Application of ARCH model on nutmeg price forecasting in South Aceh district

A Rusyana¹, R Ferdiana¹, M E Putri¹,*
Department of Statistics, Faculty of Mathematics and Natural Sciences, Syiah Kuala University, Banda Aceh 23111, Indonesia

E-mail: meiditaekaputri71@gmail.com

Abstract. The price of nutmeg which is changing and unstable in South Aceh results to volatility. Volatility is a variance pattern of time series, especially the time series of finances caused by variance not constant. This results the possibility of heteroscedastic data, so that it needs to be made a model of a particular approach to measure residual volatility problems. The data used in this research is the price of nutmeg in South Aceh district from January 2012 to December 2016 which has volatility. ARCH (Autoregressive Conditionals Heteroscedastic) is a model used to resolve the residuals variance which is not constant in financial time series. Then this model was developed into a Generalized Autoregressive Conditional Heteroskedastic (GARCH) to avoid too high orders on ARCH models and make the variance is always positive. The purpose of this research is to get the best ARCH/GARCH model for nutmeg price in South Aceh district and get the nutmeg price forecast for January 2016 until December 2017. The results of this research show that price of nutmeg in South Aceh district is stationary to the mean after the second differencing and the result of the ARCH Lagrange Multiplier test shows the problem in the 4 lag so the data are modelled with ARCH. The best models for forecasting price of nutmeg in South Aceh district is using model of ARIMA (2,2,0)-ARCH (4). While the results of the verification of the model are obtained the value of the MAPE of 6.12 percent.

1. Introduction
In the financial sector generally has a very high level of volatility. The high Volatility is shown by a phase where relatively high fluctuation and then followed the low fluctuation and high returns, with the other words, the mean and the variance are not constant. The presence of high volatility causes difficulty of the estimation and prediction of the movement of these variables. The model which assumes that the residual variance is not constant in time series data was developed by Engle (1982) in autoregressive model called conditionals heteroskedastic (ARCH).

Nutmeg is one of export commodities originating from South Aceh. According to BPS Aceh in year 2016, Aceh has nutmeg production of 6,620 Tonnes where the South Aceh Regency has the highest i.e. 5,747 tonnes. Changes of rise and fall of the price of nutmeg in South Aceh Regency cause occurrence of volatility. Since model of the ARCH was published by Engel in 1982, the model has been widely used to describe the behavior of a time series data volatility. Therefore, this research was

*Corresponding author: meiditaekaputri71@gmail.com
conducted to obtain the best ARCH model in forecasting the price of nutmeg in South Aceh district from January 2016 until December 2017.

2. Literature Review

2.1 Stationary Data
There are two types of stationary, namely stationary in the mean and stationary in variance. Stationary in the mean is data fluctuations around the mean value of a constant. Augmented Dickey Fuller (ADF) test is usually used to test stationery in mean [1]. Differencing is used to convert a time series data of nonstationary becomes stationary in the mean [2].

A time series data are said to be stationary in the variance if structure of the data have constant data fluctuations from time to time [3]. If the data are not stationary in the variance then the data transformation is used to stabilize or get a constant variance. This transformation is called the Box-Cox transformation [4].

2.2 Model of Autoregressive Integrated Moving Average (ARIMA)
A model with stationary data through differencing process is called ARIMA. If the data are stationary on the process of differencing d times and apply the ARMA (p,q) then these models are called ARIMA (p, d, q). The form of the equation for model of ARIMA is [3]:

\[
\phi_p(B)(1-B)^dZ_t = \theta_0 + \theta_q(B)a_t
\]  

2.3 Identification model of ARIMA
Model of ARIMA can be identified from pattern of the ACF and PACF plot if data is already stationary in the mean and variance, see table 1.

| Proses         | Autocorrelation Function (ACF)                        | Partial Autocorrelation Function (PACF) |
|----------------|------------------------------------------------------|----------------------------------------|
| AR(p)          | Exponentially down trend                             | Cut off after lag p                     |
| MA(q)          | Cut off after lag q                                  | Exponentially down trend               |
| ARMA(p,q)      | Cut off after lag (p-q)                              | Cut off after lag (p-q)                |

2.4 Model of Autoregressive Conditional Heterokedastic (ARCH) and the effects test of ARCH
Volatility is a measure of the uncertainty of time series data indicated by the presence of fluctuations. These fluctuations cause the residual variance not constant and heteroskedastic. ARCH is a model that can be used to overcome the residual variance not constant in financial time series data. Where the equation of the model can be seen as follows [5]:

\[
s_t^2 = \sigma_0 + \alpha_1s_{t-1}^2 + \alpha_2s_{t-2}^2 + \ldots + \alpha_q s_{t-q}^2
\]  

where:

- \(s_t^2\) : variance of the residual at time \(t\)
- \(\sigma_0\) : component constant
- \(\alpha_q\) : parameter of ARCH with order \(q\)
- \(s_{t-q}^2\) : residual squared at time \(t-q\)

LM ARCH test is used to find out the heteroscedasticity element in the data time series with hypothesis [5]:

- \(H_0 : \alpha_1=\alpha_2=\ldots=\alpha_q=0\) (There is no effect of heteroscedasticity)
- \(H_1 : \alpha_i \neq 0, i=1,2,\ldots,q\) (There is an effect of heteroscedasticity)
2.5 Akaike Information Criterion (AIC)
The selection of the best model is made based on the value of AIC of the smallest [6].

\[ \text{AIC} = \ln \hat{\sigma}^2 + \frac{2}{n} r \]  

where:
- \( n \) = number of error
- \( r \) = the number of parameters in the model of ARIMA
- \( \hat{\sigma}^2 \) = estimator of error variance

2.6 Mean Absolute Percentage Error (MAPE)
MAPE is a measurement of error calculated by formulae:

\[ \text{MAPE} = \frac{1}{n} \sum_{t=1}^{n} \left| \frac{Z_t - \hat{Z}_t}{Z_t} \right| \times 100 \]  

where:
- \( Z_t \) = actual value at time period \( t \)
- \( \hat{Z}_t \) = forecast value for time period \( t \)
- \( n \) = number of observations.

3. Method
This study uses data of price of nutmeg in a South Aceh District from 2012 until 2016 obtained from the Department of trade, industry, cooperatives and UKM in South Aceh District. The number of data used is 60 from January 2012 to December 2016. In this study, data are divided into two parts: the data from January 2012 to December 2015 are as training data while the data from January until December 2016 are data for testing.

4. Results and discussion

4.1 Time Series Plot
Based on figure 1, the highest nutmeg prices in South Aceh district occurred in August to September 2013 Rp 62,000 per Kg. Then the nutmeg price changes for the following months decreased monthly to reach the lowest price in June and July of 2015 Rp 35,000 per Kg.

![Figure 1. Plot of nutmeg price data in South Aceh District from January 2012 until December 2015.](image)

4.2 Stationarity test
The result of the Augmented Dickey Fuller (ADF) test gives the value of the P-value of 0.561. This shows that its decision rejects \( H_0 \) with \( \alpha = 5\% \) which it means data are not stationary, so the data need for differencing. After the first differencing is done then got P-value 0.549 from ADF test. This indicates that the data is still not stationary in mean, so data need to be the second differencing. On the second differencing, P-value is 0.015, where the result indicates that data have been stationary in mean, because P-value < \( \alpha \).
Box-Cox test acquired value of $\lambda=1.999924$. This shows that the data have been stationary in variance and do not require transformation because the value of $\lambda$ is obtained already past 1.

### 4.3 Identification of ARIMA Model

Model identification uses ACF and PACF plots before and after transformation in figure 2 and figure 3.

![Figure 2](image_url) Plot of ACF and PACF before differencing level 2.

![Figure 3](image_url) Plot of ACF and PACF after differencing level 2.

Figure 2 shows the plot of ACF Downtrend exponentially and plot the PACF cut off after the first lag. While the Fig. 3 showed that ACF and PACF plots have cut off and can be seen there are still some lag line. Based on figure 2 and 3, Five models in table 2 are selected.

![Table 2](image_url) Tentative model of ARIMA.

| Order | Parameter Model | AIC |
|-------|-----------------|-----|
| ARIMA | AR(1) | AR(2) | AR(3) | MA(1) | MA(2) | MA(3) | |
| (1,2,0) | -0.369** | - | - | - | - | - | 853,796 |
| (2,2,0) | -0.549** | -0.488** | - | - | - | - | **843,159** |
| (3,2,0) | -0.632** | -0.576** | -0.158 | - | - | - | 843,972 |
| (2,2,3) | 0.133* | -0.947** | - | -0.884** | 1,170** | -0.599** | **845,191** |
| (3,2,2) | 0.394 | -0.088 | 0.333** | -0.119** | 0.119 | - | 845,700 |

Significant at the $\alpha=10\%$ (**)

Based on Table 2, there are three significant coefficient of the whole model, namely, model of ARIMA(1,2,0), ARIMA(2,2,0), and ARIMA(2,2,3). To get the best of the three models then the selection is based by the smallest AIC value. The model of ARIMA (2,2,0) has the smallest AIC value.

### 4.4 Testing of ARCH Effect

Effect of ARCH test is used to check that there is a problem of heteroscedasticity of ARIMA (2,2,0) model residual. From testing of ARCH-LM to the residual of the model of ARIMA (2,2,0), p-value is obtained 0.00121 on lag 4. Then the decision is rejection of $H_0$ with $\alpha = 0.05$ so there is effect of ARCH or there is heteroscedasticity problem. The existence of the element of the ARCH can also be seen from the plot of ACF and PACF residual square.

ACF and PACF graph in Figure 4 shows that there is a lag that still over of barlet line, it also shows that the nutmeg data price in South Aceh district has problem of heteroscedasticity, so ARCH model is good to be applied.
4.5 Identification of ARCH Model

Testing of ARCH-LM previously can be inferred, there is the effect of the forth lag in ARCH so that the model of the ARCH is suitable for forecasting nutmeg price in South Aceh District. Then the model can also be identified with ACF and PACF residual square shown in Fig. 4, where there are some early lag over the line of bartlett.

Table 3. Tentative model of ARCH.

| Order ARCH | Parameter Model | AIC  |
|------------|-----------------|------|
|            | $\alpha_0$     | $\alpha_1$ | $\alpha_2$ | $\alpha_3$ | $\alpha_4$ |      |
| ARCH(1)    | 4.297 x10^6**  | 4.222 x10^-1 | -            | -            | -            | 856.8146 |
| ARCH(2)    | 4.071 x10^6**  | 2.434 x10^-1 | 1.144 x10^-1 | -            | -            | 839.8755 |
| ARCH(3)    | 3.845 x10^6**  | 2.044 x10^-1 | 5.091 x10^-9 | 7.729 x10^-2 | -            | 824.1773 |
| ARCH(4)    | 3.619 x10^6**  | 1.480 x10^-1 | 4.307 x10^-2 | 6.197 x10^-2 | 9.601 x10^-8 | **808,9155** |

Keterangan: Signifikan pada $\alpha=5\%$ (**)

Table 3 shows that the model of the ARCH (4) has the smallest AIC if it is compared three of the other models. So the ARCH (4) model is suitable to be used in forecasting the nutmeg price in South Aceh district to 24 period. The ARCH (4) model is

$$\sigma_t^2 = 3.7557 \times 10^5 + 1.6440 \times 10^{-1}(\varepsilon_{t-1}^2) + 3.3014 \times 10^{-2}(\varepsilon_{t-2}^2) + 2.9528 \times 10^{-2}(\varepsilon_{t-3}^2) + 1.1344 \times 10^{-8}(\varepsilon_{t-4}^2) \quad (5)$$

4.6 Forecasting of ARIMA-ARCH

Forecasting is done with combining of the ARIMA and ARCH model, they are ARIMA (2,2,0)-ARCH (4) which can be written as:

$$Z_t = 0.96313Z_{t-1} - 1.07374Z_{t-2} + 1.03687Z_{t-3} + 3.7557 \times 10^6 + 1.6440 \times 10^{-1}(\varepsilon_{t-1}^2) + 3.3014 \times 10^{-2}(\varepsilon_{t-2}^2) + 2.9528 \times 10^{-2}(\varepsilon_{t-3}^2) + 1.1344 \times 10^{-8}(\varepsilon_{t-4}^2) \quad (6)$$

After getting the best models of ARIMA(2,2,0) - ARCH(4), forecasting is conducted where forecasting is divided two, namely training and testing forecasting. Training forecasting is the nutmeg price from January until December 2016, while testing forecasting is from January until December 2017. The result of the training and testing forecasting can be seen in figure 4.
The value of the MAPE obtained from actual and forecasting data from January to December 2016 is 6.12 percent. It shows the average of forecasting error rate is 6.12 percent. Based on the MAPE, the forecasting with ARIMA(2,2,0)-ARCH (4) is categorized well to forecast the price of nutmeg in South Aceh Regency.

5. Conclusions
1. The best model obtained to predict the price of nutmeg in South Aceh Regency is a ARIMA (2, 2, 0)-ARCH (4) which can be written as.
   \[
   Z_t = 0.963Z_{t-1} - 1.074Z_{t-2} + 1.037Z_{t-3} + 3.756 \times 10^6 + 1.644 \times 10^{-1}(\varepsilon_{t-1}^2)
   
   3.301 \times 10^{-2}(\varepsilon_{t-2}^2) + 2.953 \times 10^{-2}(\varepsilon_{t-3}^2) + 1.134 \times 10^{-4}(\varepsilon_{t-4}^2)
   \]
2. The result of nutmeg price forecasting in South Aceh district in January 2016 until December 2017 tends to increase each month about Rp 976,- and average error level is relatively low 6.12%.

6. References
[1] Gujarati and Porter 2009 Dasar-Dasar Ekonometrika (Jakarta: Salemba Empat)
[2] Rusyana A, Nurhasanah, Marzuki and Flancia M 2016 SARIMA Model for Forecasting Foreign Tourists at the Kualanamu International Airport Proc. Int. Conf. Mathematics, Statistics, and their Applications (ICMSA) 155
[3] Wei W W S 2006 Time Series Analysis Univariate and Multivariate Methods 2nd ed (USA: Addision-Wesley Publishing Company)
[4] Montgomery D C, Jennings C L, and Kulahci M 2008 Introduction to Time Series Analysis and Forecasting (New Jersey: John Wiley & Sons)
[5] Juanda B and Junaidi 2012 Ekonometrika Deret Waktu (Bogor: IPB Press)
[6] Brooks C 2008 Introductory Econometrics for Finance. 2nd ed (New York: Cambridge University Press)