Modelling and Forecasting Tourist Arrivals in Sri Lanka

J.H. Priyangika*, P.K.B.N.M. Pallawala, and D.J.C. Sooriyaarachchi
Department of Statistics & Computer Science,
University of Kelaniya,
Kelaniya, Sri Lanka.
*hansanipriyangika2010@gmail.com

Abstract— **Tourism is one of the major industry shows a rapid growth in the Sri Lankan economy. According to the annual tourism statistics, the international tourist arrivals shows 4.4% growth in 2015 and 27.72% growth in foreign exchange earnings in the same year compared to 2014. Therefore, understanding and examining the upcoming trends of tourist's arrivals is really important and it will be beneficial and important for stakeholders and interesting parties of the country. The purpose of this research study is to investigate and forecast the tourist's arrival in Sri Lanka based on the available past data. The collected tourist's arrivals data from 2000 January to 2014 December are used for this study. For the reason that the tourist's arrivals data follow univariate time series, the time series techniques, ARIMA and GARCH models, are proposed to use in forecasting. Since the data consists with heteroscedasticity, transformation methods are needed to use in some time series modelling approaches. In GARCH model approach, original data is used for identifying a suitable model while Box-cox transform is used in SARIMA model approach to overcome the heteroscedasticity problem. Basically, model selection is done based on AIC values and MAPE, MAE and RMSE values used for measure the performance of selected models. Among the proposed time series models, none of the SARIMA models are fitted well for the data as they are not diagnostic. Finally, ARCH (1) model with optimal lag (2, 7, and 12) is identified as the best model to forecast the future values of tourist's arrivals in Sri Lanka.**

Keywords: ARIMA, Tourist arrivals Forecast, GARCH

I. INTRODUCTION

In short and simply, tourism is traveling from one place to another. The seasonal and geographical changes in early ages made people, as well as animals, travel from one place to another, mainly for their survival. This tells us that the tourism industry has long history dates back to early ages of the human race. Once the people civilized and become more stable through establishing own societies, they began to travel through geographical regions, continent to continent, country to country for many reasons.

Today, tourism has turned in to a very important industry in every region in the world. People travel for leisure, business, education and religious purposes and many more reasons. In the modern age, tourism experiences a rapid growth in worldwide. People are traveling more while finding more money for traveling, especially for leisure and adventure. Traveling for leisure has become a world phenomenon as people are attracted by nature, beauty, location, culture, religion and other fabulous things inherent to particular places all over the world. In this circumstances, Sir Lanka is a key target attracts the tourists in worldwide due to its beauty and nature. According to the World Economic Forum's index Sri Lanka [17] is ranked in 63rd place for the most tourist's friendly countries, out of total 141 countries. According to the world travel & tourism council report [18], travel and tourism are direct contributions to GDP is 1.3% and total contribution to GDP is 2.5% in Sri Lanka. On the other hand, tourism has promoted an international economical market which opens the door for new industries while developing the existing's in local and foreign context. Some of the national and international industries influenced by tourist industry are travel services, hospitality services, hotel sector, and small scale industries.

II. LITERETURE REVIEW

In literature, many research studies can be found on forecasting tourist arrivals in national and international context references. Those studies have used different methodologies to forecast the tourist arrivals based on the nature of the data. The study was done by P.C.Padhan [4] has fitted Seasonal Auto Regressive Integrated Moving Average (SARIMA) models for forecasting future tourist's arrivals in India. Fitted models are compared mainly using Mean Absolute Percentage Error (MAPE) values. Another study on forecasting tourist arrivals to Australia has used Box-Jenkins' Auto Regressive Integrated Moving Average (ARIMA) and SARIMA models in their study [1]. The authors have fitted two models for forecasting tourist arrivals from Hong Kong, Malaysia and

Singapore and fitted models were compared using MAPE and Root Mean Square Error (RMSE) values. Another study [2] used exponential smoothing and ARIMA models to forecast tourists' arrival in Kenya and they have identified the exponential smoothing model as the best model to predict the values. Two studies, one based in Taiwan [3] and other on Hong Kong [5] have forecasted tourist arrivals using ARIMA, exponential smoothing and advanced time series models, Artificial Neural Network (ANN) techniques and compared fitted models using RMSE, MAPE, and Mean Absolute Deviation (MAD) values. Econometrics models such as GARCH (Generalized Autoregressive Conditional Heteroscedasticity), GJR–GARCH (Glosten, Jagannatham and Runkle GARCH) and EGARCH (Exponential GARCH) are used in [6] to forecast tourist arrivals and fitted models are compared using MAPE, MAD, and RMSE values to identify the best model for forecast tourist arrivals in China. They have concluded the GARCH models are accurate than SARIMA models for their study.

Forecasting models can be found for some other aspects too. For example, in the study [7] has fitted SARIMA models to forecast future values of rubber yield. Another study [8] based on forecast expected dengue patients has used SARIMA models in their study. Another study [9] forecasted Malaysian gold price by fitting ARIMA and GARCH models and MAPE values are used to identify the best model. Finally, GARCH (1,1) model is selected as the best model for forecast gold price in Malaysia. The study by Vincenzo Pacelli [10] has used several advanced time series techniques, ANN, ARCH, and GARCH models, for forecast exchange rates Euro/US dollar while S.Z.S. Abdalla modelling exchange rate volatility [11] using GARCH models. In [12], volatility of USD/MUR exchange rates were forecasted using GARCH models and compared fitted models using Mean Absolute Error (MAE) and RMSE values. The study [13] done by K.A.D.S.A.Nanayakkara, N.V. Chandrasekara and D.D.M. Jayasundara has fitted GARCH and ANN models for forecast exchange rates and compared fitted models usingNormalized Mean Square Error (NMSE), MAE, Directional Symmetry (DS), Correct Up trend (CU) and Correct Down trend (CD) values.

Literature emphasis that [14] Box-cox transform is the best transformation method for normalizing the data and removing heteroscedasticity compared to traditional transformations methods such as square root, log and inverse method. Also, evidence can be found

in the literature [15] to conclude that the GARCH model can be used when hetroscedasticity errors present in the original time series data. Therefore, based on the literature review and statistical analysis, this study used Box-cox transformation before applying the SARIMA model techniques while raw data is used when GARCH models are identified as a suitable forecasting model.

III. MODEL TECHNIQUES

A. SARIMA model

Seasonal AutoRegressive Integrated Moving Average Process is a time series contain a periodic component, which repeats every observation. A general SARIMA model can be written as

$$\phi_p(B)\Phi_p(B^s)W_t = \theta_q(B)\,\Theta_q(B^s)\varepsilon_t \qquad (1)$$

Where B denotes the Backward shift operator, ϕ_p, Φ_p , θ_q , Θ_q are polynomials of order p, P, q, Q respectively. ε_t denotes a purely random process and $W_t = \nabla_d \nabla_s^d Y_t$ denotes the differenced series.

B. GARCH model

The GARCH model, on the other hand, has the ability to model time-varying conditional variances. For univariate series let,

$$y_t = \mu + a_t \qquad (2)$$

be a mean equation at a time, where μ is the conditional mean of the y_t and a_t is the shock at time t and $a_t = \varepsilon_t \sigma_t$, where, $\varepsilon_t \sim$iid(0, 1). Then a_t follows a GARCH (p,q) model if

$$\sigma_t^2 = \alpha_0 + (\alpha_1 \varepsilon_{t-1}^2 + \alpha_2 \varepsilon_{t-2}^2 + \cdots) + (\beta_1 h_{t-1} + \beta_2 h_{t-2} + \cdots)$$

$$= \alpha_0 + \sum_{i=1}^{q} \alpha_i\, a_{t-i}^2 + \sum_{i=1}^{p} \beta_i\, \sigma_{t-i}^2 \qquad (3)$$

Where $\alpha_0 > 0$, σ_t^2 is the conditional variance of y_t, $\sum_{i=1}^{\max(p,q)} \alpha_{1i} + \beta_i < 1$, α_i, β_i are the coefficients of the parameters ARCH and GARCH respectively.

C. Box-Cox transform

It is envisioned as a panacea for simultaneously correcting normality, linearity and reducing homoscedasticity [13]. The formula of the Box-Cox transformation is

$$y_t *= \begin{cases} \dfrac{y_t^\lambda - 1}{\lambda} & ; \text{For } \lambda = 0 \\ \log_e y_t & ; \text{For } \lambda \neq 0 \end{cases} \qquad (4)$$

Where y_t is the actual data at the time t, $y_t *$ is the transformed data at time t, and λ is the minimum residual mean square error value. This transformation is valid only for positive series, $y_t > 0$.

IV. RESULTS AND DISCUSSIONS

Figure 1 displays the monthly tourist arrivals in Sri Lanka for January 2000 to December 2014.According to Figure 1; tourist arrivals data show an upward trend and seasonal pattern. Also, data display non constant variance. Therefore, based on the time plot it can be concluded that tourists' arrivals increasing with the time. It is noted that after May 2009 (time point 113) pattern has significantly changed and represents a multiplicative data pattern in the time plot. Further, data before the aforesaid time point shows an additive data pattern. Therefore, the tourists' arrivals data show combined additive and multiplicative features with seasonal patterns.

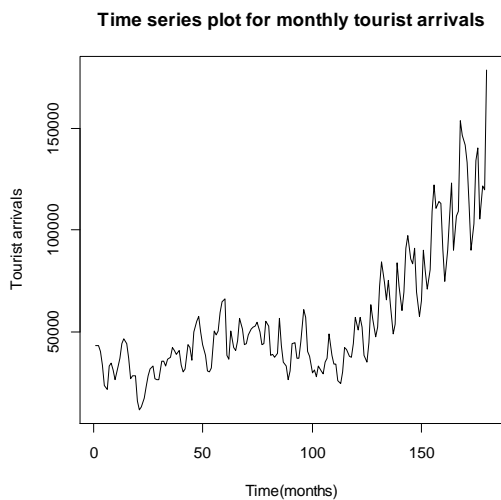**Time series plot for monthly tourist arrivals**

Figure 2. Time series plot of tourist arrivals data

Since the preliminary data show a non-stationary pattern, Box-Cox transform was performed. In this regards, unit root test, ADF and KPSS tests, were used to test the stationarity of the raw data. According to those tests, transformed tourist arrivals data are not stationary at 1% level of significance, therefore, the first difference of
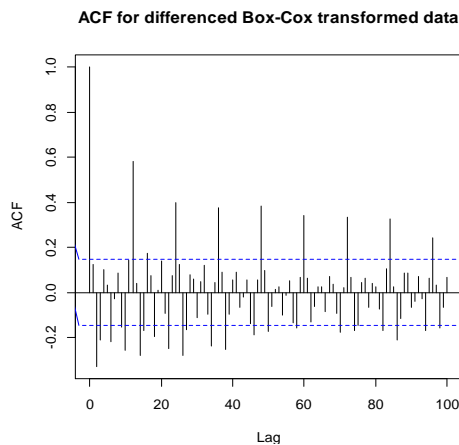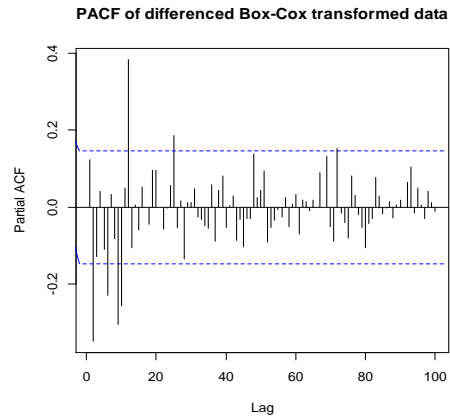
ACF for differenced Box-Cox transformed data

Figure 3. ACF of differenced Box-Cox transformed data

PACF of differenced Box-Cox transformed data

Figure 1. PACF of differenced Box-Cox transformed data

TABLE II. SELECTED SARIMA MODELS

Model	AIC value
SARIMA$(3,1,3)(1,0,1)_{12}$	211.14
SARIMA$(3,1,4)(1,0,1)_{12}$	213.17
SARIMA$(0,1,3)(1,0,1)_{12}$	234.19

transformed data was considered. Unit root tests confirmed that the differenced transformed data are stationary at 1% level of significance.

Figure 2 and Figure 3 depict several SARIMA model combinations. Three models were selected based on the minimum AIC values and it is represented in Table 1. SARIMA $(3, 1, 3) (1, 0, 1)_{12}$ model selected as the best model as it has the minimum AIC value. Jarque-Bera, Lujng-Box and ARCH-LM tests were used to test the diagnostics of the model. According to above tests, residuals of the selected model normally distributed with nearly zero mean and constant variance but they are not independently distributed. Therefore, this selected SARIMA model cannot be used to forecast tourist arrivals in Sri Lanka. Also, other selected SARIMA models' diagnostic checking was failed and hence suggested that SARIMA models are not suitable for forecasting the tourist arrivals in Sri Lanka based on the selected data set.

For fitting GARCH model, original tourist arrivals data was used. The presence of ARCH effect in the data is a basic requirement in applying the GARCH model. Therefore, using ARCH-LM test it was identified that the data has ARCH effect. Then the identification of GARCH models for the data is continued. Since data were collected in monthly, lag (1), lag (2), ... , lag (12) were used. Several GARCH models for several lag combinations were identified and finally, three significant models were selected based on the AIC values listed in Table 2.

16

xTABLE II.   SELECTED ARCH MODELS

Model	AIC value
ARCH(1) with LAG (2,7,12)	21.3026
ARCH(1) with LAG (2,12)	21.3046
ARCH(1) with LAG (4,7,12)	21.30839

ARCH (1) with optimal LAG (2, 7, and 12) model selected as the best model as it has minimum AIC value. The diagnostics tests were done for the residuals of the selected model. According to test results, residuals of selected model is normally distributed, residuals have no serial correlation and residuals have no ARCH effect at 1% level of significance which satisfies all the assumption of the technique. Therefore, selected model is used to forecast tourist arrivals in Sri Lanka for January 2015 to September 2015. Eq. 5 represent the selected model.

Mean equation: $Y_{t} = -5169.959 + 0.220243Y_{t-2} + 0.092979Y_{t-7} + 0.898348Y_{t-12} + \varepsilon_t$

Variance equation:

$$\sigma^2 = 62252667 + 0.439517\ \varepsilon_{t-1} \qquad (5)$$

Figure 4 represents the behaviour of actual values and the forecasted values using best ARCH model. Forecasted values nearly show a pattern of actual values.

Table 3 represents the accuracy measurement of the ARCH (1) with optimal lag (2, 7, and 12) model.

TABLE III.   FORECAST ACCURACY MEASUREMENTS OF SELECTED MODEL

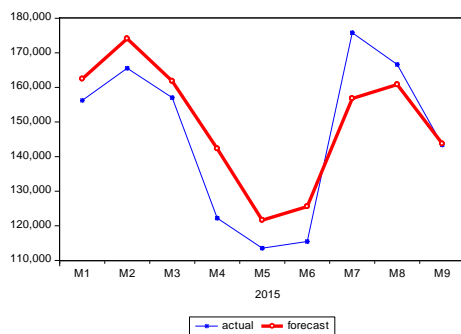| Model	MAPE	MAE	RMSE
ARCH(1) with lag(2,7,12)	6.555842	9215.111	11051.42363

Figure 4. Plot of actual and forecast values

V. CONCLUSION

Tourist arrivals data consisted with hetroscedasticity errors. Therefore, Box-Cox transform was used to transform data to remove the hetroscedasticity errors. SARIMA model was fitted to the transformed tourist arrivals data and

GARCH model was fitted for the original tourist arrivals data. However, suitable SARIMA model could not be fitted for the data and finally ARCH (1) with optimal lag (2, 7, and 12) model was identified as the best model for forecast tourist arrivals in Sri Lanka. Tourist arrivals from January 2015 to September 2015 were used to test the model adequacy. Using selected model tourist arrivals can be forecasted with 6.556% accuracy.

REFERENCES

[1]   C. Lim and M. McAleer, "Time series forecasts of international tourism demand for Australia", Discussion paper no 533, 2001.

[2]   A.O. Akuno, M.O. Otieno, C.W. Mwangi and L.A. Bichanga, "Statistical Models for Forecasting Tourists' Arrival in Kenya", Open Journal of Statistics,60-65, 2015.

[3]   C.J. Lin, H.F. Chen and T.S. Lee, "Forecasting Tourism Demand using Time series, Artificial Neural Networks and multivariate Adaptive Regression Splines:Evidence from Taiwan", International Journal of Business Administration, 2011.

[4]   P.C. Padhan, "Forecasting tourists' footfalls in India:An Assortment of Completing Models", International journal of business and Management,Vol 6, No 5, 2010.

[5]   V. Cho, "A comparison of three different approaches to tourist arrivals forecasting", Tourism Management, Vol.24, No 3, 323-330, 2003.

[6]   Y.C. Chen, H.H. Kang, and T.C. Yang," Impacts of Volatility on Forecasting Inbound tourists into China", Research gate Publication.

[7]   P.K.B.N.M. Pallawala and D.D.M. Jayasundara, "Forecast future value of rubber yeild and cost of Products by fitting the best Time Series models", European International journal of Science and Technology, Vol 2, No 5, 2013.

[8]   H.W.B. Kavinga, D.D.M. Jayasundara and D.N.K. Jayakody, "A new dengue outbreak Statistical Model using the Time Series Analysis", European International Journal of Science and Technology,Vol 2, No 10, 2013.

[9]   P. Y.Ping, N.H. Miswan and M. H. Ahamad, "Forecasting Malaysian Gold using Garch model", Applied Mathematical Sciences, Vol. 7, No 58, 2879-2884, 2013.

[10] V. Pecelli, "Forecasting Exchange rates: a Comparative Analysis", International Journal of Business and Social Science, Vol. 3, No. 10, 2012.

[11] S.Z.S Abdalla, "Modelling exchange rate volatility using Garch models, Emparical evidance from Arab Countries", International Journal of Economics and Finance, Vol. 4, No. 3, 216-229, 2012.

[12] D.N.C. Vee, P.N. Ganpot and N. Sookia, "Forecasting Volatility of USD/MUR Excahge rate using a GARCH(1,1) model with GED and student's-t errors", University of Mauritius research journal, Vol. 17, 2011.

[13] K. A.D.S.A. Nanayakkara, N.V Chandrasekara and D.D.M Jayasundara," Forecast Exchage rates using Time Series and Neural Network Approach", European

International Journal of Science and Technology, Vol. 3, No. 2, 2014.

[14] J.W. Osborne, "Improving your data transformations: Applying the Box-Cox transformation", patrical assessment, Research & evaluation,Vol 15, No 12.

[15] T. Bollerslev, Genaralized Autoregressive Consitional Hetriskedasticity, Journal of Econometrics 31, 394-419.

[16] Annual Statistical Report 2015, retrived from:http://www.sltda.lk/sites/default/files/tourist_boar d_annual_report_new-1.pdf.

[17] Global competitiveness report (1/11/2015), retrived from world economic forum site:http://reports.weforum.org/global-competitiveness-report-2014-2015.

[18] UNWTO reports retrieved from the UNWTO official website:http://dtxtq4w60xqpw.cloudfront.net/sites/all/fi les/pdf/unwto_barom15.