Foreign Exchange Rates forcasting with LSTM

Huy Phung, Tashi Choden, Sahil Pasricha
University of Konstanz

February 2019

Contents

1 Abstract

Foreign Exchange (abbv *Forex*) is the market for curency investment. It is the second most important market, after stock market. Forex rate has been surveyed carefully by financial in order to explain the market behaviour or forcast the future price. We evaluate the performance and compare of LSTM to the widely used statistical models (ARIMA and VARMA) in forcasting of financial timeseries.

2 Data Description

2.1 Forex rates

Foreign Exchange Rates is the rate at which a pair of two currencies will be exchanged. Forex rates are updated within milliseconds, normally by financial protocol FIX. Important terms - Open, High, Low, Close - Bid, Ask and Spread In this project, the

2.2 Dataset

Dataset OHLC of BID price (lack of ASK price)

2.3 Forcasting

In this project, we concern about the prediction of future Open, High, Low, Close prices. Other features, either originally exists (volume) or later added (mean, median, momentum), are only considered as supporting features. These features are only used for prediction of OHLC features.

2.4 Model selecion

2.4.1 Akaike Information Criterion (AIC)

For statistical model selection, we use Akaike Information Criterion (AIC). As we can see, AIC penalize a model by its number of parameters. The number of parameters of a neural network based model is at least total number of elements contained in all weights in its topology. Therefore, we should not use AIC as a measure of performance between statistical models and deep learning models.

$$AIC = 2k - 2\ln(\hat{L})$$

2.4.2 Root Mean Squared Error (RMSE)

Root Mean Squared Error is widely used to measure the difference between values predicted by a model and the actually observed values. Given y represents the actually observed values and $\hat{y}$ represents the values predicted, $RMSE$ is given by:

$$RMSE = \left(\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2\right)^{\frac{1}{2}}$$

Root Mean Squared Error shows absolute difference between y and $\hat{y}$. However, since it does not take the range of possible values into account, it would be difficult to intepret the $RMSE$ result without knowing the possible range of predicted and actual values.

2.4.3 Mean Absolute Percentage Error (MAPE)

In order to measure the difference between predicted values and actual values with regarding to the scale, we use Mean Absolute Percentage Error (MAPE)

$$MAPE = \frac{100\%}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right|$$

3 Statistical Models

3.1 Autocorrelation

3.2 ARIMA

3.2.1 Model description

ARIMA consists of 2 sub models: AR(p) and MA(q)

3.2.2 Parameters selection

It is important for ARIMA model that we select the proper parameter (p, d, q) so that it covers all the past values which has effect on the current value. In order to do so, we first look at survey the ACF and PACF plot. For $AR(p)$ model , we find the furthest lag with significant autocorrelation in ACF plot. For $MA(q)$, we find at which lag the autocorrelation start to decay.

3.2.3 Result

3.3 VAR

3.3.1 Model description

VAR is applied to multivariate data

3.3.2 Parameters selection

Parameters selection for VAR consists only testing a range of lag with AIC.

3.3.3 Result

4 Deep Learning Model

4.1 Recurrent Neural Network

Recurrent Neural Network (RNN) is introduced by [4] to process sequential input. In RNN, each state connects to the followed state to form a directed graph. The structure of RNN makes it capable of handling sequential data with temporal dynamic behaviour.

One problem that may occurs within RNN is *vanishing gradient*https://www.youtube.com/watc [2]

4.2 Long-Short Term Memory

Hochreiter and Schmidhuber (1997) [1] presented a network architecture to solve the vanishing gradient problem from RNN Cummins presented

4.3 Proposed Network Topology

Kim and Elsaftawy [3] proposed a LSTM structure Our survey on the data implies that significant autocorrelation may appears even further than 20 timesteps into the past. In our proposed network topology, we add more hidden units into each LSTM layers, so that the network can learn from data points of higher lags.

4.4 Training and Validation

4.5 Results

4.5.1 Single Variate

4.5.2 Multivariate

5 Conclusion

As the results have shown, ...

However, computational effort spent for training LSTM deep learning model is much higher than ARIMA and VAR. Furthermore the, adding components to neural network topology results in higher computational cost.

References

[1] Felix A Gers, Jürgen Schmidhuber, and Fred Cummins. Learning to forget: Continual prediction with LSTM. 1999.

[2] Sepp Hochreiter, Yoshua Bengio, Paolo Frasconi, Jürgen Schmidhuber, et al. Gradient flow in recurrent nets: the difficulty of learning long-term dependencies, 2001.

[3] Dave Y. Kim and Mahmoud Elsaftawy. Eurusd 15 minute interval price prediction.

[4] David E Rumelhart, Geoffrey E Hinton, Ronald J Williams, et al. Learning representations by back-propagating errors. *Cognitive modeling*, 5(3):1, 1988.