Analysis of international visitor arrivals in Bali: modeling and forecasting with seasonality and intervention

Joy Melchisedec Pierre Mangindaan1,2,a and Tipaluck Krityakierne1,2,b,*

1Department of Mathematics, Faculty of Science, Mahidol University, Bangkok 10400, Thailand
2Center of Excellence in Mathematics, Commission on Higher Education, Bangkok 10400, Thailand
*Corresponding author
E-mail: ajoymelchisedec.man@student.mahidol.ac.th, btipaluck.kri@mahidol.edu

Abstract. This work aims to develop a time series forecasting model for the number of international visitor arrivals to Bali. Tourism sector in Bali is expected to grow since it is considered as the most prepared sector in terms of facilities and infrastructure compared to other sectors. However, Bali has faced some problems related to government policies, unrests, economic and political instability, natural disasters and terrorism which obstruct the growth of the number of international visitor arrivals. Uncovering these social and environmental influences, and incorporating them into the model can lead to a better forecasting model. The data obtained from 2000 to 2016 and in 2017 were used as modeling and validation periods, respectively. The best fitted ARIMAX model is then used to forecast the visitor arrivals for the period 2018-2023. The results show that the number of international visitor arrivals in Bali will continue to grow.

1. Introduction

Since Indonesia was the colony of the Kingdom of the Netherlands in 1920s, Bali has become a popular tourist destination (Yamashita, 2004). Bali is well known for its completeness in tourism, offering its beautiful nature, amazing culinary and unique culture at once for its tourists. Tourism sector in Bali is expected to grow since it is considered as the most prepared sector in Bali (Rukini, 2015). The number of international visitor arrivals in Bali continues to increase every year. In 2016, the number of international visitor arrivals in Bali has reached 4.9 million (Bali Statistics Center, 2017).

However, Bali has faced fierce competition among other international visitor destinations in South East Asia e.g. Phuket in Thailand, Langkawi in Malaysia, Da Nang in Vietnam, Boracay in the Philippines, as well as Lombok, its sister island in Indonesia. In addition, Bali has also faced some problems related to disease outbreak, natural disasters and terrorism obstructing the growth of the number of international visitor arrivals.

Forecasting visitor arrivals with time series models has got attention of many researchers. Ruth Caroline (2012) proposed ARIMA model to forecast the number of international tourist arrivals in Bali Airport, but without considering any interventions. Sri Rezeki et al. (2013) proposed ARIMA model to forecast the number of international tourist arrivals in Bali considering Indonesian Monetary Crisis and two Bali Bombings. The results showed that Asian financial crisis and Bali bombings yield negative impacts on the number of international tourist arrivals in Bali. Wai Hong Kan Tsui et al.
(2014) employed the SARIMA and ARIMAX models to forecast air passenger throughput of Hong Kong International Airport. The results showed that the ARIMAX model performed better than the SARIMA model. In this work, we will develop a more complete time series model to forecast the number of international visitor arrivals in Bali by incorporating interventions as many as possible.

2. Methodology

2.1. Seasonal ARIMA Model

ARIMA model is a generalization of the ARMA model, which in turn is a combination of Autoregressive (AR) and Moving Average (MA) model. The non-seasonal ARIMA can be written as

\[\phi_p(B)(1 - B)^dZ_t = \alpha + \theta_q(B)\varepsilon_t\]  \hspace{1cm} (1)

where \(\phi_p(B) = 1 - \sum_{j=1}^{p} \phi_j B_j\) and \(\theta_q(B) = 1 - \sum_{j=1}^{q} \theta_j B_j\), \(\alpha\) is a constant, \(p\) is the order of non-seasonal AR term, \(d\) is the order of differencing for non-seasonal process and \(q\) is the order of non-seasonal MA term. The ARIMA model is usually denoted as ARIMA\((p, d, q)\). A more general ARIMA model, the seasonal ARIMA (SARIMA) model, can be written as

\[\phi_p(B)\phi_p(B^S)(1 - B)^d(1 - B^S)^DZ_t = \alpha + \theta_q(B)\theta_q(B^S)\varepsilon_t\]  \hspace{1cm} (2)

where \(\phi_p(B^S) = 1 - \sum_{j=1}^{P} \phi_j (B^S)_j\), \(\theta_q(B^S) = 1 - \sum_{j=1}^{Q} \theta_j (B^S)_j\), \(P\) is the order of seasonal AR term, \(D\) is the order of differencing for seasonal process, \(Q\) is the order of seasonal MA term and \(S\) is the seasonal pattern. The SARIMA model is usually denoted as SARIMA\((p, d, q) \times (P, D, Q)_S\).

2.2. ARIMA with Intervention Model

Time series observations are often affected by some unexpected interruptions that produce major changes such as government policies, unrests, economic and political instability, natural disasters and terrorism. These events are called interventions in time series. ARIMAX (ARIMA with intervention model, also known as transfer function model) extends the ARIMA model so that it can handle interventions. To represent the presence of an intervention at time \(T\) in an ARIMA model, a dummy variable is introduced. This variable is known as an intervention variable, \(I^{(T)}_t\), which is defined as

\[I^{(T)}_t = \begin{cases} 1, & t = T \\ 0, & t \neq T \end{cases}\]  \hspace{1cm} (3)

There are four types of the interventions in the time series: additive intervention, innovational intervention, level shift intervention and transient change intervention.

Additive Intervention. An additive intervention appears as surprisingly large or small in a single observation. The ARIMA model with the additive pattern is defined as

\[Z_t = \frac{\theta(B)}{\phi(B)}\varepsilon_t + \omega_A I^{(T)}_t\]  \hspace{1cm} (4)

where \(\omega_A\) is the weight for the additive intervention.

Innovational Intervention. An innovational intervention is characterized by an initial impact whose effects linger over subsequent observations. The ARIMA model with the innovational pattern is defined as
\[ Z_t = \frac{\theta(B)}{\phi(B)} \varepsilon_t + \frac{\theta(B)}{\phi(B)} \omega_t I_t^{(T)} \] (5)

where \( \omega_t \) is the weight for the innovational intervention.

**Level Shift Intervention.** A level shift intervention makes all observations after the intervention move to a new level. The ARIMA model with the level shift pattern is defined as
\[ Z_t = \frac{\theta(B)}{\phi(B)} \varepsilon_t + \frac{1}{1-B} \omega_L I_t^{(T)} \] (6)

where \( \omega_L \) is the weight for the level shift intervention.

**Transient Change Intervention.** A transient change intervention is similar to the innovational intervention but the effects decrease exponentially over subsequent observations. The ARIMA model with the transient change pattern is defined as
\[ Z_t = \frac{\theta(B)}{\phi(B)} \varepsilon_t + \frac{1}{1-\delta B} \omega_c I_t^{(T)} \] (7)

where \( \omega_c \) is the weight for the transient change intervention and \( \delta \) is the decay factor.

### 2.3. Data Description
The data used in our work is the monthly data of the number of international visitor arrivals in Bali from January 2000 to August 2017. The data were obtained from the Bali Statistics Center. Figure 1 shows the number of international visitor arrivals in Bali from January 2000 to August 2017 along with external interventions recorded, namely the 9/11 Attacks, Bali Bombing 1 in October 2002, SARS Outbreak in April and May 2003, Bali Bombing 2 in October 2005, Mount Raung Eruption in August 2015 and Mount Rinjani Eruption in November 2015. We can observe that the number of tourist arrivals suddenly dropped after these events took place.

![Figure 1. Monthly Data of the Number of International Visitor Arrivals in Bali](image)
3. Results

IBM SPSS Statistics® was used for data analysis to fit SARIMA and ARIMAX models to the data of the number of international visitor arrivals in Bali from January 2000 to December 2016. The remaining period from January 2017 to August 2017 were held out for validation.

The logarithmic transformation was applied to stabilize the time series since the plot of the time series exhibits evidence of non-stationarity. The best fit time series models for the number of international visitor arrivals in Bali were estimated based on the following criteria: R², Root Mean Square Error (RMSE) and Mean Absolute Percentage Error (MAPE).

Table 1 presents a summary of statistical results of the best ARIMAX and SARIMA forecasting models.

| Variable          | Remark                        | Coefficients | ARIMAX (0,1,1) × (0,1,1)₁₂ | SARIMA (0,1,8) × (0,1,1)₁₂ |
|-------------------|-------------------------------|--------------|-----------------------------|----------------------------|
| AR                | Lag 1                         | -            | -                           | -                          |
|                   | Lag 2                         | -            | -                           | -                          |
| MA                | Lag 1                         | 0.539***     | 0.175**                     |                            |
|                   | Lag 2                         | -            | 0.190**                     |                            |
|                   | Lag 3                         | -            | -0.022                      |                            |
|                   | Lag 4                         | -            | 0.037                       |                            |
|                   | Lag 5                         | -            | 0.058                       |                            |
|                   | Lag 6                         | -            | -0.007                      |                            |
|                   | Lag 7                         | -            | 0.004                       |                            |
|                   | Lag 8                         | -            | 0.209***                    |                            |
| Seasonal MA       | Lag 1                         | 0.724***     | 0.767***                    |                            |
| 9/11 Attacks      | Transient (ωₑ)                | -0.278***    | -                           |                            |
|                   | Decay Factor (δ)              | 0.906***     | -                           |                            |
| Bali Bombing 1    | Level Shift (ωₐₑ)             | -0.614***    | -                           |                            |
| SARS 1            | Additive (ωₐ)                 | -0.350***    | -                           |                            |
| SARS 2            | Additive (ωₐ)                 | -0.495***    | -                           |                            |
| Bali Bombing 2    | Level Shift (ωₐₑ)             | -0.626***    | -                           |                            |
| Mt. Raung Eruption| Additive (ωₐ)                 | -0.219***    | -                           |                            |
| Mt. Rinjani Eruption| Additive (ωₐ)             | -0.251***    | -                           |                            |
| R²                |                               | 98%          | 95%                         |                            |
| RMSE              |                               | 0.071        | 0.111                       |                            |
| MAPE              |                               | 0.432        | 0.650                       |                            |

Remarks: ** and *** indicate that the explanatory variable is significant at the 0.05 and 0.01 significance level, respectively.

According to the results in Table 1, SARIMA (0,1,8) × (0,1,1)₁₂ and ARIMAX (0,1,1) × (0,1,1)₁₂ models were the best fit time series models. The compact notations of these two models are given in Eqs. (8) and (9) respectively.

\[
(1 - B)(1 - B^{12})Z_t = (1 - 0.175B - 0.190B^2 + 0.022B^3 - 0.037B^4 - 0.058B^5 + 0.07B^6 - 0.004B^7 - 0.209B^8)(1 - 0.767B^{12})\epsilon_t
\]  

\[(8)\]
\[ Z_t = \frac{(1 - 0.539B)(1 - 0.724B^{12})}{(1 - B)(1 - B^{12})} \varepsilon_t - \frac{0.278}{1 - 0.906B} t^{9/11}_{\text{Attacks}} \]

Furthermore, Table 1 shows that ARIMAX model outperforms the SARIMA model as the ARIMAX model has a higher $R^2$ and lower RMSE and MAPE.

In this work, three types of intervention are used, namely additive, level shift, and transient change, in order to incorporate explicitly the effect of interventions in the ARIMAX model. SARS outbreak and two volcano eruptions are of additive type, two Bali bombings are of level shift type and 9/11 Attacks is of transient change type. Table 1 indicates that all interventions are statistically significant at 0.01 significance level.

After transforming the data back to the original scale, the two models were validated using the out-of-sample data from January 2017 to August 2017. The results of the forecasting accuracy are shown in Table 2 and in Figure 2. Observe that for the ARIMAX model, there was no error exceeding 10%. Besides, the values of Root Mean Squared Percentage Error (RMSPE) and MAPE of the ARIMAX model are lower than those of SARIMA model, indicating that the forecast performance of the ARIMAX model is better than the SARIMA model.

### Table 2. Forecasting Accuracy

| Period       | Forecasting Error (%) |
|--------------|-----------------------|
|              | ARIMAX $(0,1,1) \times (0,1,1)_{12}$ | SARIMA $(0,1,8) \times (0,1,1)_{12}$ |
| January 2017 | -9.01                  | -10.86                           |
| February 2017| -5.31                  | -5.37                            |
| March 2017   | 0.61                   | 0.19                             |
| April 2017   | -6.82                  | -8.06                            |
| May 2017     | -8.70                  | -9.22                            |
| June 2017    | -2.66                  | -4.00                            |
| July 2017    | -1.78                  | -4.72                            |
| August 2017  | -6.74                  | -11.69                           |
| RMSPE        | 0.599                  | 0.766                            |
| MAPE         | 5.179                  | 6.728                            |

The forecast results of the number of international visitor arrivals in Bali using the ARIMAX model are displayed in Figure 3. From the figure, we can see that the number of international visitor arrivals in Bali will continue to increase in the future. According to forecast results, the number will reach 6,384,241 visitors in 2018, 7,272,772 visitors in 2019, 8,284,965 visitors in 2020, 9,438,031 visitors in 2021, 10,751,575 visitors in 2022 and 12,247,932 visitors in 2023.
Figure 2. Forecasting Accuracy

Figure 3. Forecast Results of ARIMAX(0,1,1) \times (0,1,1)_{12} from September 2017 to December 2023
4. Conclusion

In this work, we model and forecast the number of international visitor arrivals in Bali using Autoregressive Integrated Moving Average with Interventions. The monthly data of the number of international visitor arrivals in Bali from January 2000 to December 2016 were used to develop the ARIMAX model, and the remaining period from January 2017 to August 2017 were used for validating the performance of the forecasting models.

The best fit ARIMAX model for the number of international visitor arrivals in Bali is ARIMAX(0,1,1) × (0,1,1)_{12}, which also outperforms the SARIMA model. The results also suggest that the interventions – 9/11 Attacks, Bali Bombing 1 in October 2002, SARS Outbreak in April and May 2003, Bali Bombing 2 in October 2005, Mount Raung Eruption in August 2015 and Mount Rinjani Eruption in November 2015 – are all statistically significant. Besides, the number of international visitor arrivals in Bali is expected to increase despite a fierce competition and undesirable events related to terrorist attacks, disease outbreak and natural disaster faced by tourism in Bali in the past. According to the forecast results, the number of international visitor arrivals in Bali will reach at least 12 million in 2023.

Acknowledgement

The first author would like to thank Ministry of Finance, Republic of Indonesia for the support through the Indonesian Endowment Fund for Education (LPDP).

References

[1] Bali Statistics Center, https://bali.bps.go.id/
[2] Caroline R 2012 The Analysis of International Tourist Arrivals in Indonesia Undergrad. Thes. (Bogor: Institut Pertanian Bogor)
[3] Montgomery D C et al 2015 Introduction to Time Series Analysis and Forecasting (New Jersey: Wiley)
[4] Rezeki S, Suhartono and Suyadi 2013 Multi input intervention model for analyzing the impact of the Asian crisis and terrorist attacks on tourist arrivals in Bali J. of Appl. Math. Sci. 7 6715–27
[5] Rukini and Arini P S 2015 Forecasting the number of international tourist arrivals in Bali in 2019: ARIMA methods J. of Appl. Quant. Econ. 8 136–141
[6] Tsui W H K, Balli H O, Gilbey A, and Gow H 2014 Forecasting of Hong Kong airport’s passenger throughput Tour. Manage. 42 62–76
[7] Wei W W S 2006 Time Series Analysis: Univariate and Multivariate Methods (New York: Pearson)
[8] Yamashita S 2004 Bali and Beyond: Explorations in the Anthropology of Tourism (New York: Berghahn Books)