Gold Price Forecast by Different Models

Wenjing Fang*

School of Mathematical and Statistical Sciences, Ludong University, Shandong, China

*Corresponding author: fangxinzhang@sepco3.com

Abstract. In recent years, the gold investment market has a dramatic effect on the world. So being able to predict gold price accurately has become an urgent demand for economic development. This paper gives three different models including the regression analysis model, back propagation neural network (BP neural network) model, and time series analysis model to realize the prediction of gold price. Firstly, this paper verifies that the data applies to each model respectively. Secondly, through training the models, the corresponding gold price prediction formulas are output. Finally, through the test of the models, it can be concluded that the three models all have high reliability and accuracy, which means that they are suitable for the prediction of gold prices. In the end, this paper compares the prediction effects of the three models, and finds that compared with the regression analysis model, the BP neural network model and the time series analysis model have certain errors and limitations for long-term prediction, while they are more suitable for the short-term forecast. Since there are certain risks in gold investment, grasping the trend and fluctuation of the gold price accurately is the concern of both all gold investors and academic circles.

Keywords: Gold price, Regression analysis model, BP neural network model, Time series analysis model

1. Introduction

In recent years, gold, as an investment tool for a long time, is widely favored by different investors in the trading markets. Investing in gold can help investors avoid possible problems in the economic environment, such as being eroded by inflation, while avoiding the potential collapse risk of stock investment and real estate investment. With the continuous improvement of social policies and the continuous development of science, the role of gold in investment and preservation of value is becoming more and more obvious [1]. Since the early 1970s, the price of gold in the world has changed dramatically. Especially in the past 30 years, the price of gold has experienced great changes. Gold is a special precious metal that has dual functions of money and commodities. So, there are more factors affecting the price trend of gold than ordinary commodities, including the US dollar index, commodity prices, Brent crude oil prices, the exchange rate of the euro against the US dollar, the Dow Jones index, and so on.

As an important choice to avoid risks, gold has attracted more and more attention from all walks of life. In addition, there are much more scholars analyze and predict the price of gold futures. For example, Grudnitski and Osburn applied neural networks to forecast Standard & Poor's index and gold futures prices in 2006 [2]. Meanwhile, Ismail, Yahya and Shabri used a multiple linear regression model to predict gold prices by introducing indicators such as bulk commodities, exchange rate and inflation rate in 2009 [3]. In 2014, Sharma and Baby predicted and analyzed the gold price through Autoregressive Integrated Moving Average model (ARIMA), which is based on the data from February 1990 to September 2014 [4].

This paper summarizes 3 different prediction models of gold price, including regression analysis model, back propagation neural network model and time series analysis model. In the regression analysis model, the partial least square method is introduced to avoid the multicollinearity problem prevalent in multiple linear regression and guarantee the comprehensiveness of gold price analysis. In the BP neural network model, based on principal component analysis, a nonlinear prediction model of gold price is established. After training, by looking at the value of R^2, it can be known that the model has very high reliability and accuracy especially when making short-term predictions. In the

time series analysis model, after analysing the main characteristics of the data, the autoregressive moving average model (ARMA) is selected to solve the problem. By looking at

the value of R^2, it can be found that the model has better prediction effect.

Through the reasonable prediction of gold price, it can better seek advantages and avoid disadvantages as well as help participants to choose the more favourable way to conduct gold transactions in the gold market.

2. Main Body

2.1 Regression Analysis Model

2.1.1. Variable Selection

Through analysis, it is found that there will be multicollinearity when doing ordinary multiple regression. In order to avoid the decline in the comprehensiveness of gold price analysis caused by eliminating variables, this paper chooses to use the partial least squares method. This method can be used for regression modeling under the condition of serious multiple correlations of independent variables. All the original independent variables will be included in the final model which can fully explain and analyze the gold price.

Firstly, the international gold market is mainly priced in US dollars, and the trend of US dollars has a high reverse linkage relationship with gold prices. Secondly, as the prices of gold and oil are affected by some common factors, there is a very close correlation between them. Thirdly, the euro is the second major international currency besides the US dollar. It not only has a wide variety of applications, but also has some impact on gold price. And finally, in the long run, the price change of the stock market will affect the change in gold price to a certain extent.

Therefore, this paper selects four factors related to the price of gold including the US dollar index, Brent crude oil price, the exchange rate of the European dollar against the US dollar and the Dow Jones index.

2.1.2. Model Building and Problem Solving

This paper uses the weekly data of gold spot (USD/oz), US Dollar index, Brent crude oil price (USD/barrel), euro to dollar (EUR/USD) exchange rate and Dow Jones index, from July 11, 2008 to May 25, 2012.

2.1.2.1. Correlation Test of Variables

First, let's carry out the correlation analysis of variables in data and get Table 1.

It can be seen from the Table1. that each independent variable is linearly related to the gold price, which means that multiple linear regression can be used for predictive purposes. However, there is also a certain autocorrelation between the four self-compiled quantities at the same time, which shows that there will be multicollinearity if using a simple multivariate statistical regression method. Therefore, this paper uses the partial least squares method to solve this problem.

2.1.2.2. Partial Least Square Method

To begin with, let's find the linear functions of independent variables which have strong correlation

Table 1. Correlation between variables [5]

	Spot gold / USD	**USD index**	**Brent crude oil**	**EUR / USD**	**DJIA**
Spot gold / USD Pearson correlation	1	-.425**	.585**	-.092**	.786**
Significance test (Bilateral)		.000	.000	.192	.000
N	203	203	203	203	203
USD index Pearson correlation	-.425**	1	-.508**	-.840**	-.600**
Significance test (Bilateral)	.000		.000	.000	.000
N	203	203	203	203	203
Brent crude oil Pearson correlation	.585**	-.508**	1	.158*	.819**
Significance test (Bilateral)	.000	.000		.025	.000
N	203	203	203	203	203
EUR / USD Pearson correlation	-.092	-.840**	.158*	1	.146*
Significance test (Bilateral)	.192	.000	.025		0.37
N	203	203	203	203	203
DJIA Pearson correlation	786**	-.600**	.819**	.146*	1
Significance test (Bilateral)	.000	.000	.000	.037	
N	203	203	203	203	203

**. Significant correlation at .01 level (Bilateral).

*. Significant correlation at .01 level (Bilateral).

with dependent variables and can easily calculate the linear function of independent variables. Its algorithm is the least square, which only selects the variables related to the dependent variable and the dependent variable, and only considers the part biased to the dependent variable. The method is as follows [5]:

Firstly, we need to center the dependent variable matrix of gold price and the self-variable matrix X of each independent variable into y and x. It is known in multiple linear regression that:

$$\hat{y}(x_i) = \frac{x_i' y}{x_i' x_i} x_i, X_i = \begin{pmatrix} x_{1i} \\ ... \\ x_{ni} \end{pmatrix}, i = 1, 2, ..., k \tag{1}$$

It is the predicted value of the centralized independent variable. So later we can make that:

$$t_1 = \sum_{i=1}^{k} (x_i' y) x_i \tag{2}$$

Secondly, we take t_1 as an independent variable and regress y:

$$\hat{y}(t_i) = \frac{t_1' y}{t_1' t_1} t_1 \tag{3}$$

After that, we can get the residual:

$$y^{(1)} = y - \hat{y}(t_i) \tag{4}$$

Because this formula doesn't contain the value of t_1, so we choose to eliminate it, which means that to regress each independent variable x_i to t_1:

$$\hat{x}_i(t_i) = \frac{t_1' x_i}{t_1' t_1} t_1, i = 1, 2, ..., k \tag{5}$$

Thirdly, based on the formula, we can get the corresponding residual of x_i is:

$$x_i^{(1)} = x_i - \hat{x}_i(t_i), i = 1, 2, ..., k \tag{6}$$

Then let's repeat the above steps with $y^{(1)}, x_1^{(1)}, ..., x_k^{(1)}$ as the new original variable, and gradually obtain $t_1, t_2, ..., t_r$, where r is the rank of $X'X$.

Finally, we use y to regress $t_1, t_2, ..., t_r$ by ordinary least squares. After the conversion of variable keys, we can obtain the regression equation of y to $t_1, t_2, ..., t_k$.

With the help of the software, we can get the gold price prediction equation of partial least squares regression [5]:

$$Y = -1480.711 + 10.1X_1 - 2.33X_2 - 0.972X_3 + 0.2X_4 \tag{7}$$

2.2 BP Neural Network Model

With the continuous expansion of the gold market, people tend to study more and more price fluctuations in the market. Many people use BP neural network to develop a nonlinear model for predicting the price of gold. Experimental results show that the network has good predictive accuracy, while there is room to optimize and improve.

This paper uses BP neural network to predict the London gold price of LIFFE London International Financial Futures Exchange. Based on principal component analysis, this paper assumes that there is a mapping relationship among gold price, US dollar index and crude oil price. The nonlinear mapping relationship is established to realize the optimal approximation of F value.

2.2.1. Working Principle of BP Neural Network

In 1986, Rumelhart and Celland formally proposed BP neural network [6]. It is one of the most widely used neural network models, which is mainly composed of forward propagation of input signal and back propagation of error signal. Forward propagation passes through the input layer input and through the hidden layer output after non-linear mapping. The back propagation of error signal is to compare the training value with the output result of the model, feed back the error layer by layer, and constantly adjust the threshold and weight of each layer.

As shown in Figure 1 is the picture of BP neural network topology. The specific principle of the model is described as follows [7]:

First of all, let's make a training input vector：

$$X^p = (x_1^p, ..., x_k^p, ..., x_m^p), (p = 1, 2, 3, ..., q) \tag{8}$$

We can get the expected output vector is [7]:

$$T^p = (t_1^p, ..., t_k^p, ..., t_n^p), (p = 1, 2, ..., q); \tag{9}$$

as well as the actual output vector is:

$$Y^p = (y_1^p, ..., y_k^p, ..., y_n^p), (p = 1, 2, ..., q) \tag{10}$$

After that, we need to initialize ownership and threshold to the minimum value.

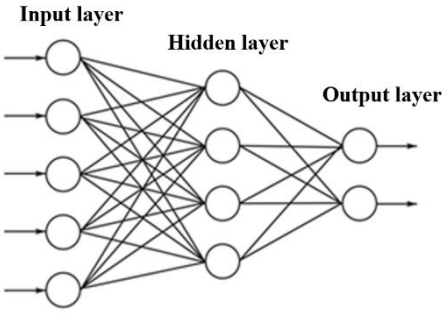

Figure 1. BP neural network topology.

After vector X enters the input layer, we can get the activation value of neurons in each hidden layer and output layer is:

$$S_j^p = \sum_{i=1}^{m} w_{ij}^1 x_i^p - \theta_j^1, (i = 1, 2, ..., n; j = 1, 2, ..., m; p = 1, 2, ..., q) \tag{11}$$

In the above formula, w is the weight and θ is the threshold. Then after sorting, we get the output values of node k are:

$$y = f(S_k^p) = f(\sum_{i=1}^{l} w_{jk}^2 h_k^p - \theta_k^2) \tag{12}$$

In the back propagation process, the correction error of the output layer as well as the weight and threshold of output layer correction are:

$$\begin{cases} \delta_k^p = (t_k^p - y_k^p) y_k^p (1 - y_k^p) \\ \Delta w_{jk}^2 = \eta \cdot \delta_k^p \cdot h_j^p \\ \Delta \theta_k^1 = \eta \cdot \delta_k^p \end{cases} \tag{13}$$

Finally, we need to calculate the total error as the process in the Figure 2.

2.2.2. Model Establishment and Result Analysis

2.2.2.1. Data Pre-processing

This paper uses the data of the dollar index futures of IMM international money market, the crude oil price of NYMEX New York Mercantile Exchange and the London gold price of LIFFE London International Financial Futures Exchange for analysis and modelling.

Data preprocessing will directly affect the performance of trained BP neural network. Therefore, before neural network prediction, in order to eliminate order-of-magnitude differences between data across dimensions, this paper normalizes the original data. Expand all variables by 10^4 times to facilitate later observation and calculation.

Meanwhile, in order to avoid the insufficient training ability or over fitting of the model caused by

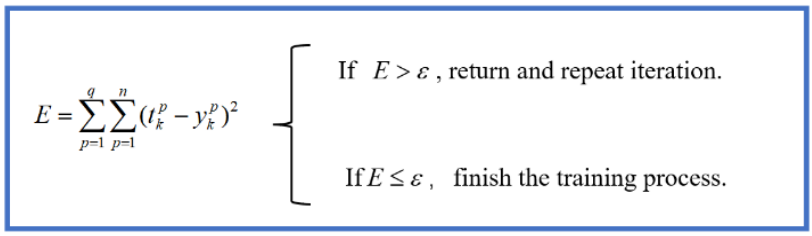

Figure 2. Total error calculation.

too many or too few nodes, this paper chooses to use the final nodes of 7 and 13.

2.2.2.2. Stationary Test

In order to eliminate the phenomenon of pseudo regression, the stationary line test was carried out. First, through the unit root test, we can get the result that the series is non-stationary.

Then, we performed autocorrelation and partial autocorrelation tests, and we can get that this is a (partial) autocorrelation coefficient tailing. So this sample is suitable for BP network analysis.

In the long memory test, this paper uses the HURST Index to test. The test result table shows that the H index of this series is 0.915. It shows that the gold market has a very obvious long memory and it has a partial random walk market as well as the characteristics of chaotic separation, which is very suitable for using BP neural network for analysis and prediction.

2.2.2.3. Construction of The Network Model

In the design of BP neural network topology, we can obtain a multi-layer neuron network from a single-layer network structure with S neurons and R inputs [7]:

$$a^3 = f^3(W^3 f^2(W^2 f^1(W^1 p + b^1) + b^2) + b^3) \tag{14}$$

Under normal circumstances, the model often adopts a three-layer network with a single hidden layer. In addition, the method of determination of the number of hidden layer nodes in the model adopts the experimental method. For the problem of time series prediction, the sequence is firstly divided into a training part and a test part. Then the final optimal value is obtained by continuously debugging the number of input nodes and testing its accuracy.

By training the BP network and continuously adjusting the initial weights of the network as well as the topology of the network, we obtained the final project establishment effect and output the training result indicators in Table 2.

After comparison, we select the third training result with the best effect. From Table 2. , it can be found that the model can explain 98.731% of the observations. It can be seen that the BP neural network has a high reliability after a sufficient amount of training. Then we

Table 2. Indicators of drilling results [7]

	MSE	NMSE	R
1	0.0131729	0.0626835	0.9682226
2	0.0086619	0.0412178	0.9792335
3	0.0053104	0.0252697	0.9873100
4	0.0077414	0.0368380	0.9814474
5	0.0897220	0.0426945	0.9785380
6	0.0780107	0.0371215	0.9813160

use this weight for testing. The test results are shown in Figure 3, and the error term scatter is shown in Figure 4.

As can be seen from the Table 2. that the model has a high accuracy. Although subsequent forecasts have varied numerically over time, according to the forecast error term, nearly 80% of forecasts are within 2 standard deviations. In summary, the model has high accuracy and is very suitable for short-term forecasting, but there will be a certain deviation in predicting long-term trends.

2.3 Time Series Analysis Model

The price of gold is affected by many factors and the generation process is extremely complex. It is therefore of major practical importance to study the dynamic evolution of gold price. Combining the time series correlation theory, this paper establishes the ARMA model of gold price prediction, and finds that it can dynamically describe the generation process of gold price data and better predict the gold price.

2.3.1 Concepts Related to Time Series

Time series analysis is a method that uses mathematical statistics to preprocess a series of data sequences generated over time [8]. Additionally, it can predict future development. It consists primarily of three core models, including Autoregressive Model (AR), Medium Mobile Model (MA) and Medium Mobile Autoregressive Model (ARMA).

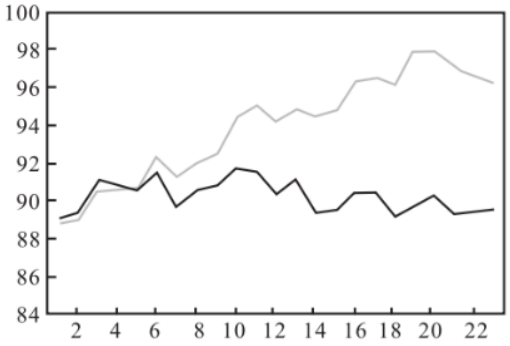

Figure 3. Line chart of predicated value and actual value [7].

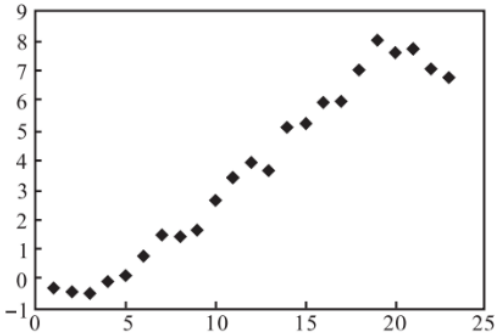

Figure 4. Error terms of predicted value [7].

The autoregressive moving average model used in

this paper is a two-part random process comprising autoregression and moving average [9]. It is denoted as $ARMA(p,q)$, whose definition is [10]:

$$\begin{cases} x_t = \phi_0 + \phi_1 x_{t-1} + \phi_2 x_{t-2} + ... + \phi_p x_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - ... - \theta_p \varepsilon_{t-p} \\ \phi_p \neq 0, \theta_q \neq 0 \\ E(\varepsilon_t) = 0, Var(\varepsilon_t) = \delta_\varepsilon^2 . E(\varepsilon_t \varepsilon_s) = 0, s \neq t \\ E(x_s \varepsilon_t) = 0, \forall s < t \end{cases} \tag{15}$$

Leaving out the default conditions, the centralized model can be written as [10]:

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \ldots + \phi_p x_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \ldots - \theta_p \varepsilon_{t-p}. \quad (16)$$

After introducing the delay operator, $ARMA(p,q)$ can be abbreviated as [10]:

$$\Phi(B) x_t = W(B) \varepsilon_t \quad (17)$$

The invertible condition of the ARMA process is that the roots of the polynomial $W(B) = 0$ are all outside the unit circle.

2.3.2 Model Establishment and Solution

Under the assumption that the national policy stable gold price is not related to political factors during the forecast period, the model framework we established is shown in Figure 5.

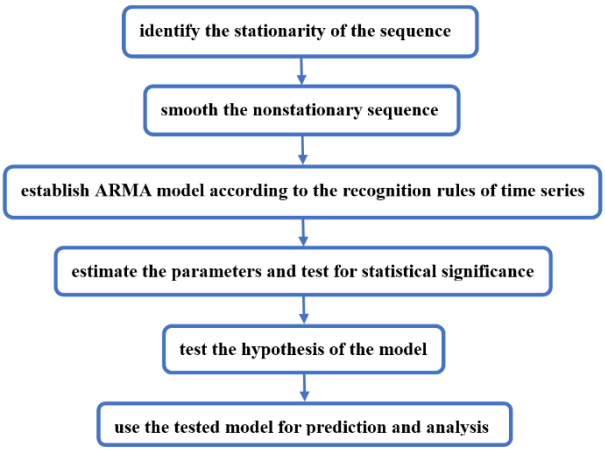

Figure 5. Model framework.

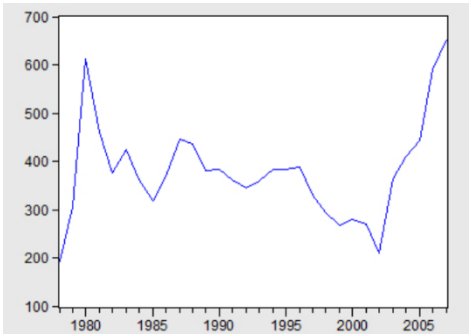

Figure 6. Time series trend chart [10].

Firstly, we draw the trend diagram of the sequence in Figure 6. We can find that the time series of gold prices is a stationary time series, which means that we can establish an ARMA model of the annual average value of gold prices. Then we try the model from high to low order according to the autocorrelation and partial correlation plots and then we test the model. From this, we can judge that the annual average price of gold in the world is a fourth-order autoregressive model, which means that we can confirm that the $ARMA$ model is in the form of $AR(4)$.

Secondly, we use the least squares to estimate the model, and the specific form of the model is [10]:

$$X = 397.57 + 1.230 X_{t-1} - 0.576 X_{t-2} - 0.439 X_{t-3} + u_t, R^2 = 0.75 \quad (18)$$

Therefore, the model can explain 75% of the observations, which shows that the effect is good. Finally, we perform a white noise test on the random error term. The Q values obtained by the test are all smaller than the critical value of the χ^2 distribution with the test level of 0.05, so it can be considered that the random error term is a white noise sequence. Therefore, we accept using the model to predict gold price.

3. Conclusions

This paper has presented three different methods for gold price forecasting, which are regression analysis model, BP neural network model and time series analysis model. After getting the corresponding prediction model from the training data, it can be found that the prices predicted by the models all have high degrees of fitting with the actual prices, which means that the models all have high reliability and prediction accuracy. Furthermore, by comparing the prediction effects of the three models, it can be summarized that compared in comparison to the regression analysis model, the BP neural network model and the time series analysis model are more suitable for short-term prediction.

As a special commodity with financial attributes, gold's price changes are closely related to the financial situation at home and abroad, as well as the value behavior of investors and producers. The ability to predict and analyze the price of gold with the help of computer and mathematical models can not only help a country's macro-control but also help investors better understand the gold market.

The main influencing variables of gold price from the short-term, medium-term and long-term perspectives are different. In addition, the factors affecting the price of gold are complex and changeable, so it is necessary to take into account the effects of various variables and the degree of influence of various variables in different periods. Therefore, the flexibility of the model can be strengthened by enhancing the consolidation correction ability of the model parameters, so that the model can correct the parameters of the new model according to the changes in the external economic environment to better fit the trend of gold prices.

References

[1] W. Ye, L. Sun, L. Qi, Research on dynamic cointegration of gold and bitcoin -- Based on semi parametric Midas quantile regression model, in: Systems Science and Mathematics, vol. 40, 2020, pp. 1270-1285, DOI: CNKI: SUN: STYS.0.2020-07-009

[2] G. Grudnitski, L. Osbum, Forecast S&P and gold futures prices: An application of neural networks, in: Journal of Futures Markets, vol. 13, 2006, pp. 631-643. DOI: 10.1002/fut.3990130605

[3] Z. Ismail, A. Yahya, A. Shabri, Forecasting gold price using multiple linear regression method, in: American Journal of Applied Sciences, vol. 6, 2009, pp. 1509-1514. DOI: 10.3844/ajassp.2009.1509.1514

[4] A. Sharma, S. Baby, Forecasting gold prices using time series analysis, in: Second International Conference on Business Analytic & Intelligence, 2014.

[5] Y. Li, Gold price prediction based on partial least squares regression, in: Goods and Quality, 2012, pp. 108-110. DOI: CNKI: SUN: SPZN.2012-S7-085

[6] H. Qiao, Application of BP and RBF network model in gold price prediction, in: Jinan University, 2016. DOI: CNKI: CDMD:2.1016.733351

[7] C. Ling, H. Zhang, Prediction and analysis of gold price based on BP neural network, in: Tianjin Technology, vol. 41, 2014, pp. 68-73. DOI: 10.14099/j.cnki.tjkj.2014.01.013

[8] X. Lin, Application of time series analysis in building deformation monitoring. In: Jilin University, 2005.

[9] B. Ma, B. Zhen, Empirical analysis of gold futures price prediction model based on time series analysis, in: Business, 2015, pp. 152. DOI: CNKI: SUN: SHNG.0.2015-07-125

[10] J. Xi, Gold price forecast based on time series model, in: Modern Computer (Pro), 2014, pp. 9-14. DOI: CNKI: SUN: XDJS.0.2014-28-002