Price Prediction and Investment Decision Model Based on LSTM

Jiaji Chen [1, #, *], Lintao Fang [1, #], Zhenghui Xu [1, #], Jie Yang [2]

[1] College of Harbour, Coastal and Offshore Engineering, Hohai University, Nanjing, Jiangsu Province, 210098, China

[2] Business School, NanJing University, Jiangsu Province, 210093&210023, China

* Corresponding author: 1369698828@qq.com

[#]These authors contributed equally.

Abstract. In today's market trading, quantitative trading has benefited investors through frequent trans- actions. Our team will operate at $1,000, forecasting by processing data of previous years and calculating the final earnings. For Problem 1, we based the prediction model of the LSTM neural network, trained the previous bitcoin and gold transaction price of the network algorithm, and predicted the trading market price of the next day. By comparison with the actual value, the model's accuracy in predicting gold reached 92.3%, and the accuracy of bitcoin reached 70.5%, which are good results. Next, our team used a dynamic programming model, with the bear-bull market index reasonable divergence rate as the decision variable, to establish a system of equations for analysis. At the same time, we conducted a linear regression analysis of risk and return and used the characterization of the residual between the predicted value and the actual value to calculate the weight as the target function to describe the investment strategy under different people's investment personalities. For Problem 2, we performed sensitivity tests for previous models, and we randomly selected ten sets of perturbation data for prediction within a range of 5% near the decision parameters. Finally, we find that if the investment scheme is perturbated, the maximum return of the target function will decrease, so the given scheme can be considered an optimal scheme. This paper uses LSTM neural network model for price prediction with extended memory function and good predictive. At the same time, through the suitable formulation of the bear market and bull market model and investment risk model, the rationality of decision-making is increased, and the scheme is optimal.

Keywords: LSTM Neural Network, Bull Market, Risk Prediction, Quantitative Trading.

1. Introduction

1.1. Background

With a goal to maximize the total return, more and more market speculators jump on the bandwagon of buying and selling volatile assets. But the fluctuating financial rates, like the growing and dropping rate of bitcoin and gold, troubles the traders a lot. Short-term fluctuations in financial markets are difficult to predict, and profits and losses are hard to control through technical operations [1]. Even if the short-term investment is highly profitable, if the compound interest effect is not produced after a certain period, its income is still limited, and once the loss will lead to short-term financial goals unachievable, resulting in certain risks. Meanwhile, with the tremendous capital movement of the financial giants, the market situation has become more unpredictable.

Hence, an accurate and highly improved model is in urgent need to predict current financial situations, and give the near perfect daily trade strategy accordingly to maximize revenue and minimize risk.

1.2. Problem Restatement

We've been requested by a trader to develop a model based upon the past stream of daily prices to date to determine each day whether the trader should buy, hold, or sell their assets in their portfolio. The given trading period is from 9/11/2016 to 9/10/2021.

To achieve our goals, specifically, we need to:

- Develop a model that gives the best strategy where the final assets are maximized, then prove it.
- Test the sensitivity of the strategy to the transaction cost, and impacts of the cost.
- Communicate our strategy, model, and results to the trader in a memorandum.

1.3. Assumptions

- Assume that cash holdings will grow at a fixed interest rate of 1% per year
- Assume that when you trade on day N, you know the price of gold/bitcoin up to day N, and buy at that price, and start calculating your return from day N+1.
- Assume that the rates for gold and bitcoin transactions are settled on the same day, and generate revenue from the next day.
- Assume that the relevant data reviewed are true and reliable
- Assume that relative importance is quantifiable.
- Assume that instability has a consistent effect on each indicator.

2. Indicator Calculation Model

In order to predict daily bitcoin and gold prices and plan the best daily strategy, we programmed new indicators by using Python. We first convert the time in the file provided to us to a date data type. This resulted in two tables of data (Table 1), and we stitched the two tables together [2]. The price of gold on a non-deal day is supplemented by the price on the previous deal day, and marks whether the current deal day is a gold deal day. Deal Day of 1 means it is a gold trading day, 0 means it is not a gold trading day.

Table 1. part of the date table

Date	USD(PM)	Value	Deal Day
2016-09- 11	NaN	621.65	0
2016-09- 12	1324.60	609.67	1
2016-09- 13	1323.65	610.92	1
2016-09- 14	1321.75	608.82	1
2016-09- 15	1310.80	610.38	1

Then we calculate the daily increase of gold and bitcoin (without distinguishing whether it is a gold trading day or not). And we calculate the average increase of the 5, 10, 15, 20 and 30 days and plot them on a graph for comparison (Figures 1 and 2).

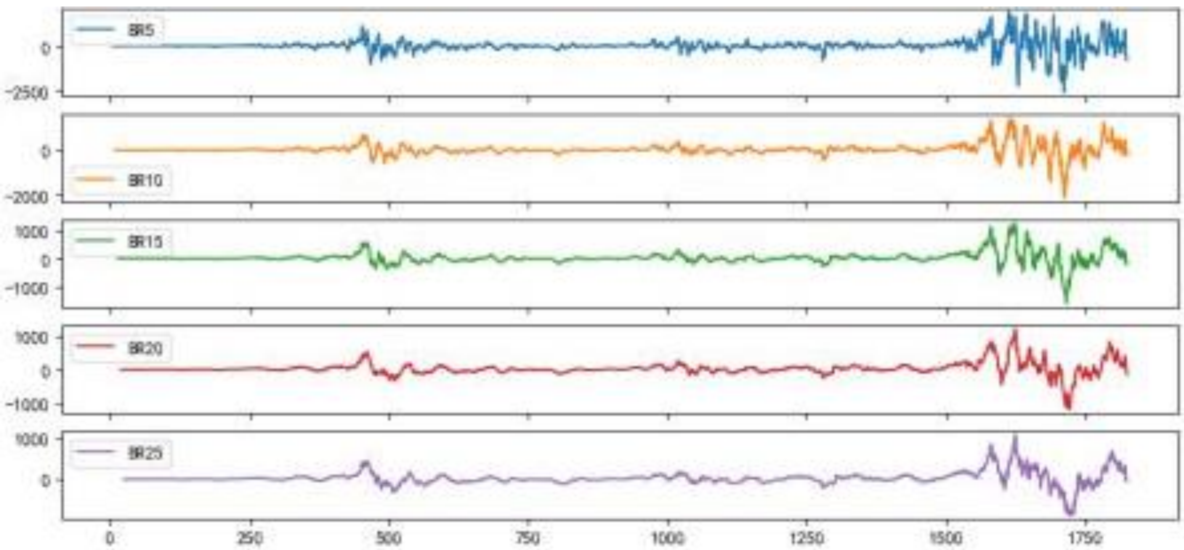

Figure 1. The average increase of bitcoin price

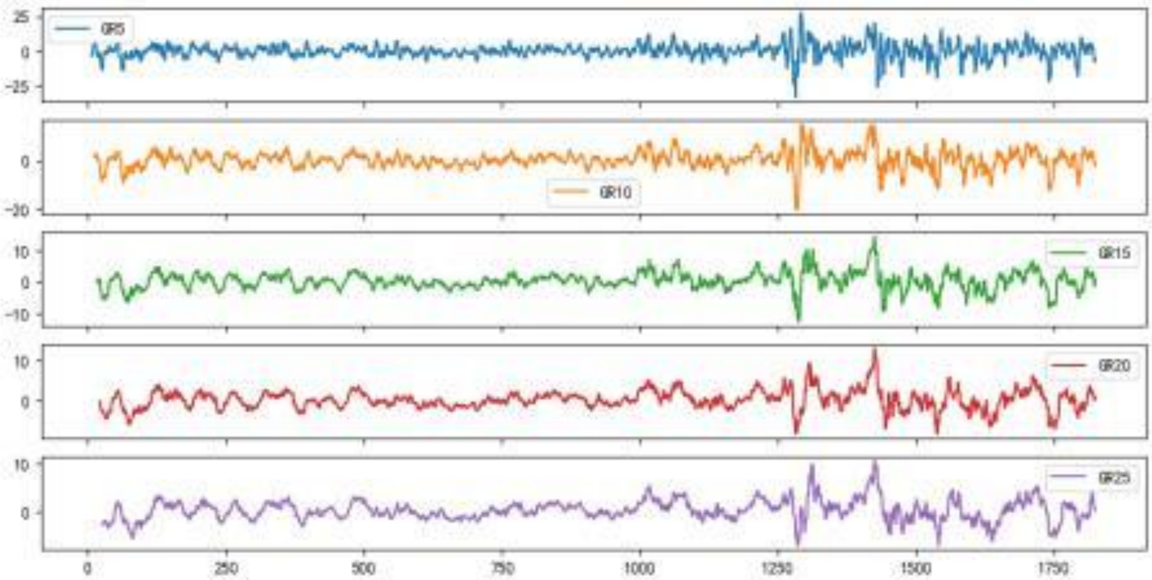

Figure 2. The average increase of gold price

From the above chart, according to the five-day increase in the calculation of the average value, the maximum increase of bitcoin is 2000. But according to the first ten days of the increase in the calculation of the average value, the maximum increase drops to 1000, where it's easy to miss the dramatic changes. So, we evaluate based on the five-day increase. And the gold price rise is small, so we evaluate according to the 15 days of the rise [3].

We then calculated the BIAS of the data. Specifically, we calculate the percentage difference between the closing price and the moving average, to reflect the degree of deviation of the price from its MA over a certain period of time.

Here is the formula:

$$BIAS = \frac{CP-AP}{AP} \tag{1}$$

Where:

BIAS is the n-day BIAS.CP is the current price. AP is the n-day average price. Then, we normalize the metrics that have been computed.

Here is the normalization formula:

$$nor = \frac{cur-min}{max-min} \tag{2}$$

Where:

Nor Represents normalized value, cur represents current value, min represents minimum value, max represents maximum value.

Table 2 and Table 3 are obtained by the above calculation.

Table 2. Part of normalization results of gold

Date	USD(PM)	Value	Deal Day	Price Increase	15-day Average Price	15-day BIAS
2016-09-11	0.00	621.65	0.0	0.566145	0.0	0.571622
2016-09-12	0.00	609.67	1.0	0.566145	0.0	0.571622
2016-09-13	0.00	610.92	1.0	0.561015	0.0	0.571622
2016-09-14	0.00	608.82	1.0	0.555886	0.0	0.571622
2016-09-15	0.00	610.38	1.0	0.507019	0.0	0.571622

Table 3. Part of normalization results of bitcoin

Date	USD(PM)	Value	Deal Day	Price Increase	5-day Average Price	5-day BIAS
2016-09-11	0.00	621.65	0.0	0.510080	0.000000	0.550392
2016-09-12	0.00	609.67	1.0	0.509356	0.000000	0.550392
2016-09-13	0.00	610.92	1.0	0.510146	0.000000	0.550392
2016-09-14	0.00	608.82	1.0	0.509949	0.000000	0.550392
2016-09-15	0.00	610.38	1.0	0.510146	0.009834	0.545356

After that, we use the increase data to build a bull market assessment indicator.

Here is the formula:

$$\mathrm{GInd} = \mathrm{AveTg1} \times 0.6 + \mathrm{AveTgbias} \times 0.3 + \mathrm{AveTg2} \times 0.1 \quad (3)$$

Where:

GInd represents bull market assessment indicator of gold, $AveT_{g1}$ represents first 90-day average of gold gains, $AveT_{gbias}$ represents first 90-day average of gold 15-day BIAS, $AveT_{g2}$ represents first 15-day average of gold gains. The numbers 0.6,0.3 and 0. 1 are the weights set by us, and have been checked by AHP.

A bull market is defined as greater than the mean value of the bull market assessment indicator, while a bear market is defined as less than that [4].

Then, we use voting method of ensemble learning to determine time of bull and bear market based on indicators. The voting method is an integrated learning model that follows the principle of minority rule to reduce the variance by integrating multiple models, thus improving the robustness and generalization ability of the model.

We use the voting method of prediction to obtain the prediction with the highest number of occurrences in the model prediction results.

For example, in the model, if today's market is defined as a bull market based upon the indicator but yesterday's is calculated as a bear market and tomorrow's is also calculated as a bear market, it is possible that the error of the result calculated today is larger. To solve this error, we set the initial value as 0 for all times, and if the current calculation obtains a bull market, the value of the previous quarter is added by 1, and if it is a bear market, it is subtracted by 1. The final result is greater than 0 for a bull market and less than 0 for a bear market. According to this method, we finally draw the final distribution map of the bull market for gold through calculation (Figure 3).

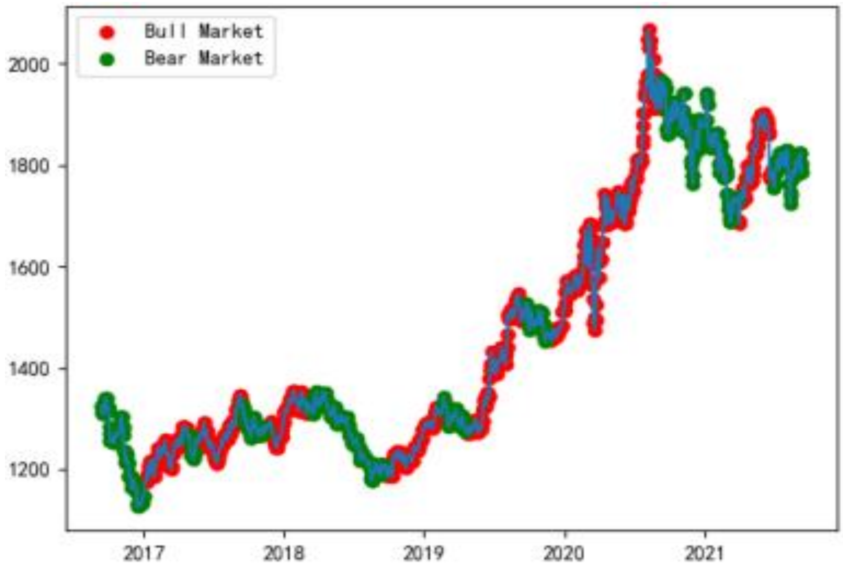

Figure 3. Gold bull market final distribution chart

Approximately, we can obtain the formula for the bitcoin bull market assessment indicator:

$$\mathrm{BInd} = \mathrm{AveTb1} \times 0.6 + \mathrm{AveTbbias} \times 0.3 + \mathrm{AveTb2} \times 0.1 \tag{4}$$

Where:

BInd represents bull market assessment indicator of bitcoin, $AveT_{b1}$ represents first 30-day average of bitcoin gains, $AveT_{bias}$ represents first 30-day average of bitcoin 5-day BIAS, $AveT_{b2}$ represents first 5-day average of bitcoin gains [5]. The numbers 0.6,0.3 and 0. 1 are the weights set by us, and have been checked by AHP (Figure 4).

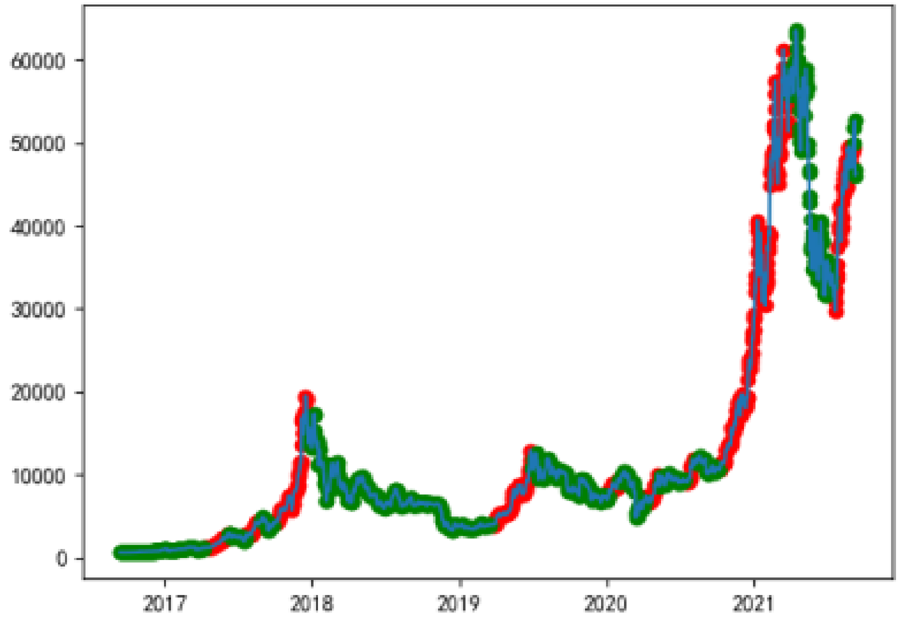

Figure 4. Bitcoin bull market final distribution chart

And, based on the bullish indicator and the deviation ratio, we can calculate the purchase risk ratio. The formula is as follows:

$$\mathrm{Risk} = \mathrm{Ind} \times 0.7 + \mathrm{Averbias} \times 0.3 \tag{5}$$

Where:

Risk represents purchase risk ratio. Ind represents bull market indicator. For gold, Averbias represents gold 15-day BIAS. But for bitcoin, Averbias represents bitcoin 5-day BIAS. The numbers 0.6 and 0.3 are the weights set by us, and have been checked by AHP.

Then we can get two purchasing risk charts (Figures 5 and 6).

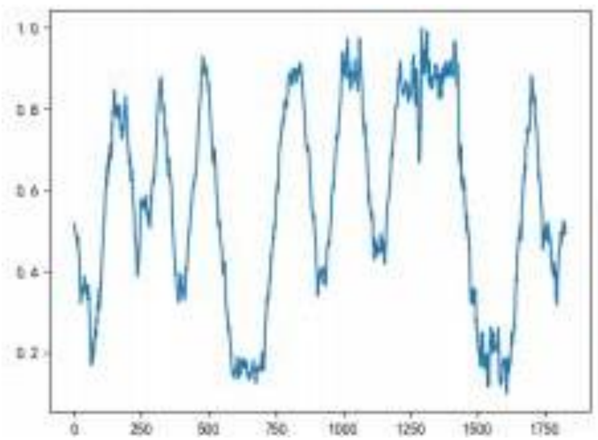

Figure 5. Gold purchasing risk chart

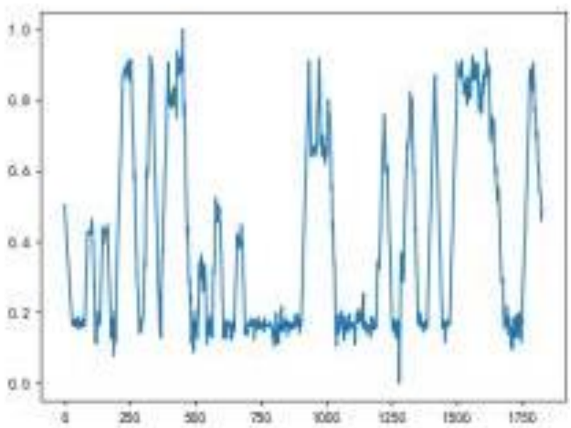

Figure 6. Bitcoin purchasing risk chart

3. LSTM-based Price Prediction Model

3.1. Analysis of the Predicting Problems

Among the forecasting methods for the financial time series of stock prices, there are generally two types. One is the financial time series forecasting models based on machine learning algorithms and the other is the financial time series forecasting models based on statistical methods. Among the categories of statistical methods, ARIMA and GARCH are the most widely accepted, but the primary premise of these methods is the assumption of linearity for the variable of financial time series, which is inherently very complex [6]. Financial time series arises from the economic and financial activities of people. Naturally, there will be obvious perturbations in the data, which make the financial time series not meet the assumption of linearity and smoothness. Therefore, traditional statistical methods are often ineffective in the practical application to the financial market. At the same time, machine learning algorithms have been widely used in financial time series and have been proven to have better predictive effects. Among them, RNN has the problem of long-term dependence in practical application. But, as one of the variants of RNN, LSTM effectively solves the long-term dependence problem of RNN and has good performance in such problems as speech recognition, image recognition and time series [1]. Therefore, considering the above, we use LSTM neural network for prediction.

3.2. Data Pre-processing

The training set for the experimental data is the daily U.S. gold settlement price from 7/3/2007 to 4/27/2018, and the daily bitcoin settlement price from 17/3/2012 to 2/7/2018. The validation is then performed using the period from 9/11/2018 to 9/10/2021 given in the question. Prices have different metrics, such a situation can affect the results of data analysis. In order to eliminate the influence of metrics between prices, data normalization is needed to solve the comparability problems between data. Then, for gold and bitcoin prices, we use today's price/yesterday's price instead of today's price for subsequent forecasts, which makes the data more comparable.

3.3. Training Network

Data used in the LSTM neural network should be the time-series features of the time series data to be extracted. We set the window length to 20 days, the rolling window length to 1 day and the prediction period to 1 day [6]. This means that we will predict the stock price 1 day later by consulting the data of the previous 20 days based on the current time point.

LSTM is a recurrent neural network suitable for extracting time-series features from time series and has the ability to learn long-term time series dependencies. The structure of the LSTM is shown in Figure 7. Generally, LSTM consists of input gates i_t, forgetting gates f_t, and output gates O_t. Ct is the state of the current cell, h_t is the state of the hidden layer, and x_t is the input data.

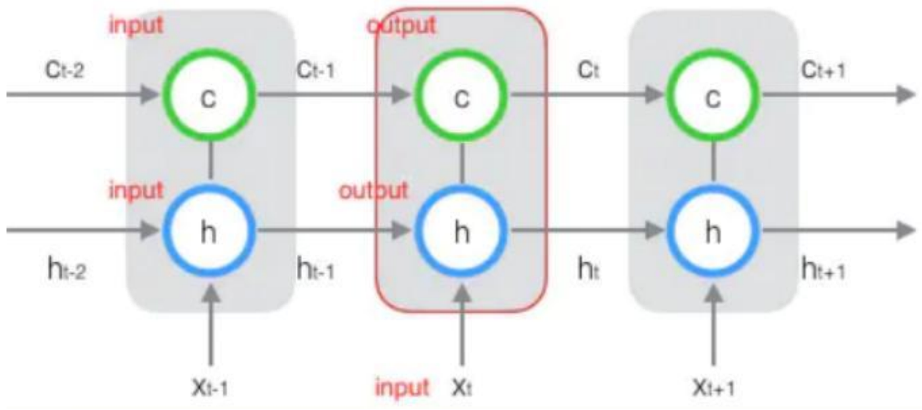

Figure 7. LSTM forward calculation chart

f_t determines how much of the cell state C_{t-1} from the previous moment is retained to the current state C_t. f_t uses an S-type function that compresses each value from a cell state between zero and one. In this case, a value close to zero means it should be discarded, while a value close to one means it should be kept.

it determines how much of the input xt to the network is saved to the cell state Ct at the current moment. it is actually two layers, the S-shape layer and the tanh layer. S-shape takes the previous hidden state and the current input, and selects the value to be updated. tanh layer also takes the hidden state and the current input, and then compresses the values to minus one and one. Then the outputs of the S-shaped and tanh layers are multiplied together [7].

Ot controls how much of the unit state Ct is output to ht. Ot takes the previous hidden state and the current input, and make them pass through the S-type layer. Then, your newly updated cell state will be distributed through the tanh function. Similar to it, the output of the S-type layer is multiplied by the output of the tanh layer, which ultimately determines what the output will contain [2].

The training set is trained by using the LSTM neural network model with the following parameter settings:

Time step. This parameter is where the LSTM neural network model is distinguished from other models. LSTM requires the input data of the price in the training data input. We input the time step 1, the length of the loop unit, into the initial settings. Each time, we only predict the data of the next one day, and then use all the data to predict the next one day of the next.

● LSTM units. Similar to other neural network models, LSTM neural network also needs to define the number of neurons for each neural layer. And we should take the input data shape into account when defining the number of neurons. Then, we should make the number of neurons in the LSTM layer slightly larger than the shape of the input data to ensure that the model can run properly. But it should not be set too large, otherwise it will have an impact on the performance of the model. Considering the above problem and the input factor 1, we initially define the number of neurons in the LSTM layer as 20.

● Optimizer. The purpose of the optimizer is to minimize the loss function. In all the neural network optimizers, Adam is the most commonly used optimizer, so we set the initial optimizer as Adam.

● Input dimension. Because only gold price / bitcoin price can be used for investment prediction according to the question, so we set the input data dimension as 1-dimensional.

3.4. Predicted Results

Depending on the model we have trained, we can predict the next one-day data based on data up to today, so that we can adjust our portfolio based on the results.

Based on our prediction results, we predict the price trend for the next one day, and then repeat the operation for each day, and select the results for one day. In order to verify the accuracy of the forecasting model, we can calculate the error using the results of day N+1. Thus, we draw the error rate charts and comparison charts of predicted value and true value (Figures 8 and 9). As we can see, the forecast error is roughly 0.99 for gold (the closer to 1, the better) and 0.92 for bitcoin, so the forecast is accurate [8].

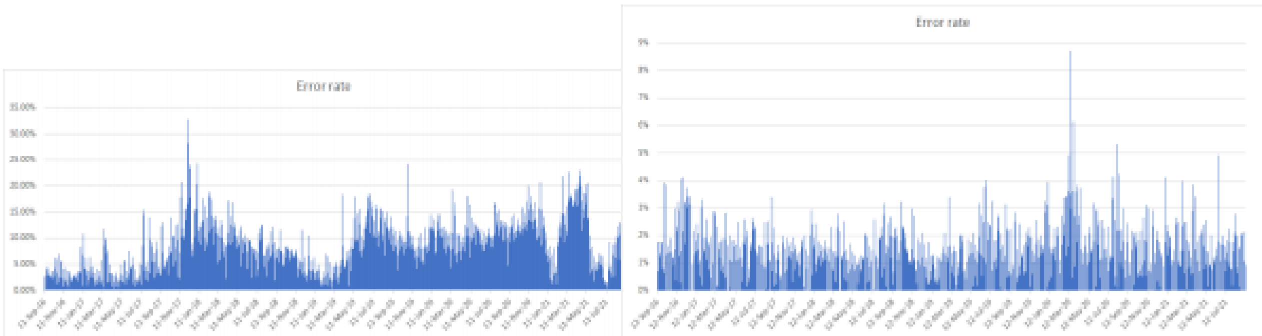

Figure 8. Error rate of gold and Error rate of bitcoin

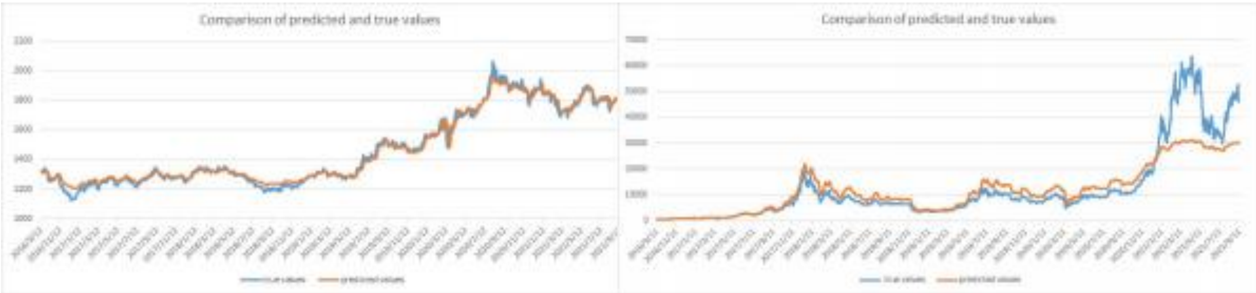

Figure 9. Prediction error of gold and prediction error of bitcoin

4. Model Improvement

In the process of the above model, when we adopt the buy score as the objective function, the investment strategy is somewhat conservative, resulting in a large risk threshold of bitcoin. Therefore, when the variance of bitcoin is large, the program does not adopt the investment strategy, which leads to the limitation of trading [9].

In fact, for investors, different personalities lead to different investment strategies. In addition to using the buy score as the objective function, there are other ways to characterize different people's mental expectations of risk and return [10]. We set three types of people as aggressive (considering less risk), stable (using buy score as the basis for investment), and intermediate (considering risk and return to have other weights), and develop different investment plans.

In the model building, for the aggressive personality, we weight the bitcoin purchases higher; for the stable personality, we weigh more on the more stable gold; and finally, for the intermediate personality, we perform calculation according to the original scheme. Ultimately, the total value of investment for the aggressive type is 31841464.05, for the inter- mediate type 771877.79, and for the stable type 4253.99.

References

[1] Wu Haiping and Liu Hui. Non-intrusive load transient identification based on multivariate LSTM neural network and time series data augmentation [J]. Sustainable Energy, Grids and Networks, 2021, 100490.

[2] Shen Feifan et al. Online local modeling and prediction of batch process trajectories using just-in-time learning and LSTM neural network [J]. Journal of Computational Methods in Sciences and Engineering, 2020, 20 (3): 1 - 12.

[3] He Xingqiang, Zhou Kaiguo. Bull and bear market cycles and the cycle synergy between stock markets [J]. Manage the world, 2006 (04): 35 - 40. DOI: 10. 19744/j.cnki.11-1235/f.2006.04.0 05.

[4] He Bate, Kita Eisuke. The Application of Sequential Generative Adversarial Networks for Stock Price Prediction [J]. The Review of Socionetwork Strategies, 2021, 15 (2).

[5] Zheng Minggui, Cao Tianqi, Zeng Jianlin. Research on the Influencing Factors of International Gold Futures Prices [J]. Gold Science and Technology, 2021, 29(04): 510 - 524.

[6] Han Xiaolin, Niu Danchen. Analysis of the influencing factors of gold price [J]. Wealth Life, 2019 (02): 109.

[7] Chen Zheshi. Research on Internet and Financial Futures Influencing Factors and Prediction Methods of Bitcoin Price [D]. Harbin Institute of Technology, 2021. DOI:10.27061/d.cnki.ghgdu.2021.000149.

[8] Liu Lijun, Liang Guopeng, Wang Haitao. Research on Quantitative Trading Strategy System Model [J]. Shanghai Business, 2021 (09): 158 - 159.

[9] Florensa Marcelo, Giuliodori David, Rodriguez Alejandro. The price elasticity of casino gambling: a case study in Argentina [J]. Applied Economics, 2022, 54 (56).

[10] Umezawa Masashi. Behavior-based price discrimination in a horizontally and vertically differentiated duopoly with switching costs [J]. Information Economics and Policy, 2022, 61.