Prediction of international tourist arrivals to Sri Lanka using state space modelling method: An empirical study after the civil war

SR Gnanapragasam[1#] TMJA Cooray[2] and R. A. Dissanayake[2]

[1] Department of Mathematics and Computer Science, The Open University of Sri Lanka, Sri Lanka
[2] Department of Mathematics, University of Moratuwa, Sri Lanka
[#] Corresponding author; <srgna@ou.ac.lk>

Abstract— *Tourism plays a big role in development of a country in terms of economics. It contributes in large scale to the national revenue in Sri Lanka through foreign exchange. Tourists had less interest of visiting Sri Lanka, mainly due to the uncertainty of security during the civil war. Previous studies identified some models to forecast the arrivals to Sri Lanka but none of them have used state space modelling method. In Sri Lanka the war is over now and the security is certain in all over the island. Thus the tourist arrivals are getting increase in recent past. Therefore, the objective of the study is to fit the model to forecast the international tourist arrivals, after the civil war, to Sri Lanka by using state space modelling method.*

Prior to model fittings, preliminary testing is carried out to study the behaviour of the series. Augmented Dickey- Fuller (ADF) and Kruskal- Wallis tests are applied to confirm the stationary of the series. In state space modelling, Kalman filter technique is used to estimate the parameters of the model. Diagnostic checking for the fitted model is also carried out. To check the accuracy of fitted model mean absolute percentage error (MAPE) statistics is employed.

After the civil war a dramatic increase is clearly observed with positive growth rates every year. Until March 2016, the total arrivals are over 7.961 million. The MAPE statistics of the fitted state space model is 5.95%. By ex-post forecast, the estimated arrivals for the next 3 quarters of the year 2016 is approximately 1.613 million and which is nearly 22.2% increase in growth with the year 2015. Therefore, the fitted model can be used to predict the future arrivals to Sri Lanka and accordingly the policy maker can take necessary action to cater them in future.

Keywords— **Civil war, State space, Tourist arrivals**

I. INTRODUCTION

Tourism is one of the largest and fastest-growing economic sectors in the world, over the past six decades. Its contribution to Gross Domestic Product (GDP) by foreign exchange generation is very large. Forecasting plays a major role in tourism planning. Prediction of tourist arrivals is essential for planning, policy making and budgeting purposes by tourism operators. Sri Lanka entered the international tourism market in the year 1960s. Since then, government involvement has been a key factor in tourism development in Sri Lanka. There were several set-backs during last few decades in the tourism development process such as global economic crisis in 2009, Tsunami in 2004 and civil war from years 1983 to 2009.

Previously two studies related to forecasting tourist arrivals to Sri Lanka are available. First is a comprehensive study of time series behaviour of the post war international tourist arrivals to Sri Lanka by Kurukulasooriya & Lelwala (2014). In the modelling exercise, classical time series decomposition approach is employed. The second is by Kurukulasooriya & Lelwala (2011) to select the most accurate time series model for forecasting international tourist arrivals to Sri Lanka and to generate short – term forecasts using exponential smoothing models.

As per the literature, using state space modelling method, a model has been fitted by Athanasopoulos & Hyndman (2006) to forecast Australian domestic tourism demand. It shows that these models outperform alternative approaches for short-term forecasting and also produce sensible long-term forecasts. However, Song & Li (2008) provides comprehensive reviews of published studies on tourism demand modelling and forecasting since 2000. It shows that there is no single model that consistently outperforms other models in all situations. In addition, there is no study has carried out to forecast tourist arrivals to Sri Lanka using state space modelling method.

In Sri Lanka the civil war is over now and the tourist arrivals is getting increase in recent past. Therefore, the objective of the study is to fit the model to forecast the international tourist arrivals, after the civil war, to Sri Lanka by using state space modelling method.

II. METHODOLOGY

The data for the study is extracted from the annual statistical reports of the Sri Lanka Tourism Development Authority.

A. Preliminary testing

Prior to model fittings, the following techniques are carried out. For this purpose the monthly arrivals from June 2009 to December 2015 are used.

1) Time series plot: Plot of time series is used to inspect the data and its behaviour such as trend, seasonality, cyclic pattern or irregular variations as elements of non-stationary.

2) Augmented Dickey- Fuller (ADF) test: ADF test is used to check whether the series has a unit root. That is to check the stationary of series in terms of trend availability.

Test statistics for the model $Y_t = \rho Y_{t-1} + u_t$ is

$$DF = \frac{\hat{\rho}}{SE(\hat{\rho})} \sim t_{n-1} \quad \text{for} \quad -1 < \rho < 1$$

where u_t is the white noise and n is the number of observations.

3) Kruskal- Wallis test: To confirm the seasonality in the series this test is used. The test statistics of Kruskal- Wallis test is:

$$H = \frac{12}{N(N+1)} \sum_{i=1}^{n_i} \frac{R_i^2}{n_i} - 3(N+1) \sim \chi_{L-1}^2$$

where N is the total number of rankings, R_i is the sum of the rankings in a specific season, n_i is the number of rankings in a specific season and L is the length of season.

4) Differencing method: Method of differencing is used to transform the non- stationary series to stationary series and it is defined as $W_t = Y_t - Y_{t-L}$, where L is the length of season.

In time series analysis, a process of examining the Auto Correlation Function (ACF) and Partial Auto Correlation Function (PACF) is to determine the nature of the process under consideration.

5) Autocorrelation function (ACF): The first several autocorrelations are persistently large in the graph of ACF and trailed off to zero rather slowly, it can be assumed that a trend exists and the time series is non-stationary. If the series is stationary, then ACF graph must be decay exponentially. ACF at lag k is defined by

$$\rho_k = \frac{\operatorname{cov}\left[\left(Y_t - \hat{Y}_t\right)\left(Y_{t+k} - \hat{Y}_{t+k}\right)\right]}{\sqrt{\operatorname{var}\left(Y_t - \hat{Y}_t\right)\operatorname{var}\left(Y_{t+k} - \hat{Y}_{t+k}\right)}}$$

6) Partial autocorrelation function *(PACF):* PACF between Y_t and Y_{t+k} is the conditional correlation between Y_t and Y_{t+k} and defined as follows:

$$\phi_{kk} = corr\left(Y_t , Y_{t+k} \middle| Y_{t+1} , Y_{t+2} , Y_{t+k-1}\right)$$

The PACF between Y_t and Y_{t+k} is the autocorrelation between Y_t and Y_{t+k} after adjusting for $Y_{t-1}, Y_{t-2}, \ldots, Y_{t-k+1}$.

In this study, *MINITAB* is used to get the above test results and relevant graphs for preliminary testing.

B. Model development using State Space method

For the purpose of model development the monthly data from June 2009 to December 2015 are used.

State Space (SS) models are models of jointly stationary multivariate time series processes that have dynamic interactions and that are formed from two basic equations.

$$S_{t+1} = BS_t + Gw_t \ , \ w_t \sim N\left(0, \textstyle\sum_w\right)$$

The state transition equation consists of a state vector of auxiliary variables as a function of a transition matrix and an input matrix.

$$Y_t = HS_t + v_t \ , \ v_t \sim N\left(0, \textstyle\sum_v\right)$$

The measurement equation consists of a state vector canonically extracted from observable variables. These vector models are estimated with a recursive protocol and can be used for multivariate forecasting, where H , B and G are system matrices (known matrices), v_t and w_t are uncorrelated sequences of white noises.

C. Parameter estimation of State Space model

In practical applications, the system matrices H and B together with the variance-covariance matrices $\sum_v$ and $\sum_w$ are unknown and have to be estimated. Obviously, whenever the explanatory variables are not observable, least squares estimation is not a way to go. The Kalman filter allows constructing the likelihood function associated with a state space model. *Kalman filter* is a set of recursive equation that allows us to update the information in a state space model.

D. Forecasting using State Space modelling method

Having obtained the maximum likelihood estimates, it can be used the state space model to forecast the observables. In particular, the final state predictor implied by the maximum likelihood estimates together with the measurement and transition equation can be used to forecast the observations.

In this study, *SAS* programme is used to get the estimation of the parameters of SS models using Kalman filter and to get the forecasted values.

E. Diagnostic checking for fitted model

Before using the model for forecasting, it must be checked for adequacy. Diagnostic checks are performed to determine the adequate of the model.

1) Anderson- Darling (AD) test: AD test is used to test if a sample of data comes from a population with a specific distribution. Here the hypotheses are H_0: The data follow normal distribution versus H_1: The data do not follow normal distribution.

The test statistic of AD test is:

$$A^2 = -N - \sum_{i=1}^{N} \frac{(2i-1)}{N}\left[\ln F(Y_i) + \ln\left(1 - F(Y_{N+1-i})\right)\right]$$

where F is the cumulative distribution function of the specified distribution, Y_i are the ordered data and N is the total number of observations.

2) Lagrange's Multiplier (LM) test: LM test is used to test the auto correlation among residuals. The null hypothesis to be tested is that, H_0: there is no serial correlation of any order. The individual residual autocorrelations should be small. Significant residual autocorrelations at low lags or seasonal lags suggest the model is inadequate.

$$W = nR^2 \sim \chi^2_{df}$$

Where, df is the number of regressors in the auxiliary regression (only linear terms of the dependent variable are in the auxiliary regression), R^2 is the determination of coefficients and n is the number of observations.

3) White's General test: This test is used in order to check constant variance of residuals. Accordingly the null hypothesis is H_0: Homoscedasticity against the alternative hypothesis H_1: Heteroscedasticity.

Test statistics of White's General test is $W = nR^2 \sim \chi^2_{df}$

Where, df is the number of regressors in the auxiliary regression (squared terms of the dependent variable are also included in addition to terms in the LM test in auxiliary regression), R^2 is the determination of coefficients and n is the number of observations.

F. Accuracy of the fitted model

For the purpose of model accuracy the monthly arrivals from January 2016 to March 2016 are compared with the relevant predicted values.

It is important to evaluate performance of fitted model on the basis of the fit of the forecasting. Measure of forecast accuracy should always be evaluated as part of a model validation effort.

1) Mean Absolute Percentage Error (MAPE): MAPE is the average of the sum of the absolute values of the percentage errors. It is generally used for evaluation of the forecast against the validation sample. To compare the average forecast accuracy of different models, MAPE statistics is used. It is defined as follows:

$$MAPE = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{Y_t - \hat{Y}_t}{Y_t}\right| \times 100$$

Generally if MAPE is less than 10% then the fitted model is highly recommended for forecasting.

III. RESULTS AND DISCUSIONS

A. Preliminary analysis

1) *Behaviour of the series:* The Figure 1 shows the yearly tourist arrivals from the year 1967 to 2015 in Sri Lanka.

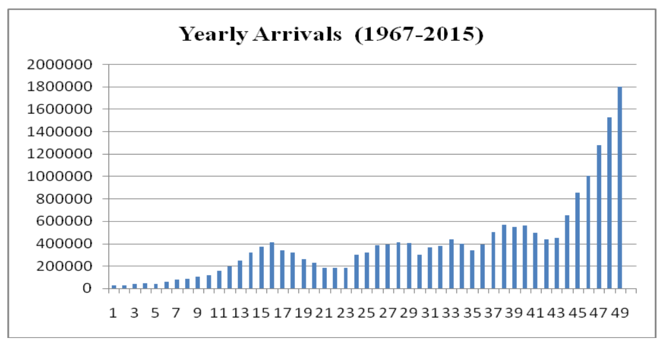

Figure 1. Plot of yearly tourist arrivals

From Figure 1, it is clearly observed that, up to first 16 points (from the years 1967 to 1983) there is an upward trend, there after ups and downs in total arrivals till 43rd point (that is from the years 1983 to 2009). Meantime, from points 43 to 49 (from the years 2009 to 2015), there is a remarkable upward increase in the total number of international tourist yearly arrivals to the island, after the civil war.

Table 1. Descriptive statistics of tourist arrivals

Year	Arrivals	Growth Rate	Year	Arrivals	Growth Rate
2009	260161	2.15%	**2013**	1274593	26.75%
2010	654476	46.12%	**2014**	1527153	19.81%
2011	855975	30.79%	**2015**	1798380	17.76%
2012	1005605	17.48%	**Total arrivals**		**7,376,343**

In this study, it is decide to consider the time frame of after the civil war from June 2009 to December 2015. As per the Table 1, the total number of international tourists had visited the island is 7,376,343 in this time frame. Also, it can be clearly observed that there is dramatic increase in yearly tourist arrivals in nearly six and half years. Also only in the year 2015 nearly 1.8 million tourists had visited the island and which is the biggest hit in tourism history of Sri Lanka.

Further it is noted a sudden change in growth rate soon after the war in 2010. Also a positive growth rate can be observed every year. It indicates that the international tourist show more interest of visiting Sri Lanka than in the past.

From the time series plot appeared in the Figure 2, it can be clearly observed a positive trend with a seasonal pattern. Further it is proved statistically by taking the p- values of ADF and Kruskal- Wallis tests.

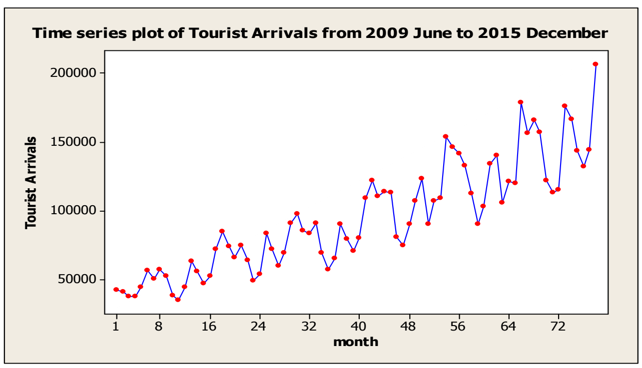

Figure 2. Tourist arrivals after the civil war

Since the p-value of ADF test is 0.99, it can be concluded with 95% confidence that there is a trend in the series. At the same time, the p- value of Kruskal- Wallis test (0.00) confirms the existence of seasonality in the series.

2) *Converting to stationary series:* The series of the time frame is named as Y in the rest of the part in this study.

Figure 3. ACF graph of the series Y

The ACF graph is not decaying exponentially with lags in Figure 3. It can be said that the series after the civil war is non-stationary at the same time there is a seasonal pattern suggests the existence of seasonality in the series. Thus the series has to be transformed as a stationary series before developing a model by removing trend and the seasonal pattern. Generally this transformation can be done by the method of differencing.

The series has a positive trend and hence it is non-stationary. To remove the trend, the regular difference is carried out. Accordingly, the new series is named as D1Y (the series after 1 difference).

The p- value of ADF test is 0.00 for the series DIY. Therefore it can be concluded with 95% confidence that, the D1Y series has no trend. Thus the ACF graph is obtained to check the existence of seasonality.

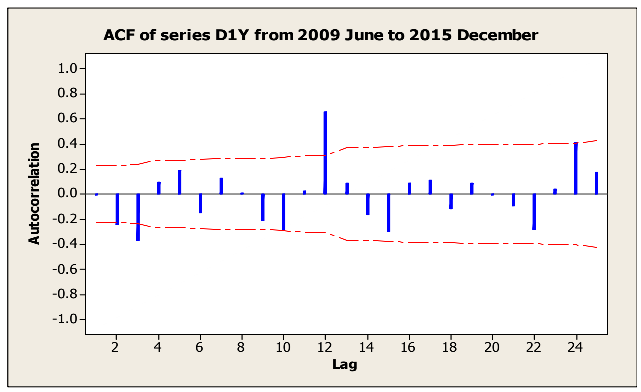

Figure 4. ACF graph of the series D1Y

From the ACF graph in Figure 4, it can be observed the high spikes at lags 12 and 24. It indicates that obviously there is a seasonal pattern in D1Y series with seasonality 12. Therefore to remove the seasonal pattern from the D1Y series, the 12^{th} difference is also taken for the series DIY and it is named as D12D1Y.

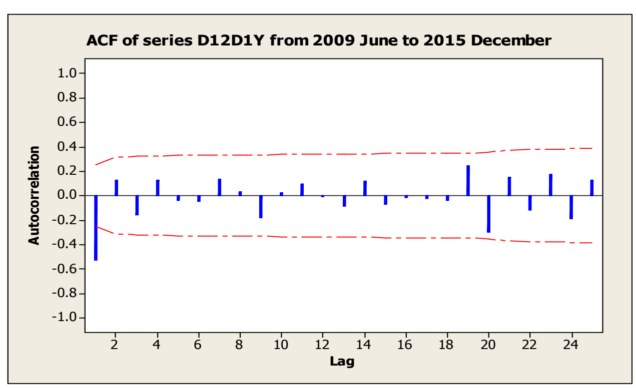

Figure 5. ACF graph of the series D12D1Y

Now it seems that the ACF in Figure 5 has no high spikes after the first spike and all others are significant. Thus it can be claimed that the D12D1Y has no seasonal pattern. Now it is necessary to check the PACF graph as well to further decide the stationary condition.

Since almost all the spikes (except few at beginning) are significant of PACF graph in Figure 6. This also suggests that the series D12D1Y is stationary. Moreover, from the p-value of Kruskal- Wallis test (0.536), it can be concluded that the series D12D1Y has no seasonality. Therefore, it can be confirmed that, the series D12D1Y is stationary.

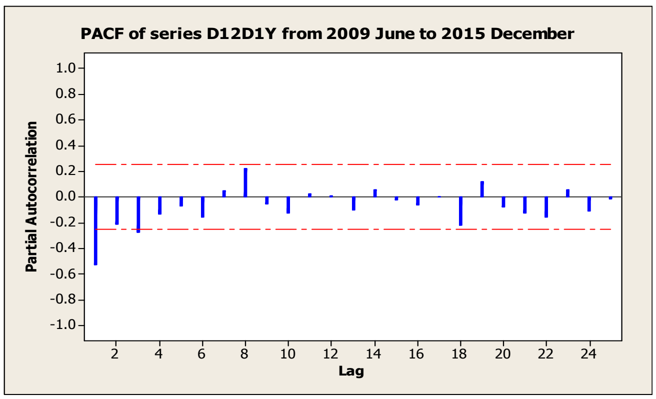

Figure 6. PACF graph of D12D1Y series

B. Model development using State Space method

After some iteration from preliminary estimates in *SAS*, the final estimated parameters of the selected fitted model are as follows in the Table 2:

Table 2: Estimates of parameters of SS model

Term	Parameter Estimates	t- value
Y_{t-1}	0.08	2.27
Y_{t-2}	-0.02	-3.24
e_{t-1}	-0.81	-5.95

From the t-values in the Table 2, it is observed by comparing with the critical t- value 1.96 (number of observations is more than 60 hence normal distribution value at 5% level of significant is considered), all three parameters are significant. Thus it can be concluded with 95% confidence that the parameters of SS models are significant. Thus the fitted model to forecast tourist arrivals is $Y_t = 0.08 * Y_{t-1} - 0.02 * Y_{t-2} - 0.81 * e_{t-1} + e_t$

C. Diagnostic checking for the fitted model

Since the graph of the normal probability is almost linear in the Figure 7, it is suggested that the residuals are normal. Further the p-value (0.780) of Anderson Darling test confirms that the residuals follow normal distribution.

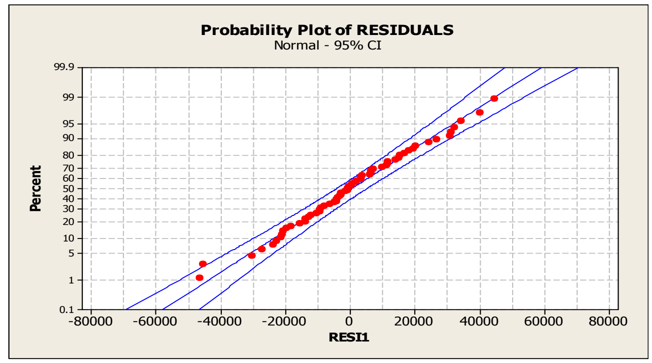

Figure 7. Normal probability plot of residuals

It can be seen from the plots of residuals versus predicted values in Figure 8, that the residuals are randomly scatted. It is suggested that the residuals are independent. Further, the p-value of LM test (0.64) confirms that the residuals have no serial correlation.

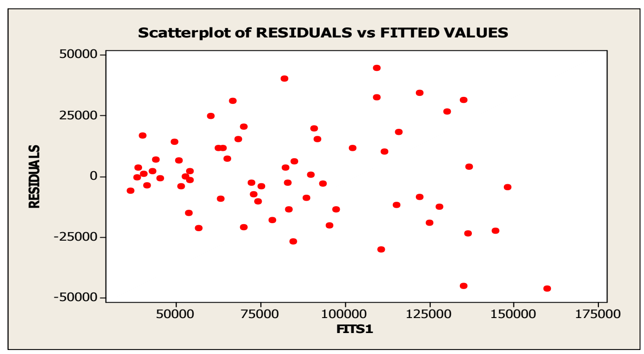

Figure 8. Plot of Residuals versus Predicted values

It is observed from the plot of residuals versus order of observations in Figure 9, that there is no any systematic pattern and points are symmetric about 0. It suggests that the residuals have constant variance. Further the p- value of White's General test (0.12) confirms that the heteroscedasticity does not exist among the residuals.

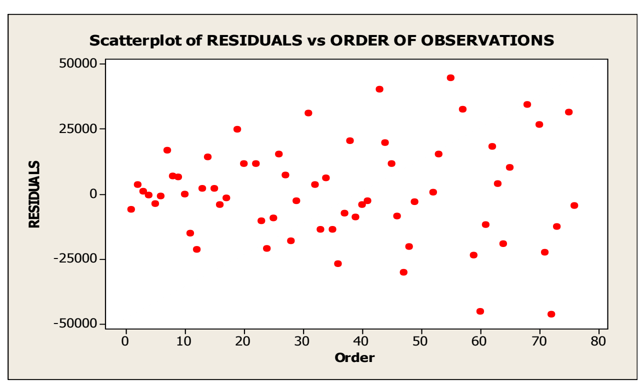

Figure 9. Plot of Residuals versus Orders

Therefore it can be concluded that, the fitted model using State Space method satisfies the diagnostic checking. Hence it can be concluded that the fitted model is adequate and can be used for forecasting purpose.

D. Accuracy of the fitted model

Based on the ex-ante forecast value in the Table 3, the calculated MAPE statistics is 5.95%. Hence, the accuracy of the fitted model is more that 94% indicates that, the model is better and it can be recommended for ex-post forecast.

Table 3. Ex-ante forecast

Month in 2016	**January**	**February**	**March**	**MAPE**
Observed	194280	197697	192841	5.95
Forecasted	179927	189355	180789	

IV. CONCLUSIONS

A. Trend of the series

A positive increasing trend in arrivals reveals that the international tourists show more interest to visit Sri Lanka. Therefore more development of the tourism industry will definitely increase the economics of Sri Lanka.

B. Fitted model to forecast the tourist arrivals

According to the model accuracy MAPE value (5.95), based on the ex-ante forecast, the fitted State Space model for the period of after the civil war is

$$\hat{Y}_t = 0.08 * Y_{t-1} - 0.02 * Y_{t-2} - 0.81 * e_{t-1}$$

where $\hat{Y}_t$ is the estimated tourist arrivals at time t

Y_{t-1} and Y_{t-2} are preceding arrivals at time $t-1$ and $t-2$ respectively

e_{t-1} is residual at one preceding period $t-1$

This model can be used for ex-post forecast the future arrivals of international tourists to Sri Lanka.

C. Ex-post forecast for last 3 quarters of the year 2016

The monthly-wise ex-post forecast for last 3 quarters of the year 2016 is summarised in the Table 4.

Table 4. Ex-post monthly forecast for the year 2016

Month in 2016	**Ex-post Forecast**	**Month in 2016**	**Ex-post Forecast**
April	153548.58	September	176228.87
May	145537.87	October	165440.68
June	147387.28	November	177616.66
July	208071.10	December	239892.03
August	199152.46	**Total**	**1,612,876**

From the Table 4, it can be concluded that in last 3 quarters of the year 2016 approximately 1.613 million international tourist arrivals can be expected to Sri Lanka. It is approximately about 22.2% increase in growth with the year 2015.

V. REFERENCES

Athanasopoulos, G. and Hyndman, R. J. (2006). Modelling and forecasting Australian domestic tourism. Preprint submitted to *Tourism Management,* Retrieved from: http://www.robjhyndman.com/papers/austourism.pdf.

Kurukulasooriya, N. and Lelwala, E. I. (2011). Short-term forecasting of international tourist arrivals to Sri Lanka: Comparative evidence using exponential smoothers. *The 6th International Research Conference on Management and Finance 2011*. Retrieved from: http://ssrn.com/abstract=2018372.

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

Song, H. and Lib, G. (2008). Tourism demand modelling and forecasting- A review of recent research. *Tourism Management, 23*, 203-220.