

The Application of the Arima – Garch Method in Predicting the Indosat Stock Price

Maktisen Ena¹, Erna Tri Herdiani², Georgina Maria Tinungki³

^{1,2,3}Department of Statistics, Faculty of Mathematics and Natural Sciences, Hasanuddin University, Makassar 90245, Indonesia
Corresponding Author's : herdiani.erna@gmail.com and enatisen06@gmail.com

Abstract— Indosat stock price data has a trend that fluctuates from time to time, therefore the GARCH model can predict and answer these fluctuating stocks. This study discusses the prediction of Indosat stock prices, the data used are 2399 data divided into 2 parts, namely 1919 training data and 480 testing data. The purpose of this study is to model Indosat stock using the ARIMA-GARCH model. The ARIMA (0,1,9) and GARCH (1,2) models are the best models for predicting Indosat shares with an RMSE value of 0.1697524.

Keywords— Indosat Stocks, ARIMA, GARCH, Predictions.

I. BACKGROUND

According to Darmadji and Fakhruddin (2012) stock prices are prices that occur on the stock exchange at a certain time. Stock prices can change up or down in minutes or even seconds so fast. This is possible because it depends on the demand and supply between the stock buyers and the stock sellers.” Stock prices are volatile. Fluctuating stock prices are determined by the company’s ability to gain profits. If the profits obtained for the company are relatively high, it will affect the share price to be paid high as well.

Time series data can be modeled using time series models (ARIMA and ARCH/GARCH), Black-Scholes model, MIMIC model (Multiple Indicators and Multiple Causes), Markov Switching model, and Time-Varying Coefficient model. In this paper, the model used is the ARCH/GARCH time series model. Time series data on stock price movements have a variance that is not constant at any time. Such time series data conditions are called conditional heteroscedasticity, the assumptions for general time series models such as Autoregressive (AR), Moving Average (MA), Autoregressive Moving Average (ARMA), and Autoregressive Integrated Moving Average (ARIMA) are not met. One of the time series models that can overcome heteroscedasticity is the Autoregressive conditional heteroscedasticity (ARCH) model introduced by Engle (1982). This model can describe all the characteristics of financial market variables. However, in financial problems with a higher level of volatility, this model requires a large order to obtain the right model. To avoid this, Bollerslev (1986) developed the ARCH model into Generalized Autoregressive Conditional Heteroscedasticity (GARCH). According to Ariefianto (2012), the GARCH model has the characteristics of a symmetrical volatility response to shocks. In other words, if the nominal is the same, then the volatility response to shocks is the same, both positive (good news) and negative (bad news) shocks. On diagnostic examination, residual variance is assumed to be constant. The data used is the closing value of Indosat shares from March 6, 2012, to January 18, 2021, totaling 2399 data,

which will be divided into two parts, namely 80% training data to 1919 data and 20% testing data to 480 data. The data is classified as high volatility because it has a high conditional variance, one method to calculate the high volatility data model is Generalized Autoregressive Conditional Heteroscedasticity (GARCH).

II. THEORETICAL FOUNDATION

2.1. ARIMA

The ARIMA model is often used in analyzing non-stationary time series data, the general model without seasonal patterns is as follows:

$$\phi_p(B)(1 - B)^d Z_t = \theta_0 + \theta_q(B)a_t \tag{1}$$

2.2. GARCH

The Generalized Autoregressive Heteroscedasticity (GARCH) model is a development of the ARCH model. This model is able to avoid orders that are too high in the ARCH(m) model. In this model, the conditional variance is not only influenced by previous errors but also by the conditional variance itself (Ariefianto, 2012). The conditional variance in the GARCH model consists of two components, namely the squared error component in the previous time and the conditional variance component in the previous time. Thus, the general form of the GARCH(p,q) model is. According to Tsay, (2010) the GARCH model is used to overcome orders that are too large in the ARCH model. The general form of the GARCH model (p,q) :

$$\sigma_t^2 = \alpha_0 + \alpha_1 e_{t-1}^2 + \alpha_1 e_{t-2}^2 + \dots + \alpha_p e_{t-p}^2 + \beta_1 \sigma_{t-1}^2 + \dots + \beta_q \sigma_{t-q}^2 \tag{2}$$

writable:

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^p \alpha_i e_{t-i}^2 + \sum_{j=1}^q \beta_j \sigma_{t-i}^2$$

2.3. GARCH effect test

In testing the presence or absence of the GARCH effect, the Lagrange Multiplier (LM test) was carried out.

2.4. Estimation of Model Parameters

Estimated parameters in the equation of the mean and conditional variance, by estimating the parameters α_0, α_i and β_j , with the regression model :

$$y_t = \gamma_0 + \gamma_1 x_i + e_t \quad t = 1, \dots, T \tag{3}$$

$$X_t = Z_t \sqrt{h_t} \tag{4}$$

$$h_t = \alpha_0 + \alpha_i X_{t-1}^2 + \beta_j h_{t-1} \tag{5}$$

With parameter simplification

$$\tilde{\theta} (\gamma_0, \gamma_1, \alpha_0, \alpha_i, \beta_j)' = (\tilde{\gamma}', \tilde{\delta}')$$

Vector of parameters $\tilde{\theta}$

$$\tilde{\delta} = \begin{bmatrix} \alpha_0 \\ \alpha_i \\ \beta_j \end{bmatrix} \text{ dan } \tilde{\gamma} = \begin{bmatrix} \gamma_0 \\ \gamma_1 \end{bmatrix}$$

2.5. Best Model Selection

The best model chosen is to look at the smallest AIC (Wei, 1990).

$$AIC = \log(\hat{\sigma}^2) + \frac{2l}{T}$$

2.6. Stationary Test

Ekananda (2015), one method to test for stationary is the unit root test. Unit root test is a term that indicates the eigenvalue of a data is one. To obtain an overview of the unit root test, it will be shown in the following AR(1) process:

$$Z_t = \omega Z_{t-1} + \varepsilon_t$$

The following is a stationary test hypothesis using the unit root test (Dickey Fuller Test): Hypothesis:

$H_0: \omega = 1$ (data has unit root/data is non-stationary)

$H_1: \omega < 1$ (data has no unit root/data is stationary)

Test statistics:

$$DF = \frac{\hat{\omega}}{SE(\hat{\omega})}$$

with

$$SE = \sqrt{\frac{s_d^2}{n}}$$

and

$$s_d^2 = \frac{1}{n} \sum_{i=1}^n (z_t - \bar{z})^2$$

2.7. Selection of the Best Model and Diagnostic Model

The selection of the best model is done by looking at the average residual (white noise), then it can be seen that the model obtained is feasible according to the influence of ARCH, to see the best model, it is seen from the smallest AIC value. For model diagnostics, look at the ACF and PACF and see the best ARIMA model. The best ARIMA model is taken and then predicted using the GARCH model.

III. RESULTS AND DISCUSSION

3.1. Data analysis

The data used in the calculation is data on the closing price of Indosat shares from 2012 to 2021.

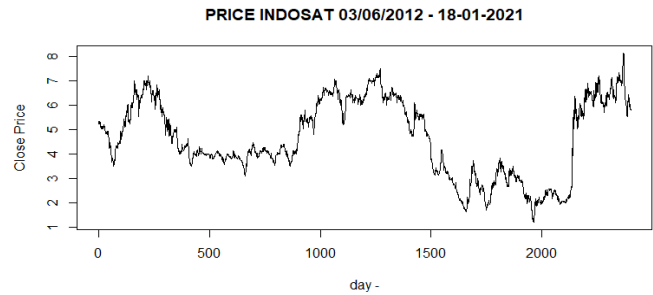


Fig. 1. Plot of Indosat Closing Stock Price Data 6 March 2012 – 18 January 2021

From the data plot, it can be seen that the closing price of Indosat shares fluctuates from time to time. Summary of Indosat stock statistics can be seen in the table below :

TABLE 1. Summary of Indosat Closing Stock Data Statistics

Statistics	Value
Mean	4.626
Median	4.350
Maximum	8.150
Minimum	1.200
Standard Deviation	1.564

Based on table 1, it can be seen that the statistical summary value of the Indosat stock data is the mean 4.626, the median is 4.350 and the standard deviation is 1.564.

3.2. Data stationarity test

The stationarity test of the data used is the Augmented Dickey Fuller Test to determine the average stationarity.

TABLE 2. Augmented Dickey Fuller test on the first differencing for train data stock prices

Lag	ADF test	p. value
12	-11.345	0.01

Based on table 2 above, the data is stationary in the first differencing and the $p - value$ is 0.01, because the $p - value < \alpha = 0.05$, then the data is stationary on the average.

3.3. Model Identification

The identification of the temporary model is done by looking at the ACF and PACF plots of data that have been differencing once.

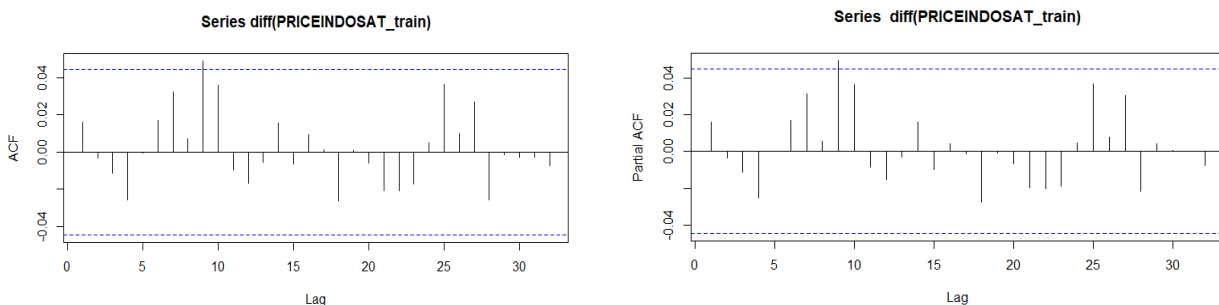


Fig. 2. Plots of ACF and PACF after differencing

Based on Figure 2 above, it can be seen that the ACF and PACF plots both have a cut off at lag-9, so the ARIMA models formed are ARIMA (0,1,9), ARIMA (9,1,0) and ARIMA (9,1,9).

3.4. Parameter Estimation and Selection of the Best Model

In the ACF and PACF plots above there is a cut off at lag-9 after the first differencing, the ARIMA models formed are ARIMA (0,1,9), ARIMA (9,1,0) and ARIMA (9,1,9). To select the best model, an assessment is carried out based on the lowest AIC value.

TABLE 3. AIC value of ARIMA

Model	AIC Value
ARIMA (0,1,9)	-3065.7527
ARIMA (9,1,0)	-2930.9463
ARIMA (9,1,9)	-3051.6845

Based on table 3 above, the lowest AIC value is found in the ARIMA model (0,1,9) with an AIC value of -3065.7527, then the ARIMA model parameter estimation (0,1,9) will be estimated.

TABLE 4. Estimation of ARIMA model parameters (0,1,9)

Parameter	Coefficient	z - value	Parameter significant
MA1	-0.9849	-42.7434	Significant
MA2	-0.0208	-0.6501	Significant
MA3	-0.0080	-0.2503	Significant
MA4	-0.0119	-0.3719	Significant
MA5	0.0293	0.8870	Not Significant
MA6	0.0176	0.5595	Significant
MA7	0.0122	0.3812	Significant
MA8	-0.0279	-0.8843	Significant
MA9	-0.0054	-0.2402	Significant

$$y_t = \mu + \phi_1(y_{t-1} - y_{t-2}) + \dots + \phi_9(y_{t-9})$$

$$y_t = \mu + (1 - 0.9849)y_{t-1} + (-0.9849 + 0.0208)y_{t-2} + (-0.0208 + 0.0080)y_{t-3} + (-0.0080 + 0.0119)y_{t-4} + (-0.0119 - 0.0293)y_{t-5} + (0.0293 - 0.0176)y_{t-6} + (0.0176 - 0.0122)y_{t-7} + (0.0122 + 0.0279)y_{t-8} + (-0.0279 + 0.0054)y_{t-9} + (-0.0054)y_{t-10}$$

$$y_t = \mu + 0.0151y_1 - 0.9641y_{t-2} - 0.0128y_{t-3} + 0.0039y_{t-4} - 0.0412y_{t-5} + 0.0116y_{t-6} + 0.0054y_{t-7} + 0.0401y_{t-8} - 0.0225y_{t-9} - 0.0054y_{t-10}$$

3.5. Model Diagnostics

For the diagnostic model, ARIMA (0,1,9) is used. In predicting stock prices, it must meet the stationary assumption, namely normality using the Shapiro-Wilk test and autocorrelation using the Box-Pierce test.

TABLE 5. Residual Normality Test Output

Shapiro – Wilk test	p - value
W	2.2 x10 ⁻¹⁶

Based on table 5 above, the p-value is 2.2x10⁻¹⁶ because the p-value 2.2x10⁻¹⁶ > a = 0.05, it is concluded that the residual follows a normal distribution. Next will be tested auto correlation.

3.6. GARCH effect test

ARCH/GARCH effect test, the test carried out is the ARCH-LM test.

TABLE 6. ARCH-LM . Test Output

ARCH-LM Test	p - value
Chi-Square	2.2 x10 ⁻¹⁶

Based on table 6 above, the Chi-Square test statistic obtained a p-value of 2.2 x10⁻¹⁶ which means that the p-value is smaller than the significant level value, it can be concluded that there is an ARCH effect on the ARIMA model (0,1,9).

3.7. Identify the ARCH/GARCH model

a. ACF and PACF Plots

The identification of the temporary model is done by looking at the ACF and PACF plots of data that have been differencing once.

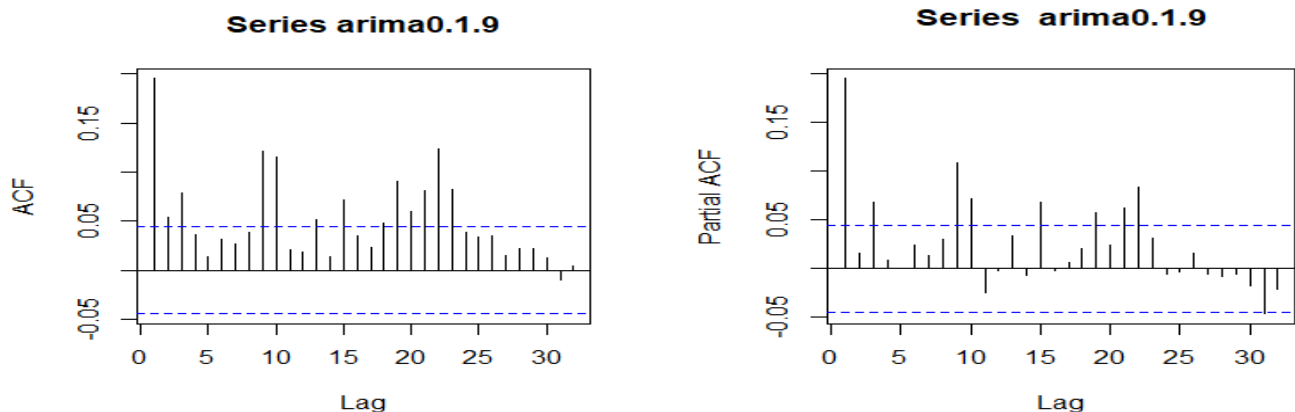


Fig. 3. Plots of ACF and PACF squared residual ARIMA (0,1,9)

Based on Figure 3 above, it can be seen that the ACF plot has cut offs at lag 1, lag-2 and lag-3 while the PACF plot has cut offs at lag-1 and lag-3, so the model that can be formed is GARCH (0,1), GARCH (1,0), GARCH (1,1), GARCH (2,1), GARCH (1,2), GARCH (3,1), GARCH (1,3), GARCH (2,3) and GARCH (3,2). To select the best model, an assessment is carried out based on the lowest AIC value.

3.8. Parameter Estimation and Selection of the Best Model

In the ACF and PACF plots above there is a cut off at lag-9 after the first differencing, the ARIMA model formed is GARCH (0,1), GARCH (1,0), GARCH (1,1), GARCH (2,1), GARCH (1,2), GARCH (3,1), GARCH (1,3), GARCH (2,3) and GARCH (3,2). To select the best model, an assessment is carried out based on the lowest AIC value.

TABLE 7. AIC value of GARCH

Model	AIC value
GARCH (0,1)	-0.57101
GARCH (1,0)	-0.64791
GARCH (1,1)	-0.83135
GARCH (1,2)	-0.83448
GARCH (2,1)	-1.6241
GARCH (1,3)	-0.83411
GARCH (3,1)	-0.82968
GARCH (3,2)	-0.83281
GARCH (2,3)	-0.83334

Based on table 7 above, the lowest AIC value is found in the GARCH model (1,2) with an AIC value of -0.83448, then the best model that can be selected is the GARCH model (1,2) with parameter estimates of :

$$\sigma^2 = 4.1922 + 0.1748\sigma_{t-1}^2 + 0.5077\sigma_{t-1}^2 + 0.3000\sigma_{t-2}^2$$

3.9. ARC/GARCH model verification

For the diagnostic model, GARCH (1,2) is used, stock price predictions must meet the stationary assumption, namely normality using the Shapiro-Wilk test and autocorrelation using the Box-Pierce test.

TABLE 8. Residual Normality Test Output

Shapiro – Wilk test		p – value
W	0.9347	2.2 x10 ⁻¹⁶

Based on table 8 above, the p-value is 2.2x10⁻¹⁶ because the p-value (2.2x10⁻¹⁶)<a=0.05, it is concluded that the residual follows a normal distribution. Next will be tested auto correlation.

TABLE 9. Autocorrelation Test Output

Box-Pierce test.		p – value
Chi-Squared	0.0004	0.983

Based on table 9 above, the p-value obtained is 0.983 > a = 0.05, it is concluded that the residuals are not autocorrelated, so that the autocorrelation assumption is fulfilled.

3.10. Forecasting

The best model obtained by GARCH (1,2) can be used to forecast Indosat stock for the next 10 days as shown in table 11.

TABLE 10. Indosat stock predictions for the next 10 days

No	Forecast	Lo 80	Hi 80	Lo 95	Hi 95	RMSE
1	5.833584	5.613735	6.053432	5.497354	6.169813	0.1697524
2	5.788989	5.482261	6.095717	5.319889	6.258089	
3	5.847412	5.470061	6.224762	5.270304	6.424519	
4	5.843566	5.403958	6.283174	5.171244	6.515889	
5	5.807581	5.320514	6.294647	5.062677	6.552485	
6	5.817853	5.262776	6.372931	4.968936	6.666771	
7	5.815417	5.202202	6.428631	4.877586	6.753247	
8	5.818077	5.155259	6.480896	4.804384	6.831771	
9	5.822140	5.113186	6.531095	4.737888	6.906393	
10	5.822140	5.072533	6.571748	4.675715	6.968566	

Based on table 10 above, it is found that the forecasting values of Indosat's shares for the next 10 days are not much different from the original data with an RMSE prediction value of 0.1697524.

IV. CONCLUSION

Indosat stock forecasting using the ARIMA and GARCH methods, the smallest AIC value is ARIMA (0,1,9) with an average mean and GARCH (1,2) with model variances are:

$$y_t = \mu + 0.0151y_1 - 0.9641y_{t-2} - 0.0128y_{t-3} + 0.0039y_{t-4} - 0.0412y_{t-5} + 0.0116y_{t-6} + 0.0054y_{t-7} + 0.0401y_{t-8} - 0.0225y_{t-9} - 0.0054y_{t-10}$$

And the variance of the GARCH(1,2) model is

$$\sigma^2 = 4.1922 + 0.1748\sigma_{t-1}^2 + 0.5077\sigma_{t-1}^2 + 0.3000\sigma_{t-2}^2$$

With an RMSE value of 0.1697524, it can be concluded that the GARCH (1,2) model is good enough to predict Indosat's stock price.

REFERENCES

- Darmadji, T. and Fakhruddin, H, "Capital Markets in Indonesia" Jakarta: Alfabeta Publishers. 2012
- Hasanah, P., Nasir, S.Q., & Subchan, S. "Gold return volatility modeling using GARCH". Indonesian Journal of Mathematics Education, 2(1), 20-26. 2019.
- Bollerslev, T. "Generalized autoregressive heteroscedasticity". Journal of Econometrics 52, 307-327. 1986.
- Hermuningsih, S. "The effect of profitability, size on firm value with capital structure as an intervening variable". Journal of business strategy, 16 (vol. 2). , 2012.
- Cryer, J.D., and K.-S. Chan. "Time Series Analysis with Applications in R" Second. Low City: Springer, 2008.
- A. Pani Desvina and Khairunisa. "Application of the Deep Arch / Garch Method Forecasting Exchange Rate Transactions (Exchange Rate) Selling Indonesian Currency (IDR) Against European Currency (GBP)," J. Sains Mat. and Stat., vol. 4, no. 2, p. 114– 123 2018.
- A. P. Raneo and F. Muthia, "Application of the GARCH Model in Volatility Prediction on the Indonesia Stock Exchange," J. Manaj. Dan Bisnis Sriwij., vol. 16, no. 3, p. 194–202. 2019.
- Ariefianto and M. Doddy. "Essence of Econometrics and Applications Using eViews". Jakarta: Erlangga. 2012.