ASM Sc. J., 13, 2020
https://doi.org/10.32802/asmscj.2020.sm26(1.14)

Multistep Forecasting for Highly Volatile Data using A New Box-Jenkins and GARCH Procedure

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

[1]*Faculty of Industrial Sciences & Technology, Universiti Malaysia Pahang, 26300 Kuantan, Pahang, Malaysia.*

[2]*School of Information Technology and Mathematical Sciences, University of South Australia, Adelaide SA 5000, Australia.*

The study of the multistep ahead forecast is significant for practical application purposes using the proposed statistical model. This study proposes a new procedure of Box-Jenkins and GARCH (or BJ-G) in evaluating the multistep forecasting performance for a highly volatile time series data. The promising results from one-step ahead out-of-sample forecast series using the BJ-G model has motivated the extension to multiple step ahead forecast. In order to achieve the objective, the procedure of multistep ahead forecast for BJ-G model is proposed using R language. In evaluating the performance of the multistep ahead forecast, the proposed procedure is employed to daily world gold price series of 5-year data. Based on the empirical results, the proposed procedure of multistep ahead forecast enhances the existing procedure of BJ-G which is able to provide a promising procedure to assess the performance of the BJ-G model in forecasting a highly volatile time series data. The procedure adds the value of BJ-G model since it allows the model to describe efficiently the characteristics of the volatile series up to *n*-step ahead forecast.

Keywords: Box-Jenkins, GARCH, highly volatile data, multistep forecast; gold price

I. INTRODUCTION

The Box-Jenkins – GARCH (BJ-G) model has been proven as a promising one in forecasting highly volatile time series data as supported by recent studies such as ARIMA-GARCH (Chen *et. al.,* 2011; Tan *et. al.,* 2010; Zhou *et. al.,* 2006), AR-GARCH (Gaglianone & Marins, 2017) and ARMA-GARCH (Liu & Shi, 2013; Wang *et. al.,* 2005). Given the overall positive results at the one-step ahead forecast in the previous studies using the BJ-G model, it motivates the extension to multistep ahead forecast. To the best of our knowledge, these studies have achieved certain effect in forecasting highly volatile time series data, however there is no study that focuses on the development of procedure of BJ-G model for multistep forecasting performance. The study of the multistep ahead forecast is significant for forecasting real data up to an *n*-step prediction period using the model of Box-Jenkins with GARCH (Babu & Reddy, 2015; Pham & Yang, 2010).

Therefore, this study is aimed at proposing a new procedure of BJ-G in evaluating the multistep forecasting performance of BJ-G model for highly volatile time series data. In order to achieve the objective, the proposed procedure is employed to forecast world gold price series as its nature exhibits highly volatile characteristics. Gold is noted as a volatile monetary asset commodity (Batten *et. al.,* 2010; Lucey *et. al.,* 2013; Yaya *et. al.,* 2016, Yaziz *et. al.,* 2017). In this study, the daily gold price of a 5-year series (2008-2013) as discussed in Yaziz *et. al.,* (2017) is used since the series is considered optimal for BJ-G model. In the proposed procedure, sets of codes are constructed in R by employing the daily gold price series in evaluating the forecasting performance up to *n*-step ahead, which is based on the proposed model of BJ-G.

*Corresponding author: roslindar@ump.edu.my

II. METHODOLOGY

The basic concepts of BJ-G modelling are described in Yaziz *et al.,* (2017) which demonstrates Stage I (Identification) to Stage IV (Forecasting). This study proposed a procedure in assessing the performance of multistep ahead forecasting of BJ-G in Stage IV as presented in Figure 1.

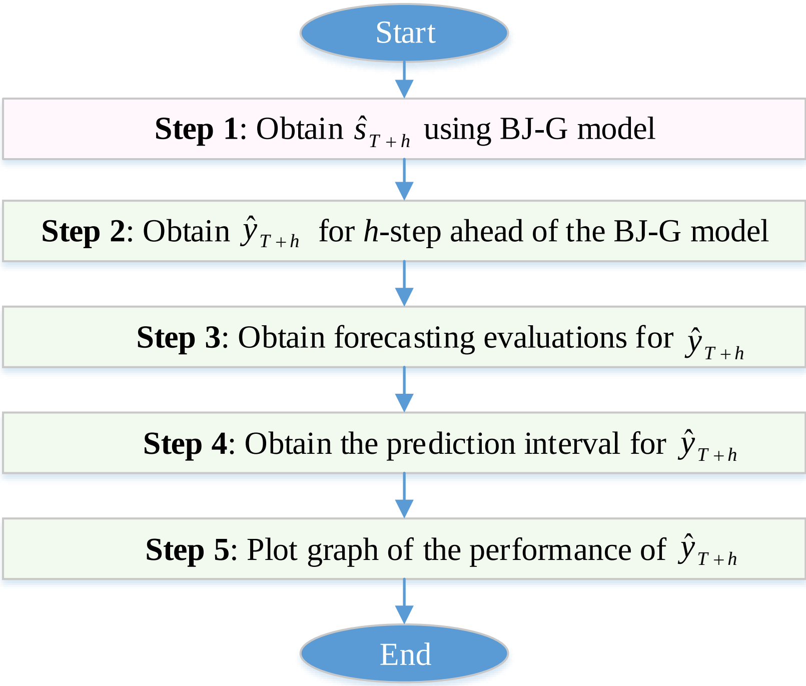

Figure 1. Proposed procedure in Step IV of BJ-G for multistep ahead forecasting

The proposed procedure in Figure 1 is explained explicitly in the following steps.

Step 1: Obtain the simulated stationary series, $\hat{s}_{T+h}$ for forecasting horizon $h = 1,2,3,\ldots,n$ using the proposed BJ-G model. Based on the proposed model for the 5-year series of daily gold price as in Yaziz *et. al.,* (2017), the $\hat{s}_{T+h}$ for ARIMA(0,1,0)-GARCH(1,1) using *t* innovations is given by Equation 1.The corresponding R codes for the proposed $\hat{s}_{T+h}$ are constructed based on the equation.

$$\begin{aligned}
\hat{s}_{T+h} &= 0.0007 + \hat{a}_{T+h} \\
\hat{a}_{T+h} &= \hat{\sigma}_{T+h}\,\hat{\varepsilon}_{T+h} \\
\hat{\sigma}^2_{T+h} &= 2.50\times10^{-6} + 0.0345\hat{a}^2_T + 0.9474\hat{\sigma}^2_T \\
\hat{\varepsilon}_{T+h} &\sim t^*_{4.81}
\end{aligned} \tag{1}$$

Step 2: Obtain the forecast data for *h*-step ahead, $\hat{y}_{T+h}$ of the BJ-G model. For the data in the case study, the forecast data $\hat{y}_{T+h}$ is given by Equation 2 since the transformed data is in logarithm.

$$\hat{y}_{T+h} = y_T \exp\left(\hat{s}_{T+h}\right) \tag{2}$$

Note that $\hat{s}_{T+h}$ is obtained from Step 1. The corresponding R codes of $\hat{y}_{T+h}$ are then constructed for one-step ahead forecast.

Step 3: Obtain forecasting evaluations of the mean absolute error (MAE), the root mean square error (RMSE) and the mean absolute percentage error (MAPE) for *h*-step ahead forecast by comparing $\hat{y}_{T+h}$ to the out-of-sample data y_{T+h}. The corresponding R codes for the forecasting evaluations of $\hat{y}_{T+h}$ are constructed for one-step ahead forecast.

Step 4: Obtain the prediction intervals (PIs) for $\hat{y}_{T+h}$. The PIs gives an interval within which the actual data, y_t is expected to lie with a specified probability by using the forecast, $\hat{y}_{T+h}$. In this study, the PIs used are 80% and 95%, which are commonly used in forecasting method as suggested by (Hyndman and Athanasopoulos, 2014). Since a_t for the series in the case study using the proposed BJ-G model follows a *t* distribution with the degrees of freedom $v = 4.81$, therefore the 80% PIs and 95% PIs for *h*-step ahead are given in Equation 3 and Equation 4, respectively.

$$\hat{y}_T(h) \pm t_{0.1,4.81}\sqrt{\text{Var}[e_T(h)]} \tag{3}$$

$$\hat{y}_T(h) \pm t_{0.025,4.81}\sqrt{\text{Var}[e_T(h)]} \tag{4}$$

In obtaining $\text{Var}[e_T(h)]$, Equation (5) is applied since the proposed model for the data series is ARIMA(0,1,0)-GARCH(1,1), which is a random walk model(Chatfield, 2001).

$$\text{Var}[e_T(h)] = h\,\text{Var}[e_T(1)] \tag{5}$$

In practice, the $\text{Var}[e_T(h)]$ is the variance of the residual for *h*-step ahead, as can be obtained from basic statistics of the residual for each forecast horizon. The R codes for PIs of 80% and 95% of $\hat{y}_{T+h}$ forecast are then constructed for one-step ahead.

Step 5: Graphical presentation for the performance of the forecast data is shown by plotting the graph of actual data in the out-of-sample series, y_{T+h} and the *h*-step ahead forecast, $\hat{y}_{T+h}$ with its prediction intervals. The R codes for plotting the performance with PIs of 80% and 95% are constructed for one-step ahead forecast.

The procedure from step 1 to step 5 for $h = 2, 3, \ldots, n$ is repeated in order to obtain the multistep ahead forecast evaluations for BJ-G model.

III. EMPIRICAL RESULTS

The empirical results of the forecasting performance of BJ-G model is based on 1250 daily world gold price series of 5-year series, starting 22 December 2008 to 17 December 2013. Given the positive results of one-day ahead using the proposed model of BJ-G, the forecasting performance of the model will be assessed at horizons greater than one day. For the 5-year data series under study, the first 1125 data are used to estimate the model while the last 125 data are defined as the out-of-sample series.

Table 1 presents the one-step to 125-step ahead forecast evaluation results with the number of data that lies outside the prediction intervals of 80% and 95% of the forecast value at the forecast origin 1125 for the daily gold price using ARIMA(0,1,0)-GARCH(1,1) with *t* innovations. Refer to Equation 1 for the model and Equation 2 for the updated point forecast, $\hat{y}_{T+h}$. Based on Table 1, the values of MAE, RMSE and MAPE are increasing as the forecast horizon increases. This is in agreement with common sense that $\hat{y}_{T+2}$ is more uncertain as compared to $\hat{y}_{T+1}$. However, it is hard to make a decision based on the forecast evaluations in order to choose the appropriate forecast horizon for the model since the values are marginally increasing, specifically in MAPE values.

Table 1. Forecast evaluation with prediction interval for the considered forecast horizon

Forecast Horizon	Forecast evaluation			Number of data outside prediction interval	
	MAE	RMSE	MAPE	80%	95%
1-step ahead	12.9301	17.8764	0.9956	1	0
2-step ahead	15.7938	21.3297	1.2132	20	1
3-step ahead	18.2953	24.4472	1.4098	25	2
4-step ahead	21.6096	28.3663	1.6716	20	1
5-step ahead	22.8394	28.9304	1.7647	22	1
7-step ahead	24.5981	30.1233	1.8941	17	2

10-step ahead	32.2870	40.1970	2.4859	15	0
15-step ahead	37.6551	46.2091	2.9068	21	3
25-step ahead	43.7949	53.0116	3.3840	36	4
125-step ahead	59.0288	76.2116	4.6135	23	2

Then, by observing the prediction interval for each horizon under consideration, it can be seen that the ten-step ahead forecast results show the lowest number of actual prices that lies out of 80% PIs with no actual prices are out of 95% PIs. The results indicate that the 10-step ahead forecast performs the best in forecasting as compared to other multistep ahead forecast horizon. Therefore, based on Table 1, it can be concluded that the model of ARIMA(0,1,0)-GARCH(1,1) with *t* innovations can be considered for forecast horizons up to 10-day ahead price for 5-year data series. However, 10-day ahead price forecast is surely weaker than for the 1-day horizon.

Figure 2 shows the corresponding out-of-sample forecasting plot of 10-step ahead using the BJ-G model for the daily gold price. The forecast and actual prices are marked by "o" and "●", which linked with red dashed line and blue solid line, respectively. The forecasting plot includes the prediction intervals of 80% and 95% which are presented by the dashed line of green and black, respectively. It can be seen that the forecasting performance of the BJ-G model for up to 10-step ahead forecast is supported graphically by the plot, since all actual prices are within 95% prediction intervals. It is observed that the trend of 10-day ahead forecast price mimics the actual price for the out-of-sample period. Table 2 presents the forecast price of the first 10-day out-of-sample period for 10-step ahead forecast using the ARIMA(0,1,0)-GARCH(1,1) associated with its PIs of 80% and 95% at the forecast origin price 24 June 2013. Based on the forecast price, only two actual data values are not within the 80% PIs while all actual data are within 95% PIs. This indicates that the proposed model of ARIMA-GARCH is able to follow the trend of actual data up to 10-day ahead, specifically within 95% PIs.

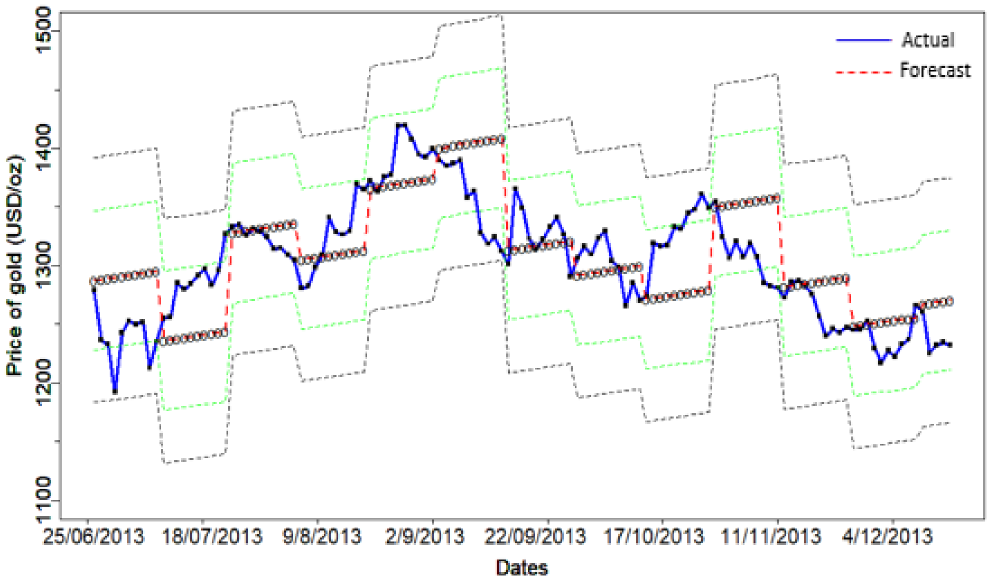

Figure 1. Plot of actual data and 10-step ahead forecast using ARIMA(0,1,0)-GARCH(1,1) with 80% (in green dashed line) and 95% (in black dashed line) prediction intervals

Table 2. Actual price and the 10-step ahead forecast price using ARIMA(0,1,0)-GARCH(1,1)

Date	**Actual Price (USD/oz)**	**Forecast Price (USD/oz)**	**Prediction Interval**	
			80%	**95%**
25 June 2013	1279.00	1287.62	(1228.12,1347.12)	(1183.37,1391.87)
26 June 2013	1236.25	1288.49	(1228.99,1347.99)	(1184.24,1392.74)
27 June 2013	1232.75	1289.36	(1229.86,1348.86)	(1185.11,1393.61)

28 June 2013	1192.00	1290.23	(1230.74,1349.73)	(1185.99,1394.48)
1 July 2013	1242.75	1291.11	(1231.61,1350.61)	(1186.86,1395.36)
2 July 2013	1252.50	1291.98	(1232.48,1351.48)	(1187.73,1396.23)
3 July 2013	1292.85	1188.60	(1233.36,1352.35)	(1188.60,1397.10)
4 July 2013	1293.73	1189.48	(1234.23,1353.23)	(1189.48,1397.98)
5 July 2013	1294.46	1190.35	(1235.11,1354.10)	(1190.35,1398.85)
8 July 2013	1295.48	1191.23	(1235.98,1354.98)	(1191.23,1399.73)

By referring to Equation 1, the significance of $c = 0.0007$ in the mean model of ARIMA(0,1,0)-GARCH(1,1) shows the upward trend of the forecast model which implies that the expected mean return of the series is positive in long term duration. The large value of $\beta_1 = 0.9474$ in the variance model reflects a long-term persistence of volatility clustering. The characteristics that is reflected from the mean and variance model can be used in analysing the actual price up to 10-day ahead.

IV. CONCLUSION

Based on the empirical results, the proposed procedure of multistep ahead forecast to the existing procedure of BJ-G in Lucey et al. (2013) provides a promising procedure to assess the performance of the BJ-G model in forecasting a highly volatile time series data. The procedure adds the value of the combination model of BJ-G since it allows the model to explain more about the characteristics of the volatile series up to *n*-step-ahead forecast.

V. ACKNOWLEDGEMENTS

The authors are grateful to Universiti Malaysia Pahang (Research Grant RDU1703198) for sponsorship of the research work. The authors also acknowledge the constructive criticism by the reviewers.

VI. REFERENCES

Babu, N. C. and Reddy, E. B. 2015, Prediction of selected Indian stock using a partitioning-interpolation based ARIMA-GARCH model, *Applied Computing and Informatics*, 11(2), pp. 130–143.

Batten J. A., Ciner C. and Lucey B. M. 2010, The macroeconomic determinants of volatility in precious metals markets, *Resources Policy,* 35(2), pp. 65–71.

Chatfield, C. 2001. *Time-series forecasting*, Boca Raton: Chapman & Hall/CRC.

Chen C., Hu J., Meng Q. and Zhang Y. 2011, Short-time traffic flow prediction with ARIMA-GARCH model, *IEEE Intelligent Vehicles Symposium (IV),* pp. 607–612.

Gaglianone W. P. and Marins J. T. M. 2017, Evaluation of exchange rate point and density forecasts: An application to Brazil, *International Journal of Forecasting*, 33(3), pp. 707–728.

Hyndman R. J. and Athanasopoulos G. 2014, *Forecasting: principles and practice,* OText.

Liu H. and Shi J. 2013, Applying ARMA–GARCH approaches to forecasting short-term electricity prices, *Energy Economics,* 37, pp. 152–166.

Lucey B. M., Larkin C. and O'Connor F. A. 2013, London or New York: where and when does the gold price originate? *Applied Economics Letters,* 20(8), pp. 813–817.

Pham, H. T. and Yang, B. S.2010. Estimation and forecasting of machine health condition using ARMA/GARCH model, *Mechanical Systems and Signal Processing*, 24(2), pp. 546–558.

Tan Z., Zhang J., Wang J. and Xu J. 2010, Day-ahead electricity price forecasting using wavelet transform combined with ARIMA and GARCH models, *Applied Energy*,87(11), pp. 3606–3610.

Wang W., Van Gelder P. H. A. J. M., Vrijling J. K. and Ma J. 2005, Testing and modelling autoregressive conditional heteroskedasticity of streamflow processes, *Nonlinear Processes in Geophysics,* 12(1), pp. 55–66.

Yaya O. S., Tumala M. M. and Udomboso C. G. 2016, Volatility persistence and returns spillovers between oil and gold prices: Analysis before and after the global financial crisis, *Resources Policy,* 49, pp. 273–281.

Yaziz S. R., Zakaria R. and Ahmad M. H. 2017, Determination of sample size for higher volatile data using new framework of Box-Jenkins model with GARCH: A case study on gold price, *IOP Conf. Series: Journal of Physics 890,* 012161.

Zhou B., He D. and Sun Z. 2006, Traffic predictability based on ARIMA/GARCH model, *2nd Conference on Next Generation Internet Design and Engineering (NGI),* pp. 200–207.