Automation and Machine Learning (2022)
Clausius Scientific Press, Canada

DOI: 10.23977/autml.2022.030104
ISSN 2516-5003 Vol. 3 Num. 1

Portfolio Investment Algorithm Based on LSTM-ARIMA Prediction Model

Xianzhuo Chen[a], Xi Wei[b], Zhitian Su[c]

Heilongjiang University, Harbin, 150006, Heilongjiang, China
[a]cxz010716@163.com, [b]819424305@qq.com, [c]3488836497@qq.com

Keywords: Trading Strategy, LSTM-ARIMA Hybrid Model, Particle Swarm, Investment Risk, Multi-Objective Programming

Abstract: When investing in bitcoin and gold, we need to be backed by predictions of future prices, which only require predictions of price data as of today. First, the ARIMA model is used for data prediction, but this model has certain limitations when dealing with nonlinear data, so the prediction results are not ideal. The LSTM model is a recurrent neural network model that can handle nonlinear data. Therefore, in this article, I use a combination of linear and nonlinear prediction models, the LSTM-ARIMA hybrid model, to predict the price of bitcoin and gold. According to the prediction results, the prediction error of the LSTM-ARIMA hybrid model has better results. Our LSTM-ARIMA hybrid model is able to predict the price of gold and bitcoin in the next 10 days. It uses the mean of predicted prices as price expectations and the variance of predicted prices as investment risk. In addition, we consider the expected risk of the forecasting model as investment risk. To determine portfolio investment strategies, we maximize post-investment value, minimize risk, and make investment decisions every day as the next day's asset holding state. In order to solve the optimal portfolio investment strategy, we use the particle swarm algorithm to solve the problem, which can ensure that each step of the solution is a local optimal solution. Finally concluded that the initial \$1000 investment can be worth \$130948420399.063 as of September 10, 2021. Next, we analyze the sensitivity of the portfolio investment strategy model to gold and bitcoin transaction costs. Plot the cumulative returns when αgold=1%, αbitcoin is 2%, 2.5%, 3%, and αbitcoin=2%, and αgold is 1%, 1.5%, and 2%, respectively. By observing the cumulative return graph, it can be found that the model has greater feedback and higher sensitivity to changes in bitcoin transaction costs, and the increase in transaction costs has a greater impact on the final return. However, the model is less sensitive to the transaction cost of gold and has strong model stability. Finally, we can draw the following conclusions: 1. The LSTM-ARIMA hybrid model outperforms the traditional time series prediction model in correlation coefficient prediction. 2. The particle swarm algorithm can converge to a local minimum at each decision. 3. With the support of prediction model and particle swarm algorithm, each step of the optimization of the portfolio investment model can provide the optimal solution, thus ensuring the upward trend of the overall income. 4. Prove that the model provides the best portfolio strategy through economic indicators such as the Sharpe ratio.

1. Introduction

With the increasing prosperity and development of the capital market, the global capital market is interconnected, and a large number of traders flood into the market, who hope to achieve the goal of wealth accumulation by buying and selling volatile assets. Determining specific trading strategies for trading at a particular time is a critical research problem in financial market trading [1]. A good trading strategy can enable traders to allocate funds reasonably, balance investment returns and investment risks. Since the market is affected by uncertain factors such as political, social, and economic environment, it is difficult for traders to give a certain value to the returns, risks, and volatility of assets such as gold and bitcoin. Therefore, while market investment brings high returns, its risks may also cause serious economic losses and other negative impacts.

How to make trading decisions in a complex and volatile market environment has always been a difficult problem for traders and financial analysts. Therefore, it is essential to rigorously and accurately predict the trend of asset flow and to determine the best portfolio trading strategy according to transaction costs, to deploy asset investment more effectively [2].

2. Assumption

Considering that real problems always contain multiple complex factors, we need to make reasonable assumptions to simplify the model:

Assumption 1: Traders use the return rate to measure the overall level of the future real return rate, and use the variance of the return rate to measure the uncertainty of the return rate (i.e., the risk rate). Therefore, traders only care about the expectation and variance of the return rate in their decisions.

Assumption 2: Traders have a greedy and risk-averse mentality, i.e., traders always want the higher return, the better, and the smaller the variance, the better.

Assumption 3: The trading market has a relatively stable trading environment.

Assumption 4: The missing dates in the file "LBMA-GOLD.csv" are the dates when the market is closed, that is, the dates on which the market is open for trading are not lost.

Assumption 5: The data of the datasets are considered reliable. We can use the historical prices in the dataset to predict future price fluctuation trends.

Assumption 6: The first day the market is open for trading, we do not buy any assets. Prediction models of asset prices need data from historical trading open days to make predictions about future asset prices.

3. Model Establishment and Solutions

In this section, the time series forecasting model is used to forecast the price of assets on each day. The multi-objective programming method is used to establish a portfolio investment model.

3.1. Predictive Models for Asset Transaction Prices

After analyzing the provided dataset, we can guarantee that the closing price of a troy ounce of gold and the price in U.S. dollars of a single bitcoin are predictable. Because the fluctuation trend of asset prices may be linear or non-linear, this paper adopts the hybrid model of Long-Short Term Memory (LSTM) and Autoregressive Integrated Moving Average model (ARIMA)as the prediction model. The hybrid model would satisfy the requirement to use only the daily price to date to determine whether a trader should buy, hold, or sell a portfolio asset as shown in Figure 1.

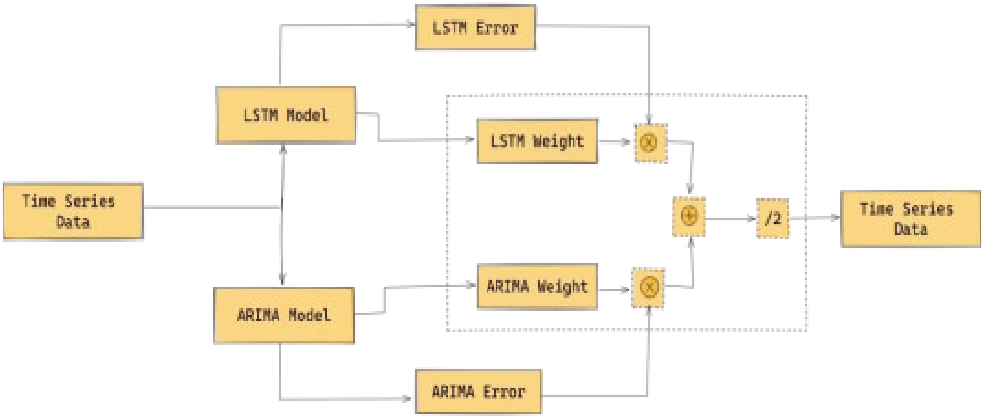


Figure 1: Workflow for Hybrid Predictive Models

3.1.1. Model 1: Hybrid model of LSTM and ARIMA

Many time series models include linear and nonlinear relationships. Although the ARIMA model is good at modeling linear relationships in time series, it is not good at modeling nonlinear relationships. LSTM can be used for both linear and Non-linear modeling[3], but the prediction results are not easy to reproduce. Therefore, in order to ensure the best prediction results, we combine LSTM and ARIMA to establish a hybrid model to model the linear and nonlinear components of the time series respectively.

The time series forecasting formulation for a hybrid model can usually be expressed as the sum of the linear and nonlinear components, as shown in Equation (1).

$$y_t = L_t + N_t \tag{1}$$

In which, L_t represents the linear component of the time series, and N_t represents the nonlinear component. In the hybrid model, the linear component Lt is first calculated by the ARIMA model, and then the nonlinear component Nt of the time series is predicted by the LSTM model. Finally, calculate the sum of the error values for both models. The formulas for L_t and N_t calculations are given in Equations (2) and (3).

$$LSTM_{error} = LSTM\_mean[error] \tag{2}$$

$$ARIMA_{error} = ARIMA\_mean[error] \tag{3}$$

The weights of the model are calculated by using the error values obtained in equations (4) and(5).

$$LSTM_{weight} = \left(1 - \left(\frac{LSTM_{error}}{LSTM_{error} + ARIMA_{error}}\right) \times 2\right) \times 2 \tag{4}$$

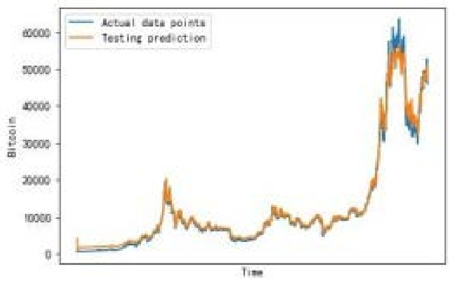
$$ARIMA_{weight} = 2 - LSTM_{error} \tag{5}$$

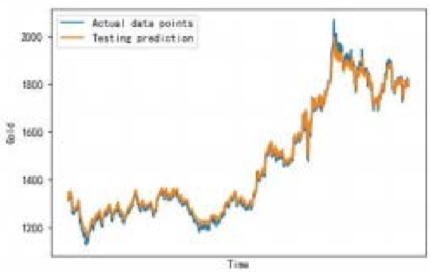
Use the given equation (6) to obtain the weight values of the model and each prediction value of the final mixture model.

$$Hybrid_{predict}[i] = \frac{LSTM_{weight}[i] \times LSTM_{error}[i] + ARIMA_{weight}[i] \times ARIMA_{error}[i]}{2} \tag{6}$$

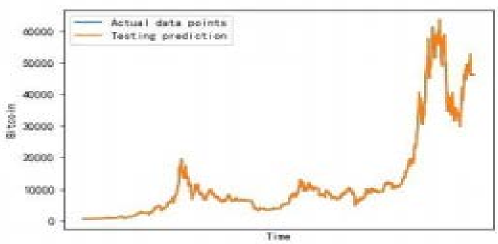
3.1.2. The Root Mean Squared Error of Model Prediction Results

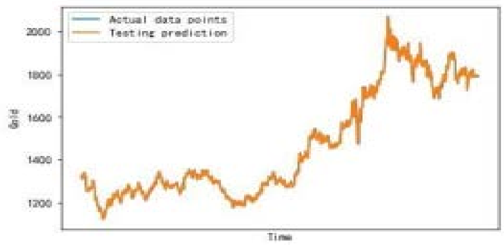
For both the ARIMA model and the LSTM-ARIMA hybrid model, future asset prices are predicted based on only knowing the price to date. That is, the data of the current time series is used to predict the data of the next day. If you want to forecast data for multiple days, you need to base the forecast data as the next day's data, and make repeated forecasts. And the mapped the data serves as the predicted price of the asset. The comparison between the prediction results of the ARIMA model and the mixed model and the real data is shown in Figure 2. Observing the above two sets of comparison charts, it is concluded that the predicted price change trend of gold and bitcoin prices by the mixed model is closer to the change trend of the real price than the predicted price of ARIMA.





(a) (b)





(c) (d)

Figure 2: Comparison of model predictions with the real price of the asset ((a) Comparison of the predicted

Results of the ARIMA model and the real price of bitcoin; (b) Comparison of the predicted results of the

ARIMA model and the real price of gold; (c) Comparison of the predicted results of the LSTM-ARIMA model

And the real price of bitcoin;(d) Comparison of the predicted results of the LSTM-ARIMA model and the real Price of gold.)

To measure the effectiveness of the model predictions, we use the mean squared error RMSE as a criterion to compare the predicted price of an asset with the actual price. RMSE is the square root of the ratio of the squared sum of the deviations of the observations to the true value to the number of observations (M)[4], which is defined as the following equation.

$$RMSE = \sqrt{\frac{1}{M}\sum_{i=1}^{M}\left(P_i - \hat{P}_i\right)^2}$$

In which, P_i is the predicted price and $\hat{P}_i$ is the actual price.

RMSE can measure the deviation between the observed value and the true value. The smaller the value of RMSE, the smaller the deviation between the observed value and the true value, and the more accurate the observed value. It is calculated that the RMSEs of the LSTM-ARIMA hybrid model for the predicted prices ofbitcoin and gold are 227.819 and 14.315, respectively; the RMSEs of the ARIMA model for the predicted prices of bitcoin and gold are 418.967 and 23.979, respectively.

Obviously, the prediction effect of the LSTM-ARIMA hybrid model is better. Therefore, in the modeling process of the portfolio investment model, we will use the prediction result of the LSTM-ARIMA hybrid model as the investment price of the asset.

3.2. The Construction of Portfolio Investment Strategy Model

The mean-variance analysis method in portfolio theory is an important method to study portfolio relationship[5]. It uses the expected return of the asset to represent the return of the asset, and the variance of the return of the asset represents the risk. Therefore, after obtaining a valid dataset, we can assess the expected returns and risks of investing in gold and bitcoin from these two dimensions.

3.2.1. Portfolio Investment Model Based on Multi-objective Programming

Denote a portfolio of cash, gold and bitcoin as, where$h_{(t)} = [c_{(t)}, g_{(t)}, b_{(t)}]$, $c_{(t)}, g_{(t)}, b_{(t)}$are cash, The holdings of gold and bitcoin, the calculation formula of cash holdings $c(t)$ is shown in equation (7) Where C is the commission generated by the asset during the transaction.

$$c(t)=c(t-1)-P_{gold}(t-1)\left[g(t-1)-g(t)\right]-P_{bitcoin}(t-1)\left[b(t-1)-b(t)\right]-C \quad (7)$$

$$C = abs\left(g(t-1)-g(t)\right)*\alpha_{gold} + abs\left(b(t-1)-b(t)\right)*\alpha_{bitcoin} \quad (8)$$

The risk ofportfolio investment strategy mainly comes from two aspects: the expected variance of the asset's return and the variance of the asset's forecast model. In order to measure the error of the prediction model on the real data, we introduce the expected risk as the variance generated by the prediction model. This paper chooses the value of empirical risk to approximate the expected risk, which is obtained by calculating the average error between the prediction results of the prediction model and the obtained training samples

For a given training set $D=\left\{\left(x^{(n)}, y^{(n)}\right)\right\}_{n=1}^{N}$, assuming that $f(x, \theta)$ is a function with Parameter θ, which is a trained prediction model, the empirical loss obtained on the training set is:

$$\delta = \frac{1}{N}\sum_{n=1}^{N}\tau\left(y^{(n)}, f\left(x^{(n)};\theta\right)\right) \quad (9)$$

It should be noted that the expected risk is related to the training set, even if sampling from the same true distribution, collecting different data sets will have different expected risks. The formulas for calculating the rate of return risk and expected risk at this time are as shown in equation (10) and(11).

$$\sigma = \sigma_{gold} + \sigma_{bitcoin} \quad (10)$$

$$\delta = \delta_{gold} + \delta_{bitcoin} \quad (11)$$

Considering that rational investors always pursue investment portfolios with as large a return as possible and as little risk as possible, this paper uses the method of multi-objective programming to establish a multi-objective portfolio investment model.

$$\max f(t) = \left[1, R_{gold}(t), R_{bitcoin}(t)\right] * H^T$$
$$\min R = \sigma + \delta \tag{12}$$
$$\text{s.t.} \begin{cases} c(t) \geq 0 \\ g(t) \geq 0 \\ b(t) \geq 0 \end{cases}$$

3.2.3. Results and Analysis

Since it is difficult to directly solve the multi-objective programming model, we convert the dual-objective programming into a single-objective programming, and then obtain a new objective function, as shown in Equation (13). The model in this paper is a typical nonlinear programming problem. After research, it is found that some traditional mathematical methods such as Lagrangian method are complicated to solve, and it is difficult to obtain an effective solution. Even if an effective solution is obtained, it is easy to fall into a local optimum; in addition, because in the model of this paper, the calculation of transaction costs can only be determined after obtaining the investment ratio, and the traditional method is also difficult to solve. Therefore, the particle swarm algorithm in this paper solves.

$$\min F\left(R, f(t)\right) = \sigma + \delta - \left[1, L_{gold}, L_{bitcoin}(t)\right] * H^T \tag{13}$$

The particle swarm algorithm relies on the model of flocking birds foraging to find the optimal value. The algorithm has a fast convergence speed, and its leap makes it easier to find the global optimal value without being trapped in the local optimal value[6].

As can be observed from the graph, almost no transactions were made in the preceding moments. In fact, this is not the case. Since the initial principal is relatively small and the number of purchases in the previous period is limited, it can only be changed at a small value, so it is difficult to see the effect. Later, with the growth of cash and the appreciation of bitcoin and gold, the amount that can be bought will increase in the future is shown in Figure 3 and 4.

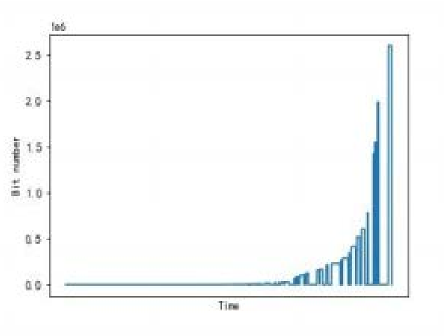


Figure 3: Changes in bitcoin holdings

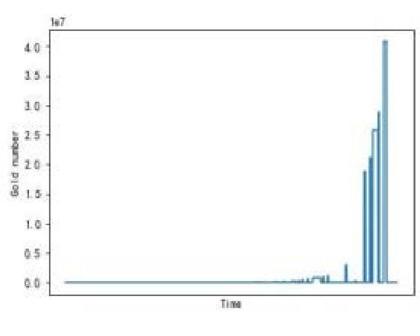


Figure 4: Changes in gold holdings

4. Analysis of the Optimality of Combination Trading Strategies

In this section, we will explore the optimality of portfolio trading strategies. The purpose of buying and selling gold and bitcoin is to reduce risks while accumulating wealth. Therefore, the quality of a trading strategy can be evaluated in terms of cumulative returns and anti-risk capabilities.

4.1. Total Value

We use the total value of cash, gold and bitcoin currently owned as an indicator to evaluate the investment efficiency of the portfolio investment strategy. The larger the total value, the better the total value of the investment. The total value is recorded:

$$V = c_t + g_t * P_{gold}(t) + b_t * P_{bitcoin}(t)$$

With $1,000 as initial funding, we simulated five years of portfolio trades and calculated the total value after each trade. The change in total value over time during these five years is shown in Figure 5. Influenced by the value of bitcoin and gold, the investment strategy given by us results in an almost exponential increase in the total value. As of September 10, 2021, the total value has increased from the initial capital to $130948420399.063, an increase of 100000000+ times, which further confirms the superiority of our strategy to obtain returns.

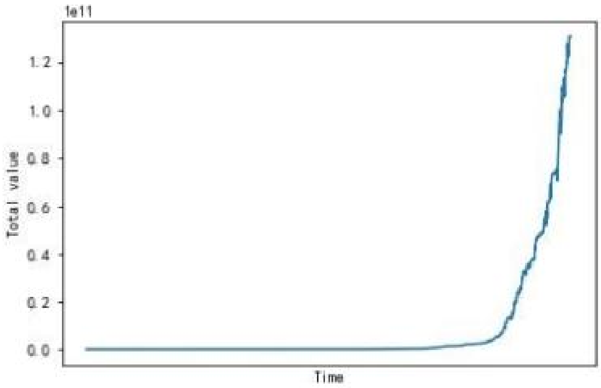


Figure 5: total value graph

4.2. Sharpe Ratio

The Sharpe Ratio (HP) is one of the ways used to evaluate the return to risk ratio of a portfolio [7]. Compared to cumulative returns, the Sharpe ratio takes into account both returns and risks, and provides a more holistic view of portfolio strategies. Generally speaking, the larger the Sharpe ratio, the larger the return-risk ratio, that is, the return is relatively more, the risk is relatively lower, and the portfolio strategy is considered to be more excellent.

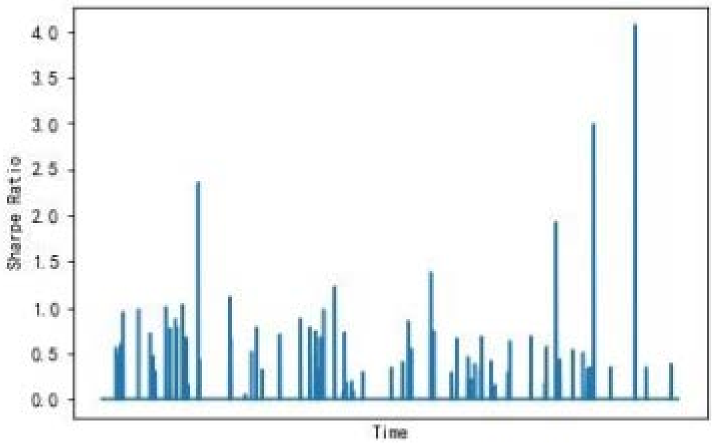


Figure 6: Sharpe ratio change graph

Based on the daily portfolio investment strategy calculated by the model, we draw the Sharpe ratio curve of the portfolio investment strategy, as shown in Figure 6. Looking at Figure 6, it can be concluded that the Sharpe ratio values are all greater than or equal to 0. There are two main reasons for this: on the one hand, because the portfolio trading strategy supports the current return every time, there will be no investment strategies with a Sharpe ratio lower than 0. On the other hand, with the improvement of the prediction model effect and the price increase of gold and bitcoin, the Sharpe ratio of the portfolio investment strategy calculated by the model has a significant increase. Our model is able to make optimal decisions by measuring risk and value based on the state of historical market open trading dates. So, in effect, every decision the model makes is the maximum Sharpe ratio that can reach the current state.

4.3. Maximum Withdrawal

The maximum withdrawal represents the maximum amount of lost wealth when the cumulative return reaches the lowest point at any time in the investment cycle. The maximum withdrawal indicates the situation where the investor's loss is the greatest, and it is an important basis for measuring the risk situation of the investment portfolio. If a portfolio strategy's maximum withdrawal is too high, the strategy is considered potentially risky and needs to be used with caution.

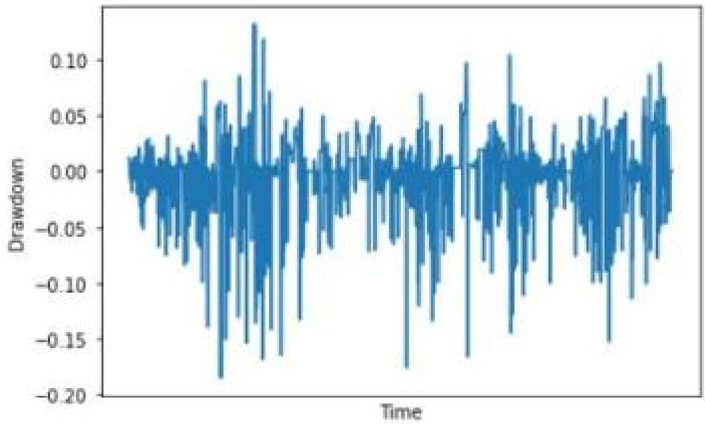


Figure 7: Maximum withdrawal amount change curve

We have plotted the withdrawal rate as a function of daily investment, as shown in Figure 7. The calculated maximum withdrawal rate is 13.19%, while the current maximum withdrawal rate for general assets is 20%[8]. Therefore, our investment strategy is relatively stable, and can even have negative withdrawal rate at some time, which means that the total value of the investment in the future will not be lower than the current total value.

Therefore, it can be concluded that our combination trading strategy can make the optimal choice for the unstable market, resist the existing risks in the market, and is an excellent trading strategy.

5. Sensitivity Analysis

Now we are to analysis of the sensitivity of the trading strategies to transaction costs.

As we all know, every time a market trader buys or sells a liquid asset, there is a certain transaction cost. Currently, there are different transaction rates in the futures trading market of various countries, generally 0.3%~3%, and the lowest transaction rate is 0.03%. Therefore, in the real world, the impact of transaction fees on the yield is very large. In addition, frequent trading will expand this influence infinitely, resulting in a decrease in yield. Classic Markowitz portfolio management models tend to ignore transaction costs, but ignoring transaction costs often leads to ineffective asset portfolios.

To examine the practical effects of model application, we compare the cumulative returns of portfolios under different transaction cost scenarios.

Case 1: bitcoin's Transaction Cost is Variable and gold's Transaction Cost is Constant.

For B:2%-G:1%, bitcoin is denoted by B, and gold is denoted by G. B:2%-G:1% means that the transaction cost of bitcoin 2%, and the transaction cost of gold is 1%.

The result can be obtained by calculation: Under the constant transaction cost of gold, when the transaction cost of bitcoin changes as: 1% →1.5% →2%, the maximum value of the cumulative income will definitely change: $130948420399.036→$75228231685.5037→ $49841230192.4623.

Obviously, when the transaction cost changes of bitcoin, it can seriously affect yield from purchasing and selling bitcoin. This is because:

The investment strategy calculated by the model, most of the assets were used to buy bitcoin after 2020-October.

Each transaction will generate commissions, and frequent purchasing and selling of bitcoin will cause a surge in transaction costs, which in turn will lead to a decline in cumulative income.

It can be concluded that when the transaction cost increases, the cumulative income will also change to a certain extent, and it should become smaller under normal circumstances. The specific change in income is determined by the purchase decision, which is made after comparing future income and transaction costs. Therefore, the model has a certain lag. In addition, the risk judgment mechanism within the model is also the fundamental reason why transaction costs affect the cumulative returns of bitcoin. When the transaction cost changes within a certain range, the cumulative income will become larger, but when the transaction cost deviates from this range, the income may become larger. This is because when the transaction cost is greater, the original selling strategies may be changed to holding, which may reduce losses and increase profits. It is worth noting that the dollar price of one bitcoin has changed a lot compared to gold. The dollar price of one bitcoin has grown from $600+ to $60,000+ in five years, and price volatility has increased by 10,000%, with a 5-year average of $12,206+, which also has a big impact on returns.

As shown in Figure 8, when bitcoin's transaction costs change, the gap in vitcoin's yield becomes more and more obvious. Therefore, on the basis of the above discussion, combined with Figure 8, it can be concluded that the model has poor stability and high sensitivity to the transaction cost of bitcoin's.

Figure 8: For B:2%,2.5%,3%-G:1%, the cumulative income change curve graph

Case 2: bitcoin's Transaction Cost is Constant and gold's Transaction Cost is Variable

It can be seen from the calculation results: when the transaction cost of bitcoin is constant, when the transaction cost of gold changes as: 2% →2.5% →3%, the maximum value of cumulative income also changes: $121746247075.351→$104303523283.813→$109494597165.664.

As shown in Figure X, when the transaction cost of bitcoin is constant and the transaction cost of gold varies, the cumulative returns vary by a small margin. This is because:

After 2020, most of the funds are used to buy bitcoins, and more bitcoins are held.

Although the transaction cost of gold has increased, the transaction cost is not very high due to the small amount of gold held, so the change in the cumulative income is not very large.

It can be concluded that, different from the change trend of bitcoin's income, when the transaction cost of gold increases, the cumulative income shows a downward trend, because the price change of gold is different from that of bitcoin. The U.S. dollar closing price for an ounce of gold is within 5 years, relatively stable compared to bitcoin. The dollar closing price of an ounce of gold has grown from 2$1300 to $1800+ in five years, and the average dollar closing price of an ounce of gold in five years is $1400, with a price volatility of 38%. Although the US dollar closing price of an ounce of gold has been fluctuating, the model and risk determination mechanism we have established are less sensitive to the transaction cost of gold, which will also lead to frequent trading of gold and eventually loss. The change in gold's return is shown in Figure 9.

As shown in Figure 9, when the transaction cost of gold changes, the range of yield changes is smaller. Therefore, combined with the above discussion, we can conclude that the model is more stable and less sensitive to the transaction cost of gold.

Figure 9: For B:2%-G:1%,2.5%,2%, the cumulative income change curve graph

6. Conclusions

In this paper, we build a comprehensive model to determine day-to-day investment strategies. The model uses the price data so far to train the LSTM-ARIMA hybrid model and predicts the expectations of future prices; the model solves the next moment's holdings by solving a dual-objective programming that maximizes the total amount of future value and minimizes risk. Quantity and then determine the investment strategy. We show that our investment strategy is the optimal strategy through economic indicators such as the Sharpe ratio. For gold and bitcoin cost changes, we find that bitcoin investment costs are more sensitive to changes in total value, while gold investment costs are less sensitive to change.

References

[1] Ponomare V E, Oseledets I, Cichocki A. Using Reinforcement Learning in the Algorithmic Trading Problem. Journal of Economic Dynamics and Control.2020.
[2] Gang Chen, Xiaomei Guo. Research on Oversampling Algorithm for Imbalanced Datasets Based On ARIMA Model. Proceedings of the 33rd China control and decision making conference. 2021:441-446.
[3] Troiano L, Mejuto Villa E, Loia V. Replicating a Trading Strategy by Means of LSTM for Financial Industry Applications. IEEE Transactions on Industrial Informatics, 2018:3226-3234.
[4] Kamble V.B., Deshmukh S.N.. Comparision between Accuracy and MSE, RMSE by Using Proposed Method with Imputation Technique. Oriental journal of computer science and technology, 2017, 10(04).
[5] Markowitz H. Portfolio selection. Journal of Finance.1952(7):77-91.
[6] Cai Biao, Zhu Xinping, Qin Yangxin. Parameters optimization of hybrid strategy recommendation based on particle swarm algorithm. Expert Systems with Applications, 2021,168:
[7] Qianhan Zhang. The Use of Sharpe Ratio in Portfolio Optimization.Proceedings of 4th International Conference on Economics and Management, Education, Humanities and Social Sciences (EMEHSS 2020). 2020:224-229.
[8] Javier Estrada. Maximum Withdrawal Rates: A Novel and Useful Tool. Journal of Applied Corporate Finance, 2017, 29(4).