**Prediction/Forecast of Forex Rate Using Data Analytic Techniques - USD to LKR**

**Group Members**

* **Sachini Jayawardena (023) / Rumesh Mohan (031) / Sajith Sanjeewa (079)**

**Keywords**

**AIC -** Akaike Information Criterion, **BIC -** Bayesian information Criterion, **HQIC -** Hannan-Quinn Information Criterion, **ARIMA -** Autoregressive Integrated Moving Average, **SARIMA -** Seasonal Autoregressive Integrated Moving Average, **FOREX** - Foreign Exchange

**Abstract**

The Forex (foreign-exchange) market is a worldwide marketplace for the exchange of national currencies. Governments, Banks, International and Local companies, Investors and Different Traders are participants in Forex. In a situation where the exchange rate fluctuates; economic decisions need to be reconsidered. As a result, forecasting exchange rates using time series data has become popular in financial market research. Researchers have discovered that exchange rate fluctuations exhibit the usual features of a nonlinear dynamic system with very complex variables affecting their changes. Deep Learning has proven effective in forecasting a variety of areas. Thus, the use of deep learning algorithms for Forex prediction has become of considerable interest.

**Introduction**

Foreign exchange rate of a country is determined and affected based on its macroeconomic factors and non-macroeconomic factors. Macro economic factors include Interest Rates, Inflation, Foreign Remittance, Trade Balance, etc. Political Considerations (Diplomatic Relations, Elections, Cabinet changes etc.), Natural Disasters, Pandemics, Terrorism Acts are considered as the major non-macroeconomic factors. The aim of this study is to incorporate these factors or indicators into consideration to develop a forecast model to predict the exchange rate for a specified month using data analytic techniques.

As this study is based on the forex market of Sri Lanka, the selected currencies of interest were the United States Dollar (USD), the Sri Lankan Rupee (LKR) and their conversion rates. Changes in rates of the Forex market is widely discussed, mostly watched and analyzed consistently. Sri Lanka has faced and is facing severe economic drawbacks in the recent years due to various political, social and economical disputes and decisions. Therefore, it is essential for the investors and other parties who are interested in the Forex market particularly in a country like Sri Lanka to understand how the market behaves in these highly volatile economic conditions.

Based on different literatures available for the study, we have selected Inflation Rate, Workers Remittance, Interest Rate, Foreign Reserves, Terms of Trade (Trade Balance), Money Supply and Political Stability as the major forcing factors that cause changes and fluctuations in the exchange rates. Application of forecasting models to predict the exchange rates is complex as sometimes these factors change drastically so that using only linear models becomes inefficient in adapting to these changes. In the recent past the deep learning models were developed as a result of advancements in Artificial Intelligence (AI). These models have been evolved and have a wide variety of research and industrial applications. One such application is predicting forex rates in foreign exchange markets. Different forecasting models can be used to predict the fluctuations in forex rates. Both linear models and deep learning models have been used in this study to predict and compare accuracies in forex rates by using multivariate forecasting models.

**Literature Review**

In the study of predicting the foreign exchange rate using data analytic techniques, we have considered the impact of macroeconomic factors such as Inflation Rate, Workers Remittance, Interest Rate, Foreign Reserves, Terms of Trade (Trade Balance), Money Supply and Non-macroeconomic factors like Political Stability. Hence, these will be selected as the features for the forecasting model.

**Macroeconomic Factors**

**Inflation**: This measures the changes in prices paid by consumers for goods and services over time, which is also known as decrease in purchasing power in an economy. Interest rate is one of the determining factors which is highly correlated with inflation rate and exchange rate/forex rate of a country. The change in interest rate of an economy by the Central Bank of a country would help in attracting new foreign capital to a country as it offers higher returns for investors or lenders, which will positively affect the exchange rate of a country. The NCPI is a key indicator of inflation and is used for socio-economic analysis. Economies with higher inflation typically tend to depreciation in their currency relative to the USD. This is also usually accompanied by higher interest rates.

**Interest Rates**, Inflation and Forex are highly correlated. Governments have a collection of tools to manipulate their local exchange rate. By manipulating interest rates, central banks intervene over both inflation and exchange rates and changing interest rates impact inflation and currency valuation. The ultimate objective is to ensure favorable conditions for a stable currency exchange rate in the country. Higher interest rates attract foreign capital and hike the exchange rate and vice versa. However, central banks mitigate the impact of higher interest rates by manipulating other factors such as adjusting interest rates, buying foreign currency, influencing local lending rate, printing money and using other tools to modulate currency exchange rate.

**Foreign Reserves** are assets held on reserve by a central bank and other institutions in foreign currencies. These reserves are used to back liabilities and influence monetary policy. It includes any foreign currencies, bonds, bills, gold reserves etc. Most foreign exchange reserves are mainly held in United States Dollars since it is the most traded currency in the world. Reserves act as a shock absorber against factors that can negatively affect a currency’s exchange rate. Central bank uses its currency reserves to maintain a steady Forex rate.

**Terms of Trade** is the relationship between the prices at which a country sells its exports and the prices paid for its imports. When the export prices increase more than the import prices, a country has positive terms of trade. This positivity or the increasing of trades shows greater demand for the country’s exports. This raises the revenue from exports, which provides increased demand for the country’s currency and higher value given for the currency.

**Remittance** is the transfer of money from foreign workers to their home countries. In many countries remittance represents a significant portion of a nation’s economy. Emerging economies or developing countries rely heavily on foreign remittance. Economists believe that since remittances are so widespread, they have implications that extend beyond an individual’s finance and also promote a country’s macroeconomic development.

**Money Supply** M2 is the reserver’s estimate of the total money supply including currency held by the public, demand deposits held by the public with commercial banks and saving deposits held by the public with commercial banks. The central bank intervenes in the market to smoothen the USD/LKR excessive volatility. An increase in a country’s money supply causes its currency to depreciate and vice versa.

**Non-Macroeconomic Factors**

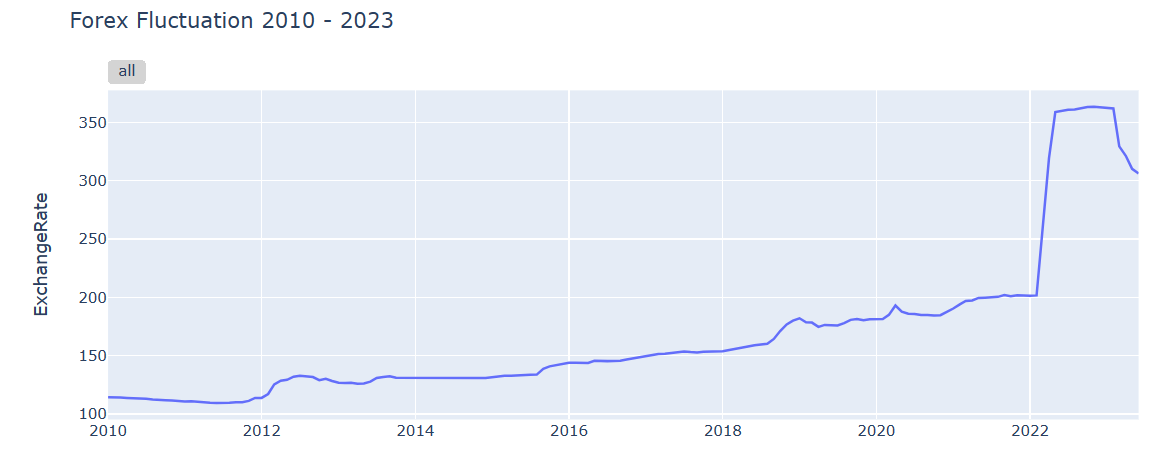
**Political Instability** is regarded by economists as a serious malaise harmful to economic performance. Political instability is a multidimensional phenomenon, not well captured by just one or two variables. It measures by the number of times in a year that the Cabinet changes occur, Gross Domestic Product (GDP) growth, Primary school enrollments, Trade openness and Investments etc. Countries with more robust political stability have stronger and higher valued currencies. We have used the political stability index ranging from **-2.5 (Weak) to +2.5 (Strong)** defined by worldbank available on ‘theglobaleconomy.com’ website up to 2021. We defined the index as -2.0 for a period of (Jan 2022 - Sep 2022), -1.5 for a period of (Oct 2022 - Jan 2023), -1 for a period of (Feb 2022 - June 2022) considering the country’s situation.

**Methodology**

Initial phase of the project was to identify the factors which have an influence in deciding the Forex according to the previous researcher. Then we collected a reasonable amount of data points regarding those factors. Next we did a statistical analysis to find the correlation between Forex and the factors found. The selected seven factors are identified as the key factors impacting Forex by most of previous studies. In the second phase we did data pre-processing, created a model and its training and made future forecasting. All models were made in Python and the libraries used are mentioned below. We have used Deep Learning methods for forecasting.

* **Data Gathering and Extraction**

Data for the above macroeconomic factors were collected from the Central Bank of Sri Lanka (CBSL) and other valid sources including World Bank, International Monetary Fund, etc. Most of them are available online with free public access. Monthly basis data from Jan 2010 to June 2023 were gathered from the above mentioned sources. Figure 1 shows the fluctuations in Foreign Exchange Rate over time from Jan 2010 to June 2023.

**Figure 1: Fluctuations in Forex (USD/LKR) from Jan 2010 - June 2023**

* **Data Preprocessing and Exploratory Data Analysis**

In this step we prepared the dataset for the models. This involves framing the dataset as a supervised learning problem. We aggregated data points and changed them to monthly intervals. Further, framed the problem as predicting the Forex for each month based on seven affecting factors identified from multiple literature reviews. These seven factors are considered as the independent variables which influence the Forex (dependent variable) in the economic domain. Few missing data values were imputed with the mean of that particular variable.

**Table 1: Statistical Analysis of Correlation between the Variables**

|  | **Exchange**  **Rate** | **Monthly NCPI** | **Workers Remittance** | **Interest Rate** | **Foreign Reserves** | **Terms of Trade(Trade Balance)** | **Sri Lanka Money Supply M2** | **Political Stability** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **ExchangeRate** |  |  |  |  |  |  |  |  |
| **Monthly NCPI** | **0.8568**  **V.S.P.** |  |  |  |  |  |  |  |
| **Workers Remittance** | **-0.2748**  **W.N.** | -0.4383 |  |  |  |  |  |  |
| **Interest Rate** | **0.7396**  **S.P.** | 0.8597 | -0.3419 |  |  |  |  |  |
| **Foreign Reserves** | **-0.7174**  **V.S.N.** | -0.6855 | 0.5630 | -0.4863 |  |  |  |  |
| **Terms of Trade(Trade Balance)** | **0.4262**  **M.P.** | 0.4645 | -0.3239 | 0.3716 | -0.3692 |  |  |  |
| **Sri Lanka Money Supply M2** | **0.4104**  **M.P.** | -0.0194 | 0.2641 | -0.2046 | -0.1870 | -0.0836 |  |  |
| **Political Stability** | **-.05368**  **V.S.N.** | -0.7362 | 0.7407 | -0.6522 | 0.6429 | -0.3857 | 0.3079 |  |

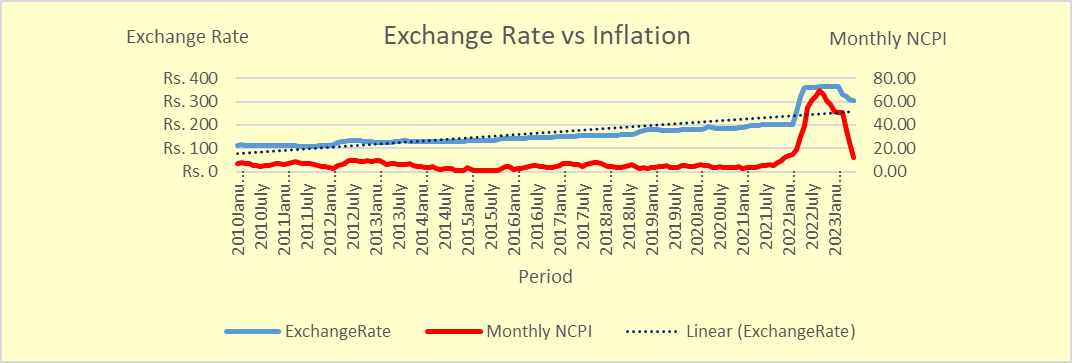
**V.S.P - Very Strong Positive S.P - Strong Positive M.P - Moderate Positive**

**V.S.N - Very Strong Negative S.N - Strong Negative W.N - Weak Negative**

Table 1 shows the correlation between each variable.

1. **Monthly** [**National Consumer Price I**](http://www.statistics.gov.lk/WebReleases/NCPI_December_2022)**ndex (NCPI)**

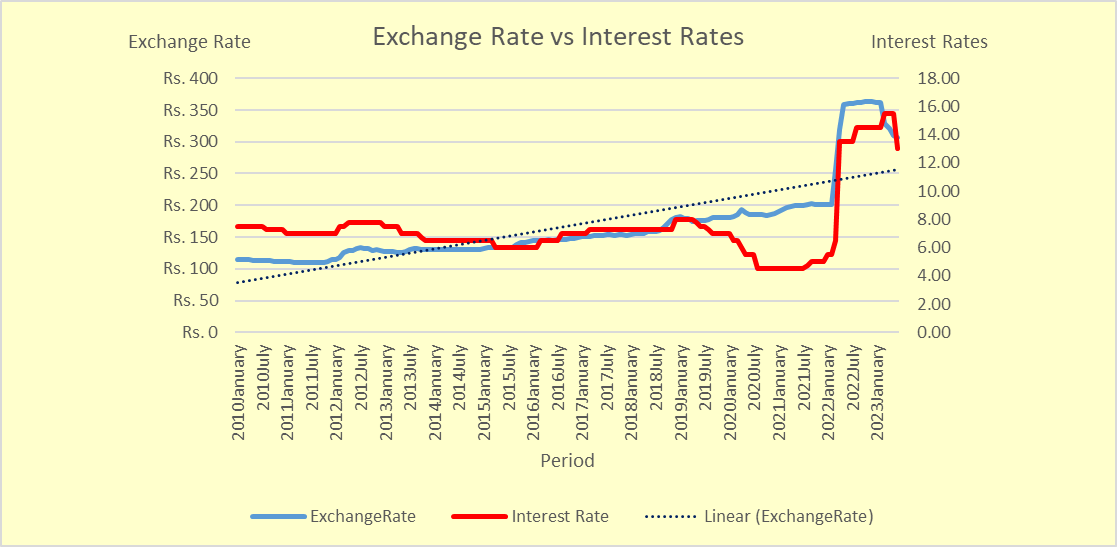
Figure 2 illustrates the relationship between Exchange rate and the Monthly NCPI during the period from 2010 to 2023. The correlation coefficient between Exchange Rate and the NCPI is 0.84768 which emphasizes a highly positive relationship.

**Figure 2: Exchange Rate (USD / LKR) and NCPI**

1. **Interest Rate**

Figure 3 shows the relationship between Exchange Rate and the Monthly Interest Rates from 2010 to 2023. The correlation coefficient between Exchange Rate and the Monthly Interest Rate is 0.7396 which emphasizes a highly positive relationship.

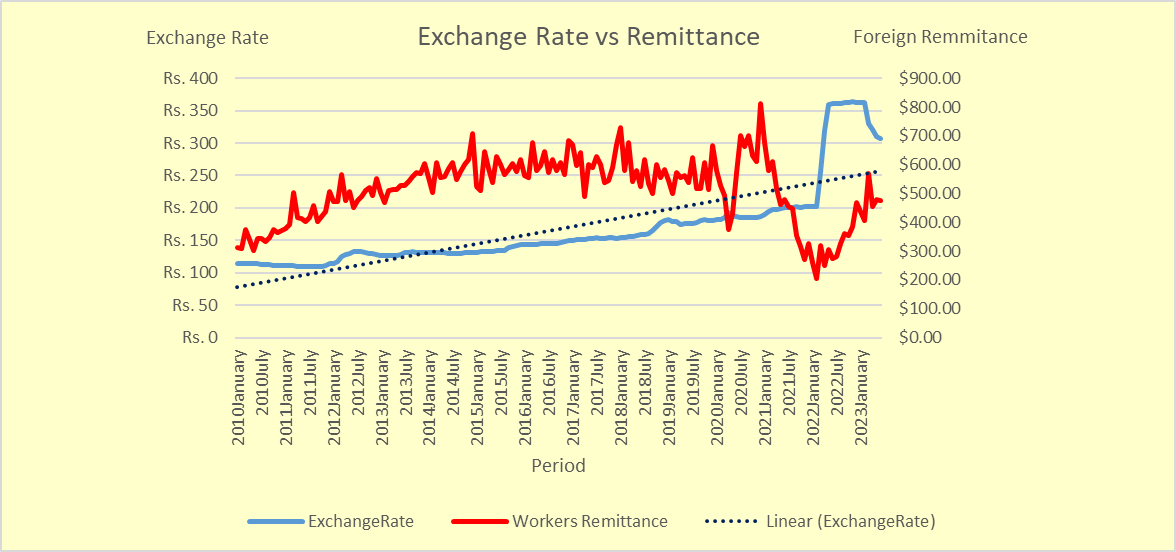
**Figure 3: Exchange Rate (USD / LKR) and Interest Rates**



1. **Workers Remittance**

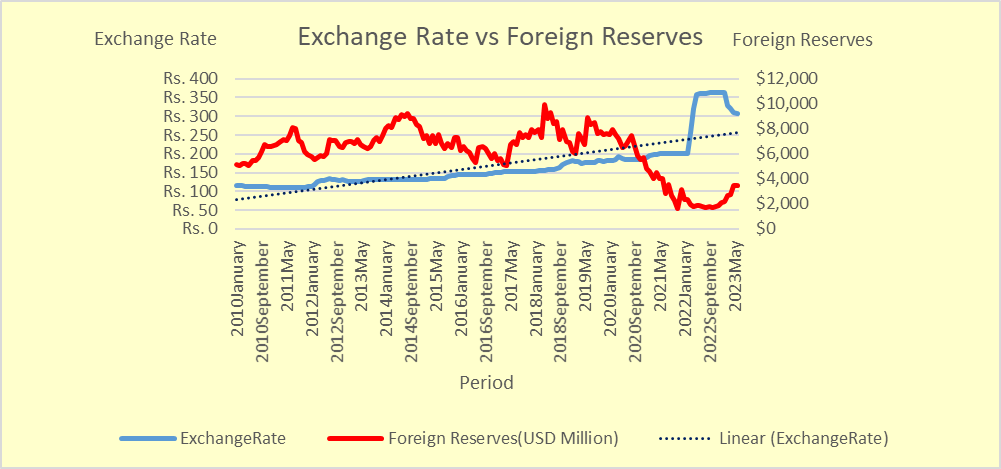
Figure 4 depicts the variation between Exchange Rate and the Foreign Remittance for the period of 2010 to 2023. The correlation coefficient between Exchange Rate and the Foreign Workers Remittance is -0.2724 which emphasizes a weak negative relationship. But in the economic perspective Foreign Workers Remittance has a high impact on emerging economies or developing countries.

**Figure 4: Exchange Rate (USD / LKR) and Foreign Workers Remittance**



1. **Foreign Reserves**

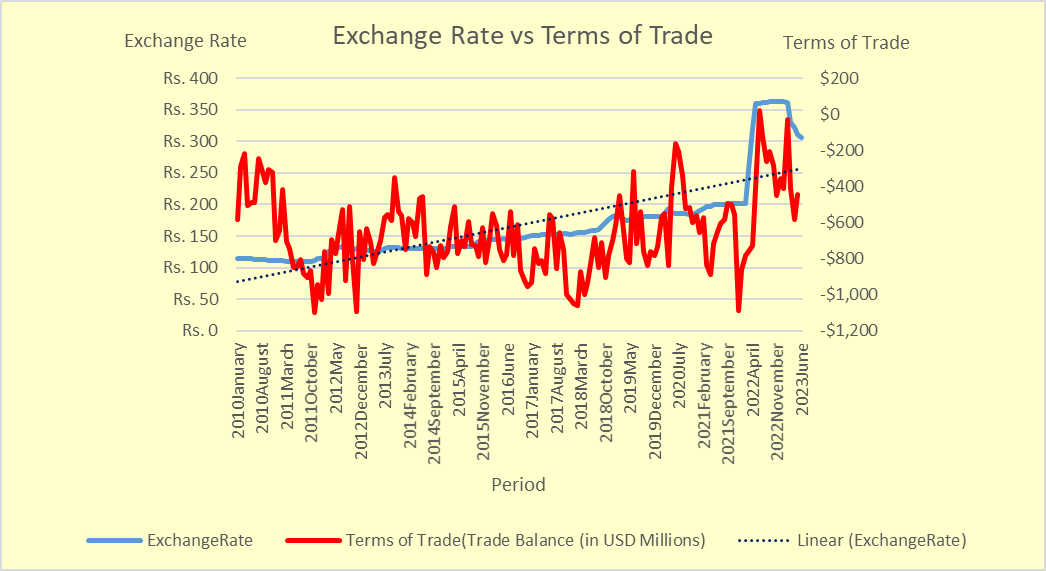
Figure 5 shows the movements of the Exchange Rates against the Foreign Reserves. The correlation between these two variables is -0.7079 which has a strong negative relationship. The negativity is due to the two factors going in two directions which clearly indicates that when the foreign reserves decreases the rates of the foreign exchange rises remarkably.

**Figure 5: Exchange Rate (USD / LKR) and Foreign Reserves**

1. **Terms of Trade (Trade Balance)**

Figure 6 represents the fluctuations of the Exchange Rates and the Terms of Trade or the Trade Balance during the past decade. According to the correlation coefficient it has a 0.4205 score of moderate positive relationship.

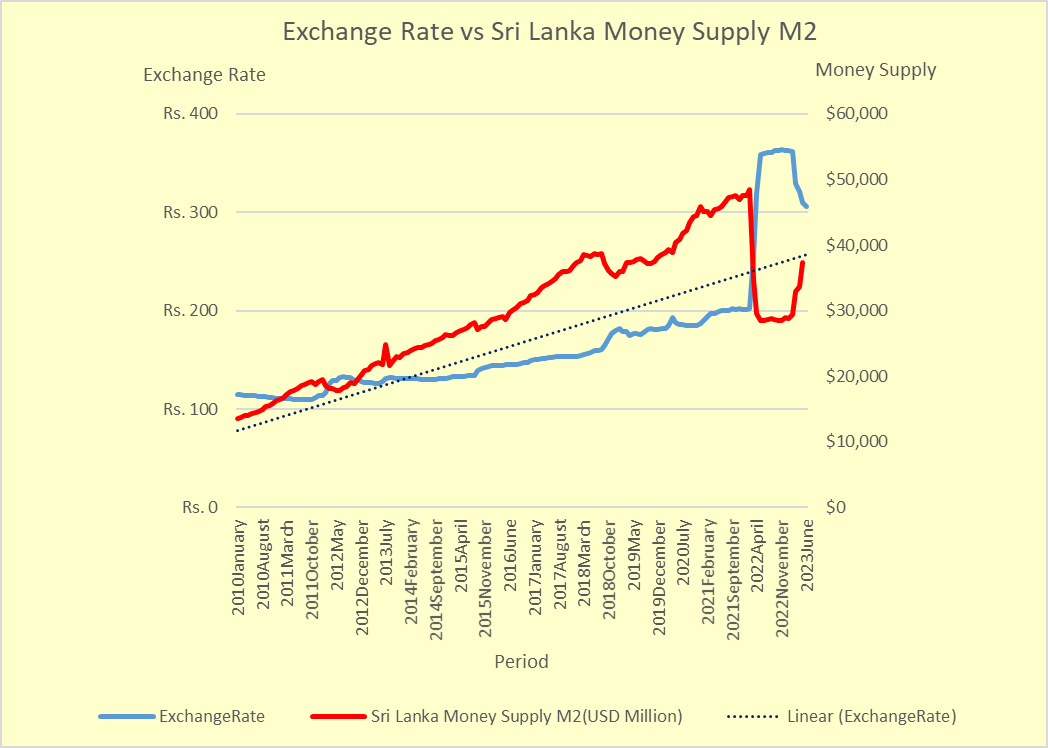
**Figure 6: Exchange Rate (USD / LKR) and Trade Balance**



1. **Sri Lanka Money Supply M2**

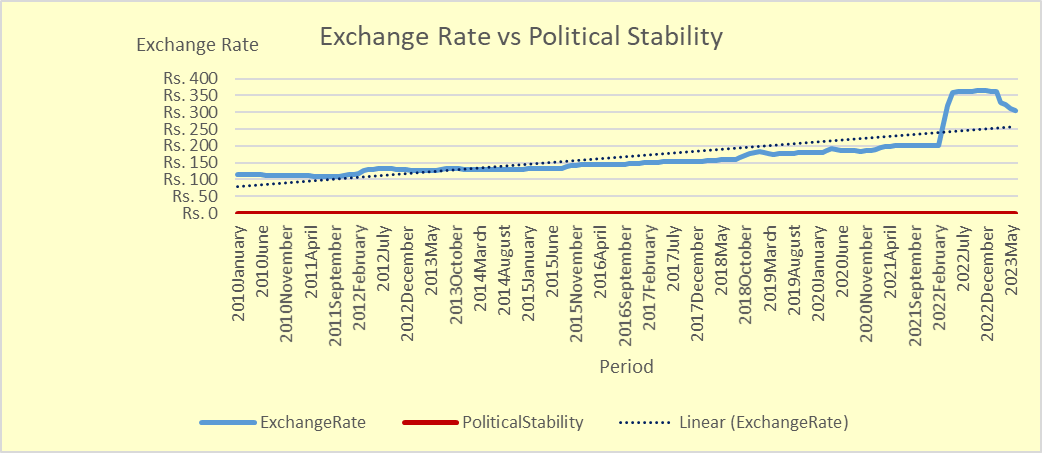
The following chart (Figure 7) shows a positive relationship between Exchange Rate and the Money Supply in (USD Mn) from 2010 to 2023 and the correlation value is 0.4058.

**Figure 7: Exchange Rate (USD / LKR) and Money Supply M2**



1. **Political Stability**

This is a debatable consideration which has a very strong negative correlation with Forex according to this dataset. The dataset was collected from an online repository which does surveys and studies around the world.

**Figure 8: Exchange Rate (USD / LKR) and Political Stability** 

**Forecasting Models**

**Time Series**

Time Series is a collection of continuous data points over time. One of the most important features of a time series is variation. Seasonal, Cyclic, Trend and Irregular fluctuations are the four (4) basic variations. It has equal intervals such as hourly, daily, weekly, monthly, yearly etc. A time series model analyzes time series values and identifies hidden patterns. Finally the model predicts future time series values based on historical data. Most of the macroeconomic variables are non-stationary because time series data are highly dependent on the actual time and do not have constant mean and variance.

* **FB Prophet Model**

FB Prophet is a time series model. Prophet Model is a widely used for forecasting time series which is developed by Facebook's Core Data Science team. It simplifies the forecasting process and provides an intriguing interface for users without much statistical expertise or knowledge.

* **SARIMA**

Seasonal Auto-Regressive Integrated Moving Average (SARIMA) is a time series model which identifies hidden patterns in time series data and makes predictions. The Auto-Regression sub model in the ARIMA model uses past values to make future predictions. Integrated sub model performs differencing to remove any non-stationarity in the time series. Moving Average sub model uses past errors to make a prediction. These sub models are parameters of the overall ARIMA model. Model initializes the parameters using unique notations p,d and q.

* **p: This is the order of the auto-regression sub model. It refers to the number of past values that the model uses to make predictions.**
* **d: This is the number of differencing done to remove the non-stationary components.**
* **Q: This is the order of the moving average sub model. It refers to the number of past errors.**

**(p,d,q) => Non - Seasonal part of the model**

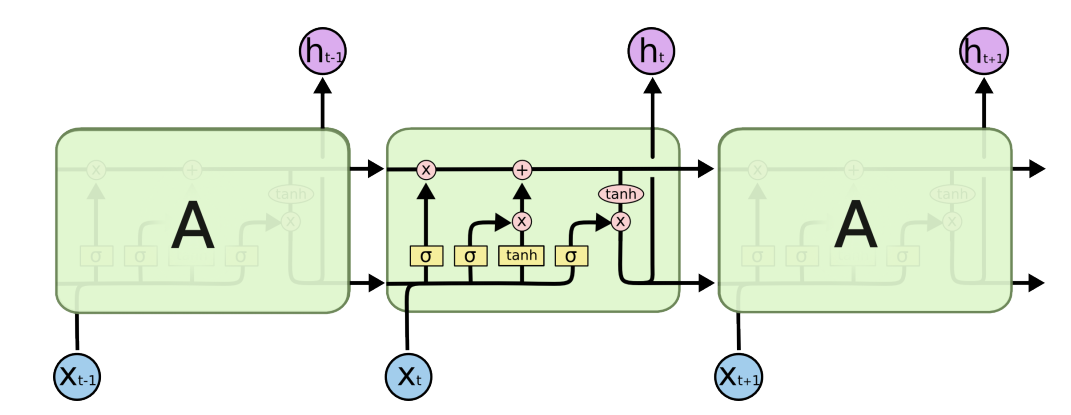
**(P,D,Q)m => Seasonal part of the model**

**m => Number of observations per year**

The Auto-ARIMA library automatically discovers the optimal order of an ARIMA model with stepwise execution of hyperparameters (p,q) and (P,Q) and parallel fitting of models. The generated values are optimized and the model will give accurate forecast results.

* **LSTM**

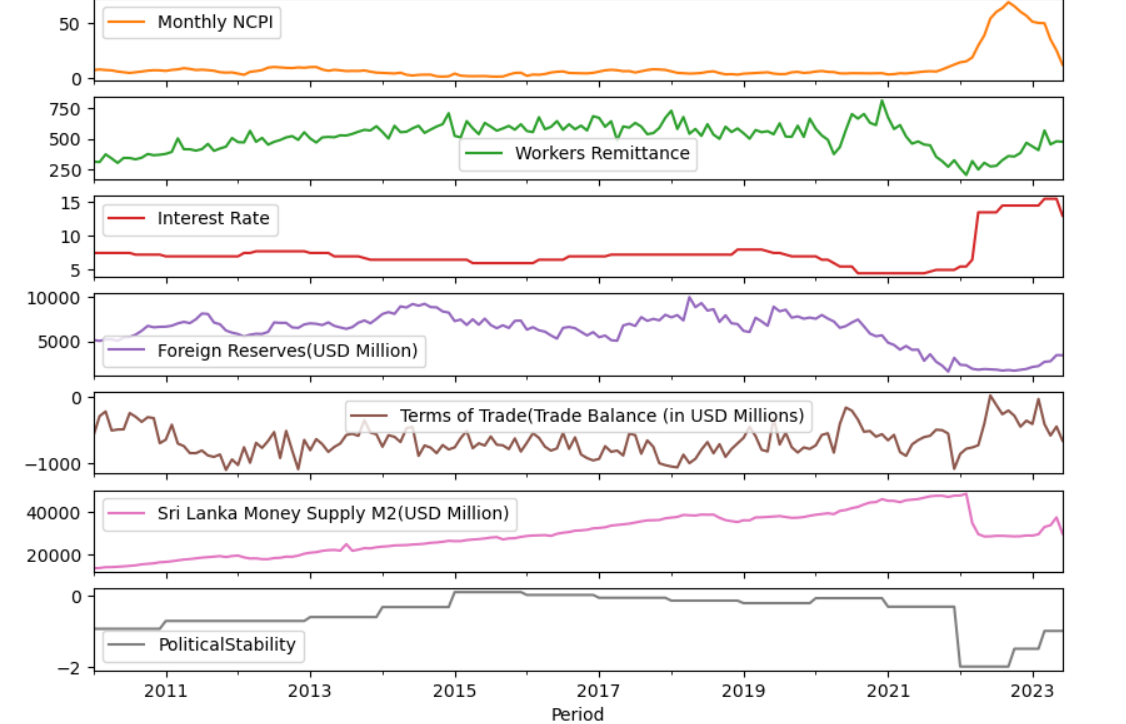
Long Short Term Memory is a special type of Recurrent Neural Network, where information about previous inputs and outputs is used to find complicated patterns in the data. Due to its ability to model complex time series structures with long range dependencies it can perform well when considering financial data.

**Figure 9: Typical Structure of a LSTM Neural Network**

The most important part of the LSTM is the flow in the top of the block which is called Ct. At time t, Ct-1 is taken as an input and an element-wise multiplication with the output of the left-most gate (Forget Gate) is performed. If all values of the forget gate’s output are closer to 1, the gate is open. This will mean that most of Ct-1 is retained at time t. The tanh layer is used to form a memory candidate which after multiplication with the middle gate’s (Input Gate’s) output is added to the part of Ct-1 which is retained to form new memory ct. The new memory is used for

1. The time t output, ht, is made by transforming Ct by an element-wise tanh operation followed by multiplication with the output of the right-most gate (Out Gate).
2. It is used as an input at time t+1. As long as the forget gate remains open through multiple time steps, the memory can be stored for a long period. This is exactly what allows the LSTM to model long-range dependencies [1].

Figure 10 which comes below shows the variance of each individual factor over the period of 2010 and 2023.

**Figure 10: Fluctuations of Other Variables from Jan 2010 - June 2023**

* **Evaluation Techniques**

Model selection is the problem of choosing one model from a set of models. An approach to model selection involves using probabilistic statistical measures that attempt to quantify both the model performance on the training data set and the complexity of the model.

**AIC -** Akaike Information Criterion is a mathematical model for evaluating how well a model fits the data it was generated from. AIC is used to compare different possible models and determine which one is the best fit for the data. AIC is calculated from

* Number of independent variables used to build the model.
* How well the model reproduces the data (maximum likelihood estimate of the model).

The best fit model according to the AIC is the one that explains the greatest amount of variation using the fewest possible independent variables. Especially when exploring a new idea, we want to know which of the independent variables we have measured explain the variation in dependent variable. By creating a set of models, each containing a different combination of the independent variables we have measured is the better approach. These combinations should be based on

* Knowledge on the study domain - avoid using parameters that are not logically connected, (weak or no correlation).
* Experimental design - create a set of procedures to systematically [test a hypothesis](https://www.scribbr.com/statistics/hypothesis-testing/).

The formula for AIC is:



**K - Number of independent variables used. Default is always 2. If one independent variable is used, K =2+1 and so on.**

**Ln (L) - Log-likelihood estimate (how likely the model fits the data)**

Models with lower AIC scores will be the better fit model.

**BIC -** Bayesian Information Criterion is a criterion for model selection of a finite set of models. It is closely related to the AIC. When fitting models, it is possible to increase the likelihood by adding parameters. But doing so may overfit the model. BIC resolves this problem by introducing a penalty term for the number of parameters in the models that the penalty term is higher than in AIC.

The formula for AIC is:

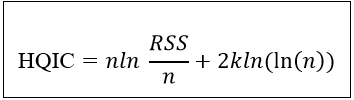
**n - Number of data points**

**Ľ - Maximized value of the likelihood function**

**k - Number of free parameters to be estimated**

Lower BIC value indicates lower penalty terms hence a better model. The only difference is BIC considers the number of observations in the formula, which AIC does not. Though BIC is always higher than AIC, lower the value of these two measures, better the model.

**HQIC -** The Hannan-Quinn Information Criterion is a measure of the goodness of fit of a model and is often used as a criterion for model selection among a finite set of models. It is also related to the AIC. The HQIC introduces a penalty term for the number of parameters in the model. The penalty is larger than one in the AIC. HQIC can be used to compare estimated models only when the numerical values of the dependent variables are identical for all estimates being compared.

The formula for HQIC is:

**n - Number of observations**

**k - Number of model parameters**

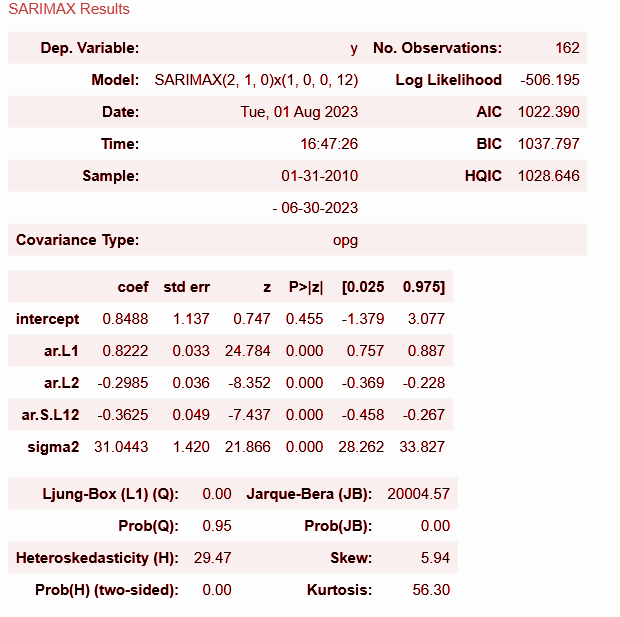
**RSS - Residual sum of squares that results from the model**

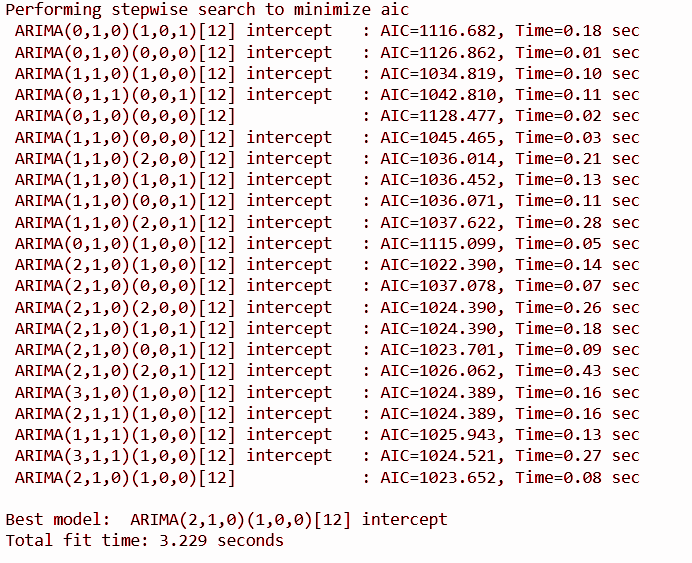
* **Train the data using forecasting models**

**Seasonal Auto Regressive Integrated Moving Average (SARIMA) Model**

To identify whether the time series variables are stationary or non-stationary we used the Augmented Dickey-Fuller (ADF) test. Initial p, q values are set to 0. If the dataset is non-stationary after the ADF test, the auto\_arima() function automatically generates the d value for differencing. If the dataset is stationary, it sets d=0 which means there is no need for differencing. It will run a random search to find the optimal parameters.

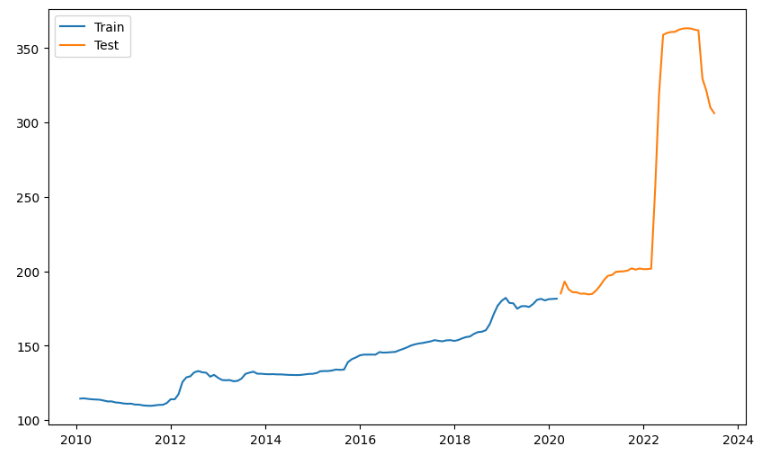
**Figure 11: Model Summary**



According to the above result AIC has the lowest value compared to BIC and HQIC. This is further explained in the section below.

**Figure 12: Stepwise Random Search (AIC) Results for Optimal Parameters**

Data points from 2010-01-31 to 2022-02-29 (75%) used for model training and data points from 2022-02-29 (25%) selected for model testing.

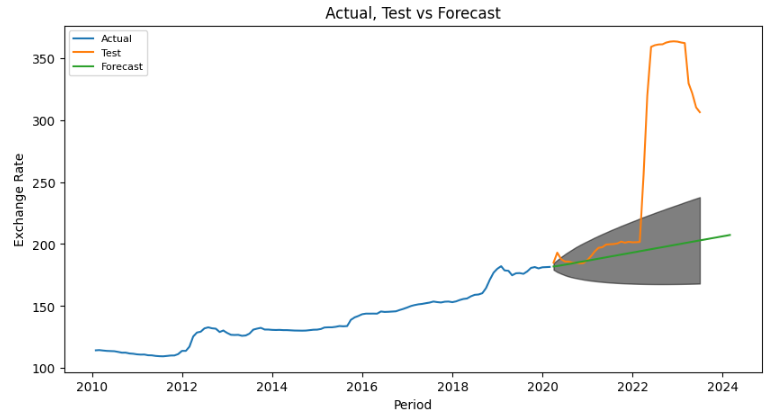
**Figure 13: Train and Test Data Plot for All Records - ARIMA**

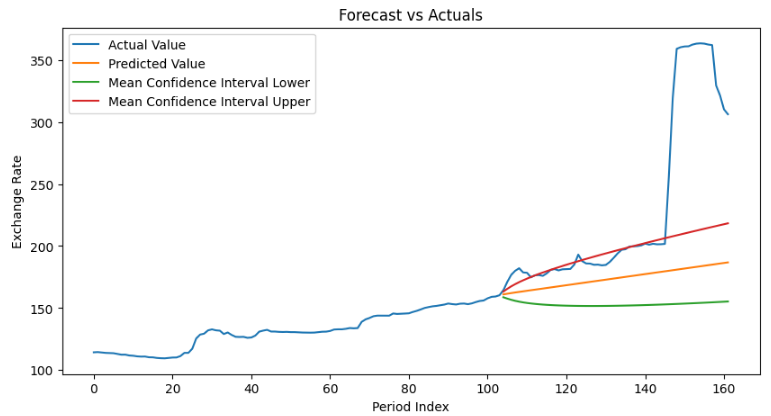
According to Table 2 (Actuals and Forecast with outliers) from March 2020 to Jan 2021 the actual and the predictive Forex have relatively closer values to each other. Since February 2021 the actual forex has gradually increased and had a sudden hike in the month of March 2022 by 54 units. Again in April 2022 and May 2023 the increment is 64 units and 37 units respectively. Therefore the forecasted is varied from the actuals.

**Table 2: Actual and Prediction Forex for 48 Months Since 2020-03-31**

| **Period** | **Actual** | **Range** | | **Forecast** | **Period** | **Actual** | **Range** | | **Forecast** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **2020-03-31** | **185.0567** | **179.1934** | **184.6851** | **181.9393** | **2022-03-31** | **255.8081** | **167.7810** | **221.7428** | **194.7619** |
| **2020-04-30** | **193.0854** | **177.4370** | **187.1276** | **182.2823** | **2022-04-30** | **319.4404** | **167.7326** | **222.8819** | **195.3072** |
| **2020-05-31** | **187.865** | **176.0729** | **189.5629** | **182.8178** | **2022-05-31** | **358.939** | **167.6968** | **224.0087** | **195.8527** |
| **2020-06-30** | **185.9545** | **174.8893** | **191.7671** | **183.3282** | **2022-06-30** | **360.2428** | **167.6726** | **225.1239** | **196.3982** |
| **2020-07-31** | **185.8487** | **173.8738** | **193.7804** | **183.8271** | **2022-07-31** | **360.8724** | **167.6595** | **226.2280** | **196.9437** |
| **2020-08-31** | **184.8939** | **173.0895** | **195.7274** | **184.4084** | **2022-08-31** | **360.9983** | **167.6569** | **227.3217** | **197.4893** |
| **2020-09-30** | **185.0317** | **172.4543** | **197.5804** | **185.0173** | **2022-09-30** | **362.3826** | **167.6642** | **228.4055** | **198.0349** |
| **2020-10-31** | **184.4379** | **171.8568** | **199.2726** | **185.5647** | **2022-10-31** | **363.1484** | **167.6808** | **229.4800** | **198.5804** |
| **2020-11-30** | **184.7882** | **171.2899** | **200.8318** | **186.0609** | **2022-11-30** | **363.394** | **167.7062** | **230.5455** | **199.1259** |
| **2020-12-31** | **187.1822** | **170.8499** | **202.3816** | **186.6158** | **2022-12-31** | **363.1652** | **167.7402** | **231.6026** | **199.6714** |
| **2021-01-31** | **190.4918** | **170.4458** | **203.8526** | **187.1492** | **2023-01-31** | **362.4228** | **167.7822** | **232.6516** | **200.2169** |
| **2021-02-28** | **194.0665** | **170.0911** | **205.2750** | **187.6831** | **2023-02-28** | **361.9766** | **167.8320** | **233.6929** | **200.7624** |
| **2021-03-31** | **196.9825** | **169.7607** | **206.6863** | **188.2235** | **2023-03-31** | **329.4709** | **167.8890** | **234.7268** | **201.3079** |
| **2021-04-30** | **197.423** | **169.4575** | **208.0684** | **188.7629** | **2023-04-30** | **321.4244** | **167.9531** | **235.7538** | **201.8535** |
| **2021-05-31** | **199.5847** | **169.1894** | **209.4269** | **189.3081** | **2023-05-31** | **310.1639** | **168.0240** | **236.7740** | **202.3990** |
| **2021-06-30** | **199.8161** | **168.9490** | **210.7562** | **189.8526** | **2023-06-30** | **306.2612** | **168.1013** | **237.7877** | **202.9445** |
| **2021-07-31** | **199.9778** | **168.7351** | **212.0584** | **190.3967** | **2023-07-31** |  | **168.1848** | **238.7952** | **203.4900** |
| **2021-08-31** | **200.4965** | **168.5484** | **213.3382** | **190.9433** | **2023-08-31** |  | **168.2743** | **239.7968** | **204.0356** |
| **2021-09-30** | **201.9858** | **168.3853** | **214.5961** | **191.4907** | **2023-09-30** | **168.3695** | **240.7926** | **204.5811** |
| **2021-10-31** | **201.0899** | **168.2414** | **215.8312** | **192.0363** | **2023-10-31** | **168.4703** | **241.7829** | **205.1266** |
| **2021-11-30** | **201.8582** | **168.1152** | **217.0455** | **192.5804** | **2023-11-30** | **168.5764** | **242.7679** | **205.6721** |
| **2021-12-31** | **201.399** | **168.0086** | **218.2437** | **193.1262** | **2023-12-31** | **168.6876** | **243.7477** | **206.2177** |
| **2022-01-31** | **201.4647** | **167.9179** | **219.4248** | **193.6713** | **2024-01-31** | **168.8038** | **244.7225** | **206.7632** |
| **2022-02-28** | **201.7362** | **167.8425** | **220.5905** | **194.2165** | **2024-02-29** | **168.9248** | **245.6926** | **207.3087** |

**Figure 14: Actual, Test and Forecasting Data Plots for All Records- ARIMA**

Figure 15 exhibits the upper and lower mean confidence distribution of the forecast with 95% confidence interval.

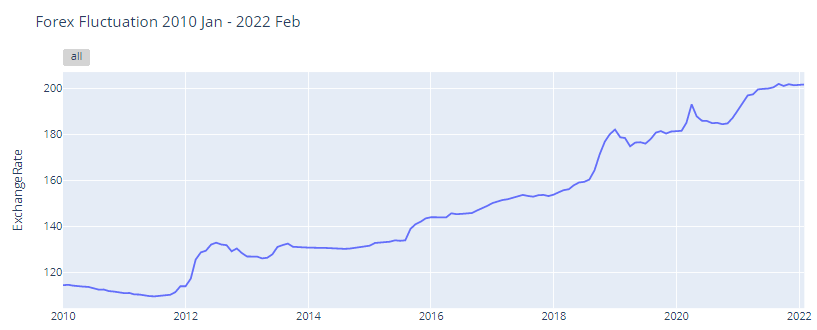
**Figure 15: Actual, Forecasting Mean Confidence Plots for All Records - ARIMA**

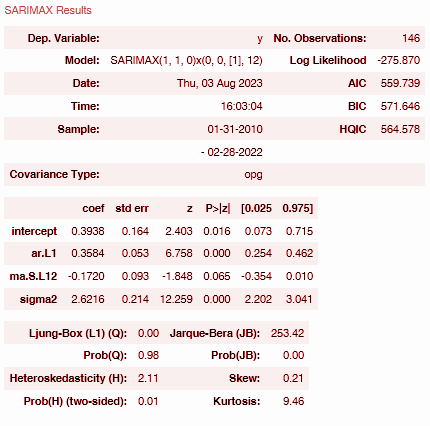
According to the gathered data we could clearly identify that Forex has skyrocketed since March 2022 (Table 3). These unexpected outliers in the dataset, the prediction would be in debate. According to the previous studies, recommendations are given as leave out the outliers and retest the models. We, too, followed that concept and removed data points from March 2022. Retested the model with data points from 2010-01-31 to 2018-10-31 and trained the model with the data points from 2018-11-30.

**Table 3: Forex Variation During the Years of Jan 2010 to June 2023**

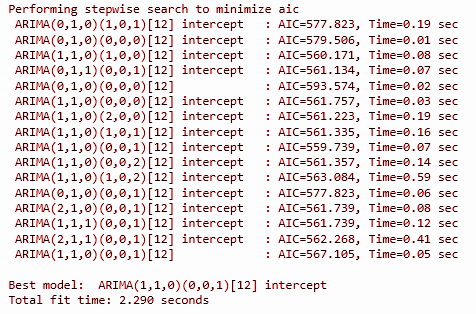
| **Year** | **Start Forex(Jan)** | **End Forex(Dec)** | **Difference** |
| --- | --- | --- | --- |
| **2010** | **114** | **111** | **-3** |
| **2011** | **110** | **113** | **3** |
| **2012** | **113** | **128** | **15** |
| **2013** | **126** | **130** | **4** |
| **2014** | **130** | **131** | **1** |
| **2015** | **131** | **143** | **12** |
| **2016** | **143** | **148** | **5** |
| **2017** | **150** | **153** | **3** |
| **2018** | **153** | **180** | **27** |
| **2019** | **182** | **181** | **-1** |
| **2020** | **181** | **187** | **6** |
| **2021** | **190** | **201** | **11** |
| **2022** | **201** | **363** | **62** |
| **2023** | **362** | **306 (June)** | **-56** |

**Figure 16: Fluctuations in Foreign Exchange Rate from Jan 2010 - Feb 2022 without Outliers**

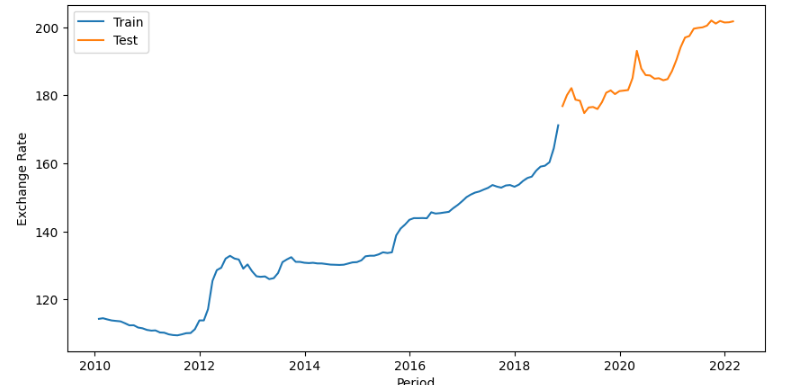


**Figure 17: Model Summary Without Outliers**

Above summary result clearly indicates that the model is improved without outliers. AIC, BIC, HQIC values are reduced approximately by fifty percent.

**Figure 18: Stepwise Random Search (AIC) Results for Optimal Parameters without Outliers**

**Figure 19: Train and Test Data Plot for Dataset Without Outliers - ARIMA**



**Table 4: Actual and Prediction Forex for 48 Months Since 2018-11-30 Without Outliers**

| **Period** | **Actual** | **Range** | | **Forecast** | **Period** | **Actual** | **Range** | | **Forecast** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **2018-11-30** | **176.8496** | **172.6127** | **177.6434** | **175.1281** | **2020-11-30** | **184.7882** | **167.7108** | **220.6180** | **194.1644** |
| **2018-12-31** | **180.098** | **172.8576** | **182.0707** | **177.4642** | **2020-12-31** | **187.1822** | **167.7321** | **221.8114** | **194.7717** |
| **2019-01-31** | **182.1277** | **172.5242** | **185.5121** | **179.0181** | **2021-01-31** | **190.4918** | **167.7657** | **222.9922** | **195.3789** |
| **2019-02-28** | **178.7331** | **171.9750** | **188.3191** | **180.1470** | **2021-02-28** | **194.0665** | **167.8110** | **224.1614** | **195.9862** |
| **2019-03-31** | **178.4298** | **171.3661** | **190.7018** | **181.0339** | **2021-03-31** | **196.9825** | **167.8672** | **225.3196** | **196.5934** |
| **2019-04-30** | **174.8012** | **170.7649** | **192.7919** | **181.7784** | **2021-04-30** | **197.423** | **167.9339** | **226.4674** | **197.2006** |
| **2019-05-31** | **176.4447** | **170.2681** | **194.7424** | **182.5052** | **2021-05-31** | **199.5847** | **168.0103** | **227.6054** | **197.8079** |
| **2019-06-30** | **176.5889** | **169.8129** | **196.5361** | **183.1745** | **2021-06-30** | **199.8161** | **168.0961** | **228.7341** | **198.4151** |
| **2019-07-31** | **175.9902** | **169.3861** | **198.1956** | **183.7909** | **2021-07-31** | **199.9778** | **168.1906** | **229.8540** | **199.0223** |
| **2019-08-31** | **177.9705** | **169.0459** | **199.8067** | **184.4263** | **2021-08-31** | **200.4965** | **168.2936** | **230.9655** | **199.6295** |
| **2019-09-30** | **180.7752** | **168.8683** | **201.4670** | **185.1677** | **2021-09-30** | **201.9858** | **168.4045** | **232.0690** | **200.2368** |
| **2019-10-31** | **181.4511** | **168.8358** | **203.1757** | **186.0058** | **2021-10-31** | **201.0899** | **168.5230** | **233.1649** | **200.8440** |
| **2019-11-30** | **180.3729** | **168.7095** | **204.7632** | **186.7363** | **2021-11-30** | **201.8582** | **168.6489** | **234.2535** | **201.4512** |
| **2019-12-31** | **181.2539** | **168.5501** | **206.2689** | **187.4095** | **2021-12-31** | **201.399** | **168.7817** | **235.3352** | **202.0584** |
| **2020-01-31** | **181.4036** | **168.3876** | **207.7162** | **188.0519** | **2022-01-31** | **201.4647** | **168.9211** | **236.4102** | **202.6656** |
| **2020-02-29** | **181.5628** | **168.2365** | **209.1193** | **188.6779** | **2022-02-28** | **201.7362** | **169.0670** | **237.4788** | **203.2729** |
| **2020-03-31** | **185.0567** | **168.1031** | **210.4873** | **189.2952** | **2022-03-31** | **255.8081** | **169.2190** | **238.5412** | **203.8801** |
| **2020-04-30** | **193.0854** | **167.9897** | **211.8259** | **189.9078** | **2022-04-30** | **319.4404** | **169.3769** | **239.5978** | **204.4873** |
| **2020-05-31** | **187.865** | **167.8966** | **213.1392** | **190.5179** | **2022-05-31** | **358.939** | **169.5404** | **240.6486** | **205.0945** |
| **2020-06-30** | **185.9545** | **167.8230** | **214.4303** | **191.1266** | **2022-06-30** | **360.2428** | **169.7095** | **241.6940** | **205.7018** |
| **2020-07-31** | **185.8487** | **167.7680** | **215.7014** | **191.7347** | **2022-07-31** | **360.8724** | **169.8838** | **242.7342** | **206.3090** |
| **2020-08-31** | **184.8939** | **167.7304** | **216.9543** | **192.3423** | **2022-08-31** | **360.9983** | **170.063** | **243.7692** | **206.9162** |
| **2020-09-30** | **185.0317** | **167.7091** | **218.1905** | **192.9498** | **2022-09-30** | **362.3826** | **170.247** | **244.7994** | **207.5234** |
| **2020-10-31** | **184.4379** | **167.7029** | **219.4114** | **193.5572** | **2022-10-31** | **363.1484** | **170.436** | **245.8247** | **208.1307** |

Table 4 (Actuals and Forecast without outliers) shows the forecasting is gradually improving for the test data, both values are closer till March 2022.

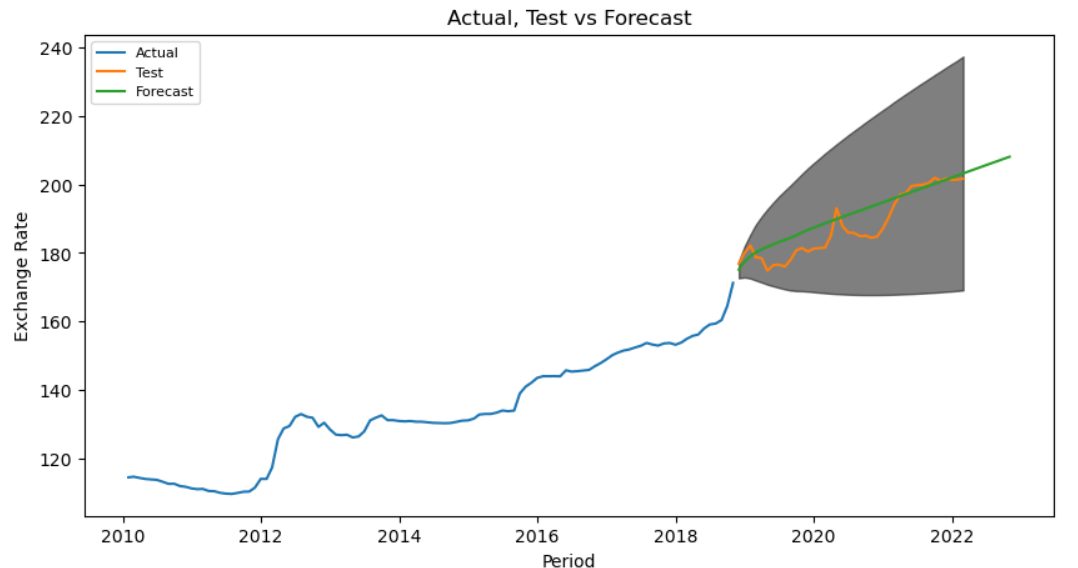
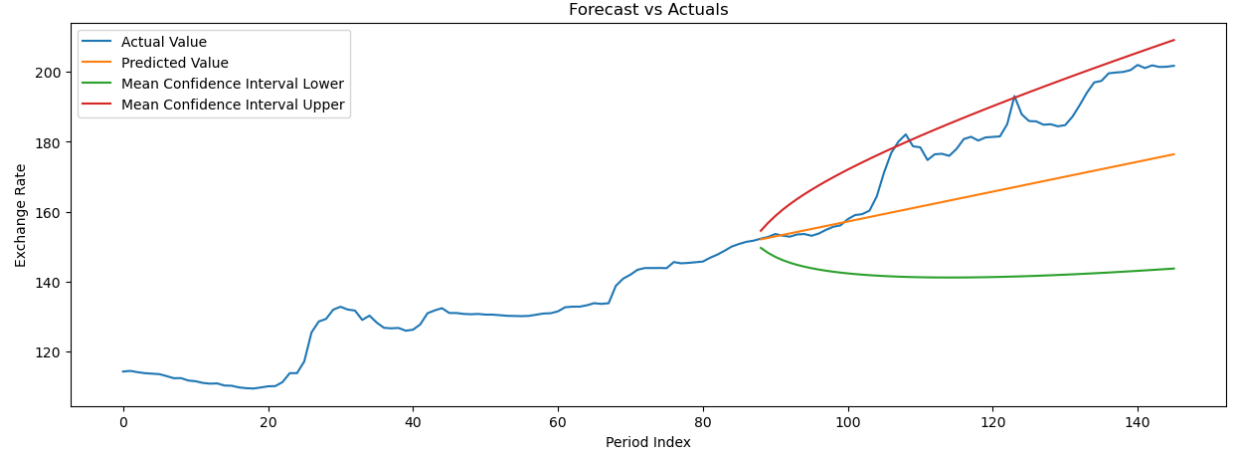
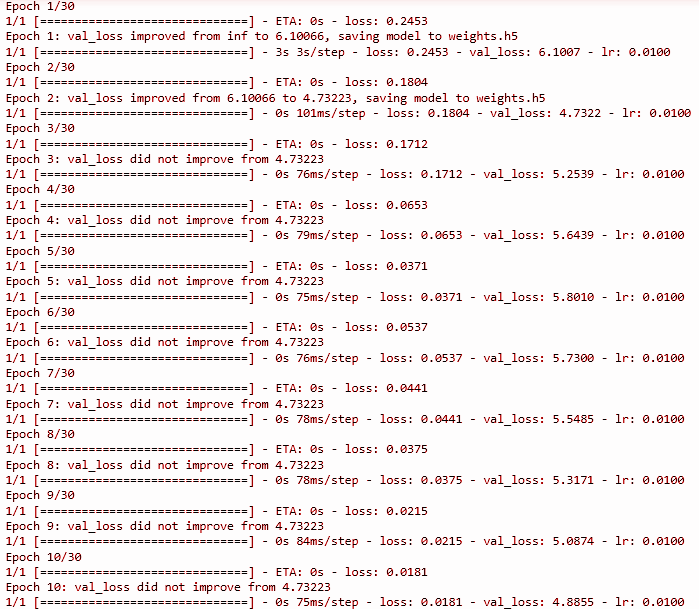
**Figure 20 : Actual, Test and Forecasting Data Plots without Outliers - ARIMA**

Figure 21 exhibits the upper and lower mean confidence distribution of the forecast with 95% confidence interval for the dataset without outliers.

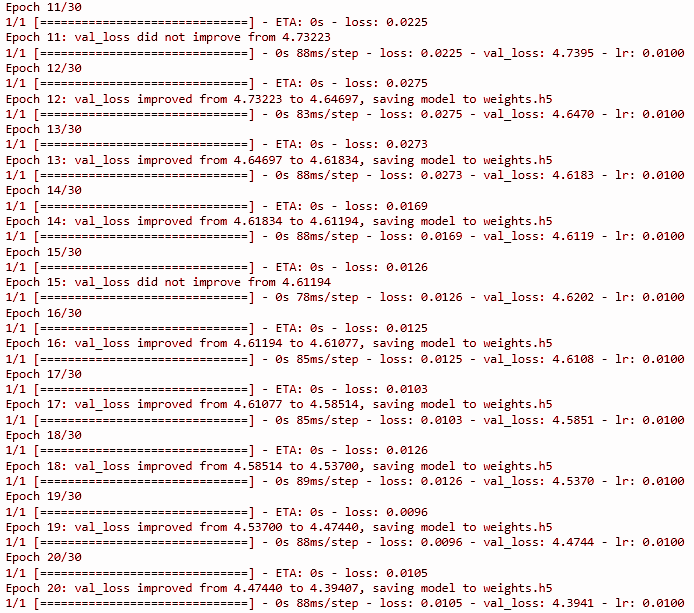
**Figure 21: Actual, Forecasting Mean Confidence Plots without Outliers - ARIMA**

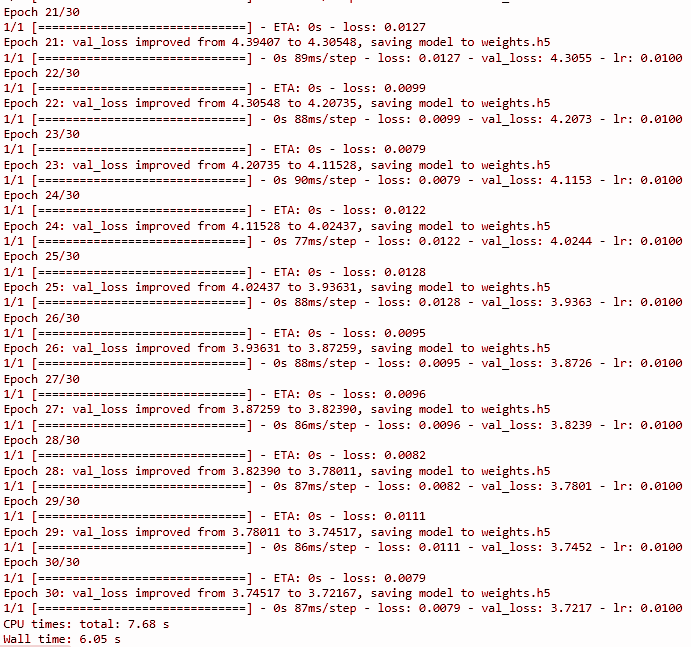
**Long Short Term Memory (LSTM) Model**

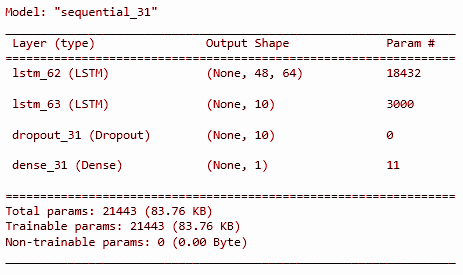
Model used 30 epochs and stopped the training when the monitored metric stopped improving for ten (10) consecutive steps. Shuffled the training dataset before each epoch to increase the performance of the model with 256 sample batch sizes.

**Figure 22: LSTM Model Training Scores All Dataset**

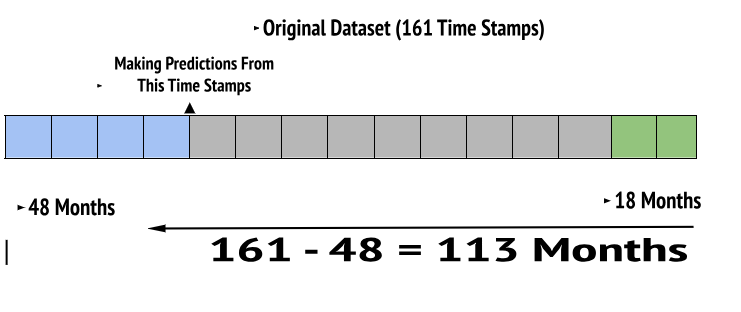
In this we trained the entire dataset and used 48 months data to forecast a 18 months period.



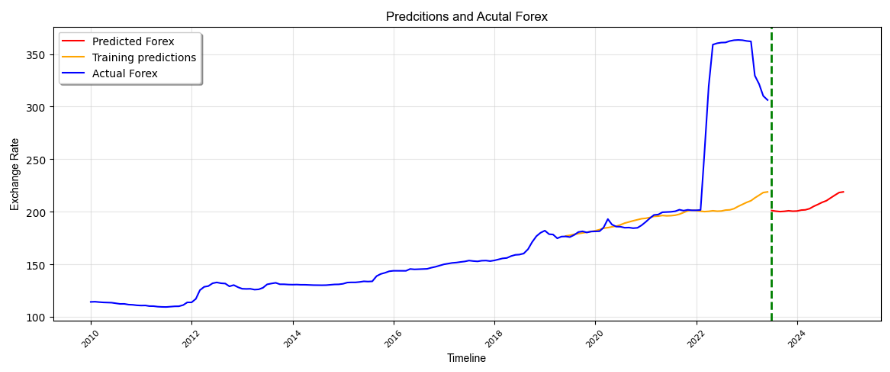


**Figure 23: LSTM Model Summary All Dataset**

**Figure 24 : Data Separation for Testing and Forecasting for All Data - LSTM**



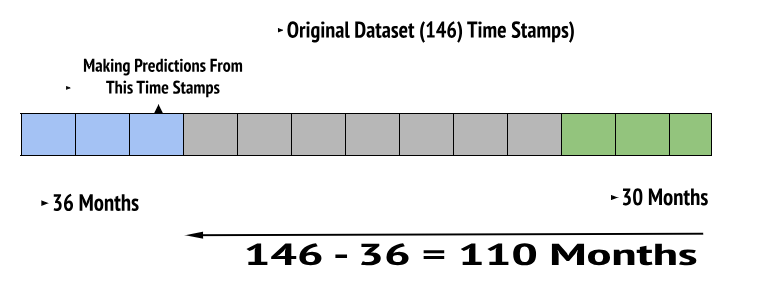
Data plot in figure 25 exhibits the training prediction and the forecast follows the continuity of actual slope.

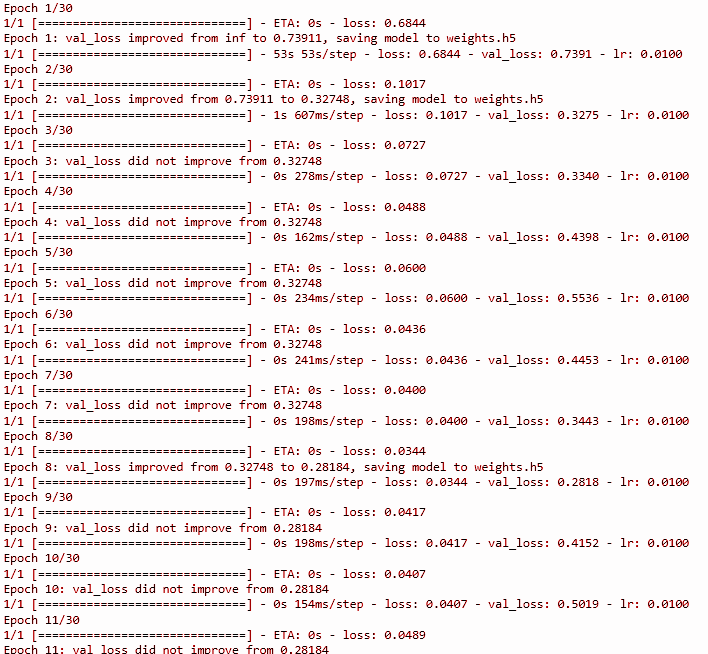
**Figure 25: Actual, Test and Forecasting Data Plots All Data - LSTM**

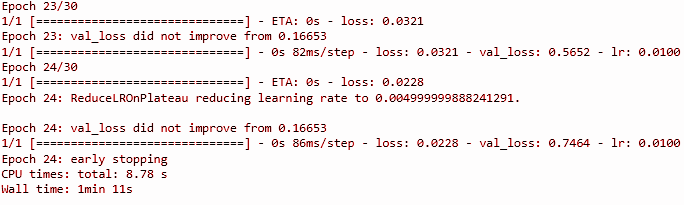
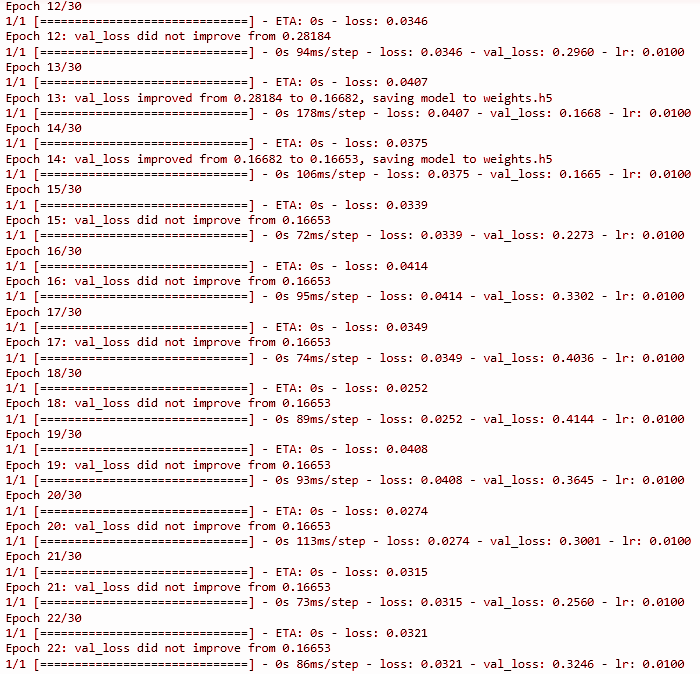
According to the results in Table 5 it clearly shows that this model more accurately predicted the forecast values which are very close to the actuals up to March 2022. It seems, forecasting follows the same continuity for the remaining period.

**Table 5: Actual and Prediction Forex for All Dataset - LSTM**

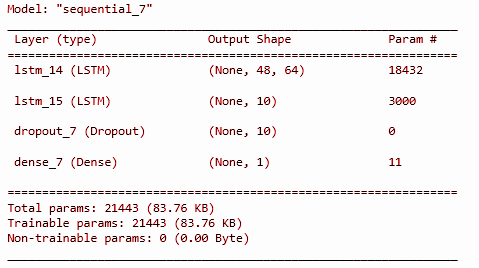
| **Period** | **Actual** | **Forecast** | **Period** | **Actual** | **Forecast** |
| --- | --- | --- | --- | --- | --- |
| **2019-06-01** | **176.5889** | **177.0419** | **2021-07-01** | **199.9778** | **196.3029** |
| **2019-07-01** | **175.9902** | **177.6842** | **2021-08-01** | **200.4965** | **196.7200** |
| **2019-08-01** | **177.9705** | **178.6239** | **2021-09-01** | **201.9858** | **197.7087** |
| **2019-09-01** | **180.7752** | **179.2799** | **2021-10-01** | **201.0899** | **199.2980** |
| **2019-10-01** | **181.4511** | **179.9376** | **2021-11-01** | **201.8582** | **200.9567** |
| **2019-11-01** | **180.3729** | **180.5690** | **2021-12-01** | **201.399** | **201.5498** |
| **2019-12-01** | **181.2539** | **181.2874** | **2022-01-01** | **201.4647** | **201.1846** |
| **2020-01-01** | **181.4036** | **182.0116** | **2022-02-01** | **201.7362** | **200.5910** |
| **2020-02-01** | **181.5628** | **183.0928** | **2022-03-01** | **255.8081** | **200.1850** |
| **2020-03-01** | **185.0567** | **184.4202** | **2022-04-01** | **319.4404** | **200.4528** |
| **2020-04-01** | **193.0854** | **184.9689** | **2022-05-01** | **358.939** | **200.9805** |
| **2020-05-01** | **187.865** | **185.7216** | **2022-06-01** | **360.2428** | **200.5777** |
| **2020-06-01** | **185.9545** | **186.3796** | **2022-07-01** | **360.8724** | **200.7310** |
| **2020-07-01** | **185.8487** | **187.6367** | **2022-08-01** | **360.9983** | **201.6122** |
| **2020-08-01** | **184.8939** | **189.2792** | **2022-09-01** | **362.3826** | **201.8609** |
| **2020-09-01** | **185.0317** | **190.3779** | **2022-10-01** | **363.1484** | **202.9828** |
| **2020-10-01** | **184.4379** | **191.4594** | **2022-11-01** | **362.4228** | **205.2007** |
| **2020-11-01** | **184.7882** | **192.4426** | **2022-12-01** | **363.1652** | **207.0413** |
| **2020-12-01** | **187.1822** | **193.4044** | **2023-01-01** | **362.4228** | **208.9864** |
| **2021-01-01** | **190.4918** | **193.8103** | **2023-02-01** | **361.9766** | **210.5523** |
| **2021-02-01** | **194.0665** | **194.6680** | **2023-03-01** | **329.4709** | **213.2360** |
| **2021-03-01** | **196.9825** | **195.6331** | **2023-04-01** | **321.4244** | **215.7986** |
| **2021-04-01** | **197.423** | **195.8878** | **2023-05-01** | **310.1639** | **218.3473** |
| **2021-05-01** | **199.5847** | **196.4792** | **2023-06-01** | **306.2612** | **218.9414** |
| **2021-06-01** | **199.8161** | **196.1606** |  |  |  |

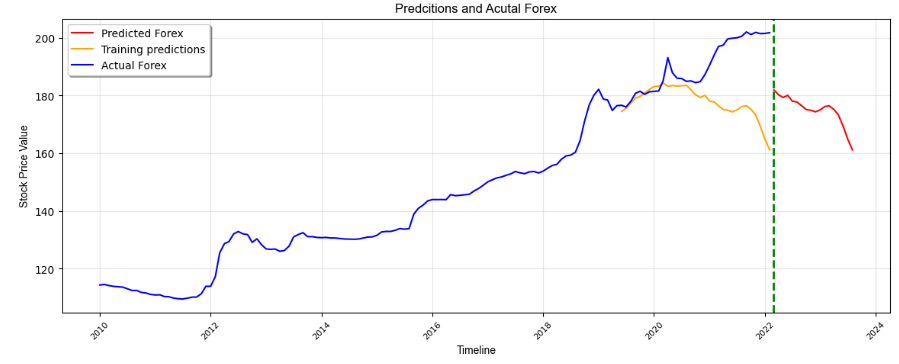
**Figure 26: Data Separation for Testing and Forecasting Without Outliers - LSTM**

**Figure 27: LSTM Model Training Scores Without Outliers**

In this we trained the entire dataset and used 36 months data to forecast a 30 months period.

**Figure 28: LSTM Model Summary Without Outliers**



**Figure 29: Actual, Test and Forecasting Data Plots Without Outliers - LSTM**

**Table 6: Actual and Forecast Forex Without Outliers - LSTM**

| **Period** | **Actual** | **Forecast** | **Period** | **Actual** | **Forecast** |
| --- | --- | --- | --- | --- | --- |
| **2019-06-01** | **176.5889** | **174.4566** | **2020-11-01** | **184.7882** | **179.2912** |
| **2019-07-01** | **175.9902** | **175.5530** | **2020-12-01** | **187.1822** | **180.0335** |
| **2019-08-01** | **177.9705** | **177.3139** | **2021-01-01** | **190.4918** | **178.0332** |
| **2019-09-01** | **180.7752** | **179.0408** | **2021-02-01** | **194.0665** | **177.7407** |
| **2019-10-01** | **181.4511** | **179.6577** | **2021-03-01** | **196.9825** | **176.4133** |
| **2019-11-01** | **180.3729** | **180.8862** | **2021-04-01** | **197.423** | **175.1115** |
| **2019-12-01** | **181.2539** | **182.0854** | **2021-05-01** | **199.5847** | **174.8459** |
| **2020-01-01** | **181.4036** | **183.0807** | **2021-06-01** | **199.8161** | **174.3064** |
| **2020-02-01** | **181.5628** | **183.1800** | **2021-07-01** | **199.9778** | **174.9980** |
| **2020-03-01** | **185.0567** | **184.3487** | **2021-08-01** | **200.4965** | **76.1678** |
| **2020-04-01** | **193.0854** | **183.1952** | **2021-09-01** | **201.9858** | **176.4264** |
| **2020-05-01** | **187.865** | **183.4294** | **2021-10-01** | **201.0899** | **175.1761** |
| **2020-06-01** | **185.9545** | **183.2416** | **2021-11-01** | **201.8582** | **173.2518** |
| **2020-07-01** | **185.8487** | **183.3709** | **2021-12-01** | **201.399** | **169.3731** |
| **2020-08-01** | **184.8939** | **183.5631** | **2022-01-01** | **201.4647** | **164.9127** |
| **2020-09-01** | **185.0317** | **181.9300** | **2022-02-01** | **201.7362** | **161.1834** |
| **2020-10-01** | **184.4379** | **180.2569** |  | | |

**FB Prophet Model with Multivariate Dataset**

Before implementing the model and predicting the exchange rate. As an initial step we checked whether the time series is stationary or non-stationary using the ADF test and the obtained results proved it to be a non stationary time series.

**Results of ADF-Statistic Test**

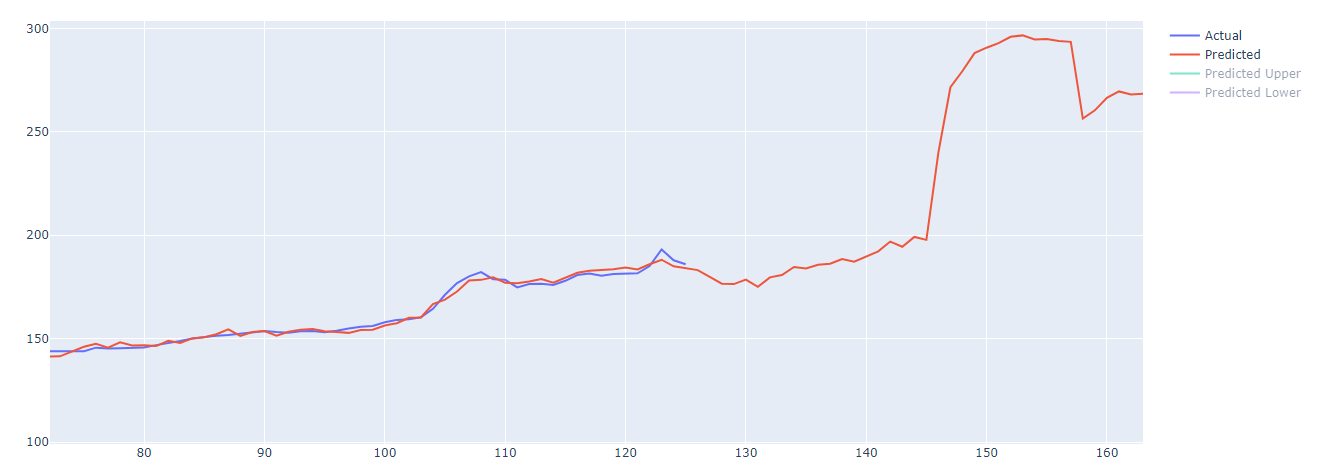
ADF p-value: 0.9975, Accepting the alternative hypothesis i.e. The time series is likely non-stationary based on the ADF test.

As FB Prophet is one of the best and efficient time based forecasting model we used it for the above problem as it has the ability to smooth the non stationary time series without any extra enhancements. We used 80% of the dataset as training data and remaining as testing data. We added all the remaining features other than the Period variable and the dependent variable (ExchangeRate) as regressor in order to predict their values based on the training data for the future dates which in return used to predict the ExchangeRate.

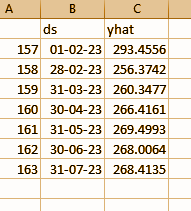
**Figure 30: Regressors and their Effects to the Final Forecast of the Model**



We predicted the entire 162 months of data using the actual train dataset of 126 months and remaining 32 months from the test dataset was used to evaluate the accuracy of the prophet model. The next 6 months of the future (till 31-07-23) from the current available dataset’s final date (01-02-23) was predicted as an output for the project’s solution findings. The entire graph of prediction with the actual train dataset is given in the figure below.

**Figure 31 : Actual and Prediction of Exchange Rate for Upcoming Months**

**Table 7: Forex Forecast - FBProphet**

The predicted future value by the model is given by the column name ‘yhat’ corresponding to the period given as ‘ds’. If we compare the current real value of USD to LKR this might differ because of the government policies that change from time to time regularly. The model mostly incorporated the measures that were relevant till Feb 2023 which is being reflected in the past 6 months results.

**FB Prophet Model Evaluation**

We evaluated the above model using both the training dataset and testing dataset separately and the following results were obtained.

**Training Dataset Evaluation Analysis**

Correlation Coefficient between Predicted Exchange Rate vs Actual Exchange Rate - 0.9969

MAE between Predicted Exchange Rate vs Actual Exchange Rate - 1.4146

RMSE between Predicted Exchange Rate vs Actual Exchange Rate - 1.7694

**Testing Dataset Evaluation Analysis**

Correlation Coefficient between Predicted Exchange Rate vs Actual Exchange Rate - 0.9951

MAE between Predicted Exchange Rate vs Actual Exchange Rate - 29.5537

RMSE between Predicted Exchange Rate vs Actual Exchange Rate - 40.8229

This shows that the model performance is very good in both scenarios. Particularly the model performs well with the train dataset compared to the test dataset.

**Python libraries used for the entire project**

* **Numpy, Pandas, Matplotlib, Plotly, Keras, Sklearn, PMDArima, Math and Statsmodels**

**Table 8: Model Evaluation Scores**

| **Model** | **MSE**  **(Mean Squared Error)** | **RMSE**  **(Root Mean Squared Error)** | **MAE**  **(Mean Absolute Error)** |
| --- | --- | --- | --- |
| **ARIMA All Records** | **10634.9987** | **103.1261** | **70.8763** |
| **ARIMA Without Outliers** | **247.6217** | **15.7360** | **12.8224** |
| **LSTM All Records** | **0.7962** | **0.8923** | **0.3881** |
| **LSTM Without Outliers** | **0.0713** | **0.2671** | **0.1614** |
| **FB Prophet (Train)**  **All Records** | **3.1308** | **1.7694** | **1.4146** |
| **FB Prophet (Test)**  **All Records** | **1666.5092** | **40.8229** | **29.5537** |

**Discussion**

There were no exactly similar previous studies to this topic. Multivariate analysis has been used for stock market price prediction and also based on different features. We followed these concepts to build our own model for forecasting Forex USD/LKR.

Selected models are sophisticated, well understood and effective on many previous problems if data are suitably prepared and methods are well configured. Traditional time series models have problems with generalization from a single study, requiring domain based identification of the actual parameters.

Large movements in the dataset can be seen as unreal movements since the end of 2020. Since the size of these movements are relatively large compared to what is observed during more liquid trading periods from 2010. Generally, financial markets are difficult to predict but most models will have an easy time detecting the pattern. The movements are so predictable, most models will actually seem to make an inferior fit as the data movements are unexpected.

One of the main difficulties of modeling is figuring out how to utilize our data in the best possible way. If we choose to use all our data set to construct the model, we would not have any additional data for testing. Therefore, we would have to use a part of it as the learning set for evaluating the models. When evaluating the model, a very good model would be produced, but when testing on unseen data, it would very likely perform poorly, which is known as Overfitting.

Predicting Forex in countries like Sri Lanka is arduous due to the irregular fluctuations in social and economical factors regularly. We can simply forecast Forex as a univariate feature over dates, months or quarters using different Deep Machine Learning techniques using the existing Forex data. But when incorporating the factors affecting on deciding the Forex required better understanding on macroeconomic factors and non-macroeconomic factors. These are loosely or tightly coupled to each other like links in a food chain. Therefore a better understanding is required of those factors affecting or behaving on an economic model when modeling a forecast. When considering the six (6) selected macroeconomic factors, inflation, interest rates, foreign remittance, foreign reserves, money supply and terms of trades they are all related to each other. For example foreign reserves are used to back liabilities and influence monetary policy. To balance such situations central banks tend to increase the interest rates to attract more foreign currency.

Further, political stability, the non-macroeconomic factors, highly affects macroeconomics as we experienced in the recent past. This is like a coincidence and due to that reason political stability is considered separately in previous research. Therefore a different model predicting the political stability index and using that dataset could be more accurate in predicting Forex.

****

We can formulate such an equation for forecasting Forex. **β** values are the intercept and slope coefficient parameters and ε is the total adjustments.

**Conclusion**

According to the results we can conclude that all the models are forecasting fairly in an accurate manner even though the factors are debatable. We can further evaluate the models by first forecasting the Forex without any factors and then gradually adding and testing the factors to see the forecasting behavior. Further, if we could utilize daily basis data instead of monthly basis, the models will perform more accurately.

As a comprehensive project we could design and develop a data repository to store the data using Data Warehouse and Big Data technologies. Furthermore, working on the model enhancement and prediction accuracy through various technical means can be done to improve our project capacity and preciseness which might in future be a game changing predictive model that can be used for real world scenarios.

**References**

[1] L. G. Nielsen, "MACHINE LEARNING FOR FOREIGN EXCHANGE RATE FORECASTING," Imperial College London, 2018.

[2] AIP Conference Proceedings 2459, 030024 (2022); https://doi.org/10.1063/5.0083533

Published Online: 05 April 2022

[3] CFI Team, “Exchange Rate,” Corporate Finance Institute, Dec. 12, 2022. <https://corporatefinanceinstitute.com/resources/economics/exchange-rate/>/

[4] A. Twin, “6 Factors that Influence Exchange Rates,” Investopedia, Jun. 08, 2021. <https://www.investopedia.com/trading/factors-influence-exchange-rates/>

[5] W. Christensen, "Model Selection Using Information Criteria (Made Easy in SAS®)," in University of California, California, 2587-2018.

[6] Jason Brownlee, Multivariate Time Series Forecasting with LSTMs in Keras, 2020, Machine Learning Mastery, accessed 20th July 2023,<<https://machinelearningmastery.com/multivariate-time-series-forecasting-lstms-keras/>>

[7] F. J. V. Ari Aisen, "How Does Political Instability Affect Economic Growth?," IMF Working Paper , 2011.

[8] T. H. A. I. K. C. R. Najia Maraoui, "How Economic, Political and Institutional Factors Influence the Choice of Exchange Rate Regimes? New Evidence from Selected Countries of the MENA Region," 2022.

[9] Central Bank of Sri Lanka, accessed 2023, <<https://www.cbsl.lk/eresearch/>>.

[10] World Bank, accessed 2023, <<https://data.worldbank.org/>>.

[11] Macrotrends. Accessed 2023, <<https://www.macrotrends.net/>>.

[12] S. Rajakaruna, "An Investigation on Factors affecting Exchange Rate Fluctuations in Sri Lanka," Central Bank of Sri Lanka – Staff Studies, Colombo, 2017.

[13] The Global Economy, accessed July 2023, <[https://www.theglobaleconomy.com/Sri-Lanka](https://www.theglobaleconomy.com/Sri-Lanka/wb_political_stability/)/>

[14] Yıldırım, D.C., Toroslu, I.H. & Fiore, U. Forecasting directional movement of Forex data using LSTM with technical and macroeconomic indicators. Finance Innov 7, 1 (2021). <https://doi.org/10.1186/s40854-020-00220-2>

[15] Data Science Garage, Accessed on June, July 2023, <[Multivariate Time Series Prediction with LSTM and Multiple features (Predict Google Stock Price)](https://www.youtube.com/watch?v=gSYiKKoREFI)>.

[16] TensorFlow, Time Series Forecasting, Accessed in June, July 2023, <<https://www.tensorflow.org/tutorials/structured_data/time_series>>.

[17] Kaushik, Manav & Giri, Arun. (2020). Forecasting Foreign Exchange Rate: A Multivariate Comparative Analysis between Traditional Econometric, Contemporary Machine Learning & Deep Learning Techniques.

[18] S. Jagadeesh, "Multivariate Multi-step Time Series Forecasting using Stacked LSTM sequence to sequence Autoencoder in Tensorflow 2.0 / Keras," Data Science Blogathon, 2023.

[19] AIEngineering, Accessed on July 2023, <[Multivariate Time series using Vector Autoregression (VAR)](https://www.youtube.com/watch?v=TpQtD7ONfxQ)>

[20] Florian Follonier, Stock Market Prediction using Multivariate Time Series and Recurrent Neural Networks in Python, Accessed on July 2023, <<https://www.relataly.com/stock-market-prediction-using-multivariate-time-series-in-python/1815/>>

.