

# **Comparative Performance of ARIMA and GARCH Models in Modelling and Forecasting Volatility of Malaysia Market Properties and Shares**

**Nor Hamizah Miswan<sup>1</sup>, Nor Azazi Ngatiman<sup>2</sup>, Khairum Hamzah<sup>3</sup>  
and Zaminor Zamzamin<sup>4</sup>**

<sup>1,2,3</sup> Fakulti Teknologi Kejuruteraan  
Universiti Teknikal Malaysia Melaka, Malaysia

<sup>4</sup> Fakulti Keusahawanan dan Perniagaan  
Universiti Malaysia Kelantan, Malaysia

Copyright © 2014 Nor Hamizah Miswan et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

## **Abstract**

Market properties and shares are important in the field of finance in order to measure the economic growth of a country. These market properties are volatile time series as they have huge price swings in a shortage or an oversupply period. In this study, we use two time series models which are Box-Jenkins Autoregressive Integrated Moving Average (ARIMA) and Generalized Autoregressive Conditional Heterocedasticity (GARCH) models in modelling and forecasting Malaysia property market. The capabilities of ARIMA and GARCH models in modelling and forecasting Malaysia property market will be evaluated by using Akaike's Information Criterion (AIC), Mean Absolute Percentage Error (MAPE) and Root Mean Squared Error (RMSE). It can be concluded that Box-Jenkins ARIMA model perform better compared than GARCH model in modelling and forecasting Malaysia market properties and shares.

**Keywords:** ARIMA, GARCH, Property, Share, Volatility

## **1 Introduction**

The investment market is influenced by various economic conditions and factors. These factors not only affect the investment market internally but also affect the global investment. At times, the market is volatile with huge swings. Volatility is a condition where the conditional variance changes between extremely high and low values.

Malaysia has maintained its positive growth in gross domestic product (GDP) and expected to evolve in 2014 as a result of higher business and customer spending. The country is now still recovering from the epidemic that has resulted an increase of the unemployment rate. This has slightly affected the property sector as the terms and conditions become tighter and subsequently the borrowing for property purchasing has become too expensive. This crisis had a major impact to the overall financial markets including the property market and share.

Although Malaysia is one of the countries that can survive from both Asian and Global financial crisis, Malaysia should be more alert of the economic changes especially in the property investment sector since it is the largest contributor to GDP. In the current study, the volatility of the Malaysia Market Properties and Shares is investigated. The monthly data used are the Kuala Lumpur Stock Exchange Properties for properties and Kuala Lumpur Composite Index (KLCI) for shares from July 1997 to July 2012.

Two time series models where one has the ability to capture the non-constant volatility throughout the observations are used. These models are Autoregressive Integrated Moving Average (ARIMA) and Generalized Autoregressive Conditional Heterocedasticity (GARCH). These models will be used to fit the data where the best model will be used to forecast the future of forecasting market properties and shares.

## **2 Methodology**

Data used for analysis using both ARIMA and GARCH models must be stationary. The data is said to be stationary if the mean, variance and autocorrelation structure are constant over the time interval. A stationary series

does not contain trend, it has no seasonality and flat looking. Stationarity of the data is important to describe the future behaviour of the process.

If the data are not stationary, we must transform them by using first difference. First differences are the data changes from one period to the next. Plotting the data of the first difference can reveal whether the data has been transformed to a stationary series or not. If it is still not stationary the second difference is taken. Model fitting can be carried out once the stationarity of the series has been achieved. In the current study, two time series models will be considered.

#### i) Box-Jenkins ARIMA models

Box-Jenkins ARIMA model has been used widely in many areas of Time Series analysis. Since ARIMA is among the earliest models, the capability of this model always being tested and widely used as a benchmark with other Time Series models. Box-Jenkins ARIMA is known as ARIMA( $p, d, q$ ) model where  $p$  is the number of autoregressive (AR) terms,  $d$  is the number of difference taken and  $q$  is the number of moving average (MA) terms. ARIMA models always assume the variance of data to be constant. ARIMA( $p, d, q$ ) model can be represented by the following equation.

$$(1 - \phi_1 B - \dots - \phi_p B^p)(1 - B)^d Y_t = \delta + (1 - \theta_1 B - \dots - \theta_q B^q) a_t$$

where

$1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p$  is the AR operator of order  $p$

$1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q$  is the MA operator of order  $q$

$\delta$  is the constant term

$a_t$  is the shock element at time  $t$

#### ii) GARCH models

GARCH model is known as a model of heterocedasticity which means not constant in variance. This model has been used widely in financial and business areas since the data of these areas tend to have variability or highly volatile throughout the time. GARCH model is written as GARCH( $q, p$ ) model where  $q$  is the number of moving average (MA) terms and  $p$  is the number of autoregressive (AR) terms. GARCH( $q, p$ ) model can be represented by the following equation.

$$\begin{aligned}
 Z_t &= \mu_t + \varepsilon_t, & \varepsilon_t &\sim N(0, h_t) \\
 \varepsilon_t &= e_t \sqrt{h_t}, & e_t &\sim N(0, 1) \\
 h_t &= \alpha + \sum_{i=1}^p \beta_i \varepsilon_{t-i}^2 + \sum_{i=1}^q \gamma_i h_{t-i}
 \end{aligned}$$

where

- $\mu_t$  is the mean or constant term
- $h_t$  is the conditional variance
- $h_{t-i}$  is the past conditional variance
- $\varepsilon_{t-i}^2$  past squared residual return
- $\alpha > 0, \beta_i \geq 0, \gamma_i \geq 0$

Before using any GARCH models, we need to check the volatility of a data. One of the methods is by computing histogram for a stationary series and check the distribution of data. Kurtosis is the measure of peakness of the data distribution and skewness is the measure of symmetrical of the distribution about the mean. When the value of kurtosis is greater than 3 and it is skewed either to the left or right, then the series is volatile. The kurtosis,  $K$  and skewness,  $S$  are defined as follows,

$$K = \frac{1}{n} \cdot \frac{\sum_{i=1}^n (y_i - \bar{y})^4}{\left(\frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2\right)^2}$$

$$S = \frac{1}{n} \cdot \frac{\sum_{i=1}^n (y_i - \bar{y})^3}{\frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^{3/2}}$$

where

- $\bar{y}$  is the mean of the data
- $y_i$  is the Time Series value of the data
- $n$  is the total number of observations

The first step in fitting the model is model identification. In order to determine the autoregressive ( $p$ ) and moving average ( $q$ ) values, correlogram of sample autocorrelation function (ACF) and partial autocorrelation function (PACF) will be computed. ACF and PACF specify the value of  $q$  and  $p$  respectively.

The next step is to estimate the parameters of the selected models by using Maximum Likelihood Estimation (MLE). MLE is suitable for both linear and nonlinear models and also satisfies all the properties of point estimator. MLE techniques will be applied to both ARIMA and GARCH models. The derivation of MLE for Box-Jenkins ARIMA model is given by the following equation.

$$l = -\frac{n}{2} \ln\{2\pi\sigma_a^2\} - \frac{\sum_{t=1}^n a_t^2}{2\sigma_a^2}$$

where

$$l \text{ is log-likelihood function of the distribution}$$

$$a_t = \theta_1 a_{t-1} + \dots + \theta_q a_{t-q} + Y_t + \phi_1 Y_{t-1} - \dots - \phi_p Y_{t-p}$$

$$\hat{\sigma}_a^2 = \frac{\sum_{t=p+1}^n a_t^2}{df}, df = n - (2p + q + 1)$$

The derivation of MLE for GARCH model is defined as follows,

$$l = -\frac{n}{2} \ln\{2\pi\} - \frac{1}{2} \sum_{t=1}^n \left\{ \ln\{h_t\} + \frac{e_t^2}{h_t} \right\} \quad (7)$$

where

$l$  is log-likelihood function of the distribution

$$h_t = \alpha + \sum_{i=1}^p \beta_i \varepsilon_{t-i}^2 + \sum_{i=1}^q \gamma_i h_{t-i}$$

To select the best model fitting, Akaike Information Criterion (AIC) can be used. The smaller the value of AIC, the better is the model fitting. AIC is defined as follows,

$$AIC = n \ln(MSE) - n \ln n + 2p$$

where

- $p$  is the number of parameters used
- $n$  is the number of observations
- $MSE$  is Mean Square Error of the model

The next step is forecasting. The purpose of forecasting is to predict the future values of the data. The accuracy of the forecasting model can be evaluated by computing the forecast accuracy criterion. Accuracy can be measured using Mean Absolute Percentage Error (MAPE) and Root Mean Square Error (RMSE) and it is defined as,

$$MAPE = \left\{ \left[ \sum_{t=1}^n \left| \frac{y_t - \hat{y}_t}{y_t} \right| \right] / n \right\} \times 100\%$$

$$RMSE = \sqrt{\{\sum_{t=1}^n (y_t - \hat{y}_t)^2\}/n}$$

where

$y_t$  is the actual value

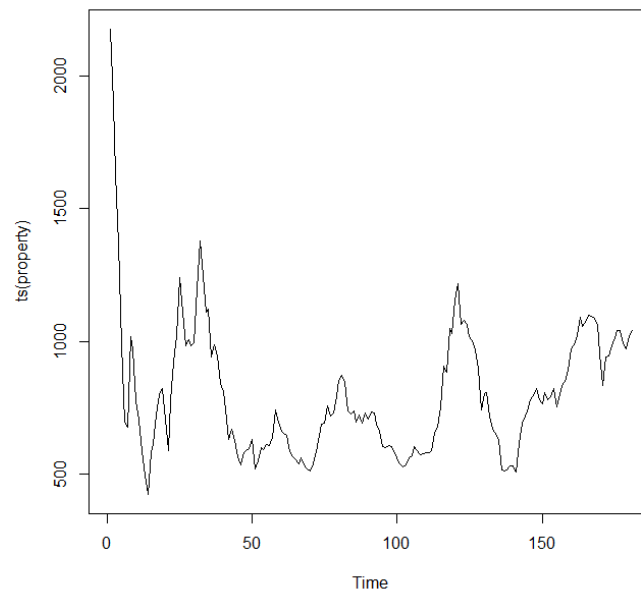
$\hat{y}_t$  is the forecast value

$n$  is the number of period.

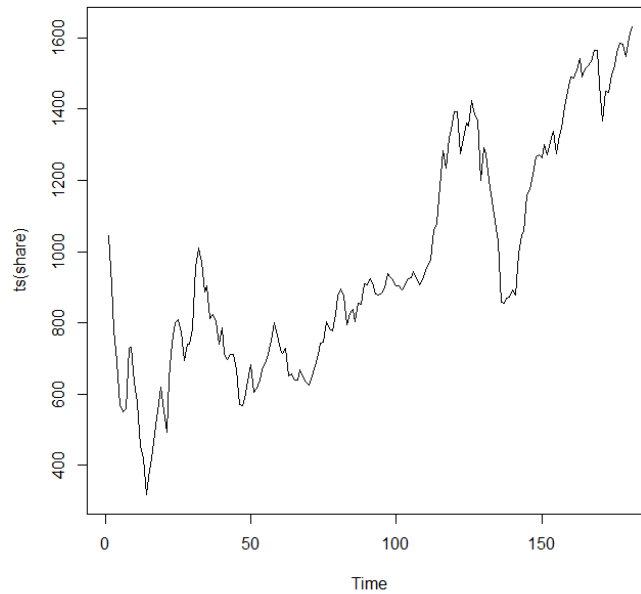
The smaller the values of MAPE and RMSE, the better are the model.

### 3 Data Analysis and Results

Data of Kuala Lumpur Stock Exchange Properties and Kuala Lumpur Composite Index (KLCI) are plotted in Figure 1 and Figure 2.

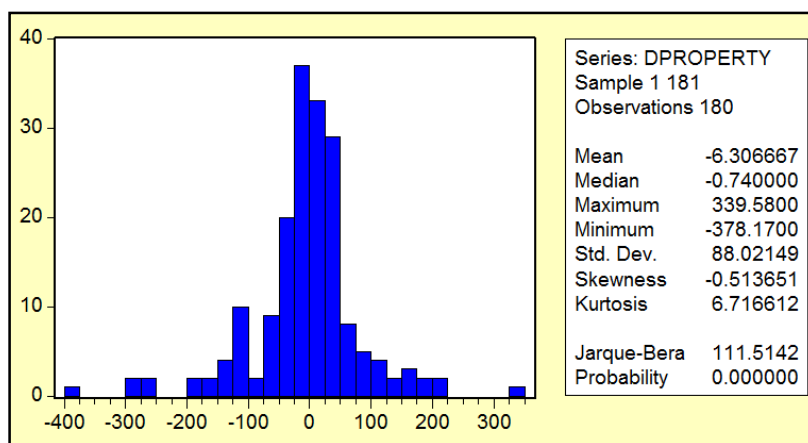


**Figure 1 : Kuala Lumpur Stock Exchange Properties from July 1997 to July 2012**

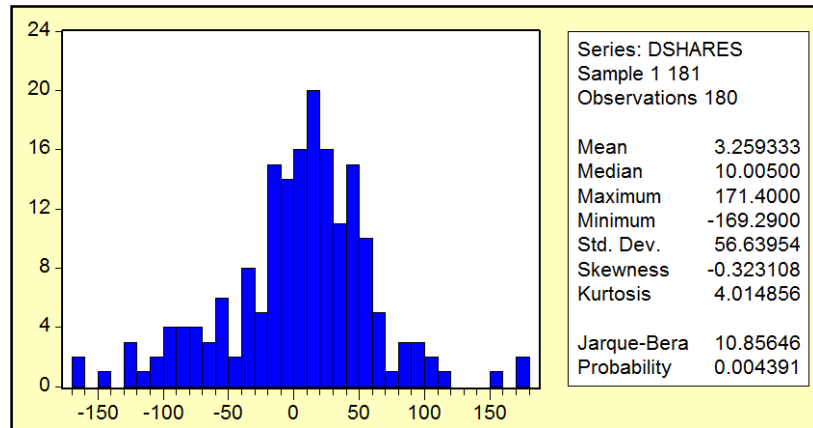


**Figure 2 : Kuala Lumpur Composite Index (KLCI) from July 1997 to July 2012**

For both observations, it is clear that both data are not stationary. Hence, first difference of the data is needed. For GARCH models, we need to plot the histogram at first difference level. GARCH can only be used when the data is volatile. Figure 3 and Figure 4 show the histogram and descriptive statistics for both data at first difference level.



**Figure 3 : Histogram for Kuala Lumpur Stock Exchange Properties at first difference level**



**Figure 4 : Histogram for Kuala Lumpur Composite Index (KLCI) at first difference level**

The values of kurtosis for both series are more than 3 and skewed to the left indicating that GARCH models can be used for both series. Next, identifying the models involved can be done by computing the sample PACF,  $p$  and ACF,  $q$ . ARIMA model is denoted by  $ARIMA(p,d,q)$  and  $GARCH(q,p)$  for GARCH model. The order of differencing,  $d$  is 1 since we are taking first difference of the series.

The corresponding value of  $p$  is 1 and values of  $q$  are 1 and 4 for properties data and for shares, the values of  $p$  and  $q$  are both 1 and 7. In order to obtain the parameters values for each models, Eviews software are used. Table 1, Table 2, Table 3 and Table 4 show the equation of  $ARIMA(p,d,q)$  and  $GARCH(q,p)$  and their corresponding AIC values.

**Table 1 : Equation of  $ARIMA(p,d,q)$  models for Kuala Lumpur Stock Exchange Properties and their corresponding AIC values**

Model	Equation	AIC
ARIMA(1,1,1)	$(1 - 0.769497B)(1 - B)y_t = 3.002597 + (1 + 0.753951B)a_t$	11.55521
ARIMA(1,1,4)	$(1 - 0.818696B)(1 - B)y_t = -1.644376 + (1 + 0.576091B + 0.375584B^2 - 0.077890B^3 + 0.244720B^4)a_t$	11.39220
ARIMA(1,1,0)	$(1 - 0.360012B)(1 - B)y_t = -3.820079 + a_t$	11.61201
ARIMA(0,1,1)	$(1 - B)y_t = -6.462743 + (1 - 0.432535B)a_t$	11.64870
ARIMA(0,1,4)	$(1 - B)y_t = -6.502618 + (1 - 0.433552B - 0.094861B^2 - 0.225968B^3 + 0.017799B^4)a_t$	11.63985



**Table 2 : Conditional variance equation of GARCH( $q,p$ ) models for Kuala Lumpur Stock Exchange Properties and their corresponding AIC values**

Model	Equation	AIC
GARCH(1,1)	$h_t^2 = 226.2584 + 0.159873\varepsilon_{t-1}^2 + 0.763929h_{t-1}^2$	11.17227
GARCH(4,1)	$h_t^2 = 438.5962 + 0.350611\varepsilon_{t-1}^2 + 1.158576h_{t-1}^2 - 1.164714h_{t-2}^2 + 0.774968h_{t-3}^2 - 0.174557h_{t-4}^2$	11.14682
ARCH(1)	$h_t^2 = 3886.305 + 0.432588\varepsilon_{t-1}^2$	11.53150

**Table 3 : Equation of ARIMA( $p,d,q$ ) models for Kuala Lumpur Composite Index (KLCI) and their corresponding AIC values**

Model	Equation	AIC
ARIMA(1,1,1)	$(1 - 0.641073B)(1 - B)y_t = 6.100240 + (1 + 0.530897B)a_t$	10.83197
ARIMA(1,1,7)	$(1 - 0.629056B)(1 - B)y_t = 5.751219 + (1 + 0.443371B + 0.066435B^2 - 0.039412B^3 + 0.142493B^4 - 0.118046B^5 + 0.095017B^6 - 0.136152B^7)a_t$	10.85835
ARIMA(7,1,1)	$(1 - 0.117693B - 0.034255B^2 - 0.037158B^3 + 0.091153B^4 - 0.064306B^5 + 0.0061907B^6 - 0.195353B^7)(1 - B)y_t = 5.662369 + (1 - 0.046419B)a_t$	10.79335
ARIMA(7,1,7)	$(1 + 0.268623B - 0.141927B^2 + 0.128353B^3 + 0.206048B^4 - 0.443837B^5 + 0.209924B^6 - 0.491756B^7)(1 - B)y_t = 6.166295 + (1 - 0.452098B + 0.046524B^2 - 0.190470B^3 - 0.235781B^4 - 0.686566B^5 - 0.382487B^6 + 0.574162B^7)a_t$	10.60028
ARIMA(1,1,0)	$(1 - 0.210693B)(1 - B)y_t = 4.318731 + a_t$	10.84947
ARIMA(7,1,0)	$(1 - 0.159747B - 0.028071B^2 - 0.035761B^3 + 0.090314B^4 - 0.068635B^5 + 0.064337B^6 - 0.198009B^7)(1 - B)y_t = 5.650164$	10.78202
ARIMA(0,1,1)	$(1 - B)y_t = 3.175129 + (1 - 0.192543B)a_t$	10.88662
ARIMA(0,1,7)	$(1 - B)y_t = 3.125076 + (1 - 0.224597B - 0.068880B^2 - 0.083819B^3 + 0.062752B^4 - 0.016080B^5 + 0.081837B^6 - 0.127904B^7)a_t$	10.89789

**Table 4 : Conditional variance equation of GARCH( $q,p$ ) models for Kuala Lumpur Composite Index (KLCI) and their corresponding AIC values**

Model	Equation	AIC
GARCH(1,1)	$h_t^2 = 103.6221 + 0.096757\varepsilon_{t-1}^2 + 0.855591h_{t-1}^2$	10.75801
GARCH(1,7)	$h_t^2 = 2380.795 + 0.063533\varepsilon_{t-1}^2 - 0.042459\varepsilon_{t-2}^2 + 0.061589\varepsilon_{t-3}^2 + 0.010853\varepsilon_{t-4}^2 + 0.044823\varepsilon_{t-5}^2 + 0.023590\varepsilon_{t-6}^2 + 0.273671\varepsilon_{t-7}^2 - 0.188178h_{t-1}^2$	10.82678
GARCH(7,1)	$h_t^2 = 1172.796 + 0.126781\varepsilon_{t-1}^2 - 0.549902h_{t-1}^2 + 0.113817h_{t-2}^2 - 0.150067h_{t-3}^2 - 0.049296h_{t-4}^2 - 0.238405h_{t-5}^2 + 0.507796h_{t-6}^2 + 0.713102h_{t-7}^2$	10.75714
GARCH(7,7)	$h_t^2 = 2320.376 + 0.074029\varepsilon_{t-1}^2 - 0.026897\varepsilon_{t-2}^2 + 0.107617\varepsilon_{t-3}^2 + 0.081705\varepsilon_{t-4}^2 + 0.061621\varepsilon_{t-5}^2 + 0.063598\varepsilon_{t-6}^2 + 0.260546\varepsilon_{t-7}^2 - 0.136159h_{t-1}^2 - 0.034856h_{t-2}^2 - 0.021197h_{t-3}^2 - 0.093275h_{t-4}^2 - 0.038293h_{t-5}^2 - 0.007106h_{t-6}^2 - 0.004816h_{t-7}^2$	10.88835
ARCH(1)	$h_t^2 = 2592.293 + 0.194692\varepsilon_{t-1}^2$	10.91190
ARCH(7)	$h_t^2 = 851.3916 + 0.091275\varepsilon_{t-1}^2 - 0.015023\varepsilon_{t-2}^2 + 0.107512\varepsilon_{t-3}^2 + 0.035505\varepsilon_{t-4}^2 + 0.068279\varepsilon_{t-5}^2 + 0.072775\varepsilon_{t-6}^2 + 0.324163\varepsilon_{t-7}^2$	10.77852

From the above tables, the lowest AIC values are considered to be the best model for modelling properties and shares data. Hence, it can be concluded that ARIMA(1,1,4) and GARCH(4,1) are the best model for Kuala Lumpur Stock Exchange Properties where the AIC values are 11.39220 and 11.14682 respectively. For KLCI series, ARIMA(7,1,7) and GARCH(7,1) are the best models where the AIC values are 10.60028 and 10.75714.

The best models will be used to forecast the data and check the accuracy of these models by using forecast accuracy criterion, MAPE and RMSE. The lowest MAPE and RMSE values are considered to best the best model for forecasting properties and shares data. Table 5 tabulates the MAPE and RMSE values when the selected models

**Table 5 : Forecasting performances of ARIMA and GARCH models**

Series	Best Model	MAPE	RMSE
Kuala Lumpur Stock Exchange Properties	ARIMA(1,1,4)	6.323847	69.66109
	GARCH(4,1)	7.142265	88.38717
Kuala Lumpur Composite Index	ARIMA(7,1,7)	3.938200	44.45587
	GARCH(7,1)	5.085875	57.14706

From Table 5, the MAPE and RMSE values for ARIMA(1,1,4) are smaller for Kuala Lumpur Stock Exchange Properties and ARIMA(7,1,7) are smaller for Kuala Lumpur Composite Index data. The differences in MAPE and RMSE values between ARIMA and GARCH models for properties data are 0.818418 and 18.72608 and for shares data are 1.147675 and 12.69119. These differences are considered large.

#### 4 Conclusion

It is known that GARCH models have been used widely for volatile data. However, for Kuala Lumpur Stock Exchange Properties and Kuala Lumpur Composite Index, GARCH models cannot give the best result. One of the reason is the data are not highly volatile based on the values of kurtosis which are only 6.716612 and 4.014856. Hence, GARCH models cannot capture variability of the data much better than ARIMA models.

As a conclusion, ARIMA is the best models in modelling and forecasting Malaysia market properties and shares as compared to GARCH model. Hence, ARIMA model can be used to predict the future values of Malaysia properties and shares. The forecasting values can help firm and investors to plan their market strategy as well as bring Malaysia economic towards positive growth.

**Acknowledgements.** The main author would like to acknowledge the support of the Faculty of Engineering Technology (FTK), Universiti Teknikal Malaysia Melaka (UTeM).

## References

- [1] D.H. Ackley, G.E. Hinton and T.J. Sejnowski, A learning algorithm for Boltzmann machine, *Cognitive Science*, **9** (1985), 147 - 169.
- [2] F.L. Crane, H. Low, P. Navas, I.L. Sun, Control of cell growth by plasma membrane NADH oxidation, *Pure and Applied Chemical Sciences*, **1** (2013), 31 - 42. <http://dx.doi.org/10.12988/pacs.2013.3310>
- [3] D.O. Hebb, *The Organization of Behavior*, Wiley, New York, 1949.
- [4] Chong Choo, Nie Lee and Nie Ung, Macroeconomics Uncertainty and Performance of GARCH Models in Forecasting Japan Stock Market Volatility. *International Journal of Business and Social Science*, **2** (2011), 200-208 (2011)
- [5] J. D. Cryer and K. S. Chan, *Time Series Analysis with Applications in R (2<sup>nd</sup> ed.)*, New York : Springer, 2008
- [6] Malaysia GDP to Grow 5.1% in 2013, 2014 : World Bank (2013)[Online]. Accessed [6<sup>th</sup> January 2014]. Available from The Sun Daily : [www.the-sundaily.my/news/751900](http://www.the-sundaily.my/news/751900)
- [7] Nor Hamizah Miswan, Modelling and Forecasting Volatile Data by using ARIMA and GARCH Models, Master Thesis, Universiti Teknologi Malaysia, 2013.
- [8] Pung Yean Ping, Nor Hamizah Miswan and Maizah Hura Ahmad, Forecasting Malaysian Gold using GARCH Model, *Applied Mathematical Sciences*, **7** (2013), 2879-2884

**Received: July 1, 2014**