Applied Mathematics & Information Sciences
An International Journal

http://dx.doi.org/10.18576/amis/110409

Malaysia Road Accidents Influences Based on Structural Time Series Analysis

Noor Wahida Md Junus[1,2,*], *Mohd Tahir Ismail*[2], *Zainudin Arsad*[2] *and Rosmanjawati Abdul Rahman*[2]

[1] Department of Mathematics, Sultan Idris Education University, 35900 Tanjung Malim, Perak, Malaysia
[2] School of Mathematical Sciences, Universiti Sains Malaysia, 11800 Minden, Penang, Malaysia

Received: 9 Sep. 2016, Revised: 28 Apr. 2017, Accepted: 2 May 2017
Published online: 1 Jul. 2017

Abstract: Traffic accidents are very common and it is such a trauma to all the road users. Decreasing the rate of road accidents is an important issue in developing countries including Malaysia. The aim of this study is to investigate the contributing factors of road accidents as well as to determine the effectiveness of structural time series in predicting Malaysian road accidents. This study applied structural time series approach as this method offers the possibilities of modeling time series component in terms of stochastic processes. The estimation process in this study was implemented by using Kalman filter approach. The factors that are incorporated include economic factors, climate factors, calendar effects and intervention policies. The study shows that effects of road accidents in different regions vary due to the unique characteristics in climate effect, the economic effect and influences of calendar effects. Hence, the study suggests that policy makers should enhance different approaches based on the region in ways to improve and to decrease the number of road accidents in Malaysia.

Keywords: Stochastic process, Structural time series, Kalman filter, Road accidents

1 Introduction

One of the aim of a developed country is to enhance the survival rate of its population by improving the community healthcare and quality of life. However, lately the increasing population and vehicle registered indirectly have become a major life threatening. Across all countries, one of the leading causes of death is attributed to road accidents. Malaysia holds the highest risk of road fatalities in the world [1]. Various efforts were executed to reduce death figure because of road accidents. In the early 1970s, the first motorcycle lane was built along Federal Highway with the aim to reduce motorcycle accidents. However, this intervention only covers the Central region. A study by [2] found that this intervention has successfully reduced motorcycle accidents by 34%.

In 1989, the Road Commission Safety Cabinet was formed with the responsibility of formulating a national road safety target. In the following year, Microcomputer Accident Analysis Package (MAAP) was introduced. The package enables Malaysia to access black spot analysis and conduct necessary treatment to the affected area. In 1996, the Malaysian government established a five-year National Road Safety Target. The government targeted to reduce the number of accident deaths by 30% by the year 2000. Various initiatives were carried out to achieve the target. In 1997, the road safety research center, which is under Universiti Putra Malaysia (UPM) was mandated to conduct research on motorcycle safety as one of the initiatives. In the next Malaysian road safety plan (2006-2010), the government targeted to reduce 52.4 % road deaths by 2010. Among the initiatives to achieve the target is enforcement of the Ops Sikap since 2001. This operation has been conducted during the festive season. It is followed by introducing the rear seat belt legislation in 2009. However, in 2010, the rate of road death stood at 3.4 per 10000 vehicles, which is higher than the expected value that is 2.0 per 10000 vehicles.

Various reasons contribute to traffic accidents. Factors that contribute to the occurrence of road accidents can be categorized into driver factor, vehicle factor and roadway factor [3]. Over the past few years, a number of studies regarding road accidents modeling has been developed. The aim of the studies is to investigate the contributing factors in road accidents as well as to identify the most accurate method to predict road accidents. There are

* Corresponding author e-mail: noor_wahida@fsmt.upsi.edu.my

numerous statistical and mathematical methods that were introduced to achieve this objective. Among the studies were by [4, 5] who applied Smeeds and Andressen equation in fitting the road accidents model. Both studies use annual series of road accidents and vehicle registered vehicle as predictor variable. Besides, the most preferable method to model road accidents is linear regression analysis, which was applied by [6–9]. At the same time, [7] compared the regression method performance with artificial neural network and [9] identified the performance of the model with Box Jenkins time series.

In addition, another potential method that has always been used in modeling road accidents is general linear model (GLM). The possible method of GLM that is likely used is Poisson regression and Negative Binomial regression. The researcher who used this method was among the pioneers of road accidents analysis in Malaysia [10]. Their study is regarding the effectiveness of the running head light intervention in motorcycle accidents. The similar method has also been used by [11] to identify the effectiveness of the enforcement of rear seatbelt in Malaysia while [12] present the prediction of motorcycle accidents at junctions on urban roads in Malaysia. Furthermore [13] and [14] apply Poisson regression analysis while [15] applies negative binomial model in developing a predictive model for road accidents.

However, the above mentioned methods may not suitable for series data. Therefore, Box-Jenkins method was used by [16–19] in modelling the road accidents series. In 2005, [18] applied the Box-Jenkins analysis to determine the effect of the recent economic crisis and the motorcycle safety program. The result found that motorcycle-related accidents are positively related to the gross domestic product. Consequently, in 2011, [19] examined the effectiveness of Ops Sikap, which is one of the intervention policies to nurture road user attitude during festive seasons. Recently, [17] forecast road traffic injuries using a few statistical techniques and found that autoregressive integrated moving average (ARIMA) performed better than damped trend. Incorporating both cross-sectional and series data made panel data analysis more preferable in developing a road accidents model. This model has been applied to determine the effectiveness of road infrastructure in reducing fatalities and injuries by [20]. On the other hand, [21] incorporated meteorological variable and traffic exposure as a potential contributor in predicting the effect of environmental factors to road safety. The study found that Poisson integer value autoregressive gives higher likelihood ratio compared to other competing models and weather conditions were significantly related to the crash count. Other researchers that use similar method is Yaacob et al. [22–24] that applied it on Malaysian road accident series.

Up to our knowledge of the road accidents study in Malaysia, the methods that is usually applied is generalized linear model (GLM), Box-Jenkins and panel data analysis. The GLM method is only suitable cross sectional data while panel data analysis is suitable if our data is combination of cross sectional and time series [25]. Box-Jenkins model is very suitable for series data but the stationary assumption were rarely fulfilled. The series could made stationary by process of deseonalizing and detrending. In this case, important information might be lost by going through this process. The method that may account for this problem would be structural time series (STS) method. This method is one of the time series class methods that model the time series component such as seasonal and trend instead of removing the component as Box-Jenkins method. The STS modeling can be applied to a variety of problems in time series such as medical, biology, engineering, marketing and many other areas. Some application of STS can be found in [26–36].

The application structural times series in road accidents is pioneered by [37] that was applied in modeling road accidents in Great Britain. The study investigated the effectiveness of seat belt usage in Britain. They applied basic structural model which includes level, slope and seasonal component that is allowed to vary overtime. After a few decades, STS model on road safety is well developed inline with the availability of many statistical software. For examples in the study of [38], which investigate the progress of road safety in 10 European countries, [39] investigate monthly frequency and severity of road traffic accidents in Belgium and [40] which examine the changes in the trend and seasonal patterns in fatal crash in New Zealand in relation to changes in economic condition. While the most recent studies on road safety based on STS approach can be found in [41–44].

Therefore, this paper aims to investigate the contributing factors of road accidents as well as to determine the effectiveness of structural time series in predicting Malaysian road accidents. The remainder of this paper is organized as follows. In section 2, the description of explanatory variable that is included in this study and the way it is handled is explained. In addition, this section also explains briefly the method that will be used in road accidents modeling. The estimation result is explained in section 3 followed by discussion and conclusion of the study in the last sections.

2 Methodology

The Structural Time Series

A structural time series is one of the time series analysis classes that have direct interpretation. It is developed based on the classical time series decomposition method that can be defined as sums of μ, trend, γ, seasonal and ε, irregular component.

$$y_t = \mu_t + \gamma_t + \varepsilon_t, \qquad t = 1, \ldots, n \tag{1}$$

where y_t denotes the observation of interest and ε_t is normally and identically distributed with mean zero and variance $\sigma_\varepsilon{}^2$. However, the classical decomposition disadvantage is that the component is treated deterministically. It makes most of the series unsuitable based on this classical decomposition.

Realizing the situation, that stochastic series fit the series well, each of the component time series was modeled as a random walk model and the stochastic trend component can be specified as in Equation (2), and named as local linear trend model.

$$\begin{aligned} \mu_t &= \mu_{t-1} + v_{t-1} + \eta_t, & \eta_t &\sim \mathrm{NID}(0, \sigma_\eta{}^2) \\ v_t &= v_{t-1} + \varsigma_t, & \varsigma_t &\sim \mathrm{NID}(0, \sigma_\varsigma{}^2) \end{aligned} \tag{2}$$

where v_t is the slope or gradient of the trend μ_t. The irregular component ε_t, level disturbance η_t, slope disturbance ς_t are mutually uncorrelated. The slope component v_t, can be excluded from the model and yield the local level model as in Equation (3).

$$\mu_t = \mu_{t-1} + \eta_t, \qquad \eta_t \sim \mathrm{NID}(0, \sigma_\eta{}^2) \tag{3}$$

Letting the variance of level disturbance equal zero reduce the model as deterministic level model where level component, μ is fixed through time. Meanwhile, fixed the irregular component equal zero produces random walk model. Letting the level disturbance or slope disturbance in Equation (2) equal zero reduce the model into smooth trend model and local level with drift respectively where level component and slope component are fixed through time. Meanwhile, if both variance disturbances (level and slope) were fixed to zero, it is reduced as deterministic linear trend model and it is equivalent to linear regression model. The specifications of structural time series trend model that discussed here can be summarised as in Table 1.

Table 1: Structural Time Series Trend Specification

Level (Without Slope)	σ_ε	σ_η	σ_ς
Deterministics Level	*	0	
Local Level	*	*	
Random Walk	0	*	
With Slope	σ_ε	σ_η	σ_ς
Deterministic linear Trend Model	*	0	0
Local Level with Drift	*	*	0
Local Linear Trend	*	*	*
Smooth Trend	*	0	*

* Indicate any positive value

Time series data is sometimes influenced by seasonal variation. Technique of handling seasonality in structural time series model is either in trigonometry form or dummy seasonal form. This study prefers dummy seasonal form in modeling seasonal variation. The state equation for seasonal variation that allows the pattern to vary over time can be modeled as:

$$\gamma_t = -\sum_{j=1}^{s-1} \gamma_{t-j} + \omega_t, \qquad \omega_t \sim \mathrm{NID}(0, \sigma_\omega{}^2) \tag{4}$$

where s is the number of season and ω_t denotes seasonal disturbance term that is independent of all disturbance terms. Meanwhile, when the seasonal effect γ_t is not allowed to change over time the state equation for seasonal variation can be defined as

$$\gamma_t = -\sum_{j=1}^{s-1} \gamma_{t-j} \tag{5}$$

In contrast with level and slope component that only requires one state equation, seasonal equation will require $s-1$ state equation. In this study that involves monthly data, 11 state equations are needed.

Some of the series may be influenced by some other external factors besides the time series component. Incorporating the influences of external factor in structural time series in ways to investigate their effect on particular series is by adding explanatory and intervention variables into the measurement equation as follows:

$$y_t = \mu_t + \sum_{j=1}^{k} \beta_{jt} x_{jt} + \lambda_t w_t + \varepsilon_t \qquad \varepsilon_t \sim \mathrm{NID}(0, \sigma_\varepsilon^2) \tag{6}$$

where x_{jt} is explanatory variable, w_t is an intervention dummy variable β_{jt} and λ_t is the unknown parameter that will be estimated.

Parameter Estimation

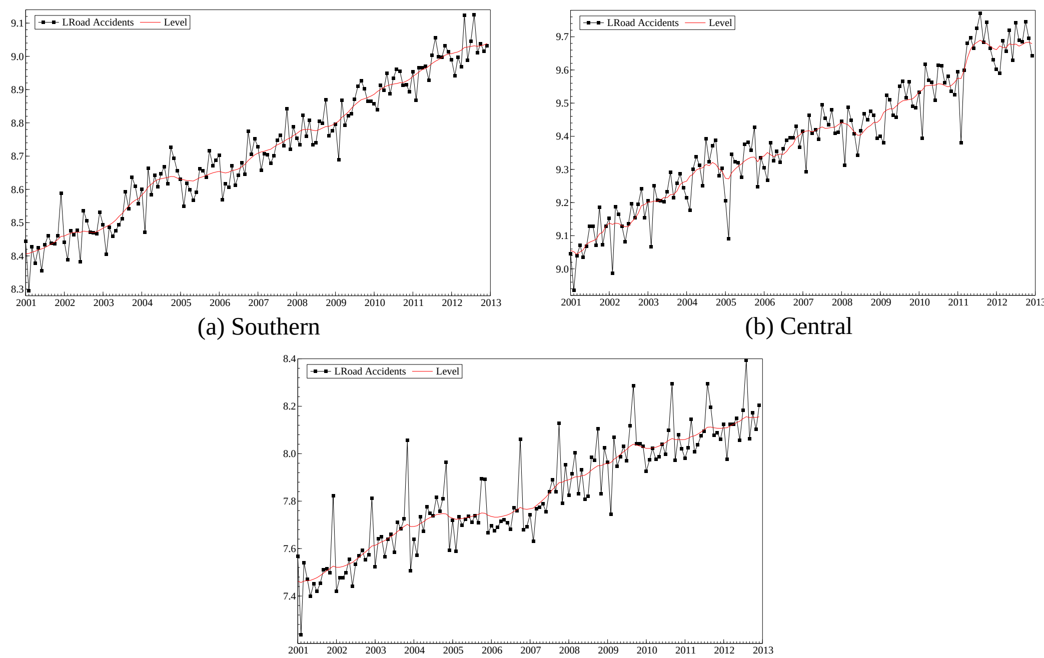
The statistical treatment of structural time series model is based on state space model form. This form allows the model discussed in previous section to be estimated by using Kalman filter technique. This technique as in [45] involves four stages of estimation procedure. In the first stage, one step ahead prediction of observation and state vector as well as corresponding mean square error were computed. The second stage involves diagnostic checking by means of one step ahead prediction errors. Diagnostic measure that was involved is similar to other linear gaussion models, which are independence of the error, homecedasticity of the error and normality of the error that are diagnosed by using Ljung Box (LB), Goldfelt Quant (GQ) and Jarque Bera (JB) test respectively. The third stage is computation of the likelihood function via one step ahead prediction error decomposition and the final stage is smoothing the output that is yielded from Kalman filter. In this study, the parameter estimation is made by using Kalman filter technique with the aid of STAMP software.

Data Considered

The data that is considered in this study involved three regions in Malaysian road accidents series which include

Table 2: Best Fitted model for Each Region

Regions	Variance Disturbance				Diagnostic test			Satisfied STS Model
	Level	Slope	Seasonal	Irregular	LB(6)	GQ	JB	
Southern	8.7×10^{-5}	0.0000	3.0×10^{-6}	0.0010	0.3904	0.8595	0.9624	Local Level Drift with Seasonal
Central	0.0003	-	0.0000	0.0008	0.1704	0.7437	0.7243	Local Level with Fixed Seasonal
East Coast	0.0005	-	-	0.0098	0.0761	0.8629	$<0.0001^{**}$	Local Level



(a) Southern (b) Central (c) East Coast

Fig. 1: Stochastic Level Component (a) Southern (b) Central (c) East Coast

Southern, East Coast and Central region. The number of road accidents in this series considered all transportations that were used the on the road which were involved in collision either with another transportation, individually or with other objects including pedestrian, animal and structure as recorded by the police. On the other hand, to investigate the influences of the occurence of road accidents the climate factor, economic factor, intervention policies as well as moving holiday variable were incorporated. The climate factors that were considered in this study include amount of rainfall (RAIN_F), number of rainy day (RAIN_D), maximum recorded temperature (TEMP), and air pollution index (API). Some of these factors were similar as in [5], [8] [21–24]. Meanwhile, the two economic factors that are believed to affect road accidents are the crude oil price (OILP) and consumer price index for transportation (CPI_T). The OILP is simple average of three spot prices which are Dated Brent, West Texas Intermediate and Dubai Fateh. The intervention variables that were included are the operation of Ops Sikap (Operation Attitude) and rear seat belt legislation (BELT). The Ops Sikap (OPSKP) is the integration of road safety operation to nurture road user behaviour during the main festival season. The first operation began on December 2001 in conjuction with celebration of Eid-ul-Fitr and Chrismast day. The duration of OPSKP usually last for 15 days. Later, rear seat belt legislation has been implemented in 2009 as an awareness safety campaign for passenger seat. In addition, the number of road accidents usually increase during festive seasons. Three main festivals were considered which is Eid-ul-Fitr, Chinese New Year and Deepavali. Hence, the occurence of these festivals does not have a fixed date according to the Gregorian calendar but moves over the years. In that case, moving holiday effect for these three main festivals was introduced and called return to village (RTV). The variables were based on one weight variable as suggested by Shuja et al. [46]. The number of road accidents and amount of rainfall were based on log transformation to reduce their variability in the series. All series cover the period from January 2001 until December 2013.

3 Estimation Results

Model Estimation Procedure
The study will involve three stages of model development. In the first stage, the best structural time series model is determined for each region followed by investigation of road accidents influences in the second stage. Finally, in the third stage, the effectiveness of structural time series is measured based on one year ahead forecast.

Estimation of Best Fitted Structural Time Series Model
The structural time series models are developed by using stepwise procedure. The analysis begins with fitted all the possible models as in Table 1. In this stage, the model was estimated without the independent variable. The best model was based on the lowest value of Akaike Information Criterion (AIC) [47, 48] with all the residual assumption satisfied. The analysis on fitting the best structural time series model for each region can be simplified as in Table 2. Road accidents model for the Southern region can be represented by seasonal local level drift. This model indicates that level and seasonal components were allowed to vary over time while slope component was fixed through time. The local level with fixed seasonal fits road accidents in the Central region. This model indicates that road accidents model in this region allowed level component to vary while the seasonal effect for this region was fixed. On the other hand, the best fitted structural time series model for the number of road accidents in East Coast region found that it is not influenced by seasonal effect. The best model to fit road accidents in East Coast region is local level model. Figure 1 shows the stochastic level component for each of the region. All the observations were above and below the estimated level. Figure 2 illustrates the individual stochastic seasonal component for the Southern region. In 2001, road accidents were higher in December followed by October and August in the following year. The fixed seasonal component for the year 2001 for the Central region is illustrated in Figure 3 The figure shows that the lowest number of road accidents fall on February for the Central region that is influenced by seasonal effect and is a little bit higher during August and October.

The irregular component or observation disturbance for each region in Figure 4 show that there is no systematic pattern which indicates that it is closer to independent random values and it is confirmed by the Ljung Box test that shows insignificant value. However, this statement does not include the East Coast region. The East Coast region observation disturbance shows some deviations of the observed time series from its mean as

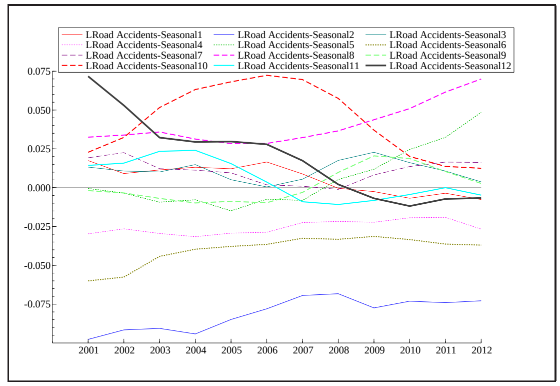


Fig. 2: Individual Seasonal Component for Southern Region

implied by the value of Ljung Box test that was significant at 10% significance level. In addition, normality residual assumption for East Coast region was significant at 1% significant level, which indicates that the residual diagnostic for both regions was not fully satisfied. However, the normality assumption is the least important assumption compared to the independent and homocedasticity assumption [49]. It may be due to the outliers or level break that were not incorporated in the model. In order of that, this study does not draw any conclusion regarding this model until the explanatory variable as discussed in previous section was included.

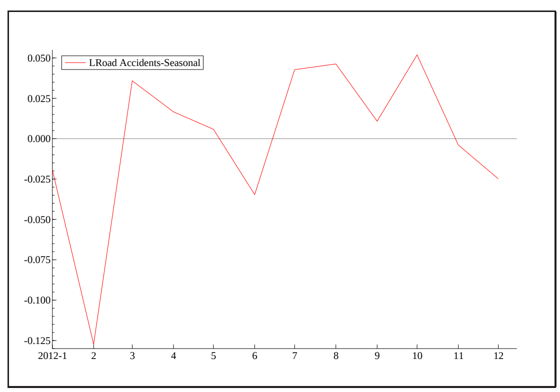
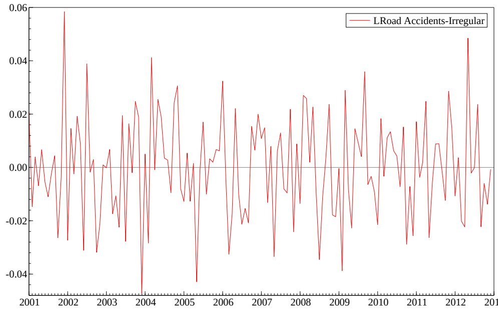
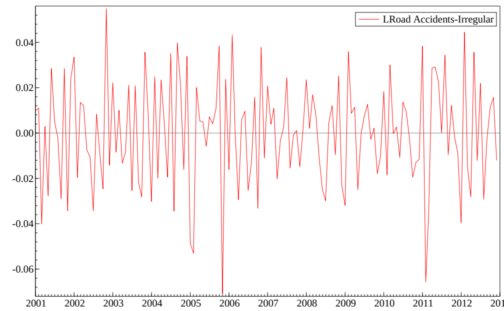


Fig. 3: Deterministic Seasonal Component for Central Region for Year 2012

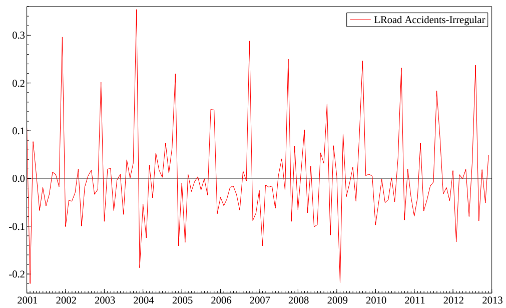
Estimation of Road Accident Influences
As the aim of this study is to investigate the influential factors that contribute to road accident, this lead to consideration of adding the best univariate structural time series that have been fitted previously with few explanatory and intervention variables. As a mean to



(a) Southern



(b) Central



(c) East Coast

Fig. 4: Irregular Component (a) Southern (b) Central (c) East Coast

Table 3: Estimate of variance Disturbance and Residual Diagnostic

Variance Disturbance	Southern	Central	East Coast
Level	0.0002	0.0008	0.0007
Slope	0.0000	-	-
Seasonal	5.1×10^{-7}	0.0000	-
Irregular	0.0006	0.0002	0.0039
LB(6)	0.2392	0.3273	0.1648
GQ	0.7648	0.5624	0.9492
JB	0.0084^{**}	0.0028^{**}	0.0046^{**}

** significant at 5% level

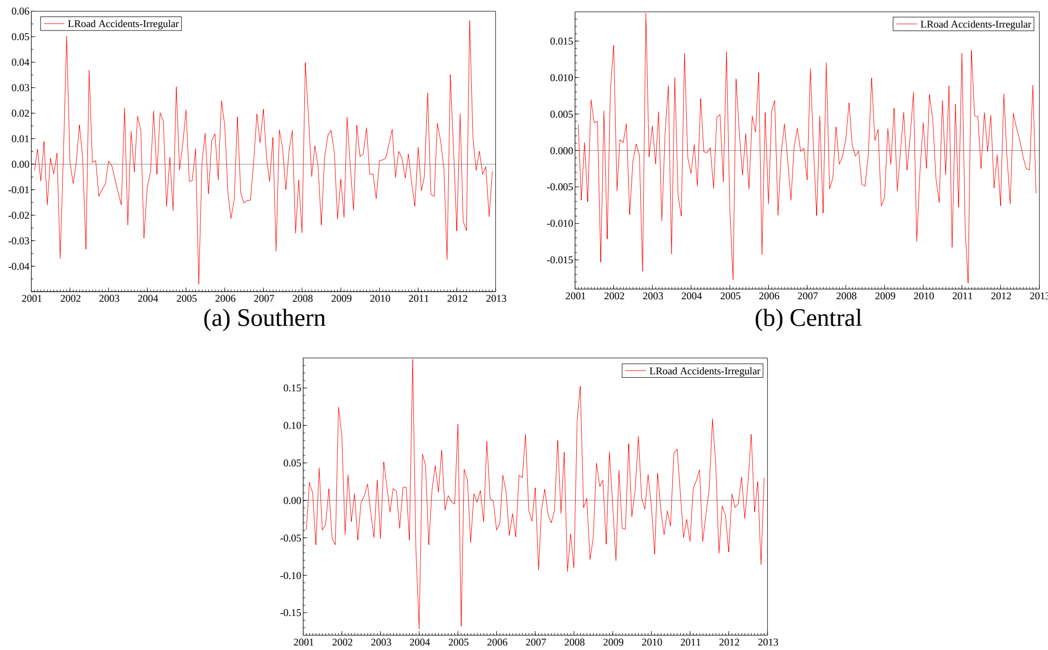
compare the performance of structural time series with time series regression that is one of the most preferable methods, the lag of dependent variable is incorporate as one of the independence variable to overcome the serial correlation. Table 3 shows the estimate of variance disturbance after explanatory variable has been added. The maximum likelihood estimate of level disturbance increase for all region after explanatory variable is added, while the observation disturbance decrease due to inclusion of the variables. It is clear that the inclusion of explanatory variable have reduce the residual (refer Figure 5). However, Southern, Central, and East Coast shows the model do not satisfy the assumption of normality at all. It is may be due to the outliers and level break that does not accounted in this study.

Table 4 shows the estimated road accidents influence based on the structural time series approach. The value in bracket indicates the standard error of the estimate. The model shows that the climate effect has a positive relationship with the increasing number of road accidents in all three regions except the number of rainy day and temperature for Central region, and API for Southern and Central region that show negative relationship with the number of road accidents. In terms of significance of relationship, amount of rainfall only has positive relationship with road accidents in the Central region. The result is quite reasonable as the Central region, which

Table 4: The Estimated Road Accident Influence

Variable	Southern	Central	East Coast
RAIN_F	0.0133	0.0378^{**}	0.0334
	(0.0110)	(0.0106)	(0.0191)
RAIN_D	0.0018	-0.0003	0.0117^{**}
	(0.0014)	(0.0012)	(0.0028)
TEMP	-0.0011	-0.0048	0.0419^{**}
	(0.0070)	(0.0061)	(0.0070)
API	-0.0001	-0.0003	0.0026^{**}
	(0.0002)	(0.0002)	(0.0013)
OILP	0.0001	-0.0003^{*}	0.0001
	(0.0001)	(0.0002)	(0.0002)
CPI_T	-0.0017^{*}	-0.0009	-0.0005
	(0.0008)	(0.0010)	(0.0019)
RTV	0.0173^{**}	-0.0263^{**}	0.0411^{**}
	(0.0090)	(0.0079)	(0.0187)
OPSKP	0.0290^{**}	-0.0061	0.1438^{**}
	(0.0119)	(0.0104)	(0.0268)
BELT	0.0133	0.0058	0.0168
	(0.0110)	(0.0352)	(0.0620)
LRA_1	0.0018	-0.1855^{**}	-0.2562^{**}
	(0.0014)	(0.0767)	(0.0586)

** significant at 5% level, * significant at 10% level



(a) Southern (b) Central

(c) East Coast

Fig. 5: The Irregular Component after Adding Explanatory Variable

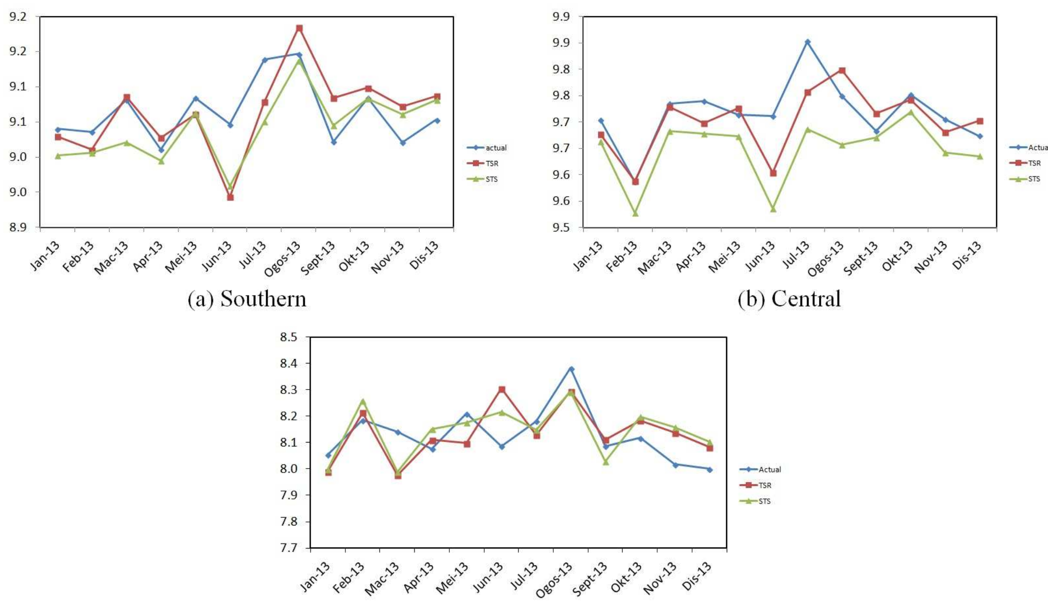
constitutes Selangor and Kuala Lumpur, is among the popular states where flash flood is one of the results of heavy rain. The flood waters that had risen up to the level of the road may result in traffic accidents. Meanwhile, the number of rainy days has significantly affected road accidents in the East Coast region. This may be due to the structure of the road of East Coast region which is near the hillside where a litte amount of rain may result in wet and slippery road condition that will contribute to road traffic accidents. Surprisingly, temperature and API have a significantly positive relationship with the East Coast region. The estimated model for the influences of economic effect in traffic accidents shows that CPI for transport has a negative relationship with the increasing number of road accidents in all three regions and it significantly affects road accidents in the Southern region. This result is also reasonable because, as the price of transportation increases, consumers who buy a new car will be reduced and subsequently the volume of traffic can be reduced and at the same time reduce traffic accidents. In terms of the region, it is reasonable since the region is among the regions that have the highest number of registered vehicles per year after the Central region but with road structures that are less developed compared to that of the Central region's. The crude oil price only has a significant effect with road accidents in the Central region. The estimated results make sense for the Central region as it is a big city and a big one-stop centre. It also agrees with [9] who stated that petrol price appeared to be quite strongly related to many accident series.

Furthermore, the estimated model shows that RTV culture is positively significantly related to the occurence of road accidents in the Southern and East Coast region but negatively related to road accidents in the Central region. Since the variable counts the weight of holiday for festive season, the results are expected since the traffic volume becomes much higher in this season as Malaysians take full advantage to go back to their hometowns and this has contributed to higher number of road accidents for those regions. At the same time, road accidents in the Central region were reduced during the festive season as many residents in this region are outsiders who are just staying in the region for livelihood and will travel to their hometown during festive seasons.

As the implementation of OPSKP operation during festive season, road accidents in all regions are estimated to be increased except for the Central Region, which is estimated to be reduced. However, this implementation only significantly affects the number of road accidents in the Southern and East Coast region. It is unexpected that the results show that the operation failed to reduce traffic accidents. However, the results for the Southern and East Coast region are parrallel with the result of RTV culture since both variables are related to festive seasons. Hence, the estimated model also shows that the rear seat belt implementation has failed to significantly reduce the number of road accidents in all regions that were studied. This may be because Malaysian citizens still lack awareness about the importance of wearing rear seat belt and there is also lack of campaigns from the authorities.

Table 5: Effectiveness of Structural Time Series

Region	RMSE		MAD	
	Time Series Regression	Structural Time Series	Time Series Regression	Structural Time Series
Southern	0.0458	0.0457	0.0370	0.0368
Central	0.0486	0.0851	0.0365	0.0695
East Coast	0.1044	0.0940	0.0870	0.0860



(a) Southern (b) Central

(c) East Coast

Fig. 6: The Forecasting Performance

In terms of seasonal effect, the estimated model does not show much difference as the univariate model in the previous section.

Effectiveness of Structural Time Series

The estimation of road accidents is modelled using the structural time series approach. The effectiveness of the model is tested by comparing their prediction accuracies with the most preferable method which is time series regression. 12 years (2001-2012) or 144 observations of the monthly number of road accidents for each year were used to predict one year ahead. Two loss functions which are root mean square error (RMSE) and mean absolute deviation (MAD) were used to determine the effectiveness of structural time series compared to time series regression as in Table 5

This table indicates that the performance of these two models is approximately similar to each other, but the structural time series shows superiority over time series regression except for the central region. To have a clear picture, Figure 6 illustrates the predicted value of training data sets for the year 2013 for both models for each region. This figure shows that the structural time series model forecast values are closer to the actual observation compared to those of time series regression. This result indicates that by letting the time series component move in random walk a prediction of road accident with better fit is made. The full estimated model is illustrated in appendix A. The estimated model shows that the diagnostic residual for time series regression was not fully satisfied compared to structural time series. Clearly, the result shows that structural time series fits the series well.

4 Conclusion

The study investigates the influences of road accidents occurence in Malaysia by using the structural time series approach. Besides that, the study also determines the effectiveness of structural time series in predicting the occurrence of road accidents. The Malaysia road accidents series of three region consisting of Southern, Central, and East Coast region was considered. Nine possible road accidents contributor for each region were incorporated. The study starts with data fitting followed by investigating the influential factor of road accidents and finally determining the effectiveness of the forecasting performance using structural time series analysis.

The study finds that the pattern of road accidents in each region is not necessarily the same. The significant factor is that related to road accidents also does not necessarily show similar results. It is due to the different structure and topography of the region. Climate effect is related to East Coast region, while economic effect was more related to central region road accidents. In terms of forecasting performance, structural time series and time series regression show approximately similar results. However, the structural time series fit the series better since residual diagnostic for the model has been fully satisfied compared to the time series regression.

The study suggests that policy makers should consider implementing a different approach in ways to overcome the increasing number of road accidents for different regions due to the unique characteristics in the transportation industry and climate effect. Some rural states may produce more accidents during the festive season while urban areas produce more accidents during working days. On the other hand, due to different topography, some states are influenced by adverse climate effect that yields a higher number of road accidents.

This study has a few limitations since it is developed only for the case of Malaysia and it may not reflect the scenario of road accidents in other countries. Besides, the model may not accurately represent the number of road accidents as the data only represents reported cases of road accidents while there might be more unreported cases of accidents. Further investigation may include other relevant variables like other calendar effect variable such as the number of working days and school holiday.

Acknowledgement

We would like to acknowledge support for this project from Universiti Sains Malaysia (USM Short-term grant 304/PMATHS/6313068) and we are grateful to the anonymous referee for a careful checking of the details and for helpful comments that improved this paper.

References

[1] G. Jacobs, A. Aeron-Thomas, A. Astrop, and G. Britain, Estimating global road fatalities, London, 2000
[2] R.U.Radin Sohadi, M.G. Mackay and L.H. Brian, IATSS Research **2**, 91-98, (1995).
[3] E. Bun, Road Traffic Accidents in Nigeria: A Public Health Problem, Short Communication **3(2)**,1-3 (2012)
[4] P. P. Valli, IATSS Research **29**, 57-65, (2010).
[5] N. Nasaruddin, Y. B. Wah, and W. S. Voon, Statistics in Science, Business, and Engineering International Conference, 1-6 (2012).
[6] T.J. Zlatoper, Joournal of Transport Economics and Policy, **18**, 263-273, (1984)
[7] G. A. Ali and C. S. Bakheit, Proceedings of the 30th Southern African Transport Conference, 202-214 (2011)
[8] M. M. Desai and A. Patel, National Conference on recent trends in Engeneering and technology, 1-8 (2011)
[9] P. P. Scott, Acccident Analysis and Prevention, textbf 18, 109-117 (1986)
[10] R.U.Radin Sohadi, M.G. Mackay and L.H. Brian, Accident Analysis and Prevention **28**, 325-332 (1996).
[11] R. Sarani, S. M. R. Sharifah Allyana, M.M.Jamilah, and S.V.Wong, Predicting Malaysian Road Fatalities for Year 2020, Research Report, Malaysia Institute of Road Safety Research, Kuala Lumpur, 2012.
[12] S. Harnen, S.V. Wong, and W. I. W. Hashim, Advances in Transportation Studies an International Journal **8**, 31-40 (2006).
[13] P. Greibe, Accident Analysis and Prevention, **35**, 273-285 (2003).
[14] S. Abusini, International Conference On Mathematical Sciences and Statistics,**1557**, 241-246 (2013)
[15] M. M. Abdul Manan, T. Johnson, and A. Vrhelyi, Safety Science,**60**, 13-20,(2013).
[16] A. Razzaghi, A. Bahrampour, M. R. Baneshi, and F. Zolala, International Journal of Health Policy and Management **1**, 51-55 (2013).
[17] T. Ofori, B. Ackah, and L. Ephraim, International Journal of Research in Environmental Sciences and Technology **2**, 143-149 (2012).
[18] T. H. Law, R. S. R. Umar, S. Zulkaurnain, and S. Kulanthayan, International Journal of Injury Control and Safety Promotion, **12**, 9-21 (2005).
[19] W. F. Wan Yaacob, W. Z. Wan Husin, N. Abd Aziz, and N. I. Nordin, Journal of Applied Science, **11**, 1105-1112 (2011).
[20] R. B. Noland, Accident Analysis and Prevention, **35**, 599-611 (2003).
[21] T. Brijs, D. Karlis, and G. Wets, Accident Analysis and Prevention, **40**, 1180-1190 (2008).
[22] W. F. Wan Yaacob, M. A. Lazim, and Y. B. Wah, Journal of Applied Science, **11**, 1185-1191 (2011).
[23] W. F. Wan Yaacob, M. A. Lazim, and Y. B. Wah, International Conference on Science and Social Research, 960-964 (2010).
[24] W. F. Wan Yaacob, M. A. Lazim, and Y. B. Wah, Proceeding of the World Congress on Engineering, **I** (2012).
[25] W. F. Wan Yaacob, M. A. Lazim, and Y. Bee Wah, Proceeding of the Regional Conference on Statistical Sciences, 176-183 (2010)
[26] V.Dordonat, S. Koopman, M. Ooms, A. Dessertaine and J. Collet, International Journal of Forecasting, **24**, 566-587 (2008)
[27] V.Dordonat, S. Koopman and M. Ooms, Computational Statistics and Data Analysis, **56**, 3134-3152 (2012)
[28] Z. Dilaver and L.C.Hunt, Energy Economics**33**, 426-436 (2011)
[29] H. Song, G. Li, S. F. Witt , G. Athanasopoulos, International Journal of Forecasting, **27**,855-869 (2011)
[30] Shepherd Estimating Price Elasticities of Supply for Cotton: A Structural Time-Series Approach, FAO Commodity and Trade Policy Research Working Paper-21, 2006
[31] G. Wang L. L. Getz, Ecological Modelling, **207**,189-196 (2007).

APPENDIX A: Full Estimation Result

Variable	Southern				Central				East Coast			
	TSR		STS		TSR		STS		TSR		STS	
	coef.	Sig	coef.	Sig	coef.	Sig	coef.	Sig	coef.	Sig	coef.	Sig
Constant	5.6185	0.0000	9.8201	0.0000	1.2203	0.0256	11.5844	0.0000	4.8740	0.0000	8.2857	0.0000
Trend	0.0029	0.0000	0.0047	0.0003								
RAIN_F	0.0196	0.1410	0.0133	0.2268	0.0479	0.0029	0.0378	0.0005	0.0351	0.1741	0.0334	0.0830
RAIN_D	0.0003	0.8257	0.0018	0.1924	0.0005	0.7687	-0.0003	0.7882	0.0111	0.0023	0.0117	0.0001
TEMP	0.0051	0.4552	-0.0011	0.8770	0.0019	0.7717	-0.0048	0.4366	0.0357	0.0000	0.0419	0.0000
API	-0.0004	0.0522	-0.0001	0.4418	0.0001	0.7180	-0.0003	0.2712	0.0044	0.0004	0.0026	0.0415
OILP	3.7×10^{-5}	0.6428	0.0001	0.5600	0.0002	0.7237	-0.0009	0.3644	0.0007	0.0000	0.0001	0.7078
CPI_T	-0.0006	0.2745	-0.0017	0.0399	0.0002	0.0440	-0.0003	0.0792	0.0044	0.0031	-0.0005	0.8003
RTV	0.0164	0.1497	0.0173	0.0553	-0.0396	0.0035	-0.0263	0.0012	0.0450	0.0799	0.0411	0.0296
OPSKP	0.0321	0.0300	0.0290	0.0166	0.0057	0.7389	-0.0061	0.5620	0.1487	0.0001	0.1438	0.0000
BELT	0.0236	0.1182	0.0026	0.9269	0.0352	0.0231	0.0058	0.8687	0.2132	0.0000	0.0168	0.7871
Seasonal1	-0.0284	0.0606	-0.0164	0.0165	0.0422	0.0159	-0.0243	0.0014				
Seasonal2	-0.0997	0.0000	-0.0241	0.0012	-0.0699	0.0005	-0.0390	0.0000				
Seasonal3	0.0299	0.1193	-0.0014	0.8038	0.1427	0.0000	-0.0124	0.0040				
Seasonal4	-0.0395	0.0315	-0.0043	0.5573	-0.0250	0.2273	-0.0143	0.0113				
Seasonal5	0.0122	0.4871	0.0154	0.0058	0.0100	0.6035	0.0196	0.0000				
Seasonal6	-0.0354	0.0299	0.0019	0.7491	-0.0087	0.6374	-0.0056	0.1302				
Seasonal7	-0.0252	0.0983	-0.0006	0.9121	0.0922	0.0000	-0.0100	0.0015				
Seasonal8	0.0347	0.0198	0.0022	0.6659	0.0441	0.0104	0.0166	0.0000				
Seasonal9	-0.0206	0.1716	0.0089	0.0957	-0.0088	0.05941	-0.0043	0.1441				
Seasonal10	0.0260	0.1351	0.0260	0.0000	0.0688	0.0003	0.0271	0.0000				
Seasonal11	-0.0419	0.0157	-0.0055	0.2062	-0.0212	0.2943	-0.0080	0.0001				
LRA _1	0.3051	0.0001	-0.0804	0.2932	0.8187	0.0000	-0.1855	0.0170	0.0572	0.4194	-0.2562	0.0000
Diagnostic Test	TSR		STS		TSR		STS		TSR		STS	
LB	0.0202**		0.2392		0.1025		0.3273		0.0198**		0.1648	
GQ	0.7152		0.7648		0.0913*		0.5624		0.0485**		0.9492	
JB	0.0265**		0.0084**		0.8035		0.0028**		0.0791*		0.0046**	

TSR=Time Series Regression, STS=Structural Time Series

** significant at 5% level, * significant at 10% level

[32] J. Knape, N. Jonzen, M. Skold, and L. Sokolov, Multivariate state space modelling of bird migration count data. In: Thomson, D.L., Cooch, E.G. and Conroy, M.J. (eds.) Modeling Demographic Processes in Marked Populations. Environmental and Ecological Statistics 3.Springer-Science + Business Media, LLC: 5979 (2009)

[33] A.R. Lawson,B. Bidisha, B. Broderick, Atmospheric Environment, **45**, 4719-4727 (2011)

[34] R. Zizza, Journal of Applied Statistics,**33**, 481-495, (2006).

[35] S. Krieg J.A. van den Brakel, Computational Statistics and Data Analysis, **56**, 2918-2933 (2012)

[36] V. A. Muscatelli and P. Tirelli, Applied Economics, **33**, 1083-1088 (2001)

[37] A. Harvey and J. Durbin, Journal of the Royal Statistical Society, **149**, 187-227 (1986).

[38] S. Lassare, Accident Analysis and Prevention **33**, 743-751 (2001)

[39] E. Hermans, G. Wets, F. Van den Bossche, Journal of Transport Statistics, **9**, 63-76 (2006)

[40] P.A. Scuffham, Applied Economics, **35**, 179-188 (2003).

[41] R. Bergel-Hayat, M. Debbarh, C.Antoniou, G.Yannis, Accident Analysis and Prevention, **60**, 456-465 (2013).

[42] R. Bergel- Hayat and J. Zukowska, Structural Time Series Modeling of the Number of Fatalities in Poland in Relation to Economic Factors, in Traffic Safety (eds G. Yannis and S. Cohen), John Wiley & Sons, 2016

[43] C. Antoniou, E. Papadimitriou, and G. Yannis, Traffic Injuries and Prevention,**15**, 598-605 (2014)

[44] C. Antoniou and G. Yannis, Accident Analysis and Prevention, **60**, 268-276 (2013).

[45] S. J. Koopman, A.Harvey, J. A. Doornik, and N. Shephard, Structural Time Series Analyser Modeller and Predictor STAMP 7. London: Timberlake Consultants Ltd, 2006.

[46] N. Shuja, M. A. Lazim, and Y. B. Wah, "Moving Holiday Effects Adjustment for Malaysian Economic Time Series," Department of Statistics. 2007.

[47] H. Akaike, Proceeding of the 2nd International Symposium on Information Theory, (1973)

[48] H. Akaike, IEEE Transactions on Automatic Control **19**, 716-723 (1974)

[49] J. Commandeur and S. Koopman, An introduction to state space time series analysis. United State: Oxford University Press, 2007.

Noor Wahida Md Junus is PhD student in School of Mathematical Sciences University Science Malaysia Her research interests are in the areas of Applied Statistics and focusing on Time Series analysis and Statistical Modeling.

Mohd Tahir Ismail is Associate Professor at School of Mathematical Sciences, Universiti Science Malaysia. His research area is on financial time series.Particularly he is keen in the modeling of structural break by using wavelet in time series analysis. He also analyses the economics issues. As of now, he has published more than 100 publications in reviewed journals and proceedings (some of them are listed in ISI, Scopus, Zentralblatt and other indices).

Zainudin Arsad is Senior Lecturer at School of Mathematical Sciences, Universiti Science Malaysia. He recieved PhD degree from Heriot-University, Edinburgh, Scotland. His research interests are in the areas of calendar effects in stocks prices, application of Kalman filter in finance issues, spillover effects and market integration, stock prices-macroeconomic variable linkages and tourism economic. He is one of the members of Universiti Science Malaysia (USM) Tourism research cluster.

Rosmanjawati Abdul Rahman is a senior lecturer in School of Mathematical Sciences, Universiti Sains Malaysia, Pulau Pinang, where she has been since 2008. Her research area is in Applied Statistics, applied to economics and financial data analysis. Currently, her research is in applying the Finite Mixture Model into the time series data.