Forecasting East Belitung Regency Rainfall Data by Reviewing Heteroscedasticity

E Kustiawan¹* and Adriyansyah²
¹Department of Mathematics, Faculty of Engineering, Universitas Bangka Belitung, Indonesia
²Department of Mathematics, Faculty of Engineering, Universitas Bangka Belitung, Indonesia

*Email : elyaskustiawan@gmail.com

Abstract. East Belitung Regency is one of the regencies located on Belitung Island. East Belitung Regency has a tropical and wet climate with a fairly high variation of rainfall. Rainfall forecasting is an important thing to model because of the many uses of rainfall forecasting results such as irrigation planning, flood prediction, erosion prediction and others. This study aims to predict rainfall for the next 5 years by using a time series model by reviewing the heteroscedasticity of the data. From the results of the analysis of rainfall in East Belitung Regency with a seasonal pattern. The best model used is ARIMA (0,1,1)(2,1,1)₁₂ with insignificant heteroscedasticity.

1. Introduction

East Belitung Regency is one of the regencies located on Belitung Island which has now been divided into Belitung Regency and East Belitung Regency. East Belitung Regency has a tropical and wet climate with variations in monthly rainfall in 2019 between 1.4 mm to 531.1 mm with the number of rainy days 2 to 28 days per month [1].

Rainfall forecasting is an important thing to model because of the many uses that can be obtained from this information, namely for irrigation planning, prediction of floods, storms, erosion, and others [2,3,4]. The method that is often used to predict rainfall is time series models such as Autoregressive Integrated Moving Average (ARIMA) and Seasonal Autoregressive Moving Average (SARIMA). The ARIMA model is effective for considering serial linear correlations between observations while SARIMA has better accuracy in predicting seasonal patterns [5,6]. SARIMA has a deficiency in the optimization to get the best parameters, so it is often combined with heuristic methods to improve the performance the model [7,8].

Rainfall in East Belitung Regency is fluctuating and has the potential to cause non-stationary variation, so the problem of heteroscedasticity must be reviewed in more detail. Rainfall data is influenced by volatility where large changes tend to follow large changes and small changes tend to follow small changes [9]. An important characteristic of rainfall data is that it has a very skewed distribution or kurtosis [10]. Performing seasonal differencing to the SARIMA model does not eliminate heteroscedasticity from the residuals, so it necessary to apply ARCH/GARCH [11, 12]. To eliminate heteroscedasticity, the ARCH/GARCH works by correcting the variance in the time series model [13, 14, 15].
Data that has high volatility, in these cases the least squares does not meet so that the time series models that can be used are the Autoregressive Conditional Heteroscedasticity (ARCH) and Generalized Autoregressive Conditional Heteroscedasticity (GARCH) models.

The purpose of this study is to apply the ARCH/GARCH to predict rainfall for the next 5 years (2021-2025) by reviewing heteroscedasticity problems in rainfall data in East Belitung Regency. The ARCH/GARCH is expected to be the best model in predicting rainfall data in East Belitung Regency.

2. Methods
2.1 ARIMA models
The general form of the ARIMA (p,d,q) is defined as

\[ \phi_p(B)(1-B)^d y_t = \theta_q(B)a_t \]  

\[ \phi_p(B) = 1 - \phi_1 B - \ldots - \phi_p B^p \]  

\[ \theta_q(B) = 1 - \theta_1 B - \ldots - \theta_q B^q \]  

where \((1-B)^d\) is d-order differencing and \(a_t\) is residual value at time \(t\).

2.2 SARIMA Models
The general form of the SARIMA (p,d,q) is defined as

\[ \phi_p(B)\Phi_p(B')(1-B)^d (1-B')^D y_t = \theta_q(B)\Theta_q(B')a_t \]  

where \(\phi_p(B)\) is non seasonal autoregressive, \(\Phi_p(B')\) is seasonal autoregressive, \((1-B)^d\) is non seasonal differencing, \((1-B')^D\) is seasonal differencing \(\theta_q(B)\) is non seasonal moving average, \(\Theta_q(B')\) is seasonal moving average.

2.3 ARCH Models
The general form of the ARCH is defined as

\[ \sigma_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \ldots + \alpha_p \varepsilon_{t-p}^2 \]  

2.4 GARCH Models
The general form of the GARCH is defined as

\[ \sigma_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \ldots + \alpha_p \varepsilon_{t-p}^2 + \beta_1 \sigma_{t-1}^2 + \ldots + \beta_q \sigma_{t-q}^2 \]  

3. Result
This simulation uses rainfall data for East Belitung Regency is used for 11 years from 2010 to 2020. East Belitung Regency rainfall data are shown in Table 1.

| Year | Jan | Feb | Mar | Apr | May | Jun | Jul | August | Sep | Oct | Nov | Dec |
|------|-----|-----|-----|-----|-----|-----|-----|--------|-----|-----|-----|-----|
| 2010 | 260.8 | 82.4 | 196.8 | 224.9 | 400.6 | 384.6 | 422.5 | 364.9 | 461 | 278.6 | 411.9 | 382.6 |
| 2011 | 263.9 | 141.2 | 233.6 | 207.9 | 285.8 | 125 | 92.8 | 0 | 26.9 | 254.6 | 502 | 498.3 |
| 2012 | 188 | 281.2 | 126.7 | 374.5 | 144.6 | 141.2 | 30.5 | 34.8 | 71.6 | 321.8 | 406.9 | 344.7 |
| 2013 | 210 | 241 | 122.4 | 345.2 | 496.4 | 192 | 279 | 152 | 59 | 355 | 444 | 732 |
| 2014 | 184 | 0 | 174 | 464 | 523 | 265 | 65 | 79 | 25 | 62 | 31.5 | 334 |
| 2015 | 263 | 257 | 166 | 453 | 253 | 93 | 17 | 0 | 0 | 116.1 | 454 | 577.3 |
| 2016 | 478.8 | 506 | 188.4 | 435.1 | 249 | 174.7 | 223.9 | 265.1 | 283.4 | 268.8 | 332 | 347.4 |
| 2017 | 440.7 | 247.3 | 271.1 | 383.6 | 284.5 | 215.7 | 540.4 | 105.9 | 111.2 | 407.4 | 218.3 | 395.7 |
| 2018 | 112.5 | 39.6 | 325.3 | 407.7 | 384.1 | 249 | 40.9 | 61.5 | 81.8 | 470.9 | 416.6 | 520.5 |
| 2019 | 466.3 | 421.7 | 54.6 | 531.1 | 351.4 | 254 | 31.4 | 1.4 | 9.5 | 182 | 360.7 | 408.8 |
| 2020 | 303.8 | 378.3 | 400.1 | 340.5 | 218.5 | 293.8 | 243 | 82.5 | 298.1 | 344.6 | 293.3 | 249.6 |
| Total | 1848.5 | 1508.8 | 1207.9 | 2504.6 | 2352.4 | 1375.5 | 1130.7 | 895.8 | 926.9 | 1656.9 | 2582.7 | 3216.3 |
| Average | 264.1 | 215.5 | 172.6 | 357.8 | 336.1 | 196.5 | 161.5 | 128 | 132.4 | 236.7 | 369 | 459.5 |
| Max | 459.5 | December | | | | | | | | | |
| Min | 128 | August |

Average per year = 252.5
Figure 1 shows a graph of East Belitung Regency rainfall data. The data will be tested stationary or not. The stationary of the data will be tested by root test. The root test used to determine the stationary of the data is using Augmented Dickey Fuller (ADF). In this case study, a significance level of 5% is used by testing $H_0$: data is not stationary and $H_1$: data is stationary. If $|ADF| < t$-statistic then fail to reject $H_0$ and if $|ADF| > t$-statistic then reject $H_0$. The following are the results of the root test using ADF.

| ADF       | t-Statistic |
|-----------|-------------|
| Critical value 1% | -3.480818   |
| Critical value 5%  | -2.883579   |
| Critical value 10% | -2.578601   |

The graph above shows seasonal pattern data. The step to eliminate the seasonal pattern is carried out by seasonal and non-seasonal differencing. From the results of seasonal and non-seasonal differencing then the differencing is combined. The following are ACF and PACF which have been combined with seasonal and non-seasonal differencing.

| Autocorrelation | Partial Correlation | AC | PAC | Q-Stat | Prob |
|-----------------|---------------------|----|-----|--------|------|
| 1               | 1                   | 0.339 | 0.339 | 14.141 | 0.000 |
| 2               | 0.145               | 0.634 | 15.757 | 0.000 |
| 3               | 0.033               | 0.653 | 17.558 | 0.001 |
| 4               | 0.053               | -0.657 | 17.256 | 0.002 |
| 5               | 0.048               | -0.602 | 17.556 | 0.004 |
| 6               | 0.045               | -0.019 | 17.015 | 0.007 |
| 7               | 0.043               | 0.025 | 18.054 | 0.012 |
| 8               | 0.051               | 0.632 | 18.398 | 0.004 |
| 9               | 0.025               | -0.016 | 18.481 | 0.030 |
| 10              | 0.090               | 0.652 | 19.113 | 0.039 |
| 11              | 1.070               | -0.227 | 19.774 | 0.048 |
| 12              | -0.359               | -0.305 | 37.196 | 0.000 |
| 13              | -0.122               | 0.140 | 39.239 | 0.000 |
| 14              | -0.096               | -0.096 | 40.555 | 0.000 |
| 15              | -0.101               | -0.117 | 41.955 | 0.000 |
| 16              | 0.117               | -0.119 | 43.002 | 0.000 |
| 17              | -0.040               | 0.036 | 44.227 | 0.000 |
| 18              | 0.054               | 0.671 | 44.544 | 0.000 |
| 19              | 0.111               | 0.644 | 44.426 | 0.000 |
| 20              | 0.090               | 0.609 | 45.958 | 0.001 |
| 21              | 0.043               | -0.127 | 47.235 | 0.001 |
| 22              | -0.196               | -0.137 | 53.071 | 0.000 |
| 23              | -0.141               | -0.250 | 59.076 | 0.000 |
| 24              | -0.187               | -0.269 | 91.425 | 0.000 |

Figure 2. ACF and PACF
By looking at the ACF and PACF, the best possible model for predicting rainfall can be estimated. The best model used is the one with the smallest Akaike Info Criterion (AIC) value. The best model used for forecasting data with the smallest AIC value is ARIMA (0,1,1) (2,1,1)$_{12}$. After getting the best model, it is checked whether the data has heteroscedasticity behavior. If the data has an indication of heteroscedasticity, it needs to be modeled again using the ARCH/GARCH model. To see if the data has heteroscedasticity, it can be seen the squared residual correlogram pattern and the ARCH-LM test.

Table 3. ARCH Test

| Variable | Coefficient | Standard Error | Z.Statistic | Probability |
|----------|-------------|----------------|-------------|-------------|
| C        | 5.451042    | 5.593555       | 0.974522    | 0.3298      |
| AR(24)   | -0.339492   | 0.111029       | -3.057704   | 0.0022      |
| MA(1)    | 0.346892    | 0.115078       | 3.014423    | 0.0026      |
| SMA(12)  | -0.876968   | 0.039048       | -22.45882   | 0.0000      |
| C        | 12895.86    | 2619.069       | 4.923832    | 0.0000      |
| RESID(-1)$^2$ | 0.099148 | 0.1921         | 0.516124    | 0.6058      |

The table above shows that resid(-1)$^2$ is not significant because the probability value is greater than 5%. The data was tested again using ARCH-LM method by looking at the correlogram of the squared residue to see the heteroscedasticity. Figure 3 shows the opportunity value is smaller than 0.05, namely at lags 1, 2, 3, and 4.

Figure 3. Correlogram Squared Residuals

In Table 5 shows that the heteroscedasticity test at lags 1, 2, 3, and 4 has a probability value > 0.05 so that it means the data does not have heteroscedasticity. Because the data does not have heteroscedasticity, the best model for predicting rainfall data in East Belitung Regency is ARIMA(0,1,1) (2,1,1)$_{12}$. Forecasting results are shown in Table 6.
Table 4. ARCH-LM Test at lags 1, 2, 3, and 4.

| Lag   | F-statistic | Obs*R-squared | Prob. F(1,93) | Prob. Chi-Square |
|-------|-------------|---------------|---------------|------------------|
| Lag 1 | 1.234901    | 1.244928      | 0.2693        | 0.2645           |
| Lag 2 | 0.770741    | 1.565777      | 0.4657        | 0.4571           |
| Lag 3 | 0.755933    | 2.310841      | 0.5218        | 0.5104           |
| Lag 4 | 1.12673     | 4.531204      | 0.3492        | 0.3389           |

Table 5. Prediction results of East Belitung Regency rainfall data

| Month | Forecast | Month | Forecast | Month | Forecast | Month | Forecast |
|-------|----------|-------|----------|-------|----------|-------|----------|
| Jan-21| 284.20   | Jan-22| 289.96   | Jan-23| 295.10   | Jan-24| 300.12   |
| Feb-21| 278.93   | Feb-22| 283.64   | Feb-23| 288.93   | Feb-24| 294.29   |
| Mar-21| 258.79   | Mar-22| 264.40   | Mar-23| 269.33   | Mar-24| 274.40   |
| Apr-21| 406.38   | Apr-22| 410.99   | Apr-23| 416.08   | Apr-24| 421.48   |
| May-21| 324.55   | May-22| 330.29   | May-23| 335.02   | May-24| 340.06   |
| Jun-21| 250.23   | Jun-22| 255.40   | Jun-23| 260.29   | Jun-24| 265.51   |
| Jul-21| 195.26   | Jul-22| 200.71   | Jul-23| 205.42   | Jul-24| 210.55   |
| Aug-21| 124.44   | Aug-22| 129.53   | Aug-23| 134.50   | Aug-24| 139.74   |
| Sep-21| 176.54   | Sep-22| 181.60   | Sep-23| 186.48   | Sep-24| 191.73   |
| Oct-21| 332.47   | Oct-22| 337.44   | Oct-23| 342.75   | Oct-24| 348.03   |
| Nov-21| 374.48   | Nov-22| 380.05   | Nov-23| 385.75   | Nov-24| 390.83   |
| Dec-21| 429.30   | Dec-22| 435.08   | Dec-23| 440.29   | Dec-24| 445.31   |

4. Conclusion
Based on the results of the analysis, it was found that the best model used to predict rainfall data in East Belitung Regency was the ARIMA(0,1,1)(2,1,1)_{12} model.

References
[1] BPS 2020 Kabupaten Belitung Dalam Angka 2020 (Bangka: Badan Pusat Statistik Kabupaten Belitung)
[2] Nwokike C C, Offorha B C, Obubu M, Ugoala C B, and Ukomah H I 2020 Comparing SANN and SARIMA for forecasting frequency of monthly rainfall in Umuahia Scientific African 10 e00621, doi: 10.1088/1755-1315/926/1/012018

[3] Karthick S M and P Arumugam 2017 Stochastic modeling base monthly rainfall prediction using seasonal artificial neural networks. ICTACT J. Soft Comput 1 doi:10.21917/ijsc.2017.0196

[4] Nasseri M, Asghari K and Abedini M J 2008 Optimized scenario for rainfall forecasting using genetic algorithm coupled with artificial neural network Expert Systems with Applications 35 1415-1421 doi : 10.1016/j.eswa.2007.08.033

[5] Wang S, Feng J and Liu G 2011 Application of seasonal time series model in the precipitation forecast Mathematical and Computer Modelling 58 677 – 683

[6] Box G E P and Jenkins G M 1976 Times Series Analysis: Forecasting and Control (San Fransisco: Holden Day)

[7] Farsi M et al 2021 Parallel genetic algorithms for optimizing the SARIMA model for better forecasting of the NCDC weather data Alexandria Engineering Journal 60 1299-1316 doi: 10.1016/j.aej.2020.10.052

[8] Maarof M Z M, Ismail Z and Fadzli M 2014 Optimization of SARIMA model using genetic algorithm method in forecasting Singapore tourist arrivals to malaysia Appl. Math. Sci 8 8481-8491, doi: 10.12988/ams.2014.410847

[9] Laux P, Vogel S, Qiu W, Knoche H R and Kunstmann H 2011 Copula-based statistical refinement of precipitation in RCM simulation over complex terrain Hydrol. Earth Syst. Sci., 15 2401-2419 doi: 10.5194/hess-15-2401-2011

[10] Villariani G, Smith J A and Napolitano F 2010 Nonstationary Modeling of a long record of rainfall and temperature over rome Advances in Water Resources 33 1256-1267 doi: 10.1016/j.advwatres.2010.03.013

[11] Modarres R and Ouarda T B M J 2012 Generalized autoregressive conditional heteroscedasticity modeling of hydrologic time series Hydrol. Process 27 3174-3191 doi: 10.1002/hyp.9452.

[12] Sun H, Zhao N, Zeng X and Yan D 2015 Study of solar radiation prediction and modelling of relationships between solar radiation and meteorological variables Energy Conversion and Management 105 880-890 doi: 10.1016/j.enconman.2015.08.045

[13] Livina V, Ashkenazy Y, Kizner Z, Strygin V, Bunde A and Havlin S 2003 A stochastic model of river discharge fluctuations Physica A 330 pp.283-290 doi:10.1016/j.physa.2003.08.012

[14] Wang W, Van Gelder P H A, Vrijling J K and Ma J 2005 Testing and modeling autoregressive conditional heteroscedasticity of streamflow processes Nonlinear Processes in Geophysics 12 55-66 doi: 10.5194/npg-12-55-2005

[15] Otache M Y, Ahaneku I E, Mohammed A S and Musa J J 2012 Conditional heterocedasticity in streamflow process: Paradox or reality? Open Journal of Modern Hydrology 2 79.90, doi: 10.4236/ojmh.2012.24010.