# Time Series Patterns of Tourist Arrivals to Sri Lanka

K.M.U.B Konarasinghe
Institute of Mathematics and Management

Review of Integrative Business & Economics Research

**ABSTRACT**

This study focused on pattern recognition of tourist arrivals to Sri Lanka from various regions in the world. Monthly time series data from January 2008 to December 2014 are used in this study. The regions selected for the study were the top four in market position. They are; Asia, Western Europe, Eastern Europe and the Middle East. Descriptive statistics, Time Series plots and Auto-Correlation Functions (ACF) were used for pattern identification and one way- Analysis of Variance (ANOVA) was used for mean comparison of tourist arrivals from selected regions. The average arrivals of Asia and Western Europe are the highest 29361 and 25982 respectively. There is no significant difference between these two regions. Eastern Europe and the Middle East have 5866 and 4300 of average respectively. Arrivals from Asia, Western Europe, and Eastern Europe were not normally distributed, all were positively skewed. Data series of all four regions were non-stationary. There is a significant difference of tourist arrival from Asian and Western Europe compared to other regions. It is recommended to test Moving Average Methods, Exponential Smoothing techniques, Decomposition techniques, linear and non-linear trend models and Circular model for forecasting arrivals.

Keywords: Keywords: Auto-Correlation Function, Analysis of Variance, Non-stationary

## 1. INTRODUCTION

Sri Lanka is considered as a wonderful tourist destination in the world. It is known by Asian poets, nothing the geographical location of the island and lauding its beauty, is called the “pearl upon the brow of India”. The country claims to a long history of over three thousand years, with one of the longest documented histories in the world. Sri Lanka consists of rich culture with many different ethnic communities, cultural, linguistic and religious diversities. Sri Lanka is loaded with lush tropical forests, white sandy beaches and panoramic and unique landscapes with rich biodiversity. This Country is a home to seven world heritage sites; namely Galle, Kandy, Sigiriya, Anuradhapura, Polonnaruwa, Dambulla cave temple and Singharaja tropical rain forest. It causes the attraction for tourism Konarasinghe (2014) and Konarasinghe and Deheragoda (2013). The growth of international tourist arrivals of every year is the strongest evidence for it. Tourism market of Sri Lanka consist all regions in the world.


ISSN: 2304-1013 (Online); 2304-1269 (CDROM); 2414-6722 (Print)

Asia and Western Europe is the highest tourist arrival regions recorded during the past (SLTDA, 2014).

## 1.1 PROBLEM STATEMENT

Sri Lankan Tourism market consists of various regions. These regions are heterogeneous by nature. As such forecasting arrivals by region is important. But this process is quite complicated. Probably the same forecasting technique may not be applicable for all the regions. In contrast, arrivals from some regions may have similar patterns. Therefore identifying the similarities and differences of the patterns of tourist arrivals in various regions is important to spotting suitable time series techniques for forecasting. This study was concerned with pattern recognition of tourist arrivals from various regions.

## 1.2 OBJECTIVES OF THE STUDY

1.2.1. Identify the patterns of arrivals to Sri Lanka, from various regions.
1.2.2. Compare the number of arrivals to Sri Lanka from various regions.

## 1.3 SIGNIFICANCE OF THE STUDY

The tourism industry is the third income generator of Sri Lanka. It creates thousands of jobs every year. Therefore, it is very important to do a proper planning, controlling and management in the industry. Konarasinghe, Abeynayake, and Gunaratne (2015) has shown the role of statistical modeling in prediction, control, and optimization of a large number of fields such as agriculture, medicine, engineering, physics, economics, and much more.

According to Konarasinghe et.al (2015); Statistical models are divided into two parts; univariate statistical models and multivariate statistical models. Univariate statistical model is an equation or set of equations that explain the behavior of single random variable over time while multivariate statistical models explain the joint behavior of two or more random variables. Univariate statistical modeling procedure is based on the past internal patterns in data to forecast the future and no external variables are required in forecasting. The basic concept of these methods is that future values of a series are a function of past values. Univariate methods include moving average smoothing, exponential smoothing, Winters' method, decomposition techniques, Fourier series analysis, Box Jenkin's Auto Regressive Integrated Moving Average (ARIMA) methods, linear and non-linear trend models.

Pattern recognition of arrivals provides guidelines to spot the suitable univariate forecasting techniques for arrivals to Sri Lanka from various regions. Also, this study will illuminate the most profitable regions to the industry. Findings of this study will be a lighthouse to develop new forecasting techniques. This study will be facilitating to identify the requirements of the tourism industry of Sri Lanka. It will ensure the proper supply and demand management to minimize the wastage strategic resources.

## 2. LITERATURE REVIEW

Singh (1999) examined Pattern Modeling and Recognition System (PMRS) as a tool for pattern recognition in the financial industry. He utilized time series data of six different


ISSN: 2304-1013 (Online); 2304-1269 (CDROM); 2414-6722 (Print)

financial indices. They are German DAX, British FTSE, French FRCAC, SWISS, Dutch EOE and the US S&P series. Morales-Esteban, Martínez-Álvarez, Troncoso, Justo and Rubio-Escudero (2010) conducted a study to find the behavior of the occurrence of medium and large earthquakes. They utilized tectonic, geological, seismic and gravimetric data from 27 areas of Spain and Portugal. Pattern recognition algorithms, time series representations, and dimensionality reduction techniques were the techniques used by them. Lin, Williamson, Borne and Barr (2012) emphasized the importance of Symbolic Aggregate approximation (SAX) in pattern recognition of a time series. Sternickel (2002) focus on the detecting of any recurrent pattern in ECG. The author used Neural Networks and wavelet transformed templates. The recognition of recurring patterns within multivariate time series was the purpose of the study of Spiegel, Gaebler, Lommatzsch, Luca and Albayrak (2011). They used agglomerative hierarchical clustering for recurring pattern recognition. Bemdt and Clifford (1994) used an algorithm based on the dynamic time warping technique for pattern recognition for stock prices or NASA telemetry data. Konarasinghe and Abeynayake (2014) emphasized that Box- plots, Time series plots and Auto-Correlation Functions (ACF) are some of the pattern recognition tools.

## 3. METHODOLOGY

"A pattern is essentially an arrangement. It is characterized by the order of the elements of which it is made, rather than by the intrinsic nature of these elements," is a definition given by Norbert Wiener, Sharma , and Kaur (2013). Sharma and Kaur (2013) emphasize that Pattern recognizing process has been gaining benefits from long term and it facilitate to developed highly sophisticated skills and techniques to take actions according to what they observed at present. Pattern Recognition provides direction to solve problems in many fields; medicine, finance, engineering agriculture and much more.

Tourist arrivals from all regions are the population of the study. Four regions with the highest number of arrivals were selected. Asian region, Western Europe, Eastern Europe and the Middle East region are the top four regions in the Sri Lankan international tourism market. Monthly tourist arrival data from January 2008 to December 2014 was obtained from annual statistical reports from 2008 to 2014, published by Sri Lanka Tourism Development Authority (SLTDA). Descriptive statistics, Box- plots, time series plots and Auto-Correlation Functions (ACF) were used for pattern identification and one way- Analysis of Variance (ANOVA) was used for mean comparison of tourist arrivals from selected regions.

## 4. RESULTS

Data analysis is organized region wise as follows;

4.1. Descriptive Statistics of all regions.

4.2 Pattern recognition tourist arrivals of all regions.

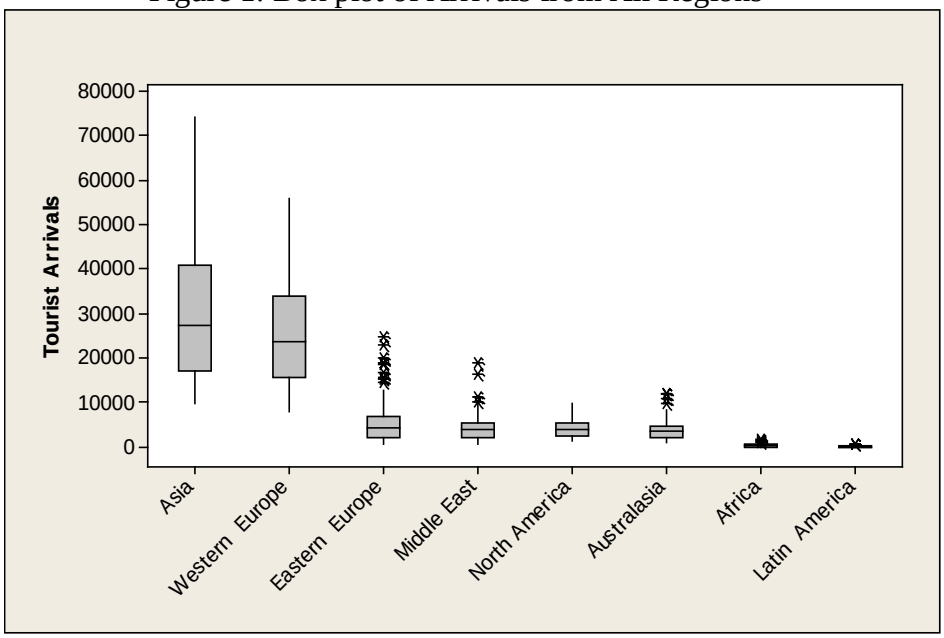
4.3. Comparison of arrivals by region


ISSN: 2304-1013 (Online); 2304-1269 (CDROM); 2414-6722 (Print)

Outliers are the extremely large or small values of a data set. They were identified with the help of Box Plot (Figure 1) and replaced by moving an average of order three. That is, if the $i^{th}$ value of a series is an outlier;

$$i^{th} value = [(i-1)^{th} value + (i-2)^{th} value + (i-3)^{th} value]/3$$

Figure 1: Box plot of Arrivals from All Regions




## 4.1 DESCRIPTIVE STATISTICS OF ALL REGIONS.

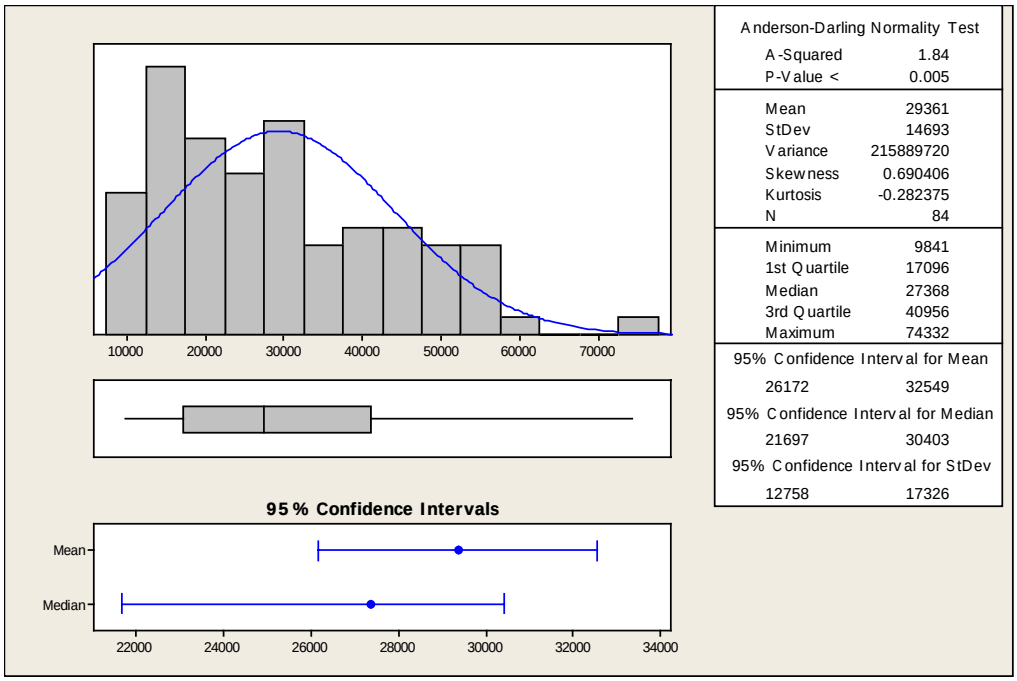
Descriptive statistics for a number of arrivals were obtained region wise; it includes graphical summary along with measures of location, dispersion and the normality of the series.

### 4.1.1 ASIAN REGION

Graphical summary of descriptive statistics is shown in Figure 2;


ISSN: 2304-1013 (Online); 2304-1269 (CDROM); 2414-6722 (Print)

Figure 2: Graphical Summary of Asian Region



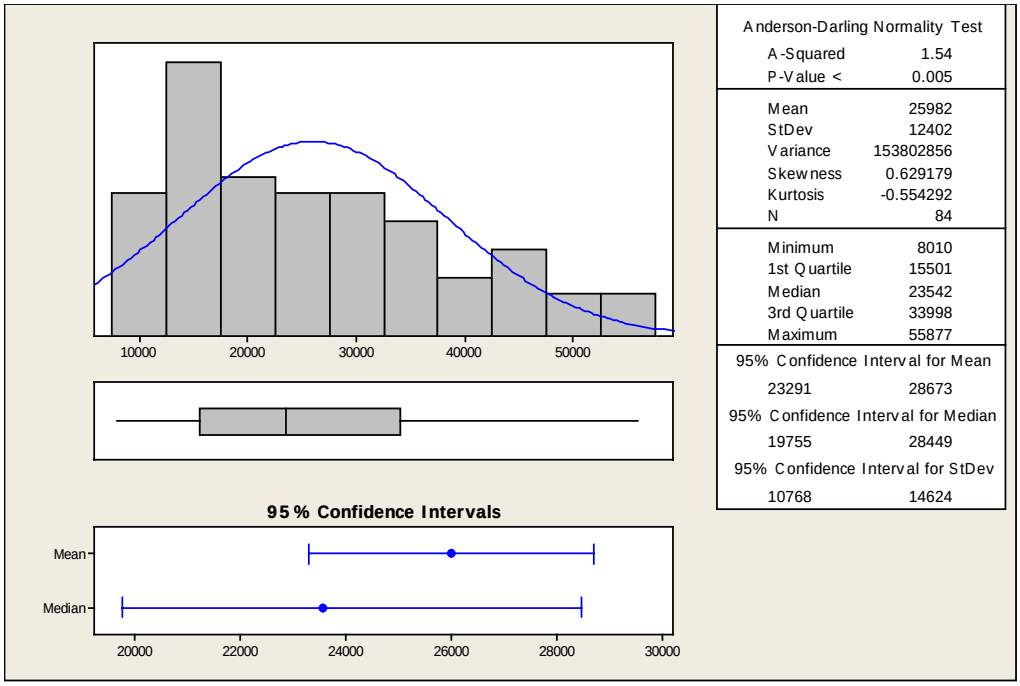

Minimum arrivals recorded from Asian region were 9841 whereas maximum were 74332 during the period. The first quartile of arrivals is 17096. It means ¼ of the months had at most 17096 arrivals. A median arrival is 27368 and the third quartile of arrival is 40956. Histogram of the arrivals does not look symmetrical. The P value of the Anderson-Darling test is less than the significance level (p-value <0.005). A number of arrivals do not follow the normal distribution. They are positively skewed.

### 4.1.2 WESTERN EUROPE REGION

The Figure 3 shows that maximum tourist arrivals from Western Europe were 55877 and minimum were 8010. The first quartile of arrivals is 15501. A median arrival is 23542 and the third quartile of arrival is 33998. Histogram of the arrivals does not look symmetrical. The P value of the Anderson-Darling test is less than the significance level (p-value <0.005). As such number of arrivals does not follow the normal distribution. They are positively skewed.□


ISSN: 2304-1013 (Online); 2304-1269 (CDROM); 2414-6722 (Print)

Figure 3: Graphical Summary of Western Europe Region

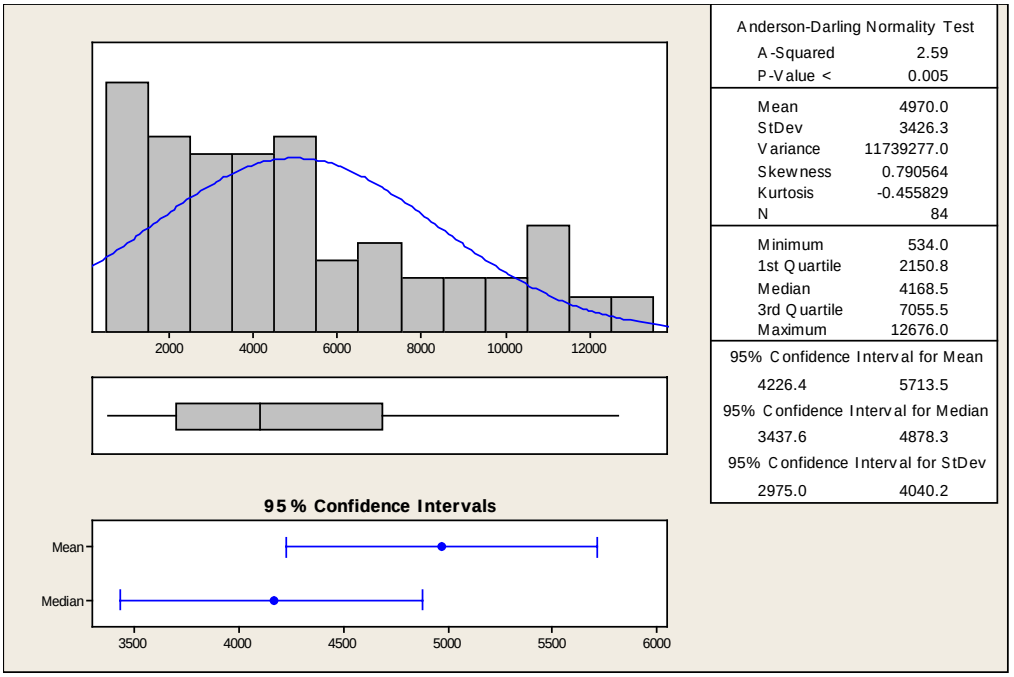



### 4.1.3 EASTERN EUROPE REGION

Graphical summary of descriptive statistics is shown in Figure 4. Minimum arrivals recorded from Eastern Europe Region were 534 where maximum 12676.The first quartile of arrivals is 2150. It means ¼ of the months had at most 2150 arrivals. A median arrival is 4168 and the third quartile of arrival is 7055. Histogram of the arrivals does not look symmetrical. The P value of the Anderson-Darling test is less than the significance level (p-value <0.005). As such number of arrivals does not follow the normal distribution. They are positively skewed.□


ISSN: 2304-1013 (Online); 2304-1269 (CDROM); 2414-6722 (Print)

Figure 4: Graphical Summary of Eastern Europe Region

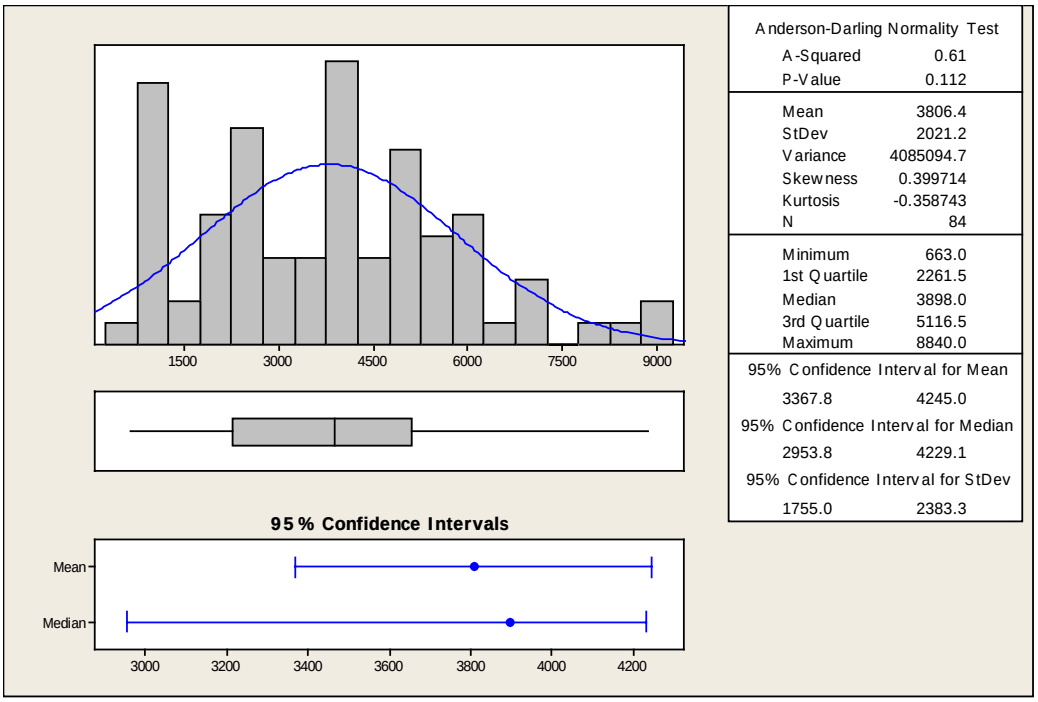



### 4.1.4 MIDDLE EAST REGION

This region is the fourth of leading four markets. According to Figure 5 of the graphical summary of descriptive statistics, minimum tourist arrivals recorded from Middle East Region were 663 where maximum 8840. The first quartile of arrivals is 2261. A median arrival is 3898 and the third quartile of arrival is 5116. Histogram of the arrivals looks symmetrical. The P value (P = 0.112) of the Anderson-Darling test is greater than the significance level (p-value > 0.05). A number of arrivals follow the normal distribution. They are positively skewed.□


ISSN: 2304-1013 (Online); 2304-1269 (CDROM); 2414-6722 (Print)

Figure 5: Graphical Summary of Middle East Region




## 4.2 PATTERN RECOGNITION OF TOURIST ARRIVALS OF ALL REGIONS

Time series plots and Auto-Correlation Functions (ACF) were obtained, region wise;

### 4.2.1 ASIAN REGION

Figure 1: Time Series Plot of Tourist arrivals from Asian Region.

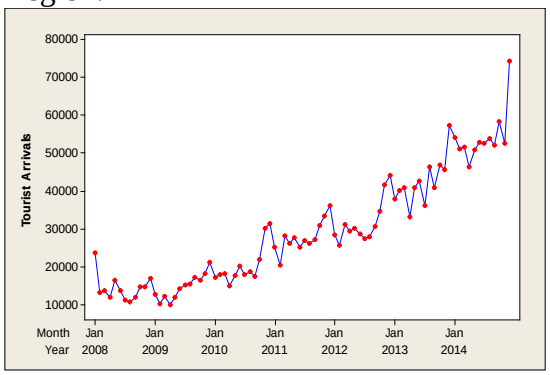
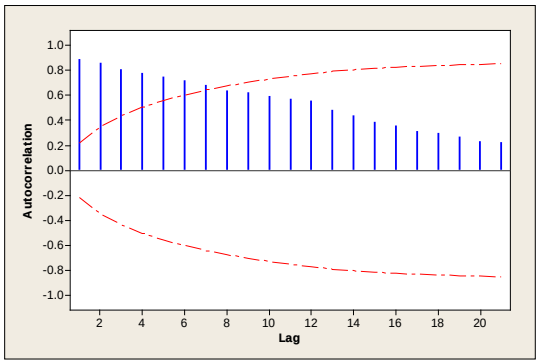



Figure 2: Autocorrelation Function of Tourist arrivals from Asian Region.




Time series plot of arrivals (Figure 1), shows the tourist arrivals from the Asian region. The behavior of the series clearly shows that there is an increasing trend from the Asian region.

A series which has constant mean and variance is known as a stationary series. Autocorrelation Function (ACF) can be used to test the stationary of a series. Figure 2 is the ACF of arrivals from the Asian region. It shows more than three significant spikes.


ISSN: 2304-1013 (Online); 2304-1269 (CDROM); 2414-6722 (Print)

It means the series is not stationary. ACF shows a decreasing pattern, suggests seasonal behavior in arrivals from the Asian region.

### 4.2.2 WESTERN EUROPE REGION

Figure 3: Time Series Plot of Tourist arrivals from Western Europe.

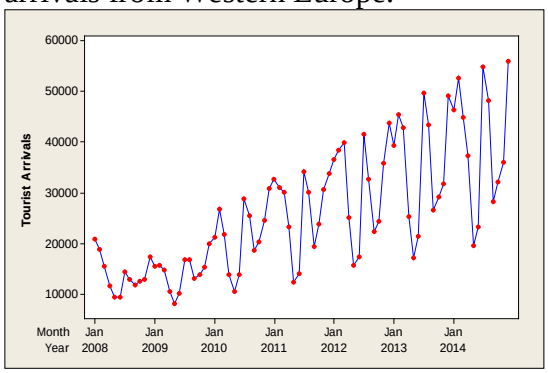
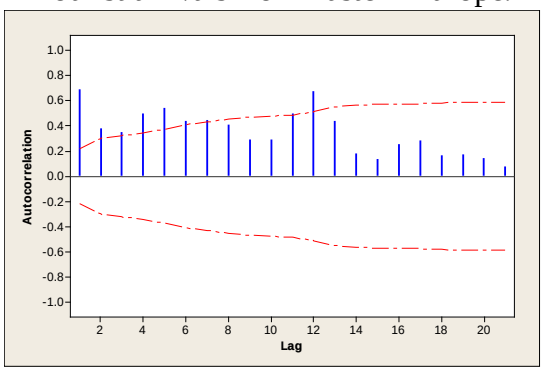



Figure 4: Autocorrelation Function of Tourist arrivals from Eastern Europe.




Time series plot of arrivals from Western Europe (Figure 3) shows an increasing trend. But fluctuation of the series is very high, increasing with time. It means arrivals from this region are not steady. Figure 4 is the Autocorrelation Function (ACF) of tourist arrivals from Western Europe. It shows that the series is not stationary.

### 4.2.3 EASTERN EUROPE REGION

Figure 5: Time Series Plot of Tourist arrivals from Eastern Europe.

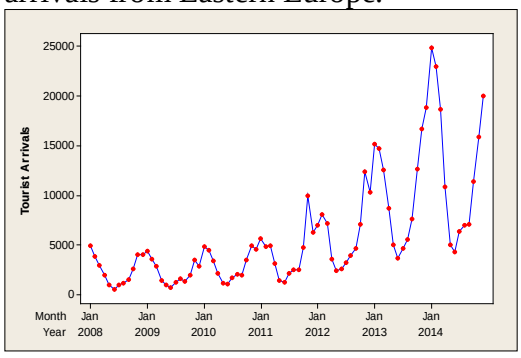
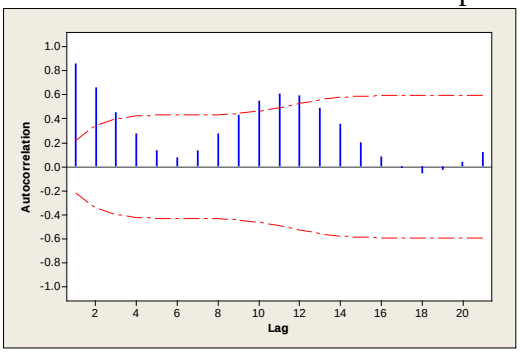



Figure 6: Autocorrelation Function for Tourist arrivals from Eastern Europe




Time series plot of arrivals from Eastern Europe (Figure 5) shows the arrivals are constant up to the year 2011 June. After that, it shows high increasing of tourist arrivals. This movement is not steady. Figure 6 is the Autocorrelation Function (ACF) of tourist arrivals from Eastern Europe. It shows that the series is not stationary.

### 4.2.4 MIDDLE EAST REGION

Figure 7: Time Series Plot of Tourist arrivals Middle East Region.

Figure 8: Autocorrelation Function for Tourist arrivals from the Middle East.


ISSN: 2304-1013 (Online); 2304-1269 (CDROM); 2414-6722 (Print)

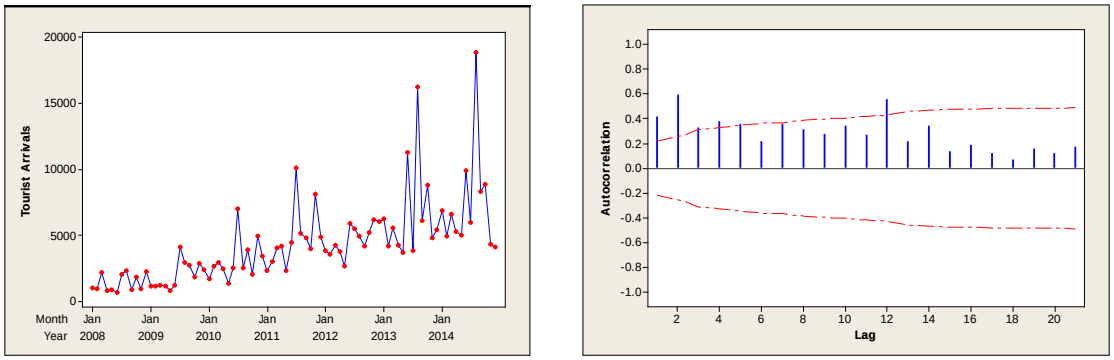



Figure 8 is the Autocorrelation Function (ACF) of tourist arrivals from the Middle East. It shows that the series is not stationary.

Figure 9: Time Series Plot of Tourist arrivals from All Regions

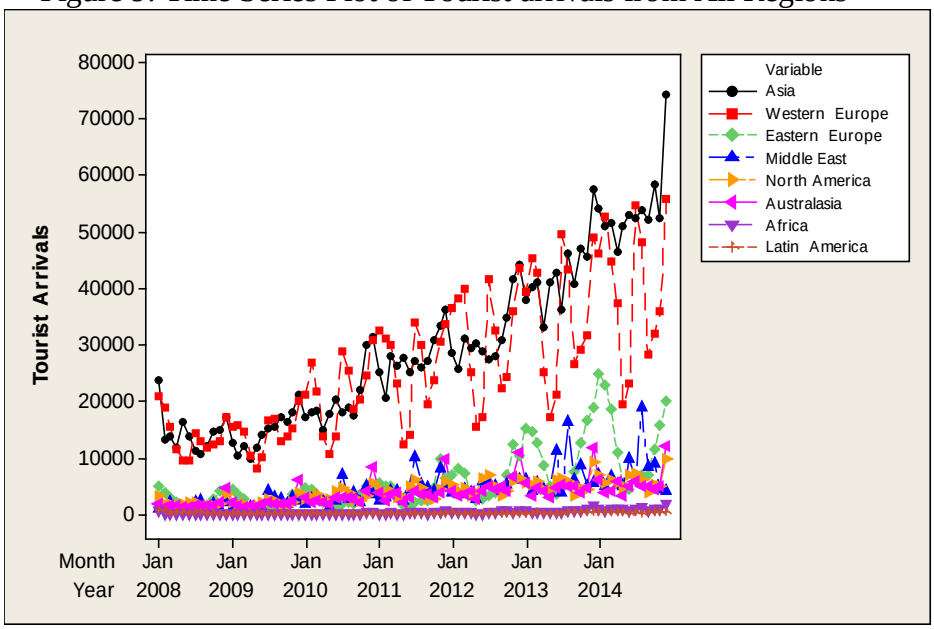



Figure 9 shows the comparison of tourist arrivals from all regions. It is clear that Asian region on top followed by Western and Eastern Europe and the Middle East.

## 4.3 COMPARISON OF ARRIVALS BY REGION

It was intended to test whether an average number of arrivals from different regions are same or not. One way - Analysis of Variance (ANOVA) was used for the purpose as follows;

$$H_0 : \mu_{Asian} = \mu_{\text{Western Europe}} = \mu_{\text{Eastern Europe}} = \mu_{\text{Middle East}}$$

$H_1$ : At least, one mean is different from the others

The P value of the ANOVA is (0.000) significant at 5% significance level. Therefore, it was concluded that at least one mean is different from others. In other words, arrivals from at least one region are different from the others. Individual 95% Confidence


ISSN: 2304-1013 (Online); 2304-1269 (CDROM); 2414-6722 (Print)

Interval (CI) for means was obtained, given in Table 1.

**Table 1: ANOVA**

| Region | N | Mean | StDev | Individual 95% CIs For Mean Based on Pooled StDev<br>-+---------+---------+---------+-------- |
|---|---|---|---|---|
| Asia | 84 | 29361 | 14693 | (*-) |
| Western Europe | 84 | 25982 | 12402 | (-*-) |
| Eastern Europe | 84 | 5866 | 5396 | (-*) |
| Middle East | 84 | 4300 | 3137 | (*-) |
| Pooled Standard Deviation | | | 7223 | -+---------+---------+---------+--------<br>0 10000 20000 30000 |

Confidence intervals (CI) for regions; Asia and Western Europe overlaps. It means there is no significant difference from arrivals of these two regions. CI's of Eastern Europe and the Middle East overlaps, confirm no significant difference from arrivals from these regions. Arrivals from both regions were below 32000. CI's for regions; Asia and Western Europe are between 20000 and 32000. They are significantly apart from the CI's of Eastern Europe and the Middle East. Whereas Eastern Europe and the Middle East are below 10000.Therefore, arrivals from Asia and Western Europe regions are higher than that of Eastern Europe and the Middle East.

## 5. CONCLUSION AND RECOMMENDATION

The purpose of this study to identify time series patterns tourist arrivals from top 4 leading regions to Sri Lanka. Patterns of tourist arrival from all the regions of Sri Lanka show increasing trends. Arrival series of all four regions were non-stationary. There was a significant autocorrelation in tourist arrivals. It means there is a significant correlation between present arrivals and the past. A significant difference of tourist arrival can be seen from Asian and Western Europe compared to other regions. The mean arrivals from both Asia and Western Europe regions fall within 24000 and 32000. Other regions were within 3000 and 6000 arrivals on average.

Based on the results; Auto Regressive Integrated Moving Average (ARIMA) methods may not be suitable in forecasting arrivals. But Seasonal ARIMA technique may appropriate Asian region. It is recommended to test Moving Average Methods, Exponential Smoothing techniques, Decomposition techniques, Circular model, linear and non-linear trend models for forecasting arrivals.

## REFERENCES

[1] Bemdt, D., & Clifford, J.(1994)."Using Dynamic Time Warping to Find Patterns in Time Series", *AAAI Technical Report WS-94-03*.


ISSN: 2304-1013 (Online); 2304-1269 (CDROM); 2414-6722 (Print)

[2] Choi, B., & Chukkapalli, R. (2009).“Applying machine learning methods for time series forecasting”, *Proceedings of the IASTED International Conference 2009 Artificial Intelligence and Applications (AIA2009)*.

[3] Konarasinghe,K.M.U.B. & Deheragoda,C.K.M. (2013). “Modeling International Tourism Demand in Sri Lanka: Parametric Approach”, *Conference Proceedings of the 2nd International Research Conference on Humanities and Social Sciences*, Faculty of Humanities and Social Sciences University of Sri Jayawardanapura, Sri Lanka,118.

[4] Konarasinghe,K.M.U.B.(2014).“International Tourism Demand on Attraction of Sri Lanka”, *Conference Proceedings of the 1st Wayamba International Research Conference*,Wayamba University of Sri Lanka,75.

[5] Konarasinghe, W .G. S., Abeynayake, N. R. (2014). “Time-series patterns of Sri Lankan stock returns", *Proceedings of the 11th International Conference on Business Management,Phd Consortium*, Faculty of management studies and commerce , University of Sri Jayewardenepura, 78-95.

[6] Konarasinghe, W.G.S., Abeynayake, N.R., & Gunaratne, L.H.P (2015). “Forecasting Share Prices of Banks of Colombo Stock Exchange”. *International Journal of Research & Review.* 2(6), 372-378.[7] Lin,J., Williamson,S., Borne, K., Barr, D.(2012). “Pattern Recognition in Time Series”.

[7] Morales-Esteban, A., Martínez-Álvarez, F., Troncoso, A., Justo, J.L., & Rubio-Escudero,C.(2010). “Pattern recognition to forecast seismic time series”, *Expert Systems with Applications* 37, 8333–8342.

[8] Singh, S.(1999). A Long Memory Pattern Modelling and Recognition System for Financial Forecasting",Pattern Analysis and Applications, 2, (3), 264-273.

[9] Sternickel, K. (2002). “Automatic pattern recognition in ECG time series”, *Computer Methods and Programs in Biomedicine*, 68,109–115.

[10] Spiegel,S., Gaebler, J., Lommatzsch,A., De Luca,E., Albayrak,S. (2011). “Pattern Recognition and Classification for Multivariate Time Series”.

[11] Sharma, P & Kaur, M (2013). “ Classification in Pattern Recognition: A Review”. *International Journal of Advanced Research in Computer Science and Software Engineering*,3(4),298-306.

[12] Sri Lanka Tourism Development Authority (SLTDA). (2014).“Summary Report 2014”, Online at http://www.sltda.lk/statistics.


ISSN: 2304-1013 (Online); 2304-1269 (CDROM); 2414-6722 (Print)