Time series modeling using Box-Jenkins model for shariah compliant healthcare sector in Malaysia

Nur Fathin Shaida Muhammad Nadhirin¹, Norizarina Ishak², Siti Raihana Hamzah³, Nurfadhlina Abdul Halim⁴ and Ahmad Fadly Nurullah Rasade⁵

¹,²,³,⁴ Faculty of Science and Technology, Universiti Sains Islam Malaysia, 71800 Nilai, Negeri Sembilan, Malaysia
⁵ Faculty of Economy and Muamalat, Universiti Sains Islam Malaysia, 71800 Nilai, Negeri Sembilan, Malaysia

E-mail: norizarina@usim.edu.my

Abstract. Most economic and financial time series are trended and not stationary and therefore the raw data need to be detrended by differencing. An appropriate and efficient model is a good practice for evaluating the stock price performance. This paper utilized the Box-Jenkins methodology for modeling the stock price for the healthcare sector in the Malaysian stock market. This study focused on one shariah compliant company, namely Hartalega Holdings Berhad, which is one of the major healthcare companies that manufactured clinical Nitrile gloves in Malaysia. The data of monthly stock price is collected from April 2008 until December 2020. The findings of this study showed that ARIMA(2,1,0) is the best model to represent the time series data according to the smallest values of AIC and SC as well as the satisfaction of Ljung-Box test statistics. Through this study, investors able to monitor the investment portfolio from estimating the stock price performance of healthcare sector, while considering the current healthcare economic situation in Malaysia.

1. Introduction

Modeling return of stock price is not an easy task, especially for the financial time series, because the data shows trend to some degree and not stationary. Apart from the volatility caused by an unexpected return, it is crucial for investors to strategize future work in order to make sound decisions on the investment market. Even there are different methods that can be used as modelling tools to predict the stock price and its performance, to identify the most reliable techniques so that the least possible error measurement is also needed. This study uses the Box-Jenkins methodology, commonly referred to as the ARIMA model that basically suitable to improve short term prediction of financial time series ([1], [2]). It is an integrated form of the ARMA(p,q) model, which is a combination of an autoregressive, AR(p) and moving average MA(q) processes.

Furthermore, the rise in Muslim population together with their demand to allocate their money in financial products that operates in a Shariah compliant manner eventually had given a huge momentum toward Islamic finance in Malaysia financial markets. While this encouraged Muslim investors to take part in the stock market with a variety of Shariah-compliant stocks, including many sectors, this study selected one of the constituents of the FTSE Bursa Malaysia Hijrah Shariah (FBMHS) index, particularly in the health care equipment and services sector. The literary constituents are prohibited from participating in any of the core activities involving interest-related
activities (riba), gambling (maysir) and excessive uncertainty (gharar) ([3], [4]). This shows that the element of preservation of wealth in maqasid shariah is being managed carefully.

On the other hand, a question has been raised about how to help Muslim investors monitor their portfolios in the midst of the Covid-19 pandemic, particularly in the healthcare sector, which appears to attract more attention from society. Having said that, we aim to model the mean return of Shariah-compliant stocks using the ARIMA model in order to assess the expected return of investors. The research flow that has been organized as follows. The second part after this first brief introduction highlights a short review of the literature from prior studies. The methodology for this study is presented in the third part. Then, the fourth part provides the findings along with the assessment of all modeling and estimation process. Lastly, the fifth part concludes this study.

2. Literature review

The use of ARIMA model that has been introduced in Box-Jenkins methodology for guiding the mean and variance stationary series and determining appropriate values of p and q using autocorrelations and partial autocorrelation coefficients through a set of computer programs. It also proposes a diagnostic check for the existence of white noise residual while considering the final order of the model.

Despite this, many researchers have been emphasizing about the accuracy of the method in empirical tests Box–Jenkins that was limited to a short term prediction. This can be seen through the research work from [5] applying the ARIMA model for banking stock market data to incorporate a fit model for forecasting accuracy. Even though the use of the method in the case of bank stock market data has validated its consistency and shown its presentation capabilities, it is not useful for long term. This creates a research gap for future research to provide other prediction horizons for stock market data, such as industrial data. [6] then used ARIMA methods to predict the return values of the S&P Bombay Stock Exchange (BSE) Sensex Index and found greater uncertainty when the forecast period is long-term and less uncertainty occurs in the case of the short-term period. This is also supported by [2] and [7] whereas the experimental results obtained with the best ARIMA model revealed the ability of the ARIMA models to forecast stock prices satisfactory on short-term basis. Therefore, this shows that ARIMA models are known to be robust and effective in the forecasting of financial time series, particularly short-term forecasts.

Apparently, since researchers always want to find the right approach for modelling stock prices, a variety of techniques are used to equate them with the ARIMA methodology. Most researchers believe that ARIMA models can compete effectively against other estimation techniques, such as artificial neural network approaches (ANN) ([7]–[10]). Some technique may help the ARIMA model to produce better predictions. According to [11] who suggested a generative model for forecasting financial time series data of Amman Stock Market (Jordan), based on Wavelet Transforms and the ARIMA model, found that the ARIMA return data after wavelet transformation model created a smaller forecast error compared to the ARIMA model for actual return data. This indicates that the Wavelet-ARIMA model is better than the ARIMA model. However, up to this date, rare to find any article that uses such complex technologies to implement. Thus, the ARIMA model is chosen for its simplicity and broad acceptability of the model ([12]).

In addition to a research that focuses on the accuracy of the expected values using the ARIMA model, some previous studies have addressed stock data for various companies from different sectors. [12] analyzed the feasibility of the ARIMA model forecasting stock prices in the automotive, banking, infrastructure, steel, FMCG, IT and power sector sectors. The FMCG sector delivers the most reliable findings from all sectors. On the other hand, the accuracy of forecasts for the banking sector and the automotive sector using the ARIMA model is lower than in other sectors. Furthermore, [13] used ARIMA model for forecasting the share price performance for oil and gas sector in Malaysia Stock Exchange and the results help investors to assess and forecast the performance of oil and gas sector in Malaysia using ARIMA(1,1,1). [1] also developed the forecasting model ARIMA for the healthcare sector in Malaysia through the stock performance of IHH Healthcare Berhad and found that
ARIMA(1,1,1) is a robust forecasting model. It is therefore in our interest to create a better model for forecasting stocks in the above-mentioned sector by using other well-established healthcare equipment and services company, which will eventually be able to support proof of the good stock performance of the healthcare sector in Malaysia.

3. Data and methodology

3.1. Type and source of data
For the purpose of the study, one of the top 10 constituent companies listed in the FBMHS Index is chosen. The company is Hartalega Holdings Berhad, which is one of the major healthcare companies that manufactured clinical Nitrile gloves in Malaysia. It was successfully listed on the main board of Bursa Malaysia Stock Exchange on 17th April 2008 and this company is listed in FBMHS as one of the Shariah compliant company. Thus, we use secondary data of monthly stock prices from April 2008 until December 2020, downloaded from Bloomberg data streaming. The statistical analysis is conducted using an econometric software, Eviews.

3.2. Stationary
Stationarity is important because if the series is non-stationary, then all the typical results of the classical regression analysis are not valid. Therefore, this study uses a unit root test to test for non-stationarity and the existence of a unit root in the original time series. The unit root test that used in this study is Augmented Dickey-Fuller (ADF) and the following are the hypothesis for ADF test:

H₀: The series is not stationary and has a unit root,
H₁: The series is stationary and has no unit root.

3.3. ARIMA model
ARMA(p,q) model initially proposed by [14] is an autoregressive, AR(p) and a moving average, MA(q) processes, which combine in a general form as follows:

\[ \Delta r_t = \phi_0 + \phi_1 r_{t-1} + \cdots + \phi_p r_{t-p} + \theta_0 u_{t-1} + \cdots + \theta_q u_{t-q} \]  

where,

- \( r_t \): stock return at time \( t \),
- \( p \): lag length of autoregressive,
- \( q \): lag length of moving average,
- \( \phi_{1-\ldots p} \): coefficient of \( r_{t-1-\ldots p} \),
- \( \theta_{1-\ldots q} \): coefficient of \( u_{t-1-\ldots q} \).

The stock return in equation (1) has a future value that is the linear composition of its past and previous residuals. Sequences of residual, \( u_t \) is normally distributed with white noise (WN) that has zero mean and variance \( \sigma^2 \) as \( u_t \sim WN(0, \sigma^2) \).

However, most economic and financial time series is trended and not stationary, which indicates non-constant mean and variance in the series. Taking the natural algorithm of the data series to eliminate its non-stationary of variance, and if the sequence is still non-stationary, according to [15], the raw data can be differen tiated one or two times to induce stationarity in the mean sense, which then followed ARIMA model. The general ARIMA model comprises ARIMA(p,d,q) whereas \( p \) is the AR terms that represent the number of lagged terms of the dependent variable, \( d \) is the number of differences needed to render the series stationary and \( q \) is the MA terms that represent the number of lags of the residuals. The first differences of a series \( r_t \) are given by the equation:

\[ \Delta r_t = r_t - r_{t-1} \]
If the series is stationary at the d-th difference, it is integrated to order d. Therefore, the ARIMA model can be formulated as follows:

$$\Delta^d r_t = \phi_0 + \phi_1 \Delta^d r_{t-1} + \cdots + \phi_p \Delta^d r_{t-p} + u_t + \theta_1 u_{t-1} + \cdots + \theta_q u_{t-q} \tag{3}$$

Box and Jenkins then introduced a three-stage technique to select a suitable (parsimonious) ARIMA model to estimate and predict a univariate time series. The three stages are: (a) identification, (b) estimation, and (c) diagnostic checking. Figure 1 displays an ARIMA process flowchart.

![Box-Jenkins methodology](image)

**Figure 1.** Box-Jenkins methodology.

3.3.1. Identification. In the identification stage, the time plot of the series Autocorrelation Function (ACF) and Partial Correlation Function (PACF) will be examined. From an economic and financial time series, a common stationary-inducing transformation are needed to make the data stationary by taking logarithms and then first differences of the series. After stationarity is achieved, the next step is to identify the p and q orders of the ARIMA model. Table 1 shows the possible combinations of ACF and PACF forms in order to detect the order of ARIMA processes. In particular, mixed processes are hard to define, so that more than one ARIMA(p,d,q) model can sometimes be estimated. That is why it is important and necessary the estimation and the diagnostic checking stages.

| Model   | ACF                                               | PACF                                               |
|---------|---------------------------------------------------|---------------------------------------------------|
| MA(1)   | Single positive spike at lag one                  | Damped sinewave or exponential decay               |
| AR(1)   | Damped sinewave or exponential decay              | Single positive spike at lag one                   |
| ARIMA(1,d,1) | Decay (exponential or sinewave) beginning at lag one | Decay (exponential or sinewave) beginning at lag one |
| ARIMA(p,d,q) | Decay (exponential or sinewave) beginning at lag q | Decay (exponential or sinewave) beginning at lag p |

3.3.2. Estimation. In the estimation stage, every preliminary model is examined and the coefficient of the parameter is evaluated. The Akaike Information Criterion (AIC) and the Schwartz Criterion (SC)
are used to compare the estimated models. For a parsimonious model, the model with relatively small of AIC and SC values and high of adjusted R-squared will be chosen.

3.3.3. Diagnostic checking. In the diagnostic checking stage, the goodness of fit of the model is examined. The special statistic is used, which is the Ljung-Box (LB) Q-statistic in order to test for autocorrelations of the residuals.

3.4 Forecasting
The final model of the ARIMA model is used to predict future values of the time series, as well as to validate the model chosen.

4. Result analysis
4.1 Stationary
In applying the ARIMA model, some techniques required the stationarity of the original time series data. Therefore, the data's autocorrelation function (ACF) and partial autocorrelation function (PACF) are drawn to identify either the data is stationary or not. For the stationary time series, the data will be analyzed using the ARMA model. Meanwhile, the ARIMA model needs to be used to deal with the non-stationary time series.

Figure 2 shows that there is a growing picture of the raw data plot to see that there is a strong pattern. Figure 3 also shows that the ACF does not die at all for all lags, which clearly means that the series is integrated and that we need to continue with logarithms and the first differences in the series.
After we take logs and then first differences of the HTHB series, we obtain the graph of DLHTHB as provided in figures 4 and 5 and notice that, there are stationary in the data series. Most of the values are deviating within the confidence interval and at least the mean and variance are more or less stable, but it is not at all random. It indicates that the trend will be eliminated as the original data series is differentiated.

Then, we conduct the unit root test of ADF and the result of the estimation is as follows:

**Table 2. ADF test result for DLHTHB.**

| ADF Test                  | Coefficient | t-statistics | p-value |
|--------------------------|-------------|--------------|---------|
| With constant            | -0.8494     | -10.3860     | 0.0000  |
| With constant and trend  | -0.8524     | -10.4117     | 0.0000  |
| Without constant and trend | -0.5483   | -5.4560      | 0.0000  |

The test for unit roots of DLHTHB series shows a significant result of p-value at any level of 1%, 5% and 10% significant level with negative coefficients, which indicates that the data does not have a unit root and stationary at the first difference. The model would then be ARIMA (p,1,q) and the next step will be to decide the p and q orders.

4.2. **Model identification**

From this correlogram in figure 5, it can be seen that there are one to two spikes on the ACF and the PACF which then die down immediately after the second lag. This suggests that we might have up to AR(2) and MA(2) specifications. So, the possible models are present below.
Table 3. Statistical results of different ARIMA parameters for DLHTHB.

| Models     | AR | MA | Adjusted R-squared | AIC    | SC     |
|------------|----|----|---------------------|--------|--------|
| ARIMA(0,1,0) | 0  | 0  | 0.0000              | -1.8353| -1.8153|
| ARIMA(1,1,0) | 1  | 0  | 0.0158              | -1.8554| -1.8152|
| ARIMA(0,1,1) | 0  | 1  | 0.0096              | -1.8384| -1.7985|
| ARIMA(1,1,1) | 1  | 1  | 0.0282              | -1.8616| -1.8013|
| ARIMA(2,1,0) | 2  | 0  | 0.0485              | -1.8760| -1.8156|
| ARIMA(0,1,2) | 0  | 2  | 0.0479              | -1.8713| -1.8113|
| ARIMA(2,1,2) | 2  | 2  | 0.0413              | -1.8555| -1.7547|
| ARIMA(2,1,1) | 2  | 1  | 0.0425              | -1.8632| -1.7825|
| ARIMA(1,1,2) | 1  | 2  | 0.0464              | -1.8739| -1.7936|

4.3. Model estimation

Through table 3, ARIMA(2,1,0) is considered to be the best model because it has the smallest Akaike Information Criterion (AIC) and Schwartz Criterion (SC) of -1.8760 and -1.8156 respectively. Nevertheless, the adjusted R-squared of 0.0485 suggested that the model is moderately fit to represent the variables, as the stock are explained by the AR(p) and MA(q) only about 50% of the return. However, ARIMA(2,1,0) was still used as the appropriate model for the stock series. Remember that a parsimonious model is needed, so there might be a problem of overfitting here. Thus, the models are tested whether or not they meet the diagnostic test. If the diagnostic test is not satisfied, it is necessary to identify the possible model again.

4.4. Model diagnostics

In figure 6, for the first 12 lags, there are no significant spikes of ACF and PACF of the residual and most of the coefficients fall within the 95% confidence interval, as well as the insignificant value of Q-statistics as equation (4):

\[ Q = 3.3428 < \chi^2_{12,0.05} = 21.026 \]  

Means the selected ARIMA model’s residuals have no autocorrelation and are white noise.

4.5. Final model

ARIMA(2,1,0) of DLHTHB shows that the current value of its stock return depends on the value of its immediate pass in lagged two. From table 4, the best ARIMA model’s equation for DLHTHB can be expressed as equation (5) below:

\[ r_t = 0.0316 + 0.1277r_{t-1} + 0.2056r_{t-2} \]
Table 4. Estimated model parameter of DLHTHB.

| Type  | Coefficient | Standard Error Coefficient | t-statistics | p-value |
|-------|-------------|-----------------------------|-------------|---------|
| Constant | 0.0316      | 0.0115                      | 2.7413      | 0.0069  |
| AR(1)  | 0.1277      | 0.0820                      | 1.5573      | 0.1216  |
| AR(2)  | 0.2056      | 0.0833                      | 2.4699      | 0.0147  |

4.6. Forecasting
Following the completion of the ARIMA model selection process, then it is used to forecast the stock returns of HTHB from January 2021 to April 2021. The forecasted values are compared to the actual values shown in table 5.

Table 5. Forecast values and actual values of DLHTHB from January 2021 to April 2021

| Month     | Forecast | Actual | Relative Error(%) |
|-----------|----------|--------|-------------------|
| January 2021 | -0.1563  | -0.1636 | 4.46              |
| February 2021 | -0.1399  | -0.1408 | 0.63              |
| March 2021   | 0.1090   | 0.1132  | 3.71              |
| April 2021   | 0.2481   | 0.2515  | 1.35              |

In January 2021, the forecast value is -0.1563. The actual value is -0.1636 and the relative error is 4.46%. Meanwhile, the relative error for the subsequent months of February, March, and April is 0.63%, 3.71%, and 1.35%, respectively. This shows that the predicted value is similar to the real value with a slight relative error for the next four months after the study period. Hence, the model is appropriate to forecast the stock returns of HTHB for the next few months.

5. Conclusion
In this paper, we came up with an appropriate model using the Box-Jenkins approach, that is, the ARIMA model. The experimental findings were obtained as follows. The original data series of HTHB stock price is not stationary and thus need to be converted into log difference transformation, which presented as DLHTHB. Through the correlogram of ACF and PACF of DLHTHB series, there are one to two spikes which then die down immediately after the second lag. So, the possible models might have up to AR(2) and MA(2) specifications. Among all the models, ARIMA(2,1,0) is chosen as the best ARIMA model with the lowest AIC and SC values with a moderate fit of R-squared value. Then, to validate the acceptance of the model, the residual diagnostics showed no significant spikes of ACF and PACF, as well as the Q-test.

From this statistical analysis, the model can be used to predict the monthly stock price of Hartalega Holdings Berhad. Therefore, it is hoped that investors will be able to acquire information in order to manage their portfolio with a profitable short-term forecast return that is compatible with Malaysia stock market.

6. References
[1] Bakar N A and Rosbi S 2017 Dynamic Forecasting method for Shariah-compliant Share Price of Healthcare sector in Malaysian Stock Exchange Int. J. Adv. Eng. Manag. Sci. 3 855–863.
[2] Adebiyi A A, Adewumi A O and Ayo C K 2014 Stock price prediction using the ARIMA model Proc. - UKSim-AMSS 16th Int. Conf. Comput. Model. Simulation 106–112
[3] Elias S M Razak M Z and Kamil K H 2014 Investment strategies in malaysian shariah-compliant equities with transaction costs Ulûm Islâmiyyah J. 13 179–196
[4] Bakar N A and Rosbi S 2016 Long term performance of Islamic share price for initial public offerings (IPO) in Malaysia: evidence from sharia-compliant companies listed on the Malaysian Stock Exchange Int. J. Manag. Sci. Bus. Adm. 2 55–66
[5] Almasarweh M and Al Wadi S 2018 *ARIMA model in predicting banking stock market data* *Mod. Appl. Sci.* **12** 309

[6] Latha M, Nageswararao K S and Venkataramanaiah M, 2018 *Forecasting time series stock returns using Arima : Evidence from S & P Bse Sensex* *Int. J. Pure Appl. Math.* **118** 1–25

[7] Pieleanu F D 2016 *Predicting the evolution of bet index, using an ARIMA model J. Inf. Syst. Oper. Manag.* **10** 1–12

[8] Saxena P, Merh N and Pardasani R 2010 *A comparison between hybrid approaches of ANN and ARIMA for Indian stock trend forecasting* *J. Bus. Intell.* **3** 23–43

[9] Wijaya B, Kom S and Napitupulu A 2010 *Stock price prediction: Comparison of Arima and artificial neural network methods—an Indonesia stock’s case* *Proc. 2nd Int. Conf. Adv. Comput. Control Telecommun. Technol.* 176–179

[10] Gao G, Lo K and Fan F 2017 *Comparison of ARIMA and ANN models used in electricity price forecasting for power market energy* *Power Eng.* **9** 120–126

[11] Al Wadi S, Ismail M T, Alkahhazaleh M H and Addul Karim S A 2011 *Selecting wavelet transforms model in forecasting financial time series data based on ARIMA model* *Appl. Math. Sci.* **5** 315–326

[12] Mondal P, Shit L and Goswami S 2014 *Study of effectiveness of time series modeling (ARIMA) in forecasting stock prices* *Int. J. Comput. Sci. Eng. Appl.* **4** 13–29

[13] Bakar N A, Rosbi S and Uzaki K 2018 *View of evaluating forecasting method using Autoregressive Integrated Moving Average (ARIMA) approach for shariah compliant oil and gas sector in Malaysia* *J. Math. Comput. Sci.* **3** 19–33

[14] Box G E P and Jenkins G M 1976 *Time Series Analysis: Forecasting and Control* Revised ed. San Francisco Holden-Day

[15] Alfaki M M A and Masih S B 2015 *Modeling and forecasting by using time series ARIMA models* *Int. J. Eng. Res.* **4** 914–918