Journal of Physics: Conference Series

**PAPER • OPEN ACCESS**

# ARIMA and Symmetric GARCH-type Models in Forecasting Malaysia Gold Price

# ARIMA and Symmetric GARCH-type Models in Forecasting Malaysia Gold Price

**Siti Roslindar Yaziz[1*], Roslinazairimah Zakaria[1] and Suhartono[2]**

[1] Centre for Mathematical Sciences, Universiti Malaysia Pahang, 26300, Malaysia.

[2] Department of Statistics, Institut Teknologi Sepuluh Nopember, Surabaya 60111, Indonesia.
*Corresponding author: roslindar@ump.edu.my (S.R. Yaziz)

**Abstract.** Gold price modelling is crucial in gold price pattern determination since the information can be used for investors to enter and exit the market. The model selection is important and corresponds to the gold price movement characteristics. This study examines the forecasting performance of autoregressive integrated moving average (ARIMA) with symmetric generalised autoregressive conditional heteroscedastic (GARCH)-type models (standard GARCH, IGARCH and GARCH-M) under three types of innovations that are Gaussian, *t* and generalized error distributions to model gold price. The proposed models are employed to daily Malaysia gold price from year 2003 to 2014. The empirical results indicate that ARIMA(0,1,0) - standard GARCH(1,1) using *t* innovations is the most preferred ARIMA with symmetric GARCH-type model.

## 1. Introduction

Gold has been considered as a safe haven asset especially during economic shocks and in global economic instability events [1–2]. Modelling and forecasting gold prices are of great interest because accurately analysing the futures prices is an important component when to enter and exit the market. Therefore, it is vital to study on gold price forecast model that is able to predict the gold price that reflects the price movement.

ARIMA is continuously used as a practical gold price model either as the forecasting, benchmark or as the hybrid model [3–6]. However, based on previous studies, gold is noted as a volatile monetary asset commodity [7–8]. [9] stated that there is a strong positive trend associated with a higher volatility price within period of 2002 to 2011. Moreover, the current trend starting from year 2011 to early year 2014 seems similar to the correction trend of year 1980-1983 which have been highly volatile in that duration. Therefore, the ARIMA model is basically inappropriate to be applied to gold price, since it violates the errors assumption of constant variance. In general, for highly volatile data such as gold price, the variance for errors is non-constant and this characteristic is known as heteroscedasticity or autoregressive conditional heteroscedastic (ARCH) effects.

Hence, incorporating ARIMA with a heteroscedastic stochastic model is the widely used approach to overcome the weakness of ARIMA in handling ARCH effects. Previous studies have shown that the generalised autoregressive conditional heteroscedasticity (GARCH)-type model is widely applied in handling volatility in a data series, including in gold market [10–11]. The standard GARCH model or simply called as GARCH has been widely used at the first instance in modelling volatility in a data series. On the basis of this idea, the incorporation of GARCH into ARIMA in modelling and forecasting a highly volatile data, including gold price, has been considered in some empirical studies [5,6,8,12–16]. The standard GARCH model is symmetric in response to the past volatility. Hence, it motivates the investigation of the performance of other types of symmetric GARCH namely the integrated GARCH (IGARCH) and GARCH in-mean (GARCH-M) in handling heteroscedasticity in a data series. These univariate symmetric GARCH-type models are incorporated to ARIMA in investigating the performance of the proposed models in forecasting daily Malaysia gold price. The performance of the proposed models in modelling the data series is analysed with three types of innovations namely Gaussian, *t* and generalized error distributions (GED) that are commonly employed when working with GARCH model.

There are various forecast models for gold price such as neural network [17–18], fuzzy system, system dynamics [19], regression model [20], jump-and-dip diffusion [4], artificial intelligence, Chaos theory, heterogeneous agents model [9], model decomposition and independent component analysis [21], fractal analysis [22] or the hybrid of the above models [23–25]. To the best of our knowledge, although these studies achieve a certain effect in gold price forecasting, no literatures focuses on the performance of ARIMA with symmetric GARCH-type models. Therefore, we are in the opinion that modelling gold price using ARIMA-GARCH in presenting the most preferred symmetric GARCH-type model in handling volatility in the series can provide a significant contribution to the existing literatures.

## 2. Methodology

This section theoretically brief the concepts and methodology used.

### *2.1 The ARIMA model*

In general, the model of ARIMA($p$,$d$,$q$) is given in equation (1),

$$\varphi_p(B)(1-B)^d(y_t-\mu)=\theta_q(B)a_t \tag{1}$$

where $y_t$ and $a_t$ be the observed value and random error at time period $t$, respectively; with $\mu$ is the mean of the model, $B$ is the backward shift operator, $\varphi_p(B)=1-\sum_{i=1}^{p}\varphi_i B^i$ and $\theta_q(B)=1-\sum_{j=1}^{q}\theta_j B^j$ are polynomials in terms of $B$ of degree $p$ and $q$, $\nabla = (1-B)$, $\varphi_1,\varphi_2,...,\varphi_p$ are the autoregressive parameters with order $p$, $\theta_1,\theta_2,...,\theta_q$ are the moving average parameters with order $q$, and $d$ is the order of differencing. Random errors, $a_t$ are assumed to be zero-mean independently and identically distributed (iid) sequences with continuous distribution with mean zero and constant variance of $\sigma^2$ or can be written as $a_t \sim iid\left(0,\sigma^2\right)$.

### *2.2 The symmetric GARCH-type models*

For a univariate series, the mean model at time $t$ is given in equation (2), where $s_t$ and $a_t$ be the stationary data and random error at time period $t$, respectively; with $\mu_t$ is the conditional mean of $y_t$ and $a_t=\sigma_t\varepsilon_t$ where $\varepsilon_t$ is the standardised error. The $\varepsilon_t$ is assumed to be zero-mean iid sequences with continuous distribution. The term $a_t$ follows a standard GARCH ($r$,$s$) if the conditional variance of $y_t$ denoted by $\sigma_t^2$ is given in equation (3), where $\alpha_i$ and $\beta_i$ are the coefficient of the parameters ARCH and GARCH, respectively; with conditions of $\alpha_0>0,\ \alpha_i\geq 0,\ \beta_i\geq 0$.

$$s_t=\mu_t+a_t \tag{2}$$

$$\sigma_t^2=\alpha_0+\sum_{i=1}^{r}\alpha_i a_{t-i}^2+\sum_{i=1}^{s}\beta_i\sigma_{t-i}^2 \tag{3}$$

If the sum of ARCH and GARCH parameters in equation (3) equals to one, the model then is called an IGARCH($r$,$s$). The IGARCH models are unit-root GARCH models due to the existence of a unit root in the autoregressive part in the GARCH models. Thus, the IGARCH behaviors might be caused by occasional level shifts in volatility. The GARCH-M model can be written as equation (4), where $s_t$ and $a_t$ be the stationary data and random error at time period $t$, respectively and $c$ is the risk premium parameter. The GARCH-M model is used to model the phenomenon of a series that may depend on its volatility.

$$s_t = \mu + c\sigma_t^2 + a_t$$
$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{r} \alpha_i a_{t-i}^2 + \sum_{i=1}^{s} \beta_i \sigma_{t-i}^2 \tag{4}$$

*2.3 The ARIMA with symmetric GARCH-type models*

The ARIMA model with symmetric GARCH-type error is used to forecast gold price. The $a_t$ of the ARIMA follows a symmetric GARCH-type of orders $r$ and $s$ in the volatility model. In this study, the performance of the ARIMA with symmetric GARCH-type models are investigated with three types of innovations for $\varepsilon_t$. In order to obtain the best innovations of $\varepsilon_t$ for the proposed model, the distributions of Gaussian, $t$ and GED are used. Figure 1 shows the flowchart of the procedure.

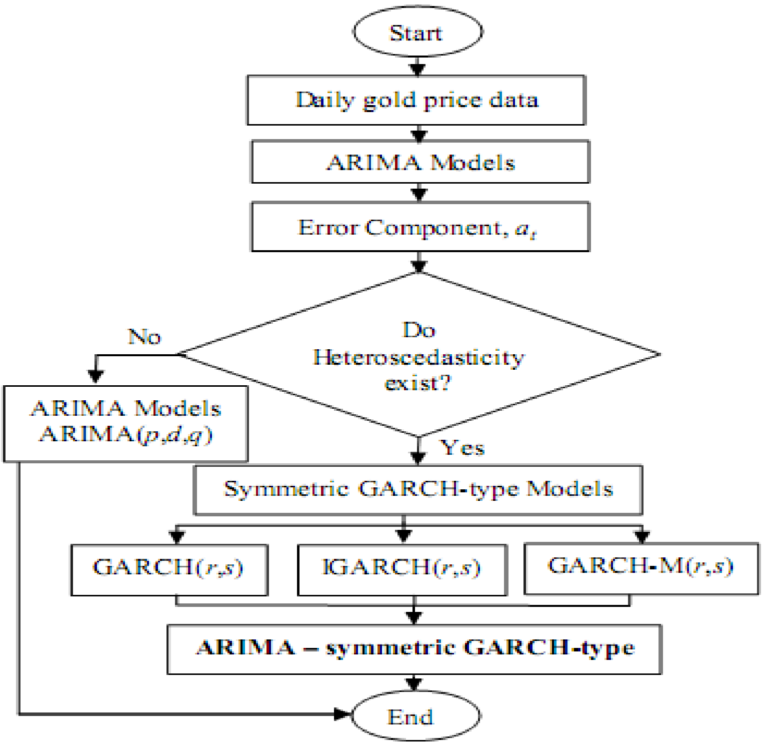


**Figure 1.** Procedure of data fitting using model of ARIMA - symmetric GARCH-type.

## 3. Empirical results using Malaysia gold price

*3.1 Data and stationarity*

A total of 2845 daily Malaysia gold price data starting from 2nd January 2003 to 12th June 2014 of 5-day-per-week is employed in this study. The data are quoted in Ringgit Malaysia per gram (RM/g). The Malaysia gold price is generated from world gold price via www.kitco.com based on London PM Fix, with the USD exchange rate at 12.00 noon from Bank Negara Malaysia. The estimation ratio (in-sample series) to forecast (out-of-sample series) is 90:10. Figure 2(a) is in-sample series graph, which shows that the price series does not vary around a fixed level, which indicates that the series is nonstationary both in-mean and in-variance, as well as exhibiting the trends of upward and non-seasonal.

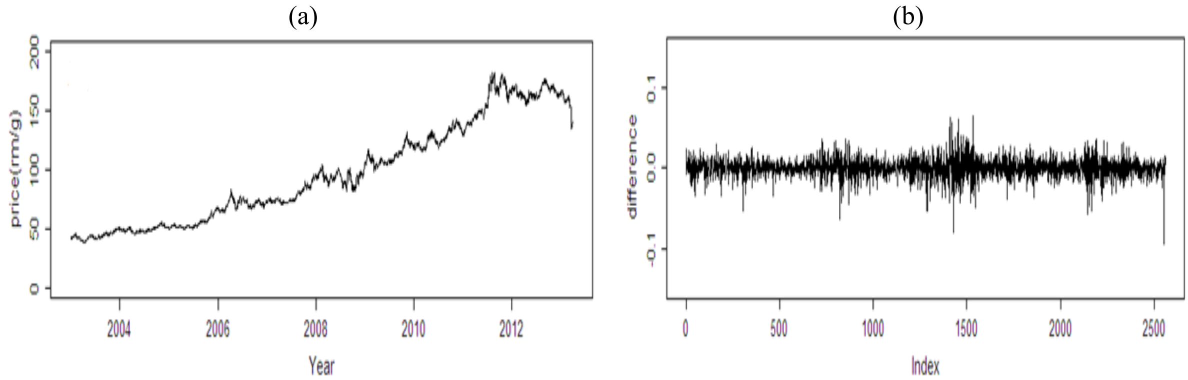


**Figure 2.** In-sample time plots of daily Malaysia gold price from 2nd January 2003 to 24th April 2013 (a) The daily data, (b) The stationary series.

In handling the nonstationary in-variance in the data series, the transformation of $y_t^* = \ln y_t$ using Box-Cox method is applied. The augmented Dickey-Fuller (ADF) test shows an insignificant result and autocorrelation function (ACF) spike demonstrates a gradual decay to zero which support the nonstationary in-mean in the log series. The ADF test for the first differenced of log series indicates no unit root, which means the series is stationary. The stationarity of the first differenced log Malaysia gold price is then supported by the ACF and partial autocorrelation function (PACF), which is suggest that Malaysia gold price exhibit random walk and this is consistent with the previous literatures of the world gold price [4,8]. Figure 2(b) illustrates the stationarity in-mean of the first differenced log gold price series since most of the data are located around the zero mean. However, there are some spikes which represents volatility clustering with high volumes specifically around 2008, 2011 and 2013 because of the instability of the global financial market [9,26].

### *3.2 ARIMA-GARCH modelling*

The daily Malaysia gold price series is nonstationary and non-seasonal. The spikes pattern of ACF and PACF of the stationary series as presented in figure 3 suggest that the possible values of both parameters $p$ and $q$ are 0,1 and 2, which produced three significant models at 5% significance level, namely ARIMA(0,1,0), ARIMA(1,1,1) and ARIMA(2,1,2). Since the values of Akaike information criterion (AIC) and Schwarz information criterion (SIC) are marginally decreased among the models, hence by applying the principle of parsimony, ARIMA(0,1,0) is preferred. However, the diagnostic test indicates that the ARIMA(0,1,0) fails in the test of heteroscedasticity. The ARCH lagrange multiplier (ARCH LM) test showed that there are strong ARCH effects in the data series. The PACF of the Ljung-Box test on squared residuals shows the insignificant results up to lag 17, hence the GARCH model is used in handling heteroscedasticity in the residuals. Therefore, various type of GARCH models can be tested to the residuals of ARIMA model including the symmetric GARCH-type namely standard GARCH, IGARCH and GARCH-M, by applying the procedure shown in figure 1.

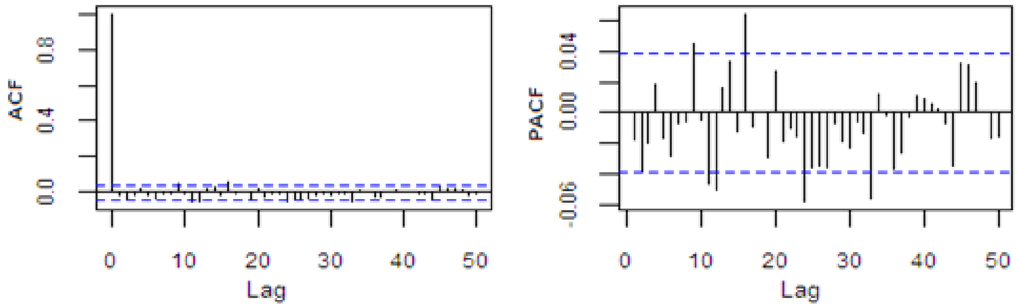


**Figure 3.** ACF and PACF of the stationary series for in-sample gold price

The PACF for the squared residuals ($a_t^2$) of ARIMA(0,1,0) as illustrated in figure 4 suggests that the possible values of both parameters $r$ and $s$ are 0, 1 and 2. The analysis conducted in the estimation stage using maximum likelihood estimation (MLE) observed that ARIMA(0,1,0) with the GARCH parameters of $r, s = 1$ showed statistically significant and preferred at 5% significance level for all symmetric GARCH models considered. Table 1 displays the estimation results for ARIMA(0,1,0)-GARCH(1,1), ARIMA(0,1,0)-IGARCH(1,1) and ARIMA(0,1,0)-GARCH(1,1)-M, respectively, with three types of innovations. Note that $\upsilon$ is the degrees of freedom for the innovations.

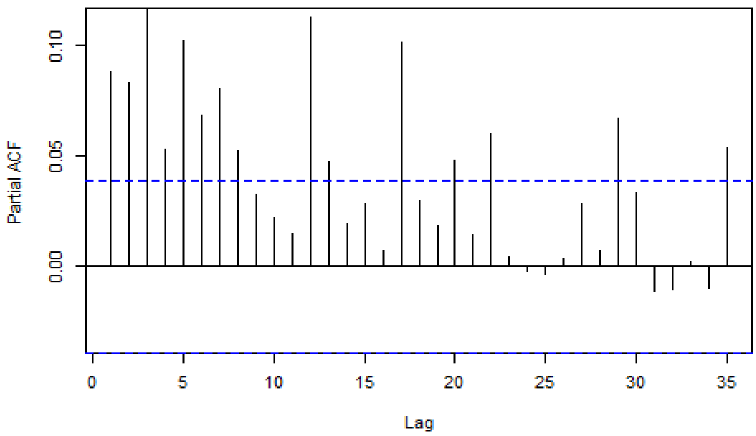


**Figure 4.** PACF of $a_t^2$ of ARIMA(0,1,0) for stationary series of in-sample gold price

It can be seen that the use of standard GARCH and IGARCH models are highly statistically significant and seemed to be justified for all types of innovations. The large values of $\beta_1$ in volatility models are reflected by the conditional standard deviation processes which demonstrate a relatively long-term persistence of volatility clustering. Furthermore, the sum of the ARCH and GARCH coefficients in the volatility model for ARIMA(0,1,0)-GARCH(1,1) that is very close to one, which is indicating that volatility shocks are quite persistent.

Meanwhile, ARIMA – GARCH-M showed insignificant results due to the estimated risk premium, $c$ values are negative and highly insignificant for all types of innovations. The values of $c$ implying that there are no serial correlations in the gold price stationary series, or in other words, although we hold extra risk for the asset, our return is indifferent with those who are not taking extra risk. Interestingly, the parameter estimates in the volatility model of the GARCH-M model are similar as standard GARCH. Therefore, all models with GARCH-M are not considered for the next stage.

Table 2 presents the joint diagnostic checking for the considered proposed models of ARIMA with the significant symmetric GARCH models. In the diagnostic test, the number of lag equals to ten is chosen for Ljung-Box test and LM ARCH test since the series is non-seasonal, as recommended by [27], to ensure that the number of lag is large enough to capture any meaningful and troublesome correlations.

The model diagnostic checking show that the all fitted models are adequate in describing the mean and volatility of the series, except for normality. Regarding on the innovations, the Jarque-Bera statistic strongly rejects the hypothesis of normal distribution for the standardised residuals for the series. Based on the table, the most negative values of AIC and SIC shows that the model of ARIMA with standard GARCH using *t* innovations is preferred.

In this context, the standardised residuals appear to be random, but the result of positive excess kurtosis indicate the characteristics of heavy tails, which support the *t* innovations. However, the *p*-value of the sample skewness for the ARIMA(0,1,0)-GARCH(1,1) with *t* innovations indicates that the series is negatively skewed. To model this skewness, a skewed *t* distribution is employed. Even though the model is significant and the SIC and AIC for the model with skewed-*t* innovations is smaller than *t*, but by applying the principle of parsimony, the model with *t* innovations is still preferred since the estimation results are marginally decreased between the models. The fitted volatility plots and QQ-plot for the models support this decision since it is hard to see any significant difference between the both innovations.

**Table 1**. Estimation results for ARIMA(0,1,0) with selected symmetric GARCH-type models.

| Model | Parameter | Gaussian | *t* | GED |
|---|---|---|---|---|
| **ARIMA(0,1,0)-GARCH(1,1)** | $\mu$ | 0.0004 | 0.0007 | 0.0006 |
| | | (0.0198) | (0.0001) | (0.0012) |
| | $\alpha_0$ | $1.8 \times 0^{-6}$ | $1.2 \times 10^{-6}$ | $1.3 \times 10^{-6}$ |
| | | (0.0000) | (0.0111) | (0.0095) |
| | $\alpha_1$ | 0.0599 | 0.0464 | 0.0503 |
| | | (0.0000) | (0.0000) | (0.0000) |
| | $\beta_1$ | 0.9302 | 0.9472 | 0.9421 |
| | | (0.0000) | (0.0000) | (0.0000) |
| | $\upsilon$ | - | 5.7636 | 1.2682 |
| | | | (0.0000) | (0.0000) |
| **ARIMA(0,1,0)-IGARCH(1,1)** | $\mu$ | 0.0004 | 0.0007 | 0.0006 |
| | | (0.0084) | (0.0001) | (0.0004) |
| | $\alpha_0$ | - | - | - |
| | $\alpha_1$ | 0.0511 | 0.0415 | 0.0442 |
| | | (0.0000) | (0.0000) | (0.0000) |
| | $\beta_1$ | 0.9489 | 0.9585 | 0.9557 |
| | | (0.0000) | (0.0000) | (0.0000) |
| | $\upsilon$ | - | 6.3785 | 1.2752 |
| | | | (0.0000) | (0.0000) |
| **ARIMA(0,1,0)-GARCH(1,1)-M** | $c$ | -1.6330 | -0.6462 | -1.0161 |
| | | (0.5760) | (0.8111) | (0.6943) |
| | $\mu$ | 0.0006 | 0.0008 | 0.0007 |
| | | (0.0963) | (0.0325) | (0.0429) |
| | $\alpha_0$ | $1.8 \times 10^{-6}$ | $1.2 \times 10^{-6}$ | $1.3 \times 10^{-6}$ |
| | | (0.0000) | (0.0111) | (0.0097) |
| | $\alpha_1$ | 0.0598 | 0.0464 | 0.0501 |
| | | (0.0000) | (0.0000) | (0.0000) |
| | $\beta_1$ | 0.9302 | 0.9473 | 0.9423 |
| | | (0.0000) | (0.0000) | (0.0000) |
| | $\upsilon$ | - | 5.7609 | 1.2682 |
| | | | (0.0000) | (0.0000) |

*Values in the parenthesis are *p*-values

Consequently, the model of ARIMA(0,1,0)-GARCH(1,1) with *t* innovations is given by equation (5) where $y_t$ is the daily gold price and $s_t$ is the stationary data.

$$y_t = y_{t-1}\exp(s_t), \quad s_t = 0.0007 + a_t, \quad a_t = \sigma_t \varepsilon_t, \quad \varepsilon_t \sim t^*_{5.76}$$
$$\sigma_t^2 = 1.2\times 10^{-6} + 0.0464 a_{t-1}^2 + 0.9472\sigma_{t-1}^2 \quad (5)$$

In the forecasting stage, the data of out-of-sample from 25th April 2013 to 12th June 2014 with a total of 284 daily gold prices is employed. The forecast evaluations using the price model are based on the basis of three evaluations criteria commonly used in the literatures that are the root mean square error (RMSE), the mean absolute error (MAE) and the mean absolute percentage error (MAPE). The RMSE, MAE and MAPE for one-step ahead forecast for the model of ARIMA(0,1,0)-GARCH(1,1) with *t* innovations are 1.2150, 1.6774 and 0.0090%, respectively. The promising performance of the proposed model in forecasting Malaysia daily gold price graphically showed by figure 5, where the trend of forecast prices mimics the actual data for the out-of-sample period. The comparison between actual and forecast price using the model of ARIMA(0,1,0)-GARCH(1,1) for the simulation last five-day out-sample period is given by table 3.

**Table 2**. Model diagnostics for ARIMA(0,1,0) with significant symmetric GARCH-type models

| Model | | Gaussian | *t* | GED |
|---|---|---|---|---|
| **ARIMA(0,1,0)-GARCH(1,1)** | $Q(10)$ | 9.6058 | 9.9229 | 9.7694 |
| | | (0.4760) | (0.4470) | (0.4610) |
| | $Q^2(10)$ | 8.4691 | 8.9797 | 8.7476 |
| | | (0.5830) | (0.5340) | (0.5560) |
| | ARCH(10) | 1.0488 | 1.1613 | 1.1173 |
| | | (0.3992) | (0.3124) | (0.3447) |
| | JB | 1845.3 | - | - |
| | | (0.0000) | | |
| | AIC | -15647 | -15815 | -15811 |
| | SIC | -15624 | -15786 | -15782 |
| **ARIMA(0,1,0)-IGARCH(1,1)** | $Q(10)$ | 8.5658 | 9.1115 | 8.9304 |
| | | (0.5740) | (0.5220) | (0.5390) |
| | $Q^2(10)$ | 11.4130 | 11.8590 | 11.6700 |
| | | (0.3260) | (0.2950) | (0.3080) |
| | ARCH(10) | 1.5151 | 1.6122 | 1.5778 |
| | | (0.1274) | (0.0969) | (0.1069) |
| | JB | 2736.6 | - | - |
| | | (0.0000) | | |
| | AIC | -15608 | -15802 | -15796 |
| | SIC | -15596 | -15785 | -15779 |

*Values in bracket denote *p*-values. $Q(10)$ is the Ljung-Box statistics for standardised residuals at lag 10, $Q^2(10)$ is the Ljung-Box statistics for squared standardised residuals at lag 10, ARCH(10) is ARCH LM test at lag 10, and JB is the Jarque-Bera test for normality.

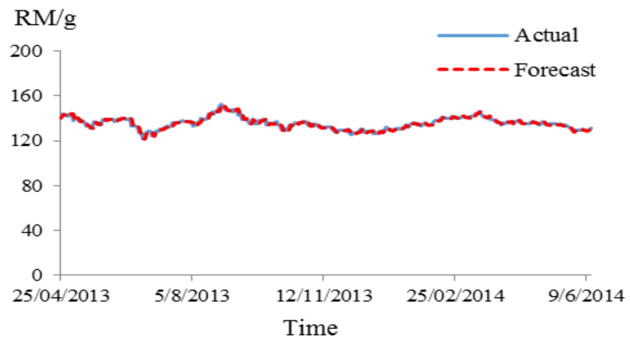


**Figure 5.** Plot of the actual and forecast using ARIMA(0,1,0)-GARCH(1,1) with *t* innovations.

**Table 3.** The comparison between actual and forecast price for out-of-sample period

| Date | Actual price (RM/g) | Forecast price (RM/g) |
|---|---|---|
| 6 June 2014 | 129.23 | 130.28 |
| 9 June 2014 | 128.86 | 129.32 |
| 10 June 2014 | 129.66 | 128.95 |
| 11 June 2014 | 129.98 | 129.75 |
| 12 Jun 2014 | 130.79 | 130.07 |

## 4. Conclusion

This study investigates the performance of ARIMA with symmetric GARCH-type models namely standard GARCH, IGARCH and GARCH-M to the daily Malaysia gold price. The empirical results indicate that the standard GARCH with *t* innovations is the most preferred symmetric GARCH model in handling heteroscedasticity in the series. The use of *t* innovations indicates that Malaysia gold price contains more extreme values and more mass on the tails. As conclusion, ARIMA(0,1,0)-GARCH(1,1) with *t* innovations is the preferred model of ARIMA-symmetric GARCH-type in forecasting daily Malaysia gold price.

**Acknowledgement**

This work is supported by Universiti Malaysia Pahang Internal Grants (RDU1703198).

## 5. References

[1] Laurence E Blose 2010 *J. Econ. Bus.* **62** 35
[2] Goodness C Aye, Hector C, Luis G-A and Rangan G 2017 *Resour. Policy* **54** 53
[3] Massarrat A K 2013 *Int. J. Emerg. Technol. Adv. Eng.* **3** 662
[4] Shahriar S and Erkan T 2010 *Resour. Policy* **35** 178
[5] Siti R Y, Noor A A, Roslinazairimah Z and Maizah H A 2013 *Proceedings of The 20th International Congress on Modelling and Simulation* 1201
[6] Siti R Y, Noor A A, Maizah H A, Roslinazairimah Z, Manju A and John B 2015 *AIP Conference proceedings* **1643** 289
[7] Olaoluwa S Y, Mohammed M T and Christopher G U 2016 *Resour. Policy* **49** 273
[8] Siti R Y, Roslinazairimah Z, Maizah H A 2017 *J. Phys. Conf. Ser.* **890** 012161
[9] Dirk G B and Kristoffer J G 2014 *J. Econ. Dyn. Control.* **40** 116
[10] Sohail C, Shahid K and Imran A 2012 *World Appl. Sci. J.* **19** 77
[11] Stefan T and Kevin L 2012 *Int. J. Bank. Finance* **9** 48
[12] Maizah H A, Pung Y P, Siti R Y, Nor H M 2014 *Int. J. Math. Anal.* **8** 1377
[13] Narendra B C and Eswara R B 2015 *Appl. Comput. Inform* **11** 130
[14] Girish G P 2016 *Energy Strategy Rev.* **11** 52
[15] Heping L and Jing S 2013 *Energy Econ.* **37** 152
[16] Tian S A L and Jia L N 2018 *Appl. Energy* **212** 244
[17] Guixia Y 2012 *J. Converg. Inf. Technol.* **7** 558
[18] Shifei Z, Kin K L and Jerome Y 2012 A *Expert Syst. Appl.* **39** 6168
[19] Wipawee T, Haruetai L and Pathompol V 2012 *Eng. J.* **16** 57
[20] Qi Z, Jun M and Yan W 2011 *Key Engineering Materials* **467** 1398
[21] E J, Le S and Ye J 2017 *Acta Phys. Pol. B.* **48** 2093
[22] Binghui W and Tingting D 2017 *Phys. Stat. Mech. Its Appl.* **474** 99
[23] Shuang G and Aiwu Z 2017 *Symmetry* **9** 207
[24] Shahrokh A, Akbar T and Seyed R H 2012 *Expert Syst. Appl.* **39** 5332
[25] Mehdi K, Seyed R H and Mehdi B 2008 *Fuzzy Sets Syst.* **159** 769
[26] Christian P, Marian R and Sebastian R 2014 *Int. Rev. Financ. Anal.* **32** 95
[27] Rob J H and George A 2014 *Forecasting: principles and practice* OTexts