The Application of the Arima Box-Jenkins Method on Forecasting the Monthly Closing Stock Price per Share of Jakarta Stock Exchange (JKSE)

Priscilia[1*], Agus Sofian Eka Hidayat[2]

[1,2] *Study Program of Actuarial Science, Faculty of Business, President University, 17550, Indonesia*

[1]*Corresponding author: priscilia@student.president.ac.id*

Abstract— Jakarta Stock Exchange (JKSE) or Indonesia Composite Index (ICI) measures the overall listed stock prices on the Indonesia Stock Exchange. This study aims to provide the closing stock price forecast of the monthly JKSE closing price per share from March 1st, 2022 to March 1st, 2023, using the ARIMA Box-Jenkins method, and compare the accuracy of forecasting 12 months of the dataset with forecasting 6 months of the dataset. The study contributes to the body of knowledge on stock market forecasting and provides investors with valuable information to make informed investment decisions**.** The study has found the best model to use is ARIMA(1,2,1) which is considered as the best model since it satisfies both Saphiro and L-jung box test, as well as having the lowest AIC among all of the models. This study also shows that its accuracy will be better by forecasting less ahead to the future. Whereby forecasting 12 months ahead resulted to 124817.291 MSE, 353.295 RMSE, 1428221815 MAE, and 4.265 MAPE. Meanwhile, forecasting 6 months ahead resulted to 34385.384 MSE, 185.433 RMSE, 1415348102 MAE, and 2.206 MAPE. The study contributes to the body of knowledge on stock market forecasting and provides investors with valuable information to make informed investment decisions.

Keywords— JKSE; forecast; model; data; ARIMA

I. INTRODUCTION

The Jakarta Stock Exchange (JKSE), also known as the Indonesia Composite Index (ICI) or IDX Composite, is an index that measures the overall listed stock prices on the Indonesia Stock Exchange [1]. It displays the average movement of listed stock shares. As a result, most of the shares listed on JKSE tend to rise when JKSE does. On the other hand, when the index is weak, the stock prices could fall. However, it's important to note that this value is an average and that there may be outlier stocks with values that deviate dramatically from those of the JKSE. JKSE can be used to measure the performance of a stock portfolio, which is a collection of stock investments owned by individuals or companies. By monitoring the performance of JKSE, investors can estimate the profits from their stock portfolio. JKSE can also serve as an indicator to monitor the development of a country's economic conditions, such as capital flows, economic growth, and tax revenues. JKSE plays a significant role, as higher investments in a country result in larger capital flows. A healthy capital flow can stimulate the economy, leading to improved economic growth and increased tax revenues paid by companies. The government can then utilize these tax revenues to create new policies to improve society's welfare.

Once the functions of JKSE are understood, it's important to learn how to interpret them. JKSE charts can be accessed through investment apps or websites. When JKSE is trending upwards, it's commonly referred to as being "green" or "bullish". During such times, investors are advised to sell stocks to gain profits. Alternatively, investors may choose to hold onto their stocks with the hope that prices will continue to increase, leading to even larger profits. However, it's essential to be cautious and not be greedy when seeing potential profits during a rising JKSE, as the index may experience a bubble and be prone to corrections. Forecasting stock prices can provide investors with valuable insights for informed decision-making, risk management, the timing of trades, portfolio optimization, and strategic planning. Accurate stock price forecasts allow investors to make informed investment decisions, manage risks, optimize their portfolios, and strategically plan their long-term investment approach. With concerns about a potential global GDP slowdown in 2023, investing in the stock market can be a strategic approach for investors to potentially mitigate the impact. By carefully selecting stocks based on accurate forecasts and market trends, investors may be able to identify opportunities for growth and capital appreciation, potentially offsetting the effects of a slowing global economy. Hence, we will provide the closing stock price forecast of the monthly JKSE closing price per share from March, 1st 2022 to March, 1st 2023 using the ARIMA Box-Jenkins method, using the data of JKSE closing stock price from March, 1st 2017 to March, 1st 2022. And we will compare the accuracy of forecasting 12 months of the dataset with forecasting 6 months of the dataset.

Studies from a journal titled "Analisis Peramalan IHSG dengan Time Series Modeling ARIMA (Analysis of Indonesia Composite Index (IHSG) Forecasting with Arima Time Series Modeling)" by Riana Susanti and Askardiya Radmoyo Adji [2] showed that the ARIMA model which has the best performance to predict JKSE was ARIMA model (7,3,1). The studies use daily closing stock price per share of JKSE from the period of January 2, 2017, to January 3, 2018. The result has RMSE of 30.33293, MAE of 22.99950, and MAPE of 0.002615. In conclusion, JKSE forecast results with ARIMA model (7,3,1) are not much different from the actual value of JKSE closing stock price. Studies from a journal titled "The Autoregressive Integrated Moving Average (ARIMA) Model for Predicting Jakarta Composite Index" by Didik Gunawan and Weni Astika [3] showed that the ARIMA model that has the best performance to predict JKSE was the ARIMA model (3,1,9). The studies use daily closing stock price per share of JKSE from the period of January 2020 to April 2021, so the sample in this study uses 324 time-series data. The result has a MAPE of 1.729. Studies from a journal titled "Analisis Model Autoregressive Integrated Moving Average Data Deret Waktu Dengan Metode Momen Sebagai Estimasi Parameter" by Santi Deviana, Nusyirwan, Dorrah Azis, and Pandri Ferdias [4] showed that the ARIMA model that has the best performance to predict JKSE was the ARIMA model (1,2,0) by testing the significance of the parameters and testing the suitability of the model. The studies use monthly stock price per share of JKSE from the period of January 2013 to June 2020, so the sample in this study uses 90 time-series data.

II. METHODOLOGY

The most essential first step of doing ARIMA modelling is to collect the data that are about to be forecasted. The forecasted data stationarity will then be identified. The stationarity test in this study will be using Augmented Dickey-Fuller (ADF) with its p-value being the focus of the test. The data will go through differencing until the p-value is equal to or lesser than 0.05. Using the autocorrelation and partial autocorrelation plots, the order of the autoregressive (AR) function, lag p, and moving average (MA) function, lag q can be determined. The ARIMA model can then be created as ARIMA(p,d,q). The ARIMA model will then be fitted into the data and assess its performance using measures like MAE, MSE, and RMSE. If required, the model will be refined by modifying the order of AR and MA and/ or the degree of differencing. The refined model will also be fitted into the data and compared with the other models to determine the best model. In this study, the best model will be determined by comparing the parameter estimation results and error parameter estimation results. Finally, the best model will be used to do forecasting. The ARIMA flow chart diagram [5] is shown in Figure 1.

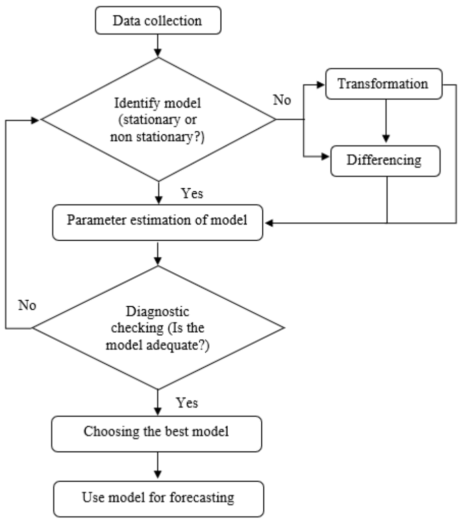

Figure. 1 ARIMA flow diagram

III. ANALYSIS AND DISCUSSION

A. Data Preparation

The data used in this analysis comprises the monthly closing stock price of the Jakarta Stock Exchange (JKSE) obtained from Yahoo Finance from the period of March 1, 2017, to March 1, 2022. A total of 61 datasets are then downloaded, processed, and analyzed as '^JKSE1.csv' in R Studio.

The main packages needed in this analysis are ggplot2, forecast, TSA, and tseries. First, ggplot2 is a package used to perform data visualization which will be helpful in this analysis to identify patterns such as trend and seasonality, as well as to show any uncertainty in the forecast. Second, the package 'forecast' is used to provide 'accuracy()' function. The function 'accuracy()' is used to provide forecast accuracy measurements, such as mean absolute error (MAE), mean squared error (MSE), and root mean squared error (RMSE). Third, the package 'TSA' stands for "Time Series Analysis" which provides functions such as 'arima()', 'acf()', pacf()", Box.test()', 'tsdiag()', and 'auto.arima()'. The function 'arima()' is the main function for fitting an ARIMA model to a time series. The function 'acf()' and 'pacf()' are used to plot the autocorrelation function (ACF) and partial autocorrelation function (PACF) of a time series, which will be useful for identifying the order of the AR (Auto-Regression) and

MA (Moving Average) terms in ARIMA model. The function 'Box.test()' is used to perform a Box-Ljung test for the presence of autocorrelation in the residuals of an ARIMA model, it is used to check the adequacy of an ARIMA model. The function 'tsdiag()' is used to produce a diagnostic plot for an ARIMA model including the plots of ACF, PACF of the residuals, as well as a histogram and a normal probability plot of the residuals. Fourth, the package 'tseries' provides 'adf.test()' function. The function 'adf.test()' is used to perform an Augmented Dickey-Fuller (ADF) test for the presence of a unit root in a time series to determine whether a time series is stationary enough to fit an ARIMA model.

The data '^JKSE1.csv' which has been obtained from Yahoo Finance will be assigned into a variable called 'data' and be read by R using 'read.csv()' function. The variable 'data' will then be assigned to create a new data frame called 'df' consisting of 'Date' which indicates dates and 'Close" which indicates the monthly closing price per share of JKSE from the 'data' variable. The first six rows of the data frame will be displayed by the function 'head()' to set a quick preview of the data structure. The data frame will then be created as a time series plot using 'ggplot2' package.

The following figure 2 shows the original plot of monthly closing price per share of JKSE, while figure 3 shows its time series plot using ggplot2 package.

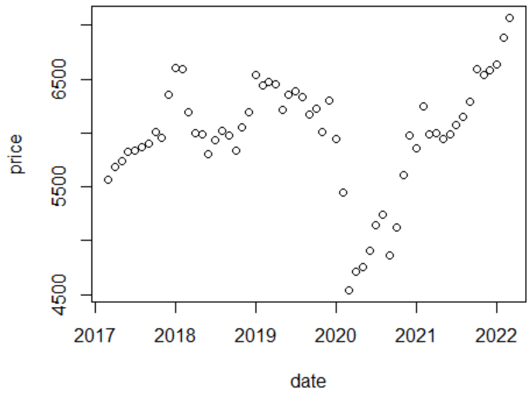

Figure. 2 Original Plot of monthly closing price per share of JKSE

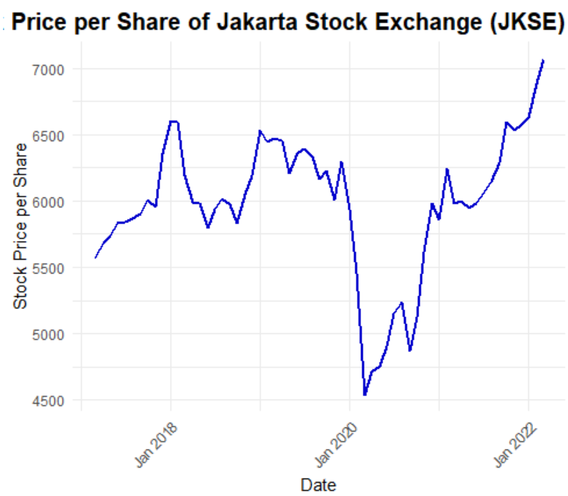

Figure. 3 Time Series Plot of monthly closing price per share of JKSE

B. Stationarity Check

After creating a time series plot, the data will then go through a stationarity check using Augmented Dickey-Fuller (ADF) test. The data will keep going through differencing if the p-value is not equal to or less than 0.05. The amount of differencing the data went through will represent the ARIMA order of d.

```
> #Stationarity check
> adf.test(data$Close)

        Augmented Dickey-Fuller Test

data:  data$Close
Dickey-Fuller = -1.2636, Lag order = 3, p-value = 0.8726
alternative hypothesis: stationary
```

Figure. 3 Stationarity check using ADF test

Since the p-value is greater than 0.05, the first differencing is needed. The first differencing is shown in Figure 5.

```
> #First differencing
> df1=diff(data$Close)
> adf.test(df1)

        Augmented Dickey-Fuller Test

data:  df1
Dickey-Fuller = -3.4761, Lag order = 3, p-value = 0.05235
alternative hypothesis: stationary
```

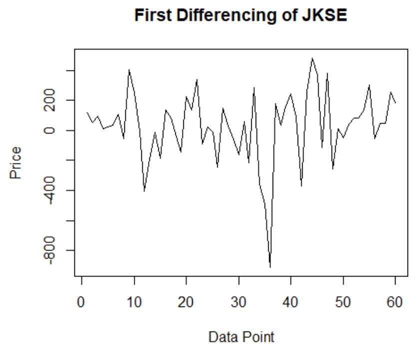

Figure. 4 First differencing output in console and as a plot

Since the p-value is still greater than 0.05 after first differencing, then a second differencing is needed. The second differencing is shown in Figure 6.

```
> #Second differencing
> df2=diff(diff(data$Close))
> adf.test(df2)

        Augmented Dickey-Fuller Test

data:  df2
Dickey-Fuller = -5.4174, Lag order = 3, p-value = 0.01
alternative hypothesis: stationary
```

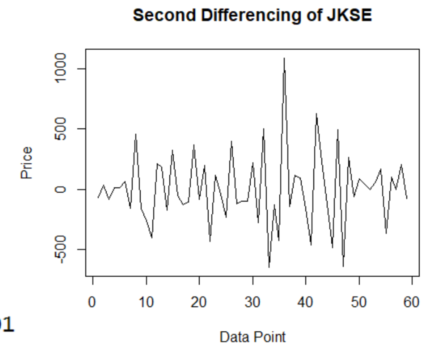

Figure. 6 Second differencing output in the console and as a plot

C. ARIMA Model Specifications

After the order d has been found and the data has become stationer, the order of the autoregressive (AR) function, lag *p*, and moving average (MA) function, lag *q*, can be determined by Partial Autocorrelation Function (PACF) and Autocorrelation Function (ACF) plot. Order p can be determined by selecting the first spike on the PACF plot that extends beyond the confidence level. The confidence level is represented by the blue dashed line in the plot. The same goes for order q. Order q can be determined by selecting the first spike on the ACF plot that extends beyond the confidence level.

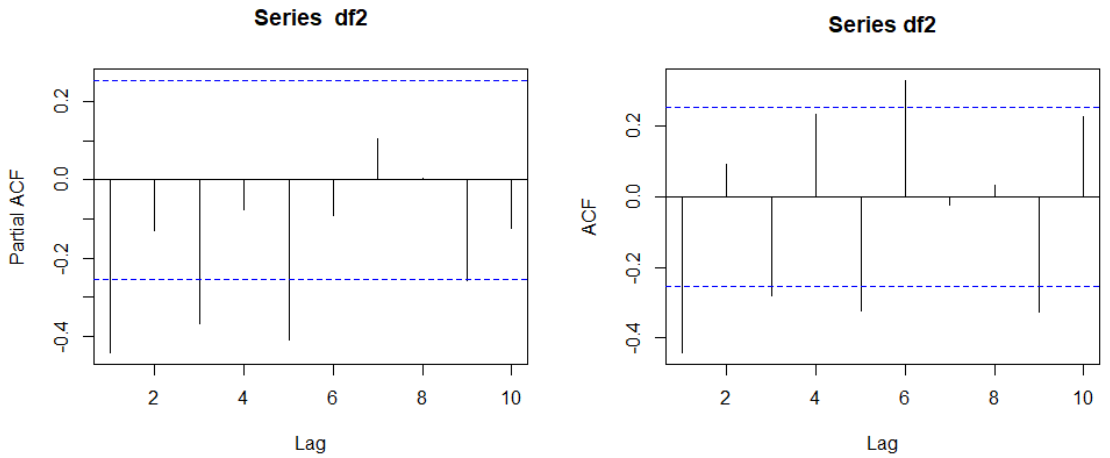

Figure. 7 PACF and ACF plots.

Both PACF and ACF plots in Figure 6 show that the plots experience first spike at lag 1, which means that both cut offs happen at lag 1. Therefore, this data has ARIMA(1,2,1) with the equation

$$Y_t = (\phi_1 + 2)Y_{t-1} + (1 + 2\phi_1)Y_{t-2} + \phi_1 Y_{t-3} + e_t - \theta_1 e_{t-1} \tag{1}$$

Some ARIMA models that can be built are shown in Table 1.

TABLE 1
ARIMA MODELS GENERATED BY THE ARIMA ORDERS OBTAINED

ARIMA Model	**P**	**D**	**Q**
ARIMA(1,2,1)	1	2	1
ARIMA(1,2,0)	1	2	0
ARIMA(1,1,1)	1	1	1
ARIMA(1,1,0)	1	1	0
ARIMA(0,1,1)	0	1	1

D. Parameter Estimation

Estimation of parameters ϕ_p of autoregressive (AR) function and θ_q of moving average (MA) function in each model can be found by looking at the summary of each model. Table 2 shows a compiled output information of each model consisting of parameter estimations and its mean squared error (MSE), log-likelihood, and AIC.

TABLE 2
PARAMETER ESTIMATION RESULTS

ARIMA Model	**AR1**	**MA1**	**MSE**	**Log Likelihood**	**AIC**
ARIMA(1,2,1)	0.1866	-1	54533	-407.32	818.65
ARIMA(1,2,0)	-0.4355	-	74644	-414.83	831.65
ARIMA(1,1,1)	0.2561	-0.0831	54036	-412.07	828.15
ARIMA(1,1,0)	0.1749	-	54071	-412.09	826.19
ARIMA(0,1,1)	0.1454	-	54333	-412.23	826.47
ARIMA(2,0,2)	-0.0246	1.2126	54533	-415.34	842.68

E. Residual Analysis

To determine which ARIMA model gives the best prediction, the normality test with Saphiro and the white noise test with Ljung-Box are needed. The model with a p-value of greater than 0.05 in both tests and low AIC will be taken as best model. Table

3 shows a compiled output information of each model consisting of its p-value in Saphiro test, Ljung-Box test, AIC, accept or reject, and description to show why it is being accepted or rejected.

TABLE 3
PARAMETER ESTIMATION RESULTS

ARIMA Model	Saphiro Test	Ljung-Box Test	AIC	Accept/Reject	Description
ARIMA(1,2,1)	0.0625	0.8615	818.65	Accept	Best Model
ARIMA(1,2,0)	0.06373	0.6135	831.65	Accept	Highest AIC
ARIMA(1,1,1)	0.01318	0.9484	828.15	Reject	Saphiro < 0.05
ARIMA(1,1,0)	0.01195	0.9174	826.19	Reject	Saphiro < 0.05
ARIMA(0,1,1)	0.007654	0.9159	826.47	Reject	Saphiro < 0.05

Since model 2 ARIMA(1,2,0) has the highest AIC value, model 1 ARIMA(1,2,1) will be picked as the best model among all of the models. Hence, the equation of model 1 ARIMA(1,2,1) is

$$Y_t = 2.1866Y_{t-1} + 1.3732Y_{t-2} + 0.1866Y_{t-3} + e_t + e_{t-1} \quad (2)$$

The plot of model 1 ARIMA (1,2,1) is shown in figure 8.

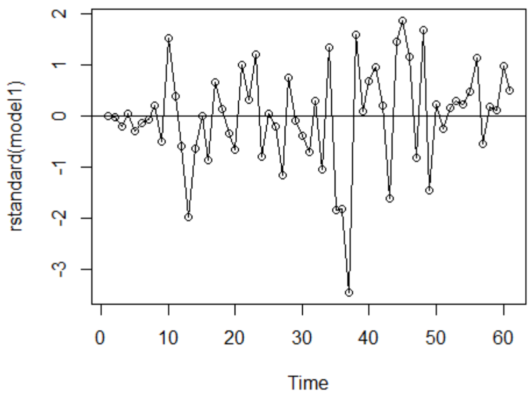

Figure. 8 Model 1 ARIMA(1,2,1) plot

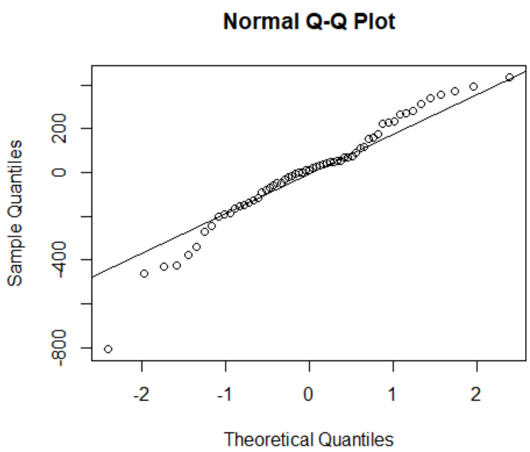

Figure. 9 Model 1 ARIMA(1,2,1) Q-Q Normal Plot

Figure 9 proved that Model 1 ARIMA(1,2,1) follows a normal distribution by showing the points on the graph approach a straight line. Figure 10 shows the plots of both residual PACF and ACF from model 1.

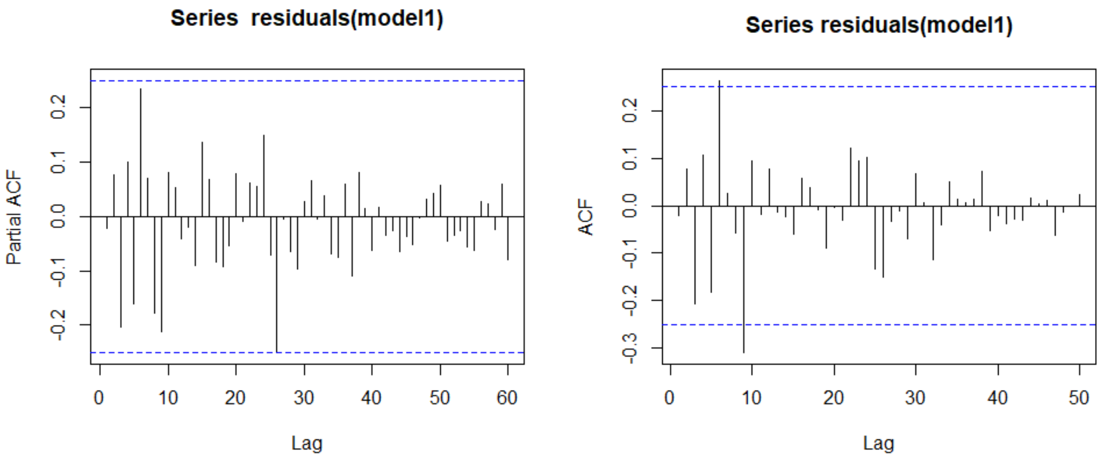

Figure. 10 Model 1 PACF and ACF Plots

F. Forecasting

The following figure 11 is the plot of the forecasted monthly stock price per share of Jakarta Stock Exchange (JKSE) for 12 months, starting from March 1, 2022, to March 1, 2023. While the following figure 12 is the plot of the forecasted monthly stock price per share of Jakarta Stock Exchange (JKSE) for 6 months, starting from March 1, 2022, to August 1, 2023. The blue line represents the predicted data, while the gray line represents the actual data.

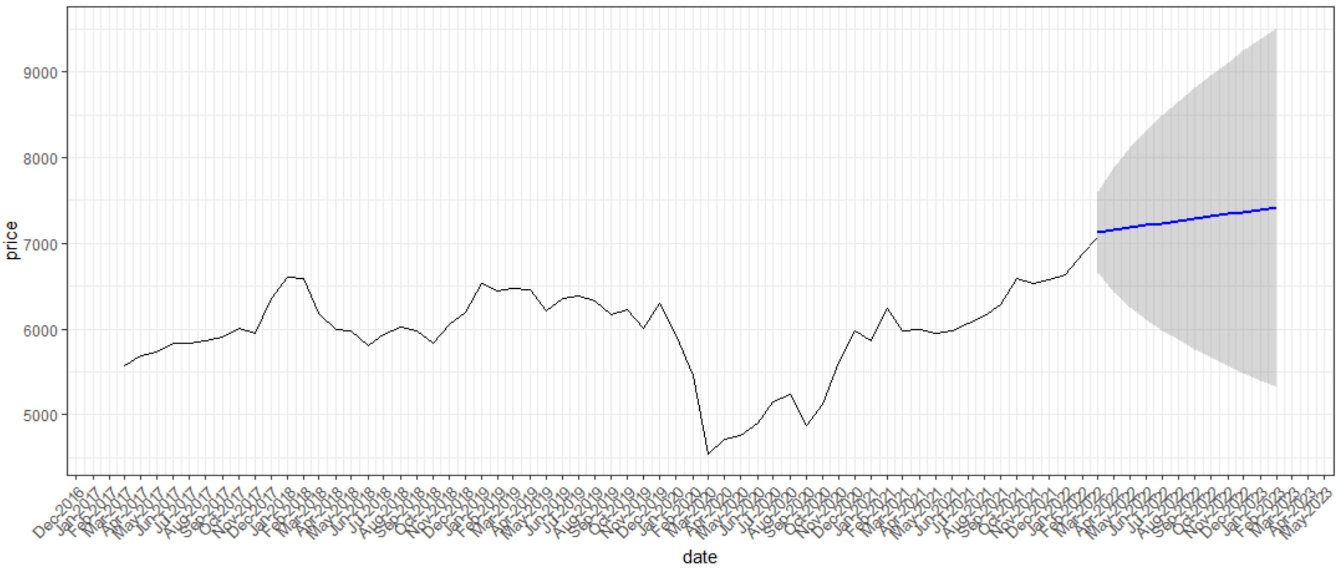

Figure. 10 12 Months Forecast Result

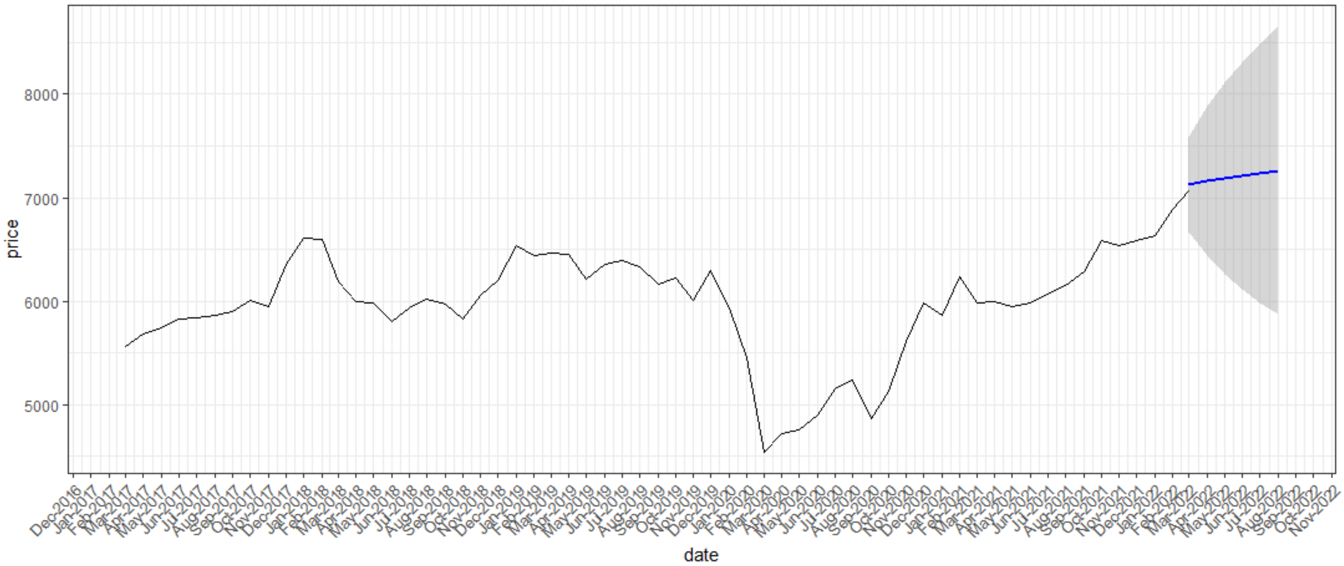

Figure. 11 6 Months Forecast Result

G. Comparison

Table 4 and 5 below will show the comparison between forecasting closing stock price per share of JKSE by 12 and 6 months in the future with four error value estimation parameters including mean squared error (MSE), root mean squared error (RMSE), mean absolute error (MAE), and mean absolute percentage error (MAPE). The error value estimation parameters are calculated by the formula given in literature review and applied to Microsoft Excel.

TABLE 4
ERROR VALUE ESTIMATION PARAMETERS RESULTS FOR 12 MONTHS FORECAST

Date	Actual	Forecasted	Error	RMSE	Abs. Error	Abs. Percent Error
01-03-22	7228.914	7126.787	1398168091	102.127063	1398168091	1.412758017
01-04-22	7148.97	7158.266	1406477243	9.295785	1406477243	0.130029707
01-05-22	6911.582	7185.292	1426604474	273.709969	1426604474	3.960163791
01-06-22	6951.123	7211.488	1425959351	260.364953	1425959351	3.745653058
01-07-22	7178.59	7237.53	1411084910	58.940156	1411084910	0.821054793
01-08-22	7040.798	7263.542	1423794544	222.744148	1423794544	3.163620838
01-09-22	7098.89	7289.549	1421750721	190.658863	1421750721	2.685755933
01-10-22	7081.313	7315.555	1425340883	234.242012	1425340883	3.307889545
01-11-22	6850.619	7341.56	1445169182	490.940859	1445169182	7.166372103
01-12-22	6839.342	7367.566	1448309234	528.224203	1448309234	7.723319271
01-01-23	6843.239	7393.572	1450372869	550.33323	1450372869	8.041999534
01-02-23	6805.277	7419.577	1455630283	614.300145	1455630283	9.026820776
			124817.291	353.295	1428221815	4.265
			MSE	**RMSE**	**MAE**	**MAPE**

TABLE 5
ERROR VALUE ESTIMATION PARAMETERS RESULTS FOR 6 MONTHS FORECAST

Date	Actual	Forecasted	Error	RMSE	Abs. Error	Abs. Percent Error
01-03-22	7228.914	7126.787	1398168091	102.127063	1398168091	1.412758017
01-04-22	7148.97	7158.266	1406477243	9.295785	1406477243	0.130029707
01-05-22	6911.582	7185.292	1426604474	273.709969	1426604474	3.960163791
01-06-22	6951.123	7211.488	1425959351	260.364953	1425959351	3.745653058
01-07-22	7178.59	7237.53	1411084910	58.940156	1411084910	0.821054793
01-08-22	7040.798	7263.542	1423794544	222.744148	1423794544	3.163620838
			34385.384	185.433	1415348102	2.206
			MSE	**RMSE**	**MAE**	**MAPE**

IV. CONCLUSION

Based on the analysis from the previous chapter on the historical data of the monthly closing stock price per share of Jakarta Stock Exchange (JKSE) from March 1st, 2022 to March, 1st 2023 shows that the best model is model 1 ARIMA(1,2,1) with the equation

$$Y_t = 2.1866Y_{t-1} + 1.3732Y_{t-2} + 0.1866Y_{t-3} + e_t + e_{t-1} \tag{3}$$

Model 1 ARIMA(1,2,1) is considered as the best model since it satisfies both Saphiro and L-jung box test, as well as having the lowest AIC among all of the models.

This study also shows that by forecasting less ahead to the future, the accuracy will be better. This has been proved by Table 4 and Table 5 comparison. Whereby forecasting 12 months ahead resulted to 124817.291 MSE, 353.295 RMSE, 1428221815 MAE, and 4.265 MAPE. Meanwhile, forecasting 6 months ahead resulted to 34385.384 MSE, 185.433 RMSE, 1415348102 MAE, and 2.206 MAPE.

REFERENCES

[1] Investments, Indonesia. "Jakarta Composite Index - IHSG - IDX Composite - Indonesia Stock Exchange | Indonesia Investments." Jakarta Composite Index - IHSG - IDX Composite - Indonesia Stock Exchange | Indonesia Investments, July 7, 2023. https://www.indonesia-investments.com/id/finance/stocks-bonds/jakarta-composite-index-ihsg/item887.

[2] Susanti, Riana, and Askardiya Radmoyo Adji. "ANALISIS PERAMALAN IHSG DENGAN TIME SERIES MODELING ARIMA." Jurnal Manajemen Kewirausahaan 17, no. 1 (June 30, 2020): 97. https://doi.org/10.33370/jmk.v17i1.393.

[3] Gunawan, Didik, and Weni Astika. "The Autoregressive Integrated Moving Average (ARIMA) Model for Predicting Jakarta Composite Index." Jurnal Informatika Ekonomi Bisnis, February 21, 2022. https://doi.org/10.37034/infeb.v4i1.114.

[4] Deviana, Santi, Nusyirwan Nusyirwan, Dorrah Azis, and Pandri Ferdias. "Analisis Model Autoregressive Integrated Moving Average Data Deret Waktu Dengan Metode Momen Sebagai Estimasi Parameter." Jurnal Siger Matematika 2, no. 2 (2021): 57-67.

[5] P. Box, George E., Gwilym M. Jenkins, Gregory C. Reinsel, and Greta M. Ljung. Time Series Analysis: Forecasting and Control. Wiley Series in Probability and Statistics Ser., 2015.