

Build Prediction Models for Gold Prices Based on Back-Propagation Neural Network

Chingpei Lin

Department of Information Management
Hwa Hsia Institute of Technology
Taipei, Taiwan R.O.C

Abstract—In recent years, international gold prices have been constantly rising, gold investment and preserve (or even appreciation) effects have been widely concerned by the market. Whether it is based on speculation, investment or hedging purposes, the gold has been incorporated into the asset allocation by many investors, which has become another important investment in addition to foreign currency, funds, stocks and securities. Therefore, this paper discusses how to construct a prediction model for gold prices to understand the future gold price trend, and to provide a reference for experts and investors. Firstly, we collect historical data of gold prices from web database, and draw a tendency chart to observe the trend of gold prices; then we use technical index formula of share price to calculate the five technical index values of gold as an independent variable and the price of gold the next day as a dependent variable, and build three prediction models including back-propagation neural network (BPN), Principal Component Regression (PCR) and Multiple Regression (MR). The study indicates that BPN's predictive ability is better than other models.

Keywords—gold price; back propagation neural network (BPN); Principal Component Regression (PCR); Multiple Regression (MR); technical index

I. INTRODUCTION

Gold has been a medium of exchange for thousands of years. It is widely used as precious metal and currency and precious jewelry to meet people's demand of decoration, storage and medium of exchange. Gold has been loved and trusted by people since ancient times due to its superior liquidity characteristics. In the eyes of most people, gold is regarded as the incarnation of wealth and power; hence there are a lot of related articles on the changes of gold prices (Wan-hsiu Cheng, 2010; Yuan-Ming Lee, 2008; Ronald, 2006). As gold is a scarce resource, gold is widely considered an inflationary hedge especially amid the economic downturn and political unrest, which could better highlight its value. As a small amount of investment would bring high returns, it is very suitable for investors who pursue long-term capital appreciation. However, gold prices have been unstable in the current domestic and foreign economic situation, and have experienced extreme price fluctuations, making it is very difficult for investors to grasp the trend of gold prices. Therefore, the paper analyses three prediction models including back-propagation neural network (BPN), Principal Component Regression (PCR) and Multiple Regression (MR) based on the historical data of gold prices.

Firstly, it uses technical index formula of share price to calculate the technical index values of gold as independent variables and the closing price of gold as dependent variables, and builds three prediction models including BPN, PCR and MR. It also compares the predictive ability of these three models to provide reference for scholars and investors. The first section introduces the purpose of this study. The second part describes BPN and PCR and relevant literature. The third part is about empirical study and the fourth part presents conclusions.

II. RESEARCH METHODS

As a multiple regression is a simple and common statistical model, readers can refer to relevant literature. This part only introduces BPN and PCR.

A. Back-Propagation Neural Network (BPN)

The structure of BPN is shown below, including Input Layer, Hidden Layer and Output Layer.

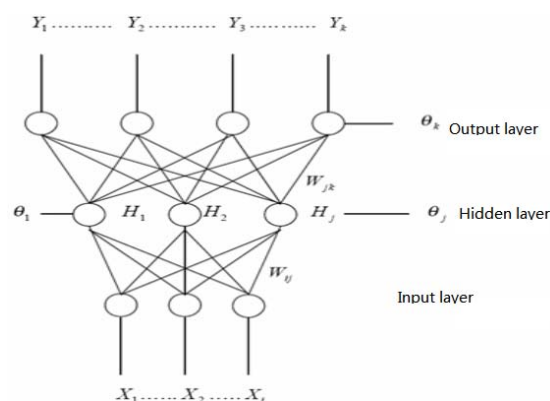


FIGURE I. STRUCTURE OF BPN

Each layer of BPN is composed of a number of neurons, the neurons among different layers are associated with each other with weighed value, but the neurons of the same layer are not connected with each other. Signals are spread from the previous layer to the latter layer, but when correcting the distribution error of the weighted value, the signals are spread oppositely from the latter layer to the previous layer. Currently neural network has been widely used in various fields of BPN (Freisleben, 1992; Chen, 2003; Refenes, 1994).

Principal Components Analysis (PCA) is an exploratory multivariate statistical technique for simplifying complex data sets. PCA is often used to reduce the dimensionality of the data set, while maintaining the feature of data concentration. This is to transform a high-dimensional dataset into a smaller-dimensional subspace. Such low-dimensional components are often able to keep the most important aspects of data. However, this is not certain, depending on the specific application. Because the principal component analysis is dependent on the data provided, the accuracy of the data has a great influence on the analysis result. The principal component regression is to reduce the variation of data sets and to reduce the complexity of the general multiple regression, and thus maintain the accuracy of prediction.

III. EMPIRICAL STUDY

A. Sample Data and Variables

This paper collects historical data of gold trading in Taiwan from domestic websites, from January 5, 2010 to December 13, 2012, about 900 transaction data were collected, and the trend chart is shown in Figure 2:

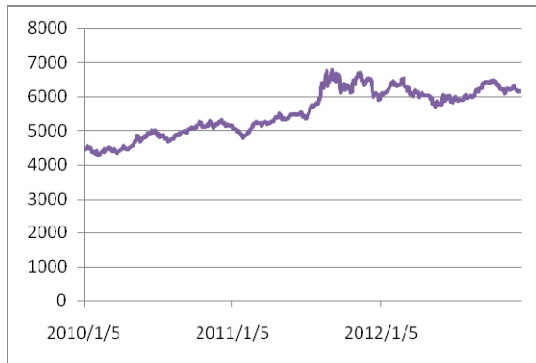


FIGURE II. TREND CHART OF CLOSING GOLD PRICES

Hence, we define the formula of technical index of five stocks as follows:

Since gold is a hedging tool in sluggish economic situation and also an investment and financial management tool amid good international economic conditions, therefore, it is seen from the Figure that gold prices increased significantly during this period. It is observed that it is difficult for ordinary investors and experts to grasp the gold price trend, which also shows the importance of the prediction model of gold prices. If one can predict accurately the price changes of gold, he/she will take profits in the gold investment increase his/her wealth. For this reason, the next section of this article will establish three predictions models for the gold closing prices. This paper will input the gold price to stock technical index formula, and then calculate the technical index of gold price as an independent variable (X), and take the closing price of oil and gold as a dependent variable (Y), and combine the both as sample information. We define the formula of technical index of five stocks as follows:

5-day MA = sum of 5-days of closing gold price total price / 5 days

5-day W%R = $100 - (\text{peak price of gold within 5 days} - \text{closing gold price of the day}) / (\text{peak price of gold within 5 days} - \text{bottom price of gold within 5 days}) \times 100$

5-day BIAS = $(\text{closing gold price of the day} - \text{average closing gold price of 5 days}) / \text{average closing gold price of 5 days} \times 100$

K value = $2 / 3 \times \text{the previous day's date, K value of } + 1 / 3 \times \text{RSV}$

Value of D = $2 / 3 \times \text{the previous day's K-value of } + 1 / 3 \times \text{date K}$

Wherein, 5-day RSV = $(\text{closing gold price of the day} - \text{bottom gold price of 5 days}) / (\text{peak price of gold within 5 days} - \text{bottom price of 5 days}) \times 100$

After calculating the technical index of gold prices, the paper collects 900 sample data to build prediction models for gold prices. Table 1 shows the technical index of gold prices.

The paper divides 900 sample data into five groups. Group 1 takes No.1 to No. 160 data as Training Data, and No. 161 to No. 180 data as Testing Data; Group 2 takes No.181 to No. 340 data as Training Data, and No. 341 to No. 360 data as Testing Data; Group 3 takes No. 361 to No. 520 data as Training Data, and No. 521 to No. 540 data as Testing Data; Group 4 takes No.541 to No. 700 data as Training Data, and No. 701 to No. 720 data as Testing Data; Group 5 takes No.721 to No. 880 data as Training Data, and No. 881 to No. 900 data as Testing Data; each group has 160 training data and 20 testing data, hence, 100 data is used for providing predictive ability analysis for all models.

TABLE I. NARRATIVE STATISTICS OF TECHNICAL INDEX OF OIL AND GOLD PRICES

Variable	X1	X2	X3	X4	X5
MAX	25.131	1.000	6.689	0.367	0.365
MIN	-99.853	0.000	4.114	0.033	0.046
AVG	0.069	0.447	5.569	0.218	0.218
STD	3.821	0.415	0.683	0.138	0.087

B. Building Three Prediction Models for Gold Closing Prices

The sample data of five groups are included into BPN, and the framework is shown in Figure 3.

Input layer has five neurons, representing five independent variables respectively; output layer has one neuron, representing predicted value; and interlayer has three neurons, which is calculated based on the suggestion of Yeh, Y.C. (2001), namely, take the sum of number neurons of input layer and output layer and divide it by 2. After completing the framework of network, we put the test data into the network to carry out model prediction. As we have 5 groups of data, the work shall be done five times, and the test results of 100 data will be obtained.

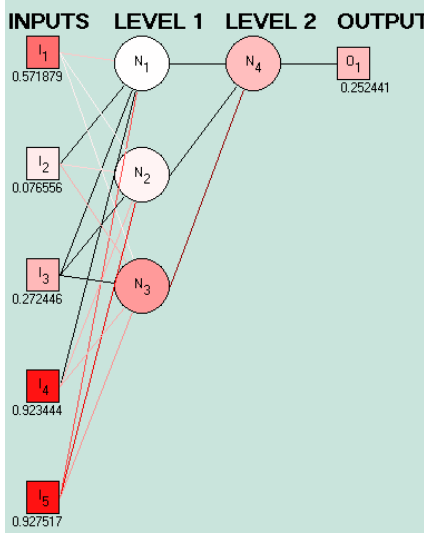


FIGURE III. FRAMEWORK OF BPN

In terms of PCR, the independent variables must be reduced with the reduction technique which is based on component with eigenvalue greater than 1. Figure 4 shows that PCA only takes 3 components as independent variables, and set up a principal component regression model. In addition, the paper also builds a multiple regression model.

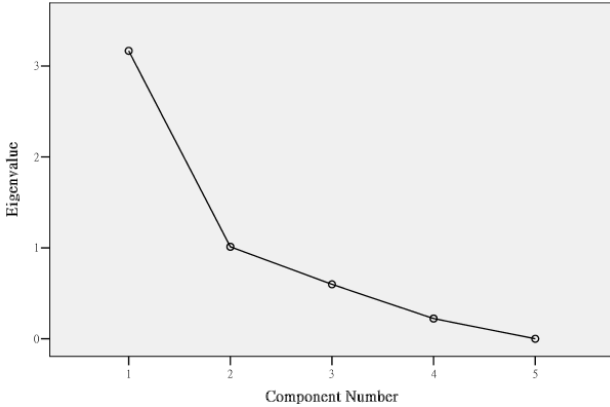


FIGURE IV. COMPONENT WITH EIGENVALUE GREATER THAN 1

C. Comprehensive Comparison of Predictive Ability of 3 Models

We input test data of 5 groups into the prediction model of gold prices to test the stability of these prediction models. In order to clearly show the predictive accuracy of these 3 models, we adopt 3 evaluation indicators to compare the predictive ability of 3 models, which are:

Root Mean Squared Error (RMSE) is formulated as:

$$RMSE = \sqrt{\frac{\sum_{i=1}^n x_i - \hat{x}_i}{N}}$$

Mean Absolute Error (MAE) is formulated as:

$$MAE = \frac{1}{M} \sum_{l=1}^M |Z_{t+1} - \hat{Z}_t(l)|$$

M is number of predictive value, Z_{t+1} is observed value and $\hat{Z}_t(l)$ is estimate value.

Mean Absolute Percentage Error (MAPE) is formulated as:

$$MAPE = \left(\frac{1}{M} \sum_{l=1}^M \left| \frac{Z_{t+1} - \hat{Z}_t(l)}{Z_{t+1}} \right| \right) \times 100\%$$

The closer the 3 evaluation indicators get to 0, the higher accuracy the models are. The result is shown in table 2:

TABLE II. EVALUATION RESULT OF 3 EVALUATION INDICATORS

Model	RMSE	MAE	MAPE
BPN	84.439	69.041	0.011
PCR	105.064	85.230	0.015
MR	122.211	93.173	0.016

We can see from Table 2 that of three evaluation indicators, BPN's RMSE is 84.439, MAE is 69.041 and MAPE is 0.011, which are lower than those of PCR and MR. Therefore, BPN has better predictive ability compared with other models.

IV. CONCLUSION AND SUGGESTION

The paper aims to provide reference for researchers by analyzing the trend of gold price and build prediction models. There were many factor affecting the price of gold, which made the closing gold price has high randomness. Therefore, we build prediction models for gold price for the purpose of accuracy. It is found that the closing gold price shows high randomness with the trend chart. The result shows that BPN has better predictive ability compared with other models.

Furthermore, paper builds models based on Back-Propagation Neural Network. It is recommended that other neural networks may be built (such as General regression neural network proposed by Professor Specht (1990, 1991) or Support Vector Machine proposed by Professor Vapnik (1995)) for comparison.

REFERENCES

- [1] Yeh, Y.C.(2001), The model application and practice of artificial neural network, Scholars Books Co., Ltd.
- [2] Wan-hsiu Cheng, Wu Yahui (2010), The Volatility Spillover Effects among Oil, Gold and US Dollar Index Futures, Journal of Risk Management, Vol. 12, No. 2, Journal of Risk Management, Vol. 12, No. 2, pp. 11-233.
- [3] Yuan-Ming Lee, Kuan-Min Wang (2008), The Dynamic Relationships between Gold Return and U.S. Dollar Depreciation, Journal of Risk Management, Vol. 10, No. 1.
- [4] Ronald and Kettering (2006), The Changung Relationships between Gold, Oil, and Stock Prices, Allied Academies International Conference, 11, No. 1, pp.37-40.
- [5] Freisleben B.(1992), Stock market prediction with backpropagation networks, Industrial and Engineering Application of Artificial Intelligence and Expert System. 5th International Conference, Paderborn, Germany, pp.451-460.

- [6] Chen, A. S. and Daouk, H. (2003), Application of Neural Networks to an Emerging Financial Market: Forecasting and Trading the Taiwan Stock Index, *Computers and Operations Research*, Vol. 30, No.6, pp. 901-923.
- [7] Refenes, A. N., Zapranis A. and Francis G.(1994), Stock performance modeling using neural networks: a comparative study with regression models, *Neural Network*, Vol. 5, pp.961-970.
- [8] Specht, D.F. (1990) 'Probabilistic neural networks and the polynomial adaline as complementary techniques for classification', *IEEE Trans. on Neural Networks*, Vol. 1, No. 1, pp.111–121.
- [9] Specht, D.F. (1991) 'A general regression neural network', *IEEE Tras. Neural Networks*, Vol. 2, No. 6, pp.568–576.
- [10] Vapnik, V. (1995), Support-vector networks. *Machine Learning*, Vol. 20, No. 3, pp. 273-297.