INTERNATIONAL JOURNAL OF ACADEMIC RESEARCH IN BUSINESS & SOCIAL SCIENCES
Vol 8, Issue 16, (2018) E-ISSN: 2222-6990

Forecasting International Tourist Arrivals in Penang using Time Series Model

Shahirah Khairudin, Nursyatiella Ahmad, Azura Razali, Ahmad Zia Ul-Saufie Mohamad Japeri, Azila Binti Azmi
Universiti Teknologi MARA, Pulau Pinang Campus, Malaysia

To Link this Article: http://dx.doi.org/10.6007/IJARBSS/v8-i16/5117 DOI:10.6007/IJARBSS/v8-i16/5117

Published Date: 31 December 2018

Abstract

Tourism forecasting plays an important role for the future development of tourism industry to accelerate the economic growth. The appropriate forecast in tourism gives benefit to both public and private sectors as the information concerning the future tourism flows is important to tourism stakeholder. However, there is no appropriate forecasting model employed in Penang. Hence, the purpose of this study is to determine the best model in forecasting international tourist arrivals in Penang for 2016–2017. Secondary data on tourist arrival to Penang for 2010-2015 was obtained from Ministry of Tourism and Culture Malaysia. The data were analysed using software QM for Windows. Trend projection model and trend projection with seasonal effect model are used in this study. The accuracy of these two models are determined using mean absolute percentage error (MAPE) and the results show that trend projection with seasonal effect model outperformed trend projection model. MAPE results are between 7.7% and 33.6%. Therefore, trend projection with seasonal effect model will be proposed to be employed in forecasting international tourist arrivals in Penang. Adequate data is important in managing resources to avoid scarcity, or over spending and excessive waste in resources. Thus, this study will be beneficial to the local authority, industry players, as well as other tourism stakeholders in Penang.

Keywords: Tourism Industry, Forecasting Model, Trend Projection Model, Trend Projection with Seasonal Effect Model, MAPE

Introduction

The tourism industry is growing and emerging fast as it has become one of the forceful sectors of the global economy. In 2015, the international tourist arrivals in Malaysia achieved its best results even in the post of highest economic crisis (Costa, Montenegro, & Gomes, 2016). Tourists tend to have a disposable income that they are able to spend during their visits to different locations and countries (Kumar & Sharma, 2016). Hence, tourism contributes significantly to the economic growth of many countries and regions (Song, Li, Witt, & Athanasopoulos, 2010).

Tourism industry is a sector that can be more resilient and recovers rapidly (Jala, 2016). The statistic shows a significant drop of tourist arrivals in Malaysia for the year of 2015. On the other hand, it is reported that the industry indicates a sign of recovery in the first half of 2016 as it registers a hike of 3.7% arrivals compared to the same period in 2015 (Tourism Malaysia , 2016).

When Malaysia was embracing the declining of tourist arrivals in 2015, Penang was facing the opposite phenomena where the number of tourist arrivals kept on increasing. The difference in arrivals trend made it unsuitable for Penang to follow the overall Malaysia forecast. Moreover, forecasting tourist arrivals in Malaysia is still in its infancy stage.

Forecasting is very important in tourism industry. Accurate forecasting provides direct assistance to the state government and industry player to help them in making important decisions, avoiding waste and inefficiency of tourism resources, thus reducing the risk and uncertainty (Chen, Lai, & Yeh, 2011; Chen, Liang, Hong, & Gu, 2014).

The forecast on international tourist arrivals is very important to the industry player as well as the local government and authority. Tourism Malaysia needs this kind of information to craft a specific marketing plan to capture a market from each region around the globe. Industry player needs the data for decision making purposes, such as hotel chain expansion and opening of new retail branch. However, the forecast of international tourist arrivals done in Malaysia is as a whole and there is no specific forecast being done in Penang. The data of international tourist profiling and future prediction of international tourist arrivals for the upcoming three years are published in a booklet form and available for sale (Deputy Director of Tourism Malaysia Penang, 2016).

However, there is no published research conducted in developing tourist arrivals forecasting model in Penang. Therefore, to fulfil this gap, this study proposes the forecasting of international tourist arrivals in Penang. It is hoped that this study is able to assist the local authority in allocating resources and making important decisions; and useful for the industry player to predict their future market.

Literature Review

Forecasting is vital in tourism industry to aid a reasonable distribution of tourism resources (Chen, Liang, Hong, & Gu, 2014). The perishable nature of tourism product makes forecasting even more important especially for the player in this industry. Unoccupied hotel room and unsold flight ticket will be lost revenue. The forecast on tourist arrivals and demands enable the business to craft a better planning. For instance, the hotel management can plan the best time to hire additional staff and the airline business can plan which route to be added or terminated. Long-term and short-term forecasting serves multiple purposes from determining staff and other resources, to the investment in public infrastructure and equipment (Gunter & Onder, 2014).

Investment in public infrastructures and equipment, such as airports, highway, and public transport hub requires a long-term financial commitment from public finances. Insufficient tourism demand may result to loss in investment. Hence, the prediction of long-term demand for tourism related infrastructures often forms an important part of project evaluations (Song & Witt, 2006).

Tourist arrivals forecasting depends on two elements: tourist arrivals and forecasting. In the context of tourism, an arrival is a statistical unit measuring the volume (number) of tourists or visitors (Ma, Liu, Li, & Chen, 2015).

The forecasting method can be generally classified into two categories: qualitative and quantitative model (Render, Stair, Hanna, & Hale, 2015). Quantitative model can be divided into three categories: econometric models, time series models, and artificial intelligence techniques (Liang, 2014). The literature on forecasting tourism is huge with various types of empirical analysis but time series models are the most popular method used in forecasting research as many studies on tourist demand and forecasting carried out for the past three decades, thus most research use pure time series models (Loganathan, Nanthakumar, & Ibrahim, 2010; Ma, Liu, Li, & Chen, 2015).

It was found that in some cases, time series models were equally or even more accurate than the causal models. Time series models have been broadly used for tourism demand forecasting in the past four decades, with the dominance of ARIMA models proposed by Box and Jenkins (1970) (Claveria & Torra, 2013).

However, in some cases, on account of the presence of non-stationary and non-linearity in the data traditional statistical, the econometric models may produce poor forecasting performance (Shabri, 2016).

Forecasting Method Worldwide

Ma, Liu, Li, and Chen (2015) develop a time series model of monthly arrivals of Chinese tourist in Australia. Data of 25 years' record (1991–2015) are used and a time series analysis with the autoregressive integrated moving average (ARIMA) model is performed. The model reflects the exponentially increasing trend and a strong seasonality of arrivals. Average correction ratio is small as 1.0257 (SD = 0.01) and the excellent result endorsed time-series model's potential in Australian tourism industry.

In forecasting Taiwanese tourism demand, Liang (2014) employs Seasonal Autoregressive Integrated Moving Average (SARIMA) and Generalize Autoregressive Conditional Heteroskedastic (GARCH) or known as SARIMA-GARCH model. SARIMA model is able to consider data involving trends and seasonality. Hence, it is widely used in forecasting. In this research, the data series extend from January 2001 until December 2013 and consist of 156 observations. The final result on mean absolute percent error (MAPE) is 0.0327 and the researcher found out that the results confirm the effectiveness of this model; and variation of Taiwanese tourism demand changes over time. The result is significant because seasonality influences tourism demand.

Parameter of econometric model is not invariant to policy changes which alter the model structure and resulting in less accurate forecast (Rufino, 2015). The modified Box-Jenkins procedure is used in forecasting tourist arrivals from ASEAN+3 countries to the

Philippines. Rufino (2015) uses actual data from January 2000 to December 2014 for this study and concludes that the target tourist arrivals for 2015 determined by the local authority is too optimistic and far from the actual figure in 2014. Partial assessment on forecasting accuracy is made using available data from the first quarter of 2015 and it appears that the best-case scenario with MAPE of 7.8% is the most reliable. The results imply that planning under optimistic demand assumptions proved to be a wise strategy.

Gunter and Onder (2014) compare the predictive accuracy of various uni- and multivariate models in forecasting international city tourism demand for Paris from its five most important foreign source markets (Germany, Italy, Japan, UK, and US). In order to achieve this, seven different forecast models are applied: EC-ADLM, classical and Bayesian VAR, TVP, ARMA, and ETS, as well as the naïve-1 model serving as a benchmark. The accuracy of the forecast models is evaluated in terms of the RMSFE and the MAE. The results indicate that for the US and UK source markets, univariate models of ARMA (1,1) and ETS are more accurate, but that multivariate models are the better predictors for the German and Italian source markets, in particular (Bayesian) VAR. For the Japanese source market, the results vary according to the forecast horizons. Overall, the naïve-1 benchmark is significantly outperformed across nearly all source markets and forecast horizons. The results are in line with the existing literature on tourism demand forecasting, namely that there is no single tourism forecast model that outperforms all others on all occasions.

Despite the pile of reports on tourism forecasting, surprisingly a little attention is paid to the nature of the out-of-sample forecasts. The degree of the individual forecasting error, as well as the possibility of predicting an outlier, hardly addressed (Chu, 2003). Chu (2003) employs a cubic polynomial model to forecast the number of tourist arrivals from January 1989 to July 1990 in Singapore. The results show that the cubic polynomial model generates relatively accurate forecast, manufactures magnitude of MAPE equal to 4.01 and it has outperformed simple linear regression. According to Lewis (1982), results within 10% range is highly accurate.

Chen, Lai, and Yeh (2011) develop a novel forecasting model by integrating empirical mode decomposition (EMD) and back-propagation neural network (BPN). Data on tourist arrivals from Japan, Hong Kong, and Macau to Taiwan from January 1971 to August 2009 are employed. Two other forecasting models are used in comparison with EMD-BPN: Single BPN and ARIMA. Results show that EMD-BPN has outperformed the other two models with MAPE 2.741%.

Tourism researchers do not primarily use qualitative forecasting method due to the criticism on the technique that lack of generalisability (Xu, Law, Chen, & Tang, 2016). Nonetheless, the qualitative method is important in forecasting when there is a lack or no availability of any historical data. Xu, Law, Chen, and Tang (2016) argue that support vector machines (SVM) performing better than ARIMA, a traditional statistical regression model. Hence, the researchers employ trained SVM by extracting fuzzy Takagi-Sugeno rules to forecast tourism demand in Hong Kong, China. Five time series forecasting models and four causal relationship models are tested to forecast tourist arrivals from nine major tourist-generating origins encompass US, Australia, Canada, France, Germany, the UK, Japan, Korea, and Taiwan, China for comparison. Results on MAPE demonstrate excellent forecasting accuracy for the data of tourist arrivals in the study where only two countries get the best MAPE results; Canada's lowest MAPE is 12.54% in a multiple regression model; and Australia's lowest MAPE is 19.63% in Naïve model, compared to the other models.

Knowing the importance of seasonality and volatility in tourism data, Claveria and Torra (2013) compare the artificial neural networks (ANN) and time series models in their research. Statistical data of inbound international tourism demand to Catalonia from 2001 to 2009 are used for this study. ARIMA and self-exciting threshold autoregression (SETAR) models are employed for time series forecast. Root mean squared forecast error (RMSFE) is calculated to rank the different method and the results show that ARIMA models outperformed ANN and SETAR models in most countries. The difference is statistically significant in 70% and 40%, respectively.

Song, Li, Witt, and Athanasopoulos (2010) develop TVP-STSM model by combining the time-varying parameter (TVP) and structural time series model (STSM) aimed to forecast seasonal tourism demand more accurately. The TVP-STSM is applied to forecasting the quarterly demand for Hong Kong tourism by tourists from China, Korea, United Kingdom, and the United States of America, as well as the data from quarter 1 in 1985 to quarter 4 in 2014 are used. The forecast accuracy in this research is evaluated based on MAPE and the root mean squared percentage error (RMSPE). The seasonal naïve "No Change 1" (constant value for each season) and "No Change 2" (constant growth rate for each season) models, the SARIMA model, basic structural model (BSM), causal structural model (CSM), the TVP model, and autoregressive distributed lag model (ADLM) are included in the comparison of the forecasting performances. Based on the findings, the proposed model TVP-STSM outperformed the other models and MAPE results from quarter 1 to quarter 4 are 0.437, 0.554, 0.637, and 0.733, respectively.

The forecast obtained in a study of tourist in-flow in Singapore is conducted by Kumar and Sharma (2016) using model SARIMA, ARIMA, and Holt Winters. Data from 2003 to 2013 are used and the results show that SARIMA outperformed the other two models with MAPE 3.21.

There is no single forecasting method consistently outperforms other model in every situation. In a competition to forecast with the best model to fit, the econometric models are emphasised when annual data are used whereas the time series models usually show their advantage for higher frequency data (Kumar & Sharma, 2016).

The summary and comparison of studies conducted in forecasting until the recent years depicted in the table below:

Vol. 8, No. 16, Special Issue on Revisiting Foodservice and Gastronomy Intersection: Business, People and Performance, 2018,
E-ISSN: 2222-6990 © 2018 HRMARS

Table 1
Empirical Findings on Tourist Arrivals Forecasting Method (Worldwide)

	Reference	Area	Methodology	MAPE	OTHERS
1	Ma, Liu, Li, and Chen (2015)	Australia	ARIMA	11.9–14.6	
2	Liang (2014)	Taiwan	SARIMA-GARCH	0.0327	
3	Rufino (2015)	Philippines	Modified Box-Jenkins	0.0780	
4	Gunter and Onder (2014)	Paris	EC-ADLM, classical and Bayesan VAR, TVP, ARMA, ETS, naïve model		MAE 0.0535–0.1056
5	Chu (2003)	Singapore	Cubic Polynomial	4.01	
6	Chen, Lai, and Yeh (2011)	Taiwan	1. EMD-BPN 2. Single BPN 3. ARIMA	1. 0.958 2. 1.378 3. 8.876	
7	Xu, Law, Chen, and Tang (2016)	Hong Kong	1. SVMRE 2. ARIMA 3. ANN 4. Multiple Regression 5. Naïve	1. 6.91–19.85 2. 14.27–36.74 3. 8.31–31.62 4. 9.11–46.29 5. 8.12–19.63	
8	Claveria and Torra (2013)	Catalonia	1. ARIMA 2. ANN 3. SETAR		**RMSFE:** 1. 0.00–45.80 2. 0.98–21.42 3. 0.17–38.23
9	Song, Li, Witt, and Athanasopoulos (2010)	Hong Kong	1. TVP-STSM 2. CSM 3. TVP 4. BSM 5. ADLM 6. SARIMA 7. Naïve 1 8. Naïve 2	1. 0.437–1.053 2. 0.528–1.660 3. 0.615–1.330 4. 0.625–1.124 5. 0.849–1.780 6. 0.744–1.250 7. 0.895–1.469 8. 1.793–1.308	
10	Kumar and Sharma (2016)	Singapore	SARIMA, ARIMA. Holt Winters	3.21	

Forecasting Method Malaysia

There are very few studies conducted on forecasting tourist and demand modelling in Malaysia. Loganathan, Nanthakumar, and Ibrahim (2010) conduct a research with aim to generate one-period-ahead forecast of international arrivals in Malaysia. ARIMA model is used in this study and data from 1995 to 2008 are employed to forecast arrivals for Q1:2009 to Q4:2009. The result shows that international tourist arrivals to Malaysia is not depending on seasonal effects as the flow of tourist arrivals is consistent in every quarter and the trend remains the same in one decade timeline.

Shabri (2016) proposes a hybrid model of least square support vector machine (LSSVM) and ensemble empirical mode decomposition (EEMD) to forecast tourist arrivals in Malaysia. The data used in this study contain a number of tourists from Singapore in monthly period from January 2000 to December 2012 in order to forecast arrivals from January 2013 to December 2014. For the purpose of evaluation in forecasting performance between different models, mean absolute error (MAE) and root mean squared error (RMSE) are applied. The comparison is made between single LSSVM and hybrid model EEMD-LSSVM and the result shows that the proposed model is capable to increase the forecasting accuracy compared to LSSVM 22.9% and 14.2% for MAE and RMSE. As a result, the proposed model enhances the forecast accuracy and produces more stable and reliable forecasting result.

Nanthakumar, Subramaniam, and Kogid (2012) employ time series modelling SARIMA to forecast ASEAN tourist arrivals to Malaysia. The study uses quarterly data from 1995 to 2009. Based on the findings, it can be concluded that in order to forecast ASEAN tourist arrivals, the study uses SARIMA model without any seasonal effect since seasonality does not provide a reliable forecast on tourism demand by ASEAN countries.

The summary and comparison of studies conducted in forecasting until the recent years depicted in the table below:

Table 2
Empirical Findings on Tourist Arrivals Forecasting Method (Malaysia)

	Reference	**Area**	**Methodology**	**MAPE**	**OTHERS**
1	Loganathan, Nanthakumar and Ibrahim (2010)	Malaysia	ARIMA	1.4319-1.6269	
2	Shabri (2016)	Malaysia	EEMD-LLSVM		MAE 22.9
3	Nanthakumar, Subramaniam and Kogid (2012)	Malaysia	SARIMA	2.92 – 3.11	

Research Method

This section focuses on the strategies used to perform the study. The figure below shows the flow of this overall study.

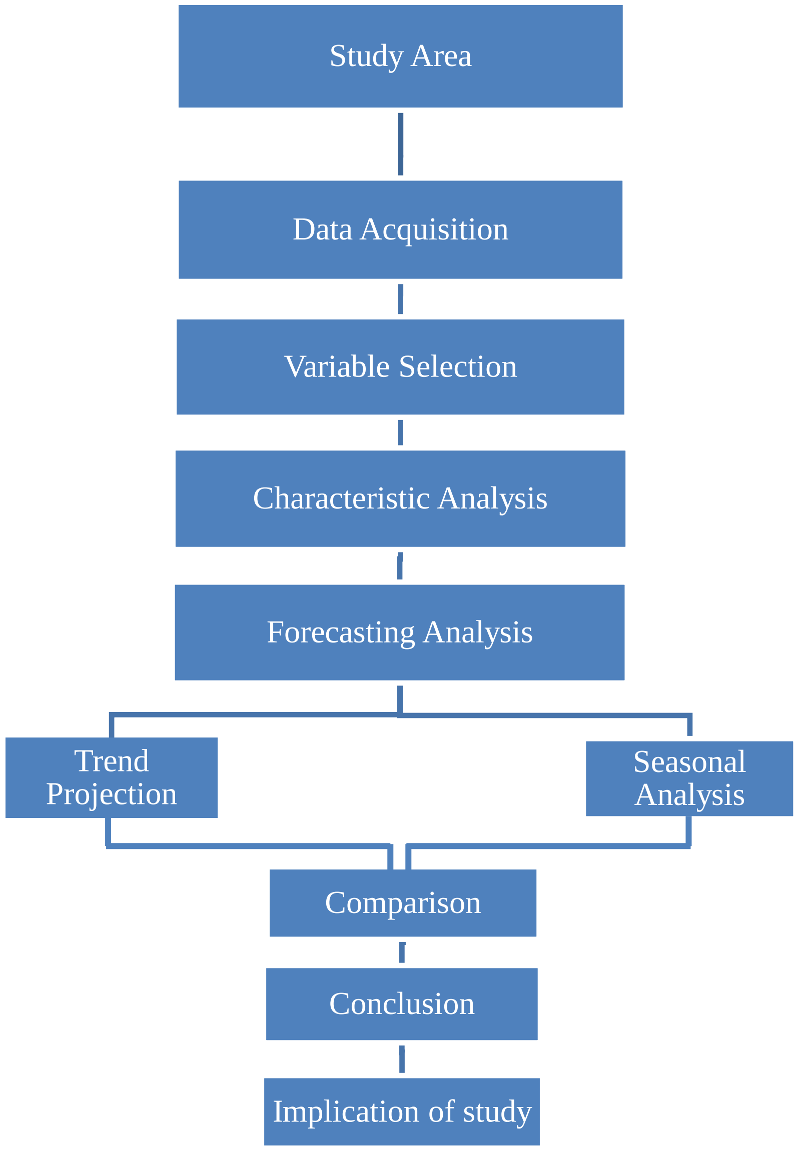

Figure 1 : Research Flow

In this study, Penang has been selected because it has been one of the leading states in tourism industry and had also dominated 40% of the overall state economy. Based on the data collected by Tourism Malaysia in 2015, Penang had received 6.3 million domestic and international tourists which represent 25% of total arrivals in Malaysia at the same year (Bernama, 2016).

Penang area is about 1,048 km2 that consists of 5 areas which are Timur Laut, Barat Daya, Seberang Perai Utara, Seberang Perai Tengah and Seberang Perai Selatan. The state has the highest population density in Malaysia with 1,505 people per square kilometre and the population of Penang is about 1,647,716 as of 2013 (Department of Statistics, Department of Survey and Mapping, Malaysia, 2013). District wise population in Penang 2009-2013 is depicted in Table 3.

Table 3
Population Distribution by District, Penang 2009-2013

	WARGANEGARA MALAYSIA				
DAERAH	2009	2010	2011	2012	2013
TIMUR LAUT	486,300	489,200	529,809	389,251	395,046
BARAT DAYA	189,500	192,200	207.004	306,440	310,982
S. PERAI UTARA	283,700	287,000	301,365	182,859	185,325
S. PERAI TENGAH	338,600	342,500	382,714	536,710	542,595
S. PERAI SELATAN	160,100	164,100	180,076	210,575	213,768
PULAU PINANG	1,458,200	1,475,000	1,600,968	1,625,835	1,647,716

Source: Department of Statistics, Malaysia (2013)

The total number of tourist arrivals is very high with frequent festival activities, international sports events and national conferences held by Penang authorities. Back in 1999, Malaysia initiated a global marketing campaign called “Malaysia, Truly Asia”. It was a successful strategy with over 7.4 million tourists was brought in. The number of tourists keeps increasing, along with the Visit Malaysia Year, as well as the Malaysia Year of Festivals and Visit Penang Year campaigns, which began in 2015 (Munan & Heidi, 2002). Consequently, Penang was chosen as a sample to forecast the total number of international tourist arrivals in Penang for the year of 2016-2017 using time series models.

Table 4
The Highest Number of International Tourist Arrived in Penang by its Regions.

Regions	**Country Selected**	**Dates of School Holiday**
ASEAN	Singapore, Brunei, Philippines, Thailand, and ***Indonesia***[a]	January, June, July, August, December.
ASIA	Hong Kong S.A.R, Hong Kong C.I, Sri Lanka, Bangladesh, India, Russia, Middle east, Taiwan, Pakistan, ***Japan***[b], China and South Korea.	April, July, August, December.
Australasia	***Australia***[c] and New Zealand	January, April, July, September, October, December.
North America	***Canada***[d] and US	July, August, December.
South America	***Latin America***[e]	January, June, July and December.
Europe	***United Kingdom***[f], German, France. Norway, Sweden Denmark, Finland, Belgium, Luxembourg, Netherlands, West Europe, East Europe	February, March, April, July, August, December.

**Selection of the country for school holidays is based on the highest number of international tourists arrived in Penang.*
** At least 7 days minimum requirement for school holiday selection.*
a: Indonesia Government (2016)
b: Japan Government (2016)
c: Australia Government *(2016)*
d: Canada Government *(2016)*
e: Brazil Government *(2016)*
f: UK government *(2016)*

Variable Selection

In this study, a dependent variable is the total number of international tourists arrived in Penang monthly. Independent variable is a seasonality or school holidays index and a trend variable. These variables are further explained below (Anaman & Looi, 2000).

$$\text{TOURIST} = B_0 + B_1{}^{*}\text{TREND} + B_2{}^{*}\text{DUMMY} \quad (1)$$

1. *Number of International Tourists to Penang (TOURIST).* This variable refers to those who entered Penang for the purpose of touring or sightseeing.
2. *Monthly Trend (TREND).* The trend variable is included to capture any sustained upward or downward movement in tourism unrelated to the other independent variable. It carries a value of 1 in January 2013 through 36 for December 2015.
3. *School Holiday (DUMMY).* This is a dummy variable that takes the value of 1 for the dates of the school holidays, and zero otherwise.

Characteristic Analysis

Deliberately to acquire a clear understanding of the population, descriptive statistical analyses are utilised on the data groups, processing measures of central tendency (means and medians) and measures of dispersion (standard deviations, minimum, maximum, and skewness).

Data analysis technique used in this study is Time Series Plots. When working with time series data, it is paramount that the data is plotted, thus the researcher is able to view the data. Time series plot displays values against time that are sufficient to show how data changes over time. The researcher samples data at a random time which signifies the stochastic nature of the measurements over time. In time series analysis, patterns are represented inside the analysis. SPSS, JMP and SAS/ETS are three statistical packages, widely available to analysed time series. Although similar, Microsoft Excel is selected to compute the data and prior to each of time series analysis, pattern has different capabilities, strengths, and weaknesses. In short, it will be more than adequately serve the researcher's purposes to perform time series analysis using QM for windows.

Forecasting

Trend Projection

The first method to forecast the international tourist arrivals in Penang using time series models is trend projection. Trend projection is a technique that suitable for historical data to fit a trend line data points and forecast the line in medium to the long run forecasts. There are a few mathematical trend equations that can be developed, such as exponential and quadratic, but for this study, only the straight line/linear is used. Linear regression equation is a simply trend line in which the independent variable (X) is the time of period. The first time period will be recognised as 1. The second time period will be recognised as 2, and so on. The last time period will be time recognised as *n* (Render, Stair, Hanna, & Hale, 2015). The form of equations is shown below:

$$\hat{Y} = b_0 + b_1X \quad (2)$$

Where

$\hat{Y}$ = predicted value
b_0= intercept
b_1= slope of line
X = time period (i.e., X= 1, 2, 3... n)

There is also another method to minimise the mean squared error (MSE) using the least square regression in order to find the coefficient that minimises the sum of the squared errors. This problem could be solved using software QM for Windows (Render, Stair, Hanna, & Hale, 2015).

Trend Projection with Seasonal Effect

The second method for this study uses seasonal variation method. It is necessary to make a seasonal adjustment in the trend line forecast; however, periodically on account of recurring variations at certain seasons of the year. A seasonal Index could be applied to indicate the way of a particular season, such as months or a quarter can be compared with an average season. Index 1 indicates a season in an average condition. While, if the index shows more than 1, the value of the time series in that season could be higher than average, meanwhile if the index is less than 1, the values of the time series in that season could be lower than average. When a seasonal component exists, the seasonal index could be applied in multiplicative time series forecasting model to make an adjustment in the forecast (Render, Stair, Hanna, & Hale, 2015).

Trend projection with seasonal effect is a process to isolate linear trend and seasonal factors to develop more accurate forecast.

When both trend and seasonal component are present in a time series, multiple regressions may be used to forecast the problem. One independent variable is categorised as time and other independent variables are categorised as dummy variables to indicate the season. An additive trend projection with seasonal effect model is the basic model that can be expressed as follows (Render, Stair, Hanna, & Hale, 2015):

$$\hat{Y} = a + b_1X_1 + b_2X_2 \quad (3)$$

Where:

X_1= time period
X_2= 1 if school holiday
= 0 otherwise

Measure of Forecasting Accuracy

In this study, mean absolute percentage error (MAPE) is used to measure the accurate results between regions, model, and then the best model will be recommended.

The criteria in selecting between forecasting models, and for keeping tabs of how well a forecast is doing once it is implemented is called measuring the accuracy or the error of the forecast. In order to do this, the average error of a forecast is computed over an appropriate period of time. Typically, the appropriate period of time would be the period of time from which data is gathered and forecasts are applied. This study uses mean absolute percentage error (MAPE) to measure the accuracy or the error of the forecast. This technique is used to determine the accuracy of a forecasting model by taking the average of the absolute errors as a percentage of the observed vales (Render, Stair, Hanna, & Hale, 2015). It measures the size of the error in percentage terms. It is calculated as the average of the unsigned percentage error. Many organisations focus primarily on the MAPE while assessing forecast accuracy. Most people are comfortable thinking in percentage terms, making the MAPE easy to interpret. It is one of the ways to convey information, if the item's demand volume is unknown (Stellwagen, 2011). The measures used to evaluate the accuracy of forecasts are also engaged, and while accuracy is not the only criteria advocated for evaluating demographic forecasts, it is generally acknowledged to be the most significant (A., Tayman, & Bryan, 2011).

The mean absolute percentage error (MAPE) is expressed in generic percentage terms and it is computed by the following formula (Mamula, 2015):

$$\text{MAPE} = \frac{1}{n} \sum_{n_{t-1}}^{n} \frac{|\,\text{Actual} - \text{Forecast}\,|}{\text{Actual}} \times 100 \qquad (4)$$

Where:

n = time periods

According to Baggio and Klobas (2011), scale for the accuracy of a model can be measured using MAPE. MAPE forecasting accuracy depicted in the table below:

Table 5

Forecasting Accuracy According to MAPE

MAPE	FORECASTING ACCURACY
Less than 10%	Highly accurate
10% – 20%	Good
21% – 50%	Reasonable
Greater than 50%	Inaccurate

Source: Baggio and Klobas (2011)

Result and Discussion

This section explains on the characteristics of international tourist arrivals in Penang for 2013-2015, followed by the findings of forecasting on international tourist arrivals in Penang for 2016-2017 using trend projection and trend projection with seasonal effect, and finally determines the best forecasting model.

Descriptive Statistics

The percentage of mean statistics for the international tourist arrivals at Penang Airport for 2013-2015 is depicted in Figure 2 below. Most international tourist arrivals are from the ASEAN region represents 74.84% (*n* = 1,538,757). Total international tourist arrivals for 6 regions are 2,056,042. The major contributors of ASEAN tourists are from Indonesia and Thailand. Tourists from Indonesia to Penang can be classified as medical tourists as most of Indonesian tourists come to Penang to seek medical treatment (Bernama, 2015).

Population in Penang as in 2013 is approximately 1.6 million. Visiting families and friends are the main reason for Singaporean tourists to visit Malaysia as well as Penang (Launch of the Penang Investment Tourism Office (PITO), Singapore, 2010).

Moreover, ASEAN Tourism Agreement signed in year 2001 recognised the strategic importance of tourism industry and stressing on the need of cooperation among ASEAN countries in making travel into and within ASEAN easier and more efficient (Agreement, 2001). Hence, the ease of travel between ASEAN countries has been a strong pulling factor that attracts ASEAN tourists to Penang.

Asia tourists represents 10.58% (*n* = 217,465) out of total international tourist arrivals followed by Europe tourists that represents 6.94% (*n* = 142,712) of total tourists. China is among Malaysia's top three highest tourists from Asia in 2013 and it is expected to increase in the future. Malaysia is well known among Chinese for its local products, and having established famous Singapore-Malaysia-Thailand travel route (Astro Awani, 2014). Westerner visitors love to visit cultural and heritage sites (Astro Awani, 2014) and according to one of the tourist guides in Penang, Europeans have a deep curiousity about culture and heritage as they always ask questions on that area (Chin, 2014).

North America, Australasia, and South America are among the lowest contributors of international tourist arrivals with 3.94% (*n* = 81,022), 3.51% (*n* = 72,212), and 0.19% (*n* = 3,874), respectively.

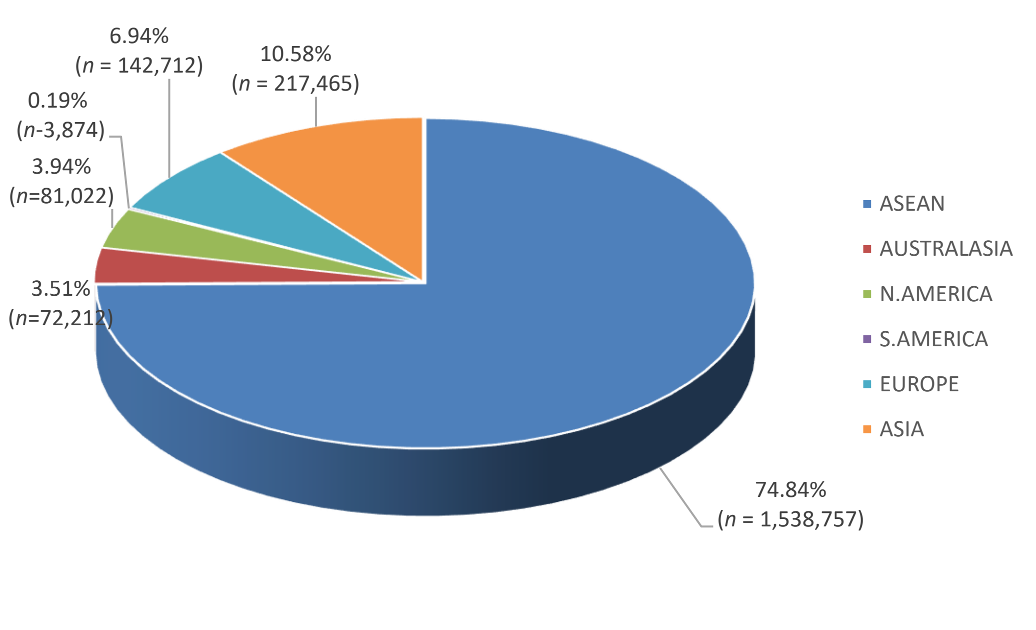

Figure 2: Percentage of Mean Statistic of International Tourist Arrivals at Penang Airport for 2013-2015

The descriptive analysis of international tourist arrivals for 2013-2015 as depicted in Table 6 shows that South America and Europe positively skewed at 3.46 and 3.29, respectively. While other regions are normally distributed as the skewness is near to zero.

The more widely spread the values are, the larger the standard deviation is. ASEAN is a group of countries that presents the highest dispersion in year 2013-2015. However, the differences in dispersion between each region are influenced by the size of data set (Statistics Canada, 2013). The total mean of international tourists from South America is 0.19% out of overall international tourist arrivals. Hence, this result describes the smallest dispersion of South America at 67.30.

Table 6

Descriptive Statistics of International Tourist Arrival at Penang Airport for 2013 – 2015

	MEAN	MEDIAN	SKEWNESS	STANDARD DEVIATION	MIN	MAX
ASEAN	42,743.25	42,581	0.43	4894.75	33,273	54,129
AUSTRALASIA	2,005.89	2,008	0.08	372.86	1,372	2,727
NORTH AMERICA	2,250.61	2,259.5	0.38	218.10	1,749	2,668
SOUTH AMERICA	101.86	93.5	3.46	67.30	34	369
EUROPE	3,964.22	3,856.5	3.29	1229.23	2,534	9,795
ASIA	6,090.69	5,950.5	0.17	855.36	3,614	8,089

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. Figure 3 below shows the time series plot of international tourist arrivals at Penang Airport for 2013-2015.

Penang tends to have a steady number of tourist arrivals and the number of arrivals keep on increasing despite the decrease of total tourist arrivals inbound Malaysia in 2015. In 2013, there are 765,445 tourist arrivals and increases to 802,913 arrivals in 2014. The numbers keep on increasing in 2015 results to 1,284,999 tourist arrivals.

Based on the time series plot as depicted in Figure 3, ASEAN tourist arrivals reach its highest plot in June and December 2013. However, the trends change with the decrease of 8895 and 2406, respectively in the same period in 2014. The tremendous reduction of ASEAN tourists in June 2014 is most probably due to the tragedy of MH370 that happened in March in the same year. In 2015, ASEAN tourist arrivals reach its highest plot in July. ASEAN tourist arrivals are significantly higher from December 2013 to 2015 and this is due to Pesta Penang that starts from early December every year. Moreover, referring to the school holidays in one of the ASEAN countries which is Indonesia, the trend also shows that ASEAN tourists optimise the school holidays to travel abroad.

Other regions such as Asia, North America, South America, and Australasia show the same trend that among the highest arrivals recorded, is in the school holiday periods at their counties, respectively. This result also explains the dropping trend in December for tourist arrivals from North America, Asia, and Europe as the school holidays end in December in these regions. However, there is unusual trend for Europe tourist arrivals with a tremendous upsurge of the number of arrivals by 5904, in April 2014.

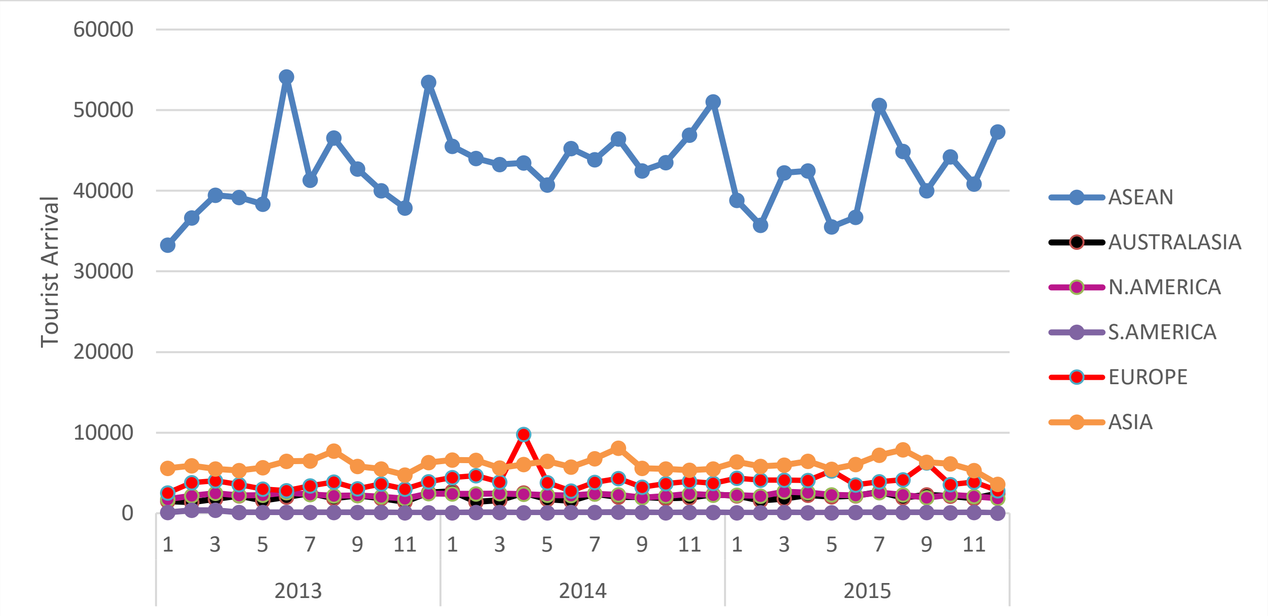

Figure 3 : Time Series Plot of International Tourist Arrivals at Penang Airport for 2013-2015

Time Series Model

The time series models look through things that have happened over a period of time and use a series of past data to make a forecast. Therefore, this study uses the past data of the number of tourist arrivals at Penang Airport for 2013-2015 to forecast the arrivals for future periods of 2016 and 2017. The time series models that are being used in this study are trend projection and trend projection with seasonal effect.

Trend Projection

This study has developed the mathematical trend equations using linear (straight line) trend. Table 7 below shows the regression model of international tourist arrivals forecast in Penang for 2016-2017.

Regression model as depicted in the Table 7 below illustrates the linear trend pattern of international tourist arrivals in Penang for a period of time, starting January 2016 until December 2017 (24 months/times). A positive result of 59.456 for ASEAN region shows the increasing number of international tourist arrivals in Penang by 59.456 per month. Australasia region shows a positive result of 10.97 that indicates an increasing number of international tourist arrivals per month. North America also shows a positive result which means the increasing number of international tourist arrivals per month by 0.096. While South America gets a negative result of -1.832 shows the decreasing number of international tourist arrivals from South America by 1.832 per month. A positive result of 22.22 for Europe region shows the increasing number of international tourist arrivals. Asia region shows a positive result of 0.149 which indicates the increasing number of tourist arrivals by 0.149 per month.

Table 7
Regression Model of International Tourist Arrivals in Penang for 2016 – 2017

REGION	**MODEL**
ASEAN	41668.32 + 59.456 * time (x)
AUSTRALASIA	1802.941 + 10.97 * time (x)
NORTH AMERICA	2248.84 + 0.096 * time (x)
SOUTH AMERICA	135.759 - 1.832 * time (x)
EUROPE	3553.151 + 22.22 * time (x)
ASIA	6037.93 + 0.149 * time (x)

Based on the forecasting result, the tourist arrivals from ASEAN, Asia, North America, Europe, and Australasia are expected to increase steadily while the arrivals from South America is estimated to drop slightly compared to the arrivals in the previous years. This result is consistent with regression shown in Table 7 that has been explained previously.

The criteria in selecting between the forecasting models, and for keeping tabs of how well a forecast is doing once it is implemented is called measuring the accuracy or the error of the forecast. To do this, the average error of a forecast over an appropriate period of time will be computed.

This study uses the mean absolute percentage error (MAPE) to measure the accuracy or the error of the forecast. This technique is used to determine the accuracy of a forecasting model by taking the average of the absolute errors as a percentage of the observed values (Render, Stair, Hanna, & Hale, 2015).

As depicted in Table 8 below, ASEAN indicates 8.8% errors of the forecast, while Australasia indicates 14.8%, North America with 7.6%, South America results to 27.2%, Europe with 15.5%, and Asia with 10.8%.

These results show that the forecasting using trend projection model is highly accurate for two regions that are ASEAN and North America as the result is less than 10%. The error measure for Australasia, Europe, and Asia is good at range between 10% and 20%. While the error measure for South America is at reasonable measure as the results is in the range of 21% to 50%.

Table 8
Result for Error Measures between Forecast and Actual Number of Tourist Arrivals in Penang for 2016 - 2017

In percentage (%)						
	ASEAN	**AUSTRAL-ASIA**	**NORTH AMERICA**	**SOUTH AMERICA**	**EUROPE**	**ASIA**
MAPE	8.8	14.8	7.6	27.2	15.5	10.8
Forecasting Accuracy	Highly Accurate	Good	Highly Accurate	Reasonable	Good	Good

Trend Projection with Seasonal Effect

Seasonality or seasonal variations in a time series is the fluctuation that occurs every month and every year. Seasonal variations tend to be repeated from year to year and always of a fixed and known period (Time Series Components, 2016). In this study, seasonality is referring to school holidays that occur each year. December is part of the summer holiday in Australia that contributes the highest number of tourists from Australasia region, meanwhile the summer holiday season for countries in the Northern Hemisphere is in July and August (Anaman & Looi, 2000).

One independent variable is time, and another independent variable is a dummy variable to indicate the season while a dependent variable is the total number of international tourist arrivals in Penang every month.

Regression model as depicted in the Table 9 below illustrates the multiple regression in forecasting the number of international tourist arrivals using trend projection with seasonal effect for a period of time starting January 2016 until December 2017. In this model, time has been assumed as a fixed variable to determine the changes in the number of international tourist arrivals by adding the school holiday season effect. ASEAN shows a positive result that represents an increasing number of international tourist arrivals by 4316.133 for a respective month that affected by school holiday season. Australasia, North America, Europe and Asia also have an increasing number of international tourist arrivals by 481.244, 49.903, 538.69 and 615.509 respectively. Whereas, South America show the negative result that represent a decreasing number of international tourist arrivals by 20.792. The regression model also shows that school holiday season does not affect the travel pattern for this region.

Table 9
Regression Model with Seasonal of Tourist Arrivals in Penang for 2016 - 2017

REGION	**MODEL**
ASEAN	39965.17 + 54.353* time (x) + 4316.133* school holiday
AUSTRALASIA	1589.819 + 9.484 * time (x) + 481.244 * school holiday
NORTH AMERICA	2241.711 - 0.193 * time (x) + 49.903 * school holiday
SOUTH AMERICA	163.718 - 2.658 * time (x) - 20.792 * school holiday
EUROPE	3260.719 + 23.468 * time (x) +538.69 * school holiday
ASIA	5876.726 - 2.227* time (x) + 615.509 * school holiday

On account of the huge gap between ASEAN and the other regions, the fluctuations in tourist arrivals for other regions are not clear. It is estimated that ASEAN tourist arrivals for June, July, August, and December for 2016 and 2017 to increase steadily because of the school holiday season in Indonesia. The trend is quite similar to the actual data in 2013-2015 that among the highest ASEAN tourist arrivals in that particular year is in the mid-year, and at the end of the year.

The tourist arrivals from Australasia is estimated to fluctuate and reach its highest point in January, April, July, September, October and December as school holidays in the region fall in these months. The forecasting results for Australasia is in line with regression and it can be interpreted that school holiday season influences the number of Australasia tourist arrivals in Penang.

School holiday season in North America falls in July, August, and December every year. However, the findings show that the school holiday season does not affect tourist arrivals. The North America tourist arrivals in December 2015 is 1854 and it is estimated that the arrivals in January 2016 will increase to 2235. It is also estimated to increase slightly every month and reach 2280 arrivals in December 2017.

Tourist arrivals from South America are estimated to decline throughout the year of 2016 and keep on declining until December 2017. The forecasting results is in line with the regression that shows the school holidays do not affect South America tourist arrivals and will decline by 20.792.

The Europe and Asia tourist arrivals show very small fluctuations. It is shown that the school holiday season has an impact to Europe and Asia tourist arrivals.

As in the international tourist arrivals forecast using trend projection, this forecast also uses mean absolute percentage error (MAPE) to measure the accuracy or the error of the forecast.

As depicted in Table 10 below, ASEAN indicates 7.7% errors of the forecast, meanwhile Australasia indicates 10.9%, North America with 7.4%, South America is 33.6%, Europe is 16.3%, and Asia is 9.9%. These results show that the forecasting using trend projection with seasonal effect model is highly accurate for ASEAN, North America and Asia as the result is less than 10%. The error measure for Australasia and Europe is good at range between10% and 20%. While the error measure for South America is at reasonable measure as the results is in the range of 21% to 50%.

Table 10
Result for Error Measures between Forecast and Actual Number of Tourist Arrivals in Penang for 2016 - 2017 with Seasonal Variable

In percentage (%)						
	ASEAN	**AUSTRAL-ASIA**	**NORTH AMERICA**	**SOUTH AMERICA**	**EUROPE**	**ASIA**
MAPE	7.7	10.9	7.4	33.6	15.1	9.9
Forecasting Accuracy	Highly Accurate	Good	Highly Accurate	Reasonable	Good	Highly Accurate

Comparison of Forecast Accuracy

This section is to determine the best tourist arrivals forecasting model. In order to determine the best forecasting model, a comparison in MAPE results is made. The comparison in MAPE results between trend projection model and trend projection with seasonal effect model is depicted in table below.

The results show that trend projection with seasonal effect model for five regions outperformed trend projection model in which the results for ASEAN, Australasia, North America, Europe and Asia are lowest at 7.7%, 10.9%, 7.4%, 15.1%and 9.9%, respectively. Error measures for these five regions are between excellent and good. This result indicates that trend projection with seasonal effect model produce more stable and reliable forecasting results. The findings also indicate that seasonality has a significant effect to international tourist arrivals in Penang. However, previous study conducted by Nanthakumar, Subramaniam, and Kogid (2012), specifies that seasonality does not provide a reliable forecast on tourism demand in Malaysia.

Table 11
Comparison of Error Measures between Trend Projection Method and Trend with Seasonal Method

	ASEAN	AUSTRALASIA	NORTH AMERICA	SOUTH AMERICA	EUROPE	ASIA
In percentage (%)						
MAPE Trend	8.8	14.8	7.6	**27.2***	15.5	10.8
MAPE Trend & Seasonal	**7.7***	**10.9***	**7.4***	33.6	**15.1***	**9.9***

* *lower MAPE selected*

Conclusion

The characteristics of international tourist arrivals in Penang are determined by the total mean of international tourist arrivals from 2013 to 2015. Based on the findings in chapter 4, the largest contributor of tourist arrivals in Penang is from ASEAN which represents 74.84% and the lowest tourist arrivals is from South America which represents 0.19% out of total international arrivals in 2013-2015. Total arrivals for 2013-2015 are 665,292, 715,200, and 675,550, respectively.

Trend projection model and trend projection with seasonal effect model are used in this study to forecast international tourist arrivals from 2016 to 2017. Based on the findings, error measure using both model are vary from excellent to reasonable with range between 7.4% and 33.6%. Nonetheless, the result is in line with previous research as the results are between range 0.0327% and 46.29%. The accuracy of MAPE is highly accurate when the result is less than 10%.

A comparison between these models has been made and it is discovered that trend projection with seasonal effect model is more appropriate to forecast international tourist's arrivals in Penang due to five out of six regions MAPE results' outperformed trend projection model.

Acknowledgement
Thank you to Universiti Teknologi MARA (UiTM) for providing financial support to carry out this study and also thanks to the Department of Tourism Malaysia for their support.

Corresponding Author
Ahmad Zia Ul-Shaufie Mohamad Japeri, Faculty Computers and Mathematic Sciences, Universiti Teknologi MARA Cawangan Pulau Pinang. Email: ahmadzia101@ppinang.uitm.edu.my

References

Anaman, K. A., & Looi, C. N. (2000). Economic impact of haze-related air pollution on the tourism industry in Brunei Darussalam. *Economic Analysis and Policy, 30* (2), 133-143.

Baggio, R., & Klobas, J. (2011). *Quantitative Methods in Tourism: A Handbook.* Bristol, UK: Channel View Publications.

Chen, C.-F., Lai, M.-C., & Yeh, C.-C. (2011). Forecasting tourism demand based on empirical mode decomposition and neural network. *Knowledge-Based Systems*.

Chen, R., Liang, C.-Y., Hong, W.-C., & Gu, D.-X. (2014). Forecasting holiday daily tourist flow based on seasonal support vector regression with adaptive genetic algorithm. *Review Article*.

Chu, F.-L. (2003). Forecasting tourism demand: a cubic polynomial approach. *Tourism Management*.

Claveria, O., & Torra, S. (2013). Forecasting tourism demand to Catalonia: Neural networks vs. time series models. *Economic Modelling*.

Costa, J., Montenegro, M., & Gomes, J. (2016). What global trends are challenging tourism organizations and destinations today? *Worldwide Hospitality and Tourism Themes*.

Gunter, U., & Onder, I. (2014). Forecasting international city tourism demand for Paris: Accuracy of uni- and multivariate models employing monthly data. *Tourism Management*.

Jala, I. (2016). *Pemandu*. Retrieved from Tourism: A Key Economic Sector: https://www.pemandu.gov.my/transformation-unplugged-tourism-a-key-economic-sector/

Kumar, M., & Sharma, S. (2016). Forecasting tourist in-flow in South East Asia: A case of Singapore. *Tourism & Management Studies*.

Liang, Y.-H. (2014). Forecasting models for Taiwanese tourism demand after allowance for Mainland China tourists visiting Taiwan. *Computers & Industrial Engineering*.

Loganathan, Nanthakumar, & Ibrahim, Y. (2010). Forecasting International Tourism Demand in Malaysia Using Box Jenkins Sarima Application. *South Asian Journal of Tourism and Heritage*.

Ma, E., Liu, Y., Li, J., & Chen, S. (2015). Anticipating Chinese tourists arrivals in Australia: A time series analysis. *Tourism Management Perspectives*.

Mamula, M. (2015). Modelling and Forecasting International Tourism Demand – Evaluation of Forecasting Performance . *International Journal of Business Administration* .

Munan, & Heidi. (2002). *Malaysia.* New York: Benchmark Books.

Nanthakumar, L., Subramaniam, T., & Kogid, M. (2012). Ia Malaysia truly Asia? Forecasting tourism demand from Asean using SARIMA approach. *An International Multidiciplinary Journal of Tourism*.

Render, B., Stair, J. M., Hanna, M. E., & Hale, T. S. (2015). *Quantitative Analysis for Management.* Harlow: Pearson Education Limited.

Rufino, C. C. (2015). Forecasting Monthly Tourist Arrivals from ASEAN+3 countries to the Philippines for 2015-2016 Using SARIMA Noise Modeling.

Shabri, A. (2016). A Hybrid of EEMD and LSSVM-PSO model for Tourist Demand Forecasting. *Indian Journal of Science and Technology*.

Song, H., & Witt, S. F. (2006). Forecasting International Tourist Flows to Macau.

Song, H., Li, G., Witt, S. F., & Athanasopoulos, G. (2010). Forecasting tourist arrivals using time-varying parameter structural time series models. *International Journal of Forecasting*.

Song, H., Witt, S. F., & Li, G. (2008). *The Advanced Econometrics of Tourism Demand.* Routledge.

Statistics Canada. (2013). Retrieved from Variance and standard deviation: http://www.statcan.gc.ca/edu/power-pouvoir/ch12/5214891-eng.htm

Stellwagen, E. (2011). *A Guide to Forecast Error Measurement Statistics and How to Use Them*. Retrieved from Forecast Pro: http://www.forecastpro.com/Trends/forecasting101August2011.html

Time Series Components. (2016). Retrieved from OTexts: https://www.otexts.org/fpp/6/1

Tourism Malaysia. (2016). Retrieved from My Tourism Data: http://mytourismdata.tourism.gov.my/

Tourism Malaysia. (2016). Retrieved from Malaysia's Jan-June 2016 tourist arrivals grow 3.7%: http://www.tourism.gov.my/media/view/malaysia-s-jan-june-2016-tourist-arrivals-grow-3-7

Xu, X., Law, R., Chen, W., & Tang, L. (2016). Forecasting tourism demand by extracting fuzzy TakagieSugeno rules from trained SVMs. *CAAI Transactions on Intelligence Technology*.