

Research on a quantitative trading strategy based on high-frequency trading

Chang Liu[#], Huilin Shan^{#,*}, Zhihao Tian[#], Hangyu Cheng

School of Statistics and Mathematics, Shandong University of Finance and Economics, Jinan, Shandong

*Corresponding author: 20191644110@mail.sdufe.edu.cn

[#]These authors contributed equally.

Abstract. This article aims to develop a quantitative trading strategy that maximizes profits while finding the best balance of risk and return. We built a high-frequency trading strategy model to maximize profits. We first used the Apriori algorithm to find frequency item sets in historical data before fitting the best daily dynamic position adjustment functions for gold and bitcoin using mathematical statistics and other methods based on price movements. Then we can trade to increase and decrease positions in gold and bitcoin based on the positions suggested by the dynamic position adjustment function. We also simulated three investors with different risk preferences trading using this high-frequency trading model for up to five years and obtained return of 266.05 %, 152.51 %, and 33.29 %, respectively.

Keywords: Quantitative Trading; High-Frequency Trading Strategy; Bitcoin; Gold.

1. Introduction

Quantitative trading is the use of advanced mathematical models to simulate the laws of financial markets, as well as computer technology to simulate historical market data [1], in order to develop strategies by selecting a variety of "high probability" events that can bring excess returns [2]. Initially proposed by Simmons, quantitative investing took off in the 1970s, particularly after introducing complex financial instruments such as options, where quantitative trading played an indispensable role and then grew rapidly in the 1990s[3]. In practice, investors analyze and build mathematical models based on historical market data, make predictions using programs and techniques such as artificial intelligence, and constantly optimize quantitative strategies [4]. The method effectively assists traders in making more accurate trading decisions and achieving higher-than-average returns [5-6].

As a high-value stand-alone resource [7], gold has always been regarded as the best tool for hedging risk in the investment market as a high-value stand-alone resource. Compared to gold [8], bitcoin has a higher yield, higher volatility, is free of regulation and taxation, and has more room for growth in the financial sector. It has the potential to replace gold as an inflation hedge, supplement gold as a hedging asset, and create a new type of hedging asset. As a result, research on bitcoin and gold is an essential parameter [9].

2. Preparation of Trading Model

In reality, the future price series is challenging to predict. The common method is to find the summary of long-term price changes from many historical price data. Future stock price changes are assumed to obey a particular distribution, and the optimal trading strategies are found based on this probability distribution [10].

For subsequent calculations and analysis, the calculation is performed using the rate of return, as shown below:

$$p_{t,i} = \frac{p_{t+i} - p_t}{p_t} \quad (i = 1, 2 \dots) \quad (1)$$

Where p_{t+i} represents the price at the moment $t+i$. p_t represents the price at the moment t . $p_{t,i}$ represents the rate of price change from day $t+i$ to day t , which is the value of the rate of increase or decrease.

It is defined $R_n = \{r_{n1}, r_{n2}, \dots\}$ as the distribution of $p_{t,i}$. When it is rising for consecutive n day and $D_n = \{r_{n1}, r_{n2}, \dots\}$ as the distribution of $p_{t,i}$, when it is falling for a consecutive n day distribution, which means that $R_n = \{r_{n1}, r_{n2}, \dots\}$ and $D_1 \{d_{11}, d_{12}, \dots\}$ it is the distribution of single-day increase and single-day decrease, it is denoted by the $X\%$ quantile of R_1 as G_R and the $X\%$ quantile of D_1 as G_D . Definition of the symbols of G and B is shown in Table 1.

Table 1. Definition of the symbols of G, B

θ	Definition
G_R	The $X_1\%$ quantile of the single-day increase R_1 in gold
G_D	The $X_2\%$ quantile of the single-day decrease D_1 in gold
B_R	The $X_3\%$ quantile of the single-day increase R_1 in Bitcoin
B_D	The $X_4\%$ quantile of the single-day decrease D_1 in Bitcoin

Denote the quantile corresponding to the distribution of consecutive increase and decrease days during the trading period μ and give the following definition. The definition of the symbol of μ is shown in Table 2.

Table 2. Definition of the symbol of μ

μ	Definition
$\mu_{G,R}$	The $X_5\%$ quantile of the distribution in consecutive days of increase in gold
$\mu_{G,D}$	The $X_6\%$ quantile of the distribution in consecutive days of decrease in gold
$\mu_{B,R}$	The $X_7\%$ quantile of the distribution in consecutive days of increase in Bitcoin
$\mu_{B,D}$	The $X_8\%$ quantile of the distribution in consecutive days of decrease in Bitcoin

n indicates the maximum number of consecutive increases or decreases in the number of days. N_i indicates the number of consecutive increases or decreases in the number of i days. The number of consecutive increase or decrease days is accumulated by size until the cumulative number of days at a day. μ reaches the $X\%$ quantile of the total increase or decrease days, which μ satisfies the following formula.

$$\left\{ \begin{array}{l} \frac{\sum_{i=2}^{\mu-1} N_i}{\sum_{i=2}^n N_i} < 90\% \\ \frac{\sum_{i=2}^{\mu} N_i}{\sum_{i=2}^n N_i} \geq 90\% \end{array} \right. \quad (2)$$

Denote the quantile of the distribution of the increase and decrease $R_\mu = \{r_{\mu 1}, r_{\mu 2}, \dots\}$ when the trade period is the consecutive increase or decrease μ days by M and give the following definition. The definition of symbol of M is shown in Table 3.

Table 3. Definition of the symbol of M

M	Definition
$M_{G,R}$	The $X_9\%$ quantile of the R_μ distribution of gold
$M_{G,D}$	The $X_{10}\%$ quantile of the D_μ distribution of gold
$M_{B,R}$	The $X_{11}\%$ quantile of the R_μ distribution of Bitcoin
$M_{B,D}$	The $X_{12}\%$ quantile of the D_μ distribution of Bitcoin

3. Construction of the Trading Model

3.1 Apriori Algorithm

The Apriori algorithm is one of the most influential algorithms for finding frequent items in data mining, which uses an iterative method of layer-by-layer search to calculate the number of occurrences and support of the itemset in the target transaction base and compare it with predefined minimum support. If the support of this itemset is not less than the minimum support min_sup , then this itemset is a frequent item. Support defines the proportion of items in the entire database. If $s\%$ of transactions in the target transaction base contains both items sets A and B , we call $s\%$ the support of $A \Rightarrow B$, denoted as:

$$\text{Support}(A \Rightarrow B) = P(A \cup B) = s\% \tag{3}$$

Where $P(A \cup B)$ denotes the probability that the transaction contains terms A and B .

3.2 The Solution of Frequent Item Sets

In order to describe it efficiently, this article discusses the purchase of gold as an example.

Traders open positions when the daily increase is greater than the cost of trading to make a profit on the trade, taking into account the cost of trading. We performed a statistical analysis of the attached price data, and the results revealed that when the price of gold increases by more than 81% of the quantile of its distribution in a single day to cover the transaction costs and for traders to make a profit. Therefore we defined $X_1\% = 90\%$ and calculated that $G_R = 1.35\%$.

We believe that the price of gold decreased more than u days continuously for the occurrence of events with low probability, which means that there is no less than 90% certainty that the number of days of consecutive decrease in gold will not exceed $\mu_{G,D}$ days. Where $\mu_{G,D}$ is the set of frequent terms satisfying the following equation:

$$\begin{cases} \frac{\sum_{i=2}^{\mu_{G,D}-1} N_i}{10} < 90\% \\ \frac{\sum_{i=2}^{\mu_{G,D}} N_i}{10} \geq 90\% \end{cases} \tag{4}$$

Where N_i denotes the number of days corresponding to the number of consecutive days of decrease i , and n denotes the maximum number of consecutive days of decrease. We use the Apriori algorithm for the solution and calculate the result as $n = 10$, $\mu_{G,R} = \mu_{G,D} = 5$

$M_{G,R}$ and $M_{G,B}$ indicate the quantile corresponding to the distribution of the increase and decrease in the price of gold when the price increases or decreases for μ consecutive days. We believe that it is less likely to happen when the price of gold continuously decreases for μ days and still decreases by more than $X_{10}\%$. Similarly, it is less likely to happen when the price of gold continuously increases for μ days and still increases by more than $1 - X_{10}\%$. The definition $X_{10} = 10\%$ means that there is more than 90% certainty that the price of gold will not decrease by more than $M_{G,R}$ when it decreases for 5 consecutive days. We use the Apriori algorithm for the solution and calculate the result $M_{G,R} \approx M_{G,D} = 2.42\%$.

3.3 Construction of The Dynamic Position Adjustment Functions

$P_i (i = 1, 2 \dots \mu_{G,R})$ represents the gold position established on the day i . It is assumed a scenario where gold is on a continuous decrease, and it is needed to keep adding to the position during the continuous decrease. Then, sell some gold for profit once there is a bounce up. Constructing a position adjustment function can achieve stable profits in a constant upward and downward movement.

The average daily decline in a continuous decline of μ days is $\frac{M_{G,D}}{\mu_{G,D}}$. The initial capital is $P_{initial}$. The transaction cost $\alpha_{gold} = 1\%$. The rebound after the end of the decrease is G_R .

Day 1: Buy gold in the amount of P_1 .

Day 2: Gold prices fell. The decline was $\frac{M_{G,D}}{\mu_{G,D}}$. Buy gold in the amount of P_2 to reduce the current cost. In order to ensure that there is no loss when the gold price rebounds higher the following day P_1 and P_2 should satisfy the equation as follows:

$$P_1(1 - \alpha_{gold}) \left(1 + \frac{M_{G,D}}{\mu_{G,D}}\right) G_R + P_2(1 - \alpha_{gold}) G_R + [P_1(1 - \alpha_{gold}) \left(1 + \frac{M_{G,D}}{\mu_{G,D}}\right) - P_1] + [P_2(1 - \alpha_{gold}) - P_2] = 0 \tag{5}$$

Day 3: Gold prices fell for two consecutive days, with cumulative losses $\frac{2M_{G,D}}{\mu_{G,D}}$. Continue to buy gold in the amount P_3 to reduce the current cost. In order to ensure that there is no loss when the gold price rebounds higher the following day P_1, P_2 and P_3 should satisfy the equation as follows:

$$P_1(1 - \alpha_{gold}) \left(1 + \frac{M_{G,D}}{\mu_{G,D}}\right)^2 G_R + P_2(1 - \alpha_{gold}) \left(1 + \frac{M_{G,D}}{\mu_{G,D}}\right) G_R + P_3(1 - \alpha_{gold}) G_R + [P_1(1 - \alpha_{gold}) \left(1 + \frac{M_{G,D}}{\mu_{G,D}}\right)^2 - P_1] + [P_2(1 - \alpha_{gold}) \left(1 + \frac{M_{G,D}}{\mu_{G,D}}\right) - P_2] + [P_3(1 - \alpha_{gold}) - P_3] = 0 \tag{6}$$

A recursive formula can be established:

$$(1 - \alpha_{gold}) (1 + G_R) \sum_{i=1}^n P_i \left(1 + \frac{M_{G,D}}{\mu_{G,D}}\right)^{n-i} - \sum_{i=1}^n P_i = 0, \quad n = 1, 2, \dots, 5 \quad (7)$$

The above solution by the Apriori algorithm gives a=5, so we have no less than 90% certainty that gold will not decrease for more than 5 consecutive days. Therefore, when the price of gold keeps decreasing for 5 consecutive days, the initial capital should be fully invested at this time, which means that the following equation is satisfied.

$$\sum_{i=1}^5 P_i = P_{initial} \quad (8)$$

Combined with Formula (10) above, the following system of equations can be obtained.

$$\begin{cases} \sum_{i=1}^5 P_i = P_{initial} \\ (1 - \alpha_{gold}) (1 + G_R) \sum_{i=1}^n P_i \left(1 + \frac{M_{G,D}}{\mu_{G,D}}\right)^{n-i} - \sum_{i=1}^n P_i = 0, \quad n = 1, 2 \dots 5 \end{cases} \quad (9)$$

Solving the above system of equations, we can obtain the amount of each position increasing when the price of gold falls $P_i (i = 1, 2 \dots \mu_{G,R})$.

Similarly, assume a scenario where the price of gold rises continuously, and when the price of gold rises continuously, it is needed to keep selling to reduce the position for profit to ensure that you can not lose money when the price of gold decreases the next day. Due to space limitations, the derivation of the gold price continuous rise scenario will not be performed, and the set of position control equations for when the gold and bitcoin prices are continuously increasing and decreasing will be given directly.

(1) When the price of gold decreases:

$$\begin{cases} \sum_{i=1}^{\mu_{G,D}} P_i = P_{initial} \\ (1 - \alpha_{gold}) (G_R + 1) \sum_{i=1}^n \left(1 + \frac{M_{G,D}}{\mu_{G,D}}\right)^{n-i} P_i - \sum_{i=1}^n P_i = 0, \quad n = 1, 2 \dots \mu_{G,D} \end{cases} \quad (10)$$

(2) When the price of gold increases:

$$\begin{cases} \sum_{i=1}^{\mu_{G,R}} P_i = P_{initial} \\ (1 - \alpha_{gold}) (G_D + 1) \sum_{i=1}^n \left(1 + \frac{M_{G,R}}{\mu_{G,R}}\right)^{n-i} P_i - \sum_{i=1}^n P_i = 0, \quad n = 1, 2 \dots \mu_{G,R} \end{cases} \quad (11)$$

(3) When the price of Bitcoin decreases:

$$\begin{cases} \sum_{i=1}^{\mu_{B,D}} P_i = P_{initial} \\ (1 - \alpha_{bitcoin}) (B_R + 1) \sum_{i=1}^n \left(1 + \frac{M_{B,D}}{\mu_{B,D}}\right)^{n-i} P_i - \sum_{i=1}^n P_i = 0, n = 1, 2 \dots \mu_{B,D} \end{cases} \quad (12)$$

(4) When the price of Bitcoin increases:

$$\begin{cases} \sum_{i=1}^{\mu_{B,R}} P_i = P_{initial} \\ (1 - \alpha_{bitcoin}) (B_D + 1) \sum_{i=1}^n \left(1 + \frac{M_{B,R}}{\mu_{B,R}}\right)^{n-i} P_i - \sum_{i=1}^n P_i = 0, n = 1, 2 \dots \mu_{B,R} \end{cases} \quad (13)$$

Because each time to calculate the number of adjusted positions needs to solve the system of equations to facilitate the calculation, this paper uses the exponential function to fit the increase or decrease and the number of adjusted positions. It is used Lingo software to solve the system of equations (10). The solution results are shown in the following Table 4.

Table 4. The result of solving the system of equations

<i>n</i>	1	2	3	4	5
<i>P_i</i>	65.3298	18.4334	107.0378	243.8172	555.3816

n is the number of consecutive declines. The following function was obtained by fitting the data in Table 4 with MATLAB.

$$y = 17.06 \cdot e^{0.8118 \cdot n} \quad (14)$$

Where $R^2 = 0.9987$ $\bar{R} = 0.9983$, and it is indicated that the exponential function is a perfect fit to the observed values.

As a result, it is possible to obtain the relationship between the increase in gold positions as a function of the magnitude of the decrease.

$$f(w)_{plus} = 17.06e^{-232.5581\omega} \quad (15)$$

where *w* is the magnitude of the decline, and $w = n \cdot \frac{M_{G,D}}{\mu_{G,D}}$

Similarly, the position reduction in gold can be obtained as a function of the magnitude of the increase.

$$f(w)_{minus} = 17.06e^{206.6116\omega} \quad (16)$$

The functions for adding and reducing positions in Bitcoin are as follows:

$$\begin{aligned} f(w)_{plus} &= 8.405e^{-54.9450\omega} \\ f(w)_{minus} &= 8.405e^{38.3436\omega} \end{aligned} \quad (17)$$

The figure 1 below shows the fitted images of the additive function for gold and Bitcoin.

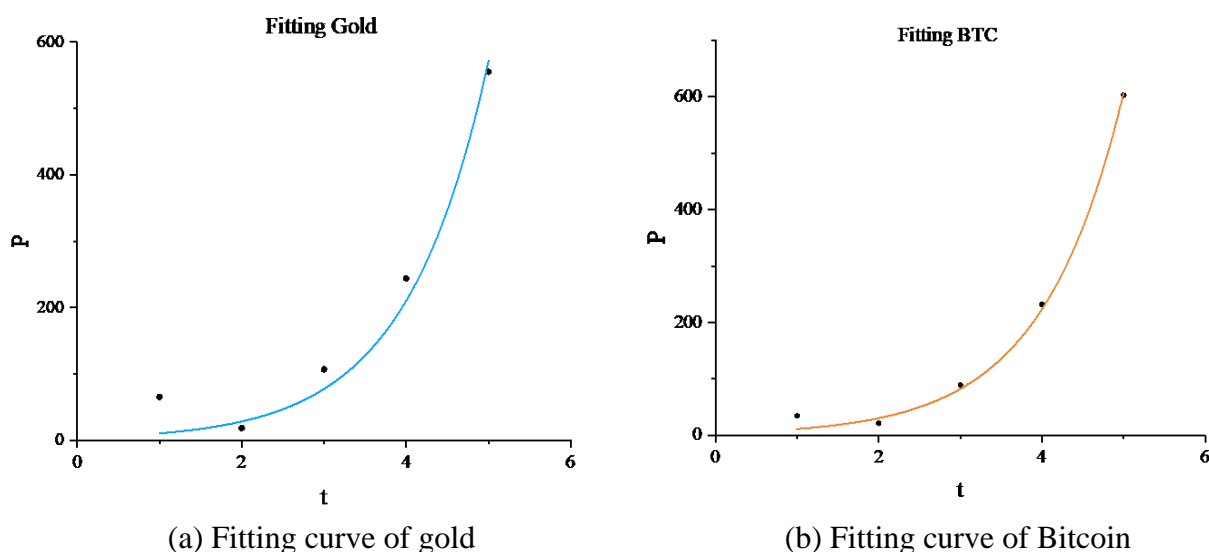


Figure 1. Fitting Curve

3.4 Construction of Trading Strategy

Based on the addition and reduction functions derived above, and using the price three minutes before the end of the trading day, it can calculate the increase or decrease of the price for the day and run the strategy model to complete the trade for the day three minutes before the end of the trading day.

- If the price decreases today, the amount of w is the price decrease, which is substituted into Formula (15) to find the increase of the position today.
- If the price is still decreased on the next day, then w is the decrease of two consecutive days, and the amount of the position increase can be calculated by substituting into Formula (15).
- If the price decreases for three consecutive days, then w is the drop for three consecutive days, and the amount of the position increases on the third day can be found by substituting Formula (15).
- If the price decreases for four consecutive days, stop trading at this point to avoid extreme losses. Wait until the uptrend before performing a sell operation.
- If there is a continuous upward trend, then w is the amount of continuous increase and the amount of daily position reduction can be found by substituting Formula (16).
- If the increase and decrease alternate, then w is the increase or decrease for that day only.

4. Simulation Trading Result

Since gold and Bitcoin trading accounts are not interchangeable, it is needed to determine the percentage of gold and Bitcoin to invest first. Because of the risk-averse properties of gold, while Bitcoin is a virtual currency, the two financial assets have different risks and returns when invested in. We ran simulation trading using data from 11 September 2016 to 10 September 2021. A 1% transaction cost for gold and a 2% transaction cost for bitcoin is assumed.

Using Markowitz's mean-variance model to quantify the respective risks and returns, we calculate that the mean bitcoin return is 0.0024, and the standard deviation is 0.0419. The mean for gold is 0.0002, and the standard deviation is 0.0087. The annual volatility of bitcoin is 0.6651, and the annual volatility of gold is 0.1381. Since the model we build is day-level high-frequency trading, the higher the volatility, the higher the return, and the higher the risk.

Therefore, we have designed different capital allocation strategies for people with different risk preferences. Based on the volatility values for Bitcoin and gold above, the initial capital allocation for people with different risk preferences are shown in Table 5. is shown in the table below.

Table 5. Initial capital allocation

Risk Appetite	Initial Amount of Bitcoin	Initial Amount of Gold
Risk Seeker	\$800	\$200
Risk Neutral	\$500	\$500
Risk Averse	\$200	\$800

Considering that the initial amount is small, it may not be able to meet the position adjustment amount calculated by the position adjustment function. So when the theoretical position increase amount is greater than the current account balance, the entire account balance is used to increase the position. When the theoretical reduction amount of the current position is held, all current positions are sold.

As of 10 September 2021, the total assets and yield for the three different risk preferences are shown in Table 6.

Table 6. Total Assets and Yield

Risk Appetite	Total Assets	Yield	Annualized Yield
Risk Seeker	\$3660.484	266.05%	53.21%
Risk Neutral	\$2525.126	152.51%	30.50%
Risk Averse	\$1332.917	33.29%	6.67%

Table 6 indicates that during the 5-year trading period, all three risk preferences achieved different levels of profitability, where the yield for the risk preferred was as high as 266.05%. Figure 2 shows the change in the net value curve during the five-year trading period for the three risk appetites.

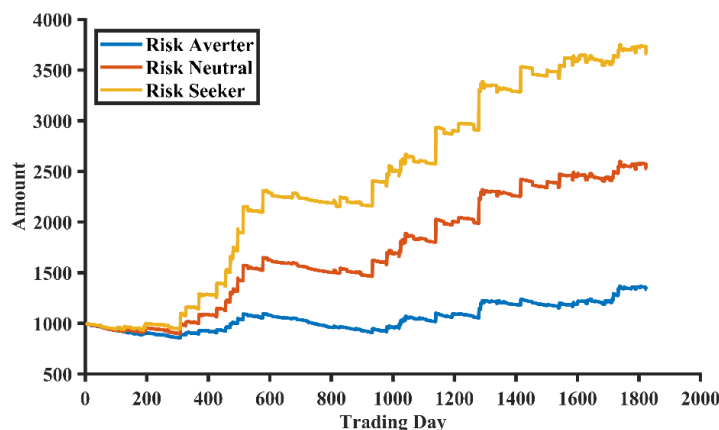


Figure 2. Net Value Curve

5. Evaluation

As shown in Figure 8, all three different risk preferences achieve a steady increase in assets using a high-frequency trading strategy model. As calculated above, bitcoin trading has high volatility and can achieve high returns with high risk. However, it can be seen by the risk preferences' Net Value Curve that assets are rising rapidly with almost no major pullbacks. Using a high-frequency trading strategy, it can allow capturing the gains from volatility in a timely manner. The model uses a dynamic

position-adjustment strategy suitable for people with different investment budgets and risk preferences.

References

- [1] Duffee G R. Stock returns and volatility a firm-level analysis[J]. *Journal of financial Economics*, 1995, 37(3): 399-420.
- [2] Argiddi R V, Apte S S. Future trend prediction of Indian IT stock market using association rule mining of transaction data[J]. *International Journal of Computer Applications*, 2012, 39(10): 30-34.
- [3] Moghar A, Hamiche M. Stock market prediction using LSTM recurrent neural network[J]. *Procedia Computer Science*, 2020, 170: 1168-1173.
- [4] Shuang Sun. Apriori association rules for analyzing financial data [J]. *National Distribution Economy*,2019(02):109-110.
- [5] Xiudong Wang, Yingchun Hu, Hui Li. A study on the application of improved Apriori algorithm in stock analysis [J] *Science and Technology Bulletin*,2013,29(03):125-128.
- [6] Shuhong Guo. Stock analysis simulation system based on Apriori algorithm [J]. *Computer Simulation*,2010,27(06):334-337.
- [7] Wenjing Li, Xiaoxiao Wang. Time series prediction method based on simplified LSTM neural network [J]. *Journal of Beijing University of Technology*,2021,47(05):480-488.
- [8] Jie Zhang. Empirical analysis of stock prediction based on LSTM [D]. *Shandong University*,2020.
- [9] Fengxia Jiao. Research on stock price prediction based on LSTM neural network [J]. *Digital Technology and Applications*,2021,39(03):220-222.
- [10] Kulikov D. A.,Platonov V. V.. Adversarial Attacks on Intrusion Detection Systems Using the LSTM Classifier[J]. *Automatic Control and Computer Sciences*,2022,55(8).