Classical Forecasting of International Tourist Arrivals to Thailand

Sukanya Intarapak¹ · Thidaporn Supapakorn²* · Witchanee Vuthipongse³

Received: 10 June 2021 / Accepted: 10 March 2022 / Published online: 19 March 2022
© The Author(s) 2022

Abstract
The objectives of this work are to find the suitable forecasting model and forecasting period of the number of foreign tourists traveling to Thailand. The monthly data is gathered during January 2008 to December 2019 and is divided into two sets. The first set is the data from January 2008 to December 2018 for the modelling by the method of decomposition, Holt–Winter’s exponential smoothing method and the Box–Jenkins. The second is the monthly data in 2019 for comparing the performance of the forecasting models via the criteria of the lowest mean absolute percentage error (MAPE) and the root mean square error (RMSE). The results show that, in term of forecasting, the multiplicative decomposition is the most accurate technique for the short-term (3 months) forecasting period with the lowest MAPE and RMSE of 1.04% and 42,054.29 international tourists, respectively.

Keywords Foreign tourist · Forecasting · Box–Jenkins · Decomposition · Exponential smoothing · Mean absolute percentage error · Root mean square error

Abbreviations
MAPE Mean absolute percentage error
RMSE Root mean square error
TAT Tourism Authority of Thailand
UNWTO World Tourism Organization
CAAT Civil Aviation Authority of Thailand
ARIMA Autoregressive integrated moving average
SARIMA Seasonal autoregressive integrated moving average
ANN Artificial neural network

* Thidaporn Supapakorn
fscitdps@ku.ac.th

¹ Department of Mathematics, Faculty of Science, Srinakharinwirot University, Bangkok 10110, Thailand
² Department of Statistics, Faculty of Science, Kasetsart University, Bangkok 10900, Thailand
³ Department of Tourism, Ministry of Tourism and Sports, Bangkok 10210, Thailand
ACF  Autocorrelation function  

PACF  Partial autocorrelation function

1 Introduction

The kingdom of Thailand is located in the center of mainland Southeast Asia. Thailand shares the border with Myanmar and Laos to the north, with Laos and Cambodia to the east, with the Gulf of Thailand and Malaysia to the south and with the Andaman Sea and Myanmar to the west. There are 76 provinces and Bangkok is the capital. The popular provinces are Phuket, Krabi, Chiang Mai, Nan, Buriram, Maehongson, Ayutthaya also Bangkok. Thailand encompasses diverse ecosystems, cultures, also, religious backgrounds. This diversity enchants tourists both Thais and foreigners. The famous tourist attractions are, for example, beaches, islands, floating markets, temples, mountains, waterfalls, national park, full moon parties and civilization of the cities. According to the data from [1], Fig. 1 clearly displays the increasing trend of both the number of foreign tourists (in million people) entering to Thailand and their expenditures (in billion baht). The percentage increase of foreign tourist arrivals of 2019 to 2008 is 172.88%. This may be due to the political factors in Thailand that are likely to improve which affect investors’ confidence, the recovery of the economy of Thailand and other countries in the world, organizing an illegal tour, the low-cost airlines, the more direct flight, or even an increase in inflation rate according to oil prices.

Tourism becomes one of the significant sectors that drives the economic growth of Thailand. It can be seen that the income from the tourism industry varies with the number of foreign tourists entering Thailand. The Thai government realizes the importance of the tourism industry as evidenced by the tourism action plan of the Tourism Authority of Thailand (TAT) which has four goals [2]. One of those

![Fig. 1 The international tourist arrivals (in million people) and their expenditures (in billion baht) in 2008–2019](image-url)
is aiming to be the popular tourist destination for sustainability and still generating the highest tourism income in the world top 7. Hence, there are several strategies to build the international market, i.e., adjusting the image of the Thai tourism brand, boosting consumer spending, and expanding the market base of special interests such as medial and wellness tourism, sport Tourism, marriage celebration, wedding anniversary including traveling to visit relatives.

Due to the COVID-19 outbreak situation, countries all over the world have implemented travel restrictions to control the outbreak within the country. For the impact of the COVID-19 outbreak on international travel, the World Tourism Organization (UNWTO) has reported that in the first quarter of 2020, the world’s international tourism has declined 22% and is expected for the full year of 2020 to drop by 60% from the previous year. In Thailand, the pursuant to the Declaration of Emergency Situations in all areas of the Kingdom of Thailand starting on 26 March 2020 including the announcement of the Civil Aviation Authority of Thailand (CAAT) to temporarily ban the inbound passenger flights result in no foreign tourists traveling to Thailand since April throughout September 2020 [1, 3].

Tourism is one of the significant contributors to Thailand’s economy. Moreover, the tourist numbers help in understanding the tourism story. The purpose of this paper is to forecast the number of foreign tourists entering Thailand which is very important for the policy planning of the country’s economy. The pre-COVID-19 data of the number of tourist arrivals to Thailand is considered to analyze under normal circumstance. In order to achieve the research goal, three classical forecasting methods: decomposition, Holt–Winter’s exponential smoothing and the Box–Jenkins, are applied to get the estimates and the details will be stated in the next section.

2 Literature Review

Time series forecasting is an important topic in practical application including business and industry, government, economics, environmental sciences, medicine, social science, politics, and finance [4]. Several forecasting approaches of modeling are employed to get accurate forecasts based on previous data. In 2012, Tularam, Wong and Nejad [5] analyzed tourist arrivals to Australia using time series analysis by an autoregressive integrated moving average (ARIMA) and Vector Auto-Regression models using Australian tourist arrival data (1956–2010). The results showed that ARIMA(2,2,2) performed better in term of the prediction in 2010.

In 2013, Keerativibool [6] applied the methods of Box–Jenkins, Winters’ multiplicative exponential smoothing, decomposition, and combined forecasting to forecast the number of international tourist arrivals to Thailand. The combined forecasting method was the most powerful in terms of forecasting the next 6 monthly values based on the criterion of maximum correlation coefficient between the actual data and the forecasts followed by the decomposition, Winters’ multiplicative exponential smoothing, and Box–Jenkins, respectively. Later, in 2014, Saothayanun et al. [7] compared the methods of Box–Jenkins and Holt–Winters’ multiplicative for forecasting the number of international tourists to Thailand. Base on the criteria of the root mean square error and mean absolute
percent error of the forecasting values, the Winters’ method performed better than the Box–Jenkins.

In 2017, Subedi [8] suggested the alternative approach of modeling combining the autoregressive model with polynomial (biquadratic) function on time series data with monthly/seasonal fluctuation to forecast the number of tourist arrivals in Nepal. The trend was represented by the autoregressive part and the monthly fluctuation/seasons was done by biquadratic part. Recently, Roshan and Jahufer [9] implemented the Holt–Winters’ Method and Seasonal Autoregressive Integrated Moving Average (SARIMA) method to forecast the tourist arrivals in Sri Lanka. The results revealed that SARIMA provided the least root mean square error and mean absolute deviation.

Recently, Xie et al. [10] proposed the decomposition-ensemble approach to enhance the predictive accuracy of tourism demand forecasting of Hong Kong from nine sources: mainland China, Korea, Japan, the USA, Philippines Singapore, Australia, the UK, and Thailand. The competing models are the naïve, ARIMA and the artificial neural network (ANN) models. Hwandee and Phumchusri [11] aimed to forecast the international tourist arrivals to Thailand from 5 major countries: mainland China, Malaysia, Korea, and Russia during Jan 2013 to September 2018. In terms of the mean absolute percentage error, the seasonal autoregressive integrated moving average outperformed the multiple regression model with several important economic factors: income, price, exchange rate, and seasonal effect. Additionally, Rahman and Lee [12] presented the artificial neural network forecasting with the missing values imputation. The accuracy performance was validated by the mean absolute error and root mean square error of two forecasting methods which are SARIMA and ANN. Based on the imputation method of the decomposition, SARIMA performed better. In the other way, based on the imputation method of the spatial weighting method, ANN outperformed. Lately, in 2021, Janjua et al. [13] adopted the univariate time series forecasting method of tourist arrival based on the data from January 1991 to March 2020. The ARIMA(12, 1, 12) was suggested to forecast the international arrival from April 2020 to December 2020. The forecasts showed that Thailand will face significant negative zone arrival of international tourists due to the COVID-19 pandemic crises, which adversely affect Thailand’s economic due to the shortfall of international tourist arrivals.

From the forecasting methods mentioned above, different models have been applied and developed for tourism demand forecasting. They corporate the classical approach [5–13], the machine learning approach [11, 13], and the combination of classical approaches and/or the machine learning approaches [6, 8, 9, 11, 13]. The objective of this paper is to apply three classical forecasting methods: decomposition, Holt–Winter’s exponential smoothing and the Box–Jenkins to get the estimate of the tourism demand base on the monthly historical time series from 2008 to 2019. The remainder of this paper is organized as follows. Section 3 describes the data sets, the classical forecasting methods; decomposition, Holt–Winters’ exponential smoothing and Box–Jenkins, and the performance evaluation based on the root mean square error and mean absolute percent error. The results of the time series forecasting analysis are in Sect. 4. Section 5 is finally the conclusions and discussion.
3 Methodology

3.1 Data Sets

The monthly data of the foreign tourist arrivals to Thailand from 2008 to 2019 [1] is divided into two segments—the fitting and the forecasting. The fitting data segment from January 2008 to December 2018 is for constructing the time series modeling and displayed in Fig. 2. The forecasting segment contains the monthly data in 2019 is for forecasting and competing forecasting models.

3.2 Forecasting Methods

3.2.1 Decomposition

The classical decomposition approach [4] is to break the time series ($Y_t$) down into the component parts of trend ($T_t$) and season ($S_t$). Typically, the model of decomposition is categorized as the additive model

$$Y_t = T_t + S_t + \epsilon_t$$

and multiplicative model

$$Y_t = T_tS_t\epsilon_t$$

where $\epsilon_t$ is the random error.

The trend component is estimated using the simple linear regression model with the least squares estimates. The seasonal factor is calculated for each period in the season by detrended time series. Herein, the multiplicative model of the

![Time series plot of international tourist arrivals (in million people) from January 2008 to December 2018](image)

**Fig. 2** Time series plot of international tourist arrivals (in million people) from January 2008 to December 2018
decomposition method is considered because the amplitude of the seasonal fluctuations varies with the level of the series as displayed in Fig. 2.

### 3.2.2 Holt–Winters’ Exponential Smoothing

The forecasting equation of (Holt) Winters’ multiplicative exponential smoothing [14], an extension of Holt’s method, is

$$\hat{Y}_{t+p} = (L_t + pT_t)S_{t-s+p},$$

where, $L_t = \alpha \frac{Y_t}{S_{t-s}} + (1 - \alpha)(L_{t-1} + T_{t-1})$ is the exponentially smooth series or level estimate, $T_t = \beta(L_t - L_{t-1}) + (1 - \beta)T_{t-1}$ is the trend estimate, $S_t = \gamma \frac{Y_t}{L_t} + (1 - \gamma)S_{t-s}$ is the seasonality estimate, $\alpha$ is smooth constant for the level, $\beta$ is smooth constant for the trend estimate, $\gamma$ is smooth constant for seasonality estimate, $p$ is the period to be forecast into the future and $0 \leq \alpha, \beta, \gamma \leq 1$.

### 3.2.3 Box–Jenkins

The Box–Jenkins approach [14] uses an iterative model-building strategy that consists of selecting an initial model (model identification), estimating the model coefficients (parameter estimation), and analyzing the residuals (model checking). The model fits well if the residuals are generally small, randomly distributed and contain no useful information.

The general class of models, representing time series containing seasonal fluctuations, is a seasonal ARIMA process. It is formed by including additional seasonal terms in the ARIMA, that is, SARIMA(p,d,q) (P,D,Q)s defined as follows

$$\phi_p(B)\Phi_p(B)(1 - B)^d(1 - B^s)^D Y_t = \phi_0 + \omega_q(B)\Omega_Q(B^s)e_t$$

where $(1 - B^s)Y_t = Y_t - Y_{t-s}$, $d$ is the order of differencing, $s$ is the number of seasons per year, and $D$ is the order of seasonal differencing. The operator polynomials are

$$\phi_p(B) = (1 - \phi_1 B - \ldots - \phi_p B^p)$$

$$\omega_q(B) = (1 - \omega_1 B - \ldots - \omega_q B^q)$$

$$\Phi_p(B^s) = (1 - \Phi_1 B^s - \ldots - \Phi_p B^{sp})$$

$$\Omega_Q(B^s) = (1 - \Omega_1 B^s - \ldots - \Omega_Q B^{sq})$$

In addition, the residuals should be independent and normally distributed.
3.3 Measures of Accuracy

3.3.1 Mean Absolute Percentage Error (MAPE)

The mean absolute percentage error [14] is computed by finding the absolute error in each period, dividing by the actual observed value for that period, and then averaging these absolute percentage errors, that is, $MAPE = \frac{1}{n} \sum_{t=1}^{n} \frac{|Y_t - \hat{Y}_t|}{Y_t} \times 100$.

3.3.2 Root Mean Squared Error (RMSE)

The RMSE [14] is the square root of the mean squared error which is the sum of the squared residuals and divided by the number of the observations. The unit of RMSE is the same as the original series and RMSE is given by $RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^{n} (Y_t - \hat{Y}_t)^2}$.

4 Results

4.1 Forecasting Methods

4.1.1 Decomposition

Based on the multiplicative decomposition model, $Y_t = T_t S_t e_t$, the forecasting equation is $\hat{Y}_t = \hat{T}_t \hat{S}_t$, where $\hat{T} = 910,099 + 17,509t$; $t = 1, 2, \ldots$ where 1 represents January of 2008 and the seasonal index is shown in Table 1. The number of foreign tourists arriving to Thailand is higher than usual during December–March, especially the New Year celebration, that is, there will be the increase in the number of tourists, representing 16%, 13%, 10% and 8%, respectively.

4.1.2 Holt–Winters’ Exponential Smoothing

Examination of the data for the international tourist arrivals from January 2008 to December 2018 in Fig. 2 indicates that trend and seasonal pattern appear to exist. Winters’ three-parameter linear and seasonal exponential smoothing method decomposes the time series into three components: level, trend, and season. The multiplicative model is demonstrated with the smoothing coefficients of level, trend, and seasonality as $\hat{\alpha} = 0.762$, $\hat{\beta} = 0.001$, and $\hat{\gamma} = 0.710$, respectively.

| Table 1 | Seasonal index of international tourist arrivals to Thailand |
|---------|-------------------------------------------------------------|
| Jan     | Feb   | Mar   | Apr   | May   | Jun   | Jul   | Aug   | Sep   | Oct   | Nov   | Dec   |
| 1.13    | 1.10  | 1.08  | 0.97  | 0.87  | 0.90  | 1.00  | 1.02  | 0.85  | 0.92  | 1.00  | 1.16  |
4.1.3 Box–Jenkins

The monthly international tourist arrivals from January 2008 to December 2018 shows non-stationary pattern because of the upward trend and seasonal components as displayed in Fig. 2. After transforming the time series data with the first season differences, the time series plot in Fig. 3 shows the fluctuation around 200,000. This suggests stationary pattern and there is not necessary to make the first differences. For the model identification, the autocorrelation coefficients trail off to zero gradually as displayed in Fig. 4a, whereas the autocorrelation coefficients drop to zero after the first time lag as shown in Fig. 4b. This suggests AR \((p = 1)\). In addition, the sample partial autocorrelations have significant spike at lag 1 and 13. This behavior suggests an MA \((Q = 1)\) term at the seasonal lag 12.

The suitable model SARIMA\((1,0,0)(0,1,1)_{12}\) in the form of backward operator is
or it can be rewritten as

\[(1 - B^{12})(1 - \phi_1 B)Y_t = \theta_0 + (1 - \theta_{12} B^{12})\epsilon_t\]

with parameter estimates in Table 2. Hence, the forecasting equation is

\[\hat{Y}_t = 206,864.98 + 0.774(Y_{t-1} - Y_{t-13}) + Y_{t-12} - 0.636\epsilon_{t-12} + \epsilon_t.\]

The model adequacy is done accordingly to the plots of ACF and PACF for the residuals of SARIMA(1,0,0)(0,1,1)_{12} as shown in Fig. 5. These plots satisfy the model assumption and reveal that the residuals are independent because there is no lag of ACF and PACF falling outside the 95% confidence interval. Also, Fig. 6 reveals equal scatter of error terms.

### 4.2 Model Comparison

According to the criteria of the performance comparison on the mean absolute percentage error and the root mean square error of the three forecasting approaches: decomposition, Holt–Winters and Box–Jenkins, Table 3 shows that the multiplicative Holt–Winters model provides the least values of MAPE and RMSE.
Fig. 6  Plots of residuals vs forecast values of international tourist arrivals to Thailand

Table 3  Comparison of model performance

| Method                  | MAPE (%) | RMSE (people) |
|-------------------------|----------|---------------|
| Decomposition           | 6.31     | 146,491.24    |
| Holt–Winters            | 5.61     | 134,694.82    |
| Box–Jenkins             | 5.88     | 138,124.91    |

Fig. 7  Actual and forecast values of international tourist arrivals to Thailand in 2019
The forecasting values of the monthly international tourist arrivals to Thailand in 2019 are computed based on the forecasting models in Sect. 4.1. Figure 7 displays the comparison of the actual values and the forecasting values of the multiplicative model of decomposition and Holt–Winters and the Box–Jenkins model of SARIMA(1,0,0)(0,1,1)_12. For the performance of forecasting, Table 4 shows that the multiplicative decomposition method performs the best for short-term forecast.

### 5 Conclusions and Discussion

The demonstration of the classical forecasting techniques and algorithms which are decomposition, Holt–Winters and Box–Jenkins are successfully done using SPSS. The time series of the international tourist arrivals of the year before the Covid-19 pandemic, 2008–2019, are applied. The performances of the model are separated into model and forecasting based on the criterial of the lowest of MAPE and RMSE. For the model comparison, the multiplicative Holt–Winters’ exponential smoothing performs the best followed by the Box–Jenkins,
SARIMA$(1,0,0)(0,1,1)_t$, and the multiplicative decomposition, respectively. In the view of forecasting performance, the short-, medium-, and long-term forecasts are focused. The multiplicative decomposition is superior for all terms of forecasting followed by Holt–Winters and Box–Jenkins. This point shows that the best model does not always support the superior forecasting.

The principle of time series forecasting is that the forecast model is created by the pattern or the behavior of the past data. This is the main reason of selecting the time series before the emergence of corona virus pandemic. The limitation of this study is that the forecasting models are valid only for Thailand during the specific period of time. As a result, the number of foreign tourists has changed greatly since February 2020 [1] (see Fig. 8) and more importantly this affects the component of time series data. This severely affects tourism revenue which is the major contributor to the GDP and economics of Thailand and exactly all over the world. The government stakeholders including the private sectors should find policies or measures whether the campaign of getting the Covid-19 vaccine shots to build up the immune response, or the epidemic prevention measures, or the policy of maintaining the tourist attractions to encourage foreign tourists to return to Thailand under normal conditions.

Acknowledgements  We would like to credit the Ministry of Tourism and Sports for providing the data. In addition, the authors acknowledge the two anonymous reviewers and the editor-in-chief Professor M. Ishaq Bhatti for their valuable comments to enrich this article. However, authors take full responsibilities of any errors there may be.

Author contributions TS carried out the study conception and design, analysis, interpret the results and draft manuscript preparation: WV provided us the data. TS and SI reviewed the results. All authors approved the final version of the manuscript.

Funding This research was funding supported by the Program Management Unit for Human Resources and Institutional Development (B05F630034).

Declarations

Conflict of interest The authors declare they have no conflicts of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

1. Ministry of Tourism and Sports. Tourism statistics (2021). https://www.mots.go.th/more_news_new.php?cid=411
2. Tourism Authority of Thailand. Policy and marketing plan (2021). https://www.tat.or.th/en/about-tat/market-plan
3. Ministry of Foreign Affairs. Declaration of an emergency situation in all areas of the Kingdom of Thailand (2020). https://www.mfa.go.th/en/content/115754-declaration-of-an-emergency?cate=5f2916be298eeb02ac2eb389
4. Montgomery, D.C., Jennings, C.L., Kulahci, M.: Introduction to Time Series Analysis and Forecasting. Wiley, New Jersey (2007)
5. Tularam, G.A., Wong, V.S.H., Nejad, S.A.S.: Modeling tourist arrivals using time series analysis. J. Math. Stat. 8(3), 348–360 (2012)
6. Keerativibool, W.: Forecasting model for the number of international tourist arrivals to Thailand. Srinakharinwirot Sci. J. 29(2), 9–26 (2013)
7. Saothayanun, L., Taweesakulvatchara, S., Kanjanasakda, Y., Somrang, B.: Forecasting methods for the number of international tourists in Thailand: Box–Jenkins method and winter’s method. Thai Sci. Technol. J. 22(1), 89–98 (2014)
8. Subedi, A.: Time series modeling on monthly data of tourist arrivals in Nepal: an alternative approach. Nepal J. Stat. 1, 41–54 (2017)
9. Roshan, A.M.F., Jahufer, A.: Forecasting Sri Lankan tourist arrivals: a comparative study of Holt–Winter’s method versus ARIMA model. Proceedings of 8th International Symposium-2018, SEUSL, pp. 11–20 (2018)
10. Xie, G., Qian, Y., Wang, S.: A decomposition-ensemble approach for tourism forecasting. Ann. Tour. Res. 81(102891), 1–16 (2020)
11. Hwandee, O., Phumchusri, N.: Advances in Electronics Engineering, pp. 115–125. Springer, Singapore (2020)
12. Rahman, N.H.A., Lee, M.H.: Artificial neural network forecasting performance with missing value imputations. IAES Int. J. Artif. Intell. (IJ-AI) 9(1), 33–39 (2020)
13. Janjua, L.R., Muhammad, F., Sukjai, P., Rehman, A., Yu, Z.: Impact of COVID-19 pandemic on logistics performance, economic growth and tourism industry of Thailand: an empirical forecasting using ARIMA. Braz J Oper Prod Manag Res 18(2), 1–13 (2021)
14. Hanke, J.E., Wichern, D.: Business Forecasting. Pearson Prentice Hall, New Jersey (2014)