Index Forecast Study Based on Amended Weighted Markov Chain in China

Yanpeng Sun

Abstract—Shanghai Composite Index is one of the most representative indexes of Chinese stock market index that is the gauge of national economy and the direction to economic development, therefore, the forecast study on Shanghai Composite index is of great significance in theory and practice. Through analysis, fluctuation of the short-term Shanghai Composite Index conforms to fundamental assumption of Markov chain forecast. This paper mainly amended weighted Markov chain model to study and forecast short-term trend of Shanghai Composite Index, hoping to facilitate investors and potential investors to make investment decisions and provide reference for them.

Index Terms—Amended weighted Markov chain, composite index in China, trend forecast.

I. INTRODUCTION

This paper aims to amend weighted Markov chain model to study and forecast short-term trend of Shanghai Composite Index by use of the MATLAB tool so that the model can give references and directions to economic individuals to make better investment decisions. The stock market index is an index describing the market price level and its change. The investors will bear the price risk due to the fluctuation of stock price. The rational investors want to know the related information about stock price before investing. It is easier for investors to understand the trend of individual stocks, but more difficult for investors to understand the trend of many stocks or even the whole market. Therefore, in order to meet their needs, some financial institutions make full use of the public information, professional knowledge and ability in the market and compile stock market index through certain rules to help investors understand the stock market more clearly. The stock market index generally synthesizes the representative stocks of various industries, which helps investors evaluate the investment effect, and helps investors make prediction and judgment on the market. In addition, the stock market reflects the current economic situation. Therefore, the government, enterprises and other institutions will take the stock market index as a reference index, and judge the current economic situation. High-quality stock market index can accurately reflect the market information and be widely used in the investment community. With the advent of the era of big data, more and more statistical learning models have been applied in the financial field. Compared with traditional artificial methods, statistical learning model has great advantages. An important application of statistical learning model in the financial field is stock market prediction. When trading stocks, rational investors usually do a lot of analysis on historical data and so on, and then buy and sell stocks. However, this method of manually analyzing data is time-consuming and laborious, and it is easy to generate many errors, which will increase investment risks and lead to investment losses. Statistical learning model can make deep and accurate analysis on a large number of data and give more accurate prediction results in a short time. Compared with manual analysis, statistical learning model improves investment efficiency and accuracy. Statistical learning model can not only process the data magnitude that cannot be processed manually, but also can browse news and social networking sites through computer technology to collect and process more information related to investment, which can not only increase the information sources of investors, but also save investors' time. In recent years, support vector machines, neural networks, genetic algorithms and other statistical pattern recognition algorithms have been widely used in stock market prediction. The application of statistical pattern recognition algorithm in stock price prediction has become a hot topic in the financial circle. A large number of conclusions show that the statistical learning model is effective in predicting stock prices, and the combination of artificial intelligence and finance is the future development trend.

With the development of the world economy, the world finance is in a stage of rapid development and the financial activities are increasing. The uncertainty of the change trend of financial activities is also increasing. How to learn and master the rules of financial activities and predict the future change trend of financial activities become the focus and the main research content of academic and financial areas. Financial prediction can effectively provide the basis for making financial plans and decisions, and then maintain the healthy development of the financial market and maximize the profit of financial organizations. The more accurately the stock price tendency is predicted, the more correctly investors make decisions on investment portfolios. This paper is based on the Markov Chain Model, but considers the fluctuation of stock price, and thus increases the accuracy to some extent.

This paper is mainly divided into the following parts. Firstly, the works that have been accomplished by scholars

Manuscript received February 2, 2020; revised July 18, 2020. This work is supported by Shandong social science planning project (19CFZJ42), Qingdao social science planning project (QDSKL1901115), and 2019 National Statistical Science Research Project (2019LY31).

Yanpeng Sun is with the School of Economics, Qingdao University, Qingdao Shandong, 266071, China (e-mail: jjxy_syp@163.com).

will be concluded briefly and the amended weighted Markov Chain Model will be introduced. Secondly, the theoretical model will be presented. Then, the main body of this paper comes. Empirical analysis will be demonstrated in detail. The overall conclusion will be showed at last.

II. LITERATURE

The origin of the Markov Model can be traced back to the second half of the 1960s, when Baum et al. presented the original prototype of the Model in a series of statistical papers. Firstly, Baum (1966) [1] published a paper on the initial prototype of Markov model, and gave a statistical inference on the probability function of finite state Markov chain. Then, Baum (1970, 1972) [2], [3] respectively gave a maximization technique for Markov chain probability function and a correlation inequality and maximization technique for statistical estimation of Markov process probability function, which was a further expansion of 1996 and laid a foundation for later work on Markov model. Later, Ryan & Nudd (1973) [4] summarized the relevant content and application field of Viterbi algorithm, and extended the algorithm. Then, Levinson et al. (1983) [5] gave the application of Markov chain probability function theory in automatic speech recognition, combined theory with practical problems, and gave a model suitable for isolated word recognition. And Rabiner & Juang (1986) [6] summarized the relevant contents of Markov Model on the basis of previous researches. Krishnalal [7] et al. combined Markov model with support vector machine and applied it to text mining and news classification.

Markov Model is also widely used in the financial field. Hassan & Nath (2005) [8], [9] first applied Markov Model to stock price prediction and gave a new method to predict stock price. This method takes the opening price, closing price, maximum price and minimum price of the stock price as the input of the model, and predicts the stock price through parameter estimation, state decoding and other steps. Later, Srivastava et al (2008) [10] applied Markov Model to credit fraud detection, and the empirical results showed that it was effective to use Markov Model to detect information fraud. And Hassan (2009) [11] combined Markov Model with fuzzy Model and applied the Model to stock price prediction. The Model USES Markov Model for data pattern recognition, then USES fuzzy logic to obtain predictions and can test stock markets from different industries. The empirical results show that the prediction accuracy of the improved model is obviously improved. Recently, Caccia & Remillard (2017) [12] proposed the multiple autoregressive Markov Model on the basis of the Markov Model. On the basis of the improved Model, the likelihood ratio test and a new goodness of fit test were conducted to test the S&P 500 daily return rate, and it was found that the improved Model had obvious advantages. Domestic research on Markov Model started relatively late, but in recent years, the application of Markov Model in the financial field has attracted more and more attention from domestic experts, scholars and investment professionals.

Using Shanghai Composite Index and Dow Jones

Industrial Average, Jiang & Xu (2013) [13] proposed a stock index performance forecast method based on grey residual model and BP neural network. Shi (2014)[14] provided an ARIMA model based on wavelet analysis based on monthly average close price of the Shanghai Composite Index and forecast long-term trend of index price. Song (2014) [15] determined share price inflection points and forecast share price trend by smoothing share price fluctuations. Cui (2016) [16]introduced normal Markov Chain to analyze the Economic market with main economic data index of China, demonstrating that Markov Chain theory is practicable for economic data. Fei (2016) [17] analyzed characters of random data in different environment and demonstrated that Markov Chain shows good precision in describing and predicting economic data in multiple random environments. Lin & Yang (2017) [18] introduced improved hybrid neural network based on fuzzy granulation to study the stock market. Their study shows that the stock index is predictable and the stock market data has the Statistical characteristics when no force majeure occurring the stock market, and they get good decision in predicting price range of stock index. Liang (2016) [19] creatively introduced the adaptive network based fuzzy inference system(ANFIS) model to the stock market, using the statistical characteristics of the stock market data, analyzing the relationship between data change and the time flow, and on empirical mode their study shows good description for stock time series data. Their researches focus on accuracy of forecast values in share price forecast researches but neglect the random fluctuation of stock market, which inevitably compromises forecast accuracy. On the contrary, the amended weight Markov chain is used for forecasting future fluctuation range and state probability of research objects. It fully considers random fluctuation of the stock market, featuring forecast results which are more scientific and practicable.

III. THEORETICAL MODEL

Define probability space (Ω , F, P) the random sequence on the $X_0, X_1,...$ It's called markov chain Chain), if the following two conditions are met:

(1) the state space of $\{X_n: n \ge 0\}$ is a countable set;

(2) for any *n* and state $i_0, i_1, \dots, i_{n+1} \in \mathbb{I}$, as long as $P\{X_0=i_0, X_1=i_1, \dots, X_n=i_n\} > 0$, the following equation is true

 $P\{X_{n+1} = i_{n+1} | X_0 = i_0, X_1 = i_1, \dots, X_n = i_n\} = P\{X_{n+1} = i_{n+1} | X_n = i_n\}$

Condition (2) is called markov property (also known as no after effect), which is the basic characteristic of markov chain. It indicates that the state at time n+1 is only related to the state at time n, independent of the state before time n. Jane By itself, Markov property means that the future has nothing to do with the past, given the present.

To set up an amended weighted Markov chain forecast model including:

1) Sequencing forecast samples from small to large and partitioning to construct state space I;

2) Determining price index states in different time

buckets;

3) Testing Markov property;

4) Calculating all-order autocorrelation coefficient r_k according to the corresponding calculation formulas,

$$r_{k} = \frac{\sum_{I=1}^{n-k} (x_{I} - \bar{x})(x_{I+k} - \bar{x})}{\sum_{I=1}^{n} (x_{I} - \bar{x})^{2}} (k \in I)$$
(1)

where in r_k stands for autocorrelation coefficient of the k order, x_1 for price index of the first bucket, \overline{x} for mean price indexvalue, and n for length of price index sequence;

5) Normalizing all-order autocorrelation coefficients, wherein m is the maximum order to be calculated in forecast;

$$w_k = \frac{|r_k|}{\sum_{k=1}^m |r_k|} \tag{2}$$

6) Statistically analyzing results of the step 5) to obtain transition probability matrixes of Markov chain in different delays;

7) Forecasting state probability $p_i(k)$ of price index of this time bucket (wherein $i \in I$, k stands for delay, and k=1, 2, ...m) using former price index as initial states and combining corresponding transition probability matrixes;

8) Forecasting weights of various forecast probabilities in the same state and forecast probabilities of price index in this state, that is,

$$p_{i} = \sum_{k=1}^{m} w_{k} p_{i}^{(k)}$$
(3)

i in Max (pi, $i \in I$) is the forecast value of the index state in this time bucket.

IV. EMPIRICAL ANALYSIS

A. Choosing Samples

This paper uses weekly close price data of Shanghai Composite index (hereinafter refers to Shanghai Composite Index) from November 28, 2014 to July 17, 2015 as research samples, and price trend of sample data is shown as Fig. 1.



Fig. 1. Shanghai composite index week line trend (from Nov. 28, 2014 to Jul. 17, 2015).

Fig.1 shows that Shanghai Composite Index has presented a stable rise from 2000 to more than 5000 points before June 2015, reflecting an inspiring vision of China's economy. During this period, the stock market was soaring, and the prices of almost all kinds of stocks increased, from which the investors would benefit a lot. Therefore, people in all walks were willing to invest in stocks. However, constant fluctuation in the rising trend always exists. The objective law also indicates that the stock price will fall when it rises to a high point, that is the market is faced with much high downturn pressure. Since June 2015, Shanghai Composite Index temporarily dropped from the high point, which exposed enormous pressure in Chinese economic recovery process. Those who purchased securities at a relatively high price confronted huge loss. Facing huge fluctuation in the stock market is of great significance in social practice, which will guide people to make relatively correct decisions on investment portfolio.

 TABLE I: SHANGHAI COMPOSITE INDEX CLOSE PRICE DATA AND THEIR

 STATES (FROM SEP.19, 2014 TO JUL. 17, 2015).

NO.	date	closing	state
1	2014-09-19	2330.83	1
2	2014-09-26	2347.72	1
3	2014-09-30	2363.87	1
4	2014-10-10	2374.54	1
5	2014-10-17	2341.17	1
6	2014-10-24	2302.28	1
7	2014-10-31	2420.17	2
8	2014-11-07	2418.16	2
9	2014-11-14	2478.82	2
10	2014-11-21	2486.79	2
11	2014-11-28	2682.83	2
12	2014-12-05	2937.64	2
13	2014-12-12	2938.16	2
14	2014-12-19	3108.60	3
15	2014-12-26	3157.60	3
16	2014-12-31	3234.67	3
17	2015-01-09	3285.40	3
18	2015-01-16	3376.50	3
19	2015-01-23	3351.76	3
20	2015-01-30	3210.36	3
21	2015-02-06	3075.90	3
22	2015-02-13	3203.83	3
23	2015-02-17	3246.90	3
24	2015-02-27	3310.30	3
25	2015-03-06	3241.18	3
26	2015-03-13	3372.90	3
27	2015-03-20	3617.32	4
28	2015-03-27	3691.10	4
29	2015-04-03	3863.92	4
30	2015-04-10	4034.31	4
31	2015-04-17	4287.29	5
32	2015-04-24	4393.69	5
33	2015-04-30	4441.66	5
34	2015-05-08	4205.92	5
35	2015-05-15	4308.69	5
36	2015-05-22	4657.60	5
37	2015-05-29	4611.74	5
38	2015-06-05	5023.10	5
39	2015-06-12	5166.35	5
40	2015-06-19	4478.36	5
41	2015-06-26	4192.87	4
42	2015-07-03	3686.92	4
43	2015-07-10	3877.80	4
44	2015-07-17	3957.35	4

B. Preprocessing Data

The Shanghai composite index sample data is sequenced from small to large and averagely divided into five sample ranges according to the maximum difference, including state 1: close price lower than 2400, state 2: close price between 2400 (included) and 3000; state 3: close price between 3000 (included) and 3600, state 4: close price 3600 (included) and

4200 and state 5 [5]: close price above 4200 (included). See Table I for details.

Т	ABLI	E II:	BASIC	STATISTIC	FEATURES	OF	SAMPLE	DAT	A

Statistic Features	Average	Median	Standard Deviation	Variance	Skewness	Kurtosis
Values	3.4420e+003	3.2979e+003	796.3420	6.3416e+005	0.2943	2.1353

C. Modeling Data



Basic statistic distribution property [6] (shown in Table II and as Fig. 2 and Fig. 3) of sample data is calculated by MATLAB software. The statistic results indicated that the sample data has no normal distribution features and their trend fluctuates about 3442. It can be judged from autocorrelation test that autocorrelation coefficient within a short delay is always positive and then negative all along. Autocorrelation distribution is not of triangular symmetric, indicating that distribution features of Shanghai Composite Index sequence live up to basic assumption of the amended weighted Markov chain model. The model can be employed in analytical prediction of Shanghai Composite Index trend.

All-order autocorrelation coefficients are calculated using MATLAB software. The operation results will be shown as follows:

Thus it can be seen that autocorrelation coefficient of five delay orders is remarkable, so take autocorrelation coefficients of five delays, that is, k=1, 2, 3, 4 and 5 (see

Table III).

Г

TABLE III: ALL-ORDER	AUTOCORRELATION COEFFICIENTS

Autocorrelation	k=1	<i>k</i> =2	<i>k</i> =3	k=4	<i>k</i> =5
r_k	0.9515	0.8800	0.8075	0.7286	0.6554

All-order autocorrelation coefficients are normalized to obtain Markov chain weights [7] of all-order delays. See Table IV for operation results.

TABLE I	V: MARKOV	CHAIN WE	EIGHTS OF A	ALL-ORDER I	DELAYS
117 1 1		1 0	1.0		

weights	<i>k</i> =1	<i>k</i> =2	<i>k</i> =3	<i>k</i> =4	<i>k</i> =3
W_k	0.2365	0.2187	0.2007	0.1811	0.1629

D. Transition Probability Matrixes of Markov Chain

Transition probability matrixes of delays can be calculated after corresponding weights of all orders are calculated (see Fig. 5, Fig. 6, Fig. 7, Fig. 8 and Fig. 9).

	0.800	0.200	0	0	0
	0.125	0.750	0.125	0	0
$P_{1} =$	0	0	1.000	0	0
	0	0	0	0.857	0.143
	0	0	0	1.000	0.900
Fig	g. 5. One-d	lelay transi	tion proba	bility matr	ix P1.
	0.600	0.400	0	0	0
	0.125	0.625	0.250	0	0
$P_2 =$	0	0	0.917	0.083	0
	0	0	0	0.667	0.333
	0	0	0	0.200	0.800
Fig	g. 6. Two-c	lelay trans	ition proba	bility mat	rix P2.
	0.400	0.600	0	0	0]
	0.125	0.500	0.375	0	0
$P_3 =$	0	0	0.833	0.167	0
Ū	0	0	0	0.400	0.600
	0	0	0	0.300	0.700
Fig	. 7. Three-	delay trans	sition prob	ability mat	rix P3.
	0.200	0.800	0	0	0
	0.125	0.375	0.500	0	0
$P_{4} = 1$	0	0	0.750	0.250	0
-	0	0	0	0	1
	0	0	0	0.400	0.600
Fig	- g. 8. Four-c	lelay trans	ition proba	bility mat	rix P4.
	0	1.000	0	0	0
	0.125	0.250	0.625	0	0
$P_{5} =$	0	0	0.667	0.333	0
	0	0	0	0	1
	0	0	0	0.444	0.556
Fig	g. 9. Five-d	lelay trans	ition proba	bility mat	rix P5.

E. Forecasting Index

From the foregoing, with weekly close price data of the last five weeks (from Jun. 19, 2015 to Jul. 17, 2015) of Shanghai Composite Index as initial data and the corresponding transition probability matrixes according to the states, close price data in future two weeks of Shanghai Composite Index and state probabilities can be forecast, and see the calculation results in Table V.

First Day	States	Delay time/	Weights		State	Spa	ce Probab	oilities
		WCCK		1	2	3	4	5
06-19	5	5	0.2365	0	0	0	0.444	0.556
06-26	4	4	0.2187	0	0	0	0	1
07-03	4	3	0.2007	0	0	0	0.400	0.600
07-10	4	2	0.1811	0	0	0	0.667	0.333
07-17	4	1	0.1629	0	0	0	0.857	0.143
weight sum			0	0	0	0.501	0.499	

TABLE V: TREND FORECAST OF SHANGHAI COMPOSITE INDEX (FROM JUL. 18, 2015 TO JUL. 24, 2015)

Weights of various prediction probabilities in the same state and forecast probability of Shanghai Composite Index in this state as well as 'state space probabilities' in Table V are displayed in MATLAB in matrixes. The results are that:

$$p1 = 0; p2 = 0; p3 = 0; p4 = 0.5013; p5 = 0.4987$$

The probability of Shanghai Composite Index within the next trade week is highest in state 4, at 50.13%, so we consider fluctuation range of Shanghai Composite Index should be between 3600 (included) and 4200 around; using the state 4 as the first prediction week, Shanghai Composite Index trend of the following week (from Jul. 25, 2015 to Jul. 31, 2015) can be predicted by repeating the above steps on this basis, in this way, the obtained prediction range still falls in the state 4. With view to the slow-rising state of Chinese stock market, we firmly believe space of the state 4 is operation space of Shanghai Composite Index in one to two weeks, and see the actual fluctuation situations shown as Fig. 4.



Fig. 4. Forecast of Shanghai composite index trend range.

F. Back Test

The actual operation results of Shanghai Composite Index market shows that Shanghai Composite Index on July 24, 2015 closed at 4070.91 while at 3633.73 on July 31, 2015, and both values drop in the state 4, indicating that actual results are highly consistent with prediction results.

V. CONCLUSION

In conclusion, we can find that the amended weighted Markov chain can effectively forecast fluctuation ranges and probabilities of short-term trend of Shanghai Composite index. The predictive results fully indicated that slow-rising status of Chinese stock market will not change in the short term but Shanghai Composite Index will fluctuate between 3600 and 4200. Meanwhile, Chinese economic resurgence will not change in the short term but still faces huge pressure in downturn. Chinese government is supposed to introduce more supporting policies given that the market plays its leading role in resource allocation, hoping to stabilize and improve investment sentiment and safeguard economic resurgence achievements. (Since the stock market can reflect the operation of real economy, it is of necessity to actively promote the development of the real economy. It is essential to encourage the development of small and medium-sized promising enterprises which are important and active parts of the market economy. It is of great importance to set up a sound and healthy environment where the capital can flow into the corporate in need and can create the real wealth. This can be accomplished by some governmental policies that regulate the capital flow from financial institutions. As to the investment sentiment, it will expand if the real economy develops at a stable or accelerating speed. And the high investment sentiment will contribute to the advancement of real economt in turn. Then, there will be a good circulation).

CONFLICT OF INTEREST

The author declares no conflict of interest.

AUTHOR CONTRIBUTIONS

Yanpeng Sun does the Conceptualization, methodology, formal analysis, investigation, resources, data curation, writing - original draft, writing - review & editing, visualization.

REFERENCES

- L. E. Baum and T. Petrie, "Statistical inference for probabilistic functions of finite state Markov chains," *Annals of Mathematical Statistics*, vol. 37, no. 6, pp. 1544-1563, 1966.
- [2] L. E. Baum, T. Petrie, G. Soules, and N. Weis, "A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains," *Annals of Mathematical Statistics*, vol. 41, no. 1, pp. 164-171, 1970.
- [3] L. E. Baum, "An inequality and associated maximization technique in statistical estimation of probabilistic functions of a Markov process," *Inequalities*, vol. 3, pp. 1-8, 1972.
- [4] M. S. Ryan and G. R. Nudd, "The viterbi algorithm," in *Proc. IEEE*, 1973, vol. 61, no. 5, pp. 268-278.
- [5] S. E. Levinson, L. R. Rabiner, and M. M. Sondhi, "An introduction to the application of the theory of probabilistic functions of a Markov process to automatic speech recognitiong," *Bell Labs Technical Journal*, vol. 62, no. 4, pp. 1035-1074, 1983.
- [6] L. R. Rabiner and B. H. Juang, "An introduction to hidden Markov models," *IEEE ASSP Mag*, vol. 3, no. 1, pp. 4-16, 1986.
- [7] G. Krishnalal, S. B. Rengarajan, and K. G. Srinivasagan, "A new text mining approach based on HMM-SVM for web news classification,"

International Journal of Computer Applications, vol. 19, pp. 98-104, 2010.

- [8] M. R. Hassan and B. Nath, "Stock market forecasting using hidden Markov mode1: A new approach," in *Proc. International Conference on Intelligent Systems Design & Applications*, 2005, pp. 192-196.
- [9] M. R. Hassan, B. Nath, and M. Kirley, "A fusion model of HMM, ANN and GA for stock market forecasting," *Expert Systems with Applications*, vol. 33, no. 1, pp. 171-180, 2007.
- [10] A. Srivastava, A. Kundu, S. Sural, and A. Majumdar, "Credit card fraud detection using hidden Markov model," *IEEE Transactions on Dependable Secure Computing*, vol. 5, no. 1, pp. 37-48, 2008.
- [11] M. R. Hassan, "A combination of hidden Markov model and fuzzy model for stock market forecasting," *Neuro Computing*, vol. 72, pp. 3439-3446, 2009.
- [12] M. Caccia and B. Remillard, "Option pricing and hedging for discrete time autoregressive hidden Markov model," *Les Cahiers du GERAD*, 2017.
- [13] H. Jiang and G. F. Xu, "Dynamic prediction of stock index based on grey system and BP neural network," *Mathematics in Practice and Theory*, vol. 22, pp. 52-59, 2013.
- [14] H. Y. Shi, "Analysis and prediction of Shanghai composite index by ARIMA model based on wavelet analysis," *Mathematics in Practice and Theory*, vol. 23, pp. 59-68, 2014.
- [15] S. Song, "Exploring the Kuznets characteristic of stock price fluctuations and its application in prediction," *Journal of Applied of Statistics and Management*, vol. 2, pp. 47-56, 2014.

- [16] L. P. Cui, "Markov chain theory and application study in the field of economic management," *China Market*, vol. 6, pp. 76-83, 2016.
- [17] S. L. Fei, "Markov chains in multiple random environments," *Journal of East China Normal University*, vol. 2, pp. 62-71, 2016.
- [18] Y. Lin and J. H. Yang, "Improved hybrid neural network based on fuzzy granulation for price range prediction of stock index futures," *South China Finance*, vol. 10, pp. 9-22, 2017.
- [19] L.-Y. Wei, "A hybrid ANFIS model based on empirical mode decomposition for stock time series forecasting," *Applied Soft Computing*, vol. 5, pp. 32-41, 2016.

Copyright © 2020 by the authors. This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited (<u>CC BY 4.0</u>).



Yanpeng Sun is currently a professor. She is now in Qingdao University, China. She obtained her B.Sc. and M.Sc. degrees in 2000 and 2005, respectively. Her current research focuses mainly involve management innovation, internet finance and big data processing.