Forecasting International Tourists Footfalls in India: An Assortment of Competing Models

Purna Chandra Padhan
Assistant Professor (Economics)
XLRI, School of Business and Human Resources
P. Box No. 222, C.H. Area (East), Jamshedpur-831035, Jharkhand, India
Tel: 91-657-398-3333(O) E-mail: pcpadhan@yahoo.com, pcpadhan@xlri.ac.in

Received: November 11, 2010 Accepted: December 8, 2010 doi:10.5539/ijbm.v6n5p190

Abstract

The paper deals with forecasting international tourists footfalls in India, applying an assortment of uni-variate time series forecasting models, for monthly data spreading over Dec 1990 to Jan 2010. The forecasting performance of various competing models is evaluated with MAPE and other criteria. Also the actual and forecasted values are compared since Feb 2010 to Sept 2010 for better evaluation. The SARIMA model performs better than other competing model for forecasting, with lowest MAPE value. In fact it has an advantage over other models as it explicates autoregressive and moving average process not only for the data series but also of seasonality. As a policy implication, SARIMA model can be used for forecasting tourists demand in India.

Keywords: Tourism, Forecasting models, Smoothing, Seasonality, SARIMA model, Model evaluation

1. Introduction

Modeling and forecasting tourists demand has received substantial attention among policy makers, hospitality management, researchers and other interest groups globally (Note 1). Tourist's footfalls immensely contribute towards the growth of economy's Gross Domestic Product (GDP). It is one of the leading sources of foreign exchange earning as well as generating employments opportunities. Despite the continued existence of number of antisocial activities and economic set backs such as terrorism, naxalite issues, inadequate infrastructure facility and recession etc, India is still a wonderful and attractive destination for international tourists. Although international tourist footfalls depend upon numerous factors such as growth of the economy, internal security, attractive tourists places, warm hospitality, infrastructure facility etc, the precedent trend demonstrate that international tourist's arrivals, at various destinations (via airways, railways, road ways, ships) have exposed exponential growth. Forecasting tourist arrival being a significant activity for its beneficiaries and stack holders', several forecasting models have been applied to estimate and forecast the tourism demand globally.

Large numbers of research papers have applied widespread time series models for forecasting tourism demand globally (Note 2). Highly structured and an extensive survey of literature of earlier studies is provided by Crouch (1994), Li et al (2005), Witt and Witt (1995). On the other hand Song and Li (2008) review the literature for post 2000 studies. This paper highlights few recent studies which applied time series model for forecasting tourism demand. For example, Song and Turner (2000) found that majority of published article have applied quantitative methods such as uni-variate, multivariate or causal forecasting methods for forecasting tourism demand. The time series methods starting from simple naïve methods to the advanced models like artificial neural network to fuzzy goal programming have been applied extensively for forecasting. The studies of Smeral and Wuger (2005) applied ARIMA, SARIMA and naïve methods for forecasting tourism demand. The result reveals that advanced models like ARIMA or SARIMA model could not even outperform the simple Naïve model. Similarly, incorporating intervention function in the multivariate SARIMA (MARIMA) model to capture the tourism demand forecast Goh and Law (2002) found multivariate SARIMA model significantly improved the forecasting performance over simple ARIMA as well as other models. However, Gustavsson and Nordstrom (2001) found opposite results to Goh and Law (2002)'s studies. Applying ARCH and GARCH model Chan et al (2005) tries to estimate and forecast volatility in tourism demand and its affect to various shocks. While Li et al (2005) and Lim

(1999) applied several econometric models, Song and Witt (2000) applied Autoregressive Distributed Lag Model (ADLM) model to forecast tourism demand. Apart from this, the error correction model is applied by Kulendran and Wilson (2000), Lim and Mc Vleer (2001), Song and Witt (2001). Advanced econometric model like VAR model is applied by Shah and Wilson (2004), Witt et al (2004). The studies of Wong et al (2006) applied Bayesian VAR Models, whereas Deaton and Muellbauer (1980) applied almost ideal demand system (AIDS) for forecasting tourism demand. Naude and Saayman (2005) and Roget and Gonzalez (2006) applied the panel data analysis. On the other hand, Turner and Witt (2001a) have applied the structural equation modeling. Cho (2003) concluded that artificial Neural Network (ANN) model outperform the exponential smoothing and ARIMA model in modeling and forecasting the tourism demand for Hong Kong. While Wang (2004) applied the fuzzy goal programming, Hernandez-Lopez (2004) applied genetic algorithms for forecasting tourism demand.

The brief survey witnessed with various issues of forecasting tourism demand such as, methodological development, forecast competition, tourism cycles, seasonality analysis etc. The survey also exposes that quite a few variables such as tourists expenditure, sightseeing expenditure, shopping, meal expenditure etc could be use as a proxy for forecasting tourism demand. Among them tourists arrival is the most popular measure. In lieu of this, the paper applies total international tourists arrivals, jointly together by air, road, railways, and ships from all destinations in India for forecasting tourism demand.

2. Trends of Tourism Industry of India

India is a country with unique feature. It is rich with cultural heritage and traditional diversity and perhaps one of the ancient cultural cradles of the world which mesmerizes foreign tourists from every corner of the world. The beauty of India lies with diversification in terms of language, religion, culture, traditions, belief and a lot more; gifted with ancient cities, temples, monuments, mosques, garden, lakes, mountains, Hill stations, and what not. The places like Khajuraho, Konark, Puri, Nalanda, Delhi, Mumbai, Darjeeling and many more are still the paradise of foreign tourist. The warm hospitality of Indian, with the words of 'Swagatam (i.e. welcome)', 'Namaskar (i.e. Greetings)' etc. touched the heart of foreign tourists since time immemorial.

The tourism industry in India is considerable and effervescent. The trend divulges fabulous and tremendous growth over the years. According to Economic Survey of India (2009-10), the number of foreign tourist's arrivals in the year 2006-07 is 46.7 lakh, which is 51.7 lakh during 2007-08, is 50.7 lakh during 2008-09, and 2009-10(April-Dec) it is 37.2 lakh. The year 2006 witnessed a stupendous growth in Indian tourism industry. In terms of percentage change of foreign tourists arrivals in India, it has increased by 6.7% from 1999 to 2000, in the year 2003, 2004….2007 it is 14.3%, 26.8%, 13.3%, 13.5%, 14.3% respectively (Sources: Ministry of Tourism GOI Annual Report 2008-09). However, in the year 2008 the rate has been dropped to 5.00%.

Although global financial crisis, along with continuous rise in terrorism, maoists violence, Kashmir issues etc. affected the growth of tourism during 2008-09, still it is one the most vibrant industry. In recent years tourism industry demonstrates exponential growth due to rise in arrivals of more and more foreign tourists from Africa, Latin America, Europe, Australia-New Zealand, South East Asia, USA and other parts of the globe. The increase in disposable income of the people, boom in IT sector, aggressive advertisement campaign, such as 'Incredible India' on various tourist destinations and rapid growth of the Indian economy are responsible for such growth.

Tourism has a significant role in the development of an economy, as it fetches foreign exchange reserve, generates employment opportunities and enhances growth in other sectors of the economy such as horticulture, handicrafts, agriculture, construction and many more. The foreign exchange earning from tourism has shown tremendous increase over the years. The foreign exchange earning during 2006-07 was 9, 123 million US dollar which has increased to 10, 543 million US dollar during 2008-09. In 2009-10 (April-Dec) it was 8,663 million US dollar. The growing tourism has created additional employment opportunity and enhanced the contributions of tourism towards GDP. In the year 2002-2003 the total contributions of tourism to GDP goes up to 5.83% which has increased to 5.90 % during 2003-04, and 7.7 % during 2008-09. The share of tourism to total jobs is 8.27% which has increased to 8.78% during 2003-04. In the year 2002-03, the total employment in the tourism sector was 21.54 million and additional 3.2 million jobs were created in 2003-04. In 2006 it was 41.8 Million.

The year 2008-09 has been a challenging year for the Indian tourism and response of international tourist arrivals continue to detoriate due to the impact of global economic crisis and aftermath of terror strike at Taj and other places. Nevertheless the tourism sector outperformed the global growth of 2 percent in international tourist arrivals in 2008. Initiatives have been taken by the government of India to develop world class tourism infrastructure through central financial assistance to state and union territories periodically.

3. Data

The empirical analysis focuses on the estimation and forecasting tourism demand in India using uni-variate time series data on tourists' arrivals. The monthly data is used spreading across Dec 2009 to January 2010 for forecasting and February 2010 to September 2010 for comparison and model evaluation. The data is secondary in nature and collected from Center for Monitoring Indian Economy (CMIE), Mumbai, India. As tourists arrivals is one of the most widely used variable for measuring demand for tourism we have used total tourists arrivals from all parts if the globe, through all possible routes (i.e. airways, railways, roadways and ships), at various destination in India. For empirical analysis we have applied Excel Spreadsheet, Eviews5.0 and Minitab 15.0 software.

4. Empirical methodology

Modeling, estimation and forecasting of tourist demand can be done applying various uni-variate and multi-variate time series models. This study applies various uni-variate time series forecasting models to estimate and forecast the tourist demand. The appropriate model is then selected based on appropriate model selection criteria, which fits the data well and forecast correctly. Then recursive forecasting technique is used to generate forecast beyond Jan 2010. Hence forth Y_t denotes the actual time series data on tourists arrival, where Y_t implies actual tourists arrival at time period t, $\hat{Y}_{t+k}$ is the k period ahead forecasting values. The various models which are applied here may be discussed as follows,

4.1 Naive Models

The simplest model for forecasting tourists demand is ***the naïve model or random walk model***, where the current value will be the future value. Mathematically the model can be expressed as,

$$\hat{Y}_{t+1} = Y_t - - - - - (1) .$$

It means best forecast of next months' value is this month value. This model is suitable when the data follows random process. Extensive data doesn't require applying this method. Minimum data point requirement is one, which is assign 100 % weightage for forecasting. However, when the random data series has some upward or down trending and therefore generally non-stationary, the ***naïve trend model*** is more suitable for forecasting. The model is expressed as,

$$\hat{Y}_{t+1} = Y_t + (Y_t - Y_{t-1}) - - - - - (2)$$

On the other hand, when the data randomly fluctuates but have some seasonal pattern, the appropriate model could be ***naive seasonality model***, which is defined as,

$$\hat{Y}_{t+1} = Y_{t-k+1} - - - - - (3)$$

Where k is the frequency of seasonal pattern e.g., k=12 for monthly, k=4 for quarterly and so on. Similarly, ***the naïve trend and seasonality model*** is suitable when the data has seasonal fluctuations with upward or downward trend, either linearly or non-linearly. The model could be expressed as,

$$\hat{Y}_{t+1} = Y_{t-k+1} + \frac{Y_t - Y_{t-k}}{k} - - - - - (4)$$

Where, k=12 for monthly and 4 for quarterly data. Suitable Naïve model may be more useful under sound theoretical judgment. Such models are often used for comparing other models.

4.2 Models Based on Average

The most frequently and commonly used methods in the past for forecasting next period is the ***simple average***. The model may be defined as,

$$\hat{Y}_{t+1} = \frac{1}{t}\sum_{i=1}^{t} Y_i - - - - - - - (5)$$

This model is more appropriate if the time series has stability or stationary and the environment of the data is unchanging. Here all the data gets equal weightage. Therefore the main limitation of this model is that extreme value or outliers can influence the forecasted values. In the simple average method, all the past observations receive equal weightage. However in the ***moving average model*** the recent observations get more weightage. The model can be expressed as,

$$\hat{Y}_{t+1} = \frac{Y_t + Y_{t-1} + ... + Y_{t-k+1}}{k} -----(6)$$

Where, k is the number of moving average process. Moving average is computed by adding the newest and dropping the oldest observations recursively. Therefore it doesn't require all the observation for calculating forecasted values of the moving average. Only the recent data are relevant. Here we have applied two moving average process namely, 3 month and 12 month moving average. All values get equal weightage in the process in this method. When the data has either trend or seasonality, this method may not produce consistent and suitable result.

When the data has linear trend, one way of forecasting is ***double moving average.*** Double moving average is calculated as the moving average of the first moving average i.e. when second set of moving average is calculated from the first set of moving average. The model may be expressed as,

$$M_t = \hat{Y}_{t+1} = \frac{Y_t + Y_{t-1} + ... + Y_{t-k+1}}{k} -----(7)$$

$$M_t' = \hat{Y}_{t+1} = \frac{M_t + M_{t-1} + ... + M_{t-k+1}}{k} -----(8)$$

The forecast is done adding the difference between the first and second moving average to the first moving average, as follows

$$\alpha_t = M_t + (M_t - M_t')$$
$$= 2M_t - M_t'$$

Adding to it is the adjusting factor as

$$\beta_t = 2/(k-1)(M_t - M_t')$$

The forecasted value for P period ahead will be

$$\hat{Y}_{t+p} = \alpha_t + \beta_t P ----(9)$$

4.3 Exponential Smoothing Models

Basically three different smoothing models are widely applied in the literature (Note 3). Those are namely simple exponential smoothing model, Holt's linear exponential smoothing model, Holt's-Winter's multiplicative seasonal model.

The essence of the ***simple exponential smoothing model*** is that the weights assign to each observation exponentially decreases over time. Here the more recent observations get higher weightage than the distance past. The model is more suitable when the data does not have any predictable upward or downward trend. It is a method which continuously renews the forecast in light of more recent experience. The model can be expressed as

$$\hat{Y}_{t+1} = \alpha Y_t + (1-\alpha)\hat{Y}_t -----(10)$$

Where, α =0< α<1 is the smoothing parameter, Y^t= is the initial level or forecasted value. The value of α is chosen to produces lowest MAPE which furnish the best model. When the predictor is to be stable and random, a smaller α value will be appropriate. But if a rapid change is required then larger value of α will be appropriate.

The ***Holt's linear exponential smoothing method*** is appropriate for forecasting the time series variable when the data is non-seasonal but revolves around local linear trend. To forecast future values initial estimate of a current level and slope of trend is required. The simple choice of initial estimates of level $L_1=Y_1$ and $T_1=0$. This model is more flexible as it allows level and trend to be flexible. The model is expressed as,

$$L_t = \alpha Y_t + (1-\alpha)(L_{t-1} + T_{t-1}) -----(11)$$

$$T_t = \beta(L_t - L_{t-1}) + (1-\beta)T_{t-1} -----(12)$$

Forecasting is done using

$$\hat{Y}_{t+p} = L_t + pT_t ------(13)$$

Where, L_t is New smoothed value (estimate of current level), α is Smoothing constant for the level (0 < α < 1), Y_t is New observation or actual value of series in period t, β is Smoothing constant for trend estimate (0< β <1), T_t is Trend estimate, p is periods to be forecast into the future. The best alternative combinations of the two smoothing constant, α, β will be that which produces lowest error (MAPE) term.

The ***Winter's multiplicative seasonality method*** is more appropriate when the data has local linear trend as well as seasonality. The method requires an estimation of the initial smoothed values, a trend estimate and seasonality applying various procedures. We have applied the Minitab 15.0 software to estimate these values. These can be estimated using the following equations

Exponentially smoothed series or level estimate

$$L_t = \alpha \frac{Y_t}{S_{t-s}} + (1-\alpha)(L_{t-1} + T_{t-1}) -----(14)$$

Trend Estimate

$$T_t = \beta(L_t - L_{t-1}) + (1-\beta)(T_{t-1}) -----(15)$$

Seasonality Estimates

$$S_t = \gamma \frac{Y_t}{L_t} + (1-\gamma)S_{t-s} -----(16)$$

Forecasted values can be obtained by the following equation

$$\hat{Y}_{t+p} = (L_t + pT_t)S_{t-s+p} -----(17)$$

Where, L_t is new smoothed value or current level estimate, α is smoothing constant for level, Y_t is new observation or actual value in period t, β is smoothing constant for trend estimate, T_t is trend estimate, ϒ is smoothing constant for seasonality estimate, S is seasonal estimate, P is number of forecasting period ahead, s is length of seasonality, $\hat{Y}_{t+p}$= forecast for p periods into future. Here the best combinations of α, β, ϒ is one which minimizes MAPE.

4.4 Seasonal Decomposition Model

A time series data could follow any of these four components namely; Trend (T) Seasonal(S), Cyclical(C) and Random (I). ***Seasonal decomposition*** is a procedure to decompose the entire series into these four different parts. Each component is estimated separately from the series and forecasted separately. The final forecasting is made by combining all of these four parts. A model that treat the time series values as the product of the components is called as multiple decomposition model which is expressed as

$$Y_t = T_t \times S_t \times C_t \times I_t ---------------------(18)$$

$$\hat{Y}_{t+1} = \hat{T}_{t+1} \times \hat{S}_{t+1} \times \hat{C}_{t+1} \times \hat{I}_{t+1} ------(19)$$

The multiplicative model is used when variability of time series being analyzed is increasing.

4.5 Regression Models

Since the data shows some kind of non-linear quadratic trend we have tried to fit a ***quadratic trend model*** not the ***simple linear regression model***. Quadratic model is a regression model with time is the independent variable and concerned series to be forecasted is the dependent variable. When the data shows quadratic behavior against time, such model is more appropriate. The coefficient of the model is estimated using OLS methods. The estimated model can be expressed as

$$\hat{Y}_t = \hat{a} + \hat{b}_1 T + \hat{b}_2 T^2 -----(20)$$

Forecasted value will be

$$\hat{Y}_{t+1} = \hat{a} + \hat{b}_1 T_{t+1} + \hat{b}_2 T_{t+1}{}^2 -----(21)$$

On the other hand, ***autoregressive moving average model*** is a class of model which is capable of representing stationary as well as non-stationary data. The autoregressive integrated moving average process, popularly know as Box-Jenkins (1970) methodology or ARIMA process, can be expressed as a combination of autoregressive and moving average process with integration of order d. The AR (P) process can be expressed as,

$$Y_t = \phi_0 + \phi_1 Y_{t-1} + \ldots + \phi_p Y_{t-p} + \varepsilon_t \quad ------(22)$$

$$t = 1,\ldots, T,$$

$$\varepsilon_t \sim NID(0, \sigma^2)$$

If Y is a moving average process of order q, then MA(q), process can be written as

$$Y_t = \gamma + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \ldots + \theta_q \varepsilon_{t-q} -----(23)$$

The combination of the two i.e. ARMA (p, q) process can be expressed as,

$$Y_t = \alpha + \phi_1 Y_{t-1} + \ldots + \phi_p Y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \ldots + \theta_q \varepsilon_{t-q} -----(24)$$

The estimation of ARIMA model involves a set up steps as discussed by Box-Jenkins (1985). First it needs identification of the model i.e. whether the series is stationary or not. As ARMA model is generally applied for stationary series, the stationary of the series can be expressed as d. If the series Y_t is stationary at level then it is called as I(0) or integrated of order 0, if not then possibly I(1). Therefore when d=0, it indicates stationary at level, and ARIMA model will be called as ARMA model. The stationarity of the series can be tested either by autocorrelation functions (ACF) or unit root tests. Model for non stationary series are called as ARIMA (p,d,q) model, whereas for stationary series they are called as ARMA(p, q) models. The maximum order of p, q can be selected based on SBC criteria. Once the order of p, q and d is selected then the parameters of the ARIMA model is estimated using non-linear least squares methods. Diagnostic checking of the model is done through $\chi 2$ and Ljung-Box Q statistics. Finally, obtain dynamic forecasts of Y from the estimated ARIMA model. Dynamic forecasts can be found recursively by,

$$Y_{t+1} = \phi_1 Y_t + \phi_2 Y_{t-1} + \varepsilon_{t+1} + \theta_1 \varepsilon_t + \theta_2 \varepsilon_{t-1} + \delta -----(25)$$

Our one period ahead forecast is the expectation of Y_{t+1} conditional upon the past history of Y, and can be obtained as

$$\hat{Y}_{t+1} = \hat{\phi}_1 Y_t + \hat{\phi}_2 Y_{t-1} + \hat{\theta}_1 t + \hat{\theta}_2 \hat{\varepsilon}_{t-1} + \hat{\delta} -----(26)$$

The ***SARIMA*** model is the extension of the ARIMA model applied for seasonal series. The model is applied when the data has seasonality. The seasonal ARIMA model contains regular autoregressive and moving average terms that accounts for the correlation at lower lags and seasonal auto regression and moving average terms that account for the correlation as the seasonal lags. When the series is non-stationary an additional difference is required to convert into stationary and estimate the model. An appropriate model expressed as SARIMA (p, d, q) $(P, D, Q)^s$ order. Where p is regular autoregressive term, d is regular differences, q is regular moving average term, P is seasonal autoregressive term, D is seasonal difference at seasonal lag, Q is seasonal moving average term and s is seasonal order. Here it is essential to choose the order of p, d, q and P, D, Q through minimum of SBC criteria.

5. Evaluation Process

There are several ways to evaluate forecasting models such as accuracy, cost associated with estimation and forecasting, easy of the estimation. The performance of the forecasting model is measured by applying forecast accuracy, which is the most dominant measure. Forecast accuracy is measured by the difference between actual value and the forecasted value at time period t. It is often called as error term. Assume that Y_t is the actual value, $Y^\wedge_t$ is the forecasted value, then $et=Y_t-Y^\wedge_t$ is the error term/residual term. The various alternative measures of forecast accuracy are mean absolute deviation (MAD), mean absolute percentage error (MAPE), mean squared error (MSE). They may be defined as,

$$MAD = \frac{1}{n}\sum_{t=1}^{n} |e_t|$$

$$MSE = \frac{1}{n}\sum_{t=1}^{n} (e_t^2)$$

$$MAPE = \frac{1}{n}\sum_{t=1}^{n} \frac{|e_t|}{Y_t}$$

MAD is the most useful measure when forecast error is measured in the same unit as the original series. MSE measures by considering the square of the error term removing the –ve deviations. The MAPE is useful when the size or magnitude of the forecast variable is important. MAPE is used to compare between different models.

6. Empirical Analysis

The time series plot of international tourists arrival in India is given in figure 1a. Two characteristics are observed from the data sets. One, there has been a non-linearly increasing trend. Two, the seasonality is clearly evident, with continuously increasing variability between the seasonal intervals. The tourist arrival is generally higher during December of every year. The fact is that December being the winter season; the weather is cool and calm favoring international tourists to spend a wonderful time. It also may be due to that fact that, December is a month of basket of holidays like Christmas, year ending etc. May to June witness lower international tourists turnover as weather is unfavorable; summer will be in peak, temperature touching around 40^0C and above in almost every part of India.

The preliminary statistical analysis in term of descriptive statistics is given in Table 1 which shows that, the average number of tourist footfall from Dec 90 to –Jan 2010 is 248122, maximum being 646000 and minimum is 103371. The skewness and kurtosis, as well as Jarque-Bera statistics demonstrate that the series is not normally distributed and probably could be non-stationary. The coefficient of variation is 0.46 indicating almost 50% variability of the series from its average value.

To identify the nature of the series, the auto correlation coefficients are estimated and the corresponding correlogram can be represented in graph 1b. From the correlogram we found that the ACF dies out slowly up to lag 6 and after that it again increases. There are ups and downs in ACFs. The ACFs also statistically significant up 19^{th} lag period. The coefficient of ACF dies out very slowly after several lag. The ACF up to lag 3 and then from 6 to 13 is greater than 0.7. The coefficient of ACF is also higher at a particular seasonal interval. It indicates that the series is clearly non-stationary with trend and seasonality. The Augmented Dickey Fuller (Dickey- Fuller, 1981) unit root test also verifies that the series is non-stationary at level and stationary at first difference. The table 2 reports the ADF test result (Note 4). Here the null hypothesis of non-stationary at level is accepted, but the same is rejected at first difference when the model has both constant and constant with trend term. The ADF test is performed to estimate the SARIMA model.

Forecast values for different foresting models are reported in table 3. As in the simple naïve model only for one period ahead forecasting is possible since past values will be the future values, the forecasted value for the month of Feb 2010 is 491000. The error is notably high with MAD is 36211 and MAPE is 0.146. However, this model may not give suitable values as it neglects the past observations. As the data might have trending pattern or seasonality or both trend with seasonality. To deal with different features of time series data we have estimated and forecasted using naïve trend, naïve seasonal and naïve trend with seasonal models. Using these three versions of the model the forecasted values are obtained. Among these models the naïve with trend and seasonality provides better forecast values than other models as MPAE is lowest i.e. 0.099.

In simple average method, the forecasted value is the average of previous data. Using this method the MAPE obtained is 0.250. We have applied four different average methods such as moving average (3), moving average (12), double moving average (3) and double moving average (12). Among these models MA (12) provides better forecasted value as the error is lowest.

Among various types of smoothing model, forecasted values using simple exponential smoothing models, Holt's methods and winter's Methods are reported in table 3. In simple exponential model the values of alpha, the smoothing parameter is chosen as 0.9, which is based on lowest MAPE. Similarly, alpha and beta for Holt's methods are equal to 0.9 and 0.2 as with this smoothing parameter the error is lowest. The values of alpha, beta and gamma are chosen as 0.9, 0.2 and 0.1 respectively for the Winter's methods. All these three methods are estimated using Minitab software. The corresponding forecasted values with the error term are reported in the table 3. Similarly the forecasted values with the actual value are denoted graphically in graph 4. Among these models Winter's methods provides better forecasted value as MAPE of 0.15 is lowest.

The above models do not forecast different components of time series separately. To separate out the seasonality, trends, cyclicality, and randomness in the data pattern, we have applied seasonal decomposition model. The model separates out entire components in the data sets. In view of the fact that, the data shows increasing variability, we applied multiplicative seasonal decomposition model. Using the seasonal decomposition the forecasted values are obtained along with the MAPE. The MAPE is 0.262 which is quite low and therefore this particular model also provides better forecasted values of tourist demand.

While the data has quadratic pattern, the quadratic equation $Y_t=164494-552_t+8.30_t^2$ is estimated. All three coefficients are statistically significant individually; the corresponding t values are 13.155, -2.208 and 7.923 respectively. Similarly p values to reject the null hypothesis are 0.000, 0.028 and 0.000 respectively. The Adjusted R^2 value is 0.705 which is reasonably good enough. The model is free from heteroscedasticity problem and although it shows some sort of autocorrelation. The monthly forecasted values are reported in table 3.

Finally we have applied Box-Jenkins (1970) approach of ARIMA model. Given that the data has seasonal pattern and non stationary the SARIMA model is applied to obtain forecast values. Since ARIMA model does not fit the data well due to seasonality, the ARIMA results are not reported here. The application of SARIMA model involves certain steps. Initially the stationary of the series is tested applying ACF or ADF unit root tests. The test results show that the series is non-stationary at level, but stationary at first difference. To make it stationary we have taken first difference of the data sets. We have run several SARIMA model with alternative combinations of (p, d, q) (P, D, Q) order with 12 month seasonality. The appropriate model chosen is of order SARIMA (1, 1, 1) (2, 1, 4) 12 based on lowest SBC criteria and the significance of the coefficients. The results are reported in table 5. Other results are not reported due to space consumption and can be obtained from the author upon request. All coefficients of SARIMA model are statistically significant as p values are very low. The diagnostic checking in terms of LBQ and Q statistics also supports the above results.

To compare among these alternative models, we have used MAPE. This is the most suitable methods to compare the model. The result is reported in table 4. The results show that MAPE equal to 0.030 is lowest corresponding to SARIMA model. It means Applying SARIMA model, the chance of actual values different from forecasted value is only 3%. Therefore this model provides better forecasting performance than the other model. The implication of this result is that, the tourists arrivals in India can be better forecasted using SARIMA model, which captures seasonality very well. A comparative study of the forecasted values and the actual values since Feb 2010 to Sept 2010 reveals that the forecasted values obtain through SARIMA model is quite close to the actual values over the months.

7. Conclusion

This paper thrashes out the tourist demand forecasting applying various alternatives forecasting models and asses the forecasting performance. Using uni-variate forecasting model, the forecasting performance of various models are evaluated, based on alternative model selection criteria, such as MAD, MAPE and MPE. For empirical analysis monthly data over the period Dec 1990 to Jan 2010 is applied. The result found that SARIMA model provides better forecast values than other competing model. The advantage of this model is also that it not only captures Autoregressive and moving average process in the series but also the seasonal process. As policy implications, for uni-variate series SARIMA model can be applied for forecasting tourists demand in India. However, other factors affecting tourist demand should also be taken into consideration while forecasting tourists demand, which perhaps is one of the limitations of this paper.

References

Box, G .E. P., & Jenkins, G. M. (1970). *Time Series Analysis: Forecasting and Control*. San Francisco: Holden day.

Chan, F., Lim C., & Mcaleerand M. (2005). Modeling multivariate international tourism demand and volatility. *Tourism Management,* 26, pp. 459-471.

Cho, V. (2003). A comparison of three different approaches to tourist arrival forecasting. *Tourism Management,* 24, pp.323-330.

Crouch, G. I. (1994). The study of international tourism demand: a review of practice. *Journal of Travel Research*, 33, pp. 41-54.

Deaton, A. S., & Muellbauer J. (1980). An almost ideal demand system. *American Economic Review*, 70, pp.312-326.

Dickey, D. A., & Fuller, W. A. (1981). Likelihood ratio statistics for autoregressive time series with a unit root. *Econometrica*, 49, pp. 1057-1072.

Goh, C., & Law, R. (2002). Modeling and forecasting tourism demand for arrivals with stochastic non-stationary seasonality and intervention. *Tourism Management*, 23, pp.499-510.

Gustavsson, P., & Nordström J. (2001). The impact of seasonal unit roots and vector ARMA modeling on forecasting monthly tourism flows. *Tourism Economics*, 7, pp.117-133.

Hanke, J. E., & Wichern, D. W. (2007). *Business Forecasting,* Pearson Education, New Delhi, Eight Editions

Hernández-López, M. (2004). Future tourists' characteristics and decisions: the use of genetic algorithms as a forecasting method. *Tourism Economics,* 10, pp.245-262.

Hylleberg, S., Engle, R. F., Granger, C. W. J., & Yoo, B. S. (1990). Seasonal integration and cointegration. *Journal of Econometrics*, Elsevier, 44(1-2), pp. 215-238.

Kulendran, N., & Wilson, K. (2000). Modeling business travel. *Tourism Economics*, 6, pp.47-59.

Li, G., Song, H., & Witt, S. F. (2005). Recent developments in econometric modeling and forecasting. *Journal of Travel Research,* 44, pp.82-99.

Lim, C. (1999). A meta-analysis review of international tourism demand. *Journal of Travel Research,* 37, pp.273-284.

Lim, C., & Mcaleer, M. (2001b). Monthly seasonal variations: Asian tourism to Australia. *Annals of Tourism Research*, 28, pp. 68-82.

Naude, W. A., & Saayman, A. (2005). Determinants of tourist arrivals in Africa: a panel data regression analysis. *Tourism Economics*, 11, pp.365-391.

Roget, F. M., & Gonzalez, X. A. R. (2006). Rural tourism demand in Galicia, Spain. *Tourism Economics,* 12, pp.21-31.

Shan, J., & Wilson, K. (2001). Causality between trade and tourism: empirical evidence from China. *Applied Economics Letters*, 8, pp.279-283.

Smeral, E., & Wager, M. (2005). Does complexity matter? Methods for improving forecasting accuracy in tourism: the case of Australia. *Journal of Travel Research*, 44, pp.100-110.

Song, H., & Li, G. (2008). Tourism demand modeling and forecasting: a review of recent research. *Tourism Management*, 29 (2), 203-220.

Song, H., & Turner, L. (2006). *Tourism demand forecasting*, Dwyer, L., and Forsyth, P. (Eds) International Handbook on the Economics of Tourism, Edward Elgar: Cheltenham

Song, H., & Witt, S. F. (2000). *Tourism demand modeling and forecasting: modern econometric approaches*, Pergamon: Cambridge.

Turner, L. W., & Witt, S. F. (2001a). Factors influencing demand for international tourism: tourism demand analysis using structural equation modeling revisited. *Tourism Economics*, 7, pp.21-38.

Wang, C. H. (2004). Predicting tourism demand using fuzzy time series and hybrid grey theory. *Tourism Management,* 25, pp.367-374.

Witt, S. F., & Witt, C. A. (1995). Forecasting tourism demand: a review of empirical research. *International Journal of Forecasting*, 11, pp. 447-475.

Witt, S. F., Song, H. & Wanhill, S. P. (2004). Forecasting tourism-generated employment: the case of Denmark. *Tourism Economics*, 10, pp.167-176.

Wong, K. K. F., Song, H., & Chon, K. S. (2006). Bayesian models for tourism demand forecasting. *Tourism Management,* 27, pp.773-780.

Notes

Note 1. Tourism is an activity whereby a person can travel and stay away from the place outside their usual environment for not more than a year for leisure, business and other purpose. Various segments of tourism are medical tourism, eco tourism, heritage tourism, adventure tourism.

Note 2. Broadly forecasting tourism demand can be categories into two competing approaches, namely quantitative and qualitative approach. The quantitative approaches examines past data by developing and employing a range of mathematical and statistical tools to estimate and test the theories pertaining to natural phenomena. On the other hand, the qualitative approaches strive to study and understand various facet of human behavior in-depth. The later models apply judgment, intuition, belief, experience, survey methods etc. for forecasting, the former apply various uni-variate or multi-variate statistical and mathematical model for forecasting the phenomena.

Note 3. For details of the various smoothing methods please refer the book, 'Business Forecasting' by John E. Hanke and Dean W. Wichern, Pearson Education, New York, 2007.

Note 4. HEGY(1990) unit root test can be also be performed to tests the seasonal unit root.

Table 1. Descriptive Statistics of Tourists Arrivals (Year wise)

Year	Mean	Median	Max	Min.	S.D.	Sk.	Ku.	J-B	P Value	C.V
1991	139792	136004	203098	109988	28430.23	0.913	3.000	1.668	0.434	0.203
1992	220411	219740	313342	139575	52278.59	0.242	2.139	0.488	0.784	0.237
1993	207703	203101	283760	151098	40772.36	0.480	2.207	0.776	0.678	0.196
1994	198697	195749	296474	134566	49102.98	0.455	2.246	0.698	0.705	0.247
1995	227184	221916	319271	141508	54988.32	0.064	1.926	0.585	0.746	0.242
1996	288123	282821	417527	185502	70712.11	0.375	2.144	0.647	0.724	0.245
1997	326551	327813	479411	225394	79742.25	0.432	2.122	0.759	0.684	0.244
1998	370597	364171	541571	255008	87513.03	0.445	2.163	0.745	0.689	0.236
1999	423484	422215	596560	277017	105551.60	0.116	1.701	0.870	0.647	0.249
2000	441599	440783	561393	300840	91759.87	-0.141	1.552	1.089	0.580	0.208
2001	426833	427000	646000	295000	99635.55	0.678	2.965	0.920	0.631	0.233
2002	155638	157250	189573	123446	21764.82	-0.126	1.802	0.750	0.687	0.140
2003	147069	147223	214408	103371	31277.96	0.750	3.003	1.126	0.569	0.213
2004	157201	158222	219501	107611	33078.29	0.252	2.224	0.428	0.807	0.210
2005	176974	179536	252887	122918	38197.70	0.372	2.477	0.414	0.813	0.216
2006	190655	186271	273310	132833	43450.30	0.343	2.202	0.555	0.758	0.228
2007	197841	197561	270954	137536	44532.52	0.122	1.707	0.865	0.649	0.225
2008	196552	189612	269810	137868	42736.62	0.196	1.865	0.721	0.697	0.217
2009	206827	204243	277504	140168	45804.42	0.072	1.643	0.931	0.628	0.221
Overall	248122	217625	646000	103371	115031.1	1.9	3.73	59.3	0.000	0.463

The abbreviations S.D. stands for standard deviation, Sk for skewness, Ku for kurtosis, J-B for Jarque Bera value, P for probability value of JB test, C.V. for coefficient of variation

Table 2. Augmented Dickey Fuller (ADF) Unit Root Test

	Level		First difference	
	C	CT	C	CT
‘t’Statistics	0.942	-1.146	-5.462*	-5.700*
‘P’ values	(0.996)	(0.918)	(0.000)	(0.000)

* Denotes 1% significance Level. C stand for constants, and CT for constant with trend. ‘P’ values are MacKinnon (1996) one sided p-values.

Table 3. Forecasted Values and Actual Values

	Forecasting Period											
	For the Year 2010											**2011**
Model	**Feb**	**Mar**	**Apr**	**May**	**Jun**	**Jul**	**Aug**	**Sep**	**Oct**	**Nov**	**Dec**	**Jan-11**
RWM	491000											
NTM	336000											
NS	502000	472000	371000	295000	341000	432000	358000	317000	448000	500000	646000	491000
NTS	507750	430167	331667	264083	316417	403583	322000	287167	421583	500000	602833	437167
Average	248122											
MA(3)	551667	568500	491000									
MA(12)	432583	426273	421700	427333	443875	458571	463000	484000	525750	600000	568500	491000
DMA(3)	551667	568500	491000									
DMA(12)	430093	444093	430473	420780	433234	468825	498713	504759	546548	600000	677216	700757
QTM	480074	483367	486676	490000	493344	496702	500078	503470	506579	510300	513740	517204
SD	475098	442341	343081	281975	308305	389808	358225	329621	428016	497943	561304	509543
ES	505137	505137	505137	505137	505137	505137	505137	505137	505137	505137	505137	505137
HM	519875	529977	540078	550180	560281	570382	580484	590585	600686	610788	620889	630991
WM	474941	442990	337191	280250	305454	387527	352611	317549	410725	482272	530676	473605
SARIMA	521677	492173	378248	322754	377709	460555	391573	341884	476731	500000	626963	524218
Actual values	**601000**	**472000**	**354000**	**345000**	**370000**	**452000**	**382000**	**369000**				

Here the above abbreviations stands for as : RWM- random walk model, NTM- Naïve Trend Model, NS- Naïve seasonal model, NTS- Naïve Trend with seasonality, MA- moving average, DMA- double moving average, TM- quadratic trend model, SD- Seasonal decomposition, ES- exponential smoothing, HM- Holt’s Methods, WM- Winter’s methods, SARIAM- Seasonal ARIMA models.

Table 4. Forecast Models Comparison

Models	MAD	MSE	MAPE
RWM	36244	2185744331	0.146
NTM	42359	3436619197	0.171
NS	27474	1336351293	0.107
NTS	25413	1167290649	0.099
Average	77889	13148342620	0.250
MA(3)	46934	3798638815	0.191
MA(12)	48555	3793478517	0.191
DMA(3)	10906	31406878608	0.370
DMA(12)	51753	4061841528	0.207
QTM	48883	3875572004	0.210
SD	33596	1947356375	0.150
ES	36948	2281620470	0.150
HM	40713	2708525708	0.160
WM	14527	433232890	0.060
SARIMA	**13752**	**235634241**	**0.030**

Table 5. SARIMA (1,1,1) $(2,1,4)^{12}$ Model Output

Type	lag	Coefficient	SE of Coefficient	t Statistics	P Values
AR	1	0.601*	0.0753	7.98	0.0000
SAR	12	-1.899*	0.0262	-72.59	0.0000
SAR	24	-0.988*	0.0261	-37.82	0.0000
MA	1	0.913*	0.0356	25.61	0.0000
SMA	12	-1.725*	0.019	-90.55	0.0000
SMA	24	-0.357*	0.0665	-5.37	0.0000
SMA	36	0.772*	0.1274	6.06	0.0000
SMA	48	0.372*	0.0898	4.15	0.0000

1. Denotes 1% significance level.
2. The abbreviation AR, SAR, MA, SMA stands for Autoregressive, seasonal autoregressive, moving average, seasonal moving average respectively.

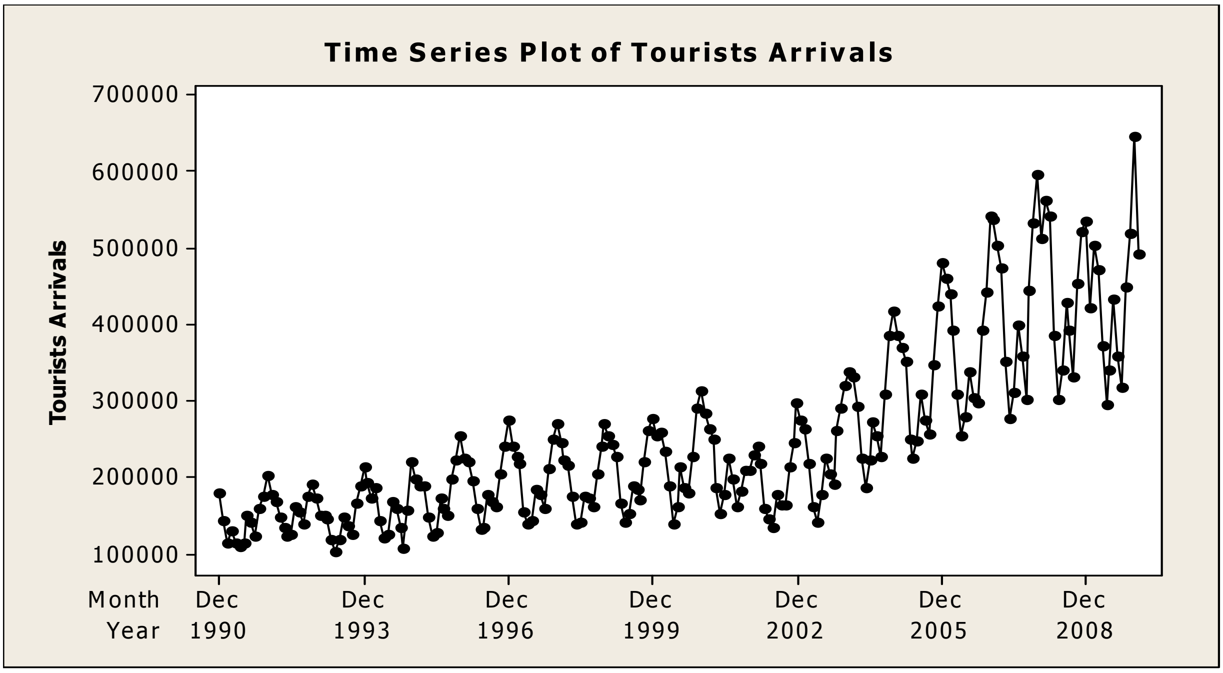

Figure 1. Time Series Plot of Tourists Arrivals

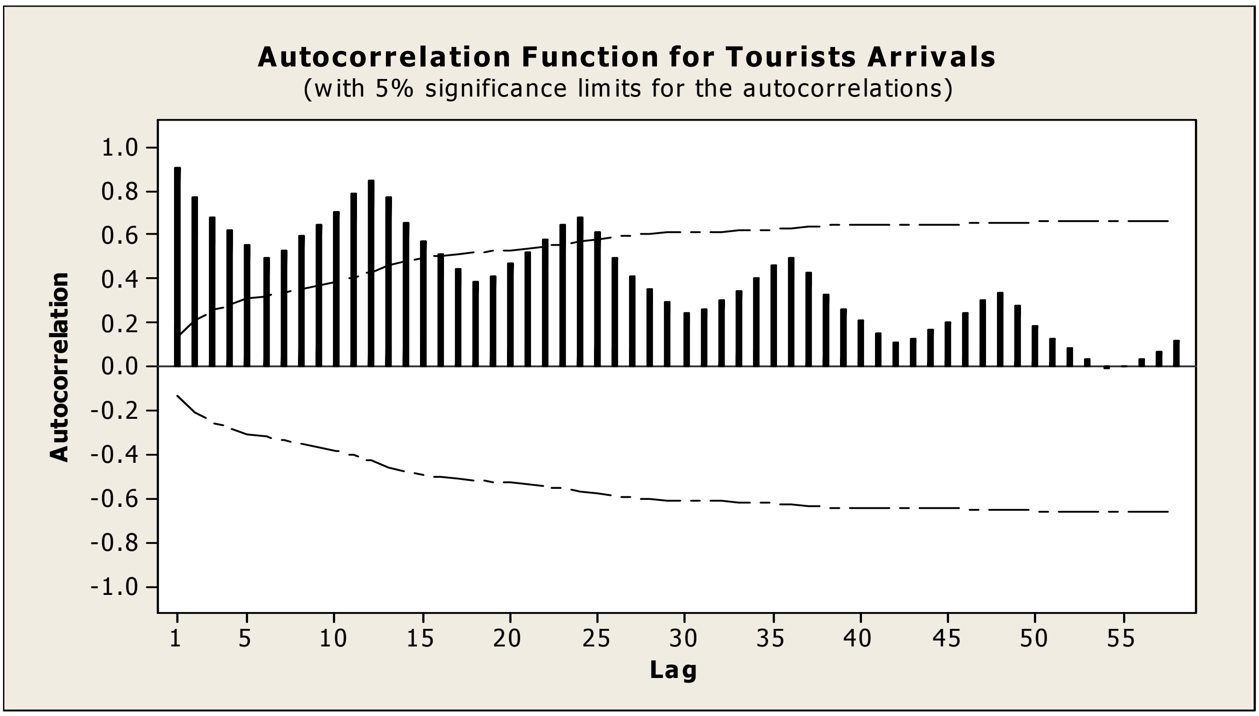

Figure 2. Autocorrelation Function of Tourists Arrivals

ISSN 1833-3850 E-ISSN 1833-8119