

Stock Price

Stock prices are rather difficult to forecast. Proponents of efficient market hypotheses argue that stock prices cannot be predicted since the market prices have already reflected all known and expected fundamentals. However, precisely because the current market price reflects all relevant information, and therefore the current and historical prices are extremely useful for predicting the future prices since some relevant information is entering the stock markets continuously. It is well documented evidence that stock prices demonstrate strong momentum, short-term continuation and long-term reversal, and other identifiable patterns.

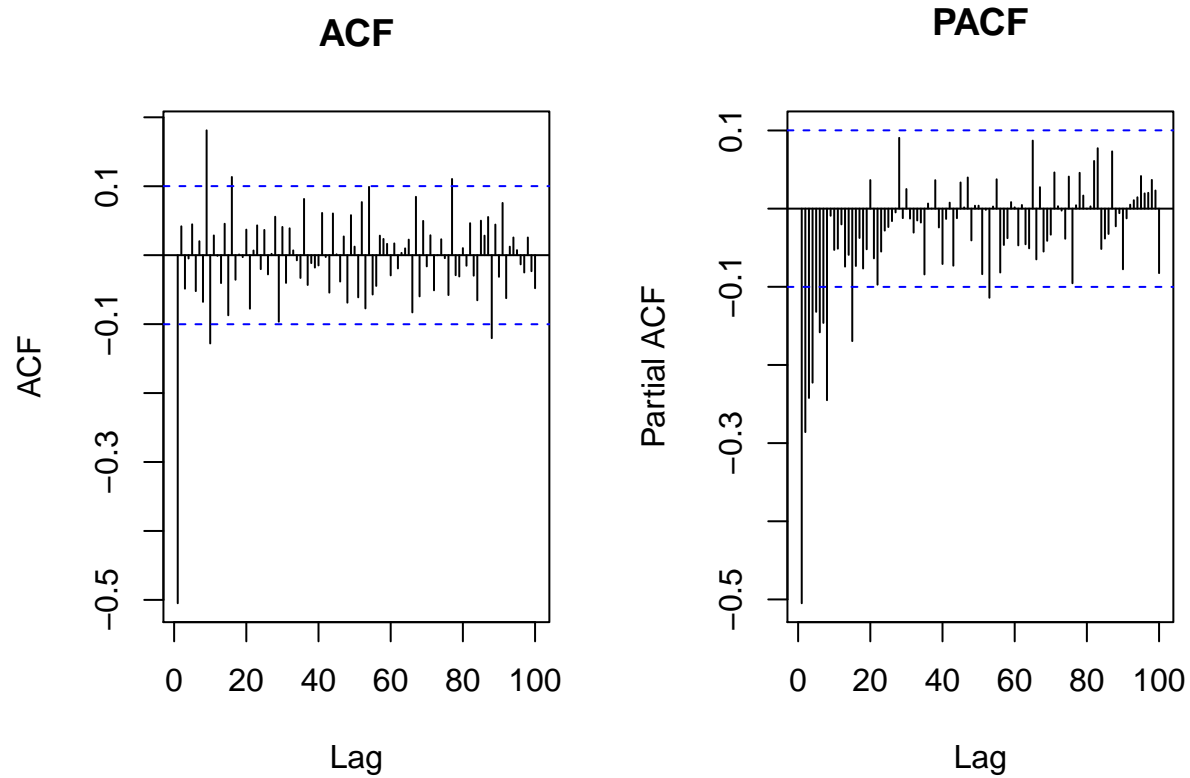
Time series Analysis: ARIMA, SARIMA and Spectral Analysis

A time series is a sequence of vectors, $x(t)$, $t = 0, 1, \dots$, where t represents elapsed time. For simplicity we will consider here only sequences of scalars, although the techniques considered generalise readily to vector series. Theoretically, x may be a value which varies continuously with t , such as a Closing price. In practice, for any given physical system, x will be sampled to give a series of discrete data points, equally spaced in time. The rate at which samples are taken dictates the maximum resolution of the model; however, it is not always the case that the model with the highest resolution has the best predictive power, so that superior results may be obtained by employing only every n th point in the series.

Time series are generally sequences of measurements of one or more visible variables of an underlying dynamic system, whose state changes with time as a function of its current state vector $u(t+1) = F(u(t))$. Such dynamic systems may evolve over time to an attracting set of points that is regular and of simple shape; any time series derived from such a system would also have a smooth and regular appearance. However another result is possible: the system may evolve to a chaotic attractor. Here, the path of the state vector through the attractor is non-periodic and because of this any time series derived from it will have a complex appearance and behaviour. In a real-world system such as a stock market, the nature and structure of the state space is obscure; so that the actual variables that contribute to the state vector are unknown or debatable. The task for a time series predictor can therefore be rephrased: given measurements of one component of the state vector of a dynamic system is it possible to reconstruct the (possibly) chaotic dynamics of the phase space and thereby predict the evolution of the measured variable?

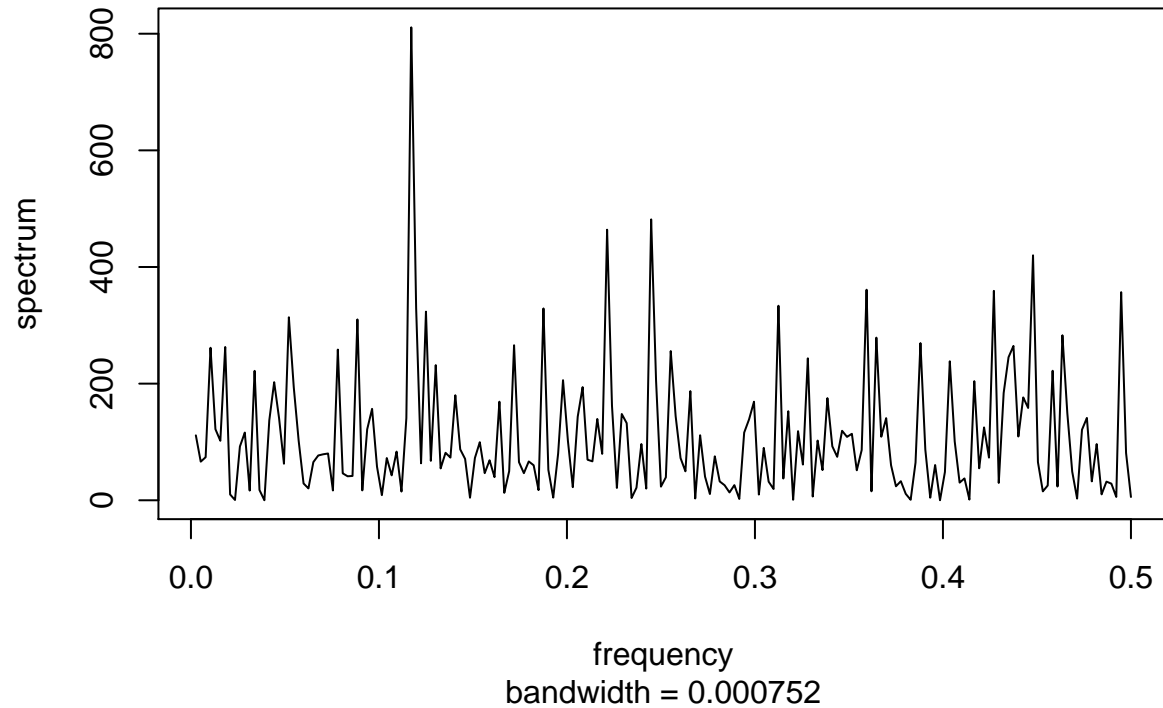
Example: Amazon For our interested stocks, we have conducted the time series analysis with similar steps to construct a single variable ARIMA model. A detailed exploration process will be shown as below.

Firstly, we check the ACF and PACF plots for the first-order or second-order difference of the original data. By observing the number of bars outside of the blue lines, which give the values beyond which the autocorrelations are significantly different from zero.



With this routine step, though there are very much unlikely for the financial data to show a pattern of seasonality, we have still tried to use spectral analysis to detect periodicity of the data by transferring the data from time-domain to frequency domain with the help of FFT(Fast Fourier Transform) and visualize the Raw Periodogram.

Raw Periodogram

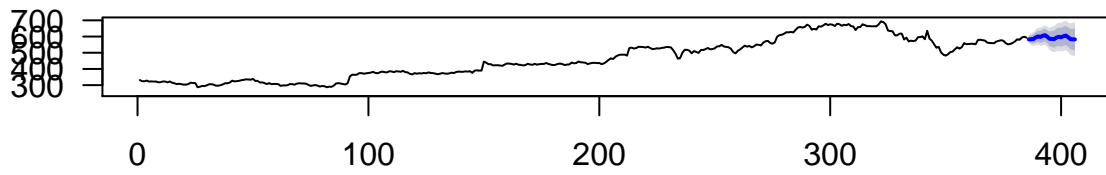


In this specific example, there is an obvious peak when the frequency is about 0.11. Thus a possible seasonality worth testing is about 8 or 9. It is worth noting that the rest of the stocks do not show such pattern. However, taking this into consideration would be an essential step to dig potential information.

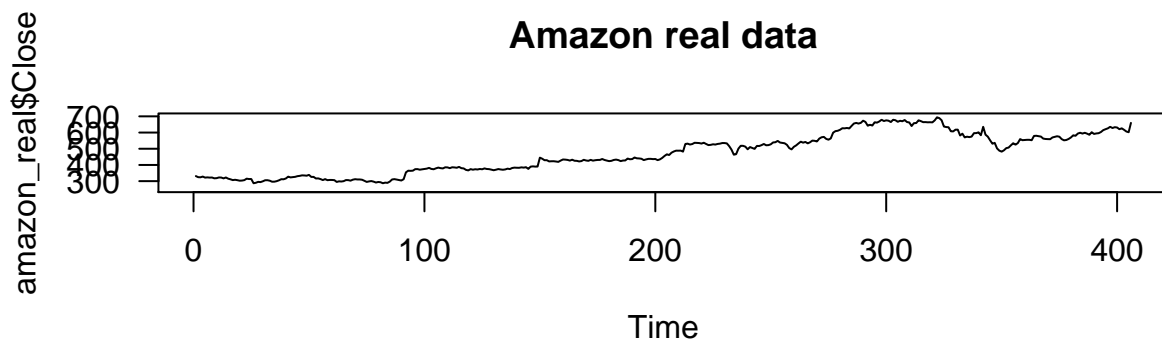
After comparing several SARIMA(Seasonal Autoregressive Integrated Moving Average) models based on their AIC, a best model is selected for making forecast. Here the best for Amazon is : $Arima(8, 0, 2) \times (0, 1, 0)_{(9)}$

The prediction plot can be seen as below with a shaded area of confidence interval of the forecast.

Amazon Forecast



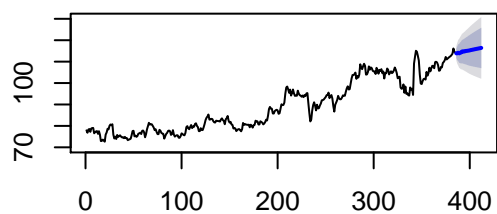
Amazon real data



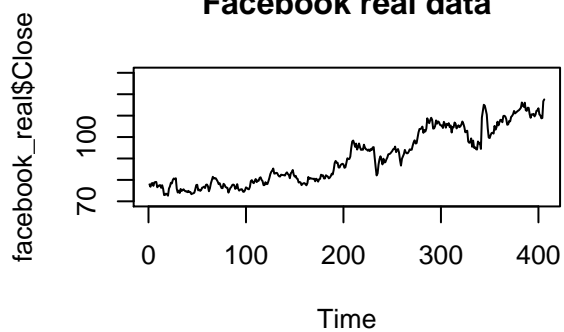
From the graph we can see that the SARIMA model shows a similar trend of the data. However, the MSE is 408.088.

Following the similar analysis procedure the rest of the forecasts and their MSEs are as below:

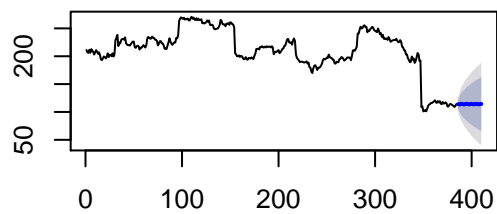
Facebook Forecast



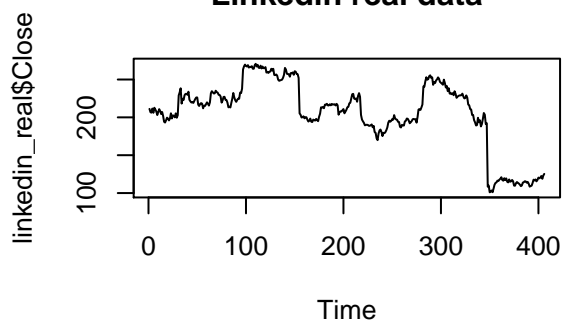
Facebook real data



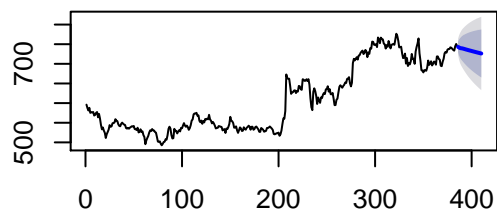
Linkedin Forecast



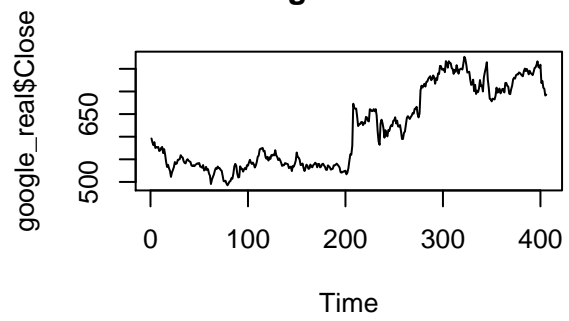
Linkedin real data



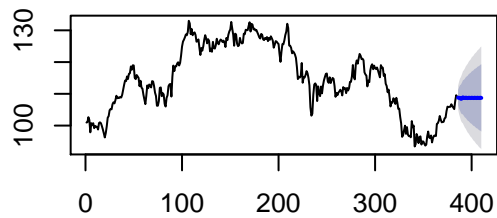
Google Forecast



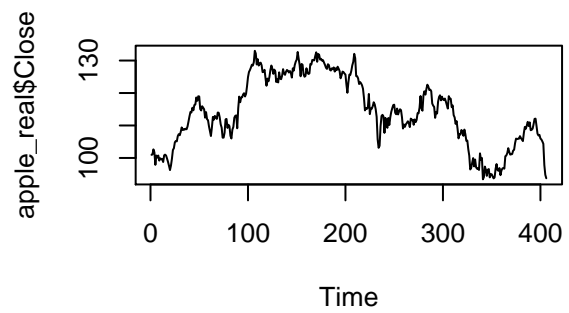
Google real data



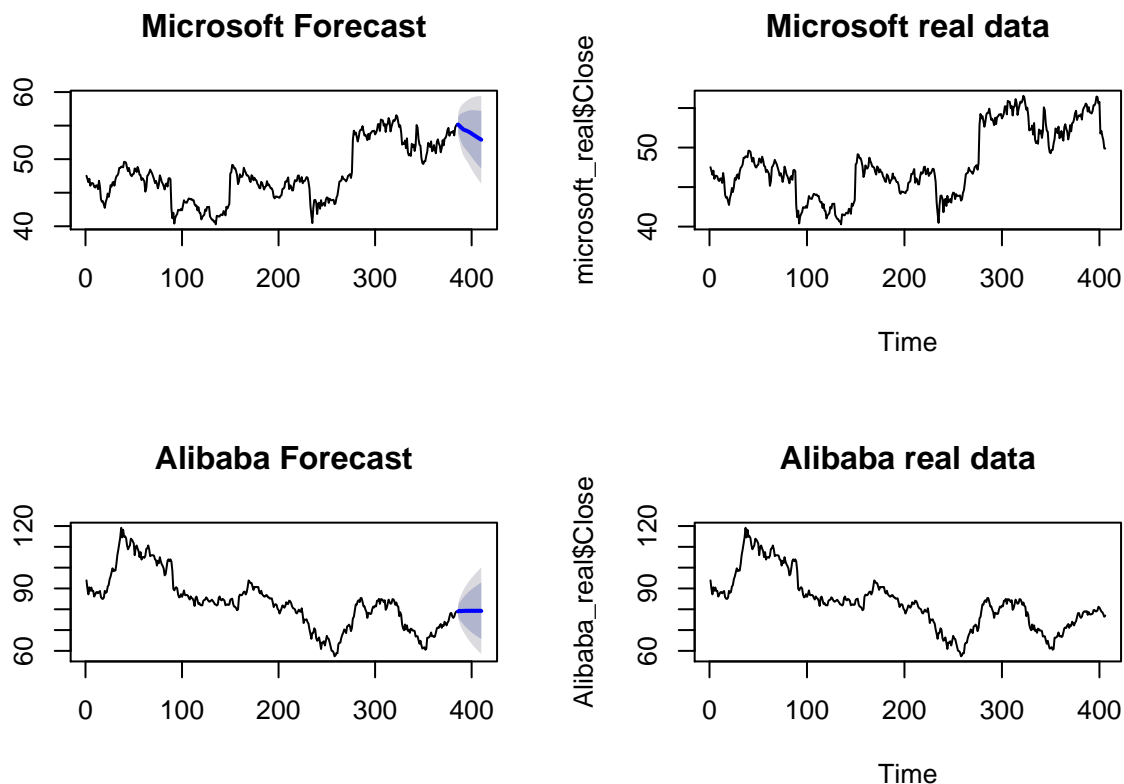
Apple Forecast



Apple real data



```
## time series starts 2014-09-19
```



Facebook: 11.405
 LinkedIn: 8.936
 Google: 234.878
 Apple: 3.895
 Microsoft: 2.027
 Alibaba: 2.239

Neural Network

Before training the models, the data is normalized before being input to the ANN. We have guaranteed the input vectors of the training data to have zero-mean and unit variance, while considering the activation function may be Unipolar sigmoid the target values are normalized between 0 and 1. We could also normalize them to between -1 and 1 and 0 and $\frac{1}{\sqrt{2\pi}\sigma}$ when the activation function is Bipolar sigmoid, Tan hyperbolic or RBF, but these go beyond the scope of this report and we stick to the Unipolar sigmoid activation function for “simplicity”. Similarly, the test data vector is also scaled by the same factors with which the training data was normalized. The output values from the ANN for this test vector are also scaled back with the same factor as the target values for the training data and plotted together with the real data for comparison.

In spite of all the features mentioned for neural networks, building a neural network for prediction is somehow complicated. In order to have a satisfactory performance, we must consider some crucial factors in designing such a prediction model. One of the main factors is the network structure including number of layers, neurons, and the connections. In the two proposed models, we followed the empirical rules when choosing the number of layer and neurons and test the model based on error.

In the first model, the inputs to the neural networks are the lowest, the highest and the open values in the d previous days. In the second model, the same variables are taken from the past few d days such that the data can be seen as longitudinal.

1. The first Neural Network model takes the data from a fixed period of time(January and February). The estimated network is shown as below.

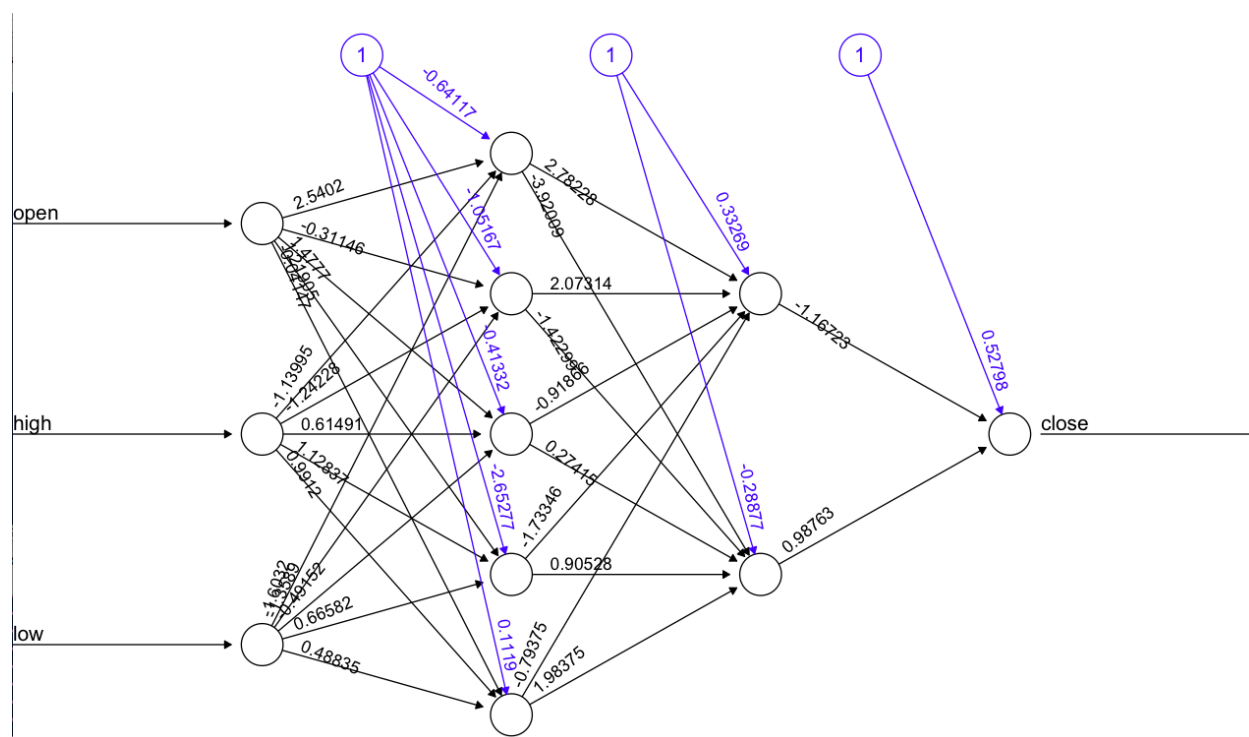
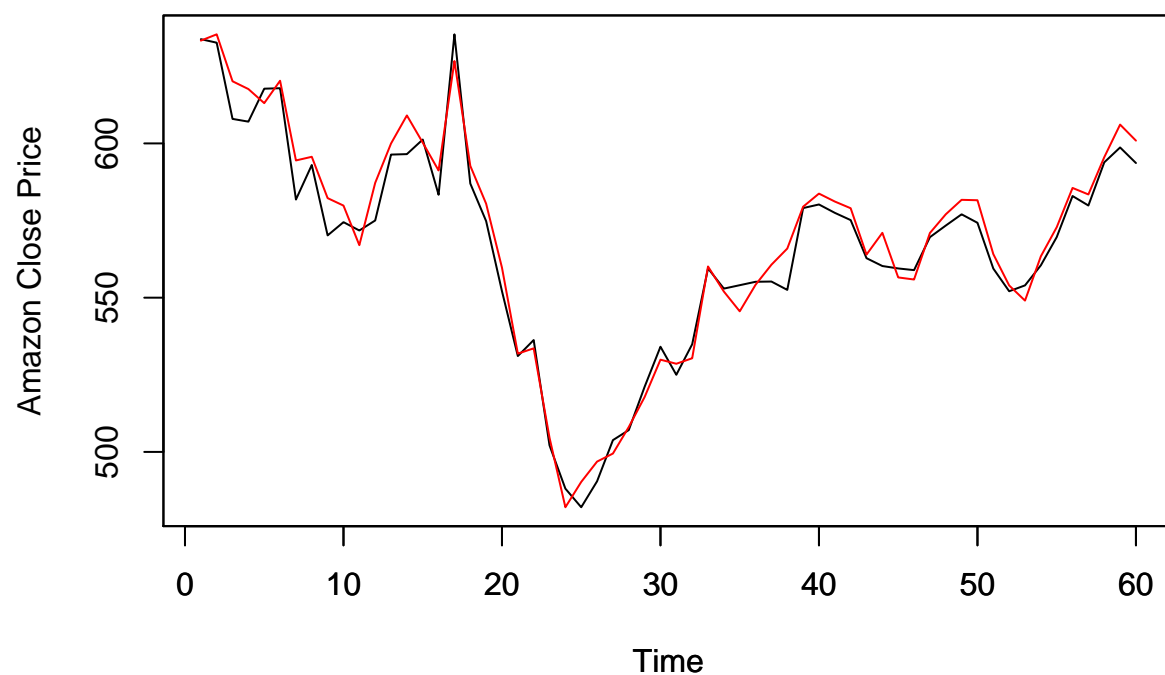


Figure 1:

With this training model, we put in the test data with the same variables for prediction(March and April). The true data is labeled as the black line and the red line is the predicted values. The pattern can be caught very well, but there may be more improvement.

Prediction based on Test data(fixed period)



Next, we try to analyze in terms of longitudinal data to improve prediction. The input will be all the variables in the past 3 periods of time (60 days). The best trained network is shown as below.

This time the red and black lines are almost overlapping. Thus by taking into account the data across a relatively long period of time, the prediction accuracy has been improved to a great extent.

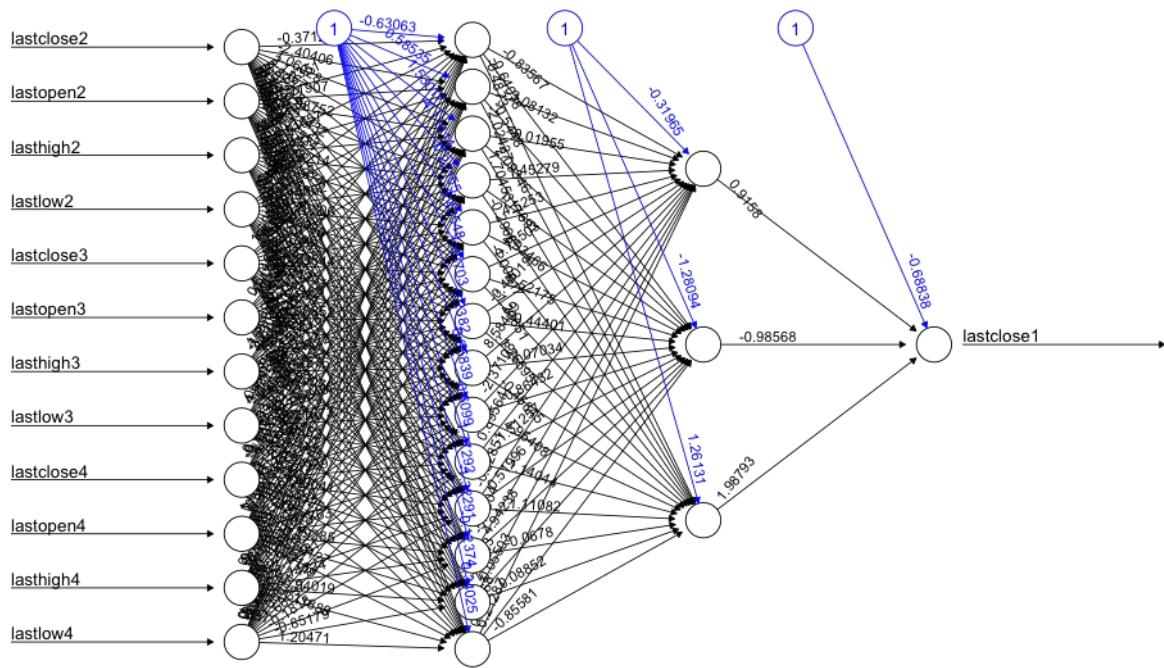
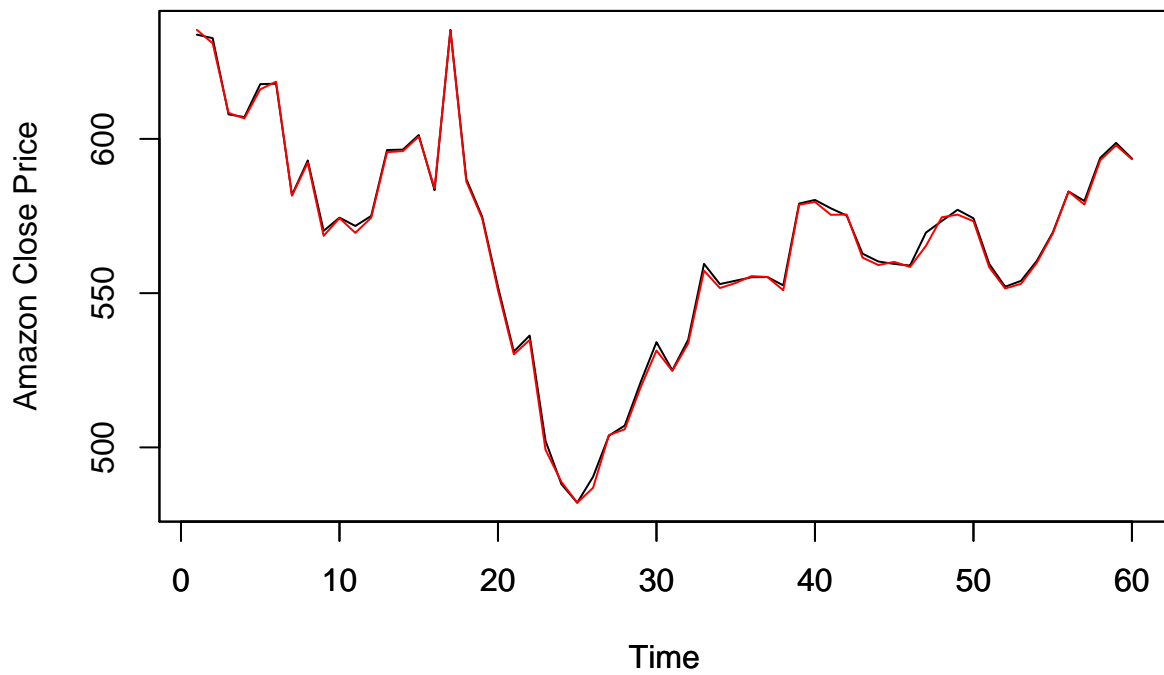


Figure 2:

Prediction based on past data(longitudinal)



Neural networks (NN) can assist in finding a relationship, or mapping, between inputs (Open, Low and High price) and an output (Closing price). Neural Networks have the advantage that can approximate any nonlinear functions without any apriori information about the properties of the data series. Comparing with traditional methods such as linear regression and logistic regression which are model based, Neural Networks are self-adjusting methods based on training data, so they have the ability to solve the problem with a little knowledge about its model and without constraining the prediction model by adding any extra assumptions. Besides, neural networks can find the relationship between the input and output of the system even if this relationship might be very complicated because they are general function approximators. Consequently, neural networks are well applied to the problems in which extracting the relationships among data is really difficult but on the other hand there exists a large enough training data sets. It should be mentioned that, although sometimes the rules or patterns that we are looking for might not be easily found or the data could be corrupted due to the process or measurement noise of the system, it is still believed that the inductive learning or data driven methods are the best way to deal with real world prediction problems. Furthermore, Neural Networks have generalization ability. After training, they can recognize the new patterns even if they haven't been in training set. Since in most of the pattern recognition problems predicting future events (unseen data) is based on previous data (training set), the application of neural networks would be very beneficial.

The application of neural networks in prediction problems is very promising due to some of their special characteristics. But we still need to be clear that the model is mostly strong in explaining about the prediction rather than forecasting. By implementing neural networks model in this report, it is further proved that including more variables other than merely the Closing price can be helpful. Also, it also shows a strong potential of making forecasting if we are able to get the required variables. Though this may require a complete reconstruction of the data frames and complexity of computation, it can be a promising model worth optimizing.

Artificial neural network: Artificial neural network (ANN) is a machine learning model to estimate the function of inputs. It is composed of many neuron-like nodes which have connections with each other.

We construct a model by utilizing the high price, low price, open price, closed price of last 3 time periods to predict all these prices of the next period. With this ANN model, stock price can be predicted by historical data. The inputs are the data of last 3 time periods and the outputs are the stock prices we would like to predict. Here we assume that the relationship between the last 3 time periods and the next time period is roughly stable so that we can utilize the model based on historical data to predict future data.