Gold Price Forecasting Based on Projection Pursuit Autoregression Model

Yan Sun
Beijing Information and Technology University, Beijing 100192, China

Abstract
A projection pursuit autoregression model is applied to forecasting gold future price in this study. Previous studies focused on fitting and forecasting the volatility of gold price to reveal the characteristics of gold market. However, based on the high dimensional data approximation capability, PPAR model can not only fit but also forecast gold price more precisely than other models. In this study, the PPAR model is applied to the daily closing price of gold futures, and the results show an overall improvement in forecasting using PPAR model as compared to a BPNN method, especially on stability, which reflects the advantage of PPAR method in dealing with a huge amounts of data.

Keywords: Gold Price, Projection Pursuit Autoregression, Forecast Model

1. Introduction
The interpretation and prediction of the macro trend of the gold market is a hot spot given its economic and financial importance. In earlier studies, the volatility of gold price has been explained with many methods. Engle [1] proposed autoregressive conditional heteroskedasticity (ARCH) model to study the law of the variation of the inflation sequence in the UK. Bollerslev [2] applied generalized autoregressive conditional heteroskedasticity (GARCH) model to describe the characteristics of the financial time series. Thereafter, the GARCH models became the main tools to study gold price volatility. Xiutian Zheng [3] used GARCH-M model to fit the trend of gold price, and analyzed the relationship between risk and return of China’s gold market. Efimova and Serletis [4] also draw the conclusion that risk and return are positively related. Werner and Marcel [5] proposed a hybrid ANN-GARCH model to forecast the gold price volatility and determine the main financial variables that influence the gold price volatility.

All these studies focused on fitting and forecasting the volatility of gold price. However, the ability to forecast the gold price in a short term is more important for investors to make investment decisions in the gold market. In this regard, GARCH models make great resultant errors. Therefore, the neural network methods have been used to build more accurate models to fit and forecast gold price. Guixia Yuan [6] and Li Chen [7] proposed a hybrid projection pursuit BP neural network model to forecast 20 days gold price. Nonetheless, the limitation of this method lies in its weak stability. They studied a short term gold price within 100 days, which realized a high accuracy in forecasting. But, when the sample size expands, the algorithm collapses fast and over fitting appears.

Projection pursuit method was proposed by Kruskal [8] and generalized by Friedman and Tukey [9]. It projects high-dimensional data to low-dimensional space and preserve the necessary information at the same time. In this study, a projection pursuit autoregression model is applied to forecast gold price for the first time. The PPAR method is focused on the gold price data, and other factors are not considered. Compared with other models, the sample size has been expanded and the accuracy of forecasting has been raised. Therefore, the conclusion of this study is important for investors to reduce risk and to get a better asset allocation.

The remainder of this work is divided into three sections. First, the realization of PPAR model is described and the performance evaluation index is given. Second, empirical analysis is conducted to test the effectiveness and accuracy of PPAR model, and compared the results with other models. Finally, the last section summarized this work and put the perspective of future research.

2. Methodology
2.1 Procedure of PPAR
PPAR model bases on projection pursuit model, and combines with time series autoregression method. Through linear projection, the high-dimensional data can be projected to the low-dimensional subspace. In the mean time, the necessary information is still preserved. Then, by using the ridge function and the approximate regression polynomial, the function is finally realized. This method has advantages in convergence rate and prediction accuracy. The modeling steps as follows:

This work is supported by Science and Technology Project of Beijing Municipal Education Commission (KM201611232019) and NSFC (71501016).
(1) Determine the projection value $z_i$:

$$z_i = \sum_{e=1}^{p} \alpha_e x_{i-e} \quad (i = p + 1, p + 2, \ldots, n) \quad (1)$$

Where, $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_p)$ is the projection direction. $p$ is the number of prediction factor, $x_{i-e}$ is the sample values.

(2) Based on orthogonal Hermit polynomial, the PPAR model is built as:

$$\hat{x} = \sum_{i=1}^{M} \sum_{j=1}^{r} v_i h_j(z) \quad (i = 1, 2, \ldots, M; \ j = 1, 2, \ldots, r) \quad (2)$$

where $\hat{x}$ is the prediction value, $M$ is the number of ridge function, and $r$ is the polynomial order. $v$ are the coefficients of the polynomial, and can be obtained by least square method. $h$ is the orthogonal Hermit polynomial.

(3) Optimize projection indicator function. Based on the real coded accelerating genetic algorithm to solve the problem of minimizing the projection indicator function is the best way to reflect a kind of structure characters of high-dimensional data as much as possible. Therefore, the best projection direction is estimated by

$$\min F(\alpha, v) = \frac{1}{n - p} \sum_{i=p+1}^{n} (x_i - \hat{x}_i)^2$$

$$s.t. \sum_{j=1}^{p} \alpha_j^2 = 1 \quad (3)$$

(4) Calculating and fitting residual error $r_i = x_i - \hat{x}_i$. If the results meet the condition, the model parameter will be returned. Otherwise, $x_i$ will be replaced by $r_i$. And the calculating will go back to step (1) for the next optimization until the condition is meet and return the final results.

2.2 Performance evaluation index

The performance of model will be evaluated by accuracy and stability. The NMSE (Normalized Mean Squared Error) and the MRE (Mean Relative Error) can indicate the accuracy of forecasting. The EF (Error Fluctuation) [10] is the overall fluctuation range, and is used to indicate the stability of the model. These indexes are calculated as:

$$NMSE = \frac{1}{NS^2} \sum_{i=1}^{N} (x_{Oi} - x_{Fj})^2$$

$$MRE = \frac{1}{M} \sum_{j=1}^{M} \left| \frac{x_{Oj} - x_{Fj}}{x_{Oj}} \right|$$

$$EF = \frac{\sum_{i=1}^{N+M-1} \sum_{p=1}^{N+M} \sqrt{\sigma_i^2 - \sigma_p^2}}{\sum_{i=1}^{N+M} t}$$

Where $N$ is the number of training period sample, $S^2$ is the sample variance of training period sample. $x_{Oi}$ is the target value of sample $i$ and $x_{Fj}$ is the training value of sample $i$ within training period. $M$ is the number of forecasting period sample. $x_{Oj}$ is the target value of sample $j$ and $x_{Fj}$ is the training value of sample $j$ within forecasting period. $\sigma_i$ and $\sigma_p$ are the relative errors.
of sample \( t \) and sample \( p \). The model has higher accuracy when the values of NMSE and MSE are lower, and the value of EF is also the smaller the better, which means the model is more stable in forecasting.

3. Empirical Results and Analysis

3.1 Results of PPAR model

The gold price data of Au(T+D) in Shanghai Gold Exchange is selected to test the performance of PPAR model. Au(T+D) is one of the gold future trading, and in this study, daily closing price from Jan 2\(^{nd}\) in 2008 to July 31\(^{th}\) in 2013 (the data come from RESSET http://www.resset.com) has been divided into 2 parts: training period sample (the front 1200 data) and forecasting period sample (the rare 256 data). The data is shown in Figure 1.

Data of training period sample are normalized and the best projection direction is returned as:

\[
\alpha = [0.098, 0.060, -0.034, -0.027, 0.046, -0.025, 0.014, 0.006, 0.092, 0.021, -0.083, 0.033, 0.013],
\]

when the training period sample is input and one ridge function is used to fit. Within the procedure, the polynomial order is selected as 15. Using R language program, the fitted value of training period are gained, and the actual value and fitted value are shown in Figure 2.

![Figure 1. Gold price data from Jan 2\(^{nd}\) in 2008 to July 31\(^{th}\) in 2013](image)
It can be seen from Figure 2 that the fitted value and actual value are approximate, and PPAR model has great advantage in fitting. Thus, the PPAR model can be used to forecast a short term of gold price. Within the procedure, the predict order is selected as 1. After calculating, the forecasting value of testing period sample is output, and the forecasting result is shown in Figure 3.

It can be seen from Figure 3 that PPAR model has permitted error in forecasting. Furthermore, the PPAR model takes one step predicting, which means within testing period, all the data are predictions except the first one. However, it shows that PPAR model is stable for making a short term forecasting, and even after 200 days, the error is still small. Altogether, the PPAR model is effective and feasible.

3.2 Compare with results of other models

Traditional forecasting methods are mainly GARCH models and BPNN model. The GARCH models
have better performance in fitting and explaining economic phenomenon than forecasting future gold price. Because of the heteroskedasticity of time series, the data always have to be taken one or two-order difference to meet the stationary condition. Therefore, it is hardly to forecast the price itself. For this reason, the root mean square errors (RMSE) of three forecast models are calculated to compare the accuracy of forecasting. The results are shown in Table 1.

<table>
<thead>
<tr>
<th>Forecast model</th>
<th>RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>BPNN</td>
<td>17.15%</td>
</tr>
<tr>
<td>GARCH(1,1)</td>
<td>57.33%</td>
</tr>
<tr>
<td>PPAR</td>
<td>12.79%</td>
</tr>
</tbody>
</table>

It can be seen from Table 1 that the RMSE of GARCH(1,1) model is the highest one, which means the GARCH(1,1) has weak ability of forecasting. Therefore, in the next part, the BPNN model is chosen to compare with PPAR model for testing their forecasting abilities.

For BPNN model, the hidden layer function is selected as Tansig function, output layer function is Purelin function and the training function is Levenberg-Marquardt function. The results of performance evaluation index are shown in Table 2.

<table>
<thead>
<tr>
<th>Model</th>
<th>NMSE</th>
<th>MRE</th>
<th>EF</th>
</tr>
</thead>
<tbody>
<tr>
<td>PPAR</td>
<td>3.82%</td>
<td>3.60%</td>
<td>20.71%</td>
</tr>
<tr>
<td>BPNN</td>
<td>1.02%</td>
<td>6.35%</td>
<td>54%</td>
</tr>
</tbody>
</table>

It is shown in Table 2 that in the forecasting period, all the performance indicators of PPAR model are smaller. The NMSE of PPAR model is higher than that of BPNN model, but the MRE of PPAR model is lower than that of BPNN model which means these two models have similar performance in accuracy ability. However, the EF of PPAR model is much lower than that of BPNN model, which means PPAR model is much more stable than BPNN model in forecasting. And the stability of model can be tested by enhance the size of sample. In that case, BPNN model loses efficiency faster than PPAR model.

4. Conclusion

In order to fitting the high nonlinear and mutable data of the gold daily closing price and forecasting precisely the future price, the PPAR model is applied to solve the problem for the first time. And by comparing with other models, it shows that the high-dimensional data approximate capability of PPAR model can better realize the purpose of predicting the future price from the inherent mechanism of financial data. The PPAR algorithm shows robust computing power for large data, so the future research may want to analyze new hybrid models to predict future data when the sample size is bigger than that in this study. And the results will help the investors making decisions and get a better asset allocation and portfolio diversification.

5. References