CAPSTONE 2 MILESTONE REPORT

Problem Statement

Foreign exchange rate forecasting is essential for managing foreign exchange risk, and is one area that financial institutions and all businesses that have exposure to foreign exchange are interested in. This topic is also essential for currency traders both at an institutional level and at the individual level.

Predicting foreign exchange rate has been a challenging task for traders and practitioners in financial markets, and this project attempts to examine and compare the effectiveness and performance of ARIMA and Neural Networks in predicting Foreign Exchange rate.

Methodology

Using foreign exchange rates from 2000 to 2019, this project attempts to predict EUR/USD exchange rate for 2019 , using ARIMA and LSTM models. Time series forecasting models support the assumption that past patterns in data can be used to forecast the future.

ARIMA (Autoregressive Integrated Moving Average) consists of three parts namely:

- Autoregressive denoted by 'p' : defined by the autocorrelation function (ACF)
- Integrated denoted by 'd' : defined by the stationary test of a time-series. We will be using the Dickey-fuller test for this.
- Moving average denoted by 'q': Defined by the partial correlation function (PACF).

LSTM: The task here will be to predict values for a time series given the history of eurusd daily closing exchange rate between year 2000 and 2019. I will be using the multi-layered LSTM recurrent neural network to predict the last couple of years of exchange rate

Dataset

Consists of daily closing exchange rate of eurusd downloaded from barchart.com

Data Cleaning

Data supplied from website is clean and complete and in the required format. I carried out the following checks to determine data quality:

I examined completeness and data types. I also examined the summary statistics of the dataset. All checks revealed that the dataset is clean and is in the required format to be used in my models.

Exploratory Data Analysis

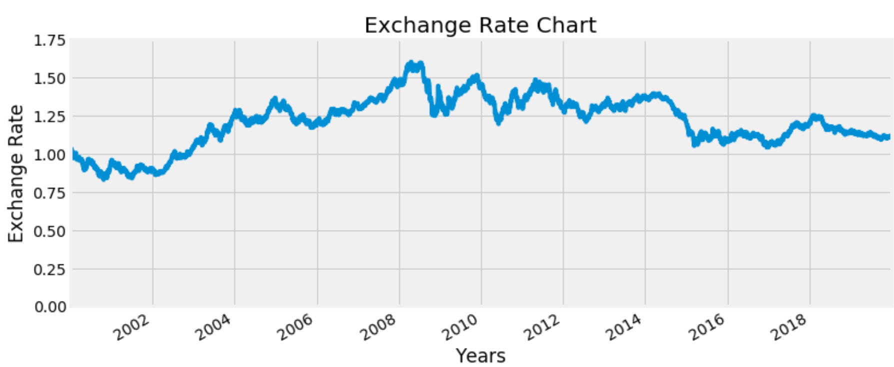
Hypothesis 1:
p-value > 0.05: Accept the null hypothesis (H0), the data has a unit root and is non-stationary.¶
p-value <= 0.05: Reject the null hypothesis (H0), the data does not have a unit root and is stationary.

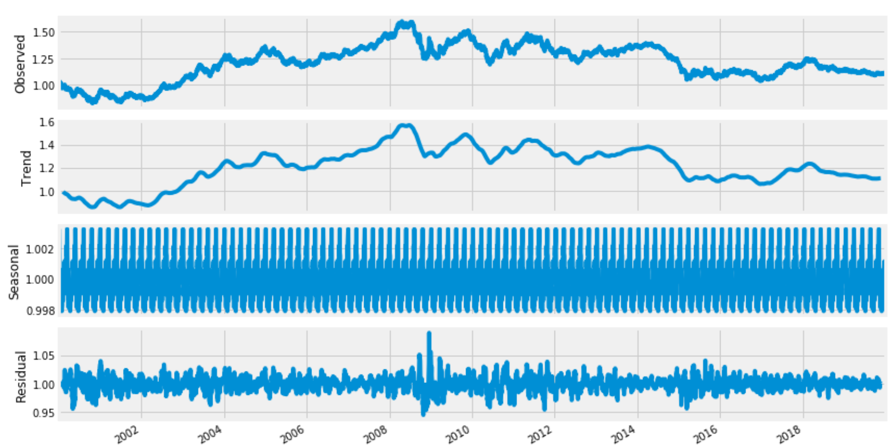
Result of hypothesis:
From the above results, it is clear that the Standard deviation is stationary, while the mean is not. We will reject the null hypothesis H0, the data does not have a unit root and is stationary.

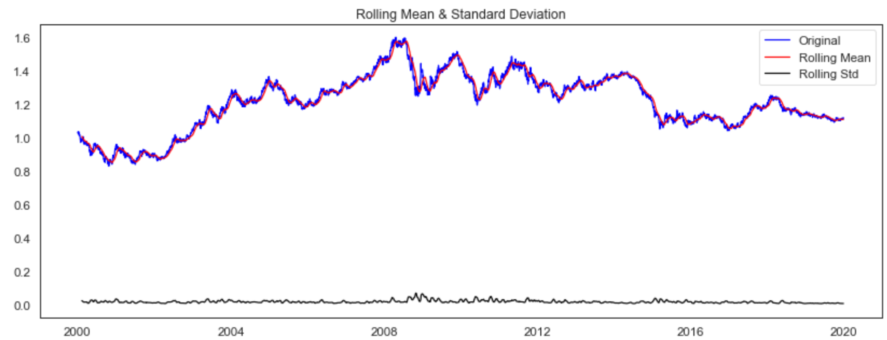
Stationarity¶
I used the Dickey–Fuller test to test the null hypothesis that a unit root is present in an autoregressive model. Stationary series has constant mean and variance over time. Rolling average and the rolling standard deviation of time series do not change over time.

Dickey-Fuller test Null Hypothesis (H0): Time series has a unit root, meaning it is non-stationary. Alternate Hypothesis (H1): Suggests the time series does not have a unit root, meaning it is stationary.

I decomposed the eurusd into its different components as shown below”







Results of Dickey-Fuller Test

Test Statistic	-1.7338
p-value	0.4138
#Lags Used	0.0000
Number of Observations Used	5186.0000
Critical Value (1%)	-3.4316
Critical Value (5%)	-2.8621
Critical Value (10%)	-2.5671

dtype: float64

For a timeseries to be considered stationary, it is expected that the mean and standard deviation should be constant which means the data should not have trend and seasonality. As mentioned earlier, I used ADF to determine the stationarity of my time series and to determine the order of differencing.

Statistical Normality Test

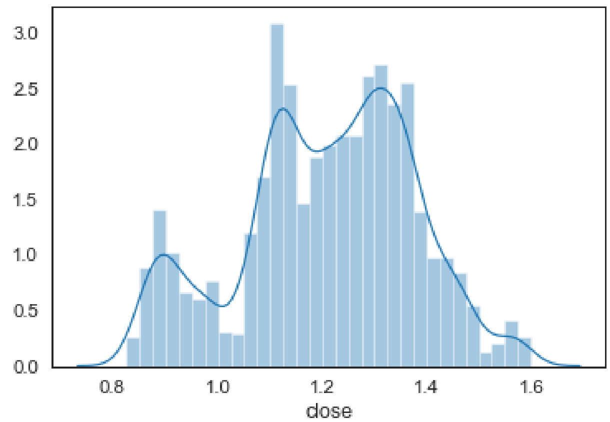
I used the statistical normality test to quantify whether data looks as though it was drawn from a Gaussian distribution, using the D'Agostino's K^2 Test.

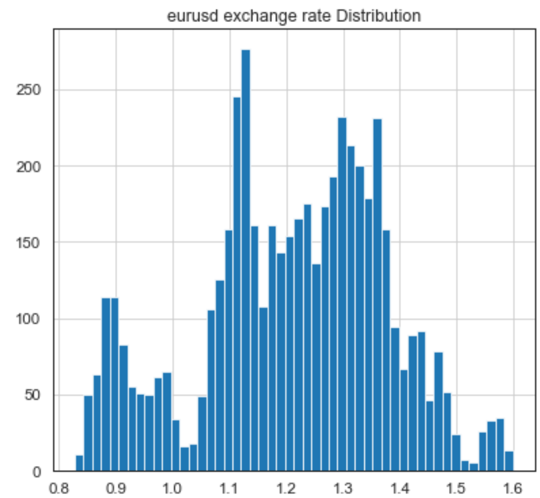
Hypothesis 2:
p <= alpha: reject H0, not normal.¶
p > alpha: fail to reject H0, normal

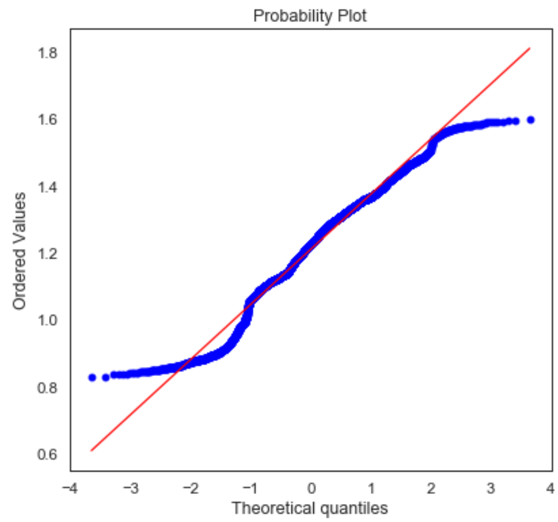
Result of hypothesis:
From the statistics below it is clear that the data does not look Gaussian, we will therefore reject the null hypothesis H0.

Statistics=137.136, p=0.000
Kurtosis of normal distribution: -0.43668157335097213
Skewness of normal distribution: -0.28722463936355497

The kurtosis of the distribution above is less than zero and is light tailed. The distribution is fairly symmetrical, as seen below:







The table below shows the summary statistics of the eurusd time series:

	count	mean	std	min	25%	50%	75%	max
Open	5187.0000	1.2093	0.1666	0.8266	1.1103	1.2258	1.3315	1.5991
High	5187.0000	1.2147	0.1672	0.8322	1.1149	1.2313	1.3372	1.6038
Low	5187.0000	1.2039	0.1658	0.8230	1.1067	1.2196	1.3254	1.5866
close	5187.0000	1.2093	0.1666	0.8271	1.1105	1.2259	1.3314	1.5990
Change	5187.0000	0.0000	0.0074	-0.0348	-0.0041	0.0001	0.0041	0.0467
Volume	5187.0000	119833.2103	155557.6128	0.0000	0.0000	0.0000	229402.5000	763921.0000