Stock Price Forecasting - Comparing Time Series techniques and Machine Learning for Forecasting*

To Cheuk Sang Jackson[1]

[1] *University of Toronto, Physics Undergraduate*

Short-term prediction of stock market trend has potential application for personal investment without the high frequency-trading infrastructure. Forecasting the stock prices has been a difficult task for many of the researchers and analyst. For many investors, there have been plenty of interest in the research of stock price prediction. For a good and successful investments, many investors in keen in knowing the future situation of the stock market. By producing a good and effective solution to predict the stock market, will help benefit traders and investors by providing supportive information like the future direction of the stock market. I propose three methods of increasing complexity, which will be used to forecast Apple Inc Stock Data from 01-2015 to 03-2019. I attempt to forecast the stock data using advanced signal techniques, which will then be compared to Recurrent neural Network (RNN). Specifically, we will be using a Long Short-Term Memory unit to process the stock data.

Keywords: Least Square Regression, Fourier Analysis, Long Short-Term Memory

I. INTRODUCTION

There are a lot of complicated financial indicators and the fluctuations of the stock market is highly violent. As technology advances, the opportunity to gain a steady fortune from the stock market is ever increasing, and it also helps analyst find out the most informative indicators to make a better prediction. The prediction of the market value is important to help in maximizing the profits of stock option purchases which keeping the risk low. Least square fit is a mathematical regression analysis for finding the best-fitting curve to a given set of data points. It is often used to check whether a given set of data points fall roughly on the best fit line. The degree of least squares become more complex as the fit function becomes more complex. The linear least square will always provide us with a straight line which best fits the data points. Any n - degrees higher, the least square will provide us with a parabola of n-degree polynomials. Once a well fit polynomial is chosen, it can be used to forecast short term data.

The concept behind Fourier analysis is that any periodic signal can be broken down into a Taylor series or sum of suitably scaled sine and cosine waveforms. A key requirement here is that the signal being processed is periodic, which means that they repeat forwards and backwards to plus and minus infinity. The fast Fourier Transform (FFT) is an algorithm which uses convolution techniques to efficiently find the magnitude and location of the tones that make up the signal of interest. By filtering particular tones that make up the signal, we can obtain a smoother version of the underlying signal via the inverse fast Fourier transform (IFFT). It is possible to forecast data by filtering out low-amplitude, high-frequency components, using these individual components shift the phase forward to obtain a forecast and reconstruct them, then run an IFFT on the filtered data to obtain the forecasted data.

Recurrent neural networks (RNN) have proved to be one of the most powerful models for processing sequential data. Long Short-Term memory (LSTM) is one of the most successful RNN's architectures. LSTM introduces the memory cell, a unit of computation that replaces traditional artificial neurons in the hidden layer of the network. With these memory cells, networks are able to effectively associate memories and input remote in time, hence suit to grasp the structure of data dynamically over time with high prediction capacity.

This paper compares the 3 models by observing the forecast results of the three methods. I have collected 4 years of historical data of Apple Inc. and used it for the training and validation purposes for the various models. The next section of the paper will be the methodology where will explain about each process in detail. Afterwards, we will have visual representations of the analysis that we have used and a comparative analysis of the experimental results.

II. METHODOLOGY

This section will discuss the methodology of the three forecasting methods, along with how we will compare them. Our forecasting method each contain individual steps, but they can be generalized into several stages:

* University of Toronto PHY408 Final Project

Raw Data: In this stage the historical stock data is collected from https://www.nasdaq.com/symbol/aapl/historical and this historical data is used for the prediction of the future stock prices.

Data Preprocessing: The pre-processing stage involves:

- Data discretization
- Data Transformation: Normalization
- Data cleaning: Fill in missing values if any
- Data Integration : Integration of Data files

After the dataset is transformed into a clean dataset, the dataset is divided into training and testing sets to allow for evaluation. I have used a standard 80/20 split, 80% is used as training data, while 20% is used for evaluation, the evaluation values are taken as the more recent values.

Feature Extraction: In this stage, only the features which we want to be forecasted are chosen. For the least square regression this is meant by which curve best fits our testing data. For the FFT method, this I meant by picking out the frequencies spectrum which best represents the data, while removing any unwanted noise. For the RNN, we will be using the feature "close" from the dataset. This is the closing price of Apple Inc. for the given day.

Training: The least square regression and FFT methods do not have a training step, as the forecast are simple phase manipulations. For the RNN, the training set is fed to the neural network and trained for prediction, by assigning random biases and weights. The LSTM model used is composed of a sequential input layer followed by 2 LSTM layers and a dense output layer with ReLU activation.

Output Generation: In this layer, the outputs are collected and plotted for visualization. The output value is generated by the output layer of the RNN is compared with the target value. The error or the difference between the target and the obtained output value is minimized by using back propagation algorithm which adjusts the weights and the biases of the network.

III. ANALYSIS

To analyzing the efficiency of the system we used the Root Mean Square Error (RMSE). The error or the difference between the target and the obtained output value is minimized by using the RMSE value. The error or difference between the target and the obtained output value is minimized by using RMSE value. RMSE is the square root of the mean/average of the square of all of the error. The use of RMSE is highly common and it makes an excellent general purpose error metric for numerical predictions. Compared to the similar Mean Absolute Error, RMSE amplifies and severely punishes large errors.

$$RMSE = \sqrt{\frac{1}{N}\sum(\hat{Y} - Y)^2)}$$

IV. EXPERIMENTAL WORK

As mentioned above we will split the collected data into a training set and validation set. Fig. 1. shows the training set. Some initial impression is that the dataset has very violent fluctuations as expected of stock data. However, it seems from this plot that there is a general a cubic trend.

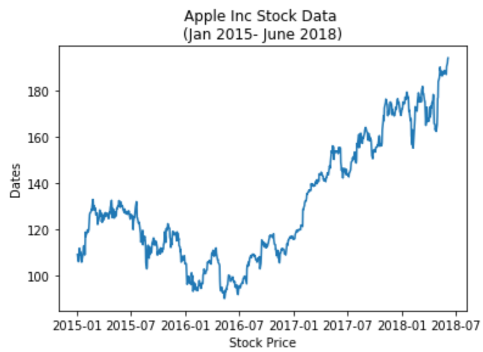


Fig. 1. Traning Set

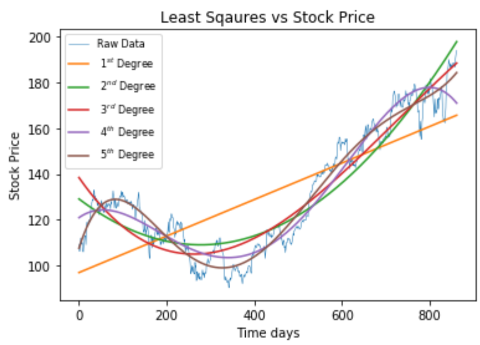


Fig. 2. Least Square Regression

The first method of forecasting is the least squares regression. This is a very simple operation, with a least square fit on different degrees of freedom. From Fig. 2, it is clear that a 5th degree least square fit best suited the training set. The quintic polynomial, is then used to forecast the remaining 20% of the dataset, by simply increasing the domain of the polynomial from 863 samples to the full 1067 samples. The forecast generated from this method is clearly very poor, it seems to expect the to continually increase in price, while in reality, Apple Inc's stock price drops. This forecast obtained a RMSE of 41.2, which further supports how big of a difference there is between the forecasted outcome and validation set.

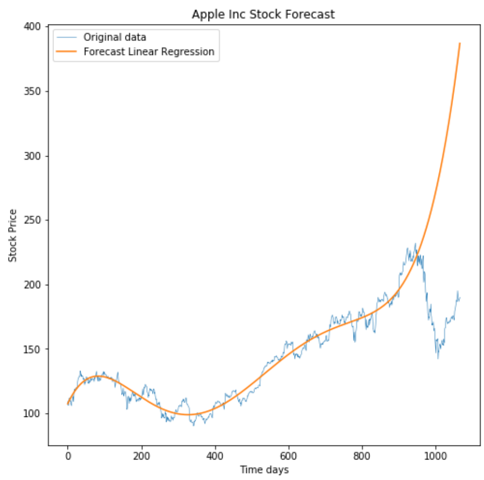


Fig. 3. Least Square Regression Forecast

The second method of forecasting is the FFT method. This method will require data manipulation to achieve the best possible forecast and can been seen as an extension to least squares regression. The trend identified by the least squares method is be removed from the data to isolate the seasonal, cyclical and irregular components of the stock data. An FFT algorithm is then applied to transform the stock data from time to frequency space. The data is then filtered by filtering out low-amplitude, high-frequency components. It seems there is a need to keep frequencies between -0.2 and 0.2 Hz, otherwise the data loses a lot of its original information. (See appendix). It is clear from this that the "noise" in the stock data, may see like random fluctuations but is actually multiple latent variables which affect the stock data. Other than setting the frequency spectrum, there was also an attempt to smooth out the stock data by applying a Hann and Gaussian window. However, it seems that both of these methods were not as effective as setting the frequency spectrum in retaining the original information. (see Appendix)

With the cleaned data, we are now able to pick out the individual sine and cosines waveforms which represent the stock data and shift the phase such the remaining 204 points are achieved. Although this method seemed to be more promising at first, it seems the forecasted price is not any better than a least square regression as evident from Fig. 3. Even with cleaned data, the prediction is barely better, which is indicative of the RMSE of 40.6. This is unsurprising as Fourier analysis works best periodic series, but financial series are aperiodic which explains why the forecast is so poor.

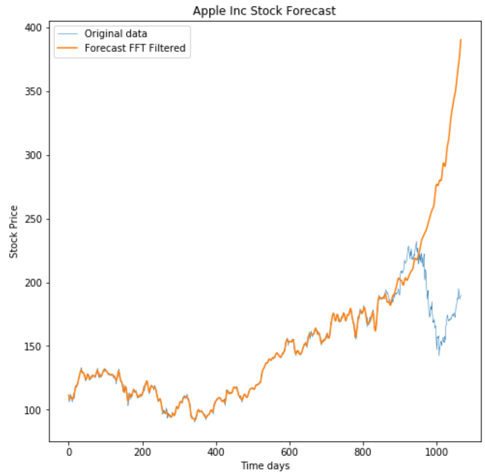


Fig. 4. FFT Forecast

The third method of forecasting is the LSTM model. In total there was 1067 sequences from 01-2015 to 03-2019. From this data set we used 863 samples for training and 204 samples for validation purposes. For training the model, Adaptive Moment Estimation (Adam) was used as the optimizer along with Keras and Tensorflow as the learning environment. *Adam is an optimization algorithm that can used instead of the classical stochastic gradient descent procedure to update network weights iterative based in training data.* For this experiment, a various set of parameters with different number of epochs was used to measure the RMSE of the Training and Testing Dataset. It was found that only 10 epochs were required to optimize the model. It is clear from the Fig. 5 that LSTM by far has the best forecasted price for Apple Inc's stock price, with a RMSE of 4.6. This is the lowest RMSE value out of the three methods indicating it's the most accurate of the three. It seems that the LSTM is able forecast stock prices very accurately and the forecasted price is almost the same validation set. The LSTM was also able to forecast any small fluctuations during this forecasting period.

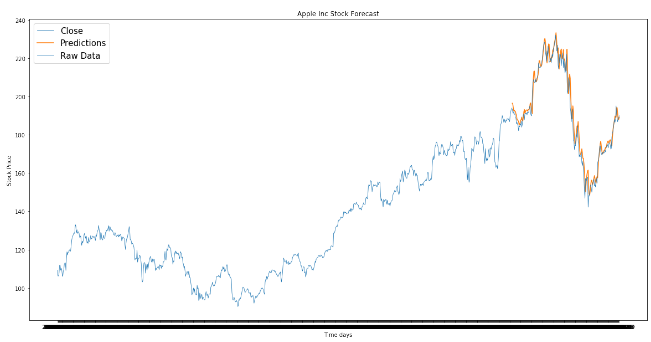


Fig. 5. LSTM Forecast

V. CONCLUSION

It is clear from this paper that least squares regression and fourier analysis have many difficulties when dealing with stock data. Although it seems, applying fourier transformation works fairly well to smoothen out the aggressive fluctuations. The two forecasting models was unable to provide a useful prediction, predicting that Apple's stock price would rise, where in reality it dropped. It was disappointing to see these two models failing to forecast the correct directional movement of the stock price. This indicates these methods would not be valuable to analyst/investors when implementing investing strategies. LSTM seems to be much more promising, providing an accurate prediction of the directional movement of the stock price, along with any small fluctuations.

The popularity of the stock market trading is becoming increasingly popular. There is a large demand for researchers to find new methods for the forecasting using new techniques. The forecasting methods used in this paper provides proof on how these techniques could help investors dealing with the stock market. In this paper, the Long Short-Term Memory unit of a Recurrent Neural Network is definitely the most promising forecast technique of the three. Its uncanny accuracy should be furthered investigated, but this paper has shown LSTM's can most definitely help investors and analyst in investing in the stock market by providing them with knowledge about the future situation of the stock market.

[1] Damian Faustryjak, Lidia Jackowska-Strumillo, and Michal Majchrowicz. Forward forecast of stock prices using lstm neural networks with statistical analysis of published messages. *2018 International Interdisciplinary PhD Workshop (IIPhDW)*, 2018.

[2] Roondiwala Harshal, Varma Murtaza, and Patel Shraddha. Predicting stock prices using lstm. *International Journal of Science and Research (IJSR)*, 6, Apr 2017.

[3] David Millán-Ruiz and J. Ignacio Hidalgo. Forecasting call centre arrivals. *Journal of Forecasting*, 32(7):628–638, 2013.

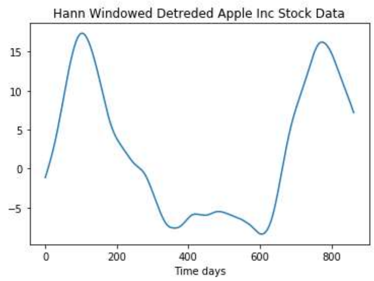
[4] David M. Q. Nelson, Adriano C. M. Pereira, and Renato A. De Oliveira. Stock markets price movement prediction with lstm neural networks. *2017 International Joint Conference on Neural Networks (IJCNN)*, 2017.

[5] Gonul Turhan-Sayan and Serdar Sayan. Use of time-frequency representations in the analysis of stock market data. *Applied Optimization Computational Methods in Decision-Making, Economics and Finance*, page 429–453, 2002.

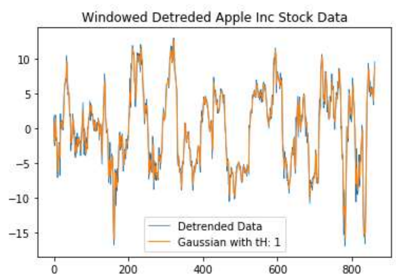
[6] Jung-Hua Wang and Jia-Yann Leu. Stock market trend prediction using arima-based neural networks. *Proceedings of International Conference on Neural Networks (ICNN96).*

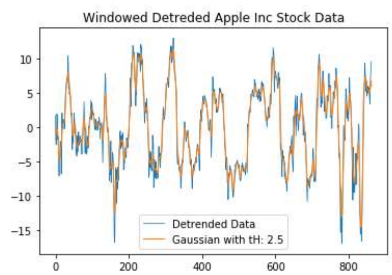
Appendix

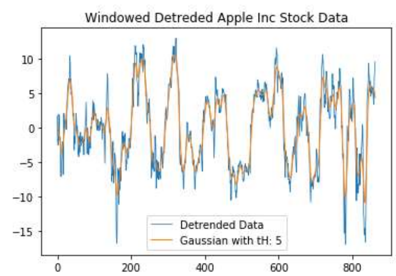
Hann Window



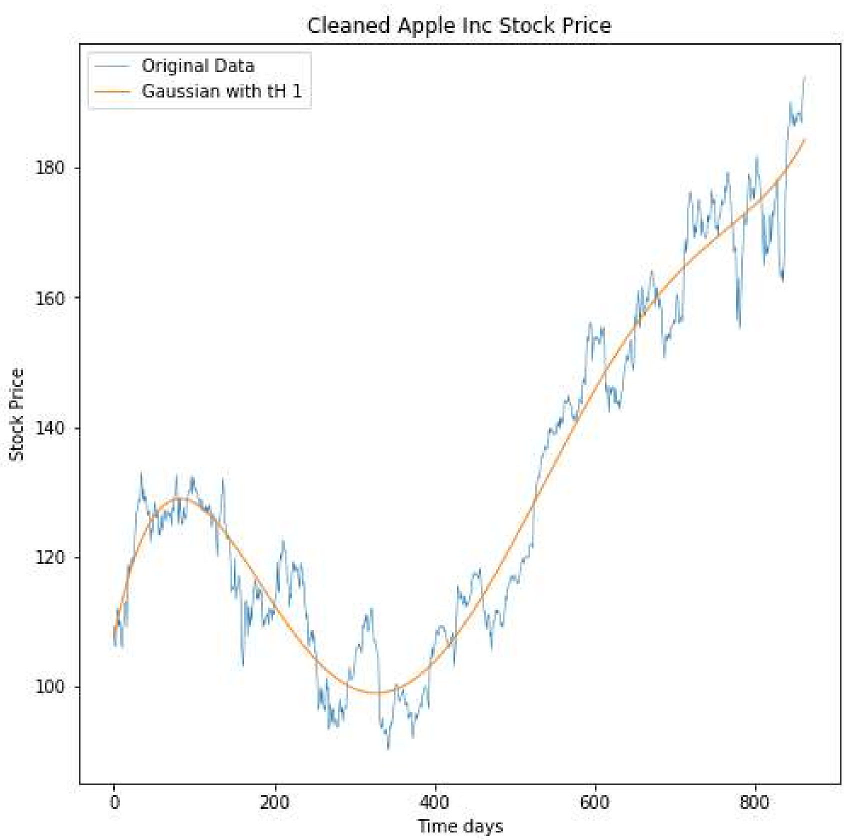
Gaussian Window



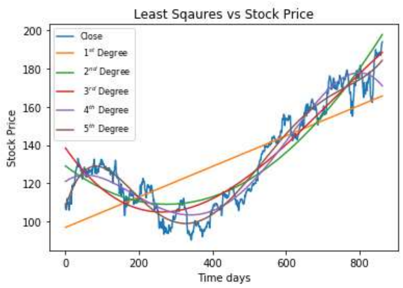


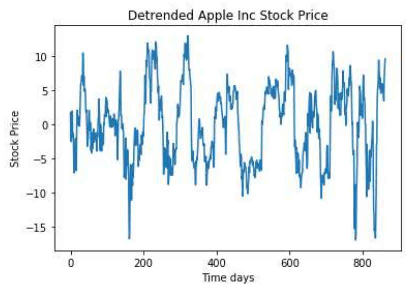


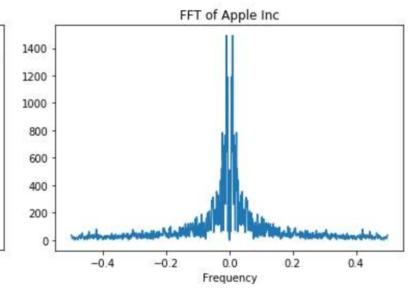
As you can see applying a gaussian window is not much better than doing a least square regression.

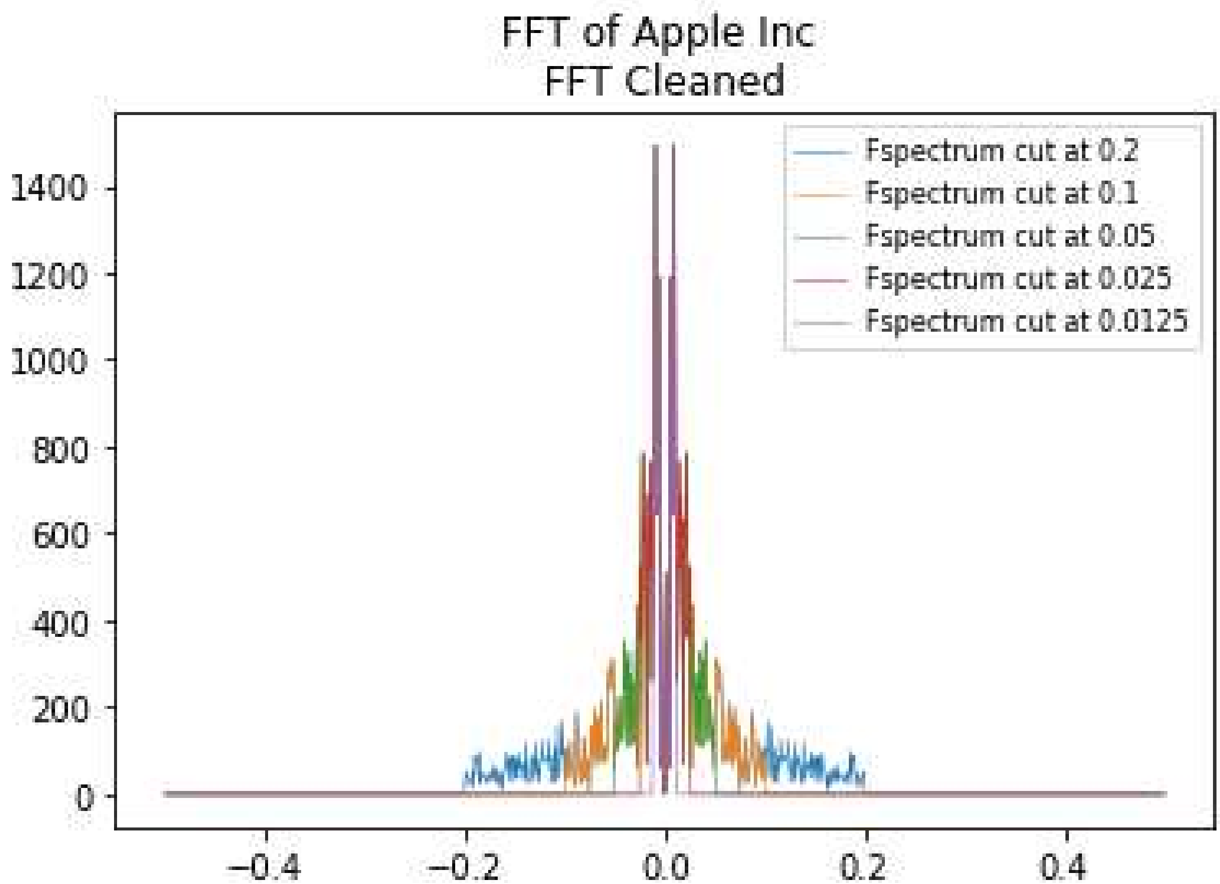


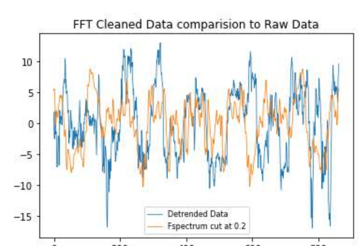
FFT Noise Cleaning

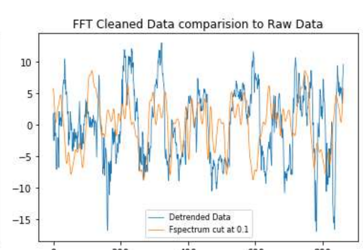


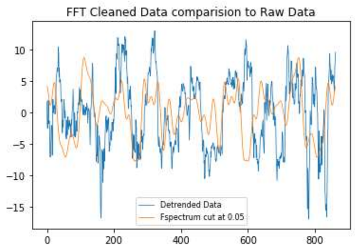


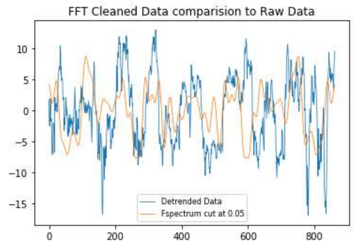


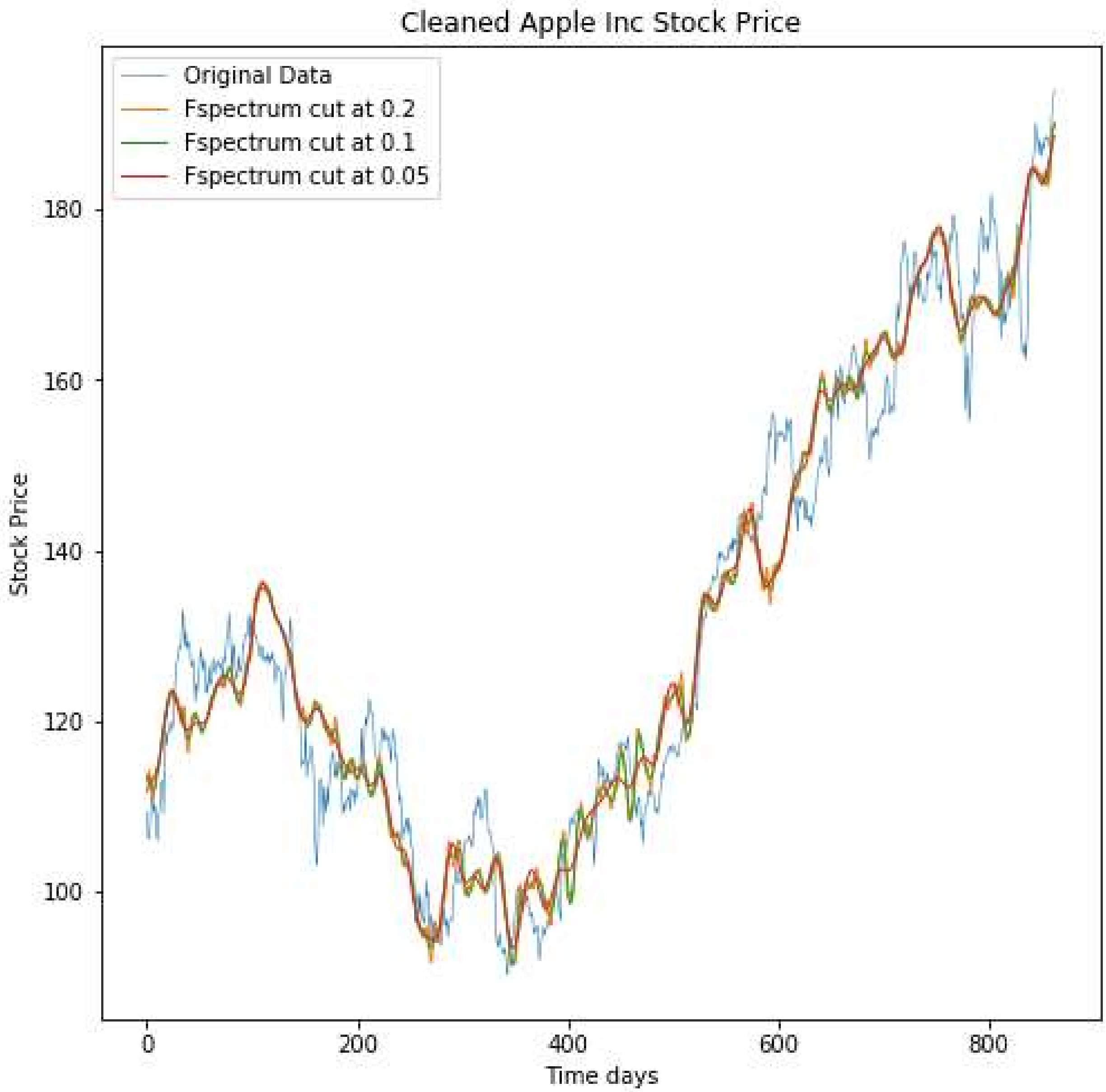












As you can see, setting the frequency spectrum between -0.05 to 0.05 losses a lot of the original information. Which indicate that the “noise” in the stock data all contribute to the financial series in some small way, therefore the frequency spectrum must be kept fairly wide.