

Forecasting Annual International Tourist Arrivals in Zambia Using Holt-Winters Exponential Smoothing

Stanley Jere^{1*}, Alick Banda¹, Bornwell Kasense¹, Ian Siluyele², Edwin Moyo³

¹Department of Mathematics and Statistics, Mulungushi University, Kabwe, Zambia

²Zambia Information and Communications Technology Authority, Lusaka, Zambia

³Department of Mathematics and Statistics, Northrise University, Ndola, Zambia

Email: *sjere@mu.ac.zm

How to cite this paper: Jere, S., Banda, A., Kasense, B., Siluyele, I. and Moyo, E. (2019) Forecasting Annual International Tourist Arrivals in Zambia Using Holt-Winters Exponential Smoothing. *Open Journal of Statistics*, 9, 258-267.

<https://doi.org/10.4236/ojs.2019.92019>

Received: November 6, 2018

Accepted: April 25, 2019

Published: April 28, 2019

Copyright © 2019 by author(s) and Scientific Research Publishing Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

Tourism is one of the major contributors to foreign exchange earnings to Zambia and world economy. Annual International tourist arrivals in Zambia from 1995 to 2014 are considered in this paper. In this study we evaluated the model performance of Auto-Regressive Integrated Moving Average (ARIMA) and Holt Winters exponential smoothing (HWES). The error indicators: Mean percentage error (MPE), Mean absolute error (MAE), Mean absolute scaled error (MASE), Root-mean-square error (RMSE) and Mean absolute percentage error (MAPE) showed that HWES is an appropriate model with reasonable forecast accuracy. The HWES ($\alpha = 1, \beta = 0.1246865$) showed smallest error than those of the ARIMA (0, 1, 2) models. Hence, the HWES ($\alpha = 1, \beta = 0.1246865$) can be used to model annual international tourist arrivals in Zambia. Further, forecasting results give a gradual increase in annual international tourist arrivals of about 42% by 2024. Accurate forecasts are key to new investors and Policymakers. The Zambian government should use such forecasts in formulating policies and making strategies that will promote the tourism industry.

Keywords

Zambia, International Tourist Arrivals, Holt-Winters Model, ARIMA, Forecasting

1. Introduction

Tourism is one of the major contributors to foreign exchange earnings for Zambia and other countries worldwide. According to [1] an international tourist

(overnight visitor) is an individual who travels to a country other than that in which they reside for a period not exceeding 12 months and whose main purpose in visiting is other than an remunerated activity within the country visited. Data collection methods for arrivals vary from one country to another. In some countries data are from border statistics. In other countries data are from tourism accommodation establishments. Tourist data refer to the number of arrivals, not to the number of people traveling. Thus a person who makes several travels to a country during a given period is counted each time as a new arrival. Tourism contributes highly to GDP, increasing the employment rate, source of revenue for local people, private sector, public sectors and government [2]. The significance of tourism has encouraged the authors to study the number of international tourist arrivals and attempt to make more accurate forecasting for future planning.

Zambia's tourist attraction includes 20 National Parks and 34 Game Management Areas (GMAs) with a total of 23 million hectares of land devoted to spectacular wildlife. Zambia has a rich array of traditional cultural festivities and events, including: Kuomboka Ceremony, Nc'wala Ceremony, Umutomboko Ceremony and LikumbiLya Mize Ceremony. One of the Seven Natural Wonders of the World is the Victoria Falls. The Falls plunge into the Zambezi River at about 550,000 cubic meters per second. The impact is so big such that falling water raises a cloud of vapor that can be seen more than 30 kilometers away. The fall has been known for centuries as Mosi-Oa-Tunya, meaning "The Smoke That Thunders" and it lie in the country's tourist capital called Livingstone, south of Zambia. It was declared a World Heritage Site for its unique geological/geomorphologic significance. Other water includes the Kalambo Falls, Ntumbachushi Falls, Ngonye Falls and the Chishimba Fall [3].

2. Literature Review

Studies by [2] were based on forecasting tourist arrival in Kenya using statistical time series modeling techniques. They used Double Exponential Smoothing and the Auto-Regressive Integrated Moving Average (ARIMA). They stated 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. Error measures such as MAPE and the RMSE were obtained in order to determine best model. Their results showed that the 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).

Studies by [4] involved applying a number of time series models to analyze Australia tourist arrivals. The models include: The Granger and impulse response analyses related to VAR model, ARIMA model and the theoretical model based on Butler. The models were calibrated using Australian tourist arrival data (1956-2010). The modified Butler model predicted growth of around 7.2 million

arrivals in 2015. The ARIMA (2, 2, 2) model improved the predictions and predicted a value of 6,016,012 in 2015. The actual value is well within the 95% range of 5,168,866 - 6,863,157. That is, a 95% range of arrival numbers predicted to lie within the range 5,618,866 to 6,863,157 for 2015. Hence the ARIMA (2, 2, 2) model gave best results. Further, they reported that a two-way causality between the tourist in Australia, Europe and World exists, while impulse response indicated different effect patterns, where tourist arrivals increase in the first period and declines in the second period but experience seasonal fluctuations in the third period.

Studies by [5] produced forecasts of international tourist arrival to Thailand during 2006-2010 using two methods: Structural and Trend Extrapolation Models. The Structural Model involved VAR model, GMM method, ARCH-GARCH method, ARCH-GARCH-M method, TARARCH method, EGARCH method and PARARCH method. The Trend Extrapolation Model involved Holt-winter method, ARIMA method, SARIMA method and Neural Network method. Structural Models gave SARIMA (0, 1, 1) (0, 1, 4) method as the best because of its low MAPE value. Trend Extrapolation Models gave VAR method because of its low MAPE value. SARIMA (0, 1, 1) (0, 1, 4) and VAR method predicted 15700656.00 million and 15985416.00 million in 2010 respectively. It was also concluded that the Thailand government tourism sector and private tourism industry sector should prepare adequately for a much more increase in number of international tourism arrival to Thailand during 2006-2010. They suggested an increase in the number of hotel, the number transportation, new tourism place, more unit of tourism polices, much more problem environment impact on tourism place, airport unit, budget for developing new tourism places and human training in tourism industry.

According to [6], tourism is a key sector and contributes significantly to foreign exchange earnings. Earnings from tourism in Kenya increased annually from Kenya Shillings 24.3 billion in 2001 to 73.7 billion in 2010. The number increased from 1,146,102 in the year 2003 to 1,822,885 in the year 2011. Major tourist attraction in Kenya includes: Nairobi, Beach, Mombasa, Coast Hinterland, Masailand, Nyanza basin. Tourism in Kenya boosts other industries like hotel and accommodation. Results in this paper also show that in order to improve tourism, a model that can give accurate forecast results is required so that hotel industry players can respond in good time to the anticipated changes in demand over time and also maximize returns on investments. The authors used the Box-Jenkins models to generate a forecasting model using quarterly data on bed occupancy by tourists visiting Kenya from 1974 to 2011. The SARIMA (1, 1, 2) (1, 1, 1) [5] was the best fit model for forecasting future quarterly demand on tourist accommodation in Kenya. They further concluded that this model should be used in forecasting future demands and maximize their returns on investment.

The study of [7] used a number of time series models of tourist arrivals and ARIMA (2, 2, 2) model was the best fit than logistic model. The models were ca-

librated using Australian tourist arrival data (1956-2010). The ARIMA (2, 2, 2) model predicted more accurately for the year 2010 with a value of 6,016,012 in 2015. The actual value was well within the 95% range of 5,168,866 - 6,863,157. That is, a 95% range of arrival numbers predicted lie within the range 5,618,866 to 6,863,157 for 2015. The results of the analysis also show close relationship between the tourist numbers recorded for Europe, Australia and the World even though Australia's tourist arrivals are very low in actual terms when compared to the other numbers of tourists. The logistic model was compared with ARIMA model and the ARIMA performed better in terms of the prediction for 2010.

The study by [8] involved forecasting international tourists footfalls in India using univariate time series forecasting models, for monthly data from Dec 1990 to Jan 2010. The forecasting performance of various models was evaluated using error measures. Also the actual and forecasted values were compared from Feb 2010 to Sept 2010. The SARIMA model performs better than other competing model for forecasting, with lowest MAPE value. They concluded that this model has an advantage over other models as it explicates autoregressive and moving average process not only for the data series but also of seasonality. Also, that the SARIMA model should be used for forecasting tourists demand in India.

A paper by [9] evaluated the forecasting performance of SARIMA, Holt Winters and Grey model for foreign tourist arrivals series in India from 2003:1 - 2015:10 using error measures. Turning point analysis and U statistic computed for SARIMA and MHW models. In addition, posterior variance ratio test to check the forecasting accuracy of Grey model was computed. Their results show that the SARIMA and MHW models outperform the other model when compared using the MAPE criterion. The forecasting performance of Grey model proved to be worse among all under MAPE and PVRT criteria. Although, it is observed that the Grey model is a better option for a de-seasonal series rather than seasonal series. However, they concluded using turning point analysis and U statistic that SARIMA and MHW are best fit models with highly forecasted accuracy and can be used in India.

A study by [10] used different time series approaches to model tourist arrivals to South Africa from its main overseas markets. Error measures were used to determine the best forecast model, these included the mean absolute percentage error (MAPE), root mean square error (RMSE) and Theill's U. The results show that seasonal ARIMA models deliver the most accurate predictions of arrivals than any other model. They concluded that one has to acknowledge that, although accurate, this method does not consider influence of external events. And therefore its application is limited to forecasting arrivals for businesses and government in the event that there are no substantial changes in the current environment. They also suggested for further research to include econometric forecasting techniques in order to address the critique above as well as the application of the current method to other sectors of tourism industry such as the accommodation industry.

3. Methodology

3.1. Holt-Winters Exponential Smoothing Model (HWES)

Holt-Winters exponential smoothing method is appropriate for forecasting non seasonal time series data. It is an extension of Simple Exponential Smoothing method and uses a linear combination of the previous values of a series for generating and modeling future values. It is applicable to time series data that has trend. Initial estimates and the slope of the trend are key to forecasting. The model for time series data Y_t is defined as:

$$L_t = \alpha Y_t + (1 - \alpha)(L_{t-1} + T_{t-1}), 0 < \alpha < 1$$

$$T_t = \beta(L_t - L_{t-1}) + (1 - \beta)T_{t-1}, 0 < \beta < 1$$

where, α is the smoothing constant, β is the trend smoothing constants, Y_t is raw data, L_t is smoothed data and T_t is the trend estimates.

The h -step-ahead forecast equation is $\hat{Y}_{t+h} = L_t + hT_t$ [9].

The main reason of choosing HWES model in this study is because Holt-Winters exponential smoothing technique can be used to forecast data containing trend.

3.2. Autoregressive Integrated Moving Average Model ARIMA

ARIMA models known as Box-Jenkins methodology have been found to be more popular, efficient and reliable even for short term forecasting. The ARIMA model consists of the following components called the order of autoregressive (AR) model (p), differencing order (d) and the order of moving average (MA) model (q). The Box-Jenkin models are denoted by ARIMA (p, d, q). “ I ” implies that the process needs to undergo differencing and when the modelling is done, the results undergo an integration process to produce forecasts and estimates. The MA, AR and ARMA are defined as follows:

$$\text{AR model: } Y_t = \sum_{i=1}^p \phi_i Y_{t-i} + \varepsilon_t,$$

$$\text{MA model: } Y_t = \sum_{i=1}^q \theta_i \varepsilon_{t-i}, \text{ and}$$

The combination of AR and MA gives

$$\text{ARMA model: } Y_t = \sum_{i=1}^p \phi_i Y_{t-i} + \varepsilon_t + \sum_{i=1}^q \theta_i \varepsilon_{t-i}$$

where ϕ_t is the autoregressive parameter at time t , ε_t is the error term at time t and θ_t is the moving-average parameter at time t [9].

The main reason of choosing ARIMA model in this study for the forecasting is because this model assumes and takes into account the non-zero autocorrelation between the successive values of the time series data.

3.3. The Error Measures for Model-Selection

There are several ways to evaluate forecasting models. The error indicators are the most used to compare how well models fit the time series. The best fit or forecasting model is one with minimal errors [9]. Forecast accuracy is measured

by the difference between actual value and the forecasted value at time period t . The error indicators considered in this paper are MPE, MAE, MASE, RMSE and MAPE defined as follows in **Table 1**.

4. Results and Discussion

Annual International tourist arrivals in Zambia from 1995 to 2014 are shown (see, **Table 2**).

The time series plot (see **Figure 1**) shows that the Zambian annual international tourist arrivals is non-stationary for $d=0$ and stationary for $d=1$.

The Holt-Winters exponential smoothing (HWES) and autoregressive integrated moving average (ARIMA) models are compared to determine the forecasting model for annual International tourist arrivals in Zambia from 1995 to 2014. HWES ($\alpha = 1, \beta = 0.1246865$) model with $a = 947,000.00, b = 39707.48$, with error measures ME = -27309.61, RMSE = 92634.37, MAE = 71,582.83, MPE = -6.240377, MAPE = 12.24099 and MASE = 0.8586324 was considered as the best fit model (see, **Table 3**).

For ARIMA model, the procedure is achieved by considering the following

Table 1. The error measures for model selection.

Criteria	Formula	Criteria	Formula
MPE	$\frac{1}{n} \sum_{i=1}^n 100 \times \frac{\varepsilon_i}{y_i}$	RMSE	$\sqrt{\frac{1}{n} \sum_{i=1}^n \varepsilon_i^2}$
MAE	$\frac{1}{n} \sum_{i=1}^n \varepsilon_i $	MAPE	$\frac{1}{n} \sum_{i=1}^n \left \frac{\varepsilon_i}{x_i} \right \times 100$
MASE	$\frac{\left \frac{\varepsilon_i}{\frac{1}{n-1} \sum_{i=1}^n Y_i - Y_{i-1} } \right }{1}$		

See [11].

Table 2. Statistics of international tourist arrivals in Zambia, 1995 to 2014.

Year	No. of Tourist Arrival	Year	No. of Tourist Arrival
1995	163,000	2005	669,000
1996	264,000	2006	757,000
1997	341,000	2007	897,000
1998	362,000	2008	812,000
1999	404,000	2009	710,000
2000	457,000	2010	815,000
2001	492,000	2011	920,000
2002	565,000	2012	859,000
2003	413,000	2013	915,000
2004	515,000	2014	947,000

Source: WTO, Yearbook of Tourism Statistics.

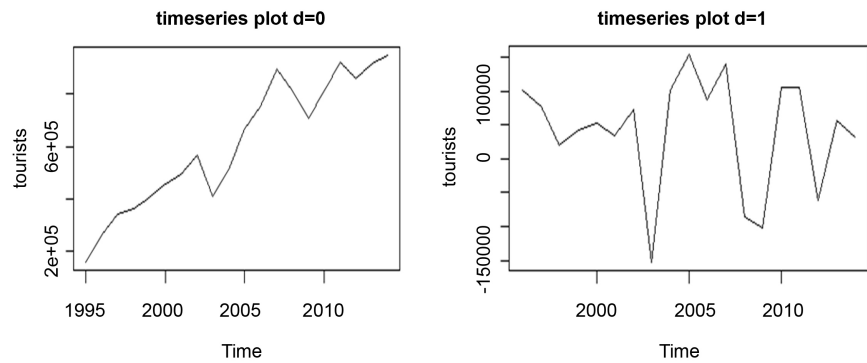


Figure 1. Time plots (original series $d = 0$ and first order difference series $d = 1$).

steps: identification, model selection, parameter estimation and diagnostic check. A code in R was used to obtain a best fit model ARIMA (0, 1, 2) model that fitted the Zambian tourist arrival data. A code in R was also used to obtain the estimated coefficients for the ARIMA (0, 1, 2) model with MA(2) = -0.2195. The ARIMA (0, 1, 2) model parameters are significant with ME = 37002.01, RMSE = 85705.65, MAE = 73770.52, MPE = 6.869655, MAPE = 13.38775 and MASE = -0.2792933. The ACF plot (original series $d = 0$) in **Figure 2** does not decay quickly indicating that the original series is non-stationary.

A code in R (version 0.99.903) was used to obtain results of residuals. The results Plots of ACF, Normal Q-Q and Histogram of Residuals for the ARIMA (0, 1, 2) analysis is shown in **Figure 3**. **Figure 3** shows that the model satisfies all required tests for a suitable model for Zambia's tourist data.

The results in **Table 3** show that HWES ($\alpha = 1$, $\beta = 0.1246865$) model performed better than the ARIMA (0, 1, 2) on tourist arrivals data for Zambia on account of smaller measures of accuracy. Hence, HWES ($\alpha = 1$, $\beta = 0.1246865$) model was selected for forecasting Zambia tourist arrivals.

5. Forecasting

Table 4 shows the forecast of Zambia tourist arrivals using the HWES ($\alpha = 1$, $\beta = 0.1246865$). Ten step forecasts up to 2024 are reported with 80% and 95% confidence limits. Forecasting results show a gradual increase in annual international tourist arrivals of about 42% by 2024 (see, **Figure 4**).

6. Conclusion

Two models of univariate time-series analysis were considered in this study: HWES and ARIMA models. The best fit of the two models used in this study was picked based on the model indicating minimum errors. The HWES ($\alpha = 1$, $\beta = 0.1246865$) showed smallest error than those of the ARIMA (0, 1, 2) models. Hence, the HWES ($\alpha = 1$, $\beta = 0.1246865$) can be used to model annual international tourist arrivals in Zambia. Forecasting results give a gradual increase in annual international tourist arrivals of about 42% by 2024 resulting in an average growth rate of 7.6% at confidence interval 95%. Accurate forecasts are key to

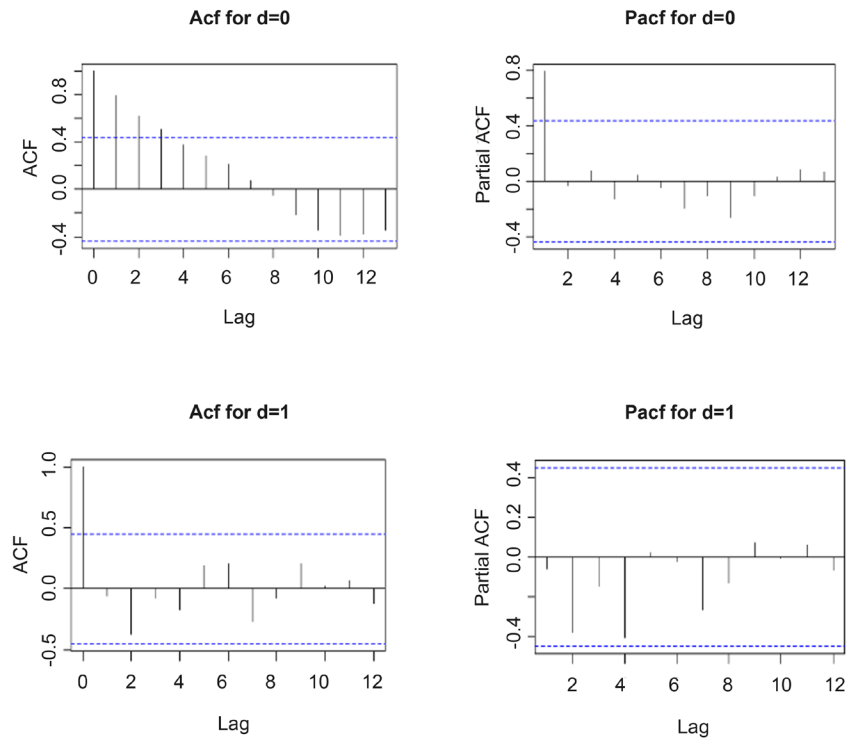


Figure 2. ACF and PACF plots (original series $d=0$ and first order difference series $d=1$).

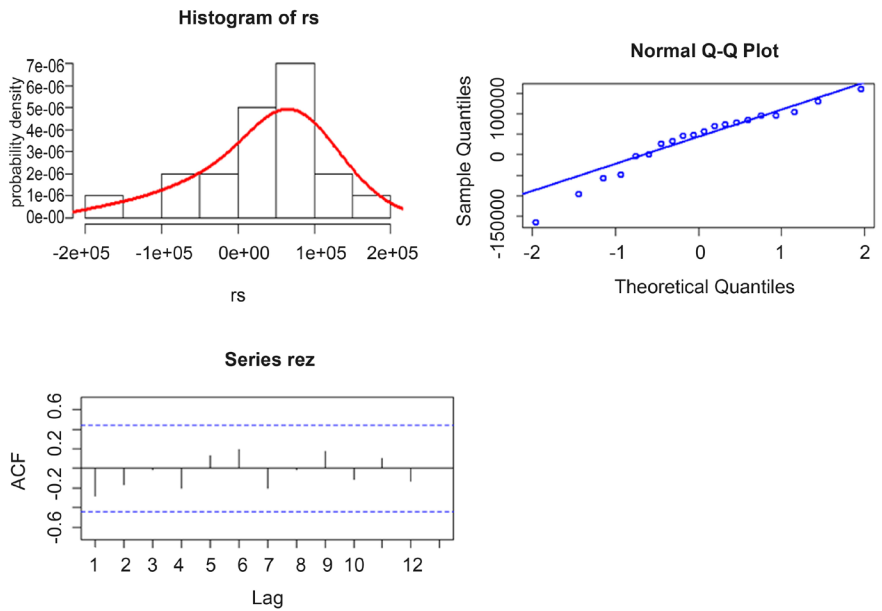


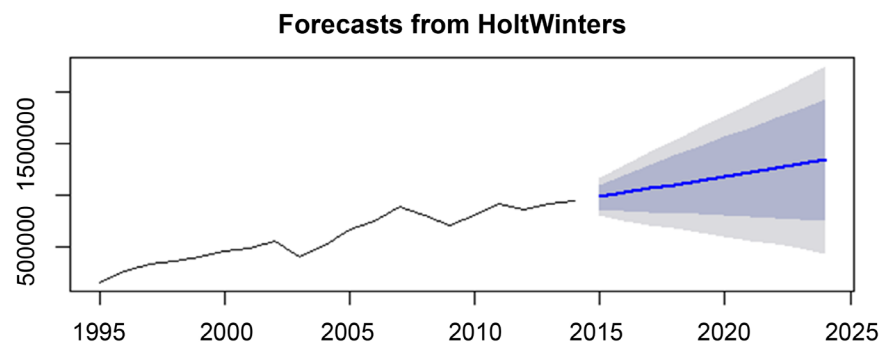
Figure 3. Plots of histogram of residuals, normal Q-Q and sample ACF for differenced series ($d=1$).

Table 3. Model diagnostics and selection.

TENTATIVE MODEL	ME	RMSE	MAE	MPE	MAPE	MASE
ARIMA (0, 1, 2)	37002.01	85705.64	73770.52	6.869655	13.38775	0.8848736
HWES ($\alpha=1, \beta=0.1246865$)	-27309.61	92634.37	71582.83	-6.240376	12.24099	0.8586324

Table 4. Forecasts and confidence bounds using HWES.

Year	Point Forecast	Low 80	High 80	Low 95	High 95
2015		986,708	1,103,436	808,187	1,165,228
2016		1,026,415	1,202,087	757,748	1,295,082
2017		1,066,122	1,294,440	716,940	1,415,305
2018		1,105,830	1,384,854	679,099	1,532,561
2019		1,145,537	1,474,871	641,865	1,649,210
2020		1,185,245	1,565,207	604,143	1,766,347
2021		1,224,952	1,656,241	565,354	1,884,551
2022		1,264,660	1,748,188	525,167	2,004,152
2023		1,304,367	1,841,177	483,388	2,125,347
2024		1,344,075	1,935,285	439,897	2,248,252

**Figure 4.** Forecasts and confidence bounds for HWES.

new investors and Policymakers. Therefore, the Zambian government should use such forecasts in formulating policies and making strategies that will promote tourism industry. Future research should go further and consider monthly and quarterly data so that seasonality models can be used. Also non-linear models such as ARCH and GARCH can be applied.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

- [1] WTO (2006) Yearbook of Tourism Statistics, Compendium of Tourism statistics and Data Files. World Tourism Organisation, Washington DC.
- [2] Akuno, A.O., Oteieno, 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. <https://doi.org/10.4236/ojs.2015.51008>
- [3] ZDA (2016) Zambia Development Agency Report. Lusaka.
- [4] AnandTularam, G., Wong, V.S.H. and AbdelhamidShobeiriNejad, S. (2012) Modeling Tourist Arrivals Using Time Series Analysis. *Journal of Mathematics and Statistics*, **3**, 348-360. <https://doi.org/10.3844/jmssp.2012.348.360>

-
- [5] Prasert, C. and Ratchanee (2008) The Times Models for Forecasting International Visitor Arrivals to Thailand. *Journal of European Economy (Special Issue)*.
- [6] Otieno, G., Mung'atu, J. and Orwa, G. (2014) Time Series Modeling of Tourist Accommodation Demand in Kenya. *Mathematical Theory and Modeling*, **4**, ISSN 2224-5804 (Paper), ISSN 2225-0522 (Online). <https://www.iiste.org/>
- [7] Gurudeo, T., Wong, V.S.H. and Nejad, S.A.S. (2012) Modeling Tourist Arrivals Using Time Series Analysis: Evidence from Australia. *Journal of Mathematics and Statistics*, **8**, 348-360. <https://doi.org/10.3844/jmssp.2012.348.360>
- [8] Purna, C.P. (2011) Forecasting International Tourists Footfalls in India: An Assortment of Competing Models. *International Journal of Business and Management*, **6**. <https://doi.org/10.5539/ijbm.v6n5p190>
- [9] Kriti, K. (2016) Forecasting Foreign Tourist Arrivals in India Using Different Time Series Models. *International Journal of Business and Management*, **15**, 38-43.
- [10] Saayman, A. and Saayman, M. (2010) Forecasting Tourist Arrivals in South Africa. *Acta Commercii*, **10**, 281-293. <https://doi.org/10.4102/ac.v10i1.141>
- [11] Tularam, G.A. and Saeed, T. (2016) Oil Price Forecasting Based on Various Univariate Time Series Models. *American Journal of Operations Research*, **6**, 226-235. <https://doi.org/10.4236/ajor.2016.63023>