Malaysian Journal of Mathematical Sciences 16(2): 175–184 (2022)
https://doi.org/10.47836/mjms.16.2.01

Malaysian Journal of Mathematical Sciences

Journal homepage: https://einspem.upm.edu.my/journal

Finite Mixture Model: Prediction of Time Series Data Using Bayesian Method

Phoong, S. Y.*[1], Phoong, S. W.[2], and Phoong, K. H.[3]

[1]*Department of Mathematics, Faculty of Science and Mathematics, Universiti Pendidikan Sultan Idris, Malaysia*
[2]*Department of Management, Faculty of Business and Economics, Universiti Malaya, Malaysia*
[3]*Institute for Advanced Studies, Universiti Malaya, Malaysia*

E-mail: phoong@fsmt.upsi.edu.my
**Corresponding author*

Received: 14 February 2020
Accepted: 26 November 2021

Abstract

The aim of this study is to measure the number of components that exhibits from the variables' series. The number of components can be affected by the time series components including trend, seasonal adjustment, and irregular changes. By using a finite mixture model, the number of components can be identifies and thereafter we can formulate a Bayesian regression equation to predict the relationship between exchange rate and international tourism expenditure in Malaysia. Identification of the number of components is an important step to weigh the probability density function for a time series data. The weight of the probability density function is then used for prediction. Besides, a Bayesian method is also used in this study to fit with the finite mixture model due to its consistency characteristic. The Bayesian parameter estimates are close to the predictive distributions because it will integrate the prior distribution with the likelihood function to produce posterior distribution. The results show that there is a two-component normal mixture model exists for the time series data. In addition, a prediction equation is obtained from the analysis.

Keywords: finite mixture model; Bayesian method; prior distribution; likelihood function; posterior distribution.

1 Introduction

In-bound tourism expenditure is considered as a significant source for economic growth [13]. These expenditures generate employment, and foreign exchange revenues and income. According to [8], international tourism expenditure can even function as an alternative to improve the balance of payments through foreign exchange revenues. Dritakis [8] further elaborated that rapid development of tourism has multiplier effect towards household incomes growth, government revenues growth, and balance of payments. Therefore, the expansion of tourism is considered as a positive contribution towards economic growth.

International tourism has been growing and becoming more integral in many countries around the world [1]. As reported by [4], the contribution from the tourism sector reached RM84.1 billion to the Malaysian economy in year 2018. This accounts for approximately 6% of Malaysia total Gross Domestic Product (GDP) which is USD 354.35 billion (approximately RM 1417.40 billion). The largest contributed was the ASEAN short haul market, accounting for RM48.5 billion of the total tourism expenditure. On the other hand, the non-ASEAN market contributed RM35.6 billion in tourism expenditure. The expenditures were concentrated mainly on the shopping segment (33.4 percent) followed by accommodation (25.7%) and food and beverages (13.4%).

However, the tourist arrivals fell 2.5% in third quarter of 2017 due to a lack of advertising and promotional activities as reflected by the lower budget allocated for 2016 and 2017. The reduction in arrivals not only affect the tourism expenditure, it also affect the GDP in Malaysia. In addition, the two airway incidents and security issues in Malaysia also caused the declining of tourism expenditure and arrivals in 2015 [16]. According to [17], the crime issues in Malaysia such as purse snatching, auto-mobile theft, fraud, and residential burglaries reported high quantity in 2014 and 2015. All these crime issues might be the concern of tourist when choose the tourist destinations.

With careful thoughts, this brings to surface another question, whether the expansion of the economic contributed to the development of tourism or tourism growth contributed to economic expansion instead in the case of Malaysia. According to tourism-led growth hypothesis (TLGH), tourism growth contributes to economic growth. This is supported by the notion that tourism increases employment, household incomes, and foreign exchange revenues. Due to the significant role of tourism expenditure towards economic growth, this question has become an essential empirical problem. Many previous studies have been conducted to identify the relationship between tourism growth and economic growth in many countries. However, the empirical findings were contradictory with one another even within the same economy [3].

Economic growth is normally estimated in terms of Gross Domestic Product (GDP), but this study will use exchange rate as a proxy for economic growth for several reasons. Firstly, according to [3], the relationship between the exchange rate and economic growth is not spurious. Real exchange rate (RER) and exchange rate alignment are critical conditions to enhance economic performance especially in Asian countries. Secondly, several past studies were conducted in the Malaysian context however exchange rate was not used as the indicator of economic growth [10], but the strong relationship between real exchange rate and economic behavior is evident in Asian countries [6].

The exchange rate plays essential role since it will influence the variables' price of all things related to international trade and investment. It is one of the essential indicators of a country's relative level of economic health. Hence, the objectives of this study is to (i) identify the number of components used to investigate the exchange rate and international tourism expenditure in

Malaysia, and (ii) formulate the regression equation for prediction using Bayesian method.

2 Literature Review

Previous researches have been executed to analyze the relationship between tourism growth and economic growths in several countries through the empirical findings were contradictory even within the same country.

One of the most notable studies was executed by [2]. Their studies identify the validity of TLGH in Spain. Results showed that tourism expenditure is causal towards Spain's economic growth unidirectionally. Johansen co-integration technique was applied due to non-stationary variables. The findings showed that there is a long-run relationship among the development of economic and tourism expansion while the Granger causality test confirms the causality of that relationship.

Another notable study was conducted by [8]. The findings revealed that there is a strong effect on international tourism expenditure and economic growth in Greece during the period 1960-2001. Both the co-integration technique and the Granger causality tests were applied in the Dristakis [8] studies. The co-integration analysis suggested the presence of a cointegration relationship between international tourism expenditure, real effective exchange rate, and real GDP. This indicated the existence of a common trend or long-run relationships among the three variables. On the other hand, Granger causality tests indicated a strong causal relationship between international tourism expenditures and economic growth. Other studies arrived at similar. Other notable studies have also found TLGH to be empirically supported in Turkey and Taiwan where the empirical results suggested that there are bidirectional causal relationships between the international tourism expenditure and economic growth in both the short and long-run [5].

However, some studies do not agree with the findings stated above. Katircioglu [9] investigated TLGH in Turkey during the period 1960 to 2006. The author suggested that TLGH was not applicable in Turkey due to a lack of cointegration among the variables. Oh [14] could not verify TLGH in the Korea economy as well. TLGH is found not applicable to all countries. Meanwhile, the exchange rate plays an essential role in economic growth since it will influence the variables' price of all things related to international trade and investment. Hence, this study is focussed on the identification of components that exists between the exchange rate and tourism expenditure in Malaysia. In addition, both the exchange rate and international tourism expenditure are essential to affect the economic growth in a country, therefore, a regression Bayesian equation is formulated to predict future observations.

Other studies discussed about the impact of the exchange rate on the tourism sectors such as [12]. Most of these studies investigate the impact of the exchange rate on tourism sectors (such as demand and supply), but very few of it investigate expenditure of the international tourism.

3 Methodology

3.1 Unit Root Test

The unit root test is an important test that aims to identify whether the sample data is non-stationary or stationary. A non-stationary data might lead to invalid analysis; therefore, a differencing process is used to transform the sample data into stationary. In this study, the unit root tests include Levin, Lin, and Chu (LLC), Im, Pesaran and Shin W-stat (IPS) test, Augmented Dickey Fuller (ADF) and Phillips–Perron (PP) test are used to measure the stationarity of the time series data. All these unit root tests have the same hypothesis that is:

$$H_0 : \text{There is a unit root,}$$
$$H_1 : \text{There is no unit root.}$$

The unit root tests mentioned above are important to measure the integration order of the time series data. However, there is a slightly difference between it. LLC test is a panel unit root test that combined the data from different individuals into a single final regression. Meanwhile, the IPS test unlike LLC test, this test allows heterogeneous coefficients. For ADF test, it allows a large and complicated time series model being analyzed and PP test changes non-parametric data into t-test statistics.

3.2 Finite Mixture Model

The finite mixture model (FMM) is a mixture of finite dimensional probabilistic model where the parameters are from different subpopulations. This model is a model based approach that used to measure the unobserved subpopulations. FMM provides efficient findings when there are two or more different subpopulations are adopted. Additionally, the finite mixture model play role in data mining [15]. This technique is widely applied in human, social, and behavioural science data in which to cluster the analysis.

Apart from that, FMM is flexible since it can model heterogeneous data with a large sample size. According to [11], the mixture of discrete and continuous data can be effectively computed by using FMM, in addition to examining group structure data. Furthermore, FMM is plausible in measuring real financial time series data due to its flexibility properties.

The finite mixture model also plays a vital role in analyzing a combination of different distributions such as nominal, ordinal, frequency responses, binary etc. This model is also used to measure the heavy-tail densities, zero-inflated, or hurdle models. For FMM, there is a k components that exists in the mixture model. The number of components aims to identify how many unobservable types exist for a dataset.

After identifying the k components, a mixture of distributions is applied for further analysis. Let x_1, x_2,..., x_n be independent, identically distributed (IID) with k-dimensional observation, the probability density function is

$$f(x;\pi) = \Sigma_{c=1}^{C} \pi_c \, f_c(x), \tag{1}$$

where π_c denotes the cth mixing probability of x_i, $f_c(x)$ is the c^{th} mixing density and C represents total number of components.

Assume that the probability density function with two components normal distribution can be written in the form

$$f(x) = \pi\phi_1(\mu_1, \sigma_1^2) + (1-\pi)\phi_2(\mu_2, \sigma_2^2), \tag{2}$$

where $\phi_i(\mu_i, \sigma_i^2)$ represents the probability density function (PDF) of a normal distribution with mean, μ and variance, σ^2. Symbol π denotes the weight of data for different components and the entire weights for all k component of the finite mixture models are equal to one.

The advantage of applied FMM in the analysis is to restore the actual observations and plays a role as a hierarchical model. In addition, FMM is used to minimize the experiment-wise error. This analysis will optimize the objective functions into the one-step approach. Lastly, this model is flexible in analysing variables with different scales.

3.3 Bayesian Method

Bayesian method is developed from the idea of Bayes' rule. This method is mainly used to measure the parameters of the variables and predict future observation [7]. There are three main inferences for this method includes prior probability, likelihood, and posterior probability.

Prior probability is the unconditional probability distribution that is assorted before it is measured. While likelihood function is the new information provided. The posterior distribution is the probability of an event that is obtained after the combination of prior probability and likelihood function.

$$\text{Posterior probability} \quad \alpha \quad \text{Prior probability} \times \text{Likelihood} \tag{3}$$

The posterior distribution is important for the Bayesian method because this distribution provides the interval estimates and point estimates for parameters, make a prediction, and evaluate the probabilistic for the hypothesis. When the Bayesian method fitted with the finite mixture model, the equation can be expresses as

$$P_c(x_i \mid \theta_c) = \Sigma_{c=1}^{C} \pi_c \; f_c(x_i \mid \mu_c, \Sigma_c), \tag{4}$$

where we assume that N observation $x_i, i = 1, ..., N$.

The advantage of using this method is to examine the small sample size of data and missing data. The Bayesian method is also used to analyse the hierarchical models and nested data, in addition to providing expected values that are closed to the observed values.

3.4 Sample and Data

In this paper, the time series data used are exchange rate and international tourism expenditure in Malaysia from 1995 to 2017. Annually data is used for the analysis. The unit for the exchange rate is U.S. per Dollar while for international tourism expenditure is in current U.S. per Dollar.

4 Results and Discussion

4.1 Unit Root Test

A group unit root tests on the exchange rate and international tourism expenditure is applied to examine the stationarity of the datasets for the years 1995 to 2017. Table 1 shows that all the unit root tests include LLC test, IPS, ADF and PP test fail to reject the null hypothesis that there is a unit root that exists. Thus, a differencing process needs to transform the non-stationary data into stationary.

Table 1: Unit root test and first differencing process.

Method	Level Statistics	Level p-value	First Difference Statistics	First Differemce p-value
LLC	0.80561	0.7898	-4.89864	0.0000
IPS	0.67951	0.7516	-3.73153	0.0001
ADF	5.16991	0.2703	19.7885	0.0005
PP	2.46851	0.6503	19.8525	0.0005

Table 1 also shows that both the exchange rate and international tourism expenditure data are stationary after the first differencing process. This can be referred to a probability value where all the unit root tests are less than 0.05. Hence, we reject the null hypothesis that there is a unit root exists.

4.2 Finite Mixture Model

After identifying the stationarity of the data, FMM is fitted with the Bayesian method to analyze the datasets. The Bayesian method is applied to investigate the mixture model of the exchange rate and international tourism expenditure in Malaysia. Table 2 shows that the exchange rate is the response variable for the analysis and a homogeneous regression mixture model is implemented in the analysis. In addition, the estimation method adopted is the Markov Chain Monte Carlo (MCMC) with two components obtained. In this study, the two components obtained represents travel and tourism trend and non-trend situation. The trend situation occurs include festivals, holidays, special event tours, etc.

Table 2: Information about model.

Dependent Variable	ExchangeRate
Model	Homogeneous Regression Mixture
Distribution	Normal
Components, k	2
Estimation Method	Markov Chain Monte Carlo

Table 3 revealed information about the Bayesian analysis. A conjugate sampling algorithm is applied with seven parameters in the sampling. The seven parameters obtained include four mean functions, two scales, and a mixing probability.

Table 3: Bayes information.

Sampling Algorithm	Conjugate
Mean Function	4
Scale	2
Mixing Probability	1

Additionally, Table 4 describes the distribution, estimate value for both the prior and posterior distributions of the time series data. Results show that there is a normal distribution which associated with the exchange rate and tourism expenditure. Meanwhile, the variance parameter in prior distribution is inverse gamma distribution and Dirichlet distribution for mixing probability. The mixing probability of posterior distribution shows that the weight for component 1 is 0.9563 while component 2 is 0.0437. Therefore, the equation for different components is $y = 0.9563f_1 + 0.0437f_2$ where f_1 and f_2 denotes the PDF for components 1 and 2.

Table 4: Prior and posterior distributions.

Component	Effect	Distribution	Prior Distribution	Posterior Distribution
1	Intercept	Normal	0.1929	0.1928
1	Expenditure	Normal	-274E-15	-273E-15
2	Intercept	Normal	0.1929	-1.22E11
2	Expenditure	Normal	-274E-15	0
1	Variance	Inverse Gamma	0.07146	0.1667
2	Variance	Inverse Gamma	0.07146	0.9998
1	Probability	Dirichlet	0.6180	0.9563

As mentioned in the Table 3, there are seven parameters obtained for the analysis. Figure 1 demonstrates the Markov Chain diagnostics graph for all seven parameters. Findings show that there is a constant alternation for the parameter estimates in trace plot. This indicates that there is a good mixing chain. The autocorrelation plot indicate the degree of autocorrelation for each posterior samples. High correlations imply slow mixing. Hence, Figure 1 revealed that there is a low autocorrelation for each posterior samples. While the density plot shows the relative variability of the posterior distribution on the data plot. Figure 1 displays a smooth, unimodal shape of posterior marginal distributions for each parameter. The findings from all diagnostics graph illustrated that the Markov Chain for parameter estimates is successfully convergence.

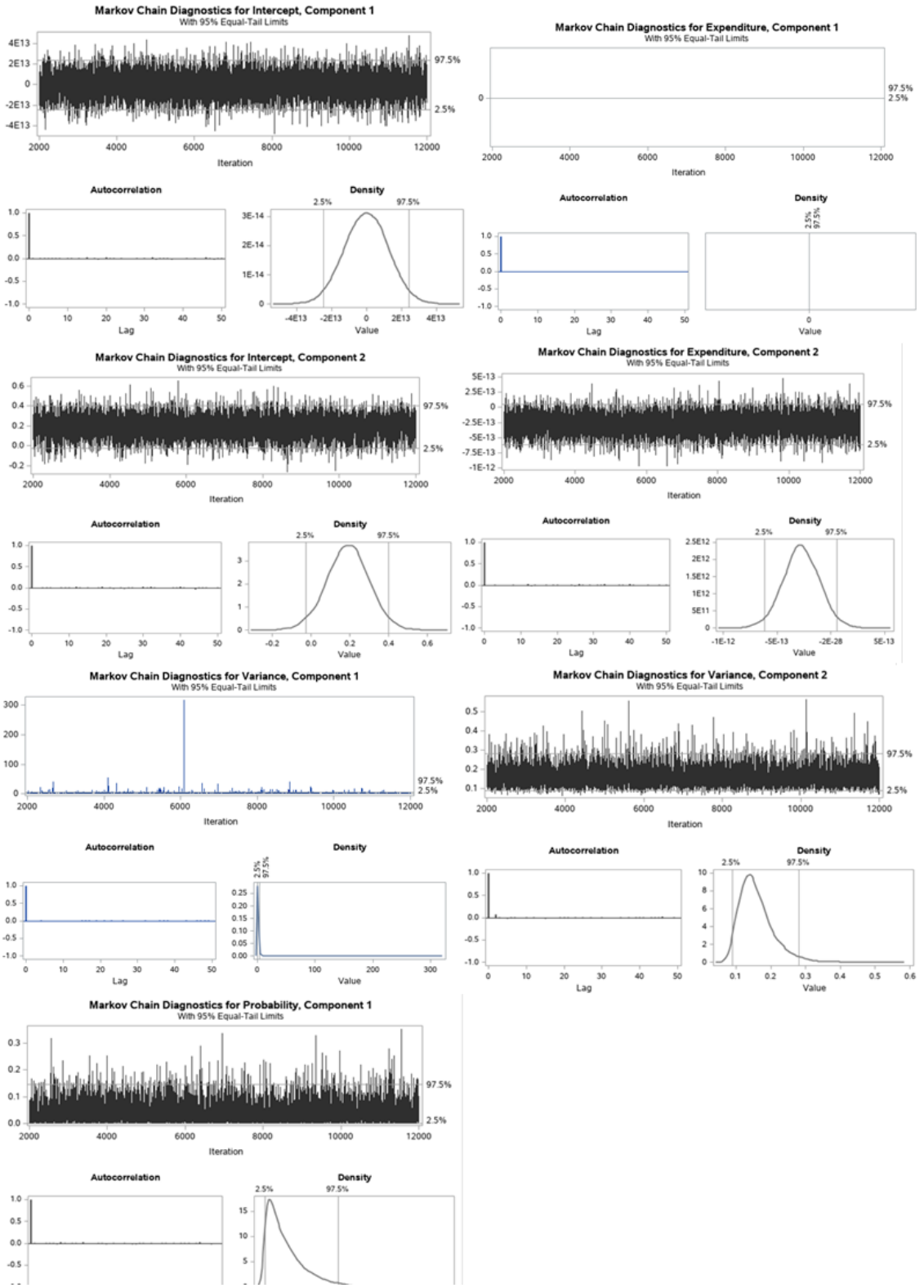

Figure 1: Markov chain diagnostics graph for all parameters.

5 Conclusions

This paper found that there is a two components mixture model between the exchange rate and international tourism expenditure in Malaysia. This model is then fitted by the Bayesian method to examine the model parameters and establish an equation to predict the future observation. Results show that Bayesian method is found suitable in examining the time series data since the posterior distribution will incorporate the existing information with the new information. This study is important to the stakeholder and government to establish relevant policy and budget to it. In addition, researchers can use Bayesian method fitted with the Bayesian method when predict the exchange rate and tourism expenditure.

Acknowledgement This research has been carried out under Fundamental Research Grants Scheme (FRGS/1/2019/STG06/UPSI/02/2) provided by Ministry of Education Malaysia. The authors would like to extend their gratitude to Universiti Pendidikan Sultan Idris (UPSI) that helped managed the grants.

Conflicts of Interest The authors declare no conflict of interest.

References

[1] Y. Arslanturk, M. Balcilar & Z. A. Ozdemir (2011). Time-varying linkages between tourism receipts and economic growth in a small open economy. *Economic Modelling*, *28*(1), 664–671. https://doi.org/10.1016/j.econmod.2010.06.003.

[2] J. Balaguer & M. Cantavella-Jorda (2002). Tourism as along-run economic growth factor the Spanish case. *Applied Economics*, *34*, 877–884. https://doi.org/10.1080/00036840110058923.

[3] M. Belloumi (2010). The relationship between tourism receipts, real effective exchange rate and economic growth in Tunisia. *International Journal of Tourism Research*, *12*(5), 550–560. https://doi.org/10.1002/jtr.774.

[4] Bernama (2019). *Tourism contributed RM84.1 billion in 2018*. Malaysia. Retrieved from: http://bernama.com/en/news.php?id=1699504.

[5] M. H. Chen & W. G. Kim (2006). The long-run equilibrium relationship between economic activity and hotel stock prices. *Journal of Hospitality Financial Management*, *14*(1), 1–15. https://doi.org/10.1080/10913211.2006.10653811.

[6] J. A. Cottani, D. F. Cavallo & M. S. Khan (1990). Real exchange rate behavior and economic performance in LDCs. *Economic Development and Cultural Change*, *39*(1), 61–76. https://www.jstor.org/stable/1154104.

[7] S. Dey (2009). Comparison of Bayes estimators of the parameter and reliability function for Rayleigh distribution under different loss functions. *Malaysian Journal of Mathematical Sciences*, *3*, 247–264.

[8] N. Dristakis (2004). Tourism as a long-run economic growth factor: An empirical investigation for Greece using causality analysis. *Tourism Economics*, *10*, 305–316. https://doi.org/10.5367/0000000041895094.

[9] S. T. Katircioglu (2009). Tourism, trade, and growth the case of Cyorus. *Applied Economics*, *41*, 2741–2750. https://doi.org/10.1080/00036840701335512.

[10] H. H. Lean, S. H. Chong & C. W. Hooy (2014). Tourism and economic growth: Comparing Malaysia and Singapore. *International Journal of Economics and Management*, *8*(1), 139–157.

[11] G. McLachlan, S. X. Lee & S. Rathnayake (2019). Finite mixture models. *Annual Review of Statistics and Its Application*, *6*(1), 355–378. https://doi.org/10.1146/annurev-statistics-031017-100325.

[12] I. A. Nugroho, S. Gunawan, A. A. Awirya & P. Nurman (2017). The effect of exchange rate fluctuations on Bali tourism sector. *Jurnal Ekonomi dan Pembangunan*, *25*(1), 15–26.

[13] R. Nunkoo (2015). Tourism development and trust in local government. *Tourism Management*, *46*, 623–634. https://doi.org/10.1016/j.tourman.2014.08.016.

[14] C. O. Oh (2005). The contribution of tourism development to economic growth in the Korean economy. *Tourism Management*, *26*(1), 39–44. https://doi.org/10.1016/j.tourman.2003.09.014.

[15] A. A. Pravitasari, S. F. Qonita, N. Iriawan, Irhamah, Fithriasari, S. W. Purnami & W. Ferriastuti (2020). MRI-Based brain tumor segmentation using Gaussian and hybrid Gaussian mixture model-spatially variant finite mixture model with expectation-maximization algorithm. *Malaysian Journal of Mathematical Sciences*, *14*(1), 77–93.

[16] C. H. Puah, M. C. Jong, N. Ayob & S. Ismail (2018). The impact of tourism on the local economy in Malaysia. *International Journal of Business and Management*, *13*(12), 151–157. https://doi.org/10.5539/ijbm.v13n12p151.

[17] C. H. Tan, X. W. Chong & S. B. Ho (2017). Significance of safety and security issues on tourism industry in Malaysia. *Science International* (*Lahore*), *29*(2), 471–475.