

Gold and Bitcoin Price Prediction Model Based on Grey Model

Zelin Qiu

School of Management, Zhengzhou University, Zhengzhou, China, 450001

1031548412@qq.com

Abstract. As a traditional equivalent, gold has long been a symbol of wealth, traded by countless people. However, with the progress of science and technology, virtual currency has gradually entered our vision and has a great impact on people's life. The purpose of this paper is to accurately predict the price of gold and bitcoin on the premise that the initial conversion is only \$1,000, and ultimately maximize the asset. Therefore, the grey model is established to make the prediction. Its advantage is that it improves the long-term dependence problem in RNN; GM generally outperforms temporal recurrent neural networks (RNN) and Hidden Markov models (HMM). By calculation, the price of gold in 2016 was \$1,325 and the price of bitcoin was \$622; Five years from now, the forecast price for gold is \$1,788 and the forecast price for bitcoin is \$46,192, with an average margin of error of just 1.7%. According to the model built in this article, investors can predict the next three days based on today's data, so that they can adjust our portfolios based on the forecast results.

Keywords: RNN, HMM, Prediction Model, Grey Model.

1. Introduction

Market traders buy and sell volatile assets frequently, with a goal to maximize their total return. There is usually a commission for each purchase and sale. Two such assets are gold and bitcoin. Bitcoin uses peer-to-peer technology to operate with no central authority or banks; managing transactions and the issuing of bitcoins is carried out collectively by the network. Bitcoin is open-source; its design is public; nobody owns or controls Bitcoin and everyone can take part. Through many of its unique properties, Bitcoin allows exciting uses that could not be covered by any previous payment system [1]. In addition, for a long time, gold has become a hot topic in the financial field because of its means of payment, means of circulation, value scale, means of storage and world currency. Based on financial knowledge, this paper uses statistical and mathematical tools to calculate data with quantitative models, and combines computer technology to conduct quantitative transactions to obtain excess returns, providing traders with the optimal strategy packaged in this paper to maximize their returns. The five-year gold price trend from September 2017 to September 2021 is shown in Figure 1, and the Bitcoin price trend is shown in Figure 2 (<https://www.comap.com/undergraduate/contests/>).



Figure 1. Gold daily prices, U.S. dollars per troy ounce



Figure 2. Bitcoin daily prices, U.S. dollars per bitcoin

Both for the economic and financial research of domestic and foreign scholars and people in the industry investment, the financial asset price forecasting problem has always been a hot research field. On the gold prices of the financial time series forecasting method, generally can be divided into two kinds, one is the financial time series forecasting model based on statistical methods, the other is a financial time series forecasting model based on machine learning algorithms [2]. Among the categories of statistical methods, differential integrated moving average autoregression model (ARIMA) and generalized autoregressive conditional heteroscedasticity model (GARCH) are the most widely accepted, but the primary premise of these methods is the linear assumption of financial time series variables, while the main task of grey prediction is analysis, modeling, prediction, decision-making and control. According to Professor Deng Julong put forward in the 1980 s of the GREY theory, the GREY prediction MODEL of typical (GREY MODEL) is the GM (1, 1) MODEL and GM (1, N) MODEL [3]. In the actual research, tend to be a factor (study) research to consider several other related factors, so the grey GM (1, N) model is more widely applied. The innovation of this paper is that prices have different dimensions and dimensional units, which will affect the results of data analysis. In order to eliminate the dimensional impact between prices, it is necessary to standardize the data and solve the comparability between the data. Therefore, for the price of gold and bitcoin, this article uses today's price/yesterday's price instead of today's price for subsequent predictions, which can make the data more substantial. All in all, the core content of the grey model is to establish the prediction model, predict the development trend, make decisions, control the system and carry out effective and reasonable evaluation of the state through data-based processing and phenomenon analysis under the background of uncertainty with little data and little information [4].

2. The basic principle of grey model

2.1. The structure of grey model

The basic idea is to use the original data to form the original sequence (0) and generate the sequence (1) by accumulative generation method, which can weaken the randomness of the original data and make it present a more obvious characteristic law. A differential equation model, namely GM model, is established for the sequence (1) generated after transformation. GM (1, 1) model represents the first-order, one-variable differential equation model. Among GM (1, 1) model group, metabolism model is the most ideal model [5]. This is because in the development process of any gray system, with the passage of time, there will be some random disturbances and driving factors into the system, so that the development of the system will be affected by it successively. When the GM (1, 1) model is used for prediction, the high accuracy is only 1-2 data after the origin data (0) (n), that is, the further the prediction time is, the weaker the prediction significance is. The basic idea of the metabolic GM (1, 1) model is that the closer the data, the greater the impact on the future. In other words, in addition to constantly adding new information, old information with little significance is re-moved. Such modeling sequence can more dynamically reflect the latest characteristics of the

system, which is actually a dynamic prediction model [6]. The grey model structure is shown in Figure 3.



Figure 3. Grey model overview

2.2. Data Processing

The training set of experimental data uses us daily gold settlement price data from January 02, 2008 to May 16, 2018 for a total of 2,290 days, and daily Bitcoin settlement price data from April 28, 2013 to August 14, 2018 [7, 8]. Then verify the data from September 11, 2018 to September 10, 2021.

Prices have different dimensional and dimensional units, which will affect the results of data analysis. In order to eliminate the dimensional influence between prices, data standardization is needed to solve the comparability between data. Therefore, for the price of gold and bitcoin, this article uses today's price/yesterday's price instead of today's price for subsequent predictions, which can make the data more substantial.

2.3. Establishment of GM (1, 1) Model

In this article, the initial relevant factor is a single primary indicator - the price of the transaction, with only one variable. Therefore, the GM (1, 1) model is to establish the grey model of this variable by using the first-order differential equation. Therefore, this article will use 30 days of initial data, assuming that the initial data is:

$$x_i^{(0)} = (x_i^{(0)}(1), x_i^{(0)}(2), \dots, x_i^{(0)}(n)) \quad (1)$$

Among them, data quantity $n = 30$; $i = 1$ means there's one variable; the superscript (0) represents the initial value.

For seemingly irregular original data, cumulative generation can be used to make the generated data change more continuous and closer to a specific function.

By adding $x_i^{(0)}$ to weaken the volatility and randomness of the random sequence, a new sequence can be obtained:

$$x_i^{(1)}(k) = \sum_{m=1}^n x_i^{(0)}(m), (k = m, 2, 3, \dots, n) \quad (2)$$

According to the above formula, the adjacent mean equal weight column of $x_i^{(1)}(k)$ is generated:

$$z_i^{(1)}(k) = (z_i^{(1)}(2), z_i^{(1)}(3), \dots, z_i^{(1)}(k)), k = 2, 2, 3, \dots, n. \quad (3)$$

Among them, $z_i^{(1)}(k) = 0.5x_i^{(1)}(k - 1) + 0.5x_i^{(1)}(k), k = 2, 3, \dots, n.$

According to the grey prediction theory, a first-order whitening differential equation about t is established for $x_i^{(1)}$:

$$\frac{dx_i^{(1)}}{dt} + a_i x^{(1)} = u_i \quad (4)$$

Among them, a_i and u_i are substitution coefficients, which are called development coefficient and grey action quantity respectively. The effective interval of a_i is $(-2, 2)$, and the matrix formed by a_i and u_i is denoted as grey parameter $\bar{a}_i = \begin{bmatrix} a_i \\ u_i \end{bmatrix}$. Only parameters a_i and u_i are required to calculate $x_i^{(1)}(t)$, and then the predicted value of $x_i^{(0)}$ can be calculated.

B and constant term vectors are generated for the mean value of the accumulated generated data:

$$B = \begin{bmatrix} -\frac{1}{2}(x^{(1)}(1) + x^{(1)}(2)) & 1 \\ -\frac{1}{2}(x^{(1)}(2) + x^{(1)}(3)) & 1 \\ \vdots & \vdots \\ -\frac{1}{2}(x^{(1)}(n - 1) + x^{(1)}(n)) & 1 \end{bmatrix}, Y_{ni} = \begin{bmatrix} x_i^{(0)}(2) \\ x_i^{(0)}(3) \\ \vdots \\ x_i^{(0)}(n) \end{bmatrix} \quad (5)$$

\bar{a}_i can be obtained according to the least square method:

$$\bar{a}_i = (B^T B)^{-1} B^T Y_{ni} \quad (6)$$

Replace grey parameter \bar{a}_i with $\frac{dx_i^{(1)}}{dt} + a_i x^{(1)} = u_i$, and $\frac{dx_i^{(1)}}{dt} + a_i x^{(1)} = u_i$ is solved to obtain:

$$\bar{x}_i^{(1)}(t + 1) = \left(x_i^{(1)} - \frac{u_i}{a_i}\right) e^{-a_i t} + \frac{u_i}{a_i} \quad (7)$$

The predicted value can be obtained by cumulative reduction of the above results:

$$\bar{x}_i^{(0)} = \left(\bar{x}_i^{(0)}(1), \bar{x}_i^{(0)}(2), \dots, \bar{x}_i^{(0)}(n + 1), \dots, \bar{x}_i^{(0)}(n + m)\right) \quad (8)$$

Among them, $\bar{x}_i^{(0)}(n + 1), \dots, \bar{x}_i^{(0)}(n + m)$ is the predicted value [9].

When the grey model experiment is implemented by programming, the predicted value must be checked. For the accuracy test of grey model, this paper uses the posterior difference test method. The posterior difference test method is publicized as follows:

$$S_1^2 = \frac{1}{n} \sum_{k=1}^n [x^{(0)}(k) - \bar{x}]^2 \quad (9)$$

$$S_2^2 = \frac{1}{n} \sum_{k=1}^n [e(k) - \bar{e}]^2 \quad (10)$$

$$C = \frac{S_2}{S_1} \quad (11)$$

The indexes C and p are two important indexes for posterior difference test. The smaller the index C , the better. A smaller C means that S_2 is large and S_1 is small. A large S_1 indicates a large variance of the original data, that is, a large degree of dispersion of the original data. Small S_2 means small residual variance and small residual dispersion. A small C indicates that although the original data is very discrete, the difference between the calculated value of the model and the actual value is not very discrete [10]. Grade judgment is made after calculating C and p , and the accuracy level is shown in Table 1.

Table 1. C and p accuracy prediction grades

Prediction accuracy/name	Excellent	Qualified	Barely qualified	Unqualified
p	>0.95	>0.8	>0.7	≤0.7
C	≤0.35	≤0.5	≤0.65	≥0.65

If it meets the criteria, it returns our predicted value.

3. Results

3.1. Calculation Results

In the grey model, three sample points are trained as a group, and these three sample points are known data. The predicted value of the next three days is calculated, and then the data of these three days is re-substituted into the gray model to get the predicted value of the next three days.

After calling the python software package, the GM (1, 1) calculation result of the gold price RMSE obtained in this paper is shown in Figure 4, and the calculation result of the bitcoin price RMSE is shown in Figure 5.

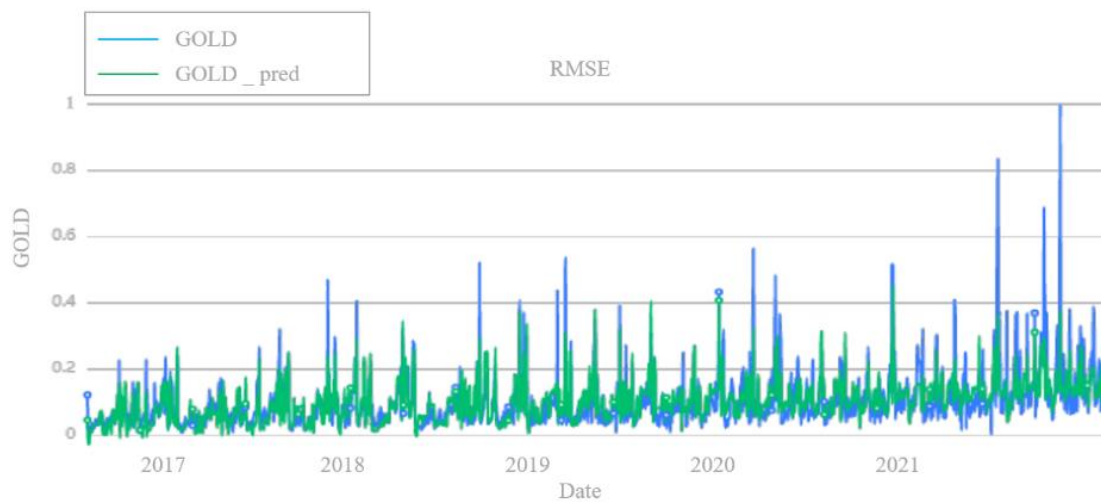


Figure 4. RMSE of gold

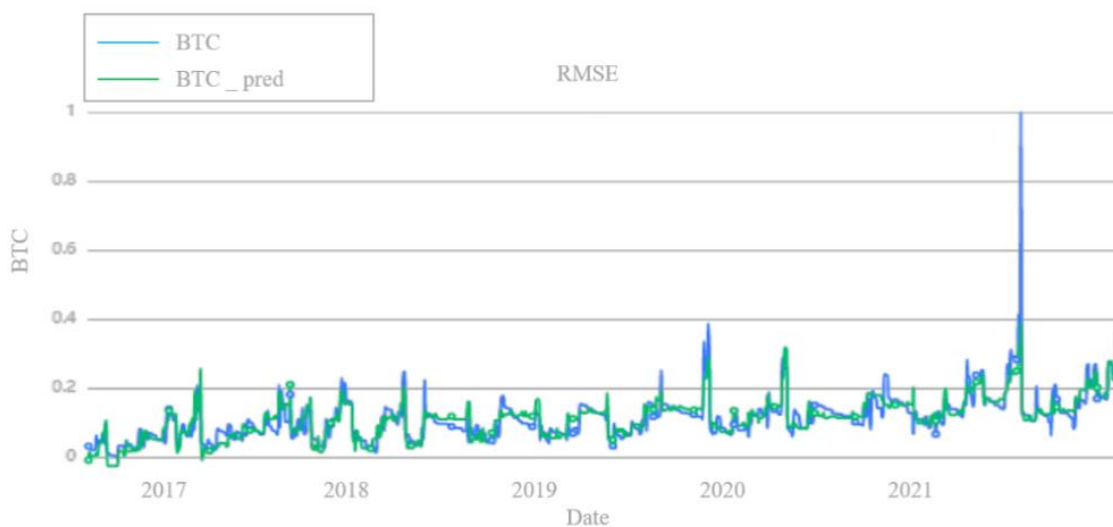


Figure 5. RMSE of bitcoin

The calculation result of gold price R^2 is shown in Figure 6, and the calculation result of bitcoin price R^2 is shown in Figure 7.

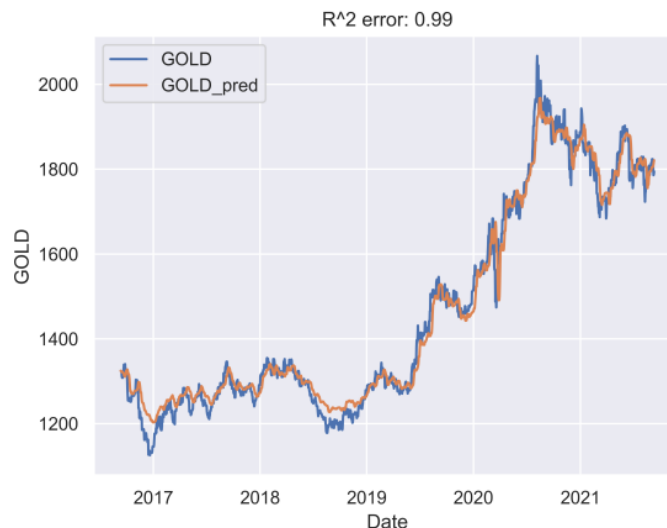


Figure 6. R^2 of gold

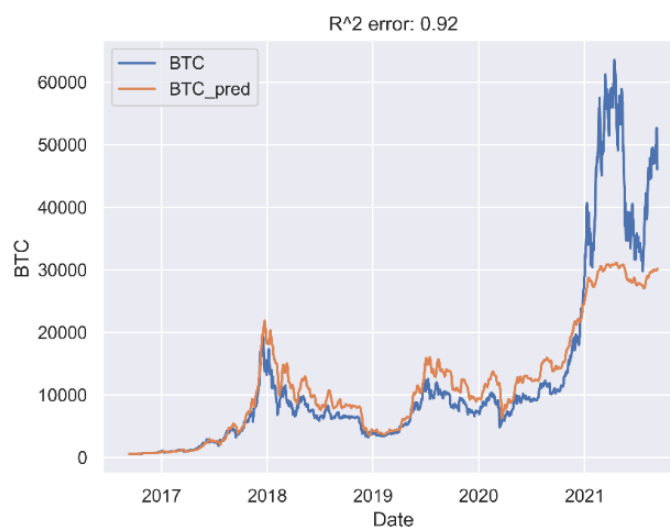


Figure 7. R^2 of bitcoin

You can see here that gold has a prediction error of about 0.99 (the closer to 1 the better) and Bitcoin has a prediction error of 0.92. Therefore, the prediction effect is good.

In addition, it can be clearly seen that the gray model predicted more accurate data in the first few years, but there were also some errors in the last two years, which gradually fell below the actual value.

3.2. Analysis of calculation results

In addition to calculating RMSE and R^2 of the prediction results, MAPE is also calculated here for convenience of comparison in order to more deeply analyze the accuracy of the prediction data. At the same time, the results predicted by GM and other models, such as LSTM neural network, are also compared. Nowadays, machine learning algorithm has been widely used in financial time series and proved to have excellent prediction effect in practice. The calculation results of gold are shown in Table 2 below.

Table 2. Gold price’s comparison of LSTM and GM

Method	Calibration			Validation		
	RMSE	MAPE	R^2	RMSE	MAPE	R^2
LSTM	0.188	18.3	0.87	0.168	15.7	0.91
GM	0.114	14.2	0.93	0.102	12.1	0.99

The calculation results of bitcoin are shown in Table 3 below.

Table 3. Bitcoin price's comparison of LSTM and GM

Method	Calibration			Validation		
	RMSE	MAPE	R ²	RMSE	MAPE	R ²
LSTM	0.251	20.1	0.82	0.202	17.3	0.88
GM	0.177	17.9	0.87	0.152	15.3	0.93

From the comparison between the forecast data of gold and bitcoin prices and the actual data, GM has better prediction performance and relatively small error, which can fully meet the demand, and the prediction speed is fast and the operation is convenient.

Under the conditions that the calculation requirements are met, the predicted value of the gold price in September 2021 is calculated to be \$1788, and the predicted value of the bitcoin gold price is \$46192, with an average margin of error of 1.7%.

4. Conclusion

The analytical trend of collecting more historical data of gold and bitcoin prices provides a basis for the establishment of price prediction models, but the traditional financial time series prediction model based on statistical methods cannot withstand such a huge consumption of time and computing resources. The primary premise of these methods is the linear assumption of financial time series variables, while the main tasks of grey forecasting are analysis, modeling, forecasting, decision making and control. Overfitting of large sample sets of time series prediction models will also affect the prediction accuracy. This paper identifies the degree of differences in development trends among system factors, that is, carries out correlation analysis, generates and processes the original data to find the law of system changes, generates data series with strong regularity, uses GM (1, 1) to establish the price prediction model of gold and bitcoin, so as to predict their future development trend. The grey prediction model is constructed with a series of quantitative values that reflect the characteristics of the predicted objects observed at isochronous intervals, and the powerful data processing function of Python program is fully utilized to prevent overfitting. The experimental results show that the grey prediction model has good predictability and robustness, and has certain practical application value.

References

- [1] Putro W A P, Prasetyo P W A, Gagas M P. Designing the Quick Cash Payment System prototype by using cash withdrawal services through minimarket as an alternative solution for secure money transfer from bank customers to non-bank customers [J]. IOP Conference Series: Materials Science and Engineering, 2020, 852 (1).
- [2] Xinchen Z, Linghao Z, Qincheng Z, et al. A Novel Bitcoin and Gold Prices Prediction Method Using an LSTM-P Neural Network Model [J]. Computational Intelligence and Neuroscience, 2022, 2022.
- [3] Yang Y, Xiong J, Zhao L, et al. A Novel Method of Blockchain Cryptocurrency Price Prediction Using Fractional Grey Model [J]. Fractal and Fractional, 2023, 7 (7).
- [4] Lan W, Nan L, Ming X. An Ensemble Learning Method for the Kernel-Based Nonlinear Multivariate Grey Model and its Application in Forecasting Greenhouse Gas Emissions [J]. Journal of Mathematics, 2022, 2022.
- [5] Yao T, Wang Z. Crude oil price prediction based on LSTM network and GM (1, 1) model [J]. Grey Systems Theory and Application, 2020, ahead-of-print(ahead-of-print).
- [6] Rathnayaka T K R, Seneviratna D. Taylor series approximation and unbiased GM (1, 1) based hybrid statistical approach for forecasting daily gold price demands [J]. Grey Systems: Theory and Application, 2019, 9 (1).
- [7] Kundu, Debasis. Stationary GE-Process and its Application in Analyzing Gold Price Data [J]. Sankhya B, 2021, 84 (2).

- [8] H. A A, H. H T A. Bitcoin Price Forecasting and Trading: Data Analytics Approaches [J]. Electronics, 2022, 11 (24).
- [9] Jin Fei. Optimization research and application of GM (1, 1), a dynamic model of grey system theory [D]. Yanshan University, 2012.
- [10] Xiao Li-Fang, Wang Hao, Duan Mei. Optimal combination method of grey model GM (1, 1) based on posterior difference test [J]. Mathematics in Practice and Understanding, 2019, 51 (12): 1-10.