Open Journal of Statistics, 2015, 5, 60-65
Published Online February 2015 in SciRes. http://www.scirp.org/journal/ojs
http://dx.doi.org/10.4236/ojs.2015.51008

Statistical Models for Forecasting Tourists' Arrival in Kenya

Albert Orwa Akuno[1], Michael Oduor Otieno[2], Charles Wambugu Mwangi[1], Lawrence Areba Bichanga[1]

[1]Department of Mathematics, Egerton University, Egerton, Kenya
[2]Department of Mathematics and Computer Science, Laikipia University, Nyahururu, Kenya
Email: orwaakuno@gmail.com, mkaili91@yahoo.com, charlesmwangi@gmail.com, lawareba@gmail.com

Received 24 January 2015; accepted 13 February 2015; published 16 February 2015

Abstract

In this paper, an attempt has been made to forecast tourists' arrival using statistical time series modeling techniques—Double Exponential Smoothing and the Auto-Regressive Integrated Moving Average (ARIMA). It is common knowledge that forecasting is very important in making future decisions such as ordering replenishment for an inventory system or increasing the capacity of the available staff in order to meet expected future service delivery. The methodology used is given in Section 2 and the results, discussion and conclusion are given in Section 3. When the forecasts from these models were validated, Double Exponential Smoothing model performed better than the ARIMA model.

Keywords

Exponential Smoothing, ARIMA Model, Tourists' Arrival Data

1. Introduction

Tourism is one of Kenya's major foreign exchange earners. This greatly depends on the arrival of various groups of tourists. The forecast of tourists' arrivals is important since it would enable the tourism related industries like airlines, hotels and other stakeholders to adequately prepare for any number of tourists at any future date. In this paper, an attempt has been made to forecast tourists' arrivals using statistical time series modeling techniques—Double Exponential Smoothing and Auto-Regressive Integrated Moving Average (ARIMA). [1] used the same models to forecast milk production in India. [2] used univariate SARIMA models to forecast tourists' demands in India.

Then data on tourists' arrival in Kenya were obtained from the Ministry of East African Affairs, Commerce

How to cite this paper: Akuno, A.O., Otieno, M.O., Mwangi, C.W. and Bichanga, L.A. (2015) Statistical Models for Forecasting Tourists' Arrival in Kenya. *Open Journal of Statistics*, **5**, 60-65. http://dx.doi.org/10.4236/ojs.2015.51008

and Tourism, Department of Tourism. Tourists' arrival for the period 1995 to 2008 was used for model fitting, and data for the remaining periods from 2009 to 2012 were used for model validation. The analysis was carried out using R-language, Excel and Minitab version 16.1.1.

2. Methodology

2.1. Selection of Appropriate Smoothing Techniques

Once the presence of trend is detected in the data, smoothing of the time series data follows. Various smoothing techniques as discussed by [3] include; Simple Exponential Smoothing (SES), Double Exponential Smoothing (DES), Triple Exponential Smoothing (TES) and Adaptive Response Rate Simple Exponential Smoothing (ARRSES) which are briefly described below:

2.1.1. Simple Exponential Smoothing (SES)

For the series $Y_1, Y_2, \cdots, Y_t$, the forecast for the preceding value Y_{t+1}, say F_{t+1}, is based on the weights α and $1-\alpha$ to the recent observation Y_t and forecast F_t respectively, where α is the smoothing constant. The form of the model is

$$F_{t+1} = F_t + \alpha\left(Y_t - F_t\right). \tag{1}$$

The size of α used has a great influence on the forecast. The best value of α corresponding to the minimum mean square error (MSE) is usually used.

2.1.2. Double Exponential Smoothing (Holt's)

The form of the model is

$$\left.\begin{aligned} L_t &= \alpha Y_t + (1-\alpha)\left(L_{t-1} + b_{t-1}\right) \\ b_t &= \beta\left(L_t - L_{t-1}\right) + (1-\beta) b_{t-1} \\ F_{t+m} &= L_t + b_t m \end{aligned}\right\} \tag{2}$$

where L_t in the model is the level of the series at time t and b_t is the slope (Trend) of the series at time t, α and β $(=0.1, 0.2, \cdots, 0.9)$ are the smoothing coefficient for level and smoothing coefficient for trend respectively. In order to fit the model, it is necessary to calculate the initial values of the level L_0 and the trend b_0. [4] suggests that the initial values can be obtained as $L_0 = Y_1$ and $b_0 = 0$, or $Y_2 - Y_1$ or $\left(Y_n - Y_1\right)/(n-1)$. In this paper, the initial values have been obtained as $L_0 = Y_1$ and $b_0 = \left(Y_n - Y_1\right)/(n-1)$.

The pair of α and β that gives a minimum Mean Square Error is preferred.

2.1.3. Triple Exponential Smoothing (Winter's)

When time series data exhibit seasonality, Triple Exponential Smoothing method is the most recommendable. It incorporates three smoothing equations; first for the level, second for trend and third for seasonality.

2.2. Auto-Regressive Integrated Moving Average (ARIMA Model)

2.2.1. Model Identification

According to Box and Jenkins two graphical procedures are used to access the correlation between the observations within a single time series data. According to [5], these devices are called an estimated autocorrelation functions and the estimated partial autocorrelation function. These two procedures measure statistical relationships within the time series data. Summarization of statistical correlation within the time series data is the other step in the identification. Box and Jenkins suggest a whole family of ARIMA models from which we may choose.

In choosing the model that seems appropriate we use the estimated ACF and PACF. This is due to the basic idea that every ARIMA model will have unique ACF and PACF associated with it. Thus we select the model whose theoretical ACF and PACF resembles the anticipated ACF and PACF of the time series data [6].

2.2.2. Estimation

An estimate of the coefficients of the model is obtained by modified least squares method or the maximum likelihood estimation method suitable to the time series data.

2.2.3. Diagnostic Checking

Diagnostic checks help to determine if the anticipated model is adequate. At this stage, an examination of the residuals from the fitted model is done and if it fails the diagnostic tests, it is rejected and we repeat the cycle until an appropriate model is achieved.

The ARIMA model is obtained by taking W_t as the first differenced time series, *i.e.* $d = 1$

$$(W_t - \mu) - \alpha_1 (W_{t-1} - \mu) - \cdots - \alpha_p (W_{t-p} - \mu) + \varepsilon_t + \beta_1 \varepsilon_{t-1} + \cdots + \beta_q \varepsilon_{t-q}. \quad (3)$$

Equation (3) is referred to as the ARIMA $(p, 1, q)$.

Different combinations of AR and MA individually yield different ARIMA models [7]. The optimal model is obtained on the basis of minimum value of Akaike Information Criteria (AIC) given by

$$\text{AIC} = -2\log L + 2m \quad (4)$$

where $m = p + q$ and L is the likelihood function. The Root Mean Square Error (RMSE) and the Mean Absolute Percentage Error (MAPE) are used to evaluate the performance of the various models and are given below.

$$\text{MAPE} = \frac{1}{n} \sum_{t=1}^{n} \left| \frac{Y_t - F_t}{Y_t} \right| \times 100 \quad (5)$$

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{t=1}^{n} (Y_t - F_t)^2} \quad (6)$$

where Y_t is the tourists' arrival in different years and F_t is the forecasted tourists' arrivals in the corresponding years and n is the number of years used as forecasting period.

3. Results and Discussions

3.1. Exponential Smoothing Model

Table 1 shows the yearly tourists' arrival in Kenya (in thousands) for the period 1995-2012. The time plot (**Figure 1**) revealed that there was increasing trend from the year 2002 to 2007. However, there was a sharp drop

Table 1. Data on tourists' arrival in Kenya for the period 1995 to 2012.

Sl. No.	Year	Observed tourists' arrival ('000)
1	1995	973.6
2	1996	1003.0
3	1997	1000.6
4	1998	894.3
5	1999	969.3
6	2000	1036.5
7	2001	993.6
8	2002	1001.5
9	2003	1146.2
10	2004	1360.7
11	2005	1479.0
12	2006	1600.7
13	2007	1816.8
14	2008	1203.2
15	2009	1490.4
16	2010	1609.1
17	2011	1822.9
18	2012	1873.8

Source: Ministry of East African affairs, Commerce and Tourism: Department of Tourism (Kenya).

in the number of tourists in the year 2008 followed by an increasing trend from the year 2009 to 2012. For smoothing the data, Holt's Double Exponential Smoothing was used. Various combinations of α and β both ranging from 0.1 to 0.9 with increments of 0.1 were tried and Mean Squared Error for the forecasts (54.186) and Mean Absolute Percentage Error (3.028) was least for $\alpha = 0.1$ and $\beta = 0.7$. The fitted model is therefore given by;

$$\left.\begin{aligned} L_t &= 0.1Y_t + 0.9\left(L_{t-1} + b_{t-1}\right) \\ b_t &= 0.7\left(L_t - L_{t-1}\right) + 0.3b_{t-1} \\ F_{t+m} &= L_t + b_t m \end{aligned}\right\} \tag{7}$$

where $m = 1, 2, 3$ and 4 the initial values for the level L_t and trend b_t are 973.6 and 17.66 respectively. **Table 2** shows the forecast of tourists' arrivals using the chosen double exponential smoothing model.

3.2. ARIMA Model

Figure 1 showed that the series was non-stationary since there was some trend component present. The data was made stationary by taking the first order difference $(d = 1)$. The time plot of the differenced data is shown in **Figure 2**.

Using R-language for different values of p and q, various ARIMA models were fitted and the best model was chosen on the basis of minimum value of the selection criteria, that is, Akaike Information Criteria (AIC) whose formula is given in Equation (4). In this way, ARIMA (1, 1, 1) was found to be the best model. The fitted model is given by

$$Y_t = 43.6319 - 0.9999\varepsilon_t + 0.4115Y_{t-1}. \tag{8}$$

The estimation of the model parameters was done by maximum likelihood estimation technique. The fitted model was then used to forecast tourists' arrival from 2009 to 2012. The forecast values are shown in **Table 3**.

Table 2. Forecast of tourists' arrival in Kenya using double exponential smoothing.

S. No.	Year	Observed tourists' arrival ('000)	Forecast of tourists' arrival
			Double exponential model
1	2009	1490.4	1560.936
2	2010	1609.1	1660.595
3	2011	1822.9	1760.254
4	2012	1873.8	1859.912

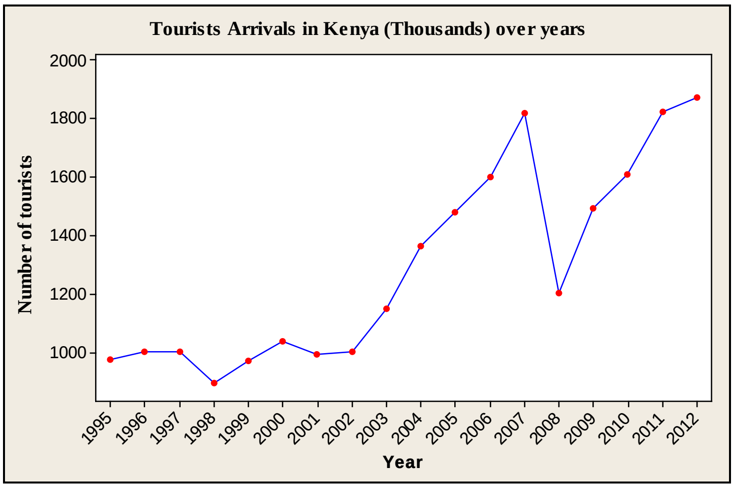

Figure 1. Time plot of tourists' arrival in Kenya between 1995 and 2009.

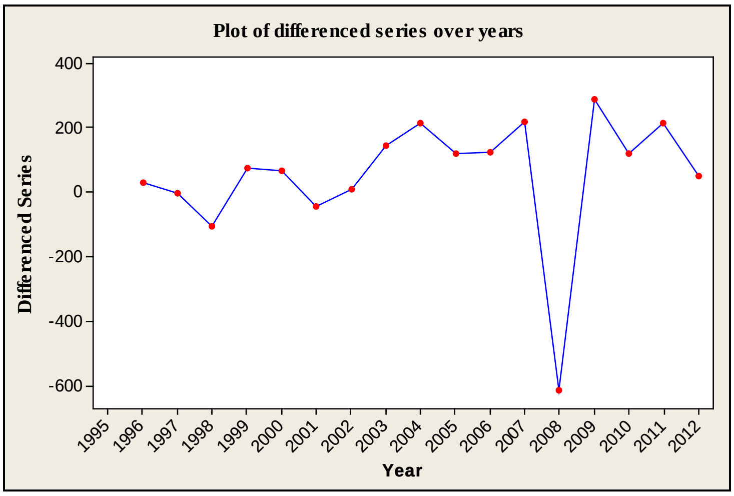

Figure 2. Plot of differenced tourists' arrival data.

Table 3. Forecast of tourists' arrival in Kenya ARIMA (1, 1, 1) models.

Sl. No.	Year	Observed tourists' arrival ('000)	Forecast of tourists' arrival
			ARIMA (1, 1, 1) model
1	2009	1600.7	1393.607
2	2010	1816.8	1497.643
3	2011	1203.2	1566.134
4	2012	1490.4	1619.997

Table 4. Forecast of tourists' arrival in Kenya using double exponential smoothing and ARIMA (1, 1, 1) models.

Sl. No.	Year	Observed tourists' arrival ('000)	Forecast of tourists' arrival	
			Double exponential model	ARIMA (1, 1, 1) model
1	2009	1490.4	1560.936	1393.607
2	2010	1609.1	1660.595	1497.643
3	2011	1822.9	1760.254	1566.134
4	2012	1873.8	1859.912	1619.997
MAPE			3.028	10.263
RMSE			54.186	195.023

3.3. Comparison and Conclusion of the Performance of the Two Models

Performance evaluation measures MAPE and the RMSE were obtained for the forecasted tourists' arrivals for the years 2009 to 2012.

The comparison of the two models based on MAPE and RMSE is as given in **Table 4**. Based on the results from the table, Double Exponential Smoothing model was the best to forecast tourists' arrival in Kenya as both its MAPE and RMSE values were least compared to those of ARIMA (1, 1, 1).

References

[1] Satya, P., Ramasubramanian, V. and Menta, S.C. (2007) Statistical Models for Forecasting Milk Production in India. *Journal of the Indian Society of Agricultural Statistics*, **61**, 80-83.

[2] Padhan, P.C. (2011) Forecasting International Tourists Footfalls in India: An Assortment of Competing Models. *International Journal of Business and Management*, **6**, 190-202.

[3] Gardener, E.S. (1985) Exponential Smoothing—The State of the Art. *Journal of Forecasting*, **4**, 1-28. http://dx.doi.org/10.1002/for.3980040103

[4] Jani, P.N. (2014) Business Statistics: Theories and Applications. PHI Learning Private Limited, Delhi.

[5] Box, G.E.P. and Jenkins, G.M. (1970) Time Series Analysis: Forecasting and Control. Holden-Day, San Francisco.

[6] Pankratz, A. (1983) Forecasting with Univariate Box-Jenkins Models: Concepts and Cases. John Wiley and Sons, New York. http://dx.doi.org/10.1002/9780470316566

[7] Makridakis, S., Wheelwright, S.C. and Hyndman, R.J. (1998) Forecasting: Methods and Applications. John Wiley & Sons, New York.