Problem Chosen	2022	Team Control Number
C	MCM/ICM Summary Sheet	2208251

The Quantitative investment Based on Machine Learning

Summary

Quantitative investment has been developing for more than 30 years, with stable investment performance, expanding market size and share, and gaining recognition from more and more investors. With the continuous update of machine learning algorithms and computing power, machine learning is becoming more and more important in the Quantitative investment.

In this paper, we aim to systematically develop a model that gives the best daily trading strategy including quantitative timing, portfolio allocation of assets based only on previous price data of bitcoins and gold.

The Random Forest Model is designed to quantify the proper time to trade. We first Separately extract 4 characteristics including MACD, HV, RSI, BOLL from the former price curves of bitcoins and gold based on financial market knowledge. Through training previous 4 characteristics in **random forest model**, we predict a series of important decision points and summary some timing strategy.

The Portfolio investment decision model is based on **the RNN-GRU model** and **Markowitz Programming model**. By training the former price data, a two-layer 260-layer 30-steps RNN-GRU model can be used to predict the future price and risk. In order to minimize risk and maximize benefits, we established a multivariate programming model based on Markowitz model.

After building the above models, we are able to make our trading strategies through the 5-year, and we finally achieve 28% average annualized rate of return and finally get $3382.

To derive models' generality, we simulate each 500 curves of bitcoins and gold by **geometric Brownian motion** and the value of the best decision benefit is always stable at about $3380.

To improve our model, we also make different strategies in different times to make up for the insufficient data when in early time. Succeeded in raising the benefit to about $3395.

In the end, in order to test the sensitivity of the model, we adjusted the transaction coefficient and found that the final transaction value of our model was always relatively stable.

Keywords： Quantitative trading model; RNN-GRU model; Random Forest model; Markowitz Programming model; Geometric Brownian Motion

Contents

1 Introduction

1.1 Problem background

With the development of economy and the continuous improvement of financial market, the transaction of volatile assets has been popular among major companies. Banks, securities companies, professional trading companies and other companies spend a lot of money to recruit excellent market traders to invest in volatile assets for them in order to chase the dividends in volatile asset trading.

Market traders buy and sell volatile assets in order to achieve the maximum total return and generate profits for their companies. If traders misbehave, the company may suffer significant losses, so it is necessary for us to develop a model to determine whether traders should buy, hold or sell assets in their portfolios every day to help traders make trading reference, in order to expand the total return.

1.2 Restatement of the Problem

In this problem, we were given two data sets, *BCHAIN-MKPRU* and *LBMA-GOLD*, and were asked to develop a model that used only the past daily trade price flows from the given two databases to date to determine whether traders should buy, hold, and sell assets in their portfolios on a daily basis.

On the basis of the above conditions, we mainly divided into the following three parts to solve the problem and optimize the model,

- **Develop a model that gives the best daily trading strategy.** We will use the original capital of $1000 to strictly execute the established trading strategy beginning November 9, 2016 and calculate the total amount of return from adopting our trading strategy for the five years ending September 1, 2021.
- Provide some model testing methods to **prove that our model is the best strategy.**
- **Considering the sensitivity of the model to transaction costs,** analyzed the influence of transaction costs on transaction strategy and total return volume**.**

At the end we will present our strategy, model and results to trader in a memorandum of two pages.

1.3 Our Work

Our quantitative trading model covers the whole process of quantitative trading, and our model can interact with traders to a certain extent. The overall model is shown in the figure below.

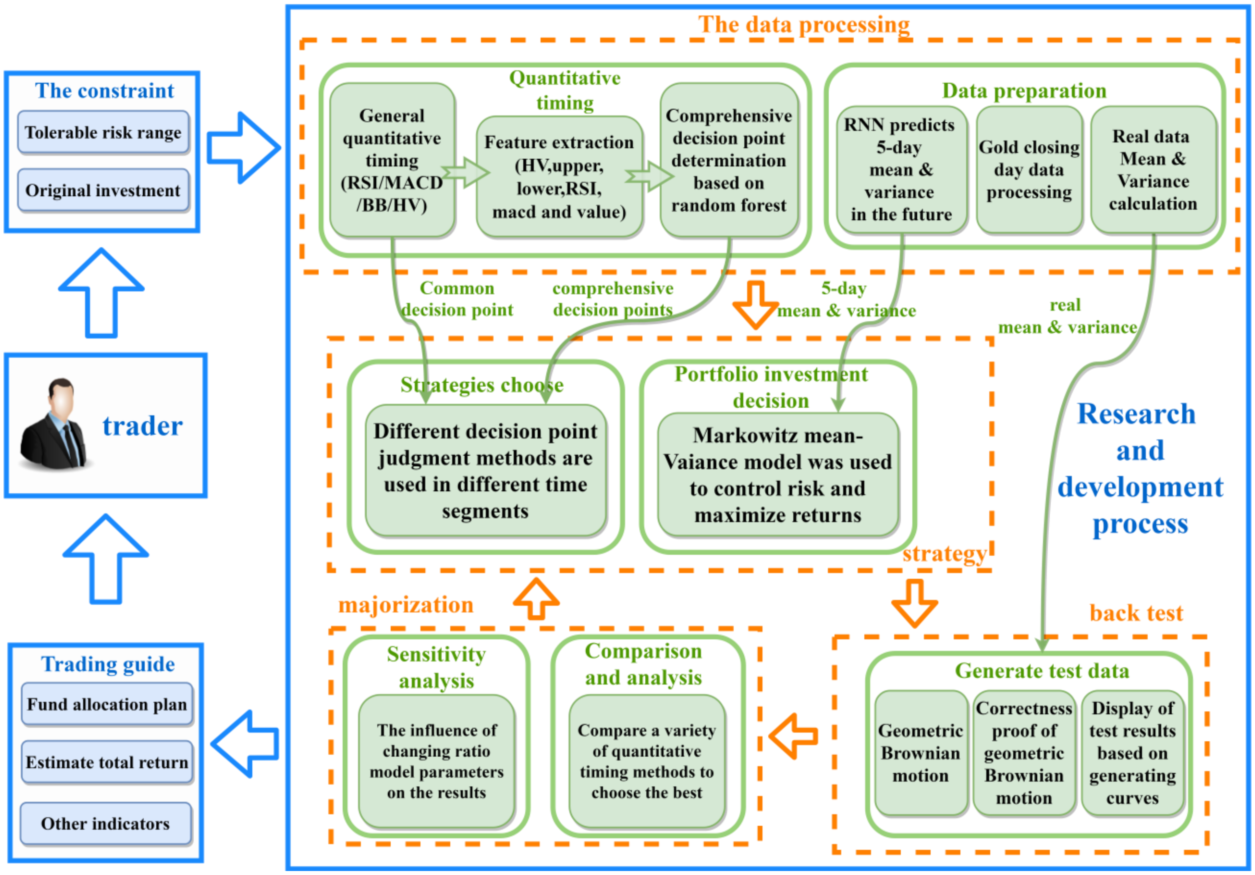


Figure 1: Work Flow

2 Notations

The following table lists the definitions of five characteristic quantities throughout the paper,

Table 1: notation explanation

characteristic quantities	Definitions
$upper$	The upper rail of the Boleyn belt
$lower$	The lower rail of the Boleyn belt
$macd$	Moving average convergence divergence
HV	Historical Volatility
$value$	The closing price

3 Assumptions

Based on the understanding of the meaning of the questions and the convenience of follow-up work, we put forward several assumptions as follows.

- In gold and bitcoin trading, short-selling is not considered, and only going-long is used. In other words, we make money by buying at the current price and selling after the price of gold or bitcoin goes up, earning the difference. Basically, buy before you sell.
- Buying and selling something in dollars during our transaction. This is because we have a small initial investment and cannot buy the entire amount of gold or bitcoin at one time. Measuring gold and bitcoin in dollars makes it easier to calculate.
- Both gold and bitcoin are bought and sold for the day based on the day's closing price. And each type of product can only be bought or sold once a day.
- Consider the dollar's own rate of return, that is, uninvested dollars produce a risk-free rate of return, analogous to a bank's call rate.

4 Quantitative timing model

4.1 Data processing of gold's closing day

Bitcoin is a virtual currency, while gold is an international currency, and the nature of both partly means that bitcoin can be traded on a daily basis, while gold is only traded on days when markets are open. Due to the absence of gold transaction data in some specific dates, for different calculations, the gold data set is processed by the following two methods，

- Delete: In risk assessment and calculation of various assessment indicators, we directly delete missing data to simplify analysis. Deleting missing data directly preserves the most authentic data features.
- Assignment: In buying and selling decisions, the lack of data may affect the consistency of gold and Bitcoin data structures, so we use the previous day's data for closed trading dates. The supplementary data is the same as the previous day. Based on the constraint conditions of the matching model established below, no gold trading can be conducted on the closing day of the gold market.

4.2 Common quantitative timing methods

We summarized several commonly used quantitative timing methods as follows, and extracted some characteristic values from these methods.

- **Relative strength index(RSI)[1][2]**

The principle of the relative strength index (RSI) is to predict the strength of the market movement trend by calculating the range of price rise and fall, and then predict the continuation or reversal of the trend. It actually shows how much gold or Bitcoin prices have moved up as a percentage of the total volatility. The calculation formula of RSI is as follows,

$$RSI_n = 100 - \frac{100}{1 + RS} \tag{1}$$

$$RS_n = \frac{gains_n}{down_n}$$

The variable $gains_n$ indicates average gains over the past n days, and the variable $down_n$ indicates average declines over the past n days.

According to the strength index theory, any market price rises or falls between 0 and 100. According to the normal distribution, RSI value is generally believed to change between 30 and 70. Usually, 80 or even 90 is considered as the market has reached the overbought state, and then the market price will naturally fall back to adjust. When a price falls below 30 it is considered oversold and the market will rebound. If the strength of RSI index is used to make buying and selling decisions, the decision points of buying and selling should comply with the provisions in the following table 2

Table 2: RSI strength determines the trading principle

RSI	operation	Decision point judgment
50-100	Sell	1
20-50	Hold	0
0-20	Buy	1

Where 1 represents the decision point and 0 represents not the decision point. Points with strong RSI values are set as sell points with high urgency to reduce risk and avoid a big drop. We plotted figure 2 using the RSI exponent,

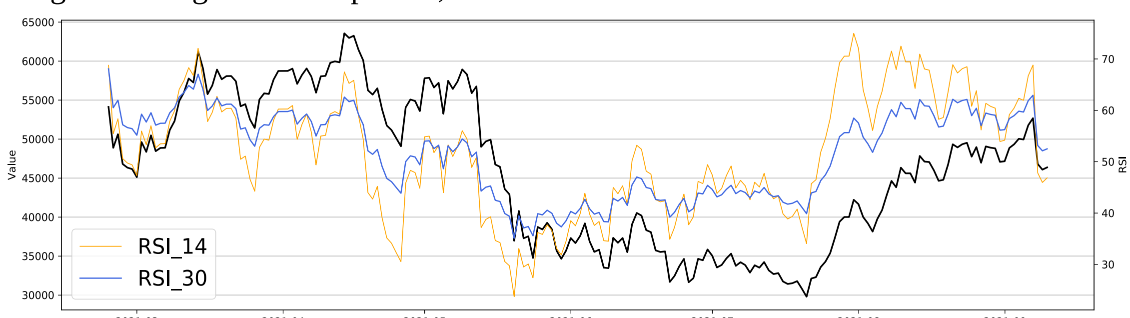


Figure 2: RSI analysis of bitcoin's daily price in 2021

The generated RS_{14} and RS_{30} curves are 14-day RSI curve and 30-day RSI curve respectively. The cross signal of two RSI curves can also be used as a criterion to judge decision points.

The judgment method of decision points based on cross signal amount is as follows:

Table 3: RSI cross signal buying and selling principle

Cross situation	operation	Decision point judgment
RS_{14} goes up through RS_{30}	Buy	1
Not cross	Hold	0
RS_{14} goes down through RS_{30}	Sell	1

The combination of the two RSI curves is: RS_{14} goes up through RS_{30}, which belongs to the long market; If RS_{14} goes down through RS_{30}, it is a bear market.

Generally, RSI index has a certain reference function for the rise and fall of prices and traders' trading strategy planning, and the union of the above two decision points is taken when judging the decision points.

- **Bolliger bands (BB)[3]**

Bolliger bands (BB) are track bands with moving average as the middle line and the mean square error of closing price as the bandwidth. Parameters include two, namely, statistical days m and width (generally 2). The calculation formula of BB is as follows,

$$BOLL = MA(close,\ m) \tag{2}$$
$$upper = BOLL + 2 \times std(close, m)$$
$$lower = BOLL - 2 \times std(close, m)$$

Where $MA(close,\ m)$ represents the m-day simple moving average of the closing price, UB represents the upper track, LB represents the lower track. Using each parameter, we can make the following figure 3,

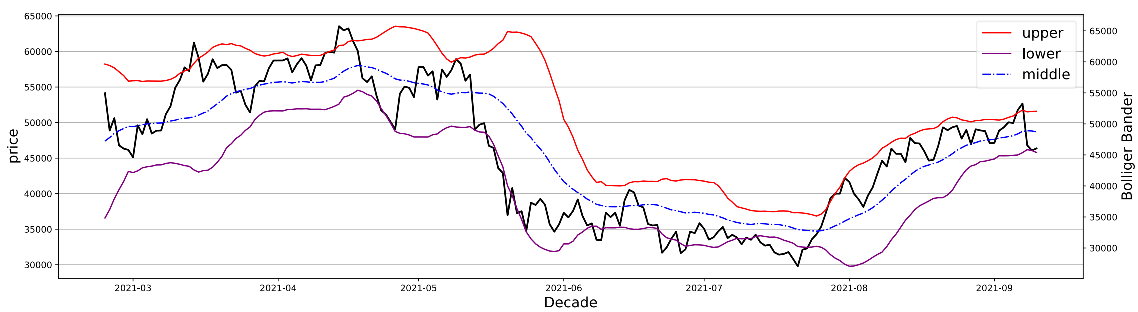


Figure 3: BB analysis of bitcoin's daily price in 2021

In this case, BB uses the wave band to display the safe high and low price of bitcoin, so as to determine the fluctuation range and future trend of bitcoin price. Our criteria for defining bollinger band buying and selling points are shown in the table 4.

Table 4: BB Summary of buying and selling principles

Bitcoin price trends	operation	Decision point judgment
Touch on the upper	Sell	1
Not touch	Hold	0
Touch the lower line	Buy	1

To put it simply, BB's bid-ask judgment principle is to buy when the price of bitcoin touch on the upper and sell when the price touch the lower line.

- **Moving average convergence divergence(MACD)[4]**

MACD is a typical trend-based indicator used to judge the timing of buying and selling by analyzing the convergence and separation between the short-term (12-day term) and long-term (26-day term) moving averages based on the construction principle of averages.

The calculation of today's n-day exponential moving average EMA_n is as follows,

$$EMA_n = \frac{(n-1) \times EMA_n^*}{n+1} + \frac{TF \times 2}{n+1}$$

Where EMA_n^* represents the n-day exponential moving average of the previous day, and TF represents today's closing price.

The calculation of DIF is defined as follows,

$$DIF = EMA_{12} - EMA_{26}$$

Where EMA_{12} is the short-term exponential moving average, and EMA_{26} is the long-term exponential moving average. The EMA of 9 days is calculated according to the deviation value DIF, that is, the average deviation value, which is the DEA value of today. DEA is defined as follows,

$$DEA = \frac{DEA^* \times 8}{10} + \frac{DIF \times 2}{10}$$

Where DEA^* is the DEA value of the previous day. The difference between the DIF and its own moving average is $macd$, $macd$ is defined as follows,

$$macd = 2 \times (DIF - DEA) \tag{3}$$

Take bitcoin as an example and calculate its $macd$ for 2021. Finally, we draw a figure based on the calculated indicators.

Table 5: Summary of MACD trading principles

Histogram color characteristics	operation	Decision point judgment
Change from red to green	Buy	1
Not change	Hold	0
Change from green to red	Sell	1

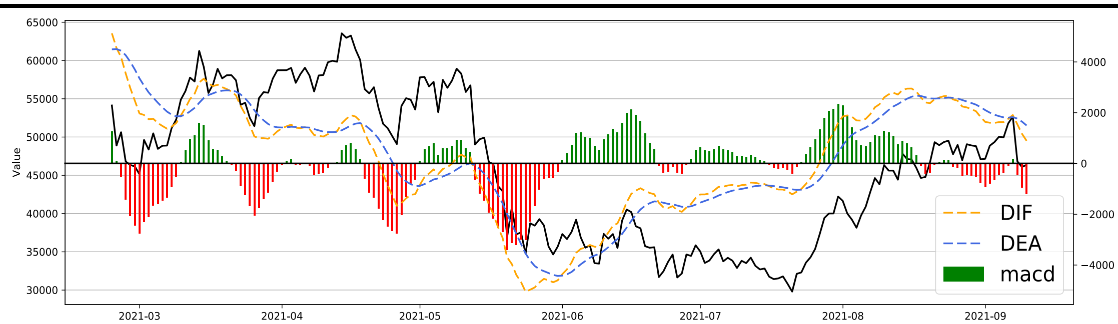


Figure 4: MACD analysis of daily bitcoin prices in 2021

The bar chart in the figure is generated by $macd$ index. From the above figure, we can more intuitively understand the buying and selling principle generated by MACD. That is, the moment when the bar chart changes color is the signal of market reversal, indicating that the overall trend of previous decline or rise basically ends, and asset allocation should be updated at this time. The trading principles are summarized in the table 5.

- **Historical Volatility(HV)**

Historical volatility (HV) refers to the volatility displayed by the market, which is reflected by the historical data of asset market prices over a period of time. Volatility can vividly reflect the volatility of the market. We assume that it is positively correlated with the risk of the market, and a volatile market usually means higher and faster gains or losses and higher risks. We use the following algorithm to calculate its 5-day historical volatility (N=5),

$$\bar{X} = \frac{1}{N}\sum X_i$$

$$X_i = \mathrm{In}\frac{P_{i+1}}{P_i}$$

$$\sigma = \sqrt{\frac{\sum(X_i - \bar{X})^2}{N-1}} \tag{4}$$

P_i represents the closing price of day i, and ， P_{i+1} represents the closing price of day i +1. Based on the above formula, we obtain the 5-day annualized historical volatility of gold $HV(G) = \sqrt{252}\sigma$, and the 5-day annualized historical volatility of Bitcoin $HV(B) = \sqrt{365}\sigma$, where 252 is the number of days in a year that gold can be traded, 365 is the number of days in a year that bitcoin can be traded。

4.3 Decision point determination based on random forest

We know that many trading systems or technical indicators are imperfect and can only be applied to specific market patterns. When they are applied to other markets, they may need to be matched by other indicators, or the parameters may need to be modified.

In response to these characteristics of technical indicators, successful investors can take the following two paths:

- One way is to choose some of the most applicable indicators, repeated tests and even perfect, and then use it in the appropriate market or opportunity, using the very accurate several opportunities to profit, many trading masters are known for being good at using a few indicators;
- Another way is to use multiple indicators comprehensively to verify each other.

In this paper, we choose to use random forest algorithm to comprehensively verify each other with a variety of indicators, so as to determine the decision point. Random forest is in line with the thinking logic of investors in actual investment, so the decision tree model has strong scalability. We can increase or decrease and change technical indicators to create personalized trading strategies.

Random forest is an ensemble learning algorithm, which combines several weak classifiers into one strong classifier. Random forest uses bootstrap to randomly extract k samples from the training set, and selects random features for each decision tree on the basis of Bagging. The k samples are used to build k decision tree models. Finally, the k decision tree model is used to vote to get the result. The specific algorithm steps[6] of random forest are as follows:

- Enter training set D.
- Bootstrap sampling was used to form k training subsets D_k.
- We randomly extract m features from the original features.
- By training the training subset D_k and making the optimal segmentation of m randomly selected features, k decision tree prediction results can be obtained.
- The result with the highest number of votes is obtained by voting according to k prediction results.

We used a post-pruning algorithm to prune each original tree from the bottom up. The algorithm starts to prune at the bottom internal node. If the node is replaced with a leaf node to improve generalization performance, the subtree is replaced with a leaf node. Figure 6 shows a pruned decision tree.

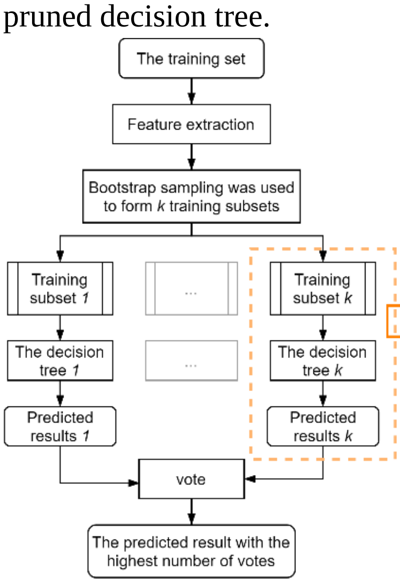


Figure 5: Flow chart of random forest algorithm

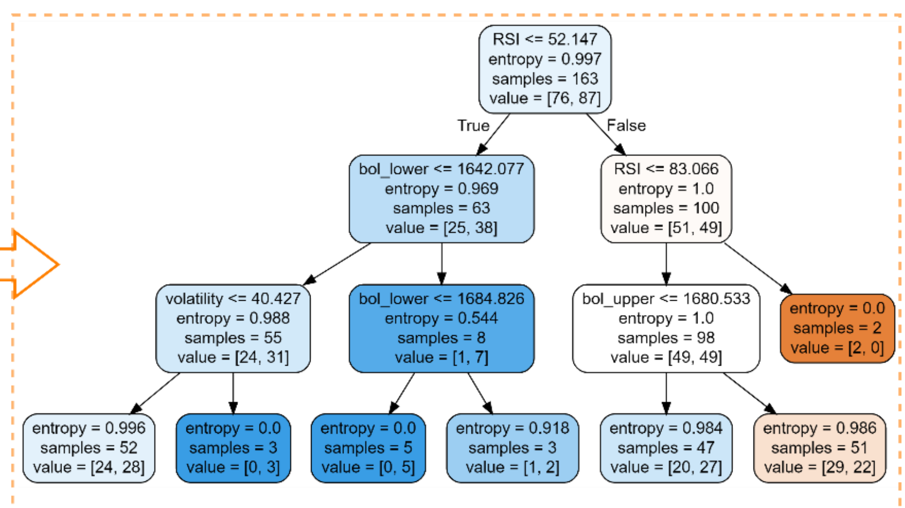


Figure 6: A pruned decision tree

According to the parameters of the decision tree, $upper, lower, RSI, HV, value$ and $macd$, as well as the number of trees and the maximum number of features in the decision tree in the random forest, data were trained and the ranking results of feature importance were obtained as shown in Figure 7. It can be found that the RSI parameter is of high importance in both gold and bitcoin markets.

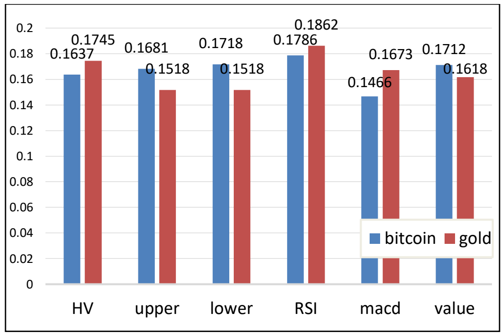
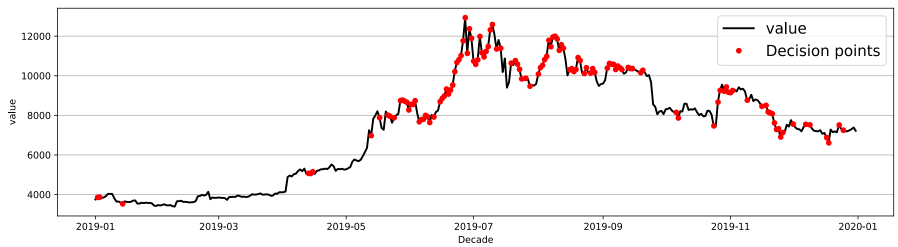
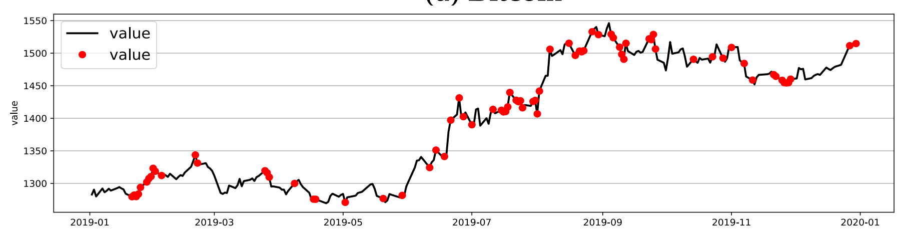


Figure 7: Feature importance display

The decision points generated by random forest voting are shown as follows,



(a) Bitcoin



(b) Gold

Figure 8: Partial decision points of 2019 generated by random forest

From the figure 8, we can clearly see the distribution of buying and selling decision points of gold and bitcoin in 2019. Direct observation shows that most of the decision points are distributed near the market reversal point and the position of sharp rises and falls. In order to avoid severe losses caused by the collapse of gold and Bitcoin as much as possible, we decided to retain all the decision points generated by the random forest, that is, if there is a decision point in a date, whether it is from the gold decision point or from the bitcoin decision point, a round of asset ratio planning will be conducted.

In this section, we adopt Markowitz's concept of "maximizing returns. The concept of risk minimization is used to establish a multi-objective programming model to study the optimization of uncertain return and risk portfolio in portfolio investment decision making. Meanwhile, the price of the next month predicted by the section RNN network is used to evaluate the return and risk. This model examines the impact of asset risk, return, correlation and diversification on portfolio return.

- **Determination of objective function**

We define the asset holdings at decision point t as: : $[C_t\ G_t\ B_t\ ]$, where C_t represents cash holdings and the unit is U.S. dollars, G_t represents gold holdings in ounces, B_t represents bitcoin holdings in units. Asset value at decision point t is defined as $[M_t\ U_t\ W_t\ ]$, where M_t is the price of U.S. dollars, which is always 1. U_t represents the value of gold in US dollars per ounce. W_t represents the value of a bitcoin in U.S. dollars. Therefore, the original portfolio value at time T of this decision point is V_t , where:

$$V_t = C_t + G_t \cdot U_t + B_t \cdot W_t$$

Where C_t represents cash holdings before adjustment, G_t represents gold holdings before adjustment, and B_t represents bitcoin holdings before adjustment. The commission costs of gold and Bitcoin transactions are respectively $\alpha\%$ and $\beta\%$ of the transaction amount, so we can get the commission costs needed to adjust assets:

$$H_t = \alpha\% \cdot |G_{t+1} - G_t| \cdot U_t + \beta\% \cdot |B_{t+1} - B_t| \cdot W_t$$

Where C_{t+1} represents the adjusted cash holdings, G_{t+1} represents the adjusted gold holdings, and B_{t+1} represents the adjusted bitcoin holdings. The expected value $E_m(U)$ and $E_m(W)$ of the next five days can be obtained from the prediction of gold and bitcoin by THE RNN network in the previous section. The objective function of maximizing the expected return of total assets in the next month can be obtained,

$$\max\ Z_t = C_{t+1} + G_{t+1} \cdot E_m(U) + B_{t+1} \cdot E_m(W) - V_t - H_t$$

In addition, the festival has been using RNN network to predict the price of five days for the next five day price variance and covariance with gold and currency for the price of gold and currency covariance matrix$\begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix}$, we used to mobilize after the value of the asset portfolio variance represents the future of investment risk, The objective function of risk minimization can be obtained:

$$\begin{aligned} Min\ D_t &= D(G_{t+1} \cdot U + B_{t+1} \cdot W) \\ &= G_{t+1} \cdot \sigma_{11} + 2\sigma_{12} \cdot G_{t+1} \cdot B_{t+1} + B_{t+1} \cdot \sigma_{22} \end{aligned}$$

σ_{11} is the variance of gold over the next five days, σ_{22} is the variance of Bitcoin over the next five days, and σ_{12} and $\cdot\sigma_{21}$ are the covariances of gold and Bitcoin prices over the next five days.

- **The constraint**

1. it is assumed previously that shorting is not allowed in asset allocation investment,

$$C_{t+1},\ G_{t+1},\ B_{t+1} \geq 0$$

2. The assets are reallocated at the decision point and the portfolio value V_t' at the decision point t is adjusted.

$$V_t' = C_{t+1} + G_{t+1} \cdot U_t + B_{t+1} \cdot W_t$$

The portfolio value obtained after redistribution shall be equal to the sum of the portfolio value before allocation and the commission cost incurred in allocating the asset investment,

$$V_t - H_t = V_t'$$

3. The gold trading market is closed on weekends and some holidays (Christmas and New Year's Day), so when the decision point coincides with the closed date, we use the data preprocessing process to assign the closed gold price to the price of the previous day as the judgment benchmark,

$$G_{t+1} = G_t ,(U_{t+1} = U_t)$$

4. Considering the psychological factors of investors themselves, general investors have a certain range of risk tolerance. Therefore, we improved the risk requirements of investors by changing the constraints in the multi-objective programming model.

$$D_t \leq threshold$$

Where, $threshold$ is the maximum risk threshold given by the customer

- **Integration of conditions and goals**

The multi-objective programming model for balancing benefits and risks is established as follows:

$$\begin{gathered} \max\ Z_t = C_{t+1} + G_{t+1} \cdot E_m(U) + B_{t+1} \cdot E_m(W) - V_t - H_t \\ \min\ D_t = G_{t+1} \cdot \sigma_{11} + 2\sigma_{12} \cdot G_{t+1} \cdot B_{t+1} + B_{t+1} \cdot \sigma_{22} \\ s.t. \begin{cases} C_{t+1}, G_{t+1}, B_{t+1} \geq 0 \\ V_t - H_t = V_t' \\ G_{t+1} = G_t ,\ (U_{t+1} = U_t) \\ D_t \leq threshold \end{cases} \end{gathered} \tag{5}$$

We take profit maximization as the first-level objective of optimization and risk minimization as the second-level objective. Multi-objective planning is carried out at decision points to reallocate asset holdings.

- **Demonstration of solution results of multi-objective programming model**

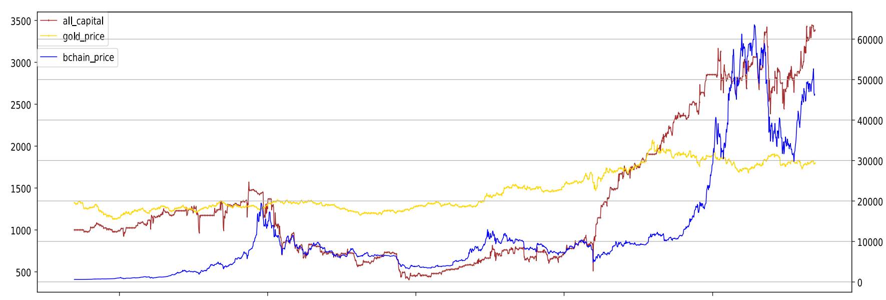
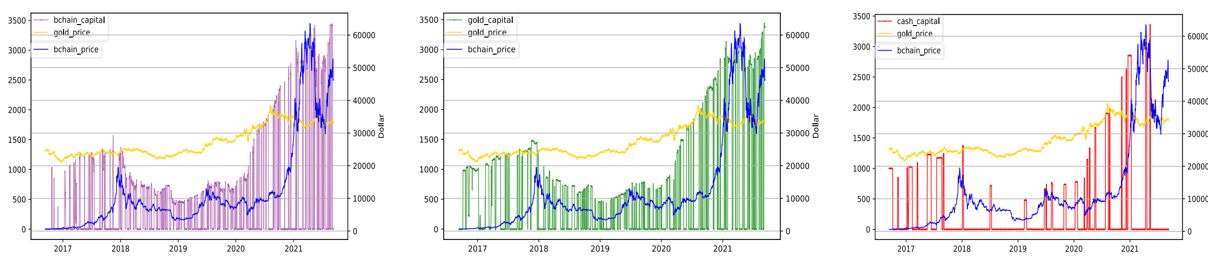


Figure 9: Trends in gold and dollar investment volumes

For five years trading assets value change with the gold, the currency of the relationship of the trend of prices we use double y axis will be the trend of drawing as shown above (the currency for the right y axis) alone, for the above changes the value of total assets and the actual price trends of comparative analysis, we preliminary analysis for the following:

- In 2018, when the market trend was not good and volatility increased, the total asset value lost, but still controlled a certain loss rate
- Bitcoin showed a continuous upward trend at the end of 2017 and 2020, and our model grasped the upward trend and ushered in a considerable return rate on the total asset value.
- At the beginning of 2017, when the gold price curve fluctuated while the bitcoin value curve fluctuated little, the total asset value fell slightly, but it also showed a trend of steady rise.



(a)bitcoin (b)gold (c)doller

Figure 10: Asset ratio and price curve trend comparison diagram

In order to study how the objective planning model balances benefits and risks to make decisions, we also make comparative analysis of the three asset values with the price trends of gold and Bitcoin, and draw the following strategic conclusions:

- Due to uncontrollable factors such as market disturbance the price of gold and currency, RNN five days of expectations and random forest prediction decision points still can not completely replace the future trend of the market, in 2018, the currency falling trend because of RNN forecast and actual price in contrast to distribute assets mainly to the currency, cause downward trend; In our optimization model, assets are transferred to gold to control a certain loss rate
- When weighing and predicting the downturn of gold and bitcoin in the future or the high investment risk, our planning model will adopt the strategy of withdrawing all positions, especially in 2020 and 2021 when the price volatility of bitcoin and gold is high, this strategy will be frequently adopted.
- The higher value intensity of bitcoin asset holdings in the figure corresponds to the rise in the price of total assets, because our optimization model grasped the rising trend while considering risk factors, thus achieving a considerable return rate on asset value.

Based on the above analysis, we established a better grasp of the trend of the model, has a certain stop loss ability and a good return.The investment started from $1000 on November 9, 2016 to $3381.94 on November 9, 2021, achieving a good five-year average annualized return of 27.6%.

5 Portfolio investment decision model

5.1 Index prediction based on GRU-RNN

Recurrent neural network (RNN)[7] is a special neural network structure, which is based on the idea that "human cognition is based on past experience and memory". It not only takes into account the input from the previous moment, but also gives the network a kind of 'memory' of the previous content. That is, the current output of a sequence is related to the previous output. Here, we use RNN to summarize the past experience and memory of gold and bitcoin in order to predict future market changes.

Based on the high volatility of gold and bitcoin market, it is easy to be affected by external environment, policy, economy and other factors. We introduce GRU gating model based on RNN network. GRU's update gate z_t and reset gate r_t prevent the memory from focusing too much on past features. Compared with LSTM model, GRU is more concise, which is conducive to repeated iterative training in the process of prediction.

Based on the transaction characteristics and rules of gold and Bitcoin markets, we simply built a prediction model consisting of three 30-step GRU layers with 80, 180 and 240 memories, respectively. The past six features $upper, lower, RSI, HV, value$ and $macd$ are used to provide the basis of parameters for the capital allocation planning model. The mean and variance of gold and Bitcoin prices over the next 5 days: $\boldsymbol{E_5}\,(U)$,$\boldsymbol{D_5}\,(U)$,$\boldsymbol{E_5}\,(W)$,$\boldsymbol{D_5}\,(W)$. Where U represents the price variable of gold in the next five days, W represents the price variable of bitcoin in the next five days.

RNN network uses the data of the previous year for model training, and takes the data of 30 days before the prediction date as the characteristic value to predict the price of the next five days, so as to obtain the expectation and variance of the next five days. After the next day, the training set and eigenvalues are rolled over and the prediction is made. Taking Bitcoin as an example, the following pseudo-code is used to illustrate:

$For\ date\ from\ 2017-9-10\ to\ 2021-9-10$

$\quad practice\ GRU\_Model(the\ features\ of\ data-365\ to\ date)$

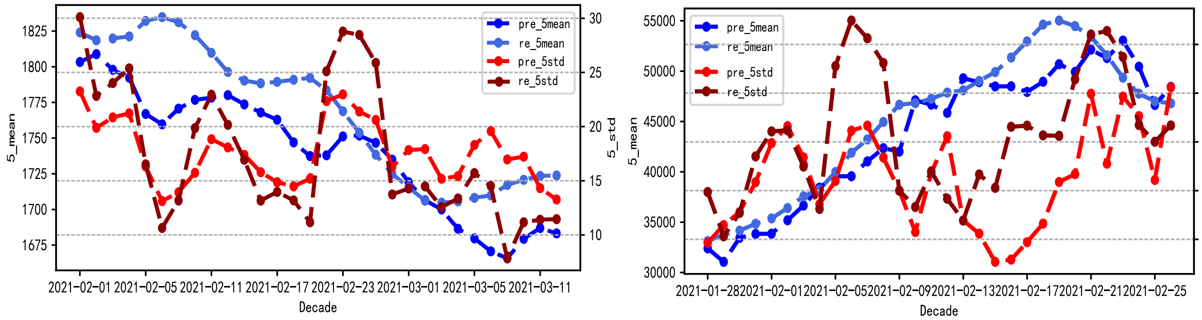
$\quad use\ Model(the\ features\ of\ data\ to\ date+30)$

$\quad predict\ the\ value\ of\ the\ data\ to\ date+5$

$\quad caculate\ the\ E(B_5)\ and\ D(B_5)$

$\quad date=date+1$

We use the mean and variance obtained in the next five days to indicate the price trend of bitcoin in the next five days. To solve the problem of gold market closing, we use the method of assigning value in "gold closing day treatment" to reasonably predict the trend and risk of gold in the next 5 days. Next, we listed the real 5-day mean and variance charts of gold and bitcoin in some months in 2021 to verify that they better represent the future price trend and risk.



(a)gold (b)bitcoin

Figure 11: Comparison of predicted and true values for the next five days

It can be seen from the figure 11 that the variance and expectation have certain similarity with the real price, which can better represent the trend of bitcoin or gold market. We respectively use the mean value of gold $\boldsymbol{E_m}\,(U)$, value variance $\boldsymbol{D_m}\,(U)$, $\boldsymbol{E_m}\,(W)$, and $\boldsymbol{D_m}\,(W)$ in the next five days.

5.2 A multi-objective programming model based on Markowitz thought

The decision points suitable for investment change have been identified above. Now we consider how to allocate funds at the decision points, that is, to study the optimal asset portfolio for the decision points. The so-called asset portfolio refers to investors allocate investment funds to several kinds of assets, and the investment amount of each kind of assets accounts for a certain proportion of the total investment of investors, so as to make the overall income of the assets held by investors as high as possible, while making the risk as low as possible.

6 Evaluation of optimal performance of trading strategy model

6.1 Geometric Brownian motion generates test data

- **Construction of geometric Brownian motion**[8]

To prove that our model can provide the best strategy, we will generate a set of future market trends to test the correctness of the strategy.

Through statistics, we found that the price fluctuations of gold and bitcoin presented uncertain results in some experiments, and showed statistical regularity in a large number of repeated experiments, which proved that the market fluctuations of gold and bitcoin were stochastic process.

A stochastic process S_t is said to follow geometric Brownian motion if it satisfies the following stochastic differential equation (SDE). Constructing a model of the price changes of gold and Bitcoin with respect to geometric Brownian motion:

$$dS_t = \mu S_t d_t + \sigma S_t dz$$

Where $dz = \varepsilon\sqrt{dt}, \varepsilon \sim N(0,1)$, dz is an increment of the standard wiener process, the discrete form of the model can be written as,

$$\Delta S = \mu S \Delta t + \sigma S \varepsilon [\Delta t]$$

Let dS_t be the change in the price of gold or Bitcoin over the same short time interval. μ is the instantaneous expected drift rate of its price, that is, the expectation of the expected price growth rate within t time. σ is the volatility of the logarithm of the price over time t, i.e. σ is the instantaneous standard deviation of the closing price increase. The relative price of gold or bitcoin x is,

$$x = \frac{s_{(t+\Delta t)}}{s_{(t)}}$$

In time t, as the price changes of gold and bitcoin are affected by a variety of external factors, such as economic policies, without regularity and interference, the superposition of various factors on the real estate market makes the real estate price $s_{(t+\Delta t)}$ a random variable and follows the normal distribution as follows,

$$\ln(x) \sim N(\mu, \sigma) \tag{6}$$

- **Feasibility test of geometric Brownian motion**

We found that $\ln(x)$calculated from the data of bitcoin showed a slight negative skew as shown in Figure 12, basically presenting a normal distribution.

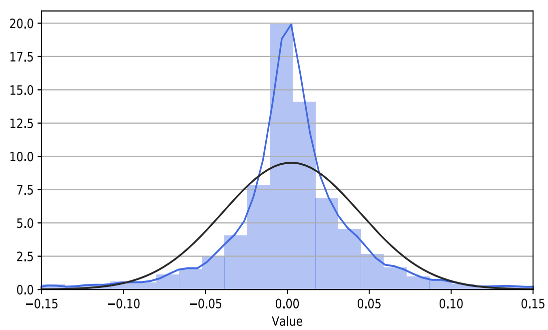


Figure 12: Test of data normality

In order to ensure that the fitted curve was basically consistent with the original market trend, we further conducted spearman's correlation coefficient test on the data obtained by geometric Browne, and obtained that the correlation coefficient between the simulated bitcoin curve and the original curve was about 0.7854825. The correlation coefficient between the gold simulation curve and the original curve is about 0.8663413, indicating that the trend is basically consistent.

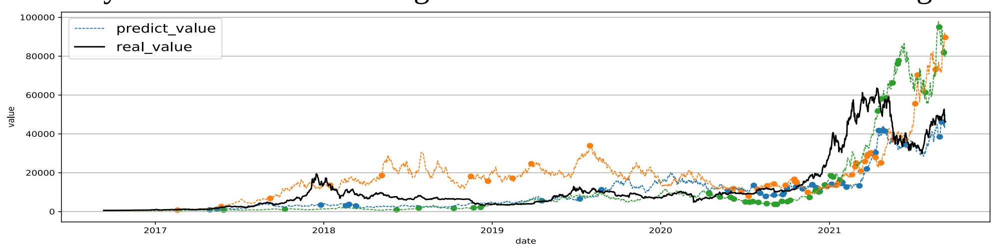
- **Geometric Brownian motion generates multiple test data**

The parameters required for Brownian motion to produce price trends for gold and Bitcoin are calculated directly from real data, the values of which are listed in Table 6,

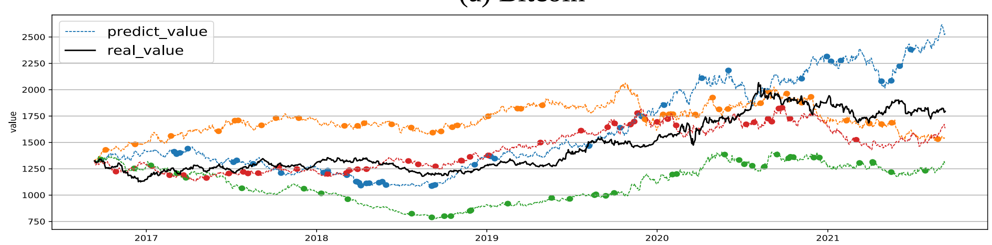
Table 6: Annual σ and μ values from 2016 to 2020

	Gold		Bitcoin	
Year	σ	μ	σ	μ
2016	0.1255066	0.0014333	1.9218475	0.7320619
2017	0.0867607	-0.1038957	0.3934960	0.9459906
2018	0.1113407	0.2368993	0.4726719	0.6886334
2019	0.1863833	0.2509477	0.0124750	0.8175075
2020	0.1555407	-0.082039821	1.5115120	0.7871302

This table covers the real σ and μ of gold and Bitcoin from 2016 to 2020. The σ and μ generated by the real values can reflect the impact of major events on the overall price trend of gold and Bitcoin in a year, which can more accurately reflect the changes of the real market. The curve generated by Brownian motion using the data in the table is shown in the figure below,



(a) Bitcoin



(b) Gold

Figure 13: Brownian motion generated test data and its decision points

The black lines in the figure 13 represent real market price movements, and the dotted lines in different colors are test curves generated by Brownian motion. Here we simply show a few test curves generated by Brownian motion. In fact, we will generate dozens or even hundreds of curves to test the total return amount, number of decision points and other parameters of our model.

6.2 Model performance test

- **Display of test results based on generating curves**

For curves generated by geometric Brownian motion, we still extracted their features $upper, lower, RSI, HV, value$ and $macd$ first, and then selected their decision points by random forest algorithm. The generated curves and curve decision points are shown in figure 13.

Due to the large amount of calculation, the GRU-RNN model and Markowitz Mean-Vaiance model were used to calculate the total return amount when 500 pairs of curves adopted the best trading strategy, and simple statistics and analysis were conducted. The calculation results are shown in the following table，

Table 7: Test results presentation

Total return volume range	Number	Average Bitcoin correlation coefficient	Average Gold correlation coefficient
(2700,3000)	32	0.6458965	0.7543264
(3000,3300)	26	0.7795451	0.8865322
(3300,3600)	44	0.7698458	0.8653253
(3600,3900)	50	0.8134784	0.8589434
(3900,4200)	48	0.5907567	0.8232145

In the statistical process, we found that our decision model had a good effect in all the transactions of the prediction curve, basically maintaining an annual return rate of 28%, the lowest about 21.374%, the highest about 33.873%. Much higher than time deposit or other financial management methods. Moreover, we find that the correlation between the simulated curve and the original curve will be reduced to some extent when the return rate is usually higher or lower, but the overall effect is still satisfactory.

In addition, our investment strategy effectively resisted risks, and its final average return rate in 100 pairs of simulated curves was about $3380, and its final average return Standard Deviation was about 54.7. This has a good ability to resist risks and gain benefits.

6.3 Optimization of trading strategies for different periods

- **The calculation of the mean value and variance**

If the GRU-RNN model is needed to make a good prediction of the mean and variance of gold and bitcoin prices in the next 5 days, we need to ensure that there are enough data sets to train the GRU-RNN model.

The results of several experiments showed that GRU-RNN model began to show good accuracy of mean and variance prediction only after a year of data training. In order to improve the improper capital allocation caused by the prediction deviation of GRU-RNN model in the early stage, the variance of gold and bitcoin is similar in a small range, so we adopt the mean and variance of 5 days before the decision point to plan the capital allocation strategy.

- **Buying a time quantifies the choice of timing methods**

If we need to use the random forest algorithm, we also need to ensure that there are enough data sets to train the random forest algorithm. However, at the beginning of the transaction, we did not have enough data to train the random forest, so we considered using other methods to select decision points at the beginning of the transaction.

Three common quantitative timing methods (BB, MACD and RSI) mentioned above were used to select the decision points in the first year. The decision point determination rules of the three common quantitative timing methods have been given in the above common quantitative timing methods.

- **Model verification after optimization**

We tested three general quantitative timing methods and found that MACD performed the best in terms of returns, and $macd$ was also the most important parameter index in the random forest verification mentioned above. Also, 500 pairs of test curves were generated by geometric Brownian motion. The final results obtained from the decision points in the first year of MACD decision making combined with the above optimization method of mean and variance are shown as follows:

Table 8: Test results presentation

Total return volume range	Number	Total return volume range	Number
(2700,3000)	6	(3600,3900)	55
(3000,3300)	32	(3900,4200)	46
(3300,3600)	54		

In the statistical process, we found that our optimized trading strategy had a better effect in most of the transactions of the forecast curve, In the 500 pairs of simulated curves, the final average return is about $3395, and the final average return standard deviation is about 54. All indexes have reached certain optimization. This has a good ability to resist risks and gain benefits.

7 Analysis of sensitivity to transaction costs

In the quantitative timing model, we use the random forest model to determine the decision point, and the change of the decision point only depends on the six feature quantities of the past historical data learned by the random forest model, so the change of commission cost will not affect the change of the decision point. However, through the analysis of the decision points, our resource allocation model will not implement the adjustment of asset holding at some of the decision points, and we call the decision points that implement the adjustment of asset holding as the real decision points, and there are about 469 real decision points in five years. By changing the commission cost, observe the change of the real number of decision points, that is, the change of trading strategy. Similarly, changes in the commission costs of gold and bitcoin will inevitably affect the final expected total asset value in the planning model of resource allocation. Therefore, we define the changeable commission cost of bitcoin as α_bitcon and gold as α_gold, and explore the sensitivity of both strategies and results:

- **The influence of parameter change on the number of real decision points**

First, we controlled α_{bitcon} or α_{gold} to explore the effect of another value increase or decrease of 0.005% on the number of real decision points. In this way we get the true number of decision points for each of the four cases. In order to specify the change of real decision points, we select the change of commission cost in some time periods to show the change of trading points.

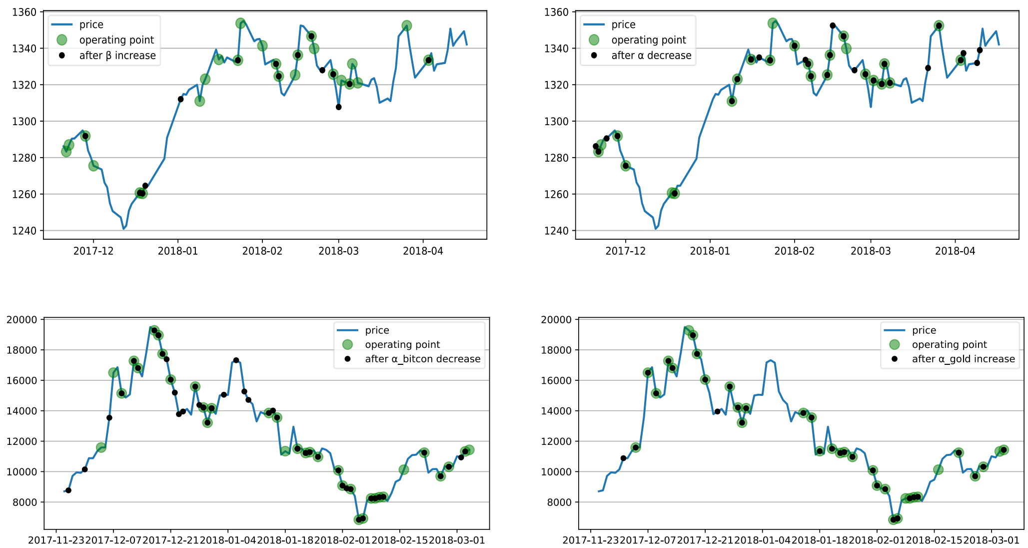


Figure 17: The effect of parameter changes on the number of operation points

In the figure, black dots are used to indicate the increase or decrease of real decision points. The following conclusions are found through observation:

- The change in α_{bitcon} significantly increased or decreased the number of trading points, increasing the number of real decision points from 469 to 511 when it decreased by 0.005%, and some of the trading points changed. When it increases, it also significantly reduces the number of buying decisions.
- The change of α_{gold} parameter has little effect on the number of trading points. When it decreases by 0.005%, the number of real decision points increases from 469 to 486, and only part of the trading points change.

Our analysis:

- Due to its high volatility, the five-year composite annual volatility of bitcoin is about 0.81, so its trading is characterized by high frequency and large volume. Volatility tends to be as high as 10 points, and as we perceive from our portfolio decision model, we tend to be big and fast buyers and sellers. The influence of α_{bitcon} parameter variation on our model accords with the characteristics of our model.
- Gold has low volatility, with a 6-year comprehensive volatility of 0.16. Therefore, in our model, it is often regarded as an asset that avoids risks and has certain income, so its trading frequency is low and the trading volume is small, and the transaction cost itself is not high.

Therefore, the increase or decrease of transaction rate has little influence on our model decision.

- The influence of $\alpha_{bitcon}, \alpha_{gold}$ parameter changes on the final trading results

Quantitative trading is a complex process. In the process of exploring the influence of α_{bitcon} and α_{gold}arameters on the final trading results, we tried to change the two parameters and listed part of the final trading results in the following table,

Table 9: Test results presentation

v	α_{gold}	Result	α_{bitcon}	α_{gold}	Result
+0.005%	+0.005%	3701	-0.005%	-0.005%	3608
-0.005%	+0.005%	3777	+0.005%	-0.005%	3770
+0.0035%	+0.0035%	3542	-0.0035%	-0.0035%	3845

We continuously adjust the above two parameters in a grid search method and get 50 final transaction results. It turns out that they do not increase as transaction costs continue to decrease. Its mean is about 3364.527, and its standard deviation is about 53.

We believe that the reduction of transaction costs means more investment transactions for investors, which can be seen from the impact of the exploration on the number of trading points. The increasing number of trades, especially in riskier assets such as Bitcoin, also leads to increased risk, and thus the probability of losses increases significantly.

Through the above two inquiries, we conclude that our model has good adaptability at least when dealing with changes in market transaction costs. It will increase trading points for bigger gains. Reduce trading points when market transaction costs rise. It has high sensitivity in selecting trading points. At the same time, it will not blindly lead to losses. For the final result, it has a certain stability.

8 Strength and weakness

8.1 Strengths

- Random forest conforms to the thinking logic of investors in actual investment, so the decision tree model has strong scalability. We can increase or decrease and change technical indicators to create personalized trading strategies.
- The method of selecting decision points by random forest greatly reduces the amount of model computation and the workload of traders. Instead of daily trading with a high rate of return, you plan to allocate money only on dates with decision points.
- Our model takes into account the interaction with traders, so trading strategies can be planned according to the range of risks traders can bear.
- Use geometric Brownian motion to generate multiple test data sets to test and analyze model parameters to obtain the optimal value, and the parameter optimization degree is higher.
- Our resource group and allocation model weigh the relationship between return rate and risk, so as to ensure high return with low risk.

8.2 Weakness

- The problem of over-fitting may occur in the classification of random forest with large noise.
- Based on the prediction and point selection model of RNN network and random forest algorithm, there is a lot of data to learn, a relatively large operation process, and a long time to program and run the program
- Risk is not rated, so investors can not necessarily quantify the maximum risk they are willing to take

9 Conclusion

we constructed our quantitative decision model based on random forest model, RNN-GRU network and Markowitz asset allocation model.Our model has a good performance in almost all times. The final annualized rate of return can reach to 26%. In the process of solving the problem, we also summarize a series of trading strategies and important features through the random forest model to help traders find decision points, quickly.To analyze the generality of the model and test its stability, we simulate geometric Brownian motion curves.

We achieve pretty returns in every simulation curve. It’s around $3380. Our model also has good stability and its’ standard deviation is about 57.4 .The optimal result of the model is less affected by transaction costs and the model is stable.

10 Memorandum to the trader

Dear trader:

As electronical information technology continues to evolve and machine learning algorithms continue to iterate, quantitative trading has become the choice of more and more traders. We are very honored to be invited by you to help you develop models to uncover the mystery of the trading market, understand the rules of the trading market, and understand how to allocate assets to achieve high returns and low risks. I hope this memorandum is intended to briefly clarify our views and insights and will be of help to you.

First of all, we have extracted some characteristic parameters, such as $MACD, RSI, HV$, that can reflect the price changes of gold or Bitcoin for you based on some existing financial knowledge. their specific meanings and functions have been introduced to you above.But financial markets are so variable that it's hard to make decisions even when we know these characteristics. So we use random forest to carry out quantitative trading timing for decision dates.

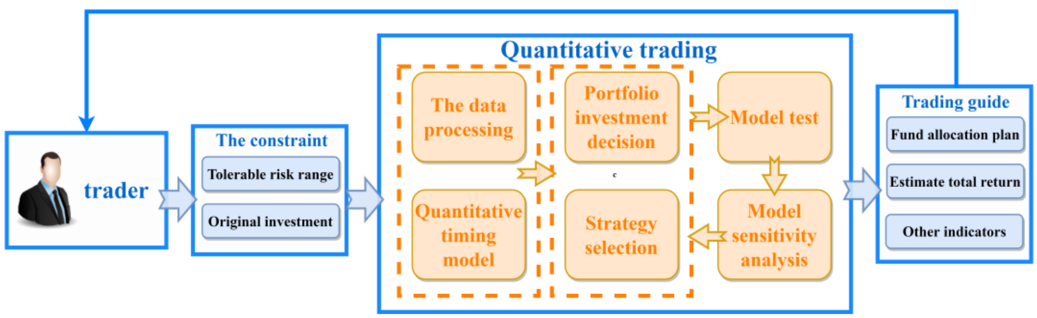


Figure : Flow chart of model

At the same time, we summarize some trading timing strategies for you. In order to comprehensively consider risk, return, market rules and other factors in portfolio investment, we established a multi-objective programming model based on Markowitz investment strategy. In this model, we aim to maximize the profit and minimize the risk of each transaction, and take the market trading rules as the restriction conditions.

In order to obtain the expectation and variance required by the model, we also built a two-layer GRU-RNN network with 260 memories to predict the future indicators. As you requested, the model we built has been tested by the simulation of geometric Brownian motion and the sensitivity test of transaction cost parameters. Its average annualized rate of return can be stably maintained in about 28%.**Achieve a change in principal from $1,000 at the beginning to $3,390 after 5 years** through a well-modeled trading strategy.

We are very grateful for this opportunity to help you understand the law of financial market development and explore the mystery of asset allocation by establishing models. We hope our model can be useful in your future transactions. Looking forward to your next contact with us. The figure shows the general framework of our model

Sincerely yours,
Team 2208251

Reference

[1] Liu Libin. *RSI index to judge price trend* [J]. Public Financial Adviser,2008(09):52-53.

[2] Yan Hongmei. *Analysis on the application of RSI in overbought and oversold type* [J]. China Securities and Futures,2013(08):260+262.

[3] Zhou X. *The application of Bren band trend breaking strategy in digital currency market* [D]. Zhejiang industry and commerce university, 2021. DOI: 10.27462 /, dc nki. GHZHC. 2021.000281.

[4] Jiang G Y. *Research on optimal parameters of MACD index -- a technical strategy of quantitative investment* [D]. Jilin University of Finance and Economics,2018.

[5] Ye Bo. *Trading Part FOUR: Gann's Rule & Bollingband Trading System* [J]. Public Financial Adviser,2008(02):70-71.

[6] Yan Zhengxu, *Qin Chao, Song Gang. Stock price Prediction based on Pearson Feature Selection in Stochastic Forest Model* [J]. Computer Engineering and Applications, 2021,57(15):286-296.

[7] Chen jia. *RNN neural network application in the stock index prediction research* [D]. Tianjin university of science and technology, 2019. The DOI: 10.27359 /, dc nki. Gtqgu. 2019.000459.

[8] Dill E A, Gru A A, Atkins K A, et al. *PD-L1 Expression and intratumoral heterogeneity across breast cancer subtypes and stages*[J]. The American journal of surgical pathology, 2017, 41(3): 334-342.

[9] Bansal T, Belanger D, McCallum A. Ask the gru: *Multi-task learning for deep text recommendations*[C]//proceedings of the 10th ACM Conference on Recommender Systems. 2016: 107-114.

[10] Michaud R O. *The Markowitz optimization enigma: Is ‘optimized’optimal?*[J]. Financial analysts journal, 1989, 45(1): 31-42.