International Journal of Engineering & Technology, 7 (4.30) (2018) 454-457

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research paper

Malaysia Tourism Demand Forecasting Using Box-Jenkins Approach

Diyana Izyan Amir Hamzah[1], Maria Elena Nor[1], Sabariah Saharan[1]*, Noor Fariza Mohd Hamdan[1], Nurul Asmaa Izzati Nohamad[1]

[1]*Department of Mathematics and Statistics, Faculty of Applied Science and Technology, Universiti Tun Hussein Onn Malaysia, Pagoh Educational Hub, 84600 Pagoh, Johor, Malaysia*
**Corresponding author E-mail: sabaria@uthm.edu.mu*

Abstract

Tourism industry in Malaysia is crucial and has contributes a huge part in Malaysia's economic growth. The capability of forecasting field in tourism industry can assist people who work in tourism-related-business to make a correct judgment and plan future strategy by providing the accurate forecast values of the future tourism demand. Therefore, this research paper was focusing on tourism demand forecasting by applying Box-Jenkins approach on tourists arrival data in Malaysia from 1998 until 2017. This research paper also was aiming to produce the accurate forecast values. In order to achieve that, the error of forecast for each model from Box-Jenkins approach was measured and compared by using Akaike Information Criterion (AIC), Mean Absolute Deviation (MAD), Mean Square Error (MSE) and Mean Absolute Percentage Error (MAPE). Model that produced the lowest error was chosen to forecast Malaysia tourism demand data. Several candidate models have been proposed during analysis but the final model selected was SARIMA $(1,1,1)(1,1,4)_{12}$. It is hoped that this research will be useful in forecasting field and tourism industry.

Keywords: *Accuracy comparison, Box-Jenkins, Malaysia tourism, SARIMA, Tourism demand forecasting.*

1. Introduction

In the past few decades, tourism industry has become crucial especially to the growth of countries' economy which is always related to the development of new jobs, the rising of money exchange rate, the increasing of Gross Domestic Product (GDP), and many more. Naturally, tourism means the activities of people visiting or traveling to other places of their interest besides their living area for a few days [1, 2, 3]. According to The Oxford English Dictionary, the synonym of tourist can also be called as visitor, traveler, sightseer, backpacker, tripper, and others.

With the rising of tourism industry, this field has become popular among academia for study purpose and analysis in various topics and methods [4]. It also involved data of tourist arrivals to various countries around the globe and other related data that influenced the tourist arrivals. However, this article generally aims to do modelling and forecasting by using Box-Jenkins approach, and also to compare the accuracy of each model in order to find a better model to forecast tourist arrival in Malaysia.

Malaysia is known as one of the popular destinations for traveling because of its attractions. For example, historical places like A'Famosa, skyscrapers like Petronas Twin Tower, islands, beautiful natures and Malaysian different cultural [5, 6]. Instead of travelling for holidays, international tourists come to Malaysia for medical treatment, business trip, sports trip and others [7, 2, 8].

Since tourism industry is crucial to boost Malaysia's economy, accurate forecasting of the future tourism demand may help people who work in tourism-related-business to make a correct judgment, plan strategy and minimize the side effect if the prediction is decreasing for the next year. They might lose capital if they prepare services excessive than the number of target tourists. Otherwise, if the prediction for the next year is increasing, then the forecasting analysis may help in optimizing the services enough for every tourist such as hotels, airplane tickets, transporters, foods, and others.

This paper provides the latest forecast research on Malaysia tourism demand by using international tourist arrival to Malaysia as data set. Box-Jenkins approach has been chosen to analyse tourism demand because it is a complex and a flexible method that able to handle time series data with several patterns.

2. Literature Review

Literatures related to this research were reviewed especially to the past researchers that were focusing on tourism demand forecasting and Box-Jenkins method.

2.1. Tourism demand forecasting

Several literatures about tourism demand forecasting were reviewed to study different methods used by past researchers. Since tourism industry is important for any country's economic growth, it has attracted a lot of attention from researchers to do forecasting on the tourism demand. For comparison, [9] and [10] used Malaysia tourism demand data but with different forecasting methods and different data of tourist arrival origin countries. The research by [9] was focusing more on the tourists from Singapore and Indonesia for periods of 2000 until 2013, while research in [10] were focusing on Singapore tourist dataset started from January 2010 to December 2014. In addition, research [9] applied artificial neural network (ANN) model, hybrid empirical mode decomposition with artificial neural network (EDM-ANN) model, and modified

Copyright © 2018 Authors. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

EDM-ANN model. Otherwise, research [10] applied artificial neural network (ANN) and support vector machine (SVM) methods. By using RMSE and MAPE, the result from [9] shows modified EDM-ANN model produced the lowest value of error whereas research [10] find out ANN produced less error than SVM.
Other than that, a few research papers that used combined model to forecast tourism demand have been found, such as the researches from [11] and [12]. Taiwan tourism data in [11] research were forecasted by using neural network model, SARIMA model and the combined model which consist of autocorrelation function (ACF) model, neural network model and genetic algorithms model. Otherwise, research [12] used Japan tourism demand data to do forecasting by using neuron model with dendritic nonlinearity (DNN) and the combined model that consist of SARIMA and DNN (SA-D) models. To compare forecasting accuracy, research [11] applied RMSE and MAD. The result shows the combined model is better than neural network and SARIMA models. While, for research in [12] the research used NMSE, APE, R and program running time (PRT) to compare the performance of forecasting. The result shows that SA-D model produce less error than DNN model.
Furthermore, another two researches that focusing on artificial neural network (ANN) for forecasting are [13] and [14], which used Mozambique tourism demand data and China tourism demand data, respectively. Researcher in [13] computed candidate models by using ANN method with different input of independent variables. The forth model contains GDP-MOZ, ER-EURO-MT, ER-RAND-MT, ER-DOLLAR-MT, and GDP_USA variables produced the lowest MAPE and the highest pearson correlation coefficient (r). Otherwise, research [14] applied single ANN (S-ANN), combined k- Means and ANN method (C-ANN), combined EMD and ANN methods (E-NN), ARIMA model, and combination clustering ANN (C-C-ANN). As a result, the proposed C-C-ANN model produced the lowest error for all accuracy methods.

2.2. Box-Jenkins approach

Autoregressive Integrated Moving Average (ARIMA) model is a combination of several time series models including autoregressive and moving average. This model is considered as one of the popular time series methods because it can be used on time series data that has pattern like trend, seasonality and cyclic. In order to obtained ARIMA model, usually researcher will apply Box-Jenkins procedure which consist of three steps such as identification step, estimation of parameter *p*, *d* and *q* step, and lastly is diagnostic checking step [15]. This forecasting method has been used widely not only in tourism field. Thus, this subsection will provide reviews on Box-Jenkins approach in various fields includes tourism field, fisheries field and gold price field.
A few literature reviews of Box-Jenkins modelling that used tourism demand data for forecasting were studied, its include researches from [16, 17,]. Based on [16], the researchers used 25 years data of China tourist arrival in Australia. Four accuracy measurements were conducted to select the best model from Box-Jenkins approach. Otherwise, the second research by [17] used Macedonia tourism demand from 1956 until 2013 for forecasting. The researcher applied RMSE, MAE, APE, MAPE and TIC to choose forecasting model. For result comparison, model selected in [16] was SARIMA $(0,1,1)(0,1,1)_{12}$ while model selected in [17] was ARIMA (1,1,1).
Another two researches that applied Box-Jenkins approach in fisheries field can be found in [18] and [19]. Researcher in [18] analysed ten years data of monthly Barramundi price by using ARIMA and Holt-Winter methods. However, researcher in [19] has applied ARIMA and intervention analysis to forecast total marine fish landings in Odisha. MSE and MAPE were used in [18] to compare forecast accuracy and the result found that ARIMA model performed better than Holt-Winter method. Otherwise, research [19] compared AIC and SBC values for each method and the result shows that ARIMA outperformed intervention analysis model.
Besides, Box-Jenkins approach has also been used to forecast gold price by investors. Forecasting is useful and important for investors so that they can be well prepared about the fluctuation of the gold prices, thus they can make a correct judgement and plan strategy before investing. Referring to [20], the researchers used ten years data of gold price in India and the accuracy of forecast between candidate models have been made by using RMSE, MAPE, MAE, BIC and Lungs Box Q statistics. Another research is by [21], 25 years data of gold prices in India were used and the researcher compared forecast performance by using RMSE, MAPE, Max APE, Max AP, and MAE. Result in [20] shows the best model selected is ARIMA (1, 1, 1) while result in [21] found ARIMA (0, 1, 1) is better than other candidate models.

3. Materials and methods

This subsection will cover the study area and data used in this research, Box-Jenkins approach, and method used to measure forecast performance.

3.1. Study area and data

In this research, the data used to forecast is a monthly tourism data of tourist arrival from January 1998 until December 2017. The data were taken from the website of Tourism Malaysia which consists of approximately 240 records. The dataset was divided into two groups, the first group dataset (in-sample) is from 1998 until 2016 which used to model and forecast 12 months ahead. The second group data (out-of-sample) is 12 months data of 2017 used to evaluate forecast accuracy. In this article, Box-Jenkins approach has been chosen as a method to forecast Malaysia tourism demand. All the data were generated and computed by using Microsoft Excel, R software and Minitab.

3.2. Box-Jenkins procedure

This approach uses a three steps iterative approach of model identification, parameter estimation and diagnostic checking to determine the best model from a general class of ARIMA models. This three steps process is repeated several times until a satisfactory model is finally selected. Then, this model can be used for forecasting future values of time series. The three steps in Box-Jenkins procedure are:
Step 1: Identification step
In the first step which is identification step, Box-Cox Transformation plot is used to stabilize the variance by transformation method while Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) plots are used to observe the stationarity of mean, and differencing method needed to stabilize the mean [15].
Step 2: Estimation of parameter
The second step is estimation of parameter, ACF and PACF plots are vital in this step to estimate the value of parameter *p* and *q*. Parameter *p* represent the number of lag cut off in ACF plot while parameter *q* represent the number of lag cut off in PACF plot. The best model usually was picked with the p-value of parameter less than 0.05 and the *p*-value of L-Jung Box value higher than 0.05 [23]. Equations (1), (2), and (3) were referred and proposed by [24].
The autoregressive process of order *p* is denoted AR (*p*) is generally defined as follows:

$$X_t = \emptyset_1 y_{t-1} + \emptyset_2 y_{t-2} + \cdots + \emptyset_p y_{t-p} + \varepsilon_t \quad (1)$$

The moving average process of order *q* is denoted MA (*q*) and defined by follows:

$$X_t = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \ldots - \theta_q \varepsilon_{t-q} \quad (2)$$

The general ARIMA model which allocates seasonality can be written as follows:

$$\emptyset_p(B)\, \Phi_p(B^S)\, \nabla^d\, \nabla_s^d\, y_t = \theta_q(B)\, \Theta_Q(B^S)\alpha_t \quad (3)$$

Where X_t and ε_t are the actual value and random error at time period t, respectively; $\emptyset_i$ (i=1, 2,…,p) and θ_j (j=0,1,2,…,q) are model parameters, p and q are integers and often referred to as orders of the model.

Step 3: Diagnostic check

The last step in Box-Jenkins procedure is diagnostic checking which requires an observation towards residual values of forecast. This will reveal if there is any autocorrelation in the residuals. Again, ACF and PACF plots can be used to check their stationarity.

3.3. Measuring of accuracy

In order to evaluate the accuracy of forecast data from Box-Jenkins method, Akaike Information Criterion (AIC), Mean Absolute Deviation (MAD), Mean Square Error (MSE) and the Mean Absolute Percentage Error (MAPE), were computed. For all four measures, smaller values usually indicate a better fitting model. These statistics are computed as follows.

$$\text{AIC} = -2\, log(L) + 2k \quad (4)$$

$$\text{MSE} = 1/n \sum |y_t - \hat{y}_t|^2 \quad (5)$$

$$\text{MAPE} = 1/n \sum |(y_t - \hat{y}_t) / y_t| \times 100 \quad (6)$$

$$\text{MAD} = 1/n \sum |y_t - \hat{y}_t| \quad (7)$$

Where, k is the number of model parameters and L is the log-likelihood to measure a model fit, y_t is the actual value at end of period t, $\hat{y}_t$ is the forecast value at end of period t, n is the number of observation.

4. Result and Discussion

This study applied Box-Jenkins approach to forecast Malaysia tourism demand for 2017. The data (in-sample) used in this study is a monthly tourist arrival in Malaysia from January 1998 until December 2016. Meanwhile, data (out-of-sample) from January 2017 till December 2017 were used to compare the accuracy of forecast. Time series plot for this dataset can be obtained from Figure 1. From Figure 1, the pattern of time series clearly shows there are trend and seasonality exist.

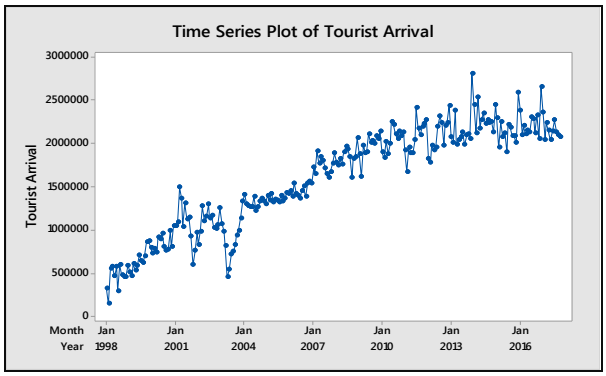


Fig. 1: Time series plot of tourist arrival from 1998 until 2017

The first step in Box-Jenkins procedure is identification step which crucial to check the stability of variance in dataset. Box-Cox transformation shows that the time series data is normal with rounded value equals to 1. Thus, there is no need to do transformation since the variance is stationary. However, ACF and PACF plots illustrated that the mean is not stationary and differencing is needed to stabilize it. Differencing for non-seasonal and seasonal terms were performed and several candidate models have been listed in Table 1.

Table 1: Candidate models

Candidate models	AIC	MSE	MAD	MAPE
SARIMA $(1,0,1)(1,1,1)_{12}$	5732.48	17123354959	91199.81	6.99
SARIMA $(1,0,1)(0,1,2)_{12}$	5732.04	17117682252	91272.97	7.00
SARIMA $(1,1,1)(1,0,1)_{12}$	6012.58	16259133007	88657.80	7.52
SARIMA $(1,1,1)(0,1,1)_{12}$	**5697.69**	16645635269	89587.61	6.84
SARIMA $(1,1,1)(1,1,4)_{12}$	5699.14	**15558714175**	**88516.81**	**6.51**

Based on Table 1, the candidate models were compared by using AIC, MSE, MAD and MAPE. According to [23], forecast value can be considered as very good if MAPE value is less than 10%. Since MAPE value in Table 1 for all candidate models showed less than 10%, it means their forecast values can also be considered as accurate. However, in order to obtain the best model to forecast tourism data, the model with the lowest value of error will be selected. Therefore, model SARIMA $(1,1,1)(1,1,4)_{12}$ has been chosen since it's value of MSE, MAD and MAPE were the lowest.

The second step in Box-Jenkins procedure is the estimation of parameters. The coefficient of parameter for the model selected (SARIMA $(1,1,1)(1,1,4)_{12}$) from the first step will be used for modelling. P-value of estimated parameters shows all values were less than 0.05, which means the parameters are significant to the model.

Based on equation (3), the full model for SARIMA $(1,1,1)(1,1,4)_{12}$ can be written as follows (8):

$$\Phi_1(B)\,(1-B)^1 \Phi^1(B^{12})(1-B^{12})^1\, Y_t = \theta_1(B)\Theta_1(B^{12})^4\, e_t \quad (8)$$

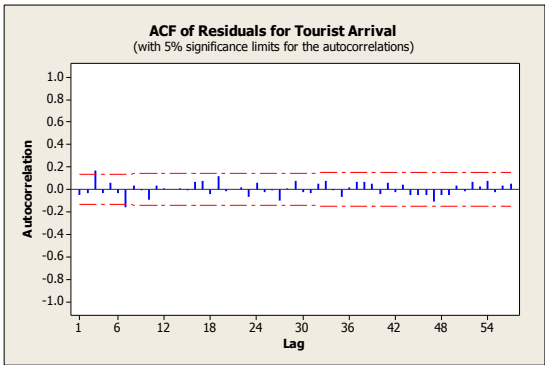


Fig. 2: ACF plot

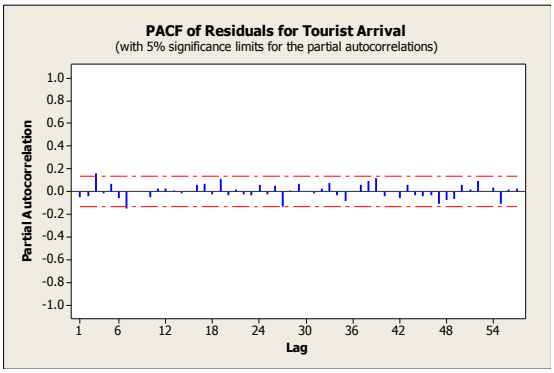


Fig. 3: PACF plot

The last step in Box-Jenkins procedure is diagnostic checking. Diagnostic plots in Figure 2 and Figure 3 show that the residuals in the model are stationary and indicate no correlations exist. Thus, SARIMA $(1,1,1)(1,1,4)_{12}$ model is fit to forecast tourism data.
Figure 4 shows the time series plot of actual data from 1998 until 2017 and the forecast values for 12 months in 2017. The red marks illustrate the upper limit and the lower limit of forecast values for 2017. As display, all forecast values are in the range of lower and upper limits and the values are not too far from the actual data.

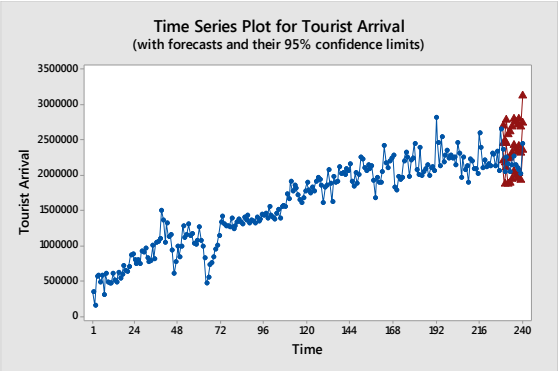


Fig. 4: Time series plot of actual and forecast data

Even though this study has provided a reliable model for forecasting, this does not mean the model is the only way to analyse future demand. This paper only provides framework technique and forecast figure in numbers, so the people who involve in tourism-related-business should not rely hundred percent on this research outcome. Planner must consider economic factors and tourism influence other than common sense and good judgement to plan future strategies.

5. Conclusion

This article was focusing on Box-Jenkins approach to forecast Malaysia tourism demand data. Three steps in Box-Jenkins procedure were carried out and the candidate models selected were SARIMA $(1,0,1)(1,1,1)_{12}$, SARIMA $(1,0,1)(0,1,2)_{12}$, SARIMA $(1,1,1)(1,0,1)_{12}$, SARIMA $(1,1,1)(0,1,1)_{12}$, and SARIMA $(1,1,1)(1,1,4)_{12}$. The best model among candidate models was chosen by comparing AIC, MSE, MAD and MAPE. The result of this study showed SARIMA $(1,1,1)(0,1,1)_{12}$ produced the lowest value if AIC (5697.69) and SARIMA $(1,1,1)(1,1,4)_{12}$ produced the lowest value of MSE (15558714175), MAD (88516.81**)** and MAPE (6.51%). Thus, SARIMA $(1,1,1)(1,1,4)_{12}$ model was selected.
This paper suggest for future research to apply advance forecasting method to obtain more precise prediction, and also to consider several other economic factors as data set.

Acknowledgement

We would like to express our thanks to Universiti Tun Hussein Onn Malaysia for providing the facilities and financial supports to do this research. This research is supported by TIER 1 Project Grant Vot. U900.

References

[1] Ismail N, Masron T & Ahmad A (2014), Cultural Heritage Tourism in Malaysia: Issues and Challenges. *SHS Web of Conferences* 12, pp. 1–8.
[2] Kumar M & Sharma S (2016), Forecasting Tourist In-Flow in South East Asia: A Case of Singapore. *Tourism & Management Studies* 12(1), pp. 107–119.
[3] Turner R & Freiermuth E (2017), Travel & Tourism Economic Impact 2017 Malaysia. *World Travel & Tourism Council*, pp. 1–19.
[4] Shinae C, Hui B, Beomcheol P & Chon K (2018), Review of Reviews : A Systematic Analysis of Review Papers in the Hospitality and Tourism Literature. *International Journal of Hospitality Management* 70, pp. 49–58.
[5] Mosbah A & Mohamed Saleh AA (2014), A Review of Tourism Development In Malaysia. *European Journal of Business and Management* 6(5), pp. 1–9.
[6] Lim WM (2015), Restoring Tourist Confidence and Travel Intentions after Disasters: Some Insights from a Rejoinder to a Series Of Unfortunate Events in Malaysian Tourism. *Current Issues in Tourism* 20(1), pp. 38–42.
[7] Aziz A, Md Yusof R, Abu Bakar NT, Taib SNH & Ayob M (2015), Travel Behavioral Intention of Choosing Malaysia as Destination for Medical Tourism. *In Theory and Practice in Hospitality and Tourism Research - Proceedings of the 2nd International Hospitality and Tourism Conference*, (2014), pp. 185-188.
[8] Seow AN, Choong YO, Moorthy K & Chan LM (2017), Intention to Visit Malaysia for Medical Tourism Using the Antecedents of Theory of Planned Behaviour: A Predictive Model. *International Journal of Tourism Research* 19(3), pp. 383–393.
[9] Yahya NA, Samsudin R & Shabri A (2017), Tourism Forecasting Using Hybrid Modified Empirical Mode Decomposition and Neural Network. *International Journal of Advances in Soft Computing and Its Applications* 9(1), pp. 14–31.
[10] Ali R & Shabri A (2017), (Modelling Singapore Tourist Arrivals to Malaysia by Using SVM and ANN. *SCIREA Journal of Mathematics* 1(2), pp. 210–216.
[11] Liang Y (2016), Using the Combined Model for Forecasting the Tourism Demand. *2016 International Conference on Machine Learning and Cybernetics*, pp. 612–615.
[12] Yu Y, Wang Y, Gao S & Tang Z (2017), Statistical Modeling and Prediction for Tourism Economy Using Dendritic Neural Network. *Computational Intelligence and Neuroscience*, pp. 1–7.
[13] Constantino HA, Fernandes PO & Teixeira J P (2016), Tourism Demand Modelling and Forecasting with Artificial Neural Network Models: The Mozambique Case Study. *Tékhne* 14(2), pp. 113–124.
[14] Jun W, Yuyan L, Lingyu T & Peng G (2018), Modeling a Combined Forecast Algorithm Based On Sequence Patterns And Near Characteristics: An Application For Tourism Demand Forecasting. *Chaos, Solitons and Fractals* 108, pp. 136–147.
[15] Hyndman RJ & Athanasopoulos G (2014), Forecasting Principles and Practice. Print edition. Otexts.com (Online, open-access textbook).
[16] Ma E, Liu Y, Li J & Chen S (2016), Anticipating Chinese Tourists' Arrivals in Australia : A Time Series Analysis. *Tourism Management Perspectives* 17, pp. 50–58.
[17] Petrevska B (2017), Predicting Tourism Demand by A.R.I.M.A. Models. *Economic Research-Ekonomska Istrazivanja* 30(1), pp. 939–950.
[18] Rusiman MS, Hau OC, Abdullah AW, Sufahani SF & Azmi NA (2017), An Analysis of Time Series for the Prediction of Barramundi (Ikan Siakap) Price in Malaysia. *Far East Journal of Mathematical Sciences (FJMS)* 102(9), pp. 2081–2093.
[19] Raman RK, Sathianandan TV, Sharma AP & Mohanty BP (2017), Modelling and Forecasting Marine Fish Production in Odisha Using Seasonal ARIMA Model. *National Academy Science Letters* 40(6), pp. 393–397.
[20] Guha B & Bandyopadhyay G (2016), Gold Price Forecasting Using ARIMA Model. *Journal of Advanced Management Science* 4(2), pp.117–121.
[21] Gharde A (2016), Influence of Factors on Clothing Sales and Its Future Trend: Regression Analysis and Time Series Forecast of Clothing Sales. *Journal of Textile and Apparel, Technology and Management* 10(2), pp. 1-11.
[22] Wiradinata SA, Yendra R, Danil M & Gamal H (2017), Multi-Input Intervention Analysis for Evaluating of the Domestic Airline Passengers in an International Airport. *Science Journal of Applied Mathematics and Statistics* 5(3), pp. 110–126.
[23] Da Veiga CP, Da Veiga CRP, Catapan A, Tortato U & Da Silva WV (2014), Demand forecasting in food retail: A comparison between the Holt-Winters and ARIMA models. *WSEAS Transactions on Business and Economics* 11(1*)*, pp. 608–614.
[24] Box GEP & Jenkins GM (1970), Time Series Analysis: Forecasting and Control. Holden Day, San Francisco.