A Method of Trading Strategies for Bitcoin and Gold

Kelei Wu(✉), Yiyi Zhu, Die Shao, Xuan Wang, and Chenyuan Ye

Zhejiang Gongshang University, Hangzhou 310038, China
wukl0113@163.com

Abstract. The price of Bitcoin has soared from $10,000 in 2021 to an all-time high of $60,000 in 2022, with a total market capitalization of more than $1 trillion, and gold has been used as the ultimate asset to protect the economy and protect against inflation. Therefore, it is predicted that gold and Bitcoin Investors can gain more income by monitoring the price trend of the currency and formulating a reasonable investment strategy. In order to obtain a better prediction effect, this paper divides the analysis of gold price prediction into two stages: before the epidemic and after the epidemic. Before the epidemic, the CEEMDAN-WD-WOA-BiLSTM model was used, and after the epidemic, the neural network was used in this paper. For Bitcoin prediction, PCA is used to select indicators with greater impact, and the ARMIA model is used to predict. After the prediction, this article combines various situations, uses the bull and bear market model, pricing index strategy and quantitative evaluation model to decide the amount of gold and Bitcoin to buy, and finally the profit obtained by this method has reached more than 30 within five years times, to provide investors with a practical and effective investment plan.

Keywords: Predictive Model · Quantitative Trading Strategy · Bull and Bear Market Judgment Model · RSI · Bollinger Bands

1 Introduction

The market state of Bitcoin (https://en.wikipedia.org/wiki/Bitcoin) and gold (https://en.wikipedia.org/wiki/Gold) is a comprehensive economic indicator, and a growing number of experiments prove that accurate identification and prediction of Bitcoin and gold market state plays a crucial role in trading strategies and decision analysis. Scaling up the market state prediction of Bitcoin and gold to stock market state prediction, two main types of methods are applied to stock market state prediction, rule-based methods and Markov zone transition models. The rule-based approach is divided into two steps, identification and prediction, and generally uses the rule set of Lunde and Timmermann (2004) centered on the magnitude of ups and downs or Pagan and Sossounov (2003) centered on the duration of bull and bear market cycles to identify bull and bear markets, while the prediction uses Markov logistic regression models. Another approach based on the Markov zone transition model allows both identification and forecasting, using

N. Radojević et al. (Eds.): ICAID 2022, AHIS 7, pp. 186–197, 2023.
https://doi.org/10.2991/978-94-6463-010-7_20

the characteristics of bull and bear market returns and volatility to predict the probability of the occurrence of bull and bear markets using the EM algorithm [1]. We found that all kinds of methods cannot judge the market state without judging and predicting the probability of occurrence of bull and bear markets, so in this paper, we developed a bull and bear market judgment model to identify or predict what state the Bitcoin and gold markets are in.

For Bitcoin and gold investments, we take market risk into trading decisions. The main common risk assessment methods are VaR risk assessment method, β coefficient method, mean-variance selection theory and mean-below risk selection theory (Considering that portfolio investors are more concerned about the lower risk, and the semi-variance is consistent with the lower risk variance, the most perfect method in theory can be obtained. However, the calculation is too complex, which reduces the effect of practical application) [11]. Therefore, this paper builds a new risk model based on these common methods to assist us in making trading strategies.

2 Price Prediction Algorithm

Gold, Bitcoin prediction based on different methods proposed in this study includes the following four steps (Fig. 1):

Step 1: Data preprocessing.
Step 2: Predict the price of gold and Bitcoin.
Step 3: Appropriate quantification strategy.
Step 4: Model performance criteria.

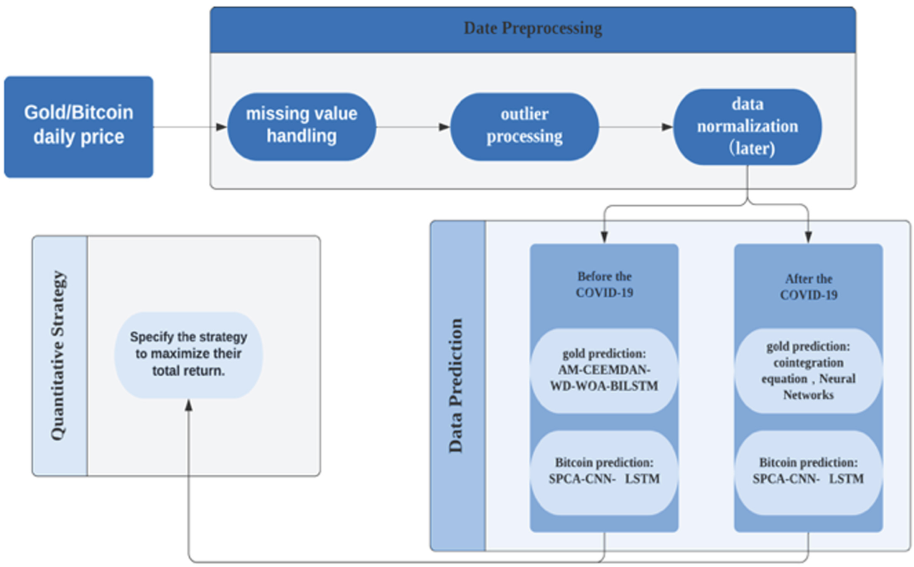

Fig. 1. Gold Bitcoin Solution Based on Historical Data

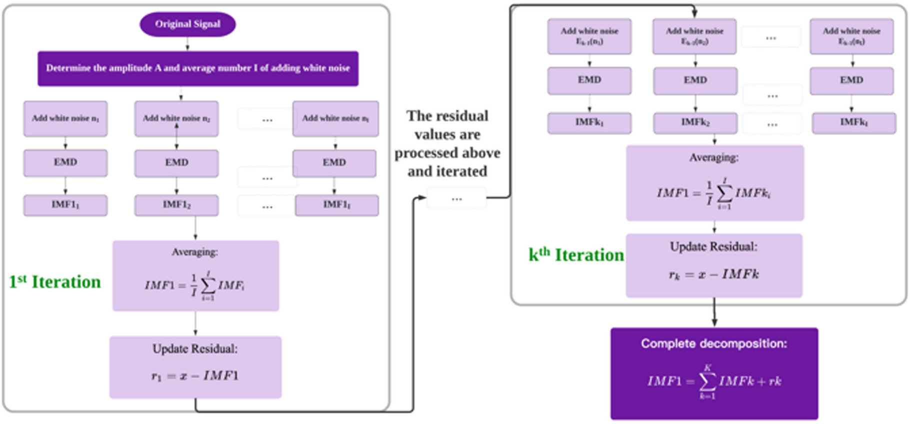

Fig. 2. CEEMDAN process

2.1 Gold Prediction Before COVID-19

Because of the complex formation mechanism of gold futures price, there are many factors affecting it, and it is difficult to find all the relevant characteristics. Therefore, starting from the gold futures price itself, this paper introduces the complete ensemble empirical mode decomposition method with adaptive noise [3] for signal processing, fully extracts the implied fluctuation information of the price sequence itself, decomposes a series of new features, and then uses the wavelet decomposition (WD) to perform the secondary decomposition of the high-frequency component [12], so as to more effectively extract the implied information in the original data. Finally, the bi-directional long-short term memory neural network BiLSTM optimized by WOA [2] is used to predict the gold futures price sequence.

2.1.1 Compete Ensemble Empirical Mode Decomposition with Adaptive Noise

The CEEMDAN method is shown in the Fig. 2.

2.1.2 Bi-directional Long Short-Term Memory

BiLSTM is a variant algorithm of LSTM, which consists of a layer of forward propagation LSTM and a layer of reverse propagation LSTM. The forward layer starts the input iteration from the beginning of the sequence, and the reverse layer starts the input iteration from the end of the sequence. Finally, the output results of the two layers are fitted to obtain the identification results. The network structure of BiLSTM is shown in Fig. 3.

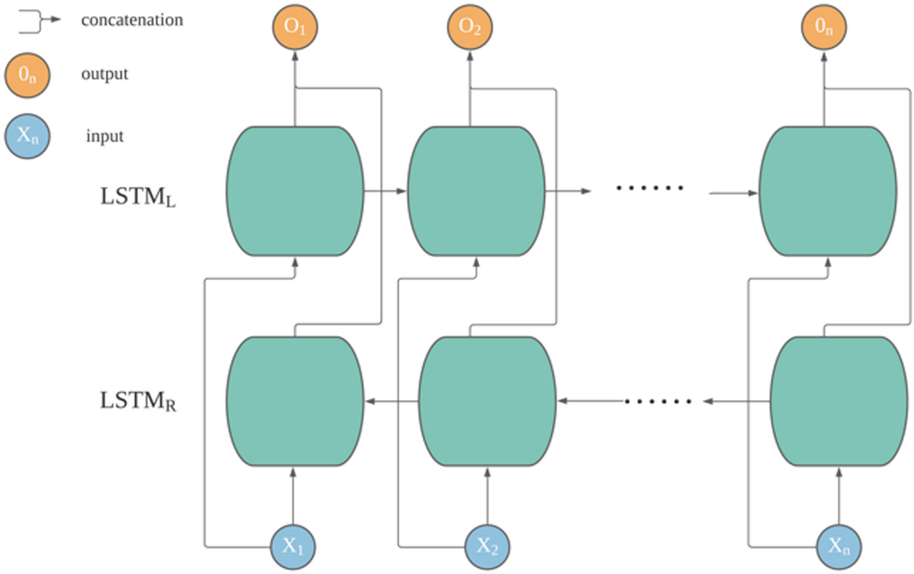

Fig. 3. The Network Structure of BiLSTM

BiLSTM can be expressed by the following formula.

$$h_t = f(W_1 x_t + W_2 h_{t-1}) \tag{1}$$

$$r_t = f(W_3 x_t + W_5 r_{t+1}) \tag{2}$$

$$y_t = g(W_4 h_t + W_6 r_t) \tag{3}$$

2.1.3 Parameter Optimization of BiLSTM Based on WOA

The prediction accuracy of BiLSTM is mainly affected by the number of hidden layer neurons m and the setting of time step c. At present, the number of hidden layer neurons can only be determined according to the following empirical formula, and the selection of the optimal number of neurons is often set by multiple tentative parameters of the number of neurons. In BiLSTM network, the time step c is the number of cyclic unit structures, indicating that the information at a certain moment can be synthesized by the information at most of the previous moments.

$$m = \sqrt{(\alpha + \beta)} + q \tag{4}$$

In the formula, α and β are the number of nodes in the output layer and the input layer, respectively, q is a constant between [0, 10].

The selection of m and c has an important influence on the accuracy of the final prediction results of the neural network. This paper optimizes the parameters m and c of BiLSTM loop network by whale optimization algorithm.

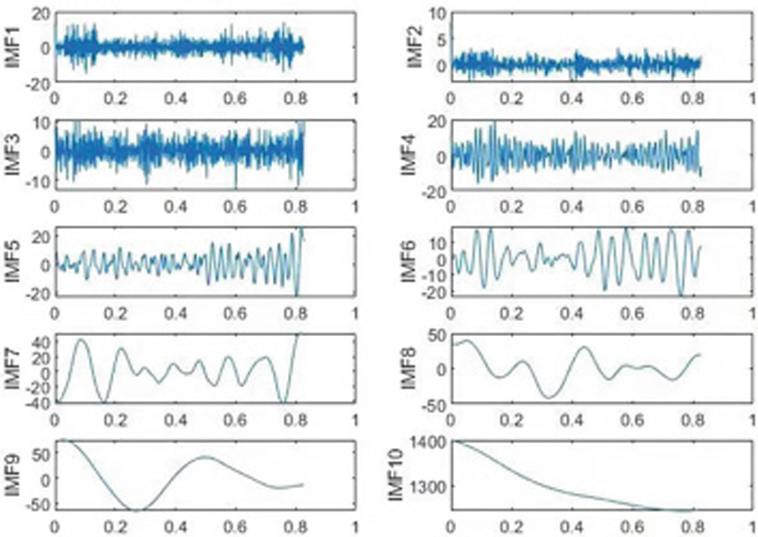

Fig. 4. CEEMDAN decomposition example diagram

2.1.4 Results and Analysis of Gold

Since the gold futures price is a nonlinear and non-stationary complex dynamic system, it is difficult to separate each component. In this paper, CEEMDAN is used to decompose the gold price. Each component can be understood as the impact of various factors on the gold futures price, and the high frequency component can be understood as noise. We further decompose the high frequency component and use the WD algorithm to process it.

CEEMDAN decomposition can decompose the original signal into several intrinsic mode functions and residual sequences with a single vibration mode. These new sequences extract the local characteristics of the original sequence, which is more helpful to grasp the implicit information of the data. Figure 4 is an example of the results of a decomposition.

The first three high-frequency variables are further decomposed by WD, and the effect after denoising is compared with the effect before denoising. It can be found that the latter is significantly better than the former. This paper has a total of 828 working days. The parameters are set as follows: epoch = 40, batch_size = 128, shuffle = False. The time span of the working day test set is from Sep. 12, 2016 to Sep. 30, 2019, and it is used as the training set.

In this paper, the prediction effect is shown in Fig. 5.

Therefore, it can be found that the algorithm used in this paper has excellent results.

2.2 Gold Prediction After COVID-19

First of all, we found that the new corona pneumonia epidemic had a great impact on the price of gold. On Dec. 30, 2019, the Chinese government announced the first case

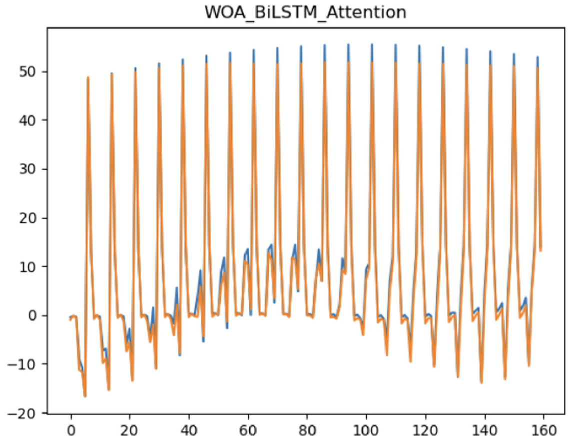

Fig. 5. BiLSTM Based on WOA Optimization Prediction Effect

of new coronavirus pneumonia. This report can be regarded as the initial starting point for the financial market to be affected by the epidemic. Therefore, this paper divides the data into two time series datasets: before and after the epidemic: the first is from Sep. 12, 2016 to December 30, 2019, with a total of 828 data. The second is from Jan. 1, 2020 to Sep. 10, 2021, with a total of 427 data.

We selected artificial neural networks for it.

2.2.1 Unit Root Test

First of all, it is easy to find that the gold price began to rise gradually from the beginning of 2020. By Aug. 2020, the gold price reached its peak. ADF unit root test $p = 0.4030$ shows that when $\alpha = 5\%$, this paper believes that gold sequence has unit root. Therefore, it is necessary to carry out the first-order difference of gold price. The difference of gold price is tested by ADF. The conclusion is that the original hypothesis is no longer accepted, and the gold price has the stability mentioned in the alternative hypothesis.

In reality, the price of gold and the new coronavirus are affected by the working day cycle, so the number of new cases also shows a cyclical upward trend. Similarly, we also conducted a unit root test on the number of newly diagnosed cases. The test showed that the number of confirmed cases does not have the stability mentioned in the alternative hypothesis. After the first-order difference, the difference of the number of cases is found to pass the ADF test, and $p \approx 0$ meets the stability requirements of the alternative hypothesis when $\alpha = 1\%$.

2.2.2 Cointegration and Error Correction Model

In order to analyze whether there is a cointegration relationship between the gold price sequence from Jan. 1, 2020 to Sep.10, 2021 and the number of newly diagnosed cases of the epidemic, this paper uses the EG two-step method to judge. Firstly, the price sequence of gold is used as the explained variable, and the number of confirmed cases

Table 1. Error Correction Model Test Results

Variable	coefficient	Std. Error	t-Statistic	Prob
DNEWCASES	1.86E−05	7.66E−05	0.242232	0.8089
C	1.922808	1.609564	1.194614	0.2338
RESID_EG2(-1)	−0.119548	0.031704	−3.770797	0.0002
R-squared	0.072779	Mean dependent var		1.903704
Adjusted R-squared	0.062809	S.D. dependent var		22.78749
S.E. of regression	22.062809	Akaike info criterion		9.041178
Sum squared resid	90517.80	Schwarz criterion		9.041178
Log likelihood	−851.3913	Hannan-Quinn criter		9.092634
F-statistic	7.299716	Durbin-Watson stat		9.062024
Prob (F-statistic)	0.000887			

is used as the explanatory variable, which is estimated by the least square method. The results are shown in Table 1.

The regression coefficient is obvious, so the estimated model is:

$$Gp_i = 1576.582 + 0.001234\mathrm{NC}_i + \mathrm{e}_i \tag{5}$$

In the formula, *Gp* is the gold price, and NC means the new cases.

Estimation error sequence can be obtained by this model:

$$\mathrm{e}_i = \mathrm{Gp}_i - 0.001234\mathrm{NC}_i - 1576.582 \tag{6}$$

The second step is to test the unit root of the estimated error e_i in the form of no trend term and no intercept term ($Y_i = \gamma Y_{i-1} + \epsilon_i$).

The above data shows that the two are in a long-term equilibrium in the system, but in the short-term, they may deviate from the equilibrium state. The temporary fluctuations of the spot gold price and the long-term trend are included together to get:

$$\Delta \mathrm{Gp}_i = \alpha + \Delta NC_i + \gamma Y_{i-1} + \epsilon_i \tag{7}$$

According to AIC, SC, HQ statistics, the error correction model results are shown as in Table 1.

At this time, it is obvious at the level of 5% dominance.

2.2.3 Neural Network Prediction Effect

After fully learning, the model is applied to the test set for testing. Under this condition, MSE and R2 had good effects, and the R2 value reached 0.9625. Neural network fitting gold price prediction has good effect, and the predicted value is close to the real value.

2.3 Bitcoin Price Forecast

2.3.1 Indicator Selection and Data Source

Indicator Selection and Data Source
The main factors influencing Bitcoin price level are Bitcoin itself and macroeconomic environment [6]. The price change caused by Bitcoin's own factors comes from the market supply and demand of Bitcoin and the factors of mining technology.

Macroeconomic Environmental Factors Selection
In general, the stock price index can reflect the impact of macroeconomic and financial developments, so the S and P 500 index and the daily yield on 30-year Treasury bonds are used as indicators. At the same time, in order to comprehensively analyze the influencing factors, two variables of gold price and oil price are added as indicators to measure the description of the global economy. Finally, this paper selects the search volume of Bitcoin keywords on Twitter as an indicator [5].

Selection of Dependent Variables
Since the dollar remains the largest currency held by central banks and the average daily trading volume of Bitcoin settled in the dollar ranks second, consider the dollar price of selected Bitcoin to reflect the price of Bitcoin.

2.3.2 Analysis of Price Trend Diagram

From Sep. 2016 to Sep. 2021, the price level of Bitcoin has a flat period and a period of rapid growth on the whole, and in some time periods, the price level of Bitcoin fluctuates greatly with time. Moreover, the price level of Bitcoin has no obvious specific rules, and the complex nonlinear trend structure contains obvious peaks and troughs.

In this paper, ARIMA is used to predict each index. The results show that the prediction effect is ideal, and the R^2 reaches 98.7%. Since it is similar to the above method, this paper will not repeat it.

3 Quantitative Trading Strategy

First of all, this article calculates the daily gains of gold and Bitcoin. This paper adopts the calculation of the 5-day increase; while the increase of gold is small and can be calculated based on the first 15 days.

This paper normalizes the deviation rate, average price, interest rate, increase and other indicators.

$$X_{norm} = \frac{\mathrm{X} - X_{min}}{X_{max} - X_{min}} \tag{8}$$

3.1 Bull and Bear Market Model

3.1.1 Stock Index Return Model

It is assumed that the random change process of the return of the stock index obeys the following two-zone Markov transition process:

$$r_t = \mu_1 S_t + \mu_0(1 - S_t) + [\sigma_1 S_t + \sigma_0(1 - S_t)]\varepsilon_t,$$

$$\mu_1 > 0, \mu_0 < 0; \sigma_1 > 0, \sigma_0 > 0. \tag{9}$$

Among them, ε_t is a list of independent and identically distributed random variables that obey the standard normal distribution. S_t represents the status of a bull and bear market, and when it is equal to 0, it means a bull market. The transition probability matrix for the two states is:

$$\begin{pmatrix} p & 1-p \\ 1-q & q \end{pmatrix} \tag{10}$$

p represents the probability that the stock market is in a bull market state and will continue to be in a bull market at the next moment, $1 - p$ represents the probability that the stock market is in a bull market state and will be converted to a bear market at the next moment; q represents that the stock market is in a bear market state and will continue at the next moment. Keeping the probability of a bear market, $1 - q$ represents the probability that the stock market is in a bear market state and will switch to a bull market at the next moment.

3.1.2 Empirical Analysis of Stock Index Bull-Bear Conversion

This paper selects the monthly data of Shanghai Composite Index and Shenzhen Component Index from June 2005 to June 2015 for empirical analysis. The stock index of the current month is represented by the closing points on the last trading day of each month, and the formula for calculating the rate of return is:

$$R_t = \frac{price_a - price_b}{price_b} \times 100\% \tag{11}$$

$price_a$ is closing price of the month, $price_b$ is closing price of the following month.

First, the MCMC method is used to estimate the model parameters, and the prior distribution of the parameters is:

$$\begin{aligned} &\mu_1 \sim N(3, 10)\mathrm{I}(0, \infty), \\ &0 \sim N(-3, 10)\mathrm{I}(-\infty, 0), \\ &\sigma_1 \sim \mathrm{IG}(0.1, 0.1), \sigma_0 \sim \mathrm{IG}(0.1, 0.1), \\ &p \sim \beta(5, 1), q \sim \beta(5, 1) \end{aligned} \tag{12}$$

This paper establishes a bull and bear market judgment model, through Bull Market Assessment Indicator = The average value of gold in the first 90 days * 0.666 + The first 90-day average of gold's 15-day deviation rate * 0.333.

This article uses the voting method to determine the time of bull and bear markets. If there is a bull market for a certain period of time, but yesterday is a bear market, and it is also a bear market tomorrow, then the value of the previous quarter will be increased by one, and if it is a bear market, it will be decreased by one. The final result is greater than 0 for a bull market, and less than 0 for a bear market. Get the final distribution map of the gold bull market. The same is true for the final distribution diagram of the Bitcoin bull and bear market.

3.2 Risk Model

When making forecast investments, we need to control the risks. Here, we introduce a risk model [7].

In this paper, the mean square error $e = \sqrt{(\mathrm{D(r)})}$ of the investment rate of return r is used to determine the total risk degree of investment. The total risk degree is composed of systematic risk degree and non-systematic risk degree.

Systematic Risk D_h. Systematic risk is caused by macro-level factors through changes in market conditions. The more intense the market changes, the greater the total risk, and thus the greater the risk of investment. D_M is used to represent the total risk, and D_H is the systemic risk, and we can get $(D_H)/(D_M) > 0$.

Unsystematic Risk D_W. For the investment of gold and Bitcoin, the degree of non-systematic risk is mainly related to the degree of deviation of its trend from the market. The higher the degree of deviation, the higher the degree of non-systematic risk. reduce, i.e. $(D_\omega)/(d^2) < 0$.

Gold buying risk is related to the gold deviation rate and the gold bull market. The gold bull market is given in this article as 0.666, and the deviation rate is 0.333. The buying risk map for gold is shown in Figs. 6 and 7. Bitcoin is the same.

3.3 Set Rating

This article simply sets the buy score as increase * 10 + bull market * 5 − residual + 1/purchase risk. Gold is the same.

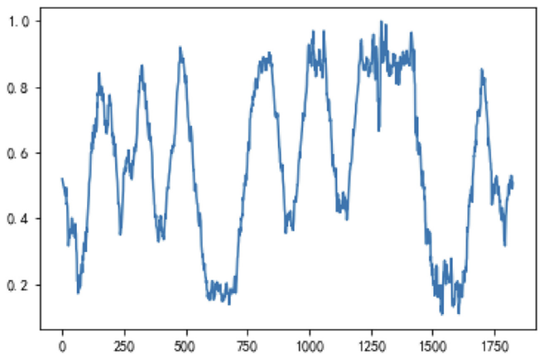

Fig. 6. Gold buying risk chart

This article sets that when the gold score is greater than 0.58 to buy and sell when it is less than 0.3, and the Bitcoin's indicators are 0.71 and 0.51. At the same time, we consider the trading rules as follows:

(1) Judging whether it is a gold trading day, gold is considered, not gold is not considered.
(2) Buy-in refers to the use of current cash * buy-in score * (1 − commission)/current price.
(3) Sale amount = holding share * (1 − score + sale standard).

If the effect is above 360k, it is better. Therefore, this paper finally solves the model and finds that the total assets will reach 3767 15.00 on Sep. 12, 2021.

4 Conclusions

In order to maximize the return on investment, this paper uses a variety of models to predict and quantify. In the prediction, we use the most appropriate model to predict gold and Bitcoin respectively. In terms of quantification, this paper adopts the quantitative trading model based on risk assessment to make the model obtain the maximum return under the minimum risk.

In the prediction, we use the bull and bear market judgment model to combine the average value of the first ninety days and the deviation rate of the fifteen days in proportion, and the index value is the classification standard of bull and bear markets. To quantify the predicted values, we compare our models with different models. The results show that our quantitative trading model based on risk assessment is significantly better than other models, and can obtain the maximum return under the minimum risk.

In the evaluation of trading strategies, this paper uses the RSI index and Bollinger bands index for decision analysis. A trading strategy is formulated by comparing the calculated RSI index with the proposed index interval, and another trading strategy is formulated by comparing the position relationship between the price line and the Brin line. Comparing the returns under the two trading strategies with the returns given by the development model, the proof of the best strategy is given.

References

1. Xu BH (2020) The identification and prediction of bull and bear markets in China. Nanjing University of Finance and Economics. https://doi.org/10.27705/d.cnki.gnjcj.2020.000173
2. Prakash DB, Lakshminarayana C (2017) Optimal siting of capacitors in radial distribution network using whale optimization algorithm. Alex Eng J 56(4):499–509
3. Yuan DF (2021) Gold futures price prediction based on CEEMDAN-PCA-LSTM model. Shandong University. https://doi.org/10.27272/d.cnki.gshdu.2021.005040
4. Gold as an investment. https://en.wikipedia.org/wiki/Gold_as_an_investment
5. Wu JR, Ding EJ, Cui R et al (2020) Methods of rotating machinery fault diagnosis using multi-scale attention mechanism. J Xi'an Jiaotong Univ 54(2):51
6. Jing PJ (2021) Comparative study of bitcoin price forecasting based on multiple models. Shanxi University. https://doi.org/10.27284/d.cnki.gsxiu.2021.000592

7. Yang SL (1998) Research on some problems of stock risk control. J Xiamen Univ (Nat Sci) (01):28–31
8. What is Bitcoin? https://en.wikipedia.org/wiki/Bitcoin
9. What is gold? https://en.wikipedia.org/wiki/Gold
10. What is RSI? https://en.wikipedia.org/wiki/RSI
11. Zheng YB (2007) Research on China's stock market system risk valuation. J Wuhan Univ Technol (Inf Manag Eng Ed) 01:122–126
12. Sang YF, Wang ZG, Liu CM (2012) Methods for selecting decomposition layers in wavelet analysis of hydrological series. Hydrology 32(4):1–7