FORECASTING SRI LANKAN TOURIST ARRIVALS: A COMPARATIVE STUDY OF HOLT- WINTER'S METHOD VERSUS ARIMA MODEL

***A.M.F. Roshan*[1] *and A. Jahufer*[2]**

[1, 2] *Department of Mathematical Sciences, South Eastern University of Sri Lanka, Sri Lanka*

roshanmajeedfas@gmail.com

ABSTRACT: *In this research study, the approach of Holt- Winter's Method and Seasonal Autoregressive Integrated Moving Average* ($SARIMA$) *method were implemented to forecast tourist arrivals in Sri Lanka. In this case, Sri Lankan monthly tourist arrivals data from January 2000 to December 2017 was considered. In the modelling implementation, data was analysed based on the two types of data such as long-term (2000-2017) and post-war (2010-2017). Because of the Sri Lankan Civil War ended in 2009, the data were categorized into two types. After the Sri Lankan civil war, tourist arrivals have increased annually. For that, forecasting Sri Lankan tourist arrivals is a necessary topic to build policy resolutions to enlarge conveniences plus additional interconnected issues in this industry. The first order difference data was concerned to make the data as stationary for the ARIMA approach. The best Holt- Winter's model was selected based on the least Root Mean Square Error (RMSE) and Mean Absolute Deviation (MAD) values meanwhile the best* $SARIMA$ *model was selected based on the minimum Akaike Information Criterion* (AIC) *value. The required statistical analysis was performed using Solver tool in Excel, Eviews9 and Minitab-16 software at 5% of significance level. The results reveal that for the long-term and post-war period,* $ARIMA$ *(3, 1, 2) (1, 0, 1)*$_{12}$ *and* $ARIMA$ *(2, 1, 3) (1, 0, 0)*$_{12}$ *are the suitable models respectively. Among the two approaches,* $ARIMA$ *(2, 1, 3) (1, 0, 0)*$_{12}$ *for post-war is the best model to sketch and to forecast the monthly tourist arrival pattern in Sri Lanka since having the least RMSE and MAD with a very precise extent by it satisfies the model assumptions. As well as, it indicates that forecasted and actual tourist arrivals are not much deviated from each other.*

Keywords: *Forecasting, HEGY test, Holt- Winter's Method, SARIMA, Tourist arrival*

1. INTRODUCTION

Tourism for pleasure has become a worldwide phenomenon since people are attracted by natural surroundings, beauty, place, culture and religious things. In Sri Lanka, travel and tourism directly contribute to GDP is 1.3% and totally contribute to GDP is 2.5% (Priyangika, J.H., Pallawala, P.K.B.N.M. & Sooriyaarachchi, D.J.C., 2016). Tourism is one of the earnings producing industries in a developing country which openly provide to the financial system.

A study has conducted for forecasting tourists' arrival in Kenya using the Double Exponential Smoothing method and the ARIMA model. The study has revealed that Double Exponential Smoothing model performed best since its Mean absolute percentage error (MAPE) and RMSE values were the minimum when compared to the ARIMA model (Akuno, A.O., Otieno, M.O., Mwangi, C.W. & Bichanga, L.A., 2015).

Another study has aimed at forecasting foreign tourists' arrival in India and to compare Holt-Winters Exponential Smoothing and ARIMA based on MAPE and RMSE. On the basis of results obtained ARIMA model is the best-fitted model for forecasting the tourist arrival in India (Sood, S. & Jain, K., 2017).

A comparative study has done for the time series behaviour of tourist arrivals using both additive and multiplicative decomposition techniques and it has revealed that multiplicative decomposition model is the best to forecast Sri Lankan tourists' arrivals due to the least MAPE value (Kurukulasooriya, N. & Lelwala, E., 2014).

Another research reveals that by considering the post-war period there was a significant increase in tourist arrivals in Sri Lanka and Holt Winter's three parameter model is the best one to forecast future arrivals (Konarasinghe, K.M.U.B., 2016).

A further comparative study shows the ARIMA model is the better predictor than VAR model for the tourism demand in Australia (Tularam, G.A., Wong, V.S.H. & Nejad, S.A.H.S., 2012).

Another comparative study has done for forecasting the tourism demand in ten countries (Japan, South Korea, Taiwan, Hong-Kong, Philippines, Indonesia, Singapore, Thailand, Australia and New Zealand) using Six types of time series approaches such as Naïve I, Naïve II, Linear Trend, Sine Wave, Holt-Winters and ARIMA. It has identified the ARIMA model as the best suitable model for prediction for nine of the ten countries with the minimum MAPE value (Chu, F., 1998).

The main objective of this study is to forecast tourist arrivals in Sri Lanka by using Holt-Winter's Method and $SARIMA$ Model. Further comparative analysis of both the methods is done on the basis of certain performance metrics. Furthermore, a comparative study of the both Holt-Winter's Method and $SARIMA$ Model is done to recommend the best method to forecast the upcoming Sri Lankan tourist arrivals.

2. METHODOLOGY

Monthly tourist arrivals of Sri Lanka from January 2000 to June 2018 were collected and the data was gained from the Annual Statistical Report of tourism research and statistics (Sri Lanka Tourism Development Authority, 2018). In this paper, the data from January 2000 to December 2017 were analysed and the forecasted period is from January – June of 2018. Holt-Winter's Method and $SARIMA$ Model have used as the optimal models for the forecasting of the number of Sri Lankan tourist arrivals from January – June of 2018. Furthermore, a comparative study of the both Holt-Winter's

Method and $SARIMA$ Model is done to recommend the best method to forecast the upcoming Sri Lankan tourist arrivals.

2.1 Holt-Winter's Method

Exponential Smoothing is one of the techniques of forecasting that inspires historical outlines such as trends and seasonal patterns into the forthcoming days. An Exponentially Weighted Moving Average mentions to a weighted moving average of the series in which the weights decrease exponentially.

Single Exponential Smoothing method is used for short-term forecasting a time series, generally just one period into the future when there is no trend or seasonal pattern, but the mean (or level) of the time series Y_t is slowly changing over time.

The form of the model is:

$$F_{t+1} = \alpha Y_t + (1-\alpha)F_t \qquad [1]$$

Where α is the smoothing constant between 0 and 1 and the best value of α matching to the least Mean Square Error (MSE) is generally used.

Double Exponential Smoothing method is used when the time series shows an increasing or decreasing trend. This method works like Single Exponential Smoothing excluding that level and trend should be added for each period.

The Double Exponential Smoothing method is given in equations [2].

$$\left.\begin{aligned} L_t &= \alpha Y_t + (1-\alpha)(L_{t-1} + T_{t-1}) \\ T_t &= \beta(L_t - L_{t-1}) + (1-\beta)T_{t-1} \end{aligned}\right\} \qquad [2]$$

$$F_{t+k} = L_t + T_t k$$

Where L_t is the level of the series at time t and T_t is the Trend of the series at time t, α and β $(0 \leq \alpha, \beta \leq 1)$ are the smoothing coefficients for level and trend respectively. Both α and β are selected based on the minimum MSE.

Holt-Winter's Method is an exponential smoothing approach for handling seasonal series with a trend. For the seasonality, another parameter should be added in addition to the parameters in Double Exponential Smoothing method. According to the seasonality style, two types of Holt-Winter's methods are designed for time series. Such as Additive Holt-Winter's method is used for time series with constant (additive) seasonal variations and Multiplicative Holt-Winter's method is used for time series with increasing (multiplicative) seasonal variations. This method is the favoured forecasting technique by most of the statisticians.

The additive Holt-Winter's Method can be clarified as:

$$\left.\begin{aligned} L_t &= \alpha(Y_t - S_{t-M}) + (1-\alpha)(L_{t-1} + T_{t-1}) \\ T_t &= \beta(L_t - L_{t-1}) + (1-\beta)T_{t-1} \end{aligned}\right\}$$

[3]

$$S_t = \gamma(Y_t - L_t) + (1-\gamma)S_{t-M}$$
$$F_{t+k} = L_t + T_t k + S_{t-M+k}$$

Where L_t is the level of the series at time t and T_t is the Trend of the series at time t α,β and γ $(0 \leq \alpha, \beta, \gamma \leq 1)$ are the smoothing coefficients for level, trend and seasonal component smoothing coefficient respectively. And M is the number of seasons in a year ($M = 12$ for monthly data, and $M = 4$ for quarterly data). α,β and γ are selected based on the minimum MSE.

As well as the multiplicative Holt-Winter's Method can be clarified as:

$$\begin{aligned} L_t &= \alpha(Y_t / S_{t-M}) + (1-\alpha)(L_{t-1} + T_{t-1}) \\ T_t &= \beta(L_t - L_{t-1}) + (1-\beta)T_{t-1} \qquad [4] \\ S_t &= \gamma(Y_t / L_t) + (1-\gamma)S_{t-M} \end{aligned}$$

$$F_{t+k} = (L_t + T_t k) * S_{t-M+k}$$

The coefficients for smoothing is the very important thing of the Exponential Smoothing Methods. In determining the smoothing coefficients for the level, trend and seasonality the Solver tool in Excel were used. And the best group of smoothing coefficients were chosen based on the minimum MSE.

2.2 Auto Regressive Integrated Moving Average Method (ARIMA)

Refer a process $\{y_t\}$ is supposed to be $ARIMA\ (p,d,q)$; $\{e_t\} \sim WN(0,\sigma^2)$: where e_t follows a white noise.

The $ARIMA\ (p,d,q)$ model can be clarified as:

$$dy_t = \alpha_0 + \alpha_1 dy_{t-1} + \cdots + \alpha_p dy_{t-p} + \beta_1 e_{t-1} + \cdots + \beta_q e_{t-q} + e_t \qquad [5]$$

The characteristic equation of $ARIMA$ model is:

$$\emptyset_p(B)\nabla^d y_t = \alpha_0 + \theta_q(B)e_t \qquad [6]$$

The $SARIMA(p,d,q)(P,D,Q)_{12}$ model can be clarified as:

$$dy_t = \alpha_0 + \alpha_1 dy_{t-1} + \cdots + \alpha_p dy_{t-p} + \beta_1 Dy_{t-1s} + \cdots + \beta_P Dy_{t-Ps} + \gamma_1 e_{t-1} + \cdots + \gamma_q e_{t-q} + \delta_1 e_{t-1s} + \cdots + \delta_Q e_{t-Qs} + e_t \qquad [7]$$

The characteristic equation of the $SARIMA$ model is:

$$\emptyset_p(B)\,\Phi(B^S)(1-B)^d(1-B^S)^D y_t = \alpha_0 + \theta_q(B)\boldsymbol{\Theta}(B^S)e_t \qquad [8]$$

The Autocorrelation Function (ACF) and Partial Autocorrelation Function $PACF$ plots were applied to find the non-seasonal and seasonal autoregressive and moving average terms $(p, q, P,$ and $Q)$ respectively. Non-seasonal unit root $(d$- non-seasonal difference) was verified using Augmented Dickey–Fuller (ADF) test and Kwiatkowski–Phillips–Schmidt–Shin $(KPSS)$ test and seasonal unit root was verified $(D$- seasonal difference) using Hylleberg-Engle-Granger-Yoo $(HEGY)$ test. The best $SARIMA$ model was selected based on the minimum AIC value. Required statistical analysis was performed using Eviews 9 and Minitab 16 software at 5% of significance level.

3. RESULTS AND DISCUSSION

3.1 Time Plot of Monthly Tourists Arrivals in Sri Lanka

From Figure 1, the monthly tourist arrival (TA) to Sri Lanka from January 2000 to December 2017 shows a non-stationary pattern because of the upward trend with the seasonal pattern. And also, the increment of tourist arrival after the Sri Lankan civil war is shown in Figure 2.

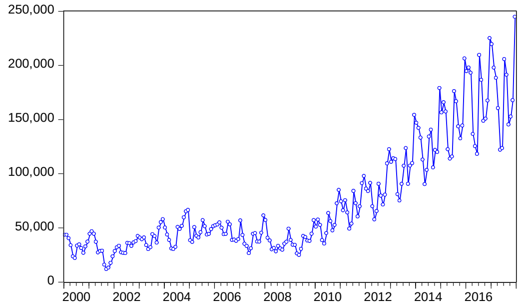


Figure 1. Monthly tourist arrivals (2000-2017)

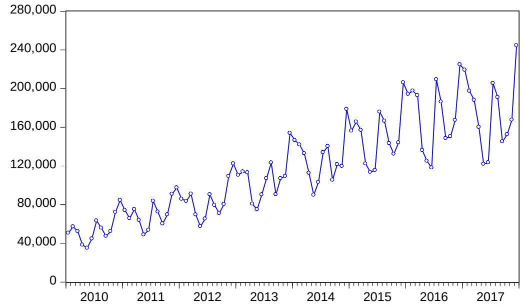


Figure 2. Post-war tourist arrivals (2010-2017)

3.2 Holt-Winter's Method

Holt-Winter's method is suitable when a trend is present in the seasonal time series data. It decomposes the data into three parts such as level, trend and seasonal factor. Both additive and multiplicative models were considered in Holt-Winter's method. The smoothing coefficients corresponding to both models with RMSE values are delineated in Table 1. According to the results, the multiplicative model of the Holt-Winter's method was selected as the best model in the 2000-2017 period with the smoothing constants that $\alpha=0.17$, $\beta=0.04$, $\gamma=0.50$. But in the post-war period, the additive model was the best model and the smoothing coefficients are $\alpha=0.81$, $\beta=0.36$, $\gamma=1$. As a final point the observed and forecasted tourist arrivals are shown in Table 2.

Table 1. Estimates of smoothing coefficients of the Holt-Winter's models

Model	(2000-2017)		(2010-2017)	
	Estimates of smoothing coefficients	RMSE	Estimates of smoothing coefficients	RMSE
Additive	$\alpha=0.45$, $\beta=0.14$, $\gamma=1$	0.2654	$\alpha=0.81$, $\beta=0.36$, $\gamma=1$	0.2160
Multiplicative	$\alpha=0.17$, $\beta=0.04$, $\gamma=0.50$	0.1988	$\alpha=0.75$, $\beta=0.19$, $\gamma=1$	2.2665

Table 2. Forecast of tourist arrival in Sri Lanka using Holt-Winter's method

Month in 2018	Observed tourists arrival	Forecasted tourists arrival	
		(2000-2017)	(2010-2017)
January	238924	225140	196511
February	235618	215902	185939
March	233382	207996	199427
April	180429	164015	185985
May	129466	135729	191956
June	146828	139101	150499

3.3 ARIMA approach

3.3.1 Non-Seasonal and Seasonal Unit Root Test

The log transformation was done to keep the seasonal effect constant. The data was transformed into stationary by concerning the first order difference. And the unit root test confirmed the stationarity of first-order differenced data at the 5% significance level.

Table 3. ADF stationary test for tourist arrival

Log(TA)	(2000-2017)		(2010-2017)	
	p-value		p-value	
	Level	First difference	Level	First difference
Constant only	0.9464	0.0004	0.0002	0.0000
Constant + Trend	0.4817	0.0021	0.9999	0.0000

According to the $HEGY$ test, the null hypothesis of the unit root at the seasonal frequency can be rejected at the 5% significance level. Hence, there is not necessary to make differences for data at the seasonal level.

Table 4. HEGY test for seasonal unit root

Auxiliary Regression	Seasonal Frequency	(2000-2017)		(2010-2017)	
		Constant	Constant + Trend	Constant	Constant + Trend
t-test :					
$\pi_1 = 0$	0	-0.397	-1.502	-1.394	0.411
$\pi_2 = 0$	$\frac{\pi}{3}$	-3.190	-3.183	-1.958	-1.921
F-test					
$\pi_3 = \pi_4 = 0$	$\frac{2\pi}{3}$	0.676	0.679	1.006	0.921
$\pi_5 = \pi_6 = 0$	$\frac{2\pi}{5}$	2.773	2.739	0.369	0.356
$\pi_7 = \pi_8 = 0$	2π	6.951	6.720	0.211	0.132
$\pi_9 = \pi_{10} = 0$	$\frac{\pi}{2}$	9.939	9.846	2.896	2.864
$\pi_{11} = \pi_{12} = 0$	π	1.215	1.206	0.398	0.384

3.3.2 Model Identification

Figure 3 and 4 reveal that the both ACF and $PACF$ lag 2, 3 and 6 are statistically significant. Therefore, this is a $SARIMA$ process.

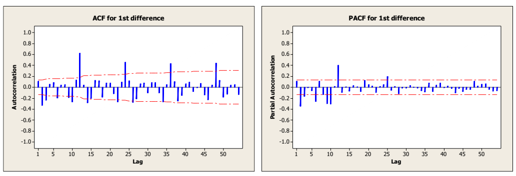


Figure 3. ACF and PACF of tourist arrivals (2000-2017)

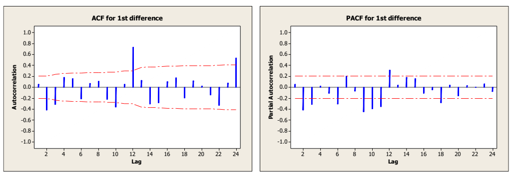


Figure 4. ACF and PACF of tourist arrivals (2010-2017)

3.3.3 Model Estimation

The estimated parameters for the best model having the least AIC values are exposed in Table 5 and 6.

Table 5. Estimated parameters for ARIMA (3, 1, 2) (1, 0, 1)12 (2000-2017)

	C	AR(1)	AR(2)	AR(3)	MA(1)	MA(2)	SAR(1)	SMA(1)
Coeff.	0.0097	1.7060	-0.8792	0.0333	-1.7612	0.8272	0.9947	-0.8825
Std. Error	0.0255	0.1262	0.1546	0.0773	0.1170	0.1051	0.0098	0.1027

σ^2 = 0.0153; Log likelihood = 123.4545, AIC=-1.1389

$Fitted\ SARIMA\ model\ is$:

$$dy_t = 0.0097 + 1.7060dy_{t-1} - 0.8792dy_{t-2} + 0.0333dy_{t-3} + 0.9947Dy_{t-12} - 1.7612e_{t-1} + 0.8272e_{t-2} - 0.8825e_{t-12} + e_t$$

Table 6. Estimated parameters for ARIMA (2, 1, 3) (1, 0, 0)12 (2010-2017)

	C	AR(1)	AR(2)	MA(1)	MA(2)	MA(3)	SAR(1)
Coefficients	0.0082	-1.9837	-1.0000	1.2658	-0.4238	-0.7178	0.9430
Std. Error	0.0220	0.0027	0.0025	0.0017	0.0007	0.0019	0.0303

σ^2 = 0.0049; Log likelihood = 84.3057, AIC=-1.8841

$Fitted\ SARIMA\ model\ is$:

$$dy_t = 0.0082 - 1.9837dy_{t-1} - 1.0000dy_{t-2} + 0.9430Dy_{t-12} + 1.2658e_{t-1} - 0.4238e_{t-2} - 0.7178e_{t-3} + e_t$$

3.3.4 Forecasting

Figure 5 shows the forecasted outcomes of the tourist arrivals over the period of January 2018 to June 2018 at 5% significance level and it indicates that forecasted and actual tourist arrivals are not much deviated from each other when performing the ARIMA process.

Table 7. Forecast of tourist arrival in Sri Lanka using ARIMA

2018 Month	Observed tourists arrival	Forecasted tourists arrival	
		ARIMA (3, 1, 2) (1, 0, 1)12 (2000-2017)	ARIMA (2, 1, 3) (1, 0, 0)12 (2010-2017)
January	238924	231636	214085
February	235618	222488	209161
March	233382	218670	186062
April	180429	173425	171028
May	129466	149322	124478
June	146828	156686	132804

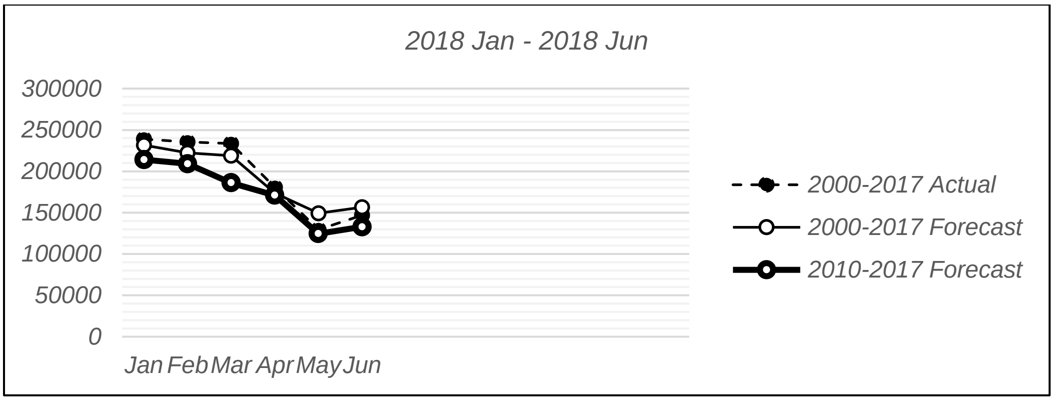


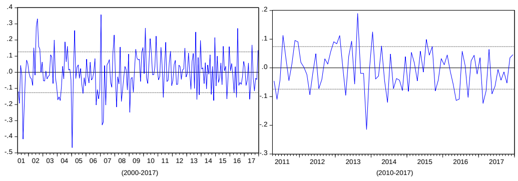
Figure 5. Actual vs Forecast

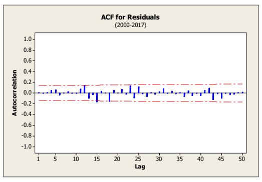
3.3.5 Model Evaluation

To check the model adequacy, the Autoregressive Conditional Heteroscedastic- Lagrange Multiplier (ARCH-LM) test was performed to check the heteroscedasticity and Table 8 shows that there is no homoscedasticity in the residuals of the model. The stationarity of the residuals for the model also has been confirmed via the Figures 6 and 7. Normality test results represent that the residuals are normally distributed. These results reveal that the modal is suitable to forecast the upcoming tourist arrivals in Sri Lanka by it satisfies the model assumptions.

Table 8. ARCH-LM Test for Heteroskedasticity

Period	Model	Prob. Chi-Square
(2000-2017)	(3, 1, 2) $(1, 0, 1)_{12}$	0.0845
(2010-2017)	(2, 1, 3) $(1, 0, 0)_{12}$	0.1320





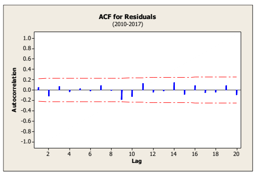


Figure 6. Time series plot of Residuals *Figure 7. ACF for Residuals*

3.4 Model Comparison on Forecasting

The comparison of forecasted tourist arrival in Sri Lanka using ARIMA and Holt-Winter's method is shown in Table 9. According to the RMSE and MAD values, for both the 2000-2017 period and post-war period ARIMA model has more forecasting capacity than Holt-Winter's method. ARIMA (2, 1, 3) $(1, 0, 0)_{12}$ model for the post-war period can be chosen as the best model to forecast tourists arrival in Sri Lanka since the both RMSE and MAD values are minimum.

Table 9. Comparison of Forecasted tourist arrival in Sri Lanka using ARIMA and Holt-Winter's method

Month in 2018	Observed tourists arrival	Forecasted tourists arrival			
		(2000-2017)		(2010-2017)	
		ARIMA (3, 1, 2) (1, 0, 1)$_{12}$	Holt-Winter's method	ARIMA (2, 1, 3) (1, 0, 0)$_{12}$	Holt-Winter's method
January	238924	231636	225140	214085	196511
February	235618	222488	215902	209161	185939
March	233382	218670	207996	186062	199427
April	180429	173425	164015	171028	185985
May	129466	149322	135729	124478	191956
June	146828	156686	139101	132804	150499
RMSE		0.1237	0.1988	0.0703	0.2160
MAD		0.0939	0.1353	0.0577	0.1691

4. CONCLUSION

In this study, monthly tourist arrivals in Sri Lanka from 2000 to 2017 were considered in two types of categorized data as (2000-2017) and post-war (2010-2017), to forecast tourist arrivals by using time series approach. The two types of models, $SARIMA$ and Holt-Winter's method were applied to both categorized data. The results reveal that for the long-term and post-war period, $ARIMA$ (3, 1, 2) (1, 0, 1)$_{12}$ and $ARIMA$ (2, 1, 3) (1, 0, 0)$_{12}$ are the suitable models respectively. Among the two approaches, $ARIMA$ (2, 1, 3) (1, 0, 0)$_{12}$ for post-war is the best model to sketch and to forecast the monthly tourist arrival pattern in Sri Lanka with a very precise extent by it satisfies the model assumptions. The both RMSE and MAD values of this model are lower than the other models. And, it indicates that forecasted and actual tourist arrivals are not much deviated from each other.

REFERENCES

Akuno, A.O., Otieno, M.O., Mwangi, C.W. & Bichanga, L.A. (2015). Statistical Models for Forecasting Tourists' Arrival in Kenya. Open Journal of Statistics, 5, 60-65.

Chu, F. (1998). Forecasting tourism: A combined approach. Tourism Management, 9(6), 515-520.

Konarasinghe, K.M.U.B. (2016). Forecasting tourist arrivals to Sri Lanka: Post-war period. International Journal of Novel Research in Physics Chemistry & Mathematics, 3(1), 57-63.

Kurukulasooriya, N. & Lelwala, E. (2014). Time series behavior of burgeoning international tourist arrivals in Sri Lanka: The post-war experience. Ruhuna Journal of Management and Finance, 1(1), 1-14.

Priyangika, J.H., Pallawala, P.K.B.N.M. & Sooriyaarachchi, D.J.C. (2016). Modelling and Forecasting Tourist Arrivals in Sri Lanka. Symposium on Statistical & Computational Modelling with Applications, (pp. 14-18).

Sood, S. & Jain, K. (2017). Comparative Analysis of Techniques for Forecasting Tourists' Arrival. Journal of Tourism & Hospitality, 6(3), 285-289.

Sri Lanka Tourism Development Authority. (2018). Retrieved August 21, 2018

Tularam, G.A., Wong, V.S.H. & Nejad, S.A.H.S. (2012). Modeling Tourist Arrivals Using Time Series Analysis: Evidence From Australia. Journal of Mathematics and Statistics, 8(3), 348-360.