ENHANCED KNOWLEDGE IN SCIENCES AND TECHNOLOGY
e-ISSN: 2773-6385

EKST

Vol. 4 No. 2 (2024) 329-339
https://publisher.uthm.edu.my/periodicals/index.php/ekst

Time Series Analysis on Tourists' Arrival to Maldives After COVID-19 Pandemic

Seh Chai Nie[1], Norhaidah Mohd Asrah[1]*

[1] *Department of Mathematics and Statistics, Faculty of Applied Sciences and Technology, UTHM Kampus Cawangan Pagoh, Hab Pendidikan Tinggi Pagoh, KM 1, Jalan Panchor, 86400 Pagoh, Muar, Johor, MALAYSIA.*

*Corresponding Author: norhaida@uthm.edu.my
DOI: https://doi.org/10.30880/ekst.2024.04.02.036

Article Info

Received: 27 December 2023
Accepted: 11 January 2024
Available online: 12 December 2024

Keywords

COVID-19, Vector Autoregression, Multivariate Singular Spectrum Analysis

Abstract

Coronavirus disease 2019, also known as COVID-19, is a highly contagious respiratory infection caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). This pandemic has significantly impacted various industries, and tourism is the one of the most affected. In Maldives, tourist arrivals play a pivotal role in the country's tourism industry. The primary aim of this research is to analyze and compare the trends and patterns in tourist arrivals to the Maldives before, during, and after the COVID-19 pandemic, by employing the graphical analysis techniques. The dataset used is based on the monthly data of tourists' arrival to Maldives, from January 2000 to February 2023. The number of tourist's arrival from five countries which are Malaysia, Singapore, Thailand, Indonesia, and Philippines to Maldives are analysed using the Vector Autoregression (VAR) method and the Multivariate Singular Spectrum Analysis (MSSA). Both methods are then compared to get the best model and used for forecasting. Using the accuracy measurements as the Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and Mean Absolute Percentage Error (MAPE), the results show that the best model for Malaysia, Indonesia and Phillipines is VAR model. Meanwhile, the best model for Singapore and Thailand is MSSA model. VAR model provide better forecasting, but MSSA show conversely. Further studies should explore the combination of VAR and MSSA models. Prediction accuracy is assessed using this approach.

1. Introduction

The coronavirus disease 2019, commonly referred to as COVID-19, is a respiratory infection of considerable contagiousness caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). In December 2019, the initial instances of COVID-19 were documented in Wuhan, Hubei Province, China. The precise source of the virus is suspected to be a marketplace that traded seafood as well as live animals. The primary mode of transmission for COVID-19 is through respiratory droplets released when an infected individual coughs, sneezes, talks, or breathes heavily. [1].

The global impact of the COVID-19 pandemic has been substantial. It has led to millions of confirmed cases and fatalities, placing immense strain on healthcare systems, disrupting economies, and giving rise to social and psychological hardships [2]. The COVID-19 pandemic impacts many industries, one of the industries is the tourism industry. In recent years, the tourism industry has become one of the most important drivers of sustainable socioeconomic development [3]. There are many aspects of tourism, including travel to the general place, local transportation, accommodations, leisure, entertainment, shopping, and nourishment. In addition to leisure travel, business travel, family travel, and friend travel [4]. Tourist arrivals to the Maldives are a crucial part of the country's economy. In 2019, before the COVID-19 pandemic, the Maldives welcomed over 1.7 million tourists [5]. However, due to the pandemic, the number of tourist arrivals in 2020 dropped significantly. According to the Maldives' Ministry of Tourism, the country received 555,494 visitors in 2020, which was a 67% decrease compared to the previous year. To attract tourists to the Maldives, the government and tourism industry have implemented various promotions such as offering visa-free entry for tourists, developing world-class infrastructure, and promoting the country's natural beauty and unique culture. Maldives is also known for its luxury tourism, with many high-end resorts offering overwater villas and exclusive experiences.

The purpose of the study is to determine the best model for tourists' arrival after the COVID-19 pandemic using Multivariate Singular Spectrum Analysis (MSSA) and Vector Autoregression (VAR) model. Then, the three measures which are MAE, RMSE, and MAPE are used to compare the forecast accuracy of the predictive models.

2. Materials and Methods

2.1 Data sources and data set

In this study, the dataset chosen is the number of tourists arrival from Southeast Asia which are Malaysia, Singapore, Thailand, Philippines, and Indonesia. The dataset has been collected from Maldives Monetary Authority website while the original dataset comes from Ministry of Maldives [6]. The dataset is from January 2000 to February 2023 and divided into three parts which are before, during and after COVID-19.

The dataset was divided into two sets which are the training set and testing set before the modeling phase. As part of training the model, in-sample data was used to capture all the characteristics of the training set. Conversely, the testing set, or out-sample data, comprised reserved historical data used to assess forecast accuracy. The training set encompassed data from January 2000 to June 2022. This set was utilized to develop a model for forecasting tourist arrivals from July 2022 to June 2023, serving as the validation period to evaluate the accuracy of each model. The objective of this research is to construct an improved forecasting model for predicting tourist arrivals from five countries to the Maldives, utilizing both Multivariate Singular Spectrum Analysis (MSSA) and the Vector Autoregression (VAR) model.

2.2 Vector Autoregression (VAR) Model

An algorithm known as the Vector Autoregression (VAR) model is used in scenarios where multiple related time series affect each other [7]. The term 'Autoregressive' is indicative of the modeling approach, where each time-series variable is expressed as a function of its own past values, and lags are employed as predictors. The formula as shown in (1). Each variable in the model has its own equation. For the general example with multiple variables, the matric notation for a VAR(p) with k variables as shown in (2).

$$Y_t = C + \Phi_1 Y_{t-1} + \Phi_2 Y_{t-2} + \dots + \Phi_p Y_{t-p} + \varepsilon_t \tag{1}$$

$$\begin{bmatrix} Y_{1,t} \\ Y_{2,t} \\ \vdots \\ Y_{k,t} \end{bmatrix} = \begin{bmatrix} C_1 \\ C_2 \\ \vdots \\ C_k \end{bmatrix} + \begin{bmatrix} \Phi^1_{1,1} & \Phi^1_{1,2} & \cdots & \Phi^1_{1,k} \\ \Phi^1_{2,1} & \Phi^1_{2,1} & \cdots & \Phi^1_{2,k} \\ \vdots & \vdots & \ddots & \vdots \\ \Phi^1_{k,1} & \Phi^1_{k,2} & \cdots & \Phi^1_{k,k} \end{bmatrix} \begin{bmatrix} Y_{1,t-1} \\ Y_{2,t-1} \\ \vdots \\ Y_{k,t-1} \end{bmatrix} + \dots + \begin{bmatrix} \Phi^p_{1,1} & \Phi^p_{1,2} & \cdots & \Phi^p_{1,k} \\ \Phi^p_{2,1} & \Phi^p_{2,1} & \cdots & \Phi^p_{2,k} \\ \vdots & \vdots & \ddots & \vdots \\ \Phi^p_{k,1} & \Phi^p_{k,2} & \cdots & \Phi^p_{k,k} \end{bmatrix} \begin{bmatrix} Y_{1,t-p} \\ Y_{2,t-p} \\ \vdots \\ Y_{k,t-p} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \\ \vdots \\ \varepsilon_{k,t} \end{bmatrix} \tag{2}$$

where Y_t is endogenous variables at time t, C is constant terms, Φ is coefficient matrices corresponding to lags 1 through p and ε_t is error term.

2.2.1 Augmented Dickey-Fuller (ADF) test

The Augmented Dickey-Fuller (ADF) test is a statistical test employed to detect the presence of unit roots in a time series dataset, which indicates non-stationarity [8]. It is widely used to assess the stationarity of a time series prior to apply VAR modelling. The ADF test helps in determining if differencing is necessary to achieve stationarity. When conducting the ADF test, if the calculated test statistic is lower than the critical value, it leads to the rejection of the null hypothesis, which indicating that the time series is stationary. On the other hand, if the test statistic exceeds the critical value, the null hypothesis cannot be rejected, implying the presence of non-stationarity in the time series [9]. The formula as shown in (3).

$$y_t' = \phi y_{t-1} + \sum_{i=1}^{k} \beta_i y_{t-i} \tag{3}$$

where y_t' is first differenced series, ϕ is coefficient of ADF, k is the number of lags to include in the regression and β_i is parameter of differenced series.

2.2.2 Kwiatkowski–Phillips–Schmidt–Shin (KPSS) test

The Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test is a statistical test used to check the stationarity of a time series. It is commonly employed to complement other stationarity tests like the Augmented Dickey-Fuller (ADF) test [10]. While the ADF test is designed to detect unit roots indicative of non-stationarity, the KPSS test focuses on the presence of a deterministic trend in the data. The formula as shown in (4).

$$\text{KPSS} = \frac{\text{T(SSE)}}{\sigma^2} \tag{4}$$

where T is number of observations in the time series, SSE is the sum of squared deviations from the estimated trend, σ^2 is an estimate of the long-run variance of the time series.

2.2.3 Model selection

Model selection in Vector Autoregression (VAR) involves choosing the appropriate order of the model, which indicates the number of lags to include in the system. Selecting the right order is crucial for obtaining accurate and reliable results. During this step, several criteria are commonly used to assess and compare the quality of different models. The multiple information criteria mentioned in this study are Akaike Information Criterion (AIC), Hannan-Quinn Criterion (HQ), Schwarz Criterion (SC), and Final Prediction Error (FPE). The model with the lowest value for these criteria is often considered the best-fitting model. Lower values indicate a better balance between model complexity and goodness of fit.

2.2.4 Diagnosis checking

To determine the suitability and validity of the VAR model, diagnostics checks must be made after model selection. These checks entail examining the model's residuals (errors) to make sure they adhere to a set of presumptions. The procedure needs to be repeated to reevaluate and improve the model if it does not meet the presumptions. The iterative process may continue until a good model is found. Typically, the assumption of a white noise process for residuals implies that they exhibit certain characteristics, which residual is not correlated with each other, the mean of the residual is zero, the variance of residuals is constant, and the residuals follow a normal distribution [11].

The Ljung-Box (LB) test is a commonly employed method to determine whether autocorrelations exist in the residuals of fitted time series models. The test evaluates the null hypothesis that the residuals are uncorrelated, while the alternative hypothesis suggests that there are correlations present in the residuals. The formula as shown in (5).

$$Q = n(n-2)\sum_{j=1}^{h}\frac{\rho_j^2}{n-j} \tag{5}$$

where n is number of observations, ρ_j^2 is estimated autocorrelation coefficient, j is the j^{th} lags and h is the number of lags.

2.3 Multivariate Singular Spectrum Analysis (MSSA)

Multivariate Singular Spectrum Analysis (MSSA) is a powerful signal processing technique used for time series analysis and forecasting. The key idea behind MSSA is to decompose a time series into a finite number of components, known as eigenvectors or principal components, which represent different patterns in the data. These components are extracted from the original time series using a sliding window technique, where the time series is divided into sub-series or segments. Once the components are obtained, they can be analysed individually to identify patterns and trends in the data. These patterns are used to forecast future values of the time series, detect anomalies or outliers, and identify underlying factors that contribute to the variation in the data. One of the advantages of MSSA is it enables to handle non-stationary and non-linear time series data [12]. It can also manage missing or incomplete data, making it useful in situations where data may be sparse or irregular.

2.3.1 Embedding

In the embedding step, time series data is transformed into a matrix called a trajectory matrix. This matrix is constructed by taking lagged vectors of the time series, and it serves as the foundation for the subsequent steps in the analysis. In this step, one-dimensional series is described as a multidimensional series. The trajectory matrix has dimensions which the integer value L is set to the window length ($2 \leq L \leq N/2$) while represents the number of columns of the trajectory matrix. The trajectory matrix is a Hankel matrix, which means that all the elements along the diagonal $i + j$ are constant. The Hankel matrix is formed as in (6).

$$T_x = (T_{i,j})_{L\times K} = \begin{bmatrix} x_1 & x_2 & \cdots & x_K \\ x_2 & x_3 & \cdots & x_{K+1} \\ \vdots & \vdots & \ddots & \vdots \\ x_L & x_{L+1} & \cdots & x_N \end{bmatrix} \quad (6)$$

2.3.2 Singular Value Decomposition (SVD)

The trajectory matrix, T_x obtained from the embedding step is decomposed using Singular Value Decomposition (SVD). The SVD can be represented as $S = T_x {T_x}^T$. A set of S eigenvalues in decreasing order of magnitude and $U_1, \dots, U_L$ the eigenvectors orthonormal system of matrix S which corresponds to the eigenvalues. The SVD trajectory matrix of T_{xi} can be written as (7),

$$T_{xi} = T_{x1} + T_{x2} + \cdots + T_{xd}, T_{xi} = \sum_{i=1}^{d} \sqrt{\lambda_i} U_i V_i^T \quad (7)$$

where $\sqrt{\lambda_i}$ is singular value, U_i is eigenvector, $V_i^T$is the principal component which is $\frac{T_x^T U_i}{\sqrt{\lambda_i}}$, and ${T_x}^T$ is the transpose of the trajectory matrix.

2.3.3 Grouping

The grouping step in MSSA involves dividing the elementary matrices into multiple groups and aggregating the matrices within each group. The grouping process is done by grouping index $\{1,2, \dots, d\}$ sets into m subset that can be represented by $I = I_1, I_2, \dots, I_m$. Then, a matrix can be formed according to the SVD for the trajectory matrix T_{xi} as (8),

$$T_{Ix} = T_{I1} + T_{I2} + \cdots + T_{Im} \quad (8)$$

2.3.4 Diagonal averaging

The primary goal of diagonal averaging is to simplify and highlight the structure of the singular vectors obtained through the Singular Value Decomposition (SVD) of the trajectory matrix. Let $\mathbf{Y} = L \times K$; $y_{ij}, 1 \le i \le L, 1 \le j \le K$. The $\mathbf{Y}$ matrix is then converted to a time series $g_0, \dots, g_{N-1}$, as a result of the diagonal averaging which shown as (9).

$$g_k = \begin{cases} \frac{1}{k+1} \sum_{m=1}^{k+1} y^*_{m,k-m+2} \, ; 0 \le k \le L^* - 1 \\ \frac{1}{L^*} \sum_{m=1}^{L^*} y^*_{m,k-m+2} \, ; L^* - 1 \le k \le K^* \\ \frac{1}{N-k} \sum_{m=k-K^*+2}^{N-K^*+1} y^*_{m,k-m+2} \, ; K^* \le k \le N \end{cases} \quad (9)$$

wher***e*** L^* is min (L, K), K^* is max (L, K), g_k is the average of the matrix elements along the diagonal $i + j = k + 2$.

2.4 Measure of the Forecast Accuracy

Three measures employed to assess and compare the accuracy of the predictive models. The forecast accuracy of the models evaluated using the Root Mean Squared Error (RMSE), Mean Absolute Percentage Error (MAPE), and Mean Absolute Error (MAE). These measures provide quantitative measures to gauge the performance of the models and enable a comprehensive comparison of their forecasting capabilities. The equation of RMSE, MAPE and MAE as shown in (10), (11) and (12).

$$\mathrm{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)} \quad (10)$$

$$\mathrm{MAPE} = \frac{1}{n} \sum_{i=1}^{n} \frac{|y_i - \hat{y}_i|}{y_i} \times 100\% \quad (11)$$

Enhanced Knowledge in Sciences and Technology Vol. 4 No. 2 (2024) p. 329-339

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|y_i - \hat{y}_i\right| \qquad (12)$$

where n is number of observations, y_i is i^{th} observed and $\hat{y}_i$ is i^{th} forecasted value.

3. Results and Discussions

3.1 Time Series Plots

The times series plots of the number of tourist arrivals from five countries which are Malaysia, Singapore, Thailand, Indonesia, and Singapore to Maldives are shown in Fig. 1 – Fig. 5.

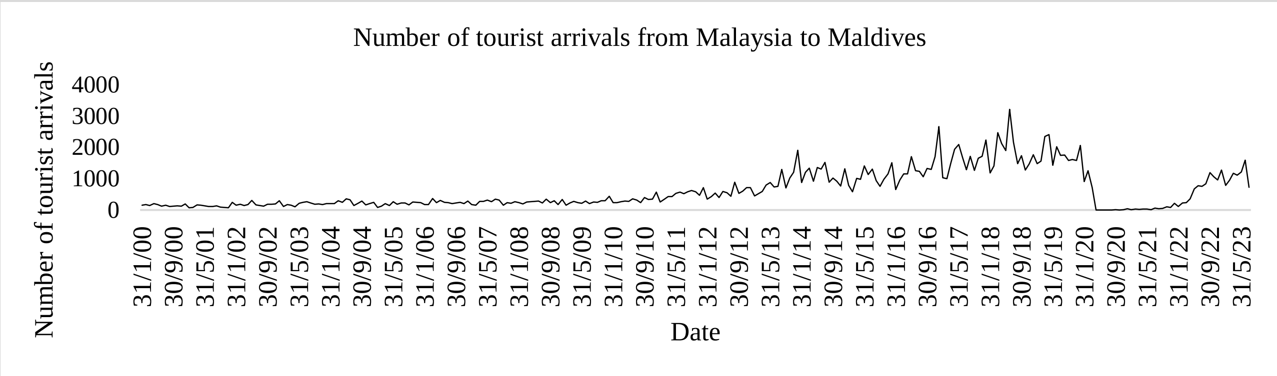

Fig. 1 Time series plot of number of tourist arrivals from Malaysia to Maldives

Based on the Fig. 1, the number of tourist arrivals from Malaysia to Maldives is an increasing trend pattern from January 2000 to December 2019. During the COVID-19 pandemic, there are an abrupt decline observed in Malaysia from January 2019 to 2020. After COVID-19 pandemic, Maldives imposing protective measure (SOP) and reopened for tourism in July 2020. The numbers of the tourist arrivals from Malaysia are gradually rising again from January 2022 to 2023.

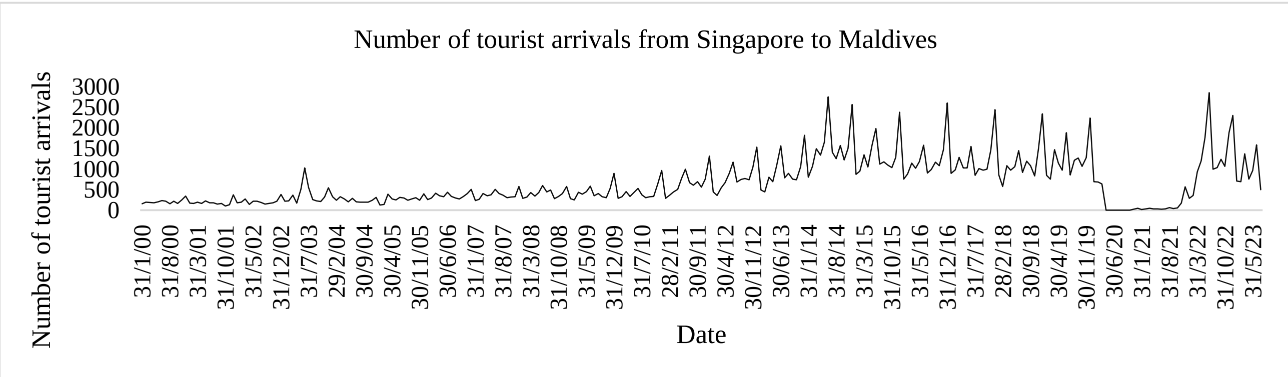

Fig. 2 Time series plot of number of tourist arrivals from Singapore to Maldives

According to Fig. 2, the trend in tourist arrivals from Singapore to Maldives displays a consistent increase from January 2000 to December 2019. However, a sharp decline is noticeable during the COVID-19 pandemic, spanning from December 2019 to February 2020. It opened for tourism again in July 2020 after Maldives implemented Standard Operating Procedures (SOP) as protective measures. Subsequently, the influx of tourists from Singapore has been steadily recovering, showing a gradual rise from January 2022 to 2023.

Based on the information presented in Fig. 3, the pattern of tourist arrivals from Thailand to Maldives exhibits a continuous growth from January 2000 to December 2019. Nevertheless, a significant decrease is evident during the COVID-19 pandemic, spanning from December 2019 to February 2020. In July 2020, the Maldives officially resumed tourism after introducing Standard Operating Procedures (SOPs) as protective measures. Since then, the number of tourists arriving from Thailand has been gradually rebounding, indicating a steady increase from January 2022 to 2023.

Fig. 3 Time series plot of number of tourist arrivals from Thailand to Maldives

Fig. 4 Time series plot of number of tourist arrivals from Indonesia to Maldives

According to the plot provided in Fig. 4, the trend in tourist arrivals from Indonesia to Maldives demonstrates consistent growth from January 2000 to December 2019. However, a notable decline is observed during the COVID-19 pandemic, extending from December 2019 to September 2020. The Maldives developed Standard Operating Procedures (SOP) as protective measures following the pandemic, and tourism officially reopened there in July 2020. Subsequently, the number of tourists arriving from Indonesia has been gradually recovering, showcasing a steady increase from September 2020 to 2023.

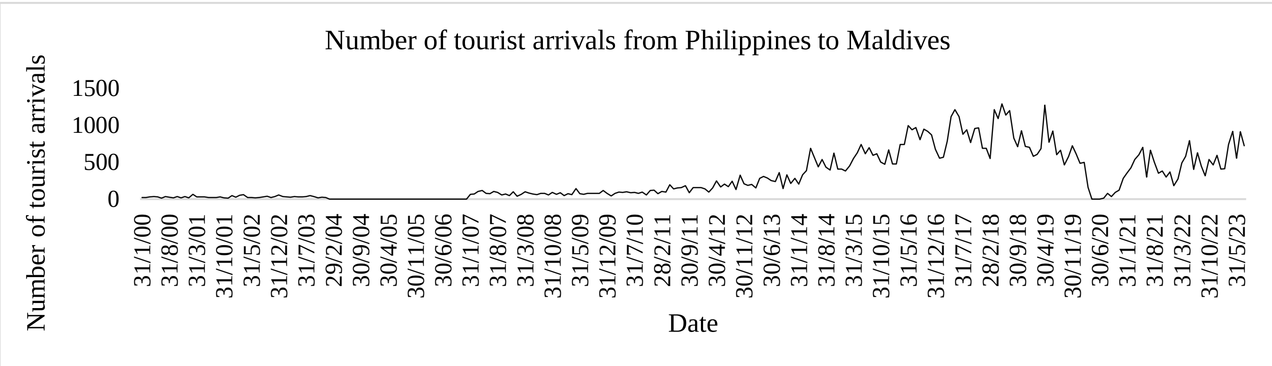

Fig. 5 Time series plot of number of tourist arrivals from Philippines to Maldives

Based on the depicted trends in Fig. 5, the number of tourists from Philippines to Maldives displays an increasing pattern from January 2000 to 2004. Subsequently, there is a period of stability observed from January 2004 to 2006, followed by another increase from January 2007 to 2019. However, a significant decline is noted during the COVID-19 pandemic, spanning from 2019 to 2020. Following the pandemic, Maldives implemented protective measures (SOP) and officially reopened for tourism in July 2020. Consequently, the number of tourist arrivals from the Philippines has been gradually recovering, indicating a rising trend from December 2020 to 2023.

3.2 Vector Autoregression (VAR) Model

In this study, statistical testing was used to assess the stationarity of the datasets from five different countries. As part of the preliminary analysis within the statistical testing phase, both the ADF test and the Kwiatkowski–

Phillips–Schmidt–Shin (KPSS) test was executed. The statistical software R Studio was utilized for the implementation of these tests. The outcomes of the statistical test are summarized in Table 1.

Table 1 : ADF and KPSS testing results.

Statistical Test / Country	Augmented Dickey Fuller (ADF)	Kwiatkowski–Phillips–Schmidt–Shin (KPSS)
Malaysia	0.6793	0.01
Singapore	0.2859	0.01
Thailand	0.4331	0.01
Indonesia	0.5642	0.01
Philippines	0.4470	0.01

According to the findings presented in Table 1, the *p*-value obtained from the ADF tests for the five distinct countries exceed the significance level of 0.05. Consequently, the null hypothesis is accepted. This implies that the data exhibits a unit root and is not stationary at a statistically significant level. In contrast, the *p*-value derived from the KPSS tests for all countries are below the significance threshold of 0.05. Hence, the null hypothesis which indicates that the data is stationary is not accepted. These results show that the data is not stationary when KPSS test is carried out.

The ADF and KPSS test are used again after the differencing process. The ADF tests for five different countries yielded all *p*-values of 0.01, below the traditional significance level of 0.05. Therefore, the null hypothesis is rejected, indicating that the data exhibit statistically significant stationarity. In contrast, the KPSS test for all countries yielded *p*-values of 0.010 which exceeded the significance threshold of 0.05. Therefore, acceptance of the null hypothesis implies that the data remains stationary when performing the KPSS test.

3.2.1 Model selection

The *p*-value from both tests consistently indicate the achievement of stationarity through differencing. The lag order selection results are shown in Table 2. Based on the Table 2, the determination of the suitable lag order for the Vector Autoregression (VAR) model involved the utilization of multiple information criteria, including the Akaike Information Criterion (AIC), Hannan-Quinn Criterion (HQ), Schwarz Criterion (SC), and Final Prediction Error (FPE). These criteria take into account both the model's goodness of fit and its complexity. Consequently, the lowest values observed across these information criteria signify the optimal balance between model fitness and simplicity. Table 2 shows the most suitable lag order for the VAR model is determined as 3, as indicated by the SC. Examining the information criteria across varying lag orders (ranging from 1 to 5), a consistent decrease in the values of AIC, HQ, and FPE is observed with an increase in lag order. Notably, the SC values follow a similar pattern, decreasing until the lag order reaches 3 and showing a subtle uptick for higher orders. Although other criteria show that lag order 5 has the lowest value, 3 also has the lowest value. Thus, choose the lowest lag order, 3 is the most suitable lag order.

Table 2 : Lag order selection results

Lag / Multiple information criteria	1	2	**3**	4	5
AIC	51.6368	51.0981	50.6427	50.5081	**50.039**
HQ	51.8001	51.3975	51.0817	51.0796	**50.7466**
SC	52.0432	51.8431	**51.7264**	51.9304	51.7999
FPE	2.66e+22	1.56e+22	9.87e+21	8.63e+21	**5.41e+21**

*Bold indicating the best results.

The VAR model is a statistical method used to analyse and forecast the relationships between the countries within the dataset over time. The coefficients in each equation represent how the current value of each country is influenced by its own past values and the past values of the other country. The estimated coefficients are essential for understanding the dynamic relationships between these variables in the VAR model. Table 3 represents the equations for each country based on the VAR model.

Table 3 : VAR equations for each station

Country	VAR Equation
$\text{Malaysia}_{(t)}$	$\text{Malaysia}_{(t)}=2.273- 0.4964\text{Malaysia}_{(t-1)} - 0.0970\text{Singapore}_{(t-1)}+ 0.0291\text{Thailand}_{(t-1)} - 0.4686\text{Indonesia}_{(t-1)}+0.3922\text{Philippines}_{(t-1)}- 0.3961\text{Malaysia}_{(t-2)} - 0.0991\text{Singapore}_{(t-2)} + 0.1413\text{Thailand}_{(t-2)}- 0.5519\text{Indonesia}_{(t-2)}+ 0.5504\text{Philippines}_{(t-2)}- 0.0299\text{Malaysia}_{(t-3)} + 0.0137\text{Singapore}_{(t-3)}+ 0.1401\text{Thailand}_{(t-3)}- 0.9892\text{Indonesia}_{(t-3)}- 0.0288\text{Philippines}_{(t-3)}$
$\text{Singapore}_{(t)}$	$\text{Singapore}_{(t)}= 10.0656- 0.1136\text{Malaysia}_{(t-1)}- 0.4393\text{Singapore}_{(t-1)}+ 0.0417\text{Thailand}_{(t-1)} - 0.4614\text{Indonesia}_{(t-1)}+0.6931\text{Philippines}_{(t-1)}- 0.1978\text{Malaysia}_{(t-2)} - 0.3605\text{Singapore}_{(t-2)} + 0.3471\text{Thailand}_{(t-2)}- 0.4263\text{Indonesia}_{(t-2)}+ 0.8230\text{Philippines}_{(t-2)}- 0.2311\text{Malaysia}_{(t-3)} + 0.1589\text{Singapore}_{(t-3)}+ 0.2034\text{Thailand}_{(t-3)}- 0.1527\text{Indonesia}_{(t-3)}- 0.4471\text{Philippines}_{(t-3)}$
$\text{Thailand}_{(t)}$	$\text{Thailand}_{(t)}= 6.4184+ 0.2126\text{Malaysia}_{(t-1)}- 0.3202\text{Singapore}_{(t-1)}- 0.4268\text{Thailand}_{(t-1)} - 0.4577\text{Indonesia}_{(t-1)}+1.1959\text{Philippines}_{(t-1)}- 0.0347\text{Malaysia}_{(t-2)} - 0.3767\text{Singapore}_{(t-2)} - 0.2036\text{Thailand}_{(t-2)}+ 0.1458\text{Indonesia}_{(t-2)}+ 0.7802\text{Philippines}_{(t-2)}+ 0.642\text{Malaysia}_{(t-3)} - 0.4852\text{Singapore}_{(t-3)}- 0.1379\text{Thailand}_{(t-3)}- 0.4354\text{Indonesia}_{(t-3)}+ 0.0770\text{Philippines}_{(t-3)}$
$\text{Indonesia}_{(t)}$	$\text{Indonesia}_{(t)}= 0.4917-0.0007\text{Malaysia}_{(t-1)}- 0.0063\text{Singapore}_{(t-1)}+ 0.0146\text{Thailand}_{(t-1)} - 0.7315\text{Indonesia}_{(t-1)}+ 0.1255\text{Philippines}_{(t-1)}- 0.0143\text{Malaysia}_{(t-2)} - 0.0031\text{Singapore}_{(t-2)} + 0.0246\text{Thailand}_{(t-2)}- 0.4919\text{Indonesia}_{(t-2)}+ 0.1929\text{Philippines}_{(t-2)}- 0.0083\text{Malaysia}_{(t-3)} + 0.0245\text{Singapore}_{(t-3)}+ 0.0517\text{Thailand}_{(t-3)}- 0.3022\text{Indonesia}_{(t-3)}+ 0.1300\text{Philippines}_{(t-3)}$
$\text{Philippines}_{(t)}$	$\text{Philippines}_{(t)}= 3.3177+0.0717\text{Malaysia}_{(t-1)}- 0.0965\text{Singapore}_{(t-1)}+ 0.0526\text{Thailand}_{(t-1)} - 0.0225\text{Indonesia}_{(t-1)}-0.4334\text{Philippines}_{(t-1)}- 0.0093\text{Malaysia}_{(t-2)} - 0.0852\text{Singapore}_{(t-2)} - 0.0121\text{Thailand}_{(t-2)}- 0.0570\text{Indonesia}_{(t-2)}+0.1021\text{Philippines}_{(t-2)}+ 0.3255\text{Malaysia}_{(t-3)} - 0.0986\text{Singapore}_{(t-3)}- 0.0124\text{Thailand}_{(t-3)}- 0.2001\text{Indonesia}_{(t-3)}+ 0.0255\text{Philippines}_{(t-3)}$

3.2.2 Diagnosis checking

Table 4 displays the results of L-Jung Box test for each country's residual. Ljung-Box tests are used to determine whether residuals at different lags of a time series model exhibit significant autocorrelation. In residuals, autocorrelation indicates that there are patterns or trends in the data that have not been accounted for by the model.

Based on the result of Table 4, the p-value of Malaysia (0.6526), Singapore (0.5163), Thailand (0.7205) and Philippines (0.9138) are greater than significant level of 0.05, it indicates that there is no autocorrelation in residual. In the case of Indonesia, the p-value (0.1606) is greater than the commonly used significance level of 0.05. This suggests that there is no significant evidence of autocorrelation in the residuals up to the specified lag.

Table 4 : Result of L-Jung Box test for each country's residuals

Country	p-value
Malaysia	0.6526
Singapore	0.5163
Thailand	0.7205
Indonesia	0.1606
Philippines	0.9138

3.3 Multivariate Singular Spectrum Analysis (MSSA)

The first stage involves the decomposition of the monthly number of tourist arrivals to Maldives while the second level involves the reconstruction of the monthly number of tourist arrivals to Maldives. Meanwhile, the reconstructed series used to forecast the future value of the monthly number of tourist arrivals to Maldives with the absence of noise.

Moreover, careful consideration should be given to determining the suitable size of the window length, denoted as L. This is crucial as the construction of the trajectory matrix relies on the chosen value of L. Opting for a longer window length enables the capture of more prolonged trends, while a shorter window length enhances sensitivity to short-term fluctuations. It is essential to note that the window length has an impact on the accuracy of forecasted values. Given the absence of specific guidelines for window length selection in a dataset, a previous researcher suggested that an appropriate window length should not exceed $N/2$. Therefore, for this study, the L value is set at 135.

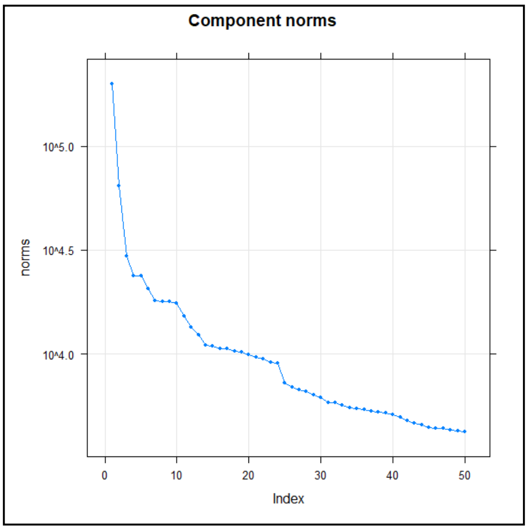

Fig. 6 Scree plot

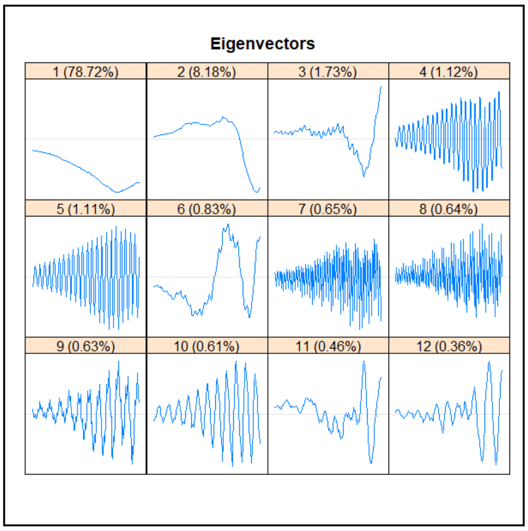

Fig. 7 Vector plot of MSSA (L=135)

Fig. 6 represents a scree plot, which displays the singular values in descending order. Singular values represent the importance of corresponding eigenvectors. The scree plot helps identify an "elbow" point where the importance of singular values diminishes. The number of eigenvectors corresponding to the plateau after the elbow is 12. Fig.7 shows the 12 eigenvectors plot that obtained from the first stage of MSSA which is the decomposition stage. The plots in the figure are grouped based on the characteristics of the components such as trend, seasonal components, and noise. From the plots, it is known that the series constructed by eigentriple 1 and eigentriple 2 contain slowly varying component, so that the two eigentriples are grouped into the trend groups. The eigentriple 6, eigentriple 11, and eigentriple 12 are grouped into the noise groups and the rest are identified as seasonal components.

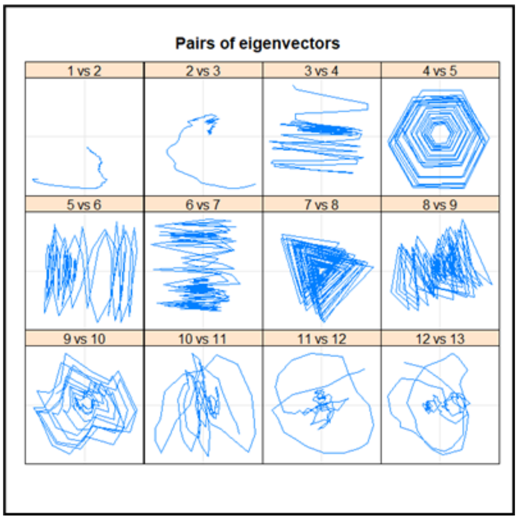

Fig. 8 Pairs of eigenvector plot of MSSA (L=135)

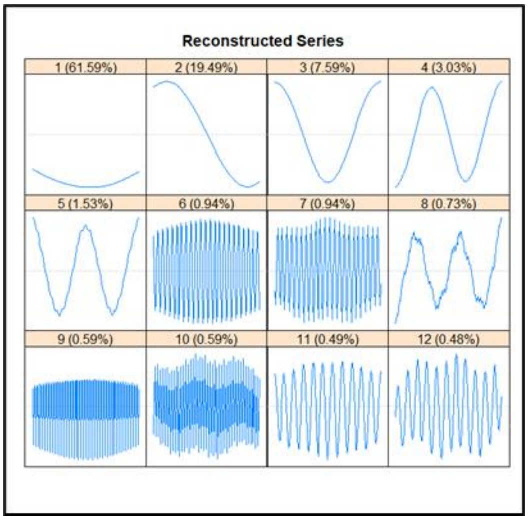

Fig. 9 Reconstructed series of MSSA (L = 135)

In Fig. 8, the plots comparing 4 and 5, as well as 7 and 8, depict p-vertex polygons. This indicates that these are pairs of sine/cosine sequences characterized by zero phase and identical amplitude. In simpler terms, these pairs generate harmonic components with varying periods. Following the identification of components in the multivariate time series, the subsequent phase involves reconstruction. As depicted in Fig.9, nearly all eigenvalues exhibit distinctions from their counterparts before the reconstruction process. Consequently, a novel dataset is created, leading to enhanced forecasting outcomes due to the reduction of noise components in the dataset. In addition, the MSSA encompasses two forecasting approaches: vector forecasting and recurrent forecasting. While recurrent forecasting is tailored for univariate time series, predicting only one step ahead, this study opts for vector forecasting. Vector forecasting is deemed more fitting for multivariate time series, as it is designed to generate joint forecasts for multiple variables within the time series data.

3.4 Comparison between VAR Model and MSSA Model

A comparative analysis is undertaken between the VAR model and MSSA. The evaluation is based on three measures: MAE (Mean Absolute Error), RMSE (Root Mean Square Error), and MAPE (Mean Absolute Percentage Error). These measures serve as criteria for selecting the superior method for forecasting the monthly number of tourist arrivals to the Maldives. Lower values of MAE, RMSE, and MAPE signify a more suitable and accurate forecast model. Table 5, 6 and 7 present the results of MAE, RMSE, and MAPE values for both methods.

According to the outcomes presented in Table 5, 6, and 7, it is evident that the multivariate VAR model demonstrates lower MAE, RMSE, and MAPE values for the countries Malaysia, Indonesia, and the Philippines in comparison to the MSSA model. This substantiates that the multivariate VAR model exhibits higher accuracy than the MSSA, as indicated by the observed lower values. Conversely, for the MSSA model, the MAE and RMSE values for Singapore and Thailand are observed to be lower than those of the VAR model. While the MAPE value for the VAR model with Thailand is marginally lower than the MSSA model, both values are comparable at 42.78% and 43.56%, respectively. Consequently, it can be concluded that the MSSA model is more fitting for forecasting the number of tourist arrivals from Singapore and Thailand to Maldives.

Table 5 : Results of MAE values of VAR Model and MSSA Model

Country	VAR Model	MSSA Model
Malaysia	**346.07**	645.33
Singapore	1348.70	**1238.99**
Thailand	848.49	**795.87**
Indonesia	**38.21**	224.54
Philippines	**147.46**	304.24

*Bold indicating the best results.

Table 6 : Results of RMSE values of VAR Model and MSSA Model

Country	VAR Model	MSSA Model
Malaysia	**410.99**	709.28
Singapore	1430.20	**1360.73**
Thailand	976.33	**917.46**
Indonesia	**40.81**	231.33
Philippines	**196.40**	348.45

*Bold indicating the best results.

Table 7 : Results of MAPE values of VAR Model and MSSA Model

Country	VAR Model	MSSA Model
Malaysia	**29.26**	59.92
Singapore	140.35	**101.64**
Thailand	**42.78**	43.56
Indonesia	**19.10**	103.52
Philippines	**24.34**	52.74

*Bold indicating the best results.

4. Conclusions

There are two methods used in this study which are the VAR model and MSSA model. All the objectives have been fulfilled and achieved. The first objective was to compare the trends and patterns in tourists' arrival to Maldives before, during and after COVID-19 pandemic using graphical analysis. The time series plot shows the increasing pattern in tourist arrivals from five different countries to Maldives before the COVID-19 pandemic. Then, it shows a decreasing line during the COVID-19. Tourist arrivals have been recovering after the COVID-19 pandemic, indicating an upward trend. The second objective was to determine the best model for tourists' arrival after COVID-19 pandemic using the Multivariate Singular Spectrum Analysis (MSSA) and Vector Autoregression (VAR) model. Both methods have been compared by using an accuracy measurement and selected the best model for the five countries. The best model by using VAR model are Malaysia, Indonesia, and Philippines. For Singapore and Thailand, the MSSA model are the best model compared to the VAR model. The third objective was to measure the performance of the techniques between MSSA and VAR model using Mean Absolute Percentage Error (MAPE), Mean Absolute Error (MAE) and Root Mean Square Error (RMSE). Lower values of the MAPE, MAE and RMSE values were selected as they represent the more precise and more accurate forecast model. Based on the result, the more accurate forecast model for the number of tourists arrival from Malaysia, Indonesia, and Philippines to

Maldives is the VAR model while the MSSA model is more precise in forecasting the number of tourists arrival from Singapore and Thailand to Maldives.

Acknowledgement

The authors would thank the Faculty of Applied Sciences and Technology, Universiti Tun Hussein Onn Malaysia for its support.

Conflict of Interest

The author declares that there is no conflict of interest that could compromise the impartiality and objectivity of this study. This declaration is made to ensure the credibility and reliability of the research findings presented in this paper.

Author Contribution

The authors confirm contribution to the paper as follows: ***study conception and design:*** *Seh Chai Nie, Norhaidah Mohd Asrah;* ***data collection:*** *Seh Chai Nie;* ***analysis and interpretation of results:*** *Seh Chai Nie, Norhaidah Mohd Asrah;* ***draft manuscript preparation:*** *Seh Chai Nie, Norhaidah Mohd Asrah. All authors reviewed the results and approved the final version of the manuscript.*

References

[1] Ciotti, M., Ciccozzi, M., Terrinoni, A., Jiang, W. C., Wang, C. B., & Bernardini, S. (2020). The COVID-19 pandemic. *Critical reviews in clinical laboratory sciences*, 57(6), 365-388. https://doi.org/10.1080/10408363.2020.1783198

[2] Shen, H., Fu, M., Pan, H., Yu, Z., & Chen, Y. (2020). The impact of the COVID-19 pandemic on firm performance. *Emerging Markets Finance and Trade*, 56(10), 2213-2230. https://doi.org/10.1080/1540496X.2020.1785863

[3] Wakimin N., Azlinaa A., & Hazman S. (2018) Tourism demand in Asean-5 countries: Evidence from panel data analysis. *Manag. Sci. Lett.* 8:677–690. http://dx.doi.org/10.5267/j.msl.2018.4.023

[4] Manzoor F, Wei L, Asif M, Haq MZ, & Rehman H. (2019). The Contribution of Sustainable Tourism to Economic Growth and Employment in Pakistan. *Int J Environ Res Public Health.16(19):3785*. https://doi.org/10.3390/ijerph16193785

[5] Makun, K., & Jayaraman, T. K. (2022). Tourism and economic growth in South asian countries: Asymmetric analysis and lessons for mitigating the adverse effects of COVID-19 pandemic. MPRA. https://mpra.ub.uni-muenchen.de/111377/

[6] Ministry of Tourism. (2021). Monthly tourism statistics report. Retrieved June 20, 2023, from https://www.tourism.gov.mv/statistics.

[7] Chen, R., Xiao, H., & Yang, D. (2021). Autoregressive models for matrix-valued time series. *Journal of Econometrics*, 222(1), 539-560. https://doi.org/10.1016/j.jeconom.2020.07.015

[8] De Assis Paiva, D., & Sáfadi, T. (2021). Study of tests for trend in time series. *Brazilian Journal of Biometrics*, 39(2), 311-333. https://doi.org/10.28951/rbb.v39i2.471

[9] Shadab, A., Sinha, A. K., & Gupta, S. K. (2021). Validating Stationarity as an Antecednt to Studying the Impacat of Exchange Rate Volatility on Bankex Values. *Indian Journal of Economics and Business*, 20(2), 511-521. https://www.ashwinanokha.com/resources/ijeb%20v20-2-34.%20ijeb.pdf

[10] Dao, P. B., & Staszewski, W. J. (2021). Lamb wave based structural damage detection using stationarity tests. *Materials*, 14(22), 6823. https://doi.org/10.3390/ma14226823

[11] Alghamdi, T., Elgazzar, K., Bayoumi, M., Sharaf, T., & Shah, S. (2019, June). Forecasting traffic congestion using ARIMA modeling. *In 2019 15th international wireless communications & mobile computing conference (IWCMC)* (pp. 1227-1232). IEEE. https://doi.org/10.1109/IWCMC.2019.8766698

[12] Malamiri, H. R. G., Rousta, I., Ólafsson, H., Zare, H., & Zhang, H. (2018). Gap-Filling of MODIS Time Series Land Surface Temperature (LST) Products Using Singular Spectrum Analysis (SSA). *Atmosphere*, *9*(9), 334. https://doi.org/10.3390/atmos9090334