Short-term Forecasting of Gold Price Based on ARMA Model

Wanle Chi
Wenzhou Vocational &Technical College, Wenzhou 325000, China.
358455713@qq.com

Keywords: Forecasting, Gold Price, ARMA, ADF.

Abstract: old price is complicated for containing many factors. So the evolution process in the gold price is important. In the paper the sequence of gold price is defined as a time series. The ARMA model is used to solve time series problems, especially in the field of finance. The ARMA model is a regression model developed by Box and Jenkins to identify, assess and diagnose dynamic time series model in which the time variable plays a key role. The paper collects the afternoon fixing price of London gold market the period of 1990-01-01 to 2018-01-12 (7315 working days) and uses ARMA model to solve the short-term forecast problem of gold price. The paper is helpful to gold investors to make more scientific decision-making.

1. Introduction

Time series is a series of chronological order data. There is a certain dependence between them. Therefore, we can forecast the future data through the study of time series. The paper collects the afternoon fixing price of London gold market the period of 1990-01-01 to 2018-01-12 (7315 working days) and uses ARMA model to solve the short-term forecast problem of gold price. The paper is helpful to for gold investors to make more scientific decision-making.

2. ARMA Model and Stability Test

2.1 ARMA Model

The time series are time dependent variables. A single sequence value of the time series is uncertain, but the variation of the whole sequence is regular. So time series can be approximated with corresponding mathematical models. The ARMA model is a commonly used stochastic time series analysis model. It is created by Box and Jenkins, also known as B-J method. There are three basic types of models: auto-regressive (AR), moving-average (MA) and auto-regressive& moving-average (ARMA) [1].

The auto-regressive (AR) model reflects the relationship between the current value and the previous values. The mathematical formula of AR model is

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \varepsilon_t \tag{1}$$

The y_t is a stationary time series. The ϕ_i (i=1,2,...,p) are undetermined coefficients of AR model. The p is the order of AR model. The ε_t is the error of AR model.

The moving-average (MA) model reflects the random error relationship between the current value and the previous values. The mathematical formula of the MA model is

$$y_t = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} \ldots - \theta_q \varepsilon_{t-q} \tag{2}$$

The y_t is a stationary time series.θ_j(j=1,2,...,q) are undetermined coefficients of MA model. The q is the order of MA model. The ε_t is the error of MA model.

The ARMA model is a combination of the AR model and the MA model. The mathematical formula of the ARMA model is the

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} \ldots - \theta_q \varepsilon_{t-q} \tag{3}$$

The parameters of formulas (3) are the same as formula (1) and formula (2).

The suitable model can be chosen, according to the autocorrelation coefficient diagram and the partial correlation coefficient diagram of the sequence.

If the autocorrelation coefficient is trailing and the partial correlation coefficient is p order truncation, the AR(p) model should be selected. If the partial correlation coefficient is trailing and the autocorrelation coefficient is q order truncation, the MA(q) model should be selected. If the partial correlation coefficient is trailing and the autocorrelation coefficient is trailing, the ARMA (p, q) model should be selected.

2.2 Stability Test

For a time series, it is necessary to test whether it is stationary. The ADF test is a commonly used unit root test method to test stationarity. A sequence which has a unit root is non-stationary. The non-stationary original sequence as a first order difference needed further processing. Then the stability of the sequence difference was re-inspected. The basic formulas of ADF test are as follows:

$$\Delta u_t = c + \delta u_{t-1} + \sum_{i=1}^{p-1} \beta_i \Delta u_{i-1} + \varepsilon_t \tag{4}$$

$$ADF = \frac{\hat{\delta}}{S(\hat{\delta})} \tag{5}$$

The $S(\hat{\delta})$ is the sample standard deviation of δ[2].

With compare the ADF critical value and ADF statistical value, we can judge whether the original hypothesis ($\delta = 0$) should be rejected. If the ADF statistical absolute value is more than the ADF critical absolute value, we do not accept the original hypothesis ($\delta = 0$). And this time series is stationary. If the ADF statistical absolute value is less than the ADF critical absolute value, we accept the original hypothesis ($\delta = 0$) , and this time series is non-stationary.

3. Empirical Analysis

3.1 Original Data & Stabilization

The paper collects the afternoon fixing price of London gold market the period of 1990-01-01 to 2018-01-12 (7315 working days) from https://www.gold.org.

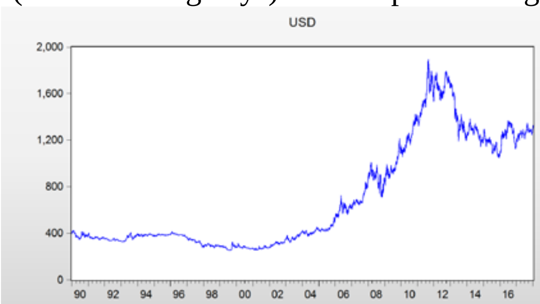

Fig. 1 Daily price

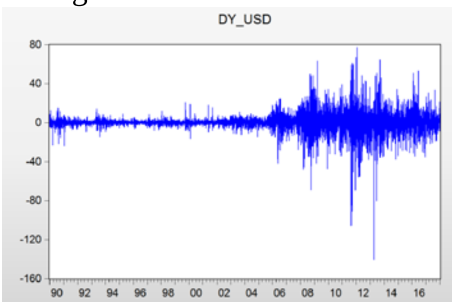

Fig. 2 First order difference data

From fig. 1, the daily gold price is a non-stationary time series. But the sequence of first order difference may be stationary (fig. 2).

The test results are as fig. 3. The ADF test are applied to test the original data sequence. It was found that the data has a unit root process. So the original hypothesis (δ=0) is accepted. And the original data sequence is non-stationary.

In addition, the ADF test should be used to test the stability of the first order difference sequence. The test results are as shown in fig. 4. The ADF statistic absolute value of the first order difference sequence is 85.96033. It is more than 1% level critical absolute value 3.431067. And its P-value (0.0001) is less than 0.05. These mean passing through the ADF test. The original hypothesis (δ=0) is not accepted. And the sequence of the first order difference is stationary [3].

Null Hypothesis: USD has a unit root
Exogenous: Constant
Lag Length: 0 (Automatic - based on SIC, maxlag=35)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-0.443863	0.8993
Test critical values:	1% level	-3.431067	
	5% level	-2.861741	
	10% level	-2.566919	

*MacKinnon (1996) one-sided p-values.

Fig. 3 The ADF of original data sequence

Null Hypothesis: DY_USD has a unit root
Exogenous: Constant
Lag Length: 0 (Automatic - based on SIC, maxlag=35)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-85.96033	0.0001
Test critical values:	1% level	-3.431067	
	5% level	-2.861741	
	10% level	-2.566919	

*MacKinnon (1996) one-sided p-values.

Fig. 4 The ADF of first order difference sequence

3.2 Autocorrelation & autocorrelation analysis

The autocorrelation coefficient of the original sequence in fig. 5 decreases very slowly. The original data sequence is not stationary. Form fig. 6, the trailing of autocorrelation diagram and partial autocorrelation diagram is very significant. And the P-values are all less than 0.05. The forecasting should be established by using the ARMA model. The P and Q values in the final model need to be determined by using the P-value test and the AIC minimum criterion.

Date: 01/17/18 Time: 19:11
Sample: 1/01/1990 1/15/2018
Included observations: 7315

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	1.000	1.000	7312.7	0.000
		2	0.999	0.005	14621.	0.000
		3	0.999	0.014	21925.	0.000
		4	0.999	0.005	29226.	0.000
		5	0.998	-0.007	36522.	0.000
		6	0.998	-0.009	43813.	0.000
		7	0.998	0.009	51101.	0.000
		8	0.997	0.008	58385.	0.000
		9	0.997	0.008	65664.	0.000
		10	0.996	-0.006	72940.	0.000
		11	0.996	-0.000	80212.	0.000
		12	0.996	0.023	87480.	0.000
		13	0.996	0.012	94744.	0.000
		14	0.995	-0.020	102004	0.000
		15	0.995	0.019	109261	0.000
		16	0.995	0.018	116514	0.000
		17	0.994	-0.009	123764	0.000
		18	0.994	0.005	131010	0.000
		19	0.994	-0.005	138253	0.000

Fig. 5 The diagram of original sequence

Date: 01/17/18 Time: 19:32
Sample: 1/01/1990 1/15/2018
Included observations: 7314

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	-0.005	-0.005	0.2077	0.649
		2	-0.023	-0.024	4.2476	0.120
		3	-0.000	-0.000	4.2476	0.236
		4	0.008	0.007	4.6907	0.321
		5	0.018	0.018	7.1128	0.212
		6	-0.016	-0.015	8.8919	0.180
		7	-0.013	-0.012	10.131	0.181
		8	-0.011	-0.012	10.968	0.203
		9	0.034	0.034	19.670	0.020
		10	-0.000	-0.000	19.670	0.033
		11	-0.043	-0.041	33.054	0.001
		12	-0.011	-0.011	33.917	0.001
		13	0.052	0.050	53.863	0.000
		14	-0.036	-0.038	63.292	0.000
		15	-0.035	-0.032	72.309	0.000
		16	0.017	0.018	74.500	0.000
		17	-0.002	-0.004	74.530	0.000
		18	0.011	0.008	75.394	0.000
		19	-0.015	-0.013	76.944	0.000

Fig. 6 The diagram of first order difference sequence

Form fig. 6, the autocorrelation coefficient is the 13 order tailing. And the partial autocorrelation coefficient is the 13 order tailing. The following ARMA models are selected [4,5].

Dependent Variable: D(USD)
Method: Least Squares
Date: 01/17/18 Time: 20:16
Sample (adjusted): 1/19/1990 1/12/2018
Included observations: 7301 after adjustments
Convergence achieved after 8 iterations
MA Backcast: 1/02/1990 1/18/1990

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.125755	0.118843	1.058157	0.2900
AR(13)	0.530897	0.150556	3.526234	0.0004
MA(13)	-0.485498	0.155340	-3.125387	0.0018

R-squared	0.003076	Mean dependent var	0.125106
Adjusted R-squared	0.002803	S.D. dependent var	9.254096
S.E. of regression	9.241118	Akaike info criterion	7.285614
Sum squared resid	623236.6	Schwarz criterion	7.288447
Log likelihood	-26593.13	Hannan-Quinn criter.	7.286588
F-statistic	11.25908	Durbin-Watson stat	2.005315
Prob(F-statistic)	0.000013		

Inverted AR Roots	.95	.84-.44i	.84+.44i	.54+.78i
	.54-.78i	.11-.95i	.11+.95i	-.34+.89i
	-.34-.89i	-.71+.63i	-.71-.63i	-.92-.23i
	-.92+.23i			
Inverted MA Roots	.95	.84-.44i	.84+.44i	.54+.78i
	.54-.78i	.11-.94i	.11+.94i	-.34-.88i
	-.34+.88i	-.71+.63i	-.71-.63i	-.92-.23i
	-.92+.23i			

Fig. 7 ARIMA (13,1,13)

Dependent Variable: D(USD)
Method: Least Squares
Date: 01/17/18 Time: 20:19
Sample (adjusted): 1/19/1990 1/12/2018
Included observations: 7301 after adjustments
Convergence achieved after 9 iterations
MA Backcast: 1/02/1990 1/18/1990

Variable	Coefficient	Std. Error	t-Statistic	Prob.
AR(13)	0.533878	0.149341	3.574888	0.0004
MA(13)	-0.488360	0.154147	-3.168143	0.0015

R-squared	0.002923	Mean dependent var	0.125106
Adjusted R-squared	0.002787	S.D. dependent var	9.254096
S.E. of regression	9.241194	Akaike info criterion	7.285493
Sum squared resid	623332.1	Schwarz criterion	7.287382
Log likelihood	-26593.69	Hannan-Quinn criter.	7.286143
Durbin-Watson stat	2.004992		

Inverted AR Roots	.95	.84-.44i	.84+.44i	.54+.78i
	.54-.78i	.11-.95i	.11+.95i	-.34-.89i
	-.34+.89i	-.71+.63i	-.71-.63i	-.93-.23i
	-.93+.23i			
Inverted MA Roots	.95	.84-.44i	.84+.44i	.54+.78i
	.54-.78i	.11-.94i	.11+.94i	-.34-.88i
	-.34+.88i	-.71+.63i	-.71-.63i	-.92-.23i
	-.92+.23i			

Fig. 8 ARIMA (13,1,13) model of removing constant C

The P-value of constant term C (0.2900) is not less than 0.05. The constant item C should be removed. And the ARIMA (13,1,13) model should be reestablished. After removing the constant C, the results of fig. 8 shows that the coefficient of the model is significant. And the AIC value becomes less. The new ARIMA (13,1,13) model is more accurate.

3.3 Residual Test

If the autocorrelation coefficient of the residual sequence is in the random interval, the residual sequence is white noise. If the residual sequence is white noise, that shows the useful information has been extracted. The test results of the residual sequence are as fig. 9. The P- values are more than 0.05. The residual sequence of model is white noise. It can be determined that the fitting effect of the model is good.

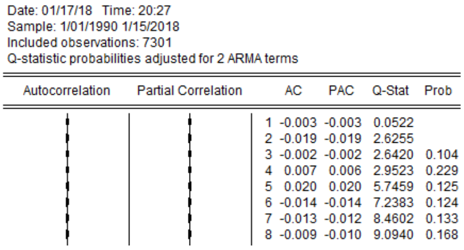

Fig. 9 Residual correlation diagram

3.4 Forecasting

Using the established model to forecast the gold price in the next 1 working days (2018-01-15). Because the dynamic forecast value is straight line and absolute percent error is 40.15768%, it shows that the effect of dynamic forecast is not good (fig. 10) [6]. The static forecasting can fit the gold price curve well. And its absolute percent error is only 0.660356%. The static forecasting model should be selected [7].

The static forecast value is 1337.345. And the actual value is 1339.25. The error rate of result is about 0.14224%. The forecasting result is as fig. 11.

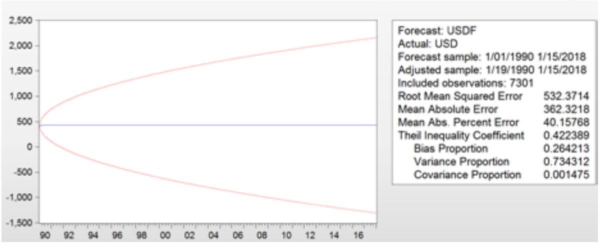

Fig. 10 The dynamic forecast graph

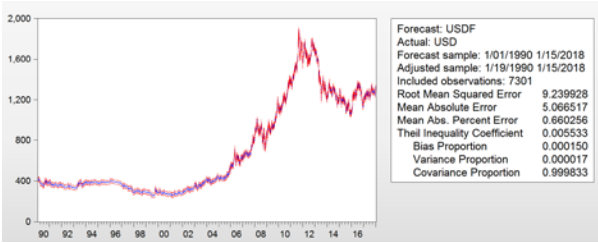

Fig. 11 The static forecast graph

4. Summary

As a crucial constituent of financial market, the gold trading market has been attracting much more interest by investors. When the gold market was established, there have been extensive researches on how to forecast the gold price. Based on the ARMA model of time series, the paper forecast the short-term gold price. The ARMA model has a good effect on the forecasting of non-stationary time series data. The establishment and solution of the ARMA model of gold price time series can provide reliable information service and decision guidance for investors.

Acknowledgments

This work was financially supported by Wenzhou vocational & technical college scientific research project fund(wzy2016019).

References

[1] Guiyang Xu, China gold spot price prediction model based on time series data analysis, China Securities and futures (2010) 12-13.

[2] Jun Deng, Xuan Yang, Wei Wang, an empirical study on stock price forecasting using ARMA model, Enterprise Guide (2010) 266-267.

[3] Men Kun, Empirical Analysis of Stock Price Based on ARMA Model, Journal of Hebei North University (2016)55-60.

[4] Jun Xu, Empirical Analysis of Gold Futures Price Based on ARMA Model, Industrial Economic Forum (2017)16-21.

[5] Ying Sun, Consumer price index prediction based on ARIMA model, Statistics & Decision

(2016)83-85

[6] Hao Ouyan, Prediction of real estate price index based on ARMA model, Statistics & Decision (2007)92-93

[7] Yongpe Hei, Improvement of ARMA model parameter estimation algorithm and its application in stock forecast, Journal of Chongqing Institute of Technology (2009)109-112