ILTC 2019 PROCEEDING

3RD INTERNATIONAL LANGUAGE AND TOURISM CONFERENCE

THEME: Sustaining Global Development Goals Through Languages, Education, and Tourism

18th – 19th October 2019 | Kulliyyah of Languages and Management, IIUM Pagoh

PARTNERS:

ORGANIZED BY:

KULIYYAH OF LANGUAGES AND MANAGEMENT
INTERNATIONAL ISLAMIC UNIVERSITY MALAYSIA PAGOH CAMPUS

TOURISM

PROCEEDING OF ILTC 2019
(TOURISM)

INTERNATIONAL LANGUAGE AND TOURISM CONFERENCE 2019

Sustaining Global Development Goals Through Languages, Education, and Tourism

FIRST PUBLISHED 2019

Disclaimer: The organizer of ILTC 2016 is not responsible or liable for any mistake and opinion presented in this proceeding and will not alter the content of the research paper. It will be published in its original form.

ISBN:978-967-467-015-3

9 789674 670153

ORGANIZED BY:
Kulliyyah of Languages and Management,
International Islamic University Malaysia,
Pagoh Campus

PUBLISHED BY:
Kulliyyah of Languages and Management,
International Islamic University Malaysia,
Pagoh Campus

CONTENT

SUB-THEME: TOURISM

MODELING AND FORECASTING: A CASE STUDY OF TOURIST ARRIVALS IN MALAYSIA

Nur Haizum Abd Rahman*[1], Nur Nabilah Jamal[2] and Nur Hidayah Abd Rahman[3]

[1, 2] Department of Mathematics, Faculty of Science, Universiti Putra Malaysia, 43400 Serdang, Selangor, MALAYSIA.
(E-mail: *nurhaizum_ar@upm.edu.my)
[3] Kulliyyah of Languages and Management, International Islamic University Malaysia, Pagoh Campus, 84600 Pagoh, Johor, MALAYSIA.
(E-mail: hidayah_rahman@iium.edu.my)

ABSTRACT

Tourism industry has become one of the main sources for Malaysia's income. It affects other sectors such as hotels, retail businesses and transportations. Thus, it is important to monitor the development of tourism industry, so that proper planning can be enforced. Therefore, forecasting by using time series analysis, which is the Box-Jenkins method will be discussed here to provide future information to support the decision-making processes. The forecast performance has been compared by using magnitude error measurements. The empirical result shown that SARIMA $(0,1,0)(0,1,1)^{12}$ is the best model in forecasting tourist arrivals with only 8% different between 12 of the actual and forecast values.

Keywords: Tourism; forecasting; Box-Jenkins; time series

1. INTRODUCTION

Malaysia is separated into two regions which are the Peninsula Malaysia and Borneo with variety of ethnics and cultures. Malaysia has been blessed with cultural diversity and the amazing background. Therefore, Malaysia is seen as one of the most perfect spots for tourists' destinations around the world. For instance, Malaysia is a well-known destination as one of the most fascinating travel destinations in South East Asia.

Tourism is one of the vigorous sectors in contributing world's economy (Kadir & Karim, 2012). This is because tourism activities can gain huge profit where it becomes an economic savior to many countries. The tourism industry in Malaysia has been started in the 1970s, after the 1972 Conference of the Pacific Areas Travel Association (PATA) in Kuala Lumpur (Marzuki, 2010). This conference gave a relatively huge impact since it provides information on the profitable gain from the tourism industry. Since then, the government had put a lot of efforts in developing tourism industry in Malaysia. As an example, the Malaysian government has highlighting the tourism strategy developments in five Malaysian Plans (MPs) (Mosbah & Salleh, 2014). The successfulness of the tourism industry has proven since 1987 where the country's economic growth has been partially depended through this sector (Ooi, Hooy, & Mat Som, 2013).

The economic growth in tourism sector can be proven with the data recorded by the Ministry of Tourism, Art and Cultural Malaysia (MOTAC). The available data where it is in the form of time series give valuable information for future planning and management. Thus, the time series analysis is suitable as statistical approach that gives inside of the recorded data. The main purpose of using time series analysis is to develop a mathematical model that can produce forecasts of future observations (Khandelwal, Adhikari, & Verma, 2015). Forecasts are crucial for planning and inventory control in many applications. For tourism, the forecasts can help in define various data types such as the numbers of tourist arrivals, length of stay and expenditures. Therefore, this paper aims to model and forecast the tourism demand based on the data of tourist arrivals in Malaysia.

2. LITERATURE REVIEW

2.1 Forecasting in Tourism

Forecasting is important in the tourism sector since it will predict the trend of tourism demand and supply in a specific destination. Therefore, forecasting is seen as an important component in the tourism planning and management process (Elena et al., 2012). Previously, there are two approaches in demand forecasting which are qualitative approach and quantitative approach (Uysal & Crompton, 1985; Peiris, 2016). According to Uysal and Crompton (1985), the qualitative approach consists of Traditional Approaches, Delphi Model and Judgement-aided Model, meanwhile, the quantitative approach requires the existence of previous data which can be categorized as Time Series, Gravity and Trip Generation Model, and Multivariate Regression. Based on Uysal and Crompton (1985), the quantitative approach is seen as the suitable approach in which, the time series methods through the Box and Jenkins model (Box & Jenkins, 1976) can provide a direct statistical result to forecast the future of tourism demand. Based on the Box-Jenkins model, the time series approach that commonly being applied in previous studies is simple ARIMA model or seasonal ARIMA model (Peiris, 2016). The SARIMA model is seen as the most suitable and direct method to forecast the tourist flows and arrivals in a specific destination (Suhartono, 2011). Hence, this study is intended to model and forecast the tourist arrivals in the context of Malaysia through the SARIMA model.

2.2 Tourism Industry in Malaysia

Malaysia has been blessed with multicultural people with various religions and unique history which can contribute to tangible and intangible heritage. Hence, Malaysia owns a Unique Selling Proposition (USP) which can be promoted to attract international tourists to visit Malaysia. The tourism industry is seen as an important sector to boost the Malaysia economy and therefore, there are various efforts have been done by the Malaysian Government through Malaysian Plans, National Tourism Policy (NTP), National Ecotourism Plan (NEP), and Malaysia Tourism Transformation Programme (MTTP) (Mosbah & Salleh, 2014). Moreover, various efforts have been done by MOTAC to increase the numbers of international tourist arrivals such as promoting the Visit Malaysia Year (VMY). Thus, it is important for the current study to forecast the tourist arrivals which in line with the target of Visit Malaysia Year 2020.

3. METHODOLOGY

3.1 Box-Jenkins Method

Box-Jenkins methodology is a class of linear models that is capable in representing stationary as well as nonstationary time series (Rahman, Lee, Suhartono, & Latif, 2016). This methodology refers to a set of procedures for model identifying, model estimating and model checking using time series data (Hanke & Wichern, 2005). Forecasts follow directly from the form of the fitted models. The Box-Jenkins aim to obtain a model that is parsimony. Parsimony referred to a model that has the smallest number of parameters needed to adequately fit the patterns in the observed data.

There are several types of Box-Jenkins model including autoregressive (AR), moving average (MA), autoregressive moving average (ARMA) and autoregressive integrated moving average (ARIMA). AR, MA and ARMA representing the model for stationary data while ARIMA represent model for the non-stationary data. The general form of ARIMA model can be written as:

$$\phi_p(B)(1-B)^d Y_t = \theta_q(B)\varepsilon_t \quad (1)$$

where $\phi_p(B) = 1-\phi_1 B - \ldots - \phi_p B^p$ and $\theta_q(B) = 1-\theta_1 B - \ldots - \theta_q B^q$.

The ARIMA model is a generalization of simple model, AR, MA and ARMA and with the integration, *I*, in data differencing. The form can be denoted as ARIMA (p,d,q). The notation *d* represents the number of times that the raw observations been differences.

There also another form of Box-Jenkins method which is seasonal autoregressive integrated moving average (SARIMA) model where the seasonal components are included in the model. The generalized form of SARIMA model can be extended from (1) which is:

$$\phi_p(B)\Phi_P\left(B^S\right)(1-B)^d\left(1-B^S\right)^D Y=\theta_q(B)\Theta_Q\left(B^S\right)\varepsilon \qquad (2)$$

where $\phi_p(B)=1-\phi_1 B-\dots-\phi_p B^p$, $\Phi_P(B)=1-\Phi_1 B-\dots-\Phi_P B^P$, $\theta_q(B)=1-\theta_1 B-\dots-\theta_q B^q$ and $\Theta_Q(B)=1-\Theta_1 B-\dots-\Theta_Q B^Q$. The seasonal in seasonal ARIMA can be abbreviated as SARIMA

3.2 Forecast Evaluations

For this study that compare between single series, the root mean squared error (RMSE) and mean absolute error (MAE) are used in the magnitude error group. MAE and RMSE are commonly used since both are the scale-dependent measures (Hyndman & Athanasopoulos, 2014).

The RMSE calculated the loss function associated with the average of the quadratic or square error loss. For MAE, it measures the magnitude of forecast errors where it takes the absolute value. Thus, it does not account for over-forecast or under-forecast. The RMSE and MAE can be written as:

$$RMSE=\sqrt{\frac{1}{n}\sum_{t=1}^{n}\left(y_t-\hat{y}_t\right)^2} \qquad (3)$$

$$MAE=\frac{1}{n}\sum_{t=1}^{n}\left|y_t-\hat{y}_t\right| \qquad (4)$$

These measurements, RMSE and MAE are widely applied since it is simple and understandable. Such error can be calculated directly from the actual and forecast values without involving any unknown parameter that needs to be estimated (Elliott & Timmermann, 2005).

4. MAIN RESULTS

The monthly tourist's arrivals data in Malaysia is used from January 2000 until December 2017. The data is divided into two parts; (a) January 2000 to December 2016 as the training data set to define the model, (b) 2017 (January to December) as the testing data set to check the forecast performance. The data is shown in Table 1 meanwhile the time series plot is given in Figure 1.

Table 1. *Tourist arrivals data from year 2000 until 2017*

Year	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
2000	731,509	786,040	737,678	916,382	894,350	960,782	811,876	764,500	778,587	990,658	806,648	1,042,572
2001	1,052,854	1,09,0219	1,492,235	1,358,776	1,038,210	1,308,788	1,125,504	1,149,987	927,206	596,202	759,321	975,771
2002	834,519	986,408	1,272,810	1,106,053	1,153,033	1,297,536	1,135,691	1,163,724	1,029,120	1,015,946	1,061,229	1,253,941
2003	1,070,428	985,343	819,376	**456,374**	541,267	719,047	753,029	826,234	943,526	994,858	1,135,493	1,328,940
2004	1,408,378	1,295,913	1,274,230	1,266,824	1,262,717	1,381,355	1,227,052	1,268,237	1,326,549	1,359,498	1,329,208	1,303,445
2005	1,394,032	1,346,550	1,419,275	1,317,749	1,351,141	1,339,051	1,317,276	1,393,666	1,335,027	1,367,735	1,432,317	1,417,236
2006	1,453,803	1,384,479	1,541,056	1,422,576	1,393,656	1,368,651	1,450,319	1,503,748	1,384,890	1,540,070	1,564,286	1,539,329
2007	1,721,786	1,652,004	1,906,304	1,764,868	1,841,870	1,803,594	1,713,951	1,642,899	1,601,016	1,673,922	1,764,586	1,886,022
2008	1,780,134	1,742,468	1,819,689	1,760,326	1,899,148	1,961,355	1,928,082	1,839,235	1,599,418	1,818,304	1,845,645	2,058,684
2009	1,871,099	1,613,309	1,975,776	1,883,873	1,894,059	2,108,328	2,003,724	2,030,337	1,997,535	2,078,485	2,048,595	2,141,071
2010	1,896,918	1,832,300	2,022,590	1,877,934	1,992,277	2,246,084	2,214,092	2,099,485	2,053,406	2,137,735	2,081,354	2,123,021
2011	1,918,751	1,669,288	1,952,343	1,891,350	1,892,095	2,039,035	2,406,100	2,166,937	2,094,104	2,189,014	2,225,534	2,269,773
2012	1,817,061	1,773,471	1,972,006	1,924,129	1,951,925	2,193,886	2,306,675	2,240,687	1,973,803	2,205,979	2,236,473	2,436,613
2013	2,070,394	2,002,709	2,376,295	1,984,982	2,044,039	2,074,312	2,129,393	1,980,960	2,093,392	2,106,569	2,045,850	**2,806,565**
2014	2,44,7397	2,118,540	2,525,496	2,175,009	2,266,417	2,342,187	2,229,920	2,273,271	2,233,545	2,247,666	2,130,022	2,447,845
2015	2,291,603	1,949,016	2,242,077	2,071,922	2,118,890	1,893,792	2,216,049	2,182,536	2,084,339	2,082,866	2,004,694	2,583,467
2016	2,376,166	2,091,098	2,198,716	2,101,280	2,144,119	2,121,396	2,296,615	2,282,173	2,118,367	2,326,487	2,054,165	2,646,810
2017	2,350,270	2,043,215	2,238,184	2,145,734	2,039,016	2,134,647	2,263,478	2,1290,13	2,092,378	2,068,995	2,007,965	2,435,564

Source: Ministry of Tourism, Art and Cultural (MOTAC)

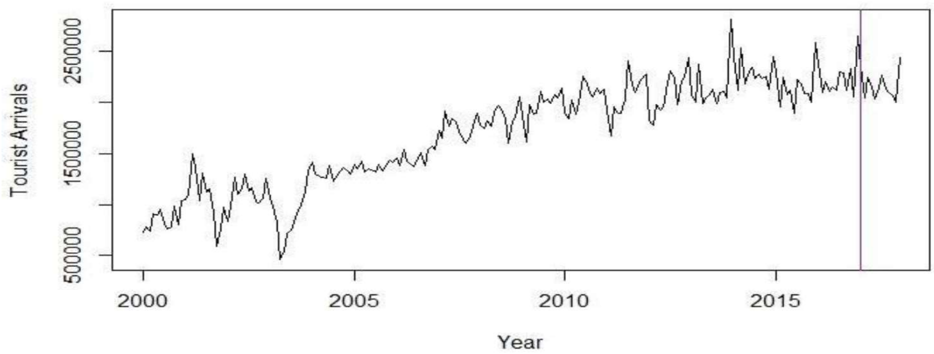

Figure 1: Time series plot for tourist arrivals from year 2000 until 2017

Roughly, based on Figure 1 and Table 1, the number of tourist arrivals is increasing from year to year. In December 2013, the highest number of tourist arrivals is recorded with a total of 2,806,565 tourists have visited Malaysia. On the other hand, the minimum tourist arrivals recorded in April 2003 with the total number of 456,374 tourists. According to Kusni, Kadir and Nayan (2014), the reason that caused huge decrement in number of tourists visited Malaysia in April and May 2003 is because of the outbreak of Severe Acute Respiratory System (SARS) disease. Since it affected human health and due to safety issues, the number of tourist arrivals decrease not only in Malaysia but also in other Asian country like Indonesia (Ooi, Hooy, & Mat Som, 2013). However, starting from June 2003, the number of tourists increased due to a media statement from the Health Ministry director general Tan Sri Mohamad Taha Arif. He had clarified Malaysia is free from SARS with no reported death or suspended case (Edwards, Mustafa, & Koh, 2003).

Besides, SARS attack, there is also other major issue in year 2008 and 2009 where the global economic crises occurred. However, Malaysia still progressively recorded increased number of tourists compared to previous year up to 5% (Mosbah & Salleh, 2014).

For forecasting purposed in the year 2017, the Box-Jenkins method is conducted where the stationary series is the important methodology. Data transformation is made accordingly, with detrended, deseasonalised and differencing for both nonseasonal (d=1) and seasonal order (D=12) (Suhartono, 2011). The models are identified from autocorrelation function (ACF) and partial autocorrelation function (PACF). The best model is chosen among the competitive models based on the smallest error. Three possible models are determined from tourist arrivals training dataset which are SARIMA (0,1,0) $(0,1,1)^{12}$, SARIMA (0,1,0) $(1,1,1)^{12}$ and SARIMA (0,1,0) $(1,1,0)^{12}$.

Table 2: *Forecast performance based on RMSE and MAE*

Evaluations	SARIMA Model		
	(0,1,0) $(0,1,1)^{12}$	(0,1,0) $(1,1,1)^{12}$	(0,1,0) $(1,1,0)^{12}$
RMSE	**193626.2**	209971.6	207697.2
MAE	**177451.7**	195240.8	192581.6

Instead of comparing the performance in training data set, this study compared the performance in testing data set due to the objective is to find the best forecast model. The result is given in Table 2. As shown, the best SARIMA model is $(0,1,0)(0,1,1)^{12}$ with the lowest value in RMSE and MAE, 193626.2 and 177451.7 respectively. Therefore, the actual and forecast plot is given in Figure 2. As shown, the developed model (forecast) can follow most of the pattern recorded in the actual values.

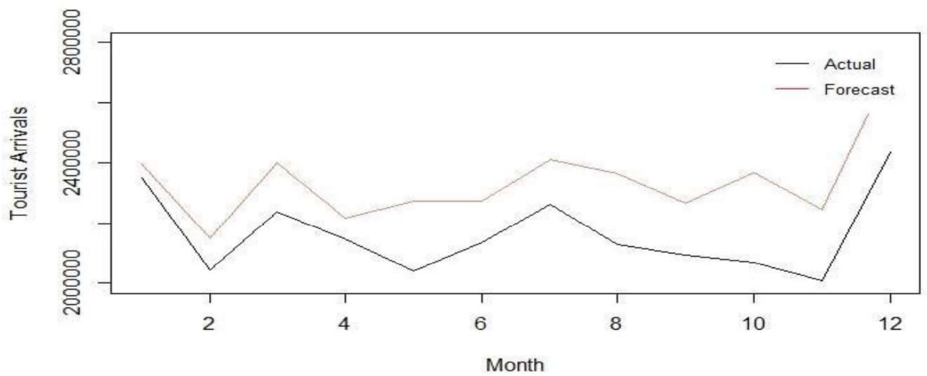

Figure 2: Actual and forecast values for Malaysia tourist arrivals in 2017

5. CONCLUSION

The findings have shown that the best model to forecast the tourist arrivals in Malaysia is SARIMA $(0,1,0)\ (0,1,1)^{12}$. Although this method known to be the simplest or conventional methods in time series research area, yet this result is acceptable since it provides high accuracy with the difference between actual and forecast values is only 8%. Several previous studies also have shown that simple method could outperform more advance or complicated methods (Makridakis, Spiliotis, & Assimakopoulos, 2018). This is because the model can immediately adapt to the change in the series.

The forecast result can be the benchmark idea since it can lead to better improvement in the tourism industry. In addition, government or any related agencies can be beneficial with the forecast values in planning tourism related programs such as providing enough accommodation or facilities to the tourist.

BIBLIOGRAPHY: Nur Haizum Abd Rahman is a senior lecturer in the Department of Mathematics, Faculty of Science, Universiti Putra Malaysia. Any questions regarding this paper, Modeling and Forecasting: A Case Study of Tourist Arrivals in Malaysia can be contacted through her email nurhaizum_ar@upm.edu.my.

ACKNOWLEDGMENT: The authors would like to express their appreciation for the support of the Putra-IPM research grant with Project No 9587700.

REFERENCES

Box, G. E. P., & Jenkins, G. M. (1976). Time series analysis: Forecasting and control San Francisco. *Calif: Holden-Day*.

Edwards, A., Mustafa, M., & Koh, C. (2003). Malaysia 'free from SARS.' *The Star*. Retrieved from https://www.thestar.com.my/news/nation/2003/04/01/malaysia-free-from-sars/

Elena, M., Lee, M. H., Suhartono, Hossein, Haizum, N., & Bazilah, N. A. (2012). Fuzzy time series and sarima model for forecasting tourist arrivals to Bali. *Jurnal Teknologi (Sciences and Engineering)*.

Elliott, G., & Timmermann, A. (2005). of Forecast Estimation and Testing Flexible Loss under Rationality, *72*(4), 1107–1125.

Hanke, J. E., & Wichern, D. W. (2005). *Business Forecasting* (8th ed.). New Jersey: Pearson/Prentice Hall.

Hyndman, R. J., & Athanasopoulos, G. (2014). *Forecasting: Principles and Practice: Notes*. *OTexts*.

Kadir, N., & Karim, M. Z. A. (2012). Tourism and Economic Growth in Malaysia: Evidence from Tourist Arrivals from Asean-S Countries. *Economic Research-Ekonomska Istraživanja*.

Khandelwal, I., Adhikari, R., & Verma, G. (2015). Time series forecasting using hybrid arima and ann models based on DWT Decomposition. *Procedia Computer Science*.

Kusni, A., Kadir, N., & Nayan, S. (2014). International Tourism Demand in Malaysia by Tourists from OECD Countries: A Panel Data Econometric Analysis. *Procedia Economics and Finance*.

Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2018). The M4 Competition: Results, findings, conclusion and way forward. *International Journal of Forecasting*, *34*(4), 802–808. Elsevier.

Marzuki, A. (2010). Tourism Development in Malaysia. *Urban Management*.

Mosbah, A., & Salleh, M. (2014). A review of tourism development in malaysia. *European Journal of Business and Management*.

Ooi, C., Hooy, C., & Mat Som, A. P. (2013). Tourism Crises and State Level Tourism. *International Journal of Business and Society*.

Peiris, H. R. (2016). A Seasonal ARIMA Model of Tourism Forecasting: The Case of Sri Lanka. *Journal of Tourism, Hospitality and Sports*, *22*, 98–109.

Rahman, N. H. A. H. A., Lee, M. H. H., Suhartono, & Latif, M. T. T. (2016). Evaluation performance of time series approach for forecasting air pollution index in Johor, Malaysia. *Sains Malaysiana*, *45*(11), 1625–1633. R

Suhartono. (2011). Time Series Forecasting by using Seasonal Autoregressive Integrated Moving Average: Subset, Multiplicative or Additive Model. *Journal of Mathematics and Statistics*, *7*(1), 20–27.

Uysal, M., & Crompton, J. L. (1985). An Overview of Approaches Used to Forecast Tourism Demand. *Journal of Travel Research*.