Information Systems and Economics (2022)
Clausius Scientific Press, Canada
DOI: 10.23977/infse.2022.030207
ISSN 2523-6407 Vol. 3 Num. 2

Strategies to Simulated Trading: Based on Sharpe Ratio and Random Forest

Xiaoyu Su, Jingyao Huang

School of Economics, Jinan University, Guangzhou, Guangdong, 511443, China

Keywords: Sharpe Ratio, Random Forest, Principal Component Analysis, Dynamic Programming

Abstract: Investors often pursue two objectives in the process of frequent trading of some assets with high price volatility: maximizing investment return and minimizing investment risk. In this paper, we use principal component analysis to capture the up and down signals that the market implies to investors from various aspects, while establishing a random forest model to realize the prediction of market prices, using the Sharpe ratio to compare the return with the risk, and finally getting the quantitative buying signal indicator; we also use the knowledge of statistics and finance to analyze the influencing factors of trading decisions. The study of this problem can provide investors with guiding advice to help them better gain returns and avoid risks.

1. Introduction

1.1 Problem Background

Since the Covid-19 in 2020, objectively the economic has been depression and the state has stimulated the economy by investing large amounts of money, leading to great volatility in the financial markets; subjectively many people's stable source of livelihood has changed and more people choose to trade in the financial markets to gain extra income. As a result, more investors want to trade frequently to obtain excess returns and avoid the risks brought by the epidemic, and this question will be studied with this strategy as the objective.

2. Bullish and Bearish Model

2.1 Evaluation Indicators

In this paper, we comprehensively consider the trends affecting the price of gold or bitcoin, so as to objectively evaluate the direction of the market. Based on the four-year daily closing prices given in the question, the influential factors are finally divided into three primary indicators and seven secondary indicators, namely RSI, MOM, BIAS, MTS, PSY, AP, and IA, and the data are organized by the definition of these indicators. These indicators reflect the magnitude of price fluctuations, investors' psychological expectations, and the subsequent direction of prices, etc.

The current rise in the price (the rate of increase) of financial instruments (gold/bitcoin) affects investors' consumption expectations and also future price trends, and often financial instruments that rise more are viewed favorably and keep prices up through positive externalities, as shown in Figure 1.

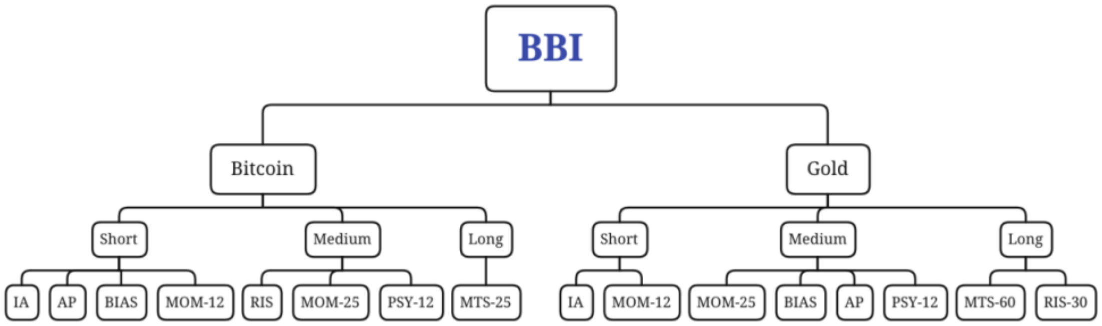


Figure 1: Indicator relationship

2.2 The Establishment of Model

This yields the index BBI for assessing price quotations.

$$BBI_{bitcoin} = 0.39 * F_1 + 0.17 * F_2 + 0.12 * F_3 + 0.1 * F4$$

$$BBI_{gold} = 0.38 * F_1 + 0.17 * F_2 + 0.12 * F_3 + 0.1 * F4$$

The average value of BBI is taken as the threshold to determine whether the day is bullish or bearish based on BBI.

$$\begin{cases} if\ \ BBI_{gold,i} > \overline{BBI_{gold}},\ i \text{ is bullish} \\ f\ \ BBI_{gold,i} \leq \overline{BBI_{gold}},\ i \text{ is bearish} \end{cases}$$

$$\begin{cases} if\ \ BBI_{bitcoin,i} > \overline{BBI_{bitcoin}},\ i \text{ is bullish} \\ f\ \ BBI_{bitcoin,i} \leq \overline{BBI_{bitcoin}},\ i \text{ is bearish} \end{cases}$$

Considering that the indicator obtained from the principal component analysis method may fluctuate repeatedly around the threshold and interfere with the trading strategy; in order to make the indicator more coherent and reflect the market over a period of time, we use the voting method correction to determine the timing of the bullish and bearish judgments. The voting method is an integrated learning model that follows the principle of minority rule. By integrating multiple models to reduce the variance, the prediction result is the sum of the prediction results of multiple other regression models. The judgment criteria are.

$$define\ list[0,0,\cdots\cdots,0]$$

$$\begin{cases} if\ \ \ i \text{ is bullish, then } list[i-1]+1, list[i-2]+1, \cdots\cdots list[i-\Delta t]+1 \\ if\ \ \text{is bearish, then } list[i-1]-1, list[i-2]-1, \cdots\cdots list[i-\Delta t]-1 \end{cases}$$

Ultimately the positive or negative result of the vote determines whether the day is bullish or bearish.

2.3 The Solution of Model 1

The results of the assessment are shown in the following chart, as shown in Figure 2.

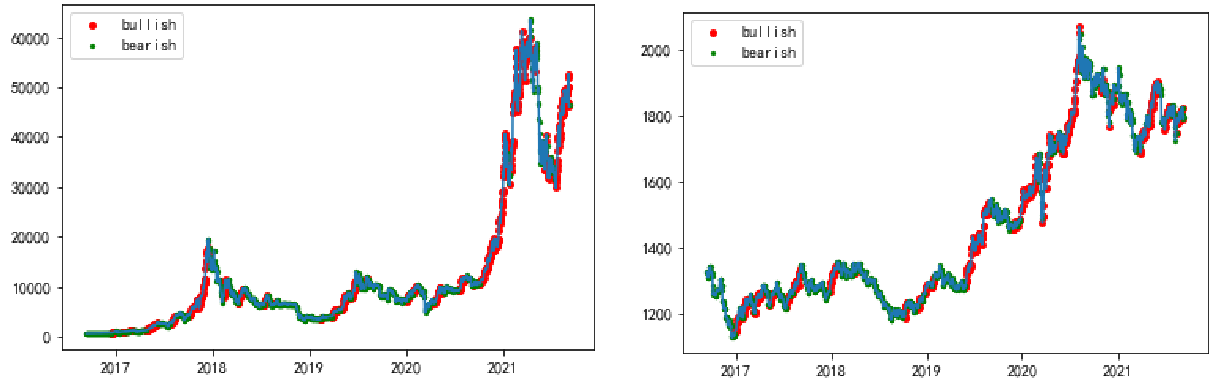


Figure 2: Bullish Bearish Indicator Trend Chart

As seen from the above graph, when the price of gold and bitcoin rises or falls, the corresponding time period can be roughly identified by BBI after the voting method, although occasionally some rising parts are not correctly identified, but overall the accuracy rate is high, so it shows that the BBI indicator created by this door has high credibility and feasibility, and is suitable for assessing the market situation, as shown in Figure 3.

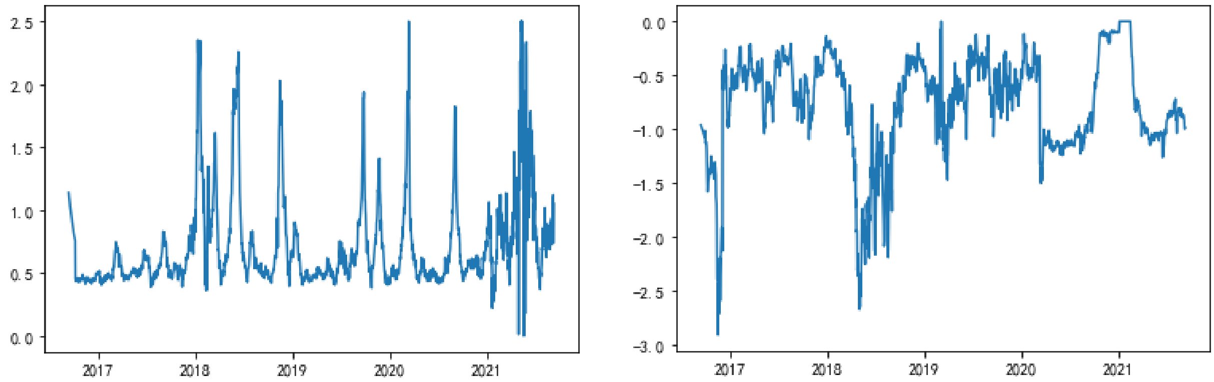


Figure 3: Final Buy Score

2.4 Trading Strategies

According to the above model, we established the rules of the simulated trading strategy:

1) The starting date of the buy signal indicator for bitcoin is October 6, while the buy signal indicator for gold is only available on 12.2. Therefore, we assume that: $1,000 between September 10 and October 6 is not used to invest, and between 10.7 and 12.1 is used to judge gold only based on the buy signal score of gold buying and selling with the amount, and on and after December 2, 2016 judging how much to buy and sell gold and bitcoin at the same time by the buying scoring signals of both currencies.

2) Buy Limit = Current Cash Limit*Buy Score*(1-Commission)/Current Price Sell amount = share held * (1 - score + sell criteria)

With the above results, the simulations simulate the returns obtained by applying the strategies we obtained, yielding the maximum return values based on the bullish and bearish judgment indicators and risk indicators.

3. Risk Model

There are many factors that affect the risk of trading. In this question, we choose the combined long-short judgment indicator and the deviation rate from the previous question to estimate the risk. Because the size of the long-short judgment indicator represents the future price trend and market sentiment, and the deviation rate, as the deviation value of the price, can also reflect the riskiness of the trade, so we customize the weight of 0.6 for long gold and 0.4 for deviation rate to build a linear risk model.

$$RI = 0.6 * BBI + 0.4 * BIAS$$

4. Random Forest Model

We choose the random forest method to predict the future direction of the stock price. Random forest algorithm can still maintain a certain accuracy in the case of uneven sample data distribution, high latitude and missing some features, and improve the accuracy under the condition that the operation amount does not increase significantly. It also shows a great advantage in processing sample data, with good noise resistance and not easy to get into overfitting. The mtry and ntree in random forest parameters can be used to coordinate the balance of classification accuracy with diversity, OOB misrate can be used for unbiased estimation of error in random forest, and variable importance estimation is able to calculate the importance of each feature variable to the classification results. Therefore, this paper tests various important parameters in random forests in empirical studies to improve the prediction accuracy [2], and the specific steps are shown in Figure 4.

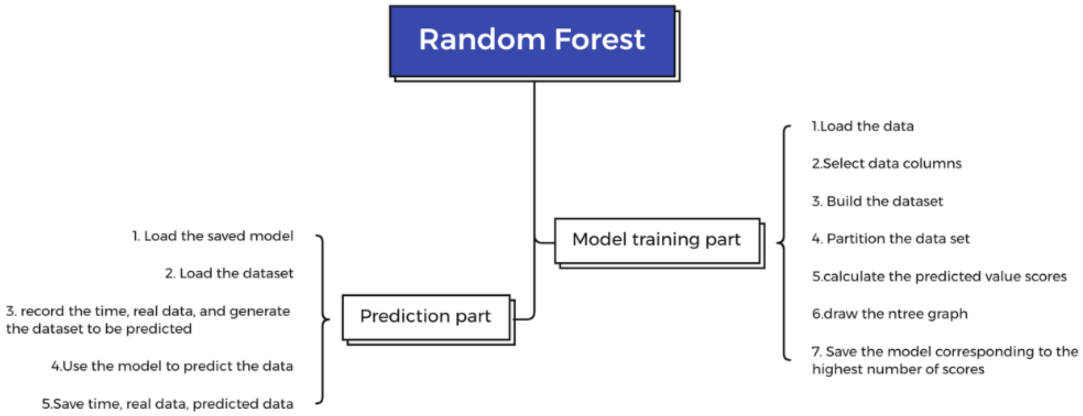


Figure 4: Random Forest Parameters

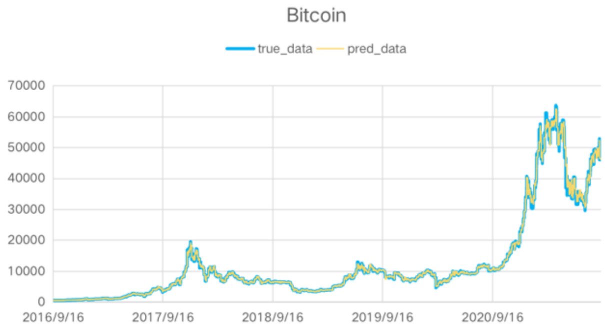
First processing data, design unknown parameters x, y, cycle 1 x (x is 15 dimensions) for 15 data, corresponding to data in article 16 is y, otherwise end; randomly using this logic x, y data, 70% data as training set, 30% as test set, training random forest prediction model.The time step was set to 5 and the number of random forest forests to 510, and the model was trained using the number of trees from 10 to 800, interval 20.

Then, the predictive value scores were calculated and the ntree plots were plotted.Adjust the picture, set the figure size width 8 height 4, per inch point 120, and call the plot function, set the parameter value, set the figure scale, call the xlabel and ylabel functions, output the model, select the number of trees to 150, output save the model, save the model corresponding to the number of the highest score in the ntree figure, and complete the program.

The prediction section is consistent with the above model training section, first loading the previously stored model and then sandwiched into the dataset, recording the time and real data in the dataset, and generating the dataset to be predicted.New tables were generated using model prediction data and saving time, real data and predicted data to be used.

Comparison of yield:

To further verify the effectiveness of the random forest model in the price trend prediction, guide the stock trading with the model prediction results to generate signals, and compare the price prediction with the trading strategy of pure technical indicators, we can find that the predicted value is very close to the real value, and the model training set and the prediction set results are shown in the Figure 5.



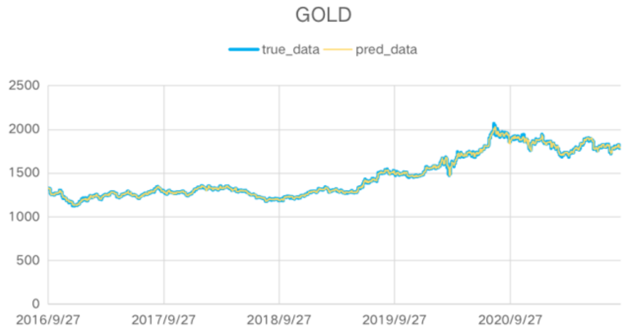


Figure 5: Model training set and test set

According to the forecast results, although the gold price fluctuated but the whole rose steadily, bitcoin peaked in 2021, and the prediction results were close to the real value, which further proves the feasibility and excellent performance of the model in predicting the price trend, which can provide reference for investors.

5. Sensitivity Analysis

The third question considers the effect of transaction costs on the sensitivity of the strategy, that is, the effect of commissions on the final total return. Taking the spacing between the two to be 0.01, gold iterates from 0.01 to 0.11 and bitcoin iterates from 0.1 to 0.21. Since there are 10 commission possibilities for gold and 20 commission possibilities for bitcoin, there are 200 possibilities for total assets to move with the indicator, and the sensitivity decreases as the possibilities increase, as shown in Figure 6.

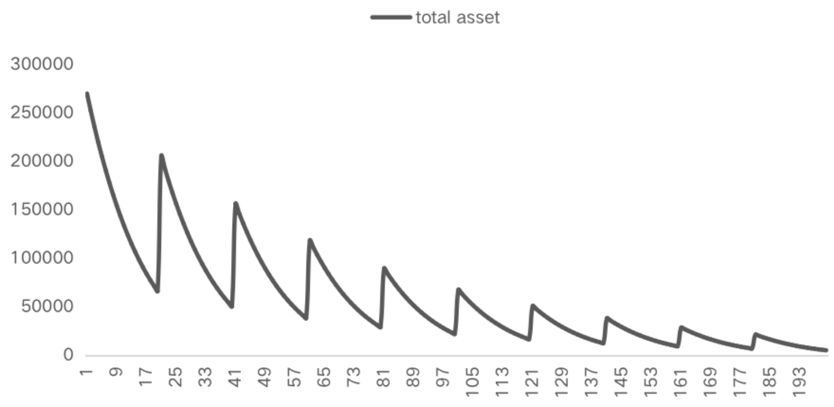


Figure 6: Sensitivity of commission to strategy

Continuing with the dynamic programming model, unlike the previous question, it is not just the zero-one indicator of whether or not to trade that affects the change in transaction costs, but also the change in commissions; also, whether or not to buy gold or bitcoin depends on the buy score, taking the same indicators as above, buying when the gold buy score is greater than 0.58 and selling when it is less than 0.3; buying when the bitcoin buy score is greater than 0.71 and selling when it is less than 0.56 Sell. In the form of a categorical discussion, nine of these scenarios are obtained (gold buy, no buy, sell; bitcoin buy, no buy, sell). [4]

Based on the total assets equal to the total gold dynamically planned investment return and the total bitcoin dynamically planned investment return, the total assets trend can be depicted, as well as the bitcoin holdings share trend and the gold holdings share trend, as shown in Figure 7.

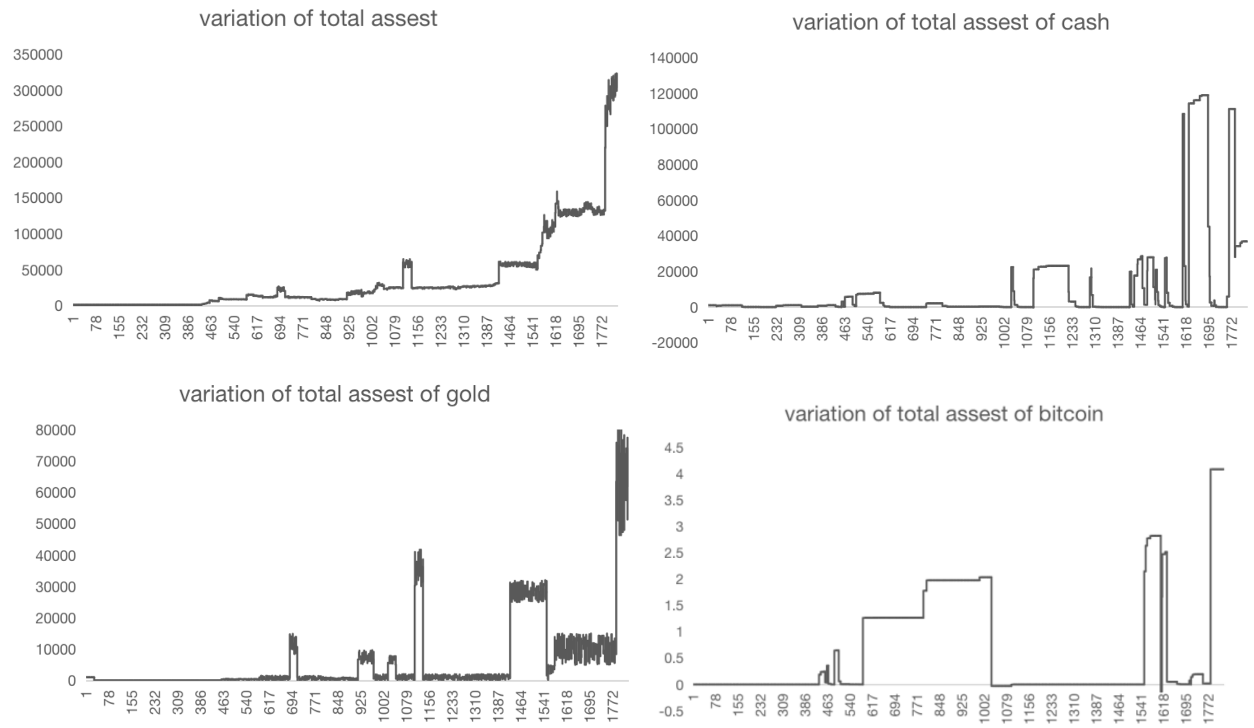


Figure 7: The variation of total assets/cash/gold/bitcoin

We can see through the individual charts that although the volatility varies in the earlier periods, they all peak at the end of the period, so the model is considered reasonable and a more substantial return can be expected on September 10, 2021.

6. Model Evaluation

The selected market sentiment factors include short, medium and long term, so that a more comprehensive and holistic measure of the market sentiment is possible. The information obtained on stops and drops is by using principal component analysis, which retains most of the information on these variables The daily closing prices are predicted using random forest, which has proven to be much more accurate than the time series results, improving the accuracy of our experiments and reflecting the feasibility of our strategy.

For the risk model, the VaR value-at-risk model can be used as an alternative to determine the minimum loss in percentage of portfolio value that is expected to occur in a certain percentage of time under current market conditions through the strategy we obtained above, which can be considered to modify the model using the indexes above.

References

[1] Nana Lin, Jiangtao Qin. Forecast of A-Share Stock Change Based on Random Forest [J]. Journal of University of Shanghai for Science and Technology, 2018, 40 (3): 267-273.

[2] Qixiang He, Yutong Ma. A Classification Study on Risk Premiums of Commodity Futures under Different Trading Strategies [J]. Journal of Industrial Engineering and Engineering Management, 2019, 33(3): 52-60.
[3] Haixiang Yao, Junwei Li, Shenghao Xia, Shumin Chen. Fuzzy Trading Decision Based on Apriori Algorithm and Neural Network [J]. Journal of Systems Science and Mathematical Sciences, 2021, 41(10): 2868-2891.
[4] Wenlan Huang. Research on Portfolio Optimization Based on Multi-objective Evolutionary Algorithm [D]. Harbin University of Commerce, 2020.