Forecasting Model with Box-Jenkins Method to Predict the Number of Tourists Visiting in Toraja

Penulis	Info Artikel

[1]Didiharyono
Universitas Andi Djemma
Email: didi.unanda@gmail.com

[2]Bakhtiar
Universitas Andi Djemma
Email: bakhtiar647@gmail.com

ABSTRACT

This study aims to determine forecasting model with Box-Jenkins method and obtain results of data forecasting the number of tourists visiting in Toraja (Tanah Toraja and North Toraja regency) the future period. Research method used is applied research with quantitative data. Research procedures include identification of model, parameter estimation in model, verification and forecasting with using Minitab computer software. Based on the research obtained four models used in forecasting the number of tourists in Toraja the future period is ARIMA(1,1,1), ARIMA(2,1,1), ARIMA(1,2,1) and ARIMA(2,2,1). The correct criteria in selecting the best model is the model that has the smallest Mean Square (MS) value. In this case the time series model with the smallest MS value is ARIMA(2,2,1) that is 736062253. Thus, this model will used in forecasting is ARIMA(2,2,1) with equations $Z_t = 1,2167Z_{t-1} + 0,2605Z_t + 0,1711Z_{t-3} + 0,3061Z_{t-4} + a_t - 1,0323a_{t-1}$. The forecasting results for January to December 2021 is 149985, 193099, 207559, 202903, 222426, 229294, 239108, 250921, 260701, 271895, 283037 and 294221.

Keywords: Forecasting, Box-Jenkins Method, Tourists.

INTRODUCTION

Forecasting is one of the most important elements in decision-making in various aspects. Basically forecasting is done based on historical data that is analyzed using certain ways (Kulendran & Wong, 2005). Such historical data is collected, analyzed and studied to be linked over time. The forecasting model is very different for each issue depending on the various factors that influence it. Although, in the forecasting will not be obtained accuracy with value of one hundred percent. However, this does not necessarily mean that forecasting is not necessary, this forecasting can at least provide a picture of a situation or event that may occur in the future, so that it can assist the decision-making process and preparation for future improvement (Ho et al., 2002).

The state of the population is increasing from year to year, causing the need for more and more tourist from time to time. Everyone will need entertainment to let go of saturation because of busyness and work done every day. The high demand level of tourism needs to be observed by the government in maintaining, protecting and preserving the attractions in order to provide comfort for the visitors. Therefore, the visitors number who come to the tourist attractions depends on the comfort of place for visitors (Baldigara & Mamula, 2015).

Tourist attractions in Toraja which includes two districts of Tanah Toraja district and North Toraja district including Ma'nene Traditions, Rambu Solo ceremony, Kete Kesu, Londa, Batutumonga, Bori Parinding and Pohon Tarra, Ranteallo, Ne' Gandeng Museum, Pallawa, Sarambu Assing Waterfall, Lolai and others. The existence of these attractions is very potential in an effort to increase economic growth in the two districts and must get special attention in improving the condition of existing attractions in Toraja (Adams, 2006) (Kausar & Gunawan, 2018).

Based on data from the Tourism Office of South Sulawesi Province The visitation rate in Toraja continues to increase from time to time and the most visit in December because at that time there was a very spectacular event that was attended by tourists from various countries, the event is known as "lovely December" (Bakhtiar & Didiharyono, 2018).

Increasing the number of tourists both local tourists and foreign tourists who continue to increase from time to time not only as something just as beneficial area (Singh, 2013) (Petrevska, 2017). But also it can only be something that happens more often in the absence of services that are less satisfactory service and infrastructure will be a disappointment for foreign tourists, resulting in a decrease in the number of tourists visiting the Toraja region and this is very detrimental to Indonesia generally and Toraja especially as a tourist destination (Crystal, 2012).

One of method that can be used in predicting the number of tourists in future is forcasting model time series analysis with Box-Jenkins method or also known as ARIMA method (Baldigara & Mamula, 2015). This method can be used with studied the pattern of previous tourist data with regard to stationary and non stationary concept, autocovariance, autocorrelation function, partial autocorrelation function, backshiff operators and different operators. So time series analysis is used in this forecast because the pattern in the past will affect the pattern in the future (Chang & Liao, 2010).

The general model representing the non-stationary time series is Autoregressive Integrated Moving Average model or ARIMA (p,d,q) with p, d, q is the level for the autoregressive process, differencing and moving average. Wei (2006) ARIMA (p,d,q) model can be written in the general form,

$$\left(1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p\right)(1 - B)^d \tilde{z}_t = (1 - \theta_1 B - \cdots - \theta_q B^q)a_t$$

where, $\phi_1 \cdots \phi_p$: parameter of AR model, (1-B)^d : *differencing* level, Z_t : forecast value periode t, $\theta_1 \cdots \theta_q$: Parameter MA model, and a_t : residual value period t.

This method also uses iterative approach in identifying the best model. The advantage of this method in general is that it can analyze all time series data. The ARIMA model is said to be precise if the residual between the forecasting model with small historical data points, distributed randomly and independently of each other. Research with ARIMA model is also used by Sarpong (2013), Paul et al (2013), Biswas et al (2013), Steve (2014), Eni and Adeyeye (2015) which focuses its research on different cases. The use of ARIMA methods in predicting or forecasting data on the number of tourists coming to Toraja in visiting all tourism places in Toraja the future

period, so that local governments can prepare as early as possible good services and infrastructure necessary in preserving the attractions in providing comfort for the visitors.

Based on the above explanation, this study aims to determine the forecasting model with Box-Jenkins method and obtain results of data forecasting the number of tourists who visit tourism in Toraja the future period. Thus, the research is expected to give an idea to the local government about the number of tourists both local tourists and foreign tourists who visit the tourism in Toraja and can respond to the number of tourists by providing better service from year to year.

METHODS

The type of research used is applied research with quantitative data. The data used from the Central Statistics agency (BPS) of Tana Toraja and North Toraja for 60 periods from January 2012 to December 2016. Aswi and Sukarna (2008) the research procedure is

1. Identification of model includes (a). Perform data plots with graphs. (b). Stationary data checking with ACF and PACF,
2. Differences data if non-stationar happens at the data, and used ACF and PACF analysis for stationary data.
3. The parameter estimation model includes (a). Determine the model coefficients, (b). Estimation the model and (c). Temporary model selection.
4. Verification and forecasting phases include (a). Determination of the smallest Mean Square (MS) model, (b). Selection of the best model and (c) forecasting.

RESULTS AND DISCUSSION

Identification of Model

The number of tourists visiting Toraja for monthly data is presented from January 2012 to December 2016, where the total period is 60 (months). Based on the research data, the graph is as follows,
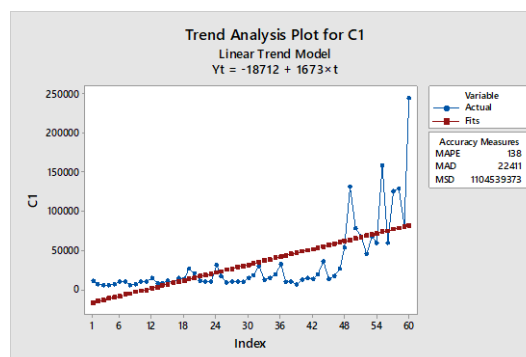

Figure 1. Trend graph a data

Based on the data plot and trend graph of data analysis in Figure 1 it can be seen that tourist visits have increased every time and the actual value is still far from linear lines and has a large variance, so this trend graph including time series that is not stationary in the average, though without seeing ACF and PACF.

In Figure 1 it shows 15 lags, in minitab if the number of lags is not requested it will automatically show the lag as much as the observation (n) ≤ 240, in this case the number of observations is 60

so 60/4 = 15 lag. In the identification of the original data is not yet formed model so that required first difference data to determine the stationary data.

Differences of Data

Figure 1 shows the data has not stationery then done the first differences so graph form is as follows
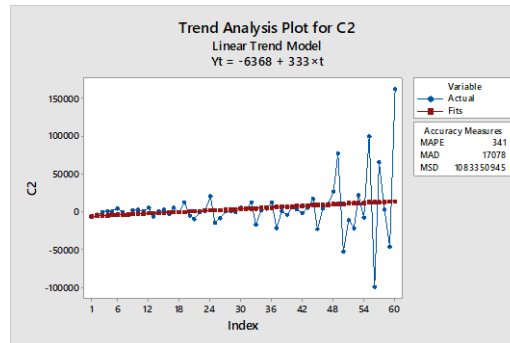

Figure 2. Graph of data analysis after differences

Based on Figure 4 we can see that the data is stationary, since the average does not move freely in a given time and has a fairly small variance and the actual value is close to a linear line.

Box-Jenkins forecasting model is identified by observing the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) value.

Autocorrelation Function: C2

| Lag	ACF	T	LBQ	Lag	ACF	T	LBQ
1	-0,422213	-3,24	11,06	8	0,008516	0,05	21,10
2	0,080420	0,53	11,47	9	0,026999	0,16	21,15
3	0,090006	0,59	11,99	10	-0,240178	-1,43	25,39
4	-0,274575	-1,79	16,92	11	0,214942	1,24	28,86
5	0,194301	1,20	19,44	12	0,016453	0,09	28,88
6	0,137676	0,83	20,73	13	0,053930	0,30	29,10
7	-0,073126	-0,44	21,10	14	0,028004	0,16	29,17
15	-0,077120	-0,43	29,65				
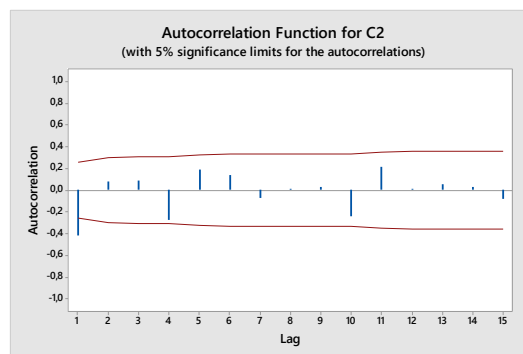
Figure 3. ACF Graph after differences

Next, determine PACF graph is as follows,

Partial Autocorrelation Function: C2

Lag	PACF	T	Lag	PACF	T
1	-0,422213	-3,24	9	0,037970	0,29
2	-0,119070	-0,91	10	-0,165956	-1,27
3	0,095954	0,74	11	-0,017758	-0,14
4	-0,232529	-1,79	12	0,068603	0,53
5	-0,020100	-0,15	13	0,191439	1,47
6	0,277918	2,13	14	0,041700	0,32
7	0,168095	1,29	15	0,026770	0,21
8	-0,076802	-0,59			
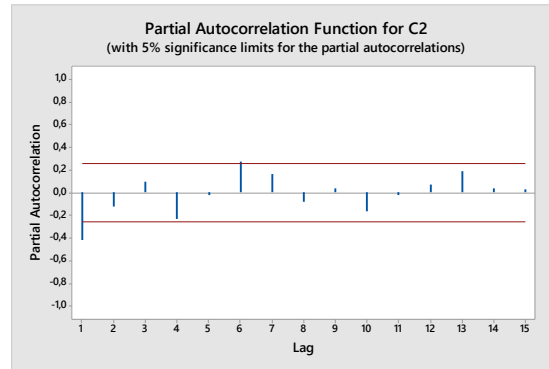
Figure 4. PACF graph after differences

Graph ACF in figure 3 looks data is stationary because the graphics dies down so it can be directly predicted the model and ρ_k cut off after lag-10 or lag-q for highest level moving average. While, the PACF graph in Figure 4 is seen following the sine graph and exponential dies down graph, with ϕ_{kk} value cut off at white noise after lag-1 and possibly on lag-2 or lag-p for highest level autoregresif.

So the approximate model is ARIMA (1,1,1) or AR (1) because the ACF graph exponential dies down and PACF is cut off on lag-1, differences 1 time, MA (1) because the ACF graph cut off at lag-1 and PACF exponentially dies down (sinusoidally), so general form ARIMA (1,1,1) can be written

$$Z_t = (1 + \phi_1)Z_{t-1} - \phi_1 Z_{t-2} + a_t - \theta_1 a_{t-1}$$

And also possible model estimation is ARIMA (2,1,1) or AR (2) because the ACF graph exponential dies down and PACF is cut off on lag-2, differences 1 time, MA (1) because the ACF graph cut off at lag-1 and PACF exponentially dies down (sinusoidally), so general form ARIMA (2,1,1) can be written

$$Z_t = (1 + \phi_1)Z_{t-1} + (\phi_2 - \phi_1)Z_{t-2} - \phi_2 Z_{t-3} + a_t - \theta_1 a_{t-1}$$

Furthermore, second difference is done with the aim of looking at the comparison between Mean Square value with the first difference result. The graph is as follows.
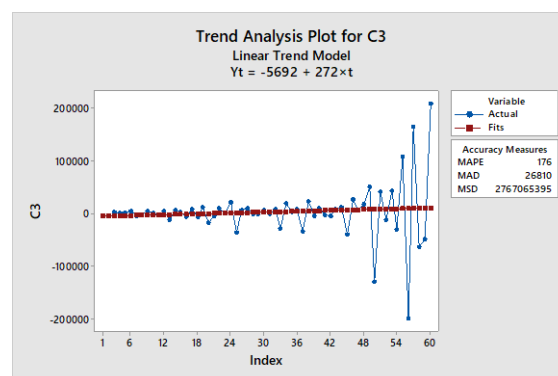
Figure 5. Graph of data analysis after second differences

Based on figure 7 shows that data is already stationary because the number of production is almost the same and not moving freely in a certain time, the value of variance is also quite small and the average actual value is close to linear line. Next is determined the value of ACF and PACF.

Autocorrelation Function: C3

Lag	ACF	T	LBQ	Lag	ACF	T	LBQ
1	-0,578791	-4,41	20,45	9	0,104674	0,58	28,92
2	0,152077	0,90	21,89	10	-0,252268	-1,38	33,53
3	0,089543	0,52	22,40	11	0,135853	0,72	34,90
4	-0,175751	-1,02	24,39	12	-0,046886	-0,25	35,06
5	0,016594	0,09	24,41	13	0,042087	0,22	35,20
6	0,168259	0,96	26,30	14	0,036553	0,19	35,31
7	-0,148727	-0,83	27,81	15	-0,044783	-0,24	35,47
8	0,068724	0,38	28,14				
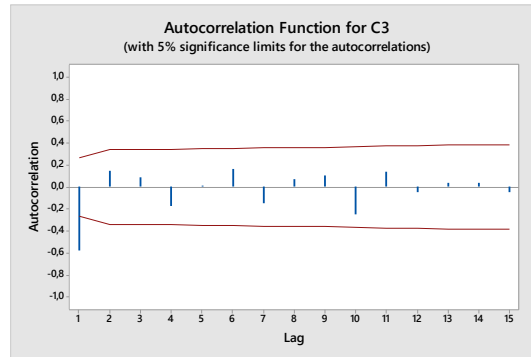
Figure 6. ACF graph after second differences

Partial Autocorrelation Function: C3

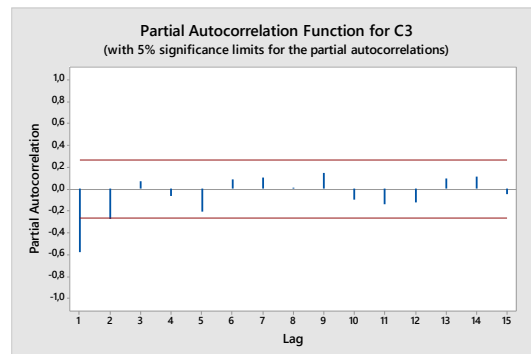| Lag	PACF	T	Lag	PACF	T
1	-0,578791	-4,41	9	0,144715	1,10
2	-0,275069	-2,09	10	-0,096190	-0,73
3	0,069252	0,53	11	-0,142189	-1,08
4	-0,060865	-0,46	12	-0,121910	-0,93
5	-0,208133	-1,59	13	0,096160	0,73
6	0,091007	0,69	14	0,118201	0,90
7	0,106055	0,81	15	-0,048933	-0,37
8	0,016395	0,12			

Figure 7. PACF graph after second differences

Graph ACF in figure 6 looks data is stationary because the graphics exponentially dies down (sinusoidally) and ρ_k cut off after lag-10 or lag-q for highest level moving average. While, the PACF graph in Figure 4 is seen following the sine graph and exponential dies down graph, with ϕ_{kk} value cut off at white noise after lag-1 and possibly on lag-2 or lag-p for highest level autoregresif.

So the approximate model is ARIMA (1,2,1) or AR (1) because the ACF graph exponential dies down and PACF is cut off on lag-1, differences 2 time, MA (1) because the ACF graph cut off at lag-1 and PACF exponentially dies down, so general form ARIMA (1,2,1) can be written

$$Z_t = (2 + \phi_1)Z_{t-1} - (1 + 2\phi_1)Z_{t-2} + \phi_1 Z_{t-3} + a_t - \theta_1 a_{t-1}$$

And also possible model estimation is ARIMA (2,2,1) or AR (2) because the ACF graph exponential dies down and PACF is cut off on lag-2, differences 2 time, MA (1) because the ACF graph cut off at lag-1 and PACF exponentially dies down (sinusoidally), so general form ARIMA (2,2,1) can be written

$$Z_t = (2 + \phi_1)Z_{t-1} - (1 + 2\phi_1 + \phi_2)Z_{t-2} + (\phi_1 + 2\phi_2)Z_{t-3} - \phi_2 Z_{t-4} + a_t - \theta_1 a_{t-1}$$

Parameter Estimation

After identifying the data and differences of data then the next step to determine parameter estimation of data.

1. Model ARIMA (1,1,1)

Based on Minitab calculation results obtained ARIMA (1,1,1) model with parameter ϕ_1 = -0,3868 and θ_1 = 0,4915, then the equation model is obtained

$$Z_t = (1 + \phi_1)Z_{t-1} - \phi_1 Z_{t-2} + a_t - \theta_1 a_{t-1}$$
$$Z_t = (1 - 0,3868\ )Z_{t-1} + 0,3868\ Z_{t-2} + a_t - 0,4915 a_{t-1}$$
$$Z_t = 0,6132 Z_{t-1} + 0,3868\ Z_{t-2} + a_t - 0,4915 a_{t-1}$$

With Mean Square (MS) value at first differences is 783047339.

2. Model ARIMA (2,1,1)

Based on Minitab calculation results obtained ARIMA (2,1,1) model with parameter ϕ_1 = -0,319, ϕ_2 = -0,0834 and θ_1 = 5572, then the equation model is obtained

$$Z_t = (1 + \phi_1)Z_{t-1} + (\phi_2 - \phi_1)Z_{t-2} - \phi_2 Z_{t-3} + a_t - \theta_1 a_{t-1}$$
$$Z_t = (1 - 0,319)Z_{t-1} + (-0,0834 + 0,319)Z_{t-2} + 0,0834 Z_{t-3} + a_t - 0,5572 a_{t-1}$$

$$Z_t = 0,681 Z_{t-1} + 0,2356 Z_{t-2} + 0,0834 Z_{t-3} + a_t - 0,5572 a_{t-1}$$

With Mean Square (MS) value at first differences is 795928310.

3. Model ARIMA (1,2,1)

Based on Minitab calculation results obtained ARIMA (1,2,1) model with parameter ϕ_1 = -0,6801 and θ_1 = 0,9586, then the equation model is obtained

$$Z_t = (2 + \phi_1)Z_{t-1} - (1 + 2\phi_1)Z_{t-2} + \phi_1 Z_{t-3} + a_t - \theta_1 a_{t-1}$$

$$Z_t = (2 - 0,6801)Z_{t-1} - \big(1 + 2(-0,6801)\big)Z_{t-2} - 0,6801 Z_{t-3} + a_t - 0,9586 a_{t-1}$$

$$Z_t = 1,3199 Z_{t-1} + 0,3602 Z_{t-2} - 0,6801 Z_{t-3} + a_t - 0,9586 a_{t-1}$$

With Mean Square (MS) value at first differences is 839223359.

4. Model ARIMA (2,2,1)

Based on Minitab calculation results obtained ARIMA (2,2,1) model with parameter ϕ_1 = -0,7833, ϕ_2 = -0,3061 and θ_1 = 1,0323, then the equation model is obtained

$$Z_t = (2 + \phi_1)Z_{t-1} - (1 + 2\phi_1 + \phi_2)Z_{t-2} + (\phi_1 + 2\phi_2)Z_{t-3} - \phi_2 Z_{t-4} + a_t - \theta_1 a_{t-1}$$

$$Z_t = (2 - 0{,}7833)Z_{t-1} - (1 - 2\,(0{,}7833) - 0{,}3061)Z_{t-2} + (0{,}7833 - 2(0{,}3061)Z_{t-3}$$
$$+ 0{,}3061 Z_{t-4} + a_t - 1{,}0323 a_{t-1}$$

$$Z_t = 1{,}2167 Z_{t-1} + 0{,}2605 Z_t + 0{,}1711 Z_{t-3} + 0{,}3061 Z_{t-4} + a_t - 1{,}0323 a_{t-1}$$

With Mean Square (MS) value at first differences is 736062253.

Verification and Forecasting

This stage aims to examine the best model that is by finding the smallest MS (Mean Square) value based on the result of parameter estimation that is as follows

- ARIMA (1,1,1) model has Mean Square value of 783047339
- ARIMA (2,1,1) model has Mean Square value of 795928310
- ARIMA (1,2,1) model has Mean Square value of 839223359
- ARIMA (1,1,1) model has Mean Square value of 736062253

Based on these four ARIMA models, it can be concluded that ARIMA (2,2,1) model has the smallest MS value. Thus the appropriate model for predicting tourist data is the ARIMA (2,2,1) model with equations as follows,

$$Z_t = 1{,}2167 Z_{t-1} + 0{,}2605 Z_t + 0{,}1711 Z_{t-3} + 0{,}3061 Z_{t-4} + a_t - 1{,}0323 a_{t-1}$$

ARIMA (2,2,1) model is the most suitable model for modeling time series data and forecasting future periods of data on the number of tourists coming to visit Tourism in Toraja. With the results of forecasting,

```
Forecasts from period 60
                 95% Limits
Period  Forecast   Lower   Upper   Actual
   61    149985    96799  203172
   62    193099   139016  247182
   63    207559   147641  267477
   64    202903   138044  267763
   65    222426   154848  290004
   66    229294   158182  300406
   67    239108   165308  312908
   68    250921   174731  327111
   69    260701   182277  339124
   70    271895   191568  352222
   71    283037   200990  365085
   72    294221   210649  377793
```

The above forcast table shows that for the period (t) after the next 60 months from January to December 2021 there is a significant increase each month with 95% limits lower and limit upper of actual data.

CONCLUSION

Based on the research results can be obtained conclusion is as follows, (1) The results obtained four time series models used in forcasting the number of tourists in Toraja in future period is ARIMA(1,1,1), ARIMA(2,1,1), ARIMA(1,2,1) and ARIMA(2,2,1). The correct criteria in selecting the best model is the model that has the smallest Mean Square (MS) value. In this case the time series model with the smallest MS value is ARIMA(2,2,1) that is 736062253. Thus, this model used in forecasting is ARIMA(2,2,1) with the general form $Z_t = 1{,}2167 Z_{t-1} + 0{,}2605 Z_t + 0{,}1711 Z_{t-3} + 0{,}3061 Z_{t-4} + a_t - 1{,}0323 a_{t-1}$; (2) the forecasting results number

of tourists in Toraja for January to December 2021 is 149985, 193099, 207559, 202903, 222426, 229294, 239108, 250921, 260701, 271895, 283037 and 294221. Results forcasting the number of tourists visiting from January to December from year to year increased. Thus, expected that the local government, especially the tourism office can formulate an ideal policy in the maximum service or best service of tourists who visit tourism places in Toraja. Local government should also be more serious to develop the existing tourist attractions there because with the increasing number of tours that visit it can automatically provide income for the improvement of the regional economy.

REFERENCES

Adams, K. M. (2006). *Art as politics: Re-crafting identities, tourism, and power in Tana Toraja, Indonesia*. University of Hawaii Press.

Aswi dan Sukarna. (2008). *Analisis Deret Waktu, Teori dan Aplikasi.* Makassar: Andhira Publisher.

Baldigara, T., & Mamula, M. (2015). Modelling international tourism demand using seasonal ARIMA models. *Tourism and hospitality management*, *21*(1), 19-31.

Bakhtiar, B., & Didiharyono, D. (2018). Perkiraan Lonjakan Jumlah Kunjungan Wisatawan di Kabupaten Toraja Utara Tahun 2021 dengan Penerapan Metode ARIMA. *Sainsmat: Jurnal Ilmiah Ilmu Pengetahuan Alam*, *7*(2), 153-162.

Biswas, R., & Bhattacharyya, B. (2013). ARIMA modeling to forecast area and production of rice in West Bengal. *Journal of crop and weed*, *9*(2), 26-31.

Chang, Y. W., & Liao, M. Y. (2010). A seasonal ARIMA model of tourism forecasting: The case of Taiwan. *Asia Pacific journal of Tourism research*, *15*(2), 215-221.

Crystal, E. (2012). Tourism in Toraja (Sulawesi, Indonesia). In *Hosts and Guests* (pp. 139-168). University of Pennsylvania Press.

Didiharyono, D., Marsal, M., & Nasruddin, N. (2017). Multivariate Regression Analysis With KICC Method In Measuring of Society Welfare In South Sulawesi. *Journal of Math Sciences*, *2*(2), 5-13.

Eni, Daniel and Fola J. Adeyeye. (2015). Seasonal ARIMA Modeling and Forecasting of Rainfall in Warri Town, Nigeria. *Journal of Geoscience and Environment Protection.* Vol. 3. August 2015, 91-98

Ho, S. L., Xie, M., & Goh, T. N. (2002). A comparative study of neural network and Box-Jenkins ARIMA modeling in time series prediction. *Computers & Industrial Engineering*, *42*(2-4), 371-375.

Kausar, D. R. K., & Gunawan, M. P. (2018). Managing heritage tourism in Toraja: strengthening local values and improving tourists' experiences. *Journal of Heritage Tourism*, *13*(6), 550-561.

Kulendran, N., & Wong, K. K. (2005). Modeling seasonality in tourism forecasting. *Journal of Travel Research*, *44*(2), 163-170.

Mondal, P., Labani, S., and Saptarsi, G. (2014). Study of Effectiveness of Time Series Modeling (ARIMA) In Forecasting Stock Prices. *International Journal of Computer Science, Engineering and Applications (IJCSEA).* Vol.4, No.2, 13-40

Paul, Jiban Chandra; Shahidul Hoque and Mohammad Morshedur Rahman. (2013). Selection of Best ARIMA Model for Forecasting Average Daily Share Price Index of Pharmaceutical Companies in Bangladesh: A Case Study on Square Pharmaceutical Ltd. *Global Journal of Management and Business Research Finance.* Volume 13, Issue 3, 14-25

Petrevska, B. (2017). Predicting tourism demand by ARIMA models. *Economic research-Ekonomska istraživanja*, *30*(1), 939-950.

Sarpong, Smart A. (2013). Modeling and Forecasting Maternal Mortality; an Application of ARIMA Models. *International Journal of Applied Science and Technology*. Vol. 3 No. 1, 19-28

Singh, E. H. (2013). Forecasting tourist inflow in Bhutan using seasonal ARIMA. *International Journal of Science and Research*, *2*(9), 242-245.

Steve, Nwankwo. (2014). Autoregressive Integrated Moving Average (ARIMA) Model for Exchange Rate (Naira to Dollar). *Academic Journal of Interdisciplinary Studies*. Vol. 3. No. 4, 429-433

Wei, William WS. (2006). *Time Series Analysis*. New York: Pearson Addison Wesle