December 14th 2018

Algorithmic Trading in the Foreign Exchange Market

Amit Pingale, Shubham Patel, Vikas Kumar

Abstract

Foreign exchange market prediction is the act of trying to determine the future value of a currency traded on the exchange. The currency of a nation is dependent on numerous inter-linked factors that contribute to the volatility of the market price. The nexus of information corresponding to a particular currency is analyzed to determine patterns exhibited in connected services for developing trading strategies. The objective of this research paper is to investigate various features that contribute to the evaluation of the exchange rate for a particular currency. These features are taken into account for making predictions for developing trading strategies.

Background

The exchange market is driven mainly by uncertainty and future expectations. The exchange rate of a country is highly complicated and is determined by a great number of variables. Global and domestic news and social media posts can hit all markets and create huge movements, all based on reaction. That makes the market difficult to predict. The analysis of this domain is complicated as there is an esoteric language used by the financial experts to explain their ideas. There are many non-statistical parameters that influence the exchange rate in the short run, and we want to explore factors such as trading volume, condition of business and market sentiments of target country.

Approach

1. Exploring various features that contribute in prediction of target currency
2. Building time series model on gathered dataset
3. Developing an ensemble model based on lowest RMSE value.
4. Using Bollinger band and double cross-over for deciding the threshold for trading

Dataset

We have implemented eight different models with features suitable for a particular use case. The dataset contains features of other currencies and ADS index for model engineering. All the currencies are converted into USD for comparison.

Below are the variables description:

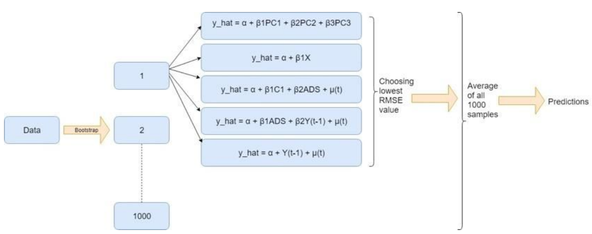
- *DEXUSUK* - Pound Sterling
- *DEXBZUS* - Brazilian Real
- *DEXUSEU* - Euro

- *DEXUSAL* - Australian dollar
- *DEXSZUS* - Swiss franc
- *ADS Index* - Aruoba-Diebold-Scotti (ADS) Index

Methodology

I. Random Forest

Random forest is an ensemble learning method for regression that operates by constructing a multiple decision for regression and another task that operates. By constructing a multitude of decision trees at training time and outputting the mean prediction of the individual trees.



The data is resampled thousand times using bootstrap technique. Each sampled dataset is passed through five different models. PCA is taken into consideration as a feature in one of the models. The model having the least RMSE values is chosen out of the five regression models. This procedure is repeated thousand times to get a better accuracy. Finally, an ensemble model is created for averaging all the selected models. This model gives a high prediction accuracy and takes into account features such as ADS index, PCA and autocorrelation values.

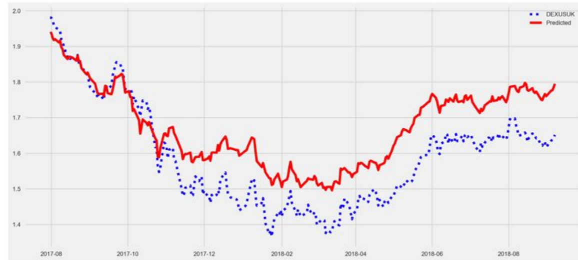


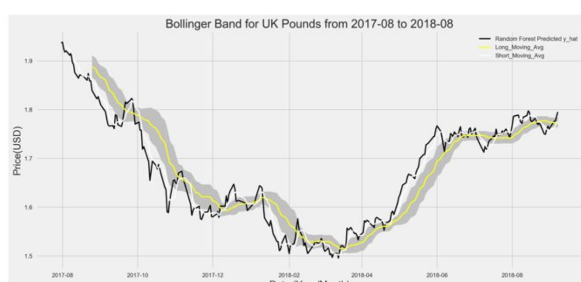
fig. Random Forest predictions

Trading Strategies

Following strategies are used for trading in foreign exchange market:

- **Bollinger Band:**

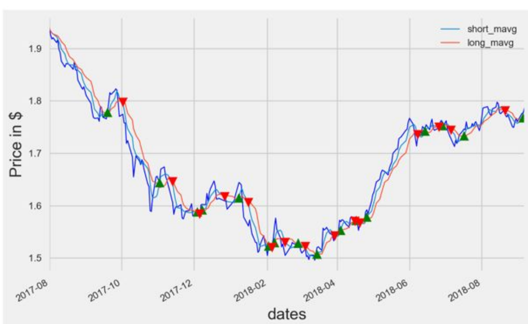
Bollinger band is a trading strategy based on the standard deviation of the distribution. Long and short moving averages are calculated for determining the optimum band values. A window size of 20 days is taken for calculating long moving average whereas a moving average of 5 days is taken into account for short moving average. The upper band is calculated by taking the sum of long and short moving average scaled with 0.75 standard deviation. Similarly, a difference of long and short moving average scaled to a factor of 0.75 is used for deciding the lower band.



When the predicted price of the currency goes below the lower band of Bollinger band the market is assumed to be under bought and a position of buy should be taken. On the other hand, when the price crosses the upper band the market is assumed to be overbought thus a signal of squaring off previous

- **Double cross-over:**

Double cross-over is a trading strategy which decides position depending upon the intersection of short and long moving average. Long moving average for a period of 20 days and short moving average over az5 day period is calculated. When the short moving average crosses the long moving average a signal sell or short is taken depending upon the previous state. Similarly, when long moving average drops below the short moving average a signal of buy is given. Apart from theses condition a hold position is taken for rest of the time.



- **Applying Trading Strategies on Predictions**

The double cross-over technique is applied from 2017 to 2018. The following is the graph of buy, hold and sell during that period.

position or taking a short position is given by the algorithm

- **Results obtained from Random forest:**

Risk Metrics	Values
Gross Profit	0.312
Gross Loss	- 0.024
Net Profit	0.2881
Profit Factor	- 13.00
Total Return	16.11

The total return expected on principal amount over a period from 2017 – 2018 is calculated to be a profit of 16.11 percent.

II. Kalman Filter

Kalman filter is an algorithm uses a series of measurements observed over time, containing statistical noise and other inaccuracies, and produces estimates of unknown variables that tend to be more accurate than those based on a single measurement alone, by estimating a joint probability distribution over the variables for each time frame.

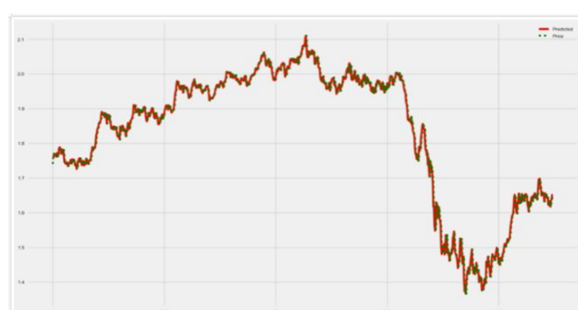


fig. Red is predicted values using kalman filter

III. ARIMA

An ARIMA(AutoRegressive Integrated Moving Average) model is a class of statistical models for analyzing and forecasting time series data.This acronym is descriptive, capturing the key aspects of the model itself. Briefly, they are:

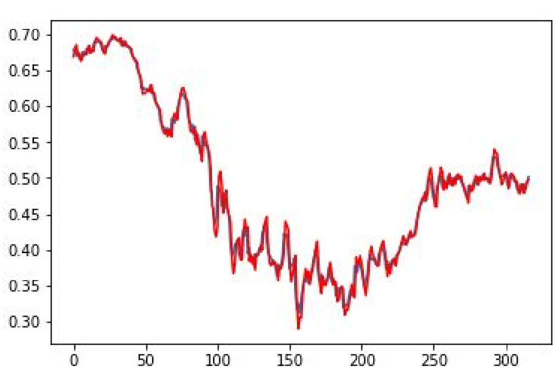
AR: Autoregression. A model that uses the dependent relationship between an observation and some number of lagged observations.

I: Integrated. The use of differencing of raw observations (e.g. subtracting an observation from an observation at the previous time step) in order to make the time series stationary.

MA: Moving Average. A model that uses the dependency between an observation and a residual error from a moving average model applied to lagged observations.

Each of these components are explicitly specified in the model as a parameter. A standard notation is used of ARIMA(p,d,q) where the parameters are substituted with integer values to quickly indicate the specific ARIMA model being used.The parameters of the ARIMA model are defined as follows:

p: The number of lag observations included in the model, also called the lag order.



d: The number of times that the raw observations are differenced, also called the degree of differencing.

q: The size of the moving average window, also called the order of moving average.

A value of 0 can be used for a parameter, which indicates to not use that element of the model. This way, the ARIMA model can be configured to perform the function of an ARMA model, and even a simple AR, I, or MA model.

Determining value of ‘p’ and ‘q’

Autocorrelation Function (ACF): It is a measure of the correlation between the the TS with a lagged version of itself. For instance at lag 5, ACF would compare series at time instant ‘t1’…’t2’ with series at instant ‘t1-5’…’t2-5’ (t1-5 and t2 being end points).

Partial Autocorrelation Function (PACF): This measures the correlation between the TS with a lagged version of itself but after eliminating the variations already explained by the intervening comparisons. Eg at lag 5, it will check the correlation but remove the effects already explained by lags 1 to 4.

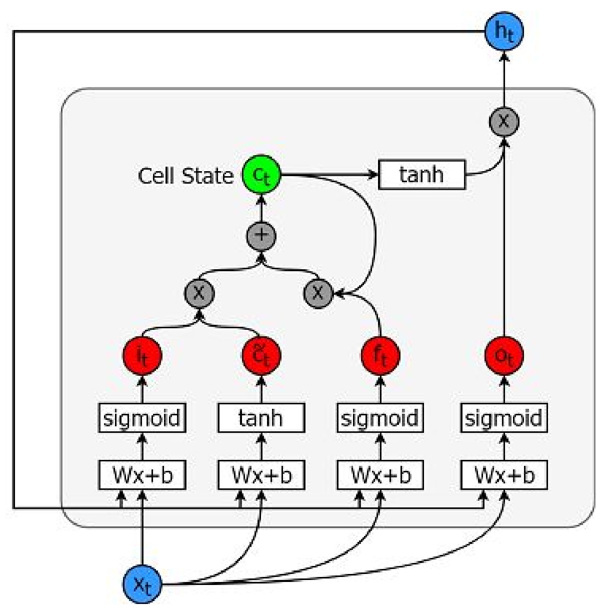
fig. Red is predicted values using ARIMA(2,2,0)

IV. LSTM

Long Short-Term Memory models are extremely powerful time-series models. They can predict an arbitrary number of steps into the future. An LSTM module (or cell) has 5 essential components which allows it to model both long-term and short-term data.

- Cell state (ctct) - This represents the internal memory of the cell which stores both short term memory and long-term memories
- Hidden state (htht) - This is output state information calculated w.r.t. current input, previous hidden state and current cell input which you eventually use to predict the future stock market prices. Additionally, the hidden state can decide to only retrive the short or long-term or both types of memory stored in the cell state to make the next prediction.
- Input gate (itit) - Decides how much information from current input flows to the cell state
- Forget gate (ftft) - Decides how much information from the current input and the previous cell state flows into the current cell state
- Output gate (otot) - Decides how much information from the current cell state flows into the hidden state, so that if needed LSTM can only pick the long-term memories or short-term memories and long-term memories

A cell is pictured below.



And the equations for calculating each of these entities are as follows.

- $i*t* = \sigma(*W*{ix}x*t* + *W*{ih}h_{t-1}+b_i)$

Conclusion

On analysis of all the models we found that Kalman filter is giving highest accuracy with an expected profit percentage of 53%. Predicting forex market is a tedious task which requires more correlated features for better predictions of the target variable.

Resources & References

1. https://pythonprogramming.net/finance-programming-python-zipline-quantopian-i ntro/
2. https://medium.com/mlreview/a-simple-deep-learning-model-for-stock-price-predi ction-using-tensorflow-30505541d877
5. Guanting Chen [guanting] 1 , Yatong Chen [yatong] 2 , and Takahiro Fushimi [tfushimi] 3"Application of Deep Learning to Algorithmic Trading".
6. https://acadgild.com/blog/streaming-twitter-data-using-kafka
7. https://www.sciencedirect.com/science/article/pii/S1877050918308433
3. https://arxiv.org/ftp/arxiv/papers/1603/1603.00751.pdf
4. Sushree Dasa, , Ranjan Kumar Beheraa, , Mukesh kumarb , Santanu Kumar Ratha," Real-Time Sentiment Analysis of Twitter Streaming data for Stock Prediction"