Modeling and Analysis of Metrics for Quantitative Investment

Yi Su *, Yumeng Zhang

School of Finance, Anhui University of Finance and Economics, Bengbu, Anhui, 233000

*Corresponding Author Email: suyi20220322@163.com

Abstract. In recent years, the issue of quantitative investment has been hot in the investment field amidst the rapid development of quantitative finance. It is necessary for different investment objects to build corresponding portfolios and set investment strategies based on historical data pairs to get maximum returns. To help investors better solve the investment problem of gold and bitcoin portfolios. This paper constructs a BP neural network price prediction model, predicts the closing price on the last day of the transaction, and then uses the greedy algorithm to solve the optimal daily trading strategy to pursue the overall optimal solution and obtain the maximum profit. Combined with grey forecasting and time series forecasting methods, the forecasting models for gold and bitcoin prices are constructed again.

Keywords: BP Neural Network, Greedy Algorithm, Entropy Method.

1. Introduction

From a global perspective, quantitative investing has been around for more than 50 years. Since the 1960s, with the continuous development of quantitative finance, the explosion of computer technology programming technology and the continuous reduction of transaction fees, quantitative finance has developed iteratively. The core of quantitative investment is mainly based on quantitative finance. Therefore, in the context of the ever-changing quantitative finance, quantitative investment has set off a wave in the investment market.

Quantitative investment refers to the realization of programmed transactions through computer programming to obtain stable returns. This investment method combines statistics, mathematics and computer technology, and uses the powerful data mining and data processing functions of computers to find investment opportunities that can generate returns from a large amount of historical data. Therefore, for different portfolios of investment objects, traders need to develop specific mathematical models to determine whether to buy, hold or sell portfolio assets based on the historical prices of investment objects, so as to obtain maximum returns.

2. Price Forecasting and Strategy Development

2.1. Data Preprocessing

According to the data attached in the title, you can know the closing price of gold in troy ounces in US dollars on the specified date. However, since the gold price has ten missing values within the tradable days, the gold price data needs to be screened. Using IBM SPSS 25.0, we processed missing values for gold prices by interpolation, retaining the remaining available values.

2.2. BP Neural Network Price Prediction Model

2.2.1 Data Selection

The pre-processed closing prices of troy ounces of gold in US dollars and the US dollar prices of individual bitcoins on the specified dates were selected as the sample data. The time interval of the data was from 2016-9-10 to 2021-9-10, for a total of 1,826 trading days.

Using Matlab to build BP neural network model, the price variables of the first five trading days of gold and bitcoin are used as the input layer, and the closing price of the day is used as the output layer respectively. The sample data are divided into two parts, 1-1160d data are used as training data,

and 1161-1260d data are used as experimental data for gold. 1-1700d data are used as training data, and 1701-1822d data are used as experimental data for bitcoin.

2.2.2 Setting of Relevant Parameters

BP neural network as an algorithm for intelligent data processing, the number of neurons (i.e., neural nodes) in the hidden layer can be set according to the actual situation. The transfer function can be tuned to meet the error requirements.

Since there exists the concept of five-day averages, i.e., the weighted average of five-day closing prices, in stock technical investment, which is indicative for investment. Therefore, in this paper,, the number of nodes in the neural network's input layer is 5, the number of nodes in the output layer is 1, and the number of neural nodes in the hidden layer l is experimented with by trial-and-error method.

Among them: m is the number of nodes in the input layer; n is the number of nodes in the output layer; a is an adjustment constant between 1 and 10. In this paper, the value of a is selected as 1, 2, 3, 4, and 5. After trial and error experiments, the number l of hidden layer nodes is initially determined to be 2, 3, 4, 5, and 6, respectively.

2.2.3 BP Neural Network Fitting and Prediction

Using Matlab software, the goodness-of-fit was obtained by setting different numbers of nodes and comparing the average absolute error. The table shows the average absolute error and the goodness of fit for different numbers of nodes in the hidden layer. The smaller the average relative error, the better the equation fits, and the better the goodness-of-fit is close to 1, the better the equation fits. According to the table, the number of nodes in the implicit layer is 4, with the highest goodness of fit and the best equation fitting.

Table 1. Gold: the average absolute error corresponding to different numbers of nodes in the hidden layer and the goodness of fit

Projects	Number of nodes in the hidden layer				
	2	3	4	5	6
Average relative error	0.0077395	0.0082372	0.0075673	0.0077161	0.0078147
Goodness of fit	0.80257	0.79699	0.85028	0.83941	0.81166

Table 2. Bitcoin: average absolute error corresponding to different number of nodes in the hidden layer and goodness-of-fit

Projects	Number of nodes in the hidden layer				
	2	3	4	5	6
Average relative error	0.038086	0.037545	0.03636	0.037817	0.041768
Goodness of fit	0.9196	0.92022	0.92025	0.91881	0.91944

The training and fitting effects of the BP neural network for gold when the number of nodes in the hidden layer is four are shown in Fig. As can be seen from the figure, when the number of neurons in the hidden layer is 4, the simulation training reaches 1000 times and meets the error requirement, and the final fitting goodness is 0.99232.

Similarly, the training and fitting results of the BP neural network for Bitcoin when the number of nodes in the hidden layer is 4. As can be seen from the figure, when the number of neurons in the hidden layer is 4, the simulation training reaches 1000 times and satisfies the error requirement, and the final goodness of fit is 0.99853.

Based on the investor's perspective, the closing prices of gold and bitcoin are unknown at the time of investing on September 10, 2021, so forecasts of gold and bitcoin prices for September 10 are needed. And the gold and bitcoin prices on the day of September 10 need to be estimated based on the predicted prices of gold and bitcoin on the previous day, September 9. By estimating the above BP neural network gold and bitcoin price prediction model, we can get the closing prices of gold and bitcoin on September 10 as 1785 and 45314, respectively.

3. Evidence of the Best Strategy

3.1. Grey Forecasting and Time Series Forecasting

3.1.1 Price Forecasting Model Based on Grey Forecasting

Gray prediction model is a forecasting method that builds mathematical models to make predictions from a small amount of incomplete information. The GM(1, 1) model is one of the gray prediction models, which is a differential equation for discrete series:

$$\frac{d_x}{d_t} + a_x = u \quad GM(1, 1) \tag{1}$$

The discrete form and prediction equations are as follows:

$$\hat{X}^{(1)}(k) = (X^{(0)}(1 - \frac{u}{a})e^{-a(k-1)} + \frac{u}{a} \tag{2}$$

Only models that pass the test can be used for forecasting. The smaller the C value, the less discrete the difference between the calculated and actual values. The relative error size test is mainly based on the average relative error alpha value. Alpha value less than 0.05 is the first level, and less than 0.10 is the second level. Predictions can be made when the grade reaches Level 1 or Level 2. Run the program in R. Input gold from 2021-7-13 to 2021-9-9 into the model and output the predicted closing price of gold for the day 2021-9-10; input bitcoin from 2021-7-10 to 2021-9-9 into the model and output the predicted closing price of bitcoin for the day 2021-9-10.

3.1.2 Price Forecasting Model Based on Time Series Forecasting

A time series forecasting model is a series of observations arranged in time order. It does not consider the causal relationship between variables, but focuses on the development of variables in time and builds a mathematical model. The prerequisite for using a time series model is a sufficiently long data series. The changes in the data series are stable and regular, and the price movements of gold and bitcoin meet the requirements of the data.

A correlation chart is made based on the price dynamics of gold and bitcoin, and a correlation analysis is performed to find the autocorrelation function. The correlation graph shows the trend and period of change and identifies jump and inflection points. Then a suitable stochastic model is identified, and curve fitting is performed, i.e., a generalized stochastic model is used to fit the time series of observations.

Using SPSS for the operation, the daily closing prices of gold and bitcoin for the period 2016-9-10-2021-9-9 were used simultaneously to predict the closing prices of gold and bitcoin for the day 2021-9-10, and the predicted results are as follows:

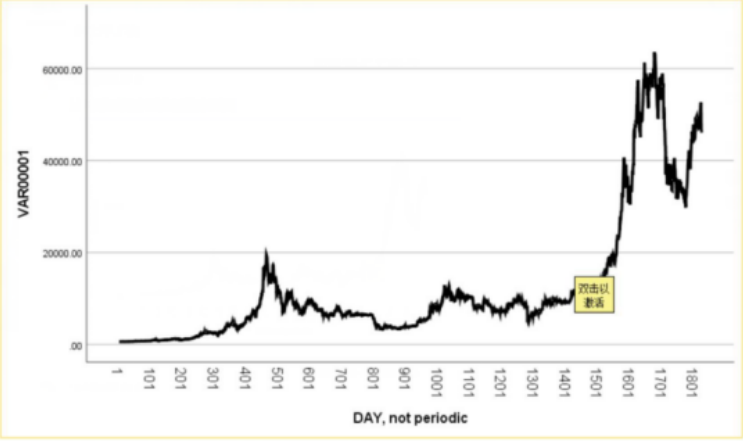


Figure 1. Gold: Time Series Forecasting Model Fitting and Forecasting

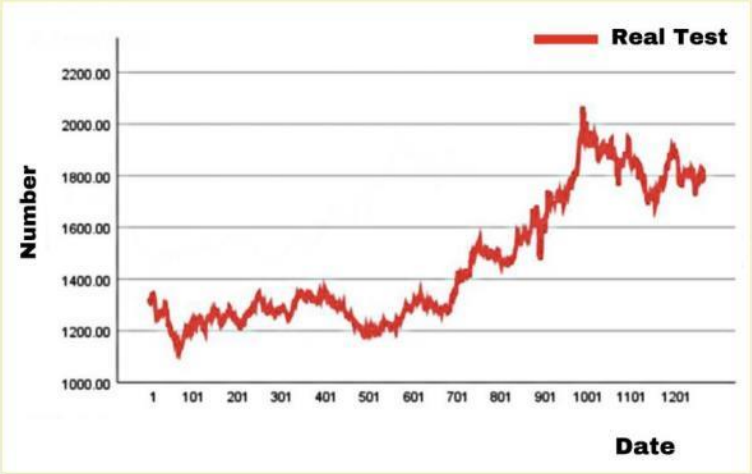


Figure 2. Bitcoin: Time Series Forecasting Model Fitting and Forecasting

3.2. Comprehensive Evaluation Model Based on Entropy Method

3.2.1 Selection of Indicators

For the three established price forecasting models, the following four indicators were selected: the coefficient of resolvability, the relative error, the difference between the forecasted and true values of gold and bitcoin on September 10, and the return values of gold and bitcoin on September 10, and the following matrices were established.

$$\begin{bmatrix} 0.9203 & 0.0364 & 54.6900 & 1.4700 \\ 0.8368 & 0.1580 & 8.4637 & 1.6900 \\ 0.9370 & 0.0266 & 215.920 & 1.4700 \end{bmatrix} \tag{3}$$

3.2.2 Entropy Method Calculation

According to the Entropy method calculation formula:

The weight of the iind sample under the jjth indicator to the indicator.

$$\rho_{ij} = \frac{x_{ij}}{\sum_{i=1}^{n} x_{ij}}, \quad (i = 1,2,3, \ldots, n;\ j = 1,2,3, \ldots, m) \tag{4}$$

Then the entropy value of the jth indicator is:

$$e_j = -k\sum_{i=1}^{n} \rho_{ij} \times \ln(\rho_{ij}), \quad (j = 1,2,3, \ldots, m) \tag{5}$$

Since there is a logarithm in equation (4), the value of k is usually taken as follows in order to determine whether it is meaningful.

$$k = \frac{1}{\ln(n)}, \quad (0 \le e_j \le 1) \tag{6}$$

The coefficient of variation for the jth indicator is

$$d_j = 1 - e_j \tag{7}$$

Then the weight of the jth indicator

$$w_j = \frac{d_j}{\sum_{j=1}^{m} d_j}, \quad (j = 1,2,3, \ldots, m) \tag{8}$$

Therefore, based on the above equations, Matlab's solution yields gold W1 and bitcoin W2.

W_1=[0.3069,0.1237,0.2434,0.3259]

W_2=[0.1375,0.2844,0.2094,0.3687]

Table 3. Comprehensive evaluation score of the model calculated by entropy method

	BP neural network model	Gray prediction model	Time series forecasting model
Gold	3.0776	2.4175	2.4085
Bitcoin	113.9016	17.632	45.8897

In summary, from the scores in the table, it can be seen that the overall evaluation scores of the BP neural network price prediction model are higher than the other two models for both gold and bitcoin, and the investment strategy calculated using the BP neural network price prediction model is the best.

4. Evaluation of the Model

4.1. Strengths

Greedy algorithm model solution process is a multi-part judgment process, and the final result is the optimal solution to the combinatorial optimization problem. The future can continue to study the nature of the PSP mathematical model and the implementation of cluster computing, and provide computing services to the majority of investors. Neural network model is suitable for solving problems with complex internal mechanisms, with strong nonlinear mapping capability.

Gray prediction model matrix-based in the prediction process, combined with MATLAB, can effectively solve the calculation problem, and the program operation is practical and straightforward, conducive to operation, and the prediction accuracy is high.

4.2. Weaknesses

Greedy algorithm model can only solve feasible solution problems that satisfy certain constraints. The model can not guarantee that the final solution is optimal, nor can it be used to solve the problem of maximum or minimum solution. Because the calculation needs huge memory space, it can be optimized further.

Neural network model whose nature is gradient descent method, convergence speed is slow, the choice of BP neural network structure is not yet clear, systematic guidance, too subjective.

The Gray prediction model has a strong dependence on historical data, fewer inherent parameters, less error tolerance, and limited timeliness, not suitable for long-term forecasting or analysis.

When forecasting, we should pay attention to the future development pattern and development level of market phenomena, which may be inconsistent with historical models and have the defects of forecasting error. Meanwhile, when the outside world changes greatly, there may be a big deviations.

References

[1] Sun, B., Li, T. K., Wang, B. L Neural network forecasting model based on stock market sensitivity analysis [J]. Computer Engineering and Applications. 2011 (01): 26-31.

[2] Luo Qixuan, Li Handong. An improved dynamic IS algorithmic trading strategy [J]. Journal of Beijing Normal University (Natural Science Edition). 2017 (03): 288-293.

[3] Agbulut, U (Agbulut, Umit); Gurel, AE (Gurel, Ali Etem); Bicen, Y (Bicen, Yunus). Prediction of daily global solar radiation using different machine learning algorithms: Evaluation and comparison. RENEWABLE & SUSTAINABLE ENERGY REVIEWS.2020-12-09.

[4] Liu Wei. Deep learning-based stock data mining methods and their applications [D]. University of Chinese Academy of Sciences (Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences). 2021 (09).

[5] Wang Bo. Weekly net value prediction of investment funds based on neural network [J]. Journal of Shanghai University of Technology. 2007 (03): 227-230.