PERFORMANCE OF KUALA LUMPUR COMPOSITE INDEX STOCK MARKET

Syazana Zakaria[1], Badrina Nur Yasmin Badrul Azhar[1],

Intan Nadia Azvilla Maulad Mohamad Rawi[1] and Noreha Mohamed Yusof[2]

[1,2]*Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA,*
Negeri Sembilan Branch, Seremban Campus, Negeri Sembilan, Malaysia
[2]noreh144@uitm.edu.my

ABSTRACT

Financial Times Stock Exchange (FTSE) Bursa Malaysia Kuala Lumpur Composite Index (KLCI) is made up of over 30 large companies listed on the Bursa Malaysia Main Market. All FTSE Bursa Malaysia data are calculated and disseminated every 15 seconds in real-time. It is believed that the volatility of the stock market has a negative impact on real economic recovery. This paper aims to describe the underlying structure and the phenomenon of the sequence of observations in the series. The information obtained, can determine the performance of time series model to fit the data series from January 2002 until December 2018. Autoregressive Integrated Moving Average (ARIMA) and Generalized Autoregressive Conditional Heteroscedasticity (GARCH) models have been shown to provide the correct trend of volatility. The objectives of this paper are to determine the overall trend of the KLCI stock return and to investigate the performance of Generalized Autoregressive Conditional Heteroscedasticity (GARCH) and Autoregressive Integrated Moving Average (ARIMA) based on KLCI stock return. Root Mean Square Error (RMSE) and Mean Absolute Percentage Error (MAPE) have been chosen to be used in this paper to measure accuracy. The results show that the best ARIMA model is ARIMA(1,1), while for the GARCH model, it is GARCH(1,1).

Keywords*: Kuala Lumpur Composite Index, ARIMA, GARCH.*

1. Introduction

Bursa Malaysia Kuala Lumpur Composite Index (KLCI) Financial Times Stock Exchange (FTSE) consists of over 30 big businesses listed on Bursa Malaysia Main Market. All big companies have completed market capitalization that meets the eligibility criteria of the FTSE Bursa Malaysia ground regulations (FTSE Russell, 2018). Based on a weighted value formula, the data is calculated and adjusted by a free float factor, using Bursa Malaysia's real time values and closing prices. Policy makers are interested in the impacts of volatility on actual activity, while the market participants are worried about the impacts of stock market volatility on asset pricing. At the same time, the KLCI index time series also retains the historical motions of the Malaysian stock market.

2. Methodology

2.1 The Data Set

This study uses secondary data obtained from Yahoo Finance through its website (https://finance.yahoo.com). The data consists of the numbers of stock returns in Malaysia from January 2002 to December 2018. Since, the annual time series data used with the model would favourably generate better outcomes with quarterly or monthly numbers where information can be easily obtained, this study uses monthly data.

2.2 Time Series Models

Data containing collection of information that generally takes place at standardized periods is recognized as Time Series data by a model of successive quantitative data pairs (Kenton, 2018). Time series forecasting is a system that predicts potential outcomes depending on current established characteristics (Brownlee, 2017). This is linked to the research conducted in China by Chen & Pan (2016) and in Korea by Han *et al.* (2015). The stock market behaviour could have an impact on investments equity as a consequence of elevated stock market volatility. Due to this, shareholders face a strong investment risk, no matter how big it may be at a comparable moment.

2.2.1 GARCH Model

GARCH model is one of the heteroscedasticity model, where there is an absence of constant variance in the model. This is an upgraded Autoregressive Conditional Heteroscedasticity (ARCH) model by Bollerslev in 1986, which included a word "smoothing-averaging" to create a more parsimonious specification. But both models have considered clustering volatility and forecast time - varying high - frequency financial data as important elements (Islam, 2013). GARCH model should be written as GARCH (q, p) model where q is the moving average number (MA) and p is the autoregressive number (AR). It is possible to follow the general GARCH (q, p) model based on the equation below.

$$y_t = \mu + g_t \tag{1}$$

where $\varepsilon_t \sim N(0,1), \ g_t = \varepsilon_t \sigma_t$

μ is smooth underlying process, g_t is known as the apparent irregularities in the process or mainly known as noise. Meanwhile,

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

where h_t represents conditional variance, h_{t-i} represents past conditional variance, ε^2_{t-i} past squared residual return and $\alpha > 0$, $\beta_i \geq 0$, $y_i \geq 0$.

2.2.2 Box-Jenkins ARIMA Model

These models are combination of Autoregressive Moving Average (ARMA) and ARIMA (Sivakumar & Mohandas, 2009). ARIMA model is one of the statistical models for analysing and forecasting time series data. A standard notation is used for ARIMA (p, d, q) where the parameters are rapidly substituted by integer numbers to show the particular model used for ARIMA. The parameters of ARIMA model shows 'p' as the number of lag observations included in the model (AR), 'd' as the number of times that the raw observations are different (I) and 'q' as the size of the moving average window (MA). With the specified number and type of terms, a linear regression model is constructed, and the data are prepared by a degree of differencing to make them stationary, instantly removing trend and seasonal structures that negatively affect the regression model. To fulfil the assumption of ARIMA models, the data variance should be constant. A general ARIMA equation (p, d, q) model such as ARIMA (1,1,1) is written as:

$$w_t = \mu + \phi_1 w_{t-1} - \theta_1 \varepsilon_{t-1} + \varepsilon_t \qquad (3)$$

where $w_t = y_t - y_{t-1}$ represents the first differences of the series and is assumed stationary. In this case, the values of $p=1$, $d=1$ and $q=1$.

2.3   Unit Root Test Procedure

The non-stationary of a series can be determined by either a simple observation of the plotted data or more accurately by using statistical test procedure. In this paper, the Augmented Dickey-Fuller test (ADF) procedure is most commonly used. The ADF test is performed by using the model:

$$\Delta y_t = \beta_0 + \beta_1 t + \phi y_{t-1} + \sum_{j=1}^{J} \phi_j \Delta y_{t-j} + \varepsilon_t \qquad (4)$$

where J is the number of lags for Δy_t with Δy_t with $\Delta y_t = y_t - y_{t-1}$ and t is the time variable. Normally, J is chosen small in order to save the degree of freedom but is large enough to ensure that ε_t is white noise, where ε_t is identically and independently distributed with mean zero and variance σ_ε^2. The hypothesis testing of ADF is to examine whether there is an existence of unit root (not stationary) or stationary. If the data are not stationary, they must be transformed by using the first difference. The first difference is the data changed from one period to the next one. By plotting the first differencing to data, it could reveal whether the data have been transformed into stationary series or not. If the data are still not stationary after the first differencing, the second differencing is required.

2.4   Model Performance

Akaike Information Criterion (AIC) is used to select the best model fitting when value of AIC is smaller. AIC is described as shown,

$$AIC = -2 \log L(\hat{\theta}) + 2k \qquad (5)$$

where θ is the set (vector) of model parameters, $L(\hat{\theta})$ is the likelihood of the candidate model given when the data are evaluated at the maximum likelihood estimate of θ and k is the number of estimated parameters in the candidate model. The Bayesian Information Criterion (BIC), proposed by Schwarz is also known as the Schwarz Information Criterion (SC). The difference between the BIC and the AIC is that the former imposes a greater penalty than the latter for the number of parameters. BIC, on the other hand, is defined as follows:

$$BIC = -2\log L(\hat{\theta}) + k\log n \tag{6}$$

where θ is the set (vector) of model parameters, $L(\hat{\theta})$ is the likelihood of the candidate model given when the data are evaluated at the maximum likelihood estimate of θ and k is the number of estimated parameters in the candidate model. One of the primary aims of forecasting is to estimate potential outcomes using the correct technique. The precision of the predictive model is often evaluated by calculating the predictive precision criteria. RMSE and MAPE are often used to measure accuracy. It is outlined as:

$$MAPE = \left\{ \frac{\left[\sum_{t=1}^{n} \frac{y_t - \hat{y}_t}{y_t} \right]}{n} \right\} \times 100\% \tag{7}$$

and

$$RMSE = \sqrt{\frac{\left\{ \sum_{t=1}^{n} (y_t - \hat{y}_t)^2 \right\}}{n}} \tag{8}$$

where y_t is the actual value, $\hat{y}_t$ is the forecast value and n is the number of period. The better model can be defined by the smaller values of RMSE and MAPE.

3. Results

The results from Figure 1 focuses on the stock price of KLCI. There had been a slight increase at the beginning of 2002 until the end of 2007 with 1393.25 levels. However, the stock market collapsed at the beginning of 2008 with 1357.400 levels until the first quarter of 2009 with 990.740 levels. After that period, the composite index rose up until 2018 with 1683.530 levels. The side way pattern started at the beginning of 2014 until the end of 2018. On the day this report is written, the KLCI has dropped at 1631.810 levels.

Figure 1. KLCI from January 2002 to December 2018.

Furthermore, trend components have been found to exist in the data series. However, no irregular, seasonal and cyclical components exist along the lines. In this graph, the data set is seen as volatile when fluctuations occurred at around 600 levels to 1500 levels from 1/1/2002 until 1/1/2012 where the trend component could clearly be seen. However, starting from 1/1/2012 until 9/1/2018, the trend still exists but the fluctuation of the data set ranges from 1400 levels to 1800 levels.

3.1   Analysis of ARIMA Model

3.1.1 Data Investigation

The data consists of 204 observations starting from January 2002 until December 2018. The estimation part consists of 144 observations and 60 observations were from the evaluation part.

Table 1 and Figure 2 show the composite index has a unit root. It can be seen that the data are not stationary since the probability value of Augmented Dickey-Fuller Test in Table 1 is 0.6076 which is more than the value of alpha, 0.05. This indicates that the non-seasonal differencing needs to be done to achieve stationary condition. Meanwhile, the graph in Figure 2 is a correlogram that shows observations which have large value of ACF and the pattern has slowly disintegrated. Moreover, PACF has one significant spike at lag 1. The probability of all observations has a significant value which is less than 0.05.

Table 1: The Unit Root Test of The Original Data.

	t-statistic	Prob.*
Augmented Dickey-Fuller test statistics	-1.346619	0.6076

```
Date: 04/08/19   Time: 23:04
Sample: 2002M01 2018M12
Included observations: 204

Autocorrelation   Partial Correlation       AC      PAC    Q-Stat   Prob

                                      1   0.986    0.986   201.26   0.000
                                      2   0.972   -0.017   397.69   0.000
                                      3   0.958   -0.002   589.40   0.000
                                      4   0.941   -0.083   775.58   0.000
                                      5   0.924   -0.061   955.71   0.000
                                      6   0.904   -0.079   1129.1   0.000
                                      7   0.884    0.002   1296.0   0.000
                                      8   0.863   -0.068   1455.8   0.000
                                      9   0.842   -0.010   1608.5   0.000
                                     10   0.818   -0.093   1753.4   0.000
                                     11   0.796    0.058   1891.3   0.000
                                     12   0.775    0.044   2022.7   0.000
                                     13   0.754   -0.005   2147.7   0.000
                                     14   0.735    0.083   2267.2   0.000
                                     15   0.717    0.005   2381.5   0.000
                                     16   0.699    0.007   2490.8   0.000
                                     17   0.682   -0.003   2595.4   0.000
                                     18   0.666    0.019   2695.8   0.000
                                     19   0.651    0.005   2792.2   0.000
                                     20   0.637   -0.023   2884.7   0.000
                                     21   0.625    0.101   2974.5   0.000
                                     22   0.614   -0.013   3061.5   0.000
                                     23   0.604    0.026   3146.1   0.000
                                     24   0.595    0.037   3228.9   0.000
                                     25   0.587    0.008   3309.9   0.000
                                     26   0.581    0.025   3389.5   0.000
                                     27   0.574   -0.017   3467.8   0.000
                                     28   0.566   -0.073   3544.3   0.000
                                     29   0.558   -0.002   3619.2   0.000
                                     30   0.550   -0.055   3692.2   0.000
                                     31   0.541   -0.038   3763.2   0.000
                                     32   0.533    0.062   3832.6   0.000
                                     33   0.525   -0.002   3900.3   0.000
                                     34   0.516    0.004   3966.2   0.000
                                     35   0.505   -0.086   4029.7   0.000
                                     36   0.494    0.006   4090.8   0.000
```
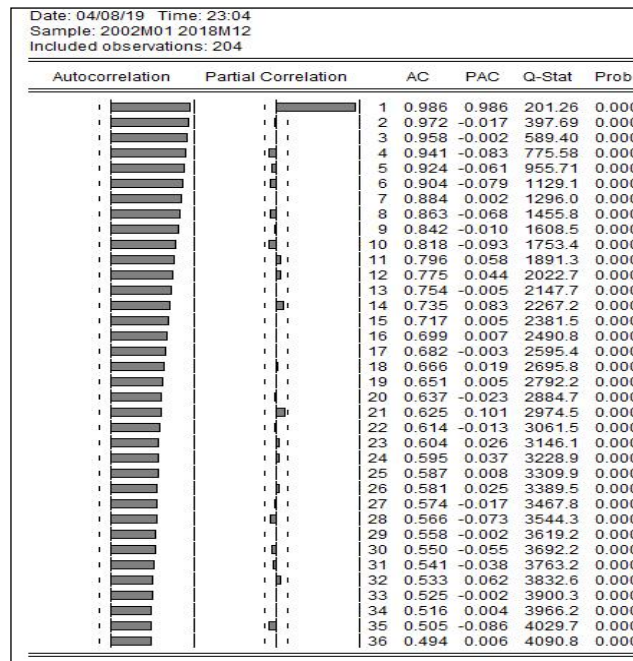
Figure 2. Original data close for KLCI.

3.1.2 Performing Non-Seasonal Differencing

Non-seasonal differencing is used to get the stationary condition. So, differencing would be used to remove systematic pattern or trend components from the data. Moreover, the non-seasonal differencing not only can form first order differencing, but it could also be run repeatedly until data are stationary.

Table 2 shows the data are already stationary after the first order of differencing. It is seen on the Augmented Dickey-Fuller test statistics. The probability is 0.000 which is less than the value of alpha, 0.05. In the correlogram in Figure 3, the value of ACF is small. Meanwhile, the PACF has one significant spike at lag 10. Therefore, 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). In the order of differencing, d is equal to 1 taking the first difference of the series. The corresponding value of p is 1 and the value of q is 1. To obtain the parameter values for each model, EViews software is used.

Table 2: The Unit Root Test After 1st Log Differencing.

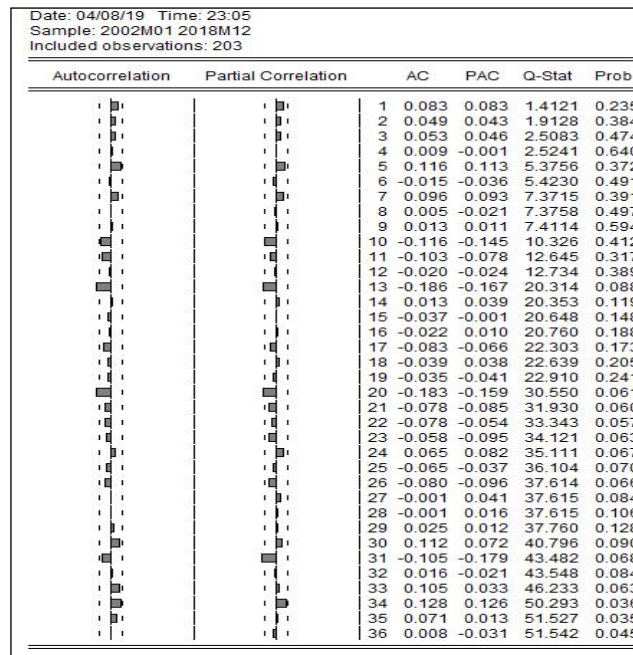|	t-statistic	Prob.*
Augmented Dickey-Fuller test statistics	-13.04413	0.0000

Figure 3. Original Data after First Differencing.

3.2 Analysis of GARCH Model

3.2.1 Data Investigation

Since KLCI data are not stationary, first differencing of the data is needed. For the line graph and histogram, the differencing levels were plotted from GARCH model. GARCH model can only be used when the data are volatile. Figure 4 shows the histogram for Kuala Lumpur Composite Index data at first differencing level.
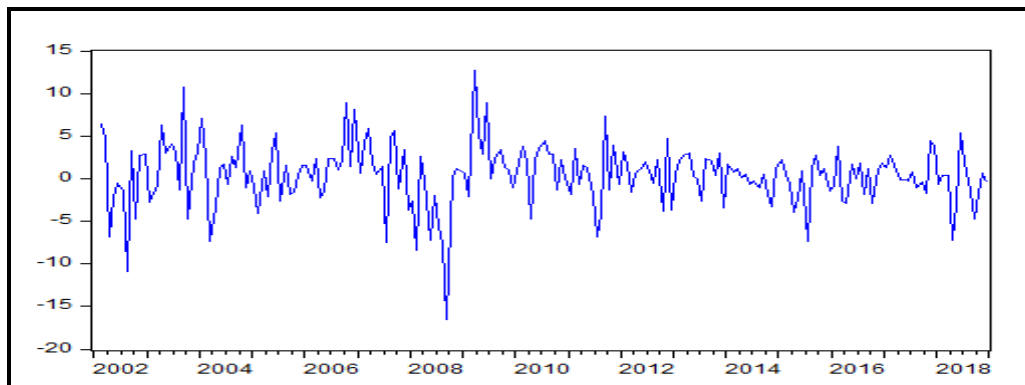

Figure 4. The line graph for volatility clustering.

The line graph shows that the data of KLCI had volatility in the year 2008. This is supported with a histogram and descriptive analysis in Figure 5.

3.2.2 Histogram for First Differencing Level of KLCI

The value of kurtosis for this analysis is 5.825304 which is more than 3 and skews to the left. From the output above, the maximum value is 12.70322 and the minimum value is -16.51417. The average for this series is 0.426069 with a standard deviation of 3.634473. The next step is to identify the model involved by computing the sample PACF, p and ACF, q. GARCH model is denoted by GARCH (q, p). The corresponding values of p and q are 1 and 2.
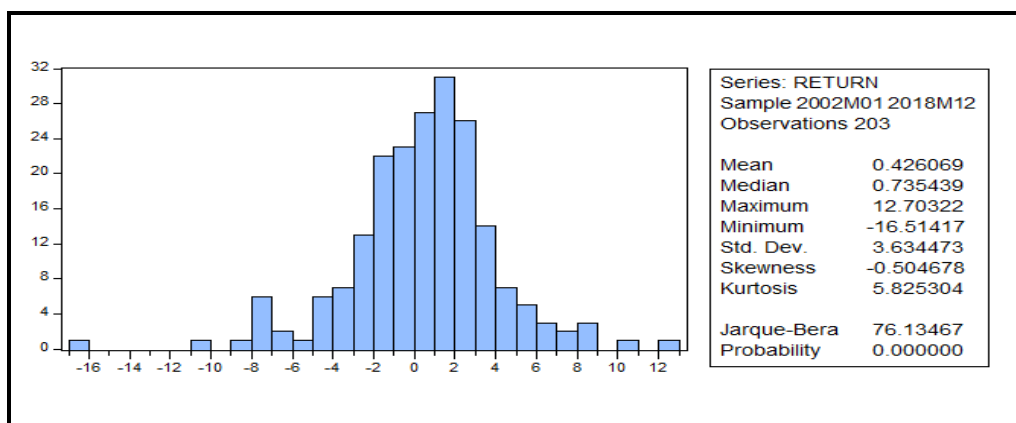
Figure 5. Histogram for KLCI at first difference level.

3.3   Error Comparison in Evaluation Part

3.3.1 Comparison between ARIMA Models

Evaluation part is used to compare the performance of all the ARIMA models. The models' performance used are AIC, SC and a number of significant variables. In order to compare the ARIMA models more efficiently, the error estimations values such as RMSE and MAPE are analysed.

Table 3. Error Comparison for ARIMA Models.

Model	AIC	SC	Error Estimation	
			RMSE	**MAPE**
ARIMA (1,1,1)	10.33554	10.44026	40.37255	1.714587
ARIMA (1,1,2)	10.32155	10.46118	39.40413	1.6674668
ARIMA (2,1,1)	10.33962	10.47924	39.77525	1.679706
ARIMA (2,1,2)	10.35449	10.52902	39.39671	1.667382
ARIMA (3,1,2)	10.38685	10.59628	39.37660	1.662223

Table 3 shows the values of all the statistical measures. The AIC and SC show almost the same figures. Hence, RMSE and MAPE were obtained to compare the errors between models and to find the least value of error. The AIC, SC, RMSE and MAPE show almost the same figures for each model. The number of significant variables and the variables involved are

considered to determine that ARIMA (1,1,1) as the best model since ARIMA (1,1,1) contains 1 out of 2 variables is significant.

3.3.2 Comparison between GARCH Models

Further analysis was done to compare with the GARCH models as shown in Table 4. The AIC, SC, a number of significant variables and also the error estimations were used to achieve the objectives. The lowest AIC and SC is identified for GARCH (1,1). The value of RMSE for GARCH (1,1) is 40.90108 and value of MAPE is 1.647338. Hence, GARCH (1,1) was chosen as the best model after considering all the statistical values and conditions.

Table 4. Error Comparison for GARCH Models.

Model	AIC	SC	Error Estimation	
			RMSE	MAPE
GARCH (1,1)	10.06112	10.23565	40.90108	1.647338
GARCH (1,2)	10.17556	10.38500	40.73866	1.648611
GARCH (2,1)	10.13362	10.34305	40.29174	1.658468
GARCH (2,2)	10.12522	10.36957	40.85229	1.644913
GARCH (3,2)	10.20034	10.30957	40.33413	1.657629

4. Conclusion

Overall, the KLCI stock return in Malaysia from January 2002 until December 2018 shows a tendency towards skewness. This paper is carried out to determine the overall trend on KLCI stock return by using two models which are GARCH and ARIMA. The time series data consists of 204 observations and it is divided into estimation and evaluation part. Based on the findings of this study, that is, by dividing into estimation part which consists of 144 observations and evaluation part consisting of 60 observations for both models, it shows that for the evaluation part, the best model for ARIMA is ARIMA (1,1,1). Meanwhile, for GARCH, the best model is GARCH (1,1).

References

Bollerslev, T. (1986). Generalized Autoregressive Conditional Heteroscedasticity. Journal of Econometrics, 31(3), 307–327.

Brownlee, J. (2017). How to Create an ARIMA Model for Time Series Forecasting in Python. Machine Learning Mastery. Retrieved from website: https://machinelearningmastery.com/arima-for-time-series-forecasting-with-python/

Chen, R., & Pan, B. (2016). Chinese Stock Index Futures Price Fluctuation Analysis and Prediction Based on Complementary Ensemble Empirical Mode Decomposition. Hindawi Publishing Corporation Mathematical Problems in Engineering Volume 2016

FTSE Russell: FTSE Bursa Malaysia Index Series. (2018). Retrieved from website: https://www.ftse.com/products/downloads/FTSE_Bursa_Malaysia_Index_Series.pdf? 32

Han, H., Kutan, A. K., & Ryu, D. (2015). Effects of the US Stock Market Return and Volatility on the VKOSPI. Economics: The Open-Access, Open-Assessment E-Journal, 9 (2015-35): 1—34

Islam, M. A. (2013). Estimating Volatility of Stock Index Returns by Using Symmetric Garch Models. Middle-East Journal of Scientific Research, 2013, 18(7)

Kenton, W. (2018). Time Series. Retrieved from Investopedia Website https://www.investopedia.com/terms/t/timeseries.asp

Sivakumar, P. B., & Mohandas, V. P. (2009). Modeling and predicting stock returns using the ARFIMA-FIGARCH. World Congress on Nature & Biologically Inspired Computing (NaBIC), 2009, 896-901.