

Modeling and Forecasting of Tourism Demand in Malaysia

Siti Syahirah Abdul Halim¹, Nora Muda²

^{1,2}Department of Mathematical Sciences, Faculty of Science and Technology, Universiti Kebangsaan Malaysia, 43600 Bangi, Selangor, Malaysia.

ABSTRACT: The pattern of foreign tourist demand to Malaysia is analysed and forecasted using time series method and non-linear technique. There are nine selected countries that contribute a lot to tourist arrivals to Malaysia, namely Australia, Brunei, China, Indonesia, India, Japan, the Philippines, South Korea, and the United Kingdom. Box-Jenkins time series method and Singular Spectrum Analysis are conducted and compared to study the best model to forecast the foreign tourist demand to Malaysia. Monthly data of tourism arrival in 1990 to 2014 were used and the forecasting were compared with 2015. Based on the results obtained, the forecasting model of Box-Jenkins time series method is the best model based on the percentage accuracy in forecasting the tourist demand to Malaysia.

KEYWORDS: Tourist, Time Series Method, Box-Jenkins (SARIMA), Singular Spectrum Analysis (SSA).

INTRODUCTION

Tourism is derived from the word tour or touring which means visiting while sightseeing. It can also be defined as the act of leaving home going to a destination with certain reason or purpose and with the intention to return to the original location. Islamic Tourism Centre under the Ministry of Tourism and Culture Malaysia has defined Islamic tourism as a field or industry related to a tour to visit places with Islamic history, getting to know Islamic culture and heritage as well as searching and understanding the Islamic community way of life.

The tourism industry, from the perspective of the national economy, is one of the most important industries, although initially, the government does not emphasize the tourism industry and focuses on the country's infrastructure development instead. In the 1960s, the tourism industry was yet to play an important role towards the country's economic development. The government also focused on natural resources sectors such as tin and rubber that contributed 45% to Gross National Product (GNP). The tourism began to improve positively in the 1970s. Therefore, the tourism industry is a growing and expanding industrial sector each year generating substantial profits to the country's economy in the form of foreign exchange and tourism is a means to introduce Malaysia globally. In Malaysia, the tourism sector is the fifth largest sector after the financial services, oil palm, wholesale and retail, gas and energy as well as petroleum sectors. According to the history of the Ministry of Tourism and Culture Malaysia, in 1987, the government showed support to the tourism sector by establishing the Ministry of Culture and Tourism. However, in April 2004, a special ministry known as the Ministry of Tourism Malaysia was set up to manage all matters concerning tourism and to proof the government's commitment to promoting the country as one of the contributing revenues to the country. After the 13th General Election, Ministry of Tourism Malaysia was renamed Ministry of Tourism and Culture Malaysia as the tourism and cultural sectors are closely linked in promoting the country as a tourist destination which is in line with the slogan "Malaysia, Truly Asia" (Official Portal of Ministry of Tourism and Culture Malaysia, May 2015). The government and the private sector are promoting Malaysia through various activities. Various government-run programs that cover culture, arts, national festivals, sports, conventions and exhibitions. The programs such as sports tournaments were introduced namely the Tour de Langkawi Bicycle Tournament and the Langkawi Maritime and Aerospace Exhibition (LIMA). With the establishment of the international sport, the government can introduce Malaysia to delegates or athletes from overseas to attract tourists to come to Malaysia and prove that the nation is able to do well in terms of international sports and excellence in tourism. As a result, tourists will keep continue to come to Malaysia for their tourism destination. The campaign of "Visit Malaysia Year 1990" was launched and was continued to the following year in 1994, 2007, 2008 and 2014. This campaign has made the numbers of tourist arrival to Malaysia increased. Fig. 1 shows the total number of tourist arrival in Malaysia from 1990 to 2014.

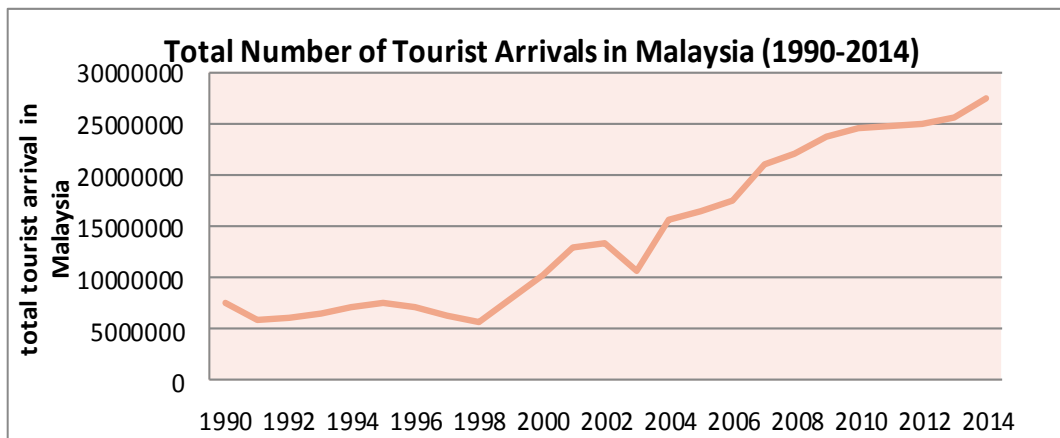


Figure 1. Total number of tourist arrivals in Malaysia from 1990 to 2014.

Source: Annual Report of Tourism Statistics 1990-2014

Fig. 2 shows the monthly tourist arrival data to Malaysia from January 2000 to December 2014. Based on Figure 2, tourist arrivals in Malaysia have a non-uniform cycle. Prior to 2003, the number of arrivals increased and decreased rapidly and unevenly. In 2003, the number of tourist arrivals in Malaysia dropped dramatically due to SARS attacks. However, overall, the number of tourist arrivals in Malaysia is increasing year by year.

Under the 9th Malaysia Plan (2006-2010), the government spent RM1.8 billion on upgrading tourism destinations, infrastructure and promoting the main tourist market. While the 10th Malaysia Plan (2011-2015), Malaysia will intensify trade promotion, enhance regional ties and continue regional integration through trade agreements, especially Asia and ASEAN. In 2012, tourist arrivals to Malaysia recorded an increase to 25 million people and RM 60 billion in tourism revenues. The activity promoting Malaysia with the theme of 'Celebrating 1 Malaysia Truly Asia' has contributed to the government's target of receiving 36 million tourists and RM 168 billion tourism revenue by 2020, as outlined in the Malaysia Tourism Transformation Plan (MTTP) 2020 (Ministry of Tourism Official Portal and Culture of Malaysia, October 2014).

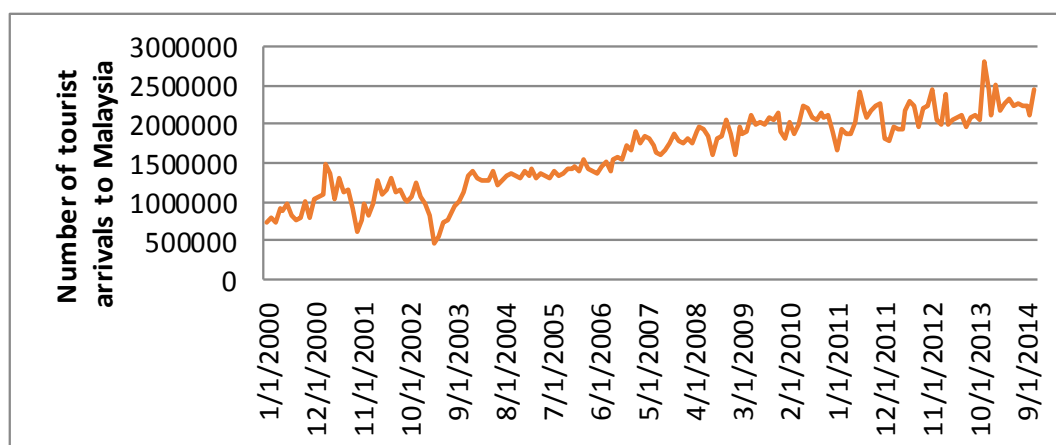


Figure 2. Total number of tourist arrivals in Malaysia from January 1990 to December 2014

Table 1 shows the number of tourist arrivals to Malaysia and tourism income. In Table 1, '+' shows an increase in the number of tourist arrivals compared to the previous year. While '-' showed a decrease in the number of tourist arrivals during the year compared to the previous year. In 1990, the number of tourist arrivals to Malaysia increased dramatically by 7.4 million people compared to 4.8 million in 1989. The growth had increased by 53.64%, and it had a positive impact on tourism revenue which also increased by 60.56%. In 1999, 2000, 2007 and 2014, respectively, it saw a steady increase in tourist arrivals to 7.9

million, 10.2 million, 20.9 million and 27.4 million respectively. However, there were several years in the middle of 1990 to 2014, a sharp decline due to economic recession, international political turmoil, natural disasters, and disease. According to the Malaysia Tourism Promotion Board, total income also increased from RM 4500 million in 1990 to generating RM 72000 million in revenue in 2014. This proved tourism industry could help the country's economic growth.

Indirectly, the tourism industry has affects Gross Domestic Product (GDP), employment opportunities, export earnings and investments. According to World Travel & Tourism Council, WTTC (2015), in 2014, the tourism sector contributed RM161 billion representing 14.9% of GDP and 1 770 000 employment opportunities. In addition, the tourism sector generates RM 74 billion representing 8.6% of total exports, in fact, investments also contribute RM 19 billion representing 6.8% of total investment in 2014. Thus, the tourism sector is one of the largest contributors to generating income country.

The main arrival of tourists to Malaysia are mostly from Asia and ASEAN countries. Countries that are the major contributors to the Malaysian market are Australia, Brunei, China, Indonesia, India, Japan, the Philippines, South Korea and the United Kingdom. Figure 3 shows the breakdown of the number of tourist arrivals to Malaysia by country in 2000-2014.

Thus, this study focuses on nine countries which are the major contributors to the Malaysian tourism market listed in the Annual Tourism Report by the Malaysian Tourism Promotion Board. The list of countries is Australia, Brunei, China, Philippines, India, Indonesia, Japan, South Korea and the United Kingdom. In this study, tourist arrivals from Singapore and Thailand were not taken into account as Singapore and Thailand frequently came to Malaysia due to the price of oil and the price of Malaysian goods and the Malaysian currency was cheaper than their country. Therefore, nine elected countries are reviewed based on their interests as tourists or representatives based on their specific goals. For example, Indonesia is the main contributor to the tourist arrivals to Malaysia followed by Brunei and the Philippines. These are the three ASEAN countries. While, the markets from East Asia are tourists from China, Japan and South Korea. While India is an important market from South Asia, Australia is an important marketplace of Australasia, and the United Kingdom is an important market from Europe.

Table.1 Arrivals and Revenues of Tourism Malaysia (1990-2014)

Year	Tourist arrivals	Growth (%)	Total income (RM mil)	Growth (%)
1990	7,476,722	+53.64	4500	+60.56
1991	5,847,231	-21.79	4300	-4.44
1992	6,016,209	+2.89	4595	+6.87
1993	6,503,860	+8.11	5065	+10.24
1994	7,197,229	+10.66	8298	+63.81
1995	7,468,749	+3.77	9174	+10.56
1996	7,138,000	-4.43	10354	+12.85
1997	6,211,000	-12.99	9699	-6.33
1998	5,551,000	-10.63	8580	-11.54
1999	7,931,000	+42.88	3450	+56.76
2000	10,222,000	+28.89	7335	+28.89
2001	2,775,000	+24.98	4222	+39.72
2002	13,292,000	+4.05	2571	+6.44
2003	10,577,00	-20.43	21291	-17.42
2004	15,703,000	+48.46	29651	+39.27
2005	16,431,000	+4.64	32054	+8.11
2006	17,547,000	+6.79	36343	+13.38
2007	20,973,000	+19.52	53400	+46.93
2008	22,052,000	+5.14	49600	-7.12
2009	23,646,000	+7.23	46100	-7.06
2010	24,580,000	+3.95	56500	+22.56
2011	24,710,000	+0.53	58300	+3.19
2012	25,030,000	+1.30	60600	+3.95
2013	25,720,000	+2.76	65400	+7.92
2014	27,440,000	+6.69	72000	+10.09

Source: Annual Report of the Malaysian Promotion Board Statistics, 1990-2014.

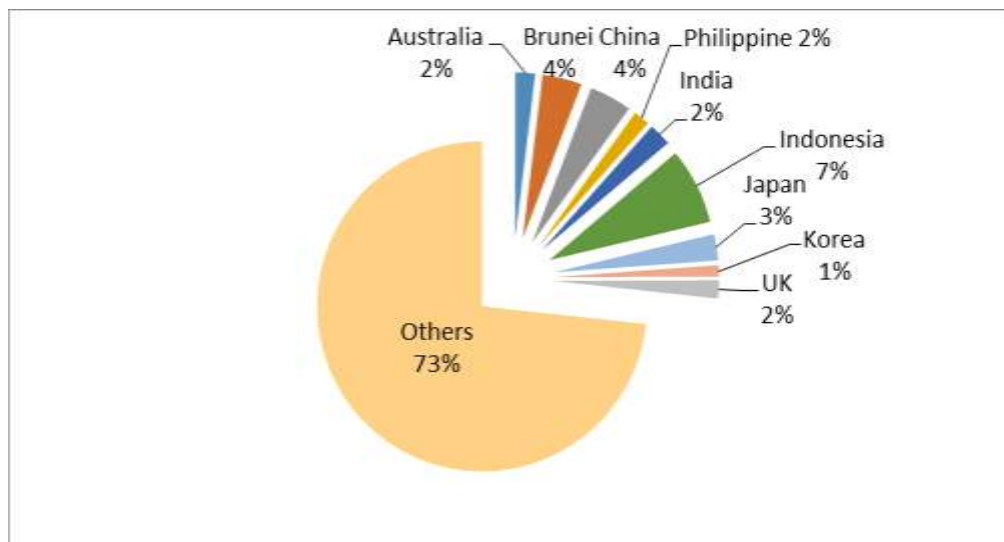


Figure 3. Tourist Arrivals to Malaysia by countries (1990-2014)

This study aims to forecast tourism demand for nine major tourism markets in Malaysia. Forecasting tourism demand in Malaysia is important as it helps generate income to the country and the findings can help the government to formulate better tourism plans and enhance the tourism promotion. The timeframe of the study dates from 1990 to 2014 while data in 2015 are kept for comparison in predictions and to test the accuracy of the model. Most data on tourism demand is non-stationary. However, the non-stationary state of the data is often negligible by many researchers in the field of tourism. Therefore, this study takes into account the non-stationary state of the data and also models the tourism demand especially the tourist arrivals using Box-Jenkins time series method such as the Autoregressive Integrated Moving Average (ARIMA) Model and Seasonal Autoregressive Integrated Moving Average (SARIMA) Model as well as Singular Spectrum Analysis (SSA). Next, this study forecasts the foreign tourist arrivals to Malaysia using Box-Jenkins time series model and non-linear approach that is the Singular Spectrum Analysis (SSA) model as well as compares Box-Jenkins time series model and SSA model to determine the best model.

MATERIALS AND METHODS

Over the last four decades, the time series model has been widely used for tourism by using the Autoregressive Integrated Moving Average model (ARIMA) proposed by Box and Jenkins (1976). In the 2000s, ARIMA was modified to the Seasonal Autoregressive Integrated Moving Average model (SARIMA). For the past few years, SARIMA is a popular time series forecasting technique due to the dominant feature of the tourism industry being seasonal in which decision makers are keen on seasonal variations in tourism demand. Performance forecasting of ARIMA and SARIMA models presents conflicting evidence in empirical studies. For example, Goh and Law (2002) suggest that the SARIMA model addresses eight other time series methods, while the ARIMA model considers average performance for all forecasting models. However, Smeral and Wuger (2005) find that the ARIMA and SARIMA models cannot cope with the Naive 1 model.

According to Witt et al. (1995), the Naive 1 model is the simplest and most often used model to generate more accurate model forecasting for the next year compared to other more sophisticated models. However, the performance of the Naive 1 model cannot make the right decision when there is a long-term structural change and prediction (Witt et al. 1995). Therefore, Model Naive 2 is introduced and widely used when there is continuous flow in data. Most of the tourism demand undertaken by previous researchers use linear techniques rather than nonlinear techniques. Among the linear techniques used include artificial neural networks, solid polynomial models, fuzzy clock series, GARCH models, transitioned parameters and varying times, arc learning models, time series structure models and nonlinear sine wave models (Song and Li 2008).

In recent years the technique known as single-spectrum analysis is a powerful technique developed in the field of time series analysis. Single Spectrum Analysis (SSA) is one of the nonparametric methods of time series analysis and projections that consist of many different but interconnected methods. There are several books devoted to SSA (Elsner & Tsonis 1996; Golyandina et al.,

2001; Golyandina & Zhigljavsky, 2013) as well as many relevant papers for SSA theory and especially for various SSA applications (see Golyandina & Zhigljavsky 2013). The scope of use of SSA is very broad, rather than parsing time series, filtering noise and rebuilding a less time-alarming time series.

According to Prideaux et al. (2003) and Schwartz and Cohen (2004), in the last 121 studies there are 2000 review of empirical studies. In 121 studies there were 72 using the time series method for the tourism demand model. In terms of integration and integration of tourism forecasts, according to Wong et al. (2007) shows that by combining various linear predictions, the risk of failure in tourism forecasts is diminishing. A variety of combined forecasts have included a long-term prediction of predictions and short-term forecasts combined with predictions of quantitative methods and qualitative methods by Song, Gao and Lin (2013), and a combination of linear and non-linear predictions by Chen (2013). Chen (2013) uses a combination of non-linear methods such as Artificial Neural Network (ANN) and Support Vector Regression (SVR).

Song and Li (2008) examine tourism demand forecasting using several non-linear methods. Non-linear methods are the Artificial Neural Network (ANN) seen in the study of Chen, Lai and Yeh (2012), Shahrabi, Hadavandi and Asadi (2013), and Claveria and Torra (2014). However, this ANN method is criticized because large data sets are needed for the model to provide accurate results and time taken for follow-up procedure training. In 2014, Claveria and Torra evaluated predictive performance for ANN and STAR model for Catalonia with the ARIMA model and found that the ARIMA model performed better, especially in short-term forecasts.

Wong (2007) carried out four types of models namely the Autoregressive Integrated Moving Average (ARIMA), Autoregressive Distributed Lag Model (ADLM), Error Correction Model (ECM) and Vector Autoregression (VAR) model to predict tourism in Hong Kong by a population of ten countries. The results show that prediction combinations do not always recognize the best single predictions. However, the model's empirical combination can reduce the risk of failure in practical predictions. Goh and Law (2002) use tourist arrivals data in Hong Kong to compare SARIMA models and MARIMA models with eight different time series models and predictive results for those models give better results using the SARIMA and MARIMA models.

The SSA model is the breakdown of time series into trends, seasonal components and noise. SSA is a different model and has the advantage over traditional time series methods where this time series model consists of noise and SSA aimed at refining noise in the forecasts of tourism demand. This SSA model was introduced in the tourism sector by Beneki et al. (2011) through sign-up extraction and UK tourist income forecasts. The SSA method is known as a powerful method, non-parametrical time series analysis and forecasting techniques. One of the key features is to use the most basic method by making optimum choice in vector SSA.

Hassani (2007) reviewed the performance of SSA techniques using the monthly death data series of mortality accidents in the United States. The forecasting results of the study have been compared with the SARIMA model, ARAR algorithm and Holt-Winter algorithm. The results show that the SSA model provides better predictive results than other methods. Hassani et al. (2015) also reviews SSA forecasts using monthly data for US tourist arrivals during the period 1996 to 2012. SSA's prediction method compared to other forecasting methods including ARIMA, exponential smoothing (ES) and neural network (ANN). The results of the study show that the SSA method yields better forecasts than alternative methods in the forecast of tourist arrivals to the United States.

Therefore, the Box-Jenkins time series model (ARIMA and SARIMA) and Singular Spectrum Analysis are more accurate models to be used for large sample. The Singular Spectrum Analysis is an alternative and extended research method that is performed to compare the performance of a Singular Spectrum Analysis with Box-Jenkins time series forecasting technique.

A. Box-Jenkins Time Series Method

Box-Jenkins method is one of the univariate methods used for time series economic forecasting. This method is able to determine the best model by using a systematic statistical procedure. This model uses variable values based on past observations and random errors to forecast future value of variables. In general, ARIMA and SARIMA series are the extension models of ARMA that include more realistic dynamics. ARIMA time series can be modelled as a combination of past values and involves parameters (p, d, q) where p is the number of autoregression parameter, d is the difference parameter, while q is the moving average parameter. There are four (4) forecasting processes using ARIMA as in Figure 4 (Newton, 1988):

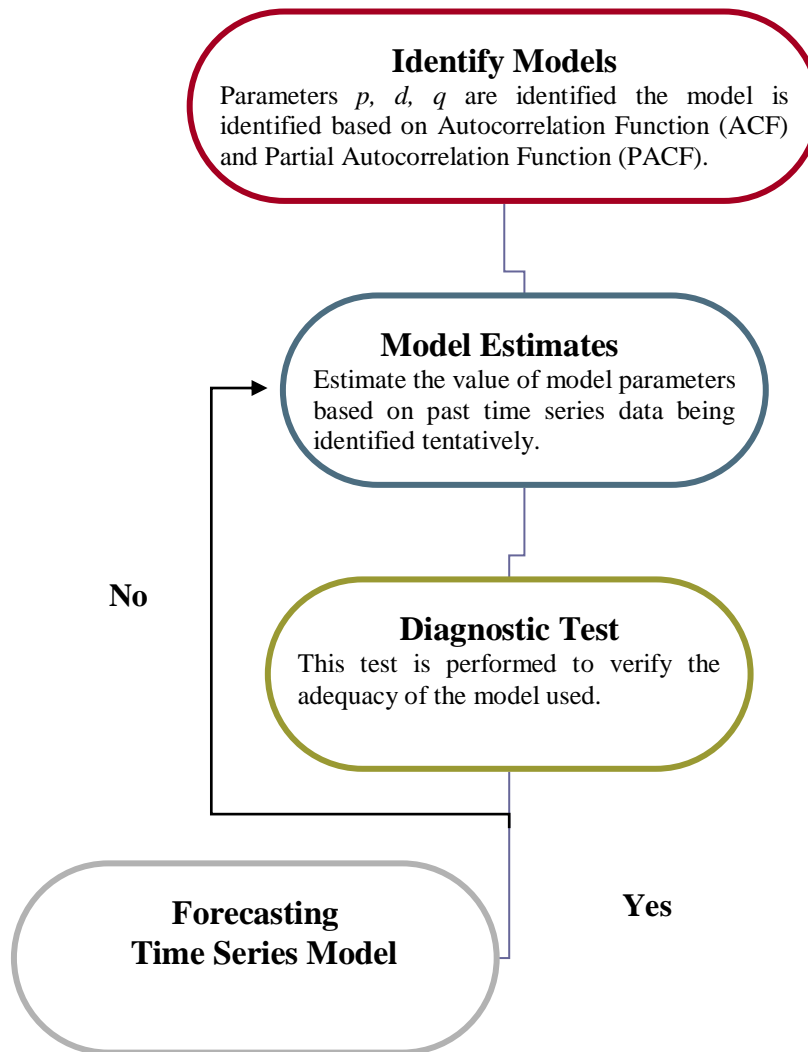


Figure 4: Box-Jenkins Time Series Methodology

The best model will be used to predict the value of the future series. According to Bowerman and Connell (1990), the classic Box Jenkins model can describe the data well. Therefore, the data used must be stationary. The data is said to be stationary in the time series, when the trend is not ascending and decreasing, but has constant mean and variance.

The ARIMA and SARIMA models usually are not stationary and has seasonal data. These model need to re-modelling by removing the source of variation which is not stationary and usually being done by making a difference in the series. When x_t is not stationary, the ARMA model is on a w_t series constructing on the decision is determined by the difference in the series (generally $d = 1$); Therefore, the ARIMA model is ARMA model defined in the original process of the difference d :

$$\Phi(B)\Delta^d X_t = \theta(B)\alpha_t$$

Whereby $\Phi(B)\Delta^d X_t$ is known as the general autoregressive operator and $\Delta^d X_t$ is a stationarity quantity through the difference. While seasonal ARIMA or SARIMA is a series of data that having seasonal components repeating each observation L . For monthly data, $L = 12$ which contains 12 months in a year while for quarterly data, $L = 4$. Therefore, the SARIMA model is

$$\Phi(B)\Delta^d X_t = \theta(B)\alpha_t$$

which α_t is;

$$\Phi(B^s)\Delta_s^D\alpha_t = \Theta(B^s)\alpha_t$$

Therefore, $\Phi(B)_s\Phi(B^s)\Delta_s^D\Delta^dX_t = \theta(B)_s\Theta(B^s)\alpha_t$, and can be written as:

$$X_t \square ARIMA(p, d, q) \times (P, D, Q)_{12}$$

In order to determine the models (p, d, q) , ACF and PACF are used to observe the form of ACF and PACF coefficients that are declining to zero. The p -value is for autoregression (AR), the d -value is the number of differences and the q -value is for moving average (MA). The p -value is determined when a stationary PACF plot is highlighted. If the partial correlation value is truncated on k intervals and in a stationary state after that; the intervals value is used. In the case of the q -value, the stationary ACF plot is being highlighted. Similar to determining the p parameter, the truncated ACF intervals value which later in a stationary state is obtained. The number of differences is performed on the d -value; if the data is non-stationary.

If both ACF and PACF plots are truncated, observations are performed to see which plots are most significantly truncated and declined (Bowerman et al., 1987). The following is the equation of non-seasonal Box-Jenkins and autoregression moving average models in order (p, q) , or ARMA (p, q) :

$$y_t = \phi_1 y_{t-1} + \dots + \phi_p y_{t-p} + \alpha_t - \theta_1 \alpha_{t-1} - \dots - \theta_q \alpha_{t-q}$$

Whereas the equation for the seasonal Box-Jenkins model, the seasonal autoregression model of p order is as follows:

$$z_t = \delta + \theta_{1,L} z_{t-L} + \theta_{2,L} z_{t-2L} + \dots + \theta_{P,L} z_{t-PL} + \alpha_t$$

Apart from that, the seasonal moving average model of q order is as the following equation;

$$z_t = \delta + \alpha_t - \theta_{1,L} \alpha_{t-L} - \theta_{2,L} \alpha_{t-2L} - \dots - \theta_{Q,L} \alpha_{t-QL}$$

Lastly, the non-seasonal and seasonal levels are combined to obtain a tentative model. Once a suitable model is obtained, the parameters are identified. Based on the smallest value of Akaike Information Criteria (AIC), the best model can be determined. Therefore, the estimation of significant parameters is included in the ARIMA or SARIMA equations.

The diagnostic test is performed to test the suitability of the model before performing the forecasting. The best way to test the entire Box-Jenkins model is by analysing the residue of the model. Before estimating the equation of Box-Jenkins model, the researcher should examine the residue to prove the serial correlation. One of the methods that can be used is Ljung Box statistics. Therefore, the model specification is sufficient if the p -value for Ljung Box statistics is greater than the significance level of 0.05.

Forecasting can be performed when an appropriate model is identified. The forecasting value is then compared with the real data. According to Lewis (1992), if the Mean Absolute Percentage Error (MAPE) is less than 10%, it can be considered as a very accurate prediction, so the forecasting model is very good.

B. Singular Spectrum Analysis (SSA)

Singular Spectrum Analysis (SSA) is a powerful time-series analysis technique that combines classical elements of time series, multivariate statistics, multivariate geometry, dynamic system and signal processing (Hassani, 2007).

The purpose of the SSA is to decompose the original time series into a smaller amount of free and assessable components such as different slow trends, oscillating and less structured noise components. SSA is a very useful method to solve the following problems:

- i. Finding trends for resolution differences;
- ii. Smoothing;
- iii. Seasonal component extraction;
- iv. Simultaneous extraction of cycles with small and large periods;
- v. Extraction of periodicities with various amplitudes;

- vi. Simultaneous extraction of complex trends and periodicities;
- vii. Finding structure in short time series;
- viii. Change-point detection.

The main purpose of the SSA is to decompose the original time series into a several time series known as trends, oscillation components or partial oscillation components which are known as amplitude modulation or noise. Then, the original time series are reconstructed. The SSA method consists of two (2) complementary stages; decomposition and reconstruction. Both include two (2) separate steps. The first stage involves the decomposition of time series and the second stage involves the reconstruction of the original time series and using a series of construction without sound for new data point forecasting. Monthly data of tourist arrivals from the country of origin to Malaysia from 1990 to 2014 are used. The data for SSA model is transformed to facilitate the comparison between models in forecasting. This is because the value of forecasting precision for the SSA model becomes larger and cannot be compared. In addition, the researcher uses the SSA method on this data with the purpose to illustrate the ability of the SSA method in the decomposition of trend, oscillation, noise, and forecast. The results are obtained through the R software.

The SSA forecasting method is divided into two namely; recurrence predictions and vector predictions. In repeated predictions, the diaphragm is applied to obtain a reconstructed series and then using linear recurrence relations (LRR) method whereby LRR are associated with autoregressive (AR) models. Whereas, in vector predictions, the SSA steps exchanged. This shows that vector predictions are more stable but have greater computation costs than recurrence predictions.

If the time series components are separated from noise and use linear recurrence function; LRF, then the two predictions coincide and provide precise continuity. In isolation, the study provides different continuity. Since LRF provides the basis for recurrence predictions, this study uses the LRR method further. It is very useful in selecting parameters and understanding the forecasting behaviour. The SSA forecast can be used for a time series that meets the linear recurrence formula (LRF).

C. Measuring Model Precision

In order to evaluate the accuracy of time series forecasting and non-linear model, the *t*-value is set as *t* = 12 in the first year. This is due to the fact that this study uses monthly data. For ex-post forecasting process of tourist arrivals to Malaysia to forecast the monthly data for 2015, the process uses the monthly data from 1990 to 2014.

The forecasting accuracy is assessed on various methods to understand the precision of the forecasting of the model. In this study, the researcher uses the MAPE method which is the most suitable to be used in assessing the performance of the tourism demand model forecasting. This measurement is defined by taking y_t as the original data or observation and \hat{y}_t is the forecasted data by the researcher at a given time (*t*) and the number of forecasts (*i*).

$$MAPE = \frac{1}{T} \sum_{i=1}^T \frac{|y_t - \hat{y}_t|}{|y_t|} \times 100\%$$

MAPE is a relative performance measure used to measure the forecasting performance. Table 2 shows the accuracy level of the MAPE test. Forecasting results with a MAPE value of <10% can be assumed generating a very precise forecasting (Lewis, 1982).

Table 2: Accuracy level of MAPE test

MAPE Value	Forecasting Precision
MAPE < 10%	Very Precise
10% ≤ MAPE ≤ 20%	Precise
20% ≤ MAPE ≤ 50%	Satisfactory
MAPE > 50%	Unacceptable

Source: Lewis (1982)

RESULTS AND DISCUSSION

A. Box-Jenkins Method

i. Model Recognition

Study has found that the autocorrelation plot (ACF) for the data is seasonal. Therefore, the logarithm transformation is used for Box-Jenkins analysis against the nine (9) major markets for their respective origin countries. The attribute of the stationary data is detected using the unit root test. In this study, the selected unit root test is Augmented Dickey-Fuller (ADF) test. Results of logarithm data for each country are tested using ADF test to determine the stationarity of the data. Table 3 shows the ADF test results on nine (9) major markets upon performing the logarithm of the tourist arrivals data in Malaysia.

In Table 3, six (6) out of the nine (9) major markets, namely Australia, Brunei, the Philippines, India, Indonesia and South Korea, have the *t*-statistics greater than the critical value at the significance level of $\alpha = 0.05$ which is -3.4524. In fact, the statistical probability value has a greater value at the significance level of 0.05. Thus, the null hypothesis that states the variable is stationary is failed to be rejected. Brunei, China, and the United Kingdom have a smaller statistical value than the critical value at the base of the logarithm, indeed the *p*-value has a value smaller than the significance level of 0.05. Thus, the null hypothesis is rejected. However, for the said three (3) countries, it is found that the autocorrelation plot (ACF) is slowly decreasing to zero which indicates the data is non-stationary. Thus, the data of tourist arrivals to Malaysia for nine (9) major markets is being differentiated and undergone the ADF test again.

Table 3: ADF testing at the base level for nine (9) major tourism markets in Malaysia

Tourism Arrivals to Malaysia Variable, LTA_t	<i>t</i> -Statistics	<i>p</i> -Value
Australia	-2.8915	0.2005
Brunei	-2.6927	0.2843
China	-3.5786*	0.0355
Philippines	-2.7056	0.2789
India	-2.8845	0.2035
Indonesia	-2.8603	0.2137
Japan	-3.893*	0.0148
South Korea	-2.2651	0.4646
United Kingdom	-3.497*	0.0434

* showing the estimations are significant at the significance level of 0.05

Table 4: ADF test results at the first differential level for nine (9) major tourism markets in Malaysia

Tourism Arrival of travel To Malaysia Variable, ΔLTA_t	<i>t</i> -Statistics	<i>p</i> -Value
Australia	-10.895	0.01
Brunei	-9.8326	0.01
China	-9.9588	0.01
Philippines	-10.383	0.01
India	-10.114	0.01
Indonesia	-11.037	0.01
Japan	-8.65	0.01
South Korea	-11.14	0.01
United Kingdom	-9.3922	0.01

Referring to Table 4, the critical value of the ADF test at the significance level of 0.05 is -3.4524. All the values shown in Table 4 indicate that all countries have a t -statistic value lesser than the critical value at the significance level of $\alpha = 0.05$. In addition, the p -value is lesser than the significance level. Thus, the null hypothesis that states the data of tourist arrivals to Malaysia is non-stationary is rejected. It can be concluded that the data of tourist arrivals to Malaysia for the nine (9) major markets are at the first differential level, $d = 1$.

i. Model Construction and Parameter Estimation

Model recognition can be carried out using the Autocorrelation plot (ACF) and the partial Autocorrelation plot (PACF) through the values being truncated and decreasing towards zero. In this study, the best-returned model based on the ACF plot that stands out at certain lag and the FAKS plot truncates rapidly towards zero. Thus, there are nine (9) best-known Box-Jenkins models recognized through autoregression and moving average significant values. The insignificant values of Autoregression (AR) and Moving Average (MA) are removed to ensure the ARIMA model is the best model and to obtain a precise forecasting model.

In order to further reinforce the ARIMA model selection, the model accuracy is tested using RMSE, MAE and ME by looking at the lowest value based on the lowest error for ME. Table 5 shows the accuracy of the ARIMA model using RMSE, MAE and ME. Based on the table, the RMSE error has an error value of about 0.2 to 0.5. The MAE error has the error value of about 0.1 to 0.3. Therefore, the model is selected as the suitable model to forecast the tourism arrivals in Malaysia.

Table 5: The Accuracy of ARIMA model

Country	ARIMA	ME	RMSE	MAE
Australia	(0,0,0)(0,0,1) ₁₂	0.0006	0.4428	0.3193
Brunei	(0,1,5)	-0.0039	0.5018	0.3705
China	(0,1,4)	-0.0030	0.4502	0.3373
Philippines	(1,1,1)(2,0,1) ₁₂	-0.0102	0.5066	0.3657
India	(0,1,2)	-0.0038	0.5288	0.3848
Indonesia	(2,1,3)	-0.0024	0.4761	0.3484
Japan	(1,1,2)(1,0,0) ₁₂	-0.0019	0.2603	0.1907
South Korea	(3,2,1)	-0.0002	0.5010	0.3818
United Kingdom	(0,1,3)(1,0,0) ₁₂	0.0003	0.3469	0.2445

Table 6 shows the equations of estimates for ARIMA model time series for tourism arrivals of nine (9) major markets.

Table 6: Equations of estimates for ARIMA model

Country	ARIMA Model
Australia	$LTA_t = 0.0018 + \alpha_t - 0.8490\alpha_{t-1} + 0.1351\alpha_{t-12}$
Brunei	$LTA_t = 0.0048 + \alpha_t - 0.7218\alpha_{t-1} + 0.0010\alpha_{t-2} - 0.1361\alpha_{t-3} - 0.1683\alpha_{t-4} + 0.1773\alpha_{t-5}$
China	$LTA_t = 0.0032 + \alpha_t - 0.7402\alpha_{t-1} + 0.045\alpha_{t-2} - 0.2721\alpha_{t-3} + 0.0873\alpha_{t-4}$
Philippines	$LTA_t = 0.0026 + 0.1583LTA_{t-1} + 0.5250LTA_{t-12} - 0.1476LTA_{t-24} + \alpha_t - 0.8648\alpha_{t-1} - 0.8059\alpha_{t-12}$
India	$LTA_t = 0.0023 + \alpha_t - 0.7033\alpha_{t-1} - 0.1631\alpha_{t-2}$
Indonesia	$LTA_t = 0.0022 + 1.1012LTA_{t-1} - 0.8176LTA_{t-2} + \alpha_t - 1.8440\alpha_{t-1} + 1.5399\alpha_{t-2} - 0.5831\alpha_{t-3}$
Japan	$LTA_t = 0.0001 + 0.8240LTA_{t-1} - 0.1972LTA_{t-12} + \alpha_t - 1.4631\alpha_{t-1} + 0.4631\alpha_{t-2}$
South Korea	$LTA_t = 0.0030 + 0.5124LTA_{t-1} - 0.060LTA_{t-2} - 0.3083LTA_{t-3} + \alpha_t - 1.2673\alpha_{t-1} + 0.4837\alpha_{t-2}$
United Kingdom	$LTA_t - 0.0022 + 0.1089LTA_{t-12} + \alpha_t - 0.7219\alpha_{t-1} - 0.0276\alpha_{t-2} - 0.1380\alpha_{t-3}$

ii. Diagnostic Inspection

Diagnostic inspection is conducted to determine the effectiveness of the model obtained for the tourism arrivals in Malaysia for the nine (9) major markets on each respective country of origin. Ljung Box test is used to see the suitability of the model in forecasting the tourism arrivals in Malaysia.

In this study, the suitable model is studied using the statistical test at lag 12 with a *p*-value of 0.05 as the forecasting is performed for a duration of 12 months for those who arrive in January 2015 to December 2015. If the *p*-value is lesser than the significance level of 0.05, then the null hypothesis that states the ARIMA model is suitable is rejected. Therefore, the ARIMA model used for future forecasting is suitable.

iii. Box-Jenkins Forecasting Model

After modelling, parameter estimation and diagnostic inspection are performed. It is found that the Box-Jenkins model is suitable to be used in this study. The model describes the forecasting of tourism arrivals in Malaysia for 2015 by using the data from January 1990 to December 2014 as a density test. The model is selected based on significant values of Autoregression (AR) and Moving Average (MA) and the lowest error compared to other models and fulfilled the Ljung Box test conditions.

In this study, the forecasting precision is measured using MAPE. The forecasting results of tourist arrivals to Malaysia in 2015 for Box-Jenkins model are illustrated in Table 7

Table 7: MAPE results for Box-Jenkins model for nine (9) major markets against their respective countries of origin

Country	Box-Jenkins Model	Real Data	Forecasting Data	Error	MAPE (%)
Australia	$ARIMA(0,1,1)(0,0,1)_{12}$	39 603	46 128	-6525	1.37
Brunei	$ARIMA(0,1,5)$	93 433	103 691	-10258	0.91
China	$ARIMA(0,1,4)$	137 999	127 601	10398	0.63
Philippines	$ARIMA(1,1,1)(2,0,1)_{12}$	45 859	48 090	-2231	0.41
India	$ARIMA(0,1,2)$	59 675	61 800	-5125	0.72
Indonesia	$ARIMA(2,1,3)$	229 808	238 590	-8782	0.32
Japan	$ARIMA(1,1,2)(1,0,0)_{12}$	39 835	41 151	-1316	0.28
South Korea	$ARIMA(3,1,2)$	34 286	31 388	2989	0.73
United Kingdom	$ARIMA(0,1,3)(1,0,0)_{12}$	33 218	35 656	-2438	0.61

According to Lewis (1982), the MAPE values that are lesser than 10% show this model is very precise in which it is suitable for future forecasting against the model. Based on Table 6, it is found that all Box-Jenkins models show MAPE results that are lesser than 10% in which they are suitable to be used to generate forecasting. This is especially for the Philippines, Indonesia, and Japan that have MAPE values close to 0%. This proves that the lower MAPE value generates smaller percentage errors and the forecasting derived from the model is more precise. This shows that all models are suitable to be used for future forecasting of tourism arrivals in Malaysia.

B. Singular Spectrum Analysis (SSA)

Singular Spectrum Analysis is used to extract trend, seasonal, noise, and forecasting. Hence, the analysis uses the monthly data from 1990 to 2014 against the nine (9) major markets that greatly influence the tourist arrivals in Malaysia. In general, the purpose of this analysis is to decompose the original series into a number of series, each of which is known as a trend, seasonal components, and noise. Next, the original series is rebuilt. This analysis is divided into two (2) levels in which each level has two (2) separate ways. In this study, the first stage involves the decomposition of the tourist arrivals series in Malaysia and the second level involves the reconstruction of the tourist arrival series in Malaysia; the series is used to forecast future tourist arrivals in

Malaysia without noise. The data is transformed into logs so that it is easy to compare the forecasting of all two models. Figure 1 represents the tourist arrivals in Malaysia for the nine (9) major markets against their respective countries of origin.

Therefore, the levels as described by Hassani (2007), there are two (2) stages of the application stage. The application of a single parameter is the length of the window, L . The periodic components are used in this study due to seasonal factors in the time series data of tourist arrivals in Malaysia. In order to obtain a good single-parameter application based on the seasonal components, the length of the window is separated with the said proportionality, therefore, $L = 144$. So, based on the length of the window and trajectory of the SVD matrix is (144×144) , the study has 144 eigentriples for this decomposition step. The SVD step is shown through the optimum components of the first 12 eigentriples of nine (9) major tourism markets in Malaysia for their respective countries of origin that show the percentage of optimum components decreases proving that the coordinates of optimum components are almost constant.

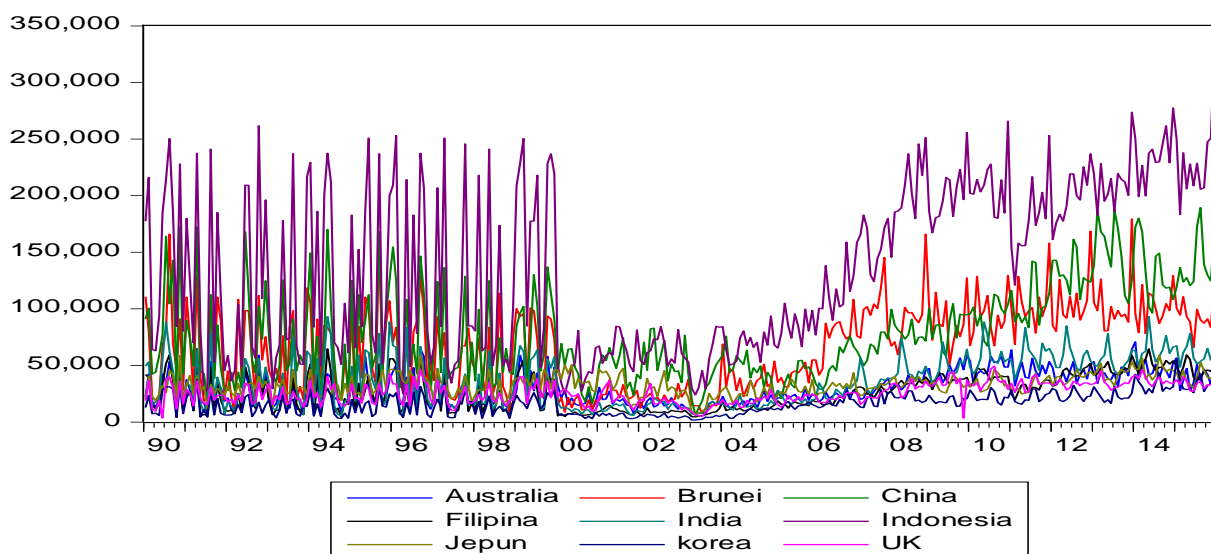


Figure 5. Nine (9) major markets influencing the tourism arrivals in Malaysia against their respective countries of origin

Additionally, proper groupings enable the process of obtaining the trends, harmony components and noise, and even increasing the ability to develop the appropriate model. As such, the additional information is known as a bridge between the decomposition and reconstruction steps. Each harmony component with different frequency yields two (2) eigentriples approaching a single value. The original noise series, as a condition, yields a single value of a sequence that is decreasing. Thus, $L = 12$ plot is used. All countries confirm that the first eigentriples correspond to the trend. Meanwhile, other eigenes contain high frequencies. Therefore, it is not related to the trend.

After identifying the trend, the pair by pair analysis of a single vector of scatter plot is used to produce visuals to identify eigentriples corresponding to seasonal components series with a condition that the seasonal components are separated with signal components. Table 8 represents the trend components and eigentriples seasonal components against the tourism arrivals series in Malaysia for nine (9) major markets:

Table 8: Trend components and eigentriples seasonal components against nine (9) selected major markets

Country	Eigen triples
Australia	1-8
Brunei	1-6
China	1-12
Philippines	1, 2-10
India	1, 5-12
Indonesia	1-4

Japan	1, 7-12
South Korea	1,5-8
United Kingdom	1, 3-6

The eigentriples listed in Table 7 correspond to the periods of 12, 6, 2.4, 3 and 4 produced by the seasonal components and are clearly explained by the periodogram analysis.

The original series of periodogram analysis and eigenvectors assist in creating proper groupings. The periodogram is to study the frequency of eigentriples that coincide with the frequency of the original series. If the periodogram eigenvector has a sharp pitch around a few frequencies, then the eigentriples are related to the signal component. Generally, eigentriples value is found to be lesser than 12 which indicates the monthly period. If pairing eigentriples have a period value of more than 12, it means they are not interpreted for monthly data and they are considered as noise, *n*. The main concept of the SSA model is separability that describes how the differences in components are separated from each other. In this study, the separability of each major market shows the *w*-correlation for the reconstruction of 30 components from black and white that correspond to the absolute value of correlation from 0-1.

i. SSA Forecasting

Forecasting in a single spectral analysis has two (2) methods, namely recurrent method and vector method. In this study, it is found that the recurrent method is closer to the actual value compared to the vector method. Hence, the forecasting of tourism arrivals to Malaysia for 2015 using the data from 1990 to 2014 is plotted using the recurrent method. The analysis uses MAPE to measure the forecasting performance. The MAPE forecasting results are shown in Table 9. The MAPE value is in percentage.

Table 9: MAPE Results of Singular Spectrum Analysis

Country	Real Data	Forecasting Data	Error	MAPE Recurrent Model (%)
Australia	39 603	51 706	-12 103	2.55
Brunei	93 433	67 846	25 587	2.28
China	137 999	136 171	1 828	0.11
Philippines	45 859	86 393	-40 534	7.37
India	59 675	59 774	-100	0.01
Indonesia	229 808	367 385	-137 577	4.99
Japan	39 835	35 745	4 090	0.86
South Korea	34 286	48 614	-14 328	3.48
United Kingdom	33 218	26 814	6 404	1.61

Lewis (1982) states that the MAPE value <10% is a very precise model for the forecasting. Based on Table 8, all SSA models are found to yield MAPE values of less than 10%. Therefore, the SSA model is considered to be a very precise forecasting model. For the studied countries, the MAPE values nearing zero indicates a very small percentage error. Therefore, the model is said to be better and suitable to be used. This proves that the model is suitable to be used to forecast future tourism in Malaysia.

C. Box-Jenkins and Singular Spectrum Analysis Comparison

In this section, the comparison between Box-Jenkins and SSA models is discussed. Therefore, the MAPE values are used to compare the two (2) models studied. The lowest MAPE value shows a more suitable and more accurate forecasting model. Table 10 shows the results of MAPE values for the two (2) models.

Table 10: Results of MAPE values of Box-Jenkins and SSA models

Country	Box-Jenkins Model	SSA Model
Australia	1.37	2.55
Brunei	0.91	2.28
China	0.63	0.11

Philippines	0.41	7.37
India	0.72	0.01
Indonesia	0.32	4.99
Japan	0.28	0.86
South Korea	0.73	3.48
United Kingdom	0.61	1.61

Based on Table 10, generally, the forecasting outcome of both models have MAPE values of less than 10% and can be considered to generate a very precise forecasting. Overall, the Box-Jenkins model has the lower MAPE value than the SSA model. Additionally, the MAPE value nearing zero proves that the model's forecasting generates a smaller percentage error, especially for the Philippines, Indonesia, and Japan. Therefore, the lower MAPE value proves that the forecasting derived from the model is more precise. In conclusion, the best model is Box-Jenkins in which it takes into account the seasonal factors. While the SSA does not take into account the seasonal factors, thus, the model accuracy can be proven.

CONCLUSIONS

The tourism industry plays an important role and makes one of the most important contributors to the Malaysian economy. The tourism industry is one of the sources of income generated by foreign exchange inflows. Malaysia is well known for its halal destinations and focuses on foreign tourists, especially the Middle East and Southeast Asia. The purpose of this study is to determine the best forecasting method and suitable for use in tourism demand in Malaysia. Overall, this study concludes that the Box Jenkins time series model is best compared to the non-linear SSA model. The result of MAPE value below 1% proves that the Box Jenkins model generates smaller percentage of errors and the predictions derived from the model are very accurate.

The best empirical prediction method used is the Box Jenkins model for most countries such as Australia, Brunei, Philippines, India, Indonesia, South Korea, and the United Kingdom when compared to the SSA model. While for China and Japan shows that the SSA model is the best model. In terms of planning, this model should be a major research model in the study of international tourism demand. Since this model has been set up, the prediction results process can help in the development and investment strategies of the tourism industry in the future. In overall, both the Box Jenkins and SSA models have MAPE forecasting values of less than 10%. It can be concluded that both models are appropriated to predict future tourism demand in Malaysia.

As a basic implication, efforts from all parties are needed to enhance the tourism industry in Malaysia. The government should coordinate and monitor the quality of their services and tourism products. Sufficient, comfortable, economical and affordable service attracts more foreign tourists to Malaysia. Additionally, the government can also reduce the goods purchase tax, immigration regulations need to be relaxed and increase the foreign exchange service centre. This is because foreign tourists can shop in Malaysia with moderate spending. In fact, the diversity of services and tourism products should be provided to meet different travelers' needs. At the same time, governments and Malaysians can join forces to organize tourism promotion campaigns both inside and outside the country. In fact, Malaysians can together help to enhance the tourism field with a friendly and helpful attitude towards tourists. The findings clearly show that this factor is very significant in attracting foreign tourists to Malaysia.

ACKNOWLEDGMENTS

This research grant is under the UKM-GUP-2016-052 grant.

REFERENCES

1. Beneki, C., Eeckels, B., & Leon, C. (2012). Signal extraction and forecasting of the UK tourism income time series: a Singular Spectrum Analysis approach. *Journal of Forecasting*, 31(5), 391-400.
2. Bowerman, B. L. & O'Connell, R.T. (1990). *Linear Statistical Models: An Applied Approach*. Boston: PWS- KENT Publishing Company.
3. Bowerman, B. L. & O'Connell, R.T. (1987). *Time Series and Forecasting: Unified Concepts and Computer Implementation*. (2nd. ed). Boston: PWS Publishers.

4. Box, G. E. P., & Jenkins, G. M. (1976). Time Series Analysis. Forecasting and Control (revised edition). San Francisco: Holden-Day.
5. Chen, C-F., Lai, M-C. & Yeh. C-C. (2012). Forecasting tourism demand based on empirical mode decomposition and neural networks. Knowledge-Based Systems, 26, 281-287.
6. Claveria, O., & Torra, S. (2014). Forecasting tourism demand to Catalonia: Neural networks vs. time series models. Economic Modelling, 36, 220-228.
7. Elsner, J.B. & Tsonis, A.A. (1996). Singular Spectrum Analysis. A new Tools in Time Series Analysis. Plenum Press
8. Goh, C., & Law, R. (2002). Modelling and Forecasting Tourism Demand for Arrivals with Stochastic Nonstationary Seasonality and Intervention. Tourism Management 23 (5), 499-510.
9. Golyandi, N., Zhigljavsky, A. (2013). Singular Spectrum Analysis for time series. Springer Briefs in Statistics. Springer
10. Golyandina, N., Nekrutkin, V., & Zhigljavsky, A. (2001). Analysis of Time Series Structure: SSA and related techniques, Chapman & Hall/CRC, New York-London
11. Hassani, H. (2007). Singular Spectrum Analysis: Methodology and Comparison, Journal of Data Science, 5(2), pp. 239-257.
12. Lewis, C.D. (1982). Industrial and Business Forecasting Method. London: Butterworths.
13. Newton, H.J., (1988). TIMESLAB: A Time Series Analysis Laboratory, Wadsworth & Brooks/Cole Publishing Company, California.
14. Official Portal of Ministry of Tourism and Culture Malaysia, May 2015.
15. Prideaux, B., Laws, E. & Faulkner, B. (2003). Events in Indo-nesia: Exploring the limits of formal tourism trend forecasting methods in complex crisis situations, Tourism Management, 24, 475-487.
16. Schwartz, Z., & Cohen, E. (2004). Subjective estimates of occupancy forecast uncertainty by hotel revenue managers. Journal of Travel & Tourism Marketing, 16, 59-66.
17. Shahrabi, J., Esmail H., & Shahrokh A. (2013). Developing a hybrid intelligent model for forecasting problems: Case study of tourism demand time series. Knowledge-Based Systems, 43, 112-122.
18. Smeral, E., & Wüger, M. (2005). Does complexity matter? Methods for improving forecasting accuracy in tourism: The case of Australia. Journal of Travel Research, 44, 100-110.
19. Song, H., & Li, G. (2008). Tourism Demand Modelling and Forecasting- A Review of Recent Research. Tourism Management, 29(2), 203-220
20. Song, H., Bastian Z. G., & Vera S. L. (2013). Combining statistical and judgemental forecasts via a web-based tourism demand forecasting system. International Journal of Forecasting, 29, 295-310.
21. Witt, C.A., Witt, S.F., & Wilson, N., (1994). Forecasting international tourist flows. Annals of Tourism Research 21 (3), 612-628.
22. Wong, K., Haiyann, F.K., Stephen F.W., & Doris C.W. (2007). Tourism Forecasting: To Combine or Not to Combine. Tourism Management. 28, 1068-1078.
23. World Tourism Organization. 2015. Yearbook of Tourism Statistics – 2015 Edition. Madrid: World Tourism Organization.

Cite this Article: Siti Syahirah Abdul Halim, Nora Muda (2020). Modeling and Forecasting of Tourism Demand in Malaysia. International Journal of Current Science Research and Review, 3(12), 230-244