ISSN: 2414-1895 DOI: 10.6919/ICJE.202205_8(5).0060

Analysis of Trading Strategy based on ARIMA and Dynamic Programming Model

Zhiduo Wang, Zixing Liu, Shuai Yin

School of Artificial Intelligence, North China University of Science and Technology, Tangshan 063000, China

Abstract

Market traders obtain the maximum value by buying and selling two volatile assets, gold and bitcoin. This paper analyzes the historical data of each day to establish different asset value prediction models and investment transaction planning models, so as to give the market The best investment strategy for traders. First, use Lagrange interpolation method to supplement the missing data in gold, predict the future value by using the effective real data of gold and bitcoin, and select the previous 30-day data of gold and bitcoin as the training set respectively, and establish an ARIMA-based gold and bitcoin model. The Bitcoin prediction model predicts the value of the 31st day, and then predicts the value of different assets on the 32nd day through the real data of the first 31 days, so as to recursively give the predicted value of gold and Bitcoin in the past five years. Secondly, introduce the Sharpe ratio that has both risk and profitability, and use the predicted Sharpe ratio of the n+1 day as the objective function, the daily profit value is greater than or equal to the transaction commission, and invest in the cash, gold and bitcoin pairs after the transaction. The proportion of total assets is a non-negative number, these four are set as constraints, and the change in the proportion of gold and bitcoin is set as a decision variable, and an asset investment transaction model based on dynamic programming is established to make reasonable decisions on daily transactions, and get the results for each day. Cash, gold, bitcoin holding rates to get the highest benefits.

Keywords

Lagrange Interpolation; ARIMA Model; Dynamic Programming Model.

1. Background

Futures investment appeared in the United States in the 1970s. After decades of development, it has grown to a large scale. Especially in the past fifteen years, with the rapid financing of the commodity futures market, futures investment funds have gradually formed, the financial market has many product portfolios for market traders to choose and buy. At the same time, due to the financialization of futures, the price fluctuations and the complexity of the cash-out process have also increased, which has increased the transaction risks in the financial market. Market traders are required to frequently buy and sell these unstable and risky assets in the investment market every day to maximize their total returns.

2. Model Establishment and Solution

2.1 Data Preprocessing

First of all, we analyze the given data and find that in addition to the date when the transaction is not open, there is also the phenomenon that there is a date but no closing price in the data. We will

Volume 8 Issue 5, 2022
DOI: 10.6919/ICJE.202205_8(5).0060

supplement the missing data. We use Lagrange interpolation [1] to supplement the missing data, we set the values given in the file table as $x_1, x_2, \cdots, x_n$, and then set the corresponding closing prices as $y_1, y_2, \cdots, y_n$, Its expression is formula (1):

$$L(x) = \sum_{m=1}^{n} y_m \left(\prod_{\substack{j=1 \\ j \neq m}}^{n} \frac{x - x_j}{x_m - x_j} \right) \tag{1}$$

The supplementary data obtained according to formula (1) are shown in Table 1 below:

Table 1. Gold price missing data supplement

12/23/16	12/30/16	12/22/17	12/29/17	12/24/18	12/31/18	12/14/19	12/31/19	12/24/20	12/31/20
1126.7	1148.3	1267.9	1302.8	1259.1	1278.9	1494.9	1516.5	1873.2	1919.6

2.2 Using ARIMA Model to Predict the Trend of Gold and Bitcoin

ARIMA model is one of the most common statistical models used for time series forecasting. The ARIMA model has three parameters: p, d, q, denoted as ARIMA(p, d, q), where p represents the lag number of the time series data itself used in the prediction model, and d represents how many orders of difference the data needs to perform in order to make the data Stationary, q represents the number of lags in the forecast error employed in the forecast model [2].

2.2.1 Model Stationarity Test

First, the data of gold and bitcoin are tested for stationarity. Stationarity requires that the mean and variance of the series do not change significantly. After the ADF test is performed on the Bitcoin data, when the difference is 0 order, the significant P value is 0.935, and the obtained series is an unstable time series, which does not satisfy the ARIMA model. When the difference is 1 order, the significant P value is 0.000084, which is significant on the level, and the series is a stationary time series. Therefore, the optimal difference sequence diagram when d=1 is obtained is Figure 1, and it shows a stationary state in the horizontal direction.

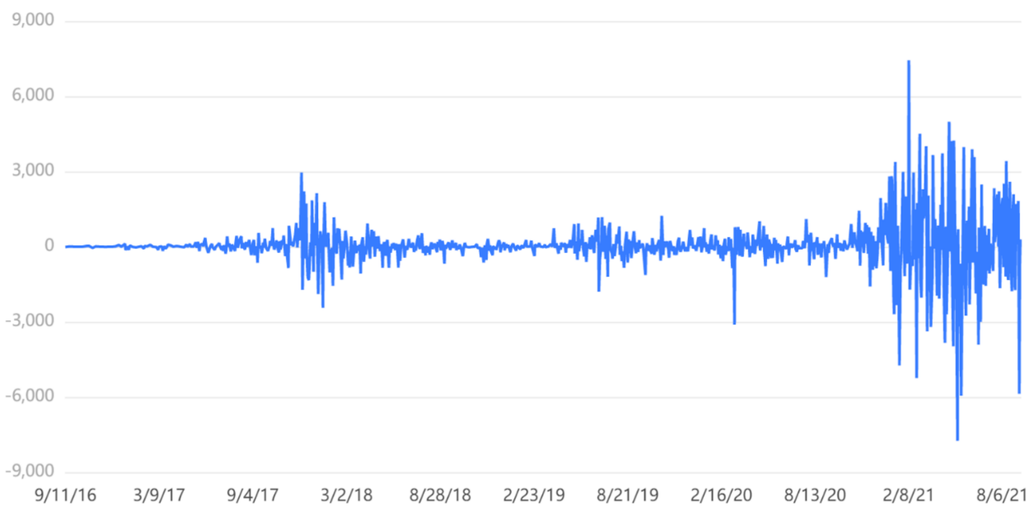

Figure 1. Timing diagram after the first-order difference of Bitcoin data

ISSN: 2414-1895 DOI: 10.6919/ICJE.202205_8(5).0060

Similarly, the ADF test is performed on the gold data. When the difference is 1 order, the significant P value is 0.000059, and it is concluded that the sequence is a stationary time series. At the same time, the best difference sequence graph when d=1 is obtained is Figure 2.

Figure 2. Timing diagram after first-order difference of golden data

2.2.2 Model Identification and Order Determination

Using the autocorrelation function ACF and the partial autocorrelation function PACF, the function image is obtained, and then the values of p and q are determined according to the following rules [3].

Table 2. p,q Determination Criterion

Model (sequence)	AR(p)	MA(q)	ARMA(p,q)
Autocorrelation function	trailing	truncate after the qth	trailing
Partial autocorrelation function	truncate after the pth	trailing	trailing

By processing the data of Bitcoin with the autocorrelation function and the partial autocorrelation function, the autocorrelation diagram and the partial autocorrelation diagram are obtained, as shown in Figure 3 and Figure 4.

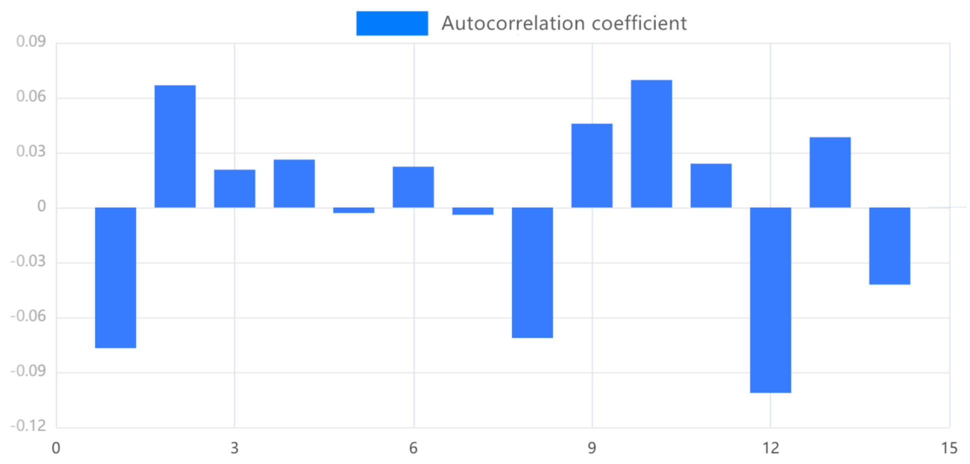

Figure 3. Bitcoin data autocorrelation diagram

ISSN: 2414-1895
DOI: 10.6919/ICJE.202205_8(5).0060

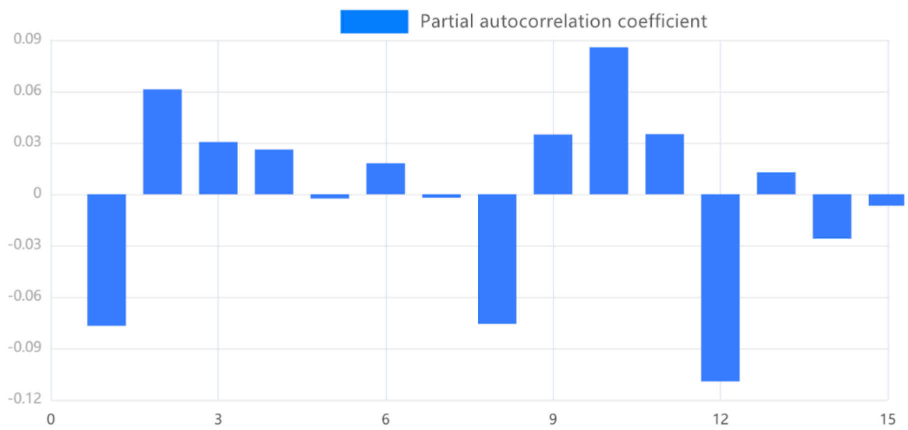

Figure 4. Bitcoin data partial autocorrelation diagram

And through the image, we can judge its tailing and censoring state, and get three values of the three fitted models of Bitcoin data, which are $ARIMA_B(0,1,2)$, $ARIMA_B(1,1,1)ARIMA_B(2,1,0)$.

In the same way, we also process the autocorrelation function and the partial autocorrelation function for the gold data and obtain the autocorrelation diagram and the partial autocorrelation diagram as shown in Figure 5 and Figure 6.

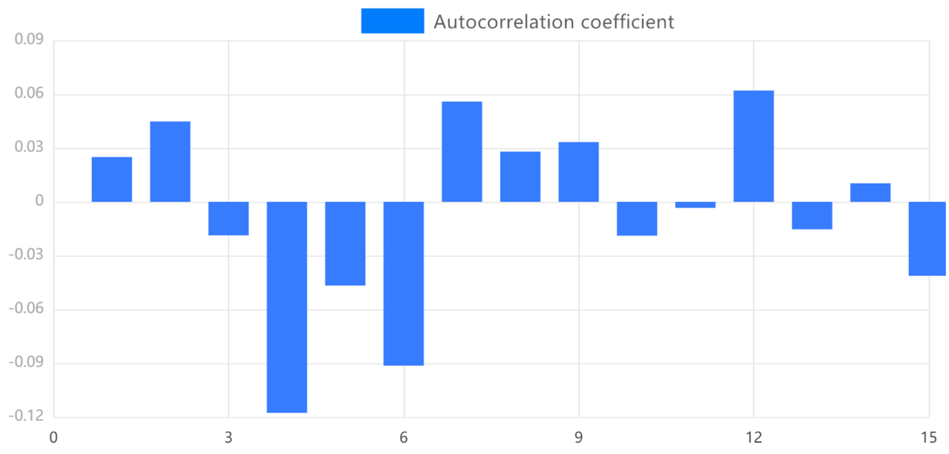

Figure 5. Autocorrelation plot of gold data

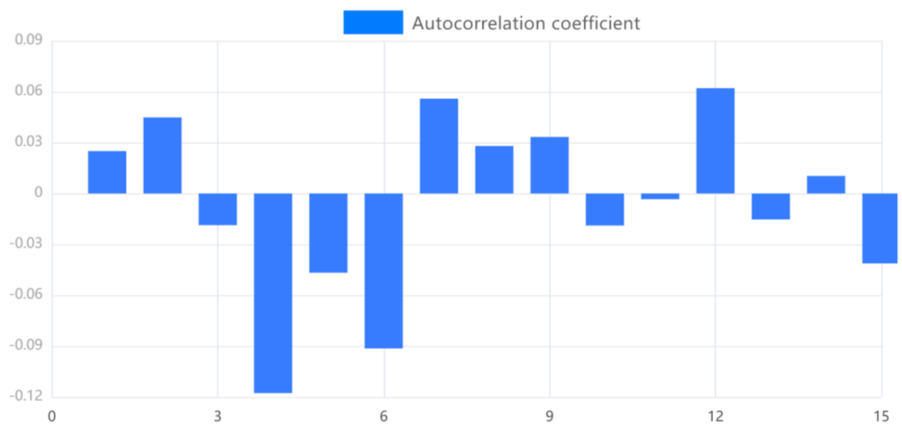

Figure 6. Partial autocorrelation plot of gold data

ISSN: 2414-1895 DOI: 10.6919/ICJE.202205_8(5).0060

Judging the gold data by Figure 5 and Figure 6, we get 3 values of the three fitting models of the gold data, which are $ARIMA_G(0,1,1)$, $ARIMA_G(1,1,0)$, $ARIMA_G(1,1,2)$.

2.2.3 Parameter Estimation

Through tailing and truncation to determine the order of the model, three possible prediction models are obtained, but human judgment is highly subjective, so in the case of the same prediction error, we use the information function criterion method AIC criterion to determine order of the model.

The full name of the AIC criterion is the minimum information criterion, and the calculation formula is as follows:

$$AIC = 2k - 2ln(L) \tag{2}$$

where k is the number of parameters and L is the likelihood function.

(1)For the selection of the Bitcoin price prediction model, by observing Figure 3 and Figure 4 and observing the tailing and truncation, different parameters were used for multiple fittings. Finally, three models passed the model test, which are $ARIMA_B(0,1,2)$, $ARIMA_B(1,1,1)$, $ARIMA_B(2,1,0)$, calculate the AIC value under each model, as shown in the following table:

Table 3. AIC values for multiple bitcoin models

Model	AIC value
$ARIMA_B(0,1,2)$	18523.931
$ARIMA_B(1,1,1)$	29604.773
$ARIMA_B(2,1,0)$	29601.956

$ARIMA_B(0,1,2)$ model has the smallest AIC value, so it is the optimal model. The model formula is as follows (3):

$$Y(t) = 25.067 - 0.076 * \varepsilon(t-1) + 0.068 * \varepsilon(t-2) \tag{3}$$

(2)For the selection of gold price prediction models, by observing Figure 4 and Figure 5, and observing tailing and truncation, different parameters were used for multiple fittings. Finally, three models passed the model test, which are $ARIMA_G(0,1,1)$, $ARIMA_G(1,1,0)$, $ARIMA_G(1,1,2)$, calculate the AIC value under each model, as shown in Table 4 below:

Table 4. AIC values for multiple golden models

Model	AIC value
$ARIMA_G(0,1,1)$	652.681
$ARIMA_G(1,1,0)$	10224.577
$ARIMA_G(1,1,2)$	10225.262

ISSN: 2414-1895 DOI: 10.6919/ICJE.202205_8(5).0060

$ARIMA_G(0,1,1)$ model has the smallest AIC value, so it is the optimal model. The model formula is as follows (4):

$$Y(t) = 0.372 + 0.023 * \varepsilon(t-1) \tag{4}$$

2.2.4 Model Checking

Model validity check:

$ARIMA_B(0,1,2)$ for Bitcoin price prediction, according to the p value of Q statistic is greater than 0.1, it can be considered that the model residual is a white noise sequence, and through the above AIC value comparison analysis, The AIC value of the model is the smallest, and the goodness-of-fit R2 is 0.996, which is close to 1. The model has excellent performance and basically meets the requirements.

$ARIMA_G(0,1,1)$for gold price prediction, according to the p value of Q statistic is significantly greater than 0.1, it can be considered that the model residual is a white noise sequence, and through the above AIC value comparison analysis, When the model $ARIMA_G(0,1,1)$ is taken, the AIC value of this model is the smallest, and the goodness of fit R2 is 0.997, which is close to 1, then the model has excellent performance and the model meets the requirements.

2.2.5 Model Results

According to the prediction model we trained, we can predict the value of gold and bitcoin in the next day according to the price data up to a certain day and establish a second investment planning model through our prediction results to adjust the value of gold and bitcoin. Invest in value for maximum benefit. The training results of Bitcoin and gold are shown in Figure 7 and Figure 8 below:

Figure 7. Bitcoin predicted value and real value comparison chart

Volume 8 Issue 5, 2022
DOI: 10.6919/ICJE.202205_8(5).0060

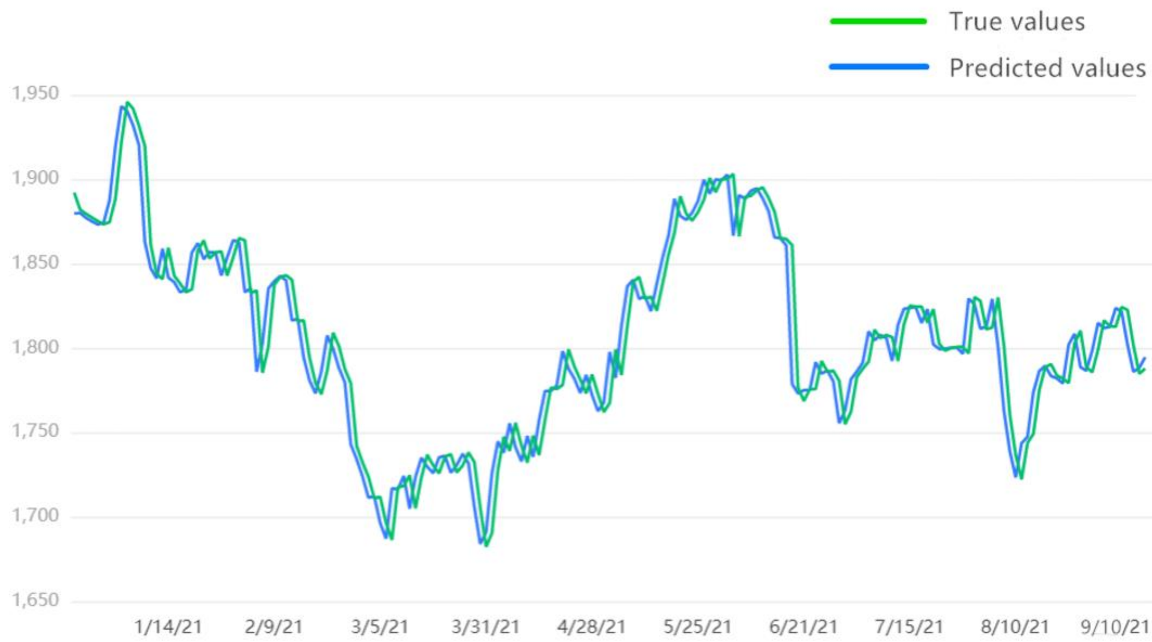

Figure 8. Gold predicted value and real value comparison chart

2.3 Investment Transaction Model based on Dynamic Programming

Through the ARIMA prediction model, we can get the value of Bitcoin and gold on the nth day. According to the forecast data on the nth day, we make a decision-making investment in gold and Bitcoin, to obtain the maximum profit on the n+1st day. On the n+1st day, we will know the real value of the n+1th day, and today's data will be historical data, and then use our prediction model to predict the n+2th day's value of gold and bitcoin, Every day is a new prediction, so we build an investment transaction model based on dynamic programming to make decisions about gold and bitcoin every day to get the maximum benefit. The dynamic programming diagram is shown in Figure 9.

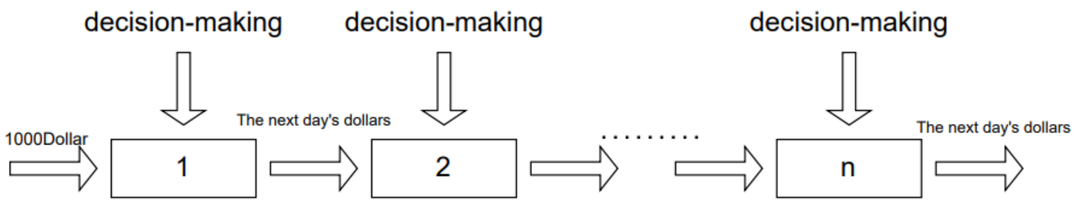

Figure 9. Dynamic programming

Through analysis, investment should have both risk and return, and an effective investment portfolio should be selected, that is, the investment portfolio with the greatest return with the lowest risk level. It means to pay attention to the risk of assets while paying attention to the return of assets. Through the ratio of investment portfolio return to risk, we can evaluate our daily return on cash, gold, and bitcoin portfolio. The mathematical formula is as formula (5): It can be considered that the larger the Sharpe ratio on that day, the better the return of the investment trading portfolio. on the contrary, the return will be worse.

$$SR_n = \frac{E(r_n) - r_f}{\sigma_n} \tag{5}$$

ISSN: 2414-1895 DOI: 10.6919/ICJE.202205_8(5).0060

In formula (5), SR_i represents the value of the Sharpe ratio of the investment trading portfolio on the nth day, $E(r_n)$ represents the expected return of the investment portfolio on the nth day, and $r_f$represents the return of the asset under the risk-free situation on the nth day, $\sigma_n$represents the standard deviation of returns.

2.3.1 Model Assumptions

On each trading day, the trader will have a portfolio [C, G, B] consisting of cash, gold, bitcoin, and we assume that on day n, the portfolio of cash, gold, bitcoin we own is C_n, G_n, B_n, we use c_n, g_n, b_n to represent the proportion of cash, gold and bitcoin in the assets of the day, then the relational formula is:

$$c_n + g_n + b_n = 1 \tag{6}$$

Assuming that on the n+1st day, compared with the nth day, the increase in the amount of assets owned by gold and Bitcoin is ΔG_n and ΔB_n respectively, so we get:

$$\Delta G_n = \frac{G_{n+1}}{G_n} \tag{7}$$

$$\Delta B_n = \frac{B_{n+1}}{B_n} \tag{8}$$

2.3.2 Establishment of Decision Variables

In the dynamic programming model, it is necessary to predict the value of gold and bitcoin on the nth day through the data as of today, and invest in existing assets through the increase in value, so that the amount of assets on the n+1th day is the largest, and the gold The change in the proportion of , and Bitcoin is set as the decision variable, Δg_n and Δb_n respectively:

$$\Delta g_n = \frac{G_{n+1}}{C_{n+1} + G_{n+1} + B_{n+1}} - \frac{G_n}{C_n + G_n + B_n} \tag{9}$$

The change in the proportion of Bitcoin is:

$$\Delta b_n = \frac{B_{n+1}}{C_{n+1} + G_{n+1} + B_{n+1}} - \frac{B_n}{C_n + G_n + B_n} \tag{10}$$

2.3.3 Determination of Objective Function

In order to get the maximum profit on the n+1th day, we adjust the position as of today, through the gold forecast value of the n+1th day, to buy and sell gold on the nth day, through the n+1th day The predicted value of Bitcoin on the nth day to buy and sell Bitcoin, we set the objective function to be the value of the Sharpe ratio on the n+1th day after the investment transaction on the nth day, and if and only if When the Sharpe Ratio is at its highest value, the day is in the lowest-risk, highest-reward situation.

ISSN: 2414-1895
DOI: 10.6919/ICJE.202205_8(5).0060

When the proportion of cash, gold, and Bitcoin on the nth day is $[c_n, g_n, b_n]$, then on the n+1th day, the proportion becomes $[c_n, (1+\Delta G_n)g_n, (1+\Delta B_n)b_n]$, after all changes, we normalize it, keeping the sum of the three proportions at 1 at all times.

Under the risk-free condition, so can get the maximum $MAX(H_n)$ of the n+1st holding value:

$$MAX(H_{n+1}) = C_{n+1} + M(m_n + \Delta m) + K(k_n + \Delta k) \tag{11}$$

With the various proportions, so can find the Sharpe ratio for the n+1st day.

2.3.4 Selection of Constraints

According to the meaning of the question, the analysis of the model and the operation of a large amount of data, we draw several constraints, and we convert them into mathematical language as follows:

1) The profit must exceed the commission

Through the topic analysis, we can conclude that "the commission for each transaction (purchase or sale) is α% of the transaction amount, we set $\alpha_{\text{Gold}} = 1\%, \alpha_{\text{bitcoin}} = 2\%$". In financial relationships, the money we make every day must exceed the commission, so we have:

$$\Delta H = MAX(H_{n+1}) - MAX(H_n) \tag{12}$$

$$\Delta\alpha = (1 \pm 1\%)\Delta g_n + (1 \pm 2\%)\Delta b_n \tag{13}$$

2) The proportion of cash to total assets is greater than or equal to 0

During the establishment of the model, we assumed that the cash held was not deposited in the bank, but kept by traders all the time, without any increase in cash, so the proportion of cash on the second day was:

$$c_{n+1} = c_n + (1 \pm 1\%)\Delta g_n + (1 \pm 2\%)\Delta b_n \tag{14}$$

3) The proportion of gold in total assets is greater than or equal to 0

Gold is only traded on the trading day, so we set today's gold ratio unchanged and continue to be g_n when gold does not trade on the second day. When gold continues to trade, then the gold ratio on the n+1st day is:

$$g_{n+1} = g_n - \Delta g_n + \Delta G_n g_n \tag{15}$$

4) The proportion of Bitcoin in total assets is greater than or equal to 0

In the question, Bitcoin can be traded every day, when the predicted value of Bitcoin on the next day becomes lower, we sell all Bitcoin today to reduce the loss on the second day, so on the second day, Bitcoin The proportion of the total value to the total value we have must be greater than or equal to 0. Therefore, the proportion of Bitcoin on the n+1st day is:

$$b_{n+1} = b_n - \Delta b_n + \Delta B_n b_n \tag{16}$$

ISSN: 2414-1895 DOI: 10.6919/ICJE.202205_8(5).0060

So get the following four constraints:

$$\begin{cases} MAX(H_{n+1}) - MAX(H_n) \geq (1 \pm 1\%)\Delta g_n + (1 \pm 2\%)\Delta b_n \\ c_n + (1 \pm 1\%)\Delta g_n + (1 \pm 2\%)\Delta b_n \geq 0 \\ g_n - \Delta g_n + \Delta G_n g_n \geq 0 \\ b_n - \Delta b_n + \Delta B_n b_n \geq 0 \end{cases} \tag{17}$$

2.3.5 Model Results

According to our planning model, based on the price data as of that day, the best daily trading strategy is given. We select five sets of data to show the number of gold and bitcoin investment transactions as shown in Table 5 below.

Table 5. Gold and Bitcoin daily trading volume

Serial number	Transaction date	Today's gold trading volume	Today's Bitcoin Trading Volume
1	2016.11.13	NO	2.384617
2	2016.11.14	1.093751	1.962831
3	2016.11.15	1.557263	2.075920
4	2016.11.16	-0.593157	2.371553
5	2016.11.17	-0.316932	1.259946

Through the above model operation process, combined with the previous prediction results, we obtain the adjustment strategy for the purchase and sale of gold and bitcoin every day, and obtain the total asset value of the day, until September 10, 2021, the initial $1,000 will rise to $182,916.539.

3. Conclusion

(1) The prediction model established in this paper adopts the ARIMA model, which is easy to operate and does not need to rely on other variables. In the process of establishment, we compare multiple models and analyze the AIC values of each model to select the most The optimal forecasting model makes the forecasting model more reliable.

(2) The dynamic programming model established in this paper introduces the Sharpe ratio, which combines risk and return, making it closer to reality, so as to better provide daily trading strategies. According to the above planning model, the initial $1,000 will rise to $182,916.539.

References

[1] BERRUT JP, TREFETHEN LN. Barycentric Lagrange interpolation[J]. SIAM Review,2004,46(3):501-517.

[2] CHEHELGERDI-SAMANI, MARYAM, SAFI-ESFAHANI, FARAMARZ. PCVM.ARIMA: predictive consolidation of virtual machines applying ARIMA method[J]. Journal of supercomputing,2021, 77(3): 2172-2206. DOI:10.1007/s11227-020-03354-3.

[3] SRIVASTAVA AKHILESH KUMAR, SRIVASTAVA ANAND, SINGH SIDDHARTHA, et al. Design of Machine-Learning Classifier for Stock Market Prediction[J]. SN Computer Science,2021,3(1). DOI: 10.1007/s42979-021-00970-5.

[4] DELSOLE, TIMOTHY, TIPPETT, MICHAEL K.. Correcting the corrected AIC[J]. Statistics & Probability Letters,2021,173. DOI:10.1016/j.spl.2021.109064.

[5] An analysis of the Hypervolume Sharpe-Ratio Indicator[J]. European Journal of Operational Research, 2020,283(2):614-629. DOI:10.1016/j.ejor.2019.11.023.