

Article

An Analysis and Implementation of the Hidden Markov Model to Technology Stock Prediction

Nguyet Nguyen

Faculty of Mathematics and Statistics, Youngstown State University, 1 University Plaza, Youngstown, OH 44555, USA; ntnguyen01@ysu.edu; Tel.: +1-330-941-1805; Fax: +1-330-941-3170

Academic Editor: Albert Cohen

Received: 20 April 2017; Accepted: 17 November 2017; Published: 24 November 2017

Abstract: Future stock prices depend on many internal and external factors that are not easy to evaluate. In this paper, we use the Hidden Markov Model, (HMM), to predict a daily stock price of three active trading stocks: Apple, Google, and Facebook, based on their historical data. We first use the Akaike information criterion (AIC) and Bayesian information criterion (BIC) to choose the numbers of states from HMM. We then use the models to predict close prices of these three stocks using both single observation data and multiple observation data. Finally, we use the predictions as signals for trading these stocks. The criteria tests' results showed that HMM with two states worked the best among two, three and four states for the three stocks. Our results also demonstrate that the HMM outperformed the naïve method in forecasting stock prices. The results also showed that active traders using HMM got a higher return than using the naïve forecast for Facebook and Google stocks. The stock price prediction method has a significant impact on stock trading and derivative hedging.

Keywords: hidden Markov model; stock prices; observations; states; predictions; AIC; BIC; likelihood; trading

1. Introduction

Stock investments can have a huge return or a significant loss due to the high volatilities of stock prices. An adaptable stock price prediction model would reduce risk and enhance potential return in financial derivative trading. Recently, researchers have applied the hidden Markov model for stock prices' forecasts. [Hassan and Nath \(2005\)](#) used HMM to predict the stock price for interrelated markets. [Kritzman, Page, and Turkington \(Kritzman et al. 2012\)](#) applied HMM with two states to predict regimes in market turbulence, inflation, and industrial production index. [Guidolin and Timmermann \(2006\)](#) used HMM with four states and multiple observations to study asset allocation decisions based on regime switching in asset returns. [Nguyen \(2014\)](#) used HMM with both single and multiple observations to forecast economic regimes and stock prices. [Nobakht, Joseph and Loni \(Nobakht et al. 2012\)](#) implemented HMM using various observation data (open, close, low, high) prices of stock to predict its close prices. In our previous work [Nguyen and Nguyen \(2015\)](#), we used HMM for single observation sequence for the S&P 500 to select stocks for trading based on performances of these stocks during the predicted regimes. In this study, we use HMM to predict stock prices and apply the results to trade stocks. We use HMM for multiple independent observation sequences in this study. Three stocks: Apple Inc., Alphabet Inc., and Facebook, Inc., were chosen to implement the model. We limit numbers of states of the HMM to a maximum of four states and use two goodness of fit tests to choose the best HMM model among HMMs with two, three, or four states. The prediction process is based on the work of [Hassan and Nath \(2005\)](#). The authors use HMM with the four observations: close, open, high, and low prices of some airline stocks to predict their future close price using four states. They used HMM to find a day in the past that was similar to

the recent day and used the price change in date and price of the current day to predict future close price. However, in the paper, the authors did not explain why they chose HMM with four states. Our approach is different from their work in the three following modifications. The first difference is that we use the Akaike information criterion (AIC) and Bayesian information criterion (BIC) to test the HMM's performances with numbers of states from two to four to find the best HMM model. The second modification is that we apply HMM for stock returns to predict future close prices and compare the results with the naïve forecast method. The modification is based on the assumption of the HMM's algorithms presented in this paper: the observation sequences are independent. Applying the HMM to stock returns, our prediction method is simpler than the method in [Hassan and Nath \(2005\)](#), which will be explained in Section 3.1. Finally, we use stock prices predicted via the HMM and the naïve method to trade these three stocks and compare the results.

The paper is organized as follows: Section 2 gives a brief introduction to HMM and its algorithms for multiple observation sequences. Section 3 describes the HMM model selections and data collections for stock price prediction. Section 4 presents the results of stock price predictions and stock trading, and Section 5 gives conclusions.

2. Hidden Markov Model and Its Algorithms

The Hidden Markov Model, HMM, is a signal detection model that was introduced in 1966 by Baum and Petrie ([Baum and Petrie 1966](#)). HMM assumes that an observation sequence was derived from a hidden state sequence of discrete data and satisfies the first order of a Markov process. HMM was developed from a model for a single observation variable to a model for multiple observation variables. The applications of HMM also were expanded to many areas such as speech recognition, biomathematics, and financial mathematics. In our previous paper [Nguyen and Nguyen \(2015\)](#), we described HMM for one observation, its algorithms, and applications. In this section, we present HMM for multiple observations and its corresponding algorithms. We assume that the multiple observations data are independent and have the same length. The basic elements of an HMM for multiple observations are:

- Observation data, $O = \{O_t^{(l)}, t = 1, 2, \dots, T, l = 1, 2, \dots, L\}$, where l is numbers of independent observation sequences and T is the length of each sequence,
- Hidden state sequence of O , $Q = \{q_t, t = 1, 2, \dots, T\}$,
- Possible values of each state, $\{S_i, i = 1, 2, \dots, N\}$,
- Possible symbols per state, $\{v_k, k = 1, 2, \dots, M\}$,
- Transition matrix, $A = (a_{ij})$, where $a_{ij} = P(q_t = S_j | q_{t-1} = S_i)$, $i, j = 1, 2, \dots, N$,
- Initial probability of being in state (regime) S_i at time $t = 1$, $p = (p_i)$, where $p_i = P(q_1 = S_i)$, $i = 1, 2, \dots, N$,
- Observation probability matrix, $B = \{b_i(k)\}$, where

$$b_i(k) \equiv b_i(O_t = v_k) \equiv P(O_t = v_k | q_t = S_i), \quad i = 1, 2, \dots, N, \quad k = 1, 2, \dots, M.$$

Parameters of an HMM are the matrices A and B and the vector p . For convenience, we use a compact notation for the parameters, given by

$$\lambda \equiv \{A, B, p\}.$$

If the observation probability assumes the Gaussian distribution, then we have a continuous HMM with $b_i(k) = b_i(O_t = v_k) = \mathcal{N}(v_k, \mu_i, \sigma_i)$, where μ_i and σ_i are the mean and variance of the distribution corresponding to the state S_i , respectively, and \mathcal{N} is Gaussian density function. For convenience, we write $b_i(O_t = v_k)$ as $b_i(O_t)$. Then, the parameters of HMM are

$$\lambda \equiv \{A, \mu, \sigma, p\},$$

where μ and σ are vectors of means and variances of the Gaussian distributions, respectively. With the assumption that the observations are independent, the probability of observation, denoted by $P(O|\lambda)$, is

$$P(O|\lambda) = \prod_{l=1}^L P(O^{(l)}|\lambda).$$

There are three main questions that readers should consider when using the HMM:

1. Given an observation data O and the model parameters λ , can we compute the probabilities of the observations $P(O|\lambda)$?
2. Given the observation data O and the model parameters λ , can we find the best hidden state sequence of O ?
3. Given the observation O , can we find the model's parameters λ ?

The first problem can be solved by using forward or backward algorithms [Baum and Egon \(1967\)](#); [Baum and Sell \(1968\)](#), the second problem was solved by using Viterbi algorithm [Forney \(1973\)](#); [Viterbi \(1967\)](#) and the Baum–Welch algorithm [Rabiner \(1989\)](#) was developed to solve the last problem. In the paper, we only use the algorithms to solve the first and the last problem. We first use the Baum–Welch algorithm to calibrate parameters for the model and the forward algorithm to calculate the probability of observation to predict trending signals for stocks. In this section, we introduce the forward algorithm and the Baum–Welch algorithm for HMM with multiple observations. These algorithms are written based on [Baum and Egon \(1967\)](#); [Baum and Sell \(1968\)](#); [Forney \(1973\)](#); [Petrushin \(2000\)](#); [Rabiner \(1989\)](#).

2.1. Forward Algorithm

We define the joint probability function as

$$\alpha_t^{(l)}(i) = P(O_1^{(l)}, O_2^{(l)}, \dots, O_t^{(l)}, q_t = S_i | \lambda), t = 1, 2, \dots, T \text{ and } l = 1, 2, \dots, L.$$

Then, we calculate $\alpha_t^{(l)}(i)$ recursively. The probability of observation $P(O^{(l)}|\lambda)$ is just the sum of the $\alpha_T^{(l)}(i)$'s.

The forward algorithm

1. Initialization $P(O|\lambda) = 1$
2. For $l = 1, 2, \dots, L$ do

(a) Initialization: for $i = 1, 2, \dots, N$

$$\alpha_{t=1}^{(l)}(i) = p_i b_i(O_1^{(l)}).$$

(b) Recursion: for $t = 2, 3, \dots, T$, and for $j = 1, 2, \dots, N$, compute

$$\alpha_t^{(l)}(j) = \left[\sum_{i=1}^N \alpha_{t-1}^{(l)}(i) a_{ij} \right] b_j(O_t^{(l)}).$$

(c) Calculate:

$$P(O^{(l)}|\lambda) = \sum_{i=1}^N \alpha_T^{(l)}(i).$$

(d) Update:

$$P(O|\lambda) = P(O|\lambda) * P(O^{(l)}|\lambda).$$

3. Output: $P(O|\lambda)$.
-

2.2. Baum–Welch Algorithm

The Baum–Welch algorithm is an algorithm to calibrate parameters for the HMM given the observation data. The algorithm was introduced in 1970 [Baum et al. \(1970\)](#), in order to estimate the parameters of HMM for a single observation. Then, in 1983, the algorithm was extended to calibrate HMM's parameters for multiple independent observations, [Levinson et al. \(1983\)](#). In 2000, the algorithm was developed for multiple observations without the assumption of independence of the observations, [Li et al. \(2000\)](#). In this paper, we use HMM for independent observations, so we will introduce the Baum–Welch algorithm for this case. The Baum–Welch method or the expectation modification (EM) method is used to find a local maximizer, λ^* , of the probability function $P(O|\lambda)$.

In order to describe the procedure, we define the conditional probability $\beta_t^{(l)}(i) = P(O_{t+1}^{(l)}, O_{t+2}^{(l)}, \dots, O_T^{(l)} | q_t = S_i, \lambda)$, for $i = 1, \dots, N, l = 1, 2, \dots, L$. Obviously, for $i = 1, 2, \dots, N$ $\beta_T^{(l)}(i) = 1$, and we have the following backward recursive:

$$\beta_t^{(l)}(i) = \sum_{j=1}^N a_{ij} b_j(O_{t+1}^{(l)}) \beta_{t+1}^{(l)}(j), \quad t = T-1, T-2, \dots, 1.$$

We then defined $\gamma_t^{(l)}(i)$, the probability of being in state S_i at time t of the observation $O^{(l)}$, $l = 1, 2, \dots, L$ as:

$$\gamma_t^{(l)}(i) = P(q_t = S_i | O^{(l)}, \lambda) = \frac{\alpha_t^{(l)}(i) \beta_t^{(l)}(i)}{P(O^{(l)} | \lambda)} = \frac{\alpha_t^{(l)}(i) \beta_t^{(l)}(i)}{\sum_{i=1}^N \alpha_t^{(l)}(i) \beta_t^{(l)}(i)}.$$

The probability of being in state S_i at time t and state S_j at time $t+1$ of the observation $O^{(l)}$, $l = 1, 2, \dots, L$ as:

$$\zeta_t^{(l)}(i, j) = P(q_t = S_i, q_{t+1} = S_j | O^{(l)}, \lambda) = \frac{\alpha_t^{(l)}(i) a_{ij} b_j(O_{t+1}^{(l)}) \beta_{t+1}^{(l)}(j)}{P(O^{(l)}, \lambda)}.$$

Clearly,

$$\gamma_t^{(l)}(i) = \sum_{j=1}^N \zeta_t^{(l)}(i, j).$$

Note that the parameter λ^* was updated in Step 2 of the Baum–Welch algorithm to maximize the function $P(O|\lambda)$ so we will have $\Delta = P(O, \lambda^*) - P(O, \lambda) > 0$.

If the observation probability $b_i(k)^*$, defined in Section 2, is Gaussian, we will use the following formula to update the model parameter, $\lambda \equiv \{A, \mu, \sigma, p\}$

$$\mu_i^* = \frac{\sum_{l=1}^L \sum_{t=1}^{T-1} \gamma_t^{(l)}(i) O_t^{(l)}}{\sum_{l=1}^L \sum_{t=1}^{T-1} \gamma_t^{(l)}(i)}$$

$$\sigma_i^* = \frac{\sum_{l=1}^L \sum_{t=1}^T \gamma_t^{(l)}(i) (O_t^{(l)} - \mu_i) (O_t^{(l)} - \mu_i)'}{\sum_{l=1}^L \sum_{t=1}^T \gamma_t^{(l)}(i)}.$$

3. Model Selections and Data Collections

The Hidden Markov Model has been widely used in financial mathematics area to predict economic regimes ([Kritzman et al. 2012](#); [Guidolin and Timmermann 2006](#); [Ang and Bekaert 2002](#); [Chen 2005](#); [Nguyen 2014](#)) or predict stock prices ([Hassan and Nath 2005](#); [Nobakht et al. 2012](#); [Nguyen 2014](#)). In this paper, we explore a new approach of HMM in predicting stock prices. In this section, we discuss how to use the Akaike information criterion, AIC, and the Bayesian information criterion, BIC, to test the HMM's performances with different numbers of states. We then will present

how to use HMM to predict stock prices and apply the results to trade stocks. First, we will describe the chosen data and the AIC and BIC for HMM with selected numbers of states.

Baum–Welch for L independent observations $O = (O^{(1)}, O^{(2)}, \dots, O^{(L)})$ with the same length T

1. Initialization: input parameters λ , the tolerance tol , and a real number Δ
2. Repeat until $\Delta < tol$
 - Calculate $P(O, \lambda) = \prod_{l=1}^L P(O^{(l)} | \lambda)$ using the forward algorithm (??)
 - Calculate new parameters $\lambda^* = \{A^*, B^*, p^*\}$, for $1 \leq i \leq N$

$$p_i^* = \frac{1}{L} \sum_{l=1}^L \gamma_1^{(l)}(i)$$

$$a_{ij}^* = \frac{\sum_{l=1}^L \sum_{t=1}^{T-1} \xi_t^{(l)}(i, j)}{\sum_{l=1}^L \sum_{t=1}^{T-1} \gamma_t^{(l)}(i)}, \quad 1 \leq j \leq N$$

$$b_i(k)^* = \frac{\sum_{l=1}^L \sum_{t=1}^T \mathbb{1}_{O_t^{(l)}=v_k} \gamma_t^{(l)}(i)}{\sum_{l=1}^L \sum_{t=1}^T \gamma_t^{(l)}(i)}, \quad 1 \leq k \leq M$$

- Calculate $\Delta = P(O, \lambda^*) - P(O, \lambda)$
- Update

$$\lambda = \lambda^*.$$

3. Output: parameters λ .
-

3.1. Overview of Data Selections

We chose three stocks that are actively trading in the stock market to examine our model: Apple Inc. (AAPL), Alphabet Inc. (GOOGL), and Facebook Inc. (FB). The daily stock prices (open, low, high, close) of these stocks and information of these companies can be found from finance.yahoo.com. We used daily historical prices of these stocks from 4 January 2010 to 30 October 2015 in this paper.

3.2. Checking Model Assumptions

The HMM’s algorithms presented in this paper are based on the assumption that the observation sequences are independent. However, the open, low, high, and close prices of a stock are highly correlated, which can be seen from the matrix of correlation in Figure 1. On the other hand, stock returns of these four series prices are independent, which are shown in Figure 2.

We use the Autocorrelation function (ACF) to calculate the paired correlation between the series and plot in Figures 1 and 2. The ACF for the Facebook and Google stocks are presented in Appendix A. We can see clearly from the figures that the return price series have low correlations while the stock price series have very high correlations.

Furthermore, we conduct the Ljung–Box test to test the independence of each time series. We use the test with $lag = 1$ for returns of the three stocks: AAPL, FB, and GOOGL, from 1 October 2014 to 1 October 2015, and present results in Table 1. Note that the stock prices are not independent, and they failed the Ljung–Box test at significance level $\alpha = 5\%$, so Table 1 only displays results for stock returns.

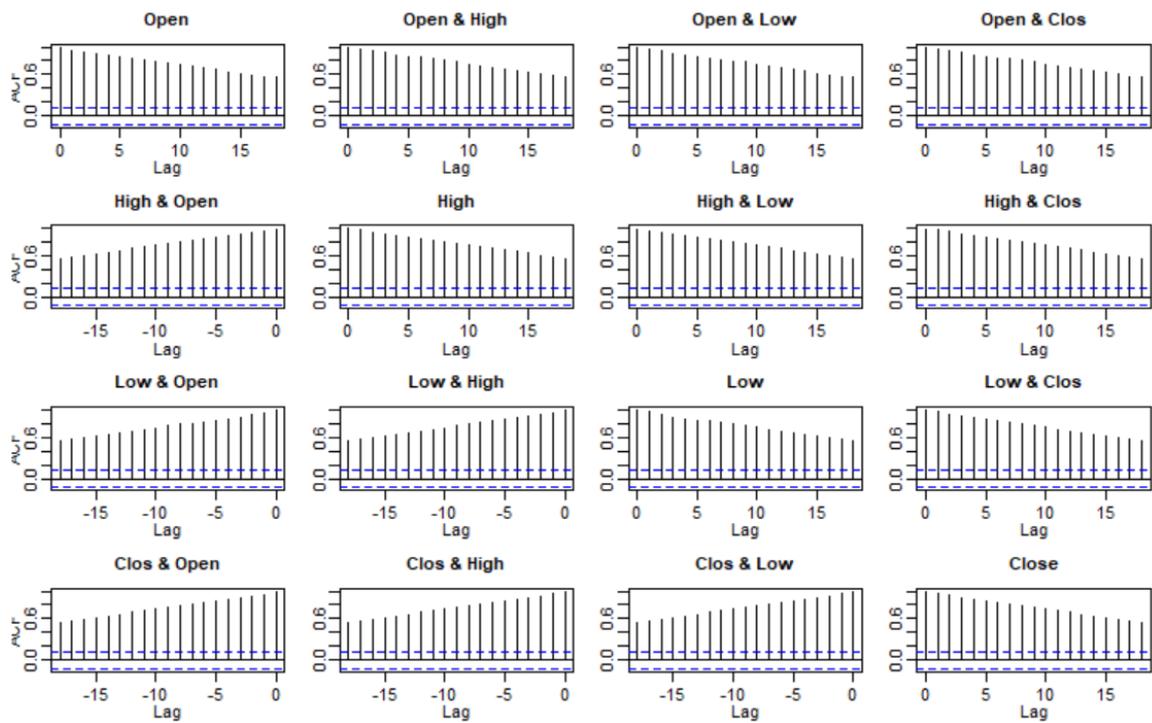


Figure 1. ACF test for correlation between open, low, high, and close of Apple stock daily prices from 1 October 2014 to 1 October 2015.

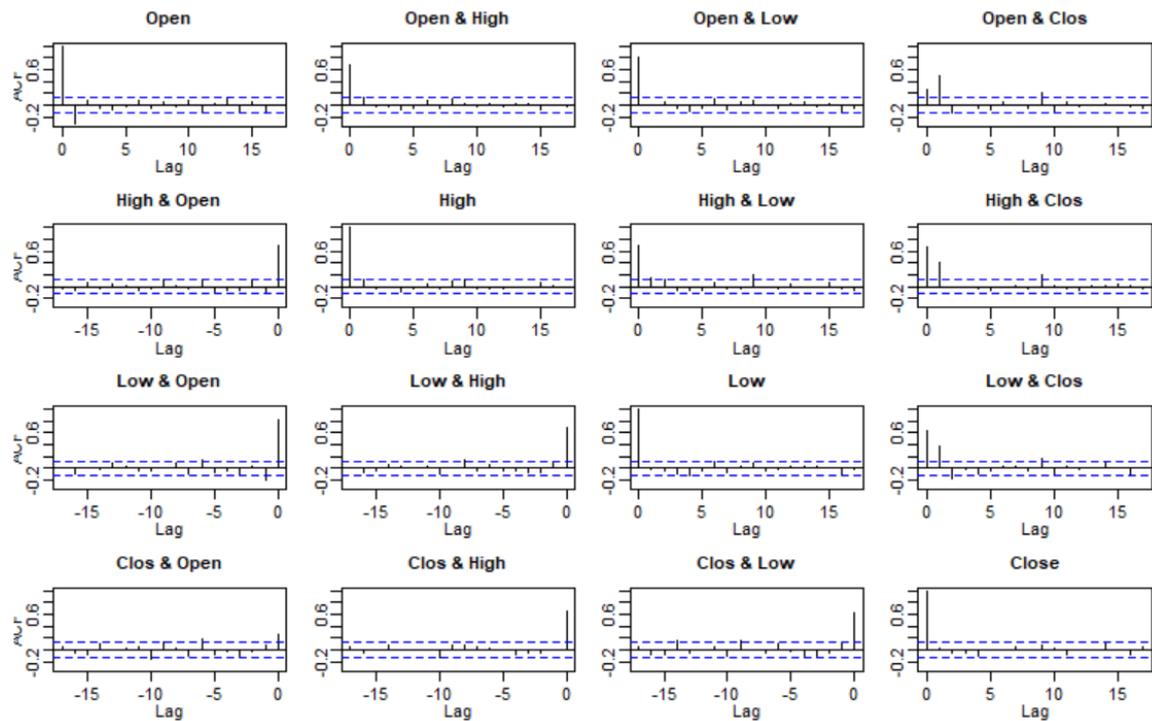


Figure 2. ACF test for correlation between open, low, high, and close of Apple stock daily return prices from 1 October 2014 to 1 October 2015.

Table 1. p -values from the Ljung–Box test for independencies of stock return series: Open, High, Low, and Close. “*” indicates that the p -value is statistically significant at $\alpha = 5\%$, “**” indicates that the p -value is statistically significant at $\alpha = 1\%$, and “***” indicates that the p -value is statistically significant at $\alpha = 0.1\%$.

Stock	Open	High	Low	Close
AAPL	0.0010 ***	0.0718	0.6584	0.6566
FB	0.2151	0.0153 *	0.3273	0.0094 **
GOOGL	0.5378	0.0214 *	0.0010 ***	0.0608

The null hypothesis of the Ljung–Box test is that the data are independently distributed. Thus, we will accept the null hypothesis if the p -value is bigger than the chosen significant level α . From Table 1, we can see that most of the stock returns series pass the independent test at the significant level $\alpha = 1\%$, and the only two series that do not pass the test at the significant level $\alpha = 0.1\%$ are APPL’s open returns and GOOGL’s low returns. The HMM works for dependent observation data with a modification in calculating probabilities of observations. We will explore the case in our future study. We will apply HMM for predicting the daily returns and then forecast future stock prices in the next section.

3.3. Model Selection

Choosing a number of hidden states for the HMM is a critical task. We first use two standard criteria: the AIC and the BIC to examine the performances of HMM with different numbers of states. The two measures are suitable for HMM because, in the model training algorithm, the Baum–Welch algorithm, the EM method was used to maximize the log-likelihood of the model. We limit numbers of states from two to four to keep the model simple and feasible for stock prediction. The AIC and BIC are calculated using the following formulas, respectively:

$$AIC = -2 \ln(L) + 2k,$$

$$BIC = -2 \ln(L) + k \ln(M),$$

where L is the likelihood function for the model, M is the number of observation points, and k is the number of estimated parameters in the model. In this paper, we assume that the distribution corresponding to each hidden state is a Gaussian distribution. Therefore, the number of parameters, k , is formulated as $k = N^2 + 2N - 1$, where N is numbers of states used in the HMM.

To train HMM’s parameters, I use historical observed data of a fixed length T ,

$$O = \{O_t^{(1)}, O_t^{(2)}, O_t^{(3)}, O_t^{(4)}, t = 1, 2, \dots, T\},$$

where $O^{(i)}$ with $i = 1, 2, 3$, or 4 represents the daily returns of open, low, high or close price of a stock, respectively. For the HMM with single observation, we use only the returns of close price data,

$$O = O_t, t = 1, 2, \dots, T,$$

where O_t is stock’s return of close price at time t . We ran the model calibrations with different time lengths, T , and saw that the model worked well for $T \geq 80$. On the results below, we used blocks of $T = 100$ trading days of stock price data, O , to calibrate HMM’s parameters and calculate the AIC and BIC numbers. Thus, the total number of observation points in each BIC calculation is $M = 400$ for four observation data and $M = 100$ for one observation data. For convenience, we did 100 calibrations for 100 blocks of data by moving the block of data forward, (we took off the price of the oldest day on the block and added the price of the following day to the recent day of the block). The calibrated

parameters of the previous step are used as initial parameters for the new calibration. The training data set is from 16 January 2015 to 30 October 2015.

The first block of stock prices of 100 trading days from 16 January 2015 to 6 June 2015 was used to calibrate HMM's parameters and calculate corresponding AIC and BIC. Let $\mu^{(O)}$ and $\sigma^{(O)}$ be the mean and standard deviation of observation data, O , respectively. We chose initial parameters for the first prediction as follows:

$$\begin{aligned} A &= (a_{ij}), \quad a_{ij} = \frac{1}{N}, \\ p &= (1, 0, \dots, 0), \\ \mu_i &= \mu^{(O)} + Z, \quad Z \sim \mathcal{N}(0, 1), \\ \sigma_i &= \sigma^{(O)}, \end{aligned} \quad (1)$$

where $i, j = 1, \dots, N$ and $\mathcal{N}(0, 1)$ is the standard normal distribution.

The second block of 100 trading day data from 17 January 2015 to 7 June 2015 was used for the second calibration and so on. The HMM calibrated parameters from the current calibration are used as initial parameters for the next estimation. We continued the process until we got 100 calibrations. We plot the AICs and BICs of the 100 calibrations of these three stocks (AAPL, FB, and GOOGL) on Figures 3–5. On Figures 3–5, the graph of AIC is located on the left and BIC is located on the right. The lower AIC or BIC is the better model calibration. However, the Baum–Welch algorithm only finds a local maximizer of the likelihood function. Therefore, we did not expect to have the same AIC or BIC if we run the calibration twice. The results on Figures 3–5 showed that the calibration performances of the model with different numbers of states differ from one simulation to others. Based on the AIC results, the performances of HMM with two, three, or four states are almost the same. However, based on the BIC, the HMM with two states is the best candidate for all three of the stocks. Therefore, we choose the HMM with two states to predict prices of the three stocks in the next section.

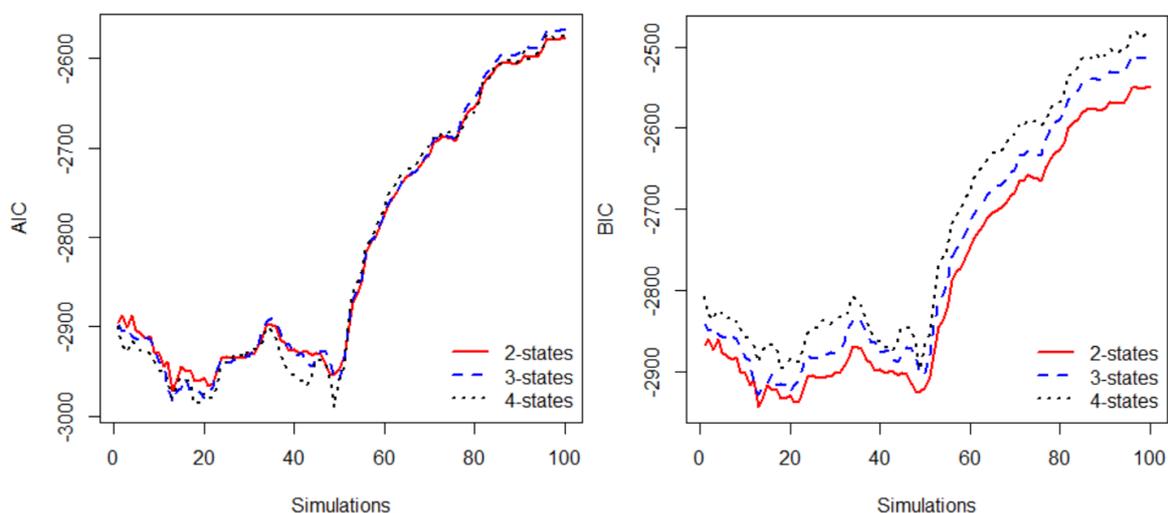


Figure 3. AIC (left) and BIC (right) for 100 parameter calibrations of HMM using Apple, AAPL, stock daily return prices.

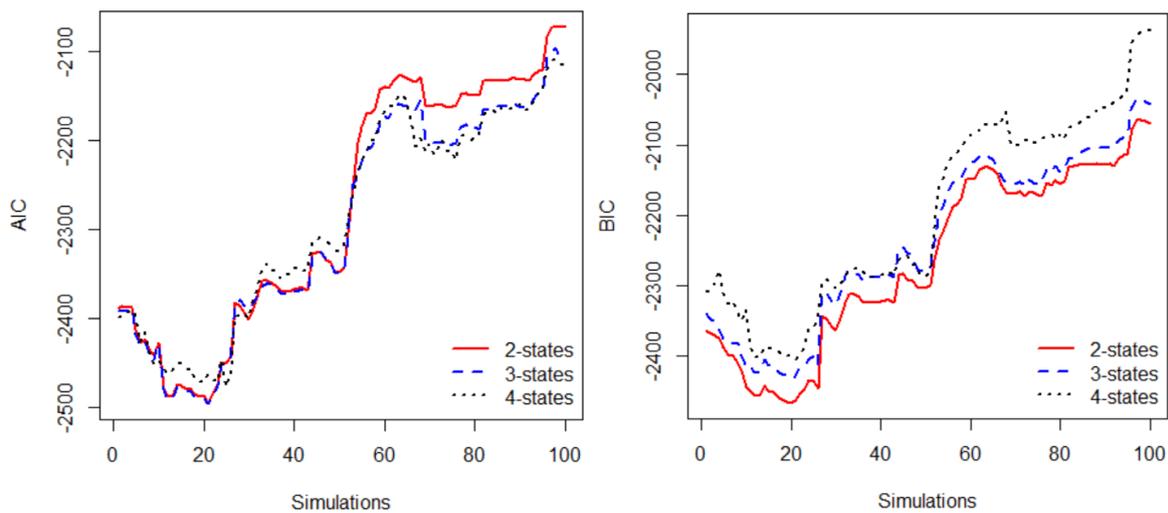


Figure 4. AIC (left) and BIC (right) for 100 parameter calibrations of HMM using Google, GOOGL, stock daily return prices.

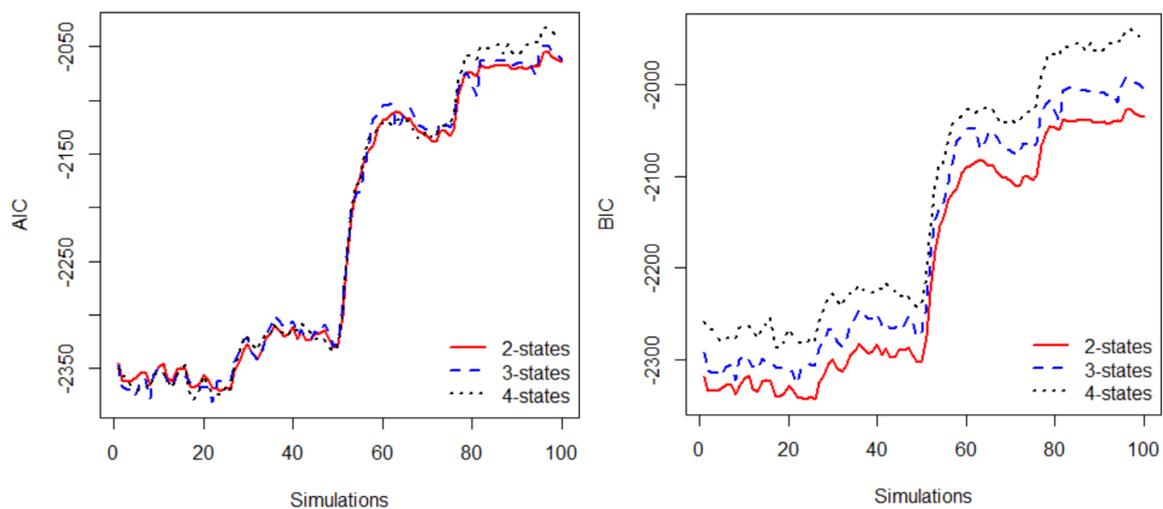


Figure 5. AIC (left) and BIC (right) for 100 parameter calibrations of HMM using Facebook, FB, stock daily return prices.

4. Stock Price Prediction and Stock Trading

In this section, we will use HMM to predict stock prices and compare the prediction with the real market prices. We will predict stock prices of GOOGL, APPL, and FB using HMM with two states, the best model selected from Section 3.3, and calculate the relative errors to the real market prices. The results will be compared with the naïve none change method. A trading strategy using HMM is also presented in this section.

4.1. Stock Price Prediction

We first introduce how to predict stock prices using HMM. The prediction process can be divided into three steps. Step 1: calibrate HMM’s parameters and calculate the likelihood of the model.

Step 2: find the day in the past that has a similar likelihood to the recent day. Step 3: use the stock returns on the day after the “similar” day in the history to be the predicted return for tomorrow price. This prediction approach is based on the work of [Hassan and Nath \(2005\)](#). However, our procedure is different from their method in that we apply HMM for the returns of open, low, high, and close prices, which are independent, while the authors used the HMM directly to open, low, high, and close prices, which are not independent. Due to applying the HMM for stock returns, our method is simpler than their method in the third step. We use HMM with the returns of the four observation sequences (open, low, high, close price), as in [Hassan and Nath \(2005\)](#).

Suppose that we want to predict tomorrow’s closing price of stock A, the prediction can be explained as follows. In the first step, we chose a block of T of the four daily return prices of stock A: open, low, high, and close, ($O = \{O_t^{(1)}, O_t^{(2)}, O_t^{(3)}, O_t^{(4)}, t = T - 99, T - 98, \dots, T\}$), to calibrate HMM’s parameters, λ , of the HMM. We then calculate the probability of observation, $P(O|\lambda)$. We assumed that the observation probability $b_i(k)$, defined in Section 2, is Gaussian distribution, so the matrix B , in the parameter $\lambda = \{A, B, p\}$, is a 2 by N matrix of means, μ , and variances, σ , of the N normal distributions, where N is numbers of states. In the second step, we move the block of data backward by one day to have new observation data $O^{new} = \{O_t^{(1)}, O_t^{(2)}, O_t^{(3)}, O_t^{(4)}, t = T - 100, T - 99, \dots, T - 1\}$ and calculate $P(O^{new}|\lambda)$. We keep moving blocks of data backward day by day until we find a data set O^* , ($O^* = \{O_t^{(1)}, O_t^{(2)}, O_t^{(3)}, O_t^{(4)}, t = T^* - 99, T^* - 98, \dots, T^*\}$) such that $P(T^*|\lambda) \simeq P(O|\lambda)$. In the third step, after finding the past “similar” day, T^* , we estimate the return of close price at time $T + 1$, by using the following formula:

$$O_{T+1}^{(4)} = O_{T^*+1}^{(4)}. \quad (2)$$

After the first prediction for stock return of day $T + 1$ we update data window, O , by moving it forward one day, $O = \{O_t^{(1)}, O_t^{(2)}, O_t^{(3)}, O_t^{(4)}, t = T - 98, T - 97, \dots, T + 1\}$, to predict stock return for the day $T + 2$. The calibrated HMM’s parameters in the first prediction were used as the initial parameters for the second prediction. We repeat the prediction process as mentioned in the first prediction for the second prediction and so on. For HMM with a single observation sequence, we use $O = O_t^{(4)}$, where $O_t^{(4)}$ is the return of close price.

The predicted close price at time $T + 1$, P_{T+1} , is calculated by the predicted stock returns:

$$P_{T+1} = P_T * (O_{T+1}^{(4)} + 1), \quad (3)$$

where P_T is close price at time T and $O_{T+1}^{(4)}$ is the return of close price calculated in (2).

The naïve none change method is applied for returns of the three stocks’ close prices. The model simply takes the return of the close price today to use as the return of the tomorrow’s close price

$$O_{T+1}^{(4)} = O_T^{(4)}.$$

After forecasting $O_{T+1}^{(4)}$, we predict the next day’s close price by using Equation (3). We use the naïve method for stock returns instead of stock prices because, for trading purposes, if we assume no change in future stock prices, then there is no trade. We present results of using the HMM to predict these three stocks’—AAPL, GOOGL, and FB—closing prices for one year trading, 252 days, in Figures 6–8. The results indicate that the HMM captures the trends of the three stocks well, while the naïve forecasts often go to the opposite directions of the real market trends. We can see from Figure 7 that the naïve forecast method had a few huge errors in predicting stock prices in February. The naïve model also showed its weakness when predicted prices of Google stock at the end of July 2015 are far from the actual prices.

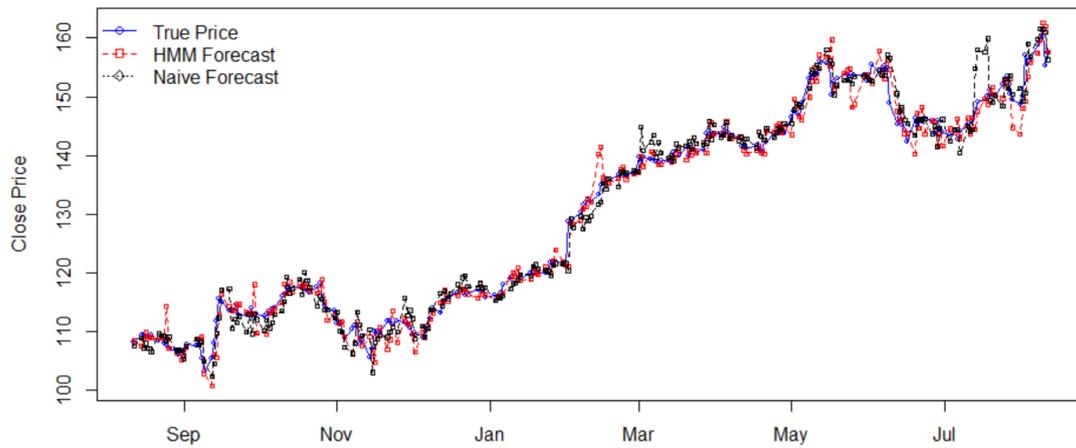


Figure 6. HMM prediction of Apple stock daily close prices from 15 August 2016 to 11 August 2017 using two-states HMM and the naïve model.

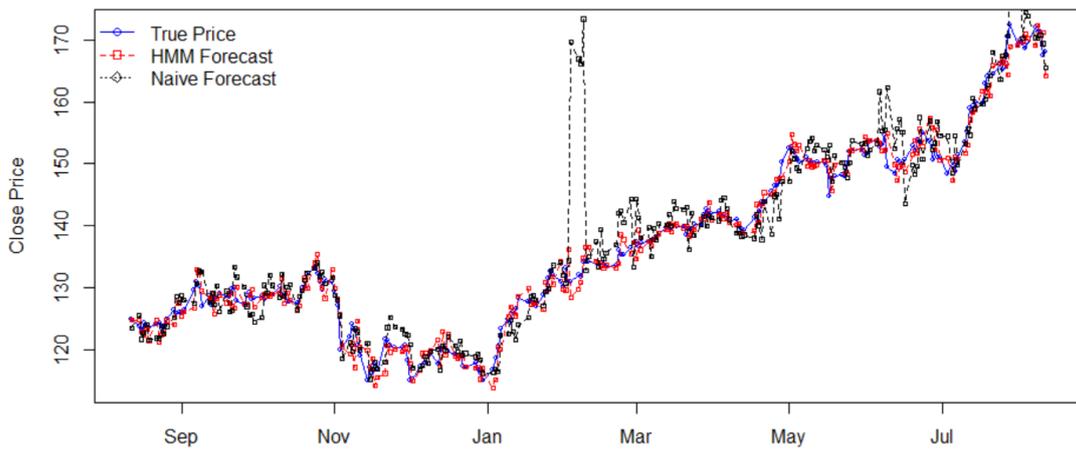


Figure 7. HMM prediction of Facebook stock daily close prices from 15 August 2016 to 11 August 2017 using two-states HMM and the naïve model.

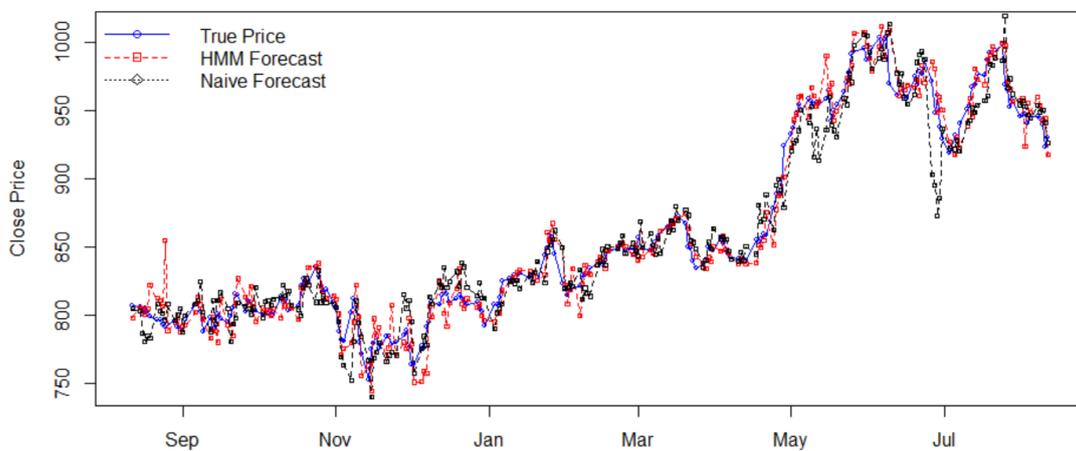


Figure 8. HMM prediction of Google stock daily close prices from 15 August 2016 to 11 August 2017 using two-states HMM and the naïve model.

We also compare the forecasting results of using the two-state HMM and the naïve method numerically by calculating the mean absolute percentage error, MAPE, of the estimations.

$$MAPE = \frac{1}{N} \sum_{i=1}^N \frac{|M_i - P_i|}{M_i},$$

where N is number of predicted points, M is market price, and P is predicted price of a stock. The results were shown in Table 2. In Table 2, the “Price Std.” and “Return Std.” are the standard deviation of the stock prices and stock returns, respectively, and the efficiencies are calculated by taking the errors of the naïve method divided by the errors of the HMM. All efficiencies in the table are bigger than one, showing that the HMM outperformed the naïve in forecasting stock prices.

Table 2. Comparison of MAPE of stock price predictions of Apple, Google, and Facebook from 15 August 2016 to 11 August 2017, between the HMM and the naïve forecast model.

Stock	Price Std.	Return Std.	HMM's MAPE	Naïve's MAPE	Efficiency
AAPL	17.0934	0.0113	0.0113	0.0133	1.1770
FB	14.4879	0.0111	0.0116	0.0213	1.8362
GOOGL	69.9839	0.0098	0.0107	0.0137	1.2804

Among these three stocks, GOOGL's prices have the highest volatility, but its returns have the lowest volatility. These factors will affect stock trading results so that we will present the results in the next section.

4.2. Stock Trading

In this section, we will use the predicted returns to trade these three stocks: AAPL, FB, and GOOGL. The trading strategy is: if HMM predicts that the stock price of AAPL will move up tomorrow, or its return is positive, we will buy this stock today and sell it tomorrow, assuming that we buy and sell with close prices. If the HMM predicts that the stock price will not increase tomorrow, then we will do nothing. We also assume that there is no trading cost. For each trade, we will buy or sell 100 shares of each of these three stocks. Based on the AIC and BIC results, we only use HMM with two states for the stock trading. Again, we will use a block of 252 trading days, one year, from 15 August 2016 to 11 August 2017 for model testing. We present the results of one year trading in the Table 3.

Table 3. Stock trading results from 15 August 2016 to 11 August 2017.

Stock	Models	Investment \$	Earning \$	Profit %
AAPL	HMM	10,908	3481	31.91
	Naïve	10,818	3513	32.47
FB	HMM	12,490	2939	23.53
	Naïve	12,488	2565	20.54
GOOGL	HMM	80,596	20,039	24.86
	Naïve	79,965	2715	3.40

In Table 3, the “Investment” is the price that we bought 100 shares of the stocks the first time. The “Earning” is the money gained, and the “Profit” is the percentage of return of the one-year trading. The results show that the HMM worked better than the naïve in trading the Facebook and Google stocks. Especially in the one year trading period, the GOOGL stock yielded a much higher return compared to the naïve forecast method. However, the results are reversed for AAPL stock. From Figures 6–8 and Table 2, we can see that, in the one period, the GOOGL prices have the highest volatility and lowest risk of returns among the three stocks. The naïve results are consistent with the

risk of return levels, the “Return Std.” in Table 2: the higher the risk, the better the return. The HMM followed close to the risk level theoretical. Based on the results in Table 3, using an HMM model, traders had returns of 32.00%, 24%, and 25% for AAPL, FB, GOOGL, respectively. Trading using HMM gave much higher returns than using the naïve for two stocks FB and GOOGL, but a likely lower return for the AAPL stock compare to the naïve.

5. Conclusions

Stock’s performances are an essential indicator of the strength or weakness of the stock’s corporation and economic viability in general. Many factors will drive stock prices up or down. In this paper, we use a Hidden Markov Model, HMM, to predict prices of three stocks: AAPL, GOOGL, and FB. We first use the AIC and BIC criteria to examine the performances of HMM numbers of states from two to four. The results showed that the HMM with two states is the best model among the two, three and four states. We then use the models to predict stock prices and compare the predictions with the naïve forecast results by plotting the forecasted prices versus the market prices and evaluating the mean absolute percentage error, MAPE. The prediction errors show that HMM worked better in predicting prices of the three stocks—AAPL, FB, and GOOGL—compared with the naïve method. In stock trading, the HMM outperformed the naïve for two stocks: FB and GOOGL. The graphs indicate that the HMM is the potential model for stock trading since it captures the trends of stock prices well.

Acknowledgments: I thank three anonymous referees at Risks, the editor Albert Cohen, and the assistant editor Shelly Liu for their comments and assistances.

Conflicts of Interest: The author declares no conflict of interest.

Appendix A

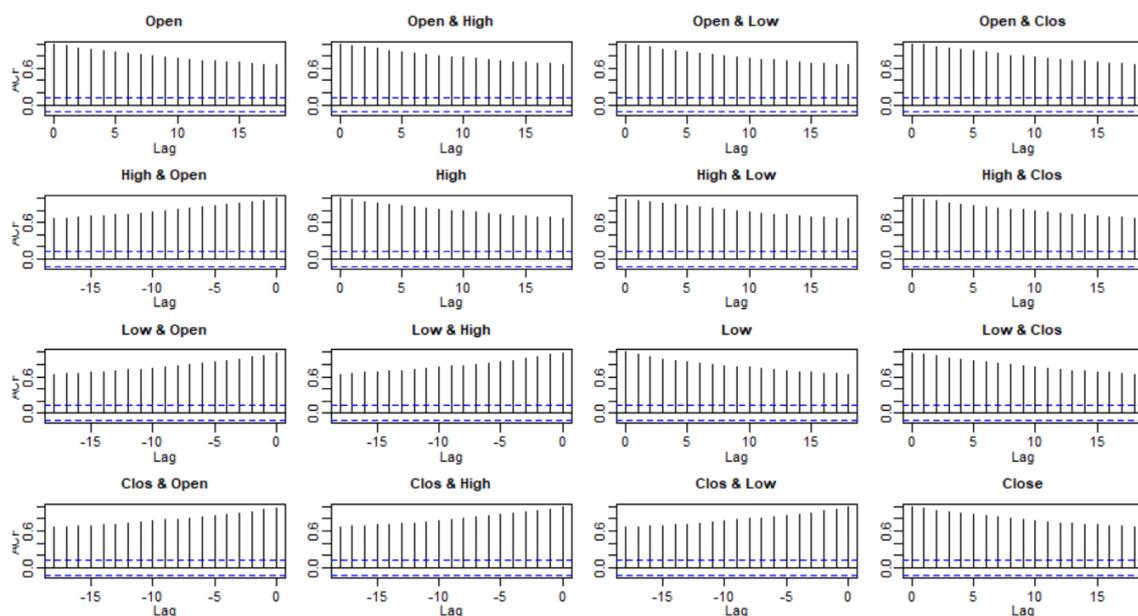


Figure A1. ACF test for correlation between open, low, high, and close of Facebook daily stock prices from 1 October 2014 to 1 October 2015.

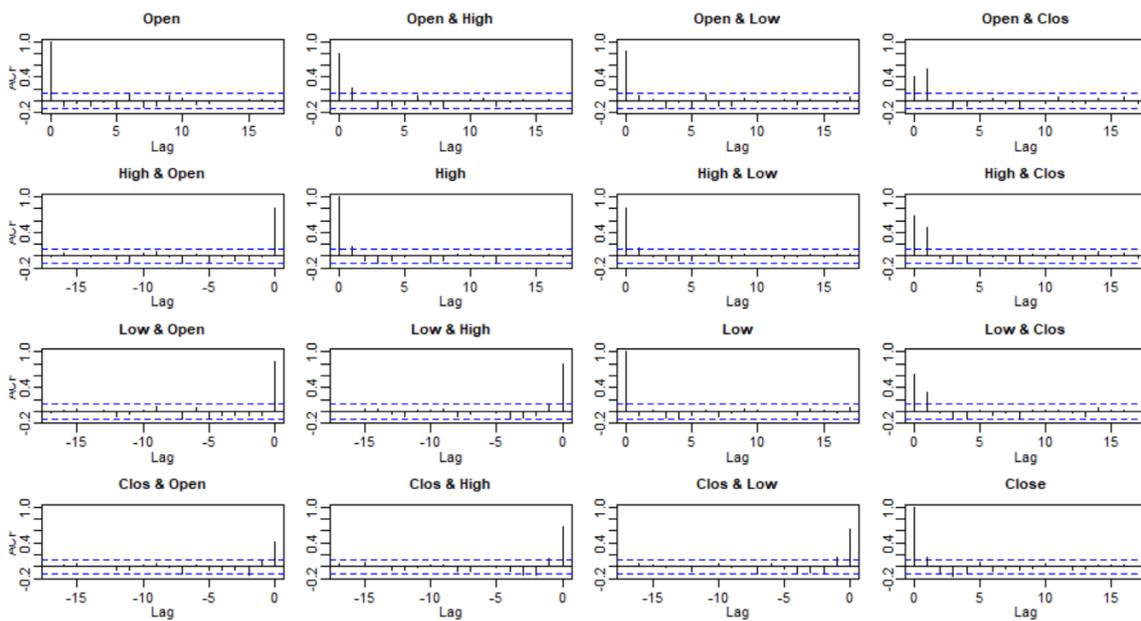


Figure A2. ACF test for correlation between open, low, high, and close of Facebook stock daily returns from 1 October 2014 to 1 October 2015.

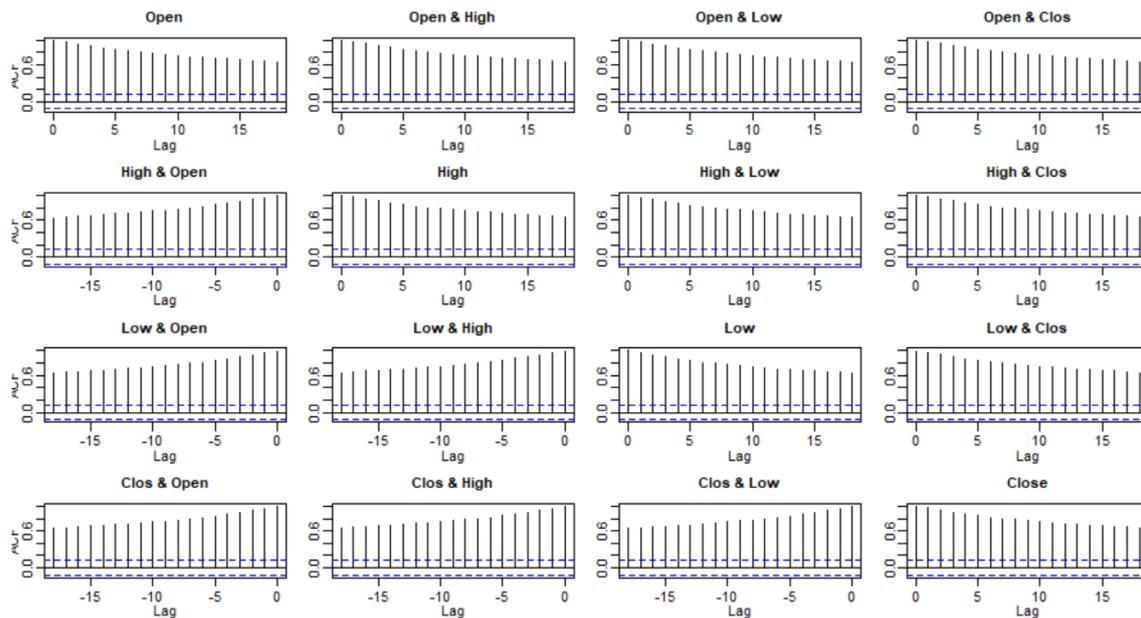


Figure A3. ACF test for correlation between open, low, high, and close of Google daily stock prices from 1 October 2014 to 1 October 2015.

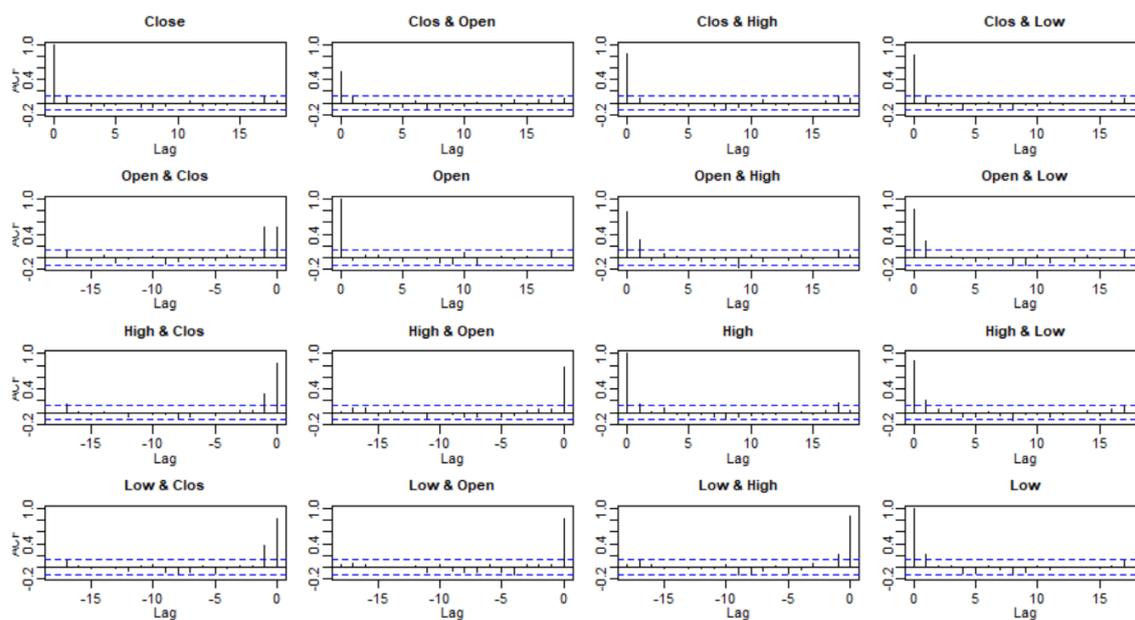


Figure A4. ACF test for correlation between open, low, high, and close of Google stock daily returns from 1 October 2014 to 1 October 2015.

References

- Ang, Andrew, and Geert Bekaert. 2002. International Asset Allocation with Regime Shifts. *The Review of Financial Studies* 15: 1137–87.
- Baum, Leonard E., and John Alonzo Eagon. 1967. An inequality with applications to statistical estimation for probabilistic functions of Markov process and to a model for ecology. *Bulletin of the American Mathematical Society* 73: 360–63.
- Baum, Leonard E., and Ted Petrie. 1966. Statistical Inference for Probabilistic Functions of Finite State Markov Chains. *The Annals of Mathematical Statistics* 37: 1554–63.
- Baum, Leonard E., and George Roger Sell. 1968. Growth functions for transformations on manifolds. *Pacific Journal of Mathematics* 27: 211–27. doi: 10.2140/pjm.1968.27.211
- Baum, Leonard E., Ted Petrie, George Soules, and Norman Weiss. 1970. A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains. *The Annals of Mathematical Statistics* 41: 164–71.
- Chen, Chunchih. 2005. *How Well Can We Predict Currency Crises? Evidence from a Three-Regime Markov-Switching Model*. Davis: Department of Economics, UC Davis.
- Forney, G. David. 1973. The Viterbi algorithm. *Proceedings of the IEEE* 61: 268–78.
- Guidolin, Massimo, and Allan Timmermann. 2006. Asset Allocation under Multivariate Regime Switching. SSRN FRB of St. Louis Working Paper No. 2005-002C, FRB of St. Louis, MO, USA.
- Hassan, Md Rafiul, and Baikunth Nath. 2005. Stock Market Forecasting Using Hidden Markov Models: A New approach. Paper presented at the IEEE fifth International Conference on Intelligent Systems Design and Applications, Warsaw, Poland, September 8–10.
- Kritzman, Mark, Sebastien Page, and David Turkington. 2012. Regime Shifts: Implications for Dynamic Strategies. *Financial Analysts Journal* 68: 22–39.
- Levinson, Stephen E., Lawrence R. Rabiner, and Man Mohan Sondhi. 1983. An introduction to the application of the theory of probabilistic functions of Markov process to automatic speech recognition. *Bell System Technical Journal* 62: 1035–74.
- Li, Xiaolin, Marc Parizeau, and Réjean Plamondon. 2000. Training Hidden Markov Models with Multiple Observations—A Combinatorial Method. *IEEE Transactions on PAMI* 22: 371–77.

- Nguyen, Nguyet, and Dung Nguyen. 2015. Hidden Markov Model for Stock Selection. Journal of Risks in special issue: Recent Advances in Mathematical Modeling of the Financial Markets. *Risks* 3: 455–73. doi:10.3390/risks3040455.
- Nguyen, Nguyet Thi. 2014. Probabilistic Methods in Estimation and Prediction of Financial Models. Electronic Theses, Treatises and Dissertations Ph.D. dissertation, The Florida State University, Tallahassee, FL, USA.
- Nobakht, B., C. E. Joseph, and B. Loni. 2012. Stock market analysis and prediction using hidden markov models. Paper presented at the 2012 Students Conference on IEEE Engineering and Systems (SCES), Allahabad, Uttar Pradesh, India, March 16–18, pp. 1–4.
- Petrushin, Valery A. 2000. Hidden Markov Models: Fundamentals and Applications (part 2-discrete and continuous hidden markov models). Online Symposium for Electronics Engineer. Available online: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.378.3099&rep=rep1&type=pdf> (accessed on 23 November 2017).
- Rabiner, Lawrence R. 1989. A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition. *Proceedings of the IEEE* 77: 257–86.
- Viterbi, Andrew J. 1967. Error bounds for convolutional codes and an asymptotically optimal decoding algorithm. *IEEE Transactions on Information Theory* IT-13: 260–69.



© 2017 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).