Forecasting Foreign Exchange rates using an ARIMA Model: A case study of USD/KSH rate

Pamba Gift

October 2025

Introduction

This document serves the purpose of explaining to the reader the procedures taken to develop a model that forecasts foreign exchange rates. The data for this project has been sourced Here. The CBK uploads forex data on `21` currencies. Out of the `21` currencies, this project works with the USD/KES pair. EDA was conducted to the data to discern the possible hypothesis tests and assumptions to be made. Anomalies such as duplicates and wrongly worded characters were identified and dealt with. No missing values were reported.

The data used to build the model runs from 1st December 2016 upto 13th June 2025. It has `2114` observations made on weekdays, excluding public holidays and weekends. All rates are the equivalent value of 1 US Dollar.

Descriptive summaries

Date

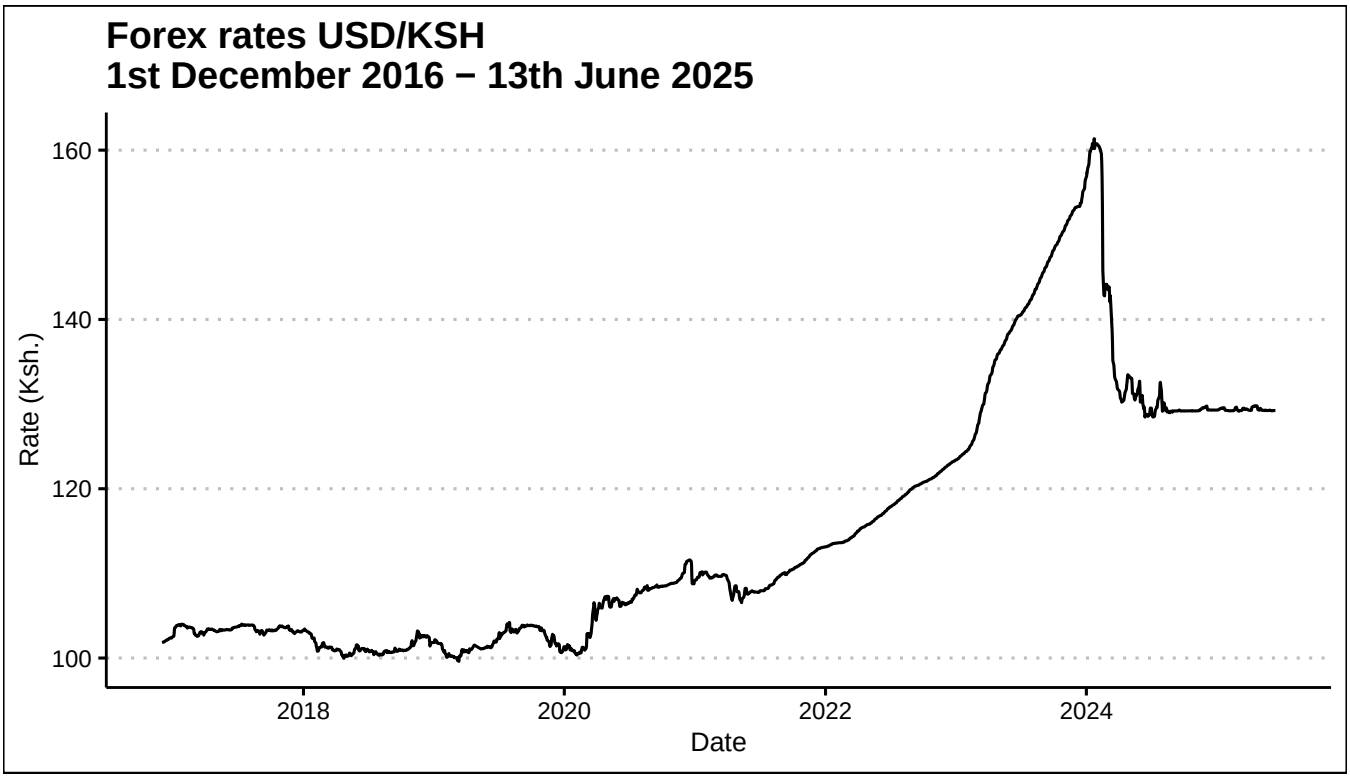
Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
2016-12-01	2019-01-16	2021-03-10	2021-03-09	2023-04-25	2025-06-13

Rates

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
99.61	103.1	108.8	115.1	129.2	161.4

Data Visualization

Time plot



Methodology

An ARIMA model was chosen to forecast the Forex rates. Fortunately, the `forecast` library has a function that returns an optimal model. This accelerated the workflow and diagnostics. An ARIMA model can have an auto-regressive coefficient(s) or moving average coefficient(s) or both. In addition ,they allow for differencing if the data to be modelled is non-stationary. The optimal model chosen for this data is the `ARIMA(1,2,1)` with an auto-regressive order of `1`, a moving average order of `1` and a differencing order of `2`. The model can be written as;

$$y_t = 2.3198y_{t-1} - 1.6396y_{t-2} + 0.3198y_{t-3} - 0.9845\epsilon_{t-1}$$

Results and Findings

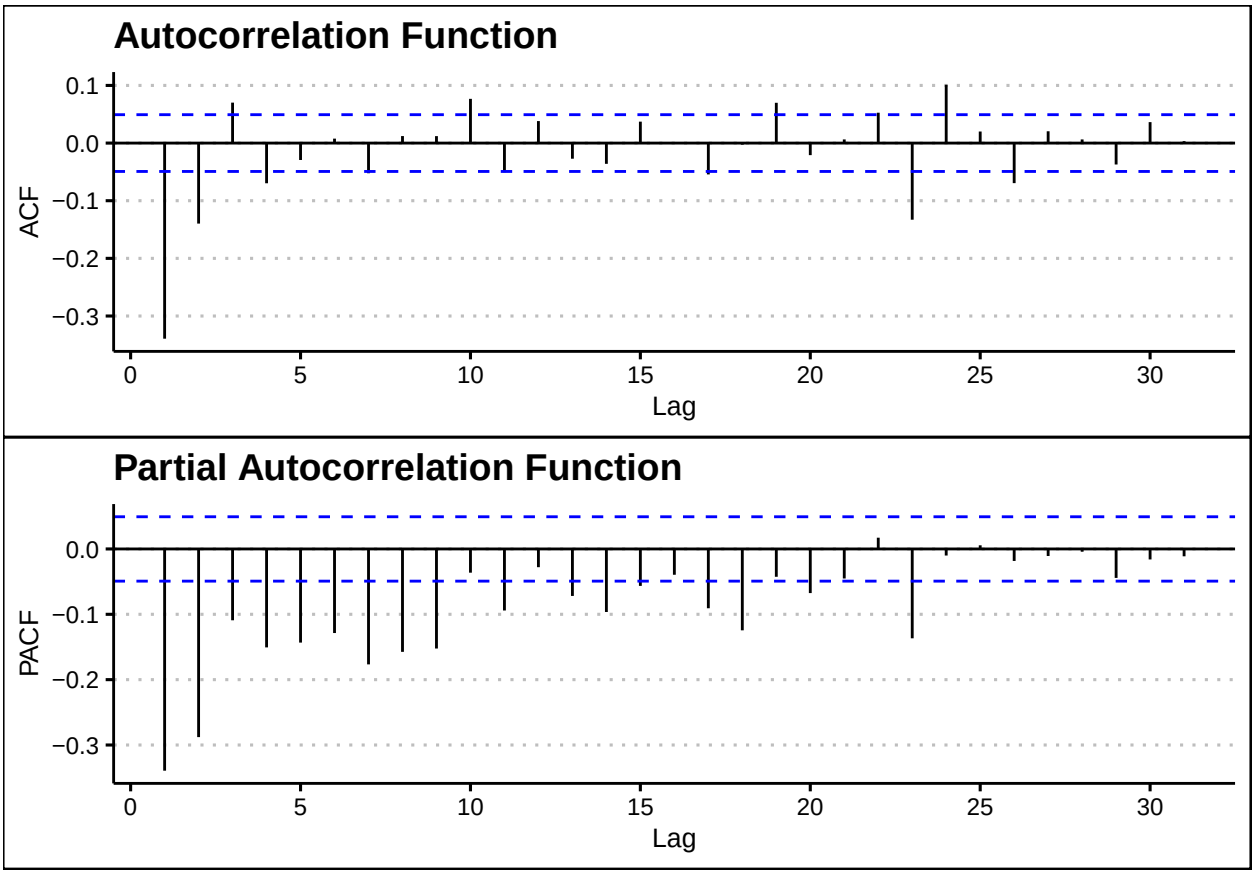
Model building

The first course of action was to split the data into a training (`75%`) and testing set (`25%`). The training set runs from 1st December 2016 upto 25th April 2023 containing `1585` observations. The testing set runs from 26th April 2023 upto 13th July 2025 with `529` observations. The training set facilitated model building and diagnostic checks while the testing set was used to evaluate the model's predictive ability.

Model validation

ACF and PACF

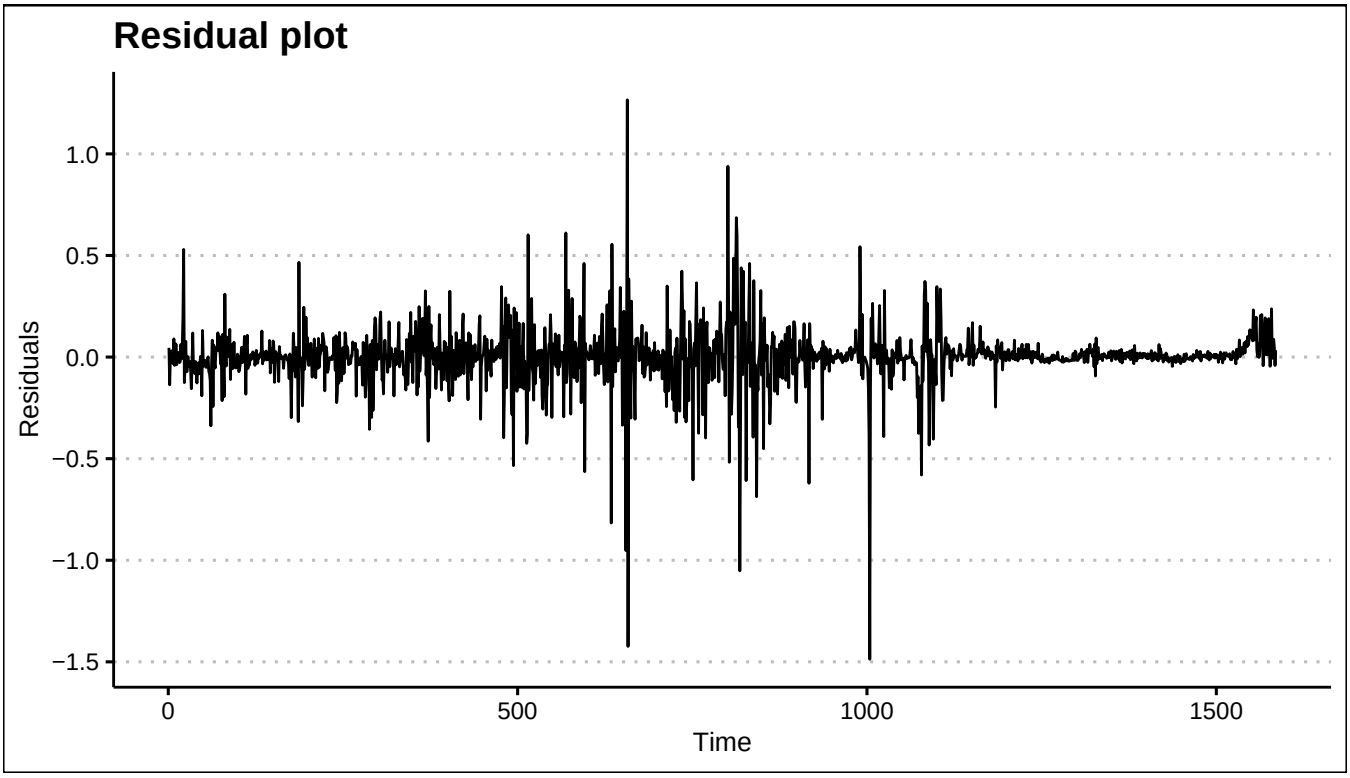
This section validates the order used under the ARIMA model `p = 1` and `q = 1`. This is made possible via autocorrelation (ACF) and partial auto-correlation plots. (PACF) The figure below shows the ACF and PACF for the training set.



Both the ACF AND PACF tail off gradually as there exists some significant spikes as the lag increases. In this case a mixed model such as an ARMA model would be of use.

Residual analysis

For an ARIMA model to hold, its residuals should have no autocorrelation (White noise). First, let us visualize the training set's residuals.



The residuals appear to be centered around `0.0` with a few extremes. Next, we use the `Ljung-Box Test` to check for independence of residuals.

Table 3: Box-Ljung test: `residuals(model)`

Test statistic	df	P value
0.0008154	1	0.9772

A p-value of `0.9772` suggests that we fail to reject the null hypothesis of residual independence.

Predictive ability

This section analyses forecast ability of our ARIMA model. The table below shows the predictive metrics for the training set;

Table 4: Table continues below

	ME	RMSE	MAE	MPE	MAPE	MASE
Training set	0.002892	0.1446	0.0782	0.00226	0.07446	0.8444

	ACF1
Training set	-0.0007166

- Summary
  - The **ME** (Mean Error) value of `0.0028921` indicates that the model has a minimal bias score.

  - A **MAPE** (Mean Absolute Percentage Error) value of `0.0744629` shows that on average ,the model's fitted values deviate by about `7.4 %` from the actual values, which is an acceptable value for forecast models.

Next, we intend to make a forecast 21 days ahead *i.e* 26th April 2023 upto 23rd May 2023 with weekends and Labour Day excluded. After making the forecast, we once again check the model's predictive metrics, this time, using the first 21 observations of the testing set. The metrics are shown below;

	ME	RMSE	MAE	MPE	MAPE
Test set	0.01289	0.06385	0.04479	0.01254	0.0438

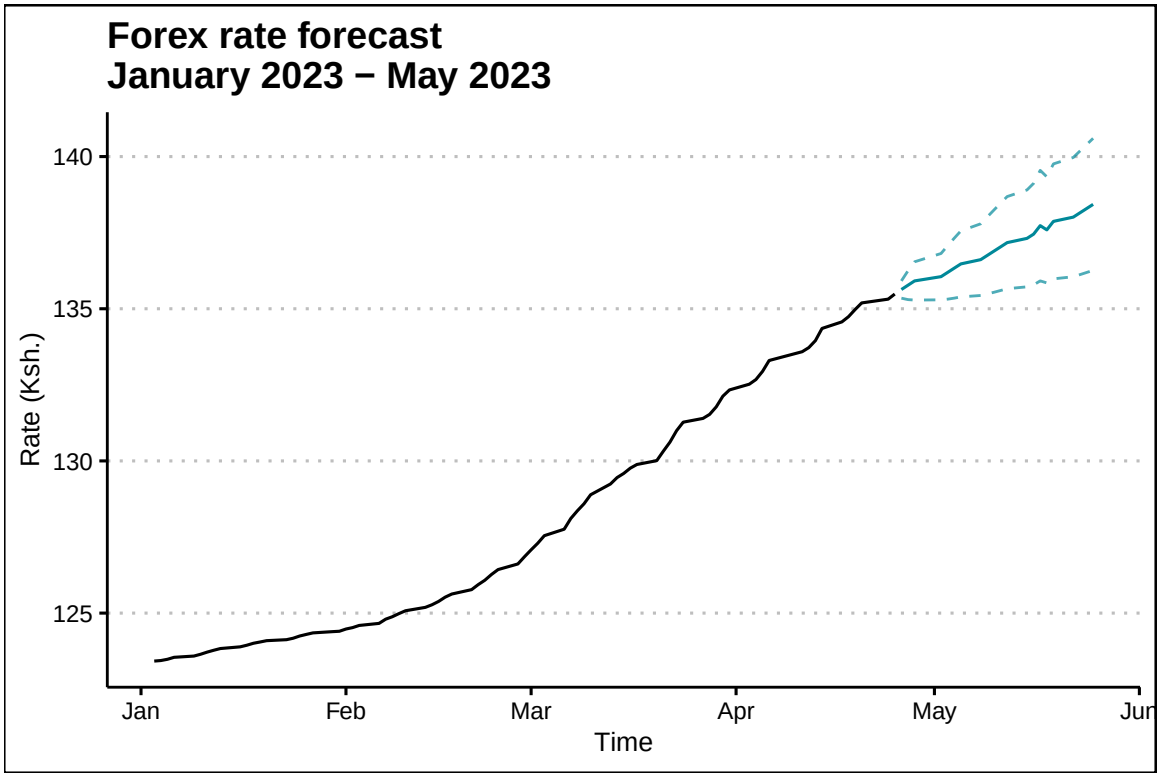
- Summary
  - The **ME** is relatively close to 0 ,implying a small prediction bias.
  - The **MAPE** value of `0.04479` shows minimal deviation between fitted values and the reported values.

Below is the table consisting of the forecast values for the period 26th April 2023 upto 25th May 2023 with 95% confidence bounds;

Date	Actual_rate	Forecast_rate	Abs_Deviation	Lower_bound	Upper_bound
2023-04-26	135.6588	135.6311	0.0277310	135.3473	135.9148
2023-04-27	135.8324	135.7737	0.0587487	135.3003	136.2470
2023-04-28	135.9118	135.9143	0.0024871	135.2856	136.5430
2023-05-02	136.0176	136.0543	0.0367002	135.2931	136.8155
2023-05-03	136.1529	136.1941	0.0412142	135.3155	137.0727
2023-05-04	136.2618	136.3339	0.0720645	135.3486	137.3192
2023-05-05	136.3971	136.4736	0.0764945	135.3893	137.5579
2023-05-08	136.4676	136.6133	0.1457180	135.4357	137.7909
2023-05-09	136.5853	136.7530	0.1677393	135.4866	138.0194
2023-05-10	136.6765	136.8928	0.2162600	135.5412	138.2444
2023-05-11	136.7912	137.0325	0.2412805	135.5986	138.4664
2023-05-12	136.8765	137.1722	0.2957009	135.6584	138.6860
2023-05-15	136.9794	137.3119	0.3325213	135.7203	138.9035
2023-05-16	137.1029	137.4516	0.3487417	135.7839	139.1194
2023-05-18	137.3735	137.5914	0.2178621	135.8490	139.3338
2023-05-17	137.2382	137.7311	0.4928825	135.9153	139.5469
2023-05-19	137.4912	137.8708	0.3796029	135.9827	139.7589
2023-05-22	137.6265	138.0105	0.3840233	136.0511	139.9700
2023-05-23	137.7618	138.1502	0.3884437	136.1202	140.1802
2023-05-24	137.9559	138.2900	0.3340640	136.1902	140.3898
2023-05-25	138.1324	138.4297	0.2972844	136.2607	140.5987

- Summary
  - All of the Actual rates fall within the 95% confidence bounds.
  - The largest deviation `0.4928825` was recorded on `2023-05-17`
  - The smallest deviation `0.0024871` was recorded on `2023-04-28`

Now, we plot the model's 21 step ahead forecast.



Conclusion & Recommendations

- Foreign exchange data is non-stationary in nature and therefore, there is need to perform a stationary check before modelling.
- ARIMA models have good predictive metrics that make them suitable for forecasting foreign exchange data.
- Other non-linear models such as GARCH and prophet can be used in place of ARIMA when it comes to forecasting forex data in order to handle their volatility.

References

- Mong T. U. (2016). Forecasting Foreign Exchange Rate by using ARIMA Model: A Case of VND/USD Exchange Rate. *Research journal of finance and accounting Vol.7(No.12)*