

Research on Investment Income of Gold and Bitcoin

Chenhao Lv^{1,a}, Anji Wang^{1,b}, Yongxuan Wang^{2,c}, Zhenyi Wei^{2,d}, JiuJun Zhang^{2,e,*}

¹ School of Mathematics and Statistics, Liaoning University, Shenyang, 110000, China

² School of Mathematics and Statistics, Liaoning University, Shenyang, 110000, China

^alch0919@foxmail.com, ^bafakesfaker@gmail.com, ^cwyx20210922@163.com, ^d3043969548@qq.com,

^ezjjly790816@163.com

*Corresponding author

Abstract: In financial markets, gold and Bitcoin are common investment assets. Under a fixed commission rate, this paper predicts price changes by referring to financial indicators and established an investment strategy model influenced by known price data. This problem is part of the 2022 U.S. Mathematical Contest in Modeling (MCM) Problem C. Using the Autoregressive Integrated Moving Average (ARIMA) model to predict prices, this paper obtains visualized prediction data sequences, establishes evaluation indicators for bull and bear markets, calculates buying scores, and verifies the accuracy of the investment strategy using a multi-objective dynamic programming model.

Keywords: Portfolio; ARIMA Model; Transaction Strategy Model; Multistage Dynamic Programming Model

1. Introduction

In the current financial market, gold and Bitcoin are two common volatile assets. An increasing number of traders believe that trading gold and Bitcoin is an excellent way to make money. However, investments often come with complex and diverse risks. According to the Basel Committee on Banking Supervision, financial risks include credit risk, market risk, and operational risk. Therefore, in the process of asset trading, traders should learn how to formulate reasonable stock trading strategies, including buying, holding, or selling their assets, while minimizing risks to maximize their overall portfolio.

In 1952, Markowitz^[1] first introduced the Portfolio Theory, which quantifies risk and return based on optimization principles and establishes the Mean-Variance model to determine the optimal asset allocation. Subsequently, the theory has been continuously extended and developed. For example, William Sharpe et al.^[2] proposed the Capital Asset Pricing Model, and Fisher Black^[3] and Robert Litterman^[4] introduced the Black-Litterman asset allocation model. In practice, investors need to continuously adjust their investment strategies to achieve maximum returns, thus transforming the portfolio selection problem into a multi-stage or dynamic investment process becomes necessary. Mossin^[5] first extended Markowitz's single-stage model to the multi-stage case using dynamic programming methods, but these approaches often face high computational complexity when dealing with large-scale portfolio problems.

Therefore, this paper presents an investment strategy model based on the Autoregressive Integrated Moving Average (ARIMA) model and comprehensive evaluation. The accuracy of the investment strategy is verified using a multi-objective dynamic programming model. This problem is part of Problem C in the 2022 Mathematical Contest in Modeling for College Students (MCM/ICM). The problem provides trading day price data from September 11, 2016, to September 10, 2021, with a commission rate of 1% for gold and 2% for Bitcoin. Further details can be found in reference^[6].

This article aims to address the following issues:

Issue One: Establishing a model based on the price data up to the current day to determine the most probable price for the next trading day.

Issue Two: Establishing a model based on the price data up to the current day to provide the optimal investment strategy for the present day. Concurrently, assuming an initial investment of \$1,000, the model will be utilized to calculate the returns after five years.

2. Model Establishment

2.1. Model Preparation

In order to address the problem more effectively, some assumptions need to be made prior to establishing the model. In mathematical modeling, assumptions are essential as they offer a means to simplify the problem, facilitating a better comprehension and resolution of the issue.

Assumption 1: Due to limited historical price data on the initial trading day, it is considered insufficient for predicting price trends and obtaining the price data for the next trading day.

Justification: Although a small number of data samples can still be predicted by time series model, it is likely to result in a large difference between the predicted results and the real value. As a result, the data of the previous days will not be predicted.

Assumption 2: In the process of asset investment, short selling is not allowed. In other words, the amount of assets sold by investors is not allowed to exceed their actual asset holdings at that time.

Justification: The context of the problem revolves around a closed investment scenario in a self-contained period without external interventions. As a result, only buying, holding, and selling behaviors for gold and Bitcoin are considered, while other investment activities are excluded.

2.2. Item-by-item ARIMA Model

The variation in price data is determined by multiple factors, such as policy changes, unforeseen events, and others. For such a complex dynamic system, time series analysis^[7,8] can effectively utilize the information within price data for predictions, disregarding external factors. Additionally, in this study, missing data is imputed using the method of sequence averaging, revealing that the provided data comprises non-stationary time series without any seasonal variations. Consequently, the suitability of the ARMA model is negated, and ARIMA model emerges as the optimal choice for prediction.

In accordance with Assumption 1, this study considers the initial ten trading days as the observation period and commences predicting the price data for the subsequent trading day starting from the eleventh day, following the algorithm as outlined below. Where, i is the variable that determines the type of asset ($i = 1$, Gold; $i = 2$, Bitcoin), j is the time from September 11, 2016, F_i^j is the factual price data for asset i on day j , and P_i^j is the predicted price data for asset i on day j .

Algorithm 1: The process of prediction of the price

Input: F_i^t (t from 1 to j)

For $j = 10$ to the penultimate day **do**

Use the ARIMA model to process the factual data up to that trading day to predict the next trading day, and get P_i^{j+1}

end

Generate $\{P_i^j\}$, $j = 11$ to the last trading day

Output: P_i

Applying the ARIMA (p, d, q) model to the price data, the general formulation for the price series j is as follows:

$$\Delta^{(d)} F_i^t = \alpha_0 + \sum_{n=1}^p \alpha_n \Delta^{(d)} F_i^{t-n} + \varepsilon_t + \sum_{n=1}^q \beta_n \varepsilon_{t-n} \tag{1}$$

where, $\Delta^{(d)}$ represents the difference operator of order, ε_t is the residual value of the model and is normally distributed white noise.

After obtaining the ARIMA model with SPSS, performing a white noise test on the residuals is needed. If the residual is white noise, it means that the calculated model can fully identify the law of price change time series data, in other words, the model is acceptable; if not, it means that there is still some price impact information which has not been identified by our model. Then the model needs to be revised to better identify this part of the information. Performing the d -th order difference on the price data sequence $\{F_i^j\}$, the processed sequence is identified as $\{x_n\}$. $\{x_n\}$ is needed to be stable, that is to

meet

$$\begin{cases} E(x_t) = E(x_{t-s}) = \mu \\ \text{Var}(x_t) = \text{Var}(x_{t-s}) = \sigma^2 \\ \text{Cov}(x_t, x_{t-s}) = \gamma_s \end{cases} \quad (2)$$

Define the sample autocorrelation coefficient:

$$r_s = \hat{\rho}_s = \frac{\sum_{t=s+1}^T (x_t - \bar{x})(x_{t-s} - \bar{x})}{\sum_{t=1}^T (x_t - \bar{x})^2} \quad (3)$$

If $\{\varepsilon_t\}$ is white noise series, then $\rho_s = \begin{cases} 1, & s = 0 \\ 0, & s \neq 0 \end{cases}$.

The Q-test is going to be used to help test whether the residuals are white noise or not, where $H_0: \rho_1 = \rho_2 = \dots = \rho_s = 0$ and $H_1: \rho_i (i = 1, 2, \dots, s)$ at least one is not 0.

2.3. Trading Strategy Model

Based on the data available up to the current day, it is essential to analyze the past price changes and divergences within appropriate time intervals to assess the current market conditions, which becomes a decisive factor in determining the investment strategy. To achieve this, it is necessary to establish relevant quantitative indicators and formulate rules to provide appropriate trading strategies^[9].

2.3.1. Bull-Bear Market Judgment Model

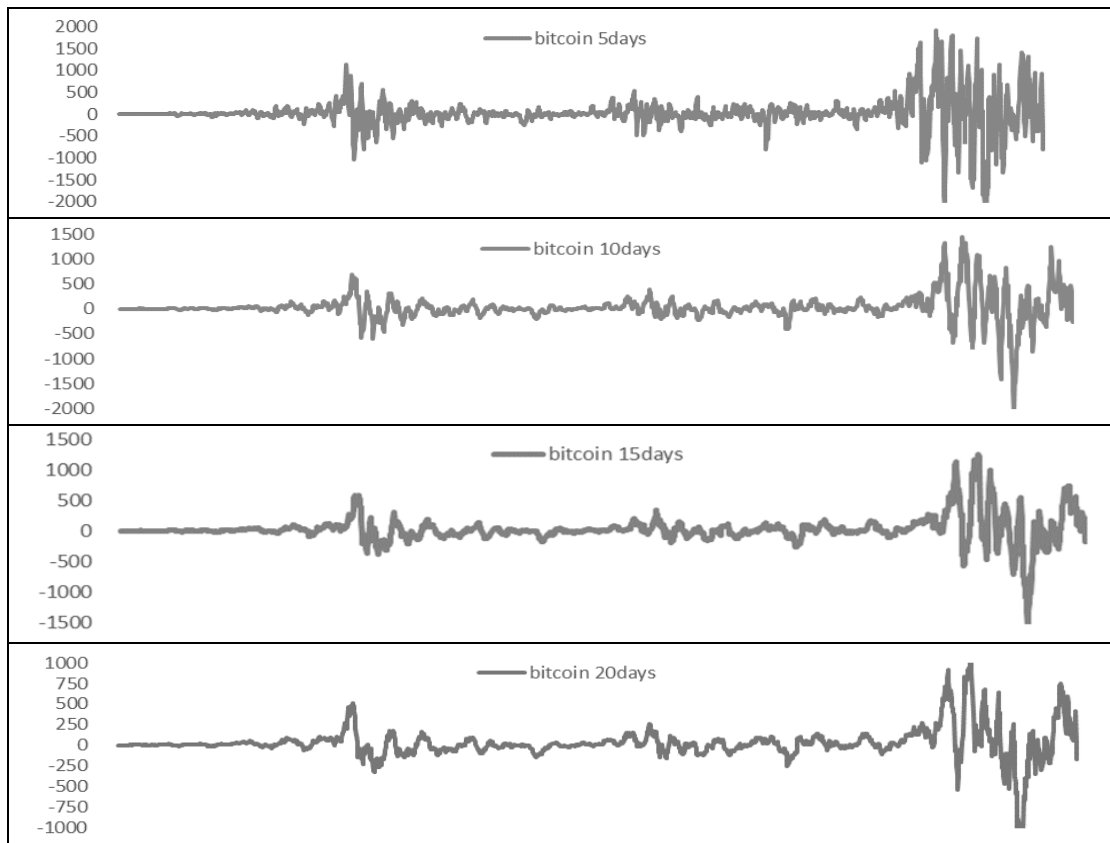


Figure 1: Bitcoin-Increase amount in the time periods of 5, 10, 15 and 20 days

Firstly, based on the five-year data of gold and bitcoin, the actual amount of increase of gold and bitcoin can be calculated, and the actual amount of increase on the day is marked as $\Delta F_i^j = F_i^j - F_i^{j-1} (j = 2, \dots, 1826; i = 1, 2)$. The comparison of gold and Bitcoin for different

time periods is illustrated in Figure 1 and Figure 2 respectively.

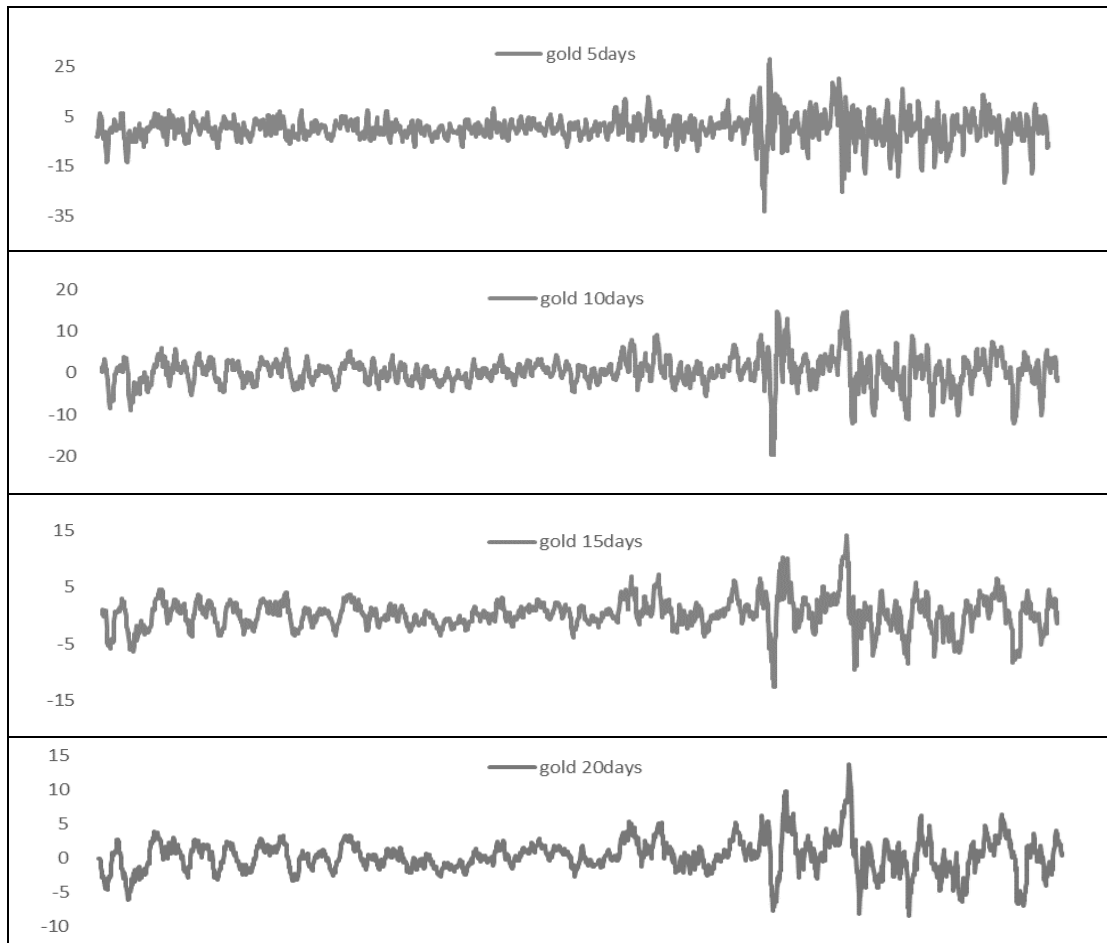


Figure 2: Gold-Increase amount in the time periods of 5, 10, 15 and 20 days

The assessment of the price fluctuations of Bitcoin and gold can be conducted based on the comparison of their respective time periods. Calculating the average price increase over a five-day time period for Bitcoin, the maximum surge is approximately 2000, whereas the average price increase over a ten-day time period yields a maximum surge of about 1000. In contrast, the rapid and substantial changes are prone to being missed, hence the selection of a five-day time period for evaluating the price fluctuations. Concurrently, it can be observed that gold exhibits smaller price fluctuations, leading to the choice of a fifteen-day time period for assessing its price movements.

To obtain the bull-bear market identification model, the n -day BIAS of asset i on day j is defined as

$$BIAS_{i,n}^j = \frac{F_i^j - \overline{F_{i,n}^j}}{\overline{F_{i,n}^j}} \tag{4}$$

where, $\overline{F_{i,n}^j}$ represents the average price of the asset i on n days before the day j . According to the calculation formula of $BIAS_{i,n}^j$, the 15-day BIAS of gold ($BIAS_{1,15}^j$) and the 5-day BIAS of

Bitcoin ($BIAS_{2,5}^j$) can be calculated. The calculated data can be normalized by formula $\frac{x - x_{\min}}{x_{\max} - x_{\min}}$, which are recorded as $BIAS_{1,15}^{j*}$ and $BIAS_{2,5}^{j*}$ respectively.

Based on the obtained data, the entropy weighting method employed to calculate the corresponding weights for the aforementioned two indicators, the bull-bear market assessment indicators for gold and Bitcoin on day j are established as follows:

$$\begin{aligned} BULL_1^j &= 0.597 \overline{\Delta F_1^j} + 0.403 \overline{BIS_{1,15}^{j*}} \\ BULL_2^j &= 0.597 \overline{\Delta F_2^j} + 0.403 \overline{BIS_{2,5}^{j*}} \end{aligned} \tag{5}$$

where, $\overline{\Delta F_i^j}$ represents the average increase of the first 90 days of the asset i on the day j ; $\overline{BIAS_{i,n}^{j*}}$ represents the n -day average BIAS of the asset i after the normalization process during the 90 days before the day j ; $BULL_i^j$ represents the bull-bear market evaluation index of the asset i on the day j . It is believed that if the daily evaluation index is bigger than the average of the bull-bear market evaluation index, it is a bull market, otherwise it is a bear market.

2.3.2. Buy-in Scoring Model

The purchase risk of gold is related to the BIAS of gold and the gold bull market. The entropy weight method used to calculate the weight of the BIAS of gold and the gold bull market, the risk model of gold and Bitcoin on the day j is established as follows:

$$\begin{aligned} R_i^j &= 0.666 BULL_i^j + 0.333 BIS_{i,15}^j \\ R_2^j &= 0.666 BULL_2^j + 0.333 BIS_{2,5}^j \end{aligned} \tag{6}$$

where, R_i^j represents the purchase risk index of the asset i on the day j .

Due to the existence of non-trading days for gold, it is necessary to handle the missing values between the actual price increase and the estimated price increase residual ε . The missing residuals are supplemented using the residuals from the previous trading day. Based on the established time series model in the preceding text, the expected price increase is computed. Specifically, the expected price increase on day j is denoted as $\Delta P_i^j = P_i^j - F_i^{j-1}$ ($j = 2, \dots, 1826$; $i = 1, 2$). Subsequently, the expected price increase is subjected to normalization, denoted as ΔP_i^{j*} . Finally, by varying the weights between different indicators for comparison, a rational purchase scoring model is established as follows:

$$S_i^j = 9.75 \Delta P_i^{j*} + 5.375 BULL_i^j - \varepsilon + \frac{1}{R_i^j} \tag{7}$$

2.4. Theoretical Maximum Return Model

In order to assess the optimality of the investment model, the theoretical maximum return value is a critical evaluation factor. As investments can be made daily and are subject to constraints such as the total amount of funds, this problem falls under the category of a multi-stage^[10] investment portfolio problem with interrelated investments at each stage. The objective is to maximize the investment return after the final investment selection. Therefore, it is necessary to study this problem, establish a model, and obtain the results under the consideration of the entire set of practical data.

Investors have two types of assets to choose from, and invest on the remaining 1825 days excluding the last day. Let's say that at the beginning of day j , the i -th asset is purchased, and the rate of return at the end of the day is $r_i^j \left(r_i^j = \frac{F_i^{j+1} - F_i^j}{F_i^j} \right)$.

Moreover, if some assets are already held at the end of day $j - 1$, at the beginning of the day j , it can be decided whether to continue to hold these assets or sell them, and the funds obtained from the sale of the assets can be used for asset purchases on that day.

Introduce the following notations:

s_i^j : At the beginning of day j , the amount of the asset i ;

t^j : At the beginning of day j , the amount of cash;

x_i^j : On day j , the transaction amount for the asset i .

When $x_i^j < 0$, it indicates selling the asset i ; when $x_i^j > 0$, it indicates buying the asset i . Then,

(s_1^j, s_2^j, t^j) represents the state of the amount of assets and free funds held on day j , which is called the investment state variable of day j , abbreviated as st_j . (x_1^j, x_2^j) represents the situation of the asset transaction on day j , which is called the investment decision variable of day j , abbreviated as X_j .

The relationship between investment decision variables and state variables and their constraints are analyzed below.

A. State variables

Let

$$s_i^1 = 0 (i = 1, 2), t^1 = 1000 \tag{8}$$

Then, according to the definition of decision variables, when the investment state of day j is st_j , the investment state of day $j+1$ will be affected by the investment decision X_j of that day. Details are as follows:

① The investment of the day j to the asset i will affect the asset holdings at the end of day j , which is $s_i^j + x_i^j$. So the return on asset i to day $j+1$ is $r_i^j (s_i^j + x_i^j)$, and the investment state st_{j+1} on day $j+1$ is:

$$s_i^{j+1} = (1 + r_i^j) (s_i^j + x_i^j), (i = 1, 2) \tag{9}$$

When it is on a non-trading day of gold, considering that gold does not trade, it is reasonable to assume that the transaction amount of the asset, together with the rate of return, is 0, so that assets owned by gold on non-trading days can be guaranteed the same as the day before.

Assumption 2 shows that

$$s_i^{j+1} \geq 0, (i = 1, 2) \tag{10}$$

② The investment x_i^j in asset i on day j affects the amount of free money held at the end of day j , which is equal to the sum of the asset amount $\sum_{i=1}^2 x_i^j$ purchased and the commission $\sum_{i=1}^2 p_i |x_i^j|$. So, the amount of cash on day $j+1$ can be calculated as

$$t^{j+1} = t^j - \sum_{i=1}^2 [x_i^j + p_i |x_i^j|] \tag{11}$$

Assumption 2 shows that $t^{j+1} \geq 0$, which means that

$$\sum_{i=1}^2 [x_i^j + p_i |x_i^j|] \leq t^j \tag{12}$$

B. Decision variables

Firstly, Assumption 2 shows that the value of the assets actually sold cannot exceed the asset holdings at this time, which means that

$$s_i^j + x_i^j \geq 0 \tag{13}$$

Let $f_k(X_k, \dots, X_{1825}) = \left(t^{1826} + \sum_{i=1}^2 s_i^{1826} \right) - \left(t^k + \sum_{i=1}^2 s_i^k \right)$, $(1 \leq k \leq 1825)$, which represents

the final actual income according to the investment decision variables X_k, \dots, X_{1825} from the investment state st_k on day k . Therefore, the investment decision is to find the daily investment decision variables to make the final total return $f_1(X_1, \dots, X_{1825})$ as large as possible under the premise of (8)-(13). Since the states of two adjacent days are related, investment decisions can be made according to the recursive relationship between the state variables.

Let $F_k = \max_{X_k, \dots, X_{1825}} f_k(X_k, \dots, X_{1825})$, $(1 \leq k \leq 1825)$, which represents the maximum total return

on the last day according to the investment decision variables X_k, \dots, X_{1825} from day k . It can be deformed as follows:

$$\begin{aligned}
 F_k &= \max_{X_k, \dots, X_{1825}} \left\{ f_{k+1}(X_{k+1}, \dots, X_{1825}) + \left[\left(t^{k+1} + \sum_{i=1}^2 s_i^{k+1} \right) - \left(t^k + \sum_{i=1}^2 s_i^k \right) \right] \right\} \\
 &= \max_{X_k} \left\{ F_{k+1} + \left[\left(t^{k+1} + \sum_{i=1}^2 s_i^{k+1} \right) - \left(t^k + \sum_{i=1}^2 s_i^k \right) \right] \right\} \\
 &= \max_{X_k} \left\{ F_{k+1} + \left[\sum_{i=1}^2 r_i^k (s_i^k + x_i^k) - \sum_{i=1}^2 p_i |x_i^k| \right] \right\}
 \end{aligned} \tag{14}$$

Therefore, the investment decision can be represented by a dynamic programming, and the basic equation of the dynamic programming can be set as:

$$\begin{cases}
 F_j = \max_{X_j} \left\{ F_{j+1} + \left[\sum_{i=1}^2 r_i^j (s_i^j + x_i^j) - \sum_{i=1}^2 p_i |x_i^j| \right] \right\} \\
 s \cdot t \cdot \begin{cases} x_i^j \geq -s_i^j \\ \sum_{i=1}^2 [x_i^j + p_i |x_i^j|] \leq t^j \end{cases} & j = 1, \dots, 1825 \\
 & i = 1, 2 \\
 s_i^{j+1} = (1 + r_i^j) (s_i^j + x_i^j) \\
 t^{j+1} = t^j - \sum_{i=1}^2 [x_i^j + p_i |x_i^j|] \\
 s_i^1 = 0, t^1 = 1000
 \end{cases} \tag{15}$$

3. Actual Measurement Process and Results

3.1. Results of the ARIMA Model

SPSS being utilized to directly apply the ARIMA (p, d, q) model to the transaction data of the first 1000 trading days to solve the prediction^[11], the obtained model is ARIMA $(0, 1, 18)$. The fitting results are shown in Table 1 below.

Table 1: Model Fit Results

Fit Statistic	Mean
Stationary R-squared	0.012520362
R-squared	0.991950171

A large R-squared indicates that the fitting effect is very good, but a small Stationary R-squared indicates that the model needs to be improved. Then we remove the outliers and recalculate to obtain the ARIMA $(1, 1, 1)$ model. The fitting results are as follows:

Table 2: Model Fit Results

Fit Statistic	Mean
Stationary R-squared	0.3670409
R-squared	0.995316565

Table 3: Q-test Results

Ljung-Box Q		
Statistics	DF	Sig.
20.96619155	16	0.179813804

Given that the R-squared increases slightly, the Stationary R-squared increases significantly, and the Q test accepts the null hypothesis. Therefore, it is believed that the obtained residuals are white noises, indicating that the model recognizes the data very well. Using this to predict the data of the next trading day, the value of one bitcoin on the 1001st trading day is 7777.1992, which is recorded as P_{1001} . Starting

from the 10th trading day, the data sequence $\{P_n\}$ from the 11th trading day to the last trading day is predicted. The curve of the factual value and the predicted value is as follows:

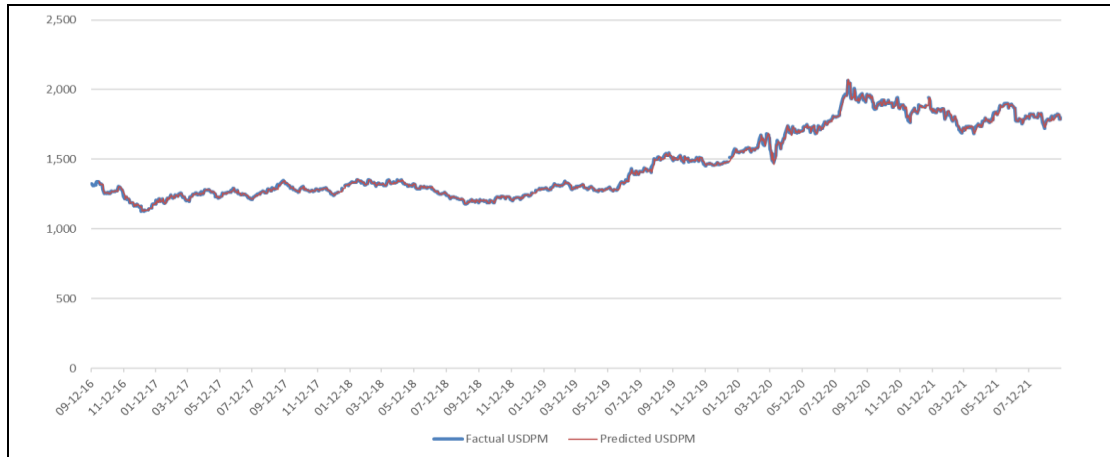


Figure 3: Gold-The curve of the factual value and the predicted value



Figure 4: Bitcoin-The curve of the factual value and the predicted value

3.2. Results of the Trading Strategy Model

3.2.1. Trading Strategy

Normalize the calculated scores and determine a threshold based on this to determine whether the transaction is buying or selling, thus formulating the following trading strategy:

- When $S_1^j > 0.56$, gold should be purchased; When $S_1^j < 0.32$, it should be sold.
- When $S_2^j > 0.73$, bitcoin should be purchased; When $S_2^j < 0.52$, it should be sold.
- Trading rules and strategies:

- 1) Judge whether it is a gold trading day. If not, gold trading should be suspended;
- 2) In order to ensure that sufficient commissions can be paid, spending all the cash you own is prohibited;

3) When bitcoin and gold can be purchased at the same time, if $buy\text{-}in\ score\ of\ gold - 0.56 > (buy\text{-}in\ score\ of\ bitcoin - 0.73) \times 2$, then only gold should be purchased, otherwise only bitcoin should be purchased (That buy-in score of gold is lower than buy-in score of bitcoin represents that gold have more room for improvement);

$$4) buy\text{-}in\ amount = [current\ cash\ amount \times buy\text{-}in\ score \times (1 - \alpha\%)] / current\ price$$

Where, $\alpha\%$ represents the commission for each transaction ($\alpha_{gold} = 1\%$, $\alpha_{bitcoin} = 2\%$).

$$5) selling\ amount = holding\ shares \times (1 + buy\text{-}in\ score - selling\ index)$$

3.2.2. Results

Based on the above suggested strategy, the value of assets owned per day (initial capital of \$1000) can be calculated and a graph of asset changes over time is created as follows simultaneously:

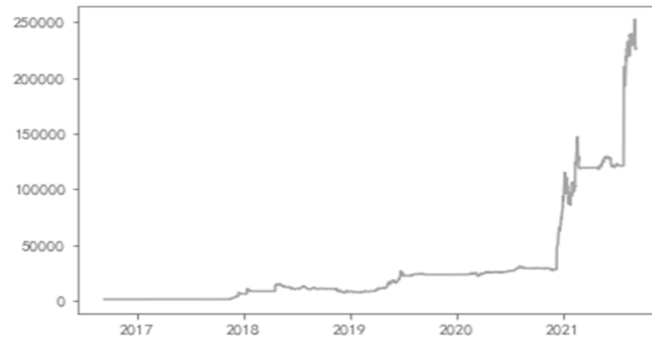


Figure 5: Changes of total assets

3.3. Results of the Theoretical Maximum Return Model

Using MATLAB to solve the model, the theoretical maximum return of \$260,304 and the optimal solution $X^* = (X_1^*, \dots, X_{1825}^*)$ are obtained. According to the significance of the model establishment, is the overall optimal strategy for investors to invest, which can maximize the total return obtained on the last day under a series of investment fund constraints. Where, X_j^* is the investment decision on day j of the overall optimal strategy.

Based on the data up to that day, according to the investment strategy described above, the final return of \$1000 is \$247,214, slightly lower than the maximum return of \$260,304 obtained by the model. Due to the fact that the model is analyzed based on all data from 5 years, the maximum return calculated by it is inevitably higher than the maximum return calculated by the trading strategy model. Therefore, the trading strategy obtained above is reasonable.

4. Conclusions

This paper firstly takes the first 10 trading days as the observation period and adopts the Item-by-Item ARIMA Model to predict future price data. Starting from the 11th trading day, the price data of the next trading day is sequentially predicted to obtain the prediction sequence. Then, the buying score calculated, the bull and bear market judgment model and the risk model are established. Based on that, the optimal investment strategy for the trading day is provided and the maximum return that can be obtained after 5 years under a certain initial investment is calculated, which is \$247,214. Finally, based on five years of actual data, a Multi-stage Dynamic Programming Model is established to obtain the theoretical maximum return, which is not much different from the maximum return calculated according to the trading strategy model, verifying the rationality of the daily trading strategy model.

References

- [1] Markowitz H. *Portfolio Selection [J]. The Journal of Finance, 1952, 7(1):77-91.*
- [2] Sharpe W F. *Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk [J]. The Journal of Finance, 1964, 19(3):425-442.*
- [3] Black F, Litterman R B. *Asset Allocation: Combining Investor Views with Market Equilibrium[J]. The Journal of Fixed Income, 1991, 1(2):7-18.*
- [4] Black F, Litterman R. *Global Portfolio Optimization[J]. Financial Analysts Journal, 1992, 48(5):28-43.*
- [5] Mossin J. *Optimal Multiperiod Portfolio Policies [J]. The Journal of Business, 1968, 41(2):215-229.*
- [6] 2022 MCM/ICM Problems[EB/OL]. [2023-07-30]. <http://www.mathmodels.org/Problems/2022/MCM-C/index.html>.
- [7] Wang Zhenlong. *Time Series Analysis[M]. Beijing: China Statistics Press, 2002.*
- [8] Zhang Shujing, etc. *Time Series Analysis Concise Course[M]. Beijing: Tsinghua University Press,*

2003.

[9] Yuan Xinsheng, Shao Dahong, Yu Shilian. *Application of LINGO and Excel in Mathematical Modeling [M]*. Beijing: Science Press, 2007. 135-173.

[10] Su Jie, Liu Jiazhuang. *Multi-stage asset investment dynamic programming decision model[J]*. *Chinese Journal of Management Science*, 2001, V (3):55-61.

[11] Jiang Qiyuan, Xie Jinxing, Ye Jun, *Mathematical Model [M]*. Beijing: Higher Education Press, 2011