Financial product investment decision model based on ultimate benefit

Yunhan Zhong*

School of Economics, Jinan University, Guangdong, China

*Corresponding author:vj162774@163.com

Abstract. With the vigorous development of the financial market, more and more people enter the financial market for investment, it is very necessary to build a quantitative analysis and decision model based on data for ordinary traders. This paper selects two time series prediction models GM(1,1) and ARIMA for comparison. For GM(1,1), we improve it to obtain unbiased GM(1,1) and unbiased metabolic GM(1,1). Unbiased metabolism GM(1,1) was selected from the three grey prediction, and three combinations were formed by ARIMA and grey prediction respectively.These three results are used separately for the decision model, which prediction model to choose is determined by the final return. The investment decision model established in this paper is helpful for financial practitioners to improve investment profit and has certain application value.

Keywords: Unbiased Metabolic GM (1, 1), ARIMA, Return on investment.

1. Introduction

With the thriving of the financial market and the continuous prosperity of people's life, more and more people will put their spare money into the financial market for investment[1], in order to maximize the value of the existing assets. For this purpose, many people invest in volatile financial assets like gold and bitcoin with the mentality of a gambler, hoping to take advantage of sharp short-term fluctuation and profit from price differences. Unfortunately, it is hard for ordinary people to predict the price changes of financial products accurately [2], so most people end up losing everything. Therefore, it is very necessary to construct a quantitative analysis and decision model based on data for common traders[3].

2. GM(1,1) Model

The best strategy is defined as maximizing daily returns. Returns are determined by commodity prices after that day, so here we need to build a predictive model to estimate future prices[4]. However, since the forecasting model usually needs to rely on a certain amount of original data, by looking for a large number of literatures, in order to ensure the good forecasting effect of the model in the short term, we consider two models[5]: GM(1,1) model and ARIMA(p,d,q) model.

2.1 Traditional GM(1,1) Model

Let the original data sequence be a strict exponential sequence.

$$x^{(0)}(k) = Ae^{-a(k-1)} \quad k = 1,2,\cdots N \tag{1}$$

After one accumulated generating operation, the sequence is as follows:

$$x^{(1)}(k) = A\frac{1-e^{ak}}{1-e^{a}} \quad k = 1,2,\cdots N \tag{2}$$

Where $x^{(0)}(k)$ is the kth value of the original sequence $x^{(1)}(k)$ is the kth value of the new sequence, a, A are the parameters to be estimated, and N is the number of data in the sequence[6]. According to the traditional GM(1,1) model, we construct the following two matrices[7]:

$$B=\begin{pmatrix} -\frac{1}{2}A\frac{2-e^{a}-e^{2a}}{1-e^{a}} & 1 \\ -\frac{1}{2}A\frac{2-e^{2a}-e^{3a}}{1-e^{a}} & 1 \\ \vdots & \\ -\frac{1}{2}A\frac{2-e^{(N-1)a}-e^{Na}}{1-e^{a}} & 1 \end{pmatrix} \tag{3}$$

$$Y_n=\begin{pmatrix} Ae^{a} \\ Ae^{2a} \\ \vdots \\ Ae^{(N-1)a} \end{pmatrix} \tag{4}$$

Then the estimated relation between parameters a and A can be represented:

$$(\hat{a}\hat{A})^T=(B^TB)^{-1}B^TY_n=\begin{pmatrix} \frac{2(1-e^{a}}{1+e^{a}} \\ \frac{2A}{1+e^{a}} \end{pmatrix} \tag{5}$$

The final fitting results are:

$$\hat{x}(1)=A$$
$$\hat{x}(k)=\frac{-Ae^{a}(1-e^{\hat{a}})}{(1-e^{a})}e^{-\hat{a}(k-1)} \quad k=2,3\cdots,N \tag{6}$$

By comparing (1) with (6), it can be seen that (1) is different from (6), that is, the traditional GM(1,1) model fails when the growth rate of the original data series is large. Therefore, it is necessary to improve it[8].

2.2 Unbiased GM(1,1) model

From (5), we have $a=\ln\left(\frac{2-\hat{a}}{2+\hat{a}}\right), A=\frac{2\hat{A}}{2+\hat{a}}$. The estimated parameters $\hat{a}$ and $\hat{A}$ can be used to present a and A.Suppose that the model established for the exponential series is

$$\hat{x}^{(0)}(k)=\hat{A}'e^{\hat{a}'(k-1)} \quad k=1,2,\cdots N \tag{7}$$

where $\hat{a}'=ln\frac{2-\hat{a}}{2+\hat{a}}$, $\hat{A}'=\frac{2\hat{A}}{2+\hat{a}}$.

Then (7) is the unbiased estimated GM(1,1) model of (1). However, as the data are always updating and changing, the information reflected by the old data become meaningless. To get rid of useless data, we can remove the old data. This leads to a more accurate model[9].

2.3 Unbiased Metabolic GM(1,1) model

Set the original series as $X^{(0)}=\left(x^{(0)}(1),x^{(0)}(2),\cdots,x^{(0)}(n)\right)$, and have the following operations on this series.

1. The GM(1,1) model built by the series $X^{(0)}=\left(x^{(0)}(1),x^{(0)}(2),\cdots,x^{(0)}(n)\right)$ is called Full-Data GM(1,1) model[10].

2. For any $k_0>1$, where k_0 is any number that is larger than 1 and smaller than n, the model constructed by the series $X^{(0)}=\left(x^{(0)}(k_0),x^{(0)}(k_0+1),\cdots,x^{(0)}(n)\right)$ is called Partial-data GM(1,1) model.

3. Set $x^{(0)}(n+1)$ as the latest data, and put $x^{(0)}(n+1)$ into $X^{(0)}$. The model constructed by $X^{(0)}=\left(x^{(0)}(1),x^{(0)}(2),\cdots,x^{(0)}(n+1)\right)$ is New-data GM(1,1) model.

4. The process of putting the latest data $x^{(0)}(n+1)$ and removing the oldest data $x^{(0)}(1)$ is metabolic. With the Metabolic theory, the unbiased model constructed by $X^{(0)} = \left(x^{(0)}(2), x^{(0)}(3), \cdots, x^{(0)}(n+1)\right)$ is Unbiased Metabolic GM(1,1) model.

2.4 Comparison between the GM(1,1) price prediction results

It is known that GM(1,1) is suitable when the dataset is small. The small dataset always contains seven to ten data. Therefore, we select eight gold price data, nine gold price data and ten gold price data from September 12th 2016 respectively as the known information, use the three GM(1,1) models to predict the gold price in the following three days and calculate the Mean Squared Error. When Mean Squared Error is large, it means the model may not be suitable. MSE(dollar/troy ounce) can be represented as follows:

$$MSE = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n}} \tag{8}$$

After using MATLAB to calculate MSE, the results are in the following table1.

Table1. Mean squared errors of different GM(1,1)

number of data for prediction	Traditional GM(1,1)	Improved GM(1,1)	Unbiased Metabolic GM(1,1) model
8	43.1731	23.1740	40.1731
9	15.8544	49.8760	14.8545
10	20.1330	57.8697	18.1328

It is shown that Unbiased Metabolic GM(1,1) model would be the best and the model is most suitable when using 9 past prices to predict 3 future prices. Therefore, in the following session, we use Unbiased Metabolic GM(1,1) model to predict prices in the short term.

3. ARIMA model

3.1 Introduction

If a time series $\{X_t\}$ is represented as follows:

$$X_t = \varphi_0 + \varphi_1 X_{t-1} + \varphi_2 X_{t-2} + \cdots + \varphi_p X_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \cdots - \theta_q \varepsilon_{t-q} \tag{9}$$

Then (9) is an Autoregressive Moving Average Model, which is denoted as ARMA(p, q).
If a time series $\{X_t\}$ is d-differenced, where $W_t = \nabla^d X_t$ and W_t is ARMA(p, q), that is

$$W_t = \varphi_1 W_{t-1} + \varphi_2 W_{t-2} + \cdots + \varphi_p W_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \cdots - \theta_q \varepsilon_{t-q} \tag{10}$$

Then (10) is an Autoregressive Integrated Moving Average Model, which is denoted as ARIMA(p, d, q).

3.2 Basic step of ARIMA(p, d, q)

(1) Use the law of the ADF unit root to test the variance, trend and seasonal variation and identify whether the series is stationary or not, according to the scatter diagram, autocorrelation function and partial autocorrelation function diagram of the time series,.

(2) Perform stationary process on non-stationary series.

(3) Establish the corresponding model according to the identification rules of the time series model.

(4) Estimate the parameters, and then test them for significance to determine whether they are statistically significant.

(5) Perform hypothesis test to diagnose whether the residual sequence is white noise. Justify whether the model is reasonable or not.

(6) Use the fitted ARIMA model for predictive analysis.

3.3 Unbiased Metabolic GM(1,1) model and ARIMA model Result Analysis

Using two models at the same time, under the condition of known gold price data volume of 30 days, 50 days, 70 days, 90 days and 100 days respectively, predict the gold prices of the next three days after the current day. The mean square errors of two models with different dataset are shown in the following table2.

Table 2. MSE of Unbiased Metabolic GM(1,1)

Amount of data used for prediction	MSE of Unbiased Metabolic GM(1,1)
10	18.1328
30	25.5882
50	34.6333
70	25.7996
90	69.7653
100	80.7821

The table shows that as the amount of data used for prediction increases, the mean square error of the model increases gradually, indicating that the applicability of gray prediction decreases with the increase of sample size, which verifies that the model is suitable for prediction when the amount of data is small (7~10 data).

Because using the ARIMA model in MATLAB requires more than 10 known data volumes, the applicability test for this model starts with a data volume of 30.

Table 3. MSE of ARIMA

Amount of data used for prediction	MSE of ARIMA
30	38.0016
50	11.8521
70	5.8976
90	13.5080
100	21.2892

As the amount of sample data increases, the mean square error of the ARIMA model first decreases and then increases, and when the amount of data is 70, the mean square error is the smallest.

In general, the ARIMA model is more suitable when the amount of data is large, while the GM(1,1) model is better when the amount of data is small. Therefore, in the subsequent use of the predictive model, we select different combinations of the two models according to different sample data volumes.

4. The construction of the price prediction models

After comparing the differences between GM(1,1) model and ARIMA model, and knowing the best application scope of each model, we can construct three different price prediction models and predict the prices of commodities in the following three days.

Prediction Model 1: From the analysis before, it is found that ARIMA model is most suitable when the number of known dataset is 70. Therefore, the first model is only using ARIMA model and assumes traders do not make trading decision during the first 70 days.

Prediction Model 2:The second model is the combination of Unbiased Metabolic GM(1,1) model and ARIMA model. Because Unbiased Metabolic GM(1,1) model is suitable for prediction when the dataset is small, especially when the number of known dataset is nine. Therefore, the second model is that from day 10 to day 70, Unbiased Metabolic GM(1,1) model is used for price prediction and after day 70, ARIMA is used for prediction.

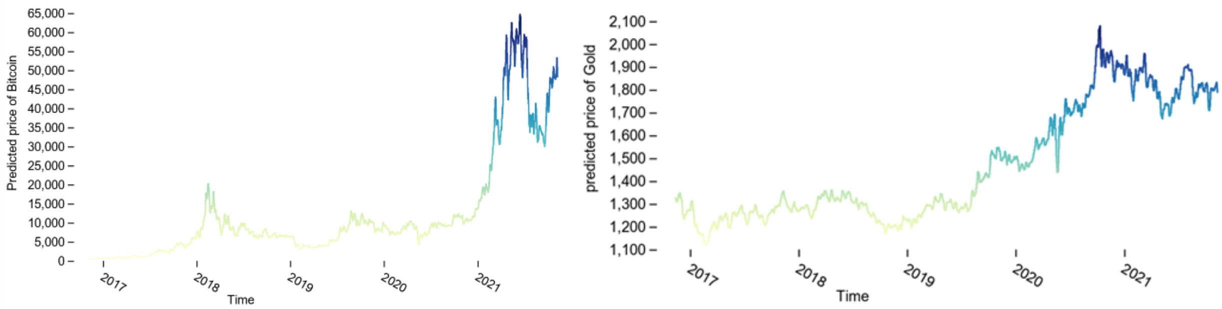
Prediction Model 3:The third model is that from day 10, the future prices are predicted by the Unbiased Metabolic GM(1,1).

The algorithms of the above three models are similar. Here the algorithm of model 2 is shown below because it is the combination of model 1 and model 3.

Algorithm 2: unbiased Metabolic GM (1,1) and ARIMA model
Input: The last trading day T_{total}, the true price of Bitcoin and Gold. i.e. GP_t and BV_t
Output: The predicted price of Bitcoin and Gold i.e. BV_t' and GP_t'

for $t = 0$ to T_{total} **do**
  if $t \geq 9$ **and** $t \leq 70$ **do**
    Using unbiased Metabolic GM (1,1) model predicted $t + 1$ day's price i.e. BV_{t+1}' and GP_{t+1}' based on the true value from $t - 9$ to t
    Calculate Gr_t, GR_t, Br_t and BR_t based on Equations (11), (12), (13) and (14)
  else if $t > 70$ **do**
    Using ARIMA model predict $t + 1$ day's price i.e. BV_{t+1}' and GP_{t+1}' based on the true value from $t - 70$ to t
    Calculate Gr_t, GR_t, Br_t and BR_t based on Equations (11), (12), (13) and (14)
  end
end

Price prediction by Unbiased Metabolic GM(1,1) is shown as figure1.



(a)Gold (b) Bitcoin

Figure 1. Price prediction by Unbiased Metabolic GM(1,1)

5. Strategy Decision Model

To select the best trading strategy each day, we solve it using a nonlinear programming model with the goal of maximizing the daily and total return.

5.1 Decision variable

Set x_{1t} as the amount of gold bought, x_{2t} as sold, x_{3t} as the number of Bitcoins bought and x_{4t} as sold. They should be all integers. $[x_{1t}, x_{2t}, x_{3t}, x_{4t}]$ is the decision variable vector.

5.2 Objective function

In order to construct the objective function, this paper creates some indicators and sets some parameters by using the forecast price. The relevant symbols, calculation formulas and explanations are shown in Table 4 and Table 5.

Table 4 Description of gold

symbol	formula	instructions
Gr_t	$Gr_t = \frac{GP'_{t+1} - GP_t}{GP_t}$	Gr_t is the expected return rate tomorrow and GR_t is the expected growth rate over next three days, where GP_t is the price on trading day t, GP'_{t+1} is the predicted price of gold tomorrow and GP'_{t+3} is the predicted price of gold after three days.
GR_t	$GR_t = \frac{GP'_{t+3} - GP_t}{GP_t}$	
α_{gold}		The transaction cost rate of gold

Table 5 Related description of Bitcoin

symbol	formula	instructions
Br_t	$Br_t = \frac{BV'_{t+1} - BV_t}{BV_t}$	As for bitcoin, denote BV_t as the price on trading day t, Br_t as the expected return rate of bitcoins tomorrow, and BR_t as the expected growth rate over next 3 days.
BR_t	$BR_t = \frac{BV'_{t+3} - BV_t}{BV_t}$	
$\alpha_{bitcoin}$		The transaction cost rate of Bitcoin

In real market, the minimum transaction volume of Bitcoin is 0.01 coin, while it is 100 troy ounces for gold. Therefore, the expected daily return per day is

$$\left(Gr_t - \alpha_{gold}\right)(x_{1t} - x_{2t}) \cdot GP_t \cdot 100 + (Br_t - \alpha_{bitcoin}) \cdot (x_{3t} - x_{4t}) \cdot BV_t \cdot 0.01 \tag{11}$$

The assumption is that traders will immediately deposit the cash they get from selling Bitcoin and gold at the bank to earn interest.By searching online, the interest rate of checking accounts in the United States, r_{bank}, is 0.03%. The interest is computed as follows.

$$(x_{2t} \cdot GP_t \cdot 100 + x_{4t} \cdot BV_t \cdot 0.01) \cdot r_{bank} \tag{12}$$

Consider that traders would take future short-term returns into account, so the expected return 3 days later is also added in the objective function but has a smaller weight showing that it has a smaller effect on daily decisions.

$$\left(GR_t - \alpha_{gold}\right)(x_{1t} - x_{2t}) + (BR_t - \alpha_{bitcoin})(x_{3t} - x_{4t}) \tag{13}$$

Set z as the expected short-term return, then the objective function is

$$\max z = \left(Gr_t - \alpha_{gold}\right)(x_{1t} - x_{2t}) \cdot GP_t \cdot 100 + (Br_t - \alpha_{bitcoin}) \cdot (x_{3t} - x_{4t}) \cdot BV_t \cdot 0.01 + (x_{2t} \cdot GP_t \cdot 100 + x_{4t} \cdot BV_t \cdot 0.01) \cdot r_{bank} + \left(GR_t - \alpha_{gold}\right)(x_{1t} - x_{2t}) + (BR_t - \alpha_{bitcoin})(x_{3t} - x_{4t}) \tag{14}$$

5.3 Constraints

This paper assumes that traders can either buy or sell at one day. Therefore, this paper set the following constraints.

$$x_{1t} \cdot x_{2t} = 0 \tag{15}$$

$$x_{3t} \cdot x_{4t} = 0 \tag{16}$$

The amount of money spent on transaction that day should be less than the remaining money after yesterday's decision.

$x_{1t} \cdot GP_t \cdot 100 + x_{3t} \cdot BV_t \cdot 0.01 \le C_{t-1}$

Besides, if traders want to sell gold or Bitcoin, they must ensure that the amount of gold or Bitcoin they hold after yesterday decision should be greater than the amount they want to sell today.

$$x_{2t} \cdot 100 \le G_{t-1} \tag{17}$$

$$x_{4t} \cdot 0.01 \le B_{t-1} \tag{18}$$

Where

$$G_t = G_{t-1} + (x_{1t} - x_{2t}) \cdot 100 \tag{19}$$

$$B_t = B_{t-1} + (x_{3t} - x_{4t}) \cdot 0.01 \tag{20}$$

$$C_t = (1000 - G_t \cdot GP_t \cdot 100 - B_t \cdot BV_t \cdot 0.01)(1 + r_{bank})^t \tag{21}$$

In conclusion, the nonlinear programming problem is

$$\max z = \left(Gr_t - \alpha_{gold}\right)(x_{1t} - x_{2t}) \cdot GP_t \cdot 100 + (Br_t - \alpha_{bitcoin}) \cdot (x_{3t} - x_{4t}) \cdot BV_t \cdot 0.01 + (x_{2t} \cdot GP_t \cdot 100 + x_{4t} \cdot BV_t \cdot 0.01) \cdot r_{bank} + \left(GR_t - \alpha_{gold}\right)(x_{1t} - x_{2t}) + (BR_t - \alpha_{bitcoin})(x_{3t} - x_{4t}) \tag{22}$$

$$st.\begin{cases} x_{1t} \cdot x_{2t} = 0 \\ x_{3t} \cdot x_{4t} = 0 \\ x_{1t} \cdot GP_t \cdot 100 + x_{3t} \cdot BV_t \cdot 0.01 \le C_{t-1} \\ x_{2t} \cdot 100 \le G_{t-1} \\ x_{4t} \cdot 0.01 \le B_{t-1} \\ (x_{1t}, x_{2t}, x_{3t}, x_{4t})^T \in \mathbb{Z}^4 \end{cases} \tag{23}$$

Since gold can be only traded on weekdays, in the subsequent process, the algorithm sets x_{1t}, x_{2t} equal to zero at weekend, that is, the trading volume of gold is equal to 0 at weekends.

Considering the risk of losing money, if the returns of buying gold or Bitcoins cannot cover the costs, then traders would not buy any gold or Bitcoin.

5.4 Model Results

Assume that traders initially have $1,000, and make daily decisions for 5 years, we calculate the total assets, using three prediction models and daily decision model. The final assets of the three models are 238006 $,268006 $, and 243494 $ respectively. The second model have the highest value of asset which means the combination of Unbiased Metabolic GM (1,1) and ARIMA model has the best return.

6. Model Evaluation

This paper considers problems from the perspective of traders in the real life. Models and algorithms embody the logical thinking process of ordinary people when making investment decisions, and abstract them into models and code languages for implementation and optimization.

When making price forecasts, this paper is not limited to a single forecasting model, but fully considers the amount of data, and uses the ARIMA model and the Unbiased Metabolic GM(1,1) model to jointly carry out price prediction. ARIMA model is for long-term prediction and improved GM(1,1) is for short-term prediction. Therefore, the prediction results are more accurate.

The assumptions of the model are closer to actual market conditions, and only a reasonable simplification of reality.The results of the price prediction models are extremely accurate.

References

[1] Kumari Anita, Goyal Ruchi, Kumar Sunil. Review on Behavioral Factors and Individual Investors Psychology Towards Investment Decision Making[J]. Electrochemical Society Transactions,2022,107(1).

[2] Chandra Atul, Hartley Peter R., Nair Gopalan. Multiple Volatility Real Options Approach to Investment Decisions Under Uncertainty[J]. Decision Analysis,2022.

[3] Faleiros Daniel Resende,Nunes da Silva Everton,Santos Andreia C,Godman Brian B,Goncalves Pereira Ramon,Guerra Junior Augusto A. Adoption of new therapies in the treatment of hepatitis: a verification of the accuracy of Budget Impact Analysis to guide investment decisions.[J]. Expert review of pharmacoeconomics & outcomes research,2022.

[4] Gambetti Elisa, Zucchelli Micaela Maria, Nori Raffaella, Giusberti Fiorella. Default rules in investment decision-making: trait anxiety and decision-making styles[J]. Financial Innovation,2022,8(1).

[5] Sun Haoyu, Zheng Huiqi,Sun Xiaoyang,Li Wei. Customized Investment Decisions for New and Remanufactured Products Supply Chain Based on 3D Printing Technology[J]. Sustainability,2022,14(5).

[6] Wu Wenqing, Huang Xuan,Wu Chia-Huei,Tsai Sang-Bing. Pricing strategy and performance investment decisions in competitive crowdfunding markets[J]. Journal of Business Research,2022,140.

[7] Yangluyuan, lizhangping Forecast and analysis of general aviation operation market demand in China -- Based on GM (1,1) model and life cycle identification model [j] Comprehensive transportation, 2022,44 (04): 61-66

[8] Gu Huxia, Hou Fang, Gao Qingtian, guoshuwei, Shiyang, Zhang Chong Application of GM (1,1) model in the prediction of incidence rate of dysentery in China [j] Medical animal control, 2022,38 (05): 419-422

[9] Zhai min, Chang Guoxia Deformation prediction practice and analysis of GM (1,1) residual error processing model [j] Shanxi architecture, 2022,48 (07): 159-160+172 DOI:10.13719/j.cnki. 1009-6825.2022.07.044.

[10] Pan Hui Application of weighted dynamic GM (1,1) model in settlement prediction of high-speed railway [j] Standardization of Surveying and mapping, 2022,38 (01): 54-58 DOI:10.20007/j.cnki. 61-1275/P.2022.01.12.