Problem Chosen	2022 MCM/ICM Summary Sheet	Team Control Number
C		2211494

A Quantitative Investment Strategy Based on MBSA Model

Summary

Gold and Bitcoin are two volatile assets, so to maximize the initial $1,000 in assets, we need to buy and sell trades through a daily strategy. In this paper, we simulate investment scenarios based on historical data of gold prices. Without knowing the future data, we build an investment strategy model (position management model) based on data characteristics after data mining, which gives direction to the trader's subsequent investments.

Considering that traders only know the price of the day's trading day and each past trading day, we need to have a preliminary forecast of future ups and downs, as we want to quantify how much buying and selling will take place. We choose a time series model to help us measure the probability of future ups and downs. To quantify trading in many ways, we combine the common indicators currently used in the market for stock trading: MA, MACD, DIF, DEA, RSI, PE, PB, and BIAS. We weight the above indicators to get the Open Interest with the highest daily returns. Since trading in Bitcoin and Gold also requires commissions, for cost reasons, buying and selling operations cannot be performed frequently. Therefore, we need to set buy and sell thresholds for the indicators. With our model, we can conclude that $1,000 can yield $2.5 million after five years.

To prove that our model is the best strategy, we fine-tune the weights and compare the obtained returns with those of the single indicator MACD, with the result that the former returns are much larger than the latter returns.

To test the sensitivity to transaction costs, we vary the percentage of commissions, in steps of 0.01, from a lower bound of 0.01 to an upper bound of 0.11, and bring different commission percentages into the model for each calculation to derive the relationship between total assets and transaction costs (commission percentage) after 5 years.

Keywords: Quantitative Trading; Time-Series Analysis; Feature Crosses; Moving Average Convergence and Divergence

Contents

1 Introduce

1.1 Background

Market traders buy and sell volatile assets to maximize their total return. Quantitative trading is a great means to achieve this.

The so-called quantitative trading means that investors use computer technology, financial engineering modeling, and other means to make investment decisions and execute trading strategies in strict accordance with the rules set to determine the amount and price of volatile assets to be bought and sold.

With the development of computer technology and modern financial theory, quantitative trading, which realizes automatic trading of securities with the help of electronic technology, has come into being. Quantitative trading has many advantages:

- Multiply efficiency by using historical data for strategy checking,
- Capture trading opportunities in real-time across the market, dramatically improving profitability,
- Allows for more objective measurement of trading results,
- Access to profit opportunities that are difficult to find by human hands alone.

With these advantages, quantitative trading has received widespread attention from the industry since its inception in the 1970s and has grown at an alarming rate.

1.2 Problem Restatement

Considering the background information and restricted condition identified in the problem statement, we need to solve the following problem:

- Traders need to use initial capital of $1000 to develop buy and sell strategies to trade based on current and past prices of gold and bitcoin. We need a quantitative analysis of past prices to develop the strategy.
- In order to prove that the strategy is optimal, given the availability of the buy and sell strategy model, we need to compare it with the normal strategy model.
- Determine how sensitive the strategy is to different transaction costs, bring different transaction costs into the model, and see how the new results change.

1.3 Overview

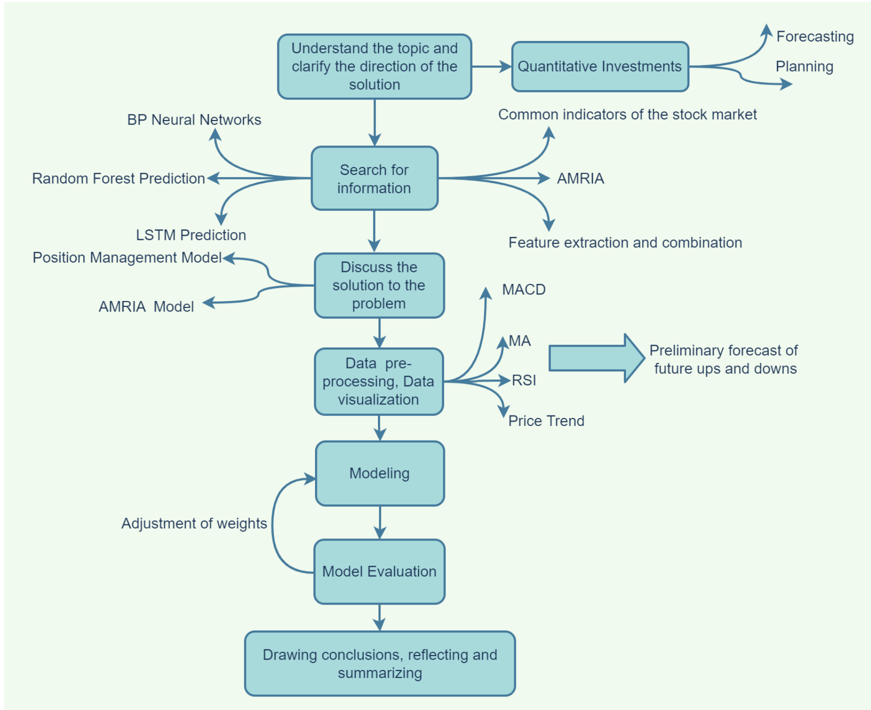


Figure 1: Overview of our work

First, we spent some time discussing and clarifying the direction of the solution and the task given by the question. We initially determined the direction of the solution - quantitative trading - after a morning of reviewing information. After that, we performed the processing of data to have a preliminary forecast of future ups and downs, which helped us to better find a suitable model. After the initial model building, we then built the scoring model that helped debug the weights, and then optimized the model step by step. Finally, we discussed the strengths and weaknesses of the model and had ideas for further improvements. The competition manager was very rewarding for us.

2 Assumptions and Justifications

The following basic assumptions are made to simplify problems:

- **The data is accurate and has not been falsified.**
- **The value of the two investment products is not affected by external subjective factors.**
- **Assume that the trader only knows the price data for the day of the trade and for the past**
- **No historical data available before 2016.**The question specifies that we cannot access and draw on other data and that the data in the given dataset all originate from after 2016, so we assume that there is no available data before 2016.
- **No economic crisis, systemic risk.**The economic crisis is a small probability event and its impact on the market is immeasurable. If the economic crisis happens suddenly, we cannot find a general pattern from it, which is not consistent with the intention of the question.
- **The two investment products conform to the universal investment law, i.e., the formula coefficients are derived from the empirical law.**We use a lot of economic and integrative indicators, such as the MACD indicator, EMA indicator, DEA indicator, etc., in order to analyze the problem better. These indicators have been constantly proven to be suitable for investment market analysis, so we assume that they are equally suitable for analyzing the two products we invest in.

Other specific assumptions, if is necessary, will be mentioned and illustrated while the model is building.

3 Notations and Glossary

3.1 Notations

Notations	definitions
P_t	price of the day
n	day number
q	rank of current day's price among all past prices
a	smoothing factor of data
EMA_t	today's exponential Moving Average
DIF_t	today's Decentralized Identity Foundation
$\alpha\%$	commission as a percentage of transaction amount
Σ_t	up rating for the day
C_t	cash share
δw	incremental update of parameter
$w_{[i-1]}$	not updated w
S_t	share of buy
P_f	single-day profit
η	learning rate

3.2 Glossary

- **MACD**:MACD, known as Moving Average Convergence and Divergence, is a technical indicator proposed by Geral Appel in 1979, which uses the convergence and divergence between the short-term (12-day) and long-term (26-day) exponential moving averages of the closing price to make a judgment on the timing of buying and selling.[1]

- **EMA**:EMA is an Exponential Moving Average, a trending indicator. It is used to determine the trend of future price movements when we want to construct an EMA as a weighted arithmetic average of the closing price. Compared to the MACD and DMA indicators, the EMA indicator is a trend analysis indicator that overcomes the price lag of the MACD indicator and, to a certain extent, eliminates the signal advance of the DMA indicator at certain times. It is a very effective indicator for analysis.[1]

- **DEA**:Data envelopment analysis (DEA) is a method in operations research and the study of economic production frontiers. The method is generally used to measure the production efficiency of some decision-making sectors.[2]

- **DIF**:The full name of DIF is Differential value, an indicator in technical analysis of the stock market, abbreviated as DIF, which is the 12-day EMA value minus the 26-day EMA value.[1]

4 Data Describe

4.1 Data Pre-processing

For data analysis problems, it is necessary to pre-process the data when faced with large amounts of raw data. We analyze the relevant conditions given in the question and the data in the two pricing data files to see that Since Bitcoin can be traded daily, there is no missing data in the BCHAIN_MKPRU.csv file and no obvious abnormal data after a simple analysis. As for the LBMA_GOLD.csv file, according to the gold trading rules provided by London Bullion Market Association, gold can only be traded on days when the market is open.

4.2 Establishment of indicators

To better quantify the trading strategy, we add several additional indicators[1] to the model: 'single-day gain', 'investment risk', 'historical quantile points', MACD indicator.

- **Single-Day Gain** :

Calculation method:Calculate it with the following formula

$$(P_t - P_{t-1})/P_{t-1} \tag{1}$$

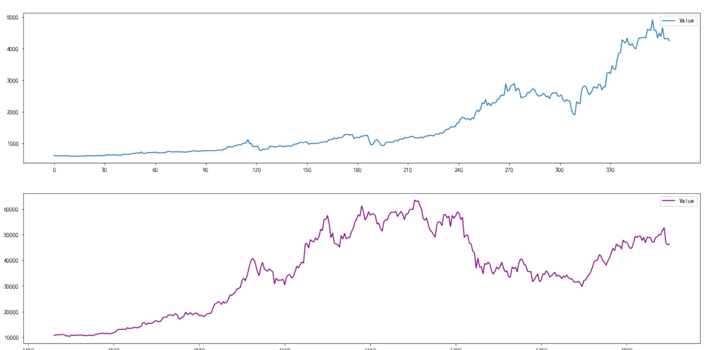


Figure 2: The price graph for the first and last year of Bitcoin

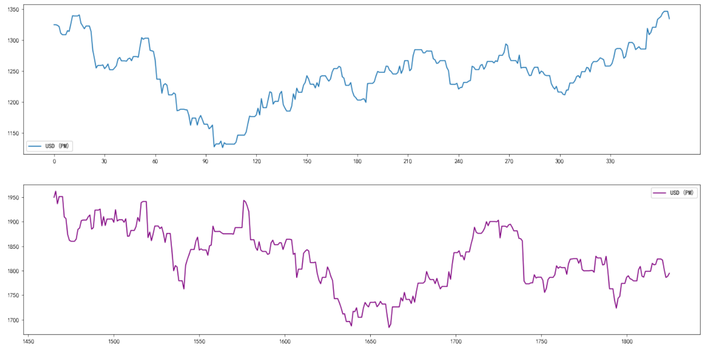


Figure 3: The price graph for the first and last year of Gold

- **Estimated quantile calculation algorithm**:

The historical quantile points are designed to show where the current valuation is in the history of all valuations and can show whether the current state is overvalued or undervalued.

Calculation method: Rank past day prices and remove duplicate items to get the day price ranking.Calculate it with the following formula

$$quantile = (q - 1) \times 100\%/(n - 1) \tag{2}$$

Overall, the lower the quantile, the lower the valuation; the higher the quantile, the higher the valuation.Since the valuation quantile of bitcoin and the valuation quantile of gold do not make a big difference in the graph, we think this indicator has less reference significance and will not use it for subsequent studies.

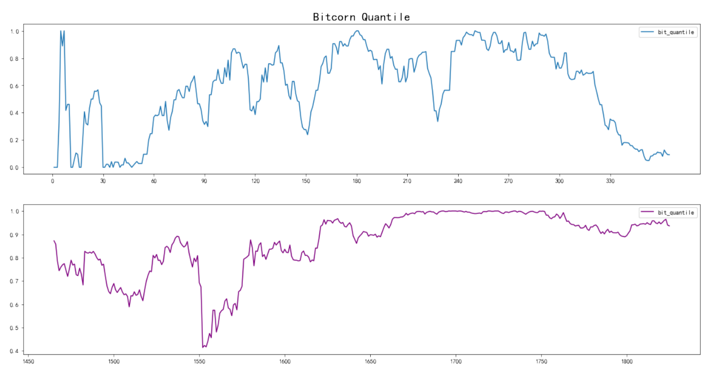


Figure 4: Valuation Quantile graph for the first and last year of Bitcoin

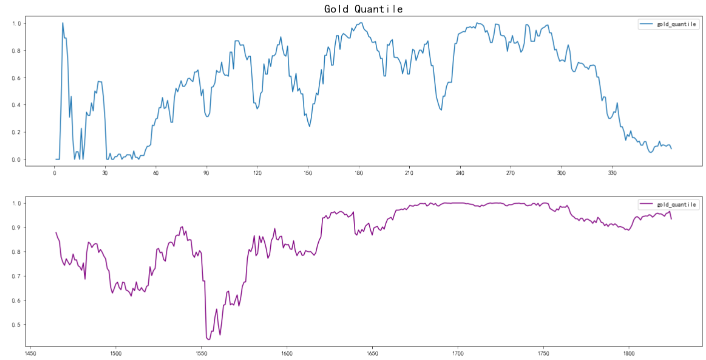


Figure 5: Valuation Quantile graph for the first and last year of Gold

- **MA indicator**:

To make the upside more informative, we need the Moving Average, the moving average, using a 5-day and 15-day average of the upside and downside.

MA uses the trading principle that if the short-term MA breaks through the long-term MA from the bottom up to form a golden cross, it is considered a time to buy, and if the long-term MA breaks through the short-term MA from the top down to form a dead cross, it is considered a time to sell, and investors can trade according to the corresponding buy and sell signals. Calculate it with the following formula

$$mat = \sum (P_t - 1 + ... + p_0)/t \tag{3}$$

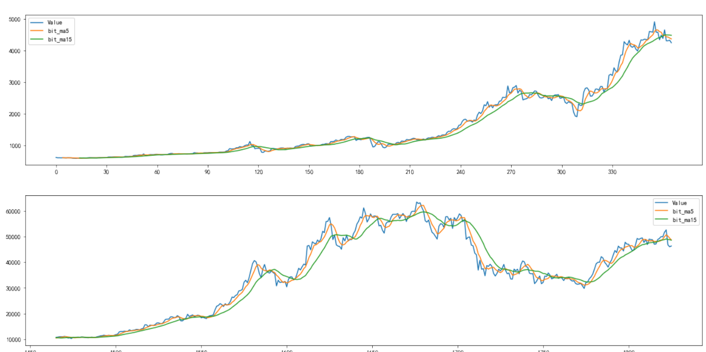


Figure 6: MA graph for the first and last year of Bitcoin

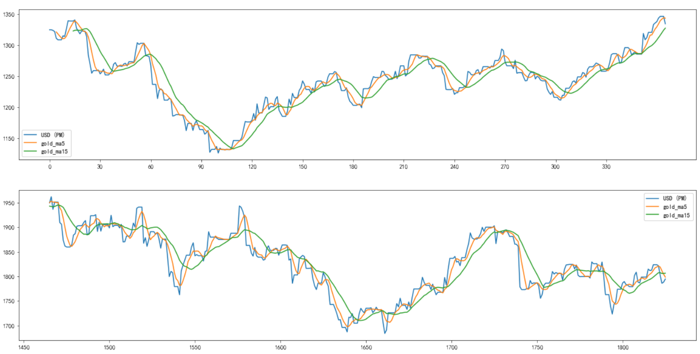


Figure 7: MA graph for the first and last year of Gold

- **MACD indicator**:

MACD indicator (Smoothed Moving Average) is derived from the double exponential moving average transformation.

MACD is obtained by subtracting the fast exponential moving average (EMA12) from the slow exponential moving average (EMA26) to get the fast line DIF, and then 2 x (fast line DIF - DIF's 9-day weighted moving average DEA) to get the MACD bar. The change in MACD represents the change in market trend, and the MACD at different K-line levels represents the buying and selling trend in the current level cycle.

The MACD has almost the same meaning as the double moving average. It characterizes the current long-short state and the possible development of the stock price trend through the dispersion and aggregation of the fast and slow averages. When MACD turns from negative to positive, it is a signal to buy. When MACD turns from positive to negative, it is a signal to sell. When the MACD changes at a large angle, it means that the gap between the fast-moving average and the slow-moving average is widening very quickly, representing a shift in the general trend of the market.

Calculation method: First, initialize EMA: usually, EMA1 is not defined, usually taken as Price1, that is, the first item of data is filled as the first item of EMA. then,Calculate it with the following formula:

$$\mathrm{EMA}_t = \alpha \times P_t + (1-\alpha) \times \mathrm{EMA}_{t-1} \tag{4}$$

$$DIF = EMA(n1) - EMA(n2) \tag{5}$$

$$DEA = (\mathrm{DEA}_{t-1} \times \alpha + \mathrm{DIF}_t \times \alpha) \tag{6}$$

$$MACD = (DIF - DEA) \times 2 \tag{7}$$

α is smoothing factor of data, which is generally taken as 2/(N+1). After our investigation, in calculating the MACD indicator for this data, we choose 12 and 26 days, i.e. n1=12, n2=26.

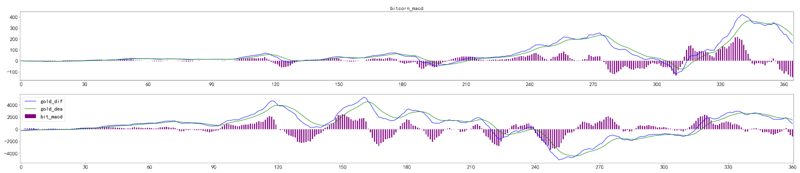


Figure 8: MACD graph for the first and last year of Bitcoin

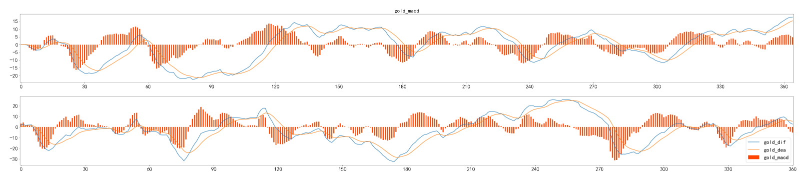


Figure 9: MACD graph for the first and last year of Gold

- **RSI indicator**:

The RSI (Relative Strength Index) was first used in futures trading to predict the direction of the market by analyzing the buying, selling, supply, and demand conditions.

Market experience shows that an RSI > 50 indicates a high probability of a higher average close and the market is called a strong market, while the opposite is called a weak market. The RSI generally fluctuates between 30 and 70. If the RSI value exceeds 80, the market is severely overflowing with buyers and is subject to extremely strong selling. This is because when the RSI reaches above 80, there is a high probability that prices will fall. If the RSI is below 20, it means that the market is heavily overflowing with sellers and the price is likely to pull back to buy. Some investors set the value at 15, 85.

Calculate it with the following formula:

$$RSI = [A \div (A + B)] \times 100\% \tag{8}$$

A is the sum of N-day closing gains (positive value) B is the sum of N-day closing losses (positive value).

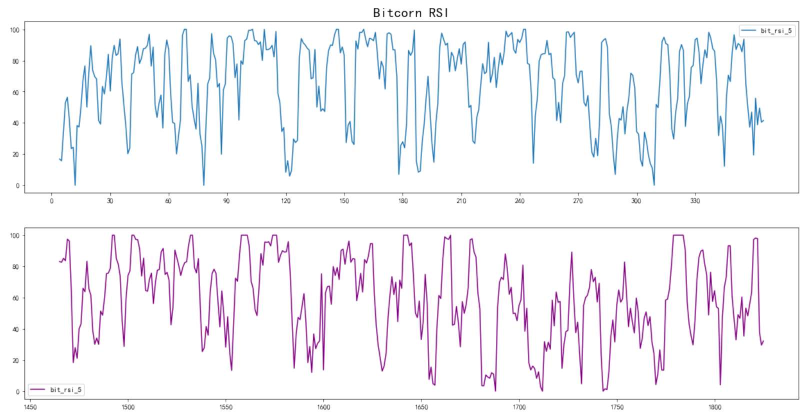


Figure 10: RSI graph for the first and last year of Bitcoin

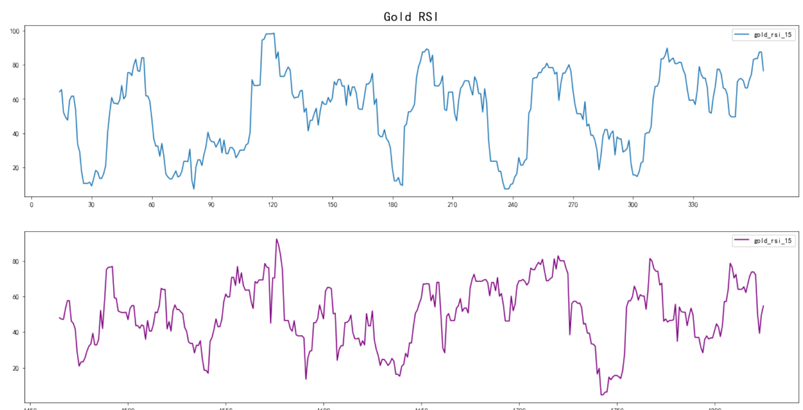


Figure 11: RSI graph for the first and last year of Gold

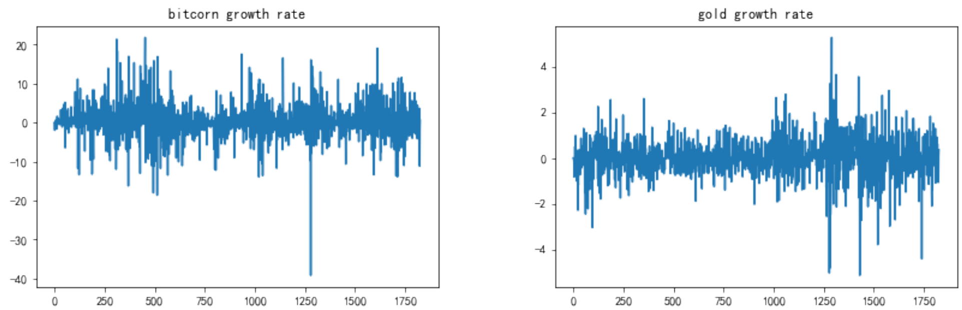
5 Models and Solutions

5.1 ARIMA Model[2]

5.1.1 Modeling

In order to obtain the likelihood of future increases or decreases, we need to forecast the future on the existing prices. By calculating the rate of increase, we roughly estimate that the rate of increase is satisfying the time series. Therefore we perform a time seriesanalysis (ARIMA) of the rise rate.

ARIMA model, the summation auto-regressive mean model, refers to the model built by transforming a non-stationary time series into a stationary time series and then regressing the dependent variable only on its lagged values and the present and lagged values of the random error term. the ARIMA model is divided into moving average process (MA),



(a) Bitcoin growth rate

(b) Gold growth rate

Figure 12: Growth Rate

auto-regressive process (AR), auto-regressive mean depending on the stationary condition of the original series, the regression moving process (ARMA), and ARIMA process.

MA uses moving average to eliminate random fluctuations in forecasts.Define a single variable timing data $\{y_t|t = 1, 2, ...\}$

$$y_t = c + \omega_t + \beta_1\omega_{t-1} + ... + \beta_p\omega_{t-p} = \omega_t + \sum_{t=1}^{p} \beta_i\omega_{t-i} \tag{9}$$

ω is white noise.

The auto-regression process is defined as follows:

$$y_t = \mu + \varepsilon_t + sum_{i=1}^{q}\theta_i\varepsilon_{t-i} \tag{10}$$

ARIMA(p,d,q)model equation is expressed as

$$\hat{y}_t = a_0 + \sum_{i=1}^{p} a_i y_{t-i} + \varepsilon_t + \sum_{t-i}^{q} \beta_i\varepsilon_{t-i} \tag{11}$$

and

$$y_t = \triangle^d y_t = (1 + L)^d y_t \tag{12}$$

$$(1 - \sum_{i-1}^{p} a_i L_i)(1 - L)^d y_t = a_0 + (1 + \sum_{i=1}^{q} \beta_i L^i)\varepsilon_t \tag{13}$$

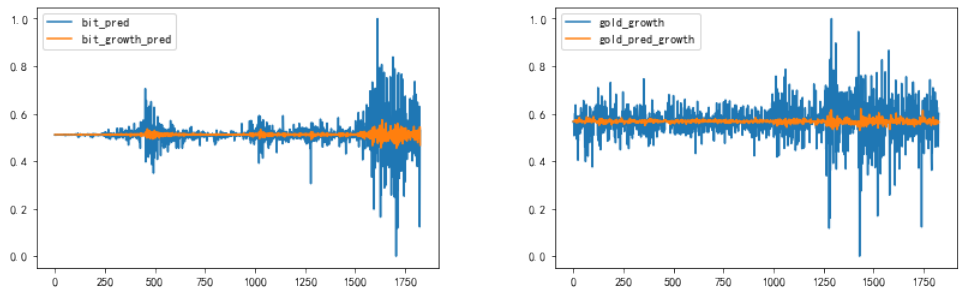
p denotes the number of lags of the time-series data taken in the prediction model, d denotes the different order of the time-series data, and q denotes the number of lags of the prediction error used in the prediction model.

We use the exhaustive method of software SPSS, we get the three parameters p, d-order difference, q of the optimal time series, and the results obtained are as follows:

The predicted increase is obtained as shown in the graph:

	p	q	r
Gold	2	0	6
Bitcoin	2	0	10

Table 1: The three parameter values of the optimal time series



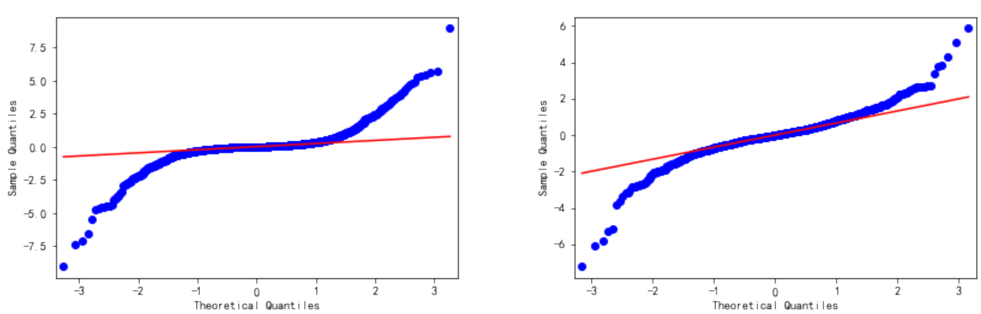
(a) Diagram of expected bitcoin rise (b) Diagram of expected gold rise

Figure 13: The predicted increase

5.1.2 Model Testing

- **Residual QQ-plot test**

We use the QQ-plot of the residuals of the model to check whether the ARIMA model has learned the appropriate features. If the residuals are white noise series, it means that the useful information in the time series has been extracted, and all that remains is random perturbation, which is impossible to predict and use. In this QQ-plot, if it is normally distributed, then it is a straight line (red line), and obviously the result is roughly consistent with white noise.



(a) Bitcoin QQ-plot (b) Gold QQ-plot

Figure 14: QQ-plot

- **DW test**

The Durbin-Watson test, also known as the DW test, is used to test the first-order auto-correlation of the residuals in regression analysis (especially for time series data). The closer the value of this statistic is to 2, the better, generally between 1 and 3 means that there is no problem, and less than 1 means that the residuals are auto-correlated.

We tested the results and obtained the following data:

Table 2

	Score
Gold	2.0382
Bitcoin	2.0348

5.2 Multivariate Buying and Selling Evaluation Model(MBSE Model)

5.2.1 Modeling

We now get the expected rise. The other MACD, MA, RSI, 5 day/15 day SMA BIAS are x_1,x_2, x_3, x_4. Since gold is a long term trade and bitcoin is a short term trade, we will update the weights every 15 days for gold and every 5 days for bitcoin. This avoids the problem of frequent trading. Using the weighting algorithm Calculate the price for the day (the first day requires an initialization of ω):

$$\Sigma_t = \sigma(w^T x_{t-1} + b) \tag{14}$$

$$\sigma(t) = \frac{1}{1 + e^{-x}} \tag{15}$$

In order to make the upside and downside evaluation metrics computed as a nonlinear mapping and as a convex function, we choose the sigmoid function.

In order to maximize the return within n days, a dynamic programming model needs to be simulated and we use a back propagation idea to update the weights to maximize the return. We make a segmentation judgment on the scoring, with a sell below 30% and a buy above 70%. This prevents buying and selling too frequently and simulates the process of trader estimation. Calculate single day earnings:

- $M_{t-1} > 0.7$

$$S_t = M_t \times C_{t-1} \times (1 - \alpha\%)/P_{t+1} + S_{t-1} \tag{16}$$

$$C_t = C_{t-1} - C_{t-1} \times M_{t-1} \times (1 - \alpha\%) \times P_{t-1} \tag{17}$$

- $M_{t-1} < 0.3$

$$S_t = S_{t-1} + C_{t-1} \times M_t \times (1 - \alpha\%)/P_{t+1} \tag{18}$$

$$C_t = C_{t-1} + C_{t-1} \times M_t \times (1 - \alpha\%) \times P_{t-1} \tag{19}$$

$$P_f(t) = S_t + C_t \tag{20}$$

The maximum value of the return within n days can be expressed as

$$L(w, b) = argmax \sum_{1}^{n} P_f(t) \tag{21}$$

To get the maximum value, we need to keep updating w. Next, the maximum value requires us to update w, using back propagation and gradient descent.

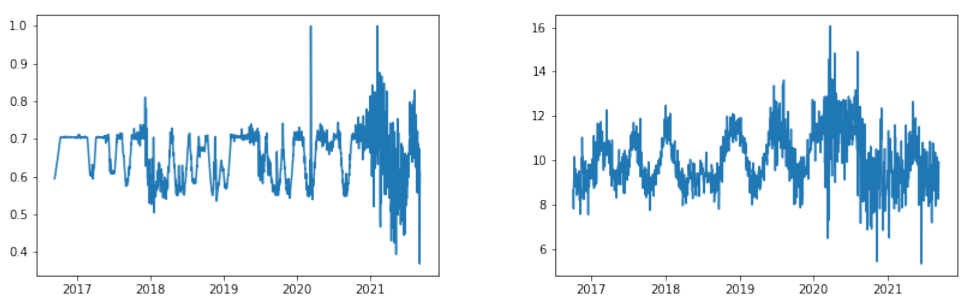
$$\delta w = \frac{\partial(w, b)}{\partial w} \tag{22}$$

$$\delta b = \frac{\partial(w, b)}{\partial b} \tag{23}$$

$$w_{[i]} = w_{[i-1]} - \eta \delta w \tag{24}$$

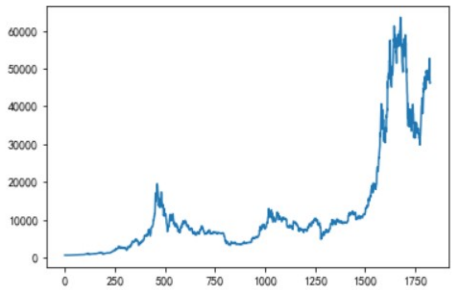
$$b_{[i]} = b_{[i-1]} - \eta \delta b \tag{25}$$

After solving, the value is brought into tomorrow's w as the initialization value of tomorrow's w.Model conclusion: We obtained the buy and sell evaluation indicators for each trading day for all five years, as shown in the figure.

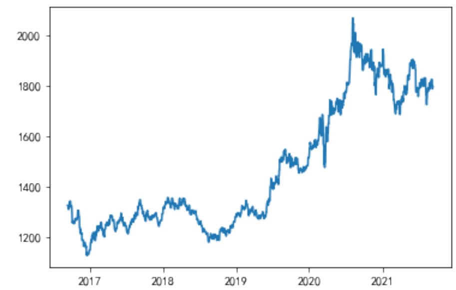


(a) The rise scoring graph of Bitcoin (b) The rise scoring graph of Bitcoin

Figure 15: Rise Scoring



(a) The price trend graph of Bitcoin



(b) The price trend graph of Bitcoin

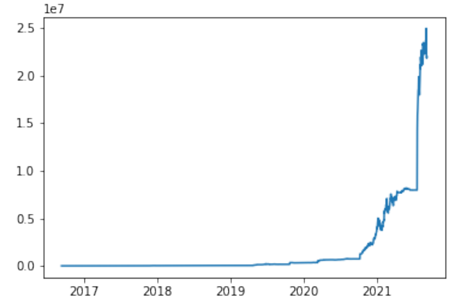
Figure 16: Price Trend Graph

5.2.2 Rational Analysis

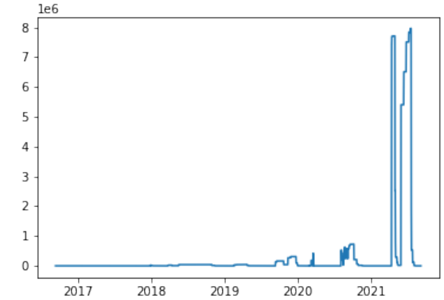
To justify the buy and sell evaluation indicators calculated by the MBSE model, we can observe that the change in volatility between the two plots should be consistent with the trading logic.When the price is at a low point, you should buy and the rating becomes higher; when the price is at a high point, you should sell and the rating becomes lower.

5.2.3 Modeling Results

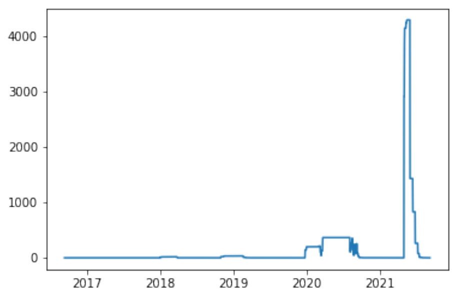
Take the solved result into the daily return calculation formulaeq. (20).Substitute the formulaeq. (16)eq. (17)eq. (18)eq. (19) to calculate the cash share, bitcoin share, and gold share for each trading day, and the total assets are their shares multiplied by their respective net values to get the result as shown in the chart.



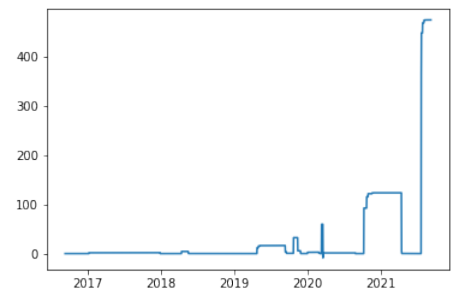
(a) Total assets trend



(b) Cash holding share trend



(c) Gold holding share trend



(d) Bitcoin holding share trend

Figure 17

6 Sensitivity Analysis and Optimality Analysis

6.1 Optimality Analysis

To demonstrate the optimality of our model, we used the more commonly used indicator macd for single-indicator strategy trading and obtained the following results for total assets: [h!] You can see that the total assets are far behind our model.

MACD model final gain approximated $60,000, up 59%. Our model $1000 final return is $24992548.649, up 24991%, a full 24932 points more than the MACD model.

At the same time, the stability of returns is essential for the best models, and our models are stable in volatile markets. Below is a diagram of the range indicators for the 2, 5 and 15 days of the test.

It can be seen that the rate of change is basically above the x-axis, indicating that the total production is more stable showing an upward trend, and the number of falls and the amount of falls is small.

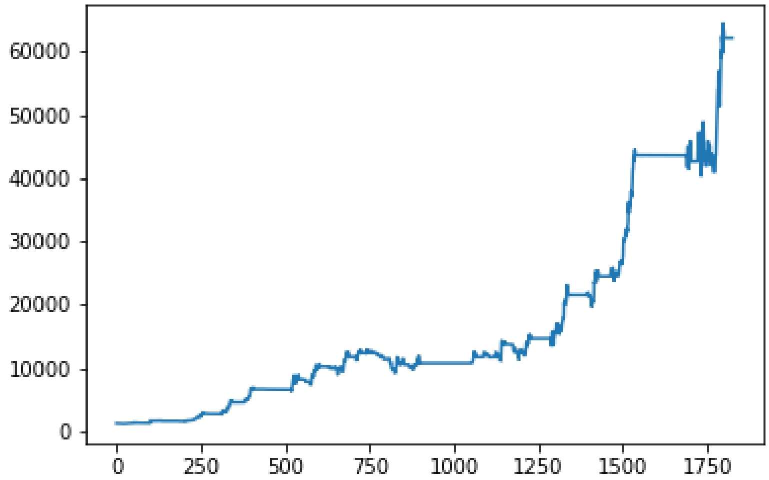


Figure 18: Total assets trend based on MACD indicator decision generation

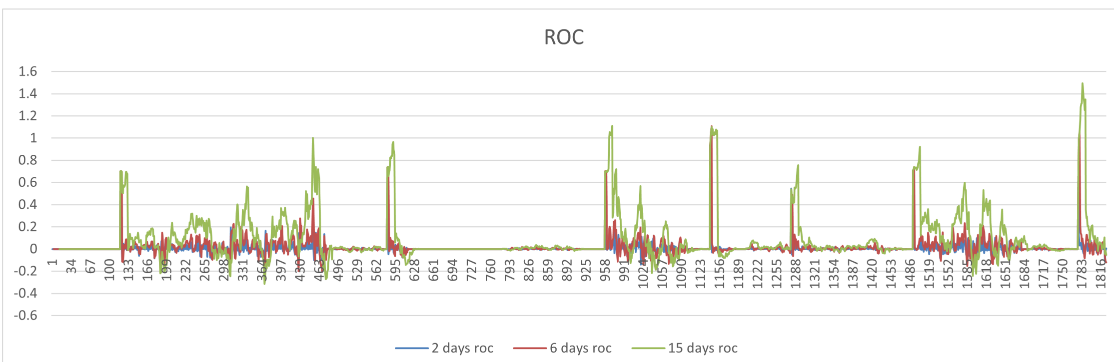


Figure 19: The range indicators for the 2, 5 and 15 days of the test

6.2 Sensitivity Analysis

We analyze the model by bringing different costs from 0.01 to 0.11 into the model and then calculate the total assets to obtain the following results.

As can be seen from the graph, the magnitude of the oscillation is very pronounced as the transaction cost rises, indicating that the total assets are and are sensitive to the transaction cost. At the same time the returns are gradually decreasing. Therefore the trader needs to re-use the increase evaluation model for the update of the indicator weights.

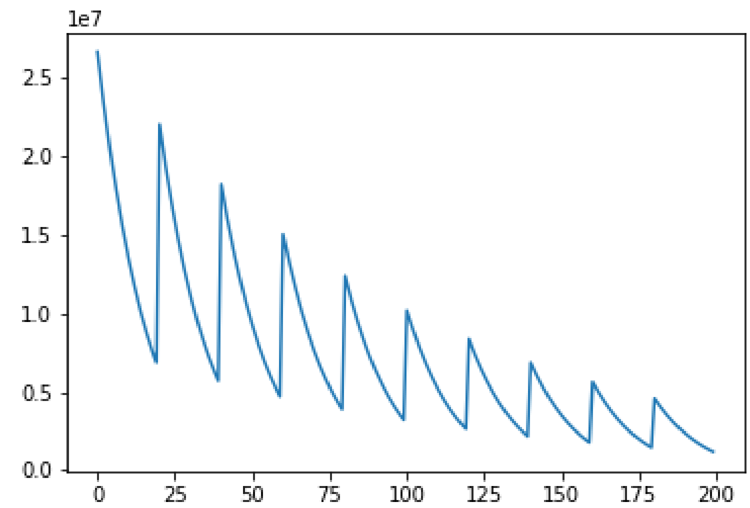


Figure 20: Schematic diagram of the change in total assets under different transaction costs

7 Strengths and weaknesses

7.1 Strengths

- The quantitative results are analyzed scientifically to be able to have stable returns under the price market of bitcoin and gold.
- It makes good use of the features of the past price and extracts the features from multiple angles to reduce information loss.

7.2 Weakness

- The number of reference indicators is large, which leads to a large amount of calculation and consumes more storage space.
- The model is highly sensitive to changes in transaction costs, and when transaction costs change, the parameters need to be updated and recalculated

7.3 Improvements

- It is possible to perform a PCA reduction method on the metrics to reduce the metrics and thus reduce a certain amount of computation.
- In terms of weight update, the greedy algorithm can be used to calculate the weights and reduce the computation time.
- We can use a deeper network for prediction of the increase to achieve more stable and accurate results.

8 Reference

References

[1] Shi, Sai-Nan. Testing the validity of MACD indicators in stock technical analysis[D]. Southwest University of Finance and Economics, 2011.

[2] Mayer A.; Zelenyuk V. (2014). "Aggregation of Malmquist productivity indexes allowing for reallocation of resources". European Journal of Operational Research. 238 (3): 774–785. doi:10.1016/j.ejor.2014.04.003

[3] Research on quantitative trading investment strategy of gold futures Liu Dongsheng.

[4] Research on the application of quantitative trading strategy models Li Y.

[5] A Study on the Effectiveness of New Momentum Trading Strategies in the A-Share Market Mingtao Wang,Shan Li.

[6] Research on High Frequency Data Statistical Arbitrage Strategy in Bitcoin Market Mei Yan.

9 Memo

MEMORANDUM

To: Mr.Traders

From: Team 2211494

Subject: Trading Strategies

Date: Monday,February21, 2022

Our team developed a model based on ARIMA and based on the data up to the current day, as per your request, to predict the increase by using the above model. Then we use the indicators often used in quantitative trading weights, MACD, MA, BIAS, historical quantile points, etc., to weight the indicators to get the evaluation. The weighting method is updated by gradient descent parameters to constantly approach the maximum rate of return in the short term. The weights are updated periodically. Ultimately our model gives more appropriate investment options based on the data up to that date, and we end up with a model return of $25 million.

Data processing: Since gold can only be traded on days when the market is open, we combine the gold and bitcoin data in a single process

Indicator calculation: A variety of common investment indicators recognized in the market are calculated and combined with the predicted values to get the investment evaluation indicators for bitcoin and gold respectively. The buying and selling waves of bitcoin and gold are controlled by the ratio between the indicators.

Investment Strategy: For the investment strategy, position changes are made for both using the investment valuation indicators for both, with sell trades made for less than 0.3 and buy trades made for greater than 0.7. This strategy is more profitable than the current more common single MACD indicator strategy 24932 percentage points

The above is a summary of our research. We hope it will provide you with useful information.

Appendices

	Average
Average R^2	.013
R^2	.013
RMSE	.062
MAPE	6.880
MaxAPE	365.955
MAE	.035
MaxAE	.529
BIC	-5.549

(a) Gold ARIMA model statistics

	Average
Average R^2	.022
R^2	.022
RMSE	.053
MAPE	5.255
MaxAPE	328.462
MAE	.024
MaxAE	.505
BIC	-5.868

(b) Bitcoin ARIMA model statistics

Figure 21: Rise Scoring

Change data of total assets at different costs (part)
26672484.77607509,
24992548.649456535,
23404222.090357,
21903290.992781986,
20485698.908093967,
19147542.50437448,
17885067.117674176,
16694662.39406001,
15572858.021371236,
14516319.549603142,
13521844.298839578,

Figure 22: Change data of total assets at different costs(part)

Date	Gold Buy Rate	Bitcorn Buy Rate
2016/10/10	0. 400035802	0. 705307206
2016/10/11	0. 361452094	0. 70556182
2016/10/12	0. 400450536	0. 704616597
2016/10/13	0. 39452343	0. 704163551
2016/10/14	0. 323146017	0. 7047415
2016/10/15	0. 367930072	0. 704927316
2016/10/16	0. 359939636	0. 705158739
2016/10/17	0. 364246853	0. 705312014
2016/10/18	0. 353411996	0. 705610318
2016/10/19	0. 378774521	0. 706045943
2016/10/20	0. 328480587	0. 706234038
2016/10/21	0. 288583076	0. 70576011
2016/10/22	0. 320141641	0. 704364217
2016/10/23	0. 319855109	0. 703922186
2016/10/24	0. 31776056	0. 704630363
2016/10/25	0. 338722657	0. 704967081
2016/10/26	0. 323984401	0. 705556141
2016/10/27	0. 299626681	0. 704281518
2016/10/28	0. 351222742	0. 703448569
2016/10/29	0. 313960414	0. 704079955
2016/10/30	0. 315259765	0. 704271252

Figure 23: Partial data prediction of ARIMA