Research of Forecasting on Tourist Arrivals to Malaysia

Nur Faezah Jamal, Nor Mariyah Abdul Ghafar, Mohd Zaki Awang Chek, Isma Liana Ismail

Abstract—Tourists get attracted towards Malaysia because of our culture and geography. Apart from heritage and culture, the tourists from all over the world visit here for various purpose. Therefore, forecasting tourist arrivals with high level of accuracy becomes important because it can ensure the development of tourism industries. So, this study focuses on tourist arrivals to Malaysia. This paper attempts to define the component of patterns exist in the time series data, to determine the most suitable model best fits in data series by using the error measure that are Mean Square Error (MSE) and Mean Absolute Deviation (MAD) and to forecast the one-step ahead forecast on the best model. In this study, data of tourist arrivals to Malaysia has been obtained from January 2000 until December 2018. All 228 monthly data were analyzed by using selected Univariate Modeling. The result found that tourist arrivals to Malaysia has a linear trend model and Double Exponential Smoothing with α = 0.17 was the best model for this time series.

Keywords: Forecasting, mean absolute error, mean square error, tourist arrivals.

I. INTRODUCTION

According to [1] said the definition of tourism consists of the activities of people travelling and living outside their usual environment.

Malaysia is 9th country in the world for tourist arrivals. Besides that, The Travel and Tourism Competitiveness Report 2017 has recognized Malaysia as the 25th out of 141 countries and ranked as the fifth best in Asia. The revenue from tourist arrivals has become an important source of income for economy, occupational, tax revenue, income and foreign exchange [2].

According to [3], Malaysia received international tourist of 25,948,459 in 2017 and has shown decline of 3% compared to the previous year. ASEAN dominated the 75.1% of the total tourist arrivals share and brought about 19,478,575 tourists to Malaysia.

Tourism industry is one of the sectors that can become more resilient and recover quickly [4]. The statistic shows that a significant drop of tourist arrivals in Malaysia for the year of 2015. Besides that, it is reported that the industry indicates a sign of recovery in the first half of 2016 as it registers increase of 3.7% arrivals compared to 2015 in the same period [5].

Forecasting is very important in tourism industry. The accuracy of forecasting provides direct assistance to the state government and industry player to assist them in making decisions, avoiding waste and inefficiency of tourism resources, thus reducing the risk and uncertainty [6] .Based on [7] is said tourism can be the largest world industry that stated in the World Travel and Tourism Report (WTTC) because the tourism is cannot only contribute to the employment sector, but contribution are also important in generating national wealth.

For this paper, the tourist arrivals analysis is focused mainly on Malaysia. There are three objectives in this study to be achieved. The first objective is to determine which component pattern exists in the data time series. The second objective is to determine the most suitable model best fits in data series. The last objective is to forecast the one-step ahead forecast on the best model.

II. LITERATURE REVIEW

According to [8] on to forecast tourist arrivals to Malaysia, they are several prediction or forecasting techniques were used such as the Autoregressive Integrated Moving Average (ARIMA) forecasting method, the naïve method, simple exponential smoothing method, Holt's method and Holts-Winter's trend and seasonal method. Based on the result obtain shows that the Moving Average 10 model was the better model rather than others.

In another study, in [9] make a comparison between Regression model and Autoregressive Fractionally Integrated Moving Average model with using Autoregressive Moving Average. It found that ARFIMA (7,0.42,10) model was better.

Based on study by [10] that focus quarterly data on forecasting international tourist arrivals to Australia using the Multiplicative Seasonal models outperform the Single, Double, The Holt-Winters Non-Seasonal Exponential Smoothing models and Holt-Winters Additive in forecasting for period 1998 to 2000, it shown that the first difference forecast of tourist arrivals performs bad performance rather than multi-stage projection.

Another study, by [11] tested for seasonal unit roots, in [12] the presence of unit roots did not mention in their tested.

The forecast obtained in a study of tourist in-flow in Singapore conducted by [13] using SARIMA model, ARIMA model and Holt Winters. The results show that SARIMA outperformed the other two models with Mean Absolute Percentage Error (MAPE) was 3.21.

There is no single forecasting method consistently perform models in every situation. In a competition to forecast with the best model to fit, the econometric models

Revised Manuscript Received on September 14, 2019.

Nur Faezah Jamal, Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, Perak Branch, Tapah Campus, Malaysia.

Nor Mariyah Abdul Ghafar, Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, Perak Branch, Tapah Campus, Malaysia.

Mohd Zaki Awang Chek, Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, Perak Branch, Tapah Campus, Malaysia.

Isma Liana Ismail, Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, Perak Branch, Tapah Campus, Malaysia.

are emphasized when the annual data are used whilethe model of time series usually show their advantage for higher frequency data.

III. METHODOLOGY

A. Data

This study used data that were obtained from a secondary source. Data on tourist arrivals to Malaysia were obtained from the Tourism Malaysia websites. The set of data consist of 228 monthly data on tourist arrivals to Malaysia from January 2000 until December 2018.

The training set contains 168 data points from January 2000 until December 2013 while the evaluation set consists of 60 data points from January 2014 until December 2018.

B. Univariate modeling

For this study, a few types of models will be used to study the patterns of component exist in the data which are Naïve with trend, Single Exponential Smoothing, Double Exponential Smoothing and Holt's Method. Naïve with trend model as in (1) can be used for a short time series and can be presented as,

$$F_{t+k} = \frac{y_t^2}{y_{t-1}} \quad (1)$$

where in model (1) can be denoted as the actual value at time t and y_{t-1} is the actual value in the preceding period. Then, a Single Exponential Smoothing can be determine by changing the current period of forecast with the difference between current forecast and actual value [13]. It can be denoted as below in (2),

$$F_{t+k} = \alpha y_t + (1-\alpha)F_t \quad (2)$$

where

F_{t+k} : smoothed value for single exponential in period $t+k$, where k =1,2,3, ...

y_t : the actual value in time period, t

α : the constant of unknown smoothing $(0 \le \alpha \le 1)$

F_t : the forecast value for period t

Second model is Double Exponential Smoothing that also known as Brown's method. A linear trend characteristic is useful to generate multiple-ahead forecast. It can be denoted as following:

Let,

P_t : smoothed value for exponential of y_t at time t

P_t' : smoothed value for double exponentially of y_t at time t

where

$$P_t = \alpha y_t + (1-\alpha)P_{t-1} \quad (3)$$

$$P_t' = \alpha P_t + (1-\alpha)P_{t-1}' \quad (4)$$

$$F_{t+k} = (2P_t - P_t') + \frac{\alpha}{1-\alpha}(P_t - P_t') * k \quad (5)$$

In model (3) is used to compute single exponential smoothed value of y_t and model as in (4) is used to compute double exponential smoothed value of y_t. Then for model as in (5), it is used to compute k-step-ahead forecast.

Lastly, to overcome the problems when estimating the trend values that sensitive to random influences Holt's Method can be. Holt's method obtained 2 parameters will provide more flexibilities. It can be denoted as the following:

$$P_t = \alpha y_t + (1-\alpha)(P_{t-1} + Q_{t-1}) \quad (6)$$

$$Q_t = \beta(P_t - P_{t-1}) + (1-\beta)Q_{t-1} \quad (7)$$

$$F_{t+k} = P_t + Q_t * k \quad (8)$$

Model as in (6) is used to compute the exponentially smoothed series, then model as in (7) is used to compute the trend estimate and model (8) is used to compute m-step-ahead into the future.

C. Error measures

In evaluating the models, MSE and MAD will be used for comparing the models. The following are the error measures that used for this study:

$$MSE = \frac{\sum_{i=1}^{n} e_t^2}{n} \quad (9)$$

$$MAD = \frac{\sum_{i=1}^{n} |e_t|}{n} \quad (10)$$

In model (9) and (10) $e_t = y_t - \hat{y}_t$ where the actual observed value at time t and $\hat{y}_t$ is the fitted actual at time t.

IV. RESULTS AND DISCUSSION

A. Descriptive analysis

In this analysis, there is only one variable since it was a univariatemodelling technique which is the value of tourist arrivals. Table I shows the average of tourist arrivals to Malaysia is 1,732,763. Then, the minimum value of tourist arrivals is 456,374 while the maximum tourist arrivals is 2,806,565. Thus, the maximum value indicates that the latest tourist arrivals is on December 2018.

Table- I: Descriptive analysis of tourist arrivals to Malaysia

Variable	Mean	Minimum	Maximum
Tourist Arrivals	1,732,763	456,374	2,806,565

B. Time series plot

A time series plot differs from other designs that collect data on the same variable at regular intervals (for instance, weeks, months, or years). Thus, this study uses monthly time series design to assess the impact of a treatment over time. Fig. 1 shows the tourist arrivals in Malaysia for January 2000-December 2018.

Tourist arrivals increase at an average rate of 10.0 per cent per annum, higher than the target of 6.9 per cent during the year 2000 to 2005. This steadily growth remained throughout that period except in 2003 when tourist arrivals decrease due to the Severe Acute Respiratory Syndrome (SARS) as well as uncertain world politics. After that, Malaysia tends to have a steady tourist arrival and the

number of arrivals keep on increasing.

Fig. 1:Tourist srrivals in Malaysia (2000 – 2018)

For determining the component exist in the time series, the data is illustrated by construct line graph. The result shown in Fig. 1. The line chart above show that there is a component exist in the time series data which is linear trend model. This is because the line chart shows that there is a general upward and downward movement along the linear line. The linear equation as follow in (11):

$$\hat{y} = 0.0068x + 0.9522 \quad (11)$$

where

$\hat{y}$: the value of tourist arrivals at time t

x : the time variable

This equation is useful to forecast the value of tourist arrivals in the future. The estimation can be changed as time is changed.

C. Comparison model for time series data

In this section, the tourist arrivals were evaluated by using Naive with Trend, Single Exponential Smoothing, Double Exponential Smoothing and Holt's Method. Microsoft Excel is employed to find the best value of parameters (α and β) using solver. The best model is determined by the lowest value of MSE and MAD. The summary of evaluation on MSE and MAD represented in Table II.

Table- II: Summary of evaluation using the MSE and MAD

Error Measure	Models	Fitted Period: January 2000 - Dec 2013	Evaluation Period: January 2014 - Dec 2018
MSE	Naïve with Trend	.0642	.1802
	Single Exponential (α=0.47)	.0255	.0330
	Double Exponential (α=0.17)	.0299	.0304
	Holt's Method (α=0.47, β=0.01)	.0254	.0320
MAD	Naïve with Trend	.1950	.3150
	Single Exponential (α=0.47)	.1140	.3140
	Double Exponential (α=0.17)	.1230	.1360
	Holt's Method (α=0.47, β=0.01)	.1130	.1370

Based on Table 2, the most suitable model to predict the value of tourist arrivals in Malaysia is Double Exponential Smoothing with α = 0.17. Since, this model has the smallest value of MSE and MAD for both evaluation parts compared to other models. On the second spot is Holt's Method with α = 0.47 and β = 0.01, followed by Single Exponential with α = 0.47; and lastly the worst model is the Naive with Trend. Thus, it can be concluded that Double Exponential Smoothing is the best model with α = 0.17. As a result, the one-step-ahead-forecast of the tourist arrivals to Malaysia for Double Exponential Smoothing with α = 0.17 is 2,175,722.

V. CONCLUSION

This paper studied the component exist in the time series data, determine the most suitable model best fits and forecast one step ahead forecast on the best model on tourist arrivals to Malaysia. Based on the analysis, the component that exist in tourist arrivals in Malaysia is linear trend model. Then, it was analyzed by using selected Univariatemodelling to achieve the second objective evaluated by using error measure which are MSE and MAD. The best model for this study is Double Exponential Smoothing. So, this model used to determine one step ahead forecast. Thus, all the objectives have been achieved. Further implementation can be focused in time series such as analyzing by using autoregressive integrated moving average (ARIMA) model.

VI. ACKNOWLEDGMENT

We would like to express our thank to UniversitiTeknologi MARA, Perak branch for their continuous supporting our research.

REFERENCES

1 N. H. Salem and S. Carvao, "World tourism organization," in Encyclopedia of Tourism, J. Jafari and H. Xiao, Eds. Cham: Springer International Publishing, 2016, pp. 1-2.
2 World Economic Forum, The travel & tourism competitiveness report 2017. [Online]. Available: http://www3.weforum.org/docs/WEF_TTCR_2017_web_0401.pdf.
3 Tourism Malaysia, Key-performance-indicator-2017. Putrajaya: Malaysia Tourism Promotion Board, 2017.
4 I. Jala, Tourism a key economic sector at these difficult times. The Star Online, 2016.
5 Tourism Malaysia, Malaysia's Jan-June 2016 tourist arrivals grow 3.7%. [Online]. Available: https://www.tourism.gov.my/media/view/malaysia-s-jan-june-2016-tourist-arrivals-grow-3-7.
6 R. Chen, C. Y. Liang, W. C. Hong, and D. X. Gu, "Forecasting holiday daily tourist flow based on seasonal support vector regression with adaptive genetic algorithm," Applied Soft Computing. 26, 2015, pp. 435-443.
7 World Travel and Tourism Council, Travel & tourism: Economic impact 2019 world. [Online]. Available: https://www.slovenia.info/uploads/dokumenti/raziskave/raziskave/world2019.pdf.

8 S. C. Chuah, Forecasting tourist arrivals to Malaysia. Master thesis, Selangor: Universiti Putra Malaysia, 2001.
9 M. Shitan, "Time series modelling of tourist arrivals to Malaysia," InterStat, 2008, pp. 1-12.
10 C. Lim and M. McAleer, "Forecasting tourist arrivals," Ann. Tour. Res., 28(4), 2001, pp. 965–977.
11 J. H. Kim, "Forecasting monthly tourist departures from Australia," Tour. Econ., 5(3), 1999, pp. 277–291.
12 L. W. Turner, "Univariatemodelling using periodic and non-periodic analysis: Inbound tourism to Japan, Australia and New Zealand compared," Tour. Econ., 3, 1997, pp. 39–56.
13 S. Safee and S. Ahmad, "Comparing the univariate modeling techniques, Box-Jenkins and Artificial Neural Network (ANN) for measuring of climate index," Appl. Math. Sci., 8(32), 2014, pp. 1557–1568.

AUTHORS PROFILE

NurFaezah Jamal is a Senior lecturer at Department of Statistics, Faculty of Computer and Mathematical Sciences, UniversitiTeknologi MARA, Perak Branch, Tapah Campus, Malaysia. Her research area is related to forecasting analysis.

N or Mariyah Abdul Ghafar is a Senior Lecturer at Department of Actuarial Science, Faculty of Computer and Mathematical Sciences, UniversitiTeknologi MARA, Perak Branch, Tapah Campus, Malaysia. Her research area is related to actuarial science and risk management.

MohdZakiAwangChek is a Senior lecturer at Department of Actuarial Sciences, Faculty of Computer and Mathematical Sciences, UniversitiTeknologi MARA, Perak Branch, Tapah Campus, Malaysia. His research area is related to Actuarial Modelling in Social Security.

Isma Liana Ismail is a Senior lecturer at Department of Statistics and Decision Science, Faculty of Computer and Mathematical Sciences, UniversitiTeknologi MARA, Perak Branch, Tapah Campus, Malaysia. Her current project is transportation scheduling.