ARIMA model to forecast international tourist visit in Bumthang, Bhutan

Choden and Suntaree Unhapipat
Department of Mathematics, Faculty of Science, Mahidol University, Bangkok
Email: chodenwangpo333@gmail.com

Abstract: The number of visitors to Bhutan increases each year making tourism industry the third largest revenue contributor to the nation. The country being the beneficiary of tourism it is very important to forecast tourist visit in the coming future. The monthly number of international visitors in Bumthang, Bhutan from January 2012 to December 2016 is modelled by using Box-Jenkins seasonal Autoregressive Integrated Moving Average (ARIMA) model. The best fitting ARIMA model is constructed based on Bayesian Information Criterion. The fitted seasonal ARIMA $(0, 0, 0) \times (1, 1, 0)_{12}$ is able to forecast international tourist visit of period Jan 2017-Jun 2017 with 91% accuracy. The forecasted result can be used as a tool to handle future challenges and to bring further development in tourism industry.

1. Introduction
Today tourism is one of the world’s largest and fastest growing service industry. It accounted for 10 percent of global Gross Domestic Product (GDP), 7 percent of total exports and one in 11 jobs across the world in 2015. The international tourist arrival reached nearly 1.2 billion during the same year. Asia and Pacific region was reported as one of the fastest growing tourism region in the world with 5 percent growth rate, above 4.3 percent global average [1].

Like other nations Bhutan is also beneficiary of tourism industry, it begun in 1974 under the dynamic leadership of fourth king Jigme Singye Wangchuck with just 287 tourist visit [2]. It was introduced with primary objective of generating revenue especially foreign exchange, publishing country’s unique culture and traditions to outside world and to contribute to country’s socio-economic development.

Bhutan being very farsighted has been aware that an unrestricted flow of tourists can have negative impact on its pristine environment and rich and unique culture therefore adopted a policy of “High value-Low impact” tourism thus controlling the type and quantity of tourism right from the start. The philosophy of Gross National Happiness (GNH), adventure pursuit, pristine nature and spiritual values are also the reasons for people to visit the country [3].

Tourism sector in Bhutan is regarded as one of the most exclusive travel destination based on GNH values in the world, this is because of clearly established tourism policies, excellent tourism resources, developing and expanding private sectors, qualified and experienced personnel and established marketing channels and contacts. Today it is a vibrant business with high potential for growth and further development [4]. For this reason, we focus on modelling and forecasting the number of tourists visiting Bhutan.
1.1 Literature review
In 2001, Lim and McAleer [5] used Box-Jenkins method to forecast international travel demand for Australia by employing Autoregressive Integrated Moving Average (ARIMA) model. They created various alternative models from the dataset of period 1975-1989 based on three major tourist markets namely Malaysia, Singapore and Hong Kong. The study shows how best fitted model is selected and its application in the field of forecast. The authors also considered unit root test and seasonal unit root test before creating the relevant model which is very rare in many studies done before. This test will integrate seasonal data and converts non-stationary data to stationary as most tourism data are seasonal in nature. Over this entire test will improve validity of forecast made by the fitted models and makes judgment at each step of modelling process. The recommend fitted models were ARIMA (3, 1, 1) for Hong Kong, ARIMA (3, 1, 2) for Malaysia and ARIMA (4, 1, 0) for Singapore respectively. The forecast outcome of last two models was not as accurate as the one first one.

In 2008, Lim et al [6] forecasted number of nights spent by guest in New Zealand based on dataset from 1997 to 2007 using Holt-Winter triple exponential smoothing (model A) and Box Jenkins Autoregressive (AR), Moving average (MA) and Seasonal Autoregressive Moving Average (SARMA) models (Model B). The authors later compared these two models. After doing modeling procedures using Box- Jenkins method various Autoregressive Moving Average (ARMA) and SARMA models were created. The selection of model was done based on Akaike Information Criterion (AIC) and Schwarz Bayesian Criterion (SBC) and it was found that ARMA (2,1) and SARMA (2, 0, 1) × (0, 1, 1) are the best fitted models. The study reveals that model A shows the downfall of guest nights whereas forecast of guest nights shown by model B was quite positive. The study also tells that there are lot of researches were done in forecasting international tourist arrivals using both time series and regression method but in this particular study they said regression models cannot be used as guest nights can be both international and domestic. It is not a suitable method due to given complexities of different demand characteristics.

China being one of the top source tourism market for Australia; it has great impact on country’s economy. Australia tries to maintain and promote Chinese tourist visit to their country and there are many studies done regarding to this context, among many in 2015, Ma et al [7] proposed a study to anticipate Chinese tourist visit to Australia by employing SARIMA model. They used monthly data set obtained from Australia Bureau Statistics and computations were done in STATA 13. ARIMA (0, 0, 1) × (0, 1, 1)_{12} was considered as best fitted model can forecast Chinese tourist visit three-period-ahead. The result confirms that February would be peak month every year for Chinese arrival to Australia and their visit will be seasonal for next two years. Based on the existing data, we are going to use the time series analysis.

2. Methodology

2.1 Data collection
For this study secondary data was used which was maintained by Tourism Council of Bhutan. The data consists of monthly international tourist visit in Bumthang from January 2012 to December 2016, “training dataset”. Although tourist visit to almost all twenty provinces we focus on Bumthang province because of accessibility and availability of data and its importance to tourism demand in the country. Figure 1 shows Bhutan map shaded at Bumthang.
2.2 Statistical analysis
In this study, the Box-Jenkins time series model is used to analyze international tourist visit to Bumthang. To apply Box-Jenkins model, we follow the following steps.

The first step is to identify all the tentative models. Identification consists of specifying appropriate AR, MA or ARMA and other order of the model. The identification is done by looking at Autocorrelation function (ACF) and Partial autocorrelation function (PACF) of the interested series.

The second step is to estimate the coefficients of the model. Coefficient of AR model can be estimated by least-square regression. The estimation of parameter usually requires more complicated iteration procedures but the computer programming automatically generate it.

Then the third step is to check the model. This step is also called diagnostic checking or verification. Two important elements in checking is to ensure that the residuals are white noise and estimated parameters are strictly significant.

Finally we select the best model among several competitive ARIMA models which might be suitable for the series thus we use Bayesian Information Criterion (BIC) for the model selection. The selected model is used for forecasting.

2.2.1 ARIMA model
If the process is not stationary, we have to take differencing term \((1 - B)^d\) in the process. When the model ARMA \((p, q)\) model on a time series which has been differenced \(d\) times we call this an ARIMA \((p, d, q)\) model [8]. Thus the general ARIMA \((p, d, q)\) written using backshift operator as:

\[
\phi_p(B)(1 - B)^d = \theta_0 + \theta_q(B)a_t
\]

where \(\phi_p(B) = 1 - \phi_1B - \phi_2B^2 - \cdots - \phi_pB^p\) is the stationary AR operator,

\[
\theta_q(B) = 1 - \theta_1B - \theta_2B^2 - \cdots - \theta_qB^q\]

is the invertible MA operator and \(\theta_0 = \mu(1 - \phi_1 - \phi_2 - \cdots - \phi_p)\) is deterministic trend term.

2.2.2 Seasonal ARIMA
The seasonal ARIMA model includes both seasonal and non-seasonal factors in multiplicative model called ARIMA \((p, d, q) \times (P, D, Q)_s\) where \(p\) is non-seasonal AR order, \(d\) is non-seasonal differencing, \(q\) is non-seasonal MA order, \(P\) is seasonal AR order, \(D\) is seasonal differencing, \(Q\) is seasonal MA order and \(s\) is a seasonal period [9]. The general equation of the model is

\[
\Phi_p(B^s)\phi_p(B)(1 - B)^d (1 - B^s)^D \hat{X}_t = \theta_q(B)\Theta_Q(B)^s a_t
\]

where \(\phi_p(B)\) is the non-seasonal autoregression component of order \(p\), \(\Phi_p(B^s)\) is the seasonal autoregression of order \(P\), \(X_t\) is the current value of the time series examined, \(B\) is the backward shift operator.
operator $X_t(B^d) = X_{t-d}$ is the non-seasonal difference term, $(1-B^s)^D$ is seasonal difference term, $\theta_q(B)$ is the non-seasonal moving average of order $q$ and $\Theta_Q(B)^s$ is the seasonal moving average of order $Q$ [9].

3. Result analysis
3.1 Model identification
To use Box-Jenkins methodology, the series should be stationary. In this study, the graphical methods and normality test based on Kolmogorov and Shapiro test was conducted to check the stationary of the series in SPSS version 18. The test tests the null hypothesis that the series is normally distributed. At significant level $\alpha=0.05$, we reject null hypothesis if $P$-value $< \alpha=0.05$ or accept otherwise. From Table 1 we reject null hypothesis since $p$-value is less than $\alpha=0.05$ i.e., series is not normally distributed. The graph of the series also shows strong dependence of variability on the level of series in which the data should be transformed to eliminate this dependence. To stabilize variance, we used log transformation. Figure 2 shows the plot of monthly international tourist visit in Bumthang. The monthly differenced logarithm transformed international tourist visit against time is plotted in Figure 3. The Table 2 shows the value of normality test of the logarithm transformed series where null hypothesis is accepted since $P$-value is greater than $\alpha=0.05$ thus series are normally distributed.

![Figure 2. Monthly international tourist visit in Bumthang from Jan 2012- Dec 2016.](image-url)
Table 1. Normality test of monthly international tourist visit to Bumthang.

|                         | Kolmogorov-Smirnov Statistic | df | Sig. | Shapiro-Wilk statistic | df | Sig. |
|-------------------------|------------------------------|----|------|------------------------|----|------|
| Tourist visit           | 0.183                        | 60 | 0.000| 0.840                  | 60 | 0.000|

Table 2. Normality test of monthly differenced logarithm transformed international tourist visit to Bumthang.

|                         | Kolmogorov-Smirnov Statistic | df | Sig. | Shapiro-Wilk statistic | df | Sig. |
|-------------------------|------------------------------|----|------|------------------------|----|------|
| Tourist visit           | 0.084                        | 60 | 0.200| 0.967                  | 60 | 0.104|

Figure 3. Monthly differenced logarithm transformed international tourist visit in Bumthang from Jan 2012- Dec 2016.

Figure 4. Sample ACF of monthly logarithm transformed international tourist visit in Bumthang.
Figure 4 and 5 shows sample ACF and PACF with 95% confidence limits. The ACF shows damp sine cosine wave and slow decay of the spikes indicating cyclic or seasonal movement of the correlation therefore a first seasonal differencing was done. Figure 6 and 7 shows sample ACF and PACF of monthly logarithm transformation of international tourist visit in Bumthang for first seasonal differencing at period 12 months. The sample ACF shows spike at lag 12 and PACF cuts off after lag 1. Therefore, the tentative models are ARIMA (0, 0, 0) × (0, 1, 1)_{12}, ARIMA (0, 0, 0) × (1, 1, 0)_{12}, ARIMA (0, 0, 0) × (1, 1, 1)_{12} and ARIMA (0, 0, 0) × (0, 1, 0)_{12}.

Figure 5. Sample PACF of monthly logarithm transformed international tourist visit in Bumthang.

Figure 6. Sample ACF for the first seasonal differencing at period 12 months of logarithm transformed international tourist visit to Bumthang.


3.2 Model estimation and selection

To check model adequacy, we perform Q statistic test where test tests the null hypothesis that residuals from ARIMA model have no autocorrelation. At significant level $\alpha=0.05$, we reject null hypothesis if P-value $< \alpha=0.05$ or accept otherwise. The values of the Q statistic and P-values are given in Table 3 where we accept null hypothesis since P-value of tentative models is greater than $\alpha=0.05$ i.e., residuals of ARIMA models have no autocorrelation thus the models are considered to be fitted models. To choose the best fitted model among the fitted models we consider the value of BIC, Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE), Mean Average Error (MAE) and $R^2$. The models are presented in Table 4. From the Table 4 the least value of BIC, RMSE, MAE, MAPE and highest $R^2$ is $\text{ARIMA (0, 0, 0)} \times (1, 1, 0)_{12}$ indicating $\text{ARIMA (0, 0, 0)} \times (1, 1, 0)_{12}$ as best model to forecast the monthly international tourist visit to Bumthang from January 2012 to December 2016.

Table 3. The Q statistic test for $k = 24$ of the tentative models for monthly international tourist visit to Bumthang.

| Model                         | Q statistic | P-value |
|-------------------------------|-------------|---------|
| $\text{ARIMA (0, 0, 0)} \times (0, 1, 1)_{12}$ | 14.02       | 0.67    |
| $\text{ARIMA (0, 0, 0)} \times (1, 1, 0)_{12}$ | 12.95       | 0.74    |
| $\text{ARIMA (0, 0, 0)} \times (1, 1, 1)_{12}$ | 11.01       | 0.81    |
| $\text{ARIMA (0, 0, 0)} \times (0, 1, 0)_{12}$ | 27.86       | 0.07    |

Table 4 Summary of BIC, RMSE, MAPE and $R^2$ of international tourist visit to Bumthang.

| Model                         | $R^2$ | RMSE | MAPE | MAE | BIC   |
|-------------------------------|-------|------|------|-----|-------|
| $\text{ARIMA (0, 0, 0)} \times (0, 1, 1)_{12}$ | 0.91  | 0.11 | 2.86 | 0.08| -4.35 |
| $\text{ARIMA (0, 0, 0)} \times (1, 1, 0)_{12}$ | **0.92** | **0.10** | **2.85** | **0.08** | **-4.38** |
| $\text{ARIMA (0, 0, 0)} \times (1, 1, 1)_{12}$ | 0.92  | 0.10 | 2.86 | 0.08| -4.36 |
| $\text{ARIMA (0, 0, 0)} \times (0, 1, 0)_{12}$ | 0.88  | 0.12 | 3.08 | 0.09| -4.13 |

Figure 7. Sample PACF for the first seasonal differencing at period 12 months of logarithm transformed international tourist visit to Bumthang.
3.3 Goodness of fit/diagnostic checking

In time series modelling the selection of best fit model is directly related to how well the residuals analysis is performed. One of the assumptions of ARIMA model is that for a good model the residual should be white noise.

From Figure 8 the sample ACF and PACF of the model shows that the autocorrelation of the residual are all close to zero which mean they are uncorrelated and reveal no apparent patterns, hence the residual assume mean of zero and constant variance. Finally the P-value (0.74) for the Ljung-Box statistic clearly exceeds 5% for all lag orders. Thus the selected model ARIMA $(0, 0, 0) \times (1, 1, 0)_{12}$ satisfies all the model assumptions and its parameter estimation is given in Table 5.

![Figure 8. Sample ACF and PACF of the residuals of ARIMA $(0, 0, 0) \times (1, 1, 0)_{12}$.](image)

### Table 5. Parameter estimation of appropriate ARIMA $(0, 0, 0) \times (1, 1, 0)_{12}$

| Model                     | Estimate | Standard Error (SE) |
|---------------------------|----------|---------------------|
| Constant                  | 0.03     | 0.01                |
| Seasonal differencing     | 1        |                     |
| MA seasonal lag 1         | 0.67     | 0.36                |

3.4 Forecasting using ARIMA $(0, 0, 0) \times (1, 1, 0)_{12}$

The forecasted values with 95% forecast limit of the ARIMA $(0, 0, 0) \times (1, 1, 0)_{12}$ of model for monthly international tourist visit in Bumthang are shown in Table 6 with standard error, lower and upper limit while actual forecast values are shown in Figure 9.

![Table 6. Forecasted value for 6 months monthly international tourist visited for 2017.](image)

| Date    | Tourist visited | Forecasted value | 95% confidence level | Error |
|---------|-----------------|------------------|-----------------------|-------|
|         |                 |                  | Lower | Upper |                 |
| Jan-2017| 201             | 251              | 159   | 398   | -50              |
| Feb-2017| 369             | 550              | 347   | 871   | -181             |
| Mar-2017| 779             | 1514             | 956   | 2399  | -735             |
| Apr-2017| 1796            | 1950             | 1202  | 3090  | -154             |
| May-2017| 915             | 1000             | 631   | 1585  | -85              |
| Jun-2017| 610             | 501              | 316   | 794   | 109              |
Conclusion and discussion
In this study, univariate time series models are selected by using the data of past monthly international tourist visit in Bumthang. The ACF and PACF of first seasonal differenced transformed data show seasonality at the period 12. Table 6 shows prominent forecasted values in March, April and May (spring in Bhutan) thus revealing seasonal movement of tourists in Bumthang. This is because the data collected was based on leisure tourist; it doesn’t include official and personal tourists. The tourist visits Bumthang to experience cultural activities, visit spiritual sites, to enjoy treks on high mountains etc. These activities happen at fixed scheduled date every year usually in spring and autumn where weather is basically warm and less rainfall which makes it favorable for tourist to visit the place.

References
[1] Executive summary: Bhutan tourism monitor; annual report 2016. Available from www.tourism.gov.bt/about-tcb/about-tcb-2 [Assessed on 27th May, 2017]
[2] Tourism in Bhutan: Wikipedia. Available from https://en.wikipedia.org/wiki/Tourism-in-Bhutan. [Assessed on 25th May, 2017]
[3] Tourism Council of Bhutan. Available from www.tourism.gov.bt/about-tcb/about-tcb-2 [Assessed on 28th May, 2017]
[4] Tourism Council of Bhutan. Available from www.tourism.gov.bt/about-tcb/about-tcb-2 [Assessed on 28th May, 2017]
[5] Lim C, McAleer M. Time series forecasts of international travel demand for Australia. Tourism Management. 2002;23:389-396
[6] Lim C, Chang C, McAleer M. Forecasting h(m)otel guest nights in New Zealand. International Journal of Hospitality Management. 2009:28:228-235
[7] Ma E, Liu Y, Li J, Chen S. Anticipating Chinese tourists’ arrivals in Australia: A time series analysis. Tourism Management Perspectives. 2016:17:50-58

Figure 9. Actual and forecasted values for ARIMA (0, 0, 0) × (1, 1, 0)_{12} of monthly international tourist visit to Bumthang from Jan 2012-Dec 2017.
[8] Nanthakumar L, Ibrahim Y. *Forecasting international tourism demands in Malaysia using Box – Jenkins’s Sarima application*. University Malaysia Terengganus, Malaysia Terengganus, Kuala Terengganus. 2010; 3(2): 50 - 60

[9] Wei WWS. *Time series analysis*, Addison-Wesley, USA; (1990), p106

**Acknowledgement**

We would like to thank Thai International Cooperation Agency (TICA) for providing the financial support. We would also like to thank Tourism Council of Bhutan (TCB) for providing us latest available data.