Estimating and Forecasting Composite Index in Pandemic Era Using ARIMA-GARCH Model

Agus Sofian Eka Hidayat¹, Gilang Primajati²

¹Actuarial Science, President University, Indonesia
²Mathematics Education, Universitas Mataram, Indonesia

ABSTRACT

Many industries have suffered financial losses as a result of the COVID-19 epidemic. The stock market’s movement has been impacted by this circumstance. Due to the influence of some people, a large number of individuals with limited trading knowledge are attempting to participate in the stock market. Market volatility should be understandable in order to gain profit instead of having losses. Therefore, it’s essential to comprehend the market of the future through analysis of the data. The purpose of this study is to use ARIMA-GARCH to predict the Indonesian stock market price during. The period covered by the dataset is January 2020-December 2022. The training data indicates that ARIMA (2,1,2) is the best model for ARIMA. The results showed that data residual fitted by ARIMA (2,1,2)-GARCH (1,2) exhibits heteroscedasticity, according to the residual analysis. The MAPE score is 2%, which is relatively small. It means that ARIMA (2,1,2)-GARCH (1,2) is good enough for forecasting the Jakarta Composite Index.

A. INTRODUCTION

The Covid-19 outbreak, first detected on March 2, 2020, in Indonesia, led to new regulations that significantly impacted economic activities. This crisis caused substantial financial losses for businesses, resulting in salary reductions and layoffs, which particularly affected millennial and Gen-Z workers (Fadly, 2021). In response, many of these individuals turned to the capital market to generate additional income. Two primary factors driving millennials’ and Gen-Z’s interest in stock investment were the new opportunities and the increased spare time to explore stock investments, along with a growing awareness of the importance of saving and investing for the future. Systematic risk cannot be mitigated through diversification, as this risk originates from fluctuations in market conditions. Market conditions are affected by external economic factors and estimating systematic risk is a key factor in investment decisions (Robbetze and Swanepoel, 2022) (Puspitaningtyas, 2017).

There are multi elements that has certain impacts on stock price of single firm, such as operation, dividend, performance, and macro elements (Thach and Huy, 2021). A notable development in the financial landscape during this period was the increased volatility in financial time series data, including stock price indexes. High volatility, marked by rapid and unpredictable stock price fluctuations, led to market instability (Aklimawati and Wahyudi, 2013; Septiana et al., 2021). During the second and third quarters of 2020, the Jakarta Composite Index (JKSE) experienced significant volatility, attributed to a surge in retail investors engaging in frequent and high-risk transactions (Fadly, 2021). Retail investors and stock influencers played a crucial role in this phenomenon by entering the market during declines, which increased the number of investors, transaction frequency, and transaction value. However,
this also introduced risks associated with high volatility driven by sentiment (Utami, 2021).

Amid the increased market risk, managing market risk for traded assets, especially stocks, became crucial by predicting and estimating future volatility. This modeling process employs Autoregressive Conditional Heteroscedasticity (ARCH) and Generalized Autoregressive Conditional Heteroscedasticity (GARCH). Given that each market has unique characteristics, performance metrics, and sizes, selecting the best model requires meticulous specification to ensure precise estimation of stock return volatility. Greater precision in volatility estimation reduces decision-making risks for investors. This research focuses on identifying the best-fitting model by comparing actual and forecasted investment values using the GARCH model, as demonstrated in previous studies (Lin, 2018; Sharma et al., 2016; Virginia et al., 2018).

The distinction of this research lies in its exploration of the COVID-19 pandemic’s impact on the capital market, and the unique characteristics of the Jakarta Composite Index (JKSE). There are many research that already focused on company stock in Indonesia (Septiana et al., 2021; Yolanda et al., 2017; Iqbal and Ningsih, 2021) but view of them discuss about Indonesia stock market. The novelty of this research aims to focus on the prediction of the Indonesian stock market to give an understanding of millennial and Gen-Z investors who turned to the stock market during the pandemic. Another objective is to provide insights into effective risk management and decision-making strategies for investors during times of heightened volatility in the financial market, particularly in developing economies like Indonesia.

B. RESEARCH METHOD

1. Time Series Analysis

Time series analysis is a technique for analyzing a set of data from the past over a period to estimate and forecast future occurrences. According to Chatfield and Xing (2019) there are two types of time series: discrete time series and continuous time series. The difference between discrete and continuous time series lies in the type of observation set. Discrete time series have a discrete set, while continuous time series have a continuous time interval. Time series is known to be denoted as \{Y_t\}. The Y is a variable for time series, subscript t refers to time and if t = 1, it becomes the first observation, and t = T is the last observation. A complete observation period is defined as t = 0, ±1, ±2, ±3, . . ., T. The observation can be measured in the terms of several intervals, such as annually, monthly, weekly, daily, hourly, etc (Mills, 2019; Tsay, 2005).

2. Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF)

Wei (2018) explained that Y_t as a stationary process has the constant mean \(E(Y_t) = \mu\), the constant variance \(Var(Y_t) = E(Y_t - \mu)^2 = \sigma^2\), along with the covariance \(Cov(Y_t, Y_s)\) which are functions impacted by time difference \((t - s)\). Therefore, the covariance between \(Y_t\) and \(Y_{t+k}\) can be written as

\[
\gamma_k = Cov(Y_t, Y_{t+k}) = E[(Y_t - \mu)(Y_{t+k} - \mu)]
\] (1)

and the autocorrelation between \(Y_t\) and \(Y_{t+k}\) is

\[
\rho_k = \frac{Cov(Y_t, Y_{t+k})}{\sqrt{Var(Y_t)Var(Y_{t+k})}} = \frac{\gamma_k}{\gamma_0}
\] (2)

where \(Var(Y_t) = Var(Y_{t+k}) = \gamma_0\).

The partial autocorrelation function (PACF) is used to measure the correlation between \(Y_t\) and \(Y_{t+k}\) after their mutual linear dependency on the intervening variables \(Y_{t+1}, Y_{t+2}, ..., Y_{t+k-1}\) has been removed. The conditional correlation is \(Covr(Y_t, Y_{t+k} | Y_{t+1}, ..., Y_{t+k-1})\). Consider \(\rho_k\) denote the partial autocorrelation between \(Y_t\) and \(Y_{t+k}\), the equation given as

\[
\rho_k = \frac{Cov[(Y_t - \hat{Y}_t), (Y_{t+k} - \hat{Y}_{t+k})]}{\sqrt{Var(Y_t - \hat{Y}_t)Var(Y_{t+k} - \hat{Y}_{t+k})}}
\] (3)

after all the assumptions related, the equation of \(\rho_k\) becomes.

\[
\rho_k = \frac{\gamma_k - \alpha_1\gamma_{k-1} - \cdots - \alpha_k\gamma_1}{\gamma_0 - \alpha_1\gamma_1 - \cdots - \alpha_k\gamma_{k-1}}
\]

\[
\rho_k = \frac{\rho_k - \alpha_1\rho_{k-1} - \cdots - \alpha_k\rho_1}{\gamma_0 - \alpha_1\rho_1 - \cdots - \alpha_k\rho_{k-1}}
\] (4)

Vol. 7, No. 2, April 2024, Hal. 115–123
DOI: https://doi.org/10.30812/varian.v7i2.2103
3. Stationary and Non-Stationary Model

3.1. Autoregressive Model (AR)

The autoregressive model has a function to identify the linear relationship between recent values and past values (Mills, 2019). The equation of the autoregressive model, \( AR(p) \) is given by

\[
Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \ldots + \phi_p Y_{t-p} + e_t
\]

(5)

where,

\( Y_t \): Linear combination of the \( p \) most recent past values
\( e_t \): Error at time \( t \) which is not explained by past values; Independent of \( Y_{t-1}, Y_{t-2}, Y_{t-3}, \ldots \)

3.2. Moving Average Model (MA)

Moving average process is where there is a finite number of -weights are nonzero only (Cryer et al., 2008; Shumway & Stoffer, 2017) It is also called a moving average of order \( q \) or \( MA(q) \). The equation is

\[
Y_t = e_t - \theta_1 e_{t-1} + \theta_2 e_{t-2} + \ldots + \theta_q e_{t-q}
\]

(6)

where,

\( Y_t \): The weights of 1, \(-\theta_1, -\theta_2, \ldots, -\theta_q\) to the variables \( e_t, e_{t-1}, e_{t-2}, \ldots, e_{t-q}\).
\( Y_{t+1} \): move the weights and apply it to \( e_{t+1}, e_t, e_{t-1}, \ldots, e_{t-q+1}\).

3.3. Mixed Autoregressive Moving Average Model (ARMA)

Shumway and Stoffer (2017) and stated that autoregressive moving average process is where the series is partly autoregressive process and moving average process. The equation of ARMA (\( p,q \)) can be obtained by

\[
Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \ldots + \phi_p Y_{t-p} + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \ldots - \theta_q e_{t-q}
\]

(7)

where \( Y_t \) is the mixed autoregressive moving average of order \( p \) and \( q \) (Cryer et al., 2008).

3.4. Integrated Autoregressive Moving Average Model (ARIMA)

The \( d \)th difference \( W_t = \nabla^d Y_t \) is an integrated model of the stationary ARMA process. It is also called the \( ARIMA(p,d,q) \) process. Shumway and Stoffer (2017) explained that the differencing process generally stops at \( d = 1 \) or at most \( d = 2 \). The equation of ARIMA (\( p,1,q \)) with \( W_t = Y_t - Y_{t-1} \) is

\[
W_t = \phi_1 W_{t-1} + \phi_2 W_{t-2} + \ldots + \phi_p W_{t-p} + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \ldots - \theta_q e_{t-q}
\]

(8)

or

\[
Y_t = (1 + \phi_1)Y_{t-1} + (\phi_2 - \phi_1)Y_{t-2} + \ldots + (\phi_p - \phi_{p-1})Y_{t-p} - \phi_p Y_{t-p-1} + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \ldots - \theta_q e_{t-q}
\]

(9)

and the characteristic polynomial is

\[
(1 - \phi_1 x - \phi_2 x^2 - \ldots - \phi_p x^p)(1-x)
\]

(Cryer et al., 2008).

4. Autoregressive Conditional Heteroscedasticity (ARCH)

The ARCH model is proposed by Engle in 1982 for modeling the changing variance of time series in the use to forecast the conditional variances in the future. The equation of this model can be obtained as (Paolella, 2018; Tsay, 2005),

\[
\sigma_t^2 = \omega + \alpha_1 r_{t-1}^2 + \alpha_2 r_{t-2}^2 + \ldots + \alpha_q r_{t-q}^2
\]

(10)

where, \( \omega > 0, \alpha_i \geq 0 \) for \( i > 0 \), and \( r_{t-j}^2; j = 1, 2, \ldots \) are the components of ARCH.
5. Generalized Autoregressive Conditional Heteroscedasticity (GARCH)

In the term of increasing the accuracy for forecasting future volatilities, not only the most recent squared returns but all past squared returns with lesser weight need to be included. It becomes the combined model called GARCH \((p,q)\) model where \(p\) is the GARCH order and \(q\) is the ARCH order (Paolella, 2018; Sharma et al., 2016):

\[
\sigma_t^2 = \omega + \sum_{i=0}^{q} \alpha_i r_{t-i}^2 + \sum_{i=0}^{p} \beta_j \sigma_{t-j}^2
\] (11)

6. Best Model Selection

The best model can be obtained by the method that is proposed by Akaike's Information Criteria. The method selects the model by minimizing:

\[
AIC = -2 \log(\text{maximum likelihood}) + 2k
\]

Where \(k = p + q + 1\) if the model is an intercept or a constant term and vice versa for \(k = p + q\) (Chatfield and Xing, 2019; Grønneberg and Hjort, 2014).

7. Forecasting

The forecast process refers to \(ARMA(P,Q) - GARCH(p,q)\) model (Mills, 2019; Paolella, 2018). The equation for forecasting \(Y_{T+h}\) as the mean equation of \(ARMA(P,Q)\):

\[
Y_{T+h} = \phi_1 Y_{T+h-1} + \ldots + \phi_{p+d} Y_{T+h+p-d} + e_{T+h} - \theta_1 e_{T+h-1} - \ldots - \theta_q e_{T+h-q}
\] (12)

and the equation to forecast the error variances of \(GARCH(p,q)\):

\[
V(e_{t,h}) = \sigma_{T+h}^2 + \psi_1^2 \sigma_{T+h-1}^2 + \ldots + \psi_{h-1}^2 \sigma_{T+1}^2
\] (13)

where,

\[
\sigma_{T+h}^2 = \omega + \alpha_1 r_{T+h-1}^2 + \ldots + \alpha_p r_{T+h-p}^2 + \beta_1 \sigma_{T+h-1}^2 + \ldots + \beta_q \sigma_{T+h-q}^2
\] (14)

8. Data Source

The data obtained is secondary data taken from the official Yahoo Finance website (https://finance.yahoo.com). The data used is the adjustment close price of Jakarta Composite Index (JKSE) data or also known as the Indeks Harga Saham Gabungan (IHSG) in the period range of January 2020 to December 2022. The data are divided into training data and test data. The training data used for building the model, the test data is used to compare the actual data with the forecast data from the model obtained.

Figure 1. Data Adjustment Closed Price JKSE 2020-2022 (Source: Yahoo Finance)
From the historical data, the price index of JKSE drop significantly in the first quarter of 2022 and continue to increase but also experienced decrease during from April 2020 until December 2022.

C. RESULT AND DISCUSSION

1. Stationary Test

Stationary test is a test to see check whether the time series has trend or seasonal component. Time series with stationary series is easier for having effective and prices prediction. In stationary test process, the data is being tested using Augmented Dickey-Fuller (ADF) test. ADF test shown that for every lag the p-value > 0.01, which means that we accept the null Hypothesis, $H_0$, that there is non-stationary in the testing data. Ljung-Box test statistics is 11766, and the p-value is < 0.01 so there is no evidence of non-zero autocorrelation in the sample data at lags 1-20. This fact also strengthen with the correlation diagram of Autocorrelation function (ACF) and Partial Autocorrelation function (PACF). From the figure 2, for every lag in ACF shows that all value shown are significantly far from zero and the only pattern is perhaps a linear decrease with increasing lag.

![Figure 2. ACF and PACF of Testing Data](image1)

To obtain stationary conditions, the first differencing of the training data is used. The first differencing gives a p-value $\leq$ 0.01, hence rejecting the null hypothesis, $H_0$, which means that the data is now stationer. In Figure 3, the correlation diagram has changed, especially in ACF, compared with the ACF in Figure 2. The output of coefficient ACF is close to 0 at lag 3 and for other lags, the values are relatively small. The cut-off happened in lag 3 while in PACF the cut-off happened in lag 3. It indicates that in the ARIMA model with $AR(p)$ and MA(q) with first differencing the estimation model of ARIMA will be tested from ARIMA (0,1,0) up to ARIMA (3,1). The best model will be selected from the smallest AIC based on the maximum likelihood formula.

![Figure 3. ACF and PACF of Testing Data](image2)
From several estimated model the best ARIMA model for this data ia ARIMA (2,1,2) with the smallest AIC=7418.37 among the other ARIMA(p,d,q) (see Table 1). ARIMA (2,1,2) has coefficient AR(1) is -1.4047, AR(2) is -0.8820, MA(1) is 1.4258, and MA(2) is 0.7851. Hence the Time series model become

\[ W_t = -1.4047W_{t-1} + (-0.8820)W_{t-2} + e_t - 1.4258e_{t-1} - 0.7851e_{t-2} \]

Table 1. Sample of AIC from Estimated Model

<table>
<thead>
<tr>
<th>ARIMA (p,d,q)</th>
<th>AIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>0,1,1</td>
<td>7426.9</td>
</tr>
<tr>
<td>1,1,2</td>
<td>7425.82</td>
</tr>
<tr>
<td>2,1,2</td>
<td>7418.37</td>
</tr>
<tr>
<td>3,1,2</td>
<td>7420.71</td>
</tr>
<tr>
<td>3,1,3</td>
<td>7420.64</td>
</tr>
</tbody>
</table>

Table 1 shows several examples of ARIMA model estimation simulations with various values with AIC values produced in ARIMA(p,d,q). The best ARIMA model for this data ia ARIMA(2, 1, 2) with the smallest AIC=7418.37 among the other ARIMA(p,d,q).

2. ARCH Effect

To continue the process of using the GARCH model, the heteroscedasticity test is performed. GARCH models are applied when the error term’s variance fluctuates, indicating heteroskedasticity. This term refers to the inconsistent variation pattern of an error term or variable within a statistical model. In essence, when heteroskedasticity is present, the observations deviate from a linear pattern. The heteroscedasticity test used is the Lagrange-Multiplier Test (LM test). If the residual (error) of the data shows heteroscedasticity then the process continues to GARCH. The null hypothesis \( H_0 \) is there is no heteroscedasticity in the residual, while the alternative \( H_a \), there is heteroscedasticity in the residual.

The residual of the data is the normal distribution with the residual plot given in Figure 4. The plot in Figure 4. Shows that there are some outliers from the residual which indicates there is heteroscedasticity. The p-value test also has a value < 0.05, which means the \( H_0 \) rejected, so there is heteroscedasticity in the residual. Further, the models have an ARCH effect. Then the estimation process can be estimated using the ARCH/GARCH Model.

3. GARCH Modeling

The selection of the best \( GARCH(p,q) \) model is determined using the ARCH-LM p-value, the smallest AIC value, and the significance of the parameter. From Table 2, we can see that the \( ARIMA(2,1,2) - GARCH(1,2) \) has the smallest AIC value with 11.031 while the maximum AIC appeared in \( ARIMA(2,1,2) - GARCH(2,2) \) with the value is 11.057. Hence the best model of the ARIMA-GARCH model is \( ARIMA(2,1,2) - GARCH(1,2) \) with the parameter \( \omega = 514.019481, \alpha_1 = 0.148498, \alpha_2 = 0.009139, \beta_1 = 0.715101 \). Hence from equation (11), the GARCH equation can be obtained as \( \sigma_t^2 = 514.019481 + 0.148499\sigma_{t-1}^2 + 0.009139\sigma_{t-2}^2 + 0.715101\sigma_{t-1}^2 \).

Table 2. Sample of AIC from Estimated Model

<table>
<thead>
<tr>
<th>ARIMA (2,1,2)-GARCH(p,q)</th>
<th>AIC</th>
<th>No Serial Correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>ARIMA(2,1,2)-GARCH(1,2)</td>
<td>11.031</td>
<td>Yes</td>
</tr>
<tr>
<td>ARIMA(2,1,2)-GARCH(2,2)</td>
<td>11.057</td>
<td>No</td>
</tr>
</tbody>
</table>

DOI: https://doi.org/10.30812/varian.v7i2.2103
Moreover the ARCH-LM test for lag [4], lag [6], and lag [8] shows the p-value 0.8857, 0.9533, and 0.9566 respectively. As the p-value $> 0.05$ then the models fulfill the assumption of homogeneity of variance or the variance is equal.

4. Forecasting

Forecasting of the time series is based on the ARIMA (2,1,2)-GARCH (1,2) model. The result of forecasting using the model has been determined in Table 3, which shows the exact forecast series 74 days ahead, from September 20, 2022, to December 30, 2022. Then we compare the result of the actual and the forecast data to obtain the Mean Absolute Percentage Error (MAPE). The MAPE score is 2% which is relatively small. It means that ARIMA (2,1,2)-GARCH (1,2) is good enough for forecasting the Jakarta Composite Index.

<table>
<thead>
<tr>
<th>Day</th>
<th>Actual</th>
<th>GARCH Forecast</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>7196.951</td>
<td>6913.78</td>
</tr>
<tr>
<td>2</td>
<td>7188.314</td>
<td>6911.284</td>
</tr>
<tr>
<td>3</td>
<td>7218.906</td>
<td>6951.955</td>
</tr>
<tr>
<td>4</td>
<td>7178.583</td>
<td>6974.915</td>
</tr>
<tr>
<td>5</td>
<td>7127.503</td>
<td>7101.185</td>
</tr>
<tr>
<td>6</td>
<td>7112.449</td>
<td>6969.412</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>72</td>
<td>6850.52</td>
<td>7189.106</td>
</tr>
<tr>
<td>73</td>
<td>6860.077</td>
<td>7117.862</td>
</tr>
<tr>
<td>74</td>
<td>6850.619</td>
<td>7140.129</td>
</tr>
</tbody>
</table>

In comparison, Figure 5 shows that the graph of actual data and the forecast data using ARIMA (2,1,2)-GARCH (1,2) shows that the forecast data can predict the actual data.

Seen in Figure 5, the research results show that the difference between the graph of actual data (blue curve) and forecast data (yellow dotted curve) is very striking. This is in line with previous research (Aklimawati and Wahyudi, 2013; Septiana et al., 2021). So the meaning of this research is good enough to predict the Composite Stock Price Index.
D. CONCLUSION AND SUGGESTION

The conclusion of this study highlights that the ARIMA (2,1,2) model, in conjunction with its variance component in GARCH(1,2) emerges as the optimal choice for modeling Jakarta Composite Index (JKSE) data during the pandemic era. The MAPE score is 2% which is relatively small. It means that ARIMA (2,1,2)-GARCH (1,2) is good enough for forecasting the Jakarta Composite Index. This selection was based on the strict criterion of minimizing the AIC value, distinguishing it from other models. This innovation holds considerable value for researchers and investors aiming to navigate volatile markets. The implications of this research impact financial decision-making. The identified model and its forecasts offer valuable insights for investors and financial professionals, helping them make more informed investment choices during times of increased uncertainty, such as a pandemic. However, it’s crucial to acknowledge the study’s limitations. Firstly, the data used in this research only goes up to December 2022, and the ongoing pandemic may have introduced dynamic changes in the financial landscape. Additionally, no model is perfect, and the selected ARIMA-GARCH model has its potential shortcomings, necessitating careful risk management.

In conclusion, this study’s innovative modeling approach, its implications for financial decision-making, and its acknowledgment of limitations highlight its importance. Future research could extend the dataset, explore alternative modeling techniques like EGARCH, GJR-GARCH, TGARCH, IGARCH, APARCH, and CGARCH, and conduct cross-country comparisons to evaluate the robustness of these findings in different developing economies, such as Malaysia, Thailand, and Vietnam. These efforts can collectively enhance our understanding of financial market behavior during crises and contribute to more effective risk management strategies.

REFERENCES


