A Markov ChainApproach in Predicting Stock Market Behaviour

Audrey Kana Kartono
Cita Hati West Campus, Surabaya, East Java, Indonesia
ssw_audrey.kartono@bchati.sch.id
APCYS 2022

Introduction

Markov Chain is a stochastic process containing random variables transitioning from one state to another depending only on certain assumptions and definite probabilistic rules—having the Markov property. Markov property is described as a memoryless process as it is only dependent on the present state and the randomness of transitioning to the next state. Simply put, the Markov Chain tends to predict the future state with the current state as the only requirement.

The aim of this research is to apply the Markov Chain in predicting the stock market. Specifically, it sought to answer this question: How effective, in terms of percentage error, is the Markov Chain as a predicting tool of stock market behavior?

Methodology

In this study, the stock market data used is Amazon (AMZN) from Yahoo! Finance. This research uses 1289 days of data from December 31, 2014 to December 30, 2019.

To find the price change, the daily closing price data is extracted using the Python programming language. The looping commands are used to look for price changes by subtracting the price of day X_t from the previous day. After the data is processed, the total data will be reduced to 1288 because there is no known difference between the oldest price data and the previous price. This process output' will be positive, negative, and zero values.

The probability distribution then can be calculated using the following formula:

$$\pi_1 = \frac{\Sigma p}{99} \qquad \pi_2 = \frac{\Sigma n}{99} \qquad \pi_3 = \frac{\Sigma z}{99}$$

(equation 1)

where
π = value in probability distribution
p = frequency of positive numbers
n = frequency of negative numbers
z = frequency of zero numbers

The next step is to find the component of the transition matrix which shows the probability of the price changes. This is done by calculating the events or frequency where the price changes from rising to rising, rising to falling, rising to the same, fall to rise, and so on. Then the number is processed as shown in the table of equations below. After the data is processed, the total data willbe reduced to 1287.

Probability	Rise	Fall	Same
Rising	$\frac{RR}{RR+RF+RS}$	$\frac{RF}{RR+RF+RS}$	$\frac{RS}{RR+RF+RS}$
Falling	$\frac{FR}{FR+FF+FS}$	$\frac{FF}{FR+FF+FS}$	$\frac{FS}{FR+FF+FS}$
Same	$\frac{SR}{SR+SF+SS}$	$\frac{SF}{SR+SF+SS}$	$\frac{SS}{SR+SF+SS}$

Table 2. Equation 2 in getting for probability component

where:
RR = frequency of rising to rising
FR = frequency of falling to rising
SR = frequency of same to rising

With these equations, the component probability P is found, $P = (P_{ij})$ where *i* is the current state and *j* is the future state. For example, P_{RF} means the probability from *R*, the current state to *F*, thefuture state.

After the first probability, π^T and the transition matrix is found, future probability distribution, *X*, is found. If X_t is the same in any two future states, then it should be treated as the stationary distribution.

The multiplication formula between π^T and transition matrix used is shown in equation 3.

$$X_1 = [\pi_a \quad \pi_b \quad \pi_c]$$
$$\pi_a = (\pi_1 * P_{RR}) + (\pi_2 * P_{FR}) + (\pi_3 * P_{SR})$$
$$\pi_b = (\pi_1 * P_{RF}) + (\pi_2 * P_{FF}) + (\pi_3 * P_{SF})$$
$$\pi_c = (\pi_1 * P_{RS}) + (\pi_2 * P_{FS}) + (\pi_3 * P_{SS})$$

(equation 3)

where:
X_1 = future state of X_0
π_a = future state of π_1
π_b = future state of π_2
π_c = future state of π_3

Result

There were 699 days where the price rose, 579 days where the price fell, and 10 days where theprice didn't change. With this, it can be seen that the probability distribution of X_0 or π^T is:

$$\pi^T = [0.5489 \quad 0.4503 \quad 0.0008]$$

The transition matrix found is as follows:

	Rise	Fall	Same
Rise	0.5508	0.4392	0.0100
Fall	0.5329	0.4619	0.0052
Same	0.6000	0.4000	0.000

Then, π^T is multiplied with the transition matrix to find the probability distribution of X_t.

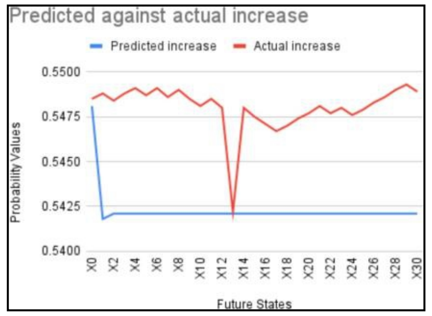
$$X_1 = [0.5428 \quad 0.4494 \quad 0.0078]$$
$$X_2 = [0.5432 \quad 0.4491 \quad 0.0077]$$
$$X_3 = [0.5431 \quad 0.4491 \quad 0.0077]$$
$$X_4 = [0.5430 \quad 0.4490 \quad 0.0077]$$
$$X_5 = [0.5430 \quad 0.4490 \quad 0.0077]$$

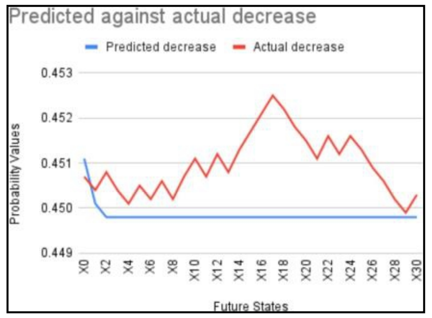
X_4 and X_5 have the same distribution probability, this means that [0.5430 0.4490 0.0077] is the stationary distribution.

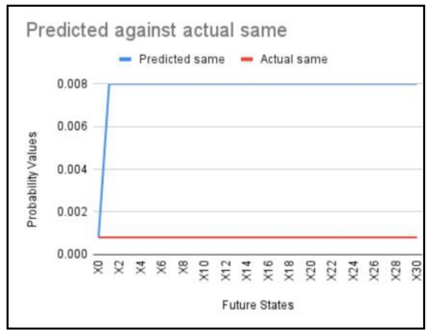
The line graphs below compares the predicted and actual probability distribution.

Conclusion

Predicting the behavior of the stock market is difficult because of its volatility. Therefore, the Markov Chain stochastic model is used. Markov Chain is applied to closing price data from five years to improve accuracy. Markov Chain is an effective predictor of volatile stock behavior (the rise, fall, and same) in a one month period, however, it is unable to follow the changing price of the stock market. It can be concluded that the Markov Chain is an effective predictor of stock behavior in a short-term period. This study can be continued by using other stock market data or with a larger amount of data.







References

Chen, J. (2021). Stock Market. https://www.investopedia.com/terms/s/stockmarket.asp

Hayes, A. (2021). Closing Price. https://www.investopedia.com/terms/c/closingprice.asp Fewster,

Rachel. (2014). Course Notes Stats 325 Stochastic Processes. Auckland.

Suhartono, D. (n.d.). Markov Chain. https://socs.binus.ac.id/2013/06/30/markov-chain/