

# Predict Gold Price Trend Based on ARIMA Model

D L Gao<sup>1,\*†</sup>, J Y Liang<sup>2,†</sup> and B H Xu<sup>3,†</sup>

<sup>1</sup>School of Mathematical and Statistical Sciences, Ludong University, Yantai, Shandong 264011(post), China;

<sup>2</sup>School of Information and Computing Sciences, Jinan University, Guangzhou, Guangdong 511436(post), China;

<sup>3</sup>School of Mathematics and Applied Mathematics, Sun Yat-sen University, Guangzhou, Guangdong 510275(post), China.

\*Corresponding author email: 3180100498@caa.edu.cn.

†These authors contributed equally.

**Abstract.** As a financial product, gold is one of the more important spot and futures trading products in the commodity market. Based on the time series model, the gold price can be fitted and predicted, in order to explore the law of gold price changes. It has positive implications for investors and government managers. This article selects the Prime Day price in 2018 as the research object. Combined with domestic and foreign research content on financial time series. First, through the time series diagram test and the unit root test, it is obtained that the gold daily price series is a cycle-free and non-stationary series. Therefore, the time series needs to be differentiated. Second, a new stationary sequence is obtained by making second-order differences. Third, after the time sequence diagram test, ADF test, and white noise test, the sequence is a non-white noise sequence. Comparing the AIC values of multiple time series models, the most ideal model for the series should be the ARIMA (2,2,2) model. The significance test of the model shows that the fitted model is significantly effective. And the significance test of the model parameters is also passed. Then make predictions with this model. Comparing the predicted value with the future real gold price, it is found that the predicted value is close to the real value. This is a good reference for the country to formulate relevant economic policies.

**Keywords:** gold price, time series, ARIMA model, white noise test, significance test.

## 1. Introduction

The first letters of the three words gold, black gold, and US dollars just form the God worshipped by Christianity. The world's financial and economic pursuit of gold, black gold, and US dollars is just like the worship of God by Christians. The reason why gold is respected by people is that it is determined by the nature of gold itself. As a precious metal, gold has excellent physical properties and stable chemical properties, and is widely used in various fields. On the other hand, gold has excellent hedging function and the gold market is difficult to control, so gold has strong investment benefits. Throughout the ages, gold has been used by people to preserve wealth. In August 2015, stock markets around the world took a hit, with stock prices plummeting, with the S&P 500 down 10%, but the

dollar-denominated gold price up 5%. This phenomenon also occurred in the "Black Monday" of 1987 and the financial crisis in 2007. There are many more examples of how gold's value continues to rise every year during times of war and economic downturns. In recent years, as an investment product, gold has attracted more and more attention. The price of gold has also become the focus of people. With the increasing role of the gold market, statistics and forecasts on the fluctuations and prices of the gold market have attracted widespread attention. Zou believes that in the long run, holding gold can effectively diversify the risk of bond market and inflation [1]. Analysis by the World Gold Council shows that during market volatility, gold can be used in investment portfolios to protect global purchasing power [2]. Gold is currently becoming a scarce resource. Because the value effect of prospecting has dropped by 40% compared with previous years. The latest report from the World Gold Council shows that central banks continue to buy gold [3].

So far, scholars in various fields have found that the fluctuations of the gold market are not irregular, but have certain predictability through their research on the gold market. Through the establishment of various models to conduct research on gold price prediction, a number of theoretical research results have been obtained. Xie et al. used the M-Copula-GJR-VaR model to effectively improve the hedging effect and asset returns [4]. Chen et al. analyzed the relevant factors that affect the price of gold [5]. Using the organic combination of neural network model and ARMA, the prediction of gold price is realized. And gives some trading suggestions for gold-related derivatives. Cao constructed a timing trading strategy after optimizing the penalty parameters in the SVM model and the parameters in the kernel function with the goal of minimizing MSE [6]. Cheng used the grey correlation degree to analyze the factors that significantly affect the price of gold, and established a multi-factor BP neural network model. The economic factors that significantly affect the price of gold are added to the forecast model, which improves the forecast accuracy [7]. Xu made a statistical analysis of the closing price of the Shanghai Gold Exchange on the last trading day of each month. A quadratic curve fitting model is established and short-term prediction is made based on it, which provides a reference for investors to invest in gold [8]. Wang used a variable coefficient regression model to predict the gold price, which greatly improved the prediction accuracy [9]. The dynamic evolution of gold prices reflects the investment decisions of economic actors in the financial market. The dynamic evolution process of gold price is also a process of data generation [10]. There are many methods used by domestic and foreign scholars to study the trend of gold price. But it has certain limitations. This paper used the time series correlation theory to establish an ARIMA model for the gold price in the London gold exchange market. and conduct empirical analysis.

In the eyes of many investors, gold is a "veritable" value preservation and appreciation product. Statistics show that over the past 12 years, the price of gold has risen continuously from nearly \$300 to nearly \$1,900 per ounce. However, with the large fluctuations in market prices, the risk of blindly purchasing gold to preserve and increase its value cannot be ignored [11]. Yin and others pointed out that the correlation between the gold spot and the average futures return and other factors changes with time, positive correlation and negative correlation appear alternately, and the time-varying is obvious. Whether it can be used as a safe-haven asset in the real economy or the stock market varies from time to time [12]. Therefore, in-depth research on the price of gold can allow us to better predict the price of gold. It not only helps the country to adjust economic policies in a timely manner, but also helps individuals and institutions to control financial risks. It has theoretical and applied value.

## 2. Methodology

The goal of this article is to analyze the gold price fluctuations and trends in 2018 based on the ARIMA model. By fitting and evaluating the corresponding models, the ideas and methods of modeling are shown in the figure 1, the following are detailed explanation.

After the data is collected, the data for 2018 is preprocessed. Through the test of the time series chart, it is observed that the data has no periodicity and has a clear downward trend. After the unit root test, a P value greater than 0.05 was obtained. It shows that the series is non-stationary, so the time series data needs to be differentiated. After the first difference, a stationary series is obtained. The

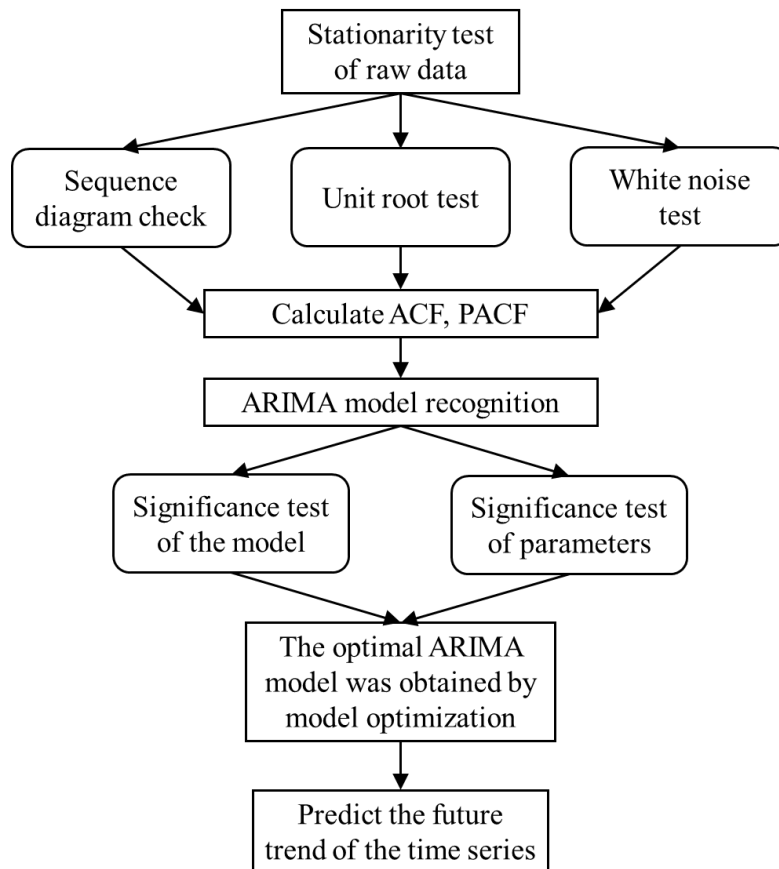
white noise test shows that the P value is less than 0.05 at the 6th and 12th order delays. Prove that the sequence is a non-white noise sequence. Then proceed to follow-up experiments.

Plot autocorrelation plots and partial autocorrelation plots of the model, as well as the results of automatic ordering of the model. Comparing the AIC values of ARIMA (1,2,1), ARIMA (2,2,2), ARIMA (3,2,3) and other related models, comprehensive test conditions and prediction effects, the most ideal model for the sequence should be ARIMA (2,2,2) model. Based on the conditional least squares estimation method, the fitting results of the ARIMA (2,2,2) model are obtained as follows:

$$\nabla^2 x_t = \frac{1-0.925B+0.929B^2}{1-0.998B-0.174B^2} \varepsilon_t \quad (1)$$

After the ideal model is obtained, the model and parameters are tested for significance. It is obtained that the P value is greater than 0.05 at the 6th and 12th order delays. Therefore, the null hypothesis is not rejected. This shows that all model-related information has been extracted, and the sequence is a purely random sequence. It is proved that the residual sequence of the fitted model belongs to the white noise sequence, that is, the fitted model is significantly effective. By constructing t-statistics, it was found that at a significance level of 0.05, the P-values for all parameters were less than 0.05. Therefore, the null hypothesis that the parameter is significantly 0 is rejected, that is, the parameter significance test of the model is passed.

After the model passes the test. By comparing with the future real value, it can be obtained that the prediction result of the corresponding model in the next 5 days is better. It has a good reference value for investors.



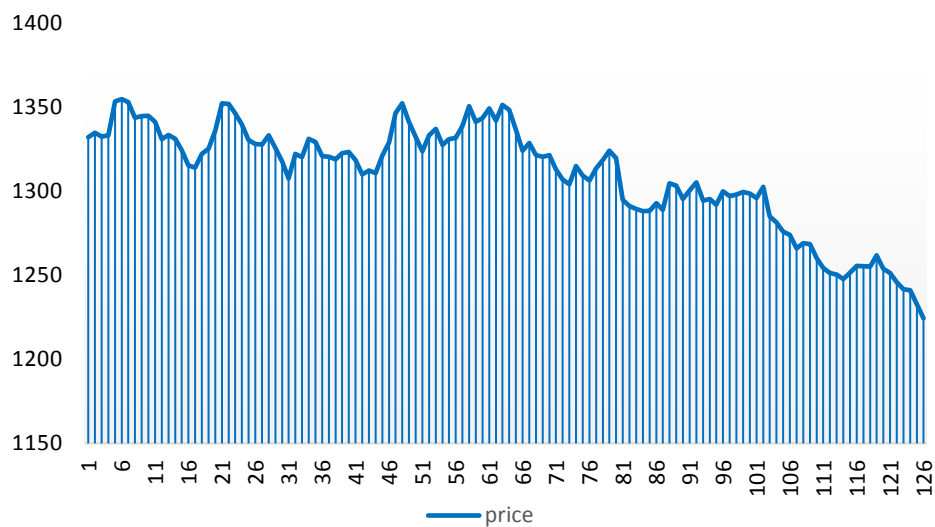
**Figure 1.** Frame diagram of paper.

### 3. Results and discussion

The following section selects the data from 2018.1.18–2018.7.18. Based on the final transaction amount on normal trading days, there are a total of 126 sample data. Analyze and compare the establishment of ARIMA (2,2,2) model, and predict the future trend of gold.

#### 3.1. Pre-processing of raw data

*3.1.1. Sequence diagram check.* The timing diagram inspection mainly refers to the inspection method that makes judgments according to the characteristics of the timing diagram. Graph tests of stationarity rely on the principle that stationary time series have constant mean and variance. This means that a log series test plot for a stationary series should show that the series fluctuates consistently around a constant value, and the fluctuation range is bounded.



**Figure 2.** Time series chart of gold price movements.

From the time series plot (Figure 2), there is a clear downward trend in the data. The series does not appear to fluctuate consistently around a constant value, and the fluctuations are not periodic. Therefore, it is preliminarily judged that the data is not stable.

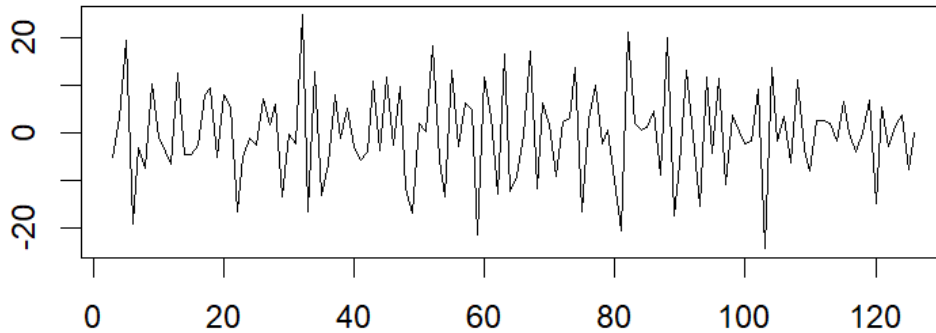
*3.1.2. Unit root (ADF) test.* The unit root test (ADF test) is the most common method for constructing a statistic for stationarity testing. In order to accurately judge whether the original series is stationary, it is necessary to further perform unit root test on the data. As shown in the table below.  $p=0.803$ , greater than 0.05, it is concluded that the series is non-stationary. The data needs to be differentially processed.

**Table 1.** ADF test of raw data.

Augmented Dickey-Fuller Test	
Dickey-Fuller	-1.454
Lag order	4.000
p-value	0.803

*3.1.3. Sequence diagram after second order difference.* The series has a clear downward trend before the difference, so the series is not stationary. It can be seen from Figure 3 that after the difference, it

can be clearly seen that the sequence has no obvious trend. It fluctuates up and down within a certain constant range, so it can be preliminarily judged that the sequence is stable. Then proceed to the following experiment.



**Figure 3.** Sequence diagram after second order difference.

**3.1.4. Unit root (ADF) test after difference.** As shown in the table below, it can be seen from the results of the ADF test after the difference.  $P < 0.01$ , less than the significant level of  $\alpha$  of 0.05, it can be considered that the series is a stationary series after the second-order difference. In order to facilitate subsequent pure randomness testing. If the sequence is stationary, the situation will be simpler. Well-established stationary series modelling methods can be used.

**Table 2.** ADF test after difference.

Augmented Dickey-Fuller Test	
Dickey-Fuller	-8.755
Lag order	4.000
p-value	<0.01

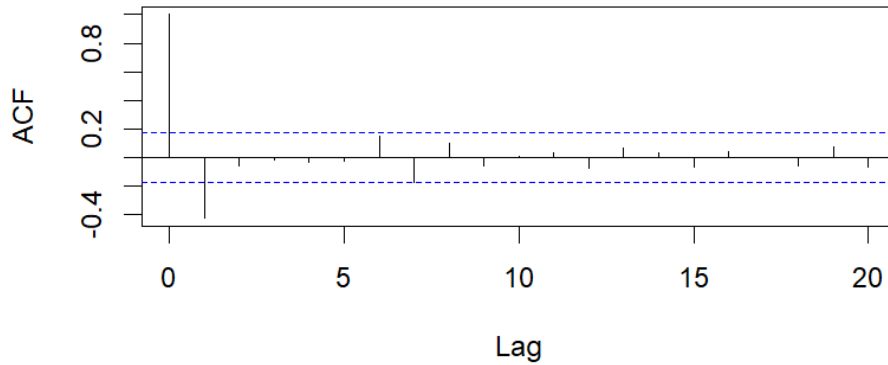
**3.1.5. White noise test.** In order to ensure the validity of the subsequent experiments, the white noise test is carried out on the time series. As shown in the table below, it is found that the P values corresponding to the LB statistic are all less than the significant level ( $\alpha=0.05$ ) after the 6th-order delay and the 12th-order delay. Therefore, the null hypothesis that the sequence is a random sequence (the null hypothesis is white noise) is rejected, and the sequence is considered to be a non-white noise sequence.

**Table 3.** White noise test.

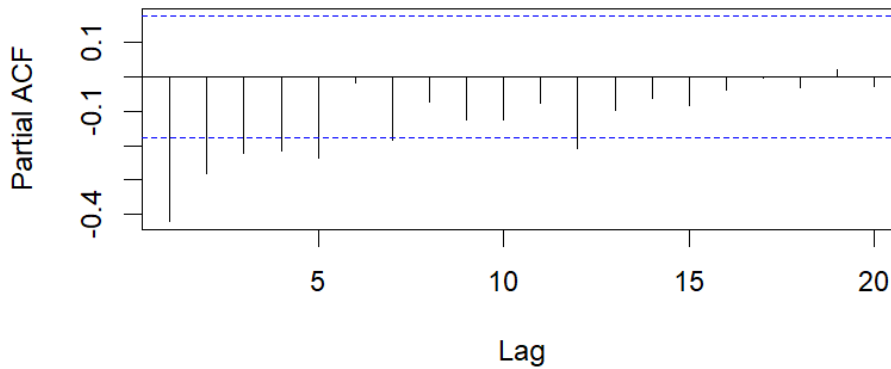
Delay in order	Value of the LB statistic	P values	Conclusion
6	26.092	<0.001	Non-white noise sequence
12	32.819	0.001	

### 3.2. Model identification

**3.2.1. Order determination and discrimination of models.** The following are the autocorrelation diagram (Fig. 4) and the partial autocorrelation diagram (Fig. 5) after the second order difference of the sequence. The model is automatically ordered.



**Figure 4.** Since the correlation diagram.



**Figure 5.** Partial autocorrelation graph.

Compare the AIC values of related models such as MA (1), AR (2), ARIMA (1,2,1), ARIMA (2,2,2), ARIMA (3,2,3). According to the comprehensive test situation and prediction results, the most ideal model for obtaining the sequence should be the ARIMA (2,2,2) model.

**Table 4.** Parameter estimation.

	Ar1	Ar2	Ma1	Ma2
Parameter	0.998	-0.174	-0.925	0.929
Standard error	0.122	0.092	0.098	0.097

ARIMA(p, d, q) model:

$$\Phi(B)\nabla^d x_t = \Theta(B)\varepsilon_t \quad (2)$$

$$E(\varepsilon_t) = 0 \quad (3)$$

$$Var(\varepsilon_t) = \sigma_\varepsilon^2 \quad (4)$$

$$E(\varepsilon_t \varepsilon_s) = 0 \quad s \neq t \quad (5)$$

$$E(x_s \varepsilon_t) = 0 \quad \forall s < t \quad (6)$$

$\Phi(B)$  is the delay operator acting on the error sequence;  $\nabla^d$  is  $d$ -order difference sequence;  $x_t$  is a sequence value at time  $t$ ;  $\Theta(B)$  is the delay operator acting on  $x_t$ ;  $\varepsilon_t$  is the error of  $t$  time;  $\varepsilon_s$  is the error of  $s$  time;  $x_s$  is a sequence value at time  $s$ ;  $\sigma_\varepsilon^2$  represents the variance of  $\varepsilon$ ;  $B$  is a form of the delay operator;  $E(\varepsilon_t \varepsilon_s)$  is the product of the expected value of  $\varepsilon_t$  and  $\varepsilon_s$ , indicating that the errors at different times are irrelevant;  $E(x_s \varepsilon_t)$  is the product of  $x_s$  and  $\varepsilon_t$  expected value.

To sum up, based on the conditional least squares estimation method, the fitting result is obtained as:

$$\nabla^2 x_t = \frac{1 - 0.925B + 0.929B^2}{1 - 0.998B - 0.174B^2} \varepsilon_t \quad (7)$$

3.2.2. *Significance test of ARIMA (2,2,2) model.* By performing a white noise test. As shown in the table below, it can be concluded that the ARIMA (2,2,2) model has p-values significantly greater than 0.05 at delays 6 and 12. Therefore, the null hypothesis is not rejected. Indicates that model-related information has been extracted. The sequence is a purely random sequence. It is proved that the residual sequence of the fitted model belongs to the white noise sequence, that is, the fitted model is significantly effective.

**Table 5.** Significance test of the model.

Delay in order	Value of the LB statistic	P values	Conclusion
6	1.270	0.973	The fitting model is remarkably effective
12	5.241	0.949	

3.2.3. *Significance test of parameters.* The T statistic was constructed and parametric test was performed, and the results were as follows:

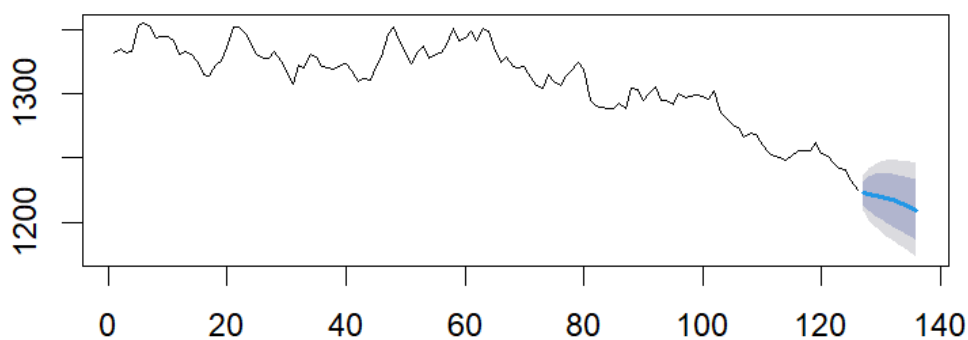
**Table 6.** Significance test of parameters.

T statistic	T <sub>1</sub>	T <sub>2</sub>	T <sub>3</sub>	T <sub>4</sub>
Parameter	0.998	-0.174	-1.925	0.929
S.E.	0.104	0.091	0.181	0.180
P values	<0.001	0.029	<0.001	<0.001

The T statistic is being constructed as shown in the table. The significance level is at 0.05. The P value for T<sub>1</sub> is less than 0.05, rejecting the null hypothesis that the parameter is significantly 0. The P value for T<sub>2</sub> is less than 0.05, rejecting the null hypothesis that the parameter is significantly 0. The P value corresponding to T<sub>3</sub> is less than 0.05, rejecting the null hypothesis that the parameter is significantly 0. The P value corresponding to T<sub>4</sub> is less than 0.05, rejecting the null hypothesis that the parameter is significantly 0. Therefore, the parametric significance test of the model is passed.

### 3.3. Model to predict

After the previous data preprocessing and the significance test of the model and parameters. The model can be further used to predict the future trend of the series, as shown in the figure below.



**Figure 6.** Sequence prediction graph.

The image above shows the forecast for this time series. It can be seen from the chart that the price of gold is still in a downward trend in the next 5 days. The tested ARIMA (2,2,2) model was used for

fitting. It can more accurately predict the price of gold for many days in the future. Here is an example of the price of gold for the next five days, as shown in the table below.

**Table 7.** Predicted value and confidence interval.

Time	Predictive value	The real value	80% confidence interval	95% confidence interval
1	1222.78	1217.55	(1208.946, 1236.610)	(1213.734, 1231.822)
2	1222.03	1228.75	(1201.738, 1242.324)	(1208.762, 1235.300)
3	1221.20	1224.95	(1196.733, 1245.659)	(1205.200, 1237.191)
4	1220.11	1222.35	(1192.714, 1247.502)	(1202.196, 1238.020)
5	1218.78	1231.50	(1189.176, 1248.376)	(1199.421, 1238.130)

By comparing the predicted value with the actual value, it is found that there is little difference between the two. The ARIMA (2,2,2) model has good prediction results and can be used to predict the price fluctuation of the US dollar against gold. The forecast has a good reference significance for the formulation of corresponding national economic policies.

#### 4. Conclusion

This paper conducts an empirical analysis of the gold price data in 2018 based on the ARIMA model. After the time series test and the unit root test, it is known that the daily price of gold is a non-stationary sequence. Then, a stationary sequence is obtained by differential processing, and the sequence is determined as a non-white noise sequence by white noise test. Finally, by comparing the AIC values of different ARIMA models, the optimal time series model ARIMA (2,2,2) is obtained. After this, the gold price forecast for the next five days is made and the result is better. This has good guiding significance for investors and the market.

Since this article selects 6 months of data in 2018 for analysis, it estimates the price trend of gold in the short term. Therefore, it is impossible to give a good forecast for the price of gold in the long term. For short-term data operations, the impact of the long-term memory of financial assets (historical events will affect the price of financial assets for a long time) can be ignored, providing a reference for short-term investors. To predict the trend of gold prices more accurately, long-term influencing factors cannot be ignored. Further research in this direction can be carried out in the future.

#### References

- [1] Zou Q 2014 *Research on the Financial Function of Gold* (Shanghai: Shanghai Academy of Social Sciences)
- [2] Duan H 2021 *Research on Gold Futures Price Prediction Based on Time Series* (Harbin: Harbin Institute of Technology)
- [3] Ou H 2016 *Statistical Analysis and Prediction of Gold Price* (Fujian: Fujian Normal University).
- [4] Xie C, Qu M and Wang G 2013 *Study on the Optimal Hedging Ratio of Gold Market Based on M-Copula-GJR-VAR Model* (Management science) **26**(02) pp 90-99
- [5] Chen X, Tian L and Han X 2018 Gold price forecast analysis and research. *Journal of Foshan University of Science and Technology* (Natural Science Edition) **36**(04) pp 6-10
- [6] Cao X 2017 *Gold Price Prediction Model and Parameter Optimization Based on SVM* (Shandong: Shandong University)
- [7] Cheng M 2020 *Research on Forecasting Method of International Gold Price* (Shandong: Shandong University)
- [8] Chen Xu 2015 *Application of Time Series in Gold Price Forecasting Business* 2015(**16**) pp 180-181



- [9] Wang Y 2010 *Research on Gold Price Forecast Based on Variable Coefficient Regression Model* (Tianjin: Tianjin University)
- [10] Pan X 2020 *Empirical Analysis of Gold Price Based on ARIMA-GARCH Model Business* 2020 (20) pp 155-156
- [11] Chen Y 2013 *Can Gold Hold its Value with Volatile Prices?* NCNA 2013-05-03(006)
- [12] Yin L and Liu Y 2015 *Is Gold a Stable Haven? — From the Perspective of Macroeconomic Uncertainty International Financial Studies* 2015(07) pp 87-96