Forecast Stock Prices with Markov Model

Ruiqing Xu[1, a, *]

[1]*Department of Statistics, University of Toronto, Toronto, Canada*
a.ruiqing.xu@mail.utoronto.ca
**corresponding author*

Keywords: stock prices, Patterns, Markova chain, Alibaba group, growth rate, transition Matrix

Abstract: Due to the tremendous interests of several large corporations, making stock market predictions has always been one of the most activated research fields. However, forecasting stock prices has been restricted by huge factors regarding its characteristics of volatility, seasonality and unpredictability. In 1991, Grant Mcqueen and Steven Thorley, who are Professors of Finance at BYU Marriott School of Management, conduct examination of predicting stock prices by using Markov Chains. In 2005, Hassan and Nath used only one HMM that is trained on the past dataset of the chosen airlines to forecast the stock prices of these companies. After reading two journals regarding different approaches, it was not hard to notice that these methods actually were conducted based on contradicted assumptions. It appears that the transition of methods used in order to predict stock prices has been well developed, and this recent study in 2005 has overturned the past premise on random walks theory, which suggested stock prices can also be predicted by using existing patterns. Finally, people can improve the prediction accuracy by combining AI technology and HMM according to Hassan and Nath, which is the future researches in this field. After reading these all five journals, this paper focus on the trading history on the stock market of Alibaba Group. Then researcher sees the history data during the last five years as training data, so it is possible to make reasonable predictions for the following days based on the transition matrix, which is founded from the training data.

1. Introduction

Due to the tremendous interests of several large corporations, making stock market predictions has always been one of the most activated research fields. There are several kinds of machine learning algorithms that have been applied to this area in the past, which achieved different levels of success. However, forecasting stock prices has been restricted by huge factors regarding its characteristics of volatility, seasonality and unpredictability [1]. Predicting stock prices will be a more challenging task if people are trying to analysis the stock market merely relied on old market data.

Hidden Markov Model (HMM) is a Markov model that can observe data in order and construct modelling concerning concealed state transition [2]. Stock market prediction can be seen as a problem that follows that same pattern. Stock prices depend on multiple elements, which have potential effects on investors. Transitions among hidden observations appear to be influenced by company policies, financial conditions and administrative decisions [3]. As a result, the stock market may show a turbulent situation, and people believe that HMM is one of the natural predictors of forecasting stock prices. My research question is how to predict stock prices by using Markov Chain models. This paper will provide technical information with the first four journals and explains several case studies with the last journals.

Eugene F. Fama, who is the 2013 Nobel Laureate in economics sciences, asserts it is not enough to predict the stock price with past behaviours of security [4]. Most data analysts share a common assumption that previous actions of securities are loaded with information regarding its future behaviours since past price behaviours is more likely to repeat itself as future behaviours. However, Fama argues this premise is contradicted by the theory of random walks, which states successive price modifications are self-dependent, and the price difference series has no memory. In other words, the past cannot forecast the future path in any eloquent way. Fama also tests the empirical validity of the random walk model.

Fama also examines the hypothesis of the random walk model, which are (1) successive price changes are self-reliant, and (2) the price differences correspond to certain probability distributions [5]. In addition to hypothesis examination, Fama introduces the "efficient" market, a market where numerous intelligent profit-maximizers are competing assiduously and where significant data is accessible to all sharers [6]. More importantly, he poses challenges to both data analysts and advocates of random walk theory. For data analysts, they need to show they can continually use their patterns to make meaningful stock price predictions, while stock advisors must demonstrate they can make a better selection based on additional fundamental analysis on new information, which has not been entirely considered in the current market [7]. For both journals, Fama discussed both the benefits and drawbacks of the theory. There are also some tests conducted on the theory which provide an objective view to the audiences. The author not only makes recommendations in terms of technical assistance but also gives criticism from all aspects.

Grant Mcqueen and Steven Thorley, who are Professors of Finance at BYU Marriott School of Management, conduct examination of predicting stock prices by using Markov Chains, which is based on Fama's journal on 1970. The random walk theory infers that there is no structure in the Markov Chain representations, and the transition probabilities are equal, which means the current returns are unrelated of the primary returns. Mcqueen and Thorley profess that this new approach solves the problem of nonlinearity by enabling the transition probabilities changes over different observations in a given set of sequences [8]. This journal also presents an investigation of the random walk hypothesis based on the statistical theory of finite-state Markov processes [9]. In particular, the authors conclude that there is a more progressive weekly serial correlation in small stocks rather than large stocks, while the large stocks appear to follow a random walk model from 1978 to 1988 [10]. This journal firstly describes the statistical issues with specific data and yields recommendations for solutions towards these issues by using asymptotic tests or Fisher's Exact Test, which provides the readers with objective opinions on this topic.

Md. Rafiul Hassan and Baikunth Nath, who are Professors of Computer Science and Software Engineering at the University of Melbourne, maintain human traders cannot always win in a stock market by its intricate and capricious nature. Hassan and Nath used only one HMM that is trained on the past dataset of the chosen airlines to forecast the stock prices of these companies, which is to detect patterns from the previous market behaviours that compete with today's market and intersperse these data with suitable connecting price component of the variable of interest [11]. The authors also point out their further research direction in developing a hybrid system using AI paradigms with HMM in terms of better accuracy [12].

Jerold B. Warner and Stephen J. Brown, who are Professors of Finance at the University of Rochester Simon School and Professor of Finance at New York University Stern School of Business, study the characteristics of the daily stock returns and show how these specific data properties influence case study methodologies for evaluating the share price influence concerning company events. This journal also reviews the fact that as the sample size of individual security grows, the statistical distribution tend to be normal [13]. This finding has led researchers to conclude that most observations cluster around the central peak and the probabilities for values further away from mean taper off equally in both directions. Specific methodological issues tend to appear when using daily data, and the authors examine an experimental design in order to analyze these issues. Warner and Brown also claim that the Raw Return methodology was misspecified, and Mean Adjusted Returns have lower power in situations with event-date clustering [14]. Moreover, the fact that samples of NYSE (New York Stock Exchange) securities demonstrating remarkably higher power than AMEX (American Stock Exchange) securities suggests that exchange-listing is an essential correlate of the power of the different tests [15].

Fama provides fundamental knowledge regarding the behaviours in the stock market and discusses the hypothesis of the random walk theory. The framework Mcqueen and Thorley have develop based on Fama's statement and test the data set using Markov Chains. They conclude that there is more positive weekly serial correlation in small stocks rather than in large stocks. Hassan and Nath examined a dataset using HMM, to predict whether the stock price is reasonable and has a sound statistical foundation. They show there is a considerable potential to use HMM in the stock market. If they can combine HMM with AI paradigms, the accuracy and efficiency of their predication can be improved dramatically. Warner and Brown examine the characteristics of daily stock returns and establish the relationship between these returns with event study methodologies. In their journal, they use data from AMEX securities and NYSE securities to show the sample formation by different trading frequency. More importantly, it appears that the explicit recognition of the properties of daily data can be advantageous in some instances.

In terms of chronology, Fama first discussed the existing patterns in the stock market in his journal article, The Behaviour of the Stock Market Prices, at the beginning of 1965. He emphasized that the conventional assumption, which is the past patterns that always repeat itself in the future, cannot be applied in the stock market. Later in the same year, Fama wrote another journal that further explained the concepts of random walk theory, which suggests that each change in the market is independent of previous changes. During the year of 1985, Warner and Brown then examined properties of the daily stock returns and interpret how these particular characteristics of these data affect event study methodologies [16].Six Years later, Mcqueen and Thorley conducted a test on stock returns using Markov Chains, which is based on the two journals Fama wrote in 1965. With the help of random walks theory, the calculation difficulties have been reduced with the help of simplified equations. Recently, Hassan and Nath used the past dataset for some chosen airline companies in order to train a Hidden Markov Model (HMM), which is helpful for researchers forecasted the stock prices in 2005. They also claim that people can combine AI technology with HMM for better accuracy in prediction. In conclusion, it appears that the transition of methods used in order to predict stock prices has been well developed, and this recent study in 2005 has overturned the past premise on random walks theory, which suggested stock prices can also be predicted by using existing patterns.

With technical development and further research in the stock market, the framework has led researchers to consider HMM in the market. As in the stock market, people can only get information about the daily stock prices and trading volume, instead of the current status of the market. Under this circumstance, the current status of the market is considered as a hidden observation of HMM. As a result, it is not hard to find out that people can construct a model base on the given information and hidden information by using HMM. By definition, HMM is a statistical model, which is used to describe a Markov Process with unobserved states [17].

Furthermore, it appears that HMM offers better results than simple Markov Chains, which allows more sequences to be found. Unlike the traditional Markov Chains, HMM assumes that the data observed is not the actual status of the case, but instead, they are generated by the underlying hidden observations. This finding has led researchers to conclude that it is complicated to understand the data from hidden information, yet the Markov property of HMM makes inferences more efficient. In addition, HMM has a stronger statistical foundation than a simple Markov Chain, and an efficient learning algorithm can take place directly from raw data.

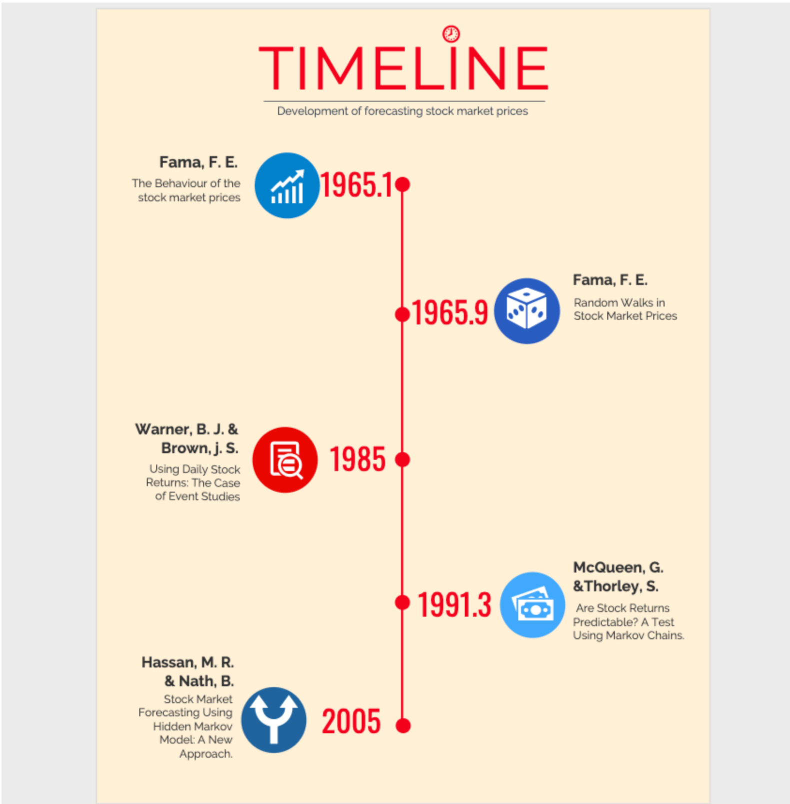

Figure 1: Timeline of published journals. From *The Journal of Business,* by E. F. Fama, (1965). From *Financial Analysts Journal,* by E. F. Fama, (1965). From *Journal of Financial Economics,* by B. J. Warner & J. S. Brown, (1985). From *5th International Conference on Intelligent Systems Design and Applications (ISDA05),* by M. R. Hassan & B. Nath, (2005).

In addition to the advantages of HMM, there are three significant problems in HMM modelling. First, it has been shown that classification and recognition problems are primary problems in HMM [18]. For example, how to find the probability that the observations are generated by HMM with a sequence of observations. Another one is called the learning problem [19], which means how researchers can modify the parameters in HMM to maximize the probability of producing the observed sequences. Last but not least, the difficulties concerning state sequence in HMM and produced the observations have been found given an HMM model and a sequence of observations.

Nevertheless, hidden Markov models have found full dedication in many distinct fields in terms of computational efficiency and flexibility. They are known for their use in temporal pattern recognition and generation, which is useful for researchers to identify the patterns and then use these patterns to predict stock prices for chosen airlines [20].

According to Hassen and Nath, people can improve the prediction accuracy by combining AI technology and HMM, which is the future researches in this field [21]. Other than the AI algorithm, machine learning techniques can be another tool to unearth patterns and insights people did not see before and be used to make unerringly accurate predictions.

2. Methods

2.1. Visual Description of Markov Chains

A Markov chain is a stochastic process, but it differs from a general stochastic process in that a Markov chain must be "memory-less." That is, the probability of future actions are not dependent upon the steps that led up to the present state, which is called the Markov property. In other words, the known present and future are independent of the past [22].

The Markov Chains prediction model is a random process, which study the states of events and transitions patterns among these states, with unique Markov memoryless property.

2.2. Markov Chain and Their Transition Probability

2.2.1 Definition 1

A (discrete-time) Markov chain with (finite or countable) state space χ is a sequence $x_0, x_1, \ldots$of χ-valued random variables such that for all states $i, j, k_0, k_1, \ldots$and all times $n = 0,1,2, \ldots$,

$$P(X_{n+1} = j || X_n = i, X_{n-1} = k_{n-1}, \ldots) = p(i,j) \quad (1)$$

where $p(i,j)$ depends only on the states i,j, and not on the time n or the previous states $k_{n-1}, k_{n-2}, \ldots$. The numbers $p(i,j)$ are called the transition probabilities of the chain.

2.2.2. Proposition 1

If X_n is a Markov chain with transition probabilities $p(x,y)$, then for every sequence fo states $x_0, x_1, \ldots, x_{n+m}$,

$$P(X_{m+i} = x_{m+i} \forall 0 < i \leq n || X_i = x_i \forall 0 \leq i \leq m) = \prod_{i=1}^{n} p(x_{m+i-1, x_{m+i}}) \quad (2)$$

Consequently, the n-step transition probabilities

$$p_n(x,y) := P(X_{n+m} = y || X_m = x) \quad (3)$$

depend only on the time lag n and the initial and terminal states x, y, but not on m.

2.3. Chapman-Kolmogorov Equations and the Transition Probability Matrix

Assume henceforth that $\{X_n\}_{n\geq 0}$ is a discrete-time Markov chain on a state space χ with transition probabilities $p(i,j)$. Define the transition probability matrix P of the chain to be the $\chi \times \chi$ matrix with entries $p(i,j)$, that is, the matrix whose ith row consists of the transition probabilities $p(i,j)$ for $j \in \chi$:

$$P = (p(i,j))_{i,j\in\chi} \tag{4}$$

If χ has N elements, then P is an $N \times N$matrix, and if χ is infinite, then P is an infinite by infinite matrix. Also, the row sums of P must all be 1, by the law of total probabilities. A matrix with this property is called stochastic.

2.3.1. Definition 2

A nonnegative matrix is a matrix with nonnegative entries. A stochastic matrix is a square nonnegative matrix all of whose row sums are 1. A sub-stochastic matrix is a square nonnegative matrix all of whose row sums are 1. A doubly stochastic matrix is a stochastic matrix all of whose column sums are 1.

2.3.2. Proposition 2

The n-step transition probabilities $p_n(i,j)$ are the entries of the nth power P^n of the matrix P. Consequently, the n-step transition probabilities $p_n(i,j)$ satisfy the Chapman-Kolmogorov equations

$$p_{n+m}(i,j) = \sum_{k\in\chi} p_n(i,k)P_m(k,j) \tag{5}$$

2.3.3. Definition 3

A probability distribution on X is stationary if

$$\pi^T = \pi^T \tag{6}$$

2.4. Accessibility and Communicating Classes

2.4.1. Definition 4

A state j is said to be accessible from state i if there is a positive-probability path from i to j, that is, if there is a finite sequence of states $k_0, k_1, k_2, \ldots, k_m$ such that $k_0 = i, k_m = j$,and $p(k_t, k_{t+1}) > 0$ for each $t = 0,1,\ldots,m-1$. States i and j are said to communicate if each is accessible from the other. This relation is denoted by $i \to j$.

2.5. Irreducible Markov chains

If the state space is finite and all states communicate (that is, the Markov chain is irreducible) then in the long run, regardless of the initial condition, the Markov chain must settle into a steady state.

2.5.1. Theorem 1

An irreducible Markov chain χ_n on a finite state space χ has a unique stationary distribution π. Furthermore, if the Markov chain is not only irreducible but also aperiodic, then for any initial distribution v,

$$\lim_{n\to\infty} p^{v}\{X_n = j\} = \pi(j) \forall j \in \chi \qquad (7)$$

2.6. The Krylov-Bogoliubov Argument

The argument turns on the fact that the probability simplex $\mathfrak{R}$ is compact. This implies that it has the Bolzano-Weierstrass property: Any sequence of vectors in $\mathfrak{R}$ has a convergent subsequence. Fix a probability vector $v \in \mathfrak{R}$, and consider the so-called Cesaro averages

$$v_n^T := n^{-1} v^T P^k \qquad (8)$$

Observe that each average v_n^T is a probability vector (because an average of probability vectors is always a probability vector), and so each v_n^T is an element of $\mathfrak{R}$. Consequently, the sequence v_n^T has a convergent subsequence:

$$\lim_{k\to\infty} v_{nk}^T = \pi^T \qquad (9)$$

Claim: The limit of any subsequence of v_n^T is a stationary distribution for P.

2.6.1. Proof

Denote the limit by π, as in (9). Since the mapping $\mu^T \to \mu^T P$ is continuous,

$$\begin{aligned}
\pi^T P &= \lim_{k\to\infty} v_{nk}^T P \\
&= \lim_{k\to\infty} n_k^{-1} \sum_{j=1}^{n_k} v^T P^j P \\
&= \lim_{k\to\infty} n_k^{-1} \sum_{j=2}^{n_k+1} v^T P^j \\
&= \lim_{k\to\infty} n_k^{-1} \{ \sum_{j=1}^{n_k} v^T P^j + v^T P^{n_k+1} - v^T P \} \\
&= \lim_{k\to\infty} n_k^{-1} \sum_{j=1}^{n_k} v^T P^j \\
&= \pi^T
\end{aligned} \qquad (10)$$

2.7. Strong Markov Property

2.7.1. Proposition 10. [Strong Markov Property][27]

Let T be a stopping time for the Markov chain $\{X_n\}_{n\geq 0}$. Then the Markov chain "regenerates" at time T, that is, the future $X_{T+1}, X_{T+2}, \ldots$is conditionally independent of the past $X_0, X_1, \ldots, X_{T-1}$given the value of T and the state $X_T = x$ at time T. More precisely, for any $m < \infty$ and all states $x_0, x_1, \ldots, x_{n+m} \in \chi$ such that $T = m$ is possible,

$$P(X_{T+i} = x_{m+i} \forall 1 \leq i \leq n || T = mand X_i = x_i \forall 0 \leq i \leq m) = \prod_{i=1}^{n} p(x_{m+i-1}, x_{m+i}) \quad (11)$$

2.8. Recurrence and Transience

2.8.1. Definition

Let $\{X_n\}_{n \geq 0}$ be a Markov chain on a finite or countable state space χ, and for any state x let $T_x = T_x^1$ be the first passage time to x. A state x is
(a) recurrent if $P^x\{T_x < \infty\} = 1$;
(b) transient if $P^x\{T_x < \infty\} < 1$;
(c) positive recurrent if $E^x T_x < \infty$; and
(d) null recurrent if it is recurrent but $E^x T_x = \infty$.

3. Results

This paper has collected daily stock market data from Alibaba Group from 2014-09-19 to 2019-11-18. Since Alibaba has made a name for itself in the last 5 years as the “Amazon of China”. The company has become a true force in the commerce space, not just as an online platform but as a partner and facilitator in every step of the way. The company has been growing at tremendous rates and its core business is incredibly profitable. It should be noticed that there are open prices, close prices, high and low prices, and, more importantly, the daily trading volume, which indicates the tendency of the stock market. Then the growth rate can be found by using the formula below,

$$\frac{[P_{current} - P_{prior}]}{P_{prior}} \quad (12)$$

where p stands for price.

With the help of excel functions, it is not hard to get the growth rate for these five years.

Similarly, the average and standard derivation can be calculated by using the functions, which are -0.000044953700549 and 0.0202264285924483, respectively. Since the lowest growth rate is -0.087760276925148 and the highest growth rate is 0.0886845473589185, it is reasonable to split the data into six different groups with the same range of 0.02, that matches the value of standard deviation and the average number.

Then, the data can be split into six ranges as shown below.

Table 1: Range States of Alibaba Group Stock Growth Rates From 2014/09/19 to 2019/11/18.

Growth rate States	1	2	3	4	5	6
range	$(-\infty, -0.04]$	$(-0.04, -0.02]$	$(-0.02, 0]$	$(0, 0.02]$	$(0.02, 0.04]$	$(0.04, \infty)$
frequency	15	135	478	457	179	31

Next, seeing the frequencies as denominator and number of transitioning states as the numerator, which gives the transition matrix below.

Table 2: Transition Matrix of Alibaba Group Stock From 2014/09/19 to 2019/11/1.

Transition Matrix	1	2	3	4	5	6
1	1/30	4/30	5/30	8/30	10/30	2/30
2	5/135	25/135	37/135	46/135	20/135	2/135
3	17/478	48/478	183/478	177/478	46/478	7/478
4	3/457	37/457	183/457	155/457	70/457	9/457
5	2/179	17/179	62/179	62/179	29/179	7/179
6	1/31	3/31	8/31	13/31	4/31	2/31

Then since this paper is desired to predict the stock prices in n days, then it is required to multiply this matrix to its nth power, which will give the following the matrixes from the first to seventh power.

Table 3: Predicted Transition Matrix in one day.

Transition Matrix	1	2	3	4	5	6
1	0.03333333	0.13333333	0.16666667	0.26666667	0.33333333	0.06666667
2	0.03703704	0.18518519	0.27407407	0.34074074	0.14814815	0.01481481
3	0.03556485	0.10041841	0.38284519	0.37029289	0.09623431	0.01464435
4	0.00656455	0.08096280	0.40043764	0.33916849	0.15317287	0.01969365
5	0.01117318	0.09497207	0.34636872	0.34636872	0.16201117	0.03910615
6	0.03225806	0.09677419	0.25806452	0.41935484	0.12903226	0.06451613

Table 4: Predicted Transition Matrix in two days.

Transition Matrix	1	2	3	4	5	6
1	0.01960234	0.10557125	0.34535021	0.34989463	0.15035522	0.02922635
2	0.02221068	0.10984486	0.35343772	0.34755943	0.14426133	0.02268598
3	0.02249898	0.10231940	0.36541095	0.35053087	0.13777405	0.02146576
4	0.02203214	0.09999271	0.37054187	0.35396526	0.13202675	0.02144126
5	0.02155390	0.10107296	0.36540375	0.35359055	0.13547475	0.02290409
6	0.02011327	0.10058675	0.35996592	0.35111635	0.14338738	0.02483033

Table 5: Predicted Transition Matrix in three days.

Transition Matrix	1	2	3	4	5	6
1	0.02205860	0.10213062	0.36543228	0.35185512	0.13649931	0.02202408
2	0.02205873	0.10213719	0.36542025	0.35185169	0.13650657	0.02202557
3	0.02205968	0.10213319	0.36542874	0.35185224	0.13650241	0.02202374
4	0.02206032	0.10213188	0.36543251	0.35185210	0.13650056	0.02202263
5	0.02206001	0.10213156	0.36543270	0.35185271	0.13650017	0.02202284
6	0.02205978	0.10212969	0.36543594	0.35185391	0.13649813	0.02202255

Table 6: Predicted Transition Matrix in four day.

Transition Matrix	1	2	3	4	5	6
1	0.02203659	0.10212547	0.36538874	0.35188352	0.13651294	0.02205273
2	0.02205026	0.10219048	0.36530892	0.35183867	0.13656420	0.02204747
3	0.02205925	0.10213807	0.36541518	0.35185049	0.13650999	0.02202702
4	0.02206516	0.10211826	0.36547020	0.35185190	0.13648236	0.02201212
5	0.02205966	0.10211887	0.36545695	0.35185911	0.13648579	0.02201962
6	0.02205305	0.10210308	0.36546843	0.35187831	0.13647135	0.02202577

Table 7: Predicted Transition Matrix in five days.

Transition Matrix	1	2	3	4	5	6
1	0.02176542	0.10227975	0.36414867	0.35208823	0.13713357	0.02258436
2	0.02200388	0.10283032	0.36411727	0.35158924	0.13722555	0.02223374
3	0.02206826	0.10218387	0.36531467	0.35178461	0.13660569	0.02204291
4	0.02210654	0.10193585	0.36594178	0.35193114	0.13619092	0.02189376
5	0.02203113	0.10199485	0.36561277	0.35194972	0.13638705	0.02202447
6	0.02190601	0.10190424	0.36540458	0.35209559	0.13646300	0.02222657

Table 8: Predicted Transition Matrix in six days.

Transition Matrix	1	2	3	4	5	6
1	0.02205983	0.10213276	0.36543004	0.35185231	0.13650168	0.02202338
2	0.02205982	0.10213280	0.36542993	0.35185230	0.13650174	0.02202341
3	0.02205983	0.10213278	0.36542997	0.35185230	0.13650172	0.02202340
4	0.02205983	0.10213278	0.36542998	0.35185229	0.13650172	0.02202339
5	0.02205983	0.10213277	0.36542999	0.35185230	0.13650171	0.02202339
6	0.02205984	0.10213276	0.36543003	0.35185230	0.13650169	0.02202338

Table 9: Predicted Transition Matrix in a week.

Transition Matrix	1	2	3	4	5	6
1	0.02205980	0.10213251	0.36543053	0.35185249	0.13650134	0.02202332
2	0.02205973	0.10213310	0.36542928	0.35185229	0.13650204	0.02202356
3	0.02205981	0.10213281	0.36542990	0.35185230	0.13650176	0.02202342
4	0.02205986	0.10213273	0.36543010	0.35185227	0.13650167	0.02202335
5	0.02205986	0.10213269	0.36543020	0.35185231	0.13650160	0.02202334
6	0.02205987	0.10213251	0.36543059	0.35185239	0.13650137	0.02202327

Expected values are needed to be found out for each transition matrixes in order to compare with the real growth rates.

Therefore, the results are shown below

Table 10: Expected Stock Values with Different Close States for one day.

Prediction Period : 1 day	
Close states	Expected Values
1	0.00866667
2	(0.00155556)
3	(0.00129707)
4	0.00221007
5	0.00340782
6	0.00419355

Table 11: Expected Stock Values with Different Close States for two days.

Prediction Period : 2 days	
Close states	**Expected Values**
1	0.00187016
2	0.00099748
3	0.00086318
4	0.00076571
5	0.00098143
6	0.00143138

Table 12: Expected Stock Values with Different Close States for three days.

Prediction Period : 3 days	
Close states	**Expected Values**
1	0.00096596
2	0.00091807
3	0.00089609
4	0.00087691
5	0.00089480
6	0.00091970

Table 13: Expected Stock Values with Different Close States for four days.

Prediction Period : 4 days	
Close states	**Expected Values**
1	0.00089738
2	0.00089637
3	0.00089390
4	0.00089209
5	0.00089303
6	0.00089378

Table 14: Expected Stock Values with Different Close States for five days.

Prediction Period : 5 days	
Close states	Expected Values
1	0.00089356
2	0.00089374
3	0.00089351
4	0.00089337
5	0.00089340
6	0.00089337

Table 15: Expected Stock Values with Different Close States for six days

Prediction Period : 6 days	
Close states	Expected Values
1	0.00089346
2	0.00089349
3	0.00089347
4	0.00089346
5	0.00089346
6	0.00089345

Table 16: Expected Stock Values with Different Close States for one week.

Prediction Period : 7 days	
Close states	Expected Values
1	0.00089347
2	0.00089347
3	0.00089347
4	0.00089347
5	0.00089347
6	0.00089347

4. Conclusions

Table 17: Actual Alibabba Group Stock Prices and Growth Rates from 2019-11-15 to 2019-12-02.

Date	Open	High	Low	Close	Volume	Growth rate
2019-11-15	184	185.600006	183.710007	185.490005	11296400	
2019-11-18	186.979996	186.979996	184.160004	184.610001	11822900	-0.00474421249813433
2019-11-19	186.309998	186.710007	183.869995	185.250000	13407200	0.003467
2019-11-20	183.669998	183.699997	181.059998	182.350006	16684600	-0.015654
2019-11-21	181.770004	184.889999	181.600006	184.860001	10254700	0.013765
2019-11-22	185.800003	186.779999	183.934998	186.779999	10541000	0.010386
2019-11-25	188.320007	190.720001	187.880005	190.449997	19157700	0.019649
2019-11-26	190.389999	195.000000	189.039993	194.699997	51832300	0.022316
2019-11-27	197.240005	200.979996	197.000000	200.820007	33040500	0.031433
2019-11-29	199.809998	200.429993	198.350006	200.000000	18593100	-0.004083
2019-12-02	198.580002	198.669998	193.509995	196.309998	19357700	-0.018450

As shown on the history data, the state on 2019-11-18 was state 3, and the actual growth rate for the following day was 0.003467. On the contrast, the one-day prediction table indicates that the expected growth rate for end state 3 should be -0.00129707. It may show a distinct difference in terms of specific numbers, but the actual growth rate is in state 4 and the expected growth rate is in 3, which means the expected growth rate should be considered be as relatively accurate. Since the range within each state is 2%, and the difference actual and expected growth rate is $0.003467 - (-0.00129707) = 0.476497\%$. It is not hard to see that the difference is much smaller that the difference of ranges. In other words, the Markov model is able to make precise prediction on stock market.

Similarly, the two-days prediction table gives the expected growth rate for end state 3 is 0.00086318, and the actual growth rate is -0.015654. the difference between the actual and the expected growth rate is $0.00086318 - (-0.015654) = 1.651718\%$, which is also within the rage of states (even though they are in different states).

The three-days prediction table gives the expected growth rate for end state 3 is 0.00089609, and the actual growth rate is 0.013765. Based on table 1, it is clear that both growth rate are in state 4.

Furthermore, the four-days prediction table shows that the expected growth rate for end stated 3 is 0.00089390, and the actual growth rate is 0.010386. It is not hard to verify from table 1 that both growth rates are from same state.

Last but the least, the seven-days prediction table shows that the expected growth rate for end state 3 is 0.00089347, and the actual growth rate is 0.019649. Both growth rates are from states 4, and their differences are within the range of states.

Finally, it worth credit to conclude that stock market can be predicted by using Markov Model with memoryless property within a certain range of error.

References

[1] *Fama, F. E. 1965. The Behaviour of STock-Market Prices. The Journal of Business. Retrieve from https://www.jstor.org/stable/pdf/2350752.pdf*

[2] *Fama, F. E. 1965. Random Walks in Stock Market Prices. Financial Analysts Journal. Retrieve from https://www.jstor.org/stable/pdf/4469865.pdf*

[3] *McQueen, Grant & Thorley, Steven. 1991. Are Stock Returns Predictable? A Test Using Markov Chains. The Journal of Finance. Retrieve from https://www.jstor.org/stable/pdf/2328695.pdf?casa_token=QNoDrjtvYncAAAAA:HCU92OseCYO3T-4MWSFX7cnqcZGiPgknIjZOPXnp1B6smwrSaAktaSRXtN11roNBl4JFl75LnkWWCV6lIqaqI-vBT8ZCRB-Klz-a17puS8Kv4NpanEwt*

[4] *Hassan, M. Rafiul & Nath, Baikunth. 2005. Stock Market Forecasting Using Hidden Markov Model: A New Approach. 5th International Conference on Intelligent Systems Design and Applications (ISDA05). Retrieve from https://s3.amazonaws.com/academia.edu.documents/1904664/5kcoxioksgt4hgw.pdf?response-content-disposition=inline%3B%20filename%3DStock_Market_Forecasting_Using_Hidden_Ma.pdf&X-Amz-Algorithm=AWS4-HMAC-SHA256&X-AmzCredential=AKIAIWOWYYGZ2Y53UL3A%2F20191111%2Fus-east-1%2Fs3%2Faws4_request&X-Amz-Date=20191111T031359Z&X-Amz-Expires=3600&X-Amz-SignedHeaders=host&X-Amz-Signature=a8a7fd24901045f634cff7473581eb5ee3ebcf8af61b9f26d5d44407f8a4d34e7*

[5] *Warner, B. J. & Brown J. S. 1985. Using Daily Stock Returns: The Case of Event studies. Journal of Financial Economics. Retrieved from http://leeds-faculty.colorado.edu/bhagat/brownwarner1985.pdf*

[6] *Serfozo, R. 2009. Basics of Applied Stochastic Processes. Probability and its applications. Retrieved from http://www.stat.yale.edu/~jtc5/251/readings/Basics%20of%20Ap plied%20Stochastic%20Processes_Serfozo.pdf*

[7] *Pitman, W. James. Recurrence and Transience. Probability Theory. Retrieved from https://www.stat.berkeley.edu/~pitman/s205s03/lecture3.pdf*

[8] *Lorenzi, Luca. 2017. Analytics Methods for Kolmogorov Equations. Monographs and Research notes in mathematics. Retrieved from https://books.google.ca/books?id=vRkNDgAAQBAJ&pg=PA430&lpg=PA430&dq=The+Krylov-Bogoliubov+Argument+formula+and+proof&source=bl&ots=vdxXERSPnv&sig=ACfU3U2MIbZXAfSZP6ik8fYhm-lxF9DzMIg&hl=zh-CN&sa=X&ved=2ahUKEwil54_RmMDnAhVQBc0KHS5rCWoQ6AEw-CHoECAoQAQ#v=onepage&q=The%20Krylov-Bogoliubov%20Argument%20formula%20and%20proof&f=false*

[9] *Heinz, Sebastian. 2018. A simple deep learning model for stock price prediction using TensorFlow. Medium. https://medium.com/mlreview/a-simple-deep-learning-model-for-stock-price-prediction-using-tensorflow-30505541d877*

[10] *Siegrist, Kyle n.d. 15. Markov Processes. Markov Processes. Retrieved from https://www.randomservices.org/random/markov/index.html*

[11] *Singh, Aishwarya. 2019. Predicting the Stock Market Using Machine Learning and Deep Learning. Analytics Vidhya. Retrieved from https://www.analyticsvidhya.com/blog/2018/10/predicting-stock-price-machine-learningnd-deep-learning-techniques-python/*

[12] *Yates, Tristan. 2020. 4 Ways to Predict Market Performance. Investopedia. Retrieved from https://www.investopedia.com/articles/07/mean_reversion_martingale.asp*

[13] *n.d. 2020. Alibaba Group Holding Limited (BABA) Stock Price, Quote, History & News. Yahoo! Finance. Retrieved from https://finance.yahoo.com/quote/BABA/*

[14] *Shreve, Ken. 2020. Is Alibaba Stock A Buy Right Now? Here's What Earnings, Chart Show. Investor's Business Daily. Retrieced from https://www.investors.com/research/alibaba-stock-buy-now/*

[15] *Sun, Leo. 2019. 8 Things to Know About Alibaba's Stock. The Motley Fool. Retrieved from https://www.fool.com/investing/2019/12/12/8-things-to-know-about-alibaba-stock.aspx*

[16] *n.d. 2020. Alibaba: Can't Ignore The Upside. Seeking Alpha. Retrieved from https://seekingalpha.com/article/4320761-alibaba-cant-ignore-upside*

[17] *Fine, S., Singer, Y. & Tishby, N. 1998, 07. The Hierarchical Hidden Markov Model: Analysis and Applications. Machine Learning 32, 41–62. Retrieved from https://doi.org/10.1023/A:1007469218079*
[18] *Bicego, M., Murino, V., Figueired, M. 2002.12. Similarity-based classification of sequences using Markov Models. Elsevier Ltd. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.115.2237&rep=rep1&type=pdf*
[19] *Nguyen, Loc. 2016.06. Tutorial on Hidden Markov Model. Applied and Computational Mathematics. Retrieved from http://article.sciencepublishinggroup.com/html/10.11648.j.acm.s.2017060401.12.html*
[20] *Kang, Eugine. 2017.08. Hidden Markov Model. Medium, Retrieved from https://medium.com/@kangeugine/hidden-markov-model-7681c22f5b9*
[21] *Rocca, Joseph. 2019.8. Introduction to Markov chains. Medium. Retrieved from http://towardsdatascience.comar/brief-introduction-to-markov-chains-2c8cab9c98ab*