Gold Price Forecasting Based on Projection Pursuit and Neural Network

Li Chen¹, Xiaofei Zhang²

¹School of mathematics and statistics, Zhengzhou Normal University, Zhengzhou, Henan, 450044
²School of information science and technology, Zhengzhou Normal University, Zhengzhou, Henan, 450044)

*Corresponding author’s e-mail: cluu6697@sina.com

Abstract. Gold price has significant nonlinear, time-varying, many influence factors and more difficult to determine. In order to improve the precision of forecast gold price, this paper proposes a gold price combination forecasting model based on pursuit algorithm and neural network. Firstly, projection pursuit algorithm is used to choose the influence factors, and then the selected impact factors are used as BP neural network input variables to learn and establish gold price forecast model, finally forecast performance is tested by simulation experiments. The test results show that combined model can well depict the gold price trend, simplify network structure, accelerate the network learning speed, and improve the prediction accuracy of gold price. It provides a new forecast method for gold price.

1. Introduction

The price of gold is the combined effect of many factors. It is characterized by randomness, uncertainty, and nonlinearity, and its fluctuation is very intense. Therefore, how to accurately forecast the price of gold has become a hot topic of concern to all countries [1].

The traditional gold price forecasting methods include: autoregressive conditional heteroskedasticity, time series analysis and other methods. The basic modeling of these methods is: gold price is a linear, deterministic change system, but the actual gold price change is extremely complicated and difficult to establish. Accurate mathematical model, so these traditional forecasts can not grasp the nonlinear phenomenon in the gold market, it is often difficult to fully reflect the changes in the price of gold, leading to larger prediction errors [2,3]. As non-linear prediction technology continues to mature, some scholars have introduced it into the prediction of gold prices. In particular, the neural network method does not need to understand the operating mechanism of the predicted object, and can effectively characterize the characteristics of the predicted object, becoming the main gold price Prediction method [4-6]. However, in the actual forecast of gold prices, there are quite a number of factors affecting the price of gold, such as the inflation rate, the effective exchange rate of the US dollar, the spot price of oil, and the world's gold reserves. If the entire gold price impact factor data is directly input into the neural network for learning, the calculation The volume and input dimensions have grown exponentially, and the network structure has become quite complex, resulting in slower convergence of the network and prone to "dimensional disasters" and other problems, which have adversely affected the forecast results of gold prices [7].
Projection Pursuit (PP) is a high-dimensional, nonlinear system analysis method that projects high-dimensional data onto low-dimensional subspaces to find projections representing high-dimensional data features in low-dimensional subspaces. Not only can we choose the information needed for prediction, but it can also effectively reduce the dimension of the feature space [8-10]. In order to improve the forecasting accuracy of gold price, this paper proposes a gold price forecasting method (PP-BPNN) based on the combination of projection pursuit and BP neural network. Through simulation experiments, an empirical analysis is performed to test the effectiveness and superiority of the combined method.

2. Gold price forecasting model

2.1 Influence factor preprocessing

The factors that determine the price of gold are:
\[ \{x(i, j)| i = 1, 2, \cdots; j = 1, 2, \cdots; p\} \]
\( n \) is the number of samples, \( p \) indicates the number of factors, \( x(i, j) \) is the j factor of the i sample. In order to make the factor dimension different influence the training efficiency of the BP neural network, the impact factor is normalized, as follows:
\[
\hat{x}(i, j) = \frac{x(i, j) - x_{\text{min}}(j)}{x_{\text{max}}(j) - x_{\text{min}}(j)}
\]

In the formula, \( \hat{x}(i, j) \) represents the normalized influence factor value, \( x_{\text{min}}(j) \) and \( x_{\text{max}}(j) \) represents the maximum and minimum value of the j influence factor.

2.2 Projection Pursuit Screening Gold Impact Factor

Step1: Construct projection index function \( Q(\alpha) \) Projection pursuit algorithm puts gold price data with p influence factors \( \{x(i, j)| j = 1, 2, \cdots; p\} \) The overall value is \( \alpha = \{\alpha(1), \alpha(2), \cdots, \alpha(p)\} \) which is the one-dimensional projection value \( Z(i) \) of the projection direction, that is:

\[
Z(i) = \sum_{j=1}^{p} \alpha(j)x(i, j), i = 1, 2, \cdots, n
\]

The projection index function can be expressed as
\[
Q(\alpha) = S_z D_z
\]

In the formula, \( S_z \) is the standard deviation of \( Z(i) \) and \( D_z \) is the local density of \( Z(i) \), \( S_z \) is defined as follows:
\[
S_z = \sqrt{\frac{\sum_{i=1}^{n} (Z(i) - E(z))^2}{n-1}}
\]

\[
D_z = \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{n} (R - r(i, j))*u(R - r(i, j))
\]

In the formula, \( E(z) \) is the average value of \( Z(i) \), \( R \) represents the window density of the local density, \( r(i, j) \) is the distance between two samples, and there are:
Step 2: Optimize projection index function

The optimal projection direction is estimated by solving the maximization problem of the projection index function, namely:

$$\max Q(\alpha) = S_z D_z.$$  \hspace{1cm} (7)

The constraints are

$$\sum_{j=1}^{p} \alpha_j^2(j) = 1$$ \hspace{1cm} (8)

Step 3: Determine the best projection direction

The solution of equation (8) is accomplished by a genetic algorithm. If the objective function of the genetic algorithm reaches an extreme value, then the projection direction at this time can be considered as the best projection direction, and the filtered gold price impact factor is obtained.

2.3 Gold Price BP Neural Network Algorithm

BP network is a feedforward neural network based on gradient descent algorithm. It consists of input layer, hidden layer and output layer. The basic structure is shown in Figure 1.

Let the output of the lth layer jth neuron be: $O_{pi}^{(l+1)}$, $W_{ij}^{(l)}$, representing two levels of neuron weights, the input and output relationships of neurons are:

$$O_{pi}^{(l+1)} = f_s[I_{pi}^{(l+1)}]$$

$$I_{pi}^{(l+1)} = \sum_{j=1}^{N_i} W_{ij}^{(l)} \cdot O_{pj}^{(l)} - \theta_{i}^{(l+1)}$$ \hspace{1cm} (9)

In the formula, $f_s[\cdot]$ is the network node function.

$$f_s[I] = \frac{1}{1 + \exp(-I / I_0)}$$ \hspace{1cm} (10)

Order $-\theta_{i}^{(l+1)} \cdot W_{pj}^{(l)} \cdot O_{pj}^{(l)} = 1$, there are
$$\begin{align}
O_{pi}^{(l+1)} &= f_i[I_{pi}^{(l+1)}] \\
I_{pi}^{(l+1)} &= \sum_{j=1}^{N_i} W_{ij}^{(l)} \cdot O_{pj}^{(l)}
\end{align}$$  \tag{11}$$

$$E_p \text{ is the sum of squared errors between the expected output and the actual output.}$$

$$E_p = \frac{1}{2} \sum_{i=1}^{N} (d_{pi} - y_{pi})^2$$  \tag{12}$$

By changing the weight coefficient $W_{ij}^{(l)}$, $E_p$ is minimized as much as possible so that the actual value approaches the desired output, and the steepest descent algorithm is used to change the weight coefficient toward the negative gradient of the error function. The adjustment formula of $W_{ij}^{(l)}$ is:

$$\Delta_p W_{ij}^{(l)} = -\alpha \frac{\partial E_p}{\partial W_{ij}^{(l)}}$$  \tag{13}$$

In the formula, $\alpha$ indicates the learning step.

$$\frac{\partial E_p}{\partial W_{ij}^{(l)}} = \frac{\partial E_p}{\partial I_{pi}^{(l+1)}} \cdot \frac{\partial I_{pi}^{(l+1)}}{\partial W_{ij}^{(l)}}$$  \tag{14}$$

Finally available,

$$\frac{\partial O_{pi}^{(l+1)}}{\partial I_{pj}^{(l+1)}} = \frac{O_{pi}^{(l+1)}(1 - O_{pi}^{(l+1)})}{I_0}$$  \tag{15}$$

2.4 Workflow of gold price forecasting model

A large number of studies have shown that for the hidden layer of BP neural network, as long as the number of hidden layer nodes is enough, only one hidden layer neural network can approximate a nonlinear function with arbitrary precision, so the hidden layer of BP neural network in this paper is 1, and because the gold price is a single output prediction problem, the output layer node of the BP neural network is one.

The gold price forecasting model includes the training phase and the prediction phase: (1) Training phase. The gold price training set is input into the BP neural network for learning, the parameters of the optimal model are found, and the gold price optimal forecasting model is established. (2) Prediction stage. Adopt the gold price forecasting model to predict the test high.

The core code of the gold price forecasting algorithm is as follows:

```matlab
Load data.txt
DD=data;% Import Gold Price Data
[n,p]=size(DD);
Np=number; number of % training samples, the first 1 to np samples are used to build the model, and the remaining samples are used for prediction
N=30; % population size
Kmax=500; % maximum iterations
Pc=0.80; % crossover probability
Pm=0.005; % variation probability%
[best_x,best_y,ALLX,ALLY]=GAUCP(K,N,Pm,LB,UB,D,Alpha)% Procedure for Invoking Genetic Algorithm to Optimize Projection Pursuit Model
Best a=(best_x(K))'; % direction vector
Disp('best projection vector');
```
Disp(best_a) ; % get gold price best projection vector 
Net=newff(best_a,output,num) ; % BP neural network training 
An=sim(net,inputn_test) ; % BP neural network prediction 

The gold price forecasting process based on projection pursuit and neural network is shown in Figure 2.

![Gold Price Forecasting Process of Combined Model](image)

**Figure 2. Gold Price Forecasting Process of Combined Model**

### 3. Gold Forecasting Example

#### 3.1 Experimental Environment and Related Parameter Settings

In order to verify the gold price combination prediction model of this paper, MATLAB 7.0 programming is used in the simulation environment of P4 dual-core 3.0G CPU, 2G memory and operating system Windows XP. The genetic algorithm for maximizing the projection index function selects an initial population size of 30, a crossover probability of 0.80, a mutation probability of 0.005, and an iteration number of 500.

#### 3.2 Sources of data

Selecting the Au999.5 price data from the Shanghai Gold Exchange for simulation experiments. The sampling time range was from July 15, 2010 to October 14, 2010 (data from Reith data [http://www.resset.com.Cn](http://www.resset.com.Cn)), a total of 100 sample data, the first 80 data as a training set to construct a gold price forecasting model, other data as a test sample set to test the establishment of model generalization capabilities, gold price data as shown in Figure 3.
3.3 Influence Factor Processing
(1) The gold price impact factor is selected initially. In order to more accurately describe the gold price trend, first of all, choose as many factors as possible to influence the price of gold. The original factors influencing the gold price selected in this article are: US dollar index, oil price, gold spot price, world gold supply, world gold demand Volume, Dow Jones index, ten-year Treasury yields in the United States, dollar exchange rate for the yuan.

(2) The projection pursuit algorithm of the factor is further chosen. Using the projection pursuit algorithm to calculate the best projection direction is:

\[ a^* = [0.145, 0.114, 0.346, 0.355, 0.447, 0.0569, 0.015, 0.075]. \]

(3) The impact factors are sorted according to the best projection direction, and five factors are selected: world gold supply, world gold demand, gold spot price, US dollar index, and oil price.

(4) The most influential factor of the above five factors on the gold price is the input node of the BP network, that is, the projection pursuit leads to five excellent input nodes.

3.4 BP Neural Network Simulation Results
The projection pursuit obtained five input nodes to the BP neural network for training. The learning process is shown in Figure 4. It can be clearly seen from Figure 4 that when the BP neural network is trained to the 80th generation, the prediction accuracy of the gold price reaches the preset requirement, which indicates that the model can be predicted at the gold price at this time. The results of fitting the model to the training sample are shown in Figure 5.
Figure 5. BPNN Gold Price Fitting Curve

As can be seen from Figure 5, the error between the fitting value and the actual value of the gold price of the BPNN is very small, which is quite close. This shows that the combination model is feasible and effective for forecasting the gold price and can predict the future gold price value.

3.5 Comparison with other models

To evaluate the quality of a predictive model, we mainly examine its predictive ability instead of the regression fitting results. At the same time, in order to make the prediction results of this model more convincing, we use PP-MLR, BPNN, and PP-BPNN to perform independent analysis. Predictions, among which PP-MLR denotes that the projection pursuit algorithm is used for factor screening first, and then multiple linear regression is used for modeling and prediction; BPNN means that without factor screening, BPNN is used for modeling prediction. The results of several models of gold price forecasts are shown in Figure 6.

Figure 6. Gold Price Forecast Results

From Figure 6, we can see that PP-BPNN can accurately capture the characteristics of gold price changes, and control the gold price forecast error within 8%, which can well meet the application requirements of gold prices.

The root mean squared error (RMSE) of the prediction results of the three prediction models was counted, and the results obtained are shown in Table 1.

| Prediction Model | RMSE   |
|------------------|--------|
| BPNN             | 12.33  |
| PP-MLR           | 27.45  |
| PP-BPNN          | 12.27  |

From Table 1, we can see that the RMSE of PP-BPNN is much lower than that of PP-MLR. This shows that for a gold price that exhibits a highly nonlinear variation law, it is difficult to establish an accurate mathematical model using the linear prediction model MLR, and the prediction error is large;
and PP - Compared with BPNN, BPNN does not reduce prediction accuracy, but learning speed increases significantly. It shows that PP is used to process and filter the impact factor of gold price, PP reduces the dimension of gold price impact factor, and extracts the most useful information. Eliminating the influence of related or repetitive information between impact factors can greatly improve the efficiency of neural network processing problems and help improve the forecast accuracy and efficiency of gold prices.

4. Conclusion
There is a nonlinear relationship between gold price and its influencing factors. Traditional linear factor screening methods and forecasting models cannot accurately describe the complex relationship between gold price and influencing factors. Therefore, this paper proposes a projection pursuit and BP neural network. Network combined gold price forecasting model. Simulation results show that the model can fully exploit the information volume between gold price and influencing factors, accurately fitting the gold price trend, and improving the accuracy of gold price forecasting. The model can also be applied to other multi-factors, Non-line prediction.

References
[1] Wang Xupeng, Zhang Hongcai, Zhao Yongqiang, He Lin. Hyperspectral image feature fusion algorithm based on projection pursuit[J]. Computer Applications and Software, 2006, 11: 1539-1541.
[2] Liu Hongjie, Huang Dongwei, Wang Yongzhao. Chaos Dynamics Analysis and Forecast of Gold Futures Price[J]. Journal of Tianjin Polytechnic University, 2011, 30(3): 85-88.
[3] Lin Yu, Kong Liuliu, Liu Pei. Gold Price Prediction Based on ARFIMA Model[J]. Journal of Nanhua University(Social Science Edition), 2010, 11(1):36-39.
[4] Liu Haoran. Analysis of Gold Price Influencing Factors and Investment Strategy[J]. Price Theory and Practice, 2007(10):57-58.
[5] Zhang Yanli, Zhang Desheng, Liu Changming, Li Jinfeng. Nonlinear combination forecasting model of gold price based on BP neural network[J]. Gold, 2011, 32(9): 5-8.
[6] Zhang Kun, Yu Yong, Li Wei. Application of Wavelet Neural Network in Gold Price Prediction [J]. Computer Engineering and Applications, 2010, 46(27): 224-227.
[7] Gu Xiaofei. Empirical research on price discovery function of China's gold futures[J]. Finance Teaching and Research, 2010,(4):64-66.
[8] LIU Zongxin, LI Xiaofeng. Multi-factor Prediction Model Based on Projection Pursuit and BP Neural Network[J]. Statistics and Decision, 2010,(1): 4-6.
[9] Li Song, Luo Yong, Zhang Mingrui. Prediction of Chaotic Time Series Based on BP Neural Network Optimized by Genetic Algorithm[J]. Computer Engineering and Applications, 2011, 47(29):52-55.
[10] Zhu Lihua, Liu Yian, Wei Min. Network intrusion detection method based on QPSO optimization projection pursuit[J]. Computer Engineering and Applications, 2011, 47(28): 80-82.