Pertanika J. Soc. Sci. & Hum. 24 (S): 101 – 108 (2016)

SOCIAL SCIENCES & HUMANITIES

Journal homepage: http://www.pertanika.upm.edu.my/

Improving Prediction of Gold Prices through inclusion of Macroeconomic Variables

Beh, W. L.[1]* and Pooi, A. H.[2]

[1]*Department of Physical and Mathematical Science, Faculty of Science, Universiti Tunku Abdul Rahman, 31900 Kampar, Perak, Malaysia*
[2]*Sunway University Business School, Sunway University, Bandar Sunway, 47500 Selangor, Malaysia*

ABSTRACT

This paper uses a method based on multivariate power-normal distribution for predicting future gold prices in Malaysia. First let $\mathbf{r}(t)$ be the vector consisting of the month-*t* values of *m* selected macroeconomic variables, and gold price. The month-(*t*+1) gold price is then modelled to be dependent on the present and *l*-1 on past values $\mathbf{r}(t)$, $\mathbf{r}(t-1)$, …, $\mathbf{r}(t-l+1)$ via a conditional distribution which is derived from a $[(m+1)l+1]$-dimensional power-normal distribution. The mean of the conditional distribution is an estimate of the month-(*t*+1) gold price. Meanwhile, the 100(α/2)% and 100(1-α/2)% points of the conditional distribution can be used to form an out-of-sample prediction interval for the month-(*t*+1) gold price. For a given value of *l*, we select various combinations of *m* variables from a pool of 17 selected macroeconomic variables in Malaysia, and obtain the combinations of which the corresponding mean absolute percentage errors (MAPE) are relatively smaller while the coverage probabilities and average lengths of the prediction interval are still satisfactory. It is found that the parsimonious model is one of which *l* = 2, *m* = 1 and involving the macroeconomic variable derived from the Gross Domestic Product, Kuala Lumpur Composite Index or Import Trade.

Keywords: Multivariate power-normal distribution, macroeconomic variables, prediction interval, parsimonious model

ARTICLE INFO

Article history:
Received: 24 February 2016
Accepted: 30 September 2016

E-mail addresses:
behwl@utar.edu.my (Beh, W. L.),
ahhinp@sunway.edu.my (Pooi, A. H.)
* Corresponding author

ISSN: 0128-7702

INTRODUCTION

Gold has long been a popular investment due to its liquidity and appreciation of value especially during periods of inflation. Apart from being used to diversify risks, it has also been used as a hedge against inflation, deflation or currency devaluation. As the gold market is subject to speculation and volatility, it is important to have a good prediction of the future price of gold together with the possible range of fluctuation of the price.

Time series models have been proposed to predict future gold prices in Malaysia. Pung, Miswan & Ahmad (2013) showed that GARCH (1,1) model was a good fit to predict Malaysian gold prices from 18th July 2001 to 25th September 2012. Ahmad & Pung (2014) showed that TGARCH(1,1) was an improvement of GARCH(1,1) model to predict the daily Malaysian gold prices. Ahmad, Pung, Yazir & Miswan (2014) on the other hand showed that the hybrid model of ARIMA(1,1,1)-GARCH(2,1) were able to improve the forecasting accuracy by using ARIMA(1,1,1) only.

A number of researchers have studied the relationship between gold price and macroeconomic factors in Malaysia and foreign countries such as Pakistan, India, and the United States of America (Bapna, Sood, Totala, & Saluju, 2012; Ernie, 2013; Sindhu, 2013; Warda, Zakaria & Farrukh, 2014; Anuar, Hazila & Saadah, 2015). Ibrahim & Baharom (2011) applied the regression method to analyse the relationship between gold price in Malaysia and the once-lagged stock return (Kuala Lumpur composite index) from August 2001 to March 2010. Ibrahim, Kamaruddin & Hasan (2014) used multiple linear regression to determine the significant relationship between gold price in Malaysia with crude oil price, inflation rate, and exchange rate. Their results show an existing significant relationship between the gold price and all the factors. However, they indicated the need to consider other factors such as unemployment rate, political risks, gross domestic product and etc. in future studies.

In this paper, we investigate the improvement in forecasting accuracy affected by inclusion of macroeconomic variables in the time series model based on multivariate power-normal distribution. It is found that a good model would be a lag-1 model involving only one of the macroeconomic variables derived from Gross Domestic Product, Kuala Lumpur composite Index and Import Trade.

METHOD BASED ON MULTIVARIATE POWER-NORMAL DISTRIBUTION

Let us begin with Yeo & Johnson's (2000) non-normal distribution. These authors introduced the following power transformation:

$$\tilde{\varepsilon} = \psi(\lambda^+, \lambda^-, z) = \begin{cases} \left[(z+1)^{\lambda^+} - 1\right] / \lambda^+, & (z \geq 0, \lambda^+ \neq 0) \\ \log(z+1), & (z \geq 0, \lambda^+ = 0) \\ -\left[(-z+1)^{\lambda^-} - 1\right] / \lambda^-, & (z < 0, \lambda^- \neq 0) \\ -\log(-z+1), & (z < 0, \lambda^- = 0) \end{cases} \quad (1)$$

If z in Eqn. (1) has the standard normal distribution, then $\tilde{\varepsilon}$ has a non-normal distribution which is derived by a type of power transformation of a random variable with normal distribution. We may say that $\tilde{\varepsilon}$ has a power-normal distribution.

Let $\boldsymbol{y}$ be a vector consisting of k correlated random variables. The vector $\boldsymbol{y}$ is said to have a k-dimensional power-normal distribution with parameters $\boldsymbol{\mu}, \mathbf{H}, \lambda_i^+, \lambda_i^-, \sigma,\ 1 \le i \le k$ if

$$\mathbf{y} = \boldsymbol{\mu} + \mathbf{H}\boldsymbol{\varepsilon} \quad (2)$$

where $\boldsymbol{\mu} = E(\mathbf{y})$, $\mathbf{H}$ is an orthogonal matrix, $\varepsilon_1, \varepsilon_2, ..., \varepsilon_k$ are uncorrelated,

$$\varepsilon_i = \sigma_i\left[\tilde{\varepsilon}_i - E(\tilde{\varepsilon}_i)\right] / \left\{\mathrm{var}(\tilde{\varepsilon}_i)\right\}^{1/2}, \quad (3)$$

$\sigma_i > 0$ is a constant, and $\tilde{\varepsilon}_i$ has a power-normal distribution with parameters λ_i^+ and λ_i^-.

When the values of $y_1, y_2, ..., y_{k-1}$ are given, we may find an approximation for the conditional probability density function (pdf) of y_k by using the numerical procedure in Pooi (2012).

Denoting t as the present time (month) and letting $x^* = \{x_1^*, x_2^*, ..., x_m^*\}$ to be a subset chosen from a pool $\{x_1, x_2, ..., x_M\}$ of M selected macroeconomic variables while x_{M+1} as the gold price variable, we may choose the variables $y_1, y_2, ..., y_k$ for the lag-(l-1) model to be those given by the values of $\{x_1^*, x_2^*, ..., x_m^*, x_{M+1}\}$ at time t, t-1, …, t-l+1 together with the value of x_{M+1} at time t+1.

Using data spanning over T months, we can form a table of T-l rows with each row representing an observed value of $(y_1, y_2, ..., y_k)$. From the table, we can form the i_w-th moving window of size n from the i_w-th row till the (i_w+n-1)-th row. We can form a total of N-l-n such windows of size n. We next find a k-dimensional power-normal distribution for $(y_1, y_2, ..., y_k)$ using the data in the i_w-th window. The procedure for parameter estimation to obtain the k-dimensional power-normal distribution can be found in Section IV of Pooi (2012).

Letting $y_1, y_2, ..., y_{k-1}$ be given by the first k-1 values in the (i_w+n)-th row immediately after the i_w-th window, we may now find a conditional distribution for y_k when $y_1, y_2, ..., y_{k-1}$ are given. The mean $\hat{y}_k^{(iw)}$ of the conditional distribution is then an estimate of the value of the gold price next month. On the other hand, the 100(α/2)% and 100(1-α/2)% which point to the conditional distribution may be regarded as the lower and upper limits of the nominally 100(1-α)% out-of-sample prediction interval for the gold price next month. The mean absolute percentage errors (MAPE) is given by

$$\mathrm{MAPE} = \left[\frac{1}{T-l-n}\sum_{iw=1}^{T-l-n}\left|\hat{y}_k^{(iw)} - y_k^{(iw)}\right| / y_k^{(iw)}\right] \times 100\% \quad (4)$$

whose $y_k^{(iw)}$ is the observed value of the gold price next month. The value of MAPE which is small (<5%) is an indication that the predictive power of the model is good (Wang, 2010).

The coverage probability of prediction interval may be estimated by the proportion of prediction intervals which include observed gold price the following month. Meanwhile, the expected length of the prediction interval may be estimated by the average length of the prediction intervals. When the estimated coverage probability is close to the target value 1-α, a small value of the average length is indicative of good predictive power of the model (Hahn & Meeker, 1991).

If we choose the variables for a lag-(l-1) model to be those given by the value of x_{M+1} at time t, t-1, …, t-l+1 together with the value of $\hat{x}_{M+1}$ at time t+1, then we can similarly find an estimate of the gold price next month and a nominally 100(1-α)% out-of-sample prediction interval for the gold price next month.

By comparing MAPE and the estimated coverage probability together with the average length of the prediction interval when x^* is used in forming $y_1, y_2, ..., y_k$ to those of the model when x^* is not used in forming $y_1, y_2, ..., y_k$, we may examine the effect of including x^* on the one-step forecasting performance of the prediction interval.

MEASURES OF PERFORMANCE OF MODELS

To investigate the performance of the method in previous section, we use the monthly Malaysian gold prices for the period January 2006 to December 2012. Data set for the macroeconomic variables is obtained from the Statistics Department of the National Bank of Malaysia. The 17 macroeconomic variables used in this study are listed in Table 1. The monthly values of m macroeconomic variables together with the monthly gold price, form the vector $\boldsymbol{r}$(t) with $1 \le t \le T = 84$. There are ${}^{17}C_m$ ways of selecting m macroeconomic variables from the pool of 17 macroeconomic variables. Each combination of m macroeconomic variables will yield a lag-(l-1) model from which we find the MAPE, and the estimated coverage probability and average length of the nominally 95% prediction interval for the gold price next month.

Table 1
Macroeconomic Variables and Their Assigned Numbers

Assigned Number	Macroeconomic Variable	Assigned Number	Macroeconomic Variable
1	Gross domestic product	10	Market indicative yield
2	Producer price indicator	11	Average discount rate on Treasury bills
3	Industrial production index	12	Consumer price index (inflation)
4	Gross domestic savings	13	Oil price
5	Unemployment rate	14	Trade (export)
6	Interbank rate	15	Trade (import)
7	Money supply M2	16	Foreign exchange rate (RM/USD)
8	Money supply M3	17	Kuala Lumpur Composite Index
9	Total reserve money		

Table 2 shows five combinations of variables of which the corresponding values of MAPE, estimated coverage probability and average interval length are relatively more satisfactory when the values of l $(1 \le l \le 3)$ and m $(1 \le m \le 2)$ are given. The corresponding performance measures when $m = 0$ are also given in the same table. In getting the results in Table 2, we choose n to be 50, which is generally recommended for time series analysis (Box, Jenkins & Reinsel, 1994).

Table 2

Performance Measures of Relatively Better Models ($\alpha = 0.05$, n = 50)

l	m	x_1^*	x_2^*	Estimated Coverage Probability	Average Length	MAPE
	0	-	-	0.84848	259.6848	4.506679
		1	-	0.84375	247.2800	4.369345
		17	-	0.84375	254.7200	4.469481
	1	2	-	0.84375	257.1600	4.563729
		13	-	0.84375	257.1200	4.572080
1		16	-	0.84375	259.6000	4.582659
		1	2	0.787879	235.7915	4.185093
		1	13	0.818182	240.0970	4.189502
	2	13	17	0.848485	236.6836	4.298899
		1	12	0.787879	237.3042	4.327562
		1	17	0.848485	249.5224	4.337109
	0	-	-	0.87500	222.8400	4.106300
		1	-	0.84375	214.2800	3.861602
		17	-	0.81250	212.4400	4.079795
	1	15	-	0.90625	219.0800	4.092122
		14	-	0.87500	222.0000	4.166314
2		13	-	0.81250	220.4000	4.188430
		1	17	0.81250	205.8000	3.824358
		1	13	0.78125	210.8800	3.898503
	2	1	15	0.84375	204.8800	3.968209
		1	12	0.84375	200.7600	3.970212
		1	3	0.84375	206.7600	3.971066
	0	-	-	0.80600	209.9200	4.265282
		13	-	0.80645	207.6077	4.165282
		15	-	0.80645	207.2774	4.190802
	1	14	-	0.80645	208.7226	4.217964
		17	-	0.80645	193.4865	4.231128
3		3	-	0.80645	208.1858	4.309082
		1	13	0.74194	198.1111	3.967482
		1	15	0.77419	199.2671	4.019773
	2	1	3	0.77419	197.2026	4.072391
		1	5	0.77419	196.4181	4.073602
		1	17	0.77419	189.0684	4.133222

From Table 2, we notice that the performance measures tend to improve when l or m increases. In order to obtain a parsimonious model, we try to keep l and m to be small while still obtaining reasonably good values of the performance measures. Table 2 shows that a parsimonious model would be obtained when l = 2, m = 1 and $x_1^* = 1$, 15 or 17. These values of x_1^* correspond respectively to the macroeconomic variables given by the Gross Domestic Product, Import Trade and Kuala Lumpur Composite Index.

The values of MAPE shown in Table 2 are obtained by using moving windows of size 50 over the fairly long period from January 2006 to December 2012. The values of MAPE which are consistently smaller than 5% indicate that the multivariate power-normal distribution based on a moving window of size 50 provides a fairly good fit to the actual distribution. Thus, the out-of-sample prediction based on the moving windows of size 50 would also be accurate for the future moving windows which involve post December 2012 data.

The lag-1 model based on the Gross Domestic Product is particularly promising with a reduction of MAPE, 14.3%, compared with the lag-0 model without any macroeconomic variables and a reduction of MAPE, 5.9%, compared with the lag-1 model without any macroeconomic variables.

CONCLUSION

The numerical results in previous section show that the inclusion of only one suitable macroeconomic variable is able to improve the predictive power of the model. The estimated coverage probability of the prediction interval rarely goes beyond 0.9. This shows that the gold price next month may not be always concordant with the distribution based on the historical data due to sudden changes in economic conditions. Further research may look into developing a more versatile model which takes into account the possible sudden changes in economic conditions.

REFERENCES

Ahmad, M. H., & Pung, Y. P. (2014). Modelling Malaysian Gold Using Symmetric and Asymmetric GARCH Models. *Applied Mathematical Science*, *8*(17), 817-822.

Ahmad, M. H., Pung, Y. P., Yazir, S. R., & Miswan, N. H. (2014). A Hybrid Model for Improving Malaysian Gold Forecast Accuracy. *International Journal of Math. Analysis*, *8*(28), 1377-1387.

Anuar, M. K., Hazila, N., & Saadah, N. (2015). The Relationship between Macroeconomic Factors and Gold Price in Malaysia. *Internatinal Journal of Business, Economics and Law*, *8*(1), 88-96.

Bapna, I., Sood, V., Totala, N. K., & Saluja, H. S. (2012). Dynamics of Macroeconomic Variables Affecting Price Innovation in Gold: A relationship analysis. *Pacific Business Review Internationa1*, *5*(1), 1-10.

Box, G. E., Jenkins, G. M., & Reinsel, G. C. (1994). *Time Series Analysis: Forecasting and Control*. Prentice Hall, Englewood Cliffs, NJ.

Ernie, N. A. (2013). The Relationship between Macroeconomic Variables and Gold Price in Malaysia. *Prosiding Perkem, 3*, 1413-1421.

Hahn, H. J., & Meeker, W. Q. (1991). *Statistical Intervals: A Guide for Practitioners*. Hoboken, NJ: JohnWiley and Sons, Inc.

İbrahim, M. H., & Baharom, A. H. (2011). The Role of Gold in Financial Investment: a Malaysian Persceptive. *Economic Computation & Economic Cybernetic Studies & Research, 45*(4), 1–12.

Ibrahim, S. N., Kamaruddin, N. I., & Hasan, R. (2014). The Determinants of Gold Prices in Malaysia. *Journal of Advanced Management Science, 2*(1), 38-41.

Pooi, A. H. (2012). A Model for Time Series Analysis. *Applied Mathematical Sciences*, *6*(115), 5735-5748.

Pung, Y. P., Miswan, N. H., & Ahmad, M. H. (2013). Forecasting Malaysian Gold Using GARCH Model. *Applied Mathematical Sciences*, *7*(58), 2879-2884.

Sindhu, D. (2013). A Study on Impact of Select Factors on the Price of Gold. *IOSR Journal of Business and Management, 8*(4), 84-93.

Warda, N., Zakaria, M., &Farrukh, N. K. (2014). Impact of Macroeconomic Factors upon Gold Prices in Pakistan. *Pakistan Journal of Social Sciences (PJSS)*, *34*(1), 383-395.

Wang, X. H. (2010). *Financial Management in the Public Sector: Tools, Applications, and Cases*. M. E. Sharpe, Armonk, New York.

Yeo, I. K., & Johnson, R. A. (2000). A New Family of Power Transformations to Improve Normality or Symmetry. *Biometrika*, *87*(4), 954-959.