

## Prediction of Sse Composite Index Based on the Markov Model Weighted by Macroeconomy on Sliding Window

Keyan Liu

School of Business, Jiangnan University, Wuxi, 214122, China

dkfjkfg@163.com

**Keywords:** Sse composite index, Weighted markov model, Macroeconomy, Sliding window

**Abstract:** Conventional Markov model could be applied to the prediction of data which has no following effects. It has certain validity for stock index prediction but the changes of the stock index will be affected by macroeconomic fundamentals, while the traditional Markov prediction model does not take into account the macroeconomy. Therefore, this paper uses the three indicators of Industrial Value Added, M1-M2, and CPI to consider the macro fundamentals and weight the conventional Markov model, and then forecasts the closing price of SSE Composite Index from the 41st to the 48-week in 2019. The prediction results show that the new method can forecast the index better, and is of some reference value for investors to make decisions.

### 1. Introduction

As an important financial instrument, stocks has always been the focus of financial institutions and investors. Financial institutions and academia have tried to achieve effective prediction of stock prices, and some have achieved good prediction results. There are three main theoretical foundations of existing prediction methods: statistical theory, artificial intelligence and stochastic process theory. However, many forecasting methods are only the pure application of pure mathematical models, which deviate from reality and lack the theoretical basis of economics. The stock market is a “barometer” of macroeconomics. Macroeconomic changes will lead to the price fluctuations of the stock market, and the unusual fluctuation of the stock market price index is also a warning to the macroeconomic situation to a certain extent. Therefore, this paper intends to build a new weighted Markov forecast model based on macroeconomic indicators, and at the same time use rolling forecasting methods to enhance the timeliness of the model, hoping to give investors in the secondary market some reference on the stock index trend.

### 2. Literature Review

Markov model is an important prediction model of stochastic process theory. Many scholars use Markov theory to make predictions. Xia Li and Huang Zhenghong (2003) used the conventional Markov model to predict the rise and fall of stock prices, and concluded that under stable conditions, stocks rose with a 50% probability, 10% probability was flat and 40% probability of falling. This is basically consistent with the situation at that time. Lu Ruirui (2009) used K-means clustering method to divide the state of the data, and used the Markov prediction model to predict the future trend of the Shanghai Stock Index, and obtained a more satisfactory conclusion. Fei Shilong and Ren Hongguang (2016) extended the traditional Markov model and introduced a multiple Markov chain model, which was applied to stock price prediction. Compared with the Markov chain model, the accuracy of stock price prediction was improved. Yang Nan and Xing Licong (2006) used the gray Markov model that introduced the advantages of interval gray prediction and Markov chain prediction to enhance the ability to predict time series data in a shorter time and applied it to the prediction of house price index. A higher degree of fit has been achieved. Zhang Enming et al. (2007) adopted the polynomial with fluctuations instead of gray GM (1,1) to solve the problem of gray GM (1,1) 's exponential features and the amount of data required by Markov, and improved gray Markov The model uses the

improved grey Markov model to predict the Shanghai Composite Index, and the accuracy of the predicted value is better than the original model. Zhang Qiansheng proposed a Markov chain prediction model weighted based on the fuzzy state transition probability matrix of different historical periods, and gave a method of using genetic algorithms to optimize the search of the weight of the fuzzy Markov chain prediction model.

In summary, the main methods for scholars to make predictions using Markov models are as follows:

(1) Improvement and promotion of Markov chain theory itself, such as multiple Markov and hidden Markov

(2) Combine with other prediction models, such as gray Markov prediction model

(3) Combined with algorithms, such as K-means clustering Markov prediction model, fuzzy Markov prediction model

However, these prediction methods are only the direct application of existing mathematical models, which may deviate from reality and lack real economic theory. Therefore, on the basis of the influence of macroeconomic fundamentals on the stock market, this paper establishes a weighted Markov prediction model based on macroeconomic indicators.

### 3. The Markov Model

#### 3.1 The Conventional Markov Model

##### 3.1.1 The Definition of Markov Chain

There is a stochastic process  $\{X_t, t \in T\}$ . Let  $T$  be a discrete set of time, that is  $T = \{0, 1, 2, 3, \dots, m\}$ . The possible values of  $X_t$  compose the state space which is a discrete set  $I = \{1, 2, \dots, n\}$ . Then if  $\{X_t\}$  satisfies the condition:

$$P(X_{t+1} = j | X_0 = i_0, X_1 = i_1, \dots, X_{t-1} = i_{t-1}, X_t = i) = P(X_{t+1} = j | X_t = i)$$

$\{X_t\}$  would called as Markov chain.

##### 3.1.2 The Definition of Transition Probability Matrix

Let  $p_{ij}(k)$ , which is called one-step transition probability, be the probability of the value of  $X_{k+1}$  would be  $j$  when the value of  $X_k$  is  $i$  at the time  $k$ , that is

$$p_{ij}(k) = P(X_{k+1} = j | X_k = i)$$

If  $p_{ij}(k)$  is independent of  $k$ , then the chain is called homogenous Markov chain. The one-step transition probability can be denoted by  $p_{ij}$ . Then we could get one-step transition probability matrix  $P$ .

$$P = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1n} \\ p_{21} & p_{22} & \cdots & p_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ p_{n1} & p_{n2} & \cdots & p_{nn} \end{bmatrix} \quad (1)$$

##### 3.1.3 The Estimation of Transition Probability Matrix

Let  $f_{ij}$  be the transfer frequency, that is, in the existing time series, the number of times that the state  $i$  is transferred to the state  $j$  in one step. The corresponding transfer frequency matrix  $f$  is as follows.

$$f = \begin{bmatrix} f_{11} & f_{12} & \cdots & f_{1n} \\ f_{21} & f_{22} & \cdots & f_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ f_{n1} & f_{n2} & \cdots & f_{nn} \end{bmatrix} \quad (2)$$

When  $n$  is very large, we can equivalently treat frequency as a probability, so we can use frequency to estimate transition probability.

$$p_{ij} = \frac{f_{ij}}{\sum_{j=1}^n f_{ij}} \quad (3)$$

### 3.1.4 The Method of Prediction

Let  $f_{ij}$  be the transfer frequency, that is, in the existing time series, the number of times that the state  $i$  is transferred to the state  $j$  in one step. The corresponding transfer frequency matrix  $f$  is as follows.

$$f = \begin{bmatrix} f_{11} & f_{12} & \cdots & f_{1n} \\ f_{21} & f_{22} & \cdots & f_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ f_{n1} & f_{n2} & \cdots & f_{nn} \end{bmatrix} \quad (4)$$

When  $n$  is very large, we can equivalently treat frequency as a probability, so we can use frequency to estimate transition probability.

$$p_{ij} = \frac{f_{ij}}{\sum_{j=1}^n f_{ij}} \quad (5)$$

Let the initial state vector be *Initial*. If the current state of the thing to be predicted is  $s$ , then the  $s$ -th element of this vector is 1, and the other elements are 0.

$$Initial = \begin{bmatrix} 0 & 0 & \cdots & 1 & \cdots & 0 & 0 \end{bmatrix} \quad (6)$$

Then the state distribution after the *times* transition starting from the original state, is *distribution*

$$distribution = initial \cdot P^{times} \quad (7)$$

## 3.2 Markov Model Weighted by Macroeconomy on Sliding Window

### 3.2.1 The Reason for Weighting

The traditional Markov prediction model assumes that the transition of the predicted object state always follows the same random pattern. Based on this assumption, the model derives a random pattern of state transition on the basis of historical data, and uses this model and the known state to predict the probability of distribution in the future. Therefore, there are two serious problems in using traditional Markov models to predict stock indexes. Firstly, the mode of state transfer of historical data may change. The stock index is affected by many external factors, so the difference of external factors may cause different random patterns of stock indexes in different periods. However, external factors of historical data with a large time span will inevitably change significantly. Then historical data at different points in time follow different random patterns for state transfer, and it is difficult for us to obtain stable random patterns through historical data. The

second problem is that the random pattern of prediction points will not necessarily be the same thing the random pattern of historical data. Compared with historical data, the external influence factors of the prediction point may have changed significantly, so the random mode of historical data cannot accurately predict the state of the prediction point.

Aiming at the first problem, that is, the instability of the state transition mode of historical data, this paper chooses to derive the transition frequency matrix or transition probability matrix representing the random mode based on the historical data with a smaller time span. The shorter the time span of the data, the closer the external factors that affect the stock index, and the closer the random mode of stock index state transition. Therefore, the resulting random mode of state transition can better reflect the reality.

For the second problem, that is, the random pattern of prediction points is not exactly the same as the random pattern of historical data, this paper divides the historical data into data with a small time span and the same data in the array, and obtains the random pattern of the state transition of each group of data. Then, according to the closeness of the external factors of each set of data and the predicted point, the random pattern of the predicted point is obtained. For the measurement of external factors, we use macroeconomic indicators and time to express. The situation of the stock market is closely related to the macroeconomic fundamentals. Therefore, the macroeconomic fundamentals are important external factors that affect the stock index. In order to simplify, this article uses limited macroeconomic indicators to represent the macroeconomic fundamentals. In addition to the macro economy, there are many external factors that affect the stock index, but the closer the general time, the closer the other external factors of the stock index, so for simplicity, this article uses time to indicate other external factors that affect the stock index. The closer the macroeconomic indicators and time of a set of data are to the forecast point, the closer the random pattern derived from it is to the random pattern of the forecast point, so the greater the weight.

### 3.2.2 The Selection of Macroeconomic Indicators

The macroeconomic indicators selected in this paper are closely related to the stock market, which are the year-on-year growth rate of monthly industrial value-added, broad money M2, narrow money M1 and CPI. Industrial value added represents the country's overall macroeconomic operating level. The money supply reflects China's monetary policy, of which the narrow sense M1 reflects the actual purchasing power in the economy, and the broad money supply M2 also reflects the potential purchasing power. The relative growth rate of the two has a very important impact on the stock market. In general, if M1 grows faster, the consumer and terminal markets are active; if M2 grows faster, investment and the middle market are active. The consumer price index indicates inflation. In general, when inflation occurs, the renminbi depreciates, the company's assets appreciate, and the stock price becomes a rising factor for the stock market.

### 3.2.3 The Method of Weighting

This paper selects the growth rate of industrial value-added in the month as  $mac_1$ , the difference between the M1 and M2 growth rates in the month as  $mac_2$ , and CPI growth rate as indicators in the month to measure the macroeconomic situation as  $mac_3$ , and introduces the time of closing price as  $mac_4$ . Then we get macro index vector of the closing price of the SSE Composite Index:

$$Mac_i = [mac_1 \ mac_2 \ mac_3 \ mac_4] \ (i = 1, 2, \dots, N) \quad (8)$$

Supposing there are  $N$  pieces of data in total, subtract the macro index vector  $Mac_0$  of the point to be predicted from the existing macro index vector, and take the absolute value to obtain the relative position vector of the existing data relative to the point to be predicted:

$$R_i = [r_1 \ r_2 \ r_3 \ r_4] \ (i = 1, 2, \dots, N) \quad (9)$$

Then get  $N$  relative position vectors. Let the difference between the maximum value and minimum value of the element at each position in the vector be  $dif_i$  ( $i = 1, 2, 3, 4$ ), then the

dimensionless processing of each element in the vector results in a new dimensionless relative position vector:

$$Rn_i = [rn_1 \ rn_2 \ rn_3 \ rn_4] \ (i=1, 2, 3, \dots, N) \quad (10)$$

And

$$rn_j = \frac{\max r_j - r_j}{dif_j} \ (j=1, 2, 3, 4) \quad (11)$$

Find the Euclidean distance of the  $N$  vectors from 0 vectors respectively:

$$d_i = (rn_1^2 + rn_2^2 + rn_3^2 + rn_4^2)^{\frac{1}{2}} \ (i=1, 2, \dots, N) \quad (12)$$

Let  $N$  pieces of data be divided into  $M$  groups of data uniformly in time sequence, each group of data has  $m$  data, then the average distance of each group of data is

$$\bar{d}_j = d_{(j-1) \times m + 1} + d_{(j-1) \times m + 2} + \dots + d_{(j-1) \times m + m} \ (j=1, 2, \dots, M) \quad (13)$$

Finally, the weights of the data transition frequency matrix of each group are obtained:

$$w_i = \frac{\bar{d}_i}{\bar{d}_1 + \bar{d}_2 + \dots + \bar{d}_M} \quad (14)$$

### 3.2.4 The Method of Prediction

After obtaining the transition frequency matrix  $f_i$  of each group, the weighted summation is performed by using the respective weights to obtain the weighted frequency transfer matrix  $f_w$ , and then the transition probability matrix is obtained, and the state prediction is performed according to the initial state. However, if static prediction is used, that is, multiple prediction values are obtained from the same set of historical data, it will be difficult to correctly derive the random pattern near the prediction point because the difference between the time index of the historical data and the prediction point is increasing. Therefore, this paper draws on the rolling prediction method of Yan Chunning et al. (2016), continuously introduces new historical data, deletes old historical data, and obtains more accurate prediction results.

## 4. Empirical Analysis

### 4.1 Data Preprocessing

Step 1: Data selection. This article selects the weekly data of the SSE Composite Index as the research object, uses the weekly closing prices of the 24th to 40th week of 2019 as historical data, and uses the traditional Markov model to predict the weekly closing prices of the 41st to 48th week of 2019. Using data from 24 to 48 weeks, a rolling forecast is adopted, and the macroeconomic index weighted Markov forecasting model is used to predict the weekly closing price from 41 to 48 weeks.

Step 2: State division. Based on the existing weekly closing price of the Index  $q$ , this article divides the state of the index evenly according to the highest closing price  $q_{\max}$ , the lowest closing price  $q_{\min}$ , and the number of required states  $n$ . The state space is  $i = \{1, 2, 3, \dots, n\}$ . The way to judge the states is as follows:

$$\begin{cases} X_t = i, & q \in \left[ q_{\min} + (i-1) \frac{q_{\max} - q_{\min}}{n}, q_{\min} + i \frac{q_{\max} - q_{\min}}{n} \right) (i = 2, 3, \dots, n-1) \\ X_t = n, & q \in \left[ q_{\min} + (n-1) \frac{q_{\max} - q_{\min}}{n}, \infty \right) \\ X_t = 1, & q \in \left( -\infty, q_{\min} + \frac{q_{\max} - q_{\min}}{n} \right) \end{cases} \quad (15)$$

#### 4.2 The Conventional Markov Model

The weekly closing price of 24 to 40 weeks in 2019 is used as the basis for prediction, and the traditional Markov model is used to predict the weekly closing price of 41 to 48 weeks in 2019.

The maximum value of the weekly closing price of 24 to 40 weeks in 2019 is 3031.235, and the minimum value is 2774.753. After testing, the number of states selected in this article is 4, and the state division is as follows:

Table.1. the State Division

State	State 1	State 2	State 3	State 4
Interval	$(-\infty, 2838.9)$	$[2838.9, 2903.0)$	$[2903.0, 2967.1)$	$[2967.1, +\infty)$

According to the state division, we can acquire the state of each week in the table below.

Table.2. the State of the Historical Data

Week	24	25	26	27	28	29
Index	2881.974	3001.98	2978.878	3011.059	2930.546	2924.201
State	2	4	4	4	3	3
Week	30	31	32	33	34	35
Index	2944.541	2867.838	2774.753	2823.824	2897.425	2886.236
State	3	2	1	1	2	2
Week	36	37	38	39	40	
Index	2999.601	3031.235	3006.447	2932.167	2905.189	
State	3	4	4	3	3	

Therefore, we can get the transfer frequency matrix, denoted by  $f$

$$f = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 \\ 0 & 1 & 3 & 1 \\ 0 & 0 & 2 & 3 \end{bmatrix}$$

Furthermore, we get a transition probability matrix denoted by  $P$

$$P = \begin{bmatrix} 0.50 & 0.50 & 0 & 0 \\ 0.25 & 0.25 & 0.25 & 0.25 \\ 0 & 0.20 & 0.60 & 0.20 \\ 0 & 0 & 0.40 & 0.60 \end{bmatrix}$$

The initial state which is the state of the price of the 40th week is:

$$Initial = [0 \ 0 \ 1 \ 0]$$

And the distribution of probability distribution of the states in week 41 to week 48 is  $Initial \cdot P$ ,  $Initial \cdot P^2$ , ...,  $Initial \cdot P^7$  respectively. All of them are shown in the table below.

Table.3. the State of the Historical Data

Week	State			
	State 1	State 2	State 3	State 4
41	0	0.2	0.6	0.2
42	0.05	0.17	0.49	0.29
43	0.0675	0.1655	0.4525	0.3145
44	0.0751	0.1656	0.4387	0.3206
45	0.079	0.1667	0.4328	0.3215
46	0.0812	0.1677	0.43	0.3211
47	0.0825	0.1685	0.4284	0.3206
48	0.0834	0.1691	0.4274	0.3202

In order to better test the accuracy of Markov interval prediction, this paper uses the weighted average method to deal with the predicted state probability, and the weighted average of the predicted state probability and the corresponding interval state to obtain the predicted value of the future weekly closing price. We assume that the lengths of the intervals  $(-\infty, 2838.9)$  and  $[2967.1, +\infty)$  are the same as other intervals, the prediction results are shown in the table below.

Table.4. the Prediction Results

Week 41	Week 42	Week 43	Week 44	Week 45	Week 46	Week 47	Week 48
2935.05425	2936.337	2935.952	2935.362	2934.849	2934.477	2934.227	2934.341

#### 4.3 The New Markov Model Weighted by Macroeconomy on Sliding Window

The new model introduces a rolling forecast method, using historical data for a length of 17 weeks, and constantly updating the data to predict the weekly closing price from weeks 41 to 48. This article uses stock index weekly closing price data for the 24th to 40th weeks to predict the closing price for the 41st week, and divides the state into four states:  $[2774.8, 2838.9]$ ,  $[2838.9, 2903.0]$ ,  $[2903.0, 2967.1]$  and  $[2967.1, 3031.2]$ . This article divides the 17 closing price data into four groups of data, which are the closing prices of 24 to 28 weeks, 28 to 32 weeks, 32 to 36 weeks, and 36 to 40 weeks, respectively. The transition frequency matrix of each group is shown as follows.

$$f_1 = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 2 \end{bmatrix} \quad f_2 = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 2 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \quad f_3 = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \quad f_4 = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 2 \end{bmatrix}$$

Then we obtain the macro vectors of each week on the basis of monthly selected macroeconomic indicators.

$$\begin{aligned} Mac_{24} &= [6.3 \quad 4.4 \quad 5.8 \quad 1] & Mac_{25} &= [6.3 \quad 4.4 \quad 5.8 \quad 2] \\ Mac_{26} &= [6.3 \quad 4.4 \quad 5.8 \quad 3] & Mac_{27} &= [4.8 \quad 3.1 \quad 5.3 \quad 4] \\ Mac_{28} &= [4.8 \quad 3.1 \quad 5.3 \quad 5] & Mac_{29} &= [4.8 \quad 3.1 \quad 5.3 \quad 6] \\ Mac_{30} &= [4.8 \quad 3.1 \quad 5.3 \quad 7] & Mac_{31} &= [4.4 \quad 3.4 \quad 5.4 \quad 8] \\ Mac_{32} &= [4.4 \quad 3.4 \quad 5.4 \quad 9] & Mac_{33} &= [4.4 \quad 3.4 \quad 5.4 \quad 10] \\ Mac_{34} &= [4.4 \quad 3.4 \quad 5.4 \quad 11] & Mac_{35} &= [4.4 \quad 3.4 \quad 5.4 \quad 12] \\ Mac_{36} &= [5.8 \quad 3.4 \quad 5.4 \quad 13] & Mac_{37} &= [5.8 \quad 3.4 \quad 5.4 \quad 14] \\ Mac_{38} &= [5.8 \quad 3.4 \quad 5.4 \quad 15] & Mac_{39} &= [5.8 \quad 3.4 \quad 5.4 \quad 16] \end{aligned}$$

The macro vector of the 40<sup>th</sup> week is  $Mac_{40} = [5.8 \quad 3.4 \quad 5.4 \quad 17]$ . By subtracting the macro vectors of the 16 week respectively, we acquire the relative position vectors, denoted as  $R_i (i = 1, 2, \dots, 16)$ , of the time from 24th week to 40th week. By using formula(3.9), we find the

dimensionless relative position vectors  $Rn_i (i=1, 2, \dots, 16)$ . Then we find the Euclidean distance, denoted by  $d_i$ , of the 16 dimensionless relative position vectors from 0 vectors respectively. According to the formula (3.11), we get the average distance of each group of data.

$$\bar{d}_1 = 0.7562 \quad \bar{d}_2 = 1.2033 \quad \bar{d}_3 = 1.5352 \quad \bar{d}_4 = 1.9277$$

Using formula (3.12), we obtain the weight of transition frequency matrix.

$$w_1 = 0.1395 \quad w_2 = 0.2219 \quad w_3 = 0.2832 \quad w_4 = 0.3555$$

By weighted calculation, the weighted transition frequency matrix is obtained.

$$f_w = \begin{bmatrix} 0.2831 & 0.2831 & 0 & 0 \\ 0.2219 & 0.2831 & 0 & 0.4226 \\ 0 & 0.2219 & 0.7993 & 0 \\ 0 & 0 & 0.4950 & 0.9999 \end{bmatrix}$$

Furthermore, we calculate the weighted transition probability matrix as follows.

$$P = \begin{bmatrix} 0.5000 & 0.5000 & 0 & 0 \\ 0.2392 & 0.3052 & 0 & 0.4556 \\ 0 & 0.2173 & 0.7827 & 0 \\ 0 & 0 & 0.3333 & 0.6667 \end{bmatrix}$$

We get the distribution of the state of the closing price in the 41st week as follows.

$$[0 \ 0.2174 \ 0.7827 \ 0]$$

Then the prediction of the closing price of the 41 week is 2921.1.

Following the same process, this paper draws the predicted weekly closing prices for the index from weeks 41 to 48:

Table.5. the Prediction Results of the New Model

Week 41	Week 42	Week 43	Week 44	Week 45	Week 46	Week 47	Week 48
2921.1	2976.7	2942.1	2944.4	2942.9	2939.9	2904.4	2893.7

#### 4.4 The Comparison of Prediction Results by Different Methods

This article uses the traditional Markov forecasting model and the Markov rolling forecasting model weighted by macroeconomic indicators to forecast the weekly closing prices of the Shanghai Stock Exchange Composite Index in the 41st to 48th weeks of 2019. The forecast results are compared in the following table.

Table.6. the Comparison of the Two Models

Week	Actual value	Prediction of new model	Prediction of conventional model
41	2973.7	2921.1	2919
42	2938.1	2976.7	2919
43	2954.9	2942.1	2921.4
44	2958.2	2944.4	2923.8
45	2964.2	2942.9	2925.6
46	2891.3	2939.9	2926.9
47	2885.3	2904.4	2927.7
48	2872.0	2893.7	2928.2

From the table above, we can get the prediction results of the two models compared with the actual value as shown below:



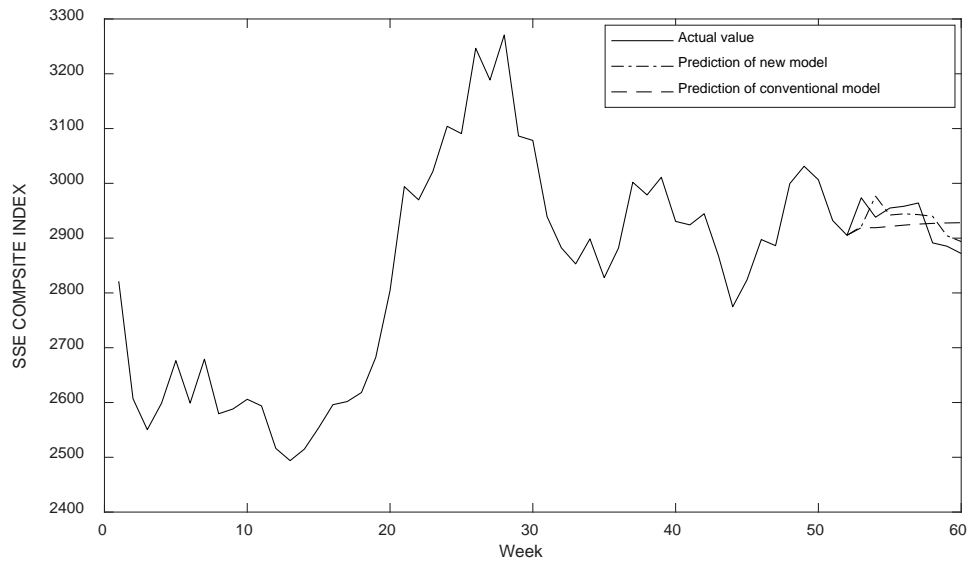


Fig.1 The Analysis of Prediction Effect

From the figure above, we can see that the curve of prediction results with the conventional Markov prediction curve is relatively stable and can not reflect the fluctuation of the stock index in the actual situation. But the fluctuation trend of the prediction curve based on the weighted Markov model using rolling prediction is very similar to the actual situation.

Compare the error between the two different results and we get the following table,

Table.7. the Comparison of the Errors of the Two Models

model	Average error	standard deviations of the two errors
New model	0.97%	0.53%
Conventional model	1.34%	0.42%

As can be seen from the table above, the average error of the conventional prediction model is 1.34%, while the average error of the new model is 0.53%, which is significantly smaller than the former. When comparing the standard deviations of the two errors, we can see that the two have similar volatility. Therefore, compared with the conventional Markov prediction model, the prediction result of the weighted Markov prediction model is more in line with the actual situation, and its change trend is close to the actual situation so its ability of prediction is better.

## 5. Conclusion

In this paper, based on the traditional Markov prediction model, the transition probability matrix is obtained by using the weight related to macroeconomic indicators, and then the new prediction model is used to predict the Shanghai Stock Exchange Index. This paper selects the data of 41th to 48th week in 2019 as the prediction object, and uses the conventional Markov model and the new model to make predictions respectively. It is found through comparison that the mean and error fluctuations of the new prediction model are obviously senior to the conventional one. And it can better reflect the trend of stock market fluctuations. Using the new model constructed in this article can provide some reference for investors to analyze the broad market index.

In terms of theoretical basis, the conventional Markov model is a pure mathematical model, lacking economic theoretical basis. The new model has a certain economic theoretical basis, that is, the macroeconomic fundamentals will affect the Chinese stock market. On this basis, this article assumes that the changes in the Shanghai Stock Exchange Index are a random process influenced by many external factors, and are based on macroeconomic fundamentals. Therefore, this article uses macroeconomic indicators for weighting, and the established model has stronger explanatory power.

Although the new model established in this paper is superior to the traditional model, there are

still many deficiencies in this model due to the complexity of the stock market itself and the limitations of my knowledge. First, the way to measure external factors is relatively simple. In addition to macroeconomic fundamentals, there are many other external factors that affect the stock market. In order to simplify the processing process, this article uses only a few economic indicators to represent the macroeconomic trends. In future research, we can measure the impact of external factors on the stock market from multiple angles. Second, the data time span required by the transfer frequency matrix is difficult to determine. If the time span is short, the transition frequency matrix uses less data, and may not reflect the real random pattern in a short time. If the time span is longer, the transfer frequency matrix can use more data, but the random pattern of the stock index may occur larger variations and the calculated transition frequency matrix may be meaningless. Third, the weighting method is relatively simple. In this paper, Euclidean distance between the forecast point macro vector and each group of historical macro vectors is calculated from large to small. And then a simple linear weighting is used to obtain the final frequency transition matrix. Therefore, the weighting method is worth continuing to study in the future.

## References

- [1] Arnesen P , Olav Kåre Malmin, Dahl E . A forward Markov model for predicting bicycle speed [J]. Transportation, 2019(7).
- [2] Tran V , Maxwell D , Fuhr N , et al. Personalised Search Time Prediction using Markov Chains[C]// Acm Sigir International Conference. ACM, 2017
- [3] Du Shiping. Theory of Hidden Markov Models and Its Applications[D].Sichuan:Sichuan University,2004.
- [4] Fei Shilong, Du Hongguang. Multiple Markov Chain Model and Its Application in Stock Market Prediction[J]. Journal of Dezhou University, 2016(4):28-31.
- [5] Guo Cunzhi, Liu Changbiao. The Method of random analysis for Forecasting the Yield Rate of Stock Investment[J]. QUANTITATIVE AND TECHNICAL ECONOMICS, 2002(02):54-57.
- [6] Lu Ruirui. The Markov Process in Price Trend Forecast Based on K-means Clustering[D], Huazhong University of Science and Technology, 2009
- [7] Wen Haibin. The application on some prediction with Markov chain model[D], Nanjing University of Posts and Telecommunications, 2012
- [8] Xia Li, Wang Yang, Li Wenhong. Improved grey Markov model and its application in stock analysis[J]. Commercial Research, 2007, 28(11):1292-1295.
- [9] Yan Chunling, Liu Ruihui, Zhang Huanghuang. The accuracy of prediction in stock market based on the improved Markov model[J]. Statistics and Decision,2016, 000(008):76-79.
- [10] Zhang Engming, Wang Yan, Li Wenhong. Improved grey Markov model and its application in stock analysis[J]. Journal of Harbin Engineering University, 2007, 28(11):1292-1295.
- [11] Yang Nan, Xing Licong. Application of Grey-Markov Model on the Prediction of Housing Price Index[J], 2006,21(5):52-55.