Stock Price Forecast with Deep Learning LSTM

and Econometric ARIMA Models

Submitted to

American University of Armenia
Manoogian Simone College of Business and Economics

In partial fulfillment of the requirements for the degree of BA in Business

By: Hakob Ohanyan

Supervisor: Dr. Gayane Barseghyan

Yerevan 2018

ABSTRACT

The forecast of stock market prices is very important information for investors who have intention to invest in stock market. There are forecast models designed to make predictions. In this paper I propose different ones from conventional econometric models to machine learning models for forecasting stock prices. At the end the results of both of them will be discussed and conclusions will be made upon which one is better as a forecasting tool for APPLE INC. In this paper the tools for prediction stock prices are ARIMA model and recurrent neural network layers such as LSTM network. The results show that LSTM outperformed ARIMA model.

.

Keywords: Epoch, dropout, dense, neurons, layers, LSTM, autocorrelation, Ljung-Box, ADF.

ACKNOWLEDGEMENTS

Before proceeding, I want to express my gratitude to the people who helped me during my undergraduate studies. I am very grateful to all the professors for their contribution in my development. Each subject and discipline was very important for me. I learnt a lot. Special thanks to my friends Robert Yenokyan and Arman Zhamharyan who gave me a hand in critical and stressful moments. Above all I am very grateful to my supervisor, Professor Gayane Barseghyan for the knowledge she gave to us and of course for her patience.

All remaining errors in this paper are mine.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

Introduction

Forecasting of stock prices encompasses broad areas such as Economy, Business, Healthcare, Finance, etc. industries. There are several types of forecasting in terms of time span of forecast. In this paper it is discussed about short-term forecasting, i.e., forecasting of short-term objectives commonly less than one year. The variable is stock price. The data can be univariate and multivariate. Univariate data contains information of only one company's stock price. In this paper the results of the predictions of stock prices are made for APPLE Incorporation. By analyzing time series data, we can identify patterns, trend and cycles hidden in the data. It helps in capturing the best performing companies for a specific period. The linear models such as AR, ARMA, ARIMA have a big disadvantage as they don't take into account the volatility existing in the data. Not to mention, the model designed for one series may not necessarily fit the other one according to [1]. Thus, these linear models are not the best option to consider to capture patterns and dynamics in the data. There are other types of models such as non-linear models. In this paper there was given an APPLE INC historical stock price and made a stock price prediction. The work was done on one company's historical price data. Nonetheless, it gave insights to make a further research in the future in order to predict the stock prices of an index. To figure out which model performed better, I made a prediction on the same stock of the same period with two different models to conclude which one outperforms the other. The comparison was done with linear ARIMA model to see whether the non-linear LSTM model outperforms it or not.

Literature Review

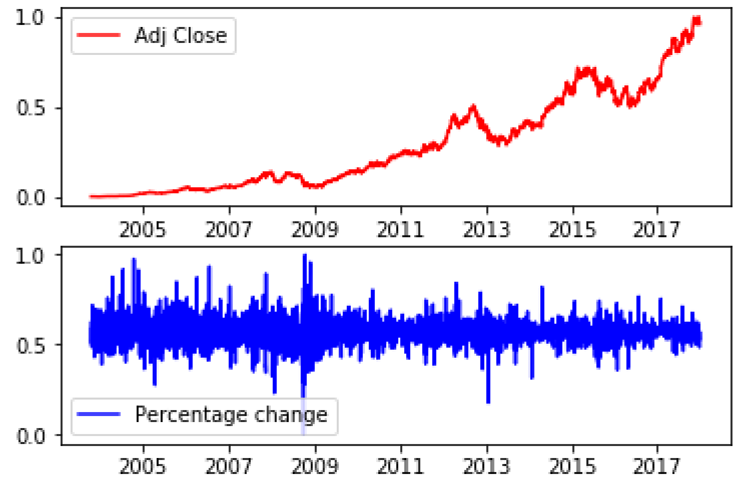
There are models designed to solve the aforementioned problems, such as dynamics, volatility and hidden patterns capture. Non-linear models such as ARCH, and models of GARCH families are bright examples. Deep learning models with various classes such as recurrent neural networks (RNN) with long short-term memory (LSTM) class of architectures, CNN (Convolutional Neural Networks) are applied in diverse disciplines and industries. These models are able to capture non-linear patterns and dynamics. [2] In stock market, the data is enormous and mostly non-linear. For modeling such data, we need models that can find and analyze the hidden patterns and dynamics. Deep learning algorithms are able to do it through a self-learning. They can give a good prediction by analyzing hidden patterns within data. [3]

Data

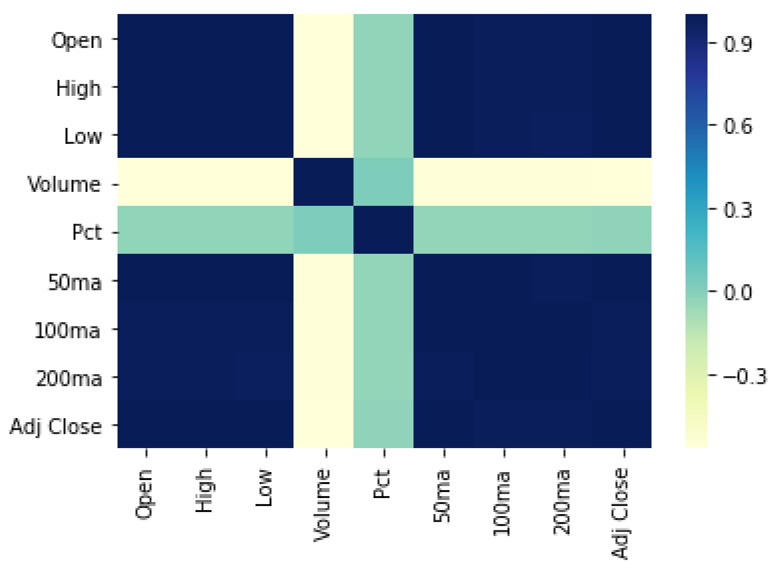
The data was obtained from Quandl (https://www.quandl.com/). It contains historical stock price daily data of APPLE Incorporation from January 1, 2003 until 31 December, 2017. 90% of the data was set for training and 10% for testing. During training of a model there was used different number of neurons, epochs and dropouts. In ARIMA model there was implemented auto-Arima package and it defined by itself the most optimal parameters for prediction.

Methodology and Results

First, LSTM model was designed to solve regression problem. I firstly visualized the data. Below on the graph, we have a look at a price change of APPLE INC stock between 2003 and 2017. On the first picture we see there is an increasing trend with ups and downs after 2009. On the second one we see percentage change in price.



Now we have a look at correlations between different moving averages (the algorithm used for smoothing short-term fluctuations and smoothing trend cycles), percentage change of prices, volume, open, high, low and adjusted close prices.



Then the data was divided into train and test. The training set was selected and trained on 5740 observations and only 1436 observations were tested. Result on the training set scored

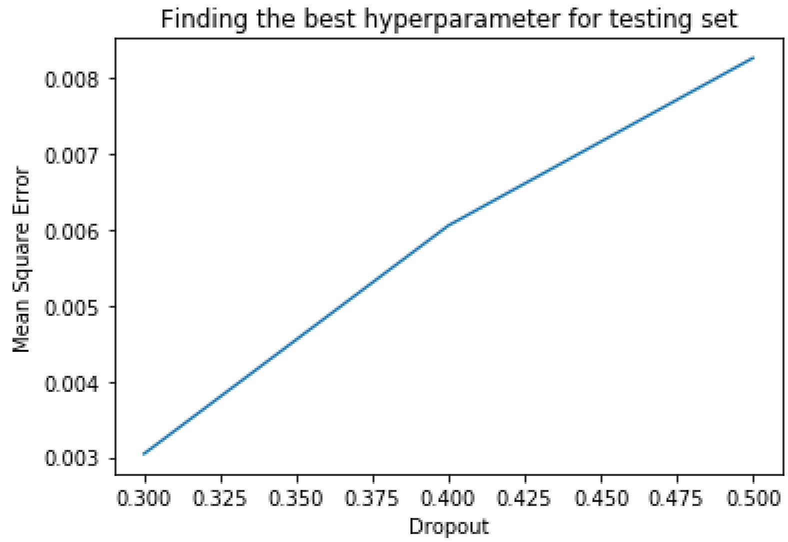
0.00025 MSE or 0.02 RMSE and 0.00817 MSE and 0.09 RMSE on test set. Then here is another plot showing us differenced then normalized values of predicted and real prices. After number of epochs were changed I did another train and test on the data. In this figure below it is shown the architecture of LSTM model. The best result was scored at 0.00381 MSE with 140 epochs taken. To avoid overfitting, different regularization techniques such as dropouts were applied. The best dropout rate was around 0.3.

```
Layer (type) Output Shape Param #

=================================================================

lstm_7 (LSTM) (None, 22, 256) 269312

_________________________________________________________________ dropout_7

(Dropout) (None, 22, 256) 0

_________________________________________________________________ lstm_8

(LSTM) (None, 256) 525312

_________________________________________________________________ dropout_8

(Dropout) (None, 256) 0

_________________________________________________________________ dense_7

(Dense) (None, 32) 8224

_________________________________________________________________ dense_8

(Dense) (None, 1) 33
```

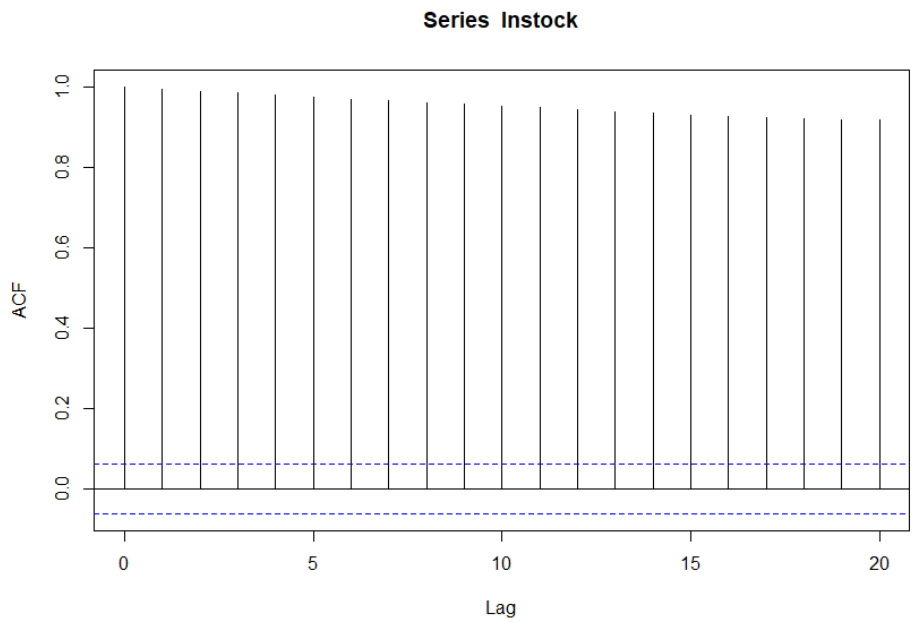
Total params: 802,881 Trainable params: 802,881 Non-trainable params: 0

____ Train Score:

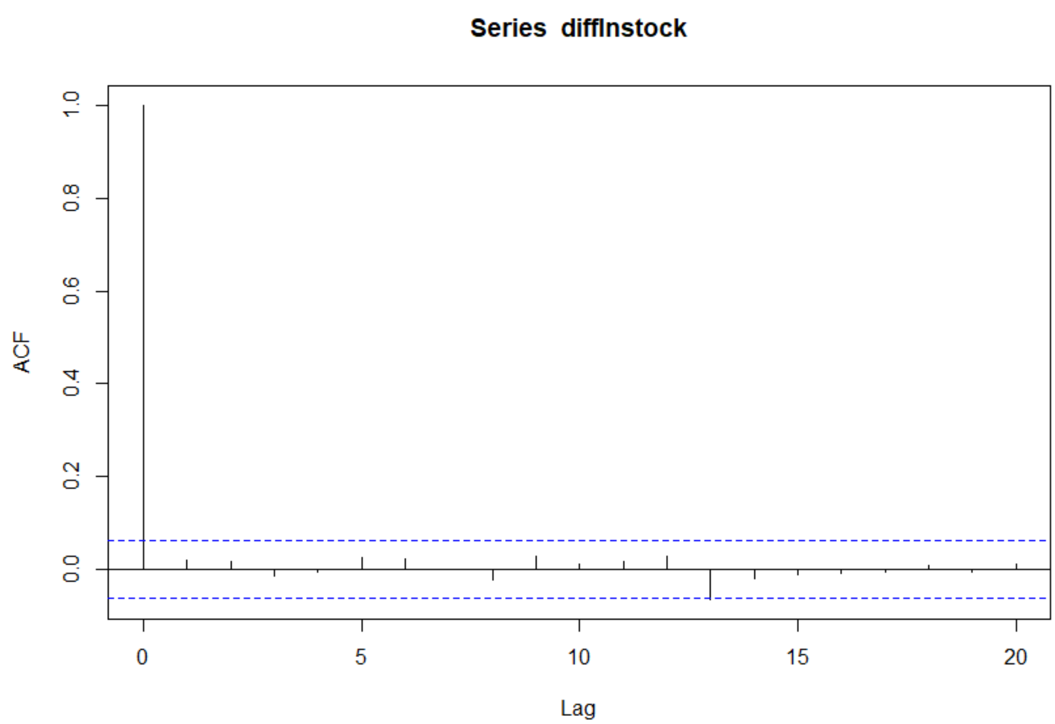
0.00009 MSE (0.01 RMSE) Test Score: 0.00381 MSE (0.06 RMSE)



While predicting stock price with ARIMA model, first, I tested whether it has strong auto and partial correlations. The initial data was represented in the actual price form. Then I transformed the data into log form and tested it for stationarity. It had a considerable correlation. According to augmented Dickey-Fuller test (ADF), the data was non-stationary. Below is the graph of autocorrelation function which shows there is a correlation in each lag.



I performed a differencing method. Then data was checked again for ADF test and there was no evidence of non-stationarity. Below in the figure you can see there was only strong correlation in the first lag. After execution of ARIMA model the achieved result stood at 0,015 MSE.



To assure there was no evidence of serial correlation of residuals, the Ljung-Box test was applied. According to the test, the alternative hypotheses of serial correlations of residuals was rejected. Below is the result of the test with different number of lags. P-values greater than 0.05 are the evidence that I cannot reject the Null Hypothesis of White Noise.

Box-Ljung test
data: fitlnstock$resid
X-squared = 3.7868, df = 5, p-value = 0.5805
> Box.test(fitlnstock$resid, lag=10, type="Ljung-Box")
Box-Ljung test
data: fitlnstock$resid
X-squared = 10.466, df = 10, p-value = 0.4006
> Box.test(fitlnstock$resid, lag=15, type="Ljung-Box")
Box-Ljung test
data: fitlnstock$resid
X-squared = 12.769, df = 15, p-value = 0.6201

Discussion

Although the data given in this research reflected only APPLE Incorporation stock prices prediction, it showed that deep learning LSTM model outperformed linear econometric model, in this case ARIMA. This evidence gives hints on making further deep researches to prove that statement. In the future research, I might include fundamental data as well. Thus, VAR model for multivariate time series analysis will be applied. To improve LSTM I might first perform CNN, which is so popular in image recognition, to reduce dimensionality and only then perform LSTM model. All these will be done on more companies.

Reference

1. Sreelekshmy Selvin, Vinayakumar R, E. R. Gopalakrishnan, Soman Kp, *"Stock price prediction using LSTM, RNN and CNN-sliding window model", (2017).*
2. Moritz and Zimmermann, *"Deep conditional portfolio sorts: The relation between past and future stock returns" (2014).*
3. *Sima Siami-Namini and Akbar Siami Namin, "Forecasting economics and financial time series: ARIMA vs. LSTM" (2018).*

Consent:

I hereby give my consent for this paper to be published for an open access in the American University of Armenia (AUA) library database.