

Research Article

Forecasting International Tourists Arrival to Nepal Using Autoregressive Integrated Moving Average (ARIMA)

Rajendra Prasad Upadhayaya

Ph.D. Scholar, Central Department of Statistics, Tribhuvan University, Kirtipur

Corresponding email: raju.nma.np@gmail.com

Article History

Received 11 April 2021

Revised 14 October 2021

Accepted 12 November 2021

ABSTRACT

This paper analyses the tourist arrival in Nepal during the last 56 years using various quantitative techniques. By the use of autoregressive integrated moving average (ARIMA) model in projecting the tourist arrival in Nepal based on historical data. Box & Jenkins methodology has been employed to forecast a variable using a database of international tourist arrival. Autocorrelation function (ACF) and partial autocorrelation function (PACF) have been used along with Ljung-Box Statistics for the test of stationary. Parameters p , d , q has been identified based on different diagnostic test statistics such as stationary R^2 , MAPE, Normalized BIC, ACF, and PACF. Conclusively, the model ARIMA (1, 1, 1) has been selected with the least BIC and tourist arrival forecasted to be 1,322,000 (LCL: 919,090 and UCL: 1,724,909) in the year 2025.

Keywords: ARIMA, international tourist, ACF, PACF, Box & Jenkins, forecasting

INTRODUCTION

Tourism is one of the world's largest industries and it represents a major area of interest due to its enormous impact on national economies and people's lives. It is a surprisingly strong and resilient sector of economic activities and a fundamental contributor to economic recovery by generating foreign currency and creating millions of jobs. The tourism sector can contribute directly and indirectly to all of the sustainable development goals (SDGs) (UNWTO, 2015). Tourism contributes highly to gross domestic product [GDP], increasing the employment rate,

source of revenue for local people, private sector, public sector, and government (Akuno, Oteieno, Mwangi, & Bichanga, 2015).

Tourism suffered the greatest crisis on record in 2020 following unprecedented health, social and economic emergency amid the outbreak of the COVID-19 pandemic. International tourism arrivals are decreased by 74% in 2020 than the previous year due to widespread travel restrictions and a massive drop in demand with the losses of USD1.3 trillion in export revenues. Asia and the Pacific saw an 84% decrease in international arrivals in 2020, about 300 million less than in the previous year. The Middle East and Africa both recorded a 75% drop in arrivals. In Europe arrivals declined by 70%, representing over 500 million fewer international tourists, while the Americas saw a drop of 69 percent (UNWTO, 2021). In Nepal, tourist arrival is decreased by 80.78% in 2020 (MoHA, 2021).

Tourism is a dynamic industry that is highly climate-sensitive and faces many challenges that affect its development. To cope with these challenges, the planner, policymaker, and the tourism industries apply the process of forecasting to furnish the information, based on which decisions are taken before the occurrence of that events. To make a robust plan as a base for formulating tourism policy, a valid and reliable estimate of the future demand must be obtained. So the main aim of the introducing forecasting process in tourism is to maximize the benefit to stakeholders by ensuring minimal negative effects, costs, and impacts (Wilkinson, 1997; Goeldner & Ritchie, 2006; Song & Li, 2008; Edgell, DelMastro Allen, Smith, & Swanson, 2008). The significance of tourism has encouraged the author to study the number of international tourist arrival and attempt to make accurate forecasting for future planning.

This paper reviews few recent studies which applied different time series models in forecasting tourist arrivals. A study carried out in India applied Holt-Winters Exponential Smoothing and ARIMA models in forecasting foreign tourist arrivals (Sood & Jain, 2017). Tourist arrival in Thailand using different time series models (Chaitip, Chaiboonsri, & Mukhjang, 2010); use of SARIMA and GARCH model in Sri- Lanka (Priyangika & Pallawala, 2016); and applied ARIMA model in an attempt to forecast a wholesale price of maize in Tanzania (Abdulqader, 2018). ARIMA model is extensively used to forecast tourist arrivals (Chang & Liao, 2010; Tularam, Wong & Nejad, 2012; Suh, Park, Song & Cho, 2014; Jeong, 2016).

DATA AND METHODS

Secondary data were obtained from the Ministry of Culture, Tourism and Civil Aviation Planning & Evaluation Division Research & Statistical Section, Kathmandu, Nepal. The data consists of yearly international tourist visits in Nepal from 1965 to 2019 B.S. long-time data of international tourist arrivals. In this study, the Box & Jenkins time series model was used to analyze and forecast international tourist arrivals in Nepal. To identify the tentative models among the AR, MA, ARMA, ARIMA, the autocorrelation function (ACF) and partial autocorrelation function (PACF) were used. In statistics and econometrics, an autoregressive integrated moving average (ARIMA) model was used to fit time series data and to predict future points. ARIMA model was used where the data show evidence of non-stationary and differencing method was used to eliminate nonstationary (Box, Jenkins & Reinsel, 1994).

Stationarity of the Time Series

A series of data is said to be stationary if it is independent of the time at which the data was recorded. The time series value is affected by the trends and seasonality whose impact was removed by differencing the time series values to make them stationary. Besides, a white noise series is stationary and should be appeared much the same at any period. Stationary time series data has mean and variance constant over time. The stationarity of the series was tested by the time plot of the given data points.

Box-Jenkins Procedure

Box and Jenkins model building consists of four steps to find the best model to fit within the class of ARIMA models. The first step consists of model identification. This step involved the tentative identification of the model order, that is identifying the values of (p, d, q). Usually, the value of (p, d, q) was selected based on two plots of ACF and PACF. The second step was focus on model identification and selection and involved in the estimation of the model parameter identified. The method of maximum likelihood was used. Bayesian information criterion (BIC) is a criterion to select the best model. The model was the best having the minimum BIC value among the selected models. The value of BIC is calculated mathematically,

$$BIC = k \ln(n) - 2 \ln(\hat{L});$$

Where \hat{L} = Maximum likelihood function's value of the model, n= number of data points, k= number of the free parameter to be estimated.

The third step consists of diagnostic checking to identify whether the fitted model

sufficiently captured the information in the time series data. Box and Jenkins (1970) had proposed an autoregressive moving average ARMA (p, q) to predict the variable using the past and present database. The model was given as

$$Y_t = a_0 + a_1 Y_{t-1} + a_2 Y_{t-2} + \dots + a_p Y_{t-p} - b_1 \varepsilon_{t-1} - b_2 \varepsilon_{t-2} - \dots - b_q \varepsilon_{t-q} + \varepsilon_t$$

Where p and q denote the order of autoregressive parts and moving average respectively and ε_t is a white noise type process.

An ARIMA model is a generalization of an ARMA model. These models were fitted to time series data either to better understand the data or to predict future points in the series. The model is usually referred to as an ARIMA (p, d, q) model where p, d, q are integer greater than or equal to zero and refer to the order of autoregressive, integrated, and moving average part of the model respectively (Mills, 1990).

Ljung- Box Test

The Ljung-box test may be defined in terms of hypothesis as:

H_0 : There is an absence of serial correlation.

H_1 : There is the presence of serial correlation.

The test statistics, $Q = n(n+2) \sum_{k=1}^h \rho^2 / \sum_{k=1}^h (n-k)$

Where, n= sample size, $\rho =$ sample autocorrelation at lag k, and h=number of lags being tested

The null hypothesis is rejected that the residuals are white noise if $Q > \chi^2_{\alpha, h-k}$ where α is the level of significance. The last step involved predicting the future international tourist arrivals using a fitted model. Based on the fitted model ten years forecast was generated from 2020 to 2029.

Model Validation and performance Evaluation

The Root Mean Square Error (RMSE) and mean absolute percentage error (MAPE) was employed for the validation and performance of the forecasting model. A model was best having a minimum value of RMSE and MAPE that is close to zero.

$$RMSE = \sqrt{[(\sum(X_t - F_t))^2 / n]}$$

$MAPE = 100/n[\sum|(X_t - F_t)/X_t|]$, Where X_t is the observed international tourist's arrival and F_t is the forecasted tourists' arrival in the corresponding years. The number of observations in the validation periods is denoted by n.

RESULTS AND DISCUSSION

ARIMA (p, d, q) model was used to predict the number of international tourists arriving in Nepal for a decade. Appropriate values of p and q were obtained by calculating the autocorrelation function (ACF) and partial autocorrelation (PACF) and by drawing the corresponding graphs of the two functions (Singh, 2013). The first step in developing the model was the identification of trends and seasonality of the given data with the help of the graph in figure 1.

Figure 1

Trends of International Tourists

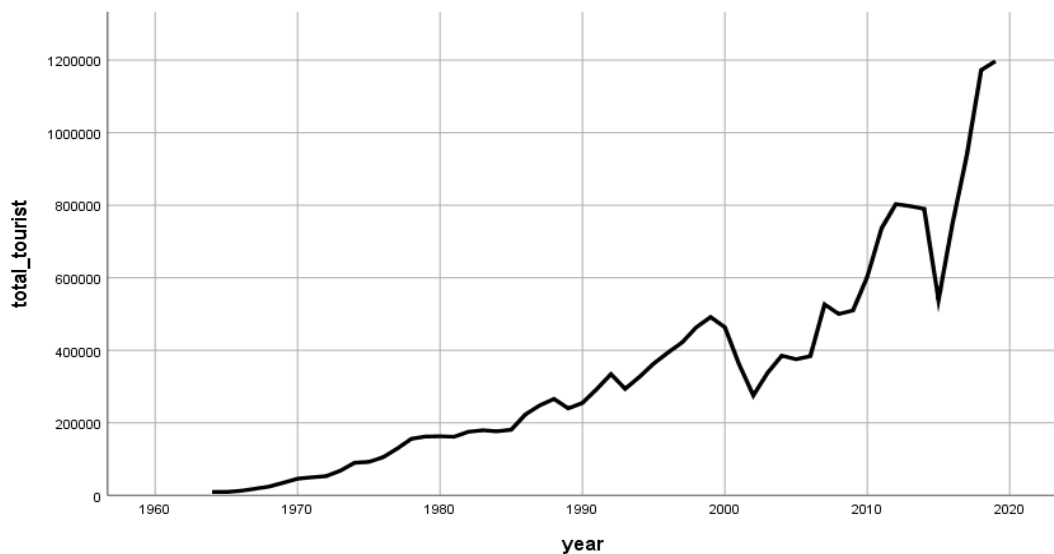


Figure 1 shows a generally increasing trend over a specified period, indicates that the series is not stationary and is essential for differentiating the datasets. The stationary feature of the series and the auto-regressive component of the data is detected with the help of autocorrelation and partial autocorrelation function of the international tourists' arrival information and are essential for further analysis. The autocorrelation function and partial autocorrelation function have been presented in figure 2.

Figure 2

ACF and PACF

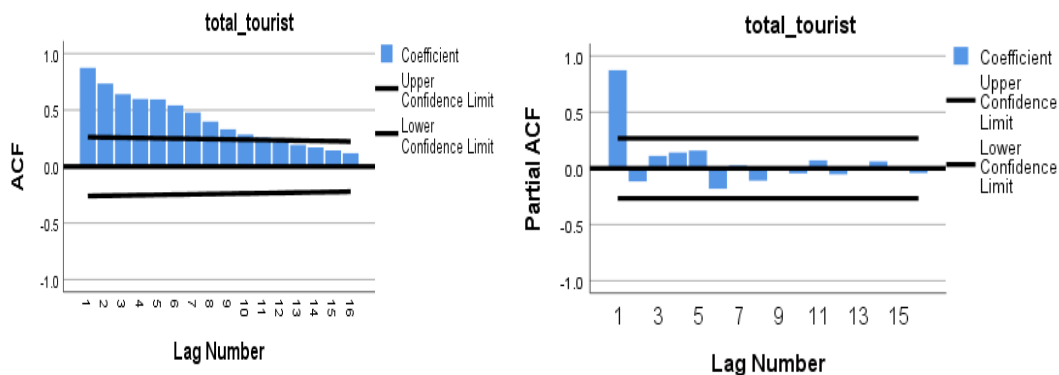


Table 1

Autocorrelation and Partial Autocorrelation

SN	Lag Auto.	Corr.	BJ Statistics	p value	Partial Auto correlation
1	.873	45.006	1	.000	.873
2	.735	77.459	2	.000	-.116
3	.640	102.593	3	.000	.111
4	.597	124.855	4	.000	.140
5	.595	147.391	5	.000	.158
6	.542	166.490	6	.000	-.182
7	.478	181.641	7	.000	.025
8	.398	192.362	8	.000	-.109
9	.330	199.908	9	.000	-.016
10	.284	205.608	10	.000	-.045
11	.257	210.376	11	.000	.072
12	.228	214.227	12	.000	-.053
13	.191	216.988	13	.000	.016
14	.169	219.195	14	.000	.062
15	.144	220.829	15	.000	-.021
16	.118	221.957	16	.000	-.041

Autocorrelation is high in the first lag and is decreasing slowly for higher lag and is statistically significant which is the indication of trends in the data. Hence the data is nonstationary. The non-stationarity of the data is also checked by Ljung-Box statistics (Table 1) which are also significant indicating the presence of non-stationary. It has to pursue to distinguish, is the data of tourist arrival is a random walk? If so, then the first difference of

the time series data should be stationary which can be presented in figure 3. The graph shows mean and variance fairly constant over time, the strong indication of stationary data.

Figure 3

Graph After First Differencing

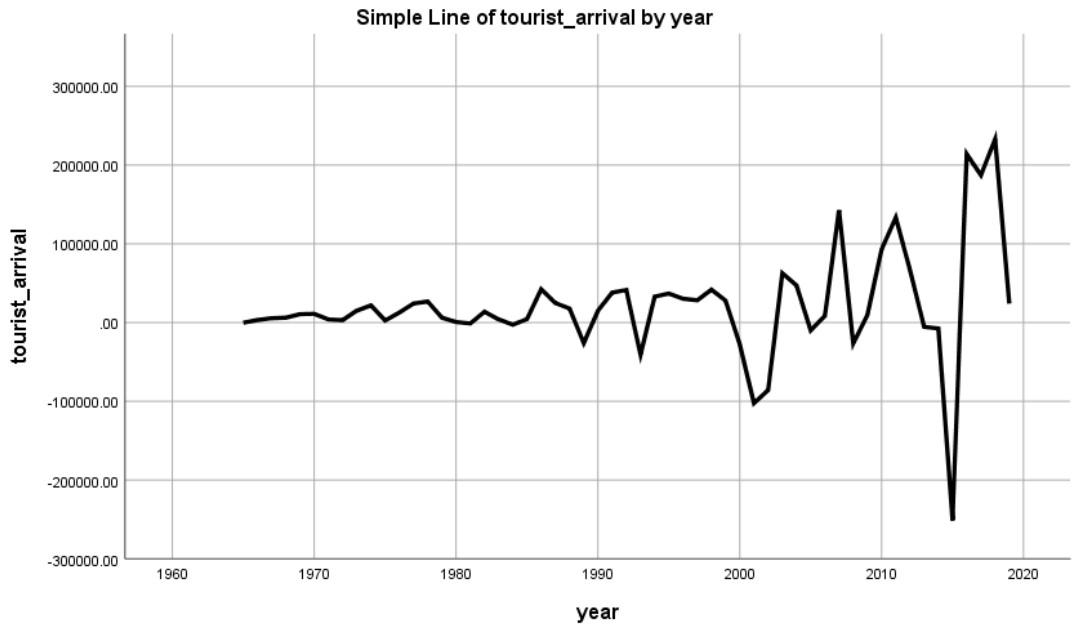


Figure 4

ACF and PACF After First Differencing

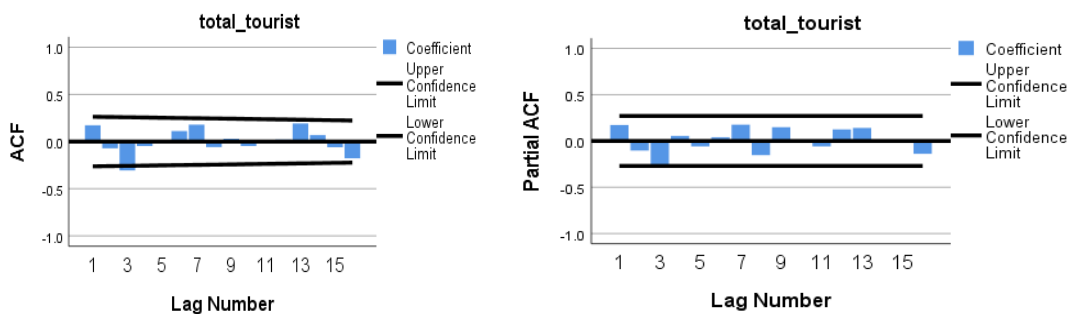


Figure 4 shows that the partial autocorrelation coefficient and autocorrelation of the differencing the original time series data lies between the specific range (confidence interval). This figure indicates the presence of stationary data and the obligation of change, suggesting the differentiation of order 1 for these variables when applying regression analysis.

Model Adequacy

There is no rigid procedure in establishing an exact ARIMA (p, d, q) model. An iterative process is necessary to complete through which 3 parameters will take different values, then will be applied a criterion for constant models found. The best model is selected based on using R², MAPE, and Normalized BIC criterion, has been chosen the following adjustment models for data series with diagnostic in table 2.

Table 2

Diagnosing the ARIMA Model for the Number of Tourists

Number of Predictors	Model Fit statistics					Ljung-Box Q(18)			Number of Outliers
	Stationary R-squared	R-squared	RMSE	MAE	Normalized BIC	Statistics	DF	Sig.	
1	.074	.94	69768.3	38292.9	22.6	17.3	16	.366	0

Best-Fitting Model has been selected based on stationary R², R², MAPE, and Normalized BIC (larger stationary R², R², smaller MAPE, and smaller Normalized BIC indicates better fit). For the international tourists’ arrival, the ARIMA (1, 1, 1) model was found to be appropriate. Diagnosing an ARIMA model is a crucial stage of the construction process of the model building that involves the verification of random distribution of residues. By analyzing the R² indicator (Table 2), it has been observed that the built model explains 94.3% of the variation in the series is explained by the given variables. Also, the Ljung-Box test is statistically insignificant. Hence the model can be considered feasible.

Figure 5

Residual for ACF and PACF

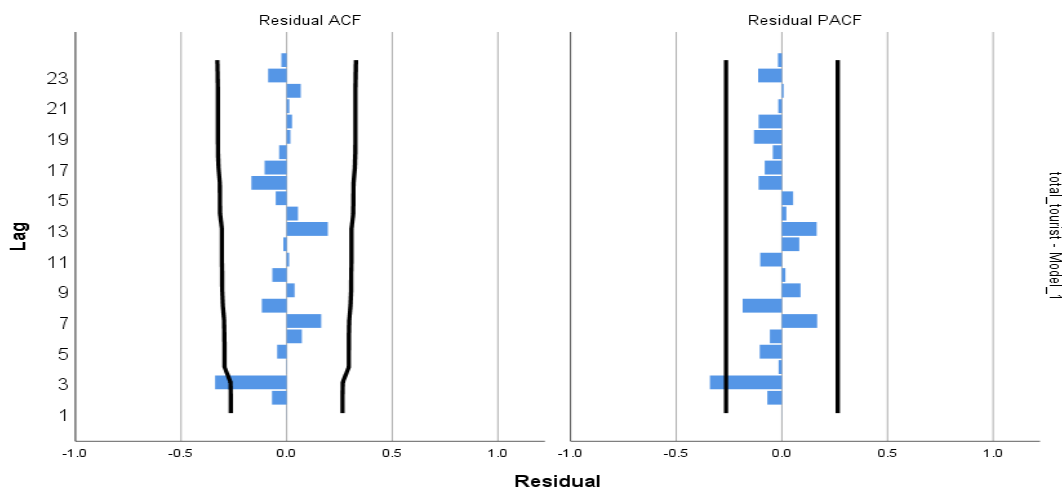
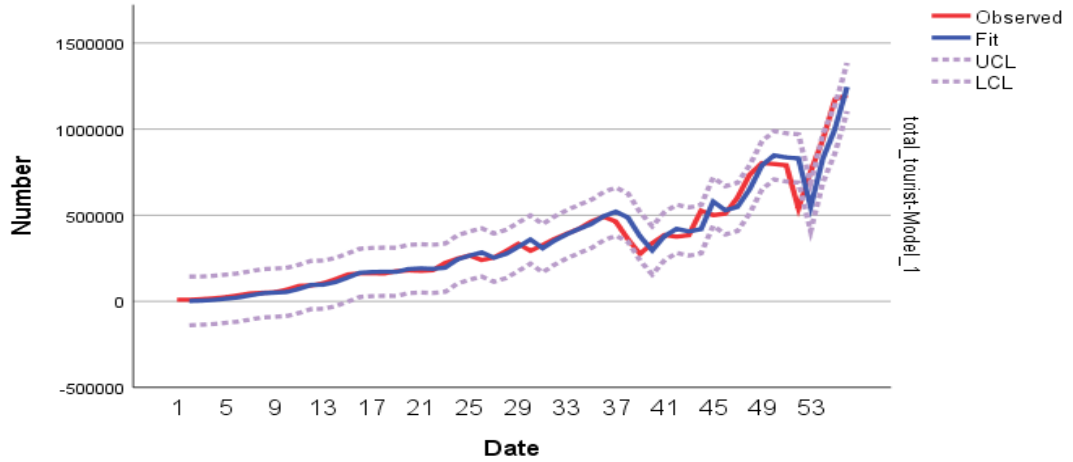


Figure 6

Observed and fitted trends by model



It can be observed the fact that the predictor models are better in estimating the observed variation in datasets during the analyzed period. For the variable tourist arrival, the univariate ARIMA (1, 1, 1) model has been used to predict the tourist arrival. After applying the model, the predicted values for a variable of interest, from 2020 to 2029 are as follows. In table 3.

Table 3

Forecast Tourist Arrival with UCL and LCL

Model \ year	2020	2021	2022	2023	2024	2025	2026	2027	2028	2029
Model_1 Forecast	1214983	1236206	1257644	1279095	1300548	1322000	1343452	1364904	1386357	1407809
UCL	1356739	1455504	1534458	1603432	1666287	1724909	1780381	1833388	1884401	1933755
LCL	1073227	1016909	980830	954758	934808	919090	906523	896420	888312	881863

DISCUSSION

Tourism plays an important role in the economic growth of Nepal, particularly in terms of job creation and GDP. Therefore, predicting the future number of international tourist’s arrival is crucial for tourism planning and marketing. As the trends are shown in figure 1, tourist arrivals are increasing, indicating the tourism industry is raising and its effects in creating a new job, investment in the tourism sector, increase the percentage of contribution to the GDP. The data indicate that the tourist arrival is decreased in 1965 due to clash between India and Pakistan; in 1978-82 democracy movement and oil crisis; in 1989 trade and transit argument between Nepal and India; in 1993 increment in visa fees, limited airlines services due to THAI and PIA

air accident, destruction due to flood and landslide, increasing air pollution in Kathmandu, and political unrest (Nepal Banda, Chakkajam and political movement); in 2000, 2001, and 2002 hijacking of Indian aeroplane on (24 Dec. 1999) Kathmandu to Delhi route, Maoist insurgency, shifting of Indian tourist to other destination; in 2008 strike in Suvarnabhumi International Airport in Thailand Bangkok; in 2015 major earthquake on 25, April 2015 and unofficial blockade of India; in 2019 COVID-19 pandemic. Hence the observed trend is increasing but abruptly decreased from 2019 during the COVID-19 pandemic.

Box–Jenkins methodology was applied to find an appropriate ARIMA model to forecast the international tourists' arrival in the country. The results show that ARIMA (1,1,1) is a best-fitted model and the model was used to generate annual forecasts from 2020 to 2029 with a 95% confidence interval. Conclusively, the tourist arrival is forecasted to be 1,322,000 (LCL: 919,090 and UCL: 1,724,909) in the year 2025 and the tourist arrival is forecasted to be 1,407,809 (LCL: 881863 and UCL: 1,933,755) in the year 2029. The forecasts indicate an upward trend in the international tourism demand will continue soon, pointing to positive impulses and outlooks of a continuous increase.

CONCLUSION

The economic growth is directly proportional to the number of tourists. Tourism can play a vital role in poverty reduction by empowering the women, youth, and ethnic group by promoting their culture, entrepreneurship, and small business (eg. tea shop, handicraft shop, cultural program) through the provision of jobs and income-generating opportunities. Based on the forecasted international tourists' arrival to Nepal the plan must be set to motivate the private sector and government agencies for the infrastructure developments, upgrading the quality of service and products, improvement in information system, reviewing the visa policies, and sustainable development of the tourism sector. Hence these results are fundamental elements in the formulation of well-coordinated policies, they cannot do the job all by themselves but outcomes may be presented as a framework. Although the precision of the proposed ARIMA model can be judged by various statistical tests and found to be good, valid, and satisfactory. However, the proposed model has some trouble in forecasting with nominal errors, implying different models would be required for testing optimality. Therefore, this research may address the application and modeling to forecasts.

REFERENCES

- Abdulqader, Q. (2018). Annual forecasting using a hybrid approach, *General Letters in Mathematics*, 4 (2), 86-95. doi:10.31559/glm2018.4.25
- Akuno, A. O., Oteieno, M. O., Mwangi, C. W., & Bichanga, L. A. (2015). Statistical model for forecasting tourist arrivals in Kenya. *Open Journal of Statistics*, 60-65. doi:https://doi.org/10.4236/ojs.2015.51008
- Box, G. E., Jenkins, G. M., & Reinsel, G. C. (1994). *Time series analysis: Forecasting and control*. (3rd ed.). Prentice-Hall. doi:ISBN-10:0130607746
- Chaitip, P., Chaiboonsri, C., & Mukhjang, R. (2010). Time series model for forecasting international visitor arrivals to Thailand, *International Conference on Applied Economics. ICAE 2008*, (pp. 159-163).
- Edgell, D. L., DelMastro Allen, M., Smith, G., & Swanson, J. R. (2008). *Tourism policy and planning: Yesterday, today, and tomorrow*. Elsevier.
- Goeldner, C. R., & Ritchie, J. R. (2006). *Tourism: Principles, practices, philosophies*. John Wiley & Sons.
- Mills, T. C. (1990). *Time series techniques for economists*. Cambridge University Press. doi:ISBN 0-521-34339-9
- MoHA. (2021). *Arrival/departure summary 2020*. Department of Immigration.
- Priyangika, j. H., & Pallawala, S. D. (2016). Modeling and forecasting tourist arrival in Sri Lanka. *Symposium on statistical & computational modeling with an application*, (pp. 14-18).
- Singh, E. H. (2013). Forecasting tourist inflow in Bhutan using seasonal ARIMA. *International Journal of Science and Research (IJSR)*. doi:ISSN: 2319-7064
- Song, H., & Li, G. (2008). Tourism demand modeling and forecasting: A review of recent research. *Tourism Management*, 29, 203-220.
- Sood, S., & Jain, K. (2017). Comparative analysis of techniques for forecasting tourist arrival, India. *Journal of Tourism & Hospitality*, 6 (3), 3-6. doi:https://doi.org/10.4172/2167-02691000285
- UNWTO. (2015). *Tourism and the sustainable development goals*. World Tourism Organization (UNWTO).
- UNWTO. (2021). *UNWTO World tourism barometer and statistical annex*. UNWTO.
- Wilkinson, P. (1997). *Tourism planning on islands*. Cognizant Communications.