Research On the Optimal Trading Strategy Based on The Grey Prediction Algorithm

Yijiao Chen[1, *], Xinyi Sun[2], Xuyuan Zhu[1]

[1]College of Internet of Things Engineering, Hohai University, Changzhou, Jiangsu, 213022

[2]Business College, Hohai University, Changzhou, Jiangsu, 213022

*Corresponding author: chenyijiao0303@163.com

Keywords: Curve fitting, Grey prediction, Consistency reliability, Pearson correlation coefficient.

Abstract: Investors can make significant gains in today's booming economy by buying and selling gold and bitcoin, two volatile assets. Given this market demand, this paper develops a model that uses daily price to date to determine whether traders should buy, hold or sell assets in their portfolios daily, based on the daily pricing of gold and Bitcoin. To maximize profits for the $1,000, it is invested over the five-year trading period from September 11, 2016, to September 10, 2021. To counter the problem of investment capital benefit maximization, this article obtains from the correlation analysis the initial data processing after using the curve fitting function. It is considered market VWAP trading strategy, taking the original data accumulation generation sequence, then set up the grey differential equation, the mathematical method of b-b computing, the price of the grey prediction model is established. Under this model, the optimal strategy is obtained in this paper: sell bitcoin shares when the amount of bitcoin shares exceeds 30% of the previous day's amount, and sell gold shares when the amount of gold shares exceeds 15% of the previous day's amount. If the price of another product drops after the transaction, use 65% of the proceeds to buy another product. Follow this strategy, and it will have a maximum investment value of $3340. To verify the optimality of the model after descriptive statistical analysis, the intrinsic consistency reliability of the model was analyzed by calculating the Kronbach coefficient, based on which the rationality of the model and its dependence on related factors were analyzed. Considering the precision of the machine learning model, this paper through the machine learning model for gold stocks and currency respectively to forecast stock trading results, which compares the results with the actual trading results. By observing its error size test of goodness of fitting model, the last of careful verification is proved the optimality of the model. Then, in order to study the strategy of the sensitive degree of transaction costs, the transaction costs on the result of policy and situation, this article calculates the correlation coefficient between stock returns and time mathematical methods, using the Pearson coefficient of deformation after as gold stocks and currency trading price impact on the strategic significance test. Then, the program is written and run through Python, and the influence degree of the rise and fall of the gold stock and bitcoin stock trading price on the strategy, the influence degree of the strategy on the stock return result, and the distribution of the final stock profit and loss are presented in the chart

1. Introduction

Recently, appropriate trading strategies have become increasingly significant for market traders to maximize total returns. Often by traders buy and sell assets, including gold and bitcoin. Based on existing data in nearly five years, significantly higher gold daily fluctuations in the prices in the past two years (Source: LondonBullionMarketAssociation, 9/11/2021). The daily price of bitcoin has fluctuated significantly over the past year (Source: NASDAQ,9/11/2021), so a reasonable trading strategy can create higher returns for market traders who buy and sell more considerable assets.

In order to analyze the given data more accurately and shorten the subsequent calculation process, we should first process the data by establishing the fitting polynomial curve function, which obtains the gold and Bitcoin stock fitting curve function, and intuitively present their characteristics by

drawing their area distribution map, histogram and line graph [1]. After the initial data processing, we should also follow this strategy and calculate the volume-weighted average price of gold and Bitcoin stocks according to its algorithm, considering the VWAP trading strategy that the market generally follows. Finally, we associate with the corresponding Price gray prediction model through the demand of price data prediction in the research question. After preliminary judgment, this model has high practicability and accuracy. The cumulative sequence is firstly generated according to the original price data of the stock to obtain the grey differential equation, establish the time corresponding function model of GM, and then carry out the reduction operation to obtain the prediction model of the original sequence.

2. Method

2.1. VWAP Trading Strategy

Volume weighted average price (VWAP) strategy is an algorithmic trading strategy that splits large orders and executes them in batches within the agreed period in order to make the final buy or sell average transaction price as close as possible to the average transaction price of the entire market in this period [2]. According to the VWAP strategy, the most important factors are as follows: historical trading volume, future trading volume forecast, market dynamics total trading volume, period of order opening.

According to the VWAP principle, the difference between the strategy's VWAP price and the market's VWAP price needs to be as small as possible. so, the mathematical formula is expressed as follows:

$$\begin{aligned} &\min_{y_1,y_2\cdots y_N} \ |(v_1^{\ s}p_1+v_2^{\ s}p_2+\cdots+v_N^{\ s}p_N)-(v_1^{\ m}p_1+v_2^{\ m}p_2+\cdots+v_N^{\ m}p_N)| \\ &st\{y_n\geq 0, x_n\geq 0, p_n\geq 0, n=1,2,\cdots,N \\ &v_n^{\ s}=y_n/(y_1+y_2+\cdots+y_N) \\ &v_n^{\ m}=x_n/(x_1+x_2+\cdots+x_N) \\ &V=y_1+y_2+\cdots+y_N\} \end{aligned} \tag{1}$$

According to the formula, in order to minimize the difference between strategy VWAP price and market VWAP price, $\mathbf{V_N^S = V_N^M}$ should be made, that is, the difference between the predicted intraday trading volume distribution of VWAP trading strategy and the actual intraday trading volume distribution of the market should be minimized. Therefore, dealers should trade at the predicted value of this volume distribution.

It is assumed that the trading time of a day is divided into N intervals, and VWAP is calculated according to the trading volume of each tick:

$$VWAP=\frac{x_1p_1+x_2p_2+\cdots x_Np_N}{x_1+x_2+\cdots+x_N}=v_1p_1+v_2p_2+\cdots+v_Np_N \tag{2}$$

$v_n=x_n/(x_1+x_2+...+x_N)$, it is indicated the proportion of the NTH transaction volume in the total transaction volume.

$$v_{t,n}^{\ h}=\frac{\hat{x}_{t,n}}{\hat{x}_{t,1}+\hat{x}_{t,2}+\cdots+\hat{x}_{t,N}}(n=1,2,\ldots N) \tag{3}$$

However, because the traditional model only analyzes the historical data to determine how to open orders, the results will appear relatively large error. Therefore, this paper mainly uses the improved dynamic model, which can introduce the facts, except in the areas where the dynamic model is not applicable, the static model is used instead.

The estimator of the correction coefficient is calculated by:

$$\hat{\varphi}_{i,tn}=v^{m}_{i,tn}\hat{v}^{h}_{i,t} \tag{4}$$

2.2. Price Grey Forecasting Model

Grey prediction is to identify the different degrees of development trend of each factor in the system by calculating the correlation degree between each factor. Its core system is GreyModel (GM), which is a modeling method that accumulates the original data (or accumulates averages and other methods) to generate approximate exponential laws [3].

Combining the application principle of the model with the demand of price data prediction in the task, it can be seen that the model has high practicability and accuracy.

The cumulative sequence is generated from the original price data x of the stock:

$$x^{(1)} = \{x^{(1)}(1), x^{(1)}(2), \cdots, x^{(1)}(k), \cdots x^{(1)}(n)\} \tag{5}$$

Among them:

$$x^{(1)}(k) = \sum_{i}^{k} x^{(1)}(i), (k = 1,2, \cdots, n) \tag{6}$$

Since the sequence $x^{(1)}(k)$ has an approximate exponential growth law, and the solution of the first-order differential equation happens to be in the exponential growth form, the sequence satisfies the first-order linear differential equation model:

$$\frac{dx^{(1)}}{dt} + ax^{(1)} = b \tag{7}$$

There are grey differential equations:

$$x^{(0)}(k) = -aZ^{(1)}(k) + b \tag{8}$$

The final strategy is

$$\hat{x}^{(1)}(k+1) = [x^{(0)}(1) - \frac{b}{a}]e^{-ak} + \frac{b}{a}, (k = 0,1,2, \cdots) \tag{9}$$

In this way, the time corresponding function model of GM can be obtained, and the prediction model of the original sequence $x^{(0)}$ can be obtained after the reduction operation:

$$\hat{x}^{(0)}(k+1) = \hat{x}^{(1)}(k+1) - \hat{x}^{(1)}(k), (k = 0,1,2, \cdots) \tag{10}$$

3. Results and discussion

3.1. Final Strategy and Results

The final strategy is to sell bitcoin as soon as it exceeds 30% of the previous day's value and gold as soon as it exceeds 15% of the previous day's value. If the price of another product drops after the transaction, use 65% of the proceeds to buy another product.

As of October 9, 2021, the initial $1000 investment would be worth $3340.

3.2. Descriptive Statistical Analysis

(1). Sensitivity analysis of rider deviations from the target power distribution

According to the requirements in the question, it is aimed to verify that the above mathematical model can provide the best strategy. We first conduct descriptive statistical analysis on the data of the application model.

Descriptive statistics refers to characterizing data by tabulating and classifying, graphing, and calculating general data. The descriptive statistical analysis describes the relevant data of all variables in the survey, including frequency analysis, central trend analysis, dispersion degree analysis, distribution, and some basic statistical graphs [4].

Table 1. The statistical results.

	USD(PM)	Values
count	1255.000000	1826.000000
mean	1464.549402	12206.068281
std	249.291812	14043.891627
min	1125.700000	594.080000
25%	1266.175000	3994.982500
50%	1329.050000	7924.460000
75%	1723.750000	11084.730000
max	2067.150000	63554.440000

On the one hand, table 1 calculates the daily closing price of gold in US dollars per ounce during this period: count, mean, arithmetic standard deviation, maximum, minimum, 25% quantile, median, and 75% quantile. On the other hand, table 1 calculates the parameters of the dollar price of a single bitcoin every day during this period: count, mean, arithmetic standard deviation, maximum, minimum, 25% quantile, median, and 75% quantile.

(2). The Reliability Analysis

Reliability refers to the degree of consistency of results obtained by repeated measuring the same object using the same method. Reliability indexes are expressed mainly by correlation coefficients, which can be roughly divided into three categories: stability coefficient (consistency across time), equivalent coefficient (consistency across forms), and internal consistency coefficient (consistency across projects) [7]. Observed value = true value + error value. The smaller the error, the more reliable. The reliability is divided into retest reliability, duplicate reliability, internal consistency reliability, and grader reliability. According to the consistency reliability, Cronbach's coefficient was calculated by calculating the direct correlation coefficient between individual score and total score and compared with the variability of each score. The calculation formula is:

$$\partial = (\frac{k}{k-1})(\frac{s_y^{\ 2} - \sum s_i^{\ 2}}{s_y^{\ 2}}) \tag{11}$$

Based on the rationality of the analysis model and dependence on relevant factors, reliability analysis is an effective analysis method to verify whether the data credibility has stability and reliability so that the strategy obtained by the model is convincing. The analysis was mainly carried out by python programming, and the statistical reliability result was 0.87.

(3). KS test

The goodness of the fitting test calculates the expected frequency of each category in the classification variables based on the overall distribution. It compares it with the observation frequency of the distribution to judge whether there is a significant difference between the expected frequency and observation frequency to achieve the purpose of analysis from the classification variable [5]. The most natural idea is to measure the distance between the empirical distribution function Fn(x) and the fitted distribution function F(x). The smaller the distance, the better the fitting effect. This distance is usually composed of its industry or quadratic norm to measure, according to the empirical distribution function Fn(x) and the fitting of the distribution function F(x) of the distance between the statistics as the empirical distribution function (EDF) statistics, remember these EDF statistics for T [6, 7]. After calculating these statistics, our goal is to determine whether the distribution F(x) is acceptable based on these statistics, that is, to test its goodness of fit with the empirical distribution.

Because the machine learning model is effective and accurate, we use the machine learning model to predict the trading results of gold and bitcoin stocks, respectively. The results are shown in Figure 1 and Figure 2.

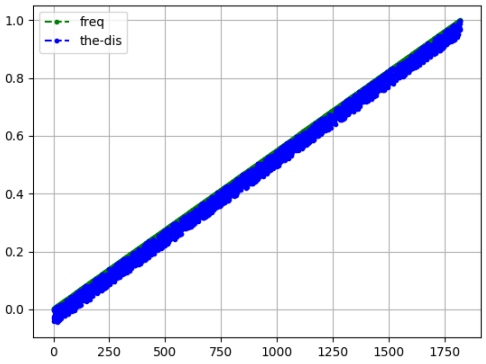

Figure 1 Bitcoin transaction result forecast.

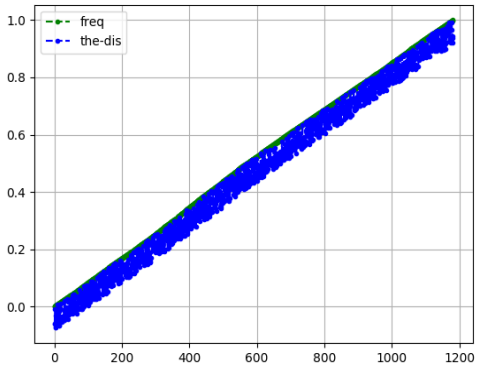

Figure 2 Gold trading result forecast.

Figure 1 shows bitcoin stocks, and Figure 2 illustrates gold stocks. The green straight line represents the transaction price prediction result obtained after machine learning, and the blue line represents the actual transaction price. The abscissa represents the gold or bitcoin stock on day i; The ordinate represents the transaction price standardization index. It can be seen from the figure that the fitting effect is excellent, which proves that our model realizes the best trading strategy.

3.3. Sensitive a Strategy is to Transaction Costs

(1). Correlation Coefficient between Stock Returns and Time

In statistics, the Kendall correlation coefficient is named after Maurice Kendall and is often denoted by the Greek letter τ. Kendall's correlation coefficient is a statistic used to measure the correlation between two random variables. A Kendall test is a parameterless hypothesis test that uses a calculated correlation coefficient to test the statistical dependence of two random variables. The Kendall correlation coefficient ranges from -1 to 1. $\tau = 1$ indicates that the two random variables have the same rank correlation. $\tau = -1$ indicates that the two random variables have an opposite rank correlation. $\tau = 0$ means that the two random variables are independent.

Suppose that the two random variables are X and Y (which can also be regarded as two sets), and their number of elements is N, the ith ($1 \le i \le N$) value of the two random variables is represented by X_i and Y_i respectively. The corresponding elements in X and Y form a pair of elements XY, which contains elements (X_i, Y_i) ($1 \le i \le N$). The two elements are considered consistent when $X_i > X_j \cup Y_i > Y_j$ or $X_i < X_j \cup Y_i < Y_j$. The two elements are considered inconsistent when $Xi > Xj \cup Yi < Yj$ or $Xi < Xj \cup Yi > Yj$. When $X_i = X_j$ or $Y_i = Y_j$, the two elements are neither consistent nor inconsistent.

The data and background conditions given in the question are consistent with the fact that neither set 1(gold stock) nor set 2(Bitcoin stock) has the same element (each element in the set is unique). Therefore, the deformed Pearson coefficient is used here for verification:

$$\rho_{X,Y} = \frac{\sum XY - \frac{\sum X \sum Y}{N}}{\sqrt{(\sum X^2 - \frac{(\sum X)^2}{N})(\sum Y^2 - \frac{(\sum Y)^2}{N})}} \tag{12}$$

Therefore, the correlation coefficient matrix between gold and Bitcoin stock returns and time is shown in Table 2 and 3:

Table 2. Gold stock returns and time.

kendallGoldprob3				
	date	value	e_rate	e_total
date	1	0.65385553	0.026535232	0.026535232
value	0.65385553	1	0.049168573	0.049168573
e_rate	0.026535232	0.049168573	1	1
e_total	0.026535232	0.049168573	1	1

Table 3. Bitcoin stock returns and time

kendallBtbprob3				
	date	value	e_rate	e_total
date	1	0.656772402	0.031310542	0.031311133
value	0.656772402	1	0.046282422	0.046283009
e_rate	0.031310542	0.046282422	1	0.9999997
e_total	0.031311133	0.046283009	0.9999997	1

According to the above matrix, the influence of the strategy on the trading price, total return, and return rate of gold and bitcoin stocks can be visually presented in the pie chart, as shown in Figure 3.

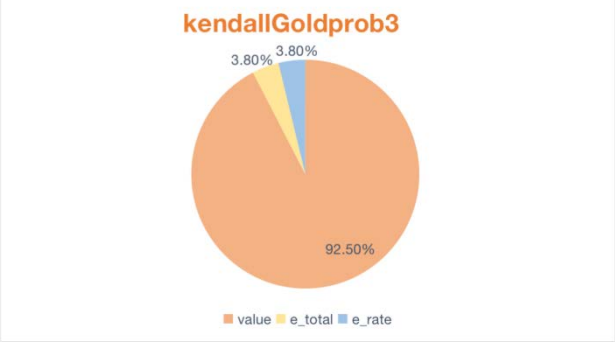

Figure 3 Degree of gold-stock return.

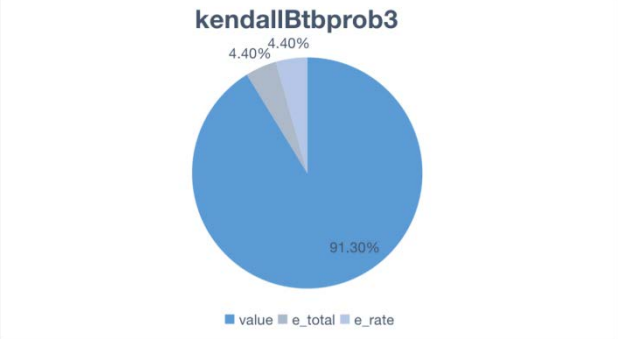

Figure 4 Degree of bitcoin stock return.

As can be seen from the figure, strategies have a 92.5% impact on the trading price of gold stocks, 3.8% on the total return of gold stocks, and 3.8% on the return rate of gold stocks. Strategies accounted for 91.3% of the impact on the trading price of bitcoin shares, 4.4% of the impact on the total return of bitcoin shares, and 4.4% of the impact on the return rate of bitcoin shares.

(2). Significance test of the impact of gold and bitcoin stocks on strategy

Through machine learning, the influence of the rise and fall of the trading price of the gold stock and bitcoin stock on the strategy can be obtained, as shown in the table:

Table 4. The extent to which the trading price of gold stocks and bitcoin stocks rises and falls affects the strategy

Significance Test①									
Model	R	R^2	R^2 after adjustment	errors in standard estimates	change the statistical				
					change in R^2	F variation	degrees of freedom 1	degrees of freedom 2	change in significance F
1	5.4 ②	25.4	0.005	98.4389	0.254	1.021	9	27	128
① Predictive variables :(constant), Gold stock price, Bitcoin stock price, Bitcoin stock price turn.									
② The dependent variable: Total earnings of bitcoin stock price on the same day.									

As shown in the figure, the accuracy rate of the results after model verification is more than 70%. The rise and fall of the transaction price significantly impact the strategy. The sharp fall of gold has a significant impact on the final return, while the rise and fall of bitcoin has a small impact on the final return because we have minimized the risk of bitcoin transactions in the strategy.

Finally, the distribution of stock profit margin and loss rate can be obtained, as shown in the figure:

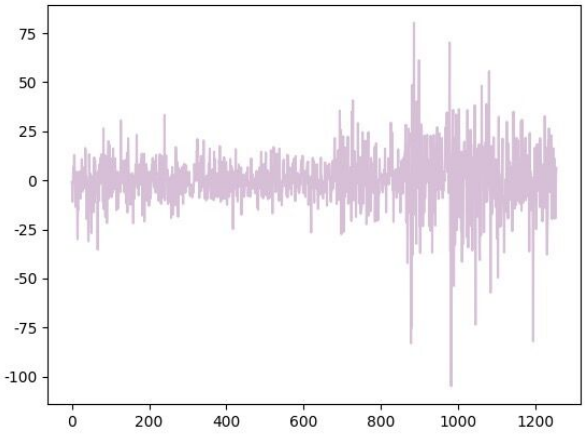

Figure 5 The distribution of profit rate and loss rate of stocks.

It can be seen from the figure that before the 850th day, the stock showed relatively stable ups and downs of profit and loss. Between the 850th day and the 1200th day, the stock showed relatively stable ups and downs of profit and loss, but it still fluctuated around the 0 lines in general, showing a balance of profit and loss in general, reflecting the law of the market and its regulating effect.

4. Conclusion

For modeling analysis, we first generate the cumulative sequence according to the original price data of the stock to obtain the grey differential equation, establish the time corresponding function model of GM, and then carry out the reduction operation to obtain the prediction model of the original sequence. By running the program, the optimal strategy can be obtained to maximize the initial $1,000 investment value by October 9, 2021.

For the optimality analysis of the test model, we need to verify whether the mathematical model established above can provide the best strategy. Based on the model established above, the most essential descriptive statistical analysis of application model data. It can take all the data and calculate the daily dollar closing price of an ounce of gold and the daily dollar price of a single bitcoin during this period: count, mean, arithmetic standard deviation, maximum, minimum, 25% quantile, median, 75% quantile. Then, we carry out a reliability analysis. By calculating the Kronbach coefficient, we get the internal consistency reliability of the model and analyze its rationality and dependence on related factors. Finally, the goodness of fit of the model is tested. Due to the good effect of the machine learning model, we predicted the transaction results of gold and bitcoin respectively through the

machine learning model and then compared with the actual daily transaction results to determine whether there is a significant difference between the expected frequency and the observed frequency, which is used to achieve the purpose of analysis from classification variables. The smaller the distance between the empirical distribution function Fn(x) and the fitted distribution function F(x) is, the better the fitting effect is. According to this standard, we will get the goodness of fit, thus completing the test of model optimality.

To analyze the impact of transaction costs on trading strategies, we need to study the sensitivity of strategies to transaction costs, that is, the impact of transaction costs on strategies. According to the requirements shown in the question, we first need to calculate the correlation coefficient between stock returns and time. As can be seen from the analysis of the data and background conditions given in the question, the deformed Pearson coefficient should be used as the significance test of the trading price of the gold stock and bitcoin stock. At the same time, the impact of the rise and fall of the trading prices of gold stocks and bitcoin stocks on the strategy, the impact of the strategy on the stock return results, and the distribution of the final stock profit and loss can be intuitively presented in charts.

References

[1] Shufeng Li,Pengyu Chen. Unbiased GM (1,1) Power Model and its Application, Statistics and Information Forum, 2010-11

[2] Weiming Chen. Curve fitting principle and its application research, Changsha University of Science and Technology, 2014-07

[3] Kai Li. The Grey Prediction Model and Its Application, Changjiang University, 2012-05

[4] Yi-lin Sun, Ari Tarigan, Owen Waygood & Dian-hai Wang. Diversity in diversification: an analysis of shopping trips in six-week travel diary data, Journal of Zhejiang University-SCIENCE A volume 18, pages234–244 (2017)

[5] Dandan Guan, Houcan Zhang, Zhongquan Li. Relience analysis of the difference scores, psychological science, 2005-01

[6] Frank J. Massey Jr. The Kolmogorov-Smirnov Test for Goodness of Fit, Journal of the American Statistical Association, Volume 46, 1951 - Issue 253

[7] Zezhi Li.Empirical and theoretical distribution functions, Journal of Xiamen University, 2012-09