Best Strategy: Lower Risk & Higher Return

Jingyi Shi

School of Electronic Engineering, XIDIAN University, Xi'an, Shannxi

916651967@qq.com

Abstract. People have an innate desire for money. Gold and bitcoin, these two high-return investments are popular in the world. We built XGBoost-based regression prediction model and mean-absolute deviation model based on the mean-variance model. We use machine learning and construct an XGBoost regression model for prediction [1]. With the forecast data from above model, we build a mean-absolute deviation model using the final assets as the objective function to maximize the goal.

Keywords: Gold; Bitcoin; Trading Strategies; Genetic Algorithm.

1. Introduction

1.1 Background

In 2009 bitcoin, a new trading currency, was born, from its inception, a bitcoin was just less than $0.01, in 2017 the huge gains in the bitcoin trading market has been comparable to stock gains, has reached $20,000 a piece, along with the huge gains, its riskiness also emerged immediately after the daily trading price reached its peak, and then a plunge of greater than 50% occurred, many people in Overnight riches, there are also many people in the nothing.

Gold, the natural metal, with its excellent non-oxidizing, lustrous and shiny natural properties already predicts that it will be an excellent item for trading. Unlike some other assets, gold is more liquid and its quotes change more rapidly, and sometimes these changes can have a substantial impact on the daily price of gold.

For both types of currencies, traders have to develop strategies in order to hedge their risks to a certain extent and maximize their profits. [2]

2. Preparation of the Models

2.1 Assumptions and Justifications

- Investors are a rational group that seeks to maximize profits.
- The investor knows before investing that the investment return is a variable that follows a normal distribution and uses the standard deviation or variance of its investment return to represent the risk of the investment.
- Investors want to minimize risks and maximize benefits.
- Return on investment, risk is the main factor that influences investors to make investment decisions.
- Investors are rational.
- Markets are time-sensitive.
- All investors are considered to be in the same single investment period.
- If the gold market is open, then gold and bitcoin must be traded at the same day.

3. Predictive Model Building

3.1 Model Building

The purpose of our model is to achieve a more accurate prediction of the daily gold and bitcoin trading prices, so that it can provide an effective reference for the actual traders, when to buy, when

to sell, when to hold, and only by predicting the prices can we make good trading decisions and thus maximize the benefits.

We use XGBoost to build the model, and certain parameters need to be set before modeling. [3]

The objective function to be optimized consists of two parts.

$$Obj = \sum_{i=1}^{m} L(y_i, \hat{y}) + \sum_{k=1}^{K} \Omega(f_k) \tag{1}$$

$$\hat{y}_i(t) = \sum_{k=1}^{t} f_k(x_i) = \sum_{k=1}^{t-1} f_k(x_i) + f_t(x_i) \tag{2}$$

XGBoost will have a prediction score on each leaf node, generally called a leaf weight, denoted by or. [4] This leaf weight is the prediction value of this tree for all samples on that leaf node. Each tree has its structure, which can be measured by the depth of the tree, the number of leaf nodes, and the position of the leaves, which can be denoted by denoting the leaf node where the sample is located, and by denoting the leaf weight of the sample falling to the first coconut node of the tth tree, whereupon there is:

$$f_t(x_i) = \omega_{q(x_i)} \tag{3}$$

A model is built with time as the independent variable and transaction unit price as the dependent variable, which is first trained using a portion of the given data and then predicted for the remaining cases.

3.2 Predicted Results

It is easy to see from the two line graphs that the results are fitted more perfectly, both for gold and bitcoin. Regarding the prediction of gold, there are almost equal data with a dozen or so at some points, and even at the place where the difference between the predicted and actual values is the largest, the difference is less than $100, which shows that the fit is almost perfect; the prediction of bitcoin is more outstanding, which shows that except for some individual points, the folds of the predicted and actual values almost coincide, which is more evidence of the perfectness of the fit, and for the differences of We will use the optimization model to make further and more accurate predictions for the individual parts.

In particular, when dealing with anomalous data for Bitcoin, we compared two processing methods, directly eliminating and retaining the anomalous data, with the following comparison of the predicted results after processing.

It is not difficult to see that the prediction results generated after retaining the outlier data fit better with the true value, which is not difficult to understand when combined with the reality. In reality, market disturbances inevitably occur, huge changes in the global economy, massive buying and selling by companies or individuals can have an extremely large impact on the price, so when these disturbances occur, certain values will show a significant over-average, i.e. become outliers, but here we choose to retain them because the gold and bitcoin trading markets are risky and highly susceptible to disturbances, and directly eliminating outliers would destroy the integrity of the data and make the processed data not reflect the change in price over time in real situations.

3.3 Model Optimization Using Genetic Algorithms

Genetic algorithm is a kind of stochastic search algorithm with the help of natural selection and natural genetic mechanism in the biological world, which can automatically acquire and accumulate knowledge about the search space during the search process, and adaptively control the search process to find the optimal solution. [5]

The genetic algorithm represents the solution of the problem as a "chromosome", which starts from a bunch of chromosomes and uses the principle of survival of the fittest to select the chromosomes

with high adaptability for replication, and generates a new generation of "chromosomes" that are more adaptable to the environment through two genetic operations: crossover and mutation. "As the genetic algorithm runs from generation to generation, those highly adapted models will grow exponentially in the offspring, and finally the most adapted chromosome will be obtained, i.e., the optimal solution of the optimization problem [6]. According to the modeling principle of xgboost, it is known that the mean square error MSE and the coefficient of determination are the main factors of prediction accuracy, so the genetic algorithm is chosen to search out the two best parameters of XGBoost. The genetic algorithm uses genetic operators to operate on individuals in the population, and by continuously exchanging chromosomal information, the population evolves so that individuals with good fitness values are retained and those with poor fitness values are gradually eliminated in the evolutionary process. [7]

4. Regression Model Building

4.1 Model Selection

Since we have a combination of cash, gold, and bitcoin at the same time, a multi-objective evolutionary algorithm is the main method to solve the multi-objective optimization problem so that each objective can reach the optimal set of equilibrium solutions as much as possible. This also means that: we sell gold or bitcoin when the predicted return is larger at this point, and we can buy it when the opposite is true. We assume that gold and bitcoin are bought and sold at the same time, so that we can determine the date of the transaction (buy and sell) based on the return combined with the risk factor, i.e., the variance (date), and it should be noted that gold is not traded on the day of rest, so the G in the triplet should be kept constant and only bitcoin is calculated.

Since the mean-variance model is a static model that cannot match the changes of the stock market in real life, we can invoke a new investment algorithm model, the mean-absolute deviation model, which uses absolute deviation instead of variance to modify the original model, which not only retains the advantages of the original model and is more adaptable to market changes, but also can use some methods to correct the new model to get the better results we want in terms of return. [8]

We used the XGBoost model for prediction and genetic algorithm for optimization, and later we chose the mean-absolute deviation model for risk analysis and evaluation to determine when the risk of the transaction is minimal, and later we also optimized this model, followed by genetic algorithm optimization of the objective function and multiple iterations of the results to achieve the closest true value.

4.2 Model Building

The mean of the historical real rate of return is used as an estimate of the expected rate of return on total income, and the variance of the expected rate of return on gold or bitcoin is used to quantify the risk assessment.

Historical real rate of return calculation:

$$r_i = \frac{S_i(t) - S_i(0)}{S_i(0)} \tag{4}$$

$$\sigma^2 = D(R_i) = \sum_t^T \left(r_i^t - \overline{R_i}\right)^2 p^t \tag{5}$$

The larger the variance, the greater the deviation between the actual return and the expected return, indicating that the investment risk in this matter is greater and it is not appropriate to buy at this time.

The expected rates of return are:

$$r = \frac{S(1) - S(0)}{S(0)} \tag{6}$$

Two market conditions:

• Bitcoin and gold markets open at the same time.

$$\text{Max}_Z = \frac{n_1}{m_1}(z_1 + x_1) + \frac{n_2}{m_2}(z_2 + x_2) \tag{7}$$

$$st \begin{cases} -z_1 - z_2 < x_1 + x_2 < z_1 + z_2 \\ -z_1 < x_1 < z_1 \\ -z_2 < x_2 < z_2 \\ 0.01 \leqslant \frac{n_1}{m_1} - 1 \\ 0.02 \leqslant \frac{n_2}{m_2} - 1 \end{cases}$$

• Gold is not open, Bitcoin is open.

$$\text{Max}_Z = z_1 + \frac{n_2}{m_2}(z_2 + x_2) \tag{8}$$

$$st \begin{cases} -z_2 < x_2 < z_2 \\ 0.02 \leqslant \frac{n_2}{m_2} - 1 \\ 0.01 \leqslant \frac{n_1}{m_1} - 1 \end{cases}$$

This is partly what happens when the market for Bitcoin opens and the market for gold closes, and the target equation is still the total return. Constraint 1 is based on the principle that the number of bitcoin transactions must not exceed the number of holdings. Constraints 2 and 3 respectively represent the gold trading when the commission may not be greater than the yield, otherwise no trading will take place. Although gold cannot be traded at this time, we still list the constraints about gold to avoid interference and improve accuracy. [9] It is worth noting that gold should continue to be the same amount as the previous day, and not change until the next available trade.

4.3 Analysis of Results

• Gold and Bitcoin Yield Comparison:

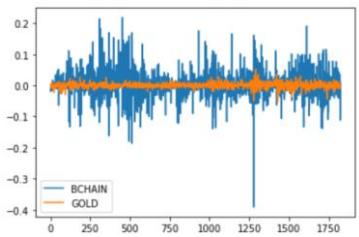
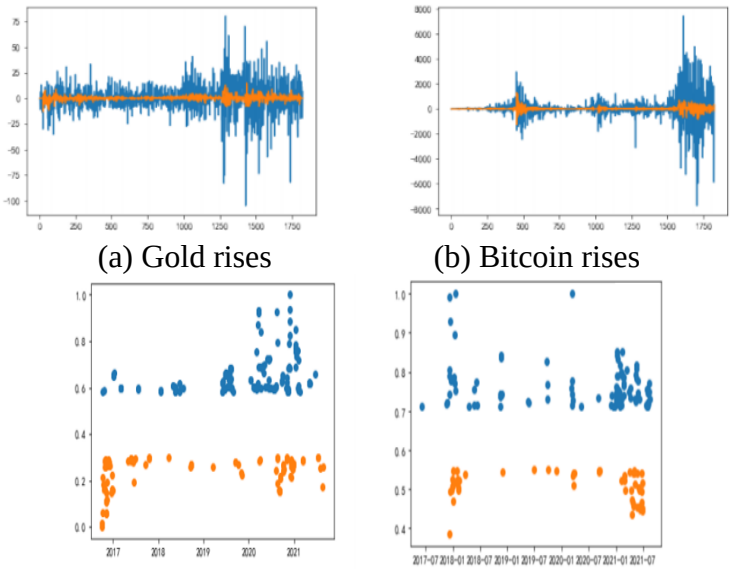


Fig. 1 Comparison of Gold and Bitcoin Daily Yields

By figure1, we calculate the daily returns to determine who is more stable in terms of change between gold and bitcoin, it is clear that gold is more stable than bitcoin.

• Comparison of Expected Increase:

According to the question, we can only use the daily prices of gold and bitcoin so far to determine the trading strategy for the next day. We did this by predicting the gold and bitcoin prices afterwards, and based on the predictions we made a chart of the future price increase of gold and bitcoin. [10] We can clearly see that the orange color is the data already given in the title, while the blue color is what we predicted based on the previous data, and we can make a comparison between the predicted return and the COMMISSION. With the data obtained from the prediction, we find that gold and bitcoin are gradually increasing. See figure 2.



(a) Gold rises (b) Bitcoin rises

(c) Gold (d) Bitcoin

Fig. 2 Blue represents purchase. Yellow means sale

• Transaction Forecast:

We have done the image analysis for gold and bitcoin buying and selling by risk prediction. We have calculated the return on gold and bitcoin, and we can observe by the expected return that we should sell when the price is high, and we should buy when the price is low, so that it is in line with the facts. For example, if I buy gold on the first day at a lower price, but the next day the price of gold rises, so that if I sell my total assets will increase, but if I buy at a high price and sell at a lower price, so that the total return will be a loss. If the absolute value of the difference between tomorrow's price and today's actual price obtained from the forecast is lower than the threshold we set, then no trade will be made.

• Total Revenue Analysis:

This is a graph of the relationship between total assets and dates. Total assets is the final return we calculate according to the algorithm. In the graph, each small color block represents a date and an amount, different colors distinguish different dates and different amounts are sorted by the size of the color block. The amount corresponding to the top left block is the largest, but it is not our total assets as of September 10, 2021. It is a very realistic investment problem because there are risks, ups and downs, and we may not always get the maximum return on September 10, 2021. See figure 3.

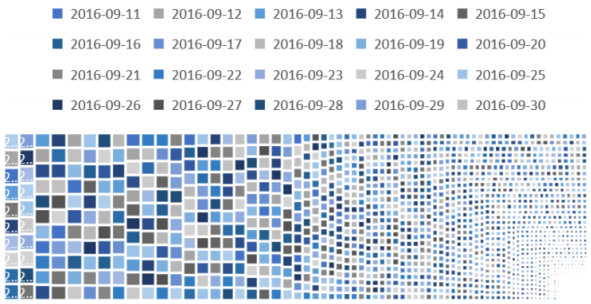


Fig. 3 Tree diagram of total revenue vs. date

References

[1] Guo Changdong. Research on stock prediction based on XGBoost model, 2021.

[2] Lin Xiumei. An empirical study of momentum investment strategies and reverse investment strategies in China's stock market, 2004.

[3] Liang Xuanzhe. A preliminary investigation of quantitative investment in A-shares based on XGBoost, 2021.

[4] Cao Wen. An empirical study of multi-factor stock selection based on XGBoost, 2021.

[5] Vaughnroy Andrae A.Smith. Interactive Genetic Algorithm for Composing Melodies, 2019.

[6] Ba Daoqi. A Multi-Agent Self-Adaptive Genetic Algorithm for Multi-Objective Optimization, 2021.

[7] Wang Puhui. Research and application of quantitative investment decision-making based on machine learning, 2021.

[8] Lin Xiumei. An empirical study of momentum investment strategies and reverse investment strategies in China's stock market, 2004.

[9] Raji Mordecai Folarin. Fast Optimization of Sparse Antenna Array Using Genetic Algorithm and Greens Function, 2018.

[10] Lou Yunjing. Design and implementation of a telecom fraud identification system based on value iteration, 2021.

[11] Zhao Mengna. Research on quantitative strategies based on SVM and BP neural networks, 2021.