Predicting Stock Prices Using Markov Chain: The Stock Price Forecast Based on Shanghai Securities

Langyu Gu[1,a,*], Kerui Zeng[2,b]

[1]Xi 'an Yuandong No. 1 Middle School, 1Xi 'an, China,
[2]Chongqing No.1 International Studies School, Chongqing, China,
a. glylex@163.com
b. sylor.zengkerui@biacademy.cn
**correspondence author*

Abstract: This study investigates and predicts the stock price of Shanghai Securities. Our analysis lemma the C-K equation，n step transition to predict the stock price of Shanghai Securities. In this paper, we have put our model into different stocks in reality to test its feasibility. Finally, we envisaged the probable scope for this approach and listed some shortages of using Markov chain in predicting stock price. A great discovery in this page is that utilizing the stock's Markov property; we concluded that Shanghai Securities is martensitic. Also, we have proved the economic benefit of this numerical model.

Keywords: stoke prediction, numerical models, markov chain, finance, probability transfer

1. Introduction

The stock market is the most critical market among the capital market. It can be said that if we can predict the stock price, we can get control of the finance. The stock price has a distinct feature: how it will change has no connection with whether it has ascended or descended before, which is why it is hard to forecast. The way of predicting stock has been available for no more than one hundred years. There are many different ways to indicate this. Basically, can be divided into three directions: the forecast model based on statistical methods, machine learning predictive models, and the prediction model based on the random procedure. The Markov model is precisely one that can identify the trend and ascertain the future state of a sure thing by studying its initial state. It can appropriately predict the stock price by giving the probability of how it will change. The Markov chain mainly indicates the product market share， sales status, and expected profit. This article uses the C-K equation and n step transition probability to show the Shanghai Securities.

2. The Model

The price of a stock, also the volume of trading, can be seen as a function of time. To a specific t it only has one particular value; its state is limited. Then we can build a model as follows. Set the closing number in the day n as Xn(n=1，2，3，…+∞）; we can divide (0,+∞）into k intervals of equal distance. The interval n is [Xn-1,Xn) (n∈[1，k]).Then Xn must exist in a specific break. Let the Xn(m)=i; this means when m∈[Xn-1,Xn], the state of the stock price is i. Then let P_{ij} represent the probability of this state turning from i to j. $P^{(n)}{}_{ij}$means the likelihood of bending from i to j in n steps. [1-4]

Then we can make such a probability transfer chart:

$$P=P_{(ij)}=\begin{bmatrix} P_{11} & P_{12} & \cdots & P_{1K} \\ P_{21} & P_{22} & \dots & P_{2K} \\ \vdots & \vdots & \ddots & \vdots \\ P_{K1} & P_{K2} & \cdots & P_{KK} \end{bmatrix}$$

The chart above is called "the transition matrix". We can know that $P_{ij}>0$ since it represents the possibility—also $\sum P_{ij}=1$.The numbers 1, 2, 3, represent the three possible stock price trends.1means ascend, 2means flat, 3means descend. According to the C-K equation, $P_{(n)}=Pn$ describes the possible distribution of the stock price transfer from one state to another. So we can predict the possible state that the stock price may transfer in n days by comparing the numerical value of i.

3. Empirical Analysis

Here we select 100 sets of the latest statistics per item to ensure timeliness. Then we build a chart about its closing price. The first object we choose is the 'RESTORE':

Table 1 The stock price of BESTORE [5].

Number	1	2	3	4	5	6	7	8	9	10
Price	37.23	35.64	36.79	36.53	37.20	37.80	38.14	39.68	39.63	39.32
Number	11	12	13	14	15	16	17	18	19	20
Price	39.95	38.31	39.59	40.21	39.98	39.92	39.96	37.00	37.80	37.10
Number	21	22	23	24	25	26	27	28	29	30
Price	35.75	35.93	35.33	35.68	36.60	35.39	35.82	34.85	33.74	33.28
Number	31	32	33	34	35	36	37	38	39	40
Price	31.01	31.5	33.83	32.53	33.06	31.99	30.29	29.70	29.80	28.05
Number	41	42	43	44	45	46	47	48	49	50
Price	27.30	29.10	27.78	27.75	26.80	27.69	28.42	28.10	28.76	28.74
Number	51	52	53	54	55	56	57	58	59	60
Price	28.86	28.74	26.13	26.41	26.40	27.18	28.03	28.30	26.80	24.36
Number	61	62	63	64	65	66	67	68	69	70
Price	25.06	25.95	26.37	28.27	28.97	29.19	28.42	27.95	28.23	27.59
Number	71	72	73	74	75	76	77	78	79	80
Price	27.19	28.12	28.26	28.61	29.70	31.00	30.29	27.54	27.69	26.29
Number	81	82	83	84	85	86	87	88	89	90
Price	23.9	23.55	24.38	24.71	24.74	24.97	25.16	24.77	24.39	24.71
Number	91	92	93	94	95	96	97	98	99	100
Price	24.77	24.60	24.43	24.25	24.31	23.45	22.98	23.15	24.06	23.77
Number	101	102								
Price	24.34	24.21								

Further, let's subtract two by two to know its state in different periods:

DOI: 10.54254/2754-1169/20/20230163

Table 2 The price difference data of the STORE.

number	difference	number	difference	number	difference	number	difference	number	difference
1	0.26	23	-0.09	45	-0.08	67	0.02	89	-0.04
2	-0.01	24	0.37	46	0.03	68	-0.05	90	0
3	0.01	25	-0.07	47	0.03	69	-0.01	91	0.02
4	-0.04	26	0.13	48	-0.04	70	0.02	92	0.04
5	-0.28	27	0.06	49	-0.01	71	0.03	93	0.09
6	0.05	28	0.05	50	0	72	-0.03	94	0.05
7	0.02	29	0.08	51	0	73	0.02	95	0.04
8	-0.14	30	-0.01	52	0.01	74	0.03	96	0.04
9	-0.14	31	0.01	53	0.05	75	-0.01	97	-0.03
10	0.13	32	0.01	54	-0.01	76	-0.01	98	-0.07
11	-0.14	33	-0.09	55	0.03	77	-0.03	99	-0.03
12	0	34	-0.04	56	0.01	78	0.06	100	0.01
13	0.12	35	-0.07	57	-0.03	79	-0.01	101	0.01
14	-0.23	36	-0.06	58	-0.04	80	0	102	0
15	0.46	37	-0.04	59	0.05	81	0.03		
16	0.02	38	0.34	60	-0.08	82	-0.05		
17	0.31	39	0.05	61	0.06	83	-0.03		
18	0.27	40	0.02	62	-0.02	84	0		
19	-0.04	41	-0.04	63	0.01	85	0.06		
20	0.01	42	0.03	64	-0.02	86	-0.04		
21	0.37	43	0.01	65	-0.03	87	0.02		
22	-0.19	44	-0.02	66	-0.08	88	-0.01		

Since two adjacent data can't be identical, we need an interval to define "flat." Here we choose 0.5 as a limit. If the difference exceeds 0.5, it would be classified as "ascend." Likely, if it is fewer than -0.5, it would be "descend," of course.

After that, we can obtain another table:

Table 3 The change situation of the data.

	State	State	State	State	State	State	State	State	State	State
State	1	3	2	3	3	2	3	2	2	3
State	1	3	3	2	2	2	1	2	1	1
State	2	1	2	3	1	2	1	1	2	1
State	2	3	1	3	1	1	1	2	1	1
State	3	1	2	1	3	3	2	3	2	2
State	2	1	2	2	3	3	2	1	1	3
State	3	2	3	3	2	1	2	2	1	2
State	3	2	2	3	3	1	1	2	1	1

DOI: 10.54254/2754-1169/20/20230163

Table 3: (continued).

State	2	3	2	2	2	2	2	2	2	2
State	2	2	2	2	1	2	2	3	2	3
State	2									

Finally, we can calculate the transfer probability of each part. Among them, P_{11} means the probability of state 1 to state 1, ascend to ascend, is0.08. The same way, P_{22}=0.2, P_{33}=0.07, P_{12}=0.13, P_{21}=0.14, P_{13}=0.06, P_{31}=0.06, P_{23}=0.14, P_{32}=0.12

Then we can get a state transition matrix:

$$\mathrm{P} = \begin{bmatrix} P_{11} & P_{12} & P_{13} \\ P_{21} & P_{22} & P_{23} \\ P_{31} & P_{32} & P_{33} \end{bmatrix} = \begin{bmatrix} 0.8 & 0.13 & 0.06 \\ 0.14 & 0.2 & 0.14 \\ 0.06 & 0.12 & 0.07 \end{bmatrix}$$

Since the latest day we chose is the last statistic, we can set it as an initial vector. The final day lies in the state of "1", then we put the vector $\pi_{(0)}$= (1, 0, 0). Then we can calculate the probability of the next day in the future:

$$\pi_{(1)} = \pi_{(0)}\mathrm{P} = (1, 0, 0) \begin{bmatrix} 0.8 & 0.13 & 0.06 \\ 0.14 & 0.2 & 0.14 \\ 0.06 & 0.12 & 0.07 \end{bmatrix} = (0.8, 0.13, 0.06).$$

We can see its state is more probably lies ini_1, which means its closing price is more likely to ascend the next day. We can even analyze $\pi_{(2)}$, $\pi_{(3)}$, etc.

But, such calculation is too complex and lacks efficiency. Also, according to a large sum of analysis, the state probability will tend toward a stable price, and this regular price has nothing to do with the initial cost. Based on the sound condition of the Markov chain:

$$\begin{cases} \pi P = \pi \\ \sum_{i=1}^{n} x_i = 1, \pi = (x_1, x_2, x_3) \end{cases}$$

Among them, x_1, x_2, x_3 represent three states (states 1, 2, and 3). Then let's juggle the equations we have:

$$(x_1, x_2, x_3) = \begin{bmatrix} 0.8 & 0.13 & 0.06 \\ 0.14 & 0.2 & 0.14 \\ 0.06 & 0.12 & 0.07 \end{bmatrix} (x_1, x_2, x_3)$$

$$x_1 + x_2 + x_3 = 1$$

$$x_1 \approx 0.5773, x_2 \approx 0.3695, \ x_3 \approx 0.0532$$

That means it has a probability of 57.73% ascending, 36.95% fat, 5.32% descending.

The result is similar to the one we had before, and they corroborate each other, showing a feasible way to make a prediction. The next object is the China Union:

DOI: 10.54254/2754-1169/20/20230163

Table 4: The stock price of the China Union [5].

number	name	price	number	name	price	number	name	price	number	name	price	number	name	price
1	China Union	4.8	23	China Union	3.99	45	China Union	3.36	67	China Union	3.56	89	China Union	3.55
2	China Union	4.54	24	China Union	4.08	46	China Union	3.44	68	China Union	3.54	90	China Union	3.59
3	China Union	4.53	25	China Union	3.71	47	China Union	3.41	69	China Union	3.59	91	China Union	3.59
4	China Union	4.52	26	China Union	3.78	48	China Union	3.38	70	China Union	3.6	92	China Union	3.57
5	China Union	4.56	27	China Union	3.65	49	China Union	3.42	71	China Union	3.58	93	China Union	3.53
6	China Union	4.84	28	China Union	3.59	50	China Union	3.43	72	China Union	3.55	94	China Union	3.44
7	China Union	4.79	29	China Union	3.54	51	China Union	3.43	73	China Union	3.58	95	China Union	3.39
8	China Union	4.77	30	China Union	3.46	52	China Union	3.43	74	China Union	3.56	96	China Union	3.35
9	China Union	4.94	31	China Union	3.47	53	China Union	3.42	75	China Union	3.53	97	China Union	3.31
10	China Union	5.08	32	China Union	3.46	54	China Union	3.37	76	China Union	3.54	98	China Union	3.34
11	China Union	4.95	33	China Union	3.45	55	China Union	3.38	77	China Union	3.55	99	China Union	3.41
12	China Union	5.09	34	China Union	3.54	56	China Union	3.35	78	China Union	3.58	100	China Union	3.44
13	China Union	5.09	35	China Union	3.58	57	China Union	3.34	79	China Union	3.52	101	China Union	3.43
14	China Union	4.97	36	China Union	3.65	58	China Union	3.37	80	China Union	3.53	102	China Union	3.42
15	China Union	5.2	37	China Union	3.71	59	China Union	3.41	81	China Union	3.53			
16	China Union	4.74	38	China Union	3.75	60	China Union	3.36	82	China Union	3.5			
17	China Union	4.72	39	China Union	3.41	61	China Union	3.48	83	China Union	3.55			
18	China Union	4.41	40	China Union	3.36	62	China Union	3.42	84	China Union	3.58			
19	China Union	4.14	41	China Union	3.34	63	China Union	3.44	85	China Union	3.58			
20	China Union	4.18	42	China Union	3.38	64	China Union	3.43	86	China Union	3.52			

Table 4: (continued).

21	China Union	4.17	43	China Union	3.35	65	China Union	3.45	87	China Union	3.56		
22	China Union	3.8	44	China Union	3.34	66	China Union	3.48	88	China Union	3.54		

As you can see, the data is here. So we can likewise do the same. Unlike the former, the data here is quite close, so we elected a smaller interval: 0.3 as a limit constant. Here is an ultimate table:

Table 5 The change situation of the data2.

	State	State	State	State	State	State	State	State	State	State
State	1	2	2	3	3	1	2	3	3	1
State	3	2	1	3	1	2	1	1	3	2
State	1	3	3	1	3	1	1	1	1	2
State	2	2	3	3	3	3	3	1	1	2
State	3	2	2	2	3	2	2	3	2	2
State	2	2	1	2	2	2	2	3	1	3
State	1	2	2	2	2	3	2	3	2	2
State	2	2	2	2	2	2	2	1	2	2
State	2	3	2	2	1	3	2	2	3	2
State	2	1	1	1	1	1	2	3	2	2
State	2									

Then comes the probability; after a series of processes, the chart tells us that P_{11}=0.09, P_{12}=0.07, P_{22}=0.3, P_{33}=0.07, P_{23}=0.11, P_{32}=0.12, P_{13}=0.08, P_{31}0.07, P_{21}=0.09.
Here the initial vector $\pi_{(0)}$= (1, 0, 0), and

$$\mathrm{P} = \begin{bmatrix} P_{11} & P_{12} & P_{13} \\ P_{21} & P_{22} & P_{23} \\ P_{31} & P_{32} & P_{33} \end{bmatrix} = \begin{bmatrix} 0.09 & 0.07 & 0.08 \\ 0.09 & 0.3 & 0.11 \\ 0.07 & 0.12 & 0.07 \end{bmatrix}$$

$$(x_1, x_2, x_3) = \begin{bmatrix} 0.09 & 0.07 & 0.08 \\ 0.09 & 0.3 & 0.11 \\ 0.07 & 0.12 & 0.07 \end{bmatrix} (x_1, x_2, x_3)$$

$$x_1 + x_2 + x_3 = 1$$

Figure out x_1, x_2, x_3 in the same way: $x_1 = 0.4340$ $x_2 = 0.1199$ x_3=0.4461. That means it is more probably descend. So we'd better buy some tomorrow. It fell from 4.8 (12.25) to 4.78 the next day (12.26).

4. Results

This way, we can estimate the probability of falling into three states (ascend, flat, and descend). After such analysis, we can obtain more profit in the stock market. It's convenient and with real significance. Through this model, we can see how the closing price will change in an obvious way. We can sell them all tomorrow if it is more likely to ascend the next day. We can even predict the day after tomorrow to earn more…If it is flat, we can just take a break. If it is going to descend, stockholders

DOI: 10.54254/2754-1169/20/20230163

may sell them as soon as possible to prevent a more significant loss. Despite the randomness of the stock market, as long as we know where it will move, we can always be the winners.

5. Conclusion

Generally, the Markov chain could adequately predict the stock price. With the help of this model, we can forecast the change situation just by inserting a set of data. This work emphasized building a stock predicting model and briefly introduced how it works. After doing this, we have proved its practicability through plenty of current statistics. Shortly, we may try to promote the way to predict stock prices in different ways--not only the Markov chain.

References

[1] *Hsu, Y. T. , Liu, M. C. , Yeh, J. , & Hung, H. F. . (2009). Forecasting the turning time of stock market based on markov–fourier grey model. Expert Systems with Applications An International Journal, 36(4), 8597-8603.*

[2] *Liu, S. , Zhang, X. , Wang, Y. , & Feng, G.. (2020). Recurrent convolutional neural kernel model for stock price movement prediction. PLOS ONE, 15..*

[3] *Kaur, S. , & Mangat, V. . (2012). Improved accuracy of pso and de using normalization: an application to stock price prediction. International Journal of Advanced Computer Science & Applications, 3(9), 115-120..*

[4] *Zhang, J. , Wang, Y. , Zhao, Y. , & Fang, H. (2021). Multi-scale flood prediction based on gm (1,2)-fuzzy weighted markov and wavelet analysis. Journal of Water and Climate Change(5)..*

[5] *Besag, J. (1994). Discussion: markov chains for exploring posterior distributions. The Annals of Statistics(4).*