

Time Series Analysis for Modeling and Forecasting International Tourist Arrivals in Sri Lanka

D. K. Ishara^{1*} and P. Wijekoon¹

¹Department of Statistics and Computer Science, Faculty of Science, University of Peradeniya, Sri Lanka.

Authors' contributions

This work was carried out in collaboration between both authors. Author PW designed and supervised the study. Author DKI managed the literature survey, performed the statistical analysis and wrote the first draft of the manuscript. Author PW edited the manuscript. Both authors read and approved the final manuscript.

Article Information

DOI: 10.9734/BJMCS/2017/31726

Editor(s):

(1) Metin Basarir, Department of Mathematics, Sakarya University, Turkey.

Reviewers:

(1) Martin Anokye, University of Cape Coast, Ghana.

(2) Irshad Ullah, Education Government of Khyber Pakhtunkhwa, Pakistan.

Complete Peer review History: <http://www.sciencedomain.org/review-history/17994>

Received: 21st January 2017

Accepted: 17th February 2017

Published: 28th February 2017

Short Research Article

Abstract

Tourism is one of the income generating industries in a developing country which directly contribute to the economy. Therefore, forecasting tourist arrivals is important for making policy decisions to improve facilities and other related factors in this industry. In this paper, an attempt has been made to forecast tourists' arrivals using time series modelling. The time span used for the study is from January 2000 to February 2016. In the modelling exercise, data has been analyzed based on the two sets of data; long-term (2000-2016), and post-war (2010-2016). This categorization was due to the significant change in the industry after the end of the civil war in 2009 in Sri Lanka. An Auto-Regressive Integrated Moving Average (ARIMA) method and Multiplicative Decomposition approach (MDA) were employed to model the data. When the forecasts from these models were validated, post-war data has more accurate results having low Mean absolute percentage error (MAPE) for MDA than the ARIMA approach. The comparison of actual data with the predicted values also confirmed that the MDA model obtained from the post-war series has high predictive ability.

Keywords: ARIMA Model; forecasts; seasonality; multiplicative decomposition method; tourist arrivals; predictions.

*Corresponding author: E-mail: isharakalpani1992@gmail.com;

Co-author: E-mail: pkwijekoon@gmail.com;

1 Introduction

Over the past six decades, tourism has become one of the most rapidly growing industries in the world. The arrivals of tourists have significant impacts on the economic growth of both developed and developing countries. Tourists' demand or simply tourist consumption is contributing to GDP, increasing the employment rate, and making a new source of revenue for local people, private and public sectors.

When considering Sri Lanka as a developing nation, tourism is one of its' major foreign exchange earners. For instance, the public sector revenue from tourism was approximately Rs. 4,017.3 million in 2011, but in 2015 it showed a smart increase of Rs. 8,282.7 million. From 2009 to 2016 the Sri Lankan tourism industry has been growing dramatically. This significant impact is enough to encourage researchers to investigate the number of tourist arrivals and attempt to make a more accurate prediction for future planning.

Sri Lankan tourism has had many drawbacks during the last three decades mainly due to the uncertain security situation due to the civil war in the country. The Tsunami in 2004 made this situation further worsened. The global economic recession has also had a major impact on the industry. At present, Sri Lanka has overcome with all these drawbacks and the industry is leading to a new era. In such a situation the forecast of tourists' arrivals is important since it would be beneficial the tourism related industries like airlines, hotels and other stakeholders to adequately prepare for the expected number of tourists in the coming year.

Over the past decades, a variety of time series modelling and forecasting techniques has been used in analyzing tourism arrivals [1,2]. These models vary from simple naive models to more complex models such as advanced econometric models. Chu [3] has applied six types of time series approaches (Naïve I, Naïve II, Linear Trend, Sine Wave, Holt-Winters and ARIMA) to forecast tourism demand in ten countries (Japan, South Korea, Taiwan, Hong-Kong, Philippines, Indonesia, Singapore, Thailand, Australia and New Zealand), and identified the ARIMA model as the most suitable model for prediction for nine of the ten countries with the minimum MAPE (Mean Absolute Percentage Error). Akuno et al. [4] have conducted a study to forecast tourists' demand in Kenya using Double Exponential Smoothing and ARIMA models. They have shown that the Double Exponential Smoothing model was the best forecasting model for tourists' arrival in Kenya since its MAPE and RMSE (Regression Mean Square Error) were minimum compared to those of the ARIMA model. Tularam et al. [5] have shown that the ARIMA model was a better predictor than VAR model [6] for the tourism demand in Australia.

Lelwala and Kurukulasooriya [7] have analyzed the arrivals of tourists in Sri Lanka using the classical decomposition technique. They used both additive and multiplicative decomposition models and have shown that multiplicative decomposition model was the best to forecast Sri Lankan tourists' arrivals due to its' minimum MAPE value. Konarasinghe [8] have studied the post-war behavior of Sri Lankan tourists arrivals using smoothing techniques. Fig. 1 of his paper [8] shows the variations of tourism arrivals in three stages; Pre-war (January 1968 – June 1983), war period (July 1983 – May 2009) and post-war (January 2010 – December 2014). According to this figure, a clear increasing trend during the pre-war period and some fluctuations during the period of civil war can be noted. However, in the post-war period, there is a remarkable increase of tourist arrivals with compared to the other two periods. Based on final results, he proposed a Holt Winter's three parameter model to forecast future arrivals. Gnanapragasam and Cooray [9] have applied dynamic transfer function (DTF) modeling method to predict tourist arrivals in Sri Lanka. Finally he concluded that the fitted DTF model explains over 90% accuracy in terms of forecasting tourist arrivals.

In this paper, an attempt has been made to forecast tourists' arrivals using Auto-Regressive Integrated Moving Average (ARIMA) and multiplicative decomposition approach [10]. The monthly data on tourists' arrival were obtained from the Annual Statistical Report 2015 [11] of tourism research and statistics. The data available for this study is from January 2000 to August 2016. Therefore, the model fitting was done separately for two sets of categorized data; long-term (January, 2000-February, 2016), and post-war period

(January, 2010-February, 2016). The data for the remaining six months were used for model validation. The analysis was carried out using R software.

2 Materials and Methods

The two time series methods; ARIMA and multiplicative decomposition approach used in this study are explained briefly in section 2.1 and 2.2. The statistical methods, used to understand the low, medium and high demand periods, are described in section 2.3.

2.1 ARIMA approach

To apply the ARIMA approach it is necessary to identify whether the time series is stationary, and whether it has any seasonal effect. The behavior of the time series data is visible in a time series plot, which shows whether any significant trend and seasonality exist.

If at least one of these factors are present the data are said to be non-stationary. The stationarity of the data can be achieved by applying differencing method and log transformation. The Augmented Dickey-Fuller (ADF) test is then applied to confirm the stationarity of the data. This test follows a unit-root process, and the test indicates whether a unit root exists or not. If the series does not have a unit root, the data can be taken as stationary.

ARIMA models can now be fitted to the stationary data. Generally an ARIMA model has two forms, a seasonal ARIMA model and a non-seasonal ARIMA model. A non-seasonal ARIMA model (X_t) is given in equation (1)

$$\Phi(L)\nabla^d X_t = \mu + \Theta(L)\varepsilon_t \forall t \geq 0 \quad (1)$$

which is denoted as "ARIMA (p, d, q)", where

- $\Phi(L) = 1 - \phi_1 L - \dots - \phi_p L^p$ and $\Theta(L) = 1 + \theta_1 L + \dots + \theta_q L^q$; $\phi_p \neq 0, \theta_q \neq 0$,
- L is the lag operator, ε_t is the weak white noise and μ is a constant term,
- p (Autoregressive parameter) is the number of autoregressive terms,
- d (Integrated parameter) is the number of non seasonal differences needed for stationarity.
- q (Moving average parameter) is the number of lagged forecast errors in the prediction equation.

Seasonality is one of the major factors affecting the variations of a time series. Seasonal ARIMA (SARIMA) model given in equation (2)

$$\Phi_p(B^s)\Phi(B)\nabla_s^D\nabla^d Z_t = \alpha + \Theta_q(B^s)\theta(B)a_t \quad (2)$$

incorporates both non-seasonal and seasonal factors in a multiplicative model, denoted by SARIMA(p, d, q)(P, D, Q) $_s$, where

- p, d, q are the parameters in non-seasonal ARIMA model as mentioned above,
- Z_t is the Auto regressive process a_t is the Moving-average process,
- s is the seasonal term, α is the constant term and B is the back shift operator,
- P is the number of seasonal autoregressive terms,
- D is the number of seasonal differences, and
- Q is the number of seasonal moving average terms.

Using the ARIMA approach a set of possible models can be identified. Then, to identify the best fitted model, Box and Jenkins two graphical procedures, estimated autocorrelation function (ACF) and the

estimated partial autocorrelation function (PACF) can be used. The optimal model is obtained on the basis of minimum value of Akaike Information Criteria (AIC). The residuals of the selected model are then tested for auto correlation by using Ljung-Box Q test for which the test statistic is given below:

$$Q = n(n + 2) \sum_{k=1}^h \frac{\hat{\rho}_k^2}{n-2}, \tag{3}$$

where $\hat{\rho}_k$ is the estimated residual auto correlation at lag k , h is the number of lags being tested, and n is the number of residuals. Finally, to measure the predictive ability of the selected model, the mean absolute percentage error (MAPE), which is also known as mean absolute percentage deviation (MAPD), is used. MAPE is defined by the formula,

$$MAPE = \frac{100}{n} \sum_{i=0}^n \left| \frac{A_t - F_t}{A_t} \right| \tag{4}$$

where A_t is the actual value, F_t is the predicted value, and n is the number of observations.

A criteria (see Table 1) based on MAPE values developed by Lewis [12] can be used to measure the predictive accuracy of the model.

Table 1. MAPE values for model selection

MAPE (%)	Evaluation
MAPE ≤ 10%	High accuracy forecasting
10% < MAPE ≤ 20%	Good forecasting
20% < MAPE ≤ 50%	Reasonable forecasting
MAPE > 50%	Inaccurate forecasting

2.2 Multiplicative decomposition method

Decomposition methods are mainly used to isolate the seasonal effects of time series data. When the variation in the seasonal pattern, or the variation around the trend-cycle, appears to be proportional to the level of the time series, then a multiplicative model is more appropriate. The multiplicative decomposition model is given by the equation,

$$Y_t = T \times S \times C \times I \tag{6}$$

where, T, S, C, I represent trend, seasonal, cyclic, and irregular components, of the data respectively.

To apply this method, first, the series is log transformed to adjust it seasonally. However, seasonally adjusted series may contain an irregular component as well as a trend-cycle. Then moving average (MA) method is applied to estimate the trend cycle and to smooth the data. Since the seasonal period is even, order 12-MA is used for the estimation.

2.3 Analyzing monthly demand periods

Identifying the tourists' seasons can be helpful to enhance the operational decisions, such as capacity planning and development in many related areas and planning the human resource for the peak demand periods. Therefore, the study also focuses on categorizing twelve months according to their arrival numbers. Here we use two graphical representations; Box plot of monthly arrivals and plot of the monthly seasonal component. Graphical results are further analyzed numerically using excel. The monthly proportion of arrivals on annual total arrivals gives the measure of demand for each month.

3 Results and Discussion

In this section, results obtained for the ARIMA approach and MDA applied to the two sets of data; long-term, and post-war were presented in section 3.1, and 3.2, respectively. In section 3.3, the fitted models were compared to identify their predictive ability.

3.1 ARIMA approach

3.1.1 Long-term analysis (January, 2000 - February, 2016)

Historical data of the tourists arrivals in Sri Lanka from January 2000 to February 2016 consist totally 194 observations of monthly data. The data was plotted (Fig. 1) to apply ARIMA approach, and it shows non-stationary effects since a significant trend and seasonal component present. The red and blue colours in the time series plot indicate the variation of tourist arrivals in the post-war and war-period. When comparing the two periods a remarkable increase in the post-war period can be noted. The data was made stationary by taking the first order difference ($d = 1$), and used log transformation to make the seasonal effect constant. The Augmented Dickey-Fuller test confirmed the stationarity of transformed and differenced data.

The parameters of the best fitted ARIMA model having the lowest AIC values are summarized in Table 2, where AR terms represent the auto regressive coefficients, MA terms represent the moving average coefficients, SAR terms represent the seasonal auto regressive coefficients, and SMA terms represent the seasonal moving average coefficients.

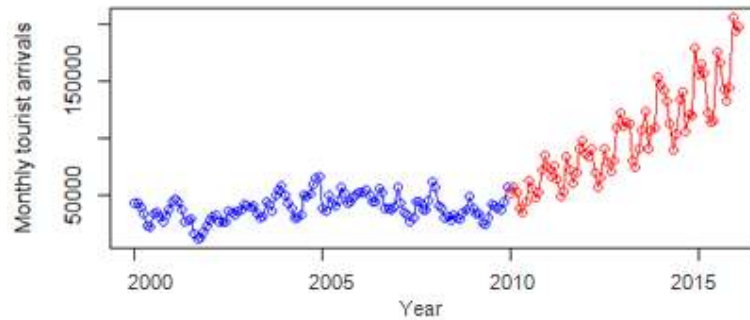


Fig. 1. Monthly tourist arrivalsof Sri Lanka (January 2000- February 2016)

Table 2. Parameters of $ARIMA(1, 1, 1)(2, 0, 2)_{12}$ model

	AR1	MA1	SAR1	SAR2	SMA1	SMA2
Coefficients	0.7425	-0.9184	0.3438	0.421	0.2461	-0.3064
Standard error	0.0962	0.0491	0.1843	0.0822	0.182	0.1618

σ^2 estimated as 0.01611; log likelihood=124.55, AIC=-231.74

3.1.2 Post-war analysis (January, 2010 - February, 2016)

Now we further proceed with the post-war data to obtain a suitable model that best describes the current situation of the industry by first applying ARIMA approach.

For this analysis 74 observations from January 2010 to February 2016 were used, and the time series plot is shown on Fig. 2. The non-stationarity of the data is visible, since the graph depicts a linear trend pattern and an obvious seasonal pattern. To make the data stationary, the first order difference ($d = 1$) and log

transformation were used. The Augmented Dickey-Fuller test confirmed ($p < 0.01$) the stationarity of transformed and differenced data. The parameters of the best fitted ARIMA model having the lowest AIC value are given in Table 3.

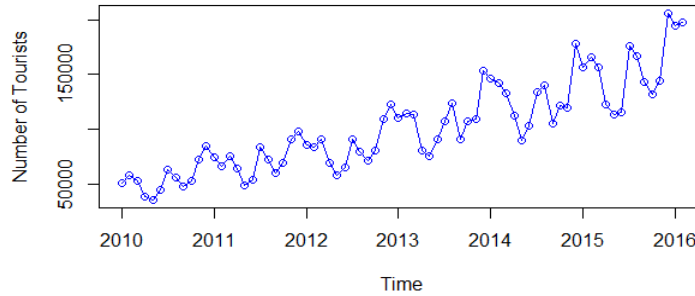


Fig. 2. Plot of Tourist arrivals: January 2010- February 2016

Table 3. The parameters of the obtained $ARIMA(3, 1, 0)(1, 0, 1)_{12}$ model

	AR1	AR2	AR3	SAR1	SMA1
Coefficients	-0.6376	-0.4829	-0.4299	0.97	-0.4228
Standard error	0.1194	0.1234	0.11	0.0217	0.1764

Sigma² estimated as 0.006713; log likelihood=66.72, AIC=-121.43

3.2 Multiplicative decomposition approach (MDA)

The MDA was applied to both the long-term data, and post-war data after adjusting the seasonality by applying the log transformation. The order 12-MA was used to estimate parameters of the model as described in section 2.2. The predicted values based on this method are shown in Table 4.

3.3 Comparison of forecasts

Using both ARIMA and MDA models, fitted for long term and post-war data, tourist arrivals were forecasted for the period from March, 2016 to August, 2016, and compared with actual values (Table 4). For both long term and post-war data analysis, MDA models have the higher predictive ability than ARIMA models since their MAPE values are lower than the respective ARIMA models. MDA model of the post-war data can be selected as the best fitted model in this case, since the MAPE value (6.6%) for this model indicates the highest accuracy for forecasting (see Table 1) than all other fitted models. These results are further confirmed graphically in Figs. 3 and 4. The blue and red colours in these figures indicate the actual and predicted values, respectively.

Since MDA forecasts of post-war data closely follow the actual ones with a high accuracy, the forecasts of the tourism demand for the next six months in 2016 are believed feasible and practical. This model is therefore suggested for further application in the Sri Lankan tourism industry. The finding of this study will be beneficial for the relevant industries and organizations such as the transportation facilities, airport management, restaurants and hotels, etc. They can be well prepared in advance to deliver their effectual services for their customers and make them satisfied with their visits.

3.4 Tourists' demands for months

Fig. 5(a) and Fig. 5(b) illustrate the plot of monthly seasonal component and box plots of monthly arrivals, respectively. Table 5 represents the monthly proportion of arrivals on annual total arrivals.

These monthly proportions show that there is a significant high demand for December, and the rest of the months have low and medium demands compared to December. According to Fig. 5(a) and 5(b), the twelve months can be categorized as in Fig. 6 with respect to their arrivals.

Table 4. Comparison of four models

2016 Month	Actual value	Forecast value			
		Long-term		Post-war	
		ARIMA	MDA	MDA	ARIMA
March	192,841	190,152	190,251	189,351	183,673
April	136,367	154,307	152,033	144,562	147,402
May	125,044	143,779	131,347	121,655	133,681
June	118,038	149,106	144,033	136,577	143,418
July	209,351	196,925	191,522	191,630	197,750
August	186,288	195,079	184,258	176,971	194,911
MAPE		11.0	8.25	6.6	8.5

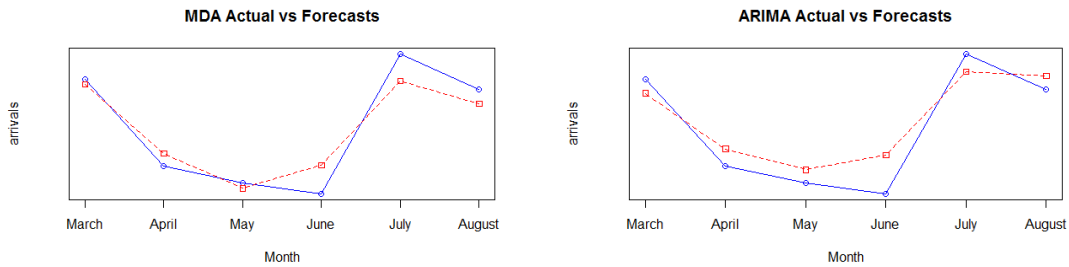


Fig. 3. Forecast of tourist arrivals: Post-war data

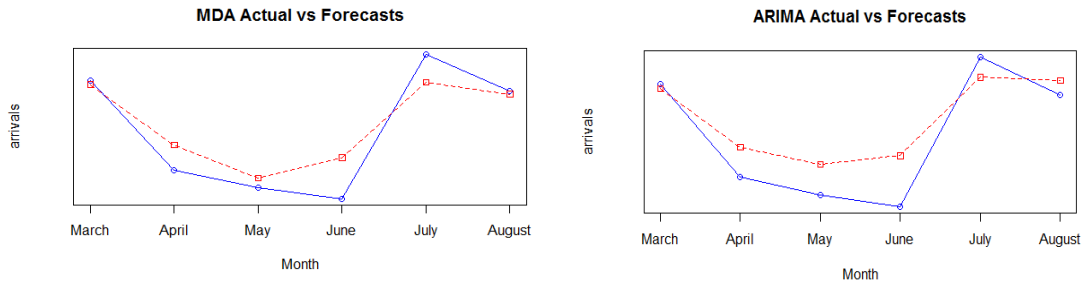


Fig. 4. Forecast of tourist arrivals: Long-term data

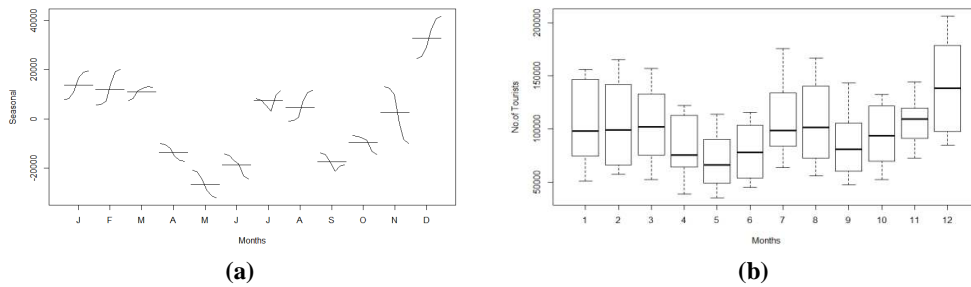


Fig. 5. Monthly seasonal effects

Table 5. Monthly proportion of tourist’s arrivals

YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC
2,010	0.08	0.09	0.08	0.06	0.05	0.07	0.10	0.09	0.07	0.08	0.11	0.13
2,011	0.09	0.08	0.09	0.07	0.06	0.06	0.10	0.08	0.07	0.08	0.11	0.11
2,012	0.09	0.08	0.09	0.07	0.06	0.06	0.09	0.08	0.07	0.08	0.11	0.12
2,013	0.09	0.09	0.09	0.06	0.06	0.07	0.08	0.10	0.07	0.08	0.09	0.12
2,014	0.10	0.09	0.09	0.07	0.06	0.07	0.09	0.09	0.07	0.08	0.08	0.12
2,015	0.09	0.09	0.09	0.07	0.06	0.06	0.10	0.09	0.08	0.07	0.08	0.11

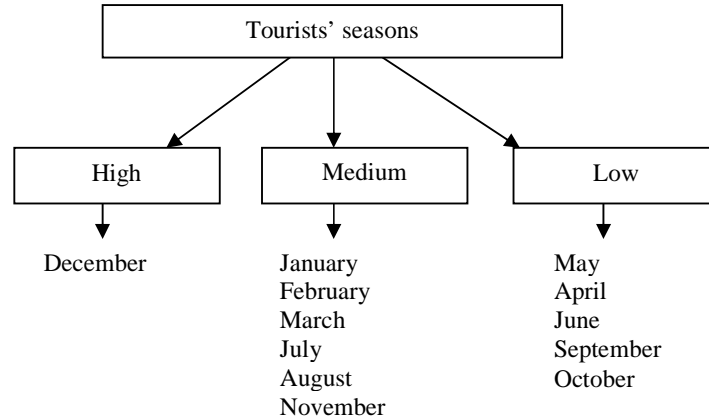


Fig. 6. Categorization of tourists’ demand

The results in Fig. 6 clearly show the tourists’ seasons and off-seasons of the year. This will be very useful for the groups who are seeking their arrival. This leads to enhance the quality of the services in peak-seasons, and attention can also be paid to off-seasonal offerings. Thus, recommendations can be made to minimize the off seasonal negative impact. One of the strategies may be to organize conventions and exhibitions during off peak periods. Further, forecasting of tourist arrivals can be utilized to develop an overall strategic plan specially focusing on the national level plan for sustainable tourism.

4 Conclusion

In this study an attempt has been made to forecast tourists’ arrivals using time series modelling by analyzing the data as long-term (2000-2016), and post-war (2010-2016) analysis. The two approaches, ARIMA and multiplicative decomposition, were used, and results revealed that the MDA model for post-war data follows better forecasts than long-term data with lower MAPE values. Among the two approaches MDA has the best fitted model with the highest accuracy to predict future tourist arrivals. The MAPE value of this model is less than the other models given in the literature ([8], [9]). The findings also indicated that the seasonality in tourist arrivals have not remained constant during the years, and the highest number of tourists are expected in the month of December based on three seasons (low, medium, high). This information can be used for a sustainable growth in the highly competitive tourism industry.

Competing Interests

Authors have declared that no competing interests exist.

References

- [1] Witt SF, Witt CA. Forecasting tourism demand: A review of empirical research. *International Journal of forecasting*. 1995;11:447–475.
- [2] Song H, Li G. Tourism demand modeling and forecasting. *Tourism Management*. 2008;29:203-220.
- [3] Chu F. Forecasting tourism: A combined approach. *Tourism Management*. 1998;9(6):515-520.
- [4] Akuno, et al. Statistical models for forecasting tourists' arrival in Kenya. 2015;5:60-65.
- [5] Tularam GA, Wong VSH, Nejad SAHS. Modeling tourist arrivals using time series analysis: Evidence from Australia. *Journal of Mathematics and Statistics*. 2010;8(3):348-360.
- [6] Song H, Witt SF. General-to-specific modeling to international tourism demand forecasting. *Journal of Travel Research*. 2003;42:65-74.
- [7] Kurukulasooriya N, Lelwala E. Time series behavior of burgeoning international tourist arrivals in Sri Lanka: The post-war experience. *Ruhuna Journal of Management and Finance*. 2014;1:1-14.
- [8] Konarasinghe KMUB. Forecasting tourist arrivals to Sri Lanka: Post-war period. *International Journal of Novel Research in Physics Chemistry & Mathematics*. 2016;3(1):57-63.
- [9] Gnanapragasam SR, Cooray TMJA. Forecasting post-war tourist arrivals to Sri Lanka using dynamic transfer function modeling method. *International Journal of Multidisciplinary Studies*. 2016;3(2):111-122.
- [10] Hyndma R, Athanasopoulos G. *Time series decomposition: Forecasting principles and practice*; 2014.
- [11] *Tourism research and statistics. Annual Statistical Report. Key Facts*; 2015.
Available:<http://www.sltda.lk/sites/default/files/>
- [12] Lewis C. *International and business Forecasting Methods*. London: Butterworths; 1982.

© 2017 Ishara and Wijekoon; This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Peer-review history:

The peer review history for this paper can be accessed here (Please copy paste the total link in your browser address bar)

<http://sciencedomain.org/review-history/17994>