BAREKENG: Journal of Mathematics and Its Applications
March 2023 Volume 17 Issue 1 Page 0285–0294
P-ISSN: 1978-7227 E-ISSN: 2615-3017
https://doi.org/10.30598/barekengvol17iss1pp0285-0294

FORECASTING TOURISM DEMAND DURING THE COVID-19 PANDEMIC: ARIMAX AND INTERVENTION MODELLING APPROACHES

Fahriza Rianda[1*], Hardius Usman[2]

[1,2]Department of Applied Statistics, Politeknik Statistika STIS
Otto Iskandardinata St., No. 64C, Jakarta Timur, 13330, Indonesia

*Corresponding author's e-mail: * **211810281@stis.ac.id***

Article History:

Received: 1st October 2022
Revised: 22nd December 2022
Accepted: 30th January 2023

Keywords:

ARIMAX model;
COVID-19;
Forecasting;
Google Trends;
Intervention model.

ABSTRACT

The tourism sector in Indonesia is one of the economic sectors severely impacted by the COVID-19 pandemic and its variants. The government is attempting to revive the economy by implementing numerous recovery policies. These economic recovery policies, particularly in the tourism sector, must be backed by a policy evaluation conducted using tourism demand data, such as the number of international visitors visiting Indonesia. However, official data from BPS-Statistics Indonesia has been released with a two-month delay and is sometimes revised the following month. Consequently, it is necessary to forecast the data for the current situation. In addition, the forecasting model must be modified to account for data conditions caused by the COVID-19 pandemic. This research proposes an ARIMAX forecasting model that utilizes Google Trends and an intervention forecasting model with data sources from BPS and Google Trends. Thus, this research aims to present an overview, develop a model and identify the best forecasting model, and calculate the influence of the intervention on the number of international visitors visiting Indonesia. Compared to the ARIMAX model, the results indicated that the intervention model provided the most accurate forecasts. Not only superior in forecasting results, but the intervention model also demonstrates the magnitude of the intervention's effect on the number of international visitors visiting Indonesia.

How to cite this article:

F. Rianda and H. Usman., "FORECASTING TOURISM DEMAND DURING THE COVID-19 PANDEMIC: ARIMAX AND INTERVENTION MODELLING APPROACHES," *BAREKENG: J. Math. & App.*, vol. 17, iss. 1, pp. 0285-0294, March 2023.

Journal homepage: *https://ojs3.unpatti.ac.id/index.php/barekeng/*
Journal e-mail: *barekeng.math@yahoo.com; barekeng.journal@mail.unpatti.ac.id*

Research Article • Open Access

1. INTRODUCTION

Tourism is one of the most desirable economic sectors for a country to develop. The tourism industry can result in an increase in foreign exchange profits, the creation of new jobs, and business-driven infrastructure development. In addition, the tourism sector has a multiplier impact, which can propel the industry and encourage investors to invest in industries that support tourism [1]. Important indices of Indonesian tourism statistics, such as the number of international visitor arrivals, contribution to GDP, and employment in the tourism sector, rose each year from 2014 to 2019 [2]–[4]. This cannot be separated from the Government of Indonesia's support for the National Tourist Development Master Plan (Riparnas) 2010-2025 and its designation of the tourism sector as a national development priority in the 2015-2019 and 2020-2024 National Medium-Term Development Plans (RPJMN).

However, the development of the Indonesian tourism sector did not continue in 2020. BPS [3] reported that the number of foreign tourists visiting Indonesia in 2020 decreased by as much as 75% compared to 2019. In addition, WTTC [4] reported that the tourist sector's contribution to Indonesia's GDP in 2020 declined from 50.7 percent in 2019 to 2.8 percent in 2020, and Indonesia lost about 1.92 million tourism workers in 2020. The presence of the Coronavirus Disease 2019 (COVID-19) pandemic, which is an infectious disease caused by the Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) virus, is the primary cause of a drop in the number of foreign tourists visiting many countries, as well as other tourism statistical indicators [5]. Because the virus spreads so quickly over the world, many countries have imposed travel and community activity limitations in order to reduce the spread and impact of COVID-19 pandemic.

The Indonesian government is attempting to repair the national economy, which was contracted due to strict mobility restrictions to reduce the spread of COVID-19. Tourism and the creative economy are two of the sectors that have received government help through the National Economic Recovery (PEN) fund to recover [6]. An examination of government policies is required for numerous government initiatives and policies related to the tourist sector to perform smoothly during the COVID-19 pandemic. Using tourism demand data, such as the number of international visitors visiting Indonesia, is one of the evaluation procedures that can be used. However, official data from the BPS has been delayed by two months, and newly released data is occasionally subject to changes in the following month. As a result, forecasting data for the current situation about foreign tourist visits to Indonesia is required so that the government can promptly implement and evaluate policies relating to the Indonesian tourism sector during the COVID-19 pandemic.

The utilized forecasting model must account for the consequences of the COVID-19 pandemic, as the pandemic has altered the pattern of data pertaining to international visitors visiting Indonesia. The initial model provided is the ARIMAX model with exogenous variables in the form of Google Trends search query data. Google Trends is one of the applications of big data that has been widely used in a variety of research over the past decade since it can provide information on currently popular subjects in real-time [7]. In addition, the search query data has the potential to aid in forecasting when events that affect foreign tourist visits to Indonesia, such as the COVID-19 pandemic [8]. This is quite useful for forecasting, given that international visitors who are planning a holiday will explore the internet for tourist information. Previous research has demonstrated that the use of Google Trends as an exogenous variable can result in more accurate forecasts compared to the absence of Google Trends [9], [10]. Even Wen et al. [11] have shown that the application of the ARIMAX forecasting model with search query data produced more accurate forecasts than the machine learning model, specifically the Neural Network model.

The ARIMA intervention model is the second forecasting model used as a comparison model without using Google Trends data and is capable of accommodating the COVID-19 pandemic. Box et al. [12] indicated that the intervention model is utilized when time series data are influenced by exceptional events or circumstances, such as political instability, holidays, or natural disasters. The intervention model is not only superior in forecasting compared to the ARIMA model in general [13], [14], but it can also determine the size of the impact of interventions on time series data, which is essential for evaluating policies [15]–[17]. This research will examine two intervention variables related to the emergence of the COVID-19 virus, namely the time when the SARS-CoV-2 virus that caused the COVID-19 pandemic was first detected and the time when the WHO first identified a variant of SARS-CoV-2.

This research will examine the description of the development of the number of international visitors visiting Indonesia and Google Trends before and after the COVID-19 pandemic, construct a forecasting model and determine the best forecasting model for predicting the number of international visitors visiting Indonesia, and calculate the magnitude of the impact of the two interventions. The purpose of the research is anticipated to aid the government in formulating and analyzing tourism sector recovery plans during the COVID-19 pandemic, particularly for international visitors visiting Indonesia.

2. RESEARCH METHODS

2.1 Google Trends Web Search Queries

Google Trends is a particular platform from Google that was introduced in May 2006 that illustrates how frequently specific search keywords are put into the Google search engine relative to the total number of searches within a specific time period and location [18]. Google Trends presents a graph of interest in a specific phrase or search query, which may be specified by time period, area, and category. Google Trends presents its data in the form of an index whose value ranges from 0 to 100. The idea of consumer behavior can be applied to the selection of Google Trends search queries utilized as exogenous variables. This is because tourism consumer behavior is related to how visitors who are buyers of products and services behave while purchasing or utilizing accessible services [19]. The selection of these queries is based on Mathieson and Wall's (1982) theory of *travel-buying behavior* [20] and is supported by Rödel's research [21] about the selection of the stages of *travel-buying behavior* theory. Table 1 provides a summary of the Google Trends search query selected following the theory of *travel-buying behavior*.

Table 1. Summary of Google Trends search query used

Stages of Behavior	Search Behavior	Web Search Queries Google Trends
Felt need/travel desire	Looking for the type of vacation Find a destination to travel	*Trip to Indonesia* (TTI) *Bali Spots* (BS) *Batam Island* (BI)
Information collection and evaluation by image	Find information on tourist destinations	*Indonesia Travel* (IT) *Holidays in Indonesia* (HsII)
Travel decision	Determining the mode of travel Determining accommodation Determining tourist activities	*Flights to Indonesia* (FTI) *Hotels in Indonesia* (HII) *Things to do in Indonesia* (TTDI)

2.2 Data Collection Methods

This research uses *time series* data for the monthly period from January 2014 to January 2022. The data are separated into two categories: *in-sample* data (January 2014 to September 2021) and out-of-sample data (October 2021 to January 2022). This study's *out-of-sample* data are limited due to changes in the BPS's methodology for collecting data on international visitor arrivals since February 2022. The number of international visitors visiting Indonesia, as reported by BPS, is then employed as the dependent variable in this study. Web search queries in the form of an index are employed as exogenous variables, and they are derived from Google Trends. Meanwhile, the intervention variables employed include the first identified COVID-19 pandemic by WHO (in January 2020) and the first detected new variant of COVID-19 by WHO (December 2020).

2.3 Analysis Methods

The analytical methods used in this research include descriptive and inference analysis. Descriptive analysis is used to provide an overview of the number of international visitor arrivals and Google Trends web search queries and provide an overview of the movement patterns between them. Meanwhile, inference analysis is used to analyze and forecast the number of international visitors to Indonesia through the ARIMAX model and multi-input intervention model.

The data used to generate the model is *in-sample* data. The steps of model development are often summarized as follows [15], [22], [23]:

1. Creating an ARIMAX model
    a. Create a time series regression model using the backward elimination method with exogenous Google Trends variables.

   b. Testing the estimated *error* (residual) of the regression model, specifically the residual stationarity test using the Augmented Dickey-Fuller (ADF) test and the *white noise* assumption on the residuals using the Ljung-Box test.
   c. When the residuals of the *time series* regression model do not match the *white noise* assumption, forming an ARIMA model from residuals (ARIMA *error)* by identifying the order *p,d,q* based on the residual ACF and PACF plots.
   d. Re-establish the regression model and the residual ARIMA model simultaneously, resulting in the ARIMAX model.
   e. All exogenous variables in the ARIMAX model have been confirmed to meet the non-multicollinearity assumption, namely, the Variance Inflation Factor (VIF) value is less than 10.
   f. Re-checking the assumption that the error is white noise through the Ljung-Box test and normally distributed through the Jarque-Bera test on the ARIMAX model.
   g. Forecasting and calculating the RMSE and symmetric MAPE (sMAPE).
   h. The ARIMAX model can be written as follows:

$$Y_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + \cdots + \beta_k X_{kt} + \frac{\theta_q(B)}{\phi_p(B)} \varepsilon_t \tag{1}$$

2. Creating a multi-input intervention model
   a. Organizing the data according to the times when the intervention occurred.
   b. Establishing an ARIMA model before the intervention by identifying the order *p,d,q* based on the ACF and PACF plot and paying attention to the criteria for selecting the best model based on significant parameters and the smallest AIC value. The best model is then checked for assumptions against errors that are white noise and normally distributed.
   c. Forecasting with the ARIMA model before the intervention till the first intervention data is reached.
   d. Calculate the residual value and plot the residual for the first intervention data. The residual plot has an upper and lower limit of $\pm 2\sigma_0$ or $\pm 3\sigma_0$, where σ_0 is the RMSE obtained through the ARIMA model before the first intervention.
   e. Based on the residual plot, determine the order of *b* (delay time), *s* (length of time to stabilize), and *r* (intervention effect pattern).
   f. In the first intervention ARIMA model, perform parameter estimation and the criteria for selecting the best model.
   g. In the first intervention ARIMA model, the assumption that the error is white noise and normally distributed is tested.
   h. Create a second intervention ARIMA model by following the same processes as the first.
   i. Forecasting and calculating the RMSE and symmetric MAPE (sMAPE).
   j. The ARIMA intervention model with two inputs can be written as follows:

$$(1-B)^d Y_t = \mu + \sum_{j=1}^{2} \frac{\omega_{s_j}(B) B^{b_j}}{\delta_{r_j}(B)} I_{j,t} + \frac{\theta_q(B)}{\phi_p(B)} \varepsilon_t \tag{2}$$

3. Comparing the forecasting accuracy results from the best selected models based on the RMSE and sMAPE. When the RMSE and sMAPE of a forecasting model are the lowest when compared to other forecasting models, the model is the best. The formulas for RMSE and sMAPE are as follows [23]:

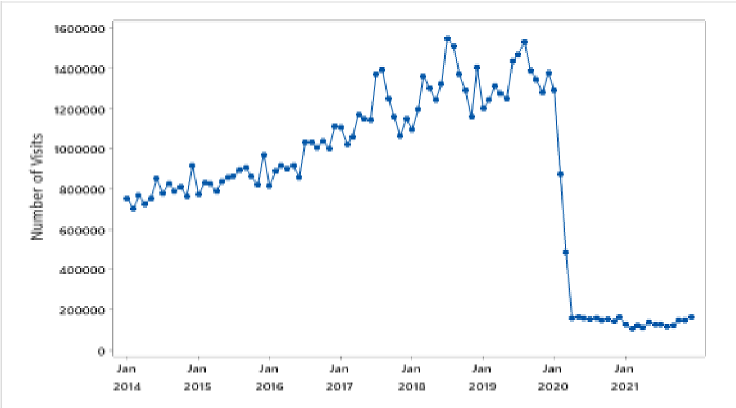
$$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^{n} (Y_t - \widehat{Y}_t)^2} \tag{3}$$

$$sMAPE = \frac{1}{n} \sum_{t=1}^{n} \frac{2\left|Y_t - \widehat{Y}_t\right|}{\left(Y_t + \widehat{Y}_t\right)} \times 100\% \tag{4}$$

3. RESULTS AND DISCUSSION

3.1 Overview of the Number of International Visitors to Indonesia

According to **Figure 1**, the number of international visitors visiting Indonesia fluctuates and tends to rise from year to year during the period 2014 to 2019. The number of international visitors visiting Indonesia in 2019 reached 16,106,954 people. This figure increased by 71 percent when compared to the 9,435,411 international visitors that came to Indonesia in 2014. However, the annual increase in the number of international visitors visiting Indonesia will not continue from 2020 to the end of the research period in 2021. In 2020 and 2021, there will be 4,052,923 and 1,557,530 international visitors in Indonesia, respectively. This figure experienced a very sharp decline when compared to 2019, which was 75 percent in 2020 and 90 percent in 2021. The decline in international visitors to Indonesia was caused by the COVID-19 pandemic and the new variant of COVID-19, which the pandemic caused many the state has imposed restrictions on travel and community activities to suppress the very rapid spread of the virus. UNWTO [5] explained that the travel restriction policies carried out by a country against visitors, the slow control of the COVID-19 virus, and the unfavorable economic conditions of a country are the primary causes of the decline in the number of international visitor arrivals experienced by some major countries in the world.

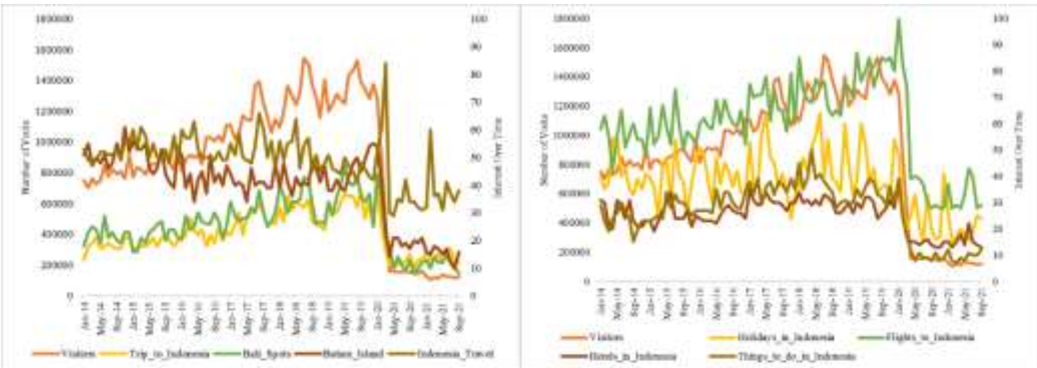


Source: BPS, processed

Figure 1. The Number of International Visitor Arrivals to Indonesia 2014-2021

3.2 Overview of Web Search Queries Index

Google Trends presents an interest graph for a certain search query, which can be specified by time period, area, and category. Google Trends presents its data in the form of an index with a range from 0 to 100.



(a) (b)

Figure 2. The number of international visitor arrivals (in-sample) and search queries (a) "*Trip to Indonesia*", "*Bali Spots*", "*Batam Island*", "*Indonesia Travel*" (b) "*Holidays in Indonesia*", "*Flights to Indonesia*", "*Hotels in Indonesia*", "*Things to do in Indonesia*"

Sumber: BPS and Google Trends, processed

The queries "*Trip to Indonesia*", "*Bali Spots*", "*Flights to Indonesia*", "*Hotels in Indonesia*", and "*Things to do in Indonesia*" tend to have the same trend as the number of international visitors visiting Indonesia, as illustrated in **Figure 2**. The movement of these web search queries fluctuated and tended to increase from 2014 to 2019, but declined after the start of 2020. Using Pearson's correlation coefficient, the correlation coefficients between each prior web search query and the number of international visitor arrivals are 0.87, 0.88, 0.86, 0.85, and 0.91. Meanwhile, the remaining web search queries, notably "*Batam Island*", "*Holidays in Indonesia*", and "*Indonesia Travel*" have fluctuated, but when compared to the number of international visitors visiting Indonesia, this movement appears inconsistent direction at the beginning of the research period. According to the Pearson correlation, the correlation coefficients between the three queries and the number of international visitor arrivals are 0.71, 0.65, and 0.47, respectively.

3.3 Development of ARIMAX Model

Before beginning the modeling, the stationarity of variance in the data on the number of international visitors visiting Indonesia must be checked. The resulting lambda (λ) value from the Box-Cox plot is 1.17. Because the value is close to one, the data can be said to be stationary on the variance, and no transformation is required. The ARIMAX model is created by combining the time series regression model with the *backward elimination* method with Google Trends web search queries as exogenous variables and the ARIMA *error* model, which are then re-estimated simultaneously, as shown in **Table 2**. **Table 2** demonstrates that the ARIMAX (2.0,0) model with exogenous variables BS, BI, and TTDI is the best.

Table 2. The Results of the Parameter Significance Test and Evaluation of the ARIMAX model (2,0,0)

Parameter Estimation		Error Independence Test (LB Test)			Normality Test Error (JB Test)	
Parameter Estimation	**P-value**	**Lag**	**LB Statistic**	**P-value**	**JB Statistic**	**P-value**
ϕ_1 = 0.6962	<0.0001	6	2.14	0.9063	4.4934	0.1057
ϕ_2= 0.2719	0.0413	12	12.69	0.3923		
BS = 4897.7228	0.0477	18	19.89	0.3391		
BI = 7780.6512	0.0002	24	22.32	0.5602		
TTDI = 5651.0396	0.0396					

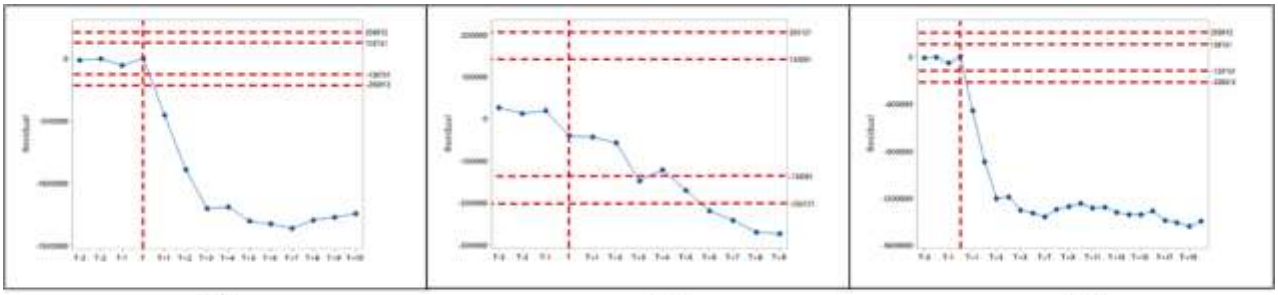
Table 2 shows the results of parameter estimation and evaluation of the ARIMAX (2,0,0) model. The model has fulfilled the assumption of white noise error and is normally distributed because the resulting p-value is greater than the 5 percent significance level. The model is the final model that can be used as a forecast. As a further indicator of forecasting accuracy, the model also has an RMSE (*in-sample*) score of 97661.69.

3.4 Development of Multi-Input Intervention Model and Impact Assessment

Data must be partitioned based on the times when the intervention occurred before modeling, namely data before the first intervention (January 2014 – December 2019), data after the first intervention (January 2020 – November 2020), and data after the second intervention (December 2020 – September 2021). The first stage is to create an ARIMA model before the intervention. Checking the stationarity of the data before the intervention needs to be done with the ADF test. The results demonstrate that the *p-value* for stationary data at the first differentiation level is 0.01, which is less than the 5 percent significance level. The ACF and PACF plots were used to identify the ARIMA model before the intervention. Based on the ACF and PACF plot analysis, the tentative models of ARIMA before intervention obtained were the ARIMA $(0,1,1)(1,0,0)^{12}$, the ARIMA $(1,1,0)(1,0,0)^{12}$, and the ARIMA $(1,1,1)(1,0,0)^{12}$ model.

According to the criteria for choosing the best model, the ARIMA $(1,1,1)(1,0,0)^{12}$ with intercept has lower AIC and RMSE values than other ARIMA tentative models, namely 1795.723 and 69870.582. The model also passed the parameter significance test with a significant level of 5 percent and the assumption of an independent error (white noise) and followed a normal distribution because both *p-values* are greater than the 5 percent significance level. The best model is used to calculate the residual value in data II, which is the data from after the first intervention to before the second intervention. This residual is helpful in determining the order of the ARIMA model's first intervention.

The first intervention in this research was the first identified COVID-19 pandemic, which began in January 2020. Based on the visualization, the intervention follows a *step* function because the number of international visitors visiting Indonesia has decreased dramatically and has an impact for a long time until the end of the research period following the COVID-19 pandemic. The first intervention ARIMA model is built by first determining the order of intervention, namely b, s, and r. The intervention order can be identified by observing the residual plot that exceeds the limit of $\pm 2\sigma_0$ or $\pm 3\sigma_0$, where σ_0 is the RMSE obtained through the best ARIMA model before the first intervention.



(a) (b) (c)

Figure 3. Residual plots on (a) first intervention (b) second intervention (c) first intervention after change

Figure 3a demonstrates that the order b or time delay is 1 because the residuals exceed the limit at time T+1 after time T. Then, there is a decrease and the order tends to stabilize after order b at time T+3, so the estimated value of order s is [2]. The order of r is estimated to be 1 based on the pattern exhibited by the residual plot, which tends to ascend. However, several possible orders were attempted, beginning with s equal to 1 to 3 and r equal to 0 or 1.

Based on the best model selection criteria, the ARIMA $(1,1,1)(1,0,0)^{12}$ model with intercept and order b = 1, s = [2], and r = 1 was selected to be the best first intervention ARIMA model. This is because the model is the only one that passes the parameter significance test with a 5 percent significance level. In addition, the model satisfies the assumption of an independent error (white noise) and a normal distribution because both of its *p-values* exceed the 5 percent significance level. The best model is used to obtain the residual value in data III, namely the data after the second intervention until the end of the research period.

In December of 2020, a new COVID-19 variant was identified, initiating the second intervention in this study. The second intervention, similar to the initial intervention, followed a *step* function. Determining the order of interventions b, s, and r is also the starting point for the development of the second intervention ARIMA model. The intervention order can be identified by observing the residual plot that exceeds the limit of $\pm 2\sigma_1$ or $\pm 3\sigma_1$, where σ_1 is the RMSE obtained through the best first intervention ARIMA model with a value of 68040.487. The researcher used a trial-and-error method to determine the b, s, and r orders based on the combinations shown in **Figure 3b**. A total of 17 combinations of models with orders b, s, and r reveal that none of the models are significant for the second intervention on orders b, s, and r. These results indicate that the intervention that identified a new variant of COVID-19 did reduce the number of foreign tourists visiting Indonesia, but the decline was not statistically significant after the first intervention, according to the data on the number of international visitors visiting Indonesia.

None of the significant models on the b, s, and r orders for the second intervention made the first intervention model change. According to **Figure 3c**, the residual plot pattern is comparable to the residual plot pattern in the first intervention before the change. However, the selection of the model is still done by looking at the combination of other models as well as the formation of the first intervention model.

Table 3. The Results of the Parameter Significance Test and Evaluation of the Best First Intervention Model after the Change

Parameter Estimation		Error Independence Test (LB Test)			Normality Test Error (JB Test)	
Parameter Estimation	**P-value**	**Lag**	**LB Statistic**	**P-value**	**JB Statistic**	**P-value**
μ = 7535.6	0.0288	6	1.03	0.7942	3.2569	0.1962
ϕ_1= 0.3902	0.0066	12	4.10	0.9050		
θ_1= 0.8612	<0.0001	18	10.21	0.8065		
Φ_1 = 0.53138	<0.0001	24	17.87	0.6574		
ω_{0_1}= -385859.2	<0.0001					
ω_{2_1}= -382384.7	<0.0001					
δ_1= 0.38351	<0.0001					

The ARIMA $(1,1,1)(1,0,0)^{12}$ with intercept and order b = 1, s = [2], and r = 1 was reselected as the first intervention ARIMA model after the change based on the criteria for selecting the best model. This is because the model is the only one that passes the parameter significance test with a 5 percent significance level. In addition, the model satisfies the assumption of an independent error (white noise) and a normal distribution because both of its *p-values* exceed the 5 percent significance level. The best model is the final model used to calculate the magnitude of the impact and forecast. The final model has an RMSE (in-sample) value as a measure of forecasting accuracy of 65168,502.

Table 3 also shows that the parameter ω_{0_1} is negative and significant, with an order of b equal to 1. This explains that the initial impact felt after the COVID-19 pandemic was a decrease in the number of international visitors visiting Indonesia, which was felt for the first time one month after COVID-19 was identified in the world or in other words, in February 2020. So that the first intervention identified as a COVID-19 pandemic was not responded to directly by international visitors visiting Indonesia. This is because COVID-19 has not spread widely throughout the world in the early days of its appearance.

The Impact of the First Intervention to the Number of Internasional Visitors to Indonesia

The impact of the intervention can be calculated by subtracting the actual value from the forecasting results of the ARIMA before the intervention model on the data after the intervention T, due to the data not transforming **[15]**. However, if the data is in the form of a transformation, the actual value can be replaced with the forecasting results of the ARIMA intervention model. This impact explains the potential loss of international visitors' visits to Indonesia due to the COVID-19 pandemic intervention.

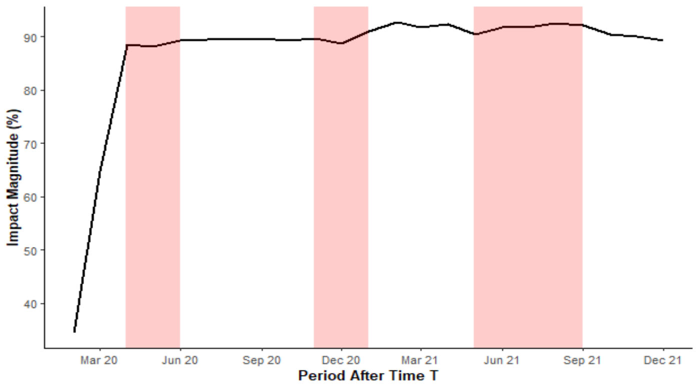


Figure 4. The Impact of Intervention on International Visitors' Visiting Indonesia

Figure 4 shows that the COVID-19 pandemic intervention had a significant effect on the decline in the number of international visitors visiting Indonesia. This impact was felt for the first time one month after COVID-19 was identified in the world, is permanent until the end of the study period, and its value is increasing along with the increasing number of positive confirmed cases of COVID-19 in Indonesia. On average, after April 2020, Indonesia has the potential to lose up to 90 percent of international visitor arrivals. Following what Gössling et al. **[24]** stated that the COVID-19 pandemic is different from various crises that have occurred in the past, such as the 2003 SARS outbreak, the 2008/2009 global economic crisis, and the 2015 MERS outbreak, where the tourism sector was able to withstand these various crises. This enormous impact indicates that the state or tourism sector businesses may experience a significant decline in tourism sector revenues, such as the state losing foreign exchange from the tourism sector and businesses losing international visitors as a source of income.

In addition, the activities of the tourism sector are also directly linked to the activities of other sectors, such as the accommodation, food and beverage, and transportation sectors [25]. Consequently, these other sectors have the potential to be impacted when tourism sector activities, such as indicators for the number of international visitors visiting, experience a moderate decline. Henseler et al. [26] demonstrated empirically that the negative impact of the COVID-19 pandemic on the tourism sector was transmitted to adjacent economic sectors such as transportation, trade, and even construction.

3.5 Comparison of Forecasting Accuracy Results

The comparison of the forecasting performance of the two models can be seen based on the results of forecasting on out-of-sample data, namely data on the number of international visitor arrivals during the period from October 2021 to January 2022.

Table 4. ARIMAX and Intervention Model Forecasting Accuracy

Model	In-sample Forecasting Results		Out-of-sample Forecasting Results	
	RMSE	sMAPE (%)	RMSE	sMAPE (%)
ARIMAX	97661.69	10.71	25569.48	12.39
Intervention ARIMA	65168.50	9.11	12923.72	8.94

Table 4 shows that, for both *in-sample* and *out-of-sample* data, the Intervention model provides the most accurate forecast of the number of international visitors visiting Indonesia. The findings of this research indicate that the intervention model is suitable for use as a forecast when data is exposed to an extraordinary event **[12]**. On the other hand, the findings of this research differ from those of previous studies in that the use of Google Trends can produce more accurate forecasts than without its use **[9]**, **[10]**.

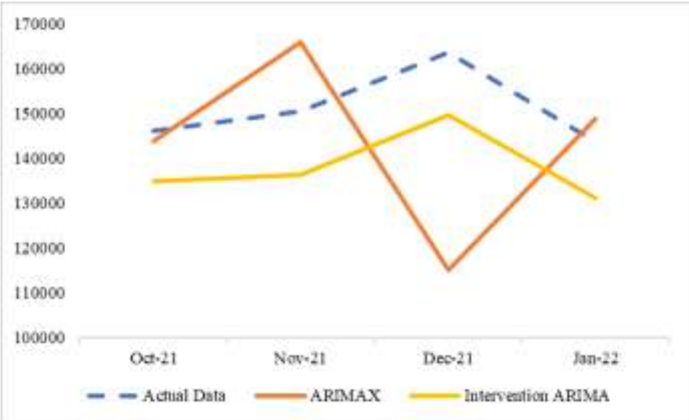


Figure 5. Out-of-sample data forecasting results for ARIMAX and intervention model

Examining the results of forecasting the *out-of-sample* data in **Figure 5**, the ARIMAX model indicated forecasting error in December 2021. The error occurred because the Google Trends web search queries used as exogenous variables demonstrated a decline in popularity in December 2021 relative to the previous month. In consequence, the ARIMAX model's forecasting results are less accurate than those of the intervention model.

4. CONCLUSIONS

Based on the results of the study, the conclusions that can be obtained are as follows:

1. The COVID-19 pandemic and its variants have had a negative impact, namely reducing the number of international visitors visiting Indonesia. Then, Google Trends web search queries have a general movement pattern that tends to be the same as the movement pattern of international visitors visiting Indonesia.
2. In the data provided, the intervention model provides better forecasting results than the ARIMAX model that utilizes Google Trends as exogenous variables for both in-sample and out-sample data.
3. The identification of the COVID-19 pandemic and its variants has led to a decline in the number of international visitors visiting Indonesia, but only the COVID-19 pandemic has had a significant impact on the number of foreign tourists visiting Indonesia. The impact was felt for the first time one month after COVID-19 was identified in the world. Since April 2020 until the end of the research period, Indonesia's potential for losing international visitors visits has averaged 90 percent.

REFERENCES

[1] I. G. B. Arjana, *Geografi Pariwisata dan Ekonomi Kreatif*. Jakarta: Rajawali Pers, 2016.

[2] Kemenparekraf/Baparekraf, "Rencana Strategis Kementerian Pariwisata dan Ekonomi Kreatif/Badan Pariwisata dan Ekonomi Kreatif 2020-2024," Jakarta, 2020.

[3] BPS, *International Visitor Arrivals Statistics 2020*. Jakarta: BPS, 2021. Accessed: Oct. 22, 2021. [Online]. Available: https://www.bps.go.id/publication/2021/06/30/ddea1823bc9cd63789d51b05/statistik-kunjungan-wisatawan-mancanegara-2020.html

[4] WTTC, "INDONESIA 2022 Annual Research: Key Highlights," 2022. Accessed: Aug. 22, 2022. [Online]. Available: https://wttc.org/Research/Economic-Impact

[5] UNWTO, "Covid-19 and Tourism 2020: A year in review," 2020. [Online]. Available: https://www.unwto.org/covid-19-and-tourism-2020

[6] Kemenko Perekonomian, "Siaran Pers: Pemerintah Dorong Pemulihan Sektor Pariwisata dan Ekonomi Kreatif," Sep. 27, 2021. https://ekon.go.id/publikasi/detail/3332/pemerintah-dorong-pemulihan-sektor-pariwisata-dan-ekonomi-kreatif

[7] E. Cebrián and J. Domenech, "Is Google Trends a quality data source?," *Appl Econ Lett*, pp. 1–5, Jan. 2022, doi: 10.1080/13504851.2021.2023088.

[8] V. Z. Eichenauer, R. Indergand, I. Z. Martínez, and C. Sax, "Obtaining consistent time series from Google Trends," *Econ Inq*, vol. 60, no. 2, pp. 694–705, Apr. 2022, doi: 10.1111/ecin.13049.

[9] B. Bokelmann and S. Lessmann, "Spurious patterns in Google Trends data - An analysis of the effects on tourism demand forecasting in Germany," *Tour Manag*, vol. 75, pp. 1–12, Dec. 2019, doi: 10.1016/j.tourman.2019.04.015.

[10] W. Höpken, T. Eberle, M. Fuchs, and M. Lexhagen, "Improving Tourist Arrival Prediction: A Big Data and Artificial Neural Network Approach," *J Travel Res*, vol. 60, no. 5, pp. 998–1017, May 2021, doi: 10.1177/0047287520921244.

[11] L. Wen, C. Liu, and H. Song, "Forecasting tourism demand using search query data: A hybrid modelling approach," *Tourism Economics*, vol. 25, no. 3, pp. 309–329, May 2019, doi: 10.1177/1354816618768317.

[12] G. E. P. Box, G. M. Jenkins, G. C. Reinsel, and G. M. Ljung, *Time Series Analysis: Forecasting and Control*, Fifth. Hoboken, New Jersey: Wiley, 2016.

[13] L. Budiarti, Tarno, and B. Warsito, "ANALISIS INTERVENSI DAN DETEKSI OUTLIER PADA DATA WISATAWAN DOMESTIK (Studi Kasus di Daerah Istimewa Yogyakarta)," *JURNAL GAUSSIAN*, vol. 2, no. 1, pp. 39–48, 2013, doi: 10.14710/j.gauss.v2i1.2742.

[14] D. al Mahkya and D. Anggraini, "Forecasting the Number of Passengers from Bakauheni Port during the Sunda Strait Tsunami Period Using Intervention Analysis Approach and Outlier Detection," *IOP Conf Ser Earth Environ Sci*, vol. 537 012009, pp. 1–9, Jul. 2020, doi: 10.1088/1755-1315/537/1/012009.

[15] M. Hisyam Lee, Suhartono, and B. Sanugi, "Multi Input Intervention Model for Evaluating the Impact of the Asian Crisis and Terrorist Attacks on Tourist Arrivals," 2010. [Online]. Available: https://oarep.usim.edu.my/jspui/handle/123456789/4216

[16] Y. Rashed, H. Meersman, E. van de Voorde, and T. Vanelslander, "Short-term forecast of container throughout: An ARIMA-intervention model for the port of Antwerp," *Maritime Economics & Logistics*, vol. 19, no. 4, pp. 749–764, Dec. 2017, doi: 10.1057/mel.2016.8.

[17] H. Prabowo and I. R. Afandy, "Intervention Analysis and Machine Learning to Evaluate the Impact of COVID-19 on Stock Prices," *Inferensi*, vol. 4, no. 1, pp. 13–27, Mar. 2021, doi: 10.12962/j27213862.v4i1.8626.

[18] R. Mulero and A. García-Hiernaux, "Forecasting Spanish unemployment with Google Trends and dimension reduction techniques," *SERIEs*, vol. 12, no. 3, pp. 329–349, Sep. 2021, doi: 10.1007/s13209-021-00231-x.

[19] S. J. Page, *Tourism Management*, Sixth. New York: Routledge, 2019.

[20] J. Fletcher, A. Fyall, D. Gilbert, and S. Wanhill, *Tourism: Principles and Practice*, Sixth. Harlow: Pearson, 2018.

[21] E. L. Rödel, "Forecasting tourism demand in Amsterdam with Google Trends: A research into the forecasting potential of Google Trends for tourism demand in Amsterdam," University of Twente, 2017. [Online]. Available: http://essay.utwente.nl/73929/

[22] B. H. Andrews, M. D. Dean, R. Swain, and C. Cole, "Building ARIMA and ARIMAX Models for Predicting Long-Term Disability Benefit Application Rates in the Public/Private Sectors," 2013. [Online]. Available: https://www.soa.org/resources/research-reports/2013/research-2013-arima-arimax-ben-appl-rates/

[23] R. J. Hyndman and G. Athanasopoulos, *Forecasting: Principles and Practice*, Third. Melbourne: OTexts, 2021. [Online]. Available: https://otexts.com/fpp3/

[24] S. Gössling, D. Scott, and C. M. Hall, "Pandemics, tourism and global change: a rapid assessment of COVID-19," *Journal of Sustainable Tourism*, vol. 29, no. 1, pp. 1–20, Jan. 2021, doi: 10.1080/09669582.2020.1758708.

[25] Wardiyanta, *Pengantar Ekonomi Pariwisata*. Yogyakarta: Pustaka Pelajar, 2020.

[26] M. Henseler, H. Maisonnave, and A. Maskaeva, "Economic impacts of COVID-19 on the tourism sector in Tanzania," *Annals of Tourism Research Empirical Insights*, vol. 3, no. 1, p. 100042, May 2022, doi: 10.1016/j.annale.2022.100042.