

Hybrid SARIMA-GARCH Model for Forecasting Indian Gold Price

¹Dileep Kumar Shetty, ²Sumithra & ³Ismail.B

^{1,2}Department of Statistics, Mangalore University, Mangalagangothri (India)

³Professor of Statistics, Department of Statistics, Mangalore University, Mangalagangothri (India)

ARTICLE DETAILS

Article History

Published Online: 07 August 2018

Keywords

GARCH; SARIMA; hybrid; accuracy; forecasting; volatility

Corresponding Author

Email: dileepshetty1991[at]gmail.com

ABSTRACT

A hybrid model has been considered an effective way to improve the forecast accuracy. This paper proposes the hybrid model of the linear seasonal autoregressive moving average (SARIMA) and the non-linear generalized autoregressive conditional heteroscedasticity (GARCH) in modeling and forecasting the Indian gold price. The goodness of fit of the model is measured using Akaike information criteria (AIC), while the forecasting performance is assessed using root mean square error (RMSE), mean absolute Error (MAE) and mean absolute percentage error (MAPE). The study concluded that SARIMA-GARCH is a more appropriate model forecasting Indian gold price. The analysis is carried out by using the R (3.2.1)-software.

1. Introduction

Gold is perhaps the most famous precious metal. Known for its use in jewelry and currency, it is unique for its durability, malleability and ability to conduct heat and electricity. In recent years, gold has become a key component in electronics manufacturing; it can be found in trace amounts in circuit boards and electrical connectors. The price of gold is determined by fluctuations in the commodities market based on supply and demand. Unlike many commodities, however, most of the gold ever mined still exists in an accessible form, such as bullion, scrap gold, and jewelry; as a result, the price of gold is mainly affected by changes in demand rather than changes in supply.

Autoregressive integrated moving average (ARIMA) models have been used for forecasting different types of time series to capture the long-term trend. In the case of financial time series that have been shown to have volatility clustering where large changes in the data tend to cluster together and resulting in persistence of the amplitudes of the changes, ARCH based models have been used. Ping [2013] the forecast the prices of Kijang Emas, the official Malaysian gold bullion. Two methods are considered, which are ARIMA and GARCH. The forecasting performance is measured using MAPE. The study concluded that GARCH is a most appropriate model. Ahmad [2014] proposed the hybrid ARIMA-GARCH model for modeling and forecasting Malaysian gold price. He found that ARIMA-GARCH model performed better than ARIMA model.

In this paper, we modeled and forecast the monthly selling price of 1 Troy ounce Indian gold using the hybrid SARIMA-GARCH model. Akaike information criterion (AIC) is used to assess the goodness of fit. RMSE, MAE and MAPE are used to evaluate the forecasting performances. All analyses are carried out using R(3.2.1) software.

The paper is organized into 4 sections. Section 2 presents the methodology of the study. Section 3 presents the data analysis. The study is concluded in Section 4.

2. Methodology

ARIMA model

ARIMA models are the most general class of models for forecasting a time series, applied in cases where data show evidence of non-stationary [Box(1970)]. Non-stationary in mean can be removed by transformations such as differencing, while non-stationary in variance can be removed by a proper variance stabilizing transformation introduced by Box and Cox.

The ARIMA (p, d, q) can be written as

$$\phi_p(B)(1-B)^d X_t = \theta_q(B)\varepsilon_t, \quad (1)$$

where $\phi_p(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p$ is the autoregressive operator of order p; $\theta_q(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q$ is the moving average operator of order q; $(1-B)^d$ is the d^{th} difference;

B is backward shift operator; and ε_t is the error term at time t. The orders are identified through the autocorrelation function (ACF) and the partial autocorrelation function (PACF) of the sample data. The error terms are generally assumed to be independent identically distributed random variables (i.i.d.) sampled from a normal distribution with zero mean and constant variance.

GARCH model

The autoregressive conditional heteroscedasticity (ARCH) class of models pioneered by Engle in 1982 and generalized by Bollerslev in 1986 are the popular class of econometric models for describing a series with time-varying conditional variance. The generalized autoregressive conditional heteroscedasticity (GARCH) family models were developed to capture volatility clustering and predict volatilities in the future [Tsay].

For a return series R_t , let $y_t = R_t - \mu_t$ be the innovation at time t. Then y_t follows a GARCH (m, s) model if

$$y_t = \sigma_t z_t, \quad \sigma_t^2 = \alpha_0 + \sum_{i=1}^m \alpha_i y_{t-i}^2 + \sum_{j=1}^s \beta_j \sigma_{t-j}^2$$

(2)

The sequence $\{z_t\}$ is a sequence of iid random variables with mean 0 and variance 1, $\alpha_0 > 0$, $\alpha_i \geq 0$, $i = 1, 2, \dots, m$, $\beta_j \geq$

0 , $j = 1, 2, \dots, s$, and $\sum_{i=1}^m \alpha_i + \sum_{j=1}^s \beta_j < 1$. The latter

constraint on $\alpha_i + \beta_j$ implies that the unconditional variance of y_t is finite, whereas its conditional variance σ_t^2 evolves over time. As mentioned before, z_t is more often assumed to follow a standard normal or standardized Student-t distribution or generalized error distribution [Tsay (2005)].

Estimation of Seasonal Component

Let $\{X_t, t=1,2, \dots, n\}$ be a time series, it contains both trend and seasonal component. It has the representation

$$X_t = m_t + s_t + \varepsilon_t \quad t = 1,2,\dots,n, \quad (3)$$

where $E(\varepsilon_t) = 0$, $s_t = s_{t+D}$, s_t is the seasonal component, m_t is the trend component and D is the seasonal period.

Step1: The trend is first estimated by applying a moving average filter, specially chosen to estimate the seasonal component.

$$\hat{m}_s = \left(\frac{1}{2} X_{t-q} + X_{t-q+1} + \dots + X_t + X_{t+1} + \dots + \frac{1}{2} X_{t+q} \right) / 2q, \quad (4)$$

where $q=D/2$.

Step2: Estimation of seasonal component for each $k=1, 2, \dots, D$

First compute the average W_k of the deviation $\{X_{k+jD} - \hat{m}_{k+jD}, \quad q < k + jD \leq n - q\}$

Next estimate the seasonal component \hat{s}_k as

$$\hat{s}_k = w_k - \frac{1}{D} \sum_{i=1}^D W_i, \quad k = 1,2,\dots,D \quad (5)$$

Step3: The de-seasonal data is

$$Y_t = X_t - \hat{s}_t, \quad t = 1,2,\dots,n \quad (6)$$

Hybrid SARIMA-GARCH model

The SARIMA-GARCH model is one in which the variance of the error term of the SARIMA model follows a GARCH process. The model can be written as:

$$\phi_p(B)\phi_p(B^S)(1-B)^d(1-B^S)^D y_t = \theta_q(B)\Theta_Q(B^S)\varepsilon_t, \quad \varepsilon_t = z_t \sigma_t, \quad (7)$$

$$\text{where } \sigma_t^2 = \omega + \sum_{i=1}^m \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^s \beta_j \sigma_{t-j}^2,$$

where y_t represents the time series;

$\phi_p(B^S) = 1 - \phi_1 B^S - \phi_2 B^{2S} - \dots - \phi_p B^{pS}$ is seasonal autoregressive part;

$\Theta_Q(B^S) = 1 - \Theta_1 B^S - \Theta_2 B^{2S} - \dots - \Theta_Q B^{QS}$ is

seasonal moving average part, S is the seasonal period; D is the seasonal difference; (m, s) is the order of GARCH process; ω, α_i and β_j parameters of GARCH

model; ε_t is the error term; σ_t^2 is the conditional variance of ε_t ; $\{z_t\}$ is the sequence of i.i.d random variables with mean zero and variance 1.

Akaike Information Criterion (AIC)

The goodness of fit of a model can be assessed using $AIC = 2k - 2\ln(L)$, where L = the maximized value of the likelihood function for the estimated model and k = the number of free and independent parameters in the model.

Mean Absolute Percentage Error (MAPE)

The model is evaluated based on its prediction errors. A successful model would give an accurate time-series forecast. The performance of the model is measured using the mean absolute percentage error (MAPE) which is defined as $(q+1) \leq t \leq (n-q)$

$$MAPE = \frac{1}{h} \sum_{t=n+1}^{n+h} \left| \frac{X_t - \hat{X}_t}{X_t} \right| \times 100$$

(8)

Where h is the number of points forecasted, X_t is the actual values and \hat{X}_t is the forecasted values from the period $t=n+1, n+2, \dots, n+h$.

3. Data Analysis and Result

The data used in the study is the monthly selling price of 1 Troy ounce Indian gold from the period 1st January 2000 to 1st December 2017 as plotted in Fig.1. The first 204 observations (Jan 1st 2000 to Dec 1st 2016) are used for parameter estimation and while the next 12 observations are used for out sample forecast evaluation. The dataset which we use for the analysis is collected from this website <https://www.investing.com/commodities>. It gives the price of precious metals in Dollars per 1 Troy ounce (1Troy ounce=31.1034768gms).

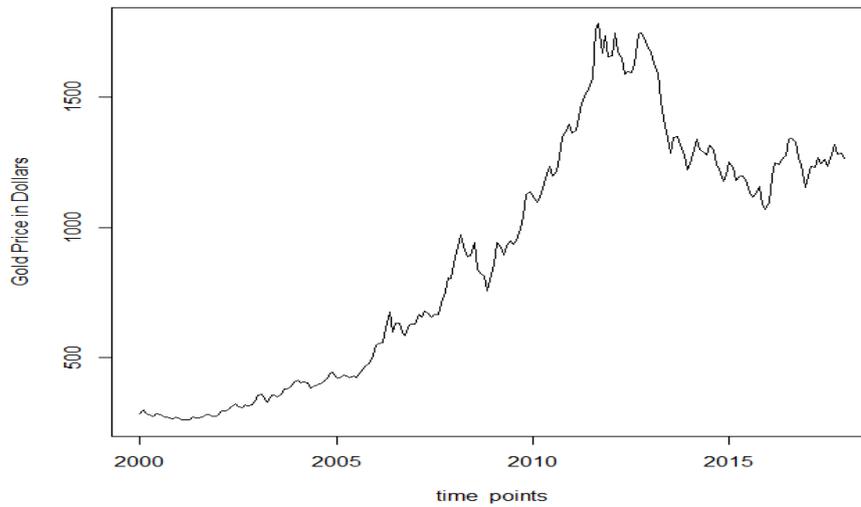


Fig.1: Monthly selling price of 1 Troy ounce Indian gold from the period 1st Jan 2000 to until 1st December 2016.

A trend and seasonality exist in the gold price data hence data is non-stationary. Fig.2 represents the decomposition of the observed time series. Decomposition of the observed series into three components, namely trend, seasonal and

random components. Here estimated random component is obtained by eliminating the estimated trend and seasonal components from observed time series.

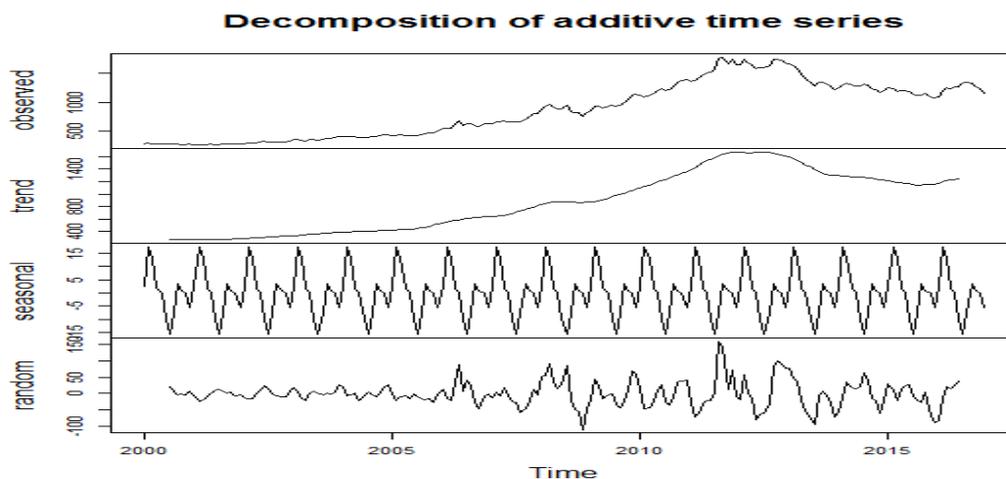


Fig.2: Decomposition of time series

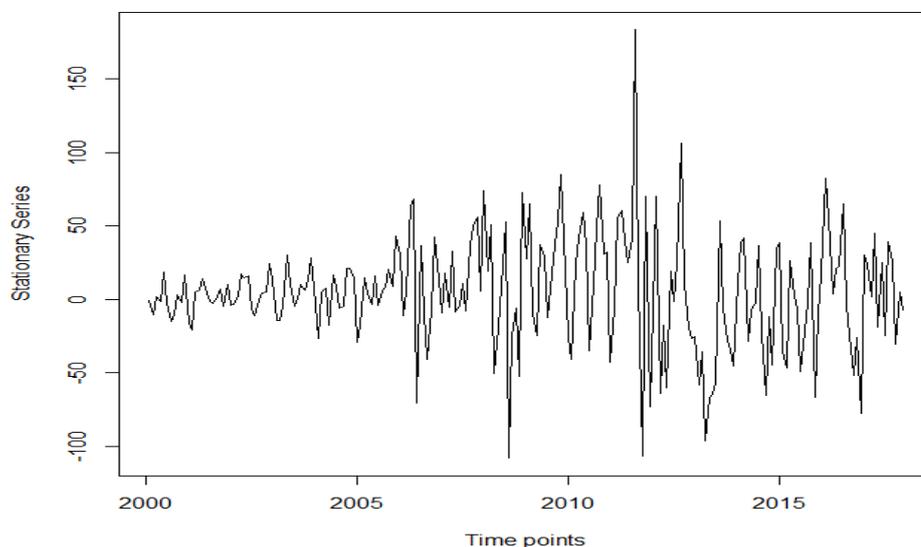


Fig.3: Stationary series

Fig.3 represents the stationary series. Here stationary series is obtained by eliminating the estimated seasonal component from the actual series. This series denoted as de-seasonal time series. Next, after one-time difference ($d=1$) of de-seasonal time series, we will get the stationary series.

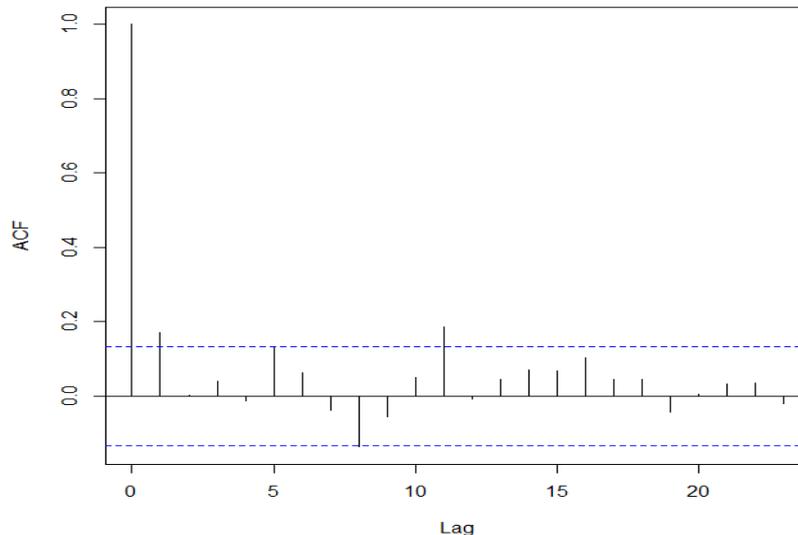


Fig.4: ACF for stationary series.

Fig.4 represents ACF for stationary. Here there are two autocorrelations lies outside the 2σ limits. Hence the maximum possible order of q and Q are 1.

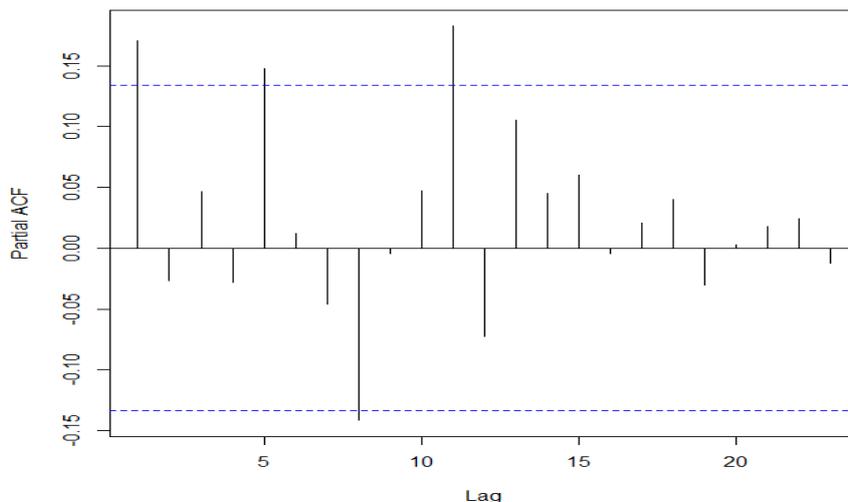


Fig.5: PACF for stationary series.

Fig.5 represents PACF for stationary. Here there are two PACF's are outside the outside the 2σ limits. Hence the maximum possible order for p and P are 1.

Table.1: SARIMA model fitting.

p	d	Q	P	D	Q	AIC	Ljung-Box p-value
0	1	1	1	0	1	2064.723	0.8357
1	1	0	1	0	1	2065.255	0.7885
1	1	1	1	0	1	2066.562	0.787

The most appropriate SARIMA model for the observed series is SARIMA (0, 1, 1) (1, 0, 1) with the minimum AIC and the highest p-value by diagnostic checking (See Table.1).

Table.2: Descriptive statistics for residuals.

Mean	3.445035
Median	1.498037

Maximum	181.4362
Minimum	-112.2679
Skewness	0.1525779
kurtosis	5.803976

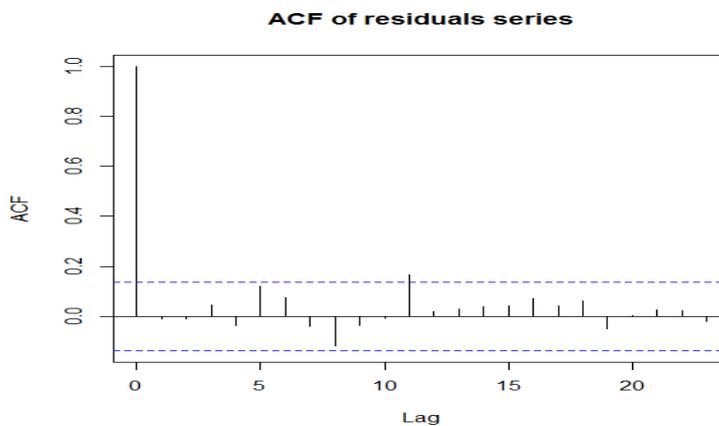


Fig.6: ACF of residuals series.

Fig.6 represents the ACF plot for residuals of fitted SARIMA (0,1,1)(1,0,1) model. The plot indicates that all the ACF's are inside the 2σ limits so residual series do not suffer from

serial correlation which suggests that fitted SARIMA model is appropriate for observed time series.

Table.3: ARCH effect test for residuals of the SARIMA model

Test	Statistics	p-value
Box-Pierce	19.375	0.9113
L-jung Box	20.747	0.8682
For squared residual series		
Box-Pierce	76.251	4.041e-06
L-Jung Box	80.244	1.067e-06

Thus, The results of Box-pierce and L-Jung Box test shows that the ARCH effect present in the residual series since the residuals are uncorrelated but squared residuals are suffering from serial correlation(see Table.3). Hence, it is

necessary to develop a better model for analysis of gold price. A GARCH model is proposed to handle heteroscedasticity in the series.

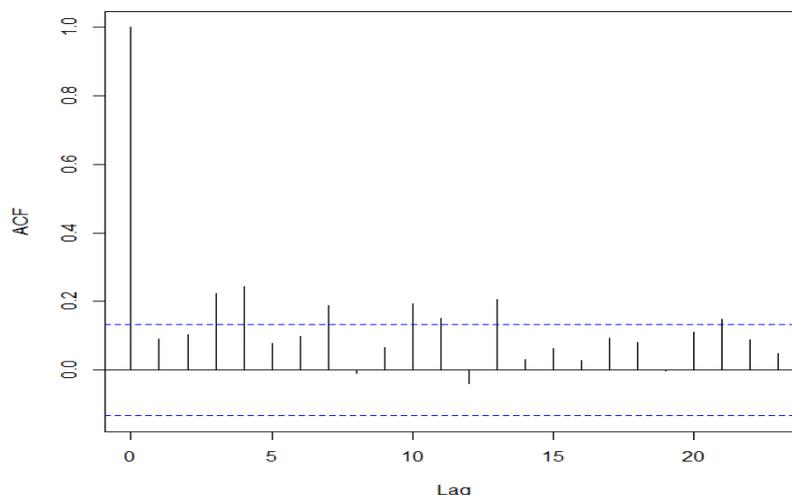


Fig.7: ACF for squared residuals of SARIMA (0, 1, 1) (1, 0, 1)

Fig.7 represents the ACF of the squared residuals of the fitted SARIMA model. It indicates that up to lag 4 autocorrelations are significant. Hence the maximum possible order for 's' is 4.

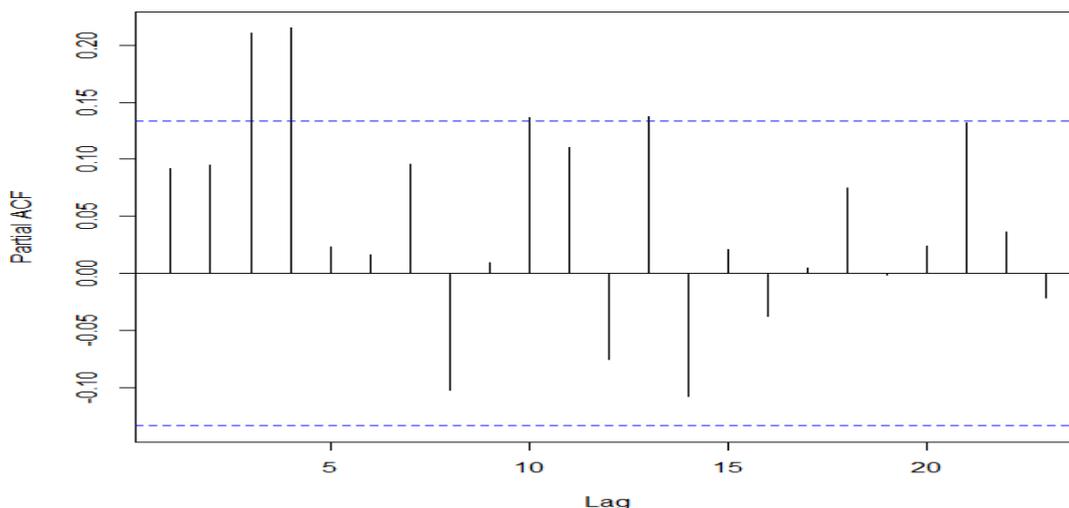


Fig.8: PACF for squared residuals of SARIMA(0,1,1)(1,0,1)

Fig.8 represents the PACF of the squared residuals of the fitted SARIMA model. It indicates that up to lag 4 PACF's are significant. Hence the maximum possible order for 'm' is 4.

Table.4: GARCH model fitting for residuals of SARIMA

Model	AIC	Significance
(1,0)	10.01647	α_1 is significant.
(1,1)	9.631851	α_1 & β_1 are significant.
(1,2)	9.653315	Both β_1 & β_2 are significant but α_1 is insignificant.
(2,0)	9.913615	Both α_1 & α_2 are significant.
(2,1)	9.638611	α_1 & α_2 are significant but β_1 is insignificant.
(3,0)	9.771105	α_1, α_2 & α_3 are significant.

From the Table.4, it is observed that the GARCH(1,1) model is appropriate since the AIC is minimum and also both α_1 & β_1 are significant.

Table.5: Fitted GARCH(1,1) model Residuals of SARIMA

Parameters	Estimate	Std. error	t-value	p-value
Omega	2.19739	4.94888	1.864	0.016570
Alpha	0.36293	0.09081	0.444	6.43e-05
Beta	0.73463	0.05190	14.154	< 2e-16

Table.6 : Heteroscedasticity test for residuals of SARIMA-GARCH model

Test	Statistics	p-value
Box-Pierce	21.398	0.8442
L-Jung Box	22.825	0.7845
For squared residual series		
Box -Pierce	13.032	0.9953
L-Jung Box	14.327	0.9896

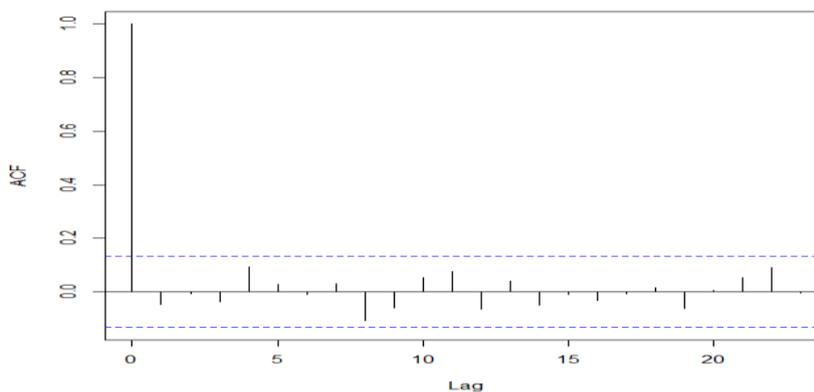


Fig. 9: ACF of squared residuals of SARIMA-GARCH.

The results of the Box-pierce and L-Jung Box test shows that ARCH effect absence in the residual series since residual and squared residuals of fitted SARIMA-GARCH model does not suffer from serial correlations (see Table 6).

However, the hybrid model is used for forecasting. The results of out-sample forecasting are presented in Fig.10 and Table.7.

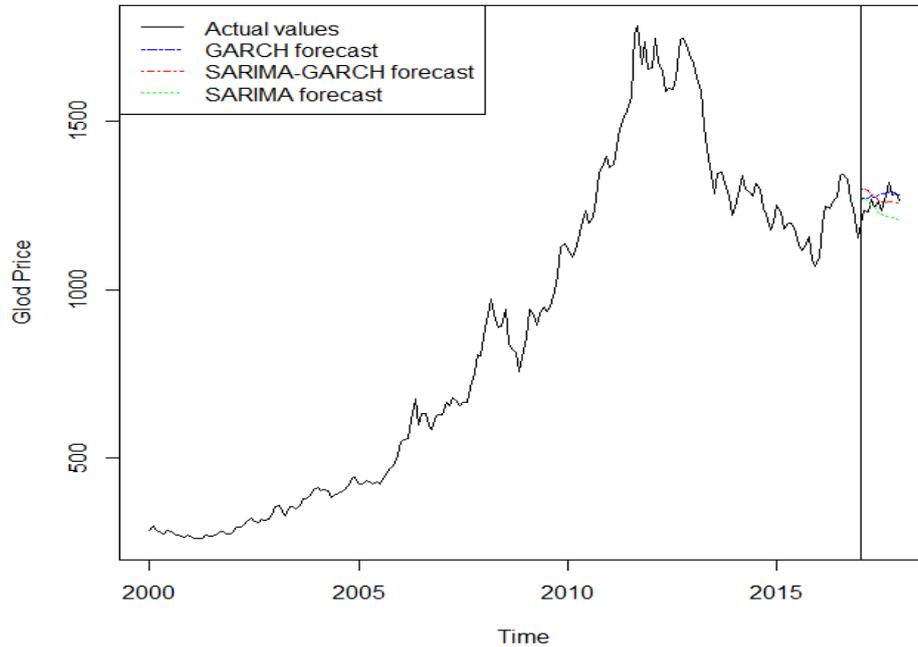


Fig.10: Forecasting Results.

4. Conclusion

Table 7 presents some results of modeling and forecasting of the monthly prices of 1 Troy ounce Indian gold recorded 1st January 2000 to 1st December 2017.

Table.7: Forecasting Results.

	GARCH	SARIMA	SARIMA-GARCH
MAPE	7.8532	10.79058	4.14353
MAE	99.4379	136.4942	58.6986
RMSE	103.552183	140.7048	53.0511

Three models were used, namely GARCH, SARIMA and SARIMA-GARCH. We compare the accuracy between the models based on error statistics such as MAPE, MAE and RMSE. Based on Table.7, the forecasts produced by SARIMA-GARCH are better since the RMSE, MAE and

MAPE are lower than those produced by ARIMA and GRACH. It can be concluded that in the case of the monthly selling prices of 1 Troy ounce Indian gold, the hybrid model of SARIMA-GARCH can be an effective way to improve the forecasting accuracy.

Reference

1. G.E.P. Box, G.M. Jenkins(1970). *Time Series Analysis, Forecasting and Control*. San Francisco:Holden-Day.
2. Maizah Hura Ahmad, Pung Yean Ping, Siti Roslindar Yaziz and Nor amizah Miswan(2014). A Hybrid Model for Improving Malaysian Gold Forecast Accuracy. *International Journal of Mathematical Analysis*. 8 (28):1377 – 1387.
3. Pung Yean Ping, Nor Hamizah Miswan and Maizah Hura Ahmad(2013). Forecasting Malaysian Gold Using GARCH Model. *Applied Mathematical Sciences*. 7(58): 2879-2884.
4. Ruey S. Tsay(2005). *Analysis of Financail Time Series*. Second Edition. U.K :John Wiley& Sons.
5. Robert H Shumway and David S Stoffer(1999). *Time Series analysis and its application with R examples*, 2nd Ed., Springer.
6. T.Bollerslev(1986). Generalized Autorregressive Conditional Heteroskedasticity. *Journal of Econometrics*. 31:307-327.
7. Robert F.Engle(1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*. 50(4):987-1007.