Quantitative trading forecasting and decision modeling and analysis

Kaixuan Xu[1], Xingji Chen[2, *]

[1]School of Automation Science and Electrical Engineering, Beihang University, Beijing, 100191, China

[2]School of Electronics and Information Engineering, Harbin Institute of Technology, Shenzhen, 518055, China.

*Corresponding author: chenxingji0312@163.com

Keywords: Grey forecast, LSTM EVaR, sensitivity binary linear regression models

Abstract: In this paper, in order to obtain the maximum benefit with the minimum risk, we construct a quantitative trade forecasting and quantitative trade decision-making model by adopting the grey forecasting model, the LSTM neural network model, the information entropy risk measurement model, and the risk-optimized threshold return model. In addition, the robustness and sensitivity of the considered models were investigated. The model we built develops the best daily trading strategy by predicting the price of gold and bitcoin. It can help market traders make better decisions every day and maximize returns within a manageable risk range.

1. Introduction

In today's world, more and more traders are joining in trading various assets. Gold has long been popular with traders as an asset with stable returns and low risk, and in recent years, with the trading of virtual currencies such as Bitcoin coming into millions of households, Bitcoin is highly sought after by the risk-averse because of its high return nature. Traders tend to hand over their money to professionals such as market traders to buy and sell, who frequently buy and sell by analyzing the asset's historical ups and downs to help maximize returns for all types of traders. The analysis of Bitcoin's data shows that Bitcoin's late-stage data fluctuates significantly, with large increases over five years that are not easily predictable. However, Bitcoin can be traded daily with complete data and predictions.

By analyzing the data for gold, we see that gold shows a volatile rise over five years, but the fluctuations are small and easy to predict. The prices of individual days are missing, and we use the average of two adjacent prices as a supplemental value. At the same time, because gold can be traded only when the market is open, there are only twelve hundred days of data in the five years of more than eighteen hundred days. At this time, if we forcefully interpolate the data, on the one hand, interpolating about two-thirds of the data will seriously affect the prediction accuracy of the model. On the other hand, we were predicting the price when the market is not open is not very useful. Therefore, we look at the more than twelve hundred data as continuous training time provided to the network.

In this paper, we develop a model for quantitative trade forecasting and quantitative trade decision-making, which combines the price data of the previous day and the commissions that each trade will generate to help market traders develop the best trading strategy for the day. Furthermore, we demonstrate with experimental data that our model can provide as much reward as possible with as little risk. We will determine the sensitivity of our strategy to transaction costs by adjusting the percentage of transaction value generated by the commission generated by the transaction.

2. Model Building and Solving

2.1 Grey Forecast Model

According to existing research, regression and neural models are not very good at predicting small data samples, while grey forecasting models [1] can give an ambiguous long-term description of the development pattern of things through a small amount of incomplete information.

Because the grey forecasting model is suitable for small samples [2], we need to divide the samples between gold and bitcoin. After repeated experiments, we arrived at a better choice. For bitcoin, we take the first 20 days to predict the next day's data, and when the actual data of the latter-day is available, we then use the nearest 20 days' data to predict the next day, and so on to obtain the price of bitcoin on the 21st day and beyond; for gold, we take every thirty data as a sample, and similarly obtain the price of gold on the 31st sample and beyond.

The C and P values in table 1 are provided to determine the number of days to forecast gold. The C and P values in table 2 are provided to judge the predicted bitcoin. For the sake of subsequent decision-making, we will uniformly start forecasting from day 37 while not forecasting for the first 36 days. The average accuracy of our forecasts for gold and bitcoin over the time period is aresummarzied in Table 3.

Table 1. The posterior difference ratio and small probability error of gold

number	posterior difference ratio C	small probability error P
23	0.35	0.65
24	0.26	0.75
79	0.40	0.70
80	0.40	0.50

Table 2. The posterior difference ratio and small probability error of bitcoins

number	posterior difference ratio C	small probability error P
36	0.33	0.65
37	0.24	0.83
89	0.38	0.80
90	0.58	0.65

Table 3. The prediction accuracy of grey forecast

	Gold	Bitcoin
Accuracy	0.96611	0.94669

2.2 LSTM Neural Network Model

LSTM is a special kind of RNN neural network, and the general structure of LSTM is as shown in Figure 1.

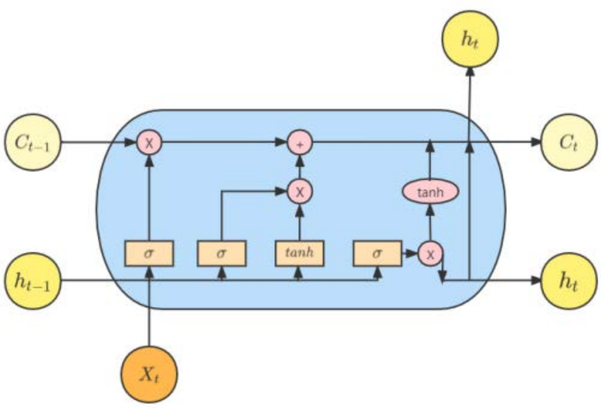

Figure 1. The structure of LSTM

However, a linear function normalization (Min-Max Scaling) operation is performed first to linearly transform the original data so that the result is mapped to the range [0, 1] to achieve isometric scaling of the original data. The normalization equation is as follows.

$$X_{\text{nom}} = \frac{X - X_{\min}}{X_{\max} - N_{\min}} \tag{1}$$

where X is the original data $X_{\max}$, $X_{\min}$ are the maximum and minimum values in the data, respectively.

The model's loss function takes MSE (Mean Square Error), and the training optimizer selects Adam. We choose 4, 12, 16, 32 as the feedforward network layers, and then the first 90 days as a training sample, followed by ten days as a validation set to compare their prediction accuracy. The results are as shown in Table 4.

Table 4. The prediction accuracy of LSTM with different number of layers

Number	Accuracy of bitcoin	Accuracy of gold
4	0.75667	0.97674
16	0.83215	0.98454
32	0.86129	0.98274
64	0.83436	0.97963

It can be seen that the predictions for both bitcoin and gold are relatively better when the number of layers is 32, so we choose 32 as the final network layer. It can also be seen that bitcoin is more unpredictable than gold due to its high volatility, and the risk and reward of purchasing it are relatively higher [3].

The prediction curve for Bitcoin is plotted against the true curve, as shown in Figure 2. Moreover, the forecast curve for gold is plotted against the true curve, as shown in Figure 3. It can see from the curves that the predictions for gold are very accurate, and for bitcoin, the predictions are mostly accurate, but at the peak around 1500 days, the model has significant errors. However, the overall trend matches the actual values and does not affect the analysis of the decision too much.

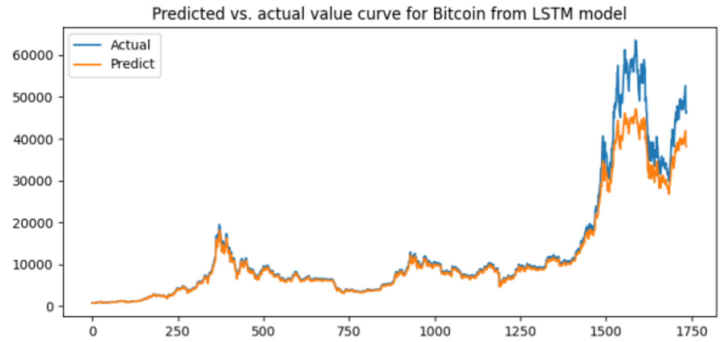

Figure 2. The predicted and actual price of bitcoin

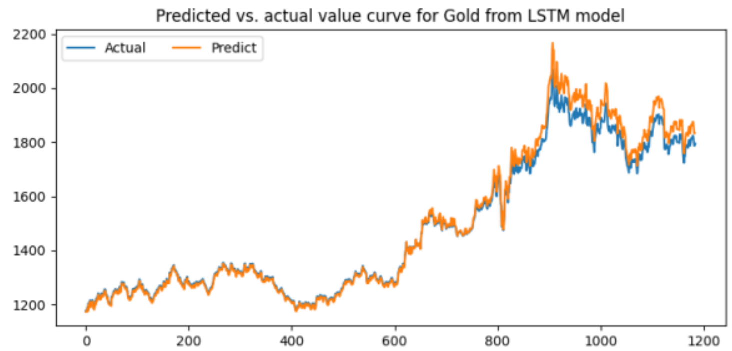

Figure 3. The predicted and actual price of gold

2.3 Risk Optimization Based Threshold Return Model

By considering the size and characteristics of the dataset, ours uses an updated, information entropy-based risk metric [4, 5]. The risk of the assets is evaluated using available price data for both assets. Set the return interval to [-10%,10%] and divide it equally into 20 small intervals, each with a step size of 1%, and calculate the frequency g f of gold and bitcoin falling in the kth subinterval, respectively, replacing the probability with the frequency [4, 6].

The entropy value-at-risk of the asset is calculated as:

$$\mathrm{H(S)} = -\sum_{k=1}^{q} p_k \ln p_k \tag{2}$$

Assuming that the weight ratio of the two assets is w_gold to $w_bitcoin$, the following relationship holds since EVaR is subadditive, which can be expressed as:

$$H\left(w_g old * G + w_b itcoin * B\right) = w_g old * H(G) + w_b itcoin * G(B) \tag{3}$$

where $\mathrm{w_gold + w_bitcoin = 1}$, $\mathrm{w_gold * H(G) = w_bitcoin * H(B)}$.

In addition, we also have to consider the cost of the trader's time and other losses that come with each transaction, so when trading, he wants to get an inevitable minimum return, which our team sets as a threshold gamma, and only if the rate of return that removes the cost is higher than γ, we will buy more assets. A schematic diagram of the decision using the rate of return is shown in Table 5.

Table 5. Operation selection interval

Interval	Trading behavior
$(-\infty,\alpha]$	Sell asset
(α, threshold]	Maintain
(threshold, ∞)	Purchase asset

where γ depends on the investor's risk appetite and also on the specific situation and has an important impact on investment decisions.

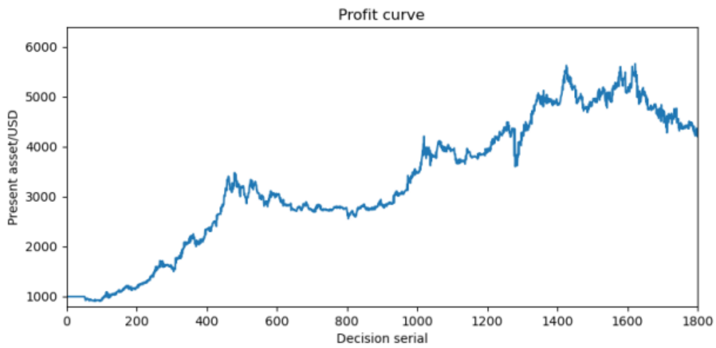

Figure 4. Profit curve

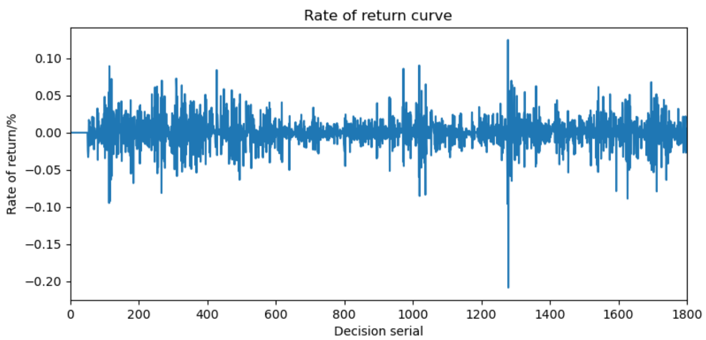

Figure 5. Profit rate curve

The model calculation gives us the following return curve for the investment strategy, as shown in Figure 4. The yield change curves are shown in Figure 5. Although losses occur, our strategy has been gaining, and the level of return has been relatively stable. During this process, the movement of positions in USD, Gold, and Bitcoin is shown in the following figure 6. Since the investor in our model is a risk-averse person, he buys less of the riskier bitcoin, buys small amounts when it is stable, and sells when it is volatile. The investor maintains a high and stable position for the less risky gold.

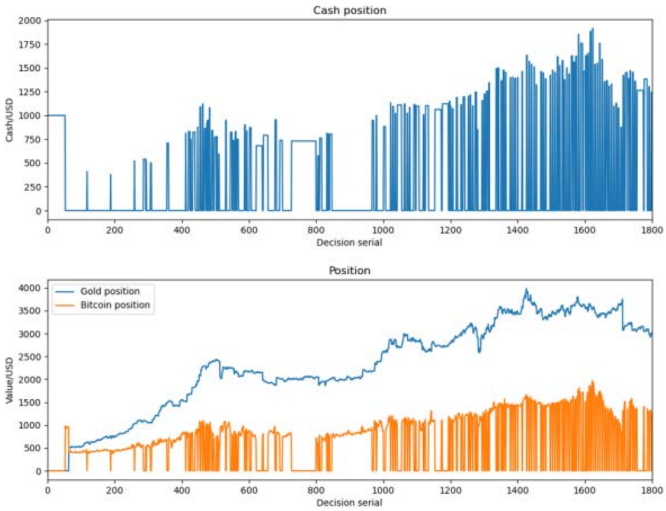

Figure 6 Asset position

3. Model Improvements

We first establish a cross-sectional comparison model by changing the threshold value and comparing the decision results to prove the superiority of the decision model. Then, the loss function MSE, Sharpe index, and other indicators demonstrate the accuracy and optimality of the model prediction results.

3.1 Modeling of cross-sectional comparisons

For the decision model, we can adjust the threshold γ to build a cross-sectional comparison model. By observing the data, we find that the change in assets mostly stays within the ±10% interval throughout the time period required by the question, so we choose multiple thresholds to obtain the results, as shown in Figure 7. The final result shows that γ= 3% is the optimal parameter taking the value of the model. Of course, in practice, we will not get all the price data, and we will not be able to obtain the value of γ by experimental methods, so we need to have a more scientific parameter determination scheme, and this method will be mentioned in the later section.

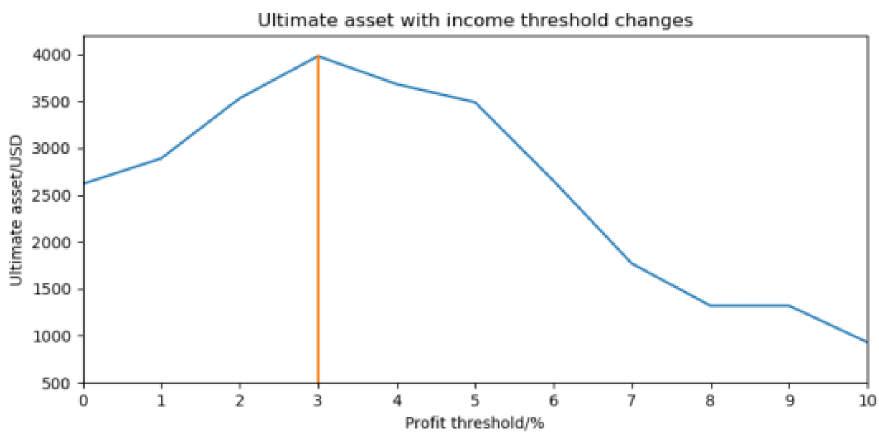

Figure 7 Ultimate asset with γ changed

3.2 Inspection and analysis of model accuracy

We have already mentioned the hierarchical rating scale for the posterior residual ratio C and the error probability P in constructing the quantitative trading prediction model.

In addition to ensuring the accuracy of the model at the time of prediction, it is also important to compare the actual results with the predicted values after they are available. The relative error formula is as follows:

$$\delta = \Delta/L \times 100\% \tag{4}$$

where Δ is the actual absolute error and L is the true value.

The kth data accuracy is defined as follows:

$$q^{(K)} = 1 - \delta^{(K)} \tag{5}$$

The average accuracy rate is

$$\mathrm{Q} = \frac{1}{n}\sum_{i=1}^{n} q^{(i)} \tag{6}$$

The accuracy of gold and bitcoin with grey prediction equals 96.611% and 94.669%.

The mean squared error (MSE) is the expected value of the squared difference between the parameter estimate and the parameter value, which can be expressed as:

$$\mathrm{MSE} = \frac{SSE}{n} \tag{7}$$

where n is the number of samples, and SSE [7] represents the sum squared variance, i.e., the sum of squares of the errors of the fitted data and the original corresponding points, calculated as:

$$\mathrm{SSE} = \sum i = 1^{m} wi(yi - \widehat{yi})^2 \tag{8}$$

where yi is the real data, $\widehat{yi}$ is the fitted data, and $wi > 0$. The loss function MSE for gold and bitcoin equals 0.00015347 and 0.000000079228.

It can be seen that the values of the MSE after training are small, indicating good accuracy of the LSTM neural network model prediction at the time of prediction.

Meanwhile, the same average accuracy of the gray prediction model Q_1 as above can be used to represent the accuracy of the LSTM model by Q_2. The accuracy of gold and bitcoin with LSTM equals 98.042% and 88.690%.

4. Sensitivity Analysis

4.1 Transaction cost sensitivity calculation results

The final returns for different cost combinations are obtained by adjusting the transaction costs $\alpha_{gold}$and $\alpha_{bitcoin}$ of the two assets separately. Accordingly, the sensitivity of the model to changes in transaction costs can be analyzed. We can see the regular from Figure 8. As transaction costs rise, investment returns show a declining trend.

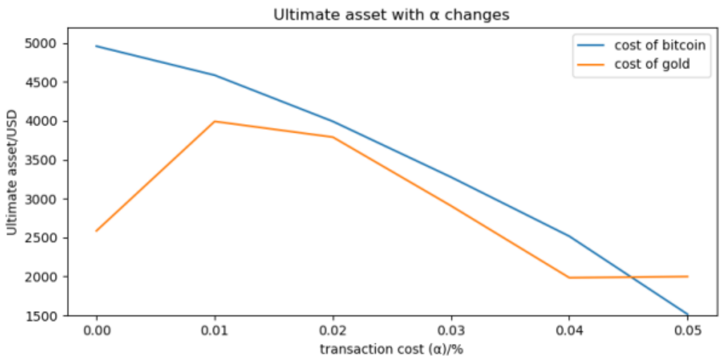

Figure 8 Ultimate asset change with transaction cost changed

4.2 Transaction Cost Sensitivity Theory Explanation

We summarize the impact of cost changes on the final benefit into two main factors: the combined effect and the cost effect. The portfolio effect refers to a change in the ratio of the two assets due to the different yields. Therefore, a significant change in the overall return occurs. The cost effect refers to the decrease in returns due to higher transaction costs. In this problem, bitcoin has a significantly higher and much higher rate of return than gold. When the cost of gold decreases, the return goes up due to the cost effect and the portfolio effect, which causes a lower weighting of bitcoin in the portfolio and ultimately a lower benefit. Thus the anomaly at α_{gold}= 0 is caused by the portfolio effect being greater than the cost effect. The other changes are also caused by the two effects going one way and the other way.

4.3 The effect of transaction costs on the return threshold γ

As the transaction cost of both assets increases, the value of γ keeps increasing and shows a more obvious regularity, of which the law can be explored by establishing a binary linear regression model [8], which can be expressed as:

$$\gamma = \mathrm{b} * \alpha_{bitcoin} + g * \alpha_{gold} + a \tag{9}$$

The model fitting effect can be shown in table 6. The coefficients of all parameters in the model pass the test

Table 6. Model fitting effect

	Coefficients	Std-error	t-Stat	P-value
Intercept	1.34	0.2343	5.7185	0.0000
b	0.42	0.0524	8.0157	0.0000
g	0.52	0.0.524	9.9242	0.0000

and this model R^2= 0.88; *Adjusted* $-R^2$= 0.87 works well. It can be concluded that:

$$\gamma = 0.42 * \alpha_{bitcoin} + 0.52 * \alpha_{gold} + 1.34 \tag{9}$$

5. Conclusion

In this paper, we predict the price of gold and bitcoin by building a gray prediction model, an LSTM neural network model, and a risk-optimized threshold return model for developing the best daily trading strategies for each day. With the application of our quantitative trading forecasting model and quantitative trading decision model. The model has great application in the financial trading process: in real-world situations, we have access to more information than just the price, which allows us to predict more accurate price trends and thus make better predictions. Gold and bitcoin are similar to financial assets such as stocks and futures, and our model can be applied to similar and more diverse areas.

References

[1] Q. Tianhua, Stock price forecasting based on gray model and ARIMA mode, Computer Age, 2021(10):83-85+89.

[2] L. Lin, Spss data regression analysis and gray prediction, Mathematical Learning and Research, 2021(29):136-137.

[3] M. Carammia, S.M. Iacus, T. Wilkin, Forecasting asylum-related migration flows with machine learning and data at scale, Sci Rep 12, 1457 (2022).

[4] H. P. Zhang, J. M. Zhu, Optimal design of multi-stock portfolio investment strategy based on risk diversification, Journal of Henan University of Science and Technology (Natural Science Edition), 2020,48(04):62-70.

[5] X. Zhang, Research on information entropy method in risk metrics, Beijing Jiaotong University,2015.

[6] J.L. Zhu, Y. Z. Fu, Y. I. Chen, Portfolio model and its application, Computer Technology and Development, 2013, 23(12):168-170+174.

[7] Y. Peng. Research on SSE 50 ETF price prediction and option strategy based on GA-LSTM, Northwestern University, 2021.

[8] S. L. Cai, Empirical analysis of multiple linear regression model application, Journal of Lang fang Normal College (Natural Science Edition), 2017, 17(04): 5-8.