Comparison of the Symmetric and Asymmetric Generalized Autoregressive Conditional Heteroscedasticity (GARCH) Models in Forecasting the 2018-2023 Jakarta Composite Index

Yenni Angraini[1]*, Adelia Putri Pangestika[2], and I Made Sumertajaya[3]

[1-3]Department of Statistics, Faculty of Mathematics and Natural Sciences,
IPB University Bogor, Indonesia 16680
[1]y_angraini@apps.ipb.ac.id; [2]adelia02putri@apps.ipb.ac.id; [3]imsjaya@apps.ipb.ac.id

Received: 2nd October 2023/ **Revised:** 10th December 2023/ **Accepted:** 11th December 2023

How to Cite: Angraini, Y., Pangestika, A. P., & Sumertajaya, I. M. (2024). Comparison of the Symmetric and Asymmetric Generalized Autoregressive Conditional Heteroscedasticity (GARCH) Models in Forecasting the 2018-2023 Jakarta Composite Index. *ComTech: Computer, Mathematics and Engineering Applications, 15*(1), 1−15. https://doi.org/10.21512/comtech.v15i1.10610

Abstract **-** The Autoregressive Integrated Moving Average with Exogenous Variables (ARIMAX) method assumes a homogeneous residual variance, but data with high volatility can cause violations of this assumption. Hence, it is interesting to compare the forecasting accuracy of symmetric and asymmetric Autoregressive Conditional Heteroskedasticity (ARCH) models in various data conditions. The research aimed to compare the accuracy of the symmetric ARCH/ Generalized Autoregressive Conditional Heteroscedasticity (GARCH) and asymmetric TGARCH models in forecasting weekly Jakarta Composite Index (JCI) data on January 1st, 2018, to April 24th, 2023, by involving the influence of COVID-19 as a covariate variable and applying several validation scenario models to training and testing data. Based on the best-selected model, forecasting was carried out from May 1st, 2023, to July 3rd, 2023. The data used were weekly JCI opening data from January 1st, 2018, to April 24th, 2023, with the COVID-19 period as a covariate variable. The analysis results show that symmetric and asymmetric methods can handle violations of the heteroscedasticity assumption in the ARIMAX model. The best model produced based on four data validation scenarios is the asymmetric ARIMAX(3,1,3)-TGARCH(1,2) model with an average MAPE value of 3.158%. In this model, the COVID-19 variable significantly influences the JCI movement. Forecasting is done with forecasting results that are stable with confidence intervals that widen in each period.

Keywords: Generalized Autoregressive Conditional Heteroscedasticity (GARCH), forecasting accuracy, Jakarta Composite Index (JCI)

I. INTRODUCTION

Forecasting is used to estimate data values in a period based on data values in previous periods. The forecasting method that is often used is the Autoregressive Integrated Moving Average (ARIMA) method. The development of the ARIMA method using other time series data as exogenous variables is called the Autoregressive Integrated Moving Average with Exogenous Variables (ARIMAX) method. The ARIMAX method assumes homoscedasticity. However, the unpredictable nature of volatility can cause heteroscedasticity (Somarajan et al., 2019). It needs a method that can be used to overcome violations of assumptions in the ARIMAX method.

The Autoregressive Conditional Heteroskedasticity (ARCH) method models the conditional variance as a function of the previous period's white noise. However, this method has a limitation in that it requires a large order to obtain a suitable model. The development of the ARCH method is the Generalized Autoregressive Conditional Heteroscedasticity (GARCH) method, which overcomes the limitation of the ARCH method by modeling the conditional variance as a function of the previous period's white noise and conditional variance

(Ekinci, 2021). However, according to Dinku and Worku (2022), the GARCH method needs to be better for modeling asymmetric volatility tendencies where positive shocks (good news) have a different effect from negative shocks (bad news). Another method that can be used is the asymmetric GARCH method, such as Threshold GARCH (TGARCH).

One data set with a tendency toward high volatility is found in financial data, such as the Jakarta Composite Index (JCI). The JCI is a crucial index that measures the performance of all shares listed on the Indonesia Stock Exchange (IDX). Investors consider JCI to determine investment in buying and selling transactions in the capital market. The JCI movement illustrates the movement of all stocks on the IDX, so an increase in the JCI will show an increase in the average stock on the IDX. Investors will use the increase in JCI to gain profits by investing (Hismendi et al., 2021). Investors must be able to analyze current economic conditions and their impact on the stocks in which they invest to maximize the investment. Therefore, an appropriate forecasting method is needed to help investors to plan their investments in the short, medium, and long term.

Several previous researchers have predicted JCI data. For example, Kasuma and Nugroho (2020) predicted daily JCI data from March 2nd, 2020, to August 5th, 2020, using the symmetric ARCH/GARCH method. The best model obtained to predict JCI data for this period was ARMA(1,1)-ARCH(1) with a Mean Absolute Percentage Error (MAPE) value of 19%. On the other hand, Saida et al. (2016) predicted the daily return of JCI data from January 2nd, 2013, to October 30th, 2015, using the asymmetric GARCH method, TGARCH. The best model obtained to predict JCI data for this period was ARMA(3.26)-TGARCH(1.1), with a MAPE value of 7.17%.

Based on the previous studies, both methods have quite good forecasting accuracy with a MAPE value of less than 20%. Hence, it is interesting to compare the forecasting accuracy of the two methods. However, both methods use the Autoregressive Moving Average (ARMA) mean model, which only considers data values from the previous period for forecasting. In reality, data values for future periods can be influenced by previous period data values and other variables outside the model. In the context of forecasting JCI data, the movement of JCI data is influenced by the JCI value for the previous period and other variables, such as COVID-19. According to Haryanto (2020), COVID-19 significantly influences the movement of the JCI, so forecasting JCI data involving the COVID-19 period should consider its influence in forecasting.

In addition, these two previous studies and most other studies using conventional methods on time series data only validate the model in one data condition. A suitable model for one data condition is not necessarily suitable for others. So, when comparing models, it is necessary to consider several data scenarios to ensure that the best model chosen is suitable for various data conditions. Based on these problems, the research aims to compare the forecasting accuracy of the symmetric ARCH/GARCH and asymmetric TGARCH models in forecasting weekly JCI data from January 1st, 2018, to April 24th, 2023, by involving the influence of COVID-19 as a covariate variable into the mean model and applying several model validation scenarios on training data and test data. Based on the best model obtained from the model validation results, JCI data forecasting is carried out in a specific period (May 1st, 2023, to July 3rd, 2023).

II. METHODS

The data used in the research are opening price data of JCI from http://finance.yahoo.com. The data are weekly from January 1st, 2018, to April 24th, 2023. The research also uses covariate variables in the form of dummies to accommodate the influence of COVID-19 on the JCI movement. The dummy variable used in this case is a value of one for the COVID-19 period and zero for others. The period of COVID-19 in the research is assumed to be between January 1st, 2020, when the first case of COVID-19 in the world was announced, until October 10th, 2021, after COVID-19 cases began to subside. The data are then analyzed using R software, starting from data imputation.

Data imputation is carried out to fill in missing data in a certain period. The imputation method used is linear interpolation. According to Noor et al. (2015), the linear interpolation formulation is shown in Equation (1). It has Y_t as the missing value in the t period, Y_A as the data before the missing value, Y_B as the data after the missing value, a as the period before the missing value, b as the period after the missing value, and t as the period of the missing value.

$$Y_t = \frac{Y_B - Y_A}{b-a}(t-a) + Y_A. \quad (1)$$

After missing data have been successfully imputed, data exploration is carried out using a time series plot to see patterns and characteristics of the data. Based on the pattern and characteristics, the data are split by selecting a split point where the data pattern before the split point is in line with the data pattern after the split point. In this case, data for July 18th, 2022, is used as the split point so that the training data used in the research range from January 1st, 2018, to July 18th, 2022. The training data are then used in ARIMAX modeling. The ARIMAX model developed by Box and Tiao in 1975 is an extension of the ARIMA model, which adds exogenous variables as covariate variables (Braz et al., 2023). The exogenous variables added are usually dummy variables in the form of calendar variations. The ARIMAX model formulation is presented in Equation (2). It consists of ϕ as the autoregressive parameter, Y_{t-i} as the $t-i$ period data, p as the autoregressive order, θ as the moving average parameter, e_{t-j} as the white noise of $t-j$ period, q as the moving average order, β_m as the exogenous variable parameter, X_{t-m} as exogenous variables, and n as the number of exogenous variables.

$$Y_t = \sum_{i=1}^{p} \phi Y_{t-1} + \sum_{j=1}^{q} \theta_j e_{t-j} + \sum_{m=1}^{n} \beta_m X_{t-m}. \quad (2)$$

The ARIMAX modeling stage begins by conducting a data stationarity test using a time series plot, Autocorrelation Function (ACF) plot, and Augmented Dickey-Fuller (ADF) test. When the data are not stationary, it will be handled by making differences. On the other hand, when the data are stationary, the ARIMA model identification stage will be carried out based on the ACF plot, Partial Autocorrelation Function (PACF) plot, and Extended Autocorrelation Function (EACF) plot. Next, the ARIMA model parameter estimation stage is carried out using the maximum likelihood method. The best ARIMA model candidate selected has all significant parameters and the smallest Akaike's Information Criterion (AIC) value.

The next stage is to enter the dummy variable for the COVID-19 period as an exogenous variable to form the ARIMAX model. The model formed then enters the model diagnostic testing stage. When all assumptions are met, modeling continues by overfitting. In this case, overfitting is adding orders to the model used to see whether there is a better model (Moffat & Akpan, 2019). After overfitting, the best ARIMAX model is selected based on all significant parameters, the smallest AIC value, and the results of model diagnostic tests that meet the assumptions.

Next, the effect of heteroscedasticity is tested on the residuals of the best ARIMAX model using the ARCH-Lagrange Multiplier (LM) test and the McLeod-Li test. The null hypothesis of the ARCH-LM test is that the residuals do not contain the ARCH effect. The test statistic used according to Fang et al. (2020) is presented in Equation (3). It has T as the residual length and R^2 as the coefficient determination of the regression between the squared residual and the sum of the squared residual up to the tested lag.

$$LM = TR^2. \quad (3)$$

This null hypothesis will be rejected at the significance level of α if it is $LM > X^2_{\alpha,K}$ where K is the maximum lag. The null hypothesis of the McLeod-Li test is that the residuals do not have a heteroscedasticity effect. According to Lekhal and El Oubani (2020), the test statistics used are listed in Equation (4). It shows r_i^2 as the autocorrelation of the squared residual. Moreover, the null hypothesis will be rejected at the significance level of α if it is $Q_\alpha > X^2_{\alpha,K}$.

$$Q = T(T+2) \sum_{i=1}^{K} \frac{r_i^2}{(T-i)}. \quad (4)$$

When the selected model violates the homoscedasticity assumption, ARCH/GARCH and TGARCH models are carried out. The ARCH models the conditional variance as a function of the squared white noise of the previous period (Raheem et al., 2020). According to Kyriazis et al. (2019), the ARCH model formulation is shown in Equation (5).

$$\sigma_t^2 = \omega + \sum_{i=1}^{p} \alpha_i e_{t-i}^2. \quad (5)$$

Equation (5) has σ_t^2 as the conditional variance of the t period, ω as a constant, α as the ARCH parameter, e_{t-i} as the $t-i$ period of white noise, and p as the ARCH order. Meanwhile, the GARCH models conditional variance as a function of the previous period's squared white noise and conditional variance. The GARCH model is more useful if the lag is large. The GARCH model formulation according to Fang et al. (2019) is shown in Equation (6). It consists of β as the GARCH parameter, and p and q are the GARCH orders.

$$\sigma_t^2 = \omega + \sum_{i=1}^{p} \alpha_i e_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2. \quad (6)$$

The TGARCH model is an extension of the GARCH model, which uses dummy variables to model the possibility of asymmetric effects on data (Shahani & Taneja, 2022). The TGARCH model formulation, according to Sheng et al. (2021), is listed in Equation (7). It includes γ as an asymmetric parameter, p and q as TGARCH orders, and $I_{[e_{t-j}<0]}$ as a dummy variable that has a zero value when $e_{t-j} \geq 0$ and a value of one when $e_{t-j} \leq 0$.

$$\sigma_t^2 = \omega + \sum_{i=1}^{p} \alpha_i e_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2 + \sum_{k=1}^{p} \gamma_i e_{t-k}^2 I_{[e_{t-k}<0]}. \quad (7)$$

The ARCH/GARCH and TGARCH modeling stages begin with identifying the model, followed by estimating model parameters using the maximum likelihood method. The best candidate model is then selected based on the criteria of all significant parameters, non-significant ARCH-LM test results, and the smallest AIC value. When the best candidate model has been selected, a model diagnostic test is carried out. If the model diagnostic test results meet the assumptions, the modeling stage continues with overfitting. After overfitting is carried out, the best model is selected by considering all significant parameters, insignificant ARCH-LM test results, the smallest AIC value, and model diagnostic test results that meet the assumptions.

The best model between the symmetric ARCH/GARCH and asymmetric TGARCH models is determined through the model validation stage using four training and testing data scenarios. These four scenarios are formed by dividing the data other than training data determined in the previous step into four equal parts. The training data for the first scenario is the same as the training data determined in the previous division. In contrast, the testing data for the first scenario is a first quarter of the data other than training data determined in the previous step. Furthermore, the second scenario training data is used in the previous scenario. In contrast, the second scenario testing data is the second quarter of the previously determined data other than training data and so on. A division of training and testing data is shown in Figure 1 (see Appendices) and Table 1 (see Appendices) in more detail.

In each scenario, data other than training and testing data are not used in the model validation process. The model validation process for each data scenario begins with estimating parameters on the training data using the best model obtained in the previous stage, followed by forecasting and calculating forecasting accuracy on test data using MAPE values. According to Montgomery et al. (2015), the formulation of the MAPE value is shown in Equation (8). In Equation (8), Y_t is the actual value, $\hat{Y}_t$ is the estimated value, and T is the amount of data.

$$MAPE = \frac{1}{T}\sum_{t=1}^{T}\left(\frac{|Y_t-\hat{Y}_t|}{Y_t} \times 100\right). \quad (8)$$

The best model chosen between the symmetric ARCH/GARCH model and the asymmetric TGARCH model is the model with the smallest average MAPE value from the four data scenarios. Next, forecasting is carried out along one test data scenario, namely 10 periods from May 1st, 2023, to July 3rd, 2023. It uses the best model obtained at the model validation stage.

III. RESULTS AND DISCUSSIONS

Data imputation is carried out to fill in the missing values for June 11th, 2018, June 3rd, 2019, and May 2nd, 2023. The data imputation method used is linear interpolation. The results of data imputation are shown in Figure 2 (see Appendices), marked by a red symbol. Figure 2 (see Appendices) shows that the JCI tended to fluctuate throughout 2018. This fluctuation can be due to domestic conditions, such as Indonesia's economic growth, the weakening of the Rupiah exchange rate and the trade balance deficit, or foreign causes, such as the occurrence of a trade war between the United States and China and a decision to increase the Fed Funds Rate (FFR) by the American central bank (Ichsani et al., 2019). Moreover, the JCI had a non-steep downward trend during 2019 and a steep downward trend in early 2020 because of the COVID-19 pandemic. Nonetheless, the JCI began to increase in the second quarter of 2020 in line with Bank Indonesia (BI) lowered the 7-Day Reverse Repo Interest Rate by 125 basis points (bps). This increase continued until late 2021 and tended to be stable in 2022 to early 2023 (Behera et al., 2023).

The JCI movement, which tends to fluctuate in each period, means that the modeling must consider the influence of data volatility. In addition, the steep downward trend at the beginning of 2020 indicates the need to consider the influence of other variables, namely COVID-19, in modeling. In this case, the influence of COVID-19 on the JCI movement is modeled using the ARIMAX mean model.

ARIMAX modeling on JCI data begins with checking the stationarity of the data. The time series plot in Figure 3 (see Appendices) shows that the data are not stationary, marked by a trend in data. This non-stationarity is supported by the ACF plot in Figure 4 (see Appendices), which decreases exponentially, and the ADF test with a p-value of 0.564, which is greater than the 5% significance level.

Nonstationary data must be handled to obtain stationarity as a prerequisite for determining autoregressive and moving average components using ACF and PACF (Hussain et al., 2023). The nonstationary data are handled by differencing (d = 1). After first differencing, the JCI data is stationary in the mean, as marked by the time series plot in Figure 5 (see Appendices), which tends to move at a constant mean, and the ACF plot in Figure 6 (see Appendices), which is cut off after the seventh lag. This stationarity is reinforced by the results of the ADF test, which has a p-value of 0.01, which is less than the 5% significance level.

Identification of the ARIMA model on already stationary data is done by looking at the ACF, PACF, and EACF plots. Based on the ACF plot in Figure 6 (see Appendices) and the PACF plot in Figure 7 (see Appendices), there are no cuts or tails off in the initial lag, so a tentative model is challenging to determine. The identification of the ARIMA model is then determined based on the EACF plot by looking at the zero triangle pattern, where the sharp triangular ends of the zero triangle pattern correspond to the tentative ARIMA model order (Kong et al., 2023). So, the tentative models formed based on Figure 8 are ARIMA(1,1,1), ARIMA(2,1,2), and ARIMA(3,1,3).

The tentative ARIMA model parameter estimates in Table 2 (see Appendices) show that the ARIMA(3,1,3) model has all significant parameters and the smallest AIC value. This model is then selected as the best ARIMA model candidate, which will be continued with ARIMAX modeling. ARIMAX modeling is done by adding a dummy variable for the COVID-19 period as a covariate variable in the ARIMA(3,1,3) model.

The estimation of the ARIMAX(3,1,3) model in Table 2 (see Appendices) shows that the ARIMAX(3,1,3) model has all significant parameters except for the dummy variables. This case is similar to Putera (2020). The previous research continues to include dummy variables that are not significant in the model to increase forecasting accuracy. It is also in line with Vukovic and Zinurova (2020) that even though it is not significant, adding covariate variables to the model will add to its goodness. The addition of dummy variables to form the ARIMAX(3,1,3) model is proven to produce a smaller AIC value than the ARIMA(3,1,3) model, so in this case, dummy variables that are not significant are still included in the model.

Next, model diagnostic tests are carried out on the residuals of ARIMAX(3,1,3) models. Exploratively, the ACF plot of residuals in Figure 9 (see Appendices) shows no significant autocorrelation in the first 20 lags, indicating the fulfillment of the assumption that residuals resemble white noise. These results are supported by formal testing using the Ljung-Box test with a p-value of 0.970, greater than the 5% significance level. The exploratory residual normality is carried out using a quantile plot. Based on the quantile plot in Figure 10, the points on the plot tend not to follow the normal line, which indicates a violation of the normality assumption. This result is supported by the Shapiro-Wilk test with a p-value of 0.000, which is less than the 5% significance level. Nonetheless, violations of this assumption can be neglected because the data used tend to be large (Schaffer et al., 2021).

Furthermore, the heterogeneity test of the variance of the residuals is carried out using the time series plot of ARIMAX(3,1,3) residual. The residual plot shown in Figure 11 shows that the residuals of the ARIMAX(3,1,3) model tend to be heterogeneous because they have different bandwidths over several periods. These results are supported by the Ljung-Box test of squared residuals with a p-value of 0.000, greater than the 5% significance level.

The heteroscedasticity effect is then carried out on the residuals of the ARIMAX(3,1,3) model. It indicates that it violates the homoscedasticity assumption. This test is carried out using the ARCH-LM and the McLeod-Li tests. Based on the ARCH-LM results in Table 3 (see Appendices), the p-value of the ARCH-LM test is significant up to the 14th lag. This result indicates a heteroscedasticity effect on the residuals of the ARIMAX(3,1,3) model. Moreover, the results of the McLeod-Li test in Figure 12 (see Appendices) are also in line with this conclusion because it has a p-value that is less than the 5% significance level up to the 15th lag. The heteroscedasticity on the residuals of the ARIMAX(3,1,3) model indicates the need for the ARCH/GARCH process to model conditional heteroscedasticity (Aliyev et al., 2020).

Then, identification of the ARCH/GARCH model on ARIMAX(3,1,3) is carried out by trial and error to get the best model. Based on the trial and error results, the ARIMAX(3,1,3)-ARCH(2) model is the best candidate model with all significant parameters and the smallest AIC value. The ARCH-LM test on this model produces a p-value of 0.325. It indicates that the violation of the homoscedasticity assumption in the ARIMAX model has been resolved. In this case, the ARCH model is better than the GARCH model because the significant lag in the ARCH-LM test tends to be less than 15 lags (Adenomon et al., 2022). The parameter estimation results for the ARIMAX(3,1,3)-ARCH(2) model are then listed in Table 4 (see Appendices).

Diagnostic tests are performed on the residuals of ARIMAX(3,1,3)-ARCH(2) models. The autocorrelation test of residuals is carried out using the Ljung-Box test. It finds that the residuals are independent with a p-value of 0.837, greater than the 5% significance level. The normality test of residuals is then carried out using the Shapiro-Wilk test. This test concludes that the residuals do not follow the normal distribution because they have a p-value of 0.000, which is less than the 5% significance level. The homogeneity of variance test is carried out using the Ljung-Box test of the squared residuals. This test concludes that the variance of the residuals is homogeneous because it has a p-value of 0.654 as it is more than the 5% significance level. Thus, the ARCH/GARCH model succeeds in overcoming violations of the assumption of homogeneity of variance in the ARIMAX model.

Overfitting is done by adding one ARCH order to the ARIMAX(3,1,3)-ARCH(2) model to see whether a better model exists. So, in this case, the candidate for the overfitting model is ARIMAX(3,1,3)-ARCH (3). Based on the results of the ARIMAX(3,1,3)-ARCH(3) model parameter estimation in Table 5 (see Appendices), it can be seen that there are model parameters that are not significant. Then, the AIC value of the ARIMAX(3,1,3)-ARCH(3) model is greater than the ARIMAX(3,1,3)-ARCH(2) model. It makes the ARIMAX(3,1,3)-ARCH(2) model the best in ARCH/GARCH modeling.

Even though ARCH/GARCH modeling has succeeded in overcoming violations of the homoscedasticity assumption, according to Lyu et al. (2021), economics tends to have asymmetric volatility. It can be seen from the histogram of the squared residuals of the ARIMAX(3,1,3) model in Figure 13 (see Appendices). It tends to slant to the right. The existence of asymmetric volatility in the data indicates that TGARCH modeling needs to be done.

Identification of the TGARCH model is carried out by trial and error using the ARIMAX(3,1,3) mean model. Based on trial and error, it is found that the tentative model selected is the ARIMAX(3,1,3)-TGARCH(1,2) model with all significant parameters and the smallest AIC value. The ARCH-LM test on this model produces a p-value of 0.893. It indicates that the violation of the homoscedasticity assumption in the ARIMAX model has been resolved. The ARIMAX(3,1,3)-TGARCH(1,2) model parameter estimates are shown in Table 6 (see Appendices).

A model diagnostic test is carried out on the residuals of ARIMAX(3,1,3)-TGARCH(1,2) models. Based on the Ljung-Box test, a p-value of 0.455 is obtained, which is greater than the 5% significance level. This result shows that the assumption of residual freedom in the ARIMAX(3,1,3)-TGARCH(1,2) model is fulfilled. Furthermore, the Shapiro-Wilk test concludes that the residuals do not follow a normal distribution because they have a p-value of 0.000, which is less than the 5% significance level. Testing the homogeneity of variance using the Ljung-Box test from the squared residual results in the conclusion that the residual variance is homogeneous because it has a p-value of 0.520, greater than the 5% significance level. Thus, the TGARCH model succeeds in overcoming the violation of the homogeneity of variance assumption in the ARIMAX model.

Overfitting is done by adding one ARCH and GARCH order to the ARIMAX(3,1,3)-TGARCH(1,2) model so that the candidate overfitting models formed are ARIMAX(3,1,3)-TGARCH(1,3) and ARIMAX (3,1,3)-TGARCH(2,2). Based on the parameter estimates in Table 7 (see Appendices), it can be seen that there are insignificant parameters in the ARIMAX(3,1,3)-TGARCH(1,3) and ARIMAX(3,1,3)-TGARCH(2,2) models. This result shows that the ARIMAX(3,1,3)-TGARCH(1,2) model is the best in TGARCH modeling.

Validation of the ARIMAX(3,1,3)-ARCH(2) model and ARIMAX(3,1,3)-TGARCH(1,2) model is then carried out by looking at the MAPE values of the four data scenarios in Table 1 (see Appendices). Table 8 (see Appendices) shows that in the first and fourth data scenarios, the ARIMAX(3,1,3)-TGARCH(1,2) model has better performance with a smaller MAPE value. However, in the second and third data scenarios, the ARIMAX(3,1,3)-ARCH(2) model performs better with smaller MAPE values. It shows that if a model is the best in a data condition, it will not necessarily remain the best in another data condition.

In this case, the best model is determined by looking at the average MAPE value of the four data scenarios. Based on the average MAPE value of the four data scenarios, it is found that the best model selected is the asymmetric ARIMAX(3,1,3)-TGARCH(1,2) model. This result follows the residual exploration of the ARIMAX model, which tends to have a histogram that sticks out as an indication of asymmetric effects. Based on the average MAPE value obtained in this validation process, it can be seen that the model used has a very small MAPE value with a very good forecasting category. This accuracy indicates that the use of the ARIMAX model can properly accommodate the influence of COVID-19 on the JCI movement, and the use of the asymmetric GARCH model, namely TGARCH, can properly accommodate the effects of asymmetric volatility on the data.

Forecasting using the ARIMAX(3,1,3)-TGARCH(1,2) model is done in the following ten periods. This period is in the range of May 1st, 2023, to July 3rd, 2023. Based on Figure 14, forecasting shows stable results with an average value of 6824.246. This result aligns with the effects of the COVID-19 pandemic, which starts to wane. Hence, the economy, which is initially fluctuating, tends to become more stable. Even though forecast results tend to be stable, JCI movements may have an upward or downward trend, characterized by forecast confidence intervals that tend to widen in each period.

IV. CONCLUSIONS

The asymmetric GARCH model has better forecasting capabilities than the symmetric GARCH model in forecasting weekly JCI data from January 1st, 2018, to April 24th, 2023, based on model validation using several data scenarios. The asymmetric GARCH model obtained in this case is ARIMAX(3,1,3)−TGARCH(1, 2). Adding a covariate variable in the form of the COVID-19 period into this model significantly influences the JCI movement even though the covariate variable parameters in the ARIMAX(3,1,3) model are not significant. This model produces outstanding forecasting accuracy because it has an average MAPE value of less than 10% from four data scenarios.

The result can be a consideration in future research to prefer an asymmetric model over a symmetric model when predicting JCI data. In this analysis, forecasting JCI data from May 1st, 2023, to July 3rd, 2023, using this model tends to be stable with confidence intervals that widen in each period. Even though the resulting forecast is excellent, the model only uses the COVID-19 variable as a covariate variable. Future research can consider COVID-19 as an intervention using other mean models, such as ARIMA intervention. Future research can also use other variance models, such as exponential GARCH and integrated GARCH. Apart from that, the model in the research is still limited in carrying out long-term forecasting. Future research can use machine learning methods, such as Recurrent Neural Networks (RNN) and Long Short-Term Memory (LSTM) to overcome the model's limitations in the research.

REFERENCES

Adenomon, M. O., Maijamaa, B., & John, D. O. (2022). The effects of COVID-19 outbreak on the Nigerian Stock Exchange performance: Evidence from GARCH models. *Journal of Statistical Modeling & Analytics (JOSMA)*, *4*(1), 25–38. https://doi.org/10.22452/josma.vol4no1.3

Aliyev, F., Ajayi, R., & Gasim, N. (2020). Modelling asymmetric market volatility with univariate GARCH models: Evidence from Nasdaq-100. *The Journal of Economic Asymmetries*, *22*. https://doi.org/10.1016/j.jeca.2020.e00167

Behera, H., Gunadi, I., & Rath, B. N. (2023). COVID-19 uncertainty, financial markets and monetary policy effects in case of two emerging Asian countries. *Economic Analysis and Policy*, *78*, 173–189. https://doi.org/10.1016/j.eap.2023.03.001

Braz, M. S., Sáfadi, T., Ferreira, R. A., Morais, M. H. F., Silva, Z., & Da Rocha, C. M. B. M. (2023). Temporal relationship between human and canine visceral leishmaniasis in an urban area in southeastern Brazil: An application of the ARIMAX model. *Preventive Veterinary Medicine*, *215*. https://doi.org/10.1016/j.prevetmed.2023.105921

Dinku, T., & Worku, G. (2022). Asymmetric GARCH models on price volatility of agricultural commodities. *SN*

Business & Economics, *2*. https://doi.org/10.1007/s43546-022-00355-7
Ekinci, A. (2021). Modelling and forecasting of growth rate of new COVID-19 cases in top nine affected countries: Considering conditional variance and asymmetric effect. *Chaos, Solitons & Fractals*, *151*. https://doi.org/10.1016/j.chaos.2021.111227
Fang, J., Gozgor, G., Lau, C. K. M., & Lu, Z. (2020). The impact of Baidu index sentiment on the volatility of China's stock markets. *Finance Research Letters*, *32*. https://doi.org/10.1016/j.frl.2019.01.011
Haryanto. (2020). Dampak COVID-19 terhadap pergerakan nilai tukar rupiah dan Indeks Harga Saham Gabungan (IHSG). *The Indonesian Journal of Development Planning*, *4*(2), 151–165.
Hismendi, Masbar, R., Nazamuddin, Majid, M. S. A., & Suriani. (2021). Sectoral stock markets and economic growth nexus: Empirical evidence from Indonesia. *The Journal of Asian Finance, Economics and Business*, *8*(4), 11–19. https://doi.org/10.13106/jafeb.2021.vol8.no4.0011
Hussain, F., Ali, Y., Li, Y., & Haque, M. M. (2023). Real-time crash risk forecasting using artificial-Intelligence based video analytics: A unified framework of generalised extreme value theory and autoregressive integrated moving average model. *Analytic Methods in Accident Research*, *40*, 1–21. https://doi.org/10.1016/j.amar.2023.100302
Ichsani, S., Mariana, C., & Andari, D. (2019). Does the Indonesia composite index get affected by the Asia composite index? *International Journal of Innovation, Creativity and Change*, *6*(7), 1–13.
Kasuma, K. A. P., & Nugroho, Y. D. (2020). Tinjauan kasus terkonfirmasi positif COVID-19 terhadap iklim investasi di Indonesia: Peramalan dan korelasi. In *Seminar Nasional Official Statistics* (pp. 190–195). https://doi.org/10.34123/semnasoffstat.v2020i1.720
Kong, Q., Han, J., Jin, X., Li, C., Wang, T., Bai, Q., & Chen, Y. (2023). Polar motion prediction using the combination of SSA and ARMA. *Geodesy and Geodynamics*, *14*(4), 368–376. https://doi.org/10.1016/j.geog.2022.12.004
Kyriazis, N. A., Daskalou, K., Arampatzis, M., Prassa, P., & Papaioannou, E. (2019). Estimating the volatility of cryptocurrencies during bearish markets by employing GARCH models. *Heliyon*, *5*(8), 1–8. https://doi.org/10.1016/j.heliyon.2019.e02239
Lekhal, M., & El Oubani, A. (2020). Does the adaptive market hypothesis explain the evolution of emerging markets efficiency? Evidence from the Moroccan financial market. *Heliyon*, *6*(7), 1–12. https://doi.org/10.1016/j.heliyon.2020.e04429
Lyu, Y., Wei, Y., Hu, Y., & Yang, M. (2021). Good volatility, bad volatility and economic uncertainty: Evidence from the crude oil futures market. *Energy*, *222*. https://doi.org/10.1016/j.energy.2021.119924
Moffat, I. U., & Akpan, E. A. (2019). White noise analysis: A measure of time series model adequacy. *Applied Mathematics,10*(11), 989–1003. https://doi.org/10.4236/am.2019.1011069
Montgomery, D. C., Jennings, C. L., & Kulahci, M. (2015). *Introduction to time series analysis and forecasting*. John Wiley & Sons.
Noor, N. M., Abdullah, M. M. A. B., Yahaya, A. S., & Ramli, N. A. (2015). Comparison of linear interpolation method and mean method to replace the missing values in environmental data set. *Materials Science Forum*, *803*, 278–281. https://doi.org/10.4028/www.scientific.net/MSF.803.278
Putera, M. L. S. (2020). Non-cash payment transaction projection using ARIMAX : Efect of calendar. *Jurnal Matematika, Statistika dan Komputasi*, *16*(3), 296–310. https://doi.org/10.20956/jmsk.v16i3.8546
Raheem, S. H., Alhusseini, F. H. H., & Alshaybawee, T. (2020). Modelling volatility in financial time series using ARCH models. *International Journal of Innovation, Creativity and Change*, *12*(7), 248–261.
Schaffer, A. L., Dobbins, T. A., & Pearson, S. A. (2021). Interrupted time series analysis using Autoregressive Integrated Moving Average (ARIMA) models: A guide for evaluating large-scale health interventions. *BMC Medical Research Methodology*, *21*, 1–12. https://doi.org/10.1186/s12874-021-01235-8
Shahani, R., & Taneja, A. (2022). Dynamics of volatility behaviour and spillover from crude to energy crops: Empirical evidence from India. *Energy Nexus*, *8*, 1–8. https://doi.org/10.1016/j.nexus.2022.100152
Sheng, C., Zhang, D., Wang, G., & Huang, Y. (2021). Research on risk mechanism of China's carbon financial market development from the perspective of ecological civilization. *Journal of Computational and Applied Mathematics*, *381*. https://doi.org/10.1016/j.cam.2020.112990
Somarajan, S., Shankar, M., Sharma, T., & Jeyanthi, R. (2019). Modelling and analysis of volatility in time series data. In J. Wang, G. Reddy, V. Prasad, & V. Reddy (Eds.), *Advances in intelligent systems and computing* (pp. 609–618). Springer. https://doi.org/10.1007/978-981-13-3393-4_62
Vukovic, B. D., & Zinurova, R. Y. (2020). Competitive advantages and sustainable development of Russian agrarian sector. In M. Radović-Marković, B. Đukanović, & N. Vuković (Eds.), *Economy and ecology: Contemporary trends and contradictions 2020* (pp. 203–207).

APPENDICES

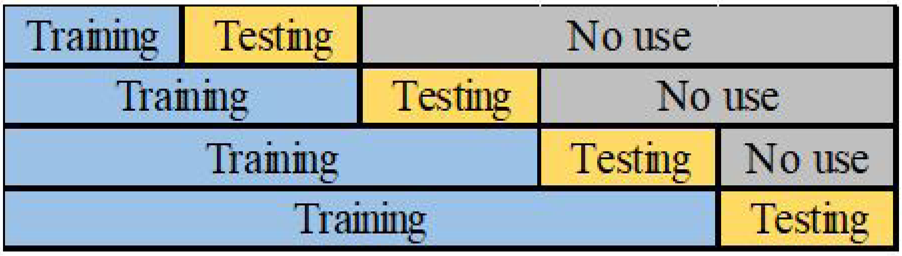

Figure 1 Illustration of the Division of Training and Testing Data

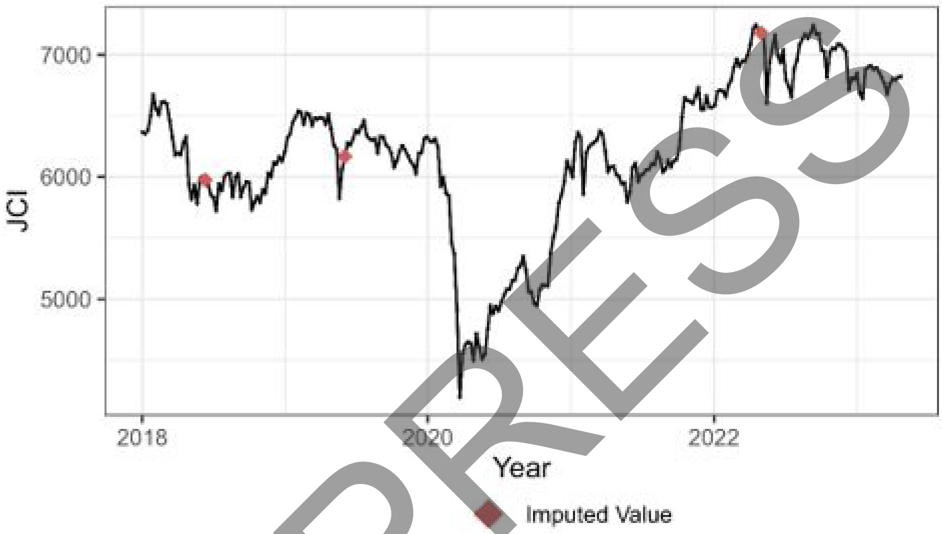

Figure 2 Imputed Time Series Plot

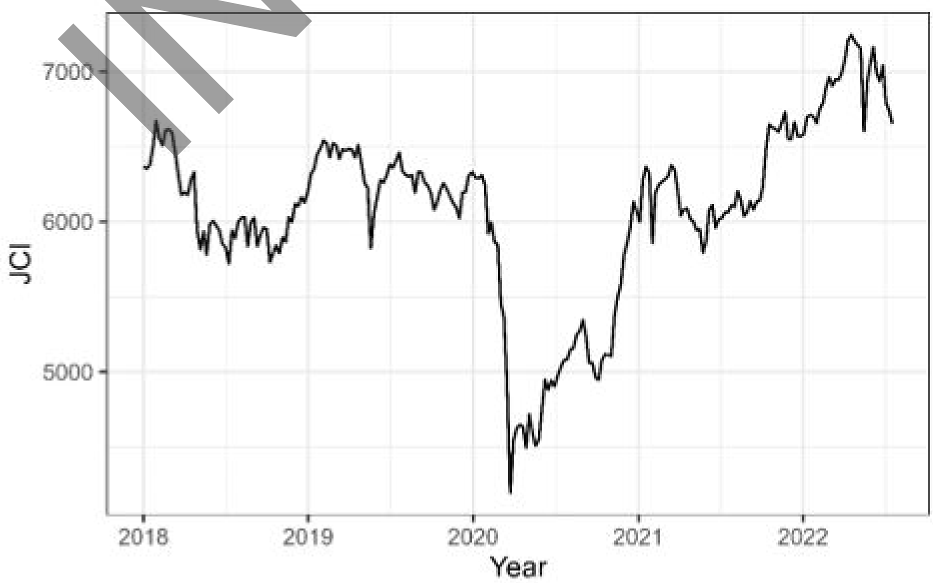

Figure 3 The Jakarta Composite Index (JCI) Time Series Plot Before Differencing

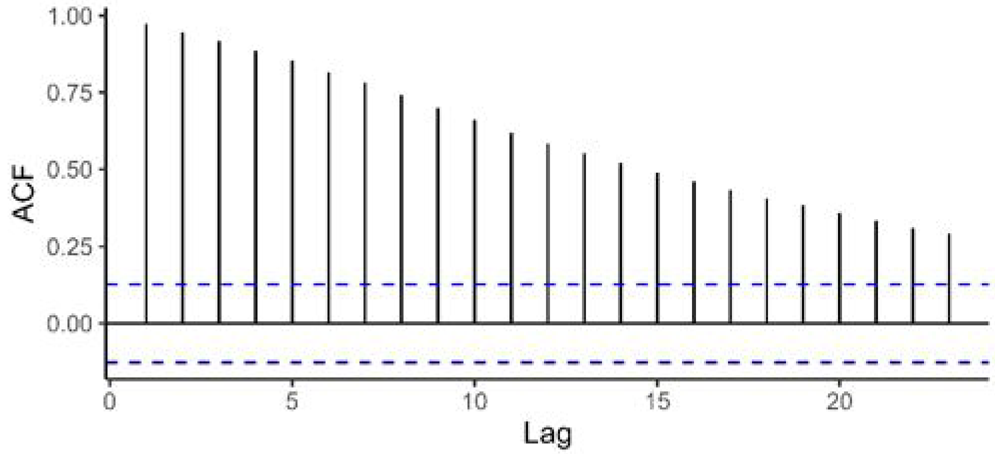

Figure 4 The Autocorrelation Function (ACF) Plot Before Differencing

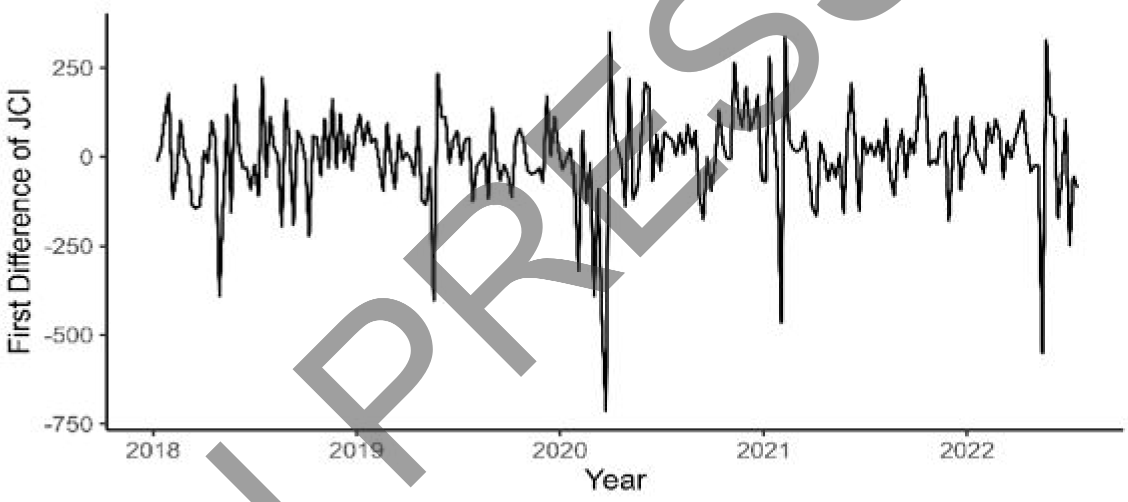

Figure 5 The Jakarta Composite Index (JCI) Time Series Plot After Differencing

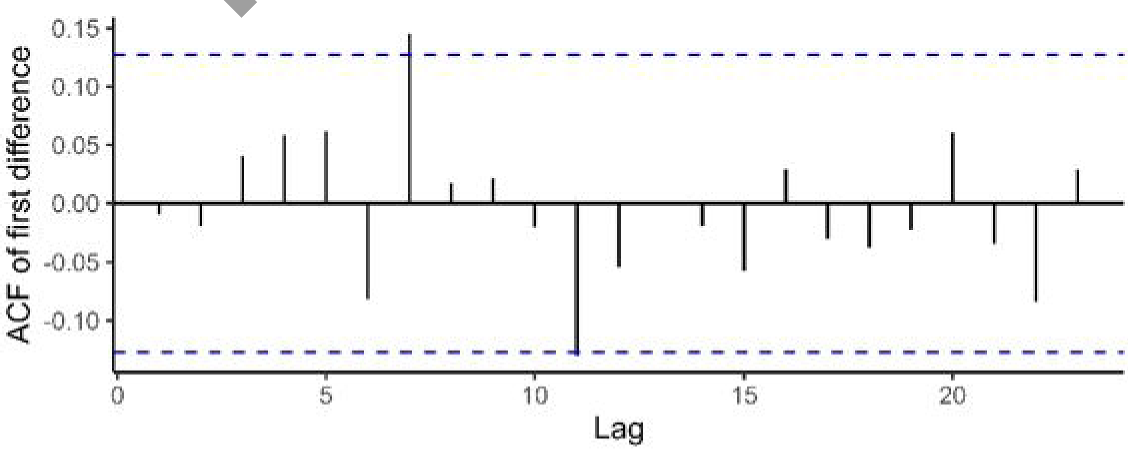

Figure 6 The Autocorrelation Function (ACF) Plot After Differencing

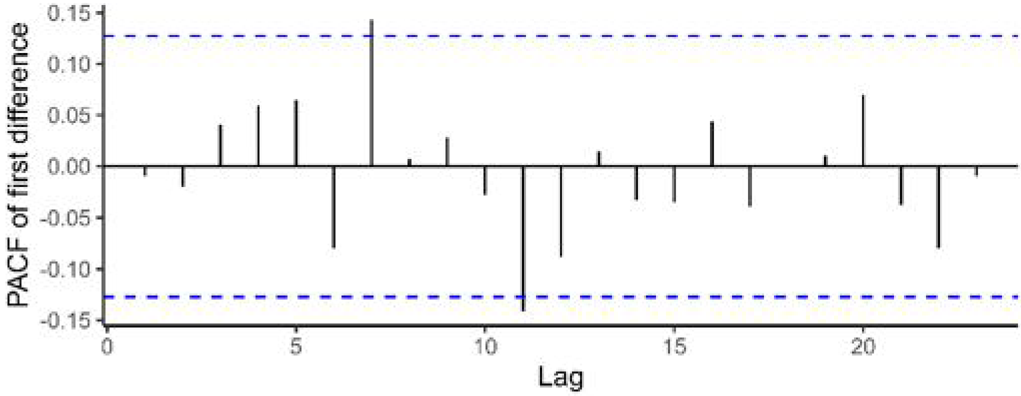

Figure 7 The Partial Autocorrelation Function (PACF) Plot After Differencing

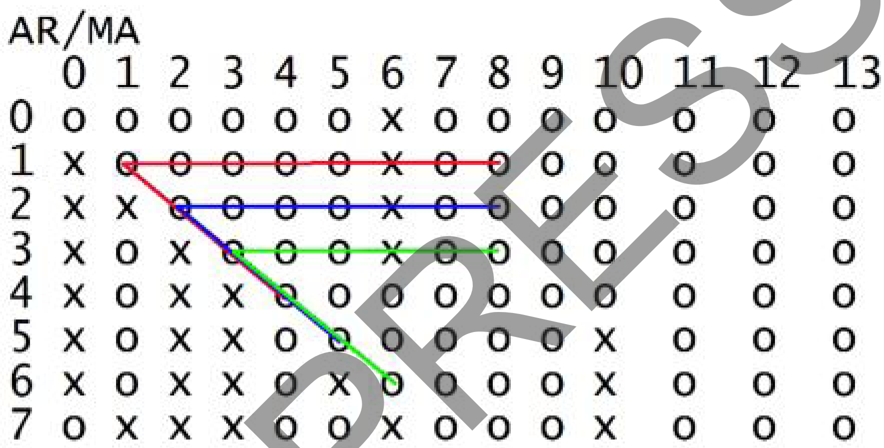

Figure 8 The Extended Autocorrelation Function (EACF) Plot After Differencing

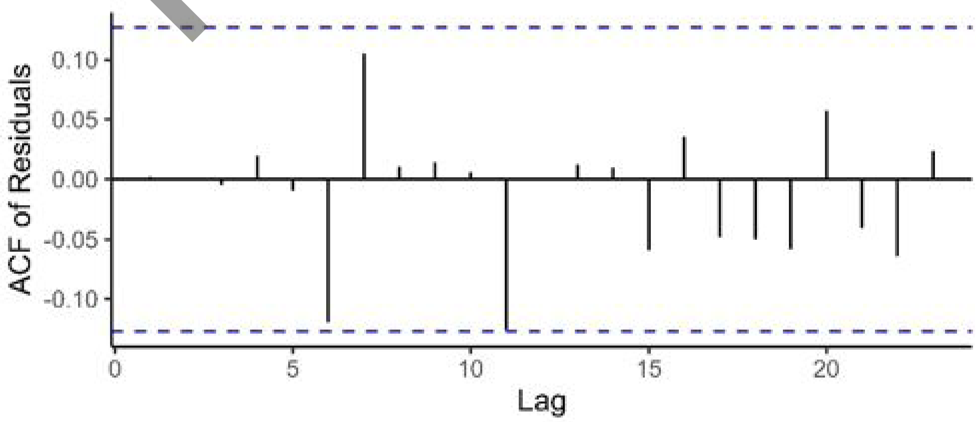

Figure 9 The Autocorrelation Function (ACF) Plot of ARIMAX(3,1,3) Residuals

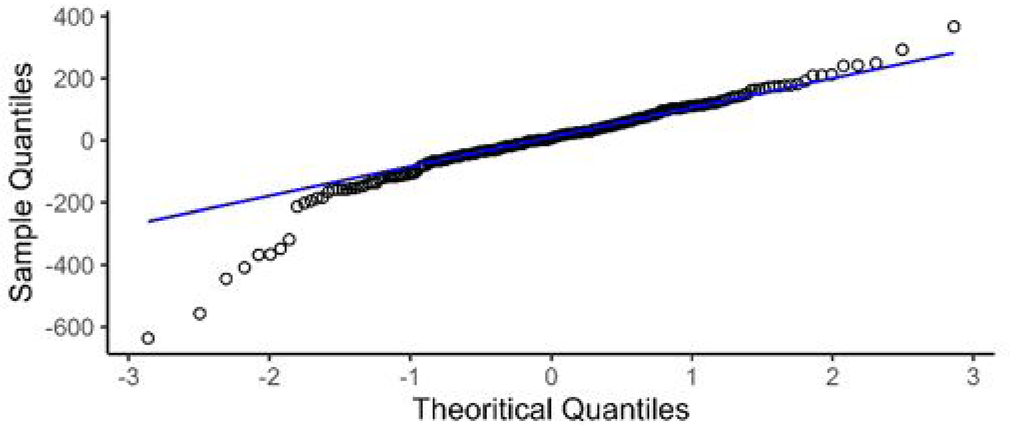

Figure 10 The Quantile Plot of ARIMAX(3,1,3) Residuals

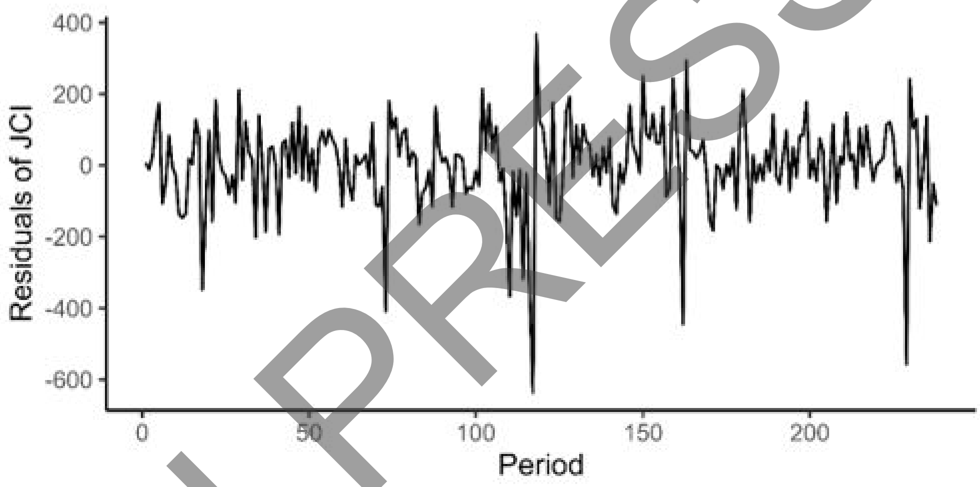

Figure 11 The Time Series Plot of ARIMAX(3,1,3) Residuals

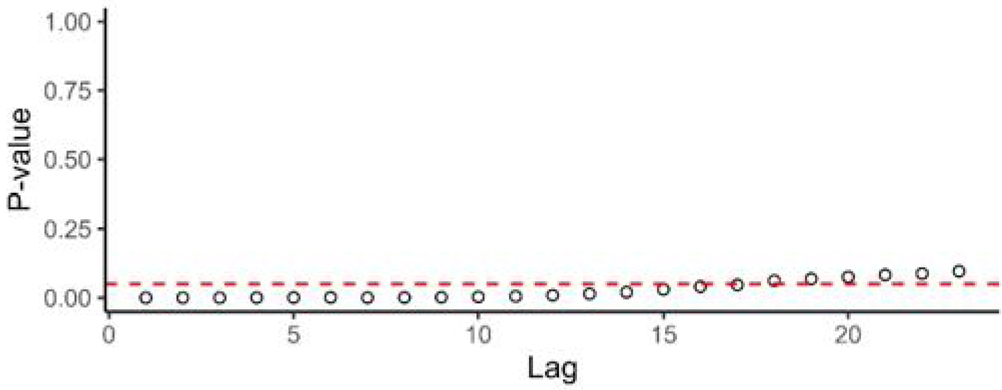

Figure 12 P-Value of McLeod-Li Test

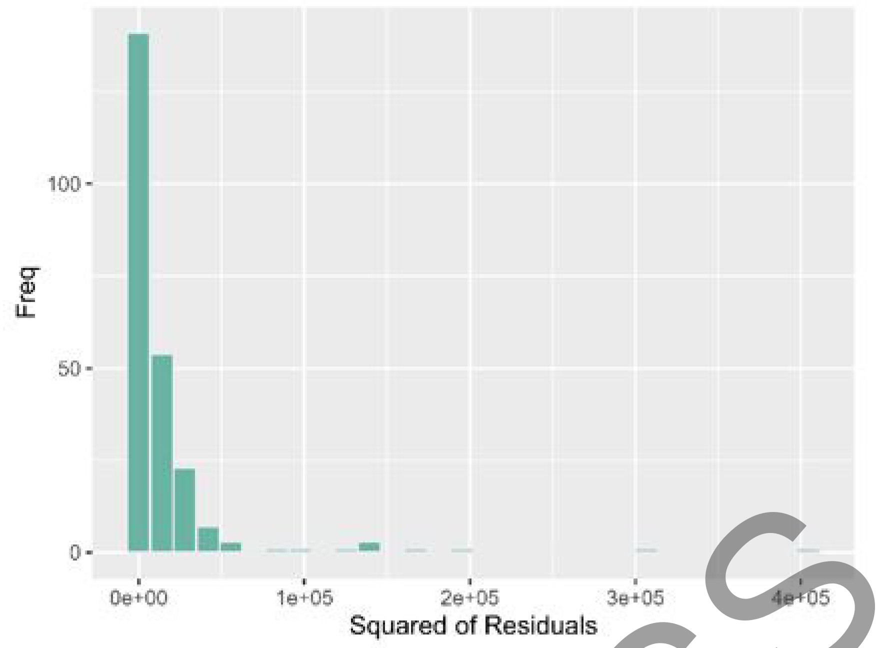

Figure 13 Histogram of Squared Residuals

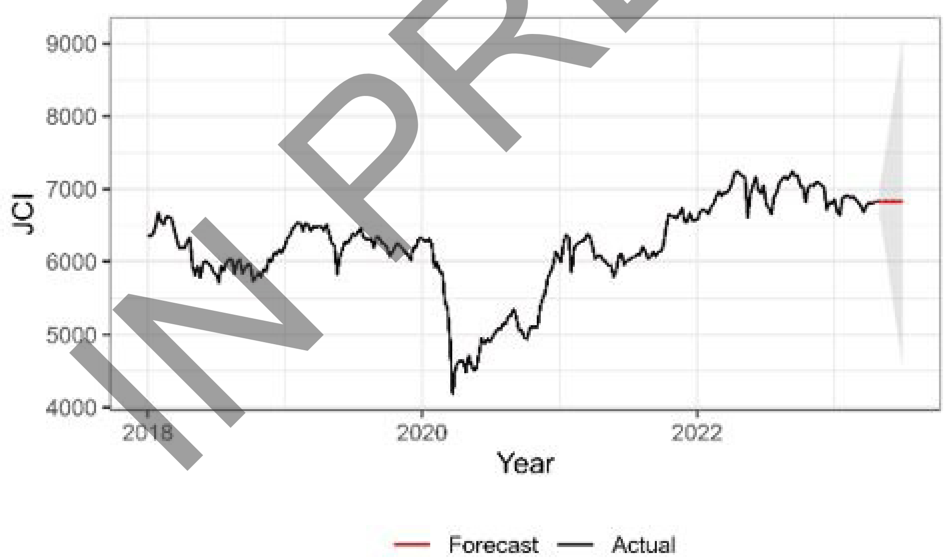

Figure 14 Result of Jakarta Composite Index (JCI) Forecasting on May 1st, 2023 to July 3rd, 2023

Table 1 Period of Training and Testing Data of Each Scenario

Scenario	Training Data	Testing Data
1	Jan. 1st, 2018–July 18th, 2018	July 25th, 2022–Sept. 26th, 2022
2	Jan. 1st, 2018–Sept. 26th, 2022	Oct. 3rd, 2022–Dec. 5th, 2022
3	Jan. 1st, 2018–Dec. 5th, 2022	Dec. 12th, 2022–Feb. 13th, 2023
4	Jan. 1st, 2023–Feb. 13th, 2023	Feb. 20, 2023–April 24th, 2023

Table 2 Parameter Estimation of ARIMA and ARIMAX Models

Model	Parameter	Coefficient	P-Value	AIC
ARIMA (1,1,1)	AR(1)	-0.004	0.999	2998.72
	MA(1)	-0.005	0.999	
ARIMA (2,1,2)	AR(1)	-0.010	0.363	2997.77
	AR(2)	-0.996	0.000	
	MA(1)	0.026	0.142	
	MA(2)	0.999	0.000	
ARIMA (3,1,3)	AR(1)	0.919	0.000	2996.29
	AR(2)	0.866	0.000	
	AR(3)	-0.930	0.000	
	MA(1)	-0.896	0.000	
	MA(2)	-0.913	0.000	
	MA(3)	0.982	0.000	
ARIMAX (3,1,3)	AR(1)	0.697	0.000	2993.50
	AR(2)	0.731	0.000	
	AR(3)	-0.856	0.000	
	MA(1)	-0.755	0.000	
	MA(2)	-0.755	0.000	
	MA(3)	0.999	0.000	
	Dummy	-122.846	0.133	

Note: Autoregressive Integrated Moving Average (ARIMA), Autoregressive Integrated Moving Average with Exogenous Variables (ARIMAX), and Akaike's Information Criterion (AIC).

Table 3 Result of Autoregressive Conditional Heteroskedasticity-Lagrange Multiplier (ARCH-LM) Test

Lag	P-Value
4	0.000
8	0.002
14	0.041
16	0.061
20	0.143

Table 4   Parameter Estimation of ARCH/GARCH Model

Model	Parameter	Coefficient	P-Value	AIC
ARIMAX(1,1,1) - ARCH(2)	AR(1)	1.871	0.000	12.561
	AR(2)	-1.757	0.000	
	AR(3)	0.863	0.000	
	MA(1)	-1.919	0.000	
	MA(2)	1.860	0.000	
	MA(3)	-0.947	0.000	
	Dummy	-8.602	0.019	
		12472.132	0.000	
		0.168	0.000	
		0.086	0.044	

Note: Autoregressive Integrated Moving Average with Exogenous Variables (ARIMAX), Akaike's Information Criterion (AIC), ω (asymmetric constant), α_1 (first order of ARCH parameter), and α_2 (second order of ARCH parameter).

Table 5 Parameter Estimation of Overfitting ARCH/GARCH Model

Model	Parameter	Coefficient	P-Value	AIC
ARIMAX(1,1,1) - ARCH(3)	AR(1)	-0.341	0.000	13.153
	AR(2)	-0.119	0.000	
	AR(3)	-0.592	0.000	
	MA(1)	0.127	0.000	
	MA(2)	0.045	0.000	
	MA(3)	0.915	0.000	
	Dummy	36.184	0.997	
		3.250	0.000	
		0.251	0.000	
		0.293	0.000	
		0.454	0.000	

Note: Autoregressive Integrated Moving Average with Exogenous Variables (ARIMAX), Akaike's Information Criterion (AIC), ω (asymmetric constant), α_1 (first order of ARCH parameters), α_2 (second order of ARCH parameters), and α_3 (third order of ARCH parameters).

Table 6 Parameter Estimation of TGARCH Model

Model	Parameter	Coefficient	P-Value	AIC
ARIMAX(1,1,1) - TGARCH(1,2)	AR(1)	2.172	0.000	12.477
	AR(2)	-1.480	0.000	
	AR(3)	0.256	0.000	
	MA(1)	-2.240	0.000	
	MA(2)	1.620	0.000	
	MA(3)	-0.325	0.000	
	Dummy	25.100	0.011	
		41.956	0.000	
		0.207	0.000	
		0.331	0.025	
		0.169	0.035	
		1.000	0.000	

Note: Autoregressive Integrated Moving Average with Exogenous Variables (ARIMAX), Threshold Generalized Autoregressive Conditional Heteroscedasticity (TGARCH), Akaike's Information Criterion (AIC), ω (asymmetric constant), β_1 (first order of GARCH parameters), β_2 (second order of GARCH parameters), and γ_1 (first order of asymmetric parameters).

Table 7 Parameter Estimation of Overfitting TGARCH Model

Model	Parameter	Coefficient	P-Value	AIC
	AR(1)	1.139	0.000	
	AR(2)	-1.219	0.000	
	AR(3)	0.913	0.000	
	MA(1)	-1.084	0.000	
	MA(2)	1.226	0.000	
	MA(3)	-0.853	0.000	
ARIMAX(1,1,1) -TGARCH(1,3)	Dummy	19.393	0.283	12.497
		48.856	0.000	
		0.262	0.000	
		0.029	0.793	
		0.000	1.000	
		0.390	0.003	
		1.000	0.000	
	AR(1)	0.932	0.000	
	AR(2)	0.882	0.000	
	AR(3)	-0.959	0.000	
	MA(1)	-0.914	0.000	
	MA(2)	-0.022	0.000	
	MA(3)	1.001	0.000	
ARIMAX(1,1,1), TGARCH(2,2)	Dummy	25.501	0.036	12.476
		38.427	0.000	
		0.224	0.000	
		0.000	0.999	
		0.208	0.115	
		0.306	0.028	
		1.000	0.000	
		0.083	0.988	

Note: Autoregressive Integrated Moving Average with Exogenous Variables (ARIMAX), Threshold Generalized Autoregressive Conditional Heteroscedasticity (TGARCH), Akaike's Information Criterion (AIC), ω (asymmetric constant), α_1 (first order of ARCH parameters), α_2 (second order of ARCH parameters), β_1 (first order of GARCH parameters), β_2 (second order of GARCH parameters), γ_1 (first order of asymmetric parameters), and γ_2 (second order of asymmetric parameters).

Table 8 MAPE Value of Each Scenario

Scenario	MAPE Value	
	ARIMAX(3,1,3)-ARCH(2)	ARIMAX(3,1,3)-TGARCH(1,2)
1	6.830%	5.994%
2	2.159%	2.193%
3	3.203%	3.245%
4	1.906%	1.199%
Average	3.525%	3.158%

Note: Autoregressive Integrated Moving Average with Exogenous Variables (ARIMAX), Autoregressive Conditional Heteroskedasticity (ARCH), Mean Absolute Percentage Error (MAPE), and Threshold Generalized Autoregressive Conditional Heteroscedasticity (TGARCH).