Applied Mathematical Sciences, Vol. 10, 2016, no. 28, 1391 - 1402
HIKARI Ltd, www.m-hikari.com
http://dx.doi.org/10.12988/ams.2016.511716

Modelling Gold Price using ARIMA – TGARCH

Siti Roslindar Yaziz[1], Noor Azlinna Azizan[2], Maizah Hura Ahmad[3] and Roslinazairimah Zakaria[4]

[1,4]Faculty of Industrial Sciences & Technology, [2]Faculty of Technology
Universiti Malaysia Pahang
Malaysia

[3]Department of Mathematical Sciences, Faculty of Science
Universiti Teknologi Malaysia, Malaysia

Abstract

Statistical models can be used to characterize numerical data so as to understand its behavior and pattern. Gold price model, for example, can give signals to investors as to when they should enter and/or exit the market. To find an appropriate gold price model, it is crucial to choose a model that reflects the pattern of the price movement so as to make the model fit and adequate. This study examines the performances of ARIMA-TGARCH with five innovations in modeling and forecasting gold prices. The innovations considered include Gaussian, Student's-*t*, skewed Student's-*t*, generalized error distribution and skewed generalized error distribution. Using daily gold price data from the years 2003 to 2014, this study concluded that a hybrid ARIMA(0,1,0)-TGARCH(1,1) with *t*-innovation was the best model due to the existence of leverage effect and heavier tail characteristics in the data.

Keywords: Gold price forecasting, ARIMA, TGARCH, Innovations

1 Introduction

Since the year 2000, historical gold price data has significantly increased by about 550% in 10 years. Since 1933, currency crises and inflation in many countries such as the Great Depression in 1993, peso Mexico crisis in 1995, ASEAN crisis in the years 1997-98, Russian Rubel crisis in 1998, and the global economic crisis in the years 2008-2010 all showed that gold reserve has been used as a hedging tool against inflation [1-2]. A significant negative relationship between inflation and the gold price over the last 40 years has also been reported [3]. Hence, since gold price movement is very vital to investors, it is necessary to develop a model that reflects the pattern of the gold price.

The autoregressive integrated moving average (ARIMA) is one of the Box-Jenkins model that is widely applied in research practice for gold price modeling and forecasting, either as a benchmark, comparison, hybrid or forecasting models [4-6][3]. However, a recent study in gold price reported that there is a strong positive trend from 2002 to 2011 that is associated with a higher volatility in that period [7]. It is thus deemed appropriate to investigate the performance of hybrid ARIMA with volatility model since the ARIMA model alone is unable to handle the volatility that exist in the data series. Previous studies showed that the type of generalized autoregressive conditional heteroscedasticity (GARCH) models are widely applied to handle gold price volatility [8-10]. The hybrid model that combines the powerful of ARIMA-GARCH showed promising approach in modeling and forecasting daily gold price [11-13].

There is a probability of the existence of significant leverage effect in certain financial and commodity time series such as gold price. The standard GARCH models are unable to model the leverage effect because the conditional standard deviation, σ_t as a function of past values of random errors, a_t in the form of a_t^2, does not consider either the positive or negative past values of a_t. However, there is a model that is commonly used to handle leverage effects, which is the threshold generalized autoregressive conditional heteroscedastic (TGARCH) model [14-15]. Some recent studies also showed that the class of TGARCH models provide the best results in modelling daily volatility of gold market [13][15].

The current study investigates the performance of ARIMA-TGARCH model, while incorporating Box-Cox transformation in analyzing and forecasting daily gold price. The performance of the hybrid ARIMA-TGARCH model is analyzed using five types of innovations that include Gaussian, Student's-*t*, skewed Student's-*t*, generalized error distribution (GED) and skewed generalized error distribution (SGED).

2 Methodology

The following are the basic concepts of the model used in the study.

ARIMA Model

Let y_t and a_t be the observed value and random error at time period *t*, respectively; δ is the standard deviation of y_t, μ is the mean of the model,

$\varphi_1, \varphi_2, ..., \varphi_p$ 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 independently and identically distributed with a zero mean and a constant variance of σ^2. The general form for ARIMA(*p*,*d*,*q*) as the model of Box-Jenkins that handles the non-stationary time series with non-seasonal characteristics is given in Eq. (1),

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

where $\varphi_p(B) = 1 - \sum_{i=1}^{p} \varphi_i B^i$, $\theta_q(B) = 1 - \sum_{j=1}^{q} \theta_j B^j$ are polynomials in terms of *B* of degree *p* and *q*, $\nabla = (1-B)$, and *B* is the backward shift operator.

TGARCH Model

The univariate TGARCH is applied in constructing a hybrid model with ARIMA in modeling daily gold price. For a univariate series, let $y_t = \mu_t + a_t$ be a mean equation at time t, where μ_t is conditional mean of y_t and $a_t = \sigma_t \varepsilon_t$ where $\varepsilon_t \sim iid\ N(0,1)$. Then a_t follows a TGARCH (*r*,*s*) if

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

where N_{t-i} is an indicator for negative a_{t-i}, that is

$$N_{t-i} = \begin{cases} 1 & \text{if } a_{t-i} < 0, \quad \text{for } i = 1,2,...,r \\ 0 & \text{if } a_{t-i} \geq 0, \end{cases}$$

and σ_t^2 is the conditional variance of y_t, $\alpha_0 > 0, \alpha_i \geq 0, \beta_i \geq 0$, $\sum_{i=1}^{\max(r,s)} (\alpha_i + \beta_i) < 1$,

and α_i and β_i are the coefficient of the parameters ARCH and GARCH, respectively. Noted that, γ_i signifies the leverage effect of a_{t-i}. It is expected that γ_i is to be negative in real applications [16]. This TGARCH model is also called the GJR model since Glosten et al. proposed essentially the same model [14].

Hybrid ARIMA – TGARCH

In this hybrid model, an ARIMA model with TGARCH error components is applied to analyze and forecast the univariate series. The error term a_t of the ARIMA model is said to follow a TGARCH process of orders *r* and *s* with the leverage effect, γ_i , considered in the volatility model. The hybrid ARIMA-TGARCH methodology includes four iterative steps, namely, model identification, parameter estimation, diagnostic checking and forecasting. In this study, the performance of the hybrid ARIMA-TGARCH is investigated with five types of innovations for the standardized error ε_t in order to find the most appropriate innovation. The innovations in this study are Gaussian, Student's-*t*, skewed Student's-*t*, GED and SGED distributions. The flowchart of this procedure is shown in Figure 1.

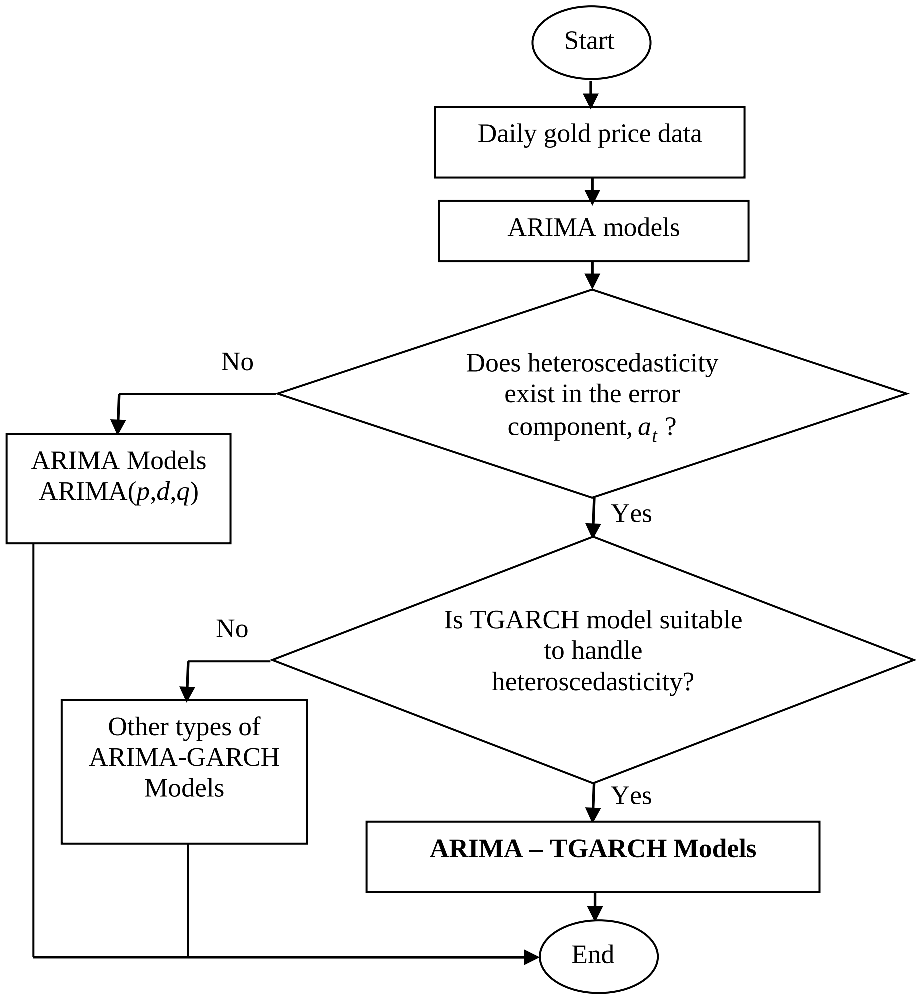

Figure 1: Procedure for Fitting an ARIMA-TGARCH Model

With reference to Figure 1, the following are the steps in selecting the best ARIMA-TGARCH model:

Step 1: Check the stationarity of the data series. Choose the possible values for parameter p,d,q of ARIMA model from autocorrelation function (ACF) and partial autocorrelation function (PACF) of the stationary series.

Step 2: In the diagnostic checking stage, perform the heteroscedasticity test by using ARCH LM test to the best ARIMA(p,d,q) model selected. If there are strong

ARCH effects in the error component (residuals), the PACF of the Ljung-Box test for the squared residuals will suggest either to choose ARCH or GARCH-type model.

Step 3: If the PACF of the squared residuals suggests GARCH-type model, then TGARCH (r,s) can be applied to the volatility model. The possible values for parameters r and s for the TGARCH model can be chosen based on ACF and PACF of the squared residuals of the best ARIMA(p,d,q) model.

Step 4: The significant and the best model of ARIMA(p,d,q)-TGARCH(r,s) is then tested under the assumptions of the distribution for ε_t, namely Gaussian, Student's-*t*, skewed Student's-*t*, GED and skewed GED, in finding the best innovation for the selected ARIMA-TGARCH model.

3 Data Analysis

A total of 2845 daily world gold price data series was used in the current study. The data dated from 2nd January 2003 to 12th June 2014 of 5-day-per-week, with some missing prices due to holiday and stock market closing day. The ratio of 90:10 was used for in-sample and out-of-sample periods. The values were quoted in US dollars per ounce and the data was obtained from www.kitco.com based on London PM Fix. Based on the plotting of the in-sample series in Fig. 2(a), it is shown that the daily gold price series did not vary around a fixed level which indicates that the series was non-stationary in both mean and variance, exhibiting an overall upward and non-seasonal trend. Using Box-Cox transformation method, $y_t^* = \ln y_t$ was the appropriate transformation to handle non-stationarity in the variance of the data series. Since trend still exists in the logarithm series, supported by the result of Augmented Dickey-Fuller (ADF) test and ACF spikes, the data was differenced to handle the non-stationarity in mean. The ADF test for the first order difference indicates no unit root, which means the series is stationary. Figure 2(b) illustrates the stationarity of the first differenced logarithm gold price series since most of the data are located around the zero mean. However, there were some spikes which represents volatility clustering specifically around 2008, 2011 and 2013 due to the U.S. subprime mortgage crisis, Lehman Brothers collapsed, the European sovereign debt and banking crisis, which affected the global financial market [7][17-18]. The pattern of ACF and PACF for the difference series indicates random walk. The characteristics of these non-stationarity and random walk in the gold price are consistent with previous studies [3] [19].

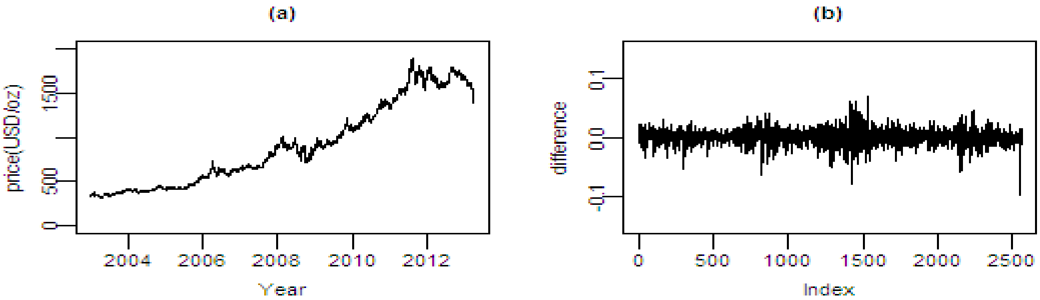

Figure 2: Time plot for in-sample series from 2nd January 2003 to 12th June 2014

(a) In-sample series of daily gold price

(b) The first differenced series of transformed daily gold price

4 ARIMA-TGARCH Modeling

The stationary series was first modeled using ARIMA model. The ACF and PACF for the stationary series suggest the values for both ARIMA parameters were 0,1 and 2. Using ordinary least square (OLS) to estimate the parameters, ARIMA(0,1,0), ARIMA(1,1,1) and ARIMA(2,1,2) were found to be significant at $\alpha = 0.05$. Even though Akaike information criterion (AIC) value for ARIMA(2,1,2) was the smallest, since the AIC values are marginally decreased between the significant models, ARIMA(0,1,0) was preferred due to parsimonious principle. This is consistent with the result of simplified Extended Autocorrelation Function (EACF) as shown in Table 1.

Table 1: The simplified Extended Autocorrelation Function (EACF) Table for the Differenced Logarithm Series

	MA Order:*q*												
p	0	1	2	3	4	5	6	7	8	9	10	11	12
0	O	O	O	O	O	X	O	O	X	O	X	O	O
1	X	O	O	O	O	X	O	O	O	O	O	X	O
2	X	X	O	O	O	X	O	O	O	O	O	O	O
3	X	X	X	O	O	X	O	O	O	O	O	O	O
4	X	X	X	X	O	X	O	O	O	O	O	O	O
5	X	X	X	X	X	X	O	O	O	O	O	O	O
6	O	X	X	X	X	X	O	O	O	O	O	O	O

The model checking indicated that ARIMA(0,1,0) failed the test of heteroscedasticity. Related tests also showed that there was strong ARCH effect in the data series. The partial autocorrelation function (PACF) of $\hat{a}_t^2$ for the test showed the insignificant results up to lag 17. Due to parsimony approach, GARCH model was used to handle the existence of heteroscedasticity in the residuals. The suggested volatility model was TGARCH that was hybridized with ARIMA to develop ARIMA-TGARCH model. The correlogram patterns of the

ACF and PACF for the squared residuals of the ARIMA(0,1,0) suggested the values for both parameters *r* and *s* to be 0,1 and 2. From the analysis conducted in the estimation stage using MLE, the hybrid model of ARIMA(0,1,0)-TGARCH(1,1), ARIMA(0,1,0)-TGARCH(1,2) and ARIMA(0,1,0)-TGARCH(2,1) were significant at 5% level. However, TGARCH(1,1) was preferred as the volatility model because of parsimonious principle when compared to other significant TGARCH(*r*,*s*) models.

Table 2 displays the estimation results for the parameters of ARIMA(0,1,0)-TGARCH(1,1). The coefficient of the mean equation for all types of innovations is highly significant. For the volatility equation, all estimates are statistically significant except for the leverage effect, γ_1. The test of leverage effect on the hybrid model showed that only the model with *t*-innovations and skewed *t*-innovations are significant at the 5% level. However, by applying the principle of parsimony, the hybrid model with *t*-innovations is preferred since the estimation results of the AIC are marginally decreased compared to the model with skewed *t*-innovations.

Table 2: ARIMA(0,1,0)-TGARCH(1,1) Estimation Results

	Conditional distributions				
	Gaussian	*t*	Skewed *t*	GED	SGED
μ	0.0006	0.0009	0.0007	0.0008	0.0005
	(0.0028)	(0.0000)	(0.0008)	(0.0000)	(0.0089)
α_0	$1.96\text{x}10^{-6}$	$9.40\text{x}10^{-7}$	$9.23\text{x}10^{-7}$	$1.14\text{x}10^{-6}$	$1.11\text{x}10^{-6}$
	(0.0000)	(0.0113)	(0.0114)	(0.0076)	(0.0077)
α_1	0.0532	0.0410	0.0404	0.0433	0.0426
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
γ_1	-0.0544	-0.1858	-0.1859	-0.1375	-0.1314
	(0.2872)	(0.0158)	(0.0151)	(0.0623)	(0.0709)
β_1	0.9357	0.9544	0.9545	0.9504	0.9507
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
υ	-	5.510	5.730	1.215	1.238
		(0.0000)	(0.0000)	(0.0000)	(0.0000)
ξ	-	-	0.9338	-	0.9520
			(0.0000)		(0.0000)
AIC	-15576.07 12	-15779.43 25	-15784.3368	-15779.44 53	-15782.60 30

P-values of the parameters are given in parentheses.

The model diagnostic checking of ARIMA(0,1,0)-TGARCH(1,1) with *t* innovations as shown in Figure 3 indicates that the model is adequate in modeling the data. The model has superior performance in describing the mean and variance equations of the series as shown by the Ljung-Box statistics for the standardized residuals and the squared standardized residuals which are all non-significant at 5% level. The standardized residuals appear to be random, but their magnitudes exhibit the characteristics of heavy tails, which support the *t*-innovations. This is proven by the rejected null hypothesis ($H_0 : k - 3 = 0$ (no excess kurtosis) in the stationary series. Since the excess kurtosis is significant and positive with a value of 1.5887, this indicates that the differenced log daily gold price exhibit heavy tails distribution. This implies that the distribution of log returns of gold price puts more mass on the tails and contains more extreme values, which is said to be leptokurtic.

Furthermore, the good fit of the QQ-plot in Figure 3 that nearly a straight line except for six outliers in the left tail, support graphically the use of *t*-innovations. The in-sample size is 2561, so the outliers are a very small fraction of the data. Consequently, the equation of the model ARIMA(0,1,0)-TGARCH(1,1) with *t* innovations is given by Eq. (3) where y_t is the daily gold prices and s_t is the stationary data (log return price) for the daily gold prices.

$$
\begin{aligned}
y_t &= y_{t-1} \exp\left(s_t\right) \\
s_t &= 0.0009 + a_t \\
a_t &= \sigma_t \varepsilon_t, \quad \varepsilon_t \sim t^{*}_{5.51} \\
\sigma_t^2 &= 9.40 \times 10^{-7} + \left(0.0410 - 0.1858 N_{t-1}\right) a_{t-1}^2 + 0.9544 \sigma_{t-1}^2
\end{aligned}
\tag{3}
$$

In the forecasting stage, the series of out-of-sample transformed data consisting of 284 daily gold prices from 25th April 2013 to 12th June 2014 are used to obtain the forecast results. The forecast evaluations are based on the basis of two evaluations criteria commonly used in the literatures that are the root mean square error (RMSE) and the mean absolute error (MAE). The forecast evaluations generated using fGarch package for MAE, RMSE and MAPE are 11.6173, 16.1576 and 0.8874, respectively. The forecasting using ARIMA(0,1,0)-TGARCH (1,1) model with *t* innovations for daily gold prices is shown in Figure 4 where the forecast data is within ±2 standard errors. Figure 4 graphically proves the promising performance of the hybrid model in forecasting daily gold price where the trend of forecast prices follows closely the actual data for the out-of-sample period. The promising performance is supported by the comparison values between actual and forecast price using the proposed model for the simulation of the last five-day out-sample period as given by Table 2.

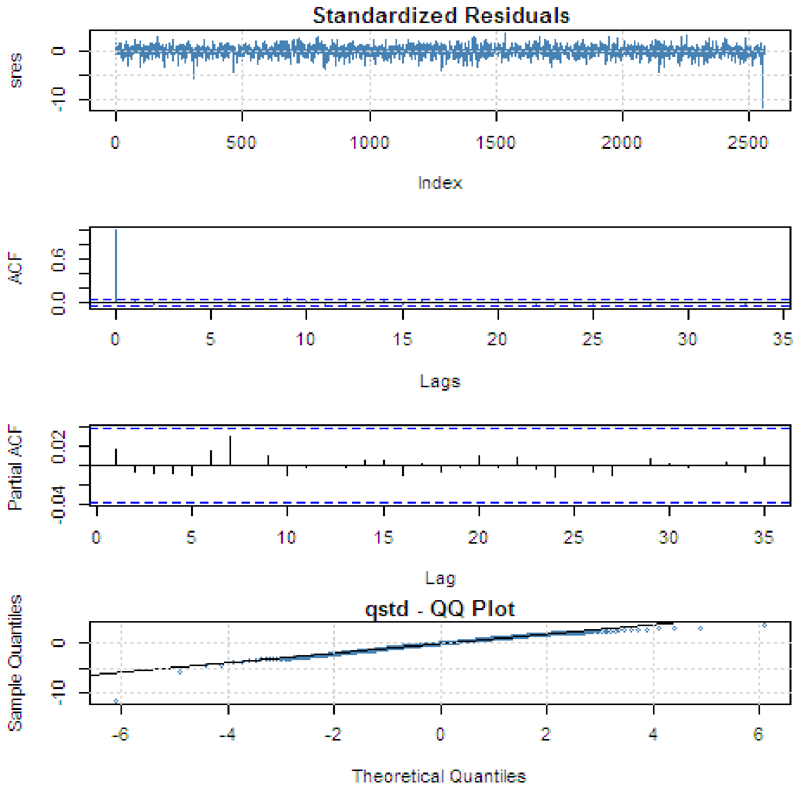

Figure 3: Diagnostic Checking on ARIMA(0,1,0)-TGARCH(1,1) Using *t*-innovations

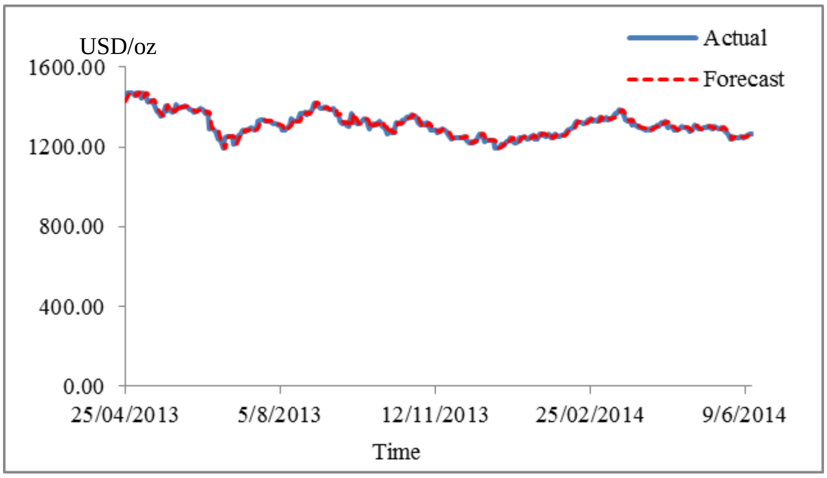

Figure 4: Graph of the Actual Data and Forecast Data Using ARIMA(0,1,0)-TGARCH(1,1) with *t*-innovation

Table 3: The Comparison between Actual Price and Forecast Price for Out-sample Period

Date	Actual price (USD/oz)	Forecast price (USD/oz)
6 June 2014	1247.50	1253.63
9 June 2014	1253.50	1248.63
10 June 2014	1259.50	1254.63
11 June 2014	1262.00	1260.64
12 June 2014	1265.75	1263.14

5 Concluding Remarks

The current study examines the performance of hybrid ARIMA-TGARCH model with five types of innovations in modeling and forecasting daily gold price series. The empirical results indicate that the hybrid model of ARIMA(0,1,0)-TGARCH(1,1) with *t*-innovations fits the series well and outperforms other models since the series is non-normal, have heavy tails characteristics and significant leverage effect. In conclusion, the hybrid ARIMA-TGARCH model with *t*-innovations is a potential approach in modeling and forecasting daily gold price.

Acknowledgements. This work was supported by Ministry of Education, Malaysia under RDU130369.

References

[1] L.E. Blose, Gold prices, cost of carry, and expected inflation, *J. Econ. Bus.*, **62** (2010), no. 1, 35 - 47. http://dx.doi.org/10.1016/j.jeconbus.2009.07.001

[2] F. Capie, T.C. Mills and G. Wood, Gold as a hedge against the dollar, *J. Int. Finance. Mark. Inst. Money*, **15** (2005), no. 4, 343 - 352. http://dx.doi.org/10.1016/j.intfin.2004.07.002

[3] S. Shafiee and E. Topal, An overview of global gold market and gold price forecasting, *Resour. Policy*, **35** (2010), 178 - 189. http://dx.doi.org/10.1016/j.resourpol.2010.05.004

[4] M.M.A. Khan, Forecasting of Gold Prices (Box Jenkins Approach), *Int. J. Emerg. Technol. Adv. Eng.,* **3** (2013), no. 3, 662 - 670.

[5] M. Khashei, M. Bijari and G.A. Raissi Ardali, Improvement of auto-regressive integrated moving average models using fuzzy logic and

artificial neural networks (ANNs), *Neurocomputing*, **72** (2009), no. 4, 956 - 967. http://dx.doi.org/10.1016/j.neucom.2008.04.017

[6] N.H. Miswan, P. Y. Ping, and M.H. Ahmad, On parameter estimation for Malaysian gold prices modelling and forecasting, *Int. J. Math. Anal.*, **7** (2013), no. 22, 1059 - 1068.

[7] D.G. Baur and K.J. Glover, Heterogeneous expectations in the gold market: Specification and estimation, *J. Econ. Dyn. Control*, **40** (2014), 116 - 133. http://dx.doi.org/10.1016/j.jedc.2014.01.001

[8] S. Hammoudeh and Y. Yuan, Metal volatility in presence of oil and interest rate shocks, *Energy Econ.*, **30** (2008), no. 2, 606 - 620. http://dx.doi.org/10.1016/j.eneco.2007.09.004

[9] M. Qadan and J. Yagil, Fear sentiments and gold price: Testing causality in-mean and in-variance, *Appl. Econ. Lett.*, **19** (2012), no. 4, 363 - 366. http://dx.doi.org/10.1080/13504851.2011.579053

[10] S. Trück and K. Liang, Modelling and forecasting volatility in the gold market, *Int. J. Bank. Finance*, **9** (2012), no. 1, 48 - 80.

[11] S.R. Yaziz, N.A. Azizan, R. Zakaria, M.H. Ahmad, The performance of hybrid ARIMA-GARCH modeling in forecasting gold price, *20th International Congress on Modelling and Simulation, Adelaide*, Australia, 2013.

[12] M. H. Ahmad, P. Y. Ping, S.R. Yaziz and N.H. Miswan, A hybrid model for improving Malaysian gold forecast accuracy, *Int. J. Math. Anal.*, **8** (2014), no. 28, 1377 - 1387. http://dx.doi.org/10.12988/ijma.2014.45139

[13] S.R. Yaziz, N.A. Azizan, M.H. Ahmad, R. Zakaria, M. Agrawal and J. Boland, Preliminary Analysis on Hybrid Box - Jenkins – GARCH Modeling in Forecasting Gold Price, *AIP Conference Proceedings*, **1643** (2015), 289-297. http://dx.doi.org/10.1063/1.4907458

[14] L.R. Glosten, R. Jagannathan and D.E. Runkle, On the relation between the expected value and the volatility of the nominal excess return on stocks, *J. Finance*, **48** (1993), no. 5, 1779 - 1801. http://dx.doi.org/10.1111/j.1540-6261.1993.tb05128.x

[15] J.M. Zakoian, Threshold heteroskedastic models, *J. Econ. Dyn. Control*, **18** (1994), no. 5, 931 - 955. http://dx.doi.org/10.1016/0165-1889(94)90039-6

[16] R.S. Tsay, *An Introduction to Analysis of Financial Data with R*, Wiley, Hoboken, New Jersey, 2013.

[17] C. Alcidi, P.D. Grauwe, D. Gros and Y. Oh, *Future of Eurozone and Gold*, Centre for European Policy Studies (CEPS), Brussels, 2010.

[18] C. Pierdzioch, M. Risse and S. Rohloff, On the efficiency of the gold market: Results of a real-time forecasting approach, *Int. Rev. Financ. Anal.*, **32** (2014), 95 - 108. http://dx.doi.org/10.1016/j.irfa.2014.01.012

[19] C.L. Dunis and A. Nathani, Quantitative trading of gold and silver using nonlinear models, *Neural Netw. World*, **17** (2007), no. 2, 93–111.

Received: December 10, 2015; Published: April 12, 2016