Article can be accessed online at http://www.publishingindia.com

Forecasting The Stock Market Values Using Hidden Markov Model

R. Sasikumar*, A. Sheik Abdullah*

Abstract

The financial market influences personal corporate financial lives and the economic health of a country. Price change of stock market is not a completely random model. The pattern of financial market has been observed by some economists, statisticians and computer scientists. This paper gives a detailed idea about the sequence and state prediction of stock market using Hidden Markov Model and also making inferences regarding stock market trend. The one day difference in close value of stock market value has been used for some period and the corresponding transition probability matrix and emission probability matrix are obtained. Seven optimal hidden states and three sequences are generated using MATLAB and then compared.

Keywords: Hidden Markov Model, Transition Probability Matrix, Emission Probability Matrix, Stock Market, States and Sequence

Introduction

The most of the trading in Indian stock market is classified in two categories, the Bombay Stock Exchange (BSE) and the National Stock Exchange (NSE). The BSE has been functioning since 1875. The NSE was founded in 1992 and started trading in 1994. Even though both exchanges have the same trading mechanism, trading hours, settlement process, etc., they are having high demand from people. The two prominent Indian market indices are Sensex and S&P CNX Nifty.

Financial market (Stock Market) is a platform for investors to own some shares of a company. Investors will then become a part of the company members and share in both gains and losses of that particular company. This is a better way for the investors to get extra income apart from their regular salary. Changes of share prices on every day make it more volatile and difficult to predict the future price. When purchasing a stock, it does not guarantees to give anything in return. Thus, it makes stocks risky in investment, but investors can also get high profit return. When investors take wrong decision in choosing the counters, it may end up in capital loss. The behavior of stock market returns has been deeply discussed over some years. In this paper, the hidden states and sequence are generated for stock market values using Hidden Markov Model (HMM) through software.

Review of related works

There are so many researches going on stock market analysis. Rabiner (1989) used precise HMMs, in which the state sequence estimation problem can be solved very efficiently by the Viterbi algorithm whose complexity is linear in the number of nodes, and quadratic in the number of states. However, this algorithm only emits a single optimal (most probable) state sequence, even in cases where there are multiple (equally probable) optimal solutions. Hassan and Baikunth Nath (2005) used HMM to predict next day closing price for some of the airlines. They considered four input attributes for a stock, and they were the opening price, highest price, lowest price and closing price. These four attributes of previous day were used to predict next day's closing price. Hassan (2009) introduced the new combination of HMM and Fuzzy model to forecast the stock market data. He classified the data set as daily opening, high, low and closing prices to predict the next day's closing price. HMM-fuzzy model is more reliable and profitable than the other model.

Jyoti Badge (2012) used Macro-Economic factor as a technical indicator, which is used to identify the patterns of the market at a particular time. For selecting technical indicator author was applying principal component

* Department of Statistics, Manonmaniam Sundaranar University, Tirunelveli, Tamil Nadu, India.
Email: sasikumarmsu@gmail.com and sheik.stat@gmail.com

analysis. Luca et al., (2013) investigated the dynamic patterns of stock markets by exploiting the potential of the HMM for defining different market regimes and providing transition probabilities for regime-switching. Tuyen (2013) used HMM to estimate the parameter of the Markov Black-Sholes model to predict the option prices in the stock market. The historical data of daily VN-Index (Vietnam Stock Market) were taken from 2009 to 2011 for finding the four hidden states corresponding to the Normal distribution $N(\mu_i, \sigma_i)$ for $i = 1,2,3,4$ with the help of HMM. Kai Cui (2014) explained that, the variation of financial time sequence for Shangai composite index was predicted by introduction of a dual state HMM. He also justified that the HMM was the best tool to predict the variation of financial time sequence. Somani et al., (2014) surveyed support vector machine, neural network and HMM in the area of stock market forecasting. HMM is more efficient in getting information from the result, showing future behavior of stock market valuesand fluctuations.

Methodology

Hidden Markov Model

HMM has been successful in analyzing and predicting phenomena's relying on a time dependence or time series. It is very effective and intuitive approach to many sequential pattern recognition tasks, such as speech recognition, protein sequence analysis, machine translation, pair wise and multiple sequence alignments, gene annotation, classification and similarity search.

A HMM is a doubly stochastic process in which an underlying stochastic process is unobservable, which means that the state is hidden. This can only be observed through another stochastic process that produces a sequence of observations. Thus, if $S = \{S_n, n=1, 2,...\}$ is a Markov process and $F = \{F_k, k=1, 2, \ldots\}$ is a function of S, then S is a hidden Markov process or HMM that is observed through F, and S is also known as the state process that is hidden and F as the observation process that can be observed. The observed event is called as a "symbol" and the invisible factor underlying the observation a "state".

A HMM is usually defined as a 5-tuple (S, F, P, ψ, π), where

$S = \{s_1, s_2, \ldots, s_n\}$ is a finite set of n states.

$F = \{o_1, o_2, \ldots, o_m\}$ is a finite set of m possible symbols.

$P = \{p_{ij}\}$ is the set of state-transition probabilities, where p_{ij} is the probability that the system goes from state s_i to state s_j.

$\psi = \{\psi_i(o_k)\}$ are the observation probabilities, where $\psi_i(o_k)$ is the probability that the symbol o_k is emitted when the system is in state s_i.

$\pi = \{\pi_i\}$ are the initial state probabilities; that is the probability that the system starts in state s_i.

As the states and the output sequence are understood, it is usually denoted by the parameters of a HMM by $\lambda = (P, \psi, \pi)$.

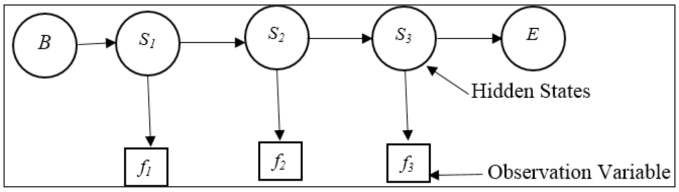

Figure 1: General Structure of a Hidden Markov Model

From the Figure 1, the $S_i$are the hidden states that is to be estimated and the F_i are the observation of the random variables from which the S_i are to be estimated. The letters B and E indicate the beginning and end of the sequence of states.

Transition Probability Matrix

The transition probability P_{jk},where $P_{jk} \geq 0$, for all j. These probabilities may be written in the matrix form,

$$P = \begin{bmatrix} p_{11} & p_{12} & p_{13} & \cdots \\ p_{21} & p_{22} & p_{23} & \cdots \\ \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots \end{bmatrix}$$

This is called the transition probability matrix (tpm). P is a stochastic matrix ie.a square matrix with non-negative elements and row total is equal to one.

Materials and Methods

In this paper, Oil India Ltd is sample share, its daily close value data for two months period is considered. Three observing symbols I, N and D are indicated. The symbol I-indicates Increasing, N-indicates No change and D-indicates Decreasing. If n^{th} day's close value – $(n-1)^{th}$ day's close value > 0, then observing symbol is I. If n^{th} day's close value – $(n-1)^{th}$ day's close value < 0, then observing symbol is D. If n^{th} day's close value – $(n-1)^{th}$ day's close value = 0, then observing symbol is N.

Seven hidden states are assumed and are denoted by the following symbols S_1, S_2, S_3, S_4, S_5, S_6, S_7

where,

S_1 - very low

S_2 - low

S_3 - moderate low

S_4 - no change

S_5 - moderate high

S_6 – high

S_7 – very high

Since the above mentioned states are not directly observable, in this situation the stock market values are considered as hidden. From the hidden state sequences, it is possible to produce the observations.

The various probability values of tpm and emission probability matrix (epm) for difference in one day, two days and three days close values are calculated as follows: tpm and epm for one day close value difference

$$
\begin{array}{c} \\ S1 \\ S2 \\ S3 \\ S4 \\ S5 \\ S6 \\ S7 \end{array}
\begin{array}{c}
\begin{array}{ccccccc} S1 & S2 & S3 & S4 & S5 & S6 & S7 \end{array} \\
\begin{bmatrix}
0 & 0 & 0 & 1 & 0 & 0 & 0 \\
1/5 & 2/5 & 1/5 & 0 & 0 & 0 & 1/5 \\
0 & 0 & 0 & 1/11 & 8/11 & 2/11 & 0 \\
0 & 1/4 & 2/4 & 1/4 & 0 & 0 & 0 \\
0 & 1/14 & 7/14 & 1/14 & 3/14 & 2/14 & 0 \\
0 & 0 & 1/6 & 0 & 4/6 & 0 & 1/6 \\
0 & 1/2 & 0 & 0 & 0 & 1/2 & 0
\end{bmatrix}
\end{array}
$$

Figure 1(a): tpm

Table 1: The Closing Value of a Stock Market Tpm and Epm for two Day Close Value Difference

S. No	Close Value	Diff in 1 day	Observation Symbol	Diff in 2 day	Observation Symbol	Diff in 3 day	Observation Symbol
1	458.00						
2	465.70	7.70	I				
3	469.00	3.30	I	-4.40	D		
4	467.35	-1.65	D	-4.95	D	-0.55	D
5	469.90	2.55	I	4.20	I	9.15	I
6	473.65	3.75	I	1.20	I	-3.00	D
7	466.00	-7.65	D	-11.40	D	-12.60	D
8	487.20	21.20	I	28.85	D	40.25	I
9	501.00	13.80	I	-7.40	D	-36.25	D
10	518.85	17.85	I	4.05	I	11.45	I
11	508.60	-10.25	D	-28.10	D	-32.15	D
12	500.15	-8.45	D	1.80	I	29.90	I
13	478.45	-21.70	D	-13.25	D	-15.05	D
14	478.45	0.00	N	21.70	I	34.95	I
15	467.50	-10.95	D	-10.95	D	-32.65	D
16	456.30	-11.20	D	-0.25	D	10.70	I
17	451.95	-4.35	D	6.85	I	7.10	I
18	457.05	5.10	I	9.45	I	2.60	I
19	459.25	2.20	I	-2.90	D	-12.35	D
20	458.50	-0.75	D	-2.95	D	-0.05	D
21	458.50	0.00	N	0.75	I	3.70	I
22	458.50	0.00	N	0.00	N	-0.75	D
23	453.65	-4.85	D	-4.85	D	-4.85	D
24	457.00	3.35	I	8.20	I	13.05	I

S. No	Close Value	Diff in 1 day	Observation Symbol	Diff in 2 day	Observation Symbol	Diff in 3 day	Observation Symbol
25	456.75	-0.25	D	-3.60	D	-11.80	D
26	467.65	10.90	I	11.15	I	14.75	I
27	470.10	2.45	I	-8.45	D	-19.60	D
28	470.95	0.85	I	-1.60	D	6.85	I
29	483.30	12.35	I	11.50	I	13.10	I
30	481.30	-2.00	D	-14.35	D	-25.85	D
31	483.45	2.15	I	4.15	I	18.50	I
32	478.65	-4.80	D	-6.95	D	-11.10	D
33	480.65	2.00	I	6.80	I	13.75	I
34	477.50	-3.15	D	-5.15	D	-11.95	D
35	485.95	8.45	I	11.60	I	16.75	I
36	487.55	1.60	I	-6.85	D	-18.45	D
37	486.00	-1.55	D	-3.15	D	3.70	I
38	492.60	6.60	I	8.15	I	11.30	I
39	491.15	-1.45	D	-8.05	D	-16.20	D
40	493.70	2.55	I	4.00	I	12.05	I
41	493.70	0.00	N	-2.55	D	-6.55	D
42	488.35	-5.35	D	-5.35	D	-2.80	D
43	490.00	1.65	I	7.00	I	12.35	I
44	499.00	9.00	I	7.35	I	0.35	I
45	501.25	2.25	I	-6.75	I	-14.10	D

$$\begin{array}{c} \\ S1 \\ S2 \\ S3 \\ S4 \\ S5 \\ S6 \\ S7 \end{array} \begin{array}{c} \begin{array}{ccc} I & N & D \end{array} \\ \left[\begin{array}{ccc} 0 & 1 & 0 \\ 1/5 & 0 & 4/5 \\ 10/11 & 1/11 & 0 \\ 0 & 1/4 & 3/4 \\ 5/14 & 1/14 & 8/14 \\ 1/6 & 0 & 5/6 \\ 0 & 0 & 1 \end{array}\right] \end{array}$$

Figure 1(b): **epm**

$$\begin{array}{c} \\ S1 \\ S2 \\ S3 \\ S4 \\ S5 \\ S6 \\ S7 \end{array} \begin{array}{c} \begin{array}{cc} I & D \end{array} \\ \left[\begin{array}{cc} 1 & 0 \\ 1 & 0 \\ 14/17 & 3/17 \\ 0 & 1 \\ 2/14 & 12/14 \\ 0 & 1 \\ 0 & 1 \end{array}\right] \end{array}$$

Figure 2(b): **epm**

$$\begin{array}{c} \\ S1 \\ S2 \\ S3 \\ S4 \\ S5 \\ S6 \\ S7 \end{array} \begin{array}{c} \begin{array}{ccccccc} S1 & S2 & S3 & S4 & S5 & S6 & S7 \end{array} \\ \left[\begin{array}{ccccccc} 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1/4 & 0 & 1/4 & 0 & 2/4 \\ 0 & 0 & 5/17 & 0 & 9/17 & 3/17 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1/14 & 2/14 & 7/14 & 1/14 & 3/14 & 0 & 0 \\ 0 & 1/3 & 2/3 & 0 & 0 & 0 & 0 \\ 0 & 1/2 & 1/2 & 0 & 0 & 0 & 0 \end{array}\right] \end{array}$$

Figure 2(a): **tpm**

tpm and epm for three day close value difference

$$\begin{array}{c} \\ S1 \\ S2 \\ S3 \\ S4 \\ S5 \\ S6 \\ S7 \end{array} \begin{array}{c} \begin{array}{ccccccc} S1 & S2 & S3 & S4 & S5 & S6 & S7 \end{array} \\ \left[\begin{array}{ccccccc} 0 & 0 & 0 & 0 & 2/4 & 1/4 & 1/4 \\ 0 & 0 & 1/9 & 0 & 5/9 & 1/9 & 2/9 \\ 0 & 0 & 3/7 & 0 & 4/7 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 2/16 & 6/16 & 3/16 & 0 & 5/16 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 2/3 & 1/3 & 0 & 0 & 0 & 0 & 0 \end{array}\right] \end{array}$$

Figure 3(a): **tpm**

$$
\begin{array}{c|cc}
 & \mathrm{I} & \mathrm{D} \\
\hline
\end{array}
\quad
\begin{array}{c}
S1 \\ S2 \\ S3 \\ S4 \\ S5 \\ S6 \\ S7
\end{array}
\begin{bmatrix}
1 & 0 \\
1 & 0 \\
5/7 & 2/7 \\
1 & 0 \\
5/16 & 11/16 \\
0 & 1 \\
0 & 1
\end{bmatrix}
$$

Figure 3(b): **epm**

From the above TPM and EPM hidden states and sequence have been generated using MATLAB software. Difference of one day, two day and three day hidden states and sequence are given below respectively. From the sequence and states we can predict the future values of stock value.

1	Sequence:	D	I	D	I	I	D	D	I	D	I
	States:	S4	S2	S7	S6	S5	S5	S5	S3	S5	S3
2	Sequence:	D	I	D	I	D	I	D	I	I	I
	States:	S5	S3	S5	S3	S6	S3	S5	S2	S3	S3
3	Sequence:	I	I	I	D	D	I	D	I	I	I
	States:	S5	S5	S3	S5	S5	S2	S6	S2	S5	S3

Conclusion

Stock market values are unpredictable because of the variation of several factors. So there is no single method which can perfectly forecast the stock price values, HMM is no exception. Even though through this paper, the HMM model easily recognized three states of the stock market and also it was used to forecast the future values. In this paper, hidden states and sequences have been generated to identify, so that, we can easily identify the future states and also easily identify the sequence whether the next day value is increasing or decreasing and increasing/ decreasing level can also be observed. We can identify whether the increasing level is moderate, high or high or very high and also decreasing level whether moderate or low or very low. This is very useful for short term as well as long term investors.

Acknowledgement

The authors acknowledge University Grants Commission for providing the infrastructure facility under the scheme of SAP (DRS-I). The second author acknowledges UGC for financial support to carry out this research work under Basic Science Research Fellowship.

References

Hassan, M. R., & Nath, B. (2005). Stock Market Forecasting Using Hidden Markov Model: A New Approach, *Proceedings of the 5th International Conference on Intelligent Systems Design and Applications*, 192-196.

Hassan, M. R., Nath, B., & Kirley, M. (2007). A fusion model of HMM, ANN and GA for stock market forecasting. *Expert Systems with Applications*, 33(1), 171-180.

Hassan, M. R. (2009). A combination of hidden Markov model and fuzzy model for stock market forecasting. *Neurocomputing*, 72, 3439-3446.

Badge, J. (2012). Forecasting of Indian Stock Market by effective macro-economic factors and stochastic model. *Journal of Statistical and Econometric Methods*, 1(2), 39-51.

Cui, K. (2014). Analysis of Financial Time Sequence using Hidden Markov Models. *Journal of Multimedia*, 9(6), 810-815.

Angelis, L. D., & Paas, L. J. (2013). A dynamic analysis of stock markets using a hidden Markov model, *Journal of Applied Statistics*, 40(8), 1682-1700.

Ibe, O. C. (2009). *Markov processes for stochastic modeling.* Elsevier Academic Press

Rabiner, L. R., & Juang, B. H. (1986). An introduction to hidden Markov models. *IEEE ASSP Magazine January*, 4-16.

Rabiner, L. R. (1989). A tutorial on HMM and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2), 257-286.

Somani, P., Talele, S., & Sawant. S. (2014). Stock market prediction using Hidden Markov model. *IEEE*, 89-92.

Tuyen, L. T. (2013). Markov financial model using hidden Markov model. *International Journal of Applied Mathematics and Statistics*, 40(10), 72-83.

Weigend, S., & Back, A. D. (1998). What drives stock returns? An independent component analysis. *In Proceedings of the IEEE/IAFE/INFORMS 1998 Conference on Computational Intelligence for Financial Engineering, IEEE*, New York., pp. 141- 156.

Zhang, D., & Xiaomin, Z. (2009). Study on forecasting the stock market trend based on stochastic analysis method. *International Journal of Business and Management*, 4(6), 163-170.