Investing Decision Model Based on Apriori Algorithm

Hanqi Chen

School of Economics and Management, Beijing Jiaotong University, Beijing, China

20711043@bjtu.edu.cn

Abstract. Quantitative investments can be effective in reducing losses. It plays an important role in the pre-diction of the market. Due to this, more and more investors are paying attention to it. Trade Strategy Model and other models are used to predict price and make the decision. In Data Mining (DM), quantile values, advance rate, and decline rate are utilized to determine relative likelihood. Furthermore, the Apriori algorithm is used to calculate the number of continuous rises and falls that occur with a probability greater than 90%. To determine the amount of dollars to be purchased or sold in the shipment space, we establish a Position-Management Model (PMM). Finally, we introduce the Decision-Making Model (DMM) as a tool for trading decisions. We then plug $1000 into the model to arrive at $1958.37 in 2021. Based on the comparison of the predicted price and the real price from 2019-9-1 to 2019-10-31, we found a slight deviation between the predicted and the real price. As part of our verification process, we used the BackTesting method and selected the indicator "annual return rate," whose test showed an average of 17.687%, an excellent result. Afterward, we provide a summary of the strategy process for assisting its quantitative trading investment. To conclude, our models are distinguished by their precise values.

Keywords: Quantitative investment; Data Mining; Apriori; Decision-Making Model.

1. Introduction

There have been many trading strategies developed in the past. The best-stop problem approach is a popular approach to trading. The practice of "buy low and sell high" [1] trading on a simulated currency pair proves profitable. However, risks like market plunges exist if a positive movement is predicted. It is not advisable to pursue the model with the highest return blindly. Our goal is to select an investment portfolio that is optimal. The term asset portfolio refers to the allocation of investment funds to assets such as gold and bitcoin by investors. Investments in various assets account for a certain percentage of investors' total investments. An investor's overall return from the purchase is as high as possible while minimizing risk.

2. Literature Review

As early as 1952, Markowitz proposed a model of optimal portfolios based on mean-variance. By using the ratio between the returns of each security as a variable, he used quadratic programming to determine the optimal portfolio boundaries for investors based on the statistical characteristics of the returns of each deposit. A portfolio's optimal allocation is determined by the investor's utility function (the trade-off between return and risk). In this case, the portfolio problem can be viewed as a one-stage process. This theory was later extended to multi-stage portfolio problems using dynamic programming methods [2].

In recent years, more and more asset management institutions have issued buy and sell orders using quantitative methods and computer programming to achieve stable returns [3]. This quantitative investment has also been recognized by more investors.

3. Decision-Making Model

3.1. Data Mining

We analyzed bitcoin and gold's Advance-Decline Rate over the past five years,

$$\text{Advance-Decline Rate} = \frac{(V_2 - V_1)}{V_1} \times 100\% \quad (1)$$

Where: V_2 = today's price, V_1 = previous days price

Bitcoin and gold Advance-Decline Rates were calculated using MATLAB. After sorting these rates, we calculate the quantile value of gold and bitcoin.

Table. 1 Bitcoin Advance-Decline Rate

Decline Rate					Advance Rate				
90 quantile	75 quantile	Median	25 quantile	10 quantile	90 quantile	75 quantile	Median	25 quantile	10 quantile
-0.06872	-0.0327	-0.0126	-0.0035	-0.00020	0.1794	0.03812	0.01874	0.0075	0.00170

Table. 2 Gold Advance-Decline Rate

Decline Rate					Advance Rate				
90 quantile	75 quantile	Median	25 quantile	10 quantile	90 quantile	75 quantile	Median	25 quantile	10 quantile
-0.01398	-0.00183	-0.0043	-0.0080	-0.00005	0.01499	0.00901	0.00423	0.0035	0.00062

The quantile value indicates the relative likelihood of the prediction. If the daily Advance-Decline Rate exceeds 90 quantiles. This is a small probability event, and we should carefully consider adding or underweighting the position. When the predicted price exceeds the median, we consider today's advance-decline rate high. A decisive addition should be made if the rate of predicted price falls between the 10th percentile and the median.

3.2. Apriori Algorithm

We then introduce the Apriori algorithm for analyzing data association rules.

We constructed a frequent subset from the number of rises and falls and counted bitcoin and gold successive rises/falls twice, three times, etc. Calculating the frequency of occurrence of each frequent subset shown in the table using MATLAB.

Table. 3 Bitcoin frequent subset

Bitcoin						
Number of successive rises	1	2	3	4	5	6
Frequency	992	510	290	100	80	30
Number of successive fall	1	2	3	4	5	6
Frequency	833	351	170	120	70	35

Table. 4 Gold frequent subset

Gold				
Number of successive rises	1	2	3	4
Frequency	652	338	175	100
Number of successive fall	1	2	3	4
Frequency	591	296	150	70

Find out that these items have accumulated frequent advancements and declines of more than 90%

$$m_2 + m_3 + ... + m_k \geq m_1 \times 90\% \quad (2)$$

$$n_2 + n_3 + ... + n_k \geq n_1 \times 90\% \quad (3)$$

where: m_k: the frequency of successive k times rises

n_k: the frequency of successive k times falls

After calculating, we can get the following information. A rise or fall in bitcoin that is successively greater than four times or greater than six times is considered to be of low probability. If gold rises or falls successively more than four times, it is considered a low probability event. Furthermore, if these cases occur, we should maintain our position

Next, we analyze bitcoin's advance-decline trend. Calculate the advance and decline rate based on the data up to today's date using T=4. To sort the advance rates, select all of them and use their quantile values. Using T=6 as the interval, calculate the advance and decline rates from today's data. To sort the decline rates, select all the decline rates and use the quantile values.

Table. 5 Bitcoin successive rise 4 day's advance rate

Bitcoin 4 days advance rate				
10 quantile	25 quantile	Median	75 quantile	90 quantile
0.013067252	0.02482123	0.049800474	0.087257128	0.149668075

Table. 6 Bitcoin successive fall 6 day's decline rate

Bitcoin 6 days decline rate				
10 quantile	25 quantile	Median	75 quantile	90 quantile
-0.009041829	-0.079075773	-0.037972775	-0.010051985	-0.164092819

The position should be underweighted for events with an advance rate of approximately 0.149668 in four days. With a six-day decline rate of approximately -0.1640928, we deemed that it was within the acceptable range and added positions.

Gold's advance-decline trend can be evaluated using the same method.

Table. 7 Gold successive rise/fall 4 day's advance/decline rate

Gold 4 days advance rate				
10 quartile	25 quartile	Median	75 quartile	90 quartile
0.00182721	0.004729694	0.010061333	0.018081563	0.0273088

Gold 4 days decline rate				
10 quartile	25 quartile	Median	75 quartile	90 quartile
-0.001791068	-0.004901	-0.009741284	-0.017076884	-0.029923399

Those events we believe are possible with an advance rate of approximately 0.0273088 in four days should be underweighted. With a six-day decline rate of approximately -0.164, we considered it within the acceptable range and added positions.

3.3. Adding and Underweight Position

We consider bitcoin and gold have the same profitability.

The Initial position:

$$\text{Bitcoin: } P_B = \frac{P_{total}}{u_1} \times \frac{1}{2} \quad P_B = P_{total} \times \frac{1}{2} \tag{4}$$

$$\text{Gold: } P_{Gold} = \frac{P_{total}}{u_2} \times \frac{1}{2} \quad P_{Gmax} = P_{total} \times \frac{1}{2} \tag{5}$$

Where: P_B: Initial position of bitcoin; P_{total}:Total assets held; u_1: Maximum number of successive advance/declines of bitcoin; P_{Bmax}: full position of bitcoin; P_{Gold}: initial position of gold; u_2: Maximum number of successive advance/declines of gold; P_{Gmax}: full position of gold

For the accuracy of our calculation, we deduct the transaction cost when adding positions. As soon as we predicted today's price would decline, we added positions in the expectation that it would rise

someday. Ideally, the decline rate would be equal to the consecutive decline $\frac{u_1}{u_2}$ time decline u_2 rate 90 quantile, and with every decline of $\frac{u_1}{u_2}$ time, we would add the position; if the price rose the next day, we would earn profit.

When adding positions, we deduct the transaction costs in order to ensure the accuracy of our calculation.

$$P_2 \times (1-\alpha\%) \times M_3 + P_1 \times \left(\frac{M_2}{u_1}\right) \times M_3 = 0 \tag{6}$$

$$P_3 \times (1-\alpha\%) \times M_3 + P_2 \times \left(\frac{M_2}{u_1}\right) \times M_3 + P_1 \times \left(\frac{M_2}{u_1}\right)^2 \times M_3 = 0 \tag{7}$$

$$P_3 \times (1-\alpha\%) \times M_3 + P_{n-1} \times \left(\frac{M_2}{u_1}\right) \times M_3 + \cdots + P_1 \times \left(\frac{M_2}{u_1}\right)^{n-1} \times M_3 = 0 \tag{8}$$

$$P_{\frac{B}{G}} + P_1 + \cdots + P_n \leq P_{max} \tag{9}$$

Where: P_1:the amount of first time add position; P_2: the amount of second time add position; P_3: the amount of third time add position; P_n: the amount of n time add position; M_3:daily advance rate's median; M_2:successive decline$\frac{u_1}{u_2}$ time 90 quantile; $P_{\frac{B}{G}}$: initial position of bitcoin or gold; P_{max}: full position of bitcoin or gold; $\alpha\%$: transaction cost

From the above expression, we can get bitcoin and gold's Adding position management model

Table. 8 Adding Position Management

Bitcoin Adding Position Managemet						
Number of successive falls	1	2	3	4	5	6
Adding Amount	0.04458	0.4674	1.6048	9.62928	57.775	330.654

Gold Adding Position Managemet				
Number of successive falls	1	2	3	4
Adding Amount	2.4038	12.01923	60.0961	300.4807

Every time we predicted that today's price would rise, we thought that there would be a possibility of decline someday in the future, so we overweight the positions during every rise.

In the ideal state, successive rise $\frac{u_1}{u_2}$ time, the advance rate is just equal to the successive $\frac{u_1}{u_2}$ time advance rate 90 quantile, and in every rise of time, we overweight the position; if the next day price falls, we can offset the loss.

For the accuracy of our calculation, we deduct the transaction cost when adding positions.

$$Q_2 \times (1-\alpha\%) \times N_3 + Q_1 \times \left(\frac{N_2}{u_1}\right) \times N_3 = 0 \tag{10}$$

$$Q_3 \times (1-\alpha\%) \times N_3 + Q_2 \times \left(\frac{N_2}{u_1}\right) \times N_3 + Q_1 \times \left(\frac{N_2}{u_1}\right)^2 \times N_3 = 0 \tag{11}$$

$$Q_3 \times (1-\alpha\%) \times N_3 + Q_{n-1} \times \left(\frac{N_2}{u_1}\right) \times N_3 + \cdots + Q_1 \times \left(\frac{N_2}{u_1}\right)^{n-1} \times N_3 = 0 \tag{12}$$

$$Q_1 + \cdots + Q_n \leq P_{\frac{B}{G}} \tag{13}$$

where: Q_1: the amount of first-time overweight position; Q_2: the amount of second time overweight position; Q_3: the amount of third time overweight position; Q_N: the amount of n time overweight position; N_3: daily decline rate's median; N_2: successive rise u_1/u_2 time 90 quantile; $P_{\frac{B}{G}}$: initial position of bitcoin or gold; $\alpha\%$: transaction cost

from the above expression, we can get bitcoin and gold's Adding position management model

Table. 9 Underweight Position Management

Bitcoin Underweight Position Management				
Number of Successive falls	1	2	3	4
Adding Amount	0.06	0.64	7.03	77.28

Gold Underweight Position Management				
Number of Successive falls	1	2	3	4
Adding Amount	0.08538	0.9392	10.3312	113.644

Calculate the parameters of a linear regression model based on the above table. Using this model, we can determine the position management model for bitcoin and gold.

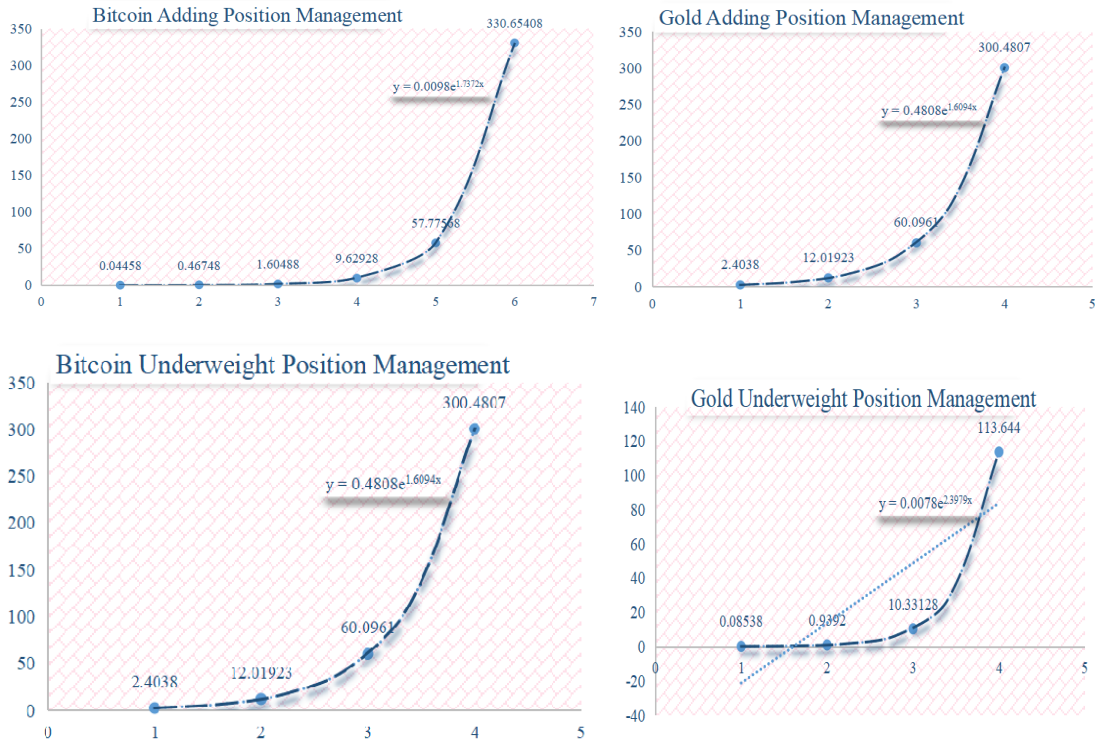

Figure. 1 Position Management Model

Numerical comparison of the sustainability indexes of the two systems
Bitcoin Position Management:

Adding Position:$y = 0.0098e^{1.7372x}$ Underweight Position:$y = 0.003e^{2.3979x}$ (14)

Gold Position Management:

Adding Position: $y = 0.4808e^{1.609x}$ Underweight Position:$y = 0.0078e^{2.3979x}$ (15)

Where: x: $\frac{daily\ advance-decline\ rate}{(\frac{M_2}{u_2})}$; y: $\frac{the\ amount\ of\ adding}{underweight\ position}$

3.4. Decision-Making Model

From above information, we build this decision-making tree, which can be used for trading.

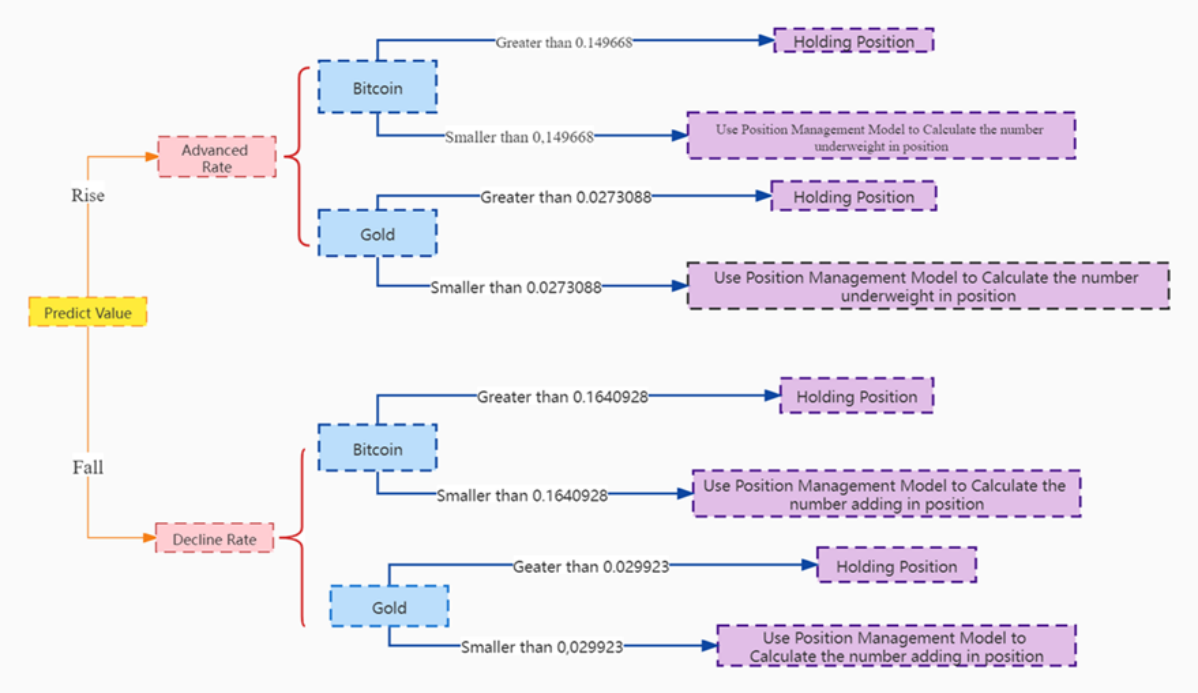

Figure. 2 Decision-Making Model

4. Effectiveness of Decision-Making Model

In efficient Decision-making, we calculate the annual return rate of the investment. And then use it to evaluate our model.

We use the BackTesting method, employed our Decision-Making model from 9/1/2019 to 10/31/2019, and then analyze its daily return rate. In addition, we transfer the daily return into the Annual return rate.

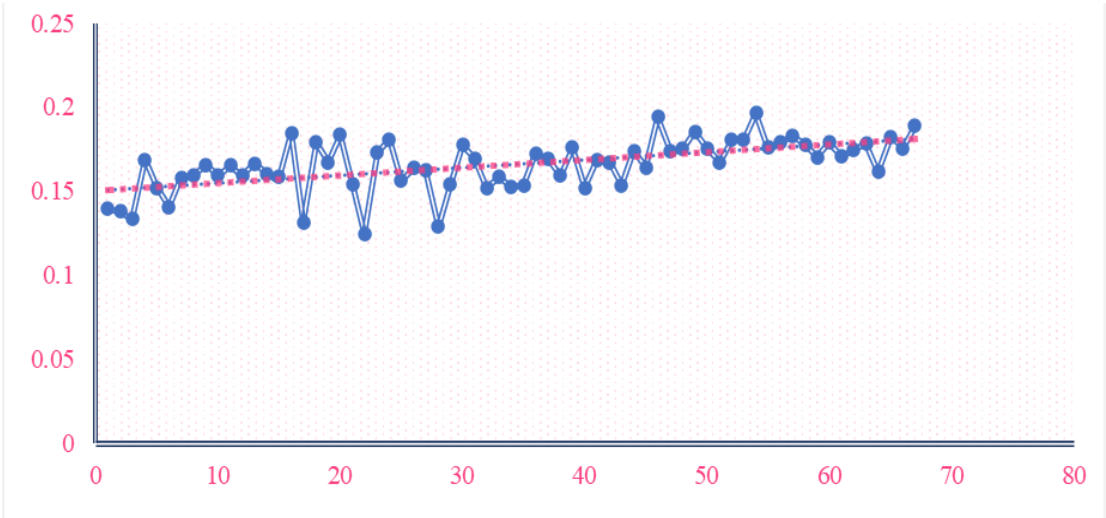

Figure. 3 Annual Return Rate Curve

From the figure, we can see that with the continuous of day, annual return rate curve shows a steady upward trend and has an average of 17.687%. In general, the valuation of the annual return of investment prediction model is shown in the following table.

Table. 10 Valuation of the Annual Return of Investment

8%-10%	10%-15%	15%-20%	20%
Pass	Good	Excellent	Distinction

Since the Annual returns rate from our investment strategy falls in the excellent range, it can prove that our model provides the best strategy.

References

[1] Cohen, Gil. “Forecasting Bitcoin Trends Using Algorithmic Learning Systems.” Entropy, vol. 22, no. 8, 2020, p. 838., https://doi.org/10.3390/e22080838.

[2] A Chunxiang, Liu Sanyang. Research on Portfolio Investment Model and Its Algorithm,2005

[3] Richard A. DeFusco; Dennis W. McLeavey; Jerald E. Pinto; David E. Runkle, Quantitative Methods for Investment Analysis, 264215, 201

[4] Guo Sihan. Research on bitcoin price prediction and trading strategy based on improved recurrent neural network [D]. Shanghai Normal University, 2021

[5] Robert J. Hilderman, Colin L. Carter, Howard J. Hamilton, And Nick Cercone, “Mining Association Rules from Market Basket Data Using Share Measures and characterized Itemsets”