

**TIME SERIES ANALYSIS OF INTERNATIONAL TOURIST ARRIVAL TO ETHIOIA
2006-2015, A STATISTICAL ANALYSIS**

Agatamudi Lakshmana Rao, Bruk Semma and Tafere Tilahun

Department of Statistics, College of Natural Sciences, Jimma University, Jimma, Ethiopia.

ABSTRACT

The aim of this research is time series analysis of international tourist arrival to Ethiopia, it focused on the monthly tourist arrival from January 2006 to December 2015 G.C. The goal of this study is to explore the number of tourist arrival to fit a Time series model for the monthly number of tourist arrival and to forecast a two year ahead of the number of tourist arrival. The analysis was done by using statistical software packages using this software's and knowledge of time series analysis, trend, seasonal, ACF, PACF and Box- Jenkins analysis computed. From the trend plot the tourist arrival is fluctuate from month to month, quarter to quarter as well as from year to year. There is arrival fluctuation from month to month (not stationary) the minimum and maximum record of tourist arrival is 19995 and 88149 observed in the year 2006 and 2015 respectively. By differencing the data one times, a seasonally adjusted autoregressive moving average (SARIMA)(1,1,2)(0,1,1) model with seasonality factor of 12 was fitted for making a one-year ahead forecast. Proper model adequacy checking was done. One year ahead forecast showed that November, January, and December are the months with the most prominent values, and Tourist arrival expected to be 931238, which is 7.24% increment from 2015 total number of tourist arrival and in 2017 tourist arrival to Ethiopia may increase that is expect to be 991554 tourist may come to Ethiopia. This is 6.1 % increment than previous year tourist arrival 2016 and 12.9 % increment form 2015. And possible recommendations are Seasonality of tourist arrival to Ethiopia implies there is no equal demand in tourist flow

Keyword: Seasonal factors, Trend component, irregular factors and Non-stationary time series.

1. INTRODUCTION

1.1. Background of the study

Tourism as an industry has been travelling with the wild pace of technological advancements and aboard is people from different places and cultures interacting with increasing since, the globe had been shrunk into a village.

Tourism being one of the biggest and fastest growing industries globally, its benefits and the challenges, keenly observed by governments affects the economic, socio-cultural, environmental and educational resources of nations (Bethapudi, 2015).

Tourism exports have become an important sector in many countries as a growing source of foreign exchange earnings. This has arisen through the rapid expansion of international tourism, which is mainly attributed to high growth rates of income in developed and newly industrialized countries and the substantial decrease in real transportation costs between countries. Besides

generating foreign exchange earnings and alleviating the balance of payments problems encountered in many countries, international tourism also creates employment. As a labor-intensive industry, it absorbs an increasing percentage of the workforce released from agriculture and the manufacturing industries, and prevents large-scale unemployment. Other benefits contributed by international tourism include increasing income, savings, investment, and economic growth (Dwyer, 2008).

Tourism is one of the largest and rapidly growing industries in the world. According to the World Tourism Organization (UNWTO, 2007), there're 846 million international tourist arrivals in 2006 only, which should an increase of 5.4% over the previous year. However, the developed world is taking the lion's share of the market with Europe, North America and East Asia claiming 76.3% the international tourists in the same year (Tadesse, 2012). Though noted for its tourism potential, Africa's underdeveloped tourism sector is attracting only 4.81% (40.7 million) of the total tourist arrivals in the world. What makes the problem severe is the fact that a considerable proportion of this number is taken by South Africa and Northern African countries (Tadesse, 2012).

Ethiopia's great potential for tourism development is mentioned everywhere. It suffices to say that it has almost all types of primary tourist products: historical attractions, national parks with endemic wild life and cultural and religious festivals. UNESCO recognizes eight world heritage sites (as many as Morocco, South Africa and Tunisia and more than any other country in Africa): Axum's obelisks, the monolithic churches of Lalibela, Gondar's castles, the Omo Valley, Hadar (where the skeleton of Lucy was discovered), Tia's carved standing stones, the Semien National Park, and the walled city of Harar (Tadesse, 2012).

The new branding 'Ethiopia is the origins of Land ', surely can attract more Foreign Direct Investment (FDI), tourist influx, entrepreneurs and enhance export and find new markets for products originated from the country. Not only that, it was also reflect the country's true image and correct the misrepresented attributes associated with it (Beyene, 2016).

The positive effects of tourism on a country's economy include the growth and development of various industries directly linked with a healthy tourism industry, such as transportation, accommodation, wildlife, arts and entertainment. This brings about the creation of new jobs and revenue generated from foreign exchange, investments and payments of goods and services provided. Though improvements in the standard of living of locals in heavily visited tourist destinations is usually little or non-existent, inflation of the prices of basic commodities, due to visiting tourists, is a constant feature of these areas (Bethapudi, 2015). A better understanding of the trends of tourist arrival help to find the highest and the lowest point of tourist arrival in terms of month which also guide the concerned body to analysis the future demand related to the number of tourist arrival. In addition to this forecasting of tourist arrival help different organization whose work are related to tourist in making their plan.

1.2. Statement of problem

Ethiopia's tourism potential offers, on top of its diverse flora and fauna, a trove of historical, cultural, religious and archaeological attractions that allow a ground for claiming 'most African countries could be no match'. Though much has been said and sung about Ethiopia's tourism resource, the nation has barely made a dent into the huge prospects (Beyene, 2016).

As far as developing the tourism infrastructure is concerned, Scholars said that the government must prioritize tourist destinations. "The tourism development should also be in advantage for the community so that it will affect its responsibility of taking care of the nearby tourism sites (Beyene, 2016).

The situation in Ethiopia is even worse. On the one hand, its tourism potential is diversified: natural attractions that include some of the highest and lowest places in Africa along with immense wild life including some endemic ones; a very old and well preserved historical traditions with fascinating stealer, churches and castles to witness that, an attractive cultural diversity of about 80 nations and nationalities; and various ceremonies and rituals of the Ethiopian Orthodox Church which open a window on the authentic world of the Old Testament On the other hand, it is one of the poorly performing countries in terms of tourist arrivals.

The under development of most of Ethiopian infrastructure and the poor service in the tourism sector has contribute to the sector not achieving its potential. There is need focus on the improvement of tourism related infrastructure and the provision of customer based services in the tourism sector. This is particularly so with restaurant and hotels since these are the main contributors to GDP and employment in the tourism sector.

The number of transit visitors in Ethiopia is directly related to airport efficiency, strong security and growth of the Ethiopian Air Lines. And except during the Ethio-Eritrean war and its aftermath (1998-2001), this number has grown steadily to register a five-fold increase in 2005 from the 1991 record. The recent growth is mainly explained by the growth of the Ethiopian Air Lines as one of the best airlines in Africa (World Bank, 2006). Almost every year, the number of visitors whose purpose was to visit relatives should a continuous but slow increase in the period under study. Still more than 10% of the tourists' purpose of visiting Ethiopia is not known (Tadese, 2012).

Some research have done related to tourist flow, but those research focused on analyzing factors affecting tourist flow like "Determinates of tourist flow in Ethiopia" (Mulualem, 2010),"The role of privet sector in tourist flow" (Mulatu ,2015).However this research focused on forecasting future tourist arrival and analyzing seasonality of tourist flow to Ethiopia.

Research question

- What are the highest and the lowest point of tourist arrival in terms of year and month?
- Is there seasonal variation in tourist arrival?
- What will be tourist arrival in next two year? Expect to be increasing or decreasing?

2. METHODOLOGY

2.1 Description of the study area and Population

This study conducted on tourist arrival to all over Ethiopia. Ethiopia was known as Abyssinia until the twentieth century. It is the oldest independent country in Africa.

Ethiopia is located at 3 degree and 14.8 degree latitude, 33 degree and 48 degree longitude in the Eastern part of Africa lying between the Equator and the Tropic of Cancer. It is bounded on the Northeast by Eritrea and Djibouti, on the east and Southeast by Somalia, on the south by Kenya and on the west and Northwest by Sudan. Ethiopia is a rugged, landlocked country split by the Great Rift Valley. With archaeological finds dating back more than 3 million years, it's a place

of ancient culture. Among its important sites are Lalibela with its rock-cut Christian churches from the 12th–13th centuries. Aksum is the ruins of an ancient city with obelisks, tombs, castles and Our Lady Mary of Zion church.

2.2 . Variables was consider in the research

The study was conducted based on quantitative variables. Quantitative variables are variables which can be expressed numerically.

Dependent variable

The number of monthly tourist arrived to Ethiopia.

Independent variable

Time measure at which tourist arrived in each month (2006-2015).

2.3.Method of Data Analysis

To analysis the data of this study the research have been used different statistical methodology.

2.3.1. Explanatory statistics

Descriptive statistics enable to determine about the general information on the tourist arrival to Ethiopia. Descriptive statistics which deals with describing(explaining) characteristics of aggregate of statistical data by method of organizing and presenting(tables, graphs) and it provides the numeric summary of central tendency and variability like mean, minimum, maximum and measures central variations.

2.3.2. Time Series Analysis

Time series is a set of data pertaining to the value of a variable at different time. In other words any sequence of measurement taken on response that is, variable over time is called time series. It is sometimes studied simply because of historical interest but mostly because of the interest in future predicting the value of the variable at the future date. A few definition of time series are given below:-

- A time series consists of statistical data which are collected, recorded or observed over successive increment.
- A set of data depending on the time is called time series.
- A time series is a set of statistical observations arranged in chronological order.

Time series analysis used for:-

The analysis of time series is of great significance not only to the economists and business man but also to the scientists, gastronomists, geologists, sociologists, biologists, research worker etc. for the reason below:-

- It helps in understanding past behavior:-By observing data over a period of time one can easily understand what changes have taken place in the past. Such analysis will be extremely helpful in predicting the future behavior.
- It helps in planning future operations:-The major use of time series analysis is in the theory of forecasting. The analysis of the past behavior enables to forecast the future. Time series forecasts are useful in planning, allocating budgets in different sectors of economy.

2.3.2.1. Time series plot

Time series plot is the most frequently used form of graphic design to show no obvious change in the variance over time and then the research says the series has constant variance and the mean, or no evidence of change in the mean over time. It is also used to examine stationarity of time series data.

2.3.2.2. Components of Time Series

Trend Component: - general tendency of a time series data to increase or decrease or stagnate during a long period of time. An upward tendency is usually observed in time series relating to population, production sales, income money in circulation while downward tendency is noticed in data of births deaths and epidemics as result of advancement in medical science, illiteracy etc. Represent the general smooth average long terms rise or fall occurring within the data. These are changes that have occurred as a result of the general tendency of the data to term movement and are also known as secular trend.

Seasonal Component: - refers that movement in a time series which are due to forces are rhythmic in nature which repeats them periodically in every season. These variations repeat themselves in less than one year time. The seasonal variation may be attributed to the result of natural forces and social customs and traditions in which uniformly and regularly rise and fall in the magnitude.

Cyclical Component: - this is the long term oscillation about the trend. This component may vary in time, in length, in external factors-like currency change and intensity but requires over periods longer than a year. These variations in time series are due to up and downs (or trend away or drift away from the mean) recurring after a period greater than one year. These are not necessarily uniformly periodic i.e., they may or may not follow exactly similar patterns after equal interval of time and one cyclic period may normally last from 5 to 10 years.

Irregular Component: - random or irregular fluctuations do not exhibit any definite pattern and there is no regular period or time of their occurrence. These are accidental changes which are purely random, unforeseen and unpredictable components. This components normally short term variations and the research assume normally distributed but sometimes their effects is so, intense that they may give rise to new cyclical or other movements.

2.3.2.3 .Model of Time Series

Additive Model

This the model in which the four components of time series are given in the form of summation.

Symbolically: $Y_t = T_t + S_t + C_t + I_t$

Where, Y_t = Observations at time t

T_t = Trend component at time t

S_t = Seasonal component time t

C_t = Cyclical component at time t

I_t = Irregular component at time t

It is an appropriate model if the research assumes that all components are independent of one another and the magnitude of the seasonal fluctuations does not vary with the level of the series.

Multiplicative model

In this model, the time series data given in the form of the product of four components. $y_t = T_t * S_t * I_t * C_t$ where T_t , S_t , C_t and I_t are trend, seasonal, cyclical and irregular component. Y_t is the

time series value at time t .

2.3.2.4. Estimation of Seasonal Component

Seasonality: - is a pattern appears with in fixed period of time (i.e. month, years ...).The study of seasonal variation is necessary for two reasons. Those are one may be interested in forecasting some futures monthly movements and the other may be interested in isolating and eliminating the effect of trend , seasonal, irregular fluctuations so as to study the effect of cycles. These fluctuations are regular in nature and tend to repeat themselves year after year.

Seasonal index:- used to measure seasonal variations (increases or decrease) that depends on seasons customs belief etc. and seasonal index are measured by different methods among thus are simple average ,link relative, ratio to moving average, and ratio to trend methods. In this study the research used Ratio-to-MA method of to measure the index. Of all the methods of measuring seasonal variations, the ratio to moving average method is the most satisfactory, flexible and widely used method and the fluctuations of indices based on ratio to moving average method is less than based on other methods.

Ratio-to-moving average method: The method of monthly totals or monthly averages does not give any consideration to the trend which may be present in the data. The ratio-to-moving-average method is one of the simplest of the commonly used devices for measuring seasonal variation which takes the trend into consideration. This technique enables you to perform two tasks:

- Easily estimate a time series' trend and seasonal indices.
- Generate forecasts of future values of the time series.

2.4. Models for non-stationary time series

Any time series without a constant mean over time is non-stationary. Models of the form $Y_t = \mu_t + X_t$ where μ_t is a non constant mean function and X_t is a zero-mean.

2.4.1. Test of randomness

The simplest time series is a random model, in which the observations vary around a constant mean, have a constant variance, and are probabilistically independent. In other words, a random time series has not time series pattern. Observations do not trend upwards or downwards, the variance does not increase over time, the observations do not tend to be bigger in some periods than in other periods. A random model can be written as $Y_t = m + e_t$

Here m is a constant, the average of the' $t Y$ s, and $t e$ is the residual (or error) term which is assumed to have a zero mean, a constant variance, and to be probabilistically independent.

There are two situations where a random time series occur.

1. The first is when the original time series in random.
2. The second is when we fit a model to a time series to obtain an equation like:

$$Y_t = \text{fitted part} + \text{residual part}$$

The second situation is the most common. What we wish to do is to model Y_t as a fitted part plus noise, where the fitted part includes any forecast able pattern in the series, and the noise is impossible to model any further.

2.4.2. Mann-Kendall Trend Test

Given n consecutive observations of a time series $z_t; t = 1, 2 \dots n$, Mann (1945) suggested using the Kendall rank correlation of Z_t with $t; t = 1, 2 \dots n$ to test for monotonic trend. The null hypothesis of no trend assumes that the $z_t; t = 1, 2 \dots n$ are independently distributed. Our S-Plus function, Mann-Kendall (z) implements the Mann-Kendall test using Kendall(x, y) to compute T and its significance level under the null hypothesis.

Mann-Kendall test outperformed the lag one autocorrelation test for detecting a variety of deterministic trends such as a step-intervention or a linear trend. In the case of no ties in the values of $z_t; t = 1, 2 \dots n$, the Mann-Kendall rank correlation coefficient T has an interesting interpretation. In this case, the Mann-Kendall rank correlation for a trend test can be written $T = S / (nc_2)$, Where $S = 2P - nc_2$

Where P is the number of times that $z_{t_2} > z_{t_1}$ for all $t_1; t_2 = 1 \dots n$ such that $t_2 > t_1$ (Zucchini, Walter, and Oleg Nenadic 2011).

2.4.3. Differencing

Differencing is the process of changing a non-stationary time series into a stationary time series.

Regularly differencing is taking successive differences of the data. The method of taking first difference of data is simply to subtract the values of two adjacent observations on time series. If the original data has n observations ($Y_1, Y_2 \dots Y_n$), then first differenced data will be $n-1$ observations.

($X_2, X_3 \dots X_n$) where, $X_2 = Y_2 - Y_1, X_3 = Y_3 - Y_2 \dots X_n = Y_n - Y_{n-1}$

Generally, $X_t = \Delta Y_t = Y_t - Y_{t-1}, Z_t = \Delta X_t = \Delta^2 Y_t = \Delta (Y_t - Y_{t-1}) = \Delta Y_t - \Delta Y_{t-1} = (Y_t - Y_{t-1}) - (Y_{t-1} - Y_{t-2})$

Seasonal differencing is used to change a non-stationary time to a stationary time series.

2.5. Modeling

In time series Analysis models are based on statistical concepts and principles and are able to model a wide spectrum of time series behavior. There are a large class of models including Box-Jenkins to choose from and a systematic approach for identifying the correct model form. There are both statistical tests for verifying model validity and statistical measures of forecast uncertainty. Therefore, in this study identify an appropriate Box-Jenkins process or model, fitting to the data and then using the fitted model for application based on the data that comes from tourist arrival to Ethiopia data from 2006-2015.

2.5.1. The identification procedure

To apply Box-Jenkins methodology on a time series data, before any analysis, the data should be checked for stationary. A stationary series is the one that one does not contain i.e. it fluctuates around a constant mean. For non-seasonal data taking first or second differences may result in a stationary time series while for seasonal data seasonal differencing is required.

2.5.2. Akaike information criterion

The Akaike information criterion (AIC) is a measure of the relative quality of a statistical model for a given set of data. As such, AIC provides a means for model selection.

AIC is deals with the trade-off between goodness of fit and complexity of the model. It is founded on information theory: it offers a relative estimate of the information lost when a given model is used to represent the process that generates the data.

For any statistical model, the AIC value is $AIC = 2k - 2 \ln(L)$

Where k is the number of parameters in the model, and L is the maximized value of the likelihood function for the model. Given a set of candidate models for the data, the preferred model is the one with the minimum AIC value.

2.5.3. Box Jenkins Parameters Estimation

The model parameters might estimate after choosing the most appropriate model from the general class of multiplicative model is to estimate the vector of parameters

$\phi = (\phi_1, \phi_2, \phi_3 \dots \phi_p)$ and $\theta = (\theta_1, \theta_2, \theta_3 \dots \theta_q)$.

2.5.4. Diagnostic checking

After identified a tentative model the next step is to determine the adequacy of the models.

Adequate model:-

- i. The errors are random.
- ii. All parameter estimated are significantly different from zero.
- iii. The model has the smallest root mean squared error.

2.5.5. The Ljung-Box

The Ljung-Box statistic, also called the modified Box-Pierce statistic, is a function of the accumulated sample autocorrelations, r_j , up to any specified time lag m . As a function of m , it is determined as

$$Q(m) = n(n+2) \sum_{j=1}^m r_j^2 / n$$

Where n = number of usable data points after any differencing operations

m = specify time lag, j = number of observation

This statistic can be used to examine residuals from a time series model in order to see if all underlying population autocorrelations for the errors may be 0 (up to a specified point).

The null hypothesis is that the model is good, so therefore need to fail to reject null hypothesis implies good model (Patrick, 2002).

2.5.6. Plot for selected models

- Plot of Standardized Residuals: The time series plot of the standardized residuals should indicate that there's no trend in the residuals, no outliers, and in general, no changing variance across time.
- Plot of ACF of Residuals: The ACF of residuals should lie between the interval, that show the model has captured the patterns in the data quite
- P-values for the Ljung-Box-Pierce statistics for each lag up to 10. These statistics consider the accumulated residual autocorrelation from lag 1 up to and including the lag on the horizontal axis. If all p-values are above it, that's a good result (OLEG, and ZUCCHINI, 2004).

2.5. Forecasting

Forecasting may represent a prediction as to what might happen to one particular such as number of tourist arrival in next year or in five year time or it may be a prediction as to the future of a much more complex entity such as the economy.

Forecasting refers to the using of knowledge use have at one moment of time to estimate what will happen at another moment of time. Tourist arrival forecasting refers to the statistical analysis of the past and current movement in a given time series to as to obtain clues about the future pattern of the movements.

2.6.1. Single exponential smoothing forecast

The simplest of the exponentially smoothing methods is naturally called “simple exponential smoothing” (SES). This method is suitable for forecasting data with no trend or seasonal pattern. Simple exponential smoothing uses a weighted moving average with weights that decrease exponentially (Sahu, Pradeep, 2013).

$$Y'_{T+1/T} = \alpha Y_t + \alpha(1 - \alpha)Y_{t-1} + \alpha(1 - \alpha)^2 Y_{t-2} + \dots \quad \text{Where } 0 \leq \alpha \leq 1$$

2.6.2. SARIMA Forecast

Once the researcher have selected the best candidate SARIMA (p,d,q)(P,D,Q)_s model for time series data, it can estimate the parameters of that SARIMA model, and use that as a predictive model for making forecasts for future values of time series.

3. RESULT AND DISCUSSION

3.1. Explanatory Analysis

Descriptive statistics enable to determine about the general information on the tourist arrival. Descriptive statistics which deals with describing(explaining) characteristics of aggregate of statistical data by method of organizing and presenting (tables, graphs) and it provides the numeric summary of central tendency and variability like mean, minimum, maximum and measures central variations.

Table 3.1: The summaries of the number of tourist arrival in Ethiopia from 2006 to 2015

year	Number of month	mean	Std.dev	minimum	maximum
2006	12	24204.83	2327.317	19995	27770
2007	12	25995.25	5803.393	20206	37460
2008	12	27513.08	1846.178	24494	29574
2009	12	35607.17	9435.132	27308	58392
2010	12	39025.50	4955.817	31263	48173
2011	12	43619.83	5182.445	37315	53971
2012	12	49695.08	5074.405	42332	58079
2013	12	56795.75	6290.339	48217	69392
2014	12	64202.33	7616.772	48513	75116
2015	12	71978.50	11581.598	54611	88149

The above descriptive statistics shows that from all the tourist arrival to Ethiopia. The mean amount of the tourist arrived in 2015 is higher than other years. Among the ten years 2015 shows that the high variability in the amount of tourist arrival than other years, this means that the amount of tourist arrived is highly different from month to month in 2015. The minimum and maximum record of tourist arrival is 19995 and 88149 observed in the year 2006 and 2015 respectively.

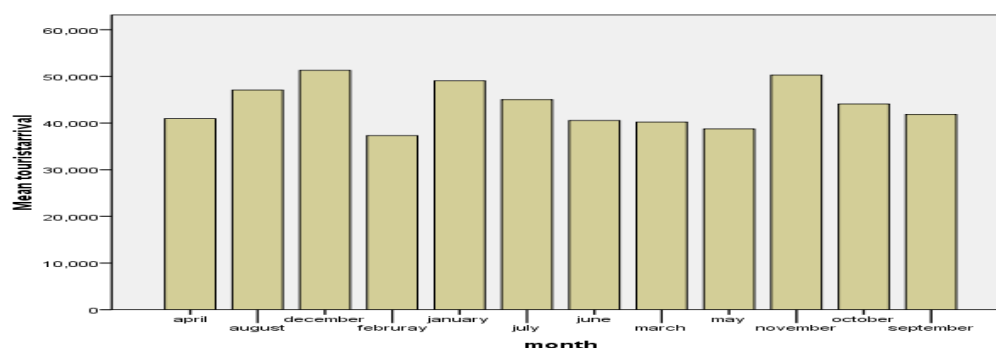


Figure 3.1 Plot of the monthly average arrival of tourists in Ethiopia from 2006 to 2015
 For above graph it is clear that high number of tourist come to Ethiopia in December, November and January whereas small number of tourist arrived to Ethiopia in February and may.



Figure 3.2 Plot of quarterly average of tourist arrival in Ethiopiain terms of corresponding percentage from 2006-2015

Generally tourist arrival was highest in the fourth quarter (October, November, and December) 48.546%, next the third quarter (July august September) follow 44.643%, then the first quarter (January, February, march) which 42.191% takes. Finally second quarter (April, may, June) 40.071% take smallest number of tourist arrival.

3.2. Time-series analysis

The first, and most important, step in any time-series analysis is to plot the observations against time. A time plot will show up important features of the series such as trend, seasonality, outliers and discontinuities. The plot is vital, both to describe the data and to help in formulating a sensible model.

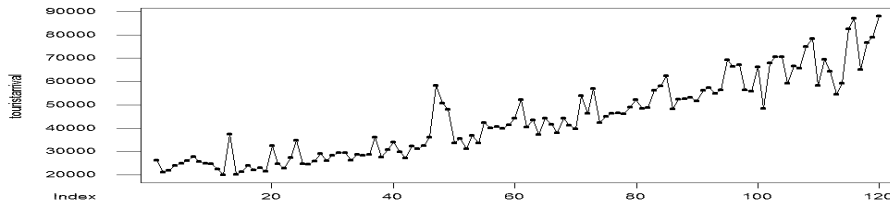


Figure 3.3 Time series plot of monthly tourist arrival to Ethiopia (2006-2015)

The above time series plot shows a clear upward trend. There may also be a slight curve in the data, because the increase in the data values seems to accelerate over time.

3.3. Trend analysis

A trend analysis is a method of analysis that allows traders to predict what will happen with a stock in the future. Trend analysis is based on historical data about the stock's performance given the overall trends of the market of tourism and particular indicators within the market.

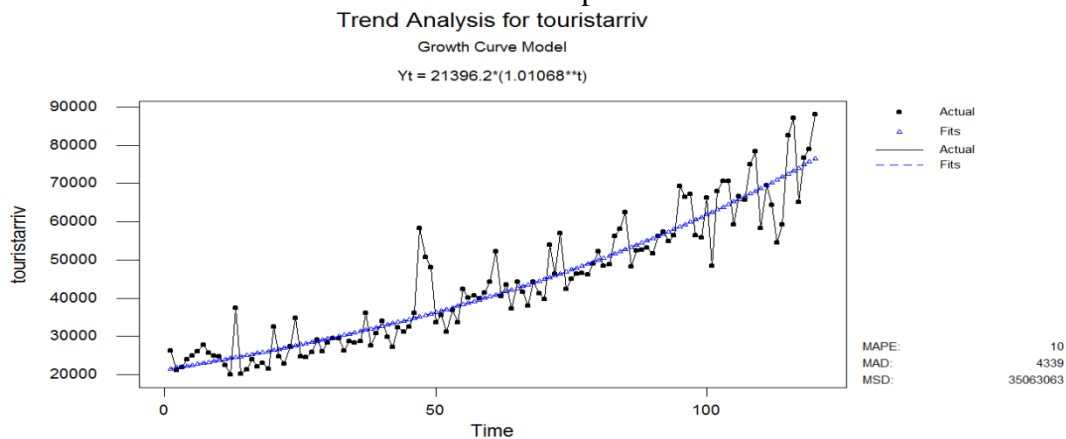


Figure 3.4 Plot of grow curve model for trend analysis of tourist arrival to Ethiopia from 2006-2015

The above figure shows that the tourist arrival data has upward trend given by equation $Y_t = 21396.2(1.01068)^t$, Where t is no of tourist arrival

This trend model type is exponential growth curve model and it's has lowest MAPE and MAD compeer to other model (linear (default), quadratic, exponential growth curve, or S-curve (Pearl-Reed logistic)).

Model type	MAPE	MAD	MSD
linear	12	5210	41795469
quadratic	10.0762	4366	34902936
Exponential growth	9.98791	4339	35063063
s-curve(Pearl-Reed logistic)	10.0041	4387	35525976

3.4.Ratio to moving average analysis

The ratio to moving average method is the most widely used method of measuring seasonal variation which takes the trend into consideration.

Table 3.2 Computation of moving average for ten year monthly tourist arrival data

Month/year	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015
January	* 3	24222. 0	27237. 8	29868. 3	38803. 8	42230. 0	46064. 3	53094. 9	60853. 1	66580.
February	* 1	24247. 0	27440. 0	30063. 5	39591. 7	42216. 9	46853. 4	53605. 4	62014. 4	67768. 0
march	* 0	24515. 9	27379. 9	30396. 5	40303. 3	42413. 9	47480. 4	54089. 4	62750. 3	68706. 8
April	* 8	24422. 4	27693. 4	30969. 8	40801. 5	42426. 1	48155. 8	54682. 3	63354. 2	69372.
may	* 9	24548. 9	27980. 0	32530. 8	40257. 1	42932. 9	48627. 1	55553. 4	63628. 7	70340.
June	* 5	25374. 2	27768. 3	34692. 8	39290. 8	43534. 5	49208. 7	56448. 5	63839. 5	71435.
July	24667.	25467.	27988.	36106.	39195.	43817.	49924.	56990.	64675.	*

	3	1	3	5	1	4	3	8	9	
August	25089. 0	25117. 6	28589. 9	36863. 1	39650. 8	44088. 5	50398. 8	57529. 7	65231. 0	*
September	25018. 4	25491. 8	28920. 0	37313. 8	40275. 9	44224. 1	50953. 5	58013. 4	65876. 9	*
October	24983. 9	25906. 6	29323. 4	37393. 0	40867. 2	44660. 6	51524. 1	58724. 1	66366. 3	*
November	24855. 0	26297. 5	29678. 0	37569. 1	41433. 7	45129. 3	52062. 5	59095. 4	66545. 4	*
December	24604. 8	26687. 3	29790. 9	38123. 1	42082. 2	45414. 6	52567. 6	59573. 5	66440. 9	*

Table 3.3 Calculation of seasonal index and percentage to moving average for each month

Percentage to moving average												
year	Januar y	Februar y	March	April	May	June	July	August	Septemb er	Octobe r	Novemb er	Decemb er
2006	*	*	*	*	*	*	112.57 8	102.69 4	99.946	98.864	90.585	81.265
2007	154.65 1	83.334	86.755	98.064	89.756	90.945	84.878	129.44 7	96.839	88.032	104.322	130.744
2008	90.994	89.264	94.821	105.48 0	93.471	101.95 8	105.66 6	102.93 2	91.041	97.830	95.872	96.570
2009	121.15 7	91.571	101.48 7	109.80 5	91.820	78.715	89.862	84.686	87.241	96.817	155.425	133.056

2010	124.14 7	85.122	88.067	76.622	91.505	85.603	108.29 2	101.22 6	101.108	97.856	100.278	105.410
2011	123.70 8	96.092	102.87 3	87.952	103.38 0	95.678	86.705	100.36 9	93.451	88.897	119.592	102.172
2012	123.70 6	90.349	95.034	96.148	95.922	93.968	98.129	103.59 6	95.020	94.701	107.999	110.484
2013	117.68 9	89.948	97.158	96.127	95.946	91.600	98.546	99.613	94.695	96.240	117.424	111.472
2014	110.37 3	91.061	89.096	104.59 4	76.244	106.46 9	109.20 6	108.37 0	89.886	100.55 4	98.806	113.057
2015	117.95 1	86.216	101.08 8	92.926	77.638	83.100	*	*	*	*	*	*
Total	1084.4	802.96	856.38	867.72	815.68	828.04	893.86	932.93	849.23	859.79	990.30	984.23
average	120.48 9	89.218	95.153	96.413	90.631	92.004	99.318	103.65 9	94.359	95.532	110.033	109.359
seasonal index	120.93 3	89.546	95.504	96.768	90.965	92.343	99.684	104.04 1	94.706	95.884	110.439	109.762

The seasonal index is highest for January (20.93%), November (10.44%), and December (9.76%) respectively and lowest for February (10.45%). This seasonal fluctuation occurred may be due weather condition in Ethiopia as well as tourist’s country, ceremony like “timket”, closed of school or break time for tourist.

3.5. Stationary time series

A *stationary* time series is one whose statistical properties such as mean, variance, autocorrelation, etc. are all constant over time. Most statistical forecasting methods are based on

the assumption that the time series can be rendered approximately stationary (i.e., "stationarized") through the use of mathematical transformations. A stationarized series is relatively easy to predict.

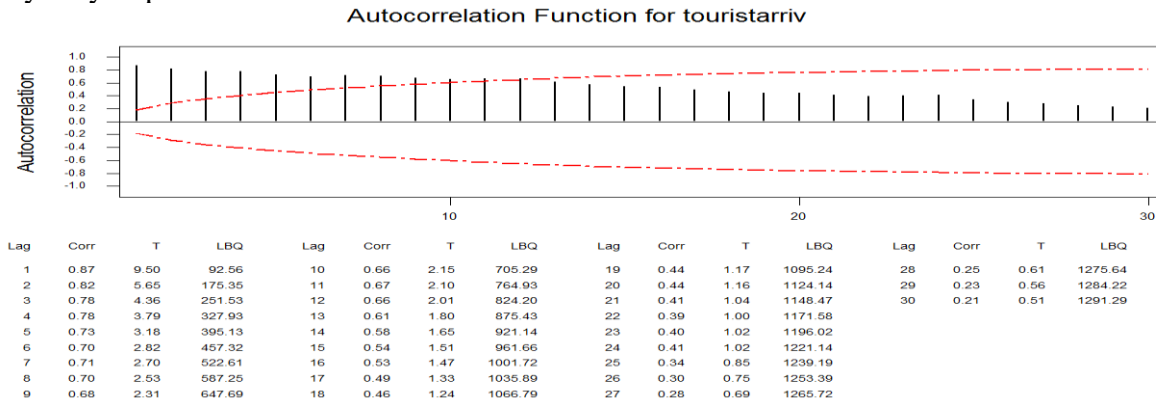


Figure 3.5 Plot of Auto correlation function for un-differenced tourist arrival data

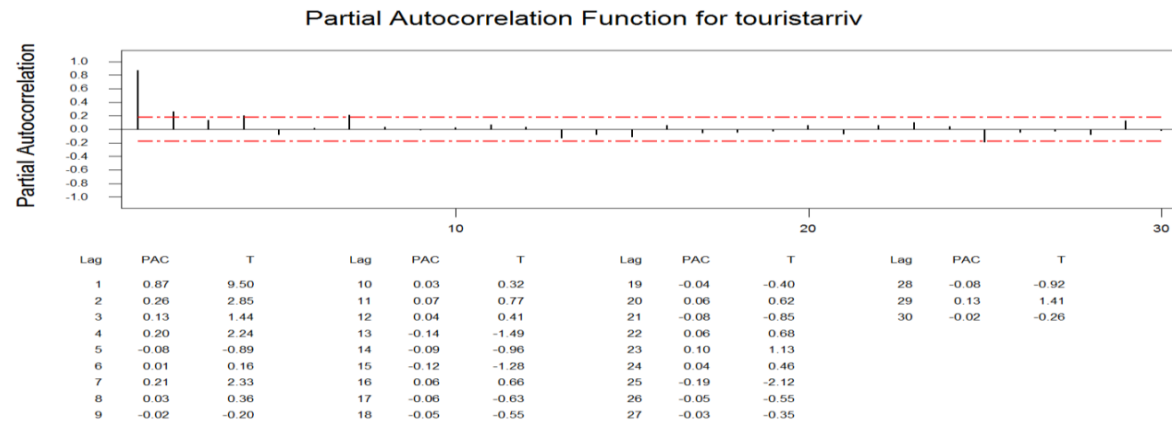


Figure 3.6 Plot of Partial autocorrelation function for un-differenced tourist arrival

Since ACF graph show very slow decay, the data is not stationary data. They are not cutting off. Each lag is quite strong. And the fact that most of them pierce the ± 1.96 standard error line is clearly proof that the series is not stationary. Since the lags in the ACF are declining very slowly, that means that terms in the series are correlated several periods in the past. Because this series is not stationary, the researcher must transform it into a stationary time series so that the it became useful to build a model.

3.6. Test for randomness

Mann-Kendall rank test

Table 3.4 Summary of Mann-Kendall Rank Test for randomness r out put

Data	Tourist arrival
Statistics	12.92
N	120
p-value	<2.2e-16

Alternative hypothesis **trend**

From the above table we reject null hypothesis which say the data has constant mean and no trend exist. Therefore the data do not have constant mean and variance; it is an indication of non-stationary of the data.

Removing Non-stationary: Differencing

The most common way to remove non-stationary is to difference the time series. The concept is the same here. Differencing a series is pretty straightforward. The researcher should subtract the first value from the second, the second value from the third, and so forth. Subtracting a period's value from its immediate subsequent period's value is called *first differencing*. The formula for a

$$Y'_t = Y_t - Y_{t-1}$$

first difference is given as:

Differencing our series, our plot of the differenced data looks like this:

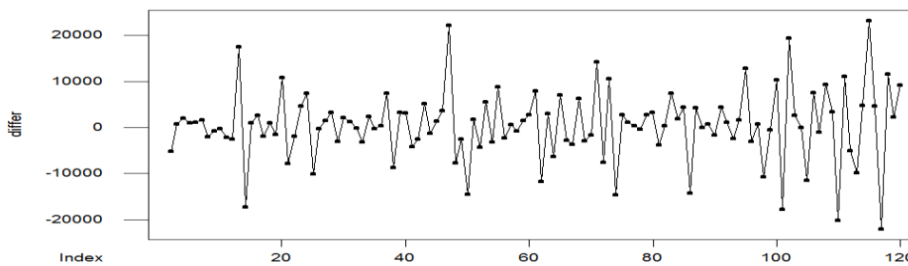
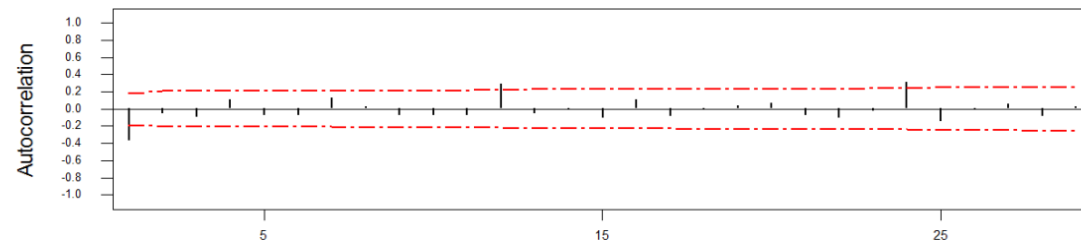


Figure 3.6 Time series plot for first differenced data tourist arrival data

As you can see, the differenced time series plot is much smoother nearly has constant mean. The ACF looks much better too: see blow ACF figure

Autocorrelation Function for differ



Lag	Corr	T	LBQ	Lag	Corr	T	LBQ	Lag	Corr	T	LBQ	Lag	Corr	T	LBQ
1	-0.37	-4.05	16.86	10	-0.08	-0.70	24.84	19	0.04	0.34	41.25	28	-0.08	-0.66	63.84
2	-0.06	-0.56	17.27	11	-0.07	-0.67	25.55	20	0.06	0.53	41.82	29	0.03	0.22	63.97
3	-0.10	-0.93	18.42	12	0.29	2.63	36.57	21	-0.07	-0.63	42.63				
4	0.10	0.99	19.76	13	-0.05	-0.45	36.93	22	-0.10	-0.87	44.19				
5	-0.08	-0.73	20.51	14	-0.01	-0.08	36.94	23	-0.03	-0.21	44.29				
6	-0.07	-0.69	21.18	15	-0.10	-0.88	38.38	24	0.31	2.63	59.15				
7	0.13	1.20	23.25	16	0.11	0.91	39.94	25	-0.14	-1.13	62.25				
8	0.03	0.28	23.37	17	-0.09	-0.74	40.98	26	-0.01	-0.06	62.26				
9	-0.07	-0.68	24.08	18	-0.02	-0.15	41.02	27	0.06	0.44	62.74				

Figure 3.7 Plot of Autocorrelation Function for first differenced tourist arrival data

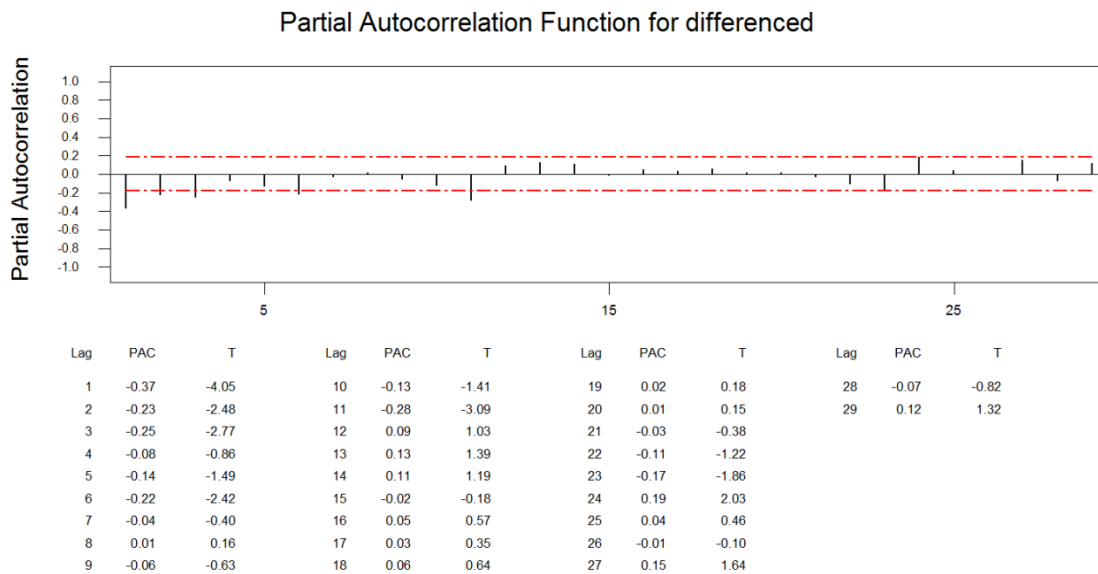


Figure 3.8 Plot of partial autocorrelation function for first differenced tourist arrival data Autoregressive processes have an exponentially declining ACF (fig 9) and spikes in the first one or more lags of the PACF (fig 10). Therefore the first difference makes the data stationary.

Table 3.5 Summary of Augmented Dickey-Fuller Test for stationary

Data	Firs differenced
Dicky-Fuller	-5.1269
Lag order	12
p-value	0.01
Alternative hypothesis	Stationary
In adf.test(dts, k = 12)	p-value smaller than printed p-value

From above r output of Augmented Dicky-Fuller test of stationary, the null hypothesis “not stationary” rejected, so first difference was enough to make the data stationary. Now it can use to build the appropriate model.

3.7. Modeling

To apply Models on a time series data, before any analysis, the data should be checked for stationary. A stationary series is the one that does not contain trend i.e. it fluctuates around a constant mean.

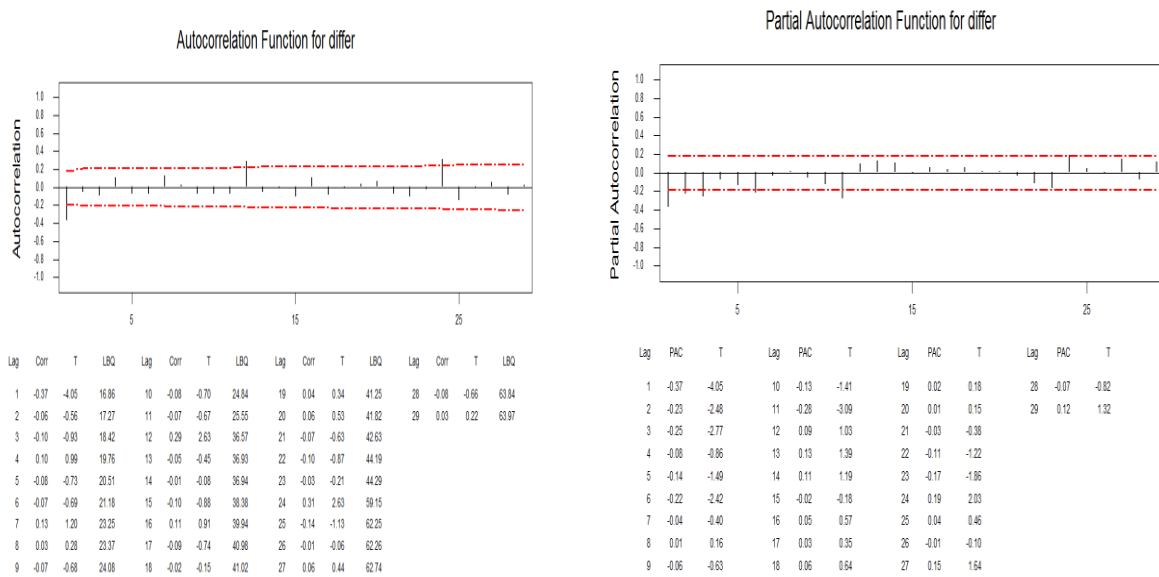


Figure 3.9 Plot of Autocorrelation function and partial autocorrelation function for first differenced tourist arrival data

The differenced series is now stationary and is shown in fig 11, the sample ACF now cuts off after lag 1 while the sample PACF tail off. The ACF lag insignificance after each 12 lag. This implies that the data has seasonal component therefore SARIMA model should be fit. Form many possible model SARIMA (1, 1,2)(0, 0, 1)₁₂ has lowest AIC value.

Table 3.6 Summary of parameters estimation for SARIMA (1, 1, 2) (0, 0, 1)₁₂

Type	Coef	SE Coef	T	P
AR 1	0.3158	0.0950	3.32	0.001
MA 1	1.0588	0.0458	23.10	0.000
MA 2	-0.1093	0.0434	-2.52	0.013
SMA 12	-0.2889	0.1031	-2.80	0.006
Constant	311.05	70.74	4.40	0.000

Differencing: 1 regular, 1 seasonal of order 12
 Differencing: 1 regular, 1 seasonal of order 12
 Number of observations: Original series 120, after differencing 107
 Residuals: SS = 4104795109 (back forecasts excluded)
 MS = 36006975
 DF = 102

SARIMA (1,1,2)(0,0,1)₁₂

$$Y_t = 311.05 + 0.6842Y_{t-1} + 0.3158Y_{t-2} + \epsilon_t - 1.0588\epsilon_{t-1} + 0.1093\epsilon_{t-2} + 0.2889\epsilon_{t-12} - 0.30589\epsilon_{t-13} + 0.03157677\epsilon_{t-14}$$

This AIC table gives for SARIMA possible model for whose box-pierce statistics is non-significance and for their corresponding coefficients p-values are significant.

Table 3.7 AIC for different candidates of SARIMA model for having significant p-values

SARIMA MODEL	AIC
(1,1,2)(0,0,1)	2439.69
(2,1,2)(0,0,1)	2441.67
(3,1,2)(0,0,1)	2443.16
(1,1,1)(0,0,1)	2440.82
(1,1,3)(0,0,1)	2440.46
(1,1,2)(1,0,1)	2442.42
(1,1,2)(0,0,2)	2442.40

The above table summarized AIC for different SARIMA model candidates who have non-significant box-pierce statistics and significant p-value for all coefficients. Form which SARIMA (1, 1, 2) (0, 0, 1)₁₂ has lowest AIC value with all estimated coefficients' are significant.

Table 3.8 Modified Box-Pierce (Ljung-Box) Chi-Square statistic for SARIMA (1, 1, 2) (0, 0, 1)₁₂ with chi-square and p-values of corresponding degree of freedom

Lag	12	24	36	48
Chi-Square	6.6	30.1	40.0	52.1
DF	7	19	31	43
P-Value	0.477	0.051	0.129	0.162

The Box-Pierce statistics are all non-significant. That means the null hypothesis which say “the model is good” will not reject. Therefore this SARIMA model is good model. Model checking by using tsdiag “R code”: which give three graphs namely “standardized residuals, ACF of residuals and p-values for ljung-Box statistic.

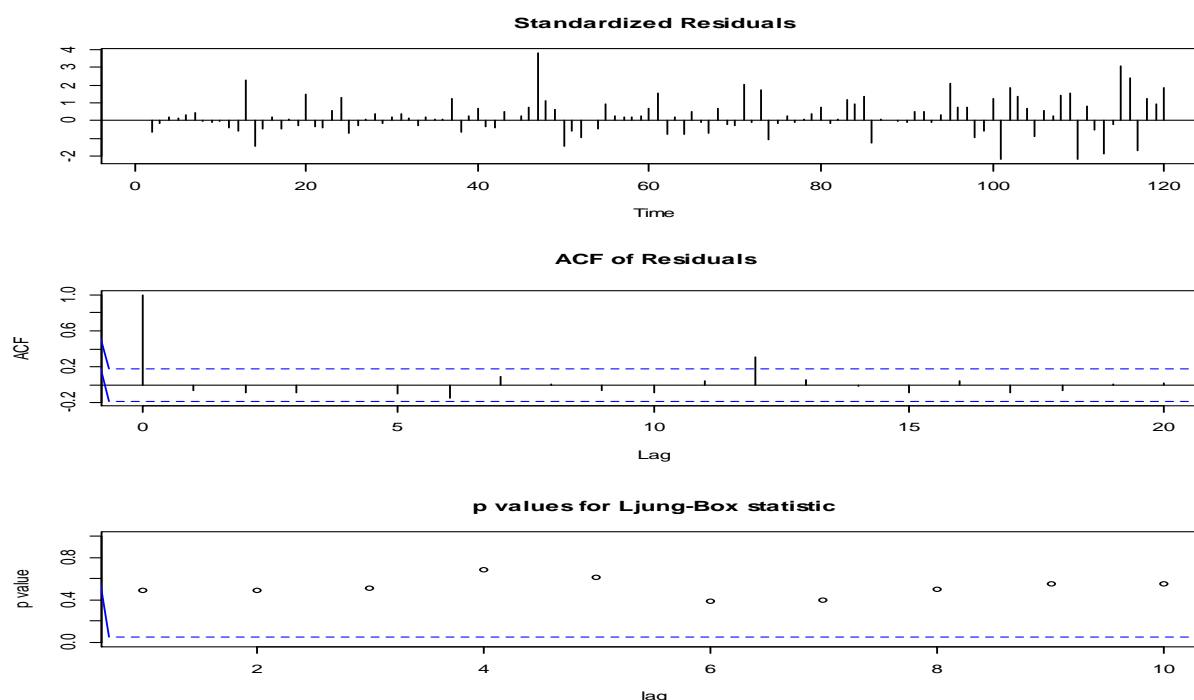


Figure 3.10 Plot of standardized Residuals, ACF of Residuals and p-values for Ljung-box statistics for diagnostic checking for SARIMA (1, 1, 2) (0, 0, 1)

The time series plot of the standardized residuals mostly indicates that there’s no trend in the residuals, no outliers, and in general, no changing variance across time.

The ACF of residuals show that the model has captured the patterns in the data quite, although there is a small amount of autocorrelation left in the residuals (seen in the significant spike in the ACF plot). This suggests that the model can be slightly improved, although it is unlikely to make much difference to the resulting forecasts.

The bottom plot gives p-values for the Ljung-Box-Pierce statistics for each lag up to 10. These statistics consider the accumulated residual autocorrelation from lag 1 up to and including the lag on the horizontal axis. The dashed blue line is at .05. All p-values are above it. That’s a good result.

3.8. Forecasting

Time series forecasting is the use of a model to predict future values based on previously observed values. Once the researcher have selected the best candidate model for time series data, you can estimate the parameters of that model, and use that as a predictive model for making forecasts for future values of your time series.

3.8.1. Single exponential smoothing

Table 3-9-Single exponential smoothing table for tourist arrival to Ethiopia from 2006-2015

Single exponential smoothing											
Janu	Febr	mar	Apri	May	Jun	Ju	Augu	Septe	Octob	Novem	Decem
ary	uary	ch	l		e	ly	st	mber	er	ber	ber

2006	245	2389	235	236	238	243	2502	251	2513	25051	24544.	23634.
	66.4	0.1	08.9	19.5	98.6	39.3	5.4	73.3	9.7	.7	4	5
2007	263	2516	243	242	238	236	2327	251	2503	24589	25158.	27105.
	99.6	0.9	82.3	95.8	43.5	90.2	5.3	23.1	5.7	.7	6	3
2008	266	2621	261	267	266	269	2749	278	2757	27796	27927.	28096.
	41.0	1.6	61.7	71.6	48.0	80.8	9.5	85.2	3.9	.6	8	1
2009	297	2927	295	304	303	297	3028	304	3088	31950	37239.	39936.
	14.5	7.4	91.7	74.6	53.4	44.4	4.7	71.3	7.7	.7	0	2
2010	415	4000	391	375	373	366	3780	382	3876	39007	39515.	40484.
	83.6	7.0	04.4	36.1	96.5	44.0	4.2	70.8	1.0	.0	4	1
2011	428	4238	426	415	421	420	4122	418	4173	41325	43854.	44363.
	35.9	2.1	32.1	68.7	31.5	35.8	7.1	31.9	1.1	.3	4	7
2012	468	4597	458	459	460	460	4667	477	4790	48083	49712.	51385.
	87.8	6.6	05.9	04.7	52.8	90.2	0.2	78.3	5.9	.5	2	6
2013	536	5252	525	525	526	524	5322	540	5422	54681	57623.	59380.
	05.6	7.9	32.7	39.2	91.5	94.6	8.1	43.9	2.3	.0	2	2
2014	609	6004	592	606	582	601	6225	639	6299	63742	64144.	66338.
	37.4	4.1	16.9	26.5	03.8	56.8	1.5	39.4	4.3	.2	0	4
2015	687	6670	672	666	642	632	6715	711	6997	71327	72860.	75918.
	77.1	7.1	56.5	98.2	80.7	97.2	3.0	73.6	4.7	.3	3	0

Single Exponential Smoothing

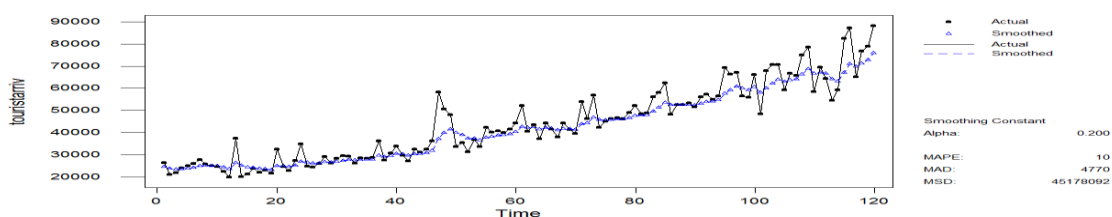


Figure 3.11 Plot of single exponential smoothing graph with smoothed line for tourist arrival to Ethiopia from 2006-2015

Single exponential smoothing is use because it has minimums MAPE, MAD and MSD value compare to other methods. Since the above graph show that there is upward trend, it cannot be used for forecast future value. (Sahu, Pradeep, 2013)

Double exponential smoothing has greater MAPE, MAD and MSD values compare to single exponential, and also it cannot use to forecast future value Because double exponential smoothing graph show upward trend (see blow fig) (Sahu, Pradeep, 2013).

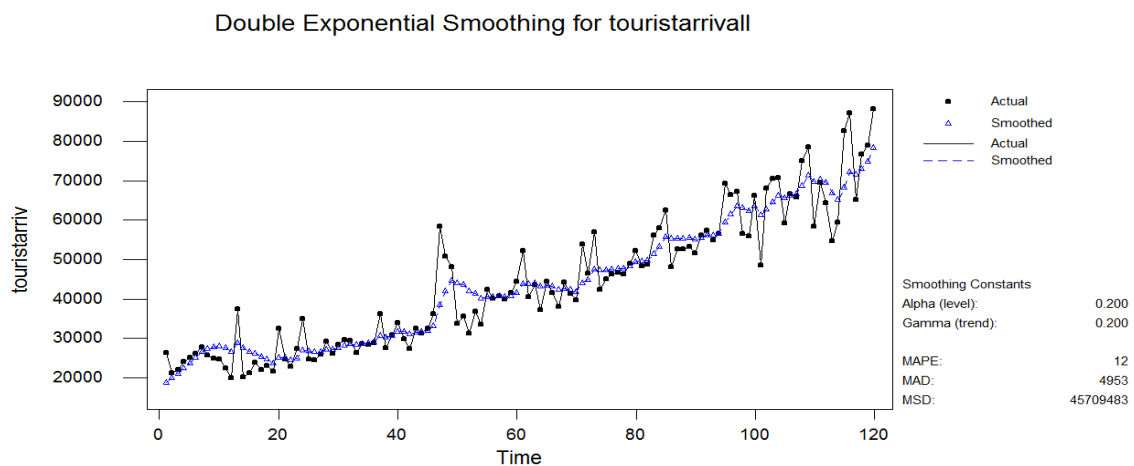


Figure 3.12 Plot of double exponential smoothing graph with smoothed line tourist arrival to Ethiopia from 2006-2015

3.8.2. SARIMA forecasting

All above works were to use the data for forecasting future tourist arrival. Since the best model was SARIMA (1, 1, 2)(0,0,1)₁₂ it is possible to use this model for forecast.

FORECASTING FOR YEAR 2016

Table 3.10 One year ahead or 12 point head SARIMA forecasting for year 2016 of tourist arrival to Ethiopia

Forecasts from period 120				
95 Percent Limits				
Period	Forecast	Lower	Upper	Actual
121	79725.0	67961.5	91488.5	
122	73098.0	60952.0	85244.0	
123	76204.2	63959.7	88448.6	
124	74053.2	61760.8	86345.6	
125	73009.7	60681.6	85337.8	
126	73053.6	60693.3	85413.9	
127	79961.6	67570.3	92352.9	
128	81678.6	69256.7	94100.5	
129	76573.8	64121.4	64121.4	
130	79671.3	67188.5	92154.0	
131	81047.5	68534.6	93560.5	
132	83161.2	70618.0	95704.3	

Now, we turn to the application (i.e., forecasting). Our objective is to predict the 12 future values of time series (monthly forecasts for time series).the table shows monthly forecasted results with confidence. November, January, and December are the months with the most prominent values, thus expressing the extension of strong seasonal movement in the number of tourist arrivals in Ethiopia. Next year tourist arrival expected to be 931238, which is 7.24% increment from 2015

total number of tourist arrival.

Forecasting for after two year (2017) are given for each month tourist arrival within the 95 percent confidence limit.

Table 3-11: 12 point a head SARIMA forecasting for year 2017 of tourist arrival to Ethiopia

95 percent limits			
Period	Forecast	Lower	Upper
133	80798.3	67548.8	94047.8
134	80731.4	67367.8	94095.1
135	81021.4	67592.9	94449.8
136	81424.0	67943.5	94904.4
137	81862.2	68333.6	95390.8
138	82311.7	68736.2	95887.1
139	82764.7	69142.8	96386.5
140	83218.8	69550.9	96886.7
141	83673.3	69959.5	97387.1
142	84127.9	70368.3	97887.4
143	84582.5	70777.4	98387.6
144	85037.1	71186.6	98887.7

The SARIAM (1,1,2)(0,0,1)₁₂ forecasting show that in 2017 tourist arrival to Ethiopia may increase that is expect to be 991554 tourist may come to Ethiopia. This is 6.1 % increment than previous year tourist arrival 2016 and 12.9 % increment form 2015.

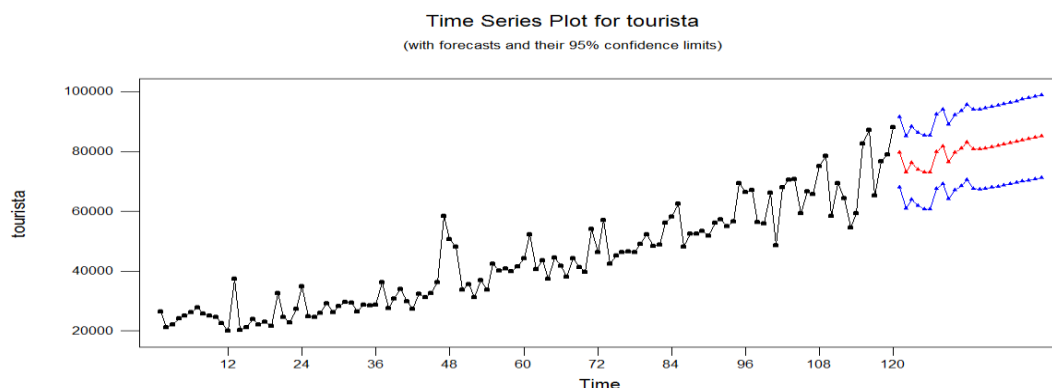


Figure 3.13 Time series plot for two a year forecast with upper and lower bound for years of 2016 and 2017

4. DISCUSSION

The present research investigated there is seasonal variation in tourist flow to Ethiopia from 2006 to 2015. The tourist flow depends on season worldwide, The fluctuations of visitor and tourist numbers to a region can be influenced by nature or by institutional interventions. Regardless of these, the impact of seasonality can be perceived as positive or negative depending on the

perspective taken by the stakeholders. (Lee, Christine, et al. 2008).

In Ethiopia tourist arrival was highest in the fourth quarter (October, November, and December) 48.546%, next the third quarter (July August September) follow 44.643%, then the first quarter (January, February, March) which 42.191% takes. Finally second quarter (April, May, June) 40.071% take smallest number of tourist arrival between the years from 2007-2016. But there were no research conducted on the same years.

In Ethiopia year 2016 tourist arrival expected to be 931238, which is 7.24% increment from 2015 total number of tourist arrival and in 2017 the 6.1 % increment than previous year tourist arrival (2016) and 12.9 % increment from 2015 whereas a 3.6% increase in tourism is expected for 2017 as scheduled flights to the island have shown according to Hermes Airports. The increase of arrivals from some destinations is significant with an expected 8% increase from the UK, 70% from Germany and 18% from Israel. Arrivals from Russia are expected to remain on par with 2016 while a slight dip is expected in arrivals from Greece (Cyprus, 2017).

5. CONCLUSION

Based on the analysis conducted in this research on the tourist arrival, the following ideas are extracted and summarized. The mean amount of the tourist arrived in 2015 is higher than other years. Among the ten years 2015 shows that the high variability in the amount of tourist arrival than other years, this means that the amount of tourist arrived is highly different from month to month in 2015. The minimum and maximum record of tourist arrival is 19995 and 88149 observed in the year 2006 and 2015 respectively.

As we have seen from the original data, there is arrival fluctuation from month to month (not stationary). But, by differencing the arrival data once the data was became insignificant from month to month (stationary). Thus, the mean tourist arrival is not significantly different from month to month. Generally the tourist arrival highly increases in 2015 G.C. So, anyone can use this series to identify the reason that make it this much different. The model that identified based on Box-Jenkins procedure is SARIMA (1, 1, 2)(0, 0, 1). This SARIMA model selected from different candidate with criteria lowest AIC value and insignificant box-pierce statistics. And the concerned body can use this model to forecast the future tourist arrival.

In this paper, the research considered one forecasting model in order to determine the size of the flows of tourism demand in Ethiopia according to the number of arrivals. Theoretical framework that the research was used the Box-Jenkins methodology for seasonal ARIMA models. After constructing the appropriate models, it utilized them to generate the forecasts of tourist arrival. The obtained results (i.e., forecasted values) can provide important information needed for an adequate destination. The SARIMA forecast showed that future tourist arrivals will be high and stable. November, January, and December are the months with the most prominent values for next year. Tourist arrival expected to be 931238, which is 7.24% increment from 2015 total number of tourist arrival and in 2017 tourist arrival to Ethiopia may increase that is expected to be 991554 tourist may come to Ethiopia. This is 6.1 % increment than previous year tourist arrival 2016 and 12.9 % increment from 2015.

REFERENCE

1. Asmare, BirtukanAtinkut. "Pitfalls of tourism development in Ethiopia: the case of Bahir Dar town and its surroundings." *Korean Social Science Journal* 43.1 (2016): 15-28.
2. AnandBethapudi. "Role of social media in marketing tourism product". *Zenon Academic Publishing* 1.1 (2015): 262-268.
3. Burns, Patrick. "Robustness of the Ljung-Box test and its rank equivalent." (2002).
4. Daniel Beyene, Ethiopia: Harnessing the Tourism Potential, *allafrica* (2016)
5. Dwyer, Larry. *International handbook on the economics of tourism*. Edward Elgar Publishing, 2007
6. HenokMitiku: *The Role of Private Sector in the Development of the Tourism Industry: The Case of Privately Owned Hotels in Addis Ababa.*(2015)
7. Kumma, Tadesse, et al. "The Ethiopian development research institute working papers." (2012).
8. Ministry of Culture and Tourism *Tourism Statistics Bulletin*, Addis Ababa, Ethiopia (2009).
9. NENADI, OLEG, and WALTER ZUCCHINI. "Statistical Analysis with R." (2004).
10. Sahu, Pradeep. "Determination of Exponential Smoothing Constant For Demand Forecasting of Milk Product (Paneer) in Chhattisgarh, India." *i-Manager's Journal on Mechanical Engineering* 3.4 (2013): 33.
11. Theobald, William F., ed. *Global tourism*. Routledge, 2005.
12. World Bank "Ethiopia: Towards a Strategy for Pro-Poor Tourism Development", (2006).
13. YabibalMulualemWalle: *Tourist Flows and Its Determinants in Ethiopia.*(2010)
- 14 Zucchini, Walter, and Oleg Nenadic. "Time series analysis with R-Part I." *Document de cours* (2011).