (CNY). Using daily exchange rate data from 2000 to 2024, model performance is evaluated based on Root Mean Squared Error (RMSE), Mean Absolute Percentage Error (MAPE), Symmetric Mean Absolute Percentage Error (sMAPE), and Mean Directional Accuracy (MDA). The results indicate that no single model consistently outperforms across all currency pairs. The ARIMAX-GARCH model excels in trend prediction, the Random Forest model balances predictive accuracy and adaptability, the Multi-Layer Perceptron model minimizes absolute errors but struggles with directional accuracy, and the Long Short-Term Memory model captures long-term dependencies but underperforms in volatile markets. These findings highlight the need for hybrid forecasting models that integrate machine learning and econometric techniques while incorporating macroeconomic indicators to enhance predictive reliability.

Contents

1. Introduction

In recent years, the ability to accurately forecast exchange rate fluctuations has become increasingly crucial for economic and managerial decision-making. Reliable exchange rate projections are essential for a wide array of activities, including government planning for import costs, managing foreign currency-denominated debt, and setting interest payments, as well as guiding central bank actions and shaping the strategic decisions of multinational corporations. These forecasts also have significant implications for trade, investment, foreign direct investment and profitability by influencing investment returns and the broader economic landscape. Reliable exchange rate predictions can help mitigate financial risks, inform monetary policy, and enhance investment strategies, playing a vital role in ensuring economic stability and growth.

The South African Rand (ZAR) serves as an exemplar of the challenges and opportunities inherent in exchange rate forecasting. Despite the relatively small size of the economy, ZAR was the 18th-most-traded currency in the world in April 2022 (Bank for International Settlements, 2022). As Eichengreen and Gupta (2014) have noted in the context of the mid-2013 'Taper Tantrum,'[1] this may be a mixed blessing, as global shocks may impact disproportionately on emerging markets with large and liquid financial systems. Characterised by significant volatility, ZAR fluctuations reflect a complex interplay of domestic factors, such as political uncertainty, economic policy decisions, commodity price changes and developments in the global economic environment (May 2015; Mlambo et al., 2013). This volatility, while presenting a risk management challenge, also offers a useful lens through which to examine the efficacy of potential forecasting models. Predicting movements of the ZAR against major currencies such as the US Dollar (USD), Euro (EUR), British Pound Sterling (GBP), Japanese Yen (JPY) and Chinese Yuan (CNY) is not only technically challenging, but also crucial for stakeholders involved in trade, investment, and policy making.

Traditionally, time-series econometric models like the Autoregressive Integrated Moving Average (ARIMA) model and its variants, such as the ARIMAX (a standard ARIMA model extended to include other relevant external (exogenous) factors such as economic indicators, market trends, or seasonal effects) and the hybrid ARIMAX-GARCH models introduced later,

[1] The term 'Taper Tantrum' refers to the 2013 market reaction to the U.S. Federal Reserve's plan to reduce its bond-buying, causing higher interest rates and capital outflows from emerging markets.

have been extensively employed for forecasting exchange rates. These models tend to be good at capturing linear relationships but often falter in modelling the inherent non-linearity and high volatility present in financial time series. In recent years, machine learning (ML) techniques, including Random Forest (RF), Multi-Layer Perceptron (MLP), and Long Short-Term Memory networks (LSTM), have gained prominence for their ability to identify complex patterns and adapt to evolving market conditions (Ayitey et al., 2023; Qureshi et al., 2023; Colombo and Pelagatti, 2020). These advanced methods leverage extensive datasets to enhance forecasting accuracy by uncovering complex relationships that traditional models may overlook. ML methods have performed well in the M4 (Makridakis, Spiliotis, and Assimakopoulos, 2020), M5 (M Open Forecasting Center, 2020) and Kaggle competitions (Bojer and Meldgaard, 2021).

Despite the growing interest in ML-based forecasting, there remains limited research comparing the effectiveness of traditional hybrid econometric models and ML approaches for forecasting exchange rates across multiple currencies, particularly for emerging market currencies and specifically for the ZAR. While some existing studies have explored the use of traditional econometric models and machine learning methods across multiple currency pairs in other contexts (Afuecheta et al., 2022; Djemo et al., 2022; Liu et al., 2023; Umoru et al., 2023), the majority have focused on predicting a single currency pair Alamneh and Mancha, 2021; Al-Gounmeein et al., 2024; Chasipanta and Sánchez-Pozo, 2024; Eniayewu et al., 2024; Inyang et al., 2023; Liu et al., 2024; Pahlavani and Roshan, 2015; Qureshi et al., 2023; Tanko et al., 2023; Zhang, 2024).

This research aims to fill this gap by conducting a comparative assessment of traditional econometric models including ARIMAX-GARCH and ML-based approaches in forecasting the exchange rate dynamics of the South African Rand against five major global currencies: the USD, EUR, GBP, JPY, and CNY. The study employs a dataset spanning from January 2000 to July 2024 and evaluates the forecasting performance of the ARIMAX-GARCH, RF, MLP, and LSTM models using rigorous validation techniques including Train-test split, Time series cross validation, Hyperparameter tuning and Regularization. By employing multiple performance metrics—including Root Mean Squared Error (RMSE), Mean Absolute Percentage Error (MAPE), Symmetric Mean Absolute Percentage Error (sMAPE), and Mean Directional Accuracy (MDA), this study seeks to determine which models offer the most reliable exchange rate predictions across different market conditions.

To preview the findings the study found that no single model consistently outperformed the other models across currency pairs. The ARIMAX-GARCH excelled in capturing trends but struggled with volatility and non-linearity. The RF model showed strong predictive accuracy and adaptability but responded slowly to extreme shifts, suggesting benefits from hybridization. The MLP minimized errors but lacked directional accuracy, while the LSTM captured long-term trends but struggled with short-term volatility.

These results are not just empirical insights; they also have meaningful theoretical implications. The varied performance of ARIMAX-GARCH, RF, MLP, and LSTM models across different exchange rate environments challenges the adequacy of relying solely on traditional linear models for forecasting. These results suggest that forecasting frameworks should be tailored to the structural characteristics of exchange rate behaviour, particularly in volatile emerging markets. By highlighting the strengths and limitations of each approach, the findings support a growing theoretical shift toward hybrid models that integrate the interpretability of econometric methods with the adaptability of machine learning. Future research should explore how such combinations can better accommodate structural breaks, non-linear dynamics, and regime changes in financial markets.

The research report is structured as follows. Section 2 provides a review of the literature that informs the research. Section 3 sets out the research methodology, including a discussion of the data and the models employed to undertake a comprehensive comparative assessment of traditional econometric and ML techniques used to forecast the Rand currency pairs, as well as model validation techniques and forecast evaluation metrics. Section 4 presents the in-sample empirical results, and Section 5 the out-of-sample results. Section 7 concludes.

2. Literature Review

Exchange rate forecasting has long posed significant challenges, with early foundational studies casting doubt on the predictive power of traditional models. Meese and Rogoff (1983) famously demonstrated that traditional structural exchange rate models were consistently outperformed by a simple random walk, sparking decades of scepticism about the usefulness of economic fundamentals in exchange rate prediction. Building on this legacy, Rossi (2013) offered a nuanced reassessment, showing that exchange rate predictability remains conditional, dependent on the forecasting horizon, sample period, model type, and evaluation method. Together, these studies expose the limitations of conventional econometric approaches and motivate the search for alternative methodologies. Recently, researchers have increasingly turned to machine learning techniques, which offer the flexibility to model nonlinear relationships and adapt to evolving financial dynamics.

Machine learning (ML) techniques such as artificial neural networks (ANNs), support vector machines (SVMs), recurrent neural networks (RNNs), and long short-term memory (LSTM) models have gained traction for their ability to capture complex, nonlinear relationships and improve forecasting accuracy in volatile financial environments.

Kamruzzaman and Sarker (2003) compared three artificial neural network models (standard backpropagation[2], scaled conjugate gradient[3], and backpropagation with Bayesian regularization[4]) against the ARIMA model for forecasting the nominal Australian Dollar's exchange rate against six currencies including USD, GBP, JPY, Singapore Dollar (SGD), New Zealand Dollar (NZD) and Swiss Franc (CHF) from January 1991 to July 2002 over 35 and 65 week-forecasting horizons. Using five moving average indicators and measuring the performance with five metrics, they found that the forecasts of all artificial neural network

---

[2] Standard backpropagation is a core algorithm for training artificial neural networks. It adjusts the network's weights and biases by propagating prediction errors back through the layers, iteratively improving the network's accuracy in mapping inputs to outputs.

[3] The Scaled Conjugate Gradient method is an optimization algorithm commonly used in training neural networks. It's a variation of the conjugate gradient method, designed to be more efficient and robust, especially for large datasets

[4] Backpropagation with Bayesian regularization enhances standard backpropagation by treating network weights as random variables with probability distributions. This approach, often using a Gaussian prior favouring smaller weight, helps prevent overfitting by penalizing large weights and effectively averaging predictions from multiple weight sets, thereby improving generalization to unseen data.

(ANN) models outperformed the ARIMA models. In particular, the scaled conjugate gradient (SCG) model excelled in terms of accuracy, supporting the effectiveness of ANNs in foreign exchange (forex) market forecasting.

Chen et al., (2008) proposed a hybrid forecasting model for foreign exchange rates that aggregates both fundamental and technical analysis over Long-Term (30 days), Medium term (15 days), and the Short-term (daily) forecasting horizons. Using daily exchange rate data for the nominal Taiwan/US Dollar (TWD/USD), spanning from 6 June 1998 to 12 February 2007, the authors employed a multi-neural network to integrate macroeconomic indicators (fundamental analysis) with historical price trends and technical indicators. Their findings suggest that this hybrid model, by capturing insights from both fundamental and technical perspectives, achieved greater forecasting accuracy compared to models relying solely on one type of analysis or a simple random walk benchmark. This study highlights the potential of combining diverse analytical approaches through machine learning for improved exchange rate prediction.

Chasipanta and Sánchez-Pozo, (2024) investigated the effectiveness of combining the traditional ARIMA model with a MLP neural network for long-term (365-day) forecasts of Euro-Dollar nominal exchange rates (EUR/USD). Using daily data for the EUR/USD spanning from March 2017 to December 2022, the authors employed a MLPs to integrate macroeconomic indicators (fundamental analysis) with historical price trends and technical indicators. Their findings suggest that this hybrid model, by capturing insights from both fundamental and technical perspectives, achieves greater forecasting accuracy compared to models relying solely on one type of analysis or a simple random walk benchmark. This research also emphasizes the promising prospect of aggregating a diverse array of analytical techniques, facilitated by machine learning, to enhance the accuracy of exchange rate forecasting.

Nanthakumaran and Tilakaratne (2017) compared different forecasting models for predicting (1-day forecasts) exchange rates between the Sri Lankan rupee (LKR) and the EUR and JPY from 2 July 2012 to 31 August 2016. They found that traditional stochastic models were ineffective. However, ML models, specifically Nonlinear Autoregressive Neural Networks

(NARNN)[5] and Support Vector Regression (SVR)[6] models, demonstrated good accuracy in predicting exchange rate variations. The SVR models outperformed the ANN models in terms of their directional accuracy. These findings are valuable for both domestic and foreign investors, and the authors suggest exploring evolutionary neural networks to improve forecasting accuracy further.

Amat et al. (2018) also considered forecasting (365-day forecasts) nominal exchange rates by employing ML techniques. Their research, utilizing monthly EUR, Canadian dollar (CAD), GBP, JPY and CFH exchange rate data from March 1973 to December 2014, centred around two primary methodologies: averaging-based methods and regret minimization-based methods. Averaging-based methods, encompassing techniques like simple moving averages and exponentially weighted moving averages, formulate predictions by calculating averages of historical exchange rate data. Conversely, regret minimization-based methods endeavour to reduce prediction errors in comparison to benchmarks set by expert predictions, which might involve strategies such as maintaining a constant prediction or aligning with trends observed in fundamental economic indicators. The authors find that these simple methods do not outperform a naive random walk model, suggesting limited predictability of exchange rates based on the explored fundamentals. They underscore the uncomplicated nature and robust theoretical underpinnings of these methods, highlighting the persistent challenge in forecasting exchange rates.

Escudero et al. (2021) compared the performance of RNN (specifically, Elman[7] and LSTM networks) against the traditional ARIMA model for forecasting the EUR/USD exchange rate. Using daily data for the Euro against the US Dollar, spanning from 2 January 1998 to 31 December 2019, over the various forecasting horizons (5, 11, 22, 35, 44, and 55 days), the study uses a rolling-window approach to evaluate short-term forecasting accuracy across different window sizes. Their findings indicate that both Elman and LSTM networks

---

[5] Nonlinear Autoregressive Neural Networks are a class of recurrent neural networks specifically designed for time series forecasting. They excel at modelling complex, nonlinear dependencies in data by leveraging past values of the time series to predict future values. This makes them particularly suitable for capturing the often-subtle patterns and dynamic behaviour observed in financial markets, such as exchange rate fluctuations.

[6] Support Vector Regression is a supervised machine learning algorithm used for regression tasks. It aims to find a hyperplane that best fits the data while staying within a specified margin of error, making it robust to outliers. SVR is particularly effective in handling non-linear relationships between variables, making it suitable for complex forecasting problems like exchange rate prediction.

[7] The Elman Recurrent Neural Network, or Elman network, is a type of recurrent neural network with feedback connections in its hidden layer, allowing it to retain information about previous inputs and learn temporal dependencies in sequential data.

outperform the ARIMA model, with LSTM demonstrating superior performance in capturing the dynamic behaviour of the exchange rate. This research highlights the potential of RNNs, particularly LSTM, as valuable tools for short-term exchange rate forecasting.

Olsen et al. (2024) compare the performance of LSTM neural networks to traditional econometric models in forecasting implied volatilities of currency options. Using daily data for the EUR/USD nominal exchange rate, spanning from 2 January 2007 to 31 August 2017, over horizons ranging from 7-day to a 365-day, the study evaluates the predictive accuracy of LSTM against benchmark models like RFs AR-GARCH, HAR, and MIDAS. Interestingly, the authors find that while LSTM shows promise for shorter maturities, traditional GARCH models outperform it for longer-term options, especially when tailored with specific specifications and residual distributions. This suggests that while machine learning models like LSTM offer valuable tools for volatility forecasting, particularly in capturing short-term dynamics, established econometric models remain competitive and even superior in certain contexts.

Lim and Zohren, (2021) provide a comprehensive survey of the use of deep learning in time-series forecasting. The authors review various deep learning architectures, including RNN, Convolutional Neural Network. (CNN)[8], and Transformers, examining their application in one-step-ahead and multi-horizon forecasting. The paper delves into how these architectures incorporate temporal information for predictions and discusses their strengths and limitations. Furthermore, the study highlights the growing trend of hybrid deep learning models, which combine statistical models with neural network components. This approach leverages the strengths of both worlds, often leading to improved forecasting performance compared to using either method in isolation. Beyond prediction accuracy, the authors recognize the importance of interpretability and counterfactual prediction in decision-making contexts. They discuss how these techniques can be applied to deep learning models for time series, making their predictions more transparent and insightful for practical applications. The paper concludes by identifying promising avenues for future research, such as continuous-time and hierarchical models, emphasizing the continuous evolution and expanding potential of deep learning in time-series analysis.

An important literature challenges the notion that deep learning methods are inherently superior, demonstrating that they do not consistently outperform traditional statistical or

[8] Convolutional Neural Networks are specialized neural networks designed for processing structured grid data like images. They employ convolutional layers that use filters to extract local patterns, making them highly effective for image recognition and other visual tasks

machine learning approaches in terms of accuracy. Makridakis et al. (2022), for example, conduct a comparative analysis of statistical, ML, and deep learning methods for time series forecasting across a range of datasets. They emphasise that while deep learning shows promise, its effectiveness is not universal and depends on the specific characteristics of the time series data being analysed. Notably, the study highlights the consistent strong performance of ensemble methods, which combine predictions from multiple models. This suggests that combining the strengths of diverse methodologies can lead to more robust and accurate forecasts. Makridakis et al., (2022) conclude by advocating for a balanced approach to time series forecasting, emphasizing the importance of considering data availability, time series characteristics, and the strengths and limitations of different forecasting techniques when selecting the most appropriate method.

Similarly, Ayitey et al. (2023) present a systematic literature review examining the application of machine learning in forecasting foreign exchange rates. Their analysis encompasses sixty primary studies published between 2010 and 2021, meticulously categorized and evaluated based on the machine learning methodologies employed, the evaluation strategies implemented, and the performance metrics used. The review reveals a distinct preference for certain currency pairs, notably EUR/USD and USD/GBP, in the existing literature. Furthermore, the authors observe a dominance of specific algorithms, such as ANNs and SVMs, highlighting their popularity in forex forecasting. The study identifies a significant gap in the standardization of evaluation techniques across the reviewed literature. The authors emphasize the need for more robust validation methods, particularly cross-validation, to ensure the reliability and generalizability of research findings. In conclusion, Ayitey et al. (2023) offer several recommendations for future research, including exploring less-studied algorithms, incorporating economic indicators into forecasting models, and focusing on longer-term forecasting horizons. These recommendations highlight potential avenues for advancing the field of forex forecasting using machine learning.

In the context of South African exchange rate forecasting, there is limited research that directly compares ML and traditional econometric approaches. Kabundi (2002) compares the forecasting performance of Artificial Neural Networks and traditional econometric models in the context of macroeconomic variables such as inflation and share prices. The analysis utilizes data from 1979 to 2000 and evaluates the models over 1-month, 6-month, and 12-month forecasting horizons. Using ANN to capture nonlinear relationships in macroeconomic and financial data, the study benchmarks their performance against traditional econometric approaches like the Vector Error Correction Model. The findings indicate that while ANNs excel at modelling complex, nonlinear patterns, the VECM model demonstrates superior capabilities, particularly in forecasting inflation rates in South Africa. Interestingly, the results suggest that while ANNs offer a powerful tool for capturing nonlinear dynamics, traditional econometric methods retain their effectiveness, especially in structured economic settings or linear tasks. The study concludes that both ANN and traditional models provide complementary insights, with ANNs offering valuable forecasts over short and long-term horizons, while traditional models like VECM continue to perform well, highlighting the complementary roles of both approaches in macroeconomic forecasting.

Botha et al. (2022 investigate the use of big data and machine learning approaches to enhance the accuracy of the South African Reserve Bank's existing inflation forecasting models. The study analyses a comprehensive dataset of 34,075 nationwide item prices collected from January 2009 to March 2021. Various statistical learning techniques, including regression trees, boosted regression trees, and neural networks, are employed to assess forecasting performance. The researchers evaluate model accuracy using root-mean-squared error and benchmark it against the SARB's official forecasts. The findings suggest that certain machine learning models, particularly those incorporating variable selection methods such as least absolute shrinkage and selection operator (LASSO)[9], adaptive LASSO[10], post-selection inference[11], and Bayesian[12] model selection, outperform the SARB's forecasts, especially for shorter time horizons of up to eight months.

Martin (2019) examined the use of ML versus autoregressive and vector autoregressive models in forecasting South African GDP growth and found that ML methods outperformed traditional techniques. This study investigates the comparative efficacy of machine learning techniques and conventional econometric models in forecasting the quarterly percentage change in South African Gross Domestic Product (GDP). Employing data from the first quarter of 1990 to the

[9] LASSO is a regression method that applies L1 regularization, forcing some coefficients to be exactly zero, effectively selecting the most important predictors while reducing overfitting.

[10] Adaptive LASSO is variation of LASSO that assigns different penalty weights to different coefficients, reducing the risk of over-selecting variables and improving model accuracy by better distinguishing between relevant and irrelevant predictors.

[11] Post-Selection Inference: A statistical technique used to refine variable selection by ensuring that only predictors with significant contributions remain in the final model, reducing the risk of including spurious relationships.

[12] Bayesian Model Selection: A probabilistic approach to selecting the best model by evaluating different sets of predictors based on their posterior probabilities, integrating prior information and model uncertainty to improve selection robustness.

last quarter of 2018, the research explores machine learning models such as Elastic-Net[13], RFs, SVMs, and RNNs, while also incorporating traditional Autoregressive and Vector Autoregressive models. The findings indicate that machine learning methods outperform their conventional counterparts in terms of forecasting accuracy, as measured by metrics such as RMSE and correlation with actual GDP trends. Notably, the Elastic-Net model demonstrated the strongest performance, highlighting the potential of machine learning approaches to provide more robust decision-making tools for policymakers. However, the study also acknowledges that machine learning techniques require greater data and computational resources compared to traditional models, which may pose a limitation in certain applications.

The existing literature suggests that machine learning methods, such as RNNs, ANNs, SVMs, and LSTMs, have frequently exhibited superior forecasting capabilities compared to traditional econometric models in various international currency markets. However, as noted by Makridakis et al. (2018), ML models are not without their limitations, including higher computational requirements and occasional underperformance, particularly when dealing with complex patterns and volatility clustering in financial time series. These challenges underscore the importance of incorporating traditional hybrid models, such as ARIMAX-GARCH, in this analysis.

While ML models have shown promise in currency forecasting, there remains a need for more research directly comparing their performance with that of traditional hybrid models, specifically in the context of forecasting the ZAR against major global currencies. By rigorously assessing both traditional econometric and ML approaches, this research aims to provide a more nuanced understanding of the most effective forecasting techniques for the highly volatile and sometimes misaligned ZAR.

[13] Elastic-Net is a regularized regression model that combines both L1 (Lasso) and L2 (Ridge) penalties. It is designed to address some limitations of Lasso regression, particularly when there are correlated predictors. Elastic-Net performs variable selection and regularization by balancing the strengths of L1 and L2 regularization, making it especially useful when dealing with high-dimensional datasets where many features may be correlated or redundant. The model automatically selects the most important features while shrinking the coefficients of less important ones toward zero, helping prevent overfitting.

3. Methodology

This study has selected LSTM networks, RFs, MLPs, and ARIMAX-GARCH models based on the extensive literature review and the distinct characteristics of foreign exchange rate forecasting. Each model was selected based on its proven effectiveness in handling different aspects of exchange rate forecasting, as identified in prior research.

LSTM networks are well-suited for time series forecasting due to their ability to capture both short-term and long-term dependencies, as well as non-linear relationships and abrupt shifts in exchange rate behaviour. The memory cells of LSTMs enable them to outperform traditional models, particularly when past behaviours influence future outcomes across different time horizons. RFs are robust models capable of handling complex, non-linear interactions without extensive data preparation. They excel in noisy environments like financial markets and mitigate overfitting[14] by aggregating predictions from multiple decision trees, making them well-positioned for exchange rate forecasting. MLPs are simpler neural networks that effectively model non-linear patterns and are especially useful for medium- to long-term forecasting. The straightforward structure of MLPs allows them to integrate fundamental and technical analysis, rendering them versatile for incorporating various data types in exchange rate forecasting. The ARIMAX-GARCH model, representing a hybrid of traditional econometric techniques, models both the mean and volatility of time series. This combination enables it to capture trends and volatility, making it a robust benchmark for financial forecasting.

By incorporating these models, the study seeks to undertake a comprehensive comparative assessment of traditional econometric approaches and advanced machine learning techniques, with the aim of providing a robust evaluation of their respective forecasting capabilities within the context of financial time series data.

[14] Overfitting in machine learning occurs when a model learns the training data too well, capturing not only the underlying patterns but also noise and irrelevant fluctuations. As a result, the model performs exceptionally well on the training set but fails to generalize to unseen data, leading to poor performance on test data. Overfitting is often caused by excessive model complexity, insufficient regularization, or too many training iterations, and it can be mitigated through techniques such as L2 regularization, dropout, early stopping, and cross-validation.

3.1 Data

This study uses daily nominal exchange rate observations for four major currencies against the South African Rand (ZAR): the US Dollar (USD/ZAR), Chine Yuan (CNY/ZAR), Japanese Yen (JPY/ZAR), Euro (EUR/ZAR), and British Pound (GBP/ZAR). The dataset is sourced from Bloomberg and the Federal Reserve Economic Data (FRED) database, spanning from 3 January 2000 to 29 July 2024, with a total of 6,411 daily observations. In addition to the exchange rates, following Alexandridis et al. (2018), the dataset also includes several economic indicators, such as gold prices, crude oil prices, S&P 500 index returns, Baltic Dry Index (BDI) values, the Chicago Board Options Exchange Volatility Index Volatility Index (VIX), and CRB commodity index returns, used as exogenous variables (predictors). Table 1 provides details

Table 1: Exchange rates, predictors, and data sources

Number	Notation	Description	Source
1	USDZAR	USD ZAR currency pair	Bloomberg
2	JPYZAR	JPY ZAR currency pair	Bloomberg
3	EURZAR	EUR ZAR currency pair	Bloomberg
4	GBPZAR	GBP ZAR currency pair	Bloomberg
5	CNYZAR	CNY ZAR currency pair	Bloomberg
6	LXAU	Gold prices	FRED
7	LOIL	Crude oil returns	FRED
8	LSP500	SP500 returns	Bloomberg
9	LBDI	Baltic Dra Index	Bloomberg
10	LVIX	VIX	Bloomberg
11	LCRB	CRB commodity Index returns	Bloomberg

This study analyses the direct exchange rates, meaning the price of 1 ZAR is expressed in terms of foreign currencies. These exchange rates are closing exchange rates, which provide a consistent basis for comparison with other daily financial data. In addition to these exchange rates, six exogenous variables that may influence exchange rates are considered. The VIX measures market expectations of near-term volatility in the S&P 500. It is often referred to as the "fear gauge" because rising VIX levels typically indicate heightened risk aversion in global markets, which may weaken the ZAR as investors seek safer assets. Typically, an increase in the VIX signals a negative financial outlook (Alexandridis et al., 2018). Furthermore, we consider gold returns as a safe haven against shocks to risky assets, incorporating them alongside the VIX as a risk measure. Gold prices are employed as an indicator of overall risk sentiment in global markets. A rise in gold prices frequently coincide with heightened risk aversion among investors, which may result in the depreciation of currencies perceived as riskier, such as the ZAR. We also include oil returns, fluctuations in crude oil prices can have far-reaching implications, as they tend to influence macroeconomic factors such as inflation, economic growth, and interest rates, all of which can subsequently impact exchange rate dynamics. Specifically, an increase in oil prices may lead to higher inflation in countries that rely on oil imports, potentially contributing to a depreciation of their domestic currencies.

To capture trade activity and future demand, we use the BDI which serves as a proxy for the cost of transporting raw materials. Elevated BDI values typically suggest heightened demand for goods, which may positively impact export-oriented economies such as South Africa. Furthermore, we analysed the predictive performance of the CRB, which is particularly relevant for countries that export or import commodities. CRB Index is a representation of the price dynamics across a basket of nineteen commodities, encompassing energy resources, metals, and agricultural products. Given South Africa's significant dependence on commodity exports, the CRB Index serves as a critical indicator of the prevailing economic conditions. The Commodity Research Bureau Index, which represents the price movements of a basket of nineteen commodities spanning the energy, metals, and agricultural sectors, serves as a crucial indicator of the broader economic environment, given South Africa's dependence on commodity exports. Movements in the S&P 500 index can be used as a gauge for the overall state of global equity markets and prevailing investor sentiment. The index's performance reflects the macroeconomic outlook, as bullish equity markets typically signal positive investor confidence, which could lead to increased capital inflows and potential appreciation of the ZAR in certain circumstances.

The dataset used in this study provides a comprehensive view of the factors influencing the ZAR by incorporating both direct exchange rates for major currencies and several key macroeconomic indicators. The selection of exogenous variables, such as the VIX, gold and oil prices, the Baltic Dry Index, and the CRB commodity index, reflects the complexity and interconnectedness of global financial markets. These indicators are particularly relevant given South Africa's position as a commodity-exporting nation and the influence of global risk sentiment on its currency. By analysing these predictors alongside daily exchange rates, this study aims to capture the multifaceted dynamics of exchange rate movements, providing a robust foundation for the forecasting models employed. This approach allows for a more nuanced understanding of how external economic factors interact with domestic currency behaviour.

3.3 Traditional Models

The ARIMAX-GARCH model is a forecasting technique that combines two statistical methods to analyse and predict time series data. The ARIMAX portion of the model considers the linear connections between past observations and external factors to estimate future values. This component involves standard ARIMA modelling of the mean equation to include the impact of previous data points and integration of the data and moving averages to integrate past forecast errors into the model. The ‘X’ represents other relevant external (exogenous) factors such as economic indicators, market trends, or seasonal effects that may influence the time series. Incorporating these variables allows the model to account for influences beyond the inherent structure of the data. The GARCH element is used to capture the changing volatility of the data series, identifying periods of increased or decreased uncertainty by examining the residuals or 'errors' from the ARIMAX model. In general, owing to its ability to capture both inherent patterns within a dataset and its evolving volatility, it presents itself as a robust choice for addressing forecasting challenges involving non-constant variance typically encountered in fields such as finance or energy consumption (Saranj and Zolfaghari, 2022; Corrêa et al., 2016).

3.3.1 ARIMAX model

Let y_t denote a non-stationary time series for each exchange rates specified showing linear auto-dependence. Additionally, assume $\left((X_{1,t})_{t=1}^{T}, \dots \left(X_{(r+1),t}\right)_{t=1}^{T}\right)$ represents a list of $r + 1$ vectors consisting of realizations from $r + 1$ stationary explanatory variables of $y_t(t = 1, \dots, T)$. According to Pankratz (2012) and Box and Tiao (1975) each realization y_t of the exchange rate can be expressed using an ARIMAX (p, d, q) model with the following mathematical representation:

$$\nabla^d y_t = \varphi_0 + \sum_{j=1}^{p} \varphi_k \, y_{t-j} + \sum_{j=i}^{q} \vartheta_i \, e_{t-i} + \sum_{i=1}^{r+1} \sum_{l_i=0}^{L_i} \gamma_{il_i} x_{i,t-l_i} + e_t \tag{1}$$

where B is the lag operator with $B^k y_t := y_{t-k}$, where $k \in \mathbb{Z}$, , $\nabla^d := (1-B)^d$represents the difference operator of order d; $(\varphi_i)_{i=0}^{p}$ and $(\vartheta_i)_{j=1}^{q}$ are ordered series of complex model parameters, with non-zero φ_p and ϑ_p. The list $\left(\gamma_{il_i}\right)_{l_{i=0}}^{L_i}$ contains complex model parameters associated with the exogenous component i, satisfying both invertibility and stationary conditions (Hamilton, 1994; Lütkepohl, 2005). e_t denotes an innovation derived from a random variable ε_t in an uncorrelated stochastic process with zero mean. Here p and q represent the orders of the Autoregressive part $\sum_{j=1}^{p} \varphi_k \, y_{t-j}$ (AR (p)) and Moving Average part $\sum_{l_i=0}^{L_i} \gamma_{il_i} x_{i,t-l_i}$. Moreover, L_i stands for maximum lag order within $\sum_{l_i=0}^{L_i} \gamma_{il_i} x_{i,t-l_i}$ involving exogenous variable $x_{i,t}$, for $i = 1, \dots, (r+1)$. Notably if $\gamma_{il_i} = 0$ for all $i = 0, \dots, (r+1)$ and for all $l_i = 0 \dots, L_i$, then equation above corresponds to a standard ARIMA(p, d, q).

However, since forecasting y_{t+1} requires the future values of exogenous variables x_{t+1}, this study adopts a recursive one-step-ahead strategy, where only current and lagged values of the exogenous variables are used. This avoids reliance on forecasted exogenous inputs and maintains model feasibility, though at the cost of discarding contemporaneous information in some cases.

In Equation (1), the innovation term e_t is assumed to represent a realization of an uncorrelated random variable ε_t with zero mean and constant conditional variance i.e. $\sigma_t^2 = \sigma^2$. However, most time series exhibit non-constant conditional variance, where the changing volatility may depend on past squared innovations or past variance values (Saranj and Zolfaghari, 2022). In such cases, an ARMA structure or its extension is required to model and project the volatility over time.

3.3.2 GARCH model

GARCH models are statistical tools used to analyse and predict volatility in time series data, particularly in finance. The core concept behind GARCH models is that volatility is not constant over time but rather exhibits clustering, where periods of high volatility tend to be followed by more high volatility, and periods of low volatility tend to be followed by more low volatility. This pattern is often observed in asset prices, where a large price swing one day increases the likelihood of another large swing the following day. GARCH models are characterized by their flexibility, using past values of volatility to predict future volatility (autoregressive) and acknowledging that volatility changes based on past information (conditional). The term "heteroskedasticity" signifies that the variance of the data is not constant over time. These models are valuable for risk management, allowing for the estimation of potential future price swings, and option pricing, where accurate volatility forecasts are crucial for determining option values. The general GARCH (R, S) model according to Bollerslev, (1986) for the conditional variance σ_t^2 of the innovation term e_t for each exchange rate is formulated as follows:

$$\sigma_t^2 = \lambda_0 + \sum_{j=1}^{R} \lambda_j \sigma_{t-j}^2 + \sum_{i=1}^{S} \mu_i \, e_t{}^2 \tag{2}$$

where the given restrictions need to be met: $\sum_{j=1}^{R} \lambda_j + \sum_{i=1}^{S} \mu_i < 1,\ \lambda_0 > 0;\ \lambda_j \geq 0\ j = 1, \dots, R)$ and $\mu_i \geq 0\ (i = 1, \dots, S)$. Additionally, a parameter φ can be employed in the equation to account for effects that are non-linear when generating forecasts of its level (conditional mean).

3.3.3 The ARIMAX – GARCH model

A model that incorporates both equations (1) and (2) and takes into consideration the constraints mentioned above is referred to as an ARIMAX-GARCH model. This model first uses the ARIMAX model to forecast the expected value of the exchange rate, accounting for past values, errors, and exogenous variables. The residuals from the ARIMAX model (e_t) are then passed into the GARCH model, which models how the volatility of these residuals (i.e., the unpredictability of the exchange rate) changes over time. The innovation process e_t, which reflects the error in the ARIMAX model, is modelled as:

$$e_t = \sigma_t z_t \tag{3}$$

Where z_t is a standardized, independent, and identically distributed $(i.i.d.)$ random variable with zero mean and unit variance. This represents the random shock or innovation that affects the exchange rate at time t. σ_t is the conditional standard deviation derived from the GARCH model. This represents the level of volatility or uncertainty at time t, which varies over time based on past errors and volatilities.

To obtain the most suitable ARIMAX-GARCH model, we need to determine the appropriate values of the model parameters, including the autoregressive (p), differencing (d), and moving average (q) orders, as well as R and S in (2). This can be achieved by leveraging a Python library such as $pmdarima$, which employs an iterative search process to select the optimal ARIMA model parameters. The library utilizes various information-theoretic criteria, including the Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), corrected Akaike Information Criterion, and Hannan-Quinn Criterion (HQIC). These criteria have distinct strengths and weaknesses. The BIC and HQIC impose a higher penalty for model complexity, and the BIC is consistent in selecting the order of the process in large sample sizes while the AIC has a nonzero probability of overfitting (Shibata, 1976). However, when model selection is required for prediction, the minimax property of the AIC that minimises the maximum possible risk means it is often preferred (Ding et al, 2018). By considering multiple evaluation metrics, a balanced approach is ensured, accounting for both model accuracy and complexity, thereby leading to the most reliable model selection for the given time series data.

Note that similar to the generation of h-step-ahead out-of-sample forecasts for conditional means, the h-step-ahead forecasts of conditional variances are derived using an estimated GARCH model. The GARCH-in-mean model is employed to generate out-of-sample volatility forecasts, which are then incorporated into the estimation of conditional mean forecasts (Hamilton, 1994). This approach leverages volatility dynamics to refine exchange rate predictions by integrating non-linear information, thereby enhancing overall forecasting accuracy (Corrêa et al., 2016).

3.4 Machine Learning Models

3.4.1 Random forest

The RF approach is a machine learning technique that constructs multiple decision trees, each trained on a random subset of the data and features, a process known as bootstrap sampling. By randomly selecting both data points and features, RFs ensure that the decision trees are less correlated with each other. This diversity among the trees reduces the likelihood of overfitting, which occurs when a model performs well on training data but fails to generalize to unseen data.

A key advantage of the RF algorithm is its ensemble approach, which combines the predictions of multiple decision trees to generate a more accurate final result. For classification problems, the algorithm employs majority voting, selecting the most common prediction from all the individual trees. This approach helps mitigate errors that may arise from individual trees trained on noisy or biased data subsets. In the case of regression tasks, instead of voting, the algorithm computes the average of all the tree predictions to arrive at the final result. This aggregation process helps to smooth out fluctuations caused by any single tree's prediction, leading to a more stable and reliable overall prediction.

This ensemble approach aggregates the different set of multiple decision trees to handle noisy data and mitigate overfitting, leading to more reliable and accurate predictions than a single decision tree. Despite being more computationally intensive and less interpretable than single decision trees, RFs offer several advantages, including improved accuracy and robustness. The model also allows for identifying feature importance, providing insights into the most influential variables for predicting outcomes. This versatility makes RFs a powerful tool for various applications, including finance, healthcare, and natural language processing.

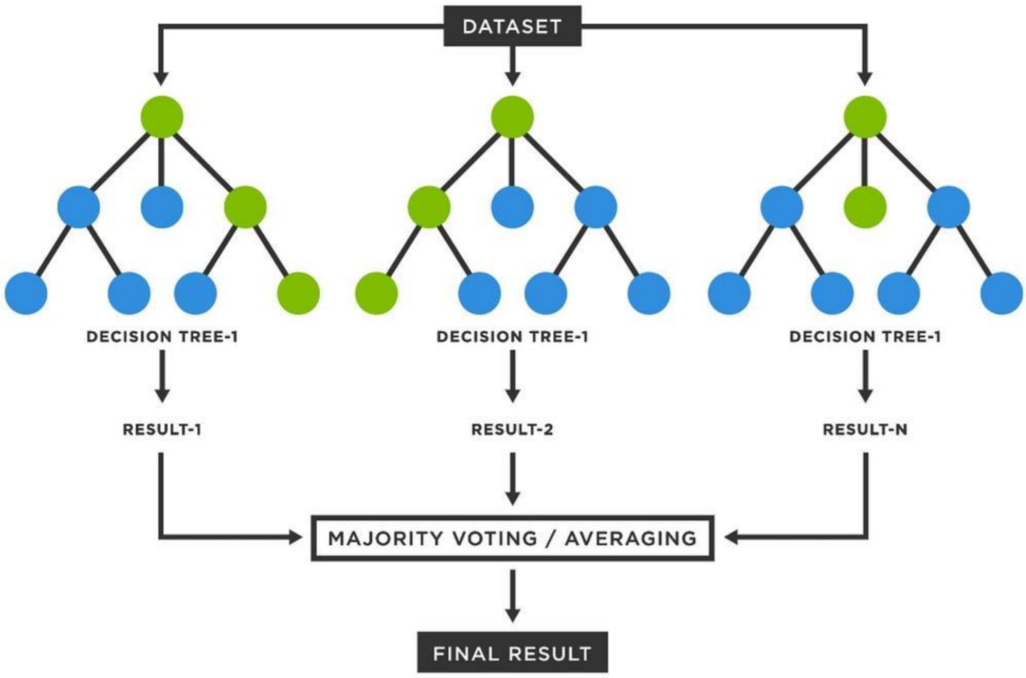

Figure 1: Random Forest Algorithm (Source: Hà, 2023)

Colombo and Pelagatti (2020) highlighted that RFs can be seen as a progression of regression trees, representing an advancement and refinement of the original concept. Construction of regression trees involves dividing the predictor space into a series of J distinct regions R_j. Subsequently, the mean response values for the training observations were calculated within each region. The determination of these regions (internal nodes) $1, \dots, J$ is based on minimizing the residual sum of squares. This process is carried out iteratively by examining all potential predictors $X_1, \dots, X_n$ and all feasible threshold values s. The goal is to identify the predictor and threshold pair that results in the most significant decrease in RSS effectively choosing the predictor j and threshold point k that leads to a reduction of

$$\sum_{i;x_i \epsilon R_1(j,k)} \left(e_i - \hat{e}_{R_1}\right)^2 + \sum_{i;x_i \epsilon R_2(j,k)} \left(e_i - \hat{e}_{R_2}\right)^2 \tag{3}$$

where R_1 and R_2 represent distinct regions of predictor space j, $\hat{e}_{R_1}$ denotes the average response of training observations in region $R_1(j,k)$ defined by cut-off point s; similarly, $\hat{e}_{R_2}$ is defined. This recursive process is repeated for each remaining predictor.

When faced with complex issues, decision trees are susceptible to becoming larger and more unstable. Small changes in data can lead to significant modifications in tree structure, resulting in high variance (Colombo and Pelagatti, 2020). It is widely recognized that averaging across independent sets of observations helps reduce variance because each set has its own variability. RFs represent an advancement from regression trees, aiming to mitigate the prediction variance by constructing multiple decision trees by bootstrapping the training samples. However, during the tree building process, only a random subset of potential predictors is considered at each split point to promote decorrelation among the trees and effectively decrease the variance.

RFs models, though computationally demanding and potentially less transparent than individual decision trees, offer substantial advantages in terms of accuracy and robustness. A key strength of this ensemble learning approach is its capacity to provide insights into feature importance, enabling users to identify the most influential variables in the prediction process. By aggregating the strengths of multiple decision trees, RFs mitigates the risk of overfitting and delivers more reliable predictions through the averaging of individual tree outputs. This ensemble technique establishes RFs as a powerful and versatile tool, well-suited for both classification and regression tasks across diverse domains, including finance, healthcare, and natural language processing.

3.4.2 Multi – linear Perceptrons(MLPs)

3.4.2.1 Perceptrons

A perceptron is a fundamental building block of artificial neural networks and one of the simplest neural network architectures. Introduced by Rosenblatt, (1958), it is a single-layer feedforward network that consists of input nodes fully connected to output nodes. The perceptron takes a set of weighted inputs, aggregates them, and applies an activation function[15] to produce a binary output (typically between 0 and 1). The weights associated with each input determine the strength of its influence on the output, and during learning, the perceptron adjusts these weights iteratively to minimize the difference between its predicted and actual target values.

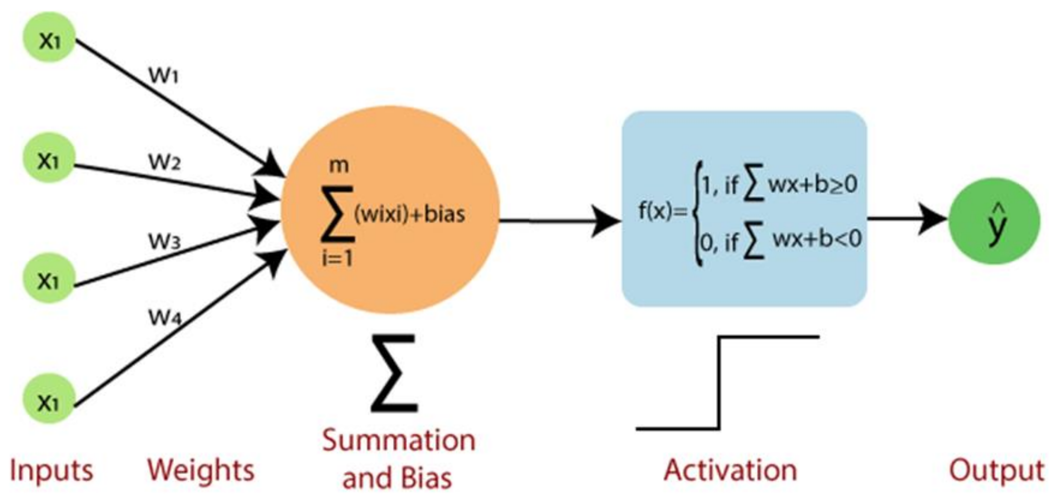

Figure 2: Perceptron architecture (Source: Yurek, I. (2023)

A perceptron learns through an iterative process of adjusting its weights and bias during training, guided by the perceptron learning algorithm. This algorithm begins by initializing the weights and bias with random values. Then, in a feedforward step, the perceptron takes the input features, multiplies them by their corresponding weights, sums them, adds the bias, and passes this weighted sum through the activation function to produce an output. This output is compared to the true label, and the difference is calculated as the error. The weights and bias

[15] An activation function in a neural network is like a gatekeeper that determines the output of a neuron based on its input. It introduces non-linearity, allowing the network to learn complex patterns. Common examples include sigmoid, ReLU, and tanh, each suited for different tasks and network architectures

are then adjusted proportionally to the error and input values to nudge the perceptron's output closer to the true label. This cycle of feedforward, error calculation, and weight update is repeated multiple times over the training data, gradually minimizing the error and allowing the perceptron to learn the relationship between input features and target output. This process effectively creates a linear decision boundary that separates different classes within the data (See Figure 2).

Although a single perceptron is limited to learning linear decision boundaries, their true potential is unleashed when multiple perceptron's are interconnected to form an artificial neural network. This network structure enables the learning and representation of more complex, non-linear relationships within data. Perceptron's are organized into layers: an input layer receiving the data, one or more hidden layers processing and transforming information, and an output layer producing the final result. Each layer can contain numerous perceptron's working simultaneously. The output of each perceptron in a layer becomes the input for perceptron's in the subsequent layer, creating a flow of information through the network. Crucially, each connection between perceptron's has an associated weight that determines the strength of the connection. During the training process, these weights are adjusted based on the discrepancy between the network's output and the desired output, allowing the network to learn and improve its accuracy over time.

3.4.2.2 Multi – Layer Perceptron (MLP)

A MLP model is a type of artificial neural network characterized by its feedforward, layered structure. This structure comprises an input layer, one or more hidden layers, and an output layer, all interconnected through weighted connections (See Figure 3). The input layer receives and represents the initial data features, which are then propagated through the network. Each hidden layer applies a series of transformations to the data, facilitated by activation functions that introduce non-linearity, enabling the MLP to model complex relationships. The output layer, in turn, generates the final prediction or classification based on the processed information. The learning process of an MLP is primarily driven by the backpropagation algorithm, which iteratively adjusts the weights of the connections between nodes to minimize the discrepancy between the network's output and the desired output. This process allows the MLP to progressively learn and refine its ability to extract meaningful patterns from the data, making it a powerful tool for a wide range of applications, including image recognition, natural language processing, and financial forecasting.

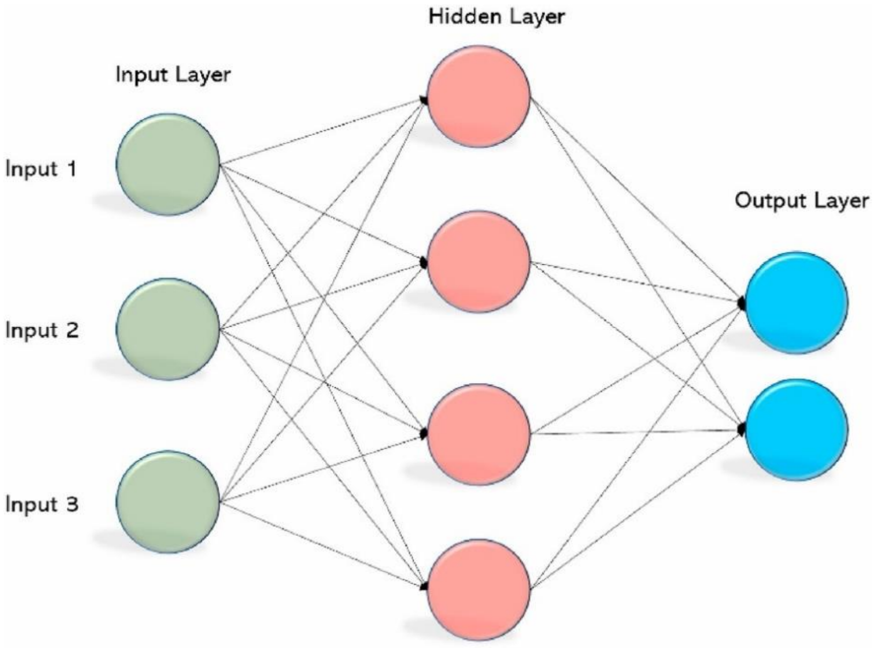

Figure 3: MLP architecture, source (Qureshi et. al, 2023)

While a variety of activation functions exist, this study employs two different activation functions, Hyperbolic tangent (Tanh)[16] and Rectified Linear Unit (Relu) activation functions for both hidden and output layers due to the anticipation of a numerical, regression. ReLU is a popular activation function due to its computational simplicity, leading to faster training, especially in deep networks. Its tendency for sparse activations promotes efficiency and feature selection by focusing on relevant inputs. ReLU also mitigates the vanishing gradient problem, but can suffer from "dying ReLU,"[17] addressed by variations like Leaky ReLU (Chasipanta and Sánchez-Pozo, 2024). The input layer, devoid of an activation function, facilitates the entry of any input variables. Subsequently, the learning process establishes weighted connections, enabling diverse input combinations to activate neurons within the hidden and output layers (see Figure 2). This mechanism empowers individual neurons, or combinations thereof, to

[16] Tanh, short for hyperbolic tangent, is an activation function that transforms input values into a range between −1 and 1. It is zero-cantered and commonly used in neural networks to model non-linear relationships.

[17] The "dying ReLU," refers to a case where neurons can get stuck in a state where they always output zero, effectively becoming inactive. Various modifications, like Leaky ReLU and Parametric ReLU, have been proposed to address this.

discern and represent non-linear patterns inherent in the data. The propagation of signals through each layer of the neural network can be formally expressed as:

$$X_j = W_{ij}.I, \tag{4}$$

$$\emptyset_j = activation(X_j) \tag{5}$$

where X_j denotes the matrix of resulting signals for a given layer j within the neural network. W_{ij} represents the matrix of weights associated with the connection between layer j and its preceding layer i. The input signal matrix is denoted as I, while ϕ_j represents the output signal matrix for each layer following the application of the activations function. The learning process of a neural network necessitates the evaluation of errors at each neuron in the final layer. This evaluation involves comparing the obtained output value with the expected value for each observation. The discrepancy between these values, termed the error, is computed using the formular $e_{out_k} = t_k - \vartheta_k$. Furthermore, this error is propagated backward through the entire neural network structure, tracing the pathways of signal origin to facilitate weight adjustments. This backpropagation process can be formally expressed as:

$$\mathcal{E}_i = W_{ij}^T.\mathcal{E}_j \tag{6}$$

where $\mathcal{E}_i$ represents the matrix of error that are propagated back to the preceding layer of the neural network, while $\mathcal{E}_j$ denotes the errors received from the subsequent layer. This backward propagated of errors facilitates the adjustment of connections weights within the network. The updated weights, in turn, enable the neural network to retain learned information from previous training examples while simultaneously incorporating new information from newly presented observations. Gradient descent, a widely employed optimization algorithm, serves to achieve the objective of weight adjustment within the neural network. This algorithm can be formally expressed as:

$$\frac{\partial \mathcal{E}}{\partial W_{jk}} = \frac{\partial \sum_n (t_n - \vartheta_n)}{\partial W_{jk}} = \frac{\partial \mathcal{E}}{\partial \emptyset_k}.\frac{\partial \emptyset_k}{\partial W_{jk}} = -2(t_n - \vartheta_n).\frac{\partial \emptyset_k}{\partial W_{jk}} = -2(t_n - \vartheta_n) \tag{7}$$

$$\frac{\partial}{\partial W_{jk}} activation\left(\sum_j W_{jk}.\emptyset_j\right) \tag{8}$$

$$W_{jk}^{(s+1)} = W_{jk}^{(s)} - \eta \frac{\partial \mathcal{E}}{\partial W_{jk}} \tag{9}$$

where $W_{jk}^{(s+1)}$ represents the newly updated weight for the connection between nodes j_k. This updated weight is calculated from its previous value, denoted as $W_{jk}^{(s)}$, by incorporating new information acquired during the learning process. The gradient term, $\frac{\partial \mathcal{E}}{\partial W_{jk}}$, quantifies the direction and magnitude of weight adjustment to minimize errors. This adjustment is moderated by the learning rate, η, a hyperparameter that governs the steps taken during weight updates. The learning rate plays a crucial role in stabilizing the optimizing the learning process.

3.4.3 Long Term Short Term Memory (LSTMs)

3.4.3.1 Recurrent Neural Networks

Recurrent Neural Networks (RNNs) distinguish themselves from traditional feedforward networks through their capacity to process sequential data by incorporating a memory element. This memory, embodied in a hidden state, is dynamically updated at each time step by integrating information from both the current input and preceding hidden state. This feedback mechanism allows RNNs to retain and contextualize information from previous steps in the sequence, mimicking the human ability to understand language or analyse time series data where context is crucial. RNNs develop a "memory" of the sequence processed thus far, enabling them to discern patterns and dependencies that span across multiple time steps (Olah, 2015). They were proposed by Hochreiter and Jürgen (1997) and improved and popularised by numerous others in subsequent studies. RNNs have built-in loops that facilitate information transfer between the neurons.

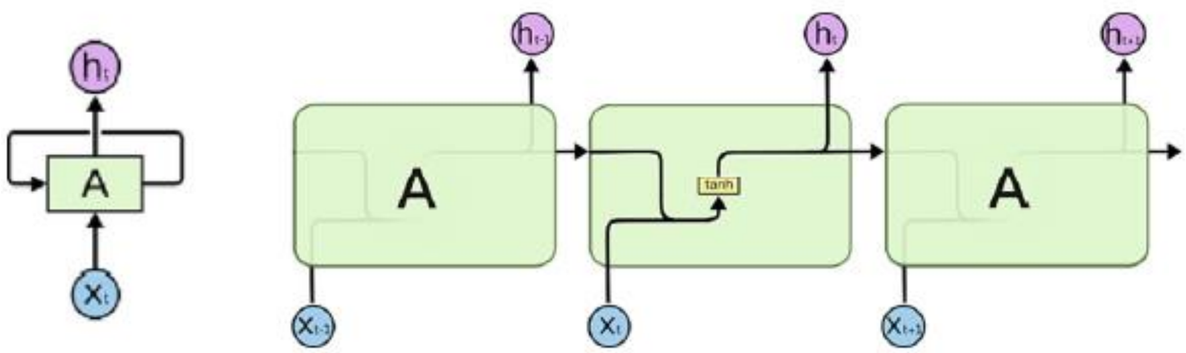

Figure 4: The Recurrent Neural Network's Architecture

Figure 4 (Left) illustrates the structure of a RNN, where A denotes the neural network model, the input and output signals are represented as Xt and h_t, respectively. The sequential nature of the RNN architecture makes it well-suited for processing time series data, as it can retain information from all previous time points. However, a significant limitation of the standard RNN model is its restricted ability to handle long-term dependencies in data processing, as shown in Figure 4, where a single neural network layer is inadequate for effective information processing. Researchers have developed enhanced RNN architectures to address this issue of long-term dependency, incorporating specialized interactions among multiple neural network layers.

3.4.3.2 LSTM

LSTM networks, a specialized variant of RNNs, address the shortcomings of traditional RNNs in capturing long-term dependencies within sequential data (Hochreiter and Jürgen, 1997). LSTMs accomplish this through a sophisticated gating mechanism integrated into their memory cells. The input, forget, and output gates regulate the flow of information into and out of the cell state, which functions as a "conveyor belt" transporting information across time steps (Zhao et al. 2023). The forget gate selectively discards irrelevant past information, while the input gate determines the new data to be stored. The output gate then controls the flow of information from the cell state to the network's output, ensuring that only the relevant information influences subsequent predictions. This intricate mechanism enables LSTMs to maintain a selective and persistent memory of past data, making them particularly adept at processing and forecasting time series data where long-term dependencies are prevalent, such as in exchange rate prediction. The chain-like structure of the LSTM model is depicted in Fig. 5.

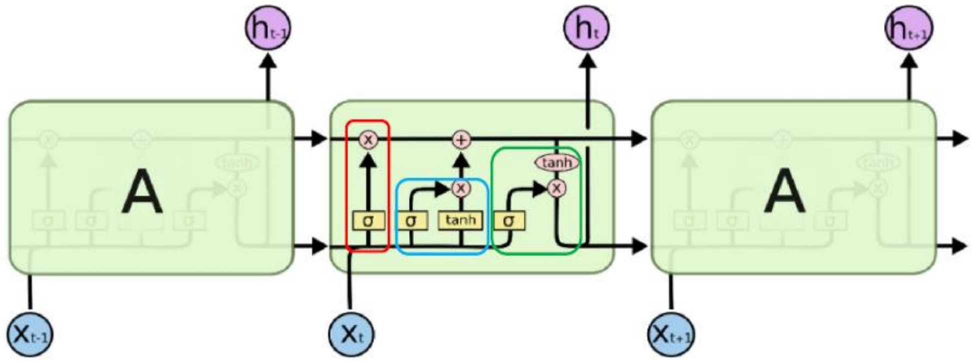

Figure 5: Architecture of the LSTM model

The rectangular box 1 (Left) represents the forgetting gate, which includes a sigmoid neural network layer[18] with parameters W_f and b_f representing the weights and bias associated with the forget gate. This layer incorporates a per-bit multiplication operation to aid in selecting and disregarding nonessential information. Its function involves taking input signal x_t at time t and the previous output signal h_{t-1} of the LSTM at time $t - 1$, combining these two signals into the sigmoid neural network, and producing an output signal $f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f)$. The value of f_t ranges between 0 and 1; it is then multiplied with the previous stage state C_{t-1} to determine which information in C_{t-1} should be discarded.

The memory gate is depicted in box 2 (the middle box) as a combination of a sigmoid neural network layer (referred to as the input gate, with parameters W_c and b_c), a tanh neural network layer (this study uses RELU activation function with parameters W_c and b_c), and a component involving element-wise multiplication. This aids in determining which of the input information x_t and h_{t-1} will be preserved. The sigmoid neural network layer takes x_t and h_{t-1} as input, producing an output between 0 and 1 that specifies which information should be updated: $i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i)$. The tanh neural network layer combines x_t with h_{t-1} to generate a new state candidate vector represented by $\check{C}_t$, where the value of $\check{C}_t$, lies between -1 and 1.

The output of the memory gate is determined jointly by the two neural network layers. This output is then used to select which information will be added to the state C_t at time t, after being modified and combined with new information. At this stage, $C_t = f_t * C_{t-1} + i_t * \check{C}_t$, illustrating that the state C_t at time t contains both discarded information from time $t - 1$ as

[18] A sigmoid layer in a neural network applies the sigmoid activation function to a weighted sum of its inputs, producing an output between 0 and 1. This introduces non-linearity and is often used for binary classification tasks

well as newly acquired information at time t. Subsequently, C_t will be transferred to the LSTM network at time $t + 1$.

The output gate is represented by a rectangular box 3 (right). x_t and h_{t-1} pass through the sigmoid neural network layer (with parameters W_o and b_o) in order to produce the value o_t, which falls between 0 and 1. This can be expressed as: $o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o$. C_t undergoes multiplication with o_t after application of the tanh function to obtain the output signal h_t, where $h_t = o_t * \tan C_t$; this signal is then forwarded as input to the subsequent stage for further processing. Zhao et al. (2023) notes the LSTM neural networks are suitable for analysing unstable time series data and understanding the characteristics of such data, leading to improved accuracy in hierarchical warning and prediction. As a result, they are well suited for the early detection of exchange rate risks.

3.5 Model Validation

Rigorous model validation is essential to ensure the reliability, accuracy, and generalizability of forecasting models, especially in financial time series analysis. This study utilizes a structured validation framework, including train-test splitting and TimeSeriesSplit cross-validation, to maintain the chronological integrity of data and prevent data leakage, ensuring that models reflect real-world forecasting scenarios. The study also optimizes key model hyperparameters, such as the number of trees in RF, the layers in MLPs, and the learning rate in LSTMs, using Randomized Search and Grid Search to enhance predictive accuracy. Additionally, L2 regularization is applied to mitigate overfitting by discouraging excessive model complexity. By integrating these techniques, the study conducts a thorough and unbiased evaluation of traditional econometric models and machine learning models for forecasting the ZAR, enhancing their practical applicability in real-world financial environments.

3.5.1 Train – test split

The train-test split is a fundamental machine learning technique that evaluates a model's ability to generalize to new, unseen data. By partitioning the dataset into distinct training and test subsets, the model learns patterns from the training data and is subsequently assessed on the test set, which it has not encountered before. This approach helps mitigate overfitting, where a model performs well on training data but fails to adapt to new observations. In the context of the ARIMAX-GARCH model, this split is particularly important, as the ARIMAX component captures linear dependencies using exogenous variables, while the GARCH component models volatility. By applying a train-test split, this study ensures that both trend dynamics and volatility patterns are rigorously evaluated on unseen data, strengthening the robustness of the forecasting models.

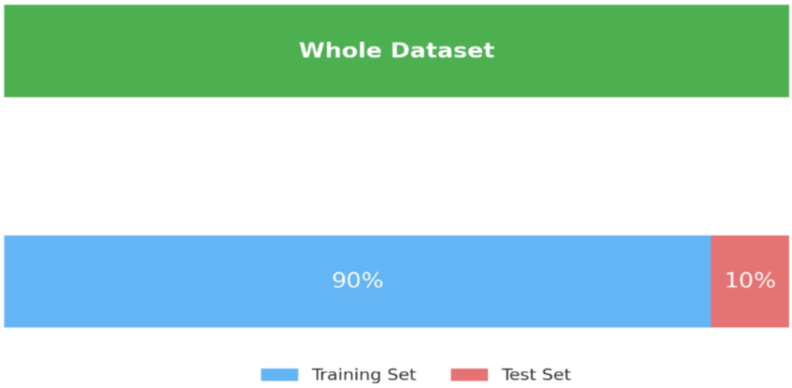

Figure 6: Train-test split (Source: Own compilation)

The whole dataset will be split into two subsets, the Training set which will be 5,128 observation (90% of the whole data) and the Testing set which will be 641 observations that is, 10% of the entire data (see Figure 6). The training period, spanning January 3, 2000, to August 28, 2019, encompasses a wide range of economic conditions, including the 2008 financial crisis and subsequent recoveries, providing the models with diverse market scenarios to learn from. This extensive training set allows advanced machine learning models like MLP, LSTM, and RFs to capture complex, nonlinear patterns, ensuring better generalization across various economic conditions. Within the training period, TimeSeriesSplit will be used to simulate multiple validation phases by progressively expanding the training set and using later data for validating the ML models. This approach helps fine-tune the model parameters and improve their robustness.

The 10% test set spanning from February 14, 2022, to July 29, 2024, is deliberately selected to rigorously evaluate the models' performance on unseen, real-world data. This period encompasses the unprecedented market disruptions stemming from the COVID-19 pandemic, geopolitical tensions, and shifts in global monetary policy. By reserving this timeframe for out-of-sample testing, the study ensures a comprehensive assessment of each model's ability to adapt to evolving economic conditions, capture exchange rate volatility, and maintain forecasting accuracy in dynamic financial environments. This robust validation strategy enhances the reliability and practical applicability of the models for exchange rate forecasting purposes.

3.5.2 Time Series Split cross – validations

TimeSeriesSplit cross-validation[19] is a specialized technique for validating time-series forecasting models. It preserves the temporal structure of the data, addressing a key limitation of traditional k-fold[20] cross-validation. Unlike standard methods that randomly shuffle and split the data, TimeSeriesSplit ensures that training always precedes validation, preventing data leakage, where future information inadvertently influences model training and leads to overly optimistic performance metrics. This approach closely mirrors real-world forecasting scenarios, where predictions are made based solely on past data.

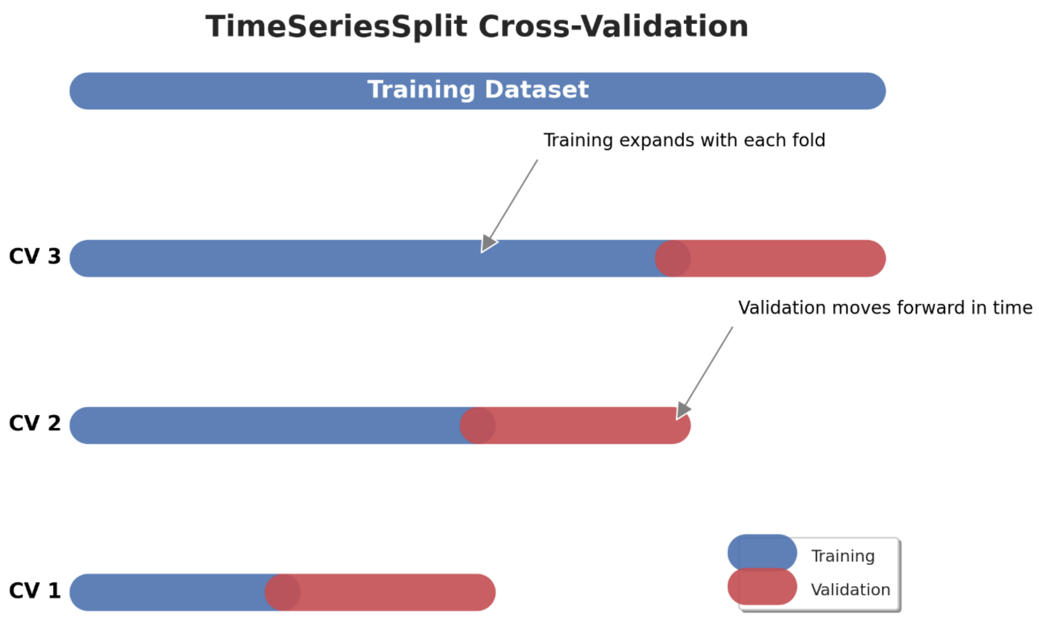

Figure 7: TimeSeriesSplit Algorithm (Source: Own compilation)

---

19 Cross-validation is a statistical technique used in machine learning to assess a model's predictive performance by partitioning the dataset into multiple subsets. The model is trained on one subset and tested on another, ensuring that it generalizes well to unseen data. This process helps detect overfitting and optimizes model selection by evaluating its performance across different data splits.

20 K-Fold cross-validation is a method used to evaluate machine learning models by dividing the dataset into **K** equally sized subsets (folds). The model is trained on **K-1** folds and tested on the remaining fold, and this process is repeated **K times**, with each fold serving as a test set once. The final performance is averaged across all iterations, providing a more reliable estimate of the model's generalization ability while reducing variance compared to a single train-test split.

A distinguishing feature of TimeSeriesSplit is its progressive expansion of the training dataset while maintaining distinct validation sets for each fold. At each step, a portion of the dataset is used for training, while the subsequent, chronologically ordered segment is reserved for validation. This sequential evaluation provides insights into the model's ability to generalize across different time horizons and adapt to changing exchange rate patterns. By testing the model on increasingly unseen data, this method helps assess its robustness, stability, and forecasting accuracy over multiple time segments.

This cross-validation approach is particularly valuable in financial time series analysis, where exchange rate movements exhibit temporal dependencies and structural shifts. Evaluating model performance over multiple sequential folds allows for a more comprehensive assessment of predictive power, identification of inconsistencies, and refinement of model parameters, leading to more reliable and robust forecasting results. As such, TimeSeriesSplit is an essential component in validating both econometric models and machine learning models for exchange rate prediction.

3.5.3 Hyperparameter tuning

Hyperparameter tuning is the process of optimizing the external settings of a machine learning model to enhance its performance. These hyperparameters guide the learning process but are not derived from the data. Instead, they must be manually configured before training. One efficient technique for hyperparameter tuning is Randomized Search. This approach involves randomly sampling from a predefined distribution of hyperparameter values, rather than exhaustively testing every possible combination as in Grid Search. Randomized Search offers computational efficiency while still exploring a wide range of possibilities, making it particularly useful for models with many hyperparameters or a large search space. This method provides flexibility and faster results, making it well-suited for complex models.

For RFs, MLPs, and LSTMs models, Randomized Search is especially useful because these models often have numerous hyperparameters to tune, such as the number of trees in RF, the number of neurons and layers in MLP, or the learning rate and dropout rates in LSTMs. Randomized Search helps find near optimal hyperparameters without the heavy computational cost of trying all combinations, and it can lead to better generalization of the model by finding hyperparameter settings that perform well across a diverse subset of data. This balance between exploration and efficiency makes it particularly well-suited for deep learning models (like LSTM) and large-scale models (like RF).

3.5.4 Regularization

L2 regularization, also referred to as Ridge regularization, is a technique employed to mitigate overfitting in machine learning models. This is achieved by incorporating a penalty term into the loss function, which is proportional to the sum of the squared model weights. The rationale behind L2 regularization is that it discourages the model from assigning disproportionately large weights to any feature, thereby enabling the model to concentrate on the most salient patterns in the data rather than fitting to noise or irrelevant variables. Fundamentally, it compels the model to learn more generalized, smooth relationships by shrinking the weight values towards, but not exactly to, zero, in contrast with the case of L1 regularization.

The use of L2 regularization is particularly advantageous for deep learning models like MLPs and LSTMs, which are prone to overfitting, especially when dealing with complex or noisy data. In MLPs, which have a multi-layered architecture with numerous neuron connections, L2 regularization helps prevent the model from becoming overly sensitive to specific training examples. Similarly, for LSTMs, which are often employed to model time-series data with lengthy sequences, L2 regularization can smooth the learning process, encouraging the model to focus on overall trends rather than individual variations. By constraining the weights, L2 regularization enables both MLP and LSTM models to generalize more effectively to unseen data, resulting in more stable and reliable predictions.

3.6 Forecast Evaluations

This research focuses on one-day-ahead forecasts, enabling an evaluation of the models' short-term performance, aligning with real-world financial applications such as currency trading, risk management, risk management, and portfolio adjustments. The employed methodology utilizes an iterative forecasting approach, where a one-step-ahead model is iteratively applied to generate predictions for the required number of periods. This technique facilitates a comparative assessment of the efficiency and accuracy of the forecasting models under investigation.

Hewamalage et al. (2022) suggested employing a variety of error metrics, such as the Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and Mean Absolute Percentage Error (MAPE), to capture different dimensions of forecast accuracy and bias. In this study, four performance metrics are used to evaluate the performance of the proposed hybrid traditional model and ML methods, namely, MAE, RMSE, and MAPE and the Symmetric Mean Absolute Percentage Error (sMAPE) which divides the absolute difference between the forecast and actual value by the average of the absolute values of the actual and forecasted data, which helps to normalize the error measure and make it symmetric, Makridakis et al. (2018). The sMAPE is a valuable metric for evaluating forecasting accuracy. It offers a balanced assessment, treating positive and negative deviations from actual values equally, thus avoiding biases towards over- or under-estimation. Additionally, sMAPE's normalization by the average of actual and predicted values makes it scale-independent, enabling comparisons of forecasting performance across diverse datasets and time series with varying magnitudes. This means it treats overestimation and underestimation equally. The lowest values of the metrics indicate the best performance. Letting $\boldsymbol{Y_i}$ and $\boldsymbol{\widehat{Y}_i}$ represent the actual value and the predicted value respectively, the following equations show how each measure is calculated:

$$\boldsymbol{MAE = \frac{1}{n}\sum_{i=1}^{n}\left|\widehat{Y}_i - Y_i\right|} \tag{10}$$

$$\boldsymbol{RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\widehat{Y}_i - Y_i)^2}} \tag{11}$$

$$\boldsymbol{MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{\widehat{Y}_i - Y_i}{Y_i}\right|} \tag{12}$$

$$\boldsymbol{sMAPE = \frac{2}{n}\sum_{i=1}^{n}\frac{\left|\widehat{Y}_i - Y_i\right|}{|Y_i| + \left|\widehat{Y}_i\right|}} \tag{13}$$

Directional forecasting accuracy is also considered by measuring the model's ability to correctly predict the direction of the exchange rate movement (up or down)**.** This can be assessed using metrics such as the MDA). It measures the proportion of correct directional predictions out of the total predictions made. The MDA is calculated by comparing the predicted and actual directions of change for each time step, with an indicator function counting correct predictions. An MDA of 0.5 suggests random guessing, while values significantly above 0.5 indicate good directional accuracy. This metric is particularly useful in financial forecasting, where predicting the trend direction is crucial and is given by

$$\boldsymbol{MDA = \frac{1}{T}\sum_{t=1}^{T} I(sign(\hat{y}_{t+1} - y_t) = sign(y_{t+1} - y_t))}$$

where $\boldsymbol{I}$ is an indicator function that equals 1 if the predicted direction matches the actual direction and 0 otherwise; $\boldsymbol{T}$ is the total number of predictions; $\boldsymbol{\hat{y}_{t+1}}$ is the predicted value; $\boldsymbol{y_t}$ and $\boldsymbol{y_{t+1}}$ are the actual values.

4. In-Sample Analysis

This section presents the in-sample analysis, which evaluates how well each forecasting model fits the training data. The purpose is to assess the models' ability to capture the underlying exchange rate dynamics before applying them to unseen data. In-sample performance offers insight into each model's stability, accuracy, and potential overfitting risk across different validation folds[21], enabling a robust comparison of their predictive capabilities. The performance assessment relies on the key metrics introduced earlier, namely RMSE, MAE, MAPE, sMAPE, and MDA. In Section 4.1, the in-sample results for each model across the currency pairs are presented, while Section 4.2 presents a comparison of the results of the models by currency pair. The findings indicate that the RF attains the lowest error metrics, while the ARIMAX-GARCH model exhibits superior directional accuracy for the USDZAR and CNYZAR currency pairs. Conversely, MLP demonstrates strong performance for GBPZAR and CNYZAR but with higher percentage errors. The LSTM model effectively captures sequential dependencies but struggles with highly volatile currency pairs like JPYZAR. These empirical insights establish a robust foundation for subsequent out-of-sample evaluation, ensuring that only well-calibrated forecasting models are employed for real-world applications.

4.1 Model fitting

The process of preparing the dataset for model training commenced with a meticulous and systematic approach, ensuring consistency and comparability across various modelling methodologies. The raw dataset encompassed daily exchange rates for the five currencies, alongside the exogenous financial and economic indicators (gold prices, Brent crude oil prices, the S&P 500 index, the Baltic Dry Index, the volatility index, and the Commodity Research Bureau index). To stabilize the variance and mitigate the effects of skewness in the data,

[21] Validation folds refer to the number of subsets into which the training data is divided during cross-validation. In k-fold cross-validation, the dataset is split into k equal parts (folds). The model is trained on k-1 folds and validated on the remaining fold, repeating the process k times with a different validation set each time. This technique helps in assessing model performance more reliably by reducing overfitting and ensuring that the model generalizes well to unseen data.

logarithmic transformations were applied to the exchange rates and the explanatory variables. Furthermore, the daily log returns were calculated for exchange rates and other variables to capture short-term changes, while 22-day rolling volatilities were computed to reflect trends and variability over a typical monthly window.

The transformations facilitated the models' ability to capture short-term dynamics and medium-term volatility. To enhance predictive performance, lagged variables were generated for exchange rates and exogenous factors, enabling the MLP, RF, and LSTM models to leverage temporal patterns. A RobustScaler was applied to standardize the features, mitigating the influence of outliers. The dataset was partitioned, with 90% allocated for in-sample training and 10% reserved for out of sample testing. The in-sample phase ensured the models effectively captured historical exchange rate trends, providing a solid foundation for evaluating their generalization to unseen data.

$4.1.1\ ARIMAX - GARCH$

The ARIMAX model selection process utilized an automated tool, the $auto\_arima$ function from the $pmdarima$ library in Python[22], to systematically determine the optimal ARIMA parameters (p, d, and q) for each currency by minimizing the AIC, a metric that balances goodness-of-fit and model complexity. The inclusion of exogenous variables enhanced the model's explanatory power, enabling it to capture the impact of external economic factors on exchange rates. To address the time-varying nature of volatility, the residuals from the fitted ARIMAX model were analysed using a GARCH(1,1) model, which estimates volatility based on past values. This combined ARIMAX-GARCH approach optimizes both the mean and variance equations. As shown in Table 2, the model exhibited strong in-sample performance across all five currency pairs, evidenced by low RMSE, MAE, MAPE, and SMAPE values, as well as high MDA.

For EURZAR, the model demonstrates moderate error performance, with RMSE of 0.0805 and MAE of 0.0218, while the MAPE and sMAPE values indicate reasonable percentage accuracy (0.1951 and 0.2081, respectively). The high Directional Movement Accuracy (87.6%)

[22] This is equivalent to R's $auto.arima$ function. See https://pypi.org/project/pmdarima/

confirms the model's strength in correctly predicting exchange rate movements. In the case of USDZAR, the model performs exceptionally well, achieving the lowest MAPE and sMAPE among all currency pairs. The low RMSE and MAE values indicate high precision, and the MDA highlights its strong directional accuracy (96.9%), making USDZAR the best-performing currency pair under ARIMAX on this metric.

For JPYZAR, the model excels in minimizing absolute errors, with an RMSE of 0.006 and an MAE of 0.0003, but struggles with higher percentage errors (the MAPE is 0.3504 and the sMAPE is 0.2466). This suggests difficulties in capturing percentage-based exchange rate fluctuations, despite maintaining a strong directional accuracy of 85.3%. Regarding GBPZAR, the model exhibits moderate error performance, with higher RMSE and MAE compared to other pairs. The MAPE and sMAPE indicate a reasonable ability to capture percentage deviations, while the MDA confirms its reliability in predicting trend directions. For CNYZAR, the model emerges as the best performer in error minimization, achieving the lowest RMSE, MAE, MAPE, and sMAPE across all currency pairs. This highlights its exceptional precision in capturing exchange rate variations, and the MDA further emphasizes its high directional forecasting accuracy.

Overall, the ARIMAX-GARCH model delivers consistent in-sample performance across all currencies. The model's ability to balance low forecasting errors with high directional accuracy suggests that it is particularly well-suited for analysing historical exchange rate trends, making it a valuable tool for capturing macroeconomic patterns and financial volatility.

Table *2*: ARIMAX-GARCH In-Sample Results

$\boldsymbol{Currency}$	$\boldsymbol{ARIMA\ Order}$	$\boldsymbol{RMSE}$	$\boldsymbol{MAE}$	$\boldsymbol{MAPE}$	$\boldsymbol{SMAPE}$	$\boldsymbol{MDA}$
$EURZAR$	$(2, 1, 3)$	0.0805	0.0218	0.1951	0.2081	87.6214
$USDZAR$	$(1, 1, 1)$	0.0749	0.0084	0.0821	0.0952	96.862
$JPYZAR$	$(1, 1, 1)$	0.006	0.0003	0.3504	0.2466	85.2982
$GBPZAR$	$(2, 1, 0)$	0.1321	0.0308	0.2080	0.2227	87.708
$CNYZAR$	$(0, 1, 2)$	0.0024	0.0011	0.0698	0.0700	95.9085

4.1.2 Multilyer Perceptron (MLP) Model

The MLP models were trained using a robust pipeline that included data preprocessing, lagged feature creation, and hyperparameter optimization. Lagged values of both target and exogenous variables were generated to account for historical dependencies, with up to ten lags included for each variable. Data were scaled using a RobustScaler to mitigate the influence of outliers, and the training set was split using time-series cross-validation (5 splits) to ensure robust model evaluation. Hyperparameter tuning was performed via $RandomizedSearchCV$, allowing the exploration of a wide range of parameter combinations, with the goal of identifying the optimal MLP configurations tailored to the unique dynamics of each currency pair. Once trained, the models were evaluated using in-sample metrics, capturing their ability to fit historical data effectively.

The $RandomizedSearchCV$ approachleverages a pre-defined parameter grid to explore various configurations and balance computational efficiency with model flexibility. The grid included key hyperparameters such as hidden layer sizes, which ranged from simpler single-layer architectures (e.g., 100 or 150 nodes) to more complex networks like (100, 50) and (200, 100, 50). The inclusion of both ReLU and tanh activation functions allowed the models to capture non-linear relationships effectively. Regularization strength, controlled through the alpha parameter, was explored over values ranging from 0.0001 to 0.1 to balance underfitting and overfitting. Learning dynamics were tuned with both constant and adaptive learning rates, and initial learning rate values were tested across 0.0001, 0.001, and 0.01 to optimize convergence. Finally, the maximum iterations parameter was fixed in 2000 to ensure sufficient training iterations for robust convergence[23].

Table 3 presents a summary of the in-sample performance metrics and the corresponding hyperparameters for the MLP models trained to predict currency exchange rates. Each currency's results are analysed based on the key evaluation metrics discussed in section 3, i.e. RMSE, MAE, MAPE, SMAPE, and MDA (%), alongside the optimal hyperparameters identified for each model. For the EURZAR currency pair, the model exhibits a low RMSE

[23] Convergence in machine learning refers to the point during training when a model's loss function stabilizes, indicating that further training will not significantly improve performance. This occurs when the model has effectively minimized the error between predicted and actual values, ensuring that the learning process has reached an optimal state. If a model does not converge, it may either underfit (stop learning too soon) or overfit (continue learning beyond optimal performance), leading to poor generalization on unseen data.

and SMAPE, suggesting a strong fit to the data and a reduction in absolute forecasting errors. However, its MDA of 48.4% implies only moderate success in capturing the directionality of exchange rate changes. The best-performing configuration for EURZAR comprises a two-layer neural network architecture, employing an adaptive learning rate, a small initial learning rate, minimal regularization, and the ReLU activation function, which enables the efficient capture of nonlinear relationships.

Regarding the USDZAR currency pair, the error metrics are slightly higher, with an RMSE of 0.0175 and an SMAPE of 0.5853. Yet, the model's directional accuracy improves slightly, reaching an MDA of 51.7%. This USDZAR-specific model adopts a similar two-layer architecture but incorporates a higher initial learning rate and stronger regularization, likely to compensate for the increased volatility characteristics of the USDZAR exchange rate. The adaptive learning rate strategy facilitates the model's convergence, enabling it to dynamically adjust to the fluctuations inherent in the exchange rate data.

For JPYZAR, the model achieved the highest directional accuracy of 74.99%, showcasing its ability to capture historical trends effectively. The error metrics, including RMSE and SMAPE of 0.0209 and 0.8230 respectively, were higher compared to other currencies, reflecting the inherent stability and complexity of this exchange rate. The optimal architecture for JPYZAR involved a deeper network with three hidden layers of sizes (200, 100, 50), a constant learning rate ($learning\_rate$: constant), an initial learning rate of 0.01, strong regularization (alpha: 0.01), and the ReLU activation function.

Regarding the GBPZAR currency pair, the model delivers the highest precision among all currency pairs in Table 3, achieving the lowest RMSE and SMAPE, along with a high MDA (74.5%). The model follows a two-layer architecture with an adaptive learning rate, an initial learning rate of 0.01, and moderate regularization. The combination of low absolute errors and strong directional accuracy suggests that the GBPZAR exchange rate exhibits predictable patterns, making it well-suited for MLP-based forecasting.

Finally, the CNYZAR model demonstrates the highest directional accuracy (76.7%), indicating its strong trend-capturing capability. However, despite low RMSE and SMAPE, the high MAPE suggests difficulties in predicting larger percentage deviations, likely due to the high volatility and extreme price swings associated with the CNYZAR exchange rate. The best configuration for this model follows a three-layer structure with a constant learning rate, an initial learning rate of 0.01, strong regularization, and ReLU activation—a setup similar to the JPYZAR model, optimized for capturing complex, nonlinear market movements.

MLP models excel in minimizing absolute errors, especially for GBPZAR and EURZAR, while achieving the highest directional accuracy for CNYZAR and JPYZAR. Deeper architectures enhance performance in volatile pairs, improving feature extraction. Adaptive learning rates benefit moderately volatile currencies, while higher regularization stabilizes forecasts in highly volatile environments. These results highlight the need to tailor MLP models to each currency pair for optimal performance.

Table *3*: MLP In-Sample Results

Currency	*RMSE*	*MAE*	*MAPE*	*SMAPE*	*MDA* (%)	*Best Hyperparameters*
EURZAR	0.0135	0.01	0.4153	0.4155	48.38	*max_iter*: 2000, *learning_rate_init*: 0.001, *learning_rate*: adaptive, *hidden_layer_sizes*: (100, 50), *alpha*: 0.0001, *activation*: relu
USDZAR	0.0175	0.0137	0.5837	0.5853	51.69	*max_iter*: 2000, *learning_rate_init*: 0.01, *learning_rate*: adaptive, *hidden_layer_sizes*: (100, 50), *alpha*: 0.01, activation: relu
JPYZAR	0.0209	0.019	0.8273	0.823	74.99	*max_iter*: 2000, *learning_rate_init*: 0.01, *learning_rate*: constant, *hidden_layer_sizes*: (200, 100, 50), *alpha*: 0.01, *activation*: relu
GBPZAR	0.0109	0.0086	0.3209	0.3209	74.49	*max_iter*: 2000, *learning_rate_init*: 0.01, *learning_rate*: adaptive, *hidden_layer_sizes*: (100, 50), alpha: 0.01, activation: relu
CNYZAR	0.0147	0.0119	10.1949	7.156	76.74	*max_iter*: 2000, *learning_rate_init*: 0.01, *learning_rate*: constant, *hidden_layer_sizes*: (200, 100, 50), *alpha*: 0.01, *activation*: relu

4.1.3 Random Forest model

The RF model employed an extensive hyperparameter optimization process to fine-tune the model's parameters for each currency pair. Using a $RandomizedSearchCV$ method, twenty candidate hyperparameter configurations were evaluated via 5-fold cross-validation, resulting in a total of 100 model evaluations per currency. The hyperparameter grid included the number of estimators ($n\_estimators$), minimum samples required for a split ($min\_samples\_split$), minimum samples required for a leaf node ($min\_samples\_leaf$), the maximum sample fraction used for training ($max\_samples$), the maximum depth of each $tree$ (max_depth), whether to use bootstrap sampling (bootstrap), and the maximum number of features considered for a split ($max\_features$). Table 4 presents a comprehensive summary of the in-sample performance metrics and the corresponding hyperparameters for the RF models trained to predict currency exchange rates.

The RF model for the EURZAR currency pair was configured with 500 estimators, a maximum depth of 30, and a training sample fraction of 0.9. Feature importance analysis revealed that CNYZAR and JPYZAR were the most influential predictors, highlighting the interconnected nature of these currency pairs. The model exhibited an exceptionally low Root Mean Squared Error and MAE, indicating its precision in minimizing absolute forecasting errors. The low MAPE and SMAPE suggest strong percentage-based predictive accuracy, while the Directional Movement Accuracy of 91.4% confirms the model's ability to effectively capture directional movements, making it particularly valuable for financial applications that require both accuracy and directional insights.

Table 4: Random forest In-Sample Results

$Currency$	$RMSE$	MAE	$MAPE$ (%)	$SMAPE$ (%)	MDA (%)	$Best\ Hyperparameters$
$EURZAR$	0.0037	0.0025	0.1063	0.1063	91.42	$n\_estimators$: 500, $min\_samples\_split$: 2, min_samples_leaf: 2, $max\_samples$: 0.9, $max\_features$: None, max_depth: 30, $bootstrap$: True
$USDZAR$	0.003	0.0017	0.0747	0.0747	94.9	$n\_estimators$: 500, $min\_samples\_split$: 5, $min\_samples\_leaf$: 1, $max\_samples$: 0.8, $max\_features$: None, max_depth: 15, $bootstrap$: True
$JPYZAR$	0.0057	0.0035	0.1495	0.1494	92.27	$n\_estimators$: 300, $min\_samples\_split$: 2, $min\_samples\_leaf$: 2, $max\_samples$: 0.8, $max\_features$: None, max_depth: 20, $bootstrap$: True
$GBPZAR$	0.006	0.0043	0.1588	0.1588	85.83	$n\_estimators$: 500, $min\_samples\_split$: 10, $min\_samples\_leaf$: 2, $max\_samples$: None, $max\_features$: sqrt, max_depth: 15, $bootstrap$: True
$CNYZAR$	0.0038	0.002	1.5038	1.3768	94.35	$n\_estimators$: 200, $min\_samples\_split$: 10, $min\_samples\_leaf$: 4, $max\_samples$: None, $max\_features$: None, max_depth: 20, $bootstrap$: True

The USDZAR model, configured with 500 estimators, a maximum depth of 15, and a training sample fraction of 0.8, delivered impressive in-sample performance. It achieved the lowest RMSE and MAE among all models, demonstrating its superior ability to minimize prediction errors. The low MAPE and SMAPE of 0.0747 further highlight the model's capability in capturing percentage deviations with high precision. With an MDA of 94.9%, the model reliably forecasts directional trends, underscoring its suitability for financial applications where understanding upward and downward movements is critical. Feature importance analysis indicated that lagged values of GBPZAR and CNYZAR were the most influential predictors, reflecting the strong interdependencies among these exchange rates.

For the JPYZAR currency pair, the RF model utilized 300 estimators, a maximum depth of 20, and a training sample fraction of 0.8, allowing it to adapt to the currency's inherent volatility and complexity. The model achieved an RMSE of 0.0057 and MAE of 0.0035, indicating effective error minimization. The MAPE and SMAPE suggest slightly higher percentage-based errors, due to the currency's sharp fluctuations and stability periods. However, the MDA of 92.3% confirms the model's strong trend-following ability.

The GBPZAR model exhibited robust in-sample performance, with an optimal configuration comprising 500 estimators, a maximum depth of 15, and the square root of feature count as the maximum number of features. This balanced complexity and generalization, enabling the model to effectively capture the intricate patterns within the GBPZAR currency pair. Feature importance analysis identified USDZAR and CNYZAR as the most significant predictors, underscoring the interconnected nature of these exchange rates. Although the model's RMSE and MAE indicate slightly higher absolute errors, the low MAPE and SMAPE of 0.1588 highlight its accuracy in capturing percentage-based changes. The Directional Movement Accuracy of 85.8% confirms the model's reliability in forecasting directional trends, a crucial capability for predicting exchange rate movements in financial markets.

The CNYZAR model demonstrated robust in-sample performance, with an optimal configuration comprising 200 estimators, a maximum depth of 20, and a minimum sample split of 10. These hyperparameters ensured a balance between model complexity and generalization, enabling the model to effectively capture the inherent nonlinear relationships within the CNYZAR exchange rate. The model achieved a low RMSE of 0.0038 and MAE of 0.002, showcasing its ability to minimize absolute forecasting errors. Despite the high MAPE and sMAPE, which indicate challenges in capturing percentage-based fluctuations, the model's Directional Movement Accuracy of 94.4% suggests exceptional proficiency in forecasting directional trends. This high Directional Movement Accuracy underscores the model's effectiveness in capturing trend movements, a critical factor for financial decision-making. Feature importance analysis highlights strong interactions with the USDZAR and EURZAR exchange rates, reflecting the interconnected nature of exchange rates and reinforcing the model's ability to incorporate broader macroeconomic influences.

4.1.4 LSTM

The LSTM model estimation process commenced with data preprocessing, including logarithmic transformations and feature scaling. Lagged features and volatility measures were incorporated to capture temporal dependencies and market dynamics. The model architecture was tailored for each currency pair, with three hidden layers optimized for units, regularization, and dropout to mitigate overfitting. The Adam optimizer[24] was employed for efficient and stable training. Early stopping prevented overfitting and ensured generalization. Evaluation metrics assessed the model's accuracy in capturing magnitude and directional trends. Hyperparameter tuning optimized units, dropout, regularization, and learning rates for balance. By training each pair independently, the LSTM model demonstrated a robust and flexible approach to complex financial data.

The LSTM model for the EURZAR currency pair was structured with a three-layer architecture, comprising 100, 150, and 200 units, respectively. Dropout rates of 0.4, 0.4, and 0.2 were employed to mitigate overfitting, while L2 regularization values of 0.000997, 0.000169, and 0.000557 were applied to enhance generalization. This configuration enabled the model to achieve a low RMSE of 0.0434 and MAE of 0.0355, indicating its strong ability to minimize prediction errors. The MAPE and SMAPE confirmed the model's capability in handling percentage-based errors. However, the Mean Directional Accuracy of 50.28% suggests moderate performance in forecasting the direction of exchange rate movements, indicating that while the model is effective in approximating exchange rate values, its ability to predict upward and downward trends requires improvement.

[24]The Adaptive Moment Estimation (Adam) optimizer is an advanced gradient descent optimization algorithm that combines the benefits of momentum and adaptive learning rate methods. It calculates individual learning rates for each parameter based on first-order momentum (mean of past gradients) and second-order momentum (uncentered variance of past gradients). Adam is widely used in deep learning due to its efficiency, adaptive learning rates, and ability to handle sparse gradients. It helps accelerate convergence while maintaining stability, making it particularly effective for training neural networks such as Multi-Layer Perceptrons (MLPs) and Long Short-Term Memory (LSTM) models.

The USDZAR LSTM model utilized a three-layer architecture, with dropout rates of 0.2, 0.1, and 0.1, and L2 regularization values of 0.001189, 0.000330, and 0.000336. This balance of complexity and regularization allowed the model to perform well in error minimization, achieving an RMSE of 0.0550 and MAE of 0.0465, showcasing its precision in numerical forecasting. However, the MAPE and SMAPE were slightly elevated, due to the heightened volatility of the USDZAR exchange rate. The MDA of 50.4% highlights the model's limited effectiveness in predicting directional trends, suggesting that while the model accurately forecasts exchange rate values, it struggles with consistent directional accuracy.

For the JPYZAR currency pair, the model employed a three-layer architecture with 150 units per layer. This configuration incorporated dropout rates of 0.1, 0.5, and 0.4, as well as L2 penalties of 0.000546, 0.000135, and 0.000104. These design choices aimed to control overfitting while ensuring the model's robustness. The model demonstrated strong performance in error minimization, as evidenced by the reported RMSE and MAE values, which reinforced its ability to accurately approximate the actual JPYZAR exchange rates. However, the MAPE and SMAPE metrics suggest that the model faced moderate challenges in capturing percentage-based fluctuations. Furthermore, the MDA of 47.1% was the lowest among all models, signalling difficulties in effectively forecasting directional trends. This could be attributed to the inherent volatility and mixed long-term trend patterns inherent in the JPYZAR currency pair.

The GBPZAR model incorporated a three-layer architecture, with dropout rates of 0.4, 0.2, and 0.3, and L2 penalties of 0.000571, 0.000331, and 0.000611. These hyperparameter configurations allowed the model to capture the intricate patterns in GBPZAR's exchange rate movements while also preventing overfitting. The model achieved the lowest RMSE and MAE among all currency pairs, indicating its superior predictive accuracy. The MAPE and SMAPE metrics further reinforce the model's ability to effectively capture percentage-based deviations. However, the MDA of 50.2% highlights the model's weakness in predicting directional trends, which is a limitation shared by other LSTM models in this study.

The CNYZAR model demonstrated excellent numerical accuracy, with the lowest RMSE and MAE among the evaluated models. This suggests the model can effectively approximate exchange rate values. However, the MAPE and SMAPE metrics indicate the model struggles to capture percentage-based fluctuations, due to the presence of outliers and extreme values in the CNYZAR exchange rate movements. Additionally, the MDA of 50.7% suggests the model's ability to forecast directional trends requires improvement.

Table 5: LSTM Results

Currency	*RMSE*	*MAE*	*MAPE*	*SMAPE*	*MDA* (%)	*Best Hyperparameters*
EURZAR	0.0434	0.0355	1.4373	1.4261	50.28	*units*_1: 100, *l2*_1: 0.000997, *dropout*_1: 0.4} *units*_2: 150, *l2*_2: 0.000169, *dropout*_2: 0.4, *units*_3: 200, l2_3: 0.000557, *dropout*_3: 0.2, *learning_rate*: 0.000272}
USDZAR	0.055	0.0465	2.0453	2.0176	50.35	*units*_1: 100, *l2*_1: 0.001189, *dropout*_1: 0.2, *units*_2: 200, l2_2: 0.000330, dropout_2: 0.1, *units*_3: 150, l2_3: 0.000336, dropout_3: 0.1, *learning_rate*: 0.000389}
JPYZAR	0.059	0.0484	2.0952	2.0761	47.06	*units*_1: 150, *l2*_1: 0.000546, *dropout*_1: 0.1, $units_2$: 150, $l2_2$: 0.000135, $dropout_2$: 0.5, units_3: 150, *l2*_3: 0.000104, *dropout*_3: 0.4, *learning_rate*: 0.000494}
GBPZAR	0.0358	0.0285	1.0471	1.0547	50.17	*units*_1: 150, *l2*_1: 0.000571, dropout_1: 0.4, *units*_2: 200, *l2*_2: 0.000331, *dropout*_2: 0.2, *units*_3: 100, l2_3: 0.000611, dropout_3: 0.3, *learning_rate*: 0.000336}
CNYZAR	0.0255	0.0207	24.3501	13.3565	50.69	*units*_1: 100, $l2_1$: 0.000657, $dropout_1$: 0.2, *units*_2: 100, $l2_2$: 0.004282, $dropout_2$: 0.1, *units*_3: 200, $l2_3$: 0.000289, $dropout_3$: 0.4, *learning_rate*: 0.000635}

The LSTM models demonstrated remarkable proficiency in minimizing absolute errors across the evaluated currency pairs, particularly for GBPZAR and CNYZAR, where the RMSE and MAE values were the lowest. However, these models exhibited limitations in capturing directional accuracy, with MDA values barely exceeding 50%, suggesting a struggle in predicting trends despite their strong performance in numerical forecasting. This observation implies that while LSTM networks effectively capture historical patterns, they may benefit from further enhancements, such as more advanced feature engineering or the incorporation of hybrid modelling approaches, to enhance their directional prediction capabilities. The crucial role of regularization techniques, including dropout and L2 penalties, in ensuring model stability is acknowledged, but additional refinements may be necessary to enhance the trend prediction performance of these models for practical financial applications.

4.2 In-Sample model comparison

Table 6 compares the in-sample results of the ARIMAX, MLP, RF, and LSTM models across the five currency pairs: EURZAR, USDZAR, JPYZAR, GBPZAR, and CNYZAR. The results demonstrate significant variations in performance metrics. Each model exhibits strengths and weaknesses depending on the exchange rate's characteristics and complexity.

The RF model emerges as the most robust forecasting technique, exhibiting the lowest Root Mean Squared Error and MAE across the studied currency pairs. It also demonstrates high Mean Directional Accuracy, exceeding 90% for several currency pairs. In contrast, while the ARIMAX-GARCH model excels in trend forecasting, with strong MDA for the USDZAR and CNYZAR pairs, it struggles with elevated error metrics, particularly in volatile market environments, limiting its precision in quantitative forecasting. This suggests that while the ARIMAX framework is valuable for modelling structured exchange rate trends, it would benefit from hybridization with non-linear models to enhance its overall forecasting accuracy. The MLP model exhibits suboptimal performance across all currency pairs, as evidenced by its high RMSE and MAE values, as well as the lowest Mean Directional Accuracy, suggesting its limited ability to capture exchange rate trends. This indicates issues such as overfitting and a constrained capacity to model time dependencies. Similarly, the LSTM, while adept at capturing long-term dependencies, struggles to address short-term market volatility, as demonstrated by its high sMAPE for the CNYZAR pair. Furthermore, its MDA hovers around

50%, implying poor directional forecasting capabilities. The in-sample results suggest that the RF model has learned salient market patterns, but its generalizability to out-of-sample data remains to be determined. In contrast, the ARIMAX model may maintain its directional reliability, yet it is likely to continue facing challenges in terms of precision. To improve real-world forecasting performance, further optimization or hybridization will be required for both the MLP and LSTM models.

Table 6 : In-Sample model comparison by currency

Currency	*Metric*	*ARIMAX*	*MLP*	*RF*	*LSTM*
EURZAR	*RMSE*	0.0805	0.0135	0.0037	0.0434
	MAE	0.0218	0.01	0.0025	0.0355
	MAPE	0.1951	0.4153	0.1063	1.4373
	sMAPE	0.2081	0.4155	0.1063	1.4261
	MDA	87.6214	48.38	91.42	50.28
USDZAR	*RMSE*	0.0749	0.0175	0.003	0.055
	MAE	0.0084	0.0137	0.0017	0.0465
	MAPE	0.0821	0.5837	0.0747	2.0453
	sMAPE	0.0952	0.5853	0.0747	2.0176
	MDA	96.862	51.690	94.90	50.35
JPYZAR	*RMSE*	0.006	0.0209	0.0057	0.059
	MAE	0.0003	0.019	0.0035	0.0484
	MAPE	0.3504	0.8273	0.1495	2.0761
	sMAPE	0.2466	0.823	0.1495	1.506
	MDA	85.2982	74.9900	92.27	47.06
GBPZAR	*RMSE*	0.1321	0.0109	0.006	0.0358
	MAE	0.0308	0.0086	0.0043	0.0285
	MAPE	0.2080	0.3209	0.1588	1.0471
	sMAPE	0.2227	0.3209	0.1588	1.0547
	MDA	87.7080	74.490	88.83	50.17
CNYZAR	*RMSE*	0.0024	0.0147	0.0038	0.0255
	MAE	0.0011	0.0119	0.002	0.0207
	MAPE	0.0698	10.1949	1.5038	24.2565
	sMAPE	0.0700	7.156	1.3768	13.3565
	MDA	95.9085	76.7400	94.35	50.69

5. Out of sample results

Extending the in-sample analysis, this section assesses each model's capacity to produce precise forecasts using previously unobserved data. The out-of-sample evaluation is critical for validating the practical applicability of the forecasting models, as it replicates their ability to generalise beyond the data on which they were trained. Specifically, this evaluation estimates the one-day-ahead exchange rate forecasts across the reserved 10% test set, encompassing a volatile period characterized by post-COVID economic adjustments and geopolitical changes. While the in-sample results emphasize the models' fit to historical data, the OOS analysis determines their forecasting efficacy in future scenarios, where concerns about overfitting and model robustness are paramount. Particular attention is directed toward models such as MLP and LSTM, which exhibited inconsistent performance during training.

This section ultimately identifies the models that sustain accuracy and directional consistency under live forecasting conditions, thereby providing the most dependable tools for exchange rate prediction. Among the currency pairs examined, the USDZAR pair exhibited the highest forecasting accuracy. The ARIMAX model achieved the highest Mean Directional Accuracy of 92.2%, effectively capturing the underlying exchange rate trends. Conversely, the RF model demonstrated the most balanced performance, maintaining low RMSE and MAE alongside a robust Mean Directional Accuracy of 78.1%. This robust performance is attributable to the economic stability of the US dollar and the well-defined exchange rate trends against the South African rand, enabling both trend-following and feature-based models to generalize effectively.

In contrast JPYZAR pair was the most challenging, with higher volatility impacting model performance. The ARIMAX model had a high MDA but elevated percentage errors, while the RF and MLP models had higher RMSE and MAE. The LSTM struggled to capture short-term volatility. Highly volatile pairs may require advanced hybrid models or additional macroeconomic indicators to improve accuracy. This suggests that forecasting JPYZAR may require the development of more advanced hybrid models or the incorporation of additional macroeconomic indicators.

Table 6 : Out of Sample model comparison

Currency	*Metric*	*ARIMAX*	*MLP*	*RF*	*LSTM*
EURZAR	*RMSE*	1.0656	0.0097	0.1307	0.0562
	MAE	0.9502	0.0076	0.1096	0.0497
	MAPE	4.8313	0.2581	3.6689	1.7033
	sMAPE	4.9749	0.2582	3.7662	1.6982
	MDA	90.48	47.81	75.66	51.35
USDZAR	*RMSE*	1.8986	0.0294	0.1272	0.1338
	MAE	1.7515	0.0276	0.1168	0.1283
	MAPE	9.6054	0.9564	4.0218	4.4466
	sMAPE	10.1706	0.9617	4.1195	4.5568
	MDA	92.2	47.42	78.09	48.49
JPYZAR	*RMSE*	0.0201	0.0302	0.1869	0.0517
	MAE	0.0186	0.0291	0.1688	0.0416
	MAPE	14.7901	1.4053	8.1404	2.0101
	sMAPE	13.6047	1.3947	8.5673	2.0055
	MDA	83.15	72.1	76.79	47.7
GBPZAR	*RMSE*	0.591	0.0763	0.083	0.0382
	MAE	0.4569	0.0712	0.0682	0.0305
	MAPE	2.1674	2.2856	2.1732	0.9975
	sMAPE	2.1288	2.2563	2.2065	0.9926
	MDA	89.86	75.35	71.92	48.81
CNYZAR	*RMSE*	0.2301	0.0357	0.0428	0.0366
	MAE	0.2104	0.0325	0.0336	0.0295
	MAPE	8.2043	3.4594	3.6335	3.195
	sMAPE	7.824	3.389	3.7474	3.2466
	MDA	85.34	72.33	79.26	46.75

5.1 Results discussion

The out-of-sample evaluation offers crucial insights into the models' ability to generalize their predictions to previously unseen data, providing a realistic assessment of their real-world forecasting performance. While the in-sample analysis underscored each model's capacity to capture historical exchange rate trends, the primary focus here is determining whether these patterns persist when applied to new, unfamiliar data. The assessment considers both minimizing errors and accurately predicting directional changes, shedding light on which models maintain their robustness across diverse market conditions.

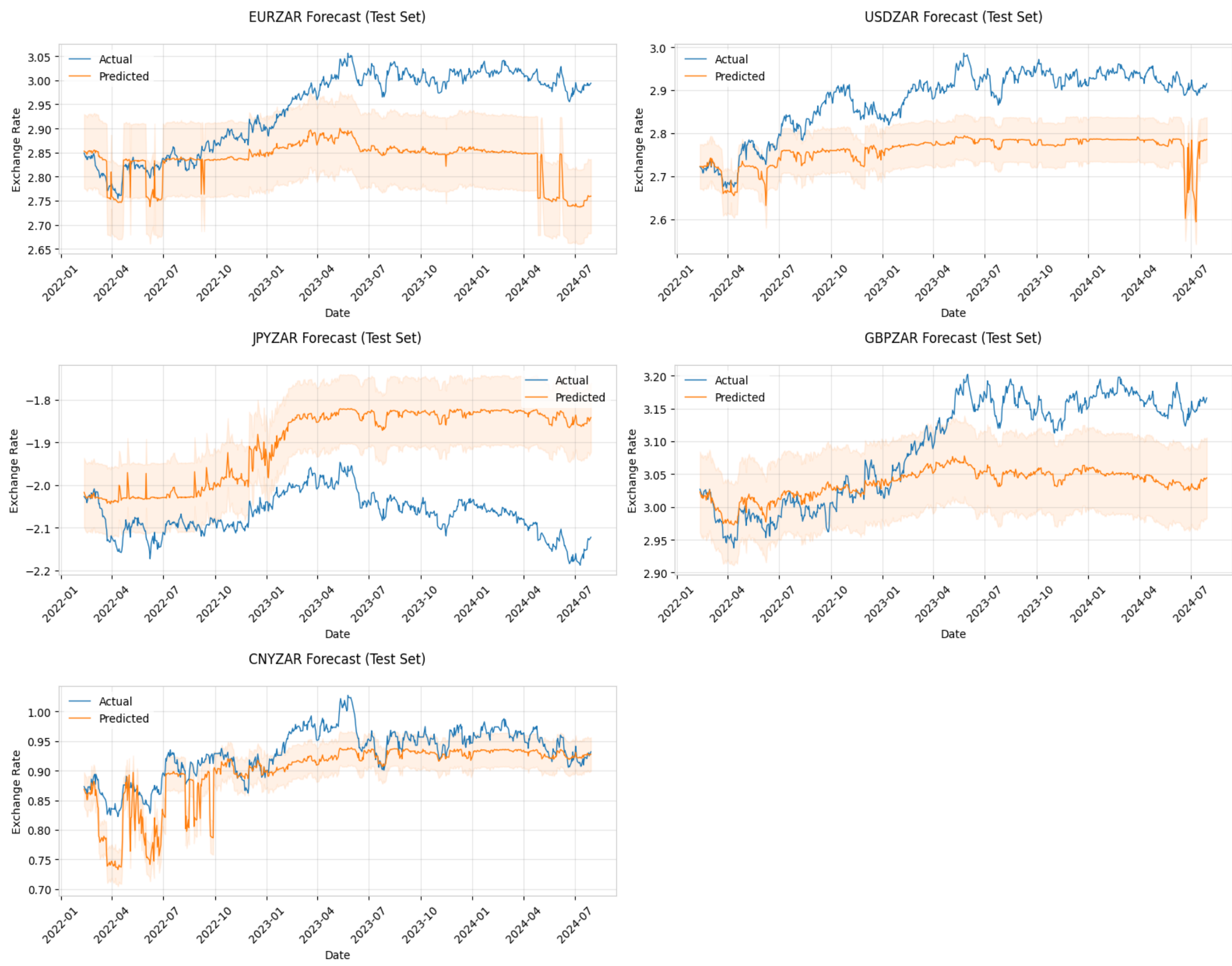

Figure 7 Random Forest Model Out of Sample results

Among the evaluated models, the RF model emerges as a well-balanced approach, exhibiting both strong predictive accuracy and robust directional forecasting capabilities. Its ability to incorporate macroeconomic factors and capture non-linear dependencies renders it particularly effective for the USDZAR and CNYZAR currency pairs, where it maintains low error metrics and high Mean Directional Accuracy aligning with the findings of Martin (2019), who demonstrated the effectiveness of RF in handling macroeconomic fluctuations when forecasting GDP growth. However, the RF model's occasional lag in responding to extreme market shifts suggests that it may benefit from hybridization with sequential learning models to enhance its adaptability to rapidly changing market conditions as also recommended by Makridakis et al. (2022).

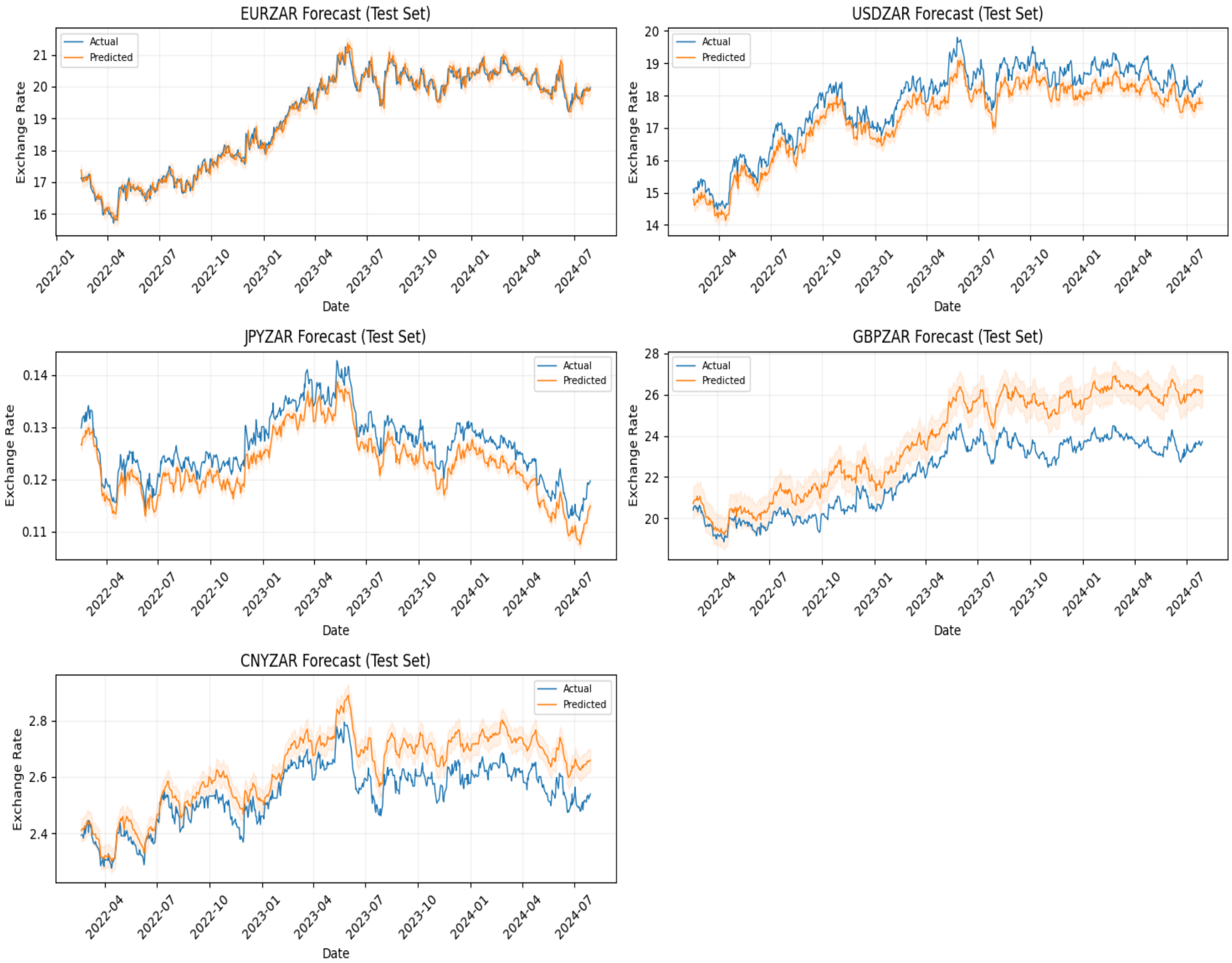

Figure 8 MLP results Out of Sample results

Conversely, the MLP model excels in minimizing absolute forecasting errors, particularly for stable exchange rates such as EURZAR and JPYZAR, this finding is consistent with Kamruzzaman and Sarker (2003), who demonstrated that ANN-based models outperform econometric models in stable currency environments, particularly due to their ability to capture complex relationships in historical data. However, the MLP model demonstrated subpar performance in predicting the directional movements of exchange rates. This limitation aligns with the findings Makridakis et al. (2022), who observed that machine learning models, including neural networks, often excessively smooth out fluctuations, leading to poor performance in capturing sharp trend reversals in volatile financial markets.

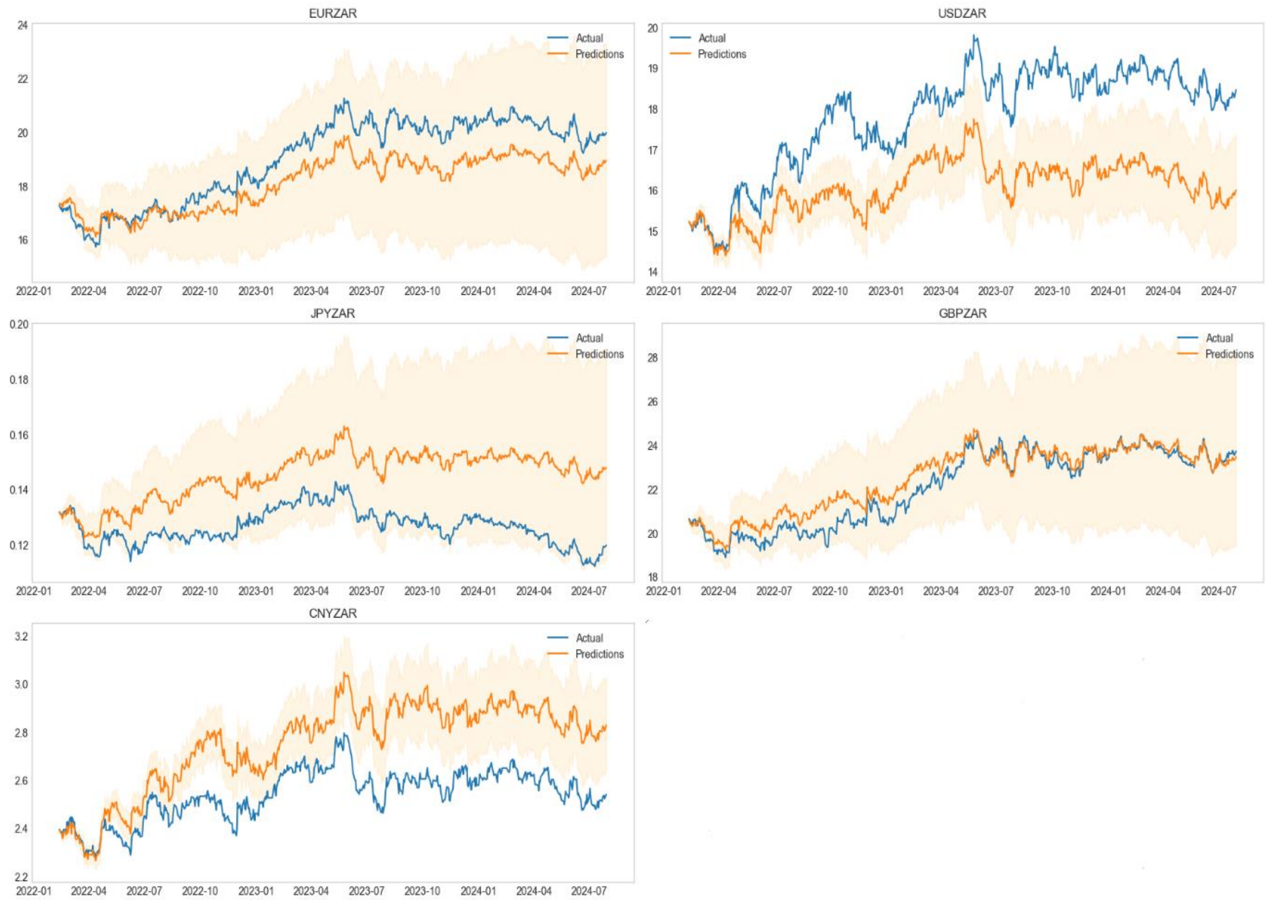

Figure 9 the ARIMAX-GARCH out of Sample result

The ARIMAX-GARCH model demonstrates exceptional directional accuracy, particularly for EURZAR and USDZAR, positioning it as a valuable tool for trend forecasting (Amat et al., 2018). However, its reliance on linear relationships constrains its adaptability to non-linear market dynamics, resulting in higher RMSE and MAPE, especially for more volatile currency pairs like JPYZAR. These findings align with the conclusions of Chen et al. (2008), which suggest that traditional econometric models perform well in capturing structured trends but struggle with non-linearity and abrupt fluctuations. This implies that while Olsen et al. (2024) highlight the competitive performance of traditional models like AR-GARCH in volatility forecasting, future research should further explore hybridization strategies to enhance the adaptability of ARIMAX-GARCH to highly volatile financial environments.

Conversely, the LSTM model exhibits strong performance in capturing long-term dependencies, making it a suitable choice for currency pairs with stable trends, such as GBPZAR. This is consistent with Escudero et al. (2021), who found that LSTM models outperform ARIMA in capturing sequential dependencies in foreign exchange rate forecasting. Nevertheless, its low Mean Directional Accuracy scores for volatile currencies like JPYZAR and CNYZAR highlight its difficulty in adapting to sudden market shifts. These findings align with those of Olsen et al. (2024), who found that while LSTM performs well in stable conditions, traditional econometric models, including AR-GARCH, remain competitive in highly volatile markets. This suggests that while LSTM models are effective in modelling long-term trends, they also require enhanced feature selection or hybridization with econometric models to improve responsiveness to abrupt market fluctuations.

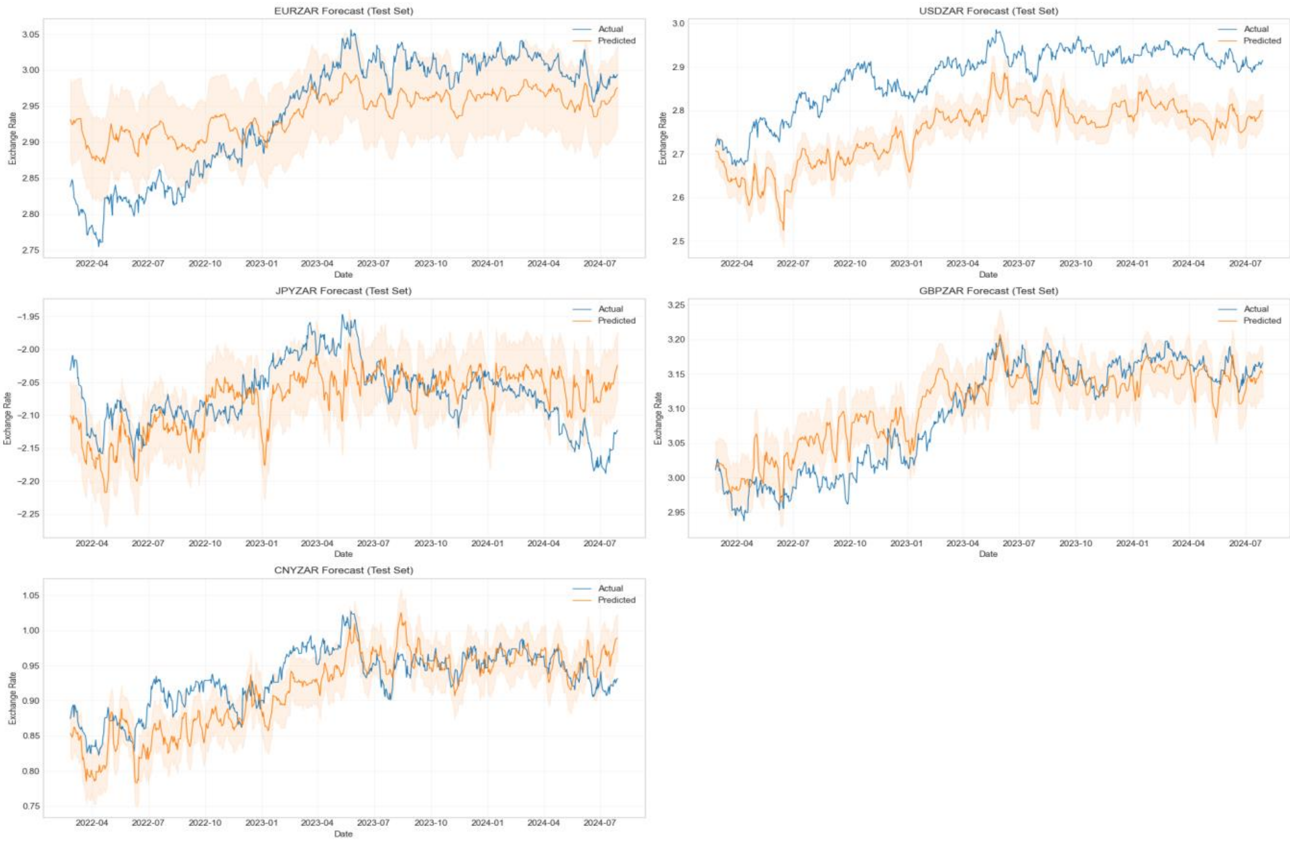

Figure 10 LSTM out of sample result

The findings show variations in model performance between in-sample and out-of-sample tests. The RF model maintains strong accuracy and adaptability but lags in extreme market shifts. The ARIMAX-GARCH continues to excel in directional accuracy but struggles with errors in volatile pairs. MLP and LSTM models face challenges out-of-sample, underscoring the need for model selection aligned with forecasting goals and potential hybrid approaches integrating trend-following and non-linear adaptability.

6. Conclusion

This study provides a comparative assessment of traditional hybrid econometric models and machine learning approaches in forecasting the exchange rate dynamics of the ZAR against five major global currencies: USD, EUR, GBP, JPY, and CNY. The research systematically evaluates 1-day ahead forecasting capabilities of ARIMAX-GARCH, RF, MLP, and LSTM models using a dataset spanning from January 2000 to July 2024, employing rigorous validation techniques, including Train-test split time-series cross-validation, Random Search cross validation and multiple error metrics including RMSE, MAE, MAPE, sMAPE as well as MDA.

The study finds that no single forecasting model consistently outperforms the others across all exchange rate pairs examined. In both in-sample and out of sample evaluations, the ARIMAX-GARCH model demonstrates superior directional accuracy, particularly for structured currency pairs like EURZAR and USDZAR, in alignment with prior work by Amat et al. (2018) and Chen et al. (2008). However, its inability to effectively handle non-linearity and high volatility reduces its effectiveness in forecasting highly dynamic currencies such as JPYZAR and CNYZAR. Conversely, the RF model emerges as the most balanced approach, providing both predictive accuracy and directional forecasting. Its capacity to capture non-linear dependencies and adapt to market volatility is consistent with the observations of Martin (2019), who found RF models effective in macroeconomic forecasting.

The performance of machine learning models varies significantly between in-sample and out-of-sample evaluations. The MLP model performs well in minimizing absolute errors for stable currencies like EURZAR and JPYZAR in-sample, aligning with Kamruzzaman and Sarker (2003), who found ANN-based models effective in low-volatility environments. However, out-of-sample results reveal a sharp increase in RMSE, MAE, and MAPE, particularly for JPYZAR and CNYZAR, indicating poor generalization and an inability to capture directional trends effectively. Similarly, the LSTM model shows strong in-sample performance for long-term dependencies, particularly in GBPZAR forecasting. However, its out-of-sample accuracy declines, with higher RMSE and MAE, reinforcing findings by Escudero et al. (2021) and Olsen et al. (2024) that LSTMs struggle with short-term volatility and rapid market shifts. The underperformance in directional accuracy of the ML models can be attributed to their lack of temporal memory, which limits their ability to adjust to abrupt regime changes or persistent shocks typical of volatile financial markets. Consequently, their directional accuracy diminishes when data diverges from previously learned behaviour (Colombo and Pelagatti, 2020). This study highlights key implications for exchange rate forecasting, emphasizing the need for tailored models to balance error minimization and trend prediction. Given South Africa's volatile currency environment, no single model consistently outperforms across all exchange rates. Hybrid approaches, such as combining RF for feature selection with LSTM for sequential learning, improve accuracy by integrating macroeconomic insights and time-series dependencies. External factors like central bank policies and geopolitical risks further enhance model robustness. While MLP effectively forecasts stable pairs like EURZAR, more volatile currencies like JPYZAR require RF or LSTM for better adaptability. Caution is advised when using MLP in high-frequency trading due to its directional limitations. Future research should explore deep learning models, including transformers, and incorporate sentiment analysis. Expanding this study to other emerging markets will validate the effectiveness of hybrid forecasting models in diverse economic contexts.

Annexure

1.1 Hypermeter turning for ML models

Table 7 : Hypermeters used for to tune the MLP model

$Hyperparameter$	$Values$
$Hidden\ Layer\ Sizes$	$(100,),(150,),(100,50),(200,100,50)$
$Activation\ Function$	$ReLU, Tanh$
$Alpha\ (L2\ Regularization)$	$0.0001, 0.001, 0.01, 0.1$
$Learning\ Rate$	$Constant, Adaptive$
$Initial\ Learning\ Rate$	$0.0001, 0.001, 0.01$
$Max\ Iterations$	2000

Table 8 : Hypermeters used to tune the Random Forest model

$Hyperparameter$	$Values$
$n\_estimators$	$[200, 300, 500, 1000]$
max_depth	$[15, 20, 30, None]$
$min\_samples\_split$	$[2, 5, 10]$
$min\_samples\_leaf$	$[1, 2, 4]$
$max\_features$	$['sqrt', 'log2', None]$
$bootstrap$	$[True, False]$
$max\_samples$	$[0.7, 0.8, 0.9, None]$

Table 9 : Hypermeters used to tune the LSTM model

$Hyperparameter$	$Values$
$units\_1$	$[50, 100, 150, 200]$
$l2\_1$	$['1e-4\ to\ 1e-2\ (log\ scale)']$
$dropout\_1$	$[0.1, 0.2, 0.3, 0.4, 0.5]$
$units\_2$	$[50, 100, 150, 200]$
$l2\_2$	$['1e-4\ to\ 1e-2\ (log\ scale)']$
$dropout\_2$	$[0.1, 0.2, 0.3, 0.4, 0.5]$
$units\_3$	$[50, 100, 150, 200]$
$l2\_3$	$['1e-4\ to\ 1e-2\ (log\ scale)']$
$dropout\_3$	$[0.1, 0.2, 0.3, 0.4, 0.5]$
$learning\_rate$	$['1e-4\ to\ 1e-2\ (log\ scale)']$

1.2 Random Forest feature importance results

References

Afuecheta, E., E. Okorie, I., Nadarajah, S., and Ejiaka Nzeribe, G. (2022). Forecasting Value at Risk and Expected Shortfall of Foreign Exchange Rate Volatility of Major African Currencies via GARCH and Dynamic Conditional Correlation Analysis. In Computational economics. Springer Science+Business Media. https://doi.org/10.1007/s10614-022-10340-9

Al-Gounmeein, R. S., and Ismail, M. T. (2020). Forecasting the exchange rate of the Jordanian Dinar versus the US dollar using a Box-Jenkins seasonal ARIMA model. *International Journal of Mathematics and Computer Science*, *15*(1), 27-40.

Alamneh, T., and Mancha, M. (2021, September 1). Impact of Exchange Rate Devaluation in Ethiopia and Forecasting Foreign Exchange Rate Using ARIMA Model. European Journal of Business and Management, 13(17), 13-23 https://doi.org/https://doi.org/10.7176/ejbm/13-17-02

Alexandridis, A. K., Panopoulou, E., and Souropanis, I. (2024). Forecasting exchange rates: An iterated combination constrained predictor approach. *Journal of Forecasting*, *43*(4), 983-1017. https://doi.org/10.1002/for.3067

Amat, C., Michalski, T., and Stoltz, G. (2018, November 1). Fundamentals and exchange rate forecastability with simple machine learning methods. 88, 1-24. https://doi.org/https://doi.org/10.1016/j.jimonfin.2018.06.003

Ayitey, M., Appiahene, P., Appiah, O., and Bombie, C N. (2023, January 28). Forex market forecasting using machine learning: Systematic Literature Review and meta-analysis. https://doi.org/10.1186/s40537-022-00676-2

Babii, A., Ball, R T., Ghysels, É., and Striaukas, J. (2023, December 1). Machine learning panel data regressions with heavy-tailed dependent data: Theory and application. Elsevier BV, 237(2), 105315-105315. https://doi.org/https://doi.org/10.1016/j.jeconom.2022.07.001

Bank for International Settlements (2022). Triennial Central Bank Survey – OTC foreign exchange turnover in April 2022, Basel: Bank for International Settlements, 2022. Available from https://www.bis.org/statistics/rpfx22.htm.

Baumeister, C., Guérin, P., and Kilian, L. (2015). Do high-frequency financial data help forecast oil prices? The MIDAS touch at work. International Journal of Forecasting, 31(2), 238–252. https://doi.org/10.1016/j.ijforecast.2014.06.005

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. Journal of econometrics, 31(3), 307-327. https://doi.org/10.1016/0304-4076(86)90063-1

Botha, B., Burger, R., Kotzé, K., Rankin, N., and Steenkamp, D. (2022). Big data forecasting of South African inflation. In Empirical Economics (Vol. 65, Issue 1, p. 149). Springer Science+Business Media. https://doi.org/10.1007/s00181-022-02329-y

Box, G. E., and Tiao, G. C. (1975). Intervention analysis with applications to economic and environmental problems. Journal of the American Statistical association, 70(349), 70-79. https://doi.org/10.1080/01621459.1975.10480264

Chasipanta, G R V., and Sánchez-Pozo, N N. (2024, February 27). Long-Term Forecasting of Euro-Dollar Exchange Rates Using the ARIMA Model and Multilayer Perceptron. Lavoisier publishing, 29(1), 125-139. https://doi.org/https://doi.org/10.18280/isi.290114

Chen, A., Hsu, Y., and Hu, H. (2008, January 1). A hybrid forecasting model for foreign exchange rate based on a multi-neural network. IEEE, 5, 293-298. https://doi.org/10.1109/ICNC.2008.298

Colombo, E., and Pelagatti, M. (2020, October 1). Statistical learning and exchange rate forecasting. Elsevier BV, 36(4), 1260-1289. https://doi.org/https://doi.org/10.1016/j.ijforecast.2019.12.007

Corrêa, J. M., Neto, A. C., Júnior, L. T., Franco, E. M. C., and Faria Jr, A. E. (2016). Time series forecasting with the WARIMAX-GARCH method. Neurocomputing, 216, 805-815. https://doi.org/10.1016/j.neucom.2016.08.046

Ding, J., Tarokh, V., and Yang, Y. (2018). Model selection techniques: An overview. *IEEE Signal Processing Magazine*, *35*(6), 16-34. https://doi.org/10.1109/MSP.2018.2867638

Djemo, C. R. T., Eita, J. H., and Mwamba, J. W. M. (2021). Predicting foreign exchange rate movements: an application of the ensemble method. *Review of Development Finance*, *11*(2), 58-70. https://hdl.handle.net/10520/ejc-rdfin_v11_n2_a4

Eichengreen, B and Gupta, P. (2014). Tapering talk: The impact of expectations of reduced Federal Reserve security purchases on emerging markets. The World Bank, Policy Research Working Paper Series 6754.

Eniayewu, P E., Samuel, G T., Dandaura, J J., Samuel, B T., Dogo, B S., Yusuf, U., Ihekuna, R O., and Mevweroso, C R. (2024, March 1). Forecasting Exchange Rate Volatility with Monetary Fundamentals: A GARCH-MIDAS Approach. *Scientific African*, *23*, e02101. https://doi.org/10.1016/j.sciaf.2024.e02101

Escudero, P., Alcocer, W., and Paredes, J. (2021, June 18). Recurrent Neural Networks and ARIMA Models for Euro/Dollar Exchange Rate Forecasting. Multidisciplinary Digital Publishing Institute, 11(12), 5658-5658. https://doi.org/https://doi.org/10.3390/app11125658

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica: Journal of the econometric society*, 987-1007. https://doi.org/10.2307/1912773

Holdgraf. C. (n.d.). *Time-delayed features and auto-regressive models* DataCamp. https://projector-video-pdf-converter.datacamp.com/6079/chapter4.pdf

Hà, V T. (2023, June 14). Experimental Study on Remaining Useful Life Prediction of Lithium-Ion Batteries Based on Three Regressions Models for Electric Vehicle Applications. https://doi.org/10.20944/preprints202306.0999.v1

Hewamalage, H., Ackermann, K., and Bergmeir, C. (2023). Forecast evaluation for data scientists: common pitfalls and best practices. *Data Mining and Knowledge Discovery*, *37*(2), 788-832. https://doi.org/10.1007/s10618-022-00894-5

Hochreiter, S., and Schmidhuber, J. (1997). LSTM. Neural computation, 9(8), 1735-1780. https://doi.org/10.1162/neco.1997.9.8.1735

Inyang, E J., Etuk, E H., Nafo, N M., and Da-Wariboko, Y A. (2023, November 8). Time Series Intervention Modelling Based on ESM and ARIMA Models: Daily Pakistan Rupee/Nigerian Naira Exchange Rate., 25(3), 1-17. https://doi.org/https://doi.org/10.9734/ajpas/2023/v25i3560

Jung, G., and Choi, S. (2021, March 31). Forecasting Foreign Exchange Volatility Using Deep Learning Autoencoder-LSTM Techniques. Hindawi Publishing Corporation, 2021, 1-16. https://doi.org/https://doi.org/10.1155/2021/6647534

Kabundi, A. N. (2002). Macroeconomic forecasting: a comparison between artificial neural networks and econometric models. University of Johannesburg (South Africa).

Kamruzzaman, J., and Sarker, R. (2003, January 1). Forecasting of currency exchange rates using ANN: A case study. IEEE, 1, 793-797. https://doi.org/10.1109/ICNNSP.2003.1279395

Lim, B., and Zohren, S. (2021). Time-series forecasting with deep learning: a survey. *Philosophical Transactions of the Royal Society A*, *379*(2194), 20200209.

https://doi.org/10.1098/rsta.2020.0209

Liu, P., Wang, Z., Liu, D., Wang, J., and Wang, T. (2023, June 1). A CNN-STLSTM-AM model for forecasting USD/RMB exchange rate. Elsevier BV, 11(2), 100079-100079. https://doi.org/https://doi.org/10.1016/j.jer.2023.100079

Liu, S., Huang, Q., Li, M., and Wei, Y. (2024, January 1). A new LASSO-BiLSTM-based ensemble learning approach for exchange rate forecasting. Elsevier BV, 127, 107305-107305. https://doi.org/https://doi.org/10.1016/j.engappai.2023.107305

Lütkepohl, H. (2005). New introduction to multiple time series analysis. Springer Science & Business Media.

Makridakis, S., Spiliotis, E., and Assimakopoulos, V. (2018, March 27). Statistical and Machine Learning forecasting methods: Concerns and ways forward. Public Library of Science, 13(3), e0194889-e0194889. https://doi.org/10.1371/journal.pone.0194889

Martin, L. (2019, January 1). Machine Learning vs Traditional Forecasting Methods: An Application to South African GDP. Federal Reserve Bank of St. Louis. https://ideas.repec.org/p/sza/wpaper/wpapers326.html

May, C. (2015). Empirical analysis of the dynamics of the South African Rand (Post-1994). Unpublished PhD thesis, University of the Witwatersrand. Available at:

https://wiredspace.wits.ac.za/bitstream/10539/20503/1/PHDMAINFinal30October2015.pdf

Meese, R. A., and Rogoff, K. (1983). Empirical exchange rate models of the seventies: Do they fit out of sample? *Journal of International Economics*, *14*(1-2), 3-24. https://doi.org/10.1016/0022-1996(83)90017-X

Mlambo, C., Maredza, A., and Sibanda, K. (2013, November 1). Effects of Exchange Rate Volatility on the Stock Market: A Case Study of South Africa. Mediterranean Center of Social and Educational Research. https://doi.org/https://doi.org/10.5901/mjss.2013.v4n14p561

Nanthakumaran, P., and Tilakaratne, C. (2017, January 1). A comparison of accuracy of forecasting models: A study on selected foreign exchange rates. IEEE, 1-8. https://doi.org/10.1109/ICTER.2017.8257789

Olsen, A., Djupskås, G., Lange, P E D., and Risstad, M. (2024, March 25). Forecasting implied volatilities of currency options with machine learning techniques and econometrics models. Springer International Publishing. https://doi.org/https://doi.org/10.1007/s41060-024-00528-7

Olah, C., 2015. Understanding LSTM networks. https://colah.github.io/posts/2015-08-Understanding-LSTMs/

Pahlavani, M., and Roshan, R. (2015, December 1). The Comparison among ARIMA and hybrid ARIMA-GARCH Models in Forecasting the Exchange Rate of Iran. Premier Publishing, 7(1), 31-50. https://doi.org/https://doi.org/10.22111/ijbds.2015.2198

Pankratz, A. (2012). Forecasting with dynamic regression models. John Wiley & Sons.

Qureshi, M., Ahmad, N., Ullah, S., and Mustafa, A R U. (2023, May 1). Forecasting real exchange rate (REER) using artificial intelligence and time series models. Elsevier BV, 9(5), e16335-e16335. https://doi.org/https://doi.org/10.1016/j.heliyon.2023.e16335

Rahman, M M., Chi, G., Gupta, A D., Hossain, M., and Abedin, M Z. (2021, December 15). Impact of early COVID-19 pandemic on the US and European stock markets and volatility forecasting. Taylor & Francis, 35(1), 3591-3608. https://doi.org/10.1080/1331677x.2021.1997626

Rosenblatt, F F. (1958, January 1). The perceptron: A probabilistic model for information storage and organization in the brain. American Psychological Association, 65(6), 386-408. https://doi.org/10.1037/h0042519

Rossi, B. (2013). Exchange rate predictability. Journal of Economic Literature, 51(4), https://doi.org/10.1257/jel.51.4.1063

Saranj, A., and Zolfaghari, M. (2022). The electricity consumption forecast: Adopting a hybrid approach by deep learning and ARIMAX-GARCH models. *Energy Reports*, *8*, 7657-7679. https://doi.org/10.1016/j.egyr.2022.06.007

Shibata, R. (1976). Selection of the order of an autoregressive model by Akaike's information criterion. *Biometrika*, *63*(1), 117-126. https://doi.org/10.1093/biomet/63.1.117

Tanko, A. N., Musa, G. K., and Auta, M. S. (2023). On the use of arima and garch in modelling nigeria's naira–us dollar monthly exchange rates. *Global Journal of Pure and Applied Sciences*, *29*(1), 73-82. https://doi.org/https://doi.org/10.4314/gjpas.v29i1.9

Umoru, D., Effiong, S E., Umar, S S., Ugbaka, M A., Iyaji, D., Okpara, E., Iyayi, D., Tizhe, A N., and Omomoh, O H. (2023, January 1). Forecasting exchange rate dynamics in developing countries. https://doi.org/10.22495/cbsrv4i2siart3

Vishwakarma, V K. (2013, January 1). Forecasting Real Estate Business: Empirical Evidence from the Canadian Market. RELX Group (Netherlands). https://papers.ssrn.com/sol3/Delivery.cfm/SSRN_ID2148507_code1332876.pdf?abstractid=2148507&mirid=1

Wang, J., Dong, Y., and Liu, J. (2023, November 1). A novel multifactor clustering integration paradigm based on two-stage feature engineering and improved bidirectional deep neural networks for exchange rate forecasting. Elsevier BV, 143, 104258-104258. https://doi.org/10.1016/j.dsp.2023.104258

Yurek, I. (2023, September 15). *Perceptron model: The foundation of neural networks*. Medium. https://medium.com/@ilyurek/perceptron-model-the-foundation-of-neural-networks-4db25b0148d

Zhang, R. (2024, January 22). RMB/USD Exchange Rate Forecasting and Analysis by SES and ARIMA Model. https://doi.org/10.54097/jryene75

Zhao, Y., Feng, C., Xu, N., Peng, S., and Liu, C. (2023). Early warning of exchange rate risk based on structural shocks in international oil prices using the LSTM neural network model. Energy Economics, 126, 106921. https://doi.org/10.1016/j.eneco.2023.106921