A Seasonal ARIMA Model of Tourism Forecasting: The Case of Sri Lanka

H. Rangika Iroshani Peiris
Department of Business management, Faculty of Business, Sri Lanka Institute of Information Technology, Malabe, PO box 10115, Sri Lanka

Abstract
Many scholars have attempted to forecast the tourist arrival series in different countries. The aim of this paper is to find a suitable SARIMA model to forecast the international tourist arrival to Sri Lanka. Monthly data of tourist arrival from January 1995 to July 2016 is used for the analysis. Seasonality in the data series is identified using the HEGY test. The Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), Mean Absolute percent Error (MAPE) are used to measure the forecasting accuracy. The result shows that the SARIMA (1, 0, 16) $(36, 0, 24)_{12}$ model is suitable to forecast the tourist arrival in Sri Lanka.
Keywords: Tourist Arrival, SARIMA model, HEGY test, Forecasting accuracy

1. Introduction

Sri Lanka formerly called as Ceylon is a beautiful island situated in the Indian Ocean southwest of the Bay of Bengal. Diverse landscapes range from sandy beaches to highlands, ancient Buddhist ruins, and indigenous wildlife species make the country a remarkable place in the global tourism industry. These resources together with the gentle hospitability of Sri Lankans brand the country as a memorable tourist destination. (Wickramasinghe & Takano, 2010). Further being a small island of 65,610 square kilometres added a value for the tourism in the country since it expedites a tourist to explore the whole country via a car in length and breadth within a short period. (Kodituwakku, Wijesundara, & Hettiarachchi, 2015)

In mid-1960s Sri Lanka started modern commercial tourism and only 18,969 tourist arrivals recorded by 1966. But this number reached to 209,351 in mid-year 2016. (SLTDA). However, the terrorist attacks which have been existing in the country since 1983 to 2009 have severely affected to the declines and growth of the tourist arrival. Though there is a growth in the tourism industry coping with the influence of terrorism for a long period time, the natural disaster “Tsunami” in year 2004 aggravated the situation. Further, the global economic crisis in 2008 also had a major impact to the country’s tourism industry.

After 26 years of civil war, by defeating tiger rebels in May 2009, the country became popular as a safe tourist destination and the tourist arrival increased significantly. Despite many issues and challenges, the Sri Lankan tourism industry has been experiencing a rapid development and expansion, to emerge as the leading and fastest booming economic sector of the post conflicts economy. In 2015 it becomes the third largest foreign exchange earner in the country. (Central Bank of Sri Lanka, 2015). In addition to the revenue generation, tourism in Sri Lanka positively contributes to generate employments for the youth and also creates business opportunities for local and foreign entrepreneurs. (Kodituwakku, Wijesundara, & Hettiarachchi, 2015)

Sri Lanka Tourism Development Authority (SLTDA) which was formally recognized as Ceylon Tourist Board was formed in 2005, as the centre for Sri Lanka Tourism. Having the intension of converting Sri Lanka to be Asia’s leading tourism hub, SLTDA aims to develop high quality, unique and diverse tourism products and services in the country.

International tourist arrivals has strong impact for the key decision makers in the business and government sectors of a country. To provide infrastructure for the rapid development of tourism and service industries, both public sector and the private sector need accurate forecasts of tourist arrival to the country. For the public sector, accurate forecasts of tourist arrivals are needed in order to make plan for tourism infrastructures, such as plans for hotel sites and plans for transportation development etc. Further, in private sector, to plan the aircraft flights, facilities for travellers and also to plan the human resources needed as hoteliers, tour operators, entertainers, good forecasting of tourist arrival is essential. (Chang & Liao, 2010)

The objective of the present research is to forecast the tourist arrival to Sri Lanka using the total number of tourist arrival for the period of 1995 January to July 2016. The next section describes the past research work relevant to the forecasting of tourist arrival. Then the methodology and the findings are discussed. Finally findings of the research with further research areas are discussed.

2. Literature review

Forecasting of tourist arrivals is a challenging process since it is severely affected by sudden changes in trends, changes in seasonal factors and unforeseen events such as natural disasters and wars.

Qualitative and Quantitative forecasting methods are the two main categories suggested by the existing forecasting literature. Qualitative methods use non-scientific techniques to forecast the future events. Quantitative methods use past information and apply mathematical rules to forecast the events considering underlying patterns and relationships in the data. Song and Li (2008) suggest that the time series methods and the econometric methods are most popular methods to forecast tourism demand considering the other models such as artificial neural network (ANN), the fuzzy time series (Fuzzy), genetic algorithms (GAs), and the rough set method.

2.1 Time series approaches

Time series approach considers the internal structure of the variable with respect to their own past data and the random variation. Historical trends and seasonal patterns are explored in time series modelling to predict the future of the variable based on the identified trends and seasonal components.

In time series approach, simple ARIMA model or seasonal ARIMA model is selected based on the time frequency used. However, the most recent literature shows that the seasonal ARIMA model is more popular than the simple ARIMA model since seasonal variations in tourism data is a main component considered by the decision markers in the tourism industry. Another reason for the popularity of the time series approach is the less cost associates with data collection and model estimation process, since it only requires past data of the variable considered. (Kodituwakku, Wijesundara, & Hettiarachchi, 2015)

Numerous studies had looked into forecasting tourist arrivals using the time series approach. Kim and Moosa (2005) use regression-based forecasting models, Structural Harvey's models and SARIMA models, to forecast the tourist flows and find that most suitable method is the indirect method compared to direct method to forecast the monthly international tourist arrival in Australia. Direct method of forecasting an aggregated tourism variable does not examine the behaviour of the time series components whereas indirect method generates forecast by aggregating the data generated from the behaviour of the time series components. Cho (2003) uses exponential smoothing, univariate ARIMA and Artificial Neural Networks model to forecast the number of tourist arrivals from six regions to Hong Kong using data for the period of January 1999 to December 2000.

Seasonality is considered as one of the most important component in tourist arrival series. Possibility of forecasting the seasonal pattern of a tourism data series depend on whether it is a deterministic or stochastic series as forecasting stochastic data series is impossible. Hence the first step of analysing the seasonal series is testing the deterministic nature of it. In the recent literature, HEGY test is highly recommended as an appropriate seasonal unit root test to test the deterministic trend of the time series.

Among all the techniques which has been used in the literature to forecast the time series with seasonality, SARIMA models is considered as the most suitable model. SARIMA model can handle both non-stationary series and stationary series together with seasonal elements and non-seasonal elements. (Chang & Liao, 2010) Chang and Liao (2010) examine an appropriate seasonal ARIMA models, in order to forecast the outbound tourist exits from Taiwan to Hong Kong, Taiwan to USA and Japan. Monthly data for the period of January 1996 to December 2006 are used in this study. Deterministic seasonality in the data is identified by using the HEGY test and the forecasting accuracy is measured using MAPE. First non-seasonal difference of all three series are used to obtain the deterministic trends for each outbound data series.

Kodituwakku et al. (2015) has used three models to forecast the tourist arrivals to Sri Lanka for the period of January 2010 to August 2014. Holt Winters Multiplicative Seasonal models was selected as the best model to forecast the data compared to SARIMA, and NARX neural network models. But this research has not considered HEGY test to test the seasonality in the SARIMA model.

Not only in the tourism forecasting, but also in other fields of studies have identified SARIMA model as an appropriate technique to forecast the time series variables. Savas (2013) use SARIMA models and Kalman filter to obtain monthly forecasts of inflation in Luxembourg, Mexico, Portugal and Switzerland. Reininger and Fingerlos (2007) uses quarterly real GDP level series in Belgian for the period of 1980 Q1 to 2006 Q4 and fit a univariate model with the purpose of explaining seasonal characteristics of the series. HEGY test and DHF tests are used to test the seasonal integration and following the results two SARIMA models are identified as best models.

Even though the SARIMA model is quite popular, testing the seasonal unit roots is neglected in the past literature. Therefore, the main purpose in this research paper is to employ SARIMA model by using HEGY test to forecast tourist arrival in Sri Lanka.

2.2 Evaluation of forecasting accuracy

A forecast may not be 100% accurate so the actual tourist arrival may be deviate from the forecasts. A model which has the minimum deviation is considered as the best forecasting model. The majority of the tourist arrival forecasting studies have used Mean Absolute Percentage Error (MAPE), Root Mean Square Error (RMSE), Root Mean Square Percentage Error (RMSPE) and Theil's U to examine the performance of forecasting in their tourism demand forecasting models.

Performance of the forecasting model depends on three factors. Forecasting horizons length, destination origin country and the frequency of data used in the model are those factors. However there are no predetermined steps to fit a forecasting model in tourism forecasting literature and the construction of the model and estimation of the model entirely relied on secondary data. (Song & Li, 2008).

Lewis (1982) develop criteria to measure the model accuracy using MAPE. Chang & Liao (2010) use these criteria to evaluate the accuracy of their forecasting model. Further Cho (2003) uses RMSE, MAPE in comparing the performance of the three forecasting procedures and select the forecasting method which has the minimum RMSE and MAPE values. Mansor and Ishak (2015) use MAE, RMSE, MAPE and acceptable output percentage (z) to evaluate the performance of Exponential Smoothing, ARIMA and ARFIMA models in forecasting the tourist arrivals in Langkawi Island. Song, Witt and Jensen (2003) forecast the demand for international tourism using six alternative econometric models and rank them over different time horizons based on MAPE and RMSE. Savas (2013) uses MSE, MAD, MAPE and Theil's U to measure the accuracy of the forecasts from Kalman filtered model and SARIMA model.

Though there are many researches in forecasting tourist arrival in worldwide, few researches have been done in Sri Lankan context. Even in those studies, attempts are made to compare several forecasting approaches and not to improve the precision of a selected model. Therefore, the present research attempts to forecast the tourist arrival in Sri Lanka by considering SARIMA model based on HEGY framework and to find the most accurate forecasting model based on forecasting accuracy measures.

3. Methodology

3.1 Seasonal ARIMA models

In time series approach, SARIMA model is becoming more popular because of the included seasonal variation component has gained more attention in the tourism industry. (Song & Li, 2008). Hence, this paper consider the SARIMA model with s (s = 12 for monthly data) to forecast the tourist arrival series as suggested by Kim and Moosa in 2005.

$$\phi_p(B)\Phi_P(B^s)\Delta^d\Delta_s^D Y_t = \mu + \theta_q(B)\Theta_Q(B^s)u_t \qquad (1)$$

In the above equation (1), u_t is a white noise and μ is a constant. B represents the back shift operator, and $\Delta^d \equiv (1-B)^d$ and $\Delta_s^D \equiv (1-B^s)^D$ are respectively the operators for the d[th] order monthly or non-seasonal differencing operator whereas the D[th] order annual or seasonal differencing operator. Polynomials of B and B^s are $\phi_p(B)$, $\Phi_P(B^s)$, $\theta_q(B)$ and $\Theta_Q(B^s)$ which are written as;

$$\begin{aligned}\phi_p(B) &= 1-\alpha_1 B - \ldots - \alpha_p B_p, \theta_q(B)\\ &= 1-\beta_1 B - \ldots - \beta_q B^q\end{aligned} \qquad (2)$$

$$\begin{aligned}\Phi_P(B^s) &= 1 \quad \delta_1 B^s \quad \ldots \quad \delta_P B^{Ps} \text{ and } \Theta_Q(B^S)\\ &= 1-\pi_1 B^S - \ldots - \pi_Q B^{QS}\end{aligned} \qquad (3)$$

Here, the unknown parameters which should be estimated are denoted as α, β, δ and π. Here the initial polynomials represent the non-seasonal AR and non-seasonal MA components. If AR component is stationary and the MA component is invertible, then characteristic roots of the polynomials are present at outside the unit circle.

In order to explain the cyclical behavior of a series that is around the mean, the AR and MA components are used. Here the mean doesn't vary with the time. If all the seasonal AR and MA polynomials are stationary and invertible, then all of seasonal AR and MA roots are expected to be placed outside the unit circle.

This model is used in various researches in forecasting time series data which are having a strong trend and seasonal components. Several authors Cho (2003), Kulendran and Wong (2005), Chan, Sriboonchitta, and Wiboonpongse (2009), Chang and Liao (2010), Premanasari, Hidayah and Bustoni (2013), Mansor and Ishak (2015) have used the above model to forecast the time series variable according to their context.

In the above expressions, time-varying mean and variance features are associated with non-stationarity of the series and removed by the monthly and annual difference operators. A series which needs "d" number of monthly differencing and "D" number of annual differencing to achieve stationarity is represented by I(d, D). Most financial and economic time series requires maximum one monthly differencing or one annual differencing represented by I(0,1).

The parameters p, P, q and Q are determined by examining the behavior of auto-correlation function (ACF) and partial auto-correlation function (PACF). If it is a stationary series, either a cut-off or fast dying pattern should be there in the ACF and PACF. (Cho, 2003)

3.2 Seasonal unit root test.

There is a challenge in ignoring the seasonality characteristic in the modeling process of monthly or quarterly tourist arrival series. Many scholars in tourism demand forecasting literature, consider seasonality either as a stochastic component or deterministic component.

In a time series where the seasonality is considered as a stochastic component, it requires seasonally differencing as a test for seasonal unit roots. But, if the seasonality is considered as a deterministic components, it is sufficient to introduce seasonal dummies into the models in account for seasonal variations. In recent researches, many scholars have used HEGY test as a seasonal unit root test. (Song & Li, 2008)

Following, Kim and Moosa, (2005) a seasonal unit root test can be specified as follows to determine the empirical values of d and D.
Regression form of the HEGY test is;

$$\Delta_{12}Y_t = \sum_{i=1}^{12}\pi_i z_{i,t-1} + \gamma t + \sum_{i=1}^{12}\delta_k D_{kt} + \sum_{i=1}^{k}\phi_i \Delta_{12}Y_{t-i} + \varepsilon_i \quad (4)$$

Where $z_{i,t-1}$'s are linear transformation of $Y_{t-1}\ldots Y_{t-12}$. The residuals of the above regression equation (4) follows a white noise process and should determine the value of k. This research follows the decision process used by Kim and Moosa (2005) for the HEGY test.

The time series has a I(0,1) process then the null hypothesis of, $\pi_i = 0$ for all i is accepted. F statistic denoted as F_{1-12} can be used to test this. The null hypothesis of the above is rejected and the failure to reject $\pi_1 = 0$ jointly with the rejection $\pi_i = 0$ for i = 2, …, 12 shows that the series follows I(1,0). The second hypothesis is tested using F_{2-12} statistic. If both F_{1-12} and F_{2-12} statistics are rejected, then the time series has a I(0,0) process.

In this research, the empirical outcomes of the HEGY test are used to find the value of d and D of the SARIMA model. Given that, visual inspection of autocorrelation function (SACF) of the residuals and the partial autocorrelation function (PACF) of the residuals are used to obtain the possible values of p, q, P and Q of the SARIMA model. Akaike Information Criterion (AIC), which is an order selection criterion with another criterian Schwarz Criterion (SCW) are used to find the possible models. (Chang, Sriboonchitta, & Wiboonpongse, 2009)

In the present research, significance of each coefficients is tested using individual asymptotic t test values. As a diagnostic check, residual analysis is used through visual inspection of ACF and PACF of residuals and Ljung-Box test. Here, the Ljung-Box test is used to test the joint significance of residual sample autocorrelation. All selected models are then go through asymptotic t-test as a further diagnostic check. The unconditional least-square method in Eviews 8.0, is used to estimate the unknown parameters of the model. (Kim & Moosa, 2005)

However, many researches in the tourism context have to deal with small samples and that leads to a problem of using the HEGY test.(Song & Li, 2008). Kulendran (2005) and Wong shows that HEGY test is not useful in model selection process and an alternative method for the HEGY test is proposed by Kim and Moosa (2005).

3.3 Forecasting accuracy measures

Following literature, the present research uses Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), Mean Absolute percent Error (MAPE) to measure the forecast accuracy.

$$RMSE = \sqrt{\frac{\sum e_t^2}{n}} \quad (5)$$

$$MAPE = \left[\frac{\sum |e_t| / Y_t}{n}\right] \times 100 \quad (6)$$

$$MAE = \frac{\sum_{t=1}^{n} |e_t|}{n} \quad (7)$$

Here e_t is the forecast error in the period t. Y_t is the real value of the time period t. Also n is the observations in the period. Criteria developed by Lewis (1982) has been used for past few decades to measure the accuracy of the model. Table 1 shows the details of the criteria.

Table 1. MAPE values for Model selection, Source: Lewis (1982)

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

4. Data

Many scholars use aggregate data series such as total tourist expenditure at the destination country, total tourist arrival to forecast the tourism data series. Total tourist arrival series from an origin level to destination level is used as a measure for tourist arrival variable in many researches over the past few years. (Song & Li, 2008)

This research uses the number of monthly international tourist arrivals in Sri Lanka obtained from Monthly Statistical Bulletin and Annual statistical reports published in the website of Sri Lanka Tourism Development Authority for the period of 2009 to 2016. Tourist arrival data for the period of 1995 to 2008 is taken from the annual statistical reports available in the library of Sri Lanka Institute of Tourism and Hotel Management. In total, 259 observation considered in the present study. In order to stabilize the variance, during the model estimation process, data series is transformed to its natural logarithms form.

Table 2 describes the sample data. Figure 1 reports the time series plot of international tourist arrival to Sri Lanka for the period considered. It is clearly shows that the data exhibit strong seasonal behaviour and upward trend.

Table 2. Descriptive Statistics

Measure	Value
Mean	57395.14
Median	41526.00
Maximum	209351.0
Minimum	11758.00
Std. Dev.	40303.53
Skewness	1.765884
Kurtosis	5.592571
Observations	259

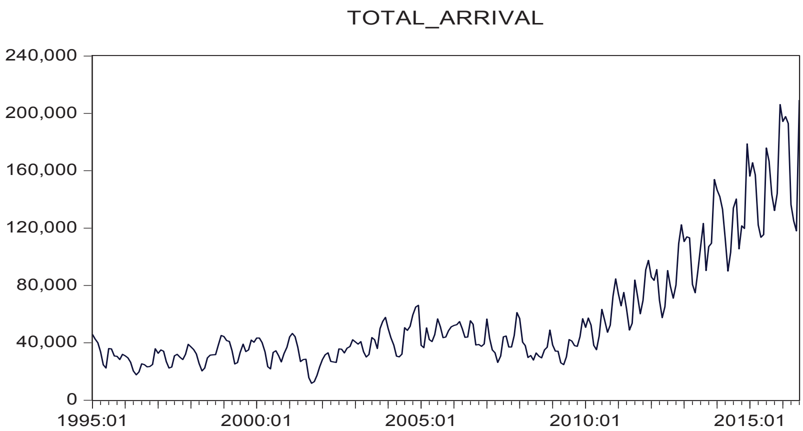


Figure 1. International Tourist arrival Jan 1995 – July 2016

5. Empirical Results

5.1 Seasonal Unit root test

Seasonality is a significant component in tourism data series. Hence HEGY test in equation (4) applied to monthly international tourist arrivals series in order to obtain accurate estimate about the seasonal component. Since the empirical data shows a trend component, trend variable is included into the estimate with the purpose of managing the deterministic trend. HEGY test involves seasonal dummies, an intercept, lags and trend of the series. Table 3 represents the estimated results for HEGY test and Wald test is used to get the required statistics.

Table 3. The HEGY test statistics (Same for log series)

	Statistic value	**Probability value (5%)**
$t(\pi_1)$	0.645922	0.5190
$t(\pi_2)$	-3.587661	0.0004*
$t(\pi_3)$	-0.512668	0.6087
$t(\pi_4)$	-0.901950	0.3681
$t(\pi_5)$	2.497389	0.0133*
$t(\pi_6)$	0.809957	0.4189
$t(\pi_7)$	0.945572	0.3455
$t(\pi_8)$	-1.952222	0.0523
$t(\pi_9)$	-4.467851	0.0000*
$t(\pi_{10})$	-5.546179	0.0000*
$t(\pi_{11})$	-0.989489	0.3236
$t(\pi_{12})$	2.078176	0.0389*
$F(\pi_1 = \ldots = \pi_{12} = 0)$	5.764333	0.0000*
$F(\pi_2 = \ldots = \pi_{12} = 0)$	6.182739	0.0000*
$F(\pi_3 = \ldots = \pi_{12} = 0)$	5.792039	0.0000*

The above results of the HEGY test is used to find the value of d and D of the Seasonal ARIMA model. Following the rules in HEGY test, it is decided that empirical series follows I(0,0) since both F_{1-12} and F_{2-12} statistics are rejected.

5.2 Seasonal ARIMA model

Various autoregressive (AR), moving average (MA) and auto regressive integrated moving average models (ARIMA) have been estimated to the logarithms of international tourist arrival series using ordinary least squares to select the best model.

The best fitting model is selected based on various criteria. Significant t values at the 5% level of significance for all the MA, AR, SAR and SMA coefficients are checked. The estimated models are tested for serial correlation using Langrange multiplier (LM) test. Models in which all of the coefficients are found to be significant and the there is no serial correlation in the residuals are selected for next step. Residuals diagnosis which are based on the correlogram provides more support for the results of the LM test for serial correlation. Out of all the selected models, best model is selected based on AIC and SBC criteria. Table 4 presents the estimated $ARIMA(1,0,16)(36,0,24)_{12}$ model which is selected as the best model since it has the minimum AIC and SBC values out of all the models considered.

Journal of Tourism, Hospitality and Sports www.iiste.org
ISSN (Paper) 2312-5187 ISSN (Online) 2312-5179 An International Peer-reviewed Journal
Vol.22, 2016

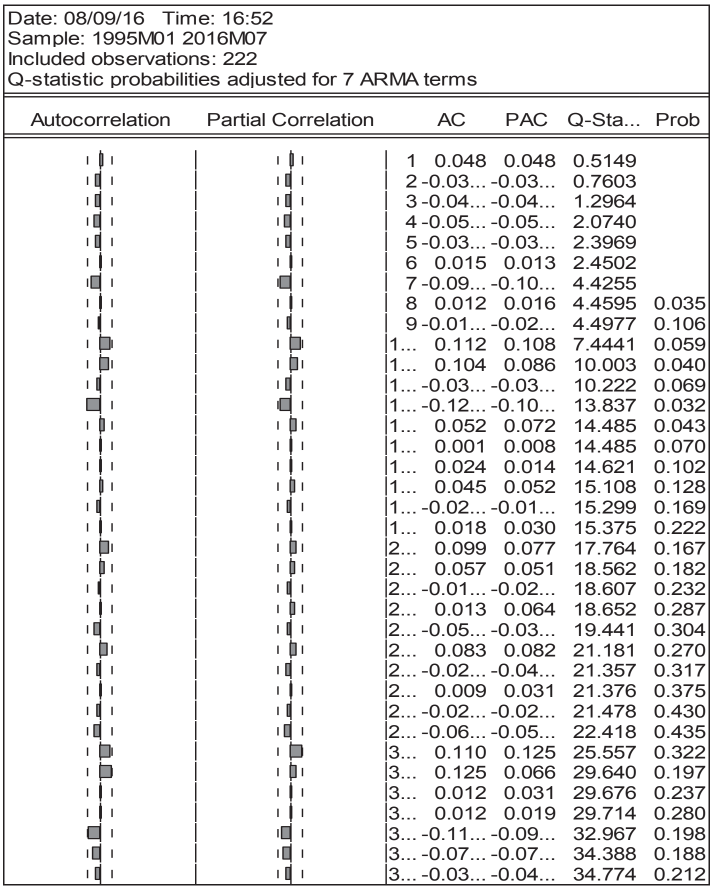
Table 4. Estimates of seasonal ARIMA model

Variable	Coefficient	t- Statistic	Probability
C	3.053425	5.044837	0.000
AR(1)	0.914654	32.17754	0.000
SAR(12)	1.985745	31.66203	0.000
SAR(24)	-1.403758	-12.12039	0.000
SAR(36)	0.430683	6.814344	0.000
MA(16)	0.146677	2.246010	0.0257
SMA(24)	0.867319	59.67877	0.000
SMA(12)	-1.797140	-120.7275	0.000
Adjusted R^2 value	0.962487		
S.E. of regression	0.048475		
AIC	-3.180162		
SCW	-3.057543		
Durbin Watson Stat	1.883789		
Serial Correlation test (LM test) probability	0.4952		
RMSE	0.047594		
MAE	0.036777		
MAPE	0.787116		
Theil inequality coefficient	0.005041		

The best fitting model to describe the international tourist arrival pattern can be expressed as follows.

$$y_t = 3.053425 + u_t,$$
$$(1 - 0.92L)(1 - 1.982L^{12})(1 + 1.402^{24})(1 - 0.43^{36})u_t = (1 - 0.15L^{16})(1 + 1.80L^{12})(1 - 0.87L^{24})\varepsilon_t$$

The model selected as the best model shows that the autocorrelations are within the 95% confidence interval (where Durbin-Watson statistic close to 2.00) which indicates that there is no significant estimated residual autocorrlations in the selected model. Figure 3 shows the actual, fitted and residuals series of the estimated model.


Date: 08/09/16 Time: 16:52
Sample: 1995M01 2016M07
Included observations: 222
Q-statistic probabilities adjusted for 7 ARMA terms

Autocorrelation	Partial Correlation		AC	PAC	Q-Sta...	Prob
		1	0.048	0.048	0.5149	
		2	-0.03...	-0.03...	0.7603	
		3	-0.04...	-0.04...	1.2964	
		4	-0.05...	-0.05...	2.0740	
		5	-0.03...	-0.03...	2.3969	
		6	0.015	0.013	2.4502	
		7	-0.09...	-0.10...	4.4255	
		8	0.012	0.016	4.4595	0.035
		9	-0.01...	-0.02...	4.4977	0.106
		1...	0.112	0.108	7.4441	0.059
		1...	0.104	0.086	10.003	0.040
		1...	-0.03...	-0.03...	10.222	0.069
		1...	-0.12...	-0.10...	13.837	0.032
		1...	0.052	0.072	14.485	0.043
		1...	0.001	0.008	14.485	0.070
		1...	0.024	0.014	14.621	0.102
		1...	0.045	0.052	15.108	0.128
		1...	-0.02...	-0.01...	15.299	0.169
		1...	0.018	0.030	15.375	0.222
		2...	0.099	0.077	17.764	0.167
		2...	0.057	0.051	18.562	0.182
		2...	-0.01...	-0.02...	18.607	0.232
		2...	0.013	0.064	18.652	0.287
		2...	-0.05...	-0.03...	19.441	0.304
		2...	0.083	0.082	21.181	0.270
		2...	-0.02...	-0.04...	21.357	0.317
		2...	0.009	0.031	21.376	0.375
		2...	-0.02...	-0.02...	21.478	0.430
		2...	-0.06...	-0.05...	22.418	0.435
		3...	0.110	0.125	25.557	0.322
		3...	0.125	0.066	29.640	0.197
		3...	0.012	0.031	29.676	0.237
		3...	0.012	0.019	29.714	0.280
		3...	-0.11...	-0.09...	32.967	0.198
		3...	-0.07...	-0.07...	34.388	0.188
		3...	-0.03...	-0.04...	34.774	0.212

Figure 2. Correlogram of residuals

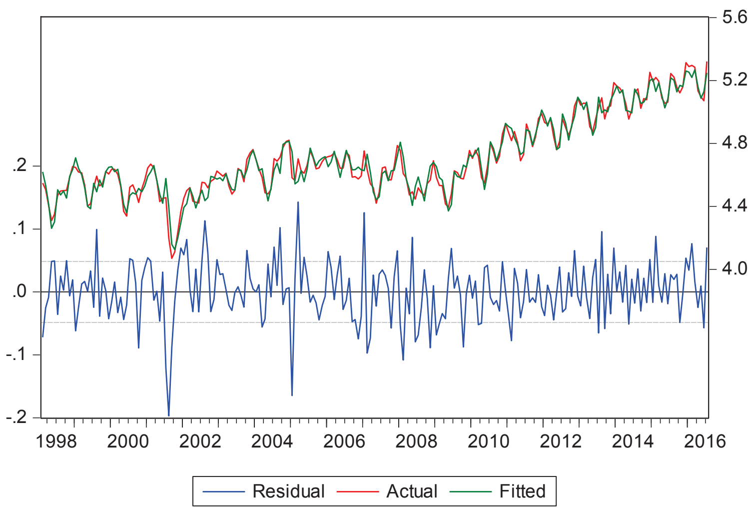
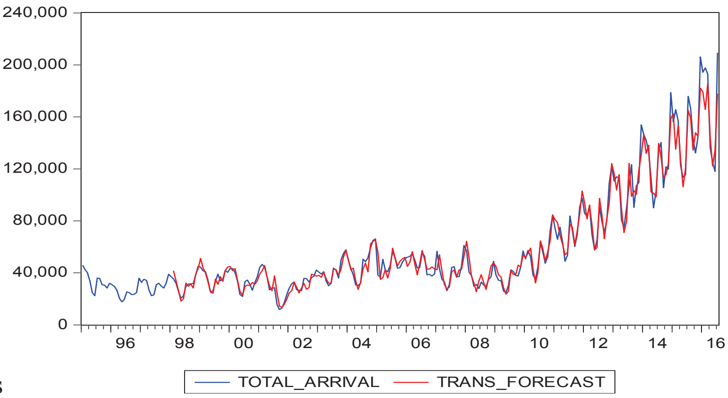


Figure 3. Actual values with fitted values and residuals from ARIMA (1,0,16)(36,0,24)$_{12}$ model

The Theils's U Statistic (0.005041) which is close to 0 shows that the model is a perfect fit. MAPE (0.787116) represents the percentage of average absolute error occurred. MAE (0.038777) measure the average absolute deviation of forecasted values from original values. However the log transformation can be affected to this MAPE and MAE values.

In order to measure the accuracy of the forecast, forecast values are transformed back to the original values. Following figure presents the graph of forecasts and the original series.



s

Figure 4. Graph of forecast values and the original values

5.3 Model accuracy

The findings of this research can be compared to other time series models in the literature to check the accuracy of the model. Table 5 provides the details of the model accuracy levels in past researches.

Table 5. Comparison of the MAPE values

Research article	Model details	MAPE
(Cho, 2003)	ARIMA model for USA, Japan, Taiwan, Korea, UK, Singapore	10.23, 44.52, 8.24, 32.13, 15.73, 14.26
(Song, Witt, & Jensen, Tourism forecasting: accuracy of alternative econometric models, 2003)	ARIMA model for Germany, Netherlands, Norway, Sweden, UK, USA	4.46, 12.12, 9.32, 11.69, 2.46, 8.79
(Chang & Liao, 2010)	SARIMA model for Hong Kong, Japan, USA, Total	5.63, 7.64, 7.20, 8.90
(Savas, 2013)	SARIMA model for Luxembourg, Mexico, Portugal, Switzerland	0.985, 0.998, 1.991, 2.080
(Premanasari, Hidayah, & Bustoni, 2013)	SARIMA model to forecast number of Malaria incidents	21.6
(Kodituwakku, Wijesundara, & Hettiarachchi, 2015)	SARIMA model from Total arrival, India, UK, Germany, Maldives, France, China	6.7786, 9.8958, 14.6411, 20.0151, 26.5687, 10.2965, 19.8696
(Mansor & Ishak , 2015)	ARIMA model	11.47

The MAPE value of the current research, 0.787% is lower than all the other time series model developed in different context. Following the details in table 1, it can be concluded that the SARIMA model developed in this research has a high accuracy level since the MAPE value is less than 10%.

6. Conclusions

This paper looked at Box–Jenkings SARIMA models based on past international tourist arrivals to Sri Lanka. In order to select a suitable SARIMA model, seasonal unit roots based on the HEGY framework are applied to the monthly data of tourist arrival from 1995 Jan to 2016 July. The findings of seasonal unit roots test evident that the SARIMA model without seasonal differencing and non- seasonal differencing is appropriate for the data.

Based on the AIC, BSC, LM tests and Correlegrams, several models are selected as the appropriate models and the best model is selected by considering the lowest value for RMSE, MAE, MAPE and U coefficients. The empirical result shows that the SARIMA (1, 0, 16)(36, 0, 24)$_{12}$ model is appropriate to capture patterns of international tourist arrival and to forecast the international tourist arrivals in Sri Lanka with a high accuracy level.

Findings of the research can be used to make better strategic plans in Sri Lanka especially in the government, hospitality and tourism industries. (Cho, 2003)

7. Further research

In future, further research can be focused in applying more sophisticated time series forecasting tools using latest technologies such as Artificial Intelligence algorithms, Genetic algorithms, Fuzzy Neural System or some advance data mining technique to forecast tourist arrival. Further, some other econometric methods should also be considered in the future.

In the past researches, considering univariate time series to a multivariate dimension has improved the performance of the ARIMA and SARIMA forecasting models. So there is a possibility to increase the forecasting accuracy by considering the tourist arrival for a given destination or by going through the number of origin countries or regions. (Song & Li, 2008). Further following Kim and Moosa (2015) there is a possibility to increase the accuracy of forecasting results by using the indirect forecasting method.

References

Central Bank of Sri Lanka. (2015). *Annual Report 2015.* CBSL.

Chaitip, P., & Chaiboonsri, C. (2014, December 31). International Tourists Arrival to Thailand: Forecasting by Non-Linear Model. *Procedia Economics and Finance*, pp. 100-109.

Chang, C.-L., Sriboonchitta, S., & Wiboonpongse, A. (2009). Modelling and forecasting tourism from East Asia to Thailand under temporal and spatial aggregation. *Mathematics and Computers in Simulation, 79*(5), 1730-1744.

Chang, Y.-W., & Liao, M.-Y. (2010). A Seasonal ARIMA Model of Tourism Forecasting: The Case of Taiwan. *Asia Pacific Journal of Tourism Research, 15*(2), 215-221.

Cho, V. (2003). A comparison of three different apporaches to tourist arrival forecasting. *Tourism Management, 24*(3), 323-330.

Kim, J. H., & Moosa, I. A. (2005). Forecasting international tourist flows to Australia: a comparison between the direct and indirect methods. *Tourism Management, 26*(1), 69-78.

Kodituwakku, W., Wijesundara, W., & Hettiarachchi, C. (2015). *Modelling and Forecasting Tourism Demand for Sri Lanka.* Colombo: University of Colombo School of Computing.

Kulendran, N., & Wong, K. K. (2005). Modeling seasonality in tourism forecasting. *Journal of travel research*, 163-170.

Lewis, C. (1982). *International and Business Foreasting Methods.* London: Butterworths.

Mansor, K. A., & Ishak , W. I. (2015). Forecasting Tourist Arrivals To Langkawi island Malaysia. *Cross-Cultural Management Journal (07)*, 69-76.

Premanasari, A. E., Hidayah, I., & Bustoni, I. A. (2013, October 07). SARIMA (Seasonal ARIMA) implementation on time series to forecast the number of Malaria incidence. *Intenrational Conference on Information Technology and Electrical Engineering (ICITEE)*, pp. 203-207.

Reininger, T., & Fingerlos, U. (2007). *Modelling Seasonality of Gross Domestic Product in Belgium.* University of Vienna Working Paper.

Savas, N. (2013). *Forecast Comparison of models based on SARIMA and the Kalman Filter for inflation.* Uppsala University.

SLTDA. (n.d.). *Annual Statistical Reports 2009 - 2015.* Retrieved from http://sltda.lk/index.html

Song, H., & Li, G. (2008). Tourism demand modelling and forecasting - A review of recent research. *Tourism Management, 29*(2), 203-220.

Song, H., Witt, S. F., & Jensen, T. C. (2003). Tourism forecasting: accuracy of alternative econometric models. *International Journal of Forecasting*, 123-141.

Wickramasinghe, V., & Takano, S.-e. (2010). Application of combined SWORT and Analytic Hierarchy Process (AHP) for tourism revival strategic marketing planning: A case of Sri Lanka tourism. *Journal of Eastern Asia Society for Transportation Studies*, 954-969.