Capstone 2 Final Report: FOREX Forecasting

Author: Siu Ling Cheung

1. Introduction

In this project, my goal is to formulate a trading strategy to trade USD/ZAR. I chose to analyze the exchange rate for South African Rand (ZAR) using Dollars (USD) in this project because of a personal interest. I will compare the performance (i.e. profitability) using 1.) technical indicators to generate trading signals 2.) build an ARIMA and LSTM model to forecast the USD/ZAR exchange rate for 29 days in the future.

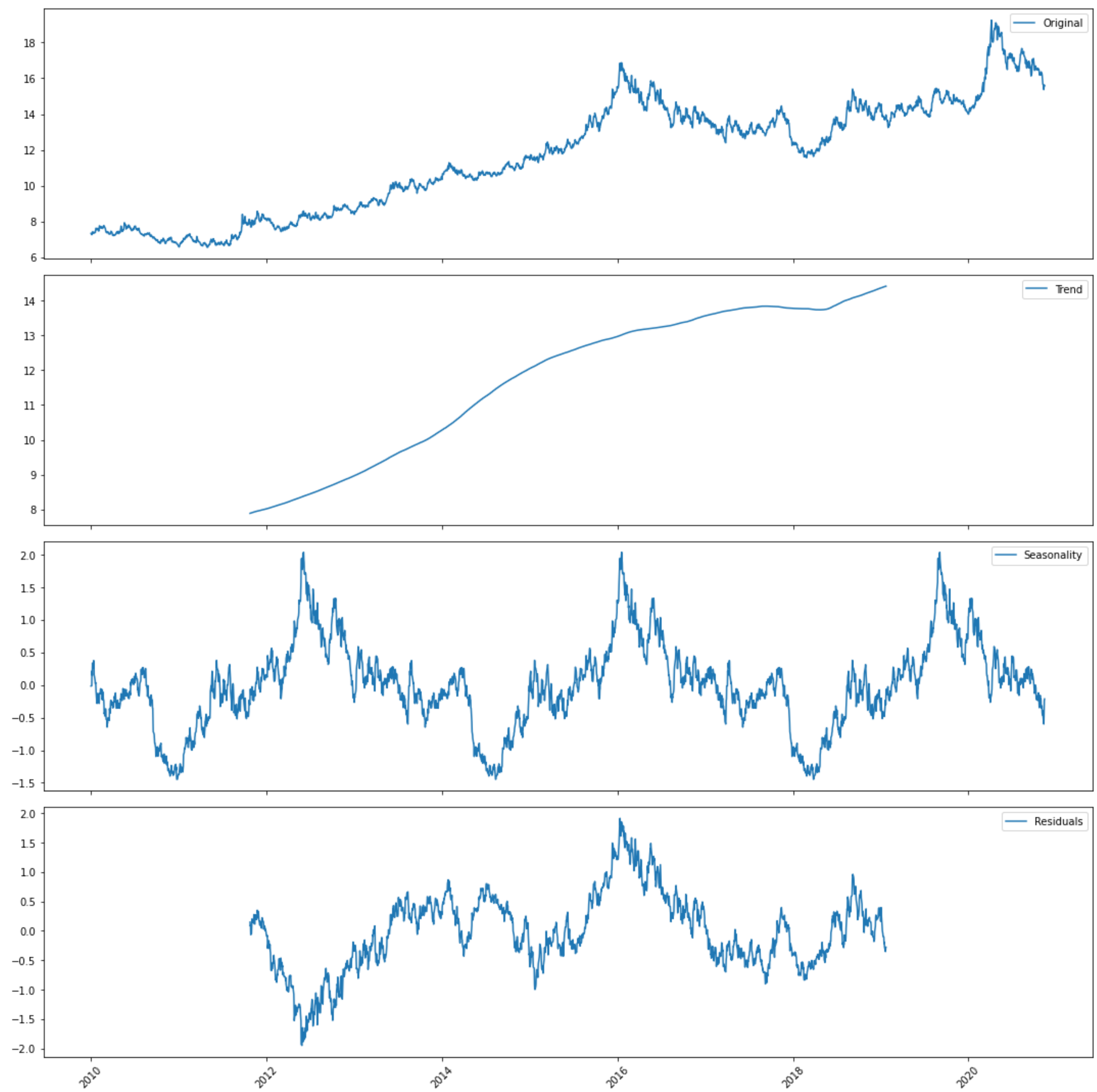
2. Data Wrangling

I downloaded the past 11 years of data from yahoo finance containing 2831 prices using yfinance and python library and dropped all the rows/days with missing data which usually correspond to holidays and weekends when the stock market is closed.

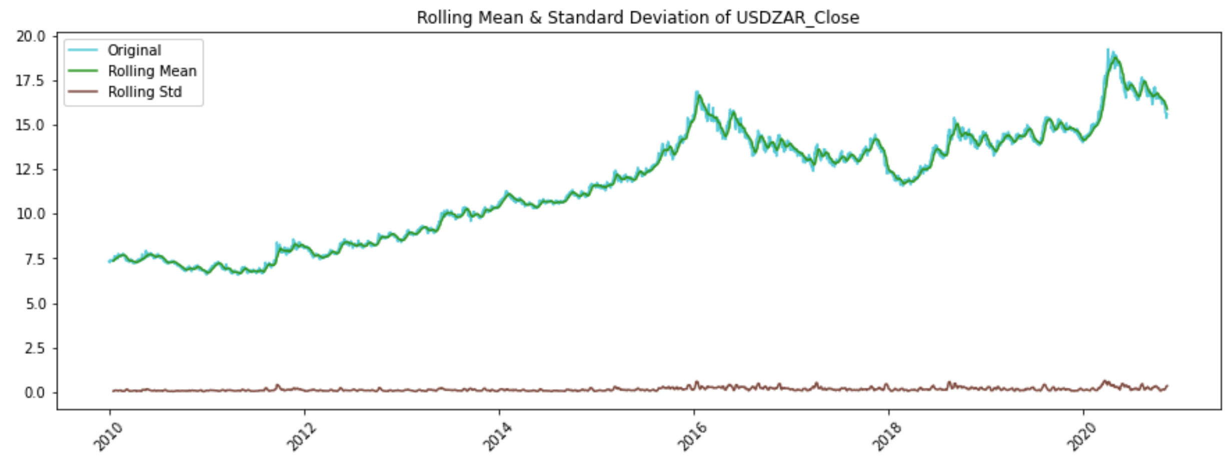
3. Exploratory Data Analysis

3a. Time Series Decomposition

Trend, Seasonal, and Residual Decomposition of USA/ZAR adjusted close



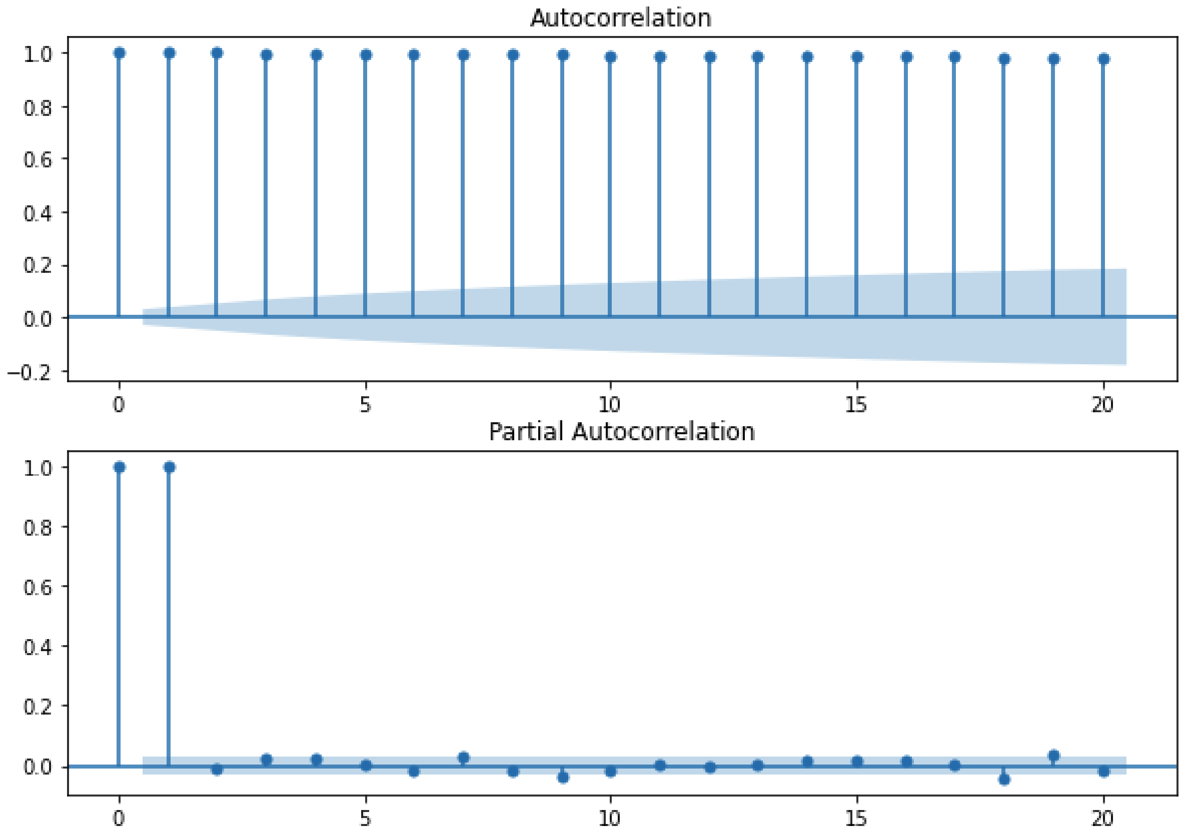
The above 3 components of the time series for USD/ZAR adjusted close price were decomposed based on the assumptions that the components are additive and it contains 3 cycles.




```
Results of Dickey-Fuller Test:
Test Statistic                -1.0361057002077456
p-value                        0.7398689192744539
# Lags Used                                  20.0
Number of Observations Used               2,808.0
Critical Value (1%)            -3.432680943017927
Critical Value (5%)            -2.862569847903434
Critical Value (10%)          -2.5673182195000446
dtype: float64
```

The adjusted close price is non-stationary because it has an upward trend and a repeating cycle every year i.e. yearly seasonality. The mean of the residuals seems to deviate from zero frequently which implies it doesn't follow a normal distribution with mean 0. Furthermore, the Adfuller test statistics is larger than the critical values and the p-value is larger than 0.05 which means we cannot reject the null hypothesis. It is likely that this data is not stationary and has some time dependent structure.

3b. Autocorrelation and Partial Autocorrelation



The autocorrelation plot with lag=20 shows that the exchange rate at time t+1 is correlated with time t. The partial autocorrelation plot shows time t-1 is the shortest lag correlated with time t controlling for all other lags in the time series.

4. Machine Learning

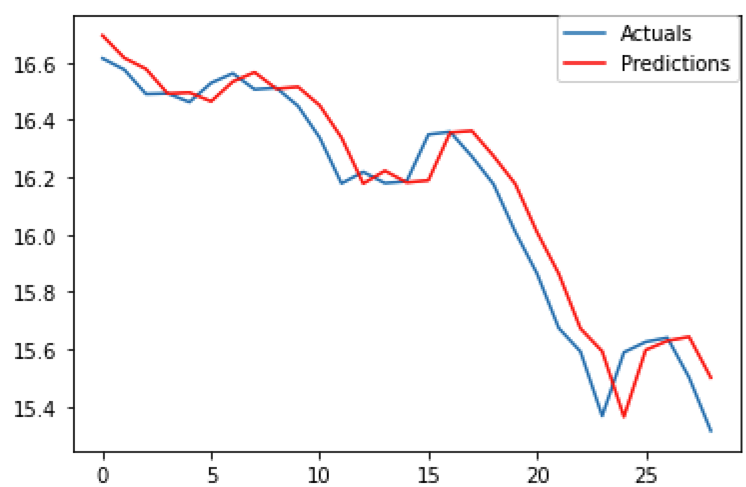
I chose 3 different forecasting methods to forecast exchange rates in the next 29 days because I want to discover which method fits my problem at hand. The methods I've chosen are completely different mathematical approaches which demonstrates how one could improve model accuracy by evaluating multiple methods. Ensemble methods (tree-based) also work well in forecasting and are proven to be able to forecast with high accuracy but I excluded them in this project because I have evaluated them extensively in my Capstone 1 project.

I have chosen Mean Absolute Percentage Error (MAPE) as the measure of prediction accuracy due to its intuitive interpretation in terms of relative error.

4b. ARIMA

ARIMA seems to be a widely used forecasting method for predicting price/returns due to its capability of capturing a suite of different standard temporal structures in time-series data and it is more efficient than neural networks.

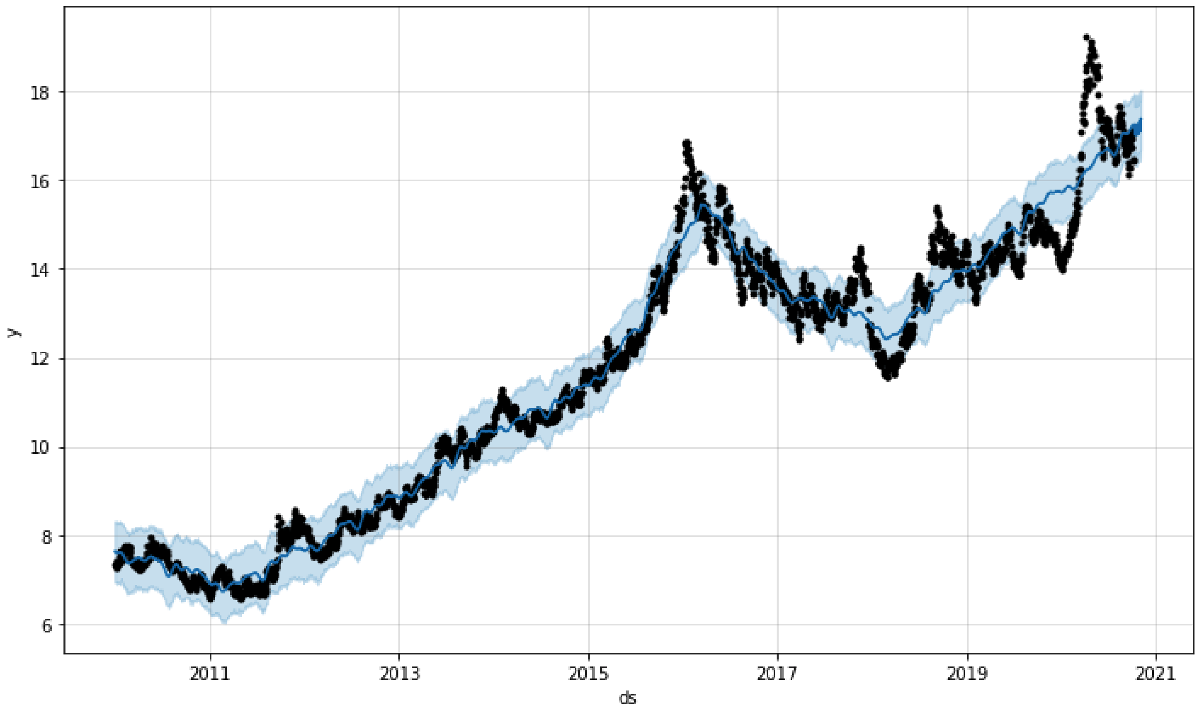
```
MAPE for Test Data: 2.7488%
```



Above chart shows the performance of the ARIMA(1,1,0) model on the validation data only. This is a model with lag=1, differencing=1 and seasonality=1. The parameters were determined by the lowest AIC score obtained in training various ARIMA models with different combinations of parameters (i.e. p=1,2 q=0,1,d=0,1).

4c. Prophet

Prophet works best for time series data that has components with piecewise linear relationship and predictable stationary seasonality.




```
MAPE for Training data:  3.02 %
MAPE for Test data:  6.49 %
```

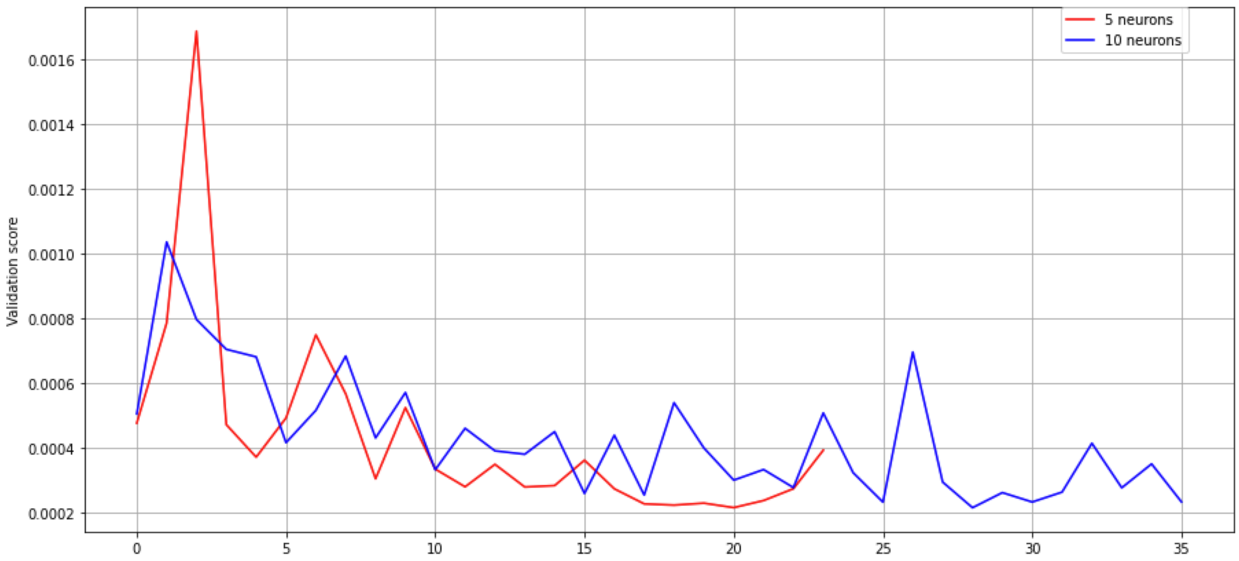
Prophet does not seem to be able to capture the seasonality of USD/ZAR exchange rate as well as other methods and that would be seen by generating a higher prediction error of 6.49% on the validation data.

4d. LSTM

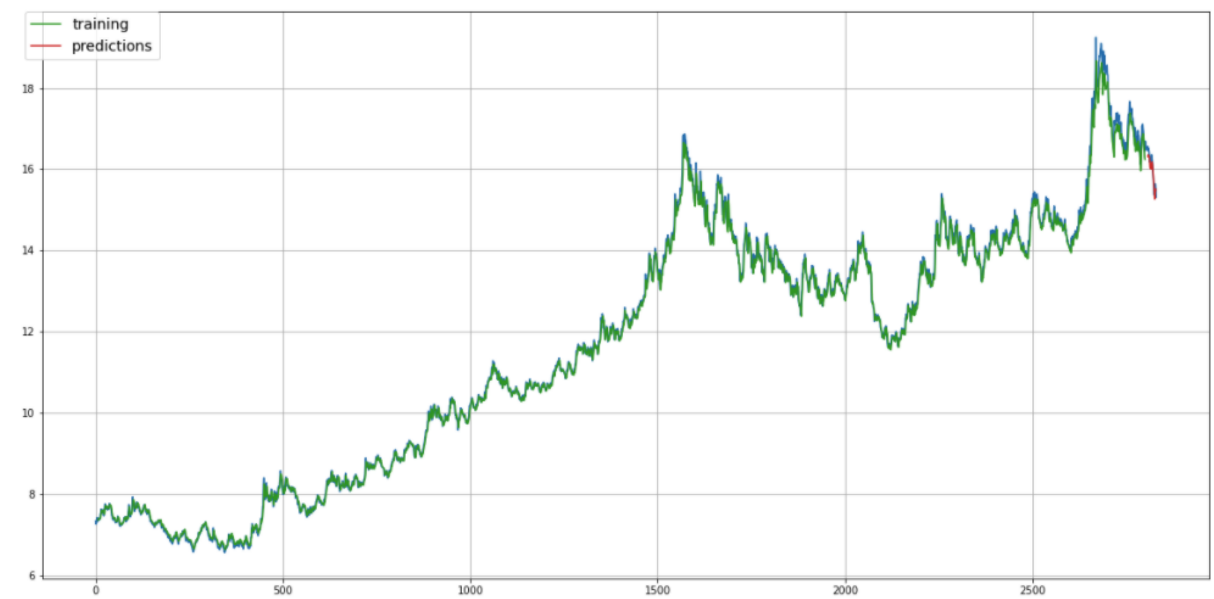
Neural Network only works when there is lots of training data and it does appear to be the best method for predicting USD/ZAR exchange rate with the lowest MAPE partly because I used almost the past 11 years as training data.

LSTM is a type of recurrent neural network RNN that works well in forecasting time series data given its capability to remember long term relationships.

Firstly, I split the training and test dataset (99% train 1% test) using a 6 days look back period. Secondly, I normalized my dataset using MinMaxScaler to transform my data so every data point falls between 0 and 1. Thirdly, I converted my dataset into a matrix so it could be fed to the neural network. Fourth, I tested the network architecture by comparing the MSE of a model with 10 neurons (model 1) and another with 15 neurons (model 2). I also used early stopping as callbacks - if accuracy did not improve after 4 iterations the training would stop. Below line plots show model 2 achieved a lower MSE (i.e. higher accuracy) with 10 neurons. Hence, I used model 2 to generate predictions on the test data.



MAPE on Train Data: 0.85%
MAPE on Test Data: 0.83%



LSTM has the lowest MAPE on the validation data is 0.83%.

Conclusion

LSTM is the winning forecast method with a MAPE on validation data 3X lower than ARIMA's and 10X lower than Prophet. This shows the power of the network ability to capture interactions between each lag of the time series and remembering long term relationships.

Future Improvement

1. Include other exogenous variables in the forecast that would likely affect the exchange rate such as Moody's rating on South African Government bonds, US interest rates, South Africa's GDP, etc.
2. Try other types of deep learning methods such as Gated Recurrent Unit Neural Network.