Research on value prediction and investment model based on machine learning

Yutong Dai[1, *], Hongyuan Zhang[2]

[1]School of Management, Jinan University, Guangzhou, China, 510632

[2]College of Cyber Security, Jinan University, Guangzhou, China, 510632

*Corresponding author: daiyutong0311@163.com

Keywords: SVR prediction model, optimal portfolio decision model, multi-objective programming

Abstract: In investment markets, high returns can be accompanied by high risks. Making more efficient use of limited information and making the best portfolio strategy is an eternal topic in the investment market. In this paper, by using the daily closing price information up to that day, a model is established to predict that day's price and make the best portfolio of gold and bitcoin according to the balance of returns and risks. First, we use the SVR prediction model to obtain information from price flow data to predict the closing price of the investment day. On this basis, we use the optimal portfolio decision model based on multi-objective programming to select the optimal portfolio by weighted distance under the careful consideration of investment return, potential maximum investment loss, and risk coefficient. According to our model, we made a total of 613 buy and sell adjustments between 2016 and 2021. Moreover, our investment strategies contribute to the appreciation of the initial principal's value from $1000 to $14,357.607 after five years. Second, from the aspects of the forecast model's accuracy and anti-interference ability, return-risk balanced viewpoint, we prove that our model provides the optimal strategy. Moreover, a disturbance verification test is made to test the strategy, and some detailed examples prove that our model of decision-making can still have a considerable income within 20% of the interference.

1. Introduction

With the development and maturity of the financial market, more and more investors participate in the financial market to invest some assets [1]. Expecting to achieve the maximum return, they make investment decisions in various assets with different risks and returns. Investors can conduct statistical analysis on the historical rate of return of assets to guide their investment behavior in the future [2].

Quantitative investment is a trading strategy based on machine learning algorithms and using a specific mathematical model to obtain excess returns from many historical data. It is increasingly used in various types of investment [3]. Gold has always been an investment tool with high value and is not limited to any country or trade market. Bitcoin is an emerging P2P form of digital currency [4], which is not issued by specific currency institutions but generated through a large number of calculations according to specific algorithms [5]. In this paper, we use the SVR prediction model to obtain information from price flow data to predict the closing price of the investment day. And then, we use the optimal portfolio decision model based on multi-objective programming to select the optimal portfolio. By building a machine learning value prediction model, an investment model to maximize the total return on assets.

2. Establishment of Models

2.1. Establishment of SVR model

SVR (Support Vector Machine Regression) is an important application of SVM (Support Vector Machine) [6]. SVM is a kind of binary classification model. Its basic model is defined as a linear classifier with the largest interval in the feature space. Its learning strategy maximizes the interval, which can be transformed into a convex quadratic programming problem. SVR creates a "spacer band" on both sides of the linear function. The loss is not calculated for all samples falling into the spacer band [7]. Only those outside the interval band are included in the loss function. Then the model is optimized by minimizing the width and total loss of the spacer band. SVR expects all sample points to fall within the "isolation zone," while SVM expects all sample points to fall outside the "isolation zone." SVM maximizes the "distance" of the nearest sample point to the hyperplane [8]. The SVR is intended to minimize the "distance" to the most distant sample point in the hyperplane.

In this paper, we expect to use the SVR model to predict the price of gold and bitcoin on that day. Find a hyperplane closest to these data based on past daily prices to predict the price of gold and bitcoin that day. We build a regression relationship f to find the relationship between the number of days and the price.

$$y = f(x) = \sum_{i=1}^{n} w_i x_i + b \tag{1}$$

Where x is the number of days from the beginning of the first transaction to the end of the day, y is the price of the day.

The relaxation factors ξ^+ and ξ^- are introduced to represent the exceeding and not exceeding the penalty interval ε respectively, and the problem of SVR is transformed into the problem of minimizing the risk function R in the presence of errors:

$$y = f(x) = \sum_{i=1}^{n} w_i x_i + b \tag{2}$$

$$f(x) - y^i \leqslant \varepsilon + \xi_i^- \;\; \text{s.t.}\, \xi^+, \xi^- \geqslant 0 \tag{3}$$

$$y^i - f(x_i) \geqslant \varepsilon + \xi_i^+,\; i = 1,2,3,\cdots \tag{4}$$

The problem is further transformed into finding solutions α_i^+ and α_i^- of the Lagrange function:

$$y = f(x, \alpha_i^+, \alpha_i^-) = \sum_{i=1}^{l} (\alpha_i^+ - \alpha_i^-) x_i \cdot x + b \tag{5}$$

The kernel mapping method is introduced, and the transformation function is used to map variable x to high-dimensional nonlinear space, and the nonlinear regression expression is obtained:

$$y = f(x, \alpha_i^+, \alpha_i^-) = \sum_{i=1}^{l} (\alpha_i^+ - \alpha_i^-) K(x_i, x) + b \tag{6}$$

The selection of kernel function determines the accuracy of the SVR model. Polynomial is selected as kernel function in this paper:

$$k(x_i, x_j) = \left(x_i^{\mathrm{T}} x_j\right)^d \quad d \geq 1 \tag{7}$$

Where d is the degree of the polynomial.

In this paper, the risk function is used to measure the quality of model prediction in the average sense, and a loss function is used to measure the degree of prediction error. The squared loss function is used here, which can be expressed as:

$$L(y, f(x)) = (y - f(x))^2 \tag{8}$$

2.2. Optimal portfolio decision model based on multi-objective programming

Investors choose a portfolio for the benefit and the maximum possible potential loss the portfolio faces. Sometimes, investors choose to invest in assets with a greater potential loss for profit, but sometimes, they give up investing in assets with a greater expected return.

We take expected to return as one of our investment objectives and take the potential maximum possible loss of investment and investment risk into our decision planning. Therefore, we establish a multi-objective programming model to make daily optimal asset strategy decisions.

First, investors tend to invest in portfolios with higher returns, so we set the first objective function based on investment returns:

$$f_1 = \left((g_1 - g_0)/g_0 - a_1\right) \times g_0(G + x_1) + \left((b_1 - b_0)/b_0 - a_2\right) \times b_0(b + x_2) \tag{9}$$

Where, f_1 represents the expected investment returns after the decision, G represents the number of current gold holdings, B represents the number of current bitcoin holdings, x_1 represents the expected amount of buying or selling gold, x_2 represents the expected amount of buying or selling bitcoin, g_0 represents yesterday's closing price per ounce of gold, b_0 represents yesterday's price per bitcoin, g_1 represents predicting price per ounce of gold on the day, b_1 represents the predicting price per bitcoin on the day, a_1 represents the transaction cost of gold, a_2 represents the transaction cost of bitcoin.

As investors are loss-averse, they tend to consider the maximum possible loss they face when investing. If it exceeds their expectations, they will not continue to invest. Therefore, we take the potential maximum possible loss function of the portfolio as our second objective function:

$$f_2 = q_1 g_0 (G + x_1) + q_2 b_0 (B + x_2) \tag{10}$$

$$P(\Delta R > VaR) = 1 - \alpha \tag{11}$$

Based on the experience of previous literature, we set the confidence level α as 95%, and then we can get the equation:

$$P(\Delta R > VaR) = 5\% \tag{12}$$

Arrange the rising and falling extent of gold and bitcoin in descending order. The data in the top 5% of gold is q_1, and the data in the top 5% of bitcoin is q_2. These two values represent the biggest risk decline. ΔR is the declining extent in the one-day holding period.

When other conditions are equal, investors prefer a less risky portfolio. Investors want to invest high return and low-risk assets. So we use risk as to the third objective function of portfolio selection. Since we only consider buying and selling gold and bitcoin in the market, the financial market setting is relatively simple. There are no risk-free assets to invest in, so it is challenging to construct a securities market line. We do not account for systematic risk of the portfolio but unsystematic risk as to its risk.

$$y_1 = \frac{G+x_1}{(G+x_1+B+x_2)} \tag{13}$$

$$y_2 = \frac{B+x_2}{(G+x_1+B+x_2)} \tag{14}$$

$$f_3 = \sqrt{y_1^2 w_1^2 + y_2^2 w_2^2 + 2 y_1 y_2 cov} \tag{15}$$

Where, w_1 represents the variance of gold price changes over past known time units, w_2 represents the variance of bitcoin price changes over past known time ranges, y_1 represents the proportion of gold in the portfolio, y_2 represents the proportion of bitcoin in the portfolio, cov represents the covariance of gold price change and bitcoin price change per unit in the past time.

Considering the actual trading situation, we think the minimum trading unit for gold is 1 ounce, and bitcoin is 0.01. Meanwhile, due to the lack of data in the first ten days, we decided not to conduct transactions in gold and bitcoin in the first ten days. Since gold can only be bought and sold on trading days, while bitcoin can be bought and sold every day, we separate trading and non-trading days to solve problems.

At the same time, we give the following constraints:

$$C - g_0 x_1 - b_0 x_2 \geqslant 0 \tag{16}$$

$$G + x_1 \geqslant 0 \tag{17}$$

$$B + x_2 \geqslant 0 \tag{18}$$

Where C represents current cash holdings.

Moreover, when gold is not traded：

$$x_2 = 0 \tag{19}$$

According to the above three objective functions, we get multiple non-inferior solutions. Then, select the best point of f_1, f_2 and f_3 to form a nonexistent optimal solution, which is (f_{1best} ,f_{2best} ,f_{3best}). Finally, find the weighted distance between multiple non-inferior solutions (f_1', f_2', f_3') and the optimal solution:

$$d = \sqrt{a(f_1' - f_{1best})^2 + b(f_2' - f_{2best})^2 + c(f_3' - f_{3best})} \quad (20)$$

The point that minimizes d is our optimal solution. The reasonable values of A, B, and C were obtained through our experiments.

3. Analysis of Results

This paper uses a price forecasting model and an investment strategy model to invest in a portfolio of gold and bitcoin from September 11, 2016, to September 10, 2021. A total of 613 investment strategy adjustments have been made. Moreover, we used our model and strategy to use the initial $1000 investment worth $14357.607 on 9/10/2021.

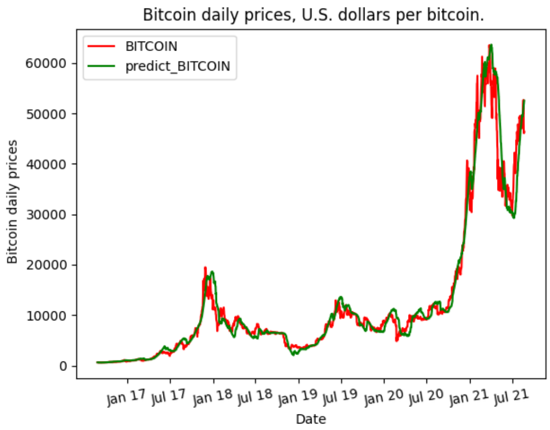

Figure 1 Daily predicting prices and actual prices for bitcoin

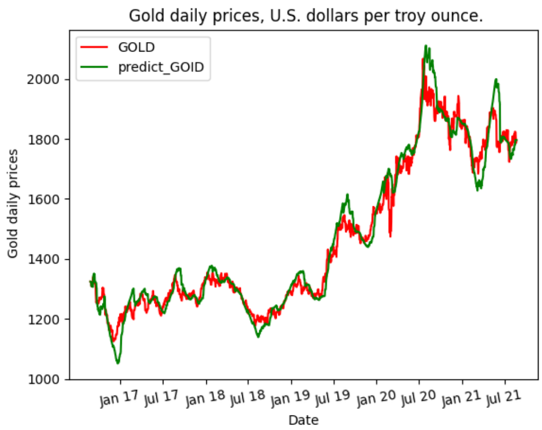

Figure 2 Daily predicting prices and actual prices for gold

The green line shows the price predicted by our SVR model for the day, and the red line shows the actual price. From the perspective of objectives, our model comprehensively considers three factors: the highest return, the lowest risk, and the maximum potential loss of the investment portfolio. Therefore, the strategy obtained by measuring the Pareto frontier points can balance returns and risks. In this way, the investment portfolio has more objective returns and has a stronger risk-bearing capacity.

As shown in Figures 1 and 2, our model fits well. The deviation between predicted and actual prices is small, and Our prediction model has high accuracy. And then we make [-10%,10%] and [-20%,20%] disturbance on actual gold and bitcoin daily price, and then make predictions.

As shown in Figures 3 and 4, after the two disturbances, our model does achieve the unity of return and risk under certain conditions.

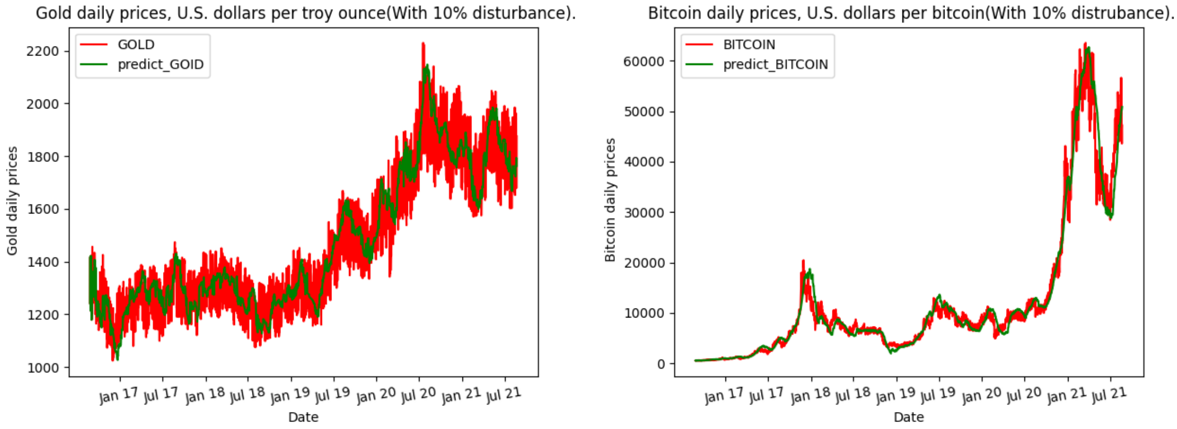

Figure 3 Price forecast with 10% disturbance

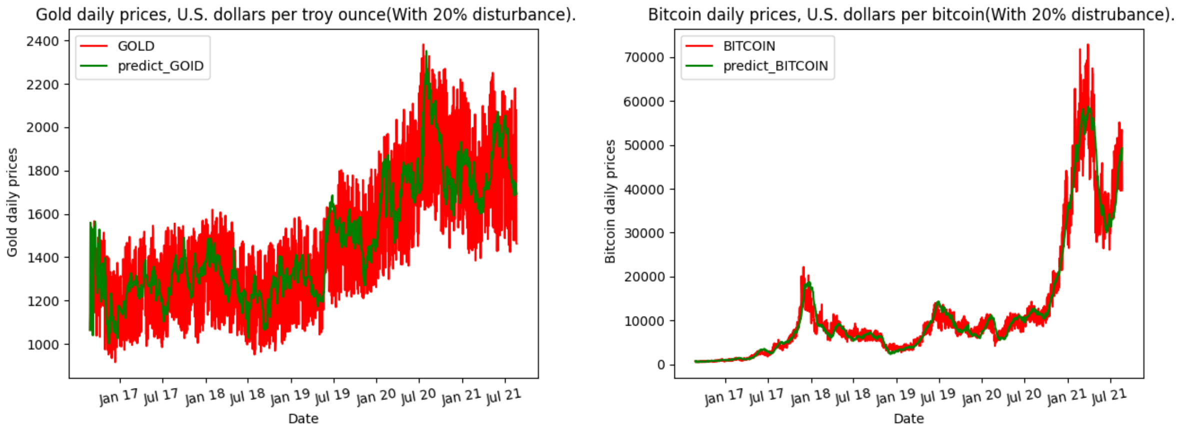

Figure 4 Price forecast with 20% disturbance

We use the same strategy to invest in the new two different price movements. The total value of our strategy on September 10, 2021, was 13666.1 and 5462.3, respectively. It can be seen that our strategy still shows good returns in dealing with the disturbance within 10% and can also achieve considerable returns in dealing with the disturbance within 20%.

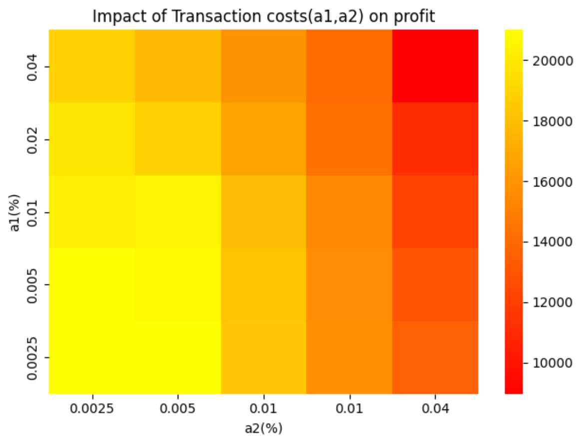

Figure 5 The impact of transaction costs on earnings

In order to test the impact of transaction costs a_1 and a_2 on our strategy, we adjusted the values of a_1 and a_2 Moreover, the strategy for the portfolio investment of gold and bitcoin is re-selected.

According to the different profits obtained in the end and combined with the specific execution process, we analyzed the sensitivity of policy pairs to transaction costs.

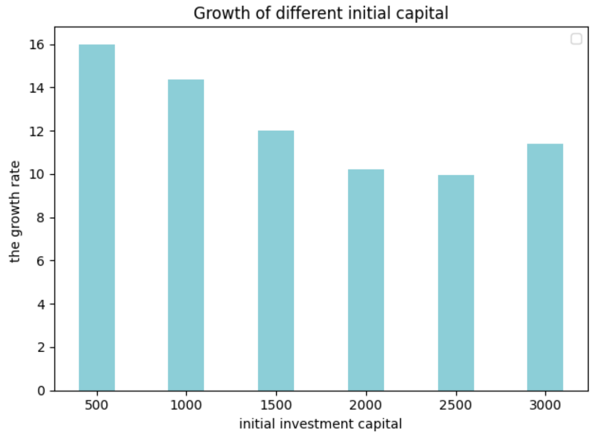

Figure 6 Growth of different initial capital

As shown in Figure 5, it can be seen that under the same cost of $1000, the final return of each investment with the different transaction costs is stable between $10,000 and $20,000. It also demonstrates that our model has strong robustness.

In order to test the sensitivity of our model to initial investment capital C, we select the initial capital in the set {500,1500,2000,2500,3000}, under the same other conditions and obtain the holding value after five years by using our model. Then calculate the ratio of the assets held five years later to their initial capital. The results are shown in Figure 6.

The results show that our model shows a good appreciation ability under different initial capital conditions. For gold and Bitcoin under the overall positive growth trend, the growth rate can reach about 10 times. It can be seen that our model is sensitive to initial capital C.

4. Conclusion

The model in this paper does not simply buy and sell according to the predicted value blindly but takes transaction costs and risks into comprehensive consideration. The initial $1000 investment worth $14357.607 is final using our model and strategy. When the price fluctuates greatly, it can also better fit the trend of price development. Our model comprehensively considers the investment return, potential investment loss, and risk and tests the model with disturb experiment. We can see that our prediction model has a specific anti-disturbance ability through the disturbing experiment. The model results are still good and can make the best portfolio decision. The model has great versatility. In addition to the prediction of gold and bitcoin, the model is also applicable to strategies of other volatile assets.

References

[1] David C. Rode, Paul S. Fischbeck Regulated equity returns: A puzzle[J].Energy Policy,2019,133(C)

[2] Yezhou Sha, Song Weijia Can Bitcoin hedge Belt and Road equity markets? [J].Finance Research Letters,2021

[3] Qiujun Ma, LI Wei. Analysis of returns and risks of Structured financial products in Chinese banks [J]. Comparison of Economic and Social Systems,2011(06):189-194.

[4] Chunfeng Wang, Haihui Wan, Zhang Wei. Financial Market Risk Measurement Model -- VaR[J]. Journal of Systems Engineering,2000(01):67-75+85.

[5] Calice Giovanni, Lin Ming. Tsung Exploring risk premium factors for country equity returns[J]. Journal of Empirical Finance,2021(prepublish)

[6] Zhongsheng Qian, Qingyuan Yu, Song Tao, Yimin Zhu, Zhu Jie, Zhao Chang. Test Case Generation and Reuse Based on Support Vector Machine Regression Model [J]. Acta Electronica Sinica, 201,49(07):1386-1391.

[7] Haibiao Tu, Yilei Yan, Karui Liu, Wenjie Li, Yuwei Fang, Yuting Hou. An efficient Regression compensation Algorithm for Support Vector Machines [J]. Information Technology and Informatization,2021(11):36-42.

[8] Zheng W T. VAR method and its application in financial risk management [J]. International Journal of Finance,1997(09):58-62. (in Chinese)