THE APPROACH OF BOX JENKINS TIME SERIES ANALYSIS FOR PREDICTING STOCK PRICE ON LQ45 STOCK INDEX

Isnaini Nuzula Agustin
Lecturer of Management Study Program
Universitas Internasional Batam
Email: isnaini.nuzula@uib.ac.id

ABSTRAK

Ketepatan prediksi harga saham merupakan aspek penting bagi investor untuk mendapatkan tingkat pengembalian saham yang maksimal. Analisis Teknikal merupakan analisis saham yang berpendapat bahwa harga saham dibentuk oleh pola harga saham itu sendiri di masa lalu. Salah satu teknik pemodelan pada analisis teknikal yang telah banyak digunakan adalah Model Box Jenkins atau secara teknis disebut Autoregressive Integrated Moving Average (ARIMA). Penelitian ini memprediksi indeks saham melalui pembentukan model ARIMA. Indeks saham LQ45 sebagai salah indeks saham *blue chip* dan memiliki tingkat likuiditas tinggi di Indonesia, akan digunakan sebagai data sampel, yaitu data saham harian LQ45 pada periode Januari 2016 – November 2018. Analisis data meliputi uji stasioneritas, pembentukan model ARIMA dan dilanjutkan dengan uji beda t-test berpasangan untuk melihat efektivitas model ARIMA yang dihasilkan. Hasil penelitian menunjukkan bahwa data tersebut stasioner pada tingkat diferensi satu, dan model yang sesuai adalah ARIMA (2,1,1). Model menghasilkan R-squared yang cukup tinggi yaitu 98.3% dengan RMSE 9.946. Model ini selanjutnya digunakan untuk memprediksi harga saham pada periode 15 hari mendatang. Hasil paired t-test menunjukkan bahwa tidak terdapat perbadaan signifikan antara data aktual dengan data hasil prediksi. Sehingga dapat disimpulkan bahwa model ARIMA yang dihasilkan telah sukses memprediksi indeks harga saham LQ45.

Kata kunci: Indeks Saham, analisis teknikal, Indeks LQ45, ARIMA

ABSTRACT

The accuracy of stock price prediction is an important aspect for investor in order to gain maximum returns. Technical analysis is one of stock analysis which argue that future stock prices are formed by price movements and the stock price pattern itself. One of modelling techniques in technical analysis that widely used is Box Jenkins Model or technically named as Autoregressive Integrated Moving Average (ARIMA). This research aims to predicting stock index through ARIMA. Having high liquidity and known as blue chip index in Indonesia, LQ45 will be used as data samples, that is daily stock index in period of Januari 2016 to November 2018. Data analysis started by running stationarity test, followed by determination of ARIMA model and paired sample t-test to measure the model's effectivity. The research result shows that the data stationary after first differencing, and the best model is ARIMA (2,1,1). Model having fairly high R-squared and minimum RMSE namely 98.3% and 9.946 resepctively. This model then used to predict future stock index in the next 15 days. Furthermore, paired t-test was run and shows that there is no significance differences between actual and predicted values, indicated that ARIMA model successfully forecast LQ45 stock index.

Keywords: Stock Index, Technical Analysis, LQ45 Index, ARIMA

A. INTRODUCTION

The capital market is a means for investors and companies to invest. Stock becomes one of the securities that are greatly demanded by investors, considering its quite high returns, with relevant risks and high liquidity compared to other types of securities. The development of capital market shows good performance and allows investors to get abnormal returns. In the last two years, namely 2016-2017, the Indonesian Composite Index (*Indeks Harga Saham Gabungan*/IHSG) showed an increasing trend of 14.21% and 20.05% respectively. Even in the last ten years, the average increase of stock index was 10.43%.

One of the leading indices on the Indonesia Stock Exchange, the LQ45 Index, is a market capitalization value of 45 stocks with high liquidity. Overall, the members of LQ45 Index are blue chip stocks, namely stocks of major issuers which have a good reputation. Historically, they recorded a profit growth from year to year and consistently distributed dividends to shareholders. Even in 2017, most of the LQ45 issuers scored double-digit profit growth (Kontan.co.id).

The efficient market hypothesis developed by Fama gives a major contribution to investment theory. Investor behavior plays a significant role in determining market efficiency. Irrational decisions taken by investors often lead to weakening levels of market efficiency. This is because investors are unable to accurately and quickly absorb and express the information. In this condition, accurate prediction techniques are needed to estimate stock prices in the future in order to gain optimal returns. There are two types of stock analysis developed by analysts, namely fundamental analysis and technical analysis. Fundamental analysis considers the intrinsic value of the stock in terms of financial condition of the company. While technical analysis predicts stock prices based on stock prices in the past (historical performance). Technical analysis has been known as an effective technique used for short-term stock investments. Technical analysts argue that future stock prices are formed by price movements and the stock price pattern itself.

There are several types of indicator for technical analysis derived from sequential (time series) stock price data, including the Box Jenkins Model, or technically referred to as the Autoregressive Integrated Moving Average (ARIMA) model.

This study aims to form an ARIMA model of the LQ45 stock index, and analyze whether the efficient market hypothesis applies to companies listed on this index or not.

B. THEORETICAL BACKGROUND
1. Technical Analysis

According to Baskarn (2014), Technical Analysis can be defined as the art and science of forecasting future prices based on past price movements. Forecasting is done based on the demand and supply analysis of commodities, stocks, futures, or other instruments traded. Technical analysis involves all stock information, such as price and volume, then applying them in a graph, as well as analyzing the patterns, trends and indicators to assess future price movements.

Technical analysis has several advantages. It is fast, can accommodate analysis needs that are in accordance with the time horizon of each investor, is able to provide a picture of the psychological market, and has the flexibility in analyzing according to the desired time period.

Kannan and Ashik (2017) conducted a technical analysis study on stock prices on the National Stock Exchange (NSE) of India. By using a sample of 50 companies listed on the Nifty index in the period January to December 2015 or as many as 245 observations. Through the application of the Autoregressive Moving Average (ARIMA) method, the results of this study indicate that the resulting ARIMA model has a predictive strength of 94% with a low Mean Absolute Percentage Error. The suitable ARIMA for this research data is ARIMA (0,1,1) which produces the most minimum BIC value.

The research conducted by Ayodele et al. (2014) used the Box Jenkins ARIMA method on two companies, namely NOKIA which is listed on the NYSE (New York Stock Exchange) and Zenith Bank which is listed on NYE (Nigeria Stock Exchange) in the period 1995 - 2011 or as many as 3987 observations. The stock prices used were closing price data on each stock exchange. The results show that the ARIMA model is suitable for predicting stock prices in the short term.

Devi et al. (2013) conducted a study of technical analysis using the Box Jenkins ARIMA method on the National Stock

Exchange (NSE) of India. The data used was closing price data in the period January 2007 to December 2011. The selected companies were five companies with the highest market capitalization values, namely Nifty 50, Reliance, OFSS, ABB and JSW Steel. The results show that the ARIMA model has a good predictive strength for short-term predictions. The best model was shown by the forecasting of Nifty 50 stock price, with an error percentage of 16.26%.

Lilipaly et al. (2014), in his research predicting the stock price of Bank Rakyat Indonesia Tbk by using daily stock data from 2011 to October 2014, showed that the ARIMA model is suitable and can be used to predict stock prices in November 2014. ARIMA model suitable for the maximum stock price is ARIMA (2,1,3) and ARIMA (2,1,3) is suitable for the minimum stock price.

Lusikooy et al. conducted a study aimed at applying the ARIMA model to predict the closing price of PT Garuda Indonesia Tbk. stocks. This study used daily stock price closing data in the period January 1, 2013 - October 24, 2016. This data was used to predict stock prices on October 25, 2016 - November 7, 2016. The corresponding model was ARIMA (3,1,3) with the smallest MSE value of 63.7%.

Abidin et al. (2016), in his research, revealed several things related to the technical factors, namely 1) that bid volume does not have a significant effect on stock prices partially 2) past stock prices have a significant effect on the future stock prices, and 3) trading volume has a significant negative effect on stock prices. This research was conducted by taking a sample of 17 companies on the IDX30 Stock Index listed on the Indonesia Stock Exchange in the period 2012-2015. Through multiple linear regression method, the resulting model was able to provide a predictive strength of 85%.

2. Box Jenkins Model

Box Jenkins model is one of the forecasting techniques for time series model that is only based on the behavior of the observed variable data (let the data speak for themselves). The Box Jenkins model is technically referred to as the Autoregressive Integrated Moving Average (ARIMA) model. ARIMA modeling technique is one indicator that is widely used for forecasting and is a method that produces predictions based on the synthesis of historical data patterns

(Widharjono, 2013). ARIMA completely ignores independent variables because this model uses the present values and past values of the dependent variable to produce accurate short-term forecasting. ARIMA requires relatively large data. According to a number of recommending literature, it requires a minimum of 72 data from a series. In various studies, ARIMA was proven to provide short-term predictions with high accuracy. In addition, the ARIMA method is also more applicable because it accommodates data that is not stationary at the level, but is stationary at the level of differentiation.

The first discussion of the ARIMA model will begin with the autoregressive (Autoregressive = AR) model. The AR model shows that the predictive value of the dependent variable Y_t is only a linear function of the previous actual Y_t velues. In general, the form of an autoregressive (AR) general model can be expressed in the following equation:

$$Y_t = \beta_0 + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \cdots + \beta_p Y_{t-p} + e_t \qquad (2.1)$$

Where: Y = dependent variable;
$Y_{t-1}, Y_{t-2}, \beta_1 Y_{t-p} = $ the lag of Y ;
p = the degree of AR
e_t = residual (error)

The residual in equation 2.1, as in the Ordinary Least Squares (OLS) model, has a characteristic of zero average value, a constant variant and not interconnected. Thus, the AR model shows that the predictive value of Y_t variable is only a linear function of the previous actual Y_t values.

The second ARIMA model is Moving Average (MA). This model states that the predictive value of Y_t variable is only influenced by the residual value of the previous period. In general, the form of Moving Average model can be expressed in an equation as follows:

$$Y_t = \alpha_0 + \alpha_1 e_{t-1} + \alpha_2 e_{t-2} + \alpha_3 e_{t-3} + \cdots \alpha_q e_{t-q}$$

Where
e_t = residual
$e_{t-1}, e_{t-2}, e_{t-q}$ = lag of the residual;
q   = the degree of MA

Often, the behavior of time series data can be explained well through the incorporation of the AR model and the MA Model. This

combined model is called as the Autoregressive-Moving Average (ARMA) model. The ARMA model is a model where the value of dependent variable Y_t is influenced by the first lag of Y_t and the first-degree lag of residual. In general, the form of ARMA model can be written in the form of an equation as follows:

$$Y_t = \beta_0 + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \cdots + \beta_p Y_{t-p} + \alpha_0 e_t + \alpha_1 e_{t-1} + \alpha_2 e_{t-2} + \cdots + \alpha_q e_{t-q}$$

The AR, MA and ARMA models previously required that time series data should have stationary properties. But in fact, time series data are often not stationary but are stationary in the differentiation process. The model with stationary data through differencing process is called ARIMA model. Thus, if a stationary data with a d times differencing process applies ARMA (p, q), then the resulting model is ARIMA (p, d, q), where p is the degree of AR, d is the degree of the differencing process and q is the degree of MA.

Simply put, the stationarity test is done by looking at the correlogram through the Autocorrelation Function (ACF). The ACF explains how much the correlation of sequential data is in terms of time. If the ACF value for each lag is zero, then the data is stationary. As a rule of thumb, the length of a lag is one third or one quarter of the time series data tested.

After detecting the data stationary problems, the next is identification of the ARIMA model. The standard methods used for selecting the ARIMA model through correlogram are the Autocorrelation Function (ACF) and the Partial Autocorrelation Function (PACF). PACF is defined as a correlation between Y_t and Y_{t-k} after removing the effect between Y which lies between the Y_t and Y_{t-k}.

Evaluation of the ARIMA model is done by using the value of Root Mean Square Error (RMSE). The smaller the RMSE value, the better the model produced.

C. RESEARCH METHOD

This Research is categorized as basic research, because the objective of the research is to developing and evaluating theretical concepts. Based on the types of data, this is a quantitative research, since it using numeric data for decision making.

The data were used is secondary data, that is daily stock price index. The population of this research is shares that incorporated in LQ45 Stock Index. The Data were taken from period of Januari 2016 to November 2018.

Data Analysis process involved these following stages:
1. Stationarity Test
   The objective of this test is to see whether time series data meet the stationaire requirement or not. This test will be done by visual analysis thorugh time series plot. If data not stationaire, it needs to differenced until stationaire.
2. Determination of ARIMA
   Determination of ARIMA model is started by determination of p,d, and q orders.
   In this research, the combination of p,d, and q orders will be tested to find the best model for data. The goodness of fit model will be determine based on Root Mean Square Error (RMSE) Value. The lowest RMSE value is the best ARIMA model which will be used for predicting LQ45 Stock Index.
3. Stock Index Prediction Values
   Best Model that obtained from previous stage then used to predicting predicted values for the next 15 days.
4. Paired samples t-test
   Furthermore, paired samples t-test will be run to see the differences between actual and predicted values. If the results shows that there is no significance differences between two samples, it can be concluded that time series model predicting LQ45 daily stock index successfully.

D. RESULTS AND DISCUSSIONS
### 1. Determination of ARIMA Model

Daily LQ45 stock price index data was done through a time series plot according to the research period, which was January 2016 - November 2018.
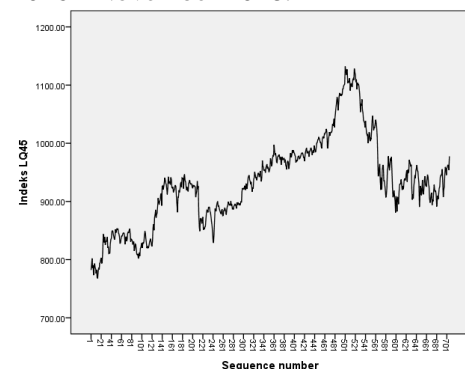

Picture 1. Time Series Plot
Source: Secondary Data, 2018

Based on Figure 1, it can be seen that the LQ45 stock price index during the study period was very volatile. This figure indicates that the data had not been stationary. Next, data plotting was done using correlogram and produced a result which is presented in Figure 2.
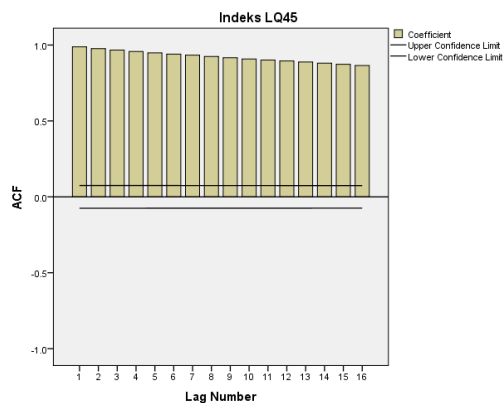
Picture 2 The Correlogram of ACF at level

Based on the correlogram plot of ACF, it can be seen that up to the 24th lag, the histogram did not quickly reach the zero value, so it can be concluded that the data had not been stationary. Therefore, differencing process was performed to make data stationary. Figure 3 shows the ACF plot results after the differencing process.
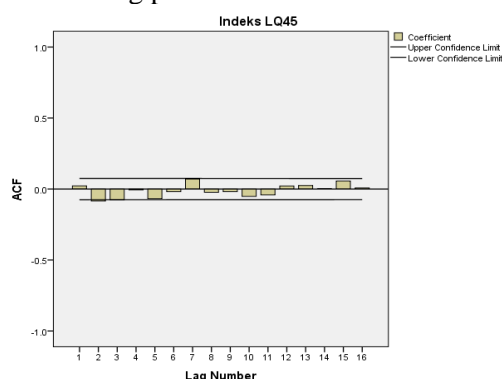

**Picture 3 Correlogram of ACF after First Differencing**

The results of correlogram in Figure 3 show that after the second lag, the level of significance was getting closer to zero, which means that the data had been stationary. So it can be concluded that the data was stationary after the first degree of differencing process had been done. The 'one' number then became the d order in ARIMA modeling.

Determination of the process of p and q orders in the ARIMA model was done through the plots of Autocorrelation (ACF) and Partial Autocorrelation (PACF). Figures 4 and 5 are the results of correlogram ACF and PACF plots.
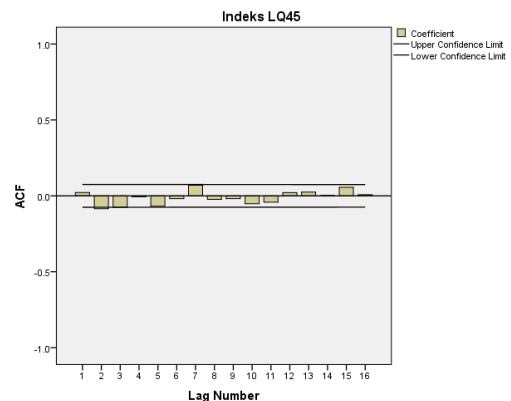
Picture 4 Correlogram of ACF for Orde q determination
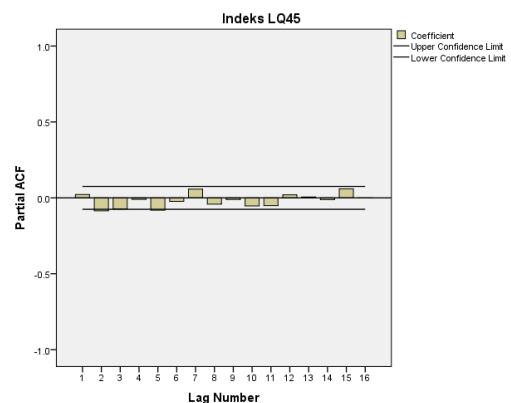
Picture 5 Correlogram of PACF for orde p determination

The figure of ACF plot above shows that there is one significant bar, which is the bar of q order. Whereas in the PACF plot, there are two significant bars which then became the p orders. So that the appropriate ARIMA tentative model for JII stock index data is ARIMA (2,1,1). However, to ensure that the model produced is the best model, modeling using several combination models with order d=1 was done. Modeling by using several combinations of p and q values at the degree of differencing 1 was conducted to determine the AR (p) and MA (q) orders. The modeling results along with the value of Root Mean Square Error (RMSE) for all models are presented in table 1.

Tabel 1 RMSE values for each tentative models

PARAMETER	RMSE
(1,1,0)	10.010
(1,1,1)	10.019
(2,1,1)	9.946
(3,1,1)	9.953
(1,1,2)	9.947
(1,1,3)	9.955
(2,1,0)	9.980
(2,1,2)	9.953
(2,1,3)	9.957
(3,1,0)	9.961
(3,1,2)	9.957
(3,1,3)	9.902

Based on table 1, there are twelve ARIMA models that were differentiated once. They are in accordance with the theory that the smaller the value of Root Mean Square Error (RMSE), the better the model. In table 1, it can be seen that the ARIMA models with order p = 2, d = 1, and q = 1 are the best model with the smallest RMSE values. Thus it can be concluded that the time series model for the LQ45 index data is ARIMA (2,1,1).

This following table shows the goodness of fit model for ARIMA (2,1,1).

Tabel 2 Significancy of model coefficients for ARIMA (2,1,1)

| Variable	Coefficient	Significance
Constant	0.268	0.340
AR (1)	0.748	0.000
AR(2)	-0.102	0.013
MA(1)	0.736	0.000

Based on modelling result shows on table 2, coefficients values can be substituted into ARIMA (2,1,1) equation as follows : $y_t = \left(1 - (0.748) - (-0.102)\right)0.268 + \left(1 + (0.748)\right)y_{t-1} + \left((-0.102) - (0.748)\right)y_{t-2} - (-0.102)y_{t-3} + 0.736e_{t-1}$

Thus, the equation model ARIMA (2,1,1) for predicting future LQ45 stock index is:

$$y_t = 0.095 + 1.748y_{t-1} - 0.85y_{t-2} + 0.102y_{t-3} + 0.102e_{t-1}$$

The value of R-squared for the resulting model is 98.3%, which means that the resulting model provides high predictive accuracy, or, in other words, the independent variable has a relatively high ability to explain the dependent variable. In the context of time series data, this means that past stock price data on the LQ45 index can be used to predict future stock prices with fairly high predictive capability.

After obtaining a model that has a significant independent variable and a high r-squared value, it is necessary to know whether the model has fulfilled the white noise requirements based on the diagnostic test. For this reason, residual data was used to form the correlogram for its ACF and PACF as shown in Figures 6 and 7 below.
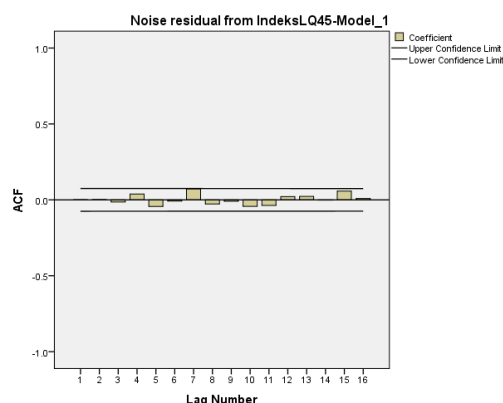
Picture 6 Diagnostic Test of ACF residual for model ARIMA (2,1,1)
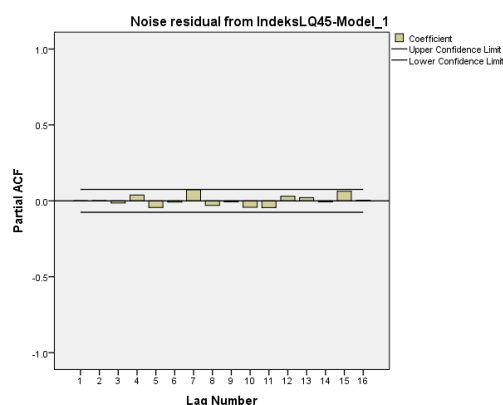
Picture 7 Diagnostic Test of PACF residual for model ARIMA (2,1,1)

Figures 6 and 7 show that all graph bars are inside the Bartlett line, so that it can be said that the nature of the residuals is white noise and it is concluded that the model obtained is quite good.

The whole test concludes that the LQ45 stock price index data for the period January 2016 - November 2018 is good and suitable to predict the future stock price index. Furthermore, the forecasting of the LQ45 stock price index for several future periods was done using the ARIMA (2,1,1) model which has been produced.

Table 3 shows the predicted values of LQ45 daily stock index starting from December 1 to December – 15 2018.

Table 3 Predicted Values of LQ45 Daily Stock Index

Date	Stock Index
December 1 2018	592.22
December 2 2018	597.47
December 3 2018	612.54
December 4 2018	598.76
December 5 2018	600.05
December 6 2018	586.51
December 7 2018	596.99
December 8 2018	603.04
December 9 2018	594.45
December 10 2018	594.34
December 11 2018	586.75
December 12 2018	592.15
December 13 2018	582.92
December 14 2018	582.58
December 15 2018	591.81

2. Paired Sample T-Test

Furthermore, on the results of the LQ45 stock price index prediction, paired sample t-tests were carried out between the actual data and the predicted results. The following Table 4 shows the probability level of the paired t-test results of the two samples.

Table 4 The Result of Paired t-Test

Pair	Significance	Conclusion
Actual and Predicted	0.988	Insignificant

Based on the results of paired t-test in Table 4, it can be seen that there is no significant difference between the predicted data and the actual data. This shows that the ARIMA model was successfully used to predict the LQ45 Stock Price Index.

E. CONCLUSION AND SUGGESION
Conclusion

Analysis that has been done for shares incorporated in LQ45 Stock Index shows that previous stock index can be used to predicting future stock index.

After testing twelve tentative ARIMA model at first differencing, the appropriate model is ARIMA (2,1,1) with this following equation:

$$y_t = 0.095 + 1.748y_{t-1} - 0.85y_{t-2} + 0.102y_{t-3} + 0.102e_{t-1}$$

This model provide predictive strength of 98.3% dan Root Mean Square Error Value of 9.946.

Model ini menghasilkan kekuatan prediksi sebesar 98.3% dan RMSE sebesar 9.946.

This model were chosen as best model because of Fairly high R-Squared, the minimum value of RMSE, and supported by paired t-test result which is shows that there is no significance differences between actual and predictive value,
For LQ45 stock Index, it is proven that technical analysis can be used to stock price prediction. Investors can use ARIMA modelling or other relevant technical analysis.

Suggesion

Based on the result of the research that has been done, it can be concluded that time series modelling can be used to predict LQ45 stock index, considering its high accuracy. The seuccesfully of ARIMA modellingshows that stock index's movement in Indonesia does not follow random walk theory, but can be predicted empirically. This is also proven that Indonesia's capital market having weak efficiency.

In this case, investor can use technical analysis to support their investment's decision making. Besides ARIMA, investor can utilize other available technical indicator to determine whether needs to sell or buy. However, ARIMA is a time series modelling which is accurate only for short term prediction, thus investor needs to consider other related factors for long term investment.

F. REFFERENCES

Abidin, Sugeng, S. Hidayat, and R. Rustam. 2016. Pengaruh Faktor-Faktor Teknikal Terhadap Harga Saham (Studi Pada Harga Saham IDX30 di Bursa Efek Indonesia Periode 2012-2015). Jurnal Administrasi Bisnis (JAB) Vol 37 No.1, pp 21-27

Artha, D.R., Achsani, N.A., & Sasongko, H. 2014. Analisis Fundamental, Teknikal dan Makroekonomi Harga Saham Sektor Pertanian. Jurnal

Manajemen dan Kewirausahaan Vol 16 No 2, pp 175-184

Ayodele, A.A.& Odewumi, Aderemi. 2014. *Stock Price Prediction Using the ARIMA Model.*UKSIM-AMSS International Conference on Computer Modelling and Simulation. Pp 105-111

Devi, B., Sundar D., & Alli. 2013. An Effective Time Series Analysis for Stock Trend Prediction Using ARIMA Model for Nifty Midcap 50. International Journal of Data Mining and Knowledge Management Process (IJDKP) Vol 3 No 1, pp 65-78

Lilipaly, S. G., Hatidja, D., Kekenusa, J.S. 2014. Prediksi Harga Saham PT BRI Tbk menggunakan Metode ARIMA (Autoregressive Integrated Moving Average). Jurnal Ilmiah Sains, Vol 14 No 2, pp 60-67

Lusikooy, J., Nainggolan, N., & Julia, T. 2017. Prediksi Harga Tutup Saham PT Garuda Indonesia Menggunakan Metode ARIMA. Jurnal MIPA UNSRAT ONLINE 6(1) 74-77

Mondal, P., Shit, L., & Goswami, S. 2014. International Journal of Computer Science, engineering and Applications (IJCSEA) Vol 4, No.2, pp 13-29

Ramadhan, Bayu Ariesta. 2015. Analisis Perbandingan Metode ARIMA dan Metode GARCH untuk Memprediksi Harga Saham. E-Proceeding of Management Vol. 2 No.1,pp 61-68

Subashini & Karthikeyan. 2018. Forecasting on Stock Market Time Series Data Using Data Mining Techniques. International Journal of Engineering Sciense Invention (IJESI), pp 6-13

Zulkarnain, Iskandar. 2015. Akurasi Peramalan Harga Saham dengan Model ARIMA dan Kombinasi Main Chart dan ICHIMOKU Chart. Management Insight 7(1), pp 59-70

Deccasari, MDD. 2015. Penerapan Analisis Teknikal dengan Metode Bollinger sebagai Salah Satu Indikator dalam Transaksi Short Time Perdagangan Saham (Studi pada PT E-Trading Securities Malang). Jurnal Dinamika Dotcom Vol. 5 No. 1, pp 64-79

Tseng, Kuo-Cheng., Kwon, O., Tjung C.L. 2012. Time Series and Neural Network Forecast of Daily Stock Prices. Investment Management and Financial Innovation, Vol.9 Issue 1, pp 32-54

Widarjono, Agus. 2013. Ekonometrika Pengantar dan Aplikasinya. UPP STIM YKPN. Yogyakarta