Forecasting the Weekly Stock Price of PT. OCBC NISP Tbk. using Auto Regressive Integrated Moving Average

Elisabeth Gloria Manurung[1*], Edwin Setiawan Nugraha[2]

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

[*]*Corresponding author: elisabeth.manurung@student.president.ac.id*

Abstract— Stocks are widely used in financial markets and can be an option for companies seeking to raise funds. Additionally, investors often opt for stocks as an investment due to their potential for providing high returns. To aid investors in making informed decisions when buying and selling stocks and mitigating risks, professionals have developed different theories and analyses to forecast stock prices. Auto Regressive Integrated Moving Average (ARIMA) (p,d,q) Within this investigation, we will employ technical analysis to forecast the weekly stock prices of PT Bank OCBC NISP Tbk (NISP.JK) for 7 weeks from Jan 7, 2022 to February 18, 2022. In this study, historical weekly stock price data for PT. Bank OCBC NISP Tbk (NISP.JK) from 1 January 2021 to 31 December 2021 was collected from Yahoo Finance website to create a forecast. The research got 12 different ARIMA models, the analysis show that the ARIMA (2,2,1) is the best.

Keywords— stocks, companies, technical analysis, ARIMA, stock price prediction, AIC value, falling stock price

I. INTRODUCTION

The objective of this document is to present a tangible and user-friendly illustration of ARIMA modeling, to forecast the stock price. Stocks represent a party's ownership stake in a company of Limited Liability Company, providing a claim to the company's income and assets, as well as the right to participate in the General Meeting of Stockholders (GMS). In the secondary market or daily stock trading, stock prices fluctuate based on supply and demand. Numerous factors, including company performance and industry-trends, as well as external factors, including company performance and industry trends, as well as external factors such as interest rates, inflation, exchange rates, and socio-political conditions, impact the demand and supply of stocks [1]. In addition, investing in stocks can provide a means of mitigating the effects of inflation. Inflation refers to the persistent and general increase in prices of goods and services over a given period of time [2]. The statement implies that the value of money decreases each year due to inflation, leading to a constant decline. Therefore, it is advisable not to limit savings to a bank that only offers 2% annual interest profit, especially considering that the annual inflation rate can range from 3% to 5%. The consequence of doing so would be that the money saved in the bank will not increase in value, since the interest earned is lower than the inflation rate. Instead, it would be more beneficial to invest the money or purchase assets that offer a higher return than the inflation rate [3].

The statement acknowledges that stock investment is generally more profitable, but it is important for an investor to carefully choose a stock portfolio that has the potential to yield positive returns. PT Bank OCBC NISP Tbk is a publicly listed bank in Indonesia that has been operating for over 75 years. It is a subsidiary of OCBC Bank, one of the largest financial services groups in Asia, and has a strong presence in Indonesia with over 300 branches and offices nationwide. The financial institution provides a variety of products and services, encompassing personal and business banking, wealth management, and insurance. Referred to as 'the Bank,' Bank OCBC NISP, formerly known as Bank NISP, holds the distinction of being the fourth oldest bank in Indonesia, founded on April 4, 1941, in Bandung under the name NV Nederlandsch Indische Spaar En Deposito Bank. Since 2005, OCBC Overseas Investments Pte. Ltd has been the majority stockholder with a stake of 85.1% by the end of 2019. Recognized for its strong financial stability, OCBC Bank has always provided full support as a stockholder to Bank OCBC NISP for managing commercial banking services in Indonesia [4].

In 2022, PT. OCBC NISP recorded a net profit of IDR 3.3 trillion. The bank's profit increased by 32% from the previous year of IDR 2.5 trillion. The increase in PT. OCBC NISP's profit was driven, among other things, by a 14% YoY increase in net interest income and a 25% YoY decrease in provision for losses. The improvement in the performance of Bank OCBC NISP was also evident in the credit disbursement, which increased by 14% YoY. The increase was supported by the credit disbursement in the business banking segment, which grew by 13% YoY, and retail banking grew by 16% YoY. One of the supporting factors for the increase in retail banking was the growth of consumer loans by 24% in 2022. The improvement in credit demand has also driven an increase in Loan to Deposits Ratio (LDR) to 77.2% at the end of 2022 [5].

The accuracy of the ARIMA model in predicting stock prices is above 85% for all industries, showing that it provides reliable predictions [6]. The study in [7], ARIMA models are used to forecast the stock prices of PT Bank Mandiri Tbk based on daily dataset from the April to July 2021. The study found that the ARIMA model provided accurate forecasts of the bank's stock prices of 9 day stock price data start from July 1, 2021, to July 18, 2021. The result demonstrates that the best model is ARIMA (1,2,1) with an accuracy of 0.95% MAPE. The recent study in [8], the forecasting Stock Price of PT. BCA Tbk show accuracy by 14.03% MAPE from ARIMA(3,20,). The utilized data is considerably dated, therefore, acquiring recent data is essential to obtain the latest forecasting.

The performance of an ARIMA model is evaluated including mean absolute error (MAE), mean absolute percentage error (MAPE), mean squared error (MSE), and root mean squared error (RMSE), to determine model accuracy. In general, lower values of these error metrics indicate better model performance. However, the acceptable range of error values can vary depending on the specific context of the analysis and the level of accuracy required for the forecasts to be useful. This research aims to use the ARIMA model and historical stock price data from January to December 2021 to forecast the changes in the stock price between 7th January 2022 and 18th February 2022.

II. LITERATURE REVIEW

A. *Introduction to Time Series*

Time series analysis is a specific method for examining a set of data points collected over a period of time. In this approach, analysts systematically record data points at regular intervals throughout a specified timeframe, as opposed to sporadically or randomly capturing them. Stock traders frequently employ time series analysis to gain insights into trends within different stock prices. Time series charts prove particularly beneficial for stock analysts and traders, providing a clear visualization of the trend and trajectory of specific stock prices. This analytical technique serves as a valuable tool for comprehending the patterns and movements in the stock market [9]. A time series is essentially composed of the following four components including trend, seasonal variation, cyclical variation and random variation [9].

B. *Stationary and Non-Stationary Model*

In time series analysis, there are two important concepts to understand, stationary and non-stationary models. Stationary models represent a time series where statistical properties such as mean, variance, and covariance do not change over time. Stationary models include autoregressive (AR), moving average (MA), and autoregressive integrated moving average (ARIMA) models, and they are widely used in various fields such as finance, economics, engineering, and meteorology. These models allow for making forecasts based on past data and identifying trends and seasonal patterns.

Non-stationary models, on the other hand, represent a time series where statistical properties change over time, such as trends, seasonal patterns, and cycles. Non-stationary models require additional processing to account for these changes, and some examples include the ARCH and GARCH models. Non-stationary models are also widely used in various fields and are important in time series analysis for identifying trends and patterns in non-stationary time series data [10]. Stationarity is classified into two types: strictly stationary and weakly stationary. The time series $\{X_t$,t ∈ Z} is said to be strictly stationary if the joint distribution of $(Xt_1, Xt_2, \dots, Xt_k)$ is the same as that of $(1 + h, Xt_2 + h, \dots, Xt_k + h)$. This mean that, strict stationarity means that the joint distribution only depends on the "difference" h, not the time $(t1, t2, \dots, tk)$. The time series {Xt, t ∈ Z} is said to be weakly stationary if fulfilled the condition : i) E $[X_t^2]$< ∞, ∀t ∈ Z; ii) E $[X_t]$ = μ, ∀t ∈ Z; and iii) γX (s,t) = γX (s + h,t + h), ∀s,t, h ∈ Z. Stated differently, three characteristics of a weakly stationary time series {X_t} are required: finite variation, constant first moment, and the second moment γX (s,t) depending only on |t − s| and not on s or t [10].

C. *Auto-Correlation Function (ACF) & Partial Auto-Correlation Function (PACF)*

The Autocorrelation Function (ACF) measures the correlation between a time series and its own lagged values. It is a plot of the correlation coefficient between the time series at different lags. The ACF can be used to identify the presence of a seasonality pattern in the time series. The Partial Autocorrelation Function (PACF) is similar to the ACF but it measures the correlation between two values of a time series while controlling for the values of the intermediate lags. It helps in identifying the order of an autoregressive (AR) process, which is a model that predicts the future values of a time series based on its past values.

The general formula for the Autocorrelation Function (ACF) at lag k for a stationary time series {Xt} is:

$$\boldsymbol{ACF(k) = Corr(X_t, X_{t-k}) = \frac{Cov(X_t, X_{t-k})}{(Var(X_t) * \sqrt{Var(X_{t-k})})}} \quad \textbf{(1)}$$

Here, Corr (X_t, X_{t-k}) represents the correlation between the observation at time t and the observation at time t-k, where k is the lag. Cov (X_t, X_{t-k}) represents the covariance between the two observations, and Var (X_t) and Var (X_{t-k}) represent the variances of the two observations. The denominator is a normalization factor that scales the covariance to lie between -1 and 1, which is the range of possible correlation coefficients. [10].

The ACF (k)gives us information about the strength and direction of the correlation between observations at lag k. If ACF (k) is positive and close to 1, it indicates a strong positive correlation between the two observations. If ACF (k) is negative and close to -1, it indicates a strong negative correlation between the two observations. If ACF (k) is close to 0, it indicates no correlation between the two observations.

The general formula for the Partial Autocorrelation Function (PACF) at lag k for a stationary time series $\{X_t\}$ is:

$$\boldsymbol{PACF\ (k) = Corr(X_t, X_{t-k} \mid X_{t-1}, X_{t-2}, X_{t-3}, \ldots, X_{t-k+1}} \quad \textbf{(2)}$$

Here, Corr $(X_t, X_t - k | X_{t-1}, X_{t-2}, \ldots, X_{t-k} + 1)$ represents the correlation between the observation at time t and the observation at time $t - k$, given the intervening values $X_{t-1}, X_{t-2}, X_{t-3}, \ldots, X_{t-k+1}$. This is called the partial correlation because it measures the direct correlation between two observations while controlling for the indirect correlation through the intervening values.

The PACF(k) gives us information about the direct correlation between observations at lag k, after removing the influence of the intervening values. If PACF(k) is positive and close to 1, it indicates a strong positive correlation between the two observations, after controlling for the intervening values. If PACF(k) is negative and close to -1, it indicates a strong negative correlation between the two observations, after controlling for the intervening values. If PACF(k) is close to 0, it indicates no direct correlation between the two observations at lag k, after controlling for the intervening values.

In practice, these formulas may need to be modified for non-stationary time series or to account for trends, seasonality, and other time series features. Additionally, there are various algorithms and software packages available to compute the ACF and PACF efficiently.

D. *Forecasting*

Making forecasts about a time series' future values based on historical observations is known as forecasting. It involves using statistical methods to identify patterns and trends in historical data, and then using these patterns to project future values of the time series. The goal of forecasting is to provide accurate and reliable estimates of future values, which can be used for planning, decision-making, and other purposes. It is a critical decision-making tool in many fields, including economics, finance, and business [11].

E. Estimated Eror-Value

The term Estimated Error Value (EEV) is used in measurement and metrology to quantify the uncertainty associated with a measurement result. The EEV is a standard deviation or a confidence interval that represents an estimate of the difference between the measured and true value of a quantity.

III. METHODOLOGY

A. *Stationery Model*

In this section we will discuss two types of stationary models including Autoregressive (AR) model and Movind Average (MA). AR that uses the dependent relationship between an observation and some number of lagged observations. AR model is a time series model that uses past values to predict future values. The AR model assumes that the future values of a variable are a linear function of its past values. The order of AR model is determined by the number of past values used to make predictions. An autoregressive model of order p, abbreviated AR (p), is of the form

$$\mathbf{X_t = \varphi_1 X_{t-1} + \varphi_2 X_{t-2} + \cdots + \varphi_p X_{t-p} + w_t = \sum_{i=1}^{p} \varphi_i X_{t-i} + w_t} \quad \textbf{(3)}$$

where X_t is stationary, $w_t \sim w_n$ $(0, \sigma_w^2)$, and $\varphi_1, \varphi_2, \ldots, \varphi_p$ $(\varphi_p \neq 0)$ are model parameters. The parameter p indicates how far the data before a time in the past determines the data at the present time

MA (Moving Average) model is a time series model that uses past errors to predict future values. The MA model assumes that the errors or residuals from the past values can be used to predict the future values. The order of MA model is determined by the number of past errors used to make predictions. A moving average model of order q, or MA(q), is defined to be

$$X_t = w_t + \theta_1 w_{t-1} + \theta_2 w_{t-2} \cdots + \theta_q w_{t-q} + w_t = w_t + \sum_{j=1}^{q} \theta_j w_{t-j} \quad (4)$$

where $w_t \sim w_n$ $(0, \sigma_w^2)$, and $\theta_1, \theta_2 . . ., \theta_q$ $(\theta_q \neq 0)$ are parameters [10] . ACF and PACF can be used for determining ARIMA model hyperparameters p and q, for more detail see at Table 1.

TABLE 1
Determining p and q from ACF and PACF plot

	AR (p)	MA (q)	ARMA (p, q)
ACF	Tails off	Cuts of after lag q	Tails off
PACF	Cuts of after lag p	Tails off	Tails off

B. Box–Jenkins Method

ARIMA (Autoregressive Integrated Moving Average) model is a combination of both AR and MA models, where integration of past values is also taken into account. ARIMA models are used to analyze and forecast time-series data, by fitting the time-series data to a linear equation that takes into account the past values, past errors and their integration. The method of creating models and predicting is explained in detail by Box and Jenkins in 1976. In conclusion, four crucial actions are advised:

(i) Model-identification: selecting an appropriate model based on the characteristics of the time series data, such as stationarity, autocorrelation, and partial autocorrelation

(ii) Model parameter estimation: estimating the parameters of the selected model using maximum likelihood estimation or a similar method

(iii) Model diagnostic checking: checking the adequacy of the model by examining the residuals, autocorrelation function, and partial autocorrelation function. If the model is found to be inadequate, the identification and estimation stages may need to be repeated until a satisfactory model is obtained, and

(iv) Model forecasting.

Through the use of ACF and PACF plot, these four steps first identify tentative model parameters; next, coefficients are determined to identify the likely model; finally, the model is validated; finally, basic statistics and confidence intervals are used to assess the forecast's validity and monitor the model's performance [10]. These steps are illustrated in Figure 1.

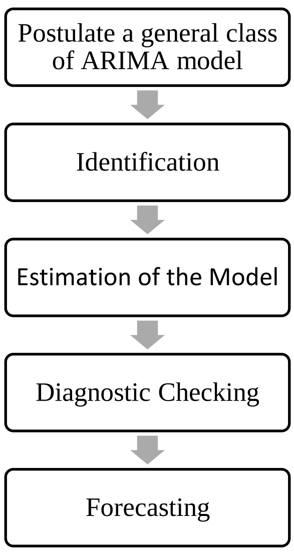


Figure 1. Four steps of Box-Jenkins Method

The Box-Jenkins ARMA model is a combination of the AR and MA models, which described below:

$$Yt = \delta + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \cdots - \theta_q e_{t-q}, \quad (6)$$

where the terms in the equation have the same meaning as given for the AR and MA model. The box-Jenkins ARIMA model is combination of the AR and MA models, but plus the differencing, which described below:

$$Wt = \delta + \phi_1 W_{t-1} + \phi_2 W_{t-2} + \cdots + \phi_p W_{t-p} + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \cdots - \theta_q e_{t-q}, \quad (7)$$

where:

$W_t = \nabla^2 Y_t = Y_t - 2Y_{t-1} + Y_{t-2}$

$W_{t-1} = \nabla^2 Y_{t-1} = Y_{t-1} - 2Y_{t-2} + Y_{t-3}$

$W_{t-2} = \nabla^2 Y_{t-2} = Y_{t-2} - 2Y_{t-3} + Y_{t-4}$

$W_{t-n} = \nabla^2 Y_{t-n} = Y_{t-n} - 2Y_{t-(n+1)} - Y_{t-(n+2)}$

The Box-Jenkins model supposes that the time series is stationary, and if it's not, Box and Jenkins suggest taking differences one or more times to achieve stationarity. By doing this, we can create an ARIMA model, where the "I" stands for "Integrated." These models are very flexible because they include both autoregressive and moving average terms.

C. Estimated Error Value

a) Mean Square Error (MSE)

It is a commonly used metric to evaluate the accuracy of a predictive model. MSE measures the average of the squared differences between the predicted and actual values of a time series. The formula is:

$$\mathbf{MSE} = \frac{\sum(\mathbf{y(i)} - \mathbf{\hat{y}(i)})^2}{\mathbf{n}} \quad \mathbf{(8)}$$

b) Root Mean Square Error (RMSE)

It is a commonly employed metric for assessing the precision of a predictive model, particularly within the context of time series analysis. RMSE (Root Mean Square Error) quantifies the disparity between predicted and actual values in a time series, often referred to as residuals. Essentially, RMSE serves as an indicator of how effectively the model aligns with the observed data. It is defined as:

$$\mathbf{RMSE} = \sqrt{\frac{\sum_{\mathbf{i=1}}^{\mathbf{N}} \|\mathbf{y(i)} - \mathbf{\hat{y}(i)}\|^2}{\boldsymbol{n}}} \quad \mathbf{(9)}$$

c) Mean Absolute Error (MAE)

MAE measures the average difference between the predicted and actual values of a time series. The equation of MAE can be written as:

$$\mathbf{MAE} = \frac{\mathbf{1}}{\mathbf{n}} \sum_{\mathbf{i=1}}^{\mathbf{n}} |\mathbf{y(i)} - \mathbf{\hat{y}(i)}| \quad \mathbf{(10)}$$

d) Mean Absolute Percentage Error (MAPE)

MAPE is a metric designed to assess the average percentage variance between predicted and actual values in a time series. This computation involves determining the average of the absolute percentage errors, where each error is the absolute value of the residual divided by the corresponding actual value. The formula is:

$$\mathbf{MAPE} = \frac{\mathbf{1}}{\mathbf{n}} \sum_{\mathbf{i=1}}^{\mathbf{n}} \frac{|\mathbf{y(i)} - \mathbf{\hat{y}(i)}|}{|\mathbf{y(i)}|} \quad \mathbf{(11)}$$

The explanation of Equation (8)-(11) are n is numbers of observation, y(i) is actual value for the i^{th} observation and $\hat{y}(i)$ = predicted value for the i^{th} observation [12].

e) Akaike Information Criterion (AIC)

AIC (Akaike Information Criterion) is a metric designed to assess how well a statistical model fits the data, considering a penalty for the number of parameters employed in the model. The calculation of AIC involves utilizing the log-likelihood of the model and incorporating the count of parameters utilized in the model. A lower AIC value indicates a better fit of the model to the data. The formula is:

$$\mathbf{AIC = -2log(maximum\ likelihood) + 2k} \quad \mathbf{(12)}$$

where k represent the number of model parameters. Log-likelihood is a measure of model fit. The higher the number, the better the fit [10].

IV. ANALYSIS AND DISCUSSION

A. THE HISTORICAL DATA

Overviewed the weekly stock price per share of PT. Bank OCBC NISP (NISP.JK) from January to December 2021, with the total 51 weeks or 51 data. The data is processed using R Studio. The table below presents the dataset.

TABLE 2
Weekly Stock Price of PT OCBC NISP Tbk (NISP.JK)

Date	Stock Price (Close)	Date	Stock Price (Close)
01/01/2021	845	09/07/2021	720
08/01/2021	890	16/07/2021	725
15/01/2021	880	23/07/2021	720
22/01/2021	810	30/07/2021	720
29/01/2021	835	06/08/2021	710
05/02/2021	890	13/08/2021	700
12/02/2021	850	20/08/2021	700
19/02/2021	870	27/08/2021	705
26/02/2021	900	03/09/2021	695
05/03/2021	890	10/09/2021	705
12/03/2021	915	17/09/2021	695
19/03/2021	880	24/09/2021	685
26/03/2021	860	01/10/2021	715
02/04/2021	870	08/10/2021	720
09/04/2021	860	15/10/2021	720
16/04/2021	845	22/10/2021	690
23/04/2021	815	29/10/2021	695
30/04/2021	805	05/11/2021	695
07/05/2021	800	12/11/2021	690
14/05/2021	800	19/11/2021	695
21/05/2021	805	26/11/2021	670
28/05/2021	810	03/12/2021	670
04/06/2021	815	10/12/2021	670
11/06/2021	800	17/12/2021	665
18/06/2021	790	24/12/2021	670
25/06/2021	775	31/12/2021	670
02/07/2021	765		

In the beginning of data processing, statistical descriptive is presented dan the result indicates that interval data between 665 and 915 with centered on 769.53 dan standard deviation is 79.03. Next step is start to processing of data to obtain the ARIMA Model by using R software. To do this, several essential packages are required, including TSA, forecast, tseries, readxl, and ggplot. The forecast package provides methods and utilities for visualizing univariate analysis of time series predictions, incorporating techniques like exponential smoothing via state-space models and automatic modeling using ARIMA. The TSA package is utilized for accessing functions such as ACF (Autocorrelation Function), PACF (Partial Autocorrelation Function), and ARIMA. The tseries package is employed to conduct the Augmented Dickey-Fuller Test (ADF Test). Using the readxl package facilitates the seamless importation of xlsx data into R Studio. Finally, ggplot is instrumental for plotting the data.

In the subsequent step, the data on the weekly stock price per share of NISP.JK is visualized, with the x-axis representing the date and the y-axis indicating the stock price.

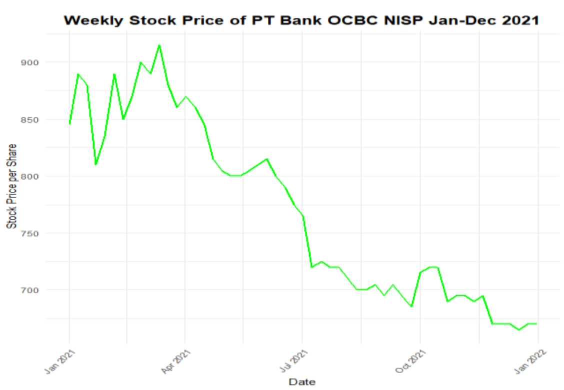


Figure 2. Plot of data NISP.JK weekly stock price

B. STATIONARITY CHECK

To verify whether the data is stationary, the "adf.test(data)" function is used in R Studio to perform the Augmented Dickey-Fuller Test (ADF Test).

The Augmented Dickey-Fuller Test (ADF Test) was performed on the original data, yielding a p-value of 0.2186, which is greater than 0.05. This result indicates that the stock price data for the NISP.JK is non-stationary. Thus, based on the first test, the data are non-stationary. We need to perform differencing to make this data stationary by calling the “diff(data)” function. For the first differencing, the result still not smaller than 0.05, that means the data not stationary yet. Then we do the second differencing. The obtained results are noteworthy because the resulting p-value of 0.01 is less than 0.05, indicating that the data is stationary and has a constant mean. Based on this, it can be inferred that the second differentiation of the weekly stock price NISP.JK has yielded data suitable for forecasting using the ARIMA model. Thus, the value of d used is 2.

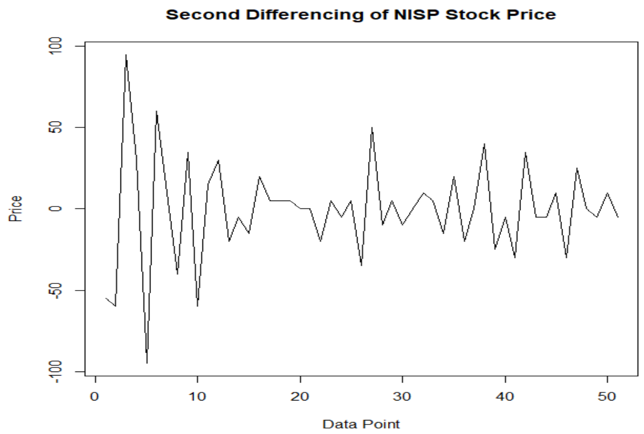


Figure 3: Plot of 2^{nd} Differencing Stock Price NISP.JK

C. ARIMA MODEL SPECIFICATION

In this section, the plot of the Auto Correlation Function (ACF) and Partial Auto Correlation Function (PACF) is presented. The appropriate ARIMA model for the NISP.JK weekly stock price data can be found by analyzing of the both plots.

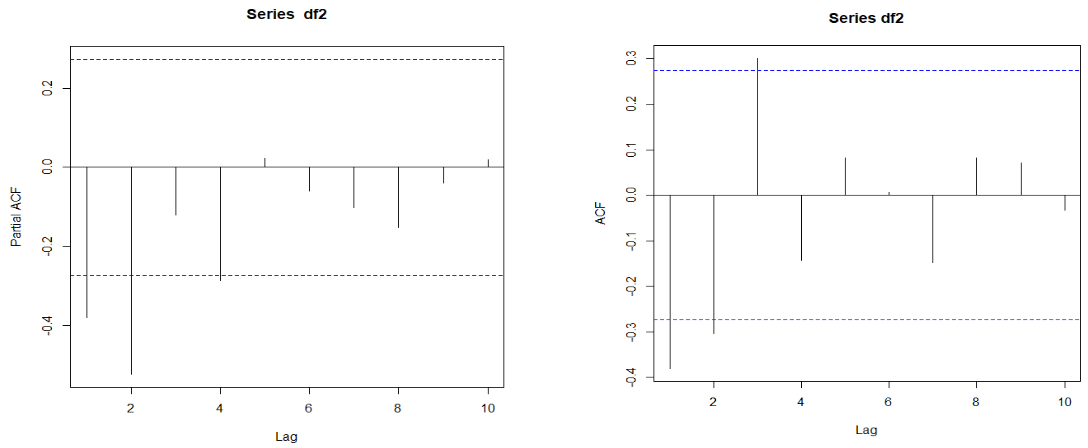


Figure 4: Plot PACF and ACF after 2^{nd} Level of Differentiation *(d = 2)*

Based on the PACF and ACF plots in Figure 4, it can be observed the initial ARIMA model for this data is ARIMA (2,2,3). After conducting further analysis on the data, several ARIMA models were proposed available at Table 3.

TABLE 3
ARIMA CONSTRUCTED

Model ARIMA	P	d	q
(2,2,3)	2	2	3
(2,2,2)	2	2	2
(2,2,1)	2	2	1
(2,2,0)	2	2	0
(1,2,3)	1	2	3
(1,2,2)	1	2	2
(1,2,1)	1	2	1
(1,2,0)	1	2	0
(0,2,3)	0	2	3
(0,2,2)	0	2	2
(0,2,1)	0	2	1
(0,2,0)	0	2	0

D. PARAMETER ESTIMATION

The parameter estimates for the autoregressive (AR) function, denoted as φ_p, and the moving average (MA) function, denoted as θ_q, can be obtained by reviewing the summary of each model. By running the code, it can obtain the values for AR1 to AR2, MA1 to MA3, MSE, Log Likelihood, and AIC. These values will be taken into account when selecting the most appropriate ARIMA model. The output of this function can be viewed in the console, as shown in Tabel 4.

TABLE 4
PARAMETER ESTIMATION FOR CANDIDATE ARIMA MODEL

No.	Model ARIMA	AR 1	AR 2	MA 1	MA 2	MA 3	Log Likelihood	AIC	MSE
1	(2,2,3)	-0.6821	-0.5013	-0.4357	-0.5029	-0.0615	-224.45	458.9	352.5
2	(2,2,2)	-0.655	-0.4489	-0.4621	-0.5379		-224.47	456.93	352.5
3	(2,2,1)	-0.2027	-0.3826	-1			-225.82	457.63	370.9
4	(2,2,0)	-0.7376	-0.7038				-230.2	464.4	472.8
5	(1,2,3)	-0.4457		-0.6188	-0.7631	0.3819	-225.37	458.74	365
6	(1,2,2)	0.267		-1.5076	0.5076		-228.58	463.17	416.6
7	(1,2,1)	-0.138		-1			-229.5	463.01	436.8
8	(1,2,0)	-0.3977					-243.67	489.34	824.2
9	(0,2,3)			-0.9878	-0.4014	0.3892	-226.62	459.24	382.9
10	(0,2,2)			-1.2768	0.2768		-228.89	461.78	422.8
11	(0,2,1)			-1			-229.94	461.89	446.8
12	(0,2,0)						-247.87	495.74	975

E. RESIDUAL ANALYSIS

Next step is find the best ARIMA model for prediction purposes, it is necessary to perform a normality test using the Shapiro test and Ljung-Box test. The model with a p-value greater than 0.05 from both tests will be selected as the best model. Table 5 is present the summary values for each model. According to Table 5, the models that meet the criteria of the Shapiro and Ljung-Box tests are Arima model with no. 1, 2, 3,5,6 and 9. To determine the best model, we must compare the four error value estimation parameters. After comparison, it can be seen that models 2 and 3 have the smallest AIC values, making them potential candidates for further comparison.

TABLE 5
RESULT OF RESIDUAL ANALYSIS

No.	Model ARIMA	Uji Saphiro	Uji Ljung-Box	AIC	Accepted
1	(2,2,3)	0.06807	0.5463	458.9	Yes
2	(2,2,2)	0.1091	0.5473	456.93	Yes
3	(2,2,1)	0.3749	0.3126	457.63	Yes
4	(2,2,0)	0.001003	0.7844	464.40	No
5	(1,2,3)	0.4347	0.8257	458.74	Yes
6	(1,2,2)	0.05014	0.4223	463.17	Yes
7	(1,2,1)	0.006753	0.8259	463.01	No
8	(1,2,0)	0.008439	0.1263	489.34	No
9	(0,2,3)	0.1058	0.6566	459.24	Yes
10	(0,2,2)	0.01412	0.3263	461.78	No
11	(0,2,1)	0.0118	0.6525	461.89	No
12	(0,2,0)	0.02812	0.004338	495.74	No

TABLE 6
ERROR VALUE ESTIMATION PARAMETERS COMPARISON BETWEEN MODEL 2 AND MODEL 3

	ARIMA (2,2,2)	ARIMA (2,2,1)
MSE	395.59871	387.13938
RMSE	19.88966	19.67586
MAE	18.08417	1796.473%
MAPE	2.795%	3.099%

According to Tabel 6, ARIMA (2,2,1) is considered the best among these models as it has the lowest values for MSE, RMSE, and MAE, and the second-lowest AIC value compared to the other models.

Below are the visualization plot and norm distribution plot for the ARIMA Model (2,2,1).

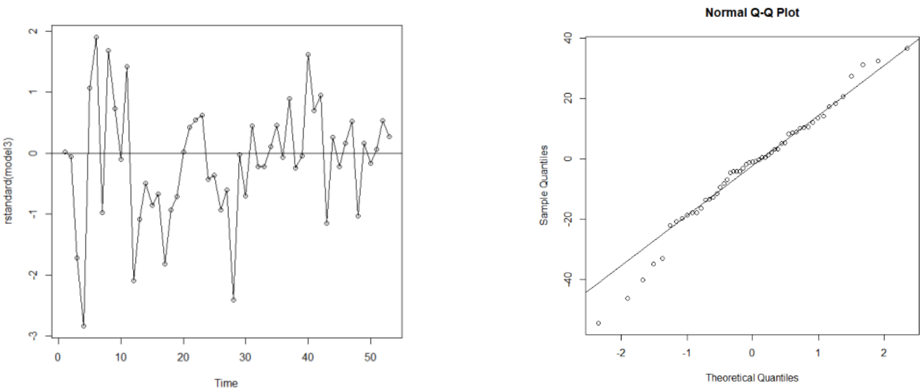
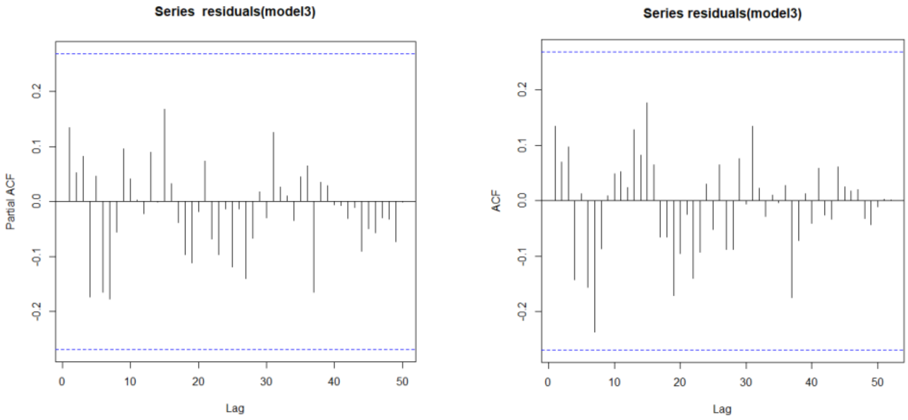


Figure 5: Residual Plot and Normal QQ-Plot of ARIMA (2,2,1)



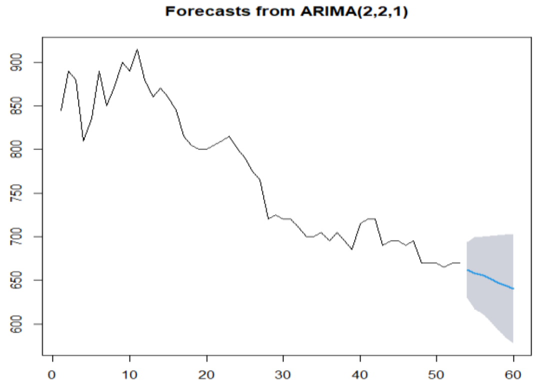


Figure 6: Plot of ACF and PACF of ARIMA (2,2,1)

Figure 7: Plot of Exchange Rate of CNY/USD Prediction Result

TABLE 7
ACTUAL DATA AND PREDICTED DATA FOR SEVEN WEEKS

DATE	ACTUAL DATA	PREDICTED DATA	LOWER BOUND	UPPER BOUND	MSE	RMSE	MAE	MAPE
2022-01-07	645	662.1425	630.1546	694.1304	293.86531	293.86531	17.14250	2.658%
2022-01-14	640	657.7910	616.4977	699.0843	316.51968	316.51968	17.79100	2.780%
2022-01-21	625	655.7349	611.5753	699.8944	944.63408	944.63408	30.73490	4.918%
2022-01-28	645	651.8721	603.1355	700.6086	47.22576	47.22576	6.87210	1.065%
2022-02-04	675	647.4973	593.0823	701.9124	756.39851	756.39851	27.50270	4.074%
2022-02-11	660	643.9176	585.1704	702.6648	258.64359	258.64359	16.08240	2.437%
2022-02-18	650	640.3725	577.7296	703.0155	92.68876	92.68876	9.62750	1.481%
		RESULT			387.13938	19.67586	17.96473	3.099%

Table 7 present the actual data and predicted data resulted from ARIMA (2,2,1). That table also calculate of error of this prediction using some measurement such as MSE, RSME, MAE and MAPE. If we compare it with previous work [7], they had a MAPE of 0.95%, which is the same; however, other error values such as MSE, RMSE, and MAE were not explained. While if it is compared with other previous study [8], the result of accuracy is better.

V. CONCLUSION

ARIMA modeling is an effective method for forecasting the stock price of NISP.JK. The best and most appropriate ARIMA model for predicting the exchange rate of CNY to USD is ARIMA (2,2,1), with an error rate of 3.099% based on the MAPE percentage. This indicates that the model's accuracy in predicting based on actual data is 96.901%. The forecast suggests that the stock price will fall from January 7, 2022, to February 18, 2022. The most suitable formula for this model is ARIMA (2,2,1), which is described below:

$$\boldsymbol{Y_t = 1.7973Y_{t-1} - 0.9772Y_{t-2} + 0.5625Y_{t-3} - 0.3826Y_{t-4} + e_t + e_{t-1}}$$

The forecasting of this stocks will give insight for the future condition. The result is can be considered for investor when they design of strategy of investment in stock. However, it should be emphasized that the limitations of this research are that residuals must follow a normal distribution, requirement of data stationarity, free from correlation among residuals, assume constant residuals. The important thing here is that the accuracy of forecasting will go worst when the time of forecasting is longer. Beside that, there are also many other factors that influence changes in stock prices, such as inflation and other economic factors. For further research, it is recommended to perform forecasting using other methods such as the ARIMA GARCH model, AFRIMA, and deep learning.

REFERENCES

[1] OJK, "Saham," [Online]. Available: https://sikapiuangmu.ojk.go.id/FrontEnd/CMS/Category/64.

[2] T. Clark, "The Balance," 12 March 2021. [Online]. Available: https://www.thebalance.com/how-inflation-affects-the-stock-market-4170135.

[3] Investopedia, "Inflation," October 2020. [Online]. Available: http://www.investopedia.com/terms/i/inflation.asp.

[4] OCBC, "Profile OCBC NISP," [Online]. Available: https://www.ocbcnisp.com/id/tentang-ocbc-nisp/profile.

[5] Kontan, "Ini Jadwal Pembayaran Dividen OCBC NISP Dengan Yield 6,86%," 12 April 2023. [Online]. Available: https://investasi.kontan.co.id/news/ini-jadwal-pembayaran-dividen-ocbc-nisp-dengan-yield-686.

[6] P. Mondhal, L. Shit and S. Gowami, "Study of effectiiveness of time series modelling (ARIMA) in forecasting stock price," *International Journal of Computer Science, Engineering and Applications,* vol. 4, no. 2, 2014.

[7] E. A. Widodo, E. S. Nugraha and D. A. Hamzah, "Forecasting PT Bank Mandiri Tbk Stock Price Using ARIMA," in *Symposium on Data Science,* Jababeka, CIkarang, 2022.

[8] A. Olivia and E. S. Nugraha, "Forecasting PT Bank Central Asia Tbk Stock Price Using ARIMA," *Journal of Actuarial, Finance and Risk Managment,* vol. 2, no. 1, pp. 1-27, 2023.

[9] J. Jose, "Introduction to time series analysis," August 2022. [Online]. Available: https://www.researchgate.net/publication/362389180_INTRODUCTION_TO_TIME_SERIES_ANALYSIS_AND_ITS_APPLICATIONS.

[10] J. D. Cryer and K.-S. Chan, Time Series Analysis With Application in R, New York: Springer, 2008.

[11] M. Suleiman, I. Muhammad, A. Z. Adamu, Y. Zakari, R. Iliyasu, A. Muhammad, I. Adamu and M. Abdu, "Modelling Nigeria Crude Oil Prices using ARIMA Time Series

Models," *Journal of Science and Technology Research*, vol. 5, no. 1, pp. 230-241, 2023.

[12] C. D. Montgomery, L. J. Cheryl and M. Kulahci, Introduction to time series analysis and forecasting, New Jersey: Jhon Wiley & Sons, 2008.