

---

## Time Series Modeling on Monthly Data of Tourist Arrivals in Nepal: An Alternative Approach

Amrit Subedi

*Submitted: 18 June 2017; Accepted: 30 July 2017*

---

### ABSTRACT

**Background:** There are various approaches of modeling on time series data. Most of the studies conducted regarding time series data are based on annual trend whereas very few concerned with data having monthly fluctuation. The data of tourist arrivals is an example of time series data with monthly fluctuation which reveals that there is higher number of tourist arrivals in some months/seasons whereas others have less number. Starting from January, it makes a complete cycle in every 12 months with 3 bends indicating that it can be captured by biquadratic function.

**Objective:** To provide an alternative approach of modeling i.e. combination of Autoregressive model with polynomial (biquadratic) function on time series data with monthly/seasonal fluctuation and compare its adequacy with widely used cyclic autoregressive model i.e. AR (12).

**Materials and Methods:** This study is based on monthly data of tourist arrivals in Nepal. Firstly, usual time series model AR (12) has been adopted and an alternative approach of modeling has been attempted combining AR and biquadratic function. The first part of the model i.e. AR represents annual trend whereas biquadratic part does for monthly fluctuation.

**Results:** The fitted cyclic autoregressive model on monthly data of tourist arrivals is  $\hat{Y}_m = 3614.33 + 0.9509Y_{m-12}$ , ( $R^2=0.80$ );  $\hat{Y}_m$  indicates predicted tourist arrivals for  $m^{\text{th}}$  month and  $Y_{m-12}$  indicates observed tourist arrivals in  $(m-12)^{\text{th}}$  month and the combined model of AR and biquadratic function is  $\hat{Y}_{t(m)} = -46464.6 + 1.0007\bar{Y}_{t-1} + 52911.56m - 17177m^2 + 2043.95m^3 - 79.43m^4$ , ( $R^2=0.78$ );  $\hat{Y}_{t(m)}$  indicates predicted tourist arrivals for  $m^{\text{th}}$  month of  $t^{\text{th}}$  year and  $\bar{Y}_{t-1}$  indicates average tourist arrivals in  $(t-1)^{\text{th}}$  year. The AR model combined with polynomial function reveals normal and homoscedastic residuals more accurately compared to first one.

**Conclusion:** The use of polynomial function combined with autoregressive model can be useful for time series data having seasonal fluctuation. It can be an alternative approach for picking up a good model for such type of data.

**Keywords:** autoregressive model, biquadratic function, monthly fluctuation, seasonal fluctuation, time series modeling, tourist arrivals.

---

**Address correspondence to the author:** Central Department of Statistics, Tribhuvan University, Kirtipur, Kathmandu, Nepal.      **Email:** tirma321@gmail.com

---

## **INTRODUCTION**

Tourism can be categorized into two sorts on the basis of the country the tourist belongs to and travels to; viz. domestic and international. If a person moves within the territory of his/her country, it falls on domestic tourism and if he/she makes a travel outside of his/her country, it is international tourism. Tourism plays a vital role in economy of most of the countries in the world. World Economic Forum (2015) reported that the Nepal has been ranked 102<sup>nd</sup> in the Travel & Tourism Competitiveness Index (TTCI) out of 141 countries. Tourism is the largest industry in Nepal and its largest source of foreign exchange and revenue. Possessing eight of the ten highest mountains in the world, Nepal is a hotspot destination for mountaineers, rock climbers and people seeking adventure. The Hindu and Buddhist heritage of Nepal and its cool weather are also strong attractions (World Travel & Tourism Council, 2015). According to Nepal Tourism Statistics 2015, highest number of tourist arrived in Nepal were from India followed by Sri Lanka, United States, Thailand, United Kingdom, etc. The number of tourist arrivals was 1,679 during the year of 1962 which reached to 790,118 in 2014 and 538,970 in 2015 with some rise and fall in several years (Ministry of Culture, Tourism & Civil Aviation, 2015). Behind the existing fluctuation, there may be various reasons like political instability, natural calamities, etc. For instance, it was badly affected, at least temporarily, by the series of earthquakes in 2015. On the other hand, there is a variation in tourist arrivals month by month. This depicts that the number of tourist arrivals is highly affected by season. Generally, the number of tourist arrivals is seen higher during the season having medium weather neither too hot nor too cold such as during March-April and October-November and minimum during the rainy season. It portrays that the same kind of infrastructure and management system is not good enough for all seasons to develop tourism industry in well manner.

Being tourism a backbone for Nepalese economy; if the number of tourist arrivals for each month of a particular year can be estimated, it is hoped that it will play a significant role for proper management of tourism industry which aids benefit for the economy. Various researchers have used different types of methods and forecasting models for predicting the number of tourist arrivals. Most of them made the studies regarding annual data whereas only a few number of studies are found regarding monthly or seasonal data. For instance, Neupane, Shrestha and Upadhyaya (2010) and Chaitip and Chaibubsri (2014) are two studies regarding monthly/seasonal data whereas Gurudeo, Victor and Seyed (2012), Akuno, Otieno, Mwangi and Bichanga (2015), Shitan (2008) and Chang, McAleer and Slottje (2009) conducted studies regarding annual data. Neupane, Shrestha and Upadhyaya (2010) used ARMA (3, 1) for forecasting the tourist arrivals in Nepal using monthly data of tourist arrivals. The empirical results showed that the long run risk or volatility is persistence in monthly international tourist arrivals and estimated coefficients are statistically significant. The

volatility can be inferred as risk or uncertainty associated with international tourist arrivals in Nepalese tourism industry. Moreover, Chaitip and Chaibunsri (2014) employed the MS-VAR model and AR model to predict tourist arrivals for future of Thailand. The empirical results from this research was concluded that in high seasonal period can be used AR (2)-MLE, AR (2)-MLE-bootstrapping, and AR (1)-ME-bootstrapping to predict the number of international tourist arrivals to Thailand for future years. However, in low seasonal period only AR (1)-ME-bootstrapping can be used to predict the number of international tourist arrival to Thailand for future years. Gurudeo, Victor and Seyed (2012) adopted ARIMA (2, 2, 2) model for predicting the tourist arrivals in Australia. Akuno et al. (2015) made an attempt to forecast tourists' arrival in Kenya using statistical time series modeling techniques—Double Exponential Smoothing and the Auto-Regressive Integrated Moving Average (ARIMA). Additionally, Shitan (2008) used ARMA and ARFIMA models for forecasting tourist arrivals in Malaysia and Chang, McAleer and Slottje (2009) applied HAR (1) model for the same in Taiwan. It is clear that most of the studies had been made regarding annual data whereas only few studies concerned for monthly/seasonal data and used cyclic ARMA model for forecasting. This paper has been trying to provide an alternative way of modeling for monthly data of tourist arrivals in Nepal comparing with usual cyclic AR model which may be beneficial for modeling on the other data like tourist arrivals regarding monthly/seasonal fluctuations.

## **MATERIALS AND METHODS**

To meet the objective of the study, secondary data published by Ministry of Culture, Tourism & Civil Aviation via Nepal Tourism Statistics 2015 has been used taking monthly tourist arrivals during 1992 to 2014 AD (Ministry of Culture, Tourism & Civil Aviation, 2015). The data of 2015 has been excluded from the study for making the results more precise because of the reason that Nepal beard sudden fall in number of tourist arrivals during the year due to the series of earthquakes. The monthly data of tourist arrivals in Nepal reveal cyclic variation making a cycle for every 12 months. Holding the nature of data, cyclic autoregressive (AR) model has been adopted regressing on data of 12 months lag i.e. AR (12). Additionally, an alternative approach of modeling to capture cyclic variation has been presented and compared the result with that of previous model. For the alternative approach, polynomial (biquadratic) function and autoregressive model are combined to incorporate cyclic variation within a year and annual variation over a time period respectively. Firstly, the average number of tourist arrivals is subtracted from monthly number of tourist arrivals for each year separately to diminish the annual trend from monthly figure then the obtained values are used in biquadratic component of the model. To hold the annual trend, autoregressive model with one lag i.e. AR (1) is used taking annual averages in another part of the model. The functional forms of both the models have been shown as follows.

Let  $Y(m)$  indicates the number of tourist arrivals in  $m^{\text{th}}$  month ( $m=1, 2, 3, \dots, 12, 13, 14, \dots$ ). Then, autoregressive model with 12 lag can be given as

$$Y(m) = AR(12) = \alpha + \beta Y(m-12) + \varepsilon \tag{1}$$

where,  $\varepsilon$  is white noise.

Additionally, let,  $Y_t(m)$  indicates number of tourist arrivals in  $m^{\text{th}}$  month of particular year  $t$ . The AR model combined with biquadratic function can be written as

$$Y_t(m) = AR(12) + \beta_0 + \beta_1 m + \beta_2 m^2 + \beta_3 m^3 + \beta_4 m^4 + \varepsilon \tag{2a}$$

$$Y_t(m) = \alpha_0 + \alpha_1 \bar{Y}_{t-1} + \beta_0 + \beta_1 m + \beta_2 m^2 + \beta_3 m^3 + \beta_4 m^4 + \varepsilon \tag{2b}$$

$$Y_t(m) = \alpha^* + \alpha_1 \bar{Y}_{t-1} + \beta_1 m + \beta_2 m^2 + \beta_3 m^3 + \beta_4 m^4 + \varepsilon \tag{2c}$$

where,  $\alpha^* = \alpha_0 + \beta_0$

To test the overall significance of the model coefficients, F-test has been used, and for testing significance of individual coefficients,  $t$ -tests have been used. To check the presence of heteroscedasticity among the residuals, the residuals are plotted against unstandardized predicted values. On the other hand, to check the normality of the residuals, Kolmogorov-Smirnov (KS) statistics and Quantile - Quantile (QQ) plots are adopted. Using these various statistical tools, two models are compared.

## RESULTS

This section reveals the results from data analysis using the data of monthly tourist arrivals in Nepal. Firstly, the annual trend of tourist arrivals in Nepal has been displayed in Fig. 1. The figure shows that the tourist arrival has been increasing with some fluctuations through the years.

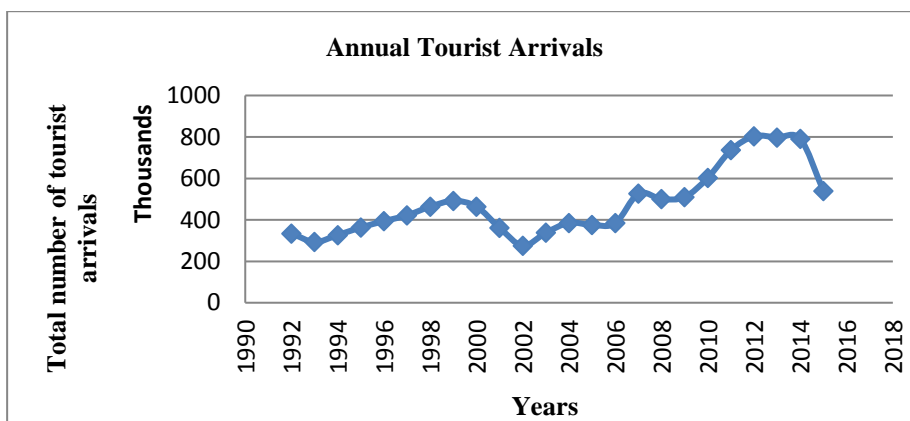


Fig. 1. Annual tourist arrivals in Nepal.

It can be seen that there is an increasing trend up to the year of 1999, then started to decline up to 2002 and again climbed up. But during the year of 2015, the tourist arrival got limited due to uncertain series of earthquakes. The minimum number of tourist i.e. 275,468 were arrived at 2002 compared to other study years whereas it met the peak at 2012 i.e. 803,092. In overall, the data reveals an increasing trend with rise and fall at various time points indicating that the data is non-stationary to some extent. The presence of non-stationarity on annual data has not been considered because of the reason that there is no such high degree of trend in one way which can be shown form correlogram as in Fig. 2.

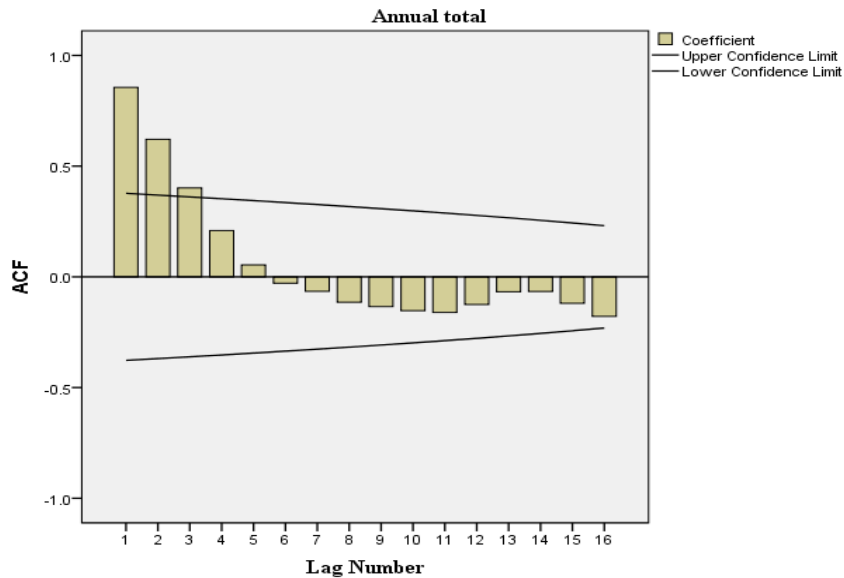


Fig. 2. Correlogram of annual data of tourist arrivals in Nepal.

Fig. 2 reveals that the value of autocorrelation function (ACF) has been fallen after 3 lags and then none of them is seen significant indicating no such high degree of autocorrelation in annual data of tourist arrivals and hence the data can be considered as stationary. And the cyclic trend on data on the basis of month has been shown in Fig. 3 taking monthly data of 2010 to 2014.

Fig. 3 shows that there is cyclic fluctuation and one cycle has been completed in each 12 months interval. In addition, the fluctuation in tourist arrivals in different months has been presented in Fig. 4 taking the average of each month separately.

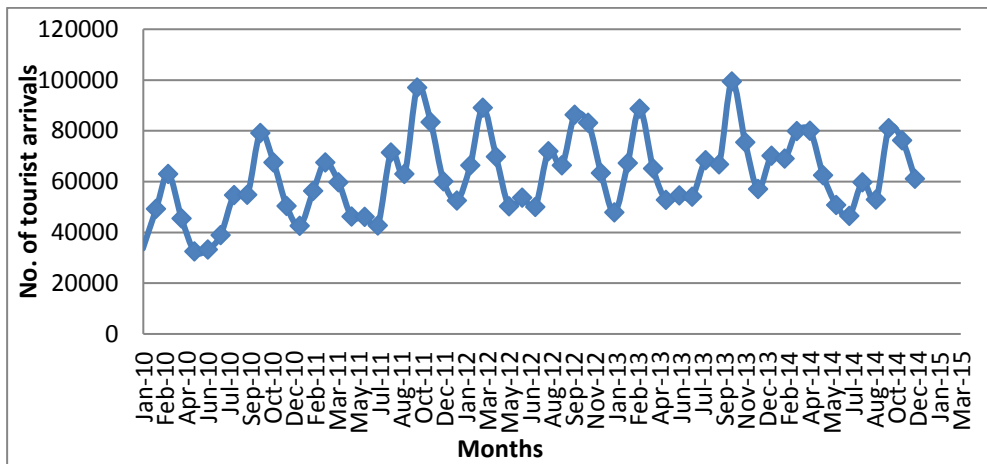


Fig. 3. Monthly tourist arrivals in Nepal (2010-2014).

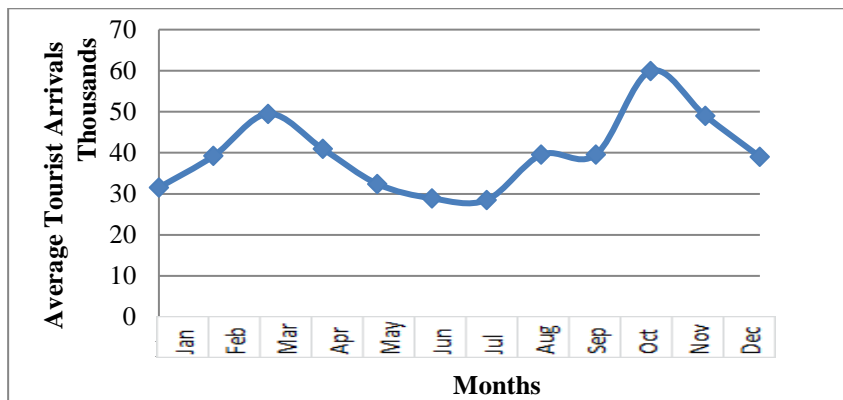


Fig. 4. Average number of monthly tourist arrivals in Nepal.

The Fig. 4 reveals that there is higher average number of tourist arrivals in March and October whereas it is less in January, June, July and December. It can be clearly seen that the graph has altogether 3 bends which indicates that the monthly tourist arrivals follows biquadratic function.

**Autoregressive model**

One of the widely used models in time series data is Autoregressive (AR) model. Since, while observing the monthly data, one cycle of fluctuation gets completed in each 12 months. It means that the value of particular month of a particular year is linked with the month of previous year indicating each value is correlated with the value of 12<sup>th</sup> lag. So, cyclic autoregressive model taking 12 lag i.e. AR (12) has been fitted. Let  $Y_m$  indicates that the number of tourist arrivals in  $m^{\text{th}}$  month from the beginning ( $m=1, 2, 3, \dots, 12, 13, \dots$ ), then the AR (12) model has been revealed as follows.

$$\hat{Y}_m = 3614.33 + 0.9509Y_{m-12}$$

$$t = (3.028) \quad (33.16)$$

$$p = (0.003) \quad (<0.001)$$

$$R^2 = 0.806 \text{ (Adjusted)}$$

$$S.E. = 7430$$

The autoregressive coefficient (0.9509) is less than 1 indicating the fitted model is wide-sense stationary. Also, it shows that the values  $Y_m$  and  $Y_{m-12}$  are highly correlated. And the regression coefficients are seen significant at 5% level of significance. The value of adjusted  $R^2$  indicates that about 81% of the variation in monthly tourist arrival has been explained by its value at 12<sup>th</sup> lag i.e. 1 year. It means that the number of tourist arrivals in a particular month of a particular year is linked with that number in same month of previous year. Additionally, the standard error of the estimate reveals that the average deviation of the observed values from the fitted line is 7430. The observed and fitted lines have been displayed in Fig. 5.

To meet the objective of the study, secondary data published by Ministry of Culture, Tourism & Civil Aviation via Nepal Tourism Statistics 2015 has been used taking monthly tourist arrivals during 1992 to 2014 AD. The data of 2015 has been excluded from the study for making the results more precise because of the reason that Nepal beard sudden fall in number of tourist arrivals during the year due to the series of earthquakes. The monthly data of tourist arrivals in Nepal reveal cyclic variation making a cycle for every 12 months. Holding the nature of data, cyclic autoregressive (AR) model has been adopted regressing on data of 12 months lag i.e. AR (12). Additionally, an alternative approach of modeling to capture cyclic variation has been presented and compared the result with that of previous model. For the alternative approach, polynomial (biquadratic) function and autoregressive model are combined to incorporate cyclic variation within a year and annual variation over a time period respectively. Firstly, the average number of tourist arrivals is subtracted from monthly number of tourist arrivals for each year separately to diminish the annual trend from monthly figure then the obtained values are used in biquadratic component of the model. To hold the annual trend, autoregressive model with one lag i.e. AR (1) is used taking annual averages in another part of the model. The functional forms of both the models have been shown in Equations 1, 2a, 2b and 2c.

To test the overall significance of the model coefficients,  $F$ -test has been used, and for testing significance of individual coefficients,  $t$ -tests have been used. To check the presence of heteroscedasticity among the residuals, the residuals are plotted against unstandardized predicted values. On the other hand, to check the normality of the residuals, Kolmogorov-Smirnov (KS) statistics and Quantile - Quantile (QQ) plots are adopted. Using these various statistical tools, two models are compared.

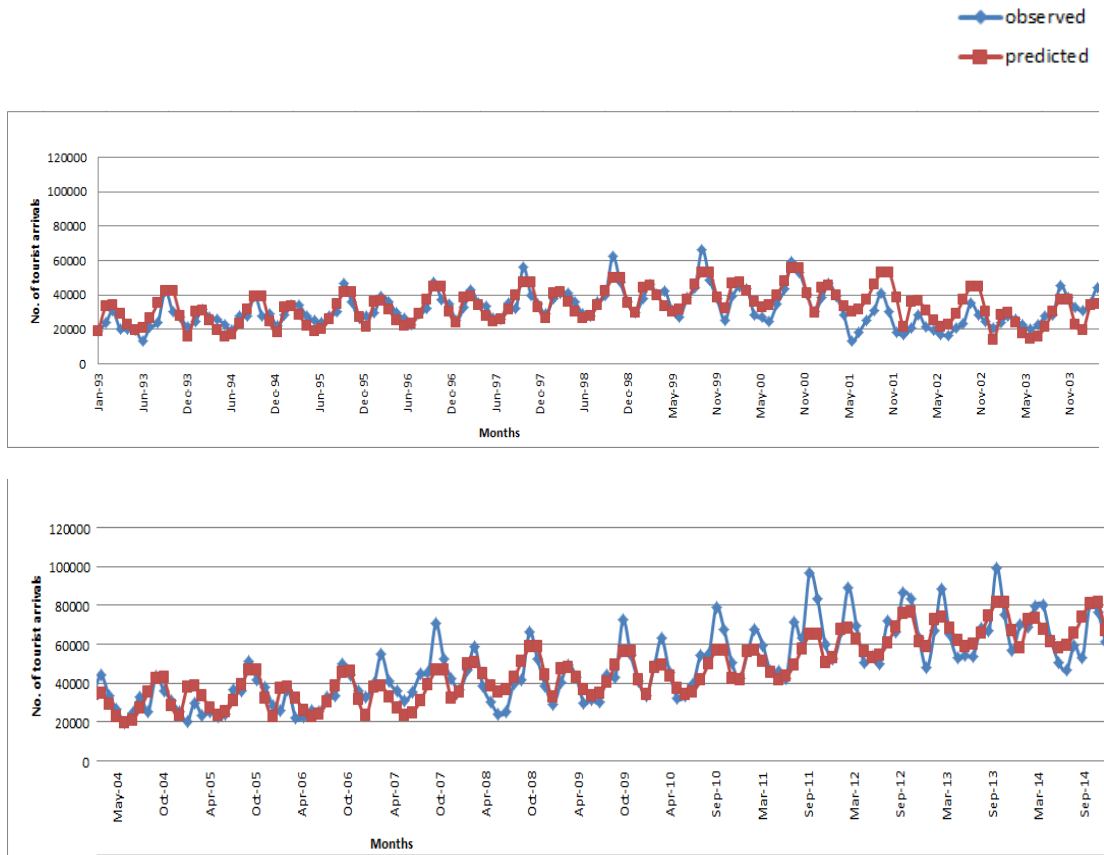


Fig. 5. Observed versus fitted lines-AR (12).

Fig. 5 depicts that the fitted model represents the observed data except in some points. To check the validity of the fitted model, assumptions about residuals have been tested as follows. The residuals are plotted against predicted values to see whether there is presence of heteroscedasticity among the residuals in Fig. 6.

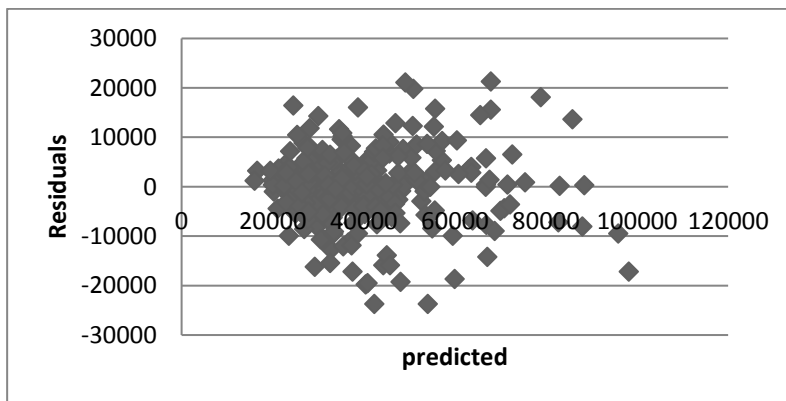


Fig. 6. Test of Homoscedasticity of residuals-AR (12).



The scatter plot displays that the residuals have higher density in the region having lower predicted values as compared to that of higher predicted values. Anyway, there is no such huge variation in variability indicating no such noticeable heteroscedasticity. Moreover, the test of normality of residuals has been made using Kolmogorov-Smirnov test which is shown in Table 1.

**Table 1.** Test of normality of residuals from AR (12)

N	264
Normal Parameters	Mean 0.0000
	Std. Deviation 7416.09
/Most Extreme Differences	Absolute 0.065
	Positive 0.065
	Negative -0.061
Kolmogorov-Smirnov Z	1.059
p-value	0.212

This result explores that the residuals are normally distributed at 5% level of significance and most extreme absolute difference is 0.065.

**Alternative Model (AR combined with Polynomial Function)**

An alternative approach of modeling in data having seasonal/monthly fluctuation has been fitted and described in this section. From the Fig. 2, a concept can be made that the biquadratic function may capture the monthly fluctuation in number of tourist arrivals in Nepal. To eliminate the annual trend from monthly figure, annual average has been deduced from the data of each month. The biquadratic function fitted for the obtained differences represents the monthly fluctuation eliminating annual trend. On the other hand, to capture the annual trend in model, Autoregressive model for 1 lag taking annual averages has been adopted. The combination of biquadratic function and AR (1) captures monthly fluctuation as well as annual trend. Let,  $Y_t(m)$  indicates number of tourist arrivals in  $m^{th}$  month of particular year  $t$ . The fitted model has been revealed as follows.

$$\hat{Y}_{t(m)} = -46464.6 + 1.0007 \cdot \bar{Y}_{t-1} + 52911.56m - 17177m^2 + 2043.95m^3 - 79.43m^4$$

t= (9.83) (25.35) (12.16) (13.34) (13.99) (14.22)

p= (<0.001) (<0.001) (<0.001) (<0.001) (<0.001) (<0.001)

R<sup>2</sup>= 0.78 (Adjusted)

S.E. = 8037

This result shows that the model coefficients are significant at 5% level of significance. The adjusted  $R^2$  indicates that about 78% of the variation in monthly tourist arrivals has been explained by average number of tourist in previous year and biquadratic function and the standard error of the estimate shows that the average deviation of the observed values from the fitted line is 8,037. The observed and fitted lines have been plotted in Fig. 7.

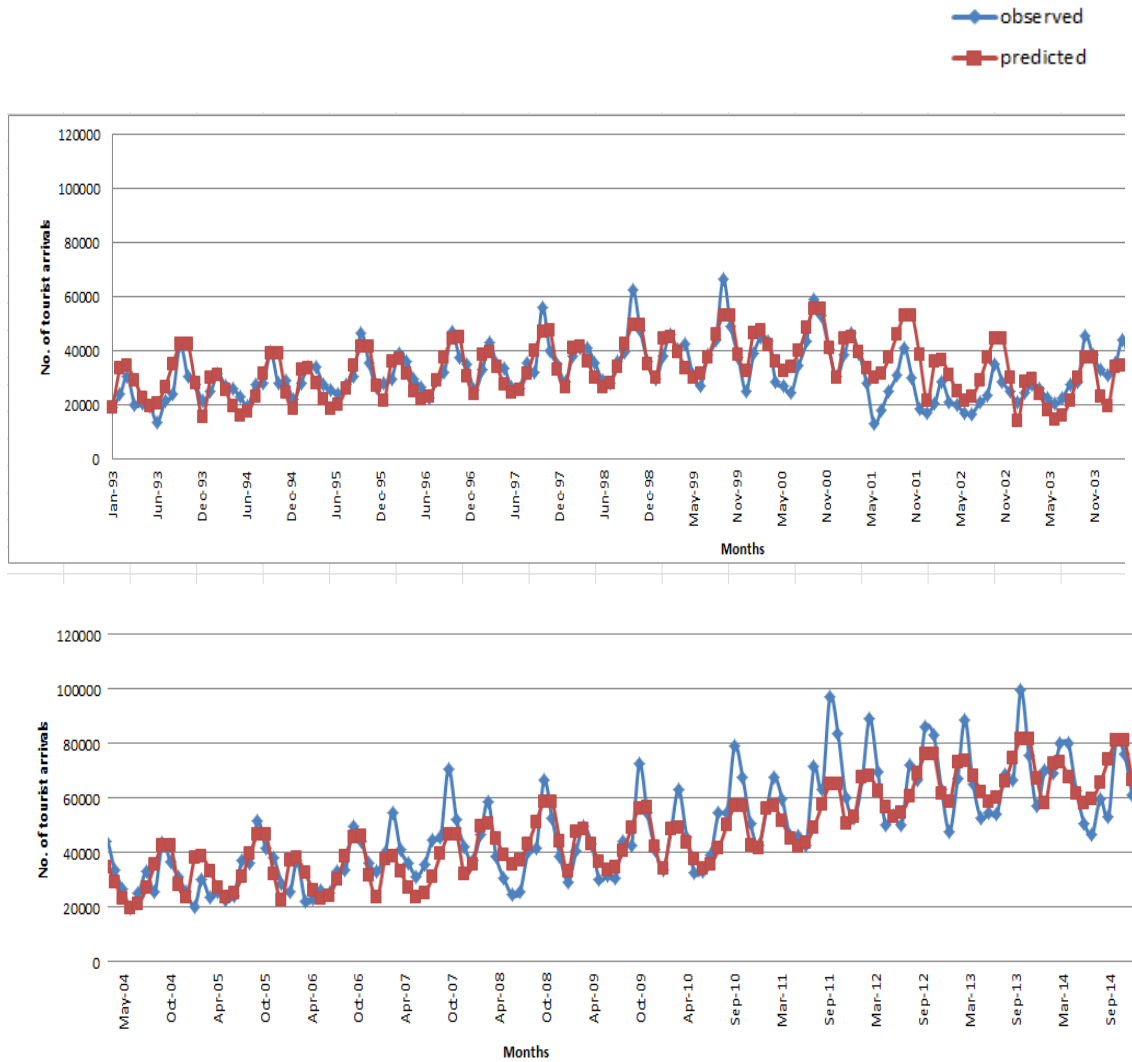


Fig. 7. Observed versus fitted lines.

The Fig. 7 reveals that the observed and fitted lines are not deviated in significant manner indicating that the fitted model represents the observed data. On the similar manner as above, the presence of heteroscedasticity is checked plotting residuals against predicted values in Fig. 8.

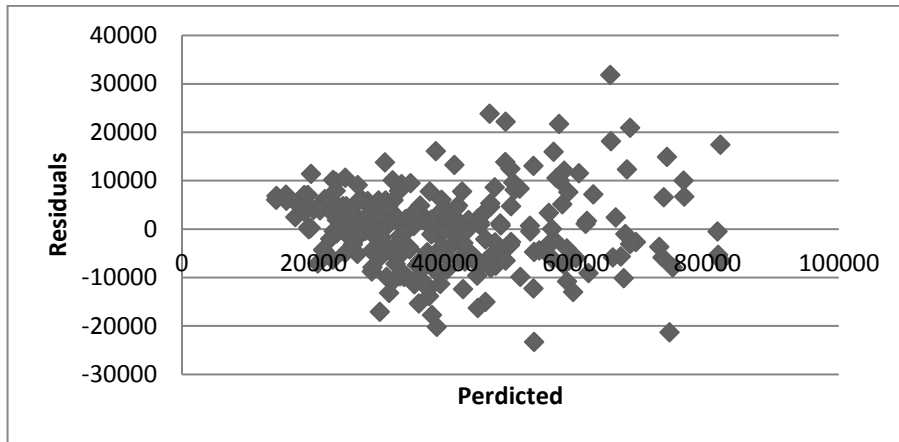


Fig. 8. Test of homoscedasticity of residuals-alternative model.

The plot of residuals versus predicted values performs that the residuals have no such huge variation in variability indicating that presence of no high degree of heteroscedasticity among the residuals. Furthermore, the normality among the residuals has been checked using Kolmogorov-Smirnov test.

Table 2: Test of normality of residuals from alternative model

N	264	
Normal Parameters	Mean	0.0000
	Std. Deviation	7960.87
Most Extreme Differences	Absolute	0.045
	Positive	0.045
	Negative	-0.034
Kolmogorov-Smirnov Z	0.732	
p-value	0.657	

The result displayed in Table 2 indicates that the residuals are normally distributed at 5% level of significance and most extreme absolute difference is 0.045. As a result, the combination of AR (1) and biquadratic function can represent the monthly fluctuation incorporating annual trend meeting all the assumptions of regression model.

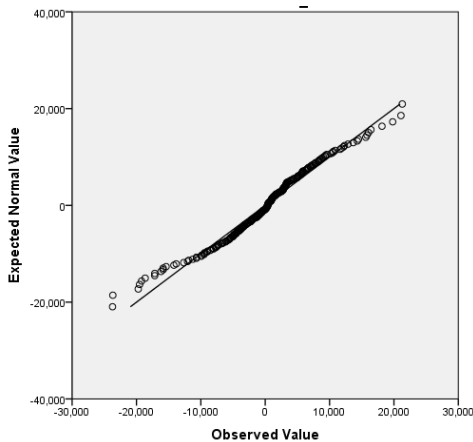
**Comparison of two models**

As it is already seen that both of the above discussed models have nearly same value of coefficient of determination ( $R^2$ ) indicating that the alternative model is not so much inferior for

explaining dependent variable as compared to the widely used model i.e. AR (12). Also, the values of standard error of the estimates are nearly same and the coefficients are seen to be significant for both the models.

While comparing the adequacy of the models regarding the assumptions about residuals, both the models reveal normal and homoscedastic residuals. More specifically, the scatter plots of residuals against predicted values displays that the model with polynomial function combined with AR deserves the assumption about homoscedasticity among the residuals more accurately compared to that of AR (12). Also, the most extreme difference is seen to be less among the residuals from alternative model compared to that from AR (12). It can also be compared using QQ-plots for normality for both the models as shown in Fig. 9.

QQ-plot of residuals from AR(12)



QQ-plot of residuals from AR model combined with biquadratic function

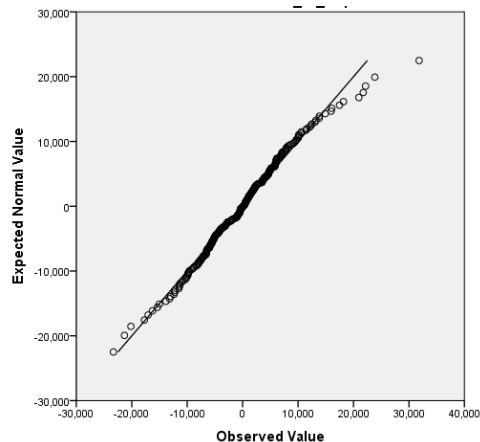


Fig. 9. Comparison of normality of residuals from two models.

The Fig. 9 shows that the plots of expected normal values of residuals are more closed to normal relatively in QQ-plot of model with polynomial function combined with AR as compared to that of AR (12) indicating that the alternative model reveals more normality among the residuals compared to that from AR (12). In addition, the plots of observed and predicted values show that both the fitted models reflect observed data and hence they can be used for forecasting. It shows that presence of some degree of serial correlation among the residuals may not lead mis-forecasting so it has been ignored. However, it does not mean that the presence of high degree serial correlation among residuals is not a serious problem. But it has not been tried in present context because of the reason that the main purpose of this paper is to provide an alternative approach of modeling in monthly time

series data. As a result, the combination of polynomial function and AR may be a suitable model for time series data having seasonal fluctuation.

## **DISCUSSION**

Most of the time series data have seasonal fluctuation. There are various approaches of modeling on time series data with such fluctuations. Various researchers have fitted different types of model on different data having various natures. The widely used models in time series data are AR, MA, ARIMA, Exponential smoothing, VAR, etc. Most of the studies conducted regarding time series data are based on only annual trend without capturing monthly fluctuations. Very few of them are conducted regarding the data with monthly/seasonal fluctuations. AR (12) is one of the models which describes annual trend with monthly fluctuation. The result of the study shows that the combination of autoregressive and biquadratic function is also capable to track annual trend with monthly fluctuation in data of tourist arrivals in Nepal. While comparing the coefficient of determination of two models, the model with polynomial function combined with AR is not seen so much inferior for explaining such scenario as compared to the AR (12). On the other hand, the model with polynomial function combined with AR is seen to be more appropriate while comparing the validity of models regarding assumptions about residuals. This all reveals that the combination of polynomial function with autoregressive model may be an alternative approach for modeling on time series data with seasonal rise and fall.

## **CONCLUSION**

To put it in a nut shell, the discussion after the result of the study makes vibrant that the use of polynomial function combined with autoregressive model may be useful on time series data having seasonal fluctuation. The use of such approach on data of monthly tourist arrivals in Nepal is only an example of such kind. As our practice, we usually compare the results of different models and choose best one for modeling data. The model described in this paper may certainly be useful and it could be the best one for certain types of data as compared to other. Hence, it is suggested that this type of approach of modeling for a time series data with seasonal fluctuation could be useful.

## **ACKNOWLEDGEMENTS**

I would like to acknowledge with deepest appreciation to Head of Department, Prof. Dr. Shankar Prasad Khanal, Prof. Dr. Srijan Lal Shrestha and Prof. Dr. Chandra Mani Paudel, Central Department of Statistics, Tribhuvan University for their kind support, encouragement and help during the course of this study. Also, my parents, brother and sisters should be thanked for their direct or indirect

support. At last but not the least, my intimates who supported me for completing the work should also be thanked.

## REFERENCES

- Akuno, A.O., Otieno, M.O., Mwangi, C.W., & 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>
- Chaitip, P., & Chaiboonsri, C. (2014). International tourists arrival to Thailand: Forecasting by non-linear model. *Procedia Economics and Finance* 14, 100 – 109.  
 doi: 10.1016/S2212-5671(14)00691-1
- Chang, C., McAleer, M., & Slottje, D. J. (2009). Modeling international tourist arrivals and volatility: An application to Taiwan. *MARCO FANNO*.  
<http://dx.doi.org/10.2139/ssrn.1355108>
- Gurudeo, A. T., Victor, S.H.W., & Seyed, A. (2012). Modeling Tourist Arrivals Using Time Series Analysis: Evidence from Australia. *Journal of Mathematics and Statistics*, 8 (3), 348-360, Retrieved from <http://thescipub.com/PDF/jmssp.2012.348.360.pdf>
- Ministry of Culture, Tourism & Civil Aviation. (2015). *Nepal Tourism Statistics 2015*. Government of Nepal, Kathmandu. Retrieved from <http://www.tourism.gov.np>
- Neupane, H. S., Shrestha, C. L., & Upadhyaya, T. P. (2012). Modelling monthly international tourist arrivals and its risk in Nepal. *NRB Economic Review*, 24(1), 28-47. Retrieved form [www.nrb.org.np](http://www.nrb.org.np)
- Shitan, M. (2008). Time series modelling of tourist arrivals to Malaysia. *InterStat*, Retrieved from <http://interstat.statjournals.net/YEAR/2008/articles/0810005.pdf>
- World Economic Forum. (2015). *The Travel & Tourism Competitiveness Report 2015*. Retrieved from [http://www3.weforum.org/docs/TT15/WEF\\_Global\\_Travel&Tourism\\_Report\\_2015.pdf](http://www3.weforum.org/docs/TT15/WEF_Global_Travel&Tourism_Report_2015.pdf)
- World Travel & Tourism Council (2015). *Travel & Tourism Economic Impact 2015 Nepal*. Retrieved from <https://www.wttc.org/-/media/files/reports/economic-impact-research/countries-2015/nepal2015.ashx>

**Reference** to this paper should be made as follows:

Subedi, A. (2017). Time series modeling on monthly data of tourist arrivals in Nepal: An alternative approach. *Nep. J. Stat.*, 1, 41-54.