International Journal of Engineering & Technology, 7 (4.30) (2018) 432-437

International Journal of Engineering & Technology

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

Research paper

Granger Causality Analysis on PLUS Highway Traffic Network

Norhaidah Mohd Asrah[1]*, Maman Abdurachman Djauhari[2], Azme Khamis[3], Rohayu Mohd Salleh[4], Sya Sya Syahira Muhammad Fitri Avtar[5]

[1,3,4,5]Faculty of Applied Science & Technology, Universiti Tun Hussein Onn Malaysia, Pagoh Campus, 84600 Pagoh, Johor, Malaysia.
[2]Centre for Research in Statistics and Data Analysis, Tjahaja Bina Statistika Indonesia, Ltd.P., Jl. Kanayakan A-15, Bandung 40135, West Java, Indonesia.
**Corresponding author E-mail: norhaida@uthm.edu.my*

Abstract

PLUS Malaysia Berhad (PMB) is the largest toll expressway operator in Malaysia and South East Asia. Based on this reputation, PLUS handled thousands of vehicles every day. It covers the in-coming and out-coming traffic burdens from the northern areas to the southern areas. In order to manage these traffic burdens, toll plazas are located along this highway. Previous studies revealed that some of these toll plazas are important in managing the traffic burdens. This study analyze the importance of Skudai (SKD) toll plaza in Johor from 2009 until 2013. The causal relationship between SKD toll plaza with other toll plazas in Johor is studied to determine if there is any potential correlation or relationship of SKD with other toll plazas by using the Granger causality analysis. The findings show that there is a bidirectional Granger causality between SKD and Tangkak (TGK) as well as Machap (MAC) toll plazas. Meanwhile, there is only a unidirectional Granger causality between SKD and Yong Peng Utara (YPU) and Yong Peng Selatan (YPS) toll plazas. These results benefit the PLUS highway management to identify which potential toll plazas that can increase the traffic burden out from SKD toll plaza.

Keywords: *Centrality Measures; Forest; Granger causality; Highway Network; Minimal Spanning Tree; Traffic Burden*

1. Introduction

Transportation is the most important facilities provided by any country. Transportation is not only to cover the vehicles facilities such as cars, buses, trains, planes, ferries and many more. It also handles the facilities for the road network. The road network plays an important role in the economic and social development. If a country can provide a better road network, it can help that country to develop more and progress rapidly. The provision of a good road network and infrastructure are crucial in order to ensure the trade activities continue.

In a developing country like Malaysia, the road network is very important since transportation gives a big impact on the economic and social development. The main purpose of providing better road network is to facilitate the transportation of goods and passengers which cover 96% of the burdens carried on the road network. The road mileage in Malaysia increased about 33.76% from 2009 to 2012 [1]. As a result, the number of registered motor vehicles also increased about 12.45% from 2010 to 2012. There are three categories of the road in Malaysia namely toll expressways, federal and states. Toll expressways have a length of about 1700 km and it is under the administration of the Malaysian Highway Authority [2].

The Projek Lebuhraya Utara Selatan (PLUS) highway development had been our interest. It is because this highway has been one of the busiest highways in Peninsular Malaysia as many vehicles utilize this highway every day. This highway involved the areas from northern to southern Peninsular Malaysia.

Based on the previous studies on PLUS highway, there are four most important toll plazas, among 63 toll plazas along the PLUS highway. The four most important toll plazas are Sungai Besi (SBI), Jalan Duta (JLD), Juru (JRU) and Skudai (SKD) toll plazas as mentioned by [3-6]. These four toll plazas really represent the northern, central and southern areas. JRU toll plaza represents the northern areas, meanwhile SBI and JLD toll plazas represent the central areas and lastly, SKD toll plaza represents the southern areas.

The connections or the relations between each toll plazas are explained by getting the network of these toll plazas. The toll plazas are treated as a node and their network are analyzed by using the minimum spanning tree (MST) and the forest of all possible minimum spanning trees (Forest). The MST is applied to filter the information or data enclosed in the highway traffic network. MST follows the idea in graph theory to filter the data enclosed in a weighted connected graph. This graph has n objects (nodes). The tree from this MST has $(n - 1)$ edges or links. It minimizes the sum of the distance of toll plazas. MST is connecting every part in a set of n together in a graph considered by a minimal distance between the nodes. The general method to construct an MST is known as Kruskal's and Prim's algorithm [7]. In terms of topological properties, MST is constructed to envision the important data contained in the network. Other studies involved with MST can be referred to [3-6] and [8-11].

The Forest was proposed by [9] since Forest is unique compared to MST that might not be unique in the network. It ensures the robustness of the filtered information because it provides a unique source of filtered information. The current practice to find MST is by using Kruskal's or Prim's algorithm on a dissimilarity matrix D of size (n x n). Then, then the results are used to find the subdomi-

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.

nant ultrametric (SDU). In Forest, the method proposed using the backward direction. The SDU will find in the first stage, then it is used to determine MST. The proposed steps can be referred to [9-11]. Then, this network can be visualized by using Pajek software. The network topology can be viewed and the connection is clearly displayed among the nodes.

The network topology cannot explain in details how the nodes are connected unless there is a measurement to describe and explain each of the connection and the importance of each node. This measurement refers to the centrality measures namely degree centrality [12], betweenness centrality [13], closeness centrality [14] and also eigenvector centrality [15]. Based on these centrality measures, the most important centrality measures are determined by introducing the effective variance (EV) and effective vector variance (EVV) [6]. The betweenness centrality measure is the most important centrality measure and again the previous four most important toll plazas are the same toll plazas with the higher scores in betweenness centrality.

It is crucial to know whether the most important toll plaza has any possibility to predict the traffic burden of another toll plaza. Based on this reason, the analysis is continue with the causal relationship between two toll plazas. The causal relationship between the most important toll plaza and other toll plazas located in Johor is analyzed by using the Granger causality analysis.

2. Methodology

The analysis of Granger causality is based on linear regression modeling of stochastic processes [16]. The Granger causality analysis is used to analyze the causal relationship between the time series data on SKD toll plaza with other toll plazas in Johor. There are several steps to be considered before it can proceed to Granger causality analysis. The steps are checking the stationarity of each time series data and to find the suitable number of lags. This procedure is discussed in details in the next section.

2.1. The PLUS toll plazas data

The Granger causality analysis is applied to the number of out-coming traffic burdens from toll plazas in Johor. The toll plazas involved were Tangkak (TGK), Pagoh (PGH), Yong Peng Utara (YPU), Yong Peng Selatan (YPS), Ayer Hitam (AHT), Machap (MAC), Simpang Renggam (SPR), Sedenak (SDK), Kulai (KLI), Senai Utara (SNU) and Skudai (SKD). The five years of data starting from the year 2009 to 2013 and cover the number of traffic burdens exit from those toll plazas. The data are provided by Toll Department of PLUS Malaysia Sdn. Bhd. The out-coming traffic burden refers to the number of vehicles that exit from each toll plaza.

2.2. Time series plot

These data were collected monthly starting from January 2009 until December 2013. The time series plot for each toll plazas can be constructed so that the pattern for each toll plazas can be identified. It can give more information and ideas on the traffic burden flows. The highest number of traffic burden in each toll plazas can be identified from this time series plot. Minitab software is used to plot the PLUS toll plazas time series data.

2.3. The stationarity

The basic procedure in Granger causality analysis is the stationarity of the toll plazas time series data. The time series data for the toll plazas must be covariance stationary. For this purpose, the unit root test of Dickey-Fuller can be applied to find the stationary [17]. We can also use the Augmented Dickey-Fuller (ADF) and Phillip-Perron unit root test as can be seen in [18]. The ADF test is used in this study to check the stationarity of the time series data for toll plazas. The formula for ADF is shown in (1) below.

$$y_t^{'} = \phi y_{t-1} + \sum_{i=1}^{k} \beta_i y_{t-i}^{'} \quad (1)$$

where

k is the number of lags to include in the regression,

$y_t^{'}$ is the first differenced series,

β is the parameter of differenced series and

ϕ is the coefficient of ADF.

$y_t^{'}$ denotes the first differenced series and k denotes the number of lags to include in the regression. If the coefficient of the ADF test, $\phi < 0$, then the time series data is stationary. Otherwise, the time series data is not stationary [19]. Based on the null hypothesis H_0 of non-stationary, it can be rejected at 5% level of confidence. The statistical packages *XL*STATS is used to help in solving the stationarity.

2.4. The number of lags

The number of suitable lags is important because the Granger-causality test is too responsive to the number of lags included in the regression [20]. The Akaike Information Criteria (AIC) and Schwarz Information Criterion (SIC) have been used to find the suitable number of lags. A study by [17] also used AIC and Bayesian Information Criterion (BIC) to find the number of lags for the environmental case study. Thus, the AIC is chosen to find the suitable number of lags for all toll plazas. The statistical packages *R* is used to find the AIC value.

2.5. The Granger causality analysis

As mention by [20], the Granger-causality analysis can be applied to assess whether one variable has any potential to predict or forecast the other variable. Back in 1969, [21] proposed the definition of Granger causality and since then a lot of economic time series datasets have been tested for causality. A variable X_{2t} is said to be causal for a time series variable X_{1t} if the former helps to improve the forecasts of the latter [22].

Thus, a time series variable X_{1t} is said to cause another variable, X_{2t} if the future values of X_{2t} can be predicted from X_{1t}. This can be done when the information contained in the series X_{1t} is included than when that information is excluded. In the test of causality, researchers must develop forecast equations for X_{2t}, with and without X_{1t} [20].

Another definition of Granger causality can be referred to [23]. Granger causality defined as a causal ordering of associated variables. The important terms are the predictability and the exploitations of the direction of time flow. This analysis also has been applied to various fields such as stock prices [20], medicine [24], environmental [16], tourism [25] and economic [18].

The Granger causality analysis can be computed after the requirements of stationary and the suitable number of lags are determined. Suppose that two stochastic processes of X_t and Y_t , said that X_t (or Y_t) Granger-(G)-causes Y_t (or X_t). When the past of X_t (or Y_t) contains information that can support to predict the future of Y_t (or X_t) over. At the same time, the above information already in the past of Y_t (or X_t) [17].

The procedures to examine the SKD toll plaza causes TGK toll plaza can be explained below:

i. SKD toll plaza is regressed on its past values excluding TGK toll plaza in the regressors. This model omits the historical values of the TGK toll plaza. This process or procedure is known as the restricted regression. The sum of squared residuals can be obtained from this procedure.

ii. The second regression is computed including the lagged of TGK toll plaza. This model has full specifications compared to the first model. This is called the unrestricted regression from which the unrestricted sum of squared residuals is obtained.

The *F*-test need to be carried out and the formula is given by

$$F = \frac{\left[\frac{SSE_r - SSE_u}{n}\right]}{\left[\frac{SSE_u}{T-(m+n+1)}\right]} \quad (2)$$

where
SSE_r is the sum of squared errors for the restricted model,
SSE_u is the sums of squared errors for the unrestricted model,
T is the elements that form the degrees of freedom,
n is the number of observations and
m are the number of lags.

The *F*-test in (2) and the significance values (*p*-value) are then calculated. In hypothesis testing, if the results are statistically significant, the null hypothesis is rejected when the *p*-value is smaller than the specified level of significance, alpha (α = 0.05). The assumption that there is no causality between the two-time series (the null hypothesis H_o) by using the statistical test of Fisher distribution or as known the *F*-test on the alternative H_1 (there is causality).
Meanwhile, there are four hypotheses can be concluded between SKD toll plaza and one of the other toll plazas (for example TGK toll plaza). These hypotheses can be concluded as:

i. H_o: There is no unidirectional Granger-causality from SKD toll plaza to TGK toll plaza (one-way).
   H_a: There is a unidirectional Granger-causality from SKD toll plaza to TGK toll plaza (one-way).
ii. H_o: There is no unidirectional Granger-causality from TGK toll plaza to SKD toll plaza (one-way).
   H_a: There is a unidirectional Granger-causality from TGK toll plaza to SKD toll plaza (one-way).
iii. H_o: There is no bidirectional Granger-causality from SKD toll plaza to TGK toll plaza (two-way).
   H_a: There is a bidirectional Granger-causality from SKD toll plaza to TGK toll plaza (two-way).
iv. H_o: There is no Granger-causality from SKD toll plaza to TGK toll plaza (independent).
   H_a: There is a Granger-causality from SKD toll plaza to TGK toll plaza (independent).

In managing the Granger causality analysis, one needs to know that there are three different situations when the Granger-causality test can be applied [20]. The three situations are:

i. simple Granger-causality test, it includes two variables and their lag.
ii. multivariate Granger-causality test, it includes more than two variables.
iii. extended multivariate Granger-causality test, to test the simultaneity of all included variables using vector autoregressive (VAR) framework.

The situation that suitable for this study refers to the first situation. The Granger causality test is applied to two variables and their lags. Two variables represent the SKD toll plaza and another one from the other toll plazas.

3. Results and Discussions

The results of the Granger analysis are discussed in this section. The results start from the time series plot for each toll plaza from 2009 until 2013, the stationary for each toll plazas, the number of suitable lags for each toll plazas and lastly the Granger causality for each pair of toll plazas.

3.1. Time series plot for all toll plazas

The time series plot for all toll plazas is showed below. Figure 1 until Figure 10 displays the time series plot for each toll plaza from 2009 until 2013. Based on these figures, the highest traffic burden is traffic burden that exit from SKD toll plaza. The pattern of this traffic burden is up and down. The traffic burden is higher during the festive seasons such as Chinese New Year, Hari Raya Puasa, Hari Raya Haji, Deepavali, and Christmas. Not only that, the traffic burden also increased during the school holidays. Since SKD toll plaza is one of the most important toll plaza [3-6], it is not possible that this toll plaza has the highest volume on traffic burden compared to other toll plazas. It also connected to all toll plazas in Johor except for TGK toll plaza. Based on this evidence, the causality relationship between SKD toll plaza to other toll plazas in Johor is crucial. It will help the PLUS highway management to identify which potential toll plazas that contribute to increasing the traffic burden in SKD toll plaza.

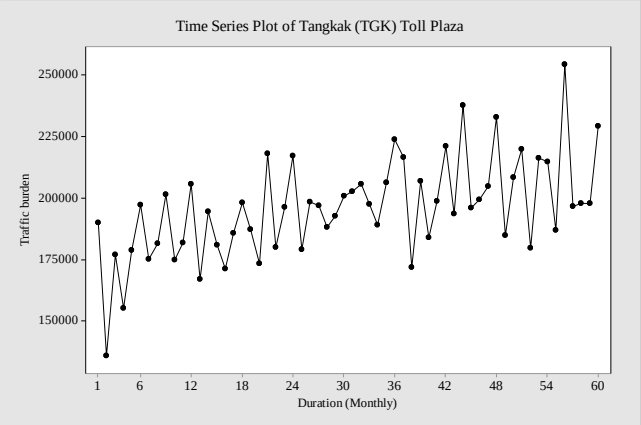


Fig. 1: The traffic burden exit from TGK toll plaza from 2009 to 2013.

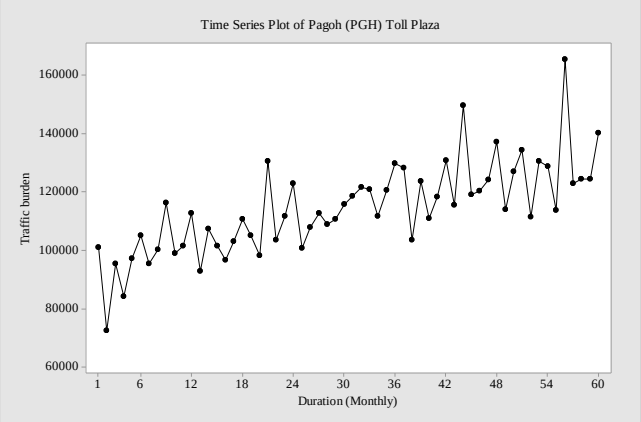


Fig. 2: The traffic burden exit from PGH toll plaza from 2009 to 2013.

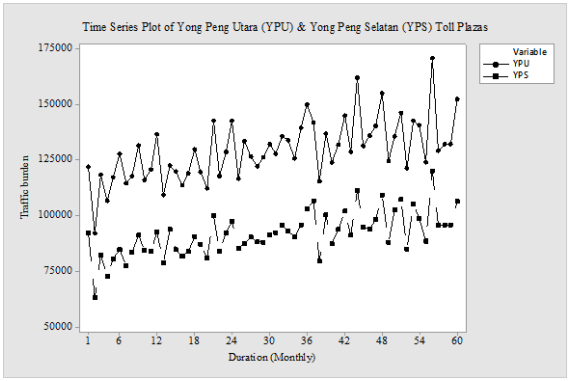


Fig. 3: The traffic burden exit from YPU and YPS toll plazas from 2009 to 2013.

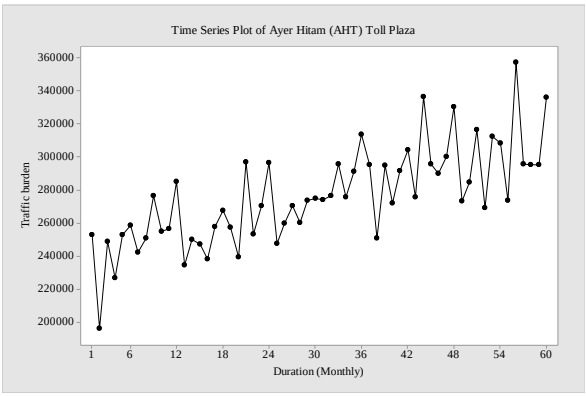


Fig. 4: The traffic burden exit from AHT toll plazas from 2009 to 2013.

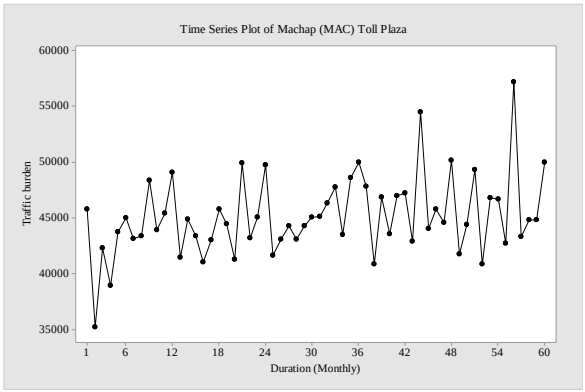


Fig. 5: The traffic burden exit from MAC toll plazas from 2009 to 2013.

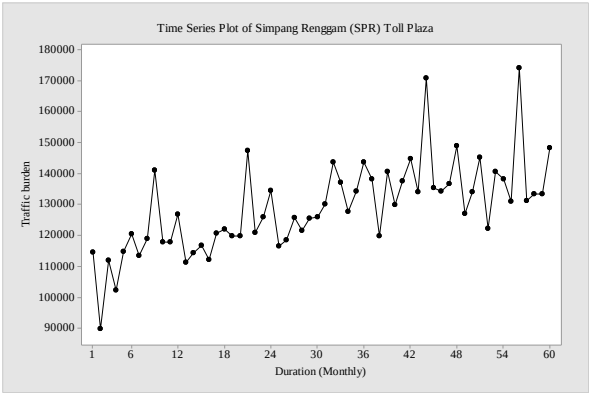


Fig. 6: The traffic burden exit from SPR toll plazas from 2009 to 2013

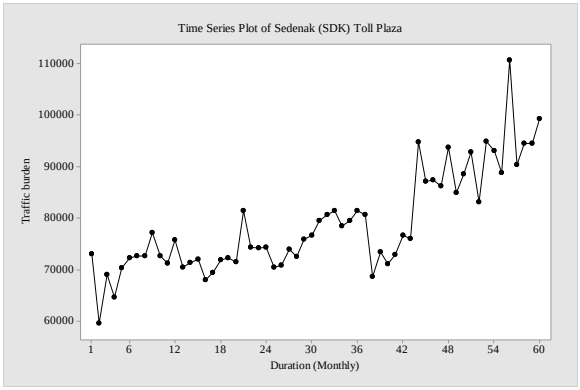


Fig. 7: The traffic burden exit from SDK toll plazas from 2009 to 2013

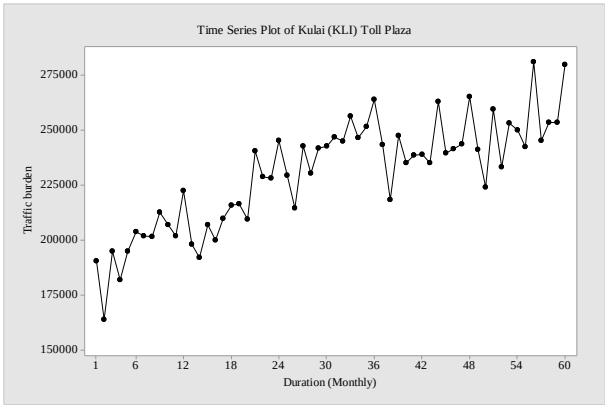


Fig. 8: The traffic burden exit from KLI toll plazas from 2009 to 2013

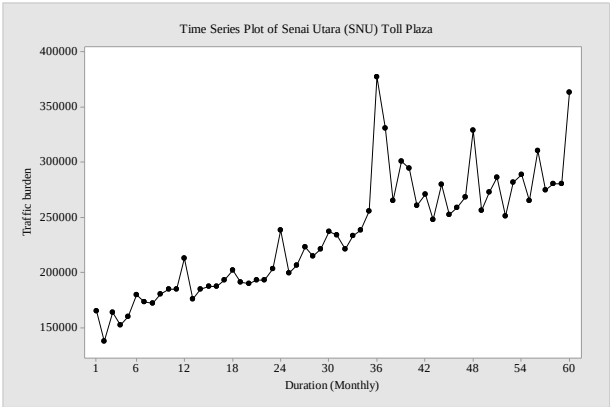


Fig. 9: The traffic burden exit from SNU toll plazas from 2009 to 2013

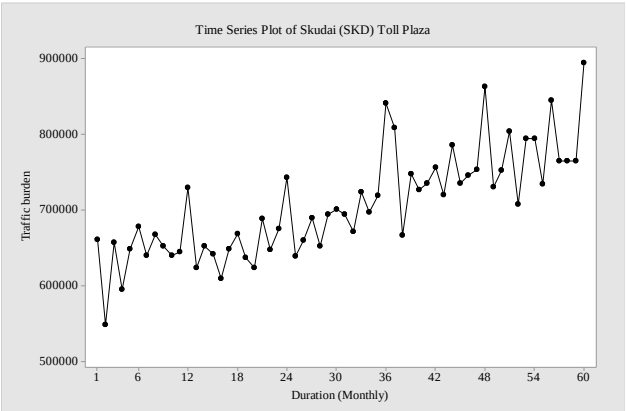


Fig. 10: The traffic burden exit from SKD toll plazas from 2009 to 2013

3.2. The stationarity

It is important to confirm that the time series data for all toll plazas are stationary. It is one of the requirements in applying the

Granger causality [16]. The stationarity for each toll plaza is calculated by using the Augmented Dickey-Fuller (ADF) test. This test is applied to all toll plazas time series data by using the *XL*STAT software. The results for each toll plaza are shown in Table 1 below.

Table 1: Augmented Dickey-Fuller (ADF) test

No.	Toll Plaza	*p*-value
1	TGK	0.0003
2	PGH	0.0017
3	YPU	0.0058
4	YPS	0.0036
5	AHT	0.0124
6	MAC	0.0043
7	SPR	0.0212
8	SDK	0.3602*
9	KLI	0.2885*
10	SNU	0.1911*
11	SKD	0.0378

The ADF test results show that all toll plazas are stationary except for SDK, KLI and SNU toll plazas. The *p*-value for these three toll plazas is more than alpha (α) value that is 5% significance level. Hence, these three toll plazas are not included in the next stage procedures.

3.3. The number of lags

It is important to find the suitable or the appropriate number of lags before using Granger causality analysis. It is because the Granger causality test is concerned about the lags to include in regression [16]. The number of lags can be found by using the Akaike Information Criteria (AIC). The results are generated from *R* software. The results for each toll plazas are shown in Table 2.

Table 2: Number of lags using Akaike Information Criteria (AIC)

No.	Toll Plaza	No. of lags
1	TGK	2
2	PGH	2
3	YPU	2
4	YPS	2
5	AHT	3
6	MAC	1
7	SPR	1
8	SKD	1

According to AIC, the suitable number of lags for TGK, PGH, YPU and YPS toll plazas is two. Meanwhile, AHT toll plaza has three suitable lags and the rest have only one suitable lag.

3.4. The Granger causality analysis

The Granger causality analysis is then conducted by using *R* and EViews software. Both software gives almost similar results. The output results for each toll plazas are shown in Table 3. Based on these results, there are several toll plazas that have significant results.

Granger causality runs two-way from SKD to TGK and TGK to SKD. The same goes from SKD to MAC and from MAC to SKD. Meanwhile, Granger causality runs one-way from SKD to YPU and not the other way. The same situation goes from SKD to YPS. This explains that there is unidirectional causality from SKD to YPU and YPS. The other pairs show that there is no Granger causality in any direction between them or they are independent.

Table 3: The Granger causality analysis results

No.	Toll Plaza	*F*-statistics	*p*-value
1	SKD – TGK	3.8336	0.0278*
	TGK – SKD	3.2988	0.0446*
2	SKD – PGH	1.2480	0.2953
	PGH – SKD	0.0081	0.9919
3	SKD – YPU	3.4809	0.0379*
	YPU – SKD	1.8724	0.1637
4	SKD – YPS	3.8593	0.0272*
	YPS – SKD	1.0866	0.3447
5	SKD – AHT	0.5060	0.6799
	AHT – SKD	1.2514	0.3011
6	SKD – MAC	4.3836	0.0408*
	MAC – SKD	14.1995	0.0004*
7	SKD – SPR	0.9188	0.3419
	SPR – SKD	0.0673	0.7962

3.5. Discussion

Based on the results, it revealed some information that needs to be considered by PLUS highway management. Among seven toll plazas (not include SKD toll plaza), three toll plazas are not having any Granger causality to SKD toll plaza. The three toll plazas are PGH, AHT, and SDK. The traffic burden exit from these toll plazas are not related or contribute to the increase of traffic burden out-coming from SKD toll plaza and vice versa.

Meanwhile, there is one-way or unidirectional Granger causality from SKD toll plaza to YPU and YPS toll plazas. This Granger causality explains that whenever the traffic burden out-coming from SKD toll plaza increase, it increases the prediction of the traffic burden in YPU and YPS toll plazas. However, the increasing traffic burden that out-coming from YPU and YPS toll plazas does not affect the prediction of traffic burden in SKD toll plaza.

Lastly, there is a two-way or bidirectional Granger causality between SKD toll plaza and TGK toll plaza. It means that when the out-coming traffic burden increase from SKD toll plaza, it will affect to the increasing of the out-coming traffic burden from TGK toll plaza and vice versa.

These findings are really interesting because the connection or relation that may contribute to the prediction in out-coming traffic burden for SKD toll plaza and other toll plazas can be identified by applying the Granger causality analysis. Perhaps in the future research, this study can continue with data involving all toll plazas for the whole areas starting from Penang to Johor. At the same time, the in-coming traffic burden can also be considered.

4. Conclusion

The Granger causality analysis shows that SKD toll plaza have one-way and two-way Granger causality with YPU, YPS and TGK toll plazas respectively. Based on this results, PLUS highway management can identified that YPU, YPS and TGK toll plazas can influence the increasing number of traffic burden in southern areas. Other toll plazas such as PGH, AHT, MAC and SPR are not given any influence in increasing the number of traffic burden. Hence, the PLUS highway management can take into consideration on these toll plazas for their future projects in reducing the traffic congestion along the highway.

Acknowledgments

The authors really appreciate the editor and anonymous reviewers for their comments and suggestions. It really helps the authors to improve the final presentation of this paper. Special appreciation goes to the Universiti Tun Hussein Onn Malaysia (UTHM) for providing the facilities and grant under UTHM TIER 1 2017 Vot U902, Tjahaja Bina Statistika Indonesia, Ltd.P. and Toll Department, PLUS Malaysia Sdn. Bhd for the data provided in this research.

References

[1] Department of Statistics Malaysia, (2013).
[2] Zakaria S & Sufian Z (2009), Public Works Department Malaysia Construction and Maintenance Ensuring Road Quality in Malaysia. Public Works Department.

[3] Asrah NM & Djauhari MA (2014), A Recommendation on PLUS Highway Development: A Social Network Analysis Approach, 2014 IEEE International Conference on Industrial Engineering and Engineering Management, 1187-1191,
[4] https://ieeexplore.ieee.org/document/7058826/
[5] Asrah NM, Djauhari MA & Mohamad I (2015), PLUS Traffic Highway: An Analysis Based on Time Series Similarity Approach. The Social Sciences 11, 11, 2753-2759.
[6] Asrah NM, Djauhari MA & Mohamad I (2017), PLUS highway network analysis: Case of in-coming traffic burden in 2013, AIP Conference Proceedings 1842(1), 030008,
[7] https://doi.org/10.1063/1.4982846
[8] Asrah NM, Djauhari MA & Mohamad I (2018), The Most Important Centrality Measure in PLUS Highway Traffic Network: An Approach Based on Effective Variance, AIP Conference Proceedings 1974(1), 040020, https://doi.org/10.1063/1.5041694
[9] Mantegna RN & Stanley HE (2000), An Introduction to Econophysics: Correlations and Complexity in Finance, Cambridge University Press.
[10] Asrah NM, Djauhari MA & Suleiman ES (2014), Work Attitude among Malaysian Academicians in the Public Universities: A Social Network Analysis. Modern Applied Science 8(5), 9-18.
[11] Djauhari MA (2012), A robust filter in stock networks analysis. Physica A: Statistical Mechanics and Its Applications 391, 5049-5057.
[12] Djauhari MA, Sharif & Djauhari H (2012), Network Analysis on Safety Culture and Worker's Behaviour: A Forest of All Minimum Spanning Trees. International Journal of Basic & Applied Sciences 12, 29-37.
[13] Djauhari MA & Lee GS (2013), Network topological property of English dialects similarity: A robust filter approach. American Journal of Applied Sciences 10, 646-653.
[14] Opsahl T, Agneessens F & Skvoretz J (2010), Node Centrality in Weighted Networks: Generalizing degree and shortest paths. Social Networks 32(3), 245-251.
[15] Park K & Yilmaz A, A Social Network Analysis Approach to Analyze Road Networks. ASPRS Annual Conference. San Diego, CA. (2010), 1-6, http://www.asprs.org/wp-content/uploads/2013/08/Park_K.pdf
[16] Newman ME (2008), The mathematics of networks. The New Palgrave Encyclopedia of Economics 2, 1-12.
[17] Borgatti SP (1995), Centrality and AIDS. Connection 8, 112-115.
[18] Charakopoulos AK, Katsouli GA & Karakasidis (2017), Dynamics and causalities of atmospheric and oceanic data identified by complex networks and Granger causality analysis. Physica A: Statistical Mechanics and its Applications 495, 436-453.
[19] Eleftherios T (2007), A Time Series Model for the Romanian Stock Market. European Research Studies XI, 3-4, 57-72.
[20] Hooi HL & Russell S (2010), On the dynamics of aggregate output, electricity consumption and exports in Malaysia: Evidence from multivariate Granger causality tests. Applied Energy 87, 1963–1971.
[21] Furuoka F & Munir Q (2014), Unemployment and Inflation in Malaysia: Evidence from Error Correction Model. Malaysian Journal of Business and Economics 1(1), 35-45.
[22] Foresti P (2006), Testing for Granger causality between stock prices and economic growth.
[23] Granger CWJ (1969), Investigating causal relations by econometric models and cross-spectral methods. Econometrica (37), 424-438.
[24] Helmut L & Markus K (2004), Applied Time Series Econometrics. Cambridge University Press, New York.
[25] Eichler M (2001), Granger causality graphs for multivariate time series.
[26] Patrick AS & Patrick LP (2017), A study of problems encountered in Granger causality analysis from a neuroscience perspective. Proceedings of the National Academy of Sciences 114(34), E7063-E7072.
[27] Oludele AA & Lydia AB (2010), International Tourism and Economic Development in South Africa: A Granger Causality Test. International Journal of Tourism Research 12, 149–163.