Deep Learning and Statistical Approaches for Forecasting the Indonesian Rupiah Exchange Rate

Neni Alya Firdausanti[1*], Veniola Forestryani[1], Husna Mir'atin Nuroini[1]
[1]Department of Statistics, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia
*Corresponding author: neni.alya@its.ac.id

Received: 16 March 2025
Revised: 6 May 2025
Accepted: 10 June 2025

ABSTRACT – Accurate forecasting of exchange rates is essential for economic stability, investment strategy, and policy formulation. This study presents a comparative analysis of two distinct modeling approaches for predicting the Indonesian Rupiah (IDR) exchange rate against the US Dollar (USD): the Markov Switching Generalized Autoregressive Conditional Heteroskedasticity (MS-GARCH) model and the Long Short-Term Memory (LSTM) network enhanced with an attention mechanism. The MS-GARCH model captures volatility clustering and regime shifts, while the LSTM-Attention model learns complex nonlinear temporal dependencies. Using historical USD/IDR exchange rate data, both models are evaluated based on Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE). Empirical results show that the LSTM-Attention model achieves higher forecasting accuracy; however, the MS-GARCH model provides superior interpretability and insight into structural volatility. These findings underscore the importance of aligning model choice with forecasting objectives—highlighting that while deep learning offers enhanced predictive capability, statistical models remain valuable for risk analysis and financial diagnostics. The results support a complementary use of both methods in financial forecasting applications.

Keywords – **Exchange Rate, MS-GARCH, LSTM, Attention Mechanism, Deep Learning.**

I. INTRODUCTION

The financial decisions and economic stability of Indonesia depend critically on the US dollar (USD) to Indonesian rupiah (IDR) exchange rate. Influencing the cost of imports, exports, foreign investments, and general economic competitiveness, the exchange rate is essentially the value one currency can be exchanged for another [1]. The stability of the IDR against the USD is often cited as a crucial macroeconomic indicator. This reflects the general resilience and strength of the Indonesian economy despite changes in world economy [2].

Many macroeconomic variables affect fluctuations in the exchange rate, including inflation rates, interest rates, government policies, trade balances, and outside world economic shocks. Furthermore, Indonesia has experienced several financial crises, including the Asian financial crisis in 1997–1998 and the worldwide financial crisis in 2008 [3]. Financial crises usually lead to increased volatility, currency devaluation, lower investor confidence, and economic instability. Therefore, an effective forecasting and risk management strategies is important to mitigate these problems [4].

One important aspect of exchange rate movements is volatility, which describes the size and regularity of changes in currency values. In financial markets, high volatility usually indicates more risk and uncertainty, which makes economic decision-making, investment strategies, and policy execution more difficult. Long-term volatility might hinder monetary policy attempts to stabilize the economy, discourage foreign investment, and increase economic vulnerabilities [5]. For policymakers, investors, and financial institutions, precisely modeling and forecasting exchange rate fluctuations has become crucial due to the substantial consequences for global commerce, investment choices, inflation management, and overall economic planning.

Numerous studies have been conducted to develop effective forecasting methodologies to mitigate risks associated with exchange rate volatility. The traditional statistical models, such as Markov Switching Generalized Autoregressive Conditional Heteroscedasticity (MS-GARCH) are frequently used due to their effectiveness in capturing volatility clustering and identifying distinct volatility regimes [5, 6]. For instance, Nunian et al. [5] compared Markov-switching and MS-GARCH models for currency exchange rate modeling and found the Markov-switching models proficient in capturing nonlinear patterns and regime changes. These models provide interpretable insights into regime-dependent volatility and are useful for identifying structural breaks. However, they often rely on assumptions such as normality and may be limited in capturing complex nonlinear relationships inherent in financial data.

In recent years, advances in deep learning, particularly Long Short-Term Memory (LSTM) networks enhanced with attention mechanisms, have provided promising alternatives. These models can capture intricate temporal patterns without relying heavily on distributional assumptions. Previous research indicates that LSTM models excel in capturing temporal dependencies and complex nonlinear relationships, outperforming traditional statistical approaches in various forecasting tasks [7]. Zyad, Lubis, and Tjandra [8] demonstrated that attention-based LSTM significantly improved forecasting accuracy for the USD/IDR exchange rate, achieving lower Root Mean Squared Error (RMSE) compared to conventional LSTM models. Similarly, Hidayat et al. [9] compared LSTM with Random Forest for currency exchange forecasting, concluding that LSTM consistently outperformed traditional machine learning methods due to its superior capability to model complex and nonlinear time series patterns. Although deep learning models often outperform traditional methods in terms of forecasting accuracy, they tend to lack transparency, making interpretation and diagnostics more challenging for policymakers and financial analysts.

Despite advancements in forecasting methodologies, comprehensive studies explicitly comparing MS-GARCH and attention-based LSTM models specifically for the IDR/USD exchange rate forecasting remain limited, highlighting a notable research gap. Moreover, the comparative analysis of traditional statistical methods and advanced deep learning approaches within the specific context of Indonesian exchange rates need to be explored more.

This study addresses the gap in comparative analysis between these two modeling paradigms, MS-GARCH and LSTM-Attention, by evaluating their performance in forecasting the USD/IDR exchange rate. In addition to reporting predictive accuracy, this research emphasizes interpretability, assumption validity, and practical implications of model outputs. By doing so, it aims to provide guidance on the strengths and trade-offs of each method, ultimately encouraging informed model selection or hybrid integration for exchange rate forecasting tasks. Utilizing historical exchange rate data, the research evaluates each model's forecasting performance based on widely recognized accuracy metrics, including Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE). The comparative analysis examines the merits and limitations of each method in mitigating the underlying complexity and volatility of financial time series.

The primary contribution of this research is to bridge the existing gap by systematically comparing MS-GARCH and attention-based LSTM models for forecasting the IDR/USD exchange rate. By conducting rigorous empirical evaluations using established accuracy metrics—Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE)—this study provides a clear assessment of the relative merits and limitations of traditional statistics method and modern deep learning approaches. Furthermore, this study's findings offer actionable insights for policymakers, investors, and financial institutions to better anticipate exchange rate movements and manage associated risks, thus enhancing decision-making and risk mitigation strategies in Indonesia’s volatile economic landscape.

The remainder of this paper is organized as follows. The literature review section critically examines previous studies and relevant methodologies. Next, the methodology section details the comparative analytical approach adopted in this study. The results and discussion section then presents empirical findings and evaluates the model performances. Lastly, the conclusion summarizes the key findings, discusses implications for practitioners and policymakers, and suggests avenues for future research.

II. LITERATURE REVIEW

Forecasting exchange rates remains a central challenge in financial econometrics due to the inherent volatility, nonlinearity, and sensitivity to macroeconomic shocks. Over the years, both statistical and deep learning approaches have been developed to tackle this problem, each with distinct strengths and limitations.

Traditional time-series models such as the Generalized Autoregressive Conditional Heteroskedasticity (GARCH) family have been widely used to model exchange rate volatility. The GARCH framework, introduced by Bollerslev [10], is effective in capturing time-varying volatility and volatility clustering, which are common in financial series. However, its inability to account for structural breaks and regime switching prompted the development of hybrid models like the Markov Switching GARCH (MS-GARCH), introduced by Hamilton [11] and expanded by Haas et al. [12].. These models incorporate a probabilistic regime-switching mechanism, enabling them to adapt to changing volatility states in the market.

Recent studies have shown that MS-GARCH models perform well in capturing volatility regimes and are particularly effective during financial crises. Klaassen [13] demonstrated that regime-switching GARCH models provide better forecasts during high-volatility periods. Similarly, Nkemnole et al. [6] validated the robustness of MS-GARCH models for exchange rate risk assessment across different estimation techniques.

Recent studies have shown MS-GARCH models performs well in capturing volatility regimes and are particularly effective during financial crises. Klaassen [13] demonstrated that GARCH forecasts for exchange rates are often too high during volatile periods and addressed this issue using a Markov Switching framework. Similarly, Nkemnole and Ebomese [6] explored estimation methods of MS-GARCH models for forecasting exchange rate volatility, highlighting their effectiveness in capturing regime-specific behaviors. Additionally, studies have compared various MS-GARCH specifications in terms of their risk forecasting ability for exchange rates, providing insights into their robustness and applicability.

However, advances in computational power and the availability of large datasets have spurred the adoption of deep learning methods in time series forecasting. Long Short-Term Memory (LSTM) networks, developed by Hochreiter and Schmidhuber [14], are capable of capturing long-term dependencies in sequential data by using gating mechanisms to retain relevant information. More recent work has incorporated attention mechanisms into LSTM architectures to dynamically focus on the most informative time steps. This enhancement improves model performance by allowing it to assign greater weight to important historical patterns, thereby increasing forecasting precision.

More recent work has incorporated attention mechanisms into LSTM architectures to dynamically focus on the most informative time steps. This enhancement improves model performance by allowing it to assign greater weight to important historical patterns, thereby increasing forecasting precision [15]. Several studies have applied LSTM networks with attention mechanisms to exchange rate forecasting. Wu [16] proposed a hybrid LSTM model based on an attention mechanism to predict stock prices, incorporating exchange rate data as an input feature, which resulted in enhanced prediction accuracy. Additionally, Saadati and Manthouri [17] developed an attention-based LSTM model for predicting Forex rates, highlighting the model's effectiveness in capturing complex patterns in currency data. Similarly, Islam and Hossain [18] implemented a GRU-LSTM hybrid network for foreign exchange currency rate prediction, demonstrating improved performance over conventional methods.

Comparative studies between traditional statistical models and deep learning approaches have yielded mixed results. While some research indicates that deep learning methods, including LSTM with attention mechanisms, can provide superior forecasting accuracy, others highlight the robustness of MS-GARCH models in capturing regime changes and volatility clustering. For instance, studies have shown that MS-GARCH models effectively capture regime changes and exhibit strong performance in volatile financial markets. Conversely, attention-based LSTM models have demonstrated improved predictive outcomes for price prediction over traditional methods [19]. However, many of these comparisons are made on major currencies (e.g., USD/EUR, USD/JPY), with fewer studies examining emerging market currencies like the Indonesian Rupiah.

Despite the rich literature on exchange rate forecasting using either MS-GARCH or attention-based LSTM models, there remains a distinct gap in direct empirical comparisons between these two paradigms, particularly in the context of emerging market currencies like the Indonesian Rupiah (IDR). Most comparative studies focus on major currency pairs (e.g., USD/EUR, USD/JPY) and overlook IDR/USD, which exhibits unique macroeconomic sensitivities and volatility patterns. This study contributes to the literature by explicitly evaluating the forecasting accuracy, volatility modeling capabilities, and practical implications of MS-GARCH versus LSTM-Attention, providing a rare side-by-side assessment tailored to Indonesian exchange rate dynamics. To our knowledge, this is one of the first works to offer such a comprehensive comparison, thereby filling a critical gap in both financial econometrics and applied deep learning research in Southeast Asian markets.

III. METHODOLOGY

A. Research Procedure

The research procedure follows a structured approach to ensure an effective comparison between MS-GARCH and LSTM with attention. The process begins with data exploration and preprocessing, including transformations to stabilize variance and ensure stationarity. Following this, the dataset is split into training and testing sets to facilitate robust model evaluation. The ARIMA and LSTM models are then developed using the training data, incorporating hyperparameter tuning and optimization techniques to enhance predictive performance. Subsequently, the models are evaluated and compared using both training and testing datasets, with the selection of the best model based on the lowest Mean Absolute Percentage Error (MAPE) and Root Mean Squared Error (RMSE) values. The chosen model is then utilized for exchange rate forecasting. The detailed research workflow is illustrated in Figure 1.

B. Data Source

This study utilizes secondary data obtained from the International Monetary Fund (IMF) website. The dataset comprises monthly USD/IDR exchange rate observations starting from January 1990 to December 2024, with total 420 data points.

C. Data Splitting

To facilitate model training and evaluation, the dataset is divided into training and testing sets. The training data consists of observations up to September 2023, while the testing data includes the last three months of 2023 and all available data for 2024. The inclusion of the last three months of 2023 in the testing set allows the models to adapt to recent exchange rate movements before making forecasts for 2024, ensuring a more reliable assessment of their predictive capabilities.

D. Markov-Switch GARCH Model

Markov Switching (MS) models are widely used in time series analysis, particularly for capturing structural changes or regime shifts in data. Unlike deterministic models, which assume fixed parameters, MS models consider regime transitions as the outcomes of an unobservable random variable, known as the state. These state-dependent changes allow the model to better capture the underlying dynamics of financial and economic data.

According to Hamilton and Susmel [20], building on Hamilton's [11] work, the Markov Switching model for conditional mean is given by the equation (1):

$$r_t = \mu_{s_t} + \widetilde{r_t} \tag{1}$$

where r_t represents the observed variable, μ_{s_t} denotes the state-dependent mean within the Markov Switching framework, and $\widetilde{r_t}$ follows an AR(p) process with a mean of zero. This formulation allows the mean of the process to vary depending on the regime.

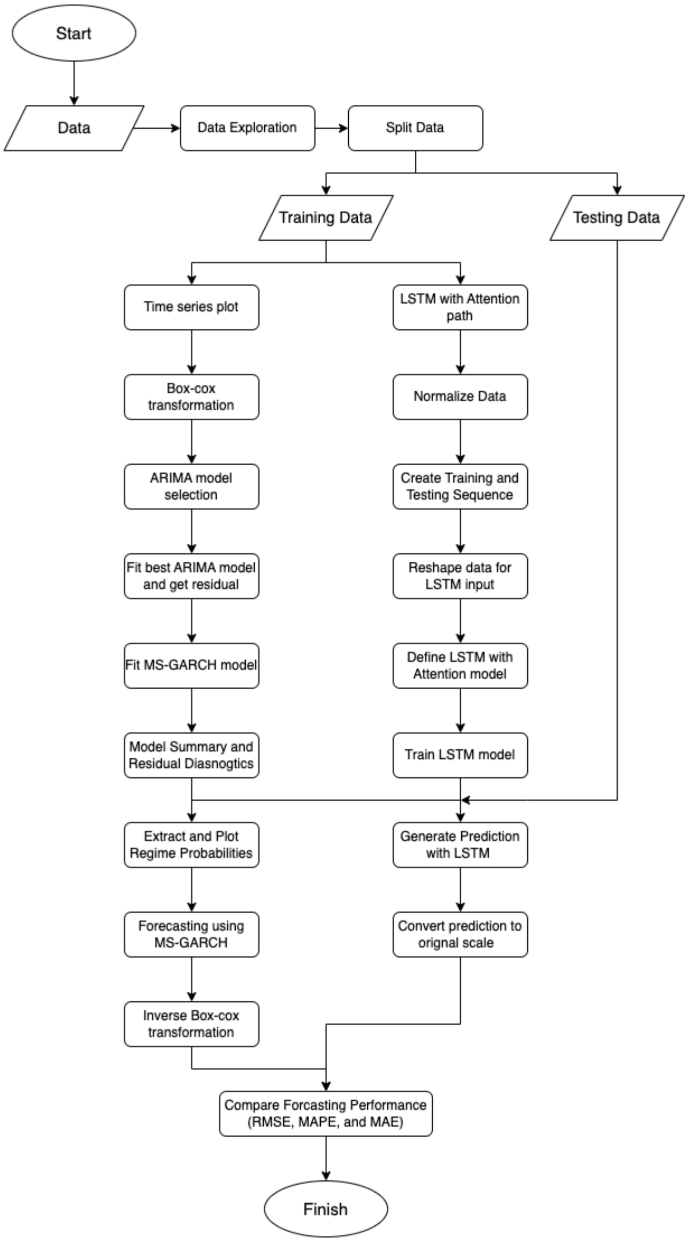

Figure 1 Research Flow Chart

To further incorporate state dynamics into the time series model, Hamilton (1989) introduced the Markov Switching autoregressive process, which is formulated as equation (2):

$$r_t - \mu_{s_t} = \sum_{i=1}^{p} \phi_i \left(r_{t-i} - \mu_{s_{t-i}}\right) + \varepsilon_t \tag{2}$$

In this equation, ε_t represents the residual term, which captures deviations from the model's conditional mean. The conditional mean is modeled as an AR(p) process, where μ_{s_t} depends on the state s_t. The state-dependent mean, μ_{s_t}, takes different values depending on the regime, meaning that when $s_t = 1$, the mean is μ_1; when $s_t = 2$, it is μ_2; and so on, up to μ_p. This flexibility enables the model to account for regime-dependent fluctuations in time series data.

The Markov Switching framework employs a first-order Markov Chain to model transitions between states. If the probability of the system being in a particular state $s_t = j$ depends only on the most recent state s_{t-1}, the transition probability can be written as equation (3):

$$P[s_t = j | s_{t-1} = i] = p_{ij} \tag{3}$$

where p_{ij} represents the probability of transitioning from state i at time $t - 1$ to state j at time t. For a two-state system, the transition probabilities can be represented as a matrix:

$$\boldsymbol{P} = \begin{bmatrix} p_{11} & p_{21} \\ p_{12} & p_{22} \end{bmatrix}$$

where each element p_{ij} represents the probability of transitioning from state i to state j. In an extension to a three-state system, the transition probability matrix expands as follows:

$$\boldsymbol{P} = \begin{bmatrix} p_{11} & p_{21} & p_{31} \\ p_{12} & p_{22} & p_{32} \\ p_{13} & p_{23} & p_{33} \end{bmatrix}$$

These transition matrices allow the model to quantify how likely it is for a process to shift from one regime to another at each time step, providing a probabilistic framework for understanding regime-switching behaviour. While the Markov Switching model effectively captures structural changes in time series, it is limited in explaining volatility dynamics. Specifically, traditional MS models do not account for volatility clustering, leverage effects, or asymmetric responses to shocks, which are common in financial data. Volatility clustering refers to the tendency of high-volatility periods to be followed by other high-volatility periods, and leverage effects describe the phenomenon where negative shocks to an asset price often lead to higher volatility than positive shocks of similar magnitude. To address these issues, the Markov Switching Generalized Autoregressive Conditional Heteroskedasticity (MS-GARCH) model was developed, incorporating both regime-dependent volatility modelling and asymmetric effects of shocks. The MS-GARCH model enhances the traditional Markov Switching framework by integrating heteroskedasticity modelling, allowing for volatility to change dynamically across different states. The Markov regime-switching generalized autoregressive conditional heteroskedasticity model is given by the equation (4):

$$\boldsymbol{\sigma}^2_{t,s_t} = \boldsymbol{\alpha}_{0,s_t} + \sum_{i=1}^{q} \boldsymbol{\alpha}_{i,s_t}\boldsymbol{\varepsilon}^2_{t-i} + \sum_{j=1}^{p} \boldsymbol{\beta}_{j,s_t}\boldsymbol{\sigma}^2_{t-j,s_t} \tag{4}$$

where $\boldsymbol{\sigma}^2_{t,s_t}$ represents the conditional variance at time t in regime s_t, and $\boldsymbol{\alpha}_{0,s_t}$ is the constant parameter that differs across regimes. The term $\boldsymbol{\alpha}_{i,s_t}$ denotes the coefficient of squared residuals at lag i in regime s_t, with q representing the number of lags considered for residual components. The squared residuals, $\boldsymbol{\varepsilon}^2_{t-i}$, capture past error fluctuations and their impact on current volatility. Additionally, $\boldsymbol{\beta}_{j,s_t}$ represents the coefficient of past conditional variance for lag j in regime s_t, while $\boldsymbol{\sigma}^2_{t-j,s_t}$ is the lagged conditional variance at $(t-j)$ within regime s_t. The number of lags for past volatility effects is given by p.

E. LSTM with Attention Model

The Long Short-Term Memory (LSTM) network with an attention mechanism is employed to capture temporal dependencies and enhance predictive accuracy. LSTM is a type of recurrent neural network (RNN) designed to mitigate the vanishing gradient problem by incorporating memory cells and gating mechanisms, enabling it to effectively model long-term dependencies in sequential data **[21]**.

The attention mechanism is integrated into the LSTM architecture to improve model focus on relevant past observations, dynamically assigning higher weights to critical time steps while reducing the influence of less significant points **[22]**. This mechanism enhances forecasting performance by allowing the model to concentrate on important historical exchange rate fluctuations.

LSTM

A distinctive feature of LSTM is the incorporation of gating mechanisms, including the input gate, output gate, and forget gate. At each time step $\boldsymbol{t}$, the input is denoted as $\boldsymbol{x_t}$, while $\boldsymbol{h_{t-1}}$ represents the hidden state from the previous time step, and $\boldsymbol{h_t}$ is the current output. The input gate ($\boldsymbol{i_t}$) regulates the information entering the cell state at time $\boldsymbol{t}$, where $\boldsymbol{W_i}$ is the corresponding weight matrix. The value of $\boldsymbol{i_t}$ is computed by applying a $\boldsymbol{tanh}$ activation function to a weighted sum of $\boldsymbol{h_{t-1}}$ and $\boldsymbol{x_t}$, followed by the addition of a bias term. This transformation determines the extent to which new information is retained in the cell. The specific mathematical formulation is presented in Equation (5).

$$\boldsymbol{i_t} = \boldsymbol{\sigma}(\boldsymbol{W_i} \times [\boldsymbol{h_{t-1}}, \boldsymbol{x_t}] + \boldsymbol{b_i}) \tag{5}$$

The output gate is associated with the weight matrix $\boldsymbol{W_o}$, and its activation at time t, denoted as $\boldsymbol{o_t}$, is determined by applying the tanh activation function to a weighted combination of the input $\boldsymbol{x_t}$ and the previous hidden state $\boldsymbol{h_{t-1}}$, followed by the addition of a bias term. Ultimately, the input gate is updated through the activation function to regulate the information flow. The precise mathematical formulation is provided in Equation (6).

$$\boldsymbol{o_t} = \boldsymbol{\sigma}(\boldsymbol{W_o} \times [\boldsymbol{h_{t-1}}, \boldsymbol{x_t}] + \boldsymbol{b_o}) \tag{6}$$

In the forget gate, the weight matrix is represented as $\boldsymbol{W_f}$, and the gate's activation at time $\boldsymbol{t}$, denoted as $\boldsymbol{f_t}$, is computed using the tanh activation function applied to a weighted sum of the input $\boldsymbol{x_t}$ and the previous hidden state $\boldsymbol{h_{t-1}}$, followed by the addition of a bias term. The final output is then processed through a sigmoid activation function ($\boldsymbol{\sigma}$), ensuring that the resulting value falls within the range of 0 to 1. A higher value indicates a lower probability of forgetting, while a value of 1 means that the input information $\boldsymbol{x_t}$ is fully retained. The corresponding mathematical formulation is presented in Equation (7).

$$\boldsymbol{f_t} = \boldsymbol{\sigma}(\boldsymbol{W_f} \times [\boldsymbol{h_{t-1}}, \boldsymbol{x_t}] + \boldsymbol{b_f}) \tag{7}$$

In the memory unit, $\boldsymbol{C_t}$ represents the state of the memory cell at time t. The forget gate activation $\boldsymbol{f_t}$ is applied to the previous memory state $\boldsymbol{C_{t-1}}$, while the input gate activation $\boldsymbol{i_t}$ is applied to the candidate cell state $\boldsymbol{\widehat{C}_t}$. These two components are then summed to compute the updated memory cell state $\boldsymbol{C_t}$, as defined in Equation (4). The weight matrix for the memory cell is denoted as $\boldsymbol{W_C}$. The candidate cell state $\boldsymbol{\widehat{C}_t}$ is computed by applying the tanh activation function to a weighted sum of the input $\boldsymbol{x_t}$ and the previous hidden state $\boldsymbol{h_{t-1}}$, followed by the addition of a bias term.

After passing through the activation function, the updated cell state $\boldsymbol{C_t}$ is obtained. The detailed mathematical formulation is provided in Equation (8).

$$\boldsymbol{C_t = f_t \times C_{t-1} + i_t \times \hat{C}_t} \tag{8}$$

$$\boldsymbol{\hat{C}_t = \tanh(W_c \times [h_{t-1}, x_t] + b_c)} \tag{9}$$

Finally, the output of the LSTM at time $\boldsymbol{t}$, denoted as $\boldsymbol{C_t}$, is obtained by taking the tanh activation of the updated memory cell state $\boldsymbol{C_t}$ and multiplying it by the output gate activation $\boldsymbol{o_t}$ at the same time step. The precise mathematical expression for this computation is presented in Equation (10).

$$\boldsymbol{h_t = o_t \times \tanh(C_t)} \tag{10}$$

Attention

The attention mechanism is a signal processing approach that originated from research on human vision in the 1990s. It is a specialized structure integrated into machine learning models, primarily designed to automatically learn and assess the relationships between input data pairs and their influence on the output. Incorporating an attention mechanism into a deep learning model enhances its ability to focus on the most relevant information, similar to how the human brain prioritizes important details while disregarding irrelevant ones. This selective focus helps improve prediction accuracy by emphasizing critical features and minimizing the influence of less significant data. The attention mechanism relies on key weight parameters, including $\boldsymbol{e_t}$, $\boldsymbol{t}$, and $\boldsymbol{C_t}$, where $\boldsymbol{e_t}$ represents the weight score assigned to different features at time $\boldsymbol{t}$. The corresponding mathematical formulation is provided in Equation (11).

$$\boldsymbol{e_t = vtanh(W_e h_t + b_e)} \tag{11}$$

In this context, $\boldsymbol{v}$ and $\boldsymbol{W_e}$ represent the weight parameters of the multilayer perceptron (MLP) used to compute the attention weights. The bias term associated with this calculation is denoted as $\boldsymbol{b_e}$. Additionally, $\boldsymbol{h_t}$ refers to the hidden layer output at time $\boldsymbol{t}$. The attention weight assigned to different features at time $\boldsymbol{t}$ is represented by $\boldsymbol{\alpha_t}$, and its corresponding mathematical formulation is provided in Equation (12).

$$\boldsymbol{\alpha_t = \frac{\exp e_t}{\sum_{j=1}^{n} e_j}} \tag{12}$$

Here, $\boldsymbol{e_j}$ represents the weight scores assigned to different features at time $\boldsymbol{j}$. The overall output of the attention mechanism at time $\boldsymbol{t}$ is denoted as $\boldsymbol{C_t}$. The specific mathematical formulation for this computation is provided in Equation (13).

$$\boldsymbol{C_t = \sum\nolimits_{j=1}^{n} \alpha_j h_j} \tag{13}$$

The attention mechanism dynamically computes and refines the hidden layer state associated with the original output feature. It prioritizes key information, ensuring that crucial patterns are effectively learned and retained while minimizing the impact of less significant data. By emphasizing essential factors and strengthening the model's ability to capture the relationships within the predicted sequence, it enhances the understanding of underlying dependencies, ultimately improving forecasting accuracy.

Model Architecture

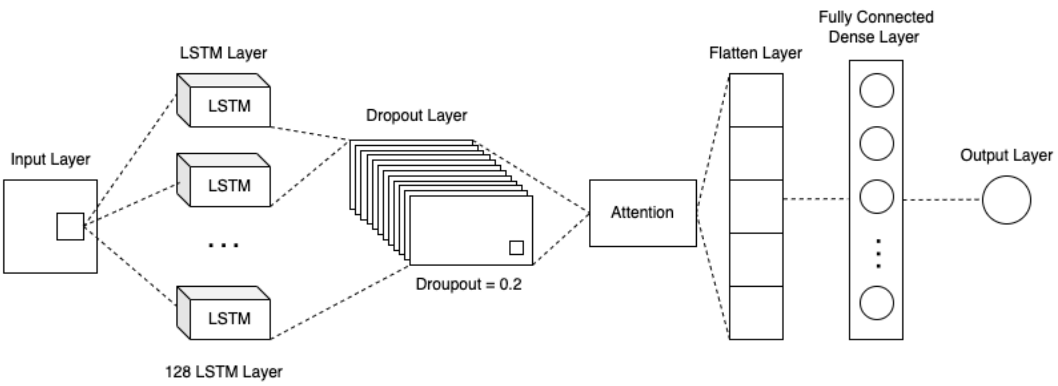

Figure 2 LSTM-Attention Architecture

The architecture of the LSTM with Attention model is shown in Figure 2. It consists of the following layers:

1) Input Layer. Accepts sequential exchange rate data as input.
2) LSTM Layer. Consist of 128 LSTM units, allowing the network to capture complex temporal dependencies.
3) Dropout Layer. Applies a dropout rate of 20% to prevent overfitting.
4) Attention Mechanism. Enhances the model's ability to focus on relevant historical time steps by computing weighted attention scores.
5) Flatten Layer. Converts the attention-enhanced LSTM outputs into a fixed-size vector.

6) Fully Connected Dense Layers with details as follows.
   a) A 64-unit dense layer with ReLU activation for feature extraction.
   b) A 32-unit dense layer with ReLU activation to refine predictions.
7) Output Layer: A single neuron for predicting the future exchange rate value.

The LSTM with attention model follows these steps:
1) Data Normalization. The exchange rate data is scaled using Min-Max normalization to improve numerical stability.
2) Sequence Construction. Sliding window sequences are created to structure the data into input-output pairs for time series forecasting.
3) LSTM Network Design.
   a) An input layer processes sequential data.
   b) LSTM layers capture temporal dependencies.
   c) An attention layer refines the focus on relevant past data.
   d) Fully connected layers transform extracted features into predictions.
4) Model Compilation and Training. The model is trained using the Adam optimizer to efficiently adjust weights and minimize loss. Mean Squared Error (MSE) is used as the loss function, optimizing numerical precision. The model is trained for 100 epochs with a batch size of 32 to ensure effective learning without excessive computational cost.
5) Forecasting and Evaluation.
   a) Predictions are generated for both training and test datasets.
   b) Inverse Min-Max Scaling is applied to convert predictions back to the original exchange rate scale.
6) Performance is assessed using Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE)

F. Hyperparameter Tuning

To improve model efficiency, several hyperparameters are optimized:
1) Number of LSTM Units: 128 in the first layer.
2) Dropout Rate: 0.2 for regularization.
3) Dense Layers: 64 and 32 units with ReLU activation.
4) Batch Size: 32.
5) Number of Epochs: 100.
6) Scaling Method: Min-Max normalization.

IV. RESULTS AND DISCUSSIONS

A. Data Exploration

The initial step in this study involves exploratory data analysis (EDA) to understand the underlying patterns, trends, and characteristics of the USD/IDR exchange rate data. The dataset consists of monthly observations spanning from January 1990 to December 2024, providing a comprehensive historical view of the exchange rate dynamics. The time series plot of the monthly USD/IDR exchange rate is shown in Figure 3.

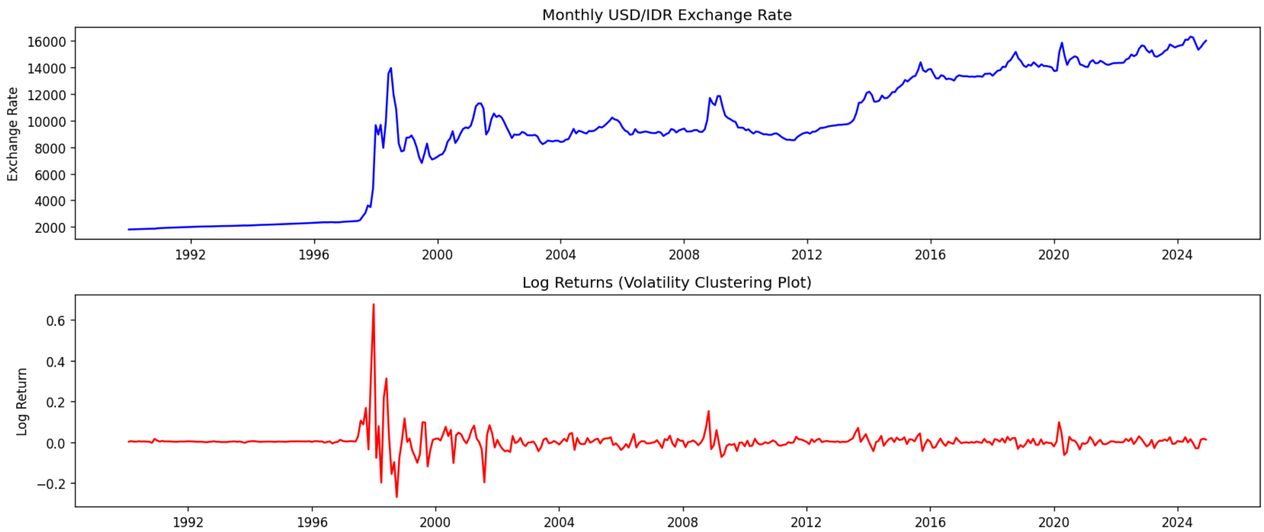

Figure 3 Time Series Plot of IDR/USD Exchange Rate and Log Returns Plot

The time series plot in Figure 3 shows notable volatility and periodic spikes in the exchange rate, reflecting significant macroeconomic events and external shocks impacting the Indonesian economy. Notably, sharp fluctuations can be observed around significant economic crises, including the Asian Financial Crisis (1997-1998) and the Global Financial

Crisis (2008). These observations align with prior studies indicating that exchange rates exhibit considerable volatility during economic turmoil periods. Figure 3 also shows the plot of log returns of the exchange rate series, which are the percentage changes between time periods, to visually demonstrate volatility clustering. The log returns plot shows that periods of large absolute returns, both positive and negative, are often followed by other periods of large movements. Conversely, calm periods with small fluctuations tends to continue for extended duration. These patterns are not random but appear in clusters, indicating that volatility is time dependent. The condition showed in the log return plot suggest the presence of conditional heteroskedasticity, meaning the variance of returns changers over time rather than remaining constant.

Table 1 Descriptive Statistics of IDR/USD Exchange Rate

Statistics	Value
Mean	9,197.543
Standard Deviation	4442.174
Skewness	-0.411
Kurtosis	-0.937

Descriptive statistics in Table 1 highlight a mean exchange rate of approximately 9,197 IDR/USD with a high standard deviation of 4442.174, along with clear volatility clustering patterns seen during the Asian Financial Crisis and Global Financial crisis shown in Figure 3, suggesting a significant volatility. The time series data also presents a skewness and kurtosis higher than normal distribution benchmarks, indicative of non-linear patterns and volatility clustering. The dataset exhibits clear volatility clustering, where periods of high volatility are followed by similar periods, a key feature suitable for GARCH modeling.

Table 2 ADF Test Results

Test	Test Statistic	p-value	Stationarity
Original Series	-1.092	0.718	Non-Stationary
First Differenced Series	-6.036	0.000	Stationary

The Augmented Dickey-Fuller (ADF) test was conducted to assess the stationarity of the exchange rate time series. The test results indicated that the series was non-stationary at the level but became stationary after first differencing, confirming the presence of dynamic fluctuation and validating the suitability of using time-series models such as MS-GARCH and LSTM.

B. MS-GARCH Model

This section presents the empirical results obtained from fitting the Markov Switching Generalized Autoregressive Conditional Heteroskedasticity (MS-GARCH) model to the USD/IDR exchange rate data. The MS-GARCH model was chosen for its ability to capture volatility clustering and regime shifts prevalent in financial time series data. The dataset comprised 420 monthly observations, spanning from January 1990 to December 2024. The dataset was divided into training (January 1990–September 2023) and testing (October 2023–December 2024) subsets for robust model evaluation.

Initially, the training dataset underwent a Box-Cox transformation to stabilize variance, resulting in an optimal lambda of $\lambda = 1.0786$, indicating a slight variance-stabilizing adjustment to the original data. The optimal ARIMA model was determined through automated selection using the Akaike Information Criterion (AIC), resulting in an ARIMA(2,1,2) specification, providing the lowest AIC value of 6717.236.

Table 3 MS-GARCH Model Estimation Results

Parameters	Coefficient	Std. Error	p-value	Significance
Regime 0 (Low Volatility)				
Constant	27.0419	17.262	0.117	Not Significant
Variance (σ^2)	81,140	9,560.05	0	Significant
Regime 1 (High Volatility)				
Constant	218.2341	210.085	0.299	Not Significant
Variance (σ^2)	3,728,000	51.138	0	Significant
Regime Transition Probabilities				
Probability (Regime 0 → Regime 0)	0.9609	0.013	0	Significant
Probability (Regime 1 → Regime 0)	0.1403	0.045	0.002	Significant

Subsequently, the residuals from the ARIMA model were modeled using the MS-GARCH approach. The MS-GARCH model was fitted with two regimes ($k = 2$), allowing it to capture different volatility states inherent in exchange rate fluctuations. The model demonstrated strong capabilities in identifying volatility regimes, aligning with known historical periods of economic volatility and relative stability. Table 3 shows the key estimation results and performance metrics, providing insights into the significance and effectiveness of your MS-GARCH model.

The fitted MS-GARCH model clearly differentiated two volatility regimes. The low volatility regime was characterized by relatively stable exchange rate movements, while the high volatility regime corresponded to periods of economic uncertainty and crisis, confirming historical events such as the Asian Financial Crisis and the global financial crisis periods. Although the constant term in both regimes was not statistically significant, this does not diminish the model's explanatory power. Instead, it highlights that the primary distinguishing factor between regimes is volatility rather than changes in mean return levels. This is expected in exchange rate dynamics, where market shocks predominantly affect volatility clustering rather than the direction of returns. Regime 0 (Low Volatility), characterized by lower variance ($\sigma^2 = 81{,}140$), statistically significant at the 1% level (p-value < 0.001). Regime 1 (High Volatility), demonstrated considerably higher variance ($\sigma^2 = 3{,}728{,}000$), also statistically significant at the 1% level (p-value < 0.001).

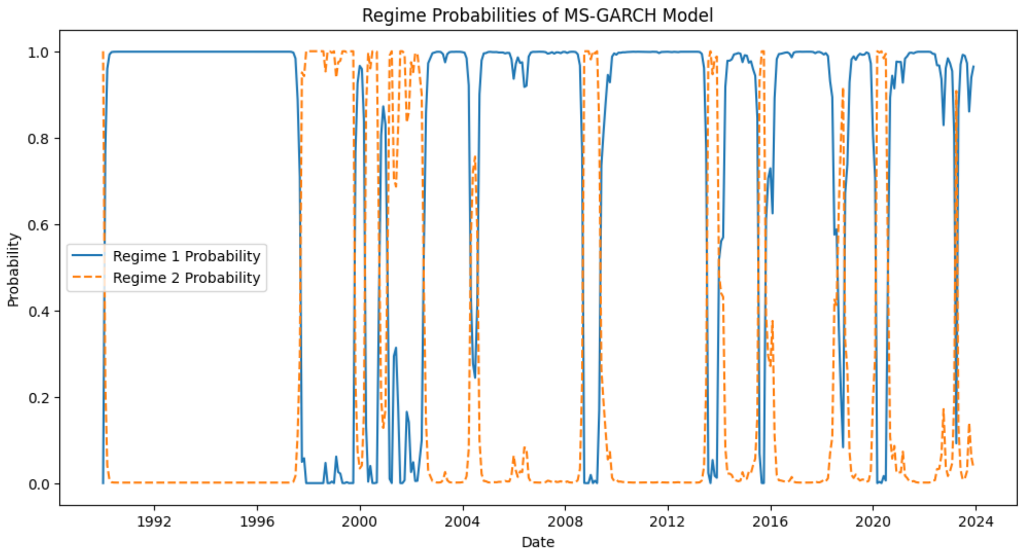

Figure 4 Regime Probabilities of MS-GARCH Model

Figure 4 displays smoothed regime probabilities across the sample period. The figure distinctly reveals regime shifts throughout the observation period (January 1990 to December 2023). High-volatility episodes, characterized by sharp spikes in the dashed orange probability line nearing values close to one, correspond clearly to historical periods of major economic instability in Indonesia. Specifically, the high-volatility regime prominently dominates during the period around 1997–1998, aligning with the Asian financial crisis when the Indonesian Rupiah experienced dramatic depreciation due to regional financial instability. Another significant regime shift is evident around 2008, coinciding with the global financial crisis marked by heightened uncertainty and volatility within global and domestic financial markets. Additionally, periodic regime shifts appear intermittently post-2008, likely reflecting domestic economic policy interventions, external economic shocks, or fluctuating market sentiments.

Transition probability parameters were significant and indicated persistent regimes. The probability of remaining in low volatility (Regime 0) was estimated at 0.9609, whereas the transition probability from high volatility (Regime 1) back to low volatility was lower at 0.1403. This pattern of transition probabilities indicates that once the exchange rate enters the high-volatility regime, it tends to persist for prolonged periods, which aligns with historical economic episodes and confirms the model's strong ability to capture regime persistence and volatility clustering.

Table 4 Residual Diagnostic Test Results for MS-GARCH Model

Test	p-value	Result
Ljung-Box Test (Autocorrelation)	0.1502	No significant autocorrelation
ARCH Test (Heteroskedasticity)	0.0000	Heteroskedasticity detected
Kolmogorov-Smirnov Test (Normality)	0.0000	Residuals deviate from normality

To assess the adequacy of the MS-GARCH model, residual diagnostic tests were conducted, including the Ljung-Box test for autocorrelation, the ARCH test for heteroskedasticity, and the Kolmogorov-Smirnov (KS) test for normality. The Ljung-Box test yielded a p-value of 0.1502, indicating that the null hypothesis of no significant autocorrelation could not be rejected. This suggests that the residuals are likely white noise, confirming that the MS-GARCH model has effectively captured the time-dependent volatility patterns in the exchange rate data. The absence of significant autocorrelation implies that the model does not leave substantial unaccounted-for temporal dependencies, affirming its suitability in modeling the volatility process.

However, the ARCH test returned a p-value of 0.0000, suggesting that heteroskedasticity remains present in the residuals. This finding indicates that volatility clustering persists, meaning that while the MS-GARCH model significantly improves over standard models in capturing time-varying volatility, some aspects of conditional heteroskedasticity may still be unexplained. This result suggests that the exchange rate data exhibits periods of heightened uncertainty, which the model recognizes but does not fully eliminate. Given that financial markets are inherently volatile, the presence of remaining heteroskedasticity does not necessarily undermine the model's usefulness but rather highlights the complexity of capturing all variations in exchange rate fluctuations.

Additionally, the Kolmogorov-Smirnov (KS) test for normality also returned a p-value of 0.0000, rejecting the null hypothesis and indicating that the residuals deviate significantly from a normal distribution. This suggests that the model does not fully capture the distributional properties of exchange rate movements, particularly extreme fluctuations. While the assumption of normally distributed residuals is often desirable in statistical modeling, financial data frequently exhibit non-normality due to the presence of fat tails and skewness, making such deviations expected.

Although the standard MS-GARCH model assumes normally distributed errors, research by Haas, Mittnik, and Paolella [12] emphasizes that this assumption may be overly restrictive. Their work proposes a more general MS-GARCH framework incorporating Student-t distributions, which significantly improves the model's ability to account for excess kurtosis and heavy-tailed behavior. This suggests that MS-GARCH models can remain robust in practice when extended to non-Gaussian specifications. Additionally, Klaassen [13] demonstrated that the regime-switching component itself enhances volatility forecasting, particularly during periods of market turbulence, even when standard distributional assumptions are retained. These findings collectively support the continued use of MS-GARCH for structural volatility analysis, while highlighting the value of alternative specifications and hybrid modeling approaches to address nonlinearity and extreme behavior more comprehensively.

Despite these assumption violations, the MS-GARCH model successfully identifies distinct volatility regimes and effectively captures the temporal dependencies in exchange rate fluctuations, which are particularly useful for risk management, policy evaluation, and scenario analysis. We emphasize that the model is more suitable for structural volatility analysis than for high-frequency point forecasting. The presence of residual heteroskedasticity and non-normality suggests that further refinements could improve statistical modeling approaches. To complement this, we applied the LSTM-Attention model, which is less assumption-dependent and better suited for capturing complex nonlinear patterns and point forecasts. The combination of both models strengthens confidence in the overall analysis: MS-GARCH provides interpretability and regime context, while LSTM offers forecasting precision.

C. LSTM-Attention Model

The LSTM with Attention Model was evaluated for its ability to predict the USD/IDR exchange rate using historical data. The model was trained using 100 epochs with a batch size of 32 and optimized with the Adam optimizer. The dataset was divided into training data (before October 2023) and testing data (October 2023 – December 2024).

The training process of the LSTM with Attention Model was evaluated by monitoring the loss function during the learning phase. The training loss curve, as illustrated in Figure 5, provides insights into how well the model learns patterns from the historical exchange rate data over multiple training epochs.

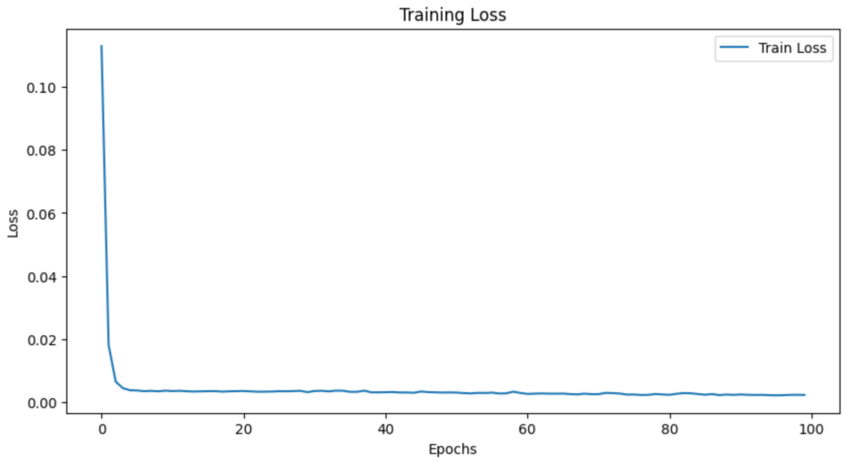

Figure 5 Training Loss Curve

From Figure 5, it is evident that the model experienced a sharp decline in loss within the first few epochs, indicating rapid learning of key temporal dependencies in the initial phase. This behavior is expected, as the model quickly adjusts its weights to minimize prediction errors. After the initial steep drop in loss, the curve transitions into a stabilization phase, where the loss gradually converges to a near-constant value. This suggests that the model has reached a point where additional training yields minimal further improvements, signifying effective convergence.

To further assess the forecasting performance of the LSTM with Attention Model, a comparison between actual and predicted exchange rates was conducted for both the training and testing periods. The results are visualized in Figure 6, which illustrates how well the model captures trends in the USD/IDR exchange rate.

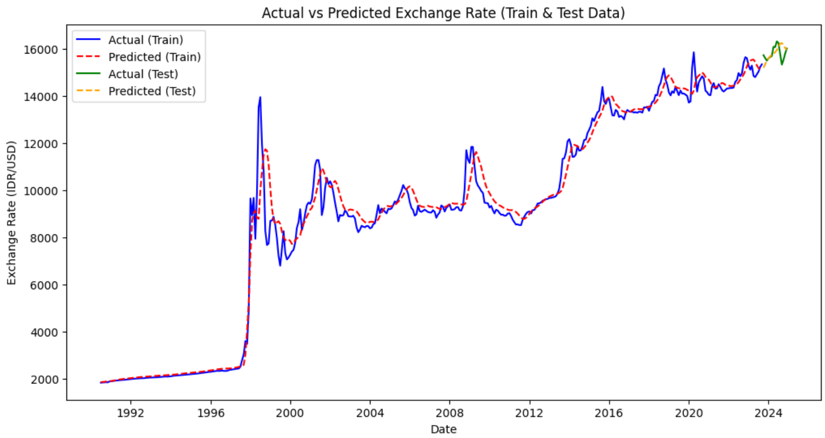

Figure 6 Actual vs Predicted Exchange Rate (Train & Test Data)

From Figure 6, it is evident that the LSTM-Attention model successfully captures both short-term fluctuations and long-term trends in the exchange rate data. The predicted values closely track the actual exchange rates in both training and testing periods, demonstrating the model's ability to effectively learn historical patterns and generalize to unseen data. The model responds well to periods of high volatility, particularly during financial fluctuations, adapting dynamically to nonlinear trends in the exchange rate. Additionally, the forecasting accuracy remains stable across different time periods, indicating that the model is robust to various market conditions and does not overfit specific trends. These results confirm that the LSTM-Attention model provides a reliable and adaptive approach to exchange rate prediction, surpassing traditional statistical methods in handling complex temporal dependencies.

D. Validation of MS-GARCH and LSTM-Attention Model

To assess the forecasting performance of the MS-GARCH and LSTM-Attention models, their prediction results are evaluated against actual exchange rate data using Root Mean Squared Error (RMSE), Mean Absolute Percentage Error (MAPE), and Mean Absolute Error (MAE). RMSE measures the goodness of fit by calculating the difference between the predicted and actual values, assigning greater weight to larger errors, thereby providing a more sensitive measure of prediction accuracy. A lower RMSE value indicates a more accurate model. MAE, on the other hand, measures the average magnitude of errors without considering their direction, making it useful for understanding the overall discrepancy between predictions and actual values. MAPE expresses the prediction error as a percentage, offering an interpretable metric for comparing deviations between predicted and actual values across different scales. By utilizing these three-evaluation metrics, this analysis provides a comprehensive quantitative comparison of both models, helping to determine which method, MS-GARCH or LSTM with Attention Mechanism, delivers superior accuracy in forecasting the USD/IDR exchange rate. The results, summarized in Table 5, show that the LSTM-Attention model outperforms the MS-GARCH model across all accuracy metrics:

Table 5 Forecasting Performance Evaluation

Metric	MS-GARCH Model	LSTM-Attention
RMSE	451.35	376.98
MAE	375.67	279.30
MAPE	2.34%	1.77%

While both models provide useful insights into exchange rate dynamics, their comparative performance reveals important trade-offs. The LSTM-Attention model outperforms the MS-GARCH model across all evaluation metrics (RMSE, MAE, and MAPE), demonstrating superior accuracy in capturing both short-term fluctuations and long-term patterns in the USD/IDR exchange rate. A MAPE of 1.77% implies that the LSTM-Attention model's predictions deviate, on average, by less than 2% from actual values. For financial institutions managing daily currency exposure, this level of precision can significantly enhance risk mitigation strategies and portfolio valuation accuracy. Moreover, the deep learning-based approach effectively captures nonlinear dependencies and long-term temporal relationships, allowing it to adapt to dynamic exchange rate fluctuations more efficiently than the statistical MS-GARCH model. While MS-GARCH is well-suited for capturing volatility clustering and regime shifts, its forecasting performance is less accurate than LSTM-Attention, particularly during high-volatility periods. The MS-GARCH model primarily focuses on structural changes in volatility rather than learning complex patterns in sequential data. In contrast, LSTM-Attention dynamically assigns weight to important historical time steps, enabling it to better predict exchange rate movements.

However, it is important to contextualize these performance improvements. The LSTM model's lower error rates do not imply it is universally superior; rather, they indicate better performance on the specific test set and forecast horizon used in this study. LSTM models are also data-hungry and require regular retraining to remain effective in dynamic financial environments. Additionally, the "black box" nature of deep learning models limits interpretability, which can be a barrier for policymakers or analysts who need to understand the drivers behind forecasted movements.

On the other hand, the MS-GARCH model, while less accurate in point forecasting, offers strong interpretability and a clear representation of regime-dependent volatility, making it particularly valuable for risk assessment and understanding macroeconomic uncertainty. The model successfully identified high-volatility regimes corresponding to known financial crises and provided useful diagnostics. However, residual diagnostics indicated the presence of heteroskedasticity and non-normality, suggesting that the model, while useful, does not fully capture all statistical properties of the data. These findings emphasize that while modern deep learning methods provide greater predictive accuracy, traditional econometric models still offer analytical transparency and diagnostic power that are essential in policy and economic analysis contexts.

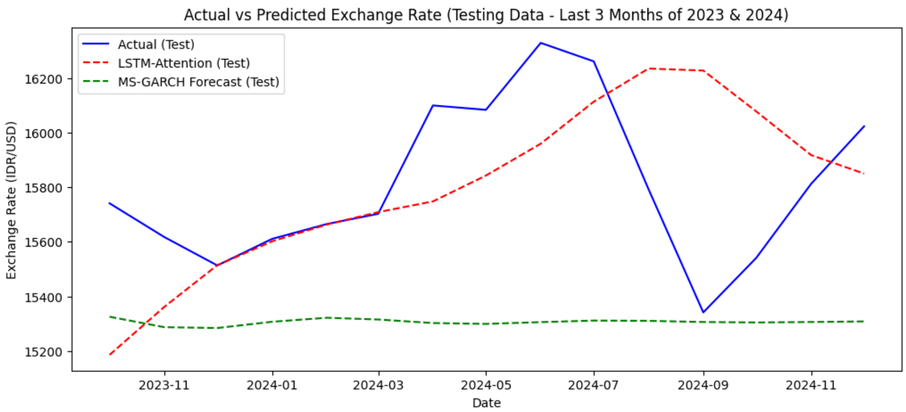

Figure 7 Comparison of Actual vs. Predicted Exchange Rate Using MS-GARCH and LSTM-Attention Models (Testing: October 2023 – December 2024)

Figure 7 illustrates the comparison between actual USD/IDR exchange rate movements and the predictions generated by the MS-GARCH and LSTM-Attention models for the last three months of 2023 and all months 2024. The LSTM-Attention model (red dashed line) closely follows the actual exchange rate fluctuations, indicating its strong ability to capture short-term variations and nonlinear dependencies in the data. In contrast, the MS-GARCH model (green dashed line) exhibits a relatively stable forecast, suggesting that while it effectively models long-term volatility regimes, it lacks the adaptability to capture sudden exchange rate fluctuations. The actual exchange rate (blue line) demonstrates significant variations, particularly in mid-2024, where the LSTM-Attention model successfully follows the trend, whereas the MS-GARCH model remains relatively constant. These results reinforce the superior predictive performance of LSTM-Attention in forecasting exchange rate movements, as it better accounts for complex temporal dependencies compared to the regime-based structure of MS-GARCH, which is more suited for volatility analysis rather than precise short-term forecasting.

The deviations observed between the predicted and actual exchange rate values, particularly in the MS-GARCH model's testing performance, underscore important limitations in its use for short-term forecasting. While MS-GARCH effectively identifies regime shifts and long-run volatility patterns, it is not primarily designed for high-frequency point prediction. This is evident in periods of heightened market fluctuation where the model fails to fully adapt to abrupt changes, resulting in diminished forecast accuracy. While the LSTM-Attention model achieved better accuracy across all metrics, it is important to note that this improvement may be partly due to the model's ability to adapt to short-term fluctuations, which are more pronounced in recent periods. In contrast, the MS-GARCH model, although slightly less accurate, provides clearer interpretability and better regime-based volatility insights. These findings suggest that while MS-GARCH is valuable for structural risk analysis and volatility regime identification, it may not be suitable for applications requiring real-time precision, such as trading, tactical hedging, or short-term monetary response. Consequently, the choice of forecasting model should be guided by the intended application: for example, policymakers focused on systemic stability may benefit from regime-based insights, whereas investors or financial institutions may prefer models like LSTM-Attention for their operational forecasts. This distinction highlights the importance of aligning model design with the forecasting horizon and use-case sensitivity.

E. Forecasting

Forecasting the USD/IDR exchange rate for the 2025 period was conducted using the LSTM-Attention model, as shown in Figure 8. The forecasted exchange rate for 2025, as illustrated in the figure, was generated using the LSTM-Attention model, leveraging historical data trends from 2024. The blue solid line represents the actual exchange rate data for 2024, while the red dashed line shows the forecasted exchange rate for 2025. The model predicts a gradual upward trend in the USD/IDR exchange rate throughout 2025, suggesting potential depreciation of the Indonesian Rupiah against the US Dollar. This indicates that, based on historical patterns, the exchange rate is expected to experience steady increases rather than extreme fluctuations. However, the initial dip in early 2025 suggests possible short-term stabilization before the projected upward trend resumes.

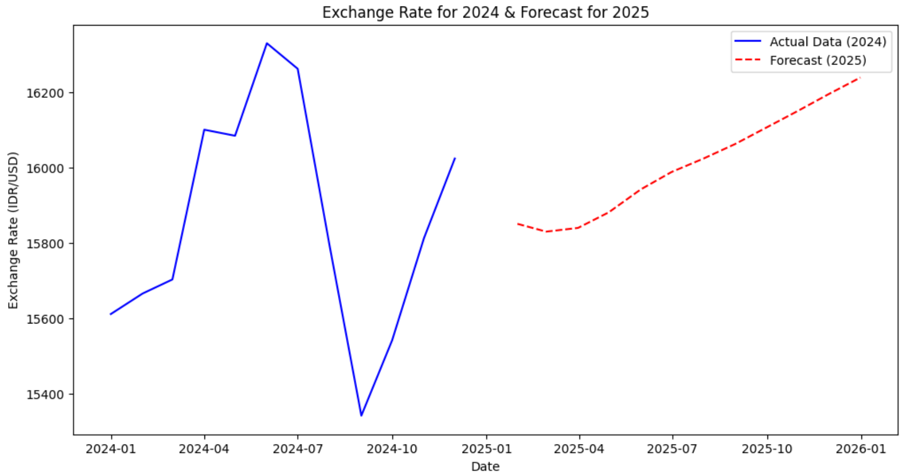

Figure 8 Actual USD/IDR Exchange Rate for 2024 and Forecast for 2025 Using the LSTM-Attention Model

These findings highlight the LSTM-Attention model's ability to capture long-term trends in exchange rate movements, making it useful for forecasting purposes. However, external macroeconomic factors and unforeseen financial events could introduce deviations from the predicted trajectory, emphasizing the need for continuous model updates and real-time adjustments in practical forecasting applications.

V. CONCLUSIONS AND SUGGESTIONS

This study conducted a comprehensive comparison of statistical (MS-GARCH) and deep learning (LSTM with Attention Mechanism) approaches for forecasting the USD/IDR exchange rate. The findings demonstrate that while both models effectively capture the volatility dynamics of the exchange rate, they exhibit different strengths.

The MS-GARCH model successfully identifies volatility clustering and regime shifts, confirming the presence of distinct low and high volatility periods in the exchange rate data. This characteristic makes MS-GARCH a valuable tool for understanding structural changes in financial markets, particularly in periods of economic uncertainty. However, the residual diagnostic tests revealed that heteroskedasticity and non-normality persist, indicating that the model does not fully capture extreme fluctuations in exchange rate movements and may limit its precision in forecasting under certain conditions. Additionally, the model’s forecasting performance, while reasonable, exhibited limitations during periods of high market turbulence.

On the other hand, the LSTM with Attention Mechanism demonstrated superior forecasting accuracy across all evaluation metrics (RMSE, MAE, and MAPE), outperforming MS-GARCH in predicting exchange rate movements. The deep learning-based approach effectively captures nonlinear dependencies and long-term temporal relationships, making it more adaptable to the dynamic nature of foreign exchange markets. The attention mechanism further enhances the model's ability to focus on relevant historical data, improving its predictive performance. Nevertheless, its limited interpretability and dependence on continuous retraining pose challenges for long-term implementation in policy environments.

Despite the higher accuracy of LSTM-Attention, the interpretability of MS-GARCH remains a crucial advantage, particularly for policymakers and financial analysts who require an understanding of market volatility patterns. Therefore, while deep learning models provide superior predictive capabilities, traditional statistical methods still hold value in financial analysis and risk management. It is essential to note that better predictive accuracy does not automatically imply better model utility. The choice of model should depend on the specific goals of the user, whether accuracy, interpretability, or regime identification is the priority.

The findings suggest that combining statistical and deep learning approaches may offer complementary benefits, balancing interpretability and predictive accuracy. Future research should consider hybrid modelling approaches that combine the interpretability of MS-GARCH with the predictive power of deep learning architectures. Additionally, incorporating external macroeconomic indicators could further enhance forecasting performance. Expanding the evaluation to include robustness across different economic scenarios and periods would also improve generalizability of the findings.

Overall, this study highlights the growing potential of deep learning models in financial forecasting, while also reaffirming the importance of traditional econometric models in understanding exchange rate volatility. The insights from this research can aid policymakers, investors, and financial institutions in making data-driven decisions regarding foreign exchange risk management and policy formulation.

REFERENCES

[1] C. Murtala, T. Putra, E. Gunawan and I. Iskandar, "Rupiah Exchange Rate Stability Towards US Dollars In Indonesia by VAR Approacch," *International Journal of Academic Research Business and Social Sciences,* no. 9(4), p. 174–191, 2019.

[2] D. Halawa, "Faktor-Faktor Yang Mempengaruhi Perubahan Nilai Tukar Rupiah Terhadap Dolar Amerika Serikat," *Journal Economics And Strategy,* vol. 4, no. 1, pp. 52-61, 2023.

[3] A. Kurnia and D. Purnomo, "Fluktuasi Kurs Rupiah terhadap Dollar Amerika Serikat pada Periode Tahun 1997.I – 2004.IV," *Jurnal Ekonomi Pembangunan,* vol. 10, no. 2, p. 234–249, 2009.

[4] L. Judijanto, Nurussama and R. Apriliani, "Financial Risk Management Strategies in Dealing With Foreign Exchange Rate Fluctuations," in *Prosiding Seminar Nasional Indonesia*.

[5] M. Nunian, S. M. Zahari and S. Shariff, "Modelling foreign exchange rates: a comparison between Markov-switching and Markov-switching GARCH," *Indonesian Journal of Electrical Engineering and Computer Science,* vol. 20, no. 2, pp. 917-923, 2020.

[6] E. Nkemnole, A. Taiwo and A. Ebomese, "Forecasting Exchange Rate Volatility with Markov-Switching GARCH Model Estimation Methods," in *Practical Statistical Learning and Data Science Methods: Case Studies from LISA 2020 Global Network, USA*, Switzerland, Springer Nature Switzerland, 2025, pp. 501-523.

[7] H. Makika, J. Romano and R. Ballini, "Deep Learning Techniques For Time Series Forecasting: An Application For Exchange Rates," in *IEEE 33rd International Workshop on Machine Learning for Signal Processing (MLSP)*, Rome, Italy, 2023.

[8] Z. Rusdi, C. Lubis and V. G. Tjandra, "Prediksi Kurs Mata Uang dengan Metode Long Short Term Memory (Lstm) Berbasis Attention," *Journal of Computer Science and Information Systems,* vol. 5, no. 2, pp. 45-51, 2021.

[9] I. Hidayat, L. A. S. I. Akbar and A. S. Rachman, "Prediksi Nilai Tukar Mata Uang Menggunakan Algoritma Long Short-Term Memory dan Random Forest," *Journal of Computer System and Informatics (JoSYC),* vol. 6, no. 1, pp. 107-116, 2024.

[10] T. Bollerslev, "Generalized autoregressive conditional heteroskedasticity," *Journal of Econometrics,* vol. 31, no. 3, pp. 307-327, 1986.

[11] J. D. Hamilton, "A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle," *Econometrica,* vol. 57, no. 2, pp. 357-384, 1989.

[12] M. Haas, S. Mittnik and M. S. Paolella, "A New Approach to Markov-Switching GARCH Models," *Journal of Financial Econometrics,* vol. 2, no. 4, p. 493–530, 2004.

[13] F. Klaassen, "Improving GARCH volatility forecasts with regime-switching GARCH," *Empirical Economics,* vol. 27, p. 363–394, 2002.

[14] S. Hochreiter and J. Schmidhuber, "LONG SHORT-TERM MEMORY," *Neural Computation,* vol. 9, no. 8, pp. 1735-1780, 1997.

[15] X. Zhang, P. Li, X. Han, Y. Yang and Y. Cui, "Enhancing Time Series Product Demand Forecasting With Hybrid Attention-Based Deep Learning Models," *IEEE Access,* vol. 12, pp. 190079-190091, 2024.

[16] X. Wu, "A Stock Price Foresting Using LSTM Based on Attention Mechanism," in *Proceedings of the 2022 International Conference on Economics, Smart Finance and Contemporary Trade (ESFCT 2022)*, 2022.

[17] S. Saadati and M. Manthouri, *Forecasting Foreign Exchange Market Prices Using Technical Indicators with Deep Learning and Attention Mechanism,* 2024.

[18] M. Islam and E. Hossain, "Foreign Exchange Currency Rate Prediction using a GRU-LSTM Hybrid Network," *Soft Computing Letters,* vol. 3, 2021.

[19] J. Qiu, B. Wang and C. Zhou, "Forecasting stock prices with long-short term memory neural network based on attention mechanism," *PloS one,* vol. 15, no. 1, 2020.

[20] J. D. Hamilton and R. Susmel, "A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle," *Journal of Econometrics,* vol. 64, no. 1-2, pp. 307-333, 1994.

[21] H. Sak, A. Senior and F. Beaufays, "Long short-term memory recurrent neural network architectures for large scale acoustic modeling," in *roceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH*, 2024.

[22] X. Ran, Z. Shan, Y. Fang and C. Lin, "An LSTM-Based Method with Attention Mechanism for Travel Time Prediction," *Sensors,* vol. 19, no. 4, 2019.

[23] F. orina, E. Hysa, U. Ergün, M. Panait and M. Voica, "The Effect of Exchange Rate Volatility on Economic Growth: Case of the CEE Countries," *Journal of Risk and Financial Management,* vol. 13, no. 8, 2020.