Forecasting of International Tourists Arrival in Nepal: An application of ARIMA

Nanda Kumar Tharu
Department of Statistics, Tri-Chandra College, Tribhuvan University., Kathmandu, Nepal

Author for Correspondence, Email: *nanda.tharu71@gmail.com*

Abstract
This paper presents the use of autoregressive integrated moving average (ARIMA) model in making projection for the tourist arrival in Nepal based on past tendency. Box & Jenkins methodology have been employed to forecast a variable using as a database of international tourist arrival. Aautocorrelation function (ACF) and partial autocorrelation function (PACF) have been used along with Ljung-Box Statistics for the test of stationery. Parameters p, d, q have discovered based on different diagnostic test statistics such as R^2, MAPE, Normalized BIC, ACF and PACF. Finally, the model ARIMA (1, 1, 1) has been picked with least BIC and tourist arrival found to be 1,353,133 (UCL: 1,670,863 – LCL: 1,035,398) in the year 2020.

Key words: ARIMA, Box & Jenkins, Forecasting, GDP, International Tourist

I. Introduction

According to World Tourism Organization, international tourism can be define as an activity of visitors who make temporary visits across international borders and remains for more than 24 hours (WTO,2008)

In the early twenty-first century, tourism is one of the leading industries in the world economy. It holds number of employees, participation in social product, national income and total consumption. The number of tourists has significant impacts on global economy and gross domestic product (GDP).Tourism sector can contribute directly or indirectly to all of the sustainable development goals (SDGs). Particulary, Goals 8, 12 and 14 has been targeted by it on inclusive and sustainable economic growth, sustainable consumption and production (SCP) and the sustainable use of oceans and marine resources, respectively(UNWTO, 2015).

Among various benefits, the two that can be generated with tourism are process of creating jobs and direct impact on the GDP. Eulalia Claros and Alessandra Di Tella (2004) studied that the European Parliament Research Service, tourism activities in the accommodation and food services sector had almost 10 million jobs (4% of total EU employment) in 2013. Tourism is an important growing profession to many country of the world and forms as growing portion of the world's economy.

Tourism industry is a very important source of economic growth and development in many countries around the world but it is more prominent in developing countries like Maldives (79.4% of the GDP was contributed by Tourism in 2016), Azerbaijan (20.33% of GDP), Cambodia (32.4% of GDP in 2017), Hong Kong, China (11.4 % of GDP in 2017), Philippines (24.7 % of GDP in 2018), Bhutan (6% of GDP in 2009 second only to Hydropower), etc. Similarly, South Asia (8.9 % of total GDP in 2016), Worldwide (10.2 % of total GDP in 2016).

Over 238 million jobs had created by the travel & tourism economy had and contributed to 9.9% of global GDP in 2008, expected to provide more than 296 million jobs and 10.5% global GDP by 2018 (WTTC, 2017). Similarly, during 2013 international tourists were 1087 million and generating 9% of the world's GDP and creating 1 in every 11 of the jobs around the world by tourism industry (Jarvis et al., 2016). This means that the tourism sector is vast enough to be able to create a massive volume of job vacancies and can contribute to growing the other macroeconomic indicators like GDP.

Tourism is one of the important industry in Nepal and it has high source of foreign receipt. This research aims to predict the international tourist arrivals to Nepal in the year 2019 and 2020 and to seek the best forecasting model for tourist arrivals based on past trends.
The results thus obtained has been useful for tourist industry of Nepal. The tourist industry has been impacted on processing of job generation, revenue, GDP and directly or indirectly to all sustainable development goals (SDGs).

II. Methodology

Data

Secondary data was obtained from Ministry of Culture, Tourism & Civil Aviation Planning & Evaluation Division Research & Statistical Section, Kathmandu, Nepal. The data was taken from 1993 to 2018 years long time series data of international tourist arrivals in Nepal.

Model

In statistics and econometrics, particularly in time series analysis, an autoregressive integrated moving average (ARIMA) model is a generalization of an autoregressive moving average (ARMA) model. Both of these models are fitted to time series data to predict future points. ARIMA models are employed in cases where data show evidence of non-stationary. An initial differencing step can be applied one or more times to eliminate the non-stationary (Box & Jenkins et al., 1994).

A stationary time series is those series that do not depend on the time at which the series is recorded. So time series with trends, or with seasonality, are not stationary **-** the trend and seasonality will affect the time series value at different times. Furthermore, a white noise series is stationary and should appear much the same at any period of time. The plots of the series will show mean and variance constant over time.

Box & Jenkins (1970) have proposed a methodology to forecast a variable using as a database only it's past and present.
An autoregressive-moving average ARMA (p,q) model has an autoregressive type component or an average moving type component. The model is given as

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

Where p is the order of the autoregressive part, q is the moving average order and is a white noise type process (a sequence of independent and identically distributed random variables with zero mean).

The ARMA models are suitable for stationary series. These were generalized for non-stationary series that become stationary by differentiation; the resulting models are called autoregressive integrated-moving average ARIMA (p, d, q) where d is the order of differentiation required for stationary series (Mills, 1990).

Ljung-Box Test

The Ljung–Box test may be defined in term of hypothesis as:
H_0: The data are independently distributed (absence of serial correlation)
H_a: The data are not independently distributed (presence of serial correlation).

The test statistic is $Q = n(n+2)\sum_{k=1}^{h} \frac{\hat{\rho}_k^2}{n-k}$

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

Bayesian Information Criterion (BIC)

Bayesian information criterion (BIC) is a criterion for model selection among a finite set of developed models. Lower the BIC value, better the model is.

Mathematically BIC can be written as: BIC = $\ln(n)k - 2\ln(\hat{L})$;

Where, $\hat{L}$ = maximum value of likelihood function of the model, n = number of data points, k = number of free parameters to be estimated.

The analysis was executed using SPSS Version 20.

III. Results

There was an attempt for retrieving an ARIMA (p, d, and q) model to forecast the values for the future periods for international tourist arrival in Nepal. The important stages of the Box-Jenkins strategy is the identification of the 3 parameters p, d, and q. Appropriate value of p and q parameters are obtained by calculating the autocorrelation function (ACF) values and partial autocorrelation function (PACF), and by drawing the corresponding graphs of the two functions (Singh, 2013).

The first step in building the model is to make graphics for data series as in figure 1.

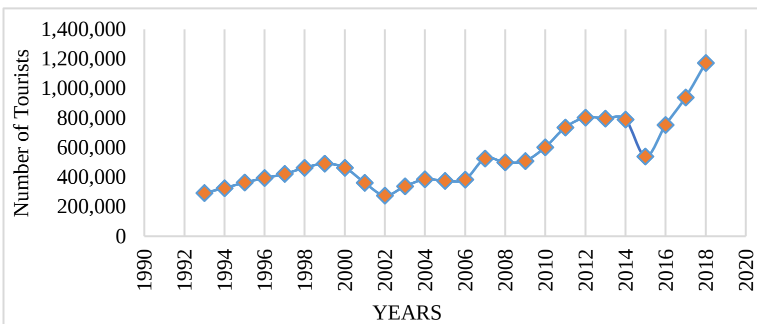

Figure 1. Trends of International Tourist

The trend component can be observed in figure 1, indicating the necessity for differentiate the data series. For lucid evidence of the stationary feature and of the auto-regressive component of data series, both the autocorrelation and the partial autocorrelation of the values of tourist arrival information is essential to analysis. The autocorrelation function and partial autocorrelation function has been seen in figure 2.

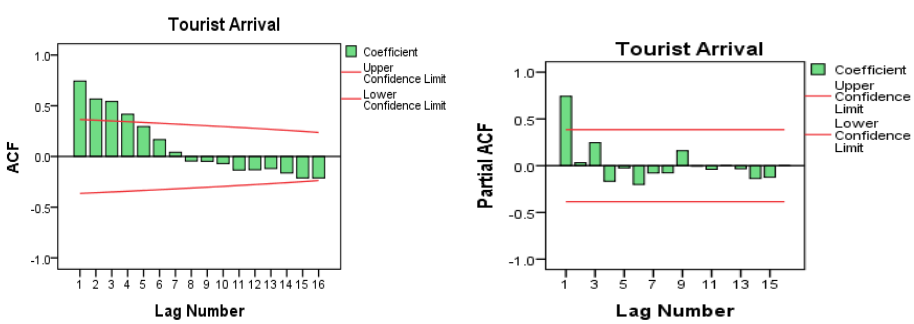

Figure 2. ACF and PACF

Table 1. Autocorrelation and partial autocorrelation

Lag	Autocorrelation	Partial Autocorrelation	Box-Ljung Statistic		
			Value	df	p-value
1	0.703	0.703	14.405	1	0.000
2	0.480	-0.030	21.381	2	0.000
3	0.359	0.065	25.469	3	0.000
4	0.410	0.279	31.031	4	0.000
5	0.340	-0.130	35.043	5	0.000
6	0.266	0.019	37.626	6	0.000
7	0.159	-0.071	38.596	7	0.000
8	0.031	-0.229	38.635	8	0.000
9	-0.042	0.006	38.711	9	0.000
10	-0.099	-0.115	39.154	10	0.000
11	-0.144	-0.082	40.157	11	0.000
12	-0.206	-0.018	42.371	12	0.000
13	-0.204	0.031	44.697	13	0.000
14	-0.240	-0.081	48.202	14	0.000
15	-0.285	-0.060	53.580	15	0.000
16	-0.305	-0.003	60.369	16	0.000

Figure 2 shows high correlations, all of them are statistically significant and sizable even at higher-lags, and slowly decreasing. This is a definite indication of non-stationary series. The Ljung-Box statistics are also statistically significant shows non-stationery series.

Next, it has seek to know, is the series of tourist arrival a random walk? If so, then the first difference of the time series should be a stationary time series which can be presented in figure 3. The plot shows mean and variance fairly constant over time, the strong indications of stationary series.

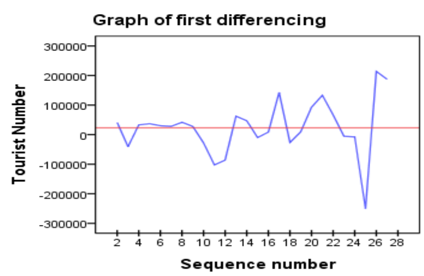

Figure 3. Graph after first differencing

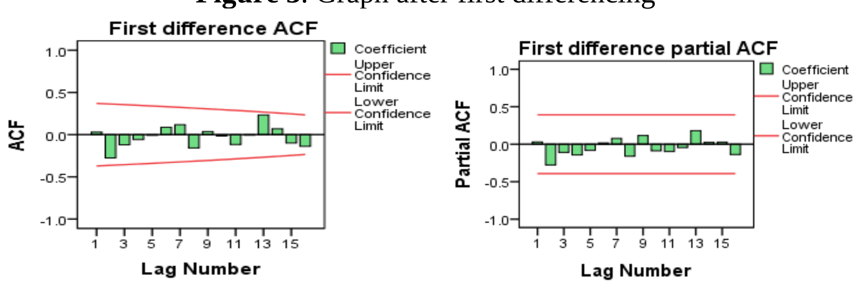

Figure 4. ACF and PACF after first differencing

Figure 4 reveals that there is no partial autocorrelation coefficient above the critical level of significance; that indicates the absence of non-stationary data series and the necessity of change, implying the differentiation of order 1 for these variables when applying a regression analysis.

Model Adequacy

However, there is no fixed algorithm in establishing an exact ARIMA (p,d,q) model optimally; it is necessary to complete an iterative process through which the 3 parameters will take different values, then will be applied a criterion for contestant models found. By refining different 8 best models using the R^2, MAPE and Normalized BIC criterion, it has been chosen the following adjustment model for data series with diagnostic in table 2.

Table 2. Diagnosing the ARIMA model for the number of tourist

Model	Model Fit statistics			Ljung-Box Statistic			Number of Outliers
	R^2	MAPE	Normalized BIC	Statistics	d.f.	p-value	
Tourist No. Model_1	0.81	13.458	23.614	13.980	16	0.600	0

Best-Fitting Models according to R-squared, MAPE and Normalized BIC (larger R^2, smaller MAPE and smaller Normalized BIC indicates better fit)

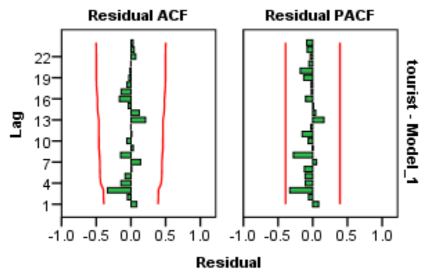

Figure 5. Residual for ACF and PACF

For the tourist arrival, the ARIMA (1, 1, 1) model was found. Diagnosing an ARIMA candidate model is a crucial stage of the construction process of the final model that involves the verification of random distribution of residues. First, by analyzing the R^2 indicator (Table 2) it has been observed that the built model explains 81% of the variation in the series. Also, the Ljung-Box test is not statistically significant, so the model can be considered feasible.

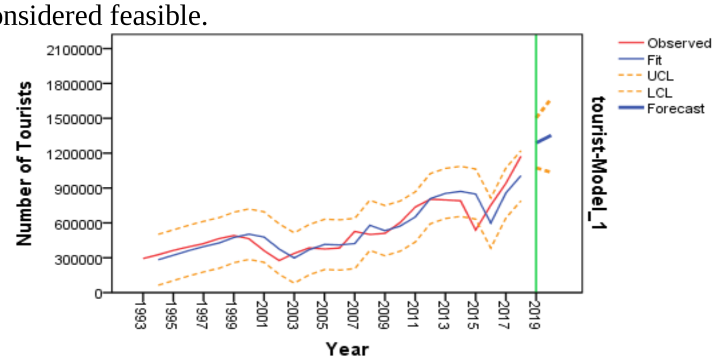

Figure 6. Observed and fitted trends by model

It can be seen clearly the fact that the predictor models are better in estimating the recorded variation in data series 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 variable of interest, during 2019, and 2020, is in table 3.

Table 3. Predicted value for the next 2 years

Model		2019	2020
Tourist No.-Model_1	Forecast	1,288,564	1,353,133
	UCL	1,503,735	1,670,868
	LCL	1,073,393	1,035,398

IV. Discussion and Conclusions

Tourism in Nepal plays a significant role in strengthening the economy, particularly in terms of job creation and contribution on GDP. As per the latest trends shown in figure 1, tourist arrivals is raising, implying tourism industry is raising and its effects contribute to the real economic growth by stimulation of the growth of GDP, creating new jobs, motivating entrepreneurship, encouraging investment etc. Hence, anticipating international tourism demand is necessary for creating favourable conditions and images of the country.

The design of the Box–Jenkins methodology, ARIMA(1,1,1) model, is built up and proposed for short-run forecasting of international tourism in Nepal based on annual data. This research forecasts that the upward trend in the international tourism demand will continue in the near future, pointing to positive impulses and outlooks of a continuous increase by 2020. Clearly, higher the tourists, higher the economic growth is. Tourism can impact at community level, can play the role of national poverty reduction by promoting entrepreneurship and small businesses, and empowering less favored groups, particularly youth and women. Tourism can empower women in multiple ways, particularly through the provision of jobs and through income-generating opportunities.

Furthermore, the paper clarifies that the proposed model does not deliver 'the solution', but only supports in finding international tourists. Even though the model results are essential elements in the preparation 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 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 modelling to forecasts.

References:

1. Box, G. E., Jenkins, G. M. & Reinsel, G. C. (1994). *Time Series Analysis: Forecasting and Control.* 3rd Edition, Prentice Hall, ISBN-10: 0130607746, pp: 592
2. Eulalia, C. & Tella, A. (2004). Tourism in EU Economy. European Parliament Research Service, Statistical spotlight.
3. Jarvis, D., Stoeckl, N., & Liu, H.B. (2016). The impact of economic, social and environmental factors on trip satisfaction and the likelihood of visitors returning. *Tourism Management*. 52(1), 1–18.
4. Mills, T. C. (1990). *Time Series Techniques for Economists.* Cambridge University Press. ISBN 0-521-34339-9.
5. Nepal Tourism Statistics (2017). Government of Nepal. Ministry of Culture, Tourism & Civil Aviation. Planning & Evaluation Division. Research & *Statistics.*
6. Singh, E. H. (2013). Forecasting Tourist Inflow in Bhutan using Seasonal ARIMA. *International Journal of Science and Research (IJSR), India Online ISSN: 2319-7064*
7. UNWTO (2015). Tourism and the Sustainable Development Goals. Published and printed by the World Tourism Organization (UNWTO), Madrid, Spain
8. WTTC (2017). The economic impact of travel & tourism Maldives
9. World Tourism Organization (2008). Youth travel matters: Understanding the global phenomenon of youth travel. *WTO: Madrid, Spain.*