Tourism Forecasting in Pahang: An Application of Homestay by using Box-Jenkins Approach

Noratikah Abu
Faculty of Science & Technology Industry,
Universiti Malaysia Pahang
Gambang, Malaysia
atikahabu@ump.edu.my

Megat Muhammad Afif Bin Megat Muainuddin
Faculty of Science & Technology Industry,
Universiti Malaysia Pahang
Gambang, Malaysia
megatafif929@gmail.com

Abstract—Tourism forecasting can lead to important elements in tourism industry to ensure that each investment by individuals, companies and government was worth it. From economy perspective, homestay is a growing business nowadays and can be an important indicator to the tourism industry. It also generates income revenue to the owner and also surrounding communities. This paper applied Box-Jenkins (ARIMA) method to make analysis and forecast for the number of domestics tourist used homestay in Pahang. This method much accurate for short and long term forecasting with high frequency of data. Results obtained by applying the proposed model and numerical calculation shows that it is effective for forecasting the number of homestay used by domestic's tourist and the number of homestay used also increase.

Keywords—forecasting; homestay; ARIMA

1. INTRODUCTION

Nowadays, tourism industry is one of the main economy indicators in the world and is growing very fast from day to day. Tourism forecasting is a method to predict number of tourist arrivals or number of tourist receipt in certain area [5]. Fewer errors in determining the total number of tourist, the country is able to inhibit gains. Tourism forecasting can lead to important elements in the tourism industry to ensure that each investment by individuals, companies and government was worth it. According to [5], the number of tourist arrivals is a most popular measure for demand tourism. Generally, there are three elements that can differentiate tourism product all along the world; destination attractiveness, facilities of destination and ease of destinations.

All states in Malaysia tend to develop their state as the major tourist destinations in Malaysia and every year different state organize tourism programmes. For 2017, Pahang had been chosen as one of the place to be visit in Malaysia. 'Visit Pahang 2017' had been declared by Pahang Regent Tengku Mahkota, Tengku Abdullah Sultan Ahmad Shah. Pahang, which is located at East Coast Malaysia and the population around 5.1% from total Malaysian population always be the tourist attraction because of its natural beauty of flora and fauna, covered by thousands of hectares of rainforest. Estimated exist more than 130 million year's old, national park located at Jerantut, gazetted as National Park since 1938, it is the largest national park more than 4000 square kilometres. Endau Rompin National provides limelight to the natural adventurous tourist. Besides that, the largest freshwater lake system in Malaysia known as Lake Bera which is at Temerloh. More than 200 species of birds, 65 mammal and thousands of freshwater fish can be main attraction in Pahang. Lake Chini which is covering more than 5000 hectares and as second largest natural freshwater also located at Pahang. The beautiful and unique natural scenery undeniable natural lovers around world. It is just a tip of an iceberg for tourist attraction in Pahang.

Based on this information, it can be seen that nature is the main source of tourism product in Pahang which lead to ecotourism product. In past, not many research about ecotourism, hence, this research attempt to study the forecasting of ecotourism product in Pahang. Ecotourism involved of natural, communities and sustainable travel. The International Ecotourism Society (TIES) defined ecotourism as responsible tourist to natural areas that maintain the environment, sustain the well-being of the local people and includes education.

[2] proposed forecasting model for the stochastic volatility data in tourism demand by using Grey forecasting model (FGM) and Fourier residual modification models. The study suggested that forecasting short term tourism demand can bring to excellent prediction. [4] have recently forecasted tourism demand from ASEAN using SARIMA model. Mean absolute percentage error (MAPE) and root mean square error (RMSE) were useful measure to compare the accuracy of models. Results showed that seasonality does not have an effect on the number of tourist in ASEAN. [5] introduced forecast the demand for Hong Kong tourism by combining statistical and judgemental forecasts via a web-based tourism demand forecasting system. They suggested that inclusion of quantitative models such as Box-Jenkins, naïve and exponential smoothing techniques can be strengthening the result. [3] used Poisson regression model to estimate international tourism demand in ASEAN countries. Data covered from 1995 to 2013, obtained from World Bank. The variables includes gross domestic product of foreign direct investment, real exchange rate, inflation and openness of trade. Results showed negative relationship between inflation, real exchange rate and international tourism demand.

2. METHODOLOGY

In this paper, we implemented Box-Jenkins (ARIMA) method to forecast number of tourist arrival in Pahang. The ARIMA method is combination of autoregressive (AR) and moving average (MA).

$$z_t = \delta + \phi_1 z_{t-1} + \phi_2 z_{t-2} + .. + \phi_p z_{t-p} + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - .. - \theta_q a_{t-q} \quad (1)$$

ARIMA method little bit different with others method because it does not assume pattern from historical data. It involves a few procedures for identifying, estimating, checking and forecasting with model. The first step in Box-Jenkins method is model identification. The purpose of this step is to determine whether the data stationary or not. The data can be converted to stationary series by using differencing. Then, parameter for model must be selected by minimizing the sum of squares of the fitting errors. Before run the model for forecasting, it must be checked for adequacy by using Ljung-Box Q statistic.

$$Q_m = n(n+2)\sum_{k=1}^{m} \frac{r_k^2(e)}{n-k} \quad (2)$$

In order to forecasting ARIMA model, the mean absolute percentage error (MAPE) is a useful measure for comparing the accuracy of forecast. Besides that, MAPE is computed by finding the absolute error in each period dividing the actual data for that period, and then averaging these absolute percentage errors. The formula as follow.

$$MAPE = \frac{1}{n}\sum_{t=1}^{n} \frac{|Y_t - \hat{Y}_t|}{Y_t} \quad (3)$$

Box-Jenkins methodology consists of five-step iterative procedures as follows:

Step 1: *Removal of Non-Stationery Components.*
The first step in model identification is to determine whether the series is stationary or non-stationary. The stationary can be determined by looking at autocorrelation function (ACF) and partial autocorrelation function (PACF).

Step 2: *Model Identification.*
The selection of an appropriate model is generally a complicated task. Analysis of the historical data is used to identify an appropriate Box-Jenkins model y looking at their ACF, PACF, and stationary and seasonality behavior if any. Once we select a tentative model, we can estimate the parameter of that model.

Step 3: *Parameter Estimation.*
We use historical data to estimate the parameters of the identified model.

Step 4: *Diagnostic Checking*.
Various diagnostic checking especially analysis of residual are used to check the adequacy of the tentatively identified model and, if need be, to suggest and improve as a tentatively new identified model.

Step 5: *Forecasting.*
Once we obtain a final model, we can use it to forecast future time series values.

3. RESULT AND DISCUSSION

Development of forecasting model was done by using Box Jenkins method with data of number of domestic tourist used homestay in Pahang from January 2009 until December 2016 gathered from tourism Pahang. There are five steps in Box Jenkins method: removal of non-stationery components, identification model, parameter estimation, model checking and forecasting. Figure 1 shows monthly number of domestic tourist used homestay in Pahang from January 2009 until December 2016.

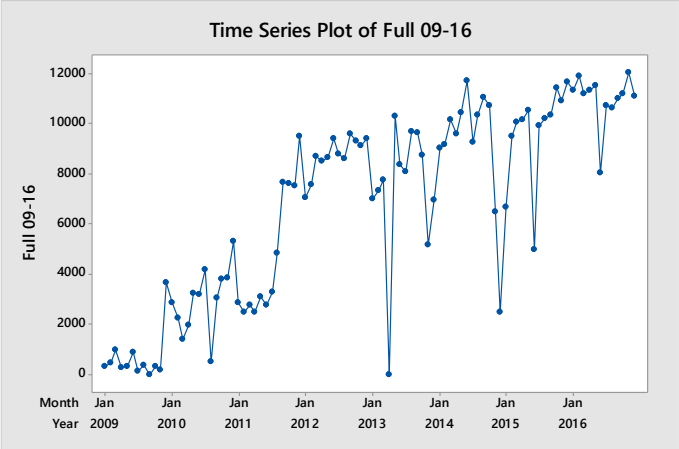

Figure 1 Time series plot for monthly number of domestic tourist used homes in Pahang

From Figure 1, the data was separated into two sets of data; training data and testing data. Separating data into training and testing sets is an important part of evaluating data. When we separate a data set into a training set and testing set, most of the data is used for training, and a smaller portion of the data is used for testing. Training data is used to fit the model while testing data is used to validate the model. Training data are data from January 2009 until January 2016, while testing data from February 2016 until December 2016. Since there is a very significant difference in data, we apply concept of transformation data for this research.

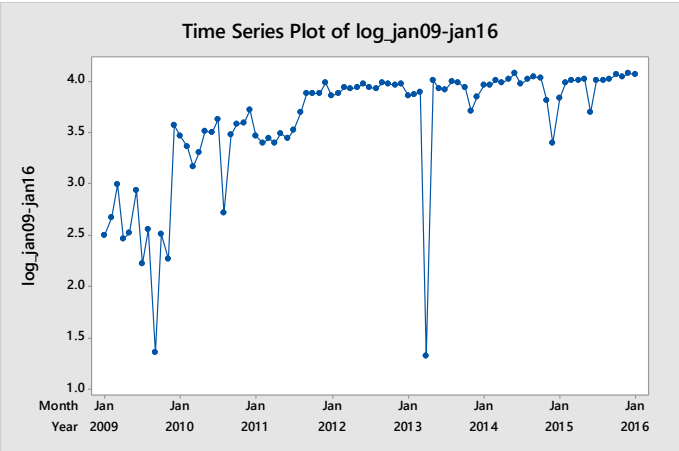

Figure 2 Time series plot for transforming data monthly number of domestic tourist used homes in Pahang

The time series plot in figure 2 shows that the transforming data from January 2009 until January 2016 by using log 10. As we can see, range for the data still significant. Thus we can start forecast the data using Box-Jenkins method.

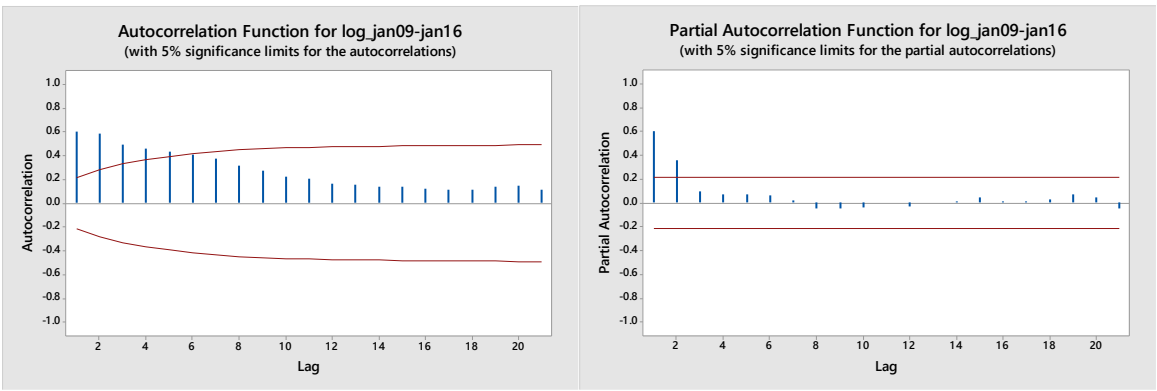

Figure 3 Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF)

The first step in Box Jenkins method is removal of non-stationery components. For this step autocorrelation (ACF) and partial autocorrelation (PACF) were plotted to determine whether the data is stationary or not. From the figure 3, we conclude that the

data is nonstationary as ACF were persistently large and trailed off to zero rather slowly and PACF cuts off after lag 2. In Box-Jenkins model the data must be stationery. If the data not stationery, it must be converted by using differencing.

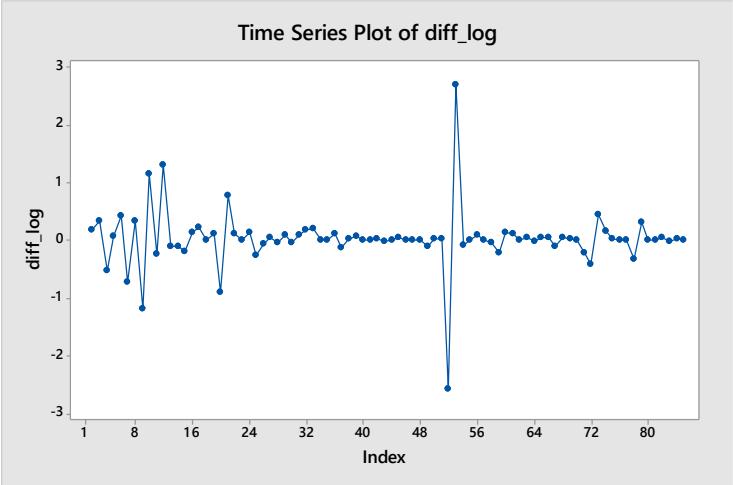

Figure 4 Time series plot non-seasonal difference

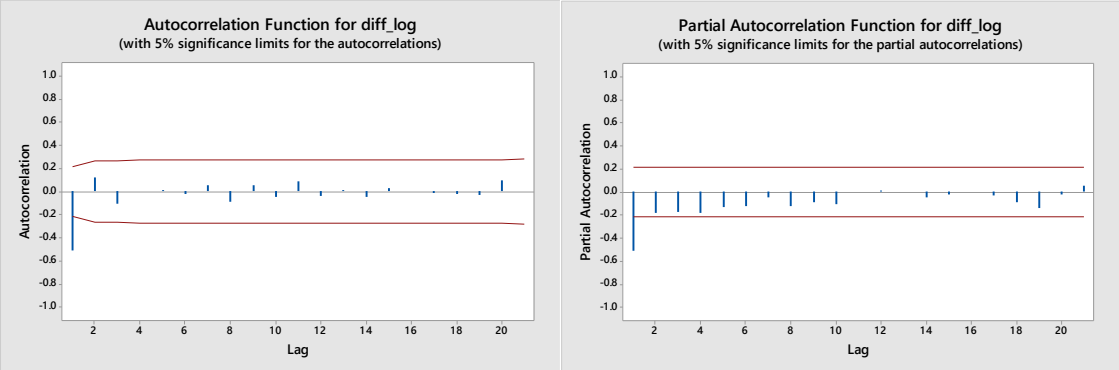

Figure 5 Non-seasonal difference Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF)

The time series plot in figure 4 shows that the differencing data and it was already satisfy the stationary condition. The ACF and PACF for first non-seasonal differencing are also plotted to verify the stationary condition of the data as stated Figure 5. Hence, this condition also proves that we are able to use ARIMA model for the data. Second step is model identification which shown in figure 5, there is a spike at lag 1 and cuts off after lag 4 for the ACF, which indicates moving average term, MA (1) behavior. For the non-seasonal pattern, the autocorrelation seems to cut off after lag1, which indicates moving average term MA behavior for the first difference series. Examining the PACF plot, autocorrelations have spikes after lag 2 and cuts off after lag 10, indicating it follows autoregressive term, AR (2). Hence, we choose ARIMA (0,1,1), ARIMA (1,1,0), ARIMA (1,0,1) ARIMA (1,1,1), ARIMA (2,1,2), ARIMA (2,2,1) and ARIMA (2,2,2) for our tentative models.

So far, we have considered forecast domestic tourist used homestay in Pahang from January 2009 until December 2016 using Box-Jenkins. In the following, we compare the p-value from each model to select which model suitable to use for forecasting number of tourist. The selected models were estimated by using least square method as for step three; parameter estimation.

Table 1 Result for ARIMA Model

Model	p-value	Result
(0,1,1)	0.624	Not Significant
(1,1,0)	0.739	Not Significant
(1,0,1)	0.610	Not Significant
(1,1,1)	0.599	Not Significant
(2,1,2)	0.810	Not Significant
(2,2,1)	0.269	Significant
(2,2,2)	0.591	Not Significant

From Table 1, the parameters are not significant if the p-value greater than 0.05 and we can eliminate the parameters from the model. We can see that only one model give p-value significant. It means the ARIMA (2,2,1) model is suitable for forecasting

domestic tourist used homestay in Pahang. After following the procedures in Box Jenkins forecasting stages, ARIMA (2,2,1) model gives a fit model that can forecast the number of domestic tourist used homestay in Pahang. Thus, we check diagnostic checking to see whether the histogram follow the normal probability curve so the model is adequate.

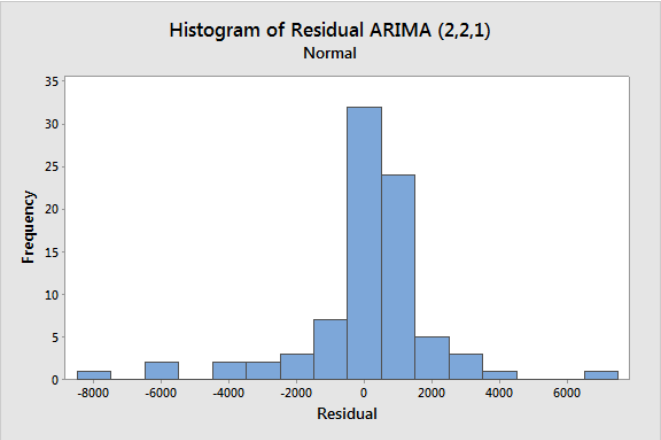

Figure 6 Histogram residual ARIMA (2,2,1)

In figure 6 it shows that the histogram residual follows the normal curve probability. Thus the model ARIMA (2,2,1) is most suitable for forecasting.
After select ARIMA (2,2,1) model, we calculated forecast error by using MAPE for both training and testing data. MAPE provides an indication of how large the forecast errors are in comparison to the actual values of the series.

Table 2 Result for MAPE

Data	**MAPE**
Training	1.5%
Testing	10.4%

The small MAPE result for both data indicates that the technique is not biased. Since the value is less than 10%. In this study, we intend to forecast for the next period of data from January 2017 December 2017. The analysis was made by the Minitab software as table 3:

Table 3 Forecasting data for domestic tourist used homestay in Pahang from January 2017 until Dec 2017.

Month	**Forecast**	**95% Limit**	
		Lower	**Upper**
January	11702.9	7861.2	15544.7
February	12306.3	7772.3	16840.3
March	12433.5	7613.3	17253.7
April	12727.1	7394.9	18059.3
May	13110.9	7336.2	18885.5
June	13410.7	7312.3	19509.2
July	13712.1	7295.9	20128.2
August	14039.1	7322.9	20755.4
Sept	14356.6	7376.6	21336.7
October	14669.5	7446.8	21892.2
November	14986.9	7538.1	22435.7
December	15304.2	7647.9	22960.6

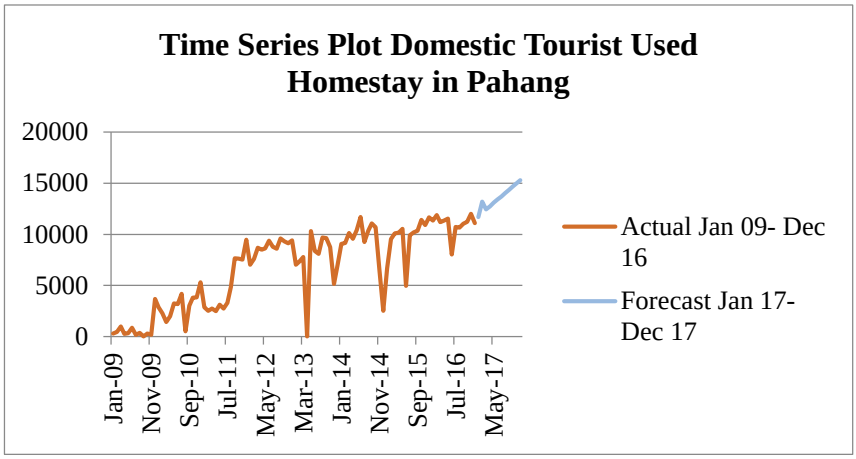

Figure 7 Time series plot for actual and forecasting value.

CONCLUSION

This paper discusses the forecasting of domestic tourist used homestay in Pahang by using Box-Jenkins method. We analyzed and discussed these data thoroughly and the plot of autocorrelation functions. However, by following the four steps in Box-Jenkins method, model identification, parameter estimation, diagnostic checking and forecasting, we conclude that ARIMA (2,2,1) model is the best model for forecasting domestic tourist used homestay in Pahang. As a conclusion, more forecasting method should be included in future studies in Malaysia.

ACKNOWLEDGMENT

This research is supported by Faculty of Industrial Sciences & Technology, Universiti Malaysia Pahang. This support is gratefully acknowledged. The authors would like to thank to lecturers and friends for many helpful ideas and discussions.

REFERENCES

[1] J. E. Hanke, D. W. Wichern, A. G. Reitsch, "Business Forecasting." New Jersey 2001

[2] Y. L. Huang, Y. H. Lee, "Accurately forecasting model for the stochastic volatility data in tourism demand." *Journal of Modern Economy*, 2, 823-829. 2011

[3] A. Karimi, P. Faroughi, K. A. Rahim, "Modeling and forecasting of international tourism demand in ASEAN countries." *American Journal of Applied Sciences*, 12(7), 479-486. 2015

[4] K. Loganathan, S. Thirunaukarasu, M. Kogid, "Is 'Malaysia truly Asia'? Forecasting tourism demand from ASEAN using SARIMA approach." *An International Multidisciplinary of Tourism*, 7(1), 367-381. 2012

[5] H. Song, B. Z. Gao, V. S. Lin, "Combining statistical and judgmental forecasts via a web-based tourism demand forecasting system." *International Journal of Forecasting*, 29, 295-310. 2013