Forecasting Gold Prices in Malaysia Using the Box-Jenkins Approach

Tengku Mardhiah Tengku Jalal, Nor Hidayah Hassim, Fadila Amira Razali, Syazwani Zainal Abidin, Nur Fatihah Haron
Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, Cawangan Pahang, Kampus Raub, Malaysia
tengkum908@pahang.uitm.edu.my

Abstract—Forecasting is an important tool that helps in making better decisions. It is a process that predicts and estimates future performance based on historical and current data. In forecasting, the Box Jenkins approach is widely used and one of the popular methods to forecast gold prices as suggested by many studies. This paper aims to identify the best Box Jenkins model for gold price in Malaysia and hence, to forecast the gold price in Malaysia for the first quarter of 2017. Thus, the time series data of gold price in Malaysia from 4th January until 30th December 2016 were used for the study. Based on the time plot and Autocorrelation Function (ACF) plots, three Box-Jenkins models were identified and applied to the data series. A portmanteau test of the L-Jung Box Q test for each model was also conducted before the comparison of Bayesian Information Criterion (BIC), Mean Square Error (MSE), and Mean Absolute Percentage Error (MAPE) was observed. It was observed that the actual data of the gold price in Malaysia is not stationary with an increasing trend pattern. As a result, the Box-Jenkins models with the first order differencing have been applied to the data series. Based on the comparison of the errors, the best Box-Jenkins model obtained was ARIMA (1, 1, 0) and an increasing trend of gold price was estimated to occur in the first quarter of 2017 in Malaysia.

Keywords—ARIMA model; Box Jenkins; forecasting; gold price; time series

I. INTRODUCTION

Throughout the past decades, gold has been one of the most valuable natural resources, which its attraction has driven people to act extremely groundless. The quest for gold has resulted in the occurrence of incidents such as wars, power struggles, and empire building particularly by the major powers. Gold becomes the most precious metal because of its ‘rare characteristics’ which are malleability and everlasting shine. Besides, its density and not being easily eroded by nature also become a factor of attraction. It is undeniable that gold is universally cherished as a symbol of wealth, beauty, and power. In almost every continent, women proudly adorn themselves with gold rings, bracelets, and necklaces. Nowadays, gold is becoming very useful in industrial application as it is a good conductor, very malleable, and highly ductile. It has been used in electronics, typically in the wiring. Currently, gold is widely used as a medium of exchange or currency. For instance, investors view it as a safe-return investment in the face of slowing global economies and recession [1].

This paper aims to identify the best Box Jenkins model for gold price in Malaysia and hence, to forecast the gold price in Malaysia for the first quarter of 2017. Therefore, these studies will provide the need for a better management of gold selling and investment purposes in order to prevent or reduce risks which may lead to financial losses or even bankruptcy.

II. LITERATURE REVIEW

A recent study by [2] gave an insight of gold price forecasting by using the ARIMA model in India. This paper used almost 10 years of data from the Multi Commodity Exchange of India Ltd (MCX). As a result, ARIMA (1, 1, 1) came out as the best model among the six different models to predict the future values of gold. This paper also suggested that the gold price can be affected by a sudden change of factors such as government policies or unstable economy.

Different from other studies, [3] used 17 methods to forecast the price of gold in their paper. Findings revealed that the Exponential Smoothing technique gave the most accurate forecast for gold prices across 24 months of forecasting horizons since it had the lowest average Root Mean Square Error (RMSE). Among the methods used in this study were the Autoregressive model, an optimised ARIMA model, Exponential smoothing (ETS), Exponential smoothing state space model with Box-Cox transformation, ARMA errors, Trend and Seasonal components (TBATS), Fractional ARIMA model (ARFIMA), Vector Auto Regression (VAR), five variations of the Bayesian Auto Regression (BAR) models, and five variations of the Bayesian Vector Auto Regression (BVAR) models. They also suggested that decision makers should take a closer look on gold price as the main indicator for price level and output when there are recessions in the economy. In general, the time series for gold price showed an exponential growth over time, although there was a sign of major shocks in the years 1980 and 2010.

A study “On Parameter Estimation for Malaysian Gold Prices Modelling and Forecasting” by [4] used the Kijang Emas
Prices to model and forecast using the Box-Jenkins methodology and believed that it can be helpful for investment purposes. The method of moment (MME), ordinary least square estimation (OLS) and maximum likelihood estimation (MLE) were used in this study. Based on the value of Akaike information criteria (AIC) and mean absolute percentage error (MAPE), the results showed that the method of OLS gave better forecasts. Based on the data plotted, there was an upward trend throughout the time.

A similar study about forecasting gold prices by using the Box Jenkins approach was also described in [5]. The data was based on the gold price per ounce in London while the Box-Jenkins and Auto Regressive Integrated Moving Average (ARIMA) were used. Findings showed that the efficient model to be used for predicting the gold price was ARIMA (0, 1, 1). Moreover, this study applied the Root Mean Square Error, Mean Absolute Error, and Mean Absolute Percentage Error to test forecasting accuracy. Based on the data plotted, it was clear that the price had an increasing trend, although it showed some fluctuations in the graph pattern.

Another study on “Finding the Best ARIMA Model to Forecast Daily Peak Electricity Demand” by [6] revealed that to forecast two to seven days ahead, the ARIMA model built based on the past three months’ data should be used, whereas to forecast one day ahead, the ARIMA model based on the past six months’ data is the most efficient model. This finding gave an idea on how much of the past data can be used to predict the peak demand in the future.

III. METHODS

The daily selling price (in ringgit) per ounce of gold for the whole year of 2016 used for the study was based on the Kijang Emas Price which was retrieved from the Bank Negara Malaysia website. The time series analysis of the gold price in Malaysia was conducted using both IBM SPSS version 23 and Microsoft Excel 2010.

For the initial data investigation, the time plot of the gold price data for the whole of 2016 was observed. The existence of four-time series components which are trend, seasonal, irregularities, and cyclical was observed based on the fluctuation of the plots. Trend components are said to exist when the time plot shows any general upward or downward movement, while the seasonal variations are identified if there exist any regular fluctuations within a specific period of time [7]. The presence of a wave-like pattern in the Partial Autocorrelation Function (PACF) was also observed in order to confirm the existence of the seasonal pattern. Meanwhile, the irregularities component is said to exist if the time chart shows any outliers or random shock event whereas the cyclical component is present if there are any changes of trend over an unspecified period of time [7].

In order to apply the suitable Box-Jenkins methodology, the ACF (Autocorrelation Function) and PACF were plotted and observed to identify the stationarity of the data series. Based on the ACF and PACF plots, it can be identified whether the data series were suitable for Mixed autoregressive Moving Average (ARMA(p, q)), Mixed Autoregressive Integrated Moving Average (ARIMA(p, d, q)), or Mixed Autoregressive Integrated Moving Average with Seasonal Component (SARIMA(p, d, q)(P, D, Q)). Note that the p is the order of lagged dependent value in Autoregressive (AR), while q is the order of lagged time periods for Moving Average (MA), and d is the level of times the variable needs to be different in order to achieve stationarity. Next, the lags and spikes for both plots were observed to identify the potential order (p, d, or q) of Box-Jenkins models that can be applied for the data series. For this study, three different orders of ARIMA models were chosen as the suitable Box-Jenkins model that can be applied to the daily gold price data series.

Before the chosen Box-Jenkins methods were applied, the data series were divided into two parts. The estimation part included the daily data series of the gold price from January 2016 until September 2016 while the evaluation part was from October 2016 up to December 2016. Next, all of the Box Jenkins models chosen were applied for the estimation part. The forecast values for the range of evaluation part were then obtained for each model for evaluation purposes. All of the process of obtaining the ACF and PACF plots and application of Box Jenkins Method were conducted using IBM SPSS version 23.

Box Jenkins assumes that the error terms are not correlated to each other [7]. Thus, a portmanteau test of Box-Pierce Q statistics (Ljung Box Q test) was conducted for each model in order to test the randomness of the error over time where the hypothesis is;

\[ H_0: \text{The errors are white noise (random)} \]

\[ H_1: \text{The errors are not white noise (not random)} \]

Next, by using the MS Excel, two types of error measures which are Mean Squared Error (MSE) and Mean Absolute Percentage Error (MAPE) were calculated for each applied model. MSE is commonly used in comparing the forecasting performance where the calculation is simply as;

\[ MSE = \frac{1}{n} \sum_{i=1}^{n} e_i^2 \]  \hspace{1cm} (1)

for which \( e_i \) is the difference between the actual observed values with the fitted or forecasted value and \( n \) is the number of effective data points. Meanwhile, MAPE was written as;

\[ MAPE = \frac{1}{n} \sum_{i=1}^{n} \left(\frac{|e_i - y_i|}{y_i}\right) * 100 \]  \hspace{1cm} (2)

where \( y_i \) is the actual observed value.

Both MSE and MAPE were compared to find the best Box Jenkins model and then selected based on the smallest value of the error measures. For obtaining the forecast values of Malaysian gold prices for the first quarter of 2017, the best models identified were applied to the whole data of 2016 and forecasting was conducted for the next 120 days.
IV. RESULTS AND DISCUSSIONS

The time series data of gold price in Malaysia from 4th January until 30th December 2016 were used for the study and depicted in Figure 1(a). Based on the observation of the time plot, the gold price in Malaysia has shown an increasing trend and non-seasonal pattern with significant changes in the volatility of the gold price. Thus, the observation showed that the actual data of the gold prices in Malaysia are not stationary, which means the assumption of stationary is not met. The ACF graph in Figure 1(b) proved the findings obtained, where the autocorrelation function showed a relatively decaying pattern. Meanwhile, the PACF correlogram showed a significant spike at the first lag, followed by other smaller spikes. This suggests that the series can be made stationary after the first differencing is done. The other two components which are cyclical and irregularities do not exist in the data series.

In order to render the original data series stationary, a simple procedure was used which is the ‘first order differencing’. The first order difference, $z_t$, was performed based on the following formula;

$$ z_t = y_t - y_{t-1} \quad (3) $$

Both ACF and PACF correlogram of $z_t = y_t - y_{t-1}$ were then plotted as in Figure 2(a) and Figure 2(b). An improved ACF plot can be observed where the decaying pattern disappeared and spikes at various lags could be seen. Since both ACF and PACF after the first order differencing showed an irregular pattern with several spikes, the process can best be captured by an ARIMA$(p, d, q)$ model.
A. Parameter Estimation

Since the best suggested Box-Jenkins model was identified as ARIMA(p, d, q), the parameters of p and q were identified by observing the significant spikes in ACF and PACF graph. Based on the ACF correlogram, a significant spike can be observed at lag 5 while the PACF plot also showed a significant spike at lag 5. Thus, both parameters p and q will be equal to 1. Therefore, three ARIMA models for the gold price were developed as ARIMA (0, 1, 1), ARIMA (1, 1, 0) and ARIMA (1, 1, 1). TABLE I shows the portmanteau test of all three models applied on the estimation part.

**TABLE I. SUMMARY OF PORTMANTEAU TEST USING LIJUNG BOX Q**

<table>
<thead>
<tr>
<th>Statistics</th>
<th>Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Calculated Q</td>
<td>ARIMA(0,1,1)</td>
</tr>
<tr>
<td>df</td>
<td>17</td>
</tr>
<tr>
<td>Tabulated Q*(0.05)</td>
<td>27.587</td>
</tr>
<tr>
<td>p-value</td>
<td>0.845</td>
</tr>
<tr>
<td>Decision</td>
<td>Accept H0, white noise</td>
</tr>
<tr>
<td>Conclusion</td>
<td>8.102</td>
</tr>
</tbody>
</table>

Based on the portmanteau test, all of the models applied to the data series met the assumption of random error over time. Therefore, the diagnostics checking of the fitted model should be applied. Since all of the models had a property of white noise, the best model was decided via comparing the error measures of the evaluation part as presented in TABLE II.

**TABLE II. ERROR MEASURES FOR EVALUATION PART**

<table>
<thead>
<tr>
<th>Errors</th>
<th>Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>MSE</td>
<td>ARIMA(0,1,1)</td>
</tr>
<tr>
<td></td>
<td>182514.0599</td>
</tr>
<tr>
<td>MAPE</td>
<td>6.8175</td>
</tr>
</tbody>
</table>

Based on the comparison of the error measures, the smallest value of error measures for both MSE and MAPE were obtained by ARIMA(1, 1, 0) with a very slight difference compared to ARIMA(0, 1, 1). Moreover, the Bayesian Criterion (BIC) for ARIMA(1, 1, 0) also had the lowest value which was 8.101 as presented in TABLE I. Therefore, it can be concluded that ARIMA(1, 1, 0) is the best appropriate model to forecast the gold prices in Malaysia.

B. Forecasting

Since ARIMA(1, 1, 0) was identified as the best Box-Jenkins model that fitted the time series data of the gold price in Malaysia, the model was applied to the historical data to forecast gold prices in Malaysia for the first quarter of 2017. Figure 3 depicts the actual historical data with its fitted and forecasted values of the first quarter of 2017. The plots showed that the gold prices in Malaysia were estimated to increase for the first quarter of year 2017.

![Time plot of actual, fitted, and forecast values for Gold Price in Malaysia](image)

**Fig. 3.** Time plot of actual, fitted, and forecast values for Gold Price in Malaysia

V. Conclusion

The secondary data from January 2016 until Dec 2016 were for analysis and forecasting. The data showed that gold prices in Malaysia had an increasing trend without seasonal effects. In this paper, ARIMA (0, 1, 1), ARIMA (1, 1, 0) and ARIMA (1, 1, 1) were applied to forecast Malaysian gold prices for the first quarter of 2017. Based on the comparison of errors, the result showed that the best Box-Jenkins model obtained was ARIMA (1, 1, 0) with the first order difference and it was used to forecast the gold price. As a result, the gold prices in Malaysia were expected to have an upward trend for the first quarter of 2017. This finding was contradictory with the previous studies by [2] and [5]. Based on the findings, gold investors can make better investment decisions hence risk will be minimised when making a decision move. Furthermore, the study presented in this paper suggests future works for the problem involving the prediction of gold prices. For future works, it is suggested that ARIMA (1, 1, 0) is combined with other time series methods. Besides, it includes more data as it can clearly identify the existence of any components compared to data within a short period of time.

**REFERENCES**


