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

EKST

Vol. 6 No. 1 (2026) 523-532
https://publisher.uthm.edu.my/periodicals/index.php/ekst

Identifying Factors Affecting the Number of Singapore Tourists' Arrivals to Malaysia using Regression Analysis

Nur Farah Aliyah Mohd Fadzil[1], Rohayu Mohd Salleh[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, 84600 Pagoh, Muar, Johor, MALAYSIA.*

*Corresponding Author: msrohayu@uthm.edu.my
DOI: https://doi.org/10.30880/ekst.2026.06.01.058

Article Info

Received: 4 January 2026
Accepted: 6 June 2026
Available online: 9 July 2026

Keywords

Tourism, Correlation Analysis, MLR, ARDL, Long-Run Relationship, Breusch-Godfrey, Breusch-Pagan, Jarque-Bera, Ramsey RESET

Abstract

Tourism has played an important role in Malaysia's economic development in term of providing to employment, foreign exchange earnings, and the development of related industries. Malaysia's diverse attractions, cultural heritage, and strategic location make it a popular destination for international tourists, particularly those from Singapore. Tourism also benefits local communities by encouraging investment in infrastructure and services. However, the number of Singapore tourist arrival has been inconsistent in some years like in year 2019 until 2022 highlight the need to understand the key factors influencing Singaporean travel patterns. This study used quarterly data from 2010 to 2024. This study examines the factors influencing Singapore tourist arrivals and their long-term relationships. Correlation analysis, Multiple Linear Regression (MLR) and Autoregressive Distributed Lag (ARDL) method were used. The correlation analysis shows that exchange rate, GDP, hotel availability, internet penetration rate, crime rate, number of events and tourist receipts are significantly associated with number of Singapore tourist arrival while inflation rate shows insignificant relationship. From MLR analysis, it has been revealed that GDP, availability of hotels and receipts from tourists are significant factors to the total tourist arrivals. 81.10% of the changes in number of Singapore tourist arrival are explained by the independent variables. From long-run analysis by Autoregressive Distributed Lag (ARDL), it has been revealed that GDP, hotel availability and tourist receipts influence positively in the long run.

1. Introduction

Tourism is an important global economic sector that contributes to employment, income generation, and national development [1]. As international travel continues to recover after the COVID-19 pandemic, global tourist arrivals reached more than 1.3 billion in 2023 [2] and continued growing in 2024–2025. Tourism plays a similar role in Malaysia, where it contributes significantly to GDP and supports various related sectors such as hospitality, transportation, and retail. Malaysia's cultural diversity, strategic location, and accessibility make it a competitive destination for international visitors.

Among all source markets, Singapore consistently contributes the highest number of tourist arrivals to Malaysia. Geographic proximity, strong economic capacity, cultural familiarity, and multiple travel modes which include land, air, and sea make Malaysia a preferred destination for Singaporeans. Government initiatives through Tourism Malaysia, including marketing campaigns, collaborations with travel partners, and promotional activities, further enhance bilateral tourism flows [3]. Despite this, Singaporean arrivals have shown inconsistency over the years as the number of tourist arrival are not consistently increased. Past studies consistently highlight the importance of Singapore as Malaysia's largest source of international tourist driven by geographical proximity, strong transport connectivity and close socio-economic ties [16][17][18]. While prior research has examined the economic and social contributions of Singaporean tourists, particularly in border regions and tourism destinations, limited attention has been given to quantitatively analysing the key economic and tourism-related factors influencing Singapore tourist arrivals over time. These fluctuations raise the need to understand the factors that influence Singaporean travel to Malaysia. It suggests that both economic variables like exchange rate, GDP, inflation, tourist receipts. Also, non-economic variables like crime rate, number of events, and infrastructure availability may affect tourism demand.

This study aims to analyse which factors positively and negatively impact the number of Singapore tourists' arrival to Malaysia using correlation analysis. To identify which factor is significant to the number of Singapore tourists' arrivals to Malaysia using multiple linear regression. Lastly, it is to conduct the cointegration analysis between the factors and number of Singapore tourist arrivals to Malaysia by using the autoregressive distributed lag (ARDL) model. The scope of this study focuses specifically on Singapore due to its position as the largest contributor to Malaysia's inbound tourism. The analysis includes economic, social, and infrastructure-related variables using quarterly data from 2010 to 2024. Eight data points from 2020 Q2 to 2022 Q1 were removed due to extreme outliers caused by pandemic-related travel restrictions, which could distort statistical modelling. Data were obtained primarily from Tourism Malaysia and the Department of Statistics Malaysia.

2. Materials and Methods

The dataset used in this study was obtained from sources like Tourism Malaysia and the Department of Statistics Malaysia (DOSM). This data is secondary as the information was extracted from data generated by government agencies such as Tourism Malaysia https://data.tourism.gov.my/frontend/pdf/Malaysia-Tourism-Key-Performance-Indicators-2023.pdf and Department of Statistics Malaysia (DOSM) https://open.dosm.gov.my/data-catalogue/crime_district.

2.1 Data Description

Table 1 shows the variables used in this study. The dependent variable is the total number of Singapore tourists visiting Malaysia, while the independent variables include three main scope which are economic, social, and infrastructure factors.

Table 1 List of Variables

Variable	Type	Description	Factor Related	Abbreviation
Singapore tourist arrival	Dependent Variable	Total number of Singapore tourist arrival to Malaysia	-	TASG
Exchange rate	Independent Variable	Average exchange rate	Economic	ER
Tourist receipt	Independent Variable	Revenue from international tourists	Economic	TR
GDP	Independent Variable	Gross domestic product	Economic	GDP
Inflation rate	Independent Variable	Percentage change in consumer price	Economic	IR
Hotel availability	Independent Variable	Total number of rooms available	Infrastructure	HA
Internet penetration rate	Independent Variable	Percentage of population with internet access	Infrastructure	IPR
Crime rate	Independent Variable	Percentage of the total number of crimes	Social	CR
Number of events	Independent Variable	Number of events conducted by Tourism Malaysia	Social	NOE

Enhanced Knowledge in Sciences and Technology Vol. 6 No. 1 (2026) 523-532

2.2 Pearson's Correlation Coefficients

Pearson's correlation coefficient is a statistical measure that evaluates the degree of relationship between two continuous variables [4]. In this study, it was applied to achieve the first objective which is to determine whether the independent variables have a positive or negative impact on the number of Singapore tourist arrivals. The Pearson's correlation coefficient formula as in equation (1)

$$r = \frac{n(\sum xy) - \sum x \sum y}{\sqrt{n(\sum x^2) - (\sum x)^2 [n(y^2) - (\sum y)^2}} \tag{1}$$

where x is independent variable, y is dependent variable, n is the sample size and $\sum$ is the summation of all values.

Pearson's correlation coefficient, denoted as r, used the range value from –1 to 1. This value indicates the linear relationship between variables where +1 indicates the perfect positive linear relationship, where both variables increase together, -1 indicates the perfect negative linear relationship, where one variable increase and the other decreases and 0 indicates that it has no linear relationship between variables.

2.3 Multiple Linear Regression

Multiple linear regression is a simple extension of linear regression [5]. The multiple linear regression was used in this study to identify which factor is significantly affects the number of tourist arrivals from Singapore to Malaysia. This method allows the estimation of the relationship between one dependent variable and many independent variables. In this study, the multiple linear regression model assumes to be as in equation (2).

$$y_i = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n + \varepsilon_i \tag{2}$$

where $\boldsymbol{y_i}$ is the number of Singapore tourist arrival (dependent variable), $\boldsymbol{\beta_0}$ is the intercept, $\boldsymbol{\beta_n}$ is the coefficient, $\boldsymbol{x_i}$ is the independent variable and $\boldsymbol{\varepsilon_i}$ is the error. The significance of each variable was estimated by using p-values [6]. If p-value are less than 0.05, it suggests that the corresponding variable is significant to the dependent variable [7].

2.3.1 Normality of Residuals

Normality assumption in linear regression states that the error term or the residual in the model follows a normal distribution. Normality assumption is important not to violate in order to perform valid hypothesis testing and accurate estimation of the coefficients. The normality assumption was tested by using residual plots, histograms, and statistical tests. Any violation of normality can be addressed by using transformation [8].

2.3.2 Multicollinearity

Multicollinearity is a statistical case where two or more predictor variables in a multiple linear regression model are highly correlated [9]. There are many ways to identify multicollinearity, and one of them including using variance inflation factors (VIF). In this study, VIF was used to identify multicollinearity. VIF is a tool to measure how much the variance is inflated. If the value of VIF is more than 10, multicollinearity will be considered highly correlated [10].

2.3.3 Stepwise Regression

The stepwise regression technique is a statistical tool used in multiple linear regression to select the variables for a model. The concept of this technique was introduced by Efroymson in the year 1960, in which variables are assessed in a step-by-step manner regarding their significance to a particular model using p-value criteria [11]. The criteria determine whether a particular variable needs to be retained or deleted.

2.3.4 Autoregressive Distributed Lag (ARDL)

ARDL is one of the methods used to examine the cointegration of variables and estimate both short-run and long-run relationships among variables when variables are integrated at different levels (I (0) or I (1)). The long run coefficients derived from ARDL refers to the equilibrium relationship in which the variables move together over time after the fluctuations

2.3.5 Augmented Dickey-Fuller (ADF) Test

The ADF test is a statistical method used to check whether a time series variable is stationary or non-stationary. The null hypothesis of test represents the existence of the unit root (non-stationary), while the alternative hypothesis suggests that the unit root does not exist (stationary). The decision to accept or reject the null hypothesis is based on the p-value. If the p-value is less than 0.05, the null hypothesis will be rejected, and it will be concluded that the series is stationary. If a time series variable is non-stationary, it may lead to inaccurate results. ADF depends on the number of lags. It reduces the ability of the test to reject the null hypothesis as it requires estimation of more parameters when the lag increases and results in loss of degree of freedom [12].

The formula for the ADF test is shown in equation (3).

$$\Delta Y_t = \alpha + \beta t + y Y_{t-1} + \sum_{i=1}^{p} \delta_i \Delta Y_{t-i} + \epsilon_t \tag{3}$$

where, ΔY_t is the variable in the first period, α is the intercept term (constant), βt is the coefficient on the time trend, Y_{t-1} is lagged level of the dependent variable, y is the coefficient of the lagged level of time series, $\sum_{i=1}^{p} \delta_i \Delta Y_{t-i}$ is the lagged difference term to correct the serial correlation in error term and ϵ_t is the error term.

2.3.6 Akaike Information Criterion

The Akaike Information Criterion (AIC) is a statistical tool that measures the relative quality of statistical models for a given dataset. It balances the model's fit and complexity by penalising the number of parameters included. Hirotugu Akaike introduced AIC in 1973 as an extension of the maximum likelihood principle [13]. The AIC is defined as in equation (4).

$$AIC = -2\ln(L) + 2k \tag{4}$$

where *ln(L)* is the logarithm of the likelihood function and k is the number of estimated parameters in the model.

2.3.7 Bayesian Information Criterion

Bayesian Information Criterion (BIC), also known as Schwarz Bayesian Criterion (SIC), is a statistical measure used to select a model from a finite set of models [14]. It also balances the model fit and complexity, but it imposes more substantial penalty for additional parameters. The BIC is defined as in equation (5).

$$BIC = -2\ln(L) + k\ln(n) \tag{5}$$

where L is the likelihood of the model given the data, k is the number of parameters in the model and n is the number of data points.

2.3.8 ARDL Bounds Testing Procedures

The ARDL bound testing procedure developed by [15] determines whether a long-run relationship exists among the variables. The bound test involves estimating the unrestricted error correction model (UECM). The step also involves conducting an F-test to examine the significance of the lagged level variables. The null hypothesis represents that no long run relationship exists, and the alternative hypothesis represents an existence of long-run relationship. The F-statistic value will be computed and compared to the critical value to know whether to accept or reject the null hypothesis.

2.3.9 Estimation of Long-Run Coefficients

The long-run coefficients are obtained from the ARDL model using the appropriate lag structure determined earlier. Following this, the Error Correction Model (ECM) is developed to show the short-run changes in the variables while staying aligned with the long-run equilibrium [15].

2.3.10 Diagnostic Checking

Diagnostic checking in ARDL helps to ensure that the model is valid and correctly specified. Few diagnostic tests were used to test the assumptions of the model. The Breusch-Godfrey test is used to test for the existence of serial correlation in the residuals, where the null hypothesis assumes no autocorrelation. On the other hand, the Breusch-Pagan test was used to test for heteroskedasticity, where the null hypothesis assumes that the residuals have constant variance.

Additionally, the Jarque-Bera test was performed to check if the residuals are normally distributed, with the null hypothesis suggesting normality. The Ramsey RESET test was employed to check if the model is properly specified, with the null hypothesis suggesting no specification error. Lastly, the CUSUM and CUSUMSQ tests were used to check the stability of the model coefficients and the residual variance. In all diagnostic checking, the results must be within acceptable limits which is p-value > 0.05 or plots within boundaries indicate that the ARDL model is valid and well-specified.

3. Results and Discussions

3.1 Correlation Analysis on Singapore tourists' arrival to Malaysia

Table 2 Result of correlation analysis

Variable	Correlation (*r*)	*p*-value
Exchange rate	0.687	0.00*
Tourist receipt	0.856	0.00*
GDP	-0.787	0.00*
Inflation rate	0.070	0.624
Hotel availability	-0.595	0.00*
Internet penetration rate	-0.821	0.00*
Crime rate	0.826	0.00*
Number of events	0.281	0.043*

Asterisk (*) shows a statistically significant relationship

Table 2 shows the result of correlation analysis. The correlation analysis between Singapore tourist arrivals to Malaysia and the selected independent variables shows both positive and negative relationships. Variables with positive correlations include exchange rate, crime rate, number of events, and tourist receipts. The exchange rate shows a moderate positive correlation ($r = 0.678$, $p < 0.05$), meaning that when the Malaysian Ringgit depreciates against the Singapore Dollar, Malaysia becomes more affordable, leading to higher tourist arrivals. Crime rate appears to have a very strong positive relationship with tourist arrivals ($r = 0.826$, $p = 0.00$), possibly because of the rise in tourism-related activities, thus providing opportunities for committing crime. The number of events also shows a positive correlation ($r = 0.281$, $p < 0.05$), implying that events and activities are the reason of Singaporean tourists because of their proximity and shared cultural interests. Tourist receipts also indicate a strong positive relationship ($r = 0.856$, $p < 0.05$), indicating that higher tourist arrivals naturally lead to increased spending.

On the other hand, GDP, hotel availability, and internet penetration rate show negative relationship with Singapore tourist arrivals. Malaysia's GDP has a strong negative correlation ($r = -0.787$, $p < 0.05$), which may be due to rising domestic prices during economic growth, making travel a bit pricey. Hotel availability shows a moderate negative correlation ($r = -0.595$, $p < 0.05$), possibly because the availability of hotels outstrips the demand or the tourists prefer alternatives like homestays and Airbnb. Internet penetration rate has a strong negative correlation ($r = -0.821$, $p < 0.05$), which could reflect shifts toward digital lifestyles or changing travel patterns. Meanwhile, inflation rate shows no significant relationship ($r = 0.07$, $p = 0.624$), indicating that price level changes do not strongly influence Singaporean tourists' travel decisions.

3.2 Significant factors influencing Singapore tourist arrivals to Malaysia using Multiple Linear Regression

3.2.1 Normality Test of Residual

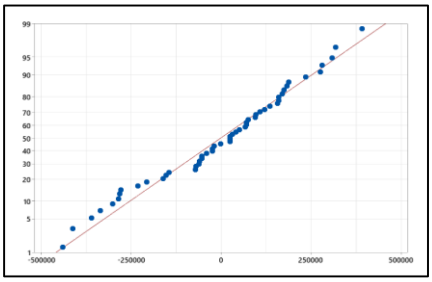

Fig. 1 Residual vs fitted plot

From Fig. 1, it shows that the residual slightly deviate from the straight line at both ends but the overall pattern follows the normal line, which indicating that the residuals are normally distributed. The number of sample size *n*=52 makes the multiple linear regression analysis robust to minor deviates from normality. Thus, the normality assumption is considered satisfied.

3.2.2 Multicollinearity Test

Table 3 Multicollinearity test

Variable	VIF before elimination	VIF after first elimination	VIF after second elimination	VIF after third elimination
Exchange rate	27.57	21.48	-	-
GDP/RM Mil.	21.61	19.05	14.76	7.21
Inflation rate	1.09	1.08	1.03	1.03
Crime rate	24.77	18.52	17.81	-
Hotel availability	12.18	12.18	6.31	5.35
Tourist receipt	5.16	3.97	3.25	3.03
Internet penetration rate	55.65	-	-	-
Events	2.07	1.94	1.80	1.63

Table 3 shows the result from multicollinearity test. From the initial output where value of VIF before elimination, several variables showed a high variance inflation factor (VIF) value which indicate strong multicollinearity among certain variables. Therefore, the variable with the highest VIF value was removed and rerun new model. This step is repeated until all variables achieved VIF value less than 10. After the third eliminations or adjustment, the values of VIF for all remaining variables reduced to a number less than 10, indicated that the issue of multicollinearity was resolved.

3.2.3 Stepwise Selection

The stepwise selection analysis was conducted to determine the most significant variables affecting the number of Singapore tourist arrivals to Malaysia. The final model retained only in the model below:

TASG = 1551224 – 0.987 GDP + 61153 IR + 268.6 HA + 44.49 TR

Table 4 Stepwise selection

Variable	*p*-value
GDP/RM Mil.	0.00
Inflation rate	0.073
Hotel availability	0.00
Tourist receipt	0.00

Table 4 shows the *p*-value of the selected variables from stepwise selection. Among these variables, GDP/mil. shows a negative coefficient which suggests an inverse relationship, indicating that higher GDP in Malaysia not necessarily lead to more tourist arrivals from Singapore. Inflation rate was removed from the model as *p*-value is more than 0.05.

3.2.4 Model Validation

Model validation was conducted to determine the predictive accuracy and reliability of the regression model selected using Best Subset Regression and Stepwise Selection. The dataset was divided into two parts which are 70% training data and 30% testing data.

Table 5 Model performance error metrics

Dataset	Error Metrics	Value
Training	MSE	$5.385 * 10^{10}$
Testing	MSPR	$9.146 * 10^{10}$
Ratio		1.7

From Table 5, the training set had a Mean Squared Error (MSE) of 53.85 billion, whereas the testing set had a higher Mean Squared Prediction Error (MSPR) of 91.46 billion. The prediction error for the testing data exceeds that of the training data, indicating greater uncertainty when forecasting unseen observations. However, the ratio between training and testing data errors is approximately 1.7 which is within an acceptable range, indicating that the model maintains reasonable predictive accuracy while minimizing significant overfitting. Overall, the validation results show that the model can explain variations in Singaporean tourist arrivals to Malaysia.

3.2.5 Model Significance

The overall significance of multiple linear regression was assessed using Analysis of Variance (ANOVA) table, which evaluates whether the independent variables have a statistically significant effect on the dependent variable. The overall regression model is statistically significant as the F-value of regression is high which is 68.66, and p-value are less than threshold. This suggests that the selected explanatory variables jointly have a significant effect on variations in tourist arrivals from Singapore to Malaysia. Overall, the model shows a good fit of model.

The goodness-of-fit results show that the Multiple Linear Regression model performs well in explaining and predicting Singapore tourist arrivals to Malaysia. The R-squared value of 81.10% indicates that most of the variation in tourist arrivals can be explained by the variables included in the model, showing a good overall fit. The adjusted R-squared remains high at 79.92% after taking into account the number of explanatory variables, suggesting that the selected variables are relevant and not unnecessarily added to the model. In addition, the predicted R-squared of 77.09% is close to the adjusted R-squared, which shows that the model is able to predict new data reasonably well and does not suffer from serious overfitting.

3.2.6 Final Model Interpretation

From the output, the final model is:

TASG = 1729551 -0.980 GDP + 260.9 HA + 43.84 TR

The estimated regression model indicates that Singapore tourist arrivals to Malaysia (TASG) are influenced by several economic and travel-related factors. The constant term of 1, 729, 551 represents the estimated number of tourist arrivals when all the other variables are equal to zero. For GDP, the coefficient is -0.98 which means 1% increase in GDP is associated with decrease of 0.987 Singapore tourist arrivals.

In addition, the coefficient of hotel availability (HA) is 260.9, implying that an increase in hotel availability leads to a rise in tourist arrivals. The travel-related variable (TR) also shows a positive relationship, with a coefficient of 43.84, meaning that higher values of this variable are associated with increased tourist arrivals from Singapore to Malaysia. Overall, the model suggests that hotel availability, interest rates, and travel-related factors have a positive effect on Singapore tourist arrivals, while GDP shows a negative relationship, and the model is useful for explaining variations in tourist arrivals from Singapore to Malaysia.

3.3 Long-term effects on number of Singapore tourist arrival to Malaysia

3.3.1 The Augmented Dickey Fuller (ADF) test

The Augmented Dickey Fuller (ADF) was applied to test the stationarity of the number of tourist arrivals and other eight variables that affects number of Singapore tourist arrivals. The ARDL model can handle the mix of stationary at two level which are stationary at level I(0) and stationary after the first differencing I(1). The hypothesis claimed that the variables that the unit root does not exist means that the variables are stationary. The result of p-value of unit root test for each variable by using ADF shown in Table 6.

Table 6 Result of ADF Test of Each Variable

Variable	Level form, I(0)	First differencing form, I(1)	Stationarity
Tourist arrival	0.825	0.003*	I(1)
GDP/RM Mil.	0.975	0.000*	I(1)
Hotel availability	0.158	0.001*	I(1)
Tourist receipt	0.424	0.001*	I(1)

Asterisk (*) denoted as p-value less than 0.05 significance level

From Table 6, it shows the result of test statistics value of ADF test. Based on the findings, all are stationary after first differencing, I(1). As its p-value were all at level less than 0.05. In ARDL, only the variables that were stationary at level form, I(0) or after first differencing, I(1) can be used in ARDL bound test. Then all variables were used to proceed with ARDL bound test.

3.3.2 Lag Length Selection

The selection of an appropriate lag length is a crucial step before estimating ARDL model as it ensures the dynamic relationships among variables are accurate. In this study, optimal lag length was determined using Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC).

The optimal lag length for ARDL model is selected as (3, 1, 3, 3) which includes variable number of tourist arrival, hotel availability, GDP and tourist receipt accordingly. This combination results in the lowest AIC value of 1348.91 and BIC value of 1377.29, which suggested that the model is the most appropriate lag structure for ARDL bound test.

3.3.3 ARDL Bound Testing

The ARDL bound test approach was employed to examine the existence of long-run relationship among the variables in the model. Also, it determined whether a cointegrating relationship exists among the dependent and independent variables.

Table 7 ARDL Bound Test Results

Critical Value (%)	Lower Boundary, I(0)	Upper Boundary, I(1)	ARDL model	*F*-statistic
1	3.15	4.43		
5	2.62	3.79	(3, 1, 3, 3)	22.117
10	2.27	3.38		

Based on the Table 7, the ARDL model with lag structure (3, 1, 3, 3) was estimated. The *F*-statistic value obtained was 22.117. The results then compared with upper boundary, I(1) at three different level of critical value bounds provided by [15]. If the *F*-statistic is more than the upper boundary value, it indicates the presents of cointegration among the variables. From the table 7 above, the calculated *F*-statistic has greater value than the upper bound critical value at all significance level. Thus, the null hypothesis of no long-run relationship is rejected. Therefore, it can be concluded that a long-run cointegration relationship exists between the number of Singapore tourist arrivals and the four selected independent variables.

3.3.4 Estimation of Long-Run Coefficients

After confirming the presence of a long-run relationship among the variables through the ARDL bounds test, the next step is to estimate the long-run coefficients to identify how each independent variable influences the number of Singapore tourist arrivals to Malaysia in the long-run. The results are presented in Table 8.

Table 8 Coefficient of the Long-Run Effect

Variable	Long-run effect
GDP/RM Mil.	1.7351
Hotel Availability	699.1339
Tourist receipt	113.2226

Based on Table 8, GDP, hotel availability, and tourist receipt all show positive long-run relationships with Singapore tourist arrivals. An increase of one unit in GDP is associated with an increase of approximately 1.7 tourists, while a one-unit rise in hotel availability increases number of arrivals by approximately 699. Lastly, tourist receipts have a positive long-run impact. A one-unit increase in tourist receipts is associated with an increase of around 113 tourists, indicating that higher spending levels support sustained tourist interest from Singapore.

3.3.5 Diagnostic Checking

Diagnostic checking was conducted to verify whether the ARDL model satisfies the assumptions of regression analysis or not.

Enhanced Knowledge in Sciences and Technology Vol. 6 No. 1 (2026) 523-532

Table 9 Diagnostic Check Result

Assumption	Test	*p*-value
Serial Correlation	Breusch-Godfrey	0.0001
Heteroscedasticity	Breusch-Pagan	0.0003
Normality	Jarque-Bera	0.8542
Model specification	Ramsey RESET	0.0065

The Breusch-Godfrey test shows a *p*-value of 0.0001, which shows the presence of autocorrelation in the residuals. Similarly, the Breusch-Pagan test returns a *p*-value of 0.0003, which shows the existing of heteroscedasticity, meaning the residual variance is not constant over time. The Ramsey RESET test produces a p-value of 0.0065, implying possible model specification errors or omitted variables. Meanwhile, the Jarque-Bera test indicates that the residuals are normally distributed since the *p*-value of 0.8542 exceeds 0.05. In conclusion, despite the normality assumption being valid, the presence of autocorrelation, heteroscedasticity, and model specification errors shows that the model is not accurate and thus requires the use of robust standard errors to make it more precise.

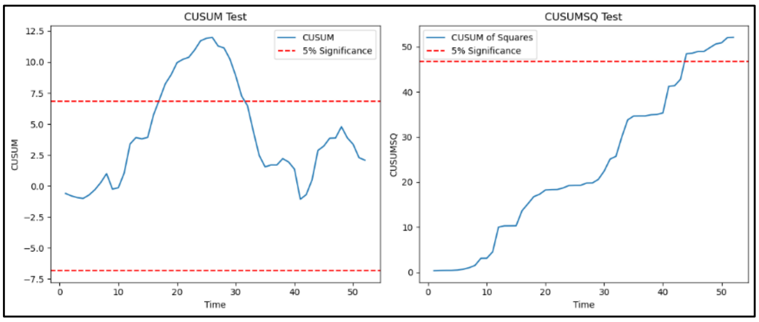

Fig. 2 CUSUM and CUSUMSQ Test

CUSUM and CUSUMSQ tests assess the stability of coefficients and the residual variance over time. Based on Fig. 2, the CUSUM test plot is crosses above the upper bound around the middle of the series. This suggests that there is some coefficient instability in the data suggesting structural break. This instability is likely due to major disruptions in year 2019 until 2021 which was during COVID-19 pandemic, especially when the international borders between Singapore and Malaysia are closed. On the other hand, the CUSUM of Squares (CUSUMSQ) Test reveals that the cumulative sum of the residuals gets closer to the 5% significance level towards the later period of the sample data. This particular assumption points to the structural instability of the variance of the residuals. The reason behind this instability could be the effect of the structural disruptions in tourism activities that resulted from the pandemic and border closures. Thus, the model coefficients are stable, but the variance of the error term is not stable over time.

Overall, these results suggest that although the model is able to capture the overall relationship, the presence of events such as the COVID-19 pandemic and border closures has resulted in structural breaks and volatility that have impacted the stability of the coefficients and variance. The behaviour of the CUSUM and CUSUMSQ plots reflects the inherent volatility of the data, particularly during periods of unprecedented shocks such as the COVID-19 pandemic, indicating that the results are consistent with real-world conditions rather than model misspecification.

4. Conclusion

The results of the Pearson's correlation, MLR, and ARDL analyses provide several significant factors that affect number of Singapore tourist arrivals to Malaysia. The positive results of the economic variables, tourist earnings, safety, and events imply that the more favourable the economy and the more interesting the events offered are, the more tourists visiting the country will be. Conversely, the negative results of the structural and infrastructure variables of hotels and the availability of the internet imply the lessening of tourist arrivals if there are more of such variables. The result of the inflation variable is insignificant, while the diagnostic tests imply the presence of problems of serial correlation, heteroscedasticity, and possibly model specification errors. This suggests a need to refine the model or use more robust techniques. Given these results, it is recommended that marketing campaigns and online promotions be improved to attract more tourists from Singapore. The government must focus on monitoring trends involving the tourism and economy of Malaysia to identify the most appropriate time for promotions. The government must ensure the price of hotel rooms is competitive. Also, there could be a need for future studies involving qualitative variables such as the effect of influence, cultural appeal, and customer satisfaction as this study is limited to quantitative variables only.

Acknowledgement

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

Conflict of Interest

Authors declare that there is no conflict of interests regarding the publication of the paper.

Author Contribution

The authors confirm contribution to the paper as follows: ***study conception and design:*** *Nur Farah Aliyah Mohd Fadzil, Rohayu Mohd Salleh;* ***data collection:*** *Nur Farah Aliyah Mohd Fadzil;* ***analysis and interpretation of results:*** *Nur Farah Aliyah Mohd Fadzil, Rohayu Mohd Salleh;* ***draft manuscript preparation:*** *Nur Farah Aliyah Mohd Fadzil, Rohayu Mohd Salleh. All authors reviewed the results and approved the final version of the manuscript.*

References

[1] R. K. Modi, "Economic Contribution and Employment Opportunities of Tourism and Hospitality Sectors," in *Economic contribution and employment opportunities of tourism and hospitality sectors.*, 2024, pp. 293–306. doi: 10.1108/978-1-83753-708-220241015.

[2] "UN Tourism World Tourism Barometer | Global Tourism Statistics." https://www.untourism.int/un-tourism-world-tourism-barometer-data

[3] M. T. Statistics, "Malaysia Tourism Statistics." https://data.tourism.gov.my/

[4] W.-M. W. Hwu, *GPU Computing Gems Emerald Edition*. 2011. doi: 10.1016/c2010-0-65709-9.

[5] N. R. Smalheiser, "Correlation and other concepts you should know," in *Elsevier eBooks*, 2017, pp. 169–185. doi: 10.1016/b978-0-12-811306-6.00013-0.

[6] J. F. Hair Jr., W. C. Black, B. J. Babin, and R. E. Anderson, *Multivariate data analysis*, Eighth. 2019.

[7] A. Field, *Discovering statistics using IBM SPSS Statistics*. SAGE Publications Limited, 2018.

[8] P. Mishra, C. M. Pandey, U. Singh, A. Gupta, C. Sahu, and A. Keshri, "Descriptive statistics and normality tests for statistical data," *Annals of Cardiac Anaesthesia*, vol. 22, no. 1, p. 67, Jan. 2019, doi: 10.4103/aca.aca_157_18.

[9] J. I. Daoud, "Multicollinearity and regression analysis," *Journal of Physics Conference Series*, vol. 949, p. 012009, Dec. 2017, doi: 10.1088/1742-6596/949/1/012009.

[10] Schreiber-Gregory and Bader, *Logistic and linear regression assumptions: violation recognition and control*. 2018. [Online]. Available: https://www.researchgate.net/publication/341354759_Logistic_and_Linear_Regression_Assumptions_Violation_Recognition_and_Control#fullTextFileContent

[11] M. A. Efroymson, "Multiple Regression Analysis," In: A. Ralston and H. S. Wilf, Eds., Mathematical Methods for Digital Computers, John Wiley, New York, 1960.

[12] S. M. Hosseini, Z. Ahmad, and Y. W. Lai, "The role of macroeconomic variables on stock market index in China and India," *International Journal of Economics and Finance*, vol. 3, no. 6, Oct. 2011, doi: 10.5539/ijef.v3n6p233.

[13] J. E. Cavanaugh and A. A. Neath, "The Akaike information criterion: Background, derivation, properties, application, interpretation, and refinements," *Wiley Interdisciplinary Reviews Computational Statistics*, vol. 11, no. 3, Mar. 2019, doi: 10.1002/wics.1460.

[14] G. Schwarz, "Estimating the dimension of a model," *The Annals of Statistics*, vol. 6, no. 2, pp. 15–18, Mar. 1978, doi: 10.1214/aos/1176344136.

[15] M. H. Pesaran, Y. Shin, and R. P. Smith, "Pooled mean group estimation of dynamic heterogeneous panels," *Journal of the American Statistical Association*, vol. 94, no. 446, pp. 621–634, Jun. 1999, doi: 10.1080/01621459.1999.10474156.

[16] Puah, C.-H., Thien, F.-T., & Affendy Arip, M. (2014). *Singaporean Demand for Tourism in Malaysia* (Unpublished bachelor's project). Universiti Malaysia Sarawak. https://ir.unimas.my/id/eprint/18019/1/Singaporean Demand for Tourism in Malaysia.pdf

[17] Ch'ng, L. G. (2010). Factors influencing the intentions of Singaporean tourists to revisit Malaysia (Master's thesis, Universiti Sains Malaysia). USM Institutional Repository. https://eprints.usm.my/28823/1/Factors influencing the intentions of Singaporean tourists to revisit Malaysia.pdf

[18] [18] Hampton, M. P. (2010). Enclaves and ethnic ties: The local impacts of Singaporean cross-border tourism in Malaysia and Indonesia. Singapore Journal of Tropical Geography, 31(2), 239–253. https://doi.org/10.1111/j.1467-9493.2010.00393.x